ریاضیات، آمار، و تدریس•

جورج کاب، دیوید مور*
ترجمهٔ مجتبی گنجعلی، محمدقاسم وحیدی‌اصل

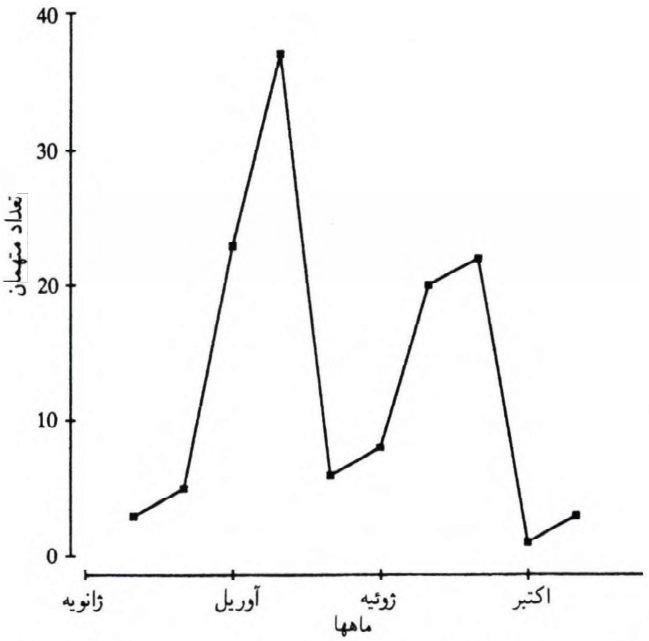
تفکر آماری چه تفاوتی با تفکر ریاضی دارد؟ نقش ریاضیات در آمار چیست؟ اگر از آمار محتوای ریاضی آن را بزدایید، چه جوهرهٔ عقلانی از آن باقی می‌ماند؟

در آنچه در زیر می‌آید، پاسخهایی به این پرسشها ارائه کرده آنها را به دنباله‌ای از مثالها ربط می‌دهیم که چشم‌اندازی از شیوهٔ عمل رایج آماری به‌دست می‌دهند. ضمن این کار، و مخصوصاً در پایان مقاله، به برخی از نتایج بحث برای تدریس آمار اشاره می‌کنیم.

۱. مقدمه: نظری اجمالی به تفکر آماری. آمار شاخه‌ای روش‌شناختی از دانش است. وجود آن برای خودش نیست بلکه برای این است که مجموعه‌ای منسجم از ایده‌ها و ابزارها برای کار با داده‌ها به شاخه‌های علمی دیگر ارائه کند. نیاز به چنین شاخه‌ای از دانش ناشی از «همه جا حاضر بودن تغییرپذیری»[1] است. افراد با هم تفاوت دارند. نتایج اندازه‌گیریهای مکرر روی فردی واحد، یکسان نیستند. در برخی موقعیتها، می‌خواهیم موارد غیرعادی را در توده‌ای انبوه از داده‌ها پیدا کنیم. در موقعیتهایی دیگر، تمرکز روی تغییر نتایج اندازه‌گیریهاست، و گاه می‌خواهیم اثرات نظام‌مند را در برابر نوفهٔ زمینه‌ای مربوط به تغییرات فردی آشکار سازیم. آمار ابزارهایی برای کار با داده‌ها به‌دست می‌دهد که همه جا حاضربودن تغییرپذیری در آنها ملحوظ شده است.

۱.۱ نقش مضمون. تمرکز بر روی تغییرپذیری، به‌طور طبیعی محتوای خاصی به آمار می‌دهد که آن‌را از خود ریاضیات و دیگر علوم ریاضی مجزا می‌سازد، اما چیزی بیش از محتوای صِرف است که تفکر آماری را از تفکر ریاضی متمایز می‌کند. آمار نیازمند نوع متفاوتی از تفکر است، زیرا داده‌ها فقط اعداد نیستند، آنها اعدادی همراه با مضمون هستند،

1. omnipresence of variability



شکل۱ تعداد افرادی که در ناحیهٔ اسکس[1] ماساچوست متهم به جادوگری شده‌اند.

مثال ۱. داز اندوور[2]. وقتی نمودار دنبالهٔ متناهی

(۳, ۵, ۲۳, ۳۷, ۶, ۸, ۲۵, ۲۲, ۱, ۳)

رسم شود (شکل ۱) الگوی مشخصی را نشان می‌دهد، اما این اعداد و الگو معنی و اهمیتی ندارند مگر آنکه مضمون آنها را بدانیم. در حقیقت این اعداد تعداد کل ماهانه افرادی هستند که از فوریهٔ ۱۶۹۲ تا چند ماه بعد در

1. Essex 2. Andover

ناحیهٔ اسکس ایالت ماساچوست، رسماً متهم به جادوگری شده‌اند. نمودار شکل ۱ دو موج از اتهام‌زدن‌ها را نشان می‌دهد که با یک نقطهٔ حضیض در تابستان ۱۶۹۲ از هم جدا شده‌اند. الگوی شکل ۱ معنی‌دارتر می‌شود وقتی‌که بدانیم دار زدن اولین جادوگر محکوم شده (بریجت بیشاپ[1]) در ۱۰ ژوئن ۱۶۹۲ صورت گرفته است. تصور اثر هشیارکنندهٔ اولین اعدام در جامعهٔ کوچک دهکدهٔ سیلم[2] (حالا دنورز[3]) سخت نیست. اما دومین موج اتهامات چرا؟ چنین معلوم می‌شود که اتهام‌زدن‌های موج اول علیه ساکنان دهکدهٔ سیلم، شهر سیلم، و هر شش شهر مجاور بجز یکی از آنها صورت گرفته است؛ در موج دوم، اکثریت اتهام‌زدن‌ها علیه ساکنان یکی دیگر از شهرهای مجاور، یعنی اندوور صورت گرفته است. منابع ما [۳ و ۴] توضیح زیادی در مورد آنچه در شهر اندوور اتفاق افتاده نمی‌دهند، ولی این الگو همراه با آنچه از مضمون می‌دانیم حداقل بخشی از ماجرا را حکایت و پرسشهای جالبی را مطرح می‌کند.

هر چند این مثالِ نخست تقریباً هیچ محتوای ریاضی ندارد، ملاحظهٔ تأثیر متقابل بین الگو و مضمون آن، نمونهٔ بخش تفسیری تفکر آماری است. برای مثالی آشناتر از نوعی بسیار متفاوت، آزمون برابری میانگینهای دو توزیع نرمال را در نظر بگیرید.

مثال ۲ الف. مدلی برای مقایسهٔ میانگینهای نرمال. مدل استانداردی متضمن دو مجموعه از متغیرهای تصادفی مستقل همتوزیع (iid) را در نظر بگیرید:

$$X_1, X_2, \cdots X_n \text{ iid } N(\mu_1, \sigma_1) \quad Y_1, Y_2, \cdots, Y_m \text{ iid } N(\mu_2, \sigma_2)$$

نتیجه می‌شود که $\overline{x} = \frac{\sum x_i}{n}$ و $s_1^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}$ آماره‌های بسنده برای μ_1 و σ_1^2 هستند. نتیجه‌ای مشابه برای Yها برقرار است. به‌طور غیررسمی، آماره‌ای برای یک پارامتر، بسنده است اگر از تمام اطلاعات در بارهٔ آن پارامتر که در نمونه موجود است، استفاده کند. به‌طور رسمیتر، توزیع شرطی داده‌ها، به شرط آمارهٔ بسنده به پارامتر وابسته نیست. قضیهٔ رائو-بلک‌ول[4] تضمین می‌کند که هیچ برآوردگر نااریبی نمی‌تواند واریانسی کوچکتر از برآوردگر نااریب مبتنی بر آمارهٔ بسنده داشته باشد. هر دوی $\overline{x}$ و s_1^2 نااریب‌اند: $E(\overline{x}) = \mu_1$ و $E(s_1^2) = \sigma_1^2$. و بالاخره توزیع توأم آنها معلوم است: میانگین نمونه‌ای $\overline{x}$ دارای توزیع نرمال با واریانس $\frac{\sigma_1^2}{n}$ است، و مستقل از آن، $\frac{(n-1)s_1^2}{\sigma_1^2}$ دارای توزیع خی دو با $(n-1)$ درجهٔ آزادی است. حال فرض کنید که می‌خواهیم فرض $H_0 : \mu_1 = \mu_2$ را آزمون کنیم. اگر $\sigma_1^2 = \sigma_2^2$ آنگاه برآوردگری بسنده و نااریب برای واریانس مشترک از طریق ادغام کردن به‌دست می‌آید:

$$s_p^2 = [(n-1)s_1^2 + (m-1)s_2^2]/(n+m-2)$$

اگر H_0 درست باشد، آنگاه $\frac{(\overline{x}-\overline{y})}{s\sqrt{\frac{1}{n}+\frac{1}{m}}}$ دارای توزیع tی استیودنت با $n+m-2$ درجهٔ آزادی است، و می‌توانیم از مقدار t که از روی داده‌ها محاسبه شده برای آزمون فرض صفر استفاده کنیم. اگر t از ۰ به‌اندازهٔ کافی دور باشد، نتیجه می‌گیریم که $\mu_2 \neq \mu_1$.

1. Bridget Bishop 2. Salem Village 3. Danvers
4. Rao-Blackwell

این مثال از دو لحاظ تفاوت بارزی با مثال اول دارد: محتوای ریاضی و نقش مضمون. مثال ۱ که اساساً هیچ محتوای ریاضی ندارد، جوهرهٔ عقلانی‌اش تقریباً به‌طور کامل از تأثیر متقابل الگو و ماجرا ناشی می‌شود. مثال ۲، که اساساً هیچ محتوای غیرریاضی ندارد، جوهرهٔ عقلانی‌اش را بدون رجوع صریح به مضمون کاربردی به‌دست می‌آورد.

هر چند ریاضیدانان اغلب بر مضمون کاربردی، هم برای ایجاد انگیزه و هم به‌عنوان منبعی از مسائل تحقیقاتی، تکیه دارند، کانون توجه نهایی در تفکر ریاضی، الگوهای مجرد است: مضمون بخشی از جزئیات نامربوطی است که باید روی شعلهٔ تجرید گداخته شوند تا بلور ساختار نابی که قبلاً ناپیدا بوده آشکار شود. در ریاضیات، مضمون، ساختار را در پردهٔ ابهام نمی‌افکند. تحلیلگران داده‌ها نیز مانند ریاضیدانان در پی یافتن الگوها هستند، اما در تحلیل داده‌ها در نهایت، اینکه الگوها دارای معنی و ارزشمند باشند، بستگی به این دارد که چگونه تار این الگوها با پود ماجرا به‌هم بافته شوند. در تحلیل داده‌ها، مضمون است که معنا می‌بخشد.

این تفاوت، پیامدهای ژرفی در تدریس دارد. برای تدریسِ خوب آمار، فهمیدن نظریهٔ ریاضی کافی نیست؛ حتی کافی نیست که علاوه بر آن، قسمت غیرریاضی آمار را فهمیده باشیم. لازم است مانند معلم ادبیات، ذخیره‌ای آماده از مثالهای واقعی داشته باشیم، و بدانیم چگونه از آنها استفاده کنیم تا دانشجو قدرت داوری نقادانه پیدا کند. در ریاضیات، که در آن مضمون کاربردی از اهمیت بسیار کمتری برخوردار است، مثالهای فی‌البداهه به‌خوبی نیاز را رفع می‌کند و معلمان ریاضی در ابداع فی‌المجلسِ مثالها مهارت به‌دست می‌آورند (به تابعی برای نمایش قاعده زنجیری نیاز دارید؟ مسأله‌ای نیست: فوراً مثالی بسازید). ولی در آمار مثالهای فی‌البداهه به درد نمی‌خورند، زیرا تأثیر متقابل الگو و مضمون را به‌طور مؤثق نشان نمی‌دهند. همان‌طور که برتراند راسل ریاضیات را به‌علت حالت تجریدی خشک و بی‌روحش به مجسمه‌سازی تشبیه می‌کند، شاید بتوان تحلیل‌داده‌ها را که در آن الگو و مضمون جداشدنی نیستند به فن شعر تشبیه کرد. تصور کنید که درسی مقدماتی در عروض تدریس می‌کنید و می‌خواهید مصرع شش رکنی داکتیلی[1] را معرفی کنید. برای این‌کار کافی نیست بگویید «TA ta ta، TA ta ta، TA ta ta ... » بلکه لازم است شاگردان شما داکتیل[2]ها را در یک شعر واقعی بشنوند [۲۰]:

This is the forest primeval. The murmuring pines and the hemlocks.

[این جنگل روزگاران آغازین است. کاجها و صنوبران در نجوا][3]. همین‌طور معلم آمار لازم است دنیای داده‌ها را بشناسد. اگر، برای مثال، وقتی نمودارهای

1. dactylic hexameter

۲. dactyl، (در شعر) کلمه‌ای دارای سه هجا که دو هجای آن کوتاه و یک هجای آن بلند باشد.

۳. مصرع آغازین شعر حماسی Evangeline از هنری وودزورث (Wadsworth).

برای خواننده شاید بهتر باشد مثالی از اوزان شعر فارسی بیاوریم:

وقتی می‌خواهید «بحر متقارب مثمن سالم» را در درس عروض به شاگردان معرفی کنید، کافی نیست بگویید «فعولن فعولن فعولن فعولن» بلکه شاگردان شما باید این وزن را در یک شعر واقعی احساس کنند:

درخت تو گر بار دانش بگیرد — به زیر آوری چرخ نیلوفری را

توزیع داده‌ها را تدریس می‌کنید از داده‌های مربوط به مدت زمان بین دو فورانِ اولد فیت‌فول[1] [۳۰] و طول مدت سلطنت پادشاهان و ملکه‌های انگلیس [۱۳] استفاده کنید، دانشجویان شما می‌توانند چیزی بیش از خود روشها فرا گیرند. شکل دو مدی زمانهای بین دو فوران، دو نوع فوران را به ذهن متبادر می‌کند و توزیع سلطنت پادشاهان چاولگی به سمت مقادیر بالا را که در زمانهای انتظار معمول است، نشان می‌دهد.

نقشهای متفاوت مضمون در ریاضیات و آمار، مخصوصاً به‌صورتی که در دو مثال افراطی نخستین نشان داده شد، ممکن است مؤید استنتاج نادرستی به‌نظر آید که در این گفتهٔ پولاک دیده می‌شود: «آماردانان زیادی اکنون مدعی‌اند که آمار چیزی کاملاً جدا از ریاضیات است، به‌طوری که درسهای آماری نیازی به هیچ نوع آمادگی در ریاضیات ندارند.» هر چند اقناع و اثبات در آمار از نوع ریاضی نیست ([۲۲] و [۲۴] را ببینید)، همهٔ درسهای آماری نیازمند قدری، و برخی از آنها نیازمند مقدار زیادی آمادگی در ریاضیات‌اند. نظریه‌های ریاضی بغرنج، داربست بخشهایی از آمارند، و مطالعهٔ این نظریه‌ها بخشی از تعلیماتی است که آماردانان می‌بینند. اما، گرچه آمار بدون ریاضیات نمی‌تواند شکوفا شود، عکس آن نادرست است. اینکه آماربخشی ضروری از تعلیمات ریاضیدانان نیست، به‌طور تلویحی از این گفتهٔ احتمال‌دان برجسته، دیوید آلدوس [۱] برمی‌آید که وی «به کاربردهای احتمال در همهٔ حوزه‌های علمی بجز آمار علاقه‌مند است».

بنابراین، به نقش ریاضیات در علم آمار چیست؟ پاسخ باید با نگاهی نظام‌مندتر به منطق تحلیل داده‌ها آغاز شود.

۲.۱ مرور کلی نموداری بر تحلیل آماری. درسی به سبک قدیم که پایبند به عرضهٔ کاربردها باشد، شاید مثال دوم را با یک تمرین کاربردی به پایان برساند. داده‌های زیر از [۲۵] گرفته شده‌اند هر چند خود تمرینْ ساختگی است؛ کل بررسی در [۲۱] توصیف شده است.

مثال ۲ ب، کلسیم و فشار خون. آیا افزایش مقدار کلسیم در رژیم غذایی ما، فشار خون را کاهش می‌دهد؟ اعداد زیر مقدار کاهش فشار خون سیستولیک را بعد از ۱۲ هفته برای ۲۱ فرد انسانی در اختیار ما قرار می‌دهد. ۱۰ فرد در گروه ۱ به‌مدت ۱۲ هفته یک مکمل کلسیم را مصرف کردند؛ ۱۱ فرد در گروه ۲ از یک دارونما[2] استفاده کردند. این فرض را که کلسیم تأثیری روی فشار خون نداشته است آزمون کنید.

گروه ۱ (کلسیم): ۷، ۴ − ، ۱۸، ۱۷، ۳ − ، ۵ − ، ۱، ۱۰، ۱۱، ۲ −

گروه ۲ (دارونما): ۱ − ، ۱۲، ۱ − ، ۳ − ، ۳، ۵ − ، ۵، ۲، ۱۱ − ، ۱ − ، ۳ −

این تمرین، اگر آن‌را به اختصار توصیف کنیم، کاریکاتوری بیش نیست، کاریکاتوری که این دیدگاه خطا را موجب می‌شود که به‌محض اینکه استنتاجات ریاضی از یک مدل تکمیل شدند، کاربردها اساساً چیزی نیستند جز محاسباتی

۱. Old Faithful، آبفشانی در پارک طبیعی هوستون آمریکا، در سال ۱۹۷۴ میلادی، ۳۰۳۴ زمان بین دو فوران آن اندازه‌گیری و معلوم شد که میانگین مدت زمان بین دو فوران ۶۶٫۵ دقیقه، کوتاهترین زمان ۳۶، و طولانیترین آن ۱۰۲ دقیقه بوده است. از ۵۰۰۰۰ فوران مشاهده و ثبت‌شده در ۱۰۵ سال گذشته، میانگین زمان بین فورانها همیشه بین ۶۰ و ۶۷ دقیقه، کمترین زمان ۳۳ دقیقه، و بیشترین آن ۱۰۲ دقیقه بوده است.م.

2. placebo

الگوها $\xrightarrow{(۲)}$ داده‌ها $\xrightarrow{(۱)}$ طرح

تعبیر $\xrightarrow{(۵)}$ نتایج $\xrightarrow{(۴)}$ روشها $\xrightarrow{(۳)}$ مدلها

شکل۲ نمایش نموداری مراحل تولید و تحلیل داده‌ها

معمولی. برای چشم‌اندازی واقعیتر، شکل ۲ را که نموداری از مراحل تحلیل آماری است، در نظر گیرید. قبل از بررسی جزئیات این طرح خام توجه به دو نکته ضروری است.

۱. این طرح خلاصه‌وار که پیشروی اکید از چپ به راست را القا می‌کند، بیش از اندازه ساده‌گرایانه است. در عالم واقع، فرایند تحلیل داده‌ها نه خطی و نه یکسویه است. درگذارهای مختلف بین مراحل، گاه دو مرحلهٔ متوالی و گاه حتی بیش از دو مرحله، بر یکدیگر تأثیر متقابل دارند. مثلاً هر چند انتخاب طرح برای تولید داده‌ها، ساختار داده‌های حاصل را تعیین می‌کند، اما داشتن اطلاعاتی براساس داده‌هایی که از قبل در دست است، می‌تواند به شکل‌گیری طرح کمک کند، همان‌گونه که اطلاع از اندازهٔ تغییرات از یک مورد به مورد دیگر به تصمیم‌گیری در بارهٔ اینکه چند مورد آزمایش لازم است کمک می‌کند. همچنین داده‌ها ممکن است مدلی را القا کنند، اما این مدل به روشهایی منجر شود که ما را دوباره به سراغ داده‌ها بفرستند تا تخلفات احتمالی از مفروضات مدل را امتحان کنیم. شاید، همان‌طور که خواهیم دید، مرحلهٔ نهاییِ تفسیر نتایج که از همه مهمتر است به‌طور حیاتی به مرحلهٔ اول که نوع طرح مورد استفاده در تولید داده‌هاست وابسته باشد.

۲. ارائهٔ این ترتیب سردستی و محدود از مراحل برای القای این نظر نیست که مباحثی که در یک درس آمار مقدماتی تدریس می‌شوند باید همین ترتیب را دنبال کنند. به دلایلی که بعداً می‌آوریم، توصیهٔ ما آغاز کردن چنین درسی با روشهای کاوش و توصیف داده‌ها، سپس بازگشت به تولید داده‌ها و از آنجا رفتن به‌سوی استنباط رسمی است.

با فرض رعایت شدن این هشدارها، این روند نما می‌تواند چارچوب مفیدی برای بازبینی نقش ریاضیات در آمار باشد و نیز عناصر جوهرهٔ غیرریاضی آمار را به اختصار بنمایاند. در این زمینه به چهار نکتهٔ واضح اشاره می‌کنیم:

۱. طرح، کاوش، و تفسیر، اجزای مرکزی تفکر آماری هستند. هر سهٔ این اجزا به‌شدت به مضمون وابسته‌اند، ولی در سطح مقدماتی متضمن ریاضیات بسیار کمی هستند. نظریهٔ (عمدتاً غیرریاضی) طرح آزمایشها چندین دهه عمر دارد و بسیار رشد کرده است. نظریهٔ کاوش جدیدتر است، و در حال حاضر هنوز جنبهٔ ابتدایی دارد، گرچه ابزارهای مبتنی بر کامپیوترِ کاوش پیشرفته‌اند؛ نظریهٔ تفسیر را در بهترین حالت می‌توان نامنسجم توصیف کرد.

۲. درس کلاسیک آمار ریاضی چنان به‌خوبی با گذار (۳) از مدلها به روشها متناظر است که «از مدلها به روشها» شاید کم‌وبیش بتواند عنوان یک درس باشد. مضمون در اینجا چندان ذی‌ربط نیست زیرا مدلها مانند مدل مثال ۲ الف به‌طور مجرد مطرح می‌شوند و معمولاً در نتیجه‌گیریها

از یکی از اصول بهینه‌سازی (کمترین مربعات، ماکسیمم درستنمایی) برای استنتاج رسمی روش استفاده می‌شود.

۳. گذار (۴) از روشها به نتایج، کانون توجه درسهای سبک قدیمی است که مانند کتابهای آشپزی تنظیم می‌شوند و در آنها هر روش در مجموعه‌ای از فرمولها خلاصه می‌شود. مضمون در اینجا نیز بی‌ربط است، از این نظر که می‌توانید الگوریتمهای محاسباتی را فرا گیرید، و درواقع آنها را بهتر بیاموزید، اگر در هیچ‌گونه دغدغهٔ خاطر در مورد اینکه این روشها به چه درد می‌خورند، به خود راه ندهید. با این حال، در برخی درسها تلاش شده است این قرص بزرگ گلوگیر را با لعابی نازک از مضمون بدلی آسانتر آمادهٔ بلع کنند. خوشبختانه، کامپیوتر این‌گونه درسها را در زباله‌دانی تاریخ برنامه‌های درسی می‌افکند.

۴. شاید طنزآمیز باشد که گذارهای (۳) و (۴) که اغلب مرکز توجه درسهای مقدماتی بوده‌اند، دقیقاً دو موردی هستند که از نظر ذهنی سرراست‌تر و مکانیکی‌تر از بقیه‌اند (با توجه به درک محدود فعلی و نظریهٔ کمتر پیشرفتهٔ گذارهای دیگر) و اغلب کمترین امکان را برای داوری و خلاقیت عرضه می‌کنند.

برای بسط این نکات با تفصیل بیشتر، به مثال کلسیم و فشار خون برمی‌گردیم. در آنچه خواهد آمد، مراحل شکل (۲) را زیر سه عنوان کلیتر ترکیب می‌کنیم: تولید داده‌ها، تحلیل داده‌ها، و استنباط.

۲. محتوای آمار

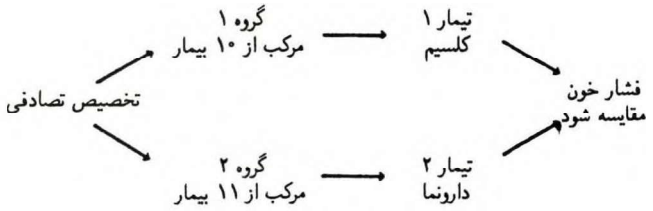
۱.۲ تولید داده‌ها. مدل استاندارد در مثال ۲ الف نقصی بسیار جدی دارد؛ این مدل تمایزی بین داده‌های مشاهده‌ای (مثلاً داده‌های یک بررسی نمونه‌ای) و داده‌های یک آزمایش مقایسه‌ای تصادفی‌شده قائل نمی‌شود. این تمایز بین مشاهده و آزمایش، یکی از مهمترین تمایزات در آمار است. پژوهشگران اغلب می‌خواهند به نتایج علّی برسند، مثلاً اینکه کلسیم علت کاهش در فشار خون است. آزمایشها اغلب اجازهٔ نتیجه‌گیریهای علّی را می‌دهند، در حالی‌که مطالعات مشاهده‌ای تقریباً همواره مسائل علیّت را حل‌وفصل نشده و در معرض مناقشه باقی می‌گذارند. با این‌حال مدلهای ریاضی نظریهٔ آمار برای داده‌های مشاهده‌ای و آزمایشی یکسان‌اند.

مطالعهٔ کلسیم درواقع یک آزمایش بود:

مثال ۲ پ، طرح بررسی کلسیم [۲۱]. بازبینی نمونه‌ای بزرگ از اشخاص رابطه‌ای بین مصرف کلسیم و فشار خون را آشکار کرد. این رابطه برای مردان سیاهپوست قویتر از دیگران بود. به همین علت پژوهشگران آزمایشی را به اجرا درآوردند.

آزمودنیهای آزمایش، ۲۱ مرد سیاهپوست سالم بودند. یک گروه ۱۰ نفره از این مردان که به‌طور تصادفی انتخاب شده بودند یک مکمل کلسیم مصرف کردند. گروه شاهد مرکب از ۱۱ مرد، یک قرص دارونما را که شبیه مکمل کلسیم بود، دریافت کردند. آزمایش از دو طرف کور[1] بود.

۱. double-blind. در این آزمایش نه آزمودنیها و نه پزشکان همکار طرح خبر ندارند که چه کسی دارو و چه کسی دارونما مصرف کرده است.ـم.



شکل۳ ساده‌ترین آزمایش مقایسه‌ای تصادفی‌شده

آیا می‌توانیم نتیجه بگیریم که کلسیم علت کاهش در فشار خون است؟ چنین استنباطی، که یک اختلاف مشاهده شده را می‌توان به‌معنای ظاهریش پذیرفت، بر سه پایه استوار است که دو تا از آنها بر تولید داده‌ها متکی هستند:

(۱) استدلالی ــ که تنها برای نمونه‌های تصادفی و آزمایشهای تصادفی‌شده جنبهٔ عادی و سرراست دارد ــ مبتنی بر اینکه مدلی احتمالاتی در مورد داده‌ها قابلِ اِعمال است.

(۲) استدلالی ــ مبتنی بر احتمال و نسبتاً سرراست ــ مبنی بر اینکه اختلاف مشاهده شده «واقعی» است، یعنی بزرگتر از آن است که انتساب آن به تغییرات شانسی موجه باشد.

(۳) استدلالی ــ که بجز در حالت آزمایشهای تصادفی‌شده اغلب پرزحمت، و امکان لغزش در آن زیاد است ــ مبنی بر اینکه اختلاف مشاهده شده ناشی از تأثیر یک عامل اختلاطی مجزا از عاملِ موردِنظر نیست.

آزمون t در مثال ۲ الف مانند همهٔ آزمونهای آماری و بازه‌های اطمینان آماری تنها به استدلال دوم می‌پردازد: «اگر فرض کنیم که مدل شانسی خاصی قابل به‌کارگیری است، چقدر محتمل است که اختلاف قابل مشاهده‌ای به این بزرگی به‌دست آوریم؟» دو استدلال دیگر به طرح وابسته هستند.

آزمایش بالینی در بارهٔ تأثیر کلسیم بر فشار خون یک آزمایش مقایسه‌ای تصادفی‌شده بود. شکل ۳ این طرح را برحسب نکات عمدهٔ آن ارائه می‌کند. حُسن عمدهٔ تخصیص تصادفی آزمودنیها، آن است که استدلالهای (۱) و (۳) را به‌صورت سرراست و مکانیکی درمی‌آورد. بنابراین مسألهٔ استنتاج علت را به امتحان برازنده بودن یک مدل، و سپس در صورت مناسب بودن برازش، به انجام محاسباتی سرراست تقلیل می‌دهد. تخصیص تصادفی آزمودنیها اریبی را در تشکیل گروههای تیمار و دارونما از بین می‌برد و گروههایی را به‌وجود می‌آورد که تنها از طریق تغییرات شانسی، پیش از اعمال تیمارها، با هم تفاوت دارند. این طرح مقایسه‌ای به یاد ما می‌آورد که با همهٔ آزمودنیها دقیقاً رفتار یکسانی بجز از نظر محتوای قرصهایی که مصرف می‌کنند به عمل می‌آید. بنابراین اگر اختلافهایی در میانگین کاهش فشار خون، بزرگتر از آنچه می‌توان انتظار داشت که ناشی از شانس باشد، مشاهده کنیم، می‌توانیم اطمینان داشته باشیم که کلسیم بوده که تأثیری را که مشاهده می‌کنیم باعث شده است.

از ابزارهای عمدهٔ دیگر تولید داده‌ها، بررسی نمونه است که در آن، نمونه‌ای را برای تولید اطلاعاتی در بارهٔ جامعه‌ای بزرگتر انتخاب و بازبینی می‌کنند. مثالهای جالب در این زمینه فراوان‌اند: نظرسنجیهای سالم و ناسالم، گردایهٔ داده‌های اجتماعی و اقتصادی دولت، منابع داده‌های دانشگاهی همچون مرکز پژوهش آرای ملی[1] در دانشگاه شیکاگو از این جمله‌اند. طرحهای

1. National Opinion Research Center

آماری برای نمونه‌گیری با اصرار بر اینکه شانس غیرشخصی باید نمونه را انتخاب کند، شروع می‌شود. ایدهٔ اصلی طرحهای آماری برای تولید داده‌ها، خواه از طریق نمونه‌گیری یا انجام آزمایش، استفادهٔ عمدی از شانس است. استفادهٔ صریح از سازوکارهای شانسی برخی از منابع اصلی اریبی را از بین می‌برد. این‌کار همچنین تضمین می‌کند که مدلهای احتمالی ساده، فرایند تولید داده‌های ما را توصیف می‌کنند، و بنابراین روشهای استنباط استاندارد قابل به‌کارگیری‌اند. با این حال، همان‌طور که مثال زیر نشان می‌دهد، برخلاف آزمایشهای تصادفی‌شده، مطالعات مشاهده‌ای به‌صورتی سرراست تن به استنباط علیّت نمی‌دهند. شرح این بررسی نخست توسط بِشت[1] و واکر[2] به‌عنوان یک مثال در [۱۲] آمده است؛ شرحی که ما از آن می‌آوریم به تأسی از [۲۶] است.

مثال ۳. سیگارکشیدن و تندرستی. در یکی از نخستین مطالعات مشاهده‌ای دربارهٔ سیگارکشیدن و تندرستی، نرخ مرگ‌ومیر برای سه گروه از مردان مقایسه شده است. این نرخها برحسب مرگ‌ومیر در هر سال برای هر ۱۰۰۰ مرد عبارت بودند از:

غیرسیگاریها ۲٫۰۲۰، سیگاریها ۵٫۰۲۰،
مصرف‌کنندگان سیگار برگ و پیپ ۵٫۰۳۵

برای آزمودن اینکه اختلافهای مشاهده شده ناشی از شانس هستند یا نه، می‌توانیم از مدلی مشابه با مدل مثال ۲ الف استفاده کنیم. اندازه‌های نمونه آن‌قدر بزرگ‌اند که می‌توانیم به‌آسانی تغییرات شانسی را به‌عنوان توضیحی برای اختلافهای مشاهده شده رد کنیم و به این نتیجه‌گیری ظاهری برسیم که سیگار کشیدن با مخاطرهٔ کمی همراه است، اما پیپ یا سیگار برگ یا هر دو کاملاً خطرناک‌اند. درواقع، این نتیجه‌گیری معتبر می‌بود هرگاه این داده‌ها از یک آزمایش تصادفی‌شدهٔ کنترل‌شدهٔ دو طرف کور، مانند مطالعهٔ کلسیم به‌دست آمده بودند. ولی این فرض در اینجا مطرح نیست. از آنجا که این بررسی مشاهده‌ای است، باید در مورد عاملهای دیگری که مرتبط با عادت سیگارکشیدن هستند و ممکن است علتِ این اختلافِ مشاهده شده باشند پرس‌وجو کنیم. در اینجا، سن یکی از چنین عوامل اصلی است: مصرف‌کنندگان سیگار برگ و پیپ معمولاً پیرتر از سیگاریها هستند و خطر مرگ با سن افزایش می‌یابد. در این مطالعه، متوسط سن برای سه گروه عبارت بوده‌اند از:

غیرسیگاریها ۵۴٫۹ سال، سیگاریها ۵۰٫۵ سال،
مصرف‌کنندگان سیگار برگ و پیپ ۶۵٫۹ سال

تنها بعد از تصحیح نرخهای مرگ با توجه به اختلاف سن است که اعدادی مطابق با آنچه انتظار داریم، به‌دست می‌آوریم:

غیرسیگاریها ۲۰٫۳، سیگاریها ۲۸٫۳،
مصرف‌کنندگان سیگار برگ و پیپ ۲۱٫۲

به نظر ما دو مثال اخیر تواماً متضمن دو درس مهم برای ریاضیدانانی است که آمار تدریس می‌کنند: نخست اینکه نتایج حاصل از یک مطالعه به‌طور قطع به چگونگی تولید داده‌ها وابسته است و دوم اینکه در مدلهای ریاضی استاندارد، تولید داده‌ها نادیده گرفته می‌شود.

ارائهٔ ایده‌های آماری برای تولید داده‌ها به‌منظور پاسخگویی به پرسشهای مشخص، از مؤثرترین خدمات آمار به معرفت بشری است. تولید داده‌ها با طرح بد، شایعترین نقیصهٔ جدّی در مطالعات آماری است. تولیدِ خوش‌طرح داده‌ها اجازه می‌دهد که روشهای استاندارد تحلیل را به‌کار برده و به نتایج روشنی برسیم. آماردانان حرفه‌ای به خاطر تخصصشان در طرح مطالعه دستمزد دریافت می‌کنند؛ اگر مطالعه خوب طرح‌ریزی شود (و هیچ حادثهٔ بد غیرمنتظره‌ای رخ ندهد)، برای تحلیل نیازی به افراد حرفه‌ای نیست. به عبارت دیگر، طرح تولید داده‌ها واقعاً مهم است. اگر صرفاً بگویید «فرض کنید X_1 تا X_n مشاهده‌هایی مستقل و همتوزیع هستند»، شما آمار تدریس نمی‌کنید.

۲.۲ تحلیل داده‌ها: کاوش و توصیف. تحلیل داده‌ها شکل نوین «آمار توصیفی» است که با ابزارهای توصیفی استادانه‌تر و بسیار بیشتر و به‌ویژه با نظریه‌ای عمدتاً متعلق به جان توکی[1] از متخصصان آزمایشگاه بل و پرینستن تجهیز شده است. این نظریه اینک به نام تحلیل کاوشی داده‌ها یا EDA[2] مشهور است. هدف EDA، نظیر کاوشگری که وارد سرزمینهای ناشناخته می‌شود، آن است که ببیند داده‌های در دست چه می‌گویند. ما این مسأله را که آیا این داده‌ها نمایندهٔ جهانی بزرگترند موقتاً (اما نه برای همیشه) کنار می‌گذاریم.

جدول ۱ تمایزات بین EDA و استنباط استاندارد را به اختصار و در حد مقدماتی [۲۵] ارائه می‌کند:

جدول ۱ مقایسهٔ تحلیل کاوشی داده‌ها با استنباط رسمی مبتنی بر احتمال.

تحلیل کاوشی داده‌ها	استنباط آماری
هدفْ کاوش نامحدود داده‌ها و جستجو برای یافتن الگوهای جالب است	هدفْ پاسخ‌دادن به پرسشهای مشخصی است که قبل از تولید داده‌ها عنوان شده‌اند.
نتایج تنها برای افراد و شرایطی که در مورد آنها داده‌هایی در دست داریم، صادق‌اند.	نتایج در مورد گروه بزرگتری از افراد یا رده‌ای گسترده‌تر از شرایط صادق‌اند.
نتایجْ غیررسمی و مبتنی بر آنچه در داده‌ها می‌بینیم، هستند.	نتایج به پشتوانهٔ حکمی که حاکی از اطمینان ما به آنهاست، رسمی‌اند.

در عمل، تحلیل کاوشی پیش‌نیازی برای استنباط رسمی است. اغلب داده‌های واقعی شامل موارد غیرمنتظره‌ای هستند که برخی از آنها ممکن است استنباط برنامه‌ریزی‌شده‌ای را نامعتبر یا ما را مجبور به اصلاح آن گرداند. این یکی از دلایلی است که به‌کارگذاشتن داده‌ها از طریق یک شیوهٔ استنباط پیچیده (و بنابراین، مکانیکی‌شده) قبل از کاوش دقیق در آنها نشانه‌ای از بی‌تجربگی در آمار است. رفت و برگشت بین داده‌ها و مدلها با ابزارهای تشخیصی پیشرفته ادامه می‌یابد به‌طوری این ابزارها امکان نقد مدلهایی خاص را با استفاده از داده‌ها می‌دهند. این ابزارها روح EDA را با نتایجی از تحلیل ریاضیِ پیامدهای مدلها درهم می‌آمیزند.

1. Best 2. Walker

1. John Tukey 2. Exploratory Data Analysis

کلسیم		دارونما
	- 1	1
5	- 0	5
234	- 0	33111
1	0	3
7	0	5
01	1	2
78	1	

شکل۴ نمودار ساقه‌ای موازی کاهش فشار خون سیستولیک برای دو گروه از مردان

همان‌گونه که قبلاً دیده‌ایم، مدل مثال ۲ الف از آن‌رو که بین مشاهده و آزمایش تمایزی قائل نمی‌شود، ناقص است. این مدل همچنین نظیر اغلب مدلهای ریاضی آرمانی برای پدیده‌های واقعی، غیرواقعی است. جمله‌ای به جورج باکس[1] آماردان منسوب است که می‌گوید «همهٔ مدلها نادرست‌اند، ولی بعضی از آنها مفیدند». استفاده‌کننده از روشهای استنباطِ مبتنی بر این مدل باید به‌دقت در بارهٔ کفایت این مدل برای وضعیت مورد بحث و داده‌ها کاوش کند. آیا نقیصه‌هایی در تولید داده‌ها (خواه نمونه‌ای باشد خواه آزمایشی) وجود داشته‌اند که استنباط را بی‌معنا کنند؟ آیا داده‌ها که یقیناً مشاهدات مستقلی روی یک توزیع کاملاً نرمال نیستند، به‌اندازهٔ کافی نرمال‌اند که اجازهٔ استفاده از روشهای استاندارد را بدهند؟ این سؤال با بررسی کاوشی خودِ داده‌ها همراه با اطلاع از اینکه تحلیل برنامه‌ریزی شده چقدر در مقابل انحراف از فرضهای مدل «استوار» است، پاسخ داده می‌شود.

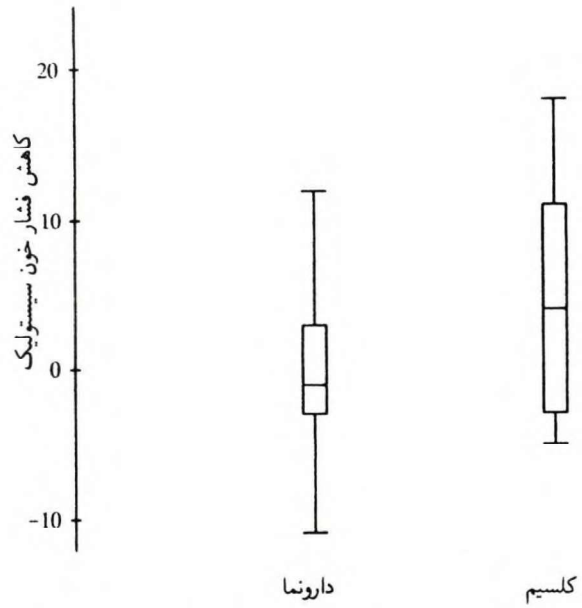
مثال ۲ د، کاوش مقدماتی داده‌های کلسیم. تحلیل ممکن است با این طرح کلی ساده آغاز شود: نمودار، شکل، مرکز و پراکندگی.

نمودار. یک نمودار ساقه‌ای هر مقدارِ داده‌ها را به یک ساقه و یک برگ تفکیک کرده، سپس برگها را روی ساقه‌های مشترک جور می‌کند. شکل ۴ یک نمودار ساقه‌ای پشت به‌پشت مفید برای مقایسهٔ دو گروه را نشان می‌دهد.

شکل. توزیع گروه دارونما تک‌مدی و متقارن است. با این حال گروه تیمار تصور ضعیفی از دومدی بودن را به‌وجود می‌آورد که امکان وجود دو نوع آزمودنی را مطرح می‌کند. آیا ممکن است کسانی باشند که به کلسیم واکنش نشان دهند و کسانی‌که به آن واکنش نشان ندهند؟ براساس این داده‌ها به‌هیچ عنوان چیزی نمی‌توان گفت، ولی این امکان ارزش توجه را دارد.

مرکز و پراکندگی. نموداری مفید برای مقایسهٔ مرکزها، پراکندگیها و تقارنها، نمودار جعبه‌ای است (شکل ۵). هر جعبه مکان چارکها و میانهٔ توزیع را مشخص می‌کند؛ «جاروبکها» از چارکها تا کرانگین‌ترین نقاط در محدودهٔ ۱٫۵ برابر دامنهٔ بین‌چارکی از نزدیکترین چارک، امتداد می‌یابند، و نقاط دورتر از میانه به‌طور مجزا نشان داده شده‌اند. ما در اینجا اختلافی بین میانه‌ها ملاحظه می‌کنیم، اما متوجه اختلاف بارزتری در پراکندگیها نیز می‌شویم، اختلافی که در فرض برابری واریانسها که برای توجیه برآورد ادغام‌شده در مثال ۲ الف به‌کار رفت ایجاد شبهه می‌کند.

1. George Box

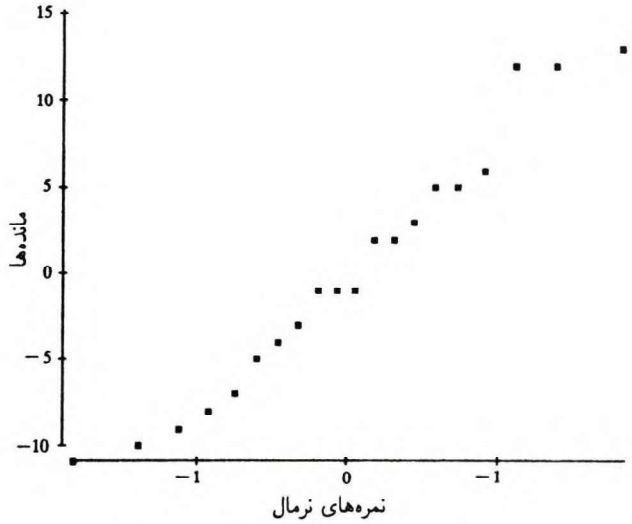


شکل۵ نمودار جعبه‌ای موازی برای کاهش فشار خون سیستولیک در دو گروه از مردان

نمودار چندکی نرمال. پیش از استفاده از آزمون t برای مقایسهٔ میانگینها، به صواب نزدیکتر است که این پرسش را مطرح کنیم. آیا داده‌ها دلیلی در اختیار ما می‌گذارند که مدل نرمال مثال ۲ الف را زیر سؤال ببریم؟ در اینجا از هر مشاهدهٔ میانگین گروه را کم می‌کنیم تا مانده‌ها را به‌دست آوریم، سپس مانده‌های مرتب شده را در مقابل چندکهای متناظر یک توزیع نرمال رسم می‌کنیم؛ (نگاه کنید به شکل ۶). عرضهای نقاط، ۲۱ ماندهٔ مرتب‌اند که خط حقیقی را به ۲۲ زیربازه تقسیم می‌کنند. طولهای نقاط ۲۱ مقداری هستند که خط حقیقی را به ۲۲ پاره‌خط که تحت مدل نرمال متساوی‌الاحتمال هستند تقسیم می‌کنند. اگر داده‌ها از تنها یک توزیع نرمال آمده باشند، می‌توانیم انتظار داشته باشیم که نقاط نزدیک یک خط قرار گیرند.

برای داده‌های کلسیم الگو تقریباً خطی است، هر چند جهش عمودی قبل از سه نقطهٔ منتهی‌الیه سمت راست، مانده‌های مشاهده‌شده‌ای را نشان می‌دهد که از مقادیر پیشگویی شده توسط مدل نرمال بزرگتر هستند، الگویی که با پراکندگیهای نابرابر در نمودار جعبه‌ای سازگار است.

در آموزشی که ساختار ریاضی دارد و معطوف به چگونگی استنتاج روشها از مدلهاست، در بسیاری موارد فقط کلیترین هشدارها در بارهٔ واقعیتهای عملی مطرح می‌شود. آمار در عمل به گفتگویی بین مدلها و داده‌ها شباهت دارد. مدلها برای فرایندی که داده‌ها را تولید می‌کند به راستی نقش اصلی را در استنباط آماری ایفا می‌کنند. بنابراین کاوش ریاضی خواص و پیامدهای مدلها مهم است (همان‌طور که در اقتصاد و فیزیک مهم است). اما داده‌ها همچنین می‌توانند مدلهای پیشنهادی را نقد یا حتی باطل بودن آنها را آشکار نمایند. در مثالهای کلسیم، تحلیل کاوشی به ما هشدار می‌دهد که خیلی زیاد به فرض برابری واریانسها متکی نباشیم، و از آزمون tی اصلاح‌شده‌ای استفاده کنیم که واریانسهای متمایزی برای دو گروه برآورد می‌کند. می‌توان اظهارنظر باکس را به این صورت تعبیر کرد که آمار صرفاً ریاضیات نیست: قضایای ریاضی درست‌اند؛ روشهای آماری اگر با مهارت به‌کار روند گاهی مؤثرند.



شکل۶ نمودار چندکی نرمال برای داده‌های فشار خون

امکانات گستردهٔ محاسبات ارزان، به‌ویژه کارهای گرافیکی، با تمایل به اینکه «ببینیم داده‌ها چه می‌گویند» درآمیخته و به تولید انبوهی از ابزارهای جدید منتهی شده است: پیشاپیش همه، نمودارهای ساقه‌ای و نمودارهای جعبه‌ای مثال ۲ پ را داریم، اما همچنین هموارکننده‌های نمودار پراکندگی آزاد-مدل، الگوریتمهای رگرسیونی مقاوم، ایده‌های هوشمندانهٔ نمایش داده‌های با بعد بالا برروی پرده‌های دوبعدی، و بسیاری ابزارهای پیشرفته‌تر تشخیصی برای وضعیتهای خاص در دست است. نرم‌افزارهای آماری متعارف بسیاری از اینها را انجام می‌دهند. کتابهای [۷] و [۹] به قلم دانشمندانی از آزمایشگاههای بل که از توکی تأثیر پذیرفته‌اند، بسیاری از مطالب گرافیکی پایه را ارائه می‌کنند. بسته‌های نرم‌افزاری S و S_PLUS که در آزمایشگاههای بل ابداع شده‌اند، بیشتر کارهای گرافیکی جدید را انجام می‌دهند، و نیز چندین رده از مدلهای جدید را به اجرا درمی‌آورند. برای بحث مفصل موضوع اخیر، ر. ک [۸].

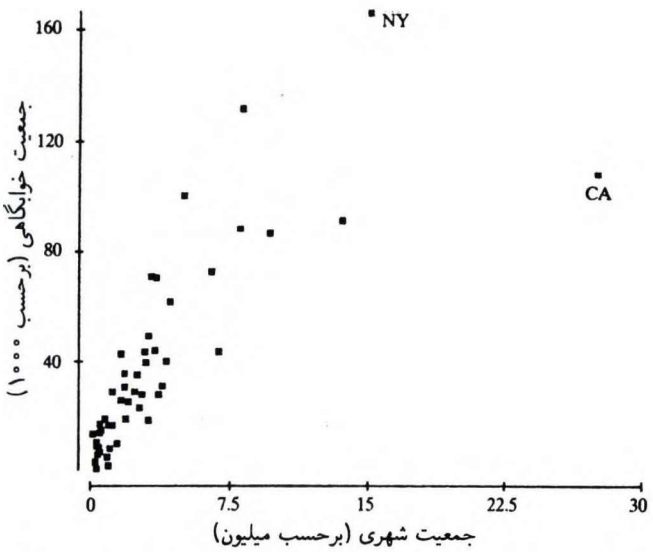
هر چند ممکن است این وسوسه در مبتدیان به‌وجود آید که تحلیل داده‌ها را صرفاً مجموعه‌ای از ابزارها و روشهای زیرکانه تلقی کنند، ولی ارزش این ابزارها در استفاده از آنها به‌روشی نظام‌مند و مطابق با استراتژیهایی است که برای سازماندهیِ بازبینی داده‌ها تنظیم شده‌اند.

۱. از ساده به پیچیده پیش بروید: ابتدا هر متغیر را جداگانه بررسی کنید، سپس رابطهٔ آنها با یکدیگر را ملاحظه کنید.

۲. از سلسله‌مراتبی از ابزارها استفاده کنید: ابتدا نمودار داده‌ها را رسم کنید، سپس توصیفهای عددی مناسبی از جنبه‌های مشخصی از داده‌ها را انتخاب کنید، بعد از آن در صورت اقتضا مدل ریاضی جمع‌وجوری برای الگوی کلی داده‌ها برگزینید.

۳. هم به الگوی کلی و هم به هرگونه انحراف چشمگیر از آن الگو توجه کنید.

بخشی از نظریهٔ یکپارچه‌ساز (اما غیرریاضی) EDA این است که این اصول در وضعیتهای مختلف به‌کار روند. اگر داده‌هایی از فقط یک متغیر کمّی در دست داشته باشیم، می‌توانیم توزیع را با یک نمودار ساقه‌ای نمایش دهیم، توجه کنیم که اگر تقریباً متقارن است، میانگین و انحراف معیار را به‌عنوان خلاصه‌های عددی محاسبه کنیم، و از یک نمودار چندکی نرمال استفاده



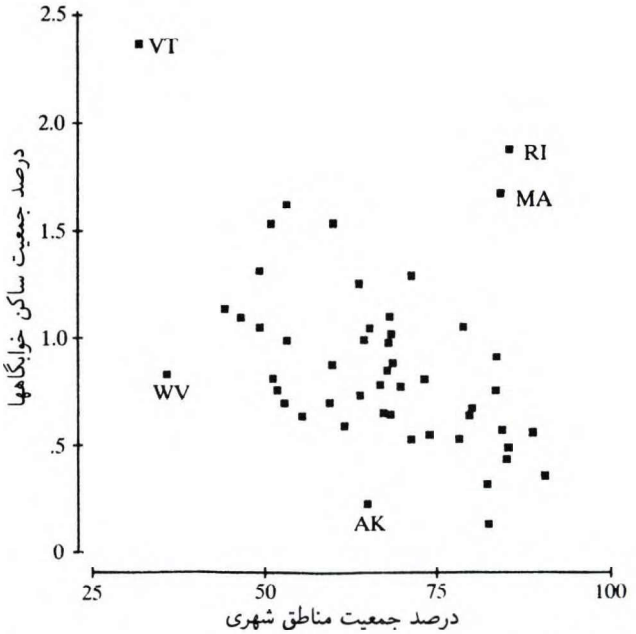
شکل۷ نمودار پراکنش نقطه‌ای جمعیت خوابگاهی در مقابل جمعیت شهری برای ۵۰ ایالت در آمریکا

کنیم تا ببینیم که آیا توزیع نرمالْ مدل جمع‌وجور مناسبی برای الگوی کلی است یا خیر. اما اگر دو متغیر کمّی در دست داشته باشیم، نمودار پراکنش را رسم می‌کنیم، سو و قوت پیوند خطی را به‌کمک همبستگی اندازه می‌گیریم، و در صورت اقتضا از یک خط راست برازش یافته به‌عنوان مدل برای الگوی کلی استفاده می‌کنیم. بنابراین دستورالعمل یک‌متغیرهٔ «نمودار، شکل، مرکز، پراکندگی» در چارچوب داده‌های دومتغیره به‌صورت «نمودار، شکل، سو، قوت» درمی‌آید.

در اینجا، مانند هر جای دیگر، تحلیل صرفاً جستجویی برای یافتن الگوها نیست، بلکه جستجویی برای یافتن الگوهای بامعنی است. همان‌طورکه مثال زیر نمایش می‌دهد، بهترین برازش لزوما مفیدترین نیست.

مثال ۳، خوابگاهها و شهرها، هر نقطه در شکل ۷ یکی از ۵۰ ایالت کشور امریکا را نشان می‌دهد که در آن مختص افقی برابر است با جمعیت شهری هر ایالت، و مختص عمودی برابر است با تعداد دانشجویان مقیم خوابگاه در آن ایالت. شکل ۷ چند جنبهٔ بارز دارد. برای مثال، نمودارْپره‌ای‌شکل است و نقاط بسیاری در سمت چپ پایین یک‌جا جمع شده‌اند: اغلب ایالتها جمعیت شهری نسبتاً کوچکی (حدود دو میلیون) و نیز جمعیت خوابگاهی نسبتاً کوچکی (زیر ۵۰۰۰۰) دارند، تنها تعداد کمی از ایالتها جمعیت شهری یا جمعیت خوابگاهی بسیار بزرگی دارند، و تغییر از ایالتی به ایالت دیگر (فضای بیشتر بین نقاط) برای ایالتهای با مقادیر بزرگتر، بیشتر است. الگوی پیوند بین دو متغیر مثبت و قوی است: جمعیتهای شهری کوچکتر با جمعیتهای خوابگاهی کمتر، و جمعیتهای شهری بزرگتر با جمعیتهای خوابگاهی بیشتر همراه است، و برای همهٔ ایالتها بجز تعداد کمی از آنها، دانستن اندازهٔ جمعیت شهری به ما اجازهٔ پیشگویی جمعیت خوابگاهی آن‌را در حدود دامنهٔ نسبتاً کوچکی می‌دهد.

علی‌رغم برازش خوب بین تصویر و ماجرا، در این تحلیل یک مشخصهٔ بسیار مهم نادیده گرفته شده است. اگر آنچه را که این الگو در ظاهر نشان می‌دهد بپذیریم که ایالتهای با جمعیت شهری بزرگ، جمعیتهای خوابگاهی



شکل۸ نمودار پراکنش داده‌های خوابگاهها و شهرها پس از تصحیح نسبت به جمعیت «متغیر پنهان»

بزرگی نیز دارند، ممکن است به این نتیجه‌گیری وسوسه شویم که شهرها دانشگاهها را بیشتر جذب کرده‌اند. هر چند نمونه‌های فراوانی دال بر تأیید این امر به ذهن می‌آید، این تفسیر ساده‌لوحانه اشتباه است: هر دو متغیر ما اندازه‌های غیرمستقیمی از اندازهٔ جمعیت ایالتها هستند، بنابراین چندان شگفت‌آور نیست که این دو مقدار، پیوند قوی مثبتی را نشان می‌دهند. برای آشکارکردن رابطهٔ بامعناتری، مجبوریم که «نسبت به متغیر پنهان تصحیح انجام دهیم»: جمعیت شهری را به جمعیت کل تقسیم کنید تا درصد شهری را به‌دست آورید، جمعیت خوابگاهی را بر جمعیت کل تقسیم کنید تا درصدی را که در خوابگاهها زندگی می‌کنند، به‌دست آورید، و نمودار نتایج را رسم کنید (شکل ۸).

حال این رابطه ضعیفتر، اما آنچه بیان می‌کند جالبتر است. سوی پیوند برعکس شده است: ایالتهای روستایی ــ آنهاکه درصد کمتری از ساکنانشان در نواحی شهری زندگی می‌کنند ــ درصد بیشتری از ساکنانشان در خوابگاههای دانشگاهی زندگی می‌کنند. این با قدری تأمل معقول به‌نظر می‌آید. به پولمن[1] در ایالت واشنگتن، یا ایمس[2] در ایالت آیوا، نورمن[3] در ایالت اکلاهما، لاورنس[4] در ایالت کانزاس بیندیشید.[5] ایالتهای روستایی ممکن است از نظر تعداد مطلق، کالجها و دانشگاههای کمتری داشته باشند، اما دانشجویان این ایالتها درصد بیشتری از کل جمعیت ایالت را تشکیل می‌دهند، و نیز احتمال بیشتری دارد که در خوابگاهها زندگی کنند.

۳.۲ استنباط رسمی: استدلال در مقابل شانس. استنباط آماریْ روشهایی آماری برای استنتاج از داده‌ها در بارهٔ جامعه یا فرایندی که داده‌ها از آن استخراج شده‌اند، به‌دست می‌دهد. در اینجا قائل شدن تمایز بین آماره‌های نمونه‌ای و پارامترهای جامعه ضرورت پیدا می‌کند (در تحلیل داده‌ها چنین

1. Pullman 2. Ames 3. Norman 4. Lawrence

۵. شهرهای کوچک دانشگاهی در آمریکا.م.

نبود. مقادیر واقعی پارامترها بر ما نامعلوم‌اند. ما آماره‌ها را در دست داریم، اما اگر تولید داده‌ها را تکرار کنیم، آنها مقادیر مختلفی به خود می‌گیرند. در استنباط باید این تغییرپذیری در نمونه به‌حساب آورده شود.

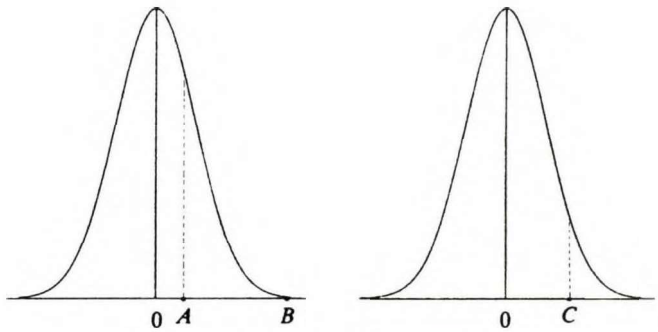
احتمالْ نوعی از تغییرپذیری، یعنی تغییرپذیری شانسی در پدیده‌های تصادفی، را توصیف می‌کند. وقتی یک سازوکار تصادفی به‌طور صریح برای تولید داده‌ها به‌کار می‌رود، احتمالْ تغییراتی را توصیف می‌کند که انتظار داریم در نمونه‌های تکراری از یک جامعه یا آزمایشهای تکراری در شرایطی یکسان، مشاهده کنیم. یعنی به این پرسش پاسخ می‌دهد که «چه اتفاقی می‌افتد اگر این کار را چندین بار انجام دهیم؟»

استنباط آماری استاندارد مبتنی بر احتمال است. این استنباطْ نتایجی را از داده‌ها همراه با اشاره‌ای به میزان اطمینان ما به این نتایج، ارائه می‌دهد. حکم مربوط به اطمینان براساس این سؤال است که «چه اتفاقی می‌افتد اگر این روش استنباط را چندین بار به‌کار بریم؟» این دقیقاً از نوع سؤالاتی است که احتمالْ می‌تواند پاسخ دهد، و به همین دلیل است که آن را مطرح می‌کنیم. این اشاره به میزان اطمینان ما به روشهای خود، که به زبان احتمال بیان می‌شود، چیزی است که استنباط رسمی را از نتایج غیررسمی مبتنی بر مثلاً، تحلیل کاوشی داده‌ها، متمایز می‌سازد.

هر شیوهٔ استنباطیِ خاص با یک آماره یا شاید چندین آماره، که از داده‌های نمونه‌ای محاسبه می‌شوند، شروع می‌شود. توزیع نمونه‌گیری، توزیع احتمالی است که توصیف می‌کند این آماره چگونه تغییر خواهد کرد هر گاه نمونه‌های بسیاری از یک جامعه استخراج شوند. در آمار مقدماتی دو نوع شیوهٔ استنباطی، بازه‌های اطمینان و آزمونهای معنی‌داری را ارائه می‌کنیم. بازهٔ اطمینانْ پارامتر نامعلومی را برآورد می‌کند. هدف از آزمون معنی‌داریْ ارزیابی شواهد حضور عامل مورد بررسی در جامعه است.

هر بازهٔ اطمینان مرکب از دستورالعملی است برای برآوردکردن پارامتری نامعلوم با استفاده از داده‌های نمونه‌ای، معمولاً به شکل «برآورد ± حاشیهٔ خطا»، و یک سطح اطمینان که احتمالِ آن است که دستورالعمل واقعاً بازه‌ای تولید کند که مقدار واقعی پارامتر را در برداشته باشد. یعنی، سطح اطمینان به این سؤال پاسخ می‌دهد که «اگر از این روش چندین بار استفاده کنم، چند بار جواب درستی به من می‌دهد؟»

آزمون معنی‌داری با این فرض آغاز می‌شود که عامل مؤثری که در جستجوی آن هستیم در جامعه حضور ندارد. طی آن سؤال می‌شود که «در چنین صورتی، آیا نتیجهٔ نمونه‌ای تعجب‌آور است یا خیر؟» یک احتمال (pـمقدار) می‌گوید که نتیجهٔ نمونه‌ای چقدر تعجب‌آور است. نتیجه‌ای که در صورت عدم حضور عاملی که در جستجوی آن هستیم به‌ندرت به‌دست می‌آید گواه خوبی است بر اینکه آن عامل درواقع حضور دارد. شکل ۹ این استدلال را در مثال پزشکی ما نشان می‌دهد. خم نرمال در آن شکل، معرفِ توزیع نمونه‌ای اختلاف $\overline{x} - \overline{y}$ بین کاهشها در میانگین فشار خون در گروههای کلسیم و دارونما برای حالتی است که اختلافی بین دو میانگین جامعه وجود ندارد. این توزیع، که تغییرپذیریِ صرفاً شانسی را نشان می‌دهد دارای میانگین ۰ است. برآمدهای بزرگتر از ۰ از آزمایشهایی می‌آیند که در آنها کلسیم فشار خون را بیشتر از دارونما کاهش می‌دهد. اگر نتیجهٔ A را مشاهده



شکل۹ ایدهٔ معنی‌داری آماری: آیا این مشاهده تعجب‌آور است؟

کنیم، تعجب نمی‌کنیم زیرا پیشامدی که با صفر این‌قدر فاصله دارد، اغلب شانسی رخ می‌دهد. این نتیجه هیچ‌گونه گواه معتبری بر این ادعا که کلسیم بر دارونما برتری دارد در اختیار نمی‌گذارد. اگر نتیجهٔ B را مشاهده کنیم، آزمایش نتیجه‌ای چنان قاطع به‌بار آورده است که تقریباً هرگز به‌طور شانسی نمی‌تواند رخ دهد. در این صورت شاهدی قوی داریم که میانگین کلسیم از میانگین دارونما بیشتر است. pـمقدار (احتمال دُم راست) برای نقطهٔ A برابر است با ۰٫۲۴ و برای نقطهٔ B برابر است ۰٫۰۰۰۵. این احتمالها میزان تعجب‌آوربودن مشاهده‌ای به این بزرگی را وقتی‌که عامل مؤثری در جامعه وجود ندارد به‌صورت کمّی درمی‌آورند. در مورد داده‌های واقعی چطور؟ نقطهٔ C مقدار مشاهدهٔ شده $\overline{x} - \overline{y} = ۵٫۲۷۳$ را نشان می‌دهد. pـمقدار متناظر ۰٫۰۵۵ است. کلسیم در ۵٫۵٪ از تعداد زیادی آزمایش، حداقل به این مقدار بر دارونما صرفاً به‌علت تغییرات شانسی برتری دارد. آزمایش شواهدی به‌دست می‌دهد که کلسیم مؤثر است اما این شواهد چندان قوی نیستند. نکته‌ای برای آنهایی که نگران جزئیات‌اند: در این محاسبات pـمقدار، تغییرات میانگین نمونه‌ای معلوم فرض شده است. در عمل، ما باید انحراف معیارها را از روی داده‌ها برآورد کنیم. آزمون حاصل pـمقدار بزرگتری دارد: $p = ۰٫۰۷۲$.

۳. تدریس: در بحث تدریس، می‌توانیم توجه خود را بر محتوا متمرکز کنیم، یعنی چیزی که می‌خواهیم دانشجویان فراگیرند، یا بر فن آموزش، یعنی کاری که می‌کنیم تا به آنها در فراگرفتن کمک کنیم. البته این دو مبحث مرتبط هستند. به‌ویژه تغییرات در فن آموزش اغلب با تغییر بعضی از اولویتها در نوع چیزهایی که می‌خواهیم دانشجویانمان یاد بگیرند به پیش برده می‌شود. با این حال مناسبتر است که محتوا و فن آموزش را جداگانه مورد بحث قرار دهیم. این بخش هماهنگ با قسمتهای دیگر مقاله، محتوا را مورد توجه قرار می‌دهد، و به‌ویژه، شامل وجهی از گفتگو بین آماردانان و ریاضیدانانی است که به تدریس آمار هم می‌پردازند.

۱.۳ آمار باید به عنوان آمار تدریس شود. آماردانان متقاعد شده‌اند که آمار با وجود آنکه وابسته به ریاضیات است، از شاخه‌های ریاضیات نیست. مانند اقتصاد و فیزیک، آمار استفاده‌ای گسترده و اساسی از ریاضیات می‌کند، اما قلمرو خاص خود را برای کاوش و مفاهیم اساسی خود را برای هدایت این کاوش دارد. با توجه به این اعتقادات، به‌طور طبیعی ترجیح می‌دهیم که آمار مقدماتی به عنوان آمار تدریس شود. انجمن آمریکایی آمار[1] (ASA) و جامعهٔ ریاضی آمریکا[2] (MAA) کمیته‌ای مشترک برای بحث در مورد برنامهٔ درسی آمار مقدماتی تشکیل داده‌اند. توصیه‌های این گروه نمودی از این نظر است که آموزش آمار باید بر ایده‌های آماری متمرکز شود. گزیده‌ای از این توصیه‌ها به نقل از [۱۰] در زیر می‌آید؛ بحثی مفصلتر در [۱۱] آمده است.

> تقریباً هر درسی در آمار با تأکید بیشتر بر داده‌ها و مفاهیم و توجه کمتر به نظریه و آوردن دستورالعملهای کمتر، بهتر می‌شود. تا سرحد امکان، به محاسبات و کارهای گرافیکی باید حالت مکانیکی و سرراست داده شود. هدف اصلی هر درس مقدماتی باید کمک به دانشجویان در یادگیری اصول تفکر آماری باشد [که مشتمل‌اند بر] نیاز به داده‌ها، اهمیت تولید داده‌ها، همه جا حاضربودن تغییرپذیری، کمّی‌سازی و توضیح تغییرپذیری.

توصیه‌های کمیتهٔ ASA/MAA منعکس‌کنندهٔ تغییراتی در حوزهٔ آمار در طی دو سه دههٔ پیشین است. آمار دانشگاهی، برخلاف ریاضیات به حیطه‌ای گسترده‌تر که کارورزی حرفه‌ای غیردانشگاهی باشد، متصل است. فناوری محاسباتی، کارورزی آماری را کاملاً تغییر داده است. پژوهشگران دانشگاهی، تا حدی بر اثر تقاضاهای کارورزی و تا حدی به علت تواناییهای فناوری نوین، سلیقه‌های خود را در پژوهش تغییر داده‌اند. روشهای خودگردان[3]، هموارسازی ناپارامتری داده‌ها، تشخیصیات[4] رگرسیونی، و رده‌هایی عامتر از مدلها که مستلزم برازشهای تکراری‌اند، ازجمله ثمرات جدید توجه مجدد به تحلیل داده‌ها و استنباط علمی‌اند. افرون و تیبشیرانی [۱۴] بخشی از این کار را برای غیرمتخصصان توصیف کرده‌اند.

۲.۳ نه ریاضیات نه جادوگری. تأکید زیاد بر استنباطِ مبتنی بر احتمال، یکی از نشانه‌های افراط در ریاضی در درس آمار مقدماتی است، و در عین حال، اینکه مدرسانِ برخوردار از تربیت و تفکر ریاضی دوست ندارند از ارائهٔ آمار براساس نظریه دست بکشند، مبنای قابل احترامی دارد: برای اجتناب از معرفی آمار به‌عنوان جادوگری است. مسلماً تدریس آمار مقدماتی به‌صورت جادوگری کار رایجی است. استفاده‌کنندهٔ آمار از جهات بسیاری شبیه شاگرد جادوگر است. سحر و جادو تأثیری خودبه‌خودی دارد که پایان‌نامه‌ها را قابل پذیرش و مطالعات را قابل چاپ می‌گرداند. منظور این نیست که ما بفهمیم سحر چگونه کار می‌کند ــ این در حیطهٔ کار خود جادوگر است. سحر باید به‌طور دقیق دستورالعمل را دنبال کند تا فاجعه‌ای پیش نیاید ــ کاوش و انعطاف‌پذیری هم، مانند فهمیدن، برای شاگردان ممنوع است. خوشبختانه جادوگران نرم‌افزاری تدارک دیده‌اند که تقلید دقیق از سحرهای پذیرفته شده را به‌صورت ماشینی و سرراست درآورده است.

خطر [تلقی] آمار به‌عنوان جادو، واقعی است. اما دفاع مناسبْ عقب‌نشینی به‌سوی عرضهٔ ریاضی مطلب نیست که برای تشریح موضوع، ناکافی است و برای دانشجویان اغلب غیرقابل درک است. درک ریاضی تنها نوع درک نیست. حتی در اغلب موضوعات علمی که از ریاضیات استفاده می‌کنند و در

1. American Statistical Association
2. Mathematical Association of America
3. Bootstrap
4. diagnostics

آنها درک پدیدهٔ هدف و مفاهیم اصلی موضوع اولویت دارند، درک ریاضی، سودمندترین نوع آن هم نیست. باید تلاش کنیم تا چارچوبی عقلانی عرضه کنیم که مجموعهٔ ابزارهایی را که آماردانان به‌کار می‌برند قابل معنا گرداند و مشوق کاربرد انعطاف‌پذیر آنها در حل مسائل باشد. دانشجویان، ریاضیات را وقتی درک می‌کنند که قدرت تجرید، استنتاج و بیان نمادین را ارج گذارند، و بتوانند به‌صورتی انعطاف‌پذیر از ابزارها و استراتژیهای ریاضی در مواجهه با مسائل گوناگون استفاده کنند. استدلال بر مبنای داده‌های تجربی ناحتمی، روشی عقلانی است که همان‌قدر قدرتمند و نافذ است. چگونه می‌توانیم به بهترین وجه دانشجویان خود را راهنمایی کنیم که این روش عقلانی را درک کنند، ارج گذارند، و شروع به جذب آن کنند؟

۳.۳ با تحلیل کاوشی داده‌ها شروع کنید. هر چند ترتیبی که با توجه به شکل ۲ به ذهن می‌آید، شروع از تولید داده‌هاست، تجربه عکس آن را می‌گوید. تحلیل کاوشی داده‌ها آغاز خوبی است زیرا ملموستر است؛ در آن نیازی به تمایز قائل شدن بین جامعه و نمونه نیست و نیازی نیست که وجوه تصادفی‌سازی که در مقابل اریبی حفاظ ایجاد می‌کنند، مورد بحث قرار گیرند. روشهای پایه‌ای از نظر مفهومی و الگوریتمی ساده‌اند، و داده‌ها در دست‌اند ــ اعداد واقعی بر روی صفحهٔ کاغذ، نه اشباح داده‌های آینده، که در طرح و آزمایش با آنها سروکار داریم. به‌علاوه، ایجاد انگیزه مطرح نیست. دانشجویان تحلیل کاوشی را دوست دارند و درمی‌یابند که می‌توانند آن‌را انجام دهند که این خود مزیتی اساسی در تدریس موضوعی است که اکثر دانشجویان از آن وحشت دارند. درگیرکردن دانشجویان در تعبیر نتایج از همان اوان کار، قبل از آنکه ایده‌های دشوارتر بر سر راهشان ظاهر شده توجه آنها را به خود جلب کنند، می‌تواند به ایجاد عادات خوبی در آنها کمک کند که وقتی به مبحث استنباط می‌رسید بهرهٔ خود را خواهد داد. سرانجام، شروع با تحلیل داده‌ها زمینه را بر این طرح و استنباط آماده می‌کند. تجربهٔ کار با توزیعهای داده‌ها، دانشجویان را با همه جا حاضربودن تغییرپذیری و وجود بالقوهٔ اریبی آشنا می‌کند که این دو جنبه، دلایل اصلی نیاز ما به طرح دقیق است. اگر طرح را قبل از تحلیل داده‌ها تدریس کنید، برای دانشجو مشکلتر خواهد بود بفهمد که چرا طرح اهمیت دارد. تجربهٔ کار با توزیعهای داده‌ها همچنین بهترین راهِ آماده شدن برای رویارویی با مفهوم دشوار توزیع نمونه‌گیری است.

کوشش کرده‌ایم این موضوع را مطرح کنیم که مجموعه‌ای منسجم (هر چند غیرریاضی) از ایده‌ها و ابزارهای وابسته برای کاوش در داده‌ها وجود دارد. دانشجویان لازم است با نوشتن شرحهای منسجمی در بارهٔ داده‌ها، استفاده از این ایده‌ها و ابزارها را تمرین کنند. برای کمک به آنها هم خلاصه‌ای از آنچه را که باید بنویسند و هم مثالهایی که می‌توانند به‌عنوان مدل به‌کار روند، فراهم آورده‌ایم. برای مثال، شکل ۱۰، طرحی برای توصیف تنها یک متغیر کمّی است.

دنبال کردن این طرح هم نیاز به اطلاع در بارهٔ ابزارهایی دارد که در آن ذکرشده است و هم نیاز به قدرت تشخیص برای انتخاب از بین شقوق و تفسیر نتایج. قدرت تشخیص از تجربهٔ کار با داده‌ها به‌دست می‌آید. دانشجویان در ابتدا نمی‌توانند آن‌طور که کلمات و معادلات را می‌فهمند، از نمودارها

الف. داده‌ها را توصیف کنید
- تعداد مشاهدات
- ماهیت متغیر
- چگونه اندازه‌گیری شده است
- واحدهای اندازه‌گیری

ب. نمودار داده‌ها را رسم کنید، با انتخاب از میان
- نمودار نقطه‌ای
- نمودار ساقه‌ای
- بافت نگار

پ. الگوی کلی را توصیف کنید
- شکل
  - شکل مشخصی وجود ندارد؟
  - چاوله یا متقارن؟
  - قله‌های یگانه یا چندگانه؟
- مرکز و پراکندگی؛ با انتخاب از میان
  - خلاصهٔ پنج عددی
  - میانگین و انحراف معیار
- آیا نرمال بودن یک مدل مناسب است (نمودار چندکی نرمال)؟

ت. انحرافات آشکار از الگوی کلی را وارسی کنید
- دورافتاده‌ها
- رخنه‌ها یا خوشه‌ها

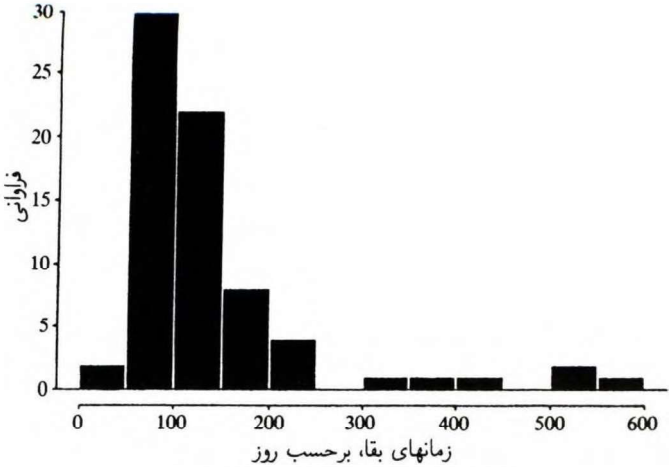
خ. به زبان زمینهٔ مسأله، یافته‌هایتان در پ و ت را تعبیر کنید. توصیفهای موجهی برای یافته‌های خود پیشنهاد کنید.

شکل ۱۰ طرح توصیف داده‌ها برای یک متغیر کمی

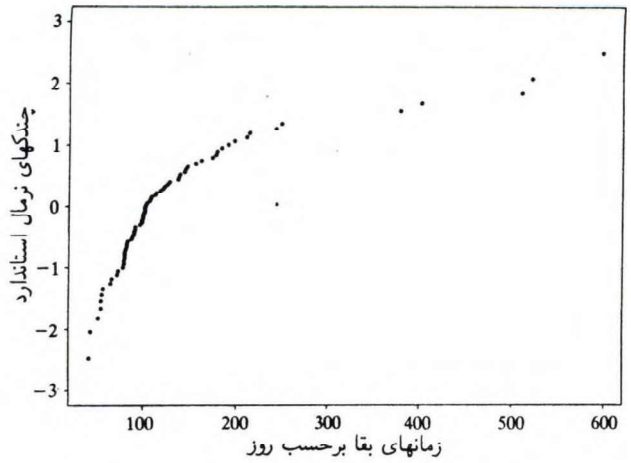
سردرآورند. در زیر مثالی از یک تحلیل بنیادی از یک متغیر ارائه می‌شود. توصیف روابط بین چندین متغیر مستلزم ابزارهای استادانه‌تر و قدرت تشخیص بیشتری است.

> در مطالعه‌ای در بارهٔ مقاومت در مقابل ابتلا به بیماری [۲]، پژوهشگرانْ باسیل سل را به ۷۲ خوکچهٔ هندی تزریق و مدت زمان زنده ماندن آنها را بعد از ابتلا به بیماری برحسب روز اندازه‌گیری کرده‌اند. هم بافت‌نگار (شکل ۱۱) و هم نمودار چندکی نرمال (شکل ۱۲) نشان می‌دهند که توزیع زمانهای بقا شدیداً چاوله به راست است. دادهٔ دورافتاده‌ای در بین نیست گرچه بعضی به مراتب بیشتر از میانگین زنده مانده‌اند، که به نظر یکی از مشخصه‌های توزیع کلی می‌آید و نه حاکی از، مثلاً، خطا در اندازه‌گیری یا ثبت این موردها.

چاولگی شدید حاکی از آن است که خلاصهٔ پنج عددی (مینیمم = ۴۳ روز،



شکل۱۱ بافت‌نگار زمانهای بقای خوکچهٔ هندی



شکل۱۲ نمودار چندکی نرمال برای زمانهای بقای خوکچهٔ هندی

چارک اول = ۸۲٫۵ روز، میانه = ۱۰۲٫۵ روز، چارک سوم = ۱۵۱٫۵ روز، و ماکسیمم = ۵۹۸ روز) خلاصهٔ عددی بهتری نسبت به میانگین و انحراف معیار ($\overline{x}$ = ۱۴۱٫۸ و s = ۱۰۹٫۲ روز) است. تغییرات بسیار بزرگی در زمانهای بقا بین موردها وجود دارد ــ برای مثال، چارک سوم تقریباً ۱۵۰٪ میانه است و بزرگترین ۶ مشاهده بیشتر از دو برابر میانه هستند. بدون اطلاعات بیشتر، نمی‌توانیم زمان بقای مورد مبتلایی را دقیقاً پیشگویی کنیم. به‌علاوه شیوه‌های tی استاندارد را نباید برای استنباط در بارهٔ زمانهای بقا به‌کار برد. در استنباط می‌توان یک توزیع غیرنرمال را به‌عنوان مدل به‌کار برد یا به دنبال تبدیل به مقیاسی بود که بیشتر نرمال باشد.

هر چند بیشتر دانشجویانی که وارد نخستین درس آمار می‌شوند، انتظار یک درس تشریفاتی توخالی را دارند، EDA به آنها این شگفتی دلپسند را عرضه می‌کند که روشهایی موجودند که در یافتن معنا به‌کار می‌آیند. این شگفتی چنان به مذاق آنها خوش می‌آید که خطر افتادن از آن طرف بام را دارد. بعضی از دانشجویان ممکن است به سمت این اعتقادِ دلخوش‌کننده رانده شوند که هر داستانی در بارهٔ داده‌ها که تطابق آن با الگوها قابل‌قبول به نظر رسد باید درست باشد. حالا زمان برای ترکیب مناسبی از طرح و تردیدگری مناسب است.

۴.۳ طرح را به‌عنوان پلی بین تحلیل داده‌ها و استنباط تدریس کنید. بحث مقدماتی در طرح تولید داده‌ها به‌طور طبیعی بین تحلیل کاوشی و استنباط جا دارد: طرح بی‌نقص چیزی است که استنباط را ممکن می‌سازد. منتظر ماندن برای معرفی توزیعهای احتمال تا زمانی که مواد پایه‌ای طرح عرضه شده باشد چند مزیت دارد. از یک نظر، این ترتیب به روشن شدن این مطلب کمک می‌کند که توجیه مدلهای احتمال باید براساس تصادفی بودن فرایند تولید داده‌ها باشد، و بنابراین حفاظی در مقابل پذیرشِ بدون تأملِ مدلهای احتمال است. از جنبهٔ دیگر، یادگیری مطالبی در بارهٔ تولید داده‌ها، دانشجویان را پیش از مواجهه با توزیع نمونه‌گیری که فی‌نفسه مشکل است، با مفاهیم ضروری مانند جمعیت و نمونه، پارامتر و آماره آشنا می‌کند.

مهمترین نکته برای دانشجویان این است که بفهمند چرا آزمایشهای مقایسه‌ای تصادفی، «محک‌زرین» برای بررسی شواهد علیت است. یک منبع غنی از حکایتهای هشداردهندهٔ مربوط به زندگی واقعی کتاب [۶] است که توسط دو پزشک یعنی بانکر و بارنس و یک آماردان یعنی موستلر ویرایش شده است و مثالهای تکان‌دهنده‌ای از شیوه‌های درمان پزشکی دربر دارد که پیش از آنکه علم پزشکیْ آزمایشهای مقایسه‌ای تصادفی‌شده را بپذیرد، به‌صورت استاندارد درآمده بودند اما پس از اینکه در معرض آزمون صحیح قرار گرفتند، بی‌ارزش بودن آنها آشکار شد.

البته جنبهٔ آماری طرح آزمایشها و بررسیهای نمونه‌ای چیزی بیش از این دربردارد. طرحهایی که درعمل به‌کار می‌روند اغلب کاملاً پیچیده‌اند، و باید بین کارآیی از یک طرف و نیاز به اطلاعاتی با دقت متفاوت در بارهٔ بسیاری از عاملها و اثر متقابل آنها از طرف دیگر، توازن برقرار شود. طرحهای ساده ــ آزمایشهای تصادفی که دو یا چند تیمار را مقایسه می‌کنند، یا نمونه‌های تصادفی ساده از یک یا چندین جامعه ــ روشنگر مهمترین ایده‌ها و مؤیّدِ استنباطی هستند که در نخستین درس آمار مقدماتی تدریس می‌شود. شما باید در بارهٔ این طرحها صحبت کنید، اما نیازی نیست که از این فراتر روید. برخی مطالب مهم دیگر، برای مثال، شیوه‌های طرح و آزمون کردن پرسشهای بررسی[1] و شیوه‌های آموزش دادن و نظارت بر مصاحبه‌کننده‌ها، معمولاً در درسهای آماری ارائه نمی‌شوند. دانشجویان آمار باید آگاه باشند که این مهارتهای عملی مهم هستند، و جریان تولید داده‌ها، حتی وقتی‌که با یک طرح بی‌نقص آماری شروع می‌کنیم، ممکن است دچار کجروی شود. اینکه در اینجا چقدر وقت صرف کنید، موضوعی است که به صلاحدید شما در مورد نیازهای مخاطبان بستگی دارد.

۵.۳ استنباط: دو مانع در برابر فهمیدن. بخش ۳.۲ مختصراً نحوهٔ کار استنباط را توصیف کرده است. از آنجا که جزئیات در عمل به‌صورت سرراست و مکانیکی درآمده‌اند، دوست داریم که دانشجویان بیشتر کوشش خود را صرف درک ایده‌ها کنند. درک این ایده‌ها آسان نیست. اولین مانعْ مفهوم توزیع نمونه‌گیری است. موردی ساده همچون استفاده از نسبت $\hat{p}$ی یک نمونه از کارگران بیکار برای برآورد نسبت pی کارگران بیکار در کل جامعه را انتخاب کنید. مثالهای فیزیکی (نمونه‌گیری مهره‌ها از یک جعبه)، شبیه‌سازی‌های

1. survey questions

کامپیوتری و آزمایشهای فکری همه به انتقال این ایده که به‌ازای نمونه‌های بسیار، مقادیر بسیار $\hat{p}$ داریم کمک می‌کنند. دائم بپرسید «چه اتفاقی خواهد افتاد اگر این کار را به دفعات بسیار انجام دهم؟» این پرسش کلید منطق استنباط آماری استاندارد است.

به محض اینکه ایدهٔ توزیع نمونه‌گیری دریافته شد، ابزارهای تحلیل داده‌ها به ما کمک می‌کنند تا گامهای بعدی را برداریم. در مواجهه با هر توزیع، دربارهٔ شکل، مرکز، و پراکندگی سؤال می‌کنیم. شکل توزیع نمونه‌گیری $\hat{p}$ تقریباً نرمال است. میانگین آن برابر با نسبت جامعه‌ای نامعلوم، p، است. این مطلب حاکی از آن است که $\hat{p}$ به‌عنوان برآوردگر p، اریبی یا خطای سیستماتیک ندارد. دقت برآوردگر با پراکندگی توزیع نمونه‌گیری توصیف می‌شود که آن را به یمن نرمال بودن برحسب انحراف معیار آن اندازه‌گیری می‌کنیم.

دومین مانع عمدهٔ استدلال، آزمونهای معنی‌داری است. هر چند ایدهٔ اصلی («آیا این برآمد تعجب‌آور است؟») بغرنج نیست، جزئیات موضوع مرعوب‌کننده است. هیچ راه گریزی از فرضهای صفر و مقابل و آزمونهای یکطرفه در مقابل دوطرفه وجود ندارد. منطق آزمون کردن که با چنین عبارتی شروع می‌شود: «موقتاً فرض کنید عامل مؤثری که در جستجوی آن هستیم، حضور ندارد»، سرراست نیست. دوست داریم که اکثر دانشجویان ما ایدهٔ توزیع نمونه‌گیری را بفهمند؛ می‌دانیم که بسیاری از آنها استدلال آزمونهای معنی‌داری را نخواهند فهمید. پس قدری کوتاه می‌آییم ولی پافشاری می‌کنیم که باید بتوانند معنای pـمقدار تولیدشده به‌وسیلهٔ نرم‌افزار یا گزارش‌شده در یک مجلهٔ علمی را بیان کنند. این بخشی از توصیهٔ کلی ماست که اصرار داریم دانشجویان شرحهای خیلی مختصری از یافته‌های آماری بنویسند. «در این بررسی، دو روش تدریس روخوانی به دانش‌آموزان کلاس سوم مقایسه می‌شود. یک آزمون tی دو نمونه‌ای که میانگین نمرات دو گروه تیمار در یک آزمون روخوانی استاندارد را با هم مقایسه می‌کند، pـمقداری برابر ۰٫۰۱۹ داشته است، یعنی در این بررسی، اثری به آن بزرگی مشاهده شده است که تنها حدود ۲٪ مواقع ممکن است به‌طور شانسی رخ دهد. این گواه بسیار قوی حاکی از آن است که روش جدید منجر به نمرهٔ میانگین بیشتری در مقایسه با روش استاندارد می‌شود.»

اما دو نکتهٔ پایانی در بارهٔ استنباط: نخست اینکه، درک مفهومیِ ایده‌ها تقریباً تصویری است، یعنی مبتنی است بر کشیدنِ شکلِ توزیع نمونه‌گیری و استفاده از تاکتیکهایی که در تحلیل داده‌ها آموخته می‌شوند. حتی مقدار بسیار زیادی ریاضیات صوری نمی‌تواند جایگزین این دید تصویری شود، و هیچ مقدار استنتاج ریاضی نمی‌تواند به دانشجویان ما کمک کنند که به این دید دست یابند. ریاضیات در دانستن حقایق جنبهٔ اساسی دارد، اما این امر به معنای آن نیست که ما باید ریاضیات را به دانشجویانمان تحمیل کنیم.

دوم اینکه می‌خواهیم آموخته‌های دانشجویان خیلی بیشتر از تصویر کلیِ مطلب و چند دستورالعمل برای اجرای آن در زمینه‌های مشخص باشد. ذیلاً چند نکتهٔ دیگر، هم عملی و هم نظری، تقریباً به ترتیب اهمیت ذکر می‌شوند. اینکه تا کجای فهرست باید پیش روید، بستگی به مخاطبان شما دارد.

- مطالعهٔ شیوه‌های مشخصی از استنباط، خصوصیاتی را آشکار می‌کند که بین همهٔ آنها مشترک است و تمام دانشجویان باید آنها را بفهمند. به‌دست آوردن اطمینان بیشتر براساس داده‌هایی واحد، به بهای حاشیهٔ خطای بزرگتر صورت می‌گیرد. حتی اثراتی آن‌قدر کوچک که عملاً بی‌اهمیت هستند، از نظر آماری بسیار بااهمیت‌اند هرگاه آزمون معنی‌داری را بر مبنای نمونه‌ای بسیار بزرگ انجام دهیم.
- بسیاری چیزها ممکن است نادرست درآیند به‌طوری که در ارزش استنباط ایجاد تردید شود. مقایسهٔ آزمودنیهایی که داوطلب مصرف کلسیم می‌شوند در مقابل دیگرانی که چنین نمی‌کنند، چیز زیادی در بارهٔ اثرات کلسیم به ما نمی‌گوید، زیرا آنها که داوطلب مصرف کلسیم می‌شوند، ممکن است عموماً از افرادِ مراقب سلامتی خود باشند. یک دادهٔ بسیار دورافتاده می‌تواند نتایج آزمایش پزشکی ما را به هر سو منحرف کند و باز استنباط را بی‌اعتبار نماید. جریان تولید داده‌ها را بازبینی کنید. نمودار داده‌ها را رسم کنید. پس از آن، اگر شد، به سراغ استنباط بروید.
- خود شیوه‌های استنباط به ما نمی‌گویند که چیزی اشتباه شده است. برای مثال، حاشیهٔ خطا در یک بازهٔ اطمینان، تنها تغییرات شانسی در نمونه‌گیری تصادفی را دربردارد. همان‌طور که روزنامهٔ نیویورک تایمز معمولاً در کادری همراه با نتایج نظرسنجی خود ذکر می‌کند، «علاوه بر خطای نمونه‌گیری، مشکلات عملی اجرای هرگونه سنجش افکار عمومی ممکن است منابع دیگری از خطا را وارد نظرسنجی کند.»
- شیوه‌های رایج استنباط درواقع بر مدلهای ریاضی مبتنی هستند، نظیر استنباطی که در مثال پزشکی ما دیده می‌شود: X_1، X_2، ...، X_n مستقل و دارای توزیع یکسان $N(\mu_1, \sigma_1)$، و Y_1، Y_2، ...، Y_m مستقل و دارای توزیع یکسان $N(\mu_2, \sigma_2)$ هستند. این مدل دقیقاً درست نیست، آیا مفید است؟ در واقع شیوه‌های tی دو نمونه‌ای که وقتی می‌خواهیم μ_1 و μ_2 را مقایسه کنیم از این مدل ناشی می‌شوند، در مقابل غیرنرمال بودن استوارند، در نتیجه این مدل عملاً به شیوه‌های مفید منجر می‌شود. اما آمارهٔ نسبت واریانسهای F برای مقایسهٔ σ_1 و σ_2 به غیرنرمال بودن بسیار حساس است تا آن حد که از ارزش عملی چندانی برخوردار نیست. حتی مبتدیان لازم است از وجود چنین مشکلاتی آگاه شوند.
- اغلب در شرایطی می‌خواهیم استنباط را انجام دهیم که داده‌هایمان از یک نمونهٔ تصادفی یا یک آزمایش مقایسه‌ای تصادفی نیامده‌اند. برای مثال به اندازه‌گیری روی قطعه‌های متوالی که در یک خط مونتاژ در حرکت‌اند، فکر کنید. استنباط براساس یک مدل احتمال برای فرایندی که داده‌هایمان را تولید می‌کنند، قابل توجیه است، و درستی مدل را می‌توان تا حدی از خود داده‌ها ارزیابی کرد. تولید تصادفی داده‌ها نمونهٔ عالی و امن‌ترین زمینه برای استنباط است، اما تنها زمینهٔ مجاز نیست.
- استنباط استقرایی از داده‌ها از نظر مفهومی پیچیده است. شگفت‌آور نیست که راههای دیگری برای اندیشیدن در بارهٔ آن وجود دارد. نظریهٔ استاندارد آماری تمایل به این برداشت از استنباط دارد که هدف از آن تصمیم‌گیری است. برای مثال، آزمون باید بین فرضهای صفر و مقابل تصمیم بگیرد. این کار فوراً به خطاهای نوع I و II و غیره منجر می‌شود. رهیافت تصمیم‌گیری به زحمت با منطق این سؤال: «آیا این برآمد تعجب‌آور است؟» که برحسب pـمقدارها ابراز می‌شود، سازگاری دارد. به نظر ما ارزیابی قدرت شواهد، هدف متداولتری است تا اتخاذ یک تصمیم، اما همه با این نظر موافق

نیستند. مکتب فکری بیزی با ارائهٔ توصیف صریحی از اطلاع پیشین موجود در هر وضعیت آماری و با ترکیب اطلاع پیشین با داده‌ها برای رسیدن به یک تصمیم، پا را فراتر از این می‌گذارد. تقریباً همهٔ آماردانان فکر می‌کنند که این ایده، گاهی خوب است. پیروان مکتب بیز فکر می‌کنند که همهٔ مسائل آماری را می‌توان با این الگو وفق داد. این موضع اقلیتی از آماردانان است (که سخت به آن باور دارند). کار مشکلی در پیش رو داریم.

۶.۳ در مورد احتمال چطور؟ احتمال بخشی اساسی از هر برنامهٔ آموزش ریاضیات است. احتمال حوزه‌ای پرظرافت و قدرتمند از ریاضیات است که تأثیرات متقابل آن با دیگر حوزه‌های ریاضیات، به کل این موضوع غنا می‌بخشد. احتمال همچنین برای مطالعهٔ جدی ریاضیات کاربردی و مدلبندی ریاضی جنبهٔ اساسی دارد. قلمرو تعیّن‌گرایی در پدیده‌های طبیعی و اجتماعی محدود است، بنابراین توصیف ریاضی رفتار تصادفی باید نقش عمده‌ای در توصیف جهان ایفا کند. خواه مذاق ریاضی ما مایل به نابگرایی باشد و خواه به مدلسازی، احتمال هر دو را ارضا می‌کند. با این حال در اینجا، بحث ما در بارهٔ آمار مقدماتی است و نه ریاضیات.

از دیدگاه منطق قیاسی که بخش زیادی از آموزش آمار را در گذشته شکل داده است، احتمال اساسیتر از آمار است: احتمال مدلهای شانسی را که تغییر پذیری در داده‌های مشاهده شده را توصیف می‌کنند، به‌دست می‌دهد. ولی از نظر فهم بهتر موضوع، ما معتقدیم که آمار اساسیتر از احتمال است: در حالی که تغییرپذیری داده‌ها را مستقیماً می‌توان دریافت، مدلهای شانسی را تنها بعد از آنکه آنها را در ذهنمان ساختیم می‌توان دریافت. در دنیای ریاضی آرمانی افلاطونی، می‌توانیم با یک جوجهٔ احتمالاتی شروع کنیم و از منطق قیاسی استفاده کنیم تا یک تخم‌مرغ آماری بگذارد، اما در دنیای درهم‌وبرهم علم تجربی، باید از تخم‌مرغ به‌عنوان دادهٔ مشاهده شده شروع کنیم و یک جوجهٔ احتمالاتی پیشین را به‌عنوان یک استنباط بسازیم. در درس آمار مقدماتی تنها ارزش جوجه در آن است که توضیح دهد تخم‌مرغها از کجا آمده‌اند. حداقل در این چارچوب، قدری نامنصفانه به نظر می‌رسد که از دانشجویان مبتدی بخواهیم مطالبی در بارهٔ تخم‌گذار یاد بگیرند پیش از اینکه با خود تخم‌مرغ آشنا شوند. با قدری مبالغه می‌توان گفت این کار شبیه آن است که مطالعهٔ شیمی با مکانیک کوانتومی آغاز شود.

بنابراین، جای احتمال در نخستین درس آمار کجا باید باشد؟ موضع ما استاندارد نیست، ولی بتدریج هوادارانی می‌یابد: نخستین درسها در آمار اساساً نباید به‌هیچ عنوان نظریهٔ احتمال رسمی را دربرداشته باشند.

چرا؟ اولاً به این دلیل که احتمالِ غیررسمی برای درک مفهومی استنباط کافی است. هر چند ساختار نظری استنباط آماریِ استاندارد مبتنی بر احتمال است، نقش احتمال محدود به پاسخ‌دادن به این سؤال است که «چه اتفاقی می‌افتد اگر این روش را دفعات بسیار زیادی به‌کار ببریم؟» پاسخ را توزیع نمونه‌گیری یک‌آماره به‌دست می‌دهد که الگوی تغییرپذیری برآمدهایِ، مثلاً، نمونه‌های تصادفی بسیاری را از جامعهٔ یکسان، ثبت می‌کند. اگر توافق کنیم که استخراج عملی این توزیعها بهتر است به درسهای پیشرفته‌تر موکول شود، می‌توان آنها را به‌عنوان توزیعهایی در نظر گرفت که با استفاده از ابزارهای تحلیل داده‌ها بدون ادوات احتمال رسمی به‌دست آمده‌اند. قواعد مربوط به $P(A \cup B)$ چیز بسیار کمی به یک درس آمار می‌افزاید.

دومین دلیل برای اجتناب از احتمال رسمی آن است که احتمال از نظر مفهومی مشکلترین موضوع در ریاضیات مقدماتی است. تاریخ ایده‌های احتمالاتی ([۱۶] و [۲۷] را ببینید) مجذوب‌کننده اما کمی ترس‌آور است. این موضوع مدتها برای ذهنهایی قویتر از ذهن ما فوق‌العاده گیج‌کننده بوده است. روانشناسانی که اولینِ آنها تورسکی[1] و همکارانش بوده‌اند، ثابت کرده‌اند که این سردرگمی حتی در میان آنهایی که می‌توانند اصول احتمال رسمی را از بر بگویند و می‌توانند تمرینات کتابهای درسی را انجام دهند، ادامه دارد. احساس شهودی ما از رفتار تصادفی شدیداً و به‌طور سیستماتیک، معیوب است، برای مثال [۲۸] و مجموعهٔ [۱۹] را ببینید. بدتر اینکه متخصصان آموزش ریاضی هیچ راه مؤثری برای تصحیح این شهود معیوب ما نیافته‌اند. گارفیلد و الگرن [۱۵] مروری بر یک پژوهش در این زمینه را با این اظهارنظر به پایان می‌رسانند که «تدرس مفهوم احتمال هنوز هم کار بسیار دشواری به‌نظر می‌آید، و مملو از ابهام و تصورات غلط است.» آنها پیشنهاد می‌کنند بررسی شود که «چگونه ایده‌های مفید استنباط آماری را می‌توان مستقل از احتمالِ درست از نظر فنی، تدریس کرد». به اعتقاد ما تمرکز بر ایدهٔ توزیع نمونه‌گیری اجازهٔ این کار را، حداقل در سطحی مناسب برای مبتدیان، به ما می‌دهد.

البته مفاهیم استنباط آماری، اگر هم مبتنی بر توزیعهای نمونه‌گیری عرضه شوند، دشوارند. باید توجه خود و توجه دانشجویان خود را ــ که بردباری محدودی نسبت به ایده‌های مشکل دارند ــ به ایده‌های اساسی آمار محدود کنیم. ما اعضای هیأت علمی تصور می‌کنیم که احتمال رسمی این ایده‌ها را روشن می‌کند. اما این امر تقریباً در مورد هیچ‌یک از دانشجویان ما صادق نیست.

۷.۳ در مورد دانشجویان رشتهٔ ریاضی چطور؟ دانشجویان رشتهٔ ریاضی به‌طور سنتی آمار را به‌عنوان دومین درس از یک سلسله درس دو ترمی که به نظریهٔ احتمال و آمار اختصاص دارد می‌گذرانند. امیدواریم واضح باشد که ما سیاحتی در آماره‌های بسنده، نااریبی، برآوردگرهای ماکسیمم درستنمایی، و قضیهٔ نیمن پیرسون[2] را راه نویدبخشی برای کمک به دانشجویان در فهمیدن ایده‌های اساسی آمار تلقی نمی‌کنیم. از طرف دیگر، دانشجویان رشتهٔ ریاضی حتماً باید بخشی از ساختارهای استنباط آماری را ببینند. پس چه باید کرد؟

ما ترجیح می‌دهیم که پیش از مطرح کردن نظریه، ایده‌ها و روشهای آماری و کاربردهای آنها با محور قراردادن داده‌ها به دانشجو معرفی شود. یعنی دانشجویان ریاضی لزوماً از این اصل که نخستین آشنایی با آمار نباید براساس احتمال رسمی باشد، مستثنی نیستند. اگر دانشجویان زمینه‌های محاسباتی قوی دارند، یک درس داده‌محور می‌تواند با سرعت کافی به سمت معرفی آمار واقعاً مفید و کاربردهای جدی پیش رود. نیاز به نظریه را می‌توان زمانی‌که با مسائل عملی مواجه می‌شویم، روشن کرد، و نظریه وقتی‌که وضعیت آن در عمل روشن باشد، بهتر معنا می‌یابد. با این حال در بسیاری مؤسسات، محدودیتها یا تردید اعضای هیأت علمی این مسیر را دشوار می‌سازد. در برخی دیگر از مؤسسات، هماهنگی کمی بین درسهای «کاربسته» و نظری وجود دارد، به‌طوری که درواقع دروس نظری مبتنی بر درسهای کاربردی نیستند.

1. Tversky 2. Neyman Pearson

بنابراین لازم است مجدداً این موضوع را مورد توجه قرار دهیم که یک درس یک‌ترمی در معرفی آمار برای دانشجویان ریاضی و دیگر دانشجویانی که پایهٔ محاسباتی قوی دارند چگونه باید باشد. راحت‌ترین کار این است که این درس پس از درسی در احتمال بیاید و معمولاً هم چنین است. در اینجا با مانعی دیگر مواجه هستیم: وجداناً نمی‌توانیم سلسله درس استاندارد احتمال-آمار را در هر دو ترم طوری برنامه‌ریزی کنیم که آشنایی با آمار به‌طور بهینه انجام گیرد. احتمال نه تنها به منظور آماده‌سازی برای نظریهٔ آماری بلکه فی‌نفسه هم مهم است. هر چقدر که یک گروه در برنامهٔ درسی رشتهٔ خود تأکید بیشتری بر کاربردها و مدلسازی داشته باشد، درس احتمال باید نقش اساسیتری در این تأکید ایفا کند. مقدمه‌ای بر احتمال که بر مدلسازی تأکید کند و شبیه‌سازی و محاسبات عددی را داشته باشد، مطمئناً زمینه را برای آمار آماده می‌کند، اما نسبت به انتقال ایده‌های صرفاً آماری به ترم احتمال در تردید هستیم. اصلاح احتمال و اصلاح آمار دو مقولهٔ متمایزند.

هدف ما باید یک درس یکپارچهٔ آمار باشد که به نوبت از درون تحلیل داده‌ها، تولید داده‌ها، و استنباط گذر کرده بر اصول سازمان‌دهندهٔ هر یک از آنها تأکید کند. حتماً باید از ظرفیتهای ریاضی دانشجویان سود برده آنها را تقویت کنیم. گرچه تحلیل داده‌ها و تولید داده‌ها نظریهٔ یکپارچه‌کننده‌ای ندارند، تحلیل ریاضی حتی می‌تواند به تحلیل داده‌ها وضوح بخشد. در اینجا چند مثال را ذکر می‌کنیم.

- الف: خواص بهینگی شاخصهای گرایش به مرکز را برای n مشاهده در نظر گیرید. میانگین، میانگین توان دوم خطا را مینیمم می‌کند؛ میانه، میانگین قدر مطلق خطا را مینیمم می‌کند (و لزومی ندارد یکتا باشد)؛ نیم دامنه، ماکسیمم مطلق (یا توان دوم) خطا را مینیمم می‌کند؛ سعی کنید میانهٔ مطلق خطا را برای $n = ۳$ مینیمم کنید و رفتار ناخوشایند شاخص حاصل را بررسی کنید.
- ب: دانشجویان ضمن مطالعهٔ احتمال به نابرابری چبیشف برمی‌خورند. سپس ممکن است همچنین با نابرابری جالب $|\mu - m| \leq \sigma$ که میانگین، میانه، و انحراف معیار هر توزیع را با هم مرتبط می‌کند [۲۹] روبه‌رو شوند. داده‌های یک نمونه‌ای را به کمک توزیع تجربی (احتمال $\frac{1}{n}$ در هر نقطهٔ مشاهده شده) برای استخراج نتایجی در بارهٔ اینکه میانگین و میانهٔ نمونه چقدر از هم فاصله دارند، توصیف کنید.
- پ: خط رگرسیونی کمترین توانهای دوم، مشابه میانگین $\overline{x}$ برای پیش‌بینی y از روی x است. آن را به‌دست آورید. سپس، اگر خواستید، به‌کمک نرم‌افزار در بارهٔ مشابه‌های سایر شاخصهای ذکر شده در الف کاوش کنید.

در مورد تولید داده‌ها به‌راحتی می‌توان محاسباتی احتمالاتی انجام داد که نشان دهند چقدر محتمل است که تخصیصهای تصادفی به طرق مشخصی نامتوازن باشند؛ مزایای نمونه‌های بزرگ به‌زودی روشن می‌شود.

بسیار خوب. می‌توانیم درس مقدماتی متوازنی در آمار به دانشجویان ارائه دهیم که در آن از معلومات ریاضی آنها استفاده شود. پیامد گریزناپذیر آن این است که کمتر در استنباط وقت صرف کنیم. باید تصمیم بگیریم چه چیزی را حفظ و چه چیزی را حذف کنیم. در این زمینه هنوز اجماعی حاصل نشده است، زیرا علی‌رغم غرولندهای فراوان، اصلاح سلسله دروس احتمال و آمارِ رشتهٔ ریاضی هنوز شروع نشده است. فکر کردن در بارهٔ چنین اصلاحی نقطه‌ای مناسب برای پایان دادن به بحث آمار، ریاضیات، و تدریس است. این تکلیفی است که باید در خانه انجام دهید: یک درس آمارِ یک ترمیِ بهتر برای دانشجویان رشتهٔ ریاضی طراحی کنید.

مراجع

1. Aldous. David (1994), Triangulating the circle. at random, *Amer. Math. Monthly* **101**, 223-233. The remark appears in the biographical note accompanying the paper.
2. Bjerkedal, T. (1960), Acquistion of resistance in guinea pigs infected with different doses of virulent tubercle bacilli, *American Journal of Hygiene* **72**, 130-148.
3. Boyer, Paul and Stephen Nissenbaum (1972). *Salem Village Witchcraft.* Belmont, CA: Wadsworth Publishing Co.
4. Boyer, Paul and Stephen Nissenbaum (1974). *Salem Possessed.* Cambridge, MA: Harvard University Press.
5. Bullock, James O. (1994), Literacy in the language of mathematics, *Amer. Math. Monthly* **101**, 735-743.
6. Bunker, John P., Benjamin A. Barnes. and Frederick Mosteller (eds.) (1977), *Costs. Risks and Benefits of Surgery.* New York: Oxford University Press.
7. Chambers, John M., William S. Cleveland, Beat Kleiner, and Paul A. Tukey (1983). *Graphical Methods for Data Analysis.* Belmont, CA: Wadsworth.
8. Chambers, John M. and Trevor J. Hastie (1992). *Statistical Model in* S. Pacific Grove, CA: Wadsworth.
9. Cleveland, William S. and Mary E. McGill (eds.) (1988), *Dynamic Graphics for Statistics.* Belmont, CA: Wadsworth.
10. Cobb, George W. (1991), Teaching statistics: more data, less lecturing, *Amstat News.* December 1991, pp. 1, 4.
11. Cobb, George W. (1992), Teaching statistics, in L. A. Steen (ed.) *Heeding the Call for Change: Suggestions for Curricular Action*, MAA Notes 22. Washington, DC: Mathematical Association of America.
12. Cochran, W. G. (1968), The effectiveness of adjustment by subclassification in removing bias in observational studies, *Biometrics* **24**, 205-213.
13. Crystal, David (ed.) (1994), *The Cambridge Factfinder.* Cambridge: Cambridge University Press. pp. 174-175.
14. Efron, Bradley and Rob Tibshirani (1991), Statistical data analysis in the computer age. *Science* **253**, 390-395.
15. Garfield, Joan and Andrew Ahlgren (1988), Difficulties in learning basic concepts in probability and statistics: implications for research. *Journal for Research in Mathematics Education* **19**, 44-63.

DC: Mathematical Association of America.

25. Moore, David S. (1995), *The Basic Practice of Statistics.* New York: W. H. Freeman.

26. Rosenbaum, Paul R. (1995), *Observational Studies.* New York: Springer-Verlag, p. 60.

27. Stigler, S. M. (1986), *The History of Statistics: The Measurement of Uncertainty Before 1900.* Cambridge, Mass: Belknap.

28. Tversky, Amos and Daniel Kahneman (1983), Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment, *Psychological Review* **90**, 293-315.

29. Waston, G. S. (1994), letter to the editor, *The American Statistician* **48**, p. 269. This is the last in a sequence of comments on this inequality, and contains references to the earlier contributions.

30. Weisberg, Sanford (1985), *Applied Linear Regression*, 2nd edition. New York: John Wiley and Sons, p. 230.

******

• George W. Cobb and David S. Moore, "Mathematics, statistics, and teaching", *Amer. Math. Monthly*, (9) **104** (1997) 801-823.

* جورج کاب، کالج مانت هالیوک، آمریکا

gcobb@mtholyoke.edu

* دیوید مور، دانشگاه پُردو، آمریکا

dsm@stat.purdue.edu

16. Gigerenzer, G., Z. Swijtink. T. Porter, L. Daston, J. Beatty, and L. Krüger (1989) *The Empire of Chance.* Cambridge: Cambridge University Press.

17. Hoaglin, D. C. (1992), Diagnostics. in D. C. Hoaglin and D. S. Moore (eds.), *Perspectives on Contemporary Statistics*, MAA Notes 21. Washington, DC: Mathematical Association of America, pp. 123-144.

18. Hoaglin, David C. and David S. Moore (eds.) (1992). *Perspectives on Contemporary Statistics*, MAA Notes 21. Washington, DC: Mathematical Association of America.

19. Kapadia, R. and M. Borovcnik (eds.) (1991), *Chance Encounters: Probability in Education.* Dordrecht: Kluwer.

20. Longfellow, Henry Wadsworth (1847). *Evangeline*, Introduction, 1.1.

21. Lyle, Roseann M. et al. (1987), Blood pressure and metabolic effects of calcium supplementation in normotensive white and black men, *Journal of the American Medical Association* **257**, 1772-1776. Dr. Lyle provided the data in the example.

22. Moore, David S. (1988), Should mathematicians teach statistics (with discussion). *College Math. Journal* **19**, 3-7.

23. Moore, David S. (1992), What is statistics? in David C. Hoaglin and David S. Moore (eds.), *Perspectives on Contemporary Statistics*, MAA Notes 21. Washington, DC: Mathematical Association of America, pp. 1-18.

24. Moore, David S. (1992), Teaching statistics as a respectable subject, in Florence Gordon and Sheldon Gordon (eds.), *Statistics for the Twenty-First Century*, MAA Notes 26. Washington,