



مبانی آمار در علوم رفتاری

a deeper understanding of retirement decisions, retirement processes, and retirement policies. I argue that, in order to make
add rigorous qualitative studies to their quantitative studies, scholars should consider the importance of the
moving elements and the role of the individual in the decision-making process.

نویسنده: دیوید سی هاوول
مترجم: دکتر محمود بحرانی
عضو هیات علمی دانشگاه شیراز

REDMI NOTE 8T
AI QUAD CAMERA



مبانی آمار در علوم رفتاری

مترجم: دکتر محمود بحرانی

دانشگاه شیراز

سرشناسه	هاول، دیوید Howell, David C.
عنوان و نام پدیدآور	مبانی آمار در علوم رفتاری / دیوید سی هوول؛ مترجم محمود بحرانی.
مشخصات نشر	تهران: انتشارات ماهواره، ۱۳۹۶.
مشخصات ظاهری	۴۳۴ ص.
شابک	978-600-459-218-5
وضعیت فهرست نویسی	فیا
یادداشت	عنوان اصلی: Fundamental statistics for the behavioral sciences, 2014.
یادداشت	کتاب حاضر در سال ۱۳۷۸ تحت عنوان "اصول آمار برای علوم رفتاری" با ترجمه کیمیا شمیرانی توسط نشر چشمه منتشر شده است.
عنوان دیگر	اصول آمار برای علوم رفتاری.
موضوع	علوم اجتماعی -- روش های آماری، Social sciences -- Statistical methods
موضوع	روان سنجی، Psychometrics
شناسه افزوده	بحرانی، محمود، ۱۳۴۲ -، مترجم
رده بندی کنگره	۱۳۹۶ الف ۶ ۲۴ هـ ۲۹ HA
رده بندی دیویی	۵۱۹/۵
شماره کتابشناسی ملی	۴۸۳۸۸۷۱

عنوان: مبانی آمار در علوم رفتاری، Fundamental Statistics for the behavioral sciences

تألیف: David C. Howell، دیوید سی هوول

مترجم: دکتر محمود بحرانی

ناشر: تهران، ماهواره

نوبت چاپ: اول، ۱۳۹۶

شابک: 978-600-459-218-5

قیمت: ۲۵۰۰۰ تومان

تیراژ: ۵۰۰ جلد

دفتر مرکزی انتشارات: میدان فاطمی، خ چهلستون، پ ۵۰، و ۴، ۰۲۱۸۸۳۹۲۲۶۷، ۰۹۱۹۰۶۲۱۴۱۸ مسئولیت صحت مطالب و حقوق مادی و معنوی کتاب بر عهده مولف می باشد. کلیه حقوق برای مولف محفوظ است.

فهرست کتاب

صفحه	عنوان
۱	فصل ۱: مقدمه
۱۰	فصل ۲: مفاهیم اساسی آمار
۲۳	فصل ۳: نمایش داده‌ها
۴۳	فصل ۴: رتبه‌های درصدی
۵۴	فصل ۵: شاخص‌های گرانش مرکزی
۶۰	فصل ۶: شاخص‌های تغییرپذیری
۸۵	فصل ۷: توزیع بهنجار
۱۰۴	فصل ۸: مفاهیم اساسی احتمالات
۱۱۷	فصل ۹: همبستگی
۱۴۴	فصل ۱۰: رگرسیون
۱۶۲	فصل ۱۱: توزیع‌های نمونه‌گیری و فرضیه‌آزمایی
۱۸۴	فصل ۱۲: آزمون‌های فرضیه برای میانگین‌های یک نمونه
۲۰۵	فصل ۱۳: آزمون‌های فرضیه برای میانگین‌ها - دو میانگین وابسته
۲۱۵	فصل ۱۴: آزمون‌های فرضیه برای میانگین‌ها - دو میانگین مستقل
۲۳۰	فصل ۱۵: توان
۲۴۸	فصل ۱۶: تحلیل واریانس یک طرفه
۲۸۰	فصل ۱۷: تحلیل واریانس عاملی
۳۰۵	فصل ۱۸: طرح‌های اندازه‌گیری مکرر
۳۱۷	فصل ۱۹: مجذور خی
۳۲۸	فصل ۲۰: آزمون‌های آماری غیرپارامتریک و توزیع - آزاد
۳۵۰	فصل ۲۱: انتخاب تحلیل مناسب
۳۵۶	پیوست ۱: مرور عملیات ریاضی

۳۶۱	پیوست ۲: نمادها و نشانه‌ها
۳۶۴	پیوست ۳: جدول‌ها
۳۸۴	پیوست ۴: پاسخ تمرین‌های انتخابی
۴۰۷	فرمول‌های اصلی آمار

فصل ۱

مقدمه

در گذشته وقتی در مهمانی‌ها و موقعیت‌های اجتماعی دیگر از من پرسیده می‌شد چه کاره هستی، می‌گفتم روانشناس هستم. بعد از چندین سال مقاومت در برابر اظهار نظرهای عجیب و غریبی که این موضوع ایجاد می‌کرد، سرانجام روش خود را تغییر دادم و تصمیم گرفتم که بگویم آمار درس می‌دهم، - پاسخی که کاملاً هم درست بود. این پاسخ یک مشکل را حل کرد- مردم دیگر به من بدبینانه نگاه نمی‌کردند، - اما مشکل دیگری را دامن زد. اکنون آنها یا مرا یک دایره المعارف سیار از "اطلاعات بی‌فایده" می‌پندارند (همچون میزان تن فولاد بارگیری شده از بازار مشترک به کشورهای عضو پیمان ورشو) یا به من می‌گویند که چطور توانستند از گرفتن درس آمار فرار کنند (چون "هیچوقت ریاضیات من خوب نبود"- یک دلیل واقعا بی ربط).

مشکل اساسی این است که بیشتر مردم نمی‌دانند آمار در باره چه چیز هست و در باره چه چیز نیست. از این رو مجبوریم با بعضی تعریف‌های اساسی شروع کنیم. واژه آمار^۱ اصطلاحی است که برای معانی مختلفی به کار می‌رود، در اینجا می‌توان دست‌کم به سه شیوه اشاره کرد. آمار در یک معنا چنان که در عنوان این کتاب آمده است به مجموعه قواعد و شیوه‌هایی برای کاهش حجم زیاد داده‌ها گفته می‌شود که امکان نتیجه‌گیری از آنها را فراهم می‌سازند. این کتاب در اصل در باره این موضوع است. یک معنای دیگر و خیلی معمول واژه آمار را می‌توان با این مثال توضیح داد که "آمارها نشان می‌دهند که شمار افراد مشمول بیمه بیکاری برای سومین ماه متوالی کاهش یافته است." آمار در این معنا، به جای کلمه داده‌ها^۲ به کار رفته است که خود واژه داده‌ها خیلی بهتر است. ما آمار را به این معنا به کار نخواهیم برد.

1. statistics
2. data

سومین معنای عمومی این واژه اشاره به نتایج بعضی دستکاری‌های حسابی و جبری داده‌ها دارد. بنابر این میانگین (معدل) یک مجموعه نمرات یک آماره است. این کاربرد کاملاً مجاز است و در این کتاب بطور مکرر مورد استفاده قرار خواهد گرفت.

بنابراین ما دو استفاده مناسب از این واژه داریم: (۱) یک مجموعه قواعد و شیوه‌ها و (۲) نتایج به کارگیری آن قواعد و شیوه‌ها برای نمونه‌هایی از داده‌ها. خواننده می‌تواند از متن دریابد کدام یک از این دو معنی مد نظر است.

۱-۱ آمار توصیفی و استنباطی

شیوه‌های آماری را بطور کلی به دو بخش توصیفی^۱ و استنباطی^۲ می‌توان تقسیم کرد. چند فصل نخست این کتاب مربوط به آمار توصیفی و فصل‌های دیگر بیشتر پیرامون آمار استنباطی هستند.

۱-۱-۱ آمار توصیفی

هرگاه هدف شما در اصل توصیف یک مجموعه از داده‌ها باشد شما آمار توصیفی را به کار می‌برید. بررسی نمرات رژیم غذایی از یک مقیاس محدودیت خوردن، نرخ جرائم گزارش شده از سوی دادگستری، و جمع‌بندی از اطلاعات معین مربوط به نمرات امتحان یک درس، همگی نمونه‌هایی از آمار توصیفی هستند.

متخصصان آمار توصیفی شیوه‌های زیادی را برای این کار دارند، البته بسیاری از این شیوه‌ها فراتر از چارچوب این کتاب است. مهمترین شیوه‌ها و مقیاس‌ها مانند روش‌های ترسیم نموداری داده‌ها، تحلیل اکتشافی داده‌ها، میانگین‌ها، انحراف معیار و درصد، در فصل ۳ تا ۶ توضیح داده خواهند شد، چون اینها برای فهم آمار استنباطی مطالب اساسی هستند.

۱-۱-۲ آمار استنباطی

همه ما گاهی از این احساس گناه کرده‌ایم که چرا بر اساس داده‌های خیلی محدود تعمیم‌های غیرمنطقی صورت داده‌ایم. مثلاً اگر شنیده یا خوانده باشید که افراد قدبلند خوش اندام‌تر از افراد قدکوتاه هستند، ممکن است با آن موافق باشید چون شما یک بار یک هم‌اتاقی قد بلند داشتید که به طرز خاصی خوش اندام بود. معمولاً می‌خواهید فرد قد دراز کم‌فهمی که برای پوشیدن شلوارش مشکل دارد را فراموش کنید. همین‌طور، مردی که می‌گوید مهارت‌های حرکتی در دختران زودتر از پسران

1. descriptive
2. inferential

رشد می‌کنند چون دختر او در ده ماهگی راه افتاده اما پسرش تا ۱۴ ماهگی نتوانسته راه برود از اشتباهی مشابه- تعمیم از مشاهدات موردی (یا اندک)- شرمسار می‌شود.

مشاهدات موردی زمانی می‌تواند خوب باشد که موضوع مورد مطالعه ما تغییرپذیری کمی داشته باشد. اگر بخواهیم بدانیم گاوها چند پا دارند می‌توانیم یک گاو را پیدا کنیم و پاهایش را بشماریم. نیازی به یک گله نیست، یکی بس است. اما اگر بخواهیم چیزی را اندازه بگیریم که از یک مورد به مورد دیگر تغییر می‌کند (مثل طول دم یا سن زایمان گاوها)، ما نمی‌توانیم به مشاهدات موردی بسنده کنیم. اگر بخواهیم بدانیم دختران عموما در چه سنی راه می‌افتند باید به بیش از یک دختر نگاه کنید، اما احتمالا همه دختران دنیا را نمی‌توانیم مورد مشاهده قرار دهیم. باید چیزی بین این دو کار انجام دهیم.

همین مطلب در مورد اثرات یک روش تازه برای درمان اعتیاد به الکل هم صادق است. اینکه این درمان باعث شود یک آزمودنی ترک اعتیاد کند خوب است، اما مسلما این واقعیت استفاده گسترده از این شیوه درمان را بیش از این توجیه نمی‌کند که برای اینکه مشتری را برای یک رژیم غذایی جلب کنید جار بزنیم "من در شش ماه ۱۸۲ پوند وزن کم کردم و از هر دقیقه آن لذت بردم". از سوی دیگر ما نمی‌توانیم پیش از آن که در این مورد نتیجه‌گیری‌هایی داشته باشیم، هر فرد معتاد به الکل یا دارای اضافه وزن را درمان کنیم. اینجا هم ما بایستی حدود میانی را پیدا کنیم. این همان جایی است که آمار استنباطی وارد می‌شود.

بیشتر کارهای آماری در روانشناسی و علوم اجتماعی با آمار استنباطی سروکار دارند. برای توضیح این موضوع بایستی مفاهیم جمعیت‌ها و نمونه‌ها را تعریف کنیم، زیرا حوزه آمار استنباطی با کاربرد نمونه‌ها برای رسیدن به استنباطی از جمعیت سروکار دارد.

۱-۱-۳- جمعیت‌ها، نمونه‌ها، پارامترها و آمارها

یک جمعیت^۱ را می‌توان به عنوان مجموعه کلی پدیده‌هایی که مورد نظر شما هستند تعریف کرد (مانند نمره‌های دانش‌آموزان متوسطه آمریکایی در یک مقیاس خودارزشمندی، حراج‌های انحصاری تمام اقلام تولیدی در این کشور در ۲۰ سال گذشته، و نظایر آن).

بنابراین اگر مثلا ما بخواهیم از نمرات حساسیت اجتماعی همه نوجوانان آمریکایی اطلاع پیدا کنیم، در این صورت مجموعه همه این نمرات حساسیت اجتماعی نوجوانان آمریکایی یک جمعیت را تشکیل می‌دهند- در این حال جمعیتی بالغ بر ۵۰ میلیون عضو. از سوی دیگر اگر تنها علاقه‌مند به مجموعه نمرات حساسیت اجتماعی کلاس‌های پایه دوم فایرفوکس، ورمونت (شهری با حدود ۱۸۰۰ نفر جمعیت) این جمعیت مورد نظر ممکن است حدود ۳۰ عضو داشته باشد و می‌توان کل آن را با راحتی و به سرعت بدست آورد. در نهایت مجموعه‌ای از نتایج ریختن یک تاس را در نظر بگیرید. بدیهی است، بعد از ریختن

1. population

یک تاس شما می‌توانید آن را دو باره تکرار کنید، و فرض کنید که این کار پایانی ندارد و می‌توان بارها آن را انجام داد، که بی‌نهایت جمعیت بزرگ از نتایج حاصل شود (ریاضیدانان واژه بی‌شمار را ترجیح می‌دهند، اما واژه بی‌نهایت به کار می‌رود).

منظور این است که جمعیت می‌تواند از یک مجموعه نسبتاً کوچک اعداد که به راحتی قابل جمع کردن هستند، تا یک مجموعه بی‌شمار و بزرگ اعداد که هرگز نتوان آنها را کاملاً شمرد، گسترده باشد. متأسفانه، جمعیت‌های مورد نظر ما کاملاً بزرگ هستند. در نتیجه عملاً بندرت یا هرگز نمی‌توانیم کل جمعیت را جمع‌آوری کنیم. به جای آن مجبور خواهیم بود که یک **نمونه**^۱ را از آن بیرون بکشیم و مورد استفاده قرار دهیم تا در باره ویژگی‌های جمعیت چیزی را استنباط کنیم.

وقتی ما نمونه‌ای از مشاهدات بیرون می‌کشیم، معمولاً اندازه‌های عددی (مثل میانگین) را محاسبه می‌کنیم که داده‌ها را در نمونه خلاصه می‌کنند. وقتی چنین اندازه‌هایی براساس نمونه باشند، به آنها **آماره**^۲ گفته می‌شود. همین اندازه‌ها را که از مطالعه کل جمعیت بدست آمده باشند (مثل میانگین جمعیت) **پارامتر**^۳ گویند. یک هدف اصلی شیوه‌های آمار استنباطی این است که از آماره‌ها (ویژگی‌های نمونه‌ای) به استنتاج‌هایی در باره پارامترها (ویژگی‌های جمعیت) برسیم.

اگر فرض کنیم که نمونه یک نمونه واقعا تصادفی باشد، یعنی هر عضو و همه اعضای جمعیت از شانس برابری برای قرار گرفتن در نمونه برخوردار باشند، در این صورت نه تنها ما می‌توانیم پارامترهای جمعیت را برآورد کنیم بلکه دلایل خوبی هم برای درستی برآوردهای خود خواهیم داشت. در صورتی که نمونه یک نمونه تصادفی نباشد برآوردهای ما ممکن است بی‌معنا باشند، چون نمونه به درستی نمودی از کل جمعیت نیست.

اجازه دهید یک نکته را روشن کنیم که بسیاری را به اشتباه می‌اندازد. مشکل این است که نمونه یک شخص ممکن است جمعیت شخص دیگری باشد. مثلاً اگر بخواهیم این موضوع را بررسی کنیم که آیا این کتاب برای تدریس ابزار مناسبی است، نمرات یک کلاس در یک امتحان را می‌توان به عنوان یک نمونه‌ی هر چند غیر تصادفی از کل جمعیت نمرات دانشجویانی که احتمالاً یا واقعا از کتاب استفاده می‌کنند در نظر گرفت. ازسوی دیگر، استاد کلاس که فقط روی دانشجویان خودش تمرکز دارد، ممکن است همین مجموعه از نمرات را به عنوان یک جمعیت در نظر بگیرد. در مقابل یک فرد علاقه‌مند به تدریس آمار ممکن است جمعیت ما را (نمرات کسانی از این کتاب استفاده کرده‌اند) به عنوان جمعیت غیرتصادفی از یک جمعیت بزرگتر (نمرات همه کسانی که از یک کتاب آماری استفاده کرده‌اند) در نظر بگیرد. به این ترتیب تعریف یک جمعیت وابسته به آن چیزی است که شما در تحقیق خود مد نظر قرار

1. sample
2. statistics
3. parameter

می‌دهید. همچنین توجه کنید که وقتی ما از جمعیت سخن می‌گوییم، جمعیت نمرات مد نظر است نه جمعیت افراد. این موضوع را بعداً توضیح خواهیم داد.

این واقعیت که ما نمونه‌های غیر تصادفی را به کار بردیم نباید خواننده را دچار این اشتباه کند که نمونه‌گیری تصادفی مهم نیست. برعکس این سنگ زیربنای بیشتر شیوه‌های آماری است. در واقع جمعیت مربوط به آن را می‌توان به عنوان مجموعه اعدادی که نمونه به طور تصادفی از آنها بیرون کشیده شده است تعریف کرد.

برآورد^۱ پیش از این ما آمار استنباطی را به عنوان شاخه‌ای از آمار تعریف کردیم که با استنباط ویژگی‌های جمعیت از روی ویژگی‌های نمونه‌ها سروکار دارد. این بیان به خودی خود کافی نیست چنان‌که خواننده را به این گمان می‌رساند که همه‌ی آنچه مورد نظر ماست همانا تعیین پارامترهای جامعه از قبیل متوسط قد مردان یا متوسط سرعت دویدن دانش‌آموزان پایه دوم ابتدائی است. البته زمان‌هایی وجود دارد که توجه ما معطوف به پارامترهای واقعی جمعیت است. بطور مثال ما اغلب می‌خوانیم و ممکن است شگفت‌زده شویم از ساعت‌های زیادی که کودکان طی روز در برابر تلویزیون می‌گذرانند. اما اگر این همه‌ی چیزی باشد که در آمار استنباطی صورت می‌گیرد، این یک موضوع جدا بسیار تأسفانگیز است و مردم حق دارند به کسی که می‌پذیرد آمار تدریس کند با تعجب نگاه کنند. هدف آمار استنباطی فراتر از این است.

مشکلی که بسیاری دانشجویان با آمار دارند این است که خیلی دیر متوجه این می‌شوند که وقتی ما سعی می‌کنیم میانگین سرعت خواندن بچه‌های سال سوم ابتدائی که تحت یک روش خاص آموزش دیده‌اند را برآورد کنیم، ما معمولاً زیاد علاقه‌مند به این نیستیم که میانگین چه اندازه است. ما فقط می‌خواهیم بدانیم این از بعضی میانگین‌های دیگر بزرگتر یا کوچکتر است. مثلاً از میانگین سرعت خواندن بچه‌های کلاس سوم که با روش دیگری آموزش دیده‌اند. بنابراین در موارد بسیاری آمار استنباطی یک ابزار برای برآورد پارامترهای دو یا چند جمعیت است، به ویژه برای دریافتن این که آیا این پارامترها با هم تفاوت دارند یا نه. این نشان می‌دهد که چرا فردی تحقیقی را در باره سرعت دویدن موش‌های آزمایشگاهی تحت بعضی شرایط تقویت رفتار انجام می‌دهد. بدیهی است که این یک موضوع مهم ملی نیست که یک موش با چه سرعتی می‌تواند بدود. اما ممکن است در سطح پایین‌تری مهم باشد که آیا موش‌هایی که تحت یک برنامه تقویتی متفاوت آموزش دیده‌اند تندتر خواهند دوید.

۱-۲ انتخاب از بین شیوه‌های آماری

همان‌گونه که پیش‌تر گفته شد برای انتخاب شیوه آماری، یکی از وجوه مهم تمایز، فرق بین آمار توصیفی و استنباطی است. بخش نخست این کتاب به آمار توصیفی می‌پردازد چون لازم است یک مجموعه داده‌ها ابتدا توصیف شوند، قبل از اینکه آنها را برای برآورد به کار ببریم. زمانی که ما وارد آمار استنباطی شدیم به‌هرحال نیاز به چند تمایز دیگر داریم که در انتخاب شیوه آماری مناسب به ما کمک می‌کنند. در ابتدای این کتاب می‌توانید شکلی که به عنوان نمودار درختی معروف است را مشاهده می‌کنید که راهنمای انتخاب شیوه‌های آماری ارائه شده در این کتاب می‌باشد. این درخت تصمیم‌گیری نه تنها یک بیان کلی از سازمان این کتاب است بلکه، همچنین برخی موضوعات اساسی را هم مورد اشاره قرار می‌دهد که در ابتدا باید آنها را توضیح دهیم. در باره این موضوعات در نظر داشته باشید که در اینجا نمی‌خواهیم به این بپردازیم که کدام آزمون آماری برای چه هدفی به کار می‌رود. این را بعداً توضیح خواهیم داد. ما ابتدا به انواع سؤال‌هایی خواهیم پرداخت که وقتی یک کار آماری با داده‌ها می‌کنیم دخالت دارند- خواه در باره آمار توصیفی صحبت کنیم یا آمار استنباطی. این موضوعات در بالای درخت تصمیم‌گیری قید شده‌اند. سه مورد اول آنها در اینجا به طور فشرده مورد بحث قرار می‌گیرند، بقیه به زمان مناسب دیگری واگذار می‌شوند.

۱-۱-۲-۱ انواع داده‌ها

داده‌های عددی عموماً دو نوع هستند، که از آنها به عنوان داده‌های اندازه‌ای و مقوله‌ای یاد می‌شود. منظور از داده‌های اندازه‌ای^۱ (گاهی داده‌های کمی^۲ هم خوانده می‌شوند) داده‌های حاصل از هر نوع اندازه‌گیری است- بطور مثال، یک نمره در یک آزمون، وزن یک شخص، سرعتی که یک نفر می‌تواند این صفحه را بخواند، نمره یک فرد در یک مقیاس اقتدارگرایی و غیره. در همه این موارد نوعی ابزار (در اطلاق کلی آن) برای اندازه‌گیری چیزی به کار برده شده است.

در مقابل داده‌های مقوله‌ای^۳ (که به عنوان داده‌های فراوانی^۴ یا شمارشی^۵ هم شناخته می‌شوند) شامل عباراتی از این قبیل است "یک صد و پنجاه و هشت نفر از افراد شکلات بادامی قلمی را دوست داشتند و ۲۶ نفر نوع پهن را ترجیح می‌دهند" یا "برای برنامه جدید ۲۳۸ رأی موافق و ۱۱۸ رأی مخالف وجود داشت." در اینجا ما چیزها را می‌شماریم، و داده‌های ما شامل کل یا فراوانی‌های هر مقوله است (نام مقوله‌ای به همین خاطر است). چند صد نفر عضو ممکن است به یک برنامه جدید رأی دهند،

-
1. measurement data
 2. quantitative data
 3. categorical data
 4. frequency data
 5. count data

اما نتایج (داده‌ها) شامل فقط دو عدد خواهد بود- عدد رأی‌های موافق و عدد رأی‌های مخالف آن. با داده‌های اندازه‌ای ما ممکن است دور شکم صدها نفر را اندازه بگیریم. در اینجا ما به صدها عدد برسیم- برای هر نفر یک عدد. گاهی می‌توان یک متغیر را برای ایجاد داده‌های اندازه‌ای یا مقوله‌ای اندازه‌گیری کرد. اگر ما با قد هر فرد بر حسب سانتیمتر کار داریم، با داده‌های اندازه‌ای سر و کار داریم. اما اگر بطور ساده افراد را به قدبلند، متوسط و کوتاه تقسیم کنیم ما با داده‌های مقوله‌ای کار داریم. تشخیص بین داده‌های اندازه‌گیری و مقوله‌بندی مانند خیلی از تمایزها می‌تواند طول و تفضیل پیدا کند ولی در عمل همواره تصمیم‌گیری تقریباً روشن است.

با این دو نوع داده‌ها به دو روش کاملاً متفاوت رفتار می‌شود. در فصل ۱۹ داده‌های مقوله‌ای را بررسی می‌کنیم تا ببینیم چگونه می‌توان تعیین کرد آیا مثلاً بین ترجیح دو نوع شکلات تفاوت پایایی وجود دارد یا نه. در فصل‌های ۹ و ۱۰، ۱۲ تا ۱۴، ۱۶ تا ۱۸ و ۲۰ ما به طور اختصاصی به داده‌های اندازه‌ای خواهیم پرداخت. در استفاده از داده‌های اندازه‌ای تمایز دیگری هم قائل خواهیم شد، نه از لحاظ نوع داده‌ها بلکه از این لحاظ که تفاوت بین گروه‌های آزمودنی‌ها بررسی می‌شود یا اینکه روابط بین متغیرها را مطالعه می‌کنیم.

۲-۱-۲- تفاوت‌ها در مقابل رابطه‌ها

بیشتر پرسش‌های آماری بطور کلی در دو بخش تفاوت‌ها و رابطه‌ها قرار می‌گیرند اگرچه آنها همپوشی هم دارند. برای مثال یک محقق ممکن است در اصل علاقه‌مند باشد که ببیند آیا بین افراد سیگاری و غیرسیگاری از لحاظ عملکردشان در یک آزمون خاص تفاوتی وجود دارد یا نه. محقق دیگری ممکن است بخواهد بداند که آیا بین تعداد نخ‌های سیگار کشیده شده در روز و نمره همان آزمون رابطه‌ای وجود دارد یا نه. هر چند این دو پرسش آشکاراً همپوشی دارند، با آنها به دو روشی که به نظر می‌آید کاملاً با هم متفاوت هستند عمل می‌شود. فصل‌های ۱۲ تا ۱۴ و ۱۶ تا ۱۸ به این موضوع می‌پردازند، که تفاوت بین دو یا چند گروه مد نظر قرار می‌گیرد، در حالی که فصل‌های ۹ و ۱۰ با مواردی سر و کار دارند که در آنها ما روابط بین دو یا چند متغیر را بررسی می‌کنیم. این شیوه‌های آماری به ظاهر متفاوت، به یک شیوه زیربنایی یکسان بر می‌گردند، هر چند پرسش‌های متفاوتی را مطرح می‌کنند و صورت‌بندی پاسخ‌های آنها فرق دارد.

۲-۱-۳- تعداد گروه‌ها یا متغیرها

همان‌گونه که در فصل‌های بعد خواهیم دید، یک تمایز آشکار بین شیوه‌های آماری مربوط به تعداد گروه‌ها و تعداد متغیرهایی است که این شیوه‌ها روی آنها اجرا می‌شود. بنابراین مثلاً خواهیم دید که چیزی که به عنوان آزمون t مستقل از آن یاد می‌شود محدود به مواردی است که داده‌های مربوط به دو

گروه از آزمودنی‌ها مقایسه می‌شوند. ازسوی دیگر، تحلیل واریانس برای هر تعداد گروه، نه فقط دو گروه قابل استفاده است. تمایز سوم در نمودار درختی ما، مربوط به تعداد گروه‌ها یا متغیرهای درگیر است. سه تمایزی که از آنها یاد کردیم (نوع داده‌ها، تفاوت‌ها در مقابل رابطه‌ها، و تعداد گروه‌ها یا متغیرها) در شیوه نگاه ما به داده‌ها و روش‌های آماری، یک امر اساسی است که در تفسیر آن داده‌ها به ما کمک می‌کند. معیار دیگری که بعضی کتاب‌ها برای طرح انواع آزمون‌های آماری و توضیح و دستکاری داده‌ها به کار می‌برند شامل مقیاس سنجشی به کار رفته که در داده‌ها است. این موضوع را ما در فصل بعد توضیح خواهیم داد چون مفهوم مهمی را بازگو می‌کند که هر دانشجویی باید با آن آشنا باشد.

۳-۱ خلاصه

در این فصل ما دو شاخه اصلی آمار (آمار توصیفی و آمار استنباطی) را بیان کردیم. سپس به تفاوت بین جمعیت و نمونه پرداختیم و مفهوم نمونه تصادفی را تعریف کردیم. در آخر هم به چند بعد که شیوه‌های مختلف آماری در طول آنها از یکدیگر متمایز می‌شوند اشاره کردیم. بعضی از مطالب مهم فصل عبارتند از:

○ جمعیت	○ نمونه تصادفی
○ نمونه	○ درخت تصمیم‌گیری
○ آمارها	○ داده‌های اندازه‌ای
○ پارامترها	○ داده‌های مقوله‌ای

۴-۱ تمرین‌ها

- ۱-۱ تحت چه شرایطی تمام دانشجویان دانشکده یا دانشگاه شما به عنوان یک جمعیت به حساب می‌آیند؟
- ۱-۲ تحت چه شرایطی تمام دانشجویان دانشکده یا دانشگاه شما به عنوان یک نمونه به حساب می‌آیند؟
- ۱-۳ اگر دانشجویان دانشکده یا دانشگاه شما را مانند تمرین ۲-۱ به عنوان یک نمونه به حساب آوریم این نمونه یک نمونه تصادفی خواهد بود یا غیرتصادفی و چرا؟
- ۱-۴ چرا انتخاب نام از یک دفترچه تلفن محلی، نمونه‌ای تصادفی از ساکنان آن شهر فراهم نمی‌کند؟ چه کسانی کمتر و چه کسانی بیش از حد، شانس انتخاب شدن خواهند داشت؟
- ۱-۵ آیا می‌توانید راه‌هایی را پیشنهاد کنید که ما از طریق آن بتوانیم یک نمونه تصادفی (یا نزدیک به آن) از افراد یک شهر کوچک را انتخاب کنیم؟
- ۱-۶ ممکن است شما حتی قبل از اینکه این درس را شروع کنید از بعضی نمونه‌های آماری اطلاع داشته‌اید. دو مورد را نام ببرید.

- ۷-۱ مثالی از یک تحقیق بزنید که در آن برآورد نمره میانگین یک جمعیت مد نظر باشد.
- ۸-۱ مثالی از یک تحقیق بزنید که در آن مقدار عددی واقعی یک میانگین جمعیتی مد نظر باشد اما هدف دانستن این موضوع باشد که آیا میانگین یک جمعیت بزرگتر از میانگین جمعیت دیگر است.
- ۹-۱ سه نمونه از داده‌های مقوله‌ای را بیان کنید.
- ۱۰-۱ سه نمونه از داده‌های اندازه‌ای را بیان کنید.
- ۱۱-۱ مثالی بزنید که در آن چیزی که مورد مطالعه است یک متغیر اندازه‌ای یا مقوله‌ای باشد.
- ۱۲-۱ دو نمونه از مطالعاتی که هدف اصلی آنها بررسی روابط بین متغیرها است را بیان کنید.
- ۱۳-۱ دو نمونه از مطالعاتی که هدف اصلی آنها بررسی تفاوت گروه‌ها باشد را بیان کنید.

فصل ۲

مفاهیم اساسی آمار

۲-۱- مقیاس‌های اندازه‌گیری

۲-۲- متغیرها

۲-۳- نمونه‌گیری تصادفی

۲-۴- نمادسازی

در فصل قبل با شماری از واژه‌های آماری (مثل پارامتر، آماره، جمعیت، نمونه و نمونه تصادفی) آشنا شدیم که برای فهم تحلیل داده‌های آماری اساسی هستند. در این فصل به مفاهیم دیگری می‌پردازیم که به آنها نیاز خواهیم داشت. ما با مفاهیم مقیاس‌های اندازه‌گیری شروع می‌کنیم زیرا هر تحقیق با اندازه‌گیری چیزهای مورد مطالعه آغاز می‌شود. **اندازه‌گیری^۱** به کرات به عنوان اختصاص دادن اعداد به اشیاء تعریف شده است، که واژه‌های اعداد و اشیاء خیلی کلی تفسیر می‌شوند. بطور مثال وقتی ما یک آزمون اقتدارگرایی را به کار می‌گیریم (مثل، مقیاس اقتدارگرایی آدرنو) برای اینکه یک نمره اقتدارگرایی برای شخصی بدست آوریم، ما اقتدارگرایی را با اختصاص یک عدد (نمره) به یک شیء (یک شخص) اندازه می‌گیریم. بسته به آنچه که ما اندازه می‌گیریم و نحوه‌ی اندازه‌گیری آن، عددی که بدست می‌آوریم می‌تواند ویژگی‌های متفاوتی داشته باشد و این ویژگی‌های متفاوت اعداد^۲ اغلب تحت عنوان خاصی بنام مقیاس‌های اندازه‌گیری^۲ مورد بحث قرار می‌گیرند.

۱. measurement

۲- در اندازه‌گیری باید بین اعداد و علائم و پدیده‌هایی که علائم به آنها نسبت داده می‌شود همگونی وجود داشته باشد.

برای روشن شدن اهمیت این همگونی لازم است خصوصیات مهم اعداد را مرور کنیم:

مقیاس‌های اندازه‌گیری

عنوان مقیاس‌های اندازه‌گیری یکی از عنوان‌هایی است که بعضی نویسندگان آن را مهم و بعضی نامربوط تلقی می‌کنند. گرچه این کتاب گرایش به رویکرد دوم دارد ولی لازم است با کلیت آن آشنایی داشته باشید. سودمندی دیگر این بحث این است که شما به تدریج با این واقعیت آشنا می‌شوید که آمار به عنوان یک موضوع درسی تنها یک مجموعه وقایع یقینی و غیرقابل تغییر نیست، بلکه بیشتر واقعیت‌های کنار هم گذاشته شده با تفسیرها و نظرات گوناگون است.

به نظر می‌رسد اس. اس. استیونز مشهورترین فرد پیش‌گام است که مقیاس‌های اندازه‌گیری را برای انتخاب شیوه‌های آماری مهم دانست. استیونز در اصل چهار نوع مقیاس را تعریف کرده است: اسمی^۱، رتبه‌ای^۲، فاصله‌ای^۳ و نسبی (نسبتی)^۴. این مقیاس‌ها بر اساس روابط فرضی بین عناصر مقیاس که دارای ارزش‌های مختلفی هستند از هم متمایز می‌شوند. در این مجموعه، مقیاس آخر دربردارنده تمام ویژگی‌های سه مقیاس دیگر بعلاوه ویژگی‌های اضافی است.

مقیاس‌های اسمی

از یک لحاظ مقیاس‌های اسمی اصلاً مقیاس نیستند، زیرا عناصر را در طول هیچ بعدی نمی‌سنجند، بلکه آنها را برچسب می‌زنند. نمونه کلاسیک یک مقیاس اسمی مجموعه اعداد اختصاص داده شده به

[برای تکمیل این بحث بهتر است در اینجا خصوصیات انتزاعی اعداد را مرور کنیم: یکی از خصوصیات منطقی هر عدد مشخص بودن آن است، یعنی هر عدد برحسب تعریف دارای مفهوم مشخص و مستقلاً است و با هیچ عدد دیگری قابل اشتباه نیست. خصوصیت دوم مرتب بودن آنهاست، یعنی هر عدد از عدد خاص دیگری کوچکتر و از عدد خاصی بزرگتر است. خصوصیت سوم ترکیبی بودن اعداد است. از ترکیب دو عدد همیشه می‌توان عدد سومی بدست آورد. قواعد چهار عمل اصلی از این خصوصیت سوم ناشی می‌شود که با آنها اعداد را محاسبه می‌کنیم، چون اگر بتوان اعداد را با هم ترکیب کرد (جمع کرد) می‌توان آنها را از هم تفکیک کرد (تفریق کرد)؛ جمع و تفریق اعداد لازم و ملزوم یکدیگرند. ضرب و تقسیم هم از جمع و تفریق ناشی می‌شوند (ضرب یعنی جمع مکرر و تقسیم یعنی تفریق مکرر).

اگر پدیده‌ای که اندازه‌گیری می‌شود همه خصوصیات انتزاعی اعداد را بپذیرد، و روش سنجش پدیده هم این خصوصیات را تماماً آشکار کند، اندازه‌گیری کامل کمتی انجام گرفته است. اندازه‌گیری طول، مسافت، جرم و حجم از این نوع هستند. اما اندازه‌گیری بسیاری از پدیده‌های اجتماعی، رفتاری و تربیتی کامل کمتی نیست بلکه ناقص و کیفی است. چون یا ماهیت پدیده‌ها یا وسایل اندازه‌گیری فقط قسمتی از این خصوصیات را می‌توانند آشکار سازند، و اعدادی که برای نشان دادن اندازه‌ها به کار می‌روند معانی کلی و مبهم‌تر و ناقص‌تری به خود می‌گیرند. هرچه در اندازه‌گیری مقدار بیشتری از خصوصیات ریاضی و منطقی اعداد مصداق پیدا کند در اصطلاح تئوری اندازه‌گیری، قوی‌تر خوانده می‌شود.]

۱۵. nominal

۱۶. ordinal

۱۷. interval

۱۸. ratio

بازیکنان فوتبال است. این اعداد معنایی بیش از برچسب‌هایی که بازیکنان را از هم تشخیص دهد ندارند. ما می‌توانیم به آسانی با به کارگیری حروف یا عکس حیوانات از آنها استفاده کنیم. در واقع جنسیت یک مقیاس اسمی است که واژه‌ها (زن و مرد) را به جای اعداد به کار می‌گیرد. مقیاس‌های اسمی عموماً برای اهداف طبقه‌بندی استفاده می‌شوند. داده‌های مقوله‌ای که در فصل ۱ مورد بحث قرار دادیم براساس مقیاس اسمی سنجیده می‌شوند، چون ما فقط برچسب‌های مقوله‌ای (مثل زن یا مرد؛ جمهوری خواه، دموکرات یا مستقل) را به مشاهدات تخصیص می‌دهیم. داده‌های کمی (اندازه‌ای) را با یکی از سه نوع مقیاس زیر می‌توان سنجید.

مقیاس‌های رتبه‌ای (رتیبی)

ساده‌ترین مقیاس حقیقی یک **مقیاس رتبه‌ای** است که مردم، اشیاء یا پدیده‌ها را در طول یک پیوستار رتبه‌بندی می‌کند. یک نمونه از مقیاس رتبه‌ای شامل جایگاه افراد در یک مسابقه ورزش انفرادی است. در اینجا مقیاس به ما می‌گوید چه کسی بالاترین رتبه را دارد، چه کسی در مرتبه بعدی است و الی آخر. مثال دیگر مقیاس هولمز و راو (۱۹۶۷) در باره فشار زندگی است. با به کارگیری این مقیاس شما به طور ساده تعداد تغییرات زندگی شخص (ازدواج، جابجایی، شغل جدید، و غیره) طی شش ماه گذشته را می‌شمارید (گاهی با وزن‌های مختلفی که به آنها داده می‌شود). افرادی که نمره ۲۰ داشته باشند به عنوان کسانی که دارای فشار زندگی بیشتر نسبت به کسانی که نمره ۱۵ دارند، و اینها به نوبه خود دارای فشار بیشتر از کسانی که نمره ۱۰ دارند به حساب می‌آیند. بنابر این ما افراد را از لحاظ فشار روانی از روی تغییرات زندگی رتبه‌بندی می‌کنیم.

توجه داشته باشید که این دو مثال از جنبه اعدادی که به کار می‌گیرند فرق دارند. در مورد اول ما رتبه‌های ۱، ۲، ۳، ... را تخصیص می‌دهیم در حالی که در مورد دوم، نمرات بیانگر تعداد تغییرات هستند نه رتبه‌ها. هر دو نمونه‌ای از مقیاس رتبه‌ای هستند، چون در باره تفاوت بین نقاط مقیاس اطلاعاتی بدست نمی‌دهند. این یک ویژگی مهم مقیاس‌های رتبه‌ای است. ما فرض نمی‌گیریم که مثلاً تفاوت بین تغییرات زندگی ۱۰ و ۱۵ همان اندازه افزایش را نشان می‌دهد که تفاوت بین تغییرات ۱۵ و ۲۰ نشان می‌دهد. این نوع تشخیص‌ها را باید به مقیاس بعدی واگذار کرد.

مقیاس فاصله‌ای

با **مقیاس فاصله‌ای** ما نوعی اندازه‌گیری داریم که می‌توانیم به درستی از تفاوت بین نقاط مقیاس صحبت کنیم. یک مثال عمومی مقیاس فاصله‌ای، دمای سانتیگراد است، که در آن ۱۰ نقطه تفاوت، در هر جای مقیاس به یک معناست. بنابراین تفاوت بین ۱۰ و ۲۰ درجه به اندازه تفاوت بین ۳۰ و ۴۰ درجه است. توجه داشته باشید که این مقیاس ویژگی دو مقیاس قبلی را هم در بر دارد. اما چیزی که ما در مقیاس فاصله‌ای نداریم این است که نمی‌توانیم از نسبت سخن بگوییم. از این رو ما نمی‌توانیم مثلاً بگوییم ۲۰

درجه سانتیگراد به اندازه نصف ۴۰ درجه سانتیگراد داغ است، یا ۳۰ درجه دو برابر ۱۵ درجه است، زیرا نقطه صفر روی مقیاس، قراردادی (دلخواه) است. مثلاً ۷- درجه سانتیگراد و ۴ درجه سانتیگراد به ترتیب معادل ۲۰ و ۴۰ درجه فارنهایت هستند و روشن است که این دو مجموعه نسبت‌ها، کاملاً متفاوت و قراردادی هستند.

مقیاس نسبی (نسبتی) مقیاس نسبی

مقیاسی است که دارای مبدأ صفر حقیقی است. توجه داشته باشید که مبدأ صفر بایستی یک مبدأ صفر حقیقی باشد، نه قراردادی همچون درجه فارنهایت یا درجه سانتیگراد^۱ (سلسیوس). (یک مبدأ صفر نقطه‌ای است که در آن چیز مورد اندازه‌گیری وجود ندارد، یعنی مقدار آن صفر است. دلیل آن به این خاطر است که صفر درجه سانتیگراد و صفر درجه فارنهایت نشان دهنده عدم وجود دما نیست، اینها صفر حقیقی نیستند؛ صفر درجه کلون به عنوان یک نقطه صفر حقیقی به حساب می‌آید [که برابر ۲۷۳/۱۵- درجه سلسیوس است]. نمونه‌های مقیاس نسبی شامل مقیاس‌های معمول فیزیکی مانند وزن، طول، حجم و زمان است. این مقیاس‌ها نه تنها ویژگی‌های مقیاس‌های قبلی را دارا هستند بلکه همچنین می‌توان در آنها از نسبت‌ها سخن گفت. به زبان فیزیکی می‌توان گفت مدت زمان ۱۰ ثانیه دو برابر ۵ ثانیه است، ۱۰۰ کیلو یک سوم ۳۰۰ کیلو است، و نظیر آن.

انواع مقیاسی که ما با آنها کار می‌کنیم در نظر اول روشن به نظر می‌رسند. اما، بویژه در مورد انواع اندازه‌هایی که در علوم اجتماعی ما با آن سروکار داریم این موضوع به ندرت درست است. امتحان گروهی از دانشجویان در یک درس را در نظر بگیرید، ممکن است من تصور کنم این یک مقیاس نسبی از دانش آن ماده درسی است. به این معنی که کسی که نمره صفر گرفته هیچ چیز نمی‌داند، یک نمره ۸۰ گویای دانش دو برابر نمره ۴۰ است. اما فرد دیگری ممکن است ادعا کند که نمرات درسی یک مقیاس فاصله‌ای است و نقطه مبدأ آن قراردادی است (کسی که نمره صفر گرفته یک حداقلی را می‌داند اما سؤال‌ها آن را نشان نمی‌دهند)، تفاوت‌های یکسان در نمرات بازگو کننده تفاوت یکسان در دانش است. و بالاخره یک حالت بیشتر قابل دفاع می‌تواند این باشد که نمرات نشان دهنده یک مقیاس ترتیبی هستند. نمره ۹۵ بهتر از ۸۵ و این نمره بهتر از ۷۵ است اما تفاوت یکسان نمرات نمی‌تواند تفاوت یکسان در دانش را نشان دهد.

به عنوان نمونه‌ای از یک اندازه‌گیری که در آن مقیاس وابسته به کاربرد آن است، به دمای منزل خودتان توجه کنید. ما معمولاً از سانتیگراد به عنوان یک مقیاس فاصله‌ای نام می‌بریم. این تنها به عنوان

[۱۰. سانتیگراد و سلسیوس معادل هم هستند. آندرس سلسیوس (۱۷۰۱-۱۷۴۴) نام دانشمند سوئدی ابداع کننده این درجه است. سانتیگراد واحد تقسیمات ۱۰۰ درجه حرارت است که نقطه اوج آن نقطه جوش آب و نقطه صفر نقطه یخ‌زدگی آب در فشار یک اتمسفر است (نقل از دانشنامه آزاد ویکی‌پدیا).]

نمونه‌ای از یک مورد است و تردیدی نیست که برای یک فیزیکدان تفاوت بین ۱۷ درجه و ۱۸ درجه دقیقاً برابر با تفاوت بین ۲۲ درجه و ۲۳ درجه است. اما اگر دما را به عنوان شاخصی از راحتی اندازه بگیریم نه فعالیت مولکولی، این اعداد دیگر یک مقیاس فاصله‌ای را تشکیل نمی‌دهند. برای کسی که در یک اتاق با ۱۷ درجه سانتیگراد است پرش ناگهانی به درجه ۱۸ قابل توجه (و مطلوب) است. اما همین موضوع را در مورد اختلاف دمای اتاق ۲۲ و ۲۳ درجه نمی‌توان گفت. نکات یاد شده گویای این واقعیت است که این متغیر مورد اندازه‌گیری (مثل راحتی) است که مقیاس را تعریف می‌کند نه خود اعداد. از آنجا که توافق همگانی درباره مقیاس اندازه‌گیری به کار رفته وجود ندارد، برعهده استفاده کننده از شیوه‌های آماری است که بهترین تصمیم را با توجه به ماهیت داده‌ها بگیرد. آنچه که می‌توان از او خواست این است که قبل از آنکه تصمیم بگیرد، به دقت در مسأله اندیشه کند، و به سادگی فرض نگیرد که جواب استاندارد حتماً بهترین جواب است.

نقش مقیاس‌های اندازه‌گیری

پیش‌تر بیان شد که دیدگاه‌های متفاوتی نسبت به اهمیت مقیاس‌های سنجش وجود دارد. برخی مؤلفان به کلی مسأله را نادیده می‌گیرند، در حالیکه دیگران کل متن را پیرامون مقیاس‌های مختلف سازمان می‌دهند. برای مؤلف این کتاب مسأله اصلی این ضرورت مطلق است که ما اعدادی را که جمع‌آوری می‌کنیم در ذهن خود از اشیاء و پدیده‌هایی که به آنها رجوع می‌کنند تفکیک کنیم. اگر یک آزمودنی در یک مطالعه یادگیری کلامی ۲۰ ایتام و آزمودنی دیگر ۱۰ ایتام را به یاد آورد، تعداد کلمات به یاد آورده برای آزمودنی اول دو برابر است. اما ما ممکن است تمایل نداشته باشیم که بگوییم آزمودنی نخست مواد مورد مطالعه را دو برابر به یاد آورده است.

سومین بحث در مورد مثال مربوط به دمای اتاق هم مطرح است، که مقیاس (فاصله‌ای یا ترتیبی) بستگی به این دارد که ما علاقه‌مند به اندازه‌گیری یک اسناد فیزیکی از دما هستیم یا اثر آن روی افراد. تفاوت ۲° در هر جای مقیاس به لحاظ فیزیکی یکسان است، اما یک تفاوت دو درجه‌ای وقتی اتاقی در حال حاضر گرم است ممکن است به آن اندازه احساس نشود که اتاق واقعاً سرد است. به عبارت دیگر ما یک مقیاس فاصله‌ای از واحدهای فیزیکی داریم اما در مورد مطلوبیت دما چیزی بیش از یک مقیاس ترتیبی نداریم.

از آنجا که آزمون‌های آماری اعداد را بدون توجه به اشیاء یا پدیده‌هایی که به آنها رجوع می‌کنند به کار می‌برند، ما ممکن است هر یک از عملیات‌های استاندارد ریاضی (جمع، ضرب و غیره) را بدون توجه به ماهیت مقیاس زیربنایی به اجرا درآوریم. مشکل وقتی بروز می‌کند که به سراغ تفسیر نتایج حاصل از بعضی کارهای آماری برویم. در این میان بایستی بپرسیم که آیا نتایج آماری ارتباط بامعنایی با اشیاء و پدیده‌های مورد بررسی نشان می‌دهند. اینجا دیگر ما با یک مسأله آماری سروکار نداریم بلکه

یک موضوع روش‌شناسی است. هیچ شیوه آماری نمی‌تواند به ما بگوید آیا این واقعیت که یک گروه که در درس تاریخ نمرات بالاتری از گروه دیگر گرفته چیزی را در باره اختلاف دو گروه در دانش آن ماده درسی نشان می‌دهد. از این گذشته، دلخوش بودن به اینکه امتحان نمراتی را بدست می‌دهد که شکل یک مقیاس نسبی از ایتهم‌های درست را دارد (۵۰ ایتهم درست دو برابر ۲۵ ایتهم درست است) نادیده گرفتن این واقعیت است که ما در پی اندازه‌گیری دانش تاریخ بوده‌ایم، که ممکن است با هیچ قاعده منظمی با افزایش نمرات افزایش نیابد. آزمون‌های آماری را می‌توانیم روی اعدادی که بدست می‌آوریم به کار بگیریم، و درستی عبارات در باره اشیاء و پدیده‌هایی که فکر می‌کنیم در حال اندازه‌گیری آنها هستیم در اصل بر پایه دانش ما در باره اشیاء و پدیده‌ها استوار است نه مقیاس اندازه‌گیری. بهترین کار این است که اطمینان حاصل کنیم که اندازه‌های ما تا حد ممکن با آنچه که می‌خواهیم بسنجیم رابطه نزدیکی دارند، اما در نهایت نتایج ما همان عددهایی است که بدست می‌آوریم و ایمان به ارتباط بین آن عددها و اشیاء یا پدیده‌های زیربنایی آنها می‌باشند.

متغیرها

به ویژگی‌های اشیاء و پدیده‌ها که می‌توانند اندازه‌ها یا ارزش‌های مختلفی به خود بگیرند **متغیر**^۱ گفته می‌شود. برای مثال رنگ مو یک متغیر است چون ویژگی یک شیئی (مو) است و می‌تواند ارزش‌های مختلفی (مشکی، قهوه‌ای، بور و غیره) را بخود بگیرد. ویژگی‌هایی مانند طول، ارتفاع، و سرعت به همین دلیل متغیر هستند. همچنین ما می‌توانیم متغیرهای گسسته و متغیرهای پیوسته را از هم جدا کنیم. در **متغیرهای گسسته**^۲ (مثل جنسیت، وضعیت تأهل، و تعداد دستگاه تلویزیون در یک منزل شخصی) متغیر فقط یکی از تعداد محدود ارزش‌های ممکن را می‌تواند به خود بگیرد. **متغیرهای پیوسته**^۳ آنهایی هستند که دست کم از لحاظ نظری می‌توانند هر ارزشی را بین پایین‌ترین و بالاترین نقطه روی مقیاس به خود بگیرند، مانند سرعت، زمان، طول دم یک بزغاله و غیره. چنانکه بعداً در این کتاب خواهیم دید این تمایز نقش مهمی را در بعضی از شیوه‌های آماری بازی می‌کند.

در آمار انواع متغیرها را به روش دیگری نیز از هم تفکیک می‌کنند. منظور **متغیرهای مستقل**^۴ (آنهایی که بوسیله آزمایشگر دستکاری یا کنترل می‌شوند تا اثرشان بررسی شود) و **متغیرهای وابسته**^۵ (که تغییرات آنها را تحت تأثیر متغیر مستقل جستجو می‌کنیم - داده‌ها) هستند. در تحقیقات آزمایشی آزمایشگر مایل است اثرات متغیرهای مستقل بر متغیرهای وابسته را مورد مطالعه قرار دهد. نمونه‌ای از

1. variable
2. discrete variables
3. continuous
4. Independent
5. dependent

متغیرهای مستقل شامل روش‌های مختلف تدریس، بازخوردهای مختلف تقویت رفتار و الگوهای متفاوت تعامل اجتماعی هستند و متغیرهای وابسته نیز می‌توانند همانند پیشرفت تحصیلی، گرایش به ترک اعتیاد و نگرش به پرخاشگری باشند. در اصل، تحقیق پیرامون متغیر مستقل به عمل می‌آید و نتایج (داده‌ها) شامل اندازه‌هایی از متغیر وابسته هستند. بطور مثال محقق ممکن است میزان پرخاشگری کلامی در دو گروه کودکانی که یکی فیلم پرخاشگرانه و دیگری فیلم غیرپرخاشگرانه تماشا کرده‌اند را بسنجد. در اینجا نوع فیلم متغیر مستقل و میزان پرخاشگری کلامی کودکان متغیر وابسته است. متغیرهای مستقل ممکن است کیفی (مثل مقایسه سه روش تدریس) یا کمی (مانند سن) باشند، در حالی که متغیرهای وابسته عموماً - اما نه همیشه - کمی هستند.

۲-۳ نمونه‌گیری تصادفی

در فصل ۱ گفته شد یک نمونه در صورتی یک نمونه تصادفی است که هر عنصر یعنی همه عناصر جمعیت شانس برابری برای انتخاب شدن به عنوان نمونه داشته باشند. علاوه بر این گفتیم که مفهوم نمونه‌گیری تصادفی برای فرایند استفاده از آماره‌های محاسبه شده بر اساس نمونه، برای استنباط مقادیر پارامترهای جمعیتی یک امر اساسی است. باید روشن شده باشد که احمقانه است اگر سعی کنیم میانگین سطح رفتار بزهکارانه همه دانش‌آموزان دبیرستانی را بر اساس داده‌های حاصل از یک گروه دانش‌آموزان کلاس نهم برآورد کنیم که در یک سالن مطالعه بوده‌اند. بدیهی است که این داده‌ها برآورد پایینی از اندازه میانگین بدست می‌دهد نسبت به وقتی که یک نمونه تصادفی از کل دانش‌آموزان دبیرستانی را بررسی کنیم.

روش‌های متعددی برای بدست آوردن یک نمونه تصادفی از جمعیت‌های نسبتاً کوچک وجود دارد. می‌توانیم به هر فرد یک شماره اختصاص دهیم و از جدول اعداد تصادفی (که معمولاً ضمیمه کتاب‌های آماری است) برای انتخاب تعدادی که در نمونه قرار می‌گیرند استفاده کنیم. یا اگر یک نمونه تقریباً تصادفی کفایت کند می‌توانیم اسم افراد را در یک کلاه بریزیم و چشم بسته تعدادی را بیرون بکشیم. نکته اساسی این است که هر نمره در جمعیت، شانس برابر برای انتخاب شدن داشته باشد.

اغلب خوب است که برای بیرون کشیدن نمونه‌های تصادفی، اختصاص آزمودنی‌ها به گروه‌ها و کارهای دیگر یک جدول اعداد- مانند جدول اعداد تصادفی پیوست داشته باشیم. این جدول شامل اعداد تصادفی همگون است. منظور از همگونی^۱ این است که احتمال بروز همه اعداد برابر است. (بطور مثال اگر شما تعداد اعداد ۱، ۵، و ۸ را بشمارید، خواهید دید که همه آنها تقریباً به یک اندازه تکرار شده‌اند. استفاده از این جدول کاملاً ساده است. برای مثال اگر شما بخواهید اعداد تصادفی بین ۰ تا ۹ را بیرون بکشید، بطور ساده با چشم بسته، انگشت خود را روی جدول قرار داده سپس روی ستون اعداد

1. uniform

زیر انگشت خود پایین می‌روید و اعدادی که برخورد می‌کنید را می‌خوانید. وقتی به انتهای ستون رسیدید به ستون بعدی می‌روید و این کار را ادامه می‌دهید تا تعدادی را که می‌خواهید بیابید. اگر اعداد بین ۰ تا ۹۹ را بخواهید همین کار را باید انجام دهید اما این بار باید دو رقم را بخوانید. در نهایت اگر اعداد تصادفی بین ۱ تا ۶۵ را بخواهید بایستی دو رقم را بخوانید اما عدد ۰۰ و اعداد بین ۶۶ تا ۹۹ را نادیده بگیرید.

اگر به جای جمع‌آوری یک مجموعه از اعداد تصادفی، بخواهید جدول اعداد تصادفی را برای تخصیص آزمودنی‌ها به دو گروه به کار بگیرید، از هر نقطه‌ای از جدول شروع می‌کنید و اگر عدد فرد بود آزمودنی را به گروه I و اگر زوج بود به گروه II اختصاص دهید. بطور منطقی، این شیوه به شما امکان تخصیص تصادفی آزمودنی‌ها به هر تعداد گروه را می‌دهد.

برای جمعیت‌های بزرگ بسیاری از شیوه‌های استاندارد تصادفی جواب‌گو نخواهد بود. ما نه می‌توانیم نام همه زنان ۲۱ تا ۳۰ ساله یک کشور را داخل یک کلاه بریزیم (حتی اگر یک کلاه خیلی بزرگ داشته باشیم)، و نه می‌توانیم به همه آنها یک شماره اختصاص دهیم و بعد به جدول اعداد تصادفی مراجعه کنیم. چنین کاری کلاً غیرعملی است. مگر اینکه منبع مالی خیلی زیادی داشته باشیم، تقریباً بهترین کاری که می‌توانیم انجام دهیم این است که تا حد ممکن منابع سوگیری را حذف کنیم (مثلاً سطح رفتار بزهکارانه را فقط براساس یک نمونه که تحت پوشش حمایتی خاصی بوده‌اند برآورد نکنیم)، نتیجه‌گیری خود را محدود به منابع سوگیری کنیم که براحتی قابل کنترل نیستند (مثلاً اعلام کنیم که داده‌ها از طریق پرسشنامه‌های افراد داوطلب جمع‌آوری شده‌اند). هر سوگیری که وجود داشته باشد میزان تعمیم نتایج به جمعیت کلی را محدود خواهد کرد. در باره شیوه‌های نمونه‌گیری، مطالب زیاد دیگری برای معرف بودن نمونه می‌توان اضافه نمود که از حوصله این کتاب خارج است.

۴-۲ نمادها

هر بحثی در باره شیوه‌های آماری نیازمند یک نظام نمادگذاری برای بیان عبارات ریاضی است. از این رو جای تعجب است که هیچ‌گونه نظام استاندارد در این باره طرح نشده است. گرچه تلاش‌های زیادی صورت گرفته تا یک شیوه فراگیر طرح شود، اما هنوز دو کتاب پیدا نمی‌شود که دقیقاً نمادهای یکسانی را به کار ببرند.

نظام‌های نمادگذاری مرسوم، از خیلی ساده تا خیلی پیچیده گسترده‌اند. نظام‌های پیچیده‌تر از مزیت دقت برخوردارند ولی آسانی درک مطلب را فدا می‌کنند، درحالی‌که نظام‌های ساده‌تر، فهم آسان را جایگزین دقت می‌کنند. ما در این کتاب سادگی را ترجیح داده‌ایم.

نمادگذاری متغیرها

قاعده عمومی این است که یک متغیر به صورت کلی با یک حرف بزرگ انگلیسی معمولاً X و Y نشان داده می‌شود. یک مقدار منفرد از آن متغیر با همان حرف و یک اندیس نشان داده می‌شود. فرض کنید مثلاً ما نمره‌های زیر را برای مدت زمانی (به ثانیه) که کودکان سوم ابتدائی بتوانند نفس خود را حبس کنند داشته باشیم:

۴۵ ۴۲ ۳۵ ۲۳ ۵۲

این مجموعه نمرات با X نشان داده می‌شوند. نمره اول این مجموعه (۴۵) را می‌توان به صورت X_1 دومی (۴۲) به صورت X_2 و الی آخر، نشان داد. وقتی که به یک نمره منفرد بدون توجه به اینکه کدام نمره است اشاره می‌کنیم به صورت X_i نشان داده می‌شود که i را می‌توان در مثال بالا بین ۱ تا ۵ در نظر گرفت. استفاده از اندیس برای دقیق ساختن توصیف شیوه آماری اساسی است. اما در عمل استفاده از اندیس بیشتر مشکل‌زا است تا اینکه کمک کننده باشد. در این کتاب هر جا که معنا روشن بوده عموماً این اندیس‌ها حذف شده‌اند.

نشانگان جمع

یکی از نمادهای معمول در آمار، حرف یونانی سیگما (Σ) است که نماد استاندارد برای جمع است. این نشانه به عنوان "حاصل جمع، یا مجموع آنچه در جلو آن می‌آید" معنا می‌شود. بنابراین ΣX_i به صورت "مجموع X_i ها" خوانده می‌شود. به بیان دقیق، نماد حاصل جمع همه N مقدار X عبارت است از:

$$\sum_{i=1}^N X_i$$

که چنین بازگو می‌شود: "مجموع X_i ها از $i=1$ تا $i=N$." در عمل به ندرت نیاز به این کار هست، و در بسیاری موارد همه اندیس‌ها حذف می‌شوند، و نماد مجموع X_i ها بطور ساده ΣX خواهد بود.

لازم است به چند شکل از این حالت ساده ΣX توجه کنیم تا فهم آنها آسان شود. یکی از آنها ΣX^2 است که به این صورت خوانده می‌شود: "مجموع مجذور مقادیر X " (مثلاً، $۵۲^۲ + ۲۳^۲ + ۳۵^۲ + ۴۲^۲ + ۴۵^۲$). عبارت معمول دیگر ΣXY است، که به معنای "مجموع حاصل ضرب مقادیر X و Y است". کاربرد این عبارت‌ها در مثال زیر نشان داده شده است.

یک آزمایش ساده را در نظر بگیرید که در آن سرعت دویدن (X) پنج موش و همچنین تعداد دفعات تقویت شدن (Y) آنها ثبت شده است. داده‌ها و حاصل جمع ساده عملیات حسابی روی آنها در جدول ۱-۲ آمده است. برخی از عملیات‌ها در اینجا و بقیه در فصل‌های بعد مورد بحث قرار می‌گیرند. جدول ۱-۲ مجموعه عملیات‌های دیگری شامل پرانتزهایی مانند $(\Sigma X)^2$ را نشان می‌دهد. قاعده کلی که همیشه اعمال می‌شود این است که عملیات درون پرانتزها قبل از عملیات بیرون پرانتز انجام شود.

بنابراین ما مقادیر X را جمع کرده و سپس حاصل را به توان دو می‌رسانیم، برخلاف $\sum X^2$ که ما ابتدا X ها را به توان دو می‌رسانیم قبل از آن که آنها را با هم جمع کنیم. خواننده باید با امتحان کردن روی عددهای ساده‌ای مانند ۲، ۳، و ۴ دقت کند که $\sum X^2$ برابر نیست ($\neq$) با $(\sum X)^2$.

جدول ۱-۲ نمایش عملیات نماد حاصل جمع

سرعت دویدن		دفعات تقویت			
XY	$X-Y$	Y^2	X^2	Y	X
۳۰	۷	۹	۱۰۰	۳	۱۰
۶۰	۱۱	۱۶	۲۲۵	۴	۱۵
۱۲	۱۱	۱	۱۴۴	۱	۱۲
۹	۸	۱	۸۱	۱	۹
۳۰	۷	۹	۱۰۰	۳	۱۰
—	—	—	—	—	—
۱۴۱	۴۴	۳۶	۶۵۰	۱۲	۵۶

حاصل
جمع

$$\sum X^2 = (10 + 15 + 12 + 9 + 10) = 56$$

$$\sum Y^2 = (3 + 4 + 1 + 1 + 3) = 12$$

$$\sum X^2 = (10^2 + 15^2 + 12^2 + 9^2 + 10^2) = 560$$

$$\sum Y^2 = (3^2 + 4^2 + 1^2 + 1^2 + 3^2) = 36$$

$$\sum (X - Y) = (7 + 11 + 11 + 8 + 7) = 44$$

$$\sum XY = (10)(3) + (15)(4) + (12)(1) + (9)(1) + (10)(3) = 141$$

$$(\sum X)^2 = 56^2 = 3136$$

$$(\sum Y)^2 = 12^2 = 144$$

$$(\sum (X - Y))^2 = (44)^2 = 1936$$

$$(\sum X)(\sum Y) = (56)(12) = 672$$

برای یادگیری اکثر شیوه‌های آماری، فهم دقیق و همه جانبه نمادها لازم است. شما بایستی جدول ۱-۲ را آنقدر مطالعه کنید که همه روندهای آن را کاملاً فرا بگیرید.

قواعد حاصل جمع

برای استفاده از علامت حاصل جمع ($\sum$) آشنایی با سه قاعده دیگر بسیار سودمند خواهد بود که در ادامه بیان می‌شود. کاربرد آنها با مثال‌های خیلی ساده روشن می‌شود که به خودتان واگذار می‌شود.

$$1. \sum (X - Y) = \sum X - \sum Y$$

$$۲. \sum CX = C \sum X$$

نماد $\sum CX$ به این معناست که هر ارزش X را در مقدار ثابت C ضرب نموده و سپس آنها را با هم جمع کنید.

$$۳. \sum_{i=1}^N (X + C) = \sum X + NC$$

(N نشان‌دهنده تعداد عناصر است)

۲-۵ خلاصه

در این فصل ما مفهوم اندازه‌گیری را بطور خلاصه مطالعه کردیم و چهار سطح متفاوت یا مقیاس سنجش را مورد بحث و بررسی قرار دادیم. همچنین مفهوم متغیر، انواع متغیر و نظام نمادهای مورد استفاده در کتاب را توضیح دادیم. در این مرحله شما واژه‌شناسی لازم را برای آغاز مشاهده داده‌ها فرا گرفته‌اید. برخی از مطالب مهم فصل عبارتند از:

- اندازه‌گیری
- مقیاس‌های سنجش
- مقیاس اسمی
- مقیاس ترتیبی
- مقیاس فاصله‌ای
- مقیاس نسبی
- متغیرها
- متغیرهای گسسته
- متغیرهای پیوسته
- متغیرهای مستقل
- متغیرهای وابسته
- حاصل جمع
- ثابت

۲-۶-تمرین‌ها

۱-۱ یک مثال برای هر نوع از مقیاس‌ها بیان کنید.

۲-۲ یک نمونه از متغیرهایی که می‌توان گفت برای بعضی هدف‌ها در مقیاس نسبی و برای هدف‌های دیگر در مقیاس فاصله‌ای یا رتبه‌ای اندازه‌گیری شده‌اند بیان کنید.

۲-۳ ما به موش‌ها آموزش دادیم تا در یک ماز بدون تا با غذا تقویت شوند. ناگهان یکی از موش‌ها در نیمه راه دراز کشید و خوابید. وقتی سرعت به عنوان شاخص یادگیری باشد این اتفاق در باره مقیاس اندازه‌گیری چه چیز را بیان می‌کند.

۲-۴ تمرین ۲-۳ در باره سرعت به عنوان یک شاخص انگیزش چه چیز را بیان می‌نماید؟

۲-۵ دو نمونه از متغیرهای مستقل و دو نمونه از متغیرهای وابسته را بیان کنید.

۲-۶ یک جمله بنویسید که بیانگر یک آزمایش شامل یک متغیر مستقل و یک متغیر وابسته باشد.

۲-۷ سه نمونه از متغیرهای پیوسته را بیان کنید.

۲-۸ سه نمونه از متغیرهای گسسته را بیان کنید.

۲-۹ در یک آزمایش فرضی هریس، پی‌بادی و اسمیت (۱۹۹۲) به رتبه بندی ۱۰ فرد اروپایی و ۱۰ نفر از آمریکای شمالی در یک مقیاس ۱۲ درجه‌ای مهارت موسیقی پرداختند. داده‌ها برای اروپایی‌ها چنین است:

۷ ۲ ۸ ۷ ۱۱ ۱۰ ۵ ۹ ۸ ۱۰

با نامگذاری این متغیر با X .

الف) X_5 ، X_3 و X_8 کدام نمره‌ها هستند؟

ب) مقدار $\sum X$ را حساب کنید.

ج) نماد حاصل جمع را برای مورد (ب) در پیچیده‌ترین حالت آن بیان کنید.

۲-۱۰ بازگشت به تمرین ۲-۹، داده‌های مربوط به افراد آمریکای شمالی عبارتند از:

۲ ۵ ۶ ۶ ۴ ۸ ۳ ۵ ۹ ۹

با نامگذاری این متغیر با Y .

الف) Y_1 و Y_{10} کدامند؟

ب) مقدار $\sum Y$ را حساب کنید.

۲-۱۱ با استفاده از داده‌های تمرین ۲-۹،

الف) مقدار $(\sum X)^2$ و $\sum X^2$ را حساب کنید.

ب) مقدار $\sum X/N$ را در جایی که N تعداد نمرات است حساب کنید.

ج) آنچه که اکنون محاسبه کردید چه نامیده می‌شود؟

۲-۱۲ با استفاده از تمرین ۲-۱۰،

الف) مقدار $(\sum Y)^2$ و $\sum Y^2$ را حساب کنید.

ب) مقدار کسر زیر را حساب کنید:

$$\frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N-1}$$

ج) ریشه دوم جواب (ب) را حساب کنید.

د) واحدهای اندازه‌گیری (ب) و (ج) چه هستند؟

(در فصل ۶ دوباره با این محاسبات سروکار خواهیم داشت.)

۲-۱۳ با داده‌های تمرین ۲-۹ و ۲-۱۰،

الف) $\sum XY$ را حساب کنید.

ب) $\sum X \sum Y$ را حساب کنید.

ج) کسر زیر را حساب کنید:

$$\frac{\sum XY - \frac{\sum X \sum Y}{N}}{N-1}$$

(در فصل ۹ دوباره با این محاسبات سروکار خواهیم داشت.)

۲-۱۴ داده‌های قبلی را به کار بگیرید که نشان دهید:

$$\sum (X + Y) = \sum X + \sum Y \quad \text{الف)}$$

$$\sum XY \neq \sum X \sum Y \quad \text{ب)}$$

ج) $\sum CX = C \sum X$ وقتی که C یک مقدار ثابت است (مثل ۳).

$$\sum X^2 \neq (\sum X)^2 \quad \text{د)}$$

۲-۱۵ داده در نظر بگیرید و با استفاده از قاعده سوم حاصل جمع نشان دهید که اگر ۱۰ نمره به نمره

هر شخص اضافه کنید مقدار کل چه تغییری می‌کند.

۲-۱۶ در نسخه قبلی این کتاب ما (بدرستی) به این مثال که "تعداد موهای بزغاله" نمونه ای از متغیر

پیوسته است انتقاد کردیم. چرا این متغیر گسسته است؟

فصل ۳

نمایش داده‌ها

- ۱-۳- رسم نمودار داده‌ها
- ۲-۳- دسته‌بندی داده‌ها
- ۳-۳- توزیع‌های تجمعی
- ۴-۳- نمایش‌های ساقه و برگ
- ۵-۳- روش‌های مختلف رسم داده‌ها
- ۶-۳- توصیف توزیع‌ها
- ۷-۳- استفاده از برنامه‌های رایانه‌ای برای نمایش داده‌ها

یک مجموعه از داده‌های خام، به خودی خود چیزی جالب‌تر یا آگاهی بخش‌تر از آگهی‌های تبلیغاتی روز قبل از انتخابات نیست. خواه داده‌ها به صورت منظم در ردیف‌های فرم داده‌ها قرار گرفته باشند، یا اینکه چون دفتر یادداشت‌تان را فراموش کرده بودید به‌طور نامنظم آن را پشت یک تکه کاغذپاره تاریخ گذشته تابلو اعلانات نوشته باشید، در هر حال چیزی بیش از یک مجموعه اعداد نیست. آنها در صورتی قابل تفسیر هستند که ابتدا به شکلی منطقی سازمان داده شوند.

به طور نمونه رتبه‌بندی ۲۰۰ دانشجو در باره مطلوبیت یک مرکز خدمات دانشجویی را در نظر بگیرید. مطلوبیت روی یک مقیاس از ۰ تا ۱۰ طرح شده که نمره بالاتر نشانگر مطلوبیت بیشتر است. در اینجا ما ۲۰۰ نمره داریم، هر نمره برای یک دانشجو. روشن است که از نگاه به این مجموعه اعداد، چیزی نخواهیم فهمید، مگر برداشتی کلی و مبهم مبنی بر اینکه این مرکز از نظر افراد عموماً مطلوب یا نامطلوب است. ما بایستی ابتدا داده‌ها را سازمان داده و سپس آنها را به چند عدد که بیشترین اطلاعات را در خود داشته باشند کاهش دهیم.

۳-۱ رسم کردن نمودار داده ها

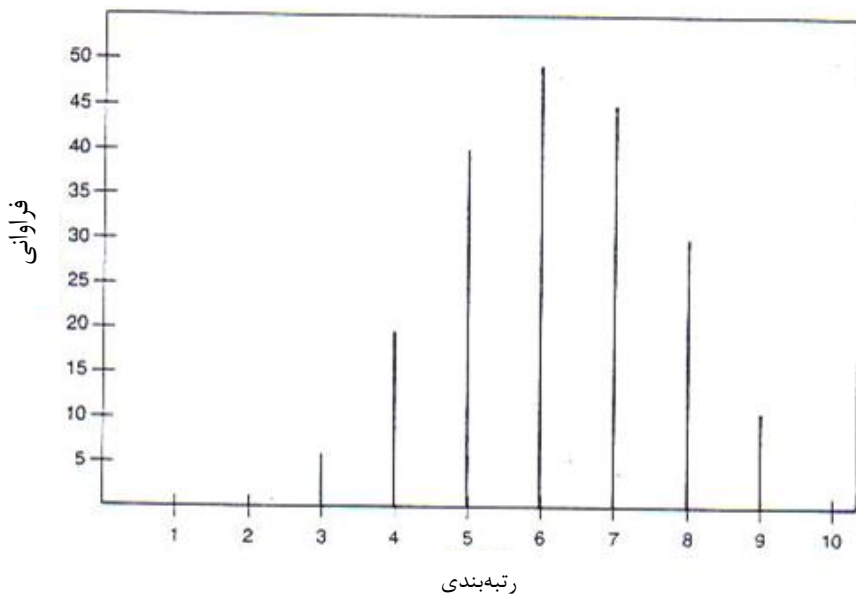
یکی از ساده ترین راه های سازماندهی برای اینکه داده ها را قابل فهم تر کنیم این است که نوعی نمودار گرافیکی از آنها ترسیم کنیم. چند شیوه برای این کار معمول است.

توزیع های فراوانی

به عنوان نخستین مرحله ممکن است ما بخواهیم یک **توزیع فراوانی**^۱ از داده ها بسازیم. در مورد رتبه بندی دانشجویان از مطلوبیت یک مرکز، ما تعداد دفعات تکرار هر یک از ۱۱ رتبه عددی توسط آزمودنی ها را شمارش می کنیم. در جدول ۳-۱ یک توزیع فراوانی احتمالی برای این مطالعه نشان داده شده و در شکل ۳-۱ به نمایش در آمده است. واژه توزیع فراوانی اغلب برای بیان آنچه شما در جدول ۳-۱ و شکل ۳-۱ مشاهده می کنید به کار می رود.

جدول ۳-۱ توزیع فراوانی رتبه بندی مطلوبیت یک مرکز دانشجویی فرضی

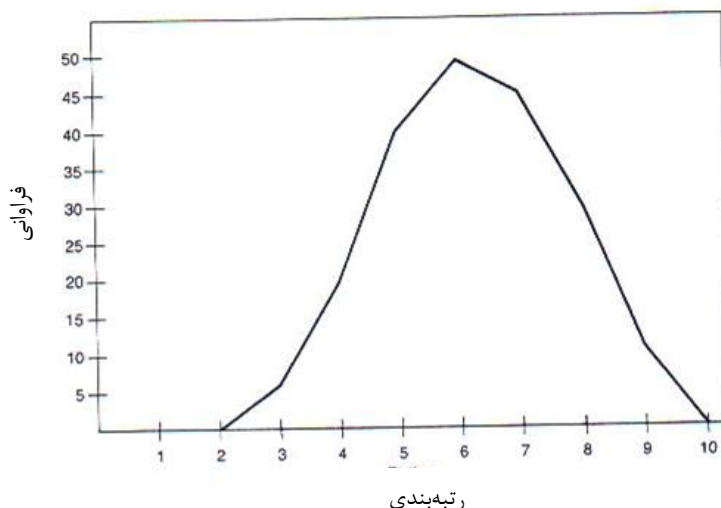
رتبه بندی	فراوانی
(X)	(f)
۰	۰
۱	۰
۲	۰
۳	۶
۴	۱۹
۵	۴۰
۶	۴۹
۷	۴۵
۸	۳۰
۹	۱۱
۱۰	۰



شکل ۳-۱ توزیع فراوانی ۲۰۰ رتبه مطلوبیت یک مرکز دانشجویی فرضی

از توزیع نشان داده شده در شکل ۳-۱ روشن می‌شود که یک توزیع گسترده از دیدگاه‌ها در باره مطلوبیت این مرکز وجود دارد، که ۶ نفر آن را در حد پایین‌تر از ۳، و ۱۱ نفر دیگر در حد بالاتر از ۹ رتبه‌بندی کرده‌اند. گرایش عمومی در داده‌ها به سمت خوشه نزدیک به نقطه ای است که کمی بالاتر از میانه مقیاس قرار دارد- نمره ۶ عمومی‌ترین مقدار است که ۴۹ نفر آن را برگزیده‌اند.

یک روش دیگر برای ترسیم یک توزیع فراوانی این است که رأس میله‌های شکل ۳-۱ را با خط به هم وصل کنیم. آنگاه می‌توانید خط‌های عمودی را حذف کنید. چنین نموداری را چندضلعی فراوانی گویند و نمونه‌ای از آن در شکل ۳-۲ نشان داده شده است.

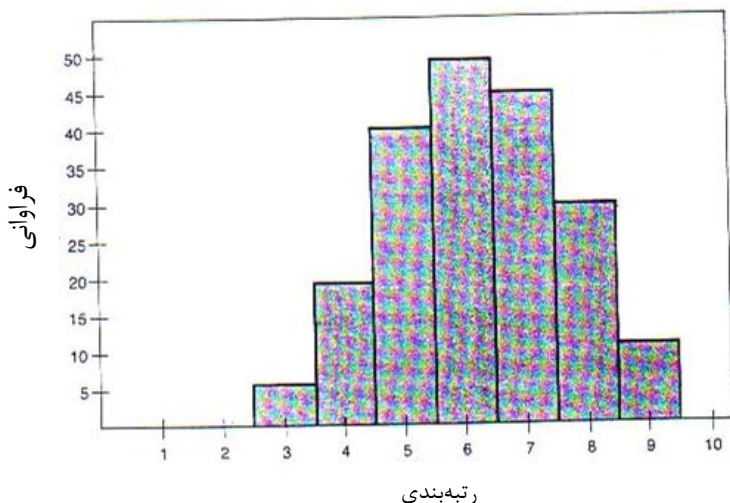


شکل ۲-۳ چندضلعی فراوانی ۲۰۰ رتبه مطلوبیت فرضی یک مرکز دانشجویی

هیستوگرام‌ها

از آنجا که اجبار کردن یک آزمودنی به انتخاب یک مقدار صحیح برای رتبه‌بندی (یعنی نمی‌تواند مثلاً رتبه $3/68$ در نظر بگیرد) او را وادار می‌سازد برای مقادیر بین $5/5$ و $6/5$ یک رتبه ۶ را به هر آزمودنی (یا شخص) اختصاص دهد، ما ممکن است نموداری را انتخاب کنیم که داده‌ها را به صورت یک مجموعه ستون‌های عمودی ترسیم کند. چنین نموداری **هیستوگرام**^۱ نامیده می‌شود، و یک نمونه از آن در شکل ۳-۳ نشان داده شده است.

هیستوگرام یک خاصیت جالب دارد که به سایر روش‌های رسم نمودار فراوانی سرایت می‌کند. چنانچه ما به دل‌خواه پهنای یک فاصله را یک واحد تعریف کنیم، آنگاه سطح هر فاصله یا هر مجموعه از فاصله‌ها برابر با تعداد نمراتی است که در آن فاصله یا مجموعه فاصله‌ها قرار می‌گیرند. علاوه بر این کل سطح زیر هیستوگرام برابر با N ، تعداد نمونه (تعداد نمرات) است. همین تعمیم را می‌توان با اصلاح مناسبی در نگارش، در باره بسیاری از توزیع‌هایی که رسم می‌کنیم بیان کنیم. دلیل اینکه آماردانان مرتباً به “سطح زیر منحنی” اشاره می‌کنند همین است – سطح و تعداد (یا درصد) نمرات، مفاهیمی هستند که می‌توان آنها را بجای یکدیگر بکار برد. اگر ما سطح کلی زیر منحنی را برابر $1/00$ تعریف کنیم، آنگاه سطحی که در محدوده یک بخش معین از هیستوگرام واقع می‌شود معادل نسبت یا درصدی از نمرات است که در آن بخش واقع شده‌اند.



شکل ۳-۳ هیستوگرام ۲۰۰ رتبه مطلوبیت یک مرکز دانشجویی

۳-۲ دسته‌بندی داده‌های

ما در مبحث پیش پیرامون توزیع‌های فراوانی و هیستوگرام‌ها، فرض کردیم که داده‌ها به سبک گسسته با تعداد نسبتاً کمی ارزش‌های متفاوت ارائه شده‌اند. اما فرض بگیرید که با داده‌های مربوط به مدت زمانی (به دقیقه) سروکار داشته باشیم که هریک از ۱۰۰ دانشجو در یک روز معین صرف بازی الکترونیکی می‌کنند. به استثنای نمره صفر (برای دانشجویانی که هرگز بازی الکترونیکی نمی‌کنند)، هر یک از نمرات منفرد به احتمال زیاد توسط تعداد کمی بدست می‌آیند، شاید هم هیچ کس. به عبارت دیگر شخصی ممکن است ۱۸ دقیقه در روز بازی کند، هیچ‌کس نباشد که ۱۹ دقیقه بازی کند، و دیگری ۲۰ دقیقه و غیره. برای این نوع داده‌ها چندان سودمند نخواهند بود که یک توزیع فراوانی استاندارد طرح‌ریزی و ارائه کنیم، زیرا چنین نموداری خیلی مسطح و گسترده خواهد بود. مفیدتر آن خواهد بود که داده‌ها را در فواصل ۱۰ دقیقه‌ای گروه‌بندی کنیم و آن توزیع را رسم کنیم. یک نمونه از این توزیع در جدول ۲-۳ نشان داده شده است.

در نیمه پایین سمت چپ جدول ۲-۳ فاصله‌ها قرار دارند. در این مورد ما حدود بالایی و پایینی فاصله‌ها را به صورت عدد صحیح ارائه کرده‌ایم به این دلیل ساده که جدول برای خواندن آسان‌تر باشد. البته شما بایستی بدانید که حدود واقعی یک فاصله معین (معروف به **حد بالایی واقعی**^۱ و **حد پایینی واقعی**^۲)

1. real upper limit
2. real lower limit

مقادیر اعشاری هستند که در حد میانی بالای یک فاصله و پایین فاصله بعدی واقع شده‌اند. حد پایینی واقعی یک فاصله کمترین مقدار قرار گرفته در آن فاصله به حساب می‌آید. به همین ترتیب، حد بالایی واقعی بیشترین مقدار است. از این رو مثلاً فاصله ۲۰-۲۹ دارای حدود واقعی ۱۹/۵ تا ۲۹/۵ است، زیرا هر مدت زمانی که بالای ۱۹/۵ یا پایین ۲۹/۵ قرار گیرد به آن فاصله گرد خواهد شد. (دانشجویان اغلب بسیار نگرانند که شخصی دقیقاً نمره ۱۹/۵۰۰۰۰ بگیرد که درست در نقطه انفصال دو فاصله قرار گرفته باشد. جای نگرانی نیست. اولاً این رخداد بسیار بعید است. دوم اینکه همیشه می‌توانید به قید قرعه عمل کنید. سوم اینکه چیزهای خیلی مهم‌تری برای نگرانی وجود دارد. (این از مسائلی است که باعث می‌شود مردم فکر کنند آمار گیج کننده، کسل کننده یا هر دو است.)

جدول ۲-۳ توزیع گروه‌بندی شده زمان صرف شده برای بازی‌های الکترونیکی

داده‌های خام (نسبی)			
فراوانی (f)	دقیقه (X)	فراوانی (f)	دقیقه (X)
۱	۱۷	۱۱	۰
۰	۱۸	۰	۱
۱	۱۹	۱	۲
۰	۰	۲	۳
۱	۴۰	۰	۴
۲	۴۱	۱	۵
۱	۴۲	۳	۶
۲	۴۳	۱	۷
۱	۴۴	۱	۸
۰	۴۵	۰	۹
۳	۴۶	۲	۱۰
۱	۴۷	۰	۱۱
۱	۴۸	۱	۱۲
۲	۴۹	۱	۱۳
۰	۵۰	۰	۱۴
۰	۰	۴	۱۵
۰	۰	۰	۱۶
داده‌های گروه‌بندی			
فراوانی تجمعی	فراوانی (f)	نقاط میانی (دقیقه)	فاصله (دقیقه) (X)
۲۰	۲۰	۴/۵	۰-۹
۳۰	۱۰	۱۴/۵	۱۰-۱۹
۳۵	۵	۲۴/۵	۲۰-۲۹
۴۳	۸	۳۴/۵	۳۰-۳۹
۵۷	۱۴	۴۴/۵	۴۰-۴۹
۶۵	۸	۵۴/۵	۵۰-۵۹

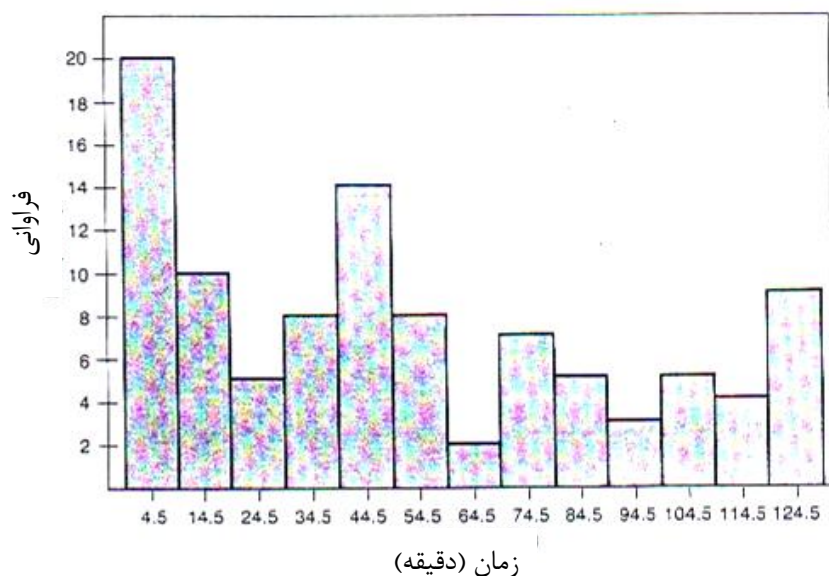
۶۰-۶۹	۶۴/۵	۲	۶۷
۷۰-۷۹	۷۴/۵	۷	۷۴
۸۰-۸۹	۸۴/۵	۵	۷۹
۹۰-۹۹	۹۴/۵	۳	۸۲
۱۰۰-۱۰۹	۱۰۴/۵	۵	۸۷
۱۱۰-۱۱۹	۱۱۴/۵	۴	۹۱
۱۲۰+	۱۲۴/۵*	۹	۱۰۰
		۱۰۰	

* توضیحات در متن آمده است.

در ستون دوم جدول ۲-۳ نقاط میانی فاصله‌ها قرار دارند. اینها همان میانگین حدود بالایی و پایینی هستند و برای آسانی ارائه شده‌اند. وقتی ما داده‌ها را به شکل نمودار در می‌آوریم، ما این نقاط را به عنوان نقاط قرار گرفته در میان فاصله مربوط به آنها رسم خواهیم کرد. بالاترین فاصله بطور دل‌خواه یک نقطه میانی ۱۲۴/۵ در نظر گرفته شده است هر چند حد بالایی واقعی بالاترین طبقه معین نشده است.

در ستون سوم جدول ۲-۳ فراوانی‌ها قرار دارند که نمرات در هر فاصله با آن فراوانی رخ داده‌اند. بنابراین مثلاً هشت نفر بین ۳۰ تا ۳۹ دقیقه در روز وقت صرف بازی‌های الکترونیکی می‌کنند. توزیع درج شده در جدول ۲-۳ به شکل هیستوگرام در شکل ۴-۳ نمایش داده شده است.

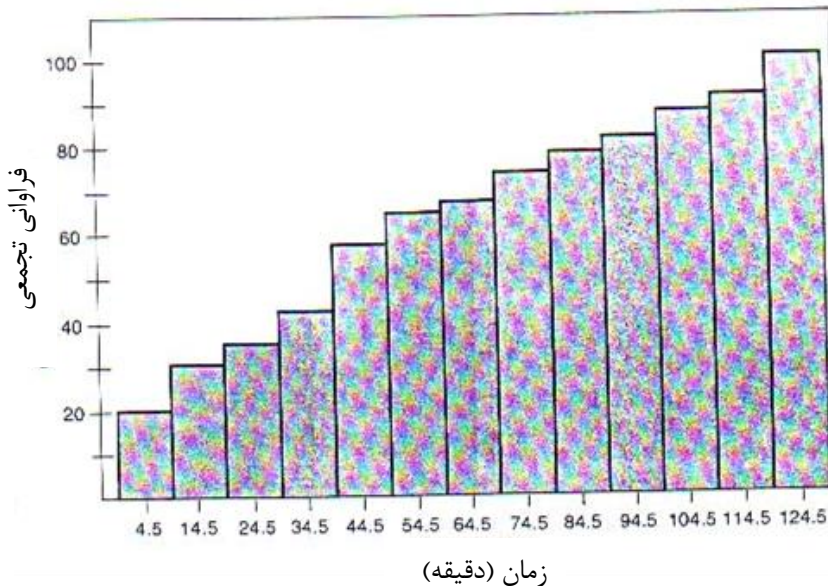
اغلب در باره عدد بهینه فاصله‌ها برای استفاده در هنگام گروه‌بندی سؤال می‌شود. گرچه یک پاسخ درست برای این سؤال وجود ندارد، برخی جاها در حدود ۱۰ فاصله معمولاً بهتر است. در مثال قبلی ما فاصله ۱۳ را در نظر گرفتیم زیرا اعداد به طور طبیعی از این طریق تقسیم می‌شدند. بطور کلی بهتر است هرگاه امکان دارد، نقاط تقسیم طبیعی نظام اعداد به کار برده شود (مثل، ۹-۰، ۱۹-۱۰، ... یا ۱۱۹-۱۰۰، ۱۳۹-۱۲۰، ۱۵۹-۱۴۰ و غیره) به جای اینکه دامنه نمرات را دقیقاً به ۱۰ فاصله تعریف شده دل‌خواه تقسیم کنیم. البته اگر در نظر گرفتن محدودیت‌های دیگری باعث می‌شود داده‌ها آسان‌تر تفسیر شوند، آن شیوه را باید به کار گرفت. بیشتر مهم است که داده‌ها بطور روشن ارائه شوند تا اینکه از قوانین از پیش طرح شده پیروی کنیم.



شکل ۳-۴ هیستوگرام داده‌های بازی الکترونیکی

۳-۳ توزیع‌های تجمعی

تمام توزیع‌هایی که ما تا به حال مورد بحث قرار دادیم مربوط به فراوانی مشاهدات در هر نمره یا درون هر فاصله بودند. یک روش دیگر نگاه به داده‌ها به شکل توزیعی است که تعدادی را که در محدوده یا پایین یک فاصله یا نمره قرار می‌گیرند ثبت می‌کند. برای مثال، شما در جدول ۳-۲ مشاهده می‌کنید که ۲۰ نفر ۹-۰ دقیقه وقت صرف بازی می‌کنند و ۱۰ نفر دیگر ۱۹-۱۰ دقیقه صرف می‌کنند. بنابراین $۲۰+۱۰=۳۰$ نفر ۱۹ دقیقه یا کمتر وقت صرف بازی الکترونیکی می‌کنند. توزیعی که **فراوانی تجمعی** را ترسیم می‌کند **توزیع تجمعی**^۱ نامیده می‌شود، و نمونه‌ای از آن در ستون سمت راست جدول ۳-۲ نشان داده شده و در شکل ۳-۵ ترسیم شده است. توزیع تجمعی نقش مهمی در توضیح رتبه‌های درصدی دارد که در فصل ۴ بیان خواهد شد.



شکل ۳-۵ توزیع تجمعی داده‌های بازی الکترونیکی

۳-۴ نمایش‌های ساقه-برگ

هر چند هر دو توزیع هیستوگرام و فراوانی روش‌های معمول ارائه داده‌ها هستند، هر یک اشکال‌های خاص خود را دارند. هیستوگرام داده‌های گروه‌بندی شده را ترسیم می‌کند، از این رو ارزش‌های عددی واقعی نمرات افراد در هر فاصله را حذف می‌کند. از سوی دیگر، توزیع‌های فراوانی ارزش‌های مشاهدات

1. cumulative distribution

منفرد را حفظ می کنند اما وقتی که داده ها را به خوبی خلاصه نمی کنند استفاده از آنها مشکل است. یک رویکرد جایگزین که از این دو انتقاد برکنار است به عنوان **نمایش ساقه-برگی**^۱ معروف است.

جان توکی (۱۹۷۷) به عنوان بخشی از رویکرد عمومی خود به تحلیل داده ها، معروف به **تحلیل اکتشافی داده ها**^۲ (EDA)، روش های متنوعی از نمایش داده ها را به شیوه های دیداری گویا توسعه داد. یکی از ساده ترین آنها نمایش ساقه-برگی است. اگر به مثال قبلی میزان وقت صرف شده برای بازی الکترونیکی برگردیم، می توان یک شاخه از نمرات را در دهگان ۴۰، دیگری را در دهگان ۵۰ و دیگری را در دهگان ۶۰ و مانند آن در نظر گرفت. این رقم های دهگان ۴، ۵، و ۶ را ما به عنوان اعداد راهنما (گاهی مهم ترین رقم ها خوانده می شوند) در این نمرات منظور می کنیم. این رقم های راهنما ساقه یا محور عمودی نمایش ما را تشکیل می دهند. در آن مجموعه از ۱۴ نمره که در دهگان ۴۰ بودند ما دو ۴۱، یک ۴۲، دو ۴۳ و مانند آن داریم (رجوع کنید به جدول ۲-۳). اینجا واحدهای عددی ۱، ۲ و ۳ را عدهای دنباله (با کم اهمیت ترین) گویند، و برگ های نمایش ما - عناصر افقی را تشکیل می دهند. در شکل ۶-۳ بعضی از داده های ردیفی مربوط به مثال بازی الکترونیکی همراه با کل نمودار ساقه - برگ حاصل شده آمده است.^۳

در شکل ۶-۳ مشاهده می کنید که مثلاً یک ۴۰، دو ۴۱، یک ۴۲، دو ۴۳، یک ۴۴، هیچ ۴۵، سه ۴۶، یک ۴۷، یک ۴۸، و دو ۴۹ وجود داشت. بنابراین برخلاف ساقه ۴ ما یک ۰، دو ۱، یک ۲، و مانند آن می بینیم. با نگاه به این نمایش می توانیم داده های اصلی را کاملاً بازسازی کنیم و به این ترتیب یک توزیع فراوانی کامل را بدرستی داشته باشیم. از سوی دیگر خود نمودار هم شکل جالبی از هیستوگرام چرخیده روی پهلوی را نشان می دهد. مقایسه شکل ۴-۳ و ۶-۳ نکته آخر را روشن می کند.

یک اشکال آشکار این ساده ترین شکل های نمایشی این واقعیت است که در مورد بعضی داده ها ممکن است منجر به گروه بندی شود که خیلی از هدف ما دور است. اگر مثلاً همه نمرات بین ۳۰ و ۳۹ بودند برای ما تنها یک ساقه با سه برگ باقی می ماند. توکی این مسأله را از طریق به کارگیری ساقه *۷ برای بازنمایی فاصله ۷۴-۷۰، و ۷/ برای بازنمایی فاصله ۷۵-۷۹ حل کرد. به همین صورت *۸ و ۸/ به ترتیب نشان دهنده فاصله ۸۴-۸۰ و ۸۹-۸۵ خواهند بود. نظام مشابهی در مورد سایر فاصله ها به کار می رود. قراردادهای دیگری نیز برای گروه بندی های مختلف وجود دارد که ما وارد آنها نمی شویم. علاقه مندان می توانند برای توضیحات تفصیلی به کتاب ولمن و هاگلین (۱۹۸۱) یا توکی (۱۹۷۷) مراجعه کنند.

1. stem-and-leaf display
2. exploratory data analysis

همیشه اینگونه نیست که عدهای دهگان تشکیل دهنده ساقه و رقم های یکان تشکیل دهنده برگ ها باشند. برای مثال اگر دامنه داده ها از ۱۰۰ تا ۱۰۰۰ باشد رقم های صدگان تشکیل دهنده ساقه و رقم های دهگان برگ ها را تشکیل خواهند داد و رقم های یکان را نادیده می گیریم.

شکل ۶-۳ نمودار ساقه-برگ داده‌های بازی الکترونیکی

برگ	ساقه	داده‌های خام
۰۸۷۶۶۶۲۳۰۰۰۰۰۰۰۰۰۰	۰	.
۹۷۵۵۳۲۲۲	۱	.
۷۷۵۳۳	۲	.
۹۹۹۸۷۲۲۲	۳	۴۰ ۴۱ ۴۱ ۴۲ ۴۳
۹۹۸۷۶۶۶۶۱۱۰	۴	۴۳ ۴۴ ۴۶ ۴۶ ۴۶ } →
۹۸۷۵۴۵۲۴	۵	۴۷ ۴۸ ۴۹ ۴۹
۷۳۷	۶	۵۹ ۵۲ ۵۴ ۵۵ ۵۷ } →
۹۸۶۵۵۱۵	۷	۵۸ ۵۹ ۵۹ } →
۹۳۷۷۳	۸	۶۳ ۶۷ } →
۶۶۴	۹	۷۱ ۷۵ ۷۵ ۷۶ ۷۶ } →
۷۷۶۳۲	۱۰	۷۸ ۷۹
۹۳۴۷	۱۱	.
۹۸۷۵۵۲	۱۲	.
۹۸	۱۳	.

نمودارهای ساقه-برگ بخصوص برای مقایسه دو توزیع مختلف سودمند هستند. این کار از طریق ترسیم دو توزیع به طور پشت به پشت بر روی یک ساقه انجام می‌گیرد. شکل ۷-۳ توزیع نمرات پسران و دختران را در یک درس عملی "شیوه‌های آزمایشی" نشان می‌دهد که شامل مؤلفه‌های آماری متعددی است. در این شکل ساقه *۶ ارزش‌های بین ۶۰ و ۶۴، و /۶ ارزش‌های بین ۶۵ و ۶۹ را نشان می‌دهد. کد ذکر شده در زیر شکل گویای آن است که *۴ دنبال شده بوسیله یک ۱ یک ۴۱ است (و نه یک ۴/۱ یا ۴۱۰).

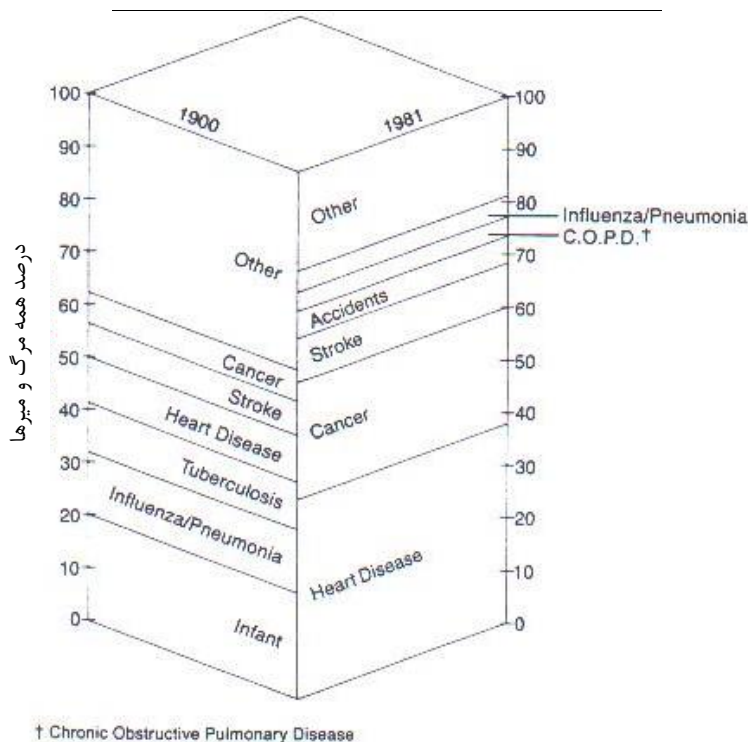
۵-۳ روش‌های دیگر ترسیم داده‌ها

در بخش‌های قبلی با تعداد کمی از روش‌های مختلف ترسیم داده‌ها آشنا شدیم. روش‌های خیلی زیاد دیگری برای این کار وجود دارد که بعضی کاملاً مبتکرانه و پرمعنا هستند. چند نمونه از آنها در شکل ۸-۳ و ۹-۳ نشان داده شده است. اینها را به عنوان نمونه آورده ایم، چون نشان می‌دهند که چگونه می‌توان از نمایش‌ها برای آشکار نمودن ویژگی‌های جالب داده‌ها استفاده کرد.

شکل ۷-۳ نمرات (به درصد) برای یک درس عملی "روش های آزمایشی"، به تفکیک جنسیت

مرد	زن
۳*	۳*
۶	۴*
۴*	۱
۴/	۴/
۵*	۵*
۵/	۵/
۶*	۰.۳
۲	۵۶۸
۶/	۰.۱۴۴
۶۶۶۶۶۶۳۲۲۰۰	۷*
۸۸۸۸۱۷۵۵	۷/
۴۴۳۲۲۱۰۰۰	۸*
۷۶۶۶۶۶۶۵۵۵	۸/
۴۲۲	۹*
	۹/
	۵۶

کد ۴۱ = ۴*



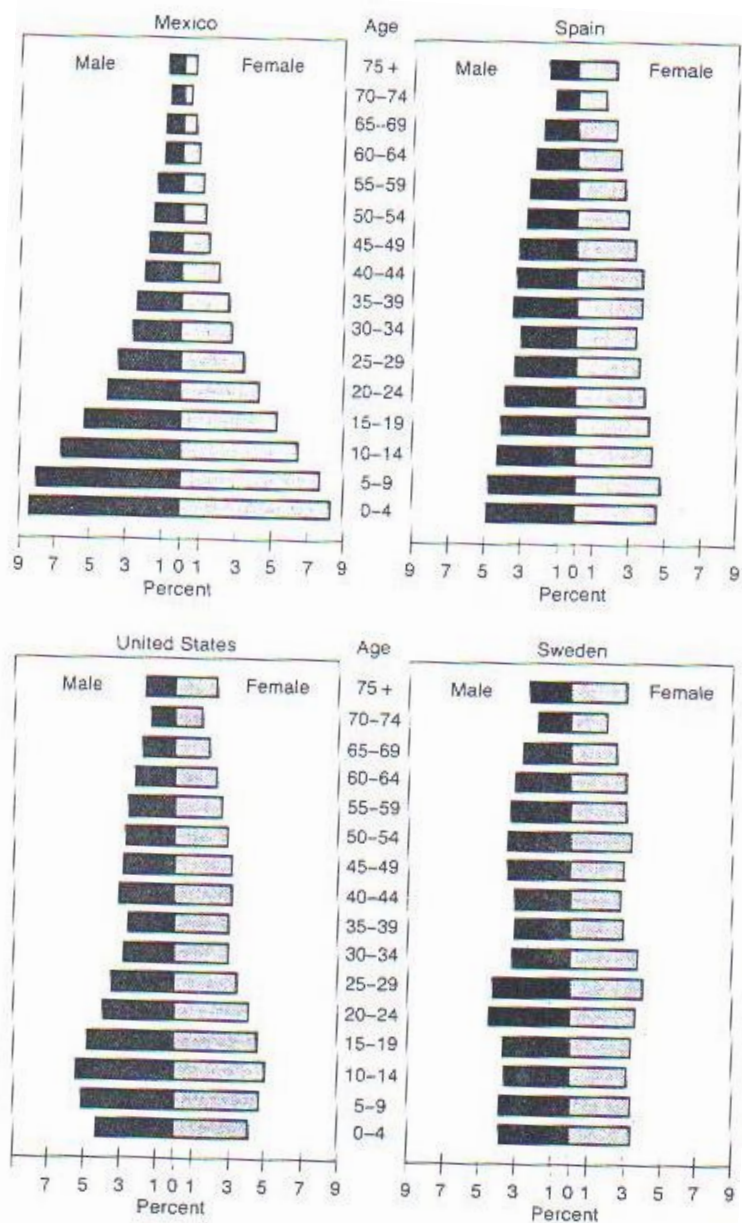
شکل ۸-۳ علل عمده مرگ و میر در ورمونت در ۱۹۰۰ و ۱۹۸۱

(به نقل از: گزارش سالانه آمار حیاتی ۱۹۸۱ ورمونت. اداره بهداشت ورمونت، ۱۹۸۲)

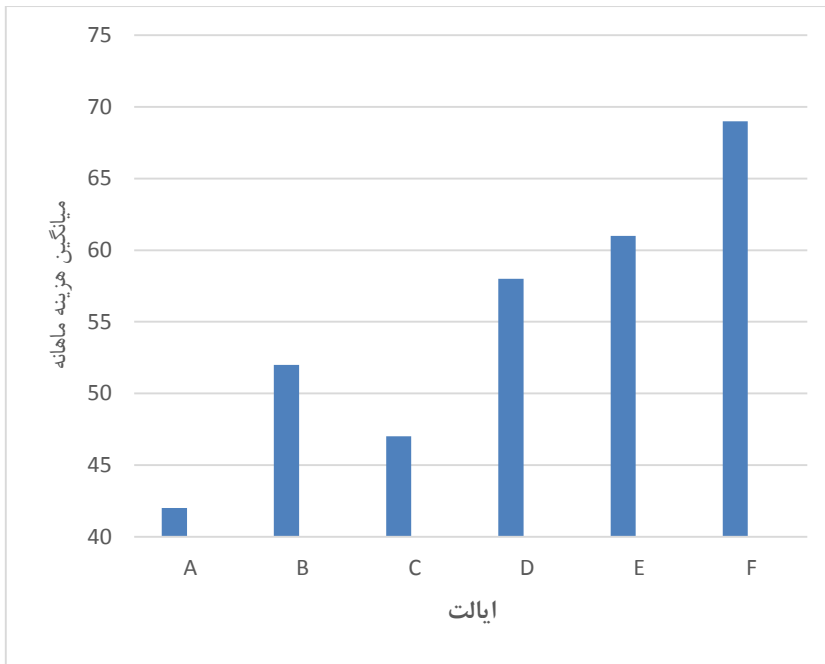
در شکل ۸-۳ مقایسه‌ای بین علل مرگ و میر در میان ساکنان ایالت ورمونت آمریکا در سال‌های ۱۹۰۰ و ۱۹۸۱ دیده می‌شود. توجه کنید که علل مختلفی از پایین به بالا، به ترتیب کاهش میزان آنها به نمایش در آمده است. در یک نگاه به این شکل آشکار می‌شود در حالی که یک سوم همه مرگ و میرها در سال ۱۹۰۰ مربوط به مرگ و میر کودکان، انفلونزا، ذات‌الریه و بیماری سل بوده، هیچ‌کدام از این منابع در نرخ مرگ و میر در ۱۹۸۱ نقش مهمی نداشته‌اند. از سوی دیگر بیماری‌های قلبی و سرطان که باهم ۶۰٪ علل همه مرگ و میرهای ۱۹۸۱ را تشکیل می‌دادند نقش خیلی کمتری در ۱۹۰۰ داشته‌اند، که کمتر از ۱۵٪ علل مرگ و میر در آن سال را تشکیل داده‌اند.

در شکل ۹-۳ توزیع جمعیت کشورهای مکزیک، اسپانیا، ایالات متحده و سوئد بر اساس سن و جنس را می‌بینیم. این شکل به روشنی تفاوت‌های بین کشورها را از لحاظ توزیع سنی آنها به تصویر می‌کشد. (برای نمونه مکزیک و سوئد را مقایسه کنید.) با ترسیم متقابل زن و مرد، اثر تفاوت‌های جنسیتی بر امید زندگی را می‌توانیم مشاهده کنیم. در همه کشورهای چهارگانه گروه‌های سنی پیرتر بیشتر زنان را شامل می‌شوند تا مردان.

شکل ۱۰-۳ در اصل نمونه‌ای است از اینکه داده‌ها چگونه گاهی به شکل درست اما اشتباه برانگیز ارائه می‌شوند. برای این که در فضایی محدود این واقعیت نشان داده شود که مشتریان در یک ایالت چگونه بهای برق کمتری نسبت به میزان یکسانی از مصرف برق (۷۵۰ کیلووات) در ایالت دیگر می‌پردازند، یک شرکت محلی نموداری مشابه شکل ۱۰-۳ را منتشر کرد. هر چند داده‌ها در چنین نموداری بطور درستی ترسیم شده‌اند، این واقعیت که مقادیر روی محور در ۴۰ دلار آشکار می‌سازد که مثلاً مشتریان در ایالت C تقریباً سه برابر (ستون آنها سه برابر بزرگتر است) مشتریان در ایالت A پرداخت می‌کنند. در حقیقت تغییرات عملاً ۴۱/۰۲ و ۴۷/۱۱ دلار بوده، که ایالت C ۱۵٪ نه ۷۰۰٪ بیشتر پرداخته است. برای ارائه چنین مواردی لازم است محور از صفر شروع شود، اگر این عملی نباشد حداقل به اندازه کافی از کوچکترین مقدار فاصله داشته باشد تا از القای تصویر نادرست پرهیز شود.



شکل ۳-۹ جمعیت براساس جنس و سن برای کشورهای برگزیده: ۱۹۷۰ (نقل از شاخصه‌های اجتماعی، ۱۹۷۶. وزارت بازرگانی ایالات متحده، اداره انتشارات دولتی، ۱۹۷۷)

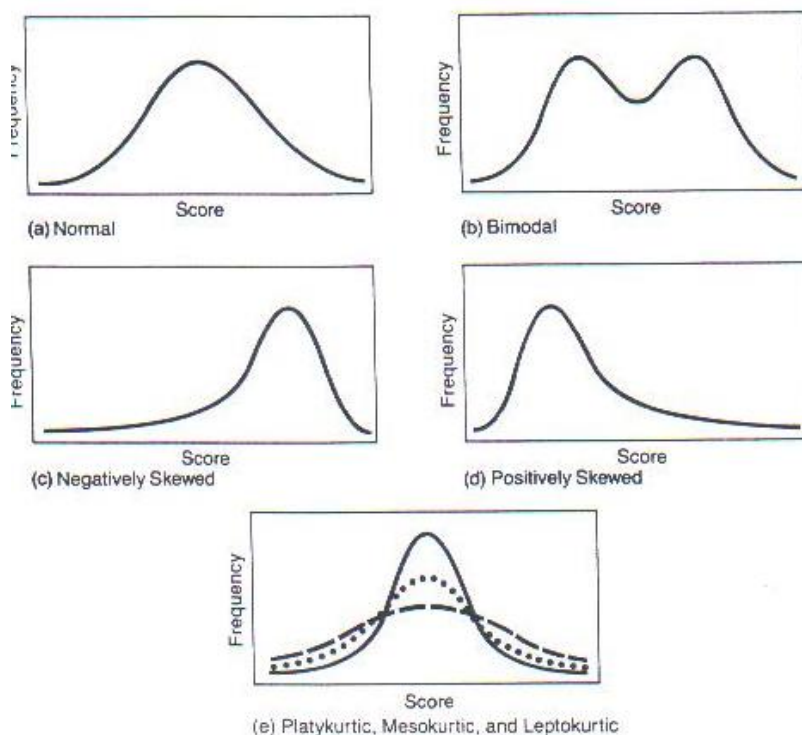


شکل ۱۰-۳ میانگین ماهانه قبض برق در نیوانگلند

۳-۶ توزیع‌های توصیفی

توزیع نمرات نمایش داده شده در شکل ۳-۱ یک توزیع کم و بیش منظم است، که به یک حد بیشینه بالا رفته و دو باره به طور ملایم کاهش یافته است. همه توزیع‌ها اینگونه نیستند، بلکه (شکل ۳-۴ را مشاهده کنید) باید دانست که این عبارت برای توصیف توزیع‌های مختلف به کار می‌رود. به دو توزیع فرضی نشان داده شده در شکل‌های ۳-۱۱ (الف) و (ب) توجه کنید. هر دوی آنها **متقارن**^۱ خوانده می‌شوند چون شکل آنها در دو سوی مرکز شبیه به هم است. توزیع نشان داده شده در شکل ۳-۱۱ (الف) چیزی است که ما بعداً به آن به عنوان توزیع بهنجار (نرمال) رجوع خواهیم کرد. توزیع شکل ۳-۱۱ (ب) را **توزیع دونمایی**^۲ گویند زیرا دارای دو قله است. واژه دونمایی برای هر نوع توزیع که دارای دو قله باشد به کار می‌رود، خواه این دو قله دارای ارتفاع دقیقاً یکسانی باشند یا نباشند. (اگر یک توزیع تنها دارای یک قله بزرگ باشد **تک نمایی**^۳ خوانده می‌شود).

1. symetric
2. bimodal
3. unimodal



شکل ۱۱-۳ انواع توزیع های فراوانی، (الف) بهنجار، (ب) دونمایی (ج) چوله منفی (د) چوله مثبت، (ه) مسطح (خط چین)، معتدل (نقطه چین) و کشیده (خط ممتد).

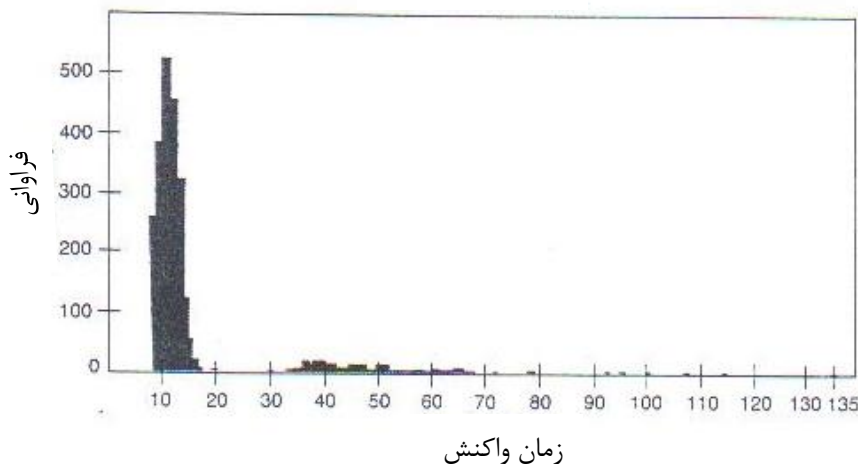
اکنون به شکل های ۱۱-۳ (ج) و (د) توجه کنید. این دو توزیع آشکارا نامتقارن هستند. توزیع شکل ۱۱-۳ (ج) دارای یک دنباله کشیده شده به چپ است در حالی که توزیع شکل ۱۱-۳ (د) در سمت راست امتداد یافته است. ما اولی را دارای **چولگی (کجی) منفی**^۱ و دومی را دارای **چولگی مثبت**^۲ می گوئیم. (نکته: برای به خاطر سپردن، توجه کنید که توزیع دارای چولگی منفی اشاره به نمرات منفی یا کوچک دارد و چولگی مثبت برعکس آن.) هرچند برای میزان چولگی یا نامتقارن بودن، اندازه های آماری وجود دارد اما در علوم اجتماعی اینها مورد استفاده قرار نمی گیرند.

یک نمونه جالب از توزیع دو نمایی چوله مثبت در شکل ۱۲-۳ نشان داده شده است. این داده ها بوسیله برادلی^۳ (۱۹۶۳) طرح شده که به آزمودنی ها یاد داد به محض اینکه نور کوچکی تابانده شد به

1. negatively skewed
2. positively skewed
3. Bradley

سرعت دکمه‌ای را فشار دهند. مشاهده می‌کنید که بیشترین نقاط داده‌ها بطور یکنواخت تقریباً بین صدم‌های ۷ و ۱۷ از یک ثانیه پخش شده‌اند اما خوشه‌ی کوچک ولی قابل توجهی از نقاط بین صدم‌های ۳۰ و ۷۰ کشیده و به سمت راست منتهی شده است. این خوشه دوم در اصل از تمرین‌های آزمودنی هنگامی که تلاش اول او بی نتیجه بوده بدست آمده است. ورود آنها به داده‌ها بطور معناداری شکل توزیع را تحت تأثیر قرار داده است. آزمایشگری که چنین مجموع‌هایی از داده‌ها را جمع کرده شدیداً به زمان‌های از دست رفته بیش از حد خاصی توجه داشته است، بر این اساس که این زمان‌ها بیشتر منعکس کننده درستی یک پاسخ روانشناسی هستند نه اینکه اندازه‌هایی از سرعت پاسخ باشند

توزیع کل تمرین‌ها



شکل ۱۲-۳ توزیع فراوانی نمرات زمان واکنش

آخرین ویژگی از یک توزیع که ما بررسی می‌کنیم کشیدگی است. کشیدگی^۱ تعریف ریاضی خاصی دارد، اما اساساً اشاره دارد به میزانی که نمرات در دنباله‌های توزیع جمع شده‌اند. توزیعی که در آن تعداد زیادی از نمرات بالا و پایین (نسبت به میانگین) وجود دارند به عنوان توزیع دنباله‌دار (یا مسطح^۲) خوانده می‌شود. توزیعی که در دنباله‌ها نسبتاً نازک باشد توزیع کشیده^۳ نامیده می‌شود. توزیع‌هایی مانند توزیع بهنجار که در دنباله‌ها نه پخ و نه نازک باشند را معتدل^۴ گویند.

1. kurtosis
2. platykurtic
3. leptokurtic
4. mesokurtic

این موضوع مهم است که پیش از آنکه بتوانیم در باره شکل یک توزیع - بخصوص کشیدگی آن، ایده خوبی داشته باشیم حجم نسبتاً زیادی از نمونه‌های داده‌ها مورد نیاز است. با نمونه‌هایی با اندازه حدود ۳۰ بهترین چیزی که بطور منطقی می‌توانیم انتظار داشته باشیم این است که آیا داده‌ها متمایل به این هستند که در دنباله‌های توزیع توده‌ای شوند یا اینکه داده‌ها بطور قابل توجهی به یک سمت یا سمت دیگر چوله (کج) شوند.

چولگی و کشیدگی گرچه عموماً در علوم اجتماعی اندازه گرفته نمی‌شوند، برچسب‌های لفظی معمولی هستند که در توزیع‌های توصیفی کاربرد دارند.

۷-۳ خلاصه

در این فصل روش‌های توصیف توزیع‌ها مورد بحث قرار گرفت. همه فنون مورد بحث در اینجا در اصل برای سازماندهی و خلاصه کردن اطلاعات موجود در یک مجموعه بزرگ از داده‌های بخش‌های جزئی‌تر هستند تا به ما امکان دهند اطلاعات را به سرعت با دیگران مبادله کنیم. در بررسی این فنون آماری ما همچنین نگاه مختصری به نمایش‌های ساقه-برگی داشتیم (یک بخش کوچک از تحلیل اکتشافی داده‌های توکی) و تعدادی از عبارات سودمند در تعیین شکل توزیع‌ها را مرور کردیم. بعضی از مطالب مهم فصل عبارتند از:

- توزیع فراوانی
- هیستوگرام
- حد واقعی پایین
- حد واقعی بالایی
- نقاط میانی
- فراوانی تجمعی
- توزیع تجمعی
- نمایش ساقه-برگ
- چولگی (کجی) منفی
- چولگی مثبت
- چولگی (کجی)
- کشیدگی
- تحلیل اکتشافی داده‌های
- رقم‌های راهنما (مهم‌ترین رقم‌ها)
- ساقه
- رقم‌های دنباله
- برگ‌ها
- متقارن
- دونمایی
- تک نمایی
- مسطح
- کشیده
- متعادل

۸-۳ تمرین‌ها

۳-۱ کودکان از این نظر با بزرگسالان متفاوتند که آنها تمایل دارند داستان‌ها را به صورت یک زنجیره اعمال تعریف کنند نه یک طرح کلی. یعنی توصیف آنها از فیلم‌ها پر از عبارات "و بعد. . . " است. یک آزمایشگر پر حوصله از ۵۰ کودک خواست در باره یک فیلم معین با او حرف بزنند. در میان سایر متغیرها او تعداد جملات "و بعد. . . " را شمرد. داده‌ها به شرح زیر است:

۲۰ ۲۰ ۲۱ ۲۰ ۱۸ ۲۳ ۲۱ ۱۷ ۱۶ ۱۶ ۱۲ ۱۷ ۲۰ ۱۸ ۱۷ ۱۸ ۱۹ ۲۲ ۱۵ ۱۸
۳۱ ۱۹ ۱۱ ۲۴ ۱۸ ۱۸ ۲۰ ۲۱ ۱۹ ۱۹ ۱۷ ۱۰ ۲۲ ۲۳ ۲۰ ۱۹ ۱۷ ۱۵ ۱۸
۱۶ ۱۹ ۱۸ ۴۰ ۱۸ ۱۸ ۲۰ ۱۹ ۱۵ ۱۷ ۱۶

الف) یک توزیع فراوانی غیرگروه‌بندی شده از این داده‌ها رسم کنید.

ب) شکل عمومی این توزیع چگونه است؟

۳-۲ یک هیستوگرام برای داده‌های تمرین ۳-۱ با بکارگیری تعداد منطقی از فاصله‌ها بکشید.

۳-۳ مشکل ایجاد یک نمودار ساقه-برگ از داده‌های تمرین ۳-۱ چیست؟

۳-۴ به عنوان بخشی از مطالعه‌ای که در تمرین قبل توضیح داده شد، آزمایشگر داده‌هایی مشابهی را از بزرگسالان بدست آورد. داده‌های آنها به این شرح است:

۹ ۱۰ ۱۴ ۱۷ ۱۲ ۱۵ ۱۱ ۹ ۹ ۱۰ ۱۱ ۱۱ ۷ ۸ ۱۲ ۱۰ ۱۳ ۸ ۵ ۱۲ ۱۰
۵ ۱۲ ۸ ۱۴ ۱۱ ۳ ۱۰ ۹ ۷ ۱۲ ۱۱ ۴ ۱ ۹ ۱۶ ۷ ۱۴ ۱۰ ۱۶ ۱۵ ۸
۹ ۱۵ ۱۰ ۱۴ ۱۱ ۷ ۹ ۱۰

الف) در باره این داده‌ها چه می‌توان گفت؟ آیا کودکان و بزرگسالان داستان‌ها را به یک صورت به یاد می‌آورند.

ب) یک توزیع فراوانی غیرگروه‌بندی شده از این داده‌ها با همان مقیاس بکار گرفته برای کودکان طرح کنید.

ج) این توزیع فراوانی را روی یکی از تمرین‌های ۳-۱ پوشش دهید.

۳-۵ یک هیستوگرام پشت به پشت (مثل شکل ۳-۹) برای مقایسه داده‌ها در دو تمرین ۳-۱ و ۳-۴ رسم کنید.

۳-۶ یک توزیع فراوانی تجمعی برای داده‌های تمرین ۳-۱ بسازید.

۳-۷ یک توزیع فراوانی تجمعی برای داده‌های تمرین ۳-۴ بسازید.

۳-۸ یک چولگی مثبت از داده طرح کنید و نمودار آن را بکشید.

۳-۹ یک حالت دو نمایی از داده‌ها طرح و نمودار آن را بکشید.

۳-۱۰ برای تعداد نخ سیگارهای کشیده شده در روز توسط ۲۰۰ نفری که از این به بعد برخورد می‌کنید چه شکلی از توزیع را پیش‌بینی می‌کنید.

- ۳-۱۳ چه واقعیت جالبی در باره جمعیت مکزیک و اسپانیا می توان در شکل ۳-۹ دید.
- ۳-۱۴ در برخی نمایش های ساقه- برگ با یک یا دو مقدار بالا، آخرین ساقه را اغلب به صورت HI و می نویسجه ند مقادیر کامل در بخش برگ می آید. علت آن ممکن است چه باشد.

فصل ۴

رتبه‌های درصدی

۴-۱- توزیع تجمعی و رتبه درصدی

۴-۲- محاسبه رتبه‌های درصدی برای نمرات

۴-۳- محاسبه نمرات برای رتبه‌های درصدی

۴-۴- مثال کار شده پایانی

در فصل قبل دیدیم که چگونه داده‌ها را برای ایجاد یک برداشت کلی در راستای هدف جمع‌بندی داده‌ها ترسیم کنیم. با نگاهی به شکل کلی توزیع می‌توانیم در باره مقدار اعداد درگیر و اندازه‌هایی که نشان دهد این عددها خیلی به هم فشرده یا پراکنده هستند ایده‌هایی بدست آوریم. بنابراین اگر ما یک مجموعه نمرات امتحانی داشته باشیم درباره آنها به عنوان یک مجموعه می‌توانیم چیزهای زیادی بگوییم. اما بیشتر افراد خود محورند. بسیاری از ما کمتر علاقه‌مندیم که بگوییم کلاس بطور کلی چگونه عمل کرده است تا اینکه نگاه کنیم به اینکه ما در رابطه با باقی کلاس چگونه عمل کرده‌ایم (بخصوص اگر فکر کنیم درست عمل کرده‌ایم). این فصل به روش‌های مشخص کردن عملکرد فرد نسبت به سایر افراد خواهد پرداخت.

۴-۱ توزیع‌های تجمعی و رتبه درصدی

در فصل ۳ دیدیم که توزیع‌های فراوانی را می‌توان بسادگی با شمردن تعداد نمرات مربوط به یک فاصله معین یا پایین‌تر از آن، به توزیع‌های تجمعی فراوانی تبدیل کرد. به عنوان یک نمونه، فرض کنید که ما یک مقیاس ۱۰۰ درجه‌ای از "دوست داشتنی بودن" طرح کرده‌ایم. نمره یک فرد در چنین مقیاسی می‌تواند از رتبه‌بندی او در تعداد زیادی ابعاد (مثل همکاری، اندیشمندی، صمیمیت، صداقت و مانند آن) و جمع کردن آنها بدست آید. با این فرض که چنین مقیاسی روی ۲۶۰ نفر اجرا شده، این داده‌ها در

جدول ۴-۱ تدوین شده و نمراتی بدست آمده است که به طور متعادل در دامنه ۰-۱۰۰، با اندکی گرایش به انباشتگی در مرکز توزیع گسترده شده‌اند.

اولین ستون جدول ۴-۱ شامل فاصله طبقاتی می‌شود که نمرات می‌توانند در آن طیف قرار بگیرند، و ستون دوم شامل فراوانی هر فاصله است. در ستون سوم، فراوانی تجمعی (CF) را مشاهده می‌کنیم. برای مثال، ۱۵ نفر نمره‌ای در فاصله ۲۵-۲۹ گرفته‌اند و نمره ۳۵ نفر در این حد یا کمتر بوده است. آخرین ستون شامل فراوانی تجمعی نسبی (RCF) است. این نسبت‌ها از تقسیم فراوانی تجمعی یک فاصله خاص بر تعداد کل مشاهدات بدست آمده است. برای فاصله ۲۵-۲۹ به این ترتیب است که ۳۵ بر ۲۶۰ تقسیم شده و ۰/۱۳ بدست آمده است. این نتیجه چنین تفسیر می‌شود که یک نسبت ۰/۱۳ یا ۱۳٪ از نمرات در این فاصله یا پایین‌تر از آن قرار گرفته‌اند.

جدول ۴-۱ داده‌های فرضی برای مقیاس دوست داشتنی بودن

فراوانی تجمعی نسبی (RCF)	فراوانی تجمعی (CF)	فراوانی (f)	فاصله طبقه (X)
۱/۰۰	۲۶۰	۵	۹۵+
۰/۹۸	۲۵۵	۱	۹۰-۹۴
۰/۹۸	۲۵۴	۱۰	۸۵-۸۹
۰/۹۴	۲۴۴	۸	۸۰-۸۴
۰/۹۱	۲۳۶	۱۳	۷۵-۷۹
۰/۸۶	۲۲۳	۱۶	۷۰-۷۴
۰/۸۰	۲۰۷	۱۷	۶۵-۶۹
۰/۷۳	۱۹۰	۱۷	۶۰-۶۴
۰/۶۷	۱۷۳	۲۵	۵۵-۵۹
۰/۵۷	۱۴۸	۳۰	۵۰-۵۴
۰/۴۵	۱۱۸	۲۴	۴۵-۴۹
۰/۳۶	۹۴	۲۵	۳۵-۳۹
۰/۲۷	۶۹	۱۶	۳۵-۳۹
۰/۲۰	۵۳	۱۸	۳۰-۳۴
۰/۱۳	۳۵	۱۵	۲۵-۲۹
۰/۰۸	۲۰	۱۰	۲۰-۲۴
۰/۰۴	۱۰	۱	۱۵-۱۹
۰/۰۴	۹	۵	۱۰-۱۴
۰/۰۲	۴	۱	۵-۹
۰/۰۱	۳	۳	۰-۴
		۲۶۰	

فرض کنید که به شما گفته شده در این شاخص دوست داشتنی بودن چه نمره‌ای دریافت کرده‌اید. منطقی است که فرض کنیم نخستین پرسش شما در باره این باشد که نمره‌تان در مقایسه با دیگران چگونه است. همچنین منطقی است که فرض کنیم شما متمایل به این هستید که فکر کنید که نسبت به خیلی از افراد دیگر از رتبه بالاتری برخوردار خواهید بود (هرچند افرادی هستند که تلاش می‌کنند مبدا در پایین مقیاس قرار گیرند).

یکی از بهترین راه‌های مقایسه نمره شما با دیگران این است که از جنبه **نقاط درصدی**^۱ و **رتبه‌های درصدی**^۲ به موضوع نگاه کنیم. نقطه‌ی درصدی نقطه‌ای روی مقیاس است که درصد معینی از نمرات در آن نقطه یا زیر آن قرار دارند. بنابراین، بطور مثال نقطه‌ای که زیر آن ۷۶٪ نمرات قرار می‌گیرند نقطه درصدی ۷۶ ام خوانده می‌شود. باید توجه داشت که یک نقطه درصدی یک نقطه روی مقیاس است. این یک نمره نیست، هرچند به لحاظ عددی برابر با یک نمره است. به عبارت دیگر، نقاط درصدی لازم نیست مانند اشیاء، مقادیر مشاهده شده‌ای باشند. برای مثال، با یک مجموعه معین از داده‌ها ممکن است بگوییم ۹۰٪ نمرات در یا زیر ۷۸/۵ (نقطه درصدی ۹۰ ام) قرار داشته‌اند هر چند کسی یک نمره واقعی معادل ۷۸/۵ نگرفته باشد.

از سوی دیگر، رتبه درصدی یک نمره، درصدی از موارد است که در محدوده آن نمره یا پایین‌تر از آن واقع شده‌اند. از این رو در مثال قبلی رتبه درصدی ۷۸/۵ معادل ۹۰ است. به بیان دیگر اگر یک نمره ۷۱/۲ در نقطه درصدی ۸۲ ام باشد آنگاه نقطه درصدی ۸۲ ام $= ۷۱/۲$ و ۷۱/۲ دارای یک رتبه درصدی ۸۲ ام است. در بخش‌های بعدی ما نخست به نحوه پیدا کردن رتبه درصدی مربوط به یک نمره می‌پردازیم، سپس به مسائل آن بر می‌گردیم و می‌بینیم که چطور یک نمره مربوط به یک رتبه درصدی معین را پیدا کنیم.

محاسبه رتبه‌های درصدی از روی نمرات

برای اینکه یک مثال عینی زده باشیم فرض کنید که نمره شما روی شاخص دوست داشتنی بودن ۷۳ بوده است. ما می‌توانیم توزیع ارائه شده در جدول ۱-۴ را برای کمک به محاسبه رتبه درصدی شما به کار بگیریم. اولین کاری که انجام خواهیم داد این است که فاصله طبقه‌ای که نمره شما در آن قرار گرفته را معین کنیم. برای یک نمره ۷۳ این فاصله ۷۴-۷۳ است کار دومی که انجام خواهیم داد ترکیب نقاط داده‌ها در سه مقوله است: آنهایی که در فاصله شما قرار می‌گیرند، آنهایی که در کل فواصل پایین‌تر قرار می‌گیرند، و آنهایی که در فواصل بالاتر قرار می‌گیرند. در جدول ۱-۴ از فراوانی‌های تجمعی نسبی درمی‌یابیم که ۸۰٪ در فواصل زیرین قرار می‌گیرند. زیرا ۸۶٪ در فاصله مشابه یا پایین‌تر و ۸۰٪ در فواصل

پایین تر، $\% = ۸۰ - ۸۶$ بایستی در فاصله شما قرار گیرند. به همین ترتیب، چون $\% ۸۶$ در یا زیر فاصله شما قرار دارند، ۱۴ درصد باقی مانده افراد بایستی بالای آن و بنابراین دوست داشتنی تر باشند (کسانی که می خواهند کامل باشند).

کل فواصل بالاتر ($\%$ بالاتر) $\% ۱۴$

فاصله مشابه ($۷۴ - ۷۰$) ($\%$ فاصله) $\% ۶$

همه فواصل پایین تر ($\%$ پایین تر) $\% ۸۰$

از این نتایج معلوم می شود که دست کم $\% ۸۰$ افراد کمتر از شما دوست داشتنی بودند. علاوه بر این، $\% ۶$ دیگر در فاصله شما بودند. مگر اینکه شما یک خودپنداری معیوب داشته باشید که تصور کنید همه این $\% ۶$ واقعا دوست داشتنی تر از شما هستند. از سوی دیگر شما آنقدر هم خودشیفته نیستید که فکر کنید همه از شما پایین ترند. بنابراین بعضی از این $\% ۶$ احتمالا نمره پایین تر و بعضی بالاتر از شما گرفته اند. به طریقی ما باید تصمیم بگیریم آن $\% ۶$ را چگونه تقسیم کنیم.

برای اینکه برآوردی منطقی از جایگاه خود در مقایسه با افراد دیگری که در این فاصله قرار دارند به عمل آورید، ما از طیف خطی ساده استفاده می کنیم. یعنی هر چه نمره شما در این فاصله بالاتر باشد، فرض می شود افراد بیشتری پایین تر از شما قرار دارند. برای استفاده از طیف خطی ما مجبوریم از حد پایین واقعی این فاصله که پیش از این (در فصل ۳) توضیح داده شد استفاده کنیم. گرچه ما بر اساس اینکه فاصله از $۷۴ - ۷۰$ است عمل می کنیم، اما حدود واقعی $۷۴/۵ - ۶۹/۵$ هستند.

با یک حد واقعی پایین $۶۹/۵$ و یک پهنه فاصله $۷۴/۵ - ۶۹/۵ = ۵$ واحدی، یک نمره ۷۳ به اندازه $۵ \div ۵ = ۱$ از حد پایین $۶۹/۵$ است. بنابراین $\% ۷۰ = ۰/۷۰ = ۳/۵ \div ۵ = ۳/۵$ بالاتر از کف فاصله است. بنابراین منطقی است که فرض کنیم شما $\% ۷۰$ از افراد را پشت سر گذاشته اید. چون $\% ۶$ افراد در فاصله شما هستند و فرض شد که شما $\% ۷۰$ آنها را پشت سر گذاشته اید به شما این اعتبار داده شد که $\% ۴/۲ = (۰/۷۰)(۰/۶)$ از کل نمونه بالای $\% ۸۰$ را در کل فاصله های پایین پشت سر گذاشته اید. بنابراین نمره ۷۳ شما در نقطه درصدی $۸۴/۲ = ۸۰ + ۴/۲$ قرار دارد. به بیان دیگر، می توان گفت یک نمره ۷۳ دارای رتبه درصدی $۸۴/۲$ است.

همه این محاسبات را می توان در یک فرمول کاملاً ساده قرار داد تا به خاطر سپردن آن آسان باشد. فهم منطق موضوع مهم تر از حفظ یک فرمول است، زیرا اگر شما منطق آن را بفهمید می توانید به جواب درست برسید خواه فرمول را بدانید یا ندانید. اما به هر حال داشتن یک فرمول حس اطمینان خاصی دارد.

$$(\% \text{ فاصله}) = \frac{\text{حد پایین واقعی} - \text{نمره}}{\text{عرض}} + \% \text{ پایین تر} = \text{رتبه درصدی}$$

درصد نمرات در کل فاصله های پایین تر = $\%$ پایین تر

در جایی که

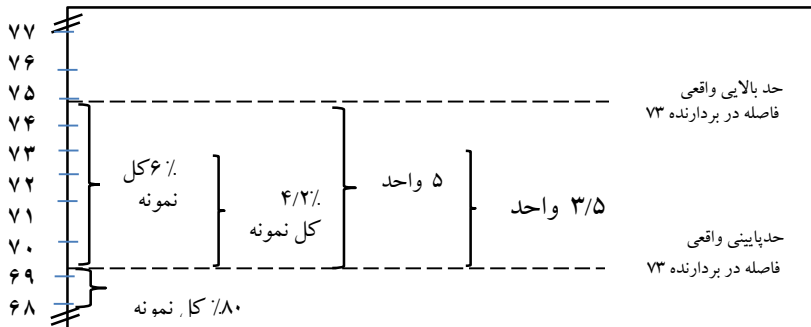
درصد نمرات در همان فاصله = $\%$ فاصله

عرض = عرض فاصله

در مثال ما

$$\begin{aligned} \text{رتبه درصدی} &= \frac{73-69}{5} (\%) + .80 \\ &= 80\% + 4/2\% = 84/2\% \end{aligned}$$

بنابراین تقریباً ۸۴٪ از افرادی که این مقیاس روی آنها اجرا شده کمتر از شما دوست‌داشتنی بوده‌اند. یک شیوه دیگر برای آنچه گفته شد در شکل ۱-۴ نشان داده شده است. در اینجا شما یک بازنمایی دیداری از فرایند مبحث قبل را مشاهده می‌کنید.



شکل ۱-۴ بازنمایی دیداری مثال شاخص دوست‌داشتنی بودن

در مثال قبلی ما با موقعیتی سروکار داشتیم که داده‌ها در فاصله‌هایی با عرض ۵ واحد گروه‌بندی شده بودند. این روند فرقی نخواهد کرد اگر داده‌ها به شکل غیرگروه‌بندی شده باشند (مثل حالتی که فاصله به اندازه یک واحد باشد). برآورد رتبه درصدی شما قدری دقیق‌تر می‌شود اما روش یکسان خواهد بود. در حالت داده‌های گروه‌بندی نشده استفاده از شیوه یاد شده دارای اثر اعتباربخشی به یک فرد با پشت‌سر گذاشتن نیمی از افرادی است که به این نمره پیوند دارند. اما این مشکلی را بوجود می‌آورد که اغلب برای دانشجویان، اساتید و نویسندگان کتاب‌ها گیج‌کننده می‌شود (و به همین دلیل ما برای مدت طولانی از کار با نقاط درصدی تنفر داشتیم). خوب است این مشکل در درس‌ساز را مورد توجه قرار دهیم. اگر در چند کتاب فصل نقاط درصدی را مطالعه کنید خواهید دید که نویسندگان با این موضوع به طور متناقضی برخورد می‌کنند (حتی در یک کتاب). ما رتبه درصدی را به عنوان درصد نمرات در آن حدود یا زیر آن تعریف کردیم. اما برخی کتاب‌ها مانند کتاب حاضر نیمی از افرادی که با آنها در آن فاصله مشترک هستید را "در پایین نمره شما" تعیین می‌کنند. (بعضی نویسندگان دیگر هر دو را انجام می‌دهند بدون اینکه

سعی کنند شیوه کار را بیان کنند). با این بی‌ثباتی تعجبی ندارد که دانشجویان گیج شوند و نسبت به نقاط درصدی حس بدی پیدا کنند. در هر حال چند راه برای خروج از این مشکل وجود دارد. اول از همه باید به این واقعیت پی ببرید که آنچه ما انجام می‌دهیم بیشتر بر اساس قرارداد است. روش "درست" برای رفع مشکلات هنگام محاسبه رتبه‌های درصدی وجود ندارد. بلکه روش‌های متفاوتی برای مواجهه با آنها وجود دارد. ما پیشنهاد می‌کنیم که فرض شود نیمی از افراد مشترک در یک نمره معین، در آن نمره یا زیر آن قرار گیرند. اگر بخواهید فرض بگیرد که همه این افراد "در یا زیر" آن نمره قرار گیرند، درست نیست - تنها یک راه متفاوت را پیش گرفته‌اید. (فقط مطمئن شوید که هرکسی که نمره‌گذاری می‌کند همان قرارداد را دنبال کند).

رویکرد دوم برای این مشکل آن است که نقطه درصدی را یک نقطه (نه یک نمره یا یک فاصله) در نظر بگیرید؛ و نقاط بطور نامحدودی کوچک باشند. اگر یک نقطه درصدی را یک نقطه در نظر بگیرید نه یک نمره، آنگاه آسان‌تر خواهد بود تا خود را متقاعد کنید که افراد دارای نمره مشابه واقعا در بالای آن نمره و دیگران واقعا در پایین آن نمره به حساب می‌آیند. دسته زیادی پری در حال رقص روی سر سنجاقی را تصور کنید که در برخی نقاط درصدی روی مقیاس چسبیده است. بعضی از پری‌ها به آرامی در سمت چپ سر سنجاق و برخی دیگر در راست آن خواهند بود. (اینکه دقیقا چند پری در کدام سمت است را نمی‌توان تعیین نمود مگر اینکه بدانیم چند پری روی آن هستند - می‌گردیم تا ببینیم!)

وقتی همه چیز گفته و انجام شد، رویکردی را که شما برای نمرات درهم تنیده پیش می‌گیرید بیش از یک اثر عملی نخواهد داشت. احتمالا شما از دانستن اینکه نمره‌تان دارای رتبه درصدی $86/3$ است چندان خوشحال‌تر نمی‌شوید تا اینکه بدانید دارای یک رتبه درصدی $84/7$ هستید. اگر این درست باشد آنگاه چرا موافق نباشیم که به توافق برسیم، که درک کنیم ممکن است دیگران توافق‌های دیگری داشته باشند، و آگاه باشیم که برخی افراد این روش‌ها را بطور متناقضی به کار می‌گیرند. وقتی چیزهای جالب زیاد و مهمی وجود دارد لازم نیست دانشجویان در مورد نقاط دقیق درصدی خیلی مته روی خشخاش بگذارند.

۳-۴ محاسبه نمرات برای رتبه‌های درصدی

ما دیدیم که چگونه رتبه درصدی برای نمره خاصی حساب می‌شود، اما گاهی می‌خواهیم بر گردیم ببینیم که یک رتبه خاص با چه نمره‌ای تطبیق دارد. برای مثال، ممکن است شما بخواهید بدانید برای اینکه در مقیاس دوست‌داشتنی بودن به رتبه درصدی ۹۵ برسید چه نمره‌ای لازم است. به عبارت دیگر، چه نمره‌ای با نقطه درصدی ۹۵ اُم تطبیق دارد؟

از جدول ۴-۱ می‌توان دریافت که ۹۴٪ نمرات در یا پایین‌تر از فاصله ۸۴-۸۰ قرار می‌گیرند، بنابراین نمره مورد نظر باید بالاتر از $84/5$ (حد بالایی واقعی این فاصله و حد پایینی واقعی فاصله بالاتر)

باشد. در واقع لازم است شما یک نقطه درصدی به فاصله بعدی بروید. همچنین می‌توانید مشاهده کنید که ۴٪ نمرات در فاصله واقعی ۸۹/۵-۸۴/۵ قرار می‌گیرند و عرض این فاصله ۵ واحد است. اگر آن ۴ نقطه درصدی حتی در این فاصله ۵ نقطه قرار گیرند هر $1/25 = 4 / 5$ نقطه برای یک نقطه درصدی اضافی به حساب می‌آید. بنابراین اگر ۹۴٪ از افراد در یا زیر نمره ۸۴/۵ قرار گیرند و اگر برای رسیدن به ۱٪ بعدی یک واحد $1/25$ اضافی نیاز داشته باشیم، آنگاه نقطه درصدی ۹۵ ام منطبق با نمره $85/75 = 1/25 + 84/5$ خواهد بود.

در مواقعی که شما بخواهید یک نمره را به نقطه درصدی خاصی اختصاص دهید، روشی که ما در اینجا به کار بردیم کفایت می‌کند. اما به هر حال ما می‌توانیم همه اینها را در یک فرمول بگذاریم تا کارمان را روشن‌تر و راحت‌تر کنیم. برای این کار می‌توان چنین تعریف کرد:

$$p \text{ (٪ پایین تر -)} = \frac{\text{عرض}}{\text{فاصله}} + \text{حد پایینی واقعی} = \text{نمره } p$$

در جایی که

نمره منطبق با نقطه درصدی p = نمره p

عرض فاصله = عرض

درصد نمرات در فاصله مورد نظر = ٪ فاصله

رتبه درصدی مورد نظر = p

درصد نمرات زیر فاصله مورد نظر = ٪ پایین‌تر

برای مثال

$$p \text{ نمره} = 84/5 + \frac{5}{4\%} (95\% - 94\%)$$

$$= 84/5 + 1/25 = 85/75$$

گروه‌های مرجع

حتما در نظر داشته باشید که رتبه‌های درصدی تنها در متن گروه مرجعی که از آن بدست آمده‌اند معنا پیدا می‌کنند. برای مثال یک نمره در نقطه درصدی ۱۵ ام مربوط به امتحان پیشرفت دبیرستان، اشاره دارد به اینکه آن فرد در مقایسه با سایر همکلاسانی که در آن امتحان شرکت داشته‌اند چه رتبه‌ای دارد. به ظاهر یک نمره نمی‌توان اکتفا کرد. رتبه ۱۵ ام برای پیشرفت انگلیسی یک دانشجوی انگلیسی زبان نمره خوبی نیست اما برای دانشجوی تازه واردی که دارای زبان مادری متفاوتی است می‌تواند نمره کاملا خوبی باشد. بنابراین هنگام تفسیر رتبه‌های درصدی مهم است که به گروه مرجعی که رتبه‌ها روی آن

اعمال شده توجه شود. یک نمره یکسان وقتی برای گروه‌های مرجع متفاوت مورد قضاوت قرار می‌گیرد دارای رتبه درصدی متفاوتی خواهد بود.

مقیاس‌های مرتبط

چنانکه دیدیم نقاط درصدی، یک توزیع نمرات را به ۱۰۰ بخش مساوی تقسیم می‌کنند. ما همچنین شماری از مقیاس‌های مشابه داریم که توزیع را به صورت متفاوتی (دشوارتر) تقسیم‌بندی می‌کنند. بنابراین مثلاً نقاط دهکی واحدهای ده‌گانه را ایجاد می‌کنند در حالی که نقاط چارکی چهار واحد برابر بوجود می‌آورند. کسی که در نقطه‌ی دهکی هشتم است دارای نمره‌ای است که برابر یا فراتر از $8 / 10 = 80\%$ همه افرادی است که آزمون داده‌اند، درحالی که نمره‌ای در نقطه‌ی چارکی سوم برابر یا فراتر از $3 / 4 = 75\%$ کل نمرات است. لازم است توجه شود که نقاط دهکی و چارکی مانند نقاط درصدی، نقطه هستند. (نقطه چارکی اول و سوم همچنین در اصل از جهاتی معادل هم هستند که در فصل‌های بعدی مورد بحث قرار خواهد گرفت.) درست همانگونه که می‌گوییم نمره یک شخص در نقطه درصدی ۴۳ اُم است، همینطور هم می‌گوییم یک شخص در نقطه‌ی چارکی دوم یا در نقطه‌ی دهکی هفتم است. ما نمی‌گوییم نمره یک شخص درون نقطه چارکی سوم است، زیرا یک نمره نمی‌تواند درون یک نقطه باشد. این خطا در مورد نقاط درصدی (که زیاد هستند) بندرت رخ می‌دهد اما در مورد نقاط چارکی خیلی پیش می‌آید. با این وجود یک خطا است. ما می‌توانیم بگوییم که نمره این شخص بین نقطه‌ی چارکی دوم و سوم قرار می‌گیرد.

۴-۴ مثال کار شده پایانی

توزیع نمرات دانشجویان در یک درس به شرح زیر است. می‌خواهیم مشخص کنیم رتبه درصدی منطبق با نمره ۷۲ چند است، و چه نمره‌ای با نقطه درصدی ۹۰ ام تطبیق می‌کند. داده‌ها به این شرح است:

نمره	فراوانی	% فاصله	% تجمعی
۹۵-۹۹	۲	۱/۶	۱۰۰
۹۰-۹۴	۱۰	۸/۰	۹۸/۴
۸۵-۸۹	۱۶	۱۲/۸	۹۰/۴
۸۰-۸۴	۲۱	۱۶/۸	۷۷/۶
۷۵-۷۹	۱۲	۹/۶	۶۰/۸
۷۰-۷۴	۲۶	۲۰/۸	۵۱/۲
۶۵-۶۹	۱۵	۱۲/۰	۳۰/۴
۶۰-۶۴	۱۶	۱۲/۸	۱۸/۴
۵۵-۵۹	۵	۴/۰	۵/۶
۵۰-۵۴	۱	۰/۸	۱/۶
۴۵-۴۹	۱	۰/۸	۰/۸
۱۲۵			

قدم نخست، محاسبه درصدها در فاصله است. چون ۱۰ دانشجو در فاصله ۹۰-۹۴ قرار گرفته‌اند، ۱۰ / ۱۲۵ = ۸٪ نمرات در آن فاصله قرار داشته‌اند. اندازه‌های همه فاصله‌ها در جدول دیده می‌شود. درصدهای تجمعی نیز در این جدول نشان داده شده و به شیوه افزودن اندازه‌های ستون ٪ در یا زیر ردیف مورد نظر بدست آمده‌اند.

برای پیدا کردن رتبه درصدی مربوط به نمره ۷۲، توجه کنید که ۳۰/۴٪ نمرات در یا زیر یک حد پایینی واقعی ۶۹/۵ قرار دارند. همچنین توجه کنید که عرض فاصله ۵ واحد است (مثلاً، ۵ = ۶۹/۵ - ۷۴/۵) و اینکه ۷۲ به اندازه ۲/۵ نقطه بالاتر از حد پایین واقعی برای این فاصله است. بنابراین ما فرض می‌گیریم که ۵۰٪ = ۵ / ۲۰.۵ از ۲۰/۸٪ نمرات در فاصله ۷۴-۷۰ واقعاً زیر ۷۲ قرار دارند. از این‌رو رتبه درصدی ۷۲ برابر با $۴۰/۸ = ۱۰/۴ + ۳۰/۴ = (۲۰/۸)(۰/۵) + ۳۰/۴$ خواهد بود. با استفاده از فرمول زیر:

$$\begin{aligned} \% \text{ فاصله} &= \left(\frac{\text{حد پایینی واقعی} - \text{نمره}}{\text{عرض}} \right) + \% \text{ پایینی} = \text{رتبه درصدی} \\ &= ۳۰/۴ + \left(\frac{۷۲ - ۶۹.۵}{۵} \right) (۲۰/۸) \\ &= ۳۰/۴ + \left(\frac{۲.۵}{۵} \right) (۲۰/۸) = ۳۰/۴ + \frac{۲۰.۸}{۲} = ۴۰/۸ \end{aligned}$$

برای برگرداندن سؤال و پرسیدن از نمره منطبق با نقطه درصدی ۹۰ ام، ما ابتدا در نظر می گیریم که ۷۷/۶٪ از نمرات زیر یک حد پایینی واقعی ۸۴/۵ و ۹۰/۴٪ نمرات زیر یک حد پایینی واقعی ۸۹/۵ قرار می گیرند. فاصله ۸۹/۵ - ۸۴/۵ شامل ۱۲/۸٪ از نقاط است. ما باید به اندازه ۱۲/۴ = ۷۷/۶٪ - ۹۰٪ نقاط درصدی در این فاصله پیش برویم تا به ۹۰٪ برسیم. اگر فاصله ۵ واحد باشد و ما نیازمند ۱۲/۸٪ (۱۲/۴ / ۸۹/۳) آن باشیم، آنگاه باید ۴/۸۴ = ۱۲/۸ / ۵ (۱۲/۴) واحد پیش برویم. پس نقطه درصدی ۹۰ ام ۸۹/۳ = ۸۴/۵ + ۴/۸۴ است. بنابراین نقطه درصدی ۹۰ ام در یک اندازه نقاط ۸۹/۳ است. انجام این کار با فرمول:

$$\begin{aligned} \% \text{ پایینی} - p &= \left(\frac{\text{عرض}}{\% \text{ فاصله}} \right) + \text{حد پایینی واقعی} = \text{نمره } p \\ &= ۸۴/۵ + \left(\frac{۵}{۱۲.۸} \right) (۹۰ - ۷۷/۶) = ۸۴/۵ + \frac{۵ (۱۲.۴)}{۱۲.۸} = ۸۹/۳ \end{aligned}$$

۴-۵ خلاصه

در این فصل ما راه های نشان دادن نمرات منفرد در مقایسه با نمرات افراد دیگر را بررسی کردیم. مشاهده کردیم که رتبه درصدی مربوط به یک نمره خاص را چگونه حساب کنیم و همچنین نحوه محاسبه نمره منطبق با یک رتبه درصدی معین را آموختیم. علاوه بر این برخی مشکلات عمومی افراد در فهم کامل نقاط درصدی را بررسی کردیم و دو مقیاس مرتبط را از نظر گذراندیم. بعضی از مطالب مهم فصل عبارتند از:

- نقطه درصدی
- نقاط دهکی
- رتبه درصدی
- نقاط چارکی

تمرین‌های ۴-۶

- ۴-۱ اگر ۸۳٪ از افرادی که در یک آزمون شرکت کرده‌اند دارای نمره کمتر از ۷۴ باشند، ۸۳
 از ۷۴ خوانده می‌شود و ۷۴ را می‌نامیم.
- ۴-۲ مثال ساده‌ای طرح کنید که فرق واژه‌های نقطه درصدی و رتبه درصدی را بطور روشن نشان دهد.
- ۴-۳ با استفاده از داده‌های جدول ۴-۱ نقطه درصدی ۳۶م را پیدا کنید.
- ۴-۴ با استفاده از داده‌های جدول ۴-۱ نقطه درصدی ۵۰م را پیدا کنید.
- ۴-۵ اولین نقطه چارکی (Q_1) داده‌های جدول ۴-۱ چیست؟
- ۴-۶ رتبه درصدی ۴۶ برای داده‌های جدول ۴-۱ کدام است؟
 داده‌های زیر را برای تمرین‌های بعد به کار ببرید:
- | | | | | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| ۲۰ | ۱۳ | ۱۳ | ۱۶ | ۱۶ | ۱۴ | ۱۴ | ۱۸ | ۱۲ | ۱۳ | ۱۹ | ۱۲ | ۸ | ۱۴ | ۱۳ | ۱۴ | ۱۵ | ۱۳ | ۱۸ |
| ۱۰ | ۱۱ | ۱۵ | ۱۱ | ۱۱ | ۱۵ | ۱۸ | ۱۲ | ۶ | ۳ | ۶ | ۱۴ | ۱۶ | ۱۲ | ۱۰ | ۱۳ | ۱۵ | ۱۶ | ۱۰ |
| ۲۱ | ۱۰ | ۱۶ | ۱۵ | ۱۳ | ۱۰ | ۱۲ | ۱۶ | ۱۴ | ۶ | ۳ | ۱۲ | ۱۸ | ۱۴ | ۱۶ | ۱۵ | ۱۱ | ۱۱ | ۱۵ |
| ۸ | ۱۵ | ۱۰ | ۱۰ | ۱۲ | ۱۳ | ۹ | ۱۳ | ۱۲ | ۱۳ | ۷ | | | | | | | | |
- ۴-۷ توزیع فراوانی تجمعی این داده‌ها را رسم کنید.
- ۴-۸ نقطه درصدی ۵۰م را پیدا کنید. نام دیگر آن چیست؟
- ۴-۹ نقطه درصدی ۲۱م را پیدا کنید.
- ۴-۱۰ رتبه درصدی نمره ۲۰ چیست؟
- ۴-۱۱ رتبه درصدی نمره ۱۳ کدام است؟
- ۴-۱۲ اگر یک توزیع خیلی چوله باشد، آیا نقطه درصدی ۴۰م و ۶۰م به یک اندازه از میانه فاصله دارند؟
- ۴-۱۳ نموداری شبیه به شکل ۴-۱ رسم کنید تا محاسبات بخش ۴-۴ را به تصویر بکشد.

فصل ۵

شاخص‌های گرایش مرکزی

۱-۵- نما

۲-۵- میانه

۳-۵- میانگین

۴-۵- کاستی‌ها و برتری‌های نسبی نما، میانه

در فصل ۳ دیدیم که چطور داده‌ها را به روشی نمایش دهیم که تا بتوانیم در باره چگونگی آنها نتیجه‌گیری کنیم. رسم داده‌ها شکل کلی توزیع را نشان می‌دهد و یک برداشت دیداری از مقدار عدهای درگیر به ما می‌دهد.

در این فصل آماره‌های مختلفی را خواهیم دید که می‌توانیم برای نمایش "مرکز" توزیع به کار ببریم. این آماره‌ها شاخص‌های گرایش مرکزی^۱ نامیده می‌شوند، که یک برجسته کاملاً توصیفی است. در فصل بعدی ما یک گام جلوتر رفته و نگاهی خواهیم انداخت به شاخص‌هایی که با چگونگی پراگندگی نمرات در اطراف یک شاخص گرایش مرکزی سر و کار دارند، اما ابتدا بایستی به سراغ پرسش قبلی در باره شناسایی مرکز توزیع برویم.

سه نوع شاخص معمول برای گرایش مرکزی وجود دارد (نما، میانه و میانگین)، که به نوبت آنها را توضیح خواهیم داد. با چیزی شروع می‌کنیم که احتمالاً کمترین کاربرد (و کمترین فایده) را دارد.

۵-۱ نما

نما^۱ (MO) را به سادگی می‌توان به عنوان عمومی‌ترین نمره-نمره‌ای که از تعداد زیادی از آزمودنی‌ها بدست آمده است تعریف کرد. بنابراین نما آن مقدار X است که با بالاترین نقطه روی توزیع منطبق است. در مثال فصل ۳ مربوط به رتبه‌بندی مطلوبیت یک مرکز دانشجویی (جدول ۱-۳)، این مقدار ۶ است، زیرا بیشتر افراد، به جای هر مقدار دیگری این مرکز را در حد ۶ رتبه‌بندی کرده بودند.

اگر دو رتبه مجاور با فراوانی برابر (و بیشترین فراوانی) رخ دهند، یک شیوه معمول این است که معدل آن دو مقدار را بدست آوریم و به عنوان نما بخوانیم. از سوی دیگر چنانچه دو رتبه‌بندی غیر مجاور با فراوانی برابر یا تقریباً برابر رخ دهند، ما توزیع را دونمایی خوانده و تمایل داریم که هر دو نما را گزارش کنیم. بنابراین مثلاً داده‌های بازی الکترونیکی دو نمایی با قله‌هایی در ۹- و ۴۹-۴۰ دقیقه هستند. (ممکن است بگویید این توزیع سه‌نمایی است با قله سوم در ۱۲۰+، اما چون این فاصله مربوط به مجموع "سایر مقادیر" است، به عنوان یک ارزش نمایی در نظر گرفته نمی‌شود.

۵-۲ میانه

میانه به عنوان نمره‌ی متناظر با نقطه‌ای که ۵۰٪ از نمرات در بالا یا زیر آن قرار گرفته‌اند تعریف می‌شود. در حالی که داده‌ها به صورت عددی مرتب شده‌اند. با این تعریف میانه همچنین نقطه درصدی ۵۰ ام است. برای مثال عددهای (۵ ۸ ۳ ۷ ۱۵) را در نظر بگیرید. اگر این عددها به ترتیب عددی مرتب شوند (۱۵ ۸ ۷ ۵ ۳)، نیمی از آنها زیر ۷ قرار می‌گیرند و ۷ را میانه خواهیم نامید. فرض کنید که تعداد عددها زوج باشد - مثل (۵ ۱۱ ۳ ۷ ۱۵ ۱۴). با مرتب کردن آنها خواهیم داشت (۵ ۱۴ ۱۵ ۱۱ ۷ ۳)، و نمره‌ای که ۵۰٪ زیر آن باشد وجود ندارد. آن نقطه در واقع بین ۷ و ۱۱ قرار دارد. در این حالت میانگین آن دو نمره میانی (۷ و ۱۱) به عنوان میانه منظور می‌شود (یعنی عدد ۹).

یک اصطلاح که به آن نیاز داریم **جایگاه میانه** است. جایگاه میانه‌ی N عدد به این صورت تعریف می‌شود

$$\text{جایگاه میانه} = \frac{N+1}{2}$$

بنابراین برای ۵ نمره جایگاه میانه $= 3 = (5+1)/2$ ، یعنی میانه عدد سوم در یک مجموعه مرتب شده است. برای ۱۲ عدد جایگاه میانه $= 6.5 = (12+1)/2$ ، و میانه در اینجا میانگین عددهای ششم و هفتم است.

برای داده‌های مرکز دانشجویی (نگاه کنید به جدول ۱-۳) جایگاه میانه $= 100/5 = (200+1)/2$ است. با کمی دقت مشخص می‌شود که هر دو نمره‌ی ۱۰۰ ام و ۱۰۱ ام ۶ خواهند بود، که میانه است. برای داده‌های الکترونیکی ۱۰۰ نمره وجود دارد و جایگاه میانه $50/5$ است. از نمایش ساقه-برگ

در شکل ۶-۳ می توان گفت که نمره ۵۰ اُم یک ۴۴ و نمره ۵۱ اُم یک ۴۶ است. میانه ۴۵ خواهد بود، که میانگین این دو مقدار است.

۳-۵ میانگین

عمومی ترین شاخص گرایش مرکزی، و چیزی که در واقع به توضیح کمی نیاز دارد، میانگین ($\bar{X}$) است، یا چیزی که عموماً مردم آن را معدل می نامند. میانگین به عنوان کل نمرات تقسیم بر تعداد آنها تعریف می شود، و معمولاً به صورت $\bar{X}$ (بار X) نشان داده می شود. میانگین (با استفاده از نماد حاصل جمع که در فصل ۲ آمد) به این صورت تعریف می شود:

$$\bar{X} = \frac{\sum X}{N}$$

که $\sum X$ حاصل جمع تمام ارزش های X ، و N تعداد X ها است. بطور مثال، میانگین عددهای ۳، ۵، ۱۲ و ۵ می شود

$$\frac{3+5+12+5}{4} = \frac{25}{4} = 6.25$$

برای داده های بازی های الکترونیکی، میانگین را با گرفتن مقادیر خام داده ها از نمایش ساقه-برگ، جمع کردن آنها و تقسیم آنها بر ۱۰۰ حساب می کنیم. برای آن مثال میانگین می شود $52.04 / 100 = 52.04$. برنامه های کامپیوتری مثل SPSS و Minitab محاسبه میانگین مجموعه های بزرگ داده ها را آسان ساخته است.

۴-۵ کاستی ها و برتری های نسبی نما، میانه و میانگین

تنها زمانی که توزیع متقارن است میانگین و میانه برابر خواهند بود، و تنها زمانی که این توزیع متقارن و یک نمایی باشد همه این سه شاخص یکسان می باشند. در تمام موارد دیگر- و این شامل تقریباً تمام شرایطی است که ما با آن سر و کار خواهیم داشت - باید دست به انتخاب از میان شاخص های گرایش مرکزی زد. مجموعه ای از قواعد در تعیین زمان مناسب بودن استفاده از یک شاخص خاص گرایش مرکزی حکم می کنند. البته چنین قواعدی وجود ندارند. اگر می خواهید به گزینش خردمندانه ای در میان این سه شاخص دست بزنید، لازم است در باره قوت و ضعف هر یک اطلاعاتی داشته باشید.

نما

چنانکه دیدیم نما پربسامدترین نمره است. به لحاظ تعریف نمره‌ای است که در عمل رخ داده، در حالی که میانگین و گاهی میانه ممکن است مقادیری باشند که هرگز در داده‌ها ظاهر نشوند. نما همچنین دارای این برتری آشکار است که بیانگر بیشترین تعداد افراد است. کسی که یک فروشگاه کوچک را اداره می‌کند با توجه کردن به نما بهتر عمل می‌کند. اگر ۸۰٪ از مشتریان سبزه‌خانی بزرگ را بخواهند و ۲۰٪ سبزه‌خانی ریزه و کوچک را بخوانند، چنانچه شما حد میانگین را در نظر بگیرید و موجودی کالای خود را در حد عادی قرار دهید احتمالا تمام مشتریان خود را از دست خواهید داد.

با توجه به همین دو برتری واقعیت این است که از نظر تعریفی احتمال اینکه یک مشاهده تصادفی (X_i) با نما برابر باشد بیشتر از آن خواهد بود که با هر نمره معین دیگری برابر باشد در بیان جبری آن می‌توان گفت (هر نمره دیگر $P(X_i = \text{نما}) > P(X_i = \text{نما})$ و در نهایت، نما دارای این برتری است که برای داده‌های اسمی قابل کاربرد است، که اگر به آن توجه کنید آشکار خواهد شد که این موضوع در مورد میانه یا میانگین صدق نمی‌کند.

یک اشکال نما این است که ممکن است بازنمایی کننده ویژه‌ای از کل مجموعه نمرات نباشد. این اشکال در داده‌های بازی‌های الکترونیکی (جدول ۲-۳) نشان داده شده است، که در آن فاصله‌نمایی برابر ۹-۰ بوده و نشان دهنده این واقعیت است که احتمالا شماری از افراد بازی‌های کامپیوتری نمی‌کنند. به کارگیری این فاصله به عنوان نما باعث خواهد شد همه آن افرادی که بازی می‌کنند نادیده گرفته شوند. اشکال دیگر نما که زمان گروه‌بندی افراد بروز می‌کند این است که فاصله مدل به شدت وابسته به چگونگی گروه‌بندی داده‌ها است. با مجموعه‌ای از داده‌های ساده براحتی می‌توانید آن را برای خود روشن سازید.

میانه

بزرگترین مزیت میانه که نما هم دارد این است که تحت تأثیر نمرات انتهایی قرار نمی‌گیرد. بنابراین میانه‌های هر دو مجموعه نمرات (۵، ۸، ۹، ۱۵، ۱۶) و (۰، ۸، ۹، ۱۵، ۲۰۶) معادل ۹ است. آزمایشگران زیادی این ویژگی را در مطالعاتی که نمرات انتهایی گاهی رخ می‌دهد بسیار مفید یافته‌اند اما خیلی مهم نیست. برای مثال، میانگین مدت زمانی که موش آموزش دیده به سرعت یک مجرای کوتاه را طی می‌کند حدود ۱ تا ۲ ثانیه است. در هر بار، زمانی این موش به هر دلیل در نیمه راه می‌ایستد، خود را می‌خاراند، بینی‌اش را به سلول‌های نوری می‌مالد و دراز می‌کشد. در این مواقع عملاً مهم نیست که او طی ۳۰ ثانیه به آخر برسد یا ۱ دقیقه طول بکشد - حتی ممکن است آزمایشگر با سیخک زدن او را وادار به حرکت کند. اگر ما موشی را طی سه بار در یک روز بدونیم و زمان‌های آن (۱/۲، ۱/۳ و ۲۰ ثانیه باشد)، معنایش در باره دانش موش از این کار برای ما همچون موقعی است که زمان‌ها (۱/۳، ۱/۲ و ۱۳۶/۴)

باشد. در هر دو بار میانه $1/3$ خواهد بود. میانگین روزانه دو بار تمرین موش آشکارا متفاوت خواهد بود ($7/5$ در مقابل $46/3$ ثانیه). این مشکل، آزمایشگران را وادار می‌سازد در بسیاری مواقع میانه را بر میانگین زمان در روز ترجیح دهند.

میانه دارای نقطه قوت دیگری در برابر میانگین است که نویسندگان علاقه‌مند به مقیاس‌های سنجش آن را مورد اشاره قرار داده‌اند. محاسبه میانه نیازمند هیچگونه پیش‌فرضی در باره ویژگی‌های فاصله‌ای مقیاس نیست. با نمرات ۵، ۸، ۱۱ و ۱۱، موردی که در وسط معرفی می‌شود عدد ۸ است، بدون توجه به اینکه چقدر از مواردی که با اعداد ۵ و ۱۱ نشان داده شده‌اند دور یا نزدیک است. اما وقتی می‌گوییم میانگین روی ۸ است ما این فرض ضمنی را بنا گذاشته‌ایم که فاصله بین موارد ۵ و ۸ به اندازه فاصله بین موارد ۸ و ۱۱ است. اینکه این فرض منطقی است یا نه، به آزمایشگر مربوط است. ما ترجیح می‌دهیم بر این اصل تأکید کنیم که اگر این یک فرض غیرمنطقی است، آزمایشگر آن را تشخیص خواهد داد و گام‌های مناسب را بر می‌دارد. اگر واقعا غیرمنطقی نیست، آنگاه به احتمال زیاد، اثر اساسی آن بر نتایج نادیده انگاشته خواهد شد. این مسأله‌ی مقیاس‌های سنجش در فصل ۲ با جزئیات بیشتر مورد بحث قرار گرفته است.

میانگین

از سه شاخص اصلی گرایش مرکزی، میانگین فراگیرتر از همه است. چندان اغراق‌آمیز نخواهد بود اگر گفته شود که برای بسیار افراد (متأسفانه) آمار تقریباً معادل مطالعه میانگین است. چنان‌که پیش‌تر گفته شد، میانگین دارای کاستی‌های خاصی است. میانگین تحت تأثیر نمرات انتهایی قرار می‌گیرد، مقدار آن ممکن است عملاً در داده‌ها وجود نداشته باشد، و تفسیر آن از لحاظ متغیر زیربنایی اندازه‌گیری شده ممکن است قدری نیازمند اطمینان از ویژگی‌های فاصله‌ای داده‌ها باشد. با وجود این اشکالات، ممکن است بخواهیم این آماره را به فراموشی بسپاریم. اما میانگین دارای امتیازهای متعددی است که بر اشکال‌های آن فزونی دارد.

مهم‌ترین برتری میانگین از جنبه تاریخی احتمالاً قابلیت آن برای دستکاری جبری است. این باعث می‌شود میانگین کاربردهای گسترده‌ای داشته باشد. دومین امتیاز میانگین این است که دارای چندین ویژگی در رابطه با کاربرد آن به عنوان برآوردی از میانگین جمعیت است. به طور خاص اگر ما چندین نمونه از برخی جمعیت‌ها بگیریم، میانگین‌های نمونه‌ای باثبات‌ترین (کم تغییرترین) برآوردهای گرایش مرکزی آن جمعیت در مقایسه با میانه‌ها و نماها خواهند بود. این ویژگی دلیل کاربردهای زیاد میانگین است.

۵-۵ خلاصه

در این فصل ما به چند شاخص مورد استفاده برای توصیف مرکز یک توزیع توجه کردیم. هر یک از این شاخص‌ها نقاط قوت و ضعف خاص خود را دارند. یکی از آنها، یعنی میانگین، اساس بیشتر مواردی است که در باقی این کتاب بحث خواهد شد. برخی از مطالب مهم فصل عبارتند از:

- شاخص‌های گرایش مرکزی
- مکان میانه
- نما (Mo)
- میانگین ($\bar{X}$)
- میانه (Med)

۵-۶ تمرین‌ها

- ۵-۱ نما، میانه، و میانگین را برای داده‌های تمرین ۱-۳ حساب کنید.
 - ۵-۲ نما، میانه، و میانگین را برای داده‌های تمرین ۳-۴ حساب کنید.
 - ۵-۳ جواب‌های تمرین‌های ۱-۵ و ۲-۵ را مقایسه کنید. نظر شما در باره رفتار داستان‌گویی در بزرگسالان و کودکان چیست؟
 - ۵-۴ یک مجموعه داده‌ها طرح کنید که میانگین بزرگتر از میانه باشد.
 - ۵-۵ یک مجموعه از داده‌های دارای چولگی مثبت طرح کنید. میانگین بالا یا پایین میانه قرار می‌گیرد؟
 - ۵-۶ آیا می‌توانید یک مجموعه تک‌نمایی از داده‌ها بسازید که میانگین و میانه برابر اما متفاوت از نما باشند؟
 - ۵-۷ یک گروه ۱۵ تایی از موش‌ها در طول یک ماز مستقیم می‌دوند به میزان تمرین زیر برای اجرای یک ملاک تعیین شده نیاز دارند. فراوانی توزیع به این شرح است.
- | | | | | | | | |
|----|----|----|----|----|----|----|---|
| ۲۴ | ۲۳ | ۲۲ | ۲۱ | ۲۰ | ۱۹ | ۱۸ | ۱ |
| ۱ | ۳ | ۳ | ۳ | ۴ | ۰ | ۱ | |
- تعداد موش‌ها (فراوانی)
- میانگین و میانه تمرین‌ها را برای این گروه حساب کنید.
- ۵-۸ با مجموعه داده‌های زیر نشان دهید که تفریق یک مقدار معین (مثلاً ۵) از هر نمره موجب کاهش همان مقدار از تمام شاخص‌های گرایش مرکزی می‌شود. ۷ ۳ ۱۴ ۱۲ ۷ ۸
 - ۵-۹ با مجموعه داده‌های زیر نشان دهید که ضرب هر نمره در یک عدد معین موجب چند برابر شدن تمام شاخص‌های گرایش مرکزی به همان مقدار می‌شود. ۲ ۶ ۵ ۵ ۳ ۸
 - ۵-۱۰ نمونه‌ای از ده عدد بسازید که میانگین آن ۸/۶ باشد. دقت کنید که چگونه این مجموعه ساخته می‌شود- این کار بعداً به درک شما از مفهوم درجات آزادی کمک خواهد کرد.
 - ۵-۱۱ چرا نما یک شاخص قابل قبول برای داده‌ها اسمی است ولی میانگین و میانه نیستند.

فصل ۶

شاخص‌های تغییرپذیری

- ۱-۶- دامنه
- ۲-۶- دامنه بین‌چارکی و آماره‌های دیگر دامنه
- ۳-۶- انحراف متوسط
- ۴-۶- انحراف مطلق
- ۵-۶- انحراف معیار
- ۶-۶- میانگین و واریانس به عنوان برآوردکننده
- ۷-۶- نمودارهای جعبه‌ای - نمایش‌های گرافیکی پراکندگی و نمرات انتهایی
- ۸-۶- واریانس
- ۹-۶- مثال کار شده پایانی

در فصل ۵ به شماری از شاخص‌های مربوط به مرکز یک توزیع پرداختیم. اما یک مقدار متوسط برای توزیع (خواه نما باشد، میانه یا میانگین) نمی‌تواند کل داستان را نشان دهد. ما به شاخص (شاخص‌های) دیگری نیاز داریم تا میزان انباشتگی پیرامون آن شاخص متوسط یا انحراف نمرات از آن را نشان دهد. مقدار متوسط بیانگر موقعیت کلی بیشتر نمرات است. از سوی دیگر نمرات ممکن است در دامنه‌ی گسترده‌ای از مقادیر پخش شده باشند و "مقدار متوسط" ممکن است معرف دقیقی از تمام مجموعه مشاهدات نباشد. احتمالاً هر کسی تجربیاتی دارد از امتحاناتی که در آن همه دانشجویان نمره تقریباً

یکسانی گرفته‌اند و امتحاناتی که در آن دامنه نمرات از خیلی عالی تا خیلی بد پراکنده بوده‌اند. شاخص‌هایی که اشاره به تفاوت‌های بین این دو نوع موقعیت دارد همان چیزهایی هستند که وقتی ما از پراکندگی^۱ یا تغییرپذیری در اطراف میانه، نما یا هر نقطه دیگری صحبت می‌کنیم در ذهن داریم. بطور کلی ما از آن به عنوان پراکندگی در اطراف میانگین یاد می‌کنیم.

به عنوان یک نمونه از شرایطی که در آن ما ممکن است انتظار تغییرپذیری متفاوتی از یک گروه به گروهی دیگر داشته باشیم، به موردی توجه کنید که در آن دو گروه درس کامپیوتر امتحان مشابهی داده‌اند. به یک گروه گفته شده که نمره امتحان ۲۰٪ کل نمره نهایی را دارد، به گروه دیگر گفته شده که ۹۰٪ نمره را دارد. منطقی است که انتظار داشته باشیم دو گروه میانگین تقریباً یکسانی داشته باشند. اما چون "فشار" بر عملکرد برخی افراد اثر بازدارنده و بر برخی دیگر اثر پیش‌برنده دارد، انتظار داریم در مورد گروهی که این امتحان ۹۰٪ نمره نهایی را دارد تغییرپذیری بیشتری را شاهد باشیم.

به عنوان مثالی دیگر، به مجموعه‌های داده‌ها در زیر توجه نمایید:

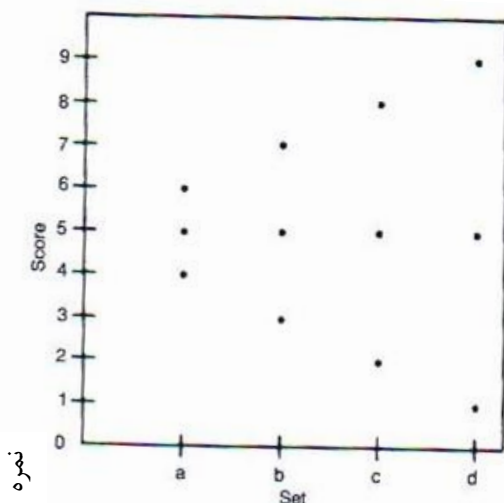
(الف) ۶ ۵ ۴

(ب) ۷ ۵ ۳

(ج) ۷ ۵ ۲

(د) ۹ ۵ ۱

یک محاسبه کوچک نشان خواهد داد که هر چهار مجموعه، میانگین یکسانی دارند (۵)، اما هر چه از مورد (الف) به سمت مجموعه (د) پیش روید تغییرپذیری بیشتر و بیشتر می‌شود. این موضوع را به صورت دیداری با نمایش مقادیر روی یک نمودار می‌توان مشاهده کرد (شکل ۱-۶).



شکل ۶-۱ نمودار چهار مجموعه از داده‌ها با میزان تغییرپذیری مسوت

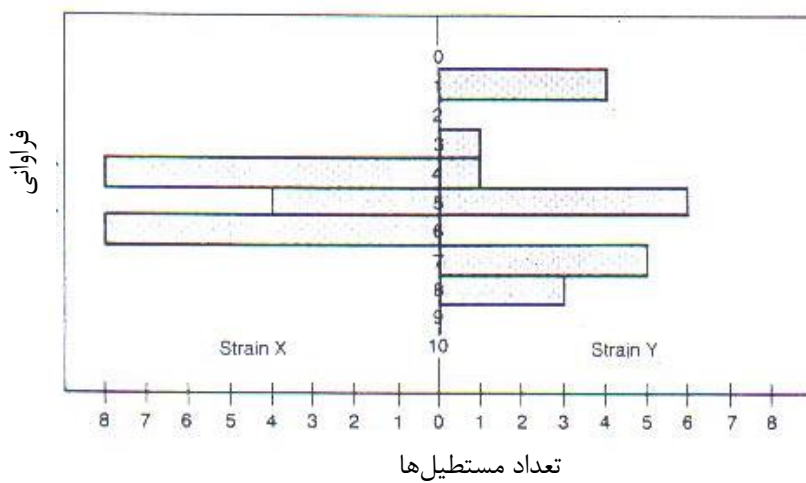
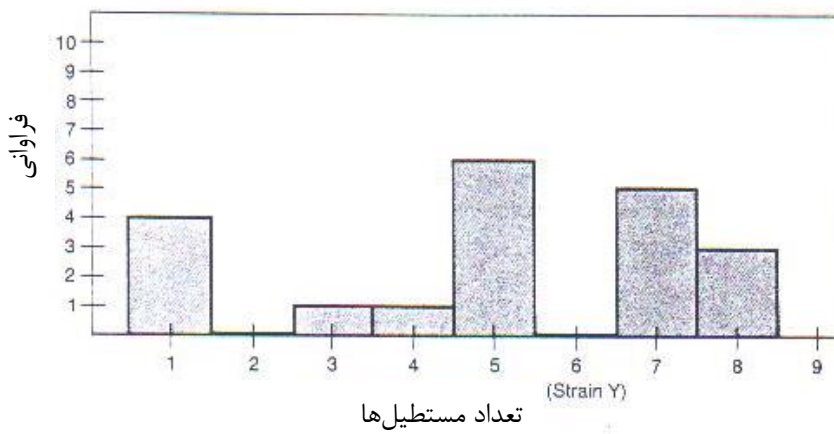
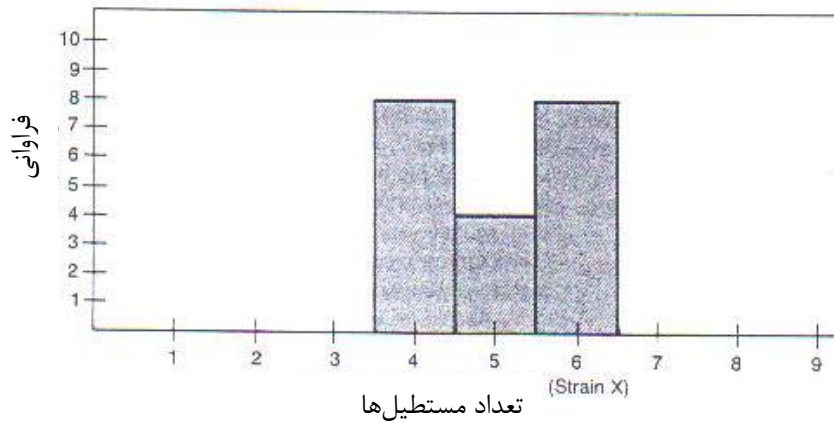
برای تصور بهتر موضوع، توجه شما را به دو مجموعه داده‌های زیر جلب می‌نماییم. ما ۲۰ موش را از دو نژاد گرفته و آنها را به صورت انفرادی روی کف یک جعبه دایره‌ای شکل قرار دادیم. کف جعبه شبکه‌ای بود و ما تعداد خانه‌هایی که موش در یک ف مجموعه طی می‌کرد را شمرده‌ایم. این یک شاخص معمول برای رفتار جستجوگری (آزمون محیط بر) است، و ما می‌خواستیم بدانیم آیا حیوانات نژاد X در رفتار جستجوگری بیش از نژاد Y با یکدیگر متفاوتند (تغییرپذیری بیشتری دارند). این مطالعه چندان ساده‌انگارانه نیست که در ابتدا به نظر می‌رسد. نژادهایی که رفتار خیلی با ثباتی از یک فرد به فرد دیگر نشان می‌دهند (یعنی تغییرپذیری کمی دارند) به عنوان منابع حیوانات آزمایشی از اهمیت خاصی برخوردارند زیرا تغییرپذیری خودبه‌خودی تصادفی کمتری در داده‌های ما خواهند داشت. سطح میانگین رفتار جستجوگری برای ما مهم نیست، و در واقع دو نژاد دارای میانگین یکسانی هستند.

نژاد X ۴ ۶ ۴ ۶ ۶ ۵ ۵ ۶ ۴ ۴ ۶ ۵ ۴ ۶ ۴ ۶ ۶ ۴ ۴ ۵

نژاد Y ۳ ۵ ۷ ۵ ۱ ۷ ۷ ۷ ۵ ۵ ۱ ۵ ۱ ۸ ۸ ۴ ۷ ۸ ۱ ۵

ما می‌توانیم توزیع‌های این نمرات را رسم کنیم تا تفاوت بین نژادها معلوم شود. دو توزیع در شکل ۶-۲ به عنوان توزیع‌های استاندارد فراوانی و هیستوگرام‌های پشت به پشت نشان داده شده است.

در حالی که پیداست که موش‌های نژاد Y بیش از نژاد X با هم فرق دارند، نوعی شاخص برای نشان دادن این تفاوت لازم است. شماری از شاخص‌ها را می‌توان به کار گرفت که در جای خود به ساده‌ترین صورت مورد بحث قرار خواهند گرفت.



شکل ۲-۶ توزیع‌های نمرات برای نژاد X و Y به صورت تعداد خانه‌های گذرانده

۱-۶ دامنه

دامنه^۱ شاخصی از فاصله است، بنام فاصله‌ی پایین‌ترین تا بالاترین نمره. برای داده‌های ما دامنه نژاد X برابر $2 = (4-6)$ و برای نژاد Y برابر $7 = (1-8)$ است. دامنه یک شاخص خیلی معمول در زندگی روزمره است و با عباراتی مانند "قیمت دلار در هفته گذشته بیش از ۲۰۰ تومان نوسان داشت و از ۳۸۰۰ تومان تا ۴۱۰۰ تومان متغیر بود" دیده می‌شود. اِشکال دامنه در وابستگی کامل به مقادیر یا ارزش‌های انتهایی است، یا، چنانچه ارزش‌ها بطور غیرعادی پرت باشند که به آنها نمرات انتهایی یا **دورافتاده**^۲ گفته می‌شود. در نتیجه، دامنه ممکن است یک تصویر غیر واقعی از تغییرپذیری بدست دهد. به عنوان نمونه‌ای از بعضی مشکلات مربوط به دامنه، به سه مجموعه داده‌ها زیر توجه نمائید:

گروه I	گروه II	گروه III
۸	۸	۶
۹	۹	۶
۹	۹	۶
۹	۹	۹
۹	۹	۹
۹	۹	۹
۹	۹	۱۰
۹	۹	۱۰
۹	۹	۱۰
۹	۹	۱۰
۱۱	۱۱	۱۱
۱۱	۱۱	۱۱
۱۱	۱۱	۱۱
۱۱	۱۱	۱۴
۱۱	۱۱	۱۴
۲۴	۱۶	۱۴
میانگین ۱۰/۵	۱۰	۱۰
دامنه ۱۶	۸	۸

دو مجموعه نخست داده‌ها دقیقاً شبیه هم هستند به جز در مورد نمره آخر، و میانگین تقریباً یکسانی دارند. اما، دامنه برای گروه اول دو برابر گروه دوم است. گروه دوم و سوم دارای میانگین و دامنه دقیقاً یکسانی هستند، اما تقریباً همه نمرات برای گروه II اعداد ۹ و ۱۱ ولی برای گروه III از ۶ تا ۱۴ بطور متعادل پخش شده‌اند.

۲-۶ دامنه بین چارکی و آماره‌های دیگر دامنه

دامنه بین چارکی^۱ که رابطه خیلی نزدیکی دارد با آنچه که ما به اختصار گستره‌ی H-خواهیم خواند، بازتابی از یک تلاش برای رهایی از مشکل وابستگی زیاد دامنه به نمرات انتهایی است. یک دامنه بین چارکی از کنار گذاشتن ۲۵٪ بالا و پایین توزیع و در نظر گرفتن دامنه‌ای که باقی می‌ماند بدست می‌آید. به این ترتیب، شامل دامنه ۵۰٪ مشاهدات میانی یا اختلاف بین نقطه درصدی ۷۵ ام و ۲۵ ام است. مثلاً ما می‌توانیم دامنه بین چارکی را برای رفتار جستجوگری با حذف پنج نمره پایین و پنج نمره بالا و تعیین دامنه باقی‌مانده حساب کنیم. در این مورد دامنه برای نژاد X بازهم $4 - 2 = 6$ خواهد بود. اما برای نژاد Y دامنه اکنون برابر $3 - 4 = 7$ خواهد بود.

دامنه چارکی نقش مهمی در یک روش گرافیکی خیلی سودمند به نام نوار جعبه‌ای دارد. این روش بعداً در این فصل توضیح داده خواهد شد.

از بسیاری جهات دامنه چارکی دارای مشکلاتی است که دقیقاً بر خلاف دامنه هستند. بویژه اینکه داده‌های بسیاری را نادیده می‌گیرد. اگر بخواهیم بدانیم آیا یک نژاد متنوع‌تر از دیگری است، به نظر نمی‌رسد درست باشد نمراتی که خیلی در انتها هستند را دور بریزیم و پراکندگی را در اطراف میانگین بجوئیم.

مزیتی در حذف ۲۵٪ درصد بالا و پایین توزیع قبل از محاسبه دامنه وجود ندارد. در واقع ما هر درصدی را که بخواهیم می‌توانیم حذف کنیم، تا اینکه خود و دیگران را در مورد کاری که می‌کنیم توجیه کنیم. آنچه که واقعاً ما می‌خواهیم این است که نمراتی را حذف کنیم که احتمال می‌دهیم شانسی باشند، بدون اینکه تغییرپذیری مورد مطالعه را از بین ببریم. نمونه‌هایی که درصد معینی (مثلاً ۱۰٪) از ارزش‌ها را در هر دنباله دارند **نمونه‌های آراسته^۲** نامیده می‌شوند و آماره‌هایی که روی چنین نمونه‌هایی اجرا می‌شود آماره‌های آراسته خوانده می‌شوند (مثل میانگین‌های آراسته یا دامنه‌های آراسته).

1. interquartile range
2. trimmed samples

۳-۶ انحراف متوسط

در نگاه اول اگر بخواهیم اندازه بگیریم که نمرات چگونه در اطراف میانگین پراکنده شده‌اند (یعنی از میانگین چقدر انحراف دارند)، به نظر می‌رسد منطقی‌ترین کار این خواهد بود که تمام انحرافات (یعنی $X_i - \bar{X}$) را بدست آوریم و میانگین آنها را حساب کنیم. هرچه نمرات با هم فاصله بیشتری داشته باشند، انحرافات بیشتر و بنابراین میانگین آنها بزرگتر است. داده‌های مثال قبلی مربوط به نژاد X و نژاد Y را در نظر بگیرید. این داده‌ها در ادامه همراه با انحراف هر نمره از میانگین آمده‌اند. شما مشاهده خواهید کرد که اگر میانگین این انحرافات را حساب کنید

$$\frac{\sum(X - \bar{X})}{N} \text{ یا } \frac{\sum(Y - \bar{Y})}{N}$$

حاصل جمع صفر خواهد شد. انحرافات زیر میانگین با انحرافات بالای میانگین هم وزن هستند. مجموع این انحرافات اطراف میانگین همیشه صفر خواهد بود، که معدل آنها را صفر می‌کند. در نتیجه این شاخص مسلماً نمی‌تواند یک شاخص رضایت‌بخشی برای پراکندگی باشد.

نژاد X		نژاد Y	
X	(X - $\bar{X}$)	Y	(Y - $\bar{Y}$)
۵	۰	۵	۰
۴	-۱	۱	-۴
۴	-۱	۸	۳
۶	۱	۷	۲
۶	۱	۴	-۱
۴	-۱	۸	۳
۶	۱	۸	۳
۴	-۱	۱	-۴
۵	۰	۵	۰
۶	۱	۱	-۴
۴	-۱	۵	۰
۶	۱	۷	-۴
۵	۰	۷	۲
۵	۰	۷	۲
۶	۱	۱	-۴
۶	۱	۵	۰
۴	-۱	۷	۲
۶	۱	۵	۰
۴	-۱	۳	-۲
۵		۵	
میانگین نمره خام =	۵		
جمع انحرافات =	۰		۰
معدل انحرافات =	۰		۰

۶-۴ انحراف مطلق

اگر به دشواری پیش آمده هنگام تلاش برای برگرفتن چیز سودمندی از معدل انحرافات فکر می‌کنید، ممکن است بخوبی به این پیشنهاد برسید که می‌توانیم با در نظر گرفتن مقدار مطلق انحرافات مشکل را به کلی حل کنیم. (مقدار مطلق یک عدد برابر است با مقدار آن عدد با حذف هر علامت منفی آن. مقدار مطلق با دو خط عمودی دو طرف عدد نشان داده می‌شود. مثلاً $3 = |+3| = |-3|$). این پیشنهاد استفاده از مقدار مطلق منطقی است زیرا ما می‌خواهیم بدانیم نمرات چقدر از میانگین انحراف دارند،

بدون توجه به اینکه آیا در بالای میانگین یا زیر آن هستند. این شاخص به نام میانگین انحراف مطلق^۱ (m. a. d) خوانده می شود.

$$m.a.d = \frac{\sum |X - \bar{X}|}{N}$$

حاصل جمع انحرافات مطلق تقسیم بر N (تعداد نمره ها) می شود تا یک انحراف متوسط (میانگین) بدست آید. برای داده های مربوط به دو نژاد موش ها میانگین انحرافات مطلق در جدول زیر آمده است:

نژاد Y		نژاد X	
$(Y - \bar{Y})$	Y	$(X - \bar{X})$	X
۰	۵	۰	۵
۴	۱	۱	۴
۳	۸	۱	۴
۲	۷	۱	۶
۱	۴	۱	۶
۳	۸	۱	۴
۳	۸	۱	۶
۴	۱	۱	۴
۰	۵	۰	۵
۴	۱	۱	۶
۰	۵	۱	۴
۰	۵	۱	۴
۴	۷	۱	۶
۲	۷	۰	۵
۲	۷	۰	۵
۴	۱	۱	۶
۰	۵	۱	۶
۲	۷	۱	۴
۰	۵	۱	۶
۲	۳	۱	۴
	۵		۵
۳۸		۱۶	
۱/۹		۰/۸۰	

میانگین نمره خام =

جمع مطلق انحراف =

میانگین انحراف مطلق

=

برای نژاد X مقدار $m.a.d = 0/80$ و برای نژاد Y برابر $1/9$ است. بنابراین $m.a.d$ بازگو کننده اختلافات پراکندگی است، که با افزایش پراکندگی بزرگتر می شود.

1. mean absolute deviation [در فارسی به اختصار انحراف مطلق خوانده می شود]

با همه سادگی و قابل فهم بودن، میانگین انحراف مطلق نقش مهمی در آمار ایفا نمی‌کند. به جای آن معمولاً شاخص‌های سودمندتری مانند واریانس (پراش) و انحراف معیار (انحراف استاندارد) استفاده می‌شود.

۵-۶ واریانس

شاخصی که در این بخش مد نظر قرار می‌دهیم، **واریانس**^۱ نمونه (S^2)، رویکرد متفاوتی را برای مشکل صفر شدن معدل انحرافات فراهم می‌نماید. (وقتی به واریانس جمعیت رجوع می‌کنیم از نماد σ^2 (حرف کوچک سیگما) استفاده می‌کنیم). در این مورد ما از امتیاز مثبت شدن یک عدد منفی وقتی به توان دو می‌رسد استفاده می‌کنیم. بنابراین ما به جای جمع کردن انحرافات، مجذور آنها را جمع می‌کنیم. چون ما یک مقدار متوسط می‌خواهیم، بعد باید این حاصل جمع را بر N ، تعداد نمره‌ها تقسیم کنیم. در واقع ما آن را بر $(N-1)$ ، نه خود N تقسیم می‌کنیم. ما از $N-1$ به عنوان یک مقسوم‌علیه برای واریانس نمونه استفاده می‌کنیم زیرا چنانکه خواهیم دید این واریانسی از نمونه بدست می‌دهد که برآورد بهتری از واریانس جمعیت مربوط به آن است. (واریانس جمعیت با تقسیم مجموع مجذورات انحراف نمرات از میانگین، بر N نه $N-1$ بدست می‌آید. اما بندرت ما واریانس جمعیت را محاسبه می‌کنیم؛ تقریباً همیشه آن را از یک واریانس نمونه برآورد می‌کنیم.) این شاخص که توضیح دادیم واریانس نمونه نامیده می‌شود و با S^2 نشان داده می‌شود، یا برای مشخص‌تر بودن، S_X^2 ، که اندیس X نشان دهنده متغیری است که واریانس آن را اندازه می‌گیریم.

برای مثال ما می‌توانیم واریانس‌های نمونه را به صورت زیر حساب کنیم.

نژاد X :

$$S_X^2 = \frac{\sum (X - \bar{X})^2}{N - 1} = \frac{(\delta - \delta)^2 + (4 - \delta)^2 + \dots + (4 - \delta)^2}{20 - 1} = \frac{16}{20 - 1} = 0.84$$

نژاد Y :

$$S_Y^2 = \frac{\sum (Y - \bar{Y})^2}{N - 1} = \frac{(\delta - \delta)^2 + (1 - \delta)^2 + \dots + (3 - \delta)^2}{20 - 1} = \frac{116}{20 - 1} = 6.11$$

از این مثال درمی‌یابیم که واریانس‌ها تفاوت‌های پراکندگی را منعکس می‌نمایند. نژاد Y که بیشتر از نژاد X پراکندگی دارد، دارای واریانس 6.11 است در حالی که نژاد X فقط دارای واریانس 0.84 است. هر چند واریانس یک مفهوم خیلی مهم است و یکی از پر کاربردترین آماره‌ها است، اما آن گونه که ما می‌خواهیم

مستقیماً قابل فهم نیست. زیرا بر اساس مجذور انحرافات بنا شده و حاصل در واحدهای مجذور شده بیان می‌شود. بنابراین Y دارای یک میانگین ۵ واحدی و یک واریانس ۶/۱۱ واحدی مجذور شده است. اما واحدهای مجذور شده چیزهای خام و مبهمی برای بیان کردن هستند و در رابطه با داده‌ها معنای کمی دارند. خوشبختانه راه حل این مشکل ساده است - جذر واریانس را بدست می‌آوریم.

۶-۶ انحراف معیار (انحراف استاندارد)

انحراف معیار^۱ (S یا σ) به عنوان ریشه دوم مثبت واریانس تعریف شده است، برای یک نمونه، با نماد S (اگر لازم باشد با یک اندیس مشخص کننده متغیر) یا، گاهی با $S.D$. نشان داده می‌شود. (نماد σ زمانی به کار می‌رود که به یک انحراف معیار جمعیت اشاره می‌شود).

$$S_X = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}}$$

برای مثال ما

$$S_X = \sqrt{S_X^2} = \sqrt{0.84} = 0.92$$

$$S_Y = \sqrt{S_Y^2} = \sqrt{6.11} = 2.47$$

اگر به فرمول انحراف معیار نگاه کنید مشاهده خواهید کرد که انحراف معیار مانند انحراف مطلق، در اصل یک اندازه متوسط از انحرافات هر نمره از میانگین است. به فرض اینکه این انحرافات مجذور شده باشند، جمع شده باشند و غیره، در اصل باز هم انحراف هستند. و چنانچه ما به جای تقسیم کردن بر N آنها را بر $N-1$ تقسیم کنیم، باز هم چیزی خیلی شبیه به میانگین یا "معدل" این انحرافات را بدست می‌آوریم. بنابراین بدون ایجاد خدشه زیادی می‌توان گفت که آزمودنی‌های نژاد X به طور متوسط ۰/۹۲ واحد از میانگین خود انحراف دارند و آزمودنی‌های نژاد Y به طور متوسط ۲/۴۷ واحد از میانگین خود انحراف دارند. نگاهی چنین ساده‌انگارانه به انحراف معیار به عنوان یک نوع انحراف متوسط قدری ما را از مفهوم واقعی دور می‌سازد. در حقیقت انحراف معیار حدود ۲۰٪ بزرگتر از انحراف متوسط برای نمونه‌های با اندازه‌ی قابل قبول است.

یک روش دیگر معنا بخشیدن به انحراف معیار این است که چه تعداد نمره فراتر از یک انحراف معیار بالا یا پایین میانگین قرار نمی‌گیرند. برای بسیاری از توزیع‌های متقارن و پشته‌ای شکل می‌توان گفت که تقریباً دو سوم همه مشاهدات در فاصله یک انحراف معیار از میانگین واقع شده‌اند (برای یک توزیع بهنجار، که در فصل ۷ توضیح داده شده، دقیقاً دو سوم است). هرچند استثنائیهایی برای این قاعده وجود دارد (بویژه برای توزیع‌های خیلی کج)، با این وجود این قاعده پرکاربردی است. اگر به شما بگویم

1. standard deviation

که برای مشاغل سنتی انتظار می‌رود میانگین شروع درآمد فارغ‌التحصیلان کارشناسی در سال جاری ۱۵۰۰۰ دلار با یک انحراف معیار ۴۰۰۰ دلاری باشد، شما نتیجه خواهید گرفت که حدود دوسوم فارغ‌التحصیلانی که این مشاغل را بدست می‌آورند بین ۱۱۰۰۰ تا ۱۹۰۰۰ دلار دریافت خواهند کرد.

سومین ویژگی در خور توجه انحراف معیار این است که اندازه‌ی آن معمولاً در حدود یک پنجم تا یک ششم دامنه است. درحالی‌که این روش خیلی مفیدی برای تفسیر انحراف معیار نیست، اما یک روش ساده برای بازبینی کارتان است. اگر یک نگاه به داده‌ها نشان دهد که دامنه حدود ۲۰ و $S_x = 15$ است، توصیه می‌کنیم حتماً کارتان را بازبینی کنید. ممکن است درست عمل کرده باشید اما احتمالاً چنین نیست. از سوی دیگر اگر دیدید که $S_x = 3/8$ است، شما دلیلی برای درستی جواب خود دارید.

فرمول محاسباتی واریانس و انحراف معیار

توضیحات پیش در باره واریانس و انحراف معیار در حالی که کاملاً درست بودند اما برای هر میزان قابل قبول از داده‌ها خیلی راحت و قابل فهم نیستند. آنها همچنین در معرض خطای گرد کردن هستند، زیرا معمولاً شامل مجذور انحرافات کسری هستند. اینها از لحاظ تعریفی فرمول‌های عالی هستند، اما اکنون ما به یک مجموعه از فرمول‌های بیشتر محاسباتی می‌پردازیم. برای این کار فرمول‌های تعریفی را از نو مرور می‌کنیم. فرمول تعریفی برای واریانس نمونه چنین است:

$$S_x^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

یک فرمول عملیاتی‌تر، و از لحاظ جبری یکسان این است^۱

$$S_x^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N - 1}$$

[۱. توضیحات مترجم]: چنانکه می‌دانیم در آمار برای نشان دادن نمره انحراف متغیر X از میانگین $(X - \bar{X})$ از حرف

یونانی x استفاده می‌شود (و برای نمره انحراف متغیر Y از میانگین $(Y - \bar{Y})$ از حرف یونانی y).

بنابراین مجموع مجذورات انحراف نمرات متغیر X از میانگین این گونه محاسبه می‌شود:

$$\sum x^2 = \sum (X - \bar{X})^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2$$

در نتیجه برای محاسبه مجموع مجذورات انحراف نمرات از روی نمرات اصلی از فرمول زیر استفاده می‌کنیم:

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

به همین ترتیب مجموع مضروب‌ات نیز که برای محاسبه کوواریانس و ضریب همبستگی به کار می‌رود و معادل مجموع مجذورات برای دو متغیر است از فرمول زیر بدست می‌آید:

$$\sum xy = \sum XY - \frac{\sum X \sum Y}{N}$$

به همین ترتیب، برای انحراف معیار نمونه

$$S_X^2 = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}}$$

$$= \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}}$$

بسیار مفید است که شما این فرمول‌ها را به خاطر بسپارید زیرا این معادلات یا بخشی از آنها در طول این کتاب تکرار خواهد شد و هر جایی دیده می‌شود.

با به کار بردن این فرمول محاسباتی برای واریانس نمونه در مثال ما، حاصل چنین خواهد بود

$$S_X^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}$$

$$= \frac{5^2 + 4^2 + \dots + 6^2 + 4^2 - \frac{100^2}{20}}{19}$$

$$= \frac{516 - \frac{100^2}{20}}{19} = \frac{516 - 500}{19} = \frac{16}{19} = 0.84$$

$$S_Y^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{N}}{N-1} = \frac{5^2 + 1^2 + \dots + 5^2 + 3^2 - \frac{100^2}{20}}{19}$$

$$= \frac{616 - \frac{100^2}{20}}{19} = \frac{616 - 500}{19} = \frac{116}{19} = 6.11$$

توجه دارید که این جواب‌ها دقیقاً با آنچه قبلاً بدست آمد یکسان هستند. همچنین باید توجه داشته

باشید که چنانکه در فصل ۲ گفته شد، $\sum X^2 = 516$ کاملاً با $\sum (X)^2 = 100^2 = 10000$ فرق دارد.

۷-۶ میانگین و واریانس به عنوان برآورد کننده

قبلاً در فصل ۱ این واقعیت را خاطر نشان کردیم که ما معمولاً شاخص‌هایی مانند میانگین و واریانس را حساب می‌کنیم تا به عنوان برآوردهایی از مقادیر نظیر آنها در جمعیت مورد استفاده قرار دهیم. ویژگی‌های نمونه‌ها آماره خوانده می‌شوند و با حروف رومی (مثل $\bar{X}$) نشان داده می‌شوند. در مقابل،

ویژگی‌های جمعیت‌ها پارامتر نامیده می‌شود و با حروف یونانی نشان داده می‌شوند. بنابر این میانگین جمعیت با μ (میو) نمادگذاری می‌شود. پس بطور کلی ما آماره‌ها را به عنوان برآوردکننده‌های جمعیت‌ها به کار می‌بریم.

چنانچه هدف از بدست آوردن یک آماره به کارگیری آن به عنوان برآورد کننده یک پارامتر باشد، آنگاه تعجبی نخواهد داشت که انتخاب یک آماره (و حتی چگونگی تعریف آن) تا حدی وابسته به این است که آن آماره با چه کیفیتی به عنوان برآورد کننده آن پارامتر در نظر گرفته می‌شود. در واقع معمولاً میانگین به سایر شاخص‌های گرایش مرکزی ترجیح داده می‌شود دقیقاً به خاطر عملکرد آن به عنوان برآورد کننده μ . واریانس (S^2) به عنوان حالتی ویژه تعریف شده است چون دارای مزایایی می‌باشد که هنگام استفاده به عنوان برآورد کننده واریانس جمعیت که با σ^2 نشان داده می‌شود بروز می‌کنند.

سه ویژگی برآوردکننده‌ها برای آماردانان از اهمیت خاصی برخوردار است و در انتخاب آماره‌هایی که ما حساب می‌کنیم خیلی تأثیر دارد. این ویژگی‌ها عبارتند از فراگیری^۱، غیر سوگیرانه بودن^۲ و کارآمدی^۳. این ویژگی‌ها در اینجا به اختصار مورد بحث قرار می‌گیرند تا نشان دهیم چرا بعضی شاخص‌های گرایش مرکزی و پراکندگی بیشتر از بقیه با اهمیت تلقی می‌شوند. خیلی مهم نیست که درک کاملی از برآورد و مفاهیم مربوط به آن پیدا کنید، بلکه کافی است کلیاتی در باره مسائل آن حاصل کنید.

فراگیری

یک آماره در صورتی آماره‌ای فراگیر است که همه اطلاعات در یک نمونه را در بر بگیرد (به کار بگیرد). میانگین یک آماره فراگیر است، چون ما از همه مشاهدات معدل گیری می‌کنیم. نما اما تنها از عمومی‌ترین مشاهدات استفاده می‌کند، بقیه را نادیده می‌گیرد، و میانه فقط از موارد میانی استفاده می‌کند، بازهم ارزش‌های سایر مشاهدات را نادیده می‌گیرد. این یکی از دلایلی است که روی میانگین به عنوان شاخص گرایش مرکزی مان تأکید می‌کنیم. به همین ترتیب دامنه یک آماره فراگیر نیست تنها براساس دو مقدار انتهایی است.

-
1. sufficiency
 2. unbiasedness
 3. efficiency

غیرسوگیرانه

فرض کنید مثلاً ما جمعیتی داریم که به گونه‌ای میانگین آن (μ) را می‌دانیم، مثل قد تیم بازیکنان بسکتبال یک ایالت. اگر بخواهیم یک نمونه از این جمعیت بیرون بکشیم و میانگین نمونه ($\bar{X}_1$) را حساب کنیم، انتظار خواهیم داشت بطور منطقی $\bar{X}_1$ نزدیک به μ باشد، بخصوص اگر N بزرگ باشد، زیرا برآوردی از μ است. به این ترتیب اگر میانگین قد در این جمعیت ۱۹۰ سانتیمتر باشد ($\mu = 190$)، ما انتظار خواهیم داشت یک نمونه مثلاً ۱۰ نفری از بازیکنان نیز دارای میانگین قد تقریباً ۱۸۵ سانتیمتر باشد، هر چند احتمالاً دقیقاً برابر با ۱۸۵ نخواهد بود. (می‌توانیم بنویسیم $\bar{X}_1 \approx 190$ که $\approx$ نشانه "تقریباً برابر" است). اکنون فرض کنید نمونه دیگری بیرون کشیده و میانگین آن $\bar{X}_2$ را بدست آورده‌ایم (در اینجا اندیس به کار رفته است تا تفاوت میانگین نمونه‌های متوالی را نشان دهیم). این میانگین نیز احتمالاً بطور منطقی نزدیک به μ خواهد بود اما دقیقاً برابر با μ یا $\bar{X}_1$ نخواهد بود. اگر ما این روند را ادامه دهیم و میانگین‌های بی‌شماری را بدست آوریم، در خواهیم یافت که معدل میانگین‌های این نمونه‌ها دقیقاً برابر μ است. بنابراین ما می‌گوئیم که **مقدار مورد انتظار** میانگین نمونه (یعنی معدل تعداد خیلی خیلی زیادی نمونه) برابر با μ ، میانگین جمعیت است. به یک برآورد کننده که مقدار مورد انتظار آن با پارامتری که می‌خواهیم برآورد کنیم برابر است، **برآورد کننده بدون سوگیری** گویند و این خاصیت خیلی مهمی برای آماره است که از آن برخوردار باشد. هم میانگین نمونه و هم واریانس نمونه، برآوردکننده‌های غیرسوگیرانه‌ی پارامترهای مربوط به خود هستند. برآورد کننده‌های بدون سوگیری و نمونه‌های بزرگ متمایل به جمعیت‌های غیرسوگیرانه هستند که برای کار مناسب‌تر از موارد دارای سوگیری هستند.

کارآمدی

برآوردکننده‌ها همچنین به **کارآمدی** نیز توصیف می‌شوند. فرض کنید که جمعیت متقارن است، و بنابراین شاخص‌های میانگین و میانه برابر هستند. حال فرض کنید می‌خواهیم میانگین جمعیت (یا میانه) را برآورد کنیم. اگر نمونه‌های زیادی را بیرون کشیده و میانگین‌های آنها را حساب کنیم، خواهیم فهمید که میانگین‌ها ($\bar{X}$ ها) بطور نزدیکی در اطراف μ انباشته شده‌اند. اما میانه‌های این نمونه‌ها کمتر نزدیک به μ خواهند بود. این بدین گونه است گرچه میانه هم در این شرایط یک برآوردکننده بدون سوگیری است، زیرا مقدار مورد انتظار میانه در این حالت برابر μ است. این حقیقت که میانگین‌های نمونه‌ای بیش از میانه‌ها در نزدیکی μ انباشته هستند نشان می‌دهد که میانگین یک برآوردکننده کارآمدتر است (در واقع کارآمدترین برآورد کننده μ است). از آنجا که میانگین بیشتر متمایل به نزدیک بودن به μ می‌باشد تا میانه، آماره‌ی بهتری برای برآورد μ است.

در حالی که روشن است که کارآمدی یک لفظ نسبی است (یک آماره کمتر یا بیشتر از آماره دیگر کارآمد است)، مطالبی در باره اینکه چنین و چنین آماره‌ای "کارآمد" است، معمول است. در کاربرد این

عبارت منظور ما واقعا این است که این آماره از همه آماره‌های دیگر کارآمدتر است. هم میانگین نمونه به عنوان برآوردی از μ و هم واریانس نمونه به عنوان برآوردی از σ^2 ، از این لحاظ برآورد کننده‌های کارآمدتری هستند. این واقعیت که میانگین و واریانس هر دو غیرسوگیرانه و کارآمد هستند دلیل اصلی چنین نقش مهمی است که در آمار دارند. این دو آماره اساس اغلب شیوه‌های مورد بحث در باقی این کتاب را تشکیل می‌دهند.

واریانس نمونه به عنوان یک برآورد کننده واریانس جمعیت

واریانس نمونه یک مثال عالی از آنچه در مبحث عدم سوگیری گفتیم می‌باشد. به خاطر دارید در آنجا اشاره شد که در محاسبه واریانس و انحراف معیار به جای N از $N-1$ استفاده می‌شود. اکنون علت آن را توضیح می‌دهیم.

به روش‌های مختلفی می‌توان علت استفاده از $N-1$ به عنوان مخرج کسر در واریانس‌های نمونه را توضیح داد. شاید ساده‌ترین آن همان مطلب قبلی در باره واریانس نمونه (S^2) به عنوان برآوردی غیرسوگیرانه از واریانس جمعیت σ^2 است. در حال حاضر فرض کنید که ما شمار نامحدودی از نمونه‌های (هر یک شامل N مشاهده) یک جمعیت را داریم و واریانس جمعیت را می‌دانیم. همچنین فرض کنید ما ناآگاهانه واریانس را به صورت $\frac{\sum (X - \bar{X})^2}{N}$ حساب کنیم. (به مخرج کسر توجه کنید.) اگر معدل این واریانس‌های نمونه‌ای را بگیریم، درخواهیم یافت که

$$\text{معدل} \left(\frac{\sum (X - \bar{X})^2}{N} \right) = E \left(\frac{\sum (X - \bar{X})^2}{N} \right) = \frac{(N-1)\sigma^2}{N}$$

که $E()$ "مقدار مورد انتظار" آنچه در پرانتز است خوانده می‌شود.

نکته آخر را می‌توان با یک مثال نشان داد. فرض کنید نمونه‌ای داریم متشکل از تنها سه عدد، ۱، ۲ و ۳. چون این کل جمعیت است، ما می‌توانیم μ و σ^2 را دقیقاً حساب کنیم. اینها $\mu = 2$ و $\sigma^2 = 0.666$ هستند. (به خاطر داشته باشید که اگر، و تنها اگر، ما کل جمعیت را به جای یک نمونه داشته باشیم، $\sigma^2 = \sum (X - \mu)^2 / N$.) همچنین فرض کنید که ما می‌خواهیم واریانس این جمعیت را براساس یک نمونه شامل دو مشاهده برآورد کنیم. تنها نه نمونه مختلف از $N=2$ می‌توان از این جمعیت بیرون کشید، و یک روش ساده این است که آنها را فهرست نمائیم و میانگین و واریانس هر نمونه را حساب کنیم. برای این مثال، ما S^2 را هم با N و هم با $N-1$ به عنوان مخرج حساب می‌کنیم. داده‌ها و محاسبات در جدول ۱-۶ آمده است.

توجه داشته باشید که میانگین میانگین‌های نمونه‌ای دقیقاً برابر با μ است و میانگین واریانس جمعیت با استفاده از مخرج $N-1$ دقیقاً برابر با σ^2 است. از این گذشته، معدل $\frac{\sum (X - \bar{X})^2}{N}$ (آماره سوگیرانه)

هست

$$E \left(\frac{\sum (X - \bar{X})^2}{N} \right) = \frac{(N-1)\sigma^2}{N} = \frac{1}{2} (\sigma^2) = \frac{1}{2} (0.766) = 0.383$$

بنابراین مقدار مورد انتظار $\frac{\sum (X - \bar{X})^2}{N}$ همان σ^2 که ما امیدواریم برآورد کند نیست، بلکه $(N-1)/N \sigma^2$ است. با استفاده از N در مخرج، ما یک برآورد اشتباه صورت داده‌ایم و برآورد سوگیرانه‌ای از σ^2 داریم. به هر حال این نتیجه اساس برآورد σ^2 را برای ما فراهم می‌سازد. اگر

$$E \left(\frac{\sum (X - \bar{X})^2}{N} \right) = \frac{(N-1)\sigma^2}{N}$$

آنگاه معادله جبری ساده نشان خواهد داد که

$$E \left(\frac{\sum (X - \bar{X})^2}{N-1} \right) = \sigma^2 \quad \text{و بنابراین} \quad E \left(\frac{\sum (X - \bar{X})^2}{N} \right) \times \frac{N}{N-1} = \sigma^2$$

این فرمول آخر فرمول استاندارد تعریفی σ^2 برای واریانس است. این فرمول نه تنها به ما نشان می‌دهد که چگونه یک برآورد از σ^2 پیدا کنیم، بلکه همچنین نشان می‌دهد که آن برآورد غیرسوگیرانه است.

جدول ۶-۱ نتایج نمونه‌گیری از یک جمعیت بسیار کوچک

S^2	S^2	$\bar{X}$	نمونه	
(استفاده از N)	(استفاده از $N - 1$)			
۰/۰۰	۰/۰۰	۱/۰	۱	۱
۰/۲۵	۰/۵۰	۱/۵	۲	۱
۱/۰۰	۲/۰۰	۲/۰	۳	۱
۰/۲۵	۰/۵۰	۱/۵	۱	۲
۰/۰۰	۰/۰۰	۲/۰	۲	۲
۰/۲۵	۰/۵۰	۲/۵	۳	۲
۱/۰۰	۲/۰۰	۲/۰	۱	۳
۰/۲۵	۰/۵۰	۲/۵	۲	۳
۰/۰۰	۰/۰۰	۳/۰	۳	۳
<u>۰/۳۳۳</u>	<u>۰/۶۶۷</u>	<u>۲/۰</u>	معدل	
۰/۳۳۳	۰/۶۶۷			

۸-۶ نمودارهای جعبه‌ای-نمایش‌های گرافیکی پراکندگی و نمرات انتهایی

در فصل ۳ دیدیم که چگونه نمایش‌های ساقه-برگ را می‌توان برای بازنمایی داده‌ها همزمان به چند روش مهم به کار برد. چنین نمایش‌هایی داده‌ها را به چیزی خیلی شبیه به هیستوگرام در می‌آورند در

حالی که مقادیر منفرد مشاهدات را حفظ می‌کند. علاوه بر نمایش‌های ساقه - برگ، توکی شیوه‌های دیگری برای نگاه به داده‌ها طرح کرده است، یکی از آنها پراکندگی داده‌ها را برجسته‌تر می‌نماید. این روش به نام **نمودارهای جعبه‌ای^۱** یا گاهی، **طرح‌های جعبه و شارپ^۲** شناخته می‌شود.

داده‌های جدول ۲-۶ برگرفته از مطالعه نوزادان عادی و کم وزن است که در دانشگاه ورمونت انجام شده، و بازگو کننده داده‌های اولیه از مدت در بیمارستان ماندن ۳۸ نوزاد دارای وزن عادی است. داده‌های سه نوزاد برای این متغیر خاص حذف شده‌اند و با علامت ستاره (*) نشان داده شده‌اند. (افزودن آنها برای تأکید بر این موضوع است که نباید فقط حذف شوند). چون داده‌ها از ۱ تا ۱۰ تغییر می‌کنند، همه برگه‌ها به استثنای دو مورد، صفر هستند. صفرها در واقع فضا را پر کرده‌اند تا یک توزیع از نوع هستیوگرام را شکل دهند. بررسی داده‌ها به شکل نمودار ساقه- برگ، نشان می‌دهد که توزیع دارای چولگی مثبت با یک میانه سه روزه است. در پایین ساقه می‌توانید مورد HI و مقادیر ۲۰ و ۳۳ را مشاهده کنید. اینها مقادیر انتهایی (پرت) هستند و به این طریق تنظیم شده‌اند تا وجود آنها برجسته شود. اینکه آیا اینها به اندازه کافی بزرگ هستند که ما را متعجب کنند از سؤالاتی است که در طرح نمودار جعبه‌ای دنبال می‌شود. آخرین ردیف نمودار ساقه- برگ، تعداد مشاهدات حذف شده را نشان می‌دهد.

به منظور فهم چگونگی ساخت یک نمودار جعبه‌ای، لازم است بعضی مفاهیم قبلی را یادآوری و به آنها چیزی اضافه کنیم. در فصل ۵ میانه را به عنوان نقطه‌ای در مجموعه نمرات به صورت $(N-1)/2$ تعریف کردیم. وقتی نقطه میانه یک عدد کامل است، آنگاه میانه بسادگی مقداری است که آن نقطه را در شکل مرتب شده داده‌ها اشغال کرده است. وقتی میانه یک عدد اعشاری باشد، میانه معدل دو مقدار دو طرف نقطه میانه است. برای داده‌ها جدول ۲-۶ نقطه میانه $19/5 = (38+1)/2$ و مقدار میانه ۳ است. برای تشکیل یک نمودار جعبه‌ای، در این جهت پیش می‌رویم که چه مقدار از هر نیمه از نمودار را بگیریم و این نقاط را **مفصل^۳** می‌نامیم. مفصل‌ها ارتباط نزدیکی با **نقاط چارکی^۴** دارند (اغلب با Q_1 و Q_3 نشان داده می‌شوند)، مقادیری که ۲۵٪ بالا و پایین توزیع را جدا می‌کنند. (فرق مفصل‌ها و نقاط چارکی جزئی بوده و با افزایش حجم نمونه‌ها، از بین می‌رود). برای محاسبه مفصل‌ها ابتدا باید جایگاه مفصل را پیدا کنیم که به این صورت تعریف می‌شود

$$\text{جایگاه مفصل} = \frac{1 + \text{موقعیت میانه}}{2}$$

اگر جایگاه میانه یک مقدار کسری است، باید این کسر به محاسبه جایگاه مفصل بیانجامد. جایگاه مفصل نسبت به مفصل، همان جایگاه میانه نسبت به میانه است. به ما می‌گوید در یک مجموعه مرتب، مقادیر

-
1. boxplots
 2. box-ans-whisker plots
 3. hinges
 4. quartiles

مفصل را کجا می‌توان یافت. برای داده‌های بستری، جایگاه مفصل $10 = 2 / (1 + 19)$ است. بنابراین مفصل‌ها به فاصله ۱۰ نمره از پایین و ۱۰ نمره از بالا قرار دارند. این مقادیر به ترتیب ۲ و ۴ هستند. برای مجموعه داده‌های بدون نمرات پیوندی یا برای نمونه‌های بزرگ، مفصل‌ها ۵۰٪ میانی نمرات را جدا می‌کنند.

برای تکمیل این مفاهیم لازم درباره نمودارهای جعبه‌ای به سه مطلب توجه کنیم: گستره H^1 ، حد درونی و مقادیر مجاور^۲. **گستره H** به طور ساده دامنه بین دو مفصل است و به این ترتیب دامنه بین چارکی است. برای داده‌های ما گستره $H = 2 - 4$ است. یک **حد درونی**^۳ نقطه‌ای است که $1/5$ برابر بالاتر یا پایین‌تر از مفصل مورد نظر قرار می‌گیرد. چون گستره H برای داده‌های ما برابر ۲ است، یک حد درونی برابر $3 = 1/5 \times 2$ نقطه فراتر از آن مفصل‌ها است. از آنجا که مفصل‌های داده‌های ما ۲ و ۴ هستند پس حدود درونی در $-1 = 2 - 3$ و $7 = 4 + 3$ هستند. در نهایت ما **مقادیر مجاور** را به صورت این مقادیر در داده‌ها تعریف می‌کنیم و انتهایی‌تر (دورتر از میانه) نیستند. چون مقدار کوچک‌تری که داریم ۱ است، این نزدیک‌ترین مقدار به حد درونی و پایین‌ترین مقدار مجاور است. حد درونی بالایی ۷ است و چون ما یک ۷ در داده‌ها داریم، این مقدار مجاور بالایی خواهد بود. محاسبات همه این مطالبی که اکنون تعریف کردیم در بخش بالایی جدول ۳-۶ نشان داده شده است.

1. H spread
2. adjacent values
3. inner fence

جدول ۲- ۶ داده‌ها و نمودار ساقه- برگ‌گی مربوط به طول دوره بستری نوزادان (به روز)

ساقه - و - برگ			داده‌ها	
...	۱	۷	۱	۲
.....	۲	۲	۳۳	۱
.....	۳	۴	۳	۲
.....	۴	۴	*	۳
..	۵	۱۰	۳	۳
.	۶	۵	۲	۹
.	۷	۳	۳	۴
.	۸	۲	۶	۲۰
.	۹	۲	۵	۴
.	۱۰	*	*	۱
۲۰، ۳۳	HI	۴	۳	۳
		۴	۳	۲
		۳	۲	۳
			۴	۲

۳ = مفقود

حدود درونی و مقادیر مجاور برای برخی دانشجویان پیچیده است. گله‌ای گاو را که در اطراف یک مزرعه پخش هستند در نظر بگیرید (یا اگر مایلید گروهی از کودکان در یک زمین بازی). حصار اطراف مزرعه (زمین بازی) حد درونی نمودار جعبه‌ای را می‌رساند. گاوها (یا بچه‌ها) ی نزدیک به حصار اما درون آن، مقادیر مجاور هستند. نگران گاوها (یا بچه‌ها) ی خارج از حصار نباشید. آنها در محاسبات این نقطه دخالت ندارند.

اکنون آمادگی داریم که یک نمودار جعبه‌ای را رسم کنیم. ابتدا مقیاسی که کل دامنه مقادیر بدست آمده را پوشش دهد رسم نموده و نامگذاری می‌کنیم. در پایین جدول ۳-۶ این کار انجام شده است. سپس یک جعبه مستطیلی از یک سوی گستره به سمت دیگر آن رسم می‌کنیم همراه با یک خط عمودی که نشانگر میانه است. سپس خط‌چین‌ها (شارب‌ها) را از گستره به سمت مقادیر مجاور می‌کشیم. در آخر همه نقاط انتهایی‌تر از مقادیر مجاور را رسم می‌کنیم.

از جدول ۳-۶ می‌توان چند چیز مهم را مشاهده کرد. اول از همه بخش مرکزی توزیع بطور منطقی متقارن است. این موضوع با این واقعیت که میانه در مرکز جعبه قرار گرفته نشان داده شده و همچنین از نمایش ساقه- و-برگ معلوم است. همچنین می‌توانیم مشاهده کنیم که توزیع دارای چولگی مثبت است زیرا خط‌چین سمت راست بطور مشخصی بلندتر از سمت چپ است. این موضوع را از نمایش ساقه- برگ‌گی هم می‌توان دید گرچه خیلی مشخص نیست. در نهایت مشاهده می‌کنیم که چهار مورد پرت یا دور افتاده داریم که در اینجا به عنوان مقادیر انتهایی‌تر از خط‌چین تعریف می‌شوند (و بنابراین از

مقادیر مجاور انتهایی تر هستند). نمایش ساقه- برگی به خوبی نمودار جعبه‌ای موقعیت دور افتاده‌ها را نشان نمی‌دهد.

جدول ۳-۶ محاسبات و نمودار جعبه‌ای برای داده‌های جدول ۲-۶

موقعیت میانه	$(N+1)/2 = (38+1)/2 = 19.5$
میانه	۳
موقعیت گستره	$(19+1)/2 = 10$
گستره پایین = دهمین نمره پایین	۲
گستره بالا = دهمین نمره بالا	۴
گستره H	$H = 4 - 2 = 2$
گستره	$H \times (1/5) = 2 \times (1/5) = 0.4$
حد پایین = گستره پایین - (گستره H) $\times (1/5)$	$2 - 0.4 = 1.6$
حد بالا = گستره بالا + (گستره H) $\times (1/5)$	$4 + 0.4 = 4.4$
مقدار مجاور پایین = حد پایین $\geq$ کمترین مقدار	۱
مقدار مجاور بالا = حد بالا $\leq$ بزرگترین مقدار	۷
	0 5 10 15 20 25 30

* مقادیر کسری را کم کنید.

موارد دور افتاده نیازمند توجه خاصی هستند. یک مورد دور افتاده می‌تواند نشانگر یک خطای اندازه‌گیری در ثبت یا وارد کردن داده‌ها باشد، یا می‌تواند نشان دهنده یک مقدار درست باشد که در حدود انتهایی قرار می‌گیرد. برای مثال داده‌های ما طول مدت بستری را نشان می‌دهند و یک نوزاد ممکن است نارسایی خاصی داشته باشد که نیاز به بستری طولانی مدت باشد. چون اینها داده‌های واقعی هستند، امکان این وجود داشت که به سوابق بیمارستانی مراجعه و چهار مورد انتهایی دقیق‌تر بررسی شوند. در بررسی معلوم شد که دو نمره انتهایی‌تر در اثر خطا بوده و به سرعت تصحیح شدند. دو نمره انتهایی دیگر ناشی از مشکلات پزشکی نوزدان بود. در اینجا لازم بود که اعضای پروژه تصمیم بگیرند که مشکلات تا چه حد جدی است که باعث شوند نوزدان از مطالعه حذف شوند یا نه. (در هر دو مورد آنها نگه داشته شدند.) دو مقدار اصلاح شده ۳ و ۵ بودند (به جای ۳۳ و ۲۰)، و یک نمودار جعبه‌ای تازه به شکل زیر در آمد:



این شبیه به نمودار جعبه‌ای در جدول ۳-۶ است به استثنای فضا و دو مقدار بزرگ. (می‌توانید خودتان واریسی کنید که داده‌های اصلاح شده واقعا به این شکل در می‌آید.)

از آنچه گفته شد باید مشخص شده باشد که نمودارهای جعبه‌ای برای بررسی پراکندگی داده‌ها خیلی سودمند هستند. بخصوص برای نظارت بر خطاهای داده‌ها و برجسته سازی مشکلات احتمالی قبل از اینکه تحلیل‌های بعدی صورت گیرد. اینها در بخش‌های بعدی این کتاب به عنوان راهنمای عینی داده‌ها ارائه خواهند شد.

۹-۶ مثال کاربردی نهایی

برای این که بار دیگر محاسبات مربوط به آماره‌های اصلی فصل را نشان دهیم، به داده‌های زیر درباره اندازه استاندارد افسردگی در ده آزمودنی توجه کنید:

X	X ²
۵	۲۵
۸	۶۴
۹	۸۱
۴	۱۶
۶	۳۶
۳	۹
۷	۴۹
۸	۶۴
۷	۴۹
۳	۹
$\Sigma X = ۶۰$	$\Sigma X^2 = ۴۰۲$

میانگین:

$$\bar{X} = \frac{\sum(X)}{N} = \frac{60}{10} = 6/0$$

واریانس:

$$S_X^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1} = \frac{402 - \frac{60^2}{10}}{9} = \frac{402 - 360}{9} = 4/667$$

انحراف معیار:

$$S = \sqrt{S^2} = \sqrt{4/667} = 2/160$$

نمودارهای جعبه‌ای:

ابتدا مشاهدات را به ترتیب از کوچک به بزرگ مرتب می‌کنیم

۳ ۳ ۴ ۵ ۶ ۷ ۷ ۸ ۸ ۹

$$\text{موقعیت میانه} = \frac{N-1}{2} = \frac{11}{2} = 5/5$$

$$\text{میانه} = \frac{6+7}{2} = 2/160$$

$$\text{جایگاه میانه} = \frac{6}{5} = 6/5$$

(در صورت نیاز کسر را از موقعیت میانه کم کنید)

مفصل‌ها شامل سومین مشاهده یک مجموعه مرتب شده از بالا و پایین = ۴ و ۸ می‌باشند.

گستره H = فاصله بین مفصل بالا و مفصل پایین = ۴ - ۴ = ۸.

$$1/5 \text{ برابر گستره } H = 6 = 1/5(4)$$

حد درونی = مفصل‌ها $\pm$ (گستره H) $1/5$.

$$\text{مفصل بالا} = 14 = 8 + 6$$

$$\text{مفصل پایین} = -2 = 4 - 6$$

مقادیر مجاور = مقادیر نزدیک اما نه گذشته از حد درونی.

$$\text{مقدار مجاور پایین} = 3$$

$$\text{مقدار مجاور بالا} = 9$$

-----|-----

۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹ ۱۰

۱۰-۶ خلاصه

در این فصل چند شاخص را برای سنجش مقدار پراکندگی نمرات در اطراف میانگین توزیع مورد توجه قرار دادیم. مهم‌ترین این شاخص‌ها انحراف معیار و واریانس هستند که برای بخش‌های باقیمانده کتاب نقش مهمی دارند. همچنین دیدیم که چگونه از نمودارهای جعبه‌ای برای برجسته کردن شکل توزیع و کمک به شناسایی مقادیر دور افتاده استفاده کنیم. برخی از مطالب مهم فصل عبارتند از:

- پراکندگی
- دامنه
- دورافتاده‌ها
- نمونه‌های آراسته
- واریانس
- انحراف معیار
- مقادیر مورد انتظار
- برآوردکننده غیر سوگیرانه
- کارآمدی
- نمودار جعبه‌ای
- مفصل‌ها
- مقادیر مجاور
- خط چین‌ها (شارب)

۱۱-۶ تمرین‌ها

دامنه، واریانس، و انحراف معیار را برای داده‌های تمرین ۱-۳ حساب کنید.
 ۲-۶ دامنه، واریانس، و انحراف معیار را برای داده‌های تمرین ۴-۳ حساب کنید.
 ۳-۶ جواب‌های دو سؤال قبلی را با هم مقایسه کنید.
 ۴-۶ در تمرین ۱-۶ چند درصد از نمرات در فاصله دو انحراف معیار از میانگین قرار می‌گیرند؟
 ۵-۶ در تمرین ۲-۶ چند درصد از نمرات در فاصله دو انحراف معیار از میانگین قرار می‌گیرند؟
 ۶-۶ داده‌ها زیر را در نظر بگیرید، نشان دهید که افزودن یا کم کردن یک مقدار ثابت به هر نمره تغییری در انحراف معیار ایجاد نمی‌کند. اگر یک مقدار ثابت به هر نمره اضافه یا کم شود میانگین چه تغییری می‌کند؟

۵ ۴ ۲ ۳ ۴ ۹ ۵

۷-۶ با در نظر گرفتن داده‌های تمرین ۶-۶ نشان دهید که اگر نمرات در یک عدد ثابت ضرب یا تقسیم شوند مقدار انحراف معیار به همان نسبت تغییر می‌کند. در این شرایط میانگین چه تغییری می‌کند؟
 ۸-۶ با استفاده از داده‌های تمرین ۶-۶ و ۷-۶، مجموعه داده‌های زیر را تغییر دهید تا به انحراف معیار ۱/۰۰ برسد:

۵ ۸ ۳ ۸ ۶ ۹ ۹ ۷

- ۹-۶ از جواب های تمرین ۶-۶ و ۶-۷ استفاده کنید تا جواب تمرین ۸-۶ را برای داشتن میانگین ۰ و انحراف معیار ۱/۰۰ اصلاح کنید. (توجه: حل تمرین ۸-۶ و ۹-۶ برای فصل ۷ مهم است).
- ۱۰-۶ دو مجموعه داده بسازید که دارای دامنه یکسان اما واریانس متفاوت باشند.
- ۱۱-۶ برای تمرین ۱-۶ نمودار جعبه ای را رسم کنید.
- ۱۲-۶ برای تمرین ۲-۶ نمودار جعبه ای را رسم کنید.
- ۱۳-۶ میانگین داده های مورد استفاده در تمرین ۱-۳ برابر ۱۸/۹ است. فرض کنید کودک دیگری داشته باشیم که نمره ۱۸/۹ گرفته باشد. واریانس را برای این داده ها دوباره حساب کنید (می توانید از مراحل میانجی تمرین ۱-۶ استفاده کنید). این نمره بر جواب تمرین ۱-۶ چه اثری دارد؟
- ۱۴-۶ به جای افزودن نمره ای معادل میانگین (مانند تمرین ۱-۶) نمره ی ۴۰ را به داده های تمرین ۱-۶ اضافه کنید. این نمره بر جواب تمرین ۱-۶ چه تأثیری می گذارد؟
- ۱۵-۶ محاسبات جدول ۱-۶ را با در نظر گرفتن جمعیت متشکل از اعداد ۱، ۲، ۳، ۴، و ۵ تکرار کنید و همه نمونه های $N = 3$ را بیرون بکشید.
- ۱۶-۶ بحث یک آماره کارآمد از این زاویه بیان شد که اگر ما شماری زیادی از نمونه ها را بیرون بکشیم چه اتفاقی می افتد. اما ما معمولاً تنها یک نمونه از هر جمعیت استخراج می کنیم. پس چرا نسبت به نتایج حاصل از نمونه های زیاد احتیاط می کنیم؟
- ۱۷-۶ بار دیگر به سؤال قبل با توجه به عدم سوگیری به جای کارآمدی پاسخ دهید.
- ۱۸-۶ (الف) برای داده های زیر یک نمودار جعبه ای رسم کنید:
- ۱ ۳ ۳ ۵ ۸ ۸ ۹ ۱۲ ۱۳ ۱۶ ۱۷ ۱۷ ۱۸ ۲۰ ۲۱ ۳۰
- (ب) انحراف معیار این داده ها را حساب کنید و هر نمره را بر آن تقسیم کنید.
- (ج) یک نمودار جعبه ای برای تمرین (ب) رسم کنید.
- (د) دو نمودار را با هم مقایسه کنید.

فصل ۷

توزیع بهنجار

از آنچه در فصل‌های پیشین گفته شده مشخص می‌شود که ما توجه زیادی به توزیع‌ها داریم- توزیع‌های داده‌ها، توزیع‌های فرضی جمعیت‌ها، و توزیع‌های نمونه‌گیری (که در فصل ۱۱ در باره آن شرح مختصری داده خواهد شد). از تمام شکل‌های ممکن که توزیع‌ها می‌توانند به خود بگیرند، گروه معروف به **توزیع بهنجار**^۱ تاکنون برای اهداف ما از همه مهم‌تر بوده است.

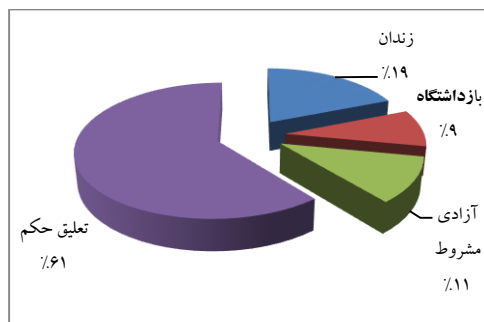
پیش از آنکه وارد بحث دقیق توزیع بهنجار (نرمال) شویم، خوب است گریزی بزنییم به اینکه بطور کلی چرا ما به توزیع‌ها این قدر توجه داریم. عامل اصلی این است که بین توزیع‌ها و احتمالات ارتباط مهمی وجود دارد. اگر در باره توزیع پدیده‌ها (یا توزیع نمونه‌ها) چیزی بدانیم، در باره احتمال رخداد یکی از این پدیده‌ها (یا آماره‌ها) چیزهایی می‌دانیم. برای اینکه این قضیه را به ساده‌ترین شکل بفهمیم به نمودار دایره‌ای توجه می‌کنیم.

نمودار دایره‌ای نشان داده شده در شکل ۱-۷ از گزارش اداره دادگستری آمریکا در باره جرائم تعلیقی و آزادی مشروط گرفته شده است. این نمودار وضعیت جاری همه کسانی که به یک جرم محکوم شده‌اند را نشان می‌دهد. در این شکل مشاهده می‌کنیم که ۹٪ در بازداشت، ۱۹٪ در زندان، ۶۱٪ در شرایط تعلیقی و ۱۱٪ دیگر بطور مشروط آزاد شده‌اند. همچنین مشاهده می‌کنید که درصد هر مقوله مستقیماً بازتاب درصد سطحی از دایره است که هر بخش به خود اختصاص داده است. سطح هر بخش نسبت مستقیمی از درصد افراد در آن قسمت است. علاوه بر این، اگر در نظر بگیریم که کل سطح برابر ۱/۰۰ واحد است، سطح هر قسمت برابر با بخشی از مشاهدات رویداده در آن بخش است.

1. normal distribution

از روی مطالب مربوط به سطوح بسادگی می‌توان در باره احتمالات صحبت کرد. مفهوم احتمال در فصل ۸ بیشتر مورد بحث قرار خواهد گرفت، اما بدون تعریف دقیق احتمالات هم می‌توان در باره سطوح یک نمودار دایره‌ای نکات مهمی را فهمید. در اینجا برای یک برداشت ساده به صورت کاربردهای روزمره احتمالات، به احتمال وقوع بعضی پدیده‌ها اشاره می‌کنیم. از این نقطه نظر منطقی است که نتیجه بگیریم که ۱۹٪ از افراد محکوم شده در زندان هستند، اگر بطور تصادفی نام یک شخص را از فهرست محکومان بیرون بیاوریم، احتمال ۰/۱۹ وجود دارد که این فرد در زندان باشد. به عبارت دیگر، اگر ۱۹٪ سطح دایره به زندان اختصاص دارد، پس احتمال اینکه یک شخص در این ناحیه قرار بگیرد ۰/۱۹ است.

این نمودار دایره‌ای همچنین امکان می‌دهد که نواحی دیگر را نیز معین کنیم. باید روشن باشد که اگر ۱۹٪ در زندان و ۹٪ در بازداشت هستند، $۱۹ + ۹ = ۲۸\%$ زندانی هستند. به عبارت دیگر، ما می‌توانیم درصد افراد در مقوله‌ها را با جمع کردن درصدهای هر مقوله پیدا کنیم. همین موضوع در باره سطوح هم صادق است، به این صورت که می‌توان درصد افراد زندانی را با جمع کردن سطوح متعلق به زندان و بازداشتگاه پیدا کرد. و در آخر، اگر ما بتوانیم درصدها را با جمع کردن سطوح پیدا کنیم، می‌توانیم احتمالات را هم با جمع کردن سطوح پیدا کنیم. بنابراین احتمال زندانی بودن برابر است با احتمال بودن در یکی از دو بخش مربوط به آنها، که می‌توان با جمع کردن این دو سطح (یا احتمالات مربوط به آنها) آن را بدست آورد.

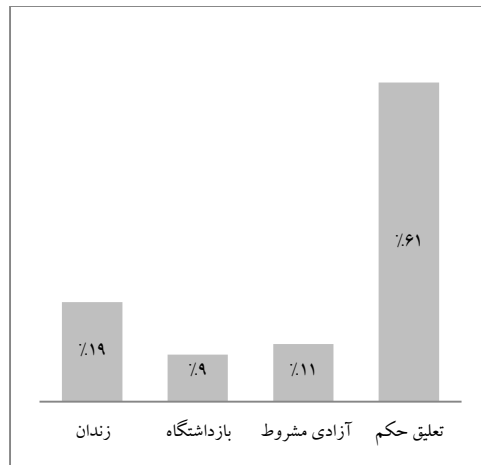


شکل ۷-۱ نمودار دایره‌ای نشان دهنده افراد تحت نظارت تأدیبی بر اساس نوع نظارت در ۳۱ دسامبر

۱۹۸۲

روش‌های دیگری هم برای نمایش داده‌ها غیر از نمودار دایره‌ای وجود دارد و یکی از ساده‌ترین آنها هیستوگرام (در فصل ۳ توضیح داده شد) یا خویشاوند نزدیک آن، نمودارستونی است. در شکل ۷-۲، بار دیگر اطلاعات شکل ۷-۱ به صورت نمودار ستونی ترسیم شده است. هر چند این نمودار اطلاعات تازه‌ای ندارد اما نسبت به نمودار دایره‌ای دو مزیت دارد. یکی اینکه مقایسه مقوله‌ها ساده‌تر است، چون تنها با

نگاه به ارتفاع ستون‌ها و نه محیط بخش‌های دایره، می‌توانیم بخش‌های مختلف را مقایسه کنیم. مزیت دوم این است که نمودار ستونی شباهت دیداری بیشتری با توزیع‌های معمول که می‌شناسیم دارد، که در آنها سطوح مختلف یا مقوله‌ها روی محور افقی توزیع و درصد هر مقوله روی محور عمودی نشان داده می‌شوند. علاوه بر این ما می‌توانیم مجموع سطوح را دقیقاً همانند نمودار دایره‌ای بدست آوریم. وقتی ما به سمت توزیع‌های معمول‌تر بخصوص توزیع نرمال پیش می‌رویم، اصول مربوط به سطوح، درصدها، احتمالات و جمع کردن سطوح یا احتمالات بدون تغییر انتقال می‌یابد.



نمودار ۲-۷ نمودار ستونی نشان دهنده افراد تحت نظارت بر اساس نوع پی‌گرد

قبلاً گفتیم که توزیع نرمال یکی از مهم‌ترین توزیع‌هایی است که با آن سروکار داریم. دلایلی برای آن وجود دارد:

۱- بسیاری از متغیرهای وابسته که معمولاً با آنها سروکار داریم فرض می‌شود که در جمعیت دارای توزیع نرمال باشند. یعنی مرتباً فرض می‌گیریم که اگر کل جمعیت را مورد مشاهده قرار می‌دادیم، توزیع حاصل به توزیع نرمال بسیار نزدیک بود.

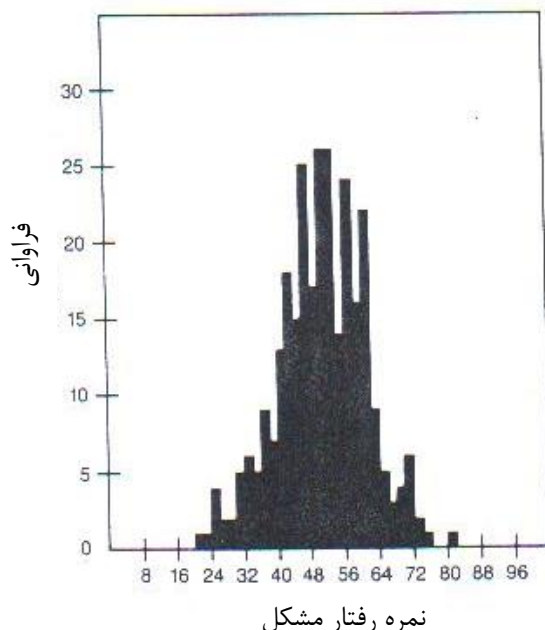
۲- اگر فرض کنیم که یک متغیر حداقل دارای توزیع نرمال است، آنگاه فنونی که در این فصل مورد بحث قرار گرفته‌اند به ما امکان نتیجه‌گیری‌های چندی را (خواه دقیق یا احتمالی) در باره اندازه آن متغیر می‌دهند.

۳- می‌توان نشان داد که توزیع نظری مجموعه فرضی میانگین‌های نمونه‌ای حاصل از بیرون کشیدن شمار نامحدودی از نمونه‌ها از یک جمعیت خاص، تحت شرایط متنوع تقریباً نرمال است. چنین توزیعی را توزیع نمونه‌گیری از میانگین گویند و در فصل‌های بعد بطور گسترده مورد استفاده قرار می‌گیرد.

۴- بسیاری از شیوه‌های آماری که ما به کار خواهیم برد، بر پیش‌فرض نرمال بودن توزیع‌های نمونه‌گیری استوار هستند.

برای ورود به بحث توزیع نرمال به یک مجموعه داده‌های دیگر نگاه می‌کنیم که تقریباً نرمال هستند (و اگر مشاهدات بیشتری داشته باشیم نرمال خواهد بود). این داده‌ها با استفاده از فرم گزارش شخصی جوانان آکنباخ^۱ (آکنباخ، ۱۹۸۶) جمع‌آوری شده است. این یک مقیاس پر کاربرد برای مشکلات رفتاری است که نمراتی را در ابعاد مختلف فراهم می‌نماید. یکی از ابعادی که ما به آن نگاه می‌کنیم بعد مشکلات کلی رفتار است که بازگو کننده تعداد کل مشکلات گزارش شده بوسیله کودک (وزن داده شده با شدت مشکل) است. (نمونه‌ای از مقوله‌های رفتاری شامل "بحث کردن"، "بی‌قراری"، "تهدید" و "دست انداختن" است.)

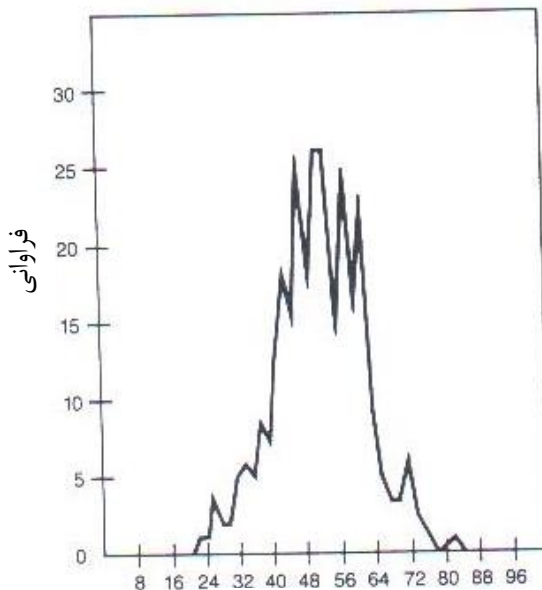
شکل ۳-۷ یک هیستوگرام از داده‌های ۳۰۹ دانش‌آموز مرحله اول دبیرستان است. یک نمره بالاتر نشان دهنده‌ی مشکلات بیشتر رفتاری است. مشاهده می‌کنید که این توزیع دارای یک مرکز ۵۰ است و خیلی به صورت متقارن در هر طرف این عدد توزیع شده است با نمراتی که دامنه آنها بین حدود ۲۵ تا ۷۵ می‌باشد. انحراف معیار این توزیع تقریباً ۱۰ است. توزیع خیلی صاف نیست - پستی و بلندی‌هایی دارد- اما روی هم رفته یک دست است، در مرکز اوج گرفته و در دو انتها کوتاه شده است. (میانگین و انحراف معیار واقعی این نمونه خاص به ترتیب ۵۰/۹۸ و ۱۰/۴۲ هستند.)



شکل ۳-۷ هیستوگرام نشان دهنده توزیع نمرات مشکل کلی رفتار

یک چیز که ممکن در این توزیع مورد توجه شما قرار گیرد این است که اگر فراوانی‌های افراد واقع در فواصل ۵۳-۵۲، ۵۴-۵۳ و ۵۵-۵۴ را جمع کنید درخواهید یافت که ۶۵ نفر نمره‌ای بین ۵۲ تا ۵۶ گرفته‌اند. از آنجا که ۳۰۹ مشاهده در این نمونه وجود دارد، $\frac{65}{309} = 21\%$ از مشاهدات در این فاصله قرار می‌گیرند. این گویای توضیحات قبلی مربوط به جمع کردن سطوح است.

چنانچه همین داده‌ها را در نظر بگیریم و آن را با یک نمودار چند ضلعی به جای هیستوگرام ارائه کنیم، شکل ۴-۷ را خواهیم داشت. اطلاعات تازه‌ای در این شکل وجود ندارد که در شکل ۳-۷ نبوده باشد. فقط نقاط بالایی ستون‌های هیستوگرام به هم وصل شده و سپس ستون‌ها کشیده شده‌اند. پس چرا باید برای شکلی که چیز جدیدی ندارد وقت تلف کرد؟ دلیلش این است که به خواننده مراحل انتقال از نمودار ستونی به نمودار خطی نشان داده شود. قدم بعدی انتقال به منحنی‌های صاف است (مثل شکل ۵-۷). تفاوت عمده نمودار خطی (چندضلعی فراوانی) با منحنی‌های صاف این است که منحنی شق خاصی است که ناهمواری‌ها را طی کرده است. اگر مایل باشید می‌توانید منحنی صافی را بر بالای یک هیستوگرام نامرئی در نظر بگیرید.

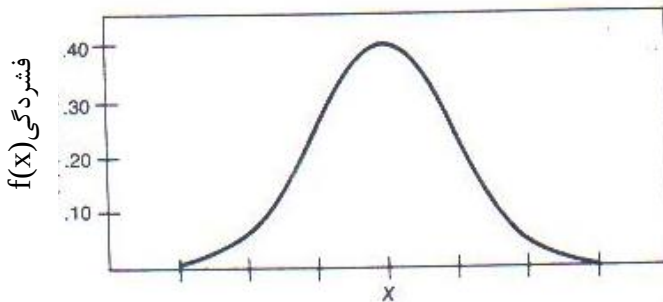


نمره رفتار مشکل

شکل ۴-۷ چندضلعی فراوانی نشان دهنده توزیع نمرات مشکل کلی رفتار

اکنون آمادگی داریم که به سراغ توزیع نرمال برویم. ابتدا به صورت خلاصه به آن نظری می‌اندازیم و سپس یک مثال عینی از آن بر اساس داده‌های نمرات مقیاس خودگزارشی آکنباخ ارائه می‌کنیم که در شکل ۷-۳ و ۷-۴ دیدیم.

توزیع ارائه شده در شکل ۷-۵ یک توزیع نرمال مشخصه است. این یک توزیع متقارن تک‌نمایی است که اغلب به عنوان توزیع “زنگوله‌ای شکل” شناخته می‌شود و حدود آن $\pm\infty$ است. ^۱ بعد افقی^۱ (محور افقی)، مقادیر مختلف ممکن برای X را نشان می‌دهد، در حالی که عرض^۲ (محور عمودی) به عنوان فشردگی منظور می‌شود و به فراوانی یا احتمال وقوع X مربوط است (اما با آن یکسان نیست). مفهوم فشردگی^۳ در فصل آینده بطور کامل توضیح داده خواهد شد.



شکل ۷-۵ یک توزیع نرمال نوعی با مقادیر X روی محور افقی و فشردگی روی محور عمودی

توزیع نرمال دارای یک تاریخچه طولانی است. در اصل بوسیله دی‌موویر (۱۷۵۴-۱۶۶۷) مورد تحقیق قرار گرفت، که علاقه‌مند به کاربرد آن برای توضیح نتایج بازی‌های شانسی (قماربازی) بود. این توزیع بطور دقیق توسط لاپلاس (۱۸۲۷-۱۷۴۹) تعریف شد و از سوی گوس (۱۸۵۵-۱۷۷۷) استفاده‌های سودمندتر آن مد نظر قرار گرفت، که هر دو علاقه‌مند به توزیع خطاهای مشاهدات ستاره‌شناسی بودند. در حقیقت توزیع نرمال بیشتر به عنوان توزیع گوسی و “قانون نرمال خطا” شناخته می‌شود. آدلف گاتلت (۱۸۷۴-۱۷۹۶) ستاره‌شناس بلژیکی اولین کسی بود که این توزیع را برای داده‌های اجتماعی و بیولوژیکی به کار گرفت.

او اندازه‌هایی از قفسه سینه سربازان اسکاتلندی و قد سربازان فرانسوی جمع‌آوری کرد و دریافت که هر دو مجموعه اندازه‌ها تقریباً دارای توزیع نرمال هستند. گاتلت این داده‌ها را تفسیر کرد تا نشان دهد میانگین این توزیع که در آن امور طبیعی هستند ایده‌آل بوده و اینکه مشاهدات در دوسوی میانگین نشان دهنده‌ی خطا (انحراف از ایده‌آل طبیعی آن) هستند. (برای مردانی مثل من با قد ۱۷۲ سانت

1. abscissa
2. ordinate
3. density

آرامش بخش است که فکر کند قد فوتبالیست ۱۸۸ سانتی خطای طبیعت است.) هر چند دیگر ما میانگین را ایده آل طبیعت فرض نمی کنیم اما این مطلب برای فهم کلامی میانگین سودمند است. در واقع، ما هنوز کلمه خطا را برای انحرافات از میانگین به کار می بریم. فرانسویس گالتون (۱۸۲۲-۱۹۱۱) ایده گالت را توسعه داد و به توزیع نرمال در نظریه روان شناختی - بخصوص نظریه توانایی های ذهنی- نقش محوری داد.

توزیع نرمال به صورت ریاضی به این صورت تعریف می شود

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X-\mu)^2/2\sigma^2}$$

که π و e ثابت ها ($\pi = 3/1416$ و $e = 2/7183$) و μ و σ به ترتیب میانگین و انحراف معیار توزیع هستند. با فرض معلوم بودن μ و σ ، محور افقی ($f(X)$) برای هر مقدار X با جانشین کردن μ ، σ و X در معادله بسادگی بدست می آید. این کار چندان که به نظر می رسد دشوار نیست اما در عمل شما هرگز مجبور به این محاسبات نیستید. این توزیع به شکلی فشرده در قالب یک جدول درآمده و به سادگی می توان اطلاعات مورد نیاز را از آن خواند.

کسانی که درس ریاضیات پایه را گذرانده باشند می توانند تشخیص دهند که سطح زیر منحنی بین هر دو مقدار X (مثلا X_1 تا X_2)، و بنابراین احتمال اینکه یک نمره تصادفی در این فاصله قرار گیرد را می توان با اعمال این تابع روی دامنه ای از X_1 تا X_2 پیدا نمود. برای سایر خوانندگان که نمی خواهند سراغ این محاسبات بروند جدول آماده ای در انتهای کتاب های آماری تحت عنوان سطوح زیر منحنی نرمال وجود دارد که برای هر مقدار z ، سطح زیر منحنی معادل آن داده شده است.

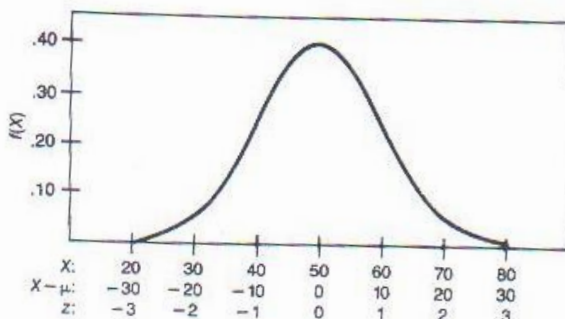
دلیل تهیه جدولی از مقادیر توزیع نرمال این است که این محاسبات برآحتی در دسترس باشند. با استفاده از این جدول بسادگی می توان احتمال اینکه یک نمره که بطور تصادفی از جمعیت بیرون کشیده می شود، مقداری باشد که بین دو نقطه خاص (X_1 و X_2) قرار می گیرد را حساب کرد. بنابراین با بکارگیری جداول آماری ما می توانیم بیانات احتمالی را در پاسخ به سؤال های متنوعی مطرح کنیم.

۷-۱ توزیع نرمال استاندارد

هنگام تلاش برای تهیه جدول توزیع طبیعی یک مشکل پیش می آید، زیرا که این توزیع وابسته به اندازه های میانگین و انحراف معیار (μ و σ) توزیع است. از این رو برای اینکه این کار را درست انجام دهیم ناچاریم برای هر توزیع ممکن از مقادیر μ و σ جدول متفاوتی طرح کنیم. راه حل این مشکل کاملاً ساده است. چیزی که ما واقعاً در این جدول داریم **توزیع نرمال استاندارد**^۱ نامیده می شود، که دارای یک میانگین ۰ و انحراف معیار و واریانس ۱ است. چنین توزیعی اغلب به صورت $N(0, 1)$ نشان داده

می‌شود، که N اشاره به واقعیت نرمال بودن آن دارد، μ مقدار μ و σ^2 مقدار σ^2 است. (بنابراین توصیف عمومی‌تر آن این است $N(\sigma^2, \mu)$) با رجوع به جدول توزیع نرمال در بخش پیوست کتاب و قواعد تبدیل توزیع نرمال به توزیع نرمال استاندارد و برعکس، می‌توان هر مقدار سطح زیر هر منحنی نرمال را پیدا کرد.

به توزیع نشان داده شده در شکل ۶-۷ توجه کنید. این یک توزیع با میانگین ۵۰ و انحراف معیار ۱۰ (واریانس ۱۰۰) است. این شکل نشان دهنده‌ی توزیع یک جمعیت کلی نمرات مشکلات کلی رفتار با فرم خودگزارشی جوانان آکنباخ است که داده‌های شکل ۳-۷ و ۴-۷ نمونه‌ای از آن هستند. چنانچه ما در باره سطوح زیر شکل ۶-۷ چیزی می‌دانستیم می‌توانستیم در باره احتمال مقادیر مختلف نمرات مشکل رفتار چیزی بگوییم و مثلاً نمرات خیلی بالا را که تنها ۵٪ یا ۱۰٪ جمعیت بدست آورده‌اند، شناسایی کنیم.



شکل ۶-۷ یک توزیع نرمال با مقیاس‌های مختلف روی محور افقی

از آنجا که فقط جداول توزیع نرمال که براحتی در دسترس هستند، توزیع نرمال استاندارد می‌باشند، پیش از آنکه بتوانیم سؤال‌های مربوط به احتمال اینکه یک فرد نمره‌ای بالاتر از یک مقدار خاص بگیرد را جواب دهیم، ما بایستی توزیع شکل ۶-۷ (یا حداقل نقاط خاص روی آن را) به یک توزیع نرمال استاندارد تبدیل کنیم. یعنی می‌خواهیم بتوانیم بگوییم که یک نمره X_i از $N(50, 100)$ (یعنی یک توزیع نرمال با میانگین ۵۰ و انحراف معیار ۱۰۰) با یک نمره Z_i از $N(0, 1)$ قابل مقایسه است. پس هر چیز که در باره Z_i درست است در باره X_i هم درست است و X و Z متغیرهای قابل مقایسه‌ای هستند.

از تمرین ۶-۶ دانستیم که با کم کردن یک مقدار ثابت از هر نمره در یک مجموعه نمرات، میانگین آن مجموعه هم به اندازه مقدار ثابت کاهش می‌یابد. بنابراین اگر ما ۵۰ (میانگین) را از همه مقادیر X کم کنیم، میانگین جدید $50 - 50 = 0$ خواهد شد. (به بیان کلی‌تر، توزیع مقادیر $(X - \mu)$ دارای میانگین ۰ است.) اثر این انتقال در مجموعه مقادیر دوم محور افقی در شکل ۶-۷ نشان داده شده است. ما در نیمه

راه هستیم چون اکنون میانگین را به ۰ کاهش داده‌ایم اما هنوز انحراف معیار (σ) ۱۰ است. ما همچنین از تمرین ۶-۶ می‌دانیم که اگر همه مقادیر یک متغیر را بر مقدار ثابتی (مثلاً ۱۰) تقسیم کنیم، انحراف معیار به همان نسبت تغییر می‌کند (بر مقدار ثابت تقسیم می‌شود). بنابر این انحراف معیار حالا $10/10 = 1$ خواهد بود، که درست همان چیزی است که ما می‌خواهیم. این توزیع تغییر شکل داده شده را توزیع z نامیده و براساس کاری که انجام داده به این صورت آن را تعریف می‌کنیم:

$$z = \frac{X - \mu}{\sigma}$$

برای مورد خاص ما، که $\mu = 50$ و $\sigma = 10$ است،

$$z = \frac{X - \mu}{\sigma} = \frac{X - 50}{10}$$

سومین مجموعه مقادیر در شکل ۶-۷ (نمرات z) اثر این انتقال را نشان می‌دهند. توجه داشته باشید که به غیر از یک انتقال خطی^۱ مقادیر عددی، داده‌ها تغییر دیگری نمی‌کنند. توزیع شکل خودش را خواهد داشت و روابط مشاهدات با یکدیگر به صورت قبل از تغییر شکل یافتن خواهد بود. تعجبی ندارد که تغییر واحد اندازه‌گیری باعث تغییر شکل توزیع یا روابط مربوط به مشاهدات نشود. اگر ما مقدار مصرف گوشت مردم در هفته را بر حسب کیلو یا پوند اندازه بگیریم در روابط نسبی مربوط به مصرف گوشت تغییری ایجاد نمی‌شود. تنها مقادیر عددی روی محور افقی تغییر می‌کنند. (پرمصرف‌ها همچنان پرمصرف هستند چه به کیلو، چه به پوند.) مهم است که بدانیم در واقع در تبدیل X به z چه کاری صورت گرفته است. یک نمره که بایستی ۶۰ باشد اینک ۱ است. یعنی یک نمره که بایستی یک انحراف معیار (۱۰ نمره) بالای میانگین باشد یک انحراف معیار بالای میانگین باقی مانده اما حالا یک مقدار جدید ۱ گرفته است. نمره ۴۵ که 0.5 انحراف معیار زیر میانگین بود، اینک مقدار -0.5 گرفته است و الی آخر. به عبارت دیگر یک نمره z ، معرف میزان انحراف معیار X_i در بالا یا پایین میانگین است.

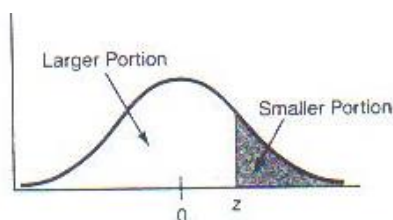
معادله z کاملاً فراگیر است. ما بسادگی با اعمال این معادله می‌توانیم هر توزیعی را به توزیع z تبدیل کنیم. در نظر داشته باشید که ما صرفاً یک تبدیل انجام می‌دهیم. شکل توزیع تحت تأثیر این تبدیل قرار نمی‌گیرد. یعنی اگر توزیع قبل از تبدیل نرمال نبوده، با این تبدیل نرمال نخواهد شد. برخی فکر می‌کنند که با تبدیل نمرات به نمره z آنها می‌توانند داده‌های خود را "نرمال" کنند (به خیال اینکه توزیع نرمال می‌سازند). چنین برداشتی کاملاً اشتباه است.

– استفاده از جداول توزیع نرمال استاندارد

چنانکه گفته شد، توزیع نرمال استاندارد به طور گسترده‌ای در قالب جدول در آمده است. چنین جدولی در انتهای این کتاب آمده است. بخشی از این جدول در جدول ۷-۱ ارائه شده است. برای اینکه ببینید

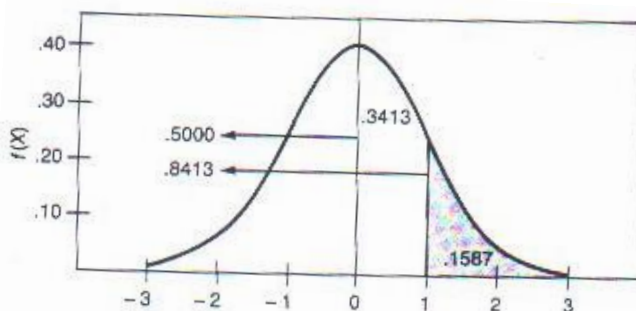
چگونه از این جدول استفاده می‌شود به توزیع نرمالی که در شکل ۷-۷ آمده است توجه کنید. این توزیع استاندارد نمرات مشکل رفتاری را نشان می‌دهد که در شکل ۷-۶ آمده بود. فرض کنید بخواهید بدانیم چه بخشی از سطح زیر منحنی بالاتر از ۱ انحراف استاندارد از میانگین است، در حالی که کل سطح زیر منحنی برابر ۱/۰۰ است. (ما در باره سطح احتیاط می‌کنیم چون اینها مستقیماً به احتمالات تبدیل می‌شوند.) قبلاً دیدیم که نمرات z انحراف استاندارد از میانگین را نشان می‌دهند و بنابراین می‌دانیم که می‌خواهیم سطح بالای $z=1$ را پیدا کنیم.

تنها نیمه مثبت توزیع نرمال در قالب جدول در آمده است. از آنجا که توزیع متقارن است، هر نوع اطلاعات مربوط به مقادیر مثبت z ، بطور یکسان در مورد مقادیر منفی z به کار می‌رود. از جدول ۱-۷، ردیف مربوط به $z=1/00$ را پیدا می‌کنیم. با دنبال کردن این ردیف می‌توانیم ببینیم که سطح از میانگین تا $z=1$ برابر ۰/۳۴۱۳ است، سطح بخش بزرگتر ۰/۸۴۱۳، و سطح بخش کوچکتر ۰/۱۵۸۷ است. اگر توزیع حد فاصل قسمت $z=1$ بالا (قسمت سایه‌دار) و $z=1$ پایین (قسمت بدون سایه) را در نظر بگیرید معنای عبارات بخش بزرگتر و بخش کوچکتر روشن می‌شود. بر این اساس جواب سؤال اولیه ما ۰/۱۵۸۷ است. چون ما قبلاً واژه‌های سطح و احتمال را معادل گرفتیم (رجوع کنید به فصل ۳)، اکنون می‌گوییم اگر ما یک کودک را از جمعیت کودکان بطور تصادفی بیرون بکشیم و اگر نمرات مشکل رفتاری دارای توزیع نرمال باشند، آنگاه احتمال اینکه این کودک نمره‌ای بالاتر از یک انحراف معیار بالای میانگین جمعیت خود (مثلاً بالای ۶۰) داشته باشد برابر ۰/۱۵۸۷ است. از آنجا که توزیع متقارن است همچنین می‌دانیم که احتمال اینکه یک کودک نمره‌ای بالاتر از یک انحراف معیار زیر میانگین بگیرد برابر ۰/۱۵۸۷ است.



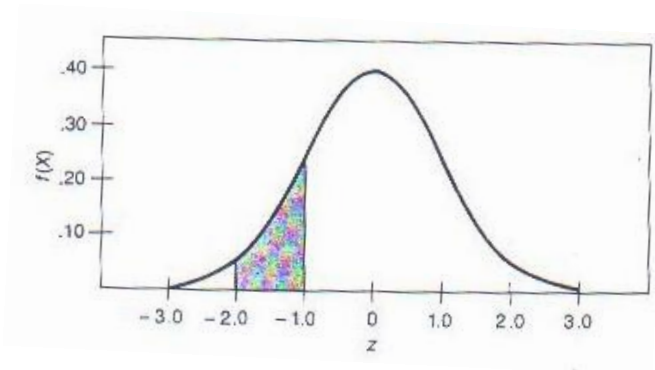
Mean to		Larger Portion	Smaller Portion	Mean to		Larger Portion	Smaller Portion
z	z			z	z		
0.00	0.0000	0.5000	0.5000	0.45	0.1736	0.6736	0.3264
0.01	0.0040	0.5040	0.4960	0.46	0.1772	0.6772	0.3228
0.02	0.0080	0.5080	0.4920	0.47	0.1808	0.6808	0.3192
0.03	0.0120	0.5120	0.4880	0.48	0.1844	0.6844	0.3156
0.04	0.0160	0.5160	0.4840	0.49	0.1879	0.6879	0.3121
0.05	0.0199	0.5199	0.4801	0.50	0.1915	0.6915	0.3085
...	...	...	...	...	...	...	...
0.97	0.3340	0.8340	0.1660	1.42	0.4222	0.9222	0.0778
0.98	0.3365	0.8365	0.1635	1.43	0.4236	0.9236	0.0764
0.99	0.3389	0.8389	0.1611	1.44	0.4251	0.9251	0.0749
1.00	0.3413	0.8413	0.1587	1.45	0.4265	0.9265	0.0735
1.01	0.3438	0.8438	0.1562	1.46	0.4279	0.9279	0.0721
1.02	0.3461	0.8461	0.1539	1.47	0.4292	0.9292	0.0708
1.03	0.3485	0.8485	0.1515	1.48	0.4306	0.9306	0.0694
1.04	0.3508	0.8508	0.1492	1.49	0.4319	0.9319	0.0681
1.05	0.3531	0.8531	0.1469	1.50	0.4332	0.9332	0.0668
...	...	...	...	...	...	...	...
1.95	0.4744	0.9744	0.0256	2.40	0.4918	0.9918	0.0082
1.96	0.4750	0.9750	0.0250	2.41	0.4920	0.9920	0.0080
1.97	0.4756	0.9756	0.0244	2.42	0.4922	0.9922	0.0078
1.98	0.4761	0.9761	0.0239	2.43	0.4925	0.9925	0.0075
1.99	0.4767	0.9767	0.0233	2.44	0.4927	0.9927	0.0073
2.00	0.4772	0.9772	0.0228	2.45	0.4929	0.9929	0.0071
2.01	0.4778	0.9778	0.0222	2.46	0.4931	0.9931	0.0069
2.02	0.4783	0.9783	0.0217	2.47	0.4932	0.9932	0.0068
2.03	0.4788	0.9788	0.0212	2.48	0.4934	0.9934	0.0066
2.04	0.4793	0.9793	0.0207	2.49	0.4936	0.9936	0.0064
2.05	0.4798	0.9798	0.0202	2.50	0.4938	0.9938	0.0062

جدول ۷-۱ خلاصه‌ای از جدول توزیع نرمال



شکل ۷-۷ نمایش سطوح زیر منحنی نرمال

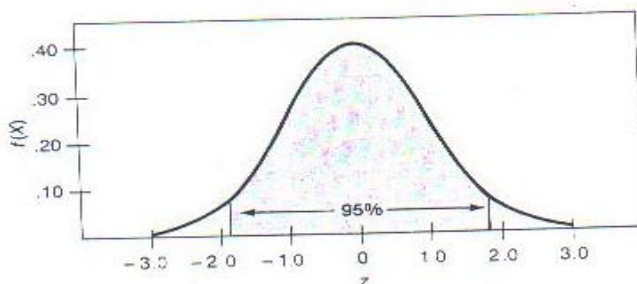
اکنون فرض کنید که بخواهیم این احتمال را بررسی کنیم که این کودک در فاصله یک انحراف معیار (۱۰ نمره) از میانگین در هر سمت قرار بگیرد. این یک موضوع ساده مربوط به جمع سطوح است. چون توزیع نرمال متقارن است، سطح زیر $z = -1$ به اندازه سطح بالای $z = +1$ است. به همین دلیل مقادیر منفی z در جدول نیامده است- نیازی به تکرار آنها نیست. اکنون می‌دانیم که سطوحی که مورد نظر ما هستند، هر یک 0.1587 می‌باشند. پس کل سطح خارج از $z = \pm 1$ بایستی $0.1587 + 0.1587 = 0.3174$ باشد. اگر سطح خارج از $z = \pm 1$ است پس سطح بین $z = +1$ و $z = -1$ برابر است با $0.6826 = 1 - 0.3174$. بنابراین احتمال اینکه یک کودک نمره‌ای بین ۴۰ و ۶۰ بگیرد 0.6826 است. برای گسترش این شیوه، توجه خواننده را به موقعیتی جلب می‌کنیم که بخواهیم احتمال قرار گرفتن یک نمره بین ۳۰ و ۴۰ را بدانیم. با یک حساب ساده مشخص می‌گردد که این احتمال بین $1/10$ تا $2/10$ انحراف معیار زیر میانگین قرار دارد. این موقعیت در نمودار شکل ۷-۸ نشان داده شده است. از جدول مربوط به سطوح زیر منحنی نرمال در می‌یابیم که سطح بین میانگین تا $z = -2/10$ برابر 0.4772 و از میانگین تا $z = -1/10$ برابر 0.3413 است. اختلاف این دو سطح معرف سطح بین $z = -2/10$ و $z = -1/10$ است. این سطح برابر $0.1359 = 0.4772 - 0.3413$ است. بنابراین احتمال اینکه نمرات مشکل رفتاری در حالت تصادفی از یک جمعیت با توزیع نرمال، بین ۳۰ تا ۴۰ باشد برابر 0.1359 خواهد بود.



شکل ۸-۷ سطوح بین ۱/۰ و ۲/۰ انحراف معیار زیر میانگین

۲-۷ تنظیم حدود احتمالی در یک مشاهده

به عنوان مثال آخر، به شرایطی توجه نمائید که بخواهیم حدودی را شناسایی کنیم که یک کودک به طور تصادفی با میزان معینی از اطمینان در آن واقع می‌شود. به عبارت دیگر ما می‌خواهیم بگوییم "اگر من کودکی را بطور تصادفی از یک جمعیت بیرون بکشم، در ۹۵٪ موارد نمره او بین $z = -2.0$ و $z = 2.0$ باشد." در شکل ۹-۷، می‌توانید حدودی که ما می‌خواهیم - که شامل ۹۵٪ نمرات جمعیت است را مشاهده کنید. وقتی بدنبال حدودی هستیم که ۹۵٪ نمرات در آن قرار می‌گیرند، همزمان در پی مواردی هستیم که ۵٪ نمرات باقیمانده قرار می‌گیرند. برای مشخص کردن این ۵٪ باقیمانده باید مقدار z مربوط به نقطه برش ۲/۵٪ (دنبالهای) توزیع را پیدا کنیم. در جدول سطوح زیر منحنی نرمال مشاهده می‌کنیم که این مقادیر $z = \pm 1.96$ هستند. بنابراین می‌توان گفت در ۹۵٪ اوقات نمره یک کودک که بطور تصادفی نمونه‌گیری شده بین ۱/۹۶ انحراف معیار بالاتر از میانگین و ۱/۹۶ انحراف معیار پایین‌تر از میانگین قرار خواهد گرفت.



شکل ۹-۷ مقادیر z که ۹۵٪ نمرات مشکلات رفتاری را در بر می‌گیرند

از آنجا که معمولا ما یلیم جواب هایمان را در قالب نمرات خام مشکل رفتاری بیان کنیم نه نمرات z ، بایستی کار کوچک دیگری انجام دهیم. برای بدست آوردن حدود نمرات خام، فرمول z برای تبدیل X به z را برعکس به کار می گیریم، بنابراین اگر بخواهیم حدودی که ۹۵٪ جمعیت بین آن قرار می گیرند را بیان کنیم، می خواهیم نمراتی را پیدا کنیم که ۱/۹۶ انحراف معیار بالا یا پایین میانگین جمعیت قرار دارند. به صورت زیر می توان آن را نوشت

$$\begin{aligned}\frac{X-\mu}{\sigma} &= z \\ \frac{X-\mu}{\sigma} &= \pm 1/96 \\ \sigma &= \pm 1/96 X - \mu \\ \sigma \pm 1/96 \mu &= X\end{aligned}$$

که مقادیر X مربوط به $\sigma + 1/96 \mu$ و $\mu - 1/96 \sigma$ بیانگر حدودی است که ما جستجو می کنیم. برای مثال ما، خواهیم داشت

$$حدود = 50 \pm (1/96)(10) = 50 \pm 19/6 = 30/4 \text{ و } 69/6$$

بنابراین به احتمال ۰/۹۵ نمره یک کودک (X) که بطور تصادفی انتخاب شده بین ۳۰/۴ و ۶۹/۶ خواهد بود. ما خیلی به نمرات پایین علاقه مند نیستیم، زیرا آنها مشکلات را نشان نمی دهند. اما هر کسی با نمره ۶۹/۶ دارای مشکل است. تنها ۲/۵٪ کودکان در این حد بالا نمره داده می شوند.

چیزی که در اینجا مورد بحث قرار گرفت با آنچه بعدا به عنوان حدود اطمینان بیان می شود ارتباط نزدیکی دارد اما کاملا یکسان نیست. تفاوت اساسی این است که در اینجا ما میانگین جمعیت را می دانستیم و سعی کردیم برآورد کنیم یک مشاهده منفرد (X) در کجا قرار خواهد گرفت. وقتی حدود اطمینان را مورد بحث قرار دهیم، یک میانگین نمونه ای (یا آماره ای دیگر) خواهیم داشت که می خواهیم حدودی را تعیین کنیم که به احتمال ۰/۹۵ میانگین جمعیت (یا پارامتر مرتبط دیگری) را در بر می گیرند. در اینجا شما اصلا نیازی ندارید که در باره حدود اطمینان چیزی بدانید. تنها به این خاطر اشاره ای به آن شد که از سردرگمی در آینده پیشگیری شود.

۷-۳ استفاده از توزیع بهنجار برای محاسبه نقاط درصدی

در مثال قبلی ما از توزیع بهنجار برای محاسبه سطح زیر، بالا، یا میان مقادیر مورد نظر X استفاده کردیم. همین روال را برای حل مسائل مربوط به نقاط درصدی می توان بکار بست.

در فصل ۴ نقاط درصدی را با نگاهی به توزیع فراوانی داده ها محاسبه کردیم و تقریبا بطور ساده نمراتی که پایین تر از برخی نقاط خاص قرار می گرفتند را می شمردیم. به هر حال اگر بدانیم که توزیع نرمال (یا تقریبا نرمال) است می توانیم از اطلاعات قبلی خود برای برآورد نقاط درصدی و رتبه های درصدی استفاده کنیم. به عنوان یک مثال، فرض کنید می خواهیم نقطه درصدی ۹۰ ام را در نمرات

مشکلات رفتاری حساب کنیم. ممکن است مثلاً بخواهیم از آن به عنوان یک نقطه برش برای شناسایی افرادی که در حدود بیمارستانی قرار می‌گیرند - یعنی مشکلات شدید رفتاری در حد نیاز به بستری شدن نشان می‌دهند استفاده کنیم. ما می‌دانیم که نقطه درصدی ۹۰ آن نقطه‌ای است که ۹۰٪ نمرات زیر آن قرار می‌گیرند. از جدول سطح زیر منحنی نرمال درمی‌یابیم که ۹۰٪ این سطح زیر $z = 1/28$ قرار دارد. چیزی که لازم است ما بدانیم این است که چه مقدار X (نمره) با $z = 1/28$ تطبیق دارد. این مقدار را می‌توان با جایگزین کردن مقادیر μ ، σ و X در فرمول z و محاسبه X بدست آورد.

$$z = \frac{X - \mu}{\sigma}$$

$$1/28 = \frac{X - 50}{10}$$

$$10(1/28) = X - 50$$

$$12/8 = X - 50$$

$$50 + 12/8 = X$$

$$X = 62/8$$

بنابراین نقطه درصدی ۹۰ آن در میان نمرات مشکل رفتاری برابر با $62/8$ است. به عنوان مثالی دیگر، فرض کنید بخواهیم نقطه درصدی ۲۵ آن را پیدا کنیم. از جدول سطوح زیر منحنی نرمال در می‌یابیم که یک نمره z معادل $-0/67$ قطع کننده ۲۵٪ پایین است. (توجه داشته باشید که z منفی است چون ما بدنبال نقطه‌ای در نیمه پایین توزیع هستیم) برای بدست آوردن ارزش منطبق با $z = -0/67$ از همان شیوه مثال قبل استفاده می‌کنیم.

$$z = \frac{X - \mu}{\sigma}$$

$$-0/67 = \frac{X - 50}{10}$$

$$10(-0/67) = X - 50$$

$$-6/7 = X - 50$$

$$50 - 6/7 = X$$

$$X = 43/3$$

بنابراین نقطه درصدی ۲۵ آن روی نمره $43/3$ قرار دارد. اکنون فرض کنید که کودکی در عمل نمره ۷۵ دارد. رتبه درصدی معادل این نمره چیست؟ در اینجا مسأله را به صورت وارونه حل می‌کنیم. به جای پیدا کردن z مربوط به سطح خاص و حل X ، ما از X شروع کرده و آن را به z تبدیل می‌کنیم، و سطح منطبق با آن مقدار z را پیدا می‌کنیم. بطور مثال

$$z = \frac{X - \mu}{\sigma}$$

$$\frac{75-50}{10} = \frac{25}{10} = 2.5$$

از جدول سطح زیر منحنی نرمال درمی‌یابیم که $99/38\%$ این سطح زیر $z = 2.5$ قرار دارد. بنابراین نتیجه می‌گیریم کودکی که نمره ۷۵ دارد دارای رتبه درصدی $99/3\%$ است که او را در مقایسه با دیگران یک کودک خیلی مشکل‌دار می‌سازد.

وقتی از توزیع نرمال برای نقاط درصدی استفاده می‌کنیم، باید در نظر داشت که چنین شیوه‌ای نیازمند این پیش‌فرض است که نمرات مورد نظر دست‌کم تقریباً نرمال هستند. اگر داده‌ها خیلی نامتقارن (چوله) یا غیر نرمال باشند، بدست آوردن نقاط درصدی با استفاده از توزیع نرمال می‌تواند نتایج اشتباه‌آمیزی را به بار آورد.

۷-۴ اندازه‌های مربوط به z

پیش از این دیدیم که فرمول z را می‌توان برای تبدیل یک توزیع با هر میانگین و واریانس به یک توزیع با میانگین ۰ و انحراف معیار (واریانس) ۱ به کار برد. ما مرتباً به این تبدیل به عنوان **نمرات استاندارد** رجوع می‌کنیم. البته نظام‌های تبدیلی دیگری هم با ویژگی‌های خاص وجود دارند، بعضی از آنها را مردم بطور روزمره استفاده می‌کنند بدون توجه به اینکه آنها چه هستند.

نمونه‌ی خوبی از این نظام نمره‌های IQ هوش است. نمرات خام یک آزمون IQ که بطور معمول تبدیل به یک توزیع با میانگین ۱۰۰ و انحراف معیار ۱۵ (یا ۱۶ در تست بینه) می‌شوند. با این آگاهی، شما می‌توانید به سرعت IQ یک فرد (مثلاً ۱۲۰) را به موقعیت او به شکل نمره انحراف بالا یا پایین میانگین تبدیل کنید (یعنی نمره z او را حساب کنید). از آنجا که نمرات IQ دارای توزیع کم و بیش نرمال هستند، پس شما می‌توانید z را به اندازه‌های درصدی با استفاده از جدول سطوح زیر منحنی نرمال تبدیل کنید. (در این مثال یک نمره ۱۲۰ تقریباً در نقطه درصدی ۹۱ ام قرار می‌گیرد.)

نمونه‌های دیگری از این نوع، بعضی آزمون‌های ملی [در آمریکا] مثل SAT (آزمون پیشرفت تحصیلی) هستند. در این موارد نمرات خام با روند آزمون تبدیل پیدا می‌کنند و به عنوان نمرات حاصل از یک توزیع با میانگین ۵۰۰ و انحراف معیار ۱۰۰ (حداقل وقتی آزمون برای اولین بار طرح شد) گزارش می‌شوند. ساختن چنین نظام‌های نمره‌گذاری بسیار آسان است. این کار با تبدیل نمرات خام به نمرات z شروع می‌شود (بر اساس میانگین و انحراف معیار نمرات خام). سپس نمرات z به نظام نمره‌گذاری مورد نظر تبدیل می‌شوند. بنابراین

$$\text{میانگین جدید} + (z) \text{ انحراف معیار جدید} = \text{نمره جدید}$$

که z نشان دهنده‌ی نمره z معادل با نمره خام فرد است. در مورد SAT

$$100(z) + 500 = \text{نمره جدید}$$

۷-۵ خلاصه

در این فصل ما توزیع بهنجار (نرمال) را بررسی کردیم. دیدیم که چگونه از توزیع بهنجار استاندارد برای بدست آوردن سطوح زیر هر توزیع بهنجار استفاده کنیم و مشاهده کردیم که این سطوح مستقیماً با احتمالات مرتبط هستند. ما همچنین دیدیم که چگونه از توزیع بهنجار برای محاسبه نقاط و رتبه‌های درصدی بهره ببریم. برخی از مطالب مهم فصل عبارتند از:

- توزیع نرمال (بهنجار)
- محور افقی
- محور عمودی
- توزیع استاندارد بهنجار
- انتقال (تبدیل) خطی
- نمرات z
- نمرات استاندارد

۷-۶- تمرین‌ها

۷-۱ فرض کنید داده‌های زیر از یک جمعیت با $\mu = 4$ و $\sigma = 1/36$ باشد:

$X = 1 \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \quad 4 \quad 4 \quad 4 \quad 4 \quad 5 \quad 5 \quad 6 \quad 6 \quad 7$

(الف) نمودار توزیع را رسم کنید.

(ب) توزیع تمرین ۷-۱ (الف) را به یک توزیع با $X - \mu$ تبدیل کنید.

(ج) توزیع تمرین ۷-۱ (ب) را به یک توزیع z تبدیل کنید.

۷-۲ با استفاده از توزیع تمرین ۷-۱، برای نمرات $2/5$ ، $6/2$ ، X ، مقدار z را حساب کنید. این نتایج را توضیح دهید.

۷-۳ فرض کنید می‌خواهیم خطاهای مربوط به اجرای یک تکلیف ساده را مطالعه کنیم. یک قلک حاوی حدود ۱۰۰۰۰ تومان را به صورت ده تومانی انتخاب نموده و از آزمودنی‌های زیادی می‌خواهیم که آنها را بشمارند. در می‌یابیم که میانگین ده تومانی‌های گزارش شده ۹۷۵ با انحراف معیار ۱۵ است. با فرض نرمال بودن توزیع شمارش‌ها:

(الف) چند درصد از شمارش‌ها بین ۹۶۰ و ۹۹۰ خواهد بود؟

(ب) چند درصد زیر ۹۷۵ خواهد بود؟

(ج) چند درصد زیر ۹۹۰ خواهد بود؟

۷-۴ با استفاده از تمرین ۷-۳:

(الف) کدام دو مقدار X (شمارش) در برگیرنده ۵۰٪ میانی نتایج خواهد بود؟

(ب) ۷۵٪ از شمارش‌ها کمتر از _____ خواهد بود.

(ج) ۹۵٪ شمارش‌ها بین ____ و ____ خواهد بود.

۷-۵ رئیس یک بخش ده تومانی‌ها را شمرده و ادعا می‌کند فقط ۹۵۰ سکه در قلم است. آیا می‌توان گفت که او مسئولانه سکه‌ها را شمارش کرده است؟

۷-۶ یک مجموعه از نمره‌های خواندن کودکان پایه چهارم دارای میانگین ۲۵ و انحراف معیار ۵ است. یک مجموعه از نمرات کودکان پایه نهم دارای میانگین ۳۰ با انحراف معیار ۱۰ است. با فرض نرمال بودن این توزیع‌ها:

(الف) یک نمودار ساده از این داده‌ها در کنار هم رسم کنید.

(ب) چند درصد از نمرات پایه چهارم از متوسط نمرات نهم بهتر بوده است؟

(ج) چند درصد نمرات نهم بدتر از متوسط نمرات چهارم بوده است؟ (منظور از این محاسبات را در فصل ۱۵ توضیح می‌دهیم).

۷-۷ تحت چه شرایطی جواب قسمت‌های (الف) و (ب) تمرین ۷-۶ یکسان خواهد بود؟

۷-۸ آزمون‌های تشخیصی خاصی نشان می‌دهند تنها در صورتی که کودکی نمره‌ای کوچکتر یا مساوی نقطه درصدی ۱۰ بگیرد نشان دهنده مشکل است. اگر میانگین نمرات ۱۵۰ و انحراف معیار ۳۰ باشد، نقطه برش تشخیصی معنادار کدام خواهد بود؟

۷-۹ رئیس اداره‌ای باید برای سال آینده افزایش حقوق‌ها را تنظیم کند. او تصمیم دارد میانگین افزایش ۲۰۰۰ دلار با انحراف معیار ۴۰۰ دلار، و توزیع نرمال باشد.

(الف) ده درصد از کارکنان افزایشی معادل یا بیشتر از ____ دلار خواهند داشت.

(ب) ۵٪ از کارکنان که کارایی نداشته‌اند بیش از ____ دلار دریافت نخواهند کرد.

۷-۱۰ در یک درس از دانشجویان خواسته می‌شود تا به مشاهده این موضوع بپردازند که آیا مردم کمربند ایمنی اتومبیل را می‌بندند. هر فرد باید ۱۰۰ اتومبیل را مشاهده و افرادی که کمربند بسته‌اند را بشمارد. تعدادی که هر دانشجو بدست می‌آورد به عنوان نمره او حساب می‌شود. میانگین نمره کلاس ۳۰ و انحراف معیار ۷ است.

(الف) این توزیع را رسم کنید. فرض کنید توزیع نرمال است.

(ب) یک دانشجو که در طول سال کم کاری کرده ۵۰ نفر (از ۱۰۰) را گزارش کرده است که کمربند بسته بودند. آیا شواهدی وجود دارد که گفته شود او فقط یک عدد را بیان کرده بدون اینکه واقعا شمارشی به عمل آورده باشد؟

۷-۱۱ چند سال قبل دوستی یک آزمون تشخیصی مشکلات گفتاری را طرح کرد. نمره مقیاس او بطور ساده تعداد جملات گفتاری (مثل جمع، منفی، مثبت) درستی که کودک در پاسخ آزمون‌گر می‌سازد را شمارش می‌کند. این آزمون دارای میانگین ۴۸ و انحراف معیار ۷ است. والدین در درک معنای نمره این

مقیاس مشکل دارند و تهیه کننده می‌خواهد میانگین را به ۸۰ و انحراف معیار را به ۱۰ تغییر دهد (تا برای استفاده والدین راحت‌تر باشد). باید چه کار کند؟

۷-۱۲ متأسفانه همه جهان بر اساس اصول توزیع نرمال ساخته نشده است. در مثال قبل توزیع واقعی بسیار چوله است چون بیشتر کودکان دارای مشکلات گفتاری نیستند و بنابراین همه جملات را درست می‌سازند.

(الف) چگونگی ظاهر شدن این توزیع را رسم کنید.

(ب) یافتن نقطه برش مربوط به ۱۰٪ پایین چگونه باید صورت بگیرید در حالی که توزیع نرمال نیست؟
۷-۱۳ میانگین و انحراف معیار یک آزمون فارغ‌التحصیلی برای تمام افراد شرکت کننده به ترتیب ۴۸۹ و ۱۲۶ بوده است. با استفاده از اطلاعات خود در باره توزیع نرمال رتبه درصدی نمره ۶۰۰ را حساب کنید.
۷-۱۴ در تمرین ۷-۱۳ چه نمره‌ای با نقطه درصدی ۷۵ آم تطبیق دارد.

۷-۱۵ اگر برای تمام دانشجویانی که در امتحان فارغ‌التحصیلی شرکت کرده‌اند میانگین و انحراف معیار به ترتیب ۵۰۷ و ۱۱۸ باشد، این موضوع بر جواب تمرین‌های ۷-۱۳ و ۷-۱۴ چه اثری می‌گذارد؟

۷-۱۶ جواب پرسش ۷-۱۵ در باره اهمیت گروه‌های مرجع چه چیزی را بیان می‌کند؟

۷-۱۷ داده‌های نمونه‌ای طرح کنید که دارای کجی (چولگی) زیادی باشند و نشان دهد که استفاده از توزیع نرمال برای نقاط درصدی تقریبی، نتایج گمراه کننده‌ای به همراه دارد.

۷-۱۸ فرض کنید نمرات مشکل رفتاری مورد بحث در این فصل از یک جمعیت با میانگین ۵۰ و انحراف معیار ۱۰ باشد. نقطه برش تشخیصی معنادار برای ۲٪ بالایی جمعیت چه خواهد بود.

فصل ۸

مفاهیم اساسی احتمالات

۸-۱- احتمال

۸-۲- واژه‌ها و قواعد اصلی

۸-۳- متغیرهای گسسته در مقابل پیوسته

۸-۴- توزیع احتمالات برای متغیرهای گسسته

۸-۵- توزیع احتمالات برای متغیرهای پیوسته

در فصل ۷ ما شروع به استفاده از مفهوم احتمالات کردیم. برای مثال دیدیم که حدود ۶۸٪ از کودکان دارای نمرات مشکلات رفتاری بین ۴۰ و ۶۰ هستند و بنابر این نتیجه گرفتیم که اگر کودکی را بطور تصادفی انتخاب کنیم، احتمال اینکه او دارای نمره بین ۴۰ تا ۶۰ باشد ۰/۶۸ می‌باشد. در فصل ۱۱ که توجه خود را بر آماره‌های استنباطی متمرکز می‌کنیم، تکیه زیادی بر بیان احتمالات خواهیم داشت. در آنجا مطالبی از این نوع خواهیم داشت "اگر این فرضیه درست باشد، فقط ۰/۱۵ احتمال وجود دارد که داده‌هایی را بدست آوریم که در عمل بدست آورده‌ایم." وقتی باید به عبارات احتمالات تکیه کنیم، پس لازم است منظور خود را از احتمال بفهمیم و بعضی قواعد اساسی محاسبه و دستکاری احتمالات را بدانیم. هدف این فصل همین است.

مطالب این فصل به دو دلیل انتخاب شده‌اند. یا این مطالب مستقیماً فهم مطالب بخش‌های بعدی کتاب را ممکن می‌سازند یا امکان محاسبات ساده احتمالات را برای خواننده فراهم می‌کنند که می‌تواند برای شما سودمند باشد. مطالبی که این دو هدف را برآورده نمی‌ساختند حذف شده‌اند. بنابراین، مثلاً ما به مواردی از قبیل احتمال بیرون کشیدن یک ورق بازی از بین ۱۴ کارت نپرداخته‌ایم. یا به احتمال

اینکه چراغ مطالعه شما در ۲۵ ساعت آینده بسوزد نپرداخته‌ایم. کسانی که به این عناوین علاقه دارند باید به کتاب‌ها و دوره‌های درسی نظریه احتمالات مراجعه کنند تا این مطالب را عمیق‌تر بیاموزند.

۸-۱ احتمال

مفهوم احتمال را به شیوه‌های مختلفی می‌توان در نظر گرفت. حتی درباره معنای کلمه احتمال^۱ هم توافق همگانی وجود ندارد. قدیمی‌ترین و معروف‌ترین تعریف احتمال چیزی است که **دیدگاه تحلیلی**^۲ نامیده می‌شود. من یک جعبه شکلات در کشو میزم پنهان دارم. (آن پنهان است زیرا یادگرفته‌ام که به همکارانم اعتماد نکنم.) این بسته دارای ۸۵ کارامل شیرین است که من دوست دارم، و ۱۵ کارامل تلخ که برای همکاران آبنبات‌خورم ذخیره کرده‌ام. وقتی گرسنه باشم سراغ جعبه شکلات می‌روم و یکی را بطور شانسی برمی‌دارم. احتمال اینکه کارامل شیرین باشد چقدر است؟ از آنجا که ۸۵ عدد از ۱۰۰ کارامل شیرین و ۱۵ عدد تلخ هستند و چون نمونه‌گیری تصادفی است، احتمال (p) بیرون کشیدن کارامل شیرین $0.85 = 85 / 100$ است. این مثال تعریف احتمال را نشان می‌دهد.

اگر رخدادی به A شیوه رخ دهد و بتواند به B شیوه رخ ندهد و اگر احتمال همه شیوه‌های ممکن به یک اندازه باشد (مثلاً هر کارامل شانسی برابر برای برداشته شدن داشته باشد)، آنگاه احتمال رخداد آن برابر $A/(A+B)$ و احتمال عدم وقوع آن برابر $B/(A+B)$ است.

چون ۸۵ شیوه برای بیرون کشیدن کارامل شیرین وجود دارد (یکی برای هر یک از ۸۵ کارامل) و ۱۵ شیوه برای انتخاب کارامل تلخ، بنابراین: $A = 85$ ، $B = 15$ و $P(A) = 85 / (85 + 15) = 0.85$. یک دیدگاه دیگر در باره احتمالات **دیدگاه فراوانی نسبی** است. در آوردن کارامل از این جعبه را با توجه به طعم (رنگ) کارامل در هر بار در نظر بگیرید. در اجرای این مطالعه نمونه‌ای، ما **نمونه‌گیری** را با **جایگزینی** انجام می‌دهیم، یعنی هر کارامل دوباره به جای خودش برمی‌گردد قبل از اینکه کارامل بعدی برداشته شود. اگر تعداد خیلی زیادی نمونه‌گیری صورت دهیم، درخواهیم یافت که (تقریباً) 85% موارد به کارامل شیرین منجر خواهد شد. بنابراین ما باید احتمال را به عنوان میزان فراوانی نسبی وقوع پدیده مورد نظر تعریف کنیم که با افزایش تعداد استخراج‌ها به آن می‌رسیم.

برداشت سومی هم از مفهوم احتمال وجود دارد که بعضی نظریه‌ها از آن دفاع می‌کنند. این شامل مفهوم **احتمال ذهنی** است. از این دیدگاه، احتمال به عنوان باور ذهنی یک نفر در باره امکان وقوع یک پدیده تعریف می‌شود. بنابراین بطور مثال، این جمله که “من فکر می‌کنم فردا هوا خوب است” یک عبارت ذهنی در باره درجه باور است که احتمالاً در باره فراوانی نسبی وقوع روزهای خوب کمتر چیزی برای گفتن دارد، و در حقیقت ممکن است اصلاً اساس محاسباتی نداشته باشد. منظور این نیست که این

1. probability
2. analytic view

دیدگاه احتمال، اساس منطقی ندارد. احتمالات ذهنی در تصمیم‌گیری‌های بشر سهم مهمی دارند و بر همه جنبه‌های رفتار ما چیره هستند. در همین حال، تصمیمات آماری به گونه‌ای که ما آنها را می‌سازیم عموماً می‌توانند به صورت رویکردهای محاسباتی‌تری بیان شوند، هر چند تفسیر آنها هم بسیار ذهنی صورت می‌گیرد.

با وجود اینکه برای هر یک از ما تعریف خاصی از احتمال ممکن است مهم باشد که من یا شما ترجیح می‌دهیم، هر کدام از آنها در باره فرضیه آزمایی اساساً به نتیجه یکسانی می‌انجامد، که در فصل ۱۱ مورد بحث قرار می‌گیرد. (گرچه آنهایی که متمایل به احتمالات ذهنی هستند ممکن است موافق رویکرد کلی فرضیه‌آزمایی نباشند). در دنیای واقعی اغلب مردم رویکردهای تعریفی را به صورت قابل جایگزین با هم به کار می‌گیرند. وقتی می‌گوییم که احتمال رخدادی $1/6$ است به این واقعیت اشاره داریم که آن رخداد ممکن است در یک مورد از شش مورد روی دهد. وقتی یک اتومبیل خاص را بر اساس گزارش مشتریان که گویای وضعیت فنی خوب آن است خریداری می‌کنیم، ما به این واقعیت پاسخ داده‌ایم که بخش زیادی از این اتومبیل‌ها بی‌عیب بوده‌اند. وقتی می‌گوییم احتمال برنده شدن یانکی‌ها زیاد است، ما باور ذهنی خود را در باره احتمال آن پدیده بیان می‌کنیم. اما وقتی برخی فرضیه‌ها را به این دلیل رد می‌کنیم که اگر آن فرضیه درست باشد احتمال خیلی کمی وجود دارد که آن داده‌ها بدست آیند، مهم نیست کدام دیدگاه احتمال را در نظر داریم.

۸-۲ واژه‌شناسی و قواعد اساسی

جزء اصلی داده‌های مربوط به نظریه‌پردازی احتمالات را **رخداد**^۱ می‌نامند. واژه رخداد عبارتی است که آماردانان تقریباً برای پوشش هر چیزی به کار می‌برند. یک رخداد می‌تواند بروز یک شاه در بازی کارتها، یک نمره از ۳۶ روی یک مقیاس دوست داشته شدن، یک طبقه‌بندی به عنوان "زن" در مورد شخص بعدی رئیس داده‌گاه عالی، یا میانگین یک نمونه باشد. هرگاه شما از احتمال چیزی صحبت می‌کنید، این چیز را رخداد گویند. وقتی ما با چیزی به سادگی انداختن یک سکه سروکار داریم- آمدن شیر یا خط، یک رخداد است. وقتی ما کارامل را از جعبه بیرون می‌کشیم، رخداد ممکن تلخ و شیرین است؛ و وقتی از یک نمره در یک درس صحبت می‌کنیم رخدادهای ممکن، الف، ب، ج و د [در نظام نمره‌گذاری با حروف] هستند.

دو رخداد را **مستقل** گویند در صورتی که وقوع یا عدم وقوع یک رخداد اثری بر وقوع یا عدم وقوع دیگری نداشته باشد. بنابراین رفتار رأی‌دهی دو آزمودنی که بطور تصادفی انتخاب شده باشند، مستقل از هم فرض خواهند شد، بخصوص در شرایط وجود پرده خصوصی، چون چگونگی رأی یک

1. event

شخص بر دیگری تأثیری نخواهد داشت. البته رفتار رأی‌دهی دو عضو یک خانواده احتمالاً رخدادهای مستقل نخواهند بود، زیرا این افراد باورها و نگرش‌های زیادی را به اشتراک می‌گذارند.

دو رخداد را **جمع‌ناپذیر** گویند در صورتی که وقوع یکی از رخدادها از وقوع دیگری جلوگیری کند. بنابراین کلاس‌های پیش‌نیاز دانشگاهی سال اول، و کلاس‌های تخصصی بعدی جمع‌ناپذیر هستند چون یک شخص نمی‌تواند همزمان در بیش از یکی از این کلاس‌ها نام‌نویسی کند. در نهایت، یک مجموعه از رخدادها را **انحصاری**^۱ گویند در صورتی که تمام پیامدهای ممکن را شامل شود. بنابراین چنین کلاس‌هایی از نظر مقررات آموزشی انحصاری هستند، که بایستی در یکی از این مقوله‌ها قرار گیرند- مگر با مجوز اداره آموزش. هم‌زمان این کلاس‌ها از جنبه ثبت نام کلی دانشگاه که دانشجویان تکمیلی، مجازی و غیره را در بر می‌گیرد، انحصاری نیستند.

همان‌گونه که قبلاً بیان شد دامنه احتمالات بین $0/00$ و $1/00$ متغیر است. اگر پدیده‌ای دارای احتمال وقوع $1/00$ باشد، پس مطمئناً اتفاق می‌افتد. اگر احتمال $0/00$ است مطمئناً اتفاق نمی‌افتد. هرچه احتمال به این حدود انتهایی نزدیک‌تر باشد، احتمال رخدادن یا رخ‌ندادن آن بیشتر است.

۲-۸-۱ قواعد اساسی احتمالات

برای هر نوع بحث در باره احتمال، دو قضیه خیلی مهم اساسی هستند. این قضایا اغلب به عنوان قواعد جمع و ضرب شناخته می‌شوند.

قاعده جمع برای نشان دادن قاعده جمع ما مثال قبلی را با خوردن چند کارامل شیرین و جایگزینی با مکعب‌ها چوبی سخت‌تر می‌کنیم. اکنون ما ۳۰ کارامل شیرین، ۱۵ کارامل تلخ و ۵۵ مکعب بدون طعم دیگر داریم. با در نظر گرفتن این فراوانی‌ها ما از تعریف تحلیلی احتمال می‌فهمیم که $0/30 = 30 / 100 = P(\text{شیرین})$ ، $0/15 = 15 / 100 = P(\text{تلخ})$ ، و $0/55 = 55 / 100 = P(\text{چوبی})$. اما احتمال اینکه یک کارامل را بیرون بکشیم که تلخ یا شیرین باشد، نه چوبی، چقدر است؟ در اینجا ما نیاز به **قانون جمع احتمال** داریم:

برای یک مجموعه پدیده‌های جمع‌ناپذیر، احتمال وقوع یک رخداد یا رخداد دیگر برابر است با حاصل جمع احتمالات جداگانه آنها.

بنابراین $P(\text{تلخ یا شیرین})$ برابر است با $P(\text{تلخ}) + P(\text{شیرین})$ ، $0/45 = 0/15 + 0/30 = P(\text{تلخ یا شیرین})$. توجه داشته باشید که ما این محدودیت را قائل شدیم که رخدادها بایستی جمع‌ناپذیر باشند، یعنی وقوع یک پدیده مانع از وقوع دیگری می‌شود. حدود نیمی از جمعیت یک کشور زنان هستند و حدود نیمی از جمعیت دارای اسامی زنانه هستند. احتمال اینکه یک شخصی که بطور تصادفی انتخاب

می‌شود زن یا دارای نام زنانه باشد آشکارا برابر $0/50 + 0/50 = 1/00$ نیست. این دو رخداد جمع‌ناپذیر نیستند. اما احتمال اینکه یک دختر که در سال ۱۹۸۱ در ورمونت بدنیا آمده نام سارا یا جسیکا داشته باشد برابر است با: $P = (13/1) + (5/1) = 18/2$ (جسیکا) + (سارا).

قاعده ضرب اکنون ما با جعبه کارامل ادامه می‌دهیم که در آن $P(\text{شیرین}) = 0/30$ و $P = 0/15$ (تلخ)، و احتمال یک مکعب چوبی $= 0/55$. فرض کنید بطور تصادفی دو کارامل بیرون بیاوریم، اولی را قبل از بیرون آوردن دومی جایگزین کنیم. احتمال اینکه در نوبت اول و در نوبت دوم یک کارامل شیرین بیرون بیاید چقدر است؟ در اینجا ما نیاز به استناد به **قانون ضرب احتمالات** داریم:

احتمال وقوع همزمان دو یا چند رخداد برابر است با حاصل ضرب احتمال تک تک آنها.

بنابراین $P(\text{شیرین}) \times P(\text{شیرین}) = 0/30 \times 0/30 = 0/09$. به همین ترتیب احتمال اینکه یک کارامل شیرین بدنال یک کارامل تلخ بیرون آید برابر است با $0/45 = 0/30 \times 0/15 = P(\text{تلخ}) \times P(\text{شیرین})$. توجه داشته باشید که ما خود را به رخدادهای مستقل محدود کردیم، یعنی وقوع یک رخداد اثری بر رخ دادن یا رخ ندادن دیگری ندارد. از آنجا که جنسیت و نام مستقل از هم نیستند نادرست است که بگوییم $0/25 = 0/50 \times 0/50$ (نام زنانه با زن بودن).

در آخر ما می‌توانیم مثالی را در نظر بگیریم که قوانین جمع و ضرب را نشان دهد. چه میزان احتمال وجود دارد که در دو کوشش (نمونه‌گیری با جایگزینی)، من یک کارامل شیرین و یک کارامل تلخ بیرون بیاورم، بدون توجه به ترتیب در آمدن آنها؟ ابتدا قاعده ضرب را به کار می‌گیریم تا حساب کنیم

$$P(\text{تلخ، شیرین}) = 0/30 \times 0/15 = 0/045$$

$$P(\text{شیرین، تلخ}) = 0/15 \times 0/30 = 0/045$$

چون این دو نتیجه نیاز ما را برآورده می‌کند (و چون آنها تنها رخدادهایی هستند که صورت می‌گیرند)، ما نیاز داریم که بدانیم احتمال اینکه یکی از این نتایج یا نتیجه دیگر رخ دهد چقدر است. در اینجا قاعده جمع را به کار می‌گیریم:

$$P(\text{تلخ، شیرین}) + P(\text{شیرین، تلخ}) = 0/045 + 0/045 = 0/09$$

بنابراین احتمال بدست آوردن یک کارامل از هر نوع در دو بار تلاش برابر $0/09$ است- یعنی، این رخداد در چیزی کمتر از یک-دهم بار اتفاق می‌افتد.

۲-۸-۲ احتمالات زوجی و شرطی

لازم است دو نوع احتمالاتی که در مبحث احتمال نقش مهمی دارند را تعریف کنیم. اینها شامل احتمالات زوجی و شرطی می‌باشند.

یک احتمال زوجی^۱ بسادگی به عنوان احتمال بروز همزمان دو یا چند رخداد تعریف می‌شود. برای مثال، احتمال اینکه قد دو نفر هر دو بیش از ۱۸۰ سانتیمتر باشد یک احتمال زوجی است. به همین ترتیب، احتمال اینکه شش نفر بعدی که وارد یک داروخانه می‌شوند مرد باشند نیز یک احتمال زوجی است. با در نظر گرفتن دو رخداد A و B ، احتمال زوجی آنها به صورت $P(A, B)$ نشان داده می‌شود. اگر این رخدادهای مستقل باشند، آنگاه احتمال وقوع زوجی آنها را می‌توان با استفاده از قاعده ضرب بدست آورد، چنانکه برای (تلخ، شیرین) P دیدیم.

یک احتمال شرطی احتمال وقوع یک رخداد در حالی است که برخی رخدادهای دیگر رخ دهند. احتمال این که برای فردی که در حال مستی رانندگی می‌کند تصادفی پیش بیاید مثالی از احتمال شرطی است. با در نظر گرفتن دو رخداد A و B ، احتمال شرطی A با وجود B به صورت $P(A|B)$ نشان داده می‌شود. فرض کنید یک ایستگاه رادیویی یک نمونه ۱۰۰ نفری گرفته که ۲۰ نفر کودک هستند. آنها دریافتند که ۳۰ نفر از نمونه از کمربند ایمنی اتومبیل استفاده می‌کنند، اما ۱۵ نفر آنان کودک هستند. پس احتمال اینکه یک شخص که بطور تصادفی نمونه‌گیری شده از کمربند استفاده کند $0/30 = 100/30 = 30/100$ است، و احتمال شرطی استفاده از کمربند با این شرط که کودک باشد $0/75 = 15/20$ است.

نباید احتمالات زوجی و شرطی را اشتباه کرد. احتمال اینکه شما یک بچه سه ساله و یک بچه شش ساله داشته باشید یک احتمال زوجی است. از سوی دیگر احتمال اینکه شما یک بچه شش ساله مشروط به داشتن یک بچه سه ساله باشید یک احتمال شرطی است. این احتمال شرطی بالاتر از احتمال زوجی است، بخشی به این خاطر که اگر شما یک بچه سه ساله داشته باشید، شما حداقل در گروه سنی که یک بچه شش ساله هم دارند قرار دارید. به عنوان مثالی دیگر، احتمال اینکه کسی مشروب بخورد و در شب دچار تصادف شود یک احتمال زوجی است. این احتمال خیلی بالا نیست در هر زمان زیرا افراد نسبتاً کمی مشروب می‌خورند و افراد کمی دچار تصادف می‌شوند. به هر حال، احتمال اینکه فردی مشروب خورده باشد مشروط به اینکه تصادف کرده باشد یک احتمال شرطی است. در شب این احتمال شرطی به $0/50$ نزدیک می‌شود زیرا در آمریکا نیمی از تصادفات شبانه با مصرف الکل درگیر است. توجه داشته باشید که هر دو (مشروب تصادف) P و (تصادف مشروب) P احتمالات شرطی درگیر با الکل و تصادفات بزرگراه هستند. اما این دو احتمال در حد برابر نیستند.

۳-۸ متغیرهای گسسته در مقابل پیوسته

در فصل ۲ بین متغیرهای گسسته و پیوسته فرق گذاشته شد. از دیدگاه ریاضی دانان یک متغیر گسسته متغیری است که می‌تواند شمار معینی از ارزش‌های متفاوت را به خود بگیرد، درحالی‌که یک متغیر

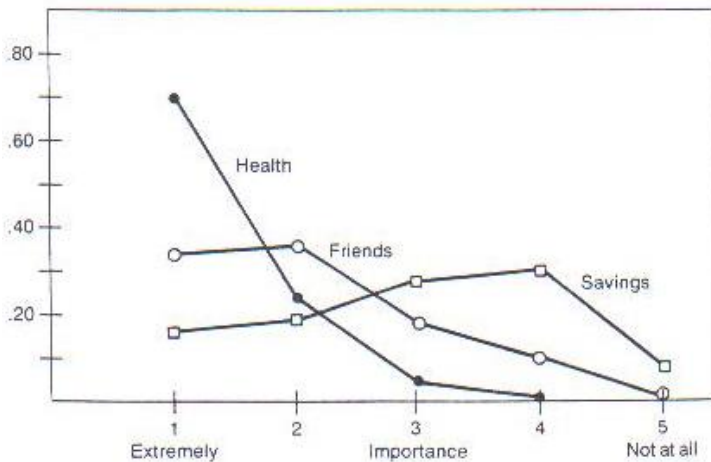
پیوسته متغیری است که می‌توان تعداد بی‌شماری از ارزش‌های مختلف به آن نسبت داد. برای مثال، تعداد افراد علاقه‌مند به یک تئاتر در هر یک از ۵۲ هفته آینده یک متغیر گسسته است زیرا تعداد افراد علاقه‌مند به تئاتر قابل شمارش است و چیزی به عنوان شخص کسری وجود ندارد. اما فاصله بین دو نفر در یک تحقیق پیرامون فضای شخصی، یک متغیر پیوسته است چون این فاصله می‌تواند ۱۰ سانتیمتر، ۱۲/۵ سانتیمتر یا ۸/۲۵۴۸ سانتیمتر باشد. هرچند این تمایز به لحاظ فنی درست است، کاربرد معمول آن قدری متفاوت است.

در عمل وقتی ما از یک متغیر گسسته حرف می‌زنیم، منظور ما معمولاً متغیری است که ارزش‌های محدود نسبتاً کمی به خود می‌گیرد (مثل تعداد سال‌های تحصیل). متغیری که یکی از ارزش‌های متعدد ممکن را می‌تواند بگیرد معمولاً به عنوان یک متغیر پیوسته به حساب می‌آید. بر این اساس ما معمولاً یک نمره IQ را به عنوان متغیر پیوسته در نظر می‌گیریم هر چند می‌دانیم که این نمرات به شکل واحدهای کامل هستند و کسی را که دارای نمره ۱۰۵/۳۱۷ باشد پیدا نمی‌کنید.

در اینجا فرق متغیرهای پیوسته و گسسته بازگویی شد چون طرز عمل توزیع‌های این دو نوع متغیر در نظریه احتمالات قدری فرق دارد. در باره متغیرهای گسسته می‌توان از احتمال یک پیامد خاص صحبت کرد. از سوی دیگر در باره متغیرهای پیوسته صحبت از احتمال بدست آوردن یک مقدار در یک فاصله معین می‌کنیم.

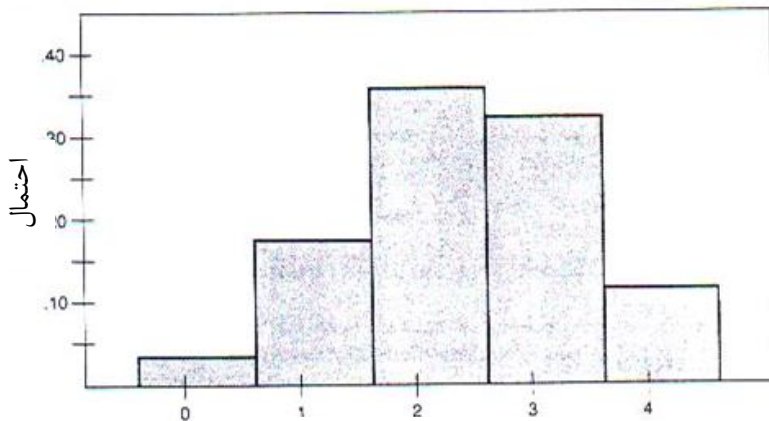
۴-۸ احتمال برای متغیرهای گسسته

مثال جالبی از یک متغیر گسسته در شکل ۸-۱ نشان داده شده است. داده‌های به نمایش درآمده در این شکل از مطالعه کمبل، کانورزس و راجرز (۱۹۷۶) گرفته شده که در آن آنها از ۲۱۶۴ پاسخگو تقاضا کردند تا اهمیتی که به جنبه‌های مختلف زندگیشان می‌دهند را روی یک مقیاس ۵ - ۱ درجه‌ای نشان دهند (۱ = بسیار مهم، ۵ = بی‌اهمیت). شکل ۸-۱ توزیع پاسخ‌ها برای چند جنبه را نشان می‌دهد. توجه دارید که ارزش‌های ممکن X (رتبه‌بندی) روی محور افقی و فراوانی‌های نسبی (احتمال) انتخاب آن پاسخ توسط افراد روی محور عمودی نشان داده شده است. از این شکل مشاهده می‌شود که توزیع پاسخ‌ها به سؤال‌های مربوط به سلامت، دوستان، و پس‌اندازها کاملاً متفاوت است. احتمال اینکه یک شخص که بطور تصادفی انتخاب می‌شود به سلامتی خود در حد بسیار مهم توجه کند ۰/۷۰ است، در حالی که احتمال اینکه همین شخص به حساب بانکی توجه کند ۰/۱۶ است.



شکل ۸-۱ توزیع‌های اهمیت رتبه بندی سه جنبه زندگی

به عنوان مثالی دیگر، به چراغ‌های راهنما توجه کنید. من در مسیر خود از منزل تا محل کار هر روز از چهار چراغ راهنمای طولانی و بنابراین کم و بیش جدا از هم، عبور می‌کنم. در نظر گرفتم که چند بار در پشت هر چراغ توقف می‌کنم. گاهی مجبورم در هر چهار مورد توقف کنم؛ معمولاً در یک مورد توقف می‌کنم، و یک‌بار بعد از مدت‌ها پشت هیچ چراغی مجبور به توقف نشدم. داده‌های این مثال در شکل ۲-۸ به نمایش در آمده است، که در آن نیز مقادیر احتمال به صورت فراوانی نسبی حساب شده است. از این شکل می‌توان مشاهده کرد که احتمال مجبور نشدن به توقف در پشت هیچ یک از چراغ‌های راهنما برای رسیدن من در فردا به اداره ۰/۰۳ است، در حالی که احتمال مجبور شدن به توقف در هر چهار مورد ۰/۱۱ است. محتمل‌ترین رخداد مجبور شدن به توقف در دو چراغ است که احتمال آن ۰/۳۶ است. این توزیع متقارن است زیرا در هر چرخه چراغ‌های قرمز یا سبز برای مدت طولانی‌تری روشن هستند تا چراغ زرد، و من هرگز در طول یک چراغ زرد شتاب نمی‌گیرم - خیلی به ندرت.



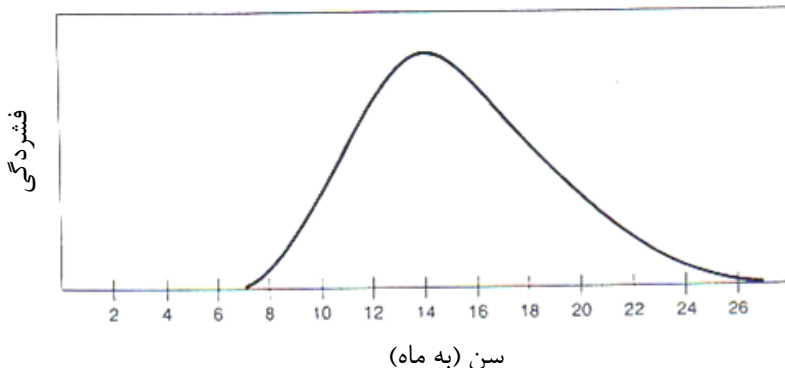
شکل ۸-۲ توزیع احتمال برای تعداد چراغ‌های راهنمای قرمز و زرد. احتمالات منفرد برای ۰-۴ چراغ به ترتیب عبارتند از ۰/۰۳۱۱، ۰/۱۷۱۹، ۰/۳۵۶۰، ۰/۳۲۷۸ و ۰/۱۱۳۲

۸-۵ توزیع‌های احتمال برای متغیرهای پیوسته

زمانی که ما از توزیع‌های احتمال گسسته به سوی پیوسته حرکت می‌کنیم به تدریج پیچیدگی‌ها بیشتر می‌شود. در فصل ۷ وقتی که به توزیع نرمال توجه داشتیم سروکار ما با یک توزیع پیوسته بود. به خاطر می‌آورید که در آن فصل ما محور عمودی توزیع را “فشردگی” نامیدیم. همچنین ما صحبت از فواصل کردیم نه پیامدهای خاص. اکنون لازم است قدری این موارد را بیشتر توضیح دهیم.

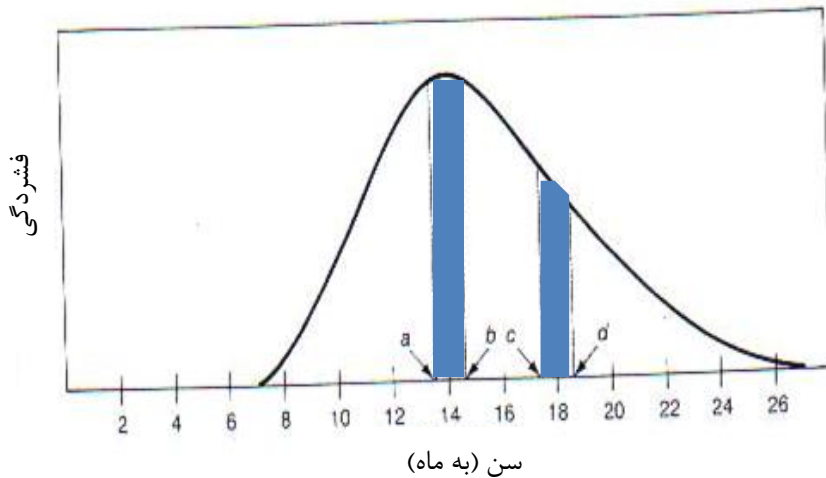
در شکل ۸-۳ توزیع تقریبی سنی که کودکان در آن سن برای اولین بار راه رفتن را یاد می‌گیرند آمده است. (شکل ۸-۳ براساس داده‌های هیندلی و همکاران، ۱۹۶۶ است.) میانگین تقریباً ۱۴ ماه است، انحراف معیار حدود سه ماه و توزیع دارای کجی (چولگی) مثبت است. توجه دارید که در این شکل برچسب محور عمودی “فشردگی” است. در حالی که در شکل ۸-۱ و ۸-۲ “احتمال” بود. فشردگی هم‌معنی احتمال نیست، و تنها ارتفاع منحنی را برای مقادیر مختلف X نشان می‌دهد. دلیل تغییر این برچسب این است که ما اکنون با یک توزیع پیوسته سروکار داریم نه گسسته. اگر دقت کنید متوجه خواهید شد که هر چند بالاترین نقطه ارتفاع منحنی در ۱۴ ماهگی است، احتمال اینکه یک کودک که بطور تصادفی انتخاب می‌شود دقیقاً در ۱۴ ماهگی (مثلاً ۱۴/۰۰۰۰۰۰۰۰ ماه) راه رفتن را آغاز کند بی‌نهایت کوچک است- به لحاظ آماری گفته می‌شود که عملاً ۰ است. به همین ترتیب احتمال آغاز به راه رفتن در ۱۴/۰۰۰۰۰۰۰۰۱ ماهگی نیز بی‌نهایت کوچک است. این نشان می‌دهد که نمی‌شود از احتمال هیچ پیامد خاصی سخن گفت. از سوی دیگر می‌دانیم که کودکان زیادی راه رفتن را تقریباً در ۱۴ ماهگی شروع می‌کنند، و این موضوع می‌رساند که می‌توان از احتمال بدست آوردن نمره‌ای که در

فاصله خاصی قرار داشته باشد صحبت کرد. مثلاً ممکن است بخواهیم احتمال راه رفتن یک نوزاد در ۱۴ ماهگی علاوه منهای نیم ماه، را بدانیم. چنین فاصله‌ای در شکل ۴-۸ نشان داده شده است. اگر ما بطور قراردادی سطح زیر منحنی را $1/00$ تعریف کنید، آنگاه سطح پر رنگ شده بین نقاط a و b در شکل ۴-۸ برابر با احتمال به راه افتادن یک نوزاد بطور تصادفی انتخاب شده، در این زمان است. مبنای محاسباتی این سطوح از طریق حساب انتگرال قابل توضیح است اما بنای ما در اینجا این محاسبات نیست. در محدوده این کتاب جداول مربوط به توزیع‌ها کفایت می‌کند همچنان که در فصل ۷ با جدول توزیع نرمال آشنا شدیم.



شکل ۳-۸ سنی که کودکان در آن سن اولین قدم‌ها را بدون کمک بر می‌دارند

ما توجه خود را معطوف به سطح بین a و b در شکل ۴-۸ می‌نماییم که روی میانگین تمرکز دارد. در باره هر فاصله‌ای همین موضوع صادق است. در شکل ۴-۸ می‌توان سطحی را مشاهده کرد که مربوط به دوره‌ای است که نیم ماه در هر سوی ۱۸ ماهگی گسترده است (ناحیه پر رنگ بین c و d). هر چند اطلاعات کافی در این مثال برای محاسبه احتمالات واقعی وجود ندارد، با دقت در شکل ۴-۸ می‌توان دریافت که فاصله یک ماهه اطراف ۱۴ ماهگی دارای احتمال بیشتری (سطح گسترده‌تر پر رنگ) نسبت به فاصله یک ماهه اطراف ۱۸ ماهگی است.



شکل ۴-۸ احتمال راه افتادن اولیه کودک در فواصل یک هفته‌ای با مرکزیت ۱۴ و ۱۸ ماهگی

یک روش خوب برای درک سطوح زیر منحنی این است که تکه‌ای کاغذ رسم شفاف برداشته و روی شکل قرار دهید (یا یک ورق کاغذ شطرنجی معمولی را از دو لبه بالا جلو نور بگیرید). آنگاه اگر تعداد مربع‌هایی را که در هر فاصله خاص قرار می‌گیرند بشمارید و بر تعداد کل مربع‌های زیر سطح کل منحنی تقسیم کنید، احتمال قرار گرفتن یک نمره تصادفی انتخاب شده در آن فاصله را می‌توانید برآورد کنید. روشن است که هر چه مربع‌های کاغذ شطرنجی کوچک‌تر باشند برآورد دقیق‌تر خواهد بود.

۸-۶ خلاصه

در این فصل ما به بررسی تعریف‌های مختلف آنچه که به عنوان احتمال از آن یاد شد پرداختیم و در این باره همچنین به تفاوت بین متغیرهای گسسته و پیوسته، و توزیع آنها توجه نمودیم. برخی از مهم‌ترین واژه‌های این فصل عبارتند از:

- دیدگاه تحلیلی
- دیدگاه فراوانی نسبی
- نمونه‌گیری با جایگزینی
- احتمال ذهنی
- رخداد
- رخداد‌های مستقل
- جمع‌ناپذیر
- انحصاری
- قانون جمع احتمالات
- قانون ضرب احتمالات
- احتمال جفتی
- احتمال شرطی
- فشردگی

۷-۸ تمرین‌ها

- ۷-۱ یک مثال از دیدگاه تحلیلی، فراوانی نسبی، و دیدگاه ذهنی بیان کنید.
- ۸-۲ فرض کنید که شما یک بلیط قرعه‌کشی خریده‌اید و برادرتان دو بلیط. بعد می‌خوانید که ۱۰۰۰ بلیط به فروش رسیده است.
- (الف) احتمال اینکه شما برنده شوید چقدر است؟
- (ب) احتمال برنده شدن برادرتان چقدر است؟
- (ج) احتمال اینکه شما یا برادرتان برنده شوید چقدر است؟
- ۸-۳ اکنون فرض کنید فقط ۱۰ بلیط فروش رفته و دو جایزه وجود دارد.
- (الف) به فرض اینکه شما جایزه اول را نبرده‌اید، احتمال اینکه جایزه دوم را ببرید چقدر است؟ (جایزه اول به صندوق باز نمی‌گردد).
- (ب) احتمال اینکه برادر شما در مرحله اول و شما در مرحله دوم برنده شوید چقدر است؟
- (ج) چقدر احتمال دارد که شما در مرحله اول و برادرتان در مرحله دوم برنده شود؟
- (د) چقدر احتمال دارد که هر دوی شما در رخداد اول و دوم قرار گیرید؟
- ۸-۴ کدام بخش از تمرین ۸-۳ با احتمالات جفتی سروکار دارد؟
- ۸-۵ کدام بخش از تمرین ۸-۳ با احتمالات شرطی سروکار دارد؟
- ۸-۶ مثال ساده‌ای از یک موقعیت مربوط به احتمالات جفتی که مورد علاقه شما باشد بیان کنید.
- ۸-۷ مثال ساده‌ای از یک موقعیت مربوط به احتمالات شرطی که مورد علاقه شما باشد بیان کنید.
- ۸-۸ به نظر می‌رسد در برخی خانه‌ها رفتار مادر مستقل از رفتار بچه‌اش است و برعکس. اگر مادر هر روز در کل ۲ ساعت به بچه‌اش نگاه کند و اگر بچه ۳ ساعت در روز به مادرش نگاه کند، در صورتی که آنها واقعا بطور مستقل این رفتار را انجام دهند چقدر احتمال دارد که آنها همزمان به یکدیگر نگاه کنند؟
- ۸-۹ در تمرین ۸-۸ فرض کنید که مادر و کودک هر دو از ساعت ۸ شب تا ۷ صبح بخوابند. در این صورت احتمال فوق چگونه خواهد بود؟
- ۸-۱۰ با استفاده از داده‌های شکل ۸-۲ احتمال اینکه فرد مجبور به توقف در پشت حداقل سه چراغ راهنما شود چقدر است؟
- ۸-۱۱ مثالی از یک توزیع پیوسته عمومی ارائه کنید که در آن ما به احتمال رخداد یک مشاهده در فاصله معینی علاقه داریم.
- ۸-۱۲ مثالی از یک متغیر پیوسته بنزید که بطور معمول آن را گسسته به حساب می‌آوریم.
- ۸-۱۳ دو نمونه از متغیرهای گسسته بیان کنید.
- ۸-۱۴ یک کمیته پذیرش دانشجویان ارشد در نهایت به این واقعیت پی برد که نمی‌تواند فرق اساسی بین داوطلبان سطح بالا در نظر بگیرد. امسال این کمیته همه ۱۰۰۰ داوطلب را رتبه‌بندی نمود و به طور

تصادفی ۱۰ نفر از کسانی که در رتبه درصدی ۸۰ یا بالاتر بودند را انتخاب کرد. احتمال اینکه هر داوطلب خاص پذیرفته شود چقدر است (فرض کنید که شما از رتبه‌بندی او اطلاعی ندارید).

۸-۱۵ با توجه به سؤال ۱۴-۸ احتمال شرطی اینکه شخص مورد پذیرش (الف) دارای بالاترین رتبه باشد و (ب) دارای پایین‌ترین رتبه باشد چقدر است؟

فصل ۹

همبستگی

- ۹-۱- نمودارهای پراکندگی
- ۹-۲- مثال - رابطه بین سرعت و دقت
- ۹-۳- کوواریانس
- ۹-۴- ضریب همبستگی گشتاوری پیرسون (r)
- ۹-۵- همبستگی با داده‌های رتبه‌ای
- ۹-۶- عواملی که بر همبستگی اثر می‌گذراند
- ۹-۷- احتیاط
- ۹-۸- ضرایب همبستگی دیگر
- ۹-۹- مثال کار شده پایانی

فصل‌های پیش به شکلی با داده‌های توصیفی مربوط به یک متغیر وابسته منفرد سر و کار داشتند. ما به بحث در باره یک متغیر پرداختیم، چگونگی بدست آوردن میانگین و انحراف معیار آن و غیره. اما بعضی تحقیقات به جای یک متغیر وابسته، بر دو متغیر یا بیشتر متمرکز هستند. در این موارد ما اغلب مایل هستیم رابطه بین دو متغیر را بدانیم نه اینکه بدانیم هر کدام به تنهایی چگونه هستند. برای نشان دادن نوع مطالعاتی که ممکن است با دو متغیر (که با X و Y نشان داده می‌شوند) سروکار داشته باشند به سؤال‌های تحقیقاتی زیر توجه کنید:

۱. آیا توانایی رانندگی (Y) به میزان الکل مصرف شده (X) بستگی دارد؟
۲. آیا پذیرش در دانشگاه (Y) با میزان فعالیت‌های فوق برنامه (X) که دانش‌آموز در دوره دبیرستان انجام داده است رابطه دارد؟

۳. آیا رتبه دوست داشتنی بودن (Y) با جذابیت فیزیکی (X) ارتباطی دارد؟

۴. آیا رفتار احتکارگری (Y) تابعی از میزان محرومیت کشیدن (X) است؟

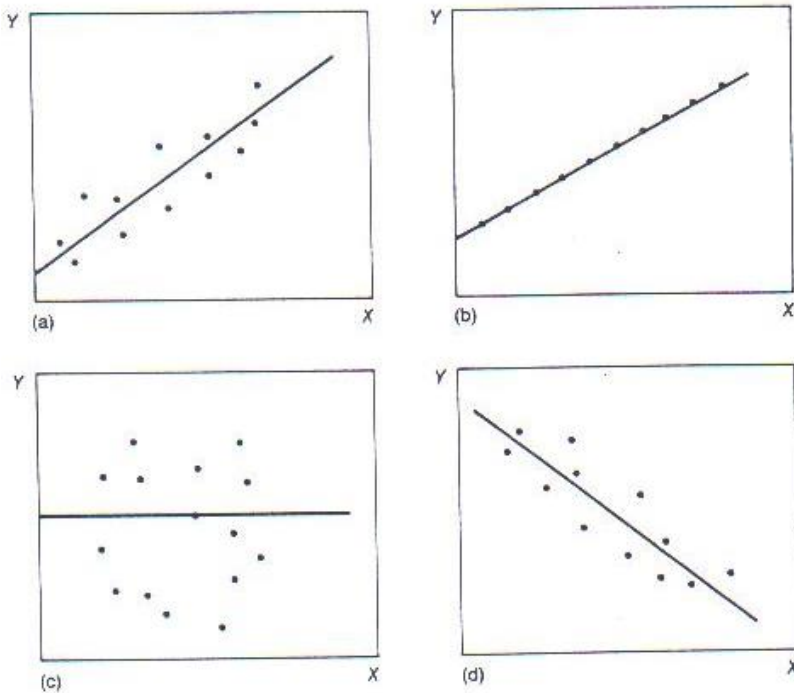
۵. آیا درستی عملکرد (Y) با افزایش سرعت (X) کاهش پیدا می‌کند؟

در هر کدام از این سؤال‌ها ما می‌پرسیم آیا یک متغیر (Y) با متغیر دیگر (X) در رابطه است. وقتی ما با رابطه بین دو متغیر سروکار داریم به همبستگی توجه داریم، و اندازه‌ی میزان یا شدت این رابطه با ضریب همبستگی^۱ نشان داده می‌شود. در حقیقت شماری از ضرایب همبستگی وجود دارند، در اصل بستگی به ماهیت اندازه‌گیری‌ها دارد، اما بعداً خواهیم دید که در موارد بسیاری فرق این ضرایب بیشتر ظاهری است تا حقیقی. فعلاً ما به معمول‌ترین ضریب همبستگی یعنی ضریب همبستگی گشتاوری پیرسون (r) می‌پردازیم.

۹-۱ نمودارهای پراکنش

وقتی اندازه‌هایی را در باره دو متغیر برای بررسی رابطه بین آنها جمع‌آوری می‌کنیم، یکی از سودمندترین شیوه‌های بدست آوردن بینشی از این رابطه تهیه یک نمودار پراکنش^۲ است (طرح پراکندگی^۳ یا پراکنش نگار^۴ هم گفته می‌شود). نمونه‌هایی از چهار نمودار در شکل ۹-۱ آمده است. در این شکل هر آزمودنی با یک نقطه در فضای دو بعدی نشان داده شد، مختصات این نقطه (Y_i, X_i) نمرات فرد (یا آزمودنی) در متغیرهای X و Y است. برای تهیه‌ی یک نمودار پراکنش معمولاً متغیر پیش‌بین^۵ روی محور افقی یا محور X ، و متغیر ملاک^۶ روی محور عمودی یا محور Y نشان داده می‌شود. اگر هدف نهایی تحقیق پیش‌بینی یک متغیر از روی اطلاعات مربوط به متغیر دیگر باشد، تشخیص ساده است چون متغیر ملاک متغیری است که پیش‌بینی می‌شود، در حالی که پیش‌بینی کننده آن متغیری است که از روی آن پیش‌بینی صورت می‌گیرد. اگر مسأله از نوع بدست آوردن ضریب همبستگی باشد، تشخیص ممکن است روشن باشد (شیوع سرطان بستگی به مقدار مصرف سیگار دارد نه برعکس، و بنابراین شیوع روی محور افقی نمایش داده می‌شود) یا روشن نباشد (نه سرعت و نه تعداد گزینه‌های درست-متغیرهای معمول در مطالعه یادگیری-آشکارا در موقعیت وابسته نسبت به دیگری قرار ندارند). جایی که تشخیص روشن نباشد، نمی‌توان گفت کدام متغیر برچسب X و کدام Y بگیرد.

-
1. correlation coefficient
 2. scatter diagram
 3. scatter plot
 4. scattergram
 5. predictor variable
 6. criterion variable



شکل ۹-۱ نمودارهای پراکنش نشان دهنده درجات مختلف رابطه

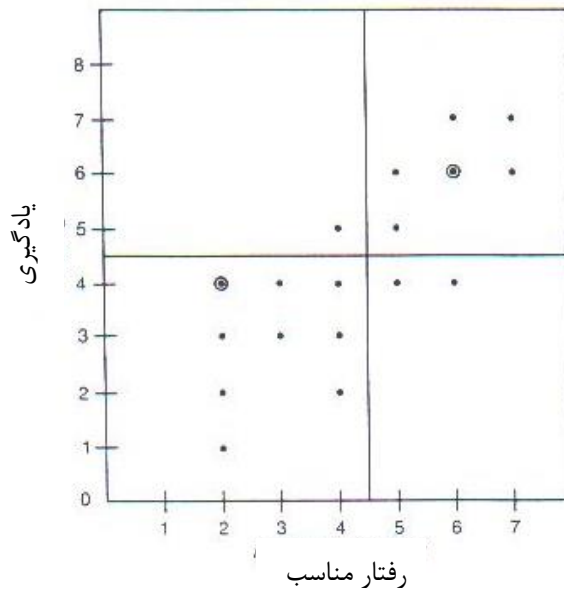
به چهار نمودار پراکنش فرضی شکل ۹-۱ توجه کنید. شکل ۹-۱ (الف) حالتی را نشان می‌دهد که رابطه نسبتاً قوی بین X و Y وجود دارد. هر چند این رابطه کامل نیست، اما این موضوع بطور کلی درست است که هر چه X افزایش پیدا می‌کند Y نیز افزایش می‌یابد. شکل ۹-۱ (ب) حالتی از یک **رابطه کامل** را نشان می‌دهد. هر افزایش در X همراه است با نسبت دقیقی از افزایش در Y ، طوری که تمام نقاط روی یک خط مستقیم واقع شده‌اند. شکل ۹-۱ (ج) حالتی را نشان می‌دهد که بین X و Y **رابطه‌ای وجود ندارد**. هیچ گرایش منظمی برای Y که با X تغییر کند وجود ندارد، و آگاهی از مقدار X چیزی در باره مقدار Y به ما نمی‌گوید. در آخر، شکل ۹-۱ (د) یک **رابطه منفی** نسبتاً قوی را بین X و Y نشان می‌دهد. در حقیقت شدت رابطه اما نه جهت آن، دقیقاً مطابق با شکل ۹-۱ (الف) است (فقط نقاط ۹۰ درجه چرخش یافته‌اند). در این حالت یک افزایش در X همراه است با یک کاهش کلی در Y .

خط‌های تعبیه شده در شکل‌های ۹-۱ (الف-د) خطوط مستقیمی را نشان می‌دهند که به بهترین حالت با داده‌ها تطبیق داده شده‌اند- روشی که این خط "بهترین" است در فصل بعد بیان خواهد شد. این خطوط اغلب در زمره یک طرح پراکنش قرار می‌گیرند چون به روشن شدن رابطه کمک می‌کنند.

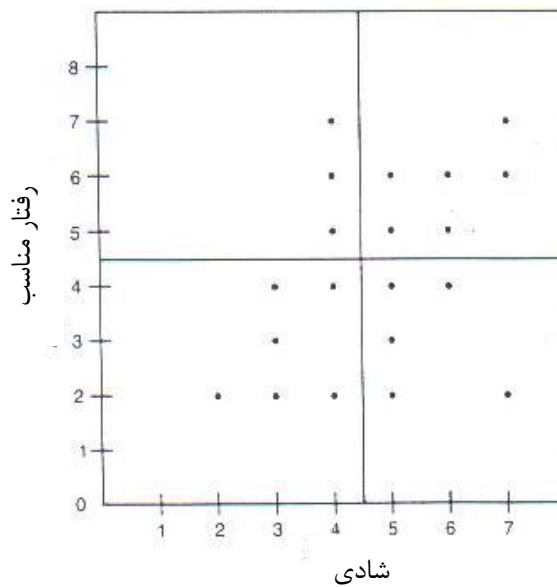
دامنه ضرایب همبستگی بین $+1/00$ و $-1/00$ است. در شکل ۹-۱ (الف) نقاط خیلی نزدیک به خط هستند و همبستگی بالاست ($r = 0/81$). در شکل ۹-۱ (ب) نقاط دقیقاً روی خط قرار گرفته‌اند و همبستگی کامل است ($r = 1/00$). در شکل ۹-۱ (ج) نقاط اصلاً در اطراف خط جمع نشده‌اند و رابطه + است. و در شکل ۹-۱ (د) میزان همبستگی مشابه شکل ۹-۱ (الف) است اما رابطه منفی است، بنابراین همبستگی منفی است ($r = -0/81$).

یک تفسیر متفاوت از نمودار پراکنش را می‌توان در شکل‌های ۹-۲ و ۹-۳ دید. این شکل‌ها بر اساس داده‌های فرم گزارش پیشرفت تحصیلی معلم بنا شده است، که یک فرم رتبه‌بندی از مشکلات رفتاری شبیه به فرم خودگزارشی جوانان اکناخ است، که قبلاً دیدیم. اما این فرم توسط معلم تکمیل می‌شود نه کودک. (داده‌ها در رابطه با این مثال قدری تغییر داده شده‌اند). در شکل ۹-۲ رابطه بین رتبه‌بندی معلم از میزان رفتار مناسب بروز داده شده از سوی دانش‌آموز با شرایط و رتبه‌بندی معلم از میزان یادگیری دانش‌آموز به نمایش در آمده است. تمام رتبه‌بندی‌ها براساس مقیاس درجه‌ای است. نقاطی که دور آن دایره کشیده شده نشان دهنده نقاط مربوط به دو دانش‌آموز هستند. در شکل ۹-۳ رتبه‌بندی معلم از میزان شادی کودک و میزان رفتار مناسب او نشان داده شده است.

در این شکل‌ها خطوط عمودی و افقی تیره برای تقسیم محورها از میانگین آنها رسم شده است. برای مثال، خط عمودی در شکل ۹-۲ کودکانی که در رفتار مناسب از میانگین پایین‌تر هستند را از کسانی که بالاتر از میانگین هستند جدا می‌سازد. به همین ترتیب، خط افقی کسانی را که در متغیر یادگیری زیر میانگین هستند از کسانی که بالاتر از میانگین هستند جدا می‌نماید. در هر دو حالت متغیری که ما فکر می‌کنیم بیشتر احتمال دارد که متغیر وابسته باشد روی محور Y نمایش داده شده، و متغیر مستقل احتمالی روی محور X نشان داده شده است. در عین حال این انتخاب دلخواه است.



شکل ۹-۲ رابطه بین یادگیری و رفتار مناسب در کودکان عادی، سنین ۱۶-۱۲



شکل ۹-۳ رابطه بین شادی و رفتار مناسب در کودکان عادی، سنین ۱۶-۱۲

اگر مثلاً رابطه مثبت قوی بین یادگیری و رفتار مناسب وجود داشته باشد، ما انتظار خواهیم داشت بیشتر کودکانی که روی یک متغیر رتبه‌های بالا (بالتر از میانگین) دارند روی متغیر دیگر هم بالا باشند. به همین ترتیب، بیشتر کسانی که در یک متغیر زیر میانگین بودند باید در متغیر دیگر هم زیر آن باشند. چنین ایده‌ای را می‌توان با یک جدول ساده نشان داد که در آن تعداد کودکانی که در هر دو متغیر بالای میانگین و تعدادی که زیر میانگین هستند و همچنین تعدادی که در یک متغیر بالا و در دیگری زیر میانگین هستند را می‌شماریم. چنین جدولی در جدول ۹-۱ برای داده‌های شکل ۹-۲ و ۹-۳ نشان داده شده است.

با وجود رابطه مثبت قوی بین دو متغیر، انتظار خواهیم داشت بیشتر نقاط داده‌ها در جدول ۱-۹ در خانه‌های ”بالا-بالا“ و ”پایین-پایین“، با تنها تعداد کمی در خانه‌های ”پایین-بالا“ و ”بالا-پایین“ قرار بگیرند. از سوی دیگر اگر دو متغیر با هم رابطه‌ای نداشته باشند انتظار داریم تقریباً تعداد برابری از نقاط داده‌ها در چهار خانه جدول (یا هر ربع نمودار پراکنش) قرار گیرند. در جدول ۹-۱ مشاهده می‌شود که برای رابطه بین رفتار مناسب و یادگیری، ۱۷ نفر از ۲۰ کودک در خانه‌های متداعی با یک رابطه مثبت بین متغیرها قرار می‌گیرند. به عبارت دیگر، اگر آنها در یک متغیر زیر میانگین باشند در دیگری هم عموماً زیر میانگین هستند، و برعکس. تنها ۳ کودک تابع این الگو نیستند. اما برای داده‌های ترسیم شده در شکل ۹-۳، جدول ۹-۱ به ما نشان می‌دهد که تنها ۱۲ کودک در جهات یکسان میانگین هر دو متغیر قرار دارند، در حالی که ۵ کودک زیر میانگین رفتار مناسب اما بالای متغیر شادی قرار دارند، و ۳ کودک الگوی عکس آن را نشان می‌دهند.

جدول ۹-۱ بررسی نمودارهای پراکنش با تقسیم آنها به ربع‌ها

رفتار مناسب در مقابل یادگیری

(شکل ۹-۲)

رفتار مناسب		یادگیری	بالای میانگین
پایین میانگین	بالای میانگین		
۱	۷	بالای میانگین	
۱۰	۲	پایین میانگین	

شادی در برابر رفتار مناسب

(شکل ۹-۳)

شادی		رفتار مناسب	بالای میانگین
پایین میانگین	بالای میانگین		
۳	۶	بالای میانگین	
۶	۵	پایین میانگین	

این دو مثال سپس به شیوه‌ای ساده تفسیر نمودارهای پراکنش و روابط بین متغیرهای را نشان می‌دهند. در مورد اول یک رابطه قوی مثبت بین متغیرها وجود دارد. در مورد دوم، رابطه هر چند مثبت، اما خیلی ضعیف است. این نتیجه در همبستگی بین متغیرها بازتاب پیدا می‌کند. در شکل ۲-۹ همبستگی ۰/۷۸ است، در حالی که همبستگی بین نقاط در شکل ۳-۹ معادل ۰/۳۸ می‌باشد. (در نظر داشته باشید که ما برای سادگی مباحث از نمونه‌های کوچک استفاده کردیم، چنانچه نمونه‌ها بزرگتر باشند همبستگی‌ها فرق خواهند کرد.)

۹-۲ یک مثال - رابطه بین سرعت و دقت

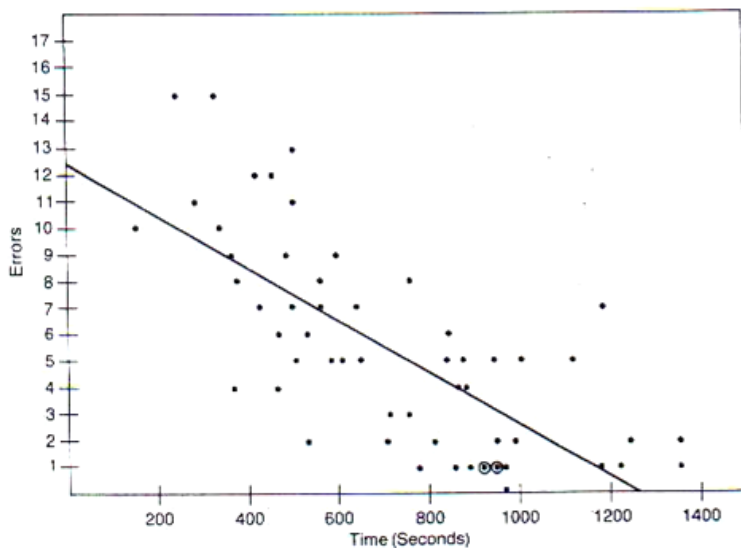
خیلی از روان‌شناسان شناختی روی حل مسأله تحقیق کرده‌اند و می‌گویند آزمودنی‌های وسواسی نسبت به آزمودنی‌های غیروسواسی از راهکارهای حل مسأله کم اثرتری استفاده می‌کنند. در یک مطالعه روی روابط بین وسواس و راهکار حل مسأله، کنر-مک‌دونالد^۱ (۱۹۸۴) به ارزیابی وسواس بر اساس تست تطبیق شکل‌های آشنای کاگان (MFFT) پرداخت. در این تکلیف به آزمودنی‌ها یک شکل ساده رسم الخطی (معیار) همراه با شش شکل مشابه نشان داده می‌شود. آزمودنی می‌بایست شکلی که دقیقاً با شکل اصلی تطبیق می‌کند را مشخص کند. (تا یافتن جواب درست پاسخگویی ادامه می‌یابد.) این تکلیف با ۱۲

شکل مختلف تکرار می‌شود. دو مورد از متغیرهای وابسته شامل زمان کل پاسخ اول (روی مجموع ۱۲ شکل) و نمره کل خطاها بودند. با ترسیم این دو متغیر در مقابل هم، امکان شناسایی گروه‌های آزمودنی دارای سبک‌های مختلف پاسخ فراهم می‌گردد. ما در اینجا برای مثال خود، داده‌های کتر-مک‌دونالد در این تکلیف را با اندکی تغییر در جدول ۹-۲ آورده و در شکل ۹-۴ نمایش داده‌ایم. در جدول ۹-۲ نیز میانگین و انحراف معیار هر متغیر همراه با بعضی محاسبات میانی که می‌تواند مفید باشد مشاهده می‌شود. از بررسی شکل ۹-۴ می‌توان یک رابطه قوی منفی بین زمان و خطاها مشاهده کرد- با افزایش زمان، خطاها کاهش می‌یابند و برعکس. این رابطه قدری **منحنی شکل**^۱ است (یعنی بهترین خط انطباق در حقیقت منحنی است)، اما باندازه کافی یک **رابطه خطی**^۲ بین دو متغیر وجود دارد طوری که یک خط مستقیم به خوبی با داده‌ها به عنوان اولین برآورد انطباق پیدا می‌کند. این خط تعبیه شده تا رابطه روشن‌تر شود. کسانی که نقاط داده‌های آنها در قسمت چپ بالایی قرار گرفته‌اند هستند که کاگان آنها را "وسواسی" می‌نامد- به طور وسواس گونه‌ای با مدت زمان کوتاه پاسخ و خطاهای زیاد عمل می‌کنند. (وسواسی‌ها همواره اطمینان دارند که "این بار درست عمل می‌کنم"، اما به نظر نمی‌رسد هیچگاه این روش جوابگو باشد.) در قسمت راست پایین نقاط داده‌ها مربوط است به "تفکری‌ها" این افراد وقت زیادی را صرف کار می‌کنند، اما در نهایت معمولاً درست عمل می‌کنند. دورتر از خط در سمت راست بالا دو یا سه نقطه داده مربوط به افرادی است که زمان زیادی را صرف می‌کنند تا تصمیم بگیرند. در نهایت در سمت چپ پایین تعداد کمی از افرادی قرار دارند که سریع و درست عمل می‌کنند و باعث می‌شوند بقیه ما احساس بدی داشته باشیم. نقاط دایره‌دار نشان‌دهنده آزمودنی‌های متعددی هستند که نمرات یکسانی در X و Y دارند.

1. curvilinear
2. linear relationship

جدول ۹-۲ داده‌های مربوط به خطاهای کل و زمان کل پاسخ (به ثانیه) برای تست تطبیق شکل‌های آشنا

Subject #	1	2	3	4	5	6	7	8	9
Time (X)	285	599	1001	324	595	363	361	870	531
Errors (Y)	11	9	5	15	5	9	4	4	2
Subject #	10	11	12	13	14	15	16	17	18
Time	526	749	852	514	856	467	449	949	929
Errors	6	3	5	7	1	6	12	2	1
Subject #	19	20	21	22	23	24	25	26	27
Time	776	348	507	640	474	497	953	575	1253
Errors	1	10	13	5	4	11	1	7	2
Subject #	28	29	30	31	32	33	34	35	36
Time	762	827	973	571	832	1352	813	603	866
Errors	8	5	0	8	6	1	2	5	4
Subject #	37	38	39	40	41	42	43	44	45
Time	1357	1220	635	1105	242	371	951	1183	1184
Errors	2	1	7	5	15	8	1	1	7
Subject #	46	47	48	49	50	51	52	53	54
Time	977	411	989	930	519	485	434	710	708
Errors	1	12	2	1	5	9	7	3	2
Subject #	55	56	57						
Time	941	170	889						
Errors	5	10	1						
$\Sigma X = 41,253$			$\Sigma Y = 305$		$\Sigma XY = 174,474$				
$\Sigma X^2 = 34,673,815$			$\Sigma Y^2 = 2483$		$N = 57$				
$\bar{X} = 723.737$			$\bar{Y} = 5.351$						
$s_x = 293.30$			$s_y = 3.90$						



شکل ۹-۴ نمودار پراکنش زمان (X) و خطاهای (Y)

۳-۹ کوواریانس

ضریب همبستگی که می‌خواهیم روی این داده‌ها محاسبه کنیم خود بر پایه یک آماره بنام **کوواریانس**^۱ است. کوواریانس اساساً یک عدد است که نشان‌دهنده درجه تغییر دو متغیر با یکدیگر است. اگر مثلاً نمرات بالا در یک متغیر متمایل به همراه بودن با نمرات بالا در دیگری باشد، کوواریانس بزرگ و مثبت خواهد بود. اگر همچون مثال ما، نمرات بالا در یک متغیر با نمرات پایین در دیگری همراه باشند، کوواریانس بزرگ اما منفی خواهد بود. و بالاخره وقتی نمرات بالا در یک متغیر تقریباً به نسبت یکسان با نمرات بالا و پایین در دیگری همراه باشند، کوواریانس نزدیک به صفر خواهد بود.

برای تعریف کوواریانس به صورت ریاضی می‌توان نوشت

$$COV_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1}$$

از این معادله روشن می‌گردد که کوواریانس به لحاظ شکلی خیلی شبیه به واریانس است. اگر ما در این معادله هر دو Y ها را به X تغییر دهیم، S_X^2 را خواهیم داشت و اگر X ها را به Y تغییر دهیم S_Y^2 را خواهیم داشت. همچنین روشن است که کوواریانس بر اساس این است که یک مشاهده چگونه از میانگین در هر متغیر انحراف دارد که در بخش ۱-۹ بیان گردید. بینش بهتری از معنای کوواریانس را از طریق توجه به آنچه از همبستگی منفی انتظار داریم می‌توان بدست آورد، همانند داده‌های جدول ۲-۹. در این مثال زمان‌های زیاد با خطاهای کوچک جفت خواهند شد. بنابراین برای یک آزمودنی کم سرعت، بادقت، $(X - \bar{X})$ مثبت، $(Y - \bar{Y})$ منفی و حاصل ضرب آنها منفی خواهد بود. در مورد یک آزمودنی سریع، غیردقیق هم این موضوع صادق است. مجموع حاصل ضرب $(X - \bar{X})(Y - \bar{Y})$ بزرگ و منفی، با یک کوواریانس منفی بزرگ خواهد بود.

سپس به یک رابطه مثبت قوی توجه کنید. در اینجا مقادیر مثبت بزرگ X و Y به احتمال زیاد با مقادیر مثبت بزرگ $(Y - \bar{Y})$ جفت خواهند بود، و برعکس. بنابراین مجموع حاصل ضرب انحرافات بزرگ و مثبت خواهد بود، که نشان از یک رابطه مثبت بزرگ دارد. چنین حالتی در شکل ۲-۹ دیده می‌شود.

در آخر، موقعیتی را در نظر بگیرید که بین X و Y رابطه‌ای وجود ندارد. در این حالت یک مقداری مثبت $(X - \bar{X})$ گاهی با یک مقدار مثبت $(Y - \bar{Y})$ و گاهی بایک مقدار منفی جفت خواهد شد. در نتیجه حاصل ضرب انحرافات در نیمی از موارد مثبت و در نیمی از موارد منفی خواهد بود که حاصل جمع نزدیک به صفر و نشان‌دهنده نبودن رابطه بین متغیرهاست. یک حالت کمتر انتهایی، در شکل ۳-۹ آمده است.

1. covariance [واریانس مشترک]

برای یک مجموعه داده‌های معین، می‌توان نشان داد که کوواریانس در بالاترین حد مثبت خود خواهد بود زمانی که X و Y به صورت مثبت با هم رابطه داشته باشند ($r = 1/00$) و در بالاترین حد منفی خواهد بود وقتی که این رابطه منفی باشد ($r = -1/00$). وقتی این دو متغیر کاملاً با هم بی‌ارتباط باشند ($r = 0$) کوواریانس صفر خواهد بود.

برای اهداف محاسباتی بیان ساده‌ای از کوواریانس به این صورت خواهد بود

$$cov_{XY} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{N-1} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N-1}$$

چنانکه از فصل ۲ بخاطر دارید، $\sum XY$ از طریق ضرب نمره X یک آزمودنی با نمره Y او و سپس جمع کردن نتایج برای کل آزمودنی‌ها حساب می‌شود. بنابراین برای داده‌های جدول ۹-۲، $\sum XY = (285)(11) + (9)(9) + (101)(5) + 000 + (889)(1)$ می‌شود^۱

$$\begin{aligned} cov_{XY} &= \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N-1} \\ &= \frac{174474 - \frac{(41253)(305)}{57}}{56} = -826/17 \end{aligned}$$

۹-۴ ضریب همبستگی گشتاوری پیرسون^۲ (r)

آنچه که پیش از این در باره کوواریانس گفتیم گویای آن است که ما می‌توانیم کوواریانس را به عنوان اندازه‌ای از درجه‌ی رابطه بین دو متغیر به کار ببریم. البته یک مشکلی که بلافاصله پیش می‌آید این است که قدر مطلق کوواریانس cov_{XY} همچنین تابعی از انحراف استانداردهای X و Y نیز می‌باشد. بنابراین، مثلاً $cov_{XY} = 20$ ، ممکن است بیانگر یک درجه همبستگی بالا باشد وقتی که انحراف استانداردها کوچک هستند، اما گویای یک درجه همبستگی پایین باشد وقتی که انحراف استانداردها بزرگ هستند. برای حل این مشکل ما کوواریانس را بر اندازه انحراف استانداردها تقسیم می‌کنیم و این را مبنای برآورد همبستگی قرار می‌دهیم. بنابراین همبستگی را چنین تعریف می‌کنیم

$$r = \frac{cov_{XY}}{S_X S_Y}$$

۱. [یادآوری: چنانکه قبلاً هم گفتیم، مجموع مضروب‌ات (مجموع حاصل ضرب‌های انحراف نمرات متغیرهای X و Y) که برای محاسبه کوواریانس و ضریب همبستگی به کار می‌رود و معادل مجموع مجذورات برای دو متغیر است از فرمول زیر بدست می‌آید:

$$\sum xy = \sum XY - \frac{\sum X \sum Y}{N}$$

از آنجا که بیشترین مقدار COV_{XY} به اندازه $\pm S_X S_Y$ می‌باشد، نتیجه می‌گیریم که حدود r برابر $\pm 1/00$ است. پس یک تفسیر r این است که اندازه‌ای از درجه نزدیکی کوواریانس به بیشترین اندازه‌ی خود است. یک روش مشابه برای نوشتن معادله قبلی این خواهد بود که واریانس‌ها و کوواریانس‌ها را با فرمول محاسباتی آنها جایگزین کنیم و سپس با حذف موارد اضافی آن را ساده کنیم. با این کار به فرمول زیر می‌رسیم

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

فرمول سومی هم برای r وجود دارد که به لحاظ جبری با دو مورد دیگر برابر است اما امتیاز آن واژه‌شناسی مشابه با مطالب فصل‌های بعد در باره رگرسیون و تحلیل واریانس (فصل‌های ۱۶ تا ۱۸) است. اگر ما مجموع مجذورات X^1 را به صورت $SS_X = \sum X^2 - (\sum X)^2 / N$ و مجموع مضروبات Y^2 و X را به صورت

$$\sum XY - \frac{\sum X \sum Y}{N} = \sum (X - \bar{X})(Y - \bar{Y}) = SP_{XY}$$

تعریف کنیم

آنگاه می‌توانیم r را به صورت زیر تعریف کنیم

$$r = \frac{SP_{XY}}{\sqrt{SS_X SS_Y}}$$

یک ویژگی جالب مجموع مجذورات و مضروبات این است که بطور مستقیم به واریانس و کوواریانس می‌انجامند. بنابراین

$$COV_{XY} = \frac{SP_{XY}}{N-1} \quad \text{و} \quad S_X^2 = \frac{SS_X}{N-1}$$

این سه معادله r دقیقاً جواب یکسانی را بدنبال دارند. انتخاب با شماست. اگر ماشین حساب شما به طور خودکار انحراف معیارها را حساب می‌کند یا اگر شما به هر دلیلی به آنها نیاز داشته باشید معادله اول بیشتر سودمند است. اگر ماشین حساب شما به طور خودکار انحراف معیارها را حساب نمی‌کند معادله دوم بهتر است. اگر شما با رگرسیون که در فصل بعد می‌آید کار دارید روشی که مجموع مجذورات و مجموع مضروبات را به کار می‌گیرد ترجیح دارد.

با به کارگیری معادله اول در داده‌های جدول ۲-۹، داریم

$$r = \frac{COV_{XY}}{S_X S_Y} = \frac{-826.17}{(293.30)(3.90)} = -.0722$$

با به کارگیری فرمول دوم، داریم

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{57(174474) - (41253)(305)}{\sqrt{[57(34673815) - 41253^2][57(2483) - 305^2]}}$$

$$= \frac{-2637147}{\sqrt{(274597450)(48506)}} = -0.722$$

در آخر به شکل مجموعه مجزورات و مضروبات داریم

$$SS_X = \sum X^2 - \frac{(\sum X)^2}{N} = 4817499.10$$

$$SS_Y = \sum Y^2 - \frac{(\sum Y)^2}{N} = 850.98$$

$$SP_{XY} = \sum XY - \frac{\sum X \sum Y}{N} = -46265.74$$

$$r = \frac{SP_{XY}}{\sqrt{SS_X SS_Y}} = \frac{-46265.74}{\sqrt{(4817499.10)(850.98)}} = -0.722$$

پس

در هر سه مورد می‌بینیم که همبستگی برابر است با -0.722 .

ضریب همبستگی را بایستی با احتیاط تفسیر نمود تا چیزی را که شامل نمی‌شود به معنای آن نسبت ندهیم. بخصوص نباید $r = -0.722$ را به این صورت تفسیر کرد که ۷۲٪ از یک رابطه (هرچه که باشد) بین زمان و خطاها وجود دارد. ضریب همبستگی به طور ساده یک نقطه روی مقیاس بین -1.00 و $+1.00$ است، و هرچه به یک انتها نزدیکتر باشد این رابطه قوی‌تر است. برای تفسیر دقیق‌تر بهتر خواهد بود که به شکل r^2 آن را تفسیر کنیم که در فصل بعد توضیح داده می‌شود.

۹-۵ همبستگی با داده‌های رتبه‌ای

در مثال پیش داده‌های مربوط به هر آزمودنی در واحدهای روزانه مثل زمان و تعداد خطاها ثبت شده بود. اما خیلی وقت‌ها، ما از افراد می‌خواهیم که تعدادی عنصر را در دو بعد رتبه‌بندی کنند و سپس می‌خواهیم بین این دو مجموعه از رتبه‌ها همبستگی بگیریم. برای مثال، ممکن است از یک نفر بخواهیم اهمیت ۱۰ مطلب را رتبه‌بندی کند. به کم اهمیت‌ترین آنها رتبه ۱، و به بعدی ۲، و الی آخر داده خواهد شد. داور دیگری ممکن است سادگی و روان بودن نگارش آن ۱۰ مطلب را رتبه‌بندی کند. سپس می‌توانیم بین دو مجموعه رتبه‌ها همبستگی بگیریم تا ببینیم آیا مطالبی که کم اهمیت هستند روان‌تر هم هستند.

وقتی چنین داده‌های رتبه‌ای (ترتیبی) داشته باشیم، اغلب از ضریب همبستگی اسپیرمن^۱ که با r_s نشان داده می‌شود استفاده می‌کنیم. (این تنها ضریب برای داده‌های رتبه‌ای نیست بلکه یکی از ساده‌ترین آنهاست).

احتمالا جالب‌ترین چیز در باره r_s اسپیرمن از نظر شما این است که مجبور نیستید چیز جدیدی یاد بگیرید. از قبل می‌دانید که چگونه آن را محاسبه کنید. بیشتر کتاب‌ها یک فرمول جالب برای انجام r_s دارند، و بعضی حتی به شما واثق می‌کنند که یک آماره جدید را فرا می‌گیرید. اما در حقیقت آن فرمول ساده از ساده کردن فرمول اصلی r بدست آمده است.

اسپیرمن فرمول خود برای همبستگی رتبه‌ای را با توجه به این موضوع استخراج کرد که اگر ما داده‌های رتبه‌ای را در باره ۱۰ چیز داشته باشیم، رتبه‌ها برای هر متغیر وابسته ۱۰ عدد صحیح متوالی نخست خواهند بود (۱، ۲، ۳، ...، ۱۰). مجموع N عدد صحیح نخست ($\sum X$) همیشه $N(N+1)/2$ و مجموع N عدد صحیح مجذور شده ($\sum X^2$) همیشه $N(N+1)(2N+1)/6$ است. تمام کاری که اسپیرمن انجام داد این بود که در فرمول پیرسون $\sum X$ و $\sum Y$ را با $N(N+1)/2$ و $\sum X^2$ و $\sum Y^2$ را با $N(N+1)(2N+1)/6$ جایگزین نمود. این گام یک ساده‌سازی واقعی برای زمانی بود که افراد دسترسی به ماشین حساب نداشتند، اما اکنون که ماشین حساب‌ها در اختیار همه هست چنین محاسباتی دیگر فایده ندارد. شما می‌توانید به سادگی از فرمول اصلی پیرسون استفاده کنید. بیشتر مردم ترجیح می‌دهند برای مجذور و حاصل جمع تعدادی عدد کار اندکی انجام دهند اما یک فرمول جدید را حفظ نکنند. این فرمول را در اینجا ارائه می‌کنیم تا کسانی که آن را حفظ کرده‌اند احساس نکنند فریب خورده‌اند.

زمانی که داده‌ها در هر دو متغیر وابسته به شکل رتبه‌ای باشد

$$r = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

که D نشان‌دهنده مجموعه اختلاف بین رتبه‌های X و Y برای هر آزمودنی است، چنانکه در جدول ۹-۳ نشان داده شده است.

به عنوان یک نمونه از کاربرد همبستگی‌های رتبه‌ای، فرض کنید که از گروهی از دانش‌آموزان بخواهیم کیفیت ۱۵ درس را رتبه‌بندی کنند و همچنین میزان کار لازم برای هر یک را نیز رتبه‌بندی کنند. (از دانش‌آموزان خواسته شده به نوعی جمع‌بندی برسند که هر درس یک رتبه برای کیفیت و یک رتبه برای کار بگیرد). داده‌ها در جدول ۹-۳ آمده است. با استفاده از یکی از فرمول‌های استاندارد برای r پیرسون ولی با نامیدن آن به عنوان r_s برای خاطر نشان کردن اینکه با داده‌های رتبه‌ای سروکار داریم

$$r_s = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{15(1112) - (120)(120)}{\sqrt{[15(1240) - 120^2][15(1240) - 120^2]}}$$

$$= \frac{2280}{\sqrt{(4200)(4200)}} = \frac{2280}{4200} = 0/543$$

با به کارگیری یک فرمول جایگزین،

$$r_s = 1 - \frac{\sum D^2}{N(N-1)} = 1 - \frac{6(256)}{15(15-1)} = 1 - \frac{1536}{15(224)} = 1 - 0/457 = 0/543$$

توجه دارید که دو جواب دقیقا یکسان هستند، باید هم چنین باشد زیرا این فرمول‌ها از لحاظ جبری معادل هم هستند. همچنین توجه خواهید نمود که رابطه متوسط قوی ($r_s = 0/543$) بین کیفیت یک درس و بار کاری آن وجود دارد.

جدول ۳-۹ رتبه‌بندی کیفیت و کار لازم برای هر یک از ۱۵ درس

درس	کیفیت	کار	D	D ²
A	۱	۳	-۲	۴
B	۲	۵	-۳	۹
C	۳	۱	۲	۴
D	۴	۶	-۲	۴
E	۵	۲	۳	۹
F	۶	۸	-۲	۴
G	۷	۱۴	-۷	۷۹
H	۸	۱۲	-۴	۱۶
I	۹	۱۱	-۲	۴
J	۱۰	۹	۱	۱
K	۱۱	۱۰	۱	۱
L	۱۲	۷	۵	۲۵
M	۱۳	۱۵	-۲	۴
N	۱۴	۱۳	۱	۱

۱۲۱	۱۱	۴	۱۵	۰
$\sum X = ۱۲۰$		$\sum Y = ۱۲۰$	$\sum D^2 = ۲۵۶$	
$\sum X^2 = ۱۲۴۰$		$\sum Y^2 = ۱۲۴۰$		
$\sum XY = ۱۱۱۲$				
توجه: رتبه پایین نشان دهنده کیفیت بالا و بار کاری زیاد است.				

۹-۵-۱ رتبه‌های همتا^۱

همتاها عموماً مایه دشواری زیادی برای افرادی هستند که با داده‌های رتبه‌ای کار می‌کنند. رتبه‌های همتا گرایش به نامعتبر کردن فرمول ما و آسیب زدن به منطق شیوه‌های ما دارند. در مورد r_s اسپیرمن، فرمول "ساده شده" با همتاها دچار خطا می‌شود زیرا $\sum X^2$ دیگر $N(N+1)/6$ نیست. البته مادامی که شما با فرمول اصلی پیرسون کار می‌کنید مشکل خاصی وجود ندارد. در این حالت به طور ساده شما به افراد یا اشیائی که نمره یکسان دارند، معدل رتبه‌هایی که در آنها همتا شده‌اند را اختصاص می‌دهید و آنگاه فرمول استاندارد را به شکل معمول آن به کار می‌بندید. به عنوان نمونه‌ای از رتبه‌های اختصاص یافته به نمرات همتا، فرض کنید داده‌های زیر را داشته باشیم:

۸ ۸ ۸ ۵ ۳ ۳ ۱

آنگاه رتبه‌ها چنین خواهند بود

۶ ۶ ۶ ۴ ۲/۵ ۲/۵ ۱

۹-۵-۲ تفسیر r_s

تفسیر r_s کمی مشکل‌تر از r پیرسون است، بخشی به خاطر ماهیت داده‌ها است. در مثال اخیر ما، داده‌ها بطور طبیعی به شکل رتبه‌ای رخ داده‌اند زیرا این تکلیفی بوده که آزمودنی‌ها را بر اساس آن تنظیم کردیم. در این شرایط r_s اندازه‌ای از رابطه خطی بین یک مجموعه و مجموعه دیگری از رتبه‌ها است. در یک شرایط دیگر داده‌ها بطور طبیعی رخ نداده بلکه ما داده‌های خام را به رتبه‌ها تبدیل می‌کنیم. مثلاً ممکن است "جامعه‌پذیری" یک شخص را با تعداد دوستانی که او دارد و اندازه‌ی جذابیت فیزیکی او را با درخواست از یک داور مستقل برای اختصاص یک رتبه در مقیاسی ۱۰ درجه‌ای بسنجیم. با اعتماد خیلی کم به ویژگی‌های زیربنایی هر مقیاس، ممکن است بسادگی داده‌های خام هر متغیر را به رتبه تبدیل کنیم و روی رتبه‌ها همبستگی را اجرا کنیم. در این حالت مقدار r_s اندازه‌ای از رابطه خطی بین رتبه‌ها است اما تنها یک اندازه‌ی **رابطه یکنواخت**^۲ بین متغیرها است (چیزی که پیوسته بالا می‌

1. tied ranks

2. monotonic relationship

رود یا پایین می‌آید). نباید از این موضوع تعجب کنید. یک ضریب همبستگی مستقیماً متغیرهایی را توضیح می‌دهد که براساس آنها حساب شده است. نمی‌توان انتظار داشت اطلاعات دقیقی در باره متغیرهایی که براساس آنها حساب نشده است بدست دهد. چنانکه در فصل ۲ گفته شد، باید در نظر داشت که بین متغیرهایی که شما واقعا اندازه‌گیری کرده‌اید (مثل تعداد دوستان) و ویژگی‌های زیربنایی که مایلید بیازمایید (مثل جامعه‌پذیری) چنین تفاوتی وجود دارد.

۹-۶ عواملی که بر همبستگی اثر می‌گذارند

ضریب همبستگی می‌تواند به گونه‌ی مهمی تحت تأثیر ویژگی‌های نمونه قرار گیرد. دو مورد از این ویژگی‌ها شامل محدودیت دامنه (یا واریانس) X و Y و استفاده از نمونه‌های فرعی ناهمگن است.

۹-۶-۱ اثر محدودیت دامنه

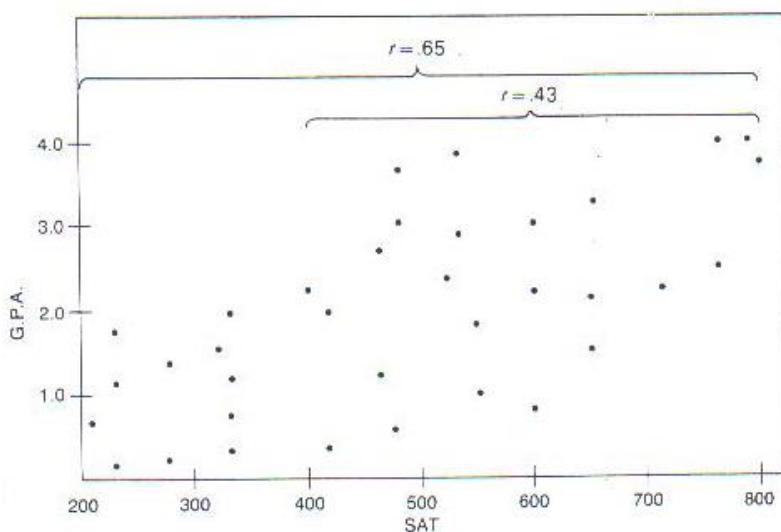
یک مشکل معمول که زیاد پیش می‌آید، مربوط به محدودیت‌های دامنه‌ای است که در آن X و Y تغییر می‌کنند. اثر چنین محدودیت‌های دامنه^۱، تغییر دادن همبستگی بین X و Y نسبت به حالتی خواهد بود که دامنه تا این حد محدود نباشد. بسته به ماهیت داده‌ها، همبستگی ممکن است افزایش یا کاهش پیدا کند، هرچند در اغلب اوقات r کاهش می‌یابد.

به استثنای موارد خیلی نادر، زمانی که محدودیت منجر به حذف برخی روابط غیرخطی شود، کاهش دامنه X باعث کاهش r می‌شود. مثلاً اگر بین قد و سن همبستگی بگیریم، در حالتی که دامنه سن از ۰ تا ۷۰ سالگی باشد، داده‌ها مسلماً منحنی شکل خواهند بود (تا ۱۷ سالگی افزایش و سپس ثابت می‌ماند یا کاهش دارد)، و همبستگی که روابط خطی را اندازه می‌گیرد کاملاً پایین خواهد بود. اما اگر دامنه را به ۰ تا ۱۷ سالگی محدود کنیم، همبستگی کاملاً بالاست، زیرا آن مقادیر Y که در اثر X بطور خطی تغییر نکرده‌اند را حذف کرده‌ایم.

با کاهش دامنه X یا Y ، معمولاً کاهش همبستگی رخ می‌دهد. این مشکل بخصوص در سطح ساخت آزمون اهمیت دارد، زیرا در این حوزه اندازه‌های ملاک (Y) ممکن است تنها برای اندازه‌های بالای X در دسترس باشند. به داده‌های فرضی شکل ۵-۹ توجه کنید. این شکل نشان دهنده رابطه بین معدل دانشگاه و نمرات یک تست پیشرفت تحصیلی (مانند SAT) برای یک نمونه فرضی از دانشجویان است. در دنیای ایده‌آلی سازنده تست همه افرادی که امتحان می‌دهند بعداً به دانشگاه فرستاده شده و یک معدل می‌گیرند، و همبستگی بین نمرات تست و معدل محاسبه خواهد شد. همانطور که از شکل ۵-۹ بر می‌آید، این همبستگی بطور منطقی بالاست. اما، در دنیای واقعی همه ثبت‌نام نمی‌کنند. دانشگاه فقط دانشجویان با استعدادتر را می‌گیرد، خواه این توانایی بوسیله تست پیشرفت یا نمرات دبیرستان یا هر چیز دیگر اندازه‌گیری شده باشد. یعنی معدل دانشگاهی عمدتاً برای دانشجویانی قابل دستیابی است که

1. range restrictions

دارای نمرات نسبتاً بالاتر در آزمون استاندارد هستند. فرض کنید که این دارای اثری باشد که به ما امکان ارزیابی رابطه بین X و Y را تنها برای مقادیری از X بدهد که بیشتر از ۴۰۰ باشند. از شکل ۵-۹ درمی‌یابیم که در این حالت رابطه نسبتاً پایین خواهد بود، نه به این خاطر که تست بی‌ارزش است بلکه به این خاطر که دامنه محدود است. به عبارت دیگر وقتی ما کل نمونه را در شکل ۵-۹ مورد استفاده قرار می‌دهیم همبستگی ۰/۶۵ است. اما وقتی نمونه را محدود به دانشجویانی می‌کنیم که نمرات SAT آنها حداقل ۴۰۰ است، همبستگی به ۰/۴۳ افت می‌کند.

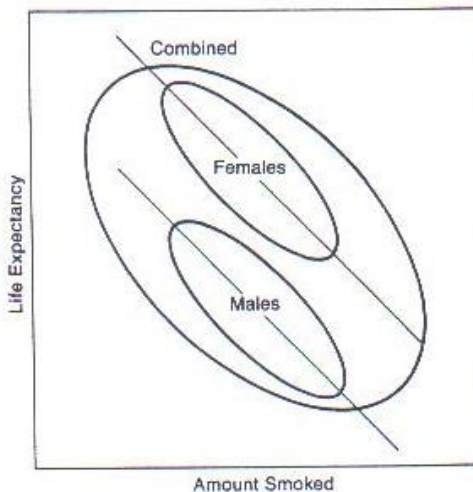


شکل ۵-۹ داده‌های فرضی نشان دهند اثر محدودیت دامنه

در هر جا که با یک ضریب اعتباری که براساس یک نمونه محدود بنا شده است سروکار داریم، بایستی اثر محدودیت دامنه به حساب آورده شود. این ضریب ممکن است به کلی برای سؤال مورد نظر نامناسب باشد. در اصل آنچه ما انجام می‌دهیم این است که می‌پرسیم یک آزمون استاندارد شده به چه خوبی مناسب بودن یک شخص را برای دانشگاه پیش‌بینی می‌کند، اما ما این سؤال را تنها با رجوع به افرادی پاسخ می‌دهیم که عملاً در دانشگاه ثبت‌نام شده‌اند. در همین حال، گاهی سودمند است که بطور عمدی دامنه یک متغیر را محدود کنیم. بطور مثال، اگر بخواهیم روشی که از طریق آن توانایی خواندن افزایش می‌یابد را بدانیم، ما احتمالاً دامنه سنی را با تمرکز بر آزمودنی‌هایی که دست کم ۵ سال و حداکثر ۲۰ سال سن دارند (یا یک محدوده منطقی بالاتر) محدود کنیم. مسلماً هرگز انتظار نداریم توانایی خواندن بطور نامحدود ادامه پیدا کند.

۶-۹-۲ اثر نمونه‌های ناهمگون

موضوع قابل توجه دیگر در ارزیابی نتایج تحلیل‌های همبستگی به نمونه‌های فرعی ناهمگون^۱ (نامتجانس) بر می‌گردد. این نکته را با یک مثال ساده می‌توان بیان نمود. یک مطالعه فرضی در باره مصرف سیگار و امید زندگی را در نظر بگیرید، که داده‌های مربوط به زنان و مردان بطور جداگانه ترسیم شده‌اند تا از مخدوش شدن تفاوت امید زندگی که معمولاً بین دو جنس دیده می‌شود جلوگیری شود. فرض کنید نتایج مطابق شکل ۶-۹ است، داده‌ها در آن بزرگ‌نمایی شده‌اند تا روشن‌تر باشد.



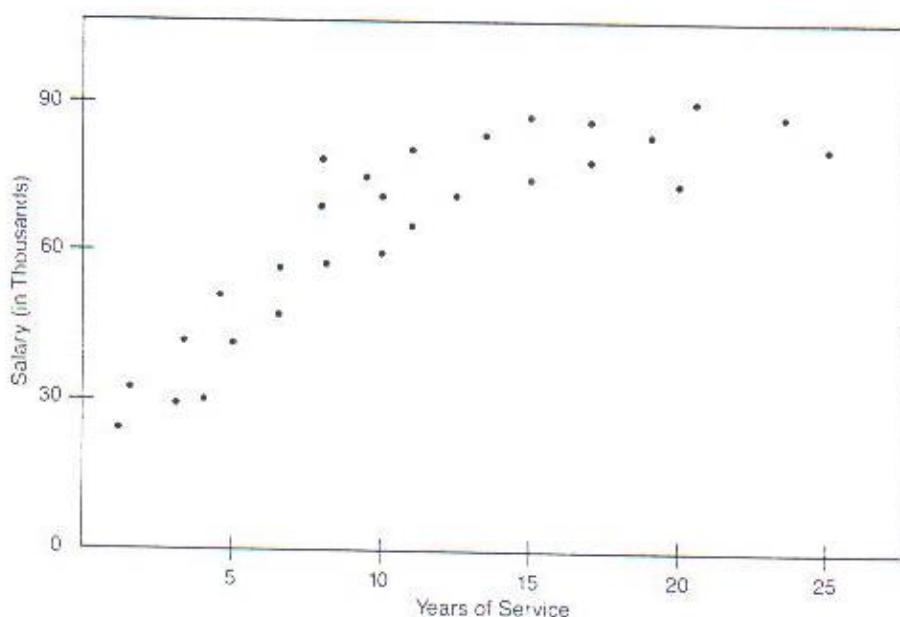
شکل ۶-۹ نمایش اثر نمونه‌های فرعی نامتجانس

در این شکل بیضی‌ها نشان دهنده خوشه‌های داده‌ها هستند. مشاهده می‌شود که برای مردان رابطه بالایی بین مصرف سیگار و امید زندگی وجود دارد. در باره زنان هم این موضوع به همین اندازه درست است. اما اگر داده‌ها را در یک نمونه بزرگ گروه‌بندی کنیم، رابطه خیلی مخدوش می‌شود. این واقعیت که همبستگی برای نمونه ترکیبی پایین است ربطی به رابطه بین سیگار و امید زندگی ندارد، بلکه به رابطه بین جنسیت و امید زندگی برمی‌گردد. نکته‌ای که در اینجا آشکار می‌شود این است که آزمایشگران وقتی داده‌های مربوط به منابع گوناگون را ترکیب می‌کنند بایستی خیلی مراقب باشند، طوری که واریانس متغیرهای نامربوط را به حساب نیاورند.

۹-۷ یک نکته احتیاطی

لازم است به دو چیز که معمولاً نادیده گرفته می‌شوند توجه شود. نخست اینکه ما در اینجا با روابط خطی سروکار داریم، مورد دوم به موضوع علت و معلولی برمی‌گردد.

ضریب همبستگی گشتاوری پیرسون فقط با روابط خطی بین دو متغیر سروکار دارد- یعنی جایی که بهترین خطی که می‌توان از درون نمودار پراکنش گذراند یک خط مستقیم است. معمولاً وقتی بدانیم یک رابطه غیرخطی است محاسبه r سودمند نیست. (همبستگی بازهم اندازه‌ای از میزان رابطه خطی خواهد بود، اما چنانچه رابطه خیلی منحنی شکل باشد، چه کسی تضمین می‌کند یک خط مستقیم تا چه اندازه با نقاط تطبیق خواهد داشت؟) به همین ترتیب این واقعیت که r نزدیک به صفر است به این معنی نیست که بین X و Y رابطه وجود ندارد. نمونه‌ای از یک رابطه غیرخطی در شکل ۹-۷ در روابط فرضی بین درآمد و میزان سنوات خدمت در مدیران میانی در یک شرکت با اندازه متوسط ارائه شده است.



شکل ۹-۷ رابطه غیرخطی بین درآمد و سنوات خدمت

نکته احتیاط‌آمیز دوم به این واقعیت بر می‌گردد که وجود رابطه بین دو متغیر به این معنی نیست که یک متغیر علت دیگری است. این حقیقت که دیدن لک‌ها روی دودکش با تغییرات نرخ زاد و ولد

همبستگی دارد دلیلی بر این نیست که لک‌ها تأثیری بر زاد و ولد دارند (مگر اینکه اطلاعات من بسیار نادرست باشد). هر دوی این متغیرها تحت تأثیر فصل سال هستند و این باعث همبستگی می‌شود.

۸-۹ ضریب‌های همبستگی دیگر

ضریب استاندارد همبستگی، r پیرسون است، که در اصل برای متغیرهایی به کار برده می‌شود که در مقیاس‌های فاصله‌ای یا نسبی سنجیده شده‌اند. ما همچنین دیدیم که از همین فرمول وقتی که متغیرها به شکل رتبه باشند یک آماره بنام r_s (یا r_s اسپیرمن) تولید می‌شود. دو ضریب همبستگی دیگر وجود دارد که لازم است با آنها آشنا شوید گرچه در اینجا هم کمتر چیز تازه‌ای دیده می‌شود.

زمانی که ما یک متغیر داریم که در مقیاس پیوسته اندازه‌گیری شده و یک متغیر که دو بخشی است (مثلاً در یکی از تنها دو سطح روی می‌دهد)، آنگاه ضریب همبستگی که ایجاد می‌شود **همبستگی دو رشته‌ای نقطه‌ای** (r_{pb}) نامیده می‌شود. مثلاً می‌توانیم یک تحلیل از سؤال‌های آزمون از طریق همبستگی نمره کل (X)، با درست/نادرست بودن یک سؤال خاص (Y) به عمل آوریم. در این حالت مقادیر X ممکن است از ۶۰ تا ۱۰۰ امتداد داشته باشند، اما مقادیر Y ، ۰ (نادرست) یا ۱ (درست) باشند. هر چند برای محاسبه ضریب همبستگی r_{pb} فرمول خاصی وجود دارد، همین کار را بسادگی با محاسبه r نیز می‌توان انجام داد. فقط ما جواب را r_{pb} به جای r می‌نامیم تا به ماهیت داده‌هایی که براساس آنها محاسبه صورت گرفته اشاره کرده باشیم.

در اینجا یک نکته در باره **متغیرهای دو بخشی**^۲ وجود دارد. در مثال قبلی ما “نادرست” را به صورت ۰ و “درست” را به صورت ۱ نمره‌گذاری کردیم. این کار محاسبات را ساده می‌کند. ما همچنین می‌توانستیم آنها را به صورت ۱ و ۲، یا حتی ۸۲ و ۹۷ نمره‌گذاری کنیم – بطوری که همه نمره‌های “درست” یک عدد و همه نمره‌های “نادرست” عدد دیگری بگیرند. همبستگی یکسان خواهد بود فرقی نمی‌کند که چه عددی به کار برده شود.

یک ضریب همبستگی اندکی متفاوت، **فی (ϕ)**، زمانی به کار برده می‌شود که هر دو متغیر به صورت دو وجهی اندازه‌گیری شده باشند. برای مثال، در مطالعه رابطه بین جنسیت و مذهبی بودن، ما بین جنسیت (کد شده به صورت مرد = ۱، زن = ۲) و گرایش معمول به کلیسا (نه = ۰، بله = ۱) همبستگی می‌گیریم. در اینجا هم فرق نمی‌کند کدام دو مقدار را ما برای کدگذاری متغیرهای دو وجهی به کار

1. point biserial correlation

2. dichotomous variables [دو وجهی یا دو ارزشی هم گفته می‌شود]

ببریم. اگرچه ϕ دارای فرمول خاصی است، استفاده از فرمول پیرسون نیز ساده و درست است، اما جواب را با ϕ نشان می‌دهیم.

شمار دیگری از ضرایب همبستگی نیز وجود دارد، اما مواردی که اینجا گفته شده بیشتر معمول است. همه آنها موارد خاصی از r پیرسون هستند، و همه را می‌توان با استفاده از فرمول r بدست آورد. سرانجام این ضرایب آنهایی هستند که معمولاً زمانی تدوین می‌شوند که یک مجموعه بزرگ از داده‌ها را وارد یک کامپیوتر کرده و یک برنامه همبستگی یا رگرسیون را اجرا می‌کنند.

۹-۹ یک مثال کاری آخر

مجموعه مشاهدات زیر داده‌های واقعی از ارزیابی ۱۵ درس هستند که از یک مجموعه داده‌های بزرگ در باره ارزیابی چند صد درس گرفته شده‌اند. ما به دو متغیر بسنده می‌کنیم. از دانش‌آموزان خواسته شده بود تا کیفیت کلی یک درس جاری خود را رتبه‌بندی و نمره خود را در آن درس پیش‌بینی کنند. آیا ما می‌توانیم از این داده‌ها نتیجه بگیریم که دروسی که عموماً نمره پایینی دارند احتمالاً همان‌هایی هستند که پایین ارزیابی می‌شوند؟

نمره مورد انتظار (X)	کیفیت کلی (Y)
۳/۵	۳/۴
۳/۲	۲/۹
۲/۸	۲/۶
۳/۳	۳/۸
۳/۲	۳/۰
۳/۲	۲/۵
۳/۶	۳/۹
۴/۰	۴/۳
۳/۰	۳/۸
۳/۱	۳/۴
۳/۰	۲/۸
۳/۳	۲/۹
۳/۲	۴/۱
۳/۴	۲/۷
۳/۷	۳/۹
$\Sigma X = 49.5$	$\Sigma Y = 50.0$

$$\sum X^2 = 164/65 \quad \sum Y^2 = 171/68$$

$$\sum XY = 166/48$$

گام نخست ما محاسبه میانگین و واریانس هر متغیر است. اینها به شرح زیر هستند:

$$\bar{Y} = 49/5 / 15 = 3/3000 \quad \bar{X} = 50/0 / 15 = 3/3333$$

$$S_X = \sqrt{\frac{164.65 - 49.5^2 / 15}{14}} = 0/3047 \quad S_Y = \sqrt{\frac{171.68 - 50.0^2 / 15}{14}} = 0/5984$$

کوواریانس این می شود

$$cov_{XY} = \frac{166.48 - (49.5)(50.0)/15}{14} = 0/1057$$

در نهایت همبستگی به این ترتیب بدست می آید

$$r = \frac{cov_{XY}}{(S_X)(S_Y)} = \frac{0.1057}{(0.3047)(0.5984)} = 0/5797 \approx 0/58$$

این یک همبستگی نسبتاً بالاست، و گویای آن است که دروسی که معدل نمرات بالاتری دارند در رتبه بندی هم معدل بالاتر دارند. این بدین معنا نیست که نمرات بالاتر باعث رتبه بندی بالاتر می شوند. تنها می توان پذیرفت که درس های پیشرفته تر بالاتر رتبه بندی می شوند، و اغلب در این درس ها است که دانش آموزان بهترین عملکرد خود را صورت می دهند. همچنین ممکن است که واقعا همبستگی وجود نداشته باشد بلکه بطور شانسی به یک همبستگی دست پیدا کرده باشیم. نحوه آزمودن این مسأله را در فصل ۱۲ خواهیم دید.

۹-۱۰ خلاصه

در این فصل با ضریب همبستگی به عنوان اندازه ای از رابطه بین دو متغیر آشنا شدیم. دیدیم که چگونه یک نمودار پراکنش برای متغیرها رسم کنیم و چگونه ضریب همبستگی را حساب کنیم. ما همچنین همبستگی داده های رتبه ای را مورد توجه قرار دادیم و دیدیم که فرمول اصلی پیرسون برای این منظور مناسب است. در انتها به عامل هایی که بر مقدار همبستگی تأثیر دارند، بعضی احتیاط ها، و متغیرهای دو وجهی پرداختیم. برخی از مطالب مهم فصل عبارتند از:

- همبستگی
- ضریب همبستگی
- ضریب همبستگی گشتاوری پیرسون (r)
- کوواریانس
- مجموع مجذورات
- ضریب همبستگی رتبه‌ای اسپیرمن (r_s)
- نمودار پراکندگی
- متغیر پیش‌بین
- متغیر ملاک
- رابطه منحنی شکل
- رابطه خطی
- رابطه یکنواخت
- همبستگی نقطه‌ای دو وجهی
- متغیرهای دو وجهی
- فی (ϕ)

۹-۱۱ تمرین‌ها

۹-۱ یکی از استان‌های کشور دارای ۱۰ شهرستان است. داده‌های روبرو درصد باروری در هر ۲۵۰۰ مورد (Y)، نرخ زاد و ولد برای زنان سنین ۱۷ تا ۳۵ سال (X_1) و درصد باروری‌های ناخواسته (X_2) در هر شهرستان را نشان می‌دهد.

الف) یک نمودار پراکنش برای Y و X_1 رسم کنید.

ب) خطی را رسم کنید (بطور دستی) که بهترین تناسب را با داده‌ها داشته باشد.

۹-۲ همبستگی بین Y و X_1 را در تمرین ۹-۱ حساب کنید.

شهرستان	Y	X_1	X_2
۱	۶/۱	۴۳/۰	۹/۲
۲	۷/۱	۵۵/۳	۱۲/۰
۳	۷/۴	۴۸/۵	۱۰/۴
۴	۶/۳	۳۸/۸	۹/۸
۵	۶/۵	۴۶/۲	۹/۸
۶	۵/۷	۳۹/۹	۷/۷
۷	۶/۶	۴۳/۱	۱۰/۹
۸	۸/۱	۴۸/۵	۹/۵
۹	۶/۳	۴۰/۰	۱۱/۶
۱۰	۶/۹	۵۶/۷	۱۱/۶

- ۹-۳ در تمرین ۹-۱ همبستگی بین Y و X_2 را حساب کنید.
- ۹-۴ برای Y و X_1 در تمرین ۹-۱ یک نمودار پراکنش رسم کنید. این نمودار را مانند شکل ۹-۲ به چهار قسمت تقسیم و مشاهدات هر بخش را شمارش کنید.
- ۹-۵ برای Y و X_2 در تمرین ۹-۱ یک نمودار پراکنش رسم کنید. این نمودار را مانند شکل ۹-۳ به چهار قسمت تقسیم و مشاهدات هر بخش را شمارش کنید.
- ۹-۶ جواب تمرین‌های ۹-۴ و ۹-۵ را باهم مقایسه کنید. آیا اینها بازتابی از تفاوت پاسخ‌های تمرین‌های ۹-۲ و ۹-۳ می‌باشند؟

۹-۷ برای داده‌های زیر نمودارهای پراکنش را رسم کنید.

۳		۲		۱	
Y	X	Y	X	Y	X
۸	۲	۴	۲	۲	۲
۶	۳	۲	۳	۴	۳
۴	۵	۸	۵	۶	۵
۲	۶	۶	۶	۸	۶

- الف) برای هر یک کوواریانس را با استفاده از فرمول تعریف کننده حساب کنید.
- ب) برای هر یک کوواریانس را با استفاده از فرمول محاسباتی حساب کنید.
- ۹-۸ همبستگی هر مجموعه داده‌ها را در تمرین ۹-۷ حساب کنید. چگونه مقادیر Y را در تمرین ۹-۷ می‌توان تغییر داد که کوچک‌ترین همبستگی مثبت بدست آید؟
- ۹-۹ داده‌های زیر درصد مسأله‌های حل شده در تکالیف آزاد منزل که توسط ۲۰ دانش‌آموز تکمیل شده است و نمره پایانی درس (تبدیل شده به مقیاس ۱۰۰ درجه‌ای) را نشان می‌دهد.
- مسأله‌های تکمیل شده

۳۰ ۸۰ ۴۰ ۷۰ ۹۵ ۸۵ ۸۰ ۴۰ ۹۵ ۵۰ ۸۰ ۹۰ ۸۵ ۱۰۰ ۴۰ ۹۰ ۷۰ ۸۰ ۶۰ ۵۰

نمره پایانی

۵۵ ۸۰ ۶۰ ۷۵ ۸۵ ۷۰ ۵۰ ۶۰ ۹۰ ۷۵ ۸۰ ۹۵ ۹۵ ۹۸ ۶۰ ۸۵ ۸۰ ۹۰ ۷۵ ۷۵

الف) نقاط داده‌ها را رسم کنید.

ب) همبستگی بین میزان مسأله‌های تکمیل شده و نمره پایانی را حساب کنید.

ج) این همبستگی را تفسیر کنید.

تمرین‌های ۹-۱۰ تا ۹-۱۵ شامل یک میزان کار غیرمعمول است. اما داده‌ها واقعی و جواب‌ها مهم هستند.

استرنگلاس و بل (۱۹۸۳) یک مقاله بحث‌برانگیز منتشر کردند که "عامل اصلی در کاهش تند نمرات پیشرفت تحصیلی دانش‌آموزان SAT در دهه ۱۹۷۰ ناشی از آزمایشات بمب‌های هسته‌ای در دهه‌های قبلی ۱۹۵۰ و ۱۹۶۰ بوده است." در حمایت از این فرضیه آنها یک همبستگی 0.55 بین سطح ماده اتمی-ید-۱۳۱ در سال ۱۹۶۲ (وقتی آزمایشات اتمی محدودی در نوادا از سر گرفته شد) و کاهش نمره مقیاس توانایی کلامی SAT از ۱۹۷۸ تا ۱۹۷۹ را گزارش کردند. ادعای آنها این بود که در معرض ید-۱۳۱ قرار گرفتن در کودکی منجر به عملکرد ضعیف بعدی در طول زندگی می‌شود. همبستگی برای کاهش در سال‌های ۱۹۷۸ و ۱۹۷۷ به ترتیب به صورت 0.2956 و 0.3300 گزارش شده است. داده‌های خام برای نمره‌های خام ریاضی و کلامی SAT در سال‌های ۱۹۷۲-۱۹۷۳ تا ۱۹۷۸-۱۹۷۹ در این ایالت‌ها همراه با سطح ید-۱۳۱ به شرح جدول ۵-۹ است. (هر ردیف نشان دهنده یک ایالت است. نمره ۱۹۷۲-۱۹۷۳ به صورت ۱۹۷۲ آمده چون آزمون در پاییز برگزار می‌گردید).

۹-۱۰ کاهش نمرات ریاضی و کلامی را در سال‌های ۱۹۷۸ تا ۱۹۷۹ حساب کنید و همبستگی این کاهش با سطح ید-۱۳۱ را بدست آورید. آیا این جواب با نتایج گزارش شده همخوانی دارد؟

۹-۱۱ یک نمایش ساقه-برگ برای متغیر ید-۱۳۱ و تغییر در نمرات ریاضی از ۱۹۷۷-۱۹۷۹ طرح کنید.

۹-۱۲ یک نمودار پراکنش از رابطه بین کاهش نمرات ریاضی SAT در ۷۸-۱۹۷۷ تا ۷۹-۱۳۷۸ و سطح ید-۱۳۱ رسم کنید. این نمودار گویای چیست؟

۹-۱۳ بیشتر داده‌های نقاط انتهایی در نمودار پراکنش در تمرین ۹-۱۲ بر اساس 4% از دانش‌آموزان دبیرستانی آن ایالت است (نقاط انتهایی دیگر بر اساس 31% آنان است). این موضوع چه مشکلی ایجاد می‌کند؟

۹-۱۴ دو نقطه انتهایی را در تمرین ۹-۱۲ حذف کنید و همبستگی را دوباره حساب کنید.

۹-۱۵ در سال ۱۹۵۷ هم افزایش قابل توجهی در آزمایشات هسته‌ای نوادا در مقایسه با سال قبل صورت گرفته است. اگر فرض کنیم که توزیع ید-۱۳۱ در ایالت‌ها شبیه به توزیع سال ۱۹۶۲ بوده است، انتظار دارید در این داده‌ها چه چیزی را مشاهده کنید؟ از جواب تمرین ۹-۱۵ برای ارزیابی این انتظار استفاده کنید.

جدول ۹-۵ داده های مربوط به نمرات SAT و پیامدهای رادیواکتیو

V72	V73	V74	V75	V76	V77	V78	M72	M73	M74	M75	M76	M77	M78	I-131
464	474	461	454	459	458	452	486	495	481	485	487	485	478	104
502	501	496	491	492	492	485	534	534	525	532	531	529	524	12
478	484	482	485	480	483	478	511	512	510	518	519	514	513	37
452	450	435	430	427	427	428	485	484	473	470	470	466	473	13
485	484	479	476	469	473	469	521	522	515	521	518	516	513	15
451	452	442	439	437	438	435	482	480	471	473	468	469	465	21
447	449	439	437	438	436	433	482	479	476	476	477	477	468	32
417	420	414	409	401	398	396	481	481	478	479	473	468	467	12
484	492	493	483	489	488	480	520	518	524	519	530	521	518	25
472	471	460	464	459	463	462	509	513	510	509	507	511	511	38
429	423	418	415	412	413	412	470	469	463	460	458	457	455	33
525	526	523	523	513	516	518	565	570	568	572	563	569	567	59
441	441	437	437	431	429	430	481	477	471	476	471	467	468	22
454	452	436	432	431	431	426	488	485	471	470	469	466	464	24
444	445	434	432	429	430	428	481	477	469	469	465	465	463	22
450	458	451	453	454	459	456	491	502	498	506	508	511	508	30
506	512	506	504	500	497	497	552	556	552	557	556	550	549	43
475	475	465	464	464	465	461	509	510	500	505	506	506	504	58
505	512	500	490	493	493	487	553	558	547	545	550	549	541	40
475	476	465	456	457	457	449	501	502	497	497	498	494	486	*
459	461	449	446	443	446	444	497	496	485	488	484	486	483	20
438	436	424	422	422	419	419	469	464	454	456	455	451	452	22
482	492	486	481	481	485	486	522	525	516	522	521	524	521	21
452	452	441	437	434	431	426	495	495	484	484	479	471	469	28
458	459	456	457	459	459	457	496	500	499	504	506	504	502	35
491	492	480	482	483	483	486	528	527	514	515	525	527	522	48
445	446	440	432	432	437	434	482	475	468	469	468	469	466	27
443	442	430	431	429	427	426	479	477	470	470	468	467	464	36
442	439	432	429	424	425	424	477	471	469	469	461	463	459	23
438	439	431	427	424	425	418	474	475	467	466	464	460	456	38
528	532	516	506	515	516	507	561	560	553	545	550	564	538	87
448	445	439	435	435	434	434	483	478	476	474	475	473	472	23
484	496	489	479	480	484	481	523	528	522	522	525	525	525	43
496	502	492	483	479	479	476	548	552	544	546	545	540	535	39
516	500	506	486	501	498	488	559	547	548	545	555	542	532	55
448	454	459	466	476	484	487	498	506	507	518	533	538	538	53
401	401	386	387	376	*	381	422	418	403	403	377	*	401	23
504	510	503	496	503	*	498	545	546	540	543	550	*	541	59
477	484	470	476	480	*	*	511	517	507	517	520	*	*	29
456	455	456	464	466	*	460	491	488	491	501	505	*	495	23
426	466	477	470	466	*	468	460	493	503	499	501	*	497	23
522	519	510	504	501	*	525	568	558	554	548	561	*	570	41
398	386	367	363	354	*	*	437	430	428	417	420	*	*	14
515	528	523	515	534	*	531	552	573	561	586	586	*	570	39
473	479	462	466	459	*	461	505	514	502	510	499	*	494	19

فصل ۱۰

رگرسیون

۱-۱۰- مثالی از سالن نهارخوری دانشگاه

۱۰-۲- خط رگرسیون

۱۰-۳- درستی پیش‌بینی

۱۰-۴- رگرسیون چندگانه

۱۰-۵- یک مثال آخر

در بسیاری از موقعیت‌هایی که ما دو متغیر داریم تنها در پی این هستیم که ببینیم آیا این متغیرها با هم رابطه دارند یا نه، و توجه خاصی به ماهیت دقیق رابطه نداریم. بطور مثال، با آگاهی از اینکه رابطه بین زمان انجام یک تکلیف و تعداد خطاها برابر $0.722 -$ است، ممکن است هر چه را که می‌خواهیم بدانیم از این رابطه بدست آوریم. از سوی دیگر در خیلی از شرایط، شکل رابطه دست کم به اندازه میزان آن در خور توجه است. برای این موضوع شرایط بسیاری وجود دارد که ما می‌خواهیم پا را از رابطه فراتر گذاشته معادله‌ای بنویسیم که امکان پیش‌بینی نمره فرد در یک متغیر را از روی اطلاعات او در متغیر دیگر فراهم سازد. هنگامی که تمایل به استخراج یک معادله برای پیش‌بینی یک متغیر از روی متغیر دیگر داریم، ما با **رگرسیون** سروکار داریم که عنوان این فصل است.

ما در رگرسیون نیز درست مانند مبحث همبستگی، خود را به مواردی محدود خواهیم کرد که خطی که بهترین انطباق را با نمودار پراکنش دارد یک خط مستقیم باشد. به این معنی که ما با **رگرسیون خطی**^۱ سر و کار داریم. این موضوع به آن اندازه که شما آن را یک محدودیت جدی بدانید نیست، چون

1. linear regression

درصد زیادی از مجموعه داده‌ها اساساً خطی هستند. و حتی در مواردی که رابطه غیرخطی است (مثل جایی که بهترین خط انطباقی یک منحنی است) اغلب یک خط مستقیم نیز تقریب خیلی خوبی را فراهم می‌سازد- بخصوص اگر ما حدود انتهایی توزیع یکی از متغیرها را حذف کنیم.

۱۰-۱ یک مثال - از سالن نهارخوری دانشگاه

به عنوان یک نمونه به مدیر یک سرویس غذاخوردی دانشگاه توجه کنید [که در آن در هر وعده غذاهای متعددی ارائه می‌شود.] تمرکز او بر این مسأله است که آیا برای همه کسانی که غذا بخواهند غذا حاضر است، اما بایستی دقت کند که غذای زیادتر از مصرف تهیه نکند. او فرض می‌کند افراد عموماً غذایی را که دوست دارند بیشتر می‌خورند و غذایی که دوست ندارند کمتر می‌خورند. بنابراین او معتقد است اگر بتواند میزان تمایل به منوهای مختلف را ارزیابی کند، آنگاه می‌تواند اطلاعاتی را جمع‌آوری و یک معادله بنویسد که تمایل نسبی را پیش‌بینی کند. هنگامی که منوهای جدیدی در آینده طرح شوند، کاری که او باید انجام دهد این است که نمونه کوچکی از دانشجویان برای رتبه‌بندی تمایل آنها تهیه کند. برای انجام این تحقیق مدیر غذاخوری از گروهی از دانشجویان تقاضا کرد برای ۳۰ منوی استاندارد تمایل خود را رتبه‌بندی کنند. برای هر منو او میانگین رتبه‌بندی تمایل (معدل همه پاسخگویان) را حساب کرد. او همچنین تعداد استفاده از هر منو را که در آخرین بار ارائه در غذاخوری ۵۵۰ نفری دانشجویان صرف شده بود ثبت کرد، که در آن گرفتن پرس دوم و سوم آزاد بود. داده‌های خام در جدول ۱-۱۰ درج شده و در شکل ۱-۱۰ ترسیم شده‌اند.

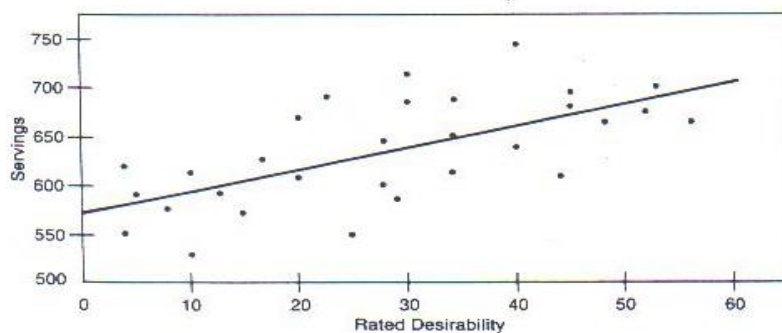
پیش از پرداختن به خط رگرسیون برای پیش‌بینی ارائه غذا از روی رتبه‌بندی بپردازیم، بهتر است ضریب همبستگی بین دو متغیر را حساب کنیم. چنانچه مشخص شود که اینها همبسته نیستند، دیگر جایی برای تلاش جهت بررسی تناسب خط رگرسیون وجود نخواهد داشت. اطلاعات جدول ۱-۱۰ امکان محاسبه همبستگی را به ما می‌دهد. در فصل ۹ ما مفهوم مجموع مجذورات و مجموع مضروب‌ات را توضیح دادیم، بخشی به این خاطر که برای محاسبه رگرسیون سودمند هستند. در اینجا آنها را برای محاسبات این فصل و در راستای فرمول‌های مرسوم مورد استفاده قرار خواهیم داد.

همبستگی بین X و Y با این فرمول بدست می‌آید

$$r = \frac{SP_{XY}}{\sqrt{SS_X SS_Y}} = \frac{15770}{\sqrt{(756.8)(85250)}} = 0.643$$

$$r = \frac{cov_{XY}}{s_X s_Y} = \frac{543.793}{(15.599)(54.219)} = 0.643$$

از روشی که در فصل ۱۲ مورد بحث قرار خواهد گرفت می‌توان دریافت که این همبستگی پایا است. مقدار خوردن یک غذا در افراد بستگی به میزان میل آنها به آن غذا دارد.



شکل ۱۰-۱ نمودار پراکنش رابطه بین تمایل به یک منو و تعداد موارد صرف غذا

جدول ۱۰-۱ داده‌های رابطه بین رتبه‌بندی تمایل به یک منو و تعداد موارد صرف غذا

منو	رتبه‌بندی (X)	صرف غذا (Y)
1	4	550
2	4	620
3	5	590
4	8	575
5	10	530
6	10	615
7	13	590
8	15	570
9	17	630
10	20	610
11	20	670
12	23	690
13	25	550
14	28	645
15	29	585
16	30	715
17	30	685
18	30	715
19	34	615
20	34	650
21	36	690
22	40	640
23	40	745
24	44	610
25	45	680
26	45	695
27	48	665
28	52	675
29	53	700
30	56	665

$$\begin{aligned}\bar{Y} &= ۶۵۳ \\ SS_Y &= ۸۵۲۵۰ & \bar{X} &= ۲۸/۲ \\ s_Y &= ۵۴/۲۱۹ & SS_X &= ۷۰۵۶/۸ \\ SP_{XY} &= ۱۵۷۷۰ & s_X &= ۱۵/۵۹۹ \\ COV_{XY} &= ۵۴۳/۷۹۳\end{aligned}$$

۲-۱۰ خط رگرسیون

چنانچه بخواهیم بدانیم چه تعدادی از یک منوی خاص صرف خواهد شد، یک راهکار مناسب این است که متوسط میزان صرف شده در دفعات گذشته آن را حساب کنیم. اما اگر با یک منوی جدید سر و کار داریم که قبلاً ارائه نشده مسلماً به رویکرد دیگری نیاز داریم. رویکردی که ما در اینجا پیش خواهیم گرفت این است که از یک نمونه از دانشجویان بخواهیم میزان تمایل به آن منو را رتبه‌بندی کنند و سپس از آن برای پیش‌بینی استفاده کنیم. اما نخست ما نیازمند یک معادله بر اساس داده‌های موجود، از یک خط هستیم، که Y (صرف غذا) را به X (رتبه‌بندی) مرتبط می‌سازد. معادله یک خط مستقیم معادله‌ای از نوع $Y = bX + a$ است. برای اهداف ما این معادله را به این صورت می‌نویسیم

$$\hat{Y} = bX + a$$

که

$\hat{Y}$ = مقدار پیش‌بینی شده Y (در اینجا، صرف غذا)

b = شیب خط رگرسیون (مقدار افزایش در $\hat{Y}$ به ازاء یک واحد تغییر در X)

a = مقدار ثابت (مقدار $\hat{Y}$ وقتی که $X = 0$)

متغیر X به طور ساده متغیر پیش‌بینی کننده است (در اینجا رتبه‌بندی). وظیفه ما حل کردن آن برای مقادیر a و b است که بهترین تابع خطی قابل انطباق را فراهم نماید. به عبارت دیگر می‌خواهیم داده‌های موجود را برای حل این مقادیر a و b به گونه‌ای به کار بگیریم که خط رگرسیون (مقادیر پیش‌بینی شده $\hat{Y}$ برای مقادیر مختلف X) تا حد ممکن به داده‌های واقعی Y نزدیک باشد.

اما عبارت "بهترین انطباق" را چگونه می‌توانیم تعریف کنیم؟ منطقی‌ترین روش از طریق **خطاهای**

پیش‌بینی است - یعنی اختلاف $\hat{Y} - Y$. از آنجا که $\hat{Y}$ مقداری است که معادله پیش‌بینی خواهد کرد و Y مقداری است که ما در عمل از آخرین بار صرف غذا بدست آورده‌ایم، $\hat{Y} - Y$ یک خطای پیش‌بینی است. ما می‌خواهیم خطی را پیدا کنیم (یعنی مجموعه‌ای از $\hat{Y}$ ها) که چنین خطاهایی را به حداقل برساند. ما نمی‌توانیم صرفاً حاصل جمع خطاها را به حداقل برسانیم، چون حاصل جمع برای شمار زیادی از خطاهای مختلف همیشه صفر است. به جای آن بدنبال خطی می‌گردیم که مجموع مجذورات خطاها را به صفر

برساند- یعنی خطی که $\sum(Y - \hat{Y})^2$ را به حداقل برساند. (در فصل ۶ در باره انحرافات چیز مشابهی گفته شد، آنجا صحبت از میانگین بود، اینجا انحرافات از خط رگرسیون مد نظر است -چیزی شبیه به یک میانگین شناور یا متغیر.)

شرح مشتق شدن معادلات مقادیر بهینه a و b دشوار نیست، اما در اینجا به آن نمی‌پردازیم. زمانی که بخواهیم آنها را طوری استخراج کنیم که مجذور خطاهای پیش‌بینی Y حداقل باشد، کافی است که بسادگی بگوییم

$$b = \frac{SP_{XY}}{SS_X} = \frac{cov_{XY}}{s_X^2}$$

و

$$a = \bar{Y} - b\bar{X} = \frac{\sum Y - b \sum X}{N}$$

(به خاطر داشته باشید که $SP_{XY} = \sum XY - \sum X \sum Y / N$ و $SS_X = \sum X^2 - (\sum X)^2 / N$ است.)

باید توجه داشته باشید که این معادله برای a شامل مقدار b است، بنابراین ابتدا باید آن را برای b حل کنید. همچنین توجه داشته باشید که این معادله برای b شبیه معادله r است به جز اینکه مخرج کسر SS_X به جای $\sqrt{SS_X SS_Y}$ (یا، هم‌ارز آن S_X^2 به جای $S_X S_Y$) است. (این موضوع در باره رابطه بین r و b هنگامی که $S_X = S_Y$ است چه چیزی را گوشزد می‌کند؟)

چنانچه این معادلات را روی داده‌ها اعمال کنیم، خواهیم داشت

$$b = \frac{SP_{XY}}{SS_X} = \frac{15770}{7056.8} = 2.235$$

$$b = \frac{cov_{XY}}{s_X^2} = \frac{543.793}{15.599^2} = \frac{543.793}{243.338} = 2.235 \quad \text{یا}$$

$$a = \bar{Y} - b\bar{X} = 635 - 2.235(28.2) = 571.973$$

اکنون می‌توان نوشت

$$\hat{Y} = 2.235X + 571.973$$

تفسیر این معادله سر راست است. ابتدا به مقدار ثابت توجه نمائید. اگر $X = 0$ باشد (منو خیلی مثبت رتبه‌بندی شده باشد)، مقدار پیش‌بینی شده $\hat{Y}$ تقریباً ۵۷۲ پرس مصرف غذا می‌شود. این احتمالاً یک شکل منطقی از ۵۵۰ دانشجوی است، با احتساب اینکه اغلب افراد موقعی که گرسنه هستند چیزی می‌خورند و برخی افراد هرچه باشد بدون توجه به لذیذ بودن آن می‌خورند. سپس به شیب توجه کنید. در این مثال b برابر ۲/۲۳۵ است. این نشان می‌دهد که برای هر افزایش در رتبه‌بندی، غذاخوری تقریباً ۲/۲ پرس بیشتر غذا ارائه خواهد کرد. به همین ترتیب برای یک افزایش ۱۰ درجه‌ای در رتبه‌بندی، غذاخوری حدود ۲۲ پرس غذا بیشتر ارائه خواهد کرد. بیشتر افراد شیب را به عنوان عدد ثابت در یک معادله ریاضی می‌دانند، اما شیب چیز بیشتری را بیان می‌کند به این صورت که مقدار تفاوت در Y به ازاء یک واحد افزایش در X انتظار دارید.

یک نکته را باید در باره ترسیم واقعی خط رگرسیون مد نظر قرار داد. برای ترسیم این خط شما می‌توانید بسادگی هر دو مقدار X را در نظر بگیرید (ترجیحا در دو سوی یک مقیاس)، $\bar{Y}$ را برای هر یک حساب کنید، این نقاط مختصاتی را روی شکل نشانه‌گذاری کنید، و آنها را با یک خط مستقیم به هم وصل کنید. برای اطمینان از درستی کار، بهتر است سه نقطه را مورد استفاده قرار دهید. برای داده‌های ما یک خط رگرسیون داریم با

$$\bar{Y} = 2/235X + 571/973$$

$$\bar{Y} = 2/235(0) + 571/973 = 571/973 \quad \text{وقتی } X = 0$$

$$\bar{Y} = 2/235(25) + 571/973 = 627/848 \quad \text{وقتی } X = 25$$

$$\bar{Y} = 2/235(50) + 571/973 = 683/723 \quad \text{وقتی } X = 50$$

بنابراین پس از سر راست کردن اعداد، می‌توانیم خطی را از نقاط $(0, 571)$ ، $(25, 628)$ و $(50, 684)$ بگذرانیم. این خط را می‌توان در شکل ۱-۱۰ دید.

لازم به ذکر است که ما خطی را ساختیم تا صرف غذا را از روی رتبه‌بندی پیش‌بینی کنیم و این خط برای به حداقل رساندن مجموع مجزورات انحراف‌های پیش‌بینی صرف غذا، از روی صرف غذای مشاهده شده است. اگر بخواهیم جهت را برگردانیم و رتبه‌بندی را از روی صرف غذا پیش‌بینی کنیم، از این خط استفاده نخواهیم کرد- برای آن هدف طرح نشده است. به جای آن بایستی خطی را پیدا کنیم که مجزورات انحرافات رتبه‌بندی پیش‌بینی شده از روی رتبه‌بندی‌های مشاهده شده را به حداقل برساند. ساده‌ترین راه انجام این کار این است که متغیرهای X و Y را به جای یکدیگر قرار دهیم. آنگاه می‌توانیم از همان فرمول قبلی استفاده کنیم.

۳-۱۰ درستی پیش‌بینی

این واقعیت که ما می‌توانیم یک خط رگرسیون را بر یک مجموعه از داده‌ها بگذرانیم به معنای این نیست که مسائل ما حل شده است. بر عکس، تازه شروع شده‌اند. نکته مهم این نیست که آیا یک خط مستقیم را می‌شود از داده‌ها عبور داد یا نه، (این کار را همیشه می‌توان کرد) بلکه سؤال این است که آیا آن خط یک انطباق منطقی با داده‌ها را نشان می‌دهد یا نه- به عبارت دیگر آیا تلاش ما ارزشمند بوده است. قبل از بحث در باره خطاهای پیش‌بینی، بهتر است توجه شما را به موقعیتی جلب کنیم که مایلیم Y را بدون اطلاع از X پیش‌بینی کنیم.

انحراف معیار به عنوان اندازه‌ای از خطا

همانطور که قبلاً گفته شد، داده‌های جدول ۱-۱۰ تعداد صرف یک غذای (Y) را به عنوان تابعی از کیفیت رتبه‌بندی آن غذا (X) نشان می‌دهد. فرض کنید که اکنون شما باید به پیش‌بینی تعداد غذای مورد نیاز

بپردازید بدون اینکه بدانید چه چیزی را باید ارائه کنید. در این حالت بهترین پیش‌بینی، میانگین صرف غذا ($\bar{Y}$) (متوسط تمام غذاها) خواهد بود، و خطای ضمن پیش‌بینی شما انحراف معیار Y (S_Y) خواهد بود، زیرا پیش‌بینی شما میانگین است و S_Y با انحرافات پیرامون میانگین سر و کار دارد.

$$S_Y = \sqrt{\frac{\sum(Y - \bar{Y})^2}{N - 1}}$$

یا به صورت واریانس،

$$S_Y^2 = \frac{\sum(Y - \bar{Y})^2}{N - 1}$$

صورت کسر مجموع مجذورات انحرافات از $\bar{Y}$ (نقطه‌ای که می‌خواهید پیش‌بینی کنید) است که ما به عنوان مجموع مجذورات Y از آن یاد خواهیم کرد (SS_Y).

خطای معیار برآورد

اکنون فرض کنید مایل هستیم دست به پیش‌بینی در باره صرف غذا برای وعده‌های تازه‌ای بزنیم که متخصصان امور دانشجویی می‌گویند دارای مرتبه ۲۶ است. در این رتبه‌بندی مقدار مربوط به X (و معادله رگرسیون)، را می‌دانیم و بهترین پیش‌بینی ما $\hat{Y}$ خواهد بود (در اینجا $X=26$ و $\hat{Y} = 63.0/0.83$). در راستای اندازه قبلی ما از خطا (انحراف معیار)، خطای همراه با پیش‌بینی حاضر باز هم تابعی از انحرافات Y از نقطه پیش‌بینی شده است، اما در این مورد نقطه پیش‌بینی شده $\hat{Y}$ است نه $\bar{Y}$. اکنون بطور خاص اندازه‌ای از خطا را می‌توان اینگونه تعریف کرد.

$$S_{Y-\hat{Y}} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{N - 2}} = \sqrt{\frac{SS_{Y-\hat{Y}}}{N - 2}}$$

و دوباره مجموع برای مجذور انحرافات از پیش‌بینی ($\hat{Y}$) است. صورت کسر باز هم یک مجموع مجذورات است و اغلب با $SS_{\text{خطا}}$ نشان داده می‌شود زیرا حاصل جمع مجذور خطاهای پیش‌بینی است. عبارت آماری $S_{Y-\hat{Y}}$ را **خطای معیار برآورد** گویند و گاهی با $S_{Y.X}$ نوشته می‌شود تا نشان دهد که انحراف معیار Y پیش‌بینی شده از روی X است. این معروف‌ترین اندازه برای خطای پیش‌بینی است (هرچند همیشه بهترین نیست). مجذور آن $S_{Y-\hat{Y}}^2$ را **واریانس باقیمانده**، یا **خطا** گویند. توجه داشته باشید که $S_{Y-\hat{Y}}^2$ ، $\sum(Y - \bar{Y})^2$ و $SS_{\text{خطا}}$ همه معادل هم هستند. عبارت اولی یک تعریف است، دومی چیز مشابهی را به زبانی دیگر بیان می‌کند، و سومی این موضوع را روشن می‌سازد که ما در باره خطای برآورد بحث می‌کنیم. آماردانان این عبارات را به جای هم به کار می‌برند، و مهم است که شما با همه آنها آشنا شوید و بدانید که همه آنها معنای یکسانی دارند.

جدول ۲- ۱۰ نشان می‌دهد که چگونه خطای معیار برآورد را مستقیماً حساب کنیم. داده‌های خام ستون ۲ و ۳، و مقادیر پیش‌بینی شده $\hat{Y}$ (بدست آمده از $\hat{Y} = 2/235X + 571/973$) در ستون ۴ آمده است. ستون ۵ شامل مقادیر $Y - \hat{Y}$ برای هر مشاهده است. توجه داشته باشید که مجموع آن ستون $(\sum Y - \hat{Y})$ برابر ۰ است، چنانکه قبلاً ذکر شد. ستون ۶ مجذور انحرافات $(\sum (Y - \hat{Y})^2)$ و حاصل جمع آنها را نشان می‌دهد. از این حاصل جمع می‌توانیم حساب کنیم که

$$S_{Y-\hat{Y}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{N - 2}} = \sqrt{\frac{50008.403}{28}} = \sqrt{1786.014} = 42.261$$

پیدا کردن خطای معیار از این راه چندان جذاب نیست. خوشبختانه راه ساده‌تری هست که نه تنها خطای معیار برآورد را بدست می‌دهد بلکه مستقیماً به موضوعات مهم‌تری هم منتهی می‌شود.

جدول ۱۰-۲ محاسبه مستقیم برآوردهای واریانس برای داده‌های جدول ۱۰-۱

منو	رتبه‌بندی (X)	صرف غذا (Y)	$\bar{Y} \bar{Y}$	$(Y - \bar{Y})$	$(Y - \bar{Y})^2$
1	4	550	۵۸۰/۹۱۳	-۳۰/۹۱۳	۹۵۵/۶۱۴
2	4	620	۵۸۰/۹۱۳	۳۹/۰۸۷	۱۵۲۷/۷۹۴
3	5	590	۵۸۳/۱۴۸	۶/۵۸۲	۴۶/۹۵۰
4	8	575	۵۸۹/۸۵۳	-۱۴/۸۵۳	۲۲۰/۶۱۲
5	10	530	۵۹۴/۳۲۳	-۶۴/۳۲۳	۴۱۳۷/۴۴۸
6	10	615	۵۹۴/۳۲۳	۲۰/۶۷۷	۴۲۷/۵۳۸
7	13	590	۶۰۱/۰۲۸	-۱۱/۰۲۸	۱۲۱/۶۱۷
8	15	570			
9	17	630			
10	20	610			
11	20	670			
12	23	690			
13	25	550			
14	28	645			
15	29	585			
16	30	715			
17	30	685			
18	30	715			
19	34	615			
20	34	650			
21	36	690			
22	40	640			
23	40	745			
24	44	610			
25	45	680			
26	45	695			
27	48	665			
28	52	675			
29	53	700			
30	56	665			
۸۴۶	۱۹۰۵۰	۱۹۰۵۰/۰۰۰	۰/۰۰۰	۵۰۰۸/۴۰۳	
۳۰۹۱۴	۱۲۱۸۲۰۰	۱۲۱۳۲۰۰/۰۰۰			

 r^2 و خطای معیار برآورد

ما واریانس باقیمانده، یا خطا، را به این صورت تعریف کردیم

$$S_{Y-\hat{Y}}^2 = \frac{\sum(Y - \hat{Y})^2}{N - 2}$$

و می‌دانیم که

$$\sum(Y - \hat{Y})^2 = S_{Y-\hat{Y}} = SS_{\text{خطا}}$$

بعد از انجام مقدار قابل توجهی دستکاری‌های جبری می‌توان $SS_{\text{خطا}}$ را چنین بیان کرد

$$SS_{\text{خطا}} = SS_Y(1-r^2)$$

که $SS_Y = \sum(Y - \bar{Y})^2$ است. بنابراین مجموع مجذور خطای باقیمانده تابعی از مجموع مجذورات اندازه‌های اصلی Y یعنی SS_Y و همبستگی بین Y و X است. این ممکن است به نظر شما ساده کردن خیلی زیادی نیاید، ولی واقعا چنین است. با اندکی جایگزینی و دستکاری جبری، از اینجا می‌رسیم به

$$S_{Y-\hat{Y}} = S_Y \sqrt{(1-r^2) \left(\frac{N-1}{N-2} \right)}$$

یا، چنانکه بسیاری از کتاب‌ها آن را نشان می‌دهند (هر چند کاملا درست نیست)،

$$S_{Y-\hat{Y}} = S_Y \sqrt{(1-r^2)}$$

این معادله آخر با در نظر گرفتن $(N-1)/(N-2)$ در حد $1/0$ بدست آمده است.

از داده‌های ما می‌توان $S_{Y-\hat{Y}}$ را به سه روش متفاوت حساب کرد.

$$S_{Y-\hat{Y}} = \sqrt{\frac{\sum(Y-\hat{Y})^2}{N-2}} \quad ۱.$$

$$= \sqrt{\frac{50008.403}{28}} \\ = 42.261$$

$$S_{Y-\hat{Y}} = S_Y \sqrt{(1-r^2) \left(\frac{N-1}{N-2} \right)} \quad ۲.$$

$$= 54.219 \sqrt{(1-0.643^2) \left(\frac{29}{28} \right)} \\ = ۴۲/۲۶۱$$

$$S_{Y-\hat{Y}} \simeq S_Y \sqrt{(1-r^2)} \quad ۳.$$

$$= 54.219 \sqrt{(1-.643^2)} \\ = ۴۲/۵۲۴$$

راه حل سوم با دو مورد دیگر فرق دارد چون بر اساس فرمولی است که مقدار ثابت $(N-1)/(N-2)$ را نادیده می‌گیرد. علت استفاده از علامت "تقریباً مساوی" ($\approx$) همین است. برای نمونه‌های بزرگ تفاوت جزئی خواهد بود، اما برای نمونه‌های کوچک چنین نیست.

اکنون که خطای معیار برآورد را محاسبه کردیم، می‌توانیم آن را به عنوان شکلی از انحراف معیار تفسیر کنیم. بنابراین منطقی است که بگوییم انحراف معیار نقاط پیرامون خط رگرسیون $42/261$ است. از آنجا که خط رگرسیون مجموعه پیش‌بینی‌های ما را نشان می‌دهد، یک انحراف معیار در حدود 42 پرس صرف غذا پیرامون این پیش‌بینی‌ها بایستی ما را کمی نگران کند. بی‌توجهی در باره میانگین 40 مورد صرف غذا در پیش‌بینی ما امر خوشایندی نیست. ما ریسک قابل ملاحظه‌ای می‌کنیم چه در باره اضافه آوردن غذا و یا گرسنه ماندن عده زیادی و زخم زبان‌های افراد. ممکن است با این موضوع کمی آرامش بگیرید، که اگر ما یک معادله رگرسیون نداشتیم بهترین حدس ما $\bar{Y}$ می‌بود، و در این حالت خطا به اندازه انحراف معیار Y ($54/219$)؛ نگاه کنید به جدول ۲-۱۰ است، که تقریباً 30% بیشتر از $SS_{Y-\bar{Y}}$ است.

ما دلیل موجهی برای مورد توجه قرار دادن خطاهای بالای پیش‌بینی در تحلیل رگرسیون داریم. بیشتر مردم فکر می‌کنند اگر ما یک معادله رگرسیون داشته باشیم، می‌توانیم پیش‌بینی کنیم و این کل ماجراست. گرچه این موضوع درست است که معادله رگرسیون پیش‌بینی را بهتر از حالتی می‌سازد که پیش‌بینی بدون آن صورت گیرد، اما خطاهای زیادی در این سیستم باقی می‌ماند (و در این حالت همبستگی مقدار قابل توجه $0/643$ می‌بود!). بهتر است که در نظر داشته باشید همه این دانشجویان گرسنه مدیر غذاخوری را مورد حمله قرار می‌دهند. توجه داشته باشید که بیش از حد از یک معادله رگرسیون هیجان زده نشوید.

r^2 به عنوان اندازه‌ای از تغییرپذیری قابل پیش‌بینی

از روی معادله پیش‌بینی که خطای باقیمانده را به شکل r^2 توضیح می‌دهد می‌توان تفسیر بسیار مهمی از ضریب همبستگی حاصل کرد. ما قبلاً آن را دیده‌ایم

$$SS_{Y-\bar{Y}} = SS_{\text{خطا}}(1 - r^2)$$

با بسط و بازنویسی آن، خواهیم داشت

$$SS_{Y-\bar{Y}} - SS_{Y-\bar{Y}}(r^2)_{\text{خطا}} = SS_{\text{خطا}}$$

$$r^2 = \frac{SS_{Y-\bar{Y}} - SS_{\text{خطا}}}{SS_{Y-\bar{Y}}}$$

در این معادله $SS_{Y-\bar{Y}}$ مجموع مجذورات Y و نشان دهنده کل دو بخش زیر است:

۱. بخش مجموع مجذورات Y که با X در رابطه است، یعنی $SS_{Y-\bar{Y}}$ و
۲. بخش مجموع مجذورات Y که مستقل از X است، یعنی $SS_{Y-\bar{Y}}$ یا $SS_{\text{خطا}}$.

در چارچوب مثال قبلی، ما در حال صحبت در باره بخشی از میزان غذایی هستیم که افراد مصرف می‌کنند که به مزه غذا مربوط است و بخشی که به چیزهای دیگر مربوط است- مانند اینکه نهار بد باشد اما افراد گرسنه باشند. $SS_{\text{خطا}}$ آن بخشی از مجموع مجزورات Y است که مستقل از X است و اندازه‌ای از میزان خطاست که حتی بعد از استفاده از X برای پیش‌بینی Y باقی می‌ماند. این مفاهیم با مثالی دیگر بهتر روشن می‌شوند.

فرض کنید ما علاقه‌مند به مطالعه رابطه بین مصرف سیگار (X) و طول عمر (Y) باشیم. همچنان که ما به مشاهده مرگ و میر افراد در طی زمان می‌پردازیم به چند چیز توجه می‌کنیم. نخست اینکه ما می‌بینیم همه دقیقاً در یک سن معین نمی‌میرند- در سن هنگام مرگ مربوط به رفتار سیگار کشیدن تغییرپذیری وجود دارد و این تغییرپذیری با $SS_Y = \sum (Y - \bar{Y})^2$ سنجیده می‌شود. ما همچنین به این واقعیت آشکار توجه داریم که بعضی افراد بیشتر از دیگران سیگار می‌کشند. این تغییرپذیری در سیگار کشیدن بدون توجه به سن هنگام مرگ است و با $SS_X = \sum (X - \bar{X})^2$ سنجیده می‌شود. از این گذشته ما دریافته‌ایم که افرادی که زیاد سیگار می‌کشند زودتر از افرادی که کم مصرف می‌کنند می‌میرند. بنابراین ما یک معادله رگرسیون برای پیش‌بینی Y از روی X می‌نویسیم. از آنجا که مردم در میزان مصرف سیگار متفاوت هستند، مدت طول عمر پیش‌بینی شده نیز تفاوت دارند، و این تغییرپذیری را تغییرات $SS_{\hat{Y}} = \sum (\hat{Y} - \bar{Y})^2 = \sum (\bar{Y} - \bar{Y})^2$ می‌نامیم. این اندازه آخری تغییرپذیری در Y است که مستقیماً به تغییرپذیری در X نسبت داده می‌شود، از این رو اندازه‌های متفاوت $\hat{Y}$ از تفاوت در اندازه‌های X ناشی می‌شود و اندازه‌های یکسان $\hat{Y}$ از اندازه‌های یکسان X بر می‌آید، یعنی $\hat{Y}$ تغییر نمی‌کند مگر اینکه X تغییر کند.

یک منبع تغییرپذیری دیگر هم وجود دارد، و آن تغییرپذیری در طول عمر افرادی است که دقیقاً به اندازه‌ی هم سیگار کشیده‌اند. این منبع با $SS_{\text{خطا}}$ سنجیده می‌شود و تغییرپذیری Y است که بوسیله تغییرات X نشان داده نمی‌شود زیرا این افراد در میزان سیگار کشیدن تفاوت نمی‌کنند. این منابع چندگانه تغییرپذیری (یعنی مجموع مجزورات) در جدول ۳-۱۰ خلاصه شده‌اند.

جدول ۳-۱۰ منابع واریانس در رگرسیون

$$SS_X = \text{تغییرپذیری در میزان مصرف سیگار} = \sum (X - \bar{X})^2$$

$$SS_Y = \text{تغییرپذیری در طول عمر} = \sum (Y - \bar{Y})^2$$

$$SS_{\hat{Y}} = \text{تغییرپذیری در طول عمر که مستقیماً به تغییرات سیگار کشیدن نسبت داده می‌شود} = \sum (\hat{Y} - \bar{Y})^2$$

$$SS_{\text{خطا}} = \text{تغییرپذیری در طول عمر که قابل استناد به سیگار کشیدن نیست} = \sum (Y - \hat{Y})^2$$

$$SS_Y - SS_{\hat{Y}}$$

اگر این حد نهایی را در نظر بگیریم که در آن همه افراد غیرسیگاری دقیقا در ۷۲ سالگی بمیرند و همه افرادی که به اندازه هم سیگار کشیده‌اند دقیقا در ۶۸ سالگی بمیرند، آنگاه تمام تغییرپذیری طول عمر مستقیما از روی تغییرپذیری رفتار سیگار کشیدن قابل پیش‌بینی است. اگر شما سیگار می‌کشید در ۶۸ سالگی و اگر نمی‌کشید در ۷۲ سالگی خواهید مرد. در اینجا $SS_Y = SS_P$ و $SS_{\text{خطا}} = 0$ است.

در یک مثال واقعی‌تر سیگاری‌ها متمایل به مرگ زودتر از غیرسیگاری‌ها هستند، اما در درون هر گروه میران معینی از تغییرپذیری در طول عمر وجود دارد. این موقعیتی است که در آن مقداری از SS_Y به قابل نسبت دادن سیگار کشیدن است (SS_P) و مقداری نیست ($SS_{\text{خطا}}$). کاری که مخصوصا ما می‌خواهیم قادر به انجام آن باشیم، تعیین درصد تغییرپذیری کلی طول عمر است که قابل استناد به رفتار سیگار کشیدن باشد. به عبارت دیگر ما اندازه‌ای را می‌خواهیم که با این فرمول نشان داده می‌شود

$$\frac{SS_P}{SS_Y} = \frac{SS_Y - SS_{\text{خطا}}}{SS_Y}$$

چنانکه قبلا دیدیم این اندازه، r^2 است. به عبارت دیگر

$$r^2 = \frac{SS_P}{SS_Y}$$

این تفسیر r^2 بسیار سودمند است. مثلا اگر همبستگی بین مدت سیگار کشیدن و طول عمر در حد غیر واقعینانه‌ای ۰/۸۰ بالا باشد، می‌توانیم بگوییم که $0.64 = 0.8^2$ از تغییرپذیری طول عمر مستقیما از طریق تغییرپذیری در رفتار سیگار کشیدن قابل پیش‌بینی است. آشکارا این میزان نسبت به دنیای واقعی بسیار مبالغه آمیز است. اگر همبستگی یک اندازه محتمل‌تر $r = 0.20$ بود، می‌گفتیم $0.04 = 0.2^2$ تغییرپذیری در طول عمر با رفتار سیگار کشیدن در رابطه است، در حالی که ۹۶٪ به عوامل دیگر مربوط است.

یکی از مسائل جنبی تمرکز بر مجذور ضریب همبستگی به فهم مناسبی از آن بر می‌گردد. اگر درست باشد که سیگار کشیدن مسئول ۴٪ تغییرات طول عمر است، ممکن است وسوسه شویم که سیگار کشیدن سهم کمی در طول عمر دارد. اما در نظر داشته باشید که شمار زیادی از متغیرها هستند که بر طول عمر اثر می‌گذارند، و یک میزان ۴٪ از این تغییرپذیری احتمالا یکی از پیش‌بینی کننده‌های اصلی است. در واقع گفته شده که ۳۰٪ از سرطان‌ها در ایالات متحده بدلیل سیگار است. متغیری که مسئول ۴٪ از تغییرات نمرات درسی باشد احتمالا جزئی است. اما نمی‌توان به سادگی چیزی که ۴٪ از تغییرپذیری طول عمر به حساب می‌آید نادیده گرفت.

لازم است توجه داشته باشیم که وقتی ما از عبارت‌های “به حساب... می‌آید”، “قابل استناد به”، “قابل پیش‌بینی از”، “تداعی شده با” استفاده می‌کنیم، نباید به عنوان عبارت علت و معلولی در نظر گرفته شوند. بنابراین شما می‌توانید بگویید که آن درد کتف شما ۱۰٪ به تغییرپذیری در وضعیت آب و هوا بر می‌گردد بدون اینکه منظورتان این باشد که ناراحتی کتف شما علت باران است، یا حتی خود باران علت

درد کتف است. بطور مثال ممکن است کتف شما موقعی که باران می‌آمده آسیب دیده باشد چون چتر بر دوشتان سنگینی کرده است.

۴-۱۰ رگرسیون چندگانه

در این فصل ما با موردی سرو کار داشتیم که در آن تنها یک متغیر پیش‌بین داشتیم. در این شرایط ما یک معادله رگرسیون از نوع $\hat{Y} = bX + a$ داریم. اما دلیلی ندارد که خود را به یک پیش‌بینی کننده محدود کنیم. جا دارد که بهر سیستم ترکیبی از دو، سه، یا چهار پیش‌بینی کننده چگونه متغیر ملاک را پیش‌بینی خواهند کرد. در اینجا با سه پیش‌بینی کننده ما معادله‌ای به این شکل خواهیم داشت

$$\hat{Y} = b_1W + b_2X + b_3Z + a$$

که b_1 ، b_2 و b_3 ضرایب شیب به ترتیب برای پیش‌بینی کننده‌های W ، X و Z هستند. a نیز همان مقدار ثابت است. انجام راه‌حل‌های رگرسیون چندگانه با دست بخصوص برای بیش از دو متغیر پیش‌بین معمولاً خیلی غیرمرسوم است، این معادلات را می‌توان با برنامه‌های کامپیوتری متنوعی محاسبه کرد.

من [هاول] چندین سال پیش داده‌هایی را پیرامون ثبت نام در دوره‌های تحصیلات تکمیلی جمع‌آوری کردم. چند صد داوطلب توسط تمام استادان روی یک مقیاس ۷ درجه‌ای از ۱ تا ۷ رتبه‌بندی شدند که ۱ معادل "رد کردن بی‌درنگ" و ۷ معادل "پذیرش بی‌درنگ" بود. محقق سعی کرد برای یک نمونه ۱۰۰ نفری از داوطلبان میانگین رتبه‌بندی را برای هر داوطلب از روی نمره مصاحبه، نمره ترکیبی توصیه‌نامه‌ها، و نمره هدفمندی پیش‌بینی کند. معادله رگرسیون بدست آمده چنین بود

$$\hat{Y} = 0.09X_1 + 0.51X_2 + 0.43X_3 - 1.87$$

و ضریب رگرسیون چندگانه^۱ (R) ۰/۷۷۵ بود. بنابراین مجذور ضریب همبستگی (R^2) ۰/۶۰ بود. تفسیر R^2 همانند r^2 با یک پیش‌بینی کننده است. به عبارت دیگر ۶۰٪ تغییرپذیری در رتبه‌بندی را می‌توان به حساب تغییرپذیری سه متغیر پیش‌بین مورد نظر با هم دانست. یعنی با استفاده همزمان از نمره مصاحبه، نمره ترکیبی توصیه‌نامه‌ها، و نمره هدفمندی، به عنوان پیش‌بینی کننده، می‌توانیم ۶۰٪ تغییرپذیری در رتبه‌بندی را به حساب آوریم.

معادله رگرسیون چندگانه (چند متغیری) عمدتاً به همان روشی که برای یک پیش‌بینی کننده است تفسیر می‌شود. هر واحد تغییر در نمره مصاحبه، افزایشی معادل ۰/۰۰۹ واحد در رتبه‌بندی خواهد داشت با این فرض که دو پیش‌بینی کننده دیگر ثابت باشند. به همین صورت، هر واحد تغییر در توصیه نامه‌ها، تغییری معادل ۰/۵۱ در رتبه‌بندی با ثابت گرفتن دو متغیر دیگر خواهد داشت.

احتمالاً شما اکنون مایل به این هستید که از معادله رگرسیون نتیجه‌گیری کنید که نمرات توصیه‌نامه‌ها و هدفمندی مهم‌تر از مصاحبه هستند زیرا ضریب رگرسیون (شیب) بیشتری دارند. اما

1. multiple correlation coefficient

این متغیرها در واحدهای بکلی متفاوتی اندازه‌گیری شده‌اند (دامنه نمرات مصاحبه از ۲۰۰ تا ۸۰۰ است، در حالی که دامنه نمرات دو متغیر دیگر از ۱/۰ تا ۳/۰ است). مقایسه بر اساس مقدار ضرایب درست نیست. تصمیم‌گیری در باره اهمیت نسبی پیش‌بینی‌کننده‌ها کار دشواری است (و برای پیش‌بینی‌کننده‌هایی که خودشان با یکدیگر خیلی همبسته هستند حتی ممکن است بی معنا باشد).

مطلب مهم‌تر در مسأله رگرسیون درجه همبستگی بین خود پیش‌بینی‌کننده‌ها است. وقتی پیش‌بینی‌کننده‌ها با یکدیگر همبستگی زیادی دارند (شرایطی که به **چندهم خطی بودن**^۱ معروف است)، معادله رگرسیون از یک نمونه به نمونه دیگر خیلی ناپایدار است. به بیان دیگر دو نمونه تصادفی از یک جمعیت ممکن است معادلات رگرسیون متفاوتی را ایجاد کنند که به کلی متفاوت از هم به نظر برسند. بنابراین با جدیت توصیه می‌شود از پیش‌بینی‌کننده‌هایی که همبستگی زیادی با هم دارند پرهیز شود و حتی در صورت امکان از آنهایی هم که همبستگی متوسطی دارند خودداری گردد. رگرسیون چندگانه شیوه‌ای بسیار سودمند است، اما کاربرد و تفسیر آن ساده نیست.

۵-۱۰ مثال کار شده پایانی

در فصل ۹ ما ضریب همبستگی را برای رابطه بین رتبه‌بندی کیفیت یک درس و دشواری آن درس (که بر اساس متوسط نمره مورد انتظار دانشجویان در آن درس نشان داده می‌شود) بدست آوردیم. این داده‌ها در اینجا بار دیگر تکرار شده تا یادآوری آسان باشد. ما در این فصل این مثال را برای معادله رگرسیون به منظور پیش‌بینی رتبه‌بندی کیفیت کلی (Y) از روی نمره مورد انتظار (X) حل می‌کنیم. سپس تفسیر این ضریب را در معادله مد نظر قرار خواهیم داد.

کیفیت کلی (Y)	نمره مورد انتظار (X)
۳/۴	۳/۵
۲/۹	۳/۲
۲/۶	۲/۸
۳/۸	۳/۳
۳/۰	۳/۲
۲/۵	۳/۲
۳/۹	۳/۶
۴/۳	۴/۰

۳/۰	۳/۸
۳/۱	۳/۴
۳/۰	۲/۸
۳/۳	۲/۹
۳/۲	۴/۱
۳/۴	۲/۷
۳/۷	۳/۹
$\sum X = 49/5$	$\sum Y = 50/0$
$\sum X^2 = 164/65$	$\sum Y^2 = 171/68$
$\sum XY = 166/48$	

گام اول محاسبه میانگین و واریانس هر متغیر است. اینها در زیر نشان داده شده‌اند:

$$\bar{X} = 49.5/15 = 3.3000$$

$$\bar{Y} = 50.0/15 = 3.333$$

$$S_X = \sqrt{\frac{164.65 - 49.5^2/15}{14}}$$

$$S_Y = \sqrt{\frac{171.68 - 50.0^2/15}{14}}$$

$$= 0.3047$$

$$S_X^2 = 0.3047^2 = 0.0928$$

$$= 0.5984$$

$$S_Y^2 = 0.5984^2 = 0.3581$$

کوواریانس به این صورت است

$$\text{COV}_{XY} = \frac{164.65 - (49.5)(50.0)/15}{14} = 0.1057$$

برای محاسبه شیب داریم

$$b = \frac{\text{COV}_{XY}}{S_X^2} = \frac{0.1057}{0.0928} = 1.1385$$

مقدار ثابت را به صورت زیر حساب می‌کنیم

$$a = \bar{Y} - b(\bar{X}) = 3.333 - (1.1385)(3.300) = -0.4238$$

آنگاه معادله ما می‌شود

$$\hat{Y} = (1.1385)(X) - 0.4238$$

ما می‌توانیم نتایج خود را به صورت زیر تفسیر کنیم. اگر ما درسی داشته باشیم که نمره مورد انتظار دانشجوی در آن * باشد، بهترین حدس از رتبه‌بندی که انتظار داریم ۰/۴۲- خواهد بود. این آماره تا زمانی که تفسیر آن مورد توجه قرار نگیرد چندان بامعنا نیست زیرا دشوار است که درسی را تصور کنیم که هر کسی انتظار شکست داشته باشد، و همچنین رتبه‌های درسی زیر ۰/۰ امکان ندارد. در این مورد مقدار ثابت تنها نقش لنگر برای خط رگرسیون دارد.

یک شیب $1/1385$ را می‌توان این‌گونه تفسیر کرد که اگر دو درس از لحاظ نمرات مورد انتظار، با هم یک نمره فرق داشته باشند، انتظار می‌رود رتبه‌بندی کلی کیفیت آنها، کمی بیش از یک نمره فرق کند. آنگاه انتظار چنین تفاوتی بین دو درس خواهد بود که نمره دانشجویان در یکی یک نمره ج پیش‌بین می‌شود و در دیگری که در آن یک نمره ب پیش‌بین می‌شود. تذکر فصل ۹ مبنی بر این که نباید از عبارات علت و معلولی استفاده کنیم را در نظر داشته باشید. ما دلیل خاصی نداریم تا نتیجه‌گیری کنیم علت رتبه‌بندی پایین انتظار نمرات پایین است.

۱۰-۶ خلاصه

در این فصل دانسته‌های خود را در باره رابطه بین دو متغیر توسعه دادیم تا در بر گیرنده شرایطی باشد که بخواهیم یک متغیر را از روی اطلاعات خود در باره دیگری پیش‌بینی کنیم. ما دیدیم که چگونه معادله خط رگرسیون را محاسبه کنیم و خطای استاندارد ($S_y - \hat{y}$) را به عنوان اندازه‌ای از درستی پیش‌بینی در نظر گرفتیم. ما دیدیم که r^2 یک شاخص خیلی مهم برای درصد تغییرپذیری در یک متغیر است که می‌توان به حساب متغیر دیگر گذاشت. استفاده از r^2 به جای r به ضریب همبستگی معنای مضاعفی می‌بخشد، هر چند که باید با احتیاط تفسیر شود. در آخر ما مبحث رگرسیون چندگانه را به اختصار توضیح دادیم. برخی از مطالب مهم فصل عبارتند از:

- رگرسیون
- خط رگرسیون
- شیب
- ثابت رگرسیون
- خطاهای پیش‌بینی
- خطای معیار برآورد
- واریانس باقیمانده
- ضریب همبستگی چندگانه (R)
- ضرایب رگرسیون
- چندهم‌خطی

۱۰-۷ تمرین‌ها

۱۰-۱ از داده‌های تمرین ۹-۱ معادله رگرسیون را برای پیش‌بینی درصد تولد نوزادان با وزن کمتر از ۲۵۰۰ گرم (Y) براساس نرخ باروری زنانی که در سنین $17 \leq X_1$ یا $35 \geq X_1$ هستند حساب کنید. X_1 به عنوان "نرخ باروری با خطر بالا" تعریف می‌شود.

۱۰-۲ خطای معیار برآورد را برای معادله رگرسیون تمرین ۱۰-۱ حساب کنید.

۱۰-۳ اگر به عنوان یک نتیجه تغییر تدریجی نقش زنان در جامعه، ما تغییری در سن بچه‌داری مشاهده کنیم طوری که نرخ باروری پرخطر در تمرین ۹-۱ به ۷۰ برسد، برای تولد نوزادان کم وزن چه چیزی را پیش‌بینی خواهیم کرد؟

۱۰-۴ چرا در پیش‌بینی تمرین ۱۰-۳ برای نرخ ۷۰ ما راحت نیستیم؟

- ۱۰-۵ یک معادله رگرسیون برای پیش‌بینی نمره نهایی در تمرین ۹-۹ حساب کنید.
- ۱۰-۶ یک معادله رگرسیون برای پیش‌بینی تعداد مسائل تکمیل شده از نمره نهایی در تمرین ۹-۹ حساب کنید.
- ۱۰-۷ با استفاده از داده‌های جدول ۱۰-۱ تعداد غذایی که برای رتبه ۲۱ ارائه می‌شود را پیش‌بینی کنید.
- ۱۰-۸ میانگین رتبه‌بندی برای داده‌های جدول ۱۰-۱ معادل $28/2$ است. پیش‌بینی شما برای ارائه غذا برای رتبه‌بندی $28/2$ چه خواهد بود؟ چگونه آن را با $\bar{Y}$ مقایسه می‌کنید؟
- ۱۰-۹ با استفاده از داده‌های تمرین ۹-۱۰ یک معادله رگرسیون مناسب برای پیش‌بینی کاهش نمرات ریاضی SAT از ۱۹۷۸ تا ۱۹۷۹ بر اساس سطح یاد ۱۳۱ طرح کنید.
- ۱۰-۱۰ با توجه به تمرین قبل، یاد ۱۳۱ را نادیده بگیرید و معادله‌ای برای پیش‌بینی یک میانگین نمره ایالتی ریاضی SAT در سال ۱۹۷۸ از روی نمرات آزمون کلامی بنویسد.
- ۱۰-۱۱ نتایج جواب تمرین ۱۰-۱۰ را مورد تفسیر قرار دهید.
- ۱۰-۱۲ از هر نمره کلامی ۱۹۷۸ در تمرین ۱۰-۱۰ به مقدار ۱۰۰ نمره کم کنید. سپس همبستگی بین MYA و هر دو ۷۷۸ قدیم و جدید را حساب کنید.
- ۱۰-۱۳ برای داده‌های جدول ۱۰-۱، $\bar{Y}$ و $Y - \bar{Y}$ را تعیین کنید.
- ۱۰-۱۴ در یک گروه ۲۰۰ نفری از اعضای هیأت علمی دانشگاه که برای پانزده سال در آنجا شناخته شده بودند معادله رابطه حقوق (به هزار دلار) با سال‌های خدمت $\bar{Y} = 0/9(X) + 15$ است. برای ۱۰۰ نفر کارکنان اداری همین دانشگاه این معادل $\bar{Y} = 1/5(X) + 10$ است. با فرض اینکه تمام تفاوت‌ها معنادار هستند، معنای این دو معادله را تفسیر کنید. چند سال باید بگذرد تا یک نفر کارمند اداری و یک عضو هیأت علمی به درآمد تقریب یکسان برسند؟

فصل ۱۱

توزیع‌های نمونه‌گیری و فرضیه‌آزمایی

۱۱-۱- دو مثال سده شامل مشکلات رفتاری و رانندگان گستاخ

۱۱-۲- توزیع‌های نمونه‌گیری

۱۱-۳- فرضیه‌آزمایی

۱۱-۴- فرض صفر

۱۱-۵- آزمون آماری و توزیع‌های نمونه‌گیری آنها

۱۱-۶- استفاده از توزیع بهنجار برای آزمون فرضیه‌ها

۱۱-۷- خطای نوع I و خطای نوع II

۱۱-۸- آزمون‌های یک دامنه و دو دامنه

۱۱-۹- مثال کار شده پایانی

در چند فصل گذشته ما شماری از آماره‌های مختلف را بررسی کردیم و با کاربرد آنها برای توصیف یک مجموعه از داده‌ها، نمایش احتمال وقوع بعضی پدیده‌ها، یا سنجش رابطه بین دو متغیر آشنا شدیم. هرچند توصیف داده‌ها مهم و اساس هر نوع تحلیلی است اما برای پاسخ بسیاری از مسائل جالبی که با آنها برخورد می‌کنیم کافی نیست. در یک آزمایش نوعی، ممکن است ما یک گروه را در معرض یک عملکرد آزمایشی قرار دهیم تا ببینیم آیا نمرات آنها با نمرات افراد درکل فرق دارد. یا ممکن است روی یک گروه، کاری انجام دهیم و گروه دیگری را به عنوان گروه گواه داشته باشیم که روی آنها کار خاصی صورت نگیرد و بخواهیم میانگین‌های دو گروه را در برخی متغیرها مقایسه کنیم. آماره‌های توصیفی به ما چیزی در این باره نمی‌گویند که مثلاً تفاوت بین یک میانگین نمونه و یک میانگین فرضی جمعیت یا

تفاوت بین دو میانگین نمونه‌ای بدست آمده آنقدر کوچک است که تنها به عامل شانس نسبت داده شود یا اینکه یک تفاوت واقعی را نشان می‌دهد که بتوان آن را به اثر عملکرد آزمایشی نسبت داد. آماردانان به کرات از عبارت‌هایی مانند "تفاوت ناشی از شانس" یا "خطای نمونه‌گیری" استفاده می‌کنند و فرض می‌گیرند که شما منظور آنان را می‌دانید. احتمالاً شما می‌دانید اما اگر ندانید شما در فهم ادامه مطالب این کتاب با مشکل روبرو خواهید شد مگر اینکه دقایقی را صرف روشن نمودن این مفاهیم کنیم. با یک مثال ساده شروع می‌کنیم. در فصل ۷ ما توزیع نمرات مشکلات کلی رفتار را از فرم خودگزارشی جوانان آکنباخ مد نظر قرار دادیم. نمرات مشکلات کلی رفتار در جمعیت به صورت بهنجار (نرمال) توزیع شده‌اند (یعنی جمعیت کامل چنین نمراتی به صورت نرمال توزیع شده است) با یک میانگین جمعیتی (μ) ۵۰ و یک انحراف معیار (σ) معادل ۱۰. ما می‌دانیم که کودکان، سطوح مختلفی از مشکلات رفتاری را نشان می‌دهند که اگر نمونه‌ای از کودکان گرفته شود میانگین نمونه احتمالاً دقیقاً برابر ۵۰ نخواهد بود. یک نمونه از کودکان ممکن است دارای میانگین ۴۹ باشد درحالی که نمونه دیگر میانگین ۵۱/۳ داشته باشد. میانگین نمونه در عمل بستگی به کودکان خاصی دارد که در آن قرار گرفته‌اند. این تغییرپذیری که انتظار داریم از نمونه‌ای به نمونه دیگر ببینیم همان چیزی است که گاهی از آن به عنوان "تغییرپذیری ناشی از شانس" سخن می‌گوییم. ما در حال اشاره به این واقعیت هستیم که آمارها (در اینجا میانگین‌ها) حاصل از نمونه‌ها بطور طبیعی از یک نمونه به نمونه دیگر فرق می‌کنند.

در همین راستا عبارت **خطای نمونه‌گیری**^۱ اغلب در این کتاب مترادف با تغییرپذیری ناشی از شانس به کار رفته است. این نشان می‌دهد که مقدار یک نمونه آماری به عنوان نتیجه مشاهدات خاصی که در نمونه قرار گرفته‌اند احتمال دارد اشتباه باشد (یعنی از پارامتر جمعیت که آن را برآورد می‌کند انحراف داشته باشد). در این زمینه "خطا" گویای بی‌دقتی یا اشتباه نیست. در مورد مشکلات رفتاری، یک نمونه تصادفی ممکن است به گونه‌ای باشد که شامل یک کودک خرابکار باشد در حالی که نمونه‌ای دیگر ممکن است شامل شمار غیرعادی از کودکان نسبتاً خوش رفتار شود.

۱۱-۱ دو مثال ساده شامل مشکلات رفتاری و رانندگان گستاخ

یکی از مقیاس‌های سودمند ضمیمه فرم خودگزارشی جوانان نمره مشکلات رفتاری است که برای یک جمعیت از کودکان عادی دارای میانگین ۵۰ و انحراف معیار ۱۰ می‌باشد. فرض کنید ما یک نمونه متشکل از ۵ کودک داریم که می‌دانیم بدلیل طلاق والدین، در خانواده تحت فشار روانی زیادی بوده‌اند، و می‌خواهیم این نمونه را مطالعه کنیم تا ببینیم اگر کودکانی که تحت فشار روانی بالایی هستند سطح بیشتری از مشکلات رفتاری را هم نشان می‌دهند. همچنین فرض کنید این پنج کودک فرم خودگزارشی جوانان را تکمیل کرده‌اند، و میانگین برای این نمونه به ۵۶ رسیده که ۶ نمره بالاتر از میانگین جمعیت

1. sampling error

کودکان عادی است. چیزی که در واقع ما می‌خواهیم بدانیم این است که آیا تفاوت بین ۵۰ و ۵۶ نمره آن قدر کوچک است که قابل نسبت دادن به تغییرات شانس باشد که اگر نمرات از یک نمونه عادی از کودکان گرفته شده باشد بطور عادی انتظارش را داریم، یا اینکه تفاوت به اندازه کافی بزرگ است که ما را به این نتیجه برساند که جمعیت کودکانی که از آنها نمونه‌گیری شده دارای میانگینی بزرگتر از میانگین جمعیت کودکان عادی می‌باشد. به بیان دیگر ما می‌خواهیم این فرضیه را بیازماییم که میانگین جمعیتی (μ) که از آن نمونه‌گیری شده ۵۰ است برخلاف فرض مقابل که ادعا می‌کند μ بزرگتر از ۵۰ است. پاسخ به این سؤال خاص در فصل بعد با جزئیات بیشتر شرح داده خواهد شد. در اینجا ما فقط به قضایای کلی مربوط به پاسخ به چنین رویکردی می‌پردازیم.

مثال دوم از تحقیق دوب و گراس (۱۹۶۸) است که تأثیر وضعیت اجتماعی ادراک شده را مطالعه کردند. آنها دریافتند که اگر در پشت چراغ راهنمایی هنگامی که چراغ سبز می‌شود یک اتومبیل کهنه (کلاس پایین) نتواند استارت بزند، در ۸۴٪ موارد راننده اتومبیل پشتی بوق می‌کشد. اما وقتی اتومبیل متوقف شده یک ماشین گران قیمت، کلاس بالا باشد فقط در ۵۰٪ موارد راننده ماشین پشتی بوق می‌کشد. این نتایج را به دو شیوه می‌توان توضیح داد.

یا

۱. تفاوت بین ۸۴٪ در یک نمونه و ۵۰٪ در یک نمونه دیگر به خطای نمونه‌گیری (تغییرپذیری

تصادفی میان نمونه‌ها) نسبت داده شود و بنابراین نمی‌توانیم نتیجه‌گیری کنیم که وضعیت

اجتماعی ادراک شده بر رفتار بوق زدن تأثیر دارد؛ یا

۲. تفاوت بین ۸۴٪ و ۵۰٪ بزرگ است، تفاوتی که نتوان به خطای نمونه‌گیری (تغییرپذیری

تصادفی میان نمونه‌ها) نسبت داد، و بنابراین مردم کمتر مایل هستند برای رانندگان کلاس

بالا بوق بکشند.

هر چند محاسبات آماری لازم برای جواب دادن به این پرسش با آنچه برای پرسش‌های قبلی به کار رفته متفاوت است (چون موارد قبلی با میانگین‌ها سروکار داشتند و این مورد با نسبت‌ها)، اما منطق زیربنایی یکی است.

مثال‌های مشکلات رفتاری و رفتار بوق کشیدن دو نوع از پرسش‌هایی است که تحت عنوان

فرضیه آزمایی قرار می‌گیرند. این فصل بدنال ارائه نظریه فرضیه آزمایی به شکلی فراگیر است، بدون

اینکه درگیر فنون خاص یا ویژگی‌های آزمون آماری خاصی شود.

نظریه فرضیه آزمایی از چنان اهمیتی برخوردار است که دنبال کردن و فهم کلی آن امری اساسی

است. بسیاری دانشجویان که یک یا دو درس آمار را گذرانده‌اند و چگونگی اجرای شماری از آزمون‌های آماری را می‌دانند، هنوز آگاهی درستی از آنچه انجام می‌دهند ندارند. در نتیجه در تفسیر جداول آماری

مشکل دارند و هر شیوه جدید را بایستی از اول - قدم به قدم به شکلی ابتدائی یاد بگیرند. این فصل برای این طرح شده که بر این مشکل به گونه‌ای کلی بدون استفاده از هیچ فرمولی چیره شویم. فرمول‌ها را بعداً فرا خواهید گرفت پس از آنکه دلیل استفاده از آنها را دانستید. آماردانان حرفه‌ای ممکن است به نبودن تعریف‌ها ایراد بگیرند، اما این کاستی در فصل‌های بعدی جبران می‌شود. دیگران ممکن است به این موضوع توجه کنند که ما فرضیه آزمایی می‌کنیم پیش از آنکه به این شیوه‌های آماری بپردازیم. هدف دقیقاً همین است. مطالبی که در اینجا ارائه می‌شود در همه آزمون‌های آماری کاربرد دارد و مستقل از آنها قابل بحث است. با جدا سازی مطالب به این شیوه، شما براحتی می‌توانید روی اصول تمرکز کنید بدون اینکه نگران راهکارهای محاسباتی باشید.

اهمیت موضوع در فرضیه آزمایی یافتن شیوه‌ای برای تصمیم‌گیری در باره این است که آیا در مثال مشکلات رفتاری، ما به نوسانات کوچک شانس در اطراف میانگین ۵۰ نگاه کنیم، یا اینکه انحراف از ۵۰ به اندازه کافی بزرگ هست که بگوییم کودکان مضطرب در مجموع دارای مشکلات رفتاری بیشتری نسبت به کودکان عادی هستند. برای پاسخ به چنین پرسش‌هایی ما مجبور به استفاده از چیزی هستیم که به **توزیع‌های نمونه‌گیری** معروف است، که به ویژه به ما می‌گوید چه میزان از تغییرپذیری از نمونه‌ای به نمونه دیگر را می‌توانیم به حساب شانس ناشی از خطای نمونه‌گیری بگذاریم.

۲-۱۱ توزیع‌های نمونه‌گیری

اساسی‌ترین مفهوم زیربنایی در همه آزمون‌های آماری توزیع نمونه‌گیری آن آماره است. بی‌جا نیست اگر گفته شود چنانچه توزیع نمونه‌گیری نداشته باشیم، هیچگونه آزمون آماری نخواهیم داشت. به بیان روشن‌تر، توزیع‌های نمونه‌گیری به ما می‌گویند چه مقادیری را می‌توانیم (یا نمی‌توانیم) برای یک آماره معین تحت یک مجموعه شرایط از قبل تعیین شده انتظار داشته باشیم (مثلاً میانگین مشاهده شده از پنج کودک ممکن است چقدر باشد اگر میانگین حقیقی جمعیتی که این کودکان از آن برگرفته شده‌اند ۵۰/۰ باشد). به این ترتیب، توزیع‌های نمونه‌گیری فرصتی را فراهم می‌کنند برای ارزیابی احتمال (با داشتن مقدار آماره نمونه) این که آیا چنین شرایط از قبل تعیین شده‌ای واقعا وجود دارد یا نه.

اساساً توزیع نمونه‌گیری یک آماره را می‌توان به عنوان توزیع مقادیر بدست آمده برای آن آماره از نمونه‌گیری مکرر (یعنی تکرار اجرای آزمایش یا استخراج نمونه‌ها به تعداد نامحدود) در نظر گرفت. هر چند توزیع‌های نمونه‌گیری تقریباً همیشه به طور حسابی مشتق می‌شوند، برای درک آسان‌تر اینکه چه چیزی را بازنمایی می‌کنند تصور کنید به لحاظ نظری اگر نمونه‌ها بطور تجربی با یک نمونه آزمایشی ساده استخراج شوند آنها چگونه خواهند بود.

ما **توزیع نمونه‌گیری از میانگین** را به عنوان نمایشی از آن در نظر می‌گیریم، چون درک آن از همه ساده‌تر است و مستقیماً با مثال مشکلات رفتاری مربوط می‌شود. توزیع نمونه‌گیری میانگین چیزی

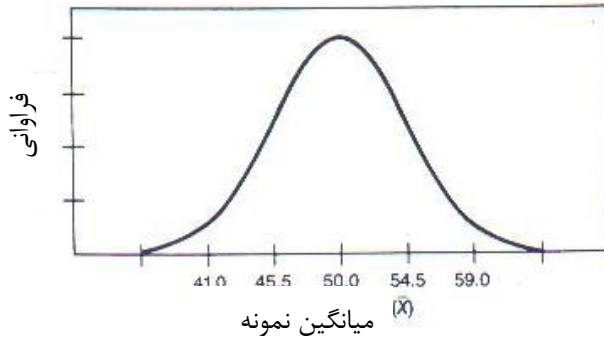
نیست جز توزیع نمونه گیری از میانگین های شمار نامحدودی از نمونه های تصادفی استخراج شده تحت شرایط معین (مثلا تحت شرایطی که میانگین حقیقی جامعه ما ۵۰ و انحراف معیار ۱۰ است). فرض کنید ما جمعیتی با میانگین شناخته شده داریم ($\mu = 50$). همچنین فرض کنید ما تعداد خیلی زیادی (به لحاظ نظری نامحدود) از نمونه های تصادفی از این جمعیت گرفته ایم که هر نمونه شامل پنج نمره است. برای هر نمونه میانگین آن را حساب می کنیم، و با استخراج تمام نمونه ها، توزیع این میانگین ها را رسم می کنیم. چنین توزیعی یک توزیع نمونه گیری از میانگین ها خواهد بود شبیه به آنچه در شکل ۱۱-۱ مشاهده می کنید. در این شکل مشاهده می کنید که وقتی ما پنج کودک را به طور تصادفی انتخاب کنیم، برای مثال میانگین های نمونه ای بین ۴۸ و ۵۲، کاملاً احتمال رخ دادن دارند. همچنین می توان دید که بسیار نامحتمل است که ما از این جمعیت یک نمونه پنج نفری بگیریم که دارای میانگینی به بزرگی ۷۰ باشد، گرچه باز هم احتمال (خیلی کمی) وجود دارد. این واقعیت که ما انواع مقادیر مورد انتظار برای میانگین یک نمونه برآمده از این جمعیت را می دانیم حالا به ما امکان می دهد که به سؤال اصلی برگردیم و بپرسیم آیا یک میانگین نمونه ای بدست آمده می تواند گواهی بر این فرضیه باشد که ما واقعا از این جامعه نمونه برداری کرده ایم.

۳-۱۱ فرضیه آزمایی

دلیل اینکه ما خواه به صورت محاسباتی و یا تجربی بدنبال بدست آوردن توزیع های نمونه ای هستیم صرفاً جالب بودن آنها نیست. ما دلایل مهمی برای این کار داریم. دلیل ساده آن این است که می خواهیم بعضی فرضیه ها را بیازماییم. ما دریافته ایم که یک نمونه تصادفی از پنج کودک پر اضطراب دارای میانگین مشکلات رفتاری ۵۶ بودند. می خواهیم این فرضیه را بیازماییم که آیا چنین میانگین نمونه ای می تواند از جمعیتی با $\mu = 50$ برآمده باشد. این راه دیگری برای بیان این منظور است که ما می خواهیم بدانیم آیا میانگین کودکان مضطرب با میانگین کودکان عادی فرق دارد. تنها راهی که می توانیم چنین فرضیه ای را آزمایش کنیم این است که ایده هایی در باره احتمال بدست آوردن یک میانگین نمونه ای ۵۶ داشته باشیم وقتی که ما مشاهدات نمونه ای واقعی از جمعیتی که میانگین آن $\mu = 50$ است به عمل آوریم. پاسخ به این پرسش دقیقاً چیزی است که یک توزیع نمونه گیری برای آن طرح شده است.

فرض کنید ما یک توزیع نمونه گیری متشکل از میانگین نمونه های پنج تایی کودکان از جمعیتی با میانگین $\mu = 50$ بدست آورده ایم. (این در واقع توزیع نمایش داده شده در شکل ۱۱-۱ است). همچنین فرض کنید سپس ما احتمال یک میانگین نمونه ای به بزرگی ۵۶ از آن توزیع را تعیین کنیم. در همین راستا همچنین فرض کنید این احتمال ۰/۱۵ باشد. آنگاه استدلال ما چنین خواهد بود: "اگر ما در عمل از جمعیتی با $\mu = 50$ نمونه گیری کنیم، احتمال بدست آوردن یک نمونه با میانگینی به بزرگی ۵۶ معادل

۰/۱۵ است- یک رخداد نسبتاً ممکن". از آنجا که یک میانگین نمونه‌ای با این اندازه از جمعیتی با میانگین ۵۰ اغلب بدست می‌آید، دلیلی ندارد که شک کنیم این نمونه از چنین جمعیتی بدست آمده است. از سوی دیگر فرض کنید که یک میانگین نمونه‌ای ۶۲ بدست آورده‌ایم و احتمال نمونه‌ای با میانگینی به بزرگی ۶۲ از توزیع نمونه‌گیری را فقط ۰/۰۰۴ یافته‌ایم. در این صورت استدلال ما این خواهد بود: "اگر در عمل از جمعیتی با $\mu = 50$ نمونه‌گیری کنیم، احتمال بدست آوردن یک نمونه با میانگینی به بزرگی ۶۲ معادل ۰/۰۰۴ است- رخدادی غیر محتمل. از آنجا که بدست آوردن یک میانگین نمونه‌ای به این بزرگی خیلی غیرمحتمل است، ما می‌توانیم بطور منطقی نتیجه‌گیری کنیم که این نمونه احتمالاً از جمعیتی دیگر آمده است (که میانگین آن ۵۰ نیست)." لازم است به طور دقیق دربابیم در این مثال چه چیزی صورت گرفته است، زیرا منطق آن برای بیشتر آزمون‌های فرضیه صادق است. آزمون واقعی شامل چند مرحله بود.



شکل ۱-۱۱ توزیع میانگین‌های نمونه‌ای مشکلات رفتاری، براساس $n = 5$ نمره برای هر نمونه.

۱. ما می‌خواستیم این فرضیه را که اغلب بنام فرضیه تحقیق از آن یاد می‌شود آزمایش کنیم، که احتمال بروز مشکلات رفتاری در کودکان تحت فشار روانی نسبت به کودکان عادی بیشتر است.
۲. ما این فرضیه (معروف به فرض صفر/پوچ، H_0) را بنا نهادیم که این نمونه در اصل از جمعیتی با میانگین معادل ۵۰، مشخص شده با نماد μ_0 ، برگرفته شده است. این فرضیه‌ای است که تأکید می‌نماید که کودکان تحت فشار روانی از لحاظ مشکلات رفتاری با کودکان عادی فرقی ندارند.
۳. سپس ما تعیین کردیم که توزیع نمونه‌گیری میانگین تحت فرض H_0 (فرض صفر) درست است (یعنی ما توزیع نمونه‌گیری را از یک جمعیت با $\mu = 50$ بدست آورده‌ایم).

۴. ما یک نمونه تصادفی از کودکان تحت فشار گرفتیم.
۵. با داشتن توزیع نمونه گیری، ما احتمال میانگینی دست کم به بزرگی میانگین نمونه واقعی را حساب کردیم.
۶. براساس این احتمال یک تصمیم گیری صورت دادیم. ما فرض H_0 را رد کردیم یا نتوانستیم رد کنیم. از آنجا که H_0 می گوید $\mu = 50$ ، رد کردن H_0 به این باور می انجامد که $\mu > 50$ است، هر چند مقدار واقعی μ نامشخص باقی می ماند.

بحث پیش خیلی موضوع را ساده می انگارد از این لحاظ که ما بطور کلی ترجیح می دهیم که به آزمون این فرضیه تحقیق بپردازیم که کودکان تحت فشار روانی با دیگر کودکان فرق دارند (نه اینکه بالاتر هستند)، ما دو باره به اختصار به این موضوع خواهیم پرداخت. همچنین از این نظر نیز ساده انگارانه است که در عمل لازم است ما مقدار σ^2 ، واریانس جمعیت، و N ، اندازه نمونه را (خواه مستقیماً یا به صورت برآورد) در نظر بگیریم. اما منطق این رویکرد بازتابی از منطق بیشتر آزمون های آماری است البته نه همه آنها. در هر مورد ما فرض صفر را بنا می کنیم، توزیع نمونه گیری آماره خاص را بر اساس اینکه H_0 درست است تشکیل داده، نمونه آماری را با آن توزیع مقایسه و سپس بسته به مقدار احتمال تحت H_0 ، به عنوان یک نمونه آماری که می توانیم در حد نهایی بدست آوریم، فرض صفر را رد می کنیم یا می پذیریم.

۴-۱۱ فرض صفر

چنانکه دیدیم مفهوم فرض صفر نقش بسزایی در آزمون فرضیه دارد. برای دانشجویان همواره این ابهام وجود دارد که چرا ما فرضیه ای را طرح می کنیم که درست بر خلاف آن چیزی است که می خواهیم نشان دهیم. برای مثال اگر بخواهیم این فرضیه تحقیق را که دانشجویان از یک جمعیت با میانگین نمره اعتماد به نفس ۱۰۰ نمی آیند را نشان دهیم، بی درنگ این فرض صفر را طرح می کنیم که آنها از این جمعیت آمده اند. یا اگر می خواهیم اعتبار یک فرضیه تحقیق را نشان دهیم که میانگین های جمعیتی که دو نمونه از آنها برگرفته شده اند (μ_1 و μ_2) متفاوت هستند، فرض صفر را به این ترتیب بنا می کنیم که میانگین های جمعیت ها مساوی هستند ($\mu_1 - \mu_2 = 0$). (معنای فرض صفر در این مثال دوم به خوبی مشخص می شود که اشاره به این فرضیه دارد که تفاوت بین دو میانگین صفر یا پوچ است). دلایل متعددی برای استفاده از فرض صفر وجود دارد. بحث فلسفی که توسط فیشر زمانی که اولین بار این مفهوم را مطرح کرد به کارگرفته شد، این است که ما هرگز نمی توانیم ثابت کنیم چیزی درست است، اما می توانیم ثابت کنیم که چیز درست نیست. مشاهده ۳۰۰۰ نفر با یک سر، ثابت کننده این عبارت نیست که "هر فرد یک سر دارد". اما پیدا کردن یک فرد با دو سر درست بودن عبارت فوق را رد می کند

بدون اینکه تردیدی پیش آید. در حالیکه می‌توان روی این موضوع با فیشر بحث کرد- خیلی‌ها کرده‌اند- اما فرض صفر در آمار جایگاه خود را حفظ کرده است.

یک دلیل دوم و عملی برای بکارگیری فرض صفر این است که نقطه آغازی برای آزمون آماری فراهم می‌کند. به موردی توجه کنید که در آن می‌خواهید نشان دهید میانگین نمره اعتماد به نفس دانشجویان بزرگتر از ۱۰۰ است. علاوه بر این فرض کنید در باره درستی بعضی فرضیه‌ها به شما تضمین داده شده است. شماچه فرضیه‌ای را می‌خواهید بیازمایید؟ آیا شما باید بیازمایید که $\mu = 101$ است یا ممکن است فرض این باشد که $\mu = 112$ است، یا $\mu = 113$ چطور؟ نکته این است که فرضیه جایگزین خاصی (فرض تحقیق) در نظر ندارید، و بدون آن نمی‌توانید توزیع نمونه‌گیری مورد نیاز خود را بسازید. اما اگر با $\mu = 100 : H_0$ شروع کنید بی‌درنگ می‌توانید در باره یافتن توزیع نمونه‌گیری برای $\mu = 100$ اقدام کنید و سپس با خوش‌شانسی فرض صفر را رد کنید و نتیجه بگیرید که نمره میانگین دانشجویان بزرگتر از ۱۰۰ است، چیزی که در ابتدا می‌خواستید آن را نشان دهید.

۵-۱۱ آماره‌های آزمون و توزیع‌های نمونه‌گیری آنها

ما در باره توزیع نمونه‌گیری میانگین بحث کردیم، اما این بحث کلی، در مورد میانه، واریانس، دامنه و هر آماره دیگری هم در اساس یکسان خواهد بود. (شکل این توزیع‌ها به لحاظ فنی فرق دارد.) اما در این فصل ما عمداً از این قضایا چشم‌پوشی می‌کنیم. آماره‌هایی که در اینجا ذکر شده‌اند به عنوان **آماره‌های نمونه‌ای**^۱ از آنها یاد می‌شود چون برای توصیف یک نمونه به کار می‌روند. یک دسته کلاً متفاوت از آماره‌ها وجود دارند بنام **آماره‌های آزمون**^۲، که با شیوه‌های آماری خاصی تداعی می‌شوند و توزیع‌های نمونه‌گیری خاص خود را دارند. اینها آماره‌هایی مانند t ، F ، χ^2 و غیره هستند، که ممکن است از قبل با آنها آشنایی داشته باشید. اگر هم آشنا نیستید نگران نباشید- در فصل‌های بعدی به آنها خواهیم پرداخت. فعلاً جای توضیح جزئیات آنها نیست. این فصل برای این منظور ارائه شده تا دریابیم توزیع‌های نمونه‌گیری برای آماره‌های آزمون بدست می‌آیند و به همان شیوه توزیع نمونه‌گیری از میانگین به کار گرفته می‌شوند. برای تصور بهتر، یک توزیع نمونه‌گیری از آماره t را در نظر بگیرید، که در سه فصل بعدی توضیح داده خواهد شد. برای کسانی که از قبل با آماره t آشنایی ندارند کافی است اشاره کنیم که این آماره اغلب برای پاسخ به این سؤال به کار می‌رود که آیا دو نمونه، از جمعیت‌هایی با میانگین‌های یکسان بیرون کشیده شده‌اند. اجازه دهید μ_1 و μ_2 نشان دهنده میانگین‌های جمعیت‌هایی باشند که دو نمونه از آنها گرفته شده‌اند. فرض صفر گویای این است که میانگین‌های دو جمعیت برابر هستند- به عبارت دیگر

نمونه‌گیری t را وقتی $H_0: \mu_1 = \mu_2$ (یا $H_0: \mu_1 - \mu_2 = 0$)، اگر خیلی با حوصله باشید شما می‌توانید به صورت تجربی توزیع جمعیت واحد بدست آورید، برای هر جفت نمونه مقدار t را حساب کنید (به روشی که بعداً گفته خواهد شد)، و مقادیر t بدست آمده را روی نمودار نشان دهید. در این مورد بایستی H_0 درست باشد زیرا نمونه‌ها از جمعیت واحدی استخراج شده‌اند. توزیع بدست آمده یک توزیع نمونه‌گیری t است در حالتی که H_0 درست است. اکنون اگر ما دو نمونه داشته باشیم که یک مقدار خاص t را تولید می‌کند، فرض صفر را با مقایسه t نمونه‌ای خود با توزیع نمونه‌گیری t می‌آزماییم. حال در صورتی فرض صفر را رد می‌کنیم که t ما مشابه آن مقادیر t که طبق توزیع نمونه‌گیری باید در حالتی که فرض صفر درست است انتظار داشته باشیم نباشد.

مطالب بند پیش را می‌توان به همین ترتیب با جایگزینی مجذور χ^2 ، یا هر آزمون آماری دیگری به جای t بازگو کرد، تنها چیزی که تغییر می‌کند نحوه محاسبه آن آماره خاص است. بنابراین دیدیم که همه توزیع‌های نمونه‌گیری را می‌توان به روش اساساً مشابهی (محاسبه و ترسیم شمار نامحدودی از آماره‌ها با نمونه‌گیری از یک جمعیت مشخص) بدست آورد. یک بار که این واقعیت خواننده را بفهمد، بقیه این کتاب عمدتاً تمرین کردن و مهارت یافتن در شیوه‌های محاسبه این آماره‌ها و شرح ویژگی‌های آن توزیع نمونه‌گیری خواهد بود.

۶-۱۱ استفاده از توزیع بهنجار برای آزمون فرضیه‌ها

بیشتر بحث ما تا به حال در ارتباط با شیوه‌های آماری بود که هنوز چگونگی استفاده از آنها را نمی‌دانید. این کار به عمد صورت گرفت تا بر این نکته تأکید شود که منطق و محاسبات پشت یک آزمون، دو قضیه جداگانه هستند. اما اکنون می‌توانیم آنچه از قبل درباره توزیع بهنجار (نرمال) می‌دانیم را برای آزمون بعضی فرضیه‌های ساده بکار ببریم. در این فرایند سر و کار ما با قضایای اساسی است که با استفاده از یک مثال عینی آسان‌تر فهمیده می‌شوند.

یکی از کاربردهای مهم توزیع نرمال آزمودن فرضیه‌ها است. از این توزیع می‌توان هم برای آزمون فرضیه‌ها درباره مشاهدات فردی هم درباره آماره‌های نمونه‌ای مانند میانگین استفاده کرد. در این فصل به سراغ مشاهدات فردی می‌رویم و آماره‌های نمونه‌ای را به فصل‌های بعد موکول می‌کنیم. اما توجه داشته باشید که در حالت کلی ما به آزمودن فرضیاتی در باره نمونه‌های آماری مانند میانگین می‌پردازیم نه مشاهدات فردی. ولی با مثالی از یک مشاهده فردی آغاز می‌کنیم چون توضیحات آن روشن‌تر است. از آنجا که سر و کار ما فقط با مشاهدات تکی است، توزیع نمونه‌گیری در اینجا شامل توزیع نمرات منفرد خواهد بود (نه توزیع میانگین‌ها). منطق زیربنایی یکسان است، و ما به سراغ مثالی از این نوع نمرات می‌رویم که از قبل هم با آن آشنایی دارید و توضیحات را ساده می‌کند.

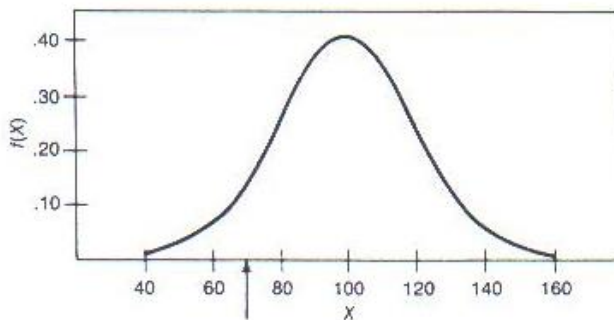
به عنوان یک مثال ساده فرض کنید میانگین نرخ ضربه زدن با انگشت در افراد سالم ۱۰۰ بار در ۲۰ ثانیه، با انحراف معیار ۲۰ است، و اینکه سرعت این کار در جمعیت دارای توزیع نرمال است. علاوه بر این فرض کنید که می‌دانیم نرخ ضربه زدن با انگشت در بین افراد دارای بعضی مشکلات عصبی پایین‌تر است. (در واقع ضربه زدن با انگشت یکی از شاخص‌های تشخیص آسیب‌های عصبی است، گرچه تفاوت نرخ‌های آن برای دست چپ و راست مهم‌تر است.) در نهایت فرض کنید فردی به ما ارجاع داده شده که نرخ ضربه زدن با انگشت در او ۷۰ بار در ۲۰ ثانیه است. آیا نمره او به اندازه کافی زیر میانگین هست تا اطمینان داشته باشیم از یک جمعیت از لحاظ عصبی سالم نیامده است؟ این شرایط در شکل ۲-۱۱ ترسیم شده است، که در آن کمان نشان دهنده موقعیت نمونه داده‌های ما (نمره شخص) است.

منطق راه‌حل این مسأله همانند منطق فرضیه آزمایی در کل است. ما با این فرض آغاز می‌کنیم که نمره شخص از یک جمعیت با نمرات سالم می‌آید. این فرض صفر است (H_0). اگر H_0 درست باشد ما خودبخود میانگین و انحراف معیار جمعیتی که از آن بیرون کشیده شده را می‌دانیم (به ترتیب ۱۰۰ و ۲۰). با این اطلاعات در موقعیتی هستیم تا این احتمال را که نمره‌ای به کوچکی نمره او از این جمعیت بدست آمده باشد را حساب کنیم. اگر این احتمال خیلی کم باشد می‌توانیم H_0 را رد کنیم و نتیجه بگیریم که او از یک جمعیت سالم نیامده است. از سوی دیگر اگر این احتمال خیلی کم نباشد، آنگاه داده‌ها یک نتیجه مستدل تحت شرایط H_0 را نشان خواهند داد، و دلیلی ندارد در باره اعتبار آن شک کنیم، و از این رو دلیلی ندارد که به سالم بودن این شخص شک کنیم. به خاطر داشته باشید که ما نمی‌خواهیم احتمال نمره مساوی ۷۰ (که بسیار کوچک است چون توزیع پیوسته است)، را بدانیم، بلکه مایلیم احتمال اینکه این نمره به کوچکی ۷۰ (یعنی کوچکتر یا مساوی ۷۰) باشد را بدانیم.

این فرد دارای نمره ۷۰ است. چیزی که ما می‌خواهیم بدانیم احتمال بدست آوردن یک نمره حداقل به کوچکی ۷۰ در صورت درست بودن H_0 است. این چیزی است که ما حالا می‌دانیم چگونه باید آن را پیدا کنیم - همان سطح زیر ۷۰ در شکل ۲-۱۱ است. همه کاری که باید انجام دهیم این است که نمره ۷۰ را به یک نمره z تبدیل کرده و سپس به سراغ جدول نمرات z در بخش پیوست، یعنی جدول سطوح زیر منحنی نرمال برویم.

$$z = \frac{X - \mu}{\sigma} = \frac{70 - 100}{20} = -1.5$$

از جدول مربوط به سطوح زیر منحنی نرمال در بخش پیوست می‌توانیم مشاهده کنیم که احتمال یک نمره z معادل -1.5 یا کمتر، ۰/۰۶۶۸ است. (در جدول پیوست این اطلاعات در ستون سمت مخالف $z = 1.5$ تحت عنوان "بخش کوچکتر" درج شده است.)



شکل ۲-۱۱ موقعیت نمره یک شخص روی یک توزیع نمرات افراد دارای سلامت عصبی

در این مرحله ما باید به جنبه تصمیم‌گیری فرضیه‌آزمایی بپردازیم. ما باید تصمیم بگیریم که آیا پدیده‌ای با احتمال $0/0668$ به اندازه کافی نامحتمل است که باعث شود H_0 را رد کنیم. در اینجا ما بر می‌گردیم سراغ قراردادهای اختیاری که به مرور زمان بنا شده‌اند. منطق این قراردادها به تدریج که پیش می‌رویم روشن خواهد شد، اما برای شروع در نظر داشته باشید که اینها صرفاً قراردادی هستند. یک قرارداد می‌گوید در صورتی H_0 را رد می‌کنیم که احتمال تحت H_0 کمتر یا برابر با $0/05$ باشد ($P \leq 0/05$), در حالی که قرارداد دیگر - محافظه کارانه‌تر در رد کردن H_0 - می‌گوید H_0 را زمانی رد می‌کنیم که احتمال تحت H_0 کمتر یا برابر با $0/01$ باشد. به این مقادیر $0/05$ و $0/01$ اغلب سطح رد کردن^۱ یا سطح معناداری^۲ آن آزمون گفته می‌شود. هر جا که احتمال بدست آمده تحت H_0 کوچکتر یا مساوی با سطح معناداری از قبل تعیین شده ما باشد، ما H_0 را رد می‌کنیم. روش دیگر برای بیان این موضوع این است که بگوییم هر پیامدی که احتمال آن تحت H_0 کوچکتر یا مساوی سطح معناداری واقع در ناحیه رد کردن^۳ باشد، زیرا چنین پیامدی به رد H_0 می‌انجامد. به منظور اینکه سطح استاندارد برای رد کردن داشته باشیم ما سطح معناداری $0/05$ را در این کتاب در نظر می‌گیریم، ضمناً به خاطر داشته باشید که برخی افراد این سطح را آسان‌گیرانه تلقی می‌کنند. وقتی به فصل ۱۵ برسیم معلوم می‌شود چرا این سطح را برگزیده‌ایم.* برای مثال قبلی، ما به مقدار احتمال $0/0668$ دست یافتیم که آشکارا بزرگتر از $0/05$ است. بنابر این نتیجه می‌گیریم که دلیلی ندارد این شخص از جمعیت افراد سالم نیامده باشد زیرا مشخص کردیم که H_0 را رد نخواهیم کرد مگر اینکه احتمال داده‌ها تحت H_0 کمتر از $0/05$ باشد. بطور دقیق‌تر،

1. rejection level
2. significance level
3. rejection region

* رویکرد سنتی این بوده که معناداری را محدود به کمتر از $0/05$ بداند و $p=0/051$ و $p=0/150$ را به یک اندازه غیرمعنادار می‌داند. اما اخیراً $p=0/051$ را در مرز معناداری به حساب می‌آورند یا $p=0/003$ را بسیار معنادار می‌خوانند. برنامه‌های کامپیوتری امروزه سطح دقیق معناداری را محاسبه می‌کنند.

ما نتیجه می‌گیریم که یک نرخ ضرب-انگشت ۷۰ می‌تواند بطور مستدل از یک جمعیت با میانگین ۱۰۰ و انحراف معیار ۲۰ آمده باشد. بسیار مهم است که توجه شود ما نشان نداده‌ایم که این فرد سالم است، بلکه فقط دلایل کافی نداریم که بگوییم او سالم نیست. می‌توان گفت این فرد فقط به بیماری مبتلا شده و بنابراین کاملاً از حد نرمال آنگونه که برای شرایط او عادی است متفاوت نیست. یا گفته شود او دارای بیماری در مرحله‌ای پیشرفته است اما فقط در یک حد غیرمعمول از ضرب-انگشت واقع شده است. این نمونه‌ای از این واقعیت است که ما هرگز نمی‌توانیم بگوییم که فرض صفر را اثبات کردیم. ما تنها می‌توانیم نتیجه بگیریم که این شخص به اندازه کافی به عنوان یک بیمار کُند نیست تا به لحاظ آماری قابل تشخیص باشد.

نظریه آزمون معناداری به گونه‌ای که بیان شد در اصل به آر. ا. فیشر در ثلث اول قرن ۲۰م بر می‌گردد. این نظریه از سوی جرجی نیمن و ایگون پیرسون بین سال‌های ۱۹۲۸ و ۱۹۳۸ گسترش یافت و به چارچوبی برای تصمیم‌گیری تبدیل شد- که مغایر با اهداف فیشر بود. کارهای آماری کنونی بیشتر بر اساس رویکرد نیمن - پیرسون است که تأکید بیشتری نسبت به فیشر بر این واقعیت دارند که ما یک فرضیه جایگزین هم داریم (H_1) که مقابل فرض صفر (H_0) است. بنابراین چنانچه فرض صفر این باشد

$$H_0 : \mu = 100$$

آنگاه فرض مقابل چنین خواهد بود

$$H_1 : \mu \neq 100$$

$$H_1 : \mu < 100 \quad \text{یا}$$

$$H_1 : \mu > 100 \quad \text{یا}$$

چیزهای بیشتری در باره این فرض مقابل برای گفتن خواهیم داشت.

۷-۱۱ خطاهای نوع I و نوع II

هرگاه ما در باره آزمون آماری به تصمیمی می‌رسیم، این شانس هم وجود دارد که تصمیم ما غلط باشد. در حالی که این موضوع تقریباً در باره همه تصمیمات صادق است، آمار یا آماردان یک امتیاز دارد که تصمیم‌گیران دیگر ندارند. او نه تنها یک تصمیم را براساس بعضی فرایندهای منطقی می‌گیرد، بلکه می‌تواند احتمال خطای آن تصمیم را نیز مشخص کند. در زندگی روزمره ما تنها با این احساس ذهنی تصمیم‌گیری می‌کنیم که احتمالاً انتخاب درست را انجام داده‌ایم. در مقابل، آماردان می‌تواند بطور کاملاً دقیقی احتمال اینکه او به اشتباه H_0 را به نفع H_1 رد کرده باشد وقتی که H_0 در حقیقت درست است را تعیین کند. این توانایی مشخص کردن خطا بطور مستقیم از منطق فرضیه آزمایی پیروی می‌کند.

به مثال قبلی ضربه انگشت توجه کنید، این بار نمره بدست آمده از آن شخص را نادیده بگیرید.

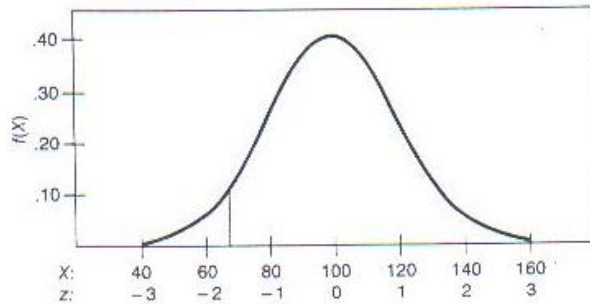
این وضعیت در شکل ۳-۱۱ نمایش داده شده است، که توزیع در اینجا توزیع نمرات آزمودنی‌های سالم

است و بخش سایه دار نشان دهنده ۵٪ پایین توزیع است. نمره واقعی که ۵٪ پایین را جدا می سازد مقدار بحرانی^۱ نامیده می شود. مقادیر بحرانی آنهایی هستند که مرز یا مرزهای ناحیه رد کردن را مشخص می کنند. برای این مثال خاص مقدار بحرانی ۶۷ است.

اگر یک قاعده تصمیم گیری داشته باشیم که بگوید H_0 جایی رد می شود که یک پیامد در ۵٪ پایین توزیع قرار گیرد، آنگاه ما H_0 را در جایی که نمره یک نفر در ناحیه سایه دار قرار بگیرد رد می کنیم، یعنی هر جا یک نمره چندان پایین باشد که احتمال آن که از یک جمعیت سالم بیاید ۰/۰۵ یا کمتر باشد. با این وجود در حالت کاملاً طبیعی شیوه ما، ۵٪ از نمرات افراد کاملاً سالم خودشان در ناحیه سایه دار قرار خواهند داشت. بنابراین اگر ما عملاً نمونه ای داشته باشیم که سالم باشد براساس شانس ۵٪ احتمال دارد که این نمره در دنباله سایه دار توزیع باشد، که این باعث اشتباه در رد فرض صفر می شود. این نوع خطا (رد H_0 وقتی در واقع درست است) **خطای نوع I** نامیده می شود، و احتمال آن با α (آلفا)، اندازه ناحیه رد کردن مشخص می شود. در ادامه هر جا ما یک احتمال را با α نشان دهیم، منظورمان احتمال خطای نوع I است.

ممکن است شما احساس کنید ۵٪ شانس خطا، خطر خیلی بزرگی است و پیشنهاد کنید که ملاک خود را سخت گیرانه تر کنیم- مثلاً رد فرض صفر را فقط به ۱٪ پایین توزیع محدود کنیم. این یک شیوه کاملاً قابل قبول است، اما به این واقعیت توجه داشته باشید که هر چه ملاک خود را سخت تر کنید، احتمال بیشتری وجود دارد که دچار نوع دیگر خطا شوید- عدم رد فرض H_0 وقتی که در حقیقت H_1 درست است. این نوع خطا را **خطای نوع II** گویند، و احتمال آن با β (بتا) نمادگذاری می شود.

مشکل اصلی در عبارات خطاهای نوع II از این واقعیت ناشی می شود که اگر H_0 نادرست باشد، ما تقریباً هیچگاه نمی دانیم برای جمعیتی که داده های ما از آن برآمده توزیع درست (توزیع تحت H_1) شبیه به چیست. ما تنها از توزیع نمرات تحت H_0 اطلاع داریم. در زمینه مثال قبلی می توان گفت که ما توزیع نمرات افراد سالم را در اختیار داریم نه افرادی که بیمار هستند. ممکن است چنین باشد که افراد مبتلا به بعضی از نارسائی های عصبی، در کل، دارای نرخ ضربه انگشت پایین تری نسبت به افراد سالم هستند، یا چنین باشد که آنها تنها کمی پایین تر هستند. این شرایط در شکل ۴-۱۱ نشان داده شده است، که در آن توزیع با برچسب H_0 نشان دهنده توزیعی از نمرات آزمودنی های برخوردار از سلامت است (مجموعه ای از مشاهدات مورد انتظار تحت فرض صفر)، و توزیع با برچسب H_1 نشان دهنده توزیع فرضی ما از نمرات ناسالم است (توزیع تحت H_1). به خاطر داشته باشید که منحنی H_1 فقط یک چیز فرضی است، ما واقعاً موقعیت توزیع ناسالم را نمی دانیم، به غیر از اینکه از توزیع H_0 پایین تر است (دارای سرعت های پایین تری است). (ما آن را بطور اختیاری با میانگین ۸۰ و انحراف معیار ۲۰ رسم کرده ایم).

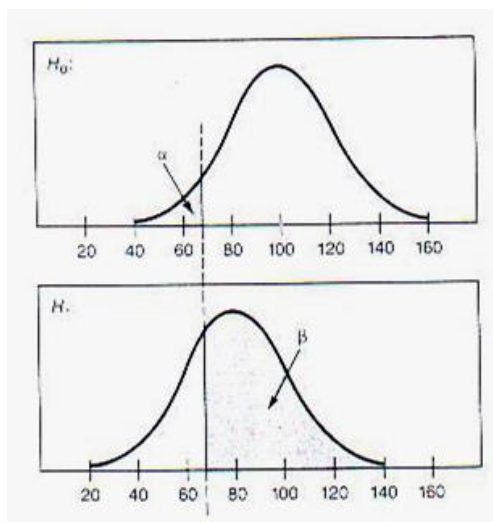


شکل ۳-۱۱ ۵ درصد پایین نمرات افراد برخوردار از سلامت

بخش تیره در نیمه بالایی شکل ۴-۱۱ ناحیه رد کردن را نشان می‌دهد. هرگونه مشاهداتی که در این ناحیه قرار گیرد (یعنی در سمت چپ نمره ۶۷) به رد کردن فرض صفر می‌انجامد. اگر فرض صفر درست باشد، ما می‌دانیم که مشاهده ما در این ناحیه واقع شده است و بنابراین در ۵٪ موارد دچار یک خطای نوع I خواهیم بود.

بخش تیره نیمه پایین شکل ۴-۱۱ [در سمت راست نمره ۶۷] نشان دهنده احتمال (β) خطای نوع II است. این در شرایطی است که در آن شخصی عملاً از جمعیت غیر سالم بیرون کشیده شده است اما نمره او چندان پایین نیست که باعث شود ما H_0 را رد کنیم.

در شرایط خاص نشان داده شده در شکل ۴-۱۱ ما در واقع می‌توانیم β را با استفاده از توزیع نرمال حساب کنیم تا احتمال بدست آمدن یک نمره بالاتر از ۶۷ (مقدار بحرانی) به شرط $\mu = 80$ و $\sigma = 20$ را محاسبه کنیم. محاسبات عملی به فهم شما از β کمکی نمی‌کند، و چون هدف این فصل مخصوصاً پرهیز از کارهای محاسباتی بوده، ما صرفاً به بیان اینکه این احتمال (یعنی سطحی که با برچسب β مشخص شده) معادل ۰/۷۴ است بسنده می‌کنیم. بدین ترتیب در این مثال در ۷۴٪ موارد وقتی که ما شخصی را داریم که واقعاً غیر سالم است (یعنی H_1 واقعاً درست است)، ما با عدم رد H_0 وقتی که نادرست است دچار خطای نوع II می‌شویم- ما به عنوان متخصص تشخیص پزشکی شانس زیادی برای بیمار نبودن در نظر گرفته‌ایم.



شکل ۴-۱۱ سطح مربوط به α و β برای مثال سرعت ضربه زدن

از شکل ۴-۱۱ می‌توان دریافت که اگر بخواهیم سطح α (احتمال خطای I) را از ۰/۰۵ به ۰/۰۱ با بردن ناحیه رد کردن به سمت چپ کاهش دهیم، احتمال خطای نوع I کاهش می‌یابد اما احتمال خطای نوع II افزایش خواهد یافت. تنظیم α در ۰/۰۱ به این معناست که $\beta = ۰/۹۰۸$. بنابراین جای زیادی برای تردید در باره اینکه چه سطح معناداری به کار برده شود وجود دارد. تصمیم‌گیری در اصل بستگی دارد به نظر شما در باره اهمیت نسبی خطاهای نوع I و نوع II در تحقیقی که انجام می‌دهید. اگر پرهیز از خطای نوع I اهمیت دارد (مثل اطلاع دادن به کسی که او مبتلا به بیماری است وقتی که نیست)، در این صورت شما سطح سخت‌گیرانه‌ای از α (یعنی سطح کوچکی) را در نظر خواهید گرفت. از سوی دیگر اگر بخواهید از خطای نوع II پرهیز کنید (به کسی بگویید برو خانه و یک آسپرین بخور در حالی که او به درمان فوری نیاز دارد)، آنگاه ممکن است یک سطح بالای α را در نظر بگیرید. (تعیین سطح α در ۰/۲۰ باعث کاهش β به ۰/۴۴ در این مثال می‌شود.) متأسفانه در عمل بیشتر افراد یک سطح اختیاری α مانند ۰/۰۵ یا ۰/۰۱ را در نظر می‌گیرند، و بسادگی β را نادیده می‌گیرند. در بعضی موارد این تنها کاری است که شما می‌توانید انجام دهید. اما در موارد دیگر کارهای بیشتری می‌توان انجام داد که در فصل ۱۵ خواهید دید.

باید دوباره تأکید کنیم که شکل ۴-۱۱ فرضی است. ما توانستیم این شکل را رسم کنیم تنها به این دلیل که بطور اختیاری تصمیم گرفتیم که سرعت افراد غیرسالم دارای توزیع نرمال با میانگین ۸۰ و انحراف معیار ۲۰ باشد. در بسیاری از موقعیت‌های روزمره ما میانگین و واریانس توزیع را نمی‌دانیم و فقط می‌توانیم حدس خردمندانه‌ای بزنیم. در عمل می‌توانیم یک مقدار μ تحت H_1 در نظر بگیریم که

نشان‌گر حداقل تفاوت مورد نظر برای فرق گذاشتن باشد، چون تفاوت‌های بیشتر حتی β های کوچک‌تر را در پی خواهد داشت. در فصل ۱۵ باز به این موضوع بر می‌گردیم.

بر اساس بحث خطاهای نوع I و II می‌توان فرایند تصمیم‌گیری را به شکل یک جدول ساده خلاصه کرد. جدول ۱۱-۱ چهار نتیجه ممکن یک تجربه را نشان می‌دهد. مطالب این جدول خود گویاست. اما خانه‌ی چپ بالا حاوی یک مفهوم است (توان) که قبلاً توضیح ندادیم. توان یک آزمون، احتمال رد H_0 است وقتی که واقعاً نادرست است. چون احتمال عدم رد یک H_0 معادل β است، بنابراین توان برابر $1 - \beta$ است. در فصل ۱۵ باره بیشتر توضیح خواهیم داد.

جدول ۱۱-۱ نتایج ممکن در فرایند تصمیم‌گیری

واقعیت موجود

تصمیم	H_0 درست	H_0 نادرست
رد شدن H_0	خطای نوع I $p = \alpha$	تصمیم درست توان $p = 1 - \beta$
رد نشدن H_0	تصمیم درست $p = 1 - \alpha$	خطای نوع II $p = \beta$

۱۱-۸ آزمون‌های یک دامنه و دو دامنه

بحث قبلی توجه ما را به آزمون‌های یک دامنه و دو دامنه جلب کرد. در مثالی که ارائه شد ما می‌دانستیم که آزمودنی‌های غیرسالم کندتر از افراد سالم ضربه می‌زدند و براین اساس ما تصمیم گرفتیم H_0 را فقط در صورتی رد کنیم که یک آزمودنی خیلی آهسته ضربه بزند. اما فرض کنید که آزمودنی ما ۱۸۰ بار در ۲۰ ثانیه ضربه بزند. هر چند مشاهده این میزان در یک آزمودنی سالم پدیده‌ای بسیار غیرمحتمل است، اما در ناحیه رد کردن که تنها شامل نرخ‌های پایین باشد قرار نمی‌گیرد. در نتیجه ما خود را در شرایط رد نشدن H_0 در روبرو شدن با بخشی از داده‌ها می‌بینیم که خیلی غیرمحتمل است، ولی در جهت مورد انتظار نیستند.

در این حالت این پرسش پیش می‌آید که در برابر این شرایط باید چه کار کنیم. پاسخ این است که پیش از اجرای آزمایش ما مشخص کنیم که می‌خواهیم درصد معینی (مثلاً ۵٪) از پیامدهای حدود نهایی - هر دو سوی بالا و پایین، را رد کنیم. اما اگر ۵٪ پایین و ۵٪ بالا را رد کنیم، آنگاه ما H_0 را در واقع با ۱۰٪ کل موارد وقتی که درست است رد خواهیم کرد، یعنی $\alpha = ۰/۱۰$. به ندرت ما مایلیم α را به بزرگی ۰/۱۰ در نظر بگیریم بلکه ترجیح می‌دهیم آن را از ۰/۰۵ بالاتر نبریم. تنها راه انجام این کار این است که ۵٪/۲ پایین و ۵٪/۲ بالا را رد کنیم و به مقدار کل ۵٪ برسیم.

شرایطی که در آن ما H_0 را فقط برای موارد پایین (یا فقط موارد بالا)ی سرعت ضربه زدن رد می کنیم به عنوان آزمون یک دامنه^۱ یا جهت دار^۲ معروف است، زیرا ما برای یک جهت پیش بینی می کنیم که فرد از میانگین متفاوت است و ناحیه رد فقط در یک دنباله توزیع قرار گرفته است. وقتی ما حد نهایی را در هر یک از دنباله ها رد می کنیم، ما به آزمون دو دامنه یا غیر جهت دار^۳ می پردازیم. باید خاطر نشان کرد که وقتی ما از امتیاز آزمون دو دامنه استفاده می کنیم (توانایی رد کردن فرض صفر برای نمرات حدود نهایی در هر جهت)، همزمان برخی چیزها را هم از دست می دهیم. نمره ای که در ناحیه رد یک آزمون یک دامنه قرار می گیرد ممکن است در ناحیه رد یک آزمون دو دامنه قرار نگیرد، زیرا در اینجا ما تنها ۲/۵٪ را در هر دنباله رد می کنیم.

در مثال ضربه انگشت تصمیم گیری بین یک آزمون یک دامنه و دو دامنه به نظر می رسد بتواند خط برش روشنی باشد. ما می دانیم که افراد مبتلا به بیماری کندتر ضربه می زنند، و بنابراین ما فقط در باره رد H_0 برای نمرات پایین دقت می کنیم- نمرات بالا اهمیت تشخیصی ندارند. (سرعت بالا می تواند ناشی از تمرین باشد). اما در موارد بسیار دیگری ما نمی دانیم کدام دنباله توزیع (یا هر دو) مهم است، و لازم است در برابر حدود نهایی در هر سوی توزیع احتیاط کنیم. این شرایط می تواند در یک جلسه متقاعد کردن کودکان به بیشتر مسواک زدن دندان ها پیش آید. ممکن است پی ببریم که این کار به افزایش رفتار مطلوب انجامیده است. از سوی دیگر ممکن است کودکان از اینکه به آنها گفته شود مسواک کنید ناخوشنود باشند و بنابراین حتی لجبازی کنند و کمتر از قبل مسواک کنند. در هر دو حالت ما می خواهیم H_0 را رد کنیم.

در مجموع آزمون های دو دامنه خیلی بیش از آزمون های یک دامنه معمول هستند. چند دلیل دارد. اول از همه اینکه محقق ممکن است ایده ای درباره اینکه داده ها چگونه خواهند بود نداشته و مجبور شود برای هر چیزی آمادگی داشته باشد. هر چند این شرایط نادر است، اما در بعضی کارهای اکتشافی رخ می دهد.

دلیل عمومی دیگر برای ترجیح آزمون های دو دامنه این است که محقق شواهدی برای اینکه داده ها از یک راه بر می آیند دارد اما می خواهد خود را از اشتباه بر حذر دارد. چنین شرایطی بیشتر از آنچه فکر می کنید پیش می آید. (فرضیه های به دقت طرح شده دارای یک عادت ناخوشایند هستند که در جهت نادرست نوشته شوند، بدلالی که بعد از آن خیلی روشن به نظر می آیند). پرسشی که زیاد پیش می آید هنگامی که داده ها از راهی دیگر به نتیجه برسند این است "چرا آزمون یک دامنه را طرح ریزی نکنیم و بعداً، اگر از داده ها روش دیگری برآمد آزمون را به دو دامنه تغییر دهیم؟" این نوع پرسش ها برای افرادی پیش می آید که نمی خواهند به انحراف بروند، بلکه فقط منطق کاری را که می کنند نمی فهمند.

-
1. one-tailed test
 2. directional
 3. nondirectional

اگر شما آزمایشی را آغاز کنید که حد نهایی ۵٪ را در دنباله چپ به عنوان ناحیه رد کردن در نظر بگیرید و سپس برگردید هر پیامدی که در حدود نهایی ۲/۵٪ دنباله سمت راست رخ می‌دهد را رد کنید، عملاً شما با سطح ۷/۵٪ کار کرده‌اید. در این وضعیت شما ۵٪ پیامدها را در یک جهت رد خواهید کرد (با فرض اینکه داده‌ها در آن سو قرار می‌گیرند)، و همچنین مایلید ۲/۵٪ از پیامدها را در سوی دیگر رد کنید (وقتی که داده‌ها در خلاف جهت مورد انتظار قرار می‌گیرند). نمی‌توان انکار کرد که $۲/۵\% = ۷/۵\% + ۵\%$. در رابطه با این روش دوم، آیا شما مایلید روی پرتاب سکه شرط‌بندی کنید وقتی که طرف مقابل "شیر" را خواهد اما در عین حال این حق را هم برای خود نگه دارد که اگر "شیر" نیامد "خط" را خواهد. این به ضرر شما خواهد شد. به همین دلیل گزینش یکی از دو آزمون یک دامنه یا دو دامنه را قبل از جمع‌آوری داده‌ها باید انجام داد. این یکی از دلایل انتخاب بیشتر آزمون دو دامنه نیز هست.

اگر چه مبحث پیش به نفع آزمون‌های دو دامنه مطرح شد و ما نیز در این کتاب خود را به چنین شیوه‌هایی محدود می‌سازیم اما قاعده قابل اتکایی در این باره وجود ندارد. تصمیم نهایی وابسته به اطلاعات شما در باره انواع خطاها است. بخاطر داشته باشید که با توجه به دنباله مورد نظر در یک توزیع، فرق آزمون یک دامنه و دو دامنه فقط این است که در دومی خط برش متفاوتی به کار می‌رود. یک آزمون دو دامنه در سطح $\alpha = ۰/۰۵$ سهل‌انگارانه‌تر از یک آزمون یک دامنه در سطح $\alpha = ۰/۰۱$ است.

اگر شما با استفاده از توزیع‌های نمونه‌گیری استدلال محکمی برای منطق فرضیه‌آزمایی داشته باشید باقی کار نسبتاً ساده است. به هر آماره تازه‌ای که برخورد می‌کنید تنها دو پرسش اساسی را باید مطرح کرد:

۱- چگونه و با چه فرضیاتی محاسبه خواهد شد؟

۲- توزیع نمونه‌گیری آن تحت H_0 چگونه خواهد بود؟

اگر پاسخ این پرسش‌ها را بدانید، شما با محاسبه آماره آزمون برای داده‌های در دست اجرا و مقایسه این آماره با توزیع نمونه‌گیری مورد نظر کار خود را انجام می‌دهید. چون توزیع‌های نمونه‌گیری مختلف در ضمیمه کتاب‌های آماری در اختیار شما قرار دارد، آنچه که شما نیاز دارید بدانید این است که چه آزمونی برای چه کار خاصی مناسب است و چگونه باید محاسبات آن را انجام دهید.

۹-۱۱ مثال کار شده پایانی

میانگین امتحان مصاحبه کارشناسی ارشد در چند سال قبل ۴۸۹ با انحراف معیار ۱۲۶ بود. این آماره بر اساس تمام دانشجویان شرکت کننده در امتحان در آن سال بود، زبان مادری اکثریت آنان انگلیسی بود. فرض کنید در آن سال ما یک تقاضا از سوی یک دانشجو با نام چینی داشته‌ایم که نمره خیلی پایینی داشته است (مثلاً، ۲۲۰). اگر زبان مادری این فرد انگلیسی بود، این نمره برای ما آن قدر پایین بود که

در مناسب بودن او برای این دوره شک کنیم، مگر اینکه سایر شرایط او خیلی بهتر باشد. به هر حال اگر دانشجو یک فرد انگلیسی زبان نباشد نمی‌توانیم نمره کم را نشان دهنده ضعف همه توانایی‌های او بدانیم. برای ما دو گزینه ممکن وجود دارد، شامل اینکه دانشجو انگلیسی زبان هست یا نیست. اگر او انگلیسی زبان باشد، ما میانگین و انحراف معیار جمعیت مشاهداتی که نمره او از آن برآمده را می‌دانیم. اینها به ترتیب ۴۸۹ و ۱۲۶ هستند. اگر انگلیسی زبان نباشد ما از جمعیتی که نمره او برآمده اطلاعی نداریم. ما فرض صفر را بر این اساس که او یک فرد انگلیسی زبان است بنا می‌کنیم، یا بطور مشخص‌تر، $\mu = 489$: H_0 و H_1 را با این فرض که این فرد انگلیسی زبان نیست تعیین می‌کنیم.

ما باید بین آزمون یک دامنه و دو دامنه دست به انتخاب بزنیم. در این مورد ما آزمون یک دامنه را بر می‌گزینیم بر این اساس که امتحان کارشناسی ارشد به زبان انگلیسی برگزار می‌شود و مشکل است که تصور کنیم یک جمعیت غیرانگلیسی زبان، میانگینی بالاتر از انگلیسی زبان‌ها در چنین امتحانی داشته باشد. چون ما یک آزمون یک دامنه را انتخاب کردیم، فرض مقابل را به این صورت طرح می‌کنیم، $H_1: \mu < 489$.

قبل از آنکه ما بتوانیم شیوه‌های آماری خود را روی این داده‌ها به کار ببندیم بایستی یک تصمیم دیگر هم بگیریم. ما باید درباره یک سطح معناداری برای آزمون تصمیم بگیریم. در این مورد تصمیم به اجرای آزمون در سطح ۵٪ می‌گیریم، بخشی از این رو که معیار را در این کتاب بر این سطح قرار دادیم، دلیل دیگر هم به خاطر نگرانی بیشتر از خطای نوع II نسبت به خطای نوع I است. چنانکه ما دچار خطای نوع I شویم و به اشتباه نتیجه‌گیری کنیم که این دانشجو یک انگلیسی زبان نیست در حالی که واقعا باشد، احتمال خیلی زیادی وجود دارد که سایر مدارک او نادیده گرفته شود. به هر حال اگر دچار خطای نوع II شویم و او را به عنوان فرد انگلیسی زبان ندانیم به او یک فرصت منصفانه داده‌ایم. در مرحله بعد ما نیاز داریم این احتمال را حساب کنیم که یک دانشجو نمره‌ای حداقل به کوچکی ۲۲۰ بدست آورد در حالی که $\mu = 489$: H_0 درست است. ابتدا نمره Z مربوط به نمره خام ۲۲۰ را حساب می‌کنیم.

$$Z = \frac{(X - \mu)}{\sigma} = \frac{(220 - 489)}{126} = \frac{-269}{126} = -2.13$$

سپس به سراغ جدول Z می‌رویم تا احتمال اینکه ما یک مقدار Z کمتر یا مساوی ۲/۱۳- بدست آوریم را حساب کنیم. از جدول خواهیم فهمید که این احتمال مساوی ۰/۰۱۷ است. از آنجا که این احتمال از سطح معناداری ۵٪ که ما برگزیدیم کمتر است، ما فرض صفر را رد خواهیم کرد، بر این اساس که احتمال زیادی وجود دارد که ما نمره‌ای به کوچکی ۲۲۰ در شرایط این امتحان بدست آوریم که نمونه ما از جمعیت انگلیسی زبان‌ها باشد. به جای آن ما به این نتیجه می‌رسیم که مشاهده‌ای از یک فرد داریم که انگلیسی زبان نیست.

توجه داشته باشید که در رد فرض صفر ما یک خطای نوع I داشتیم. ما می‌دانیم که اگر نمونه‌ای از انگلیسی زبان‌ها بگیریم $1/7\%$ آنها نمره‌ای در این حد پایین خواهند داشت. امکان دارد که داوطلب یک فرد انگلیسی زبان بوده که ضعیف عمل کرده است. تمام آنچه ما می‌گوییم این است که چنین پدیده‌ای چندان نامحتمل است که ما شرط‌بندی‌های خود را روی فرضیه مقابل خواهیم گذاشت.

۱۰-۱۱ خلاصه

هدف این فصل بررسی نظریه عمومی فرضیه‌آزمایی بوده است بدون اینکه وارد محاسبات خاص مورد نیاز برای اجرای عملی یک آزمون شویم. ما ابتدا به مفهوم توزیع نمونه‌گیری پرداختیم، که توزیعی از آماره مورد نظر است در حالی که به طور مکرر از شمار نامحدودی از نمونه‌ها تحت شرایط خاص محاسبه شود. توزیع نمونه‌گیری در اصل به ما می‌گوید وقتی شرایطی که تحت آن شرایط توزیع بدست آمده است برقرار باشد چه نوع مقادیری به طور منطقی برای آن آماره قابل انتظار است. ما سپس فرض صفر و نقش آن در فرضیه‌آزمایی را بررسی کردیم. مشاهده کردیم که می‌توان هر فرض صفری را با طرح این پرسش که اگر آن فرض درست باشد توزیع نمونه‌گیری آماره مورد نظر چگونه خواهد بود و سپس با مقایسه آماره خاص با آن توزیع، بیازماییم. بعد از آن دیدیم که چگونه با استفاده از دانش قبلی خود در باره توزیع نرمال، یک فرضیه ساده در عمل آزموده می‌شود. در پایان به خطاهای نوع I و II و آزمون‌های یک دامنه و دو دامنه پرداختیم. برخی از مطالب مهم فصل عبارتند از:

- خطای نمونه‌گیری
- فرضیه آزمایی
- توزیع‌های نمونه‌گیری
- خطای نمونه‌گیری میانگین
- فرضیه تحقیق
- فرض صفر (H_0)
- آماره‌های نمونه
- آماره‌های آزمون
- سطح رد کردن (سطح
- معناداری)
- ناحیه رد
- فرض مقابل (H_1)
- مقدار بحرانی
- خطای نوع I
- α (آلفا)
- خطای نوع II
- β (بتا)
- توان
- آزمون یک دامنه (آزمون جهت‌دار)
- آزمون دو دامنه (آزمون غیرجهت‌دار)

۱۱-۱۱ تمرین ها

۱۱-۱ فرض کنید دیشب دوست شما نتیجه بازی تیم ملی بسکتبال را به صورت نمره ۱۳-۲۶ به شما گفته باشد. احتمالاً شما در می یابید که او روزنامه را بد خوانده یا موضوع چیز دیگری بوده است. در واقع شما عملاً یک فرض صفر را آزموده و رد کردید.

(الف) فرض صفر چه بود؟

(ب) شیوه فرضیه آزمایی که بکار بردید را فهرست وار بنویسید؟

۱۱-۲ فرض کنید شما دریابید که در ترم گذشته حدود ۲ دلار برای خرید نهار روزانه خرج کردید. برای یک ترم این هزینه را در نظر بگیرید.

(الف) یک طرح اولیه از توزیع هزینه روزانه بکشید.

(ب) اگر شما بدون توجه به قبض یک اسکناس ۵ دلاری بدهید و ۲/۷۵ دلار پس بگیرید، آیا باید نگران باشید که سرتان کلاه رفته است؟

(ج) منطق مربوط به پاسخ تمرین ۱۱-۲ (ب) را توضیح دهید.

۱۱-۳ در تمرین ۱۱-۲ (ب) خطای نوع I چه خواهد بود.

۱۱-۴ در تمرین ۱۱-۲ (ب) خطای نوع II چه خواهد بود.

۱۱-۵ با استفاده از تمرین ۱۱-۲ منظور از ناحیه رد و مقدار بحرانی را بیان کنید.

۱۱-۶ چرا باید در تمرین ۱۱-۲ از آزمون یک دامنه استفاده کنیم و چه دنباله ای را باید انتخاب کنیم؟ اگر دنباله نادرست را انتخاب کنیم چه خواهد شد؟

۱۱-۷ میانگین نمره مصاحبه ورودی کارشناسی ارشد یک دانشگاه ۶۵۰ با انحراف معیار ۵۰ می باشد (با توزیع نرمال). یک دانشجوی با نمره ۴۹۰ پذیرفته شده است، که اتفاقاً مادرش از کارمندان آنجاست. آیا ممکن است پارتی بازی صورت گرفته باشد؟

۱۱-۸ چرا چنین انحراف معیار کوچکی در تمرین ۱۱-۷ منطقی به نظر می رسد؟

۱۱-۹ چرا نمرات امتحان مصاحبه کارشناسی ارشد ممکن است (یا نیست) برای نمونه انحصاری دانشجویان پذیرفته شده دارای توزیع نرمال باشند؟

۱۱-۱۰ تصور کنید که شما به تازگی یک آزمون آماری به نام آزمون نما برای بررسی این موضوع ابداع کرده اید که نمای یک جمعیت برابر مقدار خاصی است (مثلاً ۱۰۰). این آماره (M) به این صورت حساب می شود

$$M = \frac{\text{نمای نمونه}}{\text{دامنه نمونه}}$$

نحوه بدست آوردن توزیع نمونه ای M را توضیح دهید. (توجه: این یک آماره کاملاً ساختگی است.)

۱۱-۱۱ M در تمرین ۱۱-۱۰ را ما در واژه شناسی این فصل چه می نامیم؟

- ۱۱-۱۲ در زندگی روزمره موقعیتی را بیان کنید که بطور معمول فرضیه‌هایی را می‌آزماییم بدون آنکه به این واقعیت توجه داشته باشیم.
- ۱۱-۱۳ در تمرین ۱۱-۷ فرض مقابل چیست (H_1)؟
- ۱۱-۱۴ خطای نمونه‌گیری را توضیح دهید؟
- ۱۱-۱۵ تفاوت بین "توزیع" و "توزیع نمونه‌گیری" چیست؟
- ۱۱-۱۶ در جدول ۱۱-۱ کاهش α چگونه بر میزان احتمالات تأثیر می‌گذارد؟
- ۱۱-۱۷ دو نمونه از فرضیات تحقیق و فرض صفر معادل آنها را بیان کنید؟
- ۱۱-۱۸ برای توزیع شکل ۱۱-۴ گفتیم که احتمال خطای نوع II (β) معادل ۰/۷۴ است. نشان دهید این احتمال چگونه بدست آمده بود.
- ۱۱-۱۹ محاسبات تمرین ۱۱-۱۸ را با در نظر گرفتن $\alpha = 1$ از نو انجام دهید.
- ۱۱-۲۰ در مثال بخش ۱۱-۹ اگر یک آزمون دو دامنه اجرا می‌کردیم این آزمون چه تغییری می‌کرد؟

فصل ۱۲

آزمون‌های فرضیه برای میانگین‌های یک نمونه

- ۱۲-۱- توزیع نمونه‌گیری میانگین
- ۱۲-۲- آزمون فرضیه‌هایی در باره میانگین- σ معین
- ۱۲-۳- آزمون یک میانگین نمونه‌ای وقتی σ ناشناخته است (آزمون t یک نمونه‌ای)
- ۱۲-۴- یک مثال دیگر- توهم ماه
- ۱۲-۵- حدود اطمینان میانگین
- ۱۲-۶- آزمون معناداری یک ضریب همبستگی
- ۱۲-۷- عوامل مؤثر بر مقدار t یا تصمیم‌گیری در باره H_0
- ۱۲-۸- مثال کار شده پایانی

در فصل ۱۱ ما به منطق عمومی فرضیه‌آزمایی پرداختیم و محاسبات خاص آنها را نادیده گرفتیم. در این فصل می‌خواهیم نگاهی به شیوه‌های محاسباتی برای آزمون بعضی فرضیه‌های ساده بیاندازیم. بویژه تمرکز ما بر آزمون فرض صفر در باره مقدار یک میانگین جمعیتی خواهد بود. در انتها به آزمون یک فرض صفر در باره ضریب همبستگی خواهیم پرداخت.

با مثالی از فصل پیش آغاز می‌کنیم. در آنجا به موردی پرداختیم که می‌دانستیم نمرات کلی مشکلات رفتاری روی فرم خودگزارشی جوانان دارای توزیع نزدیک به نرمال با میانگین ۵۰ و انحراف معیار ۱۰ هستند. در فصل ۱۱ با آن مثال تنها به شیوه‌ای عمومی کار کردیم. اینجا آن را از سر می‌گیریم و با جزئیات خیلی بیشتری به آن می‌پردازیم.

از آنجا که در ادبیات روان‌شناسی شواهدی وجود دارد که فشار روانی در زندگی کودک ممکن است باعث مشکلات رفتاری بعدی شود، می‌توان انتظار داشت یک نمونه از کودکانی که در معرض میزان

غیرمعمول فشار بوده‌اند، سطح بالا و غیرعادی از مشکلات رفتاری را بروز دهند. از سوی دیگر ممکن است ما انتظار داشته باشیم که چنین کودکانی احساس کنند در زندگی خود به اندازه کافی وقت تلف کرده‌اند بدون هیچ پیشرفتی، و در نتیجه سطح پایین غیرعادی از مشکلات رفتاری را نشان دهند. احتمال این امکان دوم چندان زیاد نیست، اما حفاظت در برابر آن اهمیت دارد. در نتیجه ما می‌خواهیم به بررسی یک فرضیه آزمایشی (دو دامنه) پردازیم که تعداد مشکلات رفتاری در بین کودکان تحت فشار با تعداد مشکلات رفتاری در کل کودکان تفاوت دارد. به هر حال ما بطور مستقیم فرضیه آزمایشی را نمی‌آزماییم. بلکه فرض صفر (H_0) را می‌آزماییم که می‌گوید نمرات کودکان تحت فشار از جمعیت نمراتی می‌آیند که دارای میانگینی برابر با جمعیت نمرات کودکان عادی است. بطور دقیق‌تر، ما می‌خواهیم بین این دو حالت تصمیم بگیریم:

$$H_0: \mu = 100 \quad \text{و} \quad H_1: \mu \neq 100$$

ما فرض مقابل دو دامنه (H_1) را انتخاب می‌کنیم چون می‌خواهیم اگر $\mu < 50$ یا $\mu > 50$ باشد H_0 را رد کنیم. برای بررسی این مشکل، نمونه‌ای متشکل از پنج کودک بیرون کشیدیم، که هر یک از آنها تحت فشار روانی زیادی از یک نوع یا نوعی دیگر بوده‌اند. از کودکان خواسته شده تا فرم خودگزارشی جوانان را تکمیل کنند و نمرات زیر بدست آمده است

۵۱ ۶۶ ۵۳ ۶۲ ۴۸

این نمونه پنج مشاهده‌ای دارای میانگین $56/0$ و انحراف معیار $7/65$ است. بنابراین پنج کودک دارای یک نمره متوسطی هستند که از میانگین جمعیت کودکان نرمال شش نمره بالاتر است. اما از آنجا که این نتیجه بر اساس نمونه‌ای متشکل از تنها پنج کودک است، کاملاً روشن است که انحراف از 50 می‌تواند ناشی از شانس (یا به عبارتی ناشی از خطای نمونه‌گیری) باشد. پیش از آنکه هر گونه نتیجه‌گیری بتوان به عمل آوریم، لازم است ببینیم در صورتی که واقعا از یک جمعیت کودکان نرمال نمونه‌گیری کرده باشیم چه مقداری را می‌توانیم برای میانگین‌های نمونه بطور منطقی انتظار داشته باشیم.

۱-۱۲ توزیع نمونه‌گیری میانگین

چنانکه از فصل قبل به خاطر دارید توزیع نمونه‌گیری یک آماره، توزیع مقادیری از آن آماره است که انتظار داریم با بیرون کشیدن شمار نامحدودی از نمونه‌ها از آن جمعیت و محاسبه آن آماره در هر نمونه، بدست آید. چون در اینجا توجه ما روی میانگین‌های جمعیت متمرکز است، لازم است اطلاعاتی در باره توزیع نمونه‌گیری میانگین^۱ داشته باشیم. خوشبختانه همه اطلاعات مهم در باره توزیع نمونه‌گیری از

میانگین را می توان در یک قضیه مهم بنام **قضیه حد مرکزی**^۱ جمع کرد. قضیه حد مرکزی یک بیان واقعی در باره توزیع میانگین ها است. به این شرح:

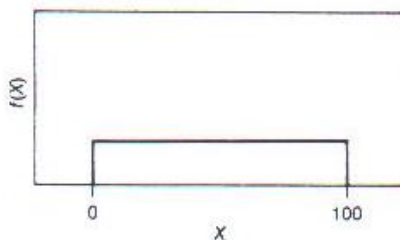
یک جمعیت معین با میانگین μ و واریانس σ^2 را در نظر بگیرید، توزیع نمونه گیری میانگین (توزیع میانگین های نمونه ای) دارای میانگینی برابر μ (یعنی $\mu_{\bar{X}} = \mu$)، و یک واریانس $(\sigma_{\bar{X}}^2)$ معادل σ^2/N (و انحراف معیاری $[\sigma_{\bar{X}}]$ معادل $(\sigma/\sqrt{N})$) است. با افزایش N ، اندازه نمونه، این توزیع به توزیع نرمال نزدیک خواهد شد.

این قضیه یکی از مهم ترین قضایای ریاضی است، زیرا نه تنها به ما می گوید میانگین و واریانس توزیع نمونه گیری از میانگین نمونه برای هر اندازه نمونه باید چقدر باشد بلکه همچنین بیان می کند که با افزایش N ، شکل این توزیع نمونه ای نیز به شکل نرمال میل می کند، خواه شکل جمعیت اصلی نرمال باشد یا نباشد. اهمیت این حقایق را به اختصار توضیح می دهیم.

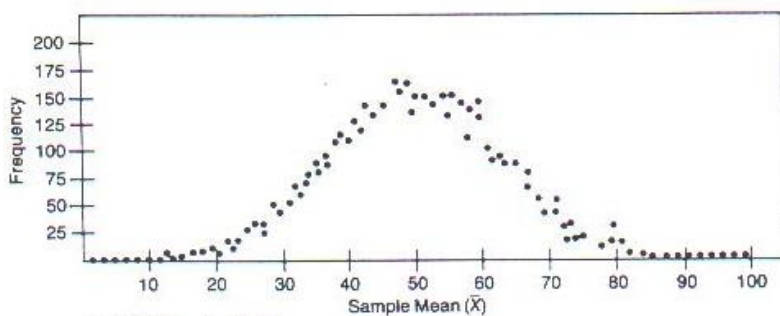
میزانی که توزیع نمونه گیری میانگین به توزیع نرمال میل می کند تابعی از شکل جمعیت اصلی است. اگر این جمعیت خودش نرمال باشد، شکل توزیع نمونه گیری میانگین هم بدون توجه به N نرمال خواهد بود. اگر این جمعیت متقارن باشد اما نرمال نباشد، توزیع نمونه گیری از میانگین، حتی برای اندازه های نمونه ای کاملاً کوچک تقریباً نرمال خواهد بود، بخصوص اگر جمعیت تک نمایی باشد. چنانچه جمعیت خیلی دارای کجی (چولگی) باشد، ممکن است لازم باشد اندازه ی نمونه ما ۳۰ یا بیشتر باشد تا میانگین ها خیلی به توزیع نرمال نزدیک شوند.

برای تجسم قضیه حد مرکزی فعلاً مثال مشکلات رفتاری را کنار می گذاریم و به سراغ یک مثال عمومی تر می رویم. فرض کنید که ما یک جمعیت بزرگ نامحدود از اعداد بطور تصادفی توزیع شده بین ۰ تا ۱۰۰ داریم. این جمعیت دارای توزیعی معروف به توزیع مستطیلی است - هر مقداری بین ۰ تا ۱۰۰ دارای احتمال یکسان است. توزیع این جمعیت در شکل ۱-۱۲ نشان داده شده است. در این مورد میانگین (μ) برابر ۵۰، انحراف معیار (σ) برابر ۲۸/۸۷، و واریانس (σ^2) برابر ۸۳۳/۳۳ است.

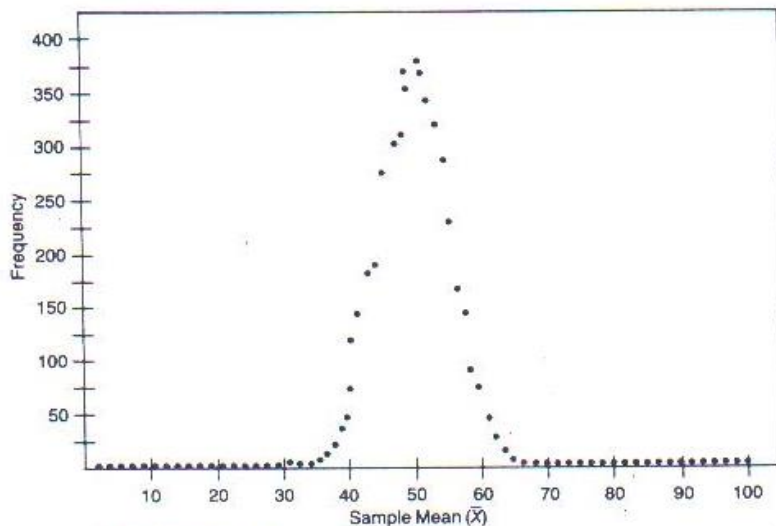
اکنون فرض کنید که ما ۵۰۰۰ نمونه ۵ تایی از این جمعیت بیرون بکشیم و نتایج میانگین های نمونه ای را رسم کنیم. چنین کاری را با یک کامپیوتر براحتی می توان انجام داد، و در حقیقت نتیجه چنین شیوه ای درست شبیه به شکل ۲-۱۲ (الف) است. این شکل نشان می دهد که میانگین ها هر چند کاملاً به صورت نرمال توزیع نشده اند، دست کم در مرکز انباشته و در دو سوی توزیع کم می شوند. اگر شما سعی کنید میانگین و واریانس این توزیع را حساب کنید در خواهید یافت که آنها بسیار نزدیک به $\mu = 50$ و $\sigma_{\bar{X}}^2 = \sigma^2/N = 833/33 = 166/3$ هستند.



شکل ۱۲-۱ توزیع مستطیلی با $\mu = 50$ و $\sigma = 28/87$



(a) 5000 Samples $N = 5$



(b) 5000 Samples $N = 30$

شکل ۱۲-۲ توزیع‌های نمونه‌گیری کامپیوتری از میانگین

حالا فرض کنید که ما کل این فرایند را تکرار کنیم- اما این بار ۵۰۰۰ نمونه، هر بار با ۳۰ مشاهده. دو باره بطور واقعی ما این کار را با یک کامپیوتر انجام داده‌ایم، و نتایج در شکل ۲-۲ (ب) رسم شده است. در اینجا شما می‌بینید که درست مانند آنچه قضیه حد مرکزی پیش‌بینی می‌کند، توزیع تقریباً نرمال است، باز هم میانگین (μ) برابر ۵۰ است، و واریانس به $27/78 = 833/33 / 30$ کاهش یافته است.

۱۲-۲ آزمون فرضیه در باره میانگین‌ها - σ معلوم

از قضیه حد مرکزی همه ویژگی‌های مهم توزیع نمونه‌گیری از میانگین را می‌فهمیم. (شکل آن، میانگین و انحراف معیار آن را می‌دانیم). براساس این اطلاعات ما در موقعیتی هستیم که شروع به آزمون فرضیه‌های مربوط به میانگین‌ها کنیم. به منظور حفظ پیوستگی بهتر است به بحث قبلی در باره توزیع نرمال برگردیم. در فصل ۱۱ دیدیم که می‌توان یک فرضیه را در باره جمعیت آزمون کرد که از آن با محاسبه z یک نمره منفرد (در آنجا یک نمره ضربه انگشت) بدست آید

$$z = \frac{(X - \mu)}{\sigma}$$

و سپس احتمال یک مقدار z به کوچکی چیزی که با استفاده از جدول توزیع استاندارد نرمال حاصل می‌شود را بدست آورد. بنابراین ما آزمون یک دامنه‌ای روی فرضیه مربوط به نرخ ضربه‌زدن با انگشت (۷۰) را برای فردی که به صورت تصادفی از یک جمعیت سالم دارای توزیع نرمال با میانگین نرخ‌های ضربه زدن ۱۰۰ و انحراف معیار ۲۰ بر آمده است اجرا می‌کنیم. این کار را با محاسبه

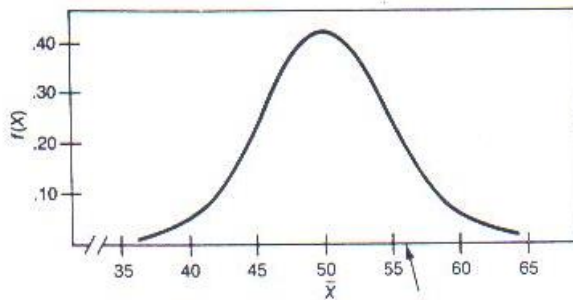
$$z = \frac{(X - \mu)}{\sigma} = \frac{70 - 100}{20} = \frac{-30}{20} = -1.5$$

انجام دادیم و سپس با استفاده از جدول سطوح زیر منحنی طبیعی (پیوست کتاب)، سطح پایین‌تر از $z = -1.5$ را پیدا کردیم. این مقدار برابر ۰/۰۶۶۸ بود. بنابراین تقریباً فقط در ۷٪ موارد ما انتظار نمره‌ای به این کوچکی را خواهیم داشت اگر نمونه‌گیری از یک جمعیت سالم به عمل آمده باشد. از آنجا که این احتمال بزرگتر از سطح معناداری مورد نظر ما $\alpha = 0.05$ بود، ما نمی‌توانستیم فرض صفر را رد کنیم. نرخ ضربه برای شخص مورد آزمایش ما، یک نرخ غیرعادی برای افراد سالم نبود. هر چند در این مثال ما به آزمون یک فرضیه در باره یک مشاهده منفرد پرداختیم، همین منطق دقیقاً برای آزمون فرضیه‌های مربوط به میانگین‌های نمونه‌ای به کار می‌رود.

در اغلب موقعیت‌هایی که ما یک فرضیه را در باره یک میانگین جمعیتی آزمایش می‌کنیم ما در باره واریانس جمعیت هیچگونه اطلاعاتی نداریم. (دلیل اصلی این که ما آزمون‌های t را داریم که تمرکز این فصل روی آنهاست همین است). اما در شمار محدودی از موقعیت‌ها ما بدلالی σ را می‌دانیم، و این فرصت یک بحث فرضیه آزمایی را برای وقتی σ معلوم است فراهم می‌نماید، که زمینه خوبی است برای انتقال از آنچه ما از قبل در باره توزیع نرمال می‌دانیم به آنچه می‌خواهیم در باره آزمون‌های t بدانیم.

مثال مشکلات رفتاری نمونه خوبی برای این منظور است چون ما میانگین و انحراف معیار جمعیت برای نمرات مشکلات کلی رفتار را می‌دانیم ($\mu = 50$ و $\sigma = 10$). ما همچنین می‌دانیم که یک نمونه تصادفی از کودکان تحت فشار روانی دارای میانگین نمره $56/0$ هستند و می‌خواهیم فرض صفر مبنی بر این که این پنج کودک یک نمونه تصادفی از جمعیت کودکان نرمال (یعنی نرمال با توجه به سطح عمومی مشکلات رفتاری) هستند را بیازماییم. به عبارت دیگر ما می‌خواهیم فرض $H_0: \mu = 50$ را در برابر فرض $H_1: \mu \neq 50$ آزمایش کنیم.

از آنجا که ما میانگین و انحراف معیار جمعیت نمرات کلی مشکلات رفتاری را می‌دانیم، ما می‌توانیم از قضیه حد مرکزی برای بدست آوردن توزیع نمونه‌گیری وقتی فرض صفر درست است استفاده کنیم. قضیه حد مرکزی می‌گوید اگر ما توزیع نمونه‌گیری از میانگین را از جمعیت بدست آوریم، این توزیع دارای میانگین 50 ، و واریانس $\sigma^2/N = 10^2/5 = 100/5 = 20$ ، و واریانس $\sigma/\sqrt{N} = 4/47 \approx 0.085$ است به خطای معیار^۱ است به خطای معیار $\sigma/\sqrt{N} = 4/47 \approx 0.085$ خواهد بود. این توزیع در شکل ۳-۱۲ به نمایش در آمده است. علامت کمان در این شکل نشان دهنده نقطه میانگین نمونه است.



شکل ۳-۱۲ توزیع نمونه‌گیری میانگین برای نمونه‌های $N = 5$ برآمده از جمعیتی با $\mu = 50$ و $\sigma = 10$

با توجه به اینکه می‌دانیم این توزیع نمونه‌گیری دارای توزیع نرمال با میانگین 50 و انحراف معیار $4/47$ است، ما می‌توانیم سطح زیر توزیع را با رجوع به جداول استاندارد توزیع نرمال پیدا کنیم. بنابراین، مثلاً چون دو خطای معیار برابر $2(4/47) = 8/47 \approx 0.17$ می‌باشد، بطور ساده سطح جانب راست $\bar{X} = 58/94$ همان سطح زیر توزیع نرمال، واری دو انحراف معیار بالاتر از میانگین است.

1. standard error

❖ انحراف معیار هر توزیع نمونه‌گیری معمولاً به عنوان خطای معیار آن توزیع به حساب می‌آید. بنابر این انحراف معیار میانگین‌ها را خطای معیار میانگین (با نماد $\sigma_{\bar{X}}$) گویند، در حالی که انحراف معیار تفاوت بین میانگین‌ها که در فصل ۱۴ مورد بحث قرار خواهد گرفت خطای معیار تفاوت بین میانگین‌ها خوانده می‌شود که به صورت $\sigma_{\bar{X}_1 - \bar{X}_2}$ نشان داده می‌شود. تغییرات جزئی واژه‌شناسی مانند نامیدن انحراف معیار بصورت خطای معیار نباید دانشجویان را گیج کند.

برای شرایط خاص ابتدا لازم است ما احتمال یک میانگین نمونه بزرگتر یا مساوی ۵۶ را بدانیم و از این رو باید سطح بالای $\bar{X} = ۵۶$ را پیدا کنیم. تنها با تغییر کوچکی در فرمول z ما می توانیم این را حساب کنیم، به شیوه ای شبیه به آنچه برای مشاهدات فردی انجام دادیم.

$$z = \frac{X - \mu}{\sigma} \quad \text{می شود} \quad z = \frac{X - \mu}{\sigma_{\bar{X}}}$$

که می توان آن را به این صورت هم نوشت

$$\frac{X - \mu}{\frac{\sigma}{\sqrt{N}}}$$

برای داده های ما این می شود

$$\frac{56 - 50}{4.47} = \frac{6}{4.47} = ۱/۳۴$$

توجه دارید که این معادله z همانند فرمول قبلی آن است. تنها فرق آنها در جایگزینی $\bar{X}$ به جای X و $\sigma_{\bar{X}}$ به جای σ است. این فرق ها به خاطر این است که ما حالا با توزیع میانگین ها سروکار داریم، و بنابراین نقاط داده ای اینک میانگین ها هستند، و انحراف معیار مورد سؤال در اینجا خطای معیار میانگین (انحراف معیار میانگین ها) می باشد. فرمول z همچنان نشان دهنده (۱) یک نقطه روی توزیع، منهای (۲) میانگین آن توزیع، حاصل تقسیم بر (۳) انحراف معیار توزیع، خواهد بود. اکنون به جای تمرکز خاص روی توزیع $\bar{X}$ ، به توضیح آن به صورت میانگین نمونه به عنوان نمرات z می پردازیم و حالا می توانیم به این سؤال با توجه به توزیع نرمال استاندارد پاسخ دهیم.

از جدول سطوح زیر منحنی (در بخش پیوست کتاب) احتمال یک نمره z به بزرگی $۱/۳۴$ را برابر $۰/۰۹۰۱$ پیدا می کنیم. چون یک آزمون دو دامنه از H_0 را در نظر داریم، لازم است احتمال را دو برابر کنیم تا احتمال یک انحراف به بزرگی $۱/۳۴$ خطای معیار در هر جهت از میانگین بدست آید. این معادل $۰/۱۸۰۲ = ۲(۰/۰۹۰۱)$ می شود. بنابراین با یک آزمون دو دامنه (که تأکید می کند که کودکان تحت فشار روانی دارای یک نمره میانگین مشکلات رفتاری هستند که در هر جهت از کودکان عادی متفاوت است) در سطح معناداری $۰/۰۵$ ما H_0 را رد می کنیم، زیرا احتمال بدست آمده بزرگتر از $۰/۰۵$ است. ما نتیجه خواهیم گرفت که شاهی برای اینکه کودکان تحت فشار روانی مشکلات رفتاری کمتر یا بیشتر از سایر کودکان داشته باشند در اختیار نداریم.

آزمودن یک میانگین نمونه ای در مقابل یک میانگین جمعیتی که در اینجا انجام دادیم، بر اساس این فرض است که میانگین های نمونه ای دارای توزیع نرمال هستند، یا حداقل این توزیع به اندازه کافی نرمال است که ما وقتی به جداول توزیع نرمال استاندارد رجوع می کنیم با خطای ناچیزی مواجه باشیم. بسیاری کتاب ها می گویند که فرض می کنیم که از یک جمعیت نرمال نمونه گیری می کنیم (مثلا نمرات مشکلات رفتاری خود دارای توزیع نرمال هستند)، اما این موضوع در عمل خیلی ضرورت ندارد. آنچه باقی می ماند این است که فرض کنیم توزیع نمونه گیری میانگین (شکل ۳-۱۲) نرمال است. این فرض

می‌تواند به دو روش تأمین گردد- به این شرط که یا (۱) جمعیتی که از آن نمونه‌گیری می‌کنیم نرمال باشد یا (۲) اندازه نمونه به قدر کافی بزرگ باشد تا حداقل نرمال بودن تقریبی را به شیوه قضیه حد مرکزی ایجاد کند. این یکی از مزیت‌های بزرگ قضیه حد مرکزی است: این قضیه به ما امکان می‌دهد که فرضیه‌های را بیازمائیم حتی اگر جمعیت اصلی نرمال نباشد، تنها چیز تأمین کننده آن است که N به قدر کافی بزرگ است.

۳-۱۲ آزمودن یک میانگین نمونه‌ای وقتی σ ناشناخته است (آزمون t تک-نمونه‌ای)

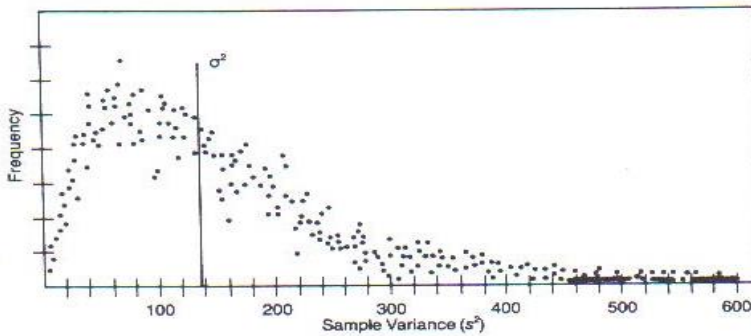
مثال قبلی را با دقت از میان تعداد محدودی موقعیت انتخاب کردیم که در آنها انحراف معیار جمعیت (σ) معلوم است. اما عموماً ما مقدار σ را نمی‌دانیم و معمولاً مجبوریم آن را از روی انحراف معیار نمونه (S) برآورد کنیم. وقتی ما σ را در فرمول با S جایگزین می‌کنیم به هر حال ماهیت آزمون تغییر می‌کند. ما دیگر نمی‌توانیم بگوییم که جواب، یک نمره z است و با مراجعه به جدول z آن را ارزیابی کنیم. بلکه ما جواب را با t نشان می‌دهیم و آن را با توجه به جدول t ارزیابی می‌کنیم که قدری متفاوت است. استدلال مربوط به جایگزینی t به جای z خیلی ساده است، با این حال بسیاری کتاب‌ها از آن طفره می‌روند. مشکل اساسی که نیاز به این تغییر را سبب می‌شود به توزیع نمونه‌گیری واریانس نمونه مربوط می‌شود.

۳-۱۲-۱ توزیع نمونه‌گیری S^2

از آنجا که آزمون t از S^2 به عنوان یک برآورد کننده σ^2 استفاده می‌کند، لازم است ابتدا نگاهی به توزیع نمونه‌گیری S^2 بیاندازیم. این توزیع نمونه‌گیری بینشی در باره مشکلاتی که در پیش رو داریم به ما می‌دهد. در فصل ۶ دیدیم که S^2 یک برآورد غیرسوگیرانه^۱ از σ^2 است، به این معنی که با تکرار نمونه‌گیری، معدل مقدار S^2 برابر با σ^2 خواهد بود. هر چند یک برآورد کننده غیرسوگیرانه چیز مناسبی است اما همه چیز نیست. مشکل آن است که توزیع نمونه‌گیری S^2 بخصوص برای نمونه‌های کوچک، به شکل کاملاً چوله‌ی مثبت است. (در واقع شبیه توزیع مجذور خی است، که در فصل ۱۹ توضیح داده خواهد شد.) نمونه‌ای از توزیع نمونه‌ای S^2 به عنوان خروجی برنامه‌های کامپیوتری در شکل ۴-۱۲ نشان داده شده است (جایی که $\sigma^2 = ۱۳۸/۸۹$). به خاطر چولگی توزیع، یک مقدار منفرد S^2 ، متمایل به برآورد پایین‌تر σ^2 است، و بخصوص برای نمونه‌های کوچک این موضوع مطرح است. (S^2 غیرسوگیرانه باقی می‌ماند زیرا وقتی σ^2 را بالا برآورد می‌کند این کار را به گونه‌ای انجام می‌دهد که موارد پرشمار را نامتعادل‌تر اما موارد کم شدت پایین‌تر برآورد شوند.) در نتیجه این چولگی، مقدار t بزرگتر از مقدار z است که با دانستن σ بدست

1 unbiased stimate

می‌آوردیم، زیرا هر واریانس نمونه (S^2) دارای احتمالی بالاتر از ۵۰:۵۰ شانس برآورد واریانس جمعیت (σ^2) است.



شکل ۱۲-۴ توزیع نمونه‌گیری S^2 از یک جمعیت دارای توزیع نرمال با $\mu = ۵۰$ ، $\sigma^2 = ۱۳۸/۸۹$ و $N = ۵$

۳-۱۲-۲ آماره t

فرمولی را که پیش‌تر برای z بسط دادیم را بازنویسی می‌کنیم

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{N}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{N}}}$$

و S^2 را جایگزین می‌کنیم

$$z = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{N}}}$$

از آنجا که ما می‌دانیم S^2 برای هر نمونه خاص، به احتمال زیاد کوچکتر از مقدار مناسب σ^2 است، می‌توان دید که فرمول t به احتمال زیاد یک جواب بزرگتر بدست می‌دهد نسبت به زمانی که برای حل کردن t از خود σ^2 استفاده کنیم. از این‌رو در واقع درست نیست که جواب را یک نمره z به حساب بیاوریم و از جدول z استفاده کنیم. چنین کاری نتایج “معنادار” خیلی زیادی می‌دهد-یعنی وقتی ما در سطح معناداری $\alpha = ۰/۰۵$ آزمون به عمل می‌آوریم، بیش از ۵٪ دچار خطای نوع I می‌شویم. (برای مثال، وقتی z را محاسبه می‌کردیم ما H_0 را در سطح معناداری $۰/۰۵$ رد کردیم در حالی که z خارج از محدوده $\pm ۱/۹۶$ واقع می‌شد. اگر ما شرایطی ایجاد کنیم که در آن H_0 درست باشد، نمونه‌های $N = ۶$ مکرر بگیریم، S^2 را به جای σ^2 به کار ببریم، و t را محاسبه کنیم، در بیش از ۱۰٪ موارد یک مقدار $\pm ۱/۹۶$ یا بزرگتر بدست خواهیم آورد).

راه حل این مشکل را ویلیام گوست ارائه کرده است که برای کمپانی آبجوسازی گینس کار می‌کرد و با نام مستعار استیودنت مقاله می‌نوشت. گوست نشان داد که استفاده از S^2 به جای σ^2 منجر به یک نمونه‌گیری خاص می‌شود، که اکنون به نام معروف **توزیع t استودنت** شناخته می‌شود. در نتیجه کار گوست، تمام کاری که باید انجام دهیم چسبیدن به S^2 ، نشان دادن پاسخ با t و ارزیابی t با توجه به توزیع خودش است، همچنان که z را با توجه به توزیع نرمال ارزیابی می‌کردیم. توزیع t نیز در بخش ضمیمه کتاب‌های آماری مانند این کتاب درج گردیده، و نمونه‌هایی از توزیع واقعی t برای اندازه‌های مختلف نمونه در شکل ۵-۱۲ ترسیم شده است.

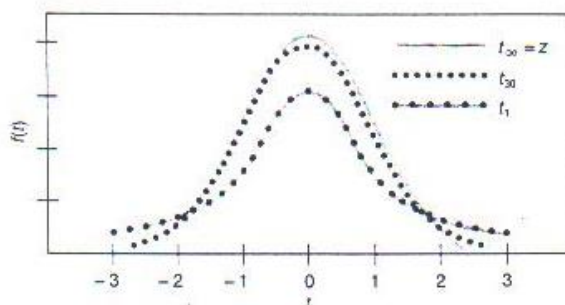
همانگونه که در شکل ۵-۱۲ مشاهده می‌شود، توزیع t تابع **درجات آزادی** $^1(df)$ تغییر می‌کند. درجات آزادی را فعلاً ما به عنوان یکی کمتر از تعداد مشاهدات در نمونه تعریف می‌کنیم. از آنجا که چولگی توزیع نمونه‌گیری S^2 با افزایش تعداد درجات آزادی از بین می‌رود، گرایش s به برآورد پایین σ نیز از بین خواهد رفت. بنابراین برای درجات آزادی بی‌نهایت، t دارای توزیع نرمال و معادل z خواهد بود.

گرچه برخی کتاب‌ها استفاده از t را وقتی $N < 30$ توصیه می‌کنند، و زمانی که $N > 30$ باشد $t=Z$ می‌گیرند، این رویکرد در این کتاب مد نظر نیست. مزیتی برای استفاده از z به عنوان یک تقریب وقتی که توزیع دقیق (مثل t) در جداول مشابهی داده شده وجود ندارد.

۳-۱۲-۳ درجات آزادی

چنان که گفته شد توزیع t تابعی از درجات آزادی است. برای حالت یک نمونه‌ای، $df = N - 1$ یک درجه آزادی کم می‌شود زیرا ما میانگین نمونه را در محاسبه S^2 به کار می‌گیریم. به بیان دقیق‌تر، ما واریانس (S^2) را با محاسبه انحراف مشاهدات از میانگین آنها ($X - \bar{X}$)، به جای انحراف از میانگین جامعه ($X - \mu$) بدست می‌آوریم. چون حاصل جمع انحرافات از میانگین ($\sum (X - \bar{X})$) همواره برابر ۰ است، تنها ۱- N مورد از انحرافات برای تغییر آزاد هستند (مورد N ام وقتی حاصل جمع صفر باشد خود بخود معلوم است). برای فهم بهتر این مطلب، پنج نمره را در نظر بگیرید که میانگین آنها ۱۰ است. چهار نمره از اینها را شما به دلخواه می‌توانید تعیین کنید (مثل، ۱۸، ۱۸، ۱۶، ۲)، اما نمره پنجم را نمی‌توان آزادانه تعیین کرد. این نمره بایستی ۴- باشد تا میانگین ۱۰ شود. به عبارت دیگر وقتی میانگین تعیین شده باشد (یعنی یک محدودیت معین شده که میانگین برابر عدد خاصی باشد) تنها می‌توان چهار نمره را در این مجموعه پنج نمره‌ای آزادانه انتخاب کرد که وقتی معین شدند عدد پنجم خود بخود معلوم است، و بنابراین ما چهار درجه آزادی داریم. به این دلیل است که در مخرج فرمول S^2 (که در فصل ۶ توضیح

داده شد) از $N-1$ استفاده شده است. به دلیل اینکه S^2 بر مبنای $N-1$ df است، برای t ما $N-1$ درجه آزادی داریم.



شکل ۵-۱۲ توزیع t برای درجات آزادی ۱، ۳۰، و ∞

۳-۱۲-۴ مثالی از کاربرد t : نمرات آزمون شایستگی تحصیل در دبیرستان

در اینجا یک مثال ساده به درک مفهوم t کمک می‌کند. اخیراً تأکید متولیان آموزشی آمریکا بر این بوده است که دانش‌آموزان پیش از دیپلم به سطح شایستگی بیشتری برسند. در این راستا و در دفاع از برنامه‌های حمایتی اولیای یک مدرسه بزرگ، نمرات ۲۰ نفر از دانش‌آموزان سال‌های بالا در آزمون استاندارد شایستگی جمع‌آوری شد. میانگین نمرات سال گذشته در کل مدرسه (جمعیت مورد نظر) $76/0$ بود. در سال جاری میانگین ($\bar{X}$) نمونه ۲۰ نفری دانش‌آموزان $79/3$ با انحراف معیار $6/4$ بود. آیا این داده‌ها می‌تواند گواهی بر این باشد که در میانگین نمره شایستگی پیشرفتی حاصل شده است، یا این تنها یک نتیجه شانس است که از جمعیتی با $\mu = 76/0$ (که میانگین سال گذشته برای نمرات جمعیت دانش‌آموزان بود) می‌توان انتظار داشت؟ برای بیان این مسأله در قالب فرضیه آزمایی،

$$H_0: \mu = 76$$

$$H_1: \mu \neq 76$$

ما یک آزمون دو دامنه به کار خواهیم برد (زیرا آموزش می‌تواند باعث انحراف افکار و دور کردن توجهات از تلاش‌های سازنده‌تر شود) و در سطح معناداری 5% کار خواهیم کرد. در مبحث قبلی داشتیم

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

توجه داشته باشید که صورت فرمول t اختلاف بین میانگین نمونه و میانگین جمعیت را بر اساس H_0 نشان می‌دهد و مخرج فرمول برآوردی از انحراف معیار توزیع میانگین‌های نمونه‌ای است. این مشابه

چیزی است که ما در باره z داشتیم، به جز اینکه واریانس (یا انحراف معیار) نمونه به جای واریانس (یا انحراف معیار) جمعیت گذاشته شده است.

برای داده‌های ما

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} = \frac{3/79 - 76}{\frac{6/40}{\sqrt{20}}}$$

$$\frac{3/3}{1/43} = 2/31$$

یک مقدار t معادل $2/31$ به خودی خود چندان اهمیت ندارد مگر آنکه بتوانیم آن را در برابر توزیع نمونه‌گیری t ارزیابی کنیم. برای این منظور این مقادیر در جدولی تحت عنوان مقادیر بحرانی t (در بخش پیوست کتاب) درج شده‌اند که بخشی از آن را در جدول ۱-۱۲ مشاهده می‌کنید. این جدول به لحاظ شکلی با جدول توزیع نرمال (z) فرق دارد. زیرا به جای ارائه سطوح بالا و پایین هر مقدار ویژه t ، که جای زیادی می‌گیرد، در این جدول مقادیر t که سطوح خاصی را نشان می‌دهد - مثل سطح $0/05$ و $0/01$ - معناداری آمده است. همچنین برخلاف z ، برای هر تعداد ممکن از درجات آزادی یک توزیع متفاوت تعریف شده است. ما می‌خواهیم در سطح $0/05$ دو دامنه کار کنیم. مقدار بحرانی معمولاً با t_{α} یا در این مورد با $t_{0.05}$ نشان داده می‌شود.

برای استفاده از جداول t ما بایستی با یک درجه آزادی مناسب وارد جدول شویم. چون ما در این مثال 20 مشاهده داریم، در نتیجه $19 = 20 - 1 = N - 1$ درجه آزادی (df) برای این داده‌ها داریم، و از آنجا که ما می‌خواهیم $t_{0.05}$ باشد، در جدول t یا جدول ۱-۱۲ مقدار بحرانی ($t_{0.05}(19)$) معادل $\pm 2/093$ است. (عدد داخل پرانتز معرف درجات آزادی است). تفسیر این نتیجه این است که اگر H_0 درست باشد، فقط در 5% موارد از نمونه‌ای با 20 مورد یک t بدست می‌آید که خارج از $\pm 2/093$ واقع شود. چون مقداری که ما محاسبه کردیم ($2/31$) بزرگتر از $2/093$ است، ما فرض صفر را رد خواهیم کرد و نتیجه خواهیم گرفت که اگر در عمل از یک جمعیت با $\mu = 76$ نمونه‌گیری کرده باشیم، احتمال یک میانگین نمونه‌ای $79/3$ وجود ندارد. پیامد رد H_0 در عمل آن است که نتیجه می‌گیریم این میانگین عملکرد در امتحان شایستگی در طی سال گذشته شکل گرفته است. اینکه چنین تغییری را باید به حساب افزایش هزینه‌ها یا عوامل ناشناخته دیگر گذاشت هنوز جواب داده نشده است، زیرا آزمون ما فقط به این می‌پردازد که آیا میانگین تغییر کرده یا نه، و با علت تغییر کاری ندارد. برای بررسی اینکه امسال کلاس یاد شده واقعاً باهوش‌تر از سال پیش بوده یا اینکه امتحان ساده‌تر بوده، نیاز به یک گروه گواه (کنترل) بود که عملکرد آزمایشی روی آن اجرا نشده باشد. در هر حال علت هر چه باشد در این سال دانش‌آموزان بطور پایا (معناداری) بهتر از سال قبل عمل کردند.

جدول ۱۲-۱ خلاصه‌ای از جدول توزیع t درج شده در بخش پیوست

df	Level of Significance for One-Tailed Test							
	.25	.20	.15	.10	.05	.025	.01	.005
	Level of Significance for Two-Tailed Test							
	.50	.40	.30	.20	.10	.05	.02	.01
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
...	...	...	...	...	...	...	...	...
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797

۱۲-۴ عوامل تأثیرگذار بر مقدار t یا تصمیم در باره H_0

بعضی عامل‌ها بر مقدار آماره t و/یا احتمال رد H_0 اثر می‌گذارند. به این شرح:

۱. اختلاف واقعی بدست آمده ($\bar{X} - \mu$)

۲. مقدار واریانس نمونه (S^2)

۳. اندازه نمونه (N)

۴. سطح معناداری (α)

۵. اینکه آزمون یک دامنه باشد یا دو دامنه

اهمیت اختلاف بدست آمده بین $\bar{X}$ و μ باید روشن باشد. این اهمیت مستقیماً به این واقعیت برمی‌گردد که هر چه صورت کسر بزرگتر باشد، مقدار t بزرگتر است. اما در عین حال باید بخاطر داشته باشیم که مقدار $\bar{X}$ در سطح وسیع تابعی است از میانگین جمعیتی که از آن نمونه‌گیری شده است. اگر این میانگین را با μ_1 و میانگین مربوط به H_0 را با μ_0 نشان دهیم، آنگاه احتمال بدست آوردن یک نتیجه معنادار با افزایش $\mu_1 - \mu_0$ افزایش می‌یابد.

اگر به فرمول t نگاه کنید خواهید دید که با کاهش S^2 یا افزایش N مخرج کسر کاهش پیدا می‌کند و در نتیجه مقدار t افزایش می‌یابد. از آنجا که تغییرپذیری تحت تأثیر خود مجموعه آزمایشی (مثلاً تحت تأثیر آموزش‌های مبهم، ثبت نادرست داده‌ها، شرایط مخدوش آزمون و غیره) نسبت به تغییرپذیری بین‌آزمودنی‌ها برتری دارد، ما تلاش می‌کنیم تا S را با کنترل هر چه بیشتر منابع تغییرپذیری کاهش

دهیم. همچنین بر اساس این واقعیت که افزایش N باعث کاهش $S_{\bar{X}}$ می‌شود، ما تعداد هرچه بیشتری از آزمودنی‌ها را به کار می‌گیریم.

در نهایت باید توجه داشت که احتمال رد H_0 بستگی به اندازه ناحیه رد کردن دارد، که خود بستگی به α و موقعیت آن ناحیه (یک دامنه بودن یا دو دامنه بودن آزمون) دارد. نقشی که همه این عامل‌ها در آزمون فرضیه دارند در فصل ۱۵ بیشتر توضیح داده خواهد شد.

۵-۱۲ یک مثال دیگر - توهم ماه

خوب است که به مثال دیگری توجه کنید که از یک مقاله قدیمی توسطه کافمن و راک (۱۹۶۲) در باره توهم ماه آورده‌ایم. کافمن و راک دریافتند که توهم ماه (واقعه همگانی بزرگتر دیدن ماه در نزدیکی افق، نسبت به وقتی که در وسط آسمان است) را می‌توان براساس فاصله ظاهری بیشتر ماه وقتی در افق است توضیح داد. به عنوان بخشی از یک مجموعه کامل آزمایشات، این محققان برای بررسی توهم ماه در ابتدا از آزمودنی‌ها خواستند تا با متغیر "ماه" در حالت ظاهر شدن در افق طوری هماهنگ شوند که با اندازه‌ی استاندارد "ماه" در وسط آسمان جور باشد، یا برعکس. (در این اندازه‌گیری آنها از ماه واقعی استفاده نکردند، بلکه یک شکل ساختگی با نمایی خاص استفاده کردند). یکی از اولین سؤال‌هایی که ممکن است پیش آید این است که آیا واقعا یک توهم ماه وجود دارد- یعنی آیا یک اندازه‌ی بزرگتر برای تطبیق ماه در افق نسبت به ماه در میان آسمان لازم است. داده‌های زیر برای ۱۰ آزمودنی از مقاله کافمن و راک گرفته شده است و نشان دهنده نسبت قطر این متغیر و ماه‌های استاندارد است. یک نسبت ۱/۰۰ نشان دهنده نبود توهم است در حالی که نسبت‌های دیگر گویای توهم است. (برای مثال، یک نسبت ۱/۵ به این معنی خواهد بود که ماه در افق با قطری ۱/۵ برابر قطر ماه در وسط آسمان به نظر آمده است). شواهد حمایت کننده از وجود توهم نیازمند آن است که ما $H_0: \mu = 1/00$ را به نفع $H_1: \mu \neq 1/00$ رد کنیم.

نسبت‌های بدست آمده: ۱/۷۳ ۱/۰۶ ۰/۳۲ ۱/۴۰ ۰/۹۵ ۱/۱۳ ۱/۴۱ ۱/۷۳ ۱/۶۳ ۱/۵۶

برای این داده‌ها $N = 10$ ، $\bar{X} = 1/463$ و $S = 0/341$ است. یک آزمون t روی $H_0: \mu = 1/00$ به شرح زیر بدست آمد

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}} = \frac{1.463 - 1.000}{\frac{0.341}{\sqrt{10}}} = \frac{0.643}{0.108} = 4/29$$

در جدول مقادیر بحرانی t پیوست کتاب برای $1 - \alpha = 0.05$ درجه آزادی با آزمون دو دامنه در سطح $\alpha = 0/05$ مقدار بحرانی $t_{0.05(9)} = \pm 2/262$ را مشاهده می‌کنیم. مقدار t بدست آمده (که با t_{obs} نمادگذاری می‌شود) برابر $4/29$ بود. چون $4/29 > 2/262$ است، ما می‌توانیم H_0 را در سطح $\alpha = 0/05$ رد کنیم و

نتیجه بگیریم که نسبت حقیقی ماه تحت این شرایط برابر با ۱/۰۰ نیست. در حقیقت این نسبت بزرگتر از ۱/۰۰ است، که همان چیزی است که ما بر اساس تجارب خود انتظار داریم.

۱۲-۶ حدود اطمینان اطراف میانگین

مثال پیش در باره توهم ماه یک نمونه عالی از مواردی است که ما علاقه مند به برآورد مقدار حقیقی μ هستیم - در اینجا نسبت حقیقی اندازه‌ی ادراک شده ماه در افق به اندازه ادراک شده ماه در میان آسمان. میانگین نمونه ($\bar{X}$) چنانکه قبلاً دانستیم یک برآورد غیرسوگیرانه از μ است. وقتی برآورد خاصی از یک پارامتر داریم، آن را یک **برآورد نقطه‌ای**^۱ می‌نامیم. همچنین **برآوردهای فاصله‌ای**^۲ نیز وجود دارند، که طی آن کوشش می‌شود حدودی را که به احتمال زیاد در برگیرنده مقادیر حقیقی (جمعیت) یک میانگین می‌باشند (میانگین، μ ، یک جمعیت کلی از مشاهدات) تعیین گردند. پس چیزی که ما می‌خواهیم **حدود اطمینان**^۳ μ است. این حدود در برگیرنده آن چیزی هستند که **فاصله اطمینان**^۴ خوانده می‌شود. در فصل ۷ مشاهده کردیم که چگونه آنچه که حدود احتمال خوانده شد را تنظیم کنیم. در اینجا هم استدلال مشابهی وجود دارد.

اگر ما بخواهیم برای μ محدودیت‌هایی در نظر بگیریم، با داده‌های موجود، کاری که انجام می‌دهیم این است که بپرسیم μ چقدر می‌تواند بزرگ یا کوچک باشد در حالتی که بدون وادار شدن به رد H_0 ، یک آزمون t را روی میانگین نمونه‌ای اجرا کنیم. به بیان دیگر، اگر μ کاملاً کوچک باشد، این احتمال وجود نخواهد داشت که ما داده‌های نمونه‌ای بگیریم. اگر μ هم خیلی بزرگ باشد همین موضوع صادق است. اما دامنه زیادی از مقادیر μ وجود دارد که ما می‌توانیم انتظار داده‌هایی مشابه آنچه بدست آوردیم در آنها داشته باشیم. ما می‌خواهیم این مقادیر را برای μ حساب کنیم. یک راه آسان برای اینکه ببینیم ما در حال انجام چه کاری هستیم این است که با فرمول t شروع کنیم.

$$t = \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

از آنجا که ما داده‌ها را جمع‌آوری کرده‌ایم، اکنون ما $\bar{X}$ ، s ، و $\sqrt{N}$ را می‌دانیم. ما همچنین می‌دانیم که مقدار بحرانی t در سطح $\alpha = 0.05$ برابر ± 2.062 است. آنچه که باید انجام دهیم جانشین کردن این مقادیر در فرمول t برای حل μ است.

-
1. point estimate
 2. interval estimates
 3. confidence limits
 4. confidence interval

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{N}}}$$

$$\pm 2/262 = \frac{1.463 - \mu}{\frac{0.341}{\sqrt{10}}} = \frac{1.463 - \mu}{0.108}$$

با بازنویسی این فرمول برای حل μ ، خواهیم داشت

$$\mu = \pm 2/262(0/108) + 1/463 = \pm 0/244 + 1/463$$

با حل جداگانه برای $+0/244$ و $-0/244$ حد بالایی و حد پایینی μ را پیدا می‌کنیم.

$$\mu_{\text{بالا}} = +0/244 + 1/463 = 1/707$$

$$\mu_{\text{پایین}} = -0/244 + 1/463 = 1/219$$

و بنابراین می‌توان ۹۵٪ حدود اطمینان را به صورت ۱/۲۱۹ و ۱/۷۰۷، و فاصله اطمینان را به صورت زیر نوشت

$$CI_{0.95} = 1/219 \leq \mu \leq 1/707$$

یا به شکل یک عبارت کلی بنویسیم،

$$CI_{0.95} = \bar{X} \pm t_{\alpha/2} s_{\bar{X}} = \bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{N}}$$

ما یک فاصله اطمینان ۹۵٪ داریم زیرا مقدار بحرانی t دو دامنه را در سطح $\alpha=0/05$ به کار بردیم. برای حدود ۹۹٪ ما $\pm 3/250 = t_{\alpha/2}$ را خواهیم گرفت. آنگاه ۹۹٪ فاصله اطمینان عبارت است از

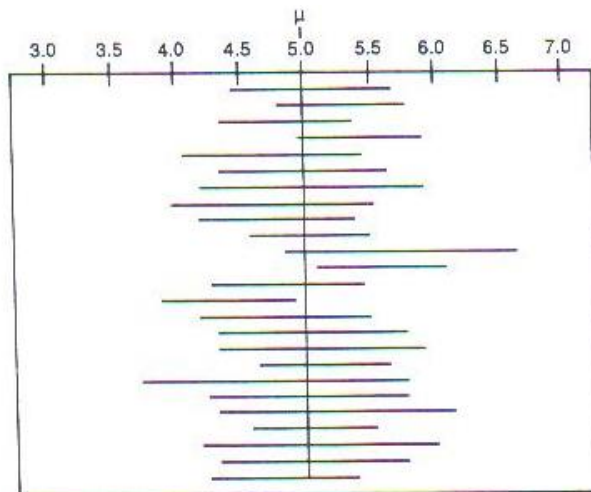
$$CI_{0.99} = \bar{X} \pm t_{\alpha/2}(s_{\bar{X}}) = 1/46 \pm (3/250) = 1/112 \leq \mu \leq 1/814$$

اکنون ما می‌توانیم بگوییم به احتمال ۹۵٪ فاصله $1/707 - 1/219$ در برگیرنده نسبت میانگین حقیقی برای توهم ماه است، در حالی که به احتمال ۹۹٪ فاصله $1/814 - 1/112$ در برگیرنده μ است. به عبارت دیگر به احتمال ۹۵٪ میانگین در فاصله اطمینان $1/707 - 1/219$ قرار می‌گیرد و اگر بخواهیم با اطمینان بیشتر سخن گفته باشیم این فاصله را افزایش داده و می‌گوییم به احتمال ۹۹٪ میانگین در فاصله $1/814 - 1/112$ قرار می‌گیرد. توجه دارید که هیچ یک از این فاصله‌ها در برگیرنده مقدار $1/00$ نیستند، که نشان دهنده نبودن توهم است. ما این موضوع را قبلاً در فاصله اطمینان ۹۵٪ یافتیم زیرا زمانی که آزمون t را اجرا کردیم فرض صفر را رد کردیم.

در شکل ۶-۱۲ یک نمایش گرافیکی از حدود اطمینان نشان داده شده است. برای تهیه این شکل ۲۵ نمونه با $N = 4$ از یک جمعیت دارای میانگین (μ) $5/0$ بیرون کشیده شده است. برای هر نمونه یک حد اطمینان ۹۵٪ روی μ محاسبه و ترسیم شده است. برای مثال، حدود ایجاد شده برای نمونه اول تقریباً $4/46$ و $5/72$ بود، در حالی که برای نمونه دوم حدود $4/83$ و $5/80$ بودند. چون در این نمونه ما می‌دانیم که مقدار μ برابر ۵ است، ما یک خط عمودی در این نقطه رسم کرده‌ایم. توجه داشته باشید که حدود برای نمونه‌های ۱۲ و ۱۴ در برگیرنده $\mu = 5$ نیستند. ما انتظار خواهیم داشت μ در ۹۵ بار از ۱۰۰

بار در این حدود اطمینان واقع شود. بنابراین دو بار از دست رفتن فرصت در ۲۵ بار منطقی به نظر می‌رسد. همچنین توجه داشته باشید که فاصله اطمینان در عرض تغییر می‌کند. این تغییرپذیری را می‌شود با این واقعیت توضیح داد که عرض یک فاصله تابعی از انحراف معیار نمونه است، و بعضی نمونه‌ها دارای انحراف معیار بزرگتری نسبت به نمونه‌های دیگر هستند.

یک نکته را باید در باره تفسیر حدود اطمینان اضافه کنیم. عباراتی به صورت $p(1/219 < \mu$ یا $0/95 = 1/707 < \mu$ به طور عادی تفسیر نمی‌شوند. پارامتر μ یک متغیر نیست. از آزمایشی به آزمایش دیگر تغییر نمی‌کند. بلکه μ یک ثابت است، و چیزی که از یک آزمایش به آزمایش دیگر تغییر می‌کند فاصله است. بنابراین ما می‌توانیم پارامتر را به عنوان یک ستون سنگی و آزمایشگر را در محاسبه حدود اطمینان به عنوان پرتاب کننده حلقه روی آن در نظر بگیریم. در نود و پنج درصد موارد حلقه‌ای با قطر معین دور پارامتر می‌افتد، و در ۵٪ اوقات نمی‌افتد. یک عبارت اطمینان، بیانی است از احتمال اینکه آن حلقه روی هدف قرار گرفته و نه از احتمال اینکه آن هدف در حلقه جای گرفته است.



شکل ۶-۱۲ فاصله محاسبه شده برای ۲۵ نمونه از جمعیتی با میانگین $\mu = 5/00$

۱۲-۷ بررسی معناداری یک ضریب همبستگی

پیش از این دیدیم که تنها به این دلیل که آماره نمونه $\bar{X}$ دقیقاً معادل یک میانگین فرضی جمعیت (μ_0) نیست بیانگر این نخواهد بود که از جمعیتی با میانگین متفاوت (μ_1) بر آمده است. به همین ترتیب، صرفاً به خاطر اینکه یک مجموعه داده‌ها یک ضریب همبستگی را بدست می‌دهند که $\neq 0$ نیست بیانگر این نیست که واقعا بین آن دو متغیر در جمعیت یک رابطه غیر صفر وجود دارد. (در واقع نمونه‌های کوچک

[مثلاً $N = 5$] می‌توانند همبستگی‌های شانس ± 0.50 یا بزرگتر را تاحدی بطور منظم تولید کنند. چیزی که ما نیاز داریم یک آزمون معناداری برای همبستگی است.

اگر ρ (رو) را نشانه همبستگی بین X و Y در نظر بگیریم در حالی که کل جمعیت اشیاء را اندازه‌گیری می‌کنیم، آنگاه می‌توانیم بررسی کنیم که آیا ρ متفاوت از 0 است. اگر $\rho = 0$ باشد، آنگاه X و Y دارای رابطه خطی با هم نیستند، و دانستن X چیزی در باره مقدار نظیر Y به ما نمی‌گوید. اگر ρ برابر 0 نباشد، در این صورت یک رابطه خطی بین X و Y وجود دارد. ضریب همبستگی نمونه‌ای (r) برآوردی از ρ است و می‌تواند برای آزمون فرض صفر که $\rho = 0$ به کار رود. از این گذشته وقتی $\rho = 0$ است توزیع نمونه‌گیری r نرمال است. برآوردی از خطای معیار توزیع نمونه‌گیری r اینگونه بدست می‌آید

$$S_r = \frac{1-r^2}{\sqrt{N-2}}$$

زمانی که سروکار ما با میانگین‌ها بود گفتیم

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}}$$

در این فرمول تفاوت بین یک آماره ($\bar{X}$) و پارامتر نظیر آن (μ) بخش بر خطای معیار آن آماره ($S_{\bar{X}}$) را داریم. در اینجا هم فرایند یکسانی را دنبال می‌کنیم تا آزمونی برای بررسی رابطه همبستگی بسازیم. بنابراین برای آزمون $H_0: \rho = 0$ داریم

$$t = \frac{r - \rho}{S_r}$$

که باز هم تفاوت بین یک آماره (r) و پارامتر نظیر آن (ρ)، بخش بر خطای معیار آن آماره (S_r) را داریم. به زبان محاسباتی این می‌شود

$$t = \frac{r - \rho}{S_r} = \frac{r - 0}{\sqrt{\frac{1-r^2}{N-2}}} = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

که r = ضریب همبستگی نمونه‌ای

N = تعداد جفت‌های مشاهدات

t = استیودنت با $N-2$ df

اگر ما مثال مربوط به سرعت و دقت را در آزمون جور کردن تصاویر در فصل ۹ بگیریم، که $r = 0.722$ و $N = 57$ است، بنابراین

$$t = \frac{-0.722\sqrt{55}}{\sqrt{1-0.521}} = \frac{-0.722\sqrt{55}}{\sqrt{.479}} = -7.737$$

از جدول مقادیر بحرانی t در می یابیم که برای ۵۵ درجه آزادی (df)، مقدار t آزمون دو دامنه در سطح ۰/۵، $t_{0.05}(55)$ برابر ۲/۰۱ است. چون مقدار t بدست آمده $(-۷/۷۳۷)$ بزرگتر از مقدار t بحرانی $(\pm ۲/۰۱)$ است، ما $H_0: \rho = 0$ را رد می کنیم و نتیجه می گیریم که رابطه معناداری بین دو متغیر وجود دارد.

۱۲-۸ مثال کار شده پایانی

در این بخش به آزمون t روی میانگین ها بر می گردیم و مثالی از آن روی فرض صفر مربوط به یک میانگین جمعیت منفرد (μ) را بررسی می کنیم. به مثال زیر توجه نمائید.

یک پژوهشگر فرض می گیرد که با تغییر در رژیم غذایی و الگوی ورزش و سایر تفاوت های زندگی بین دهه ۱۹۸۰ و ۱۹۵۰ بایستی زنان سالم تر و حتی قد بلندتر شده باشند. او می داند که در دهه ۱۹۵۰ میانگین قد زنان آمریکایی ۶۴ اینچ بوده، اما انحراف معیار آن در جداول نیامده است. برای آزمون فرضه خود او ۱۶ زن از پرونده های پزشکی بیرون می کشد و قد آنها را ثبت می کند. آیا این داده ها منجر به این خواهد شد که آزمایشگر این فرض را رد کند که می گوید زنان بطور متوسط بلندتر از ۳۰ سال قبل هستند؟

ابتدا ما نیازمند طرح نمودن فرض صفر هستیم. پژوهشگر می خواهد این فرضیه را آزمایش کند که متوسط قد زنان با آنچه در دهه ۱۹۸۰ بوده فرق دارد، بنابراین او فرض صفری را طرح ریزی می کند که مدعی است فرقی ندارد و یکسان است. پس ما داریم $\mu = 64$ ، H_0 . بنابراین ما باید در باره یک دامنه یا دو دامنه بودن آزمون و سطح معناداری تصمیم بگیریم. بدلالی که پیش تر گفته شد، ما از آزمون دو دامنه استفاده می کنیم و $\alpha = 0.05$ در نظر می گیریم.

در این داده ها ابتدا میانگین و انحراف معیار را بدست می آوریم و بعد مقدار t را حساب می کنیم.

داده ها: ۶۶ ۶۲ ۶۳ ۶۳ ۶۸ ۶۷ ۶۵ ۶۳ ۶۶ ۶۰ ۶۱ ۶۷ ۶۴ ۶۶ ۶۵ ۶۳

$$\begin{aligned}\bar{X} &= \frac{\sum X}{N} = \frac{1034}{16} = 64.625 \\ S^2 &= \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1} = \frac{66912 - \frac{(1034)^2}{16}}{15} = \frac{66912 - 66822.25}{15} = 5.9833 \\ S &= \sqrt{S^2} = \sqrt{5.9833} = 2.446\end{aligned}$$

اکنون ما می توانیم t را بدست آوریم.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{N}} = \frac{64.625 - 64}{2.446/\sqrt{16}} = \frac{0.625}{0.612} = 1.02$$

در جدول مقادیر بحرانی t در بخش پیوست مشاهده می کنیم که با $15 = 16 - 1$ N درجه آزادی در سطح $\alpha = 0.05$ ، مقدار t آزمون دو دامنه برابر $2/131$ است. بنابراین در صورتی فرض صفر رد می شود که t بدست آمده از $\pm 2/131$ بیشتر باشد. چون t_{obs} کمتر از مقدار بحرانی بود ما نمی توانیم فرض صفر

را رد کنیم. بنابراین ما دلیل کافی برای نتیجه‌گیری در باره اینکه زنان بطور متوسط قدبلندتر از ۳۰ سال قبل شده‌اند نداریم.

۹-۱۲ خلاصه

در این فصل ما به توزیع نمونه‌گیری میانگین و سودمندی اطلاعات حاصل از آن برای آزمون فرضیه پرداختیم. سپس دیدیم که چگونه می‌توانیم اطلاعات خود در باره z و توزیع نمونه‌گیری ترکیب کنیم تا یک آزمون فرض صفر مربوط به μ تدوین کنیم. بعد دیدیم چگونه می‌توانیم از z فراتر رفته، آزمون t را برای زمانی که انحراف معیار جمعیت نامعلوم است و بایستی از انحراف معیار نمونه برآورد گردد به کار گیریم. به عواملی که روی اندازه t در یک آزمایش خاص اثر می‌گذراند توجه کردیم. آنگاه به ساختار حدود اطمینان روی یک میانگین جمعیت پرداختیم. در آخر هم یک آزمون t برای معناداری ضریب همبستگی بررسی کردیم. برخی از مطالب مهم فصل عبارتند از:

- فرضیه آزمایشی
- توزیع نمونه‌گیری میانگین
- قضیه حد مرکزی
- توزیع مستطیلی
- خطای معیار
- توزیع t استیودنت
- درجات آزادی (df)
- گروه کنترل (گواه)
- برآورد نقطه‌ای
- برآوردهای فاصله‌ای
- فاصله اطمینان (حدود اطمینان)
- سطح p (احتمال)

۱۰-۱۲ تمرین

۱۲-۱ با استفاده از جدول اعداد تصادفی بخش پیوست، توزیع ۱۰۰ عدد نخست کم اهمیت‌تر (رقم‌های سمت راست) را رسم کنید.

۱۲-۲ با تکرار تمرین ۱-۱۲ میانگین‌های ۲۵ نمونه از پنج رقم کوچک‌تر را حساب و رسم کنید. (میانگین‌ها را در فواصل مناسب گروه‌بندی کنید.)

۱۲-۳ میانگین و انحراف معیار توزیع اعداد تمرین ۱-۱۲ و توزیع نمونه‌گیری میانگین در تمرین ۲-۱۲ را با هم مقایسه کنید. آیا این جواب‌ها در راستای آن چیزی هست که از قضیه حد مرکزی بر می‌آید.

۱۲-۴ اگر اندازه نمونه‌ها از ۵ بیشتر باشد از چه نظر نتیجه جواب ۲-۱۲ فرق خواهد کرد.

۱۲-۵ در سال ۱۹۷۹ تعداد ۲۳۸ دانش‌آموز یکی از ایالت‌های آمریکا در بخش شفاهی امتحان پیشرفت تحصیلی (SAT) میانگین ۵۲۵ گرفتند. انحراف معیار گزارش نشده است.

الف) آیا این نتیجه با وجود میانگین ۵۰۰ و انحراف معیار ۱۰۰ برای (SAT) همخوانی دارد؟

ب) آیا شما این H_0 را که بدنبال شواهدی برای کاهش نمرات (SAT) از میانگین ۵۰۰ در طی زمان است رد خواهید کرد؟

۱۲-۶ چرا داده‌های تمرین ۵-۱۲ در کل واقعا گویای مشکل حاد آموزشی در آمریکا نیست؟
 ۱۲-۷ در ۱۹۷۹ تعداد ۲۳۴۵ دانش‌آموز از ایالتی دیگر که بخش ریاضی SAT را امتحان دادند دارای میانگین نمره ۵۲۴ بودند. آیا این نمره با توجه به یک میانگین جمعیت ۵۰۰ اگر $\sigma = 100$ در نظر گرفته شود هم‌سو است؟

۱۲-۸ چرا جواب تمرین ۷-۱۲ اساسا با جواب ۵-۱۲ فرق دارد با وجود اینکه میانگین‌ها تقریبا یکسان هستند؟

۱۲-۹ معمولا فرض می‌شود که امتحان کارشناسی ارشد آمریکا (GRE) دارای نمراتی با میانگین ۵۰۰ است. در ۱۹۸۱ میانگین نمره کلامی ۵۷۰۱ دانشجوی و افراد پذیرفته نشده در رشته زیست‌شناسی ۵۰۳ با انحراف معیار ۱۰۴ بود. آیا باور عمومی در باره میانگین GRE برای این گروه درست است؟

۱۱-۱۰ فرض صفر (H_0) و فرض مقابل (H_1) در تمرین ۹-۱۲ چیست؟

۱۲-۱۱ آیا برای تمرین ۹-۱۲ یک آزمون یک دامنه باید اجرا کرد؟

۱۲-۱۲ اگر در تمرین ۹-۱۲ شما فرض صفر را رد کنید یک نتیجه مهم حاصل می‌شود؟ اگر بله چرا و اگر نه چرا؟

۱۲-۱۳ در یک تعاونی بزرگ میانگین حقوق مردان دارای ۳ تا ۵ سال سابقه کار ۲۷۰۰۰ دلار است. برای یک نمونه ۱۰ نفری زنان با سابقه ۳ تا ۵ سال هم به این شرح است (به هزار دلار)

۲۳ ۲۶ ۳۰ ۲۰ ۱۸ ۲۵ ۲۹ ۲۱ ۱۴ ۳۵

آیا شواهدی برای اختلاف حقوق زنان و مردان وجود دارد؟

۱۲-۱۴ هر چند بیشتر حقوق‌ها در تمرین ۱۳-۱۲ زیر ۲۷۰۰۰ دلار بود، شما نتوانستید فرض صفر را رد کنید. علت آن چیست؟

۱۲-۱۵ برای داده‌های تمرین ۵-۱۲ حدود اطمینان ۹۵٪ در اطراف μ را حساب کنید.

۱۲-۱۶ برای داده‌های تمرین ۹-۱۲ حدود اطمینان ۹۵٪ در اطراف μ را حساب کنید.

۱۲-۱۷ نظر شما در باره تفاوت تمرین ۱۵-۱۲ و ۱۶-۱۲ چیست؟

۱۲-۱۸ شیوه‌هایی که شما برای رسیدن به جواب در شکل ۴-۱۲ طی می‌کنید را توضیح دهید.

۱۲-۱۹ با یک آزمون t پاسخ به تمرین ۲-۹ و ۳-۹ را بررسی کنید.

فصل ۱۳

آزمون‌های فرضیه برای میانگین‌ها – دو میانگین وابسته

در فصل ۱۲ ما موقعیتی را در نظر گرفتیم که در آن یک میانگین نمونه‌ای ($\bar{X}$) داشتیم و می‌خواستیم بررسی کنیم که آیا می‌توانیم بطور منطقی باورر داشته باشیم که اگر از یک جمعیت با میانگین معین (بیشتر با μ_0 نشان داده می‌شود) نمونه‌گیری کرده باشیم آیا چنین میانگین نمونه‌ای رخ می‌دهد. راه دیگر برای بیان این مطلب این است که ما می‌خواهیم مشخص کنیم آیا میانگین جمعیتی که از آن نمونه‌گیری کردیم (آن را μ_1 گویند) برابر با مقدار معین است که در فرض صفر (μ_0) تعیین شده است. در این فصل ما از حالت قبلی فاصله می‌گیریم که در آن یک نمونه از داده‌ها داشتیم و می‌خواستیم روی میانگین آن آزمونی را اجرا کنیم. در اینجا به حالتی می‌پردازیم که ما دو نمونه وابسته^۱ داریم (همین تحلیل در مورد شیوه‌های دیگر هم به کار می‌رود که بنام اندازه‌های مکرر^۲، نمونه‌های جور شده^۳، نمونه‌های همبسته، نمونه‌های جفتی، یا نمونه‌های وابسته معروف هستند) و مایلیم یک آزمون روی تفاوت میانگین‌های آن دو اجرا کنیم. چنانکه خواهید دید این آزمون خیلی شبیه به آزمونی است که در فصل قبلی مورد بحث قرار گرفت.

۱۳-۱ نمونه‌های وابسته

در موقعیت‌های بسیاری که ما از آزمون t مورد بحث در این فصل استفاده می‌کنیم، اما مسلماً نه همه موقعیت‌ها، از یک عده آزمودنی‌ها ما دو مجموعه داده خواهیم داشت. برای مثال، ما ممکن است از ۲۰ نفر تقاضا کنیم تا سطح اضطراب خود را قبل و بعد از اهدای خون رتبه‌بندی کنند. یا ممکن است با

-
1. related samples
 2. repeated measures
 3. matched samples

استفاده از دو شیوه مختلف سطح ناتوانی ۲۰ فرد معلول را رتبه بندی کنیم تا ببینیم آیا اندازه هایی که از یک شیوه بدست می آید کمتر از شیوه دیگر است. در هر دو مثال ما ۲۰ جفت از نمرات خواهیم داشت، برای هر فرد دو نمره، و انتظار داریم این دو مجموعه نمره (متغیرها) همبسته باشند. در طرح آزمون t لازم است ما این همبستگی را به حساب بیاوریم. به مثال اضطراب دقت کنید، افراد از لحاظ سطح اضطراب خیلی باهم فرق دارند. بعضی به نظر می رسد در هر زمانی و هر اتفاقی که بیافتد مضطرب می شوند، و برخی دیگر هرچه را پیش آید می پذیرند و در باره هیچ چیز نگران نمی شوند. بنابراین باید بین سطح اضطراب یک فرد قبل از اهداء خون و بعد از آن رابطه ای وجود داشته باشد. به عبارت دیگر اگر بدانیم سطح اضطراب یک فرد قبل از اهداء خون چقدر است، می توانیم در باره بعد از آن حدس منطقی بزنیم. به همین ترتیب برخی افراد دچار معلولیت شدید هستند، در حالی که بعضی دیگر فقط در حد خفیف مبتلا هستند. در صورتی که ما می دانیم که یک شخص خاص با استفاده از یک شیوه اندازه ی بالایی بدست می آورد، احتمال آن می رود که این شخص با شیوه دیگر هم اندازه ی نسبتا بالایی داشته باشد. رابطه بین مجموعه های داده ها حتما لازم نیست کامل باشد- در واقع احتمالا هیچگاه چنین نباشد. این حقیقت که ما می توانیم دست به پیش بینی هایی بهتر از حالت شانسی بزنیم کافی است تا دو مجموعه از داده ها را به صورت همبسته یا جور شده طبقه بندی کنیم. (به بیان دیگر ما وقتی دو متغیر [مثلا دو مجموعه نمرات اضطراب] داریم که بطور معناداری همبسته هستند، سر و کار ما با نمونه های همبسته یا جور شده است.

در دو مثال قبل ما موقعیت های را در نظر گرفتیم که در آن هر شخص مورد مطالعه در دو نمره سهیم است. هر چند این معمول ترین شیوه بدست آوردن نمونه های همبسته است، اما تنها شیوه نیست. برای مثال، مطالعه ای روی روابط زناشوئی ممکن است شامل این باشد که از زنان و شوهران تقاضا کنید میزان رضایت از زناشوئی خود را رتبه بندی کنند، با این هدف که بخواهیم بدانیم آیا در مجموع، زنان دارای رضایت بیشتر یا کمتری نسبت به شوهران هستند. در اینجا هر فرد تنها در یک نمره سهیم است اما یک زوج به عنوان یک واحد در یک جفت نمره سهیم خواهند بود. و به احتمال زیاد اگر شوهر خیلی از زندگی زناشوئی ناراضی باشد، ممکن نیست که زن او از زناشوئی احساس خوشبختی کند - و برعکس. مثال های زیادی از طرح های آزمایشی وجود دارد که شامل نمونه های همبسته هستند. اما همه آنها در یک چیز مشترک هستند، و آن این است که اطلاع از یک عدد در یک جفت نمره، چیزی در باره نمره دیگر به ما می گوید. هرگاه چنین شرایطی باشد می گوییم نمونه ها همبسته هستند. این فصل به آزمون های t در باره تفاوت دو نمونه وابسته می پردازد.

۲-۱۳ استیودنت برای نمره‌های اختلاف

به عنوان مثالی از تحلیل نمونه‌های وابسته به سراغ یک تحقیق مداخله‌گر می‌رویم که برای توسعه مهارت‌های اجتماعی خاصی در دانش‌آموزان نوپای دبیرستانی طرح شده است (گروهی که فضای کافی برای پیشبرد هدف ما را فراهم می‌کند). برنامه با ۱۵ آزمودنی به صورت حضور در یک مرکز خرید، مهمانی در رستوران و رفتن به سینما آغاز شد. همچنین از چند ناظر خواسته شد که رفتارهای اجتماعی نامناسب سرزده از هر آزمودنی را شمارش کنند (مثل تنه زدن در مرکز خرید، داد زدن به سمت ماشین‌های عبوری، نمکدان را سس مالی کردن، انداختن ته سیگار، متلک انداختن در سینما به زوج‌ها، و سایر رفتارهای شیطنت‌آمیزی که ممکن است هر یک از ما در این سنین انجام داده باشیم). پس از یک برنامه مداخله دو هفته‌ای، که طی آن سعی شد مهارت‌های اجتماعی مطلوب به آزمودنی‌ها فهمانده شود، بار دیگر آزمودنی‌ها را در موقعیت‌های قبلی مرکز خرید و سینما قرار داده و رفتارهای آنان توسط ناظران مورد ارزیابی قرار گرفت.

داده‌های حاصل از این مطالعه در جدول ۱-۱۳ آمده است که در آن مجموعه نمرات اول با X_1 و مجموعه دوم با X_2 نشان داده شده است. فرض صفری که ما می‌خواهیم آزمایش کنیم این است که میانگین (μ_1) نمرات جمعیتی که داده‌های اول از آن بدست آمده با میانگین (μ_2) جمعیتی که از داده‌های دوم بدست آمده برابر است. به عبارت دیگر ما می‌خواهیم بیازماییم $H_0: \mu_1 = \mu_2$ (یا معادل آن، $= 0$ $\mu_1 - \mu_2$). توجه داشته باشید که مقدار μ_1 و μ_2 مهم نیست. همچنین ما این فرض صفر را با استفاده از داده‌های نمونه‌های وابسته می‌آزماییم.

نمرات اختلاف

هر چند بسیار بدیهی به نظر می‌رسد که این داده‌ها را به عنوان نمرات دو نمونه ببینیم، که یک مجموعه قبل از برنامه آموزش و دیگری بعد از آن بدست آمده باشد، همچنین می‌توان، و خیلی سودمند است که داده‌ها را به یک مجموعه نمرات تبدیل کنیم - مجموعه اختلاف یا تفاوت‌های بین X_1 و X_2 برای هر آزمودنی. این تفاوت‌ها **نمرات اختلاف** نامیده می‌شوند و در ستون سوم جدول ۱-۱۳ نشان داده شده‌اند. می‌توان آنها را به عنوان میزان پیشرفت بین دو جلسه سنجش - به عنوان نتیجه مداخله ما در نظر گرفت. اگر واقعا برنامه مداخله اثری نداشته باشد (یعنی H_0 درست باشد)، متوسط نمره از یک جلسه به جلسه دیگر تغییری نخواهد داشت. بر اساس شانس ممکن است برخی آزمودنی‌ها در X_2 نسب به X_1 نمره بالاتر و برخی نمره پایین‌تری داشته باشند، اما در مجموع تفاوتی وجود نداشته باشد.

اکنون اگر داده‌ها را به عنوان ستون نمرات اختلاف در نظر بگیریم، فرض صفر این می‌شود که میانگین یک جمعیت از نمرات اختلاف (μ_D) برابر ۰ است. از آنجا که می‌توان نشان داد که $\mu_1 - \mu_2 = \mu_D$ ، پس می‌توانیم بنویسیم $H_0: \mu_D = \mu_1 - \mu_2 = 0$. حالا مشاهده می‌شود که ما یک فرضیه صفر را با

استفاده از یک نمونه از داده ها (نمونه نمرات اختلاف) می آزماییم، و از پیش در فصل ۱۲ دانستیم چگونه آن را انجام دهیم.

جدول ۱-۱۳ داده ها و نمرات اختلاف تعداد رفتارهای نامناسب اجتماعی

قبل (X_1)	بعد (X_2)	اختلاف (D)
۱۸	۱۲	۶
۵	۴	۱
۱۹	۱۷	۲
۱۳	۱۱	۲
۴	۸	۱۲
۵	۱۲	۱۷
۲۶	۲۷	-۱
۳	۳	۰
۱	۳	-۲
۲۰	۱۴	۶
۱۵	۱۲	۳
۱۸	۱۴	۴
۱۰	۱۱	-۱
۸	۱۰	-۲
۱۵	۹	۶
میانگین	۱۳/۳۳۳	۱۱/۱۳۳
s	۶/۹۱۴	۵/۹۹۸
	۲/۹۳۳	۲/۲۰۰

آماره t

ما حالا دقیقاً در همان جایی از فصل گذشته قرار داریم که موقع داشتن یک نمونه از داده ها و یک فرض صفر ($\mu = 0$) بودیم. تنها تفاوت این است که در اینجا داده ها نمرات اختلاف هستند، و میانگین و انحراف معیار بر اساس تفاوت ها بنا شده اند. به خاطر داشته باشید که t به عنوان تفاوت بین یک میانگین نمونه و میانگین جمعیت، تقسیم بر خطای معیار میانگین تعریف شده است. آنگاه داریم

$$t = \frac{\bar{D} - 0}{s_{\bar{D}}} = \frac{\bar{D} - 0}{\frac{s_D}{\sqrt{N}}}$$

که $\bar{D}$ میانگین نمرات اختلاف و $S_{\bar{D}}$ انحراف معیار آنها است و N تعداد نمرات اختلاف است (یعنی تعداد زوجها، نه تعداد نمرات خام). برای داده‌های ما

$$t = \frac{\bar{D}-0}{\frac{S_{\bar{D}}}{\sqrt{N}}} = \frac{2.20-0}{\frac{2.933}{\sqrt{15}}} = \frac{2.20-0}{0.757} = 2.91$$

درجات آزادی

درجات آزادی برای حالت نمونه جور شده دقیقاً مانند حالت تک نمونه‌ای است. چون ما با نمرات اختلاف کار داریم، N برابر تعداد نمرات اختلاف است (یا تعداد جفت‌های مشاهدات، یا تعداد مشاهدات مستقل - همه یک چیز است). از آنجا که واریانس این نمرات اختلاف ($S_{\bar{D}}^2$) به عنوان برآوردی از واریانس یک جمعیت نمرات اختلاف (σ_D^2) به کار می‌رود و چون این واریانس نمونه با استفاده از میانگین نمونه ($\bar{D}$) بدست آمده است، ما یک df از دست داده و $N-1$ درجه آزادی داریم. به عبارت دیگر، $df =$ تعداد زوجها منهای ۱.

در این مثال ما ۱۵ نمره اختلاف داریم، پس ۱۴ درجه آزادی خواهیم داشت. در جدول t بخش پیوست برای آزمون دو دامنه در سطح ۰/۰۵، به $\pm 2.145 = t_{.05}(14)$ بر می‌خوریم. مقدار t بدست آمده ما (۲/۹۱) بزرگتر از ۲/۱۴۵ است، و بنابراین ما H_0 را رد می‌کنیم و نتیجه می‌گیریم که این نمرات اختلاف از جمعیتی از نمرات اختلاف که در آن $\mu_D = 0$ است نمونه‌گیری نشده‌اند. به شکل عملیاتی این مطلب می‌گوید که آزمودنی‌ها بعد از مداخله آزمایشی رفتارهای غیراجتماعی کمتری نسبت به قبل از آن نشان داده‌اند. هرچند ما مایلیم که این مطلب را به عنوان موفقیت برنامه خود ارزیابی کنیم، اما ممکن است فقط به این دلیل باشد که دانش‌آموزان در مرحله دوم متوجه شده باشند که رفتارهای آنها ثبت می‌شود، یا شاید به این خاطر باشد که مرحله دوم مشاهدات ما بعدازظهرها بوده و خستگی غالب شده است، یا حتی چیز دیگری که ما فکرش را نکرده‌ایم. در هر حال واقعیت فرق نمی‌کند، نمرات در مرحله دوم به هر دلیل به اندازه کافی پایین‌تر هستند تا امکان رد $\mu_D = \mu_1 - \mu_2 = 0$ را بدهند.

۳-۱۳ یک مثال دیگر - توهم ماه

به عنوان مثالی دیگر به کار کافمن و راک (۱۹۶۲) در باره توهم ماه بر می‌گردیم. یک فرض مهم در باره منشاء توهم ماه قبلاً بوسیله هالوی و بورینگ (۱۹۴۰) طرح شده بود که معتقد بودند این توهم (خطای ادراکی) ناشی از این حقیقت است که وقتی ماه در افق قرار دارد، مشاهده کننده بطور مستقیم در سطح چشم‌هایش به آن نگاه می‌کند، در حالی که وقتی در وسط آسمان (نقطه اوج) است مشاهده کننده مجبور است چشم‌ها و نیز سر خود را بالا ببرد. هالوی و بورینگ معتقد بودند این تفاوت در بالا بردن چشم‌ها علت توهم است. برای آزمودن این فرضیه کافمن و راک (۱۹۶۲) آزمایشی را طراحی کردند که به آنها

اجازه می‌داد دو حالت مصنوعی ماه (یکی در افق و یکی در وسط آسمان) و کنترل اینکه آزمودنی‌ها برای حالت ماه در وسط، چشم‌ها را بالا می‌برند یا نه را نشان دهند. در یک حالت آزمودنی وادار می‌شد تا سر خود را در وضعیتی قرار دهد که بتواند ماه را در وسط آسمان در راستای چشم‌هایش ببیند. در حالت دیگر آزمودنی مجبور بود ماه را در وسط با بالا بردن چشم‌ها ببیند (مشاهده حالت افق همیشه هم سطح با چشم‌ها بود). در هر دو حالت متغیر وابسته نسبت اندازه ادراک شده ماه در افق به اندازه ادراک شده ماه در وسط آسمان بود (یک نسبت ۱/۰۰ بیان‌گر عدم خطای ادراکی یا توهم بود). چنانچه نظر هالوی و بورینگ درست می‌بود، بایستی در حالت بالا بردن چشم نسبت به حالت هم سطحی خطای ادراکی بزرگتری (نسبت بزرگتر) بروز می‌کرد، هر چند همواره تصور می‌شود "ماه" در یک جا (وسط آسمان) باشد. داده‌های واقعی این آزمایش در جدول ۲-۱۳ آمده است.

در این مثال می‌خواهیم این فرض صفر را که میانگین‌ها برابر هستند تحت دوشرایط مشاهده بیازمائیم. چون ما با مشاهدات همبسته سروکار داریم (هر آزمودنی در هر دو شرایط حضور دارد)، ما با اختلاف نمرات کار می‌کنیم و $H_0: \mu_D = 0$ را آزمایش می‌کنیم. با به کارگیری یک آزمون دو دامنه در $\alpha = 0.05$ ، فرض مقابل $H_1: \mu_D \neq 0$ است.

از فرمول t برای نمونه‌های وابسته داریم

$$t = \frac{\bar{D} - 0}{s_{\bar{D}}} = \frac{\bar{D} - 0}{\frac{s_D}{\sqrt{N}}} = \frac{0.019 - 0}{\frac{0.137}{\sqrt{10}}} = \frac{0.019}{0.043} = 0.44$$

از جدول t بخش پیوست در می‌یابیم $t_{0.05}(9) = \pm 2.262$ چون $t_{obt} = 0.44$ کمتر از 2.262 است، ما نمی‌توانیم H_0 را رد کنیم و به این تصمیم می‌رسیم که شواهدی برای اینکه خطای ادراکی ناشی از بالا بردن چشم‌ها باشد نداریم. (در حقیقت این داده‌ها شامل آزمون فرض دوم هالوی و بورینگ نیز می‌باشد، زیرا آنها پیش‌بینی کرده‌اند که اگر آزمودنی‌ها ماه را در وسط آسمان در راستای چشم‌ها می‌بینند خطای ادراکی وجود ندارد. در واقع داده‌ها خطای ادراکی قابل توجهی را در این شرایط نشان می‌دهند. یک آزمون معناداری از خطای ادراک را می‌توان به روشی که در فصل ۱۲ بحث شد اجرا نمود، و در واقع خطای ادراکی معنادار است.)

جدول ۲-۱۳ مقدار خطای ادراک ماه هنگامی که نگرستن به ماه در وسط آسمان در راستای چشم‌ها و هنگامی که با بالا

بردن چشم‌ها صورت می‌گیرد

آزمودنی	چشم‌های بالا برده	در راستای چشم	اختلاف (D)
۱	۱/۶۵	۱/۷۳	-۰/۰۸
۲	۱/۰۰	۱/۰۶	-۰/۰۶
۳	۲/۰۳	۲/۰۳	۰/۰۰
۴	۱/۲۵	۱/۴۰	-۰/۱۵
۵	۱/۰۵	۰/۹۵	۰/۱۰
۶	۱/۰۲	۱/۱۳	-۰/۱۱
۷	۱/۶۷	۱/۴۱	۰/۲۶
۸	۱/۸۶	۱/۷۳	۰/۱۳
۹	۱/۵۶	۱/۶۳	-۰/۰۷
۱۰	۱/۷۳	۱/۵۶	۰/۱۷
$\bar{D} = ۰/۰۱۹$			
$s_D = ۰/۱۳۷$			
$s_{\bar{D}} = ۰/۰۴۳$			

۴-۱۳ برتری‌ها و کاستی‌های استفاده از نمونه‌های وابسته

در فصل بعد ما توجه خود را به طرح‌های آزمایشی معطوف می‌کنیم که در آنها به جای دو بار استفاده از آزمودنی‌های واحد (یا شیوه‌های دیگر مربوط به داده‌های نمونه‌های وابسته)، از دو گروه مستقل از آزمودنی‌ها استفاده می‌شود. نمونه‌های مستقل در موارد بسیاری سودمند هستند، اما قبل از پرداختن به آن مبحث، لازم است نقاط قوت و ضعف نمونه‌های وابسته را بفهمیم.

احتمالا مهمترین برتری یک آزمایش با نمونه‌های وابسته این است که این شیوه به ما امکان می‌دهد از مشکلاتی که از تغییرپذیری از یک آزمودنی به آزمودنی دیگر ناشی می‌شود پرهیز کنیم. مثلا در جدول ۱-۱۳ توجه کنید که بعضی آزمودنی‌ها (مثل #۸ و #۹ تقریبا رفتار نامطلوب ندارند. از سوی دیگر آزمودنی #۷ و #۱۰ نمونه بعضی کارهای ناخوشایند روی مردم دور و بر هستند. برتری طرح‌های مربوط به نمونه‌های همبسته این است که این تفاوت بین آزمودنی‌ها وارد داده‌های مورد تحلیل -نمرات اختلاف- نمی‌شوند. یک تغییر از ۲۶ به ۲۴ دقیقا مشابه تغییر از ۶ به ۴ اعمال می‌شود. با اجازه ندادن به تغییرپذیری از یک آزمودنی به آزمودنی دیگر در سطح کلی رفتار که با ایجاد واریانس بزرگ بر داده‌ها تأثیر بگذارد، طرح‌های نمونه‌های وابسته برتری قابل توجهی از جنبه توان- توانایی رد فرض صفر نادرست- نسبت به نمونه‌های مستقل دارند.

برتری دوم نمونه های وابسته بر دو نمونه مستقل این واقعیت است که نمونه های وابسته امکان کنترل متغیرهای مداخله گر را فراهم می کنند. اگر ما یک گروه آزمودنی ها را قبل از مداخله آزمایشی مورد سنجش قرار بدهیم و گروه متفاوتی را بعد از مداخله آزمایشی، ممکن است هر میزان اختلاف بین گروه ها مشاهده شود بدون اینکه ربطی به مداخله آزمایشی ما داشته باشد، ولی به هر حال بر نتایج تأثیر می گذارد. این در مطالعه ما مشکلی ایجاد نمی کند چون ما از آزمودنی های یکسانی در هر دو جلسه سنجش استفاده می کنیم.

یک برتری دیگر طرح های سنجش وابسته این است که به آزمودنی های کمتری نسبت به طرح های با نمونه مستقل برای همان اندازه توان، نیاز دارند. این یک مزیت اساسی است، که هر کس که با به کارگیری آزمودنی ها سروکار داشته است می تواند این را به شما گوشزد کند. معمولاً خیلی آسان تر است که ۲۰ نفر را برای کاری دو بار به کار بگیرید تا ۴۰ آزمودنی را یک بار.

نخستین کاستی طرح های سنجش وابسته متوجه این واقعیت است که ممکن است یک اثر ترتیبی یا اثر انتقالی از یک بخش به بخش دیگر وجود داشته باشد، یا اندازه گیری بار اول ممکن است خود از طریق فرایندهایی مانند برانگیختن حساسیت بر آزمایش اثر بگذارد. بطور مثال، اگر ما طی طرحی از دانش افراد پیرامون پدیده های جاری امتحان بگیریم و پس از آن یک دوره آموزشی در این باره ترتیب داده و سپس امتحان مجدد بگیریم، منطقی است که نتیجه گیری کنیم در بار دوم افراد با سؤال ها بیشتر آشنا هستند و حتی در فاصله بین اجرای دو آزمون ممکن است پاسخ ها را مرور کرده باشند. به همین ترتیب در مطالعات مربوط به داروها اثرات جلسه نخست ممکن است تا جلسه بعد از بین نرفته باشد. یک مشکل عمومی طرح های سنجش های وابسته زمانی برور می کند که یک آزمون اولیه ممکن است آزمودنی ها را نسبت به هدف مداخله آگاه سازد. به عنوان مثال یک پیش آزمون در باره نگرش های مربوط به تغذیه با شیر مادر ممکن است در مرحله بعدی که یک غریبه کنار شما می نشیند و شروع به صحبت درباره تغذیه با شیر مادر می کند یک ذره احساس عدم اطمینان ایجاد کند. وقتی شما آگاه باشید که تحقیق شما ممکن است تحت تأثیر اثرات مخدوش کننده انتقال قرار گیرد یا اثرات آزمایش می تواند تحت تأثیر اندازه گیری های اولیه قرار گیرد، طرح اندازه های وابسته مناسب نیست. برای کنترل کردن فنونی وجود دارد، هر چند اثرات ترتیب و انتقال را از بین نمی برد. در اینجا اینها مورد بحث قرار نمی گیرند.

۵-۱۳ خلاصه

در این فصل ما به تحلیل داده های مربوط به دو نمونه وابسته پرداختیم. ما دیدیم که داده ها را به آسانی می توان به مجموعه نمرات اختلاف کاهش داد و اینکه آزمون یک نمونه ای t را روی این نمرات اختلاف اجرا کرد تا فرض صفری که مدعی است میانگین تفاوت ها چیزی بیش از آنچه بوسیله شانس پیش بینی

می‌شود از ۵ فاصله ندارد را بررسی نمود. سپس ما برتری‌ها و کاستی‌های طرح‌های آزمایشی با سنجش‌های وابسته را بررسی کردیم. برخی از مهم‌ترین موضوع‌های این فصل عبارتند از:

- نمونه‌های وابسته (اندازه‌های وابسته)
- اثر ترتیب
- (نمونه‌های جور شده)
- نمرات اختلاف
- اثر انتقال

۱۳-۶ تمرین‌ها

۱۳-۱ در تحقیقی به ۱۵ نفر عقب مانده ذهنی برای مدت شش ماه آموزش مهارت‌های خود مراقبتی به شیوه تقلید داده شد. برای شش ماهه دوم تمرین، به این افراد راهنمایی فیزیکی داده شد. رتبه‌بندی هر شخص در هر دوره بر اساس سطح یاری مورد نیاز (بالا = بد) را در اختیار داریم. داده‌ها برای هر فرد به شرح زیر است:

آزمودنی‌ها	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
تقلید	۱۴	۱۱	۱۹	۸	۴	۹	۱۲	۵	۱۴	۱۷	۱۸	۰	۲	۸	۶
راهنمایی فیزیکی	۱۰	۱۳	۱۵	۵	۳	۶	۷	۹	۱۶	۱۰	۱۳	۱	۲	۳	۶

چگونه می‌توان فهمید آیا سطح یاری مورد نیاز آزمودنی‌ها در شش ماهه دوم کاهش یافته است؟
 ۱۳-۲ آیا مطالعه توصیف شده در تمرین ۱-۱۳ به ما پاسخ روشنی در باره کیفیت نسبی دو رویکرد می‌دهد؟ اگر نه چرا؟

۱۳-۳ چگونه می‌توان مطالعه توصیف شده در تمرین ۱-۱۳ را بهبود بخشید؟
 ۱۲-۴ سطح اطمینان ۹۵٪ برای اختلاف میانگین واقعی بین اعمال آزمایشی بیان شده در تمرین ۱-۱۳ را با کمک فنون گفته شده در فصل ۱۲ معین کنید.

۱۳-۵ به عنوان بخشی از یک تحقیق کاهش سیگار کشیدن، یک سازمان ملی یک کمپین تبلیغاتی را برای ترک مصرف سیگار اجرا نمود. در این تحقیق برای ارزیابی اثربخشی کمپین از ۱۵ آزمودنی خواسته شد در هفته قبل و بعد از تبلیغات تعداد نخ سیگار مصرفی خود در روز را ثبت نمایند. داده‌ها به این شرح است:

آزمودنی‌ها	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
قبل	۴	۱	۲	۳	۱	۳	۲	۲	۲	۴	۲	۱	۳	۲	۳
	۵	۶	۰	۳	۹	۳	۵	۶	۰	۸	۶	۵	۶	۲	۲

بعد	۴	۲	۱	۳	۲	۳	۱	۲	۲	۲	۴	۲	۴	۱	۲	۳
	۳	۰	۷	۰	۵	۹	۴	۸	۳	۱	۶	۰	۶	۳	۴	۴

آزمون t مناسب را اجرا کنید.

۱۳-۶ فرض کنید داده های تمرین ۵-۱۳ به صورت دیگری بدست آمده باشد. داده های جدید چنین است.

آزمودنی	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱	۱	۱	۱	۱	۱	۱
ا																
قبل	۴	۱	۲	۳	۳	۱	۳	۲	۲	۲	۴	۲	۳	۱	۲	۳
	۵	۶	۰	۳	۰	۹	۳	۵	۶	۰	۸	۶	۵	۶	۲	۲
بعد	۵	۳	۳	۴	۲	۱	۲	۲	۲	۳	۴	۱	۲	۱	۲	۴
	۹	۵	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰	۰

آزمون t مناسب را اجرا کنید.

۱۳-۷ نتایجی که یک آزمایشگر دقیق می تواند از تمرین های ۵-۱۳ و ۶-۱۳ بدست آورد را مقایسه کنید. ۱۳-۸ مثالی از یک آزمایش ارائه کنید که در آن استفاده از نمونه های وابسته بدلیل اثرات انتقال توصیه نادرستی است.

۱۳-۹ فرض کنید میانگین و انحراف معیار نمرات اختلاف در تمرین ۵-۱۳ در صورتی که آزمودنی ها بیشتر شوند تغییر نکنند. چه تعداد آزمودنی لازم است تا یک t معنادار بدست آوریم؟ (در فصل ۱۵ به این موضوع بر خواهیم گشت.)

۱۳-۱۰ داده ها تمرین ۵-۱۳ را اصلاح کنید طوری که در ردیف دوم نمرات بالای مربوط به "قبل" با نمرات پایین "بعد" جفت شوند. یک آزمون t اجرا کنید و اثر تغییر داده ها را بر t ببینید.

۱۳-۱۱ با روی هم ریختن جواب تمرین ۱۰-۱۳ و آنچه از قبل در باره همبستگی می دانید، چه انتظاری از اثر میزان همبستگی بین دو متغیر (دو مجموعه داده ها) بر مقدار آزمون t بین آنها دارید؟

۱۳-۱۲ در بخش ۴-۱۳ توضیح داده شد که طرح های نمونه های وابسته با حذف تغییرپذیری آزمودنی به آزمودنی از داده ها، از تأثیر این تغییرپذیری بر داده هایی که آزمون t روی آنها اجرا می شود جلوگیری می کنند. این موضوع توانایی رد فرض صفر نادرست را افزایش می دهد. برداشت خود را از این موضوع بیان کنید.

۱۳-۱۳ مثالی از اثر انتقال را بیان کنید.

فصل ۱۴

آزمون‌های فرضیه برای میانگین‌ها-

دو میانگین مستقل

در فصل ۱۳ پژوهشی را مرور کردیم که در آن یک مجموعه اندازه‌های رفتار نامناسب اجتماعی را قبل و بعد از یک برنامه مداخله بدست آوردیم. در آن مثال آزمودنی‌های واحدی را قبل و بعد از مداخله مورد مشاهده قرار دادیم. هر چند آن کار ممکن است بهترین راه برای ارزیابی اثرات آن برنامه مداخله باشد، در بسیاری آزمایش‌ها یا این کار ممکن نیست یا مناسب نیست که داده‌ها با اندازه‌گیری مکرر از آزمودنی‌های واحد بدست آید. برای مثال اگر ما بخواهیم تعیین کنیم که آیا مردان به لحاظ اجتماعی غیر منطقی‌تر از زنان هستند، واضح است که نمی‌توان افراد واحدی را به عنوان مردان و سپس به عنوان زنان آزمود. به جای آن ما به یک نمونه از مردان و سپس یک نمونه مستقل دیگر از زنان نیاز داریم.

یکی از کاربردهای خیلی معمول آزمون t شامل بررسی تفاوت بین میانگین‌های دو گروه مستقل است. ممکن است ما بخواهیم میانگین دفعات تلاش مورد نیاز را برای رسیدن به حد معینی در یک تکلیف تمایز ساده بینایی در دو گروه موش‌ها مقایسه کنیم- یک گروه تحت شرایط عادی و گروه دیگر تحت شرایط محرومیت حسی برانگیخته می‌شوند. یا در یک مطالعه حافظه ممکن است بخواهیم میانگین سطح یادآوری دو گروه را مقایسه کنیم که مثلاً یکی شامل دانشجویانی است که باید جملات اخباری معلوم و گروه دیگری که باید جملات منفی مجهول را بخاطر بیاورند. مثال آخر این است که ممکن است ما آزمودنی‌ها را در موقعیتی که فرد دیگری نیاز به کمک داشته باشد قرار دهیم. ما می‌توانیم تأخیر در کمک‌رسانی را در دو حالت مقایسه کنیم، یکی وقتی افراد تنها هستند و دیگری زمانی که در گروه هستند.

در هدایت هر آزمایشی با دو گروه مستقل ما خیلی تمایل داریم که ببینیم دو گروه به میزانی با هم تفاوت دارند. پرسش مهم این است که آیا این تفاوت به اندازه کافی بزرگ هست که این نتیجه که

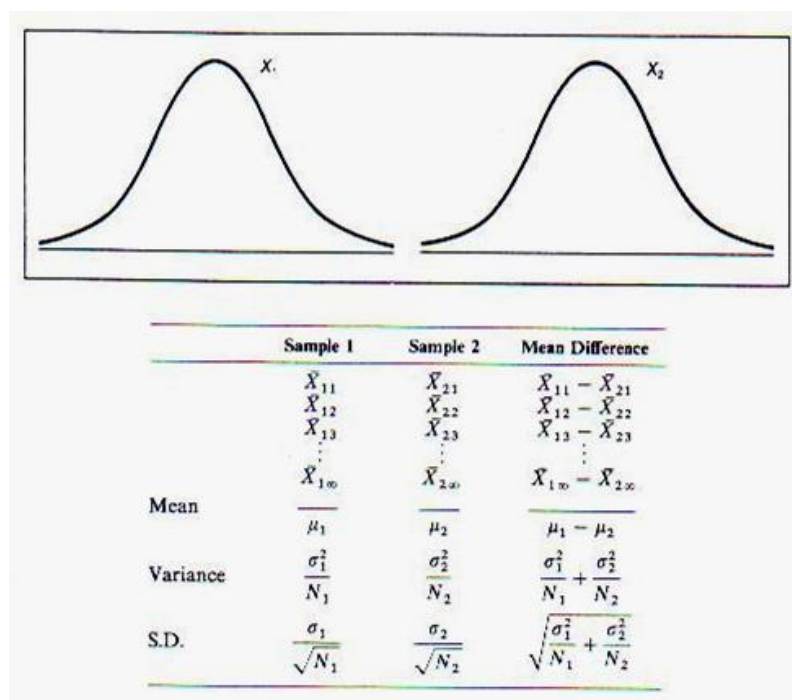
نمونه‌ها از جمعیت‌های متفاوتی برآمده‌اند را تصدیق نماید، مثلاً در مورد رفتار کمک‌رسانی، آیا میانگین تأخیر، در جمعیت آزمودنی‌های منفرد، با میانگین تأخیر در جمعیت آزمودنی‌های گروهی تفاوت دارد؟ قبل از آنکه به یک مثال خاص بپردازیم لازم است توزیع نمونه‌گیری تفاوت‌های بین میانگین‌ها و آزمون t که از آن حاصل می‌شود را بررسی کنیم.

۱۴-۱ توزیع تفاوت‌های بین میانگین‌ها

وقتی می‌خواهیم اختلاف بین میانگین یک جمعیت (μ_1) و میانگین جمعیت دیگری (μ_2) را بررسی کنیم، ما یک فرض صفر به شکل $H_0: \mu_1 - \mu_2 = 0$ یا معادل آن $\mu_1 = \mu_2$ را می‌آزماییم. چون آزمایش این فرض صفر شامل اختلاف بین میانگین‌های نمونه‌های مستقل است، خوب است قدری حوصله کنیم و **توزیع نمونه‌گیری تفاوت‌های بین میانگین‌ها** را بررسی کنیم. فرض کنید ما دو جمعیت بنام X_1 و X_2 با میانگین‌های μ_1 و μ_2 و واریانس‌های σ_1^2 و σ_2^2 داریم. حالا ما از جمعیت X_1 نمونه‌ای به اندازه N_1 و از جمعیت X_2 نمونه‌ای به اندازه N_2 بیرون می‌کشیم و میانگین‌ها و تفاوت‌های بین میانگین‌های هر جفت نمونه را ثبت می‌کنیم. چون ما از هر جمعیت بطور مستقل نمونه می‌گیریم، میانگین‌های نمونه مستقل خواهند بود. (میانگین‌ها فقط در شرایط نادر همبسته‌اند.) چون ما فقط فرض می‌کنیم، ممکن است همچنین همه راه‌ها را برویم و فرض بگیریم که این شیوه را به دفعات بی‌شمار تکرار کرده‌ایم. نتایج در شکل ۱۴-۱ به صورت نمادین آمده است. در بخش پایین این شکل دو ستون اول توزیع‌های نمونه‌گیری $\bar{X}_1$ و $\bar{X}_2$ ، و ستون سوم توزیع نمونه‌گیری تفاوت‌های میانگین‌ها $(\bar{X}_1 - \bar{X}_2)$ را نشان می‌دهند. این ستون سوم بیشتر مد نظر ما می‌باشد، چون علاقه‌مند به بررسی تفاوت‌های بین میانگین‌ها هستیم. میانگین این توزیع را می‌توان به صورت $\mu_1 - \mu_2$ نشان داد. واریانس این توزیع اختلاف‌ها با چیزی نشان داده می‌شود که معمولاً به عنوان **قاعده حاصل جمع واریانس**^۱ خوانده می‌شود، یک شکل محدود شده‌ی آن چنین است:

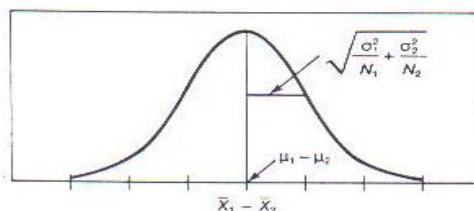
واریانس یک حاصل جمع یا اختلاف دو متغیر مستقل برابر است با حاصل جمع واریانس آنها. ما از قضیه حد مرکزی می‌دانیم که واریانس توزیع $\bar{X}_1$ برابر σ_1^2 / N_1 و واریانس توزیع $\bar{X}_2$ برابر σ_2^2 / N_2 است. از آنجا که متغیرها (میانگین نمونه‌ها) مستقل هستند، واریانس اختلاف این دو متغیر حاصل جمع واریانس‌های آنها است. بنابراین

$$\sigma_{\bar{X}_1 - \bar{X}_2}^2 = \sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2 = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}$$



شکل ۱۴-۱ مجموعه فرضی میانگین‌ها و تفاوت‌های بین میانگین‌ها هنگام نمونه‌گیری از دو جمعیت

با یافتن میانگین واریانس یک مجموعه تفاوت‌های بین میانگین‌ها، بیشتر چیزهایی که ما نیاز داریم را می‌دانیم. حالت کلی توزیع نمونه‌گیری تفاوت‌های میانگین‌ها در شکل ۱۴-۲ نشان داده شده است. آخرین نکته در باره این توزیع به شکل آن بر می‌گردد. یک قضیه مهم در آمار می‌گوید حاصل جمع یا تفریق دو متغیر مستقل دارای توزیع نرمال، خودش دارای توزیع نرمال است. چون شکل ۱۴-۲ اختلاف بین دو توزیع نمونه‌گیری میانگین‌ها را نشان می‌دهد و چون ما می‌دانیم که توزیع نمونه‌گیری از میانگین‌ها برای تعداد قابل قبول نمونه دست‌کم تقریباً نرمال است، پس توزیع شکل ۱۴-۲ خودش هم بایستی دست‌کم تقریباً نرمال باشد.



شکل ۱۴-۲ توزیع نمونه‌گیری تفاوت‌های بین میانگین‌ها

آماره t

با داشتن اطلاعاتی که ما از توزیع نمونه‌گیری از تفاوت‌های میانگین بدست آوردیم، اکنون می‌توانیم بسط شیوه آزمون مناسب را پی بگیریم. برای یک لحظه فرض کنید که آگاهی از واریانس‌های جمعیت (σ_1^2) مشکلی نباشد. پیش از این ما z را به عنوان یک آماره (یک نقطه روی توزیع) منهای میانگین توزیع، تقسیم بر خطای معیار توزیع تعریف کردیم. آماره ما در حالت فعلی ($\bar{X}_1 - \bar{X}_2$)، یعنی تفاوت مشاهده شده بین میانگین‌های نمونه‌ای است. میانگین توزیع نمونه‌گیری ($\mu_1 - \mu_2$) است، و چنانکه دیدیم خطای معیار اختلاف بین میانگین‌ها برابر است با

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$$

بنابراین می‌توانیم بنویسیم

$$\begin{aligned} z &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} \\ &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \end{aligned}$$

مقدار بحرانی برای $\alpha = 0.05$ برابر ± 1.96 است، چنانکه برای آزمون‌های یک نمونه‌ای بود که در فصل ۱۲ مورد بحث قرار گرفتند.

فرمول قبل فایده خاصی غیر از نشان دادن اساس آزمون t ندارد، زیرا ما بندرت از واریانس‌های جمعیت اطلاع داریم که در این فرمول لازم است. (چنین اطلاعی آنقدر نادر است که حتی ارزش ندارد مواردی را تصور کنیم که از آن با خبر هستیم، گرچه موارد اندکی وجود دارد.) به هر حال، همانطور که در مورد یک نمونه‌ای انجام شد، ما می‌توانیم با استفاده از واریانس‌های نمونه به عنوان برآورد کننده واریانس جمعیت بر این مشکل چیره شویم. این مورد قبلاً برای همین منظور در باره t یک نمونه‌ای توضیح داده شد، یعنی نتیجه به صورت t به جای z بحث شد.

$$\begin{aligned} t &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} \\ &= \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \end{aligned}$$

چون فرض صفر عموماً چنین بیان می‌شود که $\mu_1 - \mu_2 = 0$ ، ما معمولاً آن عبارت را از معادله حذف کرده و می‌نویسیم

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{X}_1 - \bar{X}_2}}$$

$$= \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

درآمیختن واریانس‌ها

هر چند معادله t که در بالا بدست آوردیم برای نمونه‌های برابر کاملاً مناسب است، این معادله برای نمونه‌های نابرابر نیازمند اصلاحاتی است. این اصلاحات به منظور برآورد بهتری از واریانس جمعیت طرح شده‌اند. یکی از پیش‌فرض‌های لازم برای استفاده از t برای دو نمونه مستقل این است که $\sigma_1^2 = \sigma_2^2$ (یعنی نمونه‌ها از جمعیت‌های با واریانس برابر باشند)، بدون توجه به اینکه H_0 درست باشد یا نباشد. چنین فرضی اغلب یک فرض منطقی است و به عنوان پیش‌فرض **تجانس (همگنی) واریانس**^۱ نامیده می‌شود. بیشتر وقت‌ها ما آزمایشی را با دو گروه از آزمودنی‌ها شروع می‌کنیم که معادل هم هستند و سپس روی یک (یا هر دو) گروه چیزی اعمال می‌کنیم که نمرات آزمودنی‌ها را بالاتر یا پایین‌تر می‌برد. در این حالت چنین فرض می‌شود که واریانس‌ها تحت تأثیر قرار نخواهند گرفت. (حتماً بخاطر دارید که افزودن یک مقدار ثابت به یک مجموعه نمرات یا کاستن آن، روی واریانس اثری ندارد.) از آنجا که فرض می‌شود واریانس‌های جمعیت برابر هستند، این واریانس مشترک را می‌توان با نماد σ^2 بدون اندیس^۲ نشان داد.

ما در داده‌ها دو برآورد کننده از σ^2 ، بنام s_1^2 و s_2^2 داریم. خوب است نوعی مقدار متوسط برای s_1^2 و s_2^2 بدست آوریم به گونه‌ای که مقدار بهتری از برآورد متوسط نسبت به دو برآورد کننده جداگانه فراهم کند. ما نمی‌خواهیم میانگین حسابی ساده را بدست آوریم، چون این کار وزن برابری به دو برآورد کننده خواهد داد، حتی اگر یکی بر پایه مشاهدات خیلی بیشتری باشد. چیزی که می‌خواهیم یک **مقدار متوسط وزن یافته**^۳ است، که در آن واریانس‌های نمونه بر اساس درجات آزادی $(N_i - 1)$ وزن داده شده‌اند. اگر این برآورد جدید را s_p^2 بنامیم، آنگاه

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

صورت این کسر نشان دهنده مجموع واریانس‌ها است، که هر یک بر مبنای درجات آزادی آنها وزن یافته‌اند، و مخرج مجموع وزن‌ها، یا معادل آن، درجات آزادی برای s_p^2 است.

مقدار متوسط وزن داده شده‌ی این دو واریانس نمونه را معمولاً به عنوان یک برآورد **واریانس ترکیبی** (یک نام نامأنوس اما توضیح دهند) خوانده می‌شود. با داشتن تعریف برآورد ترکیبی (s_p^2)، می‌توانیم بنویسیم

1. homogeneity of variance
2. subscript
3. weighted average

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{N_1} + \frac{s_p^2}{N_2}}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2 \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

توجه داشته باشید که هر دوی این فرمول های t و فرمولی که در بخش قبلی به کار بردیم شامل تقسیم تفاوت بین میانگین های نمونه ها بر برآوردی از خطای معیار تفاوت بین میانگین ها است. تنها فرق آنها برمی گردد به روشی که این خطا برآورد می شود. وقتی اندازه نمونه ها برابر باشد، اصلا فرقی نمی کند که واریانس ها را جمع کنید یا نکنید؛ نتیجه t یکسان خواهد بود. اما وقتی اندازه نمونه ها یکسان نباشد، جمع کردن می تواند اثر کاملا متفاوتی داشته باشد.

درجات آزادی برای t

توجه دارید که دو واریانس نمونه (s_1^2 و s_2^2) در محاسبه t آمده اند. هریک از این واریانس ها بر اساس مجذور انحرافات از میانگین خود بنا شده و بنابراین هر واریانس نمونه دارای $N_i - 1$ درجه آزادی است. برای دو نمونه، پس ما خواهیم داشت $(N_1 - 1) + (N_2 - 1) = N_1 + N_2 - 2$ درجه آزادی. بنابراین t برای دو نمونه مستقل براساس $N_1 + N_2 - 2$ درجه خواهد بود.

مثال: به یاد آوردن جملات معلوم و مجهول

برای توضیح کاربرد t بعنوان یک آزمون تفاوت بین دو میانگین مستقل، ما ابتدا به یک مثال فرضی در زمینه یادگیری کلامی می پردازیم. بگذارید فرض کنیم که یک محقق با سردبیر یک مجله که پافشاری دارد از عبارات مجهول به عنوان سوم شخص ("مشخص شد که...") به جای عبارات معلوم ("من دریافتم که...") استفاده کند درگیر شده است. محقق ما می خواهد نشان دهد که آزمودنی ها جملات معلوم ("جان ناقوس را به صدا در آورد") را بهتر از جملات منفی مجهول ("ناقوس بوسیله مری به صدا در نیامد") بیاد می آورند. آزمایشگر ۳۵ آزمودنی را انتخاب نمود و آنها را به صورت تصادفی به دو گروه تخصیص داد. بطور تصادفی گروه اول (گروه A) دارای ۱۵ آزمودنی و گروه دوم (گروه P) دارای ۲۰ آزمودنی بود. گروه A ۲۵ جمله معلوم را خوانده و یک تست یادآوری بعد از هر جمله انجام داد. گروه P شکل مجهول جملات به کار رفته برای گروه A را شنید و یادآوری کرد. متغیر وابسته تعداد جملات درست به خاطر آورده شده توسط هر آزمودنی بود و داده ها در جدول ۱-۱۴ ارائه شده است.

پیش از در نظر گرفتن هر آزمون آماری، و حتی در حالت ایده آل آن، قبل از اینکه داده ها جمع آوری شوند، ما بایستی چند ویژگی آزمون را مشخص کنیم. ابتدا ما باید فرضیه های صفر و مقابل را مشخص کنیم:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

فرض مخالف دو طرفه است (ما اگر $\mu_2 < \mu_1$ یا $\mu_2 > \mu_1$ باشد H_0 را رد می‌کنیم)، و بنابراین از یک آزمون دو دامنه استفاده می‌کنیم. برای هماهنگی با سایر مثال‌های این کتاب، ما $\alpha = 0.05$ می‌گیریم، به خاطر داشته باشیم که در باره این دو تصمیم توصیه خاصی وجود ندارد.

جدول ۱۴-۱ یادآوری جملات علوم و مجهول

معلوم	مجهول
۱۳	۱۷
۱۸	۱۷
۱۷	۲۱
۱۳	۱۸
۱۴	۲۲
۱۳	۱۸
۱۸	۱۶
۱۹	۱۵
۱۶	۱۸
۱۴	۲۰
۱۳	۲۱
۱۵	۱۶
۱۴	۱۵
۱۶	۱۶
۱۵	۲۰
۱۵	
۱۳	
۱۷	
۱۷	
۱۵	
۱۵/۲۵	۱۸/۰۰ میانگین
۳/۶۷۱	۵/۲۸۶ واریانس

با فرض صفر به شکل فوق، ما حالا می‌توانیم t را حساب کنیم.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\frac{s_{\bar{X}_1 - \bar{X}_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}}$$

چون ما $\mu_1 - \mu_2$: H_0 را می‌آزماییم، عبارت $\mu_1 - \mu_2$ از معادله حذف شده است. وقتی که واریانس‌ها را ترکیب کنیم، می‌شود

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} = \frac{14(5.286) + 19(3.671)}{15 + 20 - 2} = \frac{74.004 + 69.749}{33} = 4.356$$

توجه کنید که واریانس ترکیبی قدری به مقدار s_2^2 نزدیک‌تر است تا s_1^2 ، زیرا s_2^2 در فرمول وزن بیشتری دارد. پس

$$= \frac{(18.00 - 15.25)}{\sqrt{\frac{4.356}{15} + \frac{4.356}{20}}} = \frac{(2.75)}{\sqrt{0.5082}} = \frac{2.75}{0.713} = 3.86$$

برای این مثال ما $N_1 - 1 = 14$ درجه آزادی (df) برای گروه A، و $N_2 - 1 = 19$ درجه آزادی (df) برای گروه p داریم که حاصل جمع آن می‌شود $N_1 - 1 + N_2 - 1 = 33$ درجه آزادی (df). از توزیع نمونه‌گیری t در بخش پیوست، $\pm 2.04 = t_{0.05}(33)$ است (با ترکیب خطی). از آنجا که مقدار t_{obs} از t_a بزرگتر است، ما H_0 (در سطح $\alpha = 0.05$) رد می‌کنیم و نتیجه می‌گیریم که بین میانگین جمعیت‌هایی که مشاهدات ما از آنها بر آمده‌اند تفاوت معناداری وجود دارد. به عبارت دیگر ما نتیجه می‌گیریم که $\mu_1 \neq \mu_2$ (بطور آماری) و $\mu_1 > \mu_2$ (بطور عملی) است. به زبان متغیرهای آزمایشی، جملات اخباری معلوم بهتر از جملات مجهول به یاد آورده می‌شوند.

۲-۱۴ ناهمگنی واریانس

چنانکه دیدیم، یکی از پیش‌فرض‌های آزمون t برای دو نمونه مستقل فرض همگنی (تجانس) واریانس‌ها ($\sigma_1^2 = \sigma_2^2$) است. وقتی این فرض برقرار نباشد ما در شرایطی هستیم که به آن **ناهمگنی واریانس**^۱ ($\sigma_1^2 \neq \sigma_2^2$) گویند. کار زیادی برای آزمایش اثر عملی ناهمگن بودن واریانس در آزمون t صورت گرفته است. در نتیجه ما به بعضی نتایج کلی در باره تحلیل داده‌ها رسیده‌ایم که برای وضعیت ناهمگن بودن واریانس مناسب است. نخستین نکته‌ای که باید به خاطر داشت این است که فرض تجانس واریانس به

1. heterogeneity of variance

واریانس‌های جمعیت اشاره دارد، ما بندرت انتظار داریم واریانس‌های نمونه دقیقاً مساوی باشند حتی اگر واریانس‌های جمعیت مساوی باشند. بر اساس مطالعات نمونه‌ای که ما به عمل آورده‌ایم، قاعده کلی این است که اگر یک واریانس نمونه بیش از چهار برابر دیگری نباشد و اگر اندازه نمونه‌ها برابر یا تقریباً برابر باشند شما می‌توانید ادامه داده و t را بصورت عادی محاسبه کنید. در این شرایط به نظر نمی‌رسد ناهمگنی واریانس تأثیر جدی بر نتایج شما داشته باشد. از سوی دیگر اگر یک واریانس نمونه بیش از چهار برابر دیگری باشد یا اگر واریانس‌ها کاملاً نابرابر باشند و اندازه نمونه‌ها هم نابرابر باشد، آنگاه یک شیوه جایگزین ممکن است لازم باشد. به کارگیری این فرایند البته آسان است. بطور ساده با استفاده از برآوردهای جداگانه واریانس (یعنی ترکیب نکردن) t را حساب کنید. سپس به سراغ جداول t رفته از دومقدار درجه آزادی $N_1 - 1$ و $N_2 - 1$ هر کدام که کمتر بود به عنوان درجات آزادی استفاده کنید (به جای $N_1 + N_2 - 2$). این یک آزمون محافظه‌کارانه است، یعنی اگر H_0 درست باشد احتمال کمتری برای خطای نوع I نسبت به مقدار اسمی که α پیشنهاد می‌کند وجود دارد. به عنوان مثالی برای این روش، فرض کنید داده‌های زیر را داشته باشیم:

$$\begin{aligned}\bar{X}_1 &= 111/53 & \bar{X}_2 &= 108/38 \\ s_1^2 &= 19/65 & s_2^2 &= 3/06 \\ N_1 &= 10 & N_2 &= 18\end{aligned}$$

آنگاه

$$\begin{aligned}t &= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{111.53 - 108.38}{\sqrt{\frac{19.65}{10} + \frac{3.06}{18}}} \\ &= \frac{3.15}{1.461} = 2/16\end{aligned}$$

از آنجا که واریانس‌ها خیلی نابرابر هستند (یکی بیش از شش برابر دیگری است)، ما آنها را جمع نمی‌کنیم. مقدار $N_1 - 1$ و $N_2 - 1$ به ترتیب، ۹ و ۱۷ است و ما با توجه به ۹ درجه آزادی (مقدار کوچکتر از ۹ و ۱۷) مقدار t را از طریق جدول مربوط به مقادیر بحرانی ارزیابی می‌کنیم. در اینجا می‌بینیم که $2/16$ $\pm t_{.05}(9)$ است، که از مقدار به دست آمده بزرگتر است. بنابراین H_0 را رد نخواهیم کرد. راه‌حل‌های درست‌تری (کمتر محافظه‌کارانه‌ای) برای مسأله ناهمگنی واریانس وجود دارد، که بسته به محاسبه یک درجه آزادی متناسب بین میزان کوچکتر $N_1 - 1$ و $N_2 - 1$ از یک سو و $N_1 + N_2 - 2$ از سوی دیگر دارد. برنامه‌های کامپیوتری زیادی مانند SPSS و Minitab چنین تطبیقی را فراهم می‌کنند. روش محافظه‌کارانه‌ای که در اینجا ارائه شد از بسیاری جهات کفایت می‌کند.

۳-۱۴ نرمال نبودن توزیع ها

قبلا دیدیم که پیش فرض دیگر استفاده صحیح از آزمون t این است که جمعیت (هایی) که داده ها از آن نمونه گیری شده اند دارای توزیع نرمال باشند- یا حداقل توزیع نمونه گیری تفاوت های بین میانگین ها نرمال باشد. در کل، مادامی که توزیع های داده های نمونه ای توده ای شکل باشند (در مرکز زیاد باشند و در دو جانب کاهش یابند)، آزمون متمایل به داشتن اعتبار است. این موضوع بخصوص برای نمونه های بزرگ (بزرگتر از ۳۰ مورد) صادق است، زیرا در این صورت، قضیه حد مرکزی اغلب تضمین کننده ی نرمال بودن توزیع نمونه گیری تفاوت بین میانگین ها است.

۴-۱۴ مثالی دیگر با دو نمونه مستقل

در فصل ۱۱ بطور کوتاه به تحقیق دوب و گراس (۱۹۶۸) اشاره کردیم که اثرات وضعیت اجتماعی- اقتصادی را بر بوق کشیدن بررسی کردند. در مطالعه آنها راننده با ماشین سطح پایین یا بالا بعد از سبز شدن چراغ راهنمایی بدون حرکت باقی می ماند، و آزمایشگر رفتار راننده ماشین پشت سر او را مشاهده می کرد. یکی از سنجش های آنان ثانیه هایی بود که بین روشن شدن چراغ و شروع به بوق زدن راننده پشتی فاصله می افتاد. داده های نشان داده شده در جدول ۲-۱۴ نمونه کوچکتري از اطلاعات تحقیق دوب و گراس است.

جدول ۲-۱۴ تأخیر بوق زدن (به ثانیه) به عنوان کنش وضعیت اجتماعی- اقتصادی

گروه وضعیت بالا ($N = 20$)					گروه وضعیت پایین ($N = 15$)				
۱۳/۲۰	۵/۲۹	۱۱/۲۲	۷/۸۳	۹/۱۰	۱۰/۵۹	۶/۸۵	۸/۵۸	۶/۴۲	۱/۶۸
۱۴/۰۵	۸/۴۰	۷/۴۱	۳/۸۷	۱۱/۷۹	۱۲/۳۱	۴/۹۸	۴/۸۴	۹/۴۴	۳/۲۶
۶/۸۴	۱۱/۷۹	۹/۸۱	۸/۱۱	۴/۴۴	۸/۱۴	۶/۷۱	۷/۸۶	۶/۱۳	۹/۰۱
۹/۵۳	۹/۹۵	۱۰/۶۶	۸/۶۸	۱۲/۶۴					
$\bar{X}_2 = 9/23$ $s_{X2} = 2/82$ $s_{X2}^2 = 7/95$					$\bar{X}_1 = 7/12$ $s_{X1} = 2/77$ $s_{X1}^2 = 7/67$				

دوب و گراس علاقه مند بودند این فرض صفر که تأخیر در واکنش به توقف ماشین سطح پایین برابر است با تأخیر در واکنش به ماشین سطح بالا (یعنی $\mu_1 = \mu_2$; H_0) را بیازمایند. آنها یک فرض مقابل دو دامنه ($H_1: \mu_1 \neq \mu_2$) و $\alpha = 0.05$ را در نظر گرفتند. بررسی ظاهری داده ها گویای این است که واریانس های نمونه تقریباً برابرند ($s_1^2 = 7/67$ و $s_2^2 = 7/95$) و اینکه داده ها دست کم تک نمایی و متقارن هستند، طوری که یک آزمون t به نظر مناسب می آید. از آنجا که نمونه های ما نابرابر هستند، واریانس های نمونه را جمع می کنیم. واریانس جمع شده به این صورت است:

$$\begin{aligned}
 s_p^2 &= \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \\
 &= \frac{14(7.67) + 19(7.95)}{15 + 20 - 2} = \frac{107.38 + 151.05}{33} = \frac{258.43}{33} \\
 &= ۸۳.۷
 \end{aligned}$$

آنگاه با استفاده از واریانس‌های ترکیبی، t برابر است با

$$\begin{aligned}
 t &= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{N_1} + \frac{s_p^2}{N_2}}} \\
 &= \frac{7.12 - 9.23}{\sqrt{\frac{7.83}{15} + \frac{7.83}{20}}} \\
 &= \frac{-2.11}{\sqrt{0.9135}} \\
 &= -۲/۲۱
 \end{aligned}$$

با مراجعه به جدول t بخش پیوست در می‌یابیم که $\alpha = ۰/۰۵$ و $df = ۳۳$ داریم $\pm ۲/۰۴$ (۳۳) t . با توجه به اینکه $t_{obs} = -۲/۲۱$ در ناحیه رد قرار دارد، ما H_0 را رد کرده و نتیجه خواهیم گرفت که $\mu_1 \neq \mu_2$ است. به بیان عملیاتی ما نتیجه می‌گیریم وقتی یک راننده با ماشین سطح پایینی که راه را بند آورده روبرو می‌شود زودتر بوق می‌کشد تا وقتی که او با ماشین سطح بالا مواجه می‌شود.

۵-۱۴ حدود اطمینان $\mu_1 - \mu_2$

علاوه بر آزمون فرض صفر در باره میانگین‌های جمعیت (یعنی آزمون $H_0: \mu_1 - \mu_2 = 0$)، گاهی خوب است حدود اطمینان برای تفاوت بین μ_1 و μ_2 تعیین شود. منطق تعیین این حدود اطمینان دقیقاً همانند حالت یک نمونه‌ای در فصل ۱۲ است. محاسبات دقیقاً یکسان است به جز اینکه در اینجا از تفاوت بین میانگین‌ها و خطای معیار تفاوت بین میانگین به جای میانگین و خطای معیار استفاده می‌کنیم. بنابراین این برای سطح اطمینان ۹۵٪ در باره $\mu_1 - \mu_2$ ما داریم

$$CI_{.95} = (\bar{X}_1 - \bar{X}_2) \pm t_{.۰۵} s_{\bar{X}_1 - \bar{X}_2}$$

در مورد مثال بوق کشیدن داریم

$$\begin{aligned}
 CI_{.95} &= (۷/۱۲ - ۹/۲۳) \pm ۲/۰۴ \left(\sqrt{\frac{7.83}{15} + \frac{7.83}{20}} \right) \\
 &= -۲/۱۱ \pm (۰/۹۵۵۸) = -۲/۱۱ \pm ۱/۹۵ \\
 -۴/۰۶ &\leq (\mu_1 - \mu_2) \leq -۰/۱۶
 \end{aligned}$$

به احتمال ۰/۹۵ یک فاصله $-۴/۰۶$ تا $-۰/۱۶$ تفاوت واقعی در تأخیر بوق زدن برای ماشین‌های سطح پایین و بالا را در بر می‌گیرد.

۱۴-۶ مثال کار شده پایانی

داده‌های زیر بر اساس مطالعه ایزنک (۱۹۷۴) است که در میان چیزهای دیگر، سطح یادآوری آزمودنی‌های جوان و مسن‌تر را مقایسه کرد. ایزنک می‌خواست این فرضیه را بیازماید که در زمان نیاز به پردازش کلامی اطلاعات (فهرست واژه‌ها)، آزمودنی‌های مسن‌تر پردازش کمتری صورت می‌دهند و بنابر این واژه‌های کمتری را به یاد می‌آورند. (در این تحقیق او همچنین نشان داد زمانی که نیاز به پردازش عمیق وجود نداشت، بین دو گروه سنی از لحاظ یادآوری تفاوتی وجود نداشت.) داده‌های زیر طرح شده‌اند تا میانگین و انحراف معیارهای مشابهی همانند دو شرایط مطالعه ایزنک داشته باشد، و به گروه‌هایی اشاره دارد که به آنها گفته شده بود واژه‌ها را به گونه‌ای به خاطر بسپارند که بتوانند بعداً آنها را به یاد آورند. متغیر وابسته تعداد واژه‌های درست به یادآورده شده هستند.

	سن	
	جوان‌تر (۱۸-۳۰)	مسن‌تر (۵۵-۶۵)
	۲۱	۱۰
	۱۹	۱۹
	۱۷	۱۴
	۱۵	۵
	۲۲	۱۰
	۱۶	۱۱
	۲۲	۱۴
	۲۲	۱۵
	۱۸	۱۱
	۲۱	۱۱
$\sum X$	۱۹۳	۱۲۰
$\sum X^2$	۳۷۸۹	۱۵۶۶
N	۱۰	۱۰

ابتدا لازم است که فرض صفر، سطح معناداری و اینکه آیا از آزمون یک دامنه یا دو دامنه استفاده خواهیم کرد را مشخص کنیم. ما می‌خواهیم این فرض صفر را آزمایش کنیم که دو گروه سنی میزان اطلاعات

یکسانی را به یاد خواهند آورد، طوری که داریم $\mu_1 = \mu_2$; H_0 . ما آلفا را همچون جاهای دیگر این کتاب، در سطح $\alpha = 0.05$ تنظیم می‌کنیم. در آخر، ما یک آزمون دو دامنه را بر می‌گزینیم زیرا بطور منطقی امکان دارد که هر یک از گروه‌ها یادآوری بهتری داشته باشد. سپس باید میانگین‌ها و واریانس‌ها را حساب کنیم.

$$\bar{X}_1 = 193/10 = 19.3 \quad \bar{X}_2 = 120/10 = 12.0$$

$$S_X^2 = \frac{3789 - \frac{193^2}{10}}{9} \quad S_Y^2 = \frac{1566 - \frac{120^2}{10}}{9}$$

$$= 7.122 \quad = 14.000$$

با اندازه‌های یکسان نمونه نیازی به ترکیب واریانس‌ها نیست، زیرا t حاصل در هر دو حالت یکسان است. اما به عنوان یک مثال در اینجا آن را انجام می‌دهیم.

$$S_p^2 = \frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1 + N_2 - 2} = \frac{9(7.122) + 9(14.000)}{18} = 10.561$$

در آخر ما می‌توانیم t را با استفاده از برآورد واریانس جمعی حساب کنیم.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}} = \frac{19.3 - 12.0}{\sqrt{\frac{10.561}{10} + \frac{10.561}{10}}} = \frac{7.300}{\sqrt{2.112}} = 5.02$$

برای این مثال ما $N_1 + N_2 - 2 = 18$ درجه آزادی داریم. با مراجعه به جدول پیوست مربوط به مقادیر بحرانی t ، به مقدار $t_{0.05} = 2.101$ بر می‌خوریم. چون $5.02 > 2.101$ است، ما H_0 را رد می‌کنیم و نتیجه می‌گیریم که میانگین‌های دو جمعیت برابر نیستند. از این داده‌ها بر می‌آید که آزمودنی‌های جوان‌تر واژه‌های بیشتری را نسبت به آزمودنی‌های مسن‌تر به یاد آورده‌اند. این نتیجه بیانگر آن نیست که آزمودنی‌های مسن‌تر به اندازه جوان‌ترها با هوش نیستند، بلکه فقط می‌گویند آنها در این تکلیف به دلایلی خوب عمل نکرده‌اند (شاید چندان علاقه‌مند به اینگونه تکالیف نبوده‌اند).

۷-۱۴ خلاصه

در این فصل ما به استفاده از آزمون t برای بررسی فرض صفر یکسان بودن میانگین دو جمعیت پرداختیم. این آزمون بر اساس میانگین‌های نمونه‌ای برای دو نمونه مستقل است و معمول‌ترین شکل آزمون t است. ما همچنین پیش‌فرض‌های زیربنایی استفاده از t ، شیوه‌های مورد استفاده برای زمانی که فرض تجانس واریانس‌ها برآورده نمی‌شود، و حدود اطمینان اختلاف میانگین‌ها را بررسی کردیم. بعضی از مهمترین عناوین این فصل عبارتند از:

- توزیع نمونه‌گیری تفاوت بین میانگین‌ها
- تجانس واریانس
- قانون جمع واریانس
- واریانس ترکیب شده

- خطای معیار تفاوت های بین میانگین ها
- اندازه متوسط وزن یافته
- ناهمگنی واریانس
- ابهامات

۸-۱۴ تمرین ها

۱۴-۱ فرض کنید ما مطالعه شرح داده شده در تمرین ۱-۱۳ را از نو مطرح کنیم تا تفاوت آزمودنی ها را در شرایط آموزش تقلید و راهنمایی فیزیکی بدانیم. داده های زیر را بدست آورده ایم:

تقلید	۱۴	۱۱	۱۹	۸	۴	۹	۱۲	۵	۱۴	۱۷	۱۸	۰	۲	۸	۶
راهنمایی	۱۰	۱۴	۵	۸	۱	۱۰	۱۳	۱۴	۰	۱	۴	۲	۳	۴	۱۴
فیزیکی															

به چه نتیجه ای خواهیم رسید؟

۱۴-۲ عمده ترین مشکل داده های تمرین ۱-۱۴ از لحاظ یافتن تفاوت معنادار چیست؟

۱۴-۳ چرا طرح آزمایشی که در تمرین ۱-۱۳ به کار گرفته شد امکان ایجاد تفاوت معنادار داشت اما طرح تمرین ۱-۱۴ این امکان را ندارد؟

۱۴-۴ نه کودک بطور تصادفی به هر یک از دو گروه تخصیص داده شده اند از آنها خواسته شد با خمیردندان X یا Y مسواک کنند. در پایان ۶ ماه تعداد کرم خوردگی دندان های هر کودک شمارش شد. داده ها چنین است:

خمیردندان X	۳	۳	۲	۲	۳	۴	۳	۱	۰
خمیردندان Y	۱	۳	۲	۱	۰	۳	۱	۰	۲

آزمون t مناسب را اجرا کنید.

۱۴-۵ نقش تقسیم تصادفی در تمرین ۱۴-۴ چیست؟

۱۴-۶ در مقایسه ی برنامه های مختلف مورد حمایت یک سازمان طرفدار کاهش وزن، ۲۰ آزمودنی در برنامه A و ۲۰ آزمودنی در برنامه B ثبت نام کردند. میزان کاهش وزن در طی ۶ ماه برای آزمودنی هایی که برنامه را کامل کردند به شرح زیر است.

برنامه A	۲۵	۲۱	۱۸	۲۰	۲۲	۳۰
برنامه B	۱۵	۱۷	۹	۱۲	۱۱	۱۹
	۱۴	۱۸	۱۶	۱۰	۵	۱۳

آزمون t مناسب را پس از در نظر گرفتن واریانس های گروهی اجرا کنید.

۱۴-۷ داده های زیر نشان دهنده داده های دیگری هستند که در مطالعه تمرین ۱۴-۶ توضیح داده شد.

برنامه A	۲۵	۲۱	۱۸	۲۰	۲۲	۳۰
برنامه B	۱۵	۱۷	۹	۱۲	۱۱	۱۹
	۱۴	۱۸	۱۶	۱۰	۵	۱۳

آزمون t را پس از در نظر گرفتن دو واریانس نمونه محاسبه و ارزیابی کنید.

۱۲-۸ در تمرین ۷-۱۴ میانگین‌ها تنها آماره مهم نیستند. چه چیز دیگری را بایستی بیشتر مورد توجه قرار داد؟

۱۴-۹ در باره مفهوم سوگیری آزمایشگر خیلی صحبت شده است، اشاره به این واقعیت دارد که بسیاری آزمایشگران تمایل دارند داده‌ها در جهات مورد نظر بدست آیند. فرض کنید ما از دانشجویان به عنوان آزمایشگر استفاده کنیم. به همه گفته شده که به آزمودنی‌ها قبل از آزمایش کافئین داده می‌شود، اما به نیمی از آنها گفته شود انتظار داریم کافئین کارایی را بهتر کند، و به نیمی دیگر گفته شود انتظار داریم کارایی را بدتر کند. متغیر وابسته تعدادی برنامه ساده حساب بود که آزمودنی می‌توانست در ۲ دقیقه حل کند. داده‌های بدست آمده به شرح زیر است:

انتظار عملکرد ۲۲ ۱۵ ۱۳ ۱۸ ۱۵ ۲۰ ۲۵ ۲۲
خوب

انتظار عملکرد بد ۱۴ ۱۸ ۱۷ ۱۲ ۲۱ ۲۱ ۲۴ ۱۴
شما چه نتیجه‌ای می‌گیرید؟

۱۴-۱۰ در تمرین ۹-۱۴ سطح اطمینان ۹۵٪ را روی $\mu_1 - \mu_2$ برای داده‌ها حساب کنید.

۱۴-۱۱ در تمرین ۷-۱۴ سطح اطمینان ۹۵٪ را روی $\mu_1 - \mu_2$ برای داده‌ها حساب کنید.

۱۴-۱۲ آزمایشگری که در زمینه تصمیم‌گیری مطالعه می‌کند از ۱۲ کودک خواست هر تعداد مسأله را که می‌توانند در مدت ۱۰ دقیقه حل کنند. به یک گروه گفته شد این یک تست از توانایی ذاتی حل مسأله است و به گروه دیگر گفته شد این فقط برای سرگرمی است. آزمون t مناسب را محاسبه و تفسیر کنید. متغیر وابسته تعداد مسائل حل شده است. داده‌ها به این شرح است:

توانایی ذاتی ۷ ۳ ۸ ۵ ۴
سرگرمی ۹ ۷ ۹ ۶ ۱۱

۱۴-۱۳ آزمایشگر دیگری آزمایش تمرین ۱۴-۱۵ را تکرار کرد و دقیقاً همان نتایج را بدست آورد. اما او احساس کرد بهتر است داده‌ها را به صورت دقیقه بر مسأله ثبت کند (مثلاً، ۴ مسأله در ۱۰ دقیقه = $2/50 = 10/4$). بدین ترتیب داده چنین بودند

توانایی ذاتی ۱/۴۳ ۳/۳۳ ۱/۲۵ ۲/۰۰ ۲/۵۰
سرگرمی ۱/۱۱ ۱/۴۳ ۱/۱۱ ۱/۶۷ ۰/۹۱

این داده‌ها را با آزمون t مناسب تحلیل و تفسیر نمایید.

۱۴-۱۴ با توجه به تعریفی که در این فصل از متوسط وزن بیان گردید، معین کنید اگر دو اندازه نمونه برابر باشند برآورد واریانس ترکیبی (s_p^2) چقدر خواهد بود. (راهنمایی: N_1 و N_2 را برابر N بگیرید.)

۱۴-۱۵ در تمرین قبلی اگر بدون توجه به N_i ، $s_1^2 = s_2^2$ باشد چه خواهد شد؟

۱۴-۱۶ از مقایسه تمرین ۱۲-۱۴ و ۱۳-۱۴ چه می‌فهمی

فصل ۱۵

توان

بسیاری از کارهای آماری چنان به کار می‌روند که گویی در واقع تحلیل نتایج آزمایشی بدنبال به حداقل رساندن (یا دست کم کنترل) احتمال خطای نوع I (α) هستند. افراد وقتی آزمایشاتی را طرح می‌کنند، عموماً گرایش به نادیده گرفتن این واقعیت دارند که یک احتمال نوع دیگر خطا (β)، هم وجود دارد، که به خطای نوع II معروف است. در حالی که خطای نوع I متوجه مشکل یافتن تفاوتی است که وجود ندارد، خطای نوع II متوجه یک سری مشکلات مشابه در نیافتن تفاوتی است که وجود دارد. وقتی ما به ارزش اساسی زمان و پولی که صرف یک آزمایش می‌شود توجه می‌کنیم، کوتاه‌اندیشی آزمایشگران خواهد بود اگر این واقعیت را تشخیص ندهند که برای یافتن اثری که بدنبال آن بوده‌اند از ابتدا شانس خیلی کمی داشته‌اند، حتی اگر چنین اثری واقعا در جمعیت وجود داشته باشد- و حتی اگر این یک اثر ناچیز نباشد که ارزش یافتن را داشته باشد.

دلایل خوبی وجود دارد که چرا محققان همواره خواسته‌اند خود را به خطای نوع II مشغول نکنند. تا همین اواخر، بسیاری کتاب‌ها کلاً این مسأله را نادیده می‌گرفتند، و کتاب‌هایی هم که این بحث را مطرح می‌کردند مطالب را به گونه‌ای مورد بحث قرار می‌دادند که برای خواننده قابل فهم نبود. در ۲۰ سال گذشته روان‌شناسی بنام جاکوب کوهن این مسأله را در مقالات مختلفی به‌روشنی مورد بحث قرار داده و توضیح داده است. کوهن (۱۹۷۷) مطالب انبوه و گسترده‌ای را در این زمینه منتشر کرد. در مقاله والکویز، اوین و کوهن (۱۹۸۲) این مطالب به طور ساده‌تر به شیوه‌ی تقریب بیان شده که در این فصل از آن استفاده شده است.

صحبت از خطاهای نوع II یک شیوه منفی برای توضیح مشکل است، زیرا به ما گوشزد می‌کند که دچار خطایی شده‌ایم. رویکرد مثبت‌تر این است که از جنبه **توان** به آن بپردازیم، که به عنوان احتمال به درستی رد کردن H_0 نادرست است. به بیان دیگر، توان برابر است با $1 - \beta$. وقتی می‌گوییم توان یک

طرح آزمایشی خاص ۰/۶۵ است، منظورمان این است که چنانچه فرض صفر در سطح مورد انتظار آزمایشگر نادرست باشد، به احتمال ۰/۶۵ نتایج این آزمایش آن را رد می‌کند. یک آزمایش پرتوان‌تر آزمایشی است که احتمال قوی‌تری برای رد فرض H_0 نادرست داشته باشد.

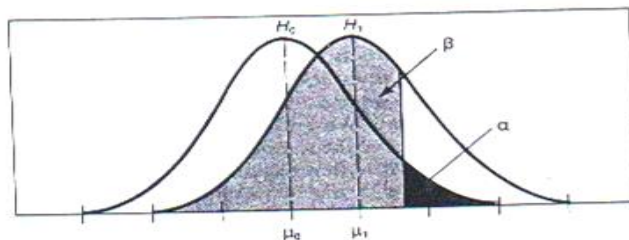
در این فصل رویکرد تقریب به توان حقیقی یک تست در پیش گرفته شده است. این تقریب یک چیز عالی است، بخصوص با توجه به این واقعیت که ما عملاً مواظب این نیستیم که توان برابر ۰/۸۵ است یا ۰/۸۳، بلکه بین ۰/۸۰ یا ۰/۳۰ بودن دقت می‌کنیم. به منظور توضیح مطالب اصلی فعلاً فرض می‌کنیم که مایل به بکارگیری یک آزمون t برای بررسی میانگین یک نمونه در برابر یک میانگین جمعیتی خاص هستیم، گرچه این شیوه بی‌درنگ به آزمون سایر فرضیات تعمیم داده می‌شود.

۱۵-۱ مفهوم اصلی

برای مرور آنچه قبلاً در فصل ۱۱ بیان شد، به دو توزیع در شکل ۱۵-۱ توجه کنید. توزیع سمت چپ (با عنوان H_0) توزیع نمونه‌گیری میانگین است در حالتی که فرض صفر درست و میانگین جمعیت برابر μ_0 است. دنباله پر رنگ این توزیع در سمت راست نشان‌دهنده α یعنی احتمال خطای نوع I است، با این پیش‌فرض که ما یک آزمون یک دامنه را به کار می‌بریم. (در غیر این صورت $\alpha/2$ را نشان می‌دهد.) این ناحیه شامل مقادیری از یک میانگین نمونه‌ای است که از مقادیر معنادار t بدست می‌آید.

توزیع دوم (H_1) توزیع نمونه‌گیری میانگین در حالتی است که H_0 نادرست و میانگین حقیقی μ_1 باشد. خیلی روشن است که حتی وقتی H_0 نادرست است بسیاری از میانگین‌های نمونه‌ای (و بنابراین مقادیر t مربوط به آنها) در این حال در سمت چپ مقدار بحرانی قرار می‌گیرند که باعث می‌شود ما نتوانیم یک H_0 نادرست را رد کنیم، و در نتیجه دچار خطای نوع II شویم. احتمال این خطا با سایه (کم‌رنگ) در شکل ۱۵-۱ نشان داده شده و به عنوان β از آن یاد می‌شود.

در نهایت، وقتی H_0 نادرست است و آزمون آماری در سمت راست ناحیه بحرانی واقع می‌شود، ما H_0 نادرست را بدرستی رد می‌کنیم. احتمال این کار همان منظور ما از توان است و در ناحیه سایه‌دار نشده توزیع H_1 نشان داده شده است.



شکل ۱۵-۱ توزیع نمونه‌گیری $\bar{X}$ تحت H_0 و H_1 .

۲-۱۵ عوامل مؤثر بر توان یک آزمون

بر اساس مطالب گذشته، توان تابعی از چند متغیر است:

۱. احتمال خطای نوع I (α)
 ۲. تفاوت واقعی بین فرض صفر و فرض مقابل ($\mu_0 - \mu_1$)
 ۳. اندازه نمونه (N) و σ^2
 ۴. آزمون خاصی که به کار گرفته می‌شود
- ما این رابطه آخر را تنها از جنبه توان نسبی نمونه‌های مستقل در برابر نمونه‌های وابسته (جور شده) مورد بحث قرار خواهیم داد. بطور کلی وقتی پیش‌فرض‌های مربوط به یک آزمون خاص برآورده شود، می‌توان نشان داد که شیوه‌های ارائه شده در این کتاب (همراه با استثنایی که در فصل ۲۰ آمده) توان بیشتری برای پاسخ به سؤال مطرح شده نسبت به سایر آزمون‌های در دسترس دارند.

۲-۱۵-۱ توان به عنوان تابعی از α

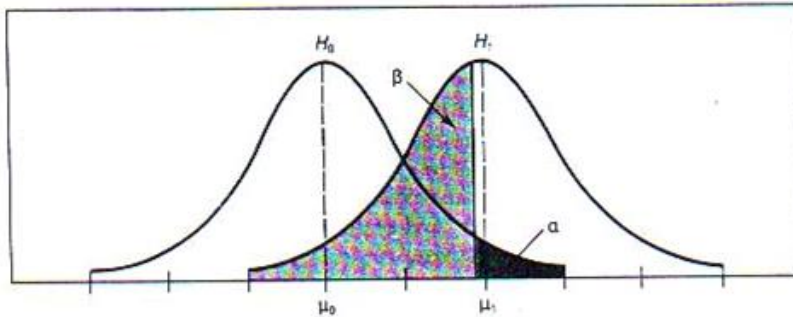
به کمک شکل ۱-۱۵ به آسانی می‌توان دید که ما چرا می‌گوییم توان تابعی از α است. اگر مایل به افزایش α هستیم، نقطه برش به سمت چپ حرکت می‌کند، از این رو همزمان β کاهش یافته و توان افزایش می‌یابد. متأسفانه این با افزایش در احتمال خطای نوع I همراه است.

۲-۱۵-۲ توان به عنوان تابعی از ($\mu_0 - \mu_1$)

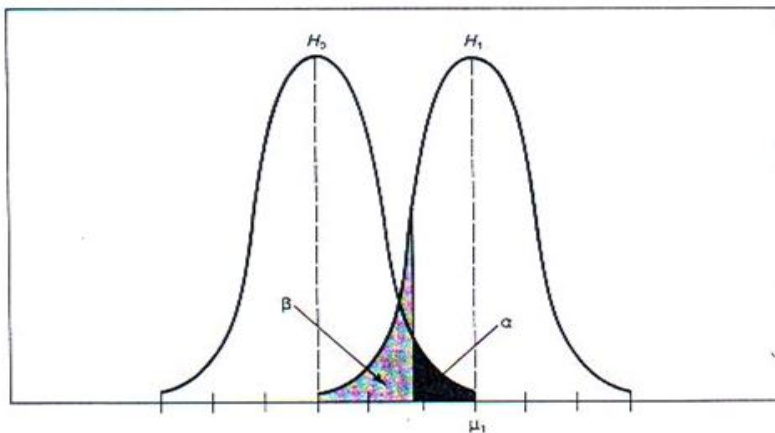
این واقعیت که توان تابعی از ماهیت فرض مخالف درست است (بطور دقیق‌تر ($\mu_0 - \mu_1$))، از مقایسه شکل ۱-۱۵ و ۲-۱۵ معلوم می‌شود. در شکل ۲-۱۵ فاصله بین μ_0 و μ_1 افزایش یافته، و در نتیجه توان هم افزایش یافته است. این جای تعجب خاصی نیست، زیرا همه می‌دانیم که شانس یافتن تفاوت بستگی به میزان بزرگی آن دارد. (تشخیص پرتقال از سیب آسان‌تر از پرتقال از نارنگی است.)

۲-۱۵-۳ توان به عنوان تابعی از اندازه نمونه (N) و σ^2

رابطه بین توان و اندازه نمونه (و بین توان و σ^2) قدری ظریف و موشکافانه است. از آنجا که ما به میانگین‌ها یا تفاوت بین میانگین‌ها توجه داریم، مستقیم یا غیرمستقیم به توزیع نمونه‌گیری میانگین‌ها توجه می‌کنیم. ما می‌دانیم که $\sigma_{\bar{X}}^2 = \sigma^2/N$ از این معادله می‌توان دریافت که واریانس توزیع نمونه‌گیری میانگین با افزایش N یا کاهش σ^2 ، کاهش می‌یابد. شکل ۳-۱۵ در مقایسه با شکل ۲-۱۵ نشان می‌دهد وقتی N را افزایش یا σ^2 را کاهش می‌دهیم، برای دو توزیع نمونه‌گیری (H_0 و H_1) چه چیزی رخ می‌دهد. در شکل ۳-۱۵ می‌بینیم که با کاهش $\sigma_{\bar{X}}^2$ ، همپوشی بین دو توزیع افزایش می‌یابد، که به افزایش توان منجر می‌شود.



شکل ۱۵-۲ تأثیرپذیری توان از اثر افزایش فاصله بین μ_0 و μ_1



شکل ۱۵-۳ تأثیرپذیری توان از اثر کاهش معیار میانگین

در صورتی که یک آزمایشگر به توان یک آزمون توجه داشته باشد، به احتمال زیاد به متغیرهای قابل دستکاری که بر توان حاکم هستند نیز توجه نشان می‌دهد. از آنجا که N را آسان‌تر از σ^2 یا $(\mu_0 - \mu_1)$ می‌توان تغییر داد و چون به هم زدن α اثرات جنبی نامساعدی دارد (افزایش احتمال خطای نوع I)، بحث روی توان عموماً متوجه اثرات تغییر اندازه نمونه می‌شود.

۱۵-۳ اندازه اثر

چنان‌که در شکل‌های ۱۵-۱ تا ۱۵-۳ مشاهده می‌شود، توان وابسته به درجه همپوشی بین توزیع‌های نمونه‌گیری تحت شرایط H_0 و H_1 است. از این گذشته، این همپوشی هم تابعی از فاصله بین μ_0 و μ_1 (میانگین جمعیت اگر H_0 درست باشد و میانگین جمعیت اگر H_1 درست باشد) و هم خطای معیار میانگین (انحراف معیار این توزیع‌های نمونه‌گیری) است. پس یک اندازه از میزانی که H_0 نادرست است

تفاوت میانگین‌های جمعیت تحت شرایط H_0 و H_1 ($\mu_1 - \mu_0$) خواهد بود که به صورت تعداد خطای معیار $(\mu_1 - \mu_0) / \sigma_{\bar{x}}$ بیان می‌شود. مشکلی که با این اندازه وجود دارد این است که مخرج کسر $(\sigma_{\bar{x}} = \sigma / \sqrt{N})$ شامل اندازه نمونه است. در عمل ما معمولاً مایلیم که N را از $(\mu_1 - \mu_0)$ و σ جدا کنیم تا بتوانیم آن را برای توان تداعی شده با یک مقدار N یا مقدار دیگر آن که برای یک سطح توان معین لازم است حل کنیم. از این رو ما اندازه فاصله یا **اندازه اثر**^۱ را بدست می‌آوریم،

$$\gamma = \frac{\mu_1 - \mu_0}{\sigma}$$

ما علامت γ (گاما) را نادیده می‌گیریم. گاما یک اندازه از درجه تفاوت μ_1 و μ_0 به زبان انحراف معیار جمعیت اصلی است. به عنوان مثال اگر ما انتظار داشته باشیم یک گروه از کودکان که از سوء تغذیه رنج می‌برند دارای میانگین IQ به اندازه ۸ نمره پایین‌تر از افراد عادی باشند (در حالی که $\sigma = 16$)، ما در باره اثری به اندازه‌ی نصف یک انحراف معیار صحبت می‌کنیم (یعنی گروه دچار سوء تغذیه، از حد متوسط به اندازه $1/2 = 8/16$ انحراف معیار پایین‌تر هستند). از این معادله درمی‌یابیم که γ بطور مستقل از N برآورد شده است، بطور ساده با برآورد μ_1 ، μ_0 و σ . ما N را هم بعداً به حساب خواهیم آورد.

۳-۱۵-۱ برآورد اندازه اثر

قدم اول برآورد γ می‌باشد، چون اساس محاسبات آینده را تشکیل می‌دهد. این کار را به یکی از سه روش زیر می‌توان انجام داد.

۱- تحقیق قبلی. ما اغلب می‌توانیم با نگاه به داده‌های گذشته، دست‌کم تقریب غیردقیقی از γ بدست آوریم. بنابراین می‌توانیم به میانگین‌ها و واریانس‌های مطالعات دیگر نگاه کنیم و حدس آگاهانه‌ای از اندازه‌های مورد انتظار برای $\mu_1 - \mu_0$ و σ صورت دهیم. در عمل این کار چندان که به نظر می‌آید مشکل نیست، بخصوص وقتی که شما آگاه باشید که یک تقریب هر چند ناشیانه بهتر از این است که هیچ تقریبی وجود نداشته باشد.

۲- اندازه‌گیری شخصی از اینکه چه تفاوتی مهم است. در بسیاری از موارد یک محقق می‌تواند بگوید، “من علاقه دارم که حداقل یک تفاوت ۱۰ درجه‌ای بین μ_1 و μ_0 تعیین کنم.” این محقق در اصل می‌گوید که تفاوت کمتر از این اهمیت یا فایده ندارد اما تفاوت بیشتر مهم است. در اینجا ما مقدار $\mu_1 - \mu_0$ را مستقیماً داده‌ایم، بدون هر گونه آگاهی لازم از مقادیر خاص μ_1 و μ_0 . آنچه باقی می‌ماند برآورد σ از روی داده‌های دیگر است. به عنوان مثال، ممکن است این محقق بگوید که علاقه‌مند است یک

راهنمای تحقیقاتی بیابد که نمرات امتحان پیشرفت تحصیلی دانشگاهی (SAT) را ۴۰ نمره بالاتر از حد معمول ببرد. ما قبلاً دیدیم که انحراف معیار این آزمون حدود ۱۰۰ است. بنابراین $0.40 = 40/100 = \gamma$. اگر محقق به جای این گفته که می‌خواهد نمرات ۴۰ نمره بالا بروند، بگوید که می‌خواهد آنها را $4/10$ یک انحراف معیار بالا ببرد، او مستقیماً γ را به ما داده است. در بعضی موارد امکان این گونه بیان وجود دارد.

۳- استفاده از قراردادهای خاص. وقتی ما به موقعیتی بر می‌خوریم که راهی برای برآورد پارامترهای جمعیتی مورد نیاز وجود ندارد، می‌توانیم به مجموعه قراردادهایی که کوهن (۱۹۷۷) وضع کرده رجوع کنیم. کوهن سه اندازه برای γ تعریف کرده است. قاعده سردستی کوهن چنین است:

اندازه	اثر
0.20	کوچک
0.50	متوسط
0.80	بزرگ

بنابراین وقتی راه‌های دیگر جواب ندهد، محقق می‌تواند تصمیم بگیرد که بدنبال یک اندازه اثر کوچک، متوسط یا بزرگ است، و γ را بدرستی تنظیم نماید. بایستی تأکید کنیم که این راه‌حل را فقط زمانی باید برگزید که روش‌های دیگر امکان‌پذیر نباشند.

ممکن است تصور کنید این عجیب است که بخواهید تفاوتی که بدنبالش هستید را قبل از انجام آزمایش تعریف کنید. بسیاری افراد در این مورد می‌گویند اگر آنها بدانند آزمایش به کجا می‌رسد دیگر نیازی به اجرای آن ندارند. گرچه آزمایشگران زیادی به این شیوه عمل می‌کنند، اما اگر با دقت به این بهانه توجه کنید، به اعتبار آن شک خواهید کرد. آیا واقعاً ما دست کم بطور مبهم هم نمی‌دانیم در آزمایش‌هایمان چه پیش خواهد آمد، و اگر چنین است چرا آنها را اجرا می‌کنیم؟ و حتی اگر ما نمی‌دانیم باید انتظار چه چیزی را داشته باشیم، حداقل باید کمترین اثری که مد نظرمان هست را بدانیم. درحالی که یک تداعی درست وجود دارد "من تعجب می‌کنم چه اتفاقی خواهد افتاد اگر" آزمایش، "من نمی‌دانم" به طور عادی تبدیل شود به "من تا به حال به این موضوع فکر نکرده‌ام".

۳-۱۵-۲ ترکیب کردن اندازه اثر و نمونه

در بحث اندازه اثر (γ) تصمیم‌گرفتنیم تا اندازه نمونه را از اندازه اثر جدا کنیم تا کار با N بطور جداگانه آسان‌تر شود. آخرین کاری که نیاز خواهیم داشت روشی برای ترکیب اندازه اثر با اندازه نمونه است که

توان یک آزمایش را برای یک N و اندازه اثر معین مشخص کنیم. برای این منظور ما نماد δ (دلتا) را به کار می‌بریم

$$\delta = \gamma[f(N)]$$

که تابع خاص N ، $f(N)$ ، برای هر آزمون منفرد بطور متفاوتی تعریف می‌شود. در این معادله $f(N)$ ، که "اف ان" خوانده می‌شود روشی است کلی برای بیان این که δ نه تنها به γ وابسته است بلکه همچنین از بعضی جنبه‌های غیرویزه به N نیز وابسته است. بطور مثال، ما خواهیم دید که در آزمون t تک نمونه‌ای δ را با جایگزینی $f(N)$ با $\sqrt{N}$ محاسبه می‌کنیم، در حالی که در آزمون t دو نمونه (N) را با $\sqrt{N/2}$ جایگزین می‌کنیم. چیزی که در این سیستم جالب است این است که به ما امکان می‌دهد از همان جدول δ برای محاسبه توان در تمام شیوه‌های آماری مورد نظر استفاده کنیم. نحوه استفاده از δ در بخش بعدی بیان خواهد شد.

لازم است دلیل تعریف کردن γ بدون توجه به N و سپس به کار گرفتن N در زمان تعریف δ را یادآور شویم. وقتی در حال طراحی یک آزمایش هستید، $(\mu_1 - \mu_0)$ و σ ، و بنابراین γ ، کم و بیش ثابت هستند. اما انتخاب N به شما بستگی دارد. ما می‌خواهیم بتوانیم بوسیله δ ، با داشتن γ ، وقتی مثلا N برابر ۲۰ و سپس برابر ۵۰ است توان را حساب کنیم. ما نمی‌خواهیم هر گاه N را تغییر می‌دهیم، مجبور به تکرار یک سری محاسبات خسته کننده شویم. با تعریف γ مستقل از N و سپس داشتن یک فرمول نمونه برای کنار هم گذاشتن آنها، در حجم زیادی از کارها صرفه‌جویی می‌کنیم.

۴-۱۵ محاسبات توان برای آزمون t تک نمونه‌ای

به عنوان اولین مثال، ما محاسبه توان برای یک آزمون t تک نمونه‌ای را بررسی می‌کنیم. در بخش قبلی دیدیم که δ براساس γ و بخشی تابع N است. برای t تک نمونه‌ای آن تابع شامل $\sqrt{N}$ خواهد بود، و آنگاه δ به صورت زیر تعریف خواهد شد

$$\delta = \gamma\sqrt{N}$$

یک تحلیل جبری کوتاه نشان خواهد داد که δ اکنون تفاوت بین μ_1 و μ_0 تقسیم بر خطای معیار میانگین است (به یاد داشته باشید که γ برابر $\mu_1 - \mu_0$ تقسیم بر انحراف معیار است).

$$\delta = \gamma\sqrt{N} = \frac{\mu_1 - \mu_0}{\sigma} \sqrt{N} = \frac{\mu_1 - \mu_0}{\left(\frac{\sigma}{\sqrt{N}}\right)} = \frac{\mu_1 - \mu_0}{\sigma_{\bar{x}}}$$

به طور مثال فرض کنید یک روان‌شناس بالینی می‌خواهد این فرض را بیازماید که افرادی که در پی درمان مشکلات روان‌شناختی هستند دارای IQ بالاتر از نرمال هستند. او می‌خواهد از IQهای ۲۵ مراجعه

کننده که بطور تصادفی از مجموعه‌های درمانگاهی زیادی انتخاب شده‌اند استفاده کند و مایل است با توان تشخیصی ۵ درجه بین میانگین جمعیت عادی و میانگین جمعیتی که نمونه‌اش را از آن بیرون می‌کشد تمایز برقرار سازد. بنابراین $\mu_1 = 105$ ، $\mu_0 = 100$ ، و $\sigma = 15$ است- دو پارامتر آخر میانگین و انحراف معیار بیشتر آزمون‌های IQ هستند وقتی که روی جمعیت کلی اجرا شوند.

$$\gamma = \frac{105-100}{15} = 0.33$$

پس

$$\delta = \gamma\sqrt{N} = 0.33\sqrt{25} = 0.33(5) = 1.65$$

گرچه آزمایشگر انتظار دارد میانگین نمونه بالاتر از میانگین عمومی جمعیت باشد، او استفاده از یک آزمون دو دانه در سطح $\alpha = 0.05$ را طرح کرده تا در برابر عوامل غیرمنتظره محافظت شود. با داشتن δ ، ما می‌توانیم بی‌درنگ توان آزمون را از جدول توان در بخش پیوست کتاب تعیین کنیم. یک بخش از این جدول در جدول ۱-۱۵ آمده است. برای استفاده از جدول یاد شده بسادگی در حاشیه سمت چپ جدول به پایین حرکت کرده تا به $\delta = 1.65$ برسیم و سپس نقطه تلاقی این ردیف را با ستون دارای سرعنوان 0.05 می‌خوانیم. البته در این جدول برای $\delta = 1.65$ سطحی معین نشده، بلکه برای دو حد $\delta = 1.60$ و $\delta = 1.70$ رقم‌هایی ذکر شده است. بدین ترتیب برای $\alpha = 0.05$ ، توان بین دو مقدار 0.36 و 0.40 است. با پیدا کردن حد میانی، خواهیم دید که توان برابر 0.38 است. به این معنی که اگر H_0 واقعا نادرست و μ_1 برابر 105 باشد، تنها در 3% اوقات این درمانگر داده‌هایی بدست خواهد آورد که هنگام آزمون تفاوت بین میانگین نمونه و آنچه با H_0 مشخص شده، یک مقدار t معنادار خواهد بود. این یک نتیجه دلسرد کننده است زیرا می‌گوید اگر میانگین واقعی 105 باشد، در $62\% = 38\% - 100\%$ اوقات این تحقیق به شکلی که طرح شده، نتیجه معناداری بدست نخواهد آورد.

جدول ۱-۱۵ نسخه کوتاهی از جدول پیوست کتاب، توان تابعی از δ و سطح معناداری (α)

آلفا برای آزمون دو دامنه				
δ	10 .	05 .	02 .	01 .
1.00	0.26	0.17	0.09	0.06
1.10	0.29	0.20	0.11	0.07
1.20	0.33	0.22	0.13	0.08
1.30	0.37	0.26	0.15	0.10
1.40	0.40	0.29	0.15	0.12
1.50	0.44	0.32	0.20	0.14
1.60	0.48	0.36	0.23	0.17
1.70	۵۲ .۰	۴۰ .۰	۲۷ .۰	۱۹ .۰
۸۰ .۱	۵۶ .۰	۴۴ .۰	۳۰ .۰	۲۲ .۰
1.90	0.60	0.48	0.34	0.25
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

از آنجا که آزمایشگر قبل از شروع آزمایش نسبت به بررسی مسأله توان هشیار بوده است، هنوز او این شانس را دارد که تغییراتی ایجاد کند تا توان را افزایش دهد. برای مثال او می‌تواند α را در سطح ۰/۱۰ تنظیم کند، تا در نتیجه توان به حدود ۰/۵۰ افزایش یابد، اما این کار احتمالاً رضایت بخش نیست. (مثلاً سردبیران مجلات علمی نمی‌پذیرند α بیشتر از ۰/۰۵ باشد.) یا به گونه‌ای دیگر او می‌تواند با افزایش N ، توان را افزایش دهد.

- برآورد اندازه‌ی نمونه مورد نیاز

اینکه گفته شود افزایش توان با افزایش N امکان‌پذیر است خوب است، اما حجم نمونه باید چقدر باشد؟ پاسخ این سؤال بستگی به سطح توان مطلوب دارد. فرض کنید مایلید در مثال قبلی توان را به ۰/۸۰ برسانید. نخستین کاری که باید انجام شود این است که با مراجعه به جدول توان، مقداری برای δ بیابیم که با مقدار توان مورد نظر هماهنگ باشد. در جدول مشاهده می‌کنیم که برای توان ۰/۸۰ بایستی δ برابر ۲/۸۰ باشد. بنابراین ما اندازه δ را داریم و بطور ساده می‌توانیم از فرمول دلتا، N را استخراج کرده و با دستکاری توان، حجم نمونه متناسب با آن را بدست آوریم.

$$\delta = \gamma \sqrt{N}$$

$$N = \left(\frac{\delta}{\gamma} \right)^2$$

$$= \left(\frac{۲/۸۰}{۰/۳۳} \right)^2 = ۸/۴۰^۲ = ۷۰/۵۶$$

چون سرو کار ما با تعداد افراد است که اعشار نمی پذیرد این عدد را به ۷۱ سر راست می کنیم. بنابراین اگر آزمایشگر بخواهد یک شانس ۸۰٪ برای رد H_0 داشته باشد، زمانی که $\gamma = 0/33$ (یعنی $\mu_1 = 105$) یا ۹۵ (است)، بایستی از ۷۱ نفر افراد تصادفی انتخاب شده تست IQ بگیرد. هر چند او ممکن است احساس کند این تعداد افراد مراجعه کننده خیلی زیادتر از آن چیزی است که او بتواند به راحتی تست کند، اما راه دیگری برای حل و فصل یک توان کم وجود ندارد.

ممکن است تعجب کنید که چرا ما در مثال قبل توان را معادل ۸۰٪ گرفتیم. به خاطر داشته باشید که با این اندازه توان، ما هنوز یک شانس ۲۰٪ برای خطای نوع II داریم. پاسخ به مشکلات عملی بر می گردد. فرض کنید مثلاً آزمایشگر می خواست توان را به ۹۵٪ برساند. اندکی محاسبات ساده نشان خواهد داد که این کار به نمونه‌ای به حجم ۱۱۹ نفر نیاز دارد، و برای توانی معادل ۹۹٪ تقریباً ۱۶۷ آزمودنی لازم است. این اندازه نمونه ممکن است در شرایط عملی یا به دلایل هزینه‌ای برای آزمایشگر قابل توجیه نباشد. در حالی که افزایش توان عموماً با افزایش نمونه بدست می آید، برای سطوح بالای توان، هزینه ممکن است خیلی زیاد باشد. علاوه بر این، این موضوع می تواند حالتی از یک نقصان بازگشت باشد چون افزایش δ تابع مجذور ریشه N است. وقتی شما از اطلاعات آماده‌ای که از قبل توسط یک موسسه جمع آوری شده استفاده می کنید وضعیت فرق دارد با وقتی که مثلاً با دوقلوهای یکسان جدا از هم پرورش یافته سروکار دارید.

۵-۱۵ محاسبات توان برای تفاوت بین میانگین‌های مستقل

عملکرد توان در موقعیتی که می‌خواهیم دو میانگین مستقل را بیازماییم خیلی شبیه به موقعیتی است که تنها یک میانگین داشتیم. در قسمت پیش ما γ را از طریق تفاوت بین μ تحت H_1 (یعنی μ_1) و μ تحت H_0 (یعنی μ_0) و تقسیم آن بر σ بدست آوردیم. در حالت فعلی ما کار مشابهی انجام خواهیم داد، هر چند اکنون ما به سمت کار با تفاوت بین میانگین‌ها پیش می‌رویم. بنابراین ما تفاوت بین دو میانگین جمعیت $(\mu_1 - \mu_2)$ تحت H_1 منهای تفاوت $(\mu_1 - \mu_2)$ تحت H_0 مجدداً تقسیم بر σ را می‌خواهیم. (بخاطر دارید که فرض کردیم $\sigma_1^2 = \sigma_2^2 = \sigma^2$) اما $(\mu_1 - \mu_2)$ تحت H_0 در تمام حالات معمولی صفر است، طوری که ما می‌توانیم این عبارت را از فرمول حذف کنیم. بنابراین

$$\gamma = \frac{(\mu_1 - \mu_2) - 0}{\sigma} = \frac{\mu_1 - \mu_2}{\sigma}$$

صورت کسر اشاره به تفاوت بین میانگین‌های مورد انتظار جمعیت تحت H_1 دارد و مخرج کسر نشان دهنده انحراف معیار مشترک دو جمعیت است.

۵-۱۵-۱ اندازه‌های نمونه برابر

به منظور طرح یک مثال، فرض کنید می‌خواهیم تفاوت بین میزان رفتار احتکارگری موش‌های عادی و آنهایی که در نوزادی دچار محرومیت بوده‌اند را بررسی کنیم. آزمایش مشابهی توسط هانت (۱۹۴۱) صورت گرفته که می‌گوید میانگین تعداد احتکارهای گروه محرومیت کشیده تقریباً ۳۵ و برای گروه غیرمحروم تقریباً ۱۵ خواهد بود. علاوه بر این داده‌های هانت گویای یک مقدار ۱۷ برای σ است. بنابراین حداقل به عنوان یک تقریب نه چندان دقیق، ما انتظار داریم $\mu_1 = 35$ ، $\mu_2 = 15$ و $\sigma = 17$ باشد. پس

$$\gamma = \frac{\mu_1 - \mu_2}{\sigma} = \frac{20}{17} = 1/18$$

ما می‌گوییم که انتظار تفاوت ۱/۱۸ انحراف معیار بین دو میانگین را داریم. ابتدا توان یک آزمایش با ۱۰ مشاهده در هر گروه را بررسی می‌کنیم. ما δ را در حالت دو-نمونه به این صورت تعریف می‌کنیم

$$\delta = \gamma \sqrt{\frac{N}{2}}$$

که N تعداد موردها در هر نمونه است (در کل $2N$ مورد وجود دارد). بنابراین

$$\begin{aligned} \delta &= \gamma \sqrt{\frac{N}{2}} = (1.18) \sqrt{\frac{10}{2}} \\ &= (1.18) \sqrt{5} = 1/18 (2/236) = 2/63 \end{aligned}$$

در جدول توان در بخش پیوست کتاب می‌بینیم که برای $\delta = 2/63$ با آزمون دو دامنه در سطح $\alpha = 0.05$ ، توان برابر ۰/۷۵ است. بنابراین اگر واقعاً این آزمایش را با ۱۰ آزمودنی در هر گروه اجرا کنیم و اگر برآورد δ درست باشد، آنگاه ما یک شانس ۷۵٪ داریم که در عمل H_0 را رد کنیم. این درجه بالای توان برای این تعداد کم آزمودنی است، ولی درعین حال ما با یک اثر واقعاً بزرگ سروکار داریم. سپس ما مایلیم پرسش را برگردانیم و بپرسیم که چه تعداد آزمودنی برای توان معادل ۰/۹۰ نیاز است. (منطقی است که در اینجا توان را تا ۰/۹۰ تقویت کنیم [β را به ۰/۱۰ کاهش دهیم] می‌دانیم که با تعداد کم ۱۰ آزمودنی در گروه، توانی معادل ۰/۷۵ با اندازه اثری که داریم به ما می‌دهد.) در جدول مشاهده می‌کنیم که برای این کار باید δ برابر ۳/۲۵ باشد.

$$\delta = \gamma \sqrt{\frac{N}{2}}$$

با مجذور کردن آن برای بازسازی عبارت‌ها، داریم

$$\delta^2 = \frac{\gamma^2 N}{2}$$

$$N = \frac{2\delta^2}{\gamma} = \frac{2(3.25)^2}{1.18^2} = \frac{2(10.5625)}{1.3924} = 15/17 \quad \text{پس}$$

چون N نشانگر تعداد آزمودنی‌ها در نمونه است، ما به ۱۵ آزمودنی در نمونه و در کل به ۳۰ آزمودنی نیاز داریم تا توان به ۰/۹۰ برسد. این اندازه نسبتاً بزرگی برای تحقیق با موش‌های آزمایشگاهی است - که بایستی با هزینه زیاد خریداری، نگهداری و تغذیه شوند. اینکه این ارزش دارد یا نه، بستگی به اهمیت تحقیق دارد.

۵-۱۵-۲ اندازه‌های نمونه نابرابر

در بخش پیش با نمونه‌هایی سروکار داشتیم که برابر بودند. اما ما آزمایشاتی را هم انجام می‌دهیم که بخاطر ماهیت آزمایش، دو نمونه دارای اندازه نابرابر هستند. این موضوع بخصوص وقتی که سعی می‌کنیم δ را حل کنیم مشکلاتی ایجاد می‌کند، چون برای وارد کردن در فرمول به یک مقدار برای N نیاز داریم. چه مقداری را می‌توانیم به کار ببریم؟

ساده‌ترین راه حل این است که به جای N میانگین هارمونیک (هم‌ساز) اندازه دو نمونه را بگذاریم. بطور کلی میانگین هارمونیک k عدد $(X_1, X_2, \dots, X_k)$ به این صورت تعریف می‌شود

$$\bar{X}_h = \frac{k}{\sum \frac{1}{X_i}}$$

مثلاً، برای اعداد ۸، ۱۲، و ۱۳ میانگین هارمونیک می‌شود

$$\bar{X}_h = \frac{3}{\frac{1}{8} + \frac{1}{12} + \frac{1}{13}} = 10/52$$

برای حالت دو اندازه نمونه (N_1 و N_2) فرمول ساده می‌شود به

$$\bar{N}_h = \frac{2}{\frac{1}{N_1} + \frac{1}{N_2}}$$

با ضرب کردن صورت و مخرج در $N_1 N_2$ یک فرمول ساده‌تر بدست می‌آید:

$$\bar{N}_h = \frac{2N_1N_2}{N_1 + N_2}$$

سپس ما می‌توانیم در محاسبه δ $\bar{N}_h$ را به جای N به کار ببریم. توجه داشته باشید که میانگین هارمونیک با عباراتی به شکل $1/N$ سروکار دارد. چون خطای معیار میانگین نیز تابع $1/N$ تغییر می‌کند، ما در محاسبه δ از میانگین هارمونیک به جای میانگین حسابی معمولی استفاده می‌کنیم.

از قرار دادن مقادیر مختلف N در فرمول قبلی چیز جالبی بدست می‌آید. برای یک تعداد کل ثابت ما بیشترین میزان توان را زمانی خواهیم داشت که آزمودنی‌ها بطور یکسان بین دو گروه تقسیم شوند. وقتی اندازه‌های نمونه نابرابر باشند و بتوان آزمودنی‌های بیشتری به تحقیق اضافه کرد، آنها را طوری اضافه کنید که توازن گروه‌ها حفظ شود.

۶-۱۵ محاسبه توان برای آزمون t نمونه‌های وابسته

زمانی که ما با شرایطی روبرو هستیم که می‌خواهیم تفاوت بین دو نمونه جور شده را بیازماییم، مسأله قدری مشکل‌تر می‌شود و یک پارامتر دیگر بایستی در نظر گرفته شود. به این دلیل تحلیل توان برای این مورد خیلی غیرعملی است. اما راه‌حل کلی برای این مسأله یک اصل مهم طرح آزمایشی را نشان می‌دهد و از این‌رو بررسی دقیق حالت نمونه وابسته را توجیه می‌کند. (این بخش به گونه‌ای نوشته نشده که هر کسی محاسبات توان را برای آزمایش مربوط به نمونه وابسته اجرا کند. محاسبه مهم نیست بلکه نتیجه‌گیری‌های بعدی مهم است.)

می‌خواهیم γ را به صورت زیر تعریف کنیم

$$\gamma = \frac{\mu_1 - \mu_2}{\sigma_D}$$

که $\mu_1 - \mu_2$ نشان‌دهنده میانگین تفاوت مورد انتظار (میانگین مورد انتظار تفاوت نمرات) است. مشکل از این واقعیت ناشی می‌شود که σ_D انحراف معیار جمعیت‌های X_1 و X_2 نیست، بلکه انحراف معیار نمرات اختلاف برآمده از این جمعیت‌ها است. گرچه ممکن است ما بتوانیم یک حدس خردمندانه از σ_1 و σ_2 به عمل آوریم، ما احتمالاً در باره σ_D ایده‌ای نداریم.

اگر بخواهیم چنین فرض کنیم که دو مجموعه نمرات دارای انحراف معیار جمعیتی یکسانی هستند، ($\sigma_1 = \sigma_2 = \sigma$) آنگاه به آسانی می‌توان نشان داد که

$$\sigma_D = \sigma\sqrt{2(1-\rho)}$$

در اینجا ρ همبستگی بین X_1 و X_2 در جمعیت است و مقادیر آن می‌تواند بین $+1$ و -1 باشد، مثبت بودن برای تقریباً تمام موقعیت‌هایی که ما بخواهیم یک t همبسته داشته باشیم. در این فرمول، σ انحراف

معیار نمرات جمعیت برای X_1 یا X_2 است (زیرا $\sigma_1 = \sigma_2 = \sigma$)

برای لحظه‌ای فرض کنید ما می‌خواهیم ρ را برآورد کنیم، از اینجا روند کار همانند حالت t تک نمونه‌ای است. ابتدا

$$\gamma = \frac{\mu_1 - \mu_2}{\sigma_D} = \frac{\mu_1 - \mu_2}{\sigma\sqrt{2(1-\rho)}}$$

$$\delta = \gamma\sqrt{N}$$

و

تعریف می‌کنیم، آنگاه به مقدار δ جدول رجوع می‌کنیم.

به طور مثال فرض کنید می‌خواهید ببینید آیا عملکرد شناختی کودکان خردسال در ابتدای روز بهتر از شب است. در پی آن ما می‌خواهیم یک آزمون پیشرفت استاندارد را روی نمونه‌ای متشکل از ۲۰ کودک

درست قبل از اینکه بخواهند اجرا کنیم، و دوباره صبح روز بعد هم آن را اجرا کنیم. (در اینجا ما از بعضی مسائل چشم‌پوشی می‌کنیم، مثل اثرات عمل، و اینکه این مطالعه بهتر است با دو گروه کودکان انجام شود، که نیمی را ابتدا در صبح و نیمی را ابتدا در شب بیازمائیم.) فرض کنید ما می‌خواهیم توان را برای یافتن یک تفاوت سه برابری بین میانگین‌ها ارزیابی کنیم. اطلاعات مربوط به آزمون‌های استاندارد را در بسیاری کتاب‌ها می‌توان یافت. از این منابع می‌شود فهمید که انحراف معیار نمرات آزمون ۱۰ است. همبستگی بین دو بار اجرای آزمون موجود نیست اما پایایی کوتاه مدت آزمون موجود است و مقدار آن ۰/۹۲ است. بنابراین

$$\begin{aligned}\sigma\sqrt{2(1-\rho)} &= 10\sqrt{2(1-0.92)} = 10\sqrt{2(0.08)} = 4.0 \\ \gamma &= \frac{\mu_1 - \mu_2}{\sigma_D} \\ &= \frac{3}{4.0} = 0.75 \\ \delta &= \gamma\sqrt{N} \\ 3/35 &= 0.75\sqrt{20} =\end{aligned}$$

توان = ۰/۹۲ برای ۰/۰۵ α .

از سوی دیگر فرض کنید ما از یک آزمون با پایایی کمتر استفاده کرده باشیم که برای آن $\rho = 0.50$ باشد. فرض می‌کنیم که σ بدون تغییر بماند. آنگاه

$$\begin{aligned}\sigma\sqrt{2(1-0.50)} &= 10\sqrt{2(1-0.50)} = 10\sqrt{1} = 10 \\ \gamma &= \frac{\mu_1 - \mu_2}{\sigma_D} \\ &= 0.75 \cdot \frac{3}{10} = 0.225 \\ \sigma &= 0.30\sqrt{20} \\ &= 1.34\end{aligned}$$

توان = ۰/۲۷ برای ۰/۰۵ α .

در اینجا مشاهده می‌کنید که با کاهش ρ ، توان هم افت می‌کند. وقتی $\rho = 0$ ، دو نمونه (بطور خطی) مستقل هستند، و بنابر این حالت نمونه‌ی وابسته به حالت نمونه‌ی مستقل تنزل پیدا می‌کند به استثنای اینکه تفاوت درجات آزادی، که در رابطه با تقریب به کار رفته در جدول توان (پیوست کتاب) ابهام‌آمیز خواهد بود. نکته مهمی که در اینجا منظور شده (و دلیل پشت همه محاسبات) این است که برای مقاصد

عملی همواره کمترین توان در مورد نمونه‌ی وابسته رخ می‌دهد زمانی که $\rho = 0$ و نمونه‌های ما مستقل هستند. بنابراین برای تمام موقعیت‌هایی که حتی ما از دور مایل به استفاده از نمونه‌های وابسته هستیم (یعنی وقتی ما انتظار یک همبستگی مثبت بین X_1 و X_2 داشته باشیم)، طرح نمونه‌ی وابسته قدرت‌مندتر از طرح گروه‌های مستقل است. این موضوع یکی از مزایای اصلی طرح‌هایی که با نمونه‌های وابسته اجرا می‌شوند را نشان می‌دهد.

۷-۱۵ ملاحظات توان از لحاظ اندازه نمونه

بحث حاضر در باره توان نشان می‌دهد چنانچه بخواهید آزمایشاتی انجام دهید که شانس خوبی برای رد H_0 نادرست دارند - بویژه اگر اثر کوچک باشد، نمونه‌های بزرگ قابل توجه تقریباً یک ضرورت است. اندکی محاسبات نشان خواهد داد که اگر بخواهیم توان برابر $0/80$ باشد و اگر تعریف کوهن را از اثرات کوچک، متوسط و بزرگ بپذیریم، نمونه‌های ما باید کاملاً بزرگ باشند. جدول ۲-۱۵ تعداد کل N مورد نیاز (برای توان $= 0/80$ ، $\alpha = 0/05$) را برای اثرات کوچک، متوسط و بزرگ در آزمون‌هایی که مورد بحث قرار دادیم نشان می‌دهد. این اطلاعات گویای آن است که توان (دست‌کم یک مقدار قابل توجه از آن) یک کالای بسیار پرهزینه است - بخصوص برای اثرات کوچک. در حالی که می‌توان گفت این چیز خوبی است، چون در غیر این صورت ادبیات تحقیق پر از مطالب کم ارزش می‌بود، اما برای آزمایشگران چندان راحت نیست. قاعده کلی این است که اثرات بزرگ جستجو شوند، از نمونه‌های بزرگ استفاده شود، یا طرح‌های آزمایشی حساس به کار برده شوند مانند آنهایی که از اندازه‌گیری‌های مکرر استفاده می‌کنند، که خطاهای آزمایشی کاهش یافته و در نتیجه تفاوت‌های کوچک، به اندازه‌ی اثر بزرگ تبدیل گردد.

جدول ۲-۱۵ کل اندازه نمونه مورد نیاز برای توان $= 0/80$ ، $\alpha = 0/05$ در آزمون دو دامنه

اندازه اثر	γ	t یک دامنه	t دو دامنه
کوچک	0/20	۱۹۶	۷۸۴
متوسط	0/50	۳۲	۱۲۶
بزرگ	0/80	۱۳	۴۹

۸-۱۵ خلاصه

در این فصل ما به عواملی پرداختیم که در توان آزمون نقش دارند. چنانکه دیدیم آسان‌ترین عامل برای دستکاری، اندازه نمونه است، و ما شیوه‌های محاسبه توان برای یک اندازه نمونه معین، و عکس آن، اندازه مورد نیاز برای یک سطح توان خاص را بیان کردیم. در آخر به این نتیجه کمتر مطلوب رسیدیم که

سطوح بالای توان اغلب ممکن است نیازمند آزمودنی‌هایی بیش از آنچه انتظار داریم باشند. برخی از مهمترین مطالب این فصل عبارتند از:

- توان
- دلتا (δ)
- اندازه اثر
- میانگین هارمونیک (هم‌ساز)
- گاما (γ)

۹-۱۵ تمرین‌های

۱-۱۵ طی ۱۰ سال گذشته یک دانشگاه اروپایی توانسته میانگین نمره پیشرفت تحصیلی (SAT) خود را (و در نتیجه، میانگین جمعیتی که دانشجویان خود را از آن استخراج کرده) در حد ۵۲۰ با انحراف معیار ۸۰ حفظ کند. یک رقیب عمده مایل است نشان دهد که معیارها لغزش داشته و این دانشگاه در حقیقت از جمعیتی با میانگین ۵۰۰ برمی‌آید. آنها در نظر دارند یک آزمون t روی میانگین نمرات SAT ورودی‌های ترم بعد اجرا کنند.

(الف) اندازه اثر مورد بحث چیست؟

(ب) اگر تعداد ورودی‌های ترم بعد ۱۰۰ نفر باشد مقدار δ چقدر می‌شود؟

(ج) توان آزمون چقدر است؟

۲-۱۵ وضعیت مورد نظر در تمرین ۱-۱۵ را در راستای شکل ۱-۱۵ رسم کنید.

۳-۱۵ در تمرین ۱-۱۵ چه اندازه نمونه به ترتیب برای رسیدن به توان ۰/۷۰، ۰/۸۰ و ۰/۹۰ لازم است؟

۴-۱۵ دانشگاه اول در تمرین ۱-۱۵ بدون اطلاع از رقابت، کمپینی را راه انداخته تا کیفیت دانشجویان تازه وارد را ارتقا بخشد. این دانشگاه امیدوار است یک افزایش ۳۰ نمره‌ای در میانگین نمره SAT برای سال آینده حاصل کند. اگر ۱۰۰ دانشجو در ترم بعد ثبت‌نام کنند، توان یک آزمون t برای بررسی معناداری هر گونه افزایش چقدر است؟

۵-۱۵ وضعیت مورد نظر در تمرین ۴-۱۵ را در راستای شکل ۱-۱۵ رسم کنید.

۶-۱۵ یک آزمایشگاه روان‌شناسی رفتار اجتنابی را در خرگوش‌ها برای مدت چند سال مورد مطالعه قرار داده و مقاله‌های متعددی منتشر کرده است. از این تحقیقات روشن شده که میانگین تأخیر پاسخ برای یک تکلیف خاص ۵/۸ ثانیه با انحراف معیار ۲ ثانیه است (براساس صدها خرگوش). اینک محققان مایلند در نواحی خاصی از آمیگدال مغز، زخمی ایجاد کنند و شرطی‌شدن اجتنابی ضعیف‌تر این حیوانات را نشان دهند. آنها انتظار دارند تأخیر به ۱ ثانیه کاهش یابد (یعنی خرگوش پاسخ تنبیه را زودتر خواهد داد)، و در نظر دارند یک آزمون t تک نمونه‌ای اجرا کنند (با $H_0: \mu_0 = 5/8$).

- (الف) چه تعداد آزمودنی نیاز دارند تا دست کم شانس موفقیت ۵۰:۵۰ داشته باشند؟
- (ب) چه تعداد آزمودنی نیاز دارند تا دست کم شانس موفقیت ۸۰:۲۰ داشته باشند؟
- ۱۵-۷ فرض کنید در آزمایش تمرین ۶-۱۵ تصمیم گرفته شود نه یک گروه و مقایسه آن با $\mu_0 = 5/8$ ، بلکه دو گروه به کار گرفته شود (یکی با ایجاد جراحت و دیگری بدون آن). به هر حال، آنها هنوز انتظار میزان تغییر مشابهی را دارند.
- (الف) اگر آنها بخواهند توان ۶۰/۰ داشته باشند چه تعداد آزمودنی (در کل) نیاز دارند؟
- (ب) اگر آنها بخواهند توان ۹۰/۰ داشته باشند چه تعداد آزمودنی (در کل) نیاز دارند؟
- ۸-۱۵ چنانکه دیدیم یک محقق در تمرین ۷-۱۵ بدون زحمت محاسبه توان تحقیق را صورت داد. او سعی کرد تا ۲۰ آزمودنی را در هر گروه قرار دهد، اما بطور اتفاقی اشکالی در کار او پیش آمد و ناچار شد ۵ موش را از گروه آزمایشی حذف کند. توان این آزمایش چقدر خواهد بود؟
- ۹-۱۵ در تحقیقی رشد شناختی نوزادان کم وزن و عادی در سال اول تولد مورد مقایسه قرار گرفته است. محقق با یک مقیاس خود ساخته دریافت که میانگین‌های دو گروه نمونه به ترتیب ۲۵ و ۳۰ بودند. اگر فرض کنیم میانگین‌ها و انحراف معیارهای واقعی به طور دقیق برآورد شده باشند، احتمال از قبل تعیین شده (قبل از اجرای آزمایش) که این مطالعه به یک تفاوت معنادار برسد چیست؟
- ۱۰-۱۵ ما تمرین ۹-۱۵ را به میانگین ۲۵ و ۲۸، با یک انحراف معیار جمعی ۸ و تعداد نمونه ۲۰ و ۲۰ تغییر خواهیم داد.
- (الف) یک توان از قبل این آزمایش چقدر خواهد بود؟
- (ب) آزمون t را روی داده‌ها اجرا کنید.
- (ج) پاسخ (الف) در باره پاسخ (ب) چه چیز را بیان می‌کند (اگر چیزی را بیان می‌کند)؟
- ۱۱-۱۵ دو دانشجوی فوق لیسانس اخیراً پایان‌نامه خود را تکمیل کرده‌اند. هر کدام یک آزمون t برای دو گروه مستقل اجرا کرده‌اند. یکی از آنها به یک تفاوت به سختی معنادار با ۱۰ آزمودنی در هر گروه دست یافت. دیگری یک تفاوت به سختی معنادار با ۴۵ آزمودنی در هر گروه پیدا کرد. کدام تحقیق شما را بیشتر تحت تأثیر قرار می‌دهد؟
- ۱۲-۱۵ نمودار (مشابه شکل ۱-۱۵) برای جواب خود به تمرین ۱۱-۱۵ را رسم کنید.
- ۱۳-۱۵ یک مثال ساده دو گروهی طرح کنید که نشان دهد وقتی آزمودنی‌ها در کل ۳۰ مورد باشند با افزایش اندازه‌ی نمونه‌ها توان تغییر زیادی نمی‌کند.
- ۱۴-۱۵ یک جدول شبیه به جدول ۲-۱۵ برای توان برابر با ۸۰/۰ با α برابر با ۰/۰۱، دو دامنه طرح کنید.
- ۱۵-۱۵ یک جدول شبیه به جدول ۲-۱۵ برای توان برابر با ۶۰/۰ با α برابر با ۰/۰۵، دو دامنه طرح کنید.

۱۵-۱۶ فرض کنید ما می‌خواهیم یک فرض صفر را پیرامون یک میانگین منفرد در سطح $\alpha = 0.05$ ، یک دامنه بی‌ازمائییم. علاوه بر این فرض کنید تمام پیش‌فرض‌های لازم برآورده شده است. آیا موردی وجود دارد که ما تمایل بیشتری به رد H_0 درست داشته باشیم نسبت به حالتی که بخواهیم H_0 را وقتی نادرست است رد کنیم؟ (به عبارت دیگر آیا توان هرگز می‌تواند کمتر از α باشد؟)

۱۵-۱۷ اگر $\sigma = 15$ ، $N = 25$ باشد و ما بخواهیم $\mu: H_0 = 100$ در برابر $\mu: H_1 > 100$ آزمایش کنیم، مقدار میانگین تحت H_1 چقدر خواهد بود در صورتی که توان برابر با احتمال یک خطای نوع II باشد؟ (راهنمایی: بهترین راه برای حل این مسأله طرح دو توزیع است. کدام ناحیه را می‌خواهید یکسان فرض کنید؟)

فصل ۱۶

تحلیل واریانس یک طرفه

تحلیل واریانس (آنوا)^۱ در حال حاضر یکی از پرکاربردترین شیوه‌های آماری است. شهرت و فایده این شیوه را می‌توان به دو واقعیت نسبت داد. اول از همه تحلیل واریانس مانند t ، با تفاوت میانگین نمونه‌ها سروکار دارد، اما برخلاف t از لحاظ تعداد میانگین‌ها محدودیتی ندارد. به جای طرح پرسش تنها از تفاوت دو میانگین، می‌توان از تفاوت ۳، ۴، ۵، یا k میانگین پرسش به میان آورد. دوم اینکه تحلیل واریانس به ما امکان می‌دهد که همزمان با دو متغیر مستقل یا بیشتر کار کنیم، و نه تنها اثرات منفرد هر متغیر به طور جداگانه، بلکه اثرات تعاملی آنها را هم بررسی کنیم.

این فصل به منطق زیربنایی تحلیل واریانس (که در واقع کاملاً ساده است) و تحلیل نتایج آزمایش‌هایی که در آنها تنها یک متغیر مستقل به کار گرفته می‌شود اختصاص دارد. علاوه بر این ما با چند موضوع مربوط دیگر سرو کار داریم که در زمینه یک تحلیل یک متغیره (آنوای یک طرفه^۲) بسادگی قابل فهم هستند. فصل‌های بعدی به تحلیل آزمایشاتی می‌پردازند که شامل دو متغیر یا بیشتر هستند و شامل طرح‌هایی است که در آنها سنجش‌های مکرر روی هر آزمودنی به عمل می‌آید.

۱۶-۱ یک مطالعه نمونه‌ای فرضی

با یک مثال فرضی سعی می‌کنیم تحلیل واریانس را به طور ساده توضیح دهیم. ما می‌خواهیم رفتار کمک‌رسانی مستقیم به افراد سنین مختلف را بررسی کنیم. برای این کار ما از آزمودنی‌هایی که ۲۰، ۴۰ یا ۶۰ سال دارند استفاده می‌کنیم و به آنها محموله‌ای همراه با یک بچه در حال گریه کردن داده می‌شود تا آن را از یک مسیر فرعی در یک ساعت شلوغ حمل کنند، و مدت زمانی که طول می‌کشد تا کسی به

1. analysis of variance (ANOVA)
2. one-way ANOVA

آنها جایی برای نشستن بدهد را ثبت می‌کنیم. فرض آزمایشی ما این است که به آزمودنی‌های مسن‌تر زودتر از افراد جوان‌تر صندلی داده می‌شود.

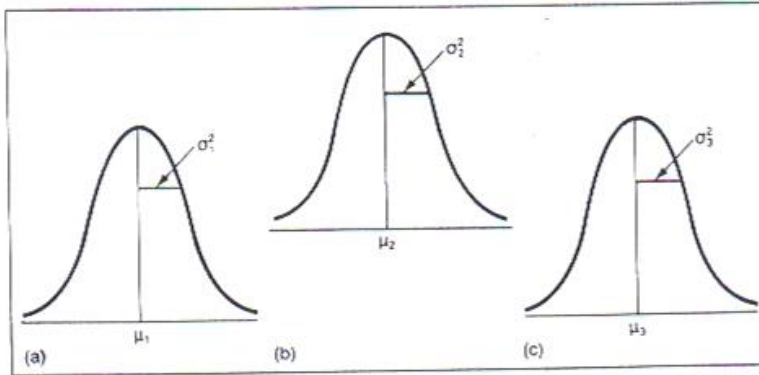
چون این یک مطالعه فرضی است، ما همه راه‌ها را بررسی می‌کنیم و فرض می‌گیریم که می‌توانیم اندازه نمونه‌های نامحدودی را بدست آوریم. به بیان دیگر فرض کنید ما تمام افراد ۲۰ ساله را در کشور به کار گرفته‌ایم و مدت زمان قبل از گرفتن صندلی را اندازه‌گیری کرده‌ایم. نتایج این بخش از آزمایش شامل توده‌ای از نقاط داده‌ها است، توزیع آنها در شکل ۱-۱۶ (الف) ارائه شده است. این توزیع در واقع جمعیت نمرات را برای همه ۲۰ ساله‌ها نشان می‌دهد و دارای یک میانگین و یک واریانس است که به ترتیب با μ_1 و σ_1^2 نشان داده شده‌اند. فرض کنید این توزیع کم و بیش نرمال باشد.

اکنون فرض کنید همه چیز را از نو با استفاده از همه افراد ۴۰ ساله کشور شروع کنیم. باز هم ما یک جمعیت کلی از نقاط داده‌ها داریم، که همانند شکل ۱-۱۶ (ب) توزیع شده‌اند. این توزیع نیز کم و بیش نرمال خواهد بود، و ما میانگین و واریانس آن را μ_2 و σ_2^2 می‌نامیم بدون اینکه در باره روابط بین μ_1 و μ_2 یا σ_1^2 و σ_2^2 فعلاً چیزی بگوییم.

در نهایت، فرض کنید که ما برای بار سوم باز آزمایش را با استفاده از تمام افراد ۶۰ ساله کشور تکرار کنیم. این داده‌ها در شکل ۱-۱۶ (ج) ترسیم شده‌اند. این توزیع نیز تقریباً نرمال خواهد بود، و دارای یک میانگین (μ_3) و واریانس (σ_3^2) می‌باشد.

تا اینجا کار ما تکمیل شده و با کمال اطمینان می‌توانیم روابط بین μ_1 ، μ_2 و μ_3 را بیان کنیم. واژه اطمینان دقیق است زیرا ما کل جمعیت را می‌سنجیم، ما μ_1 ، μ_2 و μ_3 را محاسبه می‌کنیم و مجبور به برآورد کردن نیستیم. فقط ما بایستی میانگین‌ها را پیدا کنیم تا ببینیم آیا تفاوت دارند یا نه.

چنان‌که در فصل‌های قبلی در باره آزمون t بیان کردیم، بدیهی است که ما نمی‌توانیم تحقیقات را روی کل جمعیت مورد نظر اجرا کنیم. در دنیای واقعی ما ناچاریم به نمونه‌هایی از داده‌ها بر اساس تعداد قابل قبولی از آزمودنی‌ها قناعت کنیم. پرسش اصلی یکسان است (”میانگین‌های جمعیت چه فرقی دارند، اگر فرقی داشته باشند؟“)، اما در این حالت ما باید بیشتر به سؤال مربوط به استفاده از برآوردهای μ_j بپردازیم نه خود مقادیر حقیقی μ_j . با در نظر گرفتن این آزمایش فرضی، اجازه دهید به بررسی مسأله با جزئیات بیشتر بپردازیم و سپس دو پیش‌فرض را به منظور سادگی اعمال کنیم.



شکل ۱۶-۱ نتایج آزمایش فرضی روی زمان بدست آوردن یک صندلی در طول یک راه

فرض صفر

چون ما می‌خواهیم بدانیم آیا میانگین زمان در اختیار گذاشتن یک صندلی برای هر سه گروه سنی یکسان است، در نتیجه فرض صفر به این صورت نوشته می‌شود

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu$$

که μ بدون اندیس نشان دهنده مقدار مشترک سه میانگین جمعیت است. فرض صفر به چند دلیل مختلف می‌تواند نادرست باشد، مثلاً چون $\mu_2 = \mu_1$ اما μ_3 فرق دارد، اما فعلاً تنها به این موضوع می‌پردازیم که یا کاملاً درست یا به هر دلیل نادرست است.

جمعیت

یکی از مشکلاتی که دانشجویان زیاد با آن برخورد دارند معنای واژه جمعیت است. چنان‌که در فصل ۱ ذکر شد، یک جمعیت مجموعه‌ای از اعداد است نه موش‌ها، افراد یا هر چیز دیگر. بنابراین بطور دقیق که بخواهیم بیان کنیم، ما سعی نداریم بگوییم جمعیتی از افراد سن خاص همانند جمعیتی از افراد سن دیگر هستند، بلکه می‌گوییم که جمعیت نمرات بدست آمده تحت یک شرایط دارای میانگینی بزرگتر یا کوچکتر از جمعیت نمرات بدست آمده تحت شرایط دیگر است. این ممکن است به نظر ساده برسد ولی چنین نیست. بین افراد ۲۰ و ۶۰ ساله آشکارا تفاوت‌های عمده‌ای وجود دارد، و تردیدی نیست که آنها جمعیت‌های متفاوتی از افراد را تشکیل می‌دهند. اما معلوم نیست که نمرات "زمان واگذاری یک صندلی" از جمعیت‌های متفاوتی از نمرات باشند. در این زمینه خوب است که افراد ۲۰، ۴۰، و ۶۰ ساله را به این صورت در حال حرکت به اطراف با تعداد ضربه‌زدن انگشت به بینی‌شان (نمرات تأخیر) در نظر بگیرید. ما کاری نداریم که افراد از چه جمعیتی از موجودات انسانی بیرون کشیده شده‌اند، بلکه می‌خواهیم بدانیم

خدا موقع اعمال این عددها، آیا اینها را از یک کیسه (جمعیت) نمونه‌برداری کرده یا از دو یا سه کیسه مختلف.

پیش‌فرض نرمال بودن

ما به دلایل مرتبط با آزمون معناداری آخر، این پیش‌فرض را بنا خواهیم گذاشت که زمان بدست آوردن یک صندلی برای هر جمعیت در اطراف μ_j به طور نرمال توزیع شده است. این چیزی بیش از همان پیش‌فرض نیست که مشاهدات در شکل ۱-۱۶ (الف-ج) دارای توزیع نرمال می‌باشند. همانند t ، این پیش‌فرض در اصل با توزیع نمونه‌گیری میانگین سرو کار دارد نه توزیع مشاهدات. در هر حال، حتی انحراف زیاد از نرمال بودن تحت شرایط خاص ممکن است تأثیر خیلی کمی بر نتایج نهایی داشته باشد.

پیش‌فرض تجانس واریانس

یک پیش‌فرض اساسی دیگر این است که هر جمعیت از نمرات دارای واریانس یکسان است، بطور خاص $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_e^2$ در اینجا از نشانه σ_e^2 استفاده شده تا نشان داده شود که مقدار مشترک به این سه واریانس بر می‌گردد. اندیس e مخفف خطاست، زیرا واریانس، واریانس خطا است - یعنی واریانس درون یک جمعیت خاص و بنابراین به هیچ یک از گروه‌ها (ی سنی) مربوط نیست. چنانکه بعداً خواهید دید، تحت شرایط معین، این پیش‌فرض هم می‌تواند نادیده انگاشته شود بدون اینکه به نتیجه نهایی آسیب جدی وارد شود. بایستی توجه داشت که پیش‌فرض‌های نرمال بودن و تجانس واریانس چیز تازه‌ای نیستند. اینها همان پیش‌فرض‌هایی هستند که ما برای آزمون t برای دو نمونه مستقل طرح کردیم، به استثنای اینکه در اینجا آنها را به جمعیت‌های بیشتری گسترش داده‌ایم.

۲-۱۶ منطق تحلیل واریانس

برای چند لحظه به اثر دو پیش‌فرض ما - نرمال بودن و تجانس واریانس - دقت کنید. با طرح این پیش‌فرض‌ها گفته‌ایم که سه جمعیتی که نمرات آزمودنی‌های ۲۰، ۴۰ و ۶۰ ساله از آنها گرفته شده شکل و پراکندگی یکسانی دارند. در نتیجه، تنها راه تفاوت که برای آنها باقی می‌ماند در میانگین‌های آنهاست. (در واقع اگر فرض صفر درست باشد، $\mu_1 = \mu_2 = \mu_3$ و، چون جمعیت‌ها یکسان هستند، اگر ما از یک جمعیت نمونه‌گیری می‌کردیم چیزی را از دست نمی‌دادیم.)

فرض کنید در اثر محدودیت منابع ما فقط بتوانیم نه نمره را از هر گروه سنی بدست آوریم. واریانس نه نمره برای گروه ۲۰ ساله (گروه ۱) برآوردی از واریانس جمعیت نمرات همه آزمودنی‌های ۲۰ ساله ممکن است. ما می‌توانیم این را با نوشتن $\sigma_1^2 \hat{=} s_1^2$ نشان دهیم که علامت $\hat{=}$ بسته به متن "برآورد" یا

“برآورد شده بوسیله” خوانده می‌شود. به همین ترتیب $S_2^2 \doteq \sigma_2^2$ و $S_3^2 \doteq \sigma_3^2$ از آنجا که ما فرض کردیم $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_e^2$ ، پس S_1^2 ، S_2^2 و S_3^2 برآوردهایی از σ_e^2 ، واریانس مشترک جمعیت هستند. به منظور بدست آوردن بهترین برآورد ممکن از σ_e^2 ، ما این سه برآورد را با گرفتن میانگین آنها (به فرض اینکه $n_1 = n_2 = n_3$) جمع می‌کنیم:

$$\frac{S_1^2 + S_2^2 + S_3^2}{3} = \bar{s}_f^2 = \sigma_e^2$$

این در آمیختن واریانس‌ها دقیقاً همان کاری است که هنگام آمیختن واریانس‌ها در آزمون t انجام دادیم (هرچند اینجا بیش از دو واریانس داریم). این مقدار متوسط سه واریانس نمونه ($\bar{s}_f^2$) یک برآورد از جمعیت است (σ_e^2)، چیزی که بعداً از آن به عنوان درون گروهی MS یا $MS_{\text{خطا}}$ (بخوانید “میانگین مجزورات درون گروهی” یا “میانگین مجزورات خطا”) یاد خواهیم کرد. لازم است توجه کنید که این برآورد وابسته به درستی یا نادرستی H_0 نیست، چون روی هر نمونه بطور جداگانه محاسبه می‌شود. حالا اجازه دهید فرض کنیم که H_0 درست است. اگر چنین باشد، آنگاه یک نمونه نه تایی را می‌توان به عنوان سه نمونه مستقل از جمعیت مشابه در نظر گرفت، و ما برآورد ممکن دیگری از σ_e^2 داریم. حتماً بخاطر دارید که قضیه حد مرکزی می‌گوید واریانس میانگین‌های حاصل از یک جمعیت مشابه برابر واریانس جمعیت تقسیم بر اندازه نمونه است. تحت H_0 میانگین‌های نمونه از جمعیت مشابهی (یا جمعیت‌های یکسانی، که به چیز مشابهی می‌انجامد) بیرون کشیده شده‌اند و بنابراین

$$s_{\bar{X}}^2 \doteq \frac{\sigma_e^2}{n}$$

که n اندازه هر نمونه است. ما می‌توانیم ترتیب معمول چیزها را برعکس کنیم و، به جای برآورد واریانس میانگین‌ها از واریانس جمعیت، واریانس جمعیت (ها) را از واریانس میانگین‌های نمونه ($s_{\bar{X}}^2$) برآورد کنیم. اگر کسرها را ساده کنیم، داریم

$$ns_{\bar{X}}^2 \doteq \sigma_e^2$$

این عبارت به عنوان بین گروهی MS یا ساده‌تر گروهی MS معروف است.

این مراحل کوتاه را می‌توان به گونه‌ای ساده‌تر نشان داد. این کار برای سه اندازه گروهی برابر در شکل ۲-۱۶ آمده است. این شکل تأکید دارد که متوسط واریانس‌های نمونه‌ای $MS_{\text{خطا}}$ و واریانس میانگین نمونه‌ها ضربدر اندازه نمونه، گروهی MS است.

ما اکنون دو برآورد از واریانس جمعیت (σ_e^2) داریم. یکی از این برآوردها ($MS_{\text{خطا}}$) مستقل از درست یا نادرست بودن H_0 است. دیگری (گروهی MS) فقط مادامی برآورد کننده σ_e^2 است که H_0 درست باشد (تنها تا وقتی که شرایط قضیه حد مرکزی برقرار باشد، یعنی میانگین‌ها از یک جمعیت بر گرفته شده باشند). در غیر این صورت گروهی MS علاوه بر σ_e^2 تغییرپذیری میانگین‌های گروهی را هم برآورد خواهد کرد. اگر دو برآورد کننده (گروهی MS و $MS_{\text{خطا}}$) هماهنگ باشند، ما در باره درستی H_0 و اگر

ناهمانگی باشند برای نادرستی آن حمایت می‌شویم. اما پیش از آن که به روش‌های تعریف منظور خود از "همانگی" بپردازیم، استدلالی را که شرح دادیم با دو مثال ساده نشان می‌دهیم که به عمد برای نشان دادن نتایج دلخواه تحت شرایط درست بودن و نادرست بودن H_0 طرح شده‌اند. در عمل هرگز داده‌ها این اندازه آراسته و مرتب نخواهند بود.

$$\left. \begin{array}{l} s_1^2 \leftarrow \text{نمونه 1} \rightarrow \bar{X}_1 \\ s_2^2 \leftarrow \text{نمونه 2} \rightarrow \bar{X}_2 \\ s_3^2 \leftarrow \text{نمونه 3} \rightarrow \bar{X}_3 \end{array} \right\} \bar{X}_i \text{ واریانس } = s_{\bar{X}}^2$$

$$s_j^2 = \text{متوسط} \quad n(s_{\bar{X}}^2) = MS_{\text{گروهی}}$$

$$= MS_{\text{خطا}}$$

شکل ۲-۱۶ نمایشی از معنای $MS_{\text{گروهی}}$ و $MS_{\text{خطا}}$ وقتی اندازه‌ی نمونه‌ها برابر است

مثال - حالت یک H_0 درست

همانطور که قبلاً دیدیم، زمانی که H_0 درست باشد $\mu_1 = \mu_2 = \mu_3$ ، و هر نمونه‌ای از این سه جمعیت بیرون کشیده شود به عنوان برآمده از تنها یک جمعیت به حساب می‌آید. در مثال اول سه نمونه $n = 9$ انتخاب شدند تا داده‌هایی که از یک جمعیت دارای توزیع نرمال یکسان با میانگین ۵ و واریانس ۱۰ بدست آمده‌اند را شبیه‌سازی کنند. برای مثال این داده‌ها ممکن است تعداد دستورات جستجوی اطلاعات که توسط نه آزمودنی در هر گروه قبل از یک آزمایش آموزش جامعه‌پذیری بیان شده است را نشان دهند. چون هنوز آزمایش شروع نشده، امیدی نداریم که تفاوتی مشاهده کنیم. داده‌ها در جدول ۱-۱۶ برای ۳ $k =$ گروه آمده است. در این جدول مشاهده می‌شود که واریانس متوسط در هر گروه ۹/۲۵۰ است، یک برآورد قابل قبول از $\sigma_e^2 = ۱۰$. واریانس میانگین‌های گروهی ۱/۰۰۰ است، و چون می‌دانیم H_0 درست است،

$$s_{\bar{X}}^2 = \frac{\sigma_e^2}{n}$$

$$\sigma_e^2 = n s_{\bar{X}}^2 = 9(1) = 9$$

این مقدار همچنین از نظر همانگی با σ_e^2 و برآورد دیگر ما بر اساس تغییرپذیری درون آزمایشات نیز قابل قبول است. چون این دو برآورد همانگ هستند، می‌توانیم نتیجه بگیریم که دلیلی برای شک کردن به درستی H_0 وجود ندارد. به عبارت دیگر، سه نمونه بیش از آنچه در صورت درستی H_0 انتظار داریم فرق ندارند.

گروه ۱	گروه ۲	گروه ۳
۳	۱	۵
۶	۴	۲
۹	۷	۸
۶	۴	۸
۳	۱	۲
۱۲	۱۰	۸
۶	۴	۵
۳	۱	۲
۹	۷	۸
$\bar{X}_j = 6/3333$	$4/3333$	$5/3333$
$S_j^2 = 10/000$	$10/000$	$7/750$
$(GM) = 5/3333$ میانگین کل		
$S_j^2 = \frac{\sum(\bar{X}_j - GM)^2}{k-1} = 10/000$		
$S_j^2 = \frac{10.00+10.00+7.75}{3} = 9/250$		
$MS_{\text{خطا}} = \bar{S}_j^2 = 9/25$		
$MS_{\text{گروهی}} = nS_X^2 = 9(1) = 9$		

مثال - حالت یک H_0 نادرست

حالا به مثالی توجه کنید که در آن می‌دانیم H_0 نادرست است زیرا آن را اینگونه طرح کرده‌ایم. این داده‌ها در جدول ۱۶-۲ با افزودن یا کاستن مقادیر ثابتی به داده‌های جدول ۱۶-۱ بدست آمده‌اند. این داده‌ها می‌توانند نشان دهنده تعداد دستورات جستجوی اطلاعات توسط افراد سه گروه مختلف در پایان جلسه آموزش جامعه‌پذیری باشند. ما اکنون داده‌هایی داریم که بوسیله نمونه‌گیری از جمعیت‌های دارای توزیع نرمال تهیه شده‌اند که همه دارای واریانس برابر با ۱۰ هستند. به هر حال، نمرات گروه ۱ ممکن است از جمعیت با μ معادل ۸، ولی نمرات گروه ۲ و ۳ از جمعیتی با μ معادل ۴ باشند. این نشان دهنده یک انحراف قابل ملاحظه از H_0 است.

از جدول ۱۶-۲ در خواهید یافت که واریانس درون هر عملکرد آزمایشی بدون تغییر باقی مانده است، زیرا افزودن یا کاستن یک مقدار ثابت اثری بر واریانس درون گروه‌ها ندارد. این نشان دهنده مطلب قبلی است که واریانس درون گروهی ($MS_{\text{خطا}}$) مستقل از فرض صفر است. اما واریانس در بین میانگین‌های

گروه‌ها بطور قابل ملاحظه‌ای افزایش یافته، که گویای تفاوت بین میانگین جمعیت‌ها است. در این حالت برآورد σ_e^2 بر اساس میانگین نمونه‌ها $ns_{\bar{X}}^2 = 9(6/333) = 57$ است، مقداری که انحراف از مسیر برآورد $9/25$ حاصل از واریانس درون گروه‌ها ($MS_{\text{خطا}}$) می‌باشد. منطقی‌ترین نتیجه‌گیری این خواهد بود که $ns_{\bar{X}}^2$ نه تنها واریانس جمعیت را برآورد می‌کند بلکه برآورد واریانس σ_e^2 بعلاوه واریانس خود میانگین‌های جمعیت نیز هست. به عبارت دیگر نمرات نه تنها به خاطر خطای تصادفی تغییر می‌کنند؛ بلکه همچنین بخاطر اینکه ما در آموزش آزمودنی‌هایمان موفق بوده‌ایم نیز تغییر می‌کنند. در واقع ما می‌دانیم که این حالت برقرار است چون داده‌ها عمداً برای این منظور دستکاری شده‌اند.

جدول ۱-۱۶ داده‌های معرف برای حالت H_0 درست

گروه ۱	گروه ۲	گروه ۳
۵	۰	۵
۸	۳	۲
۱۱	۶	۸
۸	۳	۸
۵	۰	۲
۱۴	۹	۸
۸	۳	۵
۵	۰	۲
۱۱	۶	۸
$\bar{X}_j = 5/3333$	$3/3333$	$8/3333$
$s_j^2 = 10/000$	$10/000$	$7/750$

$$(GM) = 5/6667 = \text{میانگین کل}$$

$$s_j^2 = \frac{\sum (\bar{X}_j - GM)^2}{k-1} = 6/3333$$

$$S_j^2 = \frac{10.00+10.00+7.75}{3} = 9/250$$

$$MS_{\text{خطا}} = \bar{S}_j^2 = 9/25$$

$$MS_{\text{گروهی}} = ns_{\bar{X}}^2 = 9(6/333) = 57$$

خلاصه منطق تحلیل واریانس

با توجه به مبحث پیش می‌توان منطق تحلیل واریانس را به طور فشرده بیان نمود. برای آزمون H_0 ما دو برآورد از واریانس جمعیت را محاسبه کردیم؛ یکی ($MS_{\text{خطا}}$) مستقل از درستی یا نادرستی H_0 است، در حالی که دیگری ($MS_{\text{گروهی}}$) وابسته به H_0 می‌باشد. اگر این دو یکسان باشند، دلیلی برای رد کردن H_0

نداریم. اگر ناهمسان باشند، نتیجه می‌گیریم که بخشی از تفاوت در میانگین عملکردهای آزمایشی باید مربوط به برآورد دوم باشد، آن را متورم کرده و باعث می‌شود تا از اولی فرق کند. بنابراین H_0 را رد می‌کنیم. این نشان می‌دهد چگونه یک تحلیل واریانس به ما امکان می‌دهد تا در باره میانگین‌ها نتیجه‌گیری کنیم.

۳-۱۶ محاسبات تحلیل واریانس

محاسبات تحلیل واریانس در عمل خیلی ساده و سراسر است. فرمول‌ها متفاوت از آنچه در آزمون‌های t دیدید به نظر می‌رسند، اما این تفاوت در حقیقت برآمده از این واقعیت است که در اینجا ما در جهت تأکید بر استفاده از مجموع کل و مجموع مجزورات به جای میانگین‌ها و واریانس‌ها پیش می‌رویم.

مجموع مجزورات

در تحلیل واریانس بیشتر محاسبات ما به مجموع مجزورات مربوط می‌شود، که در این کتاب به عنوان، مجموع مجذور انحرافات از میانگین، $\sum (X - \bar{X})^2$ یا مضربی از آن است. مزیت مجموع مجزورات و دلیلی که ما با محاسبه آنها شروع می‌کنیم این است که قابل جمع و تفریق هستند، در حالی که این عملیات را روی مجذور میانگین‌ها نمی‌توان انجام داد. در فصل ۶ ما دیدیم که می‌توان $\sum (X - \bar{X})^2$ را به صورت $\sum X^2 - \sum (X)^2 / N$ نوشت و آن را برای اهداف محاسباتی به کار برد. اینجا هم دقیقاً همان کار را می‌کنیم به استثنای اینکه برای پرهیز از تکرار علامت جمع، ما $\sum X$ را با G (برای جمع کل) عوض کرده و داریم $\sum X^2 - G^2 / N$. برای محاسبه مجموع مجزورات ما یا این فرمول یا شکل دیگری از آن را به کار می‌بریم.

حاصل جمع‌ها

هر چند ما از میانگین‌های گروهی صحبت کرده بودیم، در عمل محاسبات خود را روی حاصل جمع‌های گروهی انجام خواهیم داد. هدف از این کار برای آسان‌سازی است گرچه اصولی نیست، زیرا حاصل جمع‌ها با میانگین‌ها رابطه خطی دارند. اگر دو گروه با اندازه‌ی یکسان، دارای حاصل جمع متفاوتی باشند، روشن است که میانگین‌های آنها فرق دارد.

محاسبات

برای حفظ پیوستگی مطلب، ما تحلیل واریانس را روی داده‌های جدول ۲-۱۶ اجرا می‌کنیم. این داده‌ها و محاسبات حاصل از آن که در صفحه بعد با جزئیات بحث شده‌اند در جدول ۳-۱۶ آمده است.

در بخش (الف) جدول ۳-۱۶ مشاهدات آزمایشی، حاصل جمع‌های گروهی (T_j)، و جمع کل ($\sum X$) $G =$ را می‌بینید. استفاده از نماد T_j برای نشان دادن جمع کل گروه زام است که در تمام مبحث تحلیل

واریانس از آن پیروی خواهد شد. هر چند درست‌تر بود که به عنوان عملکرد جمع کل $\sum_i X_{ij}$ به شکل $\sum_i X_{ij}$ بیان می‌شد، اما چنین نظام نمادگذاری ممکن است روشن و قابل فهم بودن مطالب را کاهش دهد. در تحلیل بعدی که در آن بیش از یک متغیر مستقل وجود دارد، T_j را می‌توان به ستونی T_{j1} و ردیفی T_{1j} بسط داد بدون آنکه از وضوح آن کاسته شود.

جدول ۳-۱۶ محاسبه تحلیل واریانس برای داده‌های جدول ۲-۱۶

(الف) داده‌ها

گروه ۱	گروه ۲	گروه ۳
۵	۰	۵
۸	۳	۲
۱۱	۶	۸
۸	۳	۸
۵	۰	۲
۱۴	۹	۸
۸	۳	۵
۵	۰	۲
۱۱	۶	۸
$G = 153$	$\bar{T}_j = 30$	$\bar{T}_j = 48$

(ب) محاسبات

$$\begin{aligned}
 SS_{\text{کل}} &= \sum X^2 - \frac{G^2}{N} \\
 &= (2^2 + 8^2 + \dots + 8^2 + 8^2) = \frac{153^2}{27} \\
 &= 1203 - 869 = 336 \\
 SS_{\text{گروهی}} &= \frac{\sum T_j^2}{n} - \frac{G^2}{N} \\
 &= \frac{(75^2 + 30^2 + 48^2)}{9} - \frac{153^2}{27} \\
 &= \frac{8829}{9} - 867 \\
 &= 981 - 867 = 114 \\
 SS_{\text{خطا}} &= SS_{\text{کل}} - SS_{\text{گروهی}} \\
 &= 336 - 114 = 222
 \end{aligned}$$

(ج) جدول خلاصه

منبع	df	SS	MS	F
گروه‌ها	۲	۱۱۴	۵۷/۰۰	۶/۱۶۲
خطا	۲۴	۲۲۲	۹/۲۵	
کل	۲۶	۳۳۶		

بخش (ب) جدول ۳-۱۶ شامل محاسبات مورد نیاز برای انجام یک تحلیل واریانس یک طرفه است. مقداری تمرین در این بخش لازم است.

$SS_{\text{کل}}$ $SS_{\text{گ}}$ (بخوانید "مجموع مجذورات کل") نشان دهنده مجموع مجذورات همه مشاهدات است، بدون توجه به اینکه کدام عملکرد آزمایشی آن را ایجاد کرده است. مجموع همه مشاهدات به توان دو، منهای مجذور جمع کل تقسیم بر N است.

$$SS_{\text{گ}} = \sum X^2 - \frac{G^2}{N}$$

گروهی SS عبارت گروهی SS اندازه‌ای از تفاوت‌های ناشی از گروه‌ها است (در واقع تفاوت‌های بین میانگین‌های گروهی) و مستقیماً به تغییرپذیری حاصل جمع‌های گروهی ربط دارد. برای محاسبه گروهی SS ما بسادگی حاصل جمع‌های هر گروه را به توان دو رسانده و با هم جمع می‌کنیم، بر تعداد مشاهداتی که هر گروه بر آن بنا شده (در اینجا n) تقسیم کرده و از G^2/N کم می‌کنیم. برای اینکه اطلاعات حاصل از گروهی SS را بهتر بشناسیم، فرمول آن را باید به طرز دیگری بازنویسی کنیم (که k معرف تعداد گروه‌ها باشد):

$$\begin{aligned} SS_{\text{گروهی}} &= \frac{\sum T_j^2}{n} - \frac{G^2}{N} \\ &= \frac{\sum T_j^2}{n} - \frac{(\sum T_j)^2}{nk} = \frac{\sum T_j - \frac{(\sum T_j)^2}{n}}{n} = \frac{\sum (T_j - \bar{T})^2}{n} \end{aligned}$$

از این پرسش می‌توان دید که گروهی SS نشان دهنده مجموع مجذور انحرافات حاصل جمع‌های عملکرد پیرامون میانگین حاصل جمع‌ها ($\bar{T}$)، تقسیم بر n است. در این حالت n دقیقاً همانند مقسوم‌علیه است که در رابطه با قضیه حد مرکزی بحث شد. (در مخرج کسر قرار می‌گیرد نه در صورت، به این دلیل روشن که ما با حاصل جمع‌ها کار می‌کنیم نه میانگین‌ها). هدف آن در نهایت ایجاد برآوردی از σ_e^2 است. در همه این مجموع مجذورات مورد بحث در تحلیل واریانس، اصل همگانی مشابهی به کار می‌رود.

قاعده کلی محاسبه هر مجموع مجذورات پیرو این بحث گروهی SS اکنون می‌توان یک قاعده کلی برای محاسبه هر مجموع مجذورات (SS) طرح کرد، غیر از آنهایی که با تفریق حساب خواهیم کرد: برای هر SS به جز خطا SS ، جمع کل‌ها را به توان دو رسانده، بر تعداد مشاهداتی که هر حاصل جمع داشته تقسیم کنید، حاصل را جمع کرده و G^2/N را از آن کم کنید.

این قاعده به شما امکان می‌دهد که SS را برای هر مثال این کتاب حساب کنید، بدون توجه به اینکه طرح آزمایشی چقدر پیچیده است. در تحلیل واریانس یک طرفه "حاصل جمع‌های مربوط" که هم اینک به آنها اشاره شد حاصل جمع‌های مربوط به گروه‌های مختلف هستند. وقتی ما در فصل بعد به سراغ تحلیل واریانس دو طرفه می‌رویم، فقط این قاعده را تعمیم خواهیم داد تا امکان ابعاد مختلفی که گروه‌ها را شکل می‌دهند فراهم گردد. (این قاعده همچنین در محاسبه $SS_{\text{کل}}$ به کار می‌رود، اما در آنجا جمع کل‌های مربوط شامل مشاهدات منفرد است و در نتیجه مقسوم علیه ۱ می‌شود که معمولاً نشان داده نمی‌شود.)

$SS_{\text{خطا}}$ در عمل $SS_{\text{خطا}}$ معمولاً از طریق تفریق بدست می‌آید. چون بسادگی می‌توان نشان داد که

$$SS_{\text{خطا}} = MS_{\text{بین گروهی}} + MS_{\text{کل}}$$

آنگاه این نیز باید درست باشد که

$$MS_{\text{خطا}} = SS_{\text{کل}} - MS_{\text{بین گروهی}}$$

این شیوه‌ی ارائه شده در جدول ۱۶-۳ است. یک روش محاسبه دیگر هم وجود دارد. همانطور که از مبحث پیش به خاطر دارید، ما بدنبال عبارتی بودیم که تحت تأثیر تفاوت‌های بین عملکردهای آزمایشی قرار نگیرد و در نتیجه عبارتی باشد که تغییرپذیری درون هر یک از سه عملکرد را بطور جداگانه نشان دهد. تا اینجا ما توانستیم یک مجموع مجذورات درون گروهی عملکرد بین گروهی $MS - ۱$ ، و یک عبارت مشابه برای SS درون هر یک از عملکردهای دیگر را حساب کنیم:

$$\begin{aligned} SS_{\text{خطا}1} &= 5^2 + 8^2 + \dots + 11^2 - \frac{75^2}{9} = 705 - 625 = 80 \\ SS_{\text{خطا}2} &= 0^2 + 3^2 + \dots + 6^2 - \frac{30^2}{9} = 180 - 100 = 80 \\ SS_{\text{خطا}3} &= 5^2 + 2^2 + \dots + 8^2 - \frac{48^2}{9} = 318 - 256 = 62 \\ SS_{\text{خطا}} &= 222 \end{aligned}$$

زمانی که ما این عبارات منفرد را جمع کنیم مقدار ۲۲۲ بدست می‌آید، که در واقع همان جواب بدست آمده در جدول ۱۶-۳ است. این بسادگی نشان می‌دهد که $SS_{\text{خطا}}$ اندازه‌ای از تغییرپذیری درون هر گروه است.

جدول خلاصه

بخش (ج) جدول ۱۶-۳ جدول خلاصه‌ی تحلیل واریانس است. این جدول خلاصه نامیده می‌شود زیرا مجموعه محاسبات را خلاصه می‌کند، و امکان آن را فراهم می‌سازد تا با یک نگاه به داده‌ها بگوییم چه چیزی را نشان می‌دهند.

منبع تغییرات ستون اول جدول خلاصه شامل منابع تغییرات است - واژه "تغییرات" با عبارت "مجموع مجذورات" هم‌معنی است. همانطور که در این جدول مشاهده می‌کنید، سه منبع تغییرات وجود دارد: تغییرات کل، تغییرات ناشی از گروه‌ها (تغییرات بین میانگین گروه‌ها)، تغییرات ناشی از خطا (تغییرات درون گروه‌ها). این منابع به این واقعیت اشاره دارند که ما کل مجموع مجذورات را به دو بخش، یک بخش نشانگر تغییرات بین چند گروه و بخش دیگر نشانگر تغییرات درون گروه‌های منفرد تقسیم کرده‌ایم.

درجه آزادی ستون درجات آزادی نشانگر تخصیص تعداد کل درجات آزادی بین دو منبع واریانس است. محاسبه df احتمالاً ساده‌ترین بخش این کار است. درجات آزادی کل (df) همیشه $N - 1$ است، که N تعداد کل مشاهدات است. درجات آزادی بین گروهی (df بین گروهی) همیشه برابر $k - 1$ است، که k تعداد گروه‌ها است. درجات آزادی خطا (df خطا) ساده‌ترین بخش باقیمانده به حساب می‌آید، گرچه به طور سراسرتر می‌توان آن را به عنوان مجموع درجات آزادی درون هر عملکرد محاسبه نمود. در مثال ما $df = 27 - 1 = 26$. از این ۲۶ درجه آزادی، ۲ درجه به تفاوت بین گروه‌ها مربوط است و ۲۴ درجه دیگر مربوط به تغییرپذیری درون گروه‌ها می‌شوند.

به جای فراگیری یک مجموعه معادله برای محاسبه درجات آزادی، بهتر است منطق آن را بفهمیم. SS مجموع N انحراف مجذور شده در اطراف یک نقطه - میانگین اصلی - است. این واقعیت که ما انحرافات اطراف یک نقطه (برآورد شده) را بدست می‌آوریم برای ما $df = 1$ هزینه دارد، بنابراین برای ما $df = N - 1$ باقی می‌ماند. بین گروهی SS مجموع انحرافات k میانگین گروهی پیرامون یک نقطه (همان میانگین اصلی) است، و باز در برآورد این نقطه ما $df = 1$ از دست می‌دهیم و $df = k - 1$ باقی می‌ماند. خطا SS نشانگر k مجموعه از n انحراف پیرامون یک نقطه (میانگین اصلی) است، $df = 1$ برای هر گروه از دست می‌رود و $df = k(n - 1) = N - k$ باقی می‌ماند.

به بیان دیگر، کل تغییرپذیری بر مبنای N نمره است و بنابراین دارای $df = N - 1$ است. تغییرپذیری میانگین‌های عملکرد بر اساس k نمره (میانگین یا حاصل جمع) و بنابراین دارای $df = k - 1$ است. تغییرپذیری درون هر عملکرد بر اساس n نمره و در نتیجه دارای $df = n - 1$ است، اما چون ما k مورد از این عبارات درون عملکردی را جمع می‌کنیم، ما k بار $df = k(n - 1) = n - 1$ خواهیم داشت. **مجموع‌های مجذورات** در باره ستون SS چیز زیادی برای گفتن وجود ندارد. بطور ساده این ستون شامل مجموع مجذورات حاصل از بخش (ب) جدول است.

میانگین مجذورات ستون میانگین مجذورات شامل دو برآورد از σ_e^2 است. این مقدار از تقسیم مجموع مجذورات بر df خود بدست آمده است. بنابراین $57 = 114/2$ و $9.25 = 222/24$. ما معمولاً

MS را محاسبه نمی‌کنیم، چون استفاده‌ای از آن نمی‌کنیم. اگر این کار را می‌کردیم، عبارت حاصل نشان دهنده واریانس تمام N مشاهده بود.

در حالی که این درست است که میانگین مجذورات، واریانس‌ها هستند، مهم است که بدانیم این عبارات واریانس چیزی هستند. بنابراین خطای MS (متوسط) واریانس مشاهدات درون هر عملکرد است. اما، بین گروهی MS واریانس میانگین‌ها یا حاصل جمع‌های گروهی نیست، بلکه واریانس این میانگین‌ها (یا حاصل جمع‌ها) روی n تصحیح شده تا برآوردی از واریانس جمعیت (σ_e^2) حاصل شود- به عبارت دیگر برآوردی از σ_e^2 بر اساس واریانس میانگین‌های گروهی است.

آماره F ستون آخر تحت عنوان F مهمترین چیز در آزمون فرض صفر است. از تقسیم بین گروهی MS بر خطای MS بدست می‌آید. چنانکه قبلاً گفته شده خطای MS برآوردی از واریانس جمعیت (σ_e^2) است. بین گروهی MS نیز برآوردی از واریانس جمعیت (σ_e^2) است در حالی که H_0 درست باشد، اما اگر درست نباشد چنین نیست. اگر H_0 درست باشد آنگاه هر دو گروهی MS و خطای MS چیز یکسانی را برآورد می‌کنند، و به این ترتیب تقریباً باید برابر باشند. اگر چنین باشد نسبت یکی به دیگری تقریباً ۱ است، ضمن اینکه یک مقدار قابل قبول برای خطای نمونه‌گیری در نظر گرفته می‌شود. بنابراین همه آنچه که ما باید انجام دهیم این است که این نسبت را حساب کنیم و مشخص کنیم که آیا به اندازه کافی به ۱ نزدیک است یا نه، که نشان دهنده حمایت از فرض صفر باشد.

شیوه دیگر نگرستن به F این است که واریانس میانگین‌های جمعیتی ($\mu_1 = \mu_2 = \mu_3$) را به صورت σ_t^2 نشان دهیم. (بخاطر داشته باشید که ما خود واریانس‌های جمعیت را با σ_e^2 نشان می‌دادیم). سپس اگر ما یک $(\wedge)$ به عنوان نشانه پارامتر برآورد شده‌ی آن به کار ببریم

$$MS_{\text{بین گروهی}} = \hat{\sigma}_e^2 + n\hat{\sigma}_t^2 \quad \text{و} \quad MS_{\text{خطا}} = \hat{\sigma}_e^2$$

اگر H_0 درست باشد $\mu_1 = \mu_2 = \mu_3$ ، σ_t^2 برابر صفر خواهد بود، باقی می‌ماند

$$F = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2} \approx 1$$

اما اگر H_0 نادرست باشد σ_t^2 ، برابر صفر نخواهد بود، آنگاه

$$F = \frac{n\hat{\sigma}_t^2 + \hat{\sigma}_e^2}{\hat{\sigma}_e^2} > 1$$

پرسشی که باقی می‌ماند این است که مقدار F بدست آمده ما چقدر از ۱/۰ فاصله داشته باشد تا تصمیم به تفاوت بین میانگین‌های جمعیت بگیریم و H_0 را رد کنیم. جواب این سؤال به این واقعیت برمی‌گردد که اگر H_0 درست باشد، نسبت

$$\frac{MS_{\text{گروهی}}}{MS_{\text{خطا}}}$$

به صورت توزیع F در جدول مخصوص F در بخش پیوست کتاب بر اساس بین گروهی df و خطای df درجه آزادی توزیع شده است. بخشی از این جدول در جدول ۴-۱۶ نشان داده شده است.) چون شکل توزیع F و در نتیجه سطح زیر آن وابسته به درجات آزادی دو میانگین مجزورات است، این جدول با جدول‌های دیگری که دیده‌اید فرق دارد. در اینجا ما ستون درجات آزادی مربوط به میانگین مجزورات را در صورت کسر F (مثل $k-1$) و ردیف درجات آزادی مربوط به میانگین مجزورات را در مخرج (مثل $k(n-1)$) قرار می‌دهیم. محل برخورد این ردیف و ستون، مقدار بحرانی F را در سطح α در بالای جدول نشان می‌دهد. برای استفاده از جدول F پیوست، ابتدا بایستی جدول خاص مربوط به سطح α مورد نظرمان را پیدا کنیم. چون در این کتاب ما بطور معمول α را در سطح ۰/۰۵ در نظر گرفته‌ایم، با آن جدول شروع می‌کنیم. از آنجا که ما ۲ df برای صورت (گروهی MS) و ۲۴ df برای مخرج (خطای MS) داریم، ما در ستون دوم به سمت پایین پیش می‌رویم تا به ردیفی که عنوان آن ۲۴ است برسیم. در محل تقاطع این ردیف و ستون به مقدار ۳/۴۰ بر می‌خوریم. این مقدار بحرانی F است. در صورت درست بودن H_0 تنها در ۵٪ موارد انتظار خواهیم داشت که یک F از مقدار ۳/۴۰ فراتر باشد. از آنجا که مقدار بدست آمده ۶/۱۶۲ = F از $F = ۳/۴۰$ بیشتر است، ما H_0 را رد می‌کنیم و نتیجه می‌گیریم که این گروه‌ها از جمعیت‌هایی با میانگین‌های متفاوت نمونه‌گیری شده‌اند. (اگر بخواهید α را در سطح ۰/۰۱ در نظر بگیرید در جدول F پیوست خواهید دید که $F_{(2, 24)} = ۵/۶۱$ ، و نتیجه می‌گیریم باز هم H_0 رد خواهد شد.)

جدول ۴-۱۶ خلاصه جدول مقادیر بحرانی F برای $\alpha = 0.05$ = آلفا از جدول پیوست کتاب

	Degrees of Freedom for Numerator										
	1	2	3	4	5	6	7	8	9	10	15
1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	...
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	...
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	...
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	...
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	...
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	...
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	...
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	...
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	...
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	...
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	...
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	...
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	...
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	...
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	...
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	...
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	...
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	...
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	...
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	...
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	...
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	...
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	...
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	...
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	...
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	...
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	...
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	...
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	...
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	...
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	...
1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	...

۴-۱۶ نمونه‌های نابرابر

اغلب تحقیقات در اصل به گونه‌ای طرح می‌شوند که تعداد مشاهدات در هر عملکرد آزمایشی برابر باشند. اما در بسیاری وقت‌ها شرایط این گونه پیش نمی‌رود. حیوانات گاهی در طول آزمایش بدلایلی غیر از مسائل آزمایشی می‌میرند. آزمودنی‌ها خیلی ناپایا هستند و بسیاری وقت‌ها برای آزمون در دسترس نیستند یا از ادامه کار منصرف می‌شوند. حتی موردی بوده که حیوانی بدلیل اینکه مرتب آزمایشگر را گاز گرفته از تحقیق حذف شده است (اسگرو و وین‌استوک، ۱۹۶۳). از این گذشته در مطالعه گروه‌های دست‌نخورده مانند کلاس‌های مدرسه، گروه‌ها تقریباً همیشه نابرابر هستند بدلایلی که ربطی به آزمایش ندارد.

اگر اندازه نمونه‌ها برابر نباشد، تحلیلی که قبلاً مورد بحث قرار گرفت بدون اصلاح مناسب نیست. برای موردی که یک متغیر مستقل وجود دارد، اصلاح نسبتاً کمی لازم است. برای حالتی که اندازه نمونه‌ها برابر باشد ما

$$SS_{\text{گروهی}} = \frac{\sum T_j^2}{n} - \frac{G^2}{N}$$

تعریف کردیم که n تعداد مشاهدات در هر گروه است. ما می‌توانستیم هر T_j^2 (و در نتیجه $\sum T_j^2$) را بر n تقسیم کنیم، زیرا n برای همه عملکردهای آزمایشی مشترک بود. اما اگر اندازه نمونه فرق کند، ما n_j را به عنوان تعداد آزمودنی‌ها در عملکرد j ام ($\sum n_j = N$) تعریف می‌کنیم، می‌توانیم معادله را به صورت زیر بازنویسی کنیم

$$SS_{\text{گروهی}} = \frac{\sum T_j^2}{n_j} - \frac{G^2}{N}$$

که وقتی همه n_j ها برابر باشند، به شکل اولیه باز می‌گردد.

- یک مثال دیگر - سازگاری با نقش‌های مادری

مثالی دیگر از تحلیل واریانس یک طرفه موضوع نمونه‌های نابرابر را هم نشان خواهد داد. در یک تحقیق پیرامون رشد نوزادان کم وزن (نورکومب و همکاران، ۱۹۸۵)، وقتی نوزادان ۶ ماهه بودند با مادران آنها مصاحبه شد. در این آزمایش سه گروه وجود داشت - یک گروه آزمایشی کم وزن، یک گروه کنترل کم وزن و یک گروه عادی. گروه آزمایشی کم وزن بخشی از یک برنامه مداخله‌ای بودند و انتظار می‌رفت این مادران به همان خوبی مادران گروه عادی با نقش‌های خود سازگار شوند. از سوی دیگر انتظار می‌رفت مادران کودکان کم وزنی که این برنامه مداخله را نداشتند مشکلاتی در سازگاری داشته باشند. (والدین یک کودک کم وزن بودن کار آسانی نیست، بخصوص برای ماه‌های اول).

داده‌های واقعی برای این مطالعه در بخش (الف) جدول ۵-۱۶ آمده است. در بخش (ب) این جدول محاسبات تحلیل واریانس، و در بخش (ج)، جدول خلاصه ارائه شده است. توجه داشته باشید که محاسبات به همان صورتی که برای نمونه‌های برابر وجود دارند انجام شده‌اند به استثنای اینکه برای گروهی SS هر مقدار T_j^2 بر اندازه نمونه مربوط به آن تقسیم شده است، نکته‌ای که قبلاً گفته شد.

در جدول خلاصه مشاهده می‌کنیم که F بدست آمده ۵/۵۳ است و بر مبنای ۲ و ۹۰ درجه آزادی است. از جدول پیوست مقادیر بحرانی F در می‌یابیم که

$$F_{.05}(2, 90) = 3.11$$

(یعنی ۹۰ در نیمه راه بین ۶۰ و ۱۲۰ df است، بدین ترتیب به عنوان مقدار بحرانی، ما مقدار بین ۳/۱۵ و ۳/۰۷ را بر خواهیم گزید). چون $5.53 > 3.11$ است، H_0 را رد می‌کنیم و نتیجه می‌گیریم که همه نمرات هر گروه از جمعیت‌هایی با میانگین برابر بدست نیامده‌اند. در واقع چنین به نظر می‌رسد که گروه

اول و سوم تقریباً یکسان هستند، در حالی که گروه دوم میانگین بالاتری (سازگاری ضعیف‌تری) دارد. اما F فقط به ما می‌گوید که ما می‌توانیم $\mu_1: H_0 = \mu_2 = \mu_3$ را رد کنیم. به ما نمی‌گوید کدام گروه‌ها با هم فرق دارند. برای چنین نتیجه‌ای ما به شیوه‌های خاصی نیاز داریم که به شیوه‌های مقایسه چندگانه معروف هستند.

جدول ۵-۱۶ - سازگاری با نقش‌های مادری در سه گروه مادران (نمره پایین با سازگاری بهتر همراه است)

داده‌ها

(الف)

گروه ۱ آزمایشی کم وزن	گروه ۲ کنترل کم وزن	گروه ۳ نوزادان عادی	
۲۴ ۱۰ ۱۶	۲۱ ۱۷ ۱۳	۱۲ ۱۲ ۱۲	
۱۳ ۱۱ ۱۵	۱۹ ۱۸ ۲۵	۲۵ ۱۷ ۲۰	
۲۹ ۱۳ ۱۲	۱۰ ۱۸ ۱۶	۱۴ ۱۸ ۱۴	
۱۲ ۱۹ ۱۶	۲۴ ۱۳ ۱۸	۱۶ ۱۸ ۱۴	
۱۴ ۱۱ ۱۲	۱۷ ۲۱ ۱۱	۱۳ ۱۸ ۱۲	
۱۱ ۱۱ ۱۲	۲۵ ۲۷ ۱۶	۱۰ ۱۵ ۲۰	
۱۲ ۲۷ ۲۲	۱۶ ۲۹ ۱۱	۱۳ ۱۳ ۱۲	
۱۳ ۱۳ ۱۶	۲۶ ۱۴ ۲۱	۱۱ ۱۵ ۱۷	
۱۳ ۱۳ ۱۷	۱۹ ۱۷ ۱۳	۲ ۱۳ ۱۵	
۱۳ ۱۴ ۱۴		۲۳ ۱۳ ۱۱	
		۱۶ ۱۰ ۱۳	
		۲۰ ۱۲ ۱۱	
		۱۱	
۴۳۴	۴۹۵	$\sum X = ۵۴۹$	$G = ۱۴۷۸$ $N = ۹۳$
۲۹	۲۷	$n = ۳۷$	
۱۴/۹۷	۱۸/۳۳	$\bar{X} = ۱۴/۸۴$	

(ب)

محاسبات

$$\begin{aligned}
 SS_{\text{ج}} &= \sum X^2 - \frac{G^2}{N} \\
 &= ۲۴^2 + ۱۰^2 + \dots + ۱۱^2 + ۱۱^2 - \frac{1478^2}{93}
 \end{aligned}$$

$$\begin{aligned}
 &= 25652 - 23489/0.75 \\
 &= 2072/925 \\
 SS_{\text{گروهی}} &= \frac{\sum T_j^2}{n_j} - \frac{G^2}{N} \\
 &= \frac{434^2}{29} + \frac{495^2}{27} + \frac{549^2}{37} - \frac{1478^2}{93} \\
 &= 23716/0.7 - 23489/0.75 \\
 &= 226/932 \\
 SS_{\text{خطا}} &= SS_{\text{کل}} - SS_{\text{گروهی}} \\
 &= 2072/925 - 226/932 \\
 &= 1845/993
 \end{aligned}$$

(ج) جدول خلاصه

منبع	df	SS	MS	F
گروه‌ها	۲	۲۲۶/۹۳۲	۱۱۳/۴۶۶	۵/۵۳
خطا	۹۰	۱۴۸۵/۹۹۳	۲۰/۵۱۱	
کل	۹۲	۲۰۷۲/۹۲۵		

۵-۱۶ و شیوه‌های مقایسه چندگانه

قتی ما یک تحلیل واریانس را اجرا می‌کنیم و یک مقدار F معنادار بدست می‌آوریم، چیزی که به ما نشان داده می‌شود این است که فرض صفر کلی نادرست است. اما نمی‌دانیم که کدام یک از فرض‌های ممکن مقابل (مثل، $\mu_3 \neq \mu_2 \neq \mu_1$ ؛ $H_1: \mu_1 \neq \mu_2 = \mu_3$) درست است. شیوه‌های مقایسه چندگانه^۱ به ما امکان می‌دهد تا به بررسی فرضیه‌های مربوط به میانگین‌های گروه‌های منفرد یا مجموعه‌هایی از گروه‌ها بپردازیم. بنابراین برای مثال ممکن است ما بخواهیم بدانیم که آیا گروه ۱ با گروه ۲ فرق دارد یا آیا ترکیب گروه ۱ و ۲ از گروه ۳ متفاوت است.

یکی از مسائل اصلی در رابطه با انجام مقایسه بین گروه‌ها این است که استفاده نامحدود از این مقایسه‌ها می‌تواند باعث افزایش فزاینده احتمال خطای نوع I شود. برای مثال اگر ما ده گروه داشته باشیم که فرض صفر کلی درست باشد ($H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_{10}$) آزمون‌های t بین همه جفت‌های میانگین در حدود ۶۰٪ اوقات باعث ایجاد دست‌کم یک خطای نوع اول خواهد شد. به عبارت دیگر آزمایشگر که فکر می‌کند در سطح معناداری $\alpha = 0.05$ کار می‌کند در عمل در سطح $\alpha = 0.60$ کار

1. multiple comparison procedures

می‌کند. اگر چه یافتن تفاوت معنادار جالب است، یافتن چیزی که واقعی نیست جالب نخواهد بود. روان‌شناسان در توضیح همه تفاوت‌های واقعی که ما پیدا می‌کنیم به اندازه کافی مشکل دارند، گذشته از اینکه نسبت به تفاوت‌های غیرواقعی هم نگران هستند.

در تلاش برای کنترل احتمال خطاهای نوع I، آماردانان شمار زیادی از شیوه‌های مقایسه میانگین‌های منفرد را توسعه داده‌اند. خوشبختانه یکی از بهترین شیوه‌ها از ساده‌ترین آنها نیز می‌باشد، و برای بسیاری از مسائل مقایسه چندگانه که ممکن است شما بر خورد کنید کاربرد دارد. این شیوه اغلب به عنوان **t محافظت شده^۱ یا آزمون کمترین تفاوت معناداری فیشر^۲** نامیده شده و ابتدا به آن خواهیم پرداخت.

این شیوه‌ها برای استفاده از t محافظت شده خیلی ساده هستند. نخستین چیز لازم برای یک t محافظت شده این است که F کلی برای تحلیل واریانس بایستی معنادار باشد. اگر F معنادار نباشد، مجاز به مقایسه بین میانگین‌ها نیستیم. به طور ساده خواهید گفت تفاوتی بین میانگین‌ها وجود ندارد و همین‌جا کار پایان می‌یابد. از سوی دیگر اگر F کلی معنادار باشد، آنگاه شما می‌توانید هر مقایسه جفتی بین میانگین‌ها را با استفاده از یک آزمون t اصلاح شده پیگیری کنید. این اصلاح بطور ساده با جایگزینی برآورد واریانس جمعی (s_p^2) در فرمول استاندارد t با $MS_{\text{خطا}}$ از تحلیل کلی واریانس صورت می‌گیرد. این کار چندان که می‌نماید دشوار نیست. زیرا $MS_{\text{خطا}}$ به صورت میانگین واریانس‌های درون هر گروه تعریف می‌شود، اگر تنها دو گروه در آزمایش وجود داشته باشند، $MS_{\text{خطا}}$ تحلیل واریانس همانند s_p^2 آزمون t دو نمونه‌ای خواهد بود. در مقایسه بین چند گروه ما از $MS_{\text{خطا}}$ به جای s^2 استفاده می‌کنیم چون بر اساس تغییرپذیری درون همه گروه‌ها است نه درون تنها دو گروهی که ما در حال مقایسه آنها هستیم. بر این اساس احتمالاً این یک برآورد بهتری از σ_e^2 است^۳. در کنار استفاده از این عبارت خطا این مزیت وجود دارد که t حاصل دارای $df_{\text{خطا}}$ درجه آزادی خواهد بود نه فقط $n_1 + n_2 - 2$ درجه آزادی که در غیر این صورت می‌داشت.

وقتی ما s_p^2 را با $MS_{\text{خطا}}$ جایگزین کنیم فرمول t می‌شود

1. protected t

2. Fisher's least significant difference test

۳. اگر SS_j را برای بازنمایی مجموع مجدورات درون گروهی j (مثلاً $SS_j = \sum (X_{ij} - \bar{X}_j)^2$) در نظر بگیریم، آنگاه

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} = \frac{ss_1^2 + ss_2^2}{df_1 + df_2}$$

علاوه بر این

$$MS_{\text{خطا}} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + \dots + (n_k-1)s_k^2}{n_1 + n_2 + \dots + n_k - k} = \frac{ss_1^2 + ss_2^2 + \dots + ss_k^2}{df_1 + df_2 + \dots + df_k}$$

بنابراین با دو گروه

$$s_p^2 = MS_{\text{خطا}}$$

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{خطا}}}{n_i} + \frac{MS_{\text{خطا}}}{n_j}}} = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

برای نمایش کاربرد t محافظت شده می‌توانیم داده‌های سازگاری مادران را از مثال قبلی به کار بگیریم. در آن حالت ما یک F معنادار پیدا کردیم که امکان جلو رفتن در تحلیل را ممکن می‌ساخت. با وجود آن مطالعه دو سؤال مطرح می‌شود:

- ۱- آیا تفاوتی بین مادران در گروه کنترل کودکان کم وزن و مادران در گروه کودکان عادی وجود دارد؟
- ۲- آیا تفاوتی بین مادران در گروه کنترل کودکان کم وزن و گروه آزمایشی وجود دارد؟

سؤال نخست می‌پرسد که آیا مادران نوزادانی که هنگام تولد کم وزن بوده‌اند مشکل سازگاری بیشتری نسبت به مادران نوزادان عادی دارند. سؤال دوم می‌پرسد که آیا برنامه مداخله تفاوتی در سازگاری ایجاد کرده است. توجه داشته باشید که مقایسه گروه آزمایشی کم وزن با گروه عادی چیز زیادی نشان نمی‌دهد زیرا در صورت مشاهده تفاوت نمی‌توان گفت ناشی از اثرات مداخله است یا اثرات وزن هنگام تولد. در این مقایسه مداخله و وزن هنگام تولد با هم اشتباه می‌شوند.

نتایج مربوط به سازگاری مادر در جدول ۶-۱۶ آمده است، در این جدول مقادیر t بدست آمده برای دو مقایسه ۲/۷۷- و ۳/۰۴ بودند. ما یک آزمون دو دامنه در سطح $\alpha = ۰/۰۵$ به کار خواهیم برد و برای عبارت خطا ۹۰ درجه آزادی داریم.

جدول ۶-۱۶ مقایسه‌های چندگانه میان گروه‌های کم وزن و عادی

	گروه ۳	گروه ۲	گروه ۱
	نوزادان عادی	کنترل کم وزن	آزمایشی کم وزن
$\bar{X}_j =$	۱۴/۸۴	۱۸/۳۳	۱۴/۹۷
n_j	۳۷	۲۷	۲۹
$MS_{\text{خطا}}$			۲۰/۵۱۱
$df_{\text{خطا}}$			۹۰

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (\text{الف}) \mu_1 \text{ در مقابل } \mu_2$$

$$= \frac{14.97 - 18.33}{\sqrt{20.511 \left(\frac{1}{29} + \frac{1}{27} \right)}}$$

$$= \frac{-3.36}{\sqrt{1.467}} = \frac{-3.36}{1.21} = -۲/۷۷$$

(ب) μ_2 در مقابل μ_3

$$\begin{aligned}
 t &= \frac{\bar{X}_2 - \bar{X}_3}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_2} + \frac{1}{n_3} \right)}} \\
 &= \frac{18.33 - 14.84}{\sqrt{20.511 \left(\frac{1}{29} + \frac{1}{27} \right)}} \\
 &= \frac{3.49}{\sqrt{1.314}} = \frac{3.49}{1.15} = 3.04
 \end{aligned}$$

از جدول t پیوست در می‌یابیم که $1/98 \pm t_{.05}(90)$. بنابراین برای هر دو مقایسه ما می‌توانیم فرض صفر را رد کنیم، زیرا هر دو مقدار حاصل t بزرگتر از $1/98 \pm$ هستند. بر این اساس نتیجه خواهیم گرفت که بین سازگاری مادران نوزادان کم وزن و عادی تفاوت وجود دارد، مادران عادی سازگاری بهتری دارند. همچنین نتیجه خواهیم گرفت که برنامه مداخله اثر گذار است.

احتمالا شما می‌پرسید چرا این مقایسه چندگانه خاص را بدنبال یک "t محافظت شده" صورت می‌دهیم. یا، ممکن است جایی شنیده باشید که اجرای همه آزمون‌های t بین جفت میانگین‌ها ایده نادرستی است. الان وقت مناسبی است که هر دوی این مطالب را همزمان توضیح دهیم.

یکی از توجهات اولیه در اجرای یک مجموعه مقایسه‌های چندگانه این است که دست کم احتمال یک خطای نوع I را در نظر بگیریم. به عبارت دیگر اگر ما یک تحلیل واریانس و سپس سه مقایسه انجام دهیم، ما می‌خواهیم اطمینان حاصل کنیم که احتمال کمی وجود دارد در هر کجا، خواه در F اصلی یا هر یک از سه مقایسه، دچار خطای نوع I شویم. احتمال انجام چنین خطایی **میزان خطای خانوادگی**^۱ خوانده می‌شود زیرا با این احتمال سروکار دارد که خانواده‌ی مقایسه‌ها، دست کم یک خطای نوع I را شامل شود. برای میزان خطای خانوادگی، انجام ده خطای نوع I بدتر از یک خطا نیست. در هر مورد شما یک خطا می‌کنید. (به بیان دیگر، تنها یک خطا داشتن بهتر از ده خطا نیست.) اگر ما فقط آزمون‌های t بین همه جفت میانگین‌ها را اجرا کنیم، میزان خطای خانوادگی ممکن است بطور غیرقابل قبولی بزرگ شود. ما نیاز داریم شرایطی را برای جلوگیری از این اتفاق در نظر بگیریم. این چیزی است که یک آزمون t محافظت شده انجام می‌دهد با این اقتضای ساده که تا F کلی تحلیل واریانس معنادار نباشد هیچ آزمونی نمی‌تواند اجرا شود. برای مشاهده این که چرا این مرحله ساده به کار می‌رود توجه شما را به مثال‌های زیر جلب می‌کنیم.

فرض کنید که ما تنها دو میانگین داریم و فرض صفر درست باشد. احتمال انجام یک خطای نوع I به اندازه میزان این احتمال است که F اصلی بطور شانس معنادار باشد که به اندازه‌ی ۰/۰۵ است. اگر

آن F معنادار باشد ما از قبل خطای نوع I خود را انجام داده‌ایم، و حتی اگر پیش‌تر برویم و یک آزمون t اجرا کنیم، ما شرایط را بدتر خواهیم کرد. اگر F معنادار نبود ما نمی‌توانستیم t محافظت شده را اجرا کنیم و بنابراین میزان خطا افزایش نمی‌یافت^۱. بنابراین برای دو میانگین میزان خطای خانوادگی 0.05 است.

حالا فرض کنید که ما سه میانگین داشته باشیم. ابتدا فرض کنید که فرض صفر کلی درست است- یعنی فرض کنید که $\mu_1 = \mu_2 = \mu_3$. نخست ما F کلی را اجرا می‌کنیم. احتمال یافتن تفاوت معنادار (که یک خطای نوع I خواهد بود چون H_0 درست است) 0.05 است، و این اولین خطای نوع I ما خواهد بود از "دست کم یک مورد" که خطای خانوادگی به ما اجازه می‌دهد. اگر F معنادار نباشد ما در همین‌جا متوقف می‌شویم و شانس بیشتری برای انجام خطای نوع I نداریم. به عبارت دیگر وقتی فرض صفر کلی درست باشد، احتمال انجام دست کم یک خطای نوع I با قاعده 0.05 ما محدود می‌شود، همان چیزی که ما می‌خواهیم. حالا فرض کنید که فرض صفر نادرست باشد اما فقط یک میانگین با دو میانگین دیگر تفاوت داشته باشد (مثلا، $\mu_1 = \mu_2 \neq \mu_3$). آنگاه، چون فرض صفر نادرست است، امکان انجام یک خطای نوع I با آزمون F ما وجود ندارد. اگر ما یک F معنادار داشته باشیم، که امید داریم چنین باشد، ما می‌توانیم پیش‌تر رفته و مثلا هر جفت میانگین را بیازماییم- گروه ۱ در مقابل گروه ۲، گروه ۱ در مقابل گروه ۳، و گروه ۲ در برابر گروه ۳. اما فقط یکی از این آزمون‌ها هست که برای آن فرض صفر درست است و بنابراین تنها یک شانس برای خطای نوع I وجود دارد. پس در اینجا هم احتمال دست کم یک خطای نوع I فقط 0.05 است. در نهایت، فرض کنید که همه میانگین‌ها با هم متفاوت باشند. در اینجا امکان خطای نوع I وجود ندارد، زیرا فرض صفر درستی وجود ندارد که به اشتباه نادرست خوانده شود. از بررسی این احتمالات ما می‌توانیم نتیجه بگیریم که با سه میانگین میزان خطای خانوادگی حداکثر 0.05 است.

در حالتی که با چهار میانگین سرو کار داریم نیز توضیح مشابهی وجود دارد. اما در این حالت امکان دارد که بیش از یک فرض صفر درست وجود داشته باشد. برای مثال، ممکن است گروه ۱ و ۲ و گروه ۳ و ۴ برابر باشند. یا، گروه ۱، ۲ و ۳ برابر باشند اما با گروه ۴ فرق کنند. در این حالت حداکثر میزان احتمال خطای خانوادگی بیش از 0.05 خواهد بود، اما باز هم قابل توجیه است، بخصوص اگر شما در نظر نداشته باشید همه مقایسه‌های ممکن را صورت دهید. دچار خطای خانوادگی به میزان 0.10 یا 0.15 شدن به خوبی خطای 0.05 نیست. اما این پایان کار نیست. بسادگی با تقاضای F کلی معنادار قبل از اجرای مقایسه‌های چندگانه (که محافظت از آن بر می‌آید) بطور شگفت‌انگیزی در کنترل میزان خطای

۲. در حالت دو میانگین، آزمون‌های t و تحلیل واریانس معادل هم هستند، اما در اینجا این موضوع مهم نیست. منظور این است که اگر یک خطای نوع I در F وجود دارد ما در حال حاضر یک خطا داریم، و اگر F معنادار نباشد ما نمی‌توانیم آزمون را پیش‌تر ببریم.

خانوادگی دست کم وقتی که ما فقط تعداد کمی گروه داریم مؤثر است. به این دلیل ما در این فصل بر t محافظت شده تأکید نمودیم. در صورتی که تعداد گروه‌های شما نسبتاً کم باشند کنترل خطای خانوادگی کار مناسبی است، با این وجود آزمون‌ی است که بسادگی می‌توانید بکار ببرید و از توان قابل قبولی برخوردار است.

-آزمون شفه

آزمون t را همه نمی‌پسندند، و افرادی مایلند یک آزمون مقایسه چندگانه دیگر ارائه شود تا بر میزان خطای خانوادگی کنترل بیشتری اعمال کنند. ما **آزمون شفه**^۱ را برگزیدیم چون کاربرد آن آسان است و برای افرادی که کنترل دقیق‌تری می‌خواهند آشناست. نحوه ارائه آن در اینجا کمی متفاوت از سایر کتاب‌ها است، اما به لحاظ جبری یکسان است و بخوبی با آنچه پیش‌تر در باره محاسبه t محافظت شده آموختید تطبیق می‌یابد.

اول از همه، ما این ضرورت را که باید F کلی تحلیل واریانس معنادار باشد برای آزمون شفه حذف می‌کنیم. می‌توان مقایسه‌های متعددی با آزمون شفه انجام داد بدون توجه به اینکه F کلی معنادار باشد.

سپس شما هر تعداد مقایسه را که می‌خواهید بین جفت میانگین‌ها اجرا می‌کنید دقیقاً همانطور که با t محافظت شده انجام می‌دادید. به عبارت دیگر شما فرمول مشابهی را به کار می‌برید. تنها تفاوت این است که وقتی شما t را محاسبه می‌کردید، آن را به توان می‌رساندید و با آن به صورت یک مقدار مجاز برای F عمل می‌کردید. (وقتی ما دو گروه را مقایسه می‌کنیم، t^2 و F یکسان خواهند بود.)

در آخر شما F را با مقدار بحرانی F از جدول مربوط به آن در پیوست کتاب مقایسه می‌کنید. اما اینجا جایی است که تفاوت بروز می‌کند. در شرایط عادی تفاوتی ایجاد نمی‌شود اگر شما t را با $t_{.05}$ در درجات آزادی ۱ و $k(n-1)$ مقایسه کنید یا F را با $F_{.05}$ در درجات آزادی ۱ و $k(n-1)$ مقایسه کنید. نتیجه یکسان خواهد بود. اما ما نمی‌خواهیم F را با مقدار بحرانی F با ۱ و $k(n-1)$ df مقایسه کنیم. به جای آن ما می‌خواهیم به مقایسه F خود با $(k-1)$ بار F روی $(k-1)$ و $k(n-1)$ df بپردازیم. به عبارت دیگر ما داریم همان کاری را می‌کنیم که با t محافظت شده می‌کردیم (به جز ضرورت معناداری F کلی)، اما حالا مقدار بحرانی را تغییر می‌دهیم تا آزمون را بیشتر محافظه‌کارانه سازیم.

در انتهای این فصل یک مثال کار شده با استفاده از آزمون شفه را خواهیم داشت. یک نکته مهمی که باید در نظر داشت این است که شفه همچنین با پایین نگه داشتن میزان خطای خانوادگی، آزمون خود را بیشتر محافظه‌کارانه کرده است. منظور این است که با دشوارتر ساختن رد هر فرض صفری- حتی فرض‌های نادرست- رد کردن فرض صفر درست را سخت‌تر نموده است.

۱۶-۶ نقض کردن پیش فرض‌ها

چنان‌که دیدیم تحلیل واریانس براساس پیش فرض‌های نرمال بودن و تجانس واریانس است. اما در عمل تحلیل واریانس یک شیوه آماری خیلی قوی است و پیش فرض‌ها می‌توانند در بسیاری اوقات از سوی اثرات نسبتاً جزئی نقض شوند.

بطور کلی اگر بتوان فرض نمود که جمعیت‌ها به شکل متوازن یا دست‌کم از لحاظ شکل شبیه به هم (مثلاً همه دارای چولگی منفی) هستند و اگر بزرگترین واریانس بیش از چهار یا پنج برابر کوچک‌ترین واریانس نباشد، تحلیل واریانس به احتمال زیاد معتبر است. (بعضی معتقدند حتی برای واریانس‌های دارای تفاوت بیشتر هم معتبر است.) باید دقت کرد که ناهمگن بودن واریانس و نابرابر بودن اندازه نمونه‌ها با هم اشتباه نشوند. چنانچه شواهد پیش‌بینی کننده نابرابری واریانس‌ها باشد تا می‌توانید تلاش کنید که اندازه نمونه‌ها را تا حد ممکن برابر بگیرید. این بخصوص زمانی که می‌خواهید مجموعه‌ای از مقایسه‌های چندگانه انجام دهید درست است.

برای شرایطی که پیش فرض‌های زیربنایی تحلیل واریانس شدیداً نقض می‌شود، شیوه‌های جایگزینی برای دستکاری تحلیل وجود دارد. یک نوع از این شیوه‌ها شامل تغییر شکل داده‌ها (مثل تبدیل X به $\sqrt{X}$) و سپس اجرای آزمون‌های آماری استاندارد روی داده‌های تغییر شکل یافته است. شیوه‌های دیگر شامل استفاده از آزمون‌های کاملاً متفاوت است که در فصل ۲۰ روی آنها مقداری بحث خواهد شد.

۱۶-۷ اندازه اثر

اینکه ما یک تفاوت معنادار بین میانگین‌های عملکردی بدست می‌آوریم به این معنی نیست که تفاوت‌ها بزرگ یا مهم هستند. تفاوت‌های واقعی بسیاری وجود دارند که ناچیز هستند. هیچ شیوه آماری نمی‌تواند به ما بگوید که آیا یک تفاوت بدون توجه به بزرگی آن، از اهمیت عملی نسبت به موارد دیگر برخوردار است. اما شیوه‌هایی وجود دارند که به ما در این زمینه کمک می‌کنند.

یکی از ساده‌ترین سنجش‌های اندازه‌ی اثر، **مجذور اتا** (η^2) خوانده می‌شود. در حالی که مجذور اتا یک اندازه سوگیرانه است (از این لحاظ که برآورد بالاتر از مقداری بدست می‌دهد که در صورت امکان از جمعیت بدست می‌آوریم)، محاسبه آن خیلی ساده است و از این رو، مناسب است که به عنوان اولین برآورد مورد بحث قرار گیرد. در هر تحلیل واریانس، SS به ما می‌گوید چه مقدار تغییرپذیری کلی در داده‌ها وجود دارد. بخشی از این تغییرپذیری ناشی از این واقعیت است که گروه‌های مختلف آزمودنی‌ها به صورت متفاوتی عمل کرده‌اند و بنابراین نمرات متفاوتی گرفته‌اند، و بخشی از آن تنها ناشی از خطای تصادفی است - تفاوت‌های میان افرادی که در معرض عملکرد آزمایشی یکسانی قرار گرفته‌اند. تفاوت‌های

مهم تفاوت‌های بین نمراتی است که بتوان آنها را به عملکرد آزمایشی یا اثرات گروهی نسبت داد، و با گروهی SS اندازه‌گیری می‌شوند. اگر ما این نسبت را حساب کنیم

$$\eta^2 = \frac{SS_{\text{گروهی}}}{SS_{\text{کل}}}$$

می‌توانیم بگوییم که چند درصد از تغییرپذیری بین مشاهدات را می‌توان به حساب اثرات گروهی گذاشت^۱. برای داده‌های سازگاری مادرانه

$$\eta^2 = \frac{226.932}{2072.925} = 0.11$$

بنابراین می‌توانیم نتیجه‌گیری کنیم که تغییرپذیری نمرات سازگاری را می‌توان به عضویت گروهی نسبت داد. هر چند در ابتدا ممکن است درصد کوچکی به نظر بیاید، اما اگر به سطح بالای تغییرپذیری در میان مادرانی که می‌شناسید فکر نکنید، تبیین حتی ۱۰٪ آن هم یک موفقیت ارزشمند است.

هر چند η^2 یک اندازه ساده و سریع برای محاسبه است و در هنگام خواندن گزارش‌های تحقیق می‌توانید آن را به صورت ذهنی حساب کنید، اما یک آماره‌ی سوگیرانه است. این آماره متمایل به برآورد بیشتر مقدار واقعی در جمعیت است. یک برآورد کمتر سوگیرانه از طریق آماره دیگری بنام مجذور امگا (ω^2) بدست می‌آید. برای تحلیل واریانس مورد بحث در این فصل می‌توانیم آن را چنین تعریف کنیم

$$\omega^2 = \frac{SS_{\text{گروهی}} - (k-1)MS_{\text{خطا}}}{SS_{\text{کل}} + MS_{\text{خطا}}}$$

که k نشانگر تعداد گروه‌ها است. برای مثال ما

$$\omega^2 = \frac{226.932 - (3-1)(20.511)}{2072.925 + 20.511} = \frac{185.910}{2093.436} = 0.089$$

این مقدار قدری کمتر از مقداری است که برای η^2 بدست آوردیم. به هر حال هنوز هم گویای این است که ما تقریباً ۹٪ تغییرپذیری را تبیین کرده‌ایم.

۸-۱۶ مثال کار شده پایانی

مثال زیر یک تحلیل واریانس یک طرفه را با تعداد نمونه نابرابر نشان می‌دهد. این مثال همچنین استفاده از آزمون شفه را نشان می‌دهد.

هسته اکوبنس یک ساختار بخش پیشین مغز است که در فعالیت‌های حرکتی موش‌ها دخالت دارد. ارائه دوز پایین تتراهیدورکانابینول (THC ماده اصلی ماری‌جوانا) در افزایش فعالیت حرکتی تأثیر

۱. اگر شما یک متغیر (X) را با وارد کردن ۱ برای همه آزمودنی‌های گروه ۱ و یک ۲ برای همه آزمودنی‌های گروه ۲ ایجاد کنید، و اگر شما Y را متغیر وابسته بگیرید، آنگاه مجذور ضریب همبستگی (r^2) بین X و Y برابر با η^2 خواهد بود.

دارد، در حالی که دوز بالای آن منجر به کاهش فعالیت می‌شود. کونتی و موستی (۱۹۸۴) برای بررسی اینکه آیا THC در درون هسته اکومبنس برای ایجاد یک فعالیت عمل می‌کند، دوزهای متفاوتی از داروی بی‌اثر یا ۰/۱، ۰/۵، ۱ یا ۲ میکروگرم از THC را مستقیماً به هسته اکومبنس موش‌ها تزریق کردند. محققان تغییرات سطح فعالیت حیوان را پس از تزریق ثبت نمودند. انتظار می‌رفت که در دوزهای کم نسبت به زیاد، فعالیت افزایش پیدا کند. داده‌ها در جدول ۷-۱۶ مقدار تغییر (کاهش) در هر حیوان را نشان می‌دهد.

ابتدا ما فرض صفر را طرح می‌کنیم. فرض صفر فرضی است که می‌گوید همه این نمونه‌ها از جمعیت‌هایی با میانگین یکسان بدست آمده‌اند. به عبارت دیگر $\mu_1 = \mu_0 = \mu_2 = \mu_3 = \mu_4 = \mu_5$. به شیوه قبل، ما این فرض صفر را با یک سطح معناداری $\alpha = 0.05$ می‌آزماییم. سپس ما تحلیل واریانس کلی را اجرا می‌کنیم، که با محاسبه مجموع‌های مجزورات شروع می‌شود.

$$\begin{aligned}
 SS_{\text{کل}} &= \sum X^2 - \frac{G^2}{N} = (30^2 + 27^2 + \dots + 52^2 - \frac{2160^2}{47}) \\
 &= 113556 - 992689/0.85 = 14287/91 \\
 SS_{\text{گروهی}} &= \frac{\sum T_j^2}{n_j} - \frac{G^2}{N} = \frac{340^2}{10} + \frac{508^2}{10} + \frac{543^2}{9} + \frac{388^2}{8} + \frac{381^2}{10} - \frac{2160^2}{47} \\
 &= 103461/50 - 992689/0.85 = 4193/41 \\
 SS_{\text{خطا}} &= SS_{\text{کل}} - SS_{\text{گروهی}} = 14287/91 - 4193/41 = 10094/50
 \end{aligned}$$

منبع	df	SS	MS	F
گروه‌ها	۴	۴۱۹۳/۴۱	۱۰۴۸/۳۵	۴/۳۶
خطا	۴۲	۱۰۰۹۴/۵۰	۲۴۰/۳۵	
کل	۴۶	۱۴۲۸۷/۹۱		

جدول ۷-۱۶ داده‌های مطالعه کونتی و موستی (۱۹۸۴)

گروه				
داروی بی‌اثر	۰/۱ μg	۰/۵ μg	۱ μg	۲ μg
۳۰	۶۰	۷۱	۳۳	۳۶
۲۷	۴۲	۵۰	۷۸	۲۷
۵۲	۴۸	۳۸	۷۱	۶۰

	۵۱	۵۸	۵۹	۵۲	۳۸	
	۲۹	۳۵	۶۵	۲۸	۲۰	
	۳۴	۳۵	۵۸	۹۳	۲۶	
	۲۴	۴۶	۷۴	۳۲	۸	
	۱۷	۳۲	۶۷	۴۶	۴۱	
	۵۰		۶۱	۶۳	۴۹	
	۵۳			۴۴	۴۹	
کل	۳۸۱	۳۸۸	۵۴۳	۵۰۸	۳۴۰	۲۱۶۰ = G
میانگین	۱۰/۳۸	۴۸/۵۰	۶۰/۳۳	۵۰/۸۰	۳۴/۰۰	۴۵/۹۶
n	۱۰	۸	۹	۱۰	۱۰	۴۷

در آخر ما $F = ۴/۳۶$ را با مقدار بحرانی جدول F در بخش پیوست مقایسه می‌کنیم. ما ۴ df گروهی و ۴۲ df خطا داریم. در صورتی که درجات آزادی را به ۴۰ برای مخرج سر راست کنیم مقدار بحرانی در جدول ۲/۶۱ است. چون مقداری که ما بدست آورده‌ایم از ۲/۶۱ بزرگ‌تر است، فرض صفر را رد می‌کنیم و تفاوت‌های سطح فعالیت در میان پنج گروه را ناشی از دوز THC فرض می‌گیریم.

فرض آزمایشی پیش‌بینی می‌کرد که گروه‌های با دوز کم نسبت به گروه‌های با دوز زیاد افزایش بیشتری در فعالیت نشان خواهند داد. بنابراین ما ممکن است مایل باشیم تا گروه ۰/۵ میلی گرم را با گروه ۲ میلی گرم مقایسه کنیم. همچنین مقایسه گروه ۲ میلی گرم با گروهی که داروی بی‌اثر را دریافت کرده می‌تواند جالب باشد تا ببینیم آیا این دو گروه باهم تفاوت دارند. ما هر دو مقایسه را با آزمون شفه انجام خواهیم داد. چنان‌که در متن بیان شد، این آزمون را ما ابتدا با اجرای آزمون‌های t بین گروه‌ها همانند آنچه با آزمون t محافظت شده صورت گرفت اجرا می‌کنیم.

مقایسه گروه ۳ و ۵ ($۵ \mu g$) در برابر ($۲ \mu g$)

$$\begin{aligned}
 t &= \frac{\bar{X}_3 - \bar{X}_5}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_3} + \frac{1}{n_5} \right)}} \\
 &= \frac{60.33 - 38.10}{\sqrt{240.35 \left(\frac{1}{9} + \frac{1}{10} \right)}} \\
 &= \frac{22.23}{\sqrt{240.35(0.2111)}} = \frac{22.23}{\sqrt{50.74}} = \frac{22.23}{7.12} = 3.12 \\
 F &= t^2 = 3/۱۲^2 = ۹/۷۳
 \end{aligned}$$

مقایسه گروه ۱ و ۵ (داروی بی‌اثر در برابر $۲ \mu g$)

$$\begin{aligned}
 t &= \frac{\bar{X}_3 - \bar{X}_5}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_1} + \frac{1}{n_5} \right)}} \\
 &= \frac{34.00 - 38.10}{\sqrt{240.35 \left(\frac{1}{10} + \frac{1}{10} \right)}} \\
 &= \frac{-4.10}{\sqrt{240.35(0.20)}} = \frac{-4.10}{\sqrt{48.07}} = \frac{-4.10}{6.93} = 0.59 \\
 F &= t^2 = 0.59^2 = 0.35
 \end{aligned}$$

توجه داشته باشید که برای آزمون‌های t مقدار حاصل شده را به توان دو می‌رسانیم تا یک F بدست آید. به این دو مقدار F می‌توان در جدول F مربوط به آن رجوع کرد. برای این کار ابتدا لازم است مقدار بحرانی F روی درجات آزادی ۴ و ۴۲ را بدانیم. این مقدار (اگر ما درجات آزادی مخرج را به ۴۰ سر راست کنیم) $2/16$ است. سپس ما این مقدار بحرانی F را در $k-1 = 4$ ضرب می‌کنیم تا $10/44$ بدست آید. هر مقدار F که از این مقدار بیشتر باشد معنادار به حساب می‌آید. چون دو مقدار $9/73$ و $0/35$ هستند، ما فرض صفر را در هیچ مورد رد نمی‌کنیم.

ممکن است تعجب کنید چرا ما فرض صفر را برای هیچ کدام از این مقایسه‌ها رد نمی‌کنیم، هر چند برای مقایسه گروه‌های ۳ و ۵ با یک آزمون محافظت شده t آن را رد کردیم. دلیل اصلی این است که آزمون شفه یک آزمون محافظه کارانه‌تر است. در مرتبه نخست، این کار بهتری است نسبت به اینکه از t محافظت شده با داشتن میزان خطای خانوادگی، در این حالت در سطح $\alpha = 0/05$ ، استفاده کنیم. اما از این گذشته هنوز هم این آزمون خیلی محافظه کارانه‌ای است. به همین دلیل ما t محافظت شده را زمانی که تنها تعداد کمی گروه وجود دارد ترجیح می‌دهیم. ما ترجیح می‌دهیم میزان خطای خانوادگی بیشتر باشد تا اینکه با چنین آزمون محافظه کارانه‌ای به دشواری یک تفاوت معنادار پیدا کنیم. البته خیلی افراد با این نظر موافق نیستند.

۹-۱۶ خلاصه

تحلیل واریانس یکی از ابزارهای آماری بسیار قدرمند است. در این فصل ما با بررسی منطق زیربنایی این تحلیل شروع کردیم و سپس به محاسبات آن پرداختیم. بعد از توجه به محاسبات برای اندازه نمونه‌های برابر، مسأله را به نمونه‌های با اندازه نابرابر بردیم و دیدیم که برای تحلیل واریانس یک طرفه ما تنها به تغییرات جزئی در فرمول نیاز داریم. سپس ما به مسأله تفاوت‌های مجزای گروه‌ها با شیوه‌های مقایسه‌ای چندگانه پرداختیم. بعد اثرات نافرمانی از پیش فرض‌های زیربنایی آزمون را مورد بحث قرار دادیم، به مسأله برآورد مقدار اثرات آزمایشی پرداختیم. برخی از مهمترین مطالب این فصل عبارتند از:

○ تحلیل واریانس (آنوا)	○ $df_{\text{خطا}}$
○ آنوای یک طرفه	○ روش‌های مقایسه چندگانه
○ درون گروهی MS (خطا MS)	○ t محافظت شده (آزمون کمترین تفاوت معنادار)
○ بین گروهی MS (گروهی MS)	○ میزان خطای خانوادگی
○ جمع کل (G)	○ آزمون شفه
○ $SS_{\text{کل}}$	○ اندازه‌ی اثر
○ $SS_{\text{گروهی}}$	○ مجذور اتا (η^2)
○ $SS_{\text{خطا}}$	○ مذور امگا (ω^2)
○ $df_{\text{کل}}$	
○ $df_{\text{گروهی}}$	

۱۰-۱۶ تمرین

۱۶-۱ برای بررسی رفتار مادران موش‌های آزمایشگاهی ما بچه‌موش‌ها را از مادر جدا کردیم و زمان لازم (به ثانیه) برای بازگشت مادر را ثبت کردیم. این مطالعه را با بچه موش‌های ۵، ۲۰ و ۳۵ روزه اجرا کردیم چون می‌خواستیم ببینیم آیا زمان بازگشت با سن بچه‌موش‌ها فرق می‌کند. داده‌ها در زیر آمده است، که شامل ۵ بچه‌موش در هر گروه است.

۱۸	۲۰	۱۵	۲۵	۱۰	۱۵	روزه ۵
۲۰	۲۳	۲۵	۲۰	۱۵	۳۰	روزه ۲۰
۴۰	۴۵	۴۳	۵۰	۳۵	۴۰	روزه ۳۵

یک تحلیل واریانس یک طرفه با $\alpha = 0.05$ اجرا کنید.

۱۶-۲ یک آزمون t محافظت شده را با داده‌های تمرین ۱۶-۱ اجرا کنید تا تفاوت بین بچه موش‌های ۵ و ۲۰ روزه و تفاوت بین بچه‌موش‌های ۲۰ و ۳۵ روزه را ارزیابی کنید.

۱۶-۳ فرض کنید شما داده‌هایی را برای پاسخ به این سؤال دارید که آیا شخص پرداخت کننده صورت حساب غذاخوری از مهمان او غذای ارزان‌تری را سفارش می‌دهد. به منظور پرهیز از ابهام‌آمیز شدن موضوع شما در نظر می‌گیرید که تنها به جفت‌های هم جنس بپردازید و فقط قیمت غذای یک شخص را در هر جفت ثبت کنید. آزمودنی‌ها بر اساس اینکه آیا آزمودنی عملاً صورت حساب را بر می‌دارد یا نه، به گروه‌ها گمارده شدند. متغیر وابسته هزینه غذای اصلی بود.

۹/۲۵	۹/۰۰	۱۰/۲۵	۸/۷۵	۹/۵۰	میزبان
۱۲/۲۵	۱۰/۵۰	۸/۵۰	۹/۵۰	۱۰/۷۵	مهمان

(الف) تحلیل واریانس را روی این دو گروه اجرا کنید.

(ب) یک آزمون t مستقل روی همان داده‌ها اجرا کنید، t را به توان دو برسانید و نتایج را مقایسه کنید. (توجه: این رابطه بین F و t فقط زمانی برقرار است که دو گروه داشته باشیم.)

۱۶-۴ می‌توان پیش‌بینی کرد که رفتار خرید مشتریان تحت تأثیر موقعیت یک محصول در فروشگاه فرق می‌کند، حتی اگر آن محصول نام تجاری پرآوازه‌ای داشته باشد. بنابراین ما به خرید نام‌های تجاری شناخته شده و ناشناخته سیگار وقتی در قفسه‌های معمول فروشگاه قرار دارند و وقتی نزدیک صندوق قرار دارند توجه می‌کنیم. متغیر وابسته تعداد بسته‌های هر نام تجاری فروخته شده در روز است. (البته فعلاً از این موضوع که با شیوه‌ای که بعداً خواهیم آموخت این تحلیل را بهتر می‌توان انجام داد چشم‌پوشی می‌کنیم.)

نام شناخته شده/موقعیت معمولی	۱۵	۲۳	۱۸	۱۶	۲۵	۲۹	۱۷
نام شناخته شده/موقعیت ویژه	۲۴	۱۴	۱۵	۱۹	۳۰	۲۶	۱۸
نام ناشناخته/موقعیت معمولی	۱۰	۵	۸	۱۲	۱۳	۶	۱۰
نام ناشناخته / موقعیت ویژه	۱۵	۱۳	۱۰	۱۷	۱۸	۱۱	۱۵

(الف) یک تحلیل واریانس یک طرفه اجرا کنید.

(ب) حالا یک تحلیل واریانس روی گروه ۱ و ۳ به صورت ترکیبی در برابر گروه‌های ۲ و ۴ ترکیبی اجرا نمایید.

۱۶-۵ به تمرین ۱-۱۶ رجوع کنید. فرض کنید به دلایلی خارج از کنترل ما، داده‌ها برای آخرین گروه ۵ روزه‌ها را نتوانیم استفاده کنیم، همچنین نتوانیم داده‌های دو بچه‌موش آخر را در گروه ۳۵ روزه استفاده کنیم. برای باقیمانده داده‌ها تحلیل واریانس را دوباره انجام دهید.

۱۶-۶ به تمرین ۳-۱۶ رجوع کنید. فرض کنید که ما برای گروه میزبان دو نقطه داده اضافه جمع‌آوری کردیم. داده‌ها حالا به شکل زیر است:

میزبان	۹/۵۰	۸/۷۵	۱۰/۲۵	۹/۰۰	۹/۲۵	۱۱/۷۵	۹/۰۰
مهمان	۱۰/۷۵	۹/۵۰	۸/۵۰	۱۰/۵۰	۱۲/۲۵		

(الف) تحلیل واریانس را روی این دو گروه اجرا کنید.

(ب) یک آزمون t مستقل بدون جمع واریانس انجام دهید.

(ج) یک آزمون t مستقل بعد از جمع واریانس انجام دهید

(د) کدام یک از اندازه‌های t هم ارز F در بخش (الف) است (پس از مجذور کردن)؟

۱۶-۷ برای داده‌های تمرین ۳-۱۶ η^2 و ω^2 حساب کنید.

۱۶-۸ بعضی کلمات در یک متن عادی برای فهم مطلب از اهمیت خاصی برخوردارند، در حالی که کلمات دیگر چندان مهم نیستند. اگر فرض شود که افرادی که خواننده خوبی هستند در اصل کلمات مهم را می‌خوانند، آنگاه اگر این کلمات برجسته شوند ولی کلمات دیگر نشوند، می‌توان انتظار داشت که

متن با سرعت بیشتری خوانده شود. ما سه گروه را تعیین می‌کنیم که یک متن مشابه را می‌خوانند. برای گروه ۱ کلمه‌ای برجسته نشده است. برای گروه ۲ یک مجموعه تصادفی از کلمات برجسته می‌شوند. و برای گروه ۳ کلمات مهم برجسته می‌شوند. متغیر وابسته زمان خواندن آن متن (به ثانیه) است.

n	میانگین	انحراف معیار	
۱۰	۳۰/۲	۶/۲۱	گروه ۱
۱۰	۳۸/۳	۷/۵۵	گروه ۲
۱۰	۲۵/۶	۵/۷۵	گروه ۳

(الف) تحلیل را انجام دهید و نتیجه‌ای که قابل قبول به نظر می‌آید را استخراج کنید.

(ب) در طرح این آزمایش دست‌کم یک کاستی را در رابطه با این فرض که خواننده‌های خوب دنبال کلمات اصلی می‌گردند پیدا کنید.

(ج) رد فرض صفر در اینجا بیان‌کننده چیست؟

۹-۱۶ اگر در تمرین ۸-۱۶ میانگین و انحراف معیار یکسان داشته باشیم ولی n ها به جای ۱۰ هر کدام ۵ باشند چه نتیجه‌ای خواهیم گرفت؟

۱۰-۱۶ در تمرین ۸-۱۶ از آزمون‌های محافظت شده t برای روشن نمودن معنای F معنادار استفاده کنید.

۱۱-۱۶ داده‌های تمرین ۹-۱۶ نیز یک F معنادار تولید می‌نماید. آیا شما به اثر آن کم و بیش اعتماد دارید؟ چرا؟

۱۲-۱۶ دارلی و لاتان (۱۹۶۸) سرعت کمک‌رسانی آزمودنی‌ها به یک فرد دچار مشکل را ثبت نمودند. آزمودنی‌ها فکر می‌کردند که یا با یک شخص تنها هستند (گروه ۱، $n = ۱۳$)، یا شخص دیگری در آنجا هست (گروه ۲، $n = ۲۶$)، یا چهار نفر دیگر در آنجا هستند (گروه ۳، $n = ۱۳$). متغیر وابسته سرعت بود ($= ۱۰۰ \times \text{زمان}$). نمرات میانگین سرعت برای سه گروه به ترتیب ۰/۸۷، ۰/۷۲ و ۰/۵۱ بود. $MS_{\text{خطا}}$ برابر ۰/۵۳ بود. تحلیل واریانس را بازسازی کنید. (راهنمایی: ابتدا حاصل جمع‌های گروهی را بدست آورید.) چه نتیجه‌ای می‌گیرید؟

۱۳-۱۶ با استفاده از داده‌های تمرین ۱-۱۶، $MS_{\text{خطا}}$ را به طور مستقیم نه از طریق کسر کردن محاسبه کنید و نشان دهید جواب یکسان است.

۱۴-۱۶ برای داده‌های تمرین ۱-۱۶ آزمون شفه را بکار ببرید و جواب خود را با جواب تمرین ۲-۱۶ مقایسه کنید.

۱۵-۱۶ برای داده‌های تمرین ۸-۱۶ آزمون شفه را بکار بگیرید. چه نتیجه‌ای خواهید گرفت؟ این در مقایسه با جواب ۱۰-۱۶ چگونه خواهد بود؟

فصل ۱۷

تحلیل واریانس عاملی

در فصل ۱۶ ما با تحلیل واریانس یک طرفه آشنا شدیم که یک طرح آزمایشی است که تنها یک متغیر مستقل دارد. در این فصل ما در جهت بسط تحلیل واریانس پیش می‌رویم تا طرح‌های آزمایشی را پوشش دهیم که شامل دو متغیر مستقل یا بیشتر هستند. به منظور سادگی ما کار خود را فقط به طرح‌های با دو متغیر مستقل محدود می‌سازیم، گرچه بسط آن به طرح‌های پیچیده‌تر کاملاً ساده است.

به تحقیق ساده‌ای توجه کنید که طی آن ما داده‌هایی پیرامون یک متغیر وابسته با عنوان اعتماد به نفس از سه گروه مختلف از نوجوانان جمع‌آوری کرده‌ایم. دانش‌آموزان در گروه A در اصل در فعالیت‌های ورزشی شرکت داشتند، دانش‌آموزان گروه S در اصل در فعالیت‌های اجتماعی شرکت داشتند، و دانش‌آموزان در گروه U در هیچ فعالیتی شرکت نداشتند (غیر درگیر). این یک طرح یک طرفه ساده است، زیرا ما تنها یک متغیر مستقل داریم (گروه). اکنون فرض کنید که ما این طرح را با جمع‌آوری داده‌ها از دانش‌آموزان مقطع اول و دوم دبیرستان گسترش دهیم. این مطالعه اکنون شامل دو متغیر مستقل است، فعالیت طراحی شده (بسه سطح: ورزشی، اجتماعی و غیر درگیر) و مقطع (با دو سطح: اول و دوم دبیرستان). در این متن متغیرهای مستقل اغلب به عنوان عامل‌ها منظور می‌شوند، طوری که ما می‌توانیم از عامل فعالیت و عامل مقطع نام ببریم. آزمایشی که توضیح داده شد را **طرح عاملی دو طرفه** گویند.

اگر ما می‌خواستیم این آزمایش را بیشتر هم گسترش دهیم می‌توانستیم آزمودنی‌ها را نیز به دختر و پسر دسته‌بندی کنیم. آنگاه ما یک طرح معروف به عاملی سه طرفه خواهیم داشت با عامل‌های فعالیت، مقطع و جنسیت.

وقتی ما یک طرح آزمایشی داریم که هر سطحی از هر متغیر با هر سطح از متغیر دیگر جفت می‌شود، ما طرحی داریم که **طرح عاملی^۱** نامیده می‌شود. به عبارت دیگر یک طرح عاملی طرحی است

1. factorial design

که در آن ما همه ترکیب‌های سطوح متغیرهای مستقل را شامل نموده‌ایم. یک مثال از طرح عاملی در جدول ۱-۱۷ نشان داده شده است. در این فصل ما خود را محدود می‌سازیم به طرح‌های عاملی که در آنها ترکیب‌های عملکرد متفاوت به گروه‌های مستقل آزمودنی‌ها تخصیص داده شده است. وقتی طرح تحقیق آزمودنی‌های مشابهی را در بیش از یک شرایط عملکردی یا ترکیبی از عملکردها فرا می‌خواند، صحبت از طرح‌های اندازه‌گیری مکرر است. بحث طرح‌های ساده اندازه‌گیری مکرر در فصل ۱۸ خواهد آمد. بنابراین اگر ما افراد را وقتی در مقطع اول دبیرستان هستند و دوباره همان دانش‌آموزان را وقتی در مقطع دوم دبیرستان هستند مورد سنجش قرار دهیم، ما یک طرح اندازه‌گیری مکرر خواهیم داشت و تحلیلی را که در این فصل توضیح می‌دهیم در مورد آنها نمی‌توانیم به کار ببریم.

۱-۱۷ بازنمایی یک طرح عاملی برای مطالعه اعتماد به نفس روی دانش‌آموزان دبیرستانی

فعالیت

مقطع	ورزشی	اجتماعی	غیردرگیر
اول دبیرستان	دانش‌آموزان مقطع اول ورزشی	دانش‌آموزان مقطع اول اجتماعی	دانش‌آموزان مقطع اول غیردرگیر
دوم دبیرستان	دانش‌آموزان مقطع دوم ورزشی	دانش‌آموزان مقطع دوم اجتماعی	دانش‌آموزان مقطع دوم غیردرگیر

طرح‌های عاملی دارای چندین امتیاز نسب به طرح‌های یک طرفه هستند. اول اینکه تعمیم‌پذیری بیشتر نتایج را امکان‌پذیر می‌سازند. به مطالعه فعالیت و مقطع توجه کنید. اگر ما بخواهیم یک آزمایش با تنها سه شرایط فعالیت اجرا کنیم، به احتمال زیاد تنها یک مقطع را به کار می‌گیریم، و بنابراین نتایج ما تنها در آن مقطع اعمال خواهد شد. وقتی ما یک طرح عاملی با سه شرایط فعالیت و دو مقطع تحصیلی به کار می‌بریم، نتایج ما در باره فعالیت در سطح مقاطع بطور متوسط تعیین می‌شود، در نتیجه امکان تفسیر گسترده‌تری از داده‌ها فراهم می‌آید. در صورت تمایل می‌توانیم همزمان داده‌ها را برای هر مقطع جداگانه نیز بیازماییم.

ویژگی مهم دیگر طرح‌های عاملی این است که امکان بررسی تعامل^۱ را نیز به ما می‌دهند. تعامل زمانی است که اثر یک متغیر به سطح معینی از متغیر دیگر بستگی دارد. مثلاً اگر وقتی به دانش‌آموزان مقطع دوم متوسطه نگاه می‌کنیم تفاوت قابل ملاحظه‌ای در اعتماد به نفس برای سطوح مختلف فعالیت

1. Interaction

وجود داشته باشد، اما وقتی به مقطع اول نگاه می‌کنیم تفاوت نباشد، در این صورت، اثر فعالیت وابسته به مقطع تحصیلی دانش‌آموزان است. بنابراین می‌گوییم فعالیت و مقطع با هم تعامل دارند. از سوی دیگر چنانچه هر تفاوتی که در میان گروه‌های فعالیت وجود دارد برای دانش‌آموزان دو مقطع یکسان باشد، در این صورت اثر فعالیت مستقل از مقطع است و تعاملی وجود ندارد.

سومین امتیاز طرح عاملی اقتصادی بودن آن از لحاظ آزمودنی‌ها است. از آنجاکه ما اثرات یک متغیر را در همه سطوح متغیر دیگر بطور متوسط پیدا می‌کنیم، یک طرح عاملی دو متغیری به تعداد آزمودنی کمتری نسبت به دو طرح یک طرفه (یکی روی فعالیت و یکی روی مقطع) برای همان میزان توان نیاز دارد. از سوی دیگر در صورتی که هر تفاوتی که در میان گروه‌های فعالیت وجود دارد، به همان اندازه برای دانش‌آموزان مقطع اول و دوم متوسطه هم هست، پس اثر فعالیت مستقل از مقطع است و تعاملی وجود ندارد. هر آزمودنی در یک زمان به یک سطح از دو متغیر مستقل تخصیص می‌یابد - طوری که اطلاعات از دو حالت به قیمت یکی بدست می‌آید. در اصل ما چیزی را رایگان بدست می‌آوریم که به هر حال جالب است.

همانطور که قبلاً گفته شد، طرح‌های عاملی بر اساس تعداد متغیرهای درگیر در آنها نامگذاری می‌شوند. یک طرح عاملی با دو متغیر مستقل یا عامل را طرح عاملی دو طرفه می‌نامند، و یک طرح با سه عامل را طرح عاملی سه طرفه گویند. یک روش دیگر نامگذاری بر اساس تعداد سطوح هر متغیر است. مطالعه اعتماد به نفس دو سطح شرایط مقطع و سه سطح فعالیت داشت. به این ترتیب آن را **طرح عاملی ۳ × ۲** می‌نامند، که عدد اول گویای تعداد ردیف‌ها و عدد دوم نشان‌دهنده تعداد ستون‌ها در این طرح است که در جدول ۱-۱۷ نشان داده شده است. یک مطالعه با سه متغیر، که دو تای آنها سه سطحی باشند و یکی چهار سطح داشته باشد طرح عاملی $4 \times 3 \times 3$ خوانده می‌شود. عبارت دو طرفه و 3×2 هر دو برای نامگذاری طرح‌ها معمول هستند.

در بخش بعدی ما تحلیل دو طرفه را مورد توجه قرار می‌دهیم. تحلیل‌های سطح بالاتر بسط‌هایی از طرح دو طرفه هستند، و بسیاری از مسائلی که ما مورد بحث قرار خواهیم داد بیشتر به شکل دو متغیری ساده بیان می‌شوند.

۱۷-۱ نشانگان

در این فصل نشانگان بطور ساده و به دور از ابهام بیان شده‌اند. نشانه‌هایی که بیشتر مورد استفاده قرار می‌گیرند در جدول ۲-۱۷ آمده است. نکته اولی که باید توجه داشت این است که نام عامل‌ها عموماً با حرف بزرگ اول آنها، و سطح هر عامل با حرف بزرگ و اندیس مناسب نشان داده می‌شود (مثلاً برای مقطع متوسطه، دانش‌آموزان مقطع اول متوسطه با C_1 و دانش‌آموزان مقطع دوم با C_2 نشان داده می‌شوند). تعداد سطوح هر عامل با یک حرف کوچک ناظر بر آن عامل نشانه‌گذاری می‌شود. بنابراین

فعالیت (A) دارای $a = 3$ سطح است در حالی که مقطع (C) دارای $c = 2$ سطح است. هر ترکیب خاصی از سطحی از یک عامل و یک سطح از عامل دیگر (مثلا دانش‌آموزان مقطع اول اجتماعی) یک خانه^۱ نامیده می‌شوند، و تعداد مشاهدات در هر خانه با n نشانه‌گذاری می‌شود. تعداد کل مشاهدات $N = acn$ است، زیرا $a \times c$ خانه وجود دارد، هر یک دارای n مشاهده هستند.

اندیس‌های i و j به عنوان نشانگان عمومی (غیر ویژه) برای سطح ردیف‌ها و ستون‌ها به کار برده می‌شوند. بنابراین خانه با اندیس ij خانه‌ای است در ردیف i ام و ستون j ام. مثلا خانه با اندیس ۲۳ مربوط به دانش‌آموزان مقطع اول متوسطه غیردرگیر است. (رجوع کنید به جدول ۱-۱۷). حاصل جمع‌های سطوح منفرد فعالیت با T_{Aj} نشان داده می‌شوند در حالی که حاصل جمع‌های مقطع با T_{Ci} نشان داده می‌شوند. اینجا دوباره اندیس A و C اشاره به نام متغیرها دارند، و T برای "کل" است. خانه‌های کل با T_{ij} نشانه‌گذاری می‌شوند و جمع کل (جمع همه نمرات) با G یا $\sum X$ نشان داده می‌شود. اندیس‌های غیر ضروری اغلب گیج کننده‌اند و سعی خواهیم کرد هر جا ممکن باشد آنها را حذف کنیم. نشانگانی که اینجا توضیح داده شد در بحث تحلیل واریانس مورد استفاده قرار خواهند گرفت، و لازم است، این نکات را از قبل به خاطر بسپارید. مزیت این سیستم این است که به آسانی به نمونه‌های دیگر تعمیم می‌یابد. بنابراین اگر ما یک طرح عاملی جنسیت $\times$ مواد مخدر داشته باشیم، باید روشن باشد که T_{D1} و T_{S2} به ترتیب، اشاره به حاصل جمع سطح اول متغیر مواد مخدر و سطح دوم متغیر جنسیت دارند.

جدول ۲-۱۷ طرح عاملی نشان‌دهنده چگونگی نشانگان

فعالیت

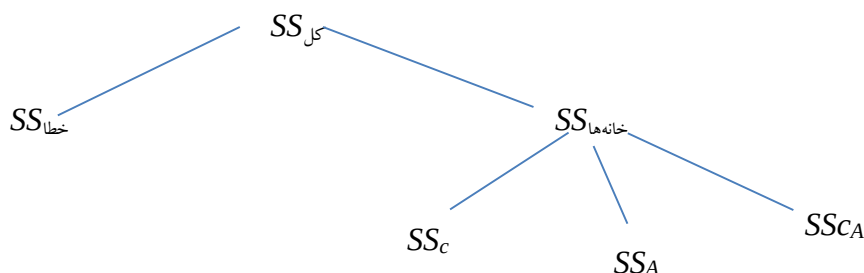
	A_3	A_2	A_1	مقطع
T_{C1}	T_{13}	T_{12}	T_{11}	C_1
T_{C2}	T_{23}	T_{22}	T_{21}	$2C$
$G = \sum X$	T_{A3}	T_{A2}	T_{A1}	

۲-۱۷ یک مثال - اعتماد به نفس دانش‌آموز

به عنوان نمونه ما یک طرح 3×2 را در نظر می‌گیریم که قبلا بحث کرده بودیم. داده‌ها و تحلیل در جدول ۳-۱۷ آمده است. میانگین‌های خانه‌ای صرفا برای اطلاع شما نشان داده شده است. متغیر وابسته رتبه‌های اعتماد به نفس هر دانش‌آموز است. مقیاس درجه‌بندی از ۰ تا ۲۵ بوده با نمرات بالاتر به عنوان اعتماد بیشتر. در هر خانه پنج مشاهده وجود دارد. قبل از هرگونه نتیجه‌گیری، نگاهی به میانگین‌های ستونی و ردیفی می‌اندازیم. این میانگین‌ها یک ایده کلی در باره آنچه از تحلیل انتظار داریم به ما می‌دهند.

از این میانگین‌ها می‌توانیم ببینیم که دانش‌آموزان مقطع دوم متوسطه (C_2)، بدون توجه به سطح فعالیت، به نظر می‌رسد اعتماد به نفس بیشتری نسبت به دانش‌آموزان مقطع اول (C_1) نشان می‌دهند، و اینکه اعتماد به نفس دانش‌آموزان غیر درگیر از همه کمتر است. علاوه بر این به نظر می‌رسد الگوی نتایج در بین خانه‌ها برای سه فعالیت وابسته به مقطع است، زیرا در مورد مقطع اول گروه اجتماعی تقریباً با گروه غیردرگیر برابر است، در حالی که در مورد مقطع دوم گروه‌های ورزشی و اجتماعی تقریباً برابر هستند. اینکه این نتیجه‌گیری‌ها برآمده از میانگین‌ها واقعی هستند یا نه چیزی است که تحلیل واریانس باید به ما بگوید.

در تحلیل واریانس یک طرفه ما مجموع مجذورات کل ($SS_{کل}$) را به تفاوت‌های بین‌گروهی ($SS_{گروهی}$) و تفاوت‌های درون‌گروهی ($SS_{خطا}$) تقسیم (دو نیمه) کردیم. در تحلیل دوطرفه ما همین کار را انجام می‌دهیم به استثنای اینکه ما اصطلاح "گروه" منفرد نداریم. چیزی که ما داریم شش خانه است، که نشان‌دهنده شش گروه مختلف از آزمودنی‌ها متشکل از افراد در فعالیت‌های گروهی متفاوت، مقاطع مختلف و ترکیب‌های فعالیت $\times$ مقاطع مختلف هستند. کاری که ما با شیوه دوطرفه انجام می‌دهیم نخست تقسیم $SS_{کل}$ به تغییرپذیری درون خانه‌ها ($SS_{خطا}$) و تغییرپذیری بین حاصل جمع‌های خانه‌ها ($SS_{خانه‌ها}$) است. سپس $SS_{خانه‌ها}$ را به تفاوت‌های ناشی از مقاطع (SS_c)، تفاوت‌های ناشی از فعالیت (SS_A)، و تعامل مقطع و فعالیت (SS_{cA}) تقسیم می‌کنیم. این تقسیمات را به صورت نمودار زیر می‌توان نشان داد:



جدول ۳-۱۷ داده‌ها و محاسبات طرح عاملی 2×3

(الف) داده‌ها

کل‌های مقاطع T_{Cj}	فعالیت			
	A_3	A_2	A_1	
	(غیردرگیر)	(اجتماعی)	(ورزشی)	
۱۳۸	۷	۱۰	۹	C_1 (مقطع اول دبیرستان)
	۱۲	۶	۱۴	
	۷	۸	۱۱	
	۶	۷	۶	
	۸	۱۲	۱۵	
	۴۰	۴۳	۵۵	
۱۸۰	۱۰	۱۵	۱۷	C_2 (مقطع دوم دبیرستان)
	۸	۱۸	۱۲	
	۴	۱۶	۱۴	
	۲	۱۳	۱۸	
	۷	۱۵	۱۱	
	۳۱	۷۷	۷۲	
$G = 318$	۷۱	۱۲۰	۱۲۷	

(ب) میانگین خانه‌ها

	فعالیت			مقطع
	A_3	A_2	A_1	
۹/۲	۸/۰	۸/۶	۱۱/۰	C_1
۱۲/۰	۶/۲	۱۲/۰	۱۴/۴	C_2
	۷/۱	۱۲/۰	۱۲/۷	میانگین‌های فعالیت

(ج) محاسبات

$$\begin{aligned}
 SS_{\text{کل}} &= \sum X^2 - \frac{G^2}{N} \\
 &= 9^2 + 14^2 + \dots + 2^2 + 7^2 - \frac{318^2}{30} \\
 &= 3900 - 3370/8 = 529/2
 \end{aligned}$$

$$SS_{\text{خانه‌ها}} = \frac{\sum T_j^2}{n_j} - \frac{G^2}{N}$$

$$\begin{aligned}
 &= \frac{55^2 + \dots + 31^2}{5} - \frac{318^2}{30} \\
 &= 370.9 - 337.0/8 = 33.8/8 \\
 SS_{\text{خانه‌ها}} - SS_{\text{کل}} &= SS_{\text{خطا}} \\
 &= 529/2 - 33.8/8 = 190/4 \\
 SS_c &= \frac{\sum T_{ci}^2}{na} - \frac{G^2}{N} \\
 &= \frac{138^2 + \dots + 180^2}{5(3)} - \frac{318^2}{30} \\
 &= 3429/6 - 337.0/8 = 58/8 \\
 SS_A &= \frac{\sum T_{Aj}^2}{nc} - \frac{G^2}{N} \\
 &= \frac{127^2 + 120^2 + 71^2}{5(2)} - \frac{318^2}{30} \\
 &= 3557/0 - 337.0/8 = 186/2 \\
 SS_{cA} &= SS_{\text{cells}} - SS_c - SS_A \\
 &= 33.8/8 - 58/8 - 186/2 = 93/8
 \end{aligned}$$

(د) جدول خلاصه

منبع	df	SS	MS	F
مقطع	۱	۵۸/۸	۵۸/۸	*۷/۴۱
فعالیت	۲	۱۸۶/۲	۹۳/۱	*۱۱/۷۴
C × A	۲	۹۳/۸	۴۶/۹	*۵/۹۱
خطا	۲۴	۱۹۰/۴	۷/۹۳	
کل	۲۹	۵۲۹/۲		
*P < ۰/۰۵				

محاسبات مجموع مجذورات در قسمت (ج) جدول ۳-۱۷ دیده می‌شوند. این محاسبات بایستی آشنا باشند، زیرا همانند شیوه‌هایی هستند که در تحلیل یک طرفه مورد استفاده قرار گرفت. برای مثال، $SS_{\text{کل}}$ دقیقاً همانند روشی است که در فصل ۱۶ محاسبه شد، و همیشه دقیقاً به همان روش حساب می‌شود. ما همه مشاهدات مجذور شده را جمع کرده و از جمع کل مجذورات تقسیم بر N ، یعنی، G^2/N کم می‌کنیم.

مجموع مجذورات خانه‌ها (SS_{cells}) را بدست نمی‌آوردیم اما اگر این یک تحلیل واریانس یک طرفه باشد (یعنی، با نادیده گرفتن ماهیت عاملی طرح) گروهی SS را بدست خواهیم آورد. به

عبارت روشن‌تر، ما بطور ساده حاصل جمع‌های مجذور شده خانه‌ها را با هم جمع نموده، بر تعداد مشاهداتی که هر جمع خانه بر اساس آن بنا شده (n) تقسیم می‌کنیم و بعد G^2/N را از آن کم می‌کنیم. مجموع مجذورات خطا نیز به شیوه مشابه با تحلیل یک طرفه بدست می‌آید وقتی خانه‌ها SS را نوعی گروهی SS در نظر بگیریم. بطور ساده خانه‌ها SS را از کل SS کم کرده، و خطا SS باقی می‌ماند. این قابل فهم است زیرا اگر کل SS متشکل از تفاوت بین حاصل جمع خانه‌ها ($SS_{\text{خانه‌ها}}$) و همچنین تفاوت‌های درون خانه‌ها ($SS_{\text{خطا}}$) باشد، آنگاه با کم کردن خانه‌ها SS از کل SS تنها تفاوت‌های درون خانه‌ها به جا می‌ماند. SS_c درست همانند حالتی بدست می‌آید که یک تحلیل یک طرفه داشته باشیم و فعالیت به عنوان یک متغیر نباشد. به عبارت دیگر بطور ساده مجموع مجذورات حاصل جمع‌های مقطع را حساب می‌کنیم. همین کار را برای فعالیت می‌توان انجام داد، در اینجا متغیر مقطع را نادیده می‌گیریم.

توجه دارید که $\sum T_{Aj}^2$ بر na و $\sum T_{Ci}^2$ بر nc تقسیم شده‌اند. اگر سعی کنید این مخرج کسرها را به عنوان فرمول به‌خاطر بسپارید، تنها موفق به انباشتن چیزهای غیرضروری خواهید شد، و بسادگی اطلاعات فراموش می‌شوند. مخرج‌ها نشان دهنده تعداد نمرات در کل هستند نه بیشتر. هر کل مقطع بر اساس $na = (6)(3) = 18$ نمره و هر فعالیت بر اساس $nc = (6)(2) = 12$ است.

خانه‌ها SS اندازه‌ای از مقدار حاصل جمع‌های خانه‌ای (و در نتیجه میانگین‌های خانه‌ای) است. دو حاصل جمع خانه‌ای به سه دلیل ممکن است متفاوت باشند. تفاوت آنها ممکن است به این دلیل باشد که از سطوح مختلف مقطع باشند؛ همچنین دلیل تفاوت ممکن است این باشد که از سطوح مختلف فعالیت باشند؛ و ممکن است به دلیل تعامل بین مقطع و فعالیت، حاصل جمع‌ها متفاوت باشند. ما می‌دانیم بین شش خانه ($SS_{\text{خانه‌ها}}$) چقدر تفاوت وجود دارد. SS_c به ما می‌گوید چه مقدار از این تفاوت‌ها را می‌توان به مقطع نسبت داد، و SS_A به ما می‌گوید چه مقدار از این تفاوت‌ها را می‌توان به تفاوت‌ها در فعالیت نسبت داد. آنچه را که نتوان به C یا A نسبت داد باید به تعامل C و A نسبت داده شود (یعنی، SS_{CA}). بنابراین برای بدست آوردن SS_{CA} بطور ساده SS_c و SS_A را از خانه‌ها SS کم می‌کنیم. آنچه باقی می‌ماند SS_{CA} است. در مثال ما

$$SS_{CA} = SS_{\text{cells}} - SS_c - SS_A$$

$$93/8 = 338/8 - 58/8 - 186/8$$

جدول خلاصه برای تحلیل واریانس در قسمت (د) جدول ۳-۱۷ آمده است. ستون منبع و ستون مجموع مجذورات معلوم هستند. ستون درجات آزادی نیز با توجه به شناختی که از تحلیل یک‌طرفه دارید برایتان آشناست. مانند همیشه $df_{\text{کل}} = N - 1$ است. درجات آزادی برای مقطع و فعالیت بطور ساده $c-1$ و $a-1$ هستند، همان‌طور که در تحلیل‌های یک‌طرفه جداگانه بودند. درجات آزادی تعامل تنها عبارت واقعا تازه هستند. ما همیشه تعامل df را به صورت حاصل ضرب درجات آزادی هر مؤلفه تعامل حساب می‌کنیم. از این رو $2 \times 2 = 4$ $df_{CA} = df_C \times df_A = 1 \times 2 = 2$. در نهایت، درجات آزادی خطا از طریق تفریق کردن راحت‌تر بدست می‌آید. بنابراین $df_{\text{خطا}} = df_{\text{کل}} - df_C - df_A - df_{CA}$. به همین صورت، از آنجا که خطا MS

میانگین واریانس‌های خانه CA است و چون هر واریانس خانه دارای $n-1$ df می‌باشد خطای MS دارای (۱- $ca(n)$ درجه آزادی است. این قواعد درجات آزادی در هر تحلیل واریانس عاملی بدون توجه به پیچیدگی آن اعمال می‌شود.

درست همانند تحلیل واریانس یک‌طرفه، میانگین مجزورات با تقسیم مجموع مجزورات بر درجات آزادی آن بدست می‌آید. این همان شیوه‌ای است که در هر تحلیل به کار خواهیم برد.

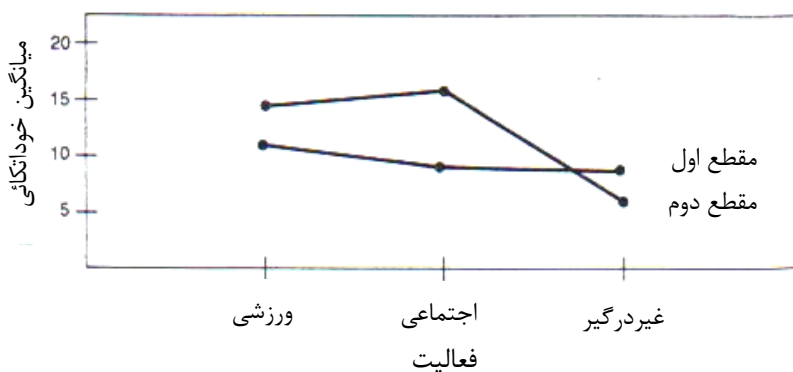
ما در نهایت، برای محاسبه F هر میانگین مجزورات را بر خطای MS تقسیم می‌کنیم. بنابراین برای مقطع، $F_C = MS_C / MS_{خطا}$ ؛ برای فعالیت، $F_A = MS_A / MS_{خطا}$ و برای $C \times A$ ، $F_{CA} = MS_{CA} / MS_{خطا}$. هر یک از این F ها بر اساس تعداد درجات آزادی برای عبارت مورد بررسی و خطای df است. پس برای مقطع بر اساس ۱ و ۲۴ df است، در حالی که F ها برای فعالیت و تعامل بر اساس ۲ و ۲۴ df است. از بخش ضمیمه، جدول F در می‌یابیم که مقادیر بحرانی $F_{0.05}(1, 24) = 4/26$ و $F_{0.05}(2, 24) = 3/40$ هستند.

- تفسیر

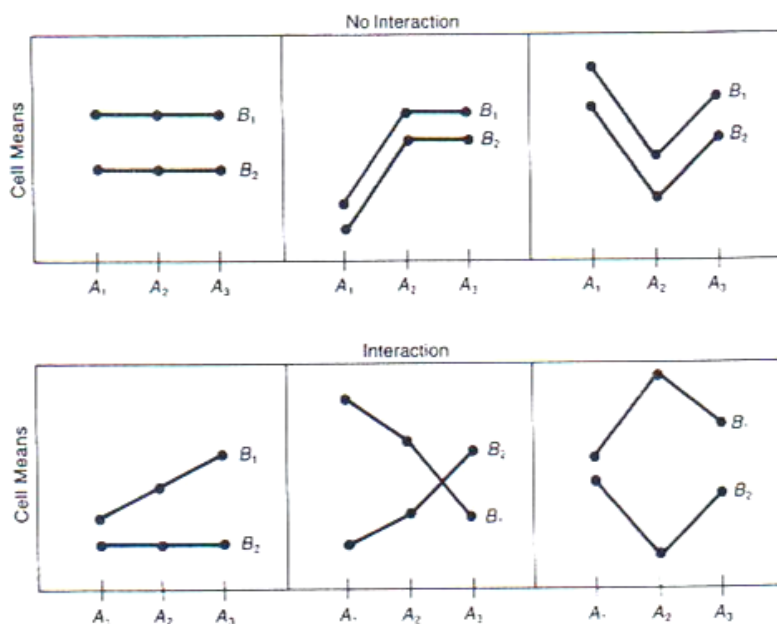
از بخش (د) جدول ۳-۱۷ در می‌یابیم که هر سه F از مقادیر بحرانی مربوط به آنها بالاتر هستند و بنابراین فرض صفر مربوط به آنها رد می‌شود. پس ما می‌توانیم نتیجه بگیریم که بطور میانگین در سطح فعالیت، دانش‌آموزان مقطع دوم متوسطه سطح بالاتری از اعتماد به نفس نسبت به دانش‌آموزان مقطع اول متوسطه دارند. ما همچنین می‌توانیم نتیجه بگیریم که بطور متوسط در سطح مقاطع تحصیلی، اعتماد به نفس با فعالیت تغییر می‌کند. بطور کلی در اینجا سطوح بالاتری از اعتماد به نفس برای گروه‌های ورزشی و اجتماعی نسبت به گروه غیردرگیر وجود دارد. دو مورد یاد شده **اثرات اصلی**^۱ نامیده می‌شوند زیرا هر یک متوجه اثر یک متغیر در یک زمان بدون توجه به متغیر دیگر است. اما، ما یک تعامل معنادار نیز در این تحلیل داریم که می‌تواند در یکی یا هر دو اثر یاد شده تردید ایجاد کند. چندان سودمند نیست که مثلاً بدانیم یک اثر فعالیت وجود دارد، وقتی که تعامل گویای آن است که اثر فعالیت برای هر مقطع یکسان نیست. بهترین راه تفسیر یک تعامل این است که میانگین‌های یک متغیر بطور جداگانه برای هر سطح متغیر دیگر ترسیم شوند. چنین نموداری را در شکل ۱-۱۷ می‌توانید ببینید، که در آن میانگین‌های فعالیت بطور جداگانه برای هر مقطع رسم شده‌اند. در اینجا می‌توانید تعامل در الگوی میانگین‌های فعالیت برای هر مقطع را مشاهده کنید که کاملاً با الگوی فعالیت در مقطع دیگر فرق دارد. در چنین حالتی نتایج مربوط به هر اثر اصلی توصیه نمی‌شود.

۳-۱۷ تعامل‌ها

دانشجویان اغلب در درک این موضوع مشکل دارند که منظور از اینکه گفته می‌شود دو متغیر تعامل دارند چیست. قبلاً گفتیم که منظور از تعامل معنادار این است که اثر یک متغیر بستگی به سطح متغیر دیگر دارد. شما این را در شکل ۱-۱۷ به صورت نموداری مشاهده کردید. وقتی نمودارهای متعددی برای میانگین‌های خانه‌ای در نظر بگیریم که وجود یا عدم وجود یک تعامل را نشان دهند وضع روشن‌تر خواهد شد. در شکل ۲-۱۷ سه نمودار بالایی حالتی را نشان می‌دهند که در آنها تعاملی وجود ندارد. توجه دارید که در هر سه حالت خطوط موازی هستند، حتی وقتی بالا و پایین می‌شوند. روش دیگر توضیح این است که گفته شود تفاوت بین B_1 و B_2 (اثر B) در A_1 همانند این تفاوت در A_2 و A_3 است. در سه نمودار قسمت پایین شکل، خط‌ها موازی نیستند. در اولی، یک خط صاف و خط دیگر بالا می‌رود. در دومی، خط‌های همدیگر را قطع می‌کنند. در سومی خط‌ها هم را قطع نمی‌کنند اما بر خلاف جهت هم حرکت می‌کنند. در هر مورد اثر B در سطوح مختلف A یکسان نیست. هر وقت خط‌ها (بطور معناداری) غیر موازی باشند، سخن از تعامل به میان می‌آید.



شکل ۱-۱۷ نمودار تعامل مقطع × فعالیت



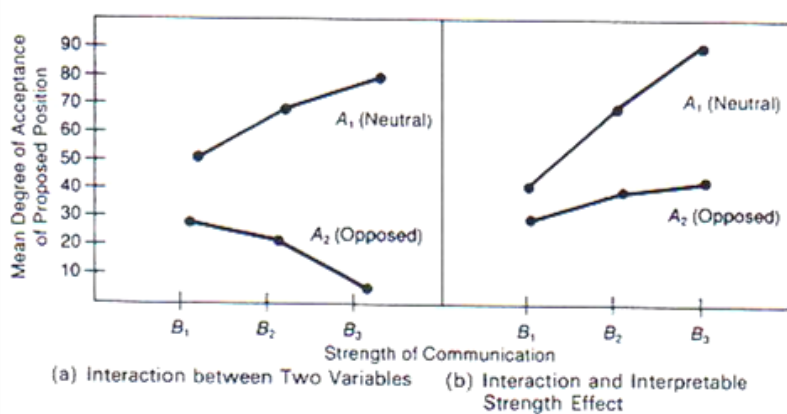
شکل ۲-۱۷ نمایش تعامل‌های ممکن و عدم تعامل

یکی از مزایای اصلی طرح‌های تعاملی این واقعیت است که به ما امکان آزمایش تعامل متغیرها را می‌دهند. در حقیقت در بسیاری موارد عبارت تعامل ممکن است بیشتر مورد توجه باشد تا اثرات اصلی (اثرات متغیرها به صورت انفرادی). مثلاً به یک مطالعه توجه کنید که در آن ما می‌خواهیم دیدگاه‌ها پیرامون برخی مسائل را تغییر دهیم. احتمالاً منطقی است که فرض کنیم اگر آزمودنی‌هایی را برگزینیم که نظر خاصی در باره موضوع ندارند، هر چه ارتباطات ما (در این محدوده) به حدود انتهایی نزدیک‌تر باشد، در جهت‌گیری ارتباطی تغییر بیشتری حاصل می‌شود. از سوی دیگر اگر آزمودنی‌ها خیلی مخالف دیدگاهی باشند که ما حمایت می‌کنیم، ارتباطات قوی‌تر و قوی‌تر می‌تواند منجر به دوری از دیدگاه ما شود. داده‌های چنین مطالعه فرضی در شکل ۳-۱۷ ارائه شده است.

در این مثال آشکارا معلوم است که سؤال از اینکه بین دو گروه تفاوت وجود دارد جایی ندارد. مسلماً وجود دارد. ما از ابتدا گروه‌های خود را بر این اساس انتخاب کرده‌ایم، نظریه ما پیش‌بینی نمی‌کند که آنها به هم نزدیک‌تر خواهند شد. همزمان ما احتمالاً به اثر اصلی متغیر توان ارتباطات کمتر علاقه‌مند هستیم. از آنجا که ما انتظار داریم یک مجموعه از نمرات افزایش و دیگری کاهش یابند، احتمالاً اصلاً مهم نیست که این نمودارهای واگرا در انتها چگونه به حد متوسط می‌رسند. این اثر تعاملی است که اهمیت خاصی دارد. ما پیش‌بینی کرده‌ایم که دو گروه به شیوه متفاوتی در سطوح مختلف متغیر ارتباطات

رفتار خواهند کرد، و این دقیقا همان چیزی است که یک تعامل معنادار به ما می‌گوید که در حال اتفاق افتادن است.

خیلی از افراد می‌گویند اگر شما یک تعامل معنادار پیدا کردید، اثرات اصلی را باید نادیده بگیرید. بعضی خوانندگان ممکن است یکی یا هر دو مثال پیش را نشان دهنده این اصل بدانند. اما دلیلی ندارد که بطور خودکار تعامل اثرات اصلی را در حضور هر تعامل معنادار غیر محتمل بدانیم. به عنوان نمونه‌ای از این مسأله، به داده‌های ترسیم شده در سمت راست شکل ۳-۱۷ توجه کنید که داده‌های دیگری را نشان می‌دهد که ممکن است از آزمایش ارتباطات بدست آورده باشیم. این داده‌ها کاملاً فرضی هستند، اما نقطه نظر فوق را بخوبی نشان می‌دهند.



شکل ۳-۱۷ داده‌های فرضی نشان‌دهنده تعامل

در اینجا می‌بینیم که دو متغیر تعامل می‌کنند، چرا که یک نمودار نسبت به دیگری دارای شیب تندتری است. در هر حال، با افزایش توان ارتباطی، هر دو گروه قدری بالا می‌روند، و اگر این اثر معنادار باشد، آگاهی از آن می‌تواند ارزشمند باشد. با یک اثر معنادار برای قدرت، ما داده‌هایی داریم که پیشنهاد کنیم اگر شما نمی‌توانید صدایتان را تعیین کنید، شما بایستی پیام خود را قوی بیان کنید نه ضعیف (دست کم در حدودی که این مطالعه طرح کرده است). برای هر دو گروه یک پیام قوی دارای اثر بیشتری است تا پیام ضعیف، هر چند این اثر برای گروه مخالف چندان چشمگیر نیست.

یک نکته مهم در این بحث این است که وقتی یک تعامل معنادار باشد آزمایشگر در مقایسه با حالت عادی باید حتی نگاه دقیق‌تری به داده‌ها بیافکند، پیش از آنکه در باره اثرات اصلی چیزی بگوید. این گفته‌ها ممکن است دقیق باشند یا نباشند، بستگی به ماهیت داده‌ها دارد.

۴-۱۷ اثرات ساده

یک شیوه خیلی مهم برای تحلیل داده‌هایی که شامل تعاملات معنادار هستند آزمایش و بررسی چیزی است که اثرات ساده نامیده می‌شود. یک اثر ساده اثر یک متغیر در یک سطح متغیر دیگر تعریف می‌شود. در مثال ما اثر متغیر ارتباطات برای گروه خنثی یک اثر ساده خواهد بود. اثر ارتباطات روی گروه مخالف اثر ساده دیگری است. یک اثر ساده سوم می‌تواند تفاوت بین گروه‌ها در ضعیف‌ترین سطح متغیر ارتباطات باشد.

محاسبات مربوط به اثرات ساده خیلی راحت است. به عنوان نمونه به داده‌های قبلی در باره اعتماد به نفس توجه کنید. تفاوت کلی بین دانش‌آموزان مقطع اول و دوم متوسطه تنها برای گروهی که درگیر فعالیت اجتماعی بودند یک اثر ساده است. برای داده‌های ما ارتباطات اجتماعی مقطع اول دارای میانگین ۸/۶ بود، در حالی که میانگین ارتباطات اجتماعی مقطع دوم ۱۵/۴ بود. برای محاسبه مجموع مجزورات این اثرات ساده ما فقط بیان می‌کنیم که تنها برای افراد درگیر در فعالیت اجتماعی داده داریم و مقطع SS را فقط برای این داده‌ها حساب می‌کنیم. فرمول کلی برای اثر ساده A در سطح اول B (B_1) اینگونه خواهد بود

$$SS_{B_1 \text{ در } A} = \frac{\sum T_{B_1 \text{ در } A}^2}{n} - \frac{(\sum T_{B_1})^2}{na}$$

برای حالت اثر ساده مقطع در سطح اجتماعی داریم

$$SS_{\text{مقطع در اجتماع}} = \frac{43^2 + 77^2}{5} - \frac{120^2}{10} = 115.6$$

این یک مجموع مجزورات است، درست شبیه به مجموع مجزورات اثر اصلی. چون این مبتنی بر دو حاصل جمع است دارای $df = 1 - 1 = 2$ خواهد بود. بنابر این

$$MS_{\text{مقطع در اجتماع}} = \frac{115.6}{2} = 57.8$$

با بکارگیری عبارت خطا برای تحلیل اولیه

$$F = \frac{MS_{\text{اجتماع در مقطع}}}{MS_{\text{خطا}}} = \frac{57.8}{7.93} = 7.29$$

از آنجا که $F_{.05}(1, 24) = 4.26$ ، ما H_0 را رد می‌کنیم و نتیجه می‌گیریم که در اعتماد به نفس بین دانش‌آموزان مقطع اول و دوم متوسطه درگیر در فعالیت‌های اجتماعی تفاوت وجود دارد. به عبارت دیگر تفاوت بین ۸/۶ و ۱۵/۴ معنادار است.

به عنوان مثالی دیگر ممکن است ما بخواهیم اطمینان پیدا کنیم که یک اثر فعالیت برای داده‌های مقطع دوم متوسطه وجود دارد. (یک اثر معنادار اصلی فعالیت ممکن است این موضوع را تضمین نکند چون این اثر می‌تواند به اثر فعالیت بزرگ برای مقطع دوم نسبت داده شود که میانگین‌های فعالیت کلی را

برای تفاوت باعث شود حتی اگر برای مقطع دوم تفاوتی نباشد. چنین نتیجه‌ای به احتمال زیاد یک تعامل معنادار ایجاد می‌کند.) برای بدست آوردن این اثر ساده داریم

$$SS_{\text{فعالیت در مقطع دوم}} = \frac{55^2 + 43^2 + 40^2}{5} - \frac{138^2}{15} = 25/2$$

با سه فعالیت ما داریم $df = 3 - 1 = 2$ ، بنابراین

$$SS_{\text{فعالیت در مقطع دوم}} = \frac{25.2}{2} = 12.6$$

$$F = \frac{12.6}{7.93} = 1.59$$

و

چون $F_{.05}(2, 24) = 3.40$ ، ما نمی‌توانیم H_0 را رد کنیم. ما مجبوریم نتیجه‌گیری کنیم که داده‌های ما در راستای این فرضیه است که وقتی ما تنها به مقطع دوم نگاه کنیم، اعتماد به نفس برای افراد درگیر در این فعالیت‌های متفاوت یکسان است.

۵-۱۷ نمونه‌های نابرابر

وقتی با یک تحلیل واریانس یک طرفه سروکار داریم، نمونه‌های نابرابر مشکل جدی ایجاد نمی‌کنند- ما بسادگی فرمول خود را تطبیق می‌دهیم. این به لحاظ تعریفی با طرح‌های عاملی سازگار نیست. زمانی که ما یک طرح عاملی با نمونه‌های نابرابر داریم، محاسبات بسیار مشکل‌تر شده و تعامل خیلی نامعلوم است. در اینجا بهترین راه حل داشتن n های نابرابر نیست. متأسفانه دنیا همیشه بر وفق مراد ما نیست و n های نابرابر اغلب رخ می‌دهند. بحث بیشتر در این زمینه را می‌توانید در کتاب هاوِل (۱۹۸۷) دنبال کنید.

۶-۱۷ اندازه‌ی اثر

روش‌های برآورد اندازه‌ی اثر برای متغیرها در یک طرح عاملی بسط ساده روش‌های مورد استفاده در طرح یک طرفه هستند. ساده‌ترین اندازه باز هم مجذور اتا (η^2) است، هر چند این برآوردی سوگیرانه از مقداری است که ما در صورت انجام مشاهدات از کل جمعیت بدست خواهیم آورد. برای هر اثر (اثرات اصلی یا تعاملات) در طرح عاملی ما η^2 را با تقسیم مجموع مجذورات آن اثر بر $SS_{\text{کل}}$ حساب می‌کنیم. برای مثال ما

$$\eta_C^2 = \frac{SS_C}{SS_{\text{کل}}} = \frac{58.8}{529.2} = 0.11$$

$$\eta_A^2 = \frac{SS_A}{SS_{\text{کل}}} = \frac{186.2}{529.2} = 0.35$$

$$\eta_{CA}^2 = \frac{SS_{CA}}{SS_{\text{کل}}} = \frac{93.8}{529.2} = 0.18$$

$$\eta^2_{\text{خطا}} = \frac{SS_{\text{خطا}}}{SS_{\text{کل}}} = \frac{190.4}{529.2} = 0.36$$

بنابراین در این آزمایش تفاوت‌های ناشی از مقطع، ۱۱٪ تغییرپذیری، تفاوت‌های ناشی از فعالیت ۳۵٪ و تعامل این دو متغیر ۱۸٪ تغییرپذیری را به حساب می‌آورند. سی و شش درصد از تغییرپذیری، واریانس خطا است و بنابراین نمی‌تواند به حساب هیچ یک از اثرات اصلی یا تعامل گذاشته شود. همانند تحلیل واریانس یک طرفه، η^2 برای برآوردهای سردستی مشارکت متغیرها ساده است. اما سوگیری خیلی کمتری در برآوردها از طریق ω^2 بدست می‌آید. محاسبات آن اگر چه کمی پیچیده‌تر اما سراسر است.

$$\begin{aligned}\omega_C^2 &= \frac{SS_C - (c-1)MS_{\text{خطا}}}{MS_{\text{کل}} + MS_{\text{خطا}}} \\ &= \frac{58.8 - (1)7.93}{529.2 + 7.93} = 0.09 \\ \omega_A^2 &= \frac{SS_A - (a-1)MS_{\text{خطا}}}{MS_{\text{کل}} + MS_{\text{خطا}}} \\ &= \frac{186.2 - (2)7.93}{529.2 + 7.93} = 0.09 \\ \omega_{CA}^2 &= \frac{SS_{CA} - (1)(1)MS_{\text{خطا}}}{MS_{\text{کل}} + MS_{\text{خطا}}} \\ &= \frac{93.8 - (1)(2)7.93}{529.2 + 7.93} = 0.15\end{aligned}$$

توجه دارید که این مقادیر اندکی کوچک‌تر از مقادیر η^2 هستند، هر چند که تفسیر آنها اساساً یکسان است.

۷-۱۷ یک مثال دیگر - سازگاری مادرانه

در فصل ۱۶ مثالی را دیدیم از داده‌های واقعی در باره سازگاری مادرانه مادران نوزادان کم وزن (گروه آزمایش و گروه گواه) و مادران نوزادان دارای وزن عادی (نورکومب و دیگران، ۱۹۸۵). در آنجا مشاهده کردیم که مادران در گروه آزمایشی (مداخله) و گروه وزن عادی بهتر از مادران نوزادان کم وزن در گروه گواه سازگاری داشتند. یک تحلیل کامل‌تر این داده‌ها می‌تواند شامل تجزیه گروه‌ها بر اساس تحصیلات مادر (دبیرستانی یا کمتر در برابر بالاتر از دبیرستان) باشد. محققان فکر می‌کردند که مادران دارای تحصیلات کمتر ممکن است از برنامه آزمایشی نفع بیشتری ببرند نسبت به مادران با تحصیلات بالاتر. اگر این درست باشد، تفاوت‌های بین گروه گواه از مادران نوزادان کم‌وزن و گروه آزمایشی آنها که دارای تحصیلات پایین هستند بزرگ‌تر از این تفاوت در دو گروه دارای تحصیلات بالا خواهد بود و یک تعامل

معنادار بدست می‌دهد. داده‌ها و محاسبات این تحلیل در جدول ۴-۱۷ آمده است. این داده‌ها بخشی از داده‌ها واقعی هستند، که تنها هشت مشاهده اول از آنها در هر خانه انتخاب شده است. (نتایج در راستای نتایج مجموعه داده است.) توجه دارید که هم یک اثر گروهی و هم اثر تحصیلی (مشخص شده با ستاره (*)) بعد از مقدار F وجود دارد، اما تعاملی وجود ندارد. (وقتی F کمتر از ۱ باشد، ما معمولاً " $F < 1$ "، را به جای مقدار واقعی آن گزارش می‌کنیم. نبودن یک تعامل به این معناست که تفاوت‌های گروهی وابسته به سطح تحصیلات نیستند، که بر خلاف فرضیه آزمایشی ما است. ما از این تحلیل نتیجه‌گیری خواهیم کرد که هم میزان تحصیلات و هم وجود یا عدم وجود برنامه مداخله بر سازگاری مادرانه اثر داشتند. اما تفاوت‌های سازگاری مشاهده شده بین گروه‌ها وابسته به سطح تحصیلات نبود.

جدول ۴-۱۷ داده‌های سازگاری مادرانه به عنوان تابعی از گروه و تحصیلات

(الف) داده‌ها			
گروه ۱	گروه ۲	گروه ۳	
گواه کم وزن آزمایشی	گواه کم وزن گواه	وزن عادی	جمع کل‌های تحصیلی
۱۴	۲۵	۱۸	تحصیلات دبیرستان یا کمتر
۲۰	۱۹	۱۴	
۲۲	۲۱	۱۸	
۱۳	۲۰	۲۰	
۱۸	۱۴	۱۴	
۱۳	۲۵	۱۷	
۱۲۷	۱۶۲	۱۳۰	
۱۱	۱۸	۱۶	تحصیلات بیشتر از دبیرستان
۱۱	۱۶	۲۰	
۱۶	۱۳	۱۲	
۱۲	۲۱	۱۴	
۱۲	۱۷	۱۸	
۱۳	۱۰	۲۰	
۱۷	۱۶	۱۲	
۱۳	۲۱	۱۳	

	۱۲۵	۱۳۲	۱۰۵	
کل‌های گروهی	۲۵۵	۲۹۴	۲۳۲	۷۸۱ = G

(ب) محاسبات

$$SS_{\text{کل}} = \sum X^2 - \frac{G^2}{N} = 14^2 + \dots + 13^2 - \frac{781^2}{48} = 13363 - 12707/53 = 655/48$$

$$SS_{\text{خانه‌ها}} = \frac{\sum T_{ij}^2}{n} - \frac{G^2}{N} = \frac{127^2 + \dots + 125^2}{8} - \frac{781^2}{48} = 12918/375 - 12707/53 = 210/86$$

$$SS_{\text{خطا}} = SS_{\text{کل}} - SS_{\text{خانه‌ها}} \\ = 655/48 - 210/86 = 444/62$$

$$SS_{\text{تحصیلات}} = \frac{\sum T_E^2}{ng} - \frac{G^2}{N} = \frac{419^2 + \dots + 362^2}{(8)(3)} - \frac{781^2}{48} \\ = 12775/21 - 12707/52 = 67/69$$

$$SS_{\text{گروه}} = \frac{\sum T_G^2}{ne} - \frac{G^2}{N} = \frac{232^2 + 294^2 + 255^2}{(8)(3)} - \frac{781^2}{48} = 122/79$$

$$SS_{\text{EG}} = SS_{\text{cells}} - SS_{\text{تحصیلات}} - SS_{\text{گروه}} = 210/86 - 67/69 - 122/79 = 20/38$$

(د) جدول خلاصه

منبع	df	SS	MS	F
تحصیلات	۱	۶۷/۶۹	۶۷/۶۹	۶/۳۹*
گروه	۲	۱۲۲/۷۹	۶۱/۴۰	۵/۸۰*
E × G	۲	۲۰/۳۸	۱۰/۱۹	< ۱
خطا	۴۲	۴۴۴/۶۲	۱۰/۵۹	
کل	۴۷	۶۶۵/۴۸		
*P < ۰/۰۵				

۸-۱۷ مثال کار شده پایانی

در فصل ۱۴ به یک مطالعه در باره یادآوری کلامی توسط آیزنک (۱۹۷۴) اشاره کردیم. در آن زمان ما تنها به بخشی از مطالعه کلی پرداختیم. در این بخش آن را به شکل کلی بررسی می‌کنیم. آیزنک به مطالعه یادآوری به صورت تابعی از (۱) سن آزمودنی‌ها و (۲) سطح پردازش علاقه‌مند بود. او تلاش کرد که این فرضیه را بیازماید که مطالبی که به صورت کامل‌تری توسط شخص پردازش می‌شوند بهتر به یاد آورده می‌شوند. علاوه بر این او فرض نمود که بطور کلی آزمودنی‌های مسن مطالب کمتری را به یاد می‌آورند، اما تفاوت اصلی در شرایطی پیش می‌آید که پردازش بیشتری نیاز است. به

عبارت دیگر در شرایط کمترین پردازش، تفاوت‌های بین آزمودنی‌های جوان و مسن کوچک خواهد بود یا وجود نخواهد داشت. در شرایطی که به پردازش بیشتری نیاز باشد، تفاوت‌های سنی قابل توجه خواهند بود. اگر این انتظارات را بپذیریم، یک تعامل سن با شرایط خواهیم داشت.

ایزنک تغییرات عمق پردازش در این تکلیف‌ها را با نوع آموزشی که برای آزمودنی‌ها در نظر گرفت اعمال نمود. در شرایط شمارشی به آزمودنی‌ها یک فهرست از لغات نشان داده می‌شد و از آنها خواسته می‌شد تا تعداد حروف هر لغت را بشمارند. در شرایط آهنگین از آزمودنی‌ها خواسته می‌شد کلمه‌ای را بسازند که با هر کلمه در آن فهرست لغات هم‌آهنگ باشد. شرایط صفت نیازمند آن بود که آزمودنی صفتی را تعیین کند که برای اصلاح لغات در آن فهرست مناسب باشد. در شرایط تصویری از آزمودنی‌ها خواسته می‌شد به تصویری فکر کنند که آن لغات بر می‌انگیزند. بالاخره، در شرایط ارادی به آزمودنی گفته می‌شد لغاتی را برای یادآوری بعدی به خاطر بسپارند. در همه شرایط پس از آنکه آزمودنی‌ها طبق برنامه آموزش، آن فهرست را طی می‌کردند از آنها خواسته می‌شد لغات را به یاد بیاورند، اما تنها در شرایط آخر به آزمودنی‌ها از پیش هشدار داده می‌شد که آنها مجبور خواهند بود لغات را به یاد بیاورند. بایستی روشن شده باشد که همچنان که ما از شرایط شمارشی دست کم به شرایط تصویری پیش می‌رویم، لازم است آزمودنی‌ها پردازش کامل‌تری روی لغات انجام دهند.

این مطالعه شامل ۵۰ آزمودنی در دامنه سنی ۱۸ تا ۳۰ و ۵۰ آزمودنی در دامنه سنی ۵۵ تا ۶۵ سال بود. داده‌های جدول ۶-۱۷ به گونه‌ای تدوین شده‌اند که میانگین‌ها و انحراف استانداردها همانند تحقیق آیزنک باشد. این جدول شامل تمام محاسبات لازم برای یک تحلیل واریانس استاندارد می‌باشد، و ما هر یک از آنها را در جای خود توضیح خواهیم داد. قبل از انجام تحلیل لازم است توجه شود که داده‌ها خود دارای توزیع تقریباً نرمال با واریانس‌های در حد قابل قبول برابر هستند. شما از روی خانه مربوط و جمع‌های کل می‌توانید بگویید که یادآوری ظاهراً با پردازش بیشتر افزایش می‌یابد، و آزمودنی‌های جوان به نظر می‌رسد بیشتر از افراد مسن به یاد می‌آورند. همچنین توجه داشته باشید که تفاوت بین آزمودنی‌های جوان و مسن ظاهراً بستگی به تکلیف دارد، که تکالیف درگیر با پردازش بیشتر تفاوت بیشتری دارند. در باره این نتایج پس از پرداختن به خود تحلیل بیشتر بحث خواهیم کرد.

از جدول خلاصه معلوم می‌شود که هم اثرات اصلی و هم تعاملی در سطح $P < 0.05$ معنادار هستند. بنابراین آزمودنی‌های جوان در کل نسبت به آزمودنی‌های مسن، لغات بیشتری را به یاد آورده‌اند؛ تکالیفی که پردازش بیشتری لازم دارند بهتر از آنهایی که نیاز به پردازش کمتر دارند به یاد آورده می‌شوند؛ و از همه مهم‌تر تفاوت‌های بین دو گروه سنی در تکالیفی که نیاز به پردازش بیشتری دارند بزرگتر است.

به عنوان یک نمایش نهایی از محاسبه اثرات ساده، ممکن است ما بخواهیم بدانیم آیا در شرایطی که نیازمند کمترین پردازش است تفاوت‌های سنی وجود دارد، و برای تکمیل اندازه‌گیری، آن را با شرایطی

که نیازمند پردازش زیاد است مقایسه کنیم. همچنین بطور نمونه، می‌توانیم بپرسیم آیا بین شرایط برای آزمودنی‌های مسن‌تر تفاوت‌های معناداری وجود دارد.

جدول ۵-۱۷ داده‌ها و محاسبات مثال آیزنگ (۱۹۷۴)

(الف) داده‌ها

شرایط یادآوری					
شمارشی	آهنگین	صفت	تصویری	ارادی	T_{Ai}
۹	۷	۱۱	۱۲	۱۰	
۸	۹	۱۳	۱۱	۱۹	
۶	۶	۸	۱۶	۱۴	
۸	۶	۶	۱۱	۵	
۱۰	۶	۱۴	۹	۱۰	مسن
۴	۱۱	۱۱	۲۳	۱۱	
۶	۶	۱۳	۱۲	۱۴	
۵	۳	۱۳	۱۰	۱۵	
۷	۸	۱۰	۱۹	۱۱	
۷	۷	۱۱	۱۱	۱۱	
۷۰	۶۹	۱۱۰	۱۳۴	۱۲۰	
۸	۱۰	۱۴	۲	۲۱	جوان
۶	۷	۱۱	۱۶	۱۹	
۴	۸	۱۸	۱۶	۱۷	
۶	۱۰	۱۴	۱۵	۱۵	
۷	۴	۱۳	۱۸	۲۲	
۹	۷	۱۲	۱۴	۱۸	
۷	۷	۱۱	۱۹	۲۱	
۶۵	۷۶	۱۴۸	۱۷۶	۱۹۳	۶۵۸
T_{Cj} ۱۳۵	۱۴۵	۲۵۸	۳۱۰	۳۱۳	$G = ۱۱۶۱$

(ب) محاسبات

$$\sum X^2 = ۱۶۱۴۷ \quad G^2 / N = ۱۱۶۱۲ / ۱۰۰ = ۱۳۴۷۹ / ۲۱$$

$$SS_{\text{کل}} = \sum X^2 - G^2 / N = ۱۶۱۴۷ - ۱۳۴۷۹ / ۲۱ = ۲۶۶۷ / ۷۹$$

$$SS_{\text{سن}} = \frac{\sum T_A^2}{nc} - \frac{G^2}{N} = \frac{503^2 + 658^2}{50} - \frac{1161^2}{100} \\ = ۱۳۷۱۹ / ۴۶ - ۱۳۴۷۹ / ۲۱ = ۲۴۰ / ۲۵$$

$$SS_{\text{شمارش}} = \frac{\sum T_C^2}{na} - \frac{G^2}{N} = \frac{135^2 + 145^2 + 258^2 + 310^2 + 313^2}{20} - \frac{1161^2}{100} \\ = ۱۴۹۹۴ / ۱۵ - ۱۳۴۷۹ / ۲۱ = ۱۵۱۴ / ۹۴$$

$$SS_{\text{خانه‌ها}} = \frac{\sum T_{ij}^2}{n} - \frac{G^2}{N} = \frac{70^2 + 69^2 + \dots + 176^2 + 193^2}{20} - \frac{1161^2}{100} \\ = ۱۵۴۲۴ / ۷۰ - ۲۴۰ / ۲۵ - ۱۵۱۴ / ۹۴ = ۱۹۴۵ / ۴۹$$

$$SS_{AC} = SS_{\text{خانه‌ها}} - SS_A - SS_C \\ = ۱۹۴۵ / ۴۹ - ۲۴۰ / ۲۵ - ۱۵۱۴ / ۹۴ = ۱۹۰ / ۳۰$$

$$SS_{\text{خطا}} = SS_{\text{کل}} - SS_{\text{خانه‌ها}} = ۲۶۶۷ / ۷۹ - ۱۹۴۵ / ۴۹ = ۷۲۲ / ۳۰$$

(د) جدول خلاصه

منبع	df	SS	MS	F
سن	۱	۲۴۰/۲۵	۲۴۰/۲۵	۲۹/۹۴*
شرایط	۴	۱۵۱۴/۹۴	۳۷۸/۷۳۵	۴۷/۱۹*
AC	۴	۱۹۰/۳۰	۴۷/۵۷۵	۵/۹۳*
خطا	۹۰	۷۲۲/۳۰	۸/۰۲۶	
کل	۹۹	۲۶۶/۷۹		

*P < ۰/۰۵

برای بدست آوردن اثرات ساده سن در سطح شرایط شمارشی و تصویری بطور جداگانه، ما نیاز داریم که جمع کل‌های هر یک از این شرایط را بدست آوریم. سپس مقدار تحلیل واریانس یک‌طرفه تحت هر یک از این دو شرایط را دنبال می‌کنیم، اما برای محاسبه Fها از $MS_{\text{خطا}}$ تحلیل واریانس کلی استفاده می‌کنیم. همین کار را برای اثر شرایط برای آزمودنی‌های مسن انجام می‌دهیم. این محاسبات به شرح زیر است:

۱- اثر ساده سن بر شمارش

$$SS_{A \text{ شمارش}} = \frac{70^2 + 65^2}{10} - \frac{135^2}{20} = ۹۱۲/۵۰ - ۹۱۱/۲۵ = ۱/۲۵$$

$$MS_{A \text{ در شمارش}} = \frac{SS_{A \text{ در شمارش}}}{1} = ۱/۲۵$$

$$F = \frac{MS_{A \text{ در شمارش}}}{MS_{\text{خطا}}} = \frac{1.25}{8.026} < 1$$

۲- اثر ساده سن در سطح تصویری

$$SS_{A \text{ تصویری}} = \frac{134^2 + 176^2}{10} - \frac{310^2}{20} = ۴۸۹۳/۲۰ - ۴۸۰۵/۰۰ = ۸۸/۲۰$$

$$MS_{A \text{ تصویری}} = \frac{SS_{A \text{ تصویری}}}{1} = ۸۸/۲۰$$

$$F = \frac{MS_{A \text{ تصویری}}}{MS_{\text{خطا}}} = \frac{88.20}{8.026} = ۱۰/۹۹$$

۳- اثر ساده شرایط در حالت مسن

$$SS_{C \text{ در مسن}} = \frac{70^2 + 69^2 + \dots + 120^2}{10} - \frac{503^2}{50}$$

$$= ۵۴۱۱/۷۰ - ۵۰۶۰/۱۸ = ۳۵۱/۵۲$$

$$MS_{A \text{ در مسن}} = \frac{SS_{C \text{ در مسن}}}{4} = \frac{351.52}{4} = ۸۷/۸۸$$

$$F = \frac{MS_{C \text{ در مسن}}}{MS_{\text{خطا}}} = \frac{87.88}{8.026} = ۱۰/۹۵$$

از این اثرات ساده می‌توان دریافت که بین آزمودنی‌های جوان و مسن برای ساده‌ترین تکلیف تفاوتی در یادآوری وجود ندارد، اما برای تکلیفی که پردازش زیادی می‌خواهد تفاوت وجود دارد. همچنین مشاهده می‌کنیم که برای آزمودنی‌های مسن تفاوت‌های معناداری در یادآوری به عنوان تابعی از عمق پردازش مورد نیاز وجود دارد.

۹-۱۷ خلاصه

در این فصل بحث تحلیل واریانس را گسترش دادیم تا طرح‌های دارای دو متغیر مستقل را شامل شود. در تحلیل عاملی ما فرض گرفتیم که در خانه‌های مختلف باز هم آزمودنی‌های متفاوتی وجود دارند. به برخی از مزایای طرح‌های عاملی توجه کردیم، بخصوص این واقعیت که به ما امکان می‌دهند تا به اثرات تعاملی دو متغیر نگاه کنیم. ما همچنین بطور کوتاه به موضوع اثرات ساده (اثر یک متغیر بر یک سطح متغیر دیگر)، مسائل ناشی از اندازه‌های نابرابر نمونه، و شیوه‌های برآورد مقادیر اثرات آزمایشی پرداختیم. برخی از مهم‌ترین مطالب این فصل عبارتند از:

○ سطوح	○ طرح عاملی 3×2
○ عامل‌ها	○ خانه
○ طرح عاملی دوطرفه	○ قسمت
○ طرح عاملی	○ اثر اصلی
○ تعامل	○ اثر ساده

۱۷-۱۰ تمرین

۱۷-۱ در یک مطالعه کامل تر رفتار غذاخوری نسبت به آنچه در فصل ۱۶ داشتیم (تمرین ۳-۱۶) ما به مشاهده مشتریان رستوران می‌پردازیم که به صورت زوج‌های همجنس می‌نشینند. ما قیمت غذای اصلی سفارش داده شده توسط یک شخص از آن زوج را ثبت نموده و آزمودنی‌ها را بر اساس جنسیت و اینکه صورت حساب را او می‌پردازد یا نه، دسته‌بندی می‌کنیم. داده‌ها به شرح زیر است:

میزبان		مهمان	
مرد	زن	مرد	زن
۸/۰۰	۸/۲۵	۹/۷۵	۸/۷۵
۷/۰۰	۸/۷۵	۱۰/۲۵	۹/۰۰
۸/۲۵	۹/۷۵	۹/۵۰	۹/۲۵
۹/۰۰	۸/۰۰	۹/۰۰	۸/۵۰
۸/۲۵	۹/۲۵	۱۰/۵۰	۸/۷۵

روی این داده‌ها یک تحلیل واریانس دوطرفه اجرا نمائید.

۱۷-۲ در یک مطالعه پیرامون رفتار مادر-کودک، مادران بر اساس کیفیت تعامل با نوزادان خود توسط یک مشاهده‌گر آموزش‌دیده رتبه‌بندی شدند. مادران بر اساس اینکه آیا این فرزند اول آنها است یا نه (زایمان اول یا چندم) و بر اساس اینکه این نوزاد کم‌وزن بوده یا عادی دسته‌بندی شدند. داده‌ها یک نمره را روی یک مقیاس ۱۲ درجه‌ای نشان می‌دهد که نمره بالاتر گویای تعامل بهتر مادر-کودک است.

فرزند اول		فرزند چندم	
کم وزن	عادی	کم وزن	عادی
۶	۸	۷	۹
۵	۷	۸	۸
۵	۷	۸	۹
۴	۶	۹	۹
۹	۷	۸	۳
۶	۲	۲	۱۰
۲	۵	۱	۹
۶	۸	۹	۸
۵	۷	۹	۷
۵	۷	۸	۱۰

یک تحلیل واریانس مناسب را اجرا و تفسیر نمائید.

۱۷-۳ با مراجعه به تمرین ۲-۱۷، به نظر می‌رسد که اندازه نمونه‌ها نشانگر فراوانی نسبی این ویژگی‌ها در جمعیت نباشند. آیا انتظار دارید که میانگین همه این مادران فرزند اول برآورد خوبی از جمعیت آنان باشد؟ چرا؟

۱۷-۴ با استفاده از شیوه‌های اثر ساده شرایط کم‌وزنی و وزن عادی را برای مادران دارای چند فرزند حساب کنید.

۱۷-۵ در یک مطالعه روی فرایندهای حافظه، حیوانات در یک تکلیف یادگیری اجتنابی تک‌مرحله‌ای مورد آزمایش قرار گرفتند. این حیوانات به محض گذشتن از یک در اتاق آزمایش با یک محرک ترس‌آور در مرحله یادگیری روبرو می‌شدند. متغیر وابسته مدت زمانی بود که طول می‌کشید تا آنها در مرحله بعدی (آزمایش) روی خط پا می‌گذاشتند. سه گروه حیوان از لحاظ ناحیه قرار گرفتن الکتروود در مغزشان (ناحیه خنثی، ناحیه الف و ناحیه ب) متفاوت بودند. هر گروه همچنین تقسیم شده و تحریک الکتریکی ترس‌آور را به فاصله زمانی ۵۰، ۱۰۰، یا ۱۵۰ میلی‌ثانیه بعد از پا گذاشتن روی خط دریافت می‌کردند. چنانچه ناحیه مغزی مورد تحریک با حافظه درگیر بود، انتظار می‌رفت تحریک با تثبیت حافظه تداخل پیدا کند و پاسخ یادگیری اجتنابی را مختل کند در نتیجه حیوان در بار دوم که پا روی خط بگذارد نباید درنگی نشان دهد. داده‌های تأخیر در پا گذاشتن روی خط در بار دوم به شرح زیر است:

منطقه تحریک								
ناحیه خنثی			ناحیه الف			ناحیه ب		
۵۰	۱۰۰	۱۵۰	۵۰	۱۰۰	۱۵۰	۵۰	۱۰۰	۱۵۰
۲۵	۳۰	۲۸	۱۱	۳۱	۲۳	۲۳	۱۸	۲۸
۳۰	۲۵	۳۱	۱۸	۲۰	۲۸	۳۰	۲۴	۲۱
۲۸	۲۷	۲۶	۲۶	۲۲	۳۵	۱۸	۹	۳۰
۴۰	۳۵	۲۰	۱۵	۲۳	۲۷	۲۸	۱۶	۳۰
۲۰	۲۳	۳۵	۱۴	۱۹	۲۱	۲۳	۳۱	۲۳

تحلیل واریانس را اجرا کنید.

۱۷-۶ در تمرین ۵-۱۷ خانه میانگین‌ها را رسم کنید.

۱۷-۷ با استفاده از آزمون محافظت شده t ناحیه خنثی را با دو ناحیه دیگر مقایسه کنید، تأخیر در تحریک را نادیده بگیرید. (راهنمایی: از شیوه‌های بیان شده در فصل ۱۶ کمک بگیرید اما مطمئن شوید که n_i را به عنوان تعداد نمراتی گرفته‌اید که بر اساس آن $\bar{X}_i$ بنا شده است.)

۱۷-۸ تمرین ۷-۱۷ را با استفاده از آزمون شفه به جای آزمون محافظت شده t انجام دهید.

۱۷-۹ با استفاده از اثرات ساده به بررسی اثر تأخیر در تحریک ناحیه الف بپردازید.

۱۷-۱۰ اگر به تمرین ۴-۱۶ مراجعه کنید، خواهید دید که در واقع یک طرح عاملی را تشکیل می‌دهد. یک تحلیل عاملی را اجرا و نتایج را تفسیر کنید.

۱۷-۱۱ در تمرین ۴-۱۶ شما یک آزمون بین گروه ۱ و گروه ۳ به صورت ترکیبی در مقابل ترکیب گروه‌های ۲ و ۴ اجرا نمودید. آن آزمون چگونه برای بررسی اثر اصلی موقعیت در تمرین ۱۰-۱۷ مقایسه می‌شود؟ آیا با هم فرق دارند؟

۱۷-۱۲ η^2 و ω^2 را برای داده‌های بخش ۹-۱۷ حساب کنید.

۱۷-۱۳ برای یک طرح 2×2 که دارای اثرات اصلی است اما اثر تعاملی ندارد یک مجموعه از داده‌ها بسازید.

۱۷-۱۴ برای یک طرح 2×2 که دارای یک اثر تعاملی باشد اما اثرات اصلی نداشته باشد یک مجموعه از داده‌ها بسازید.

۱۷-۱۵ یک آزمایش منطقی را توصیف کنید که در آن خواسته اصلی مورد نظر اثر تعاملی باشد.

۱۷-۱۶ برای تمرین ۱-۱۷، η^2 و ω^2 را حساب کنید.

۱۷-۱۷ برای تمرین ۲-۱۷، η^2 و ω^2 را حساب کنید.

۱۷-۱۸ طرحی شبیه آنچه در بخش ۱۷-۲ ترسیم شده است بکشید که نشان دهنده تقسیمات درجات آزادی باشد.

۱۷-۱۹ با مقایسه فرمول η^2 و ω^2 بیان کنید چه موقع این آمارهای متفاوت به هم نزدیک و چه موقع از هم دور خواهند شد.

۱۷-۲۰ در مطالعه ایزنک (۱۹۷۴) که در بخش ۱۷-۹ مورد تحلیل قرارگرفت، آزمون واقعی فرضیه ایزنک در باره تغییرات مربوط به سن در تعامل آن یافت شد. چرا؟

فصل ۱۸

طرح‌های اندازه‌گیری مکرر

در دو فصل قبلی به طرح‌های آزمایشی پرداختیم که آزمودنی‌های متفاوتی در هر گروه یا خانه وجود دارند. آنها را **طرح‌های بین-آزمودنی**^۱ می‌گویند زیرا شامل مقایسه‌هایی بین گروه‌های مختلف آزمودنی‌ها هستند. اما بسیاری طرح‌های آزمایشی شامل آن است که یک آزمودنی تحت بیش از یک شرایط آزمایشی قرار می‌گیرد. بطور مثال ممکن است یک سنجش خط پایه از بعضی رفتارها به عمل آوریم (اندازه‌گیری قبل از شروع هر برنامه آزمایشی)، یک اندازه‌گیری در انتهای یک برنامه آزمایشی به عمل آوریم، و سپس یک اندازه‌گیری در پایان یک دوره زمانی شش ماهه‌ی پیگیری صورت دهیم. چنین طرحی که در آن آزمودنی‌ها چند بار مورد سنجش قرار می‌گیرند **طرح اندازه‌گیری مکرر**^۲ خوانده می‌شود. این طرح‌ها موضوع این فصل هستند.

انواع مختلفی از طرح‌های اندازه‌گیری مکرر وجود دارد بسته به اینکه آیا هر آزمودنی تحت همه سطوح تمام متغیرها قرار می‌گیرد یا اینکه بعضی متغیرها گروه‌های متفاوتی از آزمودنی‌ها را در بر می‌گیرند در حالی که متغیرهای دیگر شامل آزمودنی‌های یکسان هستند. در این فصل ما تنها به ساده‌ترین مورد آن می‌پردازیم که در آن یک متغیر مستقل وجود دارد و هر آزمودنی در همه سطوح آن متغیر خدمت می‌کند. برای تحلیل‌های با پیچیدگی بیشتر به هاول (۱۹۸۱) و واینر (۱۹۷۱) مراجعه نمائید.

1. between-subjects designs
2. . repeated-measures design

۱۸-۱ یک مثال - درمان سردردهای میگرنی

به عنوان مثالی از یک طرح اندازه گیری مکرر به مطالعه‌ای در باره اثر شیوه رهاسازی^۱ در مهار سردردهای میگرنی می‌پردازیم. داده‌هایی که اینجا توضیح داده می‌شوند ساختگی هستند، اما در راستای داده‌های جمع‌آوری شده توسط بلانکارد و همکاران (۱۹۷۸) است که یک تحقیق خیلی مشابه اما پیچیده را صورت داد.

ما برای این مثال آزمایشی نه نفر را که از میگرن رنج می‌برند به کار گرفتیم. و از آنها خواستیم دفعات و مدت سردردهای خود را ثبت کنند. پس از چهار هفته ثبت خط پایه که در طی آن آموزشی داده نشد، یک دوره شش هفته‌ای آموزش رهاسازی وجود داشت. (هر آزمودنی آزمایشی در زمان‌های متفاوت در برنامه شرکت داشت، طوری که مواردی از قبیل تغییرات جوی و تعطیلات (مثلاً کریسمس) دارای اثر منظمی بر داده‌ها نباشند). ما در مثال خود داده‌های دو هفته آخر خط پایه و دو هفته آخر آموزش را مورد تحلیل قرار خواهیم داد. متغیر وابسته مدت سردردها (هفته/ساعت) در هر یک از این پنج هفته است. داده‌ها و محاسبات در جدول ۱۸-۱ نشان داده شده است.

ابتدا نگاهی بیاندازید به جدول ۱۸-۱. توجه خواهید نمود که میزان زیادی تغییرپذیری در داده‌ها وجود دارد، اما بیشتر این تغییرپذیری از این واقعیت ناشی می‌شود که برخی افراد دارای مدت سردردهای کمتر و/یا بیشتر از دیگران هستند، که عملاً با برنامه مداخله کار خیلی کمی می‌توان کرد. آنچه ما در طرح اندازه‌گیری مکرر می‌توانیم انجام دهیم که در طرح بین آزمودنی نمی‌توانیم، این است که این تغییرپذیری را از کل SS حذف کنیم. اثر این کار این است که تفاوت آزمودنی‌ها از عبارت خطا برداشته و نسبت به حالتی که این کار را نکنیم یک خطای MS کوچکتر تولید می‌کند. ما این کار را با محاسبه عبارتی که آزمودنی‌ها SS نامیده می‌شود انجام می‌دهیم، که تفاوت بین افراد را از لحاظ مدت سردرد ثبت شده آنها می‌سنجد. آنگاه عبارت آزمودنی‌ها SS از کل SS در طول هفته‌ها SS کم شده و خطای SS را حساب می‌کنیم. (در طرح قبلی هر نمره نشان‌گر یک فرد بود، اگر ما یک آزمودنی‌ها SS حساب می‌کردیم همان چیزی می‌بود که کل SS است.)

در جدول ۱۸-۱ مشاهده می‌کنید که کل SS به شیوه معمول حساب شده است. به همین ترتیب آزمودنی‌ها SS و هفته‌ها SS همیشه تنها به عنوان اثرات اصلی هستند (جمع کل‌های مربوط به آنها مجذور شده، با هم جمع، بر تعداد مشاهدات در کل تقسیم و از G^2/N کم شده است). در نهایت عبارت خطا با کم کردن آزمودنی‌ها SS و هفته‌ها SS از کل SS بدست می‌آید.

جدول خلاصه در بخش (ج) جدول ۱۸-۱ نشان داده شده است. توجه خواهید داشت که ما یک F برای هفته‌ها، اما نه برای آزمودنی‌ها حساب کرده‌ایم. دلیلش این است که خطای MS یک مخرج مناسب برای یک F روی آزمودنی‌ها نیست، چون صورت و مخرج هر دو ضرورتاً برآورد یکسانی نخواهند داشت

حتی اگر H_0 صحیح باشد. بنابراین ما نمی‌توانیم متغیر آزمودنی‌ها را آزمایش کنیم. اما چیز مهمی را از دست نمی‌دهیم، زیرا ما بندرت فرق آزمودنی‌ها را مورد توجه قرار می‌دهیم. ما تنها آزمودنی‌ها SS را حساب می‌کنیم تا امکان محاسبه یک عبارت خطای مناسب را برای آزمون هفته‌ها فراهم کنیم.

جدول ۱۸-۱ تحلیل داده‌های سردردهای میگرنی

(الف) داده‌ها

آزمودنی	خط پایه					حاصل جمع آزمودنی‌ها
	هفته ۱	هفته ۲	هفته ۳	هفته ۴	هفته ۵	
۱	۲۱	۲۲	۸	۶	۶	۶۳
۲	۲۰	۱۹	۱۰	۴	۹	۶۲
۳	۷	۵	۵	۴	۵	۲۶
۴	۲۵	۳۰	۱۳	۱۲	۴	۸۴
۵	۳۰	۳۳	۱۰	۸	۶	۸۷
۶	۱۹	۲۷	۸	۷	۴	۶۵
۷	۲۶	۱۶	۵	۲	۵	۵۴
۸	۱۳	۴	۸	۱	۵	۳۱
۹	۲۶	۲۴	۱۴	۸	۱۷	۸۹
حاصل جمع هفته	۱۸۷	۱۸۰	۸۱	۵۲	۶۱	$G = ۵۶۱$
میانگین‌های هفته	۲۰/۷۸	۲۰/۰۰	۹/۰۰	۵/۷۸	۶/۷۸	۱۲/۴۷

(ب) محاسبات

$$\begin{aligned}
 SS_{\text{کل}} &= \sum X^2 - \frac{G^2}{N} \\
 &= ۲۱^2 + ۲۰^2 + \dots + ۵^2 + ۱۷^2 - \frac{561^2}{45} \\
 &= ۱۰۴۸۳ - ۶۹۹۳/۸ = ۳۴۸۹/۲ \\
 SS_{\text{آزمودنی‌ها}} &= \frac{\sum T_S^2}{w} - \frac{G^2}{N} \\
 &= \frac{63^2 + \dots + 89^2}{5} - \frac{561^2}{45}
 \end{aligned}$$

$$\begin{aligned}
 &= 7827/4 - 6993/8 = 833/6 \\
 SS_{\text{هفته‌ها}} &= \frac{\sum T_W^2}{n} - \frac{G^2}{N} \\
 &= \frac{187^2 + \dots + 61^2}{9} - \frac{561^2}{45} \\
 &= 8928/3 - 6993/8 = 1934/5 \\
 SS_{\text{خطا}} &= SS_{\text{کل}} - SS_{\text{آزمودنی‌ها}} - SS_{\text{هفته‌ها}} \\
 &= 3489/2 - 833/6 - 1934/5 = 721/1
 \end{aligned}$$

د) جدول خلاصه

منبع	df	SS	MS	F
آزمودنی‌ها	۸	۸۳۳/۶	۸	
هفته‌ها	۴	۱۹۳۴/۵	۴۸۳/۶۲۵	۲۱/۴۶*
خطا	۳۲	۷۲۱/۱	۲۲/۵۳۴	
کل	۴۴	۳۴۸۹/۲		
*P < ۰/۰۵				

مقدار F برای هفته‌ها بر اساس درجات آزادی ۴ و ۳۲ و $F_{0.05}(4, 32) = 2/68$ است. بنابر این ما می‌توانیم $\mu_1 = \mu_2 = \dots = \mu_5$ را رد کنیم و نتیجه‌گیری کنیم که برنامه رهاسازی به کاهشی در مدت سردردهای گزارش شده توسط آزمودنی‌ها در طول هفته انجامیده است. بررسی میانگین‌ها در جدول ۱-۱۸ نشان می‌دهد که طی سه هفته آخر آموزش، مدت زمان هفتگی سردردها به حدود یک سوم آنچه که در طی دوره خط پایه بود رسیده است.

شما احتمالاً توجه کرده‌اید که تعامل آزمودنی‌ها $\times$ هفته‌ها در جدول خلاصه نشان داده نشده است. با تنها یک نمره در هر خانه، عبارت تعامل عبارت خطا است، و در واقع بعضی افراد ترجیح می‌دهند به جای خطا از $W \times S$ استفاده کنند. فرقی نمی‌کند که شما آن را خطا به حساب آورید یا $W \times S$ ، این عبارت بازهم مخرج مناسب F برای هفته‌ها است.

۱۸-۲ مقایسه‌های چندگانه

اگر ما ناگزیر باشیم تحلیل را پیش‌تر ببریم و مقایسه‌هایی بین میانگین‌ها به عمل آوریم، می‌توانیم از شیوه t محافظت شده مورد بحث در فصل ۱۶ استفاده کنیم. $MS_{\text{خطا}}$ در این تحلیل، عبارت مناسب برای

به کارگیری در t محافظت شده خواهد بود. برای داده‌های ما نتایج روشن هستند و برای انجام مقایسه‌های چندگانه نفعی وجود ندارد یا ناپذیر است. اما نشان دادن شیوه مفید است زیرا به ما امکان می‌دهد یکی از پیش‌فرض‌هایمان را بازبینی نموده و همچنین ببینیم که چگونه میانگین‌های گروه‌های ترکیبی را بررسی کنیم.

اولین مقایسه‌ای که ممکن است بخواهیم انجام دهیم مقایسه این دو میانگین برای دوره خط پایه است. ما می‌خواهیم مطمئن شویم که حتی پیش از آغاز دخالت آزمایشی پیشرفتی نبوده است که وارد شرایط آزمایشی شده باشد. چون F کلی معنادار بود، ما می‌توانیم از t محافظت شده استفاده کنیم تا این مقایسه را انجام دهیم. برای t داریم

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \\ = \frac{20.78 - 20.00}{\sqrt{22.53(1.9 + 1.9)}} = \frac{0.78}{\sqrt{5.01}} = \frac{0.78}{2.24} = 0.35$$

این t دارای $df_{\text{خطا}}$ درجه آزادی است زیرا $MS_{\text{خطا}}$ به جای واریانس جمعی به کار رفته است. یک t به اندازه 0.35 مسلماً در سطح $\alpha = 0.05$ معنادار نیست. بنابراین دلیلی ندارد پیش از دخالت یک عمل آزمایشی (درمان) انتظار پیشرفتی را داشته باشیم. توجه داشته باشید که ما می‌توانیم یک آزمون t محافظت شده را اجرا کنیم مشروط به اینکه میانگین‌ها مربوط به دو نمونه مستقل باشند چون عبارت خطا بر مبنای آن تنظیم شده است.

آشکارترین مقایسه، با توجه به ماهیت این آزمایش، بین تمرین‌های خط پایه و آموزشی است. ۱۸ مشاهده خط پایه دارای میانگین 20.39 ، و 27 مشاهده آموزشی دارای میانگین 7.19 است. بنابراین

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MS_{\text{خطا}} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \\ = \frac{20.39 - 7.19}{\sqrt{22.53(1/18 + 1/27)}} = \frac{13.20}{\sqrt{2.086}} = \frac{13.20}{1.44} = 9.17$$

این t هم دارای 32 درجه آزادی است چون از $MS_{\text{خطا}}$ استفاده می‌کند. این t به روشنی در سطح $\alpha = 0.05$ معنادار است، که نشان دهنده تفاوت در میانگین مدت سردردها بین فازهای خط پایه و آموزشی تحقیق است.

ممکن است تعجب کنید که چگونه ما می‌توانیم چیزی که ظاهراً یک آزمون t گروه‌های مستقل است را به کار بگیریم درحالی که ما می‌دانیم داده‌ها مستقل نیستند. به یاد دارید که در فصل ۱۳ ما مشاهدات وابسته را با تشکیل تفاوت‌ها و سپس گرفتن انحراف معیار تفاوت‌ها دستکاری کردیم. در یک پانوشته در فصل ۱۴ اشاره کردیم که این کار لازم بود زیرا ما بسادگی نمی‌توانستیم واریانس تفاوت‌ها (ی)

نمونه‌های وابسته) را مستقیماً از X_1 و X_2 بدست آوریم مگر آنکه همبستگی بین X_1 و X_2 را بدانیم. اما، برای یک تحلیل واریانس اندازه‌گیری مکرر، $MS_{\text{خط}}$ در حقیقت یک برآورد از خطای استاندارد تفاوت‌ها است، هرچند ما نمرات اختلاف را برای محاسبه آن به کار نمی‌بریم. شما می‌توانید بسادگی با اجرای یک تحلیل واریانس اندازه‌گیری مکرر و یک آزمون t برای دو نمونه وابسته در همان مجموعه داده‌ها (مثلاً استفاده از داده‌های خط‌پایه این آزمایش) و توجه به شباهت‌های میان این عبارت‌ها در محاسبات خود، این را برای خودتان اثبات کنید.

۳-۱۸ پیش‌فرض‌های مربوط به طرح‌های اندازه‌گیری مکرر

طرح‌های اندازه‌گیری مکرر با همان پیش‌فرض‌های نرمال بودن و همگونی (تجانس) واریانس که برای تحلیل واریانس نیاز است درگیر هستند. علاوه بر این آنها (برای اهداف بسیار عملیاتی) نیازمند این پیش‌فرض هستند که همبستگی‌های بین جفت‌های سطوح متغیر تکرار شونده ثابت می‌باشند. در مورد مثال ما به این معناست که فرض کنیم (در جمعیت) همبستگی بین هفته ۱ و هفته ۲ همانند همبستگی بین هفته ۲ و ۳ است و الی آخر. برای مثال اگر همبستگی بین مدت در هفته ۱ و مدت در هفته ۲، ۰/۵۰ است، آنگاه همبستگی بین مدت و هر جفت هفته دیگر هم بایستی در حدود ۰/۵۰ باشد. این یک پیش‌فرض خیلی سخت است و چیزی است که احتمالاً به همان میزان که به آن برخورد شود نقض می‌شود. آزمون بطور جدی تحت تأثیر قرار نمی‌گیرد مگر آنکه این پیش‌فرض بطور جدی زیر پا گذاشته شود. اگر چنین باشد دو کار می‌توان کرد تا شرایط آسان شود. اول اینکه سطوح متغیر مستقل به مواردی محدود شود که شانس برآورد شدن این پیش‌فرض فراهم شود. برای مثال اگر شما تحقیقی روی یادگیری می‌کنید که هر فردی با ندانستن هیچ چیز آغاز و با دانستن همه چیز تمام می‌کند، همبستگی بین تمرین‌های اول و آخر نزدیک صفر است، درحالی که همبستگی بین تمرین‌های میانی احتمالاً بالاست. در این مورد تمرین‌های اولی و آخری را در تحلیل وارد نکنید. آنها احتمالاً چیز زیادی به شما نخواهند گفت.

کار دومی که می‌توان کرد این است که از شیوه محافظه‌کارانه پیشنهادی گرین هوس و گیزر (۱۹۵۹) استفاده کنید. برای مثال ما برای F ، مان $(w-1)$ و $(w-1) \times (n-1)$ df داشتیم. گرین هوس و گیزر نشان دادند که اگر شما F مشابهی را بگیرید اما آن را روی ۱ و $(n-1)$ df ارزیابی کنید، یک آزمون محافظه‌کارانه خواهید داشت بدون توجه به اینکه چقدر پیش‌فرض یاد شده زیر پا گذاشته شده است. برای بحث بیشتر در باره همبستگی و برای یک کار کمتر محافظه‌کارانه (اما پیچیده‌تر)، به هاوِل (۱۹۸۷) مراجعه کنید.

۴-۱۸ برتری‌ها و کاستی‌های طرح‌های اندازه‌گیری مکرر

برتری اصلی طرح‌های اندازه‌گیری مکرر پیش از این مورد بحث قرار گرفت. جایی که تفاوت‌های فردی زیادی بین آزمودنی‌ها وجود دارد، این تفاوت‌ها منجر به تغییرپذیری زیاد در داده‌ها می‌شود. وقتی آزمودنی‌ها تنها یک بار مورد سنجش قرار می‌گیرند، ما نمی‌توانیم تفاوت‌های آزمودنی‌ها را از خطای تصادفی جدا کنیم، و هر چیزی به عبارت خطا پیوند می‌خورد. اما وقتی ما آزمودنی‌ها را بطور مکرر مورد سنجش قرار می‌دهیم، می‌توانیم تفاوت‌های آزمودنی را اندازه‌گیری و از خطا جدا کنیم. این یک طرح آزمایشی قدرتمند ایجاد می‌کند و بنابراین رد کردن H_0 را آسان‌تر می‌سازد.

کاستی‌های طرح‌های اندازه‌گیری مکرر شبیه به کاستی‌های مورد بحث در رابطه با آزمون‌های t نمونه وابسته هستند (که مورد خاصی از طرح‌های اندازه‌گیری تکراری می‌باشند). وقتی آزمودنی‌ها بطور مکرر مورد سنجش قرار می‌گیرند همیشه خطر انتقال اثرات از یک تمرین به تمرین بعد وجود دارد. برای مثال، اثر دارویی که شما در تمرین ۱ تجویز کرده‌اید ممکن است تا تمرین بعد از بین نرفته باشد. همینطور یک آزمودنی ممکن است در تمرین‌های اولیه چیزهایی آموخته باشد که به او در تمرین‌های بعد کمک کند. در بعضی شرایط این مشکل را می‌توان با «موازنه»^۱ ترتیب ارائه عملکردهای آزمایشی کاهش داد. بنابراین نیمی از آزمودنی‌ها ممکن است عملکرد آزمایشی الف را بعد از ب، و نیمی دیگر ب را بدنبال الف دریافت کنند. این موازنه اثرات انتقالی را از بین نمی‌برد، بلکه باعث می‌شود بر هر دو عملکرد اثر یکسانی داشته باشد. هر چند طرح‌های اندازه‌گیری مکرر با کاستی‌هایی همراه هستند، در بیشتر مواقع برتری‌های آنها فزونی دارد، و چنین طرح‌هایی معروف و خیلی در کارهای آزمایشی سودمند هستند.

۴-۱۸ یک مثل کاری آخر

به عنوان یک مثال آخر ما مثالی را از فصل ۱۷ طرح می‌کنیم که نشان دهنده تفاوت‌ها و شباهت‌های بین طرح‌های اندازه‌گیری مکرر و تحلیل واریانس مرسوم قبلی است. در فصل ۱۷ داده‌هایی از آیزنک (۱۹۷۴) در باره یادآوری به عنوان تابعی از عمق پردازش به کار بردیم. ما اثر ساده شرایط بر آزمودنی‌های مسن را بررسی کردیم، که در واقع به یک تحلیل واریانس با استفاده از فقط آزمودنی‌های مسن رسید، به استثنای اینکه ما از تحلیل کلی $MS_{\text{خطا}}$ را به کار می‌بریم. در این فصل ما همان مجموعه از اعداد را در مورد آزمودنی‌های مسن به کار می‌گیریم تا پیوستگی حفظ شود. البته داده‌ها را از نو مرتب کردیم تا شبیه آن چیزی باشد که انتظار داریم وقتی هر فرد تحت هر یک از پنج شرایط یادآوری قرار گیرد بدست آید. ما تنها نمرات در یک ستون را به بالا یا پایین جابجا کردیم طوری که فردی که دارای یک نمره ضعیف در یک شرایط بود در شرایط دیگر هم نمره‌اش ضعیف باشد، و به همین صورت برای آزمودنی‌که

یادآوری خوبی داشته است. نمرات در هر گروه هنوز فرق نکرده‌اند. (اگر این داده‌های جدید را به فصل ۱۷ ببرید نتایج دقیقاً یکسانی بدست خواهید آورد.) این داده‌ها در جدول زیر آمده است با یک ستون اضافه در سمت چپ که جمع‌های کل آزمودنی‌ها است:

آزمودنی	شمارش	آهنگین	صفت	تصویری	عمدی	جمع کل آزمودنی
۱	۴	۳	۶	۹	۵	۲۷
۲	۵	۶	۸	۱۲	۱۰	۴۱
۳	۶	۶	۱۰	۱۱	۱۵	۴۸
۴	۶	۸	۱۱	۱۱	۱۱	۴۷
۵	۷	۶	۱۴	۱۱	۱۱	۴۹
۶	۷	۷	۱۱	۱۰	۱۱	۴۶
۷	۸	۷	۱۳	۱۹	۱۴	۶۱
۸	۸	۶	۱۳	۱۶	۱۴	۵۷
۹	۹	۹	۱۳	۱۲	۱۰	۵۳
۱۰	۱۰	۱۱	۱۱	۲۳	۱۹	۷۴
جمع کل	۷۰	۶۹	۱۱۰	۱۳۴	۱۲۰	۵۰۳

ابتدا ما $SS_{\text{کل}}$ را حساب می‌کنیم:

$$SS_{\text{کل}} = \sum X^2 - \frac{G^2}{N} = 4^2 + 5^2 + \dots + 19^2 - \frac{503^2}{50}$$

$$= 5847 - 5060/18 = 7869/18$$

اکنون ما دو اثر اصلی برای محاسبه داریم، یکی بر اساس جمع کل شرایط و یکی بر اساس جمع کل آزمودنی‌ها:

$$SS_{\text{شرایط}} = \frac{T_C^2}{n} - \frac{G^2}{N} = \frac{70^2 + 69^2 + \dots + 120^2}{10} - \frac{503^2}{50}$$

$$= 5411/70 - 5060/18 = 351/52$$

$$SS_{\text{آزمودنی‌ها}} = \frac{T_S^2}{C} - \frac{G^2}{N} = \frac{27^2 + 41^2 + \dots + 74^2}{5} - \frac{503^2}{50}$$

$$= 5339 - 5060/18 = 278/18$$

عبارت خطا را اکنون می‌توان با کم کردن بدست آورد:

$$SS = SS_{\text{خطا}} - SS_{\text{کل}} - SS_{\text{شرایط}} - SS_{\text{آزمودنی‌ها}}$$

$$= 7869/18 - 351/52 - 278/18 = 721/1$$

این عبارت خطا نیز با تعامل شرایط $\times$ آزمودنی‌ها برابر است، چنانکه در متن بیان شد. ما حالا جدول خلاصه را تشکیل می‌دهیم:

منبع	df	SS	MS	F
آزمودنی‌ها	۹	۲۷۸/۸۲	۸	
آزمودنی‌ها	۴	۳۵۱/۵۲	۸۷/۸۸	۲۰/۲۲
خطا	۳۶	۱۵۶/۴۸	۴/۳۵	
کل	۴۹	۷۸۶/۸۲		

به منظور بررسی F برای اثر شرایط به جدول F با درجات آزادی ۴ و ۳۶ مراجعه می‌کنیم. از جدول F پیوست کتاب در می‌یابیم که مقدار F برابر ۲/۶۵ است. از آنجا که $۲/۶۵ > ۲۰/۲۲$ است فرض صفر را رد کرده و نتیجه خواهیم گرفت که یادآوری مواد کلامی تحت شرایطی که آن مواد یاد گرفته می‌شوند تغییر می‌کند.

اگر به عقب برگردید به فصل ۱۷، خواهید دید که اثر ساده شرایط برای گروه مسن را محاسبه کردیم. به دو دلیل آن F بایستی فرق داشته باشد. دلیل اول کم اهمیت‌تر است و ناشی از این واقعیت است که در فصل ۱۷ عبارت خطا بر اساس مجموعه کامل داده‌ها بود نه فقط داده‌های افراد مسن. این نباید تغییر زیادی ایجاد کند. دلیل دوم برای بدست آوردن یک تفاوت قابل توجه F این است که در این مثال ما به صورتی عمل کردیم که آزمودنی‌ها تحت تمام شرایط قرار گیرند نه فقط یک شرایط. این بدین معناست که هر تفاوت منظم آزمودنی از عبارت خطا بر گرفته خواهد شد. این دقیقاً همان چیزی است که اتفاق افتاده است. در تحلیل قبلی $MS_{خطا}$ معادل ۸/۰۲۶ بود. در اینجا ۴/۳۵ است. این تفاوت به یک F منجر می‌شود که تقریباً دو برابر F قبلی است.

لازم است بخاطر داشته باشیم که ما داده‌ها را قدری تغییر دادیم تا آزمودنی‌هایی ایجاد شوند که یکپارچه ضعیف یا خوب باشند. اما این بیش از آن چیزی نیست که وقتی شما از همین آزمودنی‌ها تحت همه شرایط استفاده می‌کنید انتظار دارید. از مقایسه F در اینجا و فصل ۱۷ می‌توانید مشاهده کنید که با استفاده از یک طرح اندازه‌گیری مکرر در صورتی که عملی یا مناسب باشد، شما بطور کلی توان یک آزمایش، و در نتیجه احتمال یافتن یک تفاوت معنادار را افزایش می‌دهید.

۱۸-۶ خلاصه

در این فصل ما دیدیم که چگونه داده‌ها را دستکاری کنیم که در آن آزمودنی‌های منفرد تحت همه سطوح یک یا چند متغیر مستقل قرار گیرند. ما یک مورد ساده از یک طرح اندازه‌گیری مکرر را بررسی

کردیم و دیدیم که چنین طرح هایی تفاوت بین آزمودنی ها را از عبارت خطا کنار می زنند. با حذف تفاوت های فردی از خطای $MS_{\text{خطا}}$ ، طرح های اندازه گیری مکرر عموماً توانمندتر از طرح های بین آزمودنی مشابه می باشند. برخی از مهمترین مطالب این فصل عبارتند از:

- طرح های بین آزمودنی
- موازنه
- طرح های اندازه گیری مکرر

۷-۱۸ تمرین ها

۱-۱۸ دست کم این بخشی از فرهنگ عامه در آمریکا است که تجربه مکرر امتحان فوق لیسانس (GRE)، حتی بدون هر گونه مطالعه در فاصله بین آنها، باعث نمرات بهتر می شود. ما هشت آزمودنی را به کار گرفتیم و طی سه هفته، هر هفته در صبح شنبه از آنها آزمون کلامی فوق لیسانس گرفتیم. داده ها در جدول زیر آمده است:

جلسه امتحان			
آزمودنی ها	۱	۲	۳
۱	۵۵۰	۵۷۰	۵۸۰
۲	۴۴۰	۴۴۰	۴۷۰
۳	۶۱۰	۶۳۰	۶۱۰
۴	۶۵۰	۶۷۰	۶۷۰
۵	۴۰۰	۴۶۰	۴۵۰
۶	۷۰۰	۶۸۰	۷۱۰
۷	۴۹۰	۵۱۰	۵۱۰
۸	۵۸۰	۵۵۰	۵۹۰

یک تحلیل واریانس مناسب اجرا کنید. در باره اثرات تمرین روی امتحان چه نتیجه ای خواهید گرفت؟

۲-۱۸ با استفاده از داده های تمرین ۱-۱۸ به این دو سؤال جواب دهید:

(الف) داده های جلسه سوم را حذف کنید و یک آزمون t (نمونه وابسته) بین جلسه ۱ و ۲ اجرا کنید.

(ب) حالا یک تحلیل واریانس اندازه گیری مکرر روی این دو ستون مشابه اجرا کنید و این F را با مجذور t قبلی مقایسه کنید.

۱۸-۳ در تلاشی برای نشان دادن استفاده‌های عملی از اصول اساسی یادگیری، یک روان‌شناس علاقه‌مند به اصلاح رفتار در مطالعه‌ای پیرامون آموزش مهارت‌های خود مراقبتی به کودکان عقب‌مانده شدید داده‌هایی را جمع‌آوری کرد. او این داده‌ها را طی مرحله خط پایه، در پایان مرحله آموزش، و در یک جلسه پیگیری در شش ماه بعد از پایان آموزش جمع‌آوری نمود. داده‌ها به شرح زیر است. تحلیل مناسب را اجرا و تفسیر کنید.

خط پایه	آموزش	پیگیری
۸	۹	۷
۵	۷	۵
۳	۲	۳
۵	۷	۲
۲	۹	۵
۶	۷	۹
۵	۸	۶
۶	۵	۷
۴	۷	۳
۴	۹	۵

۱۸-۴ با استفاده از آزمون t محافظت شده داده‌های تمرین ۱۸-۳ را تفسیر کنید. (راهنمایی: در صورتی که اینها نمونه‌های مستقل باشند می‌توانید t را محاسبه کنید زیرا $MS_{\text{خط}}$ بر این اساس - با بر داشتن تفاوت‌های آزمودنی - تنظیم شده است.)

۱۸-۵ یک مثال از شرایطی بزنید که شما می‌توانید از تحلیل واریانس اندازه‌گیری مکرر استفاده مؤثری به عمل آورید.

۱۸-۶ با استفاده از داده‌های مربوط به متغیر ۷۷۲ و ۷۷۸ برای ده ایالت در تمرین ۹-۱۰، یک تحلیل واریانس اندازه‌گیری مکرر اجرا کنید تا این فرضیه آزمایشی را بیازمایید که نمرات کلامی آزمون پیشرفت با گذشت زمان تغییر می‌کنند. (توجه داشته باشید که در این مورد اندازه‌های مکرر بر اساس ایالت‌ها هستند نه آزمودنی‌ها.)

۱۸-۷ در تمرین ۱۸-۶ چه فرض صفری را می‌آزمایید؟

۱۸-۸ چرا ما در تمرین ۱۸-۶ سال‌ها را به عنوان اندازه مکرر نه اندازه‌های بین آزمودنی در نظر می‌گیریم؟

۱۸-۹ از آزمون‌های t محافظت شده در داده‌های تمرین ۱۸-۱ استفاده کنید تا عملکرد را در موارد زیر مقایسه کنید:

(الف) خط پایه اول و آخر و

(ب) آموزش اول و آخر.

(به یادآوری تمرین ۴-۱۸ مراجعه کنید)

۱۰-۱۸ تحلیل واریانس اندازه گیری مکرر و آزمون t را برای دو نمونه وابسته در دو هفته خط پایه جدول ۱-۱۸ اجرا کنید. به شباهت های آنها توجه کنید.

۱۱-۱۸ داده های تمرین ۳-۱۸ را در سطح ستون تغییر دهید تا همبستگی بین تمرین ها به حداکثر برسد. سپس تحلیل را دوباره اجرا کنید و به تغییر F توجه کنید.

فصل ۱۹

مجدور خی

در داستان شاهزاده کوچولوی سان- اگزوپری، راوی اظهار می‌دارد که او معتقد است شاهزاده از یک شهاب سنگ آسمانی بنام بی-۶۱۲ آمده است، او توجه خود به جزئیات را با بیان شماره دقیق شهاب سنگ با این مطلب توضیح می‌دهد

تصورات دوستی بزرگترها. وقتی شما به آنها می‌گویید یک دوست جدیدی پیدا کرده‌اید، آنها هرگز در باره چیزهای اساسی نمی‌پرسند. آنها به شما نمی‌گویند ”صدایش چطور است؟ چه بازی را بیشتر دوست دارد؟ آیا کلکسیون پروانه جمع‌آوری می‌کند؟“ بلکه سؤال می‌کنند: ”چند سال دارد؟ چند برادر دارد؟ وزنش چقدر است؟ درآمد پدرش چقدر است؟“ آنها فکر می‌کنند تنها از این موارد است که آنها هر چیزی را در باره او می‌فهمند.

هیجده فصل اول این کتاب از برخی جهات تمرکزشان بر چیزهایی شبیه به تعداد افراد بزرگسالی است که سان- اگزوپری خیلی دوست دارد. این فصل به تحلیل داده‌هایی اختصاص یافته است که بیشتر غیر عددی هستند.

در فصل ۱۱ تمایزی بین داده‌های اندازه‌ای (گاهی داده‌های کمی خوانده می‌شوند) و داده‌های مقوله‌ای (گاهی داده‌های فراوانی گفته می‌شوند) قائل شدیم. وقتی ما با داده‌های اندازه‌ای سروکار داریم هر مشاهده نشانگر یک نمره روی یک طیف است، و معمول ترین آمارها میانگین و انحراف معیار هستند. وقتی با داده‌های مقوله‌ای سروکار داریم در مقابل، داده‌ها شامل فراوانی مشاهدات رخ داده در هر یک از دو یا چند مقوله است (”آیا دوست شما صدای نازک یا کلفتی دارد؟“ یا ”آیا او یک کلکسیونر پروانه، سکه یا کارت‌های بیس بال است؟“)

به عنوان مثال، ما از ۱۰۰ آزمودنی تقاضا می‌کنیم تا به طبقه‌بندی سرمقاله مبهم یک روزنامه بپردازند از این جنبه که آیا با بیان اطلاعات در باره جلوگیری از بارداری موافق یا مخالف است، (جواب‌های خنثی یا نامطمئن پذیرفته نمی‌شود). نتایج ممکن است به صورت زیر باشد:

دیدگاه سرمقاله به عنوان

موافق	مخالف	کل
۵۸	۴۲	۱۰۰

در اینجا داده‌ها تعداد مشاهداتی هستند که در هر یک از دو مقوله قرار گرفته‌اند. با داشتن چنین داده‌هایی، ممکن است ما مایل باشیم که ببینیم آیا بطور معناداری افراد بیشتری سرمقاله را به عنوان موافق مسأله یافته‌اند تا مخالف، یا آیا سرمقاله واقعا خنثی بوده و فراوانی‌ها فقط یک انحراف تصادفی از نسبت ۵۰:۵۰ است. (به یاد داشته باشید که آزمودنی‌ها مجبور بودند بین موافق و مخالف یکی را انتخاب کنند.)

یک مطالعه که طرح متفاوتی دارد داده‌های مشابهی را در باره سرمقاله روزنامه جمع‌آوری کرد اما لازم بود که پاسخ‌دهندگان نیز براساس دیدگاه‌شان پیرامون جلوگیری از بارداری طبقه‌بندی شوند. این مطالعه ممکن است به داده‌های زیر برسد:

دیدگاه سرمقاله

	موافق	مخالف	پاسخ‌گو
۷۰	۴۶	۲۴	موافق
۳۰	۱۲	۱۸	مخالف
۱۰۰	۵۸	۴۲	

در اینجا می‌بینیم که قضاوت افراد در باره سردبیر وابسته به دیدگاه شخصی آنها است، اغلب (۴۶/۷۰) کسانی که موافق آگاهی بخشی بدون محدودیت جلوگیری از بارداری هستند دیدگاه سرمقاله را به سمت خود دیده‌اند و آنهایی که مخالفند (۱۸/۳۰) عموماً دیدگاه سردبیر را در همان جهت دیده‌اند. (به عبارت دیگر، دیدگاه‌های شخصی پاسخگویان و قضاوت آنها در باره سرمقاله مستقل از یکدیگر نیست.) هر چند این دو مثال از بعضی جهات از لحاظ شیوه‌ای که داده‌ها مرتب شده‌اند و از سؤال‌های آزمایشی مطرح شده‌اند متفاوت به نظر می‌رسند، شیوه آماری یکسانی-آزمون مجذور خی^۱- را می‌توان در هر دو مورد به کار برد. اما چون در این دو موقعیت سؤال‌های تحقیقاتی که می‌پرسیم و روشی که آزمون را به کار می‌بریم فرق می‌کند، ما هر یک را جداگانه بررسی می‌کنیم.

۱۹-۱ یک متغیر طبقه‌بندی - آزمون مجذور خی نیکویی برازش

رئیس بخش زبان انگلیسی یک دانشگاه بزرگ ایالتی می‌خواست کیفیت آموزش استادان را بررسی کند. موضوع این بود که چهار استاد برای تدریس یک درس به دانشجویان تازه وارد در ساعات یکسانی از ترم پاییز در برنامه قرار گرفته بودند. در حالی که تلاشی برای متعادل کردن اندازه کلاس‌ها نشده بود، با نگاه به تقاضاهای ثبت‌نام، او داده‌های زیر را یافت:

استاد				
برون	واشنگتن	کارپ	رودریجز	کل
۲۵	۴۰	۱۵	۳۶	۱۱۶
ثبت نام				

آیا دلیلی وجود دارد که رئیس بخش احساس کند که دانشجویان بیشتر مایلند در بعضی کلاس‌ها ثبت‌نام کنند و از بعضی کلاس‌ها خودداری کنند؟ یا، با قرار دادن این موضوع در چارچوب فرض صفر، آیا این داده‌ها نشان می‌دهند که دانشجویان تمایل یکسانی برای ثبت‌نام در هر کلاس را دارند؟

این مثال شامل یک متغیر مستقل است (استاد)، و آزمون مورد استفاده **آزمون نیکویی برازش^۱** نامیده می‌شود زیرا ما مایل هستیم بررسی کنیم چگونه نوعی مدل فرضی (در این مورد احتمال برابر برای ثبت‌نام در هر کلاس) با داده‌های بدست آمده برازش (تناسب) دارد. اگر دانشجویان در حقیقت به صورت تصادفی ثبت نام شوند، نه اینکه در جستجوی استاد مناسب باشند، آنگاه ما انتظار داریم که احتمال ثبت‌نام در کلاس هر استاد همانند احتمال ثبت‌نام در کلاس استادان دیگر باشد. چون چهار استاد وجود دارد، اگر تصادفی انتخاب کنیم، احتمال انتخاب هر استاد برابر $0/25$ است. به بیان دیگر، اگر ۱۱۶ دانشجو بطور تصادفی در چهار کلاس ثبت‌نام کنند، انتظار داریم $29 = 116/4$ نفر در هر کلاس ثبت‌نام کنند. آنگاه سؤال تحقیق به این صورت ساده می‌شود که آیا تفاوت‌های بین فراوانی‌های بدست آمده (۲۵، ۴۰، ۱۵، و ۳۶) و فراوانی‌های مورد انتظار (۲۹ در هر کلاس) آنقدر کوچک هستند که به شانس نسبت داده شود (خطای تصادفی نمونه‌گیری) یا تفاوت‌ها خیلی بزرگتر از آن هستند که با عامل شانس توضیح داده شوند؟

۱۹-۱-۱ آماره آزمون خی (χ^2)

این سؤال را ما با استفاده از آماره مجذور خی (χ^2) جواب خواهیم داد که به این صورت تعریف می‌شود

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

که

O = فراوانی مشاهده شده در هر مقوله

E = فراوانی مورد انتظار در هر مقوله

و حاصل جمع روی همه مقوله‌ها حساب می‌شود.

بررسی این معادله آشکار می‌سازد که چرا این معادله برای سؤال مورد نظر ما قابل استفاده است. توجه کنید که ما در صورت کسر در حال اندازه‌گیری مستقیم این هستیم که فراوانی‌های مشاهده شده چقدر از فراوانی‌های مورد انتظار فاصله دارند. هر چه انحراف بیشتر باشد، مقدار χ^2 بزرگتر می‌شود. مخرج کسر نقش مفیدی در رابطه با به حساب آمدن انحرافات بازی می‌کند. اگر ما در یک مقوله خاص انتظار ۵ مشاهده را داشته باشیم و ۱۵ مشاهده بدست آوریم، آن تفاوت ۱۰ درجه‌ای قابل توجه است. از سوی دیگر اگر ما انتظار ۵۰۰ داشته باشیم و ۵۱۰ بدست آید، این تفاوت ۱۰ درجه‌ای را می‌توان به عنوان یک چیز بی‌اهمیت کنار گذاشت. وقتی بر E تقسیم می‌کنیم، ما اندازه مجذور انحرافات از فراوانی مورد انتظار را بر اساس اندازه فراوانی مورد انتظار وزن می‌دهیم.

به کارگیری آزمون مجذور خی در داده‌های ما به صورت زیر است:

استاد					
مشاهده شده (O)	۲۵	۴۰	۱۵	۳۶	۱۱۶
مورد انتظار (E)	۲۹	۲۹	۲۹	۲۹	۱۱۶

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(25-29)^2}{29} + \frac{(40-29)^2}{29} + \frac{(15-29)^2}{29} + \frac{(36-29)^2}{29}$$

$$= 0.55 + 4.17 + 6.76 + 1.69 = 13.17$$

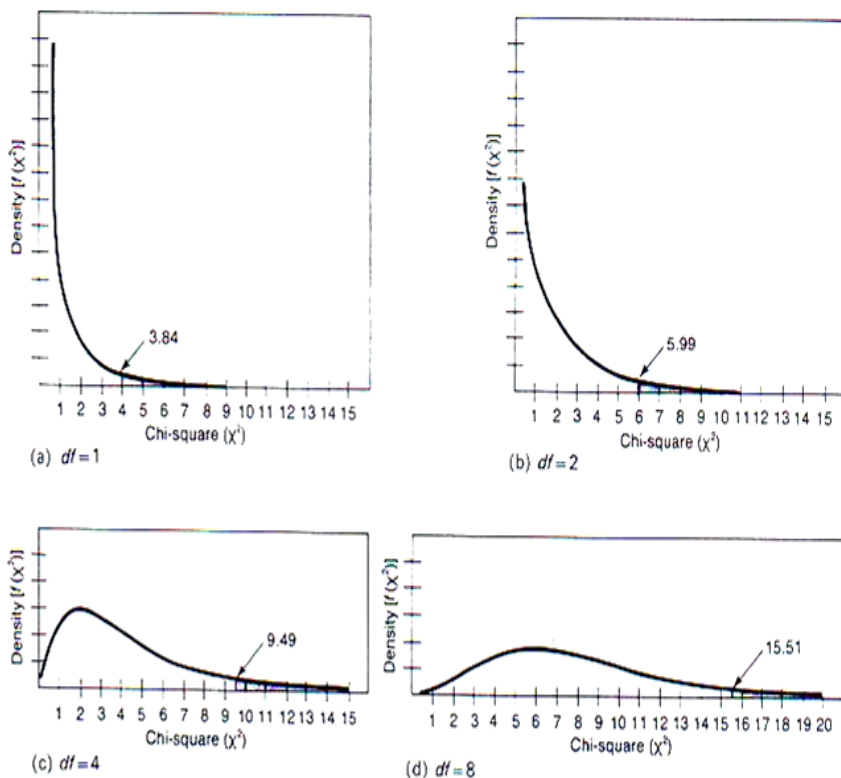
۱۹-۲ توزیع مجذور خی

برای آزمون این فرض صفر که احتمالات ثبت‌نام در کلاس‌های مختلف یکسان است، باید مقدار χ^2 بدست آمده را با توجه به جدول توزیع نمونه‌گیری مجذور خی در بخش پیوست ارزیابی کنیم. بخشی از جدول یاد شده در جدول ۱۹-۱ آمده است. توزیع مجذور خی، مانند توزیع‌های دیگری که دیده‌ایم، بستگی به درجات آزادی دارد. برای آزمون نیکویی برازش درجات آزادی به صورت $k-1$ تعریف می‌شوند که k گویای تعداد مقوله‌ها است. مثال‌هایی از توزیع مجذور خی برای چهار درجه آزادی مختلف در شکل ۱-۱۹ همراه با مقادیر بحرانی و ناحیه سایه‌دار رد فرض صفر برای $\alpha = 0.05$ نشان داده شده است. در این

شکل مشاهده می‌کنید که مقدار بحرانی برای یک سطح خاص α (مثلاً، $\alpha = 0.05$) برای درجات آزادی بزرگتر، گسترده‌تر خواهد بود. برای مثال ما چهار استاد داریم که $3 = 4 - 1$ df می‌ماند. از جدول مجذور خی پيوست کتاب یا جدول ۱-۱۹ در می‌یابیم که در $\alpha = 0.05$ ، $t_{(3)} = 3.182$ است. بنابراین اگر H_0 درست باشد، تنها در ۵٪ اوقات مقدار t را بزرگتر از 3.182 بدست خواهیم آورد. چون مقداری که ما بدست آوردیم $13/1$ است، H_0 را رد کرده و نتیجه خواهیم گرفت که چهار کلاس دانشجویان تازه وارد دارای محبوبیت یکسان نیستند.

جدول ۱-۱۹ خلاصه‌ای از نقاط درصدی بالای توزیع χ^2

[illegible]



شکل ۱۹-۱ توزیع مجذور خی برای درجات آزادی ۱، ۲، ۴ و ۸ با مقادیر بحرانی در سطح $\alpha = 0.05$

۱۹-۱-۳ یک مثال با فراوانی‌های مورد انتظار نابرابر

در مثال قبلی همه فراوانی‌های مورد انتظار برابر (یعنی ۲۹) بودند زیرا می‌خواستیم حالتی را ببینیم که دانشجویان بطور تصادفی در کلاس‌ها نام‌نویسی کنند. اما فرض کنید که بخواهیم حالت نام‌نویسی را با توجه به رتبه استاد بدانیم. این بخش دارای هشت استاد تمام، دو دانشیار و شش استادیار است، همگی کلاس ترکیبی را در ساعت یکسان از روز تدریس می‌کنند. اگر نام‌نویسی‌ها را بر اساس رتبه استاد محاسبه کنیم، نتایج به صورت زیر می‌شود:

استاد			
کل	استادیار	دانشیار	استاد
۳۹	۱۴۲	۴۶	۲۱۰
۸			مشاهده شده (O)

۳۹	۱۴۹/۲۵	۴۹/۷۵	۱۹۹	مورد انتظار (E)
۸				

مشاهده می‌کنید که تعداد خیلی زیادتری دانشجوی در کلاس‌های تدریس افراد استادتمام ثبت‌نام کرده‌اند تا افراد دانشیار یا استادیار. اما این چیزی است که برای ما قابل انتظار است زیرا تعداد بیشتری استادتمام وجود دارد تا دانشیار یا استادیار. از آنجا که ۵۰٪ از کادر آموزشی استادتمام هستند، اگر در نام‌نویسی سوگیری نباشد (یعنی H_0 درست باشد)، ما انتظار داریم ۵۰٪ از ۳۹۸ نفر (معادل ۱۹۹) دانشجوی در کلاس‌های آنها نام‌نویسی کنند. به همین صورت با داشتن ۶ دانشیار از ۱۶ نفر کادر آموزشی انتظار خواهیم داشت $۶/۱۶ = ۳۷/۵\%$ از دانشجویان در کلاس‌های آنها ثبت‌نام کنند ($۳۷۵ \times ۳۹۸ = ۱۴۹/۲۵$). بالاخره، ما انتظار داریم $۱۲/۵\% = ۲/۱۶$ از افراد در کلاس‌های با کادر مرتبه استادیار ثبت‌نام کنند ($۴۹/۷۵ \times ۳۹۸ = ۱۲۵/۰$). در جدول این فراوانی‌های مورد انتظار در زیر فراوانی‌های مشاهده شده آمده‌اند. برای این داده‌ها

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(142-149.25)^2}{149.25} + \frac{(46-49.75)^2}{49.75} + \frac{(210-199)^2}{199}$$

$$= ۰/۳۵ + ۰/۲۸ + ۰/۶۱ = ۱/۲۴$$

با سه مقوله ما $۲ = ۳ - ۱$ df داریم، و $\chi^2_{۰.۰۵}(۲) = ۵/۹۹$ ما فرض صفر را نمی‌توانیم رد کنیم. دلیلی نداریم تا در این موضوع شک کنیم که دانش‌آموزان به میزان مشابهی در کلاس‌های دارای مراتب مختلف استادان نام‌نویسی کرده‌اند. اگر این نام‌نویسی نمودی از تفاوت توانایی‌های آموزشی باشد، افراد استادتمام بهتر از دانشیاران و استادیاران به نظر نمی‌رسند. البته عوامل زیادی وجود دارند (مانند محتوای کلاس و نحوه نمره دادن) که بر نام‌نویسی تأثیر دارند. از این گذشته، تصور کنید اگر ما مثال خود را به کلاس‌های مقدماتی محدود نمی‌کردیم که در یک زمان و روز واحد برگزار می‌شوند، تفسیر داده‌های چگونه می‌بود.

۱۹-۲ دو متغیر طبقه‌بندی - تحلیل جدول توافقی

در دو مثال قبلی ما به موردی توجه کردیم که داده‌ها تنها در طول یک بعد مقوله‌بندی شده بودند (متغیر طبقه‌بندی). اما اغلب، داده‌ها با توجه به دو متغیر مقوله‌بندی شده‌اند و ما تمایل داریم ببینیم آیا این متغیرها مستقل از یکدیگر هستند. به عبارت دیگر ما می‌خواهیم بررسی کنیم که آیا توزیع پاسخ‌ها برای یک متغیر، وابسته به سطح متغیر دیگر است. در شروع این فصل ما بطور مختصر به مثالی پرداختیم که چگونگی نگاه یک آزمودنی به پیام یک سرمقاله در باره جلوگیری از بارداری، وابسته به دیدگاه شخصی آزمودنی پیرامون جلوگیری از بارداری است. مثال دیگر از این نوع، تحلیل مطالعه جکسون و پیازه (۱۹۸۲) در باره پدیده معروف به طفره اجتماعی است. طفره اجتماعی اشاره به مشاهداتی دارد در باره این که

وقتی افراد به عنوان یک تیم کار می‌کنند اغلب کمتر از زمان تنهایی کار می‌کنند. جکسون و پیازه آهنگ‌های نوشته شده بوسیله بیتل‌ها را تحلیل کردند. (هر چند ما قصد نداریم این داده‌ها را از لحاظ طفره اجتماعی بیازمائیم، مهم است که به نتیجه‌گیری نهایی جکسون و پیازه دقت کنیم که بعد از ۱۹۶۷- سالی که روابط بین این گروه شروع به از همپاشی کرد- مک کارنتی و لنون از طفره اجتماعی احساس گناه کردند که آهنگ‌هایی که با هم نوشته بودند کمتر از آهنگ‌هایی که به تنهایی نوشته بودند موفقیت‌آمیز بوده است.) ما می‌خواهیم از این داده‌ها برای پاسخ به یک پرسش مرتبط استفاده کنیم که آیا آهنگ‌های قبل از سال ۱۹۶۷ آنها، بدون توجه به سرپرستی، بیشتر یا کمتر از آهنگ‌های بعدی آنان از موفقیت برخوردار بوده است. برای اهداف تحلیلی در صورتی یک آهنگ موفقیت‌آمیز طبقه‌بندی می‌شود که همچنین به عنوان یک مورد ۴۵ rpm (دور در دقیقه) تک آهنگ هم منتشر شود. داده‌های این پرسش را می‌توان به صورت آنچه معروف به **جدول توافقی**^۱ 2×2 است نمایش داد. (نام 2×2 به این خاطر است که هر متغیر دارای دو سطح است.)

سال نوشتن			
		۱۹۶۶ یا	۱۹۶۷
		پیش‌تر	یا بعد
موفقیت‌آمیز	بلی	۳۷	۱۶
	خیر	۴۷	۶۲
		۸۴	۷۸
		۱۶۲	۵۳

در این جدول توافقی شواهدی را می‌بینیم که آهنگ‌های اولیه موفق‌تر بوده‌اند. از ۸۴ آهنگ نوشته شده در ۱۹۶۶ یا قبل از آن، ۳۷ مورد همچنین به عنوان تک آهنگ منتشر شده‌اند. اما از ۱۹۶۷ تنها ۱۶ مورد از ۷۸ آهنگ به عنوان موفقیت‌آمیز و منتشر شده به صورت تک آهنگ بودند. یک آزمون آماری فرض صفر مبنی بر این که موفقیت مستقل از سال نوشتن آهنگ است با آزمون مجذور خی مشابه آنچه در مورد نیکویی برازش به کار بردیم انجام می‌گیرد. تنها تفاوت محاسبه مقادیر مورد انتظار خواهد بود.

۲-۱۹-۱ فراوانی‌های مورد انتظار برای جداول توافقی

در یک جدول توافقی فراوانی مورد انتظار^۲ برای یک خانه معین از طریق ضرب کردن حاصل جمع‌های ستونی و ردیفی که خانه در آنها قرار گرفته و تقسیم آن بر کل تعداد نمونه (N) بدست می‌آید. (این

1. contingency table
2. expected frequency

حاصل جمع‌ها به عنوان جمع کل‌های حاشیه‌ای^۱ معروفند.) اگر E_{ij} فراوانی مورد انتظار برای خانه ردیف i و ستون j باشد، R_i و C_j جمع کل‌های ردیفی و ستونی هستند، و N تعداد کل مشاهدات است، آنگاه

$$E_{ij} = \frac{R_i C_j}{N}$$

برای مثال ما، پس

$$\begin{aligned} E_{11} &= \frac{(53 \times 84)}{162} = 27/48 \\ E_{12} &= \frac{(53 \times 78)}{162} = 25/52 \\ E_{13} &= \frac{(109 \times 84)}{162} = 56/52 \\ E_{14} &= \frac{(109 \times 78)}{162} = 52/48 \end{aligned}$$

۲-۱۹-۲ محاسبه مجذور خی

فراوانی‌های مشاهده شده و مورد انتظار برای هر خانه در جدول ۲ نشان داده شده است. محاسبه χ^2 سر راست است و همان فرمولی را که در مورد آزمون نیکویی برازش به کار بردیم مورد استفاده قرار می‌گیرد.

$$\begin{aligned} \chi^2 &= \sum \frac{(O-E)^2}{E} \\ \chi^2 &= \sum \frac{(37-27.48)^2}{27.48} + \frac{(16-25.52)^2}{25.52} + \frac{(47-56.52)^2}{56.52} + \frac{(62-52.48)^2}{52.48} \\ &= 3/30 + 3/55 + 1/60 + 1/76 = 10/18 \end{aligned}$$

جدول ۲-۱۹-۲ فراوانی‌های مشاهده شده و مورد انتظار برای آهنگ‌های بیتل‌ها

مورد انتظار			مشاهده شده			موفقیت آمیز
۱۹۶۷	یا ۱۹۶۶		۱۹۶۷	یا ۱۹۶۶		
	یا بعد	پیش تر		یا بعد	پیش تر	
۵۳	۲۵/۵۲	۲۷/۴۸	۵۳	۱۶	۳۷	بلی
۱۰۹	۵۲/۴۸	۵۶/۵۲	۱۰۹	۶۲	۴۷	خیر
۱۶۲	۷۸	۸۴	۱۶۲	۷۸	۸۴	

۳-۱۹-۲ درجات آزادی

قبل از آنکه ما بتوانیم مقدار χ^2 را با مقدار جدول مربوطه در پیوست کتاب مقایسه کنیم، می‌بایست درجات آزادی را بدانیم. برای تحلیل جداول توافقی درجات آزادی همیشه به این طریق بدست می‌آید

$$df = (R - 1)(C - 1)$$

که $R =$ تعداد ردیف‌ها

$C =$ تعداد ستون‌ها

در این مثال $R = 2$ و $C = 2$ ، و بنابراین ما $df = (2 - 1) \times (2 - 1) = 1$ داریم. داشتن یک درجه آزادی با ۴ خانه عجیب به نظر می‌آید، اما می‌توانید مشاهده کنید که وقتی حاصل جمع‌های ستونی و ردیفی را بدانید، برای تعیین آزمون تنها به فراوانی یک خانه نیاز دارید.

۱۹-۲-۴ ارزیابی χ^2

با ۱ df مقدار بحرانی χ^2 در جدول پیوست ۳/۸۴ است. چون مقدار χ^2 بدست آمده (۱۰/۱۸) از مقدار بحرانی بیشتر است، ما فرض صفر را که متغیرها مستقل از هم هستند رد می‌کنیم. در این مورد اینکه یک آهنگ موفق بوده یا نه (یعنی آیا به عنوان تک آهنگ منتشر شده) بخشی بستگی به این دارد که آیا قبل از سال ۱۹۶۷ منتشر شده یا بعد از آن. قبل از آنکه بطور قاطع نتیجه‌گیری کنیم که کیفیت آهنگ‌های بیتل‌ها بعد از ۱۹۶۷ کاهش پیدا کرده، مهم است که شماری از احتمالات دیگر را واریسی کنیم. مثلاً ممکن است صنعت ضبط پیش از سال ۱۹۶۷ کمتر از ۴۵s پخش شده یا اینکه در آن سال بیتل‌ها یک سیاست جلوگیری از p های در سطح ۴۵s اتخاذ کرده باشند. با این وجود داده‌ها نشان‌گر یک تغییر در بخش آهنگ‌های منتشر شده به صورت ۴۵s است. تنها مسأله تفسیر این نتیجه است.

۱۹-۳ حالت خاص جداول 2×2

در این کتاب از پرداختن به فرمولی که در موارد خاص به کار می‌رود وقتی فرمول عمومی آن همان کار را انجام می‌دهد خودداری شده است. ساده‌تر آن است که بجای حفظ یک فرمول اضافی، قدری بیشتر روی محاسبات کار کنیم. اما در مورد جدول 2×2 فرمولی وجود دارد که شما را از محاسبه فراوانی‌های مورد انتظار معاف می‌سازد، و در حقیقت امکان می‌دهد تا χ^2 را حتی با یک ماشین حساب بدون نوشتن نتایج میانی بسادگی حساب کنید.

فرض کنید یک جدول 2×2 با خانه‌ها و جمع کل‌های حاشیه‌ای به صورت زیر نامگذاری شده

باشد:

$A + B$	B	A
$C + D$	D	C
N	$B + D$	$A + C$

$$\chi^2 = \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)} \quad \text{آنگاه}$$

در مورد داده‌های مک‌کارتنی و لنون، داریم

موفقیت آمیز		۱۹۶۶ یا	۱۹۶۷
		پیش‌تر	یا بعد
بلی	۳۷	۱۶	۵۳
خیر	۴۷	۶۲	۱۰۹
	۸۴	۷۸	۱۶۲

$$\begin{aligned} \chi^2 &= \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)} \\ &= \frac{162[(37)(62)-(16)(47)]^2}{(53)(109)(84)(78)} = \frac{162(1542)^2}{(53)(109)(84)(78)} \\ &= 10.18 \end{aligned}$$

این دقیقاً همان مقدار χ^2 است که ما در محاسبات قبلی بدست آوردیم.

دانشجویان اغلب با این فرمول مشکل دارند زیرا فراموش می‌کنند کدام حرف به کدام خانه جدول اشاره دارد. همان کاری که ما انجام دادیم. اما به کاری که انجام می‌دهید توجه کنید. خانه‌هایی را در هم ضرب می‌کنید که نسبت به هم قطری هستند و یک حاصل ضرب را از حاصل ضرب دیگر کم می‌کنید. (حتی مهم نیست که کدام یک را از دیگری کم کنید، چون به هر حال نتیجه به توان دو می‌رسد.) سپس شما این جواب مجذور شده را در N ضرب و بعد بر جمع کل‌های تمام ردیف‌ها و ستون‌ها تقسیم می‌کنید. از این گذشته اگر شما واقعا گیج شدید و زمانی که بایستی تقسیم می‌کردید ضرب کنید یا برعکس، معمولاً به جواب غیر معقولی می‌رسید که متوجه اشتباه بودن آن می‌شوید.

تصحیح به خاطر پیوستگی بسیاری از کتاب‌ها توصیه می‌کنند که برای جدول 2×2 چیزی را اعمال کنید که تصحیح پیوستگی^۱ خوانده می‌شود (همچنین به تصحیح ییتس نیز معروف است). این تصحیح به سادگی با کم کردن نیم واحد از صورت کسر قبل از مجذور کردن آن انجام می‌گیرد. این تصحیح در گذشته کاملاً مرسوم بود، اما با آگاهی بیشتر در باره تحلیل جداول توافقی محبوبیت خود را از دست داد. چنانکه کامیل و هوپکینز (۱۹۷۸) دریافتند، چنین تصحیحی تنها معلوم می‌دارد که آیا ما چیزی

1. correction for continuity or Yates' correction

بنام "جمع کل‌های حاشیه‌ای ثابت"^۱ داریم. به بیان دیگر، ما تنها زمانی این تصحیح را به کار می‌بریم که خود را در موقعیتی ببینیم که پیش از اجرای یک آزمایش دقیقاً می‌دانیم جمع کل‌های ردیفی و ستونی چه مقدار هستند. چون چنین شرایطی بندرت پیش می‌آید، بیش از این لازم نمی‌دانیم در باره جزئیات آن بحث کنیم. برای اطلاع بیشتر می‌توانید به هاول (۱۹۸۷) رجوع کنید.

۴-۱۹ مجذور خی برای جداول توافقی بزرگتر

مثال قبلی شامل دو متغیر (موفقیت و تاریخ) بود، هر یک با دو سطح بود. این طرح خاص به عنوان یک تحلیل توافقی 2×2 خوانده می‌شود و مورد خاصی از طرح‌های عمومی تر $R \times C$ است (که باز هم R و C نشانه تعداد ردیف‌ها و ستون‌ها هستند). به عنوان مثالی از تحلیل یک جدول توافقی بزرگتر می‌توانیم به تحلیل داده‌های جمع‌آوری شده بوسیله دارلی و لاتان (۱۹۶۸) در باره "ناظران" بپردازیم. دارلی و لاتان از آزمودنی‌ها خواستند که در یک بحث که به صورت ارتباط داخلی (به منظور محرمانه ماندن) برگزار می‌شد شرکت کنند. به یک گروه از آزمودنی‌ها این باور القا شد که آنها تنها با سرپرست (بعداً قربانی خوانده می‌شود) مذاکره می‌کنند؛ گروه دیگری از آزمودنی‌ها فکر می‌کردند شخص دیگری در مذاکره حضور دارد؛ و گروه سوم فکر می‌کردند که چهار نفر در مذاکره هستند. در حقیقت آزمودنی در همه این موارد تنها بود. در میان بحث فرد قربانی در سمت دیگر تلفن وانمود می‌کرد که دچار حمله قلبی شده و درخواست کمک می‌کرد. او حتی می‌گفت که می‌ترسد بدون کمک بمیرد. یکی از متغیرهای وابسته تعداد آزمودنی‌هایی بود که در هر گروه سعی می‌کردند به قربانی کمک برسانند. قبول نتایج برای برخی از ما که مایلیم به مهربانی هموعان خود باور داشته باشیم دشوار است. گروهی از آزمودنی‌ها که هر یک فکر می‌کردند فقط آنها حضور دارند، ۸۵٪ تلاش برای کمک‌رسانی را در بر داشت (۱۵٪ دیگر ظاهراً تصمیم گرفتند بگذارند روال طبیعی طی شود). از سوی دیگر برای آزمودنی‌هایی که فکر می‌کردند یک شنونده دیگر حضور دارد، میزان پاسخگویی به ۶۲٪ افت کرد. از همه بدتر، وقتی آزمودنی‌ها فکر می‌کردند چهار شنونده دیگر هم حضور دارند میزان پاسخ به ۳۱٪ کاهش یافت. هر چند قابل تصور است که این تفاوت‌ها ناشی از شانس باشد، احتمال این موضوع زیاد نیست. ما می‌توانیم مجذور خی را به کار ببریم تا این فرض صفر را بیازمائیم که رفتار کمک‌رسانی مستقل از تعداد ناظران است. داده‌ها در جدول ۳-۱۹ آمده است. فراوانی‌های مورد انتظار به روشی همانند حالت 2×2 بدست آمدند. به عبارت دیگر فراوانی مورد انتظار در خانه بالایی سمت راست $7/75 = (31 \times 13) / 52$ است.

$$\begin{aligned} \chi^2 &= \sum \frac{(O-E)^2}{E} \\ &= \frac{(11-7.75)^2}{7.75} + \frac{(2-5.25)^2}{5.25} + \frac{(16-15.50)^2}{15.50} \end{aligned}$$

$$+ \frac{(10-10.50)^2}{10.50} + \frac{(4-7.75)^2}{7.75} + \frac{(9-5.25)^2}{5.25}$$

$$= 1/36 + 2/0.1 + 0/0.2 + 1/81 + 2/68$$

$$= 7/90$$

در این مثال ما $df = (3-1)(2-1) = 2$ داریم، و مقدار بحرانی χ^2 برابر $5/99$ است. چون مقدار بدست آمده χ^2 بزرگتر از مقدار بحرانی است، H_0 را رد می‌کنیم. اگر شما انتظار داشته باشید اتفاق ناگواری برایتان پیش آید، مطمئن باشید که افراد زیادی دور شما نیستند- چیزی که تقریباً واقعیت دارد.

جدول ۳-۱۹ فراوانی‌های مشاهده شده و مورد انتظار رفتار کمک‌رسانی به عنوان تابعی از تعداد ناظران

مشاهده شده		مورد انتظار	
جستجوی کمک		جستجوی کمک	
آری		آری	
نه		نه	
تعداد	۵	۱۱	۲
ناظران	۱	۱۶	۱۰
	۴	۴	۹
		۳۱	۲۱
		۳۱	۲۱
		۵۲	۵۲

۵-۱۹ مشکل فراوانی‌های مورد انتظار کوچک

مجذور خی یک آزمون مهم و معتبر برای بررسی نیکویی برازش یا استقلال متغیرها (جداول توافقی) است. اما این آزمون برای زمانی که فراوانی‌های مورد انتظار خیلی کوچک باشند مناسب نیست. آزمون مجذور خی بخشی بر اساس این پیش‌فرض است که اگر آزمایش در دفعات بی‌شماری با فراوانی‌های حاشیه‌ای یکسان تکرار شود، فراوانی‌های بدست آمده در هر خانه دارای توزیع نرمالی پیرامون فراوانی مورد انتظار خواهند بود. اما اگر فراوانی مورد انتظار کوچک باشد (مثلاً، $1/0$)، راهی برای این وجود ندارد که فراوانی‌های مشاهده شده توزیع نرمالی پیرامون فراوانی مورد انتظار داشته باشند. در مواردی که فراوانی‌های مورد انتظار خیلی کوچک باشند آزمون مجذور خی ممکن است یک آزمون آماری معتبر نباشد. البته مشکل این است که "خیلی کوچک" را چگونه تعریف کنیم. در کتاب‌های مختلف تعریف‌های گوناگونی شده است و تردیدها همچنان باقی است. ما در اینجا این نقطه نظر محافظه‌کارانه را پیش می‌گیریم که برای جداول توافقی کوچک (۹ خانه یا کمتر) همه فراوانی‌های مورد انتظار باید دست کم ۵ باشند. برای جداول بزرگتر این محدودیت می‌تواند تا حدودی کاهش یابد. برخی استدلال می‌کنند که این آزمون محافظه‌کارانه است و حتی با فراوانی‌های مورد انتظار خیلی کوچک‌تر، خطای نوع I اندکی

ایجاد می‌کند، اما حتی آنها هم توصیه می‌کنند وقتی حجم کلی نمونه خیلی کم است—که در بسیاری مواقع که فراوانی مورد انتظار کوچک باشد چنین است—آزمون دارای توان خیلی کمتری برای رد فرض صفر نادرست است. (برای توضیح بیشتر در این زمینه به کتاب کامیلی و هوپکینز، ۱۹۷۸ مراجعه کنید.)

۶-۱۹ استفاده از مجذور خی به عنوان یک آزمون روی نسبت‌ها

آزمون مجذور خی اغلب به عنوان یک آزمون روی نسبت‌ها یا فراوانی‌های بین دو نسبت مستقل به کار می‌رود. در هر حال این همان آزمونی است که ما از ابتدا به کار برده‌ایم. ما به طور ساده شیوه نگرش خود از نسبت‌ها را تغییر می‌دهیم (یعنی، نسبت‌ها را به فراوانی‌ها تغییر می‌دهیم). دو مثال ساده این نقطه نظر را روشن خواهد نمود.

مثال اول شامل یک نسبت منفرد است. یک فرد مسئول گسترش و نوسازی می‌خواهد یک ساختمان اداری بزرگ را در محلی که اکنون یک پارک شهری است بنا کند. این ساختمان شمار زیادی شغل ایجاد خواهد کرد، اما در عین حال پارک را تخریب خواهد کرد، و اختلاف نظرهای زیادی پیرامون این طرح وجود دارد. فرد مسئول یک آژانس نظرسنجی معروف را به استخدام درآورد، که از یک نمونه ۴۰ نفری از ساکنین محل مصاحبه به عمل آورده است. آنها دریافتند که ۶۰٪ از مردم به نفع طرح رأی دادند و ۴۰٪ مخالف آن بودند. آیا فرد مسئول می‌تواند اطمینان داشته باشد که دیدگاه اکثریت در آن محل به نفع اوست، یا اگر ساکنین بطور مساوی تقسیم شوند این نتایج ممکن است ناشی از شانس باشد. ما از مجذور خی استفاده خواهیم کرد تا این فرض صفر را بیازماییم که این جمعیت بطور برابر نصف شده‌اند—یعنی مقدار حقیقی P (نسبت موافق طرح در جمعیت) 0.50 است.

از آنجا که ۴۰ نفر رأی داده‌اند و نسبت افراد موافق 0.60 است، پس $24 = 40 \times 0.60$ نفر از افراد موافق و ۱۶ نفر مخالف هستند. اگر H_0 درست باشد ما انتظار یک فراوانی ۲۰ برای هر حالت داریم. بنابراین

موافق		مخالف	
بدست آمده	۲۴	۱۶	۴۰
مورد انتظار	۲۰	۲۰	۴۰

اکنون ما می‌توانیم حساب کنیم

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(24-20)^2}{20} + \frac{(16-20)^2}{20} = \frac{4^2 + 4^2}{20} = 1.60$$

در این مورد ما ۱ df داریم، و از جدول χ^2 می‌یابیم که در سطح $\alpha = 0.05$ مقدار بحرانی $\chi^2 = 3.84$ بنابراین ما نمی‌توانیم $H_0: P = 0.50$ را رد کنیم. فرد مسئول توسعه شواهد معتبری ندارد که بیش از نیمی از جمعیت موافق این طرح هستند.

مثال دوم شامل بررسی تفاوت بین دو نسبت مستقل و بر اساس مطالعه‌ای روی رفتار توسط لاتان و دابز (۱۹۷۵) است. در این مطالعه، آزمایش‌گران می‌آموختند که قدم به آسانسور بگذارند و، در همان لحظه یک دسته مداد یا سکه را کف آسانسور رها کنند. متغیر وابسته این بود که آیا ناظران در جمع کردن آنها کمک می‌کنند یا نه، و یکی از متغیرهای مستقل جنسیت ناظران (آزمودنی‌ها) بود. این مطالعه در سه شهر اجرا شد (کلمبوس، سیاتل و آتلانتا)، اما روی داده‌های کلمبوس تمرکز می‌کنیم، که تفاوت جنسی کمتر بود. ما همچنین اثرات جنسیت آزمایش‌گران را نادیده می‌گیریم. در اصل لاتان و دابز مشاهده کردند که ۲۳٪ از ناظران زن و ۲۸٪ از ناظران مرد اشیای ریخته را جمع کردند. (جالب است که حدود سه چهارم ناظران فقط ایستادند و تماشا کردند.) سؤال این است که آیا تفاوت بین ۲۳٪ و ۲۸٪ درصد به لحاظ آماری معنادار است. به منظور پاسخ این سؤال ما بایستی حجم کل نمونه را بدانیم. در این مورد ۱۳۰۳ زن و ۱۳۲۰ مرد وجود داشتند. (به حجم زیاد نمونه که به آسانی در این نوع آزمایش بدست می‌آید دقت کنید.) چون ما اندازه نمونه‌ها را می‌دانیم، می‌توانیم نسبت‌ها را به آسانی به فراوانی تبدیل کنیم. بنابراین

جنسیت ناظر		
مرد	زن	
۲۸٪	۲۳٪	کمک
۷۲٪	۷۷٪	عدم کمک
۱۳۲۰	۱۳۰۳	تعداد

می‌شود

جنسیت ناظر		
مرد	زن	
۶۷۰	۳۷۰	کمک
۱۹۵۳	۹۵۰	عدم کمک
۲۶۲۳	۱۳۲۰	تعداد

مقدار ۳۰۰ در گوشه راست بالای جدول از گرفتن ۲۳٪ از ۱۳۰۳ بدست آمد. مقادیر دیگر جدول هم به همین شیوه بدست آمده است.

چون این جدول 2×2 است، ما می‌توانیم محاسبات را با فرمول زیر ساده کنیم:

$$\begin{aligned} \chi^2 &= \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)} \\ &= \frac{2623[(300)(950)-(370)(1003)]^2}{(670)(1953)(1303)(1320)} \\ &= \frac{2623(-86110)^2}{(670)(1953)(1303)(1320)} = 8/64 \end{aligned}$$

از آنجا که مقدار بحرانی در $df = 1$ در سطح $\alpha = 0/05$ برابر $3/84$ است، ما H_0 را رد کرده و نتیجه‌گیری می‌کنیم که نسبت‌ها بطور معناداری متفاوت هستند. ما می‌توانیم نتیجه بگیریم که تحت شرایط این آزمایش مردان بیشتر از زنان تمایل به کمک دارند.

۷-۱۹ مشاهدات غیرمستقل

گذشته از سؤال‌های مربوط به فراوانی مورد انتظار کوچک، آزمون مجذور خی همچنین بر این پیش‌فرض استوار است که مشاهدات مستقل از هم هستند. در عمل منظور این است آزمودنی‌ها بایستی فقط و فقط در یک مشاهده از داده‌ها شرکت داشته باشند. در تحقیق دارلی و لاتان روی ناظران، اگر آنها تصمیم گرفته بودند هر آزمودنی دو یا سه بار به کار گرفته شود تا در استخدام آزمودنی‌ها صرفه‌جویی شود، تحلیل مجذور خی دیگر برای این داده‌ها مناسب نبود. خوشبختانه آنها بهتر می‌دانستند. اما این اشکال حتی در کار محققان مجرب هم وجود دارد. ساده‌ترین قاعده این است که N (مجموع جمع کل‌های ستونی و ردیفی) باید با تعداد آزمودنی‌ها (نه مشاهدات) در آزمایش برابر باشد. اگر خیلی بزرگ باشد، احتمالاً شما از هر آزمودنی در بیش از یک مشاهده استفاده کرده‌اید. اگر خیلی کوچک باشد، بعضی داده‌ها را حذف کرده‌اید (مثلاً تنها آزمودنی‌هایی را شمرده‌اید که برای قربانی درخواست کمک کرده‌اند و فراموش کرده‌اید آنهایی را که کمک‌رسانی نکرده‌اند شمارش کنید).

۸-۱۹ مثال کار شده پایانی

به عنوان مثال آخر تحقیقی را از گلر، ویتمر و اورباق (۱۹۷۶) می‌آوریم. این محققان رفتار پخش زباله را مورد مطالعه قرار دادند و در میان چیزهای دیگر علاقه‌مند بودند ببینند در صورتی که یک پیام در باره نریختن زباله روی فهرست تبلیغاتی ویژه روز یک سوپرمارکت قرار گیرد آیا اثرگذار خواهد بود. برای خیلی ساده کردن این کار تحقیقی کاملاً پیچیده، دو مورد از شرایط گلر در فهرست تبلیغاتی مورد توجه قرار گرفت. تحت یک شرایط (کنترل) فهرست تنها شامل فهرستی از اقلام ویژه‌ی روز بود. در شرایط دوم (پیام)، به فهرست تبلیغاتی این عبارت هم اضافه شده بود: "لطفاً زباله نریزید، لطفاً زباله‌ها را در جای مناسب بریزید." در انتهای یک روز، گلر و دانشجویان او نمره فهرست‌های تبلیغی را بررسی کردند. آنها تعدادی را که در سطل زباله، تعدادی که در چرخ‌های خرید، در کف فروشگاه، و در جاهای دیگری پیدا

کردند شمردند؛ و تعدادی که پیدا نشدند و ظاهراً بیرون برده شده بود را نیز مشخص کردند. داده‌های بدست آمده تحت این دو شرایط در جدول ۵-۱۹ نشان داده شده، که از یک جدول بزرگتر که توسطه گزر گزاری شده بدست آمده‌اند.

این جدول توافقی را می‌توان با استفاده از آزمون مجذور خی به خوبی تحلیل کرد زیرا ما ۱۷۷۲ مشاهده مستقل داریم که در خانه‌های غیر قابل جمع واقع شده‌اند. ما این فرض صفر را می‌آزماییم که مکان آگهی‌ها در انتهای روز مستقل از آموزش‌های روی آنها است، و ما $\alpha = 0.05$ را در نظر می‌گیریم. ما فراوانی‌های مورد انتظار را به شیوه قبل محاسبه می‌کنیم. یعنی برای یک جدول توافقی $E = RT \times CT / GT$. بنابراین تعداد آگهی‌های مورد انتظار از گروه کنترل (آگهی‌های بدون پیام) که در سطل زباله پیدا شدند، در صورتی که H_0 درست باشد $61/66 = (903)(121)/1772 = E_{23}$ در جدول ۵-۱۹ فراوانی‌های مورد انتظار در پراتنز نشان داده شده‌اند.

۵-۱۹ داده‌های حاصل از تحقیق گزر، ویتمر و اورباق (۱۹۷۶)، (داخل پراتنرها فراوانی‌های مورد انتظار)

مکان

	خروج	زباله ریختن	سطل زباله	
۹۰۳	۴۷۷ (۴۹۷/۳۶)	۳۸۵ (۳۴۳/۹۸)	۴۱ (۶۱/۶۶)	کنترل
۸۶۹	۴۹۹ (۴۷۸/۶۴)	۲۹۰ (۳۳۱/۰۲)	۸۰ (۵۹/۳۴)	پیام
۱۷۷۲	۹۷۶	۶۷۵	۱۲۱	

محاسبه χ^2 براساس همان فرمولی است که ما تا به حال استفاده کرده‌ایم.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(41-61.66)^2}{61.66} + \frac{(385-343.98)^2}{343.98} + \dots + \frac{(499-478.64)^2}{478.64} = 25.79$$

برای جدول ۵-۱۹، $df = 2$ وجود دارد، زیرا $(3-1)(2-1) = 2$. مقدار بحرانی $5/99 = \chi^2$ است، طوری که ما به رد H_0 و این نتیجه‌گیری هدایت می‌شویم که محل رها کردن آگهی‌ها بستگی به آموزش‌های داده شده دارد. به بیان دیگر، آموزش و مکان، مستقل از هم نیستند. از این داده‌ها چنین بر می‌آید که وقتی از آزمودنی‌های خواسته می‌شود زباله رها نکنند، درصد بالایی از آگهی‌ها به سطل زباله انداخته می‌شود یا از فروشگاه بیرون برده می‌شود، و تعداد کمتری در کف یا قفسه‌های فروشگاه انداخته می‌شود.

۹-۱۹ خلاصه

این فصل به استفاده از آزمون مجذور خی برای تحلیل داده‌های فراوانی می‌پردازد. ما ابتدا این آزمون را برای نیکویی برازش مد نظر قرار دادیم که در آن تنها یک متغیر طبقه‌بندی وجود داشت. سپس ما با استفاده از مجذور خی دو متغیر مستقل را مورد آزمایش قرار دادیم، که شیوه متداول‌تری است. در انتها مشکل فراوانی مورد انتظار کوچک و نیاز به مشاهدات مستقل را مورد توجه قرار دادیم. برخی از اصطلاحات مهم در این فصل عبارتند از:

- مجذور خی
- آزمون نیکوی برازش
- جداول توافقی
- فراوانی مورد انتظار
- جمع کل‌های حاشیه‌ای

۱۰-۱۹ تمرین‌ها

۱۹-۱ رئیس بخش روان‌شناسی تصور می‌کند بعضی اعضای هیئت علمی بخش معروف‌تر از بقیه هستند. سه کلاس روان‌شناسی عمومی در یک ساعت یکسان توسط اندرسون، کلانسکی و کام برگزار می‌شود. تعداد دانشجویانی که ثبت نام کرده‌اند در زیر آمده است.

اندرسون	کلانسکی	کام
۲۵	۳۲	۱۰

آزمون مجذور خی مناسب را اجرا و نتایج آن را تفسیر کنید.

۱۹-۲ از نقطه نظر طرح یک آزمایش معتبر، تفاوت مهمی بین تمرین ۱-۱۹ و مثال مشابه آن در این فصل وجود دارد. داده‌ها عملاً به سؤال رئیس بخش پاسخ نمی‌دهند. مشکل چیست و آزمایش را چگونه می‌توان اصلاح کرد؟

۱۹-۳ ما یک فرضیه داریم که اگر از آزمودنی‌ها بخواهید ویژگی‌های یک جمله از گفتار افراد (مثلاً، "من خیلی تند می‌خورم") را در پنج بخش از اصلاً شبیه من نیست تا خیلی شبیه من است، مرتب کنند، درصد مواردی که در هر بخش قرار داده می‌شود برای پنج حالت تقریباً ۱۰٪، ۲۰٪، ۴۰٪، ۲۰٪ و ۱۰٪ خواهد بود. از فرزند خود خواستیم تا ۵۰ جمله را مرتب کند و داده‌های زیر بدست آمد:

۸ ۱۰ ۲ ۱۰ ۴

آیا این داده‌ها فرضیه ما را تأیید می‌کنند؟

۱۹-۴ جواب تمرین ۳-۱۹ به چه جمعیتی تعمیم پیدا می‌کند؟

۱۹-۵ در یک مطالعه قدیمی توسط کلارک و کلارک (۱۹۳۹)، به کودکان سیاه‌پوست عروسک‌های سیاه و سفید نشان داده شد و از آنها خواستند تا یکی را برای بازی انتخاب کنند. از ۲۵۲ کودک، ۱۶۹ نفر عروسک سفید و ۸۳ نفر عروسک سیاه را انتخاب کردند. در باره رفتار این کودکان چه نتیجه‌ای می‌توان گرفت؟

۱۹-۶ بدنبال تحقیقی که در تمرین ۱۹-۵ به آن اشاره شد، هرابا و کرانت (۱۹۷۰) آن را تکرار کردند. این دو تحقیق دقیقاً یکسان نبودند اما نتایج آن جالب است. آنها دریافتند که از ۸۶ کودک سیاه‌پوست، ۲۸ نفر عروسک سفید و ۶۱ نفر عروسک سیاه را برداشتند. مجذور خی مناسب را روی داده‌ها اجرا و نتایج را تفسیر کنید.

۱۹-۷ داده‌های تمرین ۱۹-۵ و ۱۹-۶ را در یک جدول توافقی ترکیب کنید و آزمون مناسب را روی آنها اجرا کنید. سؤال مربوط به طبقه‌بندی دو طرفه با سؤال مربوط به تمرین ۱۹-۵ و ۱۹-۶ چه فرقی دارد؟ ۱۹-۸ مراکز بررسی سلامت روان جامعه مشکلات متنوعی را مشاهده می‌کنند، اما بعضی از آنها با یک نوع مشکل خاص بیشتر روبرو هستند تا دیگران. از ۱۰۰ مراجعه‌کننده‌ای که توسط هر مرکز دیده شده، شماری از آنها شامل آنهایی بوده‌اند که مشکلات سازگاری اجتماعی، مشکلات با زندگی و مشکلات دیگر را داشتند. داده‌های به شرح زیر است:

مرکز سلامت روان

	الف	ب	ج	
۱۳۰	۵۰	۴۰	۴۰	سازگاری اجتماعی
۸۰	۲۶	۳۴	۲۰	مشکلات زندگی
۹۰	۲۴	۲۶	۴۰	مشکلات دیگر
۳۰۰	۱۰۰	۱۰۰	۱۰۰	

(الف) آزمون مجذور خی عملاً چه فرضیه‌ای را روی این جدول می‌آزماید؟

(ب) تحلیل مناسب را اجرا کنید.

(ج) نتایج را تفسیر نمایید.

۱۹-۹ داده‌های تمرین ۱۹-۸ را به کار بگیرید تا مشخص کنید چگونه مجذور خی به عنوان تابعی از اندازه نمونه تغییر می‌کند.

(الف) مقدار هر خانه را نصف کنید و مجذور خی را دو باره حساب کنید.

(ب) این کار چه چیزی را در باره نقش اندازه نمونه در فرضیه آزمایی بیان می‌کند؟

۱۰-۱۹ هاوول و هوسی (۱۹۸۵) یک مقیاس رتبه‌بندی برای طبقه‌بندی کودکان به عنوان کسانی که در پایه دوم، رفتاری شبیه به اختلال نقص توجه (ADD) نشان می‌دهند یا نه به کار گرفتند. آنها سپس این کودکان را از نو در پایه‌های چهارم و پنجم طبقه‌بندی کردند. در انتهای پایه نهم آنها گزارش‌های مدرسه را واری کردند و توجه خود را به این معطوف نمودند که کدام کودک در کلاس جبرانی انگلیسی نام‌نویسی شده است. در داده‌های زیر همه کودکانی که به عنوان اختلال نقص توجه تشخیص داده شده بودند در یک گروه ترکیب شده‌اند:

طبقه‌بندی	نام‌نویسی انگلیسی	عدم نام‌نویسی انگلیسی
عادی	۲۲	۱۸۷
اختلال توجه	۱۹	۷۴
	۴۱	۲۶۱
		۳۰۲

آیا رفتار دوره ابتدائی بین آنهایی که بعداً در کلاس جبرانی انگلیسی نام‌نویسی می‌شوند و آنهایی که نام‌نویسی نمی‌شوند تمایز ایجاد می‌کند؟

۱۱-۱۹ در تمرین ۱۰-۱۹ کودکان طبقه‌بندی شدند به کسانی که حداقل یک بار در پایه دوم، چهارم یا پنجم رفتار شبه اختلال نقص توجه نشان دادند و آنهایی که نشان ندادند. اگر در عرض مقوله‌های فرو نرویم، داده‌های زیر را خواهیم داشت:

هرگز	پایه دوم	پایه چهارم	پایه دوم و چهارم
۲۲	۲	۱	۳
۱۸۷	۱۷	۱۱	۱۶

نام‌نویسی انگلیسی

عدم نام‌نویسی انگلیسی

پایه چهارم	پایه دوم و چهارم	پایه چهارم و پنجم	پایه دوم، چهارم و پنجم
۲	۴	۳	۴
۱۶	۷	۸	۶

نام‌نویسی انگلیسی

عدم نام‌نویسی انگلیسی

(الف) آزمون مجذور خی را اجرا کنید.

(ب) با نادیده گرفتن فراوانی‌های مورد انتظار کوچک چه نتیجه‌ای خواهید گرفت.

(ج) نسبت به این فراوانی‌های مورد انتظار کوچک چه احساسی دارید و چگونه می‌توانید بر این مشکل غلبه کنید؟

۱۹-۱۲ می‌توان یک آزمون یک طرفه مجذور خی روی داده‌های ردیف ۱ تمرین ۱۱-۱۹ انجام داد. اگر این کار را انجام دهید چه فرضیه‌ای را می‌آزمایید؟ این فرض چه فرقی با چیزی که در تمرین ۱۱-۱۹ آزمایش کردید دارد؟

۱۹-۱۳ اداره پست می‌خواهد سرعت تحویل نامه را بررسی کند. آنها نامه‌هایی را از واشنگتن دی‌سی به فواصل مختلف فرستادند و تعداد روزهایی که طول می‌کشید تا نامه‌ها برسند را ثبت نمودند. داده به شرح زیر است:

روزهای تحویل	۵۰ مایل	۱۵۰ مایل	۳۰۰ مایل	۳۰۰۰ مایل
یک	۵	۱۰	۱۵	۵
دو	۱۰	۱۰	۵	۵
سه یا بیشتر	۱۵	۵	۱۰	۱۰

(الف) یک مجذور خی اجرا کنید.

(ب) نتایج را تفسیر کنید.

۱۹-۱۴ در مثال آزمون نیکویی برازش با فراوانی‌های مورد انتظار نابرابر، چرا لازم بود که همه استادان در ساعت مشابهی تدریس کنند؟

۱۹-۱۵ فرض کنید که در مطالعه لاتان و دابز (۱۹۴۵) که در این فصل اشاره شد فقط ۱۰۰ مرد و ۱۰۰ زن باشند. χ^2 را در این حالت حساب کنید.

۱۹-۱۶ جواب تمرین ۱۵-۱۹ در باره اثرات اندازه نمونه بر توان یک آزمایش چه می‌گوید؟

۱۹-۱۷ در یک بررسی اخیر روی دانشجویان در یک درس نظری، ۴۵٪ فکر می‌کردند این کلاس نیاز به کار زیادی دارد و ۵۵٪ آن را راحت می‌دانستند. آیا تفاوت معنادار است؟

۱۹-۱۸ پاسخ تمرین ۱۷-۱۹ چه نکته مهمی را نادیده می‌گیرد؟

فصل ۲۰

آزمون‌های آماری غیرپارامتریک و توزیع - آزاد

۲۰-۱ آزمون من-ویتنی

۲۰-۲ آزمون جفت‌های همتای ویل کاکسون برای رتبه‌های علامت‌دار

۲۰-۳ تحلیل واریانس یک طرفه کروسکال-والیس

۲۰-۴ آزمون رتبه‌ای فریدمن برای k نمونه‌ی همبسته

بسیاری از شیوه‌های آماری که ما در فصل‌های پیش بحث کردیم هم درگیر برآورد یک پارامتر یا بیشتر از توزیع نمرات در جمعیت(ها)یی که داده‌ها از آنها نمونه‌گیری شده‌اند و هم پیش‌فرض‌های مربوط به شکل آن توزیع بودند. برای مثال آزمون t از واریانس نمونه (S^2) به عنوان برآوردی از واریانس جمعیت (σ^2) استفاده می‌کند و همچنین نیازمند این پیش‌فرض است که جمعیتی که از آن نمونه‌گیری شده نرمال باشد (یا حداقل توزیع نمونه‌گیری از میانگین نرمال است). آزمون‌هایی همچون آزمون t ، که شامل پیش‌فرض‌های مربوط به پارامترهای خاص یا برآورد دیگر آنها هستند آزمون‌های پارامتریک^۱ نامیده می‌شوند.

یک دسته آزمون‌ها هستند که متکی به برآورد پارامتر و/یا پیش‌فرض‌های توزیع نیستند. چنین آزمون‌هایی را معمولاً آزمون‌های غیر پارامتریک^۲ یا آزمون‌های توزیع-آزاد^۳ گویند. روی هم رفته اگر یک آزمون غیرپارامتریک باشد توزیع-آزاد هم هست، و در واقع این ماهیت توزیع-آزاد آزمون است

1. parametric tests
2. nonparametric tests
3. distribution-free

که برای ما خیلی مهم است. هر چند این دو نام اغلب به جای هم به کار برده می‌شوند، از این آزمون‌ها در اینجا به عنوان توزیع-آزاد یاد می‌شود.

بحث در باره ارزش آزمون‌های توزیع-آزاد سال‌ها طول کشیده و مسلماً در این فصل نمی‌توان آن را حل کرد. بسیاری احساس می‌کنند که در موارد خیلی زیادی، آزمون‌های پارامتریک چنان قوی هستند که آزمون‌های توزیع-آزاد را غیر ضروری می‌نماید. دیگران اما به همان اندازه به عدم تناسب آزمون‌های پارامتریک و برتری رویکرد توزیع-آزاد معتقدند. (بردلی، ۱۹۶۸، یک سخنگوی قدرتمند و مجرب گروه دوم است.) بدون توجه به موضع شما در این باره، مهم است که با بیشتر شیوه‌های معمول توزیع-آزاد و منطق زیربنایی آنها آشنا شوید. این آزمون‌ها چندان در ادبیات آزمایشی رایج هستند که به راحتی نمی‌توان آنها را نادیده گرفت.

برتری اصلی که برای آزمون‌های توزیع-آزاد منظور می‌شود آشکارترین آنها نیز هست- آنها متکی به هیچ پیش‌فرض محدود کننده جدی در باره شکل جمعیت(ها)ی نمونه‌گیری شده نیستند. منظور این نیست که آزمون‌های توزیع-آزاد هیچ پیش‌فرضی ندارند، بلکه منظور این است که پیش‌فرض‌های مورد نیاز آنها خیلی فراگیرتر از پیش‌فرض‌های آزمون‌های پارامتریک هستند. فرض صفر دقیق مورد آزمایش ممکن است مثلاً بستگی به این داشته باشد که آیا دو جمعیت متقارن هستند یا دارای شکل یکسانی می‌باشند. هیچ یک از این آزمون‌ها یک پیش‌فرض قبلی پیرامون شکل خاص توزیع ندارند؛ یعنی اعتبار آزمون تحت تأثیر اینکه توزیع متغیر در جمعیت نرمال است یا نه قرار نمی‌گیرد. در مقابل، یک آزمون پارامتریک معمولاً شامل بعضی پیش‌فرض‌های نرمال بودن است، و چنانچه پیش‌فرض‌ها برآورده نشوند، نتیجه‌گیری به عمل آمده بر اساس آن آزمون ممکن است نادرست باشد. ویژگی دیگر آزمون‌های توزیع-آزاد که اغلب به عنوان یک برتری به حساب می‌آید این واقعیت است که بسیاری از آنها، بخصوص آنهایی که در این فصل بحث می‌شوند، بیشتر به میانه حساس هستند تا میانگین. بنابراین اگر ماهیت داده‌های شما طوری است که شما اساساً به میانه‌ها گرایش دارید، آزمون‌هایی که اینجا ارائه می‌شود می‌توانند سودمندی خاصی برای شما داشته باشند.

کسانی که از سودمندی آزمون‌های پارامتریک در هر موردی دفاع می‌کنند منکر این نمی‌شوند که آزمون‌های توزیع-آزاد در پیش‌فرض‌های مورد نیاز خیلی آزادانه‌تر عمل می‌کنند. اما معتقدند که پیش‌فرض‌های نرمال بودن که مورد نیاز آزمون‌های پارامتریک است در عمل خیلی محدود هستند و این آزمون‌ها کمتر تحت تأثیر مخدوش شدن آنها قرار می‌گیرند.

مهمترین کاستی که عموماً به آزمون‌های توزیع-آزاد نسبت داده می‌شود توان کمتر آنها نسبت به آزمون پارامتریک نظیر آن است. در کل، وقتی پیش‌فرض‌های آزمون‌های پارامتریک برآورده شود، آزمون توزیع-آزاد برای یک سطح مشابه توان، به مشاهدات بیشتری نیاز دارد تا آزمون پارامتریک. بنابراین برای یک مجموعه معین از داده‌ها احتمال بیشتری وجود دارد که آزمون پارامتریک به رد فرض صفر

نادرست منتهی شود تا آزمون توزیع-آزاد. از این گذشته حتی وقتی پیش‌فرض‌های توزیع تا حد متوسطی زیر پا گذاشته شوند، به نظر می‌رسد آزمون پارامتریک برتری خود را حفظ می‌کند.

اغلب ادعا می‌شود شیوه‌های توزیع-آزاد به خاطر سادگی محاسبات از سودمندی خاصی برخوردارند. اما برای آزمایشگری که وقت زیادی صرف جمع‌آوری داده‌های خود کرده است، صرف چند دقیقه اضافی برای محاسبات، به سختی می‌تواند استفاده از یک آزمون کم اهمیت‌تر را توجیه کند.

برتری دیگری برای آزمون‌های توزیع-آزاد وجود دارد. چون بسیاری از آنها نمرات خام را رتبه‌بندی و روی رتبه‌ها عمل می‌کنند، آزمونی را برای تفاوت‌ها ارائه می‌کنند که تحت تأثیر یک یا تعداد اندکی نمرات خیلی انتهایی توزیع (دور افتاده‌ها) قرار نمی‌گیرد. یک نمره دور افتاده در یک مجموعه داده می‌تواند عملاً آزمون پارامتریک را کم توان سازد، زیرا واریانس را متورم نموده و از این رو عبارت خطا افزایش می‌یابد، و همچنین باعث سوگیری میانگین با کشیده شدن آن به سمت نمرات دور افتاده می‌شود (این کار می‌تواند میانگین را کاهش یا افزایش دهد).

این فصل به چهار روش از مهمترین روش‌های توزیع-آزاد می‌پردازد. دو روش اول مشابه آزمون t هستند- یکی برای نمونه‌های مستقل و یکی برای نمونه‌های جور شده (همبسته). دو آزمون بعدی مشابه توزیع-آزاد برای تحلیل واریانس هستند- اولی برای k گروه مستقل و دومی برای k اندازه‌گیری مکرر. همه این آزمون‌ها اعضای یک دسته معروف به **آزمون‌های رتبه تصادفی**^۱ هستند چون با داده‌های رتبه‌ای سروکار دارند و به عنوان توزیع آزمون آماری آنها به حساب می‌آیند وقتی فرض صفر درست است، توزیع نظری رتبه‌های دارای توزیع تصادفی است. از آنجا که این آزمون‌ها داده‌های خام را تبدیل به رتبه می‌کنند، شکل توزیع زیربنایی نمرات در جمعیت اهمیت زیادی ندارد. بنابراین هر دو مجموعه نمرات زیر

۱۱ ۱۴ ۱۵ ۱۶ ۱۷ ۲۲

(داده‌هایی که ممکن است از یک توزیع نرمال بدست آمده باشد)

۱۱ ۱۲ ۱۳ ۳۰ ۲۱ ۳۲

و

(داده‌هایی که ممکن است از یک توزیع دو نمایی بدست آمده باشد)

به این رتبه‌ها منجر می‌شوند: ۱ ۲ ۳ ۴ ۵ ۶

۲۰-۱ آزمون من-ویتنی

یکی از متداول‌ترین و شناخته‌شده‌ترین آزمون‌های توزیع-آزاد، آزمون من-ویتنی برای دو نمونه مستقل است. این آزمون به عنوان همتای توزیع-آزاد برای آزمون t دو نمونه مستقل است، هر چند فرض صفر تا اندازه‌ای متفاوت و گسترده‌تری را می‌سنجد. فرض صفر آن این فرضیه است که دو نمونه به صورت

تصادفی از جمعیت‌های یکسانی بیرون کشیده شده‌اند (نه صرفاً از جمعیت‌هایی با میانگین یکسان)، اما حساسیت خاصی نسبت به تفاوت‌های جمعیتی در گرایش مرکزی دارد. بنابراین رد H_0 عموماً به این معنی تفسیر می‌شود که دو توزیع دارای گرایش مرکزی متفاوتی هستند، اما امکان دارد رد فرض صفر عملاً ناشی از تفاوت‌های دیگر بین جمعیت‌ها باشد. توجه داشته باشید که وقتی از یک چیز نفع می‌بریم (رهایی از پیش‌فرض‌ها)، ممکن است با چیز دیگری هزینه آن را بپردازیم (از دست رفتن ویژه بودن).

آزمون من-ویتی یک تغییر روی آزمونی است که در اصل توسط ویل کاکسون طراحی شد و آزمون جمع-رتبه نام داشت. چون ویل کاکسون یک آزمون دیگر هم طرح کرده که در قسمت بعد مورد بحث قرار می‌گیرد، ما از این آزمون تحت عنوان من-ویتی نام می‌بریم تا اشتباه نشود. هر چند آزمونی که بوسیله من و ویتی طرح شده دارای آماری تاحدودی متفاوت است، از آماره مورد استفاده در این فصل (اصل جمع رتبه‌های نمرات در یکی از گروه‌ها) بیشتر حمایت شده چون خیلی راحت‌تر حساب می‌شود. هر دو آماره آزمون وقتی روی مجموعه داده‌های یکسانی اجرا شوند نتیجه مشابهی بدست می‌دهند.

فهمیدن منطق زیربنایی آزمون من-ویتی خیلی آسان است. فرض کنید ما دو گروه عملکرد مستقل داریم، با n_1 مشاهده در گروه ۱ و n_2 مشاهده در گروه ۲. همچنین فرض کنید که فرض صفر در حد قابل توجهی نادرست است و اینکه جمعیتی که نمرات گروه ۱ از آن نمونه‌گیری شده شامل مقادیری است که عموماً پایین‌تر از جمعیتی است که نمرات گروه ۲ از آن برآمده‌اند. آنگاه چنانچه ما تمام $N = n_1 + n_2$ نمره را از کوچک به بزرگ بدون توجه به عضویت گروهی مرتب کنیم، انتظار خواهیم داشت که رتبه‌های پایین‌تر عموماً در نمرات گروه ۱ و رتبه‌های بالاتر در نمرات گروه ۲ قرار گیرند. یک مرحله بعد، اگر ما رتبه‌های اختصاص یافته به هر گروه را جمع کنیم، انتظار خواهیم داشت حاصل جمع رتبه‌های گروه ۱ آشکاراً کوچکتر از حاصل جمع رتبه‌های اختصاص یافته به گروه ۲ باشد.

اکنون به موردی توجه کنید که فرض صفر درست است و نمرات برای دو گروه از جمعیت‌های یکسان بدست آمده باشد. در این حالت اگر بخواهیم همه N نمره را بدون توجه به عضویت گروهی رتبه‌بندی کنیم، انتظار خواهیم داشت تعدادی رتبه‌ی پایین و تعدادی رتبه‌ی بالا در هر گروه باشد و حاصل جمع رتبه‌های اختصاص یافته به گروه ۱ خیلی نزدیک به حاصل جمع رتبه‌های اختصاص یافته به گروه ۲ باشد. این حالت‌ها در جدول ۱-۲۰ ترسیم شده است.

من و ویتی (و ویل کاکسون) آزمون‌های خود را بر منطقی بنا کردند که اینجا مورد بحث قرار گرفت، یعنی استفاده از حاصل جمع رتبه‌ها در یکی از گروه‌ها به عنوان آزمون آماری. اگر آن حاصل جمع نسبت به دیگری خیلی کوچک باشد، ما فرض صفر را رد می‌کنیم. به بیان روشن‌تر، ما مجموع رتبه‌های اختصاص یافته به گروه کوچک‌تر را به عنوان آزمون آماری خود می‌گیریم، یا، اگر $n_1 = n_2$ ، حاصل جمعی

که کوچک تر است را به عنوان آزمون آماری خود می گیریم. با داشتن این مقدار، ما می توانیم جداول آماری من-ویننی (W_s) را برای آزمون فرض صفر به کار بگیریم.

جدول ۲۰-۱ ترسیم نوعی نتایج مورد انتظار تحت H_0 نادرست و H_0 درست

<u>H₀ نادرست</u>						گروه ۱						
گروه ۲												
۲۷	۱۸	۳۳	۲۵	۲۶	۳۰	۲۰	۱۹	۱۳	۱۷	۱۲	۱۰	داده‌های خام
۱۰	۵	۱۲	۸	۹	۱۱	۷	۶	۳	۴	۲	۱	رتبه‌ها (R _i)
۵۵						۲۳						ΣR _i
<u>H₀ درست</u>						گروه ۱						
گروه ۲												
۳۴	۲۳	۲۰	۲۹	۲۵	۱۸	۳۳	۲۴	۱۹	۳۲	۲۸	۲۲	داده‌های خام
۱۲	۵	۳	۹	۷	۱	۱۱	۶	۲	۱۰	۸	۴	رتبه‌ها (R _i)
۴۱						۳۷						ΣR _i

به منظور ارائه یک مثال ویژه، به داده های فرضی زیر در باره تعداد وقایع استرس زای اخیر زندگی گزارش شده بوسیله یک گروه از بیماران قلبی در یک بیمارستان محلی و یک گروه کنترل از بیماران ارتوپدی در یک بیمارستان کوچک توجه نمائید. این موضوع به خوبی معلوم شده که وقایع استرس زای زندگی (ازدواج، شغل جدید، مرگ همسر، و غیره) با بیماری همبستگی دارند، و منطقی است که انتظار داشته باشیم بسیاری بیماران قلبی در گذشته نزدیک وقایع استرس زای بیشتری را نسبت به بیماران ارتوپدی تجربه کرده باشند (که هنگام تخریب یک ساختمان مچ پایشان شکسته یا موقع اسکی استخوان ترقوه آنها شکسته است). از این داده ها معلوم می گردد که این انتظار حمایت می شود. از آنجا که ما شواهدی داریم که نشان می دهند نمرات استرس زندگی احتمالاً دارای توزیع متقارن نیستند (به خصوص برای بیماران قلبی اگر فرض تحقیق ما درست باشد)، ما یک آزمون توزیع-آزاد را برای استفاده برمی گزینیم. در این مورد ما آزمون من-ویننی را به کار خواهیم برد زیرا دو گروه مستقل داریم.

بیماران قلبی						بیماران ارتوپدی					
۰	۵	۲۹	۷	۸	۳۲	۶	۳	۲	۲	۱	داده های خام
۱	۶	۱۰	۸	۹	۱۱	۷	۵	۳/۵	۳/۵	۲	رتبه ها (R_i)

برای به کارگیری آزمون من-ویتی ابتدا همه ۱۱ نمره را از پایین به بالا رتبه‌بندی نموده، رتبه‌های مشترک را به نمره‌های مشترک اختصاص می‌دهیم (رجوع کنید به مبحث رتبه‌گذاری در فصل ۹). گروه ارتوپدی کوچکتر است و اگر این بیماران عموماً وقایع استرس‌زای کمتری داشته‌اند، آنگاه حاصل جمع رتبه‌های اختصاص یافته به آن گروه بایستی نسبتاً کم باشد. با در نظر گرفتن حرف W_s نشان‌گر حاصل جمع رتبه‌ها در گروه کوچک‌تر (گروه ارتوپدی)، در می‌یابیم

$$W_s = 2 + 3/5 + 3/5 + 5 + 7 = 21$$

ما می‌توانیم مقدار بدست آمده W_s را با استفاده از جدول مربوط به آن در پیوست کتاب ارزیابی کنیم که کوچکترین مقدار W_s مورد انتظار بر اثر شانس در صورت درست بودن فرض صفر را نشان می‌دهد. در جدول W_s پیوست برای $n_1 = 5$ آزمودنی در گروه کوچکتر و $n_2 = 6$ آزمودنی در گروه بزرگتر (n_1 همیشه تعداد آزمودنی‌ها در گروه کوچکتر است) در سطح $\alpha = 0.025$ (یک دامنه) مقدار ۱۸ را می‌بینیم. یعنی برای معنادار بودن تفاوت بین گروه‌ها در سطح یک دامنه 0.025 یا سطح دو دامنه 0.05 ، بایستی W_s کمتر یا مساوی ۱۸ باشد. چون ما W_s را معادل ۲۱ یافتیم H_0 را نمی‌توانیم رد کنیم. (برای مقایسه، اگر ما یک آزمون t روی همین داده‌ها اجرا کنیم، با نادیده گرفتن این واقعیت که واریانس یک نمونه ۵۰ برابر دیگری است و اینکه داده‌ها نشان می‌دهند که پیش‌بینی ما در باره شکل توزیع نمرات قبلی درست است، t معادل $1/52$ با ۹ df خواهد بود که معنادار نیست).

اندازه‌های مندرج در جدول W_s پیوست کتاب برای یک آزمون یک دامنه است و تنها در صورتی به رد فرض صفر می‌انجامد که مجموع رتبه‌ها برای گروه کوچکتر به اندازه کافی کوچک باشد. اما امکان دارد که رتبه‌های بزرگتر در گروه کوچکتر به هم بپیوندند به صورتی که اگر H_0 نادرست باشد، حاصل جمع رتبه‌ها از انتظار شانس بزرگتر شود نه از گروه کوچکتر. یک راه دشوار دیگر برای این مسئله رتبه‌بندی دوباره همه داده‌ها است که این بار از بزرگتر به کوچکتر انجام شود. اگر این کار را بکنیم آنگاه رتبه‌های کوچکتر در گروه کوچکتر ظاهر خواهند شد و می‌توانیم روال قبل را ادامه دهیم. به هر حال ما مجبور نیستیم فرایند رتبه‌بندی دوباره داده‌ها را تکرار کنیم. با استفاده از خصوصیات متقارن توزیع جمع رتبه‌ها ما می‌توانیم کار مشابهی را با محاسبه آماره‌ای بنام W'_s انجام دهیم. این آماره حاصل جمع رتبه‌ها برای گروه کوچکتر است که اگر رتبه‌ها را وارونه کنیم و از بزرگترین به کوچکترین رتبه‌بندی کنیم بدست خواهد آمد.

$$W'_s = 2\bar{W} - W_s$$

که $2\bar{W} = n_1(n_2 + n_1 + 1)$ و در جدول پیوست آمده است. سپس می‌توانیم W'_s را در مقایسه با مقدار جدول ارزیابی کرده و یک آزمون یک دامنه روی دنباله بالایی توزیع صورت دهیم. برای آزمون دو دامنه H_0 (چیزی که ما معمولاً می‌خواهیم) W_s و W'_s را حساب کرده، با هر کدام کوچکتر بود وارد جدول می‌شویم، و مقدار مربوط به α را دو برابر می‌کنیم.

برای نشان دادن W_s و W'_s ، به دو مجموعه داده های زیر توجه نمائید:

گروه ۲					گروه ۱				مجموعه ۱
۸۲	۳۷	۲۵	۲۳	۱۸	۱۹	۱۶	۱۵	۲	X
۹	۸	۷	۶	۴	۵	۳	۲	۱	رتبه ها
$W'_s = ۲۹$									$W_s = ۱۱$

گروه ۲					گروه ۱				مجموعه ۲
۴	۱۴	۱۵	۱۸	۲۳	۲۱	۲۴	۴۰	۶۰	X
۱	۲	۳	۴	۶	۵	۷	۸	۹	رتبه ها
$W'_s = ۱۱$									$W_s = ۲۹$

توجه داشته باشید که این دو مجموعه داده ها میزان یکسانی از دور افتادگی^۱ یا نمرات انتهایی را نشان می دهند، به طوری که در مجموعه اول چهار تا از پنج رتبه کوچکتر در گروه ۱ هستند، و در مجموعه دوم چهار تا از پنج رتبه بزرگتر در گروه ۱ هستند. علاوه بر این، W_s برای مجموعه ۱ برابر با W'_s برای مجموعه ۲ است، و برعکس. بنابراین اگر ما این قاعده را طرح کنیم که ما هر دو W_s و W'_s را برای گروه کوچکتر حساب کنیم و به مقدار کوچکتر W_s و W'_s در جدول مراجعه کنیم، به نتیجه یکسانی در رابطه با دو مجموعه داده ها خواهیم رسید.

تقریب نرمال

جدول W_s برای همه مواردی که در آنها n_1 و n_2 کمتر یا مساوی ۲۵ باشند مناسب است. برای مقادیر بزرگتر n_1 و n_2 می توانیم از این واقعیت استفاده کنیم که با افزایش حجم نمونه ها توزیع W_s به یک توزیع نرمال نزدیک می شود. این توزیع دارای

$$\frac{n_1(n_1+n_2+1)}{2} = \text{میانگین}$$

$$\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \text{و انحراف معیار}$$

است، چون توزیع نرمال بوده و ما میانگین و انحراف معیار آن را می دانیم (خطای معیار)، می توانیم Z را حساب کنیم:

$$Z = \frac{\text{میانگین - انحراف آماره}}{\text{انحراف معیار}}$$

$$= \frac{W_s - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

و از جدول توزیع نرمال یک مقدار نزدیک به احتمال حقیقی از یک مقدار W_s دست کم به کوچکی مقدار بدست آمده پیدا کنیم.

برای نشان دادن محاسبات مربوط به حالتی که رتبه‌های بزرگتر در گروه‌های کوچکتر قرار می‌گیرند و نشان دادن استفاده از تقریب نرمال (گرچه عملاً ما نیازی به تقریب در نمونه‌های کوچکی از این قبیل نداریم)، توجه شما را به داده‌های جدول ۲-۲۰ جلب می‌کنیم. این داده‌های فرضی (اما قابل قبول) پیرامون وزن هنگام تولد (به گرم) نوزادان مادرانی هستند که تا سه ماه آخر بارداری تحت مراقبت نبودند و نوزادان مادرانی که از ابتدای بارداری تحت مراقبت بودند.

جدول ۲-۲۰ داده‌های فرضی وزن هنگام تولد نوزادان مادران با شرایط مختلف مراقبت

شروع مراقبت			
سه ماهه اول		سه ماهه آخر	
رتبه	وزن هنگام تولد	رتبه	وزن هنگام تولد
۱۰	۲۹۴۰	۲	۱۶۸۰
۱۶	۳۳۸۰	۱۷	۴۹۰۰
۱۸	۳۸۳۰	۱۴	۳۱۱۰
۹	۲۸۱۰	۵	۲۷۶۰
۸	۲۸۰۰	۳	۱۷۰۰
۱۵	۳۲۱۰	۷	۲۷۹۰
۱۳	۳۰۸۰	۱۲	۳۰۵۰
۱۱	۲۹۵۰	۴	۲۶۶۰
		۱	۱۴۰۰
		۶	۲۷۷۵

$$W_s = \sum (\text{رتبه‌ها در گروه } ۲) = 100$$

$$W'_s = 2\bar{W} - W_s = 152 - 100 = 52$$

$$Z = \frac{W_s - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

$$= \frac{100 - \frac{8(8+10+1)}{2}}{\sqrt{\frac{8(10)(8+10+1)}{12}}} = \frac{100 - 76}{\sqrt{126.6667}} = 2/13$$

برای داده‌های جدول ۲-۲۰ جمع رتبه‌های در گروه کوچکتر برابر ۱۰۰ است. از جدول مربوط به رتبه‌ها در بخش پیوست در می‌یابیم $2\bar{W} = 152$ و بنابراین $W'_s = 2\bar{W} - W_s = 52$. چون ۵۲ کوچکتر از ۱۰۰ است، ما با $W'_s = 52$ ، $n_1 = 8$ و $n_2 = 10$ وارد می‌شویم. (به عنوان نمونه کوچکتر تعریف می‌شود). از آنجا که ما یک آزمون دو دامنه می‌خواهیم، مقدار α جدول را دو برابر می‌کنیم. مقدار بحرانی W_s (یا W'_s) برای آزمون دو دامنه در سطح $\alpha = 0.05$ برابر ۵۳ است، یعنی اینکه تنها در ۵٪ اوقات ما انتظار داریم یک مقدار W_s یا W'_s کمتر یا مساوی ۵۳ باشد در صورتی که H_0 درست باشد. مقدار بدست آمده ما از W'_s برابر ۵۲ است، که بنابراین در منطقه رد واقع شده، و ما H_0 را رد خواهیم کرد. ما نتیجه خواهیم گرفت که مادرانی که تا سه ماهه آخر تحت مراقبت قرار نگرفته‌اند متمایل به داشتن نوزادانی با وزن کمتر هستند. معنای این یافته ممکن است این نباشد که عدم مراقبت تا سه ماهه آخر بارداری باعث نوزادان کم وزن می‌شود، بلکه تنها گویای این است که متغیرهای متداعی با تأخیر مراقبت (مثل مادران جوان، ضعف تغذیه، یا فقر) نیز با کم‌وزنی همراه هستند.

استفاده از تقریب نرمال برای ارزیابی W_s در قسمت پایین جدول ۲-۲۰ نشان داده شده است. در اینجا می‌بینیم که $z = 2/13$ از جدول پیوست در می‌یابیم که احتمال W_s یا W'_s به کوچکی ۵۲ (یک z در حدود انتهای $\pm 2/13$) برابر $0.33(0.166) = 0.2$. چون این مقدار کوچکتر از خط برش سنتی ۰/۰۵ = α است، ما H_0 را رد خواهیم کرد و دو باره نتیجه‌گیری می‌کنیم که شواهد کافی برای این وجود دارد که بگوییم نداشتن مراقبت اولیه با کم‌وزنی نوزادان در رابطه است. توجه داشته باشید که هر دو راه حل دقیق و تقریب نرمال به نتیجه مشابهی در رابطه با H_0 می‌رسند. (با تقریب نرمال لازم نیست W'_s محاسبه و استفاده شود چون استفاده از W_s به مقدار z یکسانی منجر می‌شود، به استثنای اینکه علامت آن معکوس می‌شود. برای شما آموزنده خواهد بود که t را برای این مجموعه داده‌ها حساب کنید).

چگونگی کار با هم‌رتبه‌ها

وقتی داده‌ها شامل نمرات هم‌رتبه^۱ باشند، هر آزمونی که متکی به رتبه‌ها باشد متمایل به این است که تا حدی مخدوش گردد. چند راه مختلف برای رفتار با نمرات هم‌رتبه وجود دارد. شما می‌توانید رتبه‌های یکسان را به نمرات یکسان اختصاص دهید (چنانکه ما انجام دادیم)، می‌توانید با یک سکه قرعه‌کشی

کنید و براساس آن رتبه‌های متوالی را به نمرات هم‌رتبه اختصاص دهید، یا می‌توانید رتبه‌های مشترک را به هر طریقی که رد کردن H_0 را سخت‌تر کند تخصیص دهید. در عمل اغلب افراد بطور ساده رتبه‌های برابر را اختصاص می‌دهند. هر چند این کار ممکن است بهترین راه آماری نباشد، اما خیلی مرسوم است و روشی است که ما اینجا به کار می‌بریم.

فرض صفر

آزمون من-ویتی این فرض را که دو مجموعه نمرات از جمعیت‌های یکسانی نمونه‌گیری شده‌اند ارزیابی می‌کند. این گسترده‌تر از فرض صفری است که با آزمون t معادل آن آزموده می‌شود، که بطور خاص با میانگین‌ها سروکار دارد (در اصل به عنوان نتیجه پیش‌فرض‌های زیربنایی که تفاوت نمرات دیگر را غیرمحتمل می‌شمارند). اگر فرض شود دو جمعیت دارای شکل و پراکندگی یکسانی هستند، آنگاه فرض صفر مورد آزمون بوسیله آزمون من-ویتی در واقع با گرایش مرکزی (در این حالت با میانه‌ها)ی دو جمعیت سروکار دارد، و اگر جمعیت‌ها متقارن هم باشند، آزمون از نوع میانگین‌ها خواهد بود. در هر مورد آزمون من-ویتی نسبت به تفاوت‌های گرایش مرکزی بطور خاصی حساس است.

۲-۲۰ آزمون جفت‌های همتای ویل کاکسون برای رتبه‌های علامت‌دار

شهرت ویل کاکسون به خاطر توسعه معروف‌ترین آزمون توزیع-آزاد برای گروه‌های مستقل است، که آن را آزمون من-ویتی نامیدیم چون بوسیله آنها بسط یافت. او همچنین مشهورترین آزمون را برای گروه‌های همتا (یا نمرات جور شده) طرح نمود. این آزمون توزیع-آزاد معادل آزمون t برای نمونه‌های وابسته است، و این فرض صفر را می‌آزماید که دو نمونه وابسته (جور شده) یا از جمعیت‌های یکسان یا از جمعیت‌های متقارن با میانگین یکسان بدست آمده‌اند. به طور خاص‌تر، این آزمون این فرض صفر را می‌آزماید که توزیع نمرات اختلاف (در جمعیت) در اطراف صفر متقارن است. این همان فرضیه‌ای است که بوسیله آزمون t معادل هنگامی که پیش‌فرض نرمال بودن برآورده شود آزمایش می‌شود.

بسط منطق پشتوانه آزمون جفت‌های همتای ویل کاکسون برای رتبه‌های علامت‌دار^۱ سر راست بوده و با یک مثال ساده قابل توضیح است. فرض کنید ما می‌خواهیم این فرضیه را که "یک برنامه طولانی مدت دویدن، فشار خون را کاهش می‌دهد" بیازمائیم. برای آزمودن این فرضیه ما فشار خون عده‌ای از افراد را اندازه می‌گیریم، از آنها تقاضا می‌کنیم تا در یک برنامه منظم دویدن برای شش هفته شرکت کنند، و دوباره فشار خون آنها را در پایان دوره اندازه‌گیری می‌کنیم. متغیر وابسته ما تغییر در فشار خون طی یک دوره شش هفته‌ای است. اگر دویدن، فشار خون را کاهش دهد، انتظار داریم بیشتر آزمودنی‌ها در مرحله دوم فشار خون کمتر و یک تفاوت مثبت در پیش‌آزمون-پس‌آزمون را نشان دهند.

ما همچنین انتظار داریم که آنهایی که فشار خونشان بالا رفته (و در نتیجه دارای تفاوت منفی پیش-آزمون-پس آزمون هستند) تنها کمی بالاتر باشند. از سوی دیگر، اگر دویدن به عنوان روشی برای کاهش فشار خود بی‌فایده باشد، آنگاه حدود نیمی از نمرات اختلاف، مثبت و نیمی منفی خواهند بود، و نمرات مثبت اختلاف تقریباً به اندازه نمرات منفی خواهند بود. به عبارت دیگر اگر واقعاً H_0 درست باشد، دیگر انتظار نخواهیم داشت که بیشتر تغییرات در جهت پیش‌بینی شده و تنها تغییرات کوچکی بر خلاف پیش‌بینی باشند.

همانگونه که در مثال عددی زیر نشان داده شده است، برای اجرای آزمون جفت‌های همتای رتبه‌های علامت‌دار ما ابتدا نمرات اختلاف را برای هر جفت از اندازه‌ها حساب می‌کنیم. سپس همه نمرات اختلاف را بدون توجه به علامت آنها رتبه‌گذاری نموده، آنگاه علامت جبری نمرات اختلاف را به خود رتبه‌ها تخصیص می‌دهیم، و در آخر رتبه‌های مثبت و منفی را بطور جداگانه جمع می‌کنیم. آزمون آماری (T) به عنوان کوچکترین مقدار مطلق (یعنی بدون علامت) دو حاصل جمع بدست می‌آید، و در مقایسه با مقدار جدول T پیوست مورد ارزیابی قرار می‌گیرد. (لازم به توجه است که در محاسبه T ما علامت‌های جبری را تنها برای راحتی کار به رتبه‌ها می‌چسبانیم. ما می‌توانیم به طور ساده مثلاً دور رتبه‌هایی که بالا رفته‌اند دایره و زیر رتبه‌هایی که پایین آمده‌اند خط بکشیم، تلاش ما صرفاً این است که بین دو حالت تفاوت قائل شویم.)

فرض کنید تحقیقی که قبلاً توضیح داده شد داده‌های زیر را در باره فشار خون انقباضی قبل و بعد از شش هفته آموزش بدست دهد:

قبل (B)	۱۳۰	۱۷۰	۱۲۵	۱۷۰	۱۳۰	۱۴۵	۱۶۰
بعد (A)	۱۲۰	۱۶۳	۱۲۰	۱۳۵	۱۴۳	۱۳۶	۱۴۰
اختلاف (B-A)	۱۰	۷	۵	۳۵	-۱۳	-۶	۴۰
اختلاف رتبه	۵	۴	۲	۷	۶	۳	۸
رتبه علامت‌دار	۵	۴	۲	۷	-۶	-۳	۸
$T_+ = ۲۷$ (رتبه‌های مثبت)							
$T_- = -۹$ (رتبه‌های منفی)							

دو ردیف اول در بردارنده فشار خون آزمودنی‌ها در قبل و بعد از برنامه شش هفته‌گی دویدن است. ردیف سوم شامل نمرات اختلاف حاصل از کم کردن نمره "بعد" از "قبل" است. قابل توجه است که فقط دو آزمودنی تغییر منفی داشته‌اند- یعنی فشارشان افزایش یافته است. چون به نظر نمی‌رسد این نمرات اختلاف، بازتابی از توزیع نزدیک به نرمال جمعیت باشند، ما تصمیم گرفتیم از یک آزمون توزیع-آزاد

استفاده کنیم. در ردیف چهارم همه نمرات اختلاف رتبه‌بندی شده‌اند بدون توجه به جهت تغییر، و در ردیف پنجم علامت مناسب به رتبه‌ها داده شده تا بین کسانی که فشارخون آنها کاهش یافته و کسانی که فشارخون آنها افزایش یافته فرق گذاشته شود. در پایین جدول حاصل جمع رتبه‌های مثبت و منفی (T_+ و T_-) را می‌بینیم. چون T به عنوان قدر مطلق کوچکتر T_+ و T_- تعریف شده، $T = 9$ است.

برای ارزیابی T به جدول پیوست رجوع می‌کنیم که یک بخش آن در جدول ۳-۲۰ نشان داده شده است. این جدول نسبت به موارد قبلی کمی فرق دارد. راحت‌ترین راه برای فهم رقم‌هایی که در جدول آمده از راه تشابه است. فرض کنید می‌خواهید درستی سکه‌ای را که هشت بار پرتاب می‌کنید بیازمائید و فرض صفر را در سطح $\alpha = 0.05$ (یک دامنه) رد کنید، که آیا تعداد شیرها خیلی کم هستند. از هشت پرتاب یک سکه مجموعه نتایجی که دقیقاً دارای احتمال 0.05 تحت H_0 باشد وجود ندارد. احتمال یک شیر یا کمتر 0.0352 است و احتمال دو شیر یا کمتر 0.1445 (خیلی بیشتر از 0.05) است. بنابر این اگر ما بخواهیم در سطح $\alpha = 0.05$ کار کنیم، ما می‌توانیم یک شیر یا کمتر را رد کنیم که در این حالت احتمال خطای نوع I عملاً 0.0352 است (کمتر از 0.05)، یا می‌توانیم دو شیر یا کمتر را رد کنیم که در این حالت احتمال خطای نوع I عملاً 0.1445 است (خیلی بیشتر از 0.05). همین نوع مشکل در رابطه با T پیش می‌آید چون، این شبیه به توزیع دونمایی که احتمالات شیر و خط را بدست می‌دهد، یک توزیع انحصاری است.

در جدول T بخش پیوست برای آزمون یک دامنه در سطح $\alpha = 0.025$ (یا یک آزمون دو دامنه در سطح $\alpha = 0.05$) با $n = 8$ رقم‌های جدول $3(0.0195)$ و $4(0.0273)$ هستند. این به ما می‌گوید که اگر بخواهیم در یک سطح (یک دامنه) $\alpha = 0.025$ کار کنیم ما می‌توانیم یا H_0 را برای $T \leq 3$ رد کنیم (که در آن حالت α عملاً برابر 0.0195 است) یا می‌توانیم H_0 را برای $T \leq 4$ رد کنیم (که در آن حالت α عملاً برابر 0.0273 است). از آنجا که ما یک آزمون دو دامنه را می‌خواهیم، احتمالات باید دو برابر شده به $3(0.0390)$ و $4(0.05463)$ برسد. چون ما یک مقدار T معادل ۹ بدست آوردیم، هر یک از این نقاط برش را بگیریم نمی‌توانیم H_0 را رد کنیم. بنابراین نتیجه خواهیم گرفت که دلیلی نداریم که تردید کنیم که فشار خون تحت تأثیر یک دوره کوتاه (شش هفته‌ای) دویدن روزانه قرار نمی‌گیرد.

هم‌رتبه‌ها

به دو طریق مختلف در داده‌ها می‌تواند موارد هم‌رتبه پیش آید. یکی اینکه یک آزمودنی دارای نمره یکسانی در قبل و بعد باشد، که نمره اختلاف صفر می‌شود و علامتی ندارد. در این حالت ما معمولاً آن آزمودنی را کنار گذاشته و اندازه نمونه را متناسب با آن کاهش می‌دهیم، گرچه یک سوگیری در داده‌ها پیش می‌آید.

علاوه براین، ما ممکن است نمرات اختلافی داشته باشیم که باعث رتبه بندی هم مرتبه شود. اگر هر دو نمره هم مرتبه دارای علامت یکسان باشند، ما می توانیم آنها را به هر طریقی که بخواهیم جدا کنیم (یا تخصیص دهیم) بدون اینکه روی نتیجه نهایی تأثیر بگذارد. اگر نمرات دارای علامت ناهمسان باشند، ما معمولاً رتبه های هم مرتبه به آنها می دهیم و روند معمول را دنبال می کنیم.

جدول ۳-۲۰ مقادیر بحرانی دامنه پایینی T و میزان احتمال مربوط به آنها. نمونه اختصاری جدول پیوست

آلفای نرمال (یک دامنه)

N	۰/۰۵		۰/۰۲۵		۰/۰۱		۰/۰۰۵	
	T	α	T	α	T	α	T	α
۵	۰	۰/۰۳۱۳						
	۱	۰/۰۶۲۵						
۶	۲	۰/۰۴۶۹	۰	۰/۰۱۵۶				
	۳	۰/۰۷۸۱	۱	۰/۰۳۱۳				
۷	۳	۰/۰۳۹۱	۲	۰/۰۲۳۴	۰	۰/۰۰۷۸		
	۴	۰/۰۵۴۷	۳	۰/۰۳۹۱	۱	۰/۰۱۵۶		
۸	۵	۰/۰۳۹۱	۳	۰/۰۱۹۵	۱	۰/۰۰۷۸	۰	۰/۰۰۳۹
	۶	۰/۰۵۴۷	۴	۰/۰۲۷۳	۲	۰/۰۱۱۷	۱	۰/۰۰۷۸
۹	۹	۰/۰۴۸۸	۵	۰/۰۱۹۵	۳	۰/۰۰۹۸	۱	۰/۰۰۳۹
	۹	۰/۰۶۴۵	۶	۰/۰۲۷۳	۴	۰/۰۱۳۷	۲	۰/۰۰۵۹
۱۰	۱۰	۰/۰۴۲۰	۸	۰/۰۲۴۴	۵	۰/۰۰۹۸	۳	۰/۰۰۴۹
	۱۱	۰/۰۵۲۷	۹	۰/۰۳۲۲	۶	۰/۰۱۳۷	۴	۰/۰۰۶۸
۱۱	۱۳	۰/۰۴۱۵	۱۰	۰/۰۲۱۰	۷	۰/۰۰۹۳	۵	۰/۰۰۴۹
	۱۴	۰/۰۵۰۸	۱۱	۰/۰۲۶۹	۸	۰/۰۱۲۲	۶	۰/۰۰۶۸
⋮	⋮		⋮		⋮		⋮	

تقریب نرمال

زمانی که اندازه نمونه بزرگتر از ۵۰ باشد، که یک محدودیت جدول پیوست است، برای ارزیابی T یک تقریب نرمال وجود دارد. ما می دانیم که توزیع نمونه گیری برای اندازه نمونه های بزرگ تقریباً دارای توزیع نرمال است با

$$\sqrt{\frac{n(n+1)(2n+1)}{24}} = \text{انحراف معیار} \quad \text{و} \quad \frac{n(n+1)}{4} = \text{میانگین}$$

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad \text{بنابراین می‌توانیم } Z \text{ را حساب کنیم}$$

و با استفاده از جدول مربوط به آن Z را ارزیابی می‌کنیم. این شیوه شباهت زیادی به شیوه مورد استفاده در آزمون من-ویتنی دارد و در اینجا آن را تکرار نمی‌کنیم.

۳-۲۰ تحلیل واریانس یک طرفه کروسکال-والیس

تحلیل واریانس یک طرفه کروسکال-والیس^۱ یک تعمیم مستقیم از آزمون من-ویتنی برای حالتی است که در آن سه گروه مستقل یا بیشتر وجود دارد. به این ترتیب این آزمون مشابه آزمون توزیع-آزاد تحلیل واریانس یک طرفه است که در فصل ۱۶ مورد بحث قرار گرفت. این آزمون این فرضیه را می‌آزماید که همه نمونه‌ها از جمعیت یکسانی بیرون کشیده شده‌اند و نسبت به تفاوت‌های گرایش مرکزی خیلی حساس است.

برای اجرای آزمون کروسکال-والیس ما بطور ساده همه نمره‌ها را بدون توجه به گروه عضویت آنها رتبه‌بندی می‌کنیم و سپس حاصل جمع رتبه‌ها را برای هر گروه حساب می‌کنیم. حاصل جمع‌ها با R_j نشان داده می‌شوند. اگر فرض صفر درست باشد، ما انتظار داریم R_j ها کم و بیش برابر باشند (جدا از تفاوت‌های ناشی از اندازه نمونه‌ها). یک اندازه از درجه‌ای که R_j از یکدیگر فرق دارند اینگونه حاصل می‌شود

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

که

$$k = \text{تعداد گروه‌ها}$$

$$n_j = \text{تعداد مشاهدات در گروه } j$$

$$R_j = \text{حاصل جمع رتبه‌ها در گروه } j$$

$$N = \sum n_j = \text{کل اندازه نمونه}$$

سپس H براساس توزیع χ^2 با $k-1$ df ارزیابی می‌شود.

به عنوان مثال فرض کنید که داده‌های جدول ۴-۲۰ نشان دهنده تعداد مسأله‌های صحیح حل شده حساب (از ۸۵ مسأله) است که آزمودنی‌هایی که به آنها داروی افسردگی، محرک یا یک دارونما داده شده طی یک ساعت انجام داده‌اند. توجه داشته باشید که در گروه افسردگی سه نفر از آزمودنی‌ها بیش از حد افسرده بوده‌اند که بتوانند کاری بکنند و در گروه محرک سه آزمودنی هر ۸۵ مسأله را حل کرده‌اند.

این داده ها غیرنرمال در نظر گرفته می شوند، و ما داده ها را به رتبه تبدیل نموده و از آزمون کروسکال-والیس استفاده می کنیم. محاسبات در زیر جدول آمده است. مقدار H بدست آمده $۱۰/۳۶$ است، که می توان آن را به عنوان x^2 با $df = ۳-۱$ در نظر گرفت. مقدار بحرانی $x^2_{0.05}(۲)$ در جدول معادل $۵/۹۹$ است. چون $۱۰/۳۶ > ۵/۹۹$ است، ما می توانیم H_0 را رد کرده و نتیجه بگیریم که سه دارو به میزان عملکردهای متفاوتی منجر شده است.

جدول ۴-۲۰ آزمون کروسکال-والیس روی داده های حل مسأله

افسرده		محرك		دارونما	
رتبه	نمره	رتبه	نمره	رتبه	نمره
۹	۵۵	۱۵	۷۳	۱۱	۶۱
۱/۵	۰	۱۸	۸۵	۸	۵۴
۳	۱	۷	۵۱	۱۶	۸۰
۱/۵	۰	۱۲	۶۳	۵	۴۷
۶	۵۰	۱۸	۸۵		
۱۰	۶۰	۱۸	۸۵		
۴	۴۴	۱۳	۶۶		
		۱۴	۶۹		
R_j	۳۵	۱۱۵		۴۰	

$$\begin{aligned}
 H &= \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1) \\
 &= \frac{12}{19(20)} \left(\frac{35^2}{7} + \frac{115^2}{8} + \frac{40^2}{4} \right) - 3(N+1) \\
 &= \frac{12}{380} (2228/125) - 60 = 70/36 - 6 = 10/36
 \end{aligned}$$

$$x^2_{0.05}(2) = 5/99$$

۴-۲۰ آزمون رتبه ای فریدمن برای k نمونه ی همبسته

آخرین آزمون مورد بحث در این فصل همتای توزیع -آزاد تحلیل واریانس یک طرفه اندازه گیری مکرر، آزمون رتبه ای فریدمن برای k نمونه ی همبسته^۱ است. این آزمون رابطه نزدیکی با یک تحلیل

واریانس استاندارد به کار رفته روی رتبه‌ها به جای نمرات خام دارد. این آزمونی برای این فرض صفر است که نمرات هر عملکرد از جمعیت‌های مشابه برآمده، و خیلی به تفاوت‌های جمعیت در گرایش مرکزی حساس است.

فرض کنید بخواهیم این فرض را بیازمائیم که کیفیت داوری یک سخنرانی وابسته به تعداد وسایل کمک نمایشی مورد استفاده است. آزمایشگر ۱۷ نفر را که برای گروه‌های تجارت محلی سخنرانی‌های متعددی در باره موضوعات مختلف می‌کنند جمع کرده است. هر سخنران متن یکسانی را برای سخنرانی برای سه شنونده مختلف اما برابر ارائه می‌کند - یک بار با کمک‌های نمایشی، یک بار با چند اسلاید روی پروژکتور در باره نکات اصلی، و یک بار با پروژکتور و نمودارهای چرخان برای نشان دادن هر نکته‌ای که بیان می‌شود. در انتهای هر سخنرانی از شنونده خواسته می‌شود تا سخنرانی را روی یک مقیاس ۷۵ درجه‌ای نمره‌گذاری کند، و میانگین رتبه‌های همه افراد شنونده به عنوان متغیر وابسته در نظر گرفته شده است. چون برای هر سه شرایط سخنرانان مشابهی بودند، انتظار داریم داده‌ها با هم همبسته باشند. سخنرانان جدی همچنان جدی هستند فرق نمی‌کند که چه وسایل کمکی به کار ببرند. داده‌های فرضی در جدول ۵-۲۰ آمده است، که یک نمره بالاتر نشان دهنده رتبه‌ی مطلوب‌تر است. رتبه‌بندی نمرات خام برای هر آزمودنی در پرانتز نشان داده شده است.

چنانچه فرض صفر درست باشد، انتظار داریم رتبه‌بندی در هر سخنرانی دارای توزیع تصادفی باشد. بنابراین یک سخنران ممکن است بدون وسایل کمکی بهترین باشد، دیگری با وسایل زیاد کمکی بهترین باشد، و الی آخر. اگر این وضع باشد، حاصل جمع رتبه‌بندی در هر شرایط (ستون) تقریباً برابر خواهد بود. از سوی دیگر اگر تعداد کم وسایل کمک نمایشی منجر به معروف‌ترین سخنرانی شود، آنگاه بیشتر سخنرانان بالاترین رتبه خود را در این شرایط خواهند داشت، و حاصل جمع رتبه‌های این سه شرایط در عمل نابرابر خواهند بود.

برای اعمال آزمون فریدمن ما نمرات خام را برای هر سخنران بطور جداگانه رتبه‌بندی می‌کنیم و سپس رتبه‌بندی‌های هر شرایط را جمع می‌کنیم. آنگاه تغییرپذیری حاصل جمع‌ها را مورد ارزیابی قرار می‌دهیم از طریق محاسبه

$$x_F^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3N(k+1)$$

که

R_j = حاصل جمع رتبه‌ها برای شرایط j

k = تعداد شرایط

N = تعداد شرایط

این مقدار x_F^2 را می‌توان با توجه به توزیع استاندارد χ^2 با $k-1$ df ارزیابی کرد.

برای داده های جدول ۵-۲۰، $x_F^2 = ۱۰/۹۴$ روی $df = ۲$ ما H_0 را رد کرده و نتیجه می گیریم که کیفیت داوری یک سخنرانی تحت تأثیر میزان استفاده از وسایل کمک نمایشی فرق می کند. این داده ها بیانگر آن است که برخی وسایل کمکی سودمند هستند، اما میزان زیاد آنها می تواند باعث بازماندن از آنچه سخنران می گوید بشود. (تذکر: فرض صفری که ما اینک آزمودیم چیزی در باره تفاوت های میان آزمودنی ها (سخنرانان) نمی گوید، و در واقع تفاوت های آزمودنی بوسیله شیوه رتبه بندی بطور کامل ناپدید می گردد).

۵-۲۰ خلاصه

این فصل مجموعه کوتاهی از شیوه های مورد نیاز برای دوری از پیش فرض های مربوط به جمعیت هایی که داده ها از آنها نمونه گیری می شوند را ارائه می کند. ما ابتدا آزمون من-ویتنی را بحث کردیم که، مشابه توزیع-آزاد آزمون t نمونه مستقل است. برای اجرای این آزمون بطور ساده داده ها را رتبه گذاری کرده و می پرسیم آیا توزیع رتبه ها گویای توزیعی است که در صورت درست بودن فرض صفر انتظار داریم. در مورد آزمون جفت های همتای ویل کاکسون برای رتبه های علامت دار هم منطق مشابهی به کار می رود. ما سپس دو آزمون توزیع-آزاد که شبیه به یک تحلیل واریانس روی اندازه های مستقل (کروسکال-والیس) و اندازه های تکراری (فریدمن) هستند مورد بحث قرار دادیم. هر چند آشنایی با این چهار آزمون مهم است زیرا خیلی مورد استفاده قرار می گیرند، در ابتدای فصل گفته شد که مزایایی که با کاهش اثر پیش فرض ها از آن بهره مند می شویم ممکن است ارزش از دست دادن توان را نداشته باشد که اغلب استفاده از آزمون های توزیع-آزاد را قوت می بخشد. برداشت افراد از این پرسش هر چه باشد، اصل عمومی که باقی می ماند این است که ترجیح در به کارگیری و تفسیر هر شیوه آماری توجیه آماری نیست، بلکه جنبه عام قضیه و عقل جمعی مد نظر است. برخی از اصطلاحات مهم این فصل عبارتند از:

- آزمون های پارامتریک
- آزمون های غیر پارامتریک
- آزمون های رتبه تصادفی شده
- آزمون جفت های همتای ویل کاکسون
- آزمون من-ویتنی
- آزمون تحلیل واریانس یک طرفه
- آزمون رتبه های علامت دار
- آزمون رتبه فریدمن برای k نمونه همبسته
- آزمون من-ویتنی (توزیع-آزاد)
- آزمون تحلیل واریانس یک طرفه
- آزمون رتبه های علامت دار
- آزمون رتبه فریدمن برای k نمونه همبسته

۶-۲۰ تمرین‌ها

۱-۲۰ مک‌کانوفی معتقد است کودکان خردسال داستان‌ها را در عبارات توصیفی ساده (”و بعد...“) مدل‌ها سازمان‌دهی می‌کنند، در حالی که کودکان بزرگتر از جملات علت و معلولی و استنتاج اجتماعی استفاده می‌کنند. فرض کنید که شما از دو گروه از کودکان سنین مختلف خواسته‌اید یک داستان را که می‌خوانند خلاصه کنند. سپس ما تعداد جملات آن را که بتوان استنتاجی خواند شمارش کرده‌ایم. داده‌های به این شرح است:

کودکان خردسال	۰	۱	۰	۳	۲	۵	۲
کودکان بزرگتر	۴	۷	۶	۴	۸	۷	۲

(الف) این داده‌ها را با استفاده از آزمون دو دامنه جمع رتبه تحلیل کنید.

(ب) چه نتیجه‌ای می‌گیرید؟

۲-۲۰ کاپ و همکاران (۱۹۷۹) نشان دادند که آسیب به آمیگدال می‌تواند باعث کاهش پاسخ‌های معمول مربوط به ترس شود (مثلاً، کاهش در ضربان قلب). اگر ترس واقعا کاهش پیدا کند، آنگاه آموزش یک پاسخ اجتنابی به حیوان آسیب دیده مشکل‌تر خواهد شد زیرا پرهیز از محرک کاهش می‌یابد. فرض کنید دو گروه از خرگوش‌ها: یک گروه دارای آسیب آمیگدال، و دیگری یک گروه کنترل دست نخورده داشته باشیم. داده‌های زیر تعداد تمرین‌های یادگیری یک پاسخ اجتنابی برای هر حیوان را نشان می‌دهند.

گروه دارای آسیب	۱۵	۱۴	۱۵	۸	۷	۲۲
	۱۸	۱۹	۳۶	۱۷	۱۴	
گروه گواه	۹	۴	۹	۱۰	۶	۶
	۴	۵	۹			

(الف) داده‌ها را با استفاده از آزمون من-ویتنی (دو دامنه) تحلیل کنید.

(ب) چه نتیجه‌ای می‌گیرید؟

۳-۲۰ تحلیل تمرین ۲-۲۰ را با استفاده از تقریب نرمال تکرار کنید.

۴-۲۰ تحلیل تمرین ۲-۲۰ را با استفاده از آزمون یک دامنه مناسب تکرار کنید.

۵-۲۰ نورکمب و فتنزری-کور (۱۹۷۹) معتقدند که آموزش در شیوه‌های تشخیصی بایستی متخصص را به تعمیم (و آزمایش) فرضیه‌های بیشتری در رسیدن به تصمیم در باره یک بیمار هدایت کند. فرض کنید ما ده دانشجوی رزیدنت روان‌پزشکی را که تازه وارد این دوره شده‌اند انتخاب کرده و از آنها

خواسته ایم یک نوار ویدیوی مصاحبه را تماشا کنند و هر چند دقیقه یک بار نظر خود را در باره آن بیمار ثبت کنند. سپس تعداد فرضیه های هر رزیدنت شامل نقطه نظرانی که نوشته است را بشماریم. این آزمایش در پایان دوره رزیدنتی دو باره با یک نوار مشابه تکرار شده است. داده ها به این شرح است:

آزمودنی	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
قبل	۸	۴	۲	۲	۴	۸	۳	۱	۳	۹
بعد	۷	۹	۳	۶	۳	۱۰	۶	۷	۸	۷

(الف) داده ها را با استفاده از آزمون جفت های همتای ویل کاکسون برای رتبه های علامت دار تحلیل کنید.

(ب) چه نتیجه گیری خواهید داشت؟

۲۰-۶ با رجوع به تمرین ۵-۲۰

(الف) تحلیل را با استفاده از تقریب نرمال تکرار کنید.

(ب) این دو جواب چقدر با هم هماهنگ هستند؟ چرا واقعا آنها هماهنگ نیستند؟

۲۰-۷ ادعا می شود که کودکان فرزند اول خانواده بیش از فرزندان بعدی مستقل هستند. فرض کنید که یک مقیاس ۲۵ سؤالی استقلال تهیه کرده ایم و هر یک از ۲۰ فرزند اول و خواهر و برادران بعدی آنها را با آن رتبه بندی کرده ایم. این کار را وقتی آنها به سن بزرگسالی رسیدند انجام داده ایم و از این رو اثر سن خنثی شده است. داده ها استقلال به شرح زیر است (نمره بالاتر نشان دهنده استقلال بیشتر):

جفت خواهر و برادر	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
فرزند اول	۱۲	۱۸	۱۳	۱۷	۸	۱۵	۱۶	۵	۸	۱۲
فرزند بعدی	۱۰	۱۲	۱۵	۱۳	۹	۱۲	۱۳	۸	۱۰	۸
جفت خواهر و برادر	۱۱	۱۲	۱۳	۱۴	۱۵	۱۶	۱۷	۱۸	۱۹	۲۰
فرزند اول	۱۳	۵	۱۴	۲۰	۱۹	۱۷	۲	۵	۱۵	۱۸
فرزند بعدی	۸	۹	۸	۱۰	۱۴	۱۱	۷	۷	۱۳	۱۲

(الف) داده ها را با استفاده از آزمون جفت های همتای ویل کاکسون برای رتبه های علامت دار تحلیل کنید.

(ب) چه نتیجه گیری خواهید داشت؟

۲۰-۸ با رجوع به تمرین ۷-۲۰ تحلیل را با استفاده از تقریب نرمال تکرار کنید.

۲۰-۹ نتایج تمرین ۷-۲۰ خط برش روشنی به خوبی فراهم نمی کند. تفاوت ها را به صورت تابعی از نمره فرزند اول رسم کنید. این شکل چه چیزی را نشان می دهد؟

۲۰-۱۰ فرق بین فرضیه صفر مربوط به آزمون من-ویتنی و آزمون t نظیر آن چیست؟

۱۱-۲۰ چه فرقی بین فرضیه صفر مورد بررسی در آزمون جفت‌های همتای ویل کاکسون برای رتبه‌های علامت‌دار و آزمون t نظیر آن است؟

۱۲-۲۰ یکی از ادعاهای مربوط به برتری آزمون‌های توزیع-آزاد این است که برای داده‌های مقیاس ترتیبی مناسب‌تر هستند. در فصل‌های قبلی به این موضوع اشاره کردیم. چه دلیلی می‌توانید ارائه کنید که این ادعا چندان درست نیست؟

۱۳-۲۰ چرا رد کردن فرض صفر با استفاده از آزمون t بیان دقیق‌تری است تا رد کردن فرض صفر با استفاده از آزمون توزیع-آزاد مناسب؟

۱۴-۲۰ سه استاد یک درس انگلیسی پایه را تدریس می‌کنند و همه ادعا می‌کنند که بهترین دانشجویان را تربیت کرده‌اند. برای بررسی موضوع هشت دانشجو از هر کلاس بطور تصادفی انتخاب و از آنها امتحان مشابهی گرفته و توسط استاد بی‌طرفی نمره‌گذاری شده‌اند. داده‌ها به این شرح است:

استاد الف	۸۲	۷۱	۵۶	۵۸	۶۳	۶۴	۶۲	۵۳
استاد ب	۵۵	۸۸	۸۵	۸۳	۷۱	۷۰	۶۸	۷۲
استاد ج	۶۵	۵۴	۶۶	۶۸	۷۲	۷۸	۶۵	۷۳

آزمون مناسب را اجرا و نتیجه‌گیری مناسبی به عمل آورید.

۱۵-۲۰ یک روان‌شناس که در یک خانه گروهی نوجوانان بزهکار کار می‌کند می‌خواهد نشان دهد که این کار در کاهش بزهکاری آنان موفقیت‌آمیز بوده است. او ده نوجوان را که پلیس شناسایی کرده و در خانه خودشان زندگی می‌کنند، و ده نوجوان مشابه را که در خانه‌های اصلاح رفتار زندگی می‌کنند و نیز ده نوجوان را که در خانه گروهی زندگی می‌کنند نمونه‌گیری کرد. به عنوان یک متغیر شاخص او رفتار فرار از مدرسه را (تعداد روزهایی که در ترم گذشته فرار کرده‌اند)، مورد استفاده قرار داد که از دفتر مدرسه بطور آماده بدست می‌آید. براساس داده‌های زیر نتیجه‌گیری مناسب را به عمل آورید.

خانه عادی	۱۵	۱۸	۱۹	۱۴	۵	۸	۱۲	۱۳	۷
خانه اصلاح رفتار	۱۶	۱۴	۲۰	۲۲	۱۹	۵	۱۷	۱۸	۱۲
خانه گروهی	۱۰	۱۳	۱۴	۱۱	۷	۳	۴	۱۸	۲

۱۶-۲۰ به عنوان یک روش جایگزین برای ارزیابی یک خانه گروهی، فرض کنید ۱۲ نوجوان را انتخاب کرده‌ایم که بزهکاری را کاهش داده‌اند. ما تعداد روزهای فرار از مدرسه (۱) طی ماه قبل از جای گرفتن

در خانه گروهی (۲) طی ماهی که آنها در این خانه زندگی کرده‌اند (۳) طی ماه بعد از ترک آن خانه را معلوم کرده‌ایم. داده‌ها عبارتند از:

نوجوان	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
قبل	۱۰	۱۲	۱۲	۱۹	۵	۱۳	۲۰	۸	۱۲	۱۰	۸	۱۸
در طی	۵	۸	۱۳	۱۰	۱۰	۸	۱۶	۴	۱۴	۳	۳	۱۶
بعد	۸	۷	۱۰	۱۲	۸	۷	۱۲	۵	۹	۵	۳	۲

آزمون فریدمن را به کار ببرید. چه نتیجه‌ای می‌گیرید؟

۱۷-۲۰ مطالعه توضیح داده شده در تمرین ۱۶-۲۰ نسبت به مطالعه توضیح داده شده در تمرین ۱۵-۲۰ چه مزیتی دارد؟

۱۸-۲۰ می‌توان آزمون فریدمن را برای داده‌های تمرین ۵-۲۰ به کار برد. در این صورت از چه چیز غفلت شده است؟

۱۹-۲۰ برای داده‌های تمرین ۵-۲۰ می‌توان گفت که سه نفر از ده رزیدنت در بار دوم فرضیه‌های کمتر و هفت نفر فرضیه‌های بیشتری استفاده کردند. این موضوع را می‌توان با χ^2 آزمون. این چه تفاوتی با آزمون فریدمن دارد که روی آن داده‌ها اجرا شد؟

۲۰-۲۰ تاریخچه فرضیه‌آزمایی آماری در واقع با یک آزمایش آزمون-تی (فیشر، ۱۹۳۵) آغاز شد، حال خوب است برای این کتاب با همان آزمون پایان پیدا کند. مهماندار یک کافه چای کوچک فکر نمی‌کند که مردم واقعا بتوانند تفاوت بین فنجان چای اول تهیه شده با یک کیسه چای را با فنجان دوم و سوم تهیه شده با همان کیسه بفهمند. او ۸ مارک مختلف چای کیسه‌ای انتخاب کرد، با هر یک سه فنجان تهیه و سپس از یک گروه مشتریان خواست هر فنجان را روی یک مقیاس ۲۰ درجه‌ای رتبه‌گذاری کنند (بدون اینکه بدانند کدام فنجان از کدام چای است). داده‌ها اینجا آمده است، که رتبه‌بندی بالاتر نشان دهنده چای بهتر است:

فنجان			
مارک چای	اول	دوم	سوم
۱	۸	۳	۲
۲	۱۵	۱۴	۴
۳	۱۶	۱۷	۱۲
۴	۷	۵	۴

۶	۳	۹	۵
۴	۹	۸	۶
۴	۳	۱۰	۷
۲	۱۰	۱۲	۸

آزمون فریدمن را به کار ببرید، نتیجه‌گیری مناسبی به عمل آورید.

فصل ۲۱

انتخاب تحلیل مناسب

بیشتر مطالب این کتاب متوجه ارائه و توضیح شیوه‌های معمول برای توصیف و تحلیل داده‌های آزمایشی بوده است. مهم است که بدانید چگونه این شیوه‌ها را به کار بگیرید، اما همان اندازه هم مهم است که بدانید چه وقت آنها را به کار ببرید. یکی از بزرگترین مشکلاتی که دانشجویان در برابر داده‌های واقعی با آن روبرو می‌شوند این است که کدام یک از شیوه‌های متعددی که یاد گرفته‌اند برای آن مجموعه داده‌ها مناسب است.

در فصل ۱ بحث کوتاهی در باره نمودار درختی شد که داخل جلد روی کتاب درج شده است. آن نمودار برای توجه دادن به این است که چه قضایی در انتخاب آزمون آماری دخالت دارند (مثل نوع داده‌ها، سؤال از رابطه یا تفاوت، تعداد گروه‌ها، و اینکه متغیرها مستقل هستند یا وابسته). این نمودار درختی خیلی خود-توضیح دهنده است، و خوب است که شما آن را مرور کنید تا مطمئن شوید فرقهایی را که در نظر گرفته شده می‌فهمید. با این وجود به کار بردن مؤثر چنین نموداری دشوار است مگر آنکه برای این کار تمرین کنید. مثال‌هایی که در ادامه می‌آید برای این تمرین طرح شده‌اند.

در اینجا ۲۰ مثال از مطالعات تحقیقاتی آمده است که از نشریات علمی گردآوری شده‌اند. هر یک از این مطالعات نمونه‌ای از یک تحقیق واقعی هستند که به داده‌هایی رسیده که فردی آن را باید تحلیل می‌کرده است. کار شما شناسایی شیوه آماری مناسب برای استفاده در هر مورد است. در بعضی موارد ممکن است بتوان از چندین شیوه استفاده کرد که همگی برای به کارگیری مناسب باشند، و در موارد دیگر ممکن است فضای عدم توافق در باره اینکه کدام شیوه بهتر خواهد بود حاکم باشد. از این گذشته در بعضی موارد شیوه مناسب ممکن است شامل محاسبه ساده یک یا چند آماره توصیفی باشد، در حاکمی که در موارد دیگر-بیشتر آنها- نوعی آزمون فرضیه فراخوانده می‌شود. در همه این موارد شما باید فرض کنید که پیش‌فرض‌های مورد نیاز برای آزمون‌های استاندارد پارامتریک برآورده شده‌اند مگر آنکه شما ادعای دیگری داشته باشید. برای برخی مثال‌ها ما گفته‌ایم که هر چه آزمایشگر بیاورد. این تنها برای جلب توجه شماست و بخشی از سؤال به حساب نمی‌آید.

در انتخاب مثال‌های آمده در اینجا گاهی آزمایش واقعی را تبدیل به روش‌های ساده‌تری کرده‌ایم - معمولاً با حذف متغیرهای مستقل یا وابسته. سعی شده تا ماهیت آزمایش تغییر مهمی پیدا نکند. شاید مایل باشید که برخی از آنها را ادامه دهید همه آنها در بخش فهرست منابع کتاب درج شده‌اند. اگر مایلید تمرین بیشتری کنید، به منابع نشریات علمی مراجعه کنید.

جواب هر کدام از این مثال‌ها در بخش انتهایی کتاب آمده است. چنانکه گفته شد ممکن است در بعضی موارد توافقی برای تحلیل مناسب نباشد در این موارد رویکرد ما رجوع به نظر آزمایشگر اصلی بوده که اشراف بیشتری بر داده‌ها داشته است. اگر جواب شما با ما فرق داشت، اطمینان حاصل کنید که علت آن جواب ما چیست و توجه داشته باشید به اینکه آیا جواب شما تنها یک راه متفاوت پاسخ به همان سؤال است، آیا به سؤال کاملاً متفاوتی جواب می‌دهد، یا آیا شما در به حساب آوردن بعضی چیزها اشتباه کرده‌اید.

۲۱-۱ تمرین‌ها

۲۱-۱ برنت، اسکوارتز و کیزر (۱۹۸۳) می‌خواهند بدانند آیا ده مقیاس روان‌شناختی مختلف سنجش افسردگی برای استفاده روی نوجوانان و جوانان مناسب هستند. بویژه می‌خواهند بدانند آیا آزمودنی‌ها یا مراجعین می‌توانند سؤال‌ها را بفهمند. آنها با استفاده از فرمول استاندارد توانایی خواندن، نمرات توانایی خواندن را برای هر یک از ده مقیاس افسردگی نمره‌گذاری کردند. بیشتر احتمال دارد آنها چه شیوه‌های آماری را به کار بگیرند؟

۲۱-۲ هارپر و واچر (۱۹۸۳) رابطه بین نمرات آزمون قریبال‌گری رشدی دینور و نمرات اندازه‌های هوش را روی ۵۵۵ کودک سه و چهار ساله بررسی کردند. چگونه آنها می‌توانستند این رابطه را به بهترین شکلی اندازه‌گیری کنند؟

۲۱-۳ آیا افراد به عکس‌هایی که در کتاب‌های روان‌شناسی عمومی چاپ می‌شوند توجهی دارند؟ گلدشتاین، بلیس و چانس (۱۹۸۳) به ۴۷ آزمودنی تعداد زیادی عکس دادند و از آنها خواستند عکس‌هایی را که می‌شناسند بیرون بکشند. بسیاری از عکس‌های بیرون آمده از کتاب روان‌شناسی عمومی بودند که دانشجویان از آن استفاده کرده بودند. آزمایشگران همچنین عکس‌ها را به ۵۶ دانشجو دادند که از کتاب دیگری استفاده کرده بودند. برای هر آزمودنی درصد عکس‌هایی که درست شناسایی کرده بودند را ثبت کردند. علاوه بر این از آزمودنی‌های گروه نخست خواستند تا مشخص کنند چه اندازه از عکس‌های کتاب به عنوان کمک آموزشی استفاده کردند. داده‌های مربوط به تشخیص و داده‌های مربوط به گزارش استفاده عکس‌ها را چگونه مورد تحلیل قرار می‌دهید؟

۲۱-۴ نیومن، السون، هال و هورناک (۱۹۸۳) به آزمودنی‌ها یک مجموعه کلمات هدف برای یادگیری ارائه کردند. هر کلمه هدف با یک واژه سرنخ که تداعی‌کننده قوی یا ضعیف کلمه هدف بود همراه بود

(مثل، میز- صندلی یا میز- گرد). از آزمودنی‌ها طی یک تکلیف یادآوری خواسته شد تا کلمات هدف را به یاد بیاورند و به آنها مجموعه متفاوتی از سرنخ‌ها داده شد که یا تداعی کننده خیلی قوی یا ضعیف هدف بودند، یا اصلاً سرنخی داده نمی‌شد. متغیر وابسته تعداد کلمات به یاد آورده شده بود. این داده‌ها را چگونه تحلیل می‌کنید؟

۵-۲۱ در مطالعه چیزی که یادگیری نهفته فضایی نامیده شده، ساترلند و لینگارد (۱۹۸۲) به موش‌ها یاد دادند که از موقعیت‌های مختلفی در یک حوضچه آب دایره‌ای به طرف یک سکوی زیرآبی (و بنابراین نامرئی) شنا کنند. بعضی موش‌ها حوضچه را هرگز ندیده بودند، بعضی قبلاً در موقعیت متفاوتی که مستقیماً به سمت سکو بود قرار داده شده بودند، و بعضی هم در همان مکان روی سکو قرار داده شده بودند. متغیر وابسته سرعت یادگیری شنا به سمت سکو بود. مناسب‌ترین تحلیل آماری برای این داده‌ها کدام است؟

۶-۲۱ لوندبرگ (۱۹۸۳) اصول آنچه که رفتار نوع A خوانده شده را مورد مطالعه قرار داد. او یک پرسشنامه را روی ۱۵ کودک سنین ۳ تا ۶ سال اجرا کرد و با توجه به مؤلفه‌های رقابتی بودن، بی‌صبری/خشم، و خشونت‌آمیز بودن رفتار نوع A به آنها نمره داد. و آنهایی را که پایین‌تر از میانه بودند به عنوان نوع B طبقه‌بندی کرد. (به آن "نیمه کردن میانه" گویند.) او سپس ضربان قلب و فشار خون را در یک موقعیت هیجانی اندازه‌گیری کرد. او چگونه باید این داده‌ها را برای هر یک از متغیرهای وابسته تحلیل کند؟ (توجه: تفاوت‌هایی وجود داشت اما فقط در فشار سیستولیک).

۷-۲۱ سوپرمانیام (۱۹۸۳) خطاهای ویراستاری را که خواننده‌های خوب و ضعیف در متن‌های ساده و دشوار از دست داده بودند مورد بررسی قرار داد. آزمودنی‌های یکسانی هر دو متن را ویراستاری کردند. ما شیوه‌های محاسبات این داده‌ها را پوشش نداده‌ایم، اما حتماً می‌توانید نوع تحلیل مورد نیاز را توضیح دهید.

۸-۲۱ لین و هوگ (۱۹۸۲) مرکز مهار را در آزمودنی‌های بیش‌فعال و گروه گواه مورد بررسی قرار دادند. مرکز مهار اشاره دارد به میزان مثبت و منفی دیدن پیامدها از لحاظ استناد به رفتارهای قابل مهار درونی خود مانند مهارت یا کار سخت، یا نسبت دادن به عوامل خارجی مانند شانس یا دشواری تکلیف، که فرد بر آنها کنترلی ندارد. شانزده آزمودنی بیش‌فعال و ۱۶ آزمودنی گواه مقیاس مرکز مهار نویکیس-تیک‌لند و آزمون لغات تصویری پی‌بادی (آزمون هوش عمومی) را تکمیل کردند. محققان با نمرات مرکز مهار چه باید بکنند و نمرات پی‌بادی چه استفاده‌ای دارد؟

۹-۲۱ ابرازیو، هانسن و هیت (۱۹۸۲)، ۱۵۳ کودک اسپانیولی را بر اساس آزمون جور کردن تصاویر آشنا (MFFT) به عنوان ضعیف در پردازش بینایی دسته‌بندی کردند. آنها سپس کودکان را به سه گروه عملکردی گماردند. یک گروه در پردازش بینایی از آموزش خصوصی برخوردار شدند، گروه دیگر از آموزش در گروه کوچک با مواد درسی عادی برخوردار بودند، و گروه سوم گروه گواه بود که چیز خاصی دریافت

نکرد. متغیر وابسته نمره کودک در مرحله دوم اجرای آزمون جور کردن تصاویر آشنا (MFFT) بود. چه تحلیلی را مناسب می‌دانید؟

۱۰-۲۱ پلینر (۱۹۸۲) علاقه‌مند بود که بداند آیا آشنایی با یک طعم باعث گرایش بیشتر می‌شود- پدیده مزه یادگرفته. او به ۲۴ دانشجوی کارشناسی طعم آب‌میوه‌های ناآشنا را برای ۵، ۱۰ یا ۲۰ بار خورد. سپس از آنها خواسته شد میزان علاقه به مزه آب‌میوه را رتبه‌گذاری کنند. چه شیوه آماری برای تحلیل این داده‌ها مناسب است؟ اگر او یک آزمون توزیع-آزاد بخواهد چه آزمونی را به کار خواهد برد؟ (توجه: او به اثر چیزی که انتظار داشت پی برد- آشنایی بیشتر سبب گرایش بیشتر می‌شود.)

۱۱-۲۱ فاگراستروم (۱۹۸۲) اثر استفاده از تفنگ نیکوتین به عنوان کمکی برای یک برنامه استاندارد ترک سیگار را مورد مطالعه قرار داد. یک گروه از برنامه درمان روان‌شناختی معمول درمانگاه او استفاده کردند. برای گروه دوم همان برنامه همراه با آدامس نیکوتین برای زمان نیاز به سیگار استفاده شد، هر گروه شامل ۵۰ آزمودنی بود و آزمودنی‌ها براساس پرهیز از استعمال سیگار برای یک و شش ماه طبقه‌بندی شدند. محقق چگونه باید این داده‌ها را تحلیل کند. این واقعیت که گروهی که آدامس آبنباتی قدیمی ساده را دریافت کند وجود نداشت باعث چه مشکلی می‌شود؟ (توجه گروه آزمایشی (آدامس) نرخ پرهیز معادل ۹۰٪ و ۶۴٪ به ترتیب برای یک و شش ماه داشت و در گروه گواه این نرخ ۶۰٪ و ۴۵٪ بود.) ۱۲-۲۱ پین (۱۹۸۲) از آزمودنی‌های زن و مرد خواست که ده ویژگی شغلی معمول (مثل حقوق، بار کاری و غیره) را از لحاظ اهمیت شخصی برای آزمودنی و برای فردی از جنس مخالف رتبه‌بندی کنند. داده‌ها از ۹۲ آزمودنی در سال ۱۹۷۳ و ۱۴۵ آزمودنی در سال ۱۹۸۱ جمع‌آوری شد. این داده‌ها را چگونه باید تحلیل کند؟

۱۳-۲۱ کوکران و اوربانزیک (۱۹۸۲) به اثر ارتفاع یک اتاق بر فضای مطلوب آزمودنی‌ها توجه کردند. آنها ۴۸ آزمودنی را هم در اتاق سقف بلند ۱۰ فوتی و هم سقف کوتاه ۷ فوتی آزمایش کردند. آزمودنی‌ها پشت به یک دیوار در حالی که یک غریبه نزدیک می‌شد ایستادند. به آنها گفته شد هرگاه از نزدیک شدن غریبه احساس ناراحتی کردند بگویند "ایست". آزمایشگران با این داده‌ها باید چه کار کنند؟ اگر نخواهند از آزمون پارامتریک استفاده کنند باید چه کار کنند؟ (توجه: فاصله برای سقف کوتاه بیشتر بود، که گویای آن است که فضای بین فردی تنها وابسته به فاصله افقی نیست.)

۱۴-۲۱ رابینسون، بارت و اسکین (۱۹۸۳) نمرات یک مقیاس مرکز مهار را برای ۲۰ پدر نوجوان ازدواج کرده و ۲۰ ناپدری نوجوان ازدواج نکرده مورد مقایسه قرار داد. اگر آنها تمایلی به برآورده شدن پیش‌فرض‌های پارامتریک نداشته باشند چگونه باید این داده‌ها را تحلیل کنند؟ (توجه: تفاوتی مشاهده نکردند.)

۱۵-۲۱ اسمیت و پلنت (۱۹۸۲) تفاوت‌های جنسی را در رضایت شغلی مطالعه کردند. آنها ابتدا استادان زن و مرد دانشگاه را بر اساس چهار متغیر شناخته شده در ارتباط با رضایت شغلی (مثل مدت خدمت،

رتبه، بالاترین درجه، و بخش) دو به دو همتاسازی کردند. سپس یک مقیاس مداد-کاغذی رضایت شغلی را برای تعیین نمره هر آزمودنی در پنج زمینه شغلی-کار، درآمد، ارتقاء، سرپرستی و همکاران-مورد استفاده قرار دادند. برای این داده‌ها با در نظر گرفتن هر زمینه رضایت بطور جداگانه، چه شیوه آماری مناسب است؟

۱۶-۲۱ هوسک و کوپر (۱۹۸۲) نقشی که قربانی بودن نه فقط ناظر بودن در شناسایی شاهد عینی دارد را بررسی کردند. در شرایط کنترلی یک همکار آزمایشگر با یک آزمودنی وارد اتاق آزمایش می‌شود، چند فرم را تکمیل و اتاق را ترک می‌کند. در شرایط دیگر همکار همان کار را می‌کند، اما موقع رفتن ماشین حساب آزمایشگر را می‌دزد. در شرایط سوم همکار ساعت آزمودنی را که از او خواسته بودند روی میز بگذارد می‌دزد. دو متغیر وابسته وجود داشت. اول اینکه آیا آزمودنی می‌توانست همکار را از بین شش عکس بدرستی تشخیص دهد، و دوم رتبه‌بندی ذهنی آزمودنی از همکارش در مقیاس ۹ درجه‌ای شناسایی بود. آزمایشگر بیشتر علاقه‌مند بود بداند آیا یک قربانی دزدی بودن به شناسایی بهتر و مطمئن‌تر منجر می‌شود یا فقط مشاهده یک دزد. (این داده‌ها را چگونه می‌توان تحلیل کرد؟ (توجه: دو همکار دزد فرقی نداشتند و رابطه‌ای بین صداقت و اعتماد نبود.)

۱۷-۲۱ باردکلی و کجونگجی (۱۹۸۲) ادراک مثلث‌های ذهنی را مورد آزمایش قرار دادند. وقتی به سه نقطه که تشکیل یک مثلث می‌دهند نگاه می‌کنید، یک شمای ذهنی از خطوط ارتباط دهنده آن نقاط برای تشکیل شمایی از مثلث وجود دارد. باردلی و کجونگجی از آزمودنی‌های خواستند که مثلث‌های ذهنی را در حالت ایستا و دو باره در حالت چرخان نگاه کنند. به آزمودنی آموختند که بگویند آیا شکل ذهنی وقتی ایستا هست مشخص‌تر است یا زمانی که چرخان است. از ۳۷ آزمودنی، ۳۵ نفر گفتند وقتی مثلث چرخان است شمای مشخص‌تری دارد. آنها چگونه می‌توانند بررسی کنند که آیا این تفاوت معنادار است- گرچه در اینجا واقعاً به آزمونی نیاز نیست؟

۱۸-۲۱ برون و همکاران (۱۹۸۲) به بررسی فراموشی ناشی از دارو به عنوان روشی برای مشخص نمودن فراموشی ناشی از علل عضوی پرداختند. آنها به آزمودنی‌ها فهرستی از لغات ارائه کردند و به آنها داروی لورازپام (که ایجاد فراموشی می‌کند) یا سالین تزریق کردند. بعد از ۱/۵ ساعت از آزمودنی‌ها خواستند کلماتی را که یاد گرفته بودند بازگو کنند. همچنین از همان آزمودنی‌ها خواستند تا فهرستی را بعد از تزریق دارو یاد بگیرند و بعد از ۱/۵ ساعت بازگو کنند. اگر لورازپام ذخیره مواد در حافظه را مختل کند، فقط بخاطر آوردن فهرست دوم بایستی مختل شود. اگر لورازپام در بازیابی مواد از حافظه تداخل ایجاد کند نه در بخاطر سپردن آنها، آنگاه هر دو فهرست باید مختل شوند. چه آزمون آماری برای تحلیل این داده‌ها مناسب است؟ این مورد دیگری است که در آن شما نمی‌دانید چگونه یک تحلیل را انجام دهید، اما باید بتوانید طرح را توضیح دهید. (توجه: بخاطر آوردن فهرست اول تحت تأثیر قرار نگرفت، اما بخاطر آوردن فهرستی که بعد از تزریق دارو خوانده شد ضعیف بود.)

۱۹-۲۱ هیگز و گویستا (۱۸۲) از هفت آزمودنی که عادت به خواب کمتر از ۶/۵ ساعت داشتند و نه آزمودنی که عادت به خواب بیشتر از ۸/۵ ساعت داشتند خواستند تا برای ۳۰ روز در فواصل ۲ ساعته مقیاس خواب استندفورد (SSS) را تکمیل کنند. این مقیاس میزانی از هشیاری است و بطور ساده نیازمند آن است تا آزمودنی سطح هشیاری خود را بین ۱ (خیلی هوشیار) تا ۷ (به سختی بیدار مانده) رتبه‌گذاری کند. در عمل طراحان داده‌ها را به هفت زمان مختلف روز تقسیم کرده‌اند اما برای مثال حاضر فرض کنید که متغیر وابسته نمره هر آزمودنی در دوره ۳۰ روزه است. تحلیل مناسب برای این داده‌ها کدام است؟

۲۰-۲۱ سمت راست صورت هر شخص بیش از سمت چپ او کل صورتش را نمایان می‌سازد. کندی، برد و کار (۱۹۸۲) از ۹۱ آزمودنی خواستند تا کل صورت پنج چهره مختلف را ببینند. یک هفته بعد آزمون یادآوری انجام شد، که آزمودنی‌ها با ۱۲ چهره روبرو شدند و از آنها خواسته شد کسی را که قبلاً دیده‌اند شناسایی کنند. در آزمون یادآوری آزمودنی‌ها به سه گروه با اندازه برابر تقسیم شدند. به یک گروه تصویر کامل چهره، به گروه دیگر تصویر نیمه راست و گروه سوم نیمه چپ نشان داده شد. متغیر وابسته تعداد خطاها بود. این پژوهشگران از برآورده شدن پیش‌فرض‌های آزمون پارامتریک رضایت نداشتند. چگونه آنها می‌توانند این سه گروه را مقایسه کنند و چگونه می‌توانند دو گروه نیم چهره را مقایسه کنند؟ (توجه: برخلاف پیش‌بینی، سمت چپ به تشخیص بهتر از سمت راست منجر شده بود).

پایان

پیوست الف:

مرور عملیات ریاضی

مطالب زیر به عنوان یادآوری عملیات ریاضی است که در دوره‌های قبلی آموخته‌اید اما شاید از آن به بعد به کار نبرده باشید. هر چند بعضی از آنها چندان پیش پا افتاده است که از تکرار آنها در اینجا تعجب می‌کنید، اما مردم چیزهای خیلی ساده را بیشتر فراموش می‌کنند.

یکی از چیزهایی که به نظر می‌رسد دانشجویان هرگز یاد نمی‌گیرند این است که فهمیدن بسیاری از این اصول برای شما ساده است. برای مثال اگر نمی‌توانید بخاطر بسپارید که آیا می‌توان

$$\frac{18.1}{28.6} + \frac{18.1}{32.7} \quad \text{را با} \quad \frac{18.1}{28.6 + 32.7} \quad \text{عوض کرد}$$

(این کار را نمی‌توان کرد اما این یکی از چیزهایی است که گاهی نمی‌توانیم در مغز خود بگنجانیم)، سعی کنید این کار را با اعداد خیلی ساده انجام دهید. بنابراین روشن است که

$$= \frac{2}{5} = \frac{2}{1+4}$$

مانند این نیست

$$5/2 = \frac{2}{4} + \frac{2}{1}$$

بیشتر وقت‌ها واریسی یک شیوه با اعداد ساده سریع‌تر از دنبال آن گشتن است.

- علائم استاندارد و اطلاعات اساسی

$X \approx Y$ یا $X \cong Y$ تقریباً مساوی با Y است

$X \neq Y$ مساوی با Y نیست

$X > Y$ بزرگتر از Y است

$X \geq Y$ بزرگتر یا مساوی Y است

$X < Y$	X کوچک‌تر از Y است
$X \leq Y$	X کوچک‌تر یا مساوی Y است
$X < Y < Z$	X کوچک‌تر از Y و Y کوچک‌تر از Z است (یعنی Y بین X و Z است)
$ X $	قدر مطلق X - نادیده گرفتن علامت X
$\frac{1}{X}$	عکس X
X^2	مجدور X ($X \times X$)
X^n	X به توان n (X, n بار در خودش ضرب شود)
$\sqrt{X} = X^{1/2}$	جذر X (ریشه X)

- جمع و تفریق

$۸ - ۱۲ = -۴$ برای تفریق یک عدد بزرگ‌تر از یک عدد کوچک‌تر، عدد کوچک‌تر را از بزرگ‌تر کم کرده و حاصل را منفی می‌کنیم.

ترتیب عملیات مهم نیست. $۴ = ۱۲ - ۸ = ۱۲ + (-۸)$

- ضرب و تقسیم

اگر قبل از یک پرانتز () عملیاتی نباشد، در این صورت عمل ضرب انجام می‌شود.

$۲(۳)(۶) = ۲ \times ۳ \times ۶$

اعداد به هر ترتیبی می‌توانند در هم ضرب شوند.

$۲ \times ۳ \times ۶ = ۳ \times ۶ \times ۲$

تقسیم می‌تواند به هر ترتیبی صورت بگیرد.

$\frac{۲ \times ۸}{۴} = \frac{۲}{۴} \times ۸ = ۲ \times \frac{۸}{۴} = \frac{۱۶}{۴} = ۴$

در یک رشته عملیات پیوسته، ابتدا ضرب و تقسیم صورت می‌گیرد و بعد حاصل را جمع یا تفریق می‌کنیم.

$۷ \times ۳ + ۶ = ۲۱ + ۶ = ۲۷$

نتیجه ضرب یا تقسیم اعداد با علامت یکسان، عددی با علامت مثبت خواهد بود.

$$۲ \times ۳ = ۶ ; (-۲)(-۳) = ۶$$

$$\frac{۶}{۳} = ۲ ; \frac{-۶}{-۳} = ۲$$

نتیجه ضرب یا تقسیم اعداد با علامت مخالف، عددی با علامت منفی خواهد بود.

$$(-۲)۳ = -۶ ; \frac{-۶}{۳} = -۲$$

با داشتن چند علامت مختلف دو به دو کار کنید تا به علامت درست برسید

$$(-۲)۳ = (-۶)(-۴) = (-۶)(۲۴) = -۱۴۴$$

- پرانتزها

هنگام ضرب کردن، یا عملیات داخل () را قبل از ضرب انجام دهید، یا هر عنصر را در () ضرب کنید و سپس جمع کنید.

$$+ 2(-6) + 2(3) = 14 - 12 + 6 = 8$$

$$2(7 - 6 + 3) = 2(4) = 8 \quad \text{یا} \quad 2(7)$$

برای وقتی که عبارت داخل پرانتز به توان می‌رسد، عملیات داخل () را انجام دهید، حاصل را به توان برسانید، و بعد عملیات‌های دیگر را انجام دهید.

$$2(7 - 6 + 3)^2 = 2(4)^2 = 2(16) = 32$$

- کسرها

بری تبدیل کسر به اعشار، صورت را بر مخرج تقسیم کنید.

$$\frac{1}{5} = 0.2$$

برای معکوس کردن یک کسر (مثلاً $\frac{3}{4}$) جای صورت و مخرج را عوض کنید.

$$\frac{4}{3}$$

برای ضرب یک کسر در یک عدد کامل، صورت کسر را در آن عدد ضرب کنید. $\frac{3 \times 6}{5} = \frac{18}{5} = 3\frac{3}{5}$

$$3 \times \frac{6}{5} =$$

برای ضرب چند کسر در یکدیگر، صورت‌ها را در هم و مخرج‌ها را در هم ضرب کنید.

$$\frac{3}{5} \times \frac{6}{7} \times \frac{1}{2} = \frac{3 \times 6 \times 1}{5 \times 7 \times 2} = \frac{18}{70} = 0.26$$

برای جمع کردن کسرهایی که مخرج آنها یکسان است، صورت‌ها را با هم جمع کرده و بر مخرج مشترک تقسیم کنید.

$$\frac{1}{3} + \frac{4}{3} = \frac{5}{3} = 1\frac{2}{3}$$

برای جمع کردن کسرهایی که مخرج آنها فرق دارد، صورت و مخرج را در عدد ثابتی ضرب کنید

تا مخرج‌های یکسان شوند، مخرج مشترک بگیرید)، سپس قاعده قبلی را دنبال کنید.

$$\frac{1}{6} + \frac{4}{3} = \frac{1}{6} + \frac{8}{6} = \frac{9}{6} = 1/5$$

$$\frac{8}{13} + \frac{12}{25} = \left(\frac{25}{25} + \frac{8}{13} \right) + \left(\frac{13}{13} + \frac{12}{25} \right) =$$

یک نمونه ماهرانه‌تر از قاعده قبلی

$$= \frac{200}{325} + \frac{156}{325} = \frac{356}{325} = 1/0.95$$

$$\frac{8}{1/3} = 8 \left(\frac{3}{1} \right) = 24$$

برای تقسیم بر یک کسر، آن را در عکس کسر ضرب کنید.

- عملیات جبری

اغلب عملیات‌های جبری به حرکت دادن چیزها از یک طرف معادله به طرف دیگر ساده می‌شوند. به لحاظ حسابی قاعده این است که هرکاری که در یک طرف معادله انجام دادید باید در طرف دیگر هم انجام دهید.

$$3 + X = 8 \quad \text{معادله زیر را برای } X \text{ حل کنید:}$$

ما X را در یک سمت و جواب را در سمت دیگر می‌خواهیم. تمام کاری که باید انجام دهیم کم کردن ۳

$$3 + X - 3 = 8 - 3 \quad X = 5 \quad \text{از هر دو طرف است تا}$$

$$X - 3 = 8 \quad \text{اگر معادله این گونه باشد}$$

$$X - 3 + 3 = 8 + 3 \quad X = 11 \quad \text{ما ۳ را به دو طرف معادله اضافه خواهیم کرد:}$$

برای معادلاتی که شامل ضرب یا تقسیم هستند از اصول مشابهی پیروی می‌کنیم. $2X = 21$

$$\frac{2X}{2} = \frac{21}{2} \quad X = 10/5$$

هر دو طرف را بر ۲ تقسیم می‌کنیم

$$\frac{X}{7} = 13 \quad \frac{7X}{7} = 7(13) \quad X = 91$$

و

خوب است به چیزها به شیوه متفاوت اما کاملاً یکسان فکر کنیم. وقتی می‌خواهید چیزی که اضافه (یا کم) شده را از یک سوی معادله بردارید، آن را به سمت دیگر برده و علامت آن را عکس کنید.

$$X - 7 = 19 \quad \text{یا} \quad 3 + X = 12$$

$$X = 19 + 7 \quad \text{یا} \quad X = 12 - 3$$

وقتی چیزی را که می‌خواهید بردارید در صورت کسر باشد، آن را به طرف دیگر معادله ببرید و در

$$7/6X = 12 \quad X = \frac{12}{7.6}$$

مخرج قرار دهید.

وقتی چیزی را که می‌خواهید بردارید در مخرج کسر است، آن را به طرف دیگر ببرید و در صورت ضرب

$$X = 14/6(8/9) \quad \frac{X}{8.9} = 14/6$$

کنید.

توجه داشته باشید که در مورد عبارات پیچیده‌تر باید همه چیزها را در سمت دیگر معادله ضرب (یا بر

$$7/6X = 12 + 8 \quad X = \frac{12+8}{7.6} \quad \text{آن تقسیم) کنید. بنابراین}$$

برای معادلات پیچیده، در هر بار فقط یک مرحله را انجام دهید.

$$7/6(X + 8) = \frac{14}{7} - 5$$

ابتدا $7/6$ را بر می‌داریم:

$$X + 8 = \frac{14/7 - 5}{7.6}$$

بعد ۸ را بر می‌داریم

$$X = \frac{14/7 - 5}{7.6} - 8$$

حالا کسر را ساده می‌کنیم:

$$X = \frac{2-5}{7.6} - 8 = \frac{-3}{7.6} - 8 = -0.395 - 8 = -8.395$$

پیوست پ:

نمادها و نشانه‌ها

- نمادهای حروف یونانی

=====

α	سطح معناداری - احتمال خطای نوع I (آلفا)
β	احتمال خطای نوع II (بتا)
γ	اندازه اثر (گاما)
δ	پارامتر غیر مرکزی (دلتا)
η^2	مجذور اتا
μ	میانگین جمعیت (میو)
μ_x	میانگین توزیع نمونه‌گیری میانگین‌ها
σ	انحراف معیار جمعیت
σ^2	واریانس جمعیت
Σ	نماد حاصل جمع (سیگما)
$\emptyset$	ضریب فی
x^2	مجذور خی (مجذور کای)
x_F^2	مجذور خی فریدمن
ω^2	مجذور اُمگا

- نمادهای حروف انگلیسی =====

a	مقدار ثابت؛ تعداد سطوح متغیر A در تحلیل واریانس
b	شیب، که ضریب رگرسیون هم نامیده می شود
CI	فاصله اطمینان
COV_{XY}	کوواریانس X و Y
df	درجات آزادی
E	فراوانی مورد انتظار، مقدار مورد انتظار
F	آماره F
G	جمع کل - جمع تمام نمرات
GM	میانگین اصلی
H	آماره کروسکل-والیس
H_0, H_1	فرض صفر، فرض مخالف
MS	میانگین مجذورات
MS_{error}	میانگین مجذورات خطا
n, n_1, N_i	تعداد موارد یک نمونه
$N(0, 1)$	بخوانید "توزیع نرمال با $\mu = 0$ ، $\sigma^2 = 1$ "
O	فراوانی مشاهده شده
P	نماد عمومی احتمال
r, r_{XY}	ضریب همبستگی پیرسون
Γ_{pb}	ضریب همبستگی نقطه‌ای
R	ضریب همبستگی چند گانه
RLL	حد پایین واقعی
S^2, S_X^2	واریانس نمونه
S_p^2	واریانس انباشته
S_X, S	انحراف معیار نمونه
S_D	انحراف معیار تفاوت نمرات
$S_{\bar{D}}$	انحراف معیار میانگین تفاوت نمرات
$S_{\bar{X}}, S_{\bar{X}_1 - \bar{X}_2}$	خطای معیار میانگین تفاوت میانگین‌ها؛ خطای معیار میانگین
$S_{Y-\bar{Y}}$	خطای معیار برآورد
SP_{XY}	مجموع مضروب‌ات؛ مجموع حاصل ضرب‌های X و Y

SS_A	مجموع مجذورات متغیر A
SS_{AB}	مجموع مجذورات اثر تعاملی
SS_{error}	مجموع مجذورات خطا
SS_Y	مجموع مجذورات متغیر Y
$SS_{\hat{Y}}$	مجموع مجذورات مقدار پیش‌بینی شده Y
$SS_{Y-\hat{Y}}$	خطای مجموع مجذورات SS_{error}
t	آماره t
t. 05	مقدار بحرانی t
T	آماره نشانه جفت‌های همبسته ویل کاکسون
T_j	مقدار کل برای گروه j
T_{Ai}	مقدار کل برای سطح iام متغیر A
W_s, W'_s	آماره من - ویتنی
$\bar{X}, \bar{X}_i$	میانگین نمونه
$\bar{X}_h$	میانگین هارمونیک
$\hat{Y}, \hat{Y}_i$	مقدار پیش‌بینی شده Y
z	انحراف نرمال (نمره استاندارد شده نیز خوانده می‌شود)

پیوست ج:

جدول ها



TABLE 1 Upper Percentage Points of the χ^2 Distribution. (Source: The entries in this table were computed by the author.)

df	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	0.00	0.00	0.00	0.00	0.02	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.82	9.35	11.35	12.84
4	0.21	0.30	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.54	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.66	23.59
10	2.15	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.72	26.75
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.21	28.30
13	3.56	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.06	23.69	26.12	29.14	31.31
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.15
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.83	28.41	31.41	34.17	37.56	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.93	36.78	40.29	42.80
23	9.26	10.19	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.88	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.37	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.32	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.80	12.88	14.57	16.15	18.11	21.75	26.34	31.53	36.74	40.11	43.20	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.67	22.14	24.42	26.51	29.06	33.67	39.34	45.61	51.80	55.75	59.34	63.71	66.80
50	27.96	29.68	32.35	34.76	37.69	42.95	49.34	56.33	63.16	67.50	71.42	76.17	79.52
60	35.50	37.46	40.47	43.19	46.46	52.30	59.34	66.98	74.39	79.08	83.30	88.40	91.98
70	43.25	45.42	48.75	51.74	55.33	61.70	69.34	77.57	85.52	90.53	95.03	100.44	104.24
80	51.14	53.52	57.15	60.39	64.28	71.15	79.34	88.13	96.57	101.88	106.63	112.34	116.35
90	59.17	61.74	65.64	69.13	73.29	80.63	89.33	98.65	107.56	113.14	118.14	124.13	128.32
100	67.30	70.05	74.22	77.93	82.36	90.14	99.33	109.14	118.49	124.34	129.56	135.82	140.19

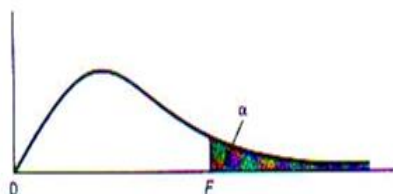


TABLE 2 Critical Values of the F Distribution Alpha = .05. (Source: The entries in this table were computed by the author.)

		Degrees of Freedom for Numerator															
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
Degrees of Freedom for Denominator	1	161.4	199.5	215.8	224.8	230.0	233.8	236.5	238.6	240.1	242.1	245.2	248.4	248.9	250.5	250.8	252.6
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43	19.44	19.46	19.47	19.48	19.48
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.63	8.62	8.59	8.58
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.77	5.75	5.72	5.70
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.52	4.50	4.46	4.44
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.83	3.81	3.77	3.75
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.40	3.38	3.34	3.32
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.11	3.08	3.04	3.02
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.89	2.86	2.83	2.80
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.73	2.70	2.66	2.64
	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72	2.65	2.60	2.57	2.53	2.51
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62	2.54	2.50	2.47	2.43	2.40
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53	2.46	2.41	2.38	2.34	2.31
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46	2.39	2.34	2.31	2.27	2.24
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.28	2.25	2.20	2.18
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35	2.28	2.23	2.19	2.15	2.12
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31	2.23	2.18	2.15	2.10	2.08
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27	2.19	2.14	2.11	2.06	2.04
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23	2.16	2.11	2.07	2.03	2.00
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.07	2.04	1.99	1.97
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.15	2.07	2.02	1.98	1.94	1.91
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.11	2.03	1.97	1.94	1.89	1.86
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.07	1.99	1.94	1.90	1.85	1.82
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.04	1.96	1.91	1.87	1.82	1.79
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.88	1.84	1.79	1.76
	40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.78	1.74	1.69	1.66
	50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.87	1.78	1.73	1.69	1.63	1.60
	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.69	1.65	1.59	1.56
	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.75	1.66	1.60	1.55	1.50	1.46
	200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.72	1.62	1.56	1.52	1.46	1.41
	500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.69	1.59	1.53	1.48	1.42	1.38
	1000	3.85	3.01	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.68	1.58	1.52	1.47	1.41	1.36

TABLE 3 Critical Values of the F Distribution Alpha = .01. (Source: The entries in this table were computed by the author.)

		Degrees of Freedom for Numerator															
		1	2	3	4	5	6	7	8	9	10	15	20	25	30	40	50
Degrees of Freedom for Denominator	1	4048	4993	5377	5577	5668	5924	5992	6096	6132	6168	6079	6168	6214	6355	6168	6213
	2	98.50	99.01	99.15	99.23	99.30	99.33	99.35	99.39	99.40	99.43	99.38	99.48	99.43	99.37	99.44	99.59
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	26.87	26.69	26.58	26.51	26.41	26.36
	4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.20	14.02	13.91	13.84	13.75	13.69
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.72	9.55	9.45	9.38	9.29	9.24
	6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.30	7.23	7.14	7.09
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	6.06	5.99	5.91	5.86
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.52	5.36	5.26	5.20	5.12	5.07
	9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.71	4.65	4.57	4.52
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.31	4.25	4.17	4.12
	11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.25	4.10	4.01	3.94	3.86	3.81
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.01	3.86	3.76	3.70	3.62	3.57
	13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.82	3.66	3.57	3.51	3.43	3.38
	14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.66	3.51	3.41	3.35	3.27	3.22
	15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.28	3.21	3.13	3.08
	16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.41	3.26	3.16	3.10	3.02	2.97
	17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.31	3.16	3.07	3.00	2.92	2.87
	18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.23	3.08	2.98	2.92	2.84	2.78
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.15	3.00	2.91	2.84	2.76	2.71	
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.84	2.78	2.69	2.64	
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.98	2.83	2.73	2.67	2.58	2.53	
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.89	2.74	2.64	2.58	2.49	2.44	
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.81	2.66	2.57	2.50	2.42	2.36	
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.75	2.60	2.51	2.44	2.35	2.30	
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.55	2.45	2.39	2.30	2.25	
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.37	2.27	2.20	2.11	2.06	
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.42	2.27	2.17	2.10	2.01	1.95	
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.35	2.20	2.10	2.03	1.94	1.88	
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.19	2.03	1.93	1.86	1.76	1.70	
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.13	1.97	1.87	1.79	1.69	1.63	
500	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.07	1.92	1.81	1.74	1.63	1.57	
1000	6.67	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.06	1.90	1.79	1.72	1.61	1.54	

TABLE 4
Power as a Function of δ and
Significance Level (α). (Source:
 The entries in this table were
 computed by the author.)

δ	Alpha for Two-Tailed Test			
	.10	.05	.02	.01
1.00	0.26	0.17	0.09	0.06
1.10	0.29	0.20	0.11	0.07
1.20	0.33	0.22	0.13	0.08
1.30	0.37	0.26	0.15	0.10
1.40	0.40	0.29	0.18	0.12
1.50	0.44	0.32	0.20	0.14
1.60	0.48	0.36	0.23	0.17
1.70	0.52	0.40	0.27	0.19
1.80	0.56	0.44	0.30	0.22
1.90	0.60	0.48	0.34	0.25
2.00	0.64	0.52	0.37	0.28
2.10	0.68	0.56	0.41	0.32
2.20	0.71	0.60	0.45	0.35
2.30	0.74	0.63	0.49	0.39
2.40	0.78	0.67	0.53	0.43
2.50	0.80	0.71	0.57	0.47
2.60	0.83	0.74	0.61	0.51
2.70	0.85	0.77	0.65	0.55
2.80	0.88	0.80	0.68	0.59
2.90	0.90	0.83	0.72	0.63
3.00	0.91	0.85	0.75	0.66
3.10	0.93	0.87	0.78	0.70
3.20	0.94	0.89	0.81	0.73
3.30	0.95	0.91	0.84	0.77
3.40	0.96	0.93	0.86	0.80
3.50	0.97	0.94	0.88	0.82
3.60	0.98	0.95	0.90	0.85
3.70	0.98	0.96	0.92	0.87
3.80	0.98	0.97	0.93	0.89
3.90	0.99	0.97	0.94	0.91
4.00	0.99	0.98	0.95	0.92
4.10	0.99	0.98	0.96	0.94
4.20	...	0.99	0.97	0.95
4.30	...	0.99	0.98	0.96
4.40	...	0.99	0.98	0.97
4.50	...	0.99	0.99	0.97
4.60	...	...	0.99	0.98
4.70	...	...	0.99	0.98
4.80	...	...	0.99	0.99
4.90	...	...	...	0.99
5.00	...	...	...	0.99

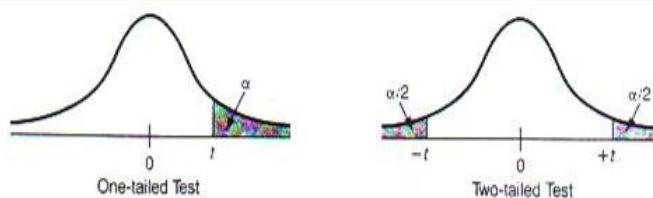


TABLE 5
Percentage Points of the *t* Distribution. (Source: The entries in this table were computed by the author.)

df	Level of Significance for One-Tailed Test							
	.25	.20	.15	.10	.05	.025	.01	.005
	Level of Significance for Two-Tailed Test							
	.50	.40	.30	.20	.10	.05	.02	.01
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

TABLE 6 Critical Lower-Tail Values of T (and Their Associated Probabilities) for Wilcoxon's Matched-Pairs Signed-Ranks Test. (Source: The entries in this table were computed by the author.)

Nominal Alpha (One-Tailed)					Nominal Alpha (One-Tailed)				
N	.05	.025	.01	.005	N	.05	.025	.01	.005
	T	T	T	T		T	T	T	T
5	0 .0313				28	130 .0496	116 .0239	101 .0096	91 .0048
	1 .0625					131 .0521	117 .0252	102 .0102	92 .0051
6	2 .0469	0 .0156			29	140 .0482	126 .0240	110 .0095	100 .0049
	3 .0781	1 .0313				141 .0504	127 .0253	111 .0101	101 .0053
7	3 .0391	2 .0234	0 .0078		30	151 .0481	137 .0249	120 .0098	109 .0050
	4 .0547	3 .0391	1 .0156			152 .0502	138 .0261	121 .0104	110 .0053
8	5 .0391	3 .0195	1 .0078	0 .0039	31	163 .0491	147 .0239	130 .0099	118 .0049
	6 .0547	4 .0273	2 .0117	1 .0078		164 .0512	148 .0251	131 .0105	119 .0052
9	8 .0488	5 .0195	3 .0098	1 .0039	32	175 .0492	159 .0249	140 .0097	128 .0050
	9 .0645	6 .0273	4 .0137	2 .0059		176 .0512	160 .0260	141 .0103	129 .0053
10	10 .0420	8 .0244	5 .0098	3 .0049	33	187 .0485	170 .0242	151 .0099	138 .0049
	11 .0527	9 .0322	6 .0137	4 .0068		188 .0503	171 .0253	152 .0104	139 .0052
11	13 .0415	10 .0210	7 .0093	5 .0049	34	200 .0488	182 .0242	162 .0098	148 .0048
	14 .0508	11 .0269	8 .0122	6 .0068		201 .0506	183 .0252	163 .0103	149 .0051
12	17 .0461	13 .0212	9 .0081	7 .0046	35	213 .0484	195 .0247	173 .0096	159 .0048
	18 .0549	14 .0261	10 .0105	8 .0061		214 .0501	196 .0257	174 .0100	160 .0051
13	21 .0471	17 .0239	12 .0085	9 .0040	36	227 .0489	208 .0248	185 .0096	171 .0050
	22 .0549	18 .0287	13 .0107	10 .0052		228 .0505	209 .0258	186 .0100	172 .0052
14	25 .0453	21 .0247	15 .0083	12 .0043	37	241 .0487	221 .0245	198 .0099	182 .0048
	26 .0520	22 .0290	16 .0101	13 .0054		242 .0503	222 .0254	199 .0103	183 .0050
15	30 .0473	25 .0240	19 .0090	15 .0042	38	256 .0493	235 .0247	211 .0099	194 .0048
	31 .0535	26 .0277	20 .0108	16 .0051		257 .0509	236 .0256	212 .0104	195 .0050
16	35 .0467	29 .0222	23 .0091	19 .0046	39	271 .0492	249 .0246	224 .0099	207 .0049
	36 .0523	30 .0253	24 .0107	20 .0055		272 .0507	250 .0254	225 .0103	208 .0051
17	41 .0492	34 .0224	27 .0087	23 .0047	40	286 .0486	264 .0249	238 .0100	220 .0049
	42 .0544	35 .0253	28 .0101	24 .0055		287 .0500	265 .0257	239 .0104	221 .0051
18	47 .0494	40 .0241	32 .0091	27 .0045	41	302 .0488	279 .0248	252 .0100	233 .0048
	48 .0542	41 .0269	33 .0104	28 .0052		303 .0501	280 .0256	253 .0103	234 .0050
19	53 .0478	46 .0247	37 .0090	32 .0047	42	319 .0496	294 .0245	266 .0098	247 .0049
	54 .0521	47 .0273	38 .0102	33 .0054		320 .0509	295 .0252	267 .0102	248 .0051
20	60 .0487	52 .0242	43 .0096	37 .0047	43	336 .0498	310 .0245	281 .0098	261 .0048
	61 .0527	53 .0266	44 .0107	38 .0053		337 .0511	311 .0252	282 .0102	262 .0050
21	67 .0479	58 .0230	49 .0097	42 .0045	44	353 .0495	327 .0250	296 .0097	276 .0049
	68 .0516	59 .0251	50 .0108	43 .0051		354 .0507	328 .0257	297 .0101	277 .0051
22	75 .0492	65 .0231	55 .0095	48 .0046	45	371 .0498	343 .0244	312 .0098	291 .0049
	76 .0527	66 .0250	56 .0104	49 .0052		372 .0510	344 .0251	313 .0101	292 .0051
23	83 .0490	73 .0242	62 .0098	54 .0046	46	389 .0497	361 .0249	328 .0098	307 .0050
	84 .0523	74 .0261	63 .0107	55 .0051		390 .0508	362 .0256	329 .0101	308 .0052
24	91 .0475	81 .0245	69 .0097	61 .0048	47	407 .0490	378 .0245	345 .0099	322 .0048
	92 .0505	82 .0263	70 .0106	62 .0053		408 .0501	379 .0251	346 .0102	323 .0050
25	100 .0479	89 .0241	76 .0094	68 .0048	48	426 .0490	396 .0244	362 .0099	339 .0050
	101 .0507	90 .0258	77 .0101	69 .0053		427 .0500	397 .0251	363 .0102	340 .0051
26	110 .0497	98 .0247	84 .0095	75 .0047	49	446 .0495	415 .0247	379 .0098	355 .0049
	111 .0524	99 .0263	85 .0102	76 .0051		447 .0505	416 .0253	380 .0100	356 .0050
27	119 .0477	107 .0246	92 .0093	83 .0048	50	466 .0495	434 .0247	397 .0098	373 .0050
	120 .0502	108 .0260	93 .0100	84 .0052		467 .0506	435 .0253	398 .0101	374 .0051

TABLE 7 Critical Lower-Tail Values of W_5 for the Mann-Whitney Test for Two Independent Samples ($N_1 \leq N_2$). (Source: Table 1 in L. R. Verdooren. Extended tables of critical values for Wilcoxon's test statistic, *Biometrika*, 1963, 50, 177-186, with permission of the author and editor.)

$N_1 = 1$								$N_1 = 2$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
2							4							10	2
3							5							12	3
4							6							14	4
5							7						3	16	5
6							8					3	4	18	6
7							9					3	4	20	7
8							10				3	4	5	22	8
9						1	11				3	4	5	24	9
10						1	12				3	4	6	26	10
11						1	13				3	4	6	28	11
12						1	14				4	5	7	30	12
13						1	15			3	4	5	7	32	13
14						1	16			3	4	6	8	34	14
15						1	17			3	4	6	8	36	15
16						1	18			3	4	6	8	38	16
17						1	19			3	5	6	9	40	17
18						1	20			3	5	7	9	42	18
19					1	2	21		3	4	5	7	10	44	19
20					1	2	22		3	4	5	7	10	46	20
21					1	2	23		3	4	6	8	11	48	21
22					1	2	24		3	4	6	8	11	50	22
23					1	2	25		3	4	6	8	12	52	23
24					1	2	26		3	4	6	9	12	54	24
25					1	2	27		3	4	6	9	12	56	25

$N_1 = 3$								$N_1 = 4$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
3						6	7	21							
4						6	7	24							
5					6	7	8	27			10	11	13	36	4
6					6	7	8	30		10	11	12	14	40	5
7					6	7	8	33		10	11	13	14	44	6
8					6	8	9	36		10	11	13	14	48	7
9					6	8	9	39		11	12	14	15	52	8
10					6	7	9	42		11	13	14	16	56	9
11					6	7	9	45		11	13	15	17	60	10
12					6	7	9	48		11	13	15	17	64	11
13					6	7	9	51		11	13	15	18	68	12
14					6	7	9	54		11	13	15	18	72	13
15					6	7	9	57		11	13	15	19	76	14
16					6	7	9	60		11	13	15	19	80	15
17					6	7	9	63		11	13	15	20	84	16
18					6	7	9	66		11	13	15	20	88	17
19					6	7	9	69		11	13	15	21	92	18
20					6	7	9	72		11	13	15	21	96	19
21					6	7	9	75		11	13	15	22	100	20
22					6	7	9	78		11	13	15	22	104	21

$N_1 = 3$								$N_1 = 4$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
23	7	10	12	15	19	23	81	14	19	22	27	31	36	112	23
24	7	10	12	16	19	24	84	15	20	23	27	32	38	116	24
25	7	11	13	16	20	25	87	15	20	23	28	33	38	120	25

$N_1 = 5$								$N_1 = 6$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
5		15	16	17	19	20	55								
6		16	17	18	20	22	60	...	23	24	26	28	30	78	6
7	...	16	18	20	21	23	65	21	24	25	27	29	32	84	7
8	15	17	19	21	23	25	70	22	25	27	29	31	34	90	8
9	16	18	20	22	24	27	75	23	26	28	31	33	36	96	9
10	16	19	21	23	26	28	80	24	27	29	32	35	38	102	10
11	17	20	22	24	27	30	85	25	28	30	34	37	40	108	11
12	17	21	23	26	28	32	90	25	30	32	35	38	42	114	12
13	18	22	24	27	30	33	95	26	31	33	37	40	44	120	13
14	18	22	25	28	31	35	100	27	32	34	38	42	46	126	14
15	19	23	26	29	33	37	105	28	33	36	40	44	48	132	15
16	20	24	27	30	34	38	110	29	34	37	42	46	50	138	16
17	20	25	28	32	35	40	115	30	36	39	43	47	52	144	17
18	21	26	29	33	37	42	120	31	37	40	45	49	55	150	18
19	22	27	30	34	38	43	125	32	38	41	46	51	57	156	19
20	22	28	31	35	40	45	130	33	39	43	48	53	59	162	20
21	23	29	32	37	41	47	135	33	40	44	50	55	61	168	21
22	23	29	33	38	43	48	140	34	42	45	51	57	63	174	22
23	24	30	34	39	44	50	145	35	43	47	53	58	65	180	23
24	25	31	35	40	45	51	150	36	44	48	54	60	67	186	24
25	25	32	36	42	47	53	155	37	45	50	56	62	69	192	25

$N_1 = 7$								$N_1 = 8$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
7	29	32	34	36	39	41	105								
8	30	34	35	38	41	44	112	40	43	45	49	51	55	136	8
9	31	35	37	40	43	46	119	41	45	47	51	54	58	144	9
10	33	37	39	42	45	49	126	42	47	49	53	56	60	152	10
11	34	38	40	44	47	51	133	44	49	51	55	59	63	160	11
12	35	40	42	46	49	54	140	45	51	53	58	62	66	168	12
13	36	41	44	48	52	56	147	47	53	56	60	64	69	176	13
14	37	43	45	50	54	59	154	48	54	58	62	67	72	184	14
15	38	44	47	52	56	61	161	50	56	60	65	69	75	192	15
16	39	46	49	54	58	64	168	51	58	62	67	72	78	200	16
17	41	47	51	56	61	66	175	53	60	64	70	75	81	208	17
18	42	49	52	58	63	69	182	54	62	66	72	77	84	216	18
19	43	50	54	60	65	71	189	56	64	68	74	80	87	224	19
20	44	52	56	62	67	74	196	57	66	70	77	83	90	232	20
21	46	53	58	64	69	76	203	59	68	72	79	85	92	240	21
22	47	55	59	66	72	79	210	60	70	74	81	88	95	248	22
23	48	57	61	68	74	81	217	62	71	76	84	90	98	256	23
24	49	58	63	70	76	84	224	64	73	78	86	93	101	264	24
25	50	60	64	72	78	86	231	65	75	81	89	96	104	272	25

$N_1 = 9$								$N_1 = 10$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
9	52	56	59	62	66	70	171	65	71	74	78	82	87	210	10
10	53	58	61	65	69	73	180	67	73	77	81	86	91	220	11
11	55	61	63	68	72	76	189	69	76	79	84	89	94	230	12
12	57	63	66	71	75	80	198	72	79	82	88	92	98	240	13
13	59	65	68	73	78	83	207	74	81	85	91	96	102	250	14
14	60	67	71	76	81	86	216	76	84	88	94	99	106	260	15
15	62	69	73	79	84	90	225	78	86	91	97	103	109	270	16
16	64	72	76	82	87	93	234	80	89	93	100	106	113	280	17
17	66	74	78	84	90	97	243	82	92	96	103	110	117	290	18
18	68	76	81	87	93	100	252	84	94	99	107	113	121	300	19
19	70	78	83	90	96	103	261	87	97	102	110	117	125	310	20
20	71	81	85	93	99	107	270	89	99	105	113	120	128	320	21
21	73	83	88	95	102	110	279	91	102	108	116	123	132	330	22
22	75	85	90	98	105	113	288	93	105	110	119	127	136	340	23
23	77	88	93	101	108	117	297	95	107	113	122	130	140	350	24
24	79	90	95	104	111	120	306	98	110	116	126	134	144	360	25
25	81	92	98	107	114	123	315								

$N_1 = 11$								$N_1 = 12$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
11	81	87	91	96	100	106	253	98	105	109	115	120	127	300	12
12	83	90	94	99	104	110	264	101	109	113	119	125	131	312	13
13	86	93	97	103	108	114	275	103	112	116	123	129	136	324	14
14	88	96	100	106	112	118	286	106	115	120	127	133	141	336	15
15	90	99	103	110	116	123	297	109	119	124	131	138	145	348	16
16	93	102	107	113	120	127	308	112	122	127	135	142	150	360	17
17	95	105	110	117	123	131	319	115	125	131	139	146	155	372	18
18	98	108	113	121	127	135	330	118	129	134	143	150	159	384	19
19	100	111	116	124	131	139	341	120	132	138	147	155	164	396	20
20	103	114	119	128	135	144	352	123	136	142	151	159	169	408	21
21	106	117	123	131	139	148	363	126	139	145	155	163	173	420	22
22	108	120	126	135	143	152	374	129	142	149	159	168	178	432	23
23	111	123	129	139	147	156	385	132	146	153	163	172	183	444	24
24	113	126	132	142	151	161	396	135	149	156	167	176	187	456	25
25	116	129	136	146	155	165	407								

$N_1 = 13$								$N_1 = 14$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
13	117	125	130	136	142	149	351	137	147	152	160	166	174	406	14
14	120	129	134	141	147	154	364	141	151	156	164	171	179	420	15
15	123	133	138	145	152	159	377	144	155	161	169	176	185	434	16
16	126	136	142	150	156	165	390	148	159	165	174	182	190	448	17
17	129	140	146	154	161	170	403	151	163	170	179	187	196	462	18
18	133	144	150	158	166	175	416	155	168	174	183	192	202	476	19
19	136	148	154	163	171	180	429	159	172	178	188	197	207	490	20
20	139	151	158	167	175	185	442	162	176	183	193	202	213	504	21
21	142	155	162	171	180	190	455	166	180	187	198	207	218	518	22
22	145	159	166	176	185	195	468								

$N_1 = 13$								$N_1 = 14$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
23	149	163	170	180	189	200	481	169	184	192	203	212	224	532	23
24	152	166	174	185	194	205	494	173	188	196	207	218	229	546	24
25	155	170	178	189	199	211	507	177	192	200	212	223	235	560	25
$N_1 = 15$								$N_1 = 16$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
15	160	171	176	184	192	200	465								
16	163	175	181	190	197	206	480	184	196	202	211	219	229	528	16
17	167	180	186	195	203	212	495	188	201	207	217	225	235	544	17
18	171	184	190	200	208	218	510	192	206	212	222	231	242	560	18
19	175	189	195	205	214	224	525	196	210	218	228	237	248	576	19
20	179	193	200	210	220	230	540	201	215	223	234	243	255	592	20
21	183	198	205	216	225	236	555	205	220	228	239	249	261	608	21
22	187	202	210	221	231	242	570	209	225	233	245	255	267	624	22
23	191	207	214	226	236	248	585	214	230	238	251	261	274	640	23
24	195	211	219	231	242	254	600	218	235	244	256	267	280	656	24
25	199	216	224	237	248	260	615	222	240	249	262	273	287	672	25
$N_1 = 17$								$N_1 = 18$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
17	210	223	230	240	249	259	595								
18	214	228	235	246	255	266	612	237	252	259	270	280	291	666	18
19	219	234	241	252	262	273	629	242	258	265	277	287	299	684	19
20	223	239	246	258	268	280	646	247	263	271	283	294	306	702	20
21	228	244	252	264	274	287	663	252	269	277	290	301	313	720	21
22	233	249	258	270	281	294	680	257	275	283	296	307	321	738	22
23	238	255	263	276	287	300	697	262	280	289	303	314	328	756	23
24	242	260	269	282	294	307	714	267	286	295	309	321	335	774	24
25	247	265	275	288	300	314	731	273	292	301	316	328	343	792	25
$N_1 = 19$								$N_1 = 20$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
19	267	283	291	303	313	325	741								
20	272	289	297	309	320	333	760	298	315	324	337	348	361	820	20
21	277	295	303	316	328	341	779	304	322	331	344	356	370	840	21
22	283	301	310	323	335	349	798	309	328	337	351	364	378	860	22
23	288	307	316	330	342	357	817	315	335	344	359	371	386	880	23
24	294	313	323	337	350	364	836	321	341	351	366	379	394	900	24
25	299	319	329	344	357	372	855	327	348	358	373	387	403	920	25
$N_1 = 21$								$N_1 = 22$							
N_2	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	N_2
21	331	349	359	373	385	399	903								
22	337	356	366	381	393	408	924	365	386	396	411	424	439	990	22
23	343	363	373	388	401	417	945	372	393	403	419	432	448	1012	23
24	349	370	381	396	410	425	966	379	400	411	427	441	457	1034	24
25	356	377	388	404	418	434	987	385	408	419	435	450	467	1056	25

TABLE 7 (continued)

N_2	$N_1 = 23$							$N_1 = 24$							N_2
	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	
23	402	424	434	451	465	481	1081								
24	409	431	443	459	474	491	1104	440	464	475	492	507	525	1176	24
25	416	439	451	468	483	500	1127	448	472	484	501	517	535	1200	25

N_2	$N_1 = 25$							$N_1 = 26$							N_2
	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	0.001	0.005	0.010	0.025	0.05	0.10	$2W$	
25	480	505	517	536	552	570	1275								

TABLE 8
Table of Uniform Random
Numbers. (Source: The entries
in this table were computed by
the author.)

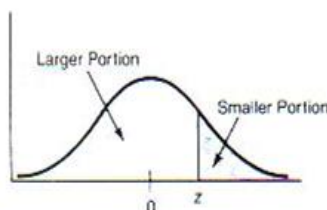
68204	38787	73304	44886	92836	43877	61049	49249	66105
61010	78345	75444	91680	33003	24128	97817	77562	62045
04604	93468	78459	27541	19672	14220	25102	42021	19252
36021	25507	64060	72923	58848	10374	63102	41534	92884
28129	43470	94097	16753	56425	75299	93688	75569	52067
09406	06584	46324	13981	06449	42604	13372	69040	95955
86423	81835	64226	20398	65772	91052	73496	14451	95967
13249	58525	81893	32894	68627	75644	45848	61511	90232
75454	17352	56548	39618	86705	50783	48388	82047	14660
06260	46176	99237	69874	84180	32005	66130	18055	99748
38507	92795	80672	00102	22980	69115	95653	05231	94996
03917	26795	59832	19014	96206	45413	76624	71219	65855
17927	32368	08177	31236	45401	26731	92256	99530	43998
26811	88937	37187	39762	29942	40091	65731	95955	23368
18480	28160	81908	30456	22462	15677	55642	67383	86884
37589	91842	76351	90585	45588	42858	37806	67969	50621
79903	34187	26952	75820	96335	90281	04269	85202	94965
46155	30200	75000	28570	47516	06744	72193	01258	85047
60916	73212	15853	28398	04721	69363	47071	65568	88519
34419	82840	88235	61966	86517	23966	45764	42177	17269
08692	26667	12941	14813	30815	26633	68184	80721	80505
92851	44185	90848	18341	77915	00177	64014	35490	02937
97909	07280	72167	10002	27374	92880	60055	94168	30742
28437	22027	07739	30905	33151	73567	82960	50104	67005
48165	28174	17909	11230	00929	54604	32435	54120	85199
99891	30913	06315	30201	72073	39589	62868	66339	15850
98022	13010	67970	99203	12536	88149	44387	20250	50798
91292	54688	47029	38970	77880	77295	11887	17628	93802
89081	34643	12988	12971	87742	57720	24438	64088	49496
32527	74239	20056	46668	94561	70111	92537	83562	11306
01870	21584	48574	09871	74453	24812	45770	95667	52377
84011	87542	96564	64256	64653	90025	61613	94168	83254
01568	29682	67489	62984	51901	30716	24513	46678	67991
40360	19206	40321	16004	64481	16130	03904	15811	19369
09392	39926	79590	23991	82492	13032	67337	54322	06058
77323	20500	52466	33008	84211	26357	79006	41178	35169

TABLE 8 (continued)

47590	01007	65376	18189	84040	39476	25383	45398	64917
29321	65783	71403	32894	32627	39067	47985	51485	27415
09530	05358	58722	31912	73356	65884	12883	36242	29646
65612	06843	72233	73352	66600	23237	71759	76881	19652
40355	85067	40788	40148	46099	48056	27858	58365	30202
24963	49571	82377	08687	73448	95484	15155	41780	71951
87273	44050	71961	48464	84084	65225	62846	11634	04853
31643	44756	12493	09024	74204	69949	67842	36141	08477
58326	55342	31419	80776	64028	59957	52969	71997	71477
02327	00460	39178	09511	92688	88585	99257	98752	39623
19377	49122	60591	79773	66289	89650	49298	13499	53623
95046	30203	47493	74395	45213	66739	45097	91670	62152
65013	71958	48360	70885	60313	44241	18740	05705	07488
86032	89018	97117	35656	20401	86438	87250	04717	67726
11799	15777	11548	45918	45706	88554	75315	70233	72575
17843	64809	00390	11980	66129	07197	36712	55062	61191
42770	65397	45010	06463	86242	06361	14293	36343	97628
02410	96933	57864	93197	88227	57139	66382	95768	60660
70939	20457	62468	68698	74875	61111	59083	09152	93625
85616	15100	26242	28677	74655	05679	56676	67224	75318
85515	33174	05496	78789	81297	73985	82120	94070	20529
73466	06254	88113	98367	22018	99372	70171	52705	61202
72255	50729	05681	37216	09363	02385	93098	09502	92589
08121	48330	86725	52922	90349	81934	14849	68005	06791
94005	85164	22994	58921	85943	67506	79730	85382	61568
09108	52299	25991	00940	22493	60987	93573	79469	97147
85687	31723	67907	55306	71748	85048	17690	04784	98470
26190	02164	95889	89712	89795	73001	82210	39357	23867
34208	07539	60907	60693	01965	43492	46688	28891	23410
13032	78798	21733	35703	71707	11931	93513	78339	74754
16801	05582	47975	25046	59220	08275	67901	94954	36662
88735	91500	41654	97225	61188	24527	35220	99794	56097
82127	17594	94217	55324	06134	25207	26758	08687	06929
29284	42271	45833	19481	56972	99042	45304	39832	40188
56300	60964	13751	72385	91180	42371	55924	95783	33096
33132	33229	39955	16779	99286	23392	24255	90856	60004
65296	94444	32091	90681	95823	73091	92912	85979	30232
11069	52931	26381	71830	50467	47783	25223	81796	97745
06720	69637	99670	58392	57943	75965	14740	74814	75598
62719	14295	16605	13146	36992	50560	50121	90278	98283
95556	36672	87202	92730	81961	38894	61358	44519	71529
12490	12304	28804	42772	27104	35518	67361	84159	52442
29865	28847	70904	96638	54226	44701	67589	27352	81078
74486	63507	92193	65022	09583	43615	59910	05301	69347
01878	56351	68618	84432	30948	65180	75446	95963	75619
65405	25720	09364	51333	03752	65756	51967	92469	47296
31711	35173	45290	49326	50368	63829	05640	26675	27367
41028	50367	01904	68068	02324	58723	96333	77032	47878
76916	55336	48767	76915	79711	05182	70489	10244	45078
16404	93068	91519	85895	34872	24701	60932	91141	33252
06776	51133	76482	14812	19777	19614	51100	52943	04068
76818	05839	26058	80972	43337	24203	72345	37967	88138
16916	64028	38968	02783	63049	12261	89587	88988	88834
33696	41621	16648	11837	08094	38217	32919	16625	91567

TABLE 8 (continued)

00143	56431	90537	95332	29879	29363	48055	86410	10594
15932	59628	00086	74633	81208	05470	56385	23601	70545
86111	14530	39958	36155	60613	73849	74842	31030	30448
46218	36313	62063	59326	93522	48983	50335	30178	42755
84153	32199	77166	63912	07984	55369	56520	14633	00252
81439	35471	29742	57110	13710	21351	29816	32783	69004
93229	82043	80136	97269	28858	03036	01304	51363	40412
78421	33809	92792	96106	95191	43514	08320	25690	76117
44265	86707	80637	44879	81457	06781	11411	88804	62551
89430	51314	76126	62672	31815	12947	76533	19761	93373
36462	19901	02919	29311	31275	83593	34933	95758	63944
55996	59605	51680	27755	06077	12797	67082	12536	64069
69338	43838	06320	63988	16549	27931	27270	94711	47834
40276	17751	72508	23027	70257	42812	87319	09160	02913
67834	93014	07816	93085	14552	10115	87740	44125	51227

TABLE 9 The Normal Distribution (z). (Source: The entries in this table were computed by the author.)

z	Mean to z	Larger Portion	Smaller Portion	z	Mean to z	Larger Portion	Smaller Portion
.00	0.0000	0.5000	0.5000	.21	0.0832	0.5832	0.4168
.01	0.0040	0.5040	0.4960	.22	0.0871	0.5871	0.4129
.02	0.0080	0.5080	0.4920	.23	0.0910	0.5910	0.4090
.03	0.0120	0.5120	0.4880	.24	0.0948	0.5948	0.4052
.04	0.0160	0.5160	0.4840	.25	0.0987	0.5987	0.4013
.05	0.0199	0.5199	0.4801	.26	0.1026	0.6026	0.3974
.06	0.0239	0.5239	0.4761	.27	0.1064	0.6064	0.3936
.07	0.0279	0.5279	0.4721	.28	0.1103	0.6103	0.3897
.08	0.0319	0.5319	0.4681	.29	0.1141	0.6141	0.3859
.09	0.0359	0.5359	0.4641	.30	0.1179	0.6179	0.3821
.10	0.0398	0.5398	0.4602	.31	0.1217	0.6217	0.3783
.11	0.0438	0.5438	0.4562	.32	0.1255	0.6255	0.3745
.12	0.0478	0.5478	0.4522	.33	0.1293	0.6293	0.3707
.13	0.0517	0.5517	0.4483	.34	0.1331	0.6331	0.3669
.14	0.0557	0.5557	0.4443	.35	0.1368	0.6368	0.3632
.15	0.0596	0.5596	0.4404	.36	0.1406	0.6406	0.3594
.16	0.0636	0.5636	0.4364	.37	0.1443	0.6443	0.3557
.17	0.0675	0.5675	0.4325	.38	0.1480	0.6480	0.3520
.18	0.0714	0.5714	0.4286	.39	0.1517	0.6517	0.3483
.19	0.0753	0.5753	0.4247	.40	0.1554	0.6554	0.3446
.20	0.0793	0.5793	0.4207	.41	0.1591	0.6591	0.3409

z	Mean to z	Larger Portion	Smaller Portion	z	Mean to z	Larger Portion	Smaller Portion
.42	0.1628	0.6628	0.3372	.93	0.3238	0.8238	0.1762
.43	0.1664	0.6664	0.3336	.94	0.3264	0.8264	0.1736
.44	0.1700	0.6700	0.3300	.95	0.3289	0.8289	0.1711
.45	0.1736	0.6736	0.3264	.96	0.3315	0.8315	0.1685
.46	0.1772	0.6772	0.3228	.97	0.3340	0.8340	0.1660
.47	0.1808	0.6808	0.3192	.98	0.3365	0.8365	0.1635
.48	0.1844	0.6844	0.3156	.99	0.3389	0.8389	0.1611
.49	0.1879	0.6879	0.3121	1.00	0.3413	0.8413	0.1587
.50	0.1915	0.6915	0.3085	1.01	0.3438	0.8438	0.1562
.51	0.1950	0.6950	0.3050	1.02	0.3461	0.8461	0.1539
.52	0.1985	0.6985	0.3015	1.03	0.3485	0.8485	0.1515
.53	0.2019	0.7019	0.2981	1.04	0.3508	0.8508	0.1492
.54	0.2054	0.7054	0.2946	1.05	0.3531	0.8531	0.1469
.55	0.2088	0.7088	0.2912	1.06	0.3554	0.8554	0.1446
.56	0.2123	0.7123	0.2877	1.07	0.3577	0.8577	0.1423
.57	0.2157	0.7157	0.2843	1.08	0.3599	0.8599	0.1401
.58	0.2190	0.7190	0.2810	1.09	0.3621	0.8621	0.1379
.59	0.2224	0.7224	0.2776	1.10	0.3643	0.8643	0.1357
.60	0.2257	0.7257	0.2743	1.11	0.3665	0.8665	0.1335
.61	0.2291	0.7291	0.2709	1.12	0.3686	0.8686	0.1314
.62	0.2324	0.7324	0.2676	1.13	0.3708	0.8708	0.1292
.63	0.2357	0.7357	0.2643	1.14	0.3729	0.8729	0.1271
.64	0.2389	0.7389	0.2611	1.15	0.3749	0.8749	0.1251
.65	0.2422	0.7422	0.2578	1.16	0.3770	0.8770	0.1230
.66	0.2454	0.7454	0.2546	1.17	0.3790	0.8790	0.1210
.67	0.2486	0.7486	0.2514	1.18	0.3810	0.8810	0.1190
.68	0.2517	0.7517	0.2483	1.19	0.3830	0.8830	0.1170
.69	0.2549	0.7549	0.2451	1.20	0.3849	0.8849	0.1151
.70	0.2580	0.7580	0.2420	1.21	0.3869	0.8869	0.1131
.71	0.2611	0.7611	0.2389	1.22	0.3888	0.8888	0.1112
.72	0.2642	0.7642	0.2358	1.23	0.3907	0.8907	0.1093
.73	0.2673	0.7673	0.2327	1.24	0.3925	0.8925	0.1075
.74	0.2704	0.7704	0.2296	1.25	0.3944	0.8944	0.1056
.75	0.2734	0.7734	0.2266	1.26	0.3962	0.8962	0.1038
.76	0.2764	0.7764	0.2236	1.27	0.3980	0.8980	0.1020
.77	0.2794	0.7794	0.2206	1.28	0.3997	0.8997	0.1003
.78	0.2823	0.7823	0.2177	1.29	0.4015	0.9015	0.0985
.79	0.2852	0.7852	0.2148	1.30	0.4032	0.9032	0.0968
.80	0.2881	0.7881	0.2119	1.31	0.4049	0.9049	0.0951
.81	0.2910	0.7910	0.2090	1.32	0.4066	0.9066	0.0934
.82	0.2939	0.7939	0.2061	1.33	0.4082	0.9082	0.0918
.83	0.2967	0.7967	0.2033	1.34	0.4099	0.9099	0.0901
.84	0.2995	0.7995	0.2005	1.35	0.4115	0.9115	0.0885
.85	0.3023	0.8023	0.1977	1.36	0.4131	0.9131	0.0869
.86	0.3051	0.8051	0.1949	1.37	0.4147	0.9147	0.0853
.87	0.3078	0.8078	0.1922	1.38	0.4162	0.9162	0.0838
.88	0.3106	0.8106	0.1894	1.39	0.4177	0.9177	0.0823
.89	0.3133	0.8133	0.1867	1.40	0.4192	0.9192	0.0808
.90	0.3159	0.8159	0.1841	1.41	0.4207	0.9207	0.0793
.91	0.3186	0.8186	0.1814	1.42	0.4222	0.9222	0.0778
.92	0.3212	0.8212	0.1788	1.43	0.4236	0.9236	0.0764

z	Mean to z	Larger Portion	Smaller Portion	z	Mean to z	Larger Portion	Smaller Portion
1.44	0.4251	0.9251	0.0749	1.95	0.4744	0.9744	0.0256
1.45	0.4265	0.9265	0.0735	1.96	0.4750	0.9750	0.0250
1.46	0.4279	0.9279	0.0721	1.97	0.4756	0.9756	0.0244
1.47	0.4292	0.9292	0.0708	1.98	0.4761	0.9761	0.0239
1.48	0.4306	0.9306	0.0694	1.99	0.4767	0.9767	0.0233
1.49	0.4319	0.9319	0.0681	2.00	0.4772	0.9772	0.0228
1.50	0.4332	0.9332	0.0668	2.01	0.4778	0.9778	0.0222
1.51	0.4345	0.9345	0.0655	2.02	0.4783	0.9783	0.0217
1.52	0.4357	0.9357	0.0643	2.03	0.4788	0.9788	0.0212
1.53	0.4370	0.9370	0.0630	2.04	0.4793	0.9793	0.0207
1.54	0.4382	0.9382	0.0618	2.05	0.4798	0.9798	0.0202
1.55	0.4394	0.9394	0.0606	2.06	0.4803	0.9803	0.0197
1.56	0.4406	0.9406	0.0594	2.07	0.4808	0.9808	0.0192
1.57	0.4418	0.9418	0.0582	2.08	0.4812	0.9812	0.0188
1.58	0.4429	0.9429	0.0571	2.09	0.4817	0.9817	0.0183
1.59	0.4441	0.9441	0.0559	2.10	0.4821	0.9821	0.0179
1.60	0.4452	0.9452	0.0548	2.11	0.4826	0.9826	0.0174
1.61	0.4463	0.9463	0.0537	2.12	0.4830	0.9830	0.0170
1.62	0.4474	0.9474	0.0526	2.13	0.4834	0.9834	0.0166
1.63	0.4484	0.9484	0.0516	2.14	0.4838	0.9838	0.0162
1.64	0.4495	0.9495	0.0505	2.15	0.4842	0.9842	0.0158
1.65	0.4505	0.9505	0.0495	2.16	0.4846	0.9846	0.0154
1.66	0.4515	0.9515	0.0485	2.17	0.4850	0.9850	0.0150
1.67	0.4525	0.9525	0.0475	2.18	0.4854	0.9854	0.0146
1.68	0.4535	0.9535	0.0465	2.19	0.4857	0.9857	0.0143
1.69	0.4545	0.9545	0.0455	2.20	0.4861	0.9861	0.0139
1.70	0.4554	0.9554	0.0446	2.21	0.4864	0.9864	0.0136
1.71	0.4564	0.9564	0.0436	2.22	0.4868	0.9868	0.0132
1.72	0.4573	0.9573	0.0427	2.23	0.4871	0.9871	0.0129
1.73	0.4582	0.9582	0.0418	2.24	0.4875	0.9875	0.0125
1.74	0.4591	0.9591	0.0409	2.25	0.4878	0.9878	0.0122
1.75	0.4599	0.9599	0.0401	2.26	0.4881	0.9881	0.0119
1.76	0.4608	0.9608	0.0392	2.27	0.4884	0.9884	0.0116
1.77	0.4616	0.9616	0.0384	2.28	0.4887	0.9887	0.0113
1.78	0.4625	0.9625	0.0375	2.29	0.4890	0.9890	0.0110
1.79	0.4633	0.9633	0.0367	2.30	0.4893	0.9893	0.0107
1.80	0.4641	0.9641	0.0359	2.31	0.4896	0.9896	0.0104
1.81	0.4649	0.9649	0.0351	2.32	0.4898	0.9898	0.0102
1.82	0.4656	0.9656	0.0344	2.33	0.4901	0.9901	0.0099
1.83	0.4664	0.9664	0.0336	2.34	0.4904	0.9904	0.0096
1.84	0.4671	0.9671	0.0329	2.35	0.4906	0.9906	0.0094
1.85	0.4678	0.9678	0.0322	2.36	0.4909	0.9909	0.0091
1.86	0.4686	0.9686	0.0314	2.37	0.4911	0.9911	0.0089
1.87	0.4693	0.9693	0.0307	2.38	0.4913	0.9913	0.0087
1.88	0.4699	0.9699	0.0301	2.39	0.4916	0.9916	0.0084
1.89	0.4706	0.9706	0.0294	2.40	0.4918	0.9918	0.0082
1.90	0.4713	0.9713	0.0287	2.41	0.4920	0.9920	0.0080
1.91	0.4719	0.9719	0.0281	2.42	0.4922	0.9922	0.0078
1.92	0.4726	0.9726	0.0274	2.43	0.4925	0.9925	0.0075
1.93	0.4732	0.9732	0.0268	2.44	0.4927	0.9927	0.0073
1.94	0.4738	0.9738	0.0262	2.45	0.4929	0.9929	0.0071

- Berndt, D. J., Schwartz, S., & Kaiser, C. F. (1983). Readability of self-report depression inventories. *Journal of Consulting and Clinical Psychology*, 51, 627-628.
- Blanchard, E. B., Theobald, D. E., Williamson, D. A., Silver, B. V., & Brown, D. A. (1978). Temperature biofeedback in the treatment of migraine headaches. *Archives of General Psychiatry*, 35, 581-588.
- Bradley, D. R., & Kjungja, L. (1982). Animated subjective contours. *Perception and Psychophysics*, 32, 393-395.
- Bradley, J. V. (1963, March). *Studies in Research Methodology: IV. A Sampling Study of the Central Limit Theorem and the Robustness of One-Sample Parametric Tests*. AMRL Technical Documentary Report 63-29, 650th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, OH.
- Bradley, J. V. (1968). *Distribution-Free Statistical Tests*. Englewood Cliffs, NJ: Prentice-Hall.
- Brown, J., Lewis, V., Brown, M., Horn, G., & Bowes, J. B. (1982). A comparison between transient amnesias induced by two drugs (diazepam and lorazepam) and amnesia of organic origin. *Neuropsychologia*, 20, 55-70.
- Camilli, G., & Hopkins, K. D. (1978). Applicability of chi-square to 2×2 contingency tables with small expected frequencies. *Psychological Bulletin*, 85, 163-167.
- Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The Quality of American Life*. New York: Russell Sage Foundation.
- Clark, K. B., & Clark, M. K. (1939). The development of consciousness of self in the emergence of racial identification in Negro pre-school children. *Journal of Social Psychology*, 10, 591-599.
- Cochran, C. D., & Urbanczyk, S. (1982). The effect of availability of vertical space on personal space. *Journal of Psychology*, 111, 137-140.
- Cohen, J. (1977). *Statistical Power Analysis for Behavioral Sciences* (Rev. ed.). New York: Academic Press.
- Darley, J. M., and Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377-383.
- Doob, A. N., & Gross, A. E. (1968). Status of frustrator as an inhibitor of horn-honking responses. *Journal of Social Psychology*, 76, 213-218.
- Eron, L. D., Huesmann, L. R., Lefkowitz, M. M., & Walden, L. O. (1972). Does television violence cause aggression? *American Psychologist*, 27, 253-263.
- Fagerström, K. (1982). A comparison of psychological and pharmacological treatment of smoking cessation. *Journal of Behavioral Medicine*, 5, 343-351.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.

- Goldstein, A. G., Bailis, K., & Chance, J. E. (1983). Do students remember pictures in psychology textbooks? *Teaching of Psychology*, 10, 23-26.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-112.
- Harper, D., & Wacker, D. P. (1983). The efficiency of the Denver Developmental Screening Test with rural disadvantaged preschool children. *Journal of Pediatric Psychology*, 8, 273-283.
- Hays, W. L. (1981). *Statistics* (3rd ed.). New York: Holt, Rinehart and Winston.
- Hicks, R. A., & Guista, M. (1982). The energy level of habitual long and short sleepers. *Bulletin of the Psychonomic Society*, 19, 131-132.
- Hindley, C. B., Filliozat, A. M., Klackenberg, G., Nicolet-Meister, D., & Sand, E. A. (1966). Differences in age of walking for five European longitudinal samples. *Human Biology*, 38, 364-379.
- Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, 11, 213.
- Holway, A. H., & Boring, E. G. (1940). The moon illusion and the angle of regard. *American Journal of Psychology*, 53, 509-516.
- Hosch, H. M., & Cooper, D. S. (1982). Victimization as a determinant of eyewitness accuracy. *Journal of Applied Psychology*, 67, 649-652.
- Howell, D. C. (1987). *Statistical Methods for Psychology*. (2nd ed.) Boston: Duxbury Press.
- Howell, D. C., & Huessy, H. R. (1985). A fifteen year follow-up of a behavioral history of Attention Deficit Syndrome (ADD). *Pediatrics*, 76, 185-190.
- Hraba, J., & Grant, G. (1970). Black is beautiful: A reexamination of racial preference and identification. *Journal of Personality and Social Psychology*, 16, 398-402.
- Huff, D. (1954). *How to Lie with Statistics*. New York: W. W. Norton.
- Hunt, J. McV. (1941). The effects of infant feeding frustration upon adult hoarding behavior. *Journal of Abnormal and Social Psychology*, 36, 338-360.
- Jackson, J. M., & Padgett, V. R. (1982). With a little help from my friend: Social loafing and the Lennon-McCartney songs. *Personality and Social Psychology Bulletin*, 8, 672-677.
- Kapp, B., Frysinger, R., Gallagher, M., & Hazelton, J. (1979). Amygdala central nucleus lesions: Effects on heart rate conditioning in the rabbit. *Physiology and Behavior*, 23, 1109-1117.
- Kaufman, L., & Rock, I. (1962). The moon illusion, I. *Science*, 136, 953-961.
- Kennedy, D., Beard, D., & Carr, W. J. (1982). Differential recognition of the left vs. right side of human faces. *Bulletin of the Psychonomic Society*, 20, 72-73.
- Knehr-McDonald, P. (1984). Carelessness: Its relationship to cognitive style and problem-solving strategies. Unpublished Ph.D. dissertation, University of Vermont, Burlington.
- Latané, B., & Dabbs, J. M., Jr. (1975). Sex, group size, and helping in three cities. *Sociometry*, 38, 180-194.
- Linn, R. T., & Hodge, G. K. (1982). Locus of control in childhood hyperactivity. *Journal of Consulting and Clinical Psychology*, 50, 592-593.
- Lord, F. M. (1953). On the statistical treatment of football numbers. *American Psychologist*, 8, 750-751.

- Lundberg, U. (1983). Note on Type A behavior and cardiovascular responses to challenge in 3-6 year old children. *Journal of Psychosomatic Research*, 27, 39-42.
- McConaughy, S. H. (1980). Cognitive structures for reading comprehension: Judging the relative importance of ideas in short stories. Unpublished Ph.D. dissertation, University of Vermont, Burlington.
- Neter, J., & Wasserman, W. (1974). *Applied Linear Statistical Models*. Homewood, IL: Richard D. Irwin.
- Newman, S. E., Olson, M. A., Hall, A. D., & Hornak, R. (1983). Effects of encoding and retrieval contexts on recall. *Bulletin of the Psychonomic Society*, 21, 4-6.
- Nurcombe, B., & Fitzhenry-Coor, I. (1979, October). Decision making in the mental health interview: I. An introduction to an education and research program. Paper delivered at the Conference on Problem Solving in Medicine, Smuggler's Notch, VT.
- Nurcombe, B., Howell, D. C., Rauh, V. A., Teti, D. M., Ruoff, P., & Brennan, J. (1984). An intervention program for mothers of low-birthweight infants: Preliminary results. *Journal of the American Academy of Child Psychiatry*, 23, 319-325.
- Obrzut, J. E., Hansen, R. L., & Heath, C. P. (1982). The effectiveness of visual information processing training with Hispanic children. *Journal of General Psychology*, 107, 165-174.
- Overall, J. E., & Klett, C. J. (1972). *Applied Multivariate Analysis*. New York: McGraw-Hill.
- Payne, S. L. (1982). Job-orientation stereotyping: Is it changing? *Journal of Psychology*, 111, 51-55.
- Pliner, P. (1982). The effects of mere exposure on liking for edible substances. *Appetite*, 3, 283-290.
- Robinson, B. E., Barrett, R. L., & Skeen, P. (1983). Locus of control of unwed adolescent fathers versus adolescent nonfathers. *Perceptual and Motor Skills*, 56, 397-398.
- Ryan, T., Joiner, B., & Ryan, B. (1985). *Minitab Student Handbook*. Boston: Duxbury Press.
- Saint-Exupery, A. de. (1943). *The Little Prince*. Tr. by Woods, K. New York: Harcourt Brace Jovanovich.
- Sgro, J. A., & Weinstock, S. (1963). Effects of delay on subsequent running under immediate reinforcement. *Journal of Experimental Psychology*, 66, 260-263.
- Smith, D. B., & Plant, W. T. (1982). Sex differences in job satisfaction of university professors. *Journal of Applied Psychology*, 67, 249-251.
- Sternglass, E. J., & Bell, S. (1983). Fallout and SAT scores: Evidence for cognitive damage during early infancy. *Phi Delta Kappan*, 64, 539-549.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: John Wiley.
- Supramaniam, S. (1983). Proofreading errors in good and poor readers. *Journal of Experimental Child Psychology*, 36, 68-80.
- Sutherland, R. J., & Linggard, R. (1982). Being there: A novel demonstration of latent spatial learning in the rat. *Behavioral and Neural Biology*, 36, 103-107.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- U.S. Department of Commerce. (1977). *Social Indicators, 1976*. Washington, D.C.: U.S. Government Printing Office.

- Velleman, P., & Hoaglin, D. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: Duxbury Press.
- Verdooren, L. R. (1963). Extended tables of critical values for Wilcoxon's test statistic. *Biometrika*, 50, 177-186.
- Vermont Department of Health. (1982). *1981 Annual Report of Vital Statistics in Vermont*. Burlington, VT.
- Wainer, H. (1984). How to display data badly. *American Statistician*, 38, 137-147.
- Welkowitz, J., Ewen, R., & Cohen, J. (1982). *Introductory Statistics for the Behavioral Sciences* (3rd ed.). New York: Academic Press.
- Winer, B. J. (1971). *Statistical Principles in Experimental Design* (2nd ed.). New York: McGraw-Hill.
- Younger, M. S. (1985). *A First Course in Linear Regression*. Boston: Duxbury Press.

پیوست د:

پاسخ تمرین‌های انتخابی

پاسخ‌های فصل ۱

- ۱-۱ وقتی کل دانشجویان یک دانشگاه یک جمعیت به حساب می‌آید که بخواهیم در باره نقطه‌نظرات، رفتار و مسائل دیگر مربوط به دانشجویان خود این دانشگاه مطالعه کنیم.
- ۱-۲ زمانی کل دانشجویان یک دانشگاه یک نمونه به حساب می‌آید که بخواهیم در باره همه دانشجویان کشور نتیجه‌گیری کنیم.
- ۱-۳ این یک نمونه تصادفی نخواهد بود زیرا همه دانشجویان عضو این جمعیت شانس یکسان برای انتخاب شدن در نمونه را ندارند.
- ۱-۴ نام همه ساکنان این شهر در فهرست تلفن وجود ندارد. افراد غیر مقیم، افراد فقیر، و بخصوص زنان و کودکان کمتر شانس انتخاب شدن دارند.
- ۱-۵ در صورتی که منازل یا آپارتمان‌ها دارای شماره باشند می‌توانیم از آنها تعدادی را بطور تصادفی انتخاب کنیم و سپس با استفاده از جدول اعداد تصادفی افراد را انتخاب کنیم.
- ۱-۶ معدل، میانگین، میانه، دامنه.
- ۱-۷ میانگین دریافت کالری آمریکایی‌های تحت پوشش تأمین اجتماعی می‌تواند از اهمیت خاصی برخوردار باشد.
- ۱-۸ آیا میانگین وزن یک گروه زنان ۳۰ ساله که از نوجوانی دارای رژیم غذایی بوده‌اند با میانگین وزن نمونه‌ای از زنان ۳۰ ساله همتا که از نوجوانی رژیم غذایی نداشته‌اند متفاوت است؟
- ۱-۹ داده‌های مقوله‌ای: (الف) تعداد کودکان یک نمونه که هر یک از چهار مرحله رشد اخلاقی را نشان می‌دهند. (ب) تعداد افراد در یک ساختمان اداری بزرگ که از وسایل نقلیه عمومی استفاده می‌کنند در مقابل کسانی که از ماشین شخصی برای رفتن به محل کار استفاده می‌کند.
- ۱-۱۰ داده‌های اندازه‌ای: (الف) نزدیک شدن به محرک برانگیزنده‌ی ترس. (ب) سرعت نبض هنگام حرکات سریع چشم (رم) در رؤیا. (ج) نمره سیاهه افسردگی بک.

۱-۱۲ یک سازه شخصیت در فلسفه تمرکز قدرت را می‌توان یا به صورت یک متغیری نسبتاً پیوسته (مثل تعداد سؤال‌های تأیید قدرت‌طلبی) یا به صورت یک طبقه‌بندی سه درجه‌ای قدرت‌طلبی، خنثی یا عدالت‌خواه طبقه‌بندی کرد.

۱-۱۳ ما ممکن است علاقه‌مند باشیم روابط بین رشد شناختی که در ۲ سالگی و نیز ۱۸ سالگی سنجیده شده را بدانیم- آیا آزمودنی‌هایی که در ۲ سالگی خوب عمل کرده‌اند در ۱۸ سالگی هم خوب عمل کرده‌اند؟ (ب) ممکن است مایل باشیم رابطه بین تعداد دفعاتی که یک آزمودنی فهرستی از لغات را تمرین کرده و تعداد لغاتی را که در یک آزمون بعدی بدرستی بیاد آورده بدانیم.

۱-۱۴ (الف) آیا والدینی که در باره مشکلات نوجوانان خدمات مشاوره‌ای دریافت کرده‌اند به کودکان خود بهتر از والدینی که دریافت نکرده‌اند پاسخ می‌دهند. (ب) آیا رشته‌های علوم پایه در یک درس منطق بهتر از دوره‌های علوم اجتماعی عمل می‌کنند؟

پاسخ‌های فصل ۲

۲-۱ (الف) اسمی- رنگ مو (ب) رتبه‌ای- ترتیب رسیدن به خط پایان در یک مسابقه دو. (ج) فاصله‌ای- مجموعه تاریخ‌های تکمیل یک تکلیف توسط آزمودنی‌ها. (د) نسبی- تعداد مسأله‌های درست حل شده در یک تکلیف.

۲-۲ میزان حقوق یک مقیاس نسبی از درآمد است، اما در بهترین حالت یک مقیاس ترتیبی از سهم یک شخص در شرکت است.

۲-۳ این سنجش ضعیفی از یادگیری است مگر آنکه فرض کنیم این حیوان که ناگهانی به خواب رفت همه چیز را در باره این کار فراموش کرده بود.

۲-۴ این به نظر می‌رسد برای انگیزش شاخص خیلی بهتری است.

۲-۵ (الف) متغیرهای مستقل- خواننده خوب در مقابل خواننده ضعیف؛ مرد در برابر زن. (ب) متغیرهای وابسته- سرعت خواندن؛ نمره میزان اضطراب.

۲-۶ آزمایش تفاوت خوانایی دست نوشته (متغیر وابسته) بین آزمودنی‌های راست دست و چپ دست (متغیر مستقل).

۲-۷ (الف) زمان کامل کردن یک کار. (ب) وزن. (ج) طول مدت بارداری.

۲-۸ (الف) جواب درست و نادرست دادن به یک پرسش. (ب) تعداد موارد درست در یک آزمون ۵ سؤالی. (ج) تعداد نمره منفی در راندگی.

۲-۹ (الف) ۹، ۱۰ و ۸ (ب) ۷۷ (ج) $\sum_{i=1}^{10} X_i$

۲-۱۰ (الف) ۹، ۲ (ب) ۵۷

۲-۱۱ (الف) ۵، ۹۲۹، ۶۵۷ (ب) ۷/۷ (ج) معدل (میانگین)

$$۲-۱۲ \quad (الف) ۳, ۲۴۹, ۳۷۷ \quad (ب) ۵/۷۸۹ \quad (ج) ۲/۴۰۶$$

$$۲-۱۳ \quad (الف) ۴۶۰ \quad (ب) ۴۳۸۹ \quad (ج) ۲/۳۴۴$$

$$۲-۱۴ \quad (لف) \sum X + \sum Y = ۱۳۴ = ۷۷ + ۵۷ = \sum X + \sum Y = \sum (X + Y) = (۱۰ + ۹) + (۸ + ۹) + \dots + (۷ + ۲)$$

$$(ب) \sum XY = ۴۶۰; \sum X \sum Y = ۴۳۸۹$$

$$(ج) \sum CX = \sum ۳X = ۳(۱۰) + ۳(۸) + \dots + ۳(۷) = ۲۳۱ = ۳(۱۰) = ۳(۷۷) = C \sum X$$

$$(د) \sum X^2 = ۶۵۷; (\sum X)^2 = ۵۹۲۹$$

پاسخ های فصل ۳

۳-۱ (ب) یک نمایی و چوله مثبت.

۳-۳ مشکل طراحی یک نمودار ساقه-برگی برای داده های تمرین ۳-۱ این است که اگر از اعداد معمول ۱۰ رقمی برای ساقه ها استفاده کنیم همه ارزش های تنها در دو برگه جا می گیرند. اگر تعداد ساقه ها را دو برابر کنیم هم این مشکل خیلی بهتر نخواهد شد. به جای آن از واحدهای عددی برای ساقه ها استفاده کنید و یک مقوله ی "کلی" برای ارزش های بالا و پایین اضافه کنید.

۳-۴ (الف) نمره های بزرگسالان به نظر می رسد خیلی کوچک تر است.

۳-۱۰ یک نمودار دو نمایی خواهد شد با یک قله در ۰ و یک قله در حدود یک بسته (۲۰) در روز.

۳-۱۳ (۱) مکزیکو دارای تعداد زیادی جمعیت جوان و تعداد کمی افراد پیر است، در حالی که اسپانیا دارای یک توزیع بیشتر مسطح است. (۲) فرق میان زنان و مردان در اسپانیا بیشتر نمودار است. (۳) نرخ بیشتر مرگ و میر نوزادان را در مکزیکو می توان دید.

۳-۱۴ ما از HI و LO در نمودار ساقه-برگی استفاده می کنیم تا از افتادن به ورطه حدود نهایی جلوگیری کنیم.

پاسخ های فصل ۴

۴-۱ رتبه درصدی، نقطه درصدی ۸۳ ام.

راهنمایی برای تمرین های ۴-۳ تا ۴-۲. همه نقاط درصدی و رتبه های درصدی که در پاسخ ها بعدی آمده رُند شده اند و به صورت تقریبی باید در نظر گرفته شوند. بسته به اینکه مقادیر را چگونه گروه بندی کنید اگر این کار را انجام دهید پاسخ دقیق ممکن است چند درجه فرق کند.

۴-۳ ۴۴ ۴-۴ ۵۲ ۴-۵ ۳۸ ۴-۶ ۳۹ ۴-۸ ۱۳، میانه ۴-۹ ۱۰ ۴-۱۰ ۹۷

۴-۱۱ ۵۱

۴-۱۳ ۸۶ ۴-۱۴ ۶۴ ۴-۱۵ ۲/۹۳ ۴-۱۶ ۶۵ ۴-۱۷ ۶۱ ۴-۱۸ ۱/۰ ۴-۱۹ ۹۳-

۴-۲۰ ۲ نه.

پاسخ‌های فصل ۵

- ۵-۱ نما = ۱۸؛ میانه = ۱۸؛ میانگین = $۱۸/۹$.
- ۵-۲ نما = ۱۰؛ میانه = ۱۰؛ میانگین = $۱۰/۲$.
- ۵-۳ بزرگسالان تقریباً به اندازه نصف کودکان می‌گویند "و بعد...".
- ۵-۵ میانگین در اطراف میانه قرار می‌گیرد.
- ۵-۶ ۱۵ ۱۵ ۱۰ ۹ ۱ ۵-۶
- ۵-۷ میانگین = $۲۱/۳۳$ ؛ میانه = ۲۱.

پاسخ‌های فصل ۶

- ۶-۱ دامنه = ۳۰؛ واریانس = $۲۰/۲۱۴$ ؛ انحراف معیار = $۴/۴۹۶$.
- ۶-۲ دامنه = ۱۶؛ واریانس = $۱۱/۵۹۲$ ؛ انحراف معیار = $۳/۴۰۵$.
- ۶-۳ دو انحراف معیار خیلی شبیه هم هستند، هر چند دامنه برای کودکان تقریباً دو برابر بزرگسالان است.
- ۶-۴ فاصله $\bar{X} \pm 2S_x = ۹/۹۰۸ - ۲۷/۲۹۸$ شامل ۹۶٪ نمرات است.
- ۶-۵ فاصله $\bar{X} \pm 2S_x = ۳/۳۹ - ۱۷/۰۱$ شامل ۹۶٪ نمرات است.
- ۶-۸ ۳/۳۳۳، ۴/۲۸۶، ۴/۲۸۶، ۲/۸۵۷، ۲/۸۰۹، ۱/۴۲۸، ۳/۸۰۹، ۲/۳۸۱
- ۶-۹ ۰/۸۹۳، ۰/۵۳۶، ۰/۸۴۵، ۰/۵۳۶، ۰/۴۱۷، ۰/۰۱۲، ۱/۰۱۲، ۰/۰۶۰
- ۶-۱۳ بر دامنه اثری ندارد. انحراف معیار و واریانس افزایش خواهند یافت.
- ۶-۱۴ هرچند معمولاً تنها یک نمونه از جامعه بیرون کشیده می‌شود، ما می‌خواهیم بدانیم آیا آماره‌هایی که از این نمونه حساب کرده‌ایم همانند آماره‌هایی می‌شد که با یک نمونه دیگر بدست می‌آید.
- ۶-۱۵ ما یک آماره غیر سوگیرانه می‌خواهیم زیرا چیزی می‌خواهیم که برآورد دقیقی از پارامتر جمعیت باشد - یعنی، فرق اساسی با آن پارامتر جمعیت نداشته باشد.

پاسخ‌های فصل ۷

- ۷-۲ برای $X = ۲/۵$ ، $z = -۰/۹۲$ ، $۱۸٪$ از توزیع پایین‌تر از $X = ۲/۵$ قرار می‌گیرد؛ برای $X = ۶/۲$ ، $۱/۳۵٪$ از توزیع پایین‌تر از $X = ۶/۲$ قرار می‌گیرد؛ برای $X = ۹$ ، $۳/۰۶٪$ از توزیع پایین‌تر از $X = ۹$ قرار می‌گیرد.
- ۷-۳ (الف) ۶۸٪ (ب) ۵۰٪ (ج) ۸۴٪
- ۷-۴ (الف) $۹۸۵/۱۲۵ \leq X \leq ۹۶۴/۸۷۵$ ؛ (ب) $X = ۹۸۵/۱۲۵$ ؛ (ج) $۱۰۰۴/۴ \leq X \leq ۹۴۵/۶$

۷-۵ $z = -1/67 = (950 - 975)/15$ ؛ با آنچه که ما از این توزیع می دانیم تنها در ۴/۷۵٪ اوقات ما انتظار یک شماره به کوچکی ۹۵۰ را خواهیم داشت.

۷-۶ (ب) ۱۵/۸۷٪ (ج) ۳۰/۸۵٪

۷-۷ وقتی دو توزیع دارای انحراف معیار برابر باشند پاسخ بخش های (الف) و (ب) تمرین ۷-۶ مساوی است.

۷-۸ $X = \bar{X} - 1/28(30) = 111/6$ ؛ $z = -1/28$

۷-۹ (الف) $z = 1/28$ ؛ $X = \bar{X} - 1/28(400) = 2512$ ؛ (ب) $z = -1/645$ ؛ $X = \bar{X} - 1/645(400) = 1342$ ؛ $X = \bar{X}$

۷-۱۰ (ب) $z = 2/86 = (50 - 30)/7$ ؛ تنها ۰/۲٪ اوقات انتظار خواهیم داشت وقتی این دانشجو آگاهانه از یک توزیع با میانگین ۳۰ و انحراف معیار ۷ نمونه گیری شده باشد به نتیجه ای به این بزرگی دست یابیم. تصور می شود او این داده ها را ساخته است.

۷-۱۱ نمرات خام را در ۱۰/۷ ضرب کنید تا انحراف معیار به ۱۰ برسد، و سپس ۱۱/۴۳ نمره به هر نمره جدید اضافه کنید تا میانگین به ۸۰ برسد.

۷-۱۲ (ب) پیشنهاد می کنیم که او مجموعه نمرات را بگیرد و به طرر تجربی (مثلا با شمردن) نقطه ای که ۱۰٪ نمرات زیر آن قرار می گیرند را تعیین کند.

۷-۱۳ $z = 0/88 = (600 - 489)/126$ ؛ $z = 0/81$ ؛ از نمرات پایین تر از این قرار می گیرند، طوری که ۶۰۰ نشان دهنده نقطه درصدی ۸۱ ام است.

۷-۱۴ $z = 0/675$ ؛ $z = 0/675 = (X - 489)/126$ ؛ $X = 574/05$

۷-۱۵ ۶۰۰ در نقطه درصدی ۷۹ ام قرار دارد، و نقطه درصدی ۷۵ ام ۵۸۶/۶۵ است.

۷-۱۶ رتبه های درصدی آشکارا به گروه مرجع بستگی دارند.

۷-۱۷ پاسخ با استفاده از توزیع نرمال بطور قابل توجهی بزرگ تر بود. این تفاوت به میزان زیادی ناشی از شکل چوله توزیع است.

پاسخ های فصل ۸

۸-۱ تحلیلی- یک موش در یک ماز که بطور تصادفی پاسخ می دهد احتمال ۰/۵۰ وجود دارد که از یک نقطه انتخابی برگردد. فراوانی نسبی- یک موش که در ۷۰۰ بار از ۱۰۰۰ تلاش خود برگشته دارای احتمال ۰/۷۰ است که این بار برگردد (با فرض اینکه روالی در داده ها برای تمرین ها وجود ندارد). ذهنی- "من این آزمایش را دارای یک شانس ۷۰٪ می دانم که به نتایج معمولی برسد."

۸-۲ (الف) $0/01 = 1/100$ ؛ (ب) $0/02 = 2/100$ ؛ (ج) $0/03 = 3/100$

۸-۳ (الف) $0/11 = 0/9$ ؛ (ب) $(1/9) \times (2/10) = 0/22$ ؛ (ج) $(2/9) \times (1/10) = 0/22$ ؛ (د) $0/44$

۸-۴ (ب)، (ج)، و بخشی از (د)

۸-۵ (الف)

۸-۸ $(2/24) \times (3/24) = 6/576 = 0/01$

۸-۹ $(2/13) \times (3/13) = 6/169 = 0/36$

۸-۱۰ $0/3278 + 0/1132 = 0/4410$

۸-۱۱ احتمال اینکه شما ۲۰ ساله باشید.

۸-۱۲ طبقه‌بندی طیف دیداری به حدود هفت رنگ.

۸-۱۳ وابستگی حزبی سیاسی؛ وضعیت اجتماعی اقتصادی به صورتی که توسط هولینگ‌شید و ریدلیچ طبقه‌بندی شده است.

۸-۱۴ $10/1000 = 0/01$

۸-۱۵ (الف) ۲۰۰ نفر بالای نقطه درصدی ۸۰ ام هستند. بنابراین $0/05 = 10/200 = p$. (ب) کسی زیر نقطه درصدی ۸۰ ام نیست. بنابراین $p = 0/00$.

پاسخ تمرین‌های فصل ۹

۹-۲ $r = 0/62$ ۹-۳ $r = 0/35$

۹-۷ (الف) $4/67$ ، $3/33$ ، $4/67$ ؛ (ب) $4/67$ ، $3/33$ ، $4/67$

۹-۸ $r = 0/99$ ، $0/71$ ، $0/99$ ؛ سه راه ممکن: ۲ ۸ ۶ ۴، ۴ ۶ ۲ ۸، و ۶ ۲ ۸ ۴.

۹-۹ (ب) $r = 0/74$

۹-۱۰ همبستگی بین ید-۱۳۱ و آزمون کلامی برابر $0/2408$ است. همبستگی بین ید-۱۳۱ و ریاضی برابر $0/5501$ است. نه. چنانچه گزارش شده همبستگی بزرگ با ریاضی SAT است و نه با آزمون کلامی.

۹-۱۲ نمودار پراکنش گویای آن است که یک یا دو حالت انتهایی نتایج را خیلی به هم ریخته کرده است.

۹-۱۳ در محاسبه همبستگی هر نقطه از داده‌ها (حالت) بطور برابر شمرده می‌شود؛ اما چون بعضی نقاط داده‌ای بر اساس تعداد خیلی کم دانشجویان است، این حالت‌ها در همبستگی‌ها بیشتر نمود پیدا می‌کند.

۹-۱۴ ید $131 \times$ آزمون کلامی $r = -0/04$ ؛ غیر معنادار. ید $131 \times$ آزمون ریاضی $r = -0/26$ ؛ غیر معنادار.

۹-۱۸ $r = 0/80$ ۹-۱۹ $r = 0/44$

۹-۲۰ بلی. همبستگی (۲) باز هم به شما می گوید یک خط مستقیم تا چه حد مطابقت دارد، حتی اگر شما فکر کنید یک خط منحنی بهتر تطبیق می کند. بیشتر وقت های تطبیق یک خط مستقیم برای اهداف ما به خوبی بسنده می کند.

۹-۲۱ وقتی می گوییم یک ضریب همبستگی پایا است منظور این است که اگر نمونه های پی در پی از همان جمعیت گرفته شود ضریب همبستگی برای این نمونه ها به اندازه همان مقدار کلی است. همبستگی هایی که بر اساس نمونه های کوچک بدست آمده باشد بیشتر وقت ها ناپایدار هستند زیرا نقاط داده ای غیرعادی می توانند اثر مهمی بر ضریب همبستگی بگذارند.

پاسخ تمرین های فصل ۱۰

$$\hat{Y} = 0.069X + 3/53 \quad 10-1$$

$$S_{Y-\hat{Y}} = 0.5796 \quad 10-2$$

$$10-3 \quad \text{پیش بینی وزن هنگام تولد} > 2500 \text{ گرم برابر } 8/36 \text{ خواهد بود.}$$

$$10-4 \quad \text{ما راهی را فراتر از دامنه داده هایی که معادله بر اساس آن بنا شده است پیش بینی می کنیم.}$$

$$\hat{Y} = 0.475X + 43/16 \quad 10-5$$

$$\bar{X} = 1/15Y + 17/65 \quad 10-6$$

$$\hat{Y} = 618/91 \quad 10-7$$

$$\hat{Y} = 1/0.5M + 17/33 \quad 10-10 \quad \hat{Y} = 0.138X - 1/54 \quad 10-9 \quad \hat{Y} = 635 = \bar{Y} \quad 10-8$$

۱۰-۱۱ یک تفاوت یک واحدی در نمره کلامی SAT با ۰/۸۸۵ تفاوت در مقدار پیش بینی شده نمره ریاضی SAT همراه است. مقدار ثابت دارای معنای قابل تفسیری نیست چون یک نمره ۰ مقدار درستی نیست.

۱۰-۱۴ بهترین برآورد حقوق اولیه برای عضو هیئت علمی ۱۵۰۰۰ دلار است. برای هر سال اضافی متوسط افزایش ۹۰۰ دلار است. برای کارکنان اداری بهترین برآورد حقوق اولیه ۱۰۰۰۰ دلار، اما برای هر سال افزایش خدمت بطور متوسط ۱۵۰۰ دلار حقوق اضاف می شود. آنها در سال ۸/۵ خدمت برابر خواهند بود.

پاسخ تمرین های فصل ۱۱

۱۱-۱ ما این فرض صفر را بنا می کنیم که بازی شب گذشته واقعا یک بازی بسکتبال بوده است. بر اساس این فرض انتظار داریم که هر تیم چیزی بین ۰ و ۶ امتیاز بدست آورد. سپس به امتیازهای واقعی

نگاه کرده و نتیجه را بررسی می‌کردیم که ببینیم آیا با آنچه انتظار داشتیم در صورت واقعی بودن بازی وجود می‌آید. بنابراین فرض صفر را رد می‌کردیم.

۱۱-۲ (ب) نه (ج) این فرض را بنا می‌کنیم که به درستی کسر شده است. بنابراین انتظار داریم که حدود ۳ دلار با ۲۵ سنت کمتر یا بیشتر پس بگیریم. مقداری که پس گرفته‌ایم در راستای مورد انتظار است، و بنابر این دلیلی برای رد H_0 نداریم.

۱۱-۳ نتیجه‌گیری این که سر ما کلاه رفته وقتی که واقعا نرفته باشد.

۱۱-۴ نتیجه‌گیری این که سر ما کلاه نرفته وقتی که واقعا رفته باشد.

۱۱-۵ مقدار بحرانی آن مقدار تغییری خواهد بود که در نظر می‌گیریم کمتر از آن را حد کلاه‌گذاری بدانیم. ناحیه رد تمام میزان تغییر کمتر از مقدار بحرانی خواهد بود- یعنی تمام مقادیری که به رد H_0 منجر می‌شود.

۱۱-۶ ما یک آزمون یک دامنه در نظر می‌گیریم چنانچه بخواهیم مورد کلاه‌گذاری قرار گرفتن را مشخص کنیم اما به گرفتن پول زیادی نپردازیم. H_0 را رد نمی‌کنیم. مهم نیست چه مقدار پول اضافی گرفته‌ایم.

۱۱-۷ $z = (۴۹۰ - ۶۵۰) / ۵۰ = -۳/۲$ احتمال اینکه یک دانشجو که بطور تصادفی از کسانی که بدرستی قبول شده‌اند دارای نمره امتحان کارشناسی ارشد کمتر از ۴۹۰ باشد ۰/۰۰۰۶ است. فرض می‌کنیم این واقعیت که مادر او از کارمندان آنجا بوده در قبولی او نقش داشته است.

۱۱-۸ ما یک نمونه تصادفی از تمام دانشجویانی که امتحان کارشناسی ارشد داده‌اند را نگاه نمی‌کنیم، بلکه نمونه‌ای انتخابی از دانشجویان دارای نمرات بالا را می‌نگریم.

۱۱-۹ این توزیع در سمت راست به آرامی افت خواهد یافت به دلیل اینکه همیشه این اتفاق می‌افتد- افرادی که نمره خیلی بالا دارند تعدادشان کم است. توزیع در سمت چپ افت خواهد کرد چون از دانشجویان مرزی تعداد کمتری پذیرفته می‌شوند (فرقی نمی‌کند که حد مرزی چقدر بزرگ است).

۱۱-۱۰ تعداد زیادی نمونه بیرون خواهیم کشید. برای هر نمونه نما، دامنه و نسبت آنها (M) را حساب خواهیم کرد. سپس مقادیر بدست آمده M را رسم می‌کنیم.

۱۱-۱۱ M یک آزمون آماری خوانده می‌شود.

۱۱-۱۳ فرض مقابل این است که این دانشجو از جمعیتی نمونه‌گیری شده که میانگین آنها برابر ۶۵۰ نیست.

۱۱-۱۴ خطای نمونه‌گیری تغییرپذیری در یک آماره از نمونه‌ای به نمونه دیگر است که در اثر آن مشاهداتی که در یک نمونه وجود دارد رخ داده است.

۱۱-۱۵ کلمه توزیع اشاره به مجموعه مقادیر بدست آمده برای هر مشاهده دارد. عبارت توزیع نمونه‌گیری برای توزیعی از نتایج (خواه نظری یا تجربی) یک نمونه آماری به کار می‌رود.

۱۱-۱۶ اگر α کاهش یابد، β افزایش خواهد یافت و توان کاهش می‌یابد.

پاسخ تمرین های فصل ۱۲

۱۲-۵ (الف) $z = 3/86 = 0.002$; $P(z \geq \pm 3/86) = 0.002$. ما این فرض صفر را رد می کنیم که این نمرات از یک جمعیت با میانگین ۱۰۰ برگرفته شده اند. (ب) اگر از یک آزمون یک دامنه استفاده کنید که $500 < \mu$ ، شما فرض صفر را رد نخواهید کرد.

۱۲-۶ این یک نمونه تصادفی نیست. (۲) ما تعریفی از معنای عبارت "یک مشکل حاد آموزشی" نداریم، همچنین نمی دانیم آیا نمرات SAT آن را می سنجد یا نه.

۱۲-۷ $z = 11/62$. ما فرض صفر را رد خواهیم کرد.

۱۲-۸ اندازه نمونه ها خیلی متفاوت است.

۱۲-۹ $t = 2/18$ با $df = 5700$. ما نمی توانیم H_0 را رد کنیم.

۱۲-۱۰ $H_1: \mu \neq 500$; $H_0: \mu = 500$

۱۲-۱۱ نه، زیرا ما می خواهیم هر گاه $\mu \neq 500$ باشد H_0 را رد کنیم. ما نمی توانیم منتظر بمانیم تا داده را ببینیم و سپس تصمیم بگیریم از کدام دامنه استفاده کنیم.

۱۲-۱۲ در این مورد این یک یافته مهم نیست زیرا تفاوت خیلی کوچک است. حتی اگر $\mu = 503$ باشد به جای ۵۰۰، تفاوت خاصی بوجود نمی آید- دشوار است این را به عنوان یک نشانه تغییر در پیشرفت امتحان کارشناسی ارشد آمریکا (GRE) بدانیم، بخصوص اینکه این بر اساس یک نمونه انتخابی بوده است.

۱۲-۱۳ $t = -1/47$. ما نمی توانیم H_0 را رد کنیم.

۱۲-۱۴ میزان متوسطی از تغییرات وجود داشت و N نسبتاً کوچک بود.

۱۲-۱۵ $512/3 \leq \mu \leq 537/7$

۱۲-۱۶ $500/3 \leq \mu \leq 505/7$

۱۲-۱۷ در تمرین ۱۲-۱۵ ما σ را می دانستیم اما s را نمی دانستیم، و برای بدست آوردن حدود اطمینان می توانستیم از انحراف معیار جمعیت استفاده کنیم. در تمرین ۱۲-۱۶ ما s را می دانستیم هر چند چیزی در باره σ گفته نشده بود، و برای حل حدود اطمینان از s استفاده کردیم. در هر دو مورد مقدار بحرانی t و z $+1/96$ بود، اما فقط بخاطر اینکه تعداد df برای t این قدر زیاد (۵۷۰۰) بود.

۱۲-۱۸ $\bar{X} = 101/82$ ، $s = 12/68$ ، $t = 0/82$ با $df = 32$. ما H_0 را رد نخواهیم کرد.

۱۲-۱۹ شما σ را نمی دانستید.

۱۲-۲۰ نخست ما نیاز به یک جدول اعداد تصادفی با یک واریانس معلوم داریم. سپس ما نمونه های زیاد ۵ نمره ای بیرون خواهیم کشید. برای هر نمونه ما واریانس نمونه را حساب می کنیم. وقتی چندین هزار واریانس نمونه بدست آوردیم توزیع فراوانی آنها را رسم می کنیم.

$$t = (0.62\sqrt{8}) / (1 - 0.62^2) = 2.85 \quad 12-21$$

$$t = (0.35\sqrt{8}) / (1 - 0.35^2) = 1.13$$

پاسخ تمرین‌های فصل ۱۳

۱۳-۱ $t = 2.23$ با $df = 14$. H_0 رد می‌شود. نتیجه می‌گیریم که راهنمایی فیزیکی میزان کمک را کاهش داده است.

۱۳-۲ شرایط راهنمایی فیزیکی همیشه در مرحله دوم آمده، و پیشرفت ممکن است فقط گویای گذشت زمان یا اثرات تأخیری تقلید باشد.

۱۳-۳ می‌توان برای نیمی از آزمودنی‌ها ترتیب عملکردها را وارونه کرد.

$$0.068 \leq \mu \leq 3.666 \quad 13-4$$

۱۳-۵ $t = 0.45$. ما نمی‌توانیم H_0 را رد کنیم.

۱۳-۶ $t, H_0 = -0.35$ را نمی‌توان رد کرد.

۱۳-۷ داده‌ها در تمرین ۱۳-۵ نشان می‌دهد که این برنامه احتمالاً موفق نبوده است. از سوی دیگر داده‌های تمرین ۱۳-۶ نشان می‌دهد که برای برخی از افراد سیگاری ممکن است موفقیت‌آمیز بوده اما باعث شود برخی دیگر حتی بیشتر سیگار بکشند. این دو اثر در ترکیب داده‌ها خیلی هم را خنثی می‌کنند.

۱۳-۹ $t = 1.39$ با $df = 19$. H_0 رد نمی‌شود.

۱۳-۱۰ برای پاسخ نیازمند این هستیم که مقدار بحرانی t را بدانیم، که آن هم نیاز به دانستن df ، و این به نوبت نیاز به دانستن N دارد. اما ما نمی‌توانیم در شکل تقریب کلی $2/0.0$ را به عنوان مقدار بحرانی استفاده کنیم. آنگاه N مورد نیاز $= [(2/0.0 \cdot 2/870) / 0.333]^2 = 297$.

۱۳-۱۲ همچنان که همبستگی بین دو متغیر افزایش می‌یابد، t هم کاهش می‌یابد.

۱۳-۱۴ اگر از آزمودنی‌ها خواسته شود یک مجموعه آنارگرام (واژه‌سازی با جایجایی حروف واژه) را در شرایطی انجام دهند و یک مجموعه دیگر را در شرایط دیگر، هر راهکاری که آنها در نیمه نخست آموخته‌اند می‌تواند به نیمه دوم منتقل شود و بر نتایج آن تأثیر بگذارد.

پاسخ تمرین‌های فصل ۱۴

۱۴-۱ $t = 1.44$ با $df = 28$. H_0 رد نمی‌شود.

۱۴-۲ در هر گروه میزان قابل توجهی واریانس وجود دارد.

۱۴-۳ با اندازه‌گیری همین موضوع تحت هر دو شرایط (مانند فصل ۱۳) ما می‌توانستیم تغییرپذیری آزمودنی به آزمودنی را حذف کنیم.

- ۱۴-۴ $t = 1/60$ با $df = 16$. H_0 رد نمی شود.
- ۱۴-۵ ما از تخصیص تصادفی استفاده می کنیم تا از این امکان گمارش نامناسب افرادی که دارای کرم خوردگی دندان زیادی هستند جلوگیری کنیم چنان که اگر اجازه دهیم افرادی که با هم دوست هستند در یک گروه قرار گیرند این اتفاق می افتد.
- ۱۴-۶ $t = 4/54$ با $df = 16$. H_0 رد می شود.
- ۱۴-۷ $t = 1/52$ با $df = 5$. H_0 رد نمی شود.
- ۱۴-۸ نرخ افت می تواند خیلی مهم باشد. تنها نیمی از افراد متعدد توانایی تکمیل برنامه A را همانند تکمیل برنامه B دارند.
- ۱۴-۹ $t = 0/59$ با $df = 15$. H_0 رد نمی شود.
- ۱۴-۱۰ $-3/03 \leq (\mu_1 - \mu_2) \leq 5/34$
- ۱۴-۱۱ $-3/61 \leq (\mu_1 - \mu_2) \leq 14/11$
- ۱۴-۱۲ $t = -2/36$ با $df = 8$. H_0 رد می شود.
- ۱۴-۱۳ $t = 2/13$ با $df = 8$. H_0 رد نمی شود.
- ۱۴-۱۴ اگر اندازه ی نمونه ها برابر باشند، برآوردهای جمع شده و جمع نشده یکسان خواهند بود.
- ۱۴-۱۵ اگر دو واریانس برابر باشند، برآوردهای جمع شده (یا جمع نشده) با مقدار مشترک واریانس جداگانه نمونه برابر خواهد بود.
- ۱۵-۱۶ مشروعیت کامل و تغییر شکل منطقی داده ها می تواند نتایج متفاوتی به بار آورد. پیش از شروع یک آزمایش لازم است به ماهیت متغیر وابسته توجه جدی نمود.

پاسخ تمرین های فصل ۱۵

- ۱۵-۱ (الف) $0/250$ (ب) $2/50$ (ج) $0/71$
- ۱۵-۳ $N = 98, 125, 169$
- ۱۵-۴ توان $= 0/965$
- ۱۵-۶ (الف) $16 \cong 15/21$ (ب) $32 \cong 31/36$ N
- ۱۵-۷ (الف) $39 \cong 38/72$ N آزمودنی در گروه (ب) $84 \cong 84/5$ N آزمودنی در گروه
- ۱۵-۸ توان $= 0/30$ ۱۵-۹ توان $= 0/51$
- ۱۵-۱۰ (الف) توان $= 0/22$ (ب) $t = 1/19$ (ج) t از لحاظ عددی با δ برابر است گرچه t از آماره های حساب می شود ولی δ از پارامترها حساب می شود.
- ۱۵-۱۱ تحقیق اول. زیرا او به یک تفاوت معنادار با یک آزمایشی که توان نسبتا پایینی داشت دست یافت، او بایستی یک اثر تقریبا بزرگ را بررسی کرده باشد.

۱۴-۱۵

اندازه اثر	γ	t یک دامنه	t دو دامنه
کوچک	۰/۲۰	۲۸۹	۱۱۵۶
متوسط	۰/۵۰	۴۷	۱۸۶
بزرگ	۰/۸۰	۱۹	۷۲

۱۵-۱۵

اندازه اثر	γ	t یک دامنه	t دو دامنه
کوچک	۰/۲۰	۱۲۱	۴۸۴
متوسط	۰/۵۰	۲۰	۷۸
بزرگ	۰/۸۰	۸	۳۲

۱۵-۱۶ چنانچه پیش فرض‌های زیربنایی آزمون برآورده شوند نمی‌تواند چنین حالتی وجود داشته باشد.

۱۵-۱۷ وقتی $1.0965 \pm \mu = \mu_1$ که توان برابر است با احتمال خطای نوع II.

پاسخ تمرین‌های فصل ۱۶

۱۶-۱ (الف)

منبع	df	SS	MS	F
گروه‌ها	۳	۶۵۵/۱۴۳	۲۱۸/۳۸۱	۱۰/۷۸*
خطا	۲۴	۴۸۶/۲۸۶	۲۰/۲۶۲	
کل	۲۷	۱۱۴۱/۴۲۹		

* $p < 0.05$

(ب)

منبع	df	SS	MS	F
گروه‌ها	۱	۵۱/۵۷۲	۵۱/۵۷۲	۱/۲۳ ns
خطا	۲۶	۱۰۸۹/۸۵۷	۴۱/۹۱۸	
کل	۲۷	۱۱۴۱/۴۲۹		

۱۶-۵

منبع	df	SS	MS	F
گروه‌ها	۲	۱۵۱۶/۱۰۰	۷۳۸/۰۵۰	۲۴/۰۱*
خطا	۱۲	۳۷۸/۸۳۳	۰/۰۶۹	
کل	۱۴	۱۸۹۴/۹۳۳		

* $p < ۰/۰۵$

۱۶-۶ (الف)

منبع	df	SS	MS	F
گروه‌ها	۱	۱/۲۶۰	۱/۲۶۰	۰/۸۷ ns
خطا	۱۰	۱۴/۵۳۲	۱/۴۵۳	
کل	۱۱	۱۵/۷۹۲		

$$(ب) \quad t = -۰/۸۸۳ = \sqrt{۰/۷۸۰} \quad \text{(غیر جمع شده)}$$

$$(ج) \quad t = -۰/۹۳۱ = \sqrt{۰/۸۶۷} \quad \text{(جمع شده)}$$

$$(د) \quad t \text{ جمع شده}$$

$$۱۶-۷ \quad \omega^2 = ۰/۰۸۷ \quad \eta^2 = ۰/۱۹۶$$

۱۶-۸ (الف)

منبع	df	SS	MS	F
گروه‌ها	۲	۸۲۶/۸۶۷	۴۱۳/۴۳۳	۹/۶۴*
خطا	۲۷	۱۱۵۷/۶۶۲	۴۲/۸۷۷	
کل	۲۹	۱۹۸۴/۵۲۹		

* $p < ۰/۰۵$

(ب) این به مقایسه خوانندگان ضعیف و قوی نمی‌پردازد- و حتی ما اصلا نمی‌دانیم که وضع خواندن آزمودنی‌ها چگونه است.

(ج) آیا برجسته بودن حروف بر سرعت خواندن افراد تأثیر می‌گذارد.

۱۶-۹

منبع	df	SS	MS	F
گروه‌ها	۲	۴۱۳/۴۳۳	۲۰۶/۷۱۷	۴/۸۲*
خطا	۱۲	۵۱۴/۵۱۲	۴۲/۸۷۶	
کل	۱۴	۹۲۷/۹۴۵		

* $p < ۰/۰۵$

هنوز بین گروه‌ها تفاوت معناداری وجود دارد. توجه داشته باشید که بیشتر مقادیر از جمله F نصف شده‌اند اما خطای MS چنین نیست.

۱۰-۱۶ ما اعتماد بیشتری داریم زیرا یک نتیجه معنادار است که با یک آزمایش دارای توان کمتر بدست آمده است.

۱۲-۱۶

منبع	df	SS	MS	F
گروه‌ها	۲	۰/۸۵۴	۰/۴۲۷	۸/۰۶*
خطا	۴۹	۲/۵۹۷	۰/۰۵۳	
کل	۵۱	۳/۴۵۱		

* $p < ۰/۰۵$

پاسخ تمرین‌های فصل ۱۷

۱۷-۱

منبع	df	SS	MS	F
قیمت	۱	۳/۸۲۸	۳/۸۲۸	۱۰/۴۳*
جنسیت	۱	۰/۰۷۸	۰/۰۷۸	< ۱
$P \times S$	۱	۳/۴۰۳	۳/۴۰۳	۹/۲۷*
خطا	۱۶	۵/۸۷۵	۰/۳۶۷	
کل	۱۹	۱۳/۱۸۴		

* $P < ۰/۰۵$

۱۷-۲

منبع	df	SS	MS	F
رفتار مادرانه	۱	۲۸/۹	۲۸/۹	۶/۰۸*
وزن	۱	۱۴/۴	۱۴/۴	۳/۰۳ ns
$P \times W$	۱	۰/۱	۰/۱	< ۱
خطا	۳۶	۱۷۱/۰	۴/۷۵	
کل	۳۹	۲۱۴/۴		

* $P < ۰/۰۵$

۱۷-۳ خیر، زیرا ۵۰٪ از جمعیت مادران فرزندان اول نوزادان کم وزن نداشتند.

۱۷-۴ $F = ۱/۷۸$ غیر معنادار.

۱۷-۵

منبع	df	SS	MS	F
ناحیه	۲	۳۵۶/۰۴۴	۱۷۸/۰۲۲	۶/۰۷*
تأخیر	۲	۱۸۸/۵۷۸	۹۴/۲۸۹	۳/۲۲ ns
S×D	۴	۳۷۱/۹۵۶	۹۲/۹۸۹	۳/۱۷*
خطا	۳۶	۱۰۵/۲۰۰	۲۹/۳۱۱	
کل	۴۴	۱۹۷۱/۷۷۸		

* $P < ۰/۰۵$

۱۷-۷ $t = ۳/۰۵$ با $df = ۳۶$ ؛ $t = ۳/۰۰$ با $df = ۳۶$ ؛ هر دو معنادار هستند.

۱۷-۸ برای ناحیه خنثی در مقایسه با A، $F = t^2 = ۳/۰۵^2 = ۹/۳۰$ مقدار بحرانی $F_{\alpha}(k-1) = ۱$ ،
(k)

$k(n-1)F_{\alpha} - (۲, ۳۶) = ۲(۳/۲۹) = ۶/۵۸$ فرض صفر رد می شود. برای ناحیه خنثی در مقابل

B.

$F = t^2 = ۳/۲۰۰ = ۹$ بازهم فرض صفر رد می شود.

۱۷-۹ $F = ۴/۳۵$ با H_0 رد می شود.

۱۷-۱۰

منبع	df	SS	MS	F
مارک	۱	۵۶۷/۰۰۰	۵۶۷/۰۰۰	۲۷/۹۸*
موقعیت	۱	۵۱/۵۷۱	۵۱/۵۷۱	۲/۵۵ ns
B×L	۱	۳۶/۵۷۱	۳۶/۵۷۱	۱/۸۸ ns
خطا	۲۴	۴۸۶/۲۸۶	۲۰/۲۶۰	
کل	۲۷	۱۱۴۱/۴۲۹		

* $P < ۰/۰۵$

یک اثر معنادار برای مارک وجود دارد اما برای موقعیت یا تعامل آنها وجود ندارد.

۱۷-۱۱ اثر موقعیت در تحلیل یک طرفه و اثر موقعیت در تحلیل دو طرفه دارای df، SS و MS یکسان هستند. اما F فرق دارد چون وقتی گروه ها را در گروه های بزرگتر ادغام می کنیم عبارت خطا را متورم

کرده ایم. (توجه: تحلیل یک طرفه روی چهار گروه و ۲×۲ همان عبارت خطا را ایجاد می کند).

۱۷-۱۲ $\eta^2_{سن} = ۰/۰۹$ ؛ $\omega^2_{سن} = ۰/۰۸$ ؛ $\eta^2 = ۰/۵۷$ ؛ $\omega^2 = ۰/۵۵$ ؛ $\omega^2 = ۰/۰۷$ ؛ $\eta^2_{A \times C} = ۰/۰۶$ ؛ $\omega^2_{A \times C} = ۰/۰۵$

۱۷-۱۶ $\eta^2 = ۰/۲۹$ ؛ $\eta^2 = ۰/۰۱$ ؛ $\eta^2 = ۰/۲۶$ ؛ $\omega^2 = ۰/۲۶$ ؛ $\omega^2 = ۰$ ؛ $\omega^2 = ۰/۲۲$ ؛ $\omega^2 = ۰$

$$\eta^2 = 0.13 \quad \omega^2 = 0.07 \quad \eta^2_{\text{وزن}} = 0.00 \quad \eta^2_{P \times W} = 0.11 \quad \omega^2_{\text{مادرانه}} = 0.04 \quad \omega^2_{\text{وزن}} = 0.00$$

$$\omega^2_{P \times W} =$$

۱۷-۲۰ وقتی عبارت خطا نسبت به مجموع مجزورات برای اثر عملکرد خیلی کوچک باشد این دو آماره خیلی نزدیک به هم خواهند بود.

۱۷-۲۱ ایزنک فکر می‌کرد در این تکالیف که شامل پردازش‌های سطح بالا بود آزمودنی‌های مسن‌تر نسبت به کوچک‌ترها فرق خواهند داشت. او انتظار نداشت تفاوت‌های تکالیف به پردازش‌های جزئی نیاز داشته باشند.

پاسخ تمرین‌های فصل ۱۸

۱۸-۱ (الف)

منبع	df	SS	MS	F
آزمودنی‌ها	۷	۱۸۹۶۶۶/۶۷		
جلسات	۲	۱۸۰۸/۳۳	۹۰۴/۱۶۵	۳/۶۶
خطا	۱۴	۳۴۵۸/۳۳	۲۴۷/۰۲۴	
کل	۲۳	۱۹۴۹۳۳/۳۳		

(ب) تفاوت معناداری بین جمع‌کل‌های جلسات وجود ندارد.

۱۸-۲ (ب) $t = 1/14$

منبع	df	SS	MS	F
آزمودنی‌ها	۷	۱۳۰۷۹۳/۷۵۰		
جلسات	۱	۵۰۶/۲۵۰	۵۰۶/۲۵۰	۱/۲۹ns
خطا	۷	۲۷۴۳/۷۵۰	۳۹۱/۹۸۰	
کل	۱۵	۱۳۴۰۴۳/۷۵۰		

۱۸-۳

منبع	df	SS	MS	F
آزمودنی‌ها	۹	۶۰/۰۰۰		
زمان	۲	۲۷/۴۶۷	۱۳/۷۳۳	۵/۰۲*
خطا	۱۸	۴۹/۲۰۰	۲/۷۳۳	
کل	۲۹	۱۳۶/۶۶۷		

* $P < 0.05$

۱۸-۴ خط پایه را با آموزش مقایسه می کنیم. $t. H_0 = -2/98$ رد می شود. با آموزش عملکرد بهبود می یابد. خط پایه را با پیگیری مقایسه می کنیم. $t. H_0 = -0/54$ رد نمی شود. بعد از دوره پیگیری، عملکرد بطور معناداری بهتر از آن در دوره خط پایه نیست.

۱۸-۶

منبع	df	SS	MS	F
ایالت ها	۹	۴۶۷۳۷/۴۴۵		
سال ها	۶	۱۹۸۷/۲۰۳	۳۳۱/۲۰۰	۱۹/۷۰*
خطا	۵۴	۹۰۷/۶۵۳	۱۶/۸۰۸	
کل	۶۹	۴۹۶۳۲/۳۰۱		

*P < ۰/۰۵

۱۸-۷ ما این فرض صفر را آزمایش می کنیم که نمرات هر سال نمونه هایی از جمعیت هایی با میانگین برابر هستند.

۱۸-۸ چون هر حالت در هر سال اندازه گیری شده - به این معنی است که ما بین ستون ها همبستگی های غیر صفر را انتظار داریم.

۱۸-۹ (الف) $t. H_0 = 0/34$ رد نمی شود. عملکرد در طی خط پایه تغییر نمی کند. (ب) $t. H_0 = 0/99$ رد نمی شود. در طی سه هفته آموزش عملکرد فرقی نکرده است. (ظاهرا از قبل طی مرحله آموزشی تغییر کرده و بعد از آن ثابت مانده است).

$$18-10 \quad F = 0/1536 \quad \sqrt{F} = 0/3919 = t.$$

پاسخ تمرین های فصل ۱۹

۱۹-۱ $x^2 = 11/313$ با $df = 2$. H_0 رد می شود و نتیجه می گیریم دانشجویان به صورت تصادفی ثبت نام نکرده اند.

۱۹-۲ ما نمی توانیم بگوییم چنانچه دانشجویان کلاس های متفاوتی انتخاب کنند بخاطر استاد است یا بخاطر زمان برگزاری کلاس است - استاد و زمان درآمیخته با هم هستند. دست کم ما باید کلاس ها را همزمان برگزار کنیم.

۱۹-۳ $x^2 = 2/4$ با $df = 4$. ما نمی توانیم H_0 رد کنیم که رفتار مرتب سازی فرزند ما با نظریه مان در راستای هم هستند.

۱۹-۴ تنها ما می توانیم این را به جمعیتی از داده ها تعمیم دهیم که با فرزندمان قابل گسترش باشد. به عبارت دیگر ما فقط یک نمونه از رفتار او را داریم. ما نمونه ای تصادفی از رفتار افراد در کل نداریم.

۱۹-۵ $x^2 = ۲۹/۳۵$ با $df = ۱$. ما می‌توانیم این H_0 را رد کنیم که کودکان عروسک‌ها را بطور تصادفی انتخاب می‌کنند (دست‌کم از لحاظ رنگ).

۱۹-۶ $x^2 = ۱۲/۲۴$ با $df = ۱$. ما باز هم می‌توانیم H_0 رد کنیم، اما این بار در جهت خلاف است.

۱۹-۷ $x^2 = ۳۴/۱۷$ با $df = ۱$. H_0 رد می‌شود و نتیجه می‌گیریم که توزیع انتخاب‌های بین عروسک‌های سیاه و سفید در دو تحقیق فرق دارد. انتخاب مستقل از تحقیق نیست. ما دیگر از این نمی‌پرسیم که آیا یک رنگ عروسک به رنگ دیگر ترجیح داده می‌شود، بلکه می‌پرسیم آیا الگوی ترجیح در طی این تحقیقات ثابت است. به زبان تحلیل واریانس ما با یک تعامل سروکار داریم.

۱۹-۸ (الف) این فرض صفر آزموده می‌شود که مرکز سلامت روان که در آن شخصی به جستجوی کمک می‌پردازد مستقل از نوع مسأله اوست.

(ب) $x^2 = ۱۰/۳۰۵$ با $df = ۴$. (ج) H_0 رد می‌شود و نتیجه می‌گیریم که دو متغیر مستقل نیستند.

۱۹-۹ $x^2 = ۵/۱۵۳$ که نصف آنچیزی است که در تمرین ۸-۱۹ بدست آمد. (ب) اندازه‌ی نمونه نقش مهمی دارد، با نمونه‌های بزرگتر احتمال بیشتری وجود دارد که نتایج معناداری ایجاد شود- چنانکه در مورد آزمون‌های دیگر چنین است.

۱۹-۱۰ $x^2 = ۵/۳۸$ با $df = ۱$. (ج) H_0 رد می‌شود و نتیجه می‌گیریم که سطح پیشرفت طی دوره دبیرستان تحت تأثیر عملکرد در دوره دبستان فرق می‌کند.

۱۹-۱۱ (الف) $x^2 = ۱۶/۴۳$ با $df = ۷$. (ب) H_0 رد می‌شود. (ج) چون نزدیک به نیمی از خانه‌ها دارای فراوانی کمتر از ۵ هستند، احساس عدم راحتی داریم. یک راه ترکیب ستون‌های مجاور هم است.

۱۹-۱۲ ما خواهیم پرسید آیا دانشجویان بطور مساوی بین مقوله‌های مختلف توزیع شده‌اند. چیزی که در تمرین ۱۱-۱۹ ما می‌آزماییم در واقع این است که آیا توزیع به ظاهر، برای آنهایی که بعداً کلاس تقویتی انگلیسی گرفته‌اند مانند کسانی است که بعداً کلاس تقویتی غیر انگلیسی گرفته‌اند.

۱۹-۱۳ (الف) $x^2 = ۱۲/۸۹۶$ با $df = ۶$. (ب) H_0 رد می‌شود. تعداد روزهای لازم برای تحویل تابعی از فاصله است، اما نه به شکلی روشن و خیلی قابل تفسیر.

۱۹-۱۴ به منظور اینکه از درآمیختگی استادان و زمان‌ها پرهیز شود.

۱۹-۱۵ $H_0: x^2 = ۰/۶۵۸$ رد نمی‌شود.

۱۹-۱۶ با افزایش تعداد نمونه با همین درصدهای خانه‌ای، توان افزایش می‌یابد.

۱۹-۱۷ $H_0: x^2 = ۱/۳۱$ رد نمی‌شود.

۱۹-۱۸ هر چند تفاوت معناداری بین درصدهای خانه‌ای وجود ندارد، این واقعیت که ۴۵٪ دانشجویان احساس می‌کنند دوره باید بهبود پیدا کند یک نتیجه مهم است.

پاسخ تمرین های فصل ۲۰

۲۰-۱ (الف) $W_s = ۲۳$ ؛ $W_{.۰۲۵} = ۲۷$ (ب) H_0 رد می شود و نتیجه می گیریم که کودکان بزرگتر از استنتاج های بیشتتری در خلاصه کردن استفاده می کنند.

۲۰-۲ (الف) $W_s = ۵۳$ ؛ $W_{.۰۲۵} = ۶۸$ (ب) H_0 رد می شود و نتیجه می گیریم که برای آزمودنی ها در گروه آسیب آمیگدال زمان بیشتتری طول می کشد تا تکلیف را یاد بگیرند، چنان که نظریه پیش بینی کرده بود.

۲۰-۳ $z = ۳/۱۵$ H_0 رد می شود.

۲۰-۴ $W_s = ۵۳$ ؛ $W_{.۰۵} = ۷۲$ دوباره H_0 رد می شود.

۲۰-۵ (الف) $T = ۸/۵$ ؛ $T_{.۰۲۵} = ۸$ H_0 رد نمی شود. (ب) ما نمی توانیم نتیجه بگیریم شواهدی داریم که از این فرضیه حمایت می کند که افزایش پایداری در تعمیم فرضیه و آزمایش در طی زمان وجود دارد. (در اینجا موردی وجود دارد که در آن روش های جایگزین برای شکستن روابط می تواند باعث نتیجه گیری های متفاوتی شود.)

۲۰-۶ (الف) $H_0 = -۱/۹۴$ z رد نمی شود. (ب) ما به همان نتیجه می رسیم. جواب ها بخوبی با هم هماهنگ هستند و حتی اگر N بزرگتر بود جواب ها به هم نزدیک تر هم بودند.

۲۰-۷ (الف) $T = ۴۶$ ؛ $T_{.۰۲۵} = ۵۲$ (ب) H_0 رد می شود و نتیجه می گیریم که کودکان فرزند اول بیشتر مستقل هستند.

۲۰-۸ $z = -۲/۲۰$ که با نتیجه گیری قبلی ما هماهنگ است.

۲۰-۹ تفاوت بین جفت ها خیلی وابسته به نمره فرزند اول دارد.

۲۰-۱۰ آزمون من-ویتنی این فرض صفر را می آزماید که نمره ها از جمعیت های یکسانی برآمده است و بویژه نسبت به تفاوت میانه ها حساس است. آزمون t این یک فرض صفر را در باره میانگین ها می آزماید و نرمال بودن و واریانس های یکسان را در جمعیت فرض می گیرد.

۲۰-۱۱ آزمون جفت های همتای ویل کاکسون برای رتبه های علامت دار این فرض صفر را می آزماید که نمرات همتا از جمعیت های یکسان یا جمعیت های متفاوت با میانگین (و میانه) یکسان بیرون کشیده شده اند. آزمون t نظیر آن این فرض صفر را می آزماید نمرات همتا از جمعیت هایی با میانگین یکسان و با پیش فرض نرمال بودن استخراج شده اند.

۲۰-۱۲ ماهیت این مقیاس برای تفسیر نتایج مهم است اما نه برای انتخاب آزمون آماری روی اعداد واقعی.

۲۰-۱۳ چون با در نظر گرفتن پیش فرض های مربوط به نرمال بودن و تجانس واریانس، آزمون t به ویژه به میانگین های جمعیت اشاره دارد.

۲۰-۱۴ $H_0: H = 5/124$ رد نمی‌شود.

۲۰-۱۵ $H_0: H = 6/757$ رد می‌شود.

۲۰-۱۶ $H_0: x_F^2 = 8/792$ رد می‌شود - میزان گریز افزایش یافته است.

۲۰-۱۷ تأثیر تفاوت‌های فردی را از بین می‌برد (تفاوت سطح کلی گریز از یک شخص به شخص دیگر).

۲۰-۱۸ ما قادر نخواهیم بود اندازه‌های نسبی تفاوت‌ها را به حساب بیاوریم.

۲۰-۱۹ در این مورد اینها آزمون‌های دقیقاً معادل هم هستند.

۲۰-۲۰ $x_F^2 = 9/00$ ما می‌توانیم فرض صفر را رد کنیم و نتیجه بگیریم که مردم چای درست شده با چای کیسه‌ای را دوست ندارند.

پاسخ تمرین‌های فصل ۲۱

۲۱-۱ این شامل آماره‌های ساده توصیفی می‌شود، احتمالاً نمودارهای جعبه‌ای نمرات توانایی خواندن پرسش‌های هر آزمون را شامل می‌شود.

۲۱-۲ آنها می‌توانند از r پیرسون برای همبستگی نمرات آزمون دینور و نمرات اندازه‌های هوشی اداره فردی استفاده کنند. ممکن است سؤال این باشد که برای برخی کودکان یک نوع مقیاس هوش و برای برخی دیگر شکل متفاوتی وجود دارد. در این مورد می‌توانید کودکان را در گروه‌ها رتبه‌بندی کنید و همبستگی‌ها را برای هر گروه حساب کنید.

۲۱-۳ آنها می‌توانند یک آزمون t برای دو گروه مستقل اجرا کنند تا دو گروه را مقایسه کنند. سپس می‌توانند همبستگی بین درصد نمره درست و سطح گزارش شده با استفاده از کمک آموزشی را بدست آورند. (به خاطر دارید که برای دو گروه فرقی ندارند که از یک آزمون t استفاده کنید یا از تحلیل واریانس یک طرفه).

۲۱-۴ این یک تحلیل واریانس 3×2 است (نوع سرنخ در زمان یادآوری $\times$ نوع سرنخ در زمان یادگیری).
۲۱-۵ این یک تحلیل واریانس با سه گروه است. آنها می‌توانند در صورت لزوم از آزمون‌های t هم برای مقایسه گروه‌های منفرد استفاده کنند.

۲۱-۶ او باید سه آزمون t جداگانه را برای دو گروه مستقل اجرا کند.

۲۱-۷ این یک تحلیل واریانس اندازه‌های مکرر پیچیده‌تر از چیزی است که در فصل ۱۸ بیان شد. یک طرح عاملی 2×2 با خواننده‌های خوب در برابر ضعیف و متن‌های دشوار در برابر آسان به عنوان عامل‌ها است. خواننده یک متغیر بین آزمودنی است (افراد مختلف در دو گروه هستند) و متن‌ها یک متغیر درون آزمودنی است (افراد یکسان هر دو نوع متن را می‌خوانند).

۲۱-۸ آنها بایستی دو گروه را در نمرات مرکز مهار مقایسه کنند، و می توانند برای آزمون لغات پی بادی از آزمون t برای گروه های مستقل استفاده کنند. آنها می توانند با استفاده از r پیرسون، همبستگی آزمون مهار نوکیس تیک لند و آزمون پی بادی را هم بدست آورند.

۲۱-۹ آنها می توانند یک تحلیل واریانس یک طرفه روی نمرات MFFT در بار دوم اجرا کنند. همچنین باید آن را روی نمرات بار اول هم اجرا کنند تا فرضیه آزمایشی را که گروه ها با هم شروع کرده اند بررسی کنند. (یک طرح پیچیده تر نیز مناسب خواهد بود، [به تمرین ۷-۲۱ رجوع کنید]، اما دو تحلیل واریانس یک طرفه را ما توصیه می کنیم به خاطر اینکه تفسیر آن راحت است.)

۲۱-۱۰ این یک تحلیل واریانس ساده اندازه گیری مکرر است. آزمون توزیع-آزاد نظیر آن آزمون رتبه فریدمن خواهد بود.

۲۱-۱۱ (الف) او باید یک طرح 2×2 (گروه در پرهیز در مقابل مصرف سیگار) در هر فاصله زمانی اجرا کند. (ب) ما نمی دانیم آیا نیکوتین درون آدامسی اثری دارد یا نه. ممکن بود با در نظر گرفتن هر نوع آدامسی برای جویدن یک عامل کنترلی داشت.

۲۱-۱۲ او نخست می تواند میانگین رتبه اختصاص یافته به هر ویژگی را (برای هر جنس و سال) پیدا کند. از آنجا که داده های خام در اصل رتبه ها بودند، توصیه می شود سپس مقدار میانگین ها را رتبه بندی کند. آنگاه او می تواند r پیرسون را بین زن و مرد برای هر سال یا بین سال ها در هر جنس حساب کند. همبستگی ها برای ده جفت نمره (یک جفت در ویژگی).

۲۱-۱۳ او باید یک آزمون t برای دو نمونه همبسته به کار ببرد. اگر نخواهند یک آزمون پارامتریک به کار ببرند، باید آزمون جفت های همتای ویل کاکسون برای رتبه های علامت دار را به کار بگیرند.

۲۱-۱۴ آنها باید آزمون من-ویتنی را به کار ببرند.

۲۱-۱۵ باید یک آزمون t برای دو نمونه همبسته به کار ببرند - نمونه ها به این دلیل همبسته هستند که اسمیت و پلنت جفت های همتا تشکیل دادند.

۲۱-۱۶ آنها باید با یک تحلیل واریانس یک طرفه شروع کنند. با فرض معنادار بودن F آنها می توانند این کار را با یک آزمون t محافظت شده دنبال کنند، دو شرایط "دزدی" را مقایسه کنند. همچنین می توانند همبستگی درستی را با اعتماد بوسیله r پیرسون بررسی کنند.

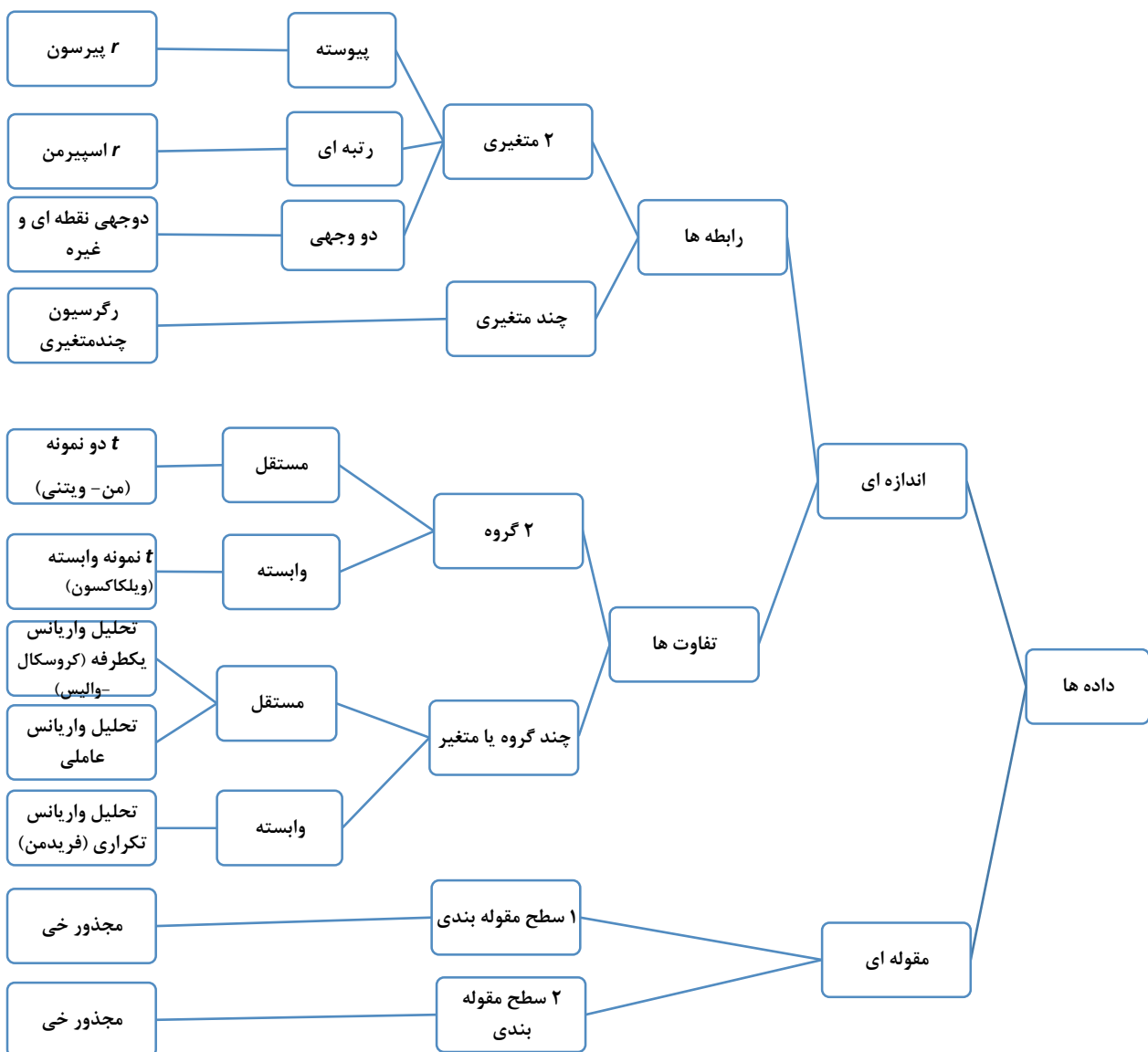
۲۱-۱۷ این شرایطی است برای یک آزمون نیکویی برازش مجذور خی.

۲۱-۱۸ این یک تحلیل پیچیده دیگر اندازه گیری مکرر است. مقایسه یادآوری دو فهرست (یکی قبل از دادن دارو و دیگری بعد از آن یاد گرفته شده است) یک اندازه گیری مکرر است زیرا آزمودنی های یکی هستند. مقایسه گروه های دارو در برابر غیردارو یک اثر بین آزمودنی ها است چون گروه ها شامل آزمودنی های متفاوتی هستند.

۲۱-۱۹ این یک آزمون t برای گروه های مستقل است.

۲۰-۲۱ آنها می‌توانند آزمون کروسکال-والیس را برای مقایسه سه گروه به کار ببرند و چنانچه معنادار بود از آزمون من-ویتنی استفاده کنند که در غیر این صورت یک آزمون t محافظت شده به کار می‌رود.

نوع داده ها سوال تعداد گروه ها یا متغیرها مقیاس وابستگی آزمون



آزمون های توزیع آزاد در پراکنش آمده اند.

فرمول‌های اصلی آمار

توان

$\gamma = (\mu_1 - \mu_o) / \sigma$	اندازه اثر (یک نمونه)
$\gamma = (\mu_1 - \mu_2) / \sigma$	اندازه اثر (دو نمونه)
$\gamma = \rho_1 - \rho_o$	اندازه اثر (همبستگی)
$\delta = \gamma \sqrt{N}$	دل‌تا (t تک نمونه‌ای)
$\delta = \gamma \sqrt{\frac{N}{2}}$	دل‌تا (t دو نمونه‌ای)
$\delta = \gamma \sqrt{N-1}$	دل‌تا (همبستگی)

همبستگی و رگرسیون

$SS_X = \sum X^2 - \frac{(\sum X)^2}{N}$	مجموع مجذورات
$SP_{XY} = \sum XY - \frac{\sum X \sum Y}{N}$	مجموع حاصل ضرب‌ها

$$cov_{XY} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N-1} = \frac{SP_{XY}}{N-1}$$

کوواریانس

$$r = \frac{cov_{XY}}{s_X s_Y} = \frac{SP_{XY}}{\sqrt{SS_X SS_Y}}$$

$$= \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

ضریب همبستگی (پیرسون)

$$b = \frac{cov_{XY}}{s_X^2} = \frac{SP_{XY}}{SS_X}$$

شیب

$$a = \frac{\sum Y - b \sum X}{N} = \bar{Y} - b \bar{X}$$

ثابت

$$S_{Y-\hat{Y}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{N-2}} = \sqrt{\frac{SS_{\text{خطا}}}{N-2}}$$

$$= S_Y \sqrt{(1-r^2) \frac{N-1}{N-2}}$$

خطای معیار برآورد

$$\sum Y^2 - \frac{(\sum Y)^2}{N}$$

SS_Y

$$\sum \hat{Y}^2 - \frac{(\sum \hat{Y})^2}{N}$$

$SS_{\hat{Y}}$

$$SS_Y - SS_{\hat{Y}} = SS_{\text{خطا}}$$

$S_{Y-\hat{Y}}$

$$SS_Y (1 - r^2)$$

 $SS_{\text{خطا}}$

تحلیل واریانس

$$\sum X^2 - \frac{(\sum X)^2}{N}$$

 $SS_{\text{کل}}$

$$\frac{\sum T_g^2}{n} - \frac{(\sum X)^2}{N}$$

 $SS_{\text{یک طرفه}} \text{ (گروهی)}$

$$SS_{\text{کل}} - SS_{\text{گروهی}}$$

 $SS_{\text{خطا}} \text{ (یک طرفه)}$

$$\sum T_{Ri}^2 - \frac{(\sum X)^2}{N}$$

 $SS_{\text{دو طرفه}} \text{ (ردیفی)}$

$$\frac{\sum T_{Ci}^2}{nc} - \frac{(\sum X)^2}{N}$$

 $SS_{\text{دو طرفه}} \text{ (ستونی)}$

$$\frac{\sum T_{\text{خانه}}^2}{nr} - \frac{(\sum X)^2}{N}$$

 $SS_{\text{دو طرفه}} \text{ (خانه ها)}$

$$\frac{\sum T_{\text{خانه}}^2}{n} - \frac{(\sum X)^2}{N}$$

 $SS_{R \times C} \text{ (دو طرفه)}$

$$SS_{\text{کل}} - SS_{\text{ردیفی}} - SS_{\text{ستونی}}$$

 $SS_{\text{خطا}} \text{ (دو طرفه)}$

$$SS_{\text{خانه ها}} - SS_{\text{کل}} - SS_{R \times C} - SS_{\text{ستونی}} - SS_{\text{ردیفی}} - SS_{\text{کل}}$$

t محافظت شده (فقط در صورتی که F معنادار باشد بکار می رود)

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{\text{خطا}}}{n_i} + \frac{MS_{\text{خطا}}}{n_j}}}$$

$$\eta^2 = \frac{SS_{\text{گروهی}}}{SS_{\text{کل}}}$$

مجذور اتا

$$\omega^2 = \frac{SS_{\text{گروهی}} - (k-1)MS_{\text{خطا}}}{SS_{\text{کل}} + MS_{\text{خطا}}}$$

مجذور امگا (یک طرفه)

مجذور خی

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

مجذور خی

$$\chi^2 = \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)}$$

مجذور خی فقط برای جدول 2×2

آماره های توزیع-آزاد

میانگین و انحراف معیار (نمونه بزرگ)
برای آماره من-ویتنی

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

میانگین و انحراف معیار (نمونه بزرگ)

$$\chi^2 = \frac{N(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)}$$

برای آماره ویل کاکسون

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

آماره H کروسکال-ویلز

$$\chi_F^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3N(k+1)$$

آماره مجذور خی فریدمن

درباره کتاب:

کتاب های آماری فراوانی در حوزه روان شناسی و علوم اجتماعی در ایران منتشر شده است، اما دلیل اینکه مترجم برگردان این اثر به فارسی را برگزیده است عمدتاً دو چیز بوده است: یکی اینکه معمولاً در بیشتر کتاب ها مبحث استنباط آماری مورد غفلت قرار گرفته و منطبق زیربنایی آن توضیح داده نشده است، در نتیجه مجموعه ای از شیوه های آماری به یادگیرنده ارائه می شود بدون اینکه اساس و استدلال استنباط آماری توضیح داده شود؛ کتاب ها اول به این موضوع توجه لازم را ننموده و فهم آن را آسان ننموده است. برتری دیگر این کتاب سادگی و قابل فهم بودن مطالب و سعی نویسنده در خنثی کردن این برداشت عمومی است که آمار درس سختی است که نیاز به مهارت در ریاضیات دارد. در این کتاب مطالب به گونه ای روان و با مثال های مناسب بیان شده است و حتی مطالب اولیه ریاضی در حاشیه شرح داده شده است. از این رو با وجود اینکه از چاپ اولیه این کتاب مدتی گذشته است اما شیوایی و جامعیت مطالب آن برای سطح دانشجویان کارشناسی و علاقه مندان به یادگیری مبانی آمار در علوم رفتاری، آن را در نقطه برتر قرار داده است.

ISBN: 978-600-459-218-5



9 786004 592185

REDMI NOTE 8T
AI QUAD CAMERA



