Fairley, D. (1986), "Cherry Trees with Cones?," *The American Statistician*, 40, 138–139.
Hakkila, P. (1989), *Utilization of Residual Forest Biomass*, Berlin: Springer-Verlag.
McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.
Ranneby, P. (ed.) (1982), *Statistics in Theory and Practice (Essays in Honour of Bertin Matérn)*, Umeå: Swedish University of Agricultural Science.
Ryan, B. F., Joiner, B. L., and Ryan, T. A. (1985), *Minitab Handbook* (2nd ed.), Boston: Duxbury Press.
Schumacher, F. X., and Hall, F. dos S. (1933), "Logarithmic Expression of Timber Tree Volume," *Journal of Agricultural Research*, 45, 741–756.
Shih, J.-Q. (1993), "Regression Transformation Diagnostics in Transform-Both-Sides Model," *Statistics and Probability Letters*, 16, 411–420.
Spurr, S. H. (1952), *Forest Inventory*, New York: Ronald Press.

How a Court Accepted an Impossible Explanation

Joseph GASTWIRTH, Abba KRIEGER, and Paul ROSENBAUM

Although it is often true that the association between two variables may be due to some unobserved third variable, the plausiblity of such arguments needs careful examination. In 1987, a Canadian court accepted an explanation offered by the National Revenue Service of Canada that, in a test for promotion, lower pass rates among women than among men were explained by differences in rates of college attendance, a variable that was not directly observed by the court. We show that this explanation is not merely implausible. It is impossible. We conclude with a brief discussion of the role of a statistician in an argument involving an unobserved variable.

KEY WORDS: Contingency table; Law and statistics, Mantel–Haenszel statistic; Sensitivity analysis; Unobserved variable.

1. A VARIABLE THAT NEVER WAS

"An association may be spurious." "The association between two variables may be due to some unobserved third variable." True as these statements are, not every such argument is plausible. We tell the tale of a court that accepted an argument involving an unobserved variable—a variable that, as we show, cannot exist. The tale is told in Section 1. A formal demonstration is given in Section 2. The moral of the tale and some practical advice are summarized in Section 3.

In the case of Maloley *v.* National Revenue Service of Canada, as described by Juriansz (1987), the Revenue Service had promoted employees to the position of collections enforcement clerk using a psychological test, the General Intelligence Test. On the basis of this test, 59% of males passed and 27% of females passed; see Table 1. Under Canadian law, the Revenue Service had to prove the test was a reliable and valid way of selecting candidates according to their merit. When challenged in court, the Revenue Service defended its use of the test with reference to Table 2, which showed that 52% of males and 25% of females had some college education. The Appeals Board accepted the Revenue Service's claim that the difference between male and female pass rates was not discriminatory because it merely reflected a difference in cognitive ability that was also evident in the data on college education. The Appeals Board concluded that the frequencies of passing were simply in line with the frequencies of college education even though the Revenue Service did not offer pass rate data by sex and education into evidence.

Table 1. Frequency of Passing

	Passed	Failed	Total	Percentage Passed
Female	68	183	251	27%
Male	68	47	115	59%
Total	136	230	366	37%

Was the Board's judgement in error given the evidence it chose to examine? Is the difference in passing rates for males and females consistent with the difference in college attendance rates? Would males and females who were the same in term of college attendance have similar passing rates? Or is the difference in passing rates too large to merely reflect a difference in college attendance?

The data are unusual. Tables 1 and 2 are two 2×2 margins of a $2 \times 2 \times 2$ table, but the Appeals Board did not request and never saw the full $2 \times 2 \times 2$ table. In legal proceedings, the parties, not the Appeals Board or jury, have the burden of producing evidence. Arguments about unobserved variables are often executed poorly. The argument the Board accepted about these tables is not merely implausible; it is impossible. The example is important because it shows that some claims about unobserved variables are not just strained or far fetched—some are just wrong.

If the $2 \times 2 \times 2$ table had been observed, the standard analysis would have compared passing rates for females and males, adjusting for education using the Mantel–

Table 2. Frequency of College Attendance

	Some College	No College	Total	Percentage College
Female	63	188	251	25%
Male	60	55	115	52%
Total	123	243	366	34%

Joseph Gastwirth is Professor, Department of Statistics, George Washington University, Washington, DC 20052. Abba Krieger and Paul Rosenbaum are Professors, Department of Statistics, Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6302. Joseph Gastwirth and Abba Krieger were supported in part by grants from the National Science Foundation.

Haenszel (1959) or MH statistic; see, for instance Bickel, Hammel, and O'Connell (1975), Fleiss (1981), Gastwirth (1988), or Agresti (1990). One can consider every possible $2 \times 2 \times 2$ table that can produce Tables 1 and 2—there are finitely many—and calculate the MH statistic for each of them. A simpler calculation producing the same result is discussed in Section 2. When this is done, the largest normal deviate that can be produced from $2 \times 2 \times 2$ tables compatible with Tables 1 and 2 is the deviate -3.11 with p value .002. Thus there is no $2 \times 2 \times 2$ table compatible with Tables 1 and 2 in which college attendance can explain the difference in passing rates for men and women. There may or may not be discrimination against women, but whether there is or is not, the explanation the Appeals Board accepted is simply wrong. The difference in passing rates is far larger than can be explained based on the difference in proportions attending college.

2. THE EXTREME MANTEL-HAENSZEL STATISTIC FROM A $2 \times 2 \times 2$ WITH GIVEN MARGINALS

Consider the $2 \times 2 \times 2$ table n_{egp} recording Education $(e)\times$ Gender $(g)\times$ Passing (p), where $e = 1$ for college and $e = 2$ otherwise, $g = 1$ for a female and $g = 2$ for a male, and $p = 1$ for passed and $p = 2$ failed. Write m for the number of passing individuals who attended college, that is, $m = n_{1 \cdot 1}$, so m cannot be determined from Tables 1 and 2. The $2 \times 2 \times 2$ table has one additional margin, in addition to Tables 1 and 2, namely, Table 3. The Mantel–Haenszel standardized deviate comparing passing rates for males and females within strata defined by college education is

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

Notice that m is the only quantity in (1) not determined from Tables 1 and 2. It is interesting to note that m determines the MH statistic and its large sample significance level, but m does not determine the exact significance level that requires the specification of additional quantities.

Although m is unknown, it is constrained by the observed data, $a \leq m \leq b$ and $a = \max(0, n_{\cdot 11} - n_{21\cdot}) + \max(0, n_{\cdot 21} - n_{22\cdot})$ and $b = \min(n_{\cdot 11}, n_{11\cdot}) + \min(n_{\cdot 21}, n_{12\cdot})$. A direct approach tries each m in this range, finding that $\max\{T(m)\} = -3.11$ with approximate significance level .002. In other words, even though m is not observed, there is no possible value of m such that an adjustment for college education would have explained the different passing rates for males and females.

Table 3. The Unobserved 2×2 Margin

	Passed	Failed	Total
College Educated	m	$n_{1\cdot\cdot} - m$	$n_{1\cdot\cdot} = 123$
Not College Educated	$n_{\cdot\cdot 1} - m$	$n_{\cdot\cdot\cdot} - n_{\cdot\cdot 1} - n_{1\cdot\cdot} + m$	$n_{2\cdot\cdot} = 243$
Total	$n_{\cdot\cdot 1} = 136$	$n_{\cdot\cdot\cdot} - n_{\cdot\cdot 1} = 230$	$n_{\cdot\cdot\cdot} = 366$

Table 4. The Table That Makes the Mantel–Haenszel Statistic as Close to Zero as Possible

	Passed	Failed	Total
College Educated	116	7	123
Not College Educated	20	223	243
Total	136	230	366

NOTE: Odds Ratio = 184.8.

Actually, it is not necessary to try each possible value of m. Viewing $T(m)$ as a differentiable function of a real argument m with derivative $T'(m)$, the following lemma can be proved.

Lemma. For $a < m < b$, the derivative $T'(m)$ is zero for at most one value of m.

Proof. Write $T(m) = N(m)/\sqrt{D(m)}$ in (1), so

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

Since $D(m) > 0$, the sign of $T'(m)$ in (2) is the same as the sign of $T'(m)\{D(m)\}^{3/2} = N'(m)D(m) - (1/2)N(m)D'(m)$. From (1), for certain constants d, e, f, g, and h that depend on the observed margins but not on $m, N(m) = d + em, D(m) = f + gm + hm^2$, so

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

which is linear in m, proving the lemma. □

The lemma implies that, for integer m between a and b, the deviate $T(m)$ is either monotone in m or else $T(m)$ is monotone from a to an integer r, with $a < r < b$, and then monotone in the opposite direction from r to b. In any event the minimum value of $T(m)$ for $a < m < b$ is $\min\{T(a), T(r), T(b)\}$ and the maximum value of $T(m)$ is $\max\{T(a), T(r), T(b)\}$.

Recall that m is the number of passing individuals who attended college, that is, $m = n_{1 \cdot 1}$; see Table 3. For the data in Section 1, $m = 116$ gives maximum Mantel–Haenszel deviate (1) of -3.11 with its associated maximum significance level of .002. Table 4 shows what Table 3 would be if $m = 116$. Notice that the odds ratio in Table 4 is 185, so a person with some college is 185 times more likely to pass the test than a person with no college, an extraordinarily strong relationship. If one were willing to assume that the odds ratio linking college with passing the test in Table 3 is somewhat smaller, then the maximum Mantel–Haenszel deviate and significance level would be smaller. For instance, if the odds ratio were no more than 10, the maximum Mantel–Haenszel deviate would be -3.97 with significance level less than .0001.

3. THE MORAL OF THE TALE

Some arguments involving unobserved variables are just wrong. Others are possible in principle but not plausible. Others are entirely plausible. Faced with an argument involving an unobserved variable, the statistician's responsibility is to clarify what that argument objectively entails and implies so listeners may appraise whether the argument is plausible.

Some general methods for appraising arguments about unobserved variables are described in Gastwirth (1992), Greenhouse (1982), Rosenbaum and Krieger (1990), and Rosenbaum (1991, 1993, in press).

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis*, New York: John Wiley.
Bickel, P., Hammel, E., and O'Connell, J.W. (1975), "Sex Bias in Graduate Admissions: Data From Berkeley," *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.
Fleiss, J. (1981), *Statistical Methods for Rates and Proportions*, New York: John Wiley.
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.
Greenhouse, S. W. (1982), "Jerome Cornfield's Contributions to Epidemiology," *Biometrics*, 28, supplement, 33–45.
Juriansz, R. G. (1987), "Recent Developments in Canadian Law: Anti-Discrimination Law Part II," *Ottawa Law Review* 19, 667–721.
Mantel, N., and Haenszel, W. (1959), "Statistical Aspects of Retrospective Studies of Diseases," *Journal of the National Cancer Institute*, 22, 719–748.
Rosenbaum, P., and Krieger, A. (1990), "Sensivity Analysis for Two-Sample Permutation Inferences in Observational Studies," *Journal of the American Statistical Association*, 85, 493–498.
Rosenbaum, P. R. (1991), "Discussing Hidden Bias in Observational Studies," *Annals of Internal Medicine*, 115, 910–905, Dec. 1, 1991.
——— (1993), "Hodges–Lehmann Point Estimates of Treatment Effect in Observational Studies," *Journal of the American Statistical Association*, 88, 1250–1253.
——— (in press), *Observational Studies*, New York: Springer-Verlag.

Simultaneous Confidence Intervals in Multiple Regression

Thomas P. LANE and William H. DUMOUCHEL

We describe a hybrid method for computing confidence intervals for linear combinations of coefficients in multiple regression, with an emphasis on intervals for fitted values. The hybrid method combines the Bonferroni and Scheffé approaches, and it is applicable when there are both continuous and discrete predictors. It often leads to intervals that are narrower than those produced by the Scheffé method. We also describe how the method can create simultaneous prediction intervals for new observations.

KEY WORDS: Bonferroni; Multiple comparisons; Scheffé.

1. INTRODUCTION

Consider the general linear regression model

$$y_i = \mathbf{x}_i^T \beta + \varepsilon_i, \tag{1}$$

where y_i is an observed response value for the ith observation, $\mathbf{x}_i$ is a p-dimensional vector of known predictor values, β is a p-dimensional vector of unknown coefficients, and ε_i is a normal random error with mean 0 and unknown variance σ^2. A statistician observes n independent response values $y_1, \ldots, y_n$, and the corresponding predictor vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, and wants to compute confidence bounds for the unknown mean $\mathbf{x}^T\beta$ so the confidence bounds hold simultaneously over all values of $\mathbf{x}$. These bounds form a simultaneous confidence region for the entire regression surface.

The vector $\mathbf{x}$ will typically contain a 1 for the intercept and may contain actual values of continuous predictors, dummy variable values for categorical predictors, and products and powers of these values. Consequently, this general regression model includes ordinary multiple regression, polynomial regression, and analysis of variance as special cases.

There are many well-known methods for computing simultaneous confidence intervals for fitted values and for other linear combinations of regression coefficients. Two commonly used procedures are the Scheffé method and the Bonferroni method. These methods are described in many books on regression such as Seber (1977), and also in books on simultaneous inference such as Miller (1981) and Hochberg and Tamhane (1987).

The Scheffé method provides simultaneous confidence over all possible linear combinations of coefficients. It can be used in any regression problem, but it is most useful when the predictors are continuous, as in polynomial regression. The Bonferroni method provides simultaneous confidence over a finite set of linear combinations. It cannot be used with continuous predictors without restricting the number of possible predictor values to a finite set. When predictors are categorical, as they are in analysis of variance, confidence intervals produced by the Bonferroni method are often narrower than those produced by the Scheffé method.

We describe a hybrid method that is useful when there are both continuous and categorical predictors. Our method includes both the Bonferroni and Scheffé methods as special cases.

In Section 2 we review the Bonferroni and Scheffé methods and describe the hybrid method. In Section 3 we examine some specific regression models and compare the relative width of intervals computed by different methods. In Section 4 we show how the hybrid method can be extended as in Carlstein (1986) to compute simultaneous prediction intervals for new observations.

Thomas P. Lane is Senior Statistician, BBN Software Products, Cambridge, MA 02140. William H. DuMouchel is Professor, Division of Biostatistics, Columbia University, New York, NY 10032.