چگونه دادگاهی توضیحی غیرممکن را پذیرفت

جوزف گست ویرث و همکاران
ترجمهٔ حمیدرضا نواب‌پور*

چکیده

گرچه غالباً این امر درست است که پیوند بین دو متغیر ممکن است به علت متغیر سومِ مشاهده نشده‌ای باشد، ولی موجه بودن چنین استدلالی نیازمند بررسی دقیق است. در سال ۱۹۸۷، یک دادگاه کانادایی توضیح ارائه شده از سوی اداره مالیات بردرآمد[1] کانادا را مبنی بر اینکه دلیل پایین بودن نرخ قبولی زنان نسبت به مردان در یک آزمون ارتقای شغلی، وجود تفاوت در میزان تحصیلات دانشگاهی آنهاست، پذیرفت و این متغیری بود که دادگاه مستقیماً آن را مورد مشاهده قرار نداده بود. در این مقاله نشان می‌دهیم این توضیح نه تنها موجه نیست، بلکه غیرممکن است. با بحثی مختصر دربارهٔ نقش آماردانان در مباحثاتی که متضمن یک متغیرِ مشاهده نشده است، مقاله را به پایان می‌بریم.

۱ متغیری که هرگز وجود نداشت

«یک پیوند شاید غیرواقعی باشد.» «پیوند بین دو متغیر شاید به علت متغیر سوم مشاهده نشده‌ای باشد.» گرچه این احکام درست‌اند، ولی هر چنین استدلالی نمی‌تواند موجه باشد. ما ماجرای دادگاهی‌را بازگو می‌کنیم که استدلالی متضمن یک متغیر مشاهده نشده را پذیرفت، به متغیری که، همچنانکه نشان می‌دهیم، نمی‌تواند وجود داشته باشد. این ماجرا در بخش ۱ نقل می‌شود، برهان رسمی آن در بخش ۲ ارائه می‌شود و پندی را که از آن می‌گیریم و چند توصیه عملی را در بخش ۳ خلاصه می‌کنیم.

در حالت مالولی[2] علیه ادارهٔ مالیات بردرآمد کانادا، آن‌گونه که جوریانتس[3] (۱۹۸۷) توصیف کرده است، این اداره مبنای ارتقای کارکنان خود به پست ممیز مالیاتی را یک آزمون روانسنجی[4] قرار داده بود. بر پایهٔ این آزمون، ۵۹% مردان و ۲۷% زنان قبول شدند، جدول ۱ را ببینید. بنابر قوانین کانادا، اداره مالیات بردرآمد می‌بایست ثابت می‌کرد که این آزمون، روشی معتبر و قابل اطمینان برای گزینش داوطلبان بر اساس شایستگیهای آنها بوده است.

جدول ۱- فراوانی قبول شدگان

	قبول شده	رد شده	مجموع	درصد قبول شدگان
زن	۶۸	۱۸۳	۲۵۱	%۲۷
مرد	۶۸	۴۷	۱۱۵	%۵۹
مجموع	۱۳۶	۲۳۵	۳۶۶	%۳۷

جدول ۲ــ فراوانی تحصیلات دانشگاهی

	دارای تحصیلات دانشگاهی	بدون تحصیلات دانشگاهی	مجموع	درصد تحصیلات دانشگاهی
زن	۶۳	۱۸۸	۲۵۱	%۲۵
مرد	۶۰	۵۵	۱۱۵	%۵۲
مجموع	۱۲۳	۲۴۳	۳۶۶	%۳۴

پس از اینکه در دادگاه علیه ادارهٔ مالیات بردرآمد اقامهٔ دعوی شد، این اداره با رجوع به جدول ۲ که نشان می‌داد ۵۲% مردان و ۲۵% از زنان تحصیلات دانشگاهی داشتند، از به کارگیری آزمون دفاع کرد. دادگاه استیناف، این ادعای ادارهٔ مالیات بردرآمد را پذیرفت، که تفاوت نرخ قبول‌شدگان در بین زنان و مردان تبعیض‌آمیز نبوده است، زیرا این تفاوت منعکس کنندهٔ اختلافی در

*) دکتر حمیدرضا نواب‌پور، گروه آمار، دانشگاه علامهٔ طباطبائی

1) National Revenue Service 2) Maloly 3) Jurianz 4) Collections enforcement clerk

قوهٔ درک آنها بوده است و داده‌های مربوط به تحصیلات دانشگاهی نیز آن را تائید می‌کرد. دادگاه استیناف نتیجه گرفت که فراوانی قبول‌شدگان صرفاً با فراوانی تحصیلات دانشگاهی هماهنگی دارد، گرچه ادارهٔ مالیات بردرآمد اطلاعات مربوط به نرخ قبولی را برحسب جنس و تحصیلات جزو مدارک ارائه نکرده بود.

آیا قضاوت دادگاه با توجه به مدرکی که برای بررسی انتخاب کرد، اشتباه بود؟ آیا تفاوت نرخ قبولی مردان و زنان با تفاوت نرخ تحصیلات دانشگاهی آنها سازگار است؟ آیا مردان و زنانی که از لحاظ تحصیلات دانشگاهی یکسان باشند، نرخ قبولی مشابهی دارند؟ یا آیا تفاوت نرخ قبولی فاحشتر از آن نیست که صرفاً منعکس کنندهٔ تفاوت تحصیلات دانشگاهی باشد؟

داده‌ها غیرعادی‌اند. جدولهای ۱ و ۲ دو جدول 2×2 حاشیه‌ای یک جدول $2 \times 2 \times 2$اند، ولی دادگاه استیناف نه جدول کامل $2 \times 2 \times 2$ را درخواست کرد و نه هرگز آن را دید. در مراحل قانونی باید دو طرف دعوا زحمت تهیهٔ مدرک را به خود بدهند، نه هیأت استیناف یا هیأت منصفه. غالباً استدلالهای مربوط به متغیرهای مشاهده نشده به صورتی ضعیف ارائه می‌شود. استدلالی را که دادگاه دربارهٔ این جدولها پذیرفت نه تنها موجه نبود بلکه غیرممکن بود. این مثال حائز اهمیت است، زیرا نشان می‌دهد که برخی ادعاها دربارهٔ متغیرهای مشاهده نشده، صرفاً متوقعانه یا دور از دسترس نیستند بلکه برخی از آنها کاملاً نادرست‌اند.

اگر جدول $2 \times 2 \times 2$ مورد مشاهده قرار می‌گرفت، تحلیل متعارف، نرخ قبولی برای زنان و مردان را که نسبت به تحصیلات با استفاده از آمارهٔ مانتل-هانتسل[5] (۱۹۵۹) تعدیل شده است، مقایسه می‌کرد. برای نمونه بیکل[6]، هَمِل و اوکانل[7] (۱۹۷۵)، فلیز[8] (۱۹۸۱)، گَشت‌ویرث[9] (۱۹۸۸) یا آگِرِستی[10] (۱۹۹۰) را ببینید.

می‌توان هر جدول $2 \times 2 \times 2$ی ممکن را که جدولهای ۱ و ۲ را تولید می‌کند، درنظر گرفت. تعداد این جدولهای $2 \times 2 \times 2$ متناهی است. برای هر یک از آنها آمارهٔ مانتل-هانتسل را محاسبه می‌کنیم. محاسبه‌ای ساده‌تر که نتیجه‌ای یکسان به دست می‌دهد، در بخش ۲ آورده می‌شود. زمانی که این کار انجام شد، بزرگترین انحراف نرمال که می‌توان از جدولهای $2 \times 2 \times 2$ هماهنگ با جدولهای ۱ و ۲ تولید کرد، انحراف ۳٫۱۱- با p- مقدارِ ۰٫۰۰۲ است. بنابراین هیچ جدول $2 \times 2 \times 2$ای هماهنگ با جدولهای ۱ و ۲ که در آن تحصیلات دانشگاهی بتواند تفاوت نرخ قبولی مردان و زنان را توضیح دهد، وجود ندارد. ممکن است آزمون برای زنان تبعیض‌آمیز باشد یا نباشد، ولی در هر حال استدلالی که دادگاه استیناف پذیرفت، کاملاً نادرست است. تفاوت نرخهای قبولی به مراتب بیشتر از آن است که بتواند با تفاوت موجود در نسبتهای تحصیلات دانشگاهی قابل توضیح باشد.

۲ آماره مانتل_هانتسل فرین از یک جدول $2 \times 2 \times 2$ با جدولهای حاشیه‌ای مفروض

جدول $2 \times 2 \times 2$، با $n_{egp} =$ تحصیلات $(e)\times$ جنس $(g)\times$ قبولی (p) مدخل را در نظر بگیرید، که در آن $e = 1$ در صورت داشتن تحصیلات دانشگاهی و $e = 2$ در غیر این صورت، $g = 1$ برای زن بودن و $g = 2$ برای مرد بودن و $p = 1$ برای قبول شدن و $p = 2$ برای رد شدن. فرض کنید m تعداد قبول شدگانی باشد که تحصیلات دانشگاهی دارند، یعنی $m = n_{1.1}$، بنابراین m را نمی‌توان از جدولهای ۱ و ۲ به دست آورد. جدول $2 \times 2 \times 2$ به‌جز جدولهای حاشیه‌ای ۱ و ۲، جدول حاشیه‌ای دیگری نیز دارد که همان جدول ۳ است.

جدول ۳- جدول حاشیه‌ای 2×2ی مشاهده نشده

	قبول شده	رد شده	مجموع
تحصیلات دانشگاهی	m	$n_{1..} - m$	$n_{1..} = 123$
تحصیلات غیردانشگاهی	$n_{1..} - m$	$n_{...} - n_{..1} - n_{1..} + m$	$n_{2..} = 243$
مجموع	$n_{..1} = 136$	$n_{...} - n_{..1} = 230$	$n_{...} = 366$

انحراف استاندارد شده مانتل_هانتسل که نرخهای قبول‌شدگان برای مردان و زنان را درون طبقات تعریف شده به وسیله تحصیلات دانشگاهی، مقایسه می‌کند به صورت زیر است:

$$T(m) = \frac{n_{.11} - \left(\frac{n_{11.}.m}{n_{1..}} + \frac{n_{21.}.(n_{..1}-m)}{n_{2..}}\right)}{\sqrt{\left[\frac{n_{11.}.n_{12.}.m(n_{1..}-m)}{n_{1..}^{2}(n_{1..}-1)} + \frac{n_{21.}.n_{22.}.(n_{..1}-m)(n_{2..}+m-n_{..1})}{n_{2..}^{2}(n_{2..}-1)}\right]}} \qquad (1)$$

توجه داشته باشید که m تنها کمیتی در برابری (۱) است که از جدولهای ۱ و ۲ تعیین نمی‌شود. جالب است توجه کنیم که m، آمارهٔ مانتل_هانتسل و سطح معنی‌داری بزرگ نمونه‌ای آن را تعیین می‌کند، ولی سطح معنی‌داری را که نیاز به مشخص‌شدن کمیتهای بیشتری است، معین نمی‌کند.

اگرچه m نامعلوم است ولی مقیّد به داده‌های مشاهده شده است، $a \leq m \leq b$،

$$a = \max(0, n_{.11} - n_{21.}) + \max(0, n_{.21} - n_{22.})$$

5) Mantel-Haenszel 6) Bickel 7) Hammel, and O'Connell 8) Fleiss 9) Gastwirth 10) Agresti

$a \le m \le b$،

$$a = \max(۰, n_{.۱۱} - n_{۲۱.}) + \max(۰, n_{.۲۱} - n_{۲۲.})$$

و $b = \min(n_{.۱۱}, n_{۱۱.}) + \min(n_{.۲۱}, n_{۱۲.})$. در یک روش مستقیم، هر m در این محدوده امتحان و نتیجه می‌شود که $\max\{T(m)\} = -۳٫۱۱$، و سطح معنی‌داری تقریبی ۰٫۰۰۲ است. به عبارت دیگر، هر چند m مشاهده نشده است، ولی هیچ مقدار ممکنی برای m به طوری که تفاوت نرخ قبولی زنان و مردان را پس از تعدیل نسبت به سطح تحصیلات توضیح دهد، وجود ندارد.

در واقع، لزومی به امتحان هر مقدار ممکن m نیست. با گرفتن $T(m)$ به عنوان تابعی مشتق‌پذیر از متغیر حقیقی m با مشتق $T'(m)$، لم زیر قابل اثبات است.

لم. برای $a \le m \le b$، مشتق $T'(m)$ حداکثر برای یک مقدار m صفر است.

برهان. در برابری (۱)، $T(m)$ را به صورت $T(m) = \frac{N(m)}{\sqrt{D(m)}}$ می‌نویسیم، بنابراین

$$T'(m) = \frac{N'(m)\sqrt{D(m)} - \frac{N(m)D'(m)}{۲\sqrt{D(m)}}}{D(m)} \qquad (۲)$$

از آنجا که $D(m) > ۰$، علامت $T'(m)$ در برابری (۲) مانند علامت عبارت

$$T'(m)[D(m)]^{۳/۲} = N'(m)D(m) - (۱/۲)N(m)D'(m)$$

است. از (۱)، برای ثابتهای معین d، e، f، g، و h که به جدولهای حاشیه‌ای مشاهده شده بستگی دارند ولی به m بستگی ندارند،

$$N(m) = d + em \qquad , \qquad D(m) = f + gm + hm^۲$$

بنابراین

$$N'(m)D(m) - (۱/۲)N(m)D'(m) = (ef - \frac{dg}{۲}) + m(\frac{eg}{۲} - dh)$$

که نسبت به m خطی است و لم ثابت می‌شود.□

از این لم نتیجه می‌شود که برای عدد صحیح m بین a و b، انحراف $T(m)$ یا نسبت به m یکنواست و یا اینکه $T(m)$ از a تا عددی صحیح مانند r یکنواست با $a < r < b$، و سپس از r تا b یکنوا در جهت مخالف است. در هر حالت حداقل مقدار $T(m)$ برای $a < m < b$، $\min\{T(a), T(r), T(b)\}$ و حداکثر مقدار آن، $\max\{T(a), T(r), T(b)\}$ است.

به یاد آورید که m تعداد قبول‌شدگانی است که تحصیلات دانشگاهی دارند، یعنی $m = n_{۱.۱}$، جدول ۳ را ببینید. برای داده‌های بخش ۱، $m = ۱۱۶$ حداکثر انحراف مانتل‌ـهانتسل در برابری (۲)، را به صورت ۳٫۱۱− و حداکثر سطح معنی‌داری ۰٫۰۰۲ متناظر با آن را به دست می‌دهد. توجه داشته باشید که نسبت بختها در جدول ۴، عدد ۱۸۵ است. بنابراین قبولی شخصی با مقداری تحصیلات دانشگاهی، ۱۸۵ برابر محتملتر از شخصی است که تحصیلات دانشگاهی ندارد که رابطه‌ای به طور خارق‌العاده قوی است. اگر کسی گرایش به این فرض داشته باشد که نسبت بختهایی که ارتباط تحصیلات دانشگاهی و قبول‌شدن در آزمون را بیان می‌کند قدری کوچکتر است، آنگاه حداکثر انحراف مانتل‌ـهانتسل و سطح معنی‌داری نیز کوچکتر می‌شد. برای مثال، اگر نسبت بختها بیشتر از ۱۰ نمی‌بود، حداکثر انحراف مانتل-هانتسل ۳٫۹۷− با سطح معنی‌داری کوچکتر از ۰٫۰۰۰۱ می‌شد.

جدول ۴- جدولی که آمارهٔ مانتل-هانتسل
را تا حد ممکن به صفر نزدیک می‌سازد

	قبول شده	ردشده	جمع
تحصیلات دانشگاهی	۱۱۶	۷	۱۲۳
بدون تحصیلات دانشگاهی	۲۰	۲۲۳	۲۴۳
جمع	۱۳۶	۲۳۰	۳۶۶

توجه: نسبت بختها = ۱۸۴٫۸.

۳ نکتهٔ پندآموز این ماجرا

پاره‌ای از مباحثه‌های مربوط به متغیرهای مشاهده نشده قطعاً نادرست‌اند. برخی، از لحاظ اصولی ممکن‌اند، ولی موجه نیستند و برخی دیگر کاملاً موجه‌اند. مسئولیت آماردان در مواجهه با استدلالی که متضمن متغیرهای مشاهده نشده است، روشن کردن پیامدهای عینی آن استدلال است تا شنوندگان موجه بودن یا نبودن آن را ارزیابی کنند.

برخی از روشهای کلی برای ارزیابی استدلالی مربوط به متغیرهای مشاهده نشده در گشت‌ویرث (۱۹۹۲)، گرین‌هاوس (۱۹۸۲)، روزن بام و کریجر (۱۹۹۰) و روزن‌بام (۱۹۹۱، ۱۹۹۳) بیان شده‌اند.

مراجع

[1] Agresti, A., (1990), *Cartegorical Data Analysis,* New York: John Wiley.

[2] Bickel, P., Hammel, E., and O'Connell J., W. (1975), "Sex Bias in Graduate Admissions: Data From Berkely," *Science,* 187, Feb. 7, 1975-398-404; reprinted in W. Fairley and F. Mosteller (eds.) (1977), Statistics and Public Policy, Reading, MA: Addision-Wesley, PP.110-130.

[3] Fleiss, J. (1981), *Statistical Methods for Rates and Proportions,* New York: John Wiley.

[4] Gastwirth, J. (1988), *Statistical Reasoning in Law and Public Policy,* New York; Academic Press. ____________ (1992), "Method for Assessing the Sensitivity of Statistical Comparisons Used in Title VII Cases to Omitted Variables, " *Jurimetrics,* 33, 19-34.

[5] Greenhouse, S.W. (1982), "Jerome Cornfield's Contributions to Epidemiology," *Biometrics,* 28, supplement, 33-45.

[6] Juriansz, R. G. (1987), "Recent Developments in Canadian Law; Anti-Discrimination Law Part II," *Ottawa Law Review* 19, 667-721.

[7] Mantel, N., and Haenszel, W. (1959), "Statistical Aspects of Retrospective Studies of Diseases," *Journal of the National Cancer Institute,* 22, 719-748.

[8] Rosenbaum, P., and Kriager, A. (1990), "Sensivity Analysis for Two Sample Permutation Inferences in Observational Studies, " *Journal of American Statistical Association,* 85, 493-498.

[9] Rosenbaum, P. R (1991), "Discussing Hidden Bias in Observational Studies," *Annals of Internal Medicine,* 115, 910-905, Dec. 1, 1991.

[10] ____________ (1993), "Hodges-Lehmann Point Estimates of Treatment Effect in Observational Studies, " *Journal of the American Statistical Association,* 88, 1250-1253.

[11] ____________ (in press), *Observational Studies,* New York: Springer-Verlag.

اصل این مقاله در

The American Statistician, November, 1994, Vol. 48, No. 4

به چاپ رسیده است.