NATIONAL INSTITUTE of STATISTICAL SCIENCES
TECHNICAL TASK FORCE REPORT

COMPARING REGRESSION COEFFICIENTS BETWEEN MODELS: CONCEPTS AND ILLUSTRATIVE EXAMPLES

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

COMPARING REGRESSION COEFFICIENTS BETWEEN MODELS

EXECUTIVE SUMMARY

One of the most common statistical procedures in quantitative social science research is to examine the association between a key predictor, X, and an outcome, Y, before and after adjusting for another predictor, Z. If the absolute value of that coefficient is reduced after adding Z, they infer that Z explains, at least in part, the relationship between X and Y. The general issue at hand, then, is "comparing regression coefficients between models." The inferential issues involved in such comparisons have arisen frequently in data analyses contracted by the National Center for Educational Statistics (NCES). Concerned about the possible subjectivity associated with comparisons using the "eyeball" method, NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of technical experts to consult with NCES on advice for contractors analyzing NCES data.

The panel met in-person at NISS in October, 1996, to consider these issues, and a sub-group of participants volunteered to serve on a Task Force to write this report. The report is restricted to the case of a continuous or approximately continuous outcome as a first step in establishing standards. The procedures recommended apply when there is no statistical interaction between X and Z. Examples are used to illustrate how to test for such interactions and how to compare coefficients across models when no such interactions are found.

NATIONAL INSTITUTE OF STATISTICAL SCIENCES TECHNICAL TASK FORCE REPORT

PREFACE

The inferential issues involved in "comparing regression coefficients between models" have arisen frequently in data analyses contracted by the National Center for Educational Statistics (NCES). Concerned about the possible subjectivity associated with comparisons using the "eyeball" method, NCES charged the National Institute of Statistical Sciences (NISS) with convening a panel of technical experts to consult with NCES on advice for contractors analyzing NCES data.

The panel met in-person at NISS in October, 1996, to consider these issues, and a sub-group of panel participants volunteered to serve on a Task Force to write this report.

COMPARING REGRESSION COEFFICIENTS BETWEEN MODELS: CONCEPTS AND ILLUSTRATIVE EXAMPLES[*]

Statement of the Problem

One of the most common statistical procedures in quantitative social science research is to examine the association between a key predictor, X, and an outcome, Y, before and after adjusting for another predictor, Z. In some cases X represents a quasi-experimental treatment, Z is a pre-treatment covariate, and the aim is to estimate the extent to which the covariate accounts for the unadjusted difference between treatments. In other cases, X is an exogenous variable known to be associated with the Y and the question is whether this association is "explained" by an endogenous Z viewed as mediating the X, Y relationship. In practice, most investigators simply "eyeball" the regression coefficient for X before and after adding Z. If the absolute value of that coefficient is reduced after adding Z, they infer that Z explains, at least in part, the relationship between X and Y. Clogg, Petkova, and Haritou (1995) and Allison (1995) advised researchers to take a more rigorous approach to this problem by computing standard errors, tests, and confidence intervals for the difference in the X, Y coefficient associated with adding Z. The general issue at hand, then, is "comparing regression coefficients between models."

Not surprisingly, the inferential issues involved in such comparisons have arisen frequently in data analyses contracted by the National Center for Educational Statistics (NCES). Concerned about the possible subjectivity associated with comparisons using the "eyeball" method, Susan Ahmed, Chief Statistician at NCES, asked the National Institute of Statistical Sciences (NISS) to consult with NCES on advice for contractors analyzing NCES data. Under the leadership of Ingram Olkin, a meeting was convened at NISS in October, 1996, to consider these issues, and a sub-group of participants volunteered to serve on a Task Force to write this report. The report is restricted to the case of a continuous or approximately continuous outcome as a first step in establishing standards. The procedures we recommend apply when there is no statistical interaction between X and Z. We illustrate by example how to test for such interactions and how to compare coefficients across models when no such interactions are found.

A Simple Example

In general, there can be one or more X variables whose relationship to Y might be "explained" by one or more Zs. However, the basic logic of the problem becomes clear in the simple case in which there is one continuous outcome variable, Y, a binary predictor X (taking on a value of 1 if a case is in group 1 and 0 if not), and a single covariate, Z. One might consider X to be either fixed or random, but we assume both Z and Y are random. Two regressions are estimated. First, we estimate a regression with Y as the outcome and X as the sole predictor:

$$Y_i = \alpha + \beta_{yx}X_i + e_i. \tag{1}$$

Next, we estimate a regression with Y as the outcome, but now we include both X and Z as predictors:

$$Y_i = \alpha^* + \beta_{yx.z}X_i + \gamma Z_i + e_i \qquad \textbf{(2)}$$

Both regressions will often be estimated under ordinary least squares (OLS), often with provisions to ensure that the standard errors take into account important features of the sampling design and possible heteroscedasticity in the errors.

The assumption of homogeneity of regression. A key assumption for the procedures described so far is that the regression coefficient for Z as a predictor of Y is identical within each of the two groups defined by the two values of X. The present report is organized around three issues: 1) how to assess the assumption of equality of slopes; 2) how to proceed when the assumption is reasonable; and 3) how to proceed when it is not.

1. Checking the assumption of equal slopes. Standard textbooks provide a test of homogeneity of regression. One creates an interaction term, the product of X and Z, adds that as a predictor, and tests the significance of the slope associated with that interaction. Of course, in large samples there will be a tendency to reject the hypothesis of homogeneity even when the difference between· the slopes is small. Thus, it will make sense to compute a confidence interval for the difference between the two slopes (which is the regression coefficient associated with the interaction term) and to assess the size of the difference between the slopes.

2. The case of parallel lines. When the data suggest that the two slopes are approximately equal, the problem at hand is to obtain a test and, more importantly, a confidence interval for the difference between the coefficients for X in the two models, that is, for

$$\delta = \beta_{yx} - \beta_{yx.z} \qquad \textbf{(3)}$$

The OLS estimators in this simple example are

$$\beta_{yx} = \bar{Y}_1 - \bar{Y}_0$$

$$\beta_{yx.z} = \bar{Y}_1 - \bar{Y}_0 - \hat{y}\,(\bar{Z}_1 - \bar{Z}_0) \qquad \textbf{(4)}$$

$$\delta = \hat{y}\,(\bar{Z}_1 - \bar{Z}_0)$$

These quantities are illustrated in Figure 1, a familiar display for those who have studied the analysis of covariance. The figure shows the regression lines for predicting Y from Z within groups 0 and 1. The estimated mean difference between groups, controlling Z, is the distance between the two regression lines. Note that there is a positive relationship between Z and Y and that group 1's mean covariate value, $\bar{Z}_1$, is larger than group 0's mean covariate value, $\bar{Z}_0$. Therefore, the adjusted mean difference, $\beta_{yx.z} = \bar{Y}_1 - \bar{Y}_0 - \hat{y}$ $(\bar{Z}_1 - \bar{Z}_0)$ is smaller than the unadjusted mean difference, $\beta_{yx} = \bar{Y}_1 - \bar{Y}_0$. The magnitude of the adjustment is $\delta = \hat{y}\,(\bar{Z}_1 - \bar{Z}_0)$. If the figure were redrawn with nonparallel lines, the adjusted mean difference would vary as a function of Z as would the magnitude of the adjustment. Statistical inference would be correspondingly more complex.

Insert Figure 1 About Here

The population quantity of interest here may be conceived in different ways. For most NCES purposes, our view is that this quantity is

$$\delta = \gamma \ (\mu_{z_1} - \mu_{z_0}).$$ (5)

Here μ_{z_1}, μ_{z_0} are the population means of Z for groups 1 and 0, respectively. In some cases these means may be known from the census, in which case inference will be quite simple. More generally, however, we recommend estimating the sampling variance of δ using Equation 7 of Allison (1995), with "design-consistent" sampling variances substituted for the usual OLS variances in that formula when necessary.[1] This sampling variance is the basis for an asymptotic unit normal test and confidence interval for δ. We shall illustrate this approach and its multivariate generalization below.

3. <u>Non-parallel lines within groups</u>. When the slopes are unequal, inference is conceptually more difficult. If the data suggest that the slopes are negligibly different, then the researcher may wish to use the procedure described above for parallel slopes. For example, one might assess the mean difference between the two groups when evaluated at selected points in the distribution of Z within the two groups. If this difference were negligible, one might proceed cautiously as described above.

When the difference between slopes is non-negligible, we suggest consulting the literature stemming from the Johnson-Neyman technique {Johnson and Neyman, 1936; see also Walker and Lev·, 1953; Rogosa, 1980) and as well as work by Oaxaca and others (Oaxaca and Ransom, 1994; Beaton, 1988). Essentially, these approaches involve evaluating the difference between groups at selected values of Z; and/or decomposing inequality between groups as a function of inequality between covariate means and inequality between group slopes.

In the case of non-parallel slopes, we do not recommend pursuing inferences about changes in the coefficient for X associated with the inclusion of Z as a covariate.

The Illustrative Data

To illustrate the procedures we recommend, we seek to compare all Catholic schools and a random sample of public schools from the base year (1980) sample of the High School and Beyond survey. The variables are all measured at the school level.

<u>The outcome</u>, Y, is "DISCLIM," the disciplinary climate of the school, a standardized composite of items indicating frequency of fighting, class cutting, tardiness, suspensions, etc. (see Lee and Bryk, 1989, for details). High values indicate a disorderly climate.

<u>The predictors</u> are "SECTOR," "MEANSES," and "PROP ACADEMIC:"

SECTOR is coded 1 for Catholic schools, 0 for public schools.

MEANSES is the mean socioeconomic status of sampled students (centered about its grand mean). The typical sample size is 45-60 per school, and the students constitute a simple random sample from the school, so the sample mean is a reasonably good estimate of the mean SES of all students in the school. Student SES is a composite of income, parent education, and parent occupation.

PROP ACADEMIC is the is the proportion of kids in the academic track in the given school.

---

[1] By "design-consistent," we mean that the variances are estimated via procedures that take into account the stratified and clustered character of the design producing the data as well as the measurement properties of Y.

Table 1 provides a description of the sample.

Table 1: Description of the Analytic Sample

Variable Name	Sample Size	Mean	Standard Deviation
DISCLIM	160	-.015	.977
SECTOR	160	.438	.498
MEANSES	160	.000	.414
PROP ACADEMIC	160	.514	.256

Example 1: The Case of a Single X and a Single Z

We begin with the example already discussed. We have a single outcome, Y = DISCLIM, a single binary predictor, X = SECTOR, and a single Z = MEANSES. The substantive question is whether the manifest mean difference in DISCLIM between public and Catholic schools reflects, at least to some degree, the relatively advantaged MEANSES of the Catholic schools. Thus, we might estimate a model in which SECTOR predicts DISCLIM and a second model in which SECTOR and MEANSES predict DISCLIM. We then might ask whether the SECTOR coefficient is significantly reduced after adding MEANSES. However, we must first examine the implicit assumption that SECTOR and MEANSES do not interact.[2]

1. Is there an interaction between X and Z? We begin by estimating the regression

$$Y_i = \alpha^+ + \beta^+ (SECTOR)_i + \gamma^+ (MEANSES)_i$$
$$+ \omega (SECTOR) * (MEANSES)_i + e_i^+$$

(6)

leading to the estimated coefficients, standard errors, and t-ratios reported in Table 2.

Table 2: Testing the MEANSES-by-SECTOR Interaction

Predictor	Coef	SE	t-ratio	p
Constant	0.573	0.076	7.52	.00
SECTOR	-1.317	0.117	-11.26	.00
MEANSES	-0.184	0.190	-0.97	.33
SECTOR*MEANSES	-0.166	0.282	-0.59	.56

Given t = -.59, p = .56 associated with the interaction between sector and MEANSES, we conclude that there is no evidence against the parallelism assumption.[3]

2 Estimating the model without Z. We now estimate the model

$$Y_i = \alpha + \beta_{yx} (SECTOR)_i + e_i$$

(7)

leading to the result reported in Table 3.

---

[2] Scatter plots and residual analyses for each group indicated the approximate linearity of the estimated relationships and the relatively "well-behaved'' character of residuals.

[3] We caution the reader that it will often be important to center one or both of the predictors and to construct the interaction term by multiplying these centered predictors. Use of the centered predictors and the resulting product term will minimize collinearity among the predictors.

Table 3: Model Estimates for X = SECTOR, without Z

Predictor	Coef	SE	t-ratio	p
Constant	0.597	0.073	8.23	.00
SECTOR	-1.399	0.110	-12.76	.00

Note that in the model with SECTOR as a dummy variable the regression coefficient for SECTOR is simply the mean difference between sectors on disclim. Thus, we see that Catholic schools have more orderly disciplinary climates than do public schools, β = -1.399, t = -12.76. The question is whether the high MEANSES of the Catholic schools can at least partially account for this difference. We therefore estimate the model

$$Y_i = \alpha^* + \beta_{yx.z} \, (SECTOR)_i + \gamma \, (MEANSES)_i + e_i^* \tag{8}$$

leading to the results reported in Table 4.

Table 4: Model Estimates for X = SECTOR, including Z = MEANSES

Predictor	Coef	SE	t-ration	p
Constant	0.563	0.074	7.59	.00
SECTOR	-1.322	0.116	-11.36	.00
MEANSES	-0.260	0.140	-1.86	.07

Clearly the "SECTOR coefficient" remains significantly less than zero. But is it now smaller than before controlling MEANSES? The estimated mean difference between coefficients for SECTOR across the two models is -1.399 - (-1.322) = -.077, but such a seemingly small difference may be attributable to chance. Moreover, a small estimate does not by itself rule out a large difference between the population coefficients. We need a confidence interval for this difference, and for that we need a standard error estimate. To compute the standard error estimate, we must first fit the regression of Z on X. Thus we estimate

$$(meanses)_i = \theta + \beta_{zx} \, (sector)_i + u_i \tag{9}$$

leading to the model estimates reported in Table 5.

Table 5: Estimated Regression of Z = MEANSES on X = SECTOR

Predictor	Coef	SE	t-ratio	p
Constant	-0.130	0.041	-3.17	0.002
SECTOR	0.296	0.062	4.78	0.000

As mentioned, these results suggest strongly that Catholic schools have higher MEANSES than do public schools on average. We now have all of the quantities needed to estimate the sampling variance of the difference between coefficients across models.

For an estimator θ of θ, let $s(\theta)$ denote the estimated standard error of θ. Then, applying Equation 7 of Allison (1995), we have

$$d = \delta = \hat{\beta}_{yx} - \hat{\beta}_{yx.z} = -1.399 - (-1.322) = -.077$$

$$s^2(d) = s^2(\hat{\gamma})\,\hat{\beta}_{zx}^2 + \hat{\gamma}^2\,s^2(\hat{\beta}_{zx}) \tag{10}$$

$$= (.1399)^2 * (0.296)^2 + (-.260)^2 * (0.619)^2 = .00197.$$

Using the standard error of $(.00197)^{\frac{1}{2}} = .0444$, we can compute a confidence interval for δ, for example

$$95\%\ CI(\delta) = [-.077 - (1.96) * (.044),\, -.077 + (1.96) * (.044)] \tag{11}$$

$$= (-.163, .009).$$

All plausible values of the difference δ are small relative to the initial estimate of the SECTOR coefficient of -1.399, although the null hypothesis of no difference would be rejected by a one-tail test at the five percent level of significance, t = -.077/.044 = -1.75.

Example 2: The Case of One X and Two Zs

The example above is strictly univariate: there is a single X and a single Z. We now consider the first of two multivariate cases. Our interest remains focused on the mean difference between Catholic and public schools in DISCLIM. However, we now ask whether this relationship can be somehow "explained" by two variables Z_1 = MEANSES and Z_2 = PROP ACADEMIC, that is, the proportion of students in the school in the academic track. It may be that Catholic schools have more favorable DISCLIM not only because they have higher MEANSES but also because more of their students are assigned to the academic track, based on the assumption that DISCLIM is generally more favorable within the academic track than the non-academic track. The logic, then, is to compare a model with only SECTOR as a predictor to a model in which SECTOR, MEANSES, and PROP ACADEMIC are predictors. However, such a model comparison assumes no interactions involving SECTOR. We therefore consider the interactions first.

1. Are any interactions with SECTOR needed in the model? To answer this question, we estimate a model with all two-way interactions and the three-way interaction (Table 6) and a second model without the interactions (Table 7).[4]

Table 6: Estimated Regression with Interactions

Predictor	Coef	SE	t-ratio	p
Constant	0.036	0.078	0.46	.65
X = SECTOR	-1.036	0.156	-6.66	.00
Z_1 = MEANSES	0.012	0.172	0.07	.95
Z_2 = P. ACAD.	-0.674	0.465	-1.45	.15
X by Z_1	0.162	0.424	0.38	.70
X by Z_2	-0.342	0.717	-0.48	.63
Z_1 by Z_1	-0.274	0.743	-0.37	.71
X by Z_1 by Z_2	-1.665	2.197	-0.76	.45

$R^2 = 54.4\%$

---

[4] Our primary concern is whether interactions exist involving X = SECTOR. Specifically, we are concerned about the two-way interactions (between MEANSES and SECTOR and between PROP ACADEMIC and SECTOR) and the three-way interaction. We also include the two-way interaction between MEANSES and PROP ACADEMIC to ensure that the interactions of interest to us are estimated without bias.

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	7	82.486	11.784	25.85	.00
Error	152	69.277	0.456		

Table 7: Estimated Regression Without Interactions

Predictor	Coef	SE	t-ratio	P
Constant	0.940	0.157	5.99	.00
SECTOR	-1.077	0.146	-7.40	.00
MEANSES	0.014	0.170	0.08	.94
P. ACAD.	-0.942	0.348	-2.71	.01

$R^2 = 54.0\%$

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	3	81.927	27.309	61.00	.00
Error	156	69.837	0.448		
Total	159	151.763			

The univariate t-tests for the three interactions involving X = SECTOR in Table 6 are non-significant. However, a more appropriate test is the omnibus F test, where

$$f = \frac{SS_{residuals|no\ interaction} - SS_{residuals|interaction}}{SS_{residuals|no\ interactions}}$$
$$* \frac{denominator\ df|interactions}{number\ of\ interactions} \tag{12}$$
$$= \frac{69.837 - 69.277}{69.837} * \frac{152}{4} = .30$$

Here f is compared to the percentiles of the F distribution with df = 4, 153, showing no evidence against the null hypothesis that all interactions are null. Thus, we proceed to the model comparison test. Note that if the omnibus null hypothesis on the interactions were rejected, we would examine interactions involving X using post hoc procedures.[5]

2. <u>How do the coefficients for SECTOR differ, before and after adding MEANSES and PROP ACADEMIC?</u> A comparison of results between Table 3 (which has SECTOR predicting DISCLIM) and Table 7 (where SECTOR, MEANSES, and PROP ACADEMIC predict DISCLIM) shows that the difference between the two coefficient estimates for SECTOR is d = -1.399 - (-1.077) = - .322. To compute a confidence interval or test, we need to compute the estimated variance of d. To compute this, it is convenient to write the two models in matrix notation.

---

[5] Alternatively, one might test the three-way interaction first. Finding it null, one might then compare two models, one with two-way interaction effects involving X and one without those interactions. A model comparison F-test could then assess the interactions involving X.

The model <u>without the Zs</u> may be written as

$$Yi = \alpha + \beta_{yx}X_i + e_i$$
$$= (1 \; Xi) \begin{pmatrix} \alpha \\ \beta_{yx} \end{pmatrix} + e_i, \qquad (13)$$

where X_i = SECTOR for school i. In full matrix notation, Equation 13 becomes $Y = X\mathbf{B}_{yx} + e$ **(14)**

The model <u>with the Zs</u> is

$$Y_i = \alpha^* + \beta_{yx.z}X_i + \gamma_1 Z_{1i} + \gamma_2 Z_{2i} + e*_i^* \qquad (15)$$
$$= (1 \; X_i) \begin{pmatrix} \alpha^* \\ \beta_{yx \cdot z} \end{pmatrix} + (Z_{1i} \; Z_{2i}) \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} e_i^*$$

where Z_{1i} = MEAN SES and Z_{2i} = PROP ACADEMIC for school i. In full, "no-subscript" matrix notation, Equation 15 becomes

$$Y = X\mathbf{B}_{yx.z} + Z\Gamma + e^* \qquad (16)$$

In order to compute the needed variance-covariance matrix of estimated differences, however, we must compute another pair of regressions, that is, the regression of MEANSES and of PROP ACADEMIC on SECTOR. The model is

$$Z_{1i} = \theta_1 + \beta_{z_1x}X_i + u_{1i} \qquad (17)$$
$$Z_{21} = \theta_2 + \beta_{z_2x}X_i + u_{2i}$$

or

$$(Z_{1i} \; Z_{2i}) = (1 \; X_i) \begin{pmatrix} \theta_1 & \beta_{z_1x} \\ \theta_2 & \beta_{z_2x} \end{pmatrix} + (u_{1i} \; u_{2i}). \qquad (18)$$

In full matrix notation we have

$$Z = X\mathbf{B}_{zx} + u. \qquad (19)$$

Tables 8 and 9 provide the results of these regressions.

Table 8: Estimated Regression of Z_1 on X

Predictor	Coef	SE	t-ratio	p
Constant	-0.130	0.041	-3.17	.00
SECTOR	0.296	0.062	4.78	.00
$s = 0.3882$	R-sq = 12.6%		R-sq (adj) = 12.1%	

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	3.441	3.441	22.83	.00
Error	158	23.808	0.151		
Total	159	27.248			

Table 9: Estimated Regression of Z_2 on X

Predictor	Coef	SE	t-ratio	p
Constant	0.363	0.020	18.11	.00
SECTOR	0.346	0.030	11.42	.00

s = 0.1900	R-sq = 45.2%	R-sq (adj) = 44.9%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	4.707	4.707	130.38	.00
Error	158	5.705	0.036		
Total	159	10.413			

The two regression outputs provide estimates of the standard deviations of the residuals. To get the covariance, we can save the residuals and compute their correlation, from which the covariance follows. In this case the correlation of the residuals is r = .593, and, thus, their covariance is (0.3882)*(0.1900)*(0.593)=0.04374.

Define $\delta = (\delta_0 , \delta_1)^T = B_{yx} - B_{yx.z}$ and d as the estimate of δ. To compute the variance of d, we apply Allison's (1995) Equation A10[6]:

$$Var(d) \approx \hat{B}_{zx} Var(\hat{\Gamma}) \hat{B}_{zx}^T + \hat{\Gamma}^T Var(u|X) \hat{\Gamma}(X^T X)^{-1}.$$

(20)

Sample estimates are substituted for unknown population quantities in Equation 20. Many statistical packages will print out the inverse of the $X^T X$ matrix. We obtain

6 One generally seeks to retain as many decimal places as possible in these computations. Hence, the numbers in Equation 6 are not rounded as in the Tables.

$$Var\begin{pmatrix} d_0 \\ d_1 \end{pmatrix} \approx \begin{pmatrix} -.1295 & .3627 \\ .2950 & .3458 \end{pmatrix} * \begin{pmatrix} .02899 & -.03511 \\ -.03511 & .12001 \end{pmatrix} * \begin{pmatrix} -.1295 & .2950 \\ 3627 & .3458 \end{pmatrix}$$

$$+ (.0138 \quad -.9421) * \begin{pmatrix} .15068 & .04374 \\ .04374 & .03610 \end{pmatrix} * \begin{pmatrix} .0138 \\ -.9421 \end{pmatrix}$$

$$= \begin{pmatrix} .02004 & .01153 \\ .01153 & .01061 \end{pmatrix}.$$

(21)

We now compute a 95% confidence interval for the difference in the coefficient for SECTOR before and after controlling for the two Zs:

$$95\% \ CI(\delta) = [-.322 - (1.96) * (.0106)^{\frac{1}{2}}, \ -.322 + (1.96) * (.0106)^{\frac{1}{2}}]$$

$$= (-.524, -.120).$$

(22)

Based on the confidence interval (Equation 22), values of o very near zero are not plausible, though the range of plausible values is quite large relative to the value of this difference.

Example 3: The Case of Two Xs and One Zs

In the last example, we had a single X, and the task was to determine the extent to which two Zs could account for the relationship between X and Y. The logic extends to the case of a single X and multiple Zs. Now we consider the case of two Xs and a single Z. The task is to determine the extent to which the relationships between each X and Y can be explained by adding Z. Of course the logic will extend to the case of multiple Zs.

In the context of our illustrative data, our interest focuses on explanatory variables X_1 = SECTOR and X_2 = MEANSES. We ask whether the relationships between these Xs and Y can, in part, be "explained" by Z = PROP ACADEMIC, that is, the proportion of students in the school in the academic track. It may be that Catholic schools have more favorable DISCLIM because more of their students are assigned to the academic track, based on the assumptions that DISCLIM is generally more favorable within the academic track than the non-academic track. It also may be that any relationship between the social composition of the school, as indicated by MEANSES, and DISCLIM, is also attributable to the proportion of students in the academic track. The reasoning here is that schools with more advantaged compositions will tend to assign more students to the academic track than will schools with less advantaged compositions, and, for that reason, those "high-MEANSES" schools will enjoy comparatively favorable DISCLIM. The logic, then, is to compare a model with SECTOR and MEANSES as predictors to a model in which SECTOR, MEANSES, and PROP ACADEMIC are predictors. However, such a model comparison assumes no interactions between Z and each X. We therefore consider the interactions first.

1. <u>Are any two-way interactions needed in the model</u>? Although the X_1 by X_2 interaction is not of central interest, we include it in the model to avoid possible bias of other estimates. The model therefore includes all two-way interactions. Results are in Table 10.

Table 10: Estimated Regression with Two-Way Interactions

Predictor	Coef	SE	t-ratio	p
Constant	0.880	0.226	3.90	.00
X_1 = SECTOR	-1.073	0.147	-7.28	.00
X_2 = MEANSES	-0.068	0.246	-0.28	.78
Z = P. ACAD.	-0.745	0.532	-1.40	.16
X_1 by X_2	0.184	0.422	0.44	.66
X_1 by Z	-0.338	0.716	-0.417	.64
X_2 by Z	-0.410	0.719	-0.57	.57

R^2 = 54.2%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	6	82.224	13.704	30.15	.00
Error	153	69.539	0.455		
Total	159	151.763			

The univariate t-tests for the interactions in Table 11 are non-significant. However, a more appropriate test is the omnibus F test, where we compare the fit of the model with the interactions to the fit of the model without them (see Table 7):

$$f = \frac{SS_{residuals|no\ interaction} - SS_{residuals|interaction}}{SS_{residuals|no\ interactions}}$$
$$* \frac{denominator\ df|interactions}{number\ of\ interactions}$$ **(23)**

$$= \frac{69.837 - 69.539}{69.837} * \frac{153}{3} = .22$$

Here f is compared to the percentiles of the F distribution with $df = 3, 153$, showing no evidence against the null hypothesis that the three interactions are null.

A note on interactions involving continuous X. In this example, X_2 = MEANSES is continuous. Our test of the significance of $X_2 * Z$ is a test of the linear-by-linear interaction. To retain the null hypothesis does not imply that the association between Z and Y is equal at every value of X_2 because the interaction may have a non-linear character. For example, it may be that the association between Z and Y is positive at high and low values of X_2 but not at moderate values of X_2. Further exploratory analyses are generally warranted. With this caveat in mind, we proceed to compare coefficients for X_1 and X_2, before and after including Z.

2. How do the coefficients for SECTOR and MEANSES compare, before and after adjusting for PROP ACADEMIC? A comparison of results between Table 4 (which has SECTOR and MEANSES predicting DISCLIM) and Table 7 (where SECTOR, MEANSES, and PROP ACADEMIC predict DISCLIM) provide point estimates of the difference between models in coefficients for SECTOR and MEANSES. For SECTOR, the difference is -1.322 - (-1.077) = -0.245. For MEANSES, the difference is -0.260 - (0.014) = -0.273. However, the point estimates alone can be misleading. To compute confidence regions and tests, we need to compute the variance-covariance matrix of the estimates. To compute those, it is convenient to represent the two models in matrix notation.

The model without Z may be written as

where X_{1i} = SECTOR and X_{2i} = MEANSES for school i. In full matrix notation, Equation 24 becomes

The model with Z is

$$Y_i = \alpha + \beta_{yx_1}X_{1i} + \beta_{yx_2}X_{2i} + e_i$$
$$= (1 \quad X_{1i} \quad X_{2i}) \begin{pmatrix} \alpha \\ \beta_{yx_1} \\ \beta_{yx_2} \end{pmatrix} + e_i,$$ **(24)**

$$Y = X\mathbf{B}_{yx} + e.$$ **(25)**

$$Y_i = \alpha^* + \beta_{yx_1.z}X_{1i} + \beta_{yx_2.z}X_{2i} + \gamma Z_i + e_i^*$$
$$= (1 \quad X_{1i} \quad X_{2i}) \begin{pmatrix} \alpha^* \\ \beta_{yx_1z} \\ \beta_{yx_2z} \end{pmatrix} + Z_i\gamma + e_i^*,$$ **(26)**

where Z_i = PROP ACADEMIC for school i. In full, "no-subscript" matrix notation, Equation 26 becomes

$$Y = X\mathbf{B}_{yz.x} + Z\Gamma + e^*.$$ **(27)**

15

In order to compute the needed variance-covariance matrix of estimated differences, however, we must compute another regression, that is, the regression of Z = PROP ACADEMIC on X_1 = SECTOR and X_2 = MEANSES. The model is

$$Z_i = \theta + \beta_{zx_1}X_{1i} + \beta_{zx_2}X_{2i} + u_i$$

$$= (1 \quad X_{1i} \quad X_{21})\begin{pmatrix}\theta \\ \beta_{zx_1} \\ \beta_{zx_2}\end{pmatrix} + u_i,$$

(28)

or, in full matrix notation,

$$Z = X\mathbf{B}_{zx} + u$$

(29)

Table 11 provides the results of this regression.

Table 11: Estimated Regression of Z on X_1 and X_2

Predictor	Coef	SE	t-ratio	p
Constant	0.400	0.017	23.99	.00
SECTOR	0.260	0.026	9.93	.00
MEANSES	0.290	0.031	9.22	.00

$R^2 = 64.5\%$

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	2	6.712	3.356	142.40	.00
Error	157	3.700	0.024		
Total	159	10.4118			

The results in Table 11 give evidence that SECTOR and MEANSES are related to PROP ACADEMIC. Thus, it is plausible that PROP ACADEMIC might "mediate" relationship between these two predictors and DISCLIM. Define $\delta = (\delta_0, \delta_1, \delta_2)^T = B_{yx} - B_{yx.z}$ and d as the estimate of δ. To compute the variance-covariance matrix of the differences between coefficients across models, we again apply Allison's (1995) Equation A10:

$$Var(d) \approx \hat{B}_{zx}Var(\hat{\Gamma})\hat{B}_{zx}^T + \hat{\Gamma}^TVar(u|X)\hat{\Gamma}(X^TX)^{-1}.$$

(30)

Substituting from Tables 7 and 12 enables us to evaluate Equation 30:

$$Var\begin{pmatrix}d_0 \\ d_1 \\ d_2\end{pmatrix} \approx \begin{pmatrix}.4002 \\ .2600 \\ .2902\end{pmatrix} * (.1210) * (.4002 \quad .2600 \quad .2902)$$

$$+ (-.9421) * (.0236) * (-.9421) * \begin{pmatrix}.01182 & -.01272 & .00544 \\ -.01272 & .02907 & -.01242 \\ .00544 & -.01242 & .04200\end{pmatrix}$$

$$= \begin{pmatrix}.01963 & .01232 & .01416 \\ .01232 & .00878 & .00887 \\ .01416 & .00887 & .01106\end{pmatrix}.$$

(31)

The variance-covariance matrix computed in Equation 31 can be used to compute confidence regions and tests, for example confidence intervals for the difference in coefficients associated with SECTOR and MEANSES after adding PROP ACADEMIC:

$$95\% \ CI(\delta_1) = -.273 \pm (2.27) * (.0088)^{1/2}$$
$$= (-.458, \ -.032)$$
$$95\% \ CI(\delta_2) = -.245 \pm (2.27) * (.011)^{1/2}$$
$$= (-.517, \ -.024).$$

(32)

Here 2.27 is the critical value of the Bonferroni-adjusted t at the five percent level of significance.

Limitations and Future Work Needed

All of the results presented here are based on the standard linear model assumptions regarding the distribution of Y given X and Z. These results, as illustrated, can be obtained using standard output of the commonly-used regression packages. In the case of a single X and a single Z (Example 1), the computation of the standard error using Allison's Equation 7 can readily be adapted to incorporate standard error estimates based on relaxed assumptions. Thus, $\text{Var}(\hat{\beta}_{xy.z})$ and $\text{Var}(\hat{\gamma})$ used above in Equation 10 could be estimated via the bootstrap (Efron, 1982), via a robust "Huber-corrected" standard error (c.f. Zeger, Liang, & Albert, 1988), or via a design-consistent estimate based on a jackknife or a Taylor-series method.

In the multivariate cases (Examples 2 and 3), we used Allison's Equation $A10$. This applies only under the standard linear model assumptions of the conditional independence and homoscedasticity of the residuals given X and Z. Allison's Equation $A6$ provides a more generally applicable but computationally more difficult formula. Raudenbush and Sampson (1996) provide methods that handle the standard linear model case but extend to multilevel designs and latent variables. The required assumption is that Y and Z are multivariate normal given X. All of the estimators we have discussed are consistent. Though little is known about their small-sample properties, the large-sample approximations we have used will be appropriate in most analyses conducted using NCES data sets, which appear uniformly to be based on large samples.

Finally, we have restricted our interest to continuous outcomes. Generally applicable methods for dichotomous data, count data, or waiting times, perhaps under the rubric of a generalized linear model, await further research. Clogg et al. (1995) provide a good start by providing methods that apply when X and Z are considered fixed rather than random.

APPENDICES

Appendix A: References

Appendix B: Figure 1. Unadjusted and Adjusted Mean Difference in the Analysis of Covariance

Appendix C: Panel Participants

Appendix A: References

REFERENCES

Allison, P. (1995). The impact of random predictors on comparisons of coefficients between models: Comment on Clogg, Petkova, and Haritou. *American Journal of Sociology*, 100(5), 1294-1305.

Beaton, A. (1988). The NAEF 1985-86 reading anomaly: A technical report. Princeton, NJ: Educational Testing Service.

Clogg, C., Petkova, E., & Haritou, A. (1995). Symposium on applied regression: Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100(5), 1261-1293.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Johnson, P., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*, 1, 57-93.

Lee, V., & Bryk, A. (1989). A multilevel model of the social distribution of educational achievement. *Sociology of Education*, 62, 172-192.

Oaxaca, R., & Ransom, M. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61, 5-21.

Raudenbush, S., & Sampson, R. (1996, October 15). Comparing regression coefficients between models with extensions to latent variables and multilevel designs. Paper prepared for workshop sponsored by the National Institute of Statistical Sciences, Research Triangle Park.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin*, 88, 819-821.

Walker, H., & Lev, J. (1953). Statistical inference. New York, NY: Holt, Rinehart & Winston.

Zeger, S., Liang, K., & Albert, P. (1988). Models for longitudinal data: A likelihood approach. Biometrics, 44, 1049-60.

Appendix B: Figure 1. Unadjusted and Adjusted Mean Difference in the Analysis of Covariance

Figure 1. Unadjusted and Adjusted Mean Difference in the Analysis of Covariance
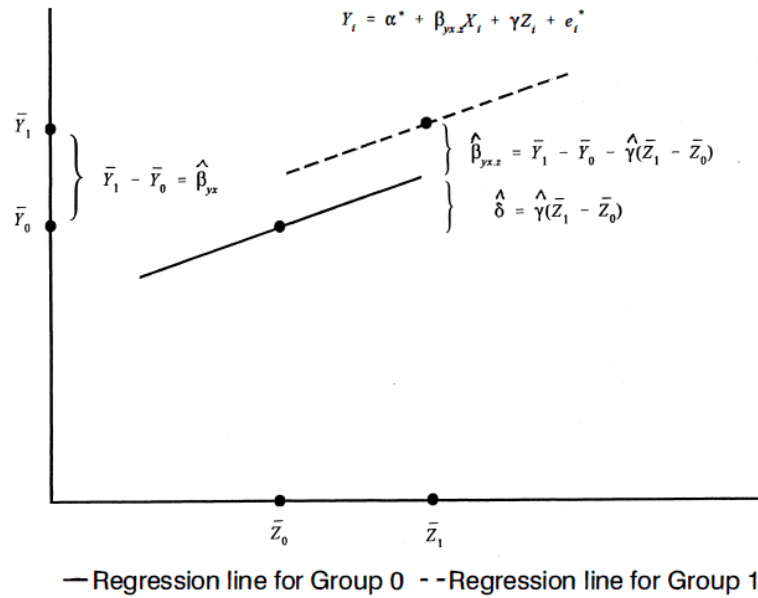
Appendix C: Technical Report Authors

Stephen W. Raudenbush (Chair)

Professor, Research Design and Statistics, Department of Counseling, Educational Psychology and Special Education, Michigan State University

Leon Gleser

Professor of Statistics, Director of Graduate Studies, Department of Statistics, University of Pittsburgh

Larry Hedges

Stella M. Rowley Professor, Departments of Education, Psychology, Sociology, and The Harris Graduate School of Public Policy Studies, The University of Chicago

Eugene G. Johnson

Principal Research Scientist, Educational Testing Service, Princeton, NJ

Eva Petkova

Departments of Biostatistics and Psychiatry at Columbia University, Director of Biostatistics, New York State Psychiatric Institute