آمار

راهنمایی به سوی ناشناخته‌ها



آمار

راهنمایی به‌سوی ناشناخته‌ها

این اثر، ترجمه‌ای است از ویرایش سوم کتاب
Statistics: A Guide to the Unknown

ویراستاران:
فرشته امیرفریار
دکتر رحیم شهلایی
دکتر علی عمیدی
دکتر محمدرضا فقیهی
دکتر محمدعلی کفایی
گیتی مختاری امیرمجدی
دکتر محمد قاسم وحیدی اصل

آمار، راهنمایی به‌سوی ناشناخته‌ها

گروه ویراستاران به زبان انگلیسی: جودیت ام. تانور و همکاران

ترجمه: گروه مترجمان

ویرایش ادبی: محمود عمادی

ناشر: انجمن آمار ایران

چاپ اول: مرداد ۱۳۷۹

شمارگان: ۲۲۰۰ نسخه

بها: ۲۵۰۰۰ ریال

چاپ: زحل چاپ

حق چاپ برای انجمن آمار ایران محفوظ است.

انجمن آمار ایران: تهران ـ خیابان فرصت ـ سازمان پژوهشهای علمی و صنعتی ـ شماره ۳۰۵ تلفن: ۸۸۲۱۲۲۶

به نام خداوند جان وخرد

پیشگفتار

هریک از رشته‌های متعدد علمی و شاخه‌های گوناگون معارف بشری، سهمی را برای پاسخ به پرسشهای گوناگون آدمی و رفع نیازهای مادی و معنوی او قائلند یا این انتظار از آنها در بین مردم وجود دارد. اما آمار از سویی نقش یک ابزار اساسی قدرتمند ـ مانند کامپیوتر ـ را در تقریباً تمامی شاخه‌های علوم پایه، تجربی و انسانی دارد و از سوی دیگر، نیاز به وقوف از باطن قضایا وکنه امور و لزوم پیشگویی و پیش‌بینی در مسائلی نظیر علوم پزشکی، امور جوی، رفتارهای اقتصادی و اجتماعی و... برای علم آمار جایگاهی خاص را پدید آورده‌است. به‌علاوه علم آمار این ویژگی را دارد که بنا به ماهیتش، روش و میزان دقت خود را در برخورد با هر مسئله واقعی شرح می‌دهد و باتوجه به ملموس‌بودن قریب به اتفاق مسائل پیش‌رو، این اجبار را دارد که بنا به اقتضا به قویترین ابزارهای نظری، بویژه علوم ریاضی و نظریه احتمال توسل جوید. بنابراین الزاماً مباحث مطروحه در آن پیچیدگیهای زیادی را طلب می‌کنند. درنتیجه آیا می‌توان برخی کاربردهای واقعی آمار در مسائل مختلف را به بیانی مطرح کرد که در عین طرح دقیق مسئله واقعی، از پیچیدگیهای فنی متعارف، اجتناب شود؟

نویسندگان کتاب "آمار، راهنمایی به‌سوی ناشناخته‌ها" نشان داده‌اند که این کار، شدنی است و این امر را چنان با ظرافت و هنرمندانه به اجرا گذاشته‌اند که برای آنها که با قابلیتهای علم آمار و پیچیدگیهای آن آشنایی دارند، شگفت‌آور به‌نظر می‌رسد. این کتاب، سه بار در ایالات متحده آمریکا تجدید ویرایش شده و مقصود از نوشتن آن، معرفی بخشی از تواناییهای این علم برای پاسخگویی به نیازهای انسان و رشته‌های مختلف علم به زبان ساده و غیرفنی بوده‌است. در مورد زیبایی، ظرافت و جذابیت مطالب کتاب، بیش از این حاجتی به گفتار نیست. چه؛ گفته‌اند که "مشک آن است که خود ببوید، نه آنکه عطار بگوید". اما در مورد ترجمه:

انجمن آمار ایران در حاشیه دومین همایش سال جهانی ریاضیات طی مباحثاتی در جمع

نمایندگان انجمن، مبادرت به تشکیل چهار کمیته تخصصی نمود که یکی از آنها به نام کمیته کاربرد آمار در تولید و خدمات به مسئولیت آقای دکتر محمدرضا فقیهی در جلسات نسبتاً منظمی که از آن زمان به بعد داشته‌است، در تبیین وظایف این کمیته، ترجمه کتاب حاضر را برای آشنایی عموم مردم با آمار به انجمن پیشنهاد کرد. شورای اجرایی انجمن، ضمن استقبال از این پیشنهاد، ترجمه و چاپ آن را تصویب و سرپرستی اجرای طرح را به آقای دکتر فقیهی واگذار کرد. خوشبختانه همزمانی این کار با فعالیتهای ستاد ملی سال جهانی ریاضیات و حمایت مالی آن، راه چاپ کتاب را از طرف انجمن آمار ـ به‌عنوان نخستین کتاب در سلسله انتشارات غیرادواری آن ـ هموار نمود.

انجمن آمار ایران وظیفه خود می‌داند که مراتب سپاس خود را از آقای دکتر توکل، دبیر ستاد ملی سال جهانی ریاضیات به دلیل حمایتهای مادی و معنوی از این طرح، آقای دکتر محمدرضا فقیهی برای نظارت بر ترجمه، ویرایش و چاپ کتاب؛ آقای محمود عمادی برای ویرایش ادبی و مدیریت انتشار و چاپ کتاب؛ و همه مترجمان و ویراستاران که باوجود ضیق وقت، ترجمه و ویرایش مقالات محوله را در رأس اولویت کارهای خود قرارداده و انجمن را در این راه یاری دادند، اعلام نماید. امید است این همکاری باب تعاون را برای اجرای طرحهایی از این نوع که قطعاً در ارتقای علم آمار و فرهنگ آماری در کشور، نقش مؤثر خواهدداشت، بگشاید.

در پایان، ذکر این نکته را ضروری می‌داند که با توجه به دستچین بودن نویسندگان مقالات اصلی از حیث تسلط برجنبه‌های فنی موضوعات مورد بحث در مقاله‌ها و هنرمندی شگرف آنها در بیان ظریف و نکته‌بینانه مطالب، ممکن است باوجود تلاش مترجمان و ویراستاران، در برگرداندن اندیشه‌های آنها به زبان فارسی در مواردی ناخواسته غفلتهایی صورت گرفته باشد. به این دلیل شمارگان چاپ اول تا حدودی اندک انتخاب شد؛ به این امید که تذکرات و راهنماییهای مشفقانه خوانندگان فهیم و علاقه‌مند، در رفع مشکلات احتمالی در چاپهای بعدی مؤثر و سودمند باشد و هر چاپ پیراسته‌تر از چاپ قبل به پیشگاه خوانندگان تقدیم شود.

محمدقاسم وحیدی اصل

دبیر شورای اجرایی دوره چهارم

انجمن آمار ایران

مقدمه

تهیه کتابی با هدف تشریح کاربردهای مهم آمار و احتمالات در زمینه بسیاری از فعالیتها، طرحی بود که به‌منظور کمک به اجرای آن، کمیته انجمن آمار آمریکا و شورای ملی معلمان ریاضی (NCTM[1] -ASA[2]) در اوایل سال ۱۹۶۹ ازمن دعوت به عمل آورد. دیدگاه کمیته چنین بود که اختصاص سهم بیشتری به آمار و زمینه آن در احتمالات در برنامه درسی دبیرستانها و مراکز آموزش عالی سودمند خواهدبود و بدین‌ترتیب نشان‌دادن چگونگی کاربرد گسترده این ابزارها نیز مطلوب می‌باشد. کمیته مزبور این کتاب را عمدتاً برای خوانندگانی بدون معلومات خاص در زمینه آمار، احتمالات یا ریاضیات طراحی نمود. این مخاطبان بویژه شامل والدین محصلین، مدیران و اعضای هیئت مدیره آموزشگاهها، معلمان ریاضی و سرپرستان آنها و سرانجام خود جوانان می‌شود. آمار، راهنمایی به‌سوی ناشناخته‌ها حاصل این طرح است. در زمان تهیه این کتاب تعدادی از ما که به‌طور همزمان روی این کتاب کار می‌کردیم و به تدریس آن اشتغال

۱ـ *National Council of Teachers of Mathematics*

۲ـ *American Statistical Association*

داشتیم، بسیاری از مطالب آن را برای دانشجویان دوره کارشناسی و فارغ‌التحصیلان مفید وحتی الهام‌بخش یافتیم. کتاب حاضر به‌نحوی غیرقابل انتظار کارکردی اضافی به‌عنوان یک کتاب کمک درسی یافت و اغلب بدین صورت مورد استفاده قرارگرفته‌است.

مقالات ارائه‌شده به‌جای آموزش روشهای فنی، به تشریح دستاوردهای گذشته و استفاده‌های جاری از آمار و احتمالات می‌پردازد. در هنگام انتخاب مقالات، هدف کمیته و من نشان‌دادن تنوع گسترده زمینه‌های کاربرد آمار بود. اما قصد ما این نبود که کار ناممکن زیرپوشش قراردادن کلیه موارد استفاده از آن را به انجام برسانیم. حتی در زمینه‌های مورد شمول نیز تعمداً از پوشش کامل اجتناب شد. ما نویسندگان را از نوشتن مقالاتی که می‌توانستند عنوان "کلیه موارد استفاده از آمار در...." را داشته باشند، برحذر داشتیم و درمقابل، از آنان درخواست نمودیم برروی یک یاچند مسئله مهم در زمینه کاری خود تأکید کنند و توضیح دهند که چگونه آمار و احتمالات به حل آنها کمک می‌کند و چرا حل این مسئله برای ملت، علم یا مردمی که در اصل آن را مطرح ساخته‌اند، سودمند است. درگذشته، برای افرادی که توانایی آن را نداشتند که ازعهده درک مطالب بسیار تخصصی برآیند، دست‌یابی به چنین مقالاتی دشوار بود.

این گستردگی کاربردها، به درک ما از یگانگی در تنوع؛ یعنی آمار مفهوم تازه‌ای بخشیده است. از یک طرف پی می‌بریم که فنون آماری یکسان یا مشابه در زمینه‌های بدون ارتباط با یکدیگر به‌کار گرفته می‌شوند. برخی نویسندگان استفاده از همبستگی یا رگرسیون را در موارد متفاوتی نظیر مطالعه خورشید، نمره‌دهی عادلانه آزمونها، تأثیر اخذ مالیات بر مصرف سیگار و یا بررسی آثار شدت عمل بر رانندگان متجاوز از سرعت مجاز را تشریح می‌کنند و برخی دیگر به کاربرد نظریه نمونه‌گیری در زمینه‌های بسیار متفاوت مانند حسابداری، اجرای بهتر سرشماری در ایالات متحده یا تخمین جمعیت نهنگها می‌پردازند و مقالاتی پی‌درپی، طرح آزمایش و ضرورت یا دشواری استنباط داده‌های آماری ناقص را مورد بحث قرار می‌دهند. بدون شک این یگانگی در تنوع به آشکار ساختن سودمندی گسترده ابزارهای آماری به عموم مردم کمک می‌کند.

از طرف دیگر، دریافتیم که مقالاتی با فنون متفاوت آماری می‌توانند براساس هماهنگی در موضوع گروه‌بندی شوند. برای مثال دو مقاله که هریک به نحوی با دیگری متفاوت است به

موضوعی مانند راضی کردن مشتری از کالای تولیدشده می‌پردازند. حداقل پنج مقاله روشهای بسیار متفاوت پژوهش درباره بیماریها، علل ودرمان آنها را شرح می‌دهند. این مقاله‌ها عبارتند از: آزمایش ارزش واکسن فلج اطفال، استفاده از دوقلوها درتحقیقات مربوط به بیماریها، استفاده از تصادفی‌سازی به‌منظور بررسی تأثیرات نوآوری در بیمارستانها، بررسی تأثیرات زیانبار یک نوع داروی بیهوشی و بررسی بیمه درمانی و آثار آن بر کاهش هزینه مراقبتهای بهداشتی.

هنگامی که مقالات گردآوری و ویرایش شدند، می‌باید درمورد ترتیب قرارگرفتن آنها تصمیم می‌گرفتیم. به نظر می‌رسید این ترتیب می‌تواند به چند صورت باشد. می‌توانستیم مقالات رابه تفکیک ابزارهای آماری مورداستفاده گروه‌بندی کرده و بدین‌ترتیب بر یکپارچگی ابزارهای آماری تأکید کنیم و تنوع معمول در این رشته از علوم را نادیده بگیریم. می‌توانستیم مقالات را به تفکیک روشهای جمع‌آوری اطلاعات، گروه‌بندی کنیم یعنی برحسب آمارگیریهای نمونه‌ای، آزمایش، مطالب حاصل از سرشماریها و نظایر آن طبقه‌بندی کنیم و یا براساس کاربردهای موضوعی آنها را در گروههای جداگانه قرار دهیم.

ازمیان روشهای یادشده، گزینش ما آخرین شیوه بود. مقالات را به چهار بخش بزرگ به تفکیک زمینه کاربرد آنها تقسیم نمودیم که برخی ازآنها دارای زیربخشهایی هستند. هر زیربخش به اندازه کافی کوچک، دارای انسجام وقابل فهم هست که به‌صورت یک واحد مجزا مورد مطالعه قرارگیرد و شرحی اجمالی از کاربردهای آمار را در یک زمینه محدود به دست دهد. البته ما مایل نبودیم از مزایای سایر روشهای محتمل طبقه‌بندی صرف‌نظر کنیم. بنابراین در پی فهرست اصلی مطالب، دو فهرست دیگر خواهد آمد. فهرست اول براساس روشهای جمع‌آوری اطلاعات و فهرست دوم به تفکیک ابزارهای آماری تنظیم شده‌است. در فهرست دوم، یک مقاله زیر عنوان ابزاری که نویسنده از آن استفاده نموده و یا ماگمان می‌کردیم خواننده ممکن است با نگریستن به مقاله چیزی در مورد آن تکنیک فراگیرد یا هر دومورد تنظیم شده‌است.

این تلاشها در هنگام طبقه‌بندی، باعث تأکید برجنبه‌ای از کتاب شد که ما عمداً به برنامه‌ریزی درمورد آن نپرداخته بودیم و در اوایل کار حتی از آن بی‌خبر بودیم. چنین معلوم شد

که ما گروه بزرگی از مقالاتی داشتیم که به خط‌مشی عمومی پرداخته‌بودند و با بسیاری از آنها تحت عنوان "دنیای اجتماعی ما" طبقه‌بندی شده‌بودند. همچنین دریافتیم که تعدادی از مقاله‌های مربوط به این گروه به ارزیابی اصلاحات یا تغییرات در خط‌مشی می‌پردازند. از یک طرف با شرح دو آزمایش میدانی در مقیاس بزرگ روبرو شدیم: آزمایش بیمه درمانی و آزمایشهای مربوط به واکسن فلج اطفال. به‌نظر می‌رسد تااین اواخر، در ایالات متحده به تعداد اندکی ازاین آزمایشهای کنترل‌شده دست‌زده‌ایم و نظر اعضای کمیته بر این است که یکی از وظایفی که آمارشناسان تاکنون از آن غفلت ورزیده‌اند، ارائه توضیح درمورد امکانات و ارزشهای این آزمایشها به عموم مردم است و مردم به چنین توضیحاتی نیاز دارند تا معقولانه تصمیم بگیرند که آیا خواهان انجام این آزمایشها هستند؟ از طرف دیگر، تعدادی از مقالات ارزیابی غیرآزمایشی (یاشبه آزمایشی) اصلاحات را مورد بحث قرارداده‌اند: آیا در کنتیکات، شدت عمل در مورد رانندگانی که از سرعت مجاز تخلف کرده‌اند، ازمیزان تصادفات کاسته‌است؟ آیا نوع خاصی از داروی بیهوشی خطرناک است؟ آیا مالیات بیشتر نوجوانان را از مصرف سیگار بازمی‌دارد؟

امیدواریم هردونوع این مقالات به درک بیشتر این موضوع کمک کنند که پی بردن به اینکه آیا برنامه به اهداف خودنایل شده‌است یا خیر تاچه اندازه دشوار است. درک این امر باعث می‌شود مردم همدلی بیشتری با مقامات دولت پیدا کنند که می‌کوشند وظایفی دشوار را علی‌رغم مشکلات و نارساییهای جدی انجام دهند.

ضرب‌المثلی قدیمی می‌گوید: شتر، اسبی است که اعضای بدنش توسط اعضای یک کمیته جفت‌وجور شده‌است. نویسندگان ما بخشهای تحلیلی بی‌نهایت جذابی را در چارچوبی بسیار خوش‌فرم فراهم نمودند اما تنظیم مناسب این کتاب عمدتاً مدیون کمیته‌ای بسیار بااستعداد و متعهد است. به‌طور کلی، دربررسیهای منتقدانه و ارائه پیشنهادها در مورد مقاله‌ها، رویکرد اعضای کمیته به وحدت نظر یک پدیده بود. اگرچه امکان داشت برخی اوقات درباره نقاط قوت یا ضعف یک مقاله خاص درمیان آنان دودستگی به‌وجود آید، معذلک همه آنها در هدف خود برای بوجود آوردن کتابی سودمند و درتوانایی خود برای یافتن چیزی که بتوانند بیش از بیست و چهار ساعت در شبانه‌روز روی آن کار کنند همواره متحد بودند. این ازخودگذشتگی همراه با

وسواسهای من بدون شک دشواریهایی برای نویسندگان بوجودآورده‌است. با این وصف آنها سرسختی نشان دادند و استحقاق سپاس بی‌اندازه از جانب من، کمیته و کلیه دست‌اندرکاران علم آمار را دارا هستند.

یک اثر هنری یا ادبی ممکن است جاودانه باشد اما یک اثر علمی چنین نیست. شانزده سال پیش نخستین چاپ این کتاب انتشار یافت و ده‌سال از انتشار دومین چاپ آن گذشته‌است. در این فاصله تحولات بسیاری دردنیای آمار رخ داده‌است و تلاشهای زیادی برای آشنایی مردم با آمار ازطریق کتاب و حتی برنامه‌های تلویزیونی صورت گرفته‌است. درواقع من نیز اخیراً مقدمه‌ای بر چاپ این کتاب به زبان چینی نوشته‌ام. بنابراین احساس کردیم زمان برای تهیه یک نسخه جدید و مؤثرتر کتاب مزبور فرارسیده‌است.

چاپ دوم با اضافه شدن دو مقاله جدید و ضمیمه نمودن فهرست مطالبی برای مطالعه در انتهای مقالات، نسبت به چاپ اول بسط بیشتری یافته‌است. در چاپ سوم تغییر عمده‌ای بوجود آمده‌است. نویسندگان مقالاتی راکه از چاپهای قبلی به این کتاب منتقل شده‌است بهنگام نموده‌اند وکتاب دارای ۱۲ مقاله جدید است. تمام مقاله‌ها همچنان دارای منابع مطالعاتی هستند.

ما از بنیاد *اسلوان* سپاسگزاریم که با پرداخت کمک مالی تهیه چاپ اول این کتاب را امکان‌پذیر نمود.

برای چاپ نخست سپاسگزار بسیاری افراد دیگر هم هستیم. *جرج. اف. پی باکس، لئو بریمن، چرچیل آیزنهارت، توماس هنسون، جی.دبلو. توکی و دبلیو. جی. یودن* فقید برای سختکوشی و رهنمودهایشان، دفتر انجمن آمار آمریکا بویژه *ادگار بیزگیر* و *جان لهمان* برای کمک ارزنده ایشان در کلیه کارهای اداری ضروری برای چاپ چنین کتابی و همچنین بخش اداری مجمع ملی معلمان ریاضی، *ادوارد میلمن* برای کمکهای خلاق و موشکافانه در ویراستاری و بسیاری افراد در هولدن‌دی بویژه *فردریک. اچ. مورفی*، *والتر سیزر*، *اریک لهمان*، و سردبیر خانم *هالی‌گرانو* به‌خاطر کفالت امور مربوط به ارتباطات راه دور و بسیاری از دوستان و همکاران سردبیر و اعضای کمیته که به صورت مشاورینی گمنام اما با عملکردی حیاتی انجام وظیفه نمودند.

باگذشت زمان، همچنان نسبت به این افراد احساس امتنان می‌نماییم. بویژه نسبت به فردریک مورفی در هولدن ـ دی که مدت چندین سال برتهیه این کتاب توجه داشته‌است. منابع تکمیلی مطالعاتی برای مقالات در چاپ دوم، به‌واسطه مساعدت ارزنده دیوید *لین*، *دونا وللاند* نوبرگ، ریک پرسون، *هایگانوش پریسلر* و *استر سید*، فراهم آمده‌است.

تهیه چاپ سوم بدون کمک ارزشمند *مارجوری اولسون* امکان‌پذیر نبود. بتی *پوند* نیز شایسته امتنان زیاد من برای توجه به بسیاری از جزئیات است. از کمک *جان کیمل* از وادزورث و *لیندا لوبا* از بروکس کول در فرایند تولید کتاب کمال امتنان را داریم. بار دیگر بیشترین قدردانی ما از نویسندگان مقالات است که تحت فشارهای شدید ناشی از کمبود وقت کارهایی با کیفیت بالا ارائه نمودند و به هنگام تقاضاهای مداوم ما برای تجدیدنظر بردباری نشان دادند.

امید من و امید همه اعضای هیئت سردبیری آن است که این کتاب همواره به‌عنوان یک کتاب درسی مکمل باقی بماند و همچنان رسالت خود را برای تشویق در منظور نمودن آمار و احتمالات در برنامه‌های درسی به انجام برساند.

جودیت. ام. تانور

گریت نک. نیویورک

۴ فوریه ۱۹۸۸

پیشگفتار چاپ سوم

انگیزه تهیه این کتاب ریشه در دگرگونی و پیشرفتهای مهمی در آموزش ریاضیات دارد که حاصل اقدامات آغازشده از سال ۱۹۵۴ است. زمانی که شورای ریاضیات هیئت مدیره آزمون ورودی کالج به‌منظور انجام پژوهشی مستمر در مورد برنامه درسی، یک گردهمایی از معلمان، استادان و مربیان ریاضی دبیرستانها، دانشکده‌های تربیت‌معلم، مدارس عالی و دانشگاهها ترتیب داد. تا قبل از این همایش، گروههای مختلف معلمان و استادان بندرت با همکاری یکدیگر برروی برنامه‌های درسی به‌کار پرداخته بودند. اما از آن پس تبادل افکار و عقاید در بسیاری جهات گسترش یافته و استمرار پیدا نموده‌است که یکی از نتایج بلندمدت آن تشکیل کمیته مشترک انجمن آمار آمریکا (ASA) و شورای ملی معلمان ریاضی (NCTM) برای تهیه برنامه درسی در زمینه آمار واحتمالات بوده‌است. تا اواخر سال ۱۹۶۷ چنین همکاریهایی میان دبیران دبیرستانها و استادان مؤسسات آموزش عالی رایج متداول بود. بنابراین برای *دنوان جانسون* رئیس وقت NTCM و من که در آن زمان ریاست ASA را به‌عهده داشتم، برپایی یک کمیته مشترک به‌منظور بازنگری امور مربوط به آموزش آمار و احتمالات با سهولت میسر شد.

هدف کمیته مشترک ترغیب به تدریس آمار در دبیرستانها و مراکز آموزش عالی است و انجمنهای این هدف به‌درستی توجه خود را به چنین آموزشی معطوف می‌کنند؛ زیرا آمار، بخشی از علوم ریاضی است که با بسیاری از موضوعات عملی و همچنین خاص سروکار دارد و بویژه بدین منظور نظام یافته که به ابهامات و پیچیدگیهای مربوط به زندگی و جامعه بپردازد.

برای توجیه این نکته که چرا نیاز به تدریس میزان بیشتری از آمار است، کمیته مشترک بر این باور بود که باید برای مردم روشن سازد آمارشناسان چه نوع کمکهایی می‌توانند به جامعه عرضه دارند؟

هنگام تشریح کار در زمینه علوم ریاضی، ابتدا باید این تصمیم اساسی اتخاذ شود که از خواننده انتظار می‌رود تا چه سطحی با ریاضیات آشنا باشد؟ سطحی از آشنایی با ریاضیات از خواننده موردانتظار است. اگرچه کمیته مشترک برای سازمانهای حرفه‌ای ارائه خدمت می‌کند که موضوع فعالیت آنها اساساً و عمدتاً ماهیتی ریاضی دارد؛ باوجود این، ما تصمیم گرفتیم اندیشه‌های آماری و کمکهایی را که این علم به انجام آنها قادر است، تشریح کنیم؛ بدون آنکه به شرح و بسط جنبه‌های ریاضی آنها بپردازیم. این حرکتی جسورانه بود و نویسندگان مقالات از اینکه ما در موضع خود پایدار ماندیم، شگفت‌زده شدند.

بخت، بی‌اندازه با کمیته مشترک یار بوده‌است. از اینکه بسیاری از محققین برجسته به شرکت در این طرح آموزشی مایل بوده‌اند؛ تمامی پاداش نویسندگان درواقع سهمی است که آنان در کمک به مردم برای پی‌بردن به ارزش آمار داشته‌اند. ما همچنین اقبال آن را داشته‌ایم که از زحمات جودیت *تانور* در مقام سردبیر این مجموعه و اعضای سختکوش کمیته به‌عنوان کارکنان او بهره‌مند باشیم.

کمیته مشترک به موازات نگارش این کتاب، مجموعه‌ای از جزوات را با عنوان "آموزش آمار با مثال[1]" برای تدریس در کلاس تهیه نموده‌است که بویژه برای دانش‌آموزانی طراحی‌شده که از پایه ریاضی قوی برخوردار نیستند. با استفاده از این جزوه‌ها که حاوی مثالهایی از زندگی واقعی است، می‌توان آمار را تدریس نمود. جزوه‌های دیگری نیز تهیه شده که با مثالهای عینی آغاز شود و دانش‌آموزان با این روش فنون ویژه، ابزار و مفاهیم آماری را فرامی‌گیرند. ناشر این جزوه‌ها بنگاه انتشاراتی Addison - Wesley است. کمیته مشترک تا به امروز همچنان فعال بوده‌است و ازمیان فعالیتهای اخیر آن، برنامه‌ای است با عنوان "سواد کلی[2]" که توسط بنیاد

۱ـ *Statistics by Examples*

۲ـ *National Science Foundation*

ملی علم[1] حمایت می‌شود و هدف آن آموزش معلمان است تا آموزش‌دهنده معلمان دیگر باشند.

برخی خوانندگان ممکن است مایل باشند بدانند چگونه می‌توان آمارشناس شد و برخی دیگر ممکن است مسئول راهنمایی دانش‌آموزان در مورد فرصتهای شغلی باشند. بروشور "فرصتهای مشاغل آماری[2]" اطلاعات لازم را درباره ماهیت کار، و آموزشهای لازم برای تخصصهای گوناگون آماری دراختیار متقاضیان قرار می‌دهد.

آنچه که اقدام ما را برای تهیه این مجموعه امکان‌پذیر ساخت، اعطای کمکهای بلاعوض از بنیاد اسلوان در آغاز کار بود. در طول سالیان، همچنین از همکاری، توجه و کمکهای مالی بنیاد راسل سیج و شورای تحقیقات علوم اجتماعی بهره‌مند بوده‌ایم. ادارات ملی ASA و NCTM نیز در هنگام تهیه سه نوبت چاپ این کتاب، ما را یاری کرده‌اند. ناشر اصلی، فرد مورفی از شرکت هولدن ـ دی چاپ اول و دوم کتاب را به‌صورتی جالب عرضه نمود و *جان کیمل* از بنگاه انتشاراتی وادزورث و بروکز/کول ادامه‌دهنده روش قبل برای چاپ سوم بود. *مارجری اولسون* مسئولیت امور مربوط به آماده‌سازی نسخه اصلی چاپ سوم را قبل از انتشار به‌عهده داشته و *کلئو یوتز* نیز برای آماده‌کردن آن مساعدت نموده‌است.

و اما سخن آخر: بر این باور نیستیم که وظیفه توضیح درباره آمار فقط درانحصار ماست. ما از دیگران نیز درخواست می‌کنیم دیدگاههای خود را درباره اهداف و روشهای این علم و نتایج حاصل ارائه کنند. باخرسندی پی‌برده‌ایم برخی چنین کرده‌اند.

فردریک ماستلر
رئیس شورای سردبیری آمار، راهنمایی به‌سوی ناشناخته‌ها
کمبریج، ماساچوست
۱۴ فوریه ۱۹۸۸

۱ـ *Quantitative Literacy*

۲ـ *Careers in Statistics*

فهرستها

مقالات رده‌بندی شده برحسب منابع داده‌ها ظ
مقالات رده‌بندی شده برحسب ابزارهای آماری ف

بخش اول
دنیای زیستی ما

سلامتی و بیماری

بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴ ۳

با کمک آمار، طرح آزمایشی آگاهانه‌ای برای اولین حمله به فلج اطفال تهیه می‌شود.

پل مه‌یر ***مترجم: دکتر زهرا موسوی جاهد***

ایمنی داروهای هوشبر ۲۱

برای انجام تحقیقی در سطح کشور، به‌منظور مقایسه ایمنی داروهای هوشبر که کاربرد گسترده‌ای در جراحی دارند، از نمونه‌گیری و استانداردسازی نرخها استفاده می‌شود.

لینکلن ای. موزس و فردریک موستلر ***مترجم: دکتر علی عمیدی***

آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران ۳۵

تصادفی‌سازی مستمر بیماران در واحدهای درمانی بیمارستان عمومی متروپولیتن کلیولند، ارزیابی سریع و قابل اعتمادی را از کوششهای انجام‌شده برای بهبود مراقبتها و کاهش هزینه‌ها امکان‌پذیر می‌سازد.

دانکن نیوهوزر ***مترجم: دکتر رحیم شهلا یی***

آزمایش بیمه درمانی ۴۵

یک آزمایش در مقیاس وسیع نشان می‌دهد که مشارکت بیمار در پرداخت مخارج بیمه درمانی موجب کاهش هزینه‌های مربوط به مراقبتهای پزشکی می‌شود، بدون آنکه به‌طور جدی از سلامت افراد بکاهد.

جوزف پی. نیوهاوس **مترجم: دکتر محمدرضا مشکانی**

قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف ۶۱

اقتصاددانان با استفاده از تحلیل رگرسیونی میزان تقاضای سیگار و آثار کاهش مالیات را بر سلامت عمومی تخمین می‌زنند و کنگره را متقاعد می‌نمایند با عدم کاهش مالیات، احتمالاً بسیاری از نوجوانان را از مصرف سیگار بازدارد.

کنت ای. وارنر **مترجم: دکتر اسدالله منجمی**

مردم و حیوانات

آیا وراثت در ایجاد بیماری مؤثر است؟ ۷۹

مقایسه دوقلوهای حاصل‌شده از یک یا دو تخمک نشان می‌دهد که چه بیماریهایی از وراثت تأثیر می‌پذیرند.

دی. دی. رید **مترجم: دکتر مقتدا هاشمی‌پرست**

وضع اسفناک نهنگها ۸۹

روشهای مربوط به تخمین تعداد جمعیت نهنگها به وضع مقررات برای حفاظت آنها کمک می‌کند.

دی. جی. چپمن **مترجم: دکتر علی عمیدی**

اهمیت انسان‌بودن ۱۰۱

استفاده از یک فن آماری به اتخاذ تصمیم در مورد تعلق قطعه‌ای استخوان فسیل‌شده به یک انسان یا میمون انسان‌نما و نیز تعیین قدمت نژاد انسانی کمک می‌کند.

دبلیو. دبلیو. هاولز **مترجم: دکتر رحیم شهلایی**

بخش دوم

دنیای سیاسی ما

برهان آماری تبعیض در استخدام ۱۱۹

روشهای آماری، ارزیابی کاربستهای استخدامی را برای گروههای تحت حفاظت بدون توجه به نیت تبعیض، توسط دادگاهها امکان‌پذیر می‌سازد.

سندی ال. زابل — **مترجم: بهرام راستین**

آمار در انتخاب هیئت منصفه ۱۳۳

بررسیهای آماری و تجزیه و تحلیل آنها به وکلا در انتخاب اعضای هیئت منصفه برای دادگاههای مدنی و جنایی کمک می‌کند.

اس. جیمز پرس — **مترجم: گیتی مختاری امیرمجدی**

اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی ۱۴۳

داده‌های جمع‌آوری شده قبل و پس از اصلاحات اجتماعی به ارزیابی آثار آن کمک می‌کنند.

دونالد تی. کمپبل — **مترجم: دکتر ناصررضا ارقامی**

شب انتخابات در تلویزیون ۱۵۹

پیشگویی صحیح درمورد برندگان، در شب انتخابات مستلزم شیوه‌هایی برای جمع‌آوری سریع اطلاعات و یک مدل ویژه آماری برای کنترل خطاست.

ریچارد اف. لینک — **مترجم: گیتی مختاری امیرمجدی**

بخش سوم

دنیای اجتماعی ما

برقراری ارتباط با دیگران

تعیین نویسنده ۱۷۵

نقش یک تحلیل آماری برای تصمیم‌گیری درمورد سئوالات تاریخی مطرح، با بررسی ۱۲ مقاله مندرج در نشریه فدرالیست که نگارش آن همزمان به الکساندر هامیلتون و جیمز مدیسون نسبت داده می‌شود.

فردریک موستلر و دیوید ال. والاس مترجم: **بهرام راستین**

یادآوری اطلاعات تصویری در کودکان ۱۹۳

یک آزمایش بدقت طراحی‌شده نشان می‌دهد که کودکان به‌طور قابل ملاحظه‌ای توانایی دارند تا تصویرهایی که قبلاً دیده‌اند را به خاطر بیاورند.

دوریس. آر آنت ویسل و دبلیو. اچ. هاگینز مترجم: **کرامت‌الله پروین جهرمی**

معنی واژه‌ها ۲۰۳

استفاده از قضاوت سوژه‌های آزمایشی در یک روش آماری برای تهیه نقشه‌ای که معنی واژه‌ها را در ابعاد گوناگون مشخص می‌کند.

جوزف ب. کروسکال مترجم: **گیتی مختاری امیرمجدی**

اندازه چیزها ۲۱۷

روشی ویژه برای ترسیم نمودار چیزهایی مانند اندازه شهرها و تناوب تکرار واژه‌ها در یک متن که منحنیهایی تولید می‌کند که شکلهایی بسیار شبیه به هم دارند.

هربرت ای. سیمون مترجم: **علی‌محمد حسینی**

مردم در عرصه‌های شغلی

حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟ ۲۳۱

آزمون روشهای نمونه‌گیری، شرکتهای راه‌آهن و خطوط هوایی را متقاعد نمودع به‌جای استفاده از تمام‌شماری برای تعیین سهم درآمد خود، از حمل بار و خدمات مسافری، از نمومه‌گیری استفاده کنند.

جان نتر — **مترجم: دکتر محمدعلی کفایی**

ارزیابی مقدماتی از یک فراورده غذایی جدید ۲۴۵

آمارگیریها و آزمایشها میزان خوش‌طعمی و پذیرش یک محصول جدید را مشخص می‌کند.

الیزابت استریت و ماویس بی. کارول — **دکتر علی عمیدی**

تعیین تحملات عددی رنگ با روشهای آماری ۲۵۷

تحلیل تشخیصی، قضاوت بصری را به اندازه‌گیری ابزاری تفاوت رنگها مرتبط می‌سازد و درنتیجه تولیدکنندگان محصولات قادر به ایجاد روشهایی مؤثر برای ایجاد سازگاری میان رنگها خواهند بود.

لونی سی. ونس — **مترجم: دکتر خلیل شفیعی**

مردم، مدرسه و بازی

منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار ۲۶۷

در یک آزمایش تعدادی مقاله بین افرادی که باید به آنها نمره بدهند، توزیع می‌شود تا اختلافهای میان آنان برای نمره‌دهی عادلانه‌تر در طول زمان تعدیل شود.

هنری آی. براون و هوارد واینر — **مترجم: دکتر احمد پارسیان و دکتر علی رجالی**

آمار، ورزش و چند چیز دیگر ۲۸۳

محاسبات انجام‌شده با استفاده از احتمالات روشن می‌سازد که در ورزش بیس‌بال در چه زمان بانتها و گامهای عمدی تدابیری سودمند است.

رابرت هوک — **مترجم: دکتر رحیم شهلایی**

شمارش مردم و اموال آنان

شاخص بهای کالاها و خدمات مصرفی ۳۰۱

شاخص قیمت مصرف‌کننده چگونه تنظیم و مورد استفاده قرار می‌گیرد؟

فیلیپ جی. مک‌کارتی — مترجم: دکتر اسدالله منجمی

چگونه بهتر بشماریم؟ استفاده از آمار برای اصلاح سرشماری ۳۱۵

روشهای آمارگیری نمونه‌ای باعث ارتقای صحت یافته‌های حاصل از سرشماری در ایالات متحده شده‌است.

موریس. اچ. هنسن و باربارا. ای. بیلر — مترجم: دکتر علی عمیدی

برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟ ۳۳۱

دفتر سرشماری و دفتر آمارهای کار، از طریق دو آمارگیری در مقیاسی گسترده تعداد افراد شاغل و بیکار را هرماه تعیین می‌کنند.

کارول بوید لئون و فیلیپ ال. رونس — مترجم: دکتر محمدرضا فقیهی

بسط و تحلیل شاخصهای اقتصادی ۳۴۳

برای پیشگویی و سنجش تحرک اقتصادی، کارشناسان به ترکیب اطلاعات از صدها شاخص اقتصادی دست می‌زنند.

جفری اچ. مور — مترجم: دکتر محمدعلی کفایی

بخش چهارم
دنیای مادی ما

بهینه‌سازی و مسئله فروشنده دوره‌گرد ۳۶۱

روشهای تصادفی در تحقیق برای دست‌یابی به ترتیبی صحیح جهت تنظیم تلسکوپ به‌طرف زنجیره‌ای از ستارگان، انتخاب کوتاهترین مسیر برای رسیدن به یک شهر و یافتن مکانهایی برای کاوش گاز طبیعی.

چارلز ای. ویتنی — مترجم: دکتر محمدرضا فقیهی

برآوردکردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن ومدل‌بندی آماری ۳۷۵

روش تاریخ‌گذاری به‌وسیله رادیوکربن برای تخمین ناشناخته‌ها، تأیید فرضیه‌ها و تخمین احتمال وقوع زلزله با استفاده از آمار.

دیوید آر. برلینجر مترجم: **دکتر محمدقاسم وحیدی اصل**

پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار ۳۹۳

با ترکیب اندازه‌های بسیاری از متغیرها و استفاده از تحلیل رگرسیونی، هواشناسان قادر به پیشگویی بهتر میدان دید برای دوره‌های زمانی کوتاه هستند که برای خلبانان هلی‌کوپتر اهمیت دارد.

رابرت جی. میلر مترجم: **گیتی مختاری و دکتر محمود طاهری**

آمار، خورشید و ستارگان ۴۰۳

همبستگی میان اندازه‌گیریهای درخشندگی خورشید به نظریه‌ای درباره رفتار درونی آن منجر شده‌است.

چارلز ای. ویتنی مترجم: **دکتر علی عمیدی**

مقالات رده‌بندی شده برحسب منابع داده‌ها

نمونه‌ها

درباره افراد و تجمع آنها

پرس: *آمار در انتخاب هیئت منصفه* ۱۳۳

چپمن: *وضع اسفناک نهنگها* ۸۹

لئون و رونس: *برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟* ۳۳۱

لینک: *شب انتخابات در تلویزیون* ۱۵۹

مک‌کارتی: *شاخص بهای کالاها و خدمات مصرفی* ۳۰۱

نیوهاوس: *آزمایش بیمه درمانی* ۴۵

وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱

هنسن و بیلر: *چگونه بهتر بشماریم؟ استفاده از آمار برای اصلاح سرشماری* ۳۱۵

درباره گزارشها و دنیای مادی

برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵

میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳

موستلر و والاس: تعیین نویسنده ۱۷۵
نتر: حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟ ۲۳۱
ویتنی: بهینه‌سازی و مسئله فروشنده دوره‌گرد ۳۶۱
ویتنی: آمار، خورشید و ستارگان ۴۰۳

داده‌های دردسترس

زابل: برهان آماری تبعیض در استخدام ۱۱۹
سیمون: اندازه چیزها ۲۱۷
کمپبل: اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی ۱۴۳
میلر: پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار ۳۹۳
مور: بسط و تحلیل شاخصهای اقتصادی ۳۴۳
موزس و موستلر: ایمنی داروهای هوشبر ۲۱
موستلر و والاس: تعیین نویسنده ۱۷۵
وارنر: قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف ۶۱
هوک: آمار، ورزش و چند چیز دیگر ۲۸۳

آمارگیریها و پرسشنامه‌ها

پرس: آمار در انتخاب هیئت منصفه ۱۳۳
رید: آیا وراثت در ایجاد بیماری مؤثر است؟ ۷۹
لئون و رونس: برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟ ۳۳۱
مک‌کارتی: شاخص بهای کالاها و خدمات مصرفی ۳۰۱
وارنر: قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف ۶۱
هنسن و بیلر: چگونه بهتر بشماریم؟ استفاده از آمار برای اصلاح سرشماری ۳۱۵

آزمایشها

آنت‌ویسل و هاگینز: *یادآوری اطلاعات تصویری در کودکان* ۱۹۳
استریت و کارول: *ارزیابی مقدماتی از یک فراورده غذایی جدید* ۲۴۵
براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷
کروسکال: *معنی واژه‌ها* ۲۰۳
مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
نیوهوزر: *آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران* ۳۵
نیوهاوس: *آزمایش بیمه درمانی* ۴۵

شبه‌آزمایشها

رید: *آیا وراثت در ایجاد بیماری مؤثر است؟* ۷۹
کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳
ونس: *تعیین تحملات عددی رنگ با روشهای آماری* ۲۵۷
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱

بررسیهای مشاهداتی

برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵
زابل: *برهان آماری تبعیض در استخدام* ۱۱۹
موزس و موستلر: *ایمنی داروهای هوشبر* ۲۱
موستلر و والاس: *تعیین نویسنده* ۱۷۵
ویتنی: *آمار، خورشید و ستارگان* ۴۰۳

مقالات رده‌بندی شده برحسب ابزارهای آماری

برآورد کردن، آزمون فرض، تحلیل بیزی، تحلیل تجربی بیزی و تحلیل داده‌ها

برآورد کردن

استریت و کارول: *ارزیابی مقدماتی از یک فراورده غذایی جدید* ۲۴۵

براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷

برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵

رید: *آیا وراثت در ایجاد بیماری مؤثر است؟* ۷۹

کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳

لئون و رونس: *برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟* ۳۳۱

مک‌کارتی: *شاخص بهای کالاها و خدمات مصرفی* ۳۰۱

میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳

نتر: *حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟* ۲۳۱

نیوهوزر: *آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران* ۳۵

نیوهاوس: *آزمایش بیمه درمانی* ۴۵

وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱

هنسن و بیلر: *چگونه بهتر بشماریم؟ استفاده از آمار برای اصلاح سرشماری* ۳۵

آزمون فرض

آنت‌ویسل و هاگینز: *یادآوری اطلاعات تصویری در کودکان* ۱۹۳
براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷
زابل: *برهان آماری تبعیض در استخدام* ۱۱۹
مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
موزس و موستلر: *ایمنی داروهای هوشبر* ۲۱
موستلر و والاس: *تعیین نویسنده* ۱۷۵
نیوهوزر: *آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران* ۳۵
نیوهاوس: *آزمایش بیمه درمانی* ۴۵
هاولز: *اهمیت انسان بودن* ۱۰۱

تحلیل بیزی و تحلیل تجربی بیزی

موستلر و والاس: *تعیین نویسنده* ۱۷۵

تحلیل داده‌ها

سیمون: *اندازه چیزها* ۲۱۷
کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳
کروسکال: *معنی واژه‌ها* ۲۰۳
مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳

مور: *بسط و تحلیل شاخصهای اقتصادی* ۳۴۳
موزس و موستلر: *ایمنی داروهای هوشبر* ۲۱
موستلر و والاس: *تعیین نویسنده* ۱۷۵
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱
ویتنی: *آمار، خورشید و ستارگان* ۴۰۳
هوک: *آمار، ورزش و چند چیز دیگر* ۲۸۳

جدولها، نمودارها و نقشه‌ها

جدولها

مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳
مور: *بسط و تحلیل شاخصهای اقتصادی* ۳۴۳
موزس و موستلر: *ایمنی داروهای هوشبر* ۲۱
موستلر و والاس: *تعیین نویسنده* ۱۷۵
هوک: *آمار، ورزش و چند چیز دیگر* ۲۸۳

نمودارها و نقشه‌ها

برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵
سیمون: *اندازه چیزها* ۲۱۷
کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳

کروسکال: *معنی واژه‌ها* ۲۰۳
مور: *بسط و تحلیل شاخصهای اقتصادی* ۳۴۳
ونس: *تعیین تحملات عددی رنگ با روشهای آماری* ۲۵۷
ویتنی: *بهینه‌سازی و مسئله فروشنده دوره‌گرد* ۳۶۱

درصدها و نرخها، استانداردسازی و تعدیل

درصدها و نرخها

آنت‌ویسل و هاگینز: *یادآوری اطلاعات تصویری در کودکان* ۱۹۳
پرس: *آمار در انتخاب هیئت منصفه* ۱۳۳
رید: *آیا وراثت در ایجاد بیماری مؤثر است؟* ۷۹
زابل: *برهان آماری تبعیض در استخدام* ۱۱۹
کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳
لئون و رونس: *برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟* ۳۳۱
مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳
مور: *بسط و تحلیل شاخصهای اقتصادی* ۳۴۳
موزس و موستلر: *ایمنی داروهای هوشبر* ۲۱
موستلر و والاس: *تعیین نویسنده* ۱۷۵
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱
هوک: *آمار، ورزش و چند چیز دیگر* ۲۸۳

استانداردسازی و تعدیل

براون و واینر: منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار ۲۶۷
لینک: شب انتخابات در تلویزیون ۱۵۹
موزس و موستلر: ایمنی داروهای هوشبر ۲۱
ونس: تعیین تحملات عددی رنگ با روشهای آماری ۲۵۷

سریهای زمانی و شاخص‌سازی

کمپبل: اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی ۱۴۳
لئون و رونس: برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟ ۳۳۱
مک‌کارتی: شاخص بهای کالاها و خدمات مصرفی ۳۰۱
مور: بسط و تحلیل شاخصهای اقتصادی ۳۴۳
ویتنی: آمار، خورشید و ستارگان ۴۰۳

احتمالات و مدلسازی

احتمالات

آنت‌ویسل و هاگینز: یادآوری اطلاعات تصویری در کودکان ۱۹۳
رید: آیا وراثت در ایجاد بیماری مؤثر است؟ ۷۹
زابل: برهان آماری تبعیض در استخدام ۱۱۹
سیمون: اندازه چیزها ۲۱۷
لینک: شب انتخابات در تلویزیون ۱۵۹
میلر: پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار ۳۹۳
موستلر و والاس: تعیین نویسنده ۱۷۵
نتر: حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟ ۲۳۱
ویتنی: بهینه‌سازی و مسئله فروشنده دوره‌گرد ۳۶۱

هوک: *آمار، ورزش و چند چیز دیگر* ۲۸۳

مدلسازی

برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵
چپمن: *وضع اسفناک نهنگها* ۸۹
رید: *آیا وراثت در ایجاد بیماری مؤثر است؟* ۷۹
سیمون: *اندازه چیزها* ۲۱۷
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱
ویتنی: *آمار، خورشید و ستارگان* ۴۰۳
ویتنی: *بهینه‌سازی و مسئله فروشنده دوره‌گرد* ۳۶۱
هاولز: *اهمیت انسان بودن* ۱۰۱

نمونه‌گیری و تصادفی‌سازی

نمونه‌گیری

استریت و کارول: *ارزیابی مقدماتی از یک فراورده غذایی جدید* ۲۴۵
براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷
برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵
پرس: *آمار در انتخاب هیئت منصفه* ۱۳۳
چپمن: *وضع اسفناک نهنگها* ۸۹
زابل: *برهان آماری تبعیض در استخدام* ۱۱۹
لئون و رونس: *برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟* ۳۳۱

لینک: *شب انتخابات در تلویزیون* ۱۵۹
مک‌کارتی: *شاخص بهای کالاها و خدمات مصرفی* ۳۰۱
موزس و موستلر: *ایمنی داروهای هوشبر* ۲۱
نتر: *حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟* ۲۳۱
ویتنی: *بهینه‌سازی و مسئله فروشنده دوره‌گرد* ۳۶۱
هنسن و بیلر: *چگونه بهتر بشماریم؟ استفاده از آمار برای اصلاح سرشماری* ۳۱۵

تصادفی‌سازی

آنت‌ویسل و هاگینز: *یادآوری اطلاعات تصویری در کودکان* ۱۹۳
براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷
مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
نیوهوزر: *آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران* ۳۵
نیوهاوس: *آزمایش بیمه درمانی* ۴۵
ویتنی: *بهینه‌سازی و مسئله فروشنده دوره‌گرد* ۳۶۱

همبستگی و رگرسیون

استریت و کارول: *ارزیابی مقدماتی از یک فراورده غذایی جدید* ۲۴۵
براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷
سیمون: *اندازه چیزها* ۲۱۷
کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳
لینک: *شب انتخابات در تلویزیون* ۱۵۹
میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳

مور: *بسط و تحلیل شاخصهای اقتصادی* ۳۴۳
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱
ویتنی: *آمار، خورشید و ستارگان* ۴۰۳
هاولز: *اهمیت انسان بودن* ۱۰۱

مقیاس‌بندی چندبعدی و تحلیل تشخیصی

مقیاس‌بندی چندبعدی

کروسکال: *معنی واژه‌ها* ۲۰۳

تحلیل تشخیصی

موستلر و والاس: *تعیین نویسنده* ۱۷۵
میلر: *پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳
ونس: *تعیین تحملات عددی رنگ با روشهای آماری* ۲۵۷
هاولز: *اهمیت انسان بودن* ۱۰۱

آزمونها و اندازه‌گیریها

براون و واینر: *منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار* ۲۶۷
آنت‌ویسل و هاگینز: *یادآوری اطلاعات تصویری در کودکان* ۱۹۳
استریت و کارول: *ارزیابی مقدماتی از یک فراورده غذایی جدید* ۲۴۵
ونس: *تعیین تحملات عددی رنگ با روشهای آماری* ۲۵۷

پیش‌بینی و پیشگویی

برلینجر: *برآورد کردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری* ۳۷۵
چپمن: *وضع اسفناک نهنگها* ۸۹
لینک: *شب انتخابات در تلویزیون* ۱۵۹
میلر: *پیش‌بینی بسیارکوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳
مور: *بسط و تحلیل شاخصهای اقتصادی* ۳۴۳
موستلر و والاس: *تعیین نویسنده* ۱۷۵
نیوهاوس: *آزمایش بیمه درمانی* ۴۵
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱

تصمیم‌گیری

پرس: *آمار در انتخاب هیئت منصفه* ۱۳۳
زابل: *برهان آماری تبعیض در استخدام* ۱۱۹
کمپبل: *اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی* ۱۴۳
مه‌یر: *بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴* ۳
میلر: *پیش‌بینی بسیارکوتاه‌مدت وضع هوا با مشاهدات خودکار* ۳۹۳
موستلر و والاس: *تعیین نویسنده* ۱۷۵
نیوهوزر: *آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران* ۳۵
نیوهاوس: *آزمایش بیمه درمانی* ۴۵
ونس: *تعیین تحملات عددی رنگ با روشهای آماری* ۲۵۷
وارنر: *قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف* ۶۱

ویتنی: *بهینه‌سازی و مسئله فروشنده دوره‌گرد* ۳۶۱
هوک: *آمار، ورزش و چند چیز دیگر* ۲۸۳
هاولز: *اهمیت انسان بودن* ۱۰۱

بخش اول

دنیای زیستی ما

سلامتی و بیماری

بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی واکسن ضدفلج اطفال سالک در سال ۱۹۵۴ ۳

ایمنی داروهای هوشبر ۲۱

آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران ۳۵

آزمایش بیمه درمانی ۴۵

قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف ۶۱

مردم و حیوانات

آیا وراثت در ایجاد بیماری مؤثر است؟ ۷۹

وضع اسفناک نهنگها ۸۹

اهمیت انسان‌بودن ۱۰۱



بزرگترین آزمایش بهداشت عمومی که تاکنون انجام شده: آزمایه میدانی[۱] واکسن ضد فلج اطفال سالک[۲] در سال ۱۹۵۴

پل مه یر[۳] دانشگاه شیکاگو

بزرگترین آزمایش پزشکی که تا دهه ۱۹۸۰ پرهزینه‌ترین آزمایش پزشکی در تاریخ بود، در سال ۱۹۵۴ انجام شد. متجاوز از یک میلیون کودک خردسال در این آزمایش شرکت داشتند و هزینه‌های مستقیم فوری آن متجاوز از ۵ میلیون دلار به ارزش دلار در نیمه قرن بود. این آزمایش

۱- *Field trial*

۲- *Salk*

۳- *Paul Meier*

از آن‌رو انجام شد که اگر واکسن سالک تأثیری داشت میزان آن تأثیر را در مصونیت در مقابل فلج یا مرگ ناشی از بیماری فلج اطفال ارزیابی کند. مطالعه از بسیاری جهات دقیق بود و بیشترین دقت در استفاده از کنترلهای دارونما[۱] (کودکانی که با محلول ساده نمک مایه‌کوبی شده بودند) به‌کار برده شد که به‌طور تصادفی تجویز می‌شد (یعنی طی فرایند شانسی که با دقت تمام به‌کار برده می‌شد و احتمال دریافت واکسن یا محلول نمک برای هر داوطلب برابر بود) و مورد ارزیابی دوسوکور[۲] قرار می‌گرفت. (یعنی ترتیبی که درآن، نه کودکان و نه پزشکانی که وضع سلامتی آنها را پس از تلقیح ارزیابی می‌کردند، نمی‌دانستند به کدام یک واکسن و به کدام یک محلول نمکی داده شده‌است).

چرا چنین دقت عملی ضرورت داشت؟ آیا واقعاً دانشی که در نتیجه این نوع مطالعه کسب می‌شد، بیشتر یا بهتر از آن بود که می‌توانست از بررسیهای به‌مراتب ساده‌تر به‌دست آید؟ اینها سئوالاتی است که بحث حاضر بر آن متمرکز شده‌است.

سابقه

فلج اطفال هرگز بیماری شایعی نبوده اما به‌طور قطع یکی از هولناکترین و از بسیاری جهات یکی از غیرقابل توضیح‌ترین بیماریها از نظر رفتاری بوده‌است.

این بیماری سخت‌ترین ضربه را به کودکان خردسال وارد کرد و اگرچه مسئول تنها ۶ درصد مرگ و میر در گروه سنی ۵ تا ۹ سال در اوایل دهه ۱۹۵۰ بود، تعداد زیادی معلول فلج به‌جای گذاشت که بعضی از آنها فقط می‌توانستند با دستگاه تنفس مصنوعی زنده بمانند. این بیماری در موجهای همه‌گیر ظاهر می‌شد و به تابستانهایی می‌انجامید که طی آن بعضی جوامع خود را مجبور می‌دیدند که استخرهای شنا را تعطیل و اجتماعات عمومی را محدود کنند. چون تعداد موارد بیماری هفته به هفته افزایش پیدا می‌کرد. جوامع دیگری که یک سال از این بیماری همه‌گیر در امان مانده بودند با نگرانی منتظر سالی بودند که نوبت آنان فرا می‌رسید. درست یا نادرست، این ترکیب حمله‌های انتخابی به بی‌دفاع‌ترین گروه سنی و بی‌نظمیهای غیرقابل

۱ـ *Placebo*

۲ـ *Double- blind*

توضیح در رفتار همه‌گیری آن منجر به نگرانیهایی درباره فلج اطفال به عنوان یکی از علل مرگ و میر شد که به‌مراتب بیشتر از نگرانی درباره سایر علل مرگ ومیر از قبیل تصادفات رانندگی بود، (که بسیار فراوانتر و از بعضی جهات برای کنترل توسط جوامع مساعدتر است).

تصمیم برای اقدام به یک تلاش پژوهشی عمده به‌منظور ریشه‌کن کردن فلج اطفال تا حد زیادی از ابتلای پرزیدنت *فرانکلین*. د. *روزولت* ناشی شد؛ هنگامی که سیاستمدار جوان موفقی بود با بیماری فلج اطفال از پای افتاد. عزم او برای غلبه بر معلولیت ناشی از فلج و تعهد *بازیل اوکونور*[1] همکار سابق حقوقی او به نبرد علیه فلج اطفال موجب گردید که مقدار زیادی توجه، تلاش و پول برای مراقبت و توانبخشی قربانیان فلج اطفال و درنهایت مهمتر از همه برای انجام تحقیق برروی علتها و پیشگیری بیماری، صرف شود.

در جریان این تحقیق کشف شد که عامل بیماری فلج اطفال یک ویروس است، و سه نوع اصلی از ویروس در ابتلا به بیماری دخالت دارند. اگرچه تظاهرات بالینی بیماری نادر هستند ولی معلوم شد که خود ویروس نادر نبوده و متداول است و اغلب بزرگسالان زمانی در طول زندگی خود بدون آنکه آگاه باشند، نوعی عفونت فلج اطفال را تجربه کرده‌اند.

این یافته کمک کرد تا وضعیت ویژه‌ای که طی آن به‌نظر می‌رسید همه‌گیریهای فلج اطفال سخت‌ترین ضربه‌ها را به کسانی وارد آورده‌است که از نظر بهداشتی بهترین وضعیت را داشتند (یعنی کسانی که بهترین تغذیه و مطلوبترین شرایط مسکن را داشتند و ظاهراً از هرلحاظ در مناسبترین شرایط به‌سر می‌بردند) توضیح داده شود. درحقیقت در کشورهایی که دارای بهداشت خیلی ضعیفی هستند، به‌نظر می‌رسید که بیماری واقعاً ناشناخته باشد. توضیح اینکه چون مقدار زیادی ویروس فلج اطفال در جمعیتهای دارای بهداشت ضعیفتر وجود دارد، تقریباً هر کودک در ابتدای زندگی هنگامی که هنوز از مصونیت مادر خود بهره‌مند است در تماس با بیماری قرار می‌گیرد. درنتیجه همه تحت شرایط محافظت شده فلج اطفال قرار گرفته و به این ترتیب مصونیت پیداکرده‌اند.

در بیماری فلج اطفال هم مانند بسیاری از بیماریهای ویروسی دیگر، فردی که به ویروس

۱ـ *Basil O´ Connor*

آلوده شده و بهبود یافته‌است، معمولاً نسبت به حمله بعدی (حداقل نسبت به همان سویه ویروسی) ایمن است. علت این امر آن است که بدن هنگام نبرد با عفونت، در برابر پادگن (قسمت پروتئینی ویروس فلج اطفال) پادتنهایی (بخشی از گاماگلوبولین خون) می‌سازد. این پادتنها سالها در جریان خون باقی می‌مانند و حتی هنگامی که سطح آنها کاهش می‌یابد تا حدی که بندرت قابل اندازه‌گیری باشند، برای پیشگیری حمله شدید همان ویروس کافی هستند.

آبله و آنفولانزا دو روش مختلف را برای تهیه یک واکسن مؤثر نشان می‌دهند. در مورد آبله که مدتهاست توسط واکسن کنترل شده‌است از یک ویروس یعنی ویروس آبله گاوی که وابستگی نزدیکی با ویروس آبله دارد، برای واکسن استفاده می‌کنیم که به‌طور معمول قادر به ایجاد بیماری جدی در انسان نیست اما باعث پیدایش پادتنهایی می‌شود که در برابر آبله نیز فرد را محافظت می‌کنند. (در تعداد معدودی از افراد این واکسن قادر به ایجاد واکنش شدید و بندرت کشنده می‌باشد. اما خطر بسیار ناچیز است، به‌طوری که قبل از اینکه بیماری آبله مغلوب شود هیچ تردیدی به خود راه نمی‌دادیم که تمام دانش‌آموزان مدارس خود را در معرض این خطر قرار دهیم تا در برابر آبله مصونیت پیدا کنند.) اما در مورد آنفلوآنزا، به‌جای استفاده از یک ویروس زنده با وابستگی نزدیک، برای واکسن از محلولی از خود ویروس آنفلوآنزا استفاده شده که از ویروسی که با درمان توسط فرمالدئید کشته شده‌است تهیه می‌شود. اگر مدت درمان با فرمالدئید زیاد طولانی نباشد، ویروس مرده هنوز فعالیت پادگنی کافی برای تولید پادتنهای مورد نیاز را خواهد داشت به‌طوری که در ضمن اینکه دیگر نمی‌تواند موجب ابتلا به بیماری شود، به قدر کافی با ویروس زنده مشابهت دارد که بتواند واکسن رضایتبخشی باشد.

برای بیماری فلج اطفال هر دو روش مورد بررسی قرار گرفته‌است. امیدواریم در افراد مایه‌کوبی شده واکسن ویروسی زنده مزیت تکثیر را داشته باشد که واکنش قوی ایجاد نماید تا این واکنش سطح بالایی از پادتنهای بادوام را تولید کند. اما ممکن است با استفاده از چنین واکسنی این خطر وجود داشته باشد که ویروس موجود در واکسن که تا این اندازه شبیه به ویروس مهلک فلج اطفال است، بتواند به شکل مهلکی جهش کند و خود عامل بیماری فلج‌کننده یا کشنده شود. واکسن ویروسی کشته‌شده باید مطمئن باشد زیرا چنین واکسنی احتمالاً نمی‌تواند مبتلا کند، اما ممکن است قادر به ایجاد یک پاسخ پادتنی کافی نباشد. این

مسئله و سایر مشکلات بر سر راه ایجاد یک واکسن موفق قرار داشتند. بعضی از تجربه‌های ناگوار قبلی نیز در نگرش احتیاط‌آمیز پژوهشگران سهیم بود. در دهه ۱۹۳۰ کوششهایی برای توسعه واکسنهای ضد فلج اطفال به‌عمل آمده‌بود. دو نوع ازاین واکسنها عملاً برای مدتی مورد استفاده قرارگرفتند. شواهدی که نشان داد حداقل یکی از این واکسنها موجب ابتلا به پولیوی فلج‌کننده بوده‌است، فوراً سبب شد که هردو واکسن بسرعت از رده خارج شوند. این تجربه به طورعمیق در ذهن پژوهشگران فلج‌اطفال اثرگذاشت وهیچ تمایلی برای تکرارخطر نشان‌ندادند.

پژوهش برای تولید واکسنهای زنده وکشته شده در اواخر دهه ۱۹۴۰ با توسعه تکنیک کشت سلول برای پرورش ویروس فلج اطفال برانگیخته شد. کسانی که با فرآورده‌های زنده کار می‌کردند، در اثر پرورش چند نسل از سویه‌های مهلک در محیط کشت سلولی مناسب سویه‌های بی‌ضرری را پرورش دادند. البته نگرانی فراوانی وجود داشت که مبادا این سویه‌ها هنگامی که به‌عنوان واکسن برای انسان مصرف می‌شوند به بیماریزایی خود برگشته و موجب فلج یا مرگ گردند. (اکنون روشن است که سویه‌های پرورش‌یافته واقعاً بی‌خطر هستند، یک واکسن ویروسی زنده که به‌صورت خوراکی مصرف می‌شود، واکسنی است که در سراسر دنیا به‌طور وسیعی مورد استفاده قرار می‌گیرد.)

کسانی که با فرآورده‌های کشته شده کار می‌کردند، نظیر *جوناس سالک*، با مشکل عمل آوردن ویروس (بافرمالدئید) مواجه بودند تا به قدرکافی عفونت‌زائی آن را از بین ببرند، اما نه به اندازه‌ای که اثر پادگنی آن زایل شود. مشکلات این کار بیش از انتظار بود و ثابت شد که بعضی از مقادیر اولیه واکسن حاوی ویروس زنده بودند که می‌توانست عامل فلج و مرگ باشد. حکایت ایمنی دارای جنبه‌های آماری است (مریس ۱۹۵۷) اما اینجا نگرانی ما درباره سنجش اثربخشی واکسن است.

سنجش تأثیرگذاری

در اوایل دهه ۱۹۵۰ همایش کمیته مشورتی که توسط بنیاد ملی فلج اطفال (NFIP) تشکیل شده‌بود به‌این نتیجه رسید که واکسن ویروسی کشته‌شده که به‌وسیله *جوناس سالک* در دانشگاه پیتسبورگ تهیه شده هم بی‌خطر است و هم قدرت القای سطوح بالای پادتن را در کودکان مورد

آزمایش دارد. این مطلب، واکسن مزبور را عامل نویدبخشی برای مصرف عموم ساخت ولی هنوز باید ثابت می‌شد که این واکسن واقعاً از ابتلا به فلج اطفال در افراد تحت تماس با ویروس جلوگیری می‌کند. دور از عدل و انصاف بود که چنین واکسنی بدون دلیل متقاعدکننده بر تأثیر آن برای مصرف عموم توزیع شود. بنابراین تصمیم گرفته شد که «آزمایه میدانی» در مقیاسی بزرگ انجام پذیرد.

واضح است که آزمایه باید در مقیاس بسیار بزرگ انجام شود. چون فرض کنید می‌خواستیم که اگر واکسن واقعاً ۵۰٪ مؤثر بود، آزمایش را متقاعدکننده بدانیم (به دلایل مختلف تأثیر ۱۰۰ درصد را نمی‌توان انتظار داشت). فرض کنید درحین آزمایش نسبت وقوع فلج اطفال حدود ۵۰ در ۱۰۰۰۰۰ باشد (که تقریباً برابر با متوسط وقوع بیماری دردهه ۱۹۵۰ در ایالات متحده بود.) با درنظرگرفتن ۴۰ هزار نفر در گروه کنترل و ۴۰۰۰۰ نفر در گروه مایه‌کوبی شده درمی‌یابیم که حدود ۲۰ نفر از افراد کنترل و ۱۰ نفر از افراد مایه‌کوبی شده مبتلا می‌شوند و اختلافی با این اندازه می‌تواند به‌راحتی به تغییرات تصادفی نسبت داده‌شود. بنابراین نتیجه می‌گیریم که واکسن ممکن است مؤثر باشد لیکن متقاعدکننده نیست. با ۱۰۰۰۰۰ نفر در هر گروه تعداد مبتلایان مورد انتظار در حدود ۵۰ و ۲۵ مورد خواهد بود و چنین نتیجه‌ای متقاعدکننده است. در عمل، مسلماً بررسی وسیعتری مورد نیاز بود. زیرا به‌دست آوردن نتایج قطعی درکمترین زمان ممکن اهمیت زیادی داشت و اگر موارد ابتلا به فلج اطفال در منطقه آزمون نسبتاً کم بود، تعداد مورد انتظار ابتلا به بیماری می‌توانست به‌راحتی کمتر از ۵۰ باشد. همچنین به دلایلی که بعداً بحث خواهیم کرد، محتمل به‌نظر می‌رسید که پولیوی فلجی نسبت به همه انواع پولیو معیار بهتری از بیماری است و تنها در حدود نیمی از موارد تشخیص داده‌شده از نوع «فلجی» طبقه‌بندی می‌شوند. بنابراین وقوع نسبتاً کم بیماری وتغییرات وسیع آن از جایی به جای دیگر و از زمانی به زمان دیگر ایجاب می‌کرد که آزمایه بر روی تعداد کثیری از افراد انجام پذیرد، و همان طورکه معلوم شد باید این تعداد متجاوز از یک میلیون نفر باشد.

روش آمارهای حیاتی

بسیاری از درمانهای جدید و واکسنها، از جمله برخی از مؤثرترین آنها نظیر واکسن آبله، به این

دلیل معرفی شده‌اند که بررسیهای مقدماتی ارزش آنها را نشان داده‌است. استفاده در مقیاس بزرگ شواهد واضحی برای کارایی آنها بوده‌است. روش ساده و طبیعی برای ارزیابی واکسن سالک می‌توانست توزیع وسیع آن تا حد ممکن، از طریق مدارس باشد تا معلوم شود که آیا نسبت موارد گزارش شده فلج اطفال به‌طور قابل ملاحظه‌ای کمتر از میزان متداول طی فصل مورد مطالعه بوده‌است؟ به طریق دیگر، این توزیع می‌توانست به یک یا چند منطقه محدود شود زیرا محدودیت ذخیره موجود مانع از پوشش مؤثر تمام کشور بود. حتی شانس نسبتاً خوبی وجود دارد که فردی واکسن مؤثری را در برابر سرماخوردگی معمولی آزمایش کند و مدارک متقاعدکننده‌ای از این راه به‌دست آید.

در مورد فلج اطفال ـ و درحقیقت دراغلب موارد ـ این چنین روشهای ساده‌ای تقریباً به‌طور یقین نمی‌تواند مدرک قاطعی بوجود آورد. اول و قبل از هرچیز باید میزان تغییرات وقوع فلج را از فصلی به فصل دیگر بررسی کنیم. حتی در مورد تعدیل آن نیز هیچگونه اقدامی انجام ندهیم. با توجه به شکل ۱ که وقوع سالانه گزارش شده این بیماری را از سال ۱۹۳۰ تا ۱۹۵۵ نشان می‌دهد، مشاهده می‌کنیم که اگر آزمایه‌ای به این طریق درسال ۱۹۳۱ انجام شده‌بود، کاهش موارد از سال ۱۹۳۱ به ۱۹۳۲ به‌خوبی تأثیر شدید واکسن را نشان می‌داد. زیرا وقوع موارد به کمتر از یک سوم سطح قبلی آن کاهش پیدا کرده‌بود. سوءتعبیرهای مشابهی می‌توانست در سالهای ۱۹۳۵، ۱۹۳۷ و سایرسالها ـ برای مثال، در سال ۱۹۵۲ ـ به‌عمل آید (در مورد مسئله کلی استنباط از این قبیل داده‌های سریهای زمانی ن.ک. مقاله کمپبل) ممکن است تصور کنیم که این قبیل خطاها با استفاده از واکسن در یک منطقه مانند ایالت نیویورک و مقایسه نسبت وقوع بیماری در آنجا با منطقه‌ای که مایه‌کوبی نشده، مثلاً ایلی‌نوی قابل اجتناب است. متأسفانه یک مورد همه‌گیری فلج اطفال ممکن است در شیکاگو در فصلی روی دهد که در نیویورک تعداد موارد وقوع بیماری بسیار پایین بوده‌است (همانگونه که در سال ۱۹۵۶ رخ داد).

مسئله دیگر که ظریفتر، حساستر ولی به همان میزان طاقت‌فرساتر است به بی‌نظمیهای تشخیص و گزارش‌دهی مربوط می‌شود. البته در تشخیص موارد فلجهای تنفسی کلاسیک مشکلی وجود ندارد، اما اکثریت قاطع حالتها کمتر مشخصند. تب و ضعف علائم عادی بسیاری از بیماریها از جمله فلج است و ناظرین مختلف ضعف و فلج خفیف موقتی را هریک به طرز

متفاوتی از یکدیگر تمیز می‌دهند. بنابراین ممکن است تصمیم‌گیری برای تشخیص پولیوی غیرفلجی از سایر بیماریها تحت تأثیر اطلاعات عمومی پزشک یا احساس او در مورد میزان شیوع بیماری فلج اطفال در آن مکان و زمان قرارگیرد.

این مشکلات از طریق تنظیم معیارهای بسیار دقیق برای تشخیص می‌تواند تاحدودی خفیفتر شود، اما وقتی در آنچه پزشک انتظار دارد بیابد، دگرگونی چشمگیری دیده می‌شود، از بین بردن کامل آنها عملاً غیرممکن است. مانند موردی که پس از ارائه یک واکسن جدید ممکن است روی دهد. این حالت بخصوص هنگامی مصداق پیدا می‌کند که تشخیص اولیه باید توسط پزشکان خانواده باشد که استفاده ازمجموعه خاص معیارهای تشخیص از جمله فلج را نمی‌توان به آنها القا کرد. البته سنجشهای بعدی متخصصین نمی‌تواند مواردی که بیماری از ابتدا چیزی غیر از فلج را تشخیص داده‌است به تصویر بکشد.

روش کنترل مشاهده

اشکالات رهیافت آمارهای حیاتی توسط تمامی کسانی که به آنها مربوط می‌شود، شناخته شده است، و اگرچه طرح مطالعه اولیه آن کاملاً رضایتبخش تلقی نمی‌شود، بسیاری از مشکلات از طریق معرفی یک گروه کنترل که از نظر خصوصیات شبیه به گروه مایه‌کوبی شده‌بودند رفع شده‌است. مخصوصاً نظر این بود که واکسن به همه بچه هایی که در سال دوم دبستانهای مشارکت‌کننده در طرح بودند، تلقیح شود و آزمایش فلج را نه‌تنها در این بچه‌ها بلکه بچه‌های کلاس اول و سوم هم پیگیری کنند. بدین‌گونه، بچه‌های سال دومی مایه‌کوبی شده گروه مورد بررسی و بچه‌های سالهای اول و سوم گروه کنترل را تشکیل می‌دادند. این طرح را **روش کنترل مشاهده** می‌نامیم.

بدیهی است که این طرح از اغلب مشکلاتی که در بالا نام بردیم جلوگیری می‌کند. هرسه کلاس از یک منطقه جغرافیایی انتخاب شده‌اند، به‌طوری که تأثیر همه‌گیری در کلاس دوم مدرسه موردنظر قطعاً در کلاس اول و سوم هم به همانگونه است. البته همه افراد در یک زمان مورد مطالعه قرار می‌گیرند. طبیعتاً افراد این سه کلاس در سنین متفاوت هستند و ابتلا به فلج اطفال با سن فرق می‌کند ولی انتظار تغییرات چندانی را از یک کلاس به کلاس دیگر نداشتیم.

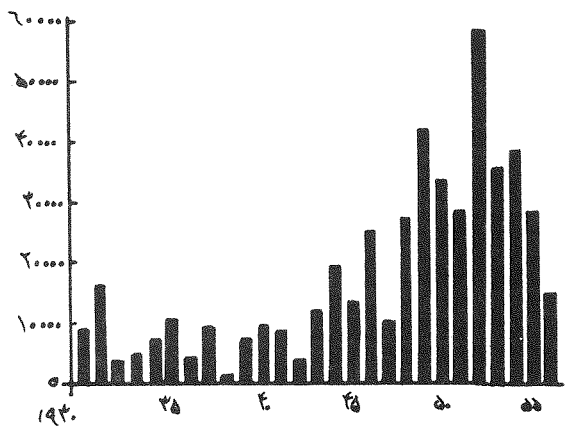
بنابراین منطقی به نظر رسید که فرض کنیم متوسط کلاسهای اول و سوم کنترل خوبی برای کلاس دوم خواهدبود.

برخلاف جذابیت نسبی این طرح و پذیرش آن توسط کمیته مشورتی NFIP اعتراضهای شدیدی از جانب برخی بخشهای بهداشتی که انتظار می‌رفت مشارکت کنند، مطرح شد. به قضاوت آنها، نتایج چنین مطالعه‌ای به دو دلیل مهم احتمال داشت به قدر کافی متقاعدکننده نباشد. یکی عدم حتمیت در فرایند تشخیص که قبلاً از آن یاد کردیم و آمادگی آن برای تأثیرپذیری از طریق انتظارات پزشکان است، و دیگری تأثیر انتخابی استفاده از داوطلبان است.

براساس طرح مطالعه پیشنهادی، پزشکان مناطق مورد بررسی از این حقیقت آگاه بودند که تنها به کلاس دومی‌ها واکسن تلقیح شده و در تشخیص بیماری هریک از این گروه کودکان به طرزی طبیعی و مناسب پرس‌وجو می‌کردند که آیا بچه‌ها واکسینه شده‌اند یا خیر.

هرگونه تمایلی برای تصمیم‌گیری در مورد یک تشخیص مشکل در حمایت از ابتلا به عارضه‌ای غیر از فلج، وقتی که معلوم است کودک مایه‌کوبی شده‌است منجر به یک فقره مدرک مجعول به حمایت از واکسن می‌شود. قضاوت قاطعانه در این مورد که آیا این چنین تأثیری واقعاً درکار بوده تقریباً غیرممکن است و نتایج درصورت مطلوب بودن برای همیشه در هاله‌ای از عدم حتمیت

شکل ۱ـ فلج اطفال در ایالات متحده، ۱۹۳۰ تا ۱۹۵۶



قرار می‌گیرند.

مشکلی که کمتر می‌توان حدس زد در اختلاف بین خانواده‌هایی است که بچه‌های خود را برای شرکت در چنین آزمایشی داوطلب می‌کنند و آنها که اصلاً داوطلب نمی‌شوند. اصلاً تعجب‌آور نیست که بعدها دریافتند، آنهاکه داوطلب می‌شوند باسوادترند و عموماً از آنهایی که شرکت نمی‌کنند، مرفه‌تر هستند و همچنین شواهدی وجود داشت مبنی بر اینکه آنها که موافق به شرکت در طرح مایه‌کوبی بودند، به‌طور قابل توجهی با فراوانی بیشتری نسبت به دیگران از مدرسه غیبت می‌کردند. جهت تأثیر چنین انتخابی بر روی وقوع پولیوی تشخیص داده‌شده در مقابل واقعیت به‌هیچ وجه روشن نیست و این تفاوت مهم میان گروه مایه‌کوبی شده و گروه کنترل نیز تفسیر نتایج را مبهم می‌سازد.

روش تصادفی کردن و کنترل دارونما

موضعگیری منتقدان طرح NFIP این بود که موضوع تأثیر واکسن به‌مراتب مهمتر از آن است که به نحوی بررسی شود که در اذهان مشاهده‌گرهای منطقی، عدم حتمیتهایی به‌جای گذارد. بدون شک اگر واکسن فوق‌العاده مؤثر به‌نظر برسد، اکثر کارکنان بهداشت عمومی و عامه مردم آن را خواهند پذیرفت. گرچه ممکن است از ابراز این مطلب خودداری کنند. اما اگر طرح کنترل مشاهده مورد استفاده قرارگیرد، تعدادی از دانشمندان صلاحیتدار بهداشت عمومی متقاعد نخواهند شد و ارزش واکسن مطمئن نخواهد بود. بنابراین منتقدان پیشنهاد کردند که مطالعه به‌عنوان یک آزمایش علمی با استفاده از روشهای تصادفی‌سازی مناسب جهت تعیین افراد مورد درمان یا کنترل و با حداکثر کوشش برای رفع اریبی مشاهده‌گر اجرا شود. این طرح آنچه را که ما **روش کنترل دارونما** می‌نامیم دنبال می‌کند.

اعتراض عمده به این طرح این بود که از والدین دانش‌آموزان به‌طور معقول نمی‌توان توقع داشت که به بچه‌هایشان اجازه دهند تا در آزمایشی شرکت کنند که ممکن بود تنها یک محلول نمک بی‌اثر به‌جای واکسنی که احتمال داشت مؤثر باشد به آنها داده‌شود. بحث بعدی این بود که تزریق دارونما از نظر اخلاقی ممکن است درست نباشد. زیرا تزریق این ماده با خطر جزیی همراه است. بویژه اگر کودک نادانسته به فلج اطفال مبتلا باشد.

طرفداران روش کنترل دارونما معتقد بودند که اگر درست عمل شود، والدین با شرکت بچه‌های خود در چنین آزمایشی موافقت خواهند کرد و براین نظر بودند که چون تزریق در طول فصل فلج انجام نمی پذیرد، خطر همراه با تزریق دارونما به‌قدری کم می‌شود که به حساب نمی‌آید. برخی از بخشهای بهداشتی سخت مقاومت کردند؛ تنها درصورتی که آزمایش به‌خوبی طراحی شده‌باشد در آن شرکت می‌کردند. نتیجه این بود که به‌طور تقریبی در نیمی از مناطق روش کنترل دارونمای تصادفی شده مورداستفاده قرار گرفت و در بقیه مناطق از روش کنترل مشاهده کلاسهای متناوب استفاده شد.

کوشش عظیمی برای ازبین بردن هرگونه امکانی در مورد آلودگی نتایج کنترل دارونما از طریق اریبیهای پیچیده مشاهده‌گر، به‌عمل آمد. تنها روش قاطع برای رسیدن به این منظور این بود که نه بچه‌ها، نه والدین و نه کارکنان مسئول تشخیص بیماری تا مادامی که همه تصمیمهای تشخیص انجام شده‌باشد، ندانند که کدام بچه‌ها مایه‌کوبی شده‌اند. روش رسیدن به نتیجه مزبور چنین بود که مواد دارونمایی تهیه شود که درست مانند واکسن به‌نظر آید اما هیچگونه فعالیت پادگنی نداشته باشد، به‌طوری که کنترلها تلقیح می‌شدند اما از هر جهت به همان روش که برای مایه‌کوبی شده‌ها صورت می‌گرفت با آنها رفتار می‌شد.

هر ظرف مایع تزریقی فقط به‌وسیله یک شماره کد، شناسایی می‌شد. به‌طوری که هیچیک از کسانی که مشغول واکسیناسیون یا فرایند سنجش تشخیص بودند نمی‌توانستند بدانند که کدام بچه واکسن گرفته‌است. چون هیچکس این مطلب را نمی‌دانست پس هیچکس نمی‌توانست به‌طور متفاوت در تشخیص افراد مایه‌کوبی شده و کنترل تأثیر گذارد. آزمایشی را که در آن هم آزمایش شونده مورد درمان و هم تشخیص دهنده‌هایی که نتایج آزمایش را ارزیابی می‌کنند از نوع درمانی که به هر فرد داده می‌شود بی‌اطلاع باشند آزمایش **دوسوکور** می‌نامند. تجربه در تحقیقات بالینی نشان داده‌است که هرگاه ارزیابی نهایی تا حدودی به‌نظر افراد بستگی داشته باشد، آزمایشهای دوسوکور تنها راه رضایتبخش برای اجتناب از اریبی جدی مشاهده‌گر است.

برای اغلب ما تکان‌دهنده است که بگویند پزشکان حاذق و متعهد باید بی‌اطلاع نگهداشته شوند مبادا که قضاوت آنها تحت تأثیر اطلاع از وضعیت درمان قرار گیرد. باید به خاطر داشته باشیم که تحریف عمدی یافته‌ها توسط پزشکان باعث نگرانی خاطر آزمایشگران طب نیست.

بلکه در بسیاری موارد مشکل حاد اتخاذ تصمیم غیرحتمی است که تجربه نشان داده‌است که باعث می‌شود بهترین آزمایشگران اندکی تحت تأثیر اطلاعاتی از این دست قرارگیرند. برای مثال، در بررسی داروهایی که برای تسکین دردهای پس از عمل مورد استفاده قرار می‌گیرند، دریافته‌اند که ب،دست آوردن یک قضاوت بدون اریبی از کیفیت تسکین درد، حتی ازمحققین باصلاحیت تقریباً غیرممکن است، مگر اینکه داور از اینکه به کدام یک از بیمارها دارو داده شده بی‌اطلاع نگهداشته شود.

دومین ویژگی اصلی روش آزمایشی اختصاص دادن افراد به درمانها با استفاده از شیوه تصادفی کردن دقیق است. همانگونه که قبلاً مشاهده کردیم، شانس رسیدن به یک مورد تشخیص داده شده فلج با عوامل زیادی از جمله سن، وضعیت اجتماعی ـ اقتصادی و مانند آن تغییر می‌کند. اگر بنا بود به‌طور عمدی کوشش کنیم که گروه مورد درمان و گروه کنترل تا حد ممکن با هم جور شوند، مجبور می‌شدیم دقت کنیم تا میان این عوامل و بسیاری از عوامل دیگر تعادل برقرار کنیم و حتی دراین صورت نیز ممکن بود بعضی ازعوامل مهم را ازدست بدهیم. بنابراین، شاید حیرت‌آور باشد که حفظ تعادل را به تعدادی پرتاب سکه دقیقاً برابر واگذار کنیم: ترتیبی می‌دهیم که هر فرد دارای شانس مساوی برای گرفتن واکسن یا دارونما باشد اما نظر خود را مطلقاً در تصمیم‌گیری در مورد افراد دخالت نمی‌دهیم و موضوع را به شانس واگذار می‌کنیم.

انجام این کار دو فایده دارد. یکی اینکه مکانیسم شانس معمولاً در همسان‌سازی تمامی متغیرها خوب عمل می‌کند (چه آنهاکه ازقبل می‌شناسیم و چه آنهاکه نمی‌شناسیم). ثانیاً اگر از مکانیسم شانس در تعیین درمان استفاده کنیم، ممکن است درباره استفاده از نظریه شانس (یعنی نظریه احتمال) برای قضاوت در مورد نتایج مطمئن باشیم. در این صورت می‌توان این احتمال را محاسبه کرد که تفاوتی به بزرگی تفاوت مشاهده‌شده می‌تواند به‌طور منطقی تنها ناشی از روشی باشد که در آن افراد برای درمان انتخاب شده‌اند، یابرعکس، این واقعاً تأثیر ناشی از اختلاف واقعی درمانها است.

مطمئناً وضعیتهایی هست که در آن یک آزمایشگر ماهر می‌تواند گروهها را به‌طرزی مؤثرتر از شیوه‌ای که انتخاب تصادفی انجام می‌دهد، متعادل کند. وقتی بعضی از عوامل ممکن است تأثیر زیادی بر روی نتایج یک آزمایش داشته‌باشند، ممکن است مطلوب یا حتی لازم باشد که از

طرح آزمایشی پیچیده‌تری که همه این عوامل را به‌حساب می‌آورد استفاده شود. اما اگر بخواهیم از نظریه احتمال برای راهنمایی خود در اظهارنظر برروی نتایج استفاده کنیم تنها درصورتی می‌توانیم از دقت نتیجه‌گیری خود مطمئن شویم که در طرح آزمایشی از تصادفی کردن در سطح مناسبی استفاده کرده‌باشیم.

تصمیمات نهایی درمورد تشخیص فلج اطفال به ترتیب زیر گرفته شده‌است: همه موارد بیماریهای شبیه به فلج که توسط پزشکان محلی گزارش شد در معرض آزمایش مخصوصی قرار گرفتند و گزارش سابقه، علائم و یافته‌های آزمایشگاهی تهیه شد. سپس یک گروه مخصوص تشخیص بیماری هر مورد را ارزیابی و به‌عنوان غیرفلجی، مشکوک به فلج ویا فلج قطعی طبقه‌بندی کرد. آخرین گروه نیز به زیرگروه‌های غیرفلجی وفلجی، تقسیم شد که فلجی خود به فلج کشنده، و غیرکشنده تقسیم شد. تنها پس از تکمیل این فرایند بود که کد شکسته شد و هر موردی که واکسن یا دارونما تزریق شده‌بود، شناخته شد.

نتایج آزمایه

نتایج اصلی در جدول ۱ نشان می‌دهد که نمایانگر اندازه جامعه آماری مورد مطالعه، تعداد مواردی که به عنوان فلج اطفال طبقه‌بندی شده‌اند ونسبت بیماری یعنی تعداد موارد در هر ۱۰۰۰۰۰ نفر جمعیت است. برای مثال، سطر دوم نشان می‌دهد که در مناطق کنترل دارونما ۴۲۸ مورد گزارش شده وجود دارد، که از بین آنها ۳۵۸ مورد به عنوان فلج اطفال تأیید شد، از آن میان ۲۷۰ مورد به‌عنوان فلج‌کننده طبقه‌بندی شد (شامل ۴ موردی که کشنده بود). ردیفهای سوم و چهارم مداخل متناظر را برای آنها که به ترتیب مایه‌کوبی شده‌اند و آنها که دارونما دریافت داشته‌اند، نشان می‌دهد. در کنار هریک از این اعداد نسبت مربوط به آن قرار دارد. با استفاده از ساده‌ترین اندازه ـ یعنی همه موارد گزارش شده ـ مشاهده می‌شود که نسبت مربوط به گروه مایه‌کوبی شده نصف نسبت مربوط به گروه کنترل (با نسبتهای قرار داده شده در مربع در جدول ۱ مقایسه کنید) برای مناطق کنترل دارونماست. این تفاوت زیادتر از آن است که به‌طور منطقی براساس محاسبه مناسب احتمالات به شانس نسبت داده شود. تأثیر آشکار واکسن همچنانکه ازموارد گزارش شده به موارد فلجی و موارد مهلک حرکت می‌کنیم، بیشتر مشخص

جدول ۱ـ خلاصه موارد مطالعه برحسب طبقه تشخیص بیماری و وضعیت واکسیناسیون (نسبت در ۱۰۰۰۰۰)

گروه مورد مطالعه	جامعه آماری مورد مطالعه	تمامی موارد گزارش‌شده		جمع		فلجی		غیرفلجی		فلج اطفال مهلک		غیر فلج اطفال	
		تعداد	نسبت	تعداد	نسبت	تعداد	نسبت	تعداد	نسبت	تعداد	نسبت	تعداد	نسبت
موارد ابتلا به بیماری فلج اطفال													
تمام مناطق: کل	۱۸۲۹۹۱۶	۱۰۱۳	۵۵	۸۶۳	۴۷	۶۸۵	۳۷	۱۷۸	۱۰	۱۵	۱	۱۵۰	۸
مناطق کنترل دارونما: کل	۷۴۹۲۳۶	۴۲۸	۵۷	۳۵۸	۴۸	۲۷۰	۳۶	۸۸	۱۲	۴	۱	۷۰	۹
واکسن‌زده	۲۰۰۷۴۵	۸۲	۴۱	۵۷	۲۸	۳۳	۱۶	۲۴	۱۲	–	–	۲۵	۱۲
دارونما	۲۰۱۲۲۹	۱۶۲	۸۱	۱۴۲	۷۱	۱۱۵	۵۷	۲۷	۱۳	۴	۲	۲۰	۱۰
تلقیح‌نشده[1]	۳۳۸۷۷۸	۱۸۲	۵۴	۱۵۷	۴۶	۱۲۱	۳۶	۳۶	۱۱	–	–	۲۵	۷
واکسیناسیون ناقص	۸۴۸۴	۲	۲۴	۲	۲۴	۱	۱۲	۱	۱۲	–	–	–	–
مناطق کنترل مشاهده	۱۰۸۰۶۸۰	۵۸۵	۵۴	۵۰۵	۴۷	۴۱۵	۳۸	۹۰	۸	۱۱	۱	۸۰	۷
واکسن‌زده	۲۲۱۹۹۸	۷۶	۳۴	۵۶	۲۵	۳۸	۱۷	۱۸	۸	–	–	۲۰	۹
کنترلها[2]	۷۲۵۱۷۳	۴۳۹	۶۱	۳۹۱	۵۴	۳۳۰	۴۶	۶۱	۸	۱۱	۲	۴۸	۶
تلقیح‌نشده کلاس دومی	۱۲۳۶۰۵	۶۶	۵۳	۵۴	۴۴	۴۳	۳۵	۱۱	۹	–	–	۱۲	۱۰
واکسیناسیون ناقص	۹۹۰۴	۴	۴۰	۴	۴۰	۴	۴۰	–	–	–	–	–	–

۱ـ *شامل ۸۵۷۷ کودک که یک یا دوبار تزریق دارونما دریافت کردند.*

۲ـ *کل جمعیت کلاس اولی و کلاس سومی.*

مأخذ: اقتباس از تی. فرانسیس. جونیور، ۱۹۵۵، جدول ۲و۳.

می‌شود، اما تعداد کم است و عاقلانه نیست که این تأثیر ظاهراً بسیار زیاد را در محافظت در مقابل موارد مهلک خیلی بزرگ کنیم. نکته اصلی این است که واکسن موفق بود. واکسن تأثیری در جلوگیری از فلج شدید از خود نشان داد که برای تضمین معرفی آن به عنوان یک شیوه استاندارد بهداشت عمومی کافی بود.

تعجب‌آور نیست که در منطقه کنترل مورد مشاهده نتایج حاصل به‌طور کلی با آنچه که در منطقه کنترل دارونما یافته شد سازگار اما تأثیر داوطلبین که قبلاً بحث شد، کاملاً روشن است (توجه: نسبت برای آنها که تلقیح نشده‌اند با نسبتهای مربوط به کنترلها درهر دو منطقه متفاوت است). اگر فقط اطلاعات کنترل مشاهده به تنهایی در دسترس قرار می‌گرفت درباره تفسیر صحیح نتایج تردید قابل ملاحظه‌ای باقی می‌ماند.

گرچه قبلاً اختلاف عقیده فراوانی درباره لزوم یا مطلوبیت طرح کنترل دارونما وجود داشت، این روش توانست پس از این واقعه رضایت زیادی حاصل کند. اگرچه اختلاف بین دوگروه اساسی و قطعی بود اما آنقدر زیاد نبود که در صورت فقدان کنترلهای دارونما از شک و تردید جلوگیری کند. در حقیقت نکات تعجب‌آور زیادی در داده‌های تفصیلی وجود داشت. برای مثال می‌دانستیم که مقادیری از واکسن دارای قدرت پادگنی بیشتری نسبت به دیگران است و ممکن بود چنین تصور شود که باید تأثیر حفاظتی بیشتری ازخود نشان دهد ولی چنین نبود. مقادیری از واکسن که از نظر قدرت پادگنی ضعیفتر تصور می‌شد درست مانند واکسن قوی عمل کرد. موضوع تعجب‌آور دیگر این بود که موارد ظاهراً مشخص پولیوی فلجی که در آزمونهای آزمایشگاهی تایید نشد فراوانی زیادی داشت. باوجوداین، هیچیک از نکات حیرت‌آور از نوعی نبود که موجب تردید جدی درباره نتیجه اصلی شود. بلافاصله پس از اعلام نتایج متخصصترین کسانی که درباره فلج اطفال تحقیق می‌کردند واکنش مطلوب خود را ابراز داشتند. خاطرنشان گردید که با انجام اینگونه مطالعات قبل ازمعرفی واکسن حقایقی رادر مورد واکسن سالک داشته‌ایم که درباره واکسن تیفوئید و سل پس از چندین دهه بعد از استفاده از آنها هنوز نداریم.

سخن آخر

گزارش یک سابقه بدون نقص و موفق برای واکسن سالک به‌دنبال یک ارزیابی ماهرانه و

موفقیت‌آمیز از تأثیر آن خشنودکننده خواهدبود. اما واقع‌بینانه‌تر این است که قبول کنیم چنین موفقیتی تنها یک قدم در جهت تداوم توسعه علوم بهداشت عمومی است. واکسن سالک اگرچه یک پیروزی قابل ذکر در نبرد علیه بیماری است اما به‌طور نسبی خام و از بسیاری جهات محصول کاملاً رضایتبخشی نبود و خیلی زود با واکسنهای بهتری جایگزین شد.

گزارش آزمایه میدانی با توزیع وسیع واکسن برای استفاده عموم دنبال و سریعاً معلوم شدکه مقادیر کمی از این واکسن واقعاً موجب موارد شدید ابتلا به فلج می‌شود. به این علت توزیع واکسن متوقف شد، درحالی که فرایند دوباره مورد ارزیابی قرار می‌گرفت. چندماه بعد توزیع مجدد آن آغازشد، اما شتاب پذیرفتن آن شکسته شده‌بود و ناپدید شدن سریع فلج که پژوهشگران به آن امیدوار بودند، پیش نیامد. در عین حال تحقیق بر روی یک واکسن ویروسی کشته‌شده خیلی خالص و چندین واکسن ویروس زنده مختلف پیشرفت کرد، و در طول چند سال واکسن سالک در آمریکا ولی نه درسوئد با یک واکسن ویروسی زنده جایگزین گردید.

درنتیجه آزمون تاریخی درازمدت واکسن سالک هرگز انجام نشد. مطمئن نیستیم که آیا آن واکسن می‌توانست به‌طور نسبتاً کامل به ازبین بردن بیماری فلج اطفال که اکنون میسر شده‌است نایل آید؟ یا نه. معهذا، این مطلب اهمیت نقش آن را در میسرساختن اولین موفقیت دلگرم‌کننده در حمله به این بیماری کم نمی‌کند، نقشی که در آن طرح آزمایش دقیقی که دارای اطلاعات آماری بود سهم بسزایی ایفا کرد.

مسائل

۱ـ با استفاده ازشکل ۱ به عنوان یک مثال توضیح دهید که چرا گروه کنترل در آزمایشهایی که قرار است تأثیر دارو یا واکسن را تعیین کند مورد نیاز است؟

۲ـ با انتقاد از بیان زیر نیاز به گروههای کنترل را شرح دهید:

«مطالعه بر روی منافع ویتامین c نشان داد که ۹۰ درصد مردمی که از سرماخوردگی رنج می‌برند، هنگامی که ویتامین c می‌خورند سرماخوردگی آنهادر فاصله یک هفته بهبود می‌یابد.»

۳ـ تفاوت بین روش کنترل مشاهده و روش کنترل دارونما را شرح دهید. کدامیک را ترجیح می‌دهید و چرا؟

۴ـ چرا داشتن آزمایش دوسوکور مهم است؟

۵ـ اگر آزمایشهای دوسوکور تنها روش رضایتبخش برای جلوگیری از اریبی مشاهده‌گر باشد، چرا در همه موارد استفاده نمی‌شود؟

۶ـ اگر در یک آزمایش به‌جای نمونه تصادفی افراد از داوطلبین استفاده شود نتیجه آزمایش هیچ اعتباری خواهدداشت؟ درباره نتایج آن چه می‌توان گفت؟

۷ـ چرا به‌نظر می‌رسید که همه‌گیریهای ویروس فلج اطفال سخت‌ترین ضربه را به کسانی که از نظر بهداشتی وضع بهتری دارند وارد می‌کند؟

۸ـ چرا آزمایه‌های میدانی در مقیاسی بزرگ برای به‌دست آوردن شواهد متقاعدکننده جهت مؤثر بودن واکسن سالک مورد نیازاست؟

۹ـ با مراجعه به شکل ۱، در کدام سال بالاترین موارد ابتلا، پایینترین موارد ابتلا، بیشترین افزایش، و کمترین افزایش در ابتلا به فلج اطفال به‌وقوع پیوست؟ مقدار تقریبی این رخدادها و افزایشها را بگویید.

۱۰ـ با مراجعه به شکل ۱، در مورد کاربرد تعداد موارد اظهارنظر کنید. می‌توانید نماگر متفاوتی از انتشار بیماری فلج اطفال را در ایالات متحده طی سالهای ۱۹۳۰ تا ۱۹۵۶ پیشنهاد کنید؟ چه موقع این دو نماگر برابرند؟

(راهنمائی: به جدول ۱ مراجعه کنید.)

منابع

K. Alexander Brownlee. 1955. "Statistics of the 1954 Polio Vaccine Trials." *Journal of the American Statistical Association 50(272): 1005-1013.*

Thomas Francis. Jr., et al. 1955. "An Evaluation of the 1954 Poliomyelities Vaccine Trials-Summary Report." *American Journal of Public Health 45(5): 1-63.*

Paul Meier. 1957. "Safety Testing of Poliomyelities Vaccin." *Science 125(3257): 1067-1071.*

D. D. Rutstein. 1957. "How Good Is Polio Vaccine?" *Atlantic Monthly 199:48-51.*



ایمنی داروهای هوشبر

لینکلن. ای. موزس[۱] دانشگاه استانفورد
فردریک موستلر[۲] دانشگاه هاروارد

در سال ۱۹۵۸، بیمارستانهای ایالات متحده آمریکا استفاده از یک داروی هوشبر جدید موسوم به هالوتان[۳] را آغاز کردند. این دارو برخلاف برخی از داروهای هوشبر متداولتر، آتش نمی‌گرفت، لذا خطر بالقوه آتش‌گرفتن یا انفجار، که درمدت عملهای جراحی موجب نگرانی

۱ـ *Lencoln E. Moses*

۲ـ *Frederick Mosteller*

۳ـ *Halothane*

بود، وجود نداشت. بیماران آن را کمتر ناخوشایند می‌یافتند و سریعتر از بیهوشی خارج می‌شدند و عوارض شدید بعدی آن کمتر بود. پژوهش جامع آزمایشگاهی و آزمایشهایی روی جراحی حیوانها و انسانها، اعتقاد بر ایمنی این دارو را تقویت کرد. بنابراین دلایل خوبی وجود داشت که از هالوتان بسرعت استفاده گسترده‌ای صورت گیرد. در ۱۹۶۲ جراحان در نیمی از عملهای خود هالوتان رابه‌کار می‌بردند.

اما، بعد از چند سال، وقتی در نوشته‌های پزشکی درباره برخی مرگهای فوق‌العاده غیرعادی ولی اغلب مشابه بیمارانی که به‌تازگی هالوتان را دریافت کرده‌بودند، گزارشهایی منتشر شد به این دارو بدگمان شدند. چند بیمار که در حال بهبودی پس از عمل جراحی بودند، ناگهان حالشان بد شد، تب کردند و مردند. کالبدشکافی بعدی تغییرات مهلک زیادی را در کبد آنها آشکار کرد. گرچه تنها تعداد کمی ازچنین گزارشهایی وجود داشتند اما طرح این سئوال طبیعی بود که آیا این حادثه‌ها به‌معنای آن است که هالوتان برای کبد انسانها سمی خطرناک است؟ و آیا باید استفاده از هالوتان را به‌عنوان داروی هوشبر در جراحی متوقف کرد؟ قبلاً می‌دانستند که بعضی از ترکیبهای شیمیایی مشابه با هالوتان موجب صدمه به کبد می‌شود. بنابراین سئوال بالا با شدت هرچه بیشتر مطرح شد. برای جمع‌آوری و آزمون شواهدی که ممکن بود در پاسخ به این سئوالها کمک کنند، کمیته‌ای ملی تعیین شد.

اطلاعات زیادی مورد نیاز بود. اولاً ممکن بود که تغییرات کبد ناشی از مصرف هالوتان با داروهای هوشبر دیگر نیز به همان فراوانی رخ دهد، اما پزشکان به دلیل توجه کمتر به داروهای هوشبری قدیمی، بندرت چنین مواردی را گزارش کرده‌بودند. ثانیاً نکته کلیدی این بود که هالوتان چه یک اثر زیانبار خاص برکبد داشته و چه نداشته باشد، این امکان وجود دارد که سایر مزایای آن یک نرخ کلی پایین مرگ پس از عمل جراحی را نتیجه دهد. غالباً چیزهایی که کاربرد آنها آسانتر است، واقعاً بهتر کار می‌کنند (این مطلب برای کاردهای تیز، ویولون‌های ظریف، و دستورالعملهایی که خواندن آن راحت است صادق می‌باشد). بنابراین، یک داروی هوشبر که استفاده از آن برای پزشک آسانتر است باید بهتر کار کند و در طول عمل جراحی، تاحدی نرخهای کمتر مرگ را نتیجه دهد. اطلاعاتی لازم بود که این امکان را تأیید یا رد کند.

به این سئوالها نمی‌شد با انجام تجربه‌های آزمایشگاهی روی موشها پاسخ داد، زیرا موشها

تحت داروی هوشبر رفتاری متفاوت با انسانها دارند. با مراجعه به کتابها نیز ممکن نبود به این سئوالها پاسخ داد، زیرا اطلاعات لازم اصلاً وجود نداشت. سئوال از خبرگان نیز مفید نبود، زیرا خبرگان مختلف، عقاید کاملاً متفاوتی داشتند.

مدارک

گردآوری مقدار زیادی از مدارک ضروری بود، لذا کمیته تصمیم گرفت یک بررسی از تجربه‌های بیمارستانی به‌عمل آورد. در آن زمان، در ایالات متحده، هالوتان را به حدود ۱۰ میلیون انسان داده‌بودند. تعداد زیادی از این بیماران در گروه خاصی متشکل از ۳۴ بیمارستانی بودند که سوابق خوبی را ضبط کرده بودند و کارکنان آنها مشتاقانه مایل بودند به سئوالهای زیادی که ما مطرح می‌کردیم، پاسخ دهند. آنها با نمونه‌گیری از سوابق عملهای جراحی خود که در طول سالهای ۱۹۶۰ تا ۱۹۶۴ انجام شده‌بود، با کمیته همکاری کردند. علاوه بر ثبت اینکه در طول ۶ هفته پس از جراحی، بیمار مرده یانمرده است، اطلاعاتی درباره داروی هوشبر مورد استفاده و همچنین حقایقی درباره شیوه جراحی، جنس، سن، و وضعیت جسمی بیمار قبل از عمل دادند. در بین ۸۵۰۰۰۰ عمل تحت بررسی، حدود ۱۷۰۰۰ مرگ وجود داشت. یعنی نرخ مرگ ومیر ۲٪ بود. نرخهای مرگ ومیر که در جدول ۱ نشان داده شده‌اند، برای هریک از چهار داروی هوشبر اصلی و برای گروه پنجم متشکل از همه داروهای هوشبر دیگر محاسبه شده‌است. توجه کنید که اینها نرخهای مرگ ومیر ناشی از همه علل، از جمله ناشی از بیماریهاست و خاص مرگ و میرهای ناشی از داروی هوشبر نیست.

نیاز به تعدیل

جدول ۱ القا می‌کند که هالوتان به اندازه هر داروی هوشبر متداول دیگر ایمن بود، اما به چنین القایی نمی‌شد صرفاً اعتماد کرد. افراد دست‌اندرکار پزشکی می‌دانند که بعضی از داروهای هوشبر (مثلاً سیکلوپروپان[۱]) بیش از سایر داروهای هوشبر (نظیر پنتوتال[۲] که خیلی کمتر

۱ـ *Cyclopropane*

۲ـ *Pentothal*

جدول ۱ـ نرخهای مرگ و میر مربوط به داروهای هوشبر مختلف

سایر	اتر	سیکلوپروپان	پنتوتال	هالوتان
٪۳	٪۱/۹	٪۳/۴	٪۱/۷	٪۱/۷

به‌کار می‌رود) در عملهای شدید و پرخطر به‌کار برده می‌شود. بنابراین قسمتی یا همه تفاوتهای بین این نرخهای مرگ ومیر ممکن است ناشی از گرایش به استفاده از یک داروی هوشبر در عملهای مشکل و استفاده از داروهای هوشبر دیگر در عملهای ساده‌تر باشد.

عملهای مختلف، مخاطره‌های خیلی متفاوتی برای مرگ به همراه دارند. البته دریافته‌اند که نرخهای مرگ و میر برای بعضی از عملها به کمی ۰/۲۵٪ و برای بعضی دیگر نزدیک ۱۴٪ است. تغییر نرخهای مرگ و میر برای رسته‌های وضعیت جسمی بیماران حتی بیشتر حیرت‌انگیز بود، و از ۰/۲۵٪ برای رسته بامساعدترین وضعیت جسمی تا بیش از ۳۰٪ برای رسته با نامساعدترین وضعیت جسمی تغییر می‌کرد. سن به میزان زیادی اهمیت داشت: مساعدترین وضع گروه ده‌ساله سنی (۱۰ تا ۱۹سال) نرخ مرگ و میری کمتر از ۰/۵٪ داشت، درحالی که نامساعدترین وضع گروه سنی (بالای ۹۰ سال) نرخ مرگ و میر ۲۶٪ را دارا بود. عامل دیگر، جنس بود که برای آن احتمال مرگ ومیر زنان حدود $\frac{2}{3}$ مقدار آن برای مردان بود.

به دلیل اینکه عاملهای سن، نوع عمل، وضع جسمانی، و جنس در تعیین نرخ مرگ ومیر مهم بودند، واضح بود که حتی فزونی نسبتاً کمی از بیماران باوضع نامساعد، در گروهی که یک داروی هوشبر خاص را دریافت می‌کردند، می‌توانست نرخ مرگ ومیر را به‌صورتی کاملاً قابل توجه افزایش دهد. پس تفاوتهای موجود در جدول ۱ می‌توانستند عمدتاً و به‌صورتی ممکن و حتی تام، ناشی از تفاوتهای انواع بیماران و انواع عملهای مربوط به داروهای هوشبر مختلف باشد. مطمئناً لازم بود قبل از تلاش برای تعیین ایمنی نسبی داروهای هوشبر، برای برابرسازی نوع عمل، جنس، وضعیت جسمی، و سن، تعدیلی انجام شود.

اگر این داده‌ها از یک آزمایش با طراحی مناسب به‌دست آمده‌بودند، محققان ممکن بود ترتیب گردآوری داده‌ها را به‌گونه‌ای داده‌باشند که بیماران دریافت‌کننده داروهای هوشبر مختلف، از نظر سن، نوع عمل، جنس و وضعیت جسمانی قابل مقایسه باشند. واقعیت این بود

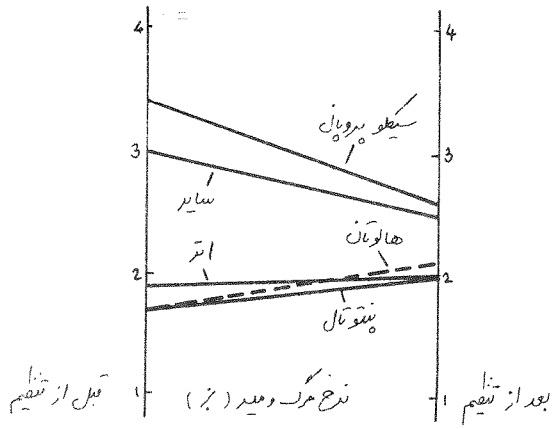
که داده‌ها با مرور سابقه‌های پیشین گردآوری شده‌بودند. بنابراین در انواع بیماران دریافت‌کننده داروهای هوشبر مختلف، انتظار تفاوتهای اساسی می‌رفت و به‌راستی این تفاوتها ظاهر شدند. مثلاً، سیکلوپروپان دو یا سه برابر بیشتر از هالوتان به بیماران با وضع جسمی بد و خیلی بیشتر از هالوتان به بیماران بالای ۶۰سال داده شده‌بود. خصوصیات عجیب این‌چنین زیاد بودند و اینها حتماً بر نرخهای مرگ ومیر جدول ۱ اثر می‌گذاشتند.

اجرای تعدیلها

کار موردنظر، برطرف کردن آثار این متغیرهای مداخله‌گر از نرخهای مرگ ومیر متناظر با پنج داروی هوشبر بود. خوشبختانه انجام این کار میسر بود. گفتیم «خوشبختانه» زیرا انجام این کار می‌توانست غیرممکن باشد. مثلاً اگر هر داروی هوشبر در مورد مجموعه خاصی از عملها به‌کار رفته‌بود و این مجموعه متفاوت از مجموعه‌هایی بود که از داروهای هوشبر دیگری استفاده کرده‌بودند، آنگاه تفاوتهایی که در نرخهای مرگ ومیر پیدا می‌شدند، می‌توانستند کاملاً از تفاوت عملهای انجام شده، ناشی شده‌باشند و راهی وجود نداشت که عمل را از داروی هوشبر جداکنند و به مسئله سروصورتی بدهند. اما در داده‌های این بررسی، در بین داروهای هوشبر مقادیر مشترک زیادی در رسته‌های سن، نوع عمل، جنس، و وضع جسمی بیمار وجود داشت، به قسمی که تعدیل آماری یا برابرسازی ممکن بود. تحلیلی که در آن از چنین تعدیلی استفاده می‌شد از راههای مختلفی انجام می‌گرفت، زیرا پیچیدگی مسئله، چند رهیافت مختلف را معقول جلوه می‌داد و نمی‌شد تنها بر یک رهیافت متکی شد. صرف‌نظر از روش به‌کاررفته، هماهنگی نزدیکی در یافته‌ها وجود داشت.

خلاصه‌ای از نتایج را درجدول ۲ نشان داده‌ایم. توجه کنید که نرخهای مرگ و میر هالوتان و پنتوتال بعد از تعدیل بزرگتر از نرخهای تعدیل‌نشده جدول ۱ هستند. همچنین، سیکلوپروپان و «سایرین» دوتایی که بالاترین نرخها را در جدول ۱ دارا هستند، اکنون کمترین نرخهای مرگ ومیر را دارند. آثار این تعدیلها کاملاً اهمیت دارند: هالوتان که بنابر پیام جدول ۱ دوبرابر ایمن‌تر از سیکلوپروپان است، اکنون تقریباً تنها یک‌پنجم ایمن‌تر به‌نظر می‌رسد. شکل ۱، به‌صورت نمودار، اثر تعدیل را نشان می‌دهد.

شکل ۱ـ آثار تعدیل نرخهای مرگ ومیر. نقاط رسم‌شده مربوط به جدولهای ۱ و ۲ هستند.



یک حالت ساده خاص[۱]

برای درک ماهیت تعدیلها، نگاهی به حالتی خیلی خاص و با داده‌های ساختگی، مفید است. فرض کنید که قرار است هالوتان و سیکلوپروپان را با هم مقایسه کنیم و تعدیل را تنها برای وضع جسمی انجام دهیم. همه بیماران را به صورت وضع جسمی خوب یابد رده‌بندی می‌کنیم. داده‌های ساختگی می‌توانند نظیر آنهایی باشند که در جدول ۳ نشان داده شده‌اند. بدین ترتیب از هر ۵ مورد استفاده از هالوتان (در این داده ساختگی)، ۴ مورد آن بر روی بیماران با وضعیت جسمی خوب بود ولی در مورد سیکلوپروپان درست عکس این موضوع برقرار بود. به این

جدول ۲ـ نرخهای مرگ ومیر تعدیل شده برای داروهای هوشبر مختلف (تعدیل شده برای سن، نوع عمل، جنس، و وضع جسمی)

سایر	اتر	سیکلوپروپان	پنتوتال	هالوتان
٪۲/۵	٪۲	٪۲/۶	٪۲	٪۲/۱

۱ـ این بخش را می‌توان بدون لطمه به پیوستگی مطلب نادیده گرفت.

معنی که از هر ۵ مورد استفاده از آن یک مورد بر روی بیماران با وضعیت جسمی خوب بود. دو داروی هوشبر دقیقاً نرخهای مرگ ومیر برابری برای هر رده‌بندی جداگانه وضع جسمی دارند، اما نسبتهای بسیار متفاوت وضع جسمی برای هردو، نرخ مرگ ومیر کلی برای سیکلوپروپان را تقریباً دوبرابر نرخ مرگ ومیر برای هالوتان می‌کند.

فرض کنید با محاسبه اینکه نرخ مرگ ومیر روی هم چه اندازه‌ای می‌شد اقدام به تعدیل کنیم. با فرض اینکه دو داروی هوشبر به جامعه‌ای از بیماران داده می‌شد که در آن ٪۷۰ دارای وضع جسمی خوب و ٪۳۰ دارای وضع جسمی بد بودند. نرخهای مرگ ومیر برای هر گروه از وضع جسمی، یکسان باقی می‌ماند. محاسبات به صورتی خواهند بود که در جدول ۴ نمایش داده‌ایم. اکنون، با این داده‌های ساختگی، برای هر داروی هوشبر، نرخ مرگ و میر روی‌هم برابر با ٪۱/۶ است.

در عمل، تعدیلهای واقعی خیلی پیچیده‌ترند که بخشی از آن به دلیل این است که باریک بودن فاصله‌های داده‌ها بسیاری از خانه‌ها را کلاً بدون درایه باقی می‌گذارد و به بسیاری از خانه‌ها درایه‌های کوچکی تخصیص می‌دهد.

جدول ۳ـ داده‌های ساختگی

	تعدادبیماران	تعدادمرگ ومیر	نرخ مرگ ومیر
هالوتان			
وضع جسمی بد	۲۰۰	۶	٪۳
وضع جسمی خوب	۸۰۰	۸	٪۱
کل بیماران	۱۰۰۰	۱۴	٪۱/۴
سیکلوپروپان			
وضع جسمی بد	۸۰۰	۲۴	٪۳
وضع جسمی خوب	۲۰۰	۲	٪۱
کل بیماران	۱۰۰۰	۲۶	٪۲/۶

جدول ۴ـ نرخهای مرگ ومیر تعدیل‌شده (داده‌های ساختگی)

	تعدادبیماران	تعدادمرگ ومیر	نرخ مرگ ومیر
هالوتان			
وضع جسمی بد	۳۰۰	۹	٪۳
وضع جسمی خوب	۷۰۰	۷	٪۱
کل بیماران	۱۰۰۰	۱۶	٪۱/۶
سیکلوپروپان			
وضع جسمی بد	۳۰۰	۹	٪۳
وضع جسمی خوب	۷۰۰	۷	٪۱
کل بیماران	۱۰۰۰	۱۶	٪۱/۶

سازگاری روی بیمارستانها

حتی بعد از تعدیلها، هنوز تغییرپذیری قابل ملاحظه‌ای از بیمارستانی به بیمارستان دیگر وجود داشت. اطمینان ما درباره اعتبار ارقام تعدیل‌شده به‌صورت آماری در جدول ۲، تحت تأثیر چگونگی سازگاری نرخهای مرگ و میر، از بیمارستانی به بیمارستان دیگر بود. با این همه، اگر بیمارستانهای زیادی وجود داشتند که در آنها سیکلوپروپان نرخ مرگ ومیر تعدیل‌شده پایینتر از هالوتان داشت، هرچند که به‌طور متوسط نرخ مرگ و میر هالوتان پایینتر بود به دلیل تفاوت در نرخهای تعدیل‌شده مرگ ومیر، احساس عدم اطمینان می‌کردیم که هالوتان واقعاً ایمن‌تر باشد. البته به‌واسطه وجود عاملهای شانس، می‌باید انتظار نوسانهایی از یک بیمارستان به بیمارستان دیگر را داشته باشیم. نباید انتظار سازگاری مطلق داشته‌باشیم.

سئوال این است: آیا مقایسه‌های بین نرخهای مرگ ومیر تعدیل‌شده داروهای هوشبر در بیمارستانها، به اندازه کافی سازگارند که به ما اجازه دهند تا تفاوتهای ظاهری جدول ۲ را جدی بگیریم؟ فنون آماری نسبتاً پیچیده‌ای برای بررسی این سئوال لازم بود، اما نتایج روشن بودند: هالوتان و پنتوتال، هردو قطعاً دارای نرخهای مرگ ومیر تعدیل‌شده پایینتری نسبت به سیکلوپروپان و «سایرین» بوده و این نرخهای مرگ ومیر تعدیل‌شده پایینتر، واقعی بودند. به‌این

مفهوم که نمی‌توانستیم آنها را با نوسانهای شانسی توجیه کنیم. تفاوتها به قدر کافی سازگار بودند که بتوان آنها را باور کرد. از سوی دیگر، اتر با همان نرخ مرگ ومیر تعدیل‌شده نظیر هالوتان و پنتوتال، الگوی سازگاری برای مقایسه نداشت. بنابراین، اتر را نمی‌توان به صورتی قابل اعتماد با داروهای هوشبر دیگر مقایسه کرد. ما نمی‌توانیم از روی گواه حاصل بگوییم که آیا اتر می‌تواند کمی از آنچه در اینجا ظاهر شده‌است یا تاحدی کمتر ایمن باشد؟ امکان دارد که به اندازه هالوتان یا بیش از آن ایمنی داشته باشد. ممکن است که ایمن‌تر از «سایرین» نباشد. یافته‌های مربوط به اتر نامعین هستند؛ زیرا اجراهای کمتری از اتر وجود داشته و این اجراها در تعداد کمی از بیمارستانها متمرکز بود. بنابراین داده‌های کمتری برای ادامه وجود دارد.

جمع‌بندی تا این مرحله

تا اینجا از یافته‌های این بررسی چه نتیجه‌ای می‌گیریم؟ اولین و مهمترین نتیجه، این نتیجه شگفت‌آور است که هالوتان که در آغاز بررسی مورد سوءظن بود، به‌صورت یک عامل هوشبر قطعاً ایمن و احتمالاً برتر ظاهر شده‌است.

دومین نتیجه این است که می‌بینیم یک بررسی آماری دقیق از ۸۵۰۰۰۰ عمل، با وجود «تجربه» خیلی بزرگتر ۱۰میلیون تجویز هالوتان و دهها میلیون از تجویز داروهای هوشبر دیگر، حرفه پزشکی را قادر می‌سازد که با قدرتی بیش از آنچه قبلاً ممکن بود به سئوالها پاسخ دهد.

بازهم تفاوتهای بیمارستانی

با مشاهده اینکه ۳۴ بیمارستان، نرخهای مرگ ومیر روی‌هم بعد از عمل جراحی بسیار مختلفی داشتند، یک سئوال کاملاً تازه مطرح شد: این نرخها از حدود ۰/۲۵٪ تا حدود ۶/۵٪ تغییر می‌کردند. به‌نظر می‌رسید که دامنه تغییر بدان مفهوم است که احتمال مرگ در طول ۶ هفته بعد از جراحی می‌توانست در یک بیمارستان بیش از ۲۵ برابر بیمارستان دیگر باشد. اما درست مثل قبل، دلایلی قوی وجود داشتند که به این اطلاع تکان‌دهنده با شک وتردید بنگرند. بعضی از بیمارستانهای تحت مطالعه، عملهای مشکل نظیر جراحی قلب باز انجام نداده بودند، درحالی که بیمارستانهای دیگر در چنین رسته‌هایی بار بسیار سنگینی داشتند. این نوع اختلاف به‌تنهایی

موجب تفاوتهایی بین نرخهای مرگ ومیر بیمارستانی می‌شد. به‌علاوه ممکن بود توزیع سن متفاوت باشد، و شاید از بیمارستانی به بیمارستان دیگر این تفاوت به‌گونه‌ای بارز وجود داشت. درواقع یکی از بیمارستانها، بیمارستان کودکان و دیگری بیمارستان افراد کهنه‌کار بود. برخی از بیمارستانها ممکن بود (نسبت به سایر بیمارستانها) افراد نیازمند به جراحی با وضعیت جسمی بد را بیشتر بپذیرند. بنابراین، همان متغیرهای مداخله‌گر قبلی مطمئناً بر تفاوتهای بین نرخ مرگ ومیر بیمارستانها اثر می‌گذاشت. اگر برای آنها کار تعدیل صورت می‌گرفت آیا تفاوتهای بزرگ بین نرخهای مرگ‌ومیر بیمارستانها از بین می‌رفتند؟ خیلی تقلیل می‌یافتند؟ بدون تغییر باقی می‌ماندند؟ یا همان‌طور که قابل تصور است واقعاً زیاد می‌شدند؟

شیوه‌های تعدیل به‌کار رفتند، و نتیجه این بود که بیمارستانهای با نرخ مرگ ومیر بالا، بعداز تعدیل، به سمت نرخ مرگ‌و میر روی‌هم ۲٪ پایین آمدند و بیمارستانهای بانرخ مرگ ومیر پایین، بعداز تعدیل، به سمت نرخ مرگ ومیر روی‌هم ۲٪ بالا آمدند. نرخهای مرگ ومیر تعدیل شده بیمارستانها دیگر از ۰/۲۵٪ تا ۶/۵٪ تغییر نمی‌کردند، بلکه بزرگترین نرخ تعدیل شده مرگ ومیر تنها حدود ۳ برابر کوچکترین نرخ تعدیل‌شده بود. حال، تقریباً هر گروه ۳۴تایی از چنین نرخهایی، مقداری پراکندگی از خود نشان خواهدداد، و نسبت بزرگترین نرخ به کوچکترین نرخ باید عددی بزرگتر از ۱ باشد. حتی اگر یک بیمارستان در چندین دوره مختلف اندازه‌گیری می‌شد نرخها تحت‌تأثیر شانس نوسان می‌کردند. نوسان نرخ باید قابل ملاحظه بوده‌باشد، زیرا نرخ مرگ ومیر، اساساً پایین بود و یک مورد مرگ ومیر بیشتر یاکمتر، تفاوتی را در نرخ مشاهده شده دوره به وجود می‌آورد. این واقعیت که دراین داده‌ها، نسبت مزبور به ۳ منتج شد به خودی خود و به‌وضوح نشان نمی‌داد که تفاوتهای توجیه نشده حقیقی بین بیمارستانها وجود داشتند.

مطالعه آماری دقیق نشان داد که احتمالاً برخی تفاوتهای واقعی در نرخهای مرگ ومیر بعداز عمل، از بیمارستانی به بیمارستان دیگر وجود دارند و این تفاوتها را نمی‌توان در کل به‌وسیله جامعه‌های بیماران بیمارستانها برحسب سن، جنس، وضع جسمی و شیوه جراحی توجیه کرد. نظریهٔ آمار نشان داد که نسبت بالاترین و پایینترین نرخ تعدیل‌شده، اگر بیمارستانها از نظر نرخ مرگ ومیر ناشی از عمل، بعد از تعدیل، یکسان باشند، باید حدود ۱/۵ شود (بنابراین نرخهای تعدیل‌شده تنها براساس نوسان حاصل از شانس باهم تفاوت دارند).

لذا وضع از این قرار است که ما با نسبت بزرگ حدود ۲۵ برای نرخهای تعدیل‌نشده شروع کردیم. نسبت نرخهای تعدیل‌شده به ۳ رسید و سپس ۳ با ۱/۵، به‌عنوان نسبت نظری، مقایسه شد. چون تعدیلها به‌زحمت می‌توانستند کامل باشند و چون بدون تردید عاملهای تعدیل نشده هم وجود داشتند که بین بیمارستانها متفاوت بودند، نتیجه می‌گیریم که درواقع نرخهای مرگ ومیر تعدیل‌شده بیمارستانها نزدیک یکدیگرند. بدین‌ترتیب آنچه برمبنای نرخهای تعدیل‌نشده مرگ ومیر بیمارستانها مسئله بهداشتی عمومی تکان‌دهنده‌ای به‌نظر می‌رسید، بعد از بررسی آماری، ثابت شد کاملاً چیز دیگری است. مسئله مشهود، عمدتاً گرچه شاید نه به‌طور کامل، یک تظاهر برجسته تفاوتها، نه در کیفیت مراقبت جراحی، بلکه در مشکل موارد جراحی انجام شده در بیمارستانهای مختلف بود.

مطالعات پیگیر

نگرانی درباره تفاوتهای بیمارستانها، انگیزه مطالعات جامع بعدی بیمارستانهای ایالات متحده شد تا تعیین کنند که آیا تفاوتهای نرخ مرگ ومیر جراحی بین بیمارستانها به همان بزرگی نرخ مرگ ومیری است که بررسی مربوط به هالوتان القا می‌کرد یانه. این مطالعات پیگیر طراحی نشدند. به‌منظور اینکه اطلاعات بیشتری درباره نرخ عوارض بعد از جراحی به‌دست آید و به منظور اینکه شروعی برای کاوش باشد، برای یافتن دلایل تفاوتهای بین بیمارستانها، اگر چنین تفاوتهایی تأیید می‌شدند. (ن.ک. *فلود و اسکات*. ۱۹۸۷)

مطالعات، نتایج بررسی مربوط به هالوتان را تأیید و برآوردهای تفصیلی بیشتری را درباره اندازه‌های تفاوتها و انواع شیوه‌های جراحی که بیشترین تغییرات را داشتند، اضافه کردند. به علاوه، باتعدیل دقیق نرخهای بیمارستانها برای مشخصه‌های بیمارانی که به‌وسیله بیمارستانها پذیرفته شده‌اند و سپس با همبسته‌کردن مشخصه‌های بیمارستانی با اندازه‌های عملکرد بیمارستانی، فرضهای جالبی برای مطالعه بعدی توسعه یافته‌اند. مثلاً، نتایج القا کردند که تعیین‌کننده بسیار مهم کیفیت نتیجه جراحی، دقت و توانی است که پزشکان بیمارستانی در تأیید جراحان برای امتیازهای بیمارستانی در بیمارستان اعمال می‌کنند.

صدمه به کبد

اکنون به سئوال صدمه به کبد برمی‌گردیم: کل مرگهای ناشی از این علت، کم بودند وعدم کالبدشکافی تقریباً نیمی ازاجساد این مرگها، کسب نتایج استوار را غیرممکن ساخت. اما، پس از اینکه بررسی انجام شد، یک متخصص بیهوشی که غالباً وقتی هالوتان را برای هوشبری بیماران به‌کار می‌برد در معرض این دارو بود، دریافت که نسبت به هالوتان حساس شده‌است. در او نشانه‌های بدکارکردن کبد ناشی از تنفس این دارو ظاهر شد.

آیا افراد، گاه‌وبیگاه به داروهای هوشبر دیگر حساسند؟ چه تعدادی از افراد چنین حساسیتی را ظاهر می‌کنند؟ پاسخ به این سئوالها سخت است و به دلیل نادربودن رخ‌دادن، بررسی آنها به‌خصوص مشکل است.

سهم آمار

سهم عمده آمار در برنامه کار چه بود؟ اولین سهم، مفهوم پایه‌ای مطالعه نرخ مرگ ومیر بود. این کار باید انجام می‌پذیرفت. برای اینکه بتوانند ایمنی داروهای هوشبر را در پرتو تجربه کلی جراحی، و نه فقط در مرگهای ناشی از علتی نادر، ملاحظه کنند. دومین سهم، گرچه از آن بحث نکردیم، استفاده از یک فن آماری ثبت نمونه‌گیریها در مطالعه بود که برای صرفه‌جویی در هزینه و تولید اطلاع باکیفیت بالا طرح‌ریزی شد. سومین سهم، تعدیلهای آماری خاصی بودند که برای ارزیابی نتایج، وقتی آن همه متغیرهای مهم ـ سن، نوع عمل، جنس، وضع جسمی بیمار و نظایر آنها – قابل کنترل نبودند، باید به‌وجود می‌آمدند. چهارمین سهم، به‌عنوان یک نتیجه، این است که پیش‌فرض اصلی مطالعه تأیید نشد. اما نتیجه جدیدی ظاهرشد: به‌نظر می‌رسید که هالوتان ایمن‌تر از سیکلوپروپان باشد. مطالعه همچنین یک امر نامسلم را نتیجه داد: مزیت مقایسه‌ای اتر نامطمئن است. نمی‌توانیم بگوییم که نرخ مرگ ومیر مربوط به اتر، بالاتر، پایینتر یا تقریباً همانند نرخ مرگ ومیر هالوتان است. گرچه قرائن حاکی از این است که نرخها تقریباً یکی هستند. پنجمین سهم، این گواه جدید درباره تفاوتهای بیمارستانی است که نشان می‌دهد تغییرات وسیع اولیه در نرخهای مرگ ومیر، وقتی به صورتی مناسب برای جامعه‌های بیماران تعدیل شود تنها تفاوتهای کوچکی در نرخها را توجیه‌نشده باقی می‌گذارد. این تفاوتهای باقیمانده منجر به این

شد که حرفه پزشکی در سال ۱۹۷۱ شروع به مطالعه تفاوتهای بین بیمارستانها در نرخهای مرگ و میر بعد از عمل بنماید، به این امید که راههایی را برای بهبودی مراقبت بعد از عمل پیدا کند.

کلام آخر

در ۱۹۸۸ بندرت از هالوتان، اتر، و سیکلوپروپان استفاده می‌شود. این داروها با داروهای هوشبر جدید جانشین شدند.

مردم، به دلیل سیستم بازپرداخت معمول در دهه ۱۹۸۰ که بیمارستانها را به ترخیص زودتر بیماران از بیمارستان یا غالباً به معالجه آنها در مرکز پزشکی سیار، به جای پذیرفتن آنها در بیمارستان، تشویق می‌کند، نگرانی تازه‌ای درباره کیفیت مراقبت بیمارستانها دارند. تقلیلهای حاصل در تعداد پذیرشها و طول مدت بستری شدن دربیمارستان به ترس از تنزل کیفیت منجر شده‌است، اما ما برای هر دومورد دلیل کمی در دست داریم. در حال حاضر از روشهایی شبیه به روشهای این آزمایش، برای روشن شدن مطلب استفاده می‌شود.

معمولاً تلاشهای مربوط به اندازه‌گیری مراقبت بیمارستانی، مبتنی بر ارقام تعدیل‌شده مرگ ومیر برای شدت بیماری و مشخصه‌های دیگر بیماران است. اما پژوهشگران زیادی بر این باورند که چنین نرخهای تعدیل‌شده مرگ ومیر، به‌طور صحیح، کیفیت مراقبت را در یک کشور پیشرفته اندازه نمی‌گیرد. درعوض، آنها برآمدهای دیگری، نظیر پیچیدگیها و پذیرشهای مجدد را مرور می‌کنند.

گرچه هالوتان جای خود را به هوشبر دیگری داده‌است، بررسی ملی هالوتان از دو جهت دیگر مهم بوده‌است: اولاً آمارشناسان را به گسترش روشهایی ترغیب کرده‌است که اکنون کاربرد گسترده‌ای، نه‌تنها در پزشکی وبهداشت، بلکه در سایر زمینه‌ها هم دارند. ثانیاً نتایج بررسی، هنوز تنها منبع اطلاعات مربوط به برخی نرخهای مرگ ومیر عملهاست که برای عاملهای مختلف مخاطره تعدیل شده‌است.

مسائل

۱ـ توضیح دهید که چگونه ممکن است هالوتان، با وجود گواه حاصل از کالبدشکافی که نشان

می‌دهد این دارو به کبد صدمه می‌زند، به اندازه سایر داروهای هوشبر ایمن است.

۲ـ برابرسازی یا تعدیل چه معنایی دارد؟

۳ـ تفاوت بین ایمنی هالوتان در جدولهای ۱ و۲ را توضیح دهید.

۴ـ چگونه می‌توانید تفاوتهای تعدیل‌نشده نرخهای کلی مرگ ومیر پس از عمل جراحی در ۳۴ بیمارستان تحت مطالعه را توضیح دهید؟

۵ـ نتایج اصلی بررسی چه بودند؟

۶ـ چه ابزار آماری در بررسی به‌کار رفتند؟

۷ـ توضیح دهید که چرا نتایج مربوط به اتر نامطمئن بودند؟

۸ـ آیا هالوتان سبب صدمه کبدی است؟ توضیح دهید.

۹ـ با رجوع به شکل ۱، تعدیل برای کدام داروی هوشبر دارای بیشترین اثر است ؟ از مقایسه جدول ۲ و۱، پاسختان را کنترل کنید.

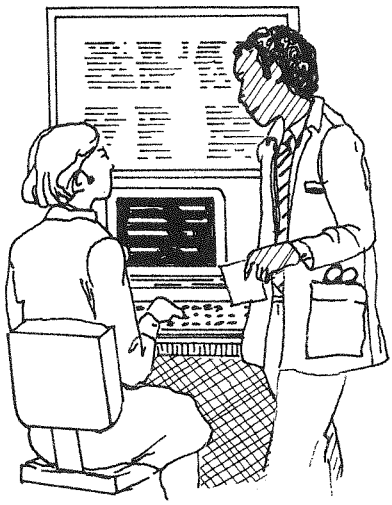
۱۰ـ توضیح دهید که درجدول ۴، چگونه تعداد مرگها در هرگروه مربوط به وضعیت جسمی، محاسبه شده‌است.

۱۱ـ مطالعه اثر یک رژیم غذایی با کلسترول زیاد را در مرگ ناشی از بیماری قلبی کرونر *(CHD)* درنظر بگیرید. نرخهای مرگ ومیر از CHD برای دو گروه مقایسه شده‌اند: یک گروه رژیم غذایی با کلسترول زیاد و دیگری رژیم غذایی با کلسترول متوسط دارد. قبل از مقایسه نرخهای مرگ ومیر، برخی از عاملهایی که برای تعدیل در گروه مورد نیازند چه هستند؟

۱۲ـ چرا مطالعه آماری که در این مقاله توصیف شده تنها متمرکز بر ۸۵۰۰۰۰ عمل بود، درحالی که بیش از ۱۰میلیون عمل بودند که در آنهااز هالوتان استفاده شده‌بود، و دهها میلیون عمل بودند که در آنها داروهای هوشبر دیگر تجویز شده‌بودند؟

منابع

John P. Bunker. William H. Forrest, Jr. Fredrick Mosteller, and Leroy D. Vandam. 1969. *The National Halothane Study,* Report of the Subcommittee on the National Halothane Study of the Committee on Anesthesia.



آزمایه‌های شرکت مترو و تصادفی‌سازی مستمر بیماران

دانکن نیوهوزر[۱] دانشگاه رزرو غربی کیس

بسیاری از بیمارستانها در اروپا و آمریکای شمالی پزشکان را به‌صورت تیمی در اختیار دارند که به مراقبت از گروههای بیماران می‌پردازند. بیماران جدید به‌صورت دوره‌ای به تیمهای پزشکی تخصیص داده می‌شوند. بیمار اول به تیم A فرستاده می‌شود، بیمار دوم به تیم B، و به همین ترتیب معرفی بیماران به تیمها ادامه می‌یابد. این شیوه به منظور تقسیم مساوی کار بین تیمها

۱ـ *Duncan Neuhauser*

صورت می‌گیرد. تیمها می‌توانند شامل پزشکان، پرستارها و کارکنان دیگر بیمارستان باشند. تیمها مسئولیت بیماران خارج ازبیمارستان، داخل بیمارستان و یا هر دو را به دوش می‌کشند. با وجود اینکه این عملکردها از ابتدای قرن بیستم رایج بوده است، تاکنون کسی از این فرصت استفاده نکرده است که به‌دقت آثار تغییر در مداوا را ارزیابی کند و شیوه ارزیابی را به‌طور مستمر برقرار نماید.

بیمارستان عمومی مرکزی کلیولند (مترو)[۱] با ۷۰۰ تختخواب، دارای چنین چهار تیمی موسوم به "شرکتها"[۲] است. هر تیم متشکل از پزشکان داخلی عمومی است که با پرستارهاو کارکنان بهداشتی وابسته دیگر کار می‌کنند. هر شرکت دارای ۱۸ پزشک، ۲۸ تختخواب برای بیماران داخلی، و یک کلینیک برای بیماران خارج از بیمارستان است. از سال ۱۹۷۶ تا ۱۹۸۱ تمام بیماران جدید به‌طور دوره‌ای به تیمها تخصیص می‌یافتند. با شروع سال ۱۹۸۱ بیماران به وسیله یک برنامه کامپیوتری که پیوسته اعداد ۱، ۲، ۳، و۴ را تصادفی تولید می‌کند، به تیمها تخصیص می‌یابند. این تخصیص تصادفی برای اینکه چهار گروه بیمار تقریباً یکسان باشند اطمینان خیلی خوبی فراهم می‌کند. اگر واحد بیماران داخلی شرکت پر باشد، بیمار در واحد دیگری نگهداری می‌شود تا اینکه یک جای خالی آزاد شده و سپس بیمار به آنجا منتقل می‌شود.

در هر بیمارستان، صدها تغییر در سازماندهی و ارائه مراقبتهای پزشکی صورت می‌گیرد به این اعتقاد که چنین تغییراتی موجب فراهم آمدن مراقبت پزشکی بهتر و یا کاراتر می‌شود. فقط درصد کوچکی ازاین تغییرات برای نشان دادن درستی این اعتقاد به‌دقت مورد ارزیابی قرار می‌گیرد.

چون بیماران متفاوتند، ارزیابی مراقبت پزشکی بخصوص مشکل است. باید اطمینان حاصل شود که تغییرات مشاهده‌شده درنتایج (مثلاً مراقبت بهتر با هزینه کمتر)، به‌واسطه خود مراقبت است ونه به‌واسطه اختلاف بین بیماران. تاریخ پزشکی شاهد روشهای درمانی ارزیابی

۱ـ *Cleveland Metropolitan General Hospital (Metro)*

۲ـ *Firms*

نشده محتملاً زیان‌آوری از قبیل روشهای درمان به‌وسیله زالوهای خون‌آشام تا جراحیهایی است که روی جمجمه برای بیماریهای مغز صورت می‌گیرد و در مورد دهها هزار بیمار به‌کار گرفته شده‌اند. برای اینکه چنین شیوه‌های بی‌حاصل و زیان‌آور در آینده کمتر رخ دهد، نیاز دائمی به ارزیابی دقیق وجود دارد.

ازسال ۱۹۸۱ در مترو، چهار شرکت پزشکی و جامعه‌های بیمار متناظر آنها برای ارزیابی یک سری تغییرات جاری به‌کار گرفته شده‌اند. در این آزمایه‌ها، یک تغییر در شیوه عملکرد دو شرکت که به‌طور تصادفی انتخاب شده‌اند، اعمال می‌شود. در حالی که دو شرکت دیگر به عنوان شاهد به‌کار می‌روند. برای افزایش بیشتر قابلیت مقایسه‌ها، از سال ۱۹۸۳ تمامی رزیدنتهای داخلی جدید به‌طور تصادفی به یکی از این چهار شرکت اختصاص یافته‌اند. این رزیدنتها برای سه سال دوره آموزشی خود، در شرکتی که به آن اختصاص یافته‌اند، باقی می‌مانند.

استفاده ازشرکتها

بااستفاده از شرکتهای مترو، یک سری آزمایه تحت هدایت دکتر دیوید کوهن انجام شده‌است. این مطالعات چنین سئوالاتی رامطرح کرد: آیا بازخور اطلاع به پزشکان رزیدنت در مورد هزینه آزمایشهای طبی که سفارش می‌دهند باعث کاهش استفاده از این آزمایشها می‌شود؟ بلی.
آیا یادآوری انجام شیوه‌های پیشگیری مناسب برای بیمارانی که قادر به راه رفتن هستند، باعث استفاده بیشتراز این شیوه‌ها می‌شود؟ بلی.
آیا سازماندهی بهتر مراقبت برای بیماران خارج از بیمارستان باعث کاهش هزینه‌های مراقبت بیماران داخل بیمارستان می‌شود؟ بلی.

به شرح جزئیات دو آزمایه می‌پردازیم. یکی از آزمایه‌ها روی کیفیت بهتر مراقبت و دیگری روی کاهش هزینه‌ها متمرکز بوده است. معمولاً در مترو شیوه تغذیه تزریقی وریدی (مقصود تغذیه بیماران از طریق مایعات به‌وسیله تزریق سرم در داخل یک رگ بازو) توسط پرستارها و پزشکان دائمی صورت می‌گیرد. عده‌ای ازپزشکان را عقیده براین بود که داشتن یک تیم تخصصی تزریق وریدی منجر به کاهش نرخ عفونتها و عواقب دیگر (تورم رگها) می‌شود.

تیمهای تزریق وریدی در واحدهای بیماران داخلی دو شرکت مستقر شدند و درصد بیماران عفونی این دو شرکت با درصد مربوط به دو شرکت دیگر که به شیوه معمولی عمل می‌کردند، مقایسه شد. تعداد موارد عفونی در بیمارانی که به‌وسیله تیمهای تخصصی تغذیه می‌شدند به‌صورتی معنی‌دار کمتر بود. درصد موارد عفونی برای تیمهای تخصصی در مقایسه با مقدار ۳۲ درصد برای گروه شاهد برابر ۱۵ بود. با استفاده ازتیمهای تخصصی درصد موارد با مشکلات خیلی جدی از ۲/۱ درصد به ۰/۲ درصد کاهش پیداکرد. به‌دلیل مشاهده نتایج بالا در چهار شرکت از تیمهای تخصصی استفاده شد. چندی بعد در یک حرکت برای کنترل هزینه‌هاپیشنهاد شد که تیمهای تزریق وریدی حذف شوند. اما با استفاده از شواهد کارایی آنها، اعضای تیمهای تزریق وریدی بر سر کار خود ماندند و بیماران پیوسته از کار آنها بهره‌مند می‌شوند.

در حال حاضر بیمارستان دارای یک سیستم اطلاعاتی کامپیوتری است که دربرگیرنده اطلاعات مربوط به بیماران، پزشکان آنها، شرکتی که در آن کار می‌کنند، آزمایشهای طبی و داروهای استفاده شده‌است. در بررسی دوم، *چارلز هرشی* و همکارانش از این شرکتها و سیستم اطلاعاتی برای اینکه ذهن پزشکان را در مورد استفاده از داروها روشن کنند، استفاده کردند. آنها یک پیام کامپیوتری به دو شرکت فرستادند و به آنها پیشنهاد دادند که پزشکان داروهای باهزینه کمتر را جایگزین داروهای گرانقیمت مشابه کنند و نشان دادند که از این طریق هر پزشک چه اندازه می‌تواند پس‌انداز کند. سپس با استفاده از همان سیستم اطلاعاتی اثر این پیام را اندازه‌گیری کردند. آنها برای شرکتهای تحت آزمایش، اثر جایگزینی معنی‌داری مشاهده کردند (ن.ک. جدول ۱). نکته بسیار جالب درباره این آزمایش هزینه بسیارکم اجرای آن بود. به‌واسطه شیوه تصادفی‌سازی مستمر و سیستم اطلاعاتی کامپیوتری، هزینه اضافی این مطالعه کمتر از هزار دلار شد. بیشترین هزینه مربوط به طول زمانی بود که دکتر هرشی صرف نوشتن مقاله برای چاپ کرد.

تعمیم یافته‌ها

در اینجا جایز است این سئوال را مطرح کنیم که آیا مترو و بیماران آن نماینده تمامی

بیمارستانهای آمریکای شمالی هستند؟ آنچه که در مترو کارا باشد ممکن است در مورد بیمارستانهای دیگر صدق نکند. حتی اگرچنین باشد، صرفاً بهبود وضعیت مراقبت در یک بیمارستان هم به خودی خود کاری است نیکو. بیمارستانهای دیگر می‌توانند برای بهره‌مندی خود، سیستمهای دائمی مشابهی برای تصادفی‌سازی بیماران بوجود آورند.

اگر مطالعات مشابهی به‌این شکل در چندین بیمارستان انجام پذیرد، می‌توانیم بیاموزیم که واقعاً مترو تاچه حد نماینده تمامی بیمارستانهاست. چنین برنامه‌ای در بیمارستانهای کلیولند شروع شده است، که در آنجا تمامی بیمارانی که قادر به راه رفتن بوده و تحت مراقبت کلینیک پزشکی عمومی قرار دارند براساس آخرین رقم شماره کارت تأمین اجتماعی خود به‌طور تصادفی به یکی ازدو تیم پزشکی تخصیص می‌یابند. شماره‌های زوج به یک تیم و شماره‌های فرد به تیم دیگر معرفی می‌شوند. درحال حاضر، مدیر کلینیک، دکتر ویکتوریا *کارگیل* در شرف اتمام آزمایه‌ای است که ببیند آیا استفاده از پرستاران متخصص به‌جای پزشکان رزیدنت برای تشخیص سرطان روده بزرگ موثرتر است یا نه؟ دکتر *کارگیل* دریافت بیمارانی که توسط پرستارها معاینه می‌شدند، احتمال بسیار زیادی داشت که نتایج آزمایشهای خودرا که در منزل انجام می‌شد برای ارزیابی به بیمارستان بفرستند تا درصورت لزوم مورد ارزیابی و درمان پیگیر بعدی قرار گیرند.

اولین عملکرد شیوه تصادفی‌سازی شناخته‌شده بیماران به‌منظور ارزیابی درمان در سال ۱۹۳۱ منتشرشد. (*امبرسون* و همکاران ۲۴ بیمار مسلول را به دوگروه قابل مقایسه ۱۲نفری تقسیم کردند، سپس آنها را دوبه‌دو جور کردند. به‌وسیله پرتاب یک سکه یک گروه ۱۲نفری تحت درمان با سانوکریسین[۱] قرار گرفتند و گروه دیگر گروه شاهد بود.)

اما در دهه ۱۹۵۰ بود که شیوه تصادفی‌سازی به‌طور گسترده به‌کار گرفته شد. آزمایه‌های تصادفی ممکن است بسیار پرهزینه باشند - بعضی از آنها بیش از یکصد میلیون دلار هزینه دارند. اما اگر چنین مطالعاتی منجر به انتشار سریعتر درمان سودمند شود یا از یک شیوه درمان بی‌حاصل پرهزینه جلوگیری کند، سودمندی آنها در سطح جهان ارزش چنین هزینه‌های زیاد را

۱ـ *Sanocrysin*

جدول ۱ـ متوسط هزینه برای داروهای استفاده شده قبل وبعد از بازخور کامپیوتری اطلاعات به پزشکان، مقایسه شرکتهای تحت آزمایش و شرکتهای شاهد مربوط به سال ۱۹۸۴ تا ۱۹۸۵

	قبل ازشروع مطالعه		درآخرزمان مطالعه	
شرکتها	*میانگین هزینه دارو	**انحراف معیار	میانگین هزینه دارو	انحراف معیار
تحت آزمایش	$۸/۴۴	±۱/۲۸	$۸/۲۲	± ۱/۰۱
شاهد	$۸/۵۶	±۱/۸۹	$۸/۷۹	± ۰/۸۲

** هزینه‌های فوق بر پایه سه داروی معمولی بود که توسط ۴۸ پزشک تجویز می‌شد. اختلاف بین شرکتهای شاهد و شرکتهای تحت آزمایش در آخر مطالعه معنی‌دار بود و بسیار بعید به‌نظر می‌رسید که این اختلاف در اثر شانس باشد (هرشی، پورتر، برسلو، وکوهن ۱۹۸۶). گرچه این اختلافهابرای یک نسخه کوچک به نظر می‌رسد، ولی درنظر گرفتن این اختلاف در مورد همه شرکتهای مترو منجر به ۶۵۰۰ دلار پس‌انداز در ماه می‌شد.*

*** انحراف معیار، یک اندازه پراکندگی یاتغییرات است. در بسیاری از توزیعها حدود دو سوم داده‌ها در داخل فاصله یک انحراف معیار از میانگین و حدود ۹۵ درصد داده‌ها در داخل دو انحراف معیار از میانگین قرار می‌گیرند.*

خواهد داشت. می‌توانیم امیدوار باشیم که اگر ارزیابی خوب راکم‌هزینه‌تر، و ساده‌تر و همچنین به‌صورت یک شیوه دائمی اعمال کنیم، منجر به‌کاربرد بیشتر این چنین ارزیابیهایی خواهد شد. ارزیابی مکرر کارایی تغییراتی که منجربه مراقبت بهتر می‌شود، حاصل انتخاب تصادفی بیماران است که به‌طور مستمر انجام می‌پذیرد. این، یک ایده کلی است که می‌توان آن را خارج از زمینه مراقبت بهداشتی به‌کار گرفت.

در جهان مجموعه بزرگی از سازمانها وجود ندارد که آمادگی داشته باشند ارزیابیهای پزشکی را از طریق انجام آزمایش در یک اعلام اضطراری دربیمارستانها انجام دهند. کم وبیش هر آزمایش را باید جداگانه با هزینه قابل ملاحظه‌ای برنامه‌ریزی کرد، و چون برای مورد خاص برنامه‌ریزی می‌شود، دقیقاً نماینده عملکرد پزشکی مورد استفاده در مؤسسه‌ها نیستند. بنابراین ارزیابی فناوریهایی که در زمینه‌های پزشکی وبهداشتی صورت می‌گیرد، مشکل و پرهزینه است.

ازطرف دیگر، این مسئله در زمینه بررسیهای نمونه‌ای نیز وجود داشت. لازم بود که ابتدا

برای هر بررسی نمونه‌ای برنامه ریزی کرد. زیرا هیچ سازمان ملی برای این کار وجود نداشت، بنابراین بررسیها به‌واسطه هزینه‌های اولیه شروع و سازماندهی بخصوص گران بودند. اکنون سازمانهای ملی بسیاری وجود دارند که قادرند بررسیها را سریع و با هزینه کم انجام دهند. زیرا سازمان و تخصص لازم وجود دارد و برای شروع آمادگی دارد. ما برای ارزیابی فناوریهای پزشکی نیاز به چنین مقدماتی داریم. این ارزیابیها اساساً باید از طریق آزمایشها صورت گیرد که نسبت به بررسیهای نمونه‌گیری، انجام آنها بسیار مشکلتر است. بدین‌ترتیب مهم است که مجموعه‌ای از سازمانها (بیمارستانها همراه با شرکتها) را در اختیار داشته باشیم که آماده باشند از طریق انجام آزمایش با فناوریهای پزشکی ارزیابیهای لازم رابه‌اجرا درآورند.

مسائل

۱ـ بیماران جدید به‌طور تصادفی به شرکتها تخصیص داده می‌شوند تا بالاترین اطمینان رابه ما بدهند که گروههای بیماران تحت مراقبت برای هر گروه مشابه باشند. اگر این بیماران به مراقبت نیاز داشته باشند، از این مرحله به بعد به‌همان شرکت مراجعه می‌کنند. فکر می‌کنید با گذشت زمان گروههای بیمار مربوط به هر شرکت مشابه باقی می‌مانند؟ پاسخ خود را توجیه کنید.

۲ـ آیا می‌توانید این مفهوم را در مورد یک دبیرستان یا دانشکده به‌کار برید؟ به چه صورت این کار را انجام می‌دهید؟

۳ـ آیا می‌توانید سازمانهای دیگری را انتخاب کنید که در آنها این مفهوم را بتوان به‌کار برد؟ کدام سازمانها؟ چه‌طوراین کار را انجام می‌دهید؟

۴ـ در ریکاویک، پایتخت ایسلند، سه بیمارستان دولتی وجود دارد. کلیه بیماران اورژانس در روز اول به بیمارستان A، در روز دوم به بیمارستان B، در روز سوم به بیمارستان C می‌روند. سپس این دور به‌طور دائم تکرار می‌شود. وزارت بهداشت از شما خواسته است که درباره کیفیت و هزینه مراقبت در بیمارستانهای آنها نظر بدهید. چه‌کار می‌کنید؟ (شما می‌توانید کیفیت مراقبت را به‌وسیله نرخ مرگ ومیر، طول دوره بیماری، یا زمان کار از دست رفته اندازه‌گیری کنید.)

۵ـ شما مسئول متوقف کردن قاچاقچیان مواد مخدر در فرودگاه بین‌المللی میامی هستید که در

آنجا هزاران مسافر هر روز از بسیاری کشورها وارد می‌شوند. شما مسئول ۲۰۰ بازرس گمرک هستید. هر مسافر از یکی از ۲۰ ایستگاه بازرسی شما عبور می‌کند. شما ایده‌های زیادی درباره راههای مختلف پیداکردن مواد مخدر در اختیاردارید، اما مطمئن نیستید که این روشها واقعاً مؤثر باشند. (برای مثال، شما می‌توانید دستور دهید که هر چمدان بزرگ را باز و بازرسی کنند. اما این کار نسبت به اینکه فقط چمدانهای مشکوک را باز کنید وقت بیشتری می‌گیرد و مسافران موقعی که مجبور باشند زمان طولانی در صف بمانند ناراحت می‌شوند.) توضیح دهید چگونه می‌توانید دریابید که ایده‌های شما ایده‌های خوبی هستند؟

۶ـ شما معاون رئیس یک شرکت بین‌المللی هستید که مالک ۲۰۰۰ مغازه فروش بستنی است. معاون بخش بازاریابی علاوه بر بستنی قیفی قهوه‌ای رنگ معمولی می‌خواهد بستنی قیفی با طعم دارچین پیشنهاد کند که دارای خطوط مارپیچ قرمز رنگ است. رئیس شرکت از شما خواسته‌است طرحی رابرای ارزیابی پیشنهاد فوق و ایده‌های جدید دیگر بهتر کنید. چه‌کار خواهید کرد؟

منابع

J.B.Amberson. Jr. B.T. McMahon. and M. Pinner. 1931. "Clinical Trial of Sanocrysin in Pulmonary Tuberculosis." *American Review of Tuberculosis 24: 401-435.*

V. Cargill. D. Cohen. K. Kroenke. and D. Neuhauser. 1986. "Ongoing Patient Randomization: An Innovation in Medical Research." *Health Services Research 21: 663-678.*

D.I. Cohen. D. Breslau. cial. 1986. "The Cost Implications of Academic Group Practice: A Randomized Controlled Trial." *New England Journal of Medicine 314: 1553-1557.*

D.I. Cohen. P.Jones. B. Littenberg and D. Neuhauser. 1982. "Dose Cost Information Availability Reduce Physician Test Usage? A Randomized Clinical Trial with Unexpected findings." *Medical Care 20: 280-292.*

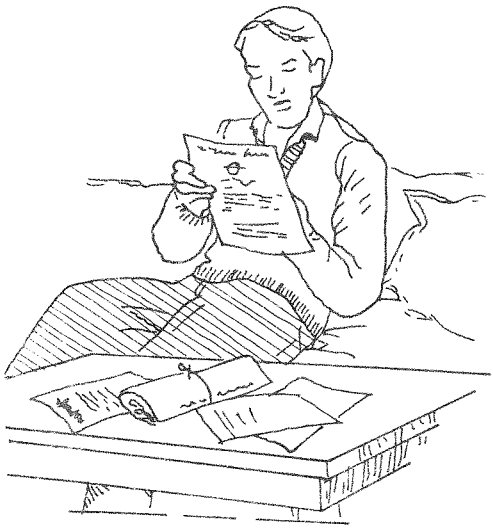
D.I. Cohen. B. Littenberg. C. Wetzel. and Neuhauser. 1982. "Improving Physician Compliance with Preventive Medicine Guidlines." *Medical Care 20: 1040-1045.*

D.I. Cohen. and D. Neuhauser. 1985. "The Metro Firm Trials: An Innovative Approach to Ongoing Randomized Clinical Trials." In Institue of Medicine. *Assessing Medical Technologies.* Washington. D.C. National Academy Press. PP. 529-534.

C. O. Hershey. D.K. Porter. D. Breslau. and D.I. Cohen. 1986. "Influence of Simple Computerized Feedback on Prescription Changes in an Ambulatory Clinic." *Medical Care 24: 472-481.*

C.O. Hershey. J.W. Tomford. etal. 1984. "The Natural History of Infavenouse Catheter Associated Phlebities." *Archives of Internal Medicine 144: 1373-1375.*

J.W. Tomford. etal. 1984. "Intravenous Therpy Team and Peripheral Venous Catheter Associated Complications: A Prospective Controlled Study." *Archives of Internal Medicine 144: 1191-1194.*



آزمایش بیمه درمانی

جوزف پی. نیوهاوس[۱] شرکت راند

اکثر مردم آمریکا از بیمه درمانی برخوردارند که به آنها کمک می‌کند تا صورتحسابهای بیمارستانی و پزشک را بپردازند. واژه مهم **کمک می‌کند** است. تنها یک بیمه‌نامه خیلی نامعمول کل صورتحسابهای مربوط به انواع خدمتها را خواهد پرداخت. اکثر بیمه‌نامه‌ها مستلزم "مشارکت درهزینه" از سوی بیمار از طریق سازوکارهایی مانند مبالغ کسرشدنی و بیمه

۱ـ *Joseph P. Newhouse*

مشارکتیند. (مبلغ کسرشدنی ۲۰۰ دلار برای هرشخص در هر سال الزام می‌کند که بیمار ۲۰۰ دلار نخست از هزینه‌های پزشکی در سال را بپردازد. بیمه مشترک ۲۰درصدی، شخص را ملزم می‌کند که ۲۰٪ از هر صورتحساب، مازاد بر هرنوع مبلغ کسرشدنی را بپردازد.) اما در برخی کشورها بیمه همگانی ممکن است همه هزینه‌های پزشکی رابدون مشارکت بیمار در هزینه پرداخت نماید.

کسی باید تصمیم بگیرد که مبالغ **کسرشدنی** و نرخهای بیمه مشارکتی چقدر باید باشند. یا اینکه اصلاً هیچ نوع مشارکت در هزینه وجود داشته باشد یا خیر. شخص تصمیم‌گیر ممکن است کارفرما یا وکیل یک اتحادیه باشد. در برنامه‌های مدیک‌ایر و مدیک‌ایدِ[۱]، تصمیم‌گیران عبارتند از: کنگره و رئیس جمهور. این اشخاص چگونه باید تصمیم بگیرند؟ چه نوع اطلاعاتی می خواهند داشته باشند؟ به نظر می‌رسد که آنها می‌خواهند بدانند با یک قرارداد بیمه‌ای سخاوتمندانه‌تر، و درنتیجه گرانتر، چه قدر تندرستی بیشتری به ارمغان آورده می‌شود؟

این موضوع برای دهها سال موجب مناقشه بوده است. برخی چنین استدلال کرده‌اند: اگر مردم مجبور باشند که هزینه مراقبتهای پزشکی را بپردازند، دریافت خدمتهای مورد نیاز رابه تعویق خواهند انداخت یا کلاً از آنها چشم خواهند پوشید، و سلامت آنها به خطر خواهد افتاد. کسانی که این گونه استدلال می‌کنند، معمولاً نتیجه می‌گیرند که مراقبت پزشکی باید رایگان باشد. آنها گرایش به این دارند که هزینه‌های اضافی مراقبتهای پزشکی را دست کم بگیرند، و این موضع را اختیار می‌کنند که مردم خوش ندارند به نزد پزشک بروند و تنها در مواقع نیاز واقعی به پزشک مراجعه می‌کنند.

دیگران درست برعکس آن استدلال می‌کنند: اگر مراقبت پزشکی رایگان باشد، مردم از آن سوءاستفاده خواهند کرد و بسیاری مراجعه‌های غیرضروری به پزشک خواهند داشت که تأثیری بر سلامتی آنها ندارد. درنتیجه، این گروه ادعا می‌کنند که مراقبت پزشکی رایگان بسیار پرهزینه خواهد بود، و آنها معمولاً نتیجه می‌گیرند که مردم باید قسمت عمده هزینه‌های مربوط

۱ـ برنامه‌های Medicare و Medicaid دونوع بیمه درمانی همگانی دولتی درایالات متحده اند.اولی برای افراد ۶۵سال به بالا ودومی برای کسانی است که توانایی مالی پرداخت هزینه های بهداشتی راندارند ـ (مترجم).

به مراقبت پزشکی خود رابپردازند، در حالی که بیمه، تنها هزینه‌های پزشکی بسیار بزرگ را باید بپردازد.

هر دو طرف بیشتر احساسی استدلال کرده‌اند تا با مدرک وسند. برای کسب اطلاعات بیشتر درباره این موضوعها، حکومت فدرال در سال ۱۹۷۴ یک آزمایش بزرگ، یعنی **آزمایش بیمه درمانی**، را به‌اجرا گذاشت. افراد شرکت‌کننده در این آزمایش به طور تصادفی به طرحهای بیمه‌ای متفاوت منتسب شدند که لازم داشت مبالغ متفاوتی رابرای مراقبت پزشکی خود بپردازند. هم تندرستی آنها و هم میزان مراقبت پزشکی مورد استفاده آنها به مدت چند سال تحت نظارت قرار گرفت. این آزمایش در زمینه سیاست بهداشتی از لحاظ اندازه وهدفهای آن، بی‌سابقه بود.

اکنون نگاهی می‌کنیم به مثالهایی از اینکه چگونه آمار برای طراحی آزمایش و تحلیل داده‌های گردآوری شده به منظور کسب بیشترین اطلاعات ممکن در ازای پولی که صرف آزمایش شده به‌کار رفت؟

سطح دانش در زمان شروع آزمایش

چرا در سال ۱۹۷۴ دانش ما تابدان حد اندک بود؟ بالاخره، بیمه درمانی اصلاً موضوع جدیدی نبود. تا سال ۱۹۷۴، زمانی که این آزمایش شروع شد، بیمه‌های خصوصی یا دولتی ۸۰٪ تا ۹۰٪ همه هزینه‌های بیمارستانی را برای دو دهه پرداخته بودند و بیش از نصف همه هزینه‌های پزشکها را بیمه پرداخت کرده بود. آدمی ممکن بود فکر کند که این مشاجره تا حد زیادی می‌توانست بدون یک آزمایش، صرفاً بابررسی کنشهای مردم با بیمه‌نامه‌های موجود، پایان یابد.

در حقیقت، می‌شد که از شرکتهای بیمه مبلغ حق بیمه‌هایی را برای قراردادهای بیمه‌ای متفاوت اخذ کرد، و مامی‌توانستیم این حق بیمه‌ها را برای برآورد کل هزینه هر شخص (شامل هرنوع مشارکت در هزینه) مربوط به گزینه‌های مختلف از برنامه‌های بیمه‌ای به‌کار ببریم. اما سعی در روشن کردن موضوع هزینه و آثار بیمه در سلامتی با استفاده ازداده‌های مربوط به بیمه‌نامه‌های موجود با سه مشکل روبه‌رو می‌شود:

۱ـ مردم قراردادهای بیمه‌ای رابا عنایت به اینکه محتملاً چقدر صرف مراقبت بهداشتی خواهند

کرد، انتخاب می‌کنند. کسانی که هزینه‌های زیادی را انتظار دارند، با انگیزه‌ای آشکار در پی قراردادهایی هستند که هزینه‌های پزشکی آنها را به‌طور نسبتاً گسترده‌ای بپوشاند. در قیاس با آن افراد خوش اقبالی که انتظار ندارند اصلاً به پزشکی مراجعه کنند، نفعی در پرداخت حق بیمه‌های زیاد برای پوششی تقریباً کامل نمی‌بینند وبدین‌ترتیب گرایش به انتخاب قراردادی با حق بیمه‌ای کم و مشارکت در هزینه زیاد دارند. در نتیجه، اگر سعی کنیم استفاده کسانی راکه اکنون در برنامه‌های سخاوتمندانه نیستند از تجربه اشخاصی که از طرحهای سخاوتمندانه برخوردارند، یا برعکس، استنباط کنیم، ممکن است به‌طور قابل توجهی آثار بیمه را در استفاده از خدمات و هزینه بیشتر برآورد کنیم.

۲ـ حتی اگر می‌توانستیم استفاده از خدمات (و بدین‌ترتیب هزینه‌ها) را تحت گزینه‌های مختلف از برنامه‌های بیمه‌ای به‌طور قابل اعتمادی ارزیابی کنیم، داده‌هایی درباره آثار گزینه‌های مختلف برنامه‌ها بر تندرستی نداشتیم. وانگهی، این نکته‌ای اساسی برای بسیاری از کسانی بود که به‌نظر آنان بیمه بهتر موجب می‌شود که مردم به‌دفعات بیشتر طالب خدمات پزشکی باشند، اما در اینکه چنین مراجعه‌ای برای گرفتن خدمت ضروری یا غیرضروری بود، باهم توافق نداشتند. حتی اگر اطلاعاتی درباره وضع سلامتی فعلی افراد به‌دست می‌آوردیم، چون وقتی مردم شغلشان را عوض می‌کنند یا بازنشسته می‌شوند، معمولاً بیمه تغییر می‌کند، بدین ترتیب نمی‌توانستیم استنباط کنیم که وضع فعلی سلامتی مرتبط با قرارداد بیمه‌ای فعلی بوده‌است یا گذشته.

۳ـ سرانجام، برخی ترتیبات بیمه‌ای اصلاً امتحان نشده بودند. مثلاً، بعضی ازپشتیبانان مشارکت در هزینه، فکر می‌کردند که مشارکت در هزینه برای فقرا باید تخفیف داده شود. اما فقرا نوعاً یا بیمه نداشتند یابه طور کامل از طریق برنامه مدیک‌اید ایالت خود تحت پوشش قرار داشتند. چگونگی واکنش فقرا نسبت به برنامه‌ای که الزام می‌کرد آنها، مثلاً تا ۵٪ درآمدشان را برای مراقبتهای پزشکی بپردازند، نمی‌شد از داده‌های موجود دریافت.

طراحی یک آزمایش اجتماعی

حدود ۷۷۰۰ نفر که در شش ناحیه مختلف کشور زندگی می‌کردند در آزمایش بیمه درمانی

شرکت کردند. این افراد از بیمه درمانی خودشان (اگر هم داشتند) یا برای مدت سه یا پنج سال استفاده نکردند و در عوض از یک برنامه بیمه درمانی آزمایشی استفاده کردند. سپس در پایان سه یا پنج سال، به بیمه بهداشتی معمولی خود برگشتند.

برخی شرکت‌کنندگان، مراقبتهای پزشکی رایگان دریافت کردند. دیگران ۲۵٪، ۵۰٪، یا ۹۵٪ از هزینه‌های پزشکیشان را، حداکثر تا ۱۰۰۰ دلار برای هر خانوار در هرسال از جیب خود پرداختند. این مقدار حداکثر برای خانواده‌های فقیرتر کاهش داده‌شد یا مشخصاً، این مبلغ نسبت به برنامه‌ای که به آن منتسب شده بودند، هرگز بیش از ۵٪، ۱۰٪، یا ۱۵٪ درآمد خانواده نبود.

شرح مجمل فوق از آزمایش بیمه درمانی، در جای خود انواع متعددی از تصمیم‌گیریها را برای طراحی آن شامل شده‌است. برخی موضوعهایی که درباره آنها تصمیم گرفته شده بود، عبارتند از:

۱) چند نفر شرکت خواهند کرد، و چند نفر به هرطرح منتسب خواهند شد؟
۲) چند مکان مورد استفاده قرار خواهند گرفت؟
۳) دقیقاً کدام افراد به کدام طرحها منتسب خواهند شد؟
۴) آزمایش چقدر طول خواهد کشید؟

تفکرآماری به طراحان کمک کرد که به این تصمیمها برسند.

شمار افراد در آزمایش

چون هدف هر آزمایش تولید اطلاعات است، طراح آزمایش در جستجوی آن است که اطلاعات حاصل در ازای پول خرج شده را ماکسیمم کند و کسانی را که هزینه آن را می‌پردازند، آگاه سازد که در ازای پول بیشتر چقدر اطلاعات بیشتر می‌توانند به‌دست آورند. یک راه برای اندازه‌گیری مقدار اطلاعات، دقت نتایج است، یا اینکه چقدر محتمل است که نتایج آزمایش (مثلاً در این آزمایش، شمار مراجعه‌ها به پزشک برای هر فرد در هرسال در یک برنامه آزمایشی) تقریبی از وضع واقعی جهان (مثلاً اگر همه افراد از برنامه بیمه خاصی برخوردار باشند، مردم چند دفعه به ازای هر شخص در هرسال خدمات درمانی خواهند خواست؟)

خواهد بود. احتمال نزدیک بودن نتایج آزمایشی به ارقام واقعی را همواره می‌توان با آوردن افراد بیشتر به درون آزمایش، افزایش داد. اما با آوردن افراد بیشتر به درون آزمایش بازده کار کاهش می‌یابد، برای مثال، افزایش نمونه از ۵۰ فرد به ۱۰۰ فرد بازده بیشتری دارد تاافزودن ۵۰ نفر دیگر و رساندن تعداد به ۱۵۰ نفر. درواقع دقت هر نمونه مفروض به‌طورکلی متناسب با ریشه دوم اندازه نمونه است.

هرچند که نمونه‌های بزرگتر نتایج دقیقتری می‌دهند، اما هزینه بیشتری هم دارند. درست مانند شخصی که خانه‌ای می‌سازد، باید تصمیم بگیرد که چقدر زیربنا می‌خواهد داشته باشد؟ شخصی هم که هزینه آزمایشی رامی پردازد، باید تصمیم بگیرد در پی چقدر دقت است؟ همچنانکه معمار نمی‌تواند بگوید؛ یک خانه چقدر باید بزرگ باشد، طرح آزمایش نیز نمی‌تواند بگوید که چقدر دقت باید حاصل شود.

گاهی به اظهارنظرهایی برمی‌خوریم که گویا نوعی آستانه بحرانی برای اندازه نمونه وجود دارد و اگر نمونه بیش از آن تعداد نباشد، بررسی "از نظر علمی" معتبر نیست. این مفهوم نادرست است، مانند این است که بگوییم ساختمانی باید بیش از چند مترمربع فضاداشته باشد وگرنه خانه نیست.

شمار افراد در هر برنامه

اکنون فرض کنید که بودجه آزمایشگر ما تعیین شده‌است. در این صورت طراح لازم است تصمیم بگیرد که چند نفر باید به هر برنامه بیمه آزمایشی اختصاص یابند. واکنش نخست ممکن است این باشد که تعداد مساوی از افراد به هر برنامه اختصاص داده شود، اما فرض کنید که یک برنامه به ازای هر شخص ثبت‌نام کرده بسیار پرهزینه‌تر از دیگری باشد.مثلاً برنامه بیمه‌ای با خدمات پزشکی رایگان ممکن است پرهزینه‌تر از برنامه‌ای باشد که افراد باید ۹۵٪ هزینه‌ها را بپردازند. در این صورت شمار برابری را به برنامه‌ها تخصیص نمی‌دهیم بلکه افراد را بر طبق ریشه دوم عکس نسبت هزینه اختصاص می‌دهیم. مثلاً اگر یک برنامه بیمه چهار برابر بیشتر از دیگری هزینه داشته باشد، به اندازه نصف برنامه دیگر، افراد به آن اختصاص داده خواهند شد. از این نظر ریشه دوم نسبت هزینه‌ها را به‌کار می‌بریم که دقت یک نمونه به تناسب ریشه دوم

اندازه نمونه افزایش می‌یابد.

شمار مکانها

ملاحظات هزینه‌ای بر تعداد مناسب مکانها نیز تأثیر می‌گذارد. صرفاً برای ورود به یک مکان برخی هزینه‌ها باید پرداخته شوند، مثلاً باید یک برنامه نمونه‌گیری تدوین شود و شاید لازم باشد فضایی برای دفتر محلی اجاره شود. بنابراین به ازای هر بودجه معین، هرقدر شمار مکانها بیشتر باشد، کل اشخاصی که می‌توان ثبت‌نام کرد، کمتر خواهد بود. نخستین فکر ما ممکن است این باشد که با انتخاب همه افراد از یک مکان واحد، شمار افراد ثبت‌نام شده را ماکسیمم کنیم، اما ممکن است وقتی می‌خواهیم تصمیم بگیریم که کدام مکان را انتخاب کنیم، نظر خود را عوض کنیم. اگر همه مکانها شبیه هم باشند، یک مکان بهینه خواهدبود. اما چون مکانها (مثلاً از لحاظ اینکه مردم چه مدت در بیمارستان می‌مانند؟) باهم فرق دارند، با داشتن مکانهای بیشتر اطلاعات بیشتر کسب می‌کنیم. به‌طور کلی، هرقدر مکانها باهم تفاوت بیشتر داشته باشند، باید مکان بیشتر و کلاً شمار افراد کمتری داشته باشیم. هزینه نیز دخیل است، هرقدر هزینه ورود به مکانی جدید نسبت به هزینه‌های افزودن شخصی دیگر در یک مکان قدیم بیشتر باشد، شمار کمتر مکانها بهینه خواهدبود.

طول مدت آزمایش

طراحی که دارای بودجه معینی است، همچنین باید تصمیم بگیرد که شمار کمتری از افراد را برای مدت زمان طولانی‌تری در برنامه بگنجاند یا از شمار بیشتری از افراد برای مدت زمانی کوتاهتر از دو نظر، ثبت‌نام افراد بیشتر برای مدت زمان کوتاهتر ارجح است. وضع رفتار شخص معمولاً در طول زمان تکرار می‌شود (مثلاً کسی با جوشهای شدید صورت ممکن است مراجعه‌هایی مکرر به پزشک داشته باشد). بنابراین دو سال مشاهده یک شخص به اندازه یک سال مشاهده دو نفر ارزش ندارد زیرا پیشاپیش از سال اول سرنخی در دست داریم که آیا شخصی زیاد به پزشک مراجعه می کند؟ یانه. افزون برآن، ثبت‌نام افراد بیشتر برای مدت زمان کوتاهتر اطلاعات حاصل از آزمایش را زودتر در دسترس قرار می‌دهد.

اما، اگر این استدلالها را تا نهایت منطقی آنهاپیش ببریم، به معنی آن خواهدبود که تاآنجاکه بودجه اجازه می‌دهد، شمار بیشتری، هریک را برای یک روز، در برنامه بگنجانیم! این کار به دلایل روشن بی‌معنی خواهدبود. وقوع برخی از رفتارهایی که می‌خواهیم آنها را بسنجیم زمان می‌برند؛ برای مثال از آنچه بر سر وضع سلامت شخص می‌آید، می‌توان نام برد. بدین ترتیب، می‌خواهیم آزمایشمان تا حد امکان کوتاه بوده و سازگار بااین شرط باشد که مدت زمانی به قدر کافی طولانی در اختیار بگذارد که امکان دهد تابرنامه بیمه‌ای اثر خودش را داشته باشد.

طراح هنگام انتخاب مدت زمان لازم برای اجرای آزمایش، باید درنظر داشته باشد که اشخاص ثبت‌نام کرده در یک آزمایش ممکن است به آن گونه‌ای که در صورت نامحدود بودن مدت بیمه عمل می‌کردند، رفتار نکنند. مثلاً افراد دارای یک برنامه بیمه آزمایشی که در تاریخ معینی پایان خواهد یافت، بسته به اینکه وضعیت آنها پس از آزمایش چه خواهدبود؟ ممکن است برای دریافت کمک پزشکی فشار بیاورند یا خدمات معینی را به تعویق اندازند. از آنجاکه این رفتار در یک برنامه (غیرآزمایشی) واقعی مشاهده نخواهد شد، می‌تواند نتایج را مغشوش کند. راهکاری که دراین آزمایش بیمه بهداشتی به‌کار بسته شد تا میزان مغشوش بودن را ارزیابی کند، عبارت بود از اینکه افراد را برای مدت زمانهای مختلف ثبت‌نام کنند، که تعدادی برای سه سال و تعدادی برای پنج سال بود. هرنوع ازدحام برای استفاده ازخدمات یا تعویق آنها در پایان آزمایش از راه مقایسه آخرین سال یک گروه سه ساله با سومین سال گروه پنج ساله‌ای که در همان زمان شروع شده بود، اندازه‌گیری شد. به طور مشابه، به‌وسیله ثبت‌نام در مکانهای مختلف اول از یک گروه پنج ساله وسپس دو سال پس ازآن یک گروه سه ساله را، هر نوع ازدحام اندازه گیری شد.

خودداری کنندگان

طراح هر آزمایشی باید درنظر داشته باشد که (به طورکلی) مردم را نمی‌توان به شرکت در آزمایشی مجبور کرد بلکه باید رضایت آنها جلب شود. افزون بر آن، طراح الزام اخلاقی و قانونی دارد که شرکت‌کنندگان بالقوه را از منافع و خطرهای شرکت کردن آگاه سازد.اگر چه رضایت شرکت‌کنندگان جلب می‌شود و به آنها آگاهی داده می‌شود، در هر حال طراح، منافع و

شاید خطرهایی را که بر تصمیمهای شرکت‌کنندگان بالقوه تأثیر دارند، خود کنترل می‌کند.

برای مینیمم کردن تعداد افراد خودداری کننده، مجری آزمایش بیمه بهداشتی مبلغ کافی به افراد پول پرداخت تا آنها را مطمئن کند که به خاطر شرکت در آزمایش از لحاظ مالی چیزی را از دست نخواهند داد. برای مثال ، در یک برنامه آزمایشی شخص باید ۵۰٪ هر هزینه بیمارستانی را حداکثر تا ۱۰۰۰ دلار در هرسال بپردازد. فرض کنید که برنامه بیمه قبلی شخص مذکور هرنوع بستری شدن در بیمارستان را کاملاً می‌پوشانده است. در این مورد مجری آزمایش، صرف‌نظر از اینکه این شخص به بیمارستان برود یا خیر، ۱۰۰۰ دلار در سال به او می‌پرداخت. اگر مبلغ پرداختی کمتر از ۱۰۰۰ دلار می‌بود، افرادی که با درجه حتمیت بالایی انتظار داشتند به بیمارستان بروند (مثلاً زنان باردار) احتمالاً از ثبت‌نام کردن خودداری می‌کردند زیرا انتظار داشتند پولی را از دست بدهند.

یقیناً، پرداخت جداگانه، خود ممکن بود که رفتار را تغییر دهد و اگر چنین می شد، اهمیت داشت که آن را بدانیم. زیرا پرداخت جداگانه بخشی از برنامه بیمه‌ای "جهان حقیقی" نخواهد بود. به‌منظور بررسی اثر پرداخت، طراح می‌تواند برخی تغییرات در پرداخت ایجاد کند، مثلاً با انتخاب تصادفی بعضی خانواده‌ها به‌جای ۱۰۰۰ دلار، ۱۵۰۰ دلار دریافت کنند. درواقع چنین تغییراتی در آزمایش بیمه خدمات درمانی گنجانده شد و معلوم گردید که پرداختها اثر قابل اغماضی بر رفتار داشتند.

تحلیل داده های آزمایشی

جدول ۱، استفاده از خدمات درمانی و میزان هزینه هر فرد را برحسب برنامه بیمه در **آزمایش بیمه خدمات درمانی** نشان می‌دهد. روشن است که موضوع اینکه شخصی تا چه حد بیمه خوبی داشته باشد، در این امر که او از چند بار خدمات پزشکی استفاده می‌کند، تأثیر فراوان دارد. اشخاصی که ۹۵٪ هزینه‌ها را می‌پرداختند ۴۰٪ کمتر از آنهایی که هیچ مبلغی نمی‌پرداختند به پزشک مراجعه کردند. به طور مشابه، استفاده از بیمارستان تا ۲۳٪ و کل هزینه تا ۳۱٪ کاهش داشت.

می‌خواهیم ارزیابی کنیم که آیا آن خدمات اضافی که توسط گروه واقع در برنامه خدمات

جدول ۱ - معیارهای استفاده از خدمات درمانی برای هر فرد در هرسال به صورت تابعی ازبیمه[۱]

هزینه[۲] (دلار ۱۹۸۴)	پذیرش در بیمارستان در هر صدنفر	مراجعه به پزشک	درصد هزینه پرداخت شده توسط شخص[۳]
۷۷۷	۱۲/۸	۴/۵۵	۰
۶۳۰	۱۰/۵	۳/۳۳	۲۵
۵۸۳	۹/۲	۳/۰۳	۵۰
۵۳۴	۹/۹	۲/۷۳	۹۵

(۱) هرچند که در اینجا نشان داده نشده‌است، دقت این اعداد کاملاً خوب است. برای آگاهی از جزئیات ن.ک. مانینگ و همکاران (۱۹۸۷).

(۲) شامل مبالغی است که از سوی بیمه و از سوی شخص پرداخت شده‌اند. این اقلام هزینه‌های خدمات دندانپزشکی و روان‌درمانی سرپایی را دربر ندارد. خدمات دندانپزشکی و روان‌درمانی سرپایی تقریباً همین درجه از پاسخ‌دهی به برنامه را نشان دادند. گنجاندن آنها در هزینه، اندکی بیش از یک سوم به هریک از ارقام خواهد افزود. عمده مبالغ اضافی برای خدمات دندانپزشکی خواهد بود. اعداد این ستون از متوسط گرفتن هزینه پیش‌بینی شده حاصل شده‌است؛ ن.ک. مانینگ و همکاران (۱۹۸۷).

(۳) درصد پرداخت شده از ماکسیمم سالانه از جیب شخص که عبارت بود از ۱۰۰۰ دلار، یا۵٪، ۱۰٪، یا۱۵٪ درآمد، هرکدام که کمتر بود؛ درصد درآمد به طورتصادفی منتسب شد.

رایگان دریافت شده، تأثیری بر سلامتی آنها داشته است؟ یا خیر. شواهد قابل اعتماد درباره این موضوع قبلاً هرگز در دسترس نبوده‌اند. داده‌های مندرج در جدول ۲ نشان می‌دهد که سلامت فرد متوسط کمتر تحت تأثیر تفاوتهای موجود درنوع استفاده‌ای است که در جدول ۱ تشریح شده‌است.

(در جدول ۲ برنامه‌های بیمه مشارکتی ۲۵٪، ۵۰٪، و ۹۵٪ را برای آسانی تعبیر در یک گروه جای داده‌است. برآمدهای این سه برنامه برحسب همه معیارها شبیه به‌هم بودند.)

مقدارهایی که درجدول ۲ نشان داده شده‌اند، سه مورد از معیارهای متعددی هستند که

گردآوری شدند؛ آنها شامل دو معیار، فشارخون و وضع بینایی هستند که در مورد آنها می‌توانیم تا حد معقولی مطمئن باشیم که خدمات اضافی به بهبود سلامتی انجامیده‌اند. (برخی معیارهای سلامت دندان نیز بهبود یافتند.) لیکن معیار کلی تندرستی، یعنی **ادراکهای تندرستی**، برای گروه افرادی که بخشی از هزینه خدمات درمانی خود را پرداختند، اندکی بهتراست (۶۸/۰ در برابر ۶۷/۴). اما مطلب مهمتر این سئوال است که این برآورد تا چه حد دقیق است؟ بااستفاده از تحلیل آماری می‌توانیم نتیجه بگیریم که هر تفاوت واقعی برحسب ادراکهای تندرستی کلی بین این دوگروه محتمل است که کوچک باشد؛ مشخصاً، محتمل است که بین ۱/۵ امتیاز (که پشتیبان مشارکت درهزینه است) و۰/۳- امتیاز (که حامی خدمات رایگان است) باشد. به‌طور مشخصتر، می‌توانیم بگوییم، اگر آزمایش مشابهی به دفعات بسیار اجرا می‌شد، و از همین شیوه برای ساختن این نوار استفاده به عمل می‌آمد، ۹۵٪ نوارهای حاصله هر تفاوت واقعی بین دو گروه را دربر می‌گرفت. (البته ممکن است هیچ تفاوتی وجود نداشته باشد.) چون حد بالایی این نوار در این مورد ۰/۳ امتیاز است، می‌توانیم به‌طور معقولی حتم داشته باشیم که نامحتمل است که برنامه رایگان بیش از ۰/۳ امتیاز بهتر از برنامه مشارکت در هزینه باشد.

البته، درباره اینکه یک تفاوت ۰/۳ امتیازی چه معنایی ممکن است داشته باشد، هیچ بینشی نداریم. یک راه برای نیل به یک بینش آن است که به همه افراد موجود در بررسی که دارای نوعی بیماری، مثلاً فشارخون بالا هستند، نگاه کنیم. اگر برای تفاوتهای قابل مشاهده بین افرادی که فشار خون بالا دارند و ندارند، تعدیل انجام دهیم (یعنی برای گروه سنی بالای دارای فشار خون تعدیل انجام دهیم) گروه دارای فشار خون بالا دارای مقداری برای ادراکهای کلی تندرستی است که در حدود ۵ امتیاز کمتر از مقدار مربوط به گروهی است که بدون فشار خون بالاست. (برای توضیح عمل تعدیل ن.ک. مقاله موزس و *ماستلر*.) چون ۵ امتیاز نسبت به ۰/۳ امتیاز، بزرگ است، یک تعبیر از این نتایج آن است که آن خدمات اضافی که اشخاص واقع در برنامه خدمات رایگان دریافت کردند، آنها را از نظر خودشان به‌طور قابل توجهی تندرست‌تر نکرد. آنها ممکن است خودشان را به‌اندازه خیلی کمی تندرست‌تر گمان کرده باشند اما اندازه‌گیریها نتوانستند آن مقدار خیلی کم را کشف کنند، ولی خلاف آن نیز می‌تواند درست باشد.

جدول ۲- معیارهای برآمد تندرستی به عنوان تابعی ازبرنامه بیمه[۱]

نوع برنامه بیمه	فشارخون دیاستولی (میلیمترجیوه)	ادراکات[۲] تندرستی کلی	معیار قدرت بینایی[۳] (خطهای اسندلن)
خدمات رایگان	۷۷/۹	۶۷/۴	۲/۴۲
مشارکت درهزینه	۷۸/۸	۶۸/۰	۲/۵۲

(۱) این مقدارها از روی معادله رگرسیونی پیشگویی شده‌اند و تا حدی دقیقتر از میانگینهای ساده‌اند. برای جزئیات ن.ک. بروک و همکاران (۱۹۸۳).

(۲) این مقیاسی است که از ۰ تا ۱۰۰ تغییر می‌کند (۱۰۰ تندرست‌ترین حالت است) و توصیف می‌کند که مردم تندرستی خود را چگونه ارزیابی می‌کنند. این مقیاس بر پایه پاسخ به ۲۲ پرسش مختلف بنانهاده شده‌است. با مساوی بودن سایر چیزها، یک تفاوت ۵ امتیازی در این مقیاس معادل با داشتن فشار خون خفیف است؛ یک تفاوت ۷ امتیازی، به شرط ثابت ماندن سایر چیزها، معادل با داشتن قند خون است.

(۳) برآمد ۲ متناظر است با قدرت دید تصحیح شده ۲۰/۲۰ در چشم باید بهتر، برآمد ۳ متناظر با ۲۰/۲۵، و برآمد ۴ با ۲۰/۴۰ است، بدین‌ترتیب اعداد کوچکتر مرجح‌اند.

برای اشخاص دارای مشکل فشارخون بالا وقدرت دید کم، به‌طور نسبی می‌توانیم مطمئن باشیم که اثرهای سودمندی رخ داده‌اند. این اثرها در بین گروه متشکل از افراد فقیر تمرکز داشتند.

اگر کمی بیندیشیم، شاید شگفت‌انگیز نباشد که خدمات اضافی در مورد این بیماری تأثیر مثبت داشته است. مشکلات فشارخون بالا و قدرت دید کم (و همچنین بیماریهای دهان و دندان) نسبت به سایر بیماریها شایعترند و تشخیص و درمان آنها بالنسبه ارزانتر است. افزون بر آن، معقول است که افراد فقیر با تفاوت محسوسی ازآن بهره گیرند. خواننده‌ای دقیق ممکن است بپرسد که اگر این اثرها بر روی فشار خون و قدرت دید وجود داشتند، چرا در معیاری مانند ادراکهای تندرستی کلی بازتاب نیافته‌اند؟ یک پاسخ آن است که نسبت افراد مبتلا به این مشکلات و مزایای خدمات رایگان نسبت به مشارکت در هزینه برای این مشکلات به قدرکافی کوچک هستند که وقتی از یک معیار تندرستی نسبت به کل جامعه متوسط گرفته می‌شود، با

اندازه نمونه موجود در آزمایش، اثرها قابل کشف نباشند. امکان دیگر این است که کسری از خدمات اضافی در برنامه خدمات رایگان با برآمدهای بد (مثلاً واکنشها نسبت به دارویی که تجویز شده بود) به هم پیوند خورده‌اند. این برآمدهای بد ممکن است سایر برآمدهای خوب را خنثی کرده باشد.

تأثیر نتایج

چون یافته‌های آزمایش مهم و منحصربه فرد بودند، مخصوصاً آنهاکه در جدول ۲ درج شده‌اند، به نظر می‌رسد اثری قابل توجه داشته باشند. نتایج مذکور در جدول ۱ در دسامبر ۱۹۸۱، و از آن جدول ۲ در دسامبر ۱۹۸۳ چاپ شدند. بین ۱۹۸۲ و ۱۹۸۴، در میزان مشارکت در هزینه‌های بیمه درمانی خصوصی درایالات متحده، بویژه برای خدمات بیمارستانی، افزایشی قابل توجه وجود داشت. نرخهای پذیرش بیمارستانی، و نیز مراجعه به پزشک نیز کاهش یافت. جدول ۳ این تغییرات را نشان می‌دهد. با استفاده از نتایج این آزمایش، می‌توانیم استنباط کنیم که کاهشها در استفاده از خدمات در واکنش به مشارکت بیشتر در هزینه بیمه خصوصی احتمالاً اثر بدی بر تندرستی نداشته یا اثر منفی آن کم بوده‌است، زیرا بسیاری از مردمی که دارای بیمه خصوصی هستند، فقیر نیستند.

البته نمی‌توانیم همه تغییرات راکه در جدول ۳ نشان داده شده‌اند به اطلاعات فراهم آمده از این آزمایش نسبت دهیم؛ درواقع هرگز نخواهیم دانست که آزمایش بالا چه نقشی در وقوع این تغییرات داشته است. اما نتایج آزمایش برای مشاورینی که طرف مشورت شرکتها درباره مزایای برنامه‌های خدمات درمانی هستند، کاملاً شناخته شده‌اند و برخی شرکتها حتی این نتایج را در بروشورهایی برای کارمندانشان نقل کرده‌اند تا تغییرات مذکور را تبیین کنند. بدین ترتیب منطقی است که بگوییم این آزمایش نقشی مهم در وقوع این تغییرات ایفا کرده‌است.

این آزمایش بااحتساب همه چیز، ۸۰ میلیون دلار هزینه داشت. آیا این اطلاعات ارزش آنقدر پول را داشت؟ اگرچه ۸۰ میلیون دلار قطعاً مبلغ معتنابهی است، در قیاس با ۲۳۰ میلیارد دلارکه ایالات متحده تنها در سال ۱۹۸۴ برای خدمات بیمارستانی و پزشکان خرج کرد، رنگ می‌بازد. جدول نشان می‌دهد که بین ۱۹۸۲ و۱۹۸۴ تعداد مرخصی از بیمارستان حدود ۱۰٪ و

جدول ۳- تغییرات بین ۱۹۸۲ و ۱۹۸۴ در بهره‌گیری از بیمه و خدمات درمانی

	۱۹۸۲	۱۹۸۴
درصد برنامه‌های بیمه خصوصی بامبلغ کسرشدنی برای خدمات بیمارستانی	۳۰	۶۳
درصد برنامه‌های بیمه خصوصی بامبلغ کسرشدنی ۲۰۰دلار یا بیشتر برای هرشخص در هر سال	۴	۲۱
مرخصی از بیمارستان به ازای هر ۱۰۰نفر زیر ۶۵سال[۱]	۱۱/۸	۱۰/۷
نرخ مراجعه به پزشک برای هر نفر زیر ۶۵ سال	۴/۹	۴/۷

(۱) زایمانها را دربر ندارد

مراجعه به پزشک حدود ۴٪ افت کرد. اگر نتایج این آزمایش حتی بخش کوچکی از این دگرگونی را باعث شده‌باشد -و بر سلامت افراد کم اثر داشته یا هیچ اثری نداشته است- بازهم آزمایش هزینه خودش را در طول چند ماه معدود درآورده است!

مسائل

۱ـ بحث زیر درباره قیمت‌گذاری خدمات تلفنی صورت گرفته است: برخی مردم گمان می‌برند که با یک نرخ ثابت در هرماه برای خدمات درون شهری، مردم زمان خیلی بیشتری را پای تلفن صرف خواهند کرد تاموقعی که برای هر دقیقه باید پول بپردازند. دیگران می‌گویند که این اثر چندان زیاد نیست. آیا شما فکر می‌کنید که یک آزمایش، راهی خوب برای ارزیابی این مطلب خواهد بود که مردم درصورتی که ناچار باشند برحسب دقیقه پول بپردازند، طرز استفاده آنها از تلفن چقدر ممکن است تغییر کند؟ چرا آری؟ چرانه؟ (در نظر داشته باشید که هم‌اکنون برخی خانوارها برحسب دقیقه می‌پردازند درحالی که خانوارهای دیگر نرخ ثابتی را می‌پردازند.)

۲ـ فرض کنید که می‌خواستید برای اینکه مقدار استفاده از تلفن چقدر نسبت به لغو حسابهای بانرخ ثابت واکنش نشان می‌دهد، آزمایشی انجام دهید. برخی تصمیمهای مربوط به طرح را که اتخاذ می‌کنید، نام ببرید. چگونه مردم را ترغیب خواهید کرد که دراین کار شرکت کنند؟

۳ـ آیا فکر می‌کنید یک آزمایش بهترین راه برای ارزیابی تأثیر گسترش سازمانهای حفاظت از سلامت بر روی قیمتهای خدمات پزشکی و هزینه‌ها خواهد بود؟ چرا آری؟ چرانه؟

۴ـ چگونه دقت با اندازه نمونه تغییر می‌کند؟ آیا اندازه‌ای از نمونه وجود دارد که یک بررسی را از لحاظ علمی معتبر سازد؟ درباره پاسخ خود توضیح دهید.

۵ـ چه موقع باید شمار مساوی از افراد رابه تیمار آزمایشی منتسب کرد؟

۶ـ تعداد مکانهای یک آزمایش چگونه باید انتخاب شوند؟

۷ـ یک آزمایش چه مدت باید اجرا شود؟ درباره پاسخ خود توضیح دهید.

۸ـ تهدیدهای مربوط به اعتبار نتایج آزمایشی که ناشی از موارد خودداری از شرکت در آزمایش هستند، چه چیزهایی می‌باشند؟

منابع

R. Archibald and J.P. Newhouse. In press. " Social Exqerimentation Some why's and How's." In E Quade and H.Miser, eds. *Handbook of systems Analysis. vol. II.* New York: North Holland.

R.H.Brook, J.E. Ware, Jr., W.H. Rogers, E.B.Keeler, A.R.Davis, C.A.Donald, G.A.Goldberg, K.N.Lohr, P.C.Masthag, and J.P.Newhouse. 1983. "Does Free Care Improve Adult´s Health? Results from a Randomised Control Trial"*New England Journal of Medicine* 309:1,426-1,434.

W.G.Manning, J.P.Newhouse, N.Duan, E.B.Keeler, A.Leibowits, and M.S.Marguis. 1987. "Health Insurance and the Demand for Medical Care: Results from a Randomised Eperiment." *American Economic Review* 77:251-277.



قیمت سیگار، مصرف سیگار و سیاست مالیات بر مصرف

کنت ای. وارنر[۱] دانشگاه میشیگان

در سال ۱۹۸۶ کنگره آمریکا دوبرابر شدن موقتی مالیات بر مصرف سیگار را به یک تغییر دائم تصویب نمود. انگیزه اصلی نیاز به درآمد بیشتر بود. ملاحظه دیگر کنگره این باور بود که مالیات بیشتر، صدهاهزار آمریکایی را از شروع به سیگارکشیدن و یا ادامه آن باز خواهدداشت. بویژه کنگره تصور می‌کرد که افزایش مالیات بیش از همه بر کودکان و نوجوانان اثر خواهد گذاشت.

۱- *Kenneth E. Warner*

اگر چنین تصوراتی درست می‌بود، افزایش مالیات درنهایت ممکن بود باعث جلوگیری از دهها هزار مرگ زودرس ناشی از مصرف سیگار و بسیاری از ناتوانیها و بیماریهای جانبی ناشی از سیگارکشیدن شود.

انتظارات کنگره در هردو مورد درآمد اضافی دولت و تأثیر مالیات برمصرف از مطالعات کشش قیمتی تقاضا ـ معیاری از واکنش مصرف‌کنندگان نسبت به تغییرات قیمت ـ توسط اقتصاددانان ناشی شده بود. با بررسی آمار مصرف سیگار و قیمت آن در زمان و برای ایالتهای مختلف، اقتصاددانان یک فن آماری را به نام تحلیل رگرسیونی چندمتغیره (به شرح ذیل) برای برآورد اثری که تغییر قیمت بر فروش سیگار دارد، به‌کار برده‌اند. از این برآوردها، سایر تحلیل‌گران، برای پیش‌بینی آثار مصرفی و درآمدی تغییر مالیات استفاده کرده‌اند. آنگاه برآوردهای آثار مصرفی جامعه با داده‌های مربوط به بیماریهای مسری ناشی از مصرف سیگار برای برآورد تعداد مرگ ومیرهای نابهنگام پیشگیری‌شده به سبب افزایش مالیات و یا در اثر کاهش مالیات تلفیق شد.

مطلب خاصی درمورد استفاده از تجزیه وتحلیل رگرسیونی در برآورد کشش تقاضابرای سیگار، همین‌طور استفاده از برآوردهای حاصل برای تأثیرگذاری سیاستی وجود ندارد. در بخش خصوصی، اقتصاددانان روزانه زحمت زیادی می‌کشند تا تقاضای مصرف‌کننده را به قیمت‌گذاری محصول مربوط نمایند. برای مثال قیمت درنظر گرفته‌شده برای یک خودرو نو تاحدودی ممکن است با تحلیل رگرسیون ـ که آیا یک افزایش (یاکاهش) قیمت اثر بیشتری بر درآمد هر واحد کالا یا بر حجم فروش خواهدداشت ـ مشخص شود. درآمد کل (TR، ارزش کل فروش خودروها) با قیمت یک خودرو (یعنی درآمد واحد، P) ضرب‌در تعداد، باکمیت (Q) خودروهای فروخته‌شده برابر است. بنابراین، به‌عنوان مثال، اگر یک کاهش ۲٪ در قیمت، فروش خودرو را بیش‌از ۲٪ (مثلاً ۳٪) افزایش دهد، درآمد کل افزایش خواهدیافت. به دلیل اینکه ۰/۹۸P×۱/۰۳ Q>PQ، زیرا ۱ < ۱/۰۰۹۴=۱/۰۳×۰/۹۸، که ۰/۹۸ P به معنای ۲٪ کاهش در قیمت و ۱/۰۳Q بیانگر ۳٪ افزایش فروش می‌باشد. اگر کاهش قیمت، برحسب درصد، بیشتر از افزایش حجم فروش شود، درآمد کل کمتر خواهد شد. درمثال پیش، اگر ۲٪ کاهش قیمت (۰/۹۸ P) باعث فقط ۱٪ افزایش فروش (۱/۰۱ Q) شود.

۰/۹۸P×۱/۰۱Q<PQ (یعنی ۱ > ۰/۹۸۹ = ۰/۹۸×۱/۰۱) اینکه اثر قیمت یاحجم فروش کدام یک بیشتر خواهدبود؟ باید بااستفاده‌از بهترین داده‌هاو روشهای موجودآماری برآورد شود.

در این مقاله به بررسی اینکه چگونه کشش تقاضا تعریف و محاسبه می‌شود؟ چگونه تجزیه و تحلیل آماری برآورد آن را مقدور می‌سازد؟ و چگونه می‌توان از آن برای پیش‌بینی درآمد مالیاتی و آثار مصرفی سیگار در اثر تغییر مالیات برمصرف سیگار فدرال استفاده نمود؟، می‌پردازیم. باتوجه به مطالعات منتشرشده درباره کششها، برآوردهای مشخصی از تعداد کودکان، نوجوانان و بزرگسالانی که در اثر تصمیم‌گیری ۱۹۸۶ مبنی بر دوبرابر نمودن دائمی مالیات سیگار از مصرف سیگار روی گردانیده‌اند، ارائه خواهیم نمود. همچنین خواهیم دید که چگونه از تحلیل رگرسیون و برآورد کشش می‌توان در اطلاع‌رسانی نسبت به سیاستهای مالیاتی و بهداشتی استفاده کرد.

تعریف و محاسبه کشش قیمتی تقاضا

کشش قیمتی تقاضا برای یک کالا سنجشی است از میزان تغییری که مصرف‌کنندگان در تقاضای خود برای آن کالا درقبال تغییر در قیمت آن می‌دهند. به‌طور فنی کشش قیمتی عبارتست از درصد تغییر در مقدار تقاضای کالا در واکنش به ۱٪ تغییر درقیمت آن.

در جایی که **کشش** یک اصطلاح فنی در اقتصاد می‌باشد، معنای کیفی آن درواقع کاملاً مشابه استفاده روزانه ما از واژه **باکشش** است: یک نوار لاستیکی باکشش است، به این معنا که هرگاه به دوطرف کشیده شود، به میزان قابل توجهی واکنش نشان می‌دهد (کش می‌آید). درمقابل یک مداد بی‌کشش است؛ بدین معنا که وقتی از دوطرف کشیده شود، واکنشی نشان نمی‌دهد. به‌طور مشابه تقاضا برای یک کالا **باکشش** است اگر به‌طور نسبی بیشتر از قیمتی که باعث تغییر آن شده‌است، تغییر یابد (قیمت در اینجا همان نیروی وارده است). تقاضا بی‌کشش است اگر به اندازه تغییر (به درصد) قیمت محصول تغییر نکند. درنهایت، تقاضا کاملاً بی‌کشش است اگر به هیچ‌وجه با تغییر قیمت، تغییر پیدا نکند. این مشابه بی‌کشش بودن یک لوله فلزی سنگین است که هرگاه شخصی بخواهد؛ آن را با دست خالی خم می‌نماید.

معمولاً وقتی که قیمت یک کالا کاهش یابد، [مقدار] تقاضا برای آن افزایش پیدا می‌کند و

برعکس با افزایش قیمت، [مقدار] تقاضا کاهش می‌یابد. این رابطه معقول به‌نظر می‌رسد. همه می‌خواهیم چیزی را بیشتر بخریم که ارزانتر است و وقتی که گرانتر می‌شود، تمایل به خرید کمتر وجود دارد. به‌وضوح نسبت به برخی از قیمتها بیشتر حساس هستیم. به‌عنوان مثال، یک افزایش ۲۵٪ در قیمت نمک ممکن است مقدار نمکی را که می‌خریم، تغییر ندهد. (تقاضایمان نسبت به قیمت کاملاً بی‌کشش خواهد بود) ولی یک افزایش ۲۵٪ در قیمت یک خودرو نو ممکن است باعث تعویق خرید خودرو نو یا بررسی خرید یک خودرو دست دوم یا خودرو نو ارزانتری نسبت به خودرو مورد نظر شود. در این حالت تقاضای ما برای خودرو موردنظر نسبت به قیمت کاملاً حساس یا باکشش است.

محاسبه کشش تقاضا کار ساده‌ای است اگر **منحنی تقاضا** برای محصول معلوم باشد. منحنی تقاضا نمودار نمایشگر مقدار محصولی است که مصرف‌کنندگان در هر قیمت خاص خواهند خرید. کشش بین هردو نقطه در روی منحنی با محاسبه درصد اختلاف بین دو مقدار تقاضای متناظر نقاط و تقسیم آن بر درصد اختلاف بین قیمتهای متناظر به‌دست می‌آید.

برآوردنمودن کشش تقاضا

دنیای واقعی نسبت به یک منحنی به‌ندرت تا این حد مهربان است. در عوض اقتصاددانانی که مایل به برآورد کشش قیمتی تقاضا برای سیگار هستند، باید داده‌هایی درمورد تعداد سیگارهای مصرف‌شده، قیمتها و دیگر متغیرهای مربوطه (مشخص شده در زیر) بیابند و سپس با استفاده از آنها تقاضا را در قالب یک تحلیل آماری مشهور به تحلیل رگرسیونی چندمتغیری (مورد بحث در زیر) برآورد نمایند.

دو روش اصلی در برآورد تقاضا برای سیگار موجوداست. یکی به تحلیل سریهای زمانی موسوم است و از داده‌های سالانه برای متغیرهای موردنظر استفاده می‌نماید. در مطالعات تقاضا برای سیگار، از داده‌های ملی در تحلیلهای سریهای زمانی استفاده شده‌است. متغیر وابسته در چنین مطالعاتی معمولاً مصرف سرانه سیگار بزرگسالان در سال می‌باشد که به‌صورت میزان فروش کل سیگار تقسیم بر میزان جمعیت بالای ۱۷ سال تعریف می‌شود. داده‌های فروش سیگار را می‌توان از وزارت کشاورزی آمریکا اخذ نمود که بر معیارهای عینی

گوناگونی ازجمله اخذ مالیات بر مصرف سیگار فدرال استوار می‌باشند. آمار جمعیت از دفتر سرشماری قابل دسترسی است. متوسط قیمت خرده‌فروشی سیگار به‌وسیله مؤسسه‌های دولتی و غیردولتی، (هردو) محاسبه شده‌است. نشریه‌ای که به‌وسیله مؤسسه تنباکو منتشر شد (۱۹۹۸) اطلاعاتی تفصیلی درباره مالیات و قیمت ارائه کرد و توسط بسیاری از اقتصاددانانی که تقاضای سیگار را مطالعه می‌نمایند، استفاده شده است . دیگر آمارهای سالانه، ازقبیل سطوح درآمدی، از منابع دولتی قابل دسترسی هستند.

دومین روش اصلی جهت برآورد تقاضا برای سیگار به آمار مقطعی متکی است. دریک مطالعه مقطعی، تحلیل‌گر داده‌های مربوط به یک سال را از تعدادی واحدهای قابل مشاهده گوناگون جمع‌آوری می‌نماید. معمولترین واحد برای مطالعات تقاضای سیگار (همین‌طور برای تحلیل تقاضاها) ایالت است. بنابراین تحلیل‌گر از اطلاعات مربوط به مقدار مصرف سیگار و متوسط قیمتهای خرده‌فروشی (و متوسط درآمد) ایالتی استفاده می‌نماید.

چون همه ایالتها مالیات برمصرف مخصوص به خود را درمورد سیگار اعمال می‌نمایند، هر ایالت آمار قابل اعتمادی درمورد فروش شامل مالیات در اختیار دارد. درحالی که این آمار معیار دقیقی از فروشهای شامل مالیات است، اما لزوماً تصویر دقیقی از میزان مصرف واقعی سیگار توسط ساکنان هر ایالت را نشان نمی‌دهند، زیرا تفاوتهای محسوس قیمت سیگار در ایالتهای مختلف که عمدتاً ناشی از تفاوت در سطوح مالیات برمصرف است، قاچاق سیگار ـ خرید سیگار در یک ایالت (بامالیات کم) برای استفاده یا فروش مجدد در یک ایالت دیگر (با مالیات زیاد) ـ را تشویق می‌نماید. قاچاقها گاهی سازمان‌یافته است، ولی بیشتر آن غیررسمی یا غیرقانونی می‌باشد که در آن ساکنان ایالت همسایه ایالتی با مالیات کم سیگارهای مصرفی خود را از ایالت همسایه تهیه می‌نمایند. شاید بهترین گواه این مدعا تفاوت میزان فروش شامل مالیات در ماساچوست و همسایه آن نیوهمپشایر باشد. در ۱۹۸۵، میزان فروش همراه با مالیات سرانه سالانه ماساچوست (با مالیات سیگار ۲۶ سنت برای هر پاکت) ۱۱۷/۲ پاکت و برای نیوهمپشایر (۱۷ سنت برای هر پاکت) ۲۰۱/۱ پاکت بیش‌از دوسوم متوسط کل کشور (انستیتو تنباکو، ۱۹۸۸) بود. هیچ دلیلی مبنی بر اینکه ساکنان نیوهمپشایر به‌طور قابل ملاحظه‌ای بیشتر از ساکنان سایر ایالتها سیگار می‌کشند وجود ندارد. ولی با وجود این شواهدی مبنی بر افزایش

قابل توجه قاچاق سیگار در مرز ماساچوست با نیوهمپشایر وجود دارد (کمیسیون مشورتی روابط بین دولتی، ۱۹۸۵).

باتوجه به اختلالاتی که قاچاق می‌تواند در مطالعات مقطعی بوجود آورد، بیشتر تحلیلها بر آمار جمع‌آوری شده از منابعی نظیر بررسی مصاحبه‌های بهداشت ملی، که سالانه توسط مرکز ملی آمار بهداشت اجرا می‌شود، تکیه دارد. این بررسیهای بزرگ، داده‌هایی را در مورد بهداشت، عادات بهداشتی و مشخصات جمعیتی آمریکائیان ارایه می‌دهند. با درنظرگرفتن عادات مصاحبه‌شوندگان، تحلیل‌گران بر این باورند که رفتار واقعی مصرف سیگار از این طریق بهتر از طریق متکی بر آمار فروش، شامل مالیات قابل درک است. تحلیل‌گران داده‌های خوداظهاری در مورد رفتار سیگارکشیدن را برای متغیر وابسته و داده‌های قیمت مشخص ایالت را به‌عنوان یکی از متغیرهای مستقل (به شرح ذیل) به‌کار می‌برند. برای درنظر گرفتن امکان حمل‌ونقل غیرقانونی مرزی، تجزیه و تحلیلهای پیچیده اخیر همچنین قیمتهای ایالتهای همسایه را وقتی که پاسخ‌دهندگان در مجاورت ایالتهایی با مالیاتهای فوق‌العاده کم زندگی می‌کنند، دخالت می‌دهند.

بررسیهای پیمایشی، به چند دلیل منبع اطلاع خوبی نیستند. مهمترین آنها تمایل مردم به کمتر گزارش دادن مقدار سیگار مصرفی آنهاست. یافته یک مطالعه، در سال ۱۹۷۵، این بود که پاسخ‌دهندگان در یک بررسی دولتی فدرال کمتر از دوسوم کل سیگارهایی را که در آن سال در ایالات متحده فروش رفته‌بود، گزارش نمودند (*وارنر*، ۱۹۸۷). بیشتر تحلیل‌گران تقاضا برای سیگار بر این مشکل واقفند و در مطالعات خود به آن اشاره می‌نمایند.

همانگونه که در بالا اشاره شد، از چنین آمارهایی برای مطالعه تقاضای سیگار در تجزیه و تحلیل رگرسیون چندمتغیره استفاده می‌شود. یک **تحلیل رگرسیون چندمتغیره** چگونگی وابستگی تعدادی از عوامل (به نام متغیرهای مستقل) به یک عامل دیگر (متغیروابسته) را می‌آزماید. همه عوامل مقادیر گوناگونی را اختیار می‌نماید (بنابراین **متغیر** نامیده می‌شوند). فرض بر این است که متغیرهای مستقل "خودبه خود" تغییر می‌کنند (بنابراین عنوان **مستقل** دارند)، درحالی که متغیر وابسته تحت تأثیر (یعنی، "وابسته به") حرکت متغیرهای مستقل است. از همین‌رو یک تحلیل رگرسیون چندمتغیره تقاضای سیگار، مقدار سیگارکشیده شده را

به‌عنوان متغیر وابسته به‌کار می‌برد و گفته می‌شود که این مقدار به مقادیر متغیرهای مستقلی **وابسته** است که معمولاً شامل قیمت سیگار و سطوح درآمدی مردم می‌باشد. آثار مستقل دیگری که معمولاً به‌عنوان متغیرهای مستقل درنظر گرفته می‌شوند، شامل شاخصهای تبلیغات ضد دخانیات (که تصور می‌شود مردم را از سیگارکشیدن نهی می‌کند) و معیارهایی از مقدارهای پیشین مصرف سیگار (که ماهیت عادی یا اعتیادی مصرف سیگار را بیان می‌کند) می‌باشد. در هرمورد، فرض می‌شود که نوسانها متغیر مستقل بر مقدار سیگارهای مصرف‌شده یا تعداد سیگاریها ـ متغیر وابسته ـ تأثیر بگذارد.

ازنظر فنی تحلیل رگرسیون چندمتغیره، با تمام آمارهای آن، **همبستگی** بین متغیرها را مشخص می‌نماید. و بیانگر آن نیست که حرکت یک متغیر **باعث** حرکت در متغیر دیگر می‌شود بلکه می‌گوید که الگوی تغییر در یک متغیر به الگوی تغییر در دیگری مربوط است. نقش تحلیل رگرسیون چندمتغیره به‌دست آوردن همبستگیهای چندگانه در میان متغیرهای مستقل و متغیر وابسته است. تاحدی که می‌توانیم علت یا احتمال آن را درک کنیم تحلیل رگرسیون نشان می‌دهد که چگونه تغییر در هریک از متغیرهای مستقل بر مقدار متغیر وابسته تأثیر می‌گذارد. به‌عنوان مثال وقتی که به رابطه بین مقدار سیگار مصرف‌شده و قیمت آن علاقه‌مندیم، تحلیل آثار دیگر متغیرهای مستقل را کنترل می‌نماید (یا ثابت نگه می‌دارد) و برآوردی از همبستگی بین مقدار سیگار مصرف‌شده و قیمت را ارائه می‌دهد.

نتایج یک تحلیل رگرسیون چندمتغیره به تحلیل‌گر را از اینکه آیا هریک از متغیرهای مستقل همبستگی معنی‌دار آماری با متغیر وابسته دارند یا نه، و دراین‌صورت، اثر برآوردشده چه میزان است، آگاه می‌سازد. معنی‌دار بودن آماری به این مفهوم است که همبستگی شناسایی‌شده بسیار نامحتمل است که شانسی اتفاق افتاده‌باشد. از نظر فنی اگر سطح معنی‌داری انتخاب شده ۰/۰۵ باشد، بدین مفهوم است که چنانچه متغیرها رابطه‌ای نداشته باشند، همبستگی اشاره‌شده یک‌بار در ۲۰ بار به‌طور تصادفی رخ می‌دهد.

یک ضریب معنی‌دار آماری متناظر با متغیر قیمت، مقدار کاهش در مصرف سیگار را که با یک واحد افزایش در قیمت رخ می‌دهد، برآورد می‌نماید. از ترکیب این برآورد با میانگینهای قیمت و مقدار تحلیل‌گران می‌توانند کشش قیمتی تقاضا برای سیگار را برآورد نمایند.

پیش‌بینی درآمد و آثار مصرفی تغییر مالیات برمصرف

همین که براوردی از کشش و آمار قیمت و مصرف جاری دردست باشد، پیش‌بینی آثار درآمدی مالیاتی و آثار مصرفی سیگار درنتیجه تغییر در مالیات برمصرف سیگار فدرال نسبتاً آسان است. مقاصد توجیهی آثار کاهش مالیات را توضیح خواهیم داد. آثار افزایش مالیات کاملاً به‌طریق مشابه محاسبه خواهد شد، بااین تفاوت که جای کلمه کاهش با افزایش ، و بالعکس، باید در طول بحث تعویض شود.

جهت برآورد اثر مصرفی، به‌خاطر می‌آوریم که کشش برابر درصد افزایش مصرف به درصد کاهش قیمت، هردو برحسب درصد، می‌باشد. همه اطلاعات اساسی به جز سطح جدید مصرف ناشی از کاهش مالیات را می‌دانیم. کشش از تحلیل رگرسیون به‌دست آمده‌بود. درصد کاهش قیمت به سادگی برابر با نسبت کاهش مالیات به متوسط قیمت است. چون سطح اولیه مصرف را داریم، به سادگی معادله کشش را برای تنها مجهول آن، سطح جدید (پس از مالیات) مصرف، حل می‌نماییم.

برآوردکردن اثر درآمدی مالیات نیز کاملاً روشن است. تغییر در درآمد مالیاتی به همراه کاهش مالیات فدرال به‌وسیله کم‌کردن برآورد درآمد در سال پس از کاهش درآمد در سال قبل از کاهش مالیات محاسبه می‌شود. در هرسال، درآمد کل مالیات سیگار برابر با حاصل ضرب مالیات هر پاکت در تعداد پاکتهای فروخته‌شده می‌باشد. در سال قبل از کاهش مالیات، درآمد برابر حاصل ضرب نرخ مالیاتی اولیه در تعداد اولیه است. در سال پس ازکاهش، درآمد به‌صورت حاصل ضرب نرخ مالیاتی جدید در تعداد برآوردشده به طریق فوق، برآورد خواهدشد.

علاقه به اثر درآمدی یک تغییر مالیاتی درکل مطرح می‌باشد. ولی در مورد آثار مصرفی، علاقه ممکن است متوجه زیرگروههایی از جامعه شود. یعنی قانونگذاران (یا متخصصان بهداشت و سایرین) ممکن است علاقه‌مند به دانستن اثر مالیات بر سیگارکشیدن نوجوانان باشند، زیرا نوجوانان نسل جدیدی از سیگاریها و نیز سیگاریهای احتمالی را تشکیل می‌دهند. ازسوی دیگر، آثار مصرفی یک تغییر مالیاتی روی سیگاریهای مسن‌تر ممکن است از توجه خاصی برخوردار باشد، زیرا این گروه، افرادی هستند که امراض مربوط به سیگارکشیدن زودتر

در آنها نشان داده خواهدشد و بنابراین با ترک سیگار ممکن است به‌سرعت سلامت خود را بازیابند. با استفاده از داده‌های مصرف از بررسیهای آماری اقتصاددانان می‌توانند توابع تقاضا برای زیرگروههای خاصی از جامعه را، با سن و جنسیت عمده‌ترین متغیرها، برآورد نمایند. همانگونه که در بخش بعدی خواهیم دید، این روش به اقتصاددانان این اجازه را داده‌است که برای گروههای سنی مختلف کشش‌های قیمتی را جدا نمایند و از آنجا آثار مصرفی جداگانه‌ای را برآورد کنند.

آثار درآمدی ومصرفی اقدام مالیاتی سیگار فدرال در ۱۹۸۶

عمل مسئولیت بودجه‌ای و برابرکردن مالیات (TEFRA)[۱] در سال ۱۹۸۲ شامل دوبرابر نمودن مالیات بر مصرف سیگار فدرال از ۸ سنت به ۱۶ سنت در پاکت بود که از اول ژانویه ۱۹۸۲ به مرحله اجرا درآمد. به‌عنوان یک عمل موقت جهت افزایش درآمد، قرار بود که این مالیات فقط تا اول اکتبر ۱۹۸۵ اخذ شود و پس از آن به سطح قبلی خود بازگردد.

به علت ازدست دادن درآمد اضافی ونیز به علت ترس از اینکه کاهش مالیات باعث تشویق به سیگارکشیدن خواهدشد، نمایندگان کنگره در اوایل ۱۹۸۶، قانونی گذرانیدند که تغییر مالیات را دائمی نمود (بین اول اکتبر تا زمان نهایی شدن قانون، ۱۶ سنت مالیات دو بار به‌طور موقت اجرا شد).

برداشت قانونگزاران از پیامدهای تغییر مالیات عمدتاً نتیجه تجزیه وتحلیل اقتصادی متکی به مطالعات کشش تقاضای سیگار جهت برآورد آثار درآمدی و مصرفی، مشروط بر اجازه اجرای مالیات جدید بود. تحلیل باداده‌هایی درمورد متوسط قیمت خرده‌فروشی یک پاکت سیگار در سال ۱۹۸۴ (۹۷/۸ سنت) در سطح کشور، معیارهای گوناگونی در مورد میزان مصرف (شامل تعداد سیگاریهای هریک ازگروههای سنی و مصرف سالانه آنها)، و برآوردهای کشش گروههای سنی گوناگون، به تفکیک کششهای **مشارکت و کل** شروع شد. کششهای مشارکت به اندازه‌گیری اینکه چگونه تصمیم به مصرف یا عدم مصرف سیگار تحت تأثیر قیمت

۱ـ *Tax Equalization and Fiscal Responsibility Act*

بود و کششهای کل به ارزیابی نگرانی سنتی مبنی بر اینکه چگونه مصرف کل سیگار (حاصل ضرب تعداد سیگاریها و متوسط مصرف سالانه آنها) به تغییرات قیمت واکنش نشان می‌داد، می‌پردازند.

جدول ۱ برآوردهای تقاضای سیگار و کششهای قیمتی تهیه شده توسط دکتر *اوگن لویت* و همکارانش را در دفتر ملی بررسی اقتصادی نشان می‌دهد. کشش کل بزرگسالان (سن ۷۴-۲۰)، ۰/۴۲ بیانگر این است که ۱۰٪ افزایش قیمت، مصرف بزرگسالان را ۴۲٪ کاهش خواهدداد. همانگونه که در جدول مشاهده می‌شود، و کاملاً منطقی به‌نظر می‌رسد، سیگاریهای مسن و خوکرده‌تر (و احتمالاً مرفه‌تر) کمتر از بزرگسالان جوان نسبت به قیمت حساس هستند. بدین معنی که کششها با افزایش سن کاهش می‌یابند. سازگار با این قالب و یا انتظارات، کشش کل برای نوجوانان، ۱/۴۰ از کشش جوانترین بزرگسالان نیز بیشتر است. برای نوجوانان، ۱۰٪ افزایش قیمت مصرف را بیش از ۱۴٪ کاهش می‌دهد.

از نظر بهداشتی، کششهای مشارکت بیشتر مورد توجه هستند، چون یک "کف" مبنی بر اینکه چند نفر تحت‌تأثیر افزایش قیمت سیگار مصرف نمی‌کنند، ارائه می‌دهد. برآورد مشارکت ۰/۲۶ برای همه بزرگسالان دال بر این است که یک افزایش ۱۰٪ در قیمت ۲/۶٪ همه بزرگسالانی را که در عدم افزایش قیمت، سیگار می‌کشیدند، نهی می‌نماید. کشش مشارکتی نیز با افزایش سن کاهش می‌یابد.

برای نوجوانان، تصمیم اصلی به سیگارکشیدن قسمت عمده کشش کل را به خود اختصاص می‌دهد. کشش مشارکت ۱/۲۰ برای نوجوانان به این مفهوم است که یک افزایش ۱۰٪ در قیمت، ۱۲٪ نوجوانانی را که در غیراین صورت سیگاری بودند از سیگارکشیدن نهی می‌نماید. توجه داشته باشید که دوگروه این ۱۲٪ را تشکیل می‌دهند: نوجوانانی که به علت افزایش قیمت وادار به ترک می‌شوند و نوجوانانی که به علت این افزایش قیمت از شروع به سیگارکشیدن خودداری می‌کنند. با چنین کشش مشارکت بالایی برای نوجوانان، افزایش قیمت ابزار مهمی برای کودکان از شروع به مصرف سیگار و یا ادامه آن است.

قدم بعدی برای تبدیل این کششهای جدول ۱ به برآوردهایی از آثار تغییر مشخص مالیات فدرال، گام ساده‌ای بود: تحلیل‌گران کاهش مالیاتی طرح‌شده اولیه را به تغییر قیمت تبدیل

جدول ۱ـ کششهای قیمت و تقاضای سیگار

گروه سنی (سال)	کششها	
	کل	مشارکت
۱۷ - ۱۲	۱/۴۰	۱/۲۰
۲۵ - ۲۰	۰/۸۹	۰/۷۴
۳۵ - ۲۶	۰/۴۷	۰/۴۴
۷۴ - ۳۶	۰/۴۵	۰/۱۵
همه بزرگسالان (۷۴ - ۲۰).....	۰/۴۲	۱/۲۶

● *کششها برای همه بزرگسالان از رگرسیونهای جداگانه محاسبه شده‌اند. این مطلب علت کمتربودن کشش کل از کشش زیرگروهها را توضیح می‌دهد.*

یاداشت نویسنده: این جداول براساس دو مطالعه مختلف می‌باشد که در منبع ذکر شده‌است. در یک مطالعه گروه سنی ۱۷-۱۲ و در دیگری گروه‌سنی ۷۴- ۲۰ مورد بررسی قرارگرفت. جدول شماره ۳ براساس برون‌یابی این دواست.

نمودند. از یک قیمت پایه ۹۷/۸ سنت، ۸ سنت کاهش مالیاتی، قیمت جدید ۸۹/۸ سنت را چنانچه سایر چیزها ثابت بمانند (به‌عنوان مثال به‌فرض اینکه تولیدکنندگان قیمت عمده‌فروشی را تغییر ندهند) ایجاد می‌کند. درصد تغییر قیمت برابر ۰/۰۸۵۳=[(۹۷/۸+۸۹/۹)×۱/۲]/۸ یا ۸/۵۳٪ خواهدبود. (مخرج کسر متوسط قیمتهای قدیم و جدید است). اگر این رقم مخرج در یک معادله کشش باشد، کششهای مختلف از جدول ۱ می‌توانند در این معادله قرار گیرند و معادله برای صورت کسر، درصد تغییر در مقدار تقاضا، حل شود.

نتایج این محاسبات در جدول ۲ عرضه شده‌است. این جدول نشان می‌دهد که ۸ سنت کاهش مالیات مصرف سیگار همه بزرگسالان را ۳/۶٪ و تعداد سیگاریهای بزرگسال را ۲/۲٪ افزایش می‌دهد. درحالی که مصرف سیگار نوجوانان را تقریباً ۱۲٪ (۱۱/۹٪) افزایش و تعداد آنها را ۱۰/۲٪ افزایش خواهدداد. برعکس، لغو حکم کاهش مالیات مانع از این افزایشها در مصرف سیگار خواهد شد.

برای تبدیل ارقام درصد تغییر در جدول ۲ به تعداد سیگاریها و سیگار، کافی است که این ارقام را در تعداد سیگاریها در هرگروه سنی و مصرف سیگار سالانه آنها ضرب نماییم. داده‌های

جدول ۲ـ درصد افزایش مورد انتظار مصرف سیگار درنتیجه ۸ سنت کاهش در مالیات برمصرف سیگار فدرال

گروه سنی (سال)	کل	مشارکت
۱۷ - ۱۲	۱۱/۹	۱۰/۲
۲۵ - ۲۰	۷/۶	۶/۳
۳۵ - ۲۶	۴/۰	۳/۷
۷۴ - ۳۶	۳/۸	۱/۳
همه بزرگسالان (۷۴ - ۲۰)	۳/۶	۲/۲

مصرف کل از بررسیهای پیمایشی دولت فدرال به‌دست آمده‌اند.

نتایج این عملیات ساده در جدول ۳ عرضه شده‌است [اضافه کردن دو گروه سنی (+ ۷۵ و ۱۹-۱۸) در جدول ۳ و داده‌های جمع‌آوری شده موردنظر در ۱۹۸۶، *وارنر* توضیح داده شده است]. این جدول پاسخ یکی از دو سئوال مورد علاقه کنگره را می‌دهد: اگر کاهش مالیاتی تصویب‌شده به اجرا گذاشته شود، چند نفر به سیگارکشیدن تشویق خواهند شد یا اگر کاهش لغو شود، چند نفر سیگار نخواهند کشید؟

اعداد شگفت‌آور هستند. درمجموع، چنانچه کاهش مالیات اعمال شود، تقریباً ۲ میلیون آمریکایی دیگر به مصرف سیگار روی خواهند آورد. از این تعداد، تقریباً نیم میلیون (۴۶۴۰۰۰) نوجوان (گروههای سنی ۱۷-۱۲ و ۱۹-۱۸) خواهند بود. بیش از نیم میلیون (۶۰۹۰۰۰) در جوانترین گروه سنی بزرگسالان، ۲۵-۲۰ خواهند بود. در گروه شامل میانسالان وبزرگسالان مسن، بزرگترین گروه سیگاریها اعداد، کوچکتر (۳۵۱۰۰۰) ولی قابل‌ملاحظه است. تحلیل نشان می‌دهد که مصرف سالانه سیگار ۲۰/۶ میلیارد نخ سیگار افزایش خواهدیافت.

همه این اعداد (از جمله صدهاهزار نفر و دهها میلیارد نخ سیگار) کاملاً بزرگ به‌نظر می‌رسند. این اعداد به‌طور مطلق نیز بزرگ هستند. آنها سیگارکشیدن اضافی عمده‌ای را همراه با افزایش قابل‌ملاحظه‌ای در امراض ناشی از سیگارکشیدن منعکس می‌نمایند. ولی به‌طور نسبی، این آثار معتدل هستند. با درک اینکه حدود ۵۶ میلیون آمریکایی سیگاری هستند، اثر

جدول ۳ـ برآورد افزایش مصرف سیگار دراثر ۸ سنت کاهش در مالیات برمصرف سیگار فدرال

گروه سنی (سال)	تعداد سیگاریها(هزار)	تعداد سیگار (میلیارد نخ)
۱۲ - ۱۷	۳۳۴	۲/۳
۱۸ - ۱۹	۱۳۰	۱/۰
۲۰ - ۲۵	۶۰۹	۴/۸
۲۶ - ۳۵	۵۰۹	۴/۱
۳۶ - ۷۴	۳۵۱	۸/۲
۷۵+	۱۳	۰/۲
کل	۱۹۴۵	۲۰/۶

مشارکت ۸ سنت تغییر مالیات تنها ۳/۵٪ جمعیت سیگاری را شامل می‌شود. به‌طور مشابه اثر پیش‌بینی شده بر مصرف کل سیگار تنها فقط ۵٪ سیگاری است که مردم مصرف آن را گزارش نموده‌اند. بنابراین اثر مصرفی ۸ سنت تغییر مالیات را در مقایسه با کل عادت مصرف سیگار باید معتدل تلقی کرد. ولی به‌طور مطلق قابل‌ملاحظه است؛ به‌خاطر اینکه سیگاریهای زیادی هستند که هرسال تعداد زیادی نخ سیگار مصرف می‌نمایند. حتی ۱٪ تغییر مختصر در مصرف سیگار می‌تواند آثار بهداشتی عمومی مهمی درپی داشته باشد.

برآورد اولیه‌ای از این اثر بهداشتی عمومی می‌تواند از این حقیقت ناشی شود که یک چهارم تا یک سوم افرادی که در طول عمرشان سیگار می‌کشند به علت عادت به سیگارکشیدنشان می‌میرند. به همین دلیل اگر ۸ سنت کاهش مالیات به اجرا گذاشته شود (و ارزش واقعی آن در طول زمان حفظ گردد)، حداقل نیم میلیون آمریکایی دیگر به‌طور نابهنگام ازبین خواهند رفت، گرچه بیشتر این مرگها چندین دهه دیگر اتفاق خواهد افتاد. هرکدام یک کاهش امید به زندگی از ۱۵ تا ۲۰ سال را دربر خواهدداشت. علاوه بر این تأسف‌بارترین نتیجه یک کاهش مالیات، افزایش عمده‌ای نیز در بیماری و ناتوانی ناشی از سیگارکشیدن درپی خواهدداشت.

باتوجه به تحلیل آثار مصرفی یک کاهش مالیاتی، برآورد پیامدهای درآمدی کاهش مالیاتی کار ساده‌ای است. قبل از تغییر، مالیات هربسته ۱۶ سنت، و مصرف کل تقریباً ۲۸/۵ میلیارد پاکت سیگار بود. پس درآمد کل مالیات بر مصرف سیگار فدرال، حاصل ضرب این دورقلم، یعنی ۴/۵۶ میلیارد دلار بود. اگر مالیات ۸ سنت کاهش داشت، نرخ جدید ۸ سنت می‌شد و

مصرف کل افزایش می‌یافت. طبق جدول ۳، مصرف ۲۰/۶ میلیارد سیگار یا کمی بیش از یک میلیارد پاکت افزایش می‌یافت. ولی این برآورد از بررسیهایی به‌دست آمد که نشان داده شده مردم مصرف روزانه خود را یک سوم کمتر گزارش می‌نمایند. همانگونه که قبلاً بحث شد، راه بهتر برآورد سطح جدید مصرف سیگار، استفاده از فرمول کشش با برآوردی از کشش قیمتی کلی تقاضا برای سیگار در سطح ملی (تهیه‌شده توسط *لویت* وهمکارانش) یعنی ۰/۴۷ است. با انجام این کار سطح جدید مصرف را ۲۹/۶۶ میلیارد پاکت سیگار، (افزایشی معادل ۱/۱۶ میلیارد پاکت) برآورد می‌کنیم. کل درآمد مالیات برمصرف با نصف کردن نرخ مالیات ۲/۳۷ میلیارد دلار خواهدبود. زیان مالیاتی ناشی از نصف کردن مالیات بر مصرف برابر ۴/۵۶ میلیارد دلار یعنی منهای ۲/۳۷ میلیارد دلار یا ۲/۱۹ میلیارد دلار خواهدبود.

همانند اثر مصرفی، اثر درآمدی به‌طور نسبی (یعنی نسبت به یک تریلیون دلار بودجه فدرال) معتدل، ولی به‌طور مطلق قابل‌ملاحظه خواهدبود. همانگونه که زمانی *سناتور اورت* دریکسن پیشنهاد نمود، اگر یک میلیارد دلار از اینجا و یک میلیارد دلار ازجای دیگر برداریم، بزودی پول قابل‌ملاحظه‌ای جمع خواهدشد.

برآورد آماری کشش، چیزی بیش از برآورد نیست. درنتیجه اعتبار برآوردهایی که از مطالعات کشش استخراج شده‌اند، به اعتبار محاسبات کشش بستگی دارد. درحالی که تضمین اعتبار با قاطعیت امکان‌پذیر نیست، اطمینان‌بخش است اگر توجه شود که در ۱۹۸۳، پس از افزایش ۸ سنت در مالیات فدرال و افزایش قیمتهای تولیدکنندگان ـ یک ترقی قیمت کل ۱۳/۴۵٪ (پس ازتورم) ـ مصرف سیگار در کل ۵/۱۳٪ کاهش یافت. این نکته بیانگر کشش کل ۰/۳۸ است که به برآورد *لویت* و همکارانش نزدیک می‌باشد. هرگاه توجه کنیم که جامعه سنی سیگاریها درفاصله ۱۹۸۳-۱۹۸۲ رشد کرده‌است، حتی برآوردهای آنها بهتر به‌نظر می‌رسد. به‌علاوه اثر درازمدت پیش‌بینی شده در مطالعات کشش ممکن است مستلزم بیش ازیک سال وقفه برای اثرگذاری باشد.

اثر تجزیه و تحلیل بر تصمیم‌گیری کنگره

هیچ‌کس نتیجه نمی‌گیرد که تجزیه و تحلیل بالا نقش مهمی در تصمیم‌گیری کنگره برای لغو

قانون کاهش مالیات سیگار داشته است. مسائل سیاسی ـ بده‌بستانهای سنتی که کنگره به‌طور توجیه‌پذیر به‌خاطر آن مشهور است ـ بدون شک نقش برجسته‌تری ایفا نمود، همانگونه که قانونگذاران در مسائل درآمدی به‌طور اعم و مالیاتهای بر مصرف (و مالیاتهای سیگار) به‌طور اخص ایفا کردند.

همچنین شواهد تاریخی دلالت بر این دارد که تحلیل تفکر برخی از کارگزاران کلیدی بحثهای کنگره را متأثر نمود. گزارش شده که یک سناتور برنامه‌ریزی کرده بود که بحث لغو کاهش مالیات را به تعویق اندازد. او احساس می‌کرد که کنگره در ۱۹۸۲ جهت برگرداندن مالیات به سطح قبل از TEFRA تعهد نموده‌است و باید به تعهداتش وفا نماید. بنا به اظهار یکی از مددیاران سناتور، او به‌قدری تحت‌تأثیر این موضوع که کاهش مالیات باعث خواهدشد که تقریباً نیم میلیون کودک سیگار بکشند، قرارگرفت که نقشه تعویق انداختن را رها نمود. طرفداران بهداشت عمومی که بحث را از نزدیک دنبال می‌کردند، گزارش دادند که چند عضو دیگر کنگره نیز تحت‌تأثیر این موضوع، که در یک مجله معروف پزشکی چاپ شده، در گزارشهای عرضه‌شده، و به‌وسیله رسانه‌ها پوشش داده شده‌بود، قرارگرفتند. بنابراین ، درحالی که غیرممکن است سهم دقیق تحلیل اقتصادی بر دائم نمودن ۱۶ سنت مالیات را دقیقاً ارزیابی نمود، منطقی است نتیجه‌گیری کنیم که تحلیل اقتصادی و آماری مربوطه، به فرایند تصمیم‌گیری درکنگره کمک نموده و عینیت مبتنی بر آمار را در فعالیتی که در غیراین صورت بسیار سیاسی بود، تزریق کرد.

مسائل

۱ـ توضیح دهید که "کشش قیمتی تقاضا" چه چیزی را اندازه می‌گیرد؟ چگونه کششها می‌توانند در تصمیم‌گیری کمک نمایند؟

۲ـ "تحلیل رگرسیون چندمتغیره" چیست؟ توضیح دهید که چگونه از آن در برآورد کششهای قیمتی استفاده می‌شود.

۳ـ کشش قیمتی تقاضای کل برای سیگار بی‌کشش است. آیا تولیدکنندگان سیگار که در پی حداکثرنمودن سود هستند، باید قیمت محصولات خودرا افزایش یاکاهش دهند؟ توضیح دهید.

۴ـ فرض کنید که کشش قیمتی تقاضای کل برای سیگار بجای ۰/۴۷، برآورد واقعی آن، ۱/۵۰ است:

الف) آثار درآمد مالیاتی و مصرف کل را درصورتی که مالیات برمصرف از ۱۶ سنت به ۸ سنت در هر پاکت کاهش یابد، برآورد نمایید.

ب) باتوجه به واقعیت بی‌کشش بودن تقاضا، ملاحظات درآمدی و بهداشت عمومی هردو به حفظ مالیات ۱۶ سنتی توصیه می‌کنند. باتوجه به محاسبات در (الف) آیا سیاستهای بهداشتی و بودجه‌ای، اگر کشش ۱/۵۰ بود، همخوانی داشتند؟ پاسخ خود را تشریح نمایید.

۵ـ فرض ضمنی در برآورد آماری کششها این است که واکنشها نسبت به افزایشها و کاهشهای قیمت متقارن هستند. این فرض ـ که دراین مقاله از آن به‌خاطر سادگی در توضیح ـ استفاده شده است این نتیجه را می‌دهد که یک افزایش ۱۰٪ در قیمت سیگار، تقاضای سیگار را به یک کاهش ۴/۷٪ خواهد رساند، همانگونه که یک کاهش ۱۰٪ در قیمت، افزایش ۴/۷٪ در تقاضا را ایجاب می‌کند. بسیاری از اقتصاددانان بر این باورند که افزایش قیمت سیگار نسبتاً اثر بیشتری بر تقاضا خواهدداشت تا کاهش قیمت.

الف) توضیح دهید که چرا این عدم تقارن درمورد مصرف سیگار منطقی به‌نظر می‌رسد؟

ب) آیا انتظار دارید این عدم تقارن برای نوجوانان و میانسالان قابل مقایسه باشد؟ پاسختان به کششهای مختلف در جدول ۱ چگونه مربوط می‌شود؟

۶ـ برآوردهای کشش از تحلیلهای رگرسیونی به‌دست می‌آیند. تحلیلهای گوناگون ـ با استفاده از خصوصیات تابعی داده‌های مختلف می‌تواند برآوردهای مختلفی از کششها را به‌دست دهد و چنین نیز می‌شود. آیا یک مخالف لغو ۱۹۸۶ ترجیح می‌دهد ازیک برآورد بالا یا پایین کشش جهت بحث علیه حفظ مالیات در ۱۶ سنت استفاده کند؟

منابع

Advisory Commision on Intergovernmental Relations. 1985, *Cigarette Tax Evasion: A Second Look.*

E. M. Lewit and D. Coate. 1982, "The Potential for Using Excise Taxes to Reduce Smoking."

Journal of Health Economics 1:121-145.

E. M. Lewit and D. Coate 1981, "The Effects of Government Regulation on Teenage Smoking." *Journal of Law and Economics* 24:545-569.

K. E. Warner 1978, "Possible Increases in the Underreporting of Cigarette Consumption." *Journal of the American Statistical Association* 73:314-318.

K. E. Warner. 1986, "Smoking and Health Implications of a Change in the Federal Cigarette Excise Tax." *Journal of the American Medical Association* 255:1028-1032.



آیا وراثت در ایجاد بیماری موثراست؟
استفاده از مطالعات روی دوقلوها در تحقیقات پزشکی

دی. دی. رید[۱] دانشکده بهداشت و طب گرمسیری

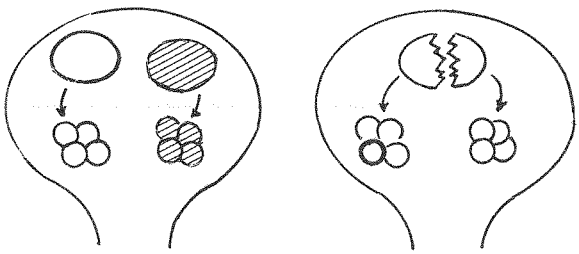
بحث در مورد اهمیت نسبی سرشت و تربیت که در زمینه‌هایی مانند روانشناسی جریان دارد، در زمینه پزشکی نیز برقرار است. سئوال اساسی این است که آیا ابتلا به جنون یا بیماریهای قلبی در ما ناشی از این است که استعداد خاصی نسبت به یک بیماری از والدین خود به ارث می‌بریم، یا این بیماریها نتیجه یک محیط پرتنش در سرتاسر زندگی ماست؟ در مورد اغلب

۱ـ *D. D. Ried*

بیماریها نمی‌توان پاسخ قطعی داد، زیرا بیماران تقریباً همیشه در طول سالهای رشد دوران کودکی هم در سرشت، یا در انتقال ژن، و هم درتربیت، یامحیط باهم تفاوت دارند، والبته هردوی اینها می‌توانند در بیماریهای خاصی مهم باشند. با وجود این مهم است بدانیم که آیا مثلاً در بیماریهای قلبی عروقی، ترغیب کردن بیمار برای تغییر محیط، با ترک مصرف سیگار ارزش دارد یانه. اگر هردوی دوقلوها زیاد سیگار بکشند و بیماری قلبی داشته باشند، چون خویی نگران و مضطرب رابه ارث برده‌اند، در این صورت ترک سیگار به طور ناامیدکننده‌ای اثر کمی دارد. بنابراین دلایل، تحقیقات پزشکی در این زمینه بر فرصتهای غیرمعمول و منحصربه فردکه با بروز بیماری در دوقلوها حاصل می‌شود، متمرکز شده‌است.

اندیشه این رهیافت در سال ۱۸۷۵ در *فرانسیس گالتون* زیست‌سنج انگلیسی به‌وجود آمد. وی متذکر شدکه دو نوع دوقلو وجود دارند، همانند و ناهمانند. دوقلوهای همانند هر دو از یک تخمک لقاح یافته تولید می‌شوند و در نتیجه دارای مجموعه ژنهای همسان هستند، که خصیصه‌های بدنی و ذهنی را تعیین می‌کنند. دوقلوهای ناهمانند نظیر نوزادان معمولی متولدشده در زمانهای مختلف، از تخمکهای جداگانه تولید می‌شوند و الگوهای ژنی آنها بیشتر (یاکمتر) ازتشابه برادرها و خواهرهای معمولی نیست (شکل ۱ را ببینید). اما هر دو نوع از دوقلوها، معمولاً در طول دوره کودکی در محیط خانوادگی مشترک هستند، و معمولاً تحت رفتار یکسان از سوی والدین خود قرار می‌گیرند. در این صورت اگر یکی از دوقلوها دارای بیماریی باشد که از طریق ژنها منتقل می‌شود، خیلی محتملتر است که در عضو دوم دوقلویی ظاهر شودکه با عضو اولی همانند است تا در عضو دوم دوقلویی که از نظر ژنی متفاوت است. از طرف دیگر در بیماریهای ناشی از رژیم غذایی یا سایر موارد زندگی خانوادگی که به معنای اکید کلمه ژنی نیست، خواهر و برادرهای دوقلوی همانند و ناهمانند، ابتلا به بیماری برای کودکان در معرض این وضع به یک اندازه محتمل است. بنابراین می‌توان تاحدی نشانه‌ای دال بر اهمیت نسبی وراثت و محیط را در ایجاد بیماری خاصی با مقایسه خطر مبتلا شدن نسبی یکی از دوقلوها وقتی که وی یکی از دوقلوهای همانند باشد با همان خطر ابتلا وقتی که یکی از عضوهای دوقلوهای ناهمانند باشد، دریابیم. تفاوت در معرض خطر بودن در بیماریهایی که توارث در آنها مهم است، بالا؛ و زمانی که توارث اهمیت ندارد، پایین است.

شکل ۱: ژنهای تعیین‌کننده موی روشن با علامت ○ و موی تیره با علامت ● مشخص می‌شوند. دوقلوهای ناهمانند از تخمکهای جداگانه‌ای تولید می‌شوند که حامل عناصر ژنی مختلفی هستند که برای مثال، رنگ مو را در دو بچه تعیین می‌کنند. در حالت دوقلوهای ناهماتند رنگ ممکن است متفاوت باشد یانباشد. دوقلوهای ـ ـانند ازتقسیم یک تخمک حاصل می‌شوند و بنابراین عنصر ژنی یکسانی را حفظ می‌کنند. بنابراین هردو بچه دارای یک رنگ مو هستند.



معیارهای هماهنگی بیماری

مخاطره‌های ظهور همزمان بیماری که در هر دونوع دوقلوها پیش می‌آید، به دو روش اندازه‌گیری می‌شود. هدف هر دو روش، ارزیابی درجه هماهنگی یا توافق بین ابتلا به بیماری برای دوقلوهای همانند و ناهمانند است. جدول ۱ مدل جمعیتی متشکل از شش مجموعه دوقلوهای همانند را نشان می‌دهد. یک معیار اندازه‌گیری مبتلا شدن هر دو عبارت از "نرخ هماهنگی زوجی" است. در این مثال، در چهار زوج از شش زوج، هردو دوقلو مبتلا شده‌اند. بنابراین نرخ هماهنگی زوجی $\frac{۴}{۶}\times ۱۰۰$ یا $۶۶\frac{۲}{۳}$ درصداست. وقتی که هیچیک از دوقلوها بیمار نباشند، زوج به جدول وارد نمی‌شود. معیار دوم بستگی به فراوانی نسبی یا مخاطره عضو دیگر دوقلوی فردی مبتلاست که او نیز دچار بیماری است. این "نرخ هماهنگی فرد منتخب" یا "نرخ هماهنگی فرد مبتلا" در مثال با نسبت تعداد دوقلوها که به صورت زوجی مبتلا شده‌اند (۸) به

جدول ۱ـ مثال هماهنگی

عضو دوم دوقلو	عضو اول دوقلو	شماره زوج
○	○	۱
○	○	۲
○	○	۳
○	○	۴
⊙	○	۵
⊙	○	۶

(○ سالم ⊙ مبتلا)

جدول ۲ـ وقوع بیماریهای جسمی انتخاب شده از دفتر ثبت دوقلوهای دانمارکی مبتنی بر بررسی ۴۳۶۸ زوج همجنس

بیماری	نرخ توافق مؤثر (درصد)			
	نرخ همانندها	تعداد *	نرخ ناهمانندها	تعداد *
سکته مغزی	۳۶	۱۲۰	۱۹	۱۶۴
انسداد سرخرگ	۳۳	۱۲۲	۲۷	۱۷۹
سل	۵۴	۱۸۵	۲۷	۳۰۹
تب روماتیسمی	۳۳	۱۷۸	۱۰	۲۳۸
روماتیسم مفصلی	۵۰	۶۳	۵	۷۳
مرگ ناشی ازعفونت حاد	۱۴	۱۳۷	۱۱	۲۳۵
آسم	۶۳	۹۴	۳۸	۱۲۵
صرع	۵۴	۳۷	۲۴	۴۹

** این تعداد نشان‌دهنده افراد مبتلا است، نه زوجهای مبتلا*

کل تعداد مبتلایان یعنی $\frac{۸}{۱۰}$ یا ۸۰ درصد مشخص می‌شود. شاید نرخ هماهنگی فرد مبتلا درمطالعه روی دوقلوها گسترده‌ترین استفاده را دارد. (بادانستن یکی از این معیارها می‌توان معیاردیگر را به آسانی به‌دست آورد، لذا این انتخاب اهمیت چندانی ندارد و اغلب سلیقه‌ای‌است.)

نتایجی از دفتر ثبت دوقلوهای دانمارکی

در جدول ۲ نرخهای هماهنگی فرد مبتلا برای زوجهای همانند و ناهمانند دوقلوهای دانمارکی‌که حداقل یکی از آنها مبتلا به یکی از بیماریهای فهرست زیراست، مقایسه شده‌اند. نرخهای‌بالا در دوقلوهای همانند در مقابل نرخهای پایین برای دوقلوهای ناهمانند برای بیماری سل (۵۴ درمقابل ۲۷)، روماتیسم مفاصل (۵۰ درمقابل ۵)، آسم (۶۳ در مقابل ۳۸)، و صرع (۵۴ درمقابل ۲۴) عنصر ژنی قوی را در این امراض نشان می‌دهد. ازطرف دیگر، مرگ ناشی از عفونتهای حاد غیر از سل و تب روماتیسمی (۱۴ درمقابل ۱۱) چنین عدم هماهنگی بین زوجهای دوقلوهای همانند و ناهمانند را نشان نمی‌دهد. بنابراین قرارگرفتن در معرض عفونت، شاید در خارج منزل، خیلی مهمتر از هرگونه استعداد توارثی به نظر می‌رسد.

غیرعادی بودن دوقلوها

این مثالها، توانایی این روش را برای تمایز بین عوامل ژنی و محیطی در بیماریهای مختلف نشان می‌دهد. اما لازم است درباره خطرهای ناشی از تعمیم از روی دوقلوها به کل جامعه هشدارداده‌شود، زیرا دوقلوها از خیلی جهات غیرعادیتر از افراد غیردوقلو هستند. دوقلوها از این نظر غیرعادیند که تولدهای چندگانه نسبتاً بندرت رخ می‌دهند. در ۱۹۶۴ تنها یک زایمان از ۹۶زایمان دوقلو بود. (سه‌قلو و چهارقلو در هر ۹۹۷۷ و ۶۶۳۴۷۰ زایمان به‌ترتیب یک بار اتفاق افتاد. در مقایسه با نرخ یک بار در ۹۲۱۶ زایمان و یک بار در ۸۸۴۷۳۰ زایمان، بنابر فرض *هلن ـ زلنی*؛ یعنی اگر دوقلو یک بار در ۹۶ زایمان رخ دهد، سه‌قلو و چهارقلو به‌ترتیب یک بار در ۹۶×۹۶ و ۹۶×۹۶×۹۶ زایمان رخ خواهد داد. در اینجا فعلاً به این فرض نمی‌پردازیم.)

مهمترین مسئله از دیدگاه پزشکی، این حقیقت است که دوقلوها در خانوارهای سیاه

آمریکایی بیشتر از خانوارهای سفید آمریکایی و مخصوصاً از مادران مسنتر که قبلاً چندین فرزند داشته‌اند، متولد می‌شوند. در وضعیتهایی نظیر مونگولیسم که بویژه دربین بچه‌هایی که توسط زنهای مسنی که فرزندان زیادی داشته‌اند، متولد شده‌اند، شانس ابتلا برای دوقلوها بیشتر است تا سایرین. زیرا دوقلوها مجبورند با مواد غذایی که تنها برای یک بچه در رحم آماده می‌شود، بسازند و بنابراین از همان ابتدا در وضعیت نامطلوب قرار می‌گیرند و تولد نارس و زایمان سخت برای آنها محتملتر است. در نتیجه تعجب‌آور نیست که نرخ مرگ در آنها هنگام طفولیت و اوان کودکی بالاتر از متوسط باشد. آنها هرچه بزرگتر می‌شوند وضع نامطلوبشان رو به بهبود می‌گذارد و در ارتباط با بیماریهای دوران بلوغ امکان ابتلای آنها به بیماری احتمالاً به بچه‌های غیردوقلو نزدیک می‌شود. بنابراین ارزیابیهایی برمبنای مشاهده بیماری در دوقلوهای میانسال احتمالاً قابل اعمال برای کل جمعیت است.

جنبه‌های عملی مطالعه روی دوقلوها

نمونه‌گیری از جامعه دوقلوها. مطالعه‌های اولیه بیماریها در دوقلوها اغلب مبتنی بر بیمارانی با بیماریهای غیرعادی بوده که به بیمارستان مراجعه می‌کردند و درمی‌یافتند که توأم او نیز از همان بیماری رنج می‌برد. از آنجا که در طب بالینی به‌طور کلی تمایل به‌این است که موارد آشکارا غیرعادی مورد توجه قرارگیرند و نتایج مطالعه منتشر شود، لذا رخ دادن همزمان اینگونه موارد در دوقلوها، در نوشته‌های پزشکی اهمیت خاصی یافته‌است. اما به‌دلیل اتفاقی بودن روش جمع‌آوری نامحتمل است که این سری موارد تصویر درستی از وقوع بیماری در جامعه دوقلوها درکل به‌دست دهد. سری افراد داوطلب نیز که شاید از طریق فراخوان در رسانه‌های عمومی به مطالعه گرفته می‌شوند محتملاً اریب‌اند. مطلوب این است که یااطلاعات در مورد همه دوقلوها که دریک نسل متولد شده‌اند یا حداقل روی نمونه‌ای تصادفی، و بنابراین یک نماینده واقعی از آنها جمع‌آوری شود. بویژه درکشورهای اسکاندیناوی، که ثبت وقایع حیاتی سابقه‌ای طولانی از دقت و کامل بودن دارد، و دفتر ملی ثبت دوقلوها براین اساس به‌وجود آمده‌است که در خدمت تحقیقات پزشکی و اجتماعی قرار بگیرد.

نتایجی از دفتر ثبت دوقلوهای دانمارکی قبلاً داده شده‌است. این دفتر ثبت شامل همه

دوقلوهایی است که در طول دوره معینی (۱۹۱۰ - ۱۸۷۰) در دانمارک متولد شده‌اند. از ۳۷۹۱۴ زایمان دوقلو که در طول این مدت رخ داده‌است، بیش ازنصف زوجها قبل از شش سالگی با مرگ یکی یا هردو از هم گسیخته‌اند و این موارد پیگیری نشده‌اند. حدود ۴۰ درصد زوجها از یک جنس نبودند و در نتیجه برای تحقیقات در مورد بیماریهایی که در یک جنس بیش از جنس دیگر ظاهر می‌شود، مفید نبوده‌اند. از باقیمانده، حدود ۶۰ درصد دوقلوها، ناهمانند و از یک جنس و ۴۰ درصد همانند و تبعاً همجنس بوده‌اند. (به عبارت دیگر، حذف کردن زوج دوقلوهای مختلط از مجموعه دوقلوهای زنده‌مانده به نسبت همانند به ناهمانند را از صورت متداول ۲۰:۸۰ به ۴۰ به ۶۰ تغییر داد.

تعیین نوع دوقلوها. پس از شناسایی و پیگیری دوقلوها از طریق دفاتر ثبت جمعیت و طرق مشابه، مسئله تعیین نوع آنها با روشهایی که بتوان به‌طور گسترده و به‌سهولت به‌کار برد، مطرح می‌شود. در مطالعه تعداد کمی از دوقلوها می‌توان از تکنیکهای بهبودیافته‌ای برای مقایسه مشخصه‌های توارثی نظیر گروههای خونی یا الگوهای پروتئین خونی استفاده کرد تا ملاحظه شود که آیا دوقلوها از کل این جهات واقعاً همانند هستند یا خیر. از اثر انگشت، لحن صدا و سایر مشخصات جسمی نیز می‌توان استفاده کرد. در مطالعاتی با مقیاس وسیع که در آن خود موردها را نمی‌توان معاینه کرد و تنها با پرسشنامه‌های پستی می‌توان مورد پرسش قرار داد، روشهای آسانتری مورد نیاز هستند. بنابراین جای خشنودی است که پاسخگویی به یک سئوال به‌طور تعجب‌آوری در تمیز بین زوجهای دوقلوهای همانند از ناهمانند بسیار تاثیر دارد. این سئوال ممکن است به این شکل باشد: "آیا شما همانند دو نخود در یک پوسته هستید؟" و توجه به این نکته موجب دلگرمی است که وقتی این سئوال پرسیده شد، به اثبات رسید که بیش از ۹۵ درصد از زوجهایی که پاسخ مثبت دادند در آزمایش خون و سایر آزمایشها، همانندند و وقتی هر دو پاسخ منفی می‌دهند، بیش از ۹۵ درصد ناهمانندند، وبالاخره در مورد ۲ درصد عدم انطباق وجود داشت.

به محض اینکه زوج دوقلوها رده‌بندی شدند، سابقه بیماری آنها باید مشخص شود. این کار را می‌توان با جمع‌آوری اطلاعات از سوابق بستری شدن در بیمارستان یا گواهی فوت در دوره‌ای شامل چند سال یابه‌طور مستقیم ازطریق پرسش در مورد بیماریهای گذشته یا وجود

علائم بیماریهای مزمن نظیر بیماری مزمن قلبی یا روماتیسم، انجام داد.

سایر کاربردهای مطالعات روی دوقلوها. مطالعه روی دوقلوها می‌تواند به تشخیص تأثیر عوامل محیطی در بیماری نیز کمک کند. برای مثال، چنین تصور می‌شود که مصرف سیگار موجب ابتلا به سرطان ریه و سایر بیماریهای ریوی می‌شود. نظیر حالت سرطان ریه، می‌توان استدلال کرد که فردی هم آمادگی گرایش به سیگار، و هم استعداد خاصی برای ابتلا به بیماریهای ریوی نظیر برونشیت مزمن را به ارث می‌برد. اگر این مطلب صحیح می‌بود، پیوند آشکار بین مصرف سیگار و ابتلا به برونشیت رابه عنوان وصفی که حداقل بخشی از آن ناشی از سایر عوامل بوده و نه دلیلی برای اینکه سیگار کشیدن موجب برونشیت شده باشد، می‌شود نادیده گرفت.

بررسی اعتیاد به سیگار در دوقلوهای همانند و ناهمانند نشان داده که در حقیقت نوعی عامل ژنتیکی در اعتیاد به سیگار تأثیرگذار است. اگر به مصرف سیگار به عنوان یک بیماری نگاه کنیم، مانند قبل می‌توانیم نرخ هماهنگی فرد مبتلا در دوقلوهای همانند را با دوقلوهای ناهمانند مقایسه کنیم. این مقایسه نشان می‌دهد، اگر یک فرد از دوقلوها سیگار بکشد، سیگاری بودن زوج دیگر در دوقلوهای همانند محتملتر از حالتی است که ناهمانند باشند (و بنابراین در این حالت، ارتباط نزدیکتر از خواهر و برادر معمولی نیست).

به دلیل این عامل ژنتیکی در سیگاری بودن، تاثیر مستقل مصرف سیگار بر روی ابتلا به برونشیت باید با مقایسه فراوانی برونشیت در دوقلوهای همانند که وضعیت ژنتیکی یکسان ۳ـ دارند ولی مقادیر متفاوتی سیگار مصرف می‌کنند، ارزیابی شود. یک بررسی بر مبنای مطالعه‌ملی دوقلوهای سوئدی دقیقاً این کار را انجام داده و نتیجه برای هر دونوع دوقلو در جدول ۳ داده شده‌است. این نتایج صرفاً به شکل نرخ شیوع تنظیم شده‌است که درصد افراد را درهرگروه که به سرفه‌های مزمن مبتلا هستند، نشان می‌دهد. بدیهی است در بین هرگروه ازدوقلوهای همانند، نرخ شیوع سرفه مزمن برای سیگاریها بیشتر از غیرسیگاریها است. به عبارت دیگر حتی وقتی که زمینه ژنتیکی یکسان باشد، به نظر می‌رسد مصرف سیگار بابرونشیت بیشتری همراه باشد و بنابراین محتملاً علت آن است.

نظریه‌های مرتبط. استفاده از دوقلوها محدود به مطالعات پزشکی یا زیست‌شناسی نیست.

جدول ۳ـ میزان شایع بودن سرفه در بین سیگاریها و غیرسیگاریها در زوج دوقلوهای ناهماهنگ

	میزان شایع بودن سرفه (درصد)		
	مجموع	غیرسیگاریها	سیگاریها
دوقلوهای همانند			
مرد	۲۷۴	۷/۷	۱۴/۶
زن	۲۶۴	۷/۶	۱۳/۶
دوقلوهای ناهمانند			
مرد	۷۳۳	۵/۵	۱۲/۳
زن	۶۵۳	۵/۷	۱۴/۵

از آنهادر مطالعاتی نظیر خواندن هم استفاده شده‌است. فکر استفاده از دوقلوهای همانند به عنوان کنترل در بررسی مصرف سیگار حالت خاصی از ایده معروف آماری استفاده از "بلوکهای" همگن برای آزمون آثار مختلف است. ابتدا در آزمایشهای کشاورزی، به موازات ایده استفاده از دوقلوها گاهی گروهی از گیاهان همزاد در معرض تیمارهای مختلف نظیر رژیم غذائی برای اصلاح دقت افزایش وزن متوسط ناشی از رژیمهای مختلف قرار می‌گیرند. البته، ممکن است به گروه همزادان زیادی نیاز پیدا کنیم، اما افراد کمتری به علت جوربودنی که گیاهان همزاد دارند به تعدادی کمتر نیاز خواهدبود.

همکاری بین المللی آینده. نتایج حاصل از دفتر ثبت دوقلوهای اسکاندیناوی نشان می‌دهد که این داده‌ها تاچه اندازه می‌توانند مفید باشند. متأسفانه، برای بیماریهایی که کمتر رواج دارند حتی دفاتر ثبت گسترده ملی نیز، موارد کافی از بیماری خاصی (مانند اشکال مختلف نارساییهای قلبی) را شامل نمی‌شوند تا امکان تحلیل آماری با جزئیات کافی موجود باشد. لذا سازمان بهداشت جهانی "دفتر ثبتی متشکل از همه دفاتر ثبت" را برای جمع‌آوری داده‌ها به سبکی هماهنگ در بسیاری از کشورها و یکجا گردآوردن و تحلیل متمرکز آنها به‌وجود آورد. به این ترتیب امیدواریم بتوانیم از فرصتهای منحصر به فردی که مطالعه دوقلوها را در بیماری و سلامت فراهم می‌آورد، استفاده کاملی کنیم.

مسائل

۱- چرا گروههای داوطلب در مطالعه روی دوقلوها به نتایج اریب منجر می‌شود؟ (راهنمایی: ن.ک. به مقاله مه‌یر)

۲- با مقایسه مخاطره ابتلای دوقلوهای همانند به بیماری با مخاطره ابتلای دوقلوهای ناهمانند آزمایشگر هر نوع عامل ــــــــــــــ را و در نتیجه هر نوع تفاوت مشاهده شده بین دو نوع دوقلو را می‌توان به عاملهای ــــــــــــــ نسبت داد.

۳- در داده‌های جدول ۳، تنها آن زوج دوقلوهایی درنظر گرفته می‌شوند که:

الف) هر زوج از دوقلوها سیگاری یا غیرسیگاری هستند.

ب) یک زوج سیگاری و زوج دیگر غیرسیگاری است.

پ) هر دو زوج یایکی از دو زوج یا هیچکدام از دو زوج سیگاری نیستند.

۴- با مراجعه به جدول ۳، آیا می‌توانید داده‌های مربوط به شایع بودن سرفه را از درصد به موارد وقوع (تعداد حالتها) تبدیل کنید؟ اگر می‌توانید این تبدیل را برای دوقلوهای ناهمانند نیز انجام دهید. اگر نتوانستید، بگویید به چه نوع اطلاعات دیگری نیاز دارید.

۵- توضیح دهید که چگونه می‌توان از مطالعه روی دوقلوها برای آشکارکردن موارد زیر استفاده کرد:

الف) اثر عوامل ژنی در بیماریها.

ب) اثر عوامل محیطی دربیماریها.

منابع

M.G.Bulmer. 1970. *The Biology of Twinning in Man.* New York: Oxford University Press.



وضع اسفناک نهنگها

دی. جی. چپمن[1] دانشگاه واشنگتن

در فاصله زمانی بین پایان جنگ جهانی اول و سال ۱۹۶۰، انواعی چند از نهنگها در اقیانوس پیرامون منطقه قطب جنوب، پایه یک صنعت مهم بودند. این پستانداران عظیم‌الجثه (بزرگترین پستاندارانی که در کره زمین وجود داشته‌اند)، برای تهیه روغن حیوانی و به میزانی کمتر، به منظور تهیه غذا و گوشت (برای مصارف انسانی) و همچنین برای تهیه فراورده‌های فرعی

۱ـ *D. G. Chapman*

بیشماری تحت جستجو بوده‌اند. در روزگاران قدیم، صیادان نهنگ در قایقهای کوچک به راه می‌افتادند و برای صید چنین منابع عمده گوشت، خطرهای بزرگی را تحمل می‌کردند. انسان به صورتی که در داستان موبی‌دیک[۱] به‌تصویر کشیده شده‌است، در قایقهای کوچک و با اسلحه‌های اولیه، به شکار نهنگها تا اواخر قرن نوزدهم ادامه داد. در قرن بیستم صید نهنگ با استفاده از زوبینهای منفجرکننده، کشتیهای بزرگ، و قایقهای صیادی توانا مجهز به رادار، بسیار پیشرفته و مدرن شد، به‌صورتی که اشتغال به صید نهنگ در اقیانوسهای طوفانی و خطرناک نزدیک ناحیه یخی قطب جنوب قابل انجام بود.

این ناحیه از جهان، درحالی که برای زندگی انسانها نامناسب است، برای نهنگها بسیار جذاب می‌باشد، زیرا در طول تابستان نیمکره جنوبی، آبها با گیاهان کوچکی شکوفا می‌شوند که هزارها جانور ریز را که عموماً به پلانکتون موسومند، تغذیه می‌کنند. بعضی انواع نهنگها این پلانکتونها را با کشیدن حجم زیادی از آب به درون دهانهای بسیار بزرگشان، به‌وسیله صافیهای غربال مانندی که به صفحه‌های بالن‌دار موسومند، صید می‌کنند (به همین دلیل این گروه از نهنگها را نهنگهای بالن‌دار[۲] می‌نامند). این نهنگها دندان ندارند و ماهی یا سایر پستانداران دریایی را نمی‌خورند. بزرگترین نهنگهای بالن‌دار و درواقع بزرگترین همه نهنگها، نهنگهای آبی رنگ هستند که ممکن است طول آنها به ۱۰۰ فوت (پا) برسد، گرچه طول ۷۰ تا ۸۰ فوت، اندازه بسیار معمولی آنهاست.

نهنگهای آبی رنگ

بلافاصله بعد از جنگ جهانی دوم، اروپا و ژاپن نیازهای مبرم زیادی ازجمله به روغن حیوانی، داشتند. لذا افزایش تعداد قایقهای صید نهنگ در منطقه قطب جنوب شگفت‌آور نبود؛ به‌علاوه پیشرفت فناوریها در طول جنگ، صنعت مربوط به استفاده از صید نهنگ را کاراتر کرده بود. نتیجه آنکه، برخی از طرفداران محیط زیست هراس داشتند که نهنگهای منطقه قطب جنوب، بویژه نهنگهای آبی کاملاً ازبین بروند. شکل ۱، صید سالیانه نهنگهای آبی را در اقیانوسهای

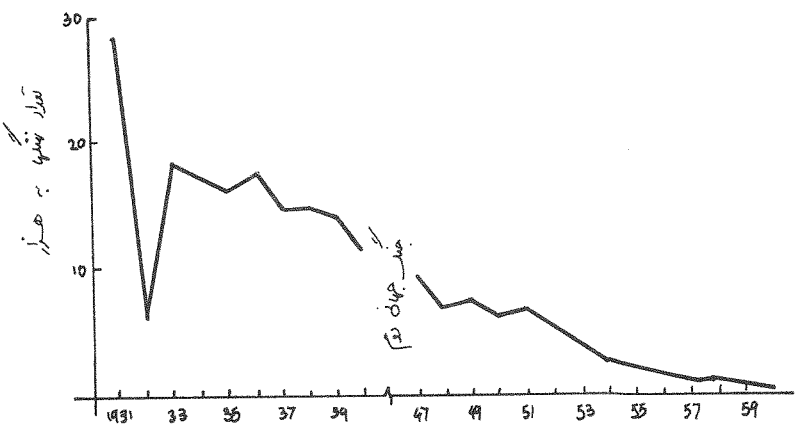
۱ـ *Moby Dick*

۲ـ *Baleen Whales*

نیمکره جنوبی، در دهه قبل از جنگ و در دوره بلافاصله بعد از جنگ تا ۱۹۶۰ نشان می‌دهد. ارائه مدرک برای نمایش دلیل نگرانی در مورد نهنگهای آبی آسان، ولی صید سایر انواع، پایدار یا درحال افزایش بود. برخی از افراد که در ارتباط با صنعت استفاده از صید نهنگ بودند، دلایلی غیر از کم‌کردن تعداد نهنگهای آبی برای کاهش صید نهنگ آبی ارائه می‌دادند و در کار صید مخالف پذیرش محدودیتها بودند. بدین‌ترتیب کمیسیون بین‌المللی ناظر بر صید نهنگ که در سال ۱۹۴۶ برای اداره منابع تشکیل شده‌بود، کانون بحث وگفتگو شد. این کمیسیون دارای نمایندگانی از همه کشورهای علاقه‌مند به این موضوع می‌باشد. آیین‌نامه‌های اندازه صید و سهمیه‌ها را مشخص می‌کند و دارای این اختیار و توانایی است که صید انواعی را که به نظر می‌رسد در خطر نابودی هستند، منع می‌نماید.

کمیسیون، گروه مطالعه‌ای برای جمع‌آوری همه داده‌ها و تهیه و گسترش روشهایی آماری برای تهیه پاسخ به سئوالهایی نظیر: چند نهنگ موجودند که در منطقه قطب جنوب زندگی می‌کنند؟ هرسال چند نوزاد نهنگ متولد می‌شوند؟ چند نهنگ به علل طبیعی هرسال می‌میرند؟ چگونه عاملهایی که انسان بر آنها کنترل دارد بر نرخهای زاد و مرگ نهنگها اثر می‌گذارند؟ تعیین نمود.

شکل ۱ـ صید نهنگهای آبی در منطقه قطب جنوب، فصلهای ۳۱ - ۱۹۳۰ تا ۶۰ - ۱۹۵۹ به استثنای دوره جنگ جهانی دوم.



روشهای شمارش

اکنون اولین و شاید اساسی‌ترین سئوال را بررسی می‌کنیم: چند نهنگ از یک نوع خاص وجود دارند؟ نهنگها متأسفانه یک‌جا ثابت نمی‌مانند که بتوان آنها را شمرد. آنها در نواحی وسیعی پراکنده‌اند و بیشتر وقت خود را زیر آب می‌گذرانند، گرچه در فواصل منظمی برای تنفس به سطح آب می‌آیند. به‌علاوه، اقیانوسهای جنوبی بخش وسیعی از جهان را شامل می‌شوند؛ ناحیه زندگی نهنگها متجاوز از ۱۰میلیون مایل مربع است، ناحیه‌ای که بزرگتر از همه آمریکای شمالی است. چندین راه استاندارد برای براوردکردن جمعیت جانوران وحشی وجود دارد که همه آنها دربرگیرنده برخی فنهای آماری هستند. ما سه‌تای آنها را شرح می‌دهیم:

روش علامت‌گذاری یا بگیر و بازبگیر. اولین روش براورد، مستلزم علامت‌گذاری تعدادی از نهنگهاست. یک استوانه فلزی به طول یک فوت به قشر چربی ضخیمی که زیر پوست نهنگ است، شلیک می‌شود. اگر بعداً برخی نهنگهای علامت‌دار شکار شوند، اطلاعاتی درباره حرکت، نرخ صید آنها، و نسبت نهنگهای علامت‌دار در کل مجموعه نهنگها فراهم می‌شود. مفید بودن اطلاعات اخیر را به‌وسیله شبیه‌سازی چنین آزمایشی با یک قوطی از مهره‌ها به آسانی می‌توان دید. فرض کنیم که مثل جمعیت نهنگها، تعداد مهره‌های درون قوطی مجهول باشد. حال به تصادف چند مهره (مثلاً ۱۰تا) از قوطی خارج کرده، روی آنها را علامت گذاشته و به قوطی بر می‌گردانیم. سپس کل مهره‌های قوطی را درهم کرده و نمونه دیگری از مهره‌های آن استخراج می‌کنیم. مهره‌های علامت‌دار و بدون علامت این نمونه را می‌شماریم. اگر در این نمونه، مهره‌های بدون علامت چهار برابر مهره‌های علامت‌دار باشند، استدلال می‌کنیم که برای کل مهره‌های قوطی نیز این نسبت درست است؛ اما چون کلاً ۱۰ مهره علامت‌دار در قوطی وجود دارد، استنباط می‌کنیم که ۴۰ مهره بدون علامت در قوطی موجود است و یا در کل ۵۰ مهره در قوطی وجود دارد.

این طرح ساده درباره بسیاری از جامعه‌های جانوران به‌کار رفته‌است. البته در عمل مشکلات بدیهی بسیاری وجود دارند، و این مطلب بخصوص برای نهنگها درست است. مثلاً چگونه بدانیم نشان فلزیی که به قشر چربی شلیک می‌شود واقعاً در این قشرنفوذ کرده، کمانه نکرده و خارج نشده است؟ آیا خدمه کشتی که نهنگ صیدشده را تکه تکه کردند با دقت به

دنبال یافتن علامت بودند؟ ممکن است حتی استوانه فلزی به درازی یک فوت هم، در شرایط کاری توأم با سرما و طوفان، وقتی حجمی تقریباً به اندازه یک خانه را تکه تکه می‌کنند بسادگی دیده‌نشود. به‌علاوه، برخلاف مهره‌ها، نهنگها در دوره‌ای از سالها متولد می‌شوند و می‌میرند. همه این پیچیدگیها و مشکلات، به دقیق‌سازی و گسترش آزمایش ساده‌ای که در اینجا طراحی شده‌است، نیاز دارند. لازم است یک سری از آزمایشها که طی سالهای متمادی بسط یافته‌اند، داشته باشیم و از شیوه‌های مقایسه‌ای استفاده کنیم. مثلاً اگر گروهی از نهنگها در سال ۱ و گروهی به همان اندازه در سال ۲ علامت‌گذاری شوند، آنگاه بعد از سال ۲، نسبت بازیافت نهنگهایی که در سال ۱ علامت‌گذاری شده‌اند به بازیافت نهنگهایی که در سال ۲ علامت‌گذاری شده‌اند، نسبت نهنگهای علامت‌داری در گروه ۱ را که در طول یک سال مرده‌اند، منعکس می‌کند. این مرگها ممکن است به‌صورت طبیعی و یا به‌دست صیادها انجام شده‌باشند. به‌علاوه، این نسبت، معیاری معتبر از این مرگ ومیر است زیرا صورت و مخرجش به‌طور مساوی تحت تأثیر خطاهای ممکنی هستند که در بالا فهرست شده‌اند. چنین بررسی تطبیقی تنها یکی از چند شیوه آماری است که در تحلیل داده‌های علامت‌گذاری نهنگها مورد استفاده قرار می‌گیرد.

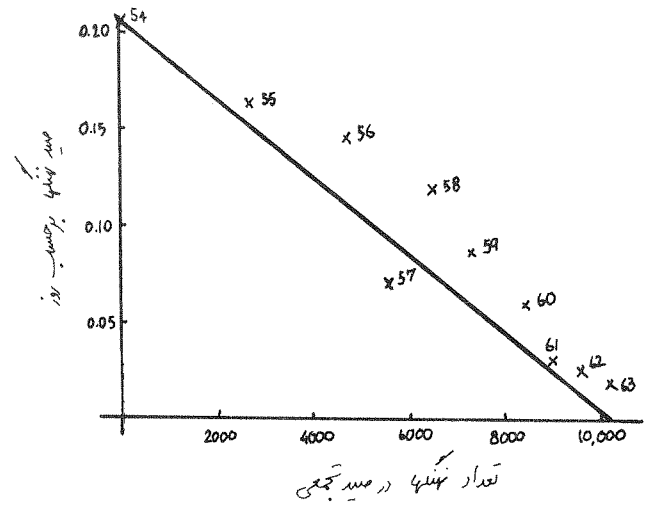
روش صید روزانه. روش دوم براورد، مبتنی بر تغییرات نرخ صید نهنگهاست. نرخ صید، عمدتاً به فراوانی مشاهده نهنگها بستگی دارد، و اگر سایر شرایط یکسان باشند، این نرخ به تراکم آنها وابسته است. پس صید روزانه، میزان تراکم را منعکس می‌کند. چگونه می‌توان این را به‌صورت اعداد مطلق برگرداند؟ اگر تغییر در صید روزانه تماماً نتیجه صید آنها به‌وسیله انسان باشد، آنگاه انجام این برگردان آسان است؛ اگر صید ۲۵۰۰۰ نهنگ در یک فصل، نرخ صید را برای فصل بعدی ۱۰٪ پایین بیاورد، آنگاه در آغاز فصل اول باید ۱۰×۲۵۰۰۰ یا ۲۵۰۰۰۰ نهنگ وجود داشته‌باشد.

ازطرف دیگر، وضع موجود، پیچیده‌تر از این مثال ساده‌است. کشتیهای صید نهنگ در ناحیه‌ای وسیع، در شرایط مشکل به صید می‌پردازند؛ به قسمی که میزان صیدها شدیداً نوسان دارد. شرکتهای صید نهنگ فناوری جدید را برای بهبود کارایی خود به‌کار می‌گیرند. به‌علاوه، مجدداً تأکید می‌کنیم که علل دیگری هم برای مرگ ومیر نهنگها وجود دارد و در این میان، زاد و ولد هم وجود دارد. در تعدیل براورد جمعیت، هردوی این عاملها را باید به‌حساب آورد. یک

راه غلبه بر بعضی ازاین مشکلات، عبارت از تعدیل تغییرات در کارایی، و همچنین پیگیری تغییر صید (تعدیل یافته) روزانه در طول دوره‌ای از چندین فصل است. شکل ۲، صید روزانه نهنگهای آبی را نشان می‌دهد که نمودار آن در مقابل صید تجمعی کشتیهای کارخانه مربوط به صنعت صید نهنگ در طول فصول ۱۹۵۴ - ۱۹۵۳ تا ۱۹۶۳ - ۱۹۶۲ رسم شده‌است؛ زمانی که تعداد زاد و مرگها کاملاً کم بوده‌اند. هرچه تعداد نهنگهای صیدشده بیشتر می‌شد، صید روزانه به‌صورتی یکنواخت کاهش می‌یافت. این نمودار القا می‌کند که در ۱۹۵۳ تنها ۱۰۰۰۰ تا ۱۲۰۰۰ نهنگ آبی وجود داشت و این تعداد در سال ۱۹۶۳ به حدود ۱۰۰۰ نهنگ نزول کرد. همان‌طور که اشاره شد، دقیق‌سازیهای آماری وجود دارند و نتایج حاصل از این راه را باید با براوردهای حاصل از راههای دیگر ترکیب کرد.

تحلیلهای سنی. روش صید روزانه، برای جوامعی که بشدت در حال کم شدن هستند، خوب عمل می‌کند، اما در وضعیتهای دیگر، وجود پیچیدگیها و تصحیحها، آن راکم‌استفاده می‌سازد. ولی روش سومی نیز وجود دارد که در آن از سن نهنگها استفاده می‌شود. همان‌طور که درختها در تنه‌های خود و ماهیها در فلسهایشان حلقه‌های سالیانه دارند؛ نهنگها هم حلقه‌هایی سالیانه در ترشح مومی گوش خود (حلقه گوش) دارند. سن نهنگهای نمونه‌ای که هرسال کشته

شکل ۲ـ صید روزانه نهنگ آبی (تعدیل‌شده برای بهبود کارایی) درمقابل صید تجمعی، ۱۹۵۳ تا ۱۹۵۴ و ۱۹۶۲ تا ۱۹۶۳



می‌شدند، به‌وسیله حلقه‌های گوش آنها تعیین می‌شد. به‌علاوه، اطلاعات مربوط به طول هر نهنگ که به دلایل تجارتی کشته می‌شد، این امکان را به ما می‌داد که رابطه سن و طول نهنگ را به‌دست آورده و براوردی برای سن هر نهنگ صیدشده محاسبه کنیم.

بنابراین ممکن بود که یک براورد آماری از تعداد نهنگهای ۴ساله در هر فصل و تعداد نهنگهای ۵ساله در فصل بعد تهیه کرد. چون نهنگهای ۵ساله یک‌سال، باقیمانده‌های ۴ساله سال قبل هستند، یک نرخ بقا یا برعکس یک نرخ مرگ و میر را می‌توان تعیین کرد. چون همه سنها براورد می‌شوند و چون انجام تعدیلهایی لازم است، نرخهای مرگ ومیر براورد شده، نوسان زیادی دارند. اما با میانگین‌گیری روی رده‌های چندین سال، و روی ناحیه‌ها و فصلها، نتایج مفیدی می‌توان به‌دست آورد. به‌علاوه، با تحلیل آماری بعدی، ممکن است حتی اندازه خطاهای ممکن در چنین براوردهایی را ارزیابی کرد.این نرخهای مرگ ومیر در پیش‌بینی کردن آینده جمعیت نهنگها به ما کمک می‌کند.

نتایج. بنابراین سه روش برای براوردکردن اندازه‌های جمعیت و نرخهای مرگ ومیر داریم: روش علامت‌گذاری، روش صید روزانه، و تحلیل سنی. نتایج روشهای مختلف نسبت به هم بازبینی شدند و خوشبختانه براوردهای مختلف با یکدیگر هماهنگ بودند. منابع خطا با دقت بازبینی و حذف شدند، به قسمی که گروه بررسی نهایتاً نتیجه گرفتند که تعداد نهنگهای آبی حداکثر بالغ بر چندهزار و شاید کمتر از هزار باشد. لذا خطر واقعی انقراض این نوع نهنگ در منطقه قطب جنوب وجود داشت (تعداد کمی از نهنگهای آبی در اقیانوسهای شمالی نیز وجود دارند.) خوشبختانه کمیسیون بین‌المللی مربوط به صید نهنگ، به محض اینکه بررسی پایان گرفت، صید نهنگهای آبی را، ابتدا در بخش بزرگی از اقیانوسهای جنوبی و سرانجام درهمه آبهای جنوب خط استوا، منع کرد. فعلاً خیلی زود است که بقای درازمدت این نوع نهنگ را پیش‌بینی کرد؛ نهنگهای آبی گهگاه دیده می‌شوند، اما اینها احتمالاً بازماندگانی هستند که در بالا ذکر کردیم (نهنگها، اگر صیدی در کار نباشد، می‌توانند بیش از ۴۰ سال عمر کنند). می‌توانیم سئوال کنیم که آیا جمعیت نهنگها به سطحی تقلیل یافته که توالد و تناسل آنها زیر سطح لازم برای ادامه حیات این نوع نهنگ است یانه؟ اما برای پاسخ به این سئوال، باید چندسالی سپری شود.

نهنگهای خاکستری[۱]

نهنگهای خاکستری نوعی از نهنگها هستند که از لحاظ بزرگی در جهان مقام دوم را دارند و بخشی از خانوده نهنگهای بالن‌دار را تشکیل می‌دهند. طول این نهنگ حدوداً ده فوت کوچکتر از نهنگ آبی است. طی دهه ۱۹۵۰ از صید این نوع نهنگ سالیانه متجاوز از یک میلیون بشکه روغن حاصل می‌شد. با کم شدن تعداد نهنگهای آبی، نهنگهای خاکستری متحمل فشار اصلی بهره‌برداری شدند. همان روشهای تحلیل برای نهنگهای آبی، برای نهنگهای خاکستری نیز به‌کار رفت. درواقع، در اینجا تحلیل، به‌صورتی بحرانی‌تر موردنیاز بود، زیرا وضعیت موجودی نهنگ خاکستری مثل مورد نهنگهای آبی آشکار نبود. به‌علاوه، صیدهای نهنگ خاکستری هنوز خیلی زیاد بودند: در فصل ۱۹۶۲ - ۱۹۶۱ بیش از ۲۷۰۰۰ نهنگ خاکستری کشته شدند. گروه بررسی توصیه کرد که برای اجتناب از انقراض نسل، صید نهنگ خاکستری باید به ۷۰۰۰ یا کمتر تقلیل یابد. پیشنهاد چنین تقلیل شدیدی، به‌منزله ضربه‌ای برای کمیسیون بود؛ گروه بررسی پیش‌بینی می‌کرد که صید فصل آینده، صرف‌نظراز سهمیه‌ها، به ۱۴۰۰۰ نزول می‌کند. وقتی ارقام واقعی ثابت کردند که پیش‌بینی درست است، اکثر مناطق می‌خواستند به سمت تقلیلهای شدید مورد نیاز حرکت کنند اما بعضی از دست‌اندرکاران صید نهنگ قادر به جلوگیری از این کار شدند. فصل مصیبت بار دیگری موجب شد که درنحوه تفکر اعضای کمیسیون، تجدیدنظری به وجود آید و در ۱۹۶۵ بایک برنامه تقلیل اساسی در سهمیه‌ها توافق شد. اما تأخیر در دست‌یابی به این توافق و تأخیر بعدی در تقلیل سهمیه‌ها، موجب شد که مجبور شوند صید مجاز را حتی بیشتر هم کاهش دهند، و درواقع در اواسط دهه ۱۹۷۰ کمیسیون حمایت کامل از نهنگهای خاکستری در منطقه قطب جنوب و در اقیانوس آرام شمالی را اعلام کرد.

مهلت قانونی قطع فعالیت صید تجارتی نهنگ

وقتی که انواع عمده نهنگها (آبی و خاکستری) دروضعی حمایت شده قرارگرفتند، صید دریایی نهنگ در منطقه قطب جنوب تنها به‌وسیله دو کشور ژاپن و شوروی سابق ادامه یافت. این

۱ـ *Fin Whales*

کشورها به صید انواع دیگری از نهنگها پرداختند. ابتدا نهنگ سی[۱] وسپس نهنگ مینک[۲]. نهنگ مینک در مقام مقایسه، نهنگی بسیار کوچک و ماکسیمم طول آن حدود ۳۰پا است. درباره این انواع، اطلاعات کمی وجودداشت و تنها تعداد کمی از آنها را علامت‌گذاری کرده بودند. روشهای تعیین سن آنها نامطمئن و قابل بحث هستند. به‌علاوه، دانشمندان سریهایی طولانی از داده‌های مربوط به صیدها را در اختیار نداشتند، به‌طوری که ذخیره داده‌ها برای تعیین وضعیت موجودی این نوع نهنگها درمجموع ناکافی بود. بدین ترتیب در تمام مدت اواخر دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰ کمیسیون بین‌المللی مربوط به صید نهنگ درگیر مباحثه‌ای بزرگ بود. بسیاری احساس می‌کردند، وقتی که اطلاعات درباره میزان موجودی نهنگها تا این حد کم است؛ کمیسیون نباید اجازه صید آنها را بدهد. در این ضمن، کشورهایی جدید که با روش محافظه‌کارانه موافق بودند، به کمسیون پیوستند. سرانجام، کمیسیون در سال ۱۹۸۰ به برقراری قطع فعالیت درباره همه صیدهای تجارتی نهنگ، نه‌تنها در منطقه قطب جنوب، بلکه در همه جهان، رأی داد. درگذر از این راه‌حل، کمیسیون در ۱۹۹۰ یک ارزیابی جامع ازهمه موجودیهای نهنگ را پذیرفت تابه صورتی امیدوارکننده، عدم حتمیتهایی راکه در مباحثات اوج گرفته بودند، رفع کند.

آینده موجودیهای نهنگ

اکنون نزدیک ۲۵ سال است که بعضی ازخانواده نهنگها تحت حمایتند، اما هنوز اطلاعاتی دردسترس نیست که تعیین کند آیا آنها شروع به بهبود و ازدیاد کرده‌اند؟ یانه. اگر ارزیابی جامع، راه‌حل وضعیت نابودی شدید موجودی نهنگها باشد و بهره‌برداریهای جدید را هم درنظر بگیریم، لازم است که روشهای نوین برای تهیه و تحلیل اطلاعات به‌دست آورد. بیشتر اطلاعاتی که اکنون درباره نهنگها دردسترس هستند، از نهنگهای مرده درکشتارهای تجارتی جمع‌آوری شده‌اند. چنین داده‌هایی محدودیتهای خود را دارند و در هرحال طی قطع فعالیت صید دردسترس نیستند. روشهای جدید، شامل اعزام هیئتهایی برای مشاهده از کشتی یا

۱ـ *Cie*

۲ـ *Minke*

هواپیماست. درحالی که کار قابل ملاحظه‌ای برای برونیابی چنین اطلاعاتی حاصل از مشاهده در تعیین اندازه جمعیت، انجام شده است، گردآوری اطلاعات زیست‌شناسی دیگری بدین طریق بسیار مشکل است. روشهای بالقوه دیگر، عکسبرداری و نمونه‌های پوستی است. نهنگها علامتهایی طبیعی دارند که می‌توان آنها را با مشاهده و یا ازروی عکسها شناسایی کرد و این علامتها را می‌توان به‌عنوان مبنای شیوه‌های «علامت بگذار و بازبگیر» به‌کار برد. این علامتها را می‌توان جانشین علامتهای (استوانه‌های فلزی) کرد که سابقاً به‌کار می‌رفت و تنها وقتی نهنگ کشته می‌شد، دوباره به‌دست می‌آمد. همه اینها، زمینه‌های نوینی از پژوهش را، هم در فنون و هم در تحلیل آماری داده‌هایی که بدین صورت به‌دست می‌آیند، می‌گشاید.

در این میان، در حال حاضر کل مسئله کشتن نهنگها برای هدفهای تجارتی، به صورتی جدی مورد بحث قرار گرفته است. بعضی موافق این مطلب می‌باشند که نهنگها منبعی هستند که باید نظیر ماهیها و درختها مورد بهره‌برداری قرار گیرند و به‌علاوه انسان تنها با صید نهنگ می‌تواند منابع بزرگ پلانکتونها را کنترل کند. دیگران اظهار می‌کنند که نهنگها حیواناتی باهوش با ساختار اجتماعی پیچیده‌اند و صید نهنگ نه‌تنها بیرحمانه است بلکه نابودکننده این ساختار اجتماعی است. این افراد بر آن باورند که محافظت از نهنگها نشانگر نیت ماست از اینکه هماهنگ با جهان طبیعی زندگی کنیم.

مسائل

۱ـ به شکل ۱ رجوع کنید. میزان تقریبی صید نهنگهای آبی در منطقه قطب جنوب، در فصل ۱۹۳۱ تا ۱۹۳۲ چقدر بوده‌است؟ در فصل ۱۹۳۷ تا ۱۹۳۸ چقدر؟ نرخ تقلیل میزان صید بین دو فصل چقدر است؟ چه موقع میزان صید بالاترین مقدار را داشته است؟ چه موقع کمترین مقدار را داشته است؟

۲ـ سه روش شمارشی که در متن مقاله از آنها بحث شده، چه هستند؟

۳ـ آزمایشی را توصیف کنید که هدف آن، براورد کردن اندازه جمعیت انسانهای ساکن در یک جزیره دورافتاده، یا استفاده از یکی از روشهای مذکور در مقاله، باشد. چه فرضهایی می‌کنید؟

۴ـ به شکل ۲ رجوع کنید. میزان (تعدیل یافته) صید روزانه نهنگهای آبی در فصل ۱۹۵۵ تا

۱۹۵۶ (که با عنوان ۱۹۵۶ رسم شده) چقدر بوده‌است؟

۵ـ به شکل ۲ رجوع کنید. آیا می‌توانید برای رقم مربوط به ۱۹۵۸ «زیاد» نسبی را توصیف کنید؟

۶ـ از شکل ۲ نتیجه‌گیری شده بود که در ۱۹۵۳ تنها ۱۰۰۰۰ تا ۱۲۰۰۰ نهنگ آبی وجود داشتند. این تعداد چگونه به‌دست آمده‌است؟

۷ـ فرض کنید با روش تحلیل سنی معین شده‌باشد که نرخ مرگ و میر سالیانه نهنگها ۲۵٪ است. به‌علاوه فرض کنید که با روش علامت‌گذاری، براورد شده‌باشد که حدود ۷۰۰۰ نهنگ آبی در آغاز ۱۹۹۰ وجودداشتند. با فرض اینکه نرخ مرگ و میر در طول سالها عوض نشود، انتظار می‌رود که در آخر سال ۱۹۶۰ چند نهنگ هنوز زنده باشند؟

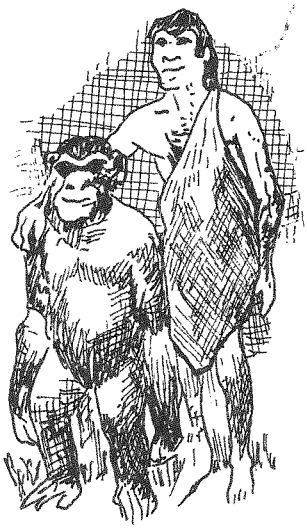
۸ـ مفهوم «اندازه اپتیمم موجودی» نهنگها چیست؟ چرا اصلاً باید اندازه اپتیممی وجود داشته باشد؟

منابع

D. G. Chapman. K. R. Allen. and S. J. Holt. 1964, "Reports of the Committee of Three Scientists on the Special Antarctic Investigations of the Antarctic Whals Stocks". *Fourteenth Report of the International Whaling Commission.* London: International Whaling Commission. pp 32-106.

J. A. Gulland. 1966, "The Effect of Regulation on Antarctic Whale Catches." *Journal du Consell, 30: 308-315.*

Scott McVay. 1966, "The Last of the Great Whals." *Scientific American 215(2): 13-21.*



اهمیت انسان بودن

دبلیو. دبلیو. هاولز[۱] دانشگاه هاروارد

در تابستان سال ۱۹۶۵، همکار دیرین‌شناس من، *برایان پاترسون*[۲]، مسئول یک هیئت اعزامی تحقیقاتی از هاروارد بود که در نزدیکی ساحل دریاچه ترکانا در شمال کنیا مشغول کار بودند. در مکانی موسوم به کاناپوی[۳]، هنگامی که او مشغول جستجو در رسوباتی بود که گمان می‌رفت

۱ـ *W. W. Howells*

۲ـ *Bryan Patterson*

۳ـ *Kanapoi*

به دوره چهارم[۱] زمین شناسی متعلق باشد (آخرین دوره زمین‌شناسی قبل از عصر کنونی)، قطعه فسیل مهمی را به‌دست آورد. این قطعه انتهای زیرین شکسته یک بازوی چپ بود که به‌سادگی تشخیص داده‌می‌شد که شبیه استخوان یک جانور انسان‌نما است؛ یعنی این فسیل، بازمانده از موجودی بود متعلق به گروهی که شامل انسانها و نزدیکترین خویشاوندان زنده آنها (میمونهای انسان‌نما) می‌شود اما متعلق به یک میمون نبود.

اهمیت خاص این فسیل چه بود؟ از شکل و اندازه آن بلافاصله می‌شد پی‌برد که این قطعه متعلق به یک گوریل، یا اورانگوتان و یا یک گیبون[۲] نبود (به‌هرحال دونوع آخر هیچوقت در آفریقا وجود نداشته‌اند). استخوان یادشده به‌طور شگفت‌انگیزی چنان با قطعه همانند خود در انسان امروزی شباهت داشت که تشخیص آن دو از یکدیگر ناممکن بود. اما قدمت رسوب مطمئناً به قبل از پیدایش وجود چیزی مثل بشر امروزی باز می‌گشت. بعد از پایان دوره عملیات، قدمت سنگهای بازالت حاصل از آتش‌فشان از بستری که بالای رسوب قرار گرفته بود، با استفاده از تاریخ‌گذاری رادیوایزوتوپی حدود $۲\frac{۱}{۲}$ میلیون سال برآورد شد. قدیمی‌ترین مرحله پیدایش موجودی انسانی که تا آن زمان به اثبات رسیده‌بود مربوط به اوسترالوپیتسینهایی[۳] است با مغز کوچک و فکهای بزرگ که به‌طور ایستاده راه می‌رفتند و قدمت فسیلهای بازمانده از آنها که به‌وسیله *لوئیس لیکی* در دره الدووی جورج[۴] پیدا شد، حدود $۱\frac{۳}{۴}$ میلیون سال محاسبه شده‌است. اگر این قطعه کوچک استخوان بازو متعلق به "انسان" یا از جنس "انسان‌سانان" بود به مفهوم اینکه به چنین موجودی تعلق داشت، در اینصورت با یک تک جهش، تاریخ تکامل مداوم بشر $\frac{۳}{۴}$ میلیون سال به دورتر می‌رسید.

اما یک مسئله وجود داشت. این قسمت از مفصل آرنج انسانها را می‌توان از قسمت همانند خود در اورانگها[۵]، گوریلها و گیبونها به آسانی تشخیص داد ولی نه مفصل آرنج شامپانزه‌ها را. گرچه بقیه استخوان یک شامپانزه کوتاهتر و کلفت‌تراست، این ناحیه به قدری در دوگونه مشابه

۱ـ *Pleistocene*

۲ـ *Gibbon*

۳ـ *Australopithecine*

۴ـ *Olduvai Gorge*

۵ـ *Orangs*

است که تشخیص بسیاری ازنمونه‌ها، را هنگام بررسی به منظور طبقه‌بندی غیرممکن می‌سازد. باوجود استفاده‌های مختلف از بازو، این قسمت ویژه چنان وجوه اختلاف اندک، ظریف و ناپایداری در اندازه و شکل دارد که روشهای بررسی معمولی را حتی به‌وسیله کارشناسان با مانع روبرو می‌کند. بنابراین مسئله به اینصورت بود: یا استخوان به اوسترالوپیتسین اولیه‌ای تعلق داشت که تا آن زمان در تاریخچه بلافصل اجدادی ما پیدا شده بود یا صرفاً بازمانده از یک شامپانزه اجدادی ما بود که دراینصورت ما می‌توانیم نفس راحتی بکشیم. اما آزمون چیزی قدیمی با یک وسیله جدید چطور است؟ آیا یک کامپیوتر می‌تواند نکته مفیدی به ما بگوید؟

البته یک کامپیوتر واقعاً چیزی به ما "نمی‌گوید"؛ بلکه صرفاً پاسخهای محتمل به سئوالات ریاضی را محاسبه می‌کند که اگرتلاش می‌کردیم با یک ماشین حساب ساده آنها رابه‌دست آوریم، عمر ما کفاف نمی‌داد. یک کامپیوتر با ظرفیتها و سرعت بی‌اندازه‌اش تلاش برای دست‌یابی به جوابهای درست را به تلاش برای عنوان کردن پرسشهای دقیق و مناسب مبدل می‌کند. آثار زیست‌شناختی ـمثل استخوانها و جمجمه‌هاـ برای طرح انواع خاصی از سئوالات قابلیت دارند. از آنجاکه ژنهای دو فرد از یک جامعه انسانی یا دو موجود از یک گونه عملاً قابلیت تعداد نامحدودی از ترکیبات مختلف را دارند، دقیقاً مثل هم نیستند (به استثنای مورد جالب دوقلوهای همانند). بنابراین، کاملاً صرف نظر از عادتهای مختلف استفاده، تغذیه، یا حادثه‌های دیگر رشد، مفصل آرنج انسانها به‌طور طبیعی از نظر جزئیات مربوط به شکل متفاوت است، گرچه این تفاوتها فقط در محدوده قالبی است که برای حرکت مفصل آرنج انسان اساسی است.

البته گونه‌های کاملاً متفاوت حیوانات دارای شکلهای کاملاً متفاوت در نواحی مختلف بدن هستند. هر مبتدی می‌تواند تفاوت بین دندان نیش یک گوشتخوار پستاندار که تاج آن به باریکی چاقو است را از دندان نیش یک گیاهخوار که دارای سطحی عریض برای خردکردن مواد گیاهی است، تشخیص دهد. اینها مربوط به وجوه اختلاف شاخص تکاملی آنها می‌باشد. در میان گیاهخوارها اختلافها کمتر است، و در میان گروههای گیاهخوار مثل خوکها (برای مثال خوکهای اهلی، گرازهای وحشی، گرازهای آفریقایی وغیره) یا فیلها، تشخیص بین گونه‌های مختلف کار متخصصین است که می‌توانند یک گنجینه اطلاعاتی درباره تاریخچه خوکها و فیلها یا درباره

گونه‌های دقیقی از حیوانات که در یک زمان خاص درگذشته در یک مکان تجمع فسیل مثل کاناپوی بوده‌اند، به‌دست آورند. بالاخره، برای اندامهای خاصی، مثل مفصل آرنج شامپانزه‌ها و انسانها، اختلافهای بین گونه‌ها ممکن است آنقدر کم باشد که (به‌طوری که قبلاً شرح داده‌شد)، تحت‌الشعاع تنوع موجود در داخل هرگونه قرار گیرد. این وضعیتی است که ما در اینجا با آن مواجه هستیم.

این مسئله فقط برداشتی از مشاهدات نیست و می‌توان آن را به‌صورت کمّی بررسی کرد. چندی پیش، پروفسور *ویلیام ال. اشتراوس*[۱] از دانشگاه جان هاپکینز، مردی با تجربه زیاد در چنین مطالعاتی، سعی کرد به این مسئله بپردازد؛ وقتی که قطعه همانندی از بازوی گونه‌ای از اوسترالوپیتسین‌ها در منطقه کرم‌درای[۲]، نزدیک پرتوریا در آفریقای جنوبی پیدا شد. در این مورد، واضح بود که استخوان متعلق به پارانتروپس[۳] یعنی گونه مورد بحث بود، زیرا قسمتهای دیگری از استخوانهای بدن و جمجمه همان گونه مشابه در آن منطقه پیدا شده‌بود و مشکل می‌شد آن را به موجود دیگری منتسب نمود. در اینجا مسئله این بود که آیا استخوان بیشتر مشابه به استخوان یک انسان بود یا یک میمون انسان‌نما؟ زیرا در جایگاه انسان قرار داشتن اوسترالوپیتسین‌ها در آن زمان چندان روشن نبود. پروفسور *اشتراوس* اقدام به چند اندازه‌گیری نمونه از استخوانهای انسان و شامپانزه به منظور پیدا کردن اختلافهای آنها کرد. او از لحاظ آماری اختلافهای[۴] معنی‌دار بین میانگینهای آنها پیدا نمود، اما اختلافهای مطلق بین آنها کم بود، و تداخل در اندازه‌های مربوط به انسان و شامپانزه به قدری زیاد بود که امکان منتسب نمودن قطعه متعلق به پارانتروپس به یکی از آنها ممکن نبود. در هیچ موردی اندازه‌های آن

۱ـ *William L. Straus*

۲ـ *Kromdraai*

۳ـ *Paranthropus*

۴ـ *خواننده ممکن است به خاطر بیاورد که آمارشناسها هنگامی می‌توانند به وجود اختلافی پی ببرند که احتمالاً نتیجه تأثیری واقعی باشد نه نتیجه تغییرات تصادفی، این اختلاف را "معنی‌دار" می‌نامند. به‌وسیله واژه "آماری" آنها به ما هشدارمی‌دهند که اندازه مطلق اختلاف ممکن است کوچک و بظاهر بی‌اهمیت باشد، به دلیل اینکه به موضوعهای موردبررسی بستگی دارد.*

خارج از دامنه اندازه‌های مربوط به انسان یا شامپانزه قرار نداشت، گرچه ارقام به‌دست آمده بیشتر نزدیک به میانگین یا حدمتوسط ارقام برای شامپانزه بود.

این راه‌حلی برای مسئله نبود و تا آنجا که استخوان بازو می‌توانست باعث روشنتر شدن ارتباط پارانتروپس با آن شود منجر به هیچ تصمیمی نشد. در چنین وضعیتی، ما به روشی نیاز داریم که به مقایسه تک‌تک اندازه‌ها محدود نباشد، و به‌نحوی شکل کلی استخوان یا قسمتی از آن را درنظر بگیرد، همان‌طور که چشم عمل می‌کند و درصورت وجود اختلافهایی واقعاً چشمگیر بین شکل دوگونه، روشی برای تأکید در مورد آنها دارد. اکنون دراینجا یک نکته مهم وجود دارد: سرانجام چنین موضوعهایی به‌صورت یک مسئله ریاضی درخواهد آمد. زیرا خود چشم نیز هرچند بندرت آگاهانه در ارزیابی متوسط اختلاف در تناسبها یا جنبه‌های پیچیده شکل یا درجه‌بندی تفاوت اهمیت آنها تلاش می‌نماید، و بالاخره سعی می‌کند درباره این احتمال نظر دهد که یک شکل کلی ویژه در یک مورد واحد، در محدوده اختلاف یک جامعه با جامعه دیگر نزدیکتر به چارچوب اساسی کدامیک از آنها قرار می‌گیرد. این سئوالها درباره کمیت و احتمال هستند، چه اندازه‌گیری شده‌باشند و چه نشده‌باشند. بدین‌ترتیب ماهیت آنها آماری است. به‌هرحال، یک نظر مبتنی بر شناخت کافی از اطلاعات موجود همواره به سبک و سنگین کردن احتمالات می‌پردازد. یک نکته مهم دیگر: زیست‌شناسها و انسان‌شناسها و دست‌اندرکاران بسیاری از علوم دیگر، اغلب در ریاضیات سطح بالاتر قوی نیستند، گرچه با فراست بسیار فقط محدوده حل مسائل به طریق خودشان را درک می‌کنند. از طرف دیگر، ریاضیدانها هرچند سرشتی بسیار مهربان دارند، معمولاً به اندازه کافی از ظرافتهای مربوط به مسائل زیستی آگاهی ندارند تا دریابند که تلاش یک زیست‌شناس برای استفاده از یک تحلیل ریاضی برای دست یافتن به چیست.

با وجود این، اگر این دو واقعاً با هم ارتباط برقرار کنند، بازده آنها از نظر دست‌یابی به راه‌حلهای جدید ممکن است زیاد باشد. و باید متذکر شویم که آموزش ریاضی در بین زیست‌شناسهایی که بهتر درک می‌کنند که چنین آموزشهایی به آنها چه عرضه می‌کند به‌طور چشمگیری در سالهای اخیر افزایش یافته‌است.

خوشبختانه، مسئله خاص فسیل کاناپوی خیلی پیچیده نیست، و چندی پیش راه‌حل آن

توسط *آر. ای. فیشر*[۱] آمارشناس و متخصص بزرگ انگلیسی ژنتیک به شکل تابع تشخیص[۲] ارائه شد. استفاده از این روش باعث کنار گذاشتن برخی کارهای بی‌حاصل می‌شود. مانند توجه کردن به تک‌تک اندازه‌ها، پی بردن به اکنونه تداخل مانع از تفاوت قائل شدن بین دومجموعه از نمونه‌ها نظیر مفاصل آرنج انسان و شامپانزه می‌شود، هرچند که متعلق به جانورانی کاملاً متفاوت باشند و یا عدم توانایی در مورد اینکه نمونه‌ای مانند پارانتروپس منطقاً نزدیکتر به یک گروه یا گروه دیگر قرار داده شود. تابع تشخیص مجموعه‌ای از وزنها را در اختیار دارد که در اندازه‌های مختلف یک نمونه ضرب می‌کند، مجموع حاصلضربها یک تک امتیاز تشخیص به‌دست می‌دهد که بهترین استفاده قابل حصول را از اطلاعات موجود در چند اندازه می‌نماید. با داشتن دو گروه، مثل انسانها و شامپانزه‌ها، محاسبات مجموعه بهینه‌ای از وزنهای ممکن را از اندازه‌های مورد استفاده توسعه می‌دهد.

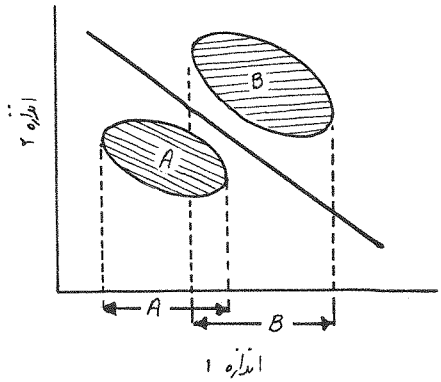
درنتیجه اختلافهای مهمی که به‌وسیله چشم قابل رؤیت نبوده و در میانگینها نیز به‌نظر نمی‌رسند به نحوی از یکدیگر تفکیک می‌شوند که جنبه‌های مربوط به شکل و اندازه که باعث تفاوت دو گروه از یکدیگر است، دقیقاً مورد تأکید قرار می‌گیرند. بدین معنی که با مقایسه تنوع موجود در محدوده یک مجموعه از مفاصل آرنج انسان یا شامپانزه، اختلافهای بین این مجموعه‌ها با روش ریاضی مورد جستجو قرار می‌گیرند؛ به‌طوری که امتیازهای تشخیص دوگروه تا حداکثر درجه ممکن از یکدیگر تفکیک شوند اما این تفکیک فقط محدود به اطلاعاتی است که در اندازه‌ها وجود دارد. بدین‌ترتیب، تداخل که مانند یک پوشش برای مخفی کردن تفاوتهای واقعی گروهها از یکدیگر عمل می‌کند، تقلیل یافته و یاحذف می‌گردد.

فکر اساسی مربوط به تابع تشخیص را می‌توان به کمک نمودار برای دو اندازه که به‌وسیله دو محور در نمودار شماره ۱ نشان داده می‌شوند، درک نمود. (این اندازه‌ها را می‌توان برای طول قد و دورکمر مردانی از دوگروه مختلف در نظر گرفت.) ناحیه‌های بیضی شکل، مربوط به گروه افراد جامعه‌های A و B می‌شوند. اگر با تصویرکردن روی محورافقی فقط به اندازه ۱ نگاه کنیم، تداخل قابل ملاحظه‌ای را خواهیم یافت. همین نکته در مورد اندازه ۲ صدق می‌کند. از

۱ـ *R. A. Fisher*

۲ـ *Discriminant Function*

شکل ۱ـ دواندازه باهم گروهها را به نحو بهتری از هم تفکیک می‌کند تا هریک از آنها به‌طور جداگانه.



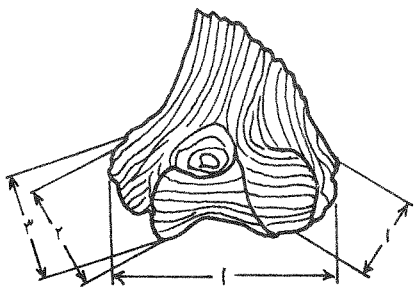
طرف دیگر، خط مایل به‌طور کامل دوجامعه را از هم جدا می‌کند. توضیح جزئیات ریاضی در اینجا مناسب نیست، اما نوشتن جمله قبلی بدین معنی‌است که بگوییم نگاه کردن به چیزی مثل (اندازه ۲)× ۲ + (اندازه ۱) امتیاز جدیدی به ما ارائه می‌کند، امتیاز تشخیص که بسیار بهتر از اندازه‌های ۱ یا ۲ به تنهایی، جدا کردن جامعه‌ها از یکدیگر را امکان‌پذیر می‌سازد. اگر بیش از دو اندازه دراختیار داشته‌باشیم، همانند مورد کنونی، در اینصورت منافع بالقوه زیادی در ترکیب کردن اندازه‌ها وجود خواهد داشت.

پروفسور *پاترسون* و من قویاً احساس می‌کردیم قطعه پیداشده در کاناپوی که از گروهی شامل جانوران انسان‌نما بازمانده است به موجودی انسانی تعلق داشت تا به میمونهای انسان‌نما. اما ما قصد داشتیم آن را به شیوه آماری نشان دهیم، نه اینکه صرفاً اظهار عقیده‌ای کرده‌باشیم که باعث شود افرادی دیگر، عقیده‌های مخالف با آن ابراز نمایند. به‌عنوان یک روش، ما استخوانهای بازوی انسان و شامپانزه را مورد بررسی قرار دادیم تا پی ببریم چه اندازه‌هایی بیشتر نشان‌دهنده تفاوتهایی بودند که در اغلب یا بعضی اوقات ما تصور می‌کردیم نمایان خواهند شد. شکل ۲ خود قطعه و برخی اندازه‌های مربوط به آن رانشان می‌دهد.

ابتدا فاصله کامل بین دو نقطه را در سراسر انتهای زیرین به‌عنوان اندازه کلی (اندازه ۱)

درنظر می‌گیریم. سپس قسمت داخلی یا زائده فوق لقمه‌ای میانی را که پیش‌آمدگی بیشتری دارد و در بعضی از شامپانزه‌ها کوتاه و پهن بوده یا کمی حالت سربالا دارد، و ما امیدوار بودیم که این حالت را به‌وسیله اندازه‌گیری از پایین‌ترین نقطه روی لبه قرقره تا زائده فوق لقمه‌ای همچنین تا نزدیک‌ترین نقطه روی شانه‌ای که درست بالای آن قرار داشت، به ثبت برسانیم (اندازه‌های ۲ و ۳). به گمان ما تفاوت نسبتاً بیشتری بین این اندازه‌ها، نشان‌دهنده یک زائده فوق لقمه‌ای با انحنای عمیقتر و حالتی بیشتر سربالا بود. همچنین ما برجستگی رو به‌عقب از قسمت مرکزی، یا قرقره‌ای، لبه مفصل، طول وعرض سطح بیضی شکل داخلی زائده فوق لقمه‌ای (هیچکدام از اینها در شکل نشان داده نشده‌است)، ارتفاع مایل زائده فوق لقمه‌ای مقابل یا جانبی را اندازه‌گیری کردیم. تصور ما بر این بود که این اندازه‌ها هرکدام به‌نوعی در انسانها و شامپانزه‌ها به تغییر گرایش دارند. اگرچه این امر در هیچکدام از دوگونه به‌صورت قاعده نیست (چنانچه تفاوتهای ثابتی وجود داشت، بدیهی است که تشخیص بین آنها دیگر چندان مسئله‌ای نبود). ما از مفهوم تابعی تفاوتهای احتمالی مطمئن نبودیم، اما به‌طور منطقی این اختلافها می‌توانند به اتصالات ماهیچه‌ای مربوط باشد که در رابطه با استفاده‌های ساده‌تر و قویتر از ماهیچه‌های خم‌کننده و بازکننده دست در شامپانزه، هنگام آویزان شدن به‌وسیله بازوها یا در نگهداری بدن

شکل ۲ـ قطعه استخوان بازوی کاناپوی و اندازه‌های برداشته‌شده.



یا راه‌رفتن بر روی زمین با شیوه تکیه بر روی بند انگشتهای میانی، که همه اینها مغایر با استفاده کلی، اما پیچیده‌تر و متنوعتر از دستها در انسانها می‌باشد.

اکنون این درست جایی است که موضوع همکاری نقش دارد. کار دیرین‌شناس یا مردم‌شناس این است که بر اساس دانش تخصصی خود اندازه‌هایی را پیدا کند که اطلاعات مهم و واقعی را در ارتباط با اختلافها دربر داشته‌باشند. کار آمارشناس این است که نشان دهد چطور اندازه‌ها را می‌توان کنار هم گذاشت تا اختلافها به منظور ارزیابی آشکار شوند. در این مورد همکاری به قدری گسترش یافته که زیست‌شناسها از قبل می‌دانند که آمارشناسها به چه صورت می‌توانند به آنها کمک کنند و بنابراین ما کار را بر آن اساس طراحی کردیم.

ما ۴۰ استخوان انسان را در موزه پی‌بادی[۱] در هاروارد و ۴۰ استخوان شامپانزه را در موزه جانورشناسی تطبیقی هاروارد و موزه آمریکایی تاریخ طبیعی در نیویورک اندازه‌گیری کردیم. همانگونه که در اندازه‌های به‌دست آمده توسط *اشتراوس* مشاهده شد، تداخل اندازه‌ها در میان انسان و شامپانزه زیاد بود اما متوسط اختلافها ناشی از انتخاب ویژه اندازه‌ها بود. مقادیر میانگین برای دوگروه و ارقام برای قطعه‌های کاناپوی و کرم‌درای (ارقام دومی از دو قالب به‌دست آمدند.) برحسب میلی‌متر در جدول ۱ ارائه شده‌است.

مقادیر مربوط به شامپانزه، به‌عنوان یک نمونه، ممکن است برحسب تصادف به‌طور متوسط اندکی بیشتر باشد. قطعه مربوط به پارانتروپس آشکارا در تمام ابعاد کوچک می‌باشد. بنابراین هنگام تماشای اجمالی فهرست، مانند استخوان انسان به نظر می‌رسد. این امر لزوماً بدین معنی نیست که ارتباطهای شکلی آنها با ارتباطهای شکلی مربوط به انسانها تطابق داشته‌باشد. تطابق اندازه‌های کاناپوی با میانگینهای مربوط به انسان (در این نمونه خاص) در تمام اندازه‌ها بسیارنزدیک است (نزدیکتر از آنچه که انتظار می‌رود استخوان یک انسان که به‌طور تصادفی برداشته‌شده‌باشد، در تمام اندازه‌هایش چنین باشد).

برای حصول اطمینان از این نزدیکی آشکار، از ارقام مربوط به انسان و شامپانزه یک تابع تشخیص محاسبه نمودیم. فقط برای هفت اندازه و نمونه‌هایی این چنین کوچک، محاسبات را

۱ـ *Peabody Museum*

جدول ۱ـ اندازه‌ها

اندازه	میانگین شامپانزه	میانگین انسان	کاناپوی	پارانتروپس	بردار مقیاس‌بندی شده
۱ـ فاصله عرض بین لقمه‌ها (اپی‌کندیل)	۶۴/۱	۵۸/۰	۶۰/۲	۵۳/۶	-۰/۰۹
۲ـ فاصله بین قرقره و زائده فوق‌لقمه‌ای	۴۴/۸	۴۰/۷	۴۱/۷	۳۳/۶	+۰/۴۰
۳ـ فاصله قرقره تاناحیه لقمه‌ای	۴۱/۳	۳۸/۸	۳۹/۴	۳۲/۱	-۰/۶۲
۴ـ لبه خلفی قرقره‌ای	۲۶/۴	۲۲/۱	۲۲/۲	۱۹/۹	+۰/۱۱
۵ـ طول زائده فوق‌لقمه‌ای میانی	۲۴/۷	۲۰/۳	۲۰/۸	۱۵/۵	+۰/۱۹
۶ـ عرض زائده فوق‌لقمه‌ای میانی	۱۲/۸	۱۲/۶	۱۳/۹	۱۰/۴	-۰/۳۲
۷ـ ارتفاع زائده فوق‌لقمه‌ای جانبی	۳۱/۵	۲۶/۷	۲۷/۶	۲۴/۹	+۰/۵۶

می‌توانستیم با دست انجام دهیم، گرچه به بهای اندکی نبود. به زبان فنی، می‌باید ماتریسهایی را از جمع حاصلضربهای دوبه‌دوی تمام اندازه‌های مربوط به همه استخوانهای موجود در هر گروه و همچنین کل آنها باهم تشکیل دهیم. مراحل دیگر مستلزم عکس کردن یک ماتریس و تعیین کردن ریشه‌های راکد دیگری بود. به‌دست آوردن عکس یک ماتریس با اندازه معمولی ۷×۷ کار خسته‌کننده‌ای است که دستخوش خطااست. انجام تمام این کارها منجربه پیدا کردن تابع تشخیص می‌شود، که هفت اندازه را از یک نمونه برمی‌دارد، هر اندازه را در یک وزن خاص مربوط به آن اندازه ضرب می‌کند، وسپس این حاصلضربها را با هم جمع نموده و امتیاز تشخیص را به‌دست می‌دهد. این کار مستلزم انجام محاسبات زیادی است و فقط می‌توانیم بگوییم که داشتن یک کامپیوتر برای انجام چنین کاری در ظرف چند دقیقه بسیار مغتنم شمرده می‌شود. زمانی را که می‌باید در انتظار باشیم تا رنگ خشک شود و یا یک فیلم ظاهر شود با مقایسه با این کار طولانی به نظر می‌رسد، و چنین محاسبه‌ای آشکارا باعث دلگرمی بسیار برای کارهایی مثل کار فوق شده‌است. آخرین ستون جدول ۱ وزنهای واقعی راکه در تابع تشخیص به‌کار برده شده است، ارائه نمی‌کند؛ بلکه یک شکل دوباره به مقیاس‌بندی مجدد از وزنها را به

ترتیب اهمیتشان از یک چشم‌انداز مناسب می‌دهد (زیرا برای مثال اندازه کوچکی مانند طول انگشت شست نسبت به اندازه بزرگی، مثل قامت یک فرد به‌منظور اینکه موثر واقع شود، ممکن است به وزن خیلی بیشتری در تابع نیاز داشته‌باشد). این ارقام نشان می‌دهد که چطور تعدادی از اندازه‌ها با هم ترکیب می‌شوند برای اینکه یک تک الگو از بزرگترین اختلاف بین دو گروه را به وجود آورند.

همان‌طور که انتظار می‌رفت، دو اندازه‌ای که حالت سربالا بودن زائده فوق لقمه‌ای میانی یا عکس این حالت را به ثبت می‌رسانند، مفید هستند و اندازه بزرگ مقادیر بردار مقیاس‌بندی شده نیز نشان‌دهنده این امر است. زمانی که زائده فوق‌لقمه‌ای بیشتر حالت سربالا دارد مقدار مثبت اندازه ۲ و مقدار منفی اندازه ۳ با یکدیگر ترکیب می‌شوند تا امتیاز تشخیص کل بیشتری به‌وجود آورند. یعنی اینکه وقتی اندازه ۲ نسبت به اندازه ۳ بالا باشد (ن.ک. شکل ۱)، تابع مقدار مثبت بزرگتری به امتیاز کل اضافه و مقدار منفی کوچکتری از آن کم می‌کند، و زمانی که عکس این موضوع برقرار باشد، و شانه استخوان لقمه‌ای دارای شیب بیشتری باشد، روی هم مقدار منفی بزرگتری در امتیاز کل قرار می‌گیرد. زمانی که زائده فوق لقمه‌ای کناری نیز بلند باشد (اندازه شماره ۷) موجب افزایش بیشتری می‌شود. درحالی که پهنای زائده فوق لقمه‌ای میانی، (اندازه ۶) موقعی که نسبتاً باریک باشد، یک مقدار مثبت اضافه می‌کند (یا بهتر است بگوییم کمترین مقدار را از مقدار کل کم می‌کند).

جدول ۱ نشان می‌دهد که اختلافهای فوق (هرچند اندک) درواقع تفاوتهای مشخصه انسان و شامپانزه در میانگینها هستند که تمام آنها به تولید مقدار امتیاز بیشتر برای شامپانزه گرایش دارند. همچنین ملاحظه می‌کنیم که هیچ اختلاف مطلقی و مطمئناً هیچگونه اختلاف معنی‌داری در اندازه ۶، مربوط به پهنای زائده فوق‌لقمه‌ای میانی وجود ندارد. با وجود این، چنین اندازه‌ای به منظور تشخیص تفاوت اهمیت دارد، زیرا در شامپانزه‌ها به نسبت باریکتر است، درحالی که اندازه‌های دیگر آنها (در این نمونه‌ها) به‌طور متوسط بزرگتر از مقادیر مربوط به انسان هستند.

همچنین پی می‌بریم که چون امتیاز تشخیص تحت تأثیر تمام اندازه‌ها قرار می‌گیرد، تغییرات شکل را در جهت یا مخالف یک الگوی اولیه درنظر می‌گیرد: اگر استخوان شامپانزه فاقد هرگونه پهنا در زائده فوق لقمه‌ای میانی باشد، این حالت ممکن است ترکیب دیگری از سطح باریک

اپی‌کوندیل یا زائده فوق لقمه‌ای جانبی بلند را نشان دهد، که در هرصورت امتیاز آن به شامپانزه تعلق می‌گیرد.

وقتی که امتیازهای تشخیص محاسبه شدند، آنها تفاوت بسیار زیادتری را میان استخوانهای انسان و شامپانزه نشان دادند تا هریک از اندازه‌ها به‌صورت جداگانه. مقادیر متوسط امتیاز و حدود مقادیر در افراد برای هرگروه به‌صورت زیر است:

	میانگین	دامنه
شامپانزه	۹۹/۷۷	۶۷ - ۱۳۰
انسان	۶۱/۴۲	۴۰ - ۸۴

بجز دو مقدار تمامی مقادیر برای شامپانزه بین ۸۰ تا ۱۲۰ قرار می‌گیرند و برای انسان تمامی مقادیر به‌استثنای یکی از آنها بین ۵۰ و ۷۵ قرار دارند، بازه‌های نامتداخل هستند. بنابراین تفکیک آنها از یکدیگر بسیارخوب بود: از ۸۰ نمونه فقط در ۳ مورد تداخل در اندازه‌ها وجود داشت که به میانگین غلط نزدیکتر قرار می‌گرفتند تا میانگین مربوط به خودشان. بدون شک، این یک شیوه موفقیت‌آمیز برای تشخیص بین استخوان انسان و شامپانزه به‌وسیله اندازه‌گیری است، که احتمال صحیح بودن آن نسبت به تشخیص از طریق چشم بسیار زیادتر است.

اکنون به امتیازهای قطعه‌های استخوان کاناپوی و پارانتروپس نگاه می‌کنیم. آنها، به‌ترتیب، برابر ۵۹/۴ و ۶۳/۹ بودند که خیلی نزدیک به مقدارمتوسط در انسان است(بسیار خوبتر از آنکه بتواند واقعیت داشته باشد، نزدیکتر از اکثر مقادیر شناخته‌شده برای افراد بشر) و البته کاملاً خارج از مقادیر مربوط به ۴۰ شامپانزه.

حال بااستفاده از نظریه آمار چنین محاسبه می‌کنیم که اگر واقعاً هرکدام از استخوانها به یک شامپانزه متعلق می‌بود، احتمال اینکه یک نمره تشخیص به این کوچکی (یا کوچکتر) از آنچه قبلاً ذکر شد، داشته باشد فقط یک در ۵۰۰ است. با چنین احتمال اندکی، ما نتیجه می‌گیریم که دو استخوان متعلق به شامپانزه‌ها نیستند، بازمانده‌ها از گروه جانوران انسان‌نما هستند، و این جواب سئوالی است که مطرح کردیم.

البته ما باید محتاط باشیم. سئوال واقعی (به دلیل آنچه که مورد استفاده قرار گرفت) بدینصورت بود: اگر از این قطعه‌ها بخواهیم که انتخابی بین استخوانهای انسان یا شامپانزه‌های

عصر حاضر صورت دهند، خود را چگونه طبقه‌بندی می‌کنند؟ اینها تنها گزینه‌هایی بود که ما به این موجودات فسیلی عرضه کردیم. موجوداتی که در زمان حیاتشان انسانهای امروزی وجود نداشتند، و شامپانزه‌های اجدادی ممکن است به‌طور معنی‌دار متفاوت از شامپانزه‌های امروزی بوده باشند. باوجود این، ما دلایل خوبی داریم برای اینکه از شکل آنها استنباط کنیم که استخوانهای بازو، به‌طور کلی، مثل قطعه همانند خود در انسانها به‌کار گرفته می‌شدند و نه مانند استخوانهای بازوی میمونهای انسان‌نمای آفریقایی. گرچه این میمونها تا حد زیادی موجوداتی زمینی هستند. با این روش می‌توان به قطعه فسیل کاناپوی نیز پرداخت و همچنین به علاوه وضعیت قطعه کرم‌درای راکه تاکنون مبهم بوده‌است نیز مشخص نمود.

جمع‌بندی

چون از بازرسی بصری نتوانستیم ثابت کنیم که فسیل کاناپوی متعلق به حیوانی مثل شامپانزه است، به اندازه‌گیری و استفاده از یک شیوه آماری روی آوردیم که با کمک یک کامپیوتر بتوان آن را به‌کار گرفت. بر اساس تجربه می‌دانستیم که چطور مسئله را بیان کنیم و اطلاعات زیادی را استخراج نماییم، اما از آمارشناسان آموختیم چگونه اطلاعات را منظم نموده، تحلیل کرده و در مورد آنها نظر دهیم. تابع تشخیص، گرچه نسبتاً پیچیده است اما برای بیان واقعیتهای زیست‌شناختی در مورد تنوع مربوط به یک موجود خاص یا تفاوتهای گروهی از لحاظ شکل کاملاً مناسب است واز شواهد معلوم جوابی را در قالب یک احتمال عددی به ما ارائه می‌کند. بنابراین، تا اواسط ۱۹۶۶، پروفسور پاترسون و من نتیجه گرفته بودیم که می‌توانیم احتمال اینکه او استخوان یک میمون انسان‌نما را یافته مردود بدانیم و براساس اطلاعات ما درباره شرق آفریقا، تنها صاحب دیگر آن فسیل از موجودات اولیه گروه جانوران انسان‌نما یعنی اوسترالوپیتسینها بوده‌است.

درحال حاضر در سال ۱۹۸۸[۱] اطلاعات خیلی بیشتری درباره اوسترالوپیتسینها و دیرین‌شناسی شرق آفریقا موجود می‌باشد. و تکوین کاناپوی به بیش از ۴ میلیون سال گذشته

۱ـ زمان نگارش مقاله ـ مترجم.

برمی‌گردد، نه به $2\frac{1}{2}$ میلیون سال.

در مأموریت بعدی، پروفسور پاترسون قسمتی ازیک فک پایین را در لوتاگام‌هیل[1] غرب دریاچه ترکانا، پیدا نمودکه قدمت آن به پنج میلیون سال پیش می‌رسید. این قسمت، همانند دو قطعه کوچکتری که در جاهای دیگر همان منطقه به‌دست آمدند، به احتمال زیاد متعلق به جانوری انسان‌نما تشخیص داده شدند. اکنون اتفاق نظر بر این است که منشاء این موجودات مربوط به ۵ تا ۸ میلیون سال پیش باشد.

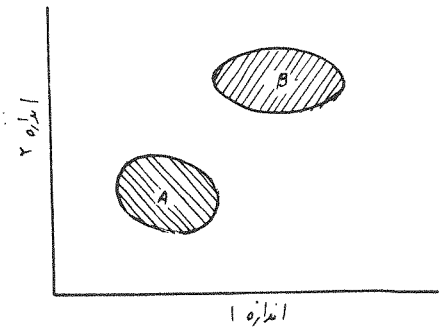
همچنین مقدار قابل توجهی مواد استخوانی در جاهای دیگری در شرق آفریقا پیدا شد که قدمت آن بیش از ۳ میلیون سال می‌باشد، بخصوص در نمونه‌ای که در حبشه پیدا شد و با نام "لوسی" شهرت فراوان یافت. اگرچه فسیل کاناپوی هنوز قدیمیترین قطعه از یک دست است که تاکنون پیدا شده، اما همانگونه که با تحلیلهای پیچیده‌تر بعدی تأیید شد، در مقایسه با قطعه لوسی و قطعات همانند آن حتماً شکلی امروزی دارد. به‌هرحال نتیجه‌گیری اصلی ما، یعنی اینکه قطعه متعلق به یک اوسترالوپیتسین است تأیید شده و اکنون کمتر باعث تعجب است. و شاید بی‌مناسبت نباشد که به این نکته توجه کنیم که تحلیل آماری سال ۱۹۶۷ ما در تأییداهمیت فسیل کاناپوی قبل از اینکه شواهد و مدارک بعدی دردسترس قرار بگیرد قطعاً سودمندبود.

مسائل

۱ـ درمثال فرضی شکل ۱، مقادیر اندازه‌های ۲ را برای گروه‌های A و B روی محور عمودی رسم کنید و فاصله‌ای را که دو جامعه از لحاظ اندازه ۲ روی هم قرار می‌گیرند، مشخص کنید. بااستفاده از شکل ۱ توضیح دهید که چرا برای اینکه بتوانیم فردی رابه‌عنوان A یا B رده‌بندی کنیم به توجه کردن به هر دو اندازه یا تابعی از این اندازه‌ها نیازمند می‌باشیم.

۲ـ فرض کنید به جای دو جامعه در شکل ۱ جامعه‌ها در شکل زیر باشند. در اینصورت خیلی ساده است که دو جامعه را از هم متمایز کنیم. چطور این کار را انجام می‌دهید؟ با استفاده از این مثال و مثال شکل ۱ توضیح دهید که تابعهای تشخیصی در چه مواقعی مورد نیاز می‌باشند؟

۱ـ *Lothagam Hill*



۳ـ درجدول ۱ تقریباً اختلافی در اندازه‌ها برای انسانها و شامپانزه‌ها وجود ندارد، لیکن این یکی از اندازه‌هایی بود که به تفکیک نمودن دو جامعه کمک نمود. توضیح دهید که اندازه شماره ۶ چطور در تفکیک کردن دو جامعه ممکن است مفید واقع شود؟

۴ـ مثال مشروحی را خودتان مطرح کنید که تحلیلی براساس تابع تشخیص ممکن است برای آن مناسب باشد.

بخش دوم

دنیای سیاسی ما

برهان آماری تبعیض در استخدام ۱۱۹

آمار در انتخاب هیئت منصفه ۱۳۳

اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی ۱۴۳

شب انتخابات در تلویزیون ۱۵۹



برهان آماری تبعیض در استخدام

سندی ال. زابل[۱] دانشگاه نورث وسترن

عنوان هفتم از قانون حقوق مدنی مصوب سال ۱۹۶۴، هرنوع تبعیض دراستخدام راکه ریشه در مبانی نژاد، رنگ، مذهب، جنسیت و یا ملیت داشته باشد، ممنوع ساخته است.تبعیض، از نوعی که توسط قانون حقوق مدنی ممنوع شده، در حال حاضر تاحدودی به مناسبت تصویب قانون مزبور و چه به‌مناسبت تحولی که در جو اجتماعی پدیدآمده، معمولاً مثل سابق آشکار نبوده و

۱ـ *Sandy L. Zabell*

اثبات وجود آن نیز به همین مقیاس نسبت به دوران پیش از تصویب این قانون مشکل شده‌است. طی دهه گذشته دادگاهها به‌طور روزافزون از روشهای آماری برای قضاوت پیرامون وجود و دامنه گسترش تبعیض یاری جسته‌اند.

دیوان عالی فدرال آمریکا در گزینش قضایی روشهای آماری برای اثبات تبعیض در استخدام نقشی کلیدی ایفا کرده‌است. هرچند برهان آماری به‌طور پراکنده در دادگاههای بدوی تقریباً بلافاصله پس از تصویب قانون حقوق مدنی شروع شده‌بود، بحث در مورد برهان آماری در برخی تصمیم‌گیریهای کلیدی دیوان عالی طی دو دهه گذشته نقش مهمی در تمایل دادگاههای بدوی برای قبول این برهان و اتکای روزافزون به آن ایفانمود.

رفتار نابرابر و آثار نابرابر

موارد تبعیض در استخدام در دو مقوله کلی دسته‌بندی می‌شوند: رفتار نابرابر و آثار نابرابر. در مورد رفتار نابرابر موضوع این است که آیا یک کارفرما به‌طور عمد علیه کسی تبعیض قائل شده است؟ در چنین مواردی آمار را می‌توان به‌عنوان مدرک اتفاقی برای تبیین یک «الگو و عملکرد» تبعیض (مشابه آنچه در مورد مسئله *هزل وود*[۱] انجام شد و شرح آن در زیر آمده‌است) به‌کار گرفت. با وجود این، در مورد شرکت برق گریگ‌وی‌دوک[۲] حکم دیوان عالی بر این قرار گرفت که کاربستهای استخدامی، که از "لحاظ صورت ظاهر بیطرف" نامیده می‌شود، ازقبیل آزمونهای استخدام و ارتقا چنانچه اثر نامطلوب یا نابرابر بر طبقه تحت حفاظت داشته باشد) از نظر جزء هفتم قانون، تبعیض‌آمیز تلقی می‌شود. حتی زمانی که یک کارفرما قصد تبعیض نداشته‌باشد چنانچه کاربستهای استخدامی منجر به استخدام یا ارتقای افراد طبقات تحت حفاظت در سطحی بسیار پایین‌تر شود، می‌توان آن را به‌عنوان مورد محکمه پسندی ارائه نمود که کارفرما به نقض قانون مبادرت کرده‌است. دیوان عالی با جدا نمودن نتیجه از هدف مقوله‌ای اساساً آماری در مورد تبعیض پدید آورده و به‌این ترتیب استفاده از برهان آماری را اجتناب‌ناپذیر نمود.

اثبات اینکه یک عملکرد ظاهراً بیطرفانه در استخدام، آثار نابرابر داشته‌است، برای اثبات

۱ـ *Hazelwood*

۲ـ *Griggs V.Duke Power Company*

تبعیض ممنوع، الزاماً کافی نیست. کارفرما می‌تواند با نشان‌دادن اینکه عملکرد او هدفی مجاز داشته‌است، تحت عنوان دفاع از «ضرورتهای کاری» از آن دفاع کند. برای مثال چنانچه بتوان نشان داد که یک آزمون استخدام، آزمون برای تعیین مهارت شغلی مجاز بوده‌است، از آن می‌توان به‌عنوان دلیلی برای دفاع استفاده کرد. تشخیص اعتبار یک آزمون استخدام (که ثابت کند آزمون برای شناختن مهارت خاصی انجام شده) نیز زمینه‌ای است که می‌توان آمار را در آن به‌کار گرفت، اما در عمل غالب اوقات برای یک کارفرما مشکل است چنین اعتباری را در زمینه‌ای تخاصم‌آمیز اثبات نماید و به‌این ترتیب اغلب اوقات نشان دادن نابرابری اساسی آماری برای تصمیم‌گیری پیرامون یک مورد کفایت می‌کند.

ماجرای هزل‌وود

هرچند دیوان عالی در موقعیتهای متعدد پیشین، آمار را به‌عنوان ابزار ارزشمندی برای تحقیق در موارد تبعیض در استخدام مورد توجه قرار داده‌بود، ولی استفاده از آزمونهای آماری معنی‌داربودن، که در پرونده بخش آموزشگاه هزل‌وود در برابر ایالات متحده انجام شد بر روشهای رسمی برهان آماری مهر تاییدی زد که این روشها قبلاً از آن برخوردار نبودند. پس از هزل وود، دادگاههای بدوی در استفاده از روشهای آماری برای تشخیص موارد تبعیض در استخدام آزادی عمل احساس کرده‌اند.

در ماجرای هزل‌وود دولت آمریکا دادخواستی علیه بخش آموزشگاه هزل‌وود اقامه نمود که در آن اظهار شده‌بود؛ آموزشگاه با به‌کارگرفتن یک «الگو یا عملکرد» تبعیض‌آمیز استخدام، عنوان هفتم قانون را نقض کرده است. هزل‌وود آموزشگاه بخشی عمدتاً روستایی در ناحیه شمالی استان سنت‌لوییز است که در آن تعدادی دانشجوی سیاهپوست به تحصیل و تعداد کمی معلم سیاهپوست در میان کارکنان آن (۲٪) به خدمت اشتغال دارند. دولت استدلال آماری خود را مبتنی براین واقعیت استوار ساخت که درصد مدرسان سیاهپوست در این بخش به‌طور اساسی کمتر از نواحی پیرامونی استان سنت‌لوییز و شهر سنت‌لوییز می‌باشد.

دادگاه بخش (که در نخستین مرحله به این مورد پرداخت)، درصد مدرسان سیاهپوست در بخش را با درصد دانش‌آموزان سیاهپوست مقایسه نمود و متذکر شد که این دو تقریباً مشابه

هستند، و حکم رابه نفع بخش آموزشگاه هزل‌وود صادر کرد. هنگامی که دعوی مورد تجدیدنظر قرار گرفت، دادگاه فدرال استیناف حکم کرد که مقایسه مدرسان با دانش‌آموزان نامربوط بوده و درست آن است که مقایسه بین درصد مدرسان سیاهپوست در آموزشگاه بخش با مدرسان سیاهپوست در بازار کار متناسبی صورت پذیرد (تقریباً همان استدلالی که دولت کرده بود) و تصمیم دادگاه بدوی را لغو نمود. پس از آن، دعوی در دیوان عالی مورد تجدیدنظر قرار گرفت.

دیوان عالی در تصمیم خود متذکر شد که از ۴۰۵ مدرس در استخدام آموزشگاه بخش هزل‌وود طی دو سال تحصیلی ۱۹۷۳-۱۹۷۲ و ۱۹۷۴-۱۹۷۳، ۱۵ نفر (یا ۳/۷٪) سیاهپوست بوده‌اند، درحالی که نسبت مدرسان سیاهپوست در استان سنت‌لوییز ۵/۷٪ بوده است. یکی از سئوالاتی که در برابر دادگاه قدعلم نمود، این بود که آیا این ۲٪، اساساً متفاوت هستند؟

برای شروع به تحلیل چنین سوالی یک راه این است که احتمال وقوع تفاوتی با این حجم یا بزرگتر از آن محاسبه شود. اگر تعداد ۴۰۵ مدرس در استخدام آموزشگاه را به‌عنوان نمونه تصادفی از مجموعه‌ای از مدرسان واجد شرایط که ۵/۷٪ آنها سیاهپوست هستند، درنظر بگیریم، در این حالت احتمال انتخاب ۱۵ مدرس سیاهپوست یا کمتر از آن به میزان ۴/۶٪ محاسبه می‌شود. این احتمال را می‌توان مستقیماً با استفاده از توزیع احتمال دوجمله‌ای محاسبه نمود (ن. ک. *موستلر، رورک، توماس* ـ ۱۹۷۰)

محاسبات مستقیم از این دست، معمولاً کسالت‌بار هستند و در عمل معنی‌داربودن چنین تفاوتهایی اغلب با روش تقریبی مورد قضاوت قرار می‌گیرد که مستلزم محاسبه یک کمیت اساسی آماری به نام انحراف معیار است. با این کمیت آماری میزان تغییرپذیری داده‌ها سنجیده می‌شود. سپس باید دید انحراف معیار میان تعداد مشاهده‌شده افراد (در اینجا ۱۵) و تعداد مورد انتظار (در اینجا ۲۳=۰/۰۵۷×۴۰۵) چقدر است؟ یک قاعده کلی مقرر می‌دارد که اگر افراد براساس تصادف انتخاب شده‌باشند، ۹۵٪ اوقات تعداد مشاهده شده میان ۲ انحراف معیار از مقدار مورد انتظار (۲/۵٪ اوقات بیش از ۲ انحراف معیار در بالا و ۲/۵٪ اوقات بیش از ۲ انحراف معیار در پایین آن) قرار می‌گیرد. به‌عکس، چنانچه تفاوت بیش از ۲ انحراف معیار باشد، در این صورت واقعه‌ای نسبتاً نادر مشاهده شده‌است و این موجب می‌شود که فرض

تصادفی بودن انتخاب افراد زیر سئوال برود. گفته می‌شود که تفاوت در این حالت از نظر آماری معنی‌دار و در سطح ۵٪ می‌باشد.

فرمول انحراف معیار در این زمینه عبارت است از:

$$\sqrt{np(1-p)}$$

که درآن n تعداد افراد و p نسبت در جامعه آماری است. برای مثال در هزل‌وود $n=۴۰۵$ و $p=۰/۰۵۷$ بوده‌است. به این ترتیب انحراف معیار برابر است با: $\sqrt{(۴۰۵)(۰/۰۵۷)(۰/۹۴۳)}$ یا ۴/۶۷. در یک پانوشت دیوان عالی متذکر شده‌است که تفاوت بین تعداد سیاهپوستان استخدام شده و تعداد مورد انتظار (۸=۱۵-۲۳) کوچکتر از ۲ انحراف معیار بوده و این بیانگر آن است که تفاوت از نظر آماری معنی‌دار نیست.

یکی از سئوالات مهمی که در ماجرای هزل‌وود، مطرح شد تعیین جامعه آماری مناسب برای مقایسه با بازار کار مربوط بوده چنانچه شهر سنت‌لوییز در مجموعه ایالت سنت‌لوییز به حساب آورده شود (به شکلی که دولت استدلال کرد)، درصد مدرسان سیاهپوست از ۵/۷٪ به ۱۵/۵٪ افزایش می‌یابد (بخشی از این امر به این علت است که در نظام آموزشی شهر سنت‌لوییز سعی شده‌است به‌طور تقریب ۵۰٪ از کارکنان، سیاهپوست باشند). واضح است که در این حالت، تفاوت از نظر آماری به سطح معنی‌دار می‌رسد. به علت آنکه تفاوت میان ۳/۷ و درصد کمتر بازار کار با ۵/۷٪، کمتر از ۲ انحراف معیار می‌باشد، دادگاه نتیجه گرفت که در این حالت «تفاوت به‌قدر کافی کوچک است و برهان دیگر دولت را تضعیف می‌کند، درحالی که تفاوت میان ۳/۷٪ و ۱۵/۵٪ به‌قدر کافی بزرگ بوده و موجب تأیید آن می‌شود». از آنجا که این موضوع هیچگاه مورد توجه دادگاههای بدوی واقع نشده‌بود، تصمیم دادگاه استیناف لغو، و دعوی برای بررسی مجدد به دادگاه بخش اعاده شد.

به این ترتیب ارزش مدرک آماری به‌عنوان برهان درمورد تبعیض در استخدام می‌تواند به نحوی بسیارحساس، بستگی پیدا کند به اینکه کدام بازار کار «مناسب» به‌عنوان معیار مقایسه انتخاب شود. در نمونه‌گیری علمی، که یک نمونه تصادفی از یک جامعه بدون ابهام و از قبل تعیین شده، انتخاب می‌شود، باید این معیار مقایسه الزاماً قبل از انتخاب نمونه به‌وضوح مشخص شود. اما در مواردی مانند هزل‌وود آنچه که با آن شروع می‌شود، نمونه است. یعنی

نسبت استخدام‌شدگان برای یک طبقه مورد حمایت و جامعه یا بازار کار مناسب، ساختاری فرضی است که تصور می‌شود کارکنان استخدام شده از آن انتخاب شده‌اند. به‌طوری که متن حکم هزل‌وود روشن می‌کند، انتخاب بازار کار مناسب، که به‌عنوان معیاری برای مقایسه مورد استفاده قرارگرفت، می‌تواند سئوال پیچیده‌ای باشد که الزاماً پاسخ منحصر به‌فرد ساده یا بدون ابهامی را نمی‌پذیرد.

هنگامی که تفاوتهای مشاهده‌شده درمورد یک استخدام تبعیض‌آمیز از آنچه که به‌راحتی در نمونه‌گیری تصادفی توجیه می‌شود تجاوزننماید، نتیجه آن مدرکی علیه وجود تبعیض در استخدام است. ازسوی دیگر چنانچه تفاوت از نظر آماری به‌عنوان معنی‌دار قلمداد شده باشد، نتیجه را باید با احتیاط تفسیر نمود. از یک جهت، برای آنکه جریان استخدام، عادلانه یابدون تعصب باشد، ضرورتی ندارد که از لحاظ آماری، تصادفی باشد. جریان عادلانه استخدام می‌تواند با روشهای ویژه نمونه‌گیری که به‌طور متوسط مقدار صحیح مورد انتظار را به‌دست می‌دهد، انطباق داشته ولی دارای انحراف معیارهای بزرگتر از انحراف معیارهای مربوط به نمونه‌گیری ساده تصادفی باشد. (ن. ک. مه‌یر، *ساکس و زابل* ۱۹۸۶)

از جهت دیگر «معنی‌دار بودن آماری» صرفاً بیانگر وجود برخی تفاوتها می‌باشد، ولی به معنای آن نیست که این تفاوت الزاماً از نظر عملی یا قانونی دارای اهمیت بوده و معنی‌دار است. تفاوتی در سطح یک یادو درصد در هریک از دوجهت برای مثال در میزان قبولی مردان و زنان در یک آزمون مهارتهای فیزیکی به‌ندرت قابل ملاحظه است. به‌طوری که دادگاه استیناف در پرونده **آمریکا در برابر آزمون** (به شماره 550 F. 2d 577, 10th Cir. 1976) متذکر شده‌است «این نتیجه‌گیری ریاضی که تفاوت میان این دو رقم از نظر آماری معنی‌دار است، مستلزم آن نیست که از قبل به این باور برسیم که این انحرافات از نظر قانونی معنی‌دار هستند».

مقایسه دونمونه‌ای داده‌های متقاضیان

از آنجاکه تعیین بازار کار مناسب می‌تواند متضمن مسائل ظریف قضایی شود، دادگاهها هرگاه امکان‌پذیر بوده‌است، ارجحیت بارزی را برای مقایسه بین داده‌های برخی متقاضیان با داده‌های سایر متقاضیان قائل شده‌اند (برای مثال، چه درصدی از متقاضیان زن در برابر چند درصد

متقاضیان مرد، استخدام شده‌اند؟). به این ترتیب از ضرورت تصمیم‌گیری درباره آنکه جامعه مناسب برای مقایسه کدام است احتراز، و سئوال بسیار مناسبتری طرح شده‌است: باتوجه به ادغام متقاضیان واقعی، طبقه مورد حمایت چگونه عمل کرده‌است؟ در بخش قبلی، ما به مقایسه درصد یک گروه با مقدار مربوط به یک جامعه کاملاً مشخص پرداختیم و به‌این ترتیب با مسئله‌ای یک نمونه‌ای سروکار داشتیم. در اینجا به مقایسه درصد بانوان و مردان متقاضی که استخدام شده‌اند، می‌پردازیم و به این ترتیب به مقایسه نتیجه برای دو نمونه دست می‌زنیم و لازم است تغییرات هردو منبع را پذیرا باشیم. در زیر به‌طور مجمل به این موضوع پرداخته شده است.

یک مثال در این زمینه، پرونده *تیل* در برابر *کنتیکات* است که یکی دیگر از دعاوی مطروحه در دیوان عالی است. در این جریان، یک امتحان کتبی که برای تعیین شایستگی ترفیع انجام می‌شد، به این دلیل مورد اعتراض قرار گرفت که آثار نابرابری بر امتحان دهندگان سیاهپوست برجای گذاشته است. از تعداد ۴۸ سیاهپوست شرکت‌کننده در امتحان، ۲۶ نفر (۵۴/۲٪) و از ۲۵۹ نفر سفیدپوست شرکت کننده، ۲۰۶نفر (۷۹/۵٪) قبول شدند. سئوالی که در این حالت مطرح می شود این است که اگر قرارباشد در درازمدت سفیدپوستهاو سیاهپوستها در امتحان با میزان برابر موفق شوند، آیا امکان وقوع تفاوت نمونه در سطحی به این بزرگی یا بزرگتراز آن وجود دارد؟ در مورد *تیل* تفاوت میزان قبولی عبارتست از: ۲۵/۳٪=۵۴/۲٪-۷۹/۵٪ که عبارتست از ۳/۷۶ انحراف معیار که به‌مراتب زیادتر از «۲یا۳ انحراف معیار»است، که دیوان عالی با استناد به آن در مورد هزل‌وود تصمیم گرفت.

در بسیاری موارد که مشاغل کوچک یا دورانهای کوتاه زمانی مطرح می‌شوند، نمونه‌های به‌کارگرفته شده کمتر از آن است که استفاده از تقریب راکه در پس نشانه انحراف معیار پنهان است، توجیه کند. در این حالت می‌توان به آزمون دقیق فیشر[۱] متوسل شد (ن. ک. موستلر، رورک، *توماس* ـ ۱۹۷۰) درآزمون دقیق فیشر به افرادی که استخدام شده یا ترفیع یافته‌اند، به‌عنوان یک نمونه تصادفی از مجموعه متقاضیان نگریسته و احتمال آنکه یک نابرابری در

۱ـ *Fisher*

میزان استخدام یا ترفیع، حداقل در سطح مشاهده‌شده اتفاق افتد، محاسبه می‌شود.

درمواردی که تعداد کم است، تحلیل آماری می‌تواند به‌خصوص مفید واقع شود، زیرا در بسیاری از مواقع درک مستقیم بسیار گمراه‌کننده است. غالباً تصور می‌شود که با نمونه‌های کم نمی‌توان اطلاعات مفیدی به‌دست آورد ولی همه اینها به میزان تفاوتهایی بستگی دارد که پیش می‌آید. برای مثال در پرونده دندی، در برابر مرکز بیمارستانی واشنگتن جدول زیر به‌دست آمد:

	انتخاب شده	ردشده	نرخ قبولی
سیاه پوستها	۴	۵	٪۴۴
سفیدپوستها	۲۶	۰	٪۱۰۰

تفاوت در نرخ قبولی دوگروه به‌طور واضح قابل توجه‌است، ولی از آنجا که فقط ۹ سیاهپوست در کل نمونه وجود دارد، با درک مستقیم ممکن است، فرض شود که این اختلاف می‌تواند به‌راحتی در اثر تصادف اتفاق افتاده‌باشد. درواقع چنین تفاوتی بسیار غیرمحتمل است: در مثال، «دندی» احتمال معنی‌دار بودن منتج از آزمون دقیق فیشر عبارتست از: ۰/۰۰۰۴ (یعنی هرگاه تیمار هر دوگروه یکسان باشد، در هر ۱۰۰۰۰ مورد فقط ۴ شانس برای پیش آمدن تفاوتی به این بزرگی یا بزرگتر از آن وجود دارد). (سرگذشت دندی جالب است: در سطح بخش، دادگاه که ریاست آن را قاضی مشهور پرونده واترگیت *جان سیریکا*[۱] به‌عهده داشت، استدلال آماری را به دلیل آنکه نمونه‌های به‌کار برده‌شده، بسیارکم بوده‌اند رد کرد، ولی بعد هنگامی که دعوی توسط دادگاه استیناف مورد تجدیدنظر قرار گرفت، نظر او کنار گذاشته شد).

هرچند لحاظ کردن داده‌های متقاضیان، مشکل تعیین بازار کار مناسب را رفع می‌کند ولی می‌تواند مسائل خاص خود را پدید آورد. به شکلی تناقض‌آمیز، برنامه‌های عملیاتی مثبت با هدف تشویق گروههای اقلیت به ثبت‌نام برای پستهای خالی می‌تواند به مجموعه‌ای از متقاضیان با سطوح متفاوت شایستگی بینجامد. برای مثال در پرونده واشنگتن در برابر دیویس که یکی

۱ـ *John Sirica*

دیگر از دعاوی دیوان عالی است، برای آزمونی که توسط اداره پلیس ناحیه کلمبیا مورد استفاده واقع شد، چنین رأیی صادر شد که عمل تبعیض‌آمیزی انجام نیافته‌است و این علی‌رغم مدارک روشنی که بر تأثیرات نابرابر گواهی می‌داد، بخشی به این دلیل بود که مجموعه، دارای سطوح شایستگی متفاوت بود.

پارادوکس سیمپسون و جمع تجمعی

شرکتها و مؤسسات عظیم تقریباً همیشه به بخشها، زیربخشها، ادارات و غیره تقسیم می‌شوند. به‌استثنای پستهای سطح بالای مدیریت، تصمیم‌گیری در مورد استخدام به‌طور معمول در سطح ادارات یا بخشها انجام می‌شود. تحلیل داده‌های تجمعی استخدامها در چنین شرکتهایی می‌تواند پدیده جالبی را باعث شود که به‌عنوان پارادوکس سیمپسون (و یا گاهی به‌نام همبستگی دروغین) خوانده می‌شود.

مثالی آموزنده و جالب از پارادوکس سیمپسون در دانشگاه برکلی کالیفرنیا در سالهای دهه ۱۹۷۰ اتفاق افتاده‌است. بررسی داده‌های متقاضیان طی یک فصل در سال ۱۹۷۳ نشان داد که میزان کلی پذیرش متقاضیان زن در این دانشگاه به‌طور اساسی کوچکتر از این میزان برای متقاضیان مرد بوده‌است (ن. ک. جدول ۱).

کدام دانشکده‌ها در دانشگاه برکلی مسئول این عدم تعادل بوده‌اند؟ به‌طور تعجب برانگیزی بررسی داده‌های پذیرش در هریک از دانشکده‌ها نشان می‌داد که میزانهای پذیرش مردها و زنها نظیر یکدیگر بوده‌اند. درواقع حتی در برخی دانشکده‌ها میزان پذیرش زنها به‌طور معنی‌داری بیشتر از مردها بود.

چنین موقعیت تناقض‌آمیزی درواقع یک توضیح بسیار ساده دارد. در برکلی تقاضای زنها بیشتر برای رشته‌هایی مانند زبان انگلیسی و تاریخ بوده‌است که تعداد کثیری متقاضی و منطبق با آن میزانهای پذیرش کمی داشته‌است، درحالی که مردها بیشتر برای رشته‌هایی مانند ریاضیات و فیزیک تقاضاداده‌اند که تعداد قلیلی متقاضی و میزان بسیار بالاتری از پذیرش دارد. یک مثال ساده فرضی پدیده مزبور را به نمایش می‌گذارد:

	ریاضیات			انگلیسی			مختلط		
	پذیرفته	ردشده	درصد	پذیرفته	ردشده	درصد	پذیرفته	ردشده	درصد
مرد	۹۰	۱۰	۹۰	۱	۹	۱۰	۹۱	۱۹	۸۳
زن	۹	۱	۹۰	۱۰	۹۰	۱۰	۱۹	۹۱	۱۷

در مثال فوق میزانهای پذیرش مرد و زن در هریک از رشته‌ها برابر است با ۹۰٪ در ریاضیات، ۱۰٪ در انگلیسی. علی‌رغم این امر، چون زنها بیشتر متقاضی انگلیسی و مردها بیشتر متقاضی ریاضیات بوده‌اند و نیز به‌این علت که میزان کل پذیرش برای دو رشته متفاوت است، میزان پذیرش در جمع تجمعی برای مردها و زنها به‌طور اساسی متفاوت شده‌است.

به دلیل وجود پارادوکس سیمپسون، مناسب است داده‌های استخدام در موارد تبعیض‌آمیز در همان سطحی تحلیل شود که عملاً اتخاذ تصمیم صورت می‌گیرد. صرف‌نظر از سهولت خوشایند آن، چنانچه فقط داده‌های تجمعی مورد ملاحظه واقع شود، بسیار گمراه‌کننده خواهد بود: عملکرد استخدامی سالم می‌تواند ناسالم و عملکرد ناسالم در استخدام می‌تواند به صورتی غیرواقعی نمایانده شود.

بحث

استفاده از آمار در پرونده‌های مربوط به تبعیض در استخدام بیان‌کننده طیف حیرت‌انگیزی از مسائل در زمینه کاربرد و تفسیر است که در اینجا امکان توضیح فقط برخی از آنها مقدور بود. جامعه مناسب برای مقایسه، مناسب بودن آزمون معنی‌دار بودن آماری مبتنی بر نمونه‌گیری

جدول ۱ـ متقاضیان ورود به دانشگاه برکلی کالیفرنیا ـ پاییز ۱۹۷۳

جنسیت	تعدادمتقاضیان			
	پذیرفته	ردشده	جمع	درصد پذیرفته شده
مرد	۳۷۳۸	۴۷۰۴	۸۴۴۲	٪۴۴
زن	۱۴۹۴	۲۸۲۷	۴۳۲۱	٪۳۵

تصادفی، تأثیرات جمع تجمعی و کنترل تفسیر مربوط به متغیرهای توصیفی مانند مهارت، تنها فهرست تعداد کمی از آنها را تشکیل می‌دهد. عدم توجه به چنین مسائلی غالب اوقات می‌تواند به تصویر بسیار گمراه‌کننده‌ای همچون مورد پارادوکس سیمپسون منجر شود. با همه اینها، استفاده دقیق و تفسیر مناسب از روشهای آماری، ابزار مؤثر و مفیدی برای انجام وظیفه در اتخاذ تصمیم در مواردی که تبعیضات در استخدام، ترفیع و اخراج اتفاق می‌افتد در اختیار دادگاه قرار می‌دهد.

مسائل

۱ـ در موردی که اخیراً برای اداره پلیس حومه یک شهر پیش آمد، نتیجه یک امتحان انجام کار فیزیکی به این دلیل مورد اعتراض واقع شد که درآن نسبت به بانوان تبعیض شده‌بود. درحالی که تقریباً ۸۰٪ از ۱۵۰ متقاضی مرد طی یک دوره شش ساله در این آزمایش موفق شدند، فقط ۳ نفر از ۱۰ نفر زن شرکت‌کننده توانستند امتحان را با موفقیت بگذرانند. اگر این تفاوت واقعی باشد، چگونه می‌توان استفاده از این امتحان را توجیه کرد؟

۲ـ در مورد پذیرشهای دانشگاه برکلی، آیا می‌توان با استفاده از تفاوت در میزان پذیرش در رشته‌های مختلف استدلال کرد که تبعیض وجود داشته است؟ دلایل له وعلیه.

۳ـ در سال ۱۹۶۸ دکتر *بنیامین اسپاک* و تعدادی از دستیاران وی در بوستون به جرم توطئه به منظور نقض قانون نظام وظیفه ۱۹۶۷ محاکمه شدند. با آنکه بیش از ۵۰٪ بزرگسالان ناحیه بوستون را بانوان تشکیل می‌دهند، هیچ زنی در هیئت منصفه دکتر *اسپاک* حضور نداشت. این نقصان به‌تدریج در سه مرحله انجام شد: به سبب عدم صلاحیت قانونی فقط ۳۰٪ داوران قابل قبول از زنها بود (مرحله اول)، در فهرست داوران که ۱۰۰ نفربود ـ و ازبین آنها هیئت منصفه انتخاب شد ـ فقط ۹ نفر زن وجود داشت (مرحله دوم)، و درنتیجه چالشهای حق تقدم، هیئت منصفه نهایی هیچ زنی را در بین خود نداشت. (مرحله سوم)

این موضوع مورد نگرانی شدید وکلای مدافع دکتر*اسپاک* در دادگاه قرار گرفت. زیرا تصور بر آن بود که خانمها نظر مساعدی به دکتر*اسپاک* داشته‌اند. هم به دلیل کتابهای معروف او در پرستاری کودکان و هم به‌دلیل اینکه نظرسنجی‌ها نشان داده‌بود که زنها بیش از مردها مخالف جنگ ویتنام

هستند.

اگر شما یکی از وکلای مدافع دکتر *اسپاک* بودید چگونه می‌توانستید تقلیل تصاعدی وجود زنها در هرمرحله از فرایند انتخاب هیئت منصفه را مورد اعتراض قرار دهید؟

هرگاه در فهرستی از ۱۰۰ نامزد عضویت در هیئت منصفه که از مجموعه‌ای شامل ۳۰٪ زن به‌طور تصادفی انتخاب شده‌است، فقط ۹ نفر یا کمتر زن باشد، تفاوت در واحدهای انحراف معیار بویژه در مرحله دوم چقدر است؟

۴ـ درهر دومورد، تبعیض در استخدام و هیئت منصفه، پیش فرض معمول محاسبات آماری بر این اساس است که افراد مطرح (افرادی که استخدام، یا داورانی که انتخاب می‌شوند) نمونه‌های تصادفی هستند که از بین متقاضیان یا مجموعه بالقوه داوران انتخاب می‌شوند. از چه راههایی به‌غیر از تبعیض می‌توان این فرض را در هر مورد نقض نمود؟ درچه مرحله‌ای از انتخاب داوران فرضیه اعمال تبعیض قابل قبول‌تر به‌نظر می‌رسد؟

۵ـ طی سالهای ۱۹۷۴ تا ۱۹۷۸ میزان مالیات بر درآمد فدرال، برای افراد در هرطبقه درآمدی، کاهش یافته‌است. با این حال میزان متوسط کل مالیاتها که عملاً جمع‌آوری می‌شود از ۱۴/۱٪ به ۱۵/۲٪ افزایش یافته‌است! توضیح دهید که چگونه این تصویر مبدل، مثالی برای پارادوکس سیمپسون است؟

منابع :

D. C. Baldus and. J .L. Cole. 1980 . *Statistical Proof of Discrimination*. Colorado Springs, Colo: Shepard´s. [The definitive reference] .

Peter j. Bickel, Eugene A. Hammel, and J. William´O Connell. 1975 "Sex Bias in Graduate Admissions: Data from Berkeley. "*Scince* 187: 398 - 404. [Reprinted, with an interesting discussion between William Kruskal and Peter Bickel, in *Statistics and public policy,* W. B Fairley and F. Mosteller, eds. 1977. Reading, Mass: Addison _ Wesley. pp. 113-130.]

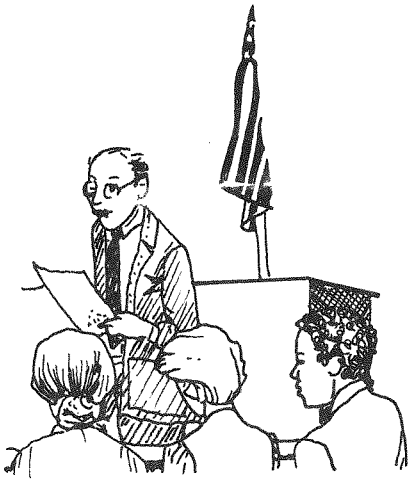
Paul Meier, Jerome Sacks and Sandy L. Zabell. 1986. "What Happened in *Hazelwood*: Employment Discrimination. Statistical Proof, and the 80% Rule: In *Statistics and Law*: M.

DeGroot, S. Fienberg, and J. Kadane, eds. 1986. New York: Wiley, pp1-40. [Discusses some of the issues raised in this article in detail. The entire volume is a valuable source of information about recent legal uses of statistics.]

Frederick Mosteller, Robert B. K. Rourke, and George B. Thomas, Jr. 1970 . *Probability with Statistical Applications,* 2nd ed. Reading, Mass: Addison _ Wesly. [An exceptionally lucid textbook with many examples.]

Clifford H. Wagner. 1982 Simpson´s Paradox in Real Life "*The American Statistician.* 36: 46-48. [Discusses several real_life instances of Simpson´s paradox.]

Hans Zeisel. 1969 "Dr. Spock and the Case of the Vanishing Women Jurors. "*The University of Chicago Law Review* 37: 1-18. [A classic jury discrimination case.]



آمار در انتخاب هیئت منصفه

اس. جیمز پرس[۱] شرکت جوراکن لوس‌آنجلس و دانشگاه کالیفرنیا، ریورسایدر

در هر دادرسی که با حضور هیئت منصفه انجام می‌شود، برای کمک به وکلای مدافع در کنارگذاشتن آن عده از اعضای هیئت منصفه که احتمال می‌رود نظر موافق نداشته‌باشند، می‌توان از روش‌شناسی آماری استفاده کرد. ابتدا به منشاء سیستم داوری با حضور هیئت منصفه می‌پردازیم و در مورد مسائل گوناگون مرتبط با انتخاب تصادفی هیئت منصفه اظهارنظر

۱ـ *S. James Press*

می‌کنیم. سپس چگونگی حق اعتراض بدون نیاز به اقامه دلیل و استفاده از آمار برای هرچه مؤثرتر ساختن این کار برای جلوگیری از حضورافرادی در هیئت منصفه را که احتمال می‌رود مخالف موکل باشند مورد بحث قرار می‌دهیم.

استعمارگران انگلیسی‌الاصل که به قاره آمریکا آمدند، روشهای دادرسی را که با آن آشنایی داشتند ازجمله سیستم داوری با حضور هیئت منصفه را به‌همراه خود به این قاره آوردند. در سال ۱۶۰۶ جیمز *اول* پادشاه انگلستان امتیازی به شرکت ویرجینیا اعطا کرد که شامل حق "دادرسی با حضور هیئت منصفه" بود. این حق بعدها به رودآیلند (۱۶۶۳)، نیویورک (۱۶۶۴)، نیوجرسی (۱۶۷۷) و مریلند(۱۶۹۳) نیز اعطا شد. بالاخره حق محاکمه با حضور هیئت منصفه طی منشور حقوق شهروندان چه در محاکمات کیفری (اصلاحیه ششم قانون اساسی) و چه در دادرسیهای حقوقی (اصلاحیه هفتم قانون اساسی) برای تمامی آمریکاییها تضمین شد. ولی قانون اساسی درمورد "تعداد افرادی" که هیئت منصفه را تشکیل می‌دهند، "نوع افرادی" که باید در هیئت منصفه باشند و "نحوه انتخاب اعضای هیئت منصفه" مبهم بود.

طی سالیان، در جریان قضاوتهای گوناگون، هیئتهای منصفه متشکل از تعداد متفاوتی از اعضا، از شش نفرگرفته تا دوازده نفر، با درجات مختلفی از موفقیت آزمایش شده‌است. (در دادرسیهای فدرال که با حضور هیئت منصفه انجام می‌گیرد، تعداد شش نفر مرسوم شده‌است.) همچنین برای تصمیم‌گیری براساس آرای هیئت منصفه، به‌جای ضرورت اتفاق آرا، حداقل دوسوم یا سه‌چهارم نیز در معرض امتحان قرار گرفته‌است.

درمورد "نوع افراد" مناسب برای عضویت در یک هیئت منصفه خاص، این عقیده که "هرکس حق دارد توسط هیئت منصفه‌ای از سنخ خود محاکمه شود" چگونه باید تعبیر و تفسیر شود؟ افراد هم‌سنخ هرکس چه کسانی هستند؟ مثلاً اگر شخصی که محاکمه می‌شود، سیاهپوست باشد آیا هیئت منصفه باید به همان نسبت دارای اعضای سیاهپوست باشد که در محل دادرسی زندگی می‌کنند؟ یا در بخش؟ یا در ایالت؟ یا در کشور؟ این مورد در دادرسیهای گوناگون در سطح کشور به طرق گوناگون حل می‌شود.

کارشناسان بین مراحل مختلف فرایند انتخاب هیئت منصفه تمایز قائل می‌شوند. ابتدا افرادی را که به‌طور بالقوه می‌توانند عضو هیئت منصفه باشند از روی فهرست ثبت‌نام

رأی‌دهندگان و فهرستهای مالکان خودروها به‌طور تصادفی انتخاب می‌کنند وبرای این کار محدودیتهایی نیز قائل می‌شوند.؛ ازقبیل اینکه "یک عضو بالقوه هیئت منصفه نباید مجبور باشد مسافتی بیش از ۳۰ مایل را از موطن خود طی کند." و برای تداخل مرزهای هر حوزه قضایی نیز تمهیداتی می‌اندیشند. آیا یک فرایند انتخاب با این صورت قیودات واقعاً تصادفی است؟ مرحله دیگری از انتخاب اعضای هیئت منصفه مستلزم حذف افرادی است که از دید قانون اساسی واجد شرایط نیستند. سپس نوبت حذف افرادی است که از نظر وکلای مدافع، مطلوب نیستند (ادای سوگند). این مرحله به انتخاب یک هیئت منصفه کامل منتهی می‌شود (که معمولاً برای یک هیئت منصفه در سطح شهر، بخش یا ایالت از ۱۲ نفر و در سطح فدرال از ۶ نفر تشکیل می‌شود) و علی‌البدلهایی نیز از میان مجموعه اعضای بالقوه برای هیئت منصفه درنظر گرفته می‌شود.

شیوه‌های آماری برای بررسی جنبه‌های گوناگون این فرایند انتخاب به‌طور مکرر مورد استفاده قرار گرفته است. برای مثال، از شیوه‌های آماری استفاده شده‌است تا بررسی شود که یک هیئت منصفه خاص تا چه اندازه "معرف" ناحیه‌ای است که جرم در آن اتفاق افتاده‌است؟ یا برخی گروههای اقلیت منظماً نادیده گرفته شده‌اند؟ از شیوه‌های آماری برای بررسی اینکه روش انتخاب تا چه اندازه تصادفی بوده نیز استفاده شده‌است. با این حال، برای جلوگیری از طولانی‌شدن این مقاله، بحث خود را به استفاده از آمار در فرایند ادای سوگند محدود می‌کنیم.

عبارت ادای سوگند، به زبان فرانسه Voir dire به‌صورت تحت‌اللفظی به معنای "گفتن حقیقت" است. در عمل، گروه بزرگی از اعضای بالقوه برای هیئت منصفه به‌طور تصادفی انتخاب می‌شوند (تا ۱۰۰نفر) که از میان آنها ۶ تا ۱۲ نفر عضو به اضافه اعضای علی‌البدل برگزیده می‌شوند. این گروه از اشخاصی تشکیل می‌شود که ازنظر قانون واجد شرایط خدمت در هیئت منصفه هستند. اما همان‌طور که همه می‌دانند اشخاص پیشداوریهایی دارند که آنها را در یک دعوای خاص دخالت می‌دهند. ایده‌آل جامعه آن است که اعضای هیئت منصفه در دعوایی که داوری می‌کنند، بی‌طرف باشند. به همین دلیل وکلای مدافع هر دو طرف معمولاً اجازه دارند از اعضای بالقوه هیئت منصفه انبوهی از سئوالات را بپرسند که به منظور شناسایی افرادی طراحی شده است که احتمال دارد بی‌طرف نباشند. (در دعاویی که در سطح فدرال

مطرح می‌شود معمولاً قاضی سئوالات را می‌پرسد ولی به وکلای مدافع اجازه داده می‌شود که فهرستی از سئوالات را به قاضی ارائه دهند. هرچند که قاضی ممکن است آن سئوالات را مطرح کند یا نکند.) پیشاپیش تعدادی حق رد صلاحیت عضویت در هیئت منصفه به وکلای مدافع اختصاص داده شده‌است. (بدین معنی که به وکلای مدافع فرصتهایی داده می‌شود تا بدون نیاز به اقامه دلیل از یک عضو بالقوه هیئت منصفه رد صلاحیت کنند.) وکلا از این حق رد صلاحیت استفاده می‌کنند تا اشخاصی را که احتمال دارد نسبت به موکل آنها در دعوای مورد بحث نظر مخالف داشته باشند از عضویت در هیئت منصفه بازدارند. فرایند پرس‌وجو از اعضای بالقوه از این راه که طی آن از آنها درخواست می‌شود تا "حقیقت را بگویند" ادای سوگند نامیده می‌شود.

وکیل مدافع از کجا می‌داند کدام عضو بالقوه هیئت منصفه را باید ردصلاحیت کند؟ تاریخ نشان داده‌است که وکلای مدافع اعضای بالقوه هیئت منصفه را براساس تجربه و احساس غریزی خود ردصلاحیت می‌کنند. اگر عکس‌العمل غریزی آنها نسبت به یک عضو بالقوه منفی باشد (یا این شخص آشکارا نظر نامساعدی نسبت به موکل آنها داشته‌باشد) عذر وی را از خدمت در هیئت منصفه می‌خواهند. امروزه این روش تا حدودی علمیتر شده و گاهی اوقات بر تحلیل آماری و علوم اجتماعی مبتنی است.

به‌کارگیری روشهای آماری

استفاده از آمار و علوم اجتماعی برای کمک به فرایند ادای سوگند به شکل رسمی آن به اوایل دهه ۱۹۷۰ برمی‌گردد. یعنی به محاکمه *میشل استانس*، محاکمه *هریس برگسون*، محاکمات *اتیکا* و سایر محاکمات. در این دعاوی برای کمک به فرایند ادای سوگند (و نیز برای کمک در سایر جوانب پرونده) از روشهای علمی به‌جای مدعی‌العموم توسط وکلای مدافع استفاده به عمل آمد. این دعاوی نوعاً متضمن مسائل سیاسی، خوانده‌های متعدد و شهره‌شدن بیش از اندازه بودند. در مواقعی که موضوع قبل از محاکمه بیش از حد برسر زبانها می‌افتاد که غالباً به ضرر وکلای مدافع پرونده بود و انتخاب افرادی را که نظر نامساعدی نداشته باشند، برای عضویت در هیئت منصفه با دشواری بیشتری همراه می‌ساخت، وکلای مدافع معمولاً اجازه می‌یافتند برای

جبران این وضعیت تعداد بیشتری را ردصلاحیت کنند. روش‌شناسی آماری برای انتخاب هیئت منصفه به شیوه علمی توسط پژوهشگران گوناگونی پیشنهاد شده‌است. (نحوه پرداختن آنها به انتخاب هیئت منصفه ناظر برجنبه های متفاوتی از دعاوی بود که از جنبه‌های آماری، اجتماعی، روان‌شناختی و اخلاقی گرفته تا جنبه‌های حقوقی را دربر می‌گرفت. ن.ک. منابع.)

در اینجا روش‌شناسی مزبور را از طریق یک مثال نشان خواهیم داد که در آن از خصوصیات جمعیت‌شناختی اعضای بالقوه برای عضویت در هیئت منصفه استفاده شده‌است تا شانس انتخاب افرادی که نظر نامساعد ندارند برای عضویت در هیئت منصفه بیشتر شود. وضعیتی که شرح داده می‌شود، شرکت مورد بحث و داده‌های مورد استفاده همه ساختگی است ولی دعوای مورد نظر ریشه در عدم توافقهای واقعی میان سازمانهای واقعی دارد: حقایق اصلی مربوط به مالکیت است. روش‌شناسی انتخاب هیئت منصفه در این دعاوی و بسیاری دعاوی دیگر با موفقیت مورد استفاده قرار گرفته‌است.

شرکت شیمیایی ماندویل طی سالها به تولید اقلامی می‌پرداخت که محتوی مواد شیمیایی بود. این مواد شیمیایی موارد استفاده سودمندی در صنعت داشت ولی معلوم شده‌بود که یکی از اقلام تولیدی این شرکت خاصیت سمی دارد و می‌تواند باعث بیماریهای گوناگونی در انسان از جمله سرطان شود. بسیاری از کارگران علیه شرکت اقامه دعوا کردند تا شرکت هزینه‌های پزشکی آنان را تأمین کند و شرکت نیز این هزینه‌ها را پرداخت کرد ولی به سهم خود علیه شرکتهای بیمه خود، پرونده‌ای تنظیم نمود که در آن ادعا کرده‌بود؛ شرکتهای بیمه خسارت وارده را جبران نکرده‌اند. شرکتهای بیمه نیز استدلال می‌کردند که شرکت شیمیایی ماندویل هرگز به آنها اطلاع نداده بود که درکار تولید محصولی خطرناک است و اگر می‌دانستند با بیمه آن شرکت موافقت نمی‌کردند. دعوا بسیار بغرنج بود. مورد ساده‌ای نبود که در آن یک "آدم معمولی" یک شرکت بزرگ را تحت تعقیب قانونی قرار داده‌باشد. مورد دعوا، تعقیب قانونی یک "غول" توسط یک "غول" دیگر بود. چه نوع آدمهایی به‌عنوان اعضای هیئت منصفه احتمال داشت نسبت به شرکت شیمیایی ماندویل نظر نامساعد (یامساعد) داشته باشند؟

شرکت تصمیم گرفت از طریق تلفن از ۸۰۰ نفر آمارگیری کند. قرار بود از این افراد سؤال شود که آیا از شهروندان آمریکا هستند؟ آیا در حوزه صلاحیت دادگاه (یعنی در حوالی ناحیه‌ای

که قرار بود محاکمه در آنجا انجام پذیرد) زندگی می‌کنند؟ و برخی سئوالات دیگر تا تعیین شود که آیا این افراد می‌توانند عضو بالقوه هیئت منصفه تلقی شوند یا خیر؟ (معلوم شد که از میان افراد مورد آمارگیری ۷۲۰ نفر عضو بالقوه هیئت منصفه‌اند.) سپس از اعضای بالقوه انبوهی از سئوالات پرسیده شد. از شماره‌گیری تصادفی استفاده شد تا کسانی که شماره تلفن آنها در کتاب راهنمای تلفن موجود نبود، مورد پرسش قرار گرفتند. در این روش سه رقم اول شماره تلفن به‌گونه‌ای انتخاب شد که با منطقه جغرافیایی موردنظر مطابقت داشته باشد و چهار رقم بعدی به‌طور تصادفی انتخاب شد. درنتیجه، شماره انتخابی ممکن بود در کتاب راهنمای تلفن موجود باشد یا نباشد. این کار به‌طور مکرر انجام شد تا تعداد زیادی از افراد آمارگیری شوند. سپس سه رقم اول تغییر داده‌شد و این روش مجدداً تکرار گردید. با نمونه‌ای به بزرگی ۷۲۰ نتایج حاصل از اینکه اعضای بالقوه هیئت منصفه در مورد پرونده چه دیدگاهی دارند با درجه بالایی از اطمینان به جامعه بزرگتر اعضای بالقوه هیئت منصفه در کل ناحیه حوزه صلاحیت دادگاه قابل تعمیم است. (این تعمیم هنگامی امکان‌پذیر است که نمونه به‌طور تصادفی از کل جامعه اعضای بالقوه هیئت منصفه در حوزه صلاحیت دادگاه انتخاب شده‌باشد.) حقایق موجود در پرونده به اطلاع پاسخگویان آمارگیری رسید و سپس از آنها خواسته شد تا ویژگیهای شخصی خود ازقبیل جنسیت، سن، قومیت، رده درآمد و امثال اینها را ذکر کنند. برخی سئوالات تکمیلی نیز پرسیده شد تا معلوم شود که نکته‌ای در پرونده بوده‌است که درک آن مشکل باشد یا خیر. (اگر چنین بود مورد دعوا از دید شاکی را می‌توان شفافتر کرد تا نکات مبهم روشن شود.) در آخر از پاسخگویان خواسته می‌شد تا اعلام کنند که اگر در هیئت منصفه‌ای که به این مورد رسیدگی می‌کرد، عضویت داشتند چگونه رأی می‌دادند.

تعبیر نتایج

معلوم شد که از ۷۲۰ نفر افراد آمارگیری شده که واجد شرایط عضویت در هیئت منصفه برای محاکمه موردنظر بودند ۶۵درصد نسبت به شرکت شیمیایی ماندویل نظر نامساعد و ۳۵درصد نظر مساعد داشتند. مسئله عبارت از این بود که چگونه می‌شد این دوگروه را از یکدیگر تمیز داد به‌طوری که از متغیرهای "نشانگر" اجتماعی ـ اقتصادی برای جانب‌گیری علیه شرکت

شیمیایی بتوان برای تعیین اعضایی که باید رد صلاحیت و حذف شوند، استفاده نمود.

جدول‌بندی دوطرفه نسبتهای پاسخگویان که نسبت به شرکت شیمیایی (شاکی) نظر مساعد و نامساعد داشتند، برحسب سن و جنسیت به‌ترتیب در جدول ۱و۲ نشان داده شده‌است.

وقتی شرکت شیمیایی ماندویل برای اولین بار به ۳۵درصد کل تعداد پاسخگویانی که به نفع شاکی نظر داده‌بودند، نگاه کرد چیزی در مورد آنها نمی‌دانست. ولی بررسی جدول ۱ نشان داد که ۶۶درصد از جمعیتی که نسبت به شرکت شیمیایی نظر مساعد داشتند، بین ۲۱ تا ۴۰ساله بودند (۰/۲۳ به ۰/۳۵) درحالی که فقط ۱۱درصد از افراد ۴۱ تا ۵۵ساله (۰/۰۴ به ۰/۳۵)، ۶درصد ۵۶ تا ۷۰ساله (۰/۰۲ به ۰/۳۵) و ۱۷درصد از افراد مسن‌تر از ۷۰ساله با شرکت شیمیایی موافق بودند. مهمتر اینکه شرکت شیمیایی دریافت که احتمال داشتن نظر مساعد نسبت به موافقت با شرکت در میان افراد گروه سنی ۲۰ تا ۴۰ساله بیشتر از افراد مسن‌تر است. اگرچه در گروه با تعداد اندکتر ۵۶ تا ۷۰ساله احتمال نظر مساعد داشتن نسبت به شرکت کمی بیشتر از آن گروه به نظر می‌رسید (۰/۴۰ در مقابل ۰/۳۸). همچنین از جدول ۱ به وضوح پیداست که احتمال جانبداری از شرکت شیمیایی در میان افراد گروه ۲۱ تا ۴۰ساله بیشتر از گروه ۴۱ تا ۵۵ ساله است (۰/۳۸ در مقابل ۰/۲۹). جدول ۲ نشان می‌دهد که ۱۳درصد از مردان نسبت به شرکت نظر مساعد داشتند (۰/۰۸ به ۰/۶۰) درحالی که احتمال جانبداری زنان

جدول ۱ـ جدول‌بندی دوطرفه برحسب سن

سن	نسبت افرادی که به شرکت نظر مساعد ندادند	نسبت افرادی که به شرکت نظر مساعد دادند	مجموع کل	نسبت گروههای سنی که به شرکت نظر مساعد دادند
۲۱ تا ۴۰ ساله	۰/۳۷	۰/۲۳	۰/۶۰	۰/۳۸
۴۱ تا ۵۵ ساله	۰/۱۰	۰/۰۴	۰/۱۴	۰/۲۹
۵۶ تا ۷۰ ساله	۰/۰۳	۰/۰۲	۰/۰۵	۰/۴۰
بالاتراز ۷۰ ساله	۰/۱۵	۰/۰۶	۰/۲۱	۰/۲۹
جمع	۰/۶۵	۰/۳۵	۱/۰۰	۰/۳۵

جدول ۲ـ جدول‌بندی دوطرفه برحسب جنسیت

جنس	نسبت افرادی که به شرکت نظر مساعد ندادند	نسبت افرادی که به شرکت نظر مساعد دادند	مجموع کل	نسبت گروههای سنی جنسیتی موافق باشرکت
مرد	۰/۵۲	۰/۰۸	۰/۶۰	۰/۱۳
زن	۰/۱۳	۰/۲۷	۰/۴۰	۰/۶۸
جمع	۰/۶۵	۰/۳۵	۱/۰۰	۰/۳۵

از شرکت شیمیایی پنج برابر مردان بود (۰/۶۸ به ۰/۱۳). به‌این ترتیب، برای شرکت شیمیایی ماندویل آشکار شد که اگر قرارباشد تمام تلاش خود را از طریق ردصلاحیت اعضای بالقوه برای حذف از عضویت در هیئت منصفه به عمل بیاورد باید سعی کند از انتخاب مردان و اشخاص مسن‌تر جلوگیری شود. از متغیرهای دیگر هم می‌شد استفاده کرد. اگرچه ارقام فوق در مورد این پرونده بخصوص به‌دست آمد، در پرونده‌های دیگر ممکن است مجموعه کاملاً متفاوتی از ارقام حاصل شود. این بدان معناست که ویژگیهای جمعیت‌شناختی مردم در جهت شکل دادن نظرات آنها برحسب محتوای هر پرونده نقش دارد.

از آنجا که اطلاعاتی در مورد قومیت، تحصیلات، درآمد، کار یدی داشتن در برابر پشت میزنشینی و سایر متغیرها نیز موجود بود؛ جدول‌بندی دوطرفه این متغیرها نیز امکان اجرا داشت. برای مثال جدول ۱ می‌توانست به جدولی با ابعاد بیشتر گسترش پیدا کند (یعنی جدولی شامل متغیرهای بیشتر) به نحوی که بتوانیم نسبت افرادی را که مثلاً، هم مرد هستند وهم سفیدپوست و هم فارغ‌التحصیل دانشگاه و از این قبیل در آن ثبت کنیم. در این‌صورت می‌توان نتایج آگاهی بخش‌تری کسب کرد. همچنین می‌توان با ساختن مدلهای آماری به شناخت آن دسته از افراد که نظرمساعدی نسبت به شاکی پرونده نداشتند، نایل آمد.

البته ویژگیهای شخصی غالباً نشانگرهای بسیار خامی برای دلالت بر رفتار رأی‌دهی اعضای هیئت منصفه است. نشانگرهای بهتر عبارتنداز: متغیرهایی که نماینده نگرشها، رفتار واقعی، اعتقادات اجتماعی و سیاسی و امثال اینهاست. این قبیل متغیرها را می‌توان بااستفاده از یک

آمارگیری تلفنی برای پرسیدن سئوالاتی از قبیل اینکه پاسخگو چه روزنامه‌هایی را می‌خواند؟ وابستگی سیاسی او کدام است؟ چه تمایلات مذهبی دارد؟ به کدام باشگاه یا سازمان تعلق دارد؟ رشته اصلی تحصیلات دانشگاهی او چه بوده‌است؟ و آیا اصلاً به دانشگاه رفته‌است؟ را مورد سنجش قرار داد. تحقیقات علوم اجتماعی نشان داده‌است که محتملاً این قبیل متغیرها نسبت به متغیرهای جمعیت شناختی، نماگرهای بهتری برای نشان دادن رفتار اعضای هیئت منصفه به هنگام رأی‌گیری باشند.

وظیفه وکیل مدافع آن است که تا آنجا که امکان دارد معرف موکل خود باشد. یعنی وکیل مدافع حامی موقعیت موکل خویش است. وکیل مدافع از روشهای انتخاب هیئت منصفه استفاده می کند تا ادعای موکل خود را پیش ببرد. اگر این روش شانس انتخاب برخی از اعضای هیئت منصفه را که پیشاپیش نسبت به دعوای موکل موضع نامساعدی اختیار خواهند کرد به حداقل برساند، شهادت شهود هرچه باشد، کل سیستم قضایی منتفع خواهدشد.

مسائل

۱ـ فرض کنید به‌جای داده‌های تفکیک شده مانند متن مقاله، داده‌های توأم در مورد سن (پیر در مقابل جوان) و جنسیت (مرد در مقابل زن) موجود است و اطلاعات توأم به شکل زیر توزیع شده‌است:

	نسبت افراد با نظر نامساعد باشرکت			نسبت افراد با نظر مساعد باشرکت			جمع کل
	مرد	زن	جمع	مرد	زن	جمع	
جوان	۰/۲۷	۰/۱۰	۰/۳۷	۰/۰۶	۰/۱۷	۰/۲۳	۰/۶۰
پیر	۰/۲۵	۰/۰۳	۰/۲۸	۰/۰۲	۰/۱۰	۰/۱۲	۰/۴۰
جمع کل	۰/۵۲	۰/۱۳	۰/۶۵	۰/۰۸	۰/۲۷	۰/۳۵	۱/۰۰

الف) در هریک از گروههای سنی جنسیتی چهارگانه، نسبت افراد با نظر مساعد با شرکت را پیدا کنید.

ب) از نسبت افراد با نظر مساعد با شرکت شیمیایی که در مسئله ۱-الف پیدا کردید، آیا هیچیک

از آنها از بزرگترین نسبت از نسبتهای چهارگانه جدول ۱ و دو نسبت جدول ۲ بزرگتر است؟ بر این اساس، وکلای مدافع شرکت شیمیایی بیش از همه از کدام دوگروه سنی ـ جنسیتی مایل خواهند بود رد صلاحیت کنند و مدعی علیه از کدام دو گروه ردصلاحیت خواهدکرد؟

۲ـ الف) در جدول مسئله ۱، فرض کنید مجموعهای جزء ثابت بماند و شما اجازه داشته باشید عدد خانه مرد ـ جوان با نظر نامساعد را کاهش دهید. کمترین مقداری که این خانه می‌تواند داشته باشد، چقدر است؟

ب) کوچکترین رقم ممکن برای خانه مرد ـ پیر با نظر مساعد را پیدا کنید به‌طوری که مجموع جزء ثابت بماند.

۳ـ در جدول ۱ کدامیک از گروههای سنی چهارگانه دارای بیشترین تعداد افرادی بود که نسبت به شرکت شیمیایی ماندویل نظر نامساعد داشتند؟

۴ـ در جدول ۱ کدام گروه سنی دارای کوچکترین نسبت از افراد با نظر نامساعد نسبت به شرکت شیمیایی ماندویل بود؟

منابع

Reid Hastie, Steven D. Penrod, and Nancy Pennington. 1983. *Inside the Jury Cambridge, Mass: Harvard University Press.*

David Kairys, ed. 1975. *The Jury System: New Methods for Reducing Prejudice.* Philadelphia: Philadelphia Resistance Print Shop.

W. McCart. 1965. *Trial by Jury: A Complete Guide to the Jury System,* 2nd ed. Phladelphia: Chilton Books.

S. J. Press. 1988. *Statistics in Jury Selection.* department of Statistics Technical Report No 163. Riverside, Calif: University of California.

J. Schulman. P. Shiver, R. Colman. B. Enrich. and R. Christie. 1973. "Reciepe for a Jury." *Psychology Today 6(May):*37-44, 77-84.



اندازه‌گیری آثار نوآوریهای اجتماعی به‌وسیله سریهای زمانی

دونالد تی. کمپبل[1] دانشگاه لیگ

ما در عصری از اصلاحات اجتماعی زندگی می‌کنیم که در آن تلاشهای پردامنه‌ای برای اصلاح مشکلات خاص اجتماعی صورت می‌گیرد. در گذشته از اکثر این تلاشها ارزیابی کافی به‌عمل نیامده است. معمولاً شواهدی که از نظر علمی معتبر باشند، دال بر اینکه مشکل مورد نظر برطرف شده یا نه؛ وجود ندارد. چون برای یک مسئله همواره راه‌حلهای پیشنهادی گوناگون

۱- *Donald T. Campbell*

وجود دارند و علاوه‌بر این مسائل متعدد دیگری هستند که نیازمند اعتبار مالی و توجهند. لذا توان جامعه برای درک میزان تأثیر هر نوآوری خاص حائز اهمیت است.

از دید آمارشناس بهترین آزمایشهای طراحی شده، چه در آزمایشگاه و چه در اجتماع، شامل فراهم کردن یک گروه برای آزمایش است و یک گروه شاهد که از هر لحاظ مشابه گروه آزمایش می‌باشد. بجز اینکه تیمار آزمایشی مربوط را دریافت نمی‌کند. روش آمارشناس برای دست‌یابی به این هم‌ارزی همه‌جانبه گروه‌های آزمایش و کنترل، تصادفی کردن است. اشخاص (کرتها یا واحدهای دیگر) به‌طور تصادفی (با انداختن یک تاس) به گروه آزمایش یا گروه شاهد تخصیص می‌یابند. پس از تیمار، دو گروه مقایسه می‌شوند و تفاوتهای با مقادیری بیشتر از آنکه شانس (تصادف) قادر به توجیه آنها باشد، به تیمار آزمایش نسبت داده می‌شوند. این روش ایده‌آل در آزمونهای راهبر در علوم اجتماعی مورد استفاده قرار می‌گیرد. برای مثال می‌توان آزمایش درآمد ـ مالیات منفی نیوجرسی (*کرشا و فیر ۱۹۷۶*، *واتس و ریس ۱۹۷۷*) را ذکر کرد که در آن چند صد خانواده شاغل با درآمد پایین که حاضر به همکاری شدند، به‌طور تصادفی به گروههای آزمایشی (که درآمدهای تکمیلی با اندازه‌های مختلف دریافت می‌داشتند) و یک گروه شاهد (که هیچ کمک مالی دریافت نمی‌کرد) تخصیص داده شدند. آنگاه آثار این کمک مالی بر مقادیر درآمدهای دیگر خانوار و بر سلامتی خانواده، تعادل خانواده و نظایر آنها مورد مطالعه قرار گرفت. (مقاله *ژوزف نیوهاوس* دراین کتاب که آزمایشی در بیمه سلامتی را شرح می‌دهد، مثالی از این رهیافت است).

متأسفانه اگرچه این طرحهای آزمایشی ایده‌آل هستند، اغلب انجام‌پذیر نیستند. استفاده از این طرحها برای مثال در مورد ارزیابی هر برنامه جدیدی که به همه شهروندان به‌طور همزمان قابل اعمال باشد، مانند اکثر تغییراتی که در قوانین می‌توان ایجاد کرد، غیرممکن است. در این موارد متداولتر، وجوهی با رضایتبخشی بسیار کمتر استنباط از آزمایش باید کفایت کند. طرح سریهای زمانی گسیخته که تمرکز این مقاله بر آن است، یکی از سودمندترین این‌گونه طرحهای شبه ـ آزمایشی است. تفسیر صحیح این‌گونه داده‌ها، مسائل آماری پیچیده‌ای پیش می‌آورد که بعضی از آنها هنوز به‌طور رضایتبخشی حل نشده‌اند. در این مقاله به بعضی از این مسائل در قالب کلمات و نمودارها، به‌جای نمادهای ریاضی، اشاره می‌شود. بحث را با درنظر گرفتن دو

مورد واقعی شروع می‌کنیم.

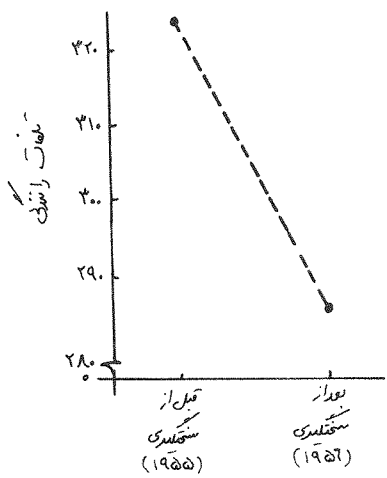
سختگیری نسبت به رانندگی سریعتر از حد مجاز در کنتیکات

در ۲۳ دسامبر ۱۹۵۵ ایالت کنتیکات سختگیری بسیار استثنایی و مستمری را درباره سرعت غیرمجاز تصویب نمود. مانند اغلب گزارشهای دولتی از اثربخشی برنامه‌ها، نتایج حاصل در قالب اندازه‌های ساده قبل و بعد، تعیین مقایسه ارقام مربوط به امسال با ارقام سال قبل، گزارش شد.

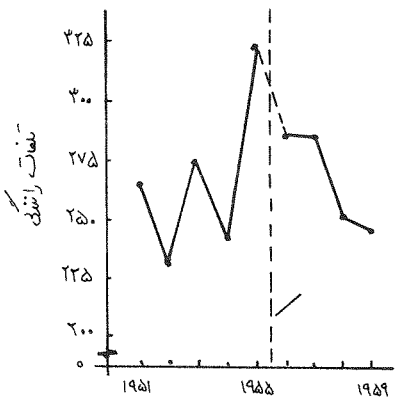
یعنی تعداد کل ۲۸۴ مرگ در سال ۱۹۵۶ با تعداد کل ۳۲۴ مرگ در سال ۱۹۵۵ مقایسه شد و فرماندار گفت؛ با نجات ۴۰ زندگی در سال ۱۹۵۶ ..... می‌توان گفت که قطعاً برنامه از ارزش لازم برخوردار است. شکل ۱ داده‌ها را به‌صورت نموداری نشان می‌دهد ولی این طرح شبه ـ آزمایشی ساده خیلی ضعیف و گمراه‌کننده است.

توجیه‌های ممکن بسیار دیگری برای تغییر شمار تلفات بزرگراهها از ۳۲۴ به ۲۸۴ وجود دارند. با نسبت دادن این تغییر به سختگیری به‌عمل آمده، آقای فرماندار به‌طور ضمنی فرض را بر این قرار داده‌است که بدون سخت‌گیری به‌عمل آمده هیچگونه تغییری بوجود نمی‌آمد.

شکل ۱ـ تلفات رانندگی کنتیکات ۱۹۵۵ تا ۱۹۵۶. منبع: کمپل راس (۱۹۶۸)



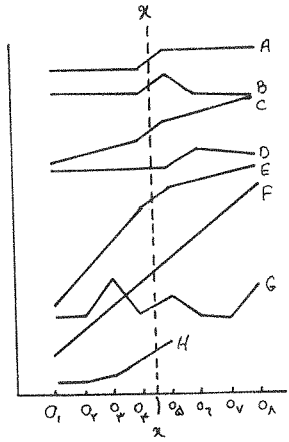
شکل ۲ـ تلفات رانندگی کنتیکات، ۱۹۵۱ تا ۱۹۵۹. منبع: کمپبل و راس (۱۹۶۸)



استفاده از تلفات ثبت شده مربوط به چند سال قبل وبعد، و ارائه آنها به‌صورت سری زمانی به‌مقدار زیادی به قدرت تحلیل می‌افزاید. شکل ۲ چنین داده‌هایی را در مورد سخت گیری کنتیکات نشان می‌دهد. در این حوزه بزرگتر، افت ۱۹۵۵ تا ۱۹۵۶ از شمار تلفات بسیارناچیز به‌نظر می‌رسد. می‌توان دید که فرض ضمنی فرماندار، که زیربنای گفته‌اش را تشکیل می‌دهد با اطمینان قریب به یقین نادرست است.

برای بررسی بیشتر این مطلب به شکل ۳ توجه کنید که به سبک خاصی نشان می‌دهد چگونه تغییر یکسان در مقادیر، قبل و بعد از تیمار، ممکن است در بعضی موارد دلیل قاطعی بر وجود اثر باشد و در موارد دیگر مدرکی دال بر وجود تغییر نباشد. لذا براساس فقط داده‌های ۱۹۵۵ و ۱۹۵۶ افت ۴۰ نفر در تعداد تلفات، در شکل ۱ نشان داده شده‌است. ممکن است قسمتی از یک افت سالیانه پایدار نزولی در شمار تلفات بوده‌باشد که از قبل وجود داشته‌است (عکس روند پایدار صعودی درخط F درشکل ۳)، یا نتیجه یک زیگزاگ ناپایدار (مانند خط G درشکل ۳) بوده باشد. شکل ۲ نشان می‌دهد که در مورد مسئله کنتیکات حالت زیگزاگ برقرار بوده است. افت ۱۹۵۵ تا ۱۹۵۶ تقریباً به همان اندازه افتهای ۱۹۵۱ تا ۱۹۵۲، ۱۹۵۳ تا ۱۹۵۴، ۱۹۵۷ تا ۱۹۵۸، که در آن زمانها هیچگونه سخت‌گیری نسبت به سرعت برای توجیه آنها وجود

شکل ۳ـ طرحهایی از برآمدهای ممکن اعمال یک تیمار در نقطه X به یک سری از اندازه‌های O1 تا O8. افزایش از O4 تا O5 برای همه سریهای زمانی بجز برای D، یکسان است، درحالی که اعتبار استنباط وجود اثر به میزان وسیعی تغییر می‌کند، که در A و B قویترین و در F و G و H کاملاً غیرقابل توجیه است. منبع: کمپبل و راس (۱۹۶۳)

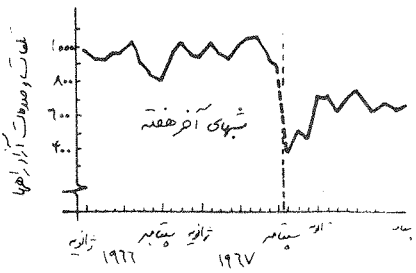


نداشته‌است، به علاوه افت ۱۹۵۵ تا ۱۹۵۶ فقط نصف افزایش ۱۹۵۴ تا ۱۹۵۵ است. بنابراین باتوجه به همه این ناپایداری قبلی در نمودار، بعید است کسی همه افت ۱۹۵۵ تا ۱۹۵۶ را معلول سخت‌گیری نسبت به سرعت غیرمجاز، آن‌طور که فرماندار تصور کرده‌است، بداند. بعداً شکل ۲ را مجدداً برای طرح سئوال مشکلتری در استنباط، بررسی خواهیم کرد. ولی قبل از آن وقت بیشتری را روی مسئله پایداری، به کمک مثالی از یک اصلاح که دیرباورترین روش‌شناسان موفقیت آن را باور دارند، صرف می‌کنیم.

سخت‌گیری تحلیلگر بازدم در بریتانیا در ۱۹۶۷

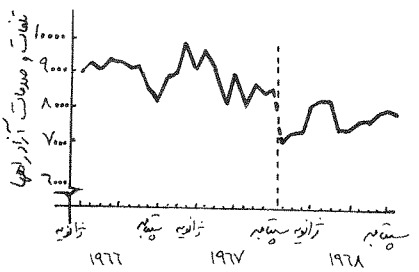
در سپتامبر ۱۹۶۷ دولت بریتانیا برنامه جدیدی را برای اجرای مقررات مربوط به رانندگی در حال مستی شروع کرد. نام معروف این برنامه مأخوذ از نام وسیله‌ای است که برای حصول اطمینان از درجه تأثیرالکل، براساس تحلیل نمونه‌ای از بازدم راننده، به‌کار می‌رود. پلیس این آزمون ساده را در مورد رانندگانی که به‌خاطر مظنون بودن متوقف شده‌بودند، به‌کار می‌برد و در

شکل ۴ـ آثار سخت‌گیری تحلیلگر بازدم سپتامبر ۱۹۶۷ انگلستان در مورد رانندگی در حالت مستی. مرگ ومیر به اضافه صدمات جدی، جمعه‌ها و شنبه‌ها، از ۱۰ بعدازظهر تا ۴ صبح، در ماه. منبع: راس، کمپبل وگلاس (۱۹۷۰)



صورت مثبت بودن آزمون، آنها را برای آزمونهای کاملتر به اداره پلیس هدایت می‌کرد. این شیوه آزمون جدید با مجازاتهای شدیدتری، از جمله تعلیق گواهینامه رانندگی، همراه بود. شکل ۴ آثار این سختگیری را بر تلفات شبهای جمعه و شنبه (مرگ و میر به‌اضافه صدمات جدی) نشان می‌دهد. اثر به‌صورتی بارز واضح است. به‌نظر می‌رسد که یک افت بلافاصله در حدود ۴۰٪ و

شکل ۵ـ آثار سخت‌گیری تحلیلگر بازدم سپتامبر ۱۹۶۷ انگلستان. تلفات به‌اضافه صدمات جدی در همه ساعات و همه روزها، درماه. منبع: راس، کمپبل وگلاس (۱۹۷۰)



ثبات تقریبی در سطحی که ۳۰٪ پایینتر از سطح قبل از سختگیری است، وجود دارد. گرچه به سختی می‌توان از این نتیجه مطمئن بود؛ زیرا نمی‌دانیم که زمان چه تغییراتی در نرخ تلفات بدون این سخت‌گیری به بار می‌آورد.

آیا همین اثر، وقتی که تلفات همه ساعات و همه روزها جمع شوند، بازهم خود را نشان خواهد داد؟ شکل ۵ چنین داده‌هایی را نشان می‌دهد. گرچه احتمالاً چنین اثری بازهم وجود دارد ولی کمتر واضح است، افت حاصل از سخت‌گیری خیلی بیشتر از ناپایداری غیرقابل توجیه دوره‌های زمانی دیگر نیست. (ولی افت ماه به ماه حاصل از سخت‌گیری نه‌تنها در دوره زمانی که در نمودار آمده‌است، بلکه برای دوره طولانی‌تر تا ۱۹۶۱، در مورد داده‌هایی که در آنها نوسانهای فصلی حذف شده‌اند، بیشترین بوده‌است.)

تحلیل آماری ناپایداری

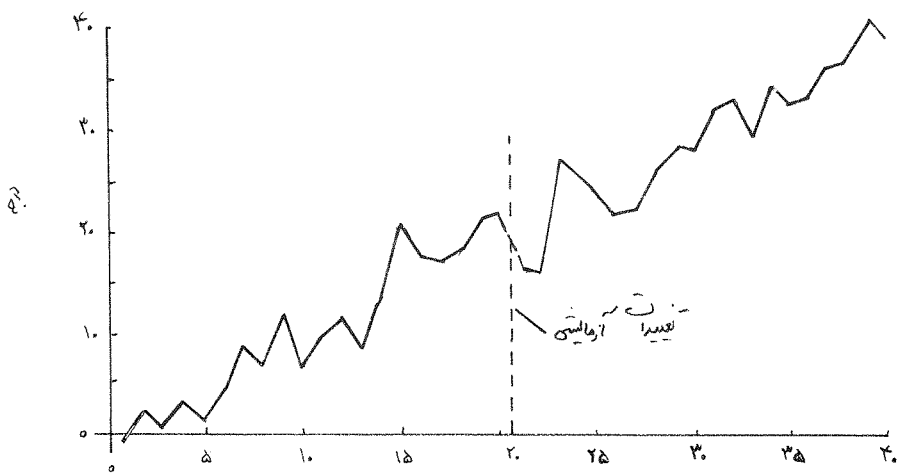
مسئله‌ای که در مقابل آمارشناس قرار دارد، صورت‌بندی زمینه‌های انجام دادن استنباطی است که ما به‌طور غیررسمی یا شهودی هنگام قضاوت درمورد نمودارهای فوق به‌کار برده‌ایم. بدیهی است که هرچه خط ناپایداری بیشتری، قبل از نقطه تغییر سیاست یا اعمال تیمار، داشته باشد، برای متقاعدکردن ما به وجود یک اثر واقعی، باید تفاوت بیشتری موجود باشد. یک رهیافت آمارشناس این است که فرض کند سری زمانی نتیجه یک روند کلی به‌اضافه انحرافات تصادفی مشخص در هر زمانی است. نظریه این نوع تحلیل، زمانی که انحرافات تصادفی در هر نقطه کاملاً مستقل از انحرافات در سایر نقاط است، به‌خوبی بررسی و مشخص شده‌است. ولی در وضعیتهای مربوط به زندگی واقعی منابع انحراف و یا آشفتگی درهر نقطه می‌تواند مشابه نقاط مجاور و نزدیک و نامشابه نقاط دور از لحاظ زمان باشد. این موضوع وضعیتهای گمراه‌کننده‌ای، هم برای آزمونهای معنی‌دار بودن و هم برای تفسیرهای مبتنی بر مشاهدات بصری ایجاد می‌کند. شکلهای ۶ و ۷ این مطالب را به‌وسیله سریهای زمانیی که به‌وسیله کامپیوتر شبیه‌سازی شده‌اند، روشن می‌کند. برای هرنقطه از زمانهای ۱ تا ۴۰ یک نمره واقعی وجود دارد. اگر نمودارهای این نمرات واقعی رسم شوند، یک خط راست قطری از نمره واقع در سمت چپ پایین تا نمره ۴۰ در سمت راست بالا به‌وجود می‌آید که در آن هیچگونه کجی یا شکستگی در

نقطه فرضی اعمال تیمار وجود ندارد. این نمرات واقعی برای شکلهای ۶ و ۷ یکسانند. به هر نمره واقعی یک انحراف، که به‌طور تصادفی انتخاب شده‌است، اضافه یا از آن کم شده‌است. در شکل ۶ انحراف در هر نقطه زمانی مستقل از هر انحراف دیگر انتخاب شده‌است. این شبیه‌سازی حالتی است که در آن تیمار فرضی اعمال‌شده بین دوره‌های زمانی ۲۰ و ۲۱ هیچ تأثیری ندارد.

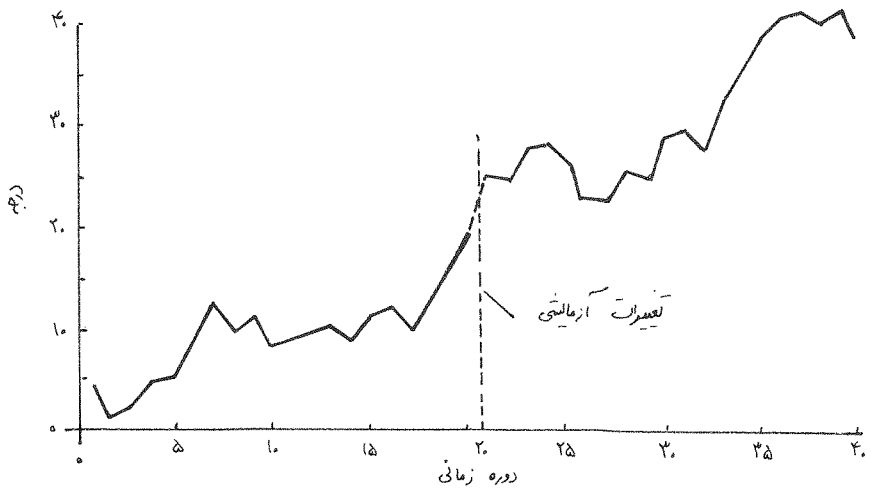
گاهی برحسب تصادف، انحرافات تصادفی با الگویی رخ می‌دهند که به‌نظر می‌رسد تیمار دارای اثربوده است. مثلاً مانند شکل ۶. این وظیفه آزمونهای معنی‌دار بودن است که برآورد کند چه موقع تفاوت قبل از تیمار و بعد از تیمار بیش از آن است که به‌حساب چنین انحرافات تصادفی گذاشته شود. در حالت انحرافات مستقل، مانندشکل ۶، فرمولهای آماری مناسب برای انجام چنین کاری به نحو احسن، به‌دست آمده‌اند.

شکل ۷ برمبنای همان خط راست قطری شکل ۶ بنانهاده شده‌است و انحرافاتی به همان اندازه به آن افزوده شده‌اند. ولی انحرافات دیگر مستقل نیستند. در عوض در هر نقطه چهار انحراف

شکل ۶ـ سری زمانی شبیه‌سازی شده با خطاهای مستقل. منبع: راس، کمپبل وگلاس (۱۹۷۰)



شکل ۷ـ سری زمانی با خطاهای همبسته (تأخیری). منبع: کمپبل (۱۹۶۹)



کوچکتر به‌صورت الگویی انباشته یا با تأخیر، اضافه شده‌است. هر انحراف جدید در یک دوره‌زمانی اضافه می‌شود و تا سه دوره زمانی بعدی باقی می‌ماند، درنتیجه هرنقطه در سه چنین

انحرافی با دوره زمانی بلافاصله قبل و در سه انحراف با دوره زمانی بلافاصله بعد مشترک است. این نقطه در دو انحراف با دوره‌های زمانی دو مرحله دورتر در هر جهت و در یک انحراف با دوره‌های زمانی سه مرحله دورتر مشترک است. انحرافات مربوط به دوره‌های زمانی چهار مرحله دورتر یا بیشتر، مستقلند. گرچه شکل ۷، مانند شکل ۶، یک خط راست است که به‌وسیله خطاهای تصادفی تغییر یافته‌است ولی توجه کنید که چقدر پویاتر و دوره‌ای‌تر به‌نظر می‌رسد. اینگونه انحرافات غیرمستقل هم قضاوت بصری را در مورد اثر و هم آزمونهای معنی‌دار بودن مبتنی بر فرض استقلال را، از این طریق که غالباً باعث قضاوت معنی‌دار بودن آماری اثر می‌شوند، گمراه می‌کند. (برای تأکید بر نبود انحراف واقعی و منظم تأکید می‌کنیم که اگر در هر مورد، شبیه‌سازی ۱۰۰۰ بار تکرار شود و از نتایج، میانگین‌گیری به‌عمل آید، هر میانگین کاملاً خط راست قطری را تقریب می‌کند. یعنی شکلهای ۶ و۷ یک خطی بودن زیربنایی را اعمال می‌کند که معمولاً نامناسب خواهدبود). آمارشناسان از راههای گوناگون

سعی دارند آزمونهای معنی‌دار بودن مناسبی برای سریهای زمانی مربوط به زندگی واقعی که روندهای کلی غیرخطی و انحرافات غیرمستقل دارند، بیابند. (احتمالاً مناسبترین آنها شیوه‌های *باکس وتیائو* ۱۹۷۵، *کوک وکمپبل* ۱۹۷۹، و *مک‌کلیری و هی* ۱۹۸۰ هستند.)

اثر برگشتی

ما از مسائل ساده استنباط به مسائل پیچیده‌تر رفته‌ایم. به‌زودی به مسائلی که فهم آنها ساده‌تر است برخواهیم گشت. ولی قبل از آن کوشش خواهیم کرد یک مشکل نهایی را که برحسب یک سنت آماری به اثر برگشتی معروف است بفهمیم. اگر بتوانیم مطمئن باشیم که تغییر سیاست مستقل از بالا و پایین رفتنهای موجود در دوره‌های زمانی به‌وقوع پیوسته است، جای نگرانی نیست. ولی اگر زمان تغییر سیاست فقط به این دلیل انتخاب شده باشد که یک مقدار کرانگین در دوره زمانی قبل رخ داده‌است، آنگاه یک اثر برگشتی برای توجیه وقوع مقادیر بعدیی که کمتر کرانگین‌اند کفایت می‌کند. برای اینکه بسنجیم آیا در مورد کنتیکات نیز یک اثر برگشتی در کار است، به شکل ۲ برمی‌گردیم. در اینجا ملاحظه می‌کنیم که بارزترین تغییر در تمام سری، افزایش ۱۹۵۴ تا ۱۹۵۵ است. با مطالعه روزنامه‌ها و اظهارات فرماندار، می‌توان گفت که این افزایش قابل توجه باعث شده‌است که وی آن سختگیری را اعمال نماید. بنابراین انجام اعمال تیمار به دلیل نقطه بسیار بالای ۱۹۵۵ بوده‌است.

در هر سری زمانی ناپایدار، بعد از هرنقطه زمانی که در آن انحراف کرانگینی از روند کلی حاصل می‌شود، نقاط بعدی، به‌طور متوسط، به روند کلی نزدیکتر خواهند بود. این ادعا را در مورد شکل ۶ امتحان کنید. چشم خود را از چپ به راست حرکت دهید و به هر نقطه که "تاکنون بیشترین" بوده‌است، توجه کنید. در این مورد از اکثر این‌گونه نقاط، نقطه بعدی پایینتر است، یعنی به‌طرف روند کلی برگشته است. این‌گونه برگشتها، پس از نقاطی که به دلیل کرانگین بودن انتخاب شده‌اند، یک ویژگی خودکار از هر واقعیت ناپایداری است و نباید مورد هیچگونه تفسیر علّی قرار گیرد. کاربرد این مطلب در مورد شکل ۲ چنین نتیجه می‌دهد که حتی بدون وجود هرگونه اثر واقعی از سخت‌گیری، انتظار داریم که نقطه ۱۹۵۶ پایینتر از نقطه کرانگین ۱۹۵۵ باشد.

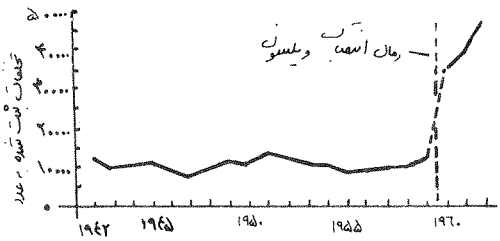
دلایل دیگر برای انتقال در سریهای زمانی

اهمیت نتیجه‌گیری از داده‌های سریهای زمانی مربوط به برنامه‌های مهم ملی، از جانب مسئولان اجرایی، قانونگذاران، رأی‌دهندگان و گروه‌های غیرآمارشناس دیگر، در حال افزایش است. به این دلیل به دو نکته دیگر که مستقیماً آماری نیستند، اشاره می‌کنیم: نخست توجه کنید که دلایل بسیار دیگری، به‌غیر از اعمال یک تغییر برنامه، برای به‌وجود آمدن انتقالهای ناگهانی در سریهای زمانی وجود دارند. یک دلیل بسیار گمراه‌کننده به‌وجود آمدن انتقالها در شیوه‌های ثبت اطلاعات است. این امکان وجود دارد که اینگونه انتقالها همزمان با تغییرات خط مشی رخ دهند. برای مثال یک تغییر عمده در سیستم پلیس شیکاگو، وقتی رخ داد که پروفسور *اورلندو ویلسون* از دانشگاه کالیفرنیا برای اصلاح فساد اداره پلیس آورده‌شد. شکل ۸ یک افزایش بارز در میزان سرقتها را که ظاهراً در اثر حضور وی بوده‌است، نشان می‌دهد. بعداً معلوم شد که این افزایش در اثر اصلاحاتی بود که وی در سیستم ثبت وقایع انجام داد و چنین افزایشی به همین دلیل پیش‌بینی شده‌بود.

در یک وضعیت واقعی، بی‌شباهت به یک آزمایشگاه منفرد، علل بسیار دیگری ممکن است همزمان با تغییر سیاست درکار باشند. بنابراین در کنتیکات یا در انگلستان، یک افت در تلفات رانندگی ممکن است ناشی از هوای بسیارخشک، تعداد کمتر خودروها در جاده‌ها یا به‌دلیل تمهیدات ایمنی جدید یا به‌دلیل تعداد زیادی از عوامل دیگر باشد. اگر می‌توانستیم آزمایشی با گروه شاهد تصادفی شده که در ابتدای این مقاله مورد بحث قرار گرفت، طراحی کنیم لزوم اینگونه توجیهات به‌طور آماری ازبین می‌رفت ولی برپا ساختن چنین آزمایشی در این دو وضعیت غیرممکن بوده‌است. در عوض باید به‌طریق دیگر اینگونه توجیهات را که رقبای توجیه اثر هستند، حذف کنیم. یک رهیافت سودمند این است که به مطالب روزنامه‌ها در مورد میزان بارندگی، تغییرات در شدت ترافیک و سایر علل ممکن انتقال نگاه کنیم.

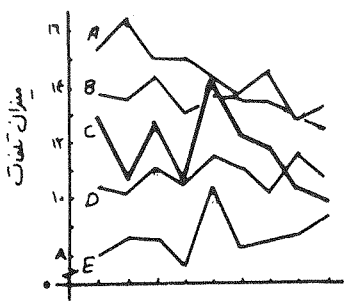
یک رهیافت دیگر این است که در جستجوی مقایسه‌های کنترلی باشیم که آثار اینگونه علل را، وقتی درکارند، در مواردی که تیمار اصلاحی موردنظر اعمال نشده‌است، نشان دهند. برای کنتیکات داده‌های چهار ایالت نزدیک، به‌صورتی که در شکل ۹ نشان داده شده مناسب هستند. همه این ایالتها باید تحت‌تأثیر تغییرات آب وهوا، تمهیدات ایمنی جدید در خودروها و نظایر آن

شکل ۸ـ سرقتهای زیر ۵۰ دلار گزارش‌شده در شیکاگو از ۱۹۴۲ تا ۱۹۶۲. منبع: کمپبل (۱۹۶۹). داده‌ها از گزارشهای یکنواخت جنایی ایالات متحده، ۱۹۶۲ - ۱۹۴۲



قرار گرفته‌باشند. گرچه این داده‌ها این معنی را تائید می‌کنند که تلفات کنتیکات در سال ۱۹۶۵ پایینتر از تلفات ۱۹۵۵ آن ایالت، حتی بدون درنظر گرفتن سخت‌گیری، بوده است ولی کاهش مستمر طی سالهای ۱۹۵۷ تا ۱۹۵۹ در کنتیکات بیشترین شیب را دارد و به‌نحوی ممکن است اثر واقعی ادامه سختگیری در این سالها باشد. گرچه به‌طور بصری مشکل زیادی در استفاده از این داده‌های تکمیلی نداریم ولی آمارشناس برای تلفیق همه آنها در یک آزمون معنی‌دار بودن مناسب با مشکلات واقعی زیادی مواجه است.

شکل ۹ـ تلفات رانندگی کانتیکات، نیویورک، نیوجرسی، رودآیلند و ماساچوست (برحسب ۱۰۰۰۰۰ نفر). منبع: کمپبل (۱۹۶۹)



A نیوجرسی
B نیویورک
C کنتیکات
D ماساچوست
E رودآیلند

برای انگلستان، کشورهای مناسبی برای مقایسه وجود نداشت ولی بارهای عمومی بریتانیا در ساعات قبل و حین اوقات رفت وآمد روزانه تعطیل هستند. بنابراین آمار تلفات مربوط به این ساعات نوعی پایه مقایسه فراهم می‌سازد (شکل ۱۰). متأسفانه مقدار زیادی ناپایداری در این داده‌ها وجود دارد. بنابراین نمی‌توان با اطمینان زیاد میزان ماندگاری اثر سختگیری اولیه را برآورد کرد.

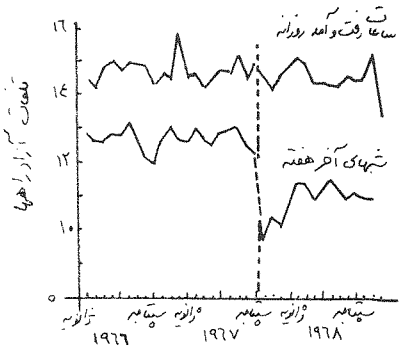
تذکرنهایی

از این مثالها و توضیحات کوتاه می‌توان درسهای کلی زیادی فراگرفت. نخستین درس متضمن تمایز بین آزمایشهای واقعی که در آن گروه‌های آزمایشی و شاهد به‌وسیله تصادفی‌کردن تشکیل می‌شوند، و شبه ـ آزمایشهاست.

آزمایشهای واقعی وقتی که امکان‌پذیر باشند، دارای قدرت ودقت استنباط بسیار بیشتری از شبه ـ آزمایشها هستند. مسئولان اجرایی نوآوریهای اجتماعی، با مشورت با آمارشناسان، باید از این طرحها، وقتی که امکان‌پذیرند، استفاده کنند.

وقتی که آزمایشهای واقعی امکان‌پذیر نباشند یا مورد استفاده قرارنگیرند، بعضی طرحهای

شکل ۱۰ـ مقایسه تلفات در طول ساعات بسته‌بودن بارها (ساعات رفت وآمد روزانه) و شبهای آخر هفته در سختگیری تحلیلگر بازدم انگلستان. منبع: راس، کمپبل وگلاس (۱۹۷۰)



شبه ـ آزمایشی، ازقبیل سریهای زمانی گسیخته، وجود دارند که می‌توانند در ارزیابی اثر تغییرات سیاستها بسیار سودمند واقع شوند. اینها نیز، برای اجتناب از نتیجه‌گیریهای گمراه‌کننده مستلزم استفاده از مهارتهای آماری هستند.ارزیابی نوآوریهای اجتماعی زمینه مهم و پرچالشی ازکاربرد آمار جدید است.

مسائل

۱ـ چرا طرحی که در این مقاله توجه برآن متمرکز شده‌است، طرح سریهای زمانی گسیخته نام دارد؟

۲ـ شکل ۲ چگونه برداشت شما را از شکل ۱ تغییر می‌دهد؟

۳ـ شکل ۳ را درنظر بگیرید. فرض کنید X کاهش نرخ برابری ارز است که منجر به افزایش قیمت بین O^4 و O^5 می‌شود. X چه‌اثری بر قیمت کالای A، کالای C و کالای F خواهد داشت؟

۴ـ شکل ۳ را درنظر بگیرید. چرامؤلف می‌گوید که می‌توان در کمال مشروعیت در A و B اثری را استنباط نمود، درحالی که هرگونه استنباط اثری در F و G و H کاملاً غیرموجه خواهد بود؟

۵ـ شکلهای ۴ و۵ را درنظر بگیرید. آیا سختگیری تحلیلگر بازدم در یکی از دو شکل نمایانتر است؟ اگر چنین است درکدامیک؟

۶ـ الف) تفاوت بین خطاهای مستقل و با تأخیر را توضیح دهید.

ب) چگونه ویژگیهای آثار دونوع خطا را در نمودارهای داده‌های شکلهای ۶ و۷ مشخص می‌کنید؟

۷ـ منظور مؤلف وقتی که می گوید شکل ۷ "پویاتر و دوره‌ای‌تر" از شکل ۶ است، چیست؟

۸ـ توضیح دهید که منظور مؤلف از اثر برگشتی چیست؟

۹ـ به شکل ۸ توجه کنید. افزایش سریع تعداد خلافها را وقتی که *اورلندو ویلسون* منصوب شد، توضیح دهید.

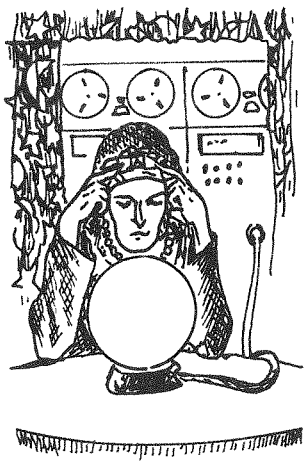
۱۰ـ آیا باید کاهش تلفات رانندگی در رودآیلندرا (شکل ۹) به سختگیری نسبت به سرعت غیرمجاز در ایالت مجاور یعنی کنتیکات نسبت داد؟ جواب خود را توجیه کنید.

۱۱ـ آزمایش واقعی چیست؟ شبه ـ آزمایش چیست؟

۱۲ـ درباره این حکم اظهارنظر کنید: شبه ـ آزمایشها برای ارزیابی تغییرات سیاستها سودمندند زیرا کاربرد طرحهای آزمایشی واقعی در مسائل اجتماعی واقعاً غیرممکن است.

منابع:

G. E. P. Box and G. C. Tiao. 1975. "Intervention Analysis With Applications to Economic and Environmental Problems." *Journal of the American Statistical Association 70:70-90.*



شب انتخابات در تلویزیون

ریچارد اف. لینک[۱] شرکت سهامی ریچارد اف. لینک و شرکا

در اولین سه‌شنبه ماه نوامبر سالهای زوج، شب هنگام، میلیونها نفر در سراسر آمریکا نمایشهای مربوط به انتخابات را که سه شبکه عمده تلویزیون تهیه کرده‌اند، تماشا می‌کنند. بینندگان جدول‌بندی خلاصه آرایی که برای تصدی پستهای اصلی سناتوری و فرمانداری به صندوقها ریخته شده‌است را می‌بینند و در سالهایی که رییس جمهور انتخاب می‌شود، جدول‌بندی

۱ـ *Richard F. Link*

خلاصه آرای ریاست جمهوری را در هر ایالت و کل کشور مشاهده می‌کنند. آنها جدول‌بندی آرای اعضای مجلس نمایندگان را نیز می‌بینند. معمولاً پس از آنکه درصد کمی از آرا گزارش شد و غالباً فقط چند دقیقه پس از پایان رأی‌گیری، نام برنده انتخابات را می‌شنوند. در ادامه شب، تحلیلهایی را ملاحظه می‌کنند که چگونگی پیروزی نامزدهای خاص را توضیح می‌دهد. یعنی شرح می‌دهد که نقاط قوت و ضعف هریک از نامزدها در کجا نهفته و چرا به‌نظر می‌رسد که برنده می‌شوند.

در پس این تلاش، دستگاه عظیمی در کار است. این دستگاه فیزیکی است. به این معنا که مستلزم یک شبکه ارتباطاتی بسیار پیچیده و استفاده وسیع از کامپیوتر است ولی در عین حال آماری و ریاضی نیز هست. یعنی مستلزم خلاصه سازیها و تعبیرهای سریع است تا یافته‌ها بتواند هرچه سریعتر به سمع و نظر توده بینندگان برسد.

در اینجا قصد نداریم که سازمان کامل ضروری برای تهیه برنامه شب انتخابات را شرح دهیم، بلکه به توصیف سه بخش اصلی از این برنامه می‌پردازیم که بشدت بر فناوری کامپیوتر و آمار تکیه دارد. یعنی جدول‌بندی آرا، پیش‌بینی برندگان، و تحلیل تفصیلی آرا. هر سه شبکه اساساً از سیستمی یکسان برای جدول‌بندی آرا استفاده می‌کنند ولی روشهای آنها برای پیش‌بینی برندگان و تحلیل آرا متفاوت است. در اینجا تنها به توصیف روش پیش‌بینی یک شبکه (یعنی ان بی سی)، که حداقل تا سال ۱۹۸۸ مورد استفاده قرار می‌گرفت، می‌پردازیم.

قبل از اینکه به بحث در مورد روشها و شیوه‌هایی بپردازیم که امروزه به‌کار می‌روند، نگاهی به تاریخچه کوتاه گزارشهای نتایج شب انتخابات می‌کنیم تا ببینیم که برنامه‌های امروزی چرا و چگونه فراهم آمده‌اند؟

تاریخچه کوتاهی از گزارش انتخابات

جامعه آمریکا، مانند سایر جوامع آزاد که انتخابات برگزار می‌کنند، همیشه علاقه شدیدی به نتایج انتخابات داشته‌است. این علاقه برای انتخابات ریاست جمهوری از همه بیشتر و برای انتخابات فرمانداری و سناتوری در حد معقولی زیاد است و حداقل تعداد جمهوریخواهان و دموکراتهایی که مجلس نمایندگان راتشکیل می‌دهند، مورد توجه مردم است. گرچه معمولاً

انتخاب یک عضو خاص از اهمیت ملی برخوردار نیست.

بنابراین، نتایج انتخابات همیشه از خبرهایی بوده که توجه و علاقه زیادی را به خود جلب کرده‌است. این خبرها تا سال ۱۹۲۸ از طریق روزنامه‌ها به مردم می‌رسید و پوشش آن به‌طور کلی نسبتاً کند و ناقص بود. رادیو وضعیت را عوض کرد و گزارش انتخابات سرعت گرفت. برای مثال، رادیو در اولین ساعات روز بعد از انتخابات در نوامبر سال ۱۹۴۸ پیروزی غیرمترقبه هری ترومن را اعلام کرد. تلویزیون در سال ۱۹۵۲ شروع به گزارش انتخابات در سطح کل کشور کرد و از آن به بعد دامنه و پوشش و سرعت خود را در جمع‌آوری آرا افزایش داده‌است. پوشش وسیع نتایج انتخابات مقدماتی از انتخابات ریاست جمهوری سال ۱۹۶۴ آغاز شد و تا به امروز همچنان از گزارشهای برجسته تلویزیونی حتی در «سالهای بدون انتخابات» است.

دو عامل بر سرعت پوشش خبری انتخابات اثر می‌گذارند: سرعت دریافت آرا از مبدأ (که اساساً حوزه انتخاباتی است) و گزارش سریع آن. سرعت گزارش‌دهی پس از جمع‌آوری آرا، به وسیله گزارش از طریق رادیو در مقایسه با گزارش از طریق روزنامه‌ها بشدت افزایش یافت. این سرعت گزارشگیری به‌وسیله تلویزیون افزایش ویژه‌ای نیافته است. رادیو و تلویزیون هردو قادرند اساساً گزارشهای لحظه‌ای ارائه نمایند ولی سرعت جمع‌آوری آرا توسط شبکه‌های تلویزیونی افزایش زیادی یافته‌است. مروری بر شیوه‌های جمع‌آوری آرا در گذشته و حال خالی از ارزش نیست.

آمریکا تقریباً ۱۷۵۰۰۰ حوزه انتخاباتی دارد. در دستگاه انتخاباتی رسمی معمولاً آرای حوزه انتخابیه به مرکز جمع‌آوری ناحیه و سپس به مرکز ایالتی و غالباً به دبیرخانه فرمانداری آنجا ارسال می‌شود که رأی‌گیری را رسماً تأیید می‌کند. جمع‌آوری نهایی آرا به‌طور رسمی و تأیید آن معمولاً چند ماه به طول می‌انجامد ولی آرای حوزه انتخاباتی با استفاده از تلفن یا پیک نسبتاً سریع به سطح ناحیه ارسال می‌شود و نتیجه رأی‌گیری در سطح ناحیه معمولاً به‌صورت غیررسمی بسرعت در اختیار قرار می‌گیرد. کار جمع‌آوری آرا در سطح ناحیه به‌مراتب از کار جمع‌آوری آرا در سطح حوزه انتخاباتی کم زحمت‌تراست. زیرا فقط در حدود ۳۰۰۰ ناحیه در کشور وجود دارد. با وجوداین اگر جمع‌آوری آرا در سطح حوزه انتخاباتی باشد سریعتر انجام خواهد شد و این همان بدعت اساسی است که تلویزیون در جمع‌آوری آرا به جای

گذاشته‌است. شبکه‌های تلویزیونی با منابع اقتصادی عظیم خود در ایجاد مکانیسمی برای جمع‌آوری آرا در سطح حوزه انتخاباتی و انتقال آن از طریق تلفن به یک جایگاه مرکزی که بتواند در آنجا به کمک کامپیوتر داده‌پردازی شود بسیار وسیله‌ساز بوده‌اند.

رقابت میان شبکه‌های تلویزیونی در زمینه جمع‌آوری آرا در سطح وسیع تا هنگام رأی‌گیریهای مقدماتی سال ۱۹۶۴ بسیار شدت یافت. در آن سال، رأی‌گیری مقدماتی در ایالت نیوهمپشایر شاهد جمع‌آوری و گزارش آرا در سطح حوزه انتخاباتی توسط هر سه شبکه تلویزیونی عمده بود. درواقع، به قول بعضی افراد شوخ طبع، طی رأی‌گیری مقدماتی در نیوهمپشایر در سال ۱۹۶۴ تعداد کارکنان تلویزیون بیشتر از رأی‌دهندگان بوده‌است یا به عبارت دیگر اگر تمام رأی‌دهندگان نیوهمپشایر را برای رأی دادن به یک محل مرکزی در نیویورک انتقال می‌دادند، ارزانتر از جمع‌آوری آرا در نیوهمپشایر تمام می‌شد. نیازی به گفتن نیست که این اظهارنظرها اغراق‌آمیز است ولی تأکیدی بر هزینه گزافی می‌باشد که دربرداشته است. رقابت در جمع‌آوری آرا در بهار آن سال شدت بیشتری گرفت و با گزارش آرای مقدماتی انتخابات ریاست جمهوری کالیفرنیا که هریک از شبکه‌های سه‌گانه برای بیش از ۳۰۰۰۰ حوزه انتخاباتی آن جمع‌آوری کرده‌بودند به اوج خود رسید. این هزینه گزاف آمیزه‌ای از خوبیها را به همراه آورد. سرویسهای خبری روزنامه‌ها نیز همچنان به شیوه سنتی به جمع‌آوری و گزارش آرا از روی عملکردهای کامل نواحی مشغول بودند به‌طوری که در روز بعد از انتخابات، پس از آنکه شبکه‌های تلویزیونی گلدواتر را برنده اعلام کردند، روزنامه‌ها همگی *راکفلر* را با فاصله زیادی برتر از رقیب نشان دادند. علت این اختلاف آن بود که گزارش کامل شمارش ناحیه لوس‌آنجلس که تقریباً یک سوم از حوزه‌های انتخاباتی ایالت را شامل می‌شود، دیرتر از مهلت تعیین شده برای روزنامه‌ها دریافت شد و گلدواتر در لوس‌آنجلس با قدرت پیشی گرفت.

این بی‌نظمی با آشفتگی دیگری ناشی از این واقعیت همراه شد که هر شبکه در هر لحظه مجموع آرای خود را گزارش می‌داد و چون جمع‌آوری آرا به‌طور مستقل انجام می‌شد، مجموعها در هر لحظه کاملاً متفاوت بودند. همه اینها به تشکیل سازمانی به نام «سرویس خبری

انتخابات[۱]» انجامید که تنها هدف آن جمع‌آوری آرا و گزارش آن به اعضای سازمان است. این سرویس با تلاش مشترک شبکه‌های سه‌گانه تلویزیونی (ای بی سی، سی بی اس و ان بی سی) و دو خبرگزاری (ای پی و یو پی آی) تشکیل شد. سرویس خبری انتخابات ارقام سود را به‌طور همزمان در اختیار اعضای خود قرار می‌دهد به‌طوری که تمام شبکه‌ها و سرویسهای خبری می‌توانند هرلحظه داده‌های اساسی یکسانی را به اطلاع عموم برسانند.

سرویس انتخاباتی جدید

عملیات وسیع سرویس خبری انتخابات به روش زیر انجام می‌پذیرد: گزارشگران که خبرنگار منطقه‌ای نامیده می‌شوند در بیش از ۱۰۰۰۰۰ حوزه بزرگتر، ۱۷۵۰۰۰ حوزه انتخاباتی کشور و در هریک از ۳۰۰۰ مرکز گزارش‌دهی نواحی به انجام وظیفه مشغولند. این گزارشگران آرا را در سطوح حوزه انتخاباتی و ناحیه جمع‌آوری و سپس نتیجه را با افزودن اطلاعات کافی به منظور شناسایی منشاء گزارش به محل مرکزی خبر می‌دهند. سپس اطلاعات رأی‌گیری وارد کامپیوتری می‌شود که اعتبار ظاهری آن را بررسی می‌کند. برای مثال، اگر اطلاعات مربوط به گزارش حوزه انتخاباتی است، بررسی می‌کند که از تعداد ثبت‌نام شده آن حوزه تجاوز نکند، یا اگر گزارش یک ناحیه است، تعداد حوزه‌های انتخاباتی آن ناحیه بیشتر نشده باشد. چون ارقام ثبت نام و داده‌های مربوط به تعداد حوزه‌های انتخاباتی در این هنگام کاملاً دقیق نیستند، این بررسی به‌جای اعلام نتیجه مطلق به تحمل آماری بستگی دارد. پس از اینکه گزارش بررسی شد، اگر گزارش یک حوزه انتخاباتی باشد به نتایج حوزه که قبلاً برای آن ناحیه گزارش شده‌است، اضافه می شود و اگر گزارش یک بخش باشد؛ جایگزین گزارش قبلی آن ناحیه می‌شود (گزارشهای نواحی برمبنای تجمعی تهیه می‌شوند). کامپیوتر در فاصله‌های زمانی منظم یک گزارش رأی‌گیری برای هر رقابت انتخاباتی مربوط به هریک از نواحی ایالت تولید می‌کند و جمع آرا را در سطح ایالت برای رقابتهای انتخاباتی ریاست جمهوری، سناتوری و فرمانداری و همچنین سرجمع آرای رقابت انتخاباتی ریاست جمهوری را در سطح کشور ارائه می‌نماید.

۱ـ *News Election Service = NES*

خلاصه گزارش با مقایسه آرای بخش که از روی گزارشهای بخش به‌دست آمده است و آرای بخش که از روی گزارشهای حوزه انتخاباتی محاسبه شده‌است، تهیه می‌شود. در این خلاصه گزارش از رقم بزرگتر به‌جای رقم ناحیه استفاده می‌شود و سپس با جمع زدن ارقام مربوط به نواحی هر ایالت رقم آرای آن ایالت به‌دست می‌آید. در سالهایی که انتخابات ریاست جمهوری برگزار می‌شود، مجموعیابی اضافی دیگری در سطح ایالتها انجام می‌پذیرد تا رقم آرا را در سطح کشور برای ریاست جمهوری به‌دست دهد. علاوه بر ارائه خلاصه مجموعهای آرا، همان‌طور که نسبت حوزه‌های انتخاباتی گزارش‌دهنده ارائه می‌شود، درصدهایی را برای هر نامزد انتخابات نیز گزارش می‌دهند. خلاصه گزارشهای مشابه برای حوزه‌های انتخاباتی در سطح نواحی انتخابات کنگره گروه‌بندی می‌شوند تا رقابتهای انتخاباتی کنگره ارزشیابی شود.

اطلاعات پس از محاسبه توسط کامپیوتر در اختیار متقاضیان قرار می‌گیرد. کامپیوتر این اطلاعات را به‌صورت چاپی هم در محل استقرار کامپیوتر و هم در استودیوی تلویزیون ارائه می‌نماید و از طریق خطوط تلفنی هم که می‌توان از آن برای انتقال اطلاعات به سیستمهای کامپیوتری شبکه‌ای گوناگون استفاده کرد، اطلاعات را در دسترس قرار می‌دهد.

قبل از انتخابات باید آماده‌سازی گسترده‌ای به عمل آید تا نه‌تنها خبرنگاران آموزشهای لازم رابینند بلکه ارقام ثبت‌نام برای حوزه‌های انتخاباتی جمع‌آوری شوند و تعداد حوزه‌های انتخاباتی هر ناحیه به‌دست آید، و سایر داده‌های پایه‌ای گردآوری شوند. این اطلاعات برای فرایند بررسی و گزارش‌دهی دقیق از نسبت آرای شمارش شده ضروری است.

آمادگیهای بیشتر برای عملیات واردکردن آرا به کامپیوتر و در برنامه‌ریزی کامپیوتر برای دریافت آرا، جمع زدن مناسب و گزارش صحیح آن لازم است. نام عمومی عملیاتی که کامپیوتر را آماده می‌سازد تا به‌طرز مناسبی کار کند، کدگذاری خوانده می‌شود و در اهمیت آن هرچه گفته شود کم است. در سال ۱۹۶۸ اشتباهاتی که در دستورالعملهای کامپیوتر وجود داشت باعث شد که کامپیوتر بد کار کند و جریان وسیع آرای دریافتی از سرویس خبری انتخابات اندکی پس از نیمه‌شب به وقت شرق آمریکا در شب انتخابات قطع و تقریباً به هیچ تبدیل شود. این بدکار کردن کامپیوتر تا حدودی باعث عدم حتمیت درباره برنده رقابت انتخاباتی ریاست جمهوری بود که تا صبح روز چهارشنبه از طریق شبکه‌های تلویزیونی اعلام نشد.

این مطلب ما را به جایی می‌برد که هنوز هم موضوع رقابت شدید بین شبکه‌های سه‌گانه است. با اینکه شبکه‌های مزبور همه به استفاده مشترک خود از مجموعهای آرای سرویس خبری انتخابات برای گزارش مجموعهای همانند آرا مقیدند، در تفسیر بین مجموعها محدودیتی ندارند. زیرا گرچه مجموع آرا در هرلحظه می‌تواند آگاهی‌دهنده باشد، ولی توجه واقعی در هر انتخاباتی معطوف به این است که چه کسی، با چه مقدار رأی و چرابرنده می‌شود.

پیش‌بینی برندگان انتخابات

جمع‌آوری آرا و گزارش‌دهی سریع آن مستلزم سازماندهی گسترده، امکانات کامپیوتری وسیع و تجهیزات ارتباطی است. با وجود این، همه این فعالیتها صرفاً برای تعیین مجموع آرا است. سئوالی که برای پیش‌بینی کننده مطرح می‌شود همیشه چنین است: کی می‌توانم درحد معقول مطمئن شوم آرایی که جدول‌بندی کرده‌ام برای تعیین برنده نهایی کافی است؟

آسانترین جواب به سئوال مذکور این است که: صبرکن تا همه آرا شمارش شوند. ولی این کار ممکن است چندین روز طول بکشد اما نظریه آمارگاهی اوقات این امکان را به ما می‌دهد که زودتر جواب بدهیم. گاهی هم به ما امکان می‌دهد هنگامی که تنها ۰/۰۱ آرا به تحلیلگر گزارش شده‌است، برنده انتخابات را مشخص کنیم. فراوان پیش آمده است که چون هنوز سرویس خبری انتخابات نتوانسته مجموع آرا را (که طبق توافق تنها اطلاعاتی است که می‌تواند در اختیار عموم قرارگیرد) ارائه نماید، پیش از آنکه حتی یک رأی به نظر بینندگان تلویزیون رسیده‌باشد، بر مبنای اطلاعاتی که توسط شبکه جمع‌آوری شده و در استودیوی تلویزیون در اختیار تحلیلگران قرار گرفته است، بتوان پیش‌بینیهایی ارائه کرد.

پیش‌بینی برندگان در شب انتخابات نیازمند ترکیبی از اطلاعات تاریخی، نظریه آمار برای ایجاد مدل ریاضی مناسب رأی‌گیری و تصمیم‌گیری درباره اینکه چه وقت اطمینان کافی برای پیش‌بینی حاصل شده است و میزان واقعی آرا در شب انتخابات است. شبکه‌ها الگوهای متفاوتی برای پیش‌بینی برندگان دارند ولی همه این الگوها دارای همان عناصر اساسی است که شرح داده شد. اکنون نگاهی کنیم به الگویی کلی که یکی از شبکه‌ها، یعنی ان بی سی به‌کار می‌برد. با پیش‌بینی برنده در یک رقابت انتخاباتی ایالتی، شروع می‌کنیم. سپس به پیش‌بینی

برنده در انتخابات ریاست جمهوری و بالاخره به پیش‌بینی ترکیب مجلس نمایندگان می‌پردازیم.

رقابتهای انتخاباتی ایالتی

اطلاعات برای پیش‌بینی برنده رقابت انتخاباتی ایالتی از چهار منبع جداگانه حاصل می‌شود. اول، برآورد درصدی آرایی که هریک از نامزدهای انتخاباتی قبل از انتخابات از نظرسنجیهای عمومی، گزارشگران روزنامه‌ها، سیاستمداران ایالت و منابع مشابه به‌دست می‌آورند. این برآورد اولیه که غالباً نسبتاً دقیق است، می‌تواند به‌طور قطع دال بر آن باشد که نتیجه رقابت انتخاباتی چه خواهد شد.

دوم، مصاحبه با رأی‌دهندگان هنگامی که محلهای مخصوص رأی‌گیری را ترک می‌کنند. نظرسنجی از رأی‌دهندگان در روز انتخابات را تأمین می‌کند. پاسخگویان خود پرسشنامه‌ها را پر می‌کنند. به قسمی که اجراکننده نظرسنجی به جوابها پی نمی‌برد.

سوم، شبکه آرای حوزه‌های منتخب خاصی را که حوزه‌های انتخاباتی کلیدی خوانده می‌شوند علاوه بر آرای جمع‌آوری شده توسط سرویس خبری انتخابات گردآوری می‌کند. شبکه نوعاً آرای ۵۰ تا ۱۵۰ حوزه انتخاباتی از اینگونه را برای هر ایالت جمع‌آوری می‌کند. به این ترتیب یک شبکه ممکن است دارای یک سیستم جمع‌آوری آرا از حوزه‌های انتخاباتی در سطح کشور با استفاده از گزارشگرانی در متجاوز از ۵۰۰۰ حوزه باشد که کاملاً مستقل از تلاشهای سرویس خبری انتخاباتند. قبلاً رفتار رأی‌دهی این حوزه‌های انتخاباتی کلیدی در انتخابات گذشته بدقت تحلیل شده‌است.

چهارم، اطلاعات سرویس خبری انتخابات که در سطح ناحیه موجود است.

اطلاعات حاصل از این منابع به‌ترتیب زمان تنظیم می‌شوند. واضح است که ابتدا برآورد اولیه در اختیار قرار می‌گیرد. زیرا قبل از روز انتخابات ارائه می‌شود. نتایج به‌دست آمده از پاسخگویان به «نظرخواهی از رأی‌دهندگان در روزانتخابات» در طول روز تلفنی اطلاع داده و با کمک کامپیوتر پردازش خواهد شد. به‌طوری که نتایج تقریبی برای یک رقابت انتخاباتی خاص در یک ایالت خاص زودتر از پایان نظرخواهی در آن ایالت مشخص خواهد شد. آرای

حوزه‌های انتخاباتی کلیدی که شبکه‌ها رأساً جمع‌آوری می‌کنند معمولاً اولین اطلاعات واقعی در مورد رای‌گیری است که در اختیار شبکه قرار می‌گیرد. براساس توافق به عمل آمده برای تشکیل سرویس خبری انتخابات، این اطلاعات مربوط به رأی‌گیری حوزه‌های انتخاباتی کلیدی را نمی‌توان از رسانه‌های رادیویی و تلویزیونی پخش کرد ولی می‌توان از آن برای پیش‌بینی برندگان استفاده نمود. اگر در ایالتی رقابت انتخاباتی یکسویه باشد، نتایج نظرخواهی روز انتخابات از رأی‌دهندگان می‌تواند خود به‌تنهایی برای اعلام اختتام رقابت انتخاباتی به محض پایان یافتن نظرسنجی مورد استفاده قرارگیرد. اگر مسابقه انتخاباتی چندان یکسویه نباشد ولی نتیجه آن نسبتاً روشن باشد، می‌توان پس از گزارش گرفتن از چند حوزه انتخاباتی کلیدی، نتیجه را تلفنی اعلام نمود.

اگر رقابت تنگاتنگی در جریان باشد، قبل از آنکه بتوان برنده‌ای را پیش‌بینی نمود، باید تعداد بیشتری از حوزه‌های انتخاباتی ویژه رأی‌گیری شوند و غالباً استفاده از اطلاعات ناحیه که از طریق سرویس خبری انتخابات دراختیار قرار می‌گیرند ضرورت خواهد داشت.

برای استفاده از اطلاعات دریافتی از نواحی، تهیه یک مدل ریاضی ضروری است. علت آن است که نواحی مختلف رفتار رأی‌دهی متفاوتی دارند. برای مثال، شهر نیویورک در آرای خود همیشه از سایر قسمتهای ایالت نیویورک دموکرات‌تراست. این نوع تفاوت در رفتار رأی‌دهی از نظرگرایشهای نسبی دموکرات یا جمهوریخواه را می‌توان در یک مدل ریاضی منظور نمود.

در مدل آماری از الگوهای رأی‌گیری گذشته نزدیک استفاده می‌شود. برای مثال، نسبت آرای ایالت نیویورک در شهر نیویورک نوعاً ۰/۴ و در سایر قسمتهای ایالت ۰/۶ است. در یک انتخابات نوعی فرماندار درگذشته نامزد انتخاباتی دموکراتها ۵۰ درصد از آرای شهر نیویورک و ۴۰ درصد آرای سایر قسمتهای آن ایالت را به خود اختصاص داد. بنابراین، میزان آرایی که در سطح ایالت به‌دست آورد، برابر می‌شد با: ۴۴ ٪ = (۴۰ ٪)۰/۶+(۵۰ ٪)۰/۴ به این ترتیب، شهر نیویورک ۶درصد بیشتر از متوسط ایالت و بقیه قسمتهای ایالت ۴درصد کمتر از متوسط کل ایالت طرفدار دموکراتها محسوب می‌شد. این نسبتها را می‌توان در یک مدل منظور نمود؛ به صورتی که اگر در همین ساده‌ترین مورد، نتایج انتخابات تازه‌ای از شهر نیویورک ۵۴ درصد به سود نامزد دموکراتها و از سایر قسمتهای ایالت ۴۸ درصد به سود نامزد دموکراتها باشد

پیش‌بینی آرا در سطح ایالت عبارت خواهدبوداز: ۰/۴(٪ ۵۴) + ۰/۶(٪ ۴۸) = ٪۵۰/۴ این رقم حاکی از آن است که نامزد دموکراتها برنده خواهدشد. هرچند که اگر نتایج بسیار زودهنگام باشند، این پیش‌بینی نمی‌تواند برای اعلام برنده به اندازه کافی دقیق تلقی شود.

عامل دیگری که در مدل آماری مدنظر قرار می‌گیرد، این است که آیا کسری از آرا که به قسمتهای مختلف ایالت اختصاص یافته است در مورد این انتخابات نیز صحت دارد یانه؟ اگر در قسمتهای بالای ایالت نیویورک برف سنگینی باریده باشد ولی در شهر نیویورک خبری از برف نباشد و درنتیجه در آرای قسمتهای فوقانی ایالت کاهشی ایجاد شود که در آرای شهر نیویورک نباشد، کسر آرا در انتخابات ممکن است ۰/۵ برای شهر نیویورک و ۰/۵ برای سایر قسمتهای ایالت باشد. بدین معنی که تعداد نسبی رأی‌دهندگان در شهر نیویورک بیش از حد عادی خواهد بود. در این صورت، پیش‌بینی به این صورت خواهد بود:

۰/۵ (٪ ۵۴) + ۰/۵ (٪۴۸) = ٪ ۵۱

و این حاکی از شانس بیشتری برای پیروزی دموکراتهاست. به این ترتیب باید برای پیش‌بینی نتایج رأی‌گیری، تعداد تفاضلی شرکت‌کنندگان نیز در مدل منظور شود.

استفاده از کامپیوتر این امکان را فراهم می‌سازد که چنین مدلی با استفاده از اطلاعات تفصیلی برای تمام بخشهای یک ایالت به جای دو منطقه‌ای که در مثال ذکر شد، ایجاد شود تا نه فقط یک پیش‌بینی بلکه نشانه‌ای از میزان دقت این پیش‌بینی را نیز ارائه نماید؛ به نحوی که بتوان در مورد زمان اعلام یک پیش‌بینی مطمئن نیز تصمیم‌گیری کرد.

گنجانیدن برآورد مقدماتی که در اختیار شبکه قرار دارد و نتایج حوزه‌های انتخاباتی کلیدی ویژه در مدل نیز برای استفاده مؤثر از تمام اطلاعاتی که در اختیار شبکه قرار داده شده سودمند است. چنین مدلی گاهی اوقات این امکان را فراهم می‌سازد که نتایج یک رقابت انتخاباتی با قطعیت قریب به‌یقین اعلام شود. حتی اگر تنها کسر کوچکی از آرا گزارش شده و رقابت نسبتاً تنگاتنگ باشد.

سیاست شبکه آن است که برنده راپیش‌بینی نکند مگر اینکه تقریباً مسلم شده باشد که برنده پیش‌بینی شده واقعاً برنده خواهدشد. میزان درستی پیش‌بینیها را می‌توان از روی این واقعیت اندازه‌گیری کرد که در طول شب معینی که بیش از صد پیش‌بینی می‌توان انجام داد،

معمولاً حداکثر یکی می‌تواند خطا باشد. استفاده از این قبیل مدلها که با نظریه آماری تهیه شده‌است به شبکه‌ها امکان می‌دهد که سیاست خود را اعمال کنند و در عین حال رقابت انتخاباتی را تنگاتنگ معرفی نمایند. زیرا همیشه دقت برآوردهای تهیه شده با استفاده از مدلها مسلم است. یکی از مهمترین بازده‌های مدل آماری در این مشکل تصمیم‌گیری مانند بسیاری موارد دیگر عبارت از برآورد دقت نتایج است.

رقابت انتخاباتی ریاست جمهوری

نظرخواهی روز انتخابات از رأی‌دهندگان، انقلابی در روش شبکه‌ها برای گزارش نتایج رقابتهای ریاست جمهوری به وجودآورد.این نظرخواهی ابزاری اساسی بود که به شبکه ان بی سی امکان داد تا در ساعت ۸/۱۵ شب (به وقت شرق آمریکا) چهارم نوامبر سال ۱۹۸۰ اعلام نماید که *رونالد* ریگان به‌عنوان چهلمین رییس جمهور ایالات متحده آمریکا انتخاب شده‌است. ان بی سی با استفاده از نظرخواهی در روز انتخابات از رأی‌دهندگان توانست تعیین کند که ریگان ۲۷۰ کرسی از آرای ایالتهایی که مکانهای رأی‌گیری آنها تا ساعت ۸ بعدازظهر (به وقت شرق آمریکا) بسته می‌شوند به‌دست می‌آورد. در سال ۱۹۸۴ همه شبکه‌هاتوانستند چنین پیش‌بینیهای زودهنگامی را اعلام کنند.

چون بسیاری از مردم در غرب آمریکا هنوز آرای خود رابرای انتخاب رییس جمهور و سایر نامزدها به این زودی یعنی ساعت ۵ بعدازظهر به‌وقت اقیانوس آرام به صندوقها نریخته بودند، این فناوری نوظهور مناقشه‌ای را برانگیخت. راه حلهای سیاسی گوناگونی پیشنهاد شده‌است از قبیل اینکه در سراسر کشور زمان پایان گرفتن رأی‌گیری یکسان باشد. راه حلهای دیگری که مستلزم سانسور اخبار است مشکلات جدی اساسنامه‌ای برانگیخته است. یک راه حل پیشنهادی این مشکل غیرعادی مضحک، این است که تازمان تشکیل جلسه هیئت انتخاب‌کننده رییس جمهور در ژانویه سال بعد نتایج انتخابات مخفی بماند!

هنگامی که رقابتهای انتخاباتی ریاست جمهوری چندان به نفع یک نفر نباشد، بازهم ممکن است تا ساعات اولیه صبح بعد از روز انتخابات نیز نتایج مشخص نشود. در سال ۱۹۶۸ کامپیوترهای سرویس خبری انتخابات مدتی از کار افتاد و چون نتایج این‌انتخابات بسیار

تنگاتنگ بود، نتیجه نهایی تا ساعتها بعد که شبکه‌ها توانستند اطلاعات کافی برای پیش‌بینی مسئولانه برنده را جمع‌آوری کنند، معلوم نشد. در این انتخابات حتی اگر فناوری «نظرخواهی روز انتخابات از رأی‌دهندگان» نیز فراهم می‌بود، باز هم پیش‌بینی برنده با تأخیر انجام می‌پذیرفت.

رقابتهای انتخاباتی مجلس

پیش‌بینی ترکیب مجلس نمایندگان نیاز به یک مدل شبیه به مدلی دارد که برای پیش‌بینی نتایج رقابتهای ریاست جمهوری به‌کار می‌رود. تنها تفاوت عمده در این است که هر کرسی به‌عنوان یک رأی حساب می‌شود. برآوردهای اولیه به‌طور معمول کمتر از برآوردهای اولیه ایالتها در انتخابات ریاست جمهوری قابل اعتمادند و گزارش آرا فقط توسط ناحیه انتخاباتی مجلس ارائه می‌شود. شبکه‌ها علاوه بر پیش‌بینی و اطلاع‌رسانی در مورد آرا، تحلیلی نیز از رأی‌گیری انجام می‌دهند. این مطلب، موضوع بعدی است که بررسی خواهیم کرد.

تحلیل اخبار

پرسشنامه «نظرخواهی روز انتخابات از رأی‌دهندگان» درباره مواردی که ممکن است نامزد انتخاباتی در طول مبارزات انتخاباتی خود مورد بحث قرار داده باشد، درباره ویژگیهای شخصی که پرسشنامه را پر می‌کند، از قبیل سن، جنس، نژاد، مذهب، درآمد و امثال آن و نیز درباره کسی که ممکن است پاسخگو به او رأی بدهد سئوالاتی دارد. با استفاده از جدول‌بندی متقاطع می‌توان آرا را برحسب عوامل جمعیت‌شناختی گوناگون (مانند جنس، نژاد و از این قبیل) و مطالب متنوع تحلیل نمود. این تحلیلها به توضیح این نکته که انتخابات چه مفهومی برای انتخاب‌کنندگان داشته‌است، کمک می‌کند. برای مثال، علی‌رغم انتخاب نامزدی که مخالف سقط جنین بوده، ممکن است اکثریت رأی‌دهندگان طرفدار آزادی انتخاب باشند. این بدان معناست که سایر عوامل در انتخاب این نامزد نقش مهمتری ایفاکرده‌اند و این نامزد نمی‌تواند بدرستی ادعا کند که موضع او در مورد سقط جنین خواسته مردم بوده‌است. اینگونه تحلیلها مفسران شبکه را قادر می‌سازد تا عقاید وبینشهای کیفی خود را با اطلاعات کمی نشان دهند و به

این ترتیب دیدگاهی از متن انتخابات به بینندگان خود ارائه نمایند که آگاهانه‌تر از دیدگاهی باشد که در غیر این صورت ممکن بود، داشته‌باشند.

نتیجه

تلاش در جهت ارسال گزارش از شبکه‌های تلویزیونی یک زمینه فعالیتی را معرفی می‌نماید که بدون کامپیوتر و بدون آمار نوین نمی‌توانست وجود داشته‌باشد. این فعالیت ترکیبی از فناوری جدید و مهارتهای سنتی گزارشگر را به نمایش می‌گذارد.

فنون آماری برای پیش‌بینی آرا ممکن است کاربردهای دیگری نیز داشته باشد. برای مثال، ممکن است بتوان با روشهای مشابه، از روی گزارشهای تاریخی، الگوی محصولات غله را بخش به بخش در ایالت آیوا تعیین نمود و محصول ایالت را از روی محصول تنها چند بخش که زودتر محصول خود را برداشت می‌کنند بدرستی برآورد کرد.

مسائل

۱ـ قسمتهای سه‌گانه برنامه تلویزیونی شب انتخابات که بیشتر از همه بر فناوری کامپیوتر و آمار تکیه دارد، کدامند؟

۲ـ مزایای جمع‌آوری آرا در سطح حوزه انتخاباتی توسط رسانه‌های جمعی کدامند؟ معایب آن کدامند؟

۳ـ چرا سرویس خبری انتخابات تشکیل شد؟ آیا تشکیل این سرویس بدان معناست که فقط داده‌های این سرویس خبری در اختیار شبکه‌ها قرار می‌گیرد؟

۴ـ نظریه آماری از دو راه در پیش‌بینی برنده دخالت دارد. آن دو راه را شرح دهید.

۵ـ حوزه‌های انتخاباتی کلیدی کدامند؟ آیا این حوزه‌ها برای تمام شبکه‌ها یکسانند؟

۶ـ فرض براین است که یک نامزد انتخابات فرمانداری نیویورک ۶۰ درصد آرای شهر نیویورک و ۵۰ درصد آرای بقیه قسمتهای ایالت را به‌خود اختصاص می‌دهد. همان‌طور که در این مقاله بیان شد، آرای شهر نیویورک معمولاً ۴۰ درصد کل آرا در سراسر ایالت است. نامزد انتخاباتی ما انتظار دریافت چند درصد از کل آرا را می‌تواند داشته باشد؟

۷ـ نامزد انتخاباتی ما ناامید است. در اولین سه‌شنبه ماه نوامبر ناگهان کولاک شدیدی شهر نیویورک را فراگرفت و تعداد کسانی که برای دادن رأی مراجعه کردند به ۳۰ درصد کل رأی‌دهندگان ایالت کاهش یافت. آیا این نامزد انتخاباتی هنوز هم ممکن است برنده شود؟

۸ـ چرا دقت برآورد برای پیش‌بینی برنده مهم است؟

۹ـ علاوه بر درصدهای برآوردشده رأی گیری، کدامیک از بازده‌های مدلهای پیش‌بینی مورد بحث به همان اندازه اهمیت دارد؟

بخش سوم

دنیای اجتماعی ما

برقراری ارتباط با دیگران

تعیین نویسنده ۱۷۵

یادآوری اطلاعات تصویری در کودکان ۱۹۳

معنی واژه‌ها ۲۰۳

اندازه چیزها ۲۱۷

مردم در عرصه‌های شغلی

حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟ ۲۳۱

ارزیابی مقدماتی از یک فراورده غذایی جدید ۲۴۵

تعیین تحملات عددی رنگ با روشهای آماری ۲۵۷

مردم، مدرسه و بازی

منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار ۲۶۷

آمار، ورزش و چند چیز دیگر ۲۸۳

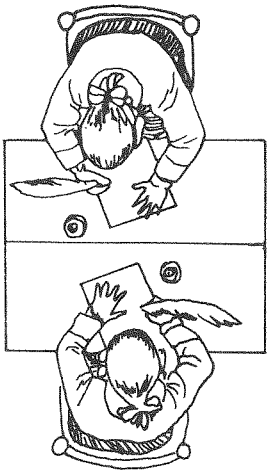
شمارش مردم و اموال آنان

شاخص بهای کالاها و خدمات مصرفی ۳۰۱

چگونه بهتر بشماریم؟ استفاده از آمار برای اصلاح سرشماری ۳۱۵

برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟ ۳۳۱

بسط و تحلیل شاخصهای اقتصادی ۳۴۳



تعیین نویسنده

فردریک موستلر[۱] دانشگاه هاروارد
دیوید ال. والاس[۲] دانشگاه شیکاگو

هنر، موسیقی، ادبیات و علوم اجتماعی، زیست‌شناسی و فیزیک از نیازی مشترک به طبقه‌بندی مطالب برخوردار هستند: کدام هنرمند این تابلو رانقاشی کرده است؟ چه کسی این قطعه (موسیقی) را ساخته‌است؟ نویسنده این مدارک چه کسی است؟ اگر زندانی با اخذ تعهد آزاد شود جرم را تکرار نخواهد کرد؟ مریض از چه بیماری رنج می‌برد؟ چه عنصر شیمیایی به فرایند

۱- *Frederick Mosteller*

۲- *David L. Wallace*

لطمه زده‌است؟ در حوزه علم آمار سئوالاتی از این دست به عنوان مسائل طبقه‌بندی یا متمایزسازی خوانده می‌شود.

از سئوالاتی که کراراً عنوان می‌شود و پاره‌ای اوقات حایز اهمیت است، سئوال در مورد نویسنده است. بیشتر مردم مطالبی درباره مناقشه شکسپیر ـ بیکن ـ *مارلو* شنیده‌اند. راجع‌به اینکه نویسنده این نمایشنامه‌های عالی، که معمولاً به شکسپیر نسبت داده می‌شود، کیست؟ سئوالاتی کمتر معروف ولی بدقت بررسی شده مربوط است به نویسنده تعدادی نوشته‌های مذهبی مسیحی معروف به پولسیها[1] که پاره‌ای از آنها کتابهایی است که در کتاب مقدس عهد جدید وجود دارد: کدامیک توسط *پولس* و کدامیک توسط سایرین نگاشته شده‌است؟ در بسیاری سئوالات پیرامون نویسنده، یافتن پاسخ، چنانچه به شمارش قاعده‌مند بپردازیم مشکل نخواهد بود. در اینجا به بررسی یک مسئله بویژه پیچیده از تاریخ آمریکا می‌پردازیم: مجادله درباره نویسنده ۱۲ مقاله مندرج در فدرالیست که به دونفر؛ *آلکساندر هامیلتون* و جیمز مدیسون نسبت داده می‌شود و نشان می‌دهیم که چگونه یک تحلیل آماری می‌تواند به حل سئوالات تاریخی کمک نماید.

مقاله‌های فدرالیست با نامهای مستعار، توسط *آلکساندر هامیلتون*، *جان جی* و جیمز مدیسون طی سالهای ۱۷۸۷ تا ۱۷۸۸ در جهت متقاعد ساختن شهروندان ایالت نیویورک برای تأیید قانون اساسی نشر یافته‌است. هفتاد و هفت مقاله به‌صورت نامه‌های وارده و با اسم مستعار «پابلیوس» در روزنامه‌های نیویورک منتشر شد. این مقاله‌ها همراه با هشت رساله دیگر به شکل یک کتاب درسال ۱۷۸۸ منتشر و پس از آن بارها چه در آمریکا و چه در خارج از آن تجدید چاپ شده‌اند. کتاب فدرالیست امروزه به‌عنوان اثری مهم در زمینه فلسفه سیاسی برجای مانده است. این اثر همچنین به‌عنوان مهمترین منبع اطلاعاتی برای بررسی نیت و هدف بنیانگذاران قانون اساسی به‌کار می‌رود. (برای مثال در تصمیم‌گیری پیرامون تسهیم نمایندگان مجمع قانونگذاری) زیرا مدیسون در مجمع قانون اساسی یادداشتهای فراوانی برداشته است.

در مورد اینکه چه اشخاصی نویسنده فدرالیست هستند؟ شناخت کلی وجود دارد، ولی

۱- *Paulines*

هیچ نوع انتساب رسمی که یک مقاله معین به نویسنده مشخصی مربوط شود، تا سال ۱۸۰۷ یعنی تا سه سال پس از درگذشت *هامیلتون* در نتیجه دوئل با *آرون بار* دیده نشده‌است. *مدیسون* فهرست خود را از نویسندگان صرفاً در سال ۱۸۱۸، پس از آنکه از ریاست جمهوری کناره گرفت، منتشر ساخت. طی یک‌ونیم قرن انواعی از فهرستها باادعاهای متناقض مورد بحث و مناقشه قرار داشت. پیرامون نویسندگان ۷۰ مقاله توافق کلی وجود دارد: ۵ مقاله توسط جی، ۱۴ مقاله توسط *مدیسون* و ۵۱ مقاله توسط *هامیلتون*. از مابقی ۱۵ مقاله، ۱۲ مقاله مورد مناقشه بین *هامیلتون* و *مدیسون* و ۳ مقاله نیز به‌عنوان کار مشترک در سطح مجادله و گفتگو می‌باشد. بدون تردید دلیل اصلی وجود مناقشه مزبور این است که *مدیسون* و *هامیلتون* شتابی برای طرح ادعای خود نداشته‌اند. طی چند سال پس از نگارش مقاله‌ها، آنها به دشمنان سخت سیاسی یکدیگر تبدیل شده و هریک بنابه مورد، موضعی مخالف با برخی نوشته‌های خود در فدرالیست اتخاذ کردند.

محتوای سیاسی رساله‌ها به‌هیچ وجه مدرک متقاعدکننده‌ای برای تعیین نویسنده به‌دست نمی‌دهد. چون *هامیلتون* و *مدیسون* که هریک به اختصار در حمایت از تصویب قانون اساسی می‌نوشتند، مانند وکلایی بودند که برای یک موکل کار می‌کردند و ضرورتی نداشت هر استدلالی راکه به نفع قانون اساسی جدید مطرح می‌کردند، باور داشته یا ظهرنویسی نمایند. در عین حال این بدان معنا نیست که آنها خود را به دردسر می‌انداخته‌اند تا موضع شخصی خود را مشتبه جلوه دهند، بلکه بدان معناست که ما نمی‌توانیم استدلال کنیم؛ «*هامیلتون* چنین نگفته است، زیرا اعتقاد وی چیز دیگری بوده‌است.» و بسیار دیده شده‌است که مواضع سیاسی اشخاص تغییر می‌یابند. بدین لحاظ محتوای سیاسی یک مقاله مورد مناقشه نمی‌تواند سند محکمی به نفع *هامیلتون* یا *مدیسون* به‌عنوان نویسنده آن باشد.

هماهنگ با تغییر جو سیاسی، پذیرش ادعاهای مختلف توسط مورخین نیز به تحول گرایش یافته است. طی نیمه دوم قرن نوزدهم ادعاهای مربوط به *هامیلتون* مورد استقبال بود، از آن پس دعاوی در مورد *مدیسون* ارجحیت یافت. در حالی که بررسیهای تاریخی دقیق مورخی چون *داگلاس آدر* طی چندین دهه، ادعاهای مربوط به *مدیسون* را تأیید می‌کرد، تمامی شواهد تاریخی در حال حاضر با آنچه مورخینی نظیر مورخ بزرگ *هنری کابوت لاج* به نفع *هامیلتون*

تفسیر کرده‌است، تفاوت چندانی ندارند. شواهد تازه مورد نیاز است تا انتساب صریح و قطعی به‌دست دهد و مدرکی که ماهیت آن بر روشهای آماری باشد، امکانی در این راستا فراهم می‌کند. تهیه این مدرک و روش شناسی تفسیر آن در محور توجه اثر حاضر قرار دارد.

نوشته‌های *هامیلتون* و *مدیسون* را مشکل می‌توان از یکدیگر تمیز داد و این بدان علت است که هر دو نویسنده در نویسندگی به سبک اسپکتیتور[1] که هم برای مردم جذابیت داشته و هم جامع‌الاطراف و خطابه‌گونه باشد، استاد بودند. برای نشان دادن این مشکلات، *فردریک ویلیامز* و *فردریک موستلر* در سال ۱۹۴۱، طول جملات را در مقاله‌های این دو، که درباره آنها مجادله‌ای وجود نداشت، شمردند و به این نتیجه رسیدند که طول هر جمله به‌طور متوسط در نوشته‌های *هامیلتون* ۳۴/۵ و در نوشته‌های *مدیسون* ۳۴/۶ کلمه است. طول جمله، مقیاسی است که برای تمیز بین سایر نویسندگان به‌طرز موفقیت‌آمیزی به‌کار بسته شده‌است ولی *هامیلتون* و *مدیسون* دراین زمینه عملاً همزاد یکدیگرند.

کلمات مشخصه

بااینکه طول جمله مقیاسی برای پیچیدگی (و متوسط ۳۵ کلمه‌ای جمله‌ها بیانگر پیچیدگی بسیار) آن است، این مقیاس از حساسیت کافی برای تمیز رضایتبخش بین دو نویسنده که با سبک مشابهی می‌نویسند، برخوردارنیست. متغیرهایی که در تعداد بررسیها درباره نویسنده مورد مناقشه به‌کار رفته عبارتست از میزان وجود کلمات مختص فردی در نوشته‌ها. مشوق ما برای این بررسی، کشف آدر یا آن‌طورکه بعداً معلوم شد کشف مجدد اوست که به این نتیجه دست یافته که *مدیسون* و *هامیلتون* در انتخاب بین کلمات بین "while" و "whilst" همواره با یکدیگر مغایرت دارند. در ۱۴ مقاله فدرالیست که نگارش آنها توسط *مدیسون* مورد تأیید قرار گرفته، کلمه "While" مطلقاً پیش نیامده در حالی که کلمه "whilst" در ۸ مورد آنها وجود دارد. در ۱۵ مقاله از ۴۸ مقاله *هامیلتون* کلمه "while" وجود دارد ولی کلمه "whilst" هرگز دیده نشده‌است. مورد مزبور مثالی است برای آنچه که به‌عنوان علائم مشخصه نامیده می‌شود. یعنی

۱- *Spectator*

اقلامی که وجود آن گواه بارزی برای متعلق بودن نوشته به یکی از این دو فرد است . به این ترتیب وجود کلمه "whilst" در ۵ مقاله مورد مناقشه ظن را به سمت آنکه مدیسون نگارنده این ۵ مقاله بوده‌است، متمایل می‌کند.

علائم مشخصه چنانچه یافت شوند، به متمایزسازی کمک زیادی می‌نمایند، در عین حال که خود نیز خالی از اشکال نیستند. نخستین اشکال این است که "while" یا "whilst" فقط در کمتر از نیمی از مقاله‌ها وجود دارند. این علائم در نیم دیگر مقاله‌ها غایب هستند واز این‌رو مدرکی از این راه عاید نمی‌شود. می‌توان امید بست که با یافتن مقدار کافی کلمات یا ساختارهای مشخصه متفاوت به‌صورتی که همیشه یک مورد یا بیشتر حضور داشته باشد، بر مشکل مزبور غلبه نمود. مشکل دوم وبسیار مهمتر این است که با مدرک قرار دادن ۱۴ مقاله مدیسون نمی‌توان اطمینان یافت که وی هرگز کلمه while را مورد استفاده قرار نمی‌داده است. سایر نوشته‌های وی مورد آزمایش قرار گرفتند و در دو مورد انحراف دیده شد. پس وجود کلمه "while" علامت خوب ولی نه اطمینان‌بخش برای تعیین *هامیلتون* به‌عنوان نویسنده مقاله است؛ در حالی که وجود کلمه "whilst" شاخص بهتر ولی ناقصی برای نویسندگی مدیسون به‌شمار می‌رود. زیرا در مورد *هامیلتون* نیز ممکن است موارد انحراف دیده شود.

یکی از وظایف محوری آمار، استنباط در حالت عدم وجود حتمیت است. دست کشیدن از نظریه علائم مشخصه کامل، ما را با یک مسئله آماری مواجه می‌سازد. مجبور هستیم ابتدا مدرکی یافته، توان آن را ارزیابی نموده و با جمع کردن آن به یک نتیجه مرکب دست یابیم. با اینکه مسائل نظری و عملی این امر خالی از اشکال نیستند. فرصت برای جمع‌کردن مدارک به‌مراتب بسیار قانع‌کننده‌تر از اطمینانی است که حتی توسط چند علامت مشخصه تقریباً کامل پدید می‌آید.

نسبت استفاده ازکلمات

به‌جای در نظر گرفتن یک کلمه به‌عنوان علامت مشخصه که وجود یا عدم وجود آن، نویسنده یک مقاله را معین کند، می‌توانیم میزان یا فراوانی نسبی استفاده از هر کلمه رابه‌عنوان مقیاسی برای متوجه شدن نسبت به این یا آن نویسنده برگزینیم. البته غالب کلمات در این مورد کمکی

نمی‌کند. زیرا استفاده از آنها توسط هر دو نویسنده تقریباً برابر است. اما چون هزاران کلمه در دسترس داریم، بعضی از آنها ممکن است در این مورد راهگشا باشند. کلمات مجموعه عظیمی از تمیز دهنده‌های ممکن را شکل می‌بخشند. با جستجوی قاعده‌مند در این مجموعه کلمات زوجهایی چون whilst و while دیگر پیدا نشد ولی کلمات مجردی را یافتیم که توسط یک نویسنده به‌طور منظم و توسط دیگری بندرت مورد استفاده واقع شده‌بود.

جدول ۱ رفتار سه کلمه «به‌طور معمول» (Commonly)، «ابتکار» (Innovation) و «جنگ» (War) را نشان می‌دهد. در این جدول داده‌های مربوط به ۴۸ مقاله سیاسی که نگارش آنها توسط *هامیلتون* و ۵۰ رساله که نگارش آنها توسط *مدیسون* مورد تأیید قرار گرفته، به‌طور خلاصه درج شده‌است. (پاره‌ای مقاله‌های سیاسی با وجودی که در منابعی غیر از فدرالیست درج شده‌اند، چون نگارش آنها توسط *هامیلتون* یا *مدیسون* مورد تأیید بود، برای یافتن مبنای وسیعتری جهت استنباط، در شمول این بررسی قرار گرفت. تمامی مقاله‌های بعدی *هامیلتون* مندرج در فدرالیست به‌طور کامل در این مجموعه وجود ندارد. در مورد *مدیسون* نیز بیشتر مقاله‌ها از منابع خارج از فدرالیست جمع‌آوری شده‌است.

نه *هامیلتون* ونه *مدیسون* «به‌طور معمول» را زیاد به‌کار نبرده‌اند. با این همه استفاده هامیلتون از این کلمه به‌مراتب فراوانتر از *مدیسون* است. به‌طوری که در جدول منعکس است در ۳۱ مقاله از ۴۸ مقاله *هامیلتون* کلمه «به‌طور معمول» هیچگاه دیده نشده ولی در بقیه ۱۷ مقاله، این کلمه یک بار یا بیشتر آمده‌است. بررسیها نشان می‌دهد که *مدیسون* در ۵۰ مقاله فقط یک بار این کلمه را به‌کار برده است. طول مقاله‌ها از ۹۰۰ کلمه تا ۳۵۰۰ کلمه متغیر بوده و متوسط آنها حدود ۲۰۰۰ کلمه می‌باشد. حتی یک رخداد در ۹۰۰ کلمه، استفاده فشرده‌تری را نسبت به دو رخداد در ۳۵۰۰ کلمه نشان می‌دهد. پس در عوض کار با تعداد رخدادها در یک مقاله، نسبت رخداد را مورد استفاده قرار می‌دهیم و ۱۰۰۰ کلمه را به‌عنوان مبنای مناسبی برای بررسی در نظر می‌گیریم. به این ترتیب، برای مثال مقاله دارای بالاترین نسبت (۱/۳۳ در هزار کلمه) استفاده از کلمه «به‌طور معمول»، مقاله‌ای است با ۱۵۰۰ کلمه، با دو رخداد. رفتار کلمه «ابتکار» نیز به‌همین ترتیب است. ولی این کلمه علامت مشخصه‌ای برای *مدیسون* است. برای هردوی این کلمات بیشترین نسبت استفاده، کمی بیشتر از ۱ در هزار است.

رفتار کلمه «جنگ» به‌طرز جالب‌توجهی متفاوت است. با وجودی که در نیمی ازمقاله‌های *هامیلتون* اثری از این کلمه نیست، ولی وقتی یافت می‌شود با استفاده فراوان توأم است؛ به‌طوری که در یک مقاله از این کلمه به نسبت ۱۴ بار در هزارکلمه استفاده شده‌است. مقاله‌های فدرالیست درباره موضوعهای ویژه پیرامون قانون اساسی است و تنوع چشمگیری در نسبت استفاده از کلماتی چون: جنگ، قانون، اجرایی، آزادی، تجارت را می‌توان متناسب با محتوای این مقالات انتظار داشت. حتی، با وجود اینکه استفاده مدیسون از کلمه «جنگ» در مقاله‌های مورد تأیید به‌مراتب بیش از *هامیلتون* است، می‌توان توضیح داد که این امر بیشتر ناشی از تقسیم وظایف بوده‌است تا رجحان مدیسون برای استفاده از کلمه «جنگ». داده‌های استفاده از کلمه‌ای چون «جنگ»، مدرکی مشابه همان نوع مدارک مسئله‌داری است که بالغ بر یکصد سال گذشته تضاد و اختلاف نظر بین مورخین را موجب شده‌است. درواقع مناقشه ادامه یافته واین بدان علت است که مدرک آوردن از موضوع و محتوا به نحو مأیوس‌کننده‌ای غیرقاطع می‌باشد.

استفاده از کلمات غیرمتنی

برای آنکه استدلال آماری از اعتبار لازم برخوردار شود، باید اطلاعات مربوط به کلمات متنی و معنایی، عمدتاً کنار گذاشته شود. به این ترتیب مطالعه و بررسی برای یافتن نویسنده، سهمی در شناخت اهمیت مقاله‌ها به‌طور مستقیم نخواهد داشت ولی مدارک دال بر نویسندگی مقاله می‌تواند هم استناد در مورد نویسنده را مستحکم سازد و هم آن را از مدرکی که از طریق تحلیلهای تاریخی تدارک دیده می‌شود بی‌نیاز سازد.

اجتناب از داوری درباره پرمعنی بودن یا اهمیت مقاله، در عرف طبقه‌بندی و شناسایی مرسوم است. هنگامی که منتقدان هنری، سعی می‌کنند اصالت یک تابلو را به‌اثبات برسانند، علاوه بر سابقه تاریخی، موضوعهای کوچکی را مورد ملاحظه قرار می‌دهند: ناخنها و گوشها چگونه نقاشی شده‌اند؟ و چه نوع رنگ و بومی به‌کار برده شده‌است؟ سهم نسبتاً کوچکی از قضاوت نهایی بر مزیت هنری نقاشی مبتنی است. به همین طریق در مسائل پلیسی، غالباً در شناسایی افراد از اثرانگشت، سوابق دندانها و آثار باقیمانده زخمها، بدون اشاره به شخصیت، شغل یا موقعیت اجتماعی آنها استفاده می‌شود. برای شناسایی آثار ادبی الزامی نیست.

جدول ۱ـ توزیع فراوانی نسبت استفاده ازکلمات «به‌طور معمول»، «ابتکار» و «جنگ» در هر هزار کلمه در ۴۸ مقاله هامیلتون و ۵۰ مقاله مدیسون

به‌طور معمول			ابتکار			جنگ		
نسبت درهزارکلمه	هامیلتون	مدیسون	نسبت درهزارکلمه	هامیلتون	مدیسون	نسبت درهزارکلمه	هامیلتون	مدیسون
۰(دقیقاً)(۱)	۳۱	۴۹	۰(دقیقاً)(۱)	۴۷	۳۴	۰(دقیقاً)(۱)	۲۳	۱۵
۰/۲ - ۰	غیرممکن(۲)	غیرممکن(۲)	۰/۲ - ۰	غیرممکن(۲)	غیرممکن(۲)	۲ - ۰	۱۶	۱۳
۰/۴ - ۰/۲	۳	۱	۰/۴ - ۰/۲		۶	۴ - ۲	۴	۵
۰/۶ - ۰/۴	۶		۰/۶ - ۰/۴	۱	۶	۶ - ۴	۲	۴
۰/۸ - ۰/۶	۲		۰/۸ - ۰/۶		۱	۸ - ۶	۱	۳
۱ - ۰/۸	۳		۱ - ۰/۸		۲	۱۰ - ۸	۱	۳
۱/۲ - ۱/۰	۲		۱/۲ - ۱/۰		۱	۱۲ - ۱۰		۳
۱/۴ - ۱/۲	۱					۱۴ - ۱۲		۲
						۱۶ - ۱۴	۱	۲
جمع	۴۸	۵۰	جمع	۴۸	۵۰	جمع	۴۸	۵۰

۱ـ هریک از بازه‌ها، به‌استثنای ۰(دقیقاً) کرانه بالایی خود را شامل نمی‌شود. بنابراین در یک مقاله ۲۰۰۰کلمه‌ای که کلمه «به‌طور معمول» دوبار دیده شود، نسبتی دقیقاً برابر ۱ در ۱۰۰۰ به‌دست می‌آید و به این ترتیب مقاله مزبور در شمار بازه ۱/۲ - ۱ درج شده‌است.

۲- با مشخص بودن طول مقاله‌های مورد بررسی، ممکن است تصادفاً چنین پیش آید که یک نسبت در این بازه نمی‌تواند رخ دهد. برای مثال اگر یک مقاله ۲۰۰۰کلمه داشته باشد، نسبت ۱ در ۱۰۰۰ به‌معنای وجود ۲کلمه مورد نظر است و یک رخداد منحصر به‌فرد به‌معنای نسبتی برابر ۰/۵ در ۱۰۰۰ است. به‌این ترتیب مقاله‌ای با ۲۰۰۰کلمه نمی‌تواند به نسبتی بزرگتر از صفر و کوچکتر از ۰/۵ بینجامد.

مأخذ: موستلر و والاس (۱۹۸۴).

که نسبت به ارزیابی سبک نگارش آثار ادبی دارای مهارت و استعداد باشیم، هرچند این امر در

پاره‌ای مسائل کمک‌کننده باشد. برای شناسایی یک شیء نیازی نیست که ارزش کامل یا مفهوم آن درک شود.

حال باید دید کدام کلمات غیرمتنی برای متمایزسازی نویسندگان، گزینه‌های مناسبی به شمار می‌روند؟ جالبترین این کلمات، واژه‌های تکمیل‌کننده در زبان هستند: حروف اضافه، حروف ربط و حروف تعریف. بسیاری حروف معنادارتر دیگر نیز وجود دارند که به‌نظر می‌رسد به‌طور نسبی فارغ از متن باشند: قیدهایی ازقبیل «در نتیجه»، «خصوصاً» و «معمولاً» یا حتی

جدول ۲- توزیع فراوانی میزان استفاده ازکلمات From, By و To در هر ۱۰۰۰ کلمه در ۴۸ مقاله هامیلتون و ۵۰ مقاله مدیسون

By			From			To		
نسبت درهزارکلمه	هامیلتون	مدیسون	نسبت درهزارکلمه	هامیلتون	مدیسون	نسبت درهزارکلمه	هامیلتون	مدیسون
۳-۱(۱)	۲		۳-۱(۱)	۳	۳	۲۵-۲۰(۱)		۳
۵-۳	۷		۵-۳	۱۵	۱۹	۳۰-۲۵	۲	۵
۷-۵	۱۲	۵	۷-۵	۲۱	۱۷	۳۵-۳۰	۶	۱۹
۹-۷	۱۸	۷	۹-۷	۹	۶	۴۰-۳۵	۱۴	۱۲
۱۱-۹	۴	۸	۱۱-۹		۱	۴۵-۴۰	۱۵	۹
۱۳-۱۱	۵	۱۶	۱۳-۱۱		۳	۵۰-۴۵	۸	۲
۱۵-۱۳		۶	۱۵-۱۳		۱	۵۵-۵۰	۲	
۱۷-۱۵		۵				۶۰-۵۵	۱	
۱۹-۱۷		۳						
جمع	۴۸	۵۰	جمع	۴۸	۵۰	جمع	۴۸	۵۰

۱ـ هریک از بازه‌ها کرانه بالایی خود را شامل نمی‌شود. به این ترتیب مقاله‌ای با نسبت دقیقاً ۳ در هزارکلمه در شمار بازه ۵-۳ درج می‌شود.

مأخذ: موستلر و والاس (۱۹۸۴)

اسامی مجرد مانند «قدرت» یا «ابتکار». ما به کلماتی برای بررسی مقاله احتیاج داریم که استفاده از آنها ارتباطی به موضوع مقاله نداشته باشد و بتوانیم آنها را به‌عنوان واکنش جزیی یا ناخودآگاه رجحانهای نویسنده تلقی نمائیم.

اکنون باید بررسی شود که با این واژه‌های تکمیل‌کننده چه می‌توان کرد؟ پاره‌ای از آنها کلماتی هستند با بیشترین کاربرد در زبان: the, and, of, to و غیره. هیچ فردی بدون استفاده از این کلمات نمی‌نویسد ولی میزان استفاده از این کلمات توسط هر نویسنده با نویسنده دیگر تفاوت دارد. در جدول ۲ توزیع نسبتهای استفاده از سه حروف اضافه by, from و to نشان داده شده‌است. ابتدا به تغییرات از مقاله‌ای به مقاله دیگر توجه کنید. مدیسون نوعاً کلمه by را تقریباً ۱۲ بار در هر ۱۰۰۰ کلمه استفاده می‌کند ولی در پاره‌ای موارد این نسبت در حدود ۱۸ بار در حد بالا و ۶ بار در حد پایین نوسان دارد. به‌هرحال با توجه به بررسیها، این تغییرات، گرایش سیستماتیک مدیسون را برای استفاده بیشتر از کلمه by نسبت به *هامیلتون* مخفی نمی‌سازد. به‌این ترتیب نسبت پایین استفاده از by دلالت بر نویسندگی *هامیلتون* و نسبت بالای استفاده از این کلمه دال بر نویسندگی مدیسون است. نسبتهای استفاده از to در جهت عکس است. نسبتهای بسیار بالای استفاده از from اشاره به نویسنده بودن مدیسون دارد ولی نسبتهای پایین استفاده از این کلمه عملاً هیچ اطلاعی به‌دست نمی‌دهد. هرچه پراکندگی توزیع وسیعتر باشد، قدرت و توان کلمات برای متمایزسازی افزونتر خواهد بود.

درمثال مورد استفاده، کلمه by بهتر از to متمایز می‌شود ولی to نیز به نوبه خود بهتر از from به‌شمار می‌رود.

الگوهای احتمالات

برای به‌کاربردن هر نظریه استنباط آماری در شواهد حاصل از نسبتهای کلمات ناگزیر هستیم یک مدل احتمالات پذیرفتنی ایجاد نماییم تا به‌وسیله آن تغییرپذیری نسبت کلمات از یک مقاله به مقاله دیگر بیان شود. ساختن یک مدل کامل برای رخداد حتی یک کلمه منحصر به‌فرد کار عبثی خواهد بود. زیرا ساختار دقیق داخل یک جمله در مقیاس وسیعی توسط عناصر غیرتصادفی دستوری، مفهومی و شیوه نگارش تعیین می‌شود. ولی چنانچه علاقه و توجه ما

محدود به نسبتهای استفاده از یک یا چند کلمه در قطعاتی از متن، متشکل از حداقل ۱۰۰ یا ۲۰۰ کلمه باشد در این صورت انتظار داریم ساختار تفصیلی عبارات و جملات چندان اهمیتی نداشته باشد. ساده‌ترین مدل را که در احتمالات کلاسیک بسیار رایج است، می‌توان به وسیله گویهای موجود در یک گلدان شرح داد. برای نشان دادن استفاده مدیسون از کلمه by فرض می‌کنیم یک نسبت ویژه استفاده از این کلمه برای مدیسون وجود دارد که حدود ۱۲ بار در ۱۰۰۰ کلمه است. حال گلدانی را تصور می‌کنیم که انباشته از هزاران گوی به رنگهای قرمز و سیاه بوده و در آن نسبت گوی قرمزرنگ به مجموعه گویهای داخل گلدان ۱۲ در هزار باشد. مدل احتمال برای رخداد by همانند مدل احتمال برای برداشتن پی‌درپی گوی از داخل گلدان است که در آن گوی قرمز نظیر by و گوی سیاه نظیر کلیه کلمات دیگر در این نوشته‌هاست. اگر خواسته باشیم این مدل رابه بررسی همزمان رخ دادن دو یا چند کلمه گسترش دهیم، به گویهایی با سه یا چند رنگ مختلف نیاز خواهیم داشت. در این مدل هیچ نوع ساختار دستوری یا مفهومی مورد توجه قرار نگرفته و قصد بر نمایاندن چگونگی رفتار در داخل جملات نیز نبوده است. آنچه مطلوب است توضیح و شرح تغییرات در نسبتها، یعنی موارد رخدادها در قطعات طولانی کلمات، در انطباق با رساله‌ها است.

مدل مزبور را با مقایسه پیش‌بینی آن و شمارش واقعی فراوانی کلمات موردنظر در مقاله‌ها مورد آزمایش قرار می‌دهیم. نتیجه‌ای که به‌دست می‌آید این است: در عین حال که طرح گلدان و گوی تغییرپذیریهای بسیاری از کلمات رابه‌خوبی منعکس می‌نماید، برای کلمات دیگر، تغییرپذیریهای اضافی موردنیازاست. تغییرات تصادفی در طرح گلدان بیان‌کننده غالب تغییرات از یک رساله به رساله دیگر می‌باشد ولی در پاره‌ای رساله‌ها، نویسندگان نسبتهای مبنای خود را اندکی تغییر می‌دهند. ناچاریم مدل نظری خود را جامعتر سازیم تا این موضوع منظور شود و مدلی که در این حالت مورد استفاده قرار می‌دهیم به نام "توزیع دوجمله‌ای منفی" خوانده می‌شود.

آزمایش همچنین نشان داد که ضمایری چون "his" و "her" متمایزکننده‌های فوق‌العاده غیرقابل اعتماد، حتی ضعیفتر از کلماتی مانند «جنگ» در زمینه تعیین نویسنده به شمار می‌روند.

استنباط و نتایج

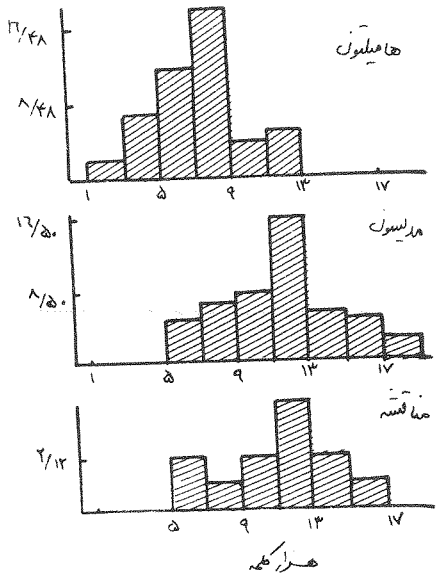
هر طریق ممکن از ایجاد مدلها گرفته تا ارزیابی کمی مواردی چون اثبات اینکه پاره‌ای مقاله‌های مورد منازعه توسط مدیسون نگاشته شده‌اند، مستلزم راه حلهایی برای مسائل نظری حاد آماری و توسعه ریاضیات جدید است. یکی از انگیزه‌های مهم، این بود که از مسئله فدرالیست به‌عنوان یک تحقیق موردی استفاده کرده، چندین رهیافت متفاوت آماری را با توجه ویژه بر یکی از آنها به‌نام روش بیزی، مورد مقایسه قرار دهیم که نتیجه نهایی را برحسب احتمالات یا بختهای نویسنده بیان می‌کند.

هر روشی که مورد استفاده واقع شود، نتیجه یکسان است: مدارک قاطع برله نگارش مقاله‌های موردنظر توسط مدیسون. مدارک فقط برای یک مقاله متعادلتر است ولی حتی در این مورد نیز یک بررسی جامعتر این نتیجه را به‌دست داد که تفاوتها ۸۰ به ۱ به نفع مدیسون است.

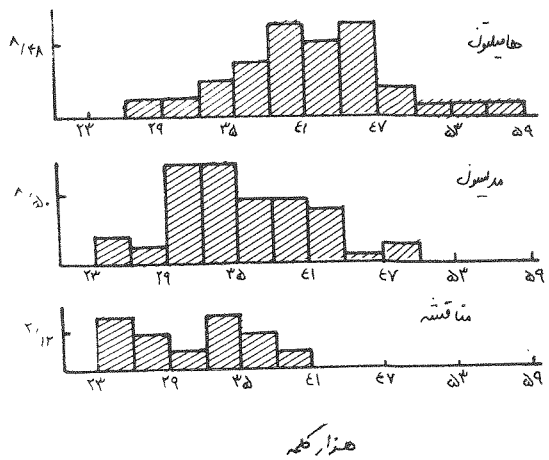
در تصاویر ۱ و ۲ چگونگی برازش ۱۲ مقاله مورد منازعه را با توزیع نسبتهای رخداد دو کلمه که نهایتاً به‌عنوان متمایزکننده در مقاله‌های *هامیلتون* و مدیسون انتخاب شده‌اند، نشان می‌دهد. در تصویر ۱، دو نمودار ستونی فوقانی، داده‌های مربوط به by را که قبلاً در جدول ۲ بیان شد، نشان می‌دهد. به‌طوری که دیده می‌شود نسبتهای مربوط به مدیسون به‌طور متوسط در سطح بالاتری قراردارد. ابتدا نمودار ستونی تحتانی مربوط به مقاله‌های مورد مناقشه را با نمودار ستونی بالایی، مربوط به مقاله‌های *هامیلتون* و سپس با نمودار بعدی مربوط به مقاله‌های مدیسون مقایسه کنید. نسبتهای ملاحظه شده در مقاله‌های مورد مشاجره، به‌طور کلی که نگاه شود، شباهت زیادی به مقاله‌های مدیسون دارد. هرچند ۳ مقاله از ۱۲ مقاله فی‌نفسه اندکی در جهت نسبتهای مربوط به *هامیلتون* هستند. تصویر ۲ داده‌های نظیر را برای کلمه «to» نشان می‌دهد. در اینجا مجدداً مقاله‌های مورد مناقشه بیشتر با توزیع مدیسون سازگار هستند ولی به‌مراتب بیشتر از آن حدی که در مورد مقاله‌های مشهور مدیسون مشخص شده‌است از مقاله‌های *هامیلتون* فاصله می‌گیرند.

جدول ۳، آن ۳۰ کلمه‌ای که در استنباط نهایی مورد استفاده قرار گرفته‌اند، همراه با نسبتهای میانگین آنها در هر هزارکلمه نوشته‌های *هامیلتون* و مدیسون درج شده‌است. گروهها در این جدول برمبنای درجه وابستگی قراین، که توسط *موستلر و والاس* در سال ۱۹۸۴ و مقدم بر این

شکل ۱ـ توزیع نسبتهای رخداد By در ۴۸ مقاله هامیلتون، ۵۰ مقاله مدیسون و ۱۲ مقاله مورد مناقشه



شکل ۲ـ توزیع نسبتهای رخداد To در ۴۸ مقاله هامیلتون، ۵۰ مقاله مدیسون و ۱۲ مقاله مورد مناقشه



در تحلیل پیش‌بینی شده‌بود، تعیین شده‌اند.

مدرک ترکیبی از هشت کلمه تکمیل‌کننده رایج که تحت گروه B قرار گرفته پرتواترترین بوده و بسیار مهمتر از مدرک ترکیبی از ۲۰ کلمه مشخصه با فراوانی کمتر از قبیل while - whilst هستند. این ۲۰ کلمه تحت گروههای E,D,C قرارگرفته‌اند.

یک کلمه باقی می‌ماند که خیلی زود به‌عنوان یک متمایزکننده پرقدرت، خودنمایی کرده و تقریباً به‌تنهایی برای این امر کفایت می‌کند. چه موقع باید نویسنده به‌جای on از upon

جدول ۳ـ کلمات مورد استفاده در تمایز نهایی و نسبتهای تعدیل‌شده کاربرد آنها توسط مدیسون و هامیلتون

کلمه	نسبت در ۱۰۰۰ کلمه		کلمه	نسبت در ۱۰۰۰ کلمه	
	هامیلتون	مدیسون		هامیلتون	مدیسون
گروه A			گروه D		
Upon	۳/۲۴	۰/۲۳	Commonly	۰/۱۷	۰/۰۵
گروه B			Consequently	۰/۱۰	۰/۴۲
Also	۰/۳۲	۰/۶۷	Considerable(ly)	۰/۳۷	۰/۱۷
An	۵/۹۵	۴/۵۸	According	۰/۱۷	۰/۵۴
By	۷/۳۲	۱۱/۴۳	Apt	۰/۲۷	۰/۰۸
Of	۶۴/۵۱	۵۷/۸۹	گروه E		
On	۳/۳۸	۷/۷۵	Direction	۰/۱۷	۰/۰۸
There	۳/۲۰	۱/۳۳	Innovation(s)	۰/۰۶	۰/۱۵
This	۷/۷۷	۶/۰۰	Language	۰/۰۸	۰/۱۸
To	۴۰/۷۹	۳۵/۲۱	Vigor(ous)	۰/۱۸	۰/۰۸
گروه C			Kind	۰/۶۹	۰/۱۷
Although	۰/۰۶	۰/۱۷	Matter(s)	۰/۳۶	۰/۰۹
Both	۰/۵۲	۱/۰۴	Particularly	۰/۱۵	۰/۳۷
Enough	۰/۲۵	۰/۱۰	Probability	۰/۲۷	۰/۰۹
While	۰/۲۱	۰/۰۷	Work(s)	۰/۱۳	۰/۲۷
Whilst	۰/۰۸	۰/۴۲			
Always	۰/۵۸	۰/۲۰			
Though	۰/۹۱	۰/۵۱			

مأخذ: موستلر و والاس (۱۹۸۴)

استفاده کند؟ حتی کتابهای موثق در مورد کاربرد زبان انگلیسی قواعد قانع‌کننده‌ای در این زمینه ارائه نمی‌دهند. *هامیلتون* و *مدیسون* در این مورد تفاوتهای فوق‌العاده دارند. on و upon در نوشته‌های *هامیلتون* تقریباً برابر به‌کار رفته است. حدود ۳ بار در هر ۱۰۰۰ کلمه. *مدیسون* از سوی دیگر به ندرت از upon استفاده می‌کند. در جدول ۴ توزیع کاربردهای کلمه upon نشان داده شده‌است. *هامیلتون* در ۴۸ مقاله هرگز از کاربرد upon کوتاهی نکرده‌است. او هرگز کمتر از دوبار از این کلمه استفاده نکرده‌است. *مدیسون* فقط در ۹ مقاله از ۵۰ مقاله خود این کلمه را به‌کار برده و آن هم در نسبتهای بسیارپایین. به این ترتیب مقاله‌های مورد مشاجره، با عنایت به اینکه کلمه upon فقط در یکی از آنها مشاهده شده‌است، به‌طور واضحی مربوط به *مدیسون* هستند. خوشبختانه این یک مقاله نیز به‌طور بسیار قدرتمندی توسط سایر کلمات در این

جدول ۴- توزیع فراوانی نسبتهای وجود کلمه Upon در ۱۰۰۰ کلمه در ۴۸ مقاله هـامیلتون، ۵۰ مقاله مدیسون و ۱۲ مقاله مورد مشاجره

نسبت استفاده در ۱۰۰۰ کلمه	هامیلتون	مدیسون	مورد مناقشه
۰ دقیقاً(۱)		۴۱	۱۱
۰+ - ۰/۴		۲	
۰/۴ - ۰/۸		۴	
۰/۸ - ۱/۲	۲	۱	۱
۱/۲ - ۱/۶	۳	۲	
۱/۶ - ۲/۰	۶		
۲/۰ - ۳/۰	۱۱		
۳/۰ - ۴/۰	۱۱		
۴/۰ - ۵/۰	۱۰		
۵/۰ - ۶/۰	۳		
۶/۰ - ۷/۰	۱		
۷/۰ - ۸/۰	۱		
جمع	۴۸	۵۰	۱۲

۱- *بازه‌ها، به‌استثنای (۰) دقیقاً، شامل کرانه بالایی نمی‌شود. به این ترتیب مقاله‌ای با نسبت دقیق ۳ در هزار کلمه در شمار فاصله ۴/۰ - ۳/۰ درج شده‌است.*

مأخذ: موستلر و والاس (۱۹۸۴)

طبقه‌بندی قرار گرفته است. مقاله مزبور مقاله‌ای نیست که تفاوتهای متعادل داشته باشد.

مسلم است که ترکیب و ارزیابی کل مدارک، یک اقدام پرحجم آماری و محاسباتی است. کامپیوترهای باسرعت بالا باید ساعتهای متمادی به محاسبه بپردازند، اعم از محاسبات ریاضی برای نظریه و محاسبات عملی برای داده‌ها.

ممکن است به فکر *جان جی* بیفتید. آیا وی نیز سهمی درمقاله‌های مورد مناقشه نداشته است؟ در جدول ۵ نسبتهای استفاده از ۹ کلمه با بیشترین فراوانی در زبان انگلیسی که در هر ۱۰۰۰ کلمه از نوشته‌های *هامیلتون*، *مدیسون*، *جی* و برای تنوع در روش در کتاب اولیس اثر جیمز جویس اندازه‌گیری شده، درج شده است. این جدول ادعای مکرر در مورد تشابه *مدیسون* و *هامیلتون* را مورد تأئید قرار می‌دهد. جویس بسیار متفاوت است ولی *جان جی* هم همین‌طور است. دو کلمه of و to با مقایسه‌های نسبتی ۶۵ به ۵۸ و ۴۱ به ۳۵ دربین متمایزکننده‌های نهایی میان *هامیلتون* و *مدیسون* قرار داشتند. همان‌طورکه دیده می‌شود به سهولت می‌توان در استفاده ازکلمات the, of, and, a و that *جی* را از *هامیلتون* یا *مدیسون* متمایز نمود. مقالات مورد مناقشه به هیچ وجه با نسبتهای جی سازگار نیستند و هیچ دلیلی ندارد که حذف وی از منازعه بر سر نگارش مقاله‌های مزبور مورد سئوال قرارگیرد.

جدول ۵ـ نسبت استفاده از کلمات با فراوانی زیاد (نسبت در ۱۰۰۰ کلمه)

	هامیلتون (۱)(۹۴۰۰۰)	مدیسون (۱)(۱۱۴۰۰۰)	جی (۱)(۵۰۰۰)	جویس (اولیس) (۱)(۲۶۰۰۰۰)
The	۹۱	۹۴	۶۷	۵۷
Of	۶۵	۵۸	۴۴	۳۰
To	۴۱	۳۵	۳۶	۱۸
And	۲۵	۲۸	۴۵	۲۸
In	۲۴	۲۳	۲۱	۱۹
A	۲۳	۲۰	۱۴	۲۵
Be	۲۰	۱۶	۱۹	۳
That	۱۵	۱۴	۲۰	۱۲
It	۱۴	۱۳	۱۷	۹

۱ـ *تعداد شمارش‌شده کلمات در متن برای محاسبه نسبتها.*

مأخذ: موستلر و والاس (۱۹۸۴) و هانلی (۱۹۳۷)

چکیده‌ای از نتایج

داده‌های ما در این بررسی، به‌طور مستقل مدرک مورخین را تکمیل می‌کند. از لحاظ درجه باور، با احتمال بسیار زیاد مدیسون نویسنده مقاله‌های مورد مناقشه فدرالیست بوده‌است. با دو استثنا مقاله شماره ۵۵، مدرکی که با تفاوت ۸۰ به ۱ در وضعیتی بسیار قوی ولی نه قاطع به نفع مدیسون است و مقاله شماره ۵۶، دومین مقاله ضعیف که با نسبت بسیار قوی ۸۰۰ به ۱ به نفع مدیسون می‌باشد. داده‌ها برای تمام مقاله‌های دیگر قاطع هستند و دو مقاله شماره ۶۲ و ۶۳ را که از نظر مورخین ضعیفترین مدارک هستند نیز شامل می‌شوند.

برای بحث بیشتر درباره این مسئله و از جمله تفصیلات تاریخی آن، بحث پیرامون روشهای واقعی و انواع تحلیلهای بدیل، و همچنین یک مرور اجمالی از بررسیهای مربوط به نویسندگان از سال ۱۹۶۹ به *موستلر و والاس* (۱۹۸۴) مراجعه شود.

مسائل

۱ـ چرا بااستفاده از شیوه نگارش ادبی و یا فلسفه سیاسی نمی‌توان نویسنده مقاله‌های مورد مناقشه را تعیین نمود؟

۲ـ الف) متمایزکننده چیست؟

ب) حداقل دو مقوله از متمایزکننده‌ها را تفکیک کنید.

پ) چرا کلمه by یک متمایزکننده خوب است ؟ (با رجوع به جدول شماره ۲)

۳ـ کلمه "غیرمتنی" چیست؟

۴ـ چرا نویسندگان به‌جای تعداد رخدادها از فراوانی کلمه در هزار استفاده می‌کنند؟

۵ـ با رجوع به جدول شماره ۱؛ در چه تعداد از مقاله‌های مورد بررسی *هامیلتون*، کلمه «به‌طور معمول» حداقل یک بار دیده شده است؟

۶ـ با رجوع به جدول شماره ۲؛ در چه درصدی از مقاله‌های مورد بررسی مدیسون، به کلمه from در فاصله ۷–۳ بار در هر ۱۰۰۰ کلمه برخورد شده‌است؟ (تذکر: بازه ۷–۳ بار استفاده مشمول قاعده ذکرشده برای بازه‌هااست.)

۷ـ با درنظر داشتن مدل «گویهای داخل گلدان »، چنانچه بخواهیم مدل احتمالات خود را به

بررسی همزمان احتمال وجود ۵ کلمه مختلف گسترش دهیم، به چند رنگ مختلف گوی احتیاج داریم؟ برای بررسی n کلمه به چند رنگ احتیاج خواهیم داشت؟

۸ـ با درنظر گرفتن تصویر شماره ۱ آیا این حکم صحیح است: در بیش از $\frac{1}{3}$ مقاله‌های مورد بررسی *هامیلتون*، کلمه by ۳ـ۷ بار در هر ۱۰۰۰ کلمه مورد استفاده واقع شده‌است؟

۹ـ با بررسی تصویر شماره ۲ مشخص کنید که نمودار مربوط به مقاله‌های مورد مشاجره، بیشتر به نمودارمربوط به مقاله‌های *هامیلتون* شبیه است یا مقاله‌های *مدیسون*؟

۱۰ـ با درنظر داشتن جدول ۳، به گروه B توجه کنید. به نظر شما چه کلمه‌ای می‌تواند بهترین متمایزکننده *هامیلتون/مدیسون* باشد؟ معیار خود را برای انتخاب کلمه شرح دهید. به سئوال مشابه درباره گروه D پاسخ دهید.

۱۱ـ در جدول شماره ۱ فراوانی نسبی کلمه «جنگ» درج شده‌است. چرا کلمه «جنگ» در جدول شماره ۳ دیده نمی‌شود؟

منابع

Miles L. Hanley. 1937. *Word Index to James Joyce's "Ulysses." Madison.* Wis University of Wisconsin.

F. Mosteller and D. l. Wallace. 1964. *Inference and Disputed Authorship. The Federalist.* Reading. Mass: Addison-Wesley.

F. Mosteller and D. l. Wallace. 1984. *Applied Bayesian and Classical Inference: The Case of "The Federalist Papers" (2nd ed. of Inference and Disputed Authorship: Federalist).* NewYork: Springer-Verlag.



یادآوری اطلاعات تصویری در کودکان

دوریس. آر آنت ویسل - دبلیو. اچ. هاگینز[1] دانشگاه جانز هاپکینز

پی بردن به چگونگی بلوغ و رشد کودکان و عواملی که موجب تسریع یا تأخیر در رشد و نمو آنها می‌شود، برای پزشکان متخصص اطفال، معلمان، والدین و بالاتراز همه برای خود کودک حائز اهمیت است. با وجود اینکه دانستنیهای ما درباره چگونگی رشد قابلیتهای کودکان بیش از گذشته است، ولی هنوز هم چندان زیاد نیست. مثلاً وقتی کودکان دوره ابتدایی راتمام می‌کنند،

۱ـ *Doris R. Entwisle - W. H. Huggins*

مهارتهای اساسی در خواندن و ریاضیات آنها دارای دامنه تغییراتی است که از مهارت یک کلاس اولی تاکلاس دهمی (دوم دبیرستان فعلی ـ مترجم)، با گستره‌ای متجاوز از ۹ سال متفاوت است. معلوم نیست چرا بعضی از کودکان تااین اندازه بهتراز دیگران رشد می‌کنند. شاید بچه‌ها در بعضی زمینه‌ها بیشتر از آنچه به آنها بها داده می‌شود، شایستگی دارند و شاید هم از بعضی استعدادهای انسانی که تاکنون شناخته نشده‌است کمک می‌گیرند. بدیهی است که اگر در ارزیابی شایستگیهای کودکان تنها، چنانکه معمول است، روی تواناییهایی که به‌وسیله آزمونهای استاندارد خواندن و ریاضیات آشکار می‌شود تمرکز کنیم، دیدگاه ضعیفی را انتخاب کرده‌ایم. چراکه این آزمونها تنها جایگاه کودکان رانسبت به این مهارتها نشان می‌دهد نه چگونگی رسیدن به آن جایگاه را.

دو مقاله‌ای که مورد توجه قرارگرفت، اشاره کرده‌اند که کودکان احتمالاً دارای بعضی قابلیتها با ماهیت بصری، بویژه در توانایی یادآوری اطلاعات تصویری می‌باشند که تابه‌حال به آنها توجه نشده‌است. یکی تحقیق هی‌بر (۱۹۷۰) است که نشان داده‌است؛ کودکان خردسال بعد از چند ثانیه قرار گرفتن در معرض بیش از ۲۵۰۰ اسلاید، توانسته‌اند بعداً ۸۵ تا ۹۵ درصد اسلایدهای قدیمی راکه با اسلایدهایی که قبلاً ندیده بودند، جفت شده بود به‌درستی تشخیص دهند. گزارش دیگر (کی‌گان ۱۹۷۰) به توسعه زودرس این توانایی که هی‌بر درکودکان کشف کرده بود اشاره‌ای گذرا دارد. به این معنی که یک کودک ۴ ساله بخش وسیعی ازعکسهای انتخاب شده از مجله هایی راکه جلوتر رویت کرده بود، به‌درستی تشخیص داده است. البته کودکان خردسال بسرعت یاد می‌گیرند که از اطلاعات تصویری برای یافتن مسیر خود در همسایگی یا اطراف خانه‌های دیگران نیز استفاده کنند.

برای شروع کار یک بررسی آزمایشی انجام دادیم و اینجاست که آمار برای اولین بار وارد می‌شود. اسلایدهای تصویری را با سرعت و پشت سرهم به ۱۰ کودک کودکستانی ۳ و ۴ ساله نشان دادیم. اسلایدها شامل منظره‌های طبیعی و صحنه‌های دیگری بودند که برای کودکان ناآشنا بود و موجودات انسانی را نشان نمی‌داد. هر کودک ۵ اسلاید را پشت سرهم در داخل یک دوربین دستی تماشا کرد. سپس بلافاصله به او ۱۰ اسلاید که مخلوطی تصادفی از ۵ اسلاید اولیه و ۵ اسلاید جدید بود به دنبال هم نشان داده شد. بعضی از کودکان توانستند ۹ عدد از آن

۱۰ اسلاید را به‌عنوان "جدید" یا "قدیم" مشخص کنند. سئوال این است که آیااین غیرعادی است؟ آیا کودکان ۳ و۴ ساله واقعاً دارای برخی تواناییها هستند یا اینکه این تعداد انتخابهای درست می‌تواند صرفاً تصادفی باشد؟

برای پاسخ به این سئوال از مدل دوجمله‌ای استفاده کردیم. انتخاب درست ۹ اسلاید از میان ۱۰ اسلاید عیناً شبیه به‌دست آوردن ۹ شیر در ۱۰ پرتاب متوالی یک سکه است. جدول ۱ طرز کار مدل دوجمله‌ای را برای ۱۰ برآمد نشان می‌دهد. از این جدول پی می‌بریم که ۹ از ۱۰ واقعاً غیرعادی است. در هر ۱۰۲۴ مرتبه تنها ۱۱ بار چنین اتفاقی رخ می‌دهد. (اگر در ۱۰ بار پرتاب سکه ۱۰ بار شیر مشاهده کنیم، بازهم تحت تأثیر قرار می‌گیریم. به همین خاطر، شانسهای ۹ انتخاب صحیح یا بیشتر را باهم جمع می‌کنیم.) حتی کودکان در سنینی که فقط می‌توانستند مورد آزمون تشخیص عکس قرار بگیرند کاملاً با موفقیت عمل کردند. سپس به طراحی آزمونی

جدول ۱ـ درستنمایی برآمدهای مختلف برای ده بار پرتاب سکه

برآمد	شانس در ۱۰۲۴ پرتاب	درصددفعاتی که برآمد مشاهده می شود
۰ شیر، ۱۰ خط	۱	۰/۱
۱ شیر، ۹ خط	۱۰	۱/۰
۲ شیر، ۸ خط	۴۵	۴/۴
۳ شیر، ۷ خط	۱۲۰	۱۱/۷
۴ شیر، ۶ خط	۲۱۰	۲۰/۵
۵ شیر، ۵ خط	۲۵۲	۲۴/۶
۶ شیر، ۴ خط	۲۱۰	۲۰/۵
۷ شیر، ۳ خط	۱۲۰	۱۱/۷
۸ شیر، ۲ خط	۴۵	۴/۴
۹ شیر، ۱ خط	۱۰	۱/۰
۱۰ شیر، ۰ خط	۱	۰/۱
جمع	۱۰۲۴	۱۰۰/۰

برای دانش آموزان کلاس اول پرداختیم که همان راهبرد کلی رادنبال می‌کرد. استفاده دوم ما از آمار در طرز طراحی مطالعه روی کلاس اولیها بود. هدف ماایـن بـود کـه بـفهمیم آیـا کـودکان می‌توانند اسلایدهایی راکه قبلاً دیده‌اند به خاطربیاورند؟ پاسخها دوباره "بلی" یا "خیر" بود. لذا مدل دوجمله‌ای برازنده بود. چون بچه‌ها می‌توانستند در نیمی از موارد جواب صحیح را حدس بزنند، ما ناچار بودیم تعدادی اسلاید که برای رسیدن به یک آزمون دقیق کافی باشد بـه آنـها نشان دهیم. برای مثال ممکن بود بچه‌ها دارای قابلیتهایی باشند که مثلاً ۷ اسلاید را از میان ۱۰ اسلاید درست تشخیص دهند اما همان‌طور که جدول ۱ نشان می‌دهد ۷ از ۱۰ می‌تواند آنقدر اتفاق بیفتد که برای قضاوت به عنوان یک پیشامد عادی کافی باشد. ولی اگر ۴۰ اسلاید نشان دهیم پایه شانس (پنجاه، پنجاه) عدد ۲۰ است و بین ۲۰ و ۴۰ جای زیـادی هست تـا کـودکان تواناییهای خود را نشان دهند. به‌دست آوردن ۴۰ از ۴۰ پیشامدی است که در هر ۲ بتوان ۴۰ دفعه فقط یکبار رخ می‌دهد. درست مانند انتخاب ۱۰ از ۱۰ که بنابر جدول ۱ هر ۲ بتوان ۱۰ (۱۰۲۴) دفعه فقط یک بار رخ می دهد اما ۲ بتوان ۴۰ برابراست با ۱۰۲۴ بتوان ۴، و بنابراین نتیجه ۴۰ از ۴۰ تنها یک بار در تعداد دفعاتی بیش از یکهزار مـیلیارد رخ مـی‌دهد. ارزشـیابیِ برآمدهای مـمکن بـا ۴۰ بـار پـرتاب سکـه کـاری مـلال‌آور است امـا خـوشبختانه جـدولهای دوجمله‌ای آماده برای تعداد مختلف آزمایه موجود است.

جدول ۲ قسمتی مربوط از یک جدول دوجـمله‌ای رانشـان مـی‌دهد کـه بـرای ۴۰ بـرآمـد محاسبه شده‌است. وقوع هر برآمدی از ۳۱ شیر یا بـیشتر نـادر خـواهـد بـود ـ ایـن بـرآمـدها روی‌هم‌رفته فقط در ۰/۰۳ درصد اوقات رخ می‌دهد ـ به همین خاطر جدول را مختصر کرده‌ایم. چنانچه شانس ۵ درصد یا کمتراز آن را به‌عنوان معیار عملکردی تعیین کنیم که به اندازه کافی نادر هست که "غیر محتمل" باشد، می‌توانیم در جدول ۲ نقطه‌ای را انتخاب کـنیم کـه چـیزی نزدیک به یک مقطع ۵ درصدی در آنجا واقع شود. مجموع شانسهای وقوع ۲۶ شیر یا بیشتر کمی از ۴ درصـد بـیشتراست (۰/۰۳+۰/۱+۰/۲+۰/۵+۱/۱+۲/۱). هـمچنین مـی‌بینیم کـه شانس وقوع ۲۵ شیر برابر ۳/۷ درصداست. پس اگر نقطه مقطع را ۲۵ انتخاب کنیم شانس وقوع احتمالی تقریباً ۴/۰+۳/۷ یا ۷/۷٪ خواهد بود که از ریسک قابل قبول ۵ درصدی بیشتر است. بنابراین تعداد ۲۶ شیر یا بیشتر را به‌عنوان نزدیکترین عدد به "ایده‌آل" خود انتخاب کردیم.

جدول ۲ـ درستنمایی برآمدهای مختلف برای ۴۰ بار پرتاب سکه

برآمد	درصد دفعاتی که برآمد مشاهده می شود
۴۰ شیر، ۰ خط	
تا	۰/۰۳
۳۱ شیر، ۹ خط	
۳۰ شیر، ۱۰ خط	۰/۱
۲۹ شیر، ۱۱ خط	۰/۲
۲۸ شیر، ۱۲ خط	۰/۵
۲۷ شیر، ۱۳ خط	۱/۱
۲۶ شیر، ۱۴ خط	۲/۱
۲۵ شیر، ۱۵ خط	۳/۷
۲۴ شیر، ۱۶ خط	۵/۷
۲۳ شیر، ۱۷ خط	۸/۱

سه آزمایش

با آمارهایی که به‌این نحو برای شکل دادن به یک طرح خوب به‌کار رفت، گامی به جلو برداشته و سه آزمایش انجام دادیم که همگی از همان طرح پیروی می‌کردند اما هرکدام به مقوله‌ای نسبتاً متفاوت اشاره داشتند:

۱- وضعیت I: اسلایدهای سیاه وسفید. به یک کلاس ۲۳ نفره از دانش آموزان کلاس اول ابتدایی که هوششان اندکی بالاتر از متوسط بود (IQ=۱۰۹)، تعداد ۴۰ اسلاید سیاه و سفید نشان دادیم. به کودکان گفته شد: "به تصویرهایی که به شما نشان می‌دهیم، نگاه کنید چون بعداً ممکن است درباره آنها از شما سئوال کنیم". اسلایدهااز منظره‌های شهری یا چشم‌اندازهای غیرمشهور انتخاب شده بودند و تا آنجاکه می‌دانیم، برای این بچه‌ها ناآشنا بودند.

هر اسلاید به مدت ۱۰ ثانیه نمایش داده شد و بین هر دو اسلاید متوالی ۳ ثانیه وقفه بود. آزمایش در ساعت ۱۰ صبح شروع شد و تا نمایش تمامی ۴۰ اسلاید ادامه داشت. در ساعت

۱۲:۳۰ بعدازظهر همان روز دو پرده نمایش درکنارهم در جلوکلاس نصب شد واز دو پروژکتور برای نمایش ۴۰ جفت اسلاید استفاده شد. مدت نمایش هرجفت اسلاید ۱۰ ثانیه و وقفه بین دو نمایش ۳ ثانیه بود. هریک از ۴۰ اسلایدی که قبلاً نشان داده شده‌بود به‌طور تصادفی در سمت راست یا سمت چپ خود با اسلایدهای جدیدی از همان نوع جفت شده بود. از کودکان خواسته شده بود که در یک دفترچه مخصوص مشخص کنند که از هر جفت اسلاید کدامیک را قبلاً دیده‌اند. همانگونه که جدول ۳ نشان می‌دهد، کودکان به‌طور متوسط ۳۱ اسلاید را شناسایی کردند. کودکان یقیناً دارای بعضی قابلیتها می‌باشند. زیرا همانگونه که قبلاً خاطر نشان کرده‌ایم، شناسایی صحیح ۲۶ مورد یا بیشتر توسط یک کودک بنابر مدل دوجمله‌ای "معنی‌دار" است.

۲- وضعیت II: اسلایدهای رنگی. وضعیت II شبیه وضعیت I بود. جزاینکه یک گروه ۳۰ نفری دیگر از دانش‌آموزان کلاس اول که هوششان اندکی بالاتر از متوسط بود (IQ=۱۰۹) در آزمون شرکت داشتند، و اسلایدها رنگی بودند. (اسلایدهای وضعیتهای I و II بجز وجود رنگ در وضعیت II یکسان بودند.) بچه‌ها بطور متوسط بیش از ۳۴ اسلاید رنگی را درست شناسایی کردند(جدول ۳).

۳ـ وضعیت III. به یک گروه ۲۹ نفری دیگر از دانش‌آموزان کلاس اول که دوباره هوششان اندکی بالاتر از متوسط بود(IQ=۱۰۹) همان ۴۰ اسلاید رنگی مورد استفاده به عنوان اسلایدهای هدف در وضعیت IIنشان داده‌شد. وضعیت III شبیه وضعیت II بود جزاینکه پیش از آزمون تشخیص یک هفته فاصله منظور شده‌بود. کودکان به‌طور متوسط بیش از ۳۱ اسلاید را درست شناسایی کردند.

نخست باید توجه داشت که در هرسه وضعیت تعداد متوسط پاسخهای درست بسیار بالاتر از حتی ریسک ۵ درصدی است که مایل به تحمل آن بودیم. اما "عملکرد متوسط" برای یک گروه ممکن است چنین معنی دهد که مثلاً برای وضعیت I از ۲۳ نفر دانش‌آموز بعضیها بسیارخوب عمل کرده‌اند و مثلاً ۴۰ درصد از ۴۰ آورده‌اند. در حالی‌که بقیه در سطح شانس عمل کردند. به‌این ترتیب اگر ۱۳ دانش‌آموز ۴۰ از ۴۰ و ۱۰ دانش‌آموز ۲۰ از ۴۰ گرفته باشند، بازهم ۱۱ اسلاید اضافه بر متوسط مشاهده شده‌است. شاید نیمی از افراد جامعه دارای چنین

جدول ۳ـ تشخیص اسلاید

	وضعیت I(اسلایدهای سیاه وسفید، آزمون وقفه دار $2\frac{1}{2}$ ساعتی ۲۳ کودک)	وضعیت II (اسلایدهای رنگی، آزمون وقفه دار $2\frac{1}{2}$ ساعتی، ۳۰ کودک)	وضعیت III (اسلایدهای رنگی، آزمون وقفه دار یک هفته‌ای، ۲۹ کودک)
متوسط تعداد اسلایدهایی که درست تشخیص داده شد	۳۱/۰	۳۲/۴	۳۱/۹
دامنه تغییر نمرات بالاتراز۲۰	۳-۱۸	۵-۲۰	۴-۱۹
تعداد و درصد دانش‌آموزانی که امتیازی بالاتر از سطح شانس ۲۰ گرفتند	۲۱ (۹۱درصد)	۲۹ (۹۷درصد)	۲۵ (۸۶ درصد)

قابلیتی باشند ونیم دیگرنباشند. اما دامنه تغییرات نمره افراد در جدول ۳ نشان می‌دهد که تقریباً همه بچه‌ها در هر سه وضعیت نمرات بالاتر از سطح شانس گرفته‌اند.

وضعیت II (اسلاید های رنگی باوقفه کوتاه) بهترین وضعیت به‌نظر می‌رسد. برای اینکه ببینیم آیا این برتری واقعی است به شیوه‌های آماری پیچیده متوسل شدیم[۱] تا معلوم شود آیا برتری وضعیت II نسبت به وضعیتهای I و III قویتر از آن است که نوسانهای شانس بتواند توضیح آن باشد؟ پاسخ سئوال "مثبت" بود. هرگاه همه داده‌ها مشترکاً باهم مطالعه شوند،

۱- *شیوه پیچیده آماری مورد نظر مولف احتمالاً شیوه تجزیه واریانس و آزمون مقایسه میانگینها بوده‌است.مترجم*

تفاوت به‌نفع وضعیت II به‌قدر کافی زیاد هست که بتواند حاکی از یک تفاوت "واقعی" باشد.

پس این آزمایش روی کلاس اولیها، با مطالعه قبلی غیر رسمی ما روی کودکستانیها هماهنگی دارد. کودکان کم سن وسال ظاهراً قادرند مقادیر نسبتاً زیاد اطلاعات تصویری پیچیده را که از هر لحاظ با فعالیتهای روزمره آنها نامرتبط است پردازش ونگهداری کنند. این یافته با مشاهدات کی‌گان و هی‌بر نیز سازگاراست . شرایطی که تحت آنها آزمایش حاضر انجام شد احتمالاً به کم برآوردکردن قابلیتهای مطلق کودکان در کار تشخیص منجر شده‌است زیرا که بعضی از کودکان در تمام مدت آزمایش توجه زیادی مبذول نمی‌کردند. اگر آنها تمرکز کرده بودند ـ مثلاً اگر ما آنها را یکی یکی در یک اتاق بدون عوامل حواس پرتی مورد آزمون قرار می‌دادیم ـ به احتمال زیاد بهتر از این عمل می‌کردند. لیکن شرایط این آزمایشها مشابه شرایط موجود در کلاسهای درس معمول است که در آنها فعالیتهای آموزشی مشابه انجام می‌شود.

تحقیقات بیشتر

یکی دیگر از پژوهشهای اخیر (فینمن و آنت ویسل) اشاره می‌کند که توانایی پردازش اطلاعات تصویری ممکن است در کودکان خردسال به همان خوبی افراد بالغ باشد. وقتی از تعدادی دانش‌آموز کلاس سومی خواسته شد که اسلایدهای چهره انسانها را با همان شیوه کلی توصیف شده در بالا درمورد آزمایش اسلایدهای مناظر طبیعی نگاه کنند و تشخیص دهند، نتیجه کارشان تقریباً با نتیجه کار دانشجویان دانشگاه برابر بود.

اینکه چگونه کودکان اطلاعات تصویری را در حافظه خود ذخیره می‌کنند؟ مطلبی مهم برای پژوهشهای بعدی است . بطور وضوح، رنگ باعث بالا رفتن نتیجه کار می‌شود. به‌این مفهوم که زمان فراموشی یک هفته‌ای را جبران می‌کند. اما این موضوع ممکن است به‌دلیل بیشتر جلب کردن توجه کودکان باشد تا به‌دلیل تغییر در چگونگی ذخیره‌سازی اطلاعات.

یک توضیح برای اینکه چرا کودکان می‌توانند اسلایدهای بسیار زیادی رابه‌درستی تشخیص دهند؛ در غنای اطلاعات موجود اسلایدها نهفته است. یک اسلاید حاوی اطلاعاتی است که می‌توان آنها را به لحاظ بسیاری از اشکال مختلف به‌یاد آورد. منظره یک بندرگاه، برای مثال، دارای آب، خط ساحل باترکیب خاص، قایقها با ترتیب مکانی خاص و غیره است. اگر این

اسلاید تنها منظره بندری از نوع خود در مجموعه اسلایدها باشد، کدگذاری آن به‌عنوان یک "منظره بندری" برای میسر ساختن تشخیص آن از میان سایراسلایدهای درون مجموعه کافی است. اما اگر تمام اسلایدها صرفاً مناظر بندری باشند، کار تشخیص بی‌اندازه مشکلتر است. می‌توانیم آثار مشاهده شده در آزمایشهای خود رابا این تذکر توضیح دهیم که اسلایدها به‌قدر کافی با یکدیگر تفاوت داشته‌اند که به‌یاد آوردن یک یا دو جنبه از هریک از آنها برای تشخیصشان کافی بوده باشد. مثلاً منظره‌های بندرگاهی به اندازه کافی تفاوت داشته‌اند که بتوانند کاملاً متمایز باشند.

نکته اصلی این مقاله جلب کردن توجه همگان به تواناییهای نسبتاً زیاد کودکان خردسال در تشخیص اسلایدها پس از مدتی نسبتاً طولانی و نیز تأکید براین نکته است که عکسها ذاتاً می‌توانند مقدار قابل توجهی از اطلاعات راکه کودکان ذخیره می‌کنند، انتقال دهند. این موضوع می‌تواند یک عامل مهم در آموزش کودکان در دوران پیش از سوادآموزی یا آموزش کودکانی باشد که به دلایل مختلف، از جمله مشکل تکلم، نمی‌توانند در کار خواندن مانند سایر بچه‌ها پیشرفت کنند.

مسائل

۱ـ چگونه می‌توان مطالب مورد استفاده در این آزمایشها را بابرنامه‌های درسی مدارس بیشتر مرتبط ساخت ؟

۲ـ آیا فکر می‌کنید دانش آموزان دبستانی کلاس اول عکس آدمها واعداد را در مقایسه با منظره‌های طبیعی بهتر به‌یاد می آورند یابدتر؟

۳ـ فکر می‌کنید تأثیرات حافظه‌ای معنی‌دار تا چه مدت بعد از یک هفته در ذهن باقی می‌ماند؟ پاسخ خود را توضیح دهید.

منابع

S. Feinman and D. R. Entwisle. 1976. "Children's Ability to Recognize Other Children's Faces".*Child Development 47: 506-510.*

R. N. Haber. 1970. "How We Remember What We See". *Scientific American 220: 104-112.*

J. Kagan. 1970. "The Determinants of Attention in the Infant." *American Scientist 58: 298-306.*



معنی واژه‌ها

جوزف ب. کروسکال[1] شرکت آزمایشگاههای تلفن بل

چگونه می‌توانیم از آمار برای به چنگ آوردن مفهوم فرّاری که "معنی" نامیده می‌شود، استفاده کنیم؟ به فرض این که چنین امکانی وجود داشته‌باشد، آیا با این کار تغزل شعر و جذابیت کلام را از بین نمی‌بریم؟

می‌توان معنی واژه‌ها را با روشهای منظم آماری مورد مطالعه قرارداد. در اینجا دو بررسی

۱ـ *Joseph B. Kruskal*

جالب از این نوع را به بحث می‌گذاریم. اما متوجه خواهید شد که تغزل واژه‌ها از علم در امان است.

اشخاصی که دارای تفکر عملی هستند ممکن است بپرسند که چرا باید این همه زحمت بکشیم تا معنی واژه‌ای را مشخص کنیم؟ آیا زبان پس از هزاران سال تکامل طبیعی نمی‌تواند خود به خوبی وظیفه‌اش را انجام دهد؟ خیلی ساده است: خیر. هر مأمور پذیرش در هر دانشکده‌ای به مشکل بودن تعبیر توصیه‌های مکتوب مدرسین شهادت خواهدداد و دانشکده‌های بسیاری هستند که برای تخفیف این مشکل از فرمهای استاندارد استفاده می‌کنند. مطالعاتی از آنگونه که در اینجا مورد بحث قرار می‌گیرد، ممکن است روزی مراوده بین مدرسین، محصلین و دانشکده‌ها را بهبود بخشد. این امر حتی می‌تواند در برقراری صلح میان ملتها نیز سهمی ادا کند.

اعداد در وضعیتهای دیگر برای مدتی مدید و غالباً به نحوی بسیار مؤثر توانسته‌اند مکمل واژه‌های وصفی باشند. استفاده از **اینچ**، **پوند** ودرجه **فارنهایت** مسلماً نسبت به **نسبتاً کوتاه**، **بسیارسبک**، و **تقریباً داغ** دارای مزیتهایی است. زمانی معنای **تب‌دار** به همان اندازه معنای **جذاب** یا **ترشرو** مبهم بود. عرضه تب‌سنج، نعمتی برای سلامتی انسان بوده‌است.

امروزه اعداد در زمینه‌های ظریفتر ازقبیل نغمه‌های موسیقی، اصوات سخنوری و هوش موارد استفاده زیادی دارند. رنگ یکی ازنمونه‌های بارز است. بسیاری از مردم تصور می‌کنند که رنگها چنان ظریفند که در توصیف عددی نمی‌گنجند. با وجود این سیستم عددی در این زمینه کاربرد روزمره دارد. ـدر بیشتر این سیستمها هررنگ با سه عدد توصیف می‌شود. ـ(ن.ک. مقاله ونس Vance درمورد تعیین تحمل رنگ.) این سیستمها در پدیدآمدن تلویزیونهای رنگی و اصلاح عکسهای رنگی نقشی حیاتی ایفا نموده‌اند و به صنعت رنگ، گریم و سایر زمینه‌ها نیز کمک کرده‌اند. بعضی از هنرمندان نیز این راه و روش علمی را برای توصیف رنگها مؤثر تشخیص داده‌اند. در این علم چیزی نیست که حیطه هنر آنان راکاهش دهد یااز لذت شما از تماشای نقاشیهای آنها، تماشای یک فیلم رنگی یا تماشای غروب خورشید بکاهد.

ولی این مثالهای قدیمی از یک نظر، یعنی روش انجام اندازه‌گیریها با مطالعاتی که درنظر است دراینجا بحث شوند، تفاوت دارند. با ابزار فیزیکی می‌توان ارتفاع، وزن و حرارت را

اندازه‌گیری نمود. با استفاده از ابزارهای دقیق ترکیب و تیره‌کردن رنگهای روشن می‌توان مسئله رنگها را رفع و رجوع کرد و هوش را با یک آزمون استاندارد می‌سنجند. ولی درکاری که ذیلاً شرح داده‌می‌شود، به جای شیء مورد مطالعه، افراد نقش ابزار اندازه‌گیری را ایفا می‌کنند. برداشت ذهنی افراد که به‌طور بسیاردقیق به‌دست آمده و تحلیل می‌شود، کار اندازه‌گیری را انجام می‌دهد.

ویژگیهای شخصیتی

مطالعه‌ای که توسط روزنبرگ، *نلسون* و *ویوکانانتان* در سال ۱۹۶۸ انجام شد به‌معنای ۶۴ واژه مربوط به ویژگیهای شخصیتی می‌پردازد؛ ازقبیل **برانگیخته، صمیمی، محتاط، زودرنج و شاد** که واژه‌هایی هستند که اشخاص را توصیف می‌کنند. این بررسی با استفاده از یک فن نوین آماری به تأمین نظم درجایی که یافتن نظم مشکل است، کمک می‌کند.

همه این واژه‌ها را به‌کار می‌برند و آنها برای همه ما بار معنایی دارند. علاوه براین حائز اهمیت هم هستند. تصور کنید که برحسب اتفاق بشنوید کسی شما را به‌عنوان "بی‌ذوق" یا "خوش قلب" توصیف کند و فکر کنید احساسی که این دو توصیف در شما ایجاد می‌کند چقدر متفاوت خواهدبود! وقتی از این واژه‌ها برای معرفی دونفر باهم برای ازدواج احتمالی یا توصیه‌هایی که معلمها می‌نویسند، استفاده می‌شود انتخاب هریک از واژه‌ها نتایج متفاوت زیادی ایجاد می‌کند.

با وجود این، استفاده مردم از این واژه‌ها متفاوت است. اگر دو نفر یک دوست مشترک را توصیف کنند یا درمورد یک شخصیت سینمایی نظر بدهند، چنانچه برای این کار از واژه‌های **دقیقاً** یکسان استفاده کنند، حیرتزده خواهیم شد. در روانپزشکی نیز همین عدم حتمیت در مورد معنا نه‌تنها برای واژه‌های عادی بلکه برای عبارات فنی ازقبیل شیزوییدی[۱] و اوتیستیک[۲] وجود دارد. مقاله‌های مفصلی درباره معنایی که این واژه‌ها دارند یا باید داشته باشند نوشته شده‌است.

۱ـ مبتلا به شیزوفرنی

۲ـ مبتلا به اوهام

مشکل از همین جا ناشی می‌شود. روشن کردن معنای واژه‌هایی که با فراوانی زیاد به‌کار برده می‌شوند و اهمیت بسیاری دارند، می‌تواند احتمالاً برای روانپزشکان و قطعاً برای مسئولان پذیرش دانشکده‌ها، مؤسساتی که قرارهای ازدواج کامپیوتری را ترتیب می‌دهند و خدا می‌داند دیگر برای چه کسانی سودمند باشد.

گریزی به مقیاس‌بندی چندبعدی

ولی چگونه می‌توان معنای واژه‌ها را روشن کرد؟ اینجاست که آمار برای اولین بار وارد می‌شود و رهیافتی برای مسئله و روشی برای تحلیل ارائه می‌نماید. ابتدا با وضوح بیشتری به ارزیابی مسئله می‌پردازیم.

۱- **چه می‌خواهیم؟** توصیف منظمی از صفتهای شخصیتی براساس معنای آنها. درصورت موفقیت توصیفی به‌دست خواهیم آورد که درست به اندازه یک نقشه شهری که در آن جاده‌ها، مسیر اتوبوسها، پارکها و امثال آن برمبنای موقعیت مکانی آنها مشخص شده، سودمند است.

۲-برای رسیدن به این توصیف چه ابزارهایی داریم؟ هیچ چیز مگر طرز استفاده مردم از واژه‌ها. چیزی نداریم که با خط‌کش یا دماسنج اندازه‌گیری کنیم. فقط می‌توانیم چگونگی استفاده مردم از واژه‌ها یا واکنش آنها را مشاهده نماییم.

روش آماری تازه توسعه یافته‌ای که **مقیاس‌بندی چندبعدی** نامیده می‌شود در مسائلی ازاین دست ارزش ویژه‌ای دارد. این حقیقت که روزنبرگ و همکارانش ازاین روش اطلاع داشته‌اند به برانگیختن آزمایش آنها کمک کرد. علم آمار نیز رهیافت آن را فراهم ساخت.

مقیاس‌بندی چندبعدی از چه چیزهایی تشکیل شده‌است؟ این روش را می‌توان در سه بخش به بهترین وجه شرح داد: آنچه داخل می‌شود، آنچه بیرون می‌آید و آنچه در این فاصله روی می‌دهد. آنچه در مقیاس‌بندی چندبعدی داخل می‌شود **تشابهات یا عدم تشابهات** میان اشیای گوناگون از یک نوع است. برای هر جفت از اشیا (در این مورد برای هرجفت واژه از ۶۴ واژه مورداستفاده) عددی میزان تشابهات یا عدم تشابهات میان دو شی را توصیف می‌کند. این قبیل اعداد به روشهای بسیاری تهیه می‌شود. برای مثال، تشابه میان دورنگ را می‌توان به این ترتیب فراهم نمود که یک نفر به رنگها نگاه کند و میزان شباهت رنگها را به مقیاس ۱ تا ۵

توصیف نماید. (اگر رقم بزرگتر حاکی از شباهت بیشتر باشد، آن اعداد را "تشابهات" و درصورت عکس آن "عدم تشابهات" می‌نامیم.) در بررسی حاضر روش پیچیده‌تری به‌کار رفت که بعداً شرح خواهم داد.

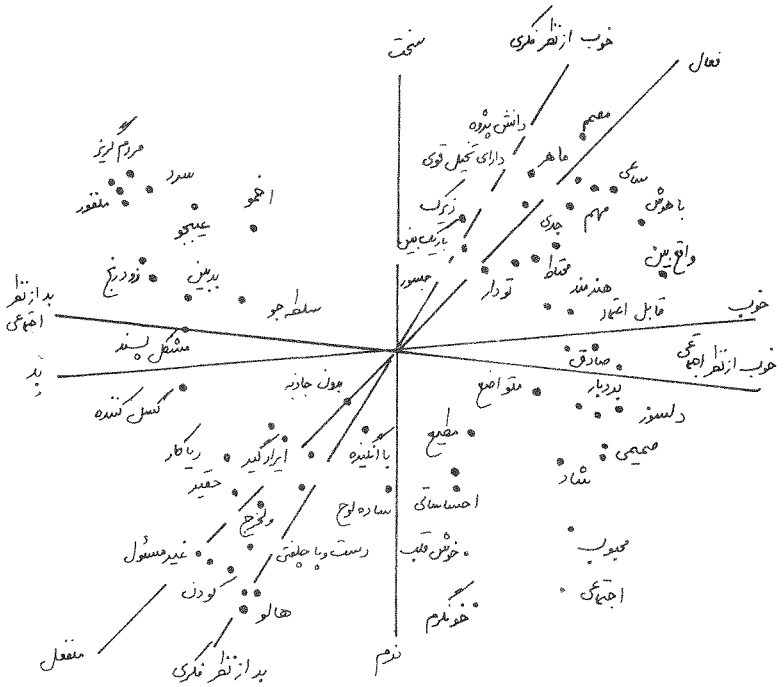
آنچه از مقیاس‌بندی چندبعدی بیرون می‌آید یک **نقشه** یا **تصویر** ازقبیل شکل ۱ است. (فعلاً خطوط موجود در شکل را نادیده بگیرید و فقط به پیکره‌بندی نقطه‌ها توجه کنید.) به‌طور خلاصه، این روش هر شیء را در موقعیت خاصی قرار می‌دهد. نقشه‌ای که با مقیاس‌بندی تهیه می‌شود اگرچه یک نقشه واقعی نیست، دارای بسیاری از ویژگیهای نقشه‌های واقعی است. برای مثال می‌توانیم آن را آزادانه بچرخانیم و از هرجهتی که دوست داریم به آن نگاه کنیم. این نقشه جهتی ندارد که واقعاً بالای نقشه باشد. همچنین می‌توانیم آن را آزادانه بزرگ (یا کوچک) کنیم یا به اصطلاح مقیاس آنرا تغییر دهیم، هرچند که نقشه‌های ما درواقع هیچ مقیاسی ندارند. بزرگی نقشه را راحتی ما تعیین می‌کند.

در فاصله ورودی و خروجی در مقیاس‌بندی چندبعدی چه اتفاقی می‌افتد؟ گفتنش آسان اما انجام دادنش سخت است. ما اشیا را با این هدف روی نقشه قرار می‌دهیم که اشیایی که نزدیک هم هستند، بسیار شبیه و اشیایی که دور ازهم قرار می‌گیرند، بسیار متفاوت باشند. به عبارت دیگر، فاصله‌های کوتاه باید باعدم تشابهات کم و فاصله‌های زیاد باعدم تشابهات زیاد متناظر باشند. درمورد تشابهات برعکس است. به‌طور خلاصه، هدف ضمن ایجاد نقشه این است که بین فاصله‌های نقشه و ورودیها (تشابهات یاعدم تشابهات) ارتباط خوبی برقرار شود.

این که ارتباط مزبور تا چه اندازه می‌تواند خوب باشد به ورودی بستگی دارد. کیفیت این ارتباط در مقیاس‌بندی چندبعدی از اهمیت زیادی برخوردار است. در مواقعی که کیفیت ارتباط خوب است، آنچه از نقشه دستگیرمان می‌شود مفید خواهد بود. وقتی کیفیت نقشه بسیاربد است، بهتر آن است که به‌کلی آنرا دور بیندازیم.

فرایند پیچیده ترسیم این نقشه، تقریباً همیشه با کامپیوتر انجام می‌پذیرد. نیازی نیست که کاربر برنامه کامپیوتر این فرایند را درک کند. درست مثل راننده‌ای که نیازی ندارد طرزکار موتور خودروی خود را درک کند. (امروزه مقیاس‌بندی چندبعدی به‌طور مکرر باکامپیوترهای کوچک شخصی اجرا می‌شود.)

شکل ۱ـ نقشه مقیاس‌بندی چندبعدی با محورهای روندها



ادامه صفتهای شخصیتی

روزنبرگ و همکارانش برای به‌دست آوردن تشابهات، مقداری نوار کاغذی در اختیار ۶۹ سوژه (دانشجوی دانشگاه) قراردادند که دارای ۶۴ واژه از صفتهای شخصیتی بود. روی هر نوار فقط یک واژه درج شده‌بود و هر سوژه یک بسته کامل ۶۴ نواری دریافت می‌کرد. از سوژه‌ها خواسته شد تا نوارهای بسته خود را در ده گروه تنظیم کنند به‌طوری که واژه‌های هر گروه بتواند به‌طور معقولی تنها یک نفر را توصیف نماید. (هر نوار را می‌شد فقط دریک گروه قرار داد.)

در میان شیوه‌های متعددی که روان‌شناسان می‌توانند برای به‌دست آوردن تشابهات به‌کار برند، این شیوه از همه سریعتر است.

پژوهشگران با شمارش سوژه‌هایی که دو واژه را باهم در یک گروه قرار داده‌بودند، برای آن دو واژه یک **نمره** توافق به‌دست آوردند. برای مثال، ۳۷ سوژه واژه‌های **قابل اعتماد** و **صادق** را در یک گروه قرار داده‌بودند، بنابراین **نمره** توافق این دو واژه ۳۷ است. چند نمره دیگر به شرح زیر است:

۱- **قابل اعتماد و صادق** امتیاز = ۳۷

۲- **دست وپاچلفتی و ساده‌لوح** امتیاز = ۱۹

۳- **احساساتی و مشکل‌پسند** امتیاز = ۲

۴- **خوش قلب و زودرنج** امتیاز = ۱

۵- **محبوب و مردم‌گریز** امیتاز = صفر

نمره‌های توافق یک آرایه مربعی ۶۴ در ۶۴ راتشکیل می‌دهند. نمره‌های توافق ۶۴گانه هر واژه با خودش مورد توجه نیست زیرا برابر با تعداد سوژه‌ها یعنی ۶۹ است. تعداد مقادیر باقیمانده ۴۰۳۲=۶۴-۶۴×۶۴ به‌صورت زوجهایی از مقدار مساوی خواهدبود؛ زیرا برای مثال نمره توافق **خونگرم** با **زیرک** درست با نمره توافق **زیرک** با **خونگرم** برابر است. به این ترتیب فقط ۲۰۱۶=۲:۴۰۳۲ نمره توافق متمایز مطرح خواهدبود.

نمره‌های توافق همان تشابهات هستند و می‌توان از آنها به‌عنوان ورودی مقیاس‌بندی چندبعدی استفاده کرد. درواقع نویسندگان مقاله این روش را در مراحل اولیه تحلیل خود امتحان کردند ولی نتایج آن روشن نبود. خوشبختانه این مولفان توانستند مشکل را دریابند: بسیاری از نمره‌های توافق به‌قدری کوچکند که مقایسه آنها قابلیت اعتماد چندانی ندارد. نمره‌هایی که در بالا ذکر شد، داد می‌زنند که تشابه جفت شماره ۲ بیشتر از جفت شماره ۳ است ولی به‌آرامی زمزمه می‌کنند که جفت شماره ۳ نسبت به جفت شماره ۴ تشابه بیشتری دارد. با بیش از نیمی از نمره‌های کمتراز ۱۰ دقت بیشتر کمک ارزنده‌ای خواهدبود. (اتفاقاً این نوع امتحان و بهبود اغلب هنگامی اتفاق می‌افتد که داده‌ها به روشی نوین تحلیل می‌شوند.)

خوشبختانه بااستفاده از داده‌ها به شیوه‌ای متفاوت، دقت بیشتری حاصل شد. برای مثال، هر دوواژه دلخواه را درنظر بگیرید، مثلاً **احساساتی و مشکل‌پسند**. حال از دو ردیف برای فهرست‌کردن نمره‌های توافق آنها با تمام ۶۴ واژه استفاده می‌کنیم:

	هالو	مبتکر	مردد	مطیع	خونسرد	بردبار	...
احساساتی	۲	۶	۵	۱۵	۳	۲۲	...
مشکل‌پسند	۶	۲	۱۸	۱۰	۹	۲	...

اگر دو واژه معنای مشابهی داشته باشند، نه‌تنها نمره توافق آنها باید زیاد باشد بلکه **نمره‌های توافق آنها با یک واژه سوم نیز باید تقریباً یکسان باشد.** به این ترتیب برای دو واژه مشابه، اعداد متناظر در این دو ردیف نباید تفاوت زیادی داشته باشند. اگر تفاضل بین این دو سطر از اعداد را به‌طور معقول اندازه‌گیری کنیم، در آن‌صورت این مقیاس ثانویه (یا مشتق) باید حاکی از عدم تشابه (به جای تشابه) در معنای این دو واژه باشد. به تبع این ایده، عدم تشابه میان **احساساتی** و **مشکل‌پسند** با جمع‌کردن مربع تفاضلها به‌دست می‌آید.

$$(۶-۲)^۲+(۲-۶)^۲+(۱۸-۵)^۲+(۱۵-۱۰)^۲+(۳-۹)^۲+(۲۲-۲)^۲+...$$

هرچه این عدم تشابه بیشتر باشد، تفاوت بین واژه‌ها بیشتر است.

نقشه شکل ۱ از کاربرد مقیاس‌بندی چندبعدی در مورد عدم تشابهات[۱] به‌دست آمد. قبل از این که بتوانیم احساس کنیم که چیزی سودمند یا روشنگر یافته‌ایم، بایستی این نقشه را به دقت بررسی، تفسیر و ارزیابی کنیم. برای اینکه ببینید چه گروههایی از واژه‌ها باهم می‌آیند و چگونه معانی در سراسر این شکل به‌گونه‌ای سیستماتیک تغییر می‌کنند، بررسی دقیق شکل ۱ سودمند خواهدبود. اگر پژوهشگری در این مرحله چیزی معقول بیابد، می‌توان ادعا نمود که موفقیتی ابتدایی حاصل شده‌است. اگر نقشه تصویر منظمی ارائه نماید که قبلاً چیزی درمورد آن نمی‌دانستیم، پیشرفت حاصل شده‌است.

ویژگیهای شخصیتی ـ مرحله دوم

نویسندگان مقاله در شکل ۱ ساختار قابل توجهی یافتند:

۱- به علت محدودیتهای برنامه کامپیوتر (که دیگر مطرح نیست) فقط ۶۰ واژه عملاً در مقیاس‌بندی چندبعدی گنجانده شد.

به‌نظر می‌رسد که وقتی از گوشه راست فوقانی به گوشه تحتانی چپ می‌رویم از مطلوبیت صفتها برای فعالیتهای ذهنی به‌طرزی سیستماتیک کاسته می‌شود. تغییر سیستماتیک دیگری با رفتن از گوشه چپ فوقانی به طرف گوشه راست تحتانی اتفاق می‌افتد که در این مورد مطلوبیت اجتماعی ویژگیها افزایش می‌یابد.

توجه داشته‌باشید که در گوشه راست فوقانی **زیرک**، **ماهر** و **دانش‌پژوه** و در گوشه مقابل آن **کودن**، **هالو** و **سبک‌سر** قرار دارند. گوشه چپ فوقانی شامل **مردم گریز**، **بی‌ذوق** و **منفور** و گوشه مقابل آن شامل **اجتماعی**، **محبوب** و **خونگرم** است. یکی از راههای مناسب برای نشان دادن این مطلب رسم محور مختصات مانند شکل ۱ است. موقعیت هر ویژگی در طول محور بر جایگاه آن ویژگی با توجه به روند موردنظر دلالت می‌کند. (برای به‌دست آوردن موقعیت یک واژه در طول محور، خطی را از موقعیت واژه در روی نقشه عمود برمحور رسم می‌کنیم. فرقی نمی‌کند که فاصله آن واژه تا محور چقدر است.)

این تعبیر باعث طرد تعبیرهای دیگر نمی‌شود. یک جهت مشخص (ازقبیل از راست فوقانی به گوشه چپ تحتانی) می‌تواند تعبیرهای متعددی داشته‌باشد و جهتهای دیگری نیز ممکن است دارای معنی باشند. درمورد بهترین راه تفسیر این شکل، همه همعقیده نیستند.

نویسندگان مقاله به تعبیرهای شخصی خود بسنده نکردند. سوژه‌هایی غیر از آنها که این نمره توافق را ارائه نموده بودند، صفتهای ۶۴گانه را براساس پنج مقیاس متفاوت رتبه‌بندی کردند. هر سوژه فقط به یک صفت می‌پرداخت. برای مثال ۳۴ سوژه هریک از صفتها را روی یک مقیاس ۷ درجه‌ای "برمبنای این که آیا شخصی که واجد هریک از این صفتهاست در فعالیتهای فکری خود خوب یا بد عمل می‌کند" درجه‌بندی کردند. مقیاس دیگری به فعالیتهای اجتماعی می‌پرداخت . سه مقیاس دیگر یعنی خوب ـ بد، نرم ـ سخت، فعال ـ منفعل نیز مورد استفاده قرارگرفت. (این مقیاسها از میان مقیاسهای افتراقی معنایی آزگود[۱]، سوسی[۲] و

۱ـ *Osgood*

۲ـ *Suci*

تاننبوم[1]، ۱۹۶۷، انتخاب شده بودند.

نویسندگان مقاله برای هر مقیاس چندین اقدام به عمل آوردند. آنها میانه درجه‌بندیهای سوژه‌ها را برای هر صفت پیدا کردند. (میانه درجه‌بندی عبارت است از مقدار میانی که از نصف مقادیر کمتر و از نصف مقادیر بیشتراست.) بااستفاده از یک روش آماری به نام **رگرسیون خطی**، محوری راکه به بهترین شکل با درجه‌بندیهای میانه جور بود، پیدا کردند. پنج محوری که در شکل ۱ ترسیم شده از این شیوه به‌دست آمده‌است. آزمایش دوم به وضوح دو روندی راکه نویسندگان شرح داده‌اند تأیید می‌کند و روندهای دیگری را نیز به نمایش می‌گذارد. (بازهم ممکن است روندهای معتبر دیگری وجود داشته‌باشد.)

آموزنده است که قدرت هر روند بازبینی شود. به عبارت دیگر، بررسی شود که جورشدن درجه‌بندیهای میانه در هر مقیاس با جایگاه واژه‌ها در طول محور مربوط به خود تا چه اندازه مناسب است؟ برای انجام این مقایسه، **ضریب همبستگی** میان نمره‌های میانه برای واژه در این مقیاس وموقعیت آنها را روی محور محاسبه می‌کنیم. این معیار آماری که کاربرد گسترده‌ای دارد حاکی ازآن است که دو مجموعه از اعداد تا چه اندازه همراه با هم تغییر می‌کنند. این ضریب همیشه بین ۱- تا ۱+ است و ۱+ بر توافق کامل در جهت تغییر اعداد باهم دلالت می‌کند. ضرایب همبستگی که در شکل ۱ نشان داده شده‌اند، عبارتنداز:

خوب ـ بد ازنظراجتماعی	۰/۹۵
خوب ـ بد ازنظر فکری	۰/۹۱
خوب ـ بد	۰/۸۸
سخت ـ نرم	۰/۷۰
فعال ـ منفعل	۰/۴۳

این اعداد نشان می‌دهند که سه مقیاس اول واقعاً بسیارخوب با نقشه جورند، مقیاس چهارم نسبتاً خوب جور شده‌است و مقیاس آخر به طرزی بسیار متوسط جورشده است (هرچند که عدد ۰/۴۳ به اندازه کافی بزرگ هست که به‌طور قطعی ثابت کند که ارتباط چندی میان این

۱ـ *Tannenbaum*

مقیاس و نقشه وجود دارد).

نتیجه آنکه این تحلیل، نقطه شروع بااهمیتی برای تدارک شرح سیستماتیکی از معانی مجموعه‌های قابل ملاحظه‌ای از واژه‌ها را برای صفتهای شخصیتی فراهم می‌نماید. نویسندگان مقاله یک نقشه منحصربه‌فرد ایجاد کرده‌اند که نشان می‌دهد چگونه سوژه‌ها این واژه‌ها را در وظایف متعددگوناگون به‌کار می‌برند و درک می‌کنند. البته فقط **بخشی** از معنای این واژه‌ها روی نقشه شرح داده شده‌است. از سایر جنبه‌ها به‌کلی صرف‌نظر شده‌است. به این ترتیب **ماهر، ساعی و زیرک** در روی نقشه بسیار نزدیک به هم قرار دارند زیرا غالباً یک شخص را توصیف می‌کنند اگرچه معنای آنها از سایر جنبه‌ها بسیار متفاوت است. همچنین نقشه فقط داده‌های **انبوهه** را برمبنای سوژه‌های متعدد شرح می‌دهد. این امر باعث می‌شود که تغییرات فردی در استفاده از این واژه‌ها ناپدید شده و نادیده گرفته شود.

با وجود این، موضوع تکاندهنده وروشنگر آن است که یک نقشه ساده می‌تواند بخش مهمی از معنای واژه‌های گوناگون بسیاری را درمورد صفتهای شخصیتی به‌چنگ آورد. پیشرفت واقعی حاصل شده‌است.

مقیاس‌بندی چندبعدی ـ تذکری دیگر

نقشه‌هایی که به روش مقیاس‌بندی چندبعدی تهیه می‌شوند، ضرورتی ندارد که نقشه‌های دوبعدی معمولی باشند. این نقشه‌ها ممکن است سه‌بعدی، چهاربعدی یا حتی دارای ابعاد بیشتر باشند (وغالباً هم چنین هستند). حقیقت آن است که روزنبرگ وهمکارانش قبل از اینکه تصمیم بگیرند نقشه مناسب کدام است؛ عملاً یک نقشه یک بعدی، یک نقشه دوبعدی (شکل ۱) و یک نقشه سه‌بعدی تهیه کردند. (شرح چگونگی رسیدن به این انتخاب ما را از مسیر خود دور می‌کند.)

قبلاً متذکر شدم که کیفیت رابطه بین فاصله‌های نقشه و عدم تشابهات از اهمیت زیادی برخوردار است. اگر رابطه بسیاربد باشد، بهتر است نقشه را دور انداخت. تصویری از این رابطه (برای نقشه شکل ۱) به‌عنوان طرح اصلی شکل ۲ دیده می‌شود. این طرح دارای ۱۷۷۰=۲:(۵۹×۶۰) نقطه یعنی نقطه برای هرجفت از واژه‌هاست. برای هرجفت، عدم

تشابهات محور عمودی مختصات یک نقطه و فاصله بین دوواژه (که ازروی شکل ۱ با استفاده از یک خط‌کش اندازه‌گیری می‌شود) محور افقی مختصات را به‌دست می‌دهد. آشکار است که رابطه بسیار خوبی عدم تشابهات و فاصله را به هم مرتبط می‌سازد. دو طرح دیگر شکل ۲ تصاویر متناظر را برمبنای نقشه‌های یک‌بعدی و سه‌بعدی ارائه می‌نماید.

ملتها

اسامی ملتها نیز مانند سایر واژه‌ها پیوندهایی را به ذهن متبادر می‌کند. دو آزمایش (یکی توسط ویش[۱] درسال ۱۹۷۰ با استفاده از ۱۲ ملت و دیگری توسط ویش ـ دویچ[۲] و باینر[۳] در سالهای ۱۹۷۰ و ۱۹۷۲ با استفاده از ۲۱ ملت) چگونگی درک مردم از ملتها و روابط متقابل میان آنها را مورد بررسی دقیق قرار داده‌اند. مهمترین سئوالاتی که برای سوژه‌ها مطرح شد راجع به این بود که ملتهای گوناگون چقدر تشابه دارند. نویسندگان مقاله‌های بالا با روشهایی که تاحدودی شبیه روشهای فوق بود و بااستفاده از دانشجویان دانشگاه کلمبیا به‌عنوان سوژه، به حقایق متعدد و جالب‌توجهی پی بردند.

تعجبی ندارد که از میان ویژگیهای متعدد ملتها که برتشابه قابل درک آنها اثر می‌گذارد دو موردی که به‌عنوان مؤثرترین ویژگی خودنمایی می‌کنند (در هر دو آزمایش) عبارت بودند از صف‌بندی سیاسی (از "متحد آمریکا" گرفته تا "متحد روسیه") و توسعه اقتصادی. ویژگی پیچیده‌تری که در آزمایش دوم اهمیتی ازخود نشان داد "موقعیت جغرافیایی، نژاد و فرهنگ" بود که ملتها برمبنای آن به چهار گروه اروپایی، اسپانیایی، افریقایی و آسیایی تفکیک می‌شوند.

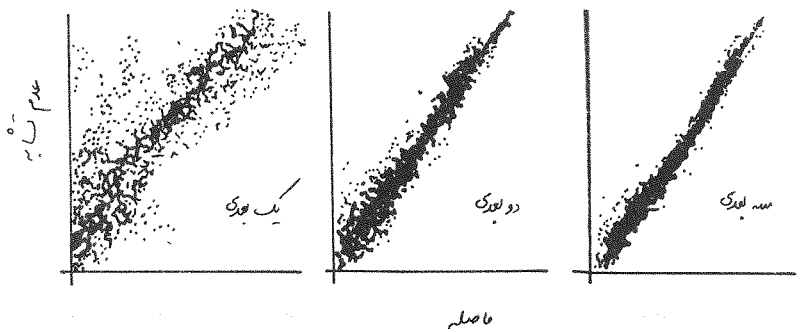
ولی اهمیت ویژگیها از نظر افراد مختلف باعث حیرت بود. در آزمایش اول ۱۸ سوژه براساس این که کدامیک از توصیه‌های پنج‌گانه را دررابطه با جنگ ویتنام انتخاب کرده‌بودند به صلح‌جویان، میانه‌روها و جنگ‌طلبان تقسیم شدند. هر شش نفر جنگ‌طلب اهمیت بیشتری (در قضاوتهای خود در مورد تشابهات) برای صف‌بندی سیاسی قائل شدند؛ درحالی که هرهفت

۱ـ *Wish*

۲ـ *Deuwech*

۳ـ *Biener*

شکل ۲ـ روابط میان فاصله‌ها در روی نقشه و عدم تشابهات واژه‌ها



نفر صلح‌جو به توسعه اقتصادی بیشتر اهمیت می‌دادند. به این ترتیب گرایشهای سیاسی حتی بر پرسشی مانند دو ملت تا چه اندازه تشابه دارند؟ نیز اثر می‌گذارد.

در آزمایش دوم ۷۵ سوژه از هشت کشور همان تأثیر سیستماتیک را به نمایش گذاشتند. همچنین زنان درمقایسه با مردان و دانشجویان کشورهای کمترتوسعه‌یافته در مقایسه با دانشجویان کشورهای توسعه‌یافته اهمیت بیشتری برای صف‌بندی سیاسی قائل بوده‌اند. دانشجویان ازنظر دانش خود درباره ۲۱ ملت نیز مورد آزمون قرارگرفتند و سپس به گروههای مطلع‌تر و کم‌اطلاع‌تر تقسیم شدند. در داخل این گروهها نیز همان آثار نشان داده‌شد؛ گرچه در میان گروه کم‌اطلاع‌تر این امر با شدت بیشتری بروز کرد.

مسائل

۱ـ مقیاسهای مورد استفاده در بررسی صفتهای شخصیتی از چه‌نظر با مقیاسهایی نظیر پوند یا اینچ تفاوت دارند؟

۲ـ کدام کیفیت در تهیه ورودیها برای مقیاس‌بندی چندبعدی به‌صورت کمیت مطرح شد؟

۳ـ فرض کنید روزنبرگ وهمکارانش (به جای ۶۴ واژه) معنای ۵۰ واژه مربوط به صفتهای شخصیتی را مورد مطالعه قرار داده‌بودند. در این صورت چند نمره توافق متمایز مطرح می‌شد؟

۴ـ آزمایشی از نوع آزمایش روزنبرگ را با استفاده از تنها ۴ واژه با نمره‌های توافق زیر درنظر بگیرید:

	خطاکار	خونگرم	پست	کامل
خطاکار	-	۱	۴	۲
خونگرم	۱	-	۰	۴
پست	۴	۰	-	۱
کامل	۲	۴	۱	-

عدم تشابه میان خونگرم و کامل چقدر است؟

۵ـ به شکل ۱ رجوع کنید.

الف) کدام ۵ واژه یک گروه کرانگینی خوب ـ نرم را تشکیل می‌دهند؟

ب) چگونه **اخمو و عیبجو** را توصیف می‌کنید؟

۶ـ جایگاه **ژرف‌اندیش و باانگیزه** را در شکل ۱ درنظر بیاورید.

الف) این دو واژه روی محور فعال ـ منفعل چگونه مقایسه می‌شوند؟

ب) آیا جواب الف دور از عقل سلیم به‌نظر می‌رسد؟ پاسخ خودرا شرح دهید.

۷ـ آیا محورهای شکل ۱ به روش مقیاس‌بندی چندمتغیره تهیه شده‌اند؟ پاسخ خود را شرح دهید.

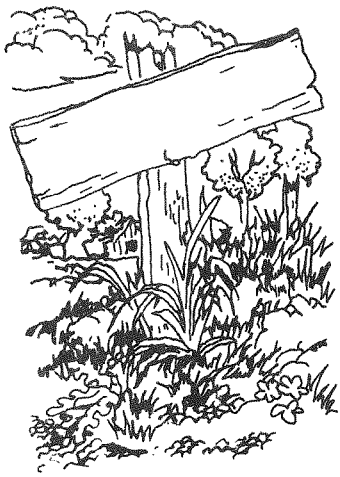
۸ـ عدد ۰/۹۱۴ (در زیر **خوب ـ روشنفکر**) در شکل ۱، یک ضریب همبستگی است. کدام دو مقیاس همبسته شده‌اند؟ از این دو مقیاس چه نتیجه‌ای می‌توانید بگیرید؟

۹ـ شکل ۲ را امتحان کنید. دلتا چگونه اندازه‌گیری می‌شود؟ d چگونه اندازه‌گیری می‌شود؟

منابع

Charls E. Osgood. Gearge J. Suci and Percy Tannenchaum. 1967, (1957). *Yhe Measurement of Meaning Urbana,* III.: University of Illinois.

Myron Wish. M. Deutsch. and L. Biener. 1972, "Differences in Perceived Similarity of Nations." *In Multidimensional Scaling Theory and Applications in the Behavioral Sciences.* New York: Academic Press.



اندازه چیزها

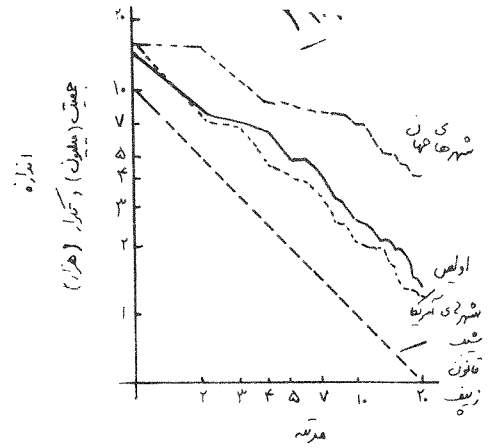
هربرت ای. سیمون[۱] دانشگاه کارنگی ـ ملون

در نمودار (۱) چهار خط رسم شده‌است. از پایینترین خط که یک خط مستقیم با زاویه ۴۵ درجه است، تنها برای مقایسه و توضیح سه خط دیگر، که موجهای ملایمی دارند، استفاده می‌شود. این سه خط موجدار و خصوصاً دوخطی که درست بالای خط مستقیم قرار گرفته‌اند، مبین حقایق جالبی درمورد جهان اطراف ما هستند. اما آیا این حقایق علاوه برجالب بودن معنی‌دار

---

۱ – *Herbert A. Simon*

نمودار ۱ـ لگاریتم اندازه در مقابل لگاریتم مرتبه برای تکرار واژه‌ها و جمعیت شهرها



هم هستند یا نه؟ مطلبی است که اینجا به بررسی آن می‌پردازیم.

خط شکسته پایین، براساس یک مقیاس لگاریتمی[۱]، جمعیت بیست شهر از بزرگترین شهرهای ایالات متحده آمریکا در سال ۱۹۸۰ را به‌رتبه‌هایی برحسب اندازه شهرها ربط می‌دهد که به‌ترتیب از نیویورک با رتبه ۱ تا سین‌سیناتی با رتبه ۲۰ تنظیم شده‌اند.

در محاسبه جمعیت، نه فقط جمعیت داخل محدوده شهرها بلکه جمعیت مناطق آماری کلان شهرها منظور شده‌است. درنمودار فوق، محور افقی رتبه شهر را در مقیاس لگاریتمی، از

۱ـ *احتمالاً بالگاریتم معمولی که روش جالبی برای چندبرابر نمودن ارقام از طریق فرایند جمع است، آشنا هستید. اما به لگاریتم از زاویه دیگری هم می‌توان نگاه کرد و آن اینکه با گرفتن لگارتیم می‌توانیم مقیاس اعداد را فشرده کرده در مقیاس جدیدی به‌وجود آوریم، به این صورت که در ازای ده برابر شدن هر عدد در مقیاس معمولی، در مقیاس جدید یک رقم به عدد (لگاریتم) قبلی‌افزوده می‌شود. برای مثال لگاریتم عدد ۱۰ مساویست با ۱، لگاریتم ۱۰۰ برابر است با ۲، لگاریتم ۱۰۰۰ برابر است با ۳، و الی آخر. لگاریتم ۲۰۰۰ مساویست با ۳/۳۰۰ ولگاریتم ۲۰۰۰۰ مساویست با ۴/۳۰۰. حال اگر یک شهر جمعیتی برابر ۵۰۰۰۰۰۰ نفر داشته باشد، لگاریتم جمعیت آن معادل ۶/۷۰ خواهد بود. و این فشرده شدن اعداد در مقیاس لگاریتم، با افزایش اعداد افزونتر می‌شود.*

۱ تا ۲۰ نشان می‌دهد و محور عمودی نشاندهنده لگاریتم جمعیت به میلیون نفراست. با نادیده‌گرفتن اولین دو شهر بزرگ (نیویورک و لوس‌آنجلس)، مشاهده می‌شود که بقیه خط تقریباً مستقیم است و یک زاویه تقریباً ۴۵ درجه می‌سازد که موازی خط مستقیم زیرین است. اگر انتهای سمت چپ منحنی را به صورت مستقیم فرض کنیم، در آن صورت جمعیت نیویورک از حدود ۲۰ میلیون به حدود ۳۰ میلیون و جمعیت لوس‌آنجلس به ۱۵ میلیون افزایش خواهد یافت (که بهای سنگینی را تحمیل می‌کند)، اما ۱۸ شهر باقیمانده به تعدیل کمی نیاز خواهند داشت، (عموماً کمتراز ۱۰٪ کمتر یا بیشتر).

خط ممتدی که با فاصله کم بالای خط مربوط به جمعیت شهرها قرار دارد، نشاندهنده (بازهم در مقیاس لگاریتمی) مراتب تکرار هریک از ۲۰ واژه‌ای است که بیش از واژه‌های دیگر در کتاب "اولیس" اثر جیمز جویس به‌کار برده شده و به‌ترتیب نزولی فراوانی رخداد تنظیم شده‌اند. برای این خط اعداد روی محور افقی مبین فراوانی رخداد در هزار می‌باشد. اولین واژه‌ای که بیش از همه در اولیس تکرار می‌شود، حرف تعریف the است با ۱۴۸۸۷ رخداد؛ و بیستمین کلمه all است با ۱۳۱۱ مورد رخداد. منحنی فراوانی واژه‌ها همچون منحنی اندازه شهرها یک خط تقریباً مستقیم است. هرچند که مستقیم‌نمودن این خط هم مستلزم تعدیل چند واژه اول است؛ به‌این ترتیب که تکرار حرف the باید به ۲۶۰۰۰ مورد، تکرار حرف of به ۱۳۰۰۰ مورد و حرف and به ۸۷۰۰ مورد افزایش یابد. تکرار ۱۷ واژه دیگر به‌گونه‌ای است که فراوانی آنها بسیار نزدیک به یک خط مستقیم با زاویه ۴۵ درجه است.

توجه داشته باشید که توزیع نقاط این خطوط به‌گونه‌ای است که نتیجه حاصلضرب رتبه در اندازه هر مورد مقداری ثابت است. مثلاً اگر اندازه اولین مورد (رتبه ۱) برابر ۱۰۰۰۰۰۰ باشد، اندازه دهمین مورد تقریباً برابر ۱۰۰۰۰۰ (۱۰۰۰۰۰۰=۱۰×۱۰۰۰۰۰) و اندازه بیستمین مورد برابر ۵۰۰۰۰ (۱۰۰۰۰۰۰=۲۰×۵۰۰۰۰) خواهدبود. حال مابایدتوضیح دهیم که چرا چنین نظمی برقراراست؟ چرا حاصلضرب تعداد و رتبه دراین دو توزیع مختلف تقریباً ثابت است؟ و سئوال رمزآمیزتر اینکه چرا توزیع اندازه شهرهای ایالات متحده و توزیع میزان فراوانی واژه‌ها در کتابی همچون اولیس که از لحاظ سبک نوشتار غیرمعمول است (یا هر کتاب دیگری) از قاعده یکسانی پیروی می‌کنند؟ اجازه دهید با واژه‌ها شروع کنیم.

واژه‌ها: متداول و نادر

در اواخر قرن نوزدهم، چند زبانشناس (از جمله دوساسور در فرانسه) به قاعده حیرت‌انگیز رتبه ـ فراوانی در کاربرد نسبی واژه‌های مختلف در هر متن پی بردند. مشخصاً، واژه‌هایی چون of در هرگونه متن انگلیسی مکرراً مشاهده می‌شود، در حالی که واژه‌های دیگری چون conundrum، بندرت و یا اصلاً در یک متن به‌کار نمی‌روند. فراوانی یک واژه خاص در دو متن مختلف می‌تواند بسیار متفاوت باشد.

هرگاه واژه‌های مختلف را به‌ترتیب فراوانی رخداد آنها در یک متن معین مرتب کنید، یعنی اول واژه‌ای که بیش از همه تکرار شده‌است، بعد واژه‌ای که بعد از واژه اول بیش ازهمه تکرار شده و الی آخر، نظمی که در نمودار ۱ به‌تصویر کشیده‌شده دوباره ظاهر می‌شود. مراتب رخداد بیستمین واژه فهرست تقریباً نصف مراتب تکرار واژه دهم خواهد بود[۱]. اگر این ترتیب عددی برای شماجالب است، می‌توانید در انتهای دیگر منحنی توزیع در بین واژه‌های نادر نیز رابطه‌هایی به همین نسبت جالب را پیدا کنید.

حدود یک دوم از تعداد کل واژه‌های مختلف یک متن دقیقاً فقط یک بار به‌کار برده می‌شوند، حدود یک ششم دقیقاً تنها دومرتبه؛ و حدود یک دوازدهم فقط سه مرتبه (ن. ک. جدول ۱). نسبت کاربرد تمامی واژه‌هایی که در یک متن n مرتبه تکرار می‌شوند را می‌توان از طریق فرمول $1/[n(n+1)]$ محاسبه نمود. مسلماً قاعده فراوانی رخداد واژه‌های نادر همان قانون رتبه ـ اندازه است که درانتهای دیگر منحنی توزیع مشاهده کردیم. چراکه رتبه یک واژه برابر با مجموع تعداد واژه‌های مختلفی است که یا به اندازه آن واژه و یا بیش از آن تکرار شده‌اند. بنابراین فرض کنید همان‌طورکه در قاعده رتبه ـ اندازه لازم است، $k/(n+1)$ واژه $n+1$ مرتبه و یابیشتر و k/n واژه n مرتبه و یابیشتر تکرار می‌شود. درنتیجه تعداد واژه‌هایی که دقیقاً n مرتبه تکرار می‌شوند، برابر خواهد بود با: $k/[n(n+1)]=k/n-k/(n+1)$.

قاعده رتبه ـ اندازه که اغلب از آن به قانون زیپف (zipf) یاد می‌شود، به احترام یک

۱ـ در اغلب موارد، فراوانی دو یا سه مورد اول کمتر از حدی است که این قاعده پیش‌بینی می‌کند، همچون وضعیتی که در نمودار (۱) مشاهده می‌شود.

زبانشناس آمریکایی به همین نام که مطالب زیادی دراین رابطه به رشته تحریر درآورد، نامیده می‌شود و تقریباً در مورد تمامی متونی که مجموعه واژه‌های آنها شمارش شده و در طیف وسیعی از زبانهای دنیا قرار داشته‌اند، ازجمله زبانهای بومی آمریکا، صادق است. اما درحالی که برای "اولیس" این قاعده درست است، در مورد کتاب دیگر جویس به‌نام "فینه گانزویک" صدق نمی‌کند (شاید به این دلیل که جویس در این کتاب در به‌کار گرفتن انواع لغات شکسته و شکلهای مختلف کلمات موجود در فرهنگ، خود را آزاد گذاشته است). در یک بررسی از خط تصویری چین ظاهراً نتیجه این شد که قاعده مذکور صادق نیست؛ اما یک بررسی جدید در مورد تکرار واژه‌های چینی (هر واژه ممکن است از یک یا دو تصویر ویا بیشتر تشکیل شود) نشان می‌دهد که این قاعده در مورد زبان چینی هم ماند دیگر زبانها جواب می‌دهد.

چرا چنین نظمی برقرار است؟ چرا باید بین واژه‌های مکرر و نادر در یک روزنامه و در "اولیس"، در کتابهای آلمانی، کتابهای انگلیسی، و یا در اغلب (نه همه) صحبتهای افراد شیزوفرنی و صحبتهای افراد عادی، توازن دقیقاً یکسانی برقرار باشد؟

در پاسخ این سئوال جوابهای متعددی ارائه شده‌است، یکی از آنها از نوع توضیحاتی است

جدول ۱ـ تعداد واژه‌های باکمترین فراوانی در اولیس اثر جیمز جویس

تعداد تکرار (n)	تعداد واژه‌های به‌کار رفته	تعداد واژه‌های پیش‌بینی شده*
۱	۱۶۴۳۲	۱۴۹۴۹
۲	۴۷۷۶	۴۹۸۳
۳	۲۱۹۴	۲۴۹۱
۴	۱۲۸۵	۱۴۹۵
۵	۹۰۶	۹۹۷
۶	۶۳۷	۷۱۲
۷	۴۸۳	۵۳۴
۸	۳۷۱	۴۱۵
۹	۲۹۸	۳۳۲
۱۰	۲۲۲	۲۷۲

K تعداد کل واژه‌های مختلف در اولیس است. * *K=۲۹۸۹۹;K/[n(n+1)]=تعداد پیش‌بینی شده*

که ازتئوری احتمالات نتیجه می‌شود. تئوری احتمالات با پذیرفتن ناتوانی خود در توضیح دقیق ترتیب امور، اغلب بر مبنای حد متوسط به تشریح آن می‌پردازد. مثلاً در توضیح قوانین گازها از ردگیری مسیر تک تک مولکولها سر باز می‌زند.

در تشریح توزیع واژه‌ها، فرضیاتی را درنظر می‌گیریم که اگر بخواهیم آنها را جزء به جزء اعمال کنیم، ممکن است بسیار غیرعادی جلوه کنند. اما اگر فقط به‌صورت کلی اعمال شوند، قابل قبول به‌نظر می‌رسند. فرض کنیم که یک نویسنده متنی را از طرفی با بهره‌گیری از تمامی محفوظات وسیع خود و از طرفی با استفاده از مجموعه به‌مراتب وسیعتر ادبیات و زبان خود، به رشته تحریر درآورد. در مورد اول این نویسنده از صنعت "تداعی معانی" و در مورد دوم از صنعت "تقلید" استفاده می‌نماید. نکته مهم اینکه فرض ما براین است که شانس انتخاب بعدی هر واژه مفروض ازطرف نویسنده، متناسب است با تعداد دفعاتی که آن واژه قبلاً در حافظه و یا در متون ادبی ذخیره شده‌است. به‌خاطر داشته باشید؛ قرار براین است که این مفروضات تنها در مقیاس کلان اعمال شوند. برای قبول این مفروضات، کافی است باور ما به اینکه شکسپیر غزلهای خود را با گرداندن چرخ رولت سروده‌است، در حد باورمان به این مطلب باشد که هریک از مولکولهای گاز مسیر خود را با ریختن تاس تعیین می‌کند.

اگر این فرضیات را قبول کنیم، موضوع مورد بحث ما، یک موضوع سرراست ریاضی می‌شود، اما مطرح نمودن توزیع احتمالاتی لازمه آن فراتر از حد این مقاله است. آنچه که از این موضوع ریاضی نتیجه می‌شود، به "توزیع یول" معروف است. در طیف واژه‌ها برای کلماتی که در بالای طیف قرار دارند و به عبارتی از تکرار زیاد برخوردارند، "توزیع یول" با قانون رتبه ـ اندازه زیپف مطابقت دارد؛ و برای کلمات پایین طیف که از تکرار کم برخوردارند، دقیقاً همان کسر $1/[n(n+1)]$ رابه‌دست می‌دهد.

حال متوجه می‌شویم که چرا توزیع "همسان" می‌تواند برای انواع گوناگون متون در ادبیات بسیاری از زبانها برازنده باشد. توزیع "همسان" می‌تواند برازنده باشد زیرا به هیچیک از ویژگیهای بسیار خاص فرایندی که متن را ایجاد کرده است، بستگی ندارد. و اینکه این فرایند به مفهوم احتمالاتی یک فرایند تداعی معانی و تقلیدی باشد، تنها به تهیه‌کننده متن بستگی دارد. حتی شاید بتوان انتظار داشت که اگر نسبت دقیق تداعی معانی و تقلید در حد قابل توجهی به‌هم

بخورد، باز هم تغییر مهمی در ماهیت توزیع رخ ندهد. تا آنجا که نتایج اعمال تغییر در مفروضات مورد بررسی قرار گرفته، چه از طریق ریاضی و چه از طریق شبیه‌سازی کامپیوتری، الگوی توزیع حقیقتاً از ثبات بسیار بالایی برخوردار بوده‌است. در واقع می‌توانیم به شکسپیر و جویس آزادی زیادی در سبک نگارش بدهیم؛ بدون آنکه در رابطه ساده رتبه ـ اندازه واژه‌های کاربردی آنها تغییر محسوسی ایجاد کنیم. اما از بررسی کتاب "فینه‌گانزویک" به این نتیجه می‌رسیم که نمی‌توانیم به آنها آزادی بی‌نهایت بدهیم.

ابرشهرها و کلان‌شهرها

حال که تاحدی پرده از رمز و راز واژه‌های متون ادبی برداشته‌ایم، شاید برای بررسی وجود نظم مشابهی در مورد اندازه شهرهای ایالات متحده آماده باشیم. همان‌طور که مشاهده کردید (نمودار ۱) جمعیت شهرها به‌تقریب خیلی نزدیک به‌یقین از قاعده رتبه ـ اندازه مشابهی تبعیت می‌کند.[۱] اگر دو شهر به‌ترتیب دارای رتبه‌های j و k در فهرست ترتیبی شهرها باشند، نسبت تقریبی جمعیت آنها k/j خواهدبود.

این‌طور نیست که این نظم یک موضوع اتفاقی خاص سرشماری ۱۹۸۰ باشد. درواقع برای تمام سرشماریهای انجام شده ازسال ۱۷۸۰ به بعد این موضوع صادق است. اما این قاعده برای شهرهایی که در آن‌دسته از مناطق جغرافیایی جهان قرار دارند که بنا به مقتضیات تعریف شده‌اند، و نسبتاً از لحاظ اقتصادی مناطق مستقلی نیستند، مصداق پیدا نمی‌کند. مثلاً برای اتریش و یا کشورهای آمریکای مرکزی به‌صورت انفرادی و یا استرالیا صدق نمی‌کند. همچنین اگر همه شهرهای جهان را کنار هم قرار دهیم باز هم قاعده فوق جواب نمی‌دهد (به بالاترین منحنی در نمودار ۱ نگاه کنید). درچنین حالتی بازهم منحنی توزیع نسبتاً هموار و منظم است. اما جمعیت با آن سرعتی که قانون زیپف اقتضا می‌کند، همراه با رتبه شهر تنزل نمی‌کند. توزیع از

۱- *در اینجا می‌توانیم یا جمعیت شهرها را براساس مرزهای مشترک تعریف شده، و یا جمعیت موجود در مناطق کلانشهر را براساس تعریف سرشماری ایالات متحده، مبنا قرار دهیم. در هر دوصورت نظم مورد بحث به‌خوبی مشهوداست. شاید استفاده از مناطق آماری کلانشهر نتیجه نسبتاً بهتری داشته‌باشد.*

یکنواختی بیشتری برخوردار است و بزرگترین کلانشهرها بسیار کوچکتر از آنچه که هستند، نمود داده می‌شوند. اما سریعاً در اینجا اضافه می‌کنیم که این عبارت را نباید تجویزی برای همه موارد مشابه دانست.

(تعریف اندازه برای شهرهای جهان با تعریف اندازه شهرهای ایالات متحده متفاوت است و بنابراین اندازه واقعی آنها کاملاً باهم قابل مقایسه نیست. و نیز فهرستهای این دوگروه مربوط به دو سال مختلف یعنی ۱۹۷۵ و ۱۹۸۰ است و این احتمال وجود دارد که در انتخاب فهرست شهرهای جهان مفهوم کلانشهر درنظر گرفته نشده‌باشد.)

بنابراین از لحاظ اندازه شهرها، باید به توضیح دو مطلب بپردازیم: یکی اینکه چرا قانون زیپف برای مدت بیش از دو قرن در مورد شهرهای ایالات متحده صادق بوده است؟ و دیگر اینکه چرا این قانون برای بسیاری از مجموعه شهرهای دیگر صادق نیست؟ اجازه دهید با اولین سئوال شروع کنیم و از خود بپرسیم که بین این موضوع و فرایندهای تداعی معانی وتقلید که مبین چگونگی توزیع واژه‌ها بودند، چه موارد تشابهی می‌تواند وجود داشته باشد؟ به بیان دقیقتر، باید از خود بپرسیم که چه فرایندهایی باعث می‌شود نرخ رشد شهرها، به‌طور متوسط، متناسب با اندازه فعلی آنها باشد؟ (بعضی اوقات از آن به "اصل جیبرات" یاد می‌شود)؛ چرا که این مطلب فرض اساسی مورد نیاز برای استخراج ریاضی است.

شهرها در اثر باقیمانده خالص تولد بر مرگ و میر و خالص افزونی مهاجرت به داخل بر مهاجرت به خارج رشد می‌کنند. از نظر تولد ومرگ، کافیست که تنها فرض را براین بگیریم که به‌طور متوسط نرخهای تولد و مرگ با اندازه شهر هیچ همبستگی ندارند. ازنظر مهاجرت فرض را براین می‌گیریم که مهاجرت به خارج، به‌طور متوسط، متناسب با اندازه شهراست (به‌عبارتی نرخهای سرانه مستقل از اندازه هستند) و مهاجرت به داخل (از مناطق روستایی، دیگر شهرها و یا خارج ازکشور) نیز متناسب با اندازه شهر است. آخرین فرض به این معناست که شهرهایی که دریک گروه اندازه‌ای مفروض قرار دارند، از لحاظ مهاجرت، یک منطقه «هدف» را تشکیل می‌دهند و هرچه جمعیتی که در شهرهای آن گروه زندگی می‌کنند بیشتر باشد، رقم کل مهاجرت هم رقم بالاتری خواهدبود. (ما این را به‌عهده خواننده می‌گذاریم که خود دلایل قابل قبول بودن این فرض را، حداقل به‌عنوان یک تقریب، مورد بررسی قرار دهد.)

با چنین مفروضاتی، بازهم جنبه ریاضی مطلب ما را به سمت قانون رتبه ـ اندازه زیپف سوق می‌دهد. اما در اینجا آموزنده است این سئوال را مطرح کنیم که تحت چه شرایطی باید انتظار داشته باشیم که مجموعه‌ای از شهرها با این مفروضات برازش داشته باشند؟ جواب این است که چنین شهرهایی باید یک منطقه "طبیعی" را تشکیل دهند که در محدوده آنها تحرک گسترده و آزاد جمعیتی و صنعتی مشاهده شود، و نیز بخش دلخواهی ازیک منطقه وسیعتر نباشند. وضعیت کشوری مانندایالات متحده کاملاً با این مفروضات برازش دارد اما منطقه‌ای که یک نقش خاص را در یک پیکره اقتصادی بزرگتر بازی می‌کند، ممکن است با این مفروضات برازش نداشته باشد (برای مثال، کشور اتریش پس از انحلال امپراتوری و یا کشوری که صادرکننده محصولات عمده کشاورزی و تنها دارای یک بندر بزرگ باشد).

اگر تعداد زیادی از توزیعها را که هرکدام جداگانه از قانون رتبه ـ اندازه تبعیت کند، در کنار هم قراردهیم، به منحنی توزیع جدیدی با همان شکل می‌رسیم؛ با این تفاوت که در ناحیه بالاتری از نمودار قرار می‌گیرد و قسمتی از بالای منحنی حذف شده‌است. انتظار ماابن است که شهرهای جهان عموماً با منحنی توزیع رتبه ـ اندازه برازش داشته باشند، به استثنای ناکارایی کلانشهرهای بسیار بزرگ در رأس رتبه‌بندی که چنین هم هست. اگر ارقام به‌چاپ رسیده را با ارزش ظاهری آنها بپذیریم (تعریف مضامین کار دشواری است ونتایج سرشماریها از دقت متفاوتی برخودار است)، چیزی در حدود ۵۰ مجموعه شهری در سطح جهان وجود دارد که هرکدام بیش از دومیلیون نفر را در خود جای داده‌اند و با این حساب قانون زیپف برای شهرهایی چون نیویورک و توکیو جمعیتی برابر ۱۰۰ میلیون را پیش‌بینی خواهد کرد، در صورتی که جمعیت فعلی هرکدام از این دو شهر تنها ۲۰ میلیون نفراست اما ناکارایی شهرهای دارای رتبه بالا (معمولاً ۱۰ مورد اول) بلافاصله توسط شهرهای بیشماری که هرکدام بیش از ۵ میلیون جمعیت دارند، تا حدود زیادی جبران می‌شود. در حال حاضر، دهمین شهر فهرست، یعنی پاریس، جمعیتی برابر ۹/۲ میلیون نفر دارد که تنها ۱۰٪ کمتر از رقمی است که قانون زیپف به‌دست می‌دهد.

اندازه شهرها برای مردم ساکن در آنها از اهمیت آشکاری برخوردار است. اما مشخصاً نمی‌توان گفت که از توزیع اندازه واقعی شهرها به چه نتایجی می‌توان رسید؟ یک نتیجه ممکن

این است که بدون وجود کنترل حکومتی و یا اقتصادی قوی بر اماکن سکونتی و شغلی، وضعیت توزیع به‌راحتی تغییر نمی‌کند و به بیان خوشایندتر ـ از آن‌رو که عموماً نمی‌خواهیم چنین کنترلهایی وجود داشته باشد ـ تحلیل ریاضی مطلب که نیروهای حاکم بر پدیده مورد نظر را مشخص می‌سازد، به ما می‌آموزد که انجام هرگونه تلاشی در جهت تغییر پدیده مفروض مستلزم برخورد زیرکانه و هشیارانه با چنین نیروهایی است.

کار و کسب بزرگ و کوچک

اقتصاددانان همواره به وسعت بنگاههای تجارتی بیش از اندازه شهرها علاقمند بوده‌اند. تمرکز صنعت در دستهای معدودی شرکتهای بزرگ از نظر اغلب صاحبنظران رقابت را مشکل می‌کند و عقیده عموم بر این است که چنین تمرکزی در طول قرن حاضر با نرخ سریعی در ایالات متحده گسترش یافته است.

سالهاست مشخص شده‌است که توزیع اندازه‌ای بنگاههای تجارتی دایر در ایالات متحده، انگلیس، و دیگر کشورها با قانون رتبه ـ اندازه زیپف همخوانی دارد، با این تفاوت که در مقایسه با مواردی که قبلاً مورد بحث قرارگرفت، اندازه با سرعت کمتری همراه با رتبه کاهش می‌یابد (به این معنی که نسبت بزرگترین بنگاه به دهمین بنگاه بزرگ عموماً کمتر از ده به‌یک است[۱]). هرچه کاهش اندازه به‌موازات افزایش رتبه سرعت کمتری داشته باشد، تمرکز فعالیتهای تجاری

۱- فرض کنیم که n و m رتبه‌های دو عضو یک توزیع ردیف - اندازه و S_m و S_n به‌ترتیب اندازه‌های این دو باشند. در اینجا براساس فرم کلی قانون رتبه-اندازه خواهیم داشت: $S_m/S_n=(n/m)k$ که در آن توان k کسری کوچکتر از یک است که وقتی به عدد یک به‌عنوان حد نزدیک می‌شود، حالت خاص توزیع "زیپف" را به‌دست می‌دهد. اقتصاددانان عموماً توزیع کلی را "توزیع پارتو" می‌نامند. حال اگر بامحاسبه لگاریتم رتبه‌ها واندازه‌ها، نموداری از توزیع لگاریتمی رسم کنیم، بازهم خط مستقیمی خواهیم داشت که زاویه آن مساوی کسر k خواهد بود و هرچه این خط مستقیم شیب تندتری داشته باشد (k بزرگترباشد)، شرکتهایی که دراول فهرست رتبه‌بندی قرار دارند در مقایسه با شرکتهایی که در مراتب پایینتر فهرست هستند، بزرگتر خواهند بود (به‌عبارت دیگر، هرچه k بزرگتر باشد، تمرکز هم بیشتر خواهد بود).

در بزرگترین بنگاهها کمتر است.

واقعیتی که اقتصاددانان را متحیر ساخته این است که نرخ کاهش اندازه با افزایش رتبه، که یکی از طرق اندازه‌گیری تمرکز صنعتی است برای شرکتهای تولیدی بزرگ ایالات متحده، در زمان حاضر و در ۳۵ سال قبل و حتی در ابتدای قرن، از وضعیت تقریباً یکسانی برخوردار بوده است. حتی در طول دوره‌هایی که ادغامهای مکرر شرکتها وجود داشته‌است، میزان تمرکز صنعتی که از طریق رابطه رتبه ـ اندازه محاسبه شده، مقدار ناچیزی نوسان داشته‌است.

باتوجه به تحلیلهایی که قبلاً داشتیم، قاعدتاً باید برای حل این معما آماده باشیم. درواقع، تحت مفروضات مناسب در مورد شرکتهایی که ادغام یا منحل شده و یا گسترش می‌یابند، می‌توان به‌صورت ریاضی نشان داد که ادغام تأثیری بر تمرکز ندارد. به‌علاوه مفروضات ضروری برای چنین استنتاج ریاضی کاملاً با داده‌های آماری ایالات متحده در مورد ادغام شرکتها برازش دارد. در مقایسه با وضعیتی که برای واژه‌ها و شهرها تشریح کردیم، می‌توان حدس زد که مفروضات موردنظر و نیز داده‌هایی که آنها را پشتیبانی می‌کند، مانند چیست؟

- احتمال اینکه یک شرکت درنتیجه ادغام «بمیرد» باید تقریباً مستقل از اندازه آن باشد.
- متوسط داراییهایی که درنتیجه ادغام نصیب شرکتهای ماندگار می‌شود باید تقریباً متناسب با اندازه فعلی آنها باشد و این پیش‌بینیها درواقع چندان دور از واقعیت نیستند.

به این ترتیب یک تکاپوی مختصر علمی که با یک مسئله زبانشناسی در مورد فراوانی واژه‌ها آغاز شده به تشریح یک دوگانگی در مورد تمرکز صنعتی در ایالات متحده منتهی شده‌است. تشریح این مطلب خطوط جدیدی از تحقیق را برای درک رشد تجارتی و ورود به سیاستگذاری عمومی برای حفظ رقابت تجاری می‌گشاید.

شیفتگی ما نسبت به توزیعهای رتبه ـ اندازه الزاماً به‌همین سه مورد که در اینجا مطرح کردیم، محدود نمی‌شود. می‌توانیم انتظار داشته‌باشیم که توزیع زیپف در موارد دیگری هم به همین صورت نمایان شود و هر رخداد جدید ما را برمی‌انگیزد که یک سری مفروضات قابل قبول (وقابل آزمون) را به‌گونه‌ای ابداع کنیم که بتوانیم قانون رتبه ـ اندازه را از آن استنتاج نموده و رخداد را توضیح دهیم. در اینجا می‌خواهیم موردی را به‌عنوان تمرین برای شما خوانندگان

مطرح کنیم. فهرستی از تمام نویسندگانی که در طول یک دوره ۲۰ساله در تهیه مطالب یک مجله علمی شرکت داشته‌اند و یااسامی آنها در یک کتابنامه کامل همچون "چکیده مقالات شیمی" چاپ شده است، تهیه نمایید. تعداد مواردی را که نام هرنویسنده ذکر شده‌است، یادداشت و سپس اسامی را برحسب این تعداد موارد رتبه‌بندی کنید. مشاهده خواهید کرد که اسامی حدوداً نیمی از نویسندگان تنها یک بار نوشته شده، یک ششم آنها دومرتبه، و الی آخر؛ ارقام حاصل تقریباً با توزیع "یول" مطابقت دارند. روش کار نویسندگانی که می‌توانند توضیحی طبیعی برای این حقیقت ارائه نمایند، چیست؟

مسائل

۱ـ با توجه به نمودار (۱) جمعیت تقریبی فیلادلفیا که پنجمین شهر بزرگ آمریکاست، چقدر است؟

۲ـ باتوجه به جدول (۱) به‌موارد زیر جواب دهید:

الف) در کتاب اولیس، ۴۸۳ واژه متفاوت هرکدام ۷ مرتبه تکرار شده‌اند. رقم پیش‌بینی شده ۵۳۴ چگونه محاسبه شده‌است؟

ب) فرض کنید تعداد واژه هایی که n مرتبه تکرار شده‌اند، برابر ۱۶۴ است. مقدار تقریبی n چقدر است؟ (توجه: باید از فرمول حل معادله درجه دوم که در جبر دبیرستان فراگرفته‌اید استفاده نمایید.)

۳ـ منظور نویسنده از واژه‌های تداعی معانی (association) و تقلید (imitation) چیست؟

۴ـ چگونه ممکن است که جمعیت شهرهای ایالات متحده آمریکا و تکرار واژه‌های یک متن دارای الگوی توزیع مشابه باشند؟

۵ـ اصل "جیبرات" چیست؟ وچه رابطه‌ای با قانون زیپف دارد؟ وچه رابطه‌ای با اندازه شهرهای ایالات متحده دارد؟

۶ـ نخست در مورد متون ادبی و سپس در مورد اندازه شهرها، مفروضات ریاضی را که منتهی به توزیع "یول" می‌شود، در قالب کلمات بیان کنید.

۷ـ می‌توانید دلیلی بیابید که چرا چند مورد اول فهرست بزرگترین شهرهای ایالات متحده

ممکن است با قانون رتبه ـ اندازه همخوانی نداشته باشد؟

۸ـ الف) توزیع زیپف و پارتو چگونه باهم رابطه دارند؟

ب) در یک نمودار لگاریتمی از توزیع پارتو پارامتر k راتعریف کنید.

۹ـ فرض کنید که شرکتهای آی بی ام و زیراکس به‌ترتیب مقامهای اول و چهارم فروش را داشته باشند و فرض کنید که فروش زیراکس برابر یک میلیارد دلار باشد و میزان فروش شرکتهای تجاری از توزیع پارتو با مقدار $k = \frac{۱}{۲}$ تبعیت کند. در این‌صورت میزان فروش آی بی ام چقدر خواهد بود؟

۱۰ـ اگر R_n رتبه چیزی با اندازه n باشد قانون زیپف را بیان کنید.

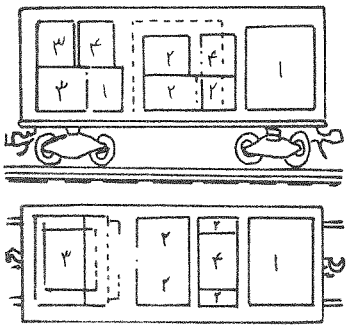
۱۱ـ فرض کنید که چهارمین شهر بزرگ، جمعیتی برابر ۱۰ میلیون نفر داشته باشد؛ براساس قانون زیپف ردیف شهری با جمعیت ۲/۵ میلیون نفر را تعیین کنید.

۱۲ـ نویسنده اظهار می‌دارد تعداد واژه‌های مختلفی که دقیقاً n مرتبه دریک متن تکرار می‌شوند برابر با $k/[n(n+۱)]$ است. k چیست؟ (راهنمایی: به جدول ۱ مراجعه کنید)

۱۳ـ باتوجه به تمرین مطرح شده در آخرین پاراگراف متن، کدام دو فرض را برای این توزیع پیشنهاد می‌کنید؟ (راهنمایی: از فرضهای ادغام شرکتها و اندازه شهرها به عنوان راهنما استفاده کنید.)

منابع

Yuji Ijiri and Herbert A. Simon. 1977. *Skew Distributions and the Sizes of Business Firms.* Amsterdam: North-Holland Publishing Company.



حسابداران چگونه با نمونه‌گیری، صرفه‌جویی می‌کنند؟

جان نتر(۱) دانشگاه جورجیا

بسیاری از مردم حسابداری و آمار را از کسل‌کننده‌ترین مباحث موجود می‌دانند. امید است که مطالب این مجموعه بتواند چنین ذهنیتی را نسبت به آمار تغییردهد. این مقاله با کاربردهای مهم آمار در حسابداری عملی سروکار دارد و ممکن است برخی از وجوه جالب حسابداری را نیز آشکار کند.

۱- *John Neter*

در هرحال همه ما می‌خواهیم از پول خود به‌طور مؤثر و کارا استفاده کنیم. ملاحظه خواهیم کرد که به‌کار گرفتن نمونه‌گیری آماری در حسابداری، وقتی که با مشکل تقسیم درآمدها بین چند شرکت حمل ونقل مواجه می‌شوند، چگونه به صرفه‌جویی شرکتهای راه‌آهن و هواپیمایی منجر می‌شود. روشهای نمونه‌گیری آماری مشابهی در دیگر زمینه‌های حسابداری و حسابرسی نیز به‌کار رفته‌اند. درواقع از این روشها در بسیاری از زمینه‌های بازرگانی، اداره کشور و علوم هم استفاده می‌شود.

حسابداران و حسابرسان به‌طور سنتی بر درستی اسناد حسابداری مؤسسات و دیگر سازمانها تأکید کرده‌اند. این پافشاری آنها را به انجام کار زیاد، به‌صورت کامل، بر مبنای ۱۰۰٪، واداشته‌است. برای مثال حسابرسان ممکن است بخواهند ارزش موجودی انبار یک شرکت را کنترل کنند. برای این کار می‌توانند موجودی انبار را به‌صورت کامل بررسی کنند. یعنی عملاً به شمارش هریک از انواع کالاهای موجود در انبار بپردازند، ارزش هرنوع کالا را معین نمایند و در آخر؛ ارزش کل موجودیهای انبار را به‌دست آورند.

به‌عنوان یک مثال دیگر، حسابرسی را درنظر بگیرید که به اطلاع از نسبت حسابهای بدهکاران با مانده ۶۰ روزه یا بیشتر علاقه‌مندی دارد. ممکن است این اطلاعات برای تعیین وثیقه بدهکاران بد ضروری باشد. حسابهایی که در مدت ۶۰ روز تسویه نشده‌اند، بیشتر از حسابهایی که در مدت کوتاهی تسویه شده‌اند، برای تحمل زیان بدحسابی مستعد هستند. به منظور تعیین نسبت حسابهای بدهکاران تسویه‌نشده طی ۶۰ روز یا بیشتر، حسابرس می‌تواند تمام بدهکاران شرکت را بررسی و مانده هر حساب با بدهی ۶۰ روز یا بیشتر را معین کند.

آیا واقعاً ضروری است که به بررسی کامل موجودی انبار حسابهای بدهکاران یا مجموعه‌های مشابه‌دست زد تا به ارقام مورد نیاز دست یافت؟ به‌طور مشخصتر آیا یک نمونه می‌تواند اطلاعات مورد نیاز حسابدار را به‌طور کامل فراهم کند، بدون آنکه تمامی کارهای خسته‌کننده برای یک بررسی کامل ۱۰۰ درصدی انجام شود؟ بگذارید روی اقلام موجودی انبار متمرکز شویم.

در واژگان آمار، مجموعه اقلام موجودی انبار که ارزش کل آن باید تعیین شود، **جامعه** مورد نظر خوانده می‌شود. یک نمونه منتخب از چنین جامعه‌ای فقط بخشی از اقلام جامعه ولی نه

تمام اقلام را دربر می‌گیرد. به منظور آگاهی از ویژگیهای جامعه و بدون بررسی تمام عناصر آن، نمونه انتخاب می‌شود.

هزینه بررسی یک نمونه نسبتاً کوچک از اقلام موجودی انبار معمولاً کمتر از هزینه بررسی کامل موجودی انبار است، زیرا نمونه به بررسی اقلام کمتری نیازمند است. اما آیا نتایج مبتنی بر یک نمونه نسبتاً کوچک، تقریباً به‌خوبی نتایج مبتنی بر یک بررسی کامل است؟

تجربه نمونه‌گیری در بسیاری از زمینه‌ها نشان داده‌است که نمونه‌های نسبتاً کوچک اغلب نتایجی را ارائه می‌کنند که تقریباً به‌خوبی نتایج به‌دست آمده از یک بررسی کامل است، درحالی که نتایج نمونه به میزان قابل ملاحظه‌ای کم هزینه‌تر است. درواقع گاهی اوقات نتایج نمونه‌گیری حتی از نتایج بررسی کامل نیز بهتراست. این حکم ممکن است شگفت‌آور به‌نظر رسد اما کار بررسی موجودی انبار یک شرکت بزرگ را درنظر بگیرید. به افراد زیادی برای این کار نیاز است. به سبب حجم کار ممکن است آموزش تمام و کمال این همه افراد دشوار و کنترل کیفیت عملیات محدود باشد. از طرف دیگر نمونه کوچکی از اقلام موجودی انبار، به افراد کمتری نیاز دارد که درنتیجه می‌توان آنان را بهتر آموزش داد. فراتر اینکه هرگاه تعداد افراد کمتری درگیر باشند، برنامه کنترل کیفی موجودی انبار را دقیقتر می‌توان انجام داد. اثر خالص ممکن است بیانگر دقیقتر بودن نتایج نمونه در مقایسه با بررسی کل جامعه باشد! یعنی فایده دقت ناشی از آموزش بهتر و کنترل کیفیت با یک نمونه کوچک ممکن است از هزینه خطای نمونه‌گیری حاصل از انتخاب فقط بخشی از اقلام موجودی انبار به‌جای همه اقلام، بیشتر باشد. البته نمونه‌گیری باید هوشمندانه و به‌طور مناسب انجام گیرد. بحث نمونه‌گیری از بخشهای مهم آمار است.

مطالعه راه‌آهن چساپیک و اوهایو[1]

احکامی از این قبیل که نمونه‌های نسبتاً کوچک می‌توانند نتایجی تقریباً به‌خوبی نتایج حاصل از بررسیهای کامل یا گاهی اوقات حتی نتایجی واقعاً بهتر از آنها را فراهم کنند، اغلب به خودی

۱ـ *Chesapeake and Ohio*

خود قانع‌کننده نیستند. بدین سبب آمارشناسان غالباً انجام مطالعاتی را مفید می‌دانند که نتایج یک نمونه را با نتایج حاصل از بررسی جامعه مقایسه می‌کنند. چنین مطالعه‌ای توسط شرکت راه‌آهن چساپیک و اوهایو به منظور تعیین میزان درآمد حاصل از حمل محموله‌های بین راهی، با بار کمتر از ظرفیت، انجام شد.اگر یک محموله توسط چند شرکت راه‌آهن حمل شود، کل هزینه حمل بین آنها تقسیم می‌شود. محاسبات لازم برای تعیین درآمد هر شرکت راه‌آهن بسیار خسته‌کننده و پرهزینه است. بدین خاطر شرکت چساپیک و اوهایو برای تعیین اینکه آیا تقسیم کل درآمد بین چند شرکت راه‌آهن براساس یک نمونه می‌تواند به‌طور دقیق و با صرفه‌جویی قابل ملاحظه‌ای در هزینه‌های اداری انجام شود؛ به آزمایشی دست زد.

این شرکت در آزمایشی به مطالعه تقسیم درآمد حمل تمام محموله‌ها با بار کمتر از ظرفیت، در بخش پرمارکت[۱] بین چساپیک و اوهایو و یک شرکت راه‌آهن دیگر (موسوم به الف برای محرمانه نگهداشتن اطلاعات) در طول یک دوره شش ماهه پرداخت. جامعه بارنامه‌های چنین محموله‌هایی بررسی شد. یک بارنامه سندی است که برای هر محموله صادر می‌شود و اطلاعات تفصیلی درباره کالا، مسیر حمل و هزینه‌ها ارائه می‌کند. بر این اساس مقادیر پرداختی به هر شرکت را می‌توان محاسبه کرد. تعداد کل بارنامه‌ها در جامعه و نیز درآمد کل حاصل از حمل کالا را از جامعه بارنامه‌ها می‌توان یافت. مسئله عبارت از تعیین سهم چساپیک و اوهایو از درآمد کل بود.

در دوره ۶ ماهه تحت بررسی، تقریباً ۲۳۰۰۰ بارنامه در جامعه وجود داشت. چون در این بارنامه‌ها، هزینه‌های حمل بشدت متفاوت بودند (برخی به‌کمی ۲ دلار و برخی به زیادی ۲۰۰ دلار بود). تصمیم گرفته شد که از یک روش نمونه‌گیری موسوم به **نمونه‌گیری با طبقه‌بندی** استفاده شود. در این روش ابتدا بارنامه‌ها را در جامعه به زیرگروههای نسبتاً همگنی موسوم به **طبقه** تقسیم می‌کنند. زیرگروهها در این مثال برحسب میزان هزینه حمل تنظیم شدند. زیرا به نظر می‌رسید که سهم چساپیک و اوهایو از هر بارنامه با مبلغ کل بارنامه متناسب باشد. یعنی هرچه مبلغ کل بارنامه‌ها بیشتر باشد، سهم چساپیک و اوهایو نیز از آن بارنامه‌ها بیشتر است.

۱- *Pere Marquette*

طبقه‌ها به‌صورت زیرتعیین شدند:

طبقه	مبلغ بارنامه
۱-	۰ تا۵دلار
۲-	۵/۰۱دلارتا۱۰دلار
۳-	۱۰/۰۱دلارتا۲۰دلار
۴-	۲۰/۰۱ دلارتا۴۰دلار
۵-	۴۰/۰۱دلاروبیشتر

توجه داشته باشید که هر طبقه شامل بارنامه‌هایی با هزینه‌های حمل تقریباً یکسانی است. نظر به اینکه مبلغ پرداختی به چساپیک و اوهایو به‌طور کلی با هزینه کل و بارنامه تغییر می‌کند، هر طبقه از نظر هزینه حمل پرداختی به چساپیک و اوهایو نسبتاً همگن است. درحالی که طبقات مختلف اساساً از یکدیگر متمایزند.

برای تصمیم‌گیری درباره بزرگی حجم نمونه‌ای که از هر طبقه باید برگزید تا درآمد پرداختی به چساپیک و اوهایو با دقت لازم و براساس نمونه‌ای به حد ممکن کوچک برآورد شود، از نظریه‌های آماری استفاده به عمل آمد. یک اطلاع مورد نیاز برای تعیین حجم چنین نمونه‌هایی، تعداد بارنامه‌ها در هر طبقه جامعه بود. نرخ نمونه‌گیری طبقه‌ای به شرح زیر تعیین شد:

طبقه	نسبتی از جامعه که باید نمونه‌گیری شود (نرخ نمونه‌گیری)
۱-	٪۱
۲-	٪۱۰
۳-	٪۲۰
۴-	٪۵۰
۵-	٪۱۰۰

توجه کنید که این نظریه به نرخهای بالاتر نمونه‌گیری از طبقاتی که دامنه هزینه‌های حمل

گسترده‌تری دارند و به نرخهای کوچکتر نمونه‌گیری از طبقاتی که دامنه هزینه‌های حمل محدودتری دارند، منجر شد. برای درک این موضوع طبقه اول را درنظر بگیرید که شامل بارنامه‌هایی با هزینه‌های حمل صفر تا ۵ دلار است. در این طبقه تغییرات مبلغ بارنامه‌ها کم است و بنابراین یک نمونه کوچک از بارنامه‌ها می‌تواند اطلاعات کافی درباره تمام بارنامه‌های این طبقه ارائه کند. از طرف دیگر طبقه ۴، شامل بارنامه‌هایی با هزینه ۲۰/۰۱ دلار تا ۴۰ دلار، تغییرات بسیار بیشتری دارد. بنابراین به یک نمونه بزرگتر از این طبقه نیاز است تا اطلاعات کافی درباره تمام بارنامه‌ها به‌دست دهد. در یک وضعیت کرانگین غیرواقعی که در آن تمام بارنامه‌های یک طبقه معین مبلغ پرداختی یکسانی را به چساپیک و اوهایو نشان می‌دهند. نمونه‌ای متشکل بر فقط یک بارنامه، تمام اطلاعات مربوط به کل بارنامه‌های آن طبقه را ارائه می‌کند.

همین که حجم نمونه‌هامعین شد، کار بعدی انتخاب نمونه‌ای از هر طبقه بود. برای اینکه یک آمارشناس قادر به ارزیابی دقت نتایج نمونه شود (یعنی چقدر نتایج نمونه به مشخصات جامعه مربوطه محتملاً نزدیک است؟) نمونه باید براساس یک سازوکار احتمالاتی مشخص انتخاب شود. شیوه‌های مختلفی برای نمونه‌گیری احتمالاتی وجود دارند که یکی از آنها نمونه‌گیری تصادفی ساده خوانده می‌شود. این‌گونه نمونه را باید به‌طور مستقیم و با استفاده از یک مولد اعداد تصادفی یا با استفاده از یک جدول اعداد تصادفی که قسمتی از آن در جدول شماره ۱ ارائه می‌شود، انتخاب کرد. چگونه می‌توان از جدول ۱ برای انتخاب یک نمونه تصادفی ساده از هریک از طبقات استفاده کرد؟ طبقه ۱ را درنظر بگیرید و فرض کنید که شامل ۹۰۰۰ بارنامه است که با اعداد ۴ رقمی از ۰۰۰۱ تا ۹۰۰۰ مشخص شده‌اند. پس ما باید به‌دنبال اعداد چهار رقمی در این جدول باشیم و می‌توانیم از گوشه بالا و سمت چپ و از ستونهای ۱ تا ۴ آغاز کنیم. اولین عدد ۱۳۲۸ است. پس نخستین بارنامه نمونه، بارنامه شماره ۱۳۲۸ است. دومین بارنامه در نمونه، بارنامه شماره ۲۱۲۲ است. عدد بعدی در جدول اعداد تصادفی عدد ۹۹۰۵ است اما فقط ۹۰۰۰ بارنامه در طبقه وجود دارد. پس، از این عدد می‌گذریم و سراغ عدد بعدی می‌رویم که ۰۰۱۹ است. این فرایند تا دست‌یابی به ۹۰ واحد نمونه مورد نیاز (یک درصد ۹۰۰۰) ادامه می‌یابد. ارقام جدول اعداد تصادفی به‌گونه‌ای ایجاد شده‌اند که همه اعداد (اعداد چهار رقمی

جدول ۱ـ بخشی از یک جدول اعداد تصادفی

خط	(۱)-(۵)	(۶)-(۱۰)	(۱۱)-(۱۵)	(۱۶)-(۲۰)	(۲۱)-(۲۵)	(۲۶)-(۳۰)	(۳۱)-(۳۵)
۱۰۱	۱۳۲۸۴	۱۶۸۳۴	۷۴۱۵۱	۹۲۰۲۷	۲۴۶۷۰	۳۶۶۶۵	۰۰۷۷۰
۱۰۲	۲۱۲۲۴	۰۰۳۷۰	۳۰۴۲۰	۰۳۸۸۳	۹۴۶۴۸	۸۹۴۲۸	۴۱۵۸۳
۱۰۳	۹۹۰۵۲	۴۷۸۸۷	۸۱۰۸۵	۶۴۹۳۳	۶۶۲۷۹	۸۰۴۳۲	۶۵۷۹۳
۱۰۴	۰۰۱۹۹	۵۰۹۹۳	۹۸۶۰۳	۳۸۴۵۲	۸۷۸۹۰	۹۴۶۲۴	۶۹۷۲۱
۱۰۵	۶۰۵۷۸	۰۶۴۸۳	۲۸۷۳۳	۳۷۸۶۷	۰۷۹۳۶	۹۸۷۱۰	۹۸۵۳۹
۱۰۶	۹۱۲۴۰	۱۸۳۱۲	۱۷۴۴۱	۰۱۹۲۹	۱۸۱۶۳	۶۹۲۰۱	۳۱۲۱۱
۱۰۷	۹۷۴۵۸	۱۴۲۲۹	۱۲۰۶۳	۵۹۶۱۱	۳۲۲۴۹	۹۰۴۶۶	۳۳۲۱۶
۱۰۸	۳۵۲۴۹	۳۸۶۴۶	۳۴۴۷۵	۷۲۴۱۷	۶۰۵۱۴	۶۹۲۵۷	۱۲۴۸۹
۱۰۹	۳۸۹۸۰	۴۶۶۰۰	۱۱۷۵۹	۱۱۹۰۰	۴۶۷۴۳	۲۷۸۶۰	۷۷۹۴۰
۱۱۰	۱۰۷۵۰	۵۲۷۴۵	۳۸۷۴۹	۸۷۳۶۵	۵۸۹۵۹	۵۳۷۳۱	۸۹۲۹۵
۱۱۱	۳۶۲۴۷	۲۷۸۵۰	۷۳۹۵۸	۲۰۶۷۳	۳۷۸۰۰	۶۳۸۳۵	۷۱۰۵۱
۱۱۲	۷۰۹۹۴	۶۶۹۸۶	۹۹۷۴۴	۷۲۴۳۸	۰۱۱۷۴	۴۲۱۵۹	۱۱۳۹۲
۱۱۳	۹۹۶۳۸	۹۴۷۰۲	۱۱۴۶۳	۱۸۱۴۸	۸۱۳۸۶	۸۰۴۳۱	۹۰۶۲۸
۱۱۴	۷۲۰۵۵	۱۵۷۷۴	۴۳۸۵۷	۹۹۸۰۵	۱۰۴۱۹	۷۶۹۳۹	۲۵۹۹۳
۱۱۵	۲۴۰۳۸	۶۵۵۴۱	۸۵۷۸۸	۵۵۸۳۵	۳۸۸۳۵	۵۹۳۹۹	۱۳۷۹۰
۱۱۶	۷۴۹۷۶	۱۴۶۳۱	۳۵۹۰۸	۲۸۲۲۱	۳۹۴۷۰	۹۱۵۴۸	۱۲۸۵۴
۱۱۷	۳۵۵۵۳	۷۱۶۲۸	۷۰۱۸۹	۲۶۴۳۶	۶۳۴۰۷	۹۱۱۷۸	۹۰۳۴۸
۱۱۸	۳۵۶۷۶	۱۲۷۹۷	۵۱۴۳۴	۸۲۹۷۶	۴۲۰۱۰	۲۶۳۴۴	۹۲۹۲۰
۱۱۹	۷۴۸۱۵	۶۷۵۲۳	۷۲۹۸۵	۲۳۱۸۳	۰۲۴۴۶	۶۳۵۹۴	۹۸۹۲۴
۱۲۰	۴۵۲۴۶	۸۸۰۴۸	۶۵۱۷۳	۵۰۹۸۹	۹۱۰۶۰	۸۹۸۹۴	۳۶۰۳۶

منبع: *جدول ۱۰۵۰۰۰ ارقام تصادفی بر مبنای ده. کمیسیون بازرگانی بین ایالتی، دفتر اقتصاد و آمار ترابری. واشنگتن دی. سی. مه ۱۹۴۹.*

در مثال ما) احتمال یکسانی دارند.

شیوه دیگر انتخاب بارنامه‌ها از هر طبقه، **شیوه نمونه‌گیری اعداد پیاپی** خوانده می‌شود. این شیوه در عمل نیز توسط راه آهن چساپیک و اوهایو به‌کار برده شد. در این شیوه نمونه هر طبقه براساس ارقام معینی از شماره‌های پیاپی بارنامه‌ها انتخاب می‌شود. در این مورد خاص دو رقم آخر عدد پیاپی بارنامه به‌کار رفته‌است. برای توضیح اینکه چرا دو رقم آخر در انتخاب

نمونه به‌کار رفته است طبقه ۱ را با نمونه ۱ درصدی مدنظر قرار دهید. تعداد اعداد دورقمی که در ۲ ستون آخر اعداد پیاپی، یعنی اعداد ۰۰ و ۰۱ و ۰۲ و...و ۹۹ ظاهر می‌شوند ۱۰۰ تااست. اگر یکی از این عددهای دورقمی را از یک جدول اعداد تصادفی انتخاب نماییم و سپس تمام بارنامه‌هایی راکه شماره پیاپی آنها به این دو رقم ختم می‌شوند، انتخاب کنیم، در این صورت تقریباً ۱٪ واحدهای طبقه اول در نمونه منتخب گنجانده شده‌است. برای طبقه اول، عدد تصادفی منتخب ۰۲ بود.

تمام بارنامه‌های با هزینه حمل ۵ دلار یا کمتر که دو رقم آخر شماره پیاپی آنها ۰۲ بود به عنوان نمونه انتخاب شدند. رقمهای دوتایی از شماره پیاپی بارنامه‌ها که برای طبقات مختلف مورد استفاده واقع شد و نیز نرخ نمونه‌گیری از آنها به شرح زیراست:

طبقه	نسبتی ازاعضای طبقه که بایدنمونه گیری شود	عددی که شماره پیاپی بارنامه های منتخب به آن ختم می‌شوند
۱	٪۱	۰۲
۲	٪۱۰	۲
۳	٪۲۰	۲یا۴
۴	٪۵۰	۰۰تا۴۹
۵	٪۱۰۰	۰۰ تا۹۹

چون تمام بارنامه‌ها شماره پیاپی دارند، استفاده از این شیوه نمونه‌گیری بسیار آسان است. به‌علاوه در این مورد تجربه نشان می‌دهد که این روش هم اساساً نمونه‌ای معادل یک نمونه تصادفی ساده از هر طبقه است.

درمجموع ۲۰۷۲ بارنامه از ۲۲۹۸۴ بارنامه موجود در جامعه (۹٪کل) براساس این روش انتخاب شدند. برای هر بارنامه در نمونه، سهم درآمد حاصل از حمل راه‌آهن چساپیک و اوهایو محاسبه شد. سپس سهم این شرکت برای تمام بارنامه‌های هر طبقه برآورد شد که با جمع کردن این برآوردها، برآوردی از سهم درآمد حاصل از حمل راه‌آهن اوهایو و چساپیک برای تقریباً ۲۳۰۰۰ بارنامه موجود در جامعه به‌دست آمد. چون این کار یک آزمایش بود از بررسی کامل جامعه نیز انجام شد، بدین‌ترتیب امکان مقایسه نتایج نمونه با نتایج حاصل از تحلیل تمام

بارنامه‌های جامعه نیز فراهم شد. یافته‌ها به قرار زیر بودند:

سهم کل راه‌آهن چساپیک و اوهایو براساس بررسی کامل جامعه: ۶۴۶۵۱ دلار

سهم کل راه‌آهن چساپیک و اوهایو براساس نمونه منتخب: ۶۴۵۶۸ دلار

اختلاف: ۸۳ دلار

بنابراین نمونه‌ای مشتمل بر فقط ۹٪ بارنامه‌ها، برآوردی از سهم درآمد چساپیک و اوهایو را به‌دست داد که بامبلغ به‌دست آمده از یک بررسی کامل بارنامه‌ها فقط به میزان ۸۳ دلار اختلاف داشت. چون نمونه‌گیری بیش از ۱۰۰۰ دلار هزینه ندارد. در حالی که هزینه بررسی کامل جامعه حدود ۵۰۰۰ دلار است، برتری نمونه‌گیری به‌خوبی آشکار می‌شود. به هیچ وجه منطقی نیست که برای رفع یک خطای ۸۳ دلاری، ۴۰۰۰ دلار هزینه کرد. فراتر اینکه اگرچه در این مورد خاص خطا به ضرر چساپیک و اوهایو بود، اما دفعه بعد ممکن است به ضرر یک شرکت دیگر باشد، به‌طوری که در بلندمدت خطای تجمعی به‌طور نسبی حتی کمتر هم خواهدشد.

بررسیهای نمونه‌ای خطوط هواپیمایی و شرکتهای راه‌آهن دیگر

چساپیک و اوهایو همین آزمون را برای درآمدهای مسافری بین شرکتی به‌کار بردند. این مطالعه شامل بلیطهای فروخته شده طی یک دوره پنج ماهه به بازرگانان مسافر در منطقه چساپیک توسط شرکتهای راه‌آهن چساپیک و اوهایو و دو راه‌آهن دیگر (الف وب) بود. یافته‌ها در جدول ۲ ارائه می‌شوند. این نتایج هم توانایی نمونه‌های به‌طور نسبی کوچک را در تهیه برآوردهای دقیق ازکل درآمد متعلق به چساپیک و اوهایو به‌خوبی نشان می‌دهند.

خطوط هواپیمایی نیز از نمونه‌گیری آماری برای برآورد سهم خود از درآمد حاصل از بلیطهای مسافران دو یا چند خط هوایی استفاده کرده‌اند. سه شرکت هواپیمایی به آزمایش نمونه‌گیری آماری در طول یک دوره چهارماهه پرداختند. در آن زمان میزان خطای برآورد مبتنی بر نمونه‌های نسبتاً کوچک از ۰/۰۷٪ (یعنی ۷۰۰ دلار از ۱۰۰۰۰۰۰ دلار) برای هریک از سه شرکت هواپیمایی تجاوز نکرد. براساس این تجربه استفاده گسترده‌تری از نمونه‌گیریهای آماری برای تنظیم حسابها بین شرکتهای هواپیمایی به‌عمل آمده‌است. یک بار که نمونه تقریباً ۱۲٪ بلیطهای فروخته شده هواپیماها را شامل می‌شد، خطای نمونه‌گیری تجمعی از ۱٪ کمتر شد.

جدول ۲ـ نتایج مطالعه بلیطهای مسافری

	بررسی‌کامل(۱۰۰درصدی)	نمونه ۵٪	اختلاف	
راه آهن (الف)			به دلار	به درصد
تعدادکل بلیطها	۱۴۱۰۹			
درآمدکل	۳۲۵۶۰۰دلار			
سهم چساپیک واوهایوازدرآمدکل	۲۱۲۱۶۴دلار	۲۱۲۰۶۳دلار	-۱۰۱	۰/۵٪
راه آهن (ب)				
تعدادکل بلیطها	۷۵۶۲			
درآمدکل	۱۲۸۵۰۳دلار			
سهم چساپیک واوهایوازدرآمدکل	۷۹۷۱۰دلار	۸۰۰۵۷دلار	۳۴۷	۰/۴۴

مأخذ: *راه‌آهن ایج*، ۹ژوئن ۱۹۵۲.

چنین برآورد شد که صرفه‌جوییهای اداری شرکتهای بزرگ هواپیمایی سالانه ۷۵۰۰۰ دلار وصرفه‌جوییهای صنعت بیش از ۵۰۰۰۰۰ دلار بود.

از نمونه‌گیری آماری در حسابداری و حسابرسی برای برآورد ارزش موجودی انبار، نسبت مانده حسابهای بدهکاران ۶۰ روزه یابیشتر و نسبت مانده حسابهای بدهکاران مورد تأیید مشتریان نیز استفاده شده‌است. در هر مورد نشان داده شده‌است که یک نمونه به‌طور نسبی کوچک که به دقت انتخاب و بررسی شده باشد، به‌خوبی می‌تواند نتایجی باکیفیت بالا و هزینه بسیار کمتر از یک بررسی کامل ارائه کند.

نمونه‌گیری برای تعیین حق امتیاز رادیویی

زمینه دیگری که در آن از نمونه‌گیری آماری با مقاصد حسابداری استفاده می‌شود، تعیین حق امتیاز برای تهیه‌کننده و توزیع‌کننده موسیقی پخش شده از رادیو است. یک مؤسسه حقوقی، BMI، برای صدور مجوز پخش موسیقی موکلان خود از ایستگاههای رادیویی حق عضویت

می‌گرفت. هر ایستگاه یک حق اشتراک عام، متناسب با درآمد خود، می‌پردازد. سپس این مبالغ به‌طور متناسب بین تهیه‌کننده و توزیع‌کننده موسیقی پخش شده، تقسیم می‌شود.

اطلاعات مربوط به موسیقی پخش شده از ایستگاههای رادیویی را نمی‌توان براساس سرشماری ۱۰۰ درصدی به‌دست آورد، زیرا بسیار پرهزینه است. حدود ۹۰۰۰ ایستگاه رادیویی به پخش موسیقی برای میانگین ۱۸ساعت در شبانه‌روز می‌پردازند. پس برای ۳۶۵ شبانه‌روز در هرسال، تقریباً ۵۹ میلیون ساعت موسیقی توسط ایستگاههای رادیویی پخش می‌شود.

درمقابل BMI نمونه‌ای از ایستگاههای رادیویی را برای تعیین موسیقی پخش شده، انتخاب می‌کند. این کار را ابتدا با طبقه‌بندی همه ایستگاههای رادیویی ایالات متحده برحسب شهر، منطقه و میزان موسیقی پخش شده در شبانه‌روز انجام می‌دهد. این کار با استفاده از یک برنامه کامپیوتری انجام می‌شود که ایستگاهها را برحسب تشابه آنها از نظر درصد موسیقی BMI پخش‌شده نیز گروه‌بندی می‌کند. این طبقه‌بندی سالانه انجام می‌شود و معمولاً به ۸ یا ۹ طبقه منجر می‌شود. سپس هر طبقه به ۲ زیرطبقه برحسب میزان حق اشتراک پرداختی به BMI تقسیم می‌شود، به‌طوری که ایستگاههایی با حق اشتراک کم در طبقه‌ای متمایز از ایستگاههایی با حق اشتراک زیاد قرار می‌گیرند. طبقه دیگری نیز علاوه براین دو طبقه، برای ایستگاههای جدیدی که در طول سال فعالیت خود را آغاز می‌کنند، تشکیل می‌شود.

سپس نمونه‌ای از ایستگاههای رادیویی به‌طور تصادفی در هر فصل از هر طبقه انتخاب می‌شود. با این فرض که روزها در طول فصل توزیع یکنواخت دارند. برای هر ایستگاه منتخب، ۲ یا ۳ شبانه‌روز از فصل به‌طور تصادفی تعیین می‌شود. هر ایستگاه منتخب نوار آهنگهای پخش‌شده در طول این روزها را نگهداری می‌کند. سپس این نوارها به BMI داده می‌شود. اطلاعات به‌کامپیوتر منتقل می‌شود و موسیقی پخش شده از نظر اینکه BMIنماینده آن است یا خیر، شناسایی می‌شود. از این اطلاعات نمونه، میزان پرداختی BMI به هر تهیه‌کننده و توزیع‌کننده موسیقی در هر فصل پیش‌بینی می‌شود. در طول یک سال از تمام ایستگاههای رادیویی خواسته می‌شود تا به ثبت و نگهداری نوار موسیقی در یک دوره خاص بپردازند.

جمع‌بندی

به‌طور خلاصه نمونه‌گیری آماری مشتمل بر انتخاب تعدادی از اقلام جامعه است و انتخاب نمونه به گونه‌ای انجام می‌شود که هر نمونه ممکن از جامعه دارای احتمال انتخاب معلومی باشد. یک نمونه آماری اغلب می‌تواند اطلاعات قابل اعتمادی را با هزینه به مراتب کمتر از هزینه یک بررسی کامل ارائه کند. همچنین یک نمونه آماری اغلب می‌تواند داده‌های بهنگامتری را نسبت به یک بررسی کامل جامعه ارائه نماید، زیرا داده‌های کمتری باید جمع‌آوری شوند و مقدار کمتری از داده‌ها را باید پردازش کرد. بالاخره گاهی اوقات یک نمونه آماری حتی می‌تواند اگر کنترل کیفی جمع‌آوری داده‌ها در این مقیاس کوچکتر به‌صورتی کاراتر انجام شود. اطلاعات دقیقتری را نسبت به یک بررسی کامل ارائه کند.

مسائل

۱ـ تفاوت نمونه‌گیری تصادفی ساده و نمونه‌گیری اعداد پیاپی را توضیح دهید.

۲ـ فرض کنید رییس دانشگاهی در حال بررسی سفارش میز و صندلیهای جدید برای کلاسهای درس است. وی نیازمند آن است که بداند چه تعداد میزهایی که فعلاً از آنها استفاده می‌شود باید عوض شوند.

الف) آیا رییس دانشگاه باید از روشهای نمونه‌گیری استفاده کند؟ دلایل لازم به نفع انجام نمونه‌گیری چه هستند؟ علیه آن کدامند؟

ب) اگر وی از روشهای نمونه‌گیری استفاده کند، جامعه کدامست؟

۳ـ چرا در بررسی حمل بار چساپیک و اوهایو از نمونه‌گیری با طبقه‌بندی استفاده شد؟

۴ـ با مراجعه به جدول ۲، جمع درآمدهای حاصل از فروش بلیط شرکتهای راه‌آهن الف و ب را به‌دست آورید و سپس درصد اختلاف نتایج نمونه ۵٪ و بررسی کامل ۱۰۰درصدی را محاسبه کنید.

۵ـ در بررسی حمل بار چساپیک و اوهایو، درصدی از درآمد کل متعلق به چساپیک و اوهایو که از خطای نمونه‌گیری ناشی شده‌بود، چقدر بود؟

۶ـ یک روانشناس ارتش می‌خواهد نمونه‌ای از ۱۰۰۰ سرباز اعزامی را برای بررسی نظر آنان

نسبت به «ارتش جدید» برگزینند. وی فهرستی از اسامی ۱۰۰۰۰ سرباز اعزامی را که برحسب گروههای نظامی مرتب شده‌اند، در اختیار می‌گیرد. هر گروه شامل ۱۰ عضو، یعنی یک گروهبان به عنوان ارشد، یک سرجوخه و ۸ سرباز عادی است.

الف) آیا شما به این روانشناس استفاده از نمونه‌گیری اعداد پیاپی (بااستفاده از رقمهای ۰ تا ۹) را برای انتخاب نمونه‌ای ۱۰۰۰تایی از فهرست ۱۰۰۰۰نفری توصیه می‌کنید. چرا؟

ب) اگر این روانشناس از نمونه‌گیری اعداد پیاپی استفاده کند، احتمال اینکه فقط گروهبانها در نمونه ظاهر شوند، چقدر است؟ اگر از نمونه‌گیری تصادفی ساده استفاده کند، چطور؟

ج) درصورتی که روانشناس از فهرستی استفاده کند که در آن نام ۱۰۰۰۰ سرباز اعزامی به ترتیب حروف الفبا درج شده باشد، به سئوالهای الف و ب پاسخ دهید.

۷ـ فرض کنید که راه‌آهنهای الف و چساپیک و اوهایو به نمونه‌گیری از بلیطها برای تعیین سهم ماهانه خود از دریافتی از مسافران راه‌آهن برای یک سال مبادرت کنند. انتظار دارید که برای چندماه خطای نمونه‌گیری به نفع چساپیک و اوهایو باشد؟

۸ـ از جدول ۱ برای انتخاب یک نمونه تصادفی مرکب از ۲۵ عدد دورقمی استفاده کنید. چند عدد زوج وجوددارند؟ چند عدد وجود دارد که هر دو رقم آنها زوج باشد؟ همین کار را برای یک نمونه تصادفی مرکب از ۱۰۰ عدد دورقمی تکرار کنید. پاسخهای خود را با آنچه دیگران یافتند، مقایسه کنید. چه نتیجه‌ای می‌توان گرفت؟

منابع

A. A. Arens and J. K. Loebbecke. 1981. *Applications of Statistical Sampling to Auditing* Englewood Cliffs, N. J.: Prentice-Hell.

"Can Scientific Sampling Techniques Be Used in Railorad Accouting ?" *Railway Age,* June9, 1952, PP. 61-64

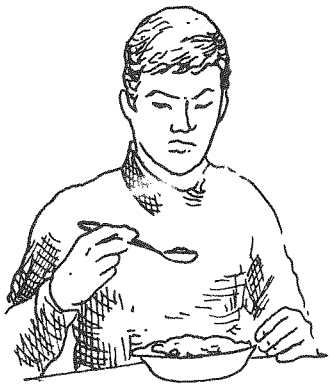
D. A. Leslie, A. D. Teitlebaum, and R. J. Anderson. 1979. *Dollar- Unit Sampling: A Practical Guide for Auditors.* Toronto: Copp Clark Pitman.

T. W. McRac. 1974. *Statistical Sampling for Aufit and Control.* London: Wiley.

M. S. Newman. 1982. *Accounting Estimates bY Computer Sampling,* 2nd ed. NewYork: Wiley.

T. M. F. Smith. 1976. *Statistical Sampling for Accountants.* London: Accountancy Age Books.

A. J. Wilburn. 1984. *Practical Statistical Sampling fot Auditors.* New York: Marcel Dekker.



ارزیابی مقدماتی از یک فراورده غذایی جدید[1]

الیزابت استریت و ماویس بی. کارول

بسیاری از آمریکاییها دوست دارند در مواقع تعجیل، غذایی دردسترس داشته‌باشند که مغذی بوده و تهیه آن آسان باشد. شرکت ما در حال پدیدآوردن چنین فراورده‌ای به‌نام H است.

تولید چنین فراورده‌ای یک ارزیابی کلی را ایجاب می‌کند. در این مقاله ما توجه خود را به دو

۱ـ *این مقاله در سال ۱۹۷۲ تهیه شده‌است و در آن زمان مؤلفان مقاله، کارمند شرکت General Foods Corporation بوده‌اند و وضعیت آن زمان را منعکس کرده‌اند ـ ویراستار.*

جنبه از ارزیابی معطوف می‌کنیم: محتوای پروتئین H وطعم آن؛ و نشان خواهیم داد که آمار چگونه درپاسخ به این سئوالها نقشی اساسی دارد.

خوش‌طعمی یک فراورده را می‌توان با طعم مردم‌پسند فراورده تهیه‌شده و ارزیابی مقبول بودن آن، هم ازنظر کلی و هم ازنظر ویژگیهای خاص آن تعیین کرد. کیفیت غذایی یک فراورده غذایی را می‌توان با دادن آن به حیواناتی که فرایندهای متابولیک آنها شبیه انسان است، معین کرد. هر دونوع آزمایش باید طراحی شوند، یعنی به‌صورتی تفصیلی از پیش طرح‌ریزی شوند. این کار به اجتناب از خطای مستمر در یک جهت (اریبی) کمک می‌کند و اطمینان می‌دهد که تعداد خاصی از انسانها و حیوانات که در هر بررسی گنجانیده شده‌اند، با درجه اطمینان مناسبی به سئوالهای مطروحه پاسخ دهند.

بعضی از مراحل طرح و تحلیل چنین آزمونهایی در زیر توصیف شده‌اند:

ارزیابی پروتئین

محتوای پروتئین H، به یک معنا، به صورتی که محاسبات مبتنی بر اجرای ترکیبی H نشان می‌دادند، رضایتبخش بود اما ما درباره کارایی پروتئین در H، تحت شرایط استفاده واقعی، اطمینان نداشتیم و به‌علاوه می‌خواستیم دو صورت H، یکی به‌صورت جامد و دیگری به‌صورت مایع را باهم مقایسه کنیم.

یک مطالعه درباره تغذیه موشها، می‌توانست پشتیبانی عملی ادعای میزان بالای پروتئین و راه مقایسه دونوع H را به‌دست دهد، زیرا پاسخ موشها تحت‌تأثیر هر اثر متقابل مواد موجود در فرمول و یا تحت‌تأثیر کمبود یک آمینواسید اساسی است که کارایی پروتئین را کم می‌کند. هیچ‌یک از این شرایط را محاسبه دستی نشان نخواهدداد.

تجربه قبلی نشان داده که یک رژیم غذایی ۲۸روزه برای ۱۰ تا ۱۵ موش، یک براورد نسبتاً قابل اعتماد ازکارایی پروتئین رژیم را به‌دست می‌دهد. در این بررسی تغذیه، مثل آنچه در همه چنین بررسیهایی صورت می‌گیرد، از موشهای نری استفاده می‌شود که تازه مکیدن شیر از پستان مادر را ترک کرده‌اند. موشهای نر سریعتر از موشهای ماده رشد می‌کنند و وقتی شیر خوردن از پستان مادر را ترک می‌کنند در دوره حداکثر رشدند. از موشهای بالغ به‌دلیل پایداری

وزنشان و به‌دلیل اینکه افزایش وزن حیوانات در درجه اول اهمیت قرار دارد، استفاده نمی‌شود. آزمایش، علاوه بر مقایسه دونوع H ، هریک از آنها را با رژیم غذایی شاهد کازئین که دو هدف را دنبال می‌کند، مقایسه می‌نماید. اولاً، این رژیم استانداردی است که با آن بسیاری از رژیمهای آزمایشی مقایسه می‌شوند؛ ثانیاً به دلیل اینکه تجربه زیادی با رژیم کازئین داشته‌ایم، روشی برای این بازبینی فراهم می‌کند که آیا چیزی در رابطه با دسته موشها یا در رابطه با جنبه‌های دیگر بررسی نادرست بوده‌است یا نه؟ گاهی کارها به راههایی اسرارآمیز به صورتی نادرست جریان دارند و داشتن نوعی راه بازبینی مهم است.

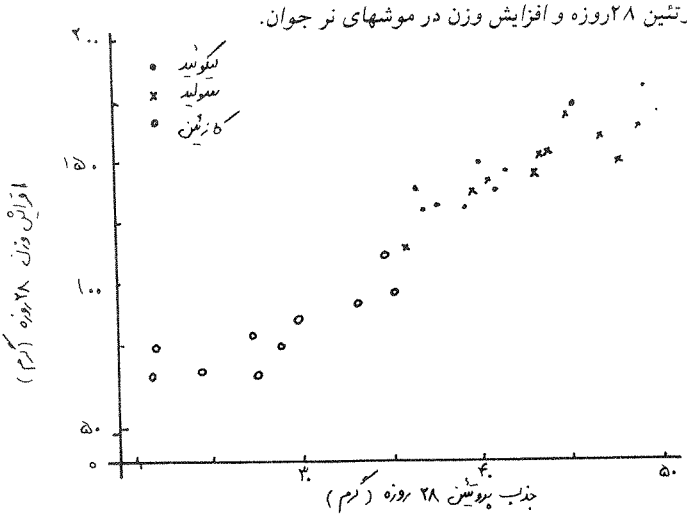
در آغاز آزمایش، از ۳۰ موش که وزن آنها از ۵۰ تا ۶۳ گرم تغییر می‌کرد، استفاده شد. آنها را برحسب مقادیر صعودی وزن مرتب کردند و هریک از ۳ موش با کمترین وزن را به یکی از سه رژیم غذایی آزمایش تخصیص دادند. چنین سه‌گانه‌ای از حیوانات تقریباً هم‌وزن را یک **بلوک** نامیدند. تخصیص موشها به رژیمهای سه‌گانه بلوک برحسب شانس یا تصادف انجام شد، یعنی اعداد تصادفی تعیین کردند که کدام موش به کدام رژیم غذایی تخصیص یابد. در بلوک دوم متشکل از ۳ موش با کمترین وزنهای بعدی، به همین طریق تخصیص ۳ رژیم غذایی صورت گرفت. این فرایند تخصیص بلوکی تصادفی‌شده تا وقتی ادامه یافت که همه ۳۰موشی که برای این بررسی معین شده‌بودند، بین سه رژیم غذایی، هر ده‌تا یک رژیم، توزیع شدند.

استفاده از بلوکهای موشهایی با وزنهای قابل‌مقایسه اطمینان می‌دهد که هررژیم غذایی سهم خاص خود را از موشهای سبک وسنگین دارد. تصادفی‌سازی، به تعادل عاملهای دیگری که ممکن است برآمد بررسی را تحت‌تأثیر قراردهد، کمک می‌کند. مثلاً مجموعه ۳۰موش احتمالاً شامل بسیاری از موشهای همزاد بود. همه موشها بتازگی شیرخوردن از پستان مادر را رها کرده و بدین طریق تقریباً همزمان متولد شده‌بودند. احتمال اینکه تصادفی‌سازی، موشهای یک زایمان را بین سه رژیم غذایی پراکنده‌کند، زیاد بود. البته برای اطمینان از اینکه موشهای همزاد به تساوی بین سه رژیم غذایی تقسیم شده‌اند، لازم بود که موشهای همزاد را شناسایی کرده و با استخراج تصادفی حیوانات، چندین سه‌تایی از موشهای همزاد بوجود آورند و موشهای باقیمانده را به تساوی و به‌صورت تصادفی به سه رژیم غذایی تخصیص دهند. اگر صفت مشخصه‌ای موروثی بر نتایج مطالعه به‌صورتی مهم اثر کند، این تخصیص باید انجام شود.

موشها به‌وسیله اعداد تصادفی به قفسها تخصیص داده شدند، زیرا مطالعات پیشین نشان داده‌بودند که موشها در طبقه بالای قفسها وزنی بیشتر از موشهای طبقه پایین قفسها به‌دست می‌آورند. این تصادفی‌سازی، محافظت علیه این امکان است که مبادا بیشتر موشهای با یک رژیم غذایی در قفسهای بالایی قرار گیرند و درنتیجه موشهای با رژیم غذایی دیگر در قفسهای پایین باشند.

در بررسی تغذیه با H، هر موش مجاز بود که هرقدر می‌خواهد، بخورد. مصرف غذایی او اندازه‌گیری و وزنش هرهفته کنترل می‌شد. افزایش نهایی وزن صرفاً تفاوت بین وزن نهایی او پس از ۲۸ روز و وزن اولیه او بود. وزنهای واسط برای بازبینی هر پیشامد ناگوار نظیر عفونت دستگاه تنفسی به‌کار می‌رفتند، زیرا وزن بدن موش نشانه‌ای حساس از سلامتی عمومی است. کل غذای مصرفی ضربدر نسبت پروتئین ـ مقدار حاصل به‌وسیله تجزیه شیمیایی نمونه‌های هر رژیم غذایی ـ پروتئین جذب‌شده هر موش را به‌دست می‌دهد. برای بررسی H، نتایج این دو اندازه در شکل ۱ نشان داده شده‌است. هر نقطه روی نمودار، معرف یک موش است. هرسه رژیم غذایی طوری تهیه شده‌بودند که یک نسبت پروتئین برابر با ۹٪ داشتند. تجزیه شیمیایی نشان داد که پروتئین آنها نزدیک هم هستند: ۹/۸۷۵٪، ۹/۸۷۵٪، و۹/۵٪. بنابراین، جذب بیشتر پروتئین برای موشهایی که با رژیمهای H تغذیه شده‌اند، بدین معناست که آنها بیشتر غذا خورده‌اند، که ممکن است معرف آن باشد که این رژیمها برای موشها مطبوعتر از رژیم غذایی

شکل ۱ـ بستگی جذب پروتئین ۲۸روزه و افزایش وزن در موشهای نر جوان.



کازئین است. اما متخصصان مجرب تغذیه اظهار می‌دارند که موشها عموماً هر غذای با پروتئین بالای متعادل را صرف‌نظر از ترکیب وطعم، با اشتها می‌خورند. همه آنچه از داده‌ها معلوم می‌شود، آن است که جذب H اساساً بیش از جذب کازئین بوده‌است. واقعیت واضحتر دیگر درباره نقاط شکل ۱ آن است که این نقاط به افتادن در نزدیکی خط شیب‌داری گرایش دارند که در میان آنها رسم می‌شود. اما با یک خط‌کش، معاینه دقیقتر شکل ۱ نشان می‌دهد خط راستی که بدون ابزار از میان نقاط H مایع رسم می‌شود، کمی بالای خطی است که ازمیان نقاط H جامد عبور می‌کند و شیب تندتری دارد. اگر خط کازئین به سمت جذب بیشتر و پروتئین ادامه داشت، خط مار بر تنها نقاط کازئین باید زیر هریک از خطوط H باشد. پس تردیدی در نتیجه‌گیری اولیه قرارگرفتن همه نقاط بر یک خط و لذا در این فرض که هر تفاوت بین افزایش وزن ناشی از رژیم غذایی را می‌توان تماماً به تفاوتهای جذب پروتئین نسبت داد، بوجود آمد. یعنی به این فکر افتادیم که تفاوت در افزایش وزنها تنها ناشی از تفاوت جذب پروتئین نیست.

ما از یک برنامه کامپیوتری برای برازش خط راست به نقاط مربوط به هر رژیم غذایی، به‌وسیله روش کمترین توانهای دوم، استفاده کردیم. یعنی خطی انتخاب شد که مجموع توانهای دوم فاصله‌های قائم نقاط از خط را مینیمم کند. این خطهای برازانده شده که به **خطهای رگرسیونی براورد شده** موسومند، به‌وسیله معادله‌هایی به‌صورت $Y=\overline{Y}+b(x-\overline{x})$، مشخص می‌شوند که در آن، Y افزایش وزن پیش‌بینی شده، و x پروتئین جذب شده‌است. $\overline{X}$ و $\overline{Y}$ متوسطهای Xها و Yها، جداگانه برای هرمجموعه از داده‌ها هستند و هر b معرف شیب آن خط راست، یعنی نمو پیش‌بینی شده، در افزایش وزن به‌ازای هرواحد پروتئین جذب شده است.

سه معادله رگرسیونی عبارت بودند از:

H مایع : $Y = ۱۵۱+۳/۷۲ (X-۴۱/۷)$

H جامد : $Y = ۱۵۰+۳/۶۶ (X-۴۳/۳)$

کازئین : $Y = ۷۹+۲/۹۱ (X-۲۸/۴)$

بررسی سه خط، که در شکل ۲ رسم شده‌اند، نشان می‌دهد که خط کازئین قطعاً پایینتر و شیب آن کمتر از دوخط H است. سپس این دوخط H باهم مقایسه شدند و دریافتیم که آنها

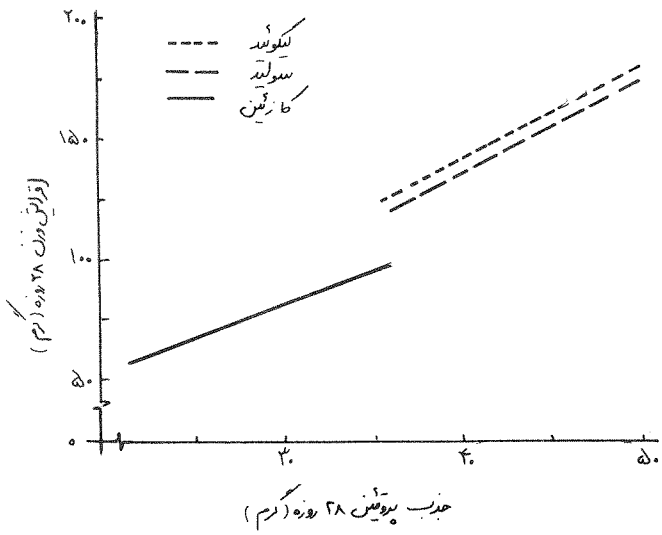
تفاوتی بیش از تفاوت تصادفی باهم ندارند. پس شک ما به این صورت تأیید شد که افزایشهای وزن بزرگتر موشهایی که با H تغذیه شدند، نسبت به افزایشهای وزن موشهایی که با کازئین تغذیه شدند، تنها ناشی از جذب بیشتر پروتئین نبوده‌است؛ زیرا شیبهای خطوط H بیشترند.

خلاصه نتایج محاسبه پروتئین اینکه: تفاوتها در پروتئین جذب‌شده بین موشهای با غذای کازئین و موشهای با غذای H، عامل همه تفاوتها در افزایش وزن نبود. برای یک افزایش جذب پروتئین مفروض، رژیمهای غذایی H، افزایش بیشتر وزنی را نسبت به رژیم غذایی کازئین نتیجه می‌دهد. بین H مایع و H جامد، از لحاظ افزایش وزن براوردشده، تفاوت کمی وجود داشت. اینکه آیا افزایش عاملهای جامدکننده مضر است یانه؟ موضوعی است که دراین سنجش آشکار نشد.

ارزیابی خوش طعمی

ضمن اینکه بررسی تعیین و مقایسه ارزش غذایی دو نوع H جریان داشت، یک آزمون مقدماتی از دید مصرف‌کنندگان برای تعیین خوش طعمی دونوع H نیز انجام می‌شد. هدف، مقایسه دو فراورده H و ملاحظه این بود که آیا آنها درکل به اندازه فراورده رقیبی که قبلاً در بازار بود وآن

شکل ۲ـ خط رگرسیونی براوردشده از روی افزایش وزن ۲۸روز نسبت به پروتئین جذب‌شده برای موشهای نر جوان



را C می‌نامیم، قابل قبول هستند یا نه؟

تجربه قبلی نشان داده‌است که چشیدن طعم غذا توسط ۵۰نفر و ارزیابی یک فراورده تحت شرایط کنترل‌شده، برای آشکارکردن مشکل عمده‌ای در استقبال از آن، در صورت وجود چنین مشکلی کافی است. شرایط کنترل شده با آوردن افراد به یک مکان مرکزی آزمون و خواستن از و ۵۰ نفر برای چشیدن هریک از دونوع H، پذیرایی کنند و با آنها مصاحبه نمایند، حاصل می‌شود. چون مردان و زنان ممکن است پاسخ متفاوت داشته‌باشند، قرار شد که در آزمون، هر فراورده را ۲۵ مرد و ۲۵ زن بچشند. چشندگان که برای شرکت در آزمون، پول دریافت می‌کردند از کلیساهای محلی یا اعضای باشگاهها استخدام شدند. پس از استخدام، تخصیص تصادفی افراد برای چشیدن هریک از سه فراورده تا وقتی ادامه یافت که ۲۵ نفر از هرجنس هریک از سه فراورده را چشیده و ارزیابی کردند.

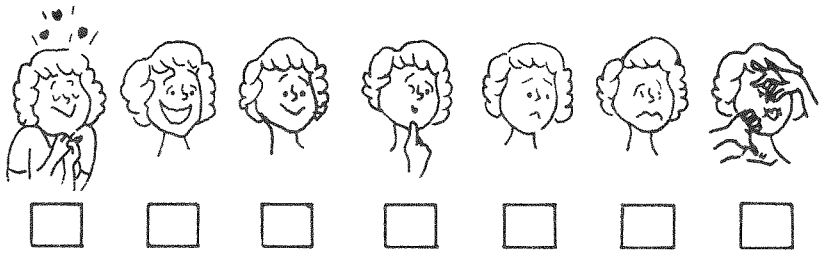
بعد از چشیدن فراورده، از هرفرد خواسته‌شد تا در یک برگه نظرخواهی علامت بزنند که قابلیت قبول کلی فراورده را رتبه‌بندی می‌کرد. شکل ۳ مقیاس تصویری مورد استفاده برای اندازه‌گیری مقبولیت را نشان می‌دهد. (به‌نظر می‌رسد که برای افراد، بیان احساسشان درباره فراورده، با استفاده از مقیاس تصویری، آسانتر از استفاده از واژه‌هایی مانند **عالی، خیلی‌خوب، خوب، متوسط، بد، خیلی بد، و وحشتناک** است.) به هر تصویر عددی (یانمره‌ای) از دنباله ۳+، ۲+، ۱+، ۰، ۱−، ۲−، ۳− تخصیص یافته‌است که ۳ برای بیشترین قابلیت قبول است.

برای هریک از سه فراورده، تعداد کل رأیهای هر تصویر به تفکیک مرد و زن شمارش شد. نتایج که **توزیعهای فراوانی** نامیده می‌شوند، در جدول ۱ آمده‌اند که در آن اعداد یا نمره‌ها جانشین تصویرها شده‌اند و M و F به‌ترتیب معرف مرد وزن است.

از توزیعهای فراوانی به‌نظر می‌رسد که همه آنها که یک فراورده را چشیدند، در مقبول بودن آن توافق نداشتند. مشکل است که با نگاه‌کردن به شش توزیع، تصمیم گرفت که آنچه را یکی از گروههای ۲۵تایی چشیده و نمره داده‌اند نسبت به آنچه پنج گروه دیگر چشیده و نمره دادند مقبولیت بیشتر یا کمتری دارد. برای تسهیل مقایسه ۶گروه ۲۵تایی، با ضرب هرنمره در فراوانی آن و تعیین مجموع نتایج، و تقسیم آنها بر ۲۵، یک نمره متوسط به‌دست آمد.

توجه می‌کنیم که وقتی چشنده‌های مرد و زن باهم مقایسه شدند، تفاوتهای میانگینهای

شکل ۳ـ برگه نظرخواهی آزمون طعم به‌صورت تصویری (به شکلها از چپ به راست، نمرات ۳- تا ۳+ تخصیص یافته‌اند.)



در مستطیل بالا تصویری را که مبین احساس شما درباره فراورده‌ای است که چشیده‌اید، با علامت مشخص کنید.

نمرات زیاد نبودند. همچنین توجه می‌کنیم که میانگینها برای H مایع وH جامد نیز تفاوت زیادی ندارند، اما هر چهار میانگین H کاملاً بالای دو میانگین C هستند. سئوالی که باید پاسخ داد، این است که باتوجه به متغیربودن نظر افراد وقتی یک فراورده را رتبه‌بندی می‌کنند، آیا تفاوتهای میانگینهای نمرات، بزرگتر از آن هستند که از شانس انتظار می‌رود یانه؟ آنچه مورد نیاز است، معیاری است که بتوانیم براساس آن بگوییم چه تفاوتی بین هردو میانگین که هریک بر پایه داده‌های ۲۵ چشنده‌است، بزرگتر از آن است که بتوان تنها از شانس انتظار داشت.

برای نیل به چنین معیاری باید ابتدا میزان متغیربودن افراد درون هریک از شش گروه را اندازه بگیریم. این اندازه که **واریانس** نامیده شده‌است، با تعیین تفاضل نمره هر فرد از میانگین، مربع کردن تفاضلها، تعیین مجموع مربعها و تقسیم نتیجه بر یکی کمتر از تعداد افراد به‌دست می‌آید؛ برای هریک از شش گروه واریانس محاسبه شد. این واریانسها در سطر زیرین جدول توزیعهای فراوانی نشان داده شده‌اند. سپس میانگین واریانسهای شش گروه به‌دست آمد که ۱/۳۲ است و اندازه خوبی از تغییرپذیری بین چشایی مردان و زنان برای یک فراورده می‌باشد. اگر افراد تغییر کنند، میانگینهای مبتنی بر افراد هم تغییر می‌کنند. اینکه میانگینها چقدر تغییر کنند، به تعداد چشنده‌ها بستگی خواهد داشت: هرچه تعداد بیشتر باشد، میانگین نماینده

جدول ۱ـ توزیع فراوانی نمره‌ها

	C		فراورده چشیده‌شده H مایع		H جامد	
نمره	M	F	M	F	M	F
+۳	۱	۰	۴	۳	۲	۳
+۲	۲	۳	۶	۷	۶	۵
+۱	۷	۸	۹	۷	۱۰	۹
۰	۸	۹	۵	۶	۶	۷
-۱	۵	۳	۰	۲	۱	۰
-۲	۲	۱	۱	۰	۰	۱
-۳	۰	۱	۰	۰	۰	۰
کل چشنده‌ها	۲۵	۲۵	۲۵	۲۵	۲۵	۲۵
میانگین نمره‌ها	۰/۲۰	۰/۲۴	۱/۲۴	۱/۱۲	۱/۰۸	۱/۰۴
واریانس	۱/۵۰	۱/۴۳	۱/۴۴	۱/۳۶	۰/۹۹۳	۱/۲۰

بهتری بوده و کمتر متغیر خواهدبود. به دلیل این تغییرپذیری، ما هرگز نمی‌توانیم مطمئن باشیم که دو میانگین معرف تفاوت واقعیند، اما باید در استخراج نتایج مخاطره‌ای را بپذیریم. بنابراین بهترین کاری که می‌توانیم انجام دهیم، منظورکردن مخاطره در تعیین مقیاس است. ما تصمیم گرفتیم که یک شانس "۱ به ۲۰" را برای امکان خطاکردن هنگامی که مقدار تفاوت بین دو میانگین را محاسبه می‌کنیم و نتیجه می‌گیریم که تفاوتهای واقعی را نشان می‌دهند، اختیار کنیم. با استفاده از واریانس میانگین ۲۵ نمره و به حساب آوردن مخاطره، برای مقایسه هردو مورد از میانگین نمرات، مقیاس ۰/۶۴ به‌دست آمد.

این، به ما می‌گوید شواهدی دردست نیست که مردان و زنانی که یک فراورده را رتبه‌بندی می‌کنند، رتبه‌بندی متفاوتی دارند، زیرا تفاوتهای میانگینها به‌صورت زیر هستند:

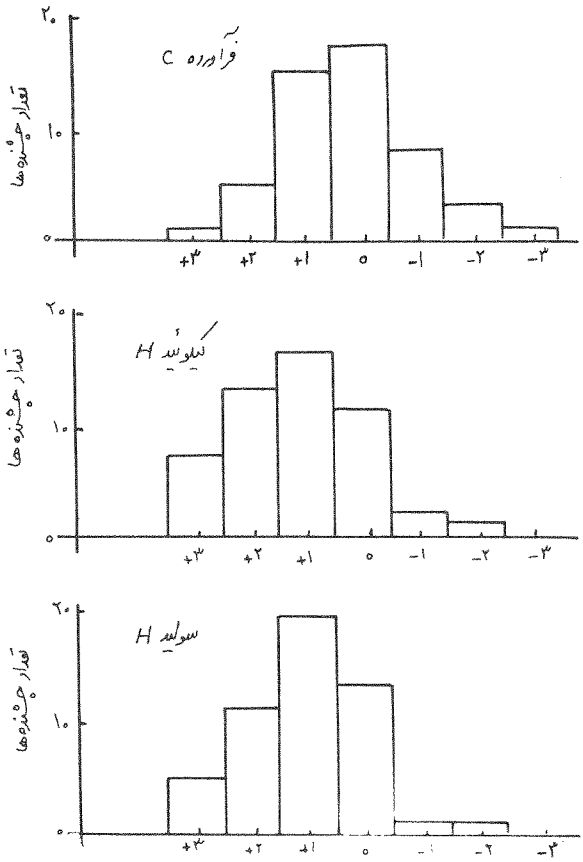
C: ۰/۲۴-۰/۲۰=۰/۰۴

H مایع : ۱/۲۴-۱/۱۲=۰/۱۲

H جامد ۱/۰۸-۱/۰۴=۰/۰۴

حال می‌توانیم توزیعهای مربوط به مردان و زنان را ترکیب کنیم. توزیعهای حاصل را در

شکل ۴ـ توزیعهای نمرات چهره ـ مقیاس از ترکیب مردان و زنان، حاصل از چشیدن فراورده، ۵۰ چشنده برای هر فراورده



شکل ۴ نشان داده‌ایم. توجه می‌کنیم که توزیع فراورده C به راست انتقال یافته که در مقایسه با نتایج H، نمراتی کمتر قابل قبولند. همپوشی قابل‌ملاحظه‌ای از توزیعها وجود دارند که معرف‌تغییراتند، بنابراین حتی بر پایه تصویر هم در شکل تعبیر مقبولیت نسبی فراورده‌های مورد آزمون مشکلاتی داریم. باید به میانگین نمرات متوسل شویم که برای ۵۰ فردی که یک فراورده را چشیده‌اند، عبارتنداز:

H جامد $= ۱/۰۶$ $\qquad$ H مایع $= ۱/۱۸$ $\qquad$ $C = ۰/۲۲$

چون این میانگینهای جدید نمرات مبتنی بر داوری ۵۰نفرند، مقیاس قضاوت درباره تفاوتها کم می‌شود. مقیاس جدید، با انتخاب شانس خطاکردن ۱ به ۲۰، برابر با ۰/۴۵ است.

آشکار است که تفاوت بین مقبولیت دو فراورده H به قدر کافی بزرگ نیست که بتوان گفت یکی مقبولتر از دیگری است. اما در نتیجه‌گیری اینکه دو فراورده H مقبولتر از C، یعنی محصول رقیب، هستند مخاطره بسیار کمی وجود دارد.

پس، از طریق آزمون‌کردن و با کاربرد روش‌شناسی آماری، نشان داده‌شد که این دو فراورده جدید مطبوع و مصداق یک ماده غذایی با پروتئین بالاست. لذا به دو ملاک اولیه در فرایند طولانی معرفی یک فراورده جدید رسیده‌ایم.

مسائل

۱ـ برای ارزیابی پروتئین، چرا موشها را به‌عنوان حیوانات آزمایشی انتخاب کردیم؟

۲ـ الف) ایده بلوکی‌کردن را بیان کنید.

ب) فرض کنید موشها از راه بلوکی‌کردن تخصیص نیافته‌اند، ولی درعوض به راه زیر تخصیص داده شده‌باشند: برای هریک از ۳۰ موش یک تاس شش‌وجهی ریخته شده‌باشد. آمدن ۱ یا ۲ به معنای تخصیص رژیم غذایی کازئین، ۳ یا ۴ به‌معنای تخصیص H مایع، وآمدن ۵ یا ۶ به‌معنای تخصیص H جامد به آن موش است. تحت این الگو چند موش به هر رژیم غذایی تخصیص یافته است؟ آیا ممکن است به همه موشها رژیم کازئین تخصیص یابد؟

۳ـ در آزمایش محاسبه پروتئین، کمترین پروتئین جذب شده برای هرموش در رژیم غذایی کازئین چقدر بود؟ بیشترین چقدر؟ برای رژیم غذایی H مایع و رژیم غذایی H جامد به همین سئوال پاسخ دهید.

۴ـ در آزمایش ارزیابی پروتئین، میانه افزایش وزن ۳۰ موش چقدر بود؟

۵ـ با استفاده از معادله مناسب حاصل از آزمایش ارزیابی پروتئین، اگر موشی ۴۵گرم پروتئین در رژیم غذایی کازئین مصرف کند، چه افزایش وزنی برای آن طی ۲۸ روز پیش‌بینی می‌کنید؟ فرضهایی راکه برای این پیشگویی درنظر می‌گیرید، بیان کنید.

۶ـ فرض کنید شما آزمایش محاسبه پروتئین را طراحی کردید. توضیح دهید که چگونه از جدول

اعداد تصادفی (جدول ۱)، برای دومورد زیر استفاده می‌کنید؟

الف) تخصیص موشها به هر بلوک برای سه رژیم غذایی متفاوت.

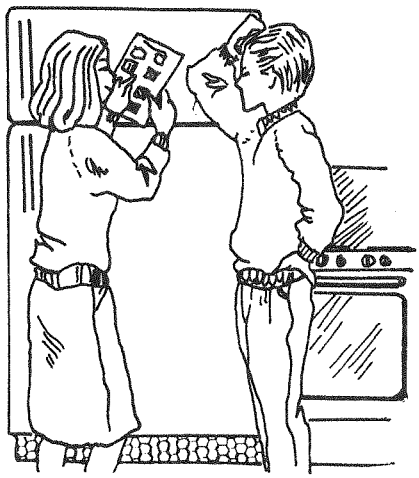
ب) تخصیص موشها به قفسهای بالا یا پایین.

۷ـ الف) برای مقایسه‌کردن میانگین نمرات چشیدن غذا بین مردان و زنان در محاسبه لذیذ بودن، مقیاس حاصل ۰/۶۴ بود؛ اما برای مقایسه میانگین نمرات بین دو فراورده غذایی متفاوت، مقیاس برابر ۰/۴۵ بود. چرا این دو مقیاس متفاوتند؟

ب) فکر می‌کنید، می‌توانید به مقیاسی کوچکتر از ۰/۱ دست یابید؟ اگر جواب شما آری است، چگونه به آن می‌رسید؟

۸ـ برای فراورده‌های B، واریانس نمرات زیر را حساب کنید.

نمره	فراوانی نمره
+۳	۲
+۲	۷
+۱	۷
۰	۷
−۱	۲۰
−۲	۰
−۳	۰
تعداد کل چشنده‌ها	۲۵
میانگین نمرات	۱



تعیین تحملات عددی رنگ با روشهای آماری

لونی سی. ونس[۱] شرکت جنرال موتورز، میشیگان

مقدمه

بعد از ده سال خدمت صادقانه، یخچال یک زوج خراب می‌شود. دیگر وسایل اصلی آشپزخانه آنها، یعنی اجاق خوراکپزی و ماشین ظرفشویی هنوز به‌طور رضایتبخش کار می‌کنند. هر سه این وسایل در یک زمان خریداری شده‌اند و ساخت یک کارخانه بوده و رنگ آنها باهم جور است.

۱_ *Lonnie C. Vance*

با اینکه وسایل این زوج، ۱۰ ساله هستند، هنگامی که درصدد خرید یخچال جدیدی برمی‌آیند، از اینکه پی می‌برند کارخانه سازنده وسایل قدیمی آنها یخچالی می‌سازد که آنها دوست دارند، خوشحال می‌شوند و به نظر دارای رنگی همانند رنگ اصلی است. تنها یک مسئله وجود دارد. رنگی «همانند» برای یخچال جدید دقیقاً با رنگ ماشین ظرفشویی و اجاق خوراکپزی آنها جور نیست. پیش‌روی آنها چندین راه چاره هست که همگی به‌مراتب کمتر از چاره مطلوب هستند. آیا اجاق خوراکپزی و ماشین ظرفشویی خود را نیز عوض کنند تا با رنگ یخچال انتخابی جدید جور باشد؟ آیا یخچال جدید را بخرند و با کمی ناجوری رنگهای وسایلشان بسازند؟ آیا اجاق خوراکپزی و ماشین ظرفشویی کنونی خود را رنگ کنند تا در صورت امکان با رنگ یخچال جدید جور شود؟

مسئله جور کردن رنگها مسئله جدیدی نیست. صنایع بسیاری مانند رنگهای نقاشی ساختمان، رنگرزی، نساجی و خودروسازی مدت زمان درازی است که با مسئله جورکردن رنگها سروکار دارند. به‌عنوان مثال، ارتش می‌خواهد یونیفرمهای سبز زیتونی‌اش کت و شلوارهایی داشته باشد که رنگشان جور بوده و یونیفرمهایی که بتازگی خریداری می‌شود، با قبلیها جور باشد.

نمای سطوح داخلی و خارجی یک خودرو برای خریدار بالقوه آن مهم است. در صورتی که قسمتهای دارای رنگ «همانند» به‌خوبی جور نباشند، مشتری ناراضی می‌شود. برای مثال، درهای جعبه دستکش و جاسیگاری ممکن است قسمتی از جلوی داشبورد باشند. مسلماً هر کسی می‌خواهد قسمتهای مختلف جلوی داشبورد جور باشند ولی این اجزای مختلف با مواد و رنگ «همانند» ممکن است در جاهای مختلف ساخته شده و بعداً در داخل خودرو و در کنار هم سوارشده باشند. چگونه یک تولیدکننده خودرو می‌تواند تا حدی اطمینان یابد که چنین قسمتهایی وقتی که در کنار هم سوار می‌شوند از نظر رنگ جور خواهند شد؟

به‌همین ترتیب، مشتری خواهان جور بودن مناسب رنگ میان سطوح مختلف رنگ شده در خارج خودرو است. بعضی از این سطوح به‌هم وصل شده ممکن است از مواد مختلفی ساخته شده و به‌صورت متفاوتی به عمل آمده‌باشند.

از نظر تاریخی، رنگها به‌صورت ذهنی به‌وسیله مشاهدات بصری افراد جور می‌شده‌اند. اما

به دلیل اهمیت ماهوی جورشدنها، مواد و قطعاتی که باید با هم جور شوند و تمایل ما برای کاهش ذهنی بودن داوریهای بصری به حداقل خود، ابزارهای سنجش رنگ می‌توانند مکمل و کمّی‌کننده ادراک بصری باشند. دانشمندان نشان داده‌اند که افزایش استفاده صحیح از ابزار، بعضی از مشکلات اندازه‌گیری و مقایسه رنگ را حل خواهدکرد.

یک امتیاز اضافی نیز به‌دست می‌آید. اندازه‌گیری رنگ به‌عنوان اساس و مبنایی برای «حافظه رنگ» قرار می‌گیرد. قابل اعتماد نبودن حافظه انسان، تا جایی که به رنگ مربوط می‌شود، به‌خوبی شناخته شده‌است (ن. ک. *بارتلسون* ـ ۱۹۶۰). استفاده صحیح از ابزار، سابقه تولید، تغییرات و حدود قابل پذیرش را مستقل از حافظه مشاهده‌گران در اختیار تولیدکنندگان مواد رنگی قرار می‌دهد.

تحملات رنگ ـ یعنی انحرافات مجاز از یک استاندارد ـ را می‌توان با مقایسه نمونه‌های مناسب با یک رنگ استاندارد تعیین نمود. علاوه بر مثالهایی که قبلاً ذکر شد و در آنها خوب جوربودن رنگها مطلوب بود، تحملات رنگ برای بازرسی مفید ضمن ساخت مواد پلاستیکی که به شکل قطعات سخت پلاستیکی تزیینی درون خودرو درمی‌آیند، مانند قطعات داشبورد و تزیینات داخلی روی درها، ضروری است. در حال حاضر این عمل معمولاً شامل یک بازرسی بصری کالای دریافتی در مقابل یک مرجع استاندارد است که توسط کسانی که مسئولیت تأیید تولیدکنندگان کالا را به‌عهده دارند، تهیه شده‌است. منظور از بازرسی در حال ساخت، معمولاً دو چیز است: یکی تصمیم‌گیری برای رد یاقبول و دیگری مرتب کردن مواد قابل قبول برحسب اختلافات کوچک در رنگ مطابق با مشخصات ازقبل تعیین‌شده. از آنجا که قضاوتهای بصری می‌توانند بسیار ذهنی باشند، تحملات عددی رنگ که بازرسی بصری را تکمیل و کمی می‌کنند مطلوب هستند.

ما توجه خود را براستفاده از مشاهده‌گران (قضاوتهای بصری) و ابزار عددی اندازه‌گیری رنگ در مقایسه با نمونه‌های رنگی قسمتهای تزیینی از پلاستیک سخت (پلی پروپیلن) از نوع استاندارد متمرکز خواهیم نمود.

تعیین نواحی تحمل عددی رنگ با روشهای آماری که برای کار بالا مفید است، شامل دو قسمت می‌شود: آماده‌سازی نمونه‌های رنگ؛ که مبنایی برای تحملات رنگ تشکیل می‌دهد، و

ساخت واقعی تحملات رنگ.

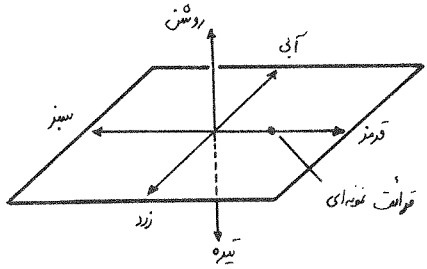
روش و آماده‌سازی نمونه

ما فرض خواهیم کرد که برای هر رنگ و ماده، یک استاندارد مرجع تهیه شده‌است. وقتی که قطعات سخت پلاستیکی تزیینی رنگی تولید می‌شوند باید از نظر رنگ به اندازه کافی با استاندارد مطابقت داشته باشند تا به‌عنوان یک رنگ جور قابل قبول درنظر گرفته شوند. استاندارد مرجع باید از همان موادی ساخته و به همان روشی تولید شده‌باشد که نمونه‌ها ساخته و تولید می‌شوند.

برای هر قطعه نمونه یک قضاوت بصری از رنگ انجام گرفته و یک قرائت اختلاف رنگ عددی از استاندارد به‌دست خواهد آمد. ابزاری که برای به‌دست آوردن قرائت عددی برای پلاستیک مورد استفاده قرار می‌گیرد، نوعی خاص از ابزار اندازه‌گیری رنگ به‌نام اسپکتروفتومتر است. قرائت شامل سه عدداست: به‌ترتیب اندازه‌های روشنی، قرمزی ـ سبزی و زردی ـ کبودی. شکل ۱ نشان می‌دهد که چگونه می‌توان این اعداد را در یک نمودار سه‌بعدی ترسیم کرد. نمونه‌ای که قرائت عددی آن در مبدأ (مرکزشکل) قرار گیرد، دقیقاً دارای همان قرائت عددی استاندارد است. بنابراین یک جورشدن کامل باید بین این دو وجود داشته باشد. نمونه‌ای با یک قرائت عددی رنگ (همان‌طور که نمایش داده شده‌است) که عدد قرمز آن بیش از حد بزرگ و روشنی و زردی – کبودیش جور هستند، قرمزتر از استاندارد است. اندازه این عدد تعیین می‌کند که نمونه در مقایسه با استاندارد قابل قبول است یا نه. به‌عبارت دیگر، اگر نمونه یک قرائت عددی داشته‌باشد که فقط مقدار کمی بیش از حد بزرگ است، به‌صورت بصری می‌تواند قابل قبول باشد.

برای تعیین تحملات عددی رنگ برای یک رنگ جدید (یعنی رنگی بدون هیچ سابقه) باید نمونه‌هایی را آماده کنیم که در مقایسه با استاندارد رنگ‌رفته باشند. دانشمندان نشان داده‌اند که عکس‌العمل مردم نسبت به رنگهای مختلف یکسان نیست. برای مثال، برای رنگهای معینی، یک نمونه می‌تواند عدد قرمزی بزرگتر از استاندارد داشته باشد ولی هنوز به‌صورت بصری قابل قبول باشد. ولی نمونه‌ای دیگر با عدد سبزی به همان بزرگی عدد قرمز می‌تواند به صورت

شکل ۱ـ تفاوت رنگها در سیستم مختصات رنگ یک قرائت نمونه‌ای که پررنگ‌تر از قرمز استاندارد است ولی از نظر عددی تفاوت زیادی در روشنی، زردی یا کبودی با استاندارد ندارد، به نمایش در آمده‌است.



بصری کاملاً غیر قابل قبول باشد. همچنین نمونه‌هایی که قرائت عددی رنگ بزرگی دارند به‌صورت بصری کاملاً غیر قابل قبول هستند. ما در نمونه‌هایمان چند نمونه می‌خواهیم که اعداد رنگ آنها از نظر بصری به‌صورت متفق‌القول توسط مشاهده‌گران قابل قبول با استاندارد باشد. چند نمونه می‌خواهیم که از نظر بصری به صورت متفق‌القول غیرقابل قبول باشند و باز چند نمونه دیگر می‌خواهیم که وقتی مشاهده‌گران آنها را به‌صورت بصری در مقابل استاندارد ارزیابی می‌کنند، عدم توافق قابل ملاحظه‌ای میان آنان وجود داشته باشد. این گروه آخر از نمونه‌های بسیار مهم است. چون به ما اطلاعاتی برای «تنظیم دقیق» مرز بین نواحی قابل قبول و غیر قابل قبول اعداد نمونه می‌دهد. همان‌طور که قبلاً ذکر شد، اندازه اعداد رنگ برای نمونه‌های قابل قبول و غیر قابل قبول بسته به رنگ خاص تحت مطالعه فرق می‌کند. اکنون می‌توان تحت شرایط نوری کنترل شده، نمونه‌ها را به‌صورت بصری در مقابل استاندارد توسط هرکدام ازچند مشاهده‌گر آموزش دیده ارزیابی کرد. کار مشاهده‌گران این است که هر نمونه را از لحاظ اینکه با استاندارد جور است یا نه؛ طبقه‌بندی کنند.

درنتیجه، برای هر نمونه یک سری از قضاوتهای بصری براساس تعداد مشاهده‌گران و تعدادی تکرار مستقل برای هر مشاهده‌گر به‌دست می‌آید. برای به‌دست آوردن یک تصمیم نهایی برای هر نمونه، داده‌های قابل قبول بودن آن یک‌کاسه می‌شوند. برای مثال، اگر اکثریت قضاوتهای مشاهده‌گران قابل قبول باشد، نمونه پذیرفته و در غیر این صورت رد می‌شود.

آخرین مسئله در اندازه‌گیری عددی رنگ ارزیابی تفاوت رنگ برحسب قابل قبول بودن

نمونه در مقایسه با استاندارد است. این‌طور نیست که همه مردم در مورد اینکه اندازه یک اختلاف چه مقدار باید باشد، توافق داشته باشند. در اینجا هم اختلاف فردی در ادراک رنگ و هم سلایق شخصی بدون شک اهمیت پیدا می‌کنند.

بیلمه‌یر و سالتزمن[1] (۱۹۸۱) اظهار می‌دارند که وقتی بناست نظر مشتری نقشی داشته باشد، بهترین شیوه آن است که اندازه‌گیری رنگ طی مدت زمانی به قدر کافی طولانی که سابقه‌ای از آن وجود دارد، انجام گیرد. اگر اصلاً مشتریان در رد یا قبول مواد به روشی سازگار عمل کنند، می‌توان در مورد یک اختلاف رنگ قابل قبول به توافق رسید. حتی اگر این اختلاف با آنچه به‌نظر اختلاف محسوس می‌آید، هیچ رابطه‌ای نداشته باشد.

قاعده تصمیم آماری

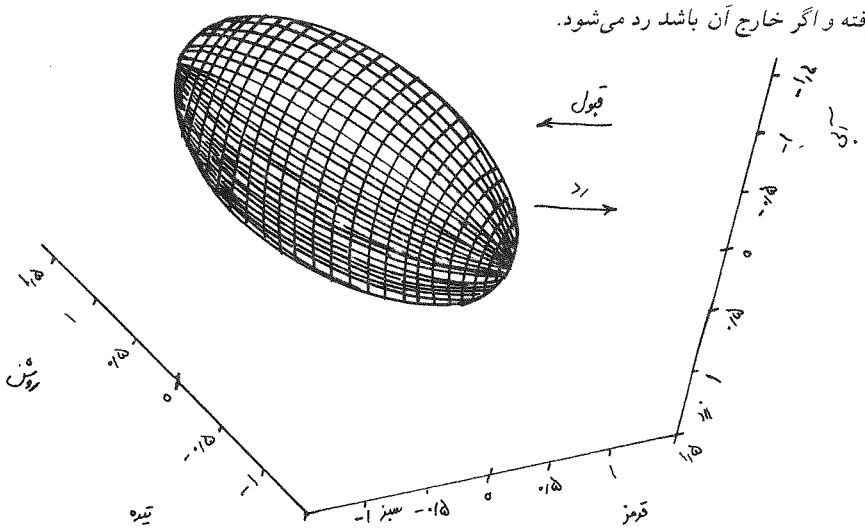
هدف عبارت است از یک قاعده تصمیم آماری عینی که براساس ارزیابی بصری و قرائتهای عددی اندازه‌گیریهای عددی اختلاف رنگ را مشخص خواهد کرد. در این مرحله می‌توانیم قضاوتهای بصری خود برای هر نمونه را به‌عنوان سابقه‌ای در نظر بگیریم که ارزیابیهای بصری را به ارزیابیهای به‌دست آمده با ابزار مرتبط می‌کند. روش مورد استفاده برای ساختن تحملات عددی رنگ آنالیز تمیز نامیده می‌شود. (ن.ک. مقاله *هاولز* برای استفاده از یک تابع تمیز.) این روش قضاوتهای بصری و اندازه‌گیری ابزاری اختلاف رنگ را مرتبط می‌کند و تحملات عددی رنگ حول هر رنگ استاندارد را مشخص می‌سازد (شکل ۲).

به طوری که ذکر شد، سه بُعد از رنگ وجود دارد. مقادیر روی محور عمودی مقدار روشنی از رنگ هستند. دومین بُعد از رنگ اندازه‌ای از قرمزی ـ سبزی و بُعد سوم اندازه‌ای از زردی ـ کبودی است.

در شکل ۲ تحمل عددی رنگ به شکل یک بیضیگون است. قابل قبول بودن بصری انحراف رنگ نمونه از استاندارد، در میان چیزهای دیگر، به بزرگی و جهت سه عدد بستگی دارد که با اسپکتروفتومتر محاسبه می‌شوند. (اگر فضای رنگ در همه جهات یکنواخت بود، آنگاه به‌جای

۱ـ *Billmeyer, Saltzman*

شکل ۲ـ ناحیه تحمل عددی رنگ برای پلاستیک قرمز تیره. اگر نقطه در داخل بیضیگون قرار گیرد، نمونه توسط دستگاه پذیرفته و اگر خارج آن باشد رد می‌شود.



بیضیگون، ناحیه تحمل عددی رنگ یک کره می شد.) برای تمام نمونه‌های بعدی این رنگ، به‌طور ساده یک اندازه‌گیری ابزاری از نمونه گرفته و بررسی می‌کنیم که آیا اعداد مشاهده‌شده به‌اندازه کافی کوچک هستند؟ اگر قرائت عددی به اندازه کافی کوچک باشد که درون بیضیگون قرار گیرد، نمونه پذیرفته‌شده در غیر این صورت رد می‌شود.

قرائت عددی از اسپکتروفتومتر سه عدد به‌دست می‌دهد. فعلاً این سه عدد را به‌عنوان سه اندازه‌گیری مجزا درنظر بگیرید. دلیل منطقی آنالیز تمیز عبارت است از یافتن یک تابع تمیز (یعنی یک قاعده تصمیم آماری) که به بهترین وجهی دو گروه (قابل قبول و غیر قابل قبول) را تمیز دهد.

تابع تمیز سه اندازه‌گیری از نمونه را گرفته، هرکدام از اندازه‌گیریها (یا حاصلضرب اندازه‌گیریها) را در یک وزن بهینه مخصوص آن اندازه‌گیری (یا حاصلضرب اندازه‌گیریها) ضرب کرده و حاصل این ضربها را برای به‌دست آوردن عدد تمیز جمع می‌کند. این محاسبات مجموعه ممکن وزنهای بهینه از اندازه‌گیریهای مورد استفاده را به‌دست می‌دهد.

منظور ما از وزنهای بهینه آن است که تمایز بین مجموعه‌های قابل قبول و غیر قابل قبول بصری به روش ریاضی پیدا می‌شود؛ به نحوی که نمرات تمیز دو گروه در حد بالایی از یکدیگر

جدا هستند.

هرچند پیدا کردن تابع تمیز به مقدار زیادی کار ریاضی نیاز دارد، یک کامپیوتر آن را کار ساده‌ای می‌کند. به محض اینکه وزنها پیدا شد، به‌دست آوردن اعداد تمیز نسبتاً ساده‌است.

مثال

برای نشان دادن روش، ۱۱۱ نمونه پلاک پلاستیکی (پلی پروپیلن) قرمز تیره آماده شد و استاندارد متناظر از همان ماده در دسترس بود. از ۱۶ مشاهده‌گر برای ارزیابی بصری هر نمونه در مقابل استاندارد استفاده شد. از ۱۶ مشاهده‌گر، ۵ نفر سه تکرار درهر نمونه داشتند. درنتیجه هر نمونه درمجموع ۲۶ بار ارزیابی بصری شد. پس برای هر نمونه ۲۶ قضاوت رد یا قبول وجود داشت. ما از معیار قانون اکثریت استفاده کردیم: اگر برای هر نمونه اکثریت ۲۶ قضاوت (۱۳یابیشتر) تصمیم به قبول بود، نمونه به‌عنوان قابل قبول بصری و در غیراین صورت به‌عنوان غیر قابل قبول بصری در نظرگرفته می شد. از ۱۱۱ نمونه، ۴۰ تا قابل قبول بصری و ۷۱ غیرقابل قبول بصری بودند.

مقادیر عددی برای ۱۱۱ نمونه پلی پروپیلن قرمز تیره در جهات و فواصل مختلف از مرکز (استاندارد) پخش شده‌اند. در مرحله بعدی، یک آنالیز تمیز با ۱۱۱ نمونه اجرا کردیم تا یک قاعده تصمیم آماری که معادله‌اش بیضیگونی را تشکیل می‌دهد به‌دست آوریم. برای هر نمونه دیگر از پلی پروپیلن قرمز تیره اگر مقدار قرائت عددی آن در داخل بیضیگون قرارگیرد، نمونه

جدول ۱ـ جدول تصمیم

	روش تصمیم آماری		
	قبول	رد	کل
روش تصمیم بصری			
قبول	۳۵	۵	۴۰
رد	۲۱	۵۰	۷۱
کل	۵۶	۵۵	۱۱۱

پذیرفته و اگر قرائت عددی خارج بیضیگون قرار گیرد، نمونه رد می‌شود.

جدول ۱ شامل خلاصه نتایج برای مشاهده‌گران (تصمیمهای بصری) و روش آماری برای ۱۱۱ نمونه قرمز تیره است. همان‌طور که قبلاً بیان شد، ۴۰ نمونه از نظر بصری قابل قبول و ۷۱ غیر قابل قبول (رد) بودند. قاعده تصمیم آماری نشان می‌دهد که ۵۶ نمونه قابل قبول و ۵۵ نمونه غیر قابل قبول بودند. یعنی عدم توافقی بین آنالیز تمیز و قضاوت بصری در طبقه‌بندی ۲۶ نمونه وجود دارد. این به معنای یک توافق ۷۷ درصداست.

ملاحظات

نگاهی دقیقتر به جدول ۱ نشان می‌دهد که نمونه‌های خیلی کمی که به روش آماری رد شده‌اند ممکن است توسط انسان پذیرفته شود ولی شانسی در حدود ۴۰ درصد وجود دارد که نمونه پذیرفته‌شده توسط روش آماری به‌وسیله انسان غیر قابل قبول تشخیص داده‌شود. بنابراین تکنیک مزبور برای بررسی مقدماتی نیزمناسب خواهدبود وقضاوت نهایی روی‌نصف موادی‌که قبلاً رد نشده‌اند، توسط انسان انجام می‌گیرد. این عمل کار بررسی را به نصف کاهش می‌دهد.

درصد توافق بین قواعد تصمیم آماری و قضاوتهای بصری برای پنج رنگ دیگر پلی‌پروپیلن از ۸۸ ٪ تا ۹۶ ٪ بوده‌است.

هرچند این درصدها به‌طور قابل ملاحظه‌ای بیشتر از درصد توافق برای پلاستیک قرمز تیره است، تعداد نمونه‌های بررسی‌شده برای پنج رنگ دیگر از ۲۳ تا ۴۰ بوده‌که مسلماً خیلی کمتر از ۱۱۱تعداد نمونه‌های قرمزتیره است. درنتیجه این درصدها ممکن‌است کمتر قابل اعتماد باشند.

این فن به‌طور موفقیت‌آمیزی برای ارزیابی بسته‌های ورودی پلی‌پروپیلن رنگی در مقابل یک استاندارد در کارخانه‌های تزیینات خودرو استفاده شده‌است. اگر این قبیل شیوه‌ها به‌صورت سازگاری به‌کار روند، سابقه‌ای نیز از تغییرات و حدود قابل قبول بودن مواد رنگی ایجاد می‌کنند که هم برای فروشندگان و هم برای کارخانه‌ها ارزشمند است.

همان‌طور که قبلاً بیان کرده‌ایم، صنایع بسیاری نیاز به جور کردن رنگها دارند. ابزارهای اندازه‌گیری رنگ می‌توانند بازرسی بصری راتکمیل و کّمی کنند. چنین ابزارهایی اندازه دقیق و جهت اختلاف بین نمونه و استاندارد را مشخص می‌کنند. بعد روش آماری از اندازه‌گیریهای

عددی برای تمیز بین نمونه‌های قابل قبول و غیرقابل قبول استفاده می‌کند و ذهنی بودن قضاوتهای بصری را به حداقل می‌رساند.

در مورد اهمیت جورکردن رنگ مبالغه نمی‌شود. ضعیف جورکردن رنگ را مشتریان به‌عنوان کیفیت پایین تلقی می‌کنند. بنابراین برای جلب رضایت مشتری باید به جورکردن رنگ قابل قبول دست یافت.

مسائل

۱ـ در این مطالعه چندین مشاهده‌گر آموزش‌دیده به‌طریق بصری هر نمونه را در مقابل استاندارد ارزیابی کرده و آن را با عنوان اینکه با استاندارد جور است یا نه دسته‌بندی کردند. آیا لازم است برای اجرای آنالیز تمیز چندین مشاهده بصری از هر نمونه داشته‌باشیم؟ امتیاز داشتن چند مشاهده‌گر چیست؟

۲ـ چرا ابزار اندازه‌گیری رنگ در جورکردن رنگ باارزش است؟

۳ـ در شکل ۲ تحمل عددی رنگ به‌صورت یک بیضیگون است. آیا شرایطی وجود دارد که شکل چیزی غیر از بیضیگون باشد؟ توضیح دهید.

۴ـ مشکل جورکردن رنگها از نمونه‌ای به نمونه دیگر، برای قبول کردن مواد بدون مراجعه به استاندارد اصلی چیست؟

۵ـ هر مشاهده‌گر یک سری قضاوت مستقل در مورد قابل قبول بودن نمونه‌ها برای جوربودن با استاندارد انجام داد. چرا این مهم است؟

۶ـ مثالی با جزئیات از خود بیاورید که در آن یک تحلیل تابع تمیز می‌تواند مفید باشد. (راهنمایی: ن. ک. مقاله دبلیو هاولز)

منابع

T. W. Anderson. 1984. *An Introduction to Multivaiate Statistical Analysis.* 2nd ed. New york: Wiley.

L. C. Vance. 1983. "Statistical Determination of Numerical Color Tolerances. " *Modern Paint and Coatings 73(11): 49-51.*



منصفانه‌تر ساختن نمرات آزمون تشریحی با استفاده از آمار

هنری آی. براون و هاورد واینر[1] خدمات آزمایشی آموزشی

مقدمه

وقتی اوراق تصحیح شده برگشت داده‌شد، طنین صدای اعتراض در کلاس درس پیچید. پس از شوک اولیه، معلم رنج‌کشیده لبخند ملایمی زد و گفت: "عملکرد ضعیف شما، نسبت به کلاسهای قبلی، نشان می‌دهد که این نوع آزمون سخت‌تر از آن بود که من تصور می‌کردم. من

۱_ *Henry I. Braun and Howard Wainer*

مجبورم نمرات را روی منحنی ببرم". رفع نگرانی دانشجویان آشکار شد.

چنین صحنه‌هایی عادی است. "روی منحنی بردن نمرات" عبارت از تبدیل قواعد عادی تناظر بین درصد جوابهای صحیح و نمره حرفی[۱] مربوطه است. در آزمونهای کلاسی، تأثیر روی منحنی بردن تقریباً همیشه اختصاص نمره‌ای بالاتر از نمره مورد انتظار است. درحالی که تقریباً همه این موضوع را می‌دانند ولی این مسئله که چرا معلمان نمرات را روی منحنی می‌برند با هاله‌ای از ابهام روبروست. جواب، با ساده‌ترین بیان این است که نمرات را روی منحنی می‌بریم (تعدیل می‌کنیم)، تا قادر باشیم مقایسه‌های منصفانه‌تری بین افرادی که در شکلهای مختلف آزمون شرکت می‌کنند، انجام دهیم.

مسئله مشابه با متعادل ساختن نمرات آزمون برای منصفانه‌تر ساختن آنها نمره‌گذاریهای آزمونهای تشریحی براساس نظر شخصی است. وقتی مجموعه بزرگی از این نوع آزمونها تصحیح و نمره‌گذاری می‌شوند، عادی است که تعداد زیادی از افراد به تصحیح اوراق بپردازند، درحالی که هریک از آنها نمونه متفاوتی از اوراق را تصحیح خواهند کرد. سختی یک سئوال تشریحی هم مشکل ذاتی سئوال است (مثلاً "فعالیتهای خود را در تعطیلات کریسمس شرح دهید" دربرابر "متافیزیک کانت را با ارسطو مقایسه کنید.") وهم به دقت مصحح بستگی دارد. می‌توانیم مشکل نوع اول را با طرح سئوالهای یکسان برای همه کنترل کنیم، ولی ملاحظات عملی، ما را از کنترل همانند برای همه مصححان باز می‌دارد. با این حال، اگر ملاکهای کار مصححی سختگیرانه‌تر از دیگران باشد، شرکت‌کنندگان به‌حد کافی بدشانسی که ورقه‌هایشان توسط این مصحح تصحیح شوند (مشابه اینکه نوع مشکلتر آزمون به آنهااختصاص یافته است) ضرر می‌کنند. انصاف این است که این تفاوتها (با روی منحنی بودن نمرات مصححان مختلف به‌قسمی که مقایسه‌پذیر باشند، از میان برود.) شرایط مصححان ممکن است با زمان تغییر کند، آنها ممکن است روز دوشنبه[۲] آسانگیرتر از روز جمعه[۳] باشند. اگر چنین تغییرپذیری

۱ـ *در بسیاری از سیستمهای نمره‌گذاری، حرفی به‌جای عددی به‌کار می‌رود. این سیستم قبلاً در برخی از دانشگاههای ایران متداول بوده‌است. به‌طور مثال از ۱۰۰نمره، ۹۰ تا ۱۰۰ الف، ۸۰ تا ۹۰ ب و... (مترجم).*

۲ـ *روز آغازین هفته*

۳ـ *روز آخر کاری هفته*

وجود داشته‌باشد، انصاف حکم می‌کند که این تفاوتهای روزبه روز هم رفع شوند.

آنگاه هدف مطلوب این است که روشی برای نمره‌گذاری آزمونهای تشریحی ارائه شود که نمرات نهایی کمتر تحت‌تأثیر زمان تصحیح و مصحح باشد. تعدیل رتبه‌های داده‌شده اصلی، به‌وسیله مصححان معقول به‌نظر می‌آید. در ادامه این مقاله راه‌حلی را شرح می‌دهیم که به تعدیل آماری متکی است. این راه‌حل در قالب برنامه آزمونی که شامل یک قسمت مهم تشریحی از برنامه پیشرفته تعیین دانشپایه AP است، شرح داده می‌شود.

برنامه پیشرفته تعیین دانشپایه محصل، (AP):[۱]

برنامه AP دوره‌های آموزشی تخصصی در موضوعات مختلف ازجمله انگلیسی، تاریخ آمریکا، تاریخ اروپا، ریاضیات، زیست‌شناسی، شیمی، فرانسه و آلمانی را ارائه می‌دهد. دانش‌آموزان دبیرستانی که در این برنامه شرکت کرده و در آزمونهای (نهایی) آن موفق شوند، شایسته دریافت واحدهای کالج[۲] هستند. در هر موضوعی، آزمون نهایی یکسان در روز شنبه خاصی در ماه مه، در سراسر آمریکا برگزار می‌شود. هر آزمون شامل قسمتی با سئوالهای چند گزینه‌ای (چندجوابی) و یک قسمت از سئوالهای تشریحی است. در سئوالهای تشریحی ریاضیات یا شیمی از دانش‌آموزان خواسته می‌شود تا مسائلی را حل کنند، درحالی که در انگلیسی یا تاریخ آمریکا از آنها می‌خواهند که مقاله بنویسند.

جوابهای این نوع سئوالهای تشریحی را انسان و نه ماشین تصحیح می‌کنند، چون برنامه‌های کامپیوتر هنوز به‌حد کافی باهوش نیستند که دست‌خط دانشجویان را بخوانند و مطالب کیفی را از نظر کمّی ارزش‌گذاری کنند. چون دهها هزار دانش‌آموز درباره یک موضوع خاص مقاله می‌نویسند، فرایند نمره‌گذاری ۴ یا ۵ روز متوالی با شرکت حدود ۱۰۰ نفر مصحح طول می‌کشد. مصححان شامل معلمان دبیرستانی دروس AP و مدرسان آن موضوعها در کالج هستند. هر مقاله (یا مسئله) فقط توسط یک مصحح که به‌تصادف از میان مصححان موجود انتخاب شده، تصحیح می‌شود. او نمره‌ای را تعیین می‌کند که قسمتی ازنمره کل است.

۱ـ *Advanced Placement*

۲ـ دوره دوساله بعداز پایان دبیرستان

مسائل نمره‌گذاری سئوالهای تشریحی

سئوال اینکه یک دانش‌آموز ناعادلانه موقعیتی خوب (یابد) به‌دلیل اینکه برگه‌اش توسط مصحح خاصی تصحیح‌شده، بهره‌مند شده‌است، یک موضوع بحرانی است. تصحیح‌کنندگان چون انسانند، نوع قضاوتشان در مورد کیفیت یک مقاله خاص متفاوت است و بنابراین نمرات اختصاص یافته به جوابهای تشریحی تاحدودی به "شانس انتخاب مصحح" بستگی دارد. این وابستگی به شانس ناخواسته است و باید تا حد امکان ازبین برود.

قبل از اینکه برای حذف این تغییرپذیری بتوانیم کاری انجام دهیم، باید بفهمیم چگونه به‌وجود می‌آیند. اولاً، تصحیح‌کنندگان مختلف ممکن است معیارهای متفاوتی را برای نمره‌گذاری به‌کار گیرند. یعنی اینکه، دو تصحیح‌کننده ممکن است در چگونگی رتبه‌بندی یک مجموعه از اوراق هماهنگ باشند، اما ممکن است یکی از آنها به‌طور سیستماتیک نمره‌هایی بالاتر از دیگری را منظور کند. ثانیاً دو تصحیح‌کننده ممکن است به‌طور متوسط نمره‌های مساوی بدهند، اما به‌طور کلی درباره اینکه کدام جواب تشریحی نمره بیشتری دارد و کدام نمره کمتر، توافق نداشته باشند. در عمل، هر دو نوع اختلاف و برخی اختلافهای دیگر تاحدودی اتفاق می‌افتند.

چون فرایند نمره‌گذاری چند روز به طول می‌کشد، نمره داده‌شده به هر جواب تشریحی ممکن است به زمان تصحیح هم بستگی داشته‌باشد. به‌طور مثال، ممکن است در زمانهایی از هفته روند کلی این باشد که تصحیح و نمره‌گذاری راحت‌تر یا سخت‌تر انجام شود. علاوه بر این روند عمومی، هر مصححی هم ممکن است روندهای خاص خود را در طول زمان داشته باشد. این‌چنین الگوهای عمومی در نمره‌گذاری هیچ ارتباطی با کیفیت جوابهای تشریحی ندارد. اگر چنین الگوهایی وجود داشته‌باشند، آنها هم در نقش شانس برای تخصیص نمره به جوابهای تشریحی سهیم هستند.

در حال حاضر ابزار غیرآماری برای مینیمم ساختن تأثیر بالقوه این عوامل بر نمرات به‌کار گرفته می‌شوند. برنامه AP، مصححان را قبل از آغاز جلسات تصحیح، آموزش می‌دهد و به‌طور پیوسته در طول جلسات بر کار آنها نظارت می‌کند. برای هر موضوع، یک تصحیح‌کننده ارشد با تجربه چندین ساله در اجرای برنامه برای پذیرش مسئولیت حفظ صحت فرایند نمره‌گذاری

منصوب می‌شوند. به‌محض دریافت دفترچه‌های پاسخ اوراق امتحانی، تصحیح‌کننده ارشد تعدادی از جوابهای تشریحی را انتخاب می‌کند تا سطوح مختلف مقیاس نمره‌گذاری را مشخص کند. پس از بحثهای زیاد با تصحیح‌کنندگان باتجربه‌تر و نهایتاً با همه تصحیح‌کنندگان، تصحیح‌کننده ارشد یک فهرست مفصل (تشریحی) از شرایط قضاوت را تنظیم می‌کند. برای تقویت این موضوع، به‌صورت دوره‌ای از همه تصحیح‌کنندگان درخواست می‌شود که به یک ورقه نمره بدهند. اگر اختلافات قابل‌توجه اتفاق بیفتد، تصحیح‌کنندگان تحت آموزش بیشتر قرار می‌گیرند. به‌نظر می‌آید که این روش به‌طور منطقی عمل می‌کند، اما همان‌طور که خواهیم دید بازهم جای اصلاح وجود دارد.

قبل از اینکه به بحث چگونگی کمک بینش آماری بپردازیم، باید در اندازه‌گیری میزان خوبی یک روش پیشنهادی، راهی برای کم‌کردن نقش شانس در نمره‌گذاری پیدا کنیم. این راه مقیاسی برای قضاوت درباره موثربودن روش جدید مهیا می‌سازد.

به چه اندازه خوب عمل می‌کنیم؟

متأسفانه هیچ راه ساده‌ای برای به‌دست آوردن "نمره واقعی" یک سئوال تشریحی وجود ندارد، لذا نمی‌توانیم نمره‌ای که به ورقه داده شده را با نمره واقعی مقایسه کنیم و تفاضل آنها را به‌عنوان شاخص تأثیر شانس به‌کار ببریم. اگر یک سئوال تشریحی به‌وسیله همه تصحیح‌کنندگان، تصحیح شود، در این‌صورت معدل همه این نمرات را می‌توان به‌جای نمره واقعی به‌کار برد. البته این غیرعملی است مگر اینکه تعداد زیادی تصحیح‌کننده برای تعداد اندکی سئوال تشریحی انتخاب شوند.

دانشمندانی که نمرات آزمونها را بررسی می‌کنند، استراتژی تقریباً متفاوتی را درنظر می‌گیرند. آنها برای قضاوت درباره اعتبار یک فرایند نمره‌گذاری، آنرا برای تصحیح دوباره یک نمونه بزرگ از سؤالات تشریحی و ارزیابی هماهنگی بین دو مجموعه نتایج، به‌کار می‌برند. در مورد AP، ممکن است نمونه‌ای ۵۰۰ تایی از سئوالهای تشریحی را برای آزمایش انتخاب کنند. آنگاه هر سئوال دوبار تصحیح می‌شود. هربار در یک روز به‌وسیله یک تصحیح‌کننده که به تصادف انتخاب می‌شود با استفاده از اولین مجموعه نمرات، ورقه‌ها از بالا به پایین (برحسب

نمره) فهرست می‌شوند. دومین رتبه‌بندی به‌وسیله دومین مجموعه نمرات به‌دست می‌آید. اگر نقش شانس نسبتاً اندک باشد، یک ورقه باید تقریباً در یک‌جا در دو فهرست قرار گیرد. اما اگر شانس سهم عمده‌ای داشته‌باشد، دو رتبه‌بندی تفاوت قابل توجهی خواهند داشت.

سطح هماهنگی بین رتبه‌بندیها معمولاً به‌وسیله کمیتی به نام ضریب *قابلیت اعتماد* اندازه‌گیری می‌شود. ضریب قابلیت اعتماد عددی است که از اطلاعات عددی موجود در دوفهرست محاسبه می‌شود. با این وضع، مقدار آن از نزدیک صفر تا نزدیک ۱ تغییر می‌کند. اگر هماهنگی اندکی بین دو فهرست باشد، این ضریب مقداری نزدیک صفر می‌گیرد که نشان می‌دهد شانس نقش مؤثر در فرایند نمره‌گذاری دارد. از طرف دیگر، اگر هماهنگی قابل توجهی بین دو فهرست باشد، این ضریب مقداری نزدیک ۱ می‌شود که نشان می‌دهد شانس نقش جزیی دارد.

مقادیری نوعی از ضریب قابلیت اعتماد برای نمره‌های سئوالهای تشریحی در علوم انسانی بین ۰/۳ تا ۰/۶ است. برای مسائل شیمی، این ضریب قابلیت اعتماد معمولاً بین ۰/۶ تا ۰/۸ قرار دارد. برای اینکه احساسی از مفهوم این اعداد به‌دست آورید، این یافته‌ها را در نظر بگیرید: اگر گروهی از پسران براساس قدشان درسن شش سالگی و مجدداً در ده سالگی رتبه‌بندی شوند، ضریب قابلیت اعتماد برای این دو فهرست معمولاً بیشتر از ۰/۸ می‌شود. اگر پسران براساس عملکردشان در یک آزمون هوش که به‌صورت عینی نمره‌گذاری می‌شود در دو سن مختلف رتبه‌بندی شوند، ضریب قابلیت اعتماد معمولاً بیشتر از ۰/۶ می‌شود. سرانجام، اگر پسران همسن، یک‌بار براساس قد و یک‌بار براساس عملکردشان در یک آزمون هوش رتبه‌بندی شوند، ضریب قابلیت اعتماد برای این دو فهرست نسبتاً غیرمربوط هستند، معمولاً فقط حدود ۰/۲ یا ۰/۳ است.

غیرعادی نیست که (همانند مطالعه‌ای که در زیر شرح داده می‌شود) بیشتر از ۲ فهرست رتبه‌بندی شده از نمرات برای مجموعه‌ای از سئوالهای تشریحی داشته‌باشیم. در چنین حالتهایی این تعداد زیاد رتبه‌بندی را به ۲ فهرست با انتخاب ۲ تا از آنها به‌تصادف تقلیل داده و ضریب قابلیت اعتماد آنها را همانند سابق محاسبه می‌کنیم. از این به‌بعد، وقتی درمورد قابلیت اعتماد یک شیوه نمره‌گذاری صحبت می‌کنیم، مقصودمان اندازه‌ای است که خیلی شبیه متوسط

قابلیت اعتماد دو به دو برای تمام جفتهای متمایز از قضاوتهاست.

واسنجی مصححان و روزها

اکنون آماده‌ایم که نحوه کمک فکر آماری را در این زمینه ببینیم. ایده اصلی کم‌کردن اثر برخی از منابع تغییرات که به آنها اشاره کرده‌ایم بر قابلیت اعتماد نمره‌گذاری است: یعنی تفاوتهای سیستماتیک بین تصحیح‌کنندگان یا بین روزها. از این مسئله مفهوم زیر را درنظر داریم: اگر می‌دانستیم که برای یک دسته از اوراق، یک مصحح به‌طور متوسط ۱۰ نمره بیشتر از مصحح دیگر نمره می‌دهد، می‌توانستیم با کم‌کردن ۱۰ نمره از هریک از نمرات (یا با اضافه کردن ۱۰ نمره به هریک از نمرات مجموعه دوم)، آنها را تعدیل کنیم. با این عمل، دو مجموعه از نمرات میانگین یکسانی خواهندداشت. این همان چیزی است که انتظار می‌رود، چون آنهامربوط به همان دسته از اوراق هستند.

دقیقاً همان نوع تعدیل می‌تواند برای اختلافهای سیستماتیک بین روزها هم به‌کار رود. اگر یک دسته از اوراق که در یک روز تصحیح می‌شوند، مثلاً به‌طور متوسط ۵ نمره کمتر از روز دیگر بگیرند، باید ۵ نمره به اولین مجموعه نمرات اضافه کنیم تا میانگینها را برابر سازیم. فرایند مساوی کردن میانگینها واسنجی نامیده می‌شود. در بحث نمره‌گذاری سئوالهای تشریحی، واسنجی تصحیح‌کنندگان و روزها، قابلیت اعتماد نمرات را به‌وسیله حذف دو منبع تغییرات شانسی بهبود می‌بخشد. درجه بهبود، علاوه بر عوامل دیگر، به میزان بزرگی اندازه این اختلافات در مرحله اول بستگی دارد.

جمع آوری داده‌ها

از کجا اطلاعات مورد نیاز برای واسنجی را به‌دست می‌آوریم؟ در تصحیح عملی هر ورقه فقط یک‌بار توسط یک مصحح خاص و در یک روز خاص تصحیح می‌شود. اگر مصححان معدلهای مختلفی در طول دوره پنج روزه تصحیح اوراق بدهند، ما نمی‌دانیم که این تفاوتها به تفاوتهای واقعی و سازگار بین تصحیح‌کنندگان مربوط است یا به اینکه تفاوتهای کیفیت جوابهای تشریحی که اتفاقاً توسط آنها تصحیح شده بستگی دارد و یا به هر دوی اینها. برای اندکی پیشروی

مجبوریم داده‌های خاصی را که اطلاعات مورد نیاز رابه ما بدهند جمع‌آوری کنیم.

نظریه آمار می‌تواند ما را به طرح آزمایشی راهنمایی کند که آن داده‌ها رابه‌طور مؤثر جمع‌آوری کنیم و به‌طور مناسب به‌کار گیریم. آزمایش زیر را درنظر بگیرید: فرض کنید نمونه‌ای کوچک از میان دهها هزار ورقه موجود را انتخاب کنیم و ترتیبی دهیم که هر ورقه توسط همه تصحیح‌کنندگان در همه روزها تصحیح شود. داده‌هایی که بدین‌وسیله تولید می‌شوند به‌ما امکان می‌دهند که متوسط تفاوتها بین مصححان ونیز بین روزها را برآورد کنیم (ما اصطلاح برآورد را به‌کار می‌بریم زیرا رفتار نمره‌گذاری مصححان را فقط برای نمونه به‌دست آورده‌ایم نه برای همه ورقه‌ها).

می‌توانیم این برآوردها را که از یک نمونه کوچک اوراق به‌دست آمده برای واسنجی تصحیح‌کنندگان و روزها به‌کار بریم. به این مفهوم که می‌توانیم نمرات همه اوراق را با درنظر گرفتن تصحیح‌کننده و روز تصحیح، براساس اطلاعات جمع‌آوری شده از این آزمایش تعدیل کنیم. اما قبل از انجام این کار باید کیفیت اطلاعاتی را که به‌کار می‌گیریم، به‌دقت مورد توجه قرار دهیم.

این آزمایش حداقل با دو مسئله روبه‌روست. به‌دلیل تعداد بسیار زیاد تصحیحهایی که باید انجام شود، محدودیت جدی برای تصحیحهای اضافی که برای انجام آزمایش لازم است وجود دارد. چون هر تصحیح‌کننده باید هر ورقه را هر روز تصحیح کند، تعداد ورقه‌ها باید خیلی‌کم باشد. مثلاً ۵ تا ۱۰ عدد. این تعداد کم سؤالاتی را درباره نماینده بودن نتایج مطرح می‌کند: آیا ما برآوردهای اساساً متفاوتی را به‌دست می‌آوریم، اگر ۵ ورقه دیگر را انتخاب می‌کردیم؟ دومین موضوع قابل طرح پیرامون تصحیح تکراری ورقه‌هاست. به همان میزان که مصحح نمره‌ای را که روز قبل به ورقه‌ای داده به‌خاطر داشته و فقط آنرا کپی کند. اطلاعات جمع‌آوری شده دقیق نیست. این‌چنین پیچیدگی در برآوردها ممکن است جهت غلطی به تعدیلهای ما بدهد، به‌طوری که واسنجی‌ها به‌جای افزایش قابلیت اعتماد، آنرا کاهش دهند.

آماربرای نجات

هدف ما برآورد دقت نسبی بین تصحیح‌کنندگان مختلف و همچنین روندهای انجام تعدیلها در

طول زمان بدون افتادن در دامهای ذکر شده‌است. خوشبختانه آمارشناسان سعی فراوان برای حل این نوع مسائل کرده‌اند. آنها روش خاصی را برای کارایی در نحوه جمع‌آوری داده‌ها ابداع

جدول ۱ـ طرح تخصیص ورقه‌ها به مصححان

	مصححان											
ورقه‌ها	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲
۱	۱	۱	۱	۴	۴	۴	۳	۳	۳	۲	۲	۲
۲	۳	۳	۴	۲	۲	۳	۱	۱	۲	۴	۴	۱
۳	۴	۲	۳	۳	۱	۲	۲	۴	۱	۱	۳	۴
۴	۲	۴	۲	۱	۳	۱	۴	۲	۴	۳	۱	۳
۵	۱	۴	۴	۴	۳	۳	۳	۲	۲	۲	۱	۱
۶	۳	۲	۱	۲	۱	۴	۱	۴	۳	۴	۳	۲
۷	۴	۳	۲	۳	۲	۱	۲	۱	۴	۱	۴	۳
۸	۲	۱	۳	۱	۴	۲	۴	۳	۱	۳	۲	۴
۹	۱	۲	۲	۴	۱	۱	۳	۴	۴	۲	۳	۳
۱۰	۲	۳	۱	۱	۲	۴	۴	۱	۳	۳	۴	۲
۱۱	۳	۴	۳	۲	۳	۲	۱	۲	۱	۴	۱	۴
۱۲	۴	۱	۴	۳	۴	۳	۲	۳	۲	۱	۲	۱
۱۳	۱	۳	۳	۴	۲	۲	۳	۱	۱	۲	۴	۴
۱۴	۳	۱	۲	۲	۴	۱	۱	۳	۴	۴	۲	۳
۱۵	۴	۴	۱	۳	۳	۴	۲	۲	۳	۱	۱	۲
۱۶	۲	۲	۴	۱	۱	۳	۴	۴	۲	۳	۳	۱
۱۷	۱	۱	۱	۴	۴	۴	۳	۳	۳	۲	۲	۲
۱۸	۳	۳	۴	۲	۲	۳	۱	۱	۲	۴	۴	۱
۱۹	۴	۲	۳	۳	۱	۲	۲	۴	۱	۱	۳	۴
۲۰	۲	۴	۲	۱	۳	۱	۴	۲	۴	۳	۱	۳
۲۱	۱	۴	۴	۴	۳	۳	۳	۲	۲	۲	۱	۱
۲۲	۳	۲	۱	۲	۱	۴	۱	۴	۳	۴	۳	۲
۲۳	۴	۳	۲	۳	۲	۱	۲	۱	۴	۱	۴	۳
۲۴	۲	۱	۳	۱	۴	۲	۴	۳	۱	۳	۲	۴
۲۵	۱	۲	۲	۴	۱	۱	۳	۴	۴	۲	۳	۳
۲۶	۲	۳	۱	۱	۲	۴	۴	۱	۳	۳	۴	۲
۲۷	۳	۴	۳	۲	۳	۲	۱	۲	۱	۴	۱	۴
۲۸	۴	۱	۴	۳	۴	۳	۲	۳	۲	۱	۲	۱
۲۹	۱	۳	۳	۴	۲	۲	۳	۱	۱	۲	۴	۴
۳۰	۳	۱	۲	۲	۴	۱	۱	۳	۴	۴	۲	۳
۳۱	۴	۴	۱	۳	۳	۴	۲	۲	۳	۱	۱	۲
۳۲	۲	۲	۴	۱	۱	۳	۴	۴	۲	۳	۳	۱

** ردیفهای ۱۷ تا ۳۲ تکرار ردیفهای ۱ تا ۱۶ هستند*

کرده‌اند که طرح آزمایشی نام دارد. مثالی از یک طرح که نیازهای ما را تأمین می‌کند در جدول ۱ آمده‌است. جدول مجموعه‌ای از دستورالعملها برای تعیین مصححان در آزمایش ۴روزه‌ای را که شامل ۱۲ مصحح و ۳۲ ورقه است (که به‌تصادف انتخاب شده‌اند) مشخص می‌کند. (یکی از دلایلی که اعداد ۱۲ و ۳۲ انتخاب شده‌اند آن است که هردو بر ۴ «طول این آزمایش خاص» قابل قسمت هستند. سایر ترکیبات نیزامکانپذیر است.)

هریک از ۳۲ ردیف به ورقه‌ای اختصاص دارد، درحالی که هریک از ۱۲ ستون به یک مصحح مربوط است. اعداد در هر سطر جدول نشان می‌دهد که کدام مصحح در چه روزی آن ورقه را تصحیح کند. به‌طور مثال مصحح شماره ۱ مقاله شانزدهم را در روز دوم تصحیح می‌کند.

این طرح طوری تنظیم شده که هر ۳۲ ورقه ۳ بار در روز توسط ۹۶ مصحح تصحیح می‌شود (دقت کنید اگر هرمصحح هر ورقه را هر روز تصحیح کند، با محدودیت کلی ۹۶ نفر مصحح، فقط ۸ورقه در آزمایش می‌تواند منظور گردد. با ضعیف کردن این شرط، قادریم ۴ برابر ورقه‌ها را در آزمایش به‌کارگیریم.) اختصاص مصححان به ورقه‌ها به طریق تصادفی نیست. درواقع، یک انتخاب ظریف از ترکیبات مصححان و اوراق است که به‌ما توان برآورد اختلافهای سیستماتیک بین مصححان را می‌دهد، اگرچه هیچ دو مصححی مجموعه‌ای یکسان از ورقه‌ها را نمی‌خوانند.

در طول مدت این آزمایش ۴ روزه، هر مصحح فقط یک‌بار هریک از ۳۲ ورقه را تصحیح می‌کند. درنتیجه، هیچ اثر تصحیح تکراری و انتقالی ما را نگران نمی‌کند. چون هر ورقه ۳بار در هر روز تصحیح می‌شود و هر مصحح ۲ ورقه را در روز تصحیح می‌کند، می‌توانیم برآوردهایی نیز از تفاوتهای سیستماتیک بین روزها داشته باشیم. چون برآوردهای ما به‌جای ۸تایی که با محدودیت طرح کامل شامل ۹۶ بار تصحیح بود، براساس نمونه‌ای ۳۲تایی است. این برآوردها بیشتر نماینده واقعیت هستند. با این طرح خاص که انتخاب کردیم، حتی ممکن است مقایسه‌های مفیدی بین مصححان براساس تصحیح روزبه‌روز هم انجام شود.

برخی نتایج

برای پی بردن به لطف این نتایج، یافته‌های یکی از این نوع آزمایشها را که برای سئوالهای تشریحی ادبیات انگلیسی و انشا اجرا شده‌است، و حدود نمرات آنها در مقیاس ۱۰۰ (پایین) و

جدول ۲ـ معدلهای روزانه و انحرافات آنها از میانگین

روز	معدل روزانه	معدل روزانه منهای معدل آزمایش
۱	۴۹۰	۷
۲	۴۷۹	-۴
۳	۴۷۸	-۵
۴	۴۸۵	۲
معدل آزمایش	۴۸۳	

۹۰۰ (بالا) است معرفی می‌کنیم. جدول ۲، برای هر روز، میانگین نمرات اوراق تشریحی تصحیح‌شده و نیز تفاضل بین این میانگینهای روزانه و میانگین کل را برای تمام آزمایش نشان می‌دهد. در روز اول، به‌طور مثال، میانگین نمرات ۴۹۰ بود که ۷ نمره، بیشتر از میانگین کل ۴۸۳، بود.

به‌طور ایده‌آل انتظار می‌رود که میانگینهای روزانه خیلی شبیه هم باشند و درحقیقت در این مورد چنین است. (بزرگترین تفاوت بین روزها ۱۲ نمره است. این کمتر از ۳درصد نمره متوسط آزمایش می‌باشد.) اما این بدان معنی است که تلاش برای تعدیل تفاوتهای سیستماتیک بین روزها، با موفقیت خیلی کمی همراه است که واقعاً چیزی نیست!

از طرف دیگر، جدول ۳ میانگین نمرات داده شده توسط هر مصحح را در طول آزمایش نمایش می‌دهد. برای مشاهده دقیق اختلافهای اساسی، تفاوتهای بین معدلهای این مصححان و معدل کل ۴۸۳ را هم نشان می‌دهیم. مصحح A آسانگیرترین مصحح، نوعاً ورقه‌ها را ۸۲ نمره بیشتر از متوسط نمره می‌دهد، درحالی که مصحح L، سخت‌گیرترین مصحح، ۵۸ نمره کمتر از متوسط می‌دهد. به خاطر داشته باشید که اینها فقط برآوردهای تفاوتها در سطح نمره‌گذاری بین مصححان براساس ۳۲ تصحیح است. به‌هر حال آنها اعتبار قابل ملاحظه‌ای دارند. چون براساس طرح توانسته‌ایم توازن لازم را برای حذف منابع تغییری که در صورت وجود برآوردهای مارا بی‌ارزش می‌ساخت، برقرار کنیم.

مطمئناً حداقل برای این سئوال به‌نظر می‌آید که روزها خیلی تأثیر ندارند ولی مصححان

اهمیت دارند. بنابراین، باید به‌خاطر داشته باشیم که ۳ بار نمرات در معدل روزانه و در معدل مصحح سهیمند، بنابراین نسبتی از تغییرات بیشتری که بین معدلهای مصححان مشاهده می‌شود (در مقایسه با معدلهای روزانه) ممکن است به‌دلیل مقوله‌های شانسی باشند. به‌هرحال، ما می‌توانیم در ارتباط با نتایج این طرح خاص و روشهای فرض آزمون آماری سرمایه‌گذاری کنیم تا مقایسه کاملی بین تغییرات نسبی مصححان علیه روزها به‌دست آوریم. وقتی این کار را انجام دهیم، آنگاه متوجه می‌شویم که نخستین گمان طبیعی ما توجیه‌پذیر شده‌است. مصححان خیلی بیشتر از روزها اهمیت دارند.

پس برای اجرای واسنجی، ۸۲ نمره از نمرات تمام اوراقی که توسط A تصحیح شده، ۶۱ نمره از نمرات تمام اوراقی که توسط B تصحیح شده و نظایر اینها کم می‌کنیم. با مقایسه قابلیت اعتماد نمرات اصلی با نمرات تعدیل‌یافته می‌توانیم درمورد تأثیر شیوه قضاوت کنیم. قابلیت اعتماد نمرات اصلی ۰/۵۷ و قابلیت اعتماد نمرات تعدیل‌یافته ۰/۶۱ با اختلاف ۰/۰۴ است. به‌نظر نمی‌آید که این همه تلاش بهبود زیادی را نتیجه داده باشد. محاسبات زیر ممکن است کمک کند تا چشم‌اندازی از نتایج مسئله را ارائه دهد.

بااستفاده از یک تحلیل ریاضی، ممکن است نشان دهیم که اگر هر مقاله به‌طور مستقل توسط دو مصحح تصحیح شده و متوسط این دونمره به‌عنوان نمره نهایی به‌کار گرفته شود، آنگاه قابلیت اعتماد این نمرات معدل حدود ۰/۷۳ می‌شود. (تصحیحهای چندگانه راه عادی برای بهبود قابلیت اعتماد است) حاصل کار ما (۰/۰۴)، بیست وپنج درصد ۰/۱۶=۰/۵۷-۰/۷۳، است که افزایش قابلیت اعتماد ممکن با دوبار تصحیح است.

به‌خاطر داشته باشید که از اطلاعات به‌دست آمده از این آزمایش کوچک، می‌توانیم نمرات تمام مجموعه اوراق را تعدیل کنیم. داده‌های ما با پرداخت نسبت کمی از هزینه به‌کارگیری مصححان بیشتر برای تصحیح دوباره دهها هزار ورقه موجود، به‌دست آمده‌است. ضریب هزینه را نوعاً حدود $\frac{1}{30}$ برآورد می‌کنیم. چون ما یک‌چهارم بهره را با $\frac{1}{30}$ هزینه به‌دست آورده‌ایم، تحلیل هزینه به بهره ضریبی حدود ۷ یا ۸ را به نفع روش واسنجی به‌دست می‌دهد. این مطلب به مفهوم این است که وقتی ۵۰۰۰ دلار برای انجام این آزمایش هزینه شود، حدود ۱۵۰۰۰۰ دلار هزینه دارد؛ اگر می‌خواستیم مصححان کافی برای یک دوباره‌خوانی به‌کارگیریم و $\frac{1}{4}$ چنین

جدول ۳ـ معدلهای مصححان وانحرافات آنها از میانگین

مصحح	معدل مصحح	معدل مصحح منهای میانگین آزمایش
A	۵۶۵	۸۲
B	۵۴۴	۶۱
C	۵۱۷	۳۴
D	۵۰۶	۲۳
E	۴۸۷	۴
F	۴۸۴	۱
G	۴۷۶	-۷
H	۴۷۳	-۱۰
I	۴۵۴	-۲۹
J	۴۳۲	-۵۱
K	۴۳۲	-۵۱
L	۴۲۵	-۵۸
میانگین آزمایش	۴۸۳	

برای راحتی درک این جدول، مصححان به‌ترتیب معدل نمراتی که می‌دهند، رتبه‌بندی شده‌اند.

عددی (۳۷۵۰۰دلار) برای به‌دست آوردن همان بهره در قابلیت اعتماد لازم بود. این مطلب توصیه می‌کند که روش واسنجی باید به‌طور جدی منظور شود.

آیا واسنجی باید به‌کار رود؟

آزمایشهای واسنجی، درحال حاضر در ۵ آزمون مختلف AP به‌کار می‌رود. به‌طور کلی، نمرات واسنجی شده قابلیت اعتماد کاملی را نمایش می‌دهند، مخصوصاً وقتی که قابلیت اعتماد نمرات اصلی در آغاز سطح بسیار پایینی باشد. در یک حالت، برآورد قابلیت اعتماد نمرات واسنجی شده واقعاً بیشتر از قابلیت اعتماد طراحی شده برای بازخوانی است! به هرحال موفقیت آشکار چنین آزمایشی برای تضمین پیاده‌سازی عملیاتی فرایند کافی نیست. مطالب بسیار زیاد دیگری نیز هستند که باید مورد توجه قرار گیرند.

یکی از این مطالب را، به‌دلیل اینکه آزمایش بیان‌شده مستلزم طراحی و تحلیل قابل توجهی است، مطرح کرده‌ایم. ما یک شیوه واسنجی دیگر را نیز مطالعه کردیم که برای موارد

محدودیت برنامه زمانی در ارائه نتایج به داوطلبان بهتر، می‌تواند هماهنگ شود. این شیوه همان است که قبلاً ذکر کردیم، یعنی تعدیلها را براساس نمرات عملی اعمال کنیم.

برای مثال، اگر میانگین نمرات یک مصحح درطول دوره تصحیح ۱۰نمره از معدل نمرات تمام مصححان بیشتر بود، ۱۰نمره از نمرات تمام ورقه‌هایی که او تصحیح کرده کم می‌کنیم. مشکل بالقوه چنین روشی این است که اوراق این مصحح ممکن است به‌صورت شانسی، در متوسط بهتر از معدل بوده و حق دریافت بهترین نمرات را داشته باشند که در اینصورت تعدیل آنها با ۱۰نمره خیلی زیاد خواهد بود.

در عمل، دفترچه‌های جواب به سئوالات، مراحل مختلف بُرزدنهایی تصادفی را می‌گذرانند تا در مقابل مصححی قرار گیرند. متأسفانه، هیچ راه مستقیمی برای تشخیص اینکه یک مصحح اوراقی را دریافت می‌کند که نوعاً نمونه‌های کاملاً تصادفی و نماینده هستند یا نه نداریم. اما این درست جایی است که آزمایش ما نقش مهمی را بازی می‌کند. می‌توانیم واسنجی را با استفاده از نمرات عملی و نمرات واسنجی را با استفاده از این آزمایش باهم مقایسه کنیم (درحالتی که نمونه اوراق کنترل‌شده و به‌طور کاملاً تصادفی انتخاب شده‌باشند) وقتی این کار را انجام دهیم، درمی‌یابیم که نتایج خیلی شبیه هم هستند. این مسئله اعتماد به روش ساده‌تر و ارزانتر را تأیید می‌کند.

ذکر این نکته هم ارزش دارد که داده‌های جمع‌آوری شده در آزمایشی شبیه آنچه شرح دادیم، به آگاهیهایی منجر می‌شود که از داده‌های عملی به‌دست نمی‌آیند. با استفاده از روشهای تحلیل که خیلی فنی‌تر از آن هستند که در این بحث مطرح شوند، می‌توانیم مطالبی بیشتر درباره آثار نسبی مصححان و روزها در غیرقابل اعتماد بودن نمرات دریابیم و آنرا به روشی انجام دهیم که مقایسه‌ها را برای آزمونهای مختلف آسان سازد. همچنین می‌توانیم حد بالای قابلیت اعتماد که ازطریق واسنجی میسر باشد را برآورد کنیم که این هدف بامعنایی را برای رسیدن به آن ارائه می‌دهد.

علاوه بر ملاحظات عملی بودن کار باید عکس‌العملهای ممکن دانش‌آموزان و مدارس را درباره مفهوم تعدیل آماری نمرات نیز به‌حساب آوریم. چون اولین مرحله این تحقیق به روشنی نشان می‌دهد آزمایشی که به‌صورت آماری طرح‌ریزی شده‌است، می‌تواند فرایند نمره‌گذاری

را منصفانه‌تر سازد. باید سایر جنبه‌ها را هم قبل از به‌کارگیری آن در سطح وسیع هموار ساخت. در حال حاضر (دسامبر ۱۹۸۷) فرایند تصمیم‌گیری آن در دست اجراست و ممکن است زمانی که آن را مطالعه می‌کنید، به‌کارگرفته شده باشد.

مسائل

۱ـ آیا آموزش مصححان منجر به نمره‌دادن یکسان ورقه‌ها می‌شود؟ چرا آری یا چرانه ؟

۲ـ آیا انتظار دارید در طول هفته مصححان مقیاس نمره‌دادن خود را تغییر دهند؟

۳ـ چرا واسنجی ورقه‌های تشریحی ضروری است؟

۴ـ چرا نمی‌توانیم همه ورقه‌ها را به‌وسیله چندین تصحیح‌کننده مختلف تصحیح کنیم؟

۵ـ فایده استفاده از طرح آزمایش پیچیده در جدول ۱ به‌جای اینکه تمام ورقه‌های تحت آزمایش به‌وسیله هریک از مصححان در هرروز تصحیح شود چیست؟

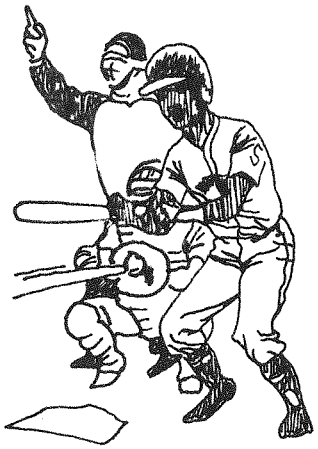
۶ـ چه مقدار دقت از تعدیل تفاوتهای بین عملکرد مصححان به‌دست می‌آید؟

منابع

H. I. Braun. 1988, "Understanding Score Reliability: Experiments in Calibrating Easy Readers," *Journal of Educational Statistics* 13: 1-18.

W. G. Cochran and G. M. Cox. 1957, *Experimental Designs,* 2nd ed. New York: Wiley, Chapter 9. A Classical work on experimental design.

H. O. Gulliksen. 1987, *Theory of Mental Tests.* Hillsdale N.J. Erlbaum. (Orginally Published in 1950 by John Wiley & Sons) This was the First Comprehensive account of mental test theory-see especially pages 211-214for description of how to grade essays.



آمار، ورزش و چند چیز دیگر

رابرت هوک[۱] آزمایشگاههای تحقیقاتی وستینگ‌هاوس

ورزشهایی که تماشاگر دارند صرفاً در تماشای ورزشکارانی که با مهارتی قابل ستایش اجرای نقش می‌کنند، خلاصه نمی‌شود. به‌عنوان مثال، نمایش یک جوان پرتاب‌کننده[۲] توپ را که سعی می‌کند یک تیم فوتبال آمریکایی حرفه‌ای را برای اولین مرتبه در مقابل ۷۰۰۰۰ بیننده

۱ـ *Robert Hooke*

۲ـ *Quarterback*

رهبری کند و یا یک توپ‌انداز[۱] کهنه‌کار که می‌خواهد از تجربه‌اش برای جبران قوای جسمانی تحلیل‌رفته خود در یک بازی حساس مربوط به مسابقه‌ای قهرمانی استفاده کند، درنظر بگیرید. از آنجاکه این نمایشها واقعاً زنده اجرا شده و غیرقابل پیش‌بینی هستند، برای بعضی از افراد نسبت به نمایشهای صحنه‌ای که ازقبل تمرین شده‌اند، از جذابیت خیلی بیشتری برخوردارند.

البته در ورزش، تمام لحظات هیجان‌آور نیستند، اما در سراسر مبارزه بین حرفه‌ایها، تماشاگر از این شانس برخوردار است که به تماشای گروهی از افراد بپردازد که تقریباً فعالیتهای کاری خود را در معرض دید عموم قرار می‌دهند، و می‌بیند که این افراد چطور با مشکلات خود برخورد می‌کنند و چطور در مقابل موفقیتها و شکستها واکنش نشان می‌دهند. بیس‌بال و فوتبال آمریکایی بخصوص فرصتهای خوبی را برای چنین تماشایی فراهم می‌آورند؛ زیرا هرکدام از این بازیها، برخلاف عملکرد نسبتاً پیوسته بسکتبال، هاکی، فوتبال، و مسابقات اتومبیل‌رانی، از دنباله‌ای از بازیها تشکیل شده‌اند. تماشاگران چیزی بیش از سوختن توپ‌زنها، رانهای کامل، پاسهای کامل[۲] و قطع پاسها[۳] را می‌بینند. آنها یک سرمربی را می‌بینند که روی یک بازی زدن ـ و ـ دویدن قمار می‌کند یا یک پرتاب‌کننده توپ را که تصمیم می‌گیرد برای کسب اولین امتیازی که بشدت به آن نیازمند است، توپ را پاس دهد. تماشاگران درباره اینکه تیم آنها در وضعیتهای مختلف چه باید بکند دیدگاههایی دارند، و برای تصمیم‌گیری دراین‌باره که آیا سرمربی آنها یک استراتژی خوب یا ضعیف دارد، به تماشا نشسته‌اند.

مدیریت در ورزشهای حرفه‌ای مشابهتهای زیادی با مدیریت در بازرگانی و صنعت دارد. بعضی از مدیران زیرکتر از بقیه هستند، و بعضی از آنها بیش از دیگران از روشهای پیشرفته استفاده می‌کنند. این نکته در ورزش همانند بازرگانی، با وجود برداشت مردم از صفحات ورزشی روزنامه‌ها که اغلب وانمود می‌شود که سرمربیها به یک نسبت حرفه‌ای وخوب هستند، واقعیت دارد. بعضی؛ از کارکنان ناکارآمد خوب کار می‌کشند، اما هیچکدام نمی‌توانند بیش از

۱ـ *Pitcher*

۲ـ *Home Runs*

۳ـ *Completed Passes*

میزان معینی بر ناتوانی افرادی که برای آنها کار می‌کنند، غلبه نمایند. بعضی از آنها به‌طور غریزی قمارباز و بعضیها همواره محافظه کار هستند، اما فقط تعداد کمی از آنها به آن اندازه باهوشند که بسته به شرایط، قمارباز یا محافظه‌کار باشند. تفاوت عمده مدیران ورزشی از مدیران در بازرگانی در این است که عملکردهای آنها خیلی بیشتر مرئی است. به‌واسطه این مرئی‌بودن، همه ما می‌توانیم از تماشای حرکتهای مختلف آنها در جام زرین مسابقات ورزشهای حرفه‌ای، مطالبی بیاموزیم.

آمار و مدیریت

مطالب فوق چه ارتباطی با آمار پیدا می‌کند؟ نگرانی واقعی آمار این است که اطلاعات کمّی قابل استفاده را بخصوص در وضعیتهای پیچیده که دربرگیرنده بسیاری از متغیرها و عدم حتمّیتهاست کسب کند. "قابل استفاده" بدین معنی که هدف آن کمک‌کردن به ما برای این است که رفتار خود را در آینده بهتر کنیم، یعنی یاری کردن ما در یادگیری این است که چطور از چنین وضعیتهایی بیش از آنچه می‌کوشیم که کسب کنیم، استنباط نماییم. بعضی از مدیران از آمار، بخوبی استفاده می‌نمایند و بعضی نه؛ و این مطلب، چه مدیر یک کارخانه باشند و چه مدیر یک تیم بیس‌بال واقعیت دارد.

به‌عنوان مثال، فرض کنید لاستیک خودرو تولید می‌کنیم. بدون شک یک متخصص می‌تواند از این مثال بفهمد که من درباره تجارت لاستیک چیزی نمی‌دانم؛ درواقع این مثال را به این دلیل انتخاب کردم که هیچ وقت در تجارت لاستیک نبوده‌ام و بنابراین اشاره به افراد واقعی نمی‌کنم. فرض کنید در مرحله‌ای از فرایند تولید، حجمی از لاستیک مایع دراختیار داریم که بالاخره به لاستیک خودرو تبدیل می‌شود. چون آگاهیم که این محموله ممکن است خوب تهیه نشده‌باشد، مایلیم به طریقی آن را آزمایش کنیم تا اگر معیوب باشد، بتوانیم آن را دور بریزیم بدون اینکه سرمایه خود را در تولید لاستیکهای معیوب هدر بدهیم. متأسفانه، آزمایش واقعی یک لاستیک خودرو، آزمایش در جاده است، و ما نمی‌توانیم محموله مایعی را در جاده آزمایش کنیم، بنابراین باید به طریقی یک آزمون مناسب مثلا" اندازه‌گیری چسبندگی انجام دهیم. این اندازه‌گیری وقت‌گیر و پرهزینه است، و دیر یا زود یک نفر این سئوال را مطرح خواهد کرد که

"آیا این آزمون ارزش سرمایه‌گذاری دارد؟" این سئوال همواره سئوال خوبی است ، و معمولاً منجر به یک بحث داغ می‌شود. این بحث دربرگیرنده استدلالهایی است برپایه ادراک، تجربه، آزمونهای آزمایشگاهی و نظریه علمی؛ که هرکدام جایی برای خود در فرایند جستجو به دنبال حقیقت دارد، اما اساساً آنها پیشگویانی هستند، و تنها راه حصول اطمینان نسبت به آنچه که در میدان می‌گذرد این است که ببینیم واقعاً در میدان چه رخ می‌دهد؟ این بدان معنی است که مابابد به گردآوری آمار بپردازیم. بعد از اندازه‌گیری یک محموله، باید آن را در سراسر فرایند تولید دنبال کرده و کیفیت نمونه‌ای از لاستیکهای ساخته‌شده از این محموله را درعمل ببینیم. این کار را روی یک محموله دیگر انجام وبه‌همین ترتیب ادامه می‌دهیم. بعد از مدت زمانی می‌توانیم رابطه بین کیفیت لاستیک و چسبندگی را برقرار کنیم، وبا استفاده از این رابطه می‌توانیم هزینه‌های آزمون، هزینه فرایند تولید متعاقب آزمون، ارزش یک لاستیک تمام‌شده خوب و غیره را حساب؛ و آنگاه تعیین کنیم که آیا به انجام آزمون ادامه دهیم یا نه. البته اکنون ممکن است بگویید که این روش از خطا مصون نیست، زیرا درگیر نمونه‌گیری است و بنابراین، دربرگیرنده خطای نمونه‌گیری است (ن. ک. مقاله نتر) زیرا ممکن است فرایند به‌طور غیرقابل پیش‌بینی تغییر کند و قس‌علی‌هذا. اما این روش تاآنجا که امکان داشته‌باشد ما را به حقیقت نزدیک می‌کند، و این اعتراض آخر صرفاً مبین این نکته است که حتی اگر از آمار هم استفاده کنید، هیچوقت نمی‌توانید حقیقت را دریابید.

استراتژی بانتینگ(۱)

دانشجوی رفتار مدیریتی، مثالهای بسیاری از این نوع مسائل را در ورزش پیدا می‌کند و اگر زرنگ باشد، از اشتباهاتی که بوضوح در زمین بازی بیس‌بال یا فوتبال رخ می‌دهند، می‌تواند بهره‌مند شود. برای مثال ساکریفایس‌بانت(۲) را در بازی بیس‌بال درنظر بگیرید. افرادی هستند که روی آن قسم می‌خورند و افرادی هم هستند که بندرت آن را به‌کار می‌گیرند. این افراد درباره خوب‌بودن این استراتژی درگیر بحثهای پرحرارتی می‌شوند. همچنانکه خواهیم دید آمار

۱ـ *The Strategy of Bunting*

۲ـ *Sacrifice Bunt*

نمی‌تواند به‌طور قطعی به بحث فیصله دهد، ولی می‌تواند تا حد زیادی مسئله را روشن کند، و اگر مردم به بعضی از این واقعیات نگاه می‌کردند، بیشتر این مجادله ممکن بود حذف شود.

ساکریفایس بانت، سوزاندن بازیکن توپ زن است که به‌منظور اینکه یک دونده پایگاهی از پایگاه اول به دوم یا از پایگاه دوم به سوم برود، به‌کار گرفته‌می‌شود، و معمولاً به فداشدن توپ‌زنی منجر می‌شود که درپایگاه اول کنارگذاشته[۱] می‌شود.

بسیاری از سرمربیها ساکریفایس‌بانت را به‌طور عادی به‌کار می‌گیرند، و این عمل خود را به صورت "بیس‌بال درصدی[۲]" عنوان می‌کنند؛ گویی درصدها را می‌دانند (که ظاهراً بیشتر آنها نمی‌دانند.) روال کار این است که شما وقتی از بانت استفاده می‌کنید که یک بازیکن در پایگاه اول یا دوم باشد، هیچ بازیکنی نسوخته باشد، و تیم شما کمی جلو یا مساوی بوده، و یا اینکه "زیاد" عقب نباشد. برای تیم میهمان عقب بودن یک ران[۳] یا بیشتر، زیاد به‌نظر می‌رسد، و برای تیم میزبان دو ران یا بیشتر زیاد به‌حساب می‌آید. اختلاف ناشی از این واقعیت است که تیم میزبان آخرین توپ را می‌زند و می‌تواند به امید این باشد که "بازی را مساوی" کند.

چرا سرمربی اقدام به بانت می‌کند؟ البته او نهایتاً به‌منظور بردن بازیهای بیشتر، این کار را انجام می‌دهد. در لحظه‌ای که این کار را انجام می‌دهد، تلاشش بر این است که شانس کسب یک ران دیگر را زیاد کند درحالی که فرصت کسب چند ران را ازدست می‌دهد. نظریه این موضوع ساده است. در پایگاه اول، حداقل به دوضربه ساده یا یک ضربه مضاعف نیاز است تا دونده امتیاز کسب کند، درحالی که یک دونده در ایستگاه دوم معمولاً می‌تواند تنها با یک ضربه ساده امتیاز کسب کند. به‌علاوه اگر بتوان یک دونده پایگاه دوم را به پایگاه سوم منتقل نمود، درحالی که فقط یک سوخته وجود دارد، آنگاه امتیازی از هرضربه خطا، پرتاب توپ همراه[۴] باخطا، سوتی، توپ بلند هوایی، خطای توپ‌انداز[۵]، یا توپ ضربه‌خورده آهسته[۶] نتیجه می‌شود. طرفداران بانتینگ به ذکر این فهرست علاقه‌مند هستند، اما این فهرست دربرگیرنده رویدادهای

۱ـ *Thrown out*

۲ـ *Percentage Baseball*

۳ـ *Run*

۴ـ *Wild Pitch*

۵ـ *Balk*

۶ـ *Slow Grounder*

نسبتاً نادری است و این نکته سئوالهای واقعی را مطرح می‌کند ـ وقتی که ما از بانت استفاده می‌کنیم، به شانس کسب امتیاز یک ران **چقدر** افزوده می‌شود،و برحسب تعداد رانهای اضافی **چه تعداد ران** را فدا می‌کنیم؟ باز، تنها راه جواب دادن به این سئوال که به بازیکنان واقعاً مهم لیگ مربوط می‌شود که تحت فشار مسابقات واقعی بازی می‌کنند؛ این است که آمارها را از بازیهای واقعی به‌دست آوریم. راهی برای ایجاد شرایط واقعی انجام یک آزمایش وجود ندارد. و نظریه (ن.ک. مقاله‌کوک،۱۹۶۶؛ هوک،۱۹۶۷) ارزشی مبهم خواهدداشت.

گرچه سوابق مسابقات انجام‌شده در بایگانی سازمان‌یافته بیس‌بال موجود است، ولی تبدیل آنها به داده‌های قابل استفاده کار بزرگی است که اگر این کار انجام شده‌باشد، تا آنجا که من اطلاع دارم به استثنای مقاله *لیندسی*(۱۹۶۳) به اطلاع عموم نرسیده‌است. *لیندسی* به بحث درباره سوابق چند صد بازی مهم لیگ می‌پردازد که در سال ۱۹۵۹ و ۱۹۶۰ انجام گرفته و نتایج بسیار جالبی را به‌دست آورده‌است که بخشی از آنها در جدول ۱ نشان داده می‌شود.

برای اینکه ببینیم ما به‌عنوان سرمربیهای نیمکت‌نشین، این جدول را چگونه درمورد بانتینگ‌به‌کار می‌گیریم، بیایید به دو خط اول نگاه کنیم. (فعلاً، فقط به موارد میانگین فکر می‌کنیم، ولی‌هیچ آمارشناس خوبی تنها به میانگینها نمی‌اندیشد؛ بنابراین درباره وضعیتهای خاص بعداً بحث خواهیم کرد.) اکنون، مثلاً اینطور شروع می‌کنیم که یک نفر را در پایگاه اول داریم و هیچ بازیکن سوخته‌ای وجود ندارد. جدول به ما می‌گوید که این وضعیت ۱۷۲۸ دفعه مشاهده شده‌است (گهگاه، شاید بیش از یک‌دفعه در هر اینینگ). درنسبتی از این موارد مساوی ۰/۶۰۴، درمدت زمان باقیمانده اینینگ هیچ امتیازی به‌دست نیامده‌است. یعنی اینکه، در ۱۰۴۴ مورد، هیچ ران امتیازداری کسب نشد، و ۰/۶۰۴ = ۱۰۴۴/۱۷۲۸. این همچنین بدان معنی است که نسبت دفعاتی که حداقل یک ران امتیازدار از این وضعیت به‌دست آمده‌است، برابر با ۰/۶۰۴ - ۱ یا ۰۰/۳۹۶ است. ما این نسبتها را به‌عنوان احتمالهای برآوردشده پیشامدهای مختلف به‌کار می‌بریم؛ بدین‌ترتیب نزدیک به آخر یک بازی دوتیم با امتیازهای نزدیک به‌هم، عدد ۰/۳۹۶، میانگین "ارزش" وجود وضعیتی است که در آن یک بازیکن در پایگاه اول است و هیچ بازیکن سوخته‌ای وجود ندارد. برای قسمتهای اولیه بازی، این مقدار مربوط می‌شود به تعداد رانهایی که به‌طور متوسط، باشروع از این وضعیت، در یک اینینگ

جدول ۱ـ رابطه بین تعداد رانهای امتیازدار به پایگاه(های) اشغالی و تعداد افراد سوخته

تعداد موارد	متوسط تعداد رانهادریک اینینگ	نسبت مواردی که حداقل یک ران در یک اینینگ کسب شده است	نسبت مواردی که هیچ ران در یک اینینگ کسب نشده	تعدادافراد سوخته	پایگاه(های) اشغال شده
۱۷۲۸	۰/۸۱۳	۰/۳۹۶	۰/۶۰۴	۰	اولین
۶۵۷	۰۶۷۱	۰/۳۹۰	۰/۶۱۰	۱	دومین
۲۹۴	۱/۱۹۴	۰/۶۱۹	۰/۳۸۱	۰	دومین
۲۰۲	۰/۹۸۰	۰/۶۹۳	۰/۳۰۷	۱	سومین
۳۶۷	۱/۴۷۱	۰/۶۰۵	۰/۳۹۵	۰	اولین ودومین
۱۷۶	۱/۵۶	۰/۷۳۰	۰/۲۷	۱	دومین وسومین

مرجع: *لیندسی (۱۹۶۳)*

کسب می‌شود؛ این مقدار برای وضعیت مورد بحث در ستون چهارم به‌صورت ۰/۸۱۳ نمایش داده شده‌است.

اکنون اگر اقدام به انجام یک ساکریفایس‌بانت بکنیم که در حالت معمولی به موفقیت بینجامد، دونده پایگاه اول به پایگاه دوم نقل مکان می‌دهد و یک نفر می‌سوزد. آیا این وضعیت بهتر از وضعیتی است که در آن قرار داشتیم؟

اگر به مفهوم متوسط تعداد رانهای به‌دست آمده، نگاه کنیم قطعاً بدتر است؛ سطرهای اول و دوم جدول ۱ نشان می‌دهد که متوسط تعداد رانهای کسب‌شده از وضعیت یک نفر در پایگاه اول و بدون بازیکن سوخته معادل ۸۱۳ درهزار است، اما از وضعیت یک نفر در پایگاه دوم و یک نفر سوخته، این مقدار فقط ۶۷۱ درهزار می‌باشد. دراینصورت، به‌طور متوسط، یک بانت موفق معمولی درهر ۱۰۰۰ دفعه که به‌کار گرفته می‌شود، ۱۴۲ ران را از دست می‌دهد. اما درباره اینینگ آخر یک بازی با نتایج نزدیک به‌هم چه می‌توان گفت؛ زمانی که برای ما فقط مهم باشد که بدانیم چه بر سر احتمال کسب حداقل یک ران آمده‌است؟ مقدار آن از ۰/۳۹۶ به ۰/۳۹۰

نزول پیدا کرده‌است؛ این اعداد به‌قدری به‌هم نزدیک هستند که اختلاف آنها را می‌توان به آسانی به‌وسیله نوسانها شانسی مربوط به نمونه توجیه کرد. بنابراین نتیجه می‌گیریم که مزیت رفتن یک دونده از پایگاه اول به دوم تقریباً به طور صحیح با ضریب سوختن یک بازیکن اضافی که معمولاً" دریک بازی بانت رخ می‌دهد خنثی می‌شود.

نتیجه‌گیری

به‌طور متوسط، بانتینگ با داشتن یک نفر درپایگاه اول منجر به ازدست دادن تعداد زیادی ران می‌شود. به‌طور متوسط، احتمال کسب حداقل یک ران در یک اینینگ را افزایش نمی‌دهد. در اینجا فرض ما براین بوده‌است که توپ‌زن همواره در پایگاه اول می‌سوزد، اما البته، در بعضی مواقع او ایمن است، که در اینصورت کارایی بانتینگ را افزایش می‌دهد. اما، سوختن دونده جلو در پایگاه دوم، احتمالاً بیشتر رخ می‌دهد، که برای تیمی که اقدام به بانت نموده یک مصیبت است. به‌نظر می‌رسد که بانتینگ بایک نفر در پایگاه اول در اوایل بازی، فقط موقعی باید انجام گیرد که دفاع را طوری غافلگیر نماید که شانس ایمن بودن توپ‌زن خیلی بالا رفته‌باشد. حتی در اواخر یک بازی با امتیازهای نزدیک به‌هم، برتری مشهودی برای چنین بانتینگی وجود ندارد مگر اینکه شرایط خاصی حکمفرما باشد.

اکنون اجازه دهید که مسئله فردی را درنظر بگیریم که در پایگاه دوم بوده و هیچ فرد سوخته‌ای وجود نداشته‌باشد. جدول به ما می‌گوید که این فرد بخصوص (یا حداقل یک نفر) در تمامی ۱۰۰۰ مورد به‌غیراز ۳۸۱ مورد، یعنی در ۶۱۹ مورد از ۱۰۰۰ مورد امتیاز کسب می‌کند. اگر ما بتوانیم بافداکردن توپ‌زن او را به پایگاه سوم نقل مکان دهیم، می‌توانیم رقم ۶۱۹ رابه ۶۹۳ افزایش دهیم. (ملاحظه کنید با انجام این کار در ۱۰۰۰ دفعه، ۲۱۴ ران را از دست می‌دهیم، اما باز وضعیتی را درنظر بگیرید که در آخر یک بازی هستیم و فقط نیاز به یک ران داریم.) در اینجا براستی از یک بازی بانت موفقیت‌آمیز سودی عایدمان می‌شود، اما وقت آن رسیده که با واقعیت روبه‌رو شویم: بازی بانت همیشه مؤثر نیست. اینکه تا چه حدودی مؤثر است، به چیزهای زیادی بستگی دارد، و ما آمار لازم برای یک نتیجه متوسط را دراختیار نداریم، اما اکنون ببینیم اگر داشتیم چطور آن را به‌کار می‌گرفتیم.

اگر توپ‌زن، توپ را کمی محکم بانت کند، تیم مدافع با خوشحالی توپ را خیلی سریع به پایگاه سوم رسانیده و دونده اصلی می‌سوزد، که باعث می‌شود تیم مهاجم یک نفر در پایگاه اول و یک بازیکن سوخته داشته باشد، احتمال کسب حداقل یک ران از ۰/۶۱۹ به ۰/۲۶۶ رسیده، که ۰/۲۶۶ از جدول کامل مقاله *لیندسی* به‌دست آمده‌است. سرمربی نوعی، احتمال چنین پیشامدی را قبول ندارد و بعد از اینکه رخ می‌دهد، با تذکر اینکه "این جوانها بلد نیستند چطور بانت کنند، مثل آن ایام که ما عمل می‌کردیم" آن را رد می‌کند. من می‌دانم که این تذکر قبل از ظاهرشدن هرکدام از سرمربیهای امروزی در اولین بازی خود به‌عنوان بازیکنهای حرفه‌ای هم داده می‌شد و احتمالاً درقرن نوزدهم به‌وسیله اولین سرمربی که خود بازی نمی‌کرد عنوان شد (این تذکر صرفاً بهانه‌ای برای بررسی نکردن مشکل است، اما بهتر است که به سرمربیان بیس‌بال سخت نگیریم؛ تا حالا اشاره کرده‌ایم حرکاتی که آنها آشکارا انجام می‌دهند، توسط انواع سرپرستهای دیگر در وضعیتهایی که کمتر آشکارند، تکرار می‌شود.)

همانطور که در بالا گفتیم، ما آمار نتایج یک بانت را با یک بازیکن در پایگاه دوم دراختیار نداریم، بنابراین من سعی می‌کنم بعضی از آنها را براساس مشاهدات شخصی ثبت‌نشده در طول سالها، با واقعگرایی ممکن سازم.

۱ـ ۶۵٪ اوقات دونده به پایگاه سوم نقل مکان می‌دهد، و توپ‌زن می‌سوزد. (حالت معمولی)

۲ـ ۱۲٪ اوقات دونده در پایگاه سوم می‌سوزد، و توپ‌زن در پایگاه اول ایمن است (با موفقیت به پایگاه اول رسیده‌است).

۳ـ ۱۰٪ از اوقات دونده باید در پایگاه دوم بماند، و توپ‌زن می‌سوزد، مثل موقعی که توپ‌زن یک ضربه کوتاه هوایی[۱] می‌زند یا اینکه می‌سوزد.

۴ـ ۸٪ اوقات توپ‌زن سالم به پایگاه اول می‌رسد، یعنی اینکه ضربه می‌زند و دونده نیز پیشروی می‌کند.

۵ـ ۵٪ اوقات زننده بانت موجب بازی دوسوخته[۲] می‌شود، بدین معنی که، او و دونده هردو می‌سوزند.

۱ـ *Pop Fly*

۲ـ *Double Play*

حال برای اینکه احتمال کل کسب حداقل یک ران را حساب کنیم، صرفاً طبق قواعد احتمال، عمل ضرب و جمع انجام می‌دهیم. اگر این قواعد را نمی‌دانید، به این صورت عمل کنید: با ۱۰۰۰ مورد شروع کنید. در ۶۵۰ مورد (یعنی ۶۵٪ موارد) نتیجه ۱ را خواهیم داشت، یعنی اینکه یک نفر در پایگاه سوم و یک نفر سوخته است. جدول چنین می‌گوید که این فرد ۶۹/۳ از اوقات موفق به کسب امتیاز می‌شود، بنابراین ۶۹/۳٪ از ۶۵۰ را اختیار می‌کنیم و عدد ۴۵۰ را به‌دست می‌آوریم. یعنی، در ۴۵۰ مورد عمل بانت طبق نتیجه ۱ منجر به موفقیت می‌شود، و یک ران بالاخره به‌دست می‌آید. حال در ۱۲۰ مورد، برآمد طبق نتیجه ۲ است، و جدول کامل *لیندسی* چنین می‌گوید که در ۲۶/۶ ٪ اوقات یک امتیاز کسب می‌شود. بنابراین ۲۶/۶٪ از ۱۲۰ را برداشته و عدد ۳۲ را به‌دست می‌آوریم. این عدد را با ۴۵۰ جمع نموده و ادامه می‌دهیم. با استفاده از جدول کامل *لیندسی* درمواقع لزوم آنچه که برای تمام پنج حالت به‌دست می‌آوریم، به‌صورت زیر است:

$$۰/۹۶۳\ (۶۵۰) + ۰/۲۶۶\ (۱۲۰) + ۰/۳۹۰\ (۱۰۰) + ۰/۸۷\ (۸۰) + ۰/۰۶۷\ (۵۰) =$$

$$۴۵۰ + ۳۲ + ۳۹ + ۷۰ + ۳ = ۵۹۴$$

به بیان دیگر در ۵۹۴ مورد از ۱۰۰۰ مورد حداقل یک ران کسب می‌کنیم. قبل از عمل بانت شانس ما ۶۱۹ از ۱۰۰۰ مورد بود، بنابراین شانس خود را کاهش داده‌ایم. البته اگر داده‌های فرضی در قسمتهای ۱ تا ۵ در بالا بیش‌از حد بدبینانه باشد، نتیجه صحیح برای بانتینگ کمی مساعدتر است، اما به‌نظر می‌رسد که هر برآورد واقع‌بینانه‌ای منجر به این نتیجه می‌شود که به‌طور متوسط بانتینگ سودمند نیست.

استفاده هوشمندانه از آمار مستلزم چیزی فراتر از یک نگاه صرف به مقادیر متوسط است. داده‌های فوق و بحثهای همراه با آن نشان می‌دهد که اگر بانتینگ را به‌طور عمومی مورد استفاده قرار دهیم، همانطور که سرمربیها عمل می‌کنند، استراتژی برنده نیست. اما این بدان معنی نیست که نباید هیچوقت اقدام به بانت کرد. فردی ممکن است توپ‌زنی ضعیف ولی بانتر بسیار خوبی باشد، وفرد بعد از او ممکن است که توپ‌زن خوبی باشد؛ توپ‌زن ممکن است توپ‌اندازی باشد که قدرت توپ‌زنی او صفر است، اما بعضی اوقات می‌تواند یک بانت خوب انجام دهد؛ و یا اینکه تیم مقابل به‌طور وضوح در انتظار یک بانت نباشد، به‌طوری که عامل

غافلگیر کردن به ما کمک کند تا عمل بانت منجر به ضربه‌ای شود که توپ‌زن، سالم به پایگاه برسد. در هرکدام از این موارد بانت می‌تواند یک عمل سودمند باشد. نقش آمار این است که نشان دهد رفتار معمولی ما باید به چه صورتی باشد. به‌طور کلی، اگر متوسط نتیجه یک استراتژی خوب باشد، باید آن را زیاد به‌کار ببریم. اگر متوسط نتیجه بد باشد، باید خیلی کم از آن استفاده کنیم، یعنی اینکه موقعیتهای خاصی که منجر به استفاده از آن می‌شوند، باید بسیار خاص باشد. افرادی هستند که می‌گویند آمارها نامربوط هستند و آنها هر موردی را یک حالت خاص درنظر می‌گیرند. این کار احتمالاً غیرممکن است، و اگر چنین افرادی رفتار خود را در طول یک مدت طولانی بررسی کنند، محتملاً درمی‌یابند که رفتار آنها از لحاظ آماری کاملاً قابل پیش‌بینی است. درواقع اگر از این دیدگاه بنگریم که کل ماجرا غافلگیر کردن است، یعنی هدف این است که غیرقابل پیش‌بینی باشیم، آنگاه یک استرتژی تصادفی‌شده توصیه می‌گردد؛ شرح جزئیات مربوط به این موضوع را می‌توان در هر کتابی درباره نظریه بازی پیدا نمود.

استراتژی گام عمدی[۱]

یک حرکت استراتژیک دیگر در بیس‌بال پایگاه عمدی روی توپهاست. تیم مقابل، مثلاً" یک نفر را در پایگاه دوم با یک یار سوخته دارد، و ما تصمیم می‌گیریم که عمداً یک نفر را در پایگاه اول بگذاریم، به این منظور که یا یک بازی دوسوخته داشته‌باشیم و یااینکه وضعیتی ایجاد کنیم که در تمام سه پایگاه امکان یک بازی اجباری وجود داشته‌باشد. آیا انجام این کار باارزش است؟ جدول *لیندسی* نشان می‌دهد که قبل از گام عمدی احتمال کسب حداقل یک ران برابر ۰/۳۹۰ است، اما بعد از گام عمدی این احتمال برابر ۰/۴۲۹ می‌باشد. واضح است که به‌طور متوسط گام عمدی یک حرکت توأم باضرر است؛ اگر بعد از آن یک بازی دوسوخته صورت گیرد، بسیارجالب است، اما اگر بعد ازآن یک گام غیرعمدی صورت گیرد، این حرکت می‌تواند به یک فاجعه منجر شود، واین احتمال دوم قسمتی از دلیل نتایج عددی است که در بالا ذکر شد. به نظر می‌رسد که استفاده گسترده از گام عمدی صرفاً برپایه خوش‌بینی است، زیرا ظاهراً آمار نشان

۱ـ *Intentional Walk*

می‌دهد که از دیدگاه تیم که در زمین است، اثرهای بد این حرکت، قطعاً به‌طور متوسط بیش‌از اثرهای خوب آن می‌باشد. اما درباره موارد خاص چه می‌توان گفت؟ اگر توپ‌زن، توپ‌زنی خوب، و توپ‌زن بعدی ضعیف باشد، آنگاه داده‌های فوق لزوماً به‌کار نمی‌روند و گام عمدی ممکن است حرکت خوبی باشد. در اوایل بازی بندرت باید اقدام به چنین حرکتی کرد، مگر اینکه توپ‌زن بعدی یک توپ‌زن ضعیف باشد، زیرا این باعث می‌شود که متوسط تعداد رانها به‌واسطه دونده اضافی پایگاه از ۶۷۱ ران در ۱۰۰۰ به ۹۳۹ ران در ۱۰۰۰ افزایش پیداکند، و مشکوک به‌نظر می‌رسد که موارد خاص زیادی وجود داشته‌باشد که آنقدر خاص باشد که این واقعیت را تحت‌الشعاع قراردهد. داده‌های جدول ۱ برای آخر بازی ممکن است بیش از اندازه به‌نفع گام عمدی خوش‌بینانه باشد، به‌دلیل اینکه فشار اضافی زیادی روی توپ‌اندازی اعمال می‌شود که برای پرکردن پایگاهها به‌وسیله ران بالقوه تمام‌کننده بازی ازپایگاه سوم به یک گام قصدی اقدام می‌نماید، و به‌نظر می‌رسد که این فشار موجب افزایش تعداد پرتابهای با خطای زیاد، توپ‌زنهایی که توپ به آنها خورده و گامهای غیرعمدی می‌شود.

گردآوری و استفاده از داده‌ها

واضح است که ارقامی مثل ارقام جدول ۱ کم‌ارزش هستند، مگر اینکه براساس تعداد موارد نسبتاً زیادی باشند. گرچه اصلاً واضح نیست که تعداد موارد چه اندازه باید بزرگ باشد. آمار ریاضی به سئوالهایی درباره میزان بزرگی اندازه نمونه پاسخ می‌دهد، اما این سئوالها باید مشخص باشند. بعضی اوقات وقتی با چنین سئوالهایی مواجهیم، نمی‌توانیم بگوییم که ۱۰۰ (یا ۱۰۰۰ یا ۲۰۰۰) یک اندازه نمونه جامع خوب می‌باشد. اما، اگر از ما بپرسند وقتی یک فرد در پایگاه اول بوده و هیچ فرد سوخته‌ای وجود ندارد، احتمال اینکه حداقل یک ران به‌دست آید چقدر است؟ می‌توانیم با برخی فرضهای ساده‌کننده معقول تعیین کنیم که اندازه نمونه چه مقدار باید بزرگ باشد تا مثلاً، ۹۰٪ مطمئن باشیم که در فاصله ۰/۰۰۵ جواب صحیح قرار بگیریم. ستون آخر جدول ۱، اندازه نمونه‌ای که برای تولید داده‌های ستونهای قبلی به‌کار برده شده‌است را نشان می‌دهد. برای فردی که سوابق را به‌عنوان سرگرمی نگه می‌دارد، این کار یک تلاش چشمگیر می‌باشد. ما چنین فکر می‌کنیم که افراد دست‌اندرکار بازی بیس‌بال که درگیر

رقابتی هستند که چند پیروزی اضافی می‌تواند باعث اختلاف مبالغ هنگفتی شود، این زحمت را به‌جان می‌خرند که حتی نمونه‌های بزرگتری را جمع‌آوری کنند. اما آنها تمایل نخواهند داشت که بیش ازحد به این کار مبادرت ورزند، زیرا اطلاعات گرایش به کهنه‌شدن دارند. تغییرات در قواعد، زمینهای بازی، و کارکنان سبب می‌شود که ماهیت بازی سال به سال کمی تغییر پیدا کند. بعضی اوقات کسب امتیاز برای چندسالی نسبتاً پایین است، و سپس برای چندسال افزایش پیدا می‌کند. داده‌هایی که در یکی از این دوره‌های زمانی گردآوری می‌شود، ممکن است روی‌هم مبنای درستی برای تصمیم‌گیری در دوره‌ای دیگر باشند.

چنین داده‌هایی نظیر آنچه مشغول صحبت درباره آنها بوده‌ایم، گاهی اوقات برخلاف داده‌ای آزمایشی یا کنترل‌شده، **داده‌های تاریخی** نامیده می‌شوند. این تفاوت دربسیاری از زمینه‌ها مهم است. برای مثال، اگر داده‌هایی به‌دست آید که نشان دهد موارد سرطان ریه در افراد سیگاری نسبت به غیرسیگاریها خیلی بیشتر است، این حقیقت تاریخی به خودی خود ثابت نمی‌کند که مصرف سیگار باعث افزایش نرخ سرطان ریه می‌شود (مثلاً، بچه‌ها بیش از بزرگسالان شیر می‌نوشند، اما این دلیل آن نیست که آنها بچه هستند). ممکن است متغیرهای دیگری نیز دخیل باشند، که به‌عنوان مثال به ابتلای سرطان ریه کمک می‌کنند و همچنین باعث سیگاری‌شدن افراد می‌شوند. لیکن آمارهای تاریخی درباره سرطان بسیار القاکننده بوده و منجر به آزمایشهای مختلفی در آزمایشگاهها شده که اعتقاد بیشتر افراد را ازنظر یک رابطه علت و معلولی مستحکم نموده‌است. به بیان دیگر ما نمی‌توانیم از داده‌های تاریخی استفاده خوبی بکنیم، اما درمورد استنباط علت و معلولی از آنها باید محتاط باشیم.

بدون شک، به‌واسطه عجز در تلاش برای نتیجه‌گیری از داده‌های تاریخی، آمارشناسان علم وهنری موسوم به طرح آزمایشها را توسعه داده‌اند. اگر بتوانیم آزمایشی را انجام دهیم که درست طراحی شده‌باشد، برای نتیجه‌گیری اینکه چه چیزی سبب چه چیزی می‌شود، در موقعیت خیلی بهتری خواهیم بود. اما امکان یک آزمایش طراحی‌شده همیشه وجود ندارد. وقتی که نمی‌توانیم آزمایش کنیم، باید ببینیم با داده‌های موجود چه می‌توانیم بکنیم، اما این بدان معنی نیست که نباید چشمهایمان را دربرابر نارساییهای این چنین داده‌هایی باز نگه نداریم.

نتیجه‌گیریها

از نگاه به آمارهای ورزشی چه آموخته‌ایم؟ آموخته‌ایم که چه کارهایی را انجام دهیم و چه کارهایی انجام ندهیم:

۱ـ اگر می‌توانید آمارهایی را گردآوری کنید که به‌طور واقع‌بینانه به سئوال پاسخ می‌دهند، با بحث درباره نکات مثبت و منفی یک چیز وقت خود را تلف نکنید.

۲ـ اگر تلاش دارید که روابط علت و معلولی برقرار کنید، حتماً این کار را با یک آزمایش انجام دهید که درست طراحی شده‌باشد.

۳ـ اگر نمی‌توانید آزمایش انجام دهید، از هر داده‌ای که می‌توانید گردآوری کنید، بهترین استفاده را بنمایید، اما نسبت به داده‌های تاریخی با شک وتردید زیاد برخورد کنید و آنها را در معرض تمام آزمونهای منطقی که می‌توانید تصور کنید، قرار دهید.

۴ـ خوب به‌خاطر داشته باشید که تجربه شخصی شما صرفاً مجموعه درهمی از آمار متشکل از مواردی هستند که برحسب اتفاق، به‌خاطر می‌آورید. چون تعداد این موارد لزوماً کم می‌باشد و چون حافظه شما ممکن است متمایل به یک نتیجه یا نتیجه دیگری باشد، لذا تجربه شما ممکن است نسبت به مجموعه‌ای از آمارها خیلی کمتر قابل اعتماد باشد. (تمایلی که در اینجا از آن سخن گفتیم، برای مثال می‌تواند از این واقعیت ناشی شود که افرادی که معتقد به عمل بانت هستند گرایش دارند مواردی را به‌خاطر بیاورند که در آنها عمل بانت موفقیت‌آمیز می‌باشد و بالعکس.

۵ـ ولی حتماً به‌خاطر داشته‌باشید که این نوع آمارها که در اینجا مورد بحث قرار گرفتند به‌صورت مقادیر متوسط می‌باشند، و موارد خاص ممکن است نیازمند عمل خاصی باشند. این بهانه‌ای نیست که همواره به دنبال احساس خود باشید، اما بدین معناست که کاربرد ۱۰۰٪ چیزی که به‌طور متوسط بهترین می‌باشد، ممکن است یک استراتژی سودمند نباشد. یک مدیر خوب راهبردی را اتخاذ می‌کند که شاید برپایه آمار بوده و دربرگیرنده بیشترین موارد تصمیم‌گیری می‌باشد. یک مدیر خیلی خوب یادگرفته است که موارد نادری را تشخیص دهد که در آنها برای کسب بیشترین سودمندی نیاز به تغییر راهبرد است.

از زمان نوشتن این مقاله، کامپیوترها به حیطه ورزش حمله‌ور شده‌اند، درست همانطور که

به حیطه‌های فعالیت بسیار دیگری هجوم برده‌اند، و منجر به استفاده افزایشی بسزایی از آمار شده‌اند. درنتیجه، بعضی از تذکرات من درباره استفاده از آمار، بخصوص در بیس‌بال، به صورتی که ذکر شد، دیگر صحت ندارند. همچنین، به‌واسطه تغییراتی که در بازی صورت گرفته و دامنه آن از ورزشگاههای جدید مخصوص بیس‌بال تا تغییرات در قوانین بازی مثل معرفی بازیگر تعیین‌شده به‌عنوان توپ‌زن گسترش یافته‌است، داده‌های جدول ۱ مربوط به سالهای ۱۹۶۰ ـ ۱۹۵۹ ممکن است برای یکی از لیگهای مهم منسوخ شده‌باشد. تمامی این نکات بدین مفهوم است که مثالها قدیمی هستند درحالی که اصول کلی صادق می‌باشند. درواقع، شباهت جدیدی بین مدیریت در ورزش و مدیریت در بازرگانی اضافه شده‌است، یعنی استفاده روزافزون از آمار لزوماً منجر به استفاده هوشمندانه‌تر از آمار نشده است. از زمان بازنشستگی من در سال ۱۹۷۹ بازیهای لیگ اصلی بیس‌بال را از نزدیک دنبال نکرده‌ام، اما از تماشای بازیهای بعد از فصل چنین نتیجه گرفته‌ام که کمتر سرمربیهایی پیدا می‌شوند (اگر اصلاً پیدا شوند) که از افزایش اطلاعات موجود به‌منظور بهترکردن استراتژی خود استفاده کنند.

مسائل

۱ـ به جدول ۱ مراجعه کنید. در چند مورد با وضعیت یک‌نفر در پایگاه سوم و یک نفر سوخته، هیچ رانی به‌دست نیامده‌است؟

۲ـ فرض کنید پایگاه دوم اشغال شده و هیچ سوخته‌ای وجود ندارد و یا یک سوخته وجود دارد، در چند حالت مربوط به این موارد درجدول ۱، هیچ رانی در یک اینینگ به‌دست نیامده‌است؟

۳ـ فرض کنید در پایگاه اول و دوم دوندگانی وجود داشته باشند و فرد سوخته‌ای موجود نبوده و اوایل بازی باشد. فرض کنید توپ‌زن بسوزد و دونده‌ها یک پایگاه جلو بروند، آیا ارقام جدول ۱ انجام یک بانت را توصیه می‌کنند؟ جواب خود را توجیه کنید.

۴ـ فرض کنید در پایگاه اول ودوم دوندگانی داشته‌باشیم، هیچ سوخته‌ای وجود نداشته‌باشد و اواخر اینینگ یک بازی با نتایج نزدیک به‌هم باشد. به فرض آنکه توپ‌زن بسوزد و دونده‌ها یک پایگاه جلو بروند، آیا ارقام جدول ۱، یک بانت را توصیه می‌کنند؟ جواب خود راتوجیه کنید.

۵ـ فرض کنید آمار مربوط به نتایج اقدام به یک بانت در وضعیت یک فرد در پایگاه دوم به‌جای

۶۵٪، ۱۲٪، ۸٪، و۵٪ که نویسنده فرض کرده‌است به‌ترتیب مقادیر ۷۰٪، ۱۳٪، ۹٪، و۳٪ باشد آیا درچنین وضعیتی به‌طور متوسط عمل بانتینگ سودمند می‌باشد؟ جواب خود را توجیه کنید.

۶ـ در چه موقعی یک بانت ساکریفایس ممکن است در یک وضعیت حرکت عاقلانه‌ای باشد در صورتی که به‌طور متوسط چنین حرکتی عاقلانه نباشد؟

۷ـ الف) اختلاف بین داده‌های **تاریخی** و داده‌های **آزمایشی** را نشان دهید.

ب) چرا لیندسی اقدام به انجام یک آزمایش کنترل شده ننمود؟

۸ـ از آمارهای اضافی زیر مربوط به **لیندسی** و درصد برآمدهایی که در متن داده‌شده، استفاده کنید. فرض نمایید یک نفر در پایگاه دوم و یک نفر سوخته وجود داشته‌باشد. توپ‌زن اقدام به انجام یک بانت می‌نماید.

پایگاه اشغال شده	تعدادسوخته ها	احتمال آنکه هیچ ران کسب نشود
۱	۲	۰/۸۸۶
۲	۲	۰/۷۸۸
۳	۲	۰/۷۳۸
۱و۳	۱	۰/۳۶۷

الف) چند دفعه (از ۱۰۰۰ مورد) حداقل یک ران کسب می‌شود؟

ب) چطور حالت ۵ (که بانتر ضربه می‌زند و دونفر می‌سوزند) در محاسبه شما وارد می‌شود؟

منابع

E. Cook. 1966, *Percentage Baseball.* Cambridge. Mass: MIT Press.

R. Hooke. 1967, Review of Cook (1966). *Journal of the American Statistical Association* 62:688-690.

R. Hooke. 1983, *Hour to Tell the Liars from the Statisticians.* New York: Marcel Dekker. (See Specially Sections 18, 25, 28, and 61)

G.R. Lindsey. 1963, "An Investigation of Sterategies in Baseball." *Operations Research 11:477-501.*

J. Thom and P. Palmer. With D. Reuther. 1985, *The Hidden Game of Baseball. Garden City.*

توضیحی درباره بازی بیس‌بال

بیس‌بال بازیی است که بین دو تیم ۹نفره با استفاده از یک راکت (Bat) و یک توپ در زمینی مربع شکل انجام می‌گیرد. هرگوشه زمین موسوم به یک پایگاه (Base) است. این پایگاهها مسیری را مشخص می‌کنند که تیم مهاجم باید طی کند تا موفق به کسب امتیاز (Run) شود. درگوشه مربوط به پایگاه اصلی یک صفحه پنج ضلعی موسوم به (Home plate) قرار دارد. پایگاه اصلی محل قرارگرفتن توپ‌زن (Balter) نیز می‌باشد. سه گوشه دیگر مربع درجهت عکس عقربه‌های ساعت به‌ترتیب به پایگاههای اول و دوم و سوم موسوم است. هر تیم به‌نوبت در وضعیت حمله و دفاع قرار می‌گیرد. بدین‌ترتیب در هرنوبت یک تیم موسوم به تیم مهاجم وتیم دیگر موسوم به تیم مدافع است. تیم مدافع چهار بازیکن در اطراف این فضای مربع شکل مستقر در زمین دارد: یک نفر نزدیک پایگاه اول، نفر دوم بین پایگاه اول و پایگاه دوم؛ نفر سوم نزدیک پایگاه سوم و نفر چهارم بین پایگاه دوم و پایگاه سوم قرار می‌گیرند. همچنین سه بازیکن در فضای خارج قرار می‌گیرند: یک نفر در سمت چپ، یک نفر در وسط و یک نفر درسمت راست زمین. بازی به این ترتیب انجام می‌گیرد که یک توپ‌انداز (pitcher) تقریباً در وسط فضای مربع شکل اصلی قرار می‌گیرد و توپ را به‌طرف یک توپ‌گیر (Catcher) که نزدیک پایگاه اصلی قرار دارد، پرتاب می‌کند. توپ‌زن (Balter) که جزو تیم مهاجم است، در نزدیکی پایگاه اصلی و توپ‌گیر قرار دارد و سعی می‌کند که به توپ ضربه بزند تا توپ دور از دسترس افراد تیم مدافع قرارگیرد و او بتواند فاصله بین پایگاهها را بدود قبل از اینکه بسوزد (Putout). به صور مختلف یک بازیکن ممکن است بسوزد:

وقتی که توپ‌زن بایک ضربه توپ را به هوا پرتاب می‌کند، طوری که قبل از اینکه توپ به زمین برسد، توسط بازیکن تیم مدافع گرفته شود. تماس بازیکن با توپ وقتی که او در پایگاه نباشد و قبل از اینکه بازیکن به پایگاه بعدی برسد، این پایگاه به‌وسیله یک بازیکن تیم مدافع که توپ همراه اوست لمس شود .توپ‌انداز می‌باید توپ را در مسیر خاص پرتاب کند به‌طوری که برای توپ‌زن میسر باشد که به‌وسیله راکت به توپ ضربه زند؛ اگر چهار پرتاب توپ‌انداز در مسیر قراردادی نباشد و توپ‌زن به آن ضربه نزند، توپ‌زن در پایگاه اول قرار می‌گیرد. توپ‌زن می‌باید به هر توپ که در مسیر قرار می‌گیرد، ضربه بزند؛ اگر سه تا از این توپها رااز دست بدهد توپ‌زن می‌سوزد (Strikesout).

وقتی که بازیگر به یک پایگاه می‌رسد، می‌تواند بدون اینکه بسوزد در آن پایگاه بماند؛ مگر اینکه

وقتی توپ‌زن ضربه می‌زند مجبور به پیشروی به سمت پایگاه بعدی شود. اگر یک بازیکن تیم مهاجم بتواند به تمام پایگاهها پیشروی کند و موفق شود پایگاه اصلی (Home plate) را لمس کند او موفق به کسب یک ران (Run) برای تیم خود می‌شود.

بازی به‌وسیله یک داور (Umpire) که در پشت توپ‌گیر قرار دارد، کنترل می‌شود. سه کمک داور که هرکدام نزدیک یکی از پایگاهها قرار دارند، نظارت بربازی داشته و به داور کمک می‌کنند.

یک بازی معمولی از ۹ اینینگ (Inning) تشکیل شده که هر اینینگ شامل یک نوبت دفاع و یک نوبت حمله برای هر تیم است. بازیکنان تیم مهاجم به‌صورت دوره‌ای به توپ‌زنی ادامه می‌دهند تا اینکه ۳نفر ازآنها بسوزند که در اینصورت جای خود را با تیم مدافع عوض می‌کنند. در آخر بازی تیمی که موفق به کسب ران بیشتر شده‌باشد، برنده اعلام می‌شود ـ (مترجم).



شاخص بهای کالاها و خدمات مصرفی

فیلیپ جی. مک‌کارتی(۱) مدرسه ایالتی روابط کاروصنعت نیویورک، دانشگاه کرنل

تقریباً همه مردم درمورد قیمت انواع کالاها نگران هستند. شاید بهترین راه برای یک فرد جهت اطلاع از چگونگی تغییرات سطح قیمت مراجعه به شاخص بهای کالاها و خدمات مصرفی (CPI) تهیه‌شده توسط دفتر آمار کار، بخشی از وزارت کار آمریکا، در روزنامه‌ها باشد. مصرف‌کنندگان، CPI را معیار خوبی از تغییرات قیمت کالاها وخدمات مصرفی خود می‌دانند.

۱- *Philip J. Mc Carthy*

آنها به مطالبی در روزنامه‌ها، واکنش نشان می‌دهند. مانند اینکه "در سپتامبر ۱۹۸۶، یک خانوار متوسط شهرنشین ۳۳/۰۲ دلار برای همان مقدار از کالاها و خدمات خرج کرد که در سال ۱۹۶۷، ۱۰ دلار هزینه داشت." آنها نمی‌دانند که افزایش درآمد جهت جبران این افزایش قیمت کافی بوده‌است.

اتحادیه‌ها و مدیریت توجه خاصی به CPI مبذول می‌دارند، زیرا می‌دانند که مقدار آن نقش حساسی در توافقهای دستمزدی خواهدداشت و برای مثال یک افزایش سالانه ۴٪ در CPI حداقل یک افزایش ۴٪ در دستمزدها را درپی خواهدداشت. به‌علاوه، در ژانویه ۱۹۸۶، دستمزد ۴/۵ میلیون کارگر شاغل مبتنی بر قراردادهایی بود که تغییر خودکار در نرخ دستمزد را باتوجه به تغییر مشخص CPI طلب می‌نمود.

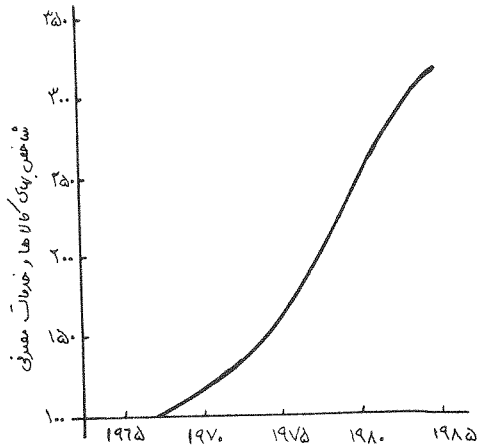
آمریکاییهای مستمری‌بگیر از تأمین اجتماعی مراقب تغییرات CPI می‌باشند، زیرا برای افزایش مستمری آنها به‌کار می‌رود. از CPI برای شاخص‌بندی پرداختها در برنامه‌های دیگر دولت نیز، مانند کوپن خوراک، استفاده می‌شود. CPI بر درآمد دولت نیز اثر می‌گذارد، زیرا از آن درتعدیل گروههای درآمدی مشمول مالیات در قانون مالیات بردرآمد استفاده می‌شود.

بالاخره، اقتصاددانان از CPI به‌عنوان یکی از نماگرهای عمده در ارزیابی عملکرد اقتصاد استفاده می‌نمایند. همانگونه که نمودار ۱ نشان می‌دهد، دامنه تورم در طول زمان به‌طور قابل ملاحظه‌ای افزایش یافته‌است.

گرچه مطالعات آماری در مورد قیمتها و شرایط زندگی در ایالات متحده در اواخر قرن نوزدهم و اوایل قرن بیستم به مرحله اجرا درآمد، ولی اولین شاخصهای کامل "هزینه زندگی" توسط دفتر آمار کار در سال ۱۹۱۹ منتشر شد. آنها ۳۲ مرکز صنعتی و کشتی‌سازی بزرگ را پوشش دادند و درحین تنظیم قرارداد بین هیئت تعدیل نیروی کار کشتی‌سازی و رهبران کارگران مطرح شدند؛ مبنی بر اینکه یکی از عواملی را که برای پایان دادن به اختلافات کارگری باید مدنظر قرار داد "تعدیل دستمزدها باتوجه به افزایش هزینه زندگی در نتیجه جنگ" می‌باشد.

از آن زمان CPI زمینه وسیعتری یافته و اهمیت آن بیشتر شده‌است. این موضوع موجب چندین ارزیابی فنی از CPI شد، که در میان آنها یکی به‌وسیله کمیته‌ای مشورتی از انجمن آمارشناسان آمریکا در ۱۹۳۴-۱۹۳۳ و یکی به‌وسیله دفتر ملی تحقیقات اقتصادی در

نمودار ۱ـ شاخص بهای کالاها و خدمات مصرفی برای مناطق شهری (CPI - U)، میانگینهای سالانه (۱۹۶۷=۱۰۰)



۱۹۶۰-۱۹۵۹ انجام شده‌است. این ارزیابیها باعث تجدیدنظرهای عمده‌ای در CPI (با فواصل زمانی تقریبی ۱۰ سال از ۱۹۴۰ به بعد)، همچنین تجدیدنظرهای مختصر بسیاری شده‌است.

مشکلات ساخت CPI

مصرف‌کنندگان واقعی از تغییراتی که در بهای کالاها و خدمات به‌طور مرتب روی می‌دهد، کاملاً آگاهی دارند. آنها می‌دانند چه زمانی قیمت یک گالن و یا قیمت یک لیتر شیر یک سنت افزایش می‌یابد. به‌علاوه آنها قادر هستند با دقتی قابل‌قبول اثر این تغییر قیمت را بر بودجه‌های ماهانه یا سالانه خود پیش‌بینی نمایند و بنابراین در موقعیتی قرار دارند که درآمد اضافی (شامل تخفیف مالیات) لازم برای جبران این افزایشها را برآورد نمایند. تغییر قیمت درمورد خریدهایی با فراوانی کمتر و پراکندگی بیشتر پوشاک، خدمات پزشکی، لوازم خانگی، خودرو و خانه)، آنقدر به چشم مصرف‌کننده نمی‌آید و ارزیابی تأثیر آنها بر بودجه به‌مراتب مشکلتر است. بنابراین یک نفر ممکن است به‌خوبی بداند که هزینه زندگی در سال جاری نسبت به سال گذشته بیشتر یا کمتر شده‌است ولی برایش تقریباً غیرممکن باشد که مقدار دقیق این تغییر در

هزینه زندگی را بیان نماید. این مشکل باتغییراتی که ممکن است در مواردی مانند وضعیت خانوادگی (تولد، مرگ، طلاق)، تمایلات خانوادگی (خودروهای بزرگ به‌جای خودروهای کوچک)، و الگوی مصرف خانواده (مصرف مرغ به جای استیک) روی دهد، تشدید خواهد شد. درست همین ارزیابی است که CPI برای کل جامعه، برای مناطق و شهرهای گوناگون، برای گروههای مختلف هزینه (خوراک، پوشاک، مسکن، حمل ونقل و امثال آن) به‌طور ماهانه ارائه می‌کند. این مقاله عمدتاً به مشارکت آمار در تهیه CPI می‌پردازد.

اگر مصرف‌کنندگان واقعاً می‌خواستند سابقه نسبتاً مشروحی را برای اندازه‌گیری میزان تغییر قیمت اقلامی که برای مثال از ژانویه ۱۹۸۶ تا ژانویه ۱۹۸۷ خریداری کرده‌اند، نگهداری نمایند، می‌توانستند به روش زیر عمل کنند: طی ژانویه ۱۹۸۶، باید سابقه‌ای از هر خرید نگه‌دارند و سپس این خریدها را برحسب مقدار (تعداد لیتر شیر، تعداد جفت کفش، تعداد دفعات اصلاح موی سر و امثال آن) و بهای یک واحد از هرقلم کالا خلاصه نمایند. درواقع، یک "سبدبازار" از کالاها و خدمات راتعریف و هزینه آن را در ژانویه ۱۹۸۶ به‌دست خواهند آورد. فرض کنید این هزینه ۷۵۰ دلاراست. در ژانویه ۱۹۸۷، دقیقاً همین سبد بازار از کالاها و خدمات قیمت‌گذاری می‌شود. فرض کنید که هزینه آن ۸۰۰ دلار است. بنابراین CPI مصرف‌کننده برای ژانویه ۱۹۸۷، با وزنهای (مقادیر) ژانویه ۱۹۸۶ و ژانویه ۱۹۸۶ به‌عنوان پایه (۱۰۰=) برابر با: $\frac{۸۰۰}{۷۵۰}\times ۱۰۰ = ۱۰۷$ است. CPI برهمین روال محاسبه می‌شود، ولی عملیات آن بسیار پیچیده‌تر است زیرا بجای یک مصرف‌کننده با جمعیت زیادی از مصرف‌کنندگان روبرو است.

حتی اگر فرض نماییم که وضعیت خانوادگی و تمایلات مصرف‌کنندگان از ژانویه ۱۹۸۶ تا ژانویه ۱۹۸۷ تغییر نکند ـ فرضی که به‌ندرت می‌تواند صحیح باشد ـ افراد مسئول تهیه و تفسیر این شاخص با مشکلات اجرایی متعدد و سئوالات آزاردهنده‌ای مواجه خواهندبود بویژه:

۱ـ حتی برای یک خانوار مصرف‌کننده، تعداد اقلام گوناگون خریداری‌شده طی یک ماه یا سال خیلی زیاد است و قیمت‌گذاری مستمر چنین فهرستی پرهزینه و وقت‌گیر خواهدبود. تهیه CPI برای گروههایی از خانوارها به‌وضوح این مشکل را تاحد غیرممکنی، بزرگتر می‌نماید.

۲ـ هر خانوار متوسط با تعدادی فروشنده کالا و خدمات روبروست و قیمت از یک فروشنده به فروشنده‌ای دیگر تغییر می‌کند. آیا هر خانوار در ژانویه ۱۹۸۷ باید سعی کند به همان مراکزی مراجعه نماید که در ژانویه ۱۹۸۶ مراجعه داشته‌است؟ در مورد فروشگاههایی که تعطیل و یا بتازگی تأسیس شده‌اند، چه بایدکرد؟

۳ـ قیمتها روزانه تغییر می‌کنند. آیا خانوار مصرف‌کننده موردنظر باید به همان فروشگاه و در همان زمان برای محاسبه هزینه سبد بازارش در ژانویه ۱۹۸۷ مراجعه نماید؟ چگونه باید این حقایق را که برخی از خانوارها از حراجیها خرید نمی‌کنند ولی برخی خرید می‌کنند و اینکه خرید برخی از خانوارها روزانه و خرید برخی با فاصله زمانی بیشتری است لحاظ کرد؟

۴ـ فرض کنید که خرید عمده‌ای، مانند خرید خودرو، به‌وسیله خانوار موردنظر در ژانویه ۱۹۸۶ صورت گرفته باشد. سبد ثابت بازار هنوز شامل این قلم در ژانویه ۱۹۸۷ است، درحالی که چنین خریدی در این زمان انجام نگرفته است. در این مورد چه باید کرد؟

۵ـ کالاها و خدمات دردسترس مصرف‌کننده از زمانی به زمان دیگر تغییر می‌یابند. اقلامی که در سبد اولیه بازار وجود داشتند، ممکن است هنگام مراجعه بعدی به فروشگاهها برای اطلاع از قیمت آنها یافت نشوند، یا فقط با کیفیت بهتر یا بدتری وجود داشته باشند. مفهوم تغییرات قیمت در این موارد چیست؟

۶ـ همچنین مشاهده می‌نماییم که خانوار مصرف‌کننده موردنظر ـ حتی اگر ترکیب، وضعیت و تمایلات خانوار تغییر نکرده‌باشد ـ ممکن است با تغییر مناسب الگوهای مصرفی خود قادر به "مقابله" با افزایش قیمت، باشد. به‌عنوان مثال فرض کنید، خانوار متوجه می‌شود که قیمت استیک بشدت افزایش یافته‌است و تمایل خانوار با مصرف مرغ ـ به‌جای مصرف استیک ـ تأمین می‌شود. اگر مرغ جانشین استیک شود، سطح ارضای خانوار تغییر نخواهد کرد، درحالی که هزینه صرف‌شده برای خوراک ممکن است برخلاف افزایش عمومی سطح قیمتها، کاهش نیز یابد. درنتیجه می‌توان به فکر جایگزینی یک سبد بازار با "ارضای ثابت" به‌جای سبد بازار با "محتوای ثابت" افتاد.

مسائل فوق هم به اندازه‌گیری و هم به مفهوم شاخص ارتباط دارد. بنابراین ممکن است متوجه شویم که کیفیت یک قلم کالای خاص تغییر کرده‌است ولی اندازه‌گیری این تغییر کیفیت و

تبدیل آن به ارزش دلاری ممکن است یکی از مشکلترین کارها باشد. دفتر آمارکار سعی می‌کند که این تغییرات کیفیت را در تهیه CPI دخالت دهد.

مورد ۶ در فوق از سطح مفهومی عمیقتری برخوردار است. بیشتر اقتصاددانان یک شاخص قیمت از نوع "با مطلوبیت ثابت" یا "ارضای ثابت" یا "رفاه" را بر شاخص قیمت از نوع "سبد ثابت بازار" که درحال حاضر به‌وسیله دفتر آمارکار تهیه می‌شود، ترجیح می‌دهند. به بیان دیگر یک خانوار با این دید که محتویات سبد بازار می‌تواند باتوجه به تغییر قیمتها، تغییر کالاها وامثال آن تغییر یابد، چه تغییری در هزینه‌هایش در طول زمان باید بدهد تا سطح مطلوبیت خویش را ثابت نگه‌دارد؟ تعدادی از پژوهشگران شاخصهای "مطلوبیت ثابت" را ساخته‌اند، ولی مشکلات مفهومی و عملی هنوز هم مانع ساخت چنین شاخصی به‌عنوان سریهای آماری ماهانه رسمی هستند. درحقیقت در سپتامبر ۱۹۴۵، عنوان رسمی شاخص از شاخص هزینه زندگی به شاخص بهای کالاها و خدمات مصرفی، جهت تأکید بر تفاوت بین این دو روش، تغییر کرد. در این مقاله با این مشکلات مفهومی کاری نخواهیم داشت، بلکه به‌جای آن به نقش آمار درکمک به رفع مشکلات اندازه‌گیری CPI توجه می‌کنیم.

نمونه‌گیری و CPI

بسیاری از مسائل فوق کنترل‌پذیر خواهندشد، چنانچه توجه از یک خانوار مصرف‌کننده به یک گروه یا جامعه‌ای از خانوارهای مصرف‌کننده معطوف شود. بنابراین گرچه برخی از اعضای جامعه در ژانویه ۱۹۸۶ به خرید خودرو دست زدند، دیگر اعضای جامعه چنین خریدی نداشتند و به‌همین ترتیب برای ژانویه ۱۹۸۷. بنابراین ترتیب خودرو را می‌توان با وزنی که بیانگر تأثیر تغییر میانگین قیمت آن (میانگین نسبت به خانوارهای جامعه) است در سبد بازار واردکرد. این تغییر تأکید از یک مصرف‌کننده به جامعه‌ای از مصرف‌کنندگان (که وجود اختلاف بین اعضای آن بدیهی است) به مفهوم آن است که تهیه CPI مستلزم **نمونه‌گیری** (و تجزیه و تحلیل) آماری از جامعه خانوارهای مصرف‌کننده است.

حرکت در این جهت همچنین انسان را ناگزیر از تفکر در زمینه **جامعه کالاهاو خدمات** قابل دسترس برای همه اعضای جامعه مصرف‌کننده و نیز در زمینه جامعه، منابعی که از محل آنها

این کالاهاو خدمات قابل خرید هستند، می‌نماید. به‌علاوه، آشکارا غیرممکن است که هر خانوار مصرف‌کننده را بررسی و هر قلم کالای خریداری‌شده را برحسب تمام منابع قیمت‌گذاری نمود. بنابراین لازم می‌شود که نمونه‌هایی از هریک از این جامعه‌ها انتخاب و نتایج به کل جامعه مربوطه تعمیم داده شود. روشهای آماری در این مراحل می‌توانند کمک نموده و راهگشا باشند.

برای شاخص بهای کالاها و خدمات مصرفی هرگز سعی نشده‌است که به اندازه‌گیری تغییر قیمت کالاها و خدمات برای همه خانوارها و افرادی که در ایالات متحده زندگی می‌کنند، توجه شود. بلکه، به‌دلیل استفاده سنتی از CPI در چانه‌زنیهای بین کارگران و مدیریت، دامنه‌اش به خانوارهای شهرنشین محدود شده‌است. تا ۱۹۷۸ جامعه خانوارهای مصرف‌کننده مشمول شاخص دربرگیرنده همه خانوارهای شهرنشینی (شامل کارگران مجرد) بود که ۵۰٪ یا بیشتر درآمد خانوار از مزدو حقوق حاصل از مشاغل دفتری به‌دست می‌آمد و حداقل یک عضو خانوار دست‌کم برای مدت ۳۷ هفته در سال شاغل بود. از ۱۹۷۸، دفتر آمار کار یک CPI برای همه مصرف‌کنندگان شهری (CPI - U)، علاوه بر ادامه محاسبه شاخص محدودتر برای مزدبگیران و کارگران دفتری شهری (CPI - W) یافته است.

به‌طور خلاصه، ابتدا نمونه‌ای از جوامع شهری انتخاب، و سپس نمونه‌ای از خانوارهای هر جامعه انتخاب می‌شود. در ۱۹۸۶ این کار در ۹۱ ناحیه اجرا شد. ۲۹ شهر بزرگ در فلات قاره‌ای ایالات متحده، به‌علاوه یک شهر در آلاسکا و یک شهر در هاوایی به‌طور خودکار در این نمونه گنجانیده شد. بقیه مناطق شهری باجمعیت ۲۵۰۰ نفر یا بیشتر در گروههای متجانسی باتوجه به اندازه و موقعیت جغرافیایی دسته‌بندی شدند و ۶۰ شهر از این گروهها، باتوجه به روشهای احتمالی انتخاب، برگزیده شدند. از این نمونه ۹۱ شهری هم به‌عنوان پایه‌ای برای بررسی الگوی هزینه‌ای خانوارهای مصرف‌کننده، وهم به‌عنوان پایه‌ای برای قیمتهای مورد استفاده در تعیین هزینه جاری سبد بازار کالاها و خدمات CPI باید مورد استفاده قرارگیرند تا تجدیدنظر عمده بعدی CPI همچنان بر این نمونه ۹۱ شهری استوار باشد.

در هریک از این ۹۱ شهر منتخب، نمونه‌ای از خانوارهای مصرف‌کننده انتخاب و مصاحبه شد. این خانوارها براساس اصول نظریه نمونه‌گیری آماری انتخاب شدند. هدف انتخاب

نمونه‌ها به طریقی بود که معیارهای عینی را برای تشخیص اندازه احتمالی انحرافات میانگین نمونه از میانگین کل جامعه (نامعلوم) ارائه نماید. این بررسی آماری خاص، بررسی هزینه مصرف‌کننده (CES) نامیده شد، زیرا هدف اولیه آن جمع‌آوری آمار در مورد هزینه‌های خانوار نسبت به کالاها و خدمات با مصرف روزانه بود.

CES درواقع دو بررسی جداگانه است. یکی بررسی پیمایشی مصاحبه‌ای در فواصل سه‌ماهه با هر واحد مصرف‌کننده در نمونه بود. این بررسی اطلاعات جامعی در مورد هزینه‌ها، داراییها، بدهیها و درآمدها جمع‌آوری می‌نماید و برای اندازه‌گیری هزینه خریدهای عمده و غیرمتداول مانند لوازم خانگی و خودرو و همین‌طور دیگر هزینه‌های خانوار مؤثر است. بررسی پیمایشی دوم براساس یادداشت روزانه‌ای بود که در آن هر واحد مصرف‌کننده، در نمونه‌ای دیگر، یک سابقه کامل از همه هزینه‌هایش را برای مدت ۲ هفته نگه می‌داشت. این روش برای اندازه‌گیری خریدهای کوچک نظیر همبرگر، ماهی، خمیردندان و فیلم عکاسی مؤثر است. هرسال، ۵۰۰۰ واحد مصرف‌کننده در هریک از دو بررسی CES بررسی می‌شوند. در ژانویه ۱۹۸۷، از اطلاعات CES برای سالهای ۱۹۸۴-۱۹۸۲ در CPI استفاده شد.

جمع‌آوری داده‌ها و CPI

داده‌های جمع‌آوری شده در مصاحبه‌های CES برای اهداف گوناگونی در تعیین ساختار فعلی CPI به‌کار می‌روند. داده‌ها جزئیات کامل و هزینه کل سبد بازار هر خانوار را طی سالهای مورد بررسی تعریف می‌نمایند. اقلام در پنج گروه عمده: خوراک و آشامیدنیها، مسکن، پوشاک و هزینه نگهداری، حمل ونقل، و مراقبتهای بهداشتی، تفریح و سایر اقلام طبقه‌بندی شده‌اند. سپس هریک از این گروهها به طبقات نهایی کوچکتری شامل ۶۹ طبقه هزینه تقسیم می‌شود. مثالهایی از این طبقات هزینه عبارتند از: گوشت گاو و گوساله، تخم‌مرغ، گاز لوله‌کشی و الکتریسیته، خدمات خانگی، کفش، تعمیرات و نگهداری خودرو، بیمارستان و خدمات مربوطه، و خواندنیها. به‌علاوه طبقات هزینه به ۲۰۷ قسمت تقسیم شده‌اند. یعنی، سبد بازار مصرف‌کننده به ۲۰۷ قسمت تقسیم شده‌است. سپس هزینه کل هر قسمت برای همه مصرف‌کنندگان نمونه که از یک شهر خاص انتخاب شده‌اند، تعیین و این هزینه‌های کل به

صورت نسبی نشان داده می‌شود؛ به‌عنوان مثال به قیمتهای دسامبر ۱۹۸۶، برآورد شده‌است که گوشت گاو و گوساله ۱/۰۱۳٪ هزینه کل سبد بازار در سطح کشور را به‌خود اختصاص می‌دهد. از سبدهای بازار گوناگونی در شهرهای مختلف استفاده می‌شود تا تنوع در مسخصه‌هایی ازقبیل آب و هوا و دسترسی به خوراکهای گوناگون لحاظ شوند. بنابراین به‌جای یک سبد بازار که به یک واحد مصرف‌کننده مربوط می‌شود به چند میانگین سبدهای بازار استفاده می‌نماییم.

همانگونه که قبلاً اشاره شد، مصاحبه‌های CES یک سبد بازار را تعریف می‌کنند که شامل همه اقلام خریداری‌شده توسط مصرف‌کنندگان است. غیرممکن است که این اقلام را باتوجه به تنوع بسیار آنها همه‌ماهه قیمت‌گذاری نمود، حتی باتوجه به اینکه این کار فقط باید در ۹۱ شهر شاخص انجام پذیرد. بنابراین مسئله نمونه‌گیری دیگری مطرح می‌شود. بازهم نظریه نمونه‌گیری آماری برای انتخاب نمونه‌ای از اقلام از هریک از طبقات ۲۰۷ کالایی برای انتخاب روش نمونه‌گیری این اجازه را می‌دهد که محتویات سبد بازار تغییر یابد. به‌عنوان مثال اگر یک قلم در بازار یافت نشود، جانشینی بانمونه‌گیری مجدد امکان‌پذیر خواهدبود. در هرحال ۲۰۷ نوع کالا و وزنهای اولیه آنها در طول زمان تا انتخاب سبد بازار جدید به‌جای آن ثابت می‌ماند.

انتخاب ۱۲۰۰۰۰ قلم کالا و مشخص نمودن منابع اطلاعی که قیمت از آنها جمع‌آوری خواهد شد، فرایند پیچیده‌ای شامل چندین مرحله است. اول اینکه بسیاری از این ۲۰۷ قلم یا قسمت (بلوکهای ساختاری ثابت شاخص) به "اقلام ورودی" یا ELIS[۱] تقسیم می‌شوند. به‌عنوان مثال قلم مربوط به قسمت تجهیزات شستشو درخانه به دو ELIS ماشین رختشویی و ماشین خشک‌کن تقسیم شده‌است. در هریک ازمناطق شهری که قیمتها جمع‌آوری می‌شوند، نمونه‌ای از ELIS در هر قسمت انتخاب می‌شود. حداقل یک ELIS از هر قسمت در هر منطقه شهری قیمت‌گذاری می‌شود. در تمام مناطق شهری، هر ELIS درجایی قیمت‌گذاری شده است که فراوانی قیمت‌گذاریش متناسب با اهمیتش در آن قسمت است. به‌عنوان مثال ماشینهای رختشویی ممکن است در پوکیپ سی[۲] و ماشینهای خشک‌کن در گراند آیلند[۳]

۱- *Entry - Level Items*

۲- *Poughkeepsie*، شهری در ایالت نیویورک

۳- *Grand Island*، شهری در ایالت نبراسکا

قیمت‌گذاری شوند، ولی در کل کشور سهم ماشینهای رختشویی وخشک‌کن در CPI تقریباً برابر سهم نسبی آنها در خرید همه مصرف‌کنندگان شهری است.

با داشتن نمونه ELIS در هرمنطقه شهری، یک نمونه علمی از فروشگاهها یا منابع اطلاع دیگر جهت قیمت‌گذاری هریک از ELIها انتخاب می‌نماییم. برای این منظور، دفتر آمار کار از بررسی مکان خرید (POPS)[۱] در هر منطقه شهری استفاده می‌نماید. این بررسی مانند CES است؛ از این نظر که در آن مشخص می‌نماید واحدهای مصرف‌کننده چقدر برای طبقات مختلف اقلام هزینه می‌نمایند و همین‌طور مشخص می‌نماید چقدر در هریک از محلها هزینه کرده‌اند. از این بررسی، دفتر چیزی را می‌سازد که آمارشناسان "چارچوب نمونه‌گیری" می‌نامند. برای هر طبقه قلم، چارچوب نمونه‌گیری شامل فهرستی از منابع اطلاعی است که طبقه قلم از آنها خریداری شده ونیز مبالغی که به‌وسیله همه واحدهای مصرف‌کننده نمونه در آنها هزینه شده است. در این صورت یک نمونه احتمالی از منابع اطلاع برای هر طبقه قلم انتخاب می‌شود، با این دید که نمونه نماینده همه انواع ومحلهای مختلف منابع اطلاع است که مصرف‌کنندگان در آن منطقه شهری اقلام را در طبقه خریداری می‌کنند.

اکنون آمارگیران دفتر می‌توانند به هریک از منابع اطلاع مراجعه نمایند، ولی مظنه از کدام ماشین لباسشویی خواهد بود؟ ممکن است دهها نام تجاری و مدل مختلف وجود داشته باشد. برای محاسبه CPI نمونه باید بیانگر همه انواع مختلف ماشینهای لباسشویی باشد. برای رسیدن به این هدف، کارکنان دفتر و فروشندگان فروشگاه حجم دلاری فروش هریک از انواع ماشینها را مرور کرده و سپس نمونه‌ای احتمالی از ماشینهای رختشویی فروخته‌شده در آن فروشگاه را انتخاب می‌نمایند. هرچه میزان فروش یک نوع ماشین بیشتر باشد، احتمال انتخاب آن در نمونه بیشتر خواهد بود، ولی هرنوع شانس برای انتخاب شدن دارد ـ هرچقدر ناچیز باشد ـ نتیجه نهایی اینکه نمونه انتخاب‌شده درسطح کشور برای هر طبقه قلم در CPI شامل انواع مختلف یک قلم است. انواعی که مصرف‌کنندگان به‌طور متوسط، برای آنها بیشتر خرج می‌کنند با فراوانی بیشتر درحالی که انواعی که برای آنها کمتر هزینه می‌شود با فراوانی کمتر در

۱- *Point of Purchase Survey*

فهرست نمونه ظاهر می‌شوند.

جمع‌آوری مظنه‌های ماهانه پیچیده است. نه‌تنها باید به تعداد زیادی آمارگیر آموزش داده شده و بر کار آنها نظارت شود، بلکه این آمارگیران باید همکاری مدیران فروشگاهها را جلب و فروشگاههایی راکه ظرف مدت ده سال بین دو تجدیدنظر عمده منحل ویا بتازگی شروع به‌کار نموده‌اند درنظر بگیرند. جهت ملاحظه عظمت کار، باید توجه نمود که مظنه‌ها هرماه از ۲۱۰۰۰ منبع اطلاع جمع‌آوری می‌شوند.

CPI یک مسئله نمونه‌گیری نهایی نیز دارد که همواره مشکل تلقی نمی‌شود. گرچه CPI ماهانه منتشر می‌شود، ولی به زمان در طول ماه اشاره نمی‌گردد. عمل مظنه‌گیری باید به‌طور مرتب انجام شود و بنابراین لازم است نمونه‌ای از زمان جمع‌آوری مظنه‌ها نیز انتخاب شود. این نمونه‌گیری به شکل رسمی مانند نمونه‌های دیگر انجام نمی‌شود. معهذا، سعی می‌شود که روزهای حراجی و غیرحراجی برای خوراک سهمهای مناسب خود را بنمایانند و درعین حال توازن مشابهی برای سهم اقلامی مانند روزنامه و ورودی سینما که بهای آنها به‌طور دوره‌ای تغییر می‌یابد، حفظ شود.

یک هدف مهم در طراحی وتجزیه وتحلیل آماری این است که معیاری عینی از دقت تحلیلهای نمونه داشته باشیم. گرچه این هدف در تنظیم پیچیده CPI هنوز به‌طور کامل تحقق نیافته ولی پیشرفت قابل ملاحظه‌ای به‌دست آمده است. تمام شواهد مؤید این هستند که عملیات نمونه‌گیری درحد قابل‌قبولی تحت کنترل است و نااطمینانی به مقدارشاخص به علت نمونه‌گیری، در مقایسه با نااطمینانیهای بوجودآمده در مراحل انجام کار، به‌عنوان مثال، تأثیر تغییرات کیفیت بر شاخص، به‌طور نسبی ناچیز است.

خلاصه

محاسبه مقادیر ماهانه شاخص بهای کالاها و خدمات مصرفی توسط دفتر آمار کار یک تعهد فوق‌العاده پیچیده است که مسائل **نظریه اساسی اقتصاد** (به‌عنوان مثال انتخاب بین یک شاخص قیمت یاشاخص ارضای ثابت)، **اندازه‌گیری و کمّی ساختن** (به‌عنوان مثال کمّی‌سازی تغییرات در کیفیت اقلام خریداری‌شده توسط مصرف‌کنندگان)، **نمونه‌گیری آماری** (تعریف و انتخاب

نمونه‌ها از انواع فراوان جوامع) و اجرا (به‌عنوان مثال آموزش آمارگیران و نظارت بر کار آنان) را شامل می‌شود.

CPI به جامعه خانوارهای مصرف‌کننده و به جامعه شهرهایی که این خانوارها در آنها زندگی می‌کنند، توجه می‌نماید. نمونه‌ای از شهرها دو هدف را برآورده می‌سازد: اول اینکه از شهرهای منتخب نمونه‌ای از مصرف‌کنندگان انتخاب می‌شود که براساس آن می‌توان الگوهای هزینه متوسط را مشخص نمود. دوم، مظنه‌ها در شهرهای منتخب به منظور تعیین مقدار فعلی CPI جمع‌آوری می‌شوند. بررسیهای پیمایشی هزینه مصرف‌کننده همچنین جامعه‌ای از کالاها و خدمات را تعریف می‌کند که درآمد خانوارهای مصرف‌کننده صرف آنها می‌شود. از این جامعه باید یک نمونه به‌منظور تعیین مظنه‌های جاری انتخاب شود. در هریک از شهرهای نمونه جوامعی از منابع اطلاع وجود دارند که اقلام از آنها خریداری می‌شود و نمونه‌هایی باید انتخاب شوند که معرف آنها باشند. بالاخره جامعه‌ای از زمان در هر ماه وجود دارد که درآن زمانها می‌توان مظنه‌های قیمت را به‌دست آورد و این جامعه نیز باید نمونه‌گیری شود.

کاملاً درست است که شاید همه از وضع فعلی CPI کاملاً راضی نباشند. در برخی از قسمتهای شاخص، مانند داده‌های اساسی درمورد الگوهای هزینه‌ای مصرف‌کننده، نمونه‌گیری از منابع اطلاع و جمع‌آوری قیمت از این منابع، تهیه شاخص برای انواع بیشتری از گروهها، و در روشهای مقابله با مشکل تغییر کیفیت، اصلاحاتی می‌تواند انجام گیرد و احتمالاً انجام خواهدگرفت. ممکن است حتی روش کاملاً جدیدی برای تهیه شاخص احتمالاً با استفاده از روشهای جدیدتر ریاضی ابداع شود. بدون شک CPI، صرف‌نظر از شکل آن به‌عنوان یکی از نماگرهای عمده وضعیت اقتصاد آمریکا باقی خواهد ماند.

مسائل

۱ـ با مراجعه به نمودار ۱، در سال ۱۹۸۰ یک خانوار متوسط شهرنشین چه مبلغ برای همان مقدار از کالاها و خدماتی که درسال ۱۹۶۷ با ۱۰ دلار می‌توانسته به‌دست آورد، پرداخت کرده است؟

۲ـ تفاوت بین یک شاخص با "مطلوبیت ثابت" و یک شاخص با "سبد بازار ثابت" را توضیح

دهید. ترجیح می‌دهید که دفتر آمار کار از کدام نوع استفاده نماید؟ چرا؟

۳ـ منظور از تقسیم‌بندی شهرها به گروههای متجانس قبل از انتخاب شهرهایی که باید مشمول بررسی قیمت شوند، چیست؟

۴ـ چرا شهرهای بزرگ به‌طور خودکار مشمول بررسی قیمت می‌شوند؟

۵ـ فرض کنید که گروهی از نمایندگان روستایی کنگره لایحه‌ای را تصویب کنند که درآن وزارت کشاورزی ملزم به تعیین هزینه زندگی خانوارهای روستایی شود:

الف) آیا وزارت کشاورزی می‌تواند آنان را به CPI ارجاع دهد؟ بلی یا خیر و چرا؟

ب) اگر نمی‌تواند چنین کند، وزارت کشاورزی چه اقداماتی را برای اجرای این قانون باید به‌عمل آورد؟

۶ـ فرض نمایید یک انسان‌شناس که سرگرم مطالعه دریک جامعه دورافتاده در آلبرتای شمالی است، تصمیم به تهیه شاخص برای این جامعه دارد. یک بررسی آماری در ژانویه ۱۹۷۵ نشان می‌دهد که هزینه جستجوگران به‌طور متوسط:

الف) ۱۰۰ اونس طلا برای آرد (که هزینه هرکیسه آن ۴ اونس است).

ب) ۲۰ اونس طلا برای الاغ که به‌طور متوسط ۱۰۰ اونس هزینه داشت.

ج) ۴۰ اونس طلا برای نوشیدنیها که هزینه هر بطری نیم لیتری آن ۵ اونس است.

در ژانویه ۱۹۷۶ این انسان‌شناس متوجه شد که هزینه هرکیسه آرد به ۶ اونس افزایش یافته است؛ الاغ به‌طور متوسط ۸۰ اونس فروخته می‌شود؛ و هر بطری نوشیدنی ۱۰ اونس هزینه دارد. CPI برای ژانویه ۱۹۷۶ (۱۰۰= ژانویه ۱۹۷۵) چقدر است؟

منابع

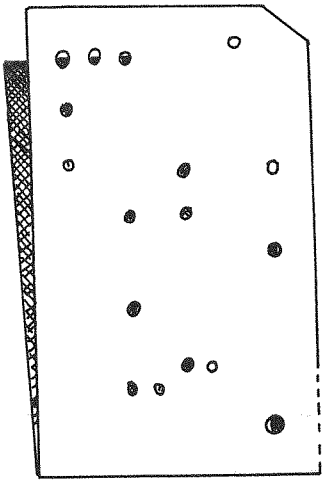
Ethel D. Hoover. "Index numbers II. Practical Applications." *In International Encyclopedia of the Social Sciences.* vol 7.

Philip J. Mc Carthy. 1968, "Index numbers III Sampling" *In International Encyclopedia of the Social Sciences.* vol 7.

Erik Ruist. 1968, "Index numbers I. Theoretical Aspects." *In International Encyclopedia of the*

Social Sciences. vol 7.

U.S. Department of Labor. Bureau of Labor Statistics. 1984, *BLS Handbook of Methods:* vol. II.



چگونه بهتر بشماریم؟
استفاده از آمار برای اصلاح سرشماری

موریس. اچ. هنسن[۱] شرکت وستت

باربارا. ای. بیلر[۲] دفتر سرشماری آمریکا

اهمیت نتایج سرشماری

در سال ۱۷۹۰ *تامس جفرسون*[۳] نتایج اولین سرشماری ایالات متحده آمریکا را به جورج *واشنگتن*[۴] اولین رئیس جمهور آمریکا ارائه داد. پس ازآن، هر ده سال یک بار، بنابر قانون،

۱ـ *Morris H. Hansen*

۲ـ *Barbara A. Bailar*

۳ـ *Thomas Jefferson*

۴ـ *George Washington*

سرشماری ده‌سالانه، اطلاعاتی اساسی درباره مردم برای ملت آمریکا فراهم کرده‌است.

البته هدف اصلی قانون سرشماری، تعیین سهم ایالتها در انتخاب اعضای مجلس نمایندگان بین ایالات است. سرشماری از آغاز اهدافی بسیارمهم، ورای هدف قانونی داشته‌است. گسترش برنامه‌های قانونگذاری برای بهبود وضع بهداشت و آموزش، کاستن از فقر، بهبود حمل ونقل و نظایر آنها به‌وسیله نتایج سرشماری هدایت شده‌اند.

این نتایج برای طراحی برنامه‌ها، اجرا وارزیابی آنها نیز به‌کار رفته‌اند. در حال حاضر، توزیع میلیاردها دلار در سال ازسوی حکومت فدرال برای ایالتها و از ایالتها به واحدهای محلی، دقیقاً مبتنی بر محاسبات سرشماری است. بازرگانی خصوصی برای اهدافی نظیر مکان تأسیس کارخانه و بازاریابی از نتایج سرشماری استفاده می‌کند. انجام اکثر پژوهشهای اجتماعی و اقتصادی، اساساً بدون استفاده از اطلاعات سرشماری، ناممکن هستند.

اهمیت سرشماری برای مسائل جاری نظیر فقر، بهداشت، آموزش، حقوق مدنی و نظایر آنها تقاضاهایی را مطرح می‌کند که برای به‌هنگام نگهداشتن اطلاعات باید سرشماری به‌جای ده سال یک بار به ۵ سال یک بار تغییر پیدا کند. این تقاضاها ازسوی فرمانداران ایالتی، شهرداران شهرهای بزرگ، دانشمندان، بازرگانان، و بسیاری دیگر مطرح می‌شود. در سال ۱۹۷۶ قانونی برای درخواست اجرای پنج سال یک بار سرشماری گذرانده‌شد اما کنگره هرگز بودجه لازم برای اجرای آن را تصویب نکرد.

کار برنامه‌ریزی و اجرای سرشماری

نیاز به سرشماری سریع و با دقتی معقول تقریباً آشکار است. آنچه که ممکن است خیلی واضح نباشد مسائل عمده درگیر در انجام سرشماری و در دسترس قراردادن فوری نتایج، وهماهنگ کردن سئوالات سرشماری برای تأمین نیازهای جاری است. در هر سرشماری، برخی سئوالها، برای پاسخ به نیازهای جدید تغییر یافته‌اند، ولی برخی اطلاعات پایه‌ای همواره درخواست شده‌اند. سئوالهای مربوط به سرشماری جمعیت ۱۹۸۰ شامل سن، جنس،نژاد، وضعیت تأهل، رابطه خانوادگی، آموزش، ثبت‌نام در مدرسه، اشتغال وبیکاری، نوع شغل و فعالیت، مهاجرت، مدت زمان لازم برای رسیدن به محل کار، تعداد افراد دارای خودرو که هرروز یکی از آنها بقیه

رابا خودروی خود به محل کار می‌برد، کشور زادگاه، *زبان تکلم فعلی*، *مهارت در زبان انگلیسی*، *اصل* و نسب، درآمد و موضوعهای دیگر بودند. (اقلام ایرانیک، در سرشماری ۱۹۸۰ اقلام جدید بودند.)

کار سازماندهی و اجرای سرشماری یک وظیفه اداری و فنی است. گرچه اکثر پرسشنامه‌ها را در یک "سرشماری به‌وسیله پست" خود پاسخگویان پر می‌کنند و کامپیوترهای مدرن و دیگر فناوریهای پیشرفته، حجم زیادی از کار با کاغذ را حذف می‌نمایند، با این حال انجام سرشماری مستلزم به‌کارگیری و کارآموزی صدها هزار نفر است که بیشتر آنها تنها چند هفته‌ای کار می‌کنند. مثلاً در سرشماری ۱۹۸۰، حدود ۴۵۸۰۰۰ نفر استخدام شدند که ۲۷۰۰۰۰ نفر طی دوره اوج عملیات فعال بودند. به‌محض اینکه هدفهای خاص، برحسب محتوای پرسشنامه و نتایج آماری مطلوب، تعیین شدند کار حجیم سازماندهی و اداری شروع می‌شود.

برای نظرخواهی، گردآوری، دریافت، پردازش و تلخیص تعداد زیادی از پرسشنامه‌های تکمیل‌شده باید سیستمی طراحی کرد. ماشینهای الکترونیک کاغذخور خاصی طراحی و در دفتر سرشماری نصب شده‌اند که به‌طور خودکار اطلاعات ثبت‌شده به‌وسیله پاسخگویان یا شمارشگران را می‌خوانند. این ماشینهای پیچیده ابتدا به‌صورت میکروفیلم عکس می‌گیرند، سپس آنها را بریده و نسخه‌های میکروفیلم فرمهای سرشماری را که در اکثر موارد پاسخگویان شخصاً پر کرده‌اند، می‌خوانند. درک حجم کار مشکل است. در سرشماری ۱۹۸۰، کمی بیش از ربع میلیارد صفحه (با احتساب هر طرف برگی نسبتاً بزرگ به‌عنوان صفحه) به‌کار رفت. بعد از بازبینی اداری، نتایج بر نوارهای مغناطیسی ثبت شدند و سپس کامپیوترها این نتایج را بررسی کردند. در پردازش، فرمها برای بعضی از انواع نواقص یا ناسازگاریها ویرایش شدند و تعدیلهایی به‌صورت خودکار صورت گرفت یا نواحی دارای مشکل خاص برای بررسیهای غیرماشینی بعدی شناسایی شد. گامهای بعدی جدول‌بندی و چاپ نتایج برای انتشار به‌وسیله کامپیوترهای الکترونیک برداشته شدند.

رهیافت کار اجرای سرشماری از رهیافت طراحی یک سیستم کاملاً جدید، از این لحاظ تفاوت دارد. چه، سرشماری قبلاً چندین بار اجرا شده و زمینه و تجربه گذشته برای راهنمایی تلاشهای جاری به‌کار می‌رود.

اما در دسترس بودن تجربه مبسوط گذشته، معایبی دارد و ممکن است رسوم دیرینه‌ای موجود باشد که به‌نظر اساسی بیاید، ولی در واقع، تنها معرف راههایی باشند که کار در گذشته به آن راهها ختم شده‌باشد. مثلاً رسم اجرای سرشماری به‌وسیله مأمور آمارگیری دریک ناحیه و پرسشگری شخصاً از هر عضو مسئول خانواده، مدتهاست که صورت می‌گیرد. این روش، به‌وسیله استفاده طولانی، ثابت شده بود که تنها راه معقول انجام سرشماری است. چون مفاهیم برخی از سئوالهای سرشماری، مشکل هستند، بر این عقیده بودند که تنها یک شمارشگر تعلیم یافته می‌تواند سئوالها را بپرسد و اطلاعات خوب را استخراج کند. اما این دیدگاه به مشکلات کارآموزی و کنترل لشگری از مصاحبه‌گران موقت توجه نداشت و تشخیص نمی‌داد. این دیدگاه تشخیص نداد که پاسخهای حاصل در موقعیتهای مصاحبه سرشماری گاهی اوقات برپایه درک نادرست از سئوالهاست و فرایند مصاحبه، وقت لازم برای اظهار جواب مورد نظر را نمی‌دهد. به‌علاوه، تصورات درست یا غلط مصاحبه‌گران می‌تواند اثر مهمی بر پاسخ داشته باشد.

در سده نوزدهم، بسیاری از مزایای بالقوه تجربه‌های قبلی از دست رفت، زیرا دفتر سرشماری تا سال ۱۹۰۲ به‌صورت یک مؤسسه دائمی و پایدار ایجاد نشده بود. لذا کار سرشماری با دارابودن کارمندان دائمی و استفاده از درسها و تجربه‌های سرشماریهای قبلی بسیار عملیتر می‌شد. وضعیت سرشماریهای سالهای ۱۹۰۰ و قبل از آن، به نقل قول زیر از تاریخچه آمار[۱] تشریح شده‌است.

آقای پورتر گزارش زیر را ازتجربه‌اش ارائه داد که باید اساساً گزارشی باشد که هر مدیر سرشماری ارائه می‌دهد.

> *در دو سرشماری گذشته (۱۸۸۰و ۱۸۹۰) مدیر سرشماری در آوریل سال قبل از سرشماری منصوب شده‌بود، ولی وقتی من منصوب شدم، تنها یک کارمند دفتری و یک نامه‌رسان و میزی با چند برگ کاغذ سفید روی آن، داشتم. من به ساختمان دفتر ثبت اختراعات پیام فرستادم تا آنچه می‌توانستم از بقایای ده سال قبل به‌دست آورم، و تعدادی کتب کهنه و*

۱ـ *Cummings, 1918,PP. 678-679*

پرسشنامه و چیزهایی نظیر آن از این کاوش به‌دست آمد.... من قادر نبودم بیش از سه پیرمرد از شهر...به‌کار بگیرم. بیشتر افراد سرشماری قدیم را می‌شناختم. بعضی از آنها مرده بودند و برخی در بخش بازرگانی خصوصی کار می‌کردند.... اما کمتر از ۶۰روز برای چاپ ۲۰ میلیون پرسشنامه و توزیع آنها، همراه با دستورات چاپی برای ۵۰ هزار شمارشگر در کل کشور باقی مانده‌بود، و با بسیاری از شمارشگران از طریق راه‌آهن یا خط تلگراف امکان تماس نبود. برای راهنمایی در به‌دست آوردن نسخه‌های سفید تنها چند آلبوم بریده جراید داشتیم که کسی از روی مآل‌اندیشی، داخل آنها فرمهای سفید سرشماری گذشته را نگهداشته بود. او هربار چند نسخه از فرمها را به خانه برده و آنها را داخل آن آلبوم قرار داده‌بود. دولت در سال ۱۸۸۵ وقتی دفتر بسته شده‌بود از این چیزها نگهداری نکرده‌بود. برخی از آنها را به صورت کاغذ باطله فروخته، بعضی را سوزانده و بقیه هم گمشده بودند.

این نقل قول، علاوه بر نشان دادن موفقیتهای بالقوه حاصل از پیوستگی و فراگیری از گذشته، بر پیچیدگی زیاد سرشماری در نیمه آخر قرن نوزدهم هم اشاره دارد. در آن زمان، در دولت فدرال کمتر برنامه آماری مداومی وجود داشت و پیامد آن، درگیری سرشماری ده‌سالانه با انبوهی از سئوالات شد که ثابت شده‌بود گردآوری پاسخها از طریق سرشماری ده‌سالانه اگر غیرممکن نباشد، مشکل است و تعداد زیاد و تنوع سئوالها و فرمها از این رو بود. بسیاری ازاین انواع اطلاعات، امروزه از راه بررسیهای نمونه‌ای یا تدوین گزارشهای ثبتی گردآوری می‌شوند.

استفاده از آمار در برنامه‌ریزی و اجرای سرشماری

در طول پنجاه سال گذشته، مفاهیم و روشهای آماری اصلاحاتی اساسی برای سرشماری فراهم کرده‌است. باید گفت که این اصلاحات یک پدیده ناگهانی فن‌شناختی را تشکیل می‌دهد. قسمتی از این تغییر ناگهانی ناشی از ورود کامپیوترهای الکترونیک بزرگ ـ مقیاس است، اما تغییر بیشتر، حاصل پیشرفتهای آماری و کاربرد بررسیهای آماری است. (درواقع، ظهور کامپیوتر

با سرعت زیاد و روشهای آماری مدرن هردو اساساً با انگیزه حل مسائل سرشماری صورت گرفتند و گسترش آنها با پشتیبانی دانشمندان دست‌اندرکار سرشماری یا دفتر سرشماری اجرا شده است.)

معرفی نمونه‌گیری به‌عنوان ابزاری برای اجرای سرشماری

نمونه‌گیری اولین بار در سرشماری ۱۹۴۰ در گردآوری اطلاعات سرشماری به‌کار رفت. دنباله‌ای از سئوالها برای یک نمونه ۵درصدی از جامعه اضافه شد. (به بیان نادقیق، این بدان معناست که برای یک نفر از هر بیست نفر، سئوالهای اضافی مطرح شد). این امر پیشرفتی عمده به‌حساب می‌آمد. زیرا پوشش عام برای هر سئوال طی یک سده و نیم به‌صورت سنت درآمده بود. سئوالاتی که در نمونه ۱۹۴۰ مطرح شدند، شامل یک سئوال درباره درآمد حاصل از مزد و حقوق (درآمد قبلاً یک سئوال سرشماری نبود) یک سئوال درباره شغل معمولی (شغل معمولی متمایز از شغل فعلی بود؛ سرشماری ۱۹۴۰ درطی دوره رکود برنامه‌ریزی شد، وقتی که نرخ بیکاری بسیار بالا بود و غالباً بین شغل معمولی شخص و شغلی که فعلاً با آن کار می‌کرد تفاوت وجود داشت)، و چند سئوال دیگر بود.

هنر و دانش نمونه‌گیری آماری مدرن، در زمان سرشماری ۱۹۴۰ در حال تکامل بود، و در همان زمان، پذیرش عموم از نمونه‌گیری افزایش یافته بود. امکان پیشروی با آگاهی و اطمینان درباره آنچه نمونه تولید می‌کند، بیشتر از آن که قبلاً میسر بود، فراهم شده‌بود. لذا، برای اندازه و طرح خاصی از نمونه، آمارشناسان می‌توانستند دامنه معقولی برای تفاوت بین نتیجه نمونه و نتیجه حاصل از سرشماری کامل به‌دست آورند. مثلاً فرض کنید که جمعیت کل شهری ۱۰۰۰۰۰ نفر بود که ۳۰۰۰۰ نفر آنها استخدام شده و دستمزد یا حقوقی داشتند و فرض کنید ۱۰۰۰۰ نفر آنها دستمزد و حقوقی کمتراز ۲۰۰۰ دلار در سال ۱۹۳۹ (سال قبل از سرشماری) داشتند. در نمونه ۵ درصدی این شهر، تقریباً ۵۰۰۰ نفر شرکت داشتند. برآورد تعداد دریافت‌کنندگان دستمزد و حقوق کمتر از ۲۰۰۰ دلار، وقتی برآورد از روی نمونه ۵٪ به‌دست آمد، با احتمال بسیار بالا در دامنه ۹۳۰۰ تا ۱۰۷۰۰ قرار داشت. این نوع دقت برای به‌کارگیری آن در بسیاری از هدفهای مهم کافی بود، درواقع این، همان اندازه دقتی بود که می‌شد در پرتو

دقتی کمتراز دقت کامل، از روی پاسخها به سئوال درآمد توجیه کرد. نه‌تنها آنچه سرشماری کامل نشان می‌داد، از روی نمونه برآورد می‌شد، بلکه علاوه بر آن، دامنه تفاوت محتمل بین برآورد حاصل از نمونه و نتیجه سرشماری کامل را می‌شد از روی نمونه برآورد کرد! نظریه نمونه‌گیری، هدایت‌کننده طراحی نمونه‌گیری برای دست‌یابی به دقت ماکسیمم نتایج، برحسب هزینه هر واحد نیز هست.

در بررسی مزایای استفاده از نمونه‌گیری، شاید به نظر بعضی چنین برسد که آنچه متضمن کار اصلی انجام سرشماری است، مدت زمان رفتن از دری به در دیگر است و وقتی سئوالاتی از خانوار می‌شود، هزینه افزایش سئوالها چه برای نمونه ۵درصدی باشد و چه برای کل جامعه، هزینه‌ای اندک است. چنین پیش فرضی دور از واقعیت است. فرض کنید مثلاً یک سئوال اضافی درباره اینکه شخص، بیماری مزمنی دارد یا نه، به‌طور متوسط برای هر شخص یک کار ۲۰ ثانیه‌ای را در سرشماری موجب شود. با یک جمعیت ۲۰۰ میلیونی، این سئوال، کاری بیش از یک میلیون ساعت را با هزینه شاید ۴ میلیون دلار در سرشماری به همراه دارد. بنابراین به‌دست آوردن اطلاع اضافی برای مثلاً ۵ سئوال، از تنها یک نمونه ۵درصدی به جای سئوال از همه افراد، می‌تواند نتایج کاملاً مفید مورد نیاز را با کسری از هزینه پوشش کامل تولید کند. سرانجام، استفاده از نمونه، جدول‌بندی و تحلیل را بسیار سریعتر از انجام آنها برای پوشش کامل، میسر می‌سازد.

بررسیهای نمونه‌ای، بادامنه وسیع آزمودنیها، در دهه ۱۹۴۰ شروع شدند، به قسمی که اطلاعات پیوسته و به‌هنگام بین سرشماریها دردسترس قرار گرفتند. مثلاً، آمارگیری جمعیت جاری، یک بررسی نمونه‌ای از جمعیت است که هرماه به‌وسیله دفتر سرشماری انجام می‌شود. این آمارگیری، هرماهه اطلاعات مربوط به اشتغال، بیکاری و دیگر مشخصه‌های نیروی کار و فعالیتهای جامعه را گردآوری می‌کند. این بررسی، برای جمع‌آوری اطلاعات درباره موضوعهای دیگر نیز با سئوالهای تکمیلی متفاوت در ماههای مختلف به‌کار می‌رود. در یک ماه هر سال، تقریباً همه سئوالهای سرشماری جمعیتی پرسیده می‌شوند. در ماههای دیگر، اطلاعات درخواستی ممکن است درباره فعالیت تفریحی، مسکن، معلولیت یا سایر موضوعها باشند. (برای بحث بیشتر آمارگیری جمعیت جاری، ن. ک. مقاله *لئون و رونس*.) آمارگیریهای

نمونه‌ای، بهداشت، خرده‌فروشی، خدمات بازرگانی و شخصی، فعالیتهای واحدهای دولتی و نظایر آنها را پوشش می‌دهد. این آمارگیریها، برای تهیه اطلاعات ملی، اطلاعاتی درباره مناطق کشور، و حتی اطلاعاتی برای ایالات بزرگ ونواحی مرکزی ایالات، نمونه‌های به‌قدرکافی بزرگ دارند. اما نمی‌توانند اطلاعاتی درباره بسیاری از شهرها و شهرستانها و برای اجتماعات نسبتاً کوچک داخل شهرها و شهرستانها فراهم کنند. برای به‌دست آوردن آن نوع از جزئیات ظریف، نمونه‌های بسیار بزرگ لازمند، نمونه‌هایی نظیر آنچه به‌عنوان بخشی از سرشماری ده‌سالانه گرفته می‌شود.

در سرشماری ۱۹۵۰، استفاده از نمونه‌گیری برای بعضی از سئوالاتی گسترش یافت که در سرشماریهای قبلی از همه افراد گردآوری شده بودند. برای این سئوالات از نمونه ۲۰٪ استفاده شد. این نمونه‌گیری در سرشماری ۱۹۶۰ به اکثر اقلام اطلاعات گسترش پیدا کرد. در سرشماریهای جمعیتی سالهای ۱۹۶۰، ۱۹۷۰ و ۱۹۸۰ تنها فهرست کردن جمعیت، با سئوالاتی درباره سن، جنس، نژاد، وضعیت تأهل و وابستگی خانوادگی بر پایه ۱۰۰٪ انجام شد.

این سئوالها غالباً مطرح شده‌اند: اگر نمونه‌گیری این‌قدر ابزار کارامدی است، چرا کل سرشماری با نمونه اجرا نمی‌شود؟ آیا سرشماری کامل، یک اتلاف تلاش نیست؟ باید به‌خاطر داشت که هدف اصلی سرشماری به‌دست آوردن اطلاعات ملی نیست، بلکه تهیه اطلاعاتی برای تک‌تک ایالتها، شهرها، شهرستانها و برای ناحیه‌های کوچک درون آنهاست. نتایج حاصل از تبدیل کل سرشماریها به یک نمونه نسبتاً بزرگ (شاید شامل ۲۰٪ جمعیت) برای برخی مقاصد کافی خواهد بود. اما چنین نمونه‌ای نمی‌تواند مانند سرشماری کامل نمایندگان را به ایالتها تخصیص دهد. همچنین تفاوتهای مهم در توزیع مبالغ زیاد سرمایه‌ها به هزاران نواحی کوچک می‌تواند وجود داشته باشد. دربعضی ایالتها وضعیت قانونی بسیاری از اجتماعات به اندازه دقیق میزان رسمی جمعیت بستگی دارد. مثلاً شهری با جمعیت ۱۰۰۰۰ نفر یابیشتر می‌تواند اوراق قرضه منتشر کند. بنابراین سرشماری کامل برای شمارش کل جمعیت وبرای برخی داده‌های پایه‌ای جمعیت مورد نیاز است اما توده بزرگ اطلاعات ممکن است از روی نمونه‌ای نسبتاً بزرگ گردآوری شود.

درآمد ـ مثالی از نمونه‌گیری در سرشماری

وقتی سرشماری ۱۹۴۰، بعد ازدوره طولانی رکود، طراحی شد، سئوالات مربوط به درآمد دستمزد و حقوق، برای راهنمایی مردم درباره برنامه‌های دولتی و بخش خصوصی، در پرسشنامه گنجانیده شدند. در سرشماری سئوال درآمد از یکایک افراد پرسیده شد، و طوفانی از اعتراض را برانگیخت. برخی مخالفتها در میان نمایندگان کنگره وجود داشت و یک سناتور نیوانگلند مبارزه‌ای را هدایت کرد تا مردم را قانع کند که به سئوال درآمد پاسخ ندهند. مطبوعات مردم را به مسئله حفظ حریم شخصی و محرمانه‌بودن موضوعها توجه دادند. مخالفت کنگره و سرمقاله روزنامه‌ها بر این موارد تهاجم به حریم خصوصی و افزایش مداخله دولت متمرکز شد. مقامات اجرایی پشتیبان این سئوال بودند. برای کاستن از ترس مردم درباره نداشتن حریم خصوصی، دفتر سرشماری ۲۰ میلیون فرم چاپ کرد که افراد می‌توانستند آن را پر کنند و مستقیماً با پست به واشنگتن بفرستند، به‌قسمی که هیچ شمارشگری نمی‌توانست پاسخ را ببیند. تنها ۲۰۰۰۰۰ نسخه از چنین فرمهایی مورد استفاده قرارگرفت و نرخ بی‌پاسخها کمتراز ۲٪ بود.

اما از سال ۱۹۴۰ به بعد، در پرسشنامه‌های سرشماری، سئوال در مورد درآمد برای نمونه‌ای از جمعیت به‌صورت یک روال درآمد. در سال ۱۹۵۰ سئوالات گسترش یافته مربوط به درآمد تنها در نمونه ۲۵٪ جمعیت پرسیده می‌شد. کار نمونه‌گیری برای سئوالات مربوط به درآمد ادامه یافت ولی بعد از سرشماری ۱۹۷۰ مشکلی را به‌وجود آورد. تسهیم درآمد در ۱۹۷۲ شروع شد و میلیاردها دلار سرمایه‌های فدرال را به دولتهای ایالتی و محلی برمبنای فرمولهایی که شامل برآوردهای درآمد محلی بود، تخصیص داد. برای نواحی کوچک، نمونه ۲۰درصدیی که در سال ۱۹۷۰ به‌کار رفت، تغییرپذیری بسیار زیادی را در خطای نمونه‌گیری بین کوچکترین و بزرگترین مکانها، که با یک نرخ نمونه‌گیری شده‌بودند، به‌دست داد. بنابراین در سال ۱۹۸۰، دو نرخ نمونه‌گیری به‌کار رفت. یکی نرخ یک به دو برای شهرستانهایی به‌کار رفت که شامل مکانها و تقسیمات شهری کوچک با جمعیت ۲۵۰۰نفر یا کمتر بودند. بقیه کشور با نرخ یک به شش نمونه‌گیری شد. این نمونه‌گیری متفاوت، نیازهای اطلاعاتی را برای نواحی کوچک تأمین کرد.

موارد استفاده از داده‌های درآمد آن‌قدر جامع بودند که اکنون درآمد یکی از مهمترین اقلام سرشماری به‌حساب می‌آید. بیش از ۲۰ برنامه فدرال وجود دارد که در آنها استفاده از داده‌های سرشماری مربوط به درآمد به‌صورت قانونی است.

استفاده از نمونه‌گیری و مطالعات تجربی در ارزیابی و اصلاح روشها و نتایج سرشماری

گامهای اساسی برای ارزیابی و اصلاح روشهای سرشماری که در دهه ۱۹۴۰ آغاز شد، ازآن پس ادامه یافته و بسیارگسترده شده‌است. برای جنبه‌های مختلف سرشماری، بررسیهای آماری انجام شده‌اند. یکی از این بررسیها، با تکرار شمارش سرشماری، در نمونه خوش طرح نواحی، کمی بعد از سرشماری اصلی، با استفاده از همان شیوه‌های سرشماری اولیه، انجام شده‌بود. بنابراین قادر بودیم ببینیم تا چه اندازه دو سرشماری که تحت شرایط و شیوه‌های یکسان انجام شده‌اند، به هم شبیه هستند.

بررسیهایی ازاین نوع، سازگاری زیاد و دقت پاسخ را برای سؤالاتی نظیر جنس، سن، نژاد، و محل تولد نشان دادند اما در پاسخهای سئوالات مشکلتر در ارتباط با شغل، بیکاری، درآمد، آموزش و غیره درجات بالاتری از ناسازگاری و عدم دقت را ارائه دادند. اطلاعات حاصل از چنین بررسیهایی، هم به اصلاح سئوالات در سرشماری بعد (مثلاً بانشان دادن اینکه کدام سئوالات موجب زحمت شده و نیاز به نگارش جدید دارند) و هم به تعبیر درستی نتایج سرشماری، وقتی سئوالات برای استفاده خاصی به‌کار می‌روند، کمک کرد.

بررسیهایی که روشها و شیوه‌های مختلف را در چارچوب آزمایشهای خوش طرح و تصادفی‌شده مقایسه می‌کنند، برای درک مؤثربودن شیوه‌های مختلف و مقایسه کردن هزینه و درستی آنها، از اهمیت زیادی برخوردار بوده‌اند. چنین مقایسه‌هایی، مثلاً برای انواع مختلف طرح پرسشنامه وتغییرات دیگر در شیوه‌ها انجام شده‌است. به‌عنوان مثال، در یک بررسی، پوشش سرشماری و سؤالات نمونه نواحی را از نو وارسی کردند، اما از شمارشگران با سطح کارآموزی بالا، مجموعه مفصلتر سؤالات و اصلاحات پرهزینه دیگری ازاین قبیل، استفاده شد. این بررسیهای نسبتاً متنوع، عموماً به این نتیجه‌گیری منجر شدند که برخی ازروشهایی که

سابقاً به‌عنوان راه حصول اصلاحات عمده مطرح بودند، در ارتباط با هزینه مؤثر نبوده‌اند (گرچه بعضی اصلاحات باارزش در طرح پرسشنامه و شیوه‌ها صورت گرفته‌بود). دریافتیم که مثلاً، صرف زمان و پول بسیار زیاد برای کارآموزی لشگری از شمارشگران موقتی، هزینه‌ها را زیاد می‌کند ولی احتمالاً به اصلاحات متناظر در درستی نتایج منجر نمی‌شود. یک مشکل در سرشماریهای ۱۹۷۰ و ۱۹۸۰ گردش کار شمارشگران بود. تعداد زیادی آموزش دیدند ولی برخی ازآنها قبل از شروع، کار را رها کردند و دیگران موارد آسان را انجام دادند و سپس کار را رها کردند. تجربه‌های پرداخت اضافی در سرشماریهای آزمایشی که منجر به سرشماری ۱۹۹۰ شد، دلگرم‌کننده بود. کارمندانی که به استانداردها دست یافتند به‌عنوان واجد شرایط برای پرداختهای اضافی معرفی شدند. در سرشماریهای آزمایشی افزایش نتایج و صرفه‌جوییهای حاصل از دستمزد شمارشگران بیشتر از میزان جبران هزینه پرداختهای اضافی بود. قبل از اخذ تصمیم نهایی در طرحهای پرداخت ۱۹۹۰، ادامه تجربه برنامه‌ریزی شد.

اثر شگفت‌آور شمارشگران

اما یک بررسی به نتایجی شگفت‌آور و سپس به اصلاح اساسی شیوه‌های سرشماری منجر شد. مدت مدیدی بود که می‌دانستند شمارشگران بر پاسخهایی که به‌دست می‌آورند، می‌توانند اثر بگذارند و احتمالاً بیشتر مواقع بدون قصد اثر می‌گذارند. اما میزان این اثرگذاری معلوم نبود. بنابراین به‌عنوان بخشی از سرشماری ۱۹۵۰، یک بررسی آماری وسیع برای اندازه‌گیری میزان اثر شمارشگر، اجرا شد.

طرح بررسی (باتوصیفی ساده) مبتنی بر تقسیم نواحی به شانزده تخصیص کاری و هشت شمارشگر بود (نواحی آن‌قدر کوچک بودند که یک شمارشگر می‌توانست دو تخصیص کاری را کاوش کند). دو تا از شانزده تخصیص کاری که به‌تصادف انتخاب شده‌بودند به اولین شمارشگر واگذار شد؛ دو تا از تخصیصهای باقیمانده هم، به‌تصادف انتخاب و به دومین شمارشگر داده شد، و قس‌علی‌هذا. البته کل آزمایش ۱۶ گانه به دفعات زیادی تکرار شد.

در اینجا، برای تعبیر نتایج به طریقی مفید، انتخاب تصادفی تخصیصهای کاری، جنبه اساسی داشت (انتخاب تصادفی به معنای انتخاب بااین روش است که اعداد ۱، ۲،....۱۶ را بر

کارتهای همانند بنویسیم، کارتهارا بُر زده یا کاملاً درهم بیامیزیم و سپس یکی پس از دیگری کارتها را با چشم بسته بیرون بکشیم.)

جنبه اساسی دیگر طرح، آن بود که هر شمارشگر انجام دوکار را به عهده داشت و تعدادی شمارشگر وجود داشتند. بدین طریق می‌شد برآوردهای خوبی از تغییرپذیری حاصل از کار هر شمارشگر (به‌صورت تقریبی، تفاوتهای عملکرد یک شمارشگر که نمی‌توان آنها را به تفاوتهای بین نواحی نسبت داد) درمقایسه باتغییرپذیری ناشی‌از تفاوتهای بین شمارشگران به‌دست‌آورد.

درباره این تحلیل شکسته مسیر نمی‌توان در اینجا جزئیات بیشتری ارائه داد، ولی می‌توان نتایج را خلاصه کرد. تفاوتهای بین شمارشگران خیلی بیشتر از آن بود که انتظار می‌رفت؛ در اقلامی نظیر سن و محل تولد نه چندان زیاد، اما درباره اقلام مشکلتر نظیر شغل، وضع اشتغال، درآمد و آموزش تفاوتها زیاد بودند. درواقع برای این اقلام، تغییرپذیری در نتایج یک سرشماری کامل (به دلیل اثر شمارشگر) به اندازه تغییرپذیری در نمونه ۲۵درصدی به شرط نبود اثر شمارشگر است.

در این‌باره چه باید کرد؟ یک رهیافت، ممکن است به‌کار بردن منابع بسیار بیشتر در انتخاب، کارآموزی و سرپرستی شمارشگران باشد. اما بررسیهای دیگری که در بالا به آنها اشاره شد، نشان داده‌اند که تحت شرایط سرشماری ملی این کار شدنی نیست.

پاسخ ممکن دیگر، حذف نیاز به شمارشگر، به‌وسیله مطرح کردن دقیق پرسشنامه‌ها با پاسخگویان و درخواست از آنهاست که پرسشنامه را پر کرده وبا پست بفرستند. (شمارشگران تنها وقتی لازم هستند که پرسشنامه‌های برگشتی، ناقص بوده یا پاسخگویان درخواست کمک کنند.) این روش در بررسیهای بیشتر آزمایش و معلوم شد که کاملاً خوب کار می‌کند. این روش در سرشماری ۱۹۶۰ برای فرمهای طولانیتر سرشماری به‌کار رفته بود و از لحاظ هزینه و افزایش دقت با موفقیت بزرگی همراه بود. بنابراین در ۱۹۷۰ و ۱۹۸۰ این روش باز بیشتر مورد استفاده قرار گرفت: بیشتر جمعیت فرمهای سرشماری را دریافت کردند و به‌وسیله پست بازگرداندند.

لذا، در فرایند استفاده از آمار برای اصلاح سرشماری، کامل بودن پوشش اصلاح شد و دقت اقلام اطلاعات گردآوری شده افزایش یافت.

خطای پوشش در سرشماری

از سال ۱۹۵۰ به بعد، دفتر سرشماری پوشش حاصل در سرشماری را اندازه‌گیری کرده‌است. نرخها خیلی بالا، بین ۹۸ تا ۹۹ درصدند. اما نرخها روی زیرگروههای جامعه تغییر می‌کنند. حدود ۵٪ از جمعیت سیاهپوست در سرشماری شرکت نکردند و در برخی گروههای سنی این نرخهابرای مردان سیاهپوست حدود ۲۰٪ است. اسپانیولی‌زبانها هم بانرخهایی پایین شمارش شدند.

بعد از سرشماری ۱۹۸۰، تعدادی از مناطق علیه دفتر سرشماری اقامه دعوی کردند. به دلیل تمرکز کم‌شماری درگروههای اقلیت، شهرهای دارای تعداد زیاد اقلیتها متقاعد شده‌بودند که کم به حساب آمده‌اند. آنها تعدیل نتایج سرشماری را درخواست کردند. در پرونده‌هایی که در محاکم قضایی مطرح شد، دفتر سرشماری استدلال کرد که گرچه روشهای آماریی وجود دارند که به‌قدر کافی در اندازه‌گیری کم‌شماری برای اهداف ارزیابی دقیقند، روشها به حد کافی برای تعدیل شمارش دقیق نیستند. آمارشناسان دیگر، که نمایندگی حوزه‌های قضایی آسیب‌دیده را برعهده داشتند، می‌گفتند که توزیعهای جمعیتی به‌وسیله تعدیلها اصلاح می‌شوند. درباره یکی از دعاوی حقوقی ـ که در ایالت و شهر نیویورک در سال ۱۹۸۰ تنظیم شده‌بود، پس از هفت سال تصمیم‌گیری شد. در ۸ دسامبر ۱۹۸۷، قاضی فدرال، *جان اسپریزو* نظری رابه نفع دولت انتشار داد. اما در زمان نگارش این مقاله، دعاوی بسیار دیگری منتظر اخذ تصمیم بوده‌اند.

دفتر سرشماری برنامه پژوهش متمرکز و فشرده‌ای را بر روشهای اندازه‌گیری کم‌شماری و توزیع آن به همه سطوح جغرافیایی متمرکز کرد. در سرشماریهای آزمایشی، تظاهراتی چند رخ داد. اما در ۳۰ اکتبر ۱۹۸۷، وزارت بازرگانی اعلام کرد که علیه انجام هر تعدیلی برای کم‌شماری در ۱۹۹۰، تصمیم گرفته است.

مسائل

۱ـ چند دلیل ارائه دهید که چرا سرشماری می‌کنند؟

۲ـ چرا در بعضی از موارد، گرفتن نمونه ۵٪ در مقابل سرشماری کامل، مطلوب است؟ در نمونه

۵٪ چه چیزی از دست می‌رود؟

۳ـ مزایا و معایب استفاده از شمارشگران را در یک سرشماری توضیح دهید.

۴ـ به‌نظر می‌رسد که پاسخهای سئوالات مربوط به جنس، سن، نژاد، و محل تولد قابل اعتمادتر از پاسخ به سئوالات شغل، بیکاری، درآمد و آموزش است. دو دلیل ممکن برای این مطلب ارائه دهید. در مورد این سؤالات فکر می‌کنید پاسخ دادن به پرسشنامه‌های پستی دقیقتر از مصاحبه شمارشگران است؟ چرا آری؟ چرانه؟

۵ـ (برای کار گروهی) نمونه تصادفی کوچکی از مردم بگیرید و پرسشنامه‌ای برای یک موضوع، مثلاً نظرخواهی درباره روزنامه‌های محلی ترتیب دهید. درباره مشکلات و وضعیتهایی که پیش می‌آیند، گزارش دهید.

۶ـ فرض کنید دفتر سرشماری تصمیم به افزودن یک مجموعه شش سئوالی درباره مصرف انرژی به پرسشنامه سرشماری ۱۹۹۰ دارد. دفتر سرشماری، اگر به‌جای پرسیدن سئوالها از هر فرد، تنها آنها را از نمونه ۱درصدی بپرسد چه مبلغی صرفه‌جویی می‌کند؟

۷ـ فرض کنید دفتر سرشماری به‌جای شمارشگران از پرسشنامه‌های پستی استفاده کند، و ۴۰٪ پرسشنامه‌های برگشتی سئوالی درباره سلامت روانی را علامت نزده باشند. آیا دفتر سرشماری توجیهی دارد که تنها پاسخهای علامت‌زده را محاسبه و نتایج را به‌صورت نمونه‌ای ۶۰درصدی منتشر کند؟ پاسختان را توجیه کنید.

۸ـ متخصصی در دفتر سرشماری استفاده از مصاحبه تلفنی را به‌جای پرسشنامه‌های پستی برای خانواده‌هایی‌که تلفن‌دارند، توصیه‌و فکر می‌کند که این‌روش باید به‌سرشماری دقیقتری منجر شود.

الف) چه استدلالهایی برای تأیید یا رد این ایده دارید؟

ب) برای تعیین اینکه این ایده صحیح است آزمایشی را طراحی کنید.

ج)چه ملاحظات دیگری (علاوه‌بردرستی) تعیین خواهندکرد که‌روش تلفنی رابابد پذیرفت یانه

منابع

Barbara A. Bailar. 1983, "Counting or Estimation in a Census - A Difficult Decision. "*1983*

Proceedings of the Social Statistics Section of the American Statistical Association. Washington. D. C. American Statistical Association, pp 42-49.

Barbara A. Bailar. 1985, "Comment." *Journal of the American StatisticalAssociation* 180: 109-114.

John Cummings. 1918, "Statistical Work of the Federal Government of the United States." *The History of Statistics.* John Koren. ed. New York: Macmillan. pp. 678-679.

Morris H. Hansen. 1987, "Some History and Reminiscences on Survey Sampling." *Statistical Science* 2(2): 180-190.

Morris H. Hansen. Leon Pritzker and Joseph Steinberg. 1959, "The Evaluation and Research Program of the 1960 Censuses." *1959 Proceeding of the Social Statistics Section. American Statistical Association.* Washington. D. C.: American Statistical Association.

Morris H. Hansen. and Benjamin J. Tepping. 1969, "Progress and Problems in Survey Methods and Theory Illustrated by the Work of the U. S. Bureau of the Census. "*In New Developments in Survey Sampling.*

Ingram Olkin. 1987, "A Conversation With Morris H. Hansen. "*Statistical Science* 2(2): 162-179.

Joseph Waksberg and Robert B. Pearl. 1965, "*New Methodological Research on Labor Force Measurements.*" *1965 Proceedings of the Social Statistical Section, American Statistical Association,* Washington. D. C. American Statistical Association, pp. 227-237.



برآوردهای اشتغال و بیکاری کشور چگونه ساخته می‌شوند؟

کارول بوید لئون (۱)
فیلیپ ال. رونس (۲) دفتر آمار نیروی کار آمریکا، اداره آمارهای اشتغال و بیکاری

هر ماه، همانند کار ساعت، روزنامه‌های سراسر کشور عناوینی مثل "اشتغال به صعود ادامه می‌دهد"، "بیکاری بشدت سقوط می‌کند" یا "نرخ بیکاری به بالاترین میزان خود می‌رسد" را چاپ می‌کنند. حتی "اشتغال و بیکاری ایالات متحده تغییری نکرده‌است" نیز خبرساز است.

۱- *Carol Boyd Leon* که قبلاً از اقتصاددانان همکار دفتر آمار نیروی کار بوده هم‌اکنون یک نویسنده مستقل در منطقه واشنگتن دی.سی. است.

۲- *Philip L.Rones*

این مقالات مبتنی بر گزارش ماهانه دفتر آمارهای نیروی کار (B.L.S) در مورد وضعیت اشتغال است. این گزارشها توجه بسیاری از سیاستگذاران، محققین، تحلیل‌گران بازرگانی، رسانه‌های گروهی، و عامه مردم را به خود جلب می‌کند؛ زیرا شرایط بازار کار کشور بر همگان تأثیر می‌گذارد.

اخبار در ساعت ۸/۳۰ صبح (به وقت شرق آمریکا) اولین جمعه هرماه همزمان با انتشار اطلاعیه خبری "وضعیت اشتغال" به‌وسیله B.L.S پخش می‌شود. برای باقیمانده روز، یکی از ۲۰ اقتصاددان دفتر به تهیه داده‌ها و کاوش درون اعداد تازه منتشرشده مشغول می‌شود. درخواستها از مطبوعات، مؤسسات مالی، سفارتخانه‌های خارجی، و تعداد زیادی منابع دیگر می‌رسد. اندکی بعد در همان صبح، رییس دفتر آمارهای نیروی کار در کمیته مشترک اقتصادی کنگره حاضر می‌شود تا درباره وقایع بیشتر توضیح داده و درحالی که دوربینهای تلویزیونی تصویربرداری می‌کنند و خبرنگاران یادداشت برمی‌دارند به سئوالات آنها پاسخ دهد.

تمامی اینها بعد از یک فرایند پیچیده جمع‌آوری داده‌ها، بررسی پاسخ مصاحبه شوندگان، انتقال اطلاعات به نوارهای کامپیوتری، ایجاد جداول داده‌های تعدیل‌شده برای نوسانها فصلی، و تحلیل اعداد برای درک جهشها و روندها اتفاق می‌افتد. این فرایند، از گردآوری داده‌ها تا پخش در اخبار صبح، کمتر از سه هفته طول می‌کشد.

برای به‌دست آوردن داده‌ها، دفتر هر ماه دو بررسی پیمایشی جداگانه اجرا می‌کند. بررسی پیمایشی اشتغال و بیکاری به ۱۹۴۰ و بررسی پیمایشی برآورد مشاغل مزد و حقوق‌بگیران برای بعضی صنایع که به ۱۹۱۳ برمی‌گردد. مردم ممکن است با نتایج آشنایی داشته باشند، لکن چگونگی ساخت این برآوردها برای بیشتر آمریکاییها یک راز است.

طرح جمعیت جاری

ممکن است یک آمارگیر برای جمع‌آوری داده‌های ماهانه اشتغال و بیکاری به خانه شما مراجعه نکرده‌باشد، زیرا دفتر سرشماری که برای دفتر آمارهای نیروی کار آمارگیری می‌کند، در هر ماه با کمتر از ۶۰۰۰۰ خانوار مصاحبه می‌کند. تعجب‌آور نیست که بیشتر آمریکاییها از کوششهای عمده‌ای که از طریق طرح جمعیت جاری (ط.ج.ج.) به عمل می‌آید تا نسبت به قابل

اعتماد و به‌هنگام بودن آمار جمع آوری شده در طول سال اطمینان حاصل شود، آگاه باشند. هرماه (در طول هفته‌ای که شامل روز نوزدهم ماه هم می‌شود) آمارگیران با یک نمونه به‌طور علمی انتخاب شده‌ای از خانوارها مصاحبه می‌کنند تا بدانند که چند نفر در آن زندگی می‌کنند و چند نفر آنها در طول هفته قبل که هفته مرجع نامیده می‌شود (هفته شامل روز دوازدهم) جزو نیروی کار بوده‌اند؟ (یعنی یا کار می‌کرده‌اند یا در جستجوی کار بوده‌اند). اولین سئوالی که آمارگیر می‌پرسد (بعد از تهیه یا تأیید یک فهرست از اعضای خانوار) این است که "بیشتر طول هفته گذشته را چه‌کار می‌کردید؟ کار می‌کردید یا چیز دیگر؟" اگر پاسخ "چیز دیگر" باشد، آمارگیر ادامه می‌دهد؛ "آیا در تمام طول هفته قبل، بدون درنظر گرفتن کارهای خانه، اصلاً کار می‌کردید؟" هر کاری که به ازای آن حقوق یا سودی پرداخت شود، فرد به‌خاطر آن شاغل قلمداد می‌شود. اگر در پاسخ بیکار گزارش شود، پرسشها وارد دور دیگری می‌شوند: "آیا در طول ۴ هفته گذشته در جستجوی کار بوده‌اید"؟ "برای یافتن کار چه کرده‌اید؟" اگر پاسخگو برای یافتن کار کوششی کرده‌باشد، بیکار به‌شمار می‌آید. اگر جستجوی کاری در میان نبوده باشد، ط.ج.ج. پاسخگو را به‌عنوان فردی که نیروی کار نیست، رده‌بندی می‌کند. آمارگیران هیچگاه مستقیماً سئوال نمی‌کنند که "شما شاغلید یا بیکار؟" در عوض، پاسخ به سئوالات فوق فرد را بر اساس آنچه انجام داده است و نه بر اساس اینکه خود را در کدام طبقه به‌حساب می‌آورد، رده‌بندی می‌کند. رده‌بندی نیروی کار بیشتر بر ملاکهای عینی مبتنی است تا بر پاسخهای ذهنی.

وقتی که آمارگیر وضعیت پایه‌ای پاسخگو را دریافت، برای اطلاع بیشتر درباره هر فرد سئوالات دیگری می‌پرسد. برای مثال، از افراد شاغل درباره ساعاتی که کار کرده‌اند، نوع شغلشان و نوع کار یا صنعتی که در آن به‌کار مشغولند سئوال می‌شود. از جویندگان کار مدتی که در جستجوی کار بوده‌اند و اینکه آیا کاریابی را به‌دلیل از دست دادن کار یا اخراج از کار یا بعضی دلایل دیگر آغاز کرده‌اند، سئوال می‌شود. افرادی که در نیروی کار نبوده‌اند، درباره علاقه فعلیشان برای یک شغل و هدفشان از جستجوی کار در آینده پرسیده می‌شود.

علاوه بر شمارش شاغلان، بیکاران، و کسانی که جزو نیروی کار نیستند، بررسی پیمایشی چندین نرخ بیکاری را برای گروههای سنی نژادی ـ جنسی خاص، نسبتهای اشتغال به جمعیت،

نرخهای مشارکت نیروی کار، و هزاران اطلاع دیگر که بیان‌کننده تعداد کسانی که کار می‌کنند، چه‌کاری می‌کنند، و چه تعداد در جستجوی کار هستند، را برای B.L.S فراهم می‌کند.

برآوردهای اشتغال و بیکاری چقدر قابل اعتمادند؟

اگرچه نسبت بسیار کوچکی از خانوارهای کشور در هر ماه مورد مصاحبه قرار می‌گیرند، ولی نمونه انتخاب شده برآوردهای منطقاً درستی را برای تمامی جمعیت ۱۶ ساله و بالاتر کشور ارائه می‌دهد. درستی داده‌ها را می‌توان با گسترش زیاد نمونه افزایش داد (که حداکثر به اجرای یک سرشماری ماهانه از کل جمعیت شبیه آنچه هر ۱۰ سال یک‌بار انجام می‌شود، منتهی می‌گردد). به‌هرحال هزینه انجام چنین کاری (بر حسب دلار یا باری که بر پاسخ‌دهندگان تحمیل می‌شود) گران است. سرشماری سال ۱۹۹۰ احتمالاً فقط برای عملیات میدانی، بیش از ۱ میلیارد دلار هزینه خواهدداشت.

اگرچه بررسیهای پیمایشی خانوار به‌منظور اطمینان از قابلیت اعتماد مطلوب، به اندازه کافی بزرگ هستند، آمارهای آن ممکن است با آنچه در سرشماری به‌دست می‌آید، متفاوت باشد. خوشبختانه بعضی از برآوردهای درستی ط.ج.ج. را در اختیار داریم.

به‌عنوان مثال، نرخ بیکاری کل معمولاً تا حدود دو دهم درصد، صحیح است. یعنی اگر نرخ "حقیقی" برای جامعه ۷٪ باشد، آنگاه ۹۰٪ احتمال دارد که نتیجه بررسی در فاصله ۶/۸٪ تا ۷/۲٪ بیفتد. در مقایسه، نرخ بیکاری جوانان، که گروه بسیار کوچکتری است، تا حدود یک درصد، درست است. به‌دلیل آنکه نرخ بیکاری هر ماه آماری در بردارنده خطاست، تفاضل بین دو ماه نیز همین‌طور است. برای آنکه تغییر ماهانه نرخ بیکاری از نظر آماری معنی‌دار باشد، باید بزرگتر یا مساوی ۰/۲ درصد باشد.

شیوه نمونه‌گیری و برآورد چگونه است؟

خانوارها از ۷۲۹ ناحیه نمونه که در میان ۱۹۷۳ شهرستان یا شهر مستقل پخش شده‌اند، انتخاب می‌شوند. این نواحی نمونه (که در هر ایالت به‌علاوه بخش کلمبیا واقع شده‌اند) آدرس در شهرهای بزرگ و کوچک، حومه شهرها، و نواحی روستایی را در برمی‌گیرد. در تبدیل یافته‌های

طرح به برآوردهای ملی، هر فردی که با وی مصاحبه شده وزنی معادل تعداد افرادی که او نماینده آنهاست، می‌گیرد. هر فرد تقریباً نماینده ۱۸۰۰ نفر از افراد کشور است. همچنین به‌دلیل آنکه خصوصیات افرادی که برای طرح انتخاب شده‌اند، ممکن است با خصوصیات کل جامعه فرق داشته باشد، برآوردهای طرح برای تطبیق با مجموعهای حاصل از سرشماری که بیانگر خصوصیات جامعه برحسب سن، جنس، نژاد و سایر عوامل هستند، تعدیل می‌شود.

خانوار ساکن در هر آدرس در نمونه برای ۴ ماه متوالی مورد مصاحبه قرار می‌گیرد. سپس برای ۸ ماه بعد حذف شده، و آنگاه مجدداً برای ۴ ماه مشابه سال بعد (قبل از حذف کامل از نمونه) مورد مصاحبه قرار می‌گیرد. این بدین معناست که سه چهارم آدرسهای نمونه از یک ماه به ماه دیگر و نیمی از آنها از یک سال به سال دیگر تغییر نمی‌کنند. بنابر این دو هدف حاصل می‌شود. الگوی ۴ـ۸ـ۴ سازگاری قابل ملاحظه‌ای در نمونه را تضمین می‌کند و به هیچ خانواری به‌خاطر مشارکت در طرح فشار زیادی نمی‌آید. (یک مصاحبه ط.ج.ج. به‌عنوان نمونه فقط ۱۵ یا ۲۰ دقیقه طول می‌کشد.)

پس از آنکه دفتر سرشماری داده‌های ط.ج.ج. را جمع‌آوری و جدول‌بندی کرد، آنها را به دفتر آمار نیروی کار، که مسئول تحلیل و انتشار داده‌های بازار کار است، منتقل می‌کند. ط.ج.ج. به ما چه می‌گوید؟ در میان خیلی چیزها، ما می‌توانیم نرخ بیکاری جوانان سیاهپوست، نرخ مشارکت نیروی کار برای زنان متأهل سنین ۲۵ تا ۳۴ سال، نسبت اشتغال به جمعیت برای مردان سنین ۶۵ تا ۶۹ سال، نرخ بیکاری کارگران صنایع خودروسازی، و تعداد افرادی که بیکار فصلی مشاغل ساختمانی هستند را بدانیم. ما همچنین می‌توانیم چیزهایی مثل تعداد افرادی که در نیروی کار نیستند، لکن خواستار یک شغل هستند و در عین حال جویای کار نیستند به‌دلیل آنکه آنها فکر می‌کنند نمی‌توانند کار پیدا کنند (که کارکنان مأیوس نامیده می‌شوند)، میانه تعداد هفته‌هایی که افراد اسپانیایی‌الاصل بیکار بوده‌اند، درآمد زنانیکه در مشاغل خدماتی کار می‌کنند، و تعداد خانواده‌های با فرزندی که عضو شاغل ندارند را بدانیم. در واقع، ط.ج.ج. ثروتی از اطلاعات درباره نیروی کار را فراهم می‌کند. به‌علاوه، در برخی ماهها سئوالات خاصی ضمیمه پرسشنامه می‌شوند. به‌عنوان مثال، در هر ماه مارس، ط.ج.ج. درباره کار در طول یک سال تقویمی گذشته سئوال می‌کند. سایر پیوستهای ط.ج.ج. کوششهایی هستند که برای

به‌دست آوردن داده روی موضوعات جاری، مثل جابجایی کارکنان یا آموزشهای شغلی، فقط یک‌بار انجام می‌شوند.

سیاست در طول زمان تغییر می‌کند، و این امر تهیه آمارهای خاص مثل داده‌های مربوط به نظامیان سابق در جنگ ویتنام، کارکنان مأیوس، یا مقایسه درآمد مردان و زنان را ایجاب می‌نماید. در هر حال، تعاریف و مفاهیم اساسی به‌کار رفته در ط.ج.ج. اساساً یکسان می‌مانند. در عین حال پافشاری حریصانه‌ای وجود دارد که آمارها به‌صورتی عینی و واقع‌بینانه ارائه شود به‌طوریکه یک نقطه‌نظر سیاسی را نرساند. همه افراد در دفتر آمار نیروی کار، از تازه‌کارترین تحلیل‌گر گرفته تا رییس دفتر، اعتقاد دارند که بیشترین خدمت، ارائه مستقیم داده‌هایی است که به سیاستگذاران امکان تصمیم‌گیری بر مبنای اطلاعات را می‌دهد. اگرچه افراد آگاه اغلب نسبت به معنی ارقام توافق ندارند، لکن بیشتر احزاب از نقش دفتر به‌عنوان گزارشگر بیطرفی که از فشارهای سیاسی بیرون آزاد است، تقدیر می‌کنند.

تعجب‌آور است که انتشار آمارهای نسبتاً خشک می‌تواند منبع تناقضات بسیار زیادی باشد. همه ما مطمئنیم که بیکاری، بد و اشتغال، خوب است. آیا انتشار اطلاعات به‌خودی خود برای نمایش آنچه درباره بازار کار، درست یا غلط است مفید نیست؟ حقیقتاً خیر. به‌عنوان مثال، اقتصاددانان بعضی سطوح بیکاری را که نتیجه ورود یا دوباره وارد شدن مردم به نیروی کار یا ترک داوطلبانه یک شغل به‌منظور جستجوی کاری دیگر است، طبیعی و حتی سالم می‌دانند. لکن متخصصین در مورد اینکه این سطح سلامت چقدر بالاست؟ توافق ندارند.

مطمئناً، بیکاری در طول دوران رکود بزرگ در سطح بالای تکاندهنده‌ای مانده بود، و خیلیها به‌دلیل آن مشقت طاقت‌فرسایی را متحمل شدند. حتی در زمانهای نسبتاً خوب، بعضی خانواده‌ها خانه خود را به‌دلیل آنکه از دست دادن یک شغل به معنای عدم توانایی در پرداخت وام است، از دست می‌دهند. لکن هر کارگر بیکاری به میزان یکسانی صدمه نمی‌بیند. به‌عنوان نمونه، دانشجویی از یک خانواده مرفه که به‌طور غیرجدی به‌منظور فراهم کردن پول توجیبی خود در جستجوی کار است، درست مثل والدین مجردی که کاری را که با آن اجاره‌خانه را می‌پرداخته‌اند از دست داده‌اند، بیکار محسوب می‌شود. آمارها به تنهایی نمی‌توانند پاسخگوی سئوالات درباره سیاستها باشند، لکن آنها واقعیتهای اساسی را ارائه می‌دهند که قبل از انجام

مذاکرات مفید پارلمانی، لازمند.

تعریف رسمی اشتغال و بیکاری چیست؟

از زمان شروع ط.ج.ج.، تعاریف نیروی کار اصلاح شده‌اند، لکن تغییر زیادی نکرده‌اند. از ۱۹۶۷ به بعد تعاریف اساسی زیر مورد استفاده قرار گرفته‌اند:

● **شاغل** به کسانی گفته می‌شود که هر کاری را در قبال مزد یا سود در طول هفته طرح انجام می‌داده‌اند یا کسانی که حداقل ۱۵ ساعت به‌عنوان کارگر بدون مزد در کسب وکار متعلق به خانواده کار می‌کرده‌اند. کسانی که به‌دلیل بیماری، مرخصی، یا سایر دلایل از این نوع، به‌طور موقت از کار غایب می‌شوند نیز شاغل به حساب می‌آیند.

● **بیکار** به کسانی اتلاق می‌شود که در طول هفته مورد بررسی هیچ شغلی نداشته، برای کار آماده بوده، و در طول چهار هفته قبل برای یافتن کار کوششهای خاصی کرده‌اند. این کوششها ممکن است رسمی بوده‌باشند، مثل ثبت‌نام در یک مؤسسه کاریابی یا مصاحبه برای یک شغل خالی، یا نسبتاً غیررسمی، مثل بررسی کردن کارهای موجود با دوستان و بستگان. دفتر همچنین کسانی را که بعد از ترک فصلی یک کار منتظر فراخوانی برای مراجعت به آن کار هستند، یا منتظرند که ظرف ۳۰ روز یک کار جدید را شروع کنند، بیکار به‌شمار می‌آورد.

داده‌های بیکاری ط.ج.ج.، افرادی که حقوق بیکاری دریافت می‌کنند را نمی‌شمارد. اگرچه اداره کار سوابق مطالبات بیمه بیکاری را نگهداری می‌کند، لکن اصطلاحات بیکاری تحت پوشش بیمه، کمتر از یک سوم کل بیکاری را شامل می‌شود. دلیل اصلی خیلی کم بودن این سهم این است که بسیاری از بیکاران یا هرگز شغلی نداشته‌اند یا پس از یک دوره غیبت، در حال ورود مجدد به بازار کار می‌باشند. سایرین به این دلیل بیمه بیکاری دریافت نمی‌کنند که کاری که از دست داده‌اند تحت پوشش بیمه بیکاری نبوده، یا به مدت کافی در یک شغل تحت پوشش بیمه کار نکرده‌اند، یا مدت طولانی بیکار بوده‌اند؛ به‌طوری که حق و حقوق آنها مستهلک شده‌است.

● **نیروی کار** مجموع شاغلان و بیکاران است. از ۱۸۰ میلیون افراد ۱۶ ساله و بیشتر در طول سال ۱۹۸۶، حدود ۱۱۸ میلیون، یا ۶۵٪، جزو نیروی کار بودند.

● **نرخ بیکاری** درصد نسبت تعداد افراد بیکار به نیروی کار است. به‌عنوان مثال، ۷ میلیون بیکار

از میان ۱۰۰ میلیون نیروی کار یک نرخ بیکاری ۷/۰ درصدی را به‌دست می‌دهد. آنچه که اغلب می‌شنویم درباره نرخ بیکاری است، و نه درباره تعداد واقعی شاغلین یا بیکاران. درواقع، این یکی از آمارهایی است که بیشترین اطلاع‌رسانی در مورد آن در کشور صورت گرفته است.

در محاسبه نرخ بیکاری غیرنظامی، فقط غیرنظامیان به‌حساب می‌آیند، درحالی که اعضای نیروهای مسلح ایالات متحده در نرخ مربوط به کل نیروی کار، به‌عنوان شاغل محسوب می‌شوند. نرخ بیکاری غیرنظامی نوعاً یک دهم درصد بیشتر است. نرخ بیکاری غیرنظامی حداکثر ۲۵٪ و حداقل ۱/۲٪ بوده است.

دفتر آمار نیروی کار دریافته است که بعضیها تعریف رسمی بیکاری را خیلی وسیع می‌دانند درحالی که دیگران آن را خیلی محدود می‌پندارند. برای به‌حساب آوردن چنین نقطه‌نظراتی، دفتر نرخهای مختلفی مبتنی بر تعاریف بیکاری و نیروی کار را منتشر می‌کند. یک نرخ فقط کسانی را بیکار محسوب می‌کند که برای ۱۵ هفته یا بیشتر بیکار بوده‌اند. از سوی دیگر، یک نرخ به‌دلایل اقتصادی نه‌تنها بیکاران غیرنظامی را بلکه کارکنان مأیوس و افرادی که به‌طور نیمه وقت کار می‌کنند را نیز به‌حساب می‌آورد.

● **نرخهای مشارکت نیروی کار** می‌گوید که چه نسبتی از جمعیتی که در سنین کار قرار دارند، جزو نیروی کار هستند. این نرخ به آرامی ولی به‌طور ثابت از نیمه دهه ۱۹۶۰ رشد کرده است، در ابتدا به‌دلیل اینکه رشد سریع اشتغال مزد و حقوق‌بگیران در میان زنان روند روبه رشد بازنشستگی زودتر میان مردان به‌طور متوسط، در حدود ۷۶٪ مردان و ۵۵٪ زنان، در نیروی کار در ۱۹۸۶ حضور داشته‌اند.

● **نسبت اشتغال به جمعیت** نشان‌دهنده سهمی از جمعیت در سنین کار است که شاغلند. اشتغال به استثنای دوران تندترین رکود فعالیتها رشد می‌کند. نگاهی به رشد اشتغال در ارتباط با رشد جمعیت، جنبه معنی‌دارتری به تغییر تعداد کارکنان می‌دهد.

اقتصاددانان این نسبتها و سایر ابزار تحلیلی را برای بررسی روندهای جاری بازار کار و نیروی کار ایالات متحده به‌کار می‌گیرند. دفتر آمار نیروی کار به تنهایی حدود ۷۰۰ مقاله و مطالعه خاص را برای تحلیل داده‌های ط.ج.ج. در طول ۲۵ سال گذشته، انتشار داده است. اینها بسیاری از موضوعات مختلف را پوشش می‌دهند و عناوینی مثل "بیشتر زنانی که خانواده را

اداره می‌کنند از بازار کار پاداش کمی دریافت می‌کنند"، "یک چهارم افراد بزرگسال نیروی کار فارغ‌التحصیلان مؤسسات عالی هستند"، و "آیا الگوی اشتغال در بحرانهای اقتصادی تغییر کرده است؟" را دارا هستند.

برنامه آمارهای اشتغال جاری

با داشتن مقدار اطلاعاتی که از ط.ج.ج. به‌دست می‌آید، چه چیزی برای اندازه‌گیری توسط برنامه آمارهای اشتغال جاری (آ.ا.ج.) باقی می‌ماند؟ آ.ا.ج. آرایه وسیعی از اطلاعات در مورد اشتغال، ساعات کار، و درآمدها بر حسب فعالیتها و مکانهای جغرافیایی را فراهم می‌کند و یک عنصر مهم در ساخت برآوردهای درآمد ملی، شاخص تولیدات صنعتی، و شاخصهای پیشرو اقتصادی است. (مقاله مور[۱] در مورد شاخصهای اقتصادی را ببینید.) برخلاف ط.ج.ج. که یک بررسی پیمایشی خانوارهاست، آ.ا.ج. بررسی پیمایشی کارگاههاست. دفتر آمار نیروی کار با همکاری مؤسسات ایالتی داده‌های آ.ا.ج. را با استفاده از صورت پرداخت کارگاهها جمع‌آوری می‌کند. نمونه آ.ا.ج. خیلی بزرگتر از بررسی پیمایشی خانوار، تقریباً ۳۰۰۰۰۰ کارگاه با بیش از ۳۸ میلیون شاغل در بر می‌گیرد. در واقع، نمونه آ.ا.ج. بزرگترین عملیات نمونه‌گیری ماهانه در زمینه آمارهای اقتصادی و اجتماعی است. برای جمع‌آوری داده‌های آ.ا.ج.، نمایندگیهای ایالتی فرمهایی را به کارگاههای نمونه پست کرده و پاسخها را بررسی می‌کنند. ایالتها از داده‌ها برای تهیه برآوردهای اشتغال ناحیه‌ای و ایالتی استفاده می‌کنند و همچنین داده‌های گزارش شده را برای تهیه آمارهای ملی به د.آ.ن.ک. در واشنگتن دی سی انتقال می‌دهند. هر ماه، همان فرم به کارگاهها پس فرستاده می‌شود تا داده‌های ماه بعد بتواند تحت آنچه در ماه قبل نوشته شده وارد شود. شرکتها تعداد کل کارکنان و تعداد زنان در بنگاه خود را به‌همراه اشتغال، فهرست حقوق، و ساعات کاری تولید یا شاغلان غیرمدیر را گزارش می‌کنند. آ.ا.ج. درباره شاغلان مندرج در فهرست حقوق در طول دوره پرداختی که شامل روز دوازدهم ماه هم می‌شود، سئوال می‌کنند، خیلی شبیه ارجاع بررسی پیمایشی خانوار به هفته‌ای که شامل

۱ـ *Moore*

روز دوازدهم ماه است. لکن آ.ا.ج. و ط.ج.ج. نسبتاً در مفاهیمشان با یکدیگر متفاوتند. داده‌های کارگاهی، کارکنان بخش کشاورزی و همچنین خویش‌فرمایان را در بر نمی‌گیرد. بنابراین شمارش مشاغل آ.ا.ج. کوچکتر از "اشتغال کل" از ط.ج.ج. است.

داده‌های مربوط به پرداختها برحسب نوع فعالیت جدول‌بندی می‌شوند. فعالیتهای ارائه‌کننده کالا شامل صنعت، معدن، و ساختمان می‌شوند. فعالیتهای ارائه‌کننده خدمات شامل حمل و نقل و خدمات عمومی[1] عمده فروشی، خرده فروشی، مالی، بیمه، و معاملات املاک، خدمات، و دولت می‌شود.

میدان عمل بررسی پیمایشی کارگاهی به دفتر آمار نیروی کار اجازه می‌دهد که اشتغال را به تفصیل هر یک از فعالیتها برآورد نماید. به‌عنوان مثال، تحت عنوان تولید کالاهای بادوام، صنایع فلزی پایه‌ای را می‌بینیم، که به پنج صنعت کوچکتر تقسیم می‌شود. هر یک از این اجزا مجدداً تا جایی تقسیم می‌شوند که بین ریخته‌گریهای آهن خاکستری و ریخته‌گریهای آهن چکش خور فرق می‌گذارد. ما همچنین می‌توانیم بیاموزیم چه تعداد از کارکنان این ریخته‌گریها کارکنان تولیدی هستند؟ متوسط ساعات کار هفتگی آنها، اضافه‌کاری، و دستمزد ساعتی و هفتگی آنها چقدر است؟

چرا برای بررسی بازار کار، این دو طرح پیچیده به‌کار گرفته می‌شوند؟ ساده است. تواناییهای یک طرح عموماً ناتواناییهای دیگری است. به‌عنوان مثال، آ.ا.ج. داده‌هایی را در مورد روند مشاغل در صدها فعالیت جزیی، ارائه می‌دهد، لکن درباره مشخصات کارکنان، خیلی کم و درباره کسانی که کار نمی‌کنند هیچ چیزی نمی‌گوید. در مقابل، ط.ج.ج. وارد جزئیات فعالیتها نمی‌شود، لکن درباره افرادی که شاغلند، بیکارند یا در نیروی کار نیستند، سرشار از اطلاعات تفصیلی است.

تعدیلات فصلی چیست؟

فرض کنیم که شما شنیده‌اید اشتغال (که در هر دو طرح اندازه‌گیری می‌شود) در ماه ژوئن

۱- مانند آب، برق و تلفن

گذشته کاهش یافته‌است. چطور چنین چیزی ممکن است حتی اگر اقتصاد ضعیف باشد؟ می‌دانید که در ماه ژوئن معمولاً مردم بیشتری نسبت به ماه مه کار می‌کنند. با این همه، دانش‌آموزان برای تابستان تعطیل می‌شوند و سیل فارغ‌التحصیلان تازه، وارد نیروی کار می‌شوند و سهم خوبی از آنها کار پیدا می‌کنند. اخبار غیرمنتظره، نتیجه رویه‌ای است که **تعدیل فصلی** خوانده می‌شود. تغییرات فصلی در اشتغال می‌تواند خیلی بزرگ باشد. برای بسیاری از داده‌های سری زمانی، تغییرات فصلی بیشتر تغییرات ماه به ماه را توضیح می‌دهد. به‌عنوان نمونه، اشتغال در بخش ساختمان همیشه در طول ماههای سرد زمستان نزول می‌کند، درحالی که مشاغل مربوط به خرده‌فروشی در طول ایام کریسمس زیاد می‌شود. نیروی کار در هر ژوئن توسعه می‌یابد. اشتغال در بخش کشاورزی در زمان برداشت محصول رشد می‌کند. تعدیل فصلی اساساً یک فن میانگین‌گیری مبتنی بر تجربیات قبلی است، و تغییرات فصلی را از داده‌ها برطرف می‌کند. کاهش میزان اشتغال بین ماههای مه و ژوئن، "بعد از تعدیل فصلی"، اغلب نشانگر این است که رشد تندی که عموماً رخ می‌دهد، در آن دوره کوچکتر از معمول بوده است. تغییرات داده‌ها بعد از تعدیل فصلی تغییرات حقیقی اقتصاد هستند، و نه بسادگی همانهایی که هر سال تقریباً در همین موقع رخ می‌داده‌اند.

بدون تعدیل فصلی، تعیین روندهای اساسی در یک سری زمانی دشوار است. نزولها و صعودها در چرخه کسب و کار، بعد از تعدیل فصلی بیشتر قابل مشاهده‌اند، و روندهای عادی درازمدت (مثل رشد اشتغال در طول زمان) آشکارترند.

اطلاعات حاصل از این دو طرح عمده دولت در زمینه اشتغال، عناوین خبری را می‌سازد، مباحثات پارلمانی را راه می‌اندازد، و اندیشمندان، سیاستمداران و جامعه تجاری را به وجد می‌آورد. اینکه چگونه آنها از این داده‌ها استفاده می‌کنند؟ کاملاً به خود آنها مربوط است، لکن اطمینان از دقت و بهنگام بودن داده‌ها وظیفه دولت است. زیرا بدون این طرحها مردم آگاهی کمتری نسبت به عملکرد خوب یا ضعیف اقتصاد خود می‌داشتند.

مسائل

۱ـ توضیح دهید که چرا آمارگیران سرشماری بجای اینکه فقط بپرسند "آیا شما بیکارید"، یک

سری سئوال درباره فعالیتهای کاری فرد می‌پرسند.

۲ـ چرا همه ماهه از خانوارهای یکسانی آمارگیری نمی‌شود؟ چرا همه ماهه با خانوارهای جدید مصاحبه نمی‌شود؟

۳ـ آیا بیکاری بد است؟ چگونه می‌توان تعریف "رسمی" بیکاری را وسیعتر یا محدودتر نمود تا به یک نرخ بیکاری بیشتر یا کمتر منجر شود؟

۴ـ چرا بسیاری از شاخصهای نیروی کار به‌صورت نسبت بیان می‌شوند؟ مثل نرخ بیکاری یا نسبت اشتغال به جمعیت.

۵ـ چرا دولت نمی‌تواند شمارش کامل بیکاران را از آمارهای ثبتی سیستم بیمه بیکاری یا با پرسش از کارفرمایان، به‌دست آورد؟

۶ـ چرا بررسی پیمایشی کارگاهی، برای فعالیتهای خاص، برآوردهای دقیقتری نسبت به بررسی پیمایشی خانوار ارائه می‌دهد؟

۷ـ در چه صورت برآوردهای اشتغال از بررسیهای پیمایشی کارگاهی و خانوار با یکدیگر متفاوت خواهند بود؟

۸ـ چرا داده‌های اشتغال و بیکاری به‌طور فصلی تعدیل می‌شوند؟ چرا داده‌های تعدیل‌نشده برای تحلیل تغییرات در طول ماه مورد استفاده قرار نمی‌گیرند؟

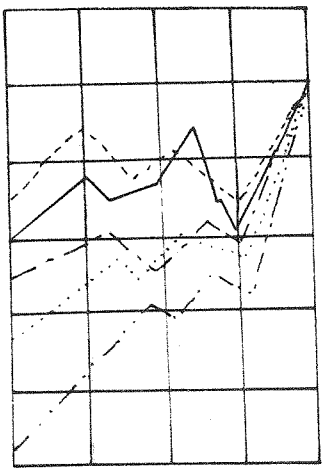
۹ـ آیا ارقام گزارش شده در مورد بیکاری در میان جوانان سیاهپوست نسبت به کل بیکاری به همان اندازه، بیشتر، یا کمتر دقیق است؟ چرا؟

۱۰ـ چرا دفتر آمار نیروی کار نرخ بیکاری کل را تقریباً تغییر نیافته می‌خواند وقتی که از ۷٪ به ۶/۹٪ تنزل می‌یابد؟

منابع

U.S. Department of Labor. Bureau of Labor Statistics, April 1988. *BLS Handbook of Methods. 2285.* Washington D.C.: U.S. Government Printing Office.

U.S. Department of Labor. Bureau of Labor Statistics, 1987. *How the Government Measures Unemployment.* Report 742. Washington D.C.: U.S. Government Printing Office.



بسط و تحلیل شاخصهای اقتصادی[۱]

جفری اچ. مور[۲] مرکز پژوهشهای بین‌المللی دوره‌های تجاری. دانشکده تحصیلات تکمیلی بازرگانی، دانشگاه کلمبیا

شاخصها و دوره‌های تجاری

به نظر می‌رسد که دوره‌های تجاری (بزرگ وکوچک) ویژگی همیشگی دورنمای اقتصادی است. کسادی و رونق اقتصادی، آشکارا حادثه‌ای با اهمیت عمده اجتماعی است. بنابراین

۱ـ *این متن برگرفته و بهنگام شده مقاله نویسنده "تحلیل شاخصهای اقتصادی" منتشره در ساینتیفیک امریکن ، ۱: ۳۳۲ (ژانویه ۱۹۷۵) است.*

۲- *Geoffrey H. Moore*

توسل به روشهایی که بتوانند یک تغییر اقتصادی را پیش‌بینی و گستره آن را برآورد کنند، به میزان قابل ملاحظه‌ای مورد توجه است. این نقشی است که **شاخصهای اقتصادی** ایفا می‌کنند. شاخصهایی که به اندازه‌گیریهای متعددی از ضربان اقتصاد متکی هستند که توسط موسسات دولتی، سازمانهای خصوصی و اقتصاددانان انجام می‌گیرند. تحلیل شاخصهای اقتصادی فنی کاملاً توسعه‌یافته برای تعیین کمیت ضربانهایی است که اوضاع اقتصادی را نشان می‌دهند.

شاخصهای اقتصادی تقریباً تمام معیارهای کمّی آن تغییرات اقتصادی را شامل می‌شوند که همواره در دسترس هستند. شاخصهای روزانه، هفتگی، ماهانه و فصلی بسیاری را می‌توان یافت که تولید، قیمتها، درآمدها، اشتغال، سرمایه‌گذاری، موجودیها، فروشها و مانند آنها را اندازه‌گیری و برنامه‌ها، تعهدات و پیش‌بینیها و همچنین معاملات اخیر را ثبت می‌کنند. برخی از شاخصها همچون نرخ بیکاری یا شاخص بهای کالاها و خدمات مصرفی را دولت مرکزی براساس بررسیهای نمونه‌ای تفصیلی که همه ماهه انجام می‌شوند، محاسبه می‌کند. برخی دیگر همچون شاخص قیمت بازار سهام، بررسی نظر نمایندگان خرید نسبت به قیمتها، سفارشها و موجودیهای انبار براساس اطلاعاتی ساخته می‌شوند که سازمانهای خصوصی جمع‌آوری می‌کنند یابه‌عنوان یک محصول جانبی به‌دست می‌آیند.

درنتیجه اقتصاددانان یا بازرگانان علاقمند به پیش‌بینی تغییراتند و همچون پیش‌بینی کنندگان هوا، با انبوهی از اطلاعات واقعی مواجهند که دائماً دریافت می‌شوند. اینان باید به‌طور منظم به ارزیابی آنچه اطلاعات درباره اوضاع فعلی و آتی می‌گوید، بپردازند. **فن تحلیل شاخص** راههای منظم مختلفی را برای ملاحظه این اطلاعات، با نگاهی به ارزیابی وضعیت فعلی و شناسایی وقایع مهم آینده در دوره تجاری ارائه می‌کند.

یکی از نظامهای اولیه از این نوع که اندکی پیش از جنگ جهانی اول تکوین یافت، به منحنیهای الفبایی هاروارد مشهور شده‌است. منحنی الف نمایه‌ای بود که سفته‌بازی و به‌طور مشخصتر، قیمتهای سهام را نشان می‌داد. منحنی ب، فعالیتهای تجاری رانشان می‌داد که با ارقام دلاری چکهای صادره از حسابهای بانکی اندازه‌گیری می‌شد. منحنی پ، بازار پول را نمایش می‌داد که با نرخ بهره وامهای تجاری کوتاه‌مدت سنجیده می‌شد. مطالعات تاریخی بویژه آنها که

توسط *وارن پرسونز*[۱] در دانشگاه هاروارد انجام گرفته، نشان دادند که نوعاً این سه منحنی به‌دنبال هم حرکت می‌کردند: نخست قیمتهای سهام، سپس بدهیهای بانکی و در آخر نرخهای بهره، با چرخشهای پسگامانه نرخهای بهره قبل از چرخشهای معکوس قیمتهای سهام. منطق اقتصادی این دنباله آن بود که انقباض پولی و نرخهای بهره بالا به افول انتظارات بازرگانی و تنزل قیمتهای سهام منجر شد که این دو نیز کاهش سرمایه‌گذاری و رکود تجارت را به‌دنبال آوردند. رکود هم به سهم خود به فراوانی پول و نرخهای بهره پایینتر منجر شد که در نهایت چشم‌انداز بازرگانی را بهبود بخشید، قیمتهای سهام را بالا برد و گسترش تازه‌ای را در فعالیتهای اقتصادی ایجاد کرد.

نظام اقتصادی در بحران بزرگ ۱۹۲۹ گرفتار شد. زیرا تفسیرکنندگان این منحنیها با نظر خوشبینانه‌ای بدانها نگریستند و از پیش‌بینی سقوط ناگهانی غافل ماندند. اقتصاددانان معمولاً از این حادثه به‌عنوان یکی از شکستهای بزرگ پیش‌بینی در تمام دوران یاد می‌کنند.

اما کنجکاوانه این دنباله از وقایع که پایه‌های نظام درابتدا بر آن استوار شده‌بود، همچنان حفظ شده است. این نکته بدان معنی نیست که اگر منحنیهای الفبایی دوباره رواج پیدا می‌کردند، هنوز هم برای مقاصد موردنظر کافی هستند. مجموعه شاخصهای بسیار جامعتری از ۱۹۲۹ به بعد تدوین شده و پایه نظری و تجربی این شاخصها عمیقاً مطالعه، مستندسازی و آزمون شده‌است.

رکود شدید ۱۹۳۷ تا ۱۹۳۸ که پیش از بهبودی کامل اقتصاد از بحران بزرگ رخ داد، انگیزش دیگر این بسط شد. در پاییز ۱۹۳۷، هنری مورگنتاو دوم[۲]، وزیر خزانه‌داری آمریکا از اداره ملی تحقیقات اقتصادی (یک موسسه تحقیقاتی خصوصی غیرانتقاعی) خواست تا نظامی از شاخصها را بسازد که هنگام پایان گرفتن رکود اقتصادی را اعلام کند. در آن زمان تحلیل کمّی عملکرد اقتصادی در ایالات متحده با استانداردهای امروزی تطابق نداشت. حسابهای دولتی تولید و درآمد ملی که مبنای بیشتر تحلیلهای اقتصادی امروزی را تشکیل می‌دهند، تازه تدوین شده بودند. دیگر آماره‌های اقتصادی مهم از جمله نرخهای بیکاری توسط موسسه‌های عمومی و در تلاش برای تدارک اطلاعات مفید در مبارزه بابحران؛ تدوین یا پالایش شده‌بودند. چند

۱- *Warren Persons*

۲- *Henry Morgenthau. Jr.*

سری آماری با تعدیل فصلی به شکل امروزی منتشر شدند. الگوهای اقتصادسنجی جامع (دستگاه معادلات بیانگر روابط کمّی بین متغیرهای اقتصادی) که اکنون به‌طور گسترده‌ای برای پیش‌بینی و ارزیابی سیاستهای اقتصادی به‌کار گرفته می‌شوند، درآن زمان تقریباً ناشناخته بودند.

اداره ملی پژوهشهای اقتصادی تحت هدایت *وزلی. سی. میچل* و *آرتور. اف. برنز*[۱] از دهه ۱۹۲۰ به بعد به پردازش و تحلیل اطلاعات ماهانه، فصلی و سالانه قیمتها، اشتغال، تولید و دیگر عوامل به‌عنوان بخشی از تلاشهای تحقیقاتی اساسی با هدف نیل به ادراک بهتر دورهای تجاری پرداخته است. این پروژه میچل و برنز را قادر ساخت تا سریهایی را انتخاب کنند که براساس عملکرد گذشته و ارتباط با دوره‌های تجاری، شاخصهای نسبتاً قابل اعتمادی برای تجدید حیات فعالیتهای تجاری به‌نظر می‌آمدند. این فهرست به وزارت خزانه‌داری در اواخر سال ۱۹۳۷، در پاسخ به درخواست مورگنتاو ارائه و در مه ۱۹۳۸ منتشرشد. بدین ترتیب نظام شاخصهای پیشگام، همگام، پسگامی پایه‌ریزی شد که امروزه در تحلیل وضعیتهای اقتصادی، تعیین عوامل مساعد و نامساعد و پیش‌بینی تحولات کوتاه‌مدت به‌طور گسترده‌ای مورد استفاده واقع می‌شوند.

از ۱۹۳۸ به بعد، فراهمی و به‌کارگیری شاخصهای اقتصادی باهدایت اداره ملی پژوهشهای اقتصادی، وزارت بازرگانی آمریکا، سازمان توسعه و همکاریهای اقتصادی در پاریس و دیگر موسسه‌های خصوصی وعمومی، بشدت گسترش یافته‌اند. در سالهای ۱۹۵۰، ۱۹۶۰، ۱۹۶۶ و ۱۹۷۵ در فهرست شاخصهای ساخته‌شده درسال ۱۹۳۷ برای لحاظ کردن سریهای اقتصادی جدید، یافته‌های پژوهشی جدید و نیز تغییر در ساختاراقتصاد، تجدیدنظر شد. بتازگی ارزیابی جدیدی نیز توسط مرکز بین‌المللی پژوهشهای دوره‌های تجاری[۲] در دانشگاه کلمبیا صورت گرفته‌است. با هر تجدیدنظر، عملکرد شاخصها قبل و بعد از زمان انتخاب آنها بدقت آزموده و جهت بررسی عموم ارائه شده‌است.

۱- *Wesley C. Mitchel. and Arthur F. Burns.*

۲- *Center for International Business Cycle Research*

در ۱۹۷۵ ریموند سال نیر[۱]، که درآن زمان رییس شورای مشاوران اقتصادی رییس جمهور آمریکا بود، از اداره سرشماری تدوین روشهایی را درخواست کرد که در ارزشیابی نوسانهای تجاری فعلی بتوانند از پردازش داده‌ها با مقیاس وسیع الکترونیکی که امکانپذیر شده بود، استفاده کنند و نتایج در گزارشهای ماهانه درج شوند. اقدامهای آزمایشی انجام شده طی چند سال بعد زیر نظر جولیوس شیسکین[۲]، که آمارشناس اقتصادی عالی‌رتبه اداره سرشماری بود، به انتشار ماهانه بیزنس سایکل دولو پمنتس[۳] در ۱۹۶۱ منجر شد. (این نشریه اکنون بیزنس کاندیشنز دایجست[۴] خوانده می‌شود ولی اقتصاددانان به آن تحت هر دو نام و به اختصار با BCD ارجاع می‌دهند.) این نشریه دسترسی به شاخصهای فعلی و ابزارهای آماری مختلفی را که در تعبیر آنها موثر است بشدت افزایش داده‌است. درنتیجه شاخصها یک وسیله عمده پیش‌بینی اقتصادی شده‌اند.

ویژگیهای سودمند شاخصها

همان‌طور که در بالا نیز اشاره‌شد، تحلیل شاخصهای اقتصادی بر بنیانهای نظری و تجربی استوار است. انتخاب شاخصهای ویژه و تاکید منسوب به آنها متاثر از درک ما از علتهای دوره‌های تجاری است. انسان آشکارا به بررسی تحولات تازه در هر فرایند اقتصادی علاقه‌مند است که باور دارد نقش مهمی در هر توضیح عموماً پذیرفته شده نوسانهای دوره‌ای ایفا می‌کند.

برای این نوسانها توضیحهای مختلف بسیاری ارائه شده‌اند. برخی از این توضیحها، تاکید اولیه را بر نوسانهای سرمایه‌گذاری، موجودی انبار و دستگاهها و تجهیزات تولیدی جدید می‌گذارند که هم توسط تغییر در تقاضای نهایی تعیین می‌شوند و هم آنها را تعیین می‌کنند. برخی دیگر نقش محوری را به عرضه پول و اعتبار یا به سیاستهای مالیاتی و هزینه‌ای دولت یا به روابط بین قیمتها، هزینه‌ها و سودها می‌دهند.

بدون شک همه این عوامل مسیر فعالیتهای بازرگانی را متاثر می‌سازند. برخی از این عوامل

۱ـ *Raymond J. Saulnier*

۲ـ *Julius Shiskin*

۳ـ *Business Cycle Developments*

۴ـ *Business Conditions Digest*

ممکن است در یک مقطع زمانی مهمتر از بقیه عوامل باشند اما هیچ اتفاق نظری نسبت به مهمترین عامل یا حتی چگونگی واکنش متقابل آنها وجود ندارد. بنابراین باید محض احتیاط با مجموعه‌ای از شاخصها کار کرد که نمایشگر طیف گسترده تاثیرات هستند. دسترسی آسان به طیف وسیع شاخصها، به فرد امکان آزمودن فرضهای رقیب یا مکمل را درباره نوسانهای فعلی اقتصاد می‌دهد.

با این ذهنیت، فعالیتهای اقتصادی را می‌توان به چند رسته گسترده از فرایندهای کاملاً وابسته تقسیم کرد که از نقطه‌نظر دوره‌های تجاری معنی‌دار باشند. از هر رسته چند شاخص انتخاب شدند. رسته‌های اصلی که اکنون در "بیزنس کاندیشنز دایجست" ارائه می‌شوند، در ستون سمت چپ جدول ۱ آمده‌اند. توجه داشته باشید که این رسته‌ها شامل تمامی جنبه‌های اقتصادی نمی‌شوند. برای مثال آمار در مورد کشاورزی، دولتهای مرکزی، ایالتی و محلی، تجارت خارجی، جمعیت و دارائیها حذف می‌شوند. با این وجوه، این رسته‌ها چارچوب عواملی را مشخص می‌کنند که در نظریه‌های دوره‌های تجاری ملحوظ می‌شوند و در ارزیابی عملکرد اقتصادی مهمند. رسته‌های حذف شده هم مهمند اما فقط چند شاخصی را که به‌طور منظم مربوط به دوره‌های تجاری هستند، شامل می‌شوند.

مطالعه روی دوره های تجاری ، شاخصهایی رادرهر رسته مشخص کرده است که رفتاری منظم دارند. این یافته‌ها، مبنای انتخاب شاخصهای معینی راونیز رده بندی آنهارابرحسب ویژگیهای رفتاری دوره ای آنها، به‌صورت جدول ۱ فراهم کرده اند. دو ویژگی اصلی موردنظر، یکی قاعده تطابق شاخصهابادوره های تجاری است ودیگری سازگاری پیشگامی ، همگامی و پسگامی شاخصهادرنقاط برگشت دوره های تجاری است . دیگرملاحظات مربوط، کفایت آماری داده ها(زیراپشتیبانی آماری ازیک شاخص بستگی به این داردکه چقدر انعکاس فرایندمفروضی راخوب معرفی می‌کند) ، هموازی داده ها(زیراسریهای بسیارپراکنده رابه سختی می توان بدرستی تفسیرکرد) وسرعت انتشارداده ها (چون ارقام قدیمی تاثیرمحدودی بر وضعیت فعلی دارند) هستند.

معیارهای تجربی این مشخصه‌هابرای تعداد زیادی شاخص ساخته شده‌اند. از چنین معیارهایی در تلاش برای کسب اطلاعاتی استفاده شده‌است که قادر به انتقال تصویری مناسب

جدول ۱ـ رده‌بندی متقاطع شاخصهای اقتصادی

فرایندااقتصادی	ارتباط بادوره های تجاری		
	پیشگام	تقریباً همگام	پسگام
اشتغال وبیکاری	نرخهای استخراج واستخدام و متوسط ساعات کار هفتگی و اضافه کاری تقاضاهای جدید بیمه بیکاری	اشتغال کل	بیکاری بلندمدت
تولید و درآمد	-	محصول ناخالص ملی واقعی تولیدصنعتی درآمدشخصی	
مصرف، تجارت، سفارشهاو حمل کالاها	سفارشهای جدید،کالاهای مصرفی عملکرد فروشنده	خرده‌فروشی، فروش صنعتی وتجاری	
سرمایه‌گذاری ثابت	تعهدات جدید سرمایه گذاری تأسیس مؤسسات تجاری ساخت اماکن مسکونی		انباشته شدن تعهدات سرمایه‌گذاری هزینه‌های سرمایه‌گذاری
موجودیهای انبار و سرمایه گذاری ذخیره‌ای قیمتها، هزینه ها و سودها	خریدو سرمایه‌گذاری ذخیره‌ای تغییر قیمت مواد صنعتی، قیمت سهام، سود و حاشیه‌های سود	تغییر در قیمتهای تولیدکننده	میزان موجودی انبار تغییر در قیمتهای مصرف‌کننده، تغییر در هزینه‌های هر واحد کار
پول و اعتبار	جریانهای پول واعتبار و شکست درکار و کسب قیمت اوراق بهادار		بدهیهای قابل توجه نرخهای بهره

از تغییرات در اقتصادی باشند که از مراحل رونق و کسادی در حال گذر است. به علاوه بر رفتار شاخصهای برگزیده از نزدیک نظارت شده‌است. بسیاری از شاخصها به‌خوبی چندین ارزیابی پیاپی را پشت سر گذاشته‌اند. برای مثال معیارهای متوسط ساعت کار هفتگی، قراردادهای

ساخت و قیمتهای سهام در تمام فهرستهایی از شاخصها که از ۱۹۳۷ به‌بعد توسط دفتر ملی پژوهشهای اقتصادی به‌طور پیاپی تنظیم شده، آمده‌اند.

این شاخصها از نظر عملکردشان نیز در دیگرکشورها آزمایش شده‌اند. هر کسادی یا تنزل جدید، خواه در این کشور یا خارج، شاهد دیگری است علیه آنچه شاخصها می‌توانند ارزیابی کنند. همان‌طور که در مورد هر صعود نیز چنین است. این آزمایشها و آزمایشهای مجدد، به انباشته شدن انبوهی از شواهد تجربی منجر شدند که هم ارزش شاخصهاو هم محدودیتهای آنها را به نمایش می‌گذارند.

تعدیل فصلی و هموارسازی

نمونه‌گیری از این شواهد در نمودارهای صفحات بعد ظاهر می‌شوند. شاخصی همچون تعداد مجوزهای ساخت صادره برای خانه‌های جدید را درنظر بگیرید (ن. ک. نمودار ۱). داده‌های خام از نظر آماری تفکیک شدند تا تغییرات معمول فصلی را که همه‌ساله تکرار می‌شوند، اندازه‌گیری و حذف نمایند. وقتی این عامل حذف شود، شاخص روند رو به کاهش مجوزها را ـ حین رکودهای ۱۹۸۰، ۱۹۸۱ و ۱۹۸۲ ـ و روند رو به افزایش آنها را (هنگامی که شرایط رونق در ۱۹۸۳ دوباره پدیدار شد) آشکارتر بیان می‌کند.

امروزه بسیاری از شاخصهای اقتصادی به‌صورت تعدیل شده فصلی در دسترس هستند. برخی از آنها بندرت به راه دیگری گزارش می‌شوند. تولید ناخالص ملی، نرخ بیکاری و شاخص تولید صنعتی مثالهایی از شاخصهایی هستند که نسبت به عوامل فصلی به‌صورت ثابتی تعدیل شده‌اند. یکی از برنامه‌های کامپیوتری که به‌طور گسترده‌ای در تعدیلهای فصلی به‌کار برده می‌شوند و توسط اداره سرشماری تدوین شده‌است، برنامه X-11 خوانده می‌شود (ن. ک. مقاله *لئون ورانز* برای بحثی پیرامون تعدیلهای فصلی).

به‌طور معمول هموارسازی تغییرات بیقاعده کمتر صورت می‌گیرد زیرا فنون آن نیز چندان معمول نیستند ولی برخی از سریهای آماری به سبب تفاوت در خطای نمونه‌گیری یا تاثیرپذیری از عواملی چون شرایط نامناسب آب وهوایی و یا اختلافات کارگری بیشتر از دیگر شاخصها تحت تاثیر تغییرات گسترده بیقاعده هستند. بنابراین مفید است که در تفسیر تغییرات فصلی،

یک معیار استاندارد از میزان نوسانهای بیقاعده در مقایسه با میزان تغییراتی داشت که بیانگر روندهای بلندمدت و پیشامدهای دوره‌های تجاری موسوم به تغییرات روند ـ دوره‌ای است.

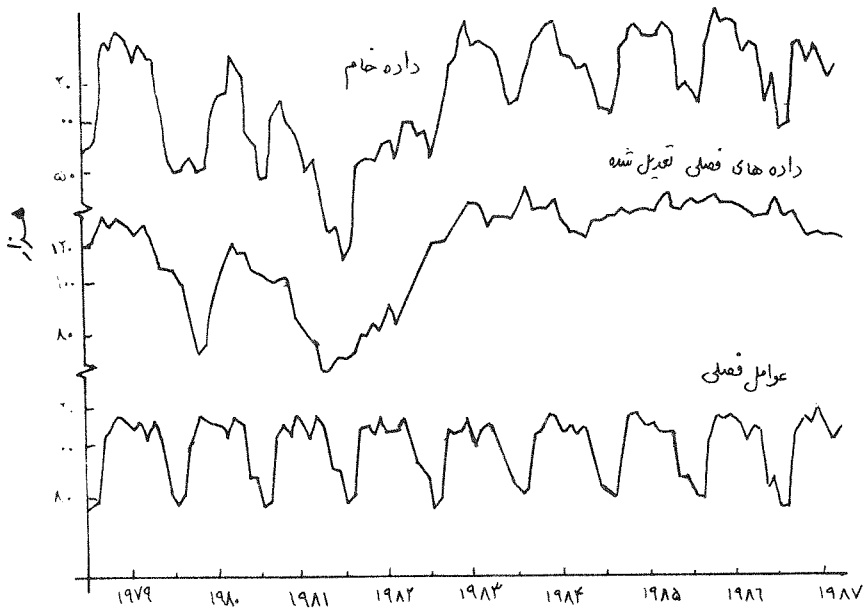
دو معیار از این نوع برای تمام شاخصهای گزارش شده در بیزنس کاندیشنز دایجست ارائه می‌شود. یکی نشان می‌دهد که متوسط تغییر ماهانه در جزء بیقاعده در مقایسه با متوسط تغییر ماهانه در جزء روند ـ دوره‌ای چقدر بزرگ است. دیگری نشان می‌دهد که به‌طور متوسط چند ماه باید بگذرد تا تغییر در جزء روند ـ دوره‌ای که در طول یک دوره زمانی ساخته می‌شود بزرگتر از جزء بیقاعده شود، که البته چنین نمی‌شود. برای مثال معیارها نشان می‌دهند که ممکن است تغییرات ماهانه در شروع خانه‌سازی تحت سلطه "نوفه" (همچون تغییرات تصادفی) درآید، اما هرگاه که این تغییرات در دوره‌های چهارماهه اندازه‌گیری شوند، "نشانه" روند ـ دوره‌ای حاکم می‌شود. از طرف دیگر شاخص تولید صنعتی از "نوفه" کمتر تاثیر می‌پذیرد به‌طوری که تغییرات ماهانه معنی‌دارترند.

پیشگامی و پسگامی

مهمترین مشخصه یک شاخص از نظر پیش‌بینی، البته ارائه شواهدی درباره تغییرات آتی فعالیتهای اقتصادی است. از این منظر، شاخصهابه‌دلایل متعدد از هم متمایزند. شاخصهایی خاص همچون شروع به خانه‌سازی، قرارداد ساخت و سفارشهای جدید ماشین آلات و ابزارها مرحله اول تصمیم‌گیری در فرایند سرمایه‌گذاری را نشان می‌دهند. چون ساخت یک خانه، یک کارخانه یا توربین زمانبر است، تولید (یاتکمیل یاحمل) آنها معمولاً بعد از سفارشها یا عقد قراردادها انجام می‌گیرد. این تاخیر علاوه بر دیگر عوامل، به حجم سفارشها یا قراردادهای تکمیل نشده بستگی دارد. هرگاه کالا برای انبار کردن و نه براساس سفارش ساخته شود، دیگر تاخیری بروز نخواهد کرد زیرا سفارشها به محض وصول، انجام می‌شوند.

نوع دیگری از رابطه پیشگامی ـ پسگامی بین تغییر در ساعت کار هفتگی از یکسو و اشتغال از سوی دیگر بروز می‌کند. در بسیاری از موسسات، کارفرمایان ساعت کار را سریعتر، ارزانتر و با تعهد کمتری در مقایسه با استخدام یا اخراج کارگران می‌توانند افزایش یا کاهش دهند. درنتیجه در بیشتر صنایع تولیدی، افزایش یاکاهش متوسط ساعت کار هفتگی معمولاً پیش از

نمودار ۱ـ مجوزهای ساخت خانه‌های جدید، قبل وبعد از تعدیلهای فصلی.

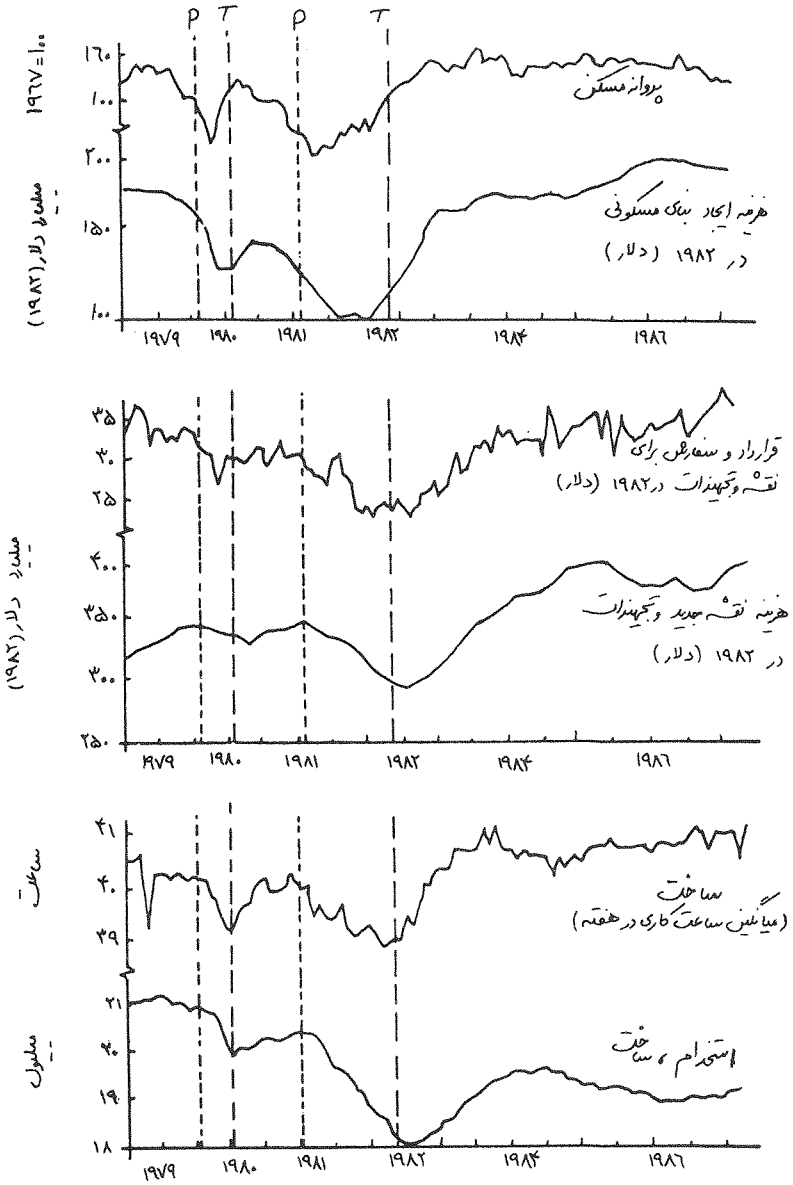


ماخذ: وزارت بازرگانی آمریکا

تغییر متناظر در میزان اشتغال آغاز می‌شود. بنابراین ساعت کار هفتگی نسبت به نرخ بیکاری یک شاخص پیشگام‌است.

چند رابطه دوجانبه از این نوع ردیابی شده‌اند. (ن.ک. شکل ۲). اما آشکارا هرگاه رابطه‌ها چندجانبه شوند، موضوع نیز پیچیده‌تر می‌گردد. برای مثال شاخصهای قیمتی بازار سهام، گرایش پایداری را در پیشگامی تغییر فعالیتهای اقتصادی نشان داده‌اند. (منحنیهای الفبایی هاروارد تاحدودی بر این گرایش متکی است) اما به‌نظر می‌رسد که توضیح آن بستگی به واکنش متقابل تغییرات نرخهای بهره و سود و همچنین دیگر عوامل داشته باشد. تنزل دوره‌ای سود اغلب پیش از پایان یافتن گسترش فعالیتهای تجاری شروع می‌شود؛ علت تقریبی معمولاً

نمودار ۲ـ سه شاخص پیشگام و فعالیتهایی که به‌دنبال آنها حرکت می‌کنند.



توجه: خطوط عمودی اوجها (p) و حضیضهای (T) دوره تجاری هستند.

افزایش سریع هزینه‌های تولیدی است. نرخهای بهره نیز محتملاً بشدت افزایش می‌یابند. هردو عامل در جهت کاهش جاذبه سهام عادی و کاهش قیمت آنها عمل می‌کنند، هرچند حجم فعالیتهای اقتصادی هنوز هم در حال افزایش است. نزدیک به انتهای یک کسادی، گرایشهای معکوس فعال می‌شوند و قیمتهای سهام را پیش از آغاز بهبودی فعالیتهای تجاری، افزایش می‌دهند. چون نرخهای بهره به‌طور کلی جزو شاخصهای پسگام محسوب می‌شوند، این موضوع نشان می‌دهد که حتی شاخصهای پسگام هم می‌توانند در پیش‌بینی یا متاثر ساختن تغییرات شاخصهای پیشگام نقشی بازی کنند.

به منظور سنجش پیشگامی و پسگامی، تاریخ‌شناسی دوره‌های تجاری سودمندی خود را اثبات می‌کند. اداره ملی پژوهشهای اقتصادی، دوره‌های تجاری را به گونه‌ای تعریف کرده است که زمان اوجها و حضیضها را بتوان با عینیت منطقی ثبت کرد. درواقع اکنون می‌توان قسمتهایی از فرایند ثبت زمانی وقایع را با کامپیوتر انجام داد. چون اکثریت غالب شاخصهای موردنظر، تغییرات دوره‌ای سازگاری را با دوره‌های تجاری نشان می‌دهند، اوجها و حضیضهای هر شاخص را می‌توان با اوجها و حضیضهای دوره تجاری تطبیق داد تا پیشگامی و پسگامی مشخصه تعیین شوند.

به‌دنبال این طرح، چند سری گروه شاخص معین می‌شوند که نوعاً پیشگام، همگام، یاهمه چرخش پسگام دوره‌های تجاری هستند، همان‌طور که در جدول ۱ بیان شده است؛ شاخصهای ترکیبی ساخته شده از این گروهها را (ن. ک. شکل۳) می‌توان ـ مانند شاخصهای مجزا ـ برای اندازه‌گیری شدت نسبی رکود اقتصادی درگذر از یک ماه به ماه بعد به‌کاربرد. با چنین شمای نظارت‌کننده، می‌توان شدت نسبی رکود فعلی را مشاهده کرد و براساس این واقعیت که معمولاً رتبه‌بندیهای شدت کسادیهای مختلف حداقل در چند ماه اول تغییر زیادی نکرده است، نکاتی را می‌توان استنباط کرد.

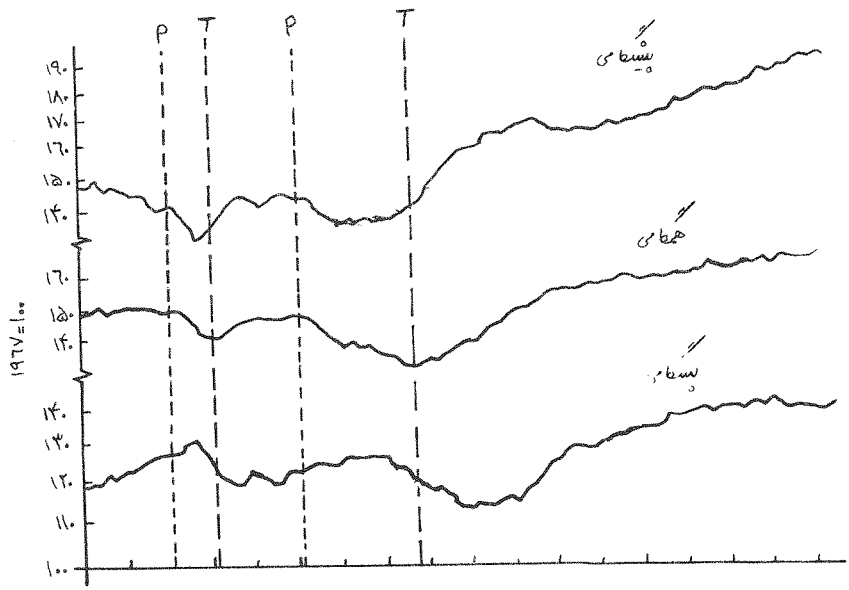
البته در هرارزیابی چشم‌انداز اقتصادی، لحاظ کردن اقدامهای سیاستی واقعی و آینده‌نگر دولت کاری بنیادین است. چنین اقدامهایی شامل افزایش یا کاهش مالیات، تغییر در سپرده‌های اجباری بانکها، تغییر در هزینه‌های نظامی و برپایی برنامه‌های اشتغال عمومی نیز می‌شود. این اقدامها اغلب به‌سادگی در قالب شاخصها تعریف نمی‌شوند، گرچه آثار آنها، به همراه تاثیر

دیگر عوامل، ممکن است فوراً در سفارشها، قراردادها شروع به ساختمان سازی، قیمتهای سهام و مانند آنها منعکس شود. با وجود این، شاخصهای خاصی به‌صورتی تقریباً پیوسته سنجشی از فعالیتهای دولتی را ارائه می‌کنند. گرچه نسبت به دوره‌های تجاری رابطه به حد کافی منظمی برایشان نیافته‌اند که گزینش شاخصهای پیشگام و همگام یا پسگام را تضمین کند.

جنبه‌های بین‌المللی

از انواع شاخصهای اقتصادی در ایالات متحده به‌صورتی گسترده‌تر از اغلب کشورهای دیگر استفاده شده‌است اما رشد تجارت، مسافرت و مالیه بین‌المللی، نیاز دیگر کشورها را به آمار

نمودار ۳ـ شاخصهای پیشگام، همگام، پسگام، برای ایالات متحده



توجه: خطوط عمودی اوجها (P) و حضیضهای (T) دوره تجاری را نشان می‌دهند.

مأخذ: وزارت بازرگانی آمریکا

بادسترسی آسان درباره معاملات بین‌المللی و توسعه اقتصادی افزایش داده‌است. در ۱۹۷۳ اداره
ملی پژوهشهای اقتصادی، برنامه ساخت و تحلیل شاخصها را برای چندین کشور صنعتی آغاز کرد. از ۱۹۷۹ به بعد این برنامه توسط مرکز پژوهشهای بین‌المللی دوره‌های تجاری دنبال شده است.

خوشبختانه شیوه شاخص به حد کافی انعطاف‌پذیر است تا به‌آسانی در وضعیتهایی به‌کار گرفته شود که در آنها مانند بسیاری از کشورها بعد از جنگ جهانی دوم کسادی اقتصادی به‌جای تنزلهای مطلق به شکل رشد عقب‌افتاده فعالیتهای انبوه درآمده است و این عقب‌افتادگی عمداً با سیاستهای دولتی برای خنثی‌سازی تورم یا برقراری تراز تجاری در حال افول القا شده‌است. فراتر اینکه این شیوه برای لحاظ کردن تفاوتهای بین کشورها از نظر نوع شاخصهایی که وجود دارند یا بیان‌کننده‌ترند، به حد کافی انعطاف پذیراست. برای مثال در اروپا، بیشتر از آمریکا بر آمار مربوط به فرصتهای شغلی تکیه می‌شود و داده‌های مربوط به مهاجرت بین‌المللی کارگران مهمترند زیرا کارگران مهاجر بخش مهمی از نیروی کار کشورهای اروپایی را تکشیل می‌دهند.

دنبال کردن تحلیل شاخصهای اقتصادی در یک مقیاس بین‌المللی توسط موسسات بین‌المللی علاوه بر موسسات داخلی، امکان استفاده از این شیوه و ارزش بالقوه آن را در مشاهده و ارزیابی نوسانهای بین‌المللی، در نرخهای رشد اقتصادی و روندهای ملازم آن در سطح قیمتها، تجارت خارجی، سرمایه‌گذاری و اشتغال نمایان ساخته‌است. می‌توان تحول یک نظام جهانی شاخص را، براساس برنامه‌ای که در ابتدا برای ایالات متحده تدوین شده‌بود، برای پشتیبانی از تحلیل شاخصهای اقتصادی در یک مقیاس جهانی در ذهن مجسم کرد.

مسائل

۱ـ چرا از داده‌های تعدیل شده فصلی در تحلیل دوره‌های تجاری استفاده می‌شود؟

۲ـ شاخصهای پیشگام چه می‌کنند؟

۳ـ یک فرایند شاخص خوب اقتصادی باید چه ویژگیهایی داشته باشد؟

۴ـ نامزد سیاسی الف می‌گوید که کسادی به سبب تنزل قیمت سهام آغاز شده‌است. نامزد ب

مخالف است و می‌گوید علت کسادی آن است که اشتغال هنوز هم درحال افزایش است. شما چگونه تصمیم‌گیری می کنید که حق با کیست؟

۵ـ ارقام زیر از نشریه بیزنس کاندیشنز دایجست، شماره ژوئن ۱۹۷۵، اقتباس شده‌است. باتوجه به این واقعیت که کسادی در نوامبر ۱۹۷۳ آغاز شد، فکر می‌کنید که پایان یافته‌است و در این صورت کی؟ چه شواهد تأییدکننده‌ای را مایلید دراختیار داشته باشید؟

	شاخصهای پیشگام	شاخصهای همگام (۱۰۰ = ۱۹۶۷)	شاخصهای پسگام
ژوئیه ۱۹۷۴	۱۴۵/۳	۱۳۸/۸	۲۱۰/۵
اوت ۱۹۷۴	۱۴۰/۴	۱۳۸/۲	۲۱۴/۵
سپتامبر ۱۹۷۴	۱۳۵/۰	۱۳۷/۴	۲۱۶/۳
اکتبر ۱۹۷۴	۱۳۰/۱	۱۳۶/۲	۲۱۹/۰
نوامبر ۱۹۷۴	۱۲۶/۰	۱۳۲/۳	۲۲۰/۴
دسامبر ۱۹۷۴	۱۲۳/۶	۱۲۸/۲	۲۲۰/۲
ژانویه ۱۹۷۵	۱۱۸/۷	۱۲۵/۲	۲۱۷/۸
فوریه ۱۹۷۵	۱۱۸/۶	۱۲۴/۱	۲۱۲/۹
مارس ۱۹۷۵	۱۲۰/۳	۱۲۲/۰	۲۱۰/۱
آوریل ۱۹۷۵	۱۲۴/۸	۱۲۲/۲	۲۰۵/۵
مه ۱۹۷۵	۱۲۷/۹	۱۲۲/۴	۲۰۱/۵
ژوئن ۱۹۷۵	۱۳۰/۷	۱۲۴/۶	۲۰۰/۲

۶ـ چگونه شاخصهای پیشگام یک کشور دیگر، مثلاً ژاپن، می‌تواند برای بنگاههای تجارتی ایالات متحده مفید باشد؟

۷ـ در تصمیم‌گیری نسبت به سرمایه‌گذاری در سهام یا اوراق قرضه، کدام شاخصهای اقتصادی را بررسی می‌کنید؟

منابع

Bureau of Economic Analysis, U. S. Department of Commerce, 1984. *Handbook of Cgelical Indicators.* Washington, D. C.

Bureau of Economic Analysis, U. S. Department of Commerce, *Business Conditions Digest,* monthly. Washington.D.C.

Philip A. Klein and Geoffrey M. Moore. 1985. *Monitoring Growth Cycles in Market. Driented Eountries.* Cambridge, MA: Balinger Publishing.

Geoffrey H. Moore. 1983. *Business Cgcles, In flation, and Forecasting, 2nd edition,* Cambridge,MA: Ballinger Publishing.

Organization for Economic Cooperation and Development. *Main Economic Indicators*, Monthly, Paris.

بخش چهارم

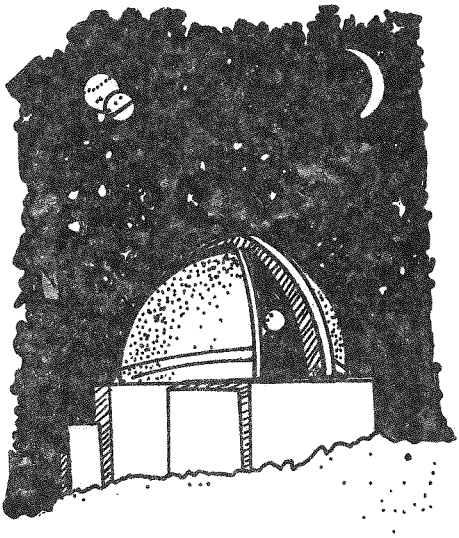
دنیای مادی ما

بهینه‌سازی و مسئله فروشنده دوره‌گرد ۳۶۱

برآوردکردن شانس زلزله‌های بزرگ ازطریق تعیین تاریخ به‌وسیله رادیوکربن ومدل‌بندی آماری ۳۷۵

پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار ۳۹۳

آمار، خورشید و ستارگان ۴۰۳



بهینه‌سازی و مسئله فروشنده دوره‌گرد

چارلز اِی. ویتنی[1] دانشگاه هاروارد

مقدمه

یک اخترشناس در مرکز فرمان یک تلسکوپ، اغلب میل دارد اندازه‌گیریهای کوتاه‌مدتی را برای فهرست بلندی از ستارگانی که در پهنه آسمان پراکنده‌اند، انجام دهد. حرکت دادن تلسکوپ از ستاره‌ای به ستاره دیگر، فرایندی وقت‌گیر است. بنابراین اخترشناس مزبور سعی می‌کند برای

۱ـ *Charls A. Whitney*

پرهیز از حرکتهای غیرضروری، کاراترین دنباله از "دیدارها" را انتخاب کند. این مسئله بخصوص در مورد یک تلسکوپ فضایی که در بالای جو در حال گردش به‌دور زمین است، تشدید می‌شود. عمل نشانه‌روی در چنین تلسکوپی اغلب از زمین به کمک یک جت گازی کوچک تنظیم می‌شود و سوختی که در دسترس قرار دارد، بشدت محدود است.

این مسئله صورتی از **مسئله فروشنده دوره‌گرد** است، و معمولاً به عنوان مسئله جستجو برای کوتاهترین مسیر بسته بین مجموعه‌ای از شهرها، به‌طوریکه هر شهر فقط یک‌بار مورد بازدید قرار گیرد، بیان می‌شود. این، الگوی نمونه‌ای از یک رده مسائل است که می‌توان آنها را به‌طور خلاصه بیان کرد اما راه حل مستقیم آنها چالش‌برانگیز است و تنها می‌توان به کمک رشته حدسهای عالمانه حلشان کرد. این حدسها اغلب بر محاسبات احتمالاتی مبتنی هستند. این خاصیت، چنین مسائلی را از اغلب مسائل طرح شده در این کتاب، متمایز و در عین حال مسائل عملی زیادی در زمینه‌های متفاوتی از اکتشافات زمین‌شناسی گرفته تا علم الکترونیک را توصیف می‌کند.

یک راه حل مستقیم برای این مسئله آن است که شهرها را روی یک نقشه مشخص کرده، تمام مسیرهای ممکن را اندازه‌گیری، و کوتاهترین آنها را انتخاب کنیم. با شمارش تعداد مسیرهای ممکن، معلوم می‌شود که گفتن این کار، آسانتر از انجام دادن آن است. اگر N شهر داشته باشیم که نقطه شروع آنها از قبل مشخص شده‌باشد، برای اولین نقطه توقف $N-1$ انتخاب، برای دومین نقطه توقف $N-2$، و به همین ترتیب الی آخر وجود دارد که تعداد کل مسیرهای ممکن برای سفر به $N-1$ شهر مزبور و بازگشت به نقطه شروع عبارت از:

$(N-1)\times(N-2)\times(N-3)\ldots 2\times 1$ خواهد بود. حتی اگر فقط ۸ شهر داشته باشیم، آنگاه باید:

$۷\times۶\times۵\times۴\times۳\times۲\times۱ = ۵۰۴۰$ مسیر ممکن مورد بررسی قرار گیرند. با افزایش تعداد شهرها این عدد به سرعت افزایش می‌یابد. اگر تنها دو گروه هشت‌تایی از شهرها، یعنی ۱۶ شهر داشته باشیم، آنگاه بیش از یک تریلیون مسیر خواهیم داشت! و این در حالی است که ۱۶، عدد خیلی بزرگی نیست. بنابراین باید آشکارا فکر محاسبه مستقیم تمام راههای ممکن را کنار گذاشت.

مثال دیگر طراحی مدارهای الکترونیک است که در آن چند دوجین قطعه باید به‌وسیله سیم به‌هم متصل شوند. طراح به روشی نیازمند است که با استفاده از آن یک آرایش موثر با مینیمم

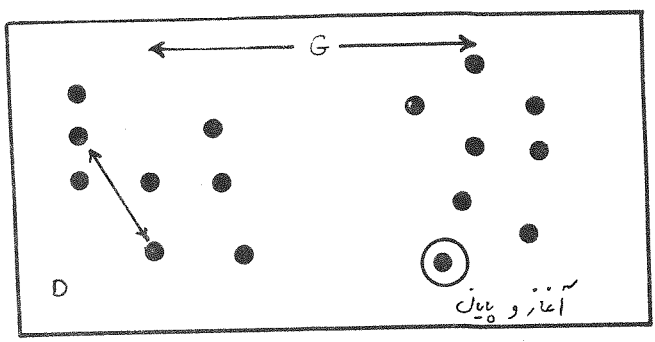
مقدار سیم مورد نیاز را بیابد. در مدارهای امروزی تعداد آرایشهای ممکن اغلب از یک تریلیون تجاوز می‌کند، و عمر این جهان هم برای بررسی تمام راههای ممکن به منظور یافتن بهترین آنها برای طراح کافی نیست.

رهیافتی به حل مسئله فروشنده دوره‌گرد

این نوع مسئله جنبه‌ای دارد که ما را برای یافتن یک راه حل عملی امیدوار می‌سازد. به‌طور شهودی می‌دانیم که علاوه بر بهترین راه‌حل، راه‌حلهای متعدد دیگری وجود دارند که آنها هم تقریباً به همان خوبی هستند. بنابراین بسیاری از راه‌حلهای خوب که بهترین را هم شامل می‌شود، تقریباً طول مسیر یکسانی دارند. به عبارت دیگر، تکنیکی که ما را به جواب بهینه نزدیک کند، بدون نگرانی زیاد از اینکه بهترین مسیر را نیافته‌ایم، برای ما رضایتبخش خواهد بود. برای مسائلی از این قبیل، اغلب می‌توانیم فرض کنیم که بهترین دوم (یا بهترین هزارم) به اندازه کافی خوب است.

درباره راه‌حلهای ممکن مسئله فروشنده دوره‌گرد بیش از این چه می‌توانیم بگوییم؟ احساس ما می‌گوید (و تجربه هم تایید می‌کند) که فروشنده باید سفر خود را در هر قسمت از

شکل ۱ـ طرح نقشه ۱۶ شهر. مسئله فروشنده دوره‌گرد، یافتن کوتاهترین مسیر از یک شهر معین (که با دایره مشخص شده است)، به‌طوری که از تمام شهرهای دیگر فقط یک بار عبور کند. در این مثال، شهرها به دو گروه با فاصله G مایل تفکیک شده‌اند. در داخل هر گروه میانگین فاصله شهرها از هم D مایل است.

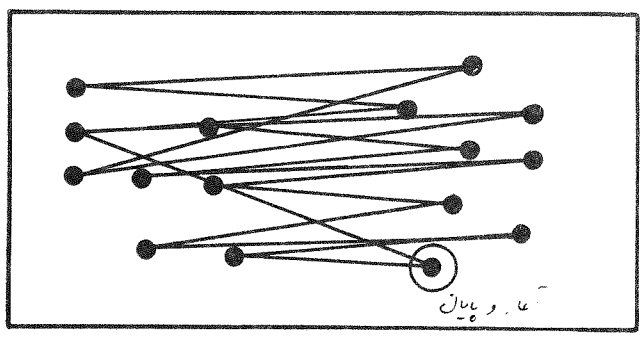


نقشه، قبل از حرکت به قسمتی دیگر، کامل کند. به عنوان یک تمثیل افراطی، فرض کنید همان‌طور که در شکل ۱ نشان داده شده، شهرها در دو گروه قرار داشته باشند ($\frac{N}{۲}$ در هر گروه)، به‌طوری که در یک فاصله G مایلی از هم قرار داشته باشند. مسیری که به نوبت از این گروه به آن گروه می‌پرد قطعاً انتخاب کارایی نخواهد بود (شکل ۲ را ببینید). طول این مسیر تقریباً برابر حاصلضرب $N \times G$ خواهد بود و لذا فروشنده بهتر است که قبل از حرکت به سمت گروه دیگر، یک گروه شامل $\frac{N}{۲}$ از شهرها را، (مانند شکل ۳) به اتمام برساند.

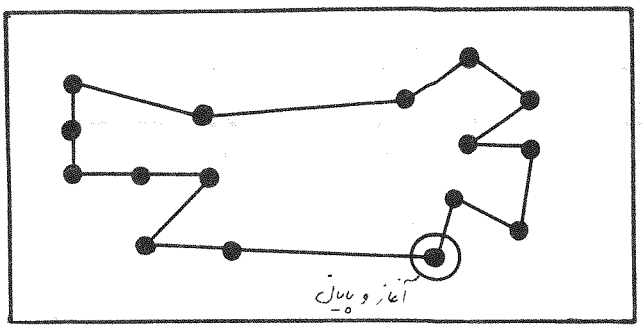
به عنوان مثال؛ فرض کنید D متوسط فاصله بین شهرهای درون هر گروه باشد و فروشنده از یکی از شهرهای معین گروه اول شروع کند. بازدید از بقیه شهرهای گروه اول مسیری به طول $D\times(\frac{N}{۲}-۱)$ ایجاد خواهد کرد. سپس او G مایل به سمت یکی از شهرهای گروه دوم رفته و پس از آن $D\times(\frac{N}{۲}-۱)$ مایل برای دیدن بقیه شهرهای آن گروه طی خواهد کرد. نهایتاً، وی برای بازگشت به گروه اول G مایل و در کل $۲G+D\times(N-۲)$ مایل طی خواهد کرد. شکلهای ۲ و ۳ طولهای این دو راه‌حل را با هم مقایسه می‌کنند و نشان می‌دهند که کمی دقت موجب کاهش قابل توجهی در طول مسیر خواهد شد.

بنابراین فروشنده باید هر ناحیه را قبل از حرکت به سمت نواحی دیگر طی کند، ولی چگونه یک مسیر خوب را در داخل هر ناحیه بیابد؟

شکل ۲ـ مثالی از یک راه‌حل نامطلوب، که در آن فروشنده بین شهرهای دو گروه به نوبت رفت و برگشت می‌کند. طول مسیر او تقریباً $N \times G$ مایل است، که در آن N تعداد کل شهرهاست. شکل ۳ راه‌حل بهتری را نشان می‌دهد.



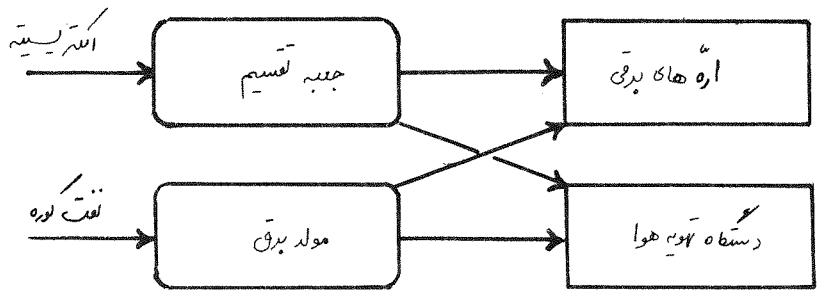
شکل ۳ـ مثالی از یک راه‌حل خوب، که در آن فروشنده از تمام شهرهای یک گروه، قبل از حرکت به سمت گروه دیگر بازدید می‌کند. همان طور که در متن نشان داده شده، طول مسیر او در این حالت تقریباً $L' = 2G + D \times (N-2)$ است که از فاصله $L = N \times G$ برای مسیر شکل ۲ کمتر است. به عنوان مثال، فرض کنید $N = 16$، $G = 8$ و D برابر ۲ مایل باشد. آنگاه در این صورت $L = 128$ مایل برای مسیر شکل ۲، و $L' = 44 = 14 \times 2 + 8 \times 2 + 2$ مایل خواهد بود. بنابراین طول مسیر بهتر حدوداً $\frac{1}{3}$ مسیر نامطلوب است.



قبل از مواجه شدن با این مسئله، گونه عمومی‌تری از این مسئله را که کاربردهای صنعتی و علمی زیادی دارد، بررسی می‌کنیم. ممکن است به‌نظر برسد که این مسئله از نوع جدیدی است، ولی صرفاً صورتی تغییریافته از مسئله فروشنده دوره‌گرد است.

هزینه سوخت عملیات شکل ساده شده‌ای از یک کارخانه تولیدی را که طرح آن در شکل ۴ نشان داده شده، در نظر بگیرید. این کارخانه، برق مورد نیاز تجهیزات برقی خود مثل اره برقی و دستگاه تهویه هوا را از یک اداره برق عمومی تأمین می‌کند. این کارخانه همچنین برای استفاده از یک مولد برق، سوخت مایع می‌خرد تا مقداری برق و مقداری گرما تولید کند که در عین حال برای کنترل دما از آن استفاده می‌کند. با معلوم بودن هزینه‌های برق و سوخت، و همچنین معلوم بودن کارایی دستگاهها (که به مقدار کار ایجاد شده به‌وسیله هر دستگاه بستگی دارد) چگونه مدیر کارخانه منابع مالی خود را برای خرید سوخت اختصاص دهد؟ و چگونه باید از انواع مختلف سوختها به بهترین نحو، در کارخانه استفاده کند؟ و اگر کارخانه به تهویه هوا با ظرفیت بیشتر نیاز دارد، چه نوع دستگاه جدیدی باید به‌کارخانه وارد شود؟ در یک مسئله پیچیده

● *شکل ۴: این نمودار از کارخانه‌ای که اره‌های برقی و یک دستگاه تهویه هوا را به‌کار می‌اندازد، یک مثال ساده شده صنعتی از مسئله توصیف شده در مقاله را ارائه می‌دهد. برق و بنزین، قیمتها و کارائیهای متفاوتی دارند و مدیر کارخانه لازم است در مورد اقتصادی‌ترین مقدارهای برق و بنزین که باید بخرد، تصمیم بگیرد. پاسخ به تصمیم در مورد مقدار برق تولید شده به‌وسیله مولد خود کارخانه بستگی خواهد داشت. در یک کارخانه واقعی، مسئله اغلب خیلی پیچیده‌تر از آن است که بتوان آن را مستقیماً حل کرد و مدیر بناچار باید به فنون جستجوی تصادفی توسل جوید.*



واقع‌گرایانه، راه‌حل چندان بدیهی نیست.

توصیف کلی مسئله بهینه‌سازی

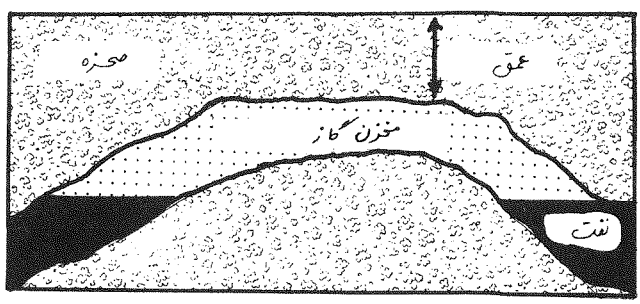
این مسئله چندین خصوصیت مشترک دارد. اولاً، برای هر کدام از آنها می‌توانیم یک تابع هزینه که به‌دنبال مینیمم کردن مقدار آن هستیم، تعریف کنیم. این تابع ممکن است طول یک مسیر یا هزینه سوخت، یا کل زمان لازم برای نشانه‌روی یک تلسکوپ به‌سوی مجموعه‌ای برگزیده از ستاره‌ها باشد. ثانیاً، هر مسئله به‌وسیله تعدادی پارامتر ثابت مشخص می‌شود که از کنترل ما خارج است، مانند موقعیت ستاره‌ها در آسمان، یا سوخت مورد نیاز دستگاههای کارخانه و سهمیه تولید که مدیریت آنها را تعیین کرده‌است. اینها قیودات مسئله هستند و راه‌حلهای ممکن را محدود می‌سازند. ثالثاً، اعداد قابل تغییری وجود دارند که انتخاب مقادیری از آنها راه‌حل مسئله را تشکیل می‌دهد. به عنوان مثال؛ ما ترتیبی از ستاره‌ها را تنظیم می‌کنیم تا زمان مینیمم شود. یا خرید سوخت و انتساب خروجی دستگاههای کارخانه را تنظیم می‌کنیم تا کل هزینه

عملیات را مینیمم کنیم.

با ظهور کامپیوترهای نسبتاً ارزان، انجام چنین جستجوهایی به اصطلاح با کورمالی در تاریکی ممکن شده است. این فرایند بسیار شبیه جستجوی تصادفی برای یافتن عمیقترین قسمت یک دریاچه است.

برای بررسی مسئله‌ای که در اکتشاف زمین‌شناسی پیش می‌آید، فرض کنید درصدد اکتشاف گاز طبیعی باشیم و بدین منظور در جستجوی نوک مخزن گنبدی شکل گاز هستیم که در بالای یک معدن نفت محبوس است (شکل ۵الف). برای به‌دست آوردن بیشترین مقدار گاز، مایلیم که آن را از بالاترین بخش مخزن گاز که گنبدی شکل است استخراج کنیم. فرض کنید وسیله‌ای در اختیار داریم که با آن می‌توانیم ارتفاع مخزن از سطح نفت را در هر نقطه‌ای اندازه بگیریم. در نتیجه در جستجوی نقطه‌ای هستیم که ماکسیمم ارتفاع را داشته باشد. برای سازگاری واژگان، می‌خواهیم مسئله را به‌صورت مینیمم کردن یک تابع هزینه بیان کنیم. پس عمق مخزن در هر گنبدی نسبت به سطح تراز معینی را به‌عنوان تابع هزینه تعریف کرده و به‌دنبال مینیمم کردن آن

شکل ۵الف) نمودار یک معدن فرضی نفت با مخزنی از گاز طبیعی بر روی آن. عمق مخزن گاز در زیر سطح زمین مشخص شده است. وقتی گاز تخلیه می‌شود، نفت در مخزن گنبدی بالا خواهد آمد. بنابراین شیر تخلیه باید در بالاترین نقطه نصب شود تا تمام گاز تخلیه شود. در متن مقاله یک عمل جستجوی تصادفی برای یافتن قله، در جایی که عمق مینیمم است توصیف شده‌است.



هستیم.

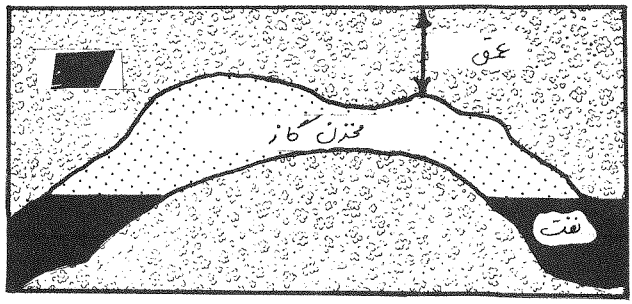
اگر اطلاعات ما درباره وضع مخزن بسیار کم باشد، از یک نقطه دلخواه شروع می‌کنیم. در نقطه شروع عمق را اندازه‌گیری کرده و سپس با طی یک مسافت کوتاه دوباره عمق را می‌سنجیم. اگر عمق کاهش یافت، در همان جهت به حرکت خود ادامه می‌دهیم. اگر عمق افزایش یابد، تغییر جهت می‌دهیم. انتخاب جهت تازه، می‌تواند بر اساس چینه‌بندی محل یا حدس کور باشد. در هر صورت، این فرایند را آنقدر ادامه می‌دهیم تا به جایی برسیم که حرکت به هر طرفی موجب افزایش عمق شود. در این صورت می‌دانیم که نزدیک قله مخزن گنبدی شکل قرار داریم. همه نقاطی که با قله فاصله کمی دارند (مثلاً چند متر) تقریباً همان عمق را خواهند داشت، زیرا می‌توانیم ناحیه بالای مخزن در حوالی قله را افقی فرض کنیم. بنابراین هر نقطه‌ای در این ناحیه را می‌توانیم به‌عنوان یک جواب بپذیریم، زیرا معنی ندارد که بر یافتن قله با تقریب یک سانتیمتر پافشاری کنیم.

اگر مخزن گنبدی، شکلی واقعاً کروی یا استوانه‌ای داشته باشد، می‌توانیم نسبتاً مطمئن باشیم که به همسایگی قله دست یافته‌ایم. اما فرض کنید شکل آن پیچیده‌تر باشد، (مثل شکل ۵ب) در آن صورت، ممکن است یک قله موضعی را یافته باشیم. چگونه می‌توان از اشتباه گرفتن یک قله موضعی با قله واقعی پرهیز کرد؟ رهیافت احتمالاتی، کلید حل مشکل گرفتار شدن در یک قله اشتباهی است.

توصیف عمل جستجوی احتمالاتی بهینه

در یک جستجوی احتمالاتی با تصحیحات پی در پی پیش می‌رویم، از یک جواب حدس زده‌شده شروع می‌کنیم و جوابهای تازه و تصادفی انتخاب شده را امتحان می‌کنیم، تا به نتیجه رضایتبخشی برسیم. برای اینکه ببینیم چگونه کار می‌کند، فرض کنید C تابع هزینه‌ای مثل طول مسیر یا عمق موضعی مخزن گنبدی گاز باشد که به‌منظور یافتن جواب باید مینیمم شود. برای اولین حدس، C را ارزیابی می‌کنیم و سپس یک جواب تازه را به‌طور تصادفی یا با استفاده از هر اطلاعی که داریم، تعیین می‌کنیم. این امر معادل حرکت به یک نقطه دیگر روی مخزن گنبدی گاز است. سپس تابع هزینه (عمق) را دوباره ارزیابی می‌کنیم، و C' را به‌دست می‌آوریم. اگر مقدار

شکل ۵ب) شبیه شکل ۵الف، با این تفاوت که مخزن، قله دومی دارد که ممکن است با قله واقعی اشتباه شود. روش جستجو باید به‌گونه‌ای طراحی شود که از توقف در قله دوم احتراز شود، که در آن صورت گاز محبوس در قله بالاتر، از دست خواهد رفت.



جدید کوچکتر یا مساوی مقدار قبلی بود یعنی $C' \leq C$، نقطه جدید را به‌عنوان نقطه شروع جدید می‌پذیریم. از طرف دیگر، اگر عمق تازه بزرگتر باشد، یعنی $C' > C$ آنگاه بر اساس یک محاسبه احتمالاتی به‌ترتیب زیر، تصمیم می‌گیریم که نقطه جدید را بپذیریم یا آن را رد کنیم. یک عدد تصادفی به‌دست می‌آوریم (مثلاً شبیه نتیجه پرتاب یک جفت تاس)، و اگر این عدد از عدد دیگری که قبلاً انتخاب کرده‌ایم کوچکتر بود، از نقطه قبلی دوباره شروع می‌کنیم. این فرایند را چندین بار با حرکت از یک نقطه به نقطه دیگر تکرار می‌کنیم.

فعلاً نگرانی در مورد جزئیات محاسباتی که ما را به قبول یا رد یک جواب جدید می‌رساند، کنار می‌گذاریم، زیرا برای درک این روش حیاتی نیست. نکته اساسی این است که تصمیم احتمالاتی به ما اجازه می‌دهد گهگاه به نقطه جدیدی برویم، حتی اگر افزایش اندکی در C ایجاد شود. این امر ما را از گیر افتادن در مینیمم موضعی، قبل از رسیدن به همسایگی مینیمم مطلق نجات می‌دهد. این راهی برای اجتناب از توقف در قله دوم مخزن گنبدی گاز است. همیشه قدمهایی که موقتاً ما را به نقطه عمیقتری از مخزن گنبدی می‌برد، مجاز می‌شماریم زیرا بعضی مواقع این تنها راه رسیدن به قله واقعی است.

علی‌رغم این واقعیت که فرایند مزبور تا حدی شبیه کورمال گشتن در تاریکی است، اما به دو

دلیل قوت زیادی دارد. اولاً، این روش در میان نمونه کوچکی از جوابهای ممکن جستجو کرده و جوابهایی را انتخاب می‌کند که تقریباً بهترین هستند. این امر زمان جستجو را به میزان قابل ملاحظه‌ای کاهش می‌دهد. ثانیاً، این روش برای یافتن یک جواب عملی در مورد هر مسئله‌ای که قابل بیان بر حسب تابع هزینه C و مجموعه قابل تنظیمی از پارامترها باشد، به‌کار می‌آید. تا زمانی که یک کامپیوتر در اختیار باشد، لزومی ندارد که نگران سادگی یا پیچیدگی محاسبه عددی تابع هزینه باشیم. هیچ محدودیتی هم وجود ندارد، و این دلیلی برای مفید بودن جستجوی تصادفی در مسائل گوناگونی همچون تولید، اکتشافات زمین‌شناسی، و نجوم است. برای کسانی که علاقه‌مند باشند، برخی از جزئیات فنی تصمیم احتمالاتی برای قبول یا رد هر جواب جدید را بیان می‌کنیم. ما آن را بر حسب جستجوی تصادفی‌شده قله مخزن گنبدی گاز، تعبیر خواهیم کرد.

پس از آنکه هر نقطه جدید انتخاب شد و عمق آن C' محاسبه شد، تغییر هزینه $C' - C$ را بر عددی مانند T که با C هم‌واحد است، تقسیم می‌کنیم. (انتخاب مقدار T بر حسب تجربه صورت می‌گیرد. این مقدار باید برای مسئله مورد نظر مناسب باشد.) سپس یک عدد تصادفی مانند r بین صفر و یک انتخاب می‌کنیم، $0 \leq r < 1$ و لگاریتم طبیعی آن را (که همیشه منفی خواهد بود) گرفته و با $\frac{(C'-C)}{T}$ مقایسه می‌کنیم. اگر

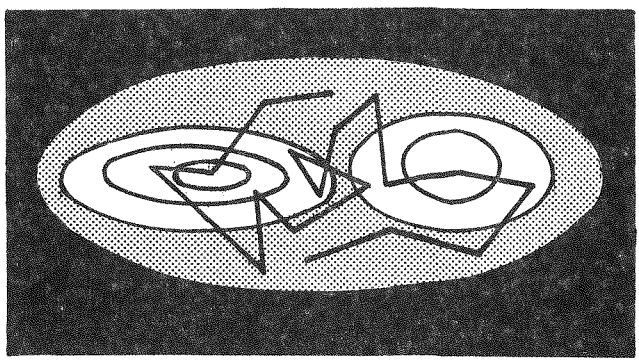
لگاریتم r $< \frac{(C'-C)}{T}$

بود، مسیر جدید را قبول و در غیر این صورت آن را رد می‌کنیم. این فرمول خاص از آن‌رو انتخاب شده که نسبت به افزایش اندک در مقدار C با مسامحه عمل می‌کند، در حالی که تنها اجازه می‌دهد C دفعه‌های بسیار کمی افزایش زیاد در مقدار داشته باشد. احتمال پذیرفتن یک افزایش در C به مقدار T بستگی دارد. اگر T بزرگ باشد، این فرمول افزایشهای بزرگ را مجاز می‌شمارد. این بدان معناست که تقریباً هر امتحان جدیدی که انجام می‌دهیم به‌عنوان نقطه شروع جدیدی پذیرفته خواهد شد. برای شروع جستجو این راه خوبی است، زیرا می‌توان آزادانه در اطراف مخزن گنبدی گشت. در همان حال که محاسبات پیش می‌رود، از اینجا و آنجای مخزن گنبدی شکل نمونه‌گیری می‌کند و از این قله به آن قله می‌رود. معیار رد کردن، بتدریج با کم کردن مقدار T تنظیم می‌شود و پذیرش یک افزایش در C را غیر محتمل‌تر

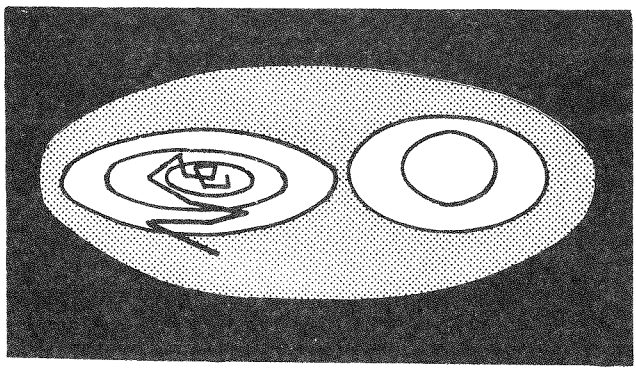
می‌سازد. بالاخره، زمین‌شناسی که قبلاً از بیشتر نقاط مخزن نمونه گرفته است، نمی‌خواهد به جایی که مخزن ارتفاع کمتری داشته است، بازگردد. ایجاد برنامه‌ای برای کاهش مقدار T امری است تجربی؛ اگر T بسرعت کاهش یابد، فرایند ممکن است در یک قله اشتباه متوقف شود و اگر خیلی کند کاهش یابد، زمین‌شناس به‌دفعات سرگردان خواهد شد. به‌طور ایده‌آل، کاهش T بتدریج جستجو را به همسایگی قله واقعی محدود خواهد ساخت. فرایند وقتی متوقف می‌شود که T کوچک شده و جستجو به ناحیه کوچکی از مخزن گنبدی محدود شود. زمین‌شناس می‌تواند با تکرار کامل جستجو و دیدن اینکه آیا به همان ناحیه می‌رسد یا نه، از واقعی بودن قله به‌دست آمده مطمئن شود.

اثر تغییر دادن T با ترسیم دو جستجوی فرضی در شکلهای ۶الف و ب تشریح شده‌است. در حالت اول که برای مراحل اولیه جستجو مناسب است، مقدار بزرگ T اجازه می‌دهد که بیشتر نقاط مخزن گنبدی بررسی شود. وقتی که جستجو به حوالی قله واقعی برسد، مقدار T کاهش می‌یابد که موجب گیرافتادن جستجو در بهترین ناحیه می‌شود.

شکل ۶الف) نقشه تراز پشت‌بام یک مخزن گنبدی گاز با دو قله. ناحیه سیاه نشان دهنده سطح نفت است. سایه کم رنگتر نشان‌دهنده سطح رویی است و قله واقعی در سمت چپ قرار دارد. یک جستجوی تصادفی با مقدار بزرگ پارامتر T، با خط شکسته نشان داده شده است. این مسیر جستجوی بخش بزرگی از مخزن گنبدی را دربر می‌گیرد، و تا زمانی که T کاهش نیابد در یکی از قله‌ها متوقف نخواهد شد.



شکل ۶ب) شبیه شکل ۶الف .. این حالت مقدار کوچکی برای T به‌کار گرفته شده و این کار جستجو را به گامهایی محدود می‌کند که به سمت نزدیکترین قله حرکت می‌کند. بنابراین ممکن است در قله اشتباهی متوقف شود. اگرچه در این مثال قله بلندتر را یافته است.



نتیجه‌گیری

روش جستجوی تصادفی، برخلاف کور بودن ظاهری آن، بسیار توانمند از کار درآمده‌است. کلید موفقیت آن دو جنبه دارد. از میان تعداد کل جوابهای ممکن یک نمونه کوچک می‌گیرد، و از اینکه در دام جوابهای اشتباه نیفتد، پرهیز می‌کند. برای هر نوع مسئله‌ای که بتوان آن را بر حسب یک تابع هزینه و مجموعه‌ای از اعداد قابل تنظیم بیان کرد، این روش یک جواب خوب به‌دست می‌آورد. اینکه جواب به‌دست آمده چقدر به بهترین جواب ممکن نزدیک شود، به پافشاری حل‌کننده مسئله بستگی دارد، لکن در بیشتر مسائل عملی، یک جواب خوب کاملاً کافی است.

مسائل

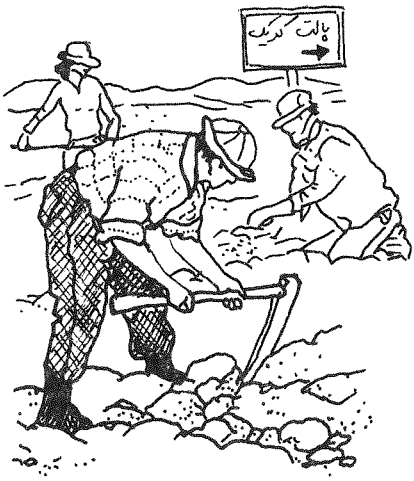
۱ـ چندین مسئله را که برای روش جستجوی تصادفی مناسب هستند، فهرست کنید. مسائلی که به این نوع راه‌حل نیاز دارند دارای چه مشخصه‌هایی هستند؟

۲ـ تصور کنید که شما توزیع‌کننده روزنامه هستید و مایلید کاراترین ترتیب توزیع را بیابید. برخی از راههایی که ممکن است به‌عنوان مسیرهای آزمایشی بسازید کدامند؟ فرض کنید شما در

حومه شهری زندگی می‌کنید که در آنجا خانه‌ها در حاشیه چندین خیابان طویل با تعداد کمی خیابان عمود به خیابان اصلی قرار دارند. این امر بر فرایندی که شما با آن مسیرهای آزمایشی را می‌سازید، چه تأثیری می‌گذارد؟

منابع:

S.Kirkpatrick, C.D.Gelatt. Jr., and M.P. Vecchi (1983), " *Optimization by Simulated Annealing.*" Scince 220: 671-680



برآوردکردن شانس زلزله‌های بزرگ از طریق تعیین تاریخ به‌وسیله رادیوکربن و مدل‌بندی آماری

دیوید آر. برلینجر(۱) دانشگاه کالیفرنیا، برکلی

ساکنان مناطق زلزله‌خیز نگران این اتفاق هستند که شاید زلزله‌ای رخ دهد و موجب خسارت مالی یا جانی برای آنها شود. آنها به بیمه توسل می‌جویند تا از آثار این مخاطره بکاهند. مقامات دولتی هم نگران این امرند؛ زیرا آنها در مورد برنامه‌ریزی برای استمرار خدمات درصورت آسیب رسیدن به تسهیلات حساس و در مورد آموزش عموم مسئولند. مقدم بر این محاسبات

۱ـ *David R. Brillinger*

حق بیمه و تخصیص منابع دولتی، برآوردهایی برای شانس زلزله‌ها و تخریب ناشی از آنهاست.

خوشبختانه، زلزله‌های بزرگ نادرند. با این حال متأسفانه، نادربودن آنها این عیب را از نظر آماری دارد که برآورد شانس وقوع آنها را با اطمینان خاطر، مشکل می‌کند. چندین شیوه عمل برای ارزیابی خطر زلزله به‌وجود آمده‌اند. این مقاله رهیافتی بین رشته‌ای را توصیف می‌کند با این جنبه اعجاب‌انگیز که بر پایه داده‌هایی برای زلزله‌هایی است که در یک مکان موردنظر در زمانی رخ داده که کسی برای ثبت آن حادثه حضور نداشته‌است. درواقع، نه مورد از ده زلزله که در این مطالعه به‌کار رفته‌اند، به پیش از تاریخ مربوطند.

پالت کریک (۱)

این ماجرا با عزیمت کری‌سی‌یه(۲)، زمین شناس دانشگاه استانفورد به صحرای کالیفرنیا، درست پس از ماه عسلش، همراه با همسر وبرادرش آغاز شد. مقصد پروفسور سی یه قطعه زمین کوچکی درحدود ۵۵ کیلومتری شمال شرقی لوس‌آنجلس بود که بر گسل سن آندریاس سوار بود (برای ملاحظه مکان در مقیاس گسترده، ن. ک. شکل ۱). نهری به نام پالت کریک در همان حوالی جریان دارد. تا سال ۱۹۱۰ یا همان حدودها، این منطقه یک باتلاق بود. طی سالیان، توربهای سیاه تشکیل شدند و به‌طور دوره‌ای روی آنهارا شن و ماسه‌ای که سیلابهای آن نهر باخود حمل می‌کردند، پوشاند. سی‌یه و همراهانش به کندن گودالها پرداختند. آنها لایه‌های ازهم گسیخته تورب، تکه‌های چوب، ذغال سنگ، وحتی لانه‌های قدیمی جانوران را یافتند. سی‌یه با امتحان کردن دیوارهای گودالها متوجه جاهایی شد که در آنها لایه‌ها درهم شکسته بودند و چنین استنباط کرد که این شکستگیها طی زلزله‌های پیش از تاریخ به‌وقوع پیوسته است. همه آموزشهای حرفه‌ای و تبحر او به‌عنوان یک زمین‌شناس وی را یاری دادند تا تصمیم بگیرد که کدامیک از گسیختگیها در لایه‌ها ممکن است بر اثر زلزله باشند. وی نمونه‌هایی از اطراف شکستگیها برگزید تا تاریخ آنها را به کمک فنون رادیوکربن تعیین کند.

براساس سوابق، متأخرترین زلزله بزرگ که بر منطقه پالت کریک اثر گذارد، در ۱۸۵۷ بود،

۱ـ *Pallett Creek*

۲ـ *Kerry Sieh*

شکل ۱ـ نقشه کالیفرنیا که مکان جایگاهی در گسل سن‌آندریاس را نشان می‌دهد و نمونه‌ها برای تعیین تاریخ به‌کمک رادیوکربن ازآنجا گردآوری شده‌اند. pc مکان پالت‌کریک، LA لوس‌آنجلس و UCB دانشگاه کالیفرنیا در برکلی را نشان می‌دهند.



(زلزله بزرگ، هر زلزله‌ای با قدرت ۷/۵ ریشتر یا بیشتراست.) همچنین مطالعه زلزله‌ها در پالت کریک در ارتباط با پیشامدهای مخربی که ممکن است منطقه لوس‌آنجلس بزرگ را مورد اصابت قراردهند، بسیار آگاهی‌دهنده است.

تجزیه وتحلیل به‌کمک رادیوکربن

پس ازبازگشت از پالت‌کریک، سی‌یه نمونه‌هارا برای پروفسور مینزه *استوی‌ور*[۱] در

۱ـ *Minze Stuiver*

"آزمایشگاه ایزوتوپ دوره چهارم زمین شناسی"[۱] در دانشگاه واشنگتن فرستاد. وظیفه پروفسور *استوی‌ور* این بود که برآوردی برای تاریخی تعیین کند که هرنمونه رسوبی شده‌بود (مرده بود). این کار در دو مرحله انجام می‌شود. در مرحله نخست، *استوی‌ور* از فنی به نام **تاریخگذاری رادیوکربن** استفاده می‌کند. این کار تقریبهای اولیه‌ای برای تاریخهای مرگ نمونه‌ها به او می‌دهد. در مرحله دوم، وی از یک فن واسنجی[۲] برای بهبود تقریب استفاده می‌کند. فنون آماری در هر دو مرحله نقشی مهم دارند.

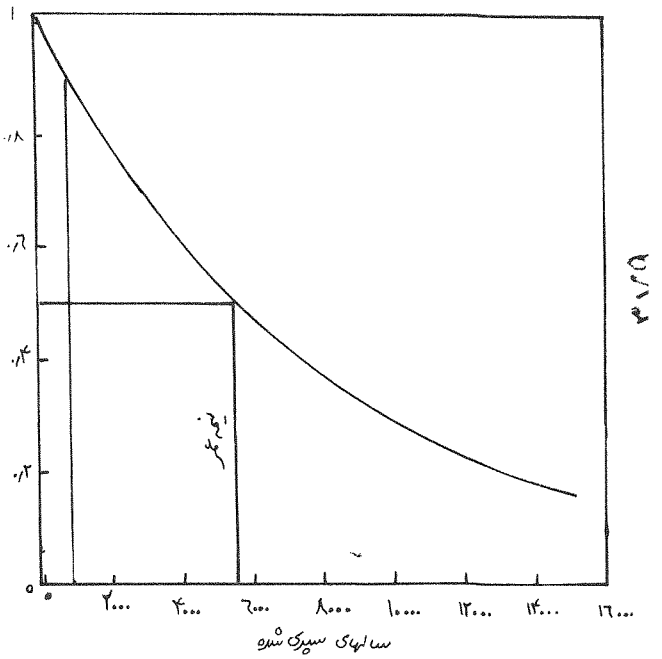
در نخستین گام، *استوی‌ور* نمونه‌های سی‌یه را به "خالصترین دی‌اکسیدکربن در سیاتل" تبدیل می‌کند و سپس سطح رادیواکتیویته آنها را اندازه‌گیری می‌نماید. وی می‌خواهد دریابد که چقدر رادیوایزوتوپ، کربن ۱۴ (رادیوکربن)، در هرنمونه وجود دارد؟ او از فنی پیروی می‌کند که پروفسور *ویلیام لیبی*[۳] از دانشگاه شیکاگو بنیاد نهاد. *لیبی* می‌دانست که ماده زنده، مانند یک درخت، مقدار تقریباً ثابتی از کربن ۱۴ را در طول دوره زنده‌بودنش دارد. با این حال به محض مردن درخت، کربن ۱۴ با نرخ معلومی به عنصر دیگری رو به تبدیل می‌گذارد. این موضوع در شکل ۲ نشان داده شده‌است. به‌عنوان مثال، کربن ۱۴ به نصف آنچه در ۵۵۶۸ سال پیش بوده و به یک‌پنجم آنچه ۱۳۰۰۰ سال پیش بوده، کاهش می‌یابد. (در شکل ۲، مقدار ۵۵۶۸ با خط عمودی پررنگ نشان داده شده‌است. این را **نیم‌عمر** می‌نامند). بنابراین، همانگونه که *لیبی* مشاهده کرد، این امکان وجود خواهدداشت که تاریخ مرگ نمونه‌ای را درصورت معلوم بودن مقدارکربن ۱۴ که در آغاز داشته و اینکه حالا چقدر دارد، پیدا کرد. یافتن مقدار اولیه به‌طور مستقیم غیرممکن است. بنابراین در مرحله اول، *استوی‌ور* این فرض را پیش می‌نهد که مقدار کربن ۱۴ در ماده زنده در طول زمان تقریباً ثابت مانده و از ماده استانداردی (اسیداوکسالیک) برای برآوردکردن اینکه چقدر رادیوکربن در هریک از نمونه‌های سی‌یه در زمان مرگش وجود داشته، استفاده می‌کند. او از این کار نسبتی را به‌دست می‌آورد. اگر این نسبت مثلاً ۰/۹ باشد، در این‌صورت با استفاده از خم شکل ۲، زمان سپری شده تقریباً ۸۵۰

۱- *Quaternary Isotope Laboratory*

۲- *Calibration*

۳- *William LibbY*

شکل ۲ـ نموداری که زوال یافتن نمایی رادیواکتیویته و نیم‌عمر رادیوکربن را نشان می‌دهد. خط عمودی پررنگ سالهای سپری شده متناظر با نسبت ۰/۵ را مشخص می‌کند. این خط نیم‌عمر ۵۵۶۸ را به‌دست می‌دهد. خط عمودی نقطه چین، سالهای سپری شده متناظر با نسبت ۰/۹ را مشخص می‌کند.



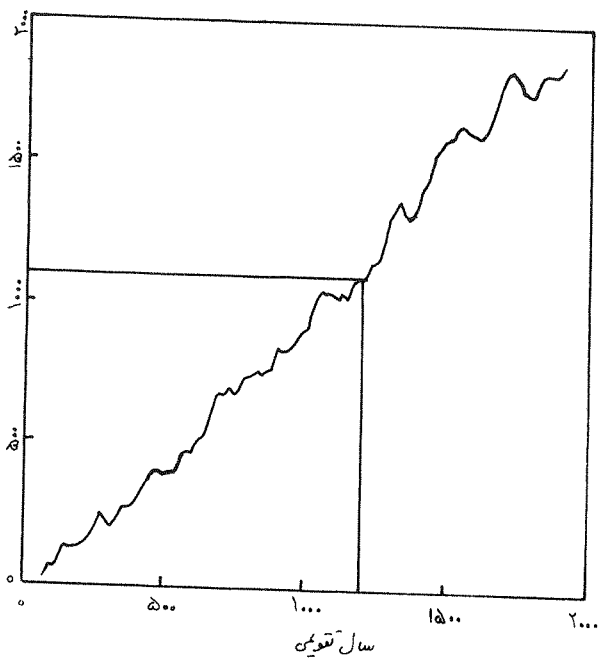
سال است و تاریخ متناظر که نمونه در آن مرده، عبارت است از: ۸۵۰-۱۹۸۷=۱۱۳۷. سال ۱۱۳۷ نخستین تقریب *استوی‌ور* برای تاریخ مرگ نمونه خواهدبود. این عدد را **تاریخ رادیوکربن** نمونه می‌نامند. باید متذکر شد که تصحیحات متنوعی، مثلاً برای پرتوپراکنی زمینه‌ای، نیز ضمن تعیین رادیواکتیویته نمونه‌ها و تاریخ رادیوکربن به‌کار می‌رود.

واسنجی

پروفسور *استوی‌ور* و دیگران مجموعه گوناگونی از تاریخگذاریهای رادیوکربن بر روی نمونه‌هایی (حلقه‌های درخت) با تاریخ معلوم به‌عمل آورده‌اند. آنها دریافته‌اند که فعالیت کربن ۱۴ در جو، همانگونه که *لیبی* در ابتدا فرض کرده‌بود، دقیقاً ثابت نمانده، بلکه تاحدی نوسان

داشته‌است. با دانستن تاریخهای رادیوکربن و تقویمی این نمونه‌ها، پژوهشگران توانستند یک **خم واسنجی** که این دو را به هم مرتبط می‌کرد، تهیه کنند. *استوی‌ور* در مرحله دوم کارش چنین خمی را برای تعیین برآورد بهبود یافته‌ای از تاریخ تقویمی نمونه‌ای ارائه شده به‌کار می‌گیرد. شکل ۳، این خم واسنجی را ارائه می‌کند (ن.ک. *استوی‌ور و پی‌یرسون*، ۱۹۸۶). برای یک تاریخ رادیوکربن مفروض، می‌توان یک تاریخ تقویمی را به‌دست آورد. فرض کنید نمونه‌ای با تاریخ رادیوکربن ۱۱۰۰ (بعد از میلاد) داشته‌باشیم، در این‌صورت تاریخ تقویمی نظیر آن حدود ۱۲۰۰ (بعد از میلاد) است (ن.ک. خطوط افقی و عمودی درشکل). عمل واسنجی بسیار مؤثر بوده تاریخ را حدود ۱۰۰ سال تغییر داده‌است.

شکل ۳ـ خم واسنجی که سالهای رادیوکربن و سالهای تقویمی متناظر را نشان می‌دهد. خط عمودی درمجاورت ۱۲۰۰ سال تقویمی متناظر با یک سال رادیوکربن ۱۱۰۰ را نشان می‌دهد.



استوی‌ور در کار خود باید به خطاهای اندازه‌گیری بپردازد و برآوردهایی از کمیتهای ناشناخته را محاسبه کند. وی همچنین میل دارد تا اندازه‌هایی برای عدم حتمیتها در برآوردهایش فراهم کند. علم آمار فنون گوناگونی برای رویارویی با این مسائل دارد.

رهیافت آماری

در میان مفاهیم بنیادی در رهیافت آماری به مسائل علمی مفهوم **توزیع** در درجه نخست قرار دارد. با فرض اینکه صحبت ازاحتمالهای منضم شده به وضعیت موردنظری معنی داشته باشد، در این صورت توزیع یک کمیت عددی مقدار، عبارت از تابعی است که احتمال آن را می‌دهد که این کمیت مقداری، بیشتر از یک عدد مشخص را اختیار می‌کند. شکل ۴ دو نمونه از تابعهای توزیع تجمعی را نشان می‌دهد، یک توزیع نرمال و یک توزیع وایبول. خانواده توزیعهای فراوانی نرمال شکل خاصی دارند ـ آنها متقارنند و یک کوهان دارند ـ و در توصیف توزیعهای فراوانیی که مشاهدات آنها ممکن است مجموع سهمهای مستقل متعددی نظیر قد، یا نمرات در یک آزمون پیشرفت[۱] را نشان دهد، مفید است. توزیعهای وایبول رده‌ای را تشکیل می‌دهند که دمهای سمت راست درازی دارند و بویژه در توصیف نتایج آزمونهای عمر، نظیر زمان ازکارافتادن لامپهای روشنایی و آزمونهای فرسودگی سودمندند. به‌طور کلی، توزیعهای وایبول می‌توانند توزیع زمان تا پیشامد بعدی را در یک سری از رویدادها توصیف کنند. شکل ۴، توزیعهای تجمعی را به‌عنوان نمونه ارائه می‌دهد، و کوهان خانوادهٔ نرمال با شیب تندی در حوالی میانگین نشان داده شده‌است. دم دراز توزیع وایبول با خیزش کندی در سمت راست شکل نشان داده شده‌است. از نمودار بالایی در شکل ۴، می‌توان دریافت که احتمال رخ دادن مقداری (در این حالت، تاریخ) که بیشتر از ۹۰۰ نیست، ۰/۷۰ است. از نمودار پایینی، برای احتمال نتیجه‌ای (دراین حالت، بازه زمانی) که بزرگتر از ۱۳۰ نیست، ۰/۴۲ به‌دست می‌آید. توزیعها در ساختن **مدلهای آماری** (توصیفات احتمالاتی قابل اداره‌کردن وضعیتهای موردنظر) به‌کار گرفته می‌شوند. با یک مدل آماری، می‌توان جمع عظیمی از پرسشهای علمی را به روالی

۱ـ *Achievement Test*

صوری مورد بررسی قرارداد.

توزیعها عموماً به‌صورت خانواده ظاهر می‌شوند که اعضای انفرادی آن برحسب **پارامترهایی** برچسب‌گذاری می‌شوند. نمودار بالایی شکل ۴ موردی از خانواده **نرمال** را با پارامترهای ۸۹۱/۷ و ۱۵/۷ نشان می‌دهد، درحالی که نمودار پایینی موردی از **وایبول** را با پارامترهای ۲/۵۵ و ۱۶۴/۴ نشان می‌دهد. (این مقادیر خاص پارامترها، در محاسبات مربوط به داده‌های بخشهای بعدی این مقاله به‌کار رفته‌اند.) برای توزیع نرمال، پارامتر اول که مقدار آن ۸۹۱/۷ است، میانگین یا مقدار متوسط، و پارامتردوم انحراف معیار است که یک معیار پراکندگی است. (برای تفصیل بیشتر درباره انحراف معیار ن.ک. مقاله *زابل*.[۱۱]) **در برخی مضامین، اصطلاح خطای معیار به‌جای انحراف معیار اندازه‌ای مانند یک میانگین یا مشاهده دیگر به‌کار می‌رود. در بقیه این مقاله خطای معیار به‌کار خواهد رفت.** پارامترهای توزیع وایبول توصیف نخواهند شد.

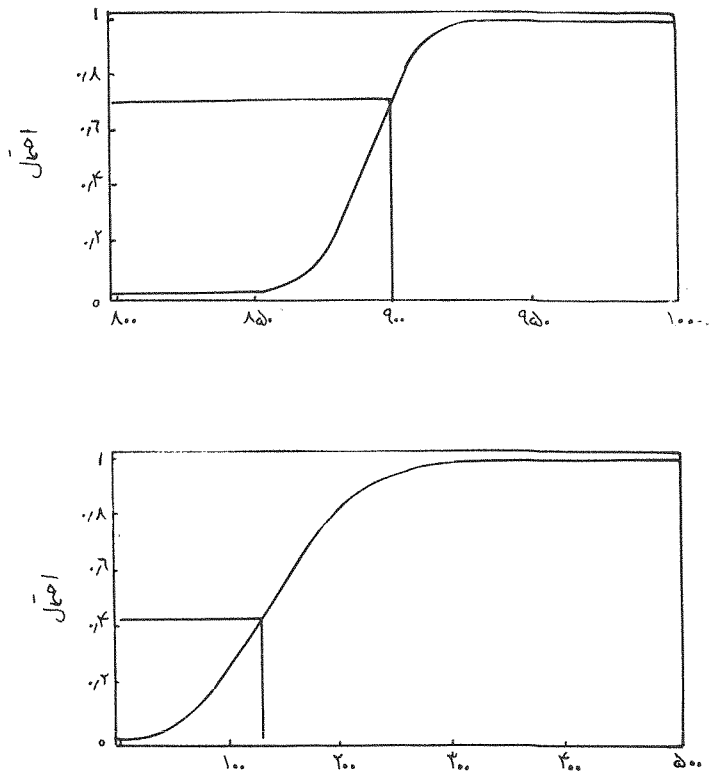
درکار آماری با داده‌ها، یکی از موضوعات اصلی، انتخاب توزیع مناسبی است که باید به‌کار گرفته شود. رسم **نمودار احتمال**، فنی برای تشخیص خانواده‌ای مناسب از توزیعها مناسب‌است. در رسم نمودار احتمال، برآوردی از تابع توزیع در مقابل عضوی از یک خانوادهٔ موردنظر بر روی یک کاغذ نمودار خاص رسم می‌شود. اگر خانواده مناسب باشد، نقطه‌های رسم شده در حول یک خط مستقیم قرار خواهندگرفت. شکل ۶ (که بعداً درباره آن بحث می‌شود)، مثالی ازیک نمودار احتمال وایبول را برای داده‌های موردنظر پروفسور سی‌یه نشان می‌دهد.

به محض اینکه خانواده‌ای از توزیعهابرگزیده شد، دانستن مقادیر پارامتر برای عضو خاصی از خانواده راکه بهترین برازش را با داده‌ها دارد، مورد نیاز است. این امرکه مستلزم برآوردکردن پارامتر است، از طریق **تابع درستنمایی** انجام‌می‌شود.

تابع درستنمایی یک مدل احتمال فرضی ویک مجموعه مفروض از داده‌ها، تابعی خاص از پارامترهای مدل است که وزن شواهد برای مقادیر ممکن مختلف پارامترها را اندازه‌گیری

۱ـ *Zabell*

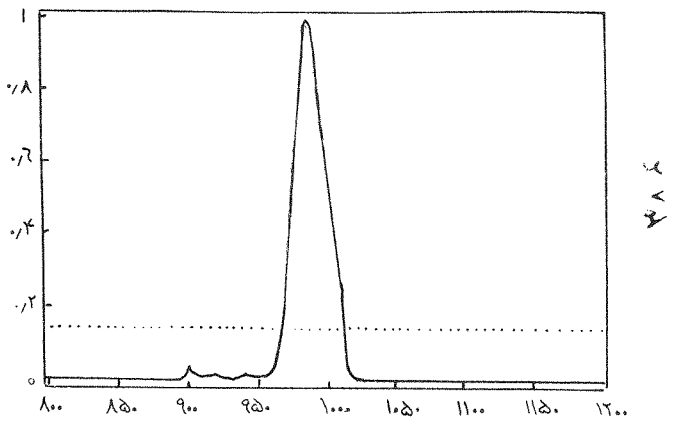
شکل ۴ـ نمونه‌هایی از تابعهای توزیع برای دو توزیع خاص (یک توزیع نرمال ویک توزیع وایبول) این خم احتمال بیشتر نشدن مقداری مشخص در امتداد محور xها را می‌دهد. در نمودار بالایی، احتمال نابیشتر از ۹۰۰ بودن نشان داده شده‌است. در نمودار پایینی، احتمال نابیشتر از ۱۳۰بودن نشان داده شده‌است.



می‌کند. شکل ۵ (که در زیر مورد بحث قرار می‌گیرد) مثالی را دراختیار می‌گذارد که در آن ناشناخته، تاریخ تقویمی یک نمونه مورد بحث است.

پس از آنکه برآوردی از یک پارامتر به‌دست آمد، رسم برآن است که نشانه‌ای از عدم حتمیت آن ارائه شود. یک اقدام خیلی راحت برای انجام این کار ازطریق یک **بازهٔ اطمینان** ۹۵٪ است. این بازه‌ها، بازه‌های عددی‌اند که به این سبک ساخته می‌شوند که در درازمدت ۹۵ ٪ آنها مقدار واقعی پارامتر را واقعاً دربر خواهندگرفت. بازه نقطه‌چین در خم شکل ۵ مثالی از چنین

شکل ۵ـ نموداری از تابع درستنمایی متناظر با نمونه‌ای که تاریخ‌گذاری رادیوکربن آن ۸۹۱/۷ با خطای استاندارد ۱۵/۷ است. قله مرکزی متناظر با برآوردی از تاریخ تقویمی مرگ نمونه، در اینجا ۹۸۶ بعد از میلاد است. خط نقطه‌چین که خم را قطع می‌کند، یک بازه اطمینان ۹۵٪ را برای تاریخ تقویمی نمونه مشخص می‌کند.



تاریخ برآوردشده ۹۸۶ محلهای برخورد با خط دو برابر خطای معیار ۹۶۵/۵، ۱۰۱۱

بازه‌ای است. مفاهیم آماری[۱] بالا نقشی مهم درتحلیلهای توصیف شده دراین مقاله دارند.

تاریخهای برآوردشده

استوی‌ور دریافت که یکی از نمونه های اخیر سی‌یه، تاریخ رادیوکربنی ۸۹۱/۷ با خطای استاندارد ۱۵/۷ دارد. شکل ۵ تابع درستنمایی تاریخ تقویمی رسوب یافتن این نمونه را نشان می‌دهد. در محاسبه این تابع درستنمایی، آمارشناس این وضع را به حساب می‌آورد که هم برآورد تاریخ رادیوکربن نمونه و هم خم واسنجی در معرض خطاهای اندازه‌گیری با توزیعهای تقریبی نرمال هستند. بخشی از خطای تاریخ رادیوکربن بستگی به این دارد که یک دوره زمانی

۱ـ برای توضیح ایده‌های رسم نمودار احتمال، توزیعهای نرمال و وایبول، و مفاهیم آماری وابسته (ن.ک. نلسون ۱۹۸۲)

سطح رادیواکتیویته نمونه طی چه مدت در آزمایشگاه اندازه‌گیری می‌شود. بخشی از خطای خم واسنجی بستگی به این دارد که چندتا از ماده‌های با تاریخ معلوم در ساختن آن منظور شده‌اند؟ تاریخی که برای آن درستنمایی بیشترین مقدار را دارد، در اینجا ۹۸۶ بعد از میلاد است (ن.ک. شکل ۵). این نمونه خاص به‌وسیله سی‌یه برای ارائه تاریخی بین زلزله‌هایی که وی آنها را در جدول ۱ با i و N نشانه‌گذاری کرده‌است، برگزیده شد. بااستفاده از بازه‌ای به طول دو برابر خطای استاندارد در هر طرف برآورد، بازه‌های اطمینان تقریبی ۹۵ ٪ را به صورتی که در جدول ۱ نشان داده شده‌است، به‌دست می‌آوریم. بازهٔ اطمینان ۹۵٪ برای تاریخ تقویمی نمونه، از ۹۶۵ تا ۱۰۱۱ بعد از میلاد تغییر می‌کند. این بازه متناظر با نقاطی است که در آنها خط نقطه‌چین درشکل، خم شکل ۵ را قطع می‌کند. برای درایهٔ اول ۱۸۵۷ هیچ بازه اطمینانی نداریم زیرا این تاریخ، بخشی از تاریخ ثبت شده‌است.

در عمل گاهی یک مشکل اضافی هم پیش می‌آید. خم واسنجی به‌عنوان تابعی از تاریخ تقویمی منظماً افزایش نمی‌یابد. افت وخیزهایی در آن به‌علت چیزهایی مانند اختلالات در حوزهٔ مغناطیسی خورشید، تغییرات درحوزه مغناطیسی زمین، و خطای اندازه‌گیری که به آن اشاره شد، پدید می‌آید. این افت وخیزها به این معنی است که گاهی نمی‌توان یک تاریخ رادیوکربن مفروض را به یک دورهٔ تقویمی منحصربه فردنسبت داد. برای پشت سرهم قرار دادن دوره‌ها، به اطلاعات تکمیلی نیاز پیدا می‌شود.

زمانهای بین رویدادها

سی‌یه (۱۹۸۴) تاریخهای تقویمی برآوردشده زیر را برای ۱۰ زلزله در پالت‌کریک فهرست می‌کند: ۱۸۵۷، ۱۷۲۰، ۱۵۵۰، ۱۳۵۰، ۱۰۸۰، ۱۰۱۵، ۹۳۵، ۸۴۵، ۷۳۵، ۵۹۰. (این تاریخهابه علاوه دوبرابر خطاهای استانداردشان در جدول ۱ نشان داده شده‌اند.) تنها یکی ازاین تاریخها، یعنی ۱۸۵۷ ازنظرتاریخی ثبت شده‌است. تاریخهای دیگر توسط سی‌یه با درونیابی بین تاریخهای برآوردشدهٔ نمونه‌های گوناگونی که وی طی کاوشهای خود برگزید، استخراج شده است.

سی‌یه، در این مرحله از مطالعات خود به یک آمارشناس برای کمک به استنباط درباره

جدول ۱ـ تاریخهای برآوردشده و دوبرابر خطاهای استانداردشان برای زلزله‌های تاریخی در پالت کریک. (برچسب رویدادها و مقدارها مربوط به سی‌یه (۱۹۸۴)است.

رویداد	تاریخ بعد از میلاد
Z	۱۸۵۷
X	۱۷۲۰ ±
V	۱۵۵۰ ±
T	۱۳۵۰ ±
R	۱۰۸۰ ±
N	۱۰۱۵ ±
I	۹۳۵ ±
F	۸۴۵ ±
D	۷۳۵ ±
C	۵۹۰ ±

منبع: سی یه (۱۹۸۴)

احتمالهای زلزله‌ها در آینده مراجعه کرد. (تاریخهای رادیوکربن قبلاً طی مراحل برآورد آنها در اختیار آمارشناسان قرارگرفته بود.) در رهیافت آماری نسبت به مسئله برآورد احتمال، باید در جستجوی تابع توزیعی برای سلسله زمانهای بین رویدادها بود. این زمانها به‌ترتیب کوچک به بزرگ عبارتنداز: ۶۵، ۸۰، ۹۰، ۱۱۰، ۱۳۷، ۱۴۵، ۱۷۰، ۲۰۰ و ۲۷۰ سال، و اکنون در سال[۱] ۱۹۸۸ از زمان رویداد سال ۱۸۵۷، ۱۳۱ سال گذشته است. آمارشناس عزم خود را جزم می‌کند تا یک مدل آماری برای این مقادیر بیابد.

خانواده وایبول اغلب برای طول عمر ماده‌های در معرض خرابی و دیگر پدیده‌های وابسته قابل کاربرد بوده‌اند. یک نمودار احتمال وایبول برای داده‌های سی‌یه آماده شد. این نمودار در شکل ۶ داده شده‌است. پاره خطهای عمودی متناظر با خطاهای تاریخگذاری زمانهای بین

۱ـ سال نگارش مقاله (توضیح مترجم)

پیشامدها هستند. اگر توزیع وایبول برای توصیف توزیع زمانهای بین زلزله‌ها مناسب باشد، در این صورت نقاط رسم شده باید نزدیک خطی راست بیفتد. به‌عنوان مرجع، خطی راست در شکل درج شده‌است. فرض وایبول در اینجا موجه به‌نظر می‌رسد.

برآوردهای مخاطره

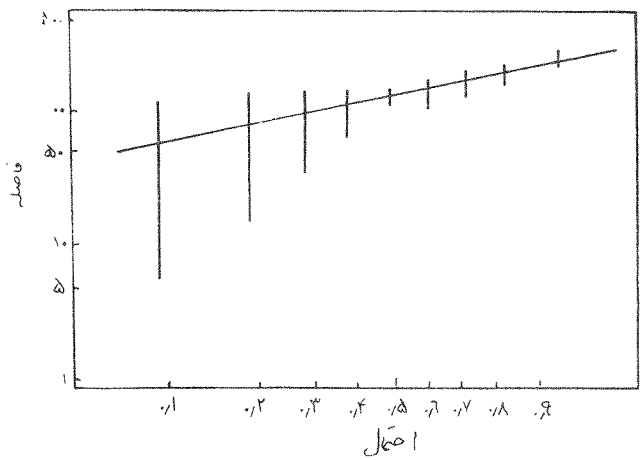
عده زیادی از مردم به سئوالهایی ازاین قبیل علاقه نشان می‌دهند: احتمال آنکه زلزله‌ای در ناحیه لوس‌آنجلس طی ۵ سال آینده رخ دهد، چقدر است؟ در ده سال بعد، و الی آخر چطور؟ این احتمالها (مخاطره‌ها) را می‌توان به محض اینکه برای زمانهای بین زلزله‌ها یک شکل توزیعی یافت شود، برآورد کرد. شکل ۷ برآوردهای اولیه‌ای از احتمالهای مخاطره را، با به‌کارگیری توزیع وایبول که به آن اشاره رفت و بااستفاده از داده‌های جدول ۱ دراختیار می‌گذارد. می‌توان از این شکل دریافت که به‌عنوان مثال احتمال زلزله‌ای بزرگ در طول ۳۰ سال آینده به فرض اینکه آخرین زلزله ۱۳۱ سال قبل رخ داده ۰/۳۲ است (یعنی ۳۲ درصد شانس برای وقوع یک زلزله موجوداست. خطهای نقطه‌چین در این شکل، نشانه‌ای ازعدم حتمیت در مقادیر احتمال برازش یافته را ارائه می‌دهد.) آنها با یک بازهٔ اطمینان ۹۵٪ متناظرند. با تعقیب خطهای افقی این شکل به یک حد پایین ۰/۱۸ ویک حد بالای ۰/۵۰ برای احتمال رخ دادن زلزله‌ای طی ۳۰ سال آینده رهنمون می‌شویم. از این نتیجه بیمه‌گران، مهندسان و برنامه‌ریزان می‌توانند در کارهایشان استفاده کنند.

حق بیمه‌ها

فرض کنید که شخصی می‌خواهد وجوهی را برای پرداخت هزینه بازسازی واحدی که ممکن است در زلزله‌ای (که در سال بعد روی می‌دهد) آسیب ببیند، کنار گذارد.

حق بیمه منصفانه برای پوشش بازسازی، درصورتی که زلزله اتفاق افتد، برای هر هزار دلار هزینه، با هزار برابر احتمال وقوع زلزله‌ای در سال آینده محاسبه می‌شود. بااستفاده از توزیع وایبول برازش یافته، برآورد احتمال زلزله‌ای درسال آینده ۰/۱۰۸ است و حق بیمه ۱۰/۸۰ دلار از کار در می‌آید. (البته، شرکتهای بیمه درحقیقت بر حق بیمه‌های خود مبالغی برای پوشش

شکل ۶ـ یک نمودار احتمال برای ارزیابی معقول بودن توزیع وایبول برای فاصله‌های بین زلزله‌ها در پالت‌کریک. نقطه‌های رسم‌شده متناظر با فاصله‌های مشاهده شده‌اند. خطهای عمودی به‌علاوه و منهای دوبرابر خطاهای استاندارد را مشخص می‌کنند. اگر توزیع معقول باشد، نقطه‌ها باید در نزدیکی یک خط راست قرارگیرند. به‌عنوان مرجع، یک خط راست اضافه شده‌است.

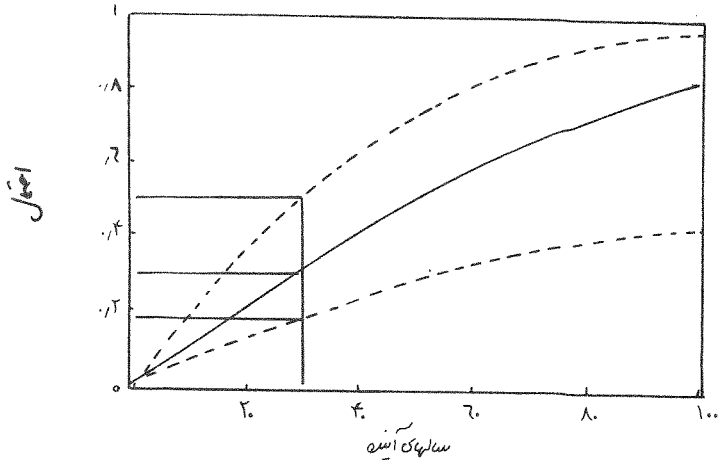


هزینه، کسب سود و حفظ خود در برابر فاجعه‌های غیرمنتظره "اضافه" می‌کنند، بنابراین بیش از ۱۰/۸۰ دلار دریافت می‌کنند.

سهم آمار در این مسئله

محصول نهایی مطلوب در یک مطالعهٔ خطر زلزله، یک احتمال است. بنابراین آمار مجبور به

شکل ۷ـ برآوردی از احتمال زلزله‌ای که در آینده در پالت کریک طی تعداد سالهای مشخص شده روی می‌دهد. خطهای نقطه‌چین مقدارهای بالایی و پایینی یک بازه اطمینان ۹۵ ٪ را برای این احتمال ارائه می‌دهند. خطهای افقی، این مقدارها را برای احتمال زلزله‌ای در ۳۰ سال آینده ارائه می‌دهند.



مداخله است؛ زیرا توزیعهای آماری اساس برآورد احتمالها هستند. در مطالعه‌ای که در بالا توصیف شد، ابزار تاریخگذاری رادیوکربن اهمیت پایه‌ای داشت. پژوهشگران این رشته مدتهاست اهمیت فنون مناسب آماری را تشخیص داده‌اند. به‌طوری که اچ. ای. پولاچ[۱] (۱۹۷۶) گفته است، کاربرد روشهای صحیح آماری به‌صورت "ابزار معاش" تاریخگذاران رادیوکربن درآمده‌است.

نقل‌قولی از *هارولد* جفریز (۱۹۷۶)، یکی از مهمترین زلزله‌شناسان و آمارشناسان این سده نیز خالی از ارزش نیست. وی گفته است؛ "یک برآورد بدون یک خطای استاندارد عملاً بی‌معنی است." این گفته به بیان نتیجه‌گیری اشاره دارد. ارائه خطاهای آماری ابزاری برای این کار، و بازه های اطمینان ابزار دیگری است. اینها هردو از مفاهیم اصلی آمارند. به خاطر داشته باشید که بازه اطمینان ۹۵ ٪ برای احتمال یک زلزله بزرگ در پالت کریک در ۳۰ سال آینده[۲] از ۰/۱۸ تا ۰/۵۰ تغییر می‌کند.

۱ـ *H. A. Polach*

۲ـ از ۱۹۸۸ به بعد (توضیح مترجم)

نتیجه گیری

علم با انباشتن خودش پیش می‌رود. در کاری که شرح داده شد، نمونه‌هایی با عمر معلوم (حلقه‌های درخت) برای ساختن یک خم واسنجی به‌کار گرفته شده‌است که خود آن در تاریخگذاری نمونه‌هایی با تاریخ ناشناخته به‌کار رفته است. علم از مفاهیم آماری برای روبروشدن با مسائل برآورد کردن ناشناخته‌ها، برای تأیید اعتبار فرضها، و برای کمّی کردن عدم حتمیتها در استنباطهای انجام شده، استفاده می‌کند.

مسائل

۱ـ از خم شکل ۲ استفاده کرده سالهای سپری شده برای پرتوافکنی را به‌طوری که به یک چهارم مقدار اولیه آن کاهش یابد، مشخص کنید.

۲ـ از خم شکل ۳ استفاده کرده سال تقویمی متناظر باسال رادیوکربن ۵۰۰ را مشخص کنید.

۳ـ از خم شکل ۳ استفاده کرده سال رادیوکربنی راکه متناظر با چندین سال تقویمی به‌جای یک سال تقویمی است، پیدا کنید. در مورد این پدیده اظهارنظر کنید.

۴ـ حول سال ۹۰۰ بعد از میلاد در خم شکل ۵ یک برآمدگی وجود دارد. درباره منشاء آن چه فکر می‌کنید؟ (راهنمایی: شکل ۳ را در نظر بگیرید.)

۵ـ حق بیمه تقریبی منصفانه پرداختی برای پوشش دادن خسارتی به ارزش ۵۰۰۰۰ دلار بر اثر زلزله‌ای که ممکن است در ۲۰ سال آینده (شروع از سال ۱۹۸۸ـ توضیح مترجم) رخ دهد، چقدر است؟ (راهنمایی: یک برآورد احتمال را از شکل ۷ مشخص کنید.)

۶ـ زمانهای بین دو رویداد متوالی برای زلزله‌های فهرست شده در جدول ۱ را محاسبه کنید. آیا به نظر می‌رسد که ساختاری در دنباله مقادیر وجود داشته باشد؟

منابع

D. R. Brillinger. 1982. "Seismic Risk Assessment: Some Statistical Aspect." *Earthquake Prediction Research 1:183-195.*

H. Jeffreys. 1967. "Seismology Statistical Methods in." *In International Dictionary of Geophysics,*

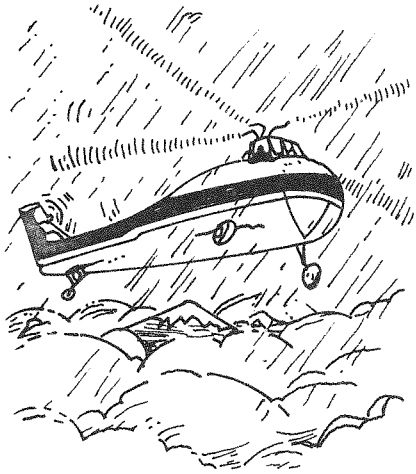
K,Runcorn, ed. London: Pergamon Press, PP. 1398-1401.

W. Nelson. 1982. *Applied Life Data Analysis.* New York: Wiley.

H. A. Polach. 1976. "Radiocarbon Dating as a Research Tool in Archeology: Hopes and Limitations." *In Proceedings of a Symposium on Scientific Methods of Research in the Study of Ancient Chinese Bronzes and Southeast Asian Metal and Other Archeological Artifacts,* N. Barnard, ed. Melbourne, Australia: National Gallery of Victoria, PP. 255-298.

K. E. Sieh.1984. "Lateral Offsets and Revised Dates of Large Prehistoric Earthquakes at Pallett Creek, Soutern California." *Journal of Geophysical Research 89:7641-7670.*

M. Stuiver and G. W. Pearson. 1986. "High-Precision Calibration of the Radiocarbon Time Scale, A.D. 1950-500 B.C." *Radiocarbon 28:805-838.*



پیش‌بینی بسیار کوتاه‌مدت وضع هوا با مشاهدات خودکار

رابرت جی. میلر[۱] خدمات هواشناسی ملی

خلبان جو لانگ از هلی‌کوپتر امدادرسانی خود برای انتقال مصدومی از جاده کمربندی به نزدیک‌ترین مرکز سوانح استفاده می‌کرد که پس از طی ۱۰ مایل از مسیر ۲۰ مایلی، هوای مقصد تغییرکرد، به‌طوری که برای فرود اصلاً دید نداشت. به‌خاطر پیداکردن محل دیگری برای فرود در آخرین لحظات، زمان گرانبهایی ازدست رفت؛ چرا که وضع هوا برای آینده بسیار نزدیک

۱ـ *Robert G. Miller*

اصلاً پیش‌بینی نشده‌بود.

این مقاله، به توصیف کوششی آماری برای پیش‌بینی بسیار کوتاه‌مدت وضع هوا می‌پردازد. به‌گونه‌ای که بتوان به خلبانان هلی‌کوپتر و دیگران در مواردی مانند مشکل بالاکمک کرد. این کار مقارن با راه‌اندازی سیستم خودکار مشاهده وضع هوا[1] (AWOS) که در آن هنگام دردست آزمایش بود، به‌اجرا در آمد. سیستم مشاهده خودکار وضع هوا به این ترتیب است که اطلاعات خوانده‌شده از روی ابزار هواشناسی (حس‌گرهایی که دما، باد و امثال آن را اندازه‌گیری می‌کنند) را دقیقه به دقیقه ثبت می‌کند. اطلاعات مربوط به آخرین ۳۰ دقیقه به‌صورت الکترونیکی به کامپیوتر برنامه‌ریزی شده‌ای ارسال می‌شود تا به تقلید از مشاهدات انسان وضع هوا را مشاهده کند.

سه روش متفاوت برای پیش‌بینیهای کوتاه‌مدت مثلا ۱۰ تا ۱۲۰ دقیقه‌ای وجود دارد:

۱ـ پیش‌بینی پایداری در این پیش‌بینی فرض بر این است که شرایط فعلی آب وهوا تا زمان کوتاهی در آینده همچنان ادامه خواهدداشت. این روش نتایج خوبی به‌بار می‌آورد و کوبیدن آن کار آسانی نیست. می‌دانیم که وضع هوا بندرت در فواصل زمانی کوتاه تغییر می‌کند. بنابراین پیش‌بینی پایداری وضع هوا بناچار نسبتاً دقیق است.

۲ـ قضاوت انسانی پیش‌بینی وضع هوا توسط انسان، روش دیگری است که از این جهت اهمیت دارد که اطلاعات مربوط به وضعیت هوای گذشته.و حال در اختیار پیش‌بینی‌کننده قرار دارد. این اطلاعات شامل مشاهدات به‌دست آمده از رادارها و ماهواره‌های هواشناسی، شرایط هوایی در سطوح مختلف جوی، پیش‌گوییهای کلان، و البته اطلاعات سیستم مشاهده خودکار محلی می‌باشند. این اطلاعات با درنظر گرفتن تجربه و آموزش پیش‌بینی‌کننده، بهترین منابع برای پیش‌بینیهای کوتاه‌مدت ۱۰ تا ۱۲۰ دقیقه‌ای به‌نظر می‌رسند. باوجود این، تجربه نشان داده است که برای پیش‌بینیهای کمتر از ۱۲۰ دقیقه‌ای، این نوع پیش‌بینیها بندرت بر پیش‌بینی پایداری ارجحیت دارند.

۳ـ پیش‌بینی آماری روش آماری که در این مقاله به‌کار گرفته شده‌است، یعنی روش پیش‌بینی

۱ـ *Automated Weather Observation System*

برمبنای داده‌های تجربی یکی دیگر از روشهاست. از مزایای این روش آن است که سیستم مشاهده خودکار محلی از جدیدترین مشاهدات موجود به‌عنوان داده استفاده می‌کند، به‌آسانی به‌صورت خودکار عمل می‌نماید و زمان پاسخ‌دهی آن بسیار سریع است. این روش برمبنای تقریباً یک میلیون مشاهده قبلی بنا شده‌است که روابط برآورد شده را ایجاد می‌کنند. چون تهیه پیش‌بینیهای کوتاه‌مدت برمبنای قضاوت انسان که نسبت به پیش‌بینیهای پایداری بهتر باشند مشکل است، آیا پیش‌بینی آماری می‌تواند کمک کند؟

تشکیل معادلات پیش‌بینی

به‌منظور تشریح دقیق مسئله، یک روش آماری را برای پیش‌بینی ۱۰ دقیقه‌ای میدان دید که به‌طور مستقیم با روش پایداری مقایسه می‌شود، شرح می‌دهیم. این روش تحلیل رگرسیونی، یا تحلیل تشخیصی دوگروهی (ن.ک. *تاتسوگا*، ۱۹۷۱) نامیده می‌شود. (مقاله *هاول* را درباره تابع تشخیص نیز مطالعه کنید). تحلیل رگرسیونی روش متداولی برای برآورد مقدار یک متغیر مجهول است که به‌صورت مقادیر موزون متغیرهای معلوم ارائه می‌شود. (بخصوص وزنها طوری تعیین می‌شوند که مجموع توانهای دوم تفاوتهای موجود میان برآوردها و مقادیر واقعی متغیرها به حداقل برسد.) این وزنها را با امتحان‌کردن وضعیتهایی انتخاب می‌کنیم که مقادیر متغیرهای معمولاً مجهول را در آن وضعیتها می‌دانیم.

میدان دید توسط یک حس‌گر برحسب مایل و کسری از آن اندازه‌گیری می‌شود. میدان دید توسط سیستم مشاهده خودکار وضع هوا از روی اطلاعات حس‌گر برای ۱۰ دقیقه آخر به‌دست می‌آید. می‌توانستیم سعی کنیم که این مقدار به‌دست آمده را نیز پیش‌بینی کنیم، ولی درعوض معلوم شد راحت‌تر آن است که ببینیم مشاهده و پیش‌بینی در کدامیک از شش رسته دید ۱۰- دقیقه‌ای واقع می‌شود؟

از صفر تا $\frac{1}{2}$ مایل ولی شامل $\frac{1}{2}$ مایل نیست
از $\frac{1}{2}$ تا ۱ مایل ولی شامل ۱ مایل نیست
از ۱ تا ۳ مایل ولی شامل ۳ مایل نیست
از ۳ تا ۵ مایل ولی شامل ۵ مایل نیست

از ۵ تا ۷ مایل ولی شامل ۷ مایل نیست

۷ مایل وبیشتر

در یک وضعیت فرضی پیش‌بینی وضع هوا، مشاهدات سیستم خودکار، مشاهده وضع هوا با یک سری صفر و یک (۰ و ۱) نشان داده‌می‌شود. برای مثال عنصر هوایی سرعت باد دارای پنج رسته سرعت است که فقط یکی از آنها در مشاهده روی می‌دهد. به این رسته عدد ۱ و به بقیه رسته‌ها صفر داده‌می‌شود. به‌طور مشابه جهت باد دارای ۸ رسته است که فقط یکی از آنها در مشاهده قرارمی‌گیرد. روی هم‌رفته ۲۶ عنصر مربوط به وضع هوا وجود دارد که کلاً با ۱۶۶ رسته نمایش داده‌می‌شوند. هر مشاهده فرضی از ۱۶۶ رسته دقیقاً ۲۶ عدد یک و ۱۴۰ عدد صفر خواهد داشت. گروه‌بندی متغیرها به این شکل دارای دو هدف مهم است. نخست اینکه گرچه بعضی عناصر هواشناسی دارای نظم هستند، بر مقیاس عددی مبتنی نیستند (میدان دید مثالی از این دست است که در آن هیچ مقدار عددی برای رسته "نامحدود" و "۷ مایل وبیشتر" متصور نیست. دوم اینکه پردازش ریاضی داده‌های رسته‌ای با عملیات منطقی کامپیوتری با کارایی بسیار زیاد انجام می‌شود وجای عملیات کندتر حسابی را می‌گیرد. (برای مثال، یک میکروکامپیوتر (ریزکامپیوتر) برای محاسبه حاصلضرب دو عدد حقیقی ۲۰ برابر بیشتر از یک عملیات منطقی بین دو عدد صحیح که هریک با تعداد زیادی صفر و یک نشان داده می‌شود، زمان می‌برد. درمجموع مقایسه درباره دو رتبه بزرگی در زمان محاسبه و یک رتبه بزرگی در فضای ذخیره‌سازی به نفع طرح پیشنهادی ۱ ـ ۰ است.)

تحلیل آماری یک شیوه رگرسیونی است که ۳۰ پیشگو را از میان ۱۶۶ متغیر انتخاب می‌کند (ن. ک. دریپر و *اسمیت*، ۱۹۸۱ و *میلر* ۱۹۶۲). چند نمونه از متغیرهای بالقوه که از رسته‌های آنان به‌عنوان پیشگو استفاده می‌شود، عبارتند از: کمترین ارتفاع مشاهده‌شده توسط حسگر ابرها در یک دقیقه، فشار، جهت باد، و مقدار نزولات آسمانی. این شیوه آماری براساس موفقیت در برآورد نتیجه ۸۱۸۹۵۳ مشاهده قبلی، ۳۰ پیشگو یا رسته را انتخاب می‌کند و وزنهایی را برای آنها محاسبه می‌نماید که نتیجه آنها را به بهترین وجه مورد استفاده قراردهد. این روش کلی برآورد یا نمره جداگانه‌ای برای هریک از رسته‌های ۱۰ دقیقه‌ای میدان دید که در بالا ذکر شده، فراهم می‌کند. در حالت مطلوب، نتیجه عبارت است از یک عدد ۱رسته‌ای که

اتفاق خواهد افتاد و یک عدد صفر برای بقیه رسته‌ها، اما در عمل این اعداد از اندکی زیر صفر تا اندکی بالای یک تغییر می‌کنند. هر رسته دارای یک معادله مخصوص به خود است. اولین برآورد رگرسیونی برای میدان دید ۱۰ دقیقه‌ای عددی به ما ارائه می‌دهد که با اولین رسته آن که در بالا فهرست شده‌است، پیوند دارد: صفر تا $\frac{1}{2}$ مایل

(حس‌گر میدان دید ۷ مایل یا بیشتر)×(۰/۰۰۱-)+۰/۰۴۷= برآورد رگرسیونی

(حس‌گر میدان دید در ۱۰ دقیقه گذشته از ۱ مایل تا کمتر از ۳مایل)×(۰/۰۰۳)+

+

.

.

.

(مقدار بارش ۰/۰۰۲ تا ۰/۱۰۰ در ۱دقیقه)×(۰/۰۰۹-) +

عبارتهای سمت راست علامت مساوی براساس اینکه آیا شرط پیشگویی مربوط به آن که در داخل پرانتز آمده‌است، اتفاق می‌افتد یاخیر جمع جبری می‌شوند. برای مثال، با ۰/۰۴۷ شروع می‌کنیم، سپس اگر حس‌گر، میدان دید را ۷ مایل یا بیشتر گزارش داد، مقدار ۰/۰۰۱- را به آن اضافه می‌کنیم ولی اگر میدان دید ۷ مایل یا بیشتر نبود، عبارت ۰/۰۰۱- حذف می‌شود. ۲۸ ضریب باقیمانده نیز به همین ترتیب و براساس شرایط مشاهده‌شده پیشگو یا اضافه می‌شوند یا نمی‌شوند. کمیت به‌دست آمده به‌عنوان برآورد رگرسیونی عبارت است از شاخصی که پیش‌بینی نهایی میدان دید از روی آن انجام می‌پذیرد و متعاقباً شرح داده خواهدشد.

به‌یاد داشته‌باشید که تحلیل رگرسیونی مقدار وزنها را در هر معادله طوری محاسبه می‌کند که جمع مربعات خطا بین مقدار پیش‌بینی و مقدار واقعی مینیمم شود.

جدول ۱ عملکرد هر دو روش را جمع‌بندی می‌کند. این جدول مجموع مربعات انحرافها را میان مقادیر پیش‌بینی شده و مشاهده‌شده برای هر رسته و با هریک از روشها با مجموع مربعات با استفاده از میانگین عملکرد هر رسته مقایسه می‌کند. بااین ترتیب مجموع مربعات انحرافها با روش پیش‌بینی پایداری برای رسته صفر تا $\frac{1}{2}$ مایل، ۲/۵۰٪ بود که درست به همان اندازه مجموع مربعات انحرافها نسبت به میانگین است. مکمل این مقدار ۴۹/۸=۵۰/۲ - ۱۰۰

جدول ۱ـ مقایسه دو روش پایدار و رگرسیونی از لحاظ "درصد پیش‌بینی پذیری"

طبقه پیش بین	رگرسیونی (٪)	پایداری(٪)	برتری رگرسیونی نسبت به پایداری (٪)
صفر تا کمتراز $\frac{1}{2}$	۵۴/۸	۴۹/۸	۵/۰
$\frac{1}{2}$ تا کمتراز ۱	۴۰/۶	۳۱/۵	۹/۱
۱ تا کمتراز ۳	۶۴/۳	۵۶/۵	۷/۸
۳ تا کمتراز ۵	۶۲/۴	۵۴/۹	۷/۵
۵ تا کمتراز ۷	۵۳/۳	۴۴/۵	۸/۸
۷ وبیشتر	۸۷/۶	۸۳/۹	۳/۷

پیش‌بینی پذیری نامیده می‌شود.

درصد پایداری کل پیش‌بینی پذیری به‌شرح زیر به‌دست آمد: پیش‌بینیهای احتمال پایداری با به‌کارگیری رهیافت رگرسیونی مشابه روش آماری که در بالا ذکر شد، به‌دست آمد؛ بجز اینکه فقط پیشگوهای پایداری (یعنی شش رسته میدان دید در زمان صفر که قبلاً فهرست شد) برای منظورشدن در معادلات خود (معادلات پایداری) انتخاب شدند. این حالت در ستون ۱ نشان داده شده‌است. بهتربودن رگرسیون نسبت به‌پایداری در آخرین ستون جدول ۱ ارائه شده است.

میزان بهتر بودن رگرسیون نسبت به پایداری حاکی ازآن است که پیشگوها در معادله رگرسیونی سهم بسزایی داشته‌اند. چه در درصد مطلق پیش‌بینی پذیری کل و چه درمقایسه با آنچه که شرایط پایداری در معادله ادا می‌کنند. این کمیتها نشان می‌دهد که هرگاه مقداری شاخص معادلات رگرسیونی در آخرین مرحله از فرایند پیش‌بینی به‌کار گرفته شوند، می‌توانیم انتظار پیش‌بینیهای بهتری را داشته‌باشیم. حال به فرایند مورد استفاده واقعی و کارکرد آن می‌پردازیم.

نتایج تطبیقی

هدف از تمام این تلاشها این است که مقدار شاخصهای حاصل از تحلیل رگرسیونی مورد استفاده قرار گیرد و پیش‌بینی شود که از حالا تا ۱۰ دقیقه دیگر، میدان دید در کدام رسته

مشاهده خواهدشد. روشهای متعددی می‌توان اتخاذ نمود. ولی معلوم شده‌است که رسته دارای بالاترین مقدار شاخص بیشترین تعداد پیش‌بینیهای صحیح را ارائه می‌دهد. متأسفانه این معیار به رسته‌هایی که کمتر روی می‌دهند، گرایش دارد و شانس کمتری برای پیش‌بینی می‌دهد و برای رسته‌هایی که بیشتر اتفاق می‌افتند، مساعدتر است. این پدیده همچنان که زمان پیش‌بینی طولانیتر می‌شود باوضوح بیشتری نمایان می‌گردد.

آنچه برای هواشناسان فعال و کاربران عملیاتی هردو مطلوب است، تعادل در فراوانی هر رسته می‌باشد ـ تعادل از این نظر که فراوانیهای پیش‌بینی با فراوانی مشاهده‌شده برای هر رسته از پیشامد مطابقت نزدیک داشته‌باشد. برای مثال میدان دید صفر تا $\frac{1}{2}$ مایل فقط در حدود ۰/۴٪ اوقات پیش می‌آید، درصورتی که میدان دید ۷ مایل و بیشتر در حدود ۷۵/۵٪ مواقع روی می‌دهد. طرح ما باید کم وبیش با همین درصدها پیش‌بینی کند وگرنه موفق نخواهیم شد. استانداردهای تعادل راکه در حرفه هواشناسی به تکامل رسیده‌اند، رعایت کنیم.

کلین و همکاران (۱۹۵۹) روشی را ابداع نموده‌اند که در شرایط تغییرپذیری که هواشناسان و کاربران ترجیح می‌دهند، صدق کند. در اینجا به شرح این روش نمی‌پردازیم ولی درعوض به اثبات خوبی آن توجه می‌کنیم.

می‌توانیم ثبات شیوه آماری خود را با به‌کارگیری یک نمونه آزمایشی از داده‌ها ارزیابی کنیم. حساب امتیازات برای نمونه‌ای ۳۶۹۸۰۲ تایی از پیش‌بینی میدان دید ۱۰ دقیقه‌ای مربوط به روش آماری در جدول ۲ ارائه شده‌است. جدول ۳ مجموعه نتایج متناظر راکه با استفاده از روش پایداری به‌دست آمده‌است، نشان می‌دهد.

نتایج نمونه آزمایشی که در این تلاش به‌دست آمد، بسیار رضایت‌بخش بود. بدین‌ترتیب از روی جدول ۲ به‌دست می‌آید که از ۱۰۳۲ شروع می‌کنیم و در امتداد قطر به طرف پایین ادامه می‌دهیم:

$$۱۰۳۲+۱۵۴۹+۱۴۵۶۲+\ldots+۲۷۴۳۶۸=۳۴۵۲۴۰$$

درحالی که طبق جدول ۳، تعداد پیش‌بینیهای صحیح برای روش پایدار برابر است با

$$۱۰۶۱+۱۳۸۲+۱۳۹۲۶+\ldots+۲۷۳۸۴۹=۳۴۲۷۳۶$$

بنابراین با روش آماری تعداد پیش‌بینیهای صحیح ۲۵۰۴ مورد بیشتر بوده‌است. باتوجه به آنکه

جدول ۲ـ پیش‌بینی آماری در برابر مشاهدات واقعی میدان دید ۱۰دقیقه‌ای برای نمونه‌ای آزمایشی شامل ۳۶۹۸۰۲ مشاهده

رسته پیش‌بینی (مایل)	۰ تا کمتراز $\frac{1}{2}$	$\frac{1}{2}$ تا کمتراز ۱	۱ تا کمتراز ۳	۳ تا کمتراز ۵	۵ تا کمتراز ۷	۷ و بیشتر	جمع
صفر تاکمتراز $\frac{1}{2}$	۱۰۳۲	۱۶۲	۵۲	۱۳	۶	۵	۱۲۷۰
$\frac{1}{2}$ تاکمتراز ۱	۲۶۱	۱۵۴۹	۶۵۷	۹۰	۳۱	۱۶	۲۶۰۴
۱ تاکمتراز ۳	۵۲	۴۷۶	۱۴۵۶۲	۲۸۸۴	۱۹۶	۱۰۳	۱۸۲۷۳
۳ تاکمتراز ۵	۵	۳۴	۲۳۴۱	۲۷۶۶۰	۴۳۱۲	۳۸۲	۳۴۷۳۴
۵ تا کمتراز ۷	۱	۳	۱۳۳	۳۴۰۵	۲۶۰۶۹	۴۲۶۴	۳۳۸۷۵
۷ وبیشتر	۰	۰	۷۶	۳۵۱	۴۲۵۱	۲۷۴۳۶۸	۲۷۹۰۴۶
جمع	۱۳۵۱	۲۲۲۴	۱۷۸۲۱	۳۴۴۰۳	۳۴۸۶۵	۲۷۹۱۳۸	۳۶۹۸۰۲

با روش پایدار تعداد ۲۷۰۶۶=۳۶۹۸۰۲-۳۴۲۷۳۶ پیش‌بینی غلط وجود داشته است، رگرسیون توانسته‌است بیش از ۹٪ از پیش‌بینیهای اشتباه را اصلاح کند. واقعیت دلگرم‌کننده دیگر آن است که طرح آماری (۱۲۷۰ ـ ۱۴۵۰) ۱۸۰ مورد از این وضعیتها را که طی آن میدان دید صفر تا $\frac{1}{2}$ مایل در زمان پیش‌بینی بود، تغییر داد و دراین رسته فقط ۲۹ مورد کمتر از روش پایداری موفقیت داشت. درصد پیش‌بینیهای صحیح برای این رده ۱۰۳۲:۱۲۷۰ یا ۸۱/۳٪ است که از درصد مربوط به روش پایدار، یعنی ۱۰۶۱:۱۴۵۰ یا ۷۳/۲٪ بهتر است.

واضح است که پیش‌بینی میدان دید وسیع و روبه‌رو شدن با میدان دید محدود به هیچ وجه مطلوب نیست. جدول ۳ دارای ۱۳۲۹۷ موقعیت پایداری است که زیر قطر قراردارند (پیش‌بینیهای نامطلوب) درحالی که (جدول ۲) رگرسیون ۱۱۳۸۹ یا ۱۹۰۸ مورد کمتر پیش‌بینی نامطلوب داشته‌است. منطقه حساس ۳ یا بیشتر در پایین قطر این جدول چشمگیرتر است؛ یعنی ۱۵۳ به روش پایداری و ۵۸ به روش رگرسیونی که حاکی از آن است که پیش‌بینی خیلی‌بد در روش رگرسیونی ۶۸ مورد کمتر است. نهایتاً اینکه آمار توانسته است نسبت به پش‌بینیهای پایداری در یکی از عناصر مشکل هواشناسی، یعنی میدان دید، با موفقیت عمل کند و این کار را در محدوده زمانی بسیار کوتاه ۱۰دقیقه‌ای انجام داده‌است. ارزش پیش‌بینی صحیح شرایط

جدول ۳ـ پیش‌بینی به روش پایدار دربرابر مشاهدات واقعی میدان دید ۱۰دقیقه‌ای برای نمونه‌ای آزمایشی شامل ۳۶۹۸۰۲ مشاهده

رسته پیش‌بینی (مایل)	۰ تا $\frac{1}{2}$ کمتراز	$\frac{1}{2}$ تا ۱ کمتراز	۱ تا ۳ کمتراز	۳ تا ۵ کمتراز	۵ تا ۷ کمتراز	۷ و بیشتر	جمع
صفر تاکمتراز $\frac{1}{2}$	۱۰۶۱	۲۴۹	۱۱۰	۱۷	۷	۶	۱۴۵۰
$\frac{1}{2}$ تاکمتراز ۱	۱۹۷	۱۳۸۲	۵۹۰	۴۰	۱۱	۵	۲۲۲۵
۱تاکمتراز ۳	۸۰	۵۳۰	۱۳۹۲۶	۲۹۹۸	۲۰۷	۱۱۲	۱۷۸۵۳
۳ تاکمتراز ۵	۸	۴۷	۲۸۲۸	۲۶۶۷۳	۴۲۵۱	۵۶۴	۳۴۳۷۱
۵ تاکمتراز ۷	۲	۱۱	۲۴۳	۴۲۰۵	۲۵۸۴۵	۴۶۰۲	۳۴۹۰۸
۷ وبیشتر	۳	۵	۱۲۴	۴۷۰	۴۵۴۴	۲۷۳۸۴۹	۲۷۸۹۹۵
جمع	۱۳۵۱	۲۲۲۴	۱۷۸۱۱	۳۴۴۰۳	۳۴۸۶۵	۲۷۹۱۳۸	۳۶۹۸۰۲

دستخوش تغییر، که روش پایداری به‌خاطر ماهیت ذاتی خود چنین پیش‌بینیهایی را ارائه می‌نماید، دستاوردی است که در اهمیت آن هرچه گفته شود، کم است.

مسائل

۱ـ فرض کنید که اگر میدان دید کمتر از ۱۰مایل بوده‌باشد، جو لانگ نمی‌توانسته هلی‌کوپتر را فرود بیاورد. اگر وی از روش آماری برای پیش‌بینی میدان دید طی زمان نمونه مورد اشاره در جدولهای ۲ و ۳ پیروی می‌کرد، چند بار مجبور می‌شد پروازهایش را لغو کند؟ اگر از روش پایداری استفاده می‌کرد، چند بار مجبور می‌شد پروازهایش را لغو کند؟ درصورتی که پیش‌بینی درست نباشد، ناچار باید به فرودگاه دیگری تغییر مسیر دهد. دو روش را باتوجه به عملیات مورد نیاز مقایسه کنید و موردی را برای انتخاب یک روش نسبت به دیگری بیان کنید.

۲ـ به نظر شما، چرا کاربران پیش‌بینی وضع هوا که به‌صورت رسته‌ای ارائه می‌شود، ترجیح می‌دهند این قبیل پیش‌بینیها با همان فراوانی که مشاهده می‌شوند، ارائه گردند؟ آیا این الزام منطقی است؟ دلایل خود را له و علیه آن ذکر کنید.

۳ـ آیا از اینکه پی ببرید پیش‌بینیهای بسیار کوتاه‌مدت بهتر است با استفاده از برنامه کامپیوتری به‌دست آید و براساس قضاوت انسان نباشد، ناراحت می‌شوید؟ نظر خود را بیان کنید.

۴ـ چرا تصور می‌کنید پیش‌بینیهای وضع هوا برای تأمین نیازهای ما به اندازه کافی دقیق نیستند؟ آیا به این خاطر است که توقعات ما بیش‌از حد زیاد است؟ درک ما از اتمسفر بیش‌از حد ناکافی است؟ داده‌های ما بیش‌از حد ناقص است؟ یا روشهایی که برای تحلیل به‌کار می‌بریم بیش‌از حد محدود است؟

منابع

N. R. Draper and H. Smith. 1981, *Applied Regression Analysis.* New York: Wiley.

R. G. Miller. 1962, *Statistical Prediction by Discriminant Analysis.* Meteorology Monograph No. 25. Boston: American Meteorological Society.

M. M. Tatsuoka. 1971, *Multivariate Analysis. Techniques for Educational and Psychological Research.* New York: Wiley.



آمار، خورشید و ستارگان

چارلز ای. ویتنی[1] دانشگاه هاروارد

به‌نظر می‌رسد که ستارگان مرئی به‌گونه‌ای تصادفی در فضا پراکنده‌اند. بنابراین طبیعی است که منجمان مایل به استفاده از روشهای آماری و استدلالهای احتمالاتی باشند. اما بهره‌گیری از این روشهاو استدلالها پیشرفت علمی نسبتاً نوینی است. نجوم بابلیها و یونانیها مبتنی بر افلاک بود، و مفاهیم احتمال، قبل از قرن بیستم به‌ندرت در نجوم جدید به‌کار می‌رفت.

۱ـ *Charles A. Whitney*

جزء اولین موارد استفاده از ایده تصادفی بودن و مفاهیم احتمال در نجوم، اثبات ریاضی *جان میچل*[۱] در قرن هیجدهم است. اثبات وی در این‌باره است که ستارگانی که در آسمان خیلی نزدیک یکدیگر به‌نظر می‌آیند، واقعاً در فضا نیز به هم نزدیکند. *ویلیام هرشل*[۲] که معاصر با میچل بود، تعداد زیادی از زوج ستارگان خیلی نزدیک به هم را کشف کرد که **ستارگان مزدوج** نامیده شدند ـ اما وی دلیلی برای اینکه نشان دهد این ستارگان واقعاً به‌صورت زوجند و یا تنها چنین به نظر می‌آیند در دست نداشت. هرشل تصور می‌کرد هر دو ستاره‌ای که به زوجی تعلق دارند کاملاً از هم جدا بوده و به‌طور تصادفی در طول یک خط دید قرار گرفته‌اند. او با مشاهده چنین زوجهایی امیدوار بود که حرکت زمین حول خورشید را آشکار، و نهایتاً فاصله ستارگان از زمین را تعیین کند. اما میچل با فرض اینکه ستارگان به تصادف در فضا پراکنده باشند، احتمال رخداد چنین جفت شدن ظاهری در ارقامی را که هرشل یافته بود محاسبه کرد. این احتمال آن‌قدر کوچک بود که میچل باور کرد این ستارگان به‌صورت فیزیکی با هم زوج شده‌اند. درستی این باور، کمی بعد به‌وسیله خود هرشل نیز ثابت شد. هرشل بسیاری از ستارگان مزدوج رایافت که اعضای آنها حول یک مرکز گرانش مشترک دوران می‌کردند و لذا این کار او بود که به اولین تشریح کاربرد مفهوم جاذبه عمومی نیوتن در خارج از منظومه شمسی منجر شد.

هرشل همچنین تلاش کرد که نقشه کهکشان راه شیری را با این مطالعه ترسیم کند که چگونه تعداد ستارگانی که در راستایی خاص دیده می‌شوند، وقتی او به کرانهای کمرنگ و کم‌رنگتر می‌نگرد، افزایش می‌یابند. وی به‌دلیل فقدان اطلاع بیشتر، فرض کرد همه ستارگان واقع در یک مسافت استاندارد باید یکسان به‌نظر آیند و دریافت که به نظر می‌رسد خورشید در مرکز راه شیری قرار دارد.اخترشناسان اکنون تشخیص داده‌اند که این تنها یک تصور است ـ فضای بین ستارگان با غبار و گازهایی پوشیده شده‌است که نور ستارگان دوردست را تار می‌کند و درعین حال رنگ آنها را تغییر می‌دهد. داده‌ها و روشهای آماری تازه، به ارائه مدلی منجر شده‌اند که در آن، خورشید از مرکز تغییر مکان یافته و در یکی از بازوهای مارپیچی کهکشان ماجای

۱- *John Michell*

۲- *William Herschel*

گرفته‌است.

در مقیاسی بزرگتر، مطالعه‌های جدید نشان داده‌اند که کهکشانها در سراسر فضا با الگویی کاملاً نامنظم به‌صورت خوشه‌هایی توزیع شده‌اند که بین خوشه‌ها مناطق تهی وسیعی شبیه به حباب وجوددارند. درک این الگو، کار عمده کیهان‌شناسی جدید است.

بابیانی کلی، بحثهای آماریی که اکنون به‌وسیله اخترشناسان، در تلاششان برای آشکار ساختن ارتباطهای اتفاقی تنیده‌شده در فضا، مورد استفاده قرار می‌گیرند از دو رده هستند: اول، تحلیلهای آماری داده‌ها، و دوم، نظریه‌های فیزیکی مبتنی بر مفاهیم آماری یا احتمالاتی. مثالی از هریک، در زیر ارائه می‌شود:

تحلیلهای آشفتگیها در جو خورشیدی

جو خورشیدی با فعالیت در حال جوش و خروش است، و برخی از آشفتگیها، خصوصاً آنها که جوش و خروش کمتر و آرامتری دارند، پیرو تحلیل آماری هستند. هوای خورشیدی (اگر بتوان چنین گفت) به‌طور کامل بی‌نظم نیست، واخترشناسان مایلند تا آنجا که می‌توانند درباره نظمهای الگوی این هوای خورشیدی جمع‌آوری کنند، زیرا چنین نظمهایی قطعاً در ساختن توجیهات نظری یاری می‌کنند.

گازهای جو خورشیدی شدیداً داغند، آنچنان داغ که درواقع نور مرئی از آنها می‌تابد. ولی گرمای آنها دور از یکنواختی است؛ با یک تلسکوپ در دقایقی کوتاه می‌توان جزئیات تغییرات زودگذری را دید که از مکانی به مکان دیگر اوج می‌گیرد و تدریجاً ناپدید می‌شود. آشفتگیها در

جدول ۱ـ همبستگی نظریات

	نقاط				
زمان	۱	۲	۳	۴	۵
۱	۱/۵	۰/۵	۰	-۰/۵	-۱/۵
۲	۱/۴	۰/۷	۰	-۰/۷	-۱/۴
۳	۱/۰	-۱/۴	۰/۴	-۱/۰	۱/۰

یک الگوی نامنظم پخش می‌شوند، نواحی سرد و گرم صدها مایل مربع را می‌پوشانند و خطوط‌مرزی آنها تقریباً به شکل شش‌گوش است. این الگو را **الگوی دانه‌ای** نامیده‌اند و عناصری که به شکل شش‌گوش هستند و به **دانه‌ها** موسومند به‌وضوح حبابهایی از گازهای داغند که از درون به بالا می‌جوشند و با انتقال حرارت از مرکز، لایه نازک خارجی را آشفته می‌کنند. حتی دید قویترین تلسکوپها نمی‌تواند به درون جو خورشیدی نفوذ کند، به قسمی که منجمان برای ارزیابی و تشخیص وضع درونی خورشید به تحلیل ریاضی تکیه می‌کنند. اما این تحلیل به یک بازبینی مشاهداتی نیاز دارد. ماهیت تفضیلی نوسانهای جوی امکان چنین بازبینی را فراهم می‌کند، و به‌اخترشناسان نیز این اجازه را می‌دهد که درون خورشید را خوب وارسی کنند، شبیه روشی که یک زلزله‌شناس می‌تواند درون زمین را بررسی نماید.

درخشندگی

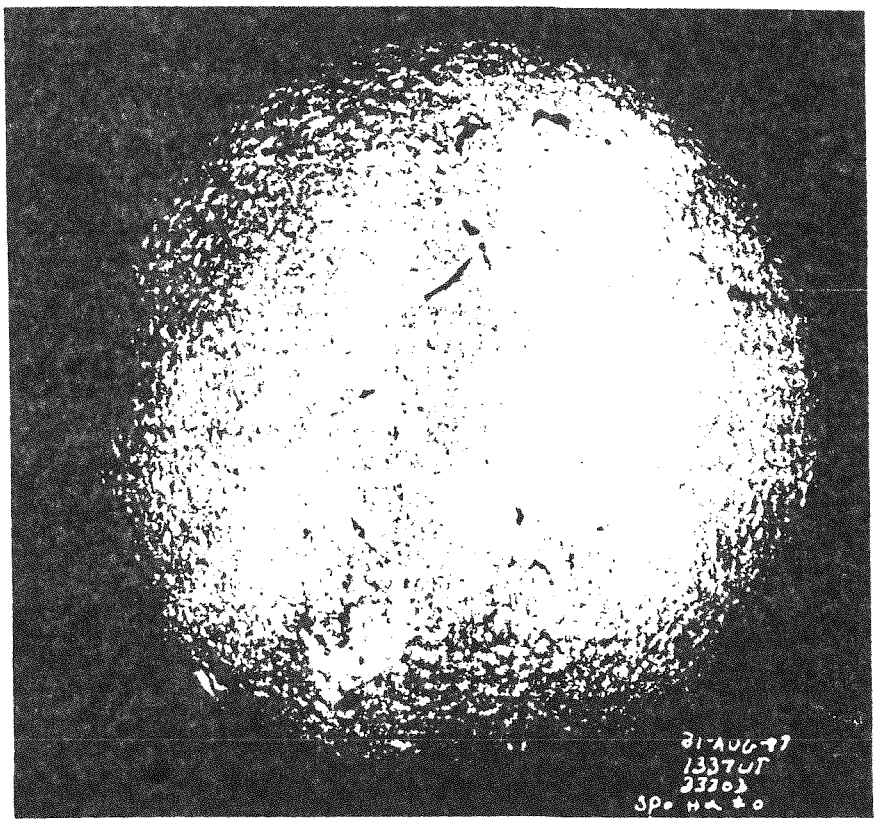
برای استنباط درباره ساختار مورد نظر، دو نوع اندازه‌گیری می‌تواند صورت گیرد: درخشندگی و سرعت گاز. مطالعات اولیه نشان می‌داد که الگوی درخشندگی هر ۵ دقیقه یا در همین حدود شدیداً تغییر می‌کرد. داده‌های کمی برپایه سریهایی از عکسها، در طول هر ۱۰ تا ۲۰ ثانیه تهیه شده‌اند. عکسها نوعاً در طول $\frac{۱}{۱۰۰}$ ثانیه به‌دست آمده‌اند و بهترین سریها آنهایی هستند که بلافاصله بعد از طلوع خورشید، وقتی زمین سرد و هوا آرام است تهیه می‌شوند. تهیه این سریها اغلب شامل چندین ساعت می‌شود، و اخترشناسان در نواحی قطبی، در طول فصلهای خورشید نیمه شب[۱]، توانسته‌اند دنباله‌هایی از عکسها را که چندین دوازده ساعت به طول انجامیده‌اند به‌دست آورند.

مقایسه تک‌تک عکسهایی (شکل ۱ مثالی از این عکسهاست) که فاصله زمانی آنها بیشتر و

۱ـ *در مناطق بالاتر از ۶۳/۵ درجه عرض جغرافیایی و درحد فاصل اواخر اردیبهشت تا اوایل مرداد در نیمکره شمالی و از اواخر آبان تا اوایل بهمن در نیمکره جنوبی، خورشید در طول شبانه‌روز همواره بالای افق بوده و غروب نمی‌کند وحتی در نیمه‌شب نیز به نورافشانی ادامه می‌دهد. به همین جهت نیز به‌آن خورشید نیمه شب می‌گویند. (ویراستار)*

شکل ۱ـ تصویر، سطح خورشید جزئیات ظریفی را که تحت تحلیلهای مختلف آماری قرار گرفته‌اند نشان می‌دهد. زمینه‌های روشن و تاریک، نواحی با شدت مغناطیسی مختلف را آشکار می‌سازد.

مأخذ: رصدخانه ملی خورشیدی



بیشتر می‌شود، تغییراتی را در درخشندگی همراه با مرور زمان آشکار می‌کند. همبستگی متقاطع الگوها درباره فیلمهای مختلف به طریق زیر تعیین می‌شود: خطی در سطح خورشید مشخص، و شدت نور در هر نقطه از خط در هرعکس را تعیین می‌کنند: (کسانی که علاقه‌مند به موضوع اندازه مشابهت نیستند، یا آنان که با ضریب همبستگی آشنایی دارند، می‌توانند به مطالعه مطلب بعدی تحت عنوان «یک مدل ممکن» بپردازند)

جدول ۱ معرف پنج نقطه در طول یک خط از سطح خورشید است. اولین سطر اعداد، درخشندگی‌های اولیه را به‌دست می‌دهد. این اعداد به طریقی استانداردشده اندازه‌گیری گردیده‌اند، به قسمی که درخشندگی در هر نقطه به‌عنوان میزان انحراف از میانگین برحسب واحد تغییرپذیری اولیه اندازه‌گیری شده‌است. عمل استانداردکردن، این ویژگیها را به اعداد می‌دهد که متوسط آنها صفر و متوسط توانهای دوم آنها برابر با ۱ می‌شود. (در جدول زیر این متوسطها به دلیل گردکردن اعداد دقیقاً چنین ویژگیهایی ندارند.) سطر دوم، درخشندگی‌های استانداردشده را در اندک زمانی بعد، اما در همان نقاط به‌دست می‌دهد. سطر سوم، درخشندگی‌های استانداردشده را در چند دقیقه بعد معرفی می‌کند.

توجه کنید که سطر دوم تقریباً نظیر سطر اول است، اما سطر سوم به میزانی زیاد با سطر اول تفاوت دارد. یک راه اندازه‌گیری میزان مشابهت بین دو مجوعه از پنج عدد این است که دو عدد استانداردشده هر زوج را در هم ضرب کنیم و پنج حاصلضرب را باهم جمع کنیم و با تقسیم این حاصل بر پنج، متوسط آنها را به‌دست آوریم. ایده اساسی آن است که اگر در این متوسط، مشابهت زیادی وجود داشته‌باشد، نزدیک ۱، و اگر ارتباط کمی موجود باشد نزدیک صفر، و اگر عدم شباهت زیاد باشد، نزدیک ۱ـ خواهد بود. متوسط حاصلضرب‌ها، **ضریب همبستگی** نام دارد.

اکنون ابتدا نظری به حالت حدی مشابهت می‌افکنیم؛ وقتی اولین مجموعه ۵ عدد را با خودش مقایسه می‌کنیم، باید مشابهتی کامل به‌دست آوریم. ضریب همبستگی همانگونه که گفتیم برابر است با

$$\frac{1}{5}[(1/5\times1/5)+(0/5\times0/5)+(5\times5)+(-0/5)(-0/5)+(-1/5)(-1/5)]=10$$

مشابهت یک مجموعه از چنین اعداد استانداردشده‌ای با خودش، وقتی این مشابهت به شیوه ذکر شده اندازه‌گیری شود، همیشه نظیر مثال بالا مقدار ۱ را به‌دست می‌دهد. ضریب همبستگی می‌تواند از ۱ـ تا ۱+ تغییر کند. اگر نتوانیم مقادیر مربوط به یک زمان را از روی مقادیر مربوط به زمان دیگر، بهتر از حدس زدن پیش‌بینی کنیم، ضریب همبستگی، مقدار صفر را به‌دست می‌دهد. از محاسبه ضریب همبستگی بین اولین مجموعه درخشندگی با دومین مجموعه، داریم

$$\frac{1}{5}[1/5(1/4)+0/5(0/7)+0(0)+(-0/5)(-0/7)+(-1/5)(-1/4)]=0/98$$

این زوج به میزان زیادی همبسته‌اند، گرچه البته، این همبستگی اندکی کوچکتر از همبستگی اعداد اولی با خودشان است. محاسبه ضریب همبستگی مجموعه اول با سوم به ما می‌دهد:

$$\frac{1}{5}[1/5(1/0)+0/5(-1/4)+0(0/4)+(-0/5)(-1/0)+(-1/5)(1/0)]=-0/04$$

یک مقدار منفی نسبتاً کوچک که چندان دور از صفر نیست. مثال بالا بیشتر برای تشریح هدف است، زیرا دراین مورد ضریب همبستگی معمولاً مثبت باقی می‌ماند.

اگر هر درخشندگی استانداردشده با همان مقدار ولی با علامت مخالف جانشین شود، بین اعداد اصلی و مقادیر جدید یک ضریب همبستگی ۱- پیدا می‌کنیم. این مطلب را برای اعداد اولین مجموعه آزمایش می‌کنیم.

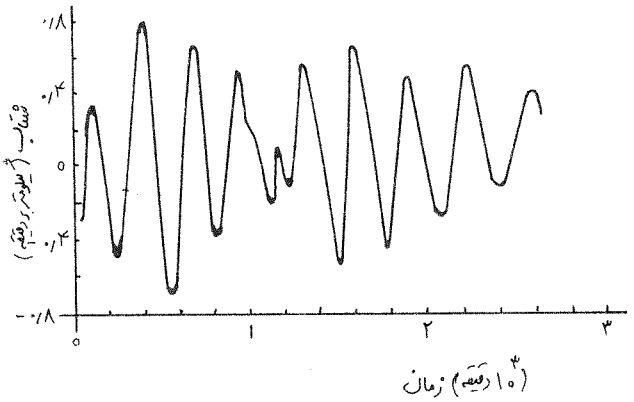
$$\frac{1}{5}[1/5\times(-1/5)+0/5(-0/5)+0(0)+(-0/5)(0/5)+(-1/5)(1/5)]=-1$$

گرچه ضریب همبستگی در دامنه ۱- تا ۱+ تغییر می‌کند، راههای دیگر استانداردسازی نیز امکان دارد. امابازه [۱ و ۱-] قراردادی و **راحت** است.

یک مدل ممکن. اگر نقاط روشن روی یک فیلم، متناظر با نقاط روشن روی فیلم دیگر باشند همبستگی مثبت خواهدبود و این متناظر باحالتی است که فیلمها فاصله زمانی کوتاهی ازهم دارند. اگر نقطه‌های روشنتر روی یکی متناظر با نقاط تاریکتر روی دیگری باشند، ضریب همبستگی منفی است و اگر نتوانیم یک مجموعه را ازروی مجموعه دیگر، بهتراز حدس زدن پیش‌بینی کنیم، همبستگی برابر صفر است.

هرچه فاصله زمانی افزایش یابد،این نتیجه حاصل می‌شود که ضریب همبستگی بدون اینکه واقعاً منفی شود، به‌طور یکنواخت به سوی صفر کاهش می‌یابد. مقدار همبستگی در یک فاصله زمانی ۵ دقیقه‌ای به حدود $\frac{1}{2}$ و در یک فاصله ده‌دقیقه‌ای به حدود کمتراز $\frac{1}{4}$ کاهش پیدا می‌کند. برای این فرایند، مدلی پیشنهاد شده‌است. مطابق این مدل فرض می‌شود که در زمانهای تصادفی به‌طور متوسط حدوداً پنج دقیقه جدا ازهم، دانه‌های جدید ظاهر و تدریجاً سرد می‌شوند. به‌جای هر دانه در زمان تصادفی بعدی دانه دیگری قرار می‌گیرد. اگر عمل سردشدن آهسته صورت گیرد و عمل جانشینی نیز به آرامی روی دهد، ضریب همبستگی به سرعت به صفر می‌گراید. کارهای نظری مربوط به این زمینه، که دراینجا ارائه نشده‌اند، نشان می‌دهند که این مدل همبستگیهایی را به‌وجود می‌آورد که به‌طور متوسط در حدود پنج دقیقه از ۱ به $\frac{1}{2}$

شکل ۲ـ در اینجا نمودار اندازه‌های مولفه قائم سرعت گاز در جو خورشیدی برای یک نقطه روی خورشید رسم شده‌است. حالت تناوبی قابل ملاحظه حرکتها، حالتی نوعی از فوتوسفر خورشیدی است که موضوع بسیاری از مطالعات مربوط به همبستگی بوده‌است.



کاهش می‌یابند. این هماهنگی با واقعیتها تأییدی بر درستی این مدل است.

سرعت

نوع دیگر اندازه‌گیری، یعنی سرعت، در نتایجی که داده بیشتر هیجان‌زا بوده‌است. به‌دست آوردن سرعت مشکلتراست ولی از این داده‌ها به میزان زیادی جمع‌آوری شده‌است واین داده‌ها رفتاری به‌طور قابل ملاحظه متفاوت از رفتار تغییرات درخشندگی همراه با طرح دانه‌ای را نشان می‌دهند. همبستگیهایی که می‌توان برای سرعت اندازه‌گیری کرد، عمدتاً به همان طریقی است که برای درخشندگی ذکر کردیم؛ همبستگیهای اخیر با زمان کاهش می‌یابند اما شدت کاهش آنها زیاد بوده و قبل از برگشت دوباره به صفر، منفی می‌شوند. درواقع یک نمودار تفصیلی از سرعت دریک نقطه خاص از سطح خورشید، نوسانی را نشان می‌دهد که ازنظر نظم ظاهری کاملاً برجسته است. شکل ۲ مثالی از این مورداست؛ این نمودار کاملاً منظم نیست، اما در اینکه واقعاً نوسانی است تردیدی وجودندارد. حال این سئوال مطرح می‌شودکه چرا درحالی که الگوی درخشندگی در فضا و زمان هردو نامنظم است، در سرعت نوسانهایی مشاهده

می‌شوند؟

پاسخ کاملی به این سئوال نشان داده نشده‌است، اما بررسیهای تفصیلی مربوط به ترکیب فراوانی این نوسانها سرآغازی بر آشکارسازی این پدیده و نشان دادن پیچیدگی واقعی آن است. این بررسیهانشان داده‌اند که نوسانها به‌گونه‌ای شگفت‌آور شبیه امواج موسیقی هستند که به‌وسیله یک رشته سیم ویولون منتشر می‌شوند. نتهای فرعی با طنین بالا نیز وجود دارند ولی به‌طور نسبی ضعیفند. درواقع نیمی از انرژی حاصل در نوسانهایی است که دوره آنها از ۲ تا ۶ دقیقه است. به‌عنوان مثالی که در ارتباط باپدیده‌های زمینی است، باید بگوییم که فیلمی با دور کند از امواجی که به ساحل می‌خورند، حدود همین درجه از تناوبی‌بودن را نشان می‌دهد.

منجمان باکشف این حالت برخوردار از نظم دچار شگفتی شدند، زیرا قبلاً جو خورشید را جایگاه محض بی‌نظمی تصور می‌کردند؛ همچنین تغییرات درخشندگی، چنین نوسانهای تناوبی را نشان نداده‌بود واخترشناسان فرض کرده‌بودند که الگوی حرکات به‌سوی بالا وپایین، باید شباهتی نزدیک به بی‌نظمی تغییرات درخشندگی داشته‌باشد.

تناوبی بودن حرکتها و تضاد آشکار آن با تصادفی‌بودن درخشندگی بی‌درنگ نشان می‌داد که منجمان دو بخش متفاوت از جو خورشیدی را مشاهده می‌کردند؛ از آن زمان تاکنون، ثابت شده است که تغییرات درخشندگی در سطح پایین جو مرئی تولید می‌شوند، درحالی که حرکتهای مشاهده‌شده در لایه‌های بالایی جو به‌وقوع می‌پیوندند. آنچه بیشتر قابل ملاحظه است، آن که ماهیت حرکتها با افزایش ارتفاع در جو تغییر می‌کند. متوسط دوره تناوب نصف می‌شود، و در بالاترین ارتفاع قابل مشاهده (چندین هزار مایل بالای سطح خورشید) افت وخیز سرعت بی‌نظم است، که شباهت به سر و صدایی شدید بدون هیچ دوره تناوبی مشخص دارد.

چرا چنین است؟اخترشناسان فرض می‌کنند که ما شاهد گریز به سوی بالای «امواج صوتی» بسیار طولانی در جو خورشیدی هستیم، امواجی که در عمق جو خورشید تولید می‌شوند و شاید به‌وسیله حرکت صعودی دانه‌ها به‌وجود می‌آیند. در آن جو بعضی از امواج، و به‌طور برجسته آن امواجی که دوره تناوبی حدود ۵ دقیقه دارند، به دام می‌افتند؛ امواج با دوره تناوب کوتاهتر، با سرعت به سطوح بالایی می‌گریزند، جایی که فراتر باشند. امواج با دوره تناوب خیلی طولانی سریعاً محو می‌شوند، و درواقع این امواج به آسانی توسط دانه‌ها برانگیخته نمی‌شوند،

به قسمی که این امواج در همه سطوح بسیار ضعیفند.

توضیح مختصر بالا، حتی این مطلب را آشکار می‌کند که مطالعه آماری درباره فراوانیهای این نوسانها ممکن است برخی جنبه‌ها و سیماهای خورشیدی را فاش نماید که عبارتنداز: ماهیت آشفتگیهای عمیقتری که این امواج صوتی را تولید می‌کنند، نرخی که بر آن اساس، امواج با دوره‌های تناوب مختلف همچنانکه انتشار می‌یابند محو می‌شوند و بالاخره اندازه و میزانی که جو خورشیدی قادر است امواج بادوره‌های متناوب مختلف را به دام اندازد.

مسائل

۱ـ محاسبات عدم احتمال، چه نقشی در نظریه ستارگان مزدوج بازی می‌کنند؟

۲ـ دونوع بحثهای آماری که‌اخترشناسان از آنها استفاده می‌کنند، چه هستند؟

۳ـ به جدول ۱ رجوع کنید. ما تنها به اندازه‌گیریهای انجام شده در زمان ۲ توجه داریم. چگونه درایه صفر در نقطه ۳ را با درخشندگی متوسط در همه نقاط مقایسه می‌کنید؟

۴ـ بااستفاده از جدول ۱، ضریب همبستگی را برای مجموعه دوم و سوم اندازه‌ها حساب کنید.

۵ـ از روی مدلی که در متن مقاله توصیف شد، انتظار دارید که ضریب همبستگی برای زمانهای ۱ و ۳ بزرگتر از ضریب همبستگی برای زمانهای ۲ و ۳ باشد؟ آیا محاسبه شما با این انتظار هماهنگ است؟ توضیح دهید.

۶ـ به شکل ۲ رجوع کنید. چندتا از نوسانها قبل از ۰ به سوی پایین هستند؟ قبل از ۶۴ ـ چندتا؟

۷ـ برای ضریب همبستگی اندازه‌های درخشندگی جو خورشیدی که در چندین نقطه خورشید انجام می‌شود، شکلی مشابه شکل ۲ رسم کنید.

۸ـ سرعتهایی که در شکل ۲ معرفی شدند، تناوبی بودن بارزی را نشان می‌دهند. آیا نمودارهای مشابهی را برای نقاط دیگر اندازه‌گیری در جو خورشیدی، انتظار خواهید داشت؟ پاسختان را توضیح دهید.

۹ـ الف) اختلافهای درخشندگی مشاهده‌شده و الگوی سرعتهای گاز در جو خورشیدی را توصیف کنید.

ب) توضیحی جزئی از این تفاوت ارائه دهید.





سال ریاضیات ۱۳۷۹

درباره این کتاب :

کاربردهای گسترده آمار در بسیاری از زمینه‌ها برای عموم مردم و حتی بعضا اهل فن، ناشناخته است و در نتیجه این علم تاکنون به منزلت شایسته آن در جامعه دست نیافته است .

کتاب حاضر، ترجمه‌ای است از ویرایش سوم کتابی که در آن به کاربردهای آمار به زبانی ساده در علوم پزشکی، اقتصاد، حسابداری، سیاست، تاریخ، قضاوت، نجوم، هواشناسی، دیرین شناسی، ورزش و ... در قالب مقاله‌هایی خواندنی و گیرا اشاره شده است .

این کتاب به روشی بدیع خوانندگانی را که دارای هیچگونه معلومات خاصی در رشته آمار نیستند با سودمندی و جذابیتهای یک رشته علمی آشنا می‌سازد. به علاوه دانشجویان و کارشناسان این رشته نیز می‌توانند گستره رشته تحصیلی مورد عمل خود را به زبانی غیرفنی اما دقیق مشاهده کنند.

بنابراین خواندن این کتاب به همه از خاص و عام توصیه می‌شود.