

چگونه دادگاهی توضیحی غیرممکن را پذیرفت

جوزف گست ویرث و همکاران
ترجمه حمیدرضا نوابپور*

چکیده

در حالت مالولی^۱ علیه اداره مالیات برآمد کانادا، آن گونه که جوریانتس^۲ (۱۹۸۷) توصیف کرده است، این اداره مبنای ارتقای کارکنان خود به پست ممیز مالیاتی را یک آزمون روانستجوی^۳ قرار داده بود. بر پایه این آزمون، ۵۹٪ مردان و ۲۷٪ زنان قبول شدند، جدول ۱ را ببینید. بنابر قوانین کانادا، اداره مالیات برآمد می‌باشد ثابت می‌کرد که این آزمون، روشی معتبر و قابل اطمینان برای گزینش داوطلبان بر اساس شایستگی‌های آنها بوده است.

جدول ۱ - فراوانی قبول شدگان

	مجموع	درصد قبول شدگان	قبول شده	رد شده	مجموع
%۲۷	۲۵۱	۱۸۳	۶۸	۱۱۵	۲۲۵
%۵۹	۱۱۵	۴۷	۶۸	۲۳۵	۳۶۶
%۳۷	۳۶۶	۱۳۶			مجموع

جدول ۲ - فراوانی تحصیلات داشتگاهی

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

پس از اینکه در دادگاه علیه اداره مالیات برآمد اقامه دعوا شد، این اداره با رجوع به جدول ۲ که نشان می‌داد ۵۲٪ مردان و ۲۵٪ زنان تحصیلات داشتگاهی داشتند، از به کارگیری آزمون دفاع کرد. دادگاه استیناف، این ادعای اداره مالیات برآمد را پذیرفت، که تفاوت نزدیک قبول شدگان در بین زنان و مردان تبعیض‌آمیز نبوده است، زیرا این تفاوت منعکس کننده اختلافی در

گرچه غالباً این امر درست است که پیوند بین دو متغیر ممکن است به علت متغیر سوم مشاهده شده‌ای باشد، ولی موجه بودن چنین استدلالی نیازمند بررسی دقیق است. در سال ۱۹۸۷، یک دادگاه کانادایی توضیح ارائه شده از سوی اداره مالیات برآمد^۱ کانادا را مبنی بر اینکه دلیل پایین بودن نزدیکی زنان نسبت به مردان در یک آزمون ارتقای شغلی، وجود تفاوت در میزان تحصیلات دانشگاهی آنهاست، پذیرفت و این متغیر 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 درباره نقش آماردانان در مباحثاتی که متنضم یک متغیر 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 شده‌ای باشد.» گرچه این احکام درست‌اند، ولی هر چنین استدلالی نمی‌تواند موجه باشد. ما ماجراهای دادگاهی را بازگو می‌کنیم که استدلالی متنضم 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.

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

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

تبیین آمیز 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.

قوه درک آنها بوده است و داده‌های مربوط به تحصیلات دانشگاهی نیز آن را تائید می‌کرد. دادگاه استیناف نتیجه گرفت که فراوانی قبول شدگان صرفاً با فراوانی تحصیلات دانشگاهی هماهنگی دارد، گرچه اداره مالیات بردرآمد اطلاعات مربوط به نرخ قبولی را بر حسب جنس و تحصیلات جزو مدارک ارائه نکرده بود.

۲ آماره مانتل-هانتسل فرین از یک جدول

۲ × ۲ با جدولهای حاشیه‌ای مفروض

جدول $2 \times 2 \times 2$ ، با $n_{epp} = \text{تحصیلات}(e) \times \text{جنس}(e) \times \text{قبولی}(p)$ مدخل را در نظر بگیرید، که در آن $e = 1$ در صورت داشتن تحصیلات دانشگاهی و $e = 2$ در غیر این صورت، $1 = g$ برای زن بودن و $2 = p$ برای مرد بودن و $1 = p$ برای قبول شدن و $2 = g$ برای رد شدن. فرض کنید m تعداد قبول شدگانی باشد که تحصیلات دانشگاهی دارند، یعنی $n_{1..} = n_{11} + n_{12}$ ، بنابراین $m = n_{1..}$ را نمی‌توان از جدولهای ۱ و ۲ به دست آورد. جدول $2 \times 2 \times 2$ به جزء جدولهای حاشیه‌ای ۱ و ۲، جدول حاشیه‌ای دیگری نیز دارد که همان جدول ۳ است.

جدول ۳ - جدول حاشیه‌ای 2×2 می‌مشاهده نشده

مجموع	رد شده	قبول شده	
$n_{1..} = 123$	$n_{1..} - m$	m	تحصیلات دانشگاهی
$n_{2..} = 243$	$n_{2..} - n_{1..} - n_{1..} + m$	$n_{1..} - m$	تحصیلات غیردانشگاهی
$n_{...} = 366$	$n_{...} - n_{1..} = 230$	$n_{1..} = 126$	مجموع

انحراف استاندارد شده مانتل-هانتسل که نرخهای قبول شدگان برای مردان و زنان را درون طبقات تعریف شده به وسیله تحصیلات دانشگاهی، مقایسه می‌کند به صورت زیر است:

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

توجه داشته باشید که m تنها کمیتی در برای (۱) است که از جدولهای ۱ و ۲ تعیین نمی‌شود. جالب است توجه کنیم که m ، آماره مانتل-هانتسل و سطح معنی داری بزرگ نموده‌ای آن را تعیین می‌کند، ولی سطح معنی داری را که نیاز به مشخص شدن کمیتهای بیشتری است، معین نمی‌کند.

اگرچه m نامعلوم است ولی محدود به داده‌های مشاهده شده است،

$$a \leq m \leq b$$

$$a = \max(0, n_{11} - n_{1..}) + \max(0, n_{21} - n_{2..})$$

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 فاقد است 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 رحمت تهیه مدرک را به خود بدھند، نه هیأت استیناف یا هیأت منصفه. غالباً استدلالهای مربوط به متغیرهای مشاهده نشده به صورتی ضعیف ارائه می‌شود. استدلالی را که دادگاه درباره این جدولها پذیرفت نه تنها موجه نبود بلکه غیرممکن بود. این مثال حائز اهمیت است، زیرا نشان می‌دهد که برخی ادعاهای درباره 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 آماره مانتل-هانتسل^۵ (۱۹۵۹) تبدیل شده است، مقایسه می‌کرد. برای نمونه بیکل^۶، بکل و اوکانل^۷ (۱۹۷۵)، فلیز^۸ (۱۹۸۱)، گشت‌ویرث^۹ (۱۹۸۸) یا آگرستی^{۱۰} (۱۹۹۰) را ببینید.

می‌توان هر جدول $2 \times 2 \times 2$ می‌مکن 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 آماره 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 $11 - 3 = 8$ با $p = \text{مقدار } 2^{10} = 1024$ 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

5) Mantel-Haenszel

6) Bickel

7) Hammel

and O'Connell

8) Fleiss

9) Gastwirth

10) Agresti

حداکثر مقدار آن، $\max\{T(a), T(r), T(b)\}$ است.

$$a \leq m \leq b$$

به یاد آورید که m تعداد قبول شدگانی است که تحصیلات دانشگاهی دارند، یعنی $m = n_{11}$ ، جدول ۳ را بینید. برای داده های بخش ۱، $m = 116$ حداکثر انحراف مانتل-هانتسل در برابری (2) ، را به صورت $-3, 11$ - و حداکثر سطح معنی داری $0,002$ متناظر با آن رابه دست می دهد. توجه داشته باشید که نسبت بختها در جدول 4 ، عدد 185 است. بنابراین قبولی شخصی با مقداری تحصیلات دانشگاهی، 185 برابر محتملتر از شخصی است که تحصیلات دانشگاهی ندارد که رابطه ای به طور خارق العاده قوی است. اگر کسی گرایش به این فرض داشته باشد که نسبت بختها که ارتباط تحصیلات دانشگاهی و قبول شدن در آزمون را بیان می کند قدری کوچکتر است، آنگاه حداکثر انحراف مانتل-هانتسل و سطح معنی داری نیز کوچکتر می شد. برای مثال، اگر نسبت بختها بیشتر از 10 نمی بود، حداکثر انحراف مانتل-هانتسل $-3,97$ - با سطح معنی داری کوچکتر از $0,0001$ می شد.

جدول ۴- جدولی که آماره مانتل-هانتسل را تا حد ممکن به صفر نزدیک می سازد

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

توجه: نسبت بختها = $184,8$.

۳ نکته پنداموز این ماجرا

پاره ای از مباحثه های مربوط به متغیرهای مشاهده نشده قطعاً نادرست است. برخی، از لحاظ اصولی ممکن اند، ولی موجه نیستند و برخی دیگر کاملاً موجه اند. مستولیت آماردان در مواجهه با استدلالی که متضمن متغیرهای مشاهده نشده است، روشن کردن پیامدهای عینی آن استدلال است تا شوندگان موجه بودن یا نبودن آن را ارزیابی کنند.

برخی از روشهای کلی برای ارزیابی استدلالی مربوط به متغیرهای مشاهده نشده در گشت ویرث (1992) ، گرین هاوس (1982) ، روزن بام و کریجر (1990) و روزن بام $(1991, 1993)$ بیان شده اند.

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

و $b = \min(n_{11}, n_{11} + \min(n_{21}, n_{21}))$. در یک روش مستقیم، هر m در این محدوده امتحان و نتیجه می شود که $m = -3, 11$ و $\max\{T(m)\} = -3, 11$ و سطح معنی داری تقریبی $0,002$ است. به عبارت دیگر، هر چند m مشاهده نشده است، ولی هیچ مقدار ممکنی برای m به طوری که تفاوت نزدیکی زنان و مردان را پس از تعديل 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 مشتقه از متغیر حقیقی m با مشتق $T'(m)$ ، لم زیر قابل اثبات است.

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

برهان. در برابری (1) ، $T(m)$ را به صورت $T(m) = \frac{N(m)}{\sqrt{D(m)}}$ می نویسیم، بنابراین

$$T'(m) = \frac{N'(m)\sqrt{D(m)} - \frac{N(m)D'(m)}{\sqrt{D(m)}}}{D(m)} \quad (2)$$

از آنجا که $D(m) > 0$ ، علامت $T'(m)$ در برابری (2) مانند علامت عبارت

$$T'(m)[D(m)]^{3/2} = N'(m)D(m) - (1/2)N(m)D'(m)$$

است. از (1) ، برای ثابتی معین d, e, f, g ، و h که به جدولهای حاشیه ای مشاهده شده بستگی دارند ولی به m بستگی ندارند،

$$D(m) = f + gm + hm^2, \quad N(m) = d + em$$

بنابراین

$$N'(m)D(m) - (1/2)N(m)D'(m) = (ef - \frac{dg}{2}) + m(\frac{eg}{2} - dh)$$

که نسبت به m خطی است و لم ثابت می شود. □

از این لم نتیجه می شود که برای عدد صحیح m بین a و b ، انحراف $T(m)$ یا نسبت به m یکنواست و یا بینکه $T(m)$ از a تا عددی صحیح مانند r یکنواست با $a < r < b$ یکنوا درجهت مخالف است. در هر حالت حداقل مقدار $T(m)$ برای $a < m < b$ و $\min\{T(a), T(r), T(b)\}$ است.

مراجع

- [1] Agresti, A., (1990), *Categorical 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: Addison-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), "Sensitivity 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.