USING THE CORRECT STATISTICAL TEST FOR THE EQUALITY OF REGRESSION COEFFICIENTS

RAYMOND PATERNOSTER
ROBERT BRAME
  University of Maryland
  National Consortium on Violence Research

PAUL MAZEROLLE
  University of Cincinnati

ALEX PIQUERO
  Temple University
  National Consortium on Violence Research

  Criminologists are often interested in examining interactive effects within a regression context. For example, "holding other relevant factors constant, is the effect of delinquent peers on one's own delinquent conduct the same for males and females?" or "is the effect of a given treatment program comparable between first-time and repeat offenders?" A frequent strategy in examining such interactive effects is to test for the difference between two regression coefficients across independent samples. That is, does $b_1 = b_2$? Traditionally, criminologists have employed a t or z test for the difference between slopes in making these coefficient comparisons. While there is considerable consensus as to the appropriateness of this strategy, there has been some confusion in the criminological literature as to the correct estimator of the standard error of the difference, the standard deviation of the sampling distribution of coefficient differences, in the t or z formula. Criminologists have employed two different estimators of this standard deviation in their empirical work. In this note, we point out that one of these estimators is correct while the other is incorrect. The incorrect estimator biases one's hypothesis test in favor of rejecting the null hypothesis that $b_1 = b_2$. Unfortunately, the use of this incorrect estimator of the standard error of the difference has been fairly widespread in criminology. We provide the formula for the correct statistical test and illustrate with two examples from the literature how the biased estimator can lead to incorrect conclusions.

A frequent task in criminological research is to determine whether an empirical relationship or causal effect that is estimated within two independent samples is equivalent. This question involves the possible

existence of an interactive effect. Illustrations of this task occur in both theoretical and applied work. For example, criminological theorists frequently, either explicitly or implicitly, limit the scope of their theory to some defined groups, such as males, or members of a particular social class (Hagan et al., 1987; Jarjoura, 1996; Smith and Paternoster, 1987). They may also allude to possible conditions when the described causal process may operate with enhanced vitality or when it may be suppressed. Researchers more interested in applied problems may want to know if the effect of a correctional treatment program is the same for different offenders, whether the same decision-making process is at work for different persons or within different organizational contexts (Albonetti, 1990; Erez, 1989; Visher, 1983), or whether the effect of a given independent variable is the same before and after some historically relevant point in time (Hagan and Palloni, 1986; Jang and Krohn, 1995; Spohn and Horney, 1993). In both theoretical and applied contexts, therefore, researchers frequently want to know if the effect of a given explanatory factor is invariant over persons, time, or organizations.

When the issue of the invariance of explanatory variables has been investigated within a regression framework, one commonly employed strategy has been to determine the significance of the difference between two regression coefficients estimated within two independent samples. For example, if b_1 reflects the effect of explanatory variable x within group 1 (say, males) and b_2 is the effect of that same variable within group 2 (females), a test of explanatory invariance has consisted of a formal hypothesis test that the difference between b_1 and b_2 is zero. Traditionally, this hypothesis test has followed the structure of a significance test between two sample means. The relevant test statistic has been a t or z test. The numerator of this test is the estimated difference between the two coefficients in the population $(b_1 - b_2)$, and the denominator is the estimated standard error of the difference. The standard error of the difference is the estimated standard deviation of the sampling distribution of coefficient differences $(\hat{\sigma}_{b1-b2})$.

While there has been considerable consensus in the criminological literature with respect to the appropriateness of this coefficient-comparison strategy in examining what is essentially an interactive effect, there has been some confusion as to the appropriate formula to employ. In reviewing the criminological literature, we have identified two similar, but distinct, formulas that have been used in testing the difference between two regression coefficients.[1] While the numerators for these two formulas are

---

1. The points we make in this note about the comparison of regression coefficients are applicable to all regression-type problems that yield maximum likelihood

identical (the difference between the sample coefficients, $b_1 - b_2$), the estimated standard error of the difference is not the same. Based upon extensive simulation and other evidence that we report in detail elsewhere (Brame et al., 1998), we have come to the conclusion that one formula is correct and the other is incorrect. The incorrect formula provides a negatively or downwardly biased estimate of the true standard deviation of the sampling distribution of coefficient differences. As a result, the probability of rejecting a false null hypothesis is greater than one's reported alpha value. Consequently, one has a greater than alpha probability of rejecting the null hypothesis that $b_1 - b_2 = 0$ when in fact it is true.

Unfortunately, the use of what we have come to believe is the incorrect formula has become common place in criminology. We have systematically examined articles published over the past 20 years in five of the most influential journals used as an outlet for the work of criminologists. In this review, we have identified at least 16 published papers in which we could confidently conclude that this incorrect formula was used (Brame et al., 1998).[2] As a result of the use of this incorrect formula, we believe researchers were led to and reported some incorrect conclusions. Our purpose in writing this note is to provide criminologists with what we think is the correct formula to apply when one is interested in testing the hypothesis about the comparability of two regression coefficients. Based upon published studies, we also provide a brief illustration of the difference that the use of the correct formula can make. Readers interested in the detailed simulation results and other evidence that we drew upon in drawing our conclusions are invited to examine Brame et al. (1998).

THE COMPETING FORMULAS

As discussed above, a frequently applied hypothesis test in criminological research for the difference between two regression coefficients is the z test with general form:

---

estimates (OLS, probit, tobit, Poisson, negative binomial, and parametric failure-time regression models).

  2. These studies are reported in Brame et al. (1998). We should note that our estimate of 16 published papers is in all likelihood a very conservative estimate of the use of the incorrect formula in criminological research. In many studies that we reviewed, because of insufficient detail provided, we were unable to determine which formula was employed. We also limited our search to five major journals to which criminologists submit their papers. We think that the estimate of 16 is, therefore, a lower bound. We would also like to note that in identifying these studies, our purpose is not to be critical, but simply to document the widespread use of the incorrect formula. We greatly appreciated the detail provided by the 16 sets of authors. Their thoroughness enabled us to replicate their reported results. This was not always the case.

$$z = \frac{b_1 - b_2}{\hat{\sigma}_{b1-b2}} \tag{1}$$

where, $\hat{\sigma}_{b1-b2}$ equals the estimated standard error of the difference. In many applications, the following denominator has been used:

$$\hat{\sigma}_{b1-b2} = \sqrt{\frac{V_1(SEb_2^2) + V_2(SEb_2^2)}{V_1 + V_2}} \tag{2}$$

where, V_1 and V_2 are the degrees of freedom and SEb_1^2 and SEb_2^2 are the coefficient variances associated with the first and second groups respectively. The z test for the difference between two regression coefficients is, then:

$$Z = \frac{b_1 - b_2}{\sqrt{\dfrac{V_1(SEb_1^2) + V_2(SEb_2^2)}{V_1 + V_2}}} \tag{3}$$

In our simulation work, we have found this to be an incorrect formula for the difference between two regression coefficients, because the estimated standard error of the difference is negatively biased. Using this formula, the probability of rejecting the null hypothesis that $b_1 = b_2$ is greater than one's reported alpha level. One would, therefore, mistakenly conclude that there are group differences in the estimated structural coefficient when in fact there is no difference. Moreover, we have found that this bias is likely to be particularly pronounced when the two groups have very unequal sample sizes.

Drawing on the work of Clogg et al. (1995), the correct formula for this statistical test should be:

$$Z = \frac{b_1 - b_2}{\sqrt{SEb_1^2 + SEb_2^2}} \tag{4}$$

As we demonstrate in some detail elsewhere (Brame et al., 1998), the estimate of the standard deviation of the sampling distribution in this formula is unbiased. We would, therefore, recommend that researchers abandon Equation 3 and use Equation 4 whenever their interest is in the difference between two regression coefficients.[3]

---

3. Our recommendation of Equation 4 applies to large sample studies. Equation 2 will continue to provide biased estimates of the standard deviation of the sampling

APPLICATIONS

We would like to provide two brief illustrations of the implications of our findings for previously published criminological research. In a paper investigating whether the same causal factors are at work for male and female delinquency, Smith and Paternoster (1987:151) concluded that with few exceptions the factors that accounted for male participation in and frequency of delinquent behavior were similar to those for females. One of the exceptions was the effect of what they called the "behavioral dimension" of differential association theory—the proportion of one's friends who report committing delinquent acts. They found that while this behavioral dimension had a significant positive effect on both male and female participation in marijuana use, the effect for males ($b = .404$, s.e. $= .094$) was significantly greater than that for females ($b = .221$, s.e. $= .091$). Using Equation 3, Smith and Paternoster rejected the null hypothesis that $b_{males} = b_{females}$ ($t = 1.98$, $p < .05$). Using Equation 4, on the same coefficients and standard errors, we find that the difference is not statistically significant:

$$z = \frac{.183}{\sqrt{(.094)^2 + (.091)^2}}.$$

$$z = 1.40.$$

On the basis of the correct statistical test, then, one would conclude that the effect of delinquent peer association on participation in delinquency is similar for males and females.

Another reported exception to the general pattern of causal invariance was the different effect of moral beliefs. Smith and Paternoster reported that the probit regression coefficient for the relationship between moral beliefs and participation in marijuana use was $-.116$ (s.e. $= .072$) within their sample of males, and $-.269$ for females (s.e. $= .071$). Using Equation 3, they reported that the obtained t statistic for the difference between these coefficients was 2.14. Accordingly, they rejected the null hypothesis of equal coefficients and concluded that "while moral beliefs decrease the propensity to use marijuana for both males and females, this effect is significantly more pronounced among females" (1987:153). When we use the correct formula (Equation 4) to conduct the same hypothesis test, we find that:

---

distribution in small sample studies. Small sample tests can be found in Cohen (1983) and Kleinbaum and Kupper (1978).

$$z = \frac{.153}{\sqrt{(.072)^2 + (.071)^2}}$$

$$z = 1.51.$$

Contrary to the original conclusion, one would now fail to reject the null hypothesis that $b_1 = b_2$, and one could not conclude that moral beliefs affect males differently than females. In this example from Smith and Paternoster, two instances in which an explanatory factor that was significant for one gender group but not for the other are found to be not significantly different when the correct statistical formula is used. The causal factors for males and females seem to be even more similar than suggested by the original study.

Our second illustration comes from Hagan (1991), who hypothesized that the effects of subcultural "drift" may vary by an adolescent's gender and social class of origin. To test this hypothesis, Hagan conducted a series of t tests, using Equation 3, of slope coefficient differences among gender and social class groups. Each slope coefficient reflected the effect of subcultural preferences on adult status attainment. The purpose of the slope coefficient difference tests was to determine if this effect was the same for groups classified by their gender and social class. His table reporting the difference in OLS regression coefficients and the associated t value for each comparison is shown in Table 1. In five of the six comparisons, the t test of coefficient differences was statistically significant. These

Table 1.   Gender-Specific and Class-Specific Direct Effects of Subcultural Preferences on Adult Status Attainment (Hagan, 1991:Table 6)

	Sons of Working-Class Fathers Compared to		
	Sons of Non-Working Class Fathers	Daughters of Working-Class Fathers	Daughters of Non-Working-Class Fathers
Subculture of Delinquency			
Difference in bs	1.755	1.774	1.348
t Value of Difference	1.844*	2.002*	1.097
Corrected z Value	1.297	1.223	.763
	Sons of Non-Working-Class Fathers Compared to		
	Sons of Working-Class Fathers	Daughters of Working-Class Fathers	Daughters of Non-Working-Class Fathers
Party Subculture			
Difference in bs	2.531	1.891	1.878
t Value of Difference	2.744**	1.914*	1.700*
Corrected z Value	1.930*	1.307	1.174

* $p < .05$ (one-tailed).
** $p < .01$ (one-tailed).

results led Hagan (1991:578) to conclude that his hypotheses were supported and that "the effects of subcultural drift are contingent on gender and class of origin." Just under Hagan's reported t value from Equation 3 in Table 1, we report the corrected z statistic obtained using Equation 4. The results change dramatically; now, only one of the six coefficient differences is statistically significant. Contrary to the original conclusion of Hagan, one would be led to conclude that the effect of adolescent subcultural preferences on adult status attainment does not significantly vary by gender or social class of origin. We suspect that there may be other instances in published studies in which the correct statistical test will make a difference.

CONCLUSIONS

Researchers in criminology who are interested in testing the significance of the difference between two regression coefficients have commonly applied a statistical test that we have found leads to incorrect conclusions. Our simulation and other evidence lead us to believe that the estimated standard deviation of the sampling distribution $(\hat{\sigma}_{b1-b2})$ of this statistical test is negatively biased. As a result, the true probability of incorrectly rejecting a true null hypothesis is greater than the reported alpha level. We have generally found this bias to be nontrivial in magnitude, although it is much more so when the sample sizes of the groups are substantially unequal. To make no mistake, we strongly recommend that researchers abandon the use of Equation 3 and use Equation 4 whenever they are interested in testing the null hypothesis that two regression coefficients are equal.

REFERENCES

Albonetti, Celesta
    1990   Race and the probability of pleading guilty. Journal of Quantitative Criminology 6:315–334.

Brame, Robert, Raymond Paternoster, Paul Mazerolle, and Alex Piquero
    1998   Testing for the equality of maximum likelihood regression coefficients between two independent equations. Journal of Quantitative Criminology 14:245–261.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou
    1995   Statistical methods for comparing regression coefficients between models. American Journal of Sociology 100:1261–1293.

Cohen, Ayala
    1983   Comparing regression coefficients across subsamples: A study of the statistical test. Sociological Methods and Research 12:77–94.

Erez, Edna
    1989   Gender, rehabilitation, and probation decisions. Criminology 27:307–327.

Hagan, John
  1991   Destiny and drift: Subcultural preferences, status attainments, and the risks
          and rewards of youth. American Sociological Review 56:567–582.

Hagan, John and Alberto Palloni
  1986   "Club fed" and the sentencing of white-collar offenders before and after
          Watergate. Criminology 24:603–621.

Hagan, John, John Simpson, and A.R. Gillis
  1987   Class in the household: A power-control theory of gender and delinquency.
          American Journal of Sociology 92:788–816.

Jang, Sung Joon and Marvin D. Krohn
  1995   Developmental patterns of sex differences in delinquency among African-
          American adolescents: A test of the sex-invariance hypothesis. Journal of
          Quantitative Criminology 11:195–222.

Jarjoura, G. Roger
  1996   The conditional effect of social class on the dropout-delinquency relation-
          ship. Journal of Research in Crime and Delinquency 33:232–255.

Kleinbaum, David G. and Lawrence L. Kupper
  1978   Applied Regression Analysis and Other Multivariate Methods. North
          Scituate, Mass.: Duxbury Press.

Smith, Douglas A. and Raymond Paternoster
  1987   The gender gap in theories of deviance: Issues and evidence. Journal of
          Research in Crime and Delinquency 24:140–172.

Spohn, Cassia and Julie Horney
  1993   Rape law reform and the effect of victim characteristics on case processing.
          Journal of Quantitative Criminology 9:383–409.

Visher, Christy A.
  1983   Gender, police arrest decisions, and notions of chivalry. Criminology
          21:5–28.

Raymond Paternoster is Professor of Criminology at the University of Maryland and Fellow with the National Consortium on Violence Research. His research interests include the testing of criminological theory and quantitative methods in criminology.

Robert Brame is Assistant Professor of Criminology at the University of Maryland and Post-Doctoral Research Fellow with the National Consortium on Violence Research. His research interests include the testing of criminological theory and applied statistics.

Paul Mazerolle is Assistant Professor in the Division of Criminal Justice at the University of Cincinnati. His research examines theoretical and empirical dimensions of crime and delinquency, including issues related to criminal careers and crime over the life course.

Alex Piquero is Assistant Professor of Criminal Justice, and Research Associate in the Center for Public Policy at Temple University. His research interests include criminological theory, criminal careers, quantitative research methods, and policing.