

فعالیت به عنوان یک آماردان: یک توصیه شخصی

دی. آر. کاکس^۱

مترجم: علیرضا خوشگویان فرد^۲

روشن نبود. در شغل دوم خود، پیوسته با مجموعه داده هایی مواجه می شدم که حاصل آزمایشهای به دقت طراحی شده، با شکلی نه کاملاً استاندارد، بودند و من از این بیم داشتم که باید تحلیلهای بسیاری را انجام دهم که از لحاظ فنی بیشتر بر امید حاصل از استقرای موارد ساده تری مبتنی بودند تا در کی مناسب.

در عین حال عجیب این که تا آن زمان یعنی ۱۹۴۵-۱۹۵۰، بسیاری از آنچه اکنون به عنوان روشهای متعارف آمار در نظر گرفته می شوند، توسعه یافته بودند ولی تصور می کنم که در هر حال، خسارچ از گروه های متخصص، به صورتی فراگیر شناخته شده نبودند. به این ترتیب، فیشر^۳ و خصوصاً یتس^۴، طرح آزمایشها و تحلیل واریانس را به شکلی بسیار هنرمندانه و پیشرفته توسعه داده بودند، حتی در آن زمان کمترین توانهای دوم و تقریباً همه ویژگیهای آن، دارای سابقه طولانی بود. سابقه مؤلفه های اصلی، حداقل به اوایل دهه ۱۹۰۰ و همبستگی کانونی به ۱۹۳۵ برمی گردد. روشهای سریهای زمانی در هر دو حوزه زمان و فراوانی توسعه یافته بودند و نیم نظریات قطعی نمونه گیری را تشریح کرده بود.

چرا از این روشها به صورتی وسیعتر استفاده نمی شد؟ بخش عمده پاسخ در این نیست که آنها شناخته شده نبودند، گرچه این موضوع می تواند در این پاسخ سهمی داشته باشد، ولی به استثنای محاسبات خیلی

پرفسور دوج^۲ پیشنهاد کرده است که برخی از نظرات شخصی خود را درباره فعالیت به عنوان یک آماردان بنویسم. من این کار را بیش از هر چیز، از منظر یک بریتانیایی و دیدگاه آماردانی که اکثراً و نه کاملاً در محیط دانشگاهی فعالیت دارد و با کاربردهایی در علوم طبیعی و اجتماعی و فن آوریهای مرتبط با آنها سروکار داشته است، انجام می دهم. جنبه ملیت زیاد مهم نیست، ولی بی شک در صورتی که در دفتر آمار دولتی کار کرده بودم، پیامدهایی که باید مورد بحث قرار می گرفتند، متفاوت بودند، اگرچه تحولات آنجا نیز مطمئناً گسترده اند. هنگامی که در سال ۱۹۴۴ برای اولین بار فعالیت آماری را آغاز کردم، افراد خیلی کمی با آموزش تخصصی دانشگاهی در زمینه آمار وجود داشتند. بیشتر فعالیتهای آماری توسط کسانی صورت می گرفت که یا مانند من در ریاضیات تحصیل کرده بودند و تعداد کمی از آنها دارای حداقل آموزش رسمی در نظریات آماری بودند یا دانشمندان علوم طبیعی به ویژه زیست شناسان یا مهندسانی بودند که خارج از نیازشان به بخشی از دانش آماری آگاهی داشتند. در زمینه آمار کتابهای خیلی کمی وجود داشتند و کتابهای موجود بیشتر به تحلیل واریانس، رگرسیون و غیره، آن هم برای استفاده غیر متخصصان، می پرداختند. این کتابها شامل نظریه نبودند و یا مقصدار کمی از نظریه ها را ارائه می دادند، بنابراین مبنای فکری برای مقابله با مسائل جدید غالباً خیلی

^۴ Fisher

^۵ Yates

^۶ Neyman

^۱ D. R. Cox

^۲ کارشناس آمار اقتصادی - اجتماعی، دانشگاه علامه طباطبائی

^۳ Dodge

محاسبه وارون کردن ماتریس تقریباً با مدت زمانی که متناسب با مکعب اندازه آن است طول می کشد. تصور می کنم بزرگترین وارون ماتریسی که تاکنون من خود به دست آوردم، ماتریس 7×7 ی بود که با بازیبنیهای لازم، بیش از یک روز طول کشید.

نتیجه تمام این کارها در تحلیل داده ها عبارت بود از: نگاه خیلی دقیق به داده ها، تشخیص خصوصیات کلی آن، نقصهای ممکن و غیره، آنگاه تفکر، و سپس محاسبه.

ایده یک رایانه الکترونیکی معلوم بود و خصوصاً اظهارات دوران دانشانه آی. جی. گود و ای. دی. بوث^۱ را به یاد می آورم. آنها به ویژه تأثیری را که رایانه ها بر زندگی روزانه خواهند داشت پیش بینی کردند. جالب است که تقریباً ۵۰ سال سپری شده است تا آنچه آنها پیش بینی کردند به وقوع بپیوندد. گمان می کنم یکی از آنها (و شاید هر دو) آینده آرامتری را تصور می کرد که در آن معجزاتی که با پیشرفت علمی ممکن روی خواهند داد، نیروی محرکی برای یک تلاش مستمر رقابتی دیوانه وار نخواهند شد، بلکه کیفیت عمومی زندگی را بهبود می بخشند.

در آمار، به منظور یک ساده سازی فراگیر در روشها، دوره ای حداکثر از ۱۹۷۵ تا ۱۹۸۰ صرف تهیه شیوه های کلاسیک و توسعه های مستقیم آنها شد به طوری که بتوانند نسبتاً آسان به کار گرفته شوند. از آن پس، با قدرت عظیم رایانه ای، افقهای تازه گشوده شدند، از جمله قابلیت کار کردن با حجم انبوهی از داده ها که در حال بررسی اند.

در مقابله با نگرشی که بیشتر مرور شد، پیامد همه این پیشرفتها در تحلیل حجمهای نسبتاً متعدلی از داده ها عبارت است از: ابتدا محاسبه، سپس تفکر.

از یک نظر، این یک رهایی است و شکی ندارم که روی هم رفته چیز خوبی است، ولی نقطه ضعفی وجود دارد که لازم است همه ما روی آن تأمل کنیم.

فعالیت آماری در حال حاضر، چقدر با زمانی که من شروع به فعالیت کردم، فرق دارد؟ اول این که افراد بسیار بیشتری با تحصیلات آماری وجود دارند و نفوذ عمیق تر ایده های آماری در موضوعات مسائل رشته ها به چشم می خورد برای مثال، پژوهشهای پزشکی را در نظر بگیرید. در بسیاری از حوزه ها، البته نه در همه آنها، مجلات

معمولی، پاسخ، بیشتر، ذکر موانع محاسباتی است. یک آماردان نوعی، در آن زمان از دو یا سه رایانه کمک می گرفت. در آن روزگار، یک رایانه یک فرد بود. چنین فردی قادر بود از یک ماشین محاسباتی الکتریکی برای انجام تحلیل واریانس متعادل استاندارد، محاسبات رگرسیونی ساده با حداکثر دو یا سه متغیر توضیحی، آزمونهای خی - دوی ساده و شاید چند محاسبه تخصصیتر دیگر استفاده کند. همچنین طیف متنوعی از انواع خاص برگه های نمودار برای شیوه های نموداری مختلف وجود داشت. البته محاسبات دیگری نیز می توانستند صورت پذیرند اما به برنامه ریزی خاصی نیاز داشتند.

امکان خطاهای عددی همیشه وجود داشت. محاسبات مهم ترجیحاً از راه دیگری بازیابی می شدند. من این شانس بزرگ را داشتم که مدتی با هنری دانیلز^۲ کار کنم. او نظریه پرداز بسیار توانایی بود و من تقریباً همه آنچه در مورد نظریه می دانستم از وی آموختم. او وسواسی نیز بود. کارهای محاسباتی می بایست با مدادی تیز در حالی که جوابها تا دو (نه سه) رقم در مربعهای کاغذ شطرنجی (با ضلعهای ۱/۴ اینچ) وارد می شدند، صورت گیرد. اشتباهات می بایست پاک (نه خط خطی) و اصلاح می شدند. جدولها و غیره می بایست دارای عناوین مکتوبی باشند.

احتمالاً ملال آورترین محاسبه، مربوط به تحلیل سریهای زمانی بود. داده ها تا دو رقم گرد و دو بار وارد دستگاه محاسبه گر (یک دستگاه قدیمی ثبت کننده پول) می شدند. از این طریق، نوار کاغذی حاوی داده ها در قالب ستونهایی که هر مقدار دو بار در آن وجود داشت، چاپ می شد. این نوار از وسط بریده می شد تا دو نوار ایجاد شود. سپس برای اتصال این نوارها در تغییر مکان h یعنی تأخیر^۳ مورد نظر برای همبستگی متوالی، از سنجاقی استفاده می شد. مقادیر متناظر به طور ذهنی کم می شدند و پس از آن به صورت ذهنی به توان دو می رسیدند و معمولاً جواب توسط فرد دومی در ماشین دیگری وارد می شد. این محاسبات، $\sum (y_{i+h} - y_i)^2$ را تولید می کرد که از آن تأخیر h همبستگی متوالی به دست می آمد. این کار به تعداد h هایی که لازم بود، تکرار می شد، مقدار h برحسب نیاز به تأمین نظر پژوهشگران محدود می شد. البته من دانستن توانهای دوم اعداد صحیح حداکثر تا مثلاً حدود ۵۰ را مسلم می دانم.

Henry Daniels^۲Lag^۳I. J. Good and A. D. Booth^۱

برای این منظور درک مناسبی از ایده‌ها و نگرشهای اساسی موضوع‌مان و نگرشی مناسب از پژوهش علمی، بیش از اطلاعات خیلی تخصصی، مورد نیاز است. غالباً جلوی پیشرفت کار خود من، بیشتر به دلیل زمینه علمی ناکافی در آن موضوع و خصوصاً در سالهای اخیر با پایه نظری نامناسب در آمار به جای عدم آشنایی با روشهای تخصصی خاص، گرفته شده است.

اگرچه امروزه اکثر پژوهشها به جای انفرادی به صورت گروهی انجام می‌شوند، ولی پژوهش خواه نظری باشد خواه کاربردی، یک فعالیت عمیقاً شخصی باقی می‌ماند و بهترین شکل فعالیت باید توسط هر یک از ما با آزمایش و خطا به دست آید. این موضوع در مرحله مهیج و دشواری است و من به کسانی که حرفه‌ای را در فعالیت آماری آغاز می‌کنند، غبطه می‌خورم.

پژوهشی پزشکی با موضوعات تخصصی بر استفاده از روشهای آماری کاملاً پیشرفته ای متکی هستند که حتی تا ۲۰ سال پیش قابل تصور نبود. قلمرو بحث، کمک گیری و مشاوره با همکاران خبره، خیلی بیشتر از گذشته است؛ در مجموع، آماردانان خیلی کمی به صورت انفرادی کار می‌کنند. در بالا به کمبود نسبی کتاب در آغاز فعالیتم، اشاره کردم. مسأله در حال حاضر، بیشتر در میزان تقریباً سرسام آور مکتوبات است و نه ظهور هر ساله تعداد زیادی کتاب که اکثراً سطح استاندارد بالایی دارند. البته این پیشرفت فنی موضوعات موجب می‌شود که آماردانان بیش از پیش در زمینه‌های خاص متخصص شوند.

با این وجود از جهتی در شگفتم که آیا امور، همه در سطحی عمیقتر از این مقدار تغییر کرده اند یا نه. در واقع، تقریباً موضوعهای جدی مسائل پژوهشی همیشه دارای ویژگیهای واحدی هستند که لازم است حداقل از طریق اصول اولیه تا حدی مورد بررسی قرار گیرند و

مقاله فوق در مجله Student، سال ۲۰۰۰، جلد سوم شماره ۴،

صفحات ۲۸۸ تا ۲۹۰ به چاپ رسیده است.

Student, (2000), Vol. 4, pp. 288-290