Published in final edited form as:

J Educ Behav Stat. 2011 December ; 36(6): 699–719. doi:10.3102/1076998610396889.

Solutions for Determining the Significance Region Using the Johnson-Neyman Type Procedure in Generalized Linear (Mixed) Models

Ann A. Lazar and
University of California, San Francisco

Gary O. Zerbe
University of Colorado, Denver

Abstract

Researchers often compare the relationship between an outcome and covariate for two or more groups by evaluating whether the fitted regression curves differ significantly. When they do, researchers need to determine the “significance region,” or the values of the covariate where the curves significantly differ. In analysis of covariance (ANCOVA), the Johnson-Neyman procedure can be used to determine the significance region; for the hierarchical linear model (HLM), the Miyazaki and Maier (M-M) procedure has been suggested. However, neither procedure can assume nonnormally distributed data. Furthermore, the M-M procedure produces biased (downward) results because it uses the Wald test, does not control the inflated Type I error rate due to multiple testing, and requires implementing multiple software packages to determine the significance region. In this article, we address these limitations by proposing solutions for determining the significance region suitable for generalized linear (mixed) model (GLM or GLMM). These proposed solutions incorporate test statistics that resolve the biased results, control the Type I error rate using Scheffé’s method, and uses a single statistical software package to determine the significance region.

Keywords

generalized estimating equations; hierarchical linear models; longitudinal analysis; multiple testing; simultaneous testing

Comparative studies often test whether the fitted regression curves significantly differ between two or more groups. More importantly, if a significant difference arises between the fitted curves, researchers often need to know the “significance region,” or the range of the continuous covariate (also known as the independent variable or predictor) where the groups statistically differ. Determining the significance region involves comparing the curves at each unique predictor value (X), thus yielding statistical tests for each X , which in educational and behavioral studies with continuous predictors can be well over 1,000 tests.

It is well-known that if each null hypothesis is tested at nominal level α , the overall Type I error rate will be substantially larger than α . When the Type I error rate is uncontrolled then significant differences between groups can be due to chance findings leading to erroneous conclusions. In a typical model with a single covariate, researchers often ignore the inflated Type I error rate and assume that the observations are normally (Gaussian) distributed. But researchers often have to assume nonnormally distributed data (e.g., Bernoulli distribution for dichotomous outcomes, Poisson distribution for counts) from models with single or multiple covariates. While researchers can fit generalized linear models (GLM or marginal models) and generalized linear mixed models (GLMM or conditional models), assuming any normal or

nonnormal distribution from the exponential family for independent and correlated observations, respectively, this does not resolve the statistical problem associated with determining the significance region.

This problem has been studied in the analysis of covariance (ANCOVA) setting, where the significance region is often determined by using the Johnson-Neyman (J-N) procedure (Johnson & Neyman, 1936). In the ANCOVA setting, Steffens (1968) and Zerbe, Archer, Banchemo, and Lechner (1982) determined the significance region by developing an explicit solution, which involves solving a quadratic inequality for two unparallel regression lines after the significance level is adjusted for multiple testing using Scheffé's method. When the fitted model includes two or more covariates of interest, the typical solution is cast within a general linear model framework that uses symbolic processing capabilities of computational software (Hunka, 1995; Hunka & Leighton, 1997) since the explicit solution is intractable (Johnson & Fay, 1950). The Miyazaki and Maier (M-M) (2005) procedure for determining the significance region for correlated Gaussian distributed data involves fitting hierarchical linear mixed models (HLM) using statistical software (Raudenbush, Bryk, Cheong, & Congdon, 2000), where special cases of GLMM include HLM and ANCOVA, and then solving for the significance region using the symbolic processing capabilities of the Mathematica software package. The M-M procedure uses the Wald test (also termed Z-test), which has a chi-square distribution, as opposed to the exact F -statistic for balanced data used in the ANCOVA setting for the J-N procedure.

However, neither the J-N nor the M-M procedure can be applied directly to nonnormally distributed data, since these procedures assume the data to be Gaussian. Furthermore, the Wald test used in the M-M procedure often underestimates the true variability in the estimated fixed effects (referred to as a downward bias), as described by Dempster, Rubin, and Tsutakawa (1981), and the M-M procedure does not control the inflated Type I error rate due to multiple testing. Moreover, when two software packages are used to determine the significance region, such as statistical and mathematical software packages as described in the M-M procedure, transcription errors can increase (e.g., copying the variance-covariance matrix from the statistical package to the mathematical package).

In this article, we address the limitations of the J-N and M-M procedures by proposing solutions to the problem of determining the significance region for marginal and conditional models with single or multiple covariates, where these models can be adjusted for potential confounders. The solutions presented in this article involve test statistics based on the t - and F -statistic, thereby resolving the biased results from the Wald test that assumes a chi-square distribution (Dempster et al., 1981). Certainly, for very large sample sizes we can assume a chi-square distribution for the special case of the F with s numerator and infinite denominator degrees of freedom. The solutions also control the Type I error rate at level α by using Scheffé's method. In so doing, the statistical software package can fit the GLM and GLMM model, and solve for the significance region without using the symbolic processing capabilities of a mathematical software package. Using a single software package to solve for the significance regions reduces the computing costs and errors associated with using multiple software packages. We illustrate this new methodology for determining the significance region by answering important questions from the High School and Beyond Survey (HSBS; Coleman, Hoffer, & Kilgore, 1982), which is the same example described in Miyazaki and Maier (2005).

1.0 GLM and GLMM

Building a GLM requires three decisions. The first decision involves the distribution of the data, which may be from any probability distribution of the exponential family. The second decision involves the systematic component in the model. This specifies the effect of the

independent variables, X , on the mean of the distribution of Y , known as a linear predictor, $\eta = X\beta$ where β is a regression parameter vector. The term linear implies that η must be linear in the regression parameters. The third decision involves the link function, which describes the relationship between the marginal mean of Y denoted by m , and η , such that, the $E(Y) = m$ and $g(m) = X\beta$ where $g(\cdot)$ is a link function. The link function relates the parameters of the distribution to the predictor. Any differentiable monotonic link function $g(\cdot)$ can be chosen. For example, the normal distribution link function is usually the identity, such that $g(m) = m$, the Poisson distribution link function is usually the log link function, such that $g(m) = \log(m)$, and the Bernoulli distribution link function is usually the logistic or logit link function, such that $g(m) = \log(m/(1-m))$ (McCullagh & Nelder, 1989).

The three decisions needed to specify a GLM are applicable to GLMM. The conditional mean of Y depends on the fixed and random effects via the linear predictor $\eta = X\beta + Zu$ where $g\{E(Y|u)\} = \eta = X\beta + Zu$ and u represents the random effects with design matrix Z , and β denotes the fixed effects with design matrix X . The probability distribution of the random effects can be from any multivariate distribution, and the multivariate normal distribution is often selected. When correlated data are involved, it may be adequate to specify a marginal distribution, which is the basis of the generalized estimation equation approach (GEE), such that $g\{E(Y)\} = \eta = X\beta$ (McCulloch & Searle, 2001).

$$\text{When testing the null hypothesis, } H_0: \underbrace{\theta}_{s \times 1} = \underbrace{C}_{s \times (p+q)} \begin{pmatrix} \underbrace{\beta}_{p \times 1} \\ \underbrace{u}_{q \times 1} \end{pmatrix} = C_1\beta + C_2\mu = 0$$

for a fixed contrast matrix C , θ represents s predictable function, which are linear combinations of the fixed and random effects. When $C_2 = 0$, then the null hypothesis is of fixed effects, otherwise, the hypothesis is a function of the conditional mean (e.g., GLMM). For example, if the researcher has to test the joint hypothesis of equality of the intercepts, α_1 and α_2 , and slopes, β_1 and β_2 , of two regression lines, they would choose $C_2 = 0$, and

$$\underbrace{\theta}_{2 \times 1} = \underbrace{C_1}_{2 \times 4} \underbrace{\beta}_{4 \times 1} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{bmatrix} \alpha_1 - \alpha_2 \\ \beta_1 - \beta_2 \end{bmatrix}.$$

The null hypothesis, $H_0: \theta = 0$ can be tested with a t - or F -statistic. When C is a single row,

$t(\hat{\theta}) = \hat{\theta} / \sqrt{\hat{V}_{\hat{\theta}}}$, where $\hat{V}_{\hat{\theta}} = C\hat{L}C'$ and $\hat{L}$ is the empirical covariance matrix of $\hat{\beta} - \beta$ and $\hat{u} - u$. When s , the rank of C , is greater than or equal to one, the null hypothesis, $\theta = 0$, can be

$$F(\hat{\theta}) = \left[\underbrace{\tilde{\theta}}_{1 \times s} \left(\underbrace{\tilde{V}_{\tilde{\theta}}}_{s \times s} \right)^{-1} \underbrace{\tilde{\theta}}_{s \times 1} \right] / s$$

tested with $F(\hat{\theta})$ often has an approximate F -distribution with s numerator degrees of freedom and v denominator degrees of freedom. The t - and F -statistic allows the investigator to make inference about θ that accounts for the variance-covariance matrix they select (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006). Certainly, the usual model assumptions must hold to assure valid inference (Snedecor & Cochran, 1989).

In general, a single null hypothesis of the form, $\underbrace{h}_{1 \times s} \underbrace{\theta}_{s \times 1} = 0$, is rejected in favor of the alternative, $h\theta \neq 0$, when $t(h\hat{\theta}) > \sqrt{F_{1-\alpha, 1, v}}$. The t -statistic associated with $h\hat{\theta}$ is $t(h\hat{\theta}) = h\hat{\theta} / \sqrt{h\hat{V}_{\hat{\theta}}h'}$ and $F_{1-\alpha, 1, v}$ is the $1 - \alpha$ percentage point of an F -distribution with 1 and v numerator and denominator degrees of freedom, respectively. Multiple null hypotheses of the form, $h\theta = 0$, are rejected in favor of the alternatives, $h\theta \neq 0$, whenever the t -statistics associated with $h\hat{\theta}$ exceed Scheffé's (1959) criterion, $\sqrt{sF_{1-\alpha, s, v}}$, that is,

$$t(h\hat{\theta}) > \sqrt{sF_{1-\alpha, s, v}} \quad (1)$$

where $F_{1-\alpha, s, v}$ is the $1 - \alpha$ percentage point of an F distribution with s and v numerator and denominator degrees of freedom, respectively. For example, the vertical difference between two regression lines at any X can be expressed as

$h_x\theta = \begin{pmatrix} 1 & X \end{pmatrix} \begin{bmatrix} \alpha_1 - \alpha_2 \\ \beta_1 - \beta_2 \end{bmatrix} = (\alpha_1 + \beta_1 X) - (\alpha_2 + \beta_2 X)$, and multiple null hypotheses can be tested simultaneously at two or more unique values of X by repeating (1) for each X . The Scheffé method of multiple comparisons controls the familywise Type 1 error rate for tests of all linear functions, $h_x\theta$ of θ , at level α .

When comparing regression curves suitable for GLM and GLMM, we carefully choose h that

identifies the subset of linear functions of $\begin{pmatrix} \beta \\ u \end{pmatrix}$, which u corresponds to the vertical differences between the curves at any X . C is chosen, such that θ is an $s \times 1$ vector whose elements are the differences in intercepts and difference in slopes of the curves being compared. Suppose θ_c represents the difference in the intercepts and θ_A denotes the difference in the slopes, then

$\theta = \begin{pmatrix} \theta_c \\ \theta_A \end{pmatrix}$. The numerator degrees of freedom, s , are always determined as the number of rows from the C matrix, and the denominator degrees of freedom, v , are often estimated from the data (Molenberghs & Verbeke, 2005), particularly for longitudinal applications (e.g., GLMM). Sometimes v are simply the sample size (n) minus the number of parameters (e.g., for comparing two regression curves $v = n - 4$). In general for most longitudinal applications, different estimation methods of v , such as residual, Satterthwaite-type or Kenward and Roger (1997), will lead to very similar p -values (Molenberghs & Verbeke, 2005).

We can determine whether the regression curves generated from GLM or GLMM statistically differ by jointly testing the equality of the intercepts and the equality of the slopes. If this joint test is not statistically significant then the curves do not differ, and the significance region is the empty set. However, if the joint test is statistically significant, then we can compare the curves at many values of X by using (1) to compare each X 's t -statistic with the Scheffé criterion, $\sqrt{sF_{1-\alpha, s, v}}$. Or, much more efficiently, we can develop an explicit solution as follows.

The significance region is determined explicitly by rewriting (1) as a quadratic inequality in X ,

$$h_x [\theta\theta' - sF_{1-\alpha, s, v} V_{\theta}] h_x' > 0.$$

Let $V_{\theta(C)}$, $V_{\theta(A)}$, and $\text{Cov}_{\theta(A)\theta(C)}$ be the empirical prediction error variances (V) and covariance (Cov) of θ_C and θ_A . Then, the parabolic (or quadratic) inequality is:

$$\begin{pmatrix} 1 & X \end{pmatrix} \left[\begin{pmatrix} \theta_C^2 & \theta_C \theta_A \\ \theta_A \theta_C & \theta_A^2 \end{pmatrix} - sF_{1-\alpha, s, v} \begin{pmatrix} V_{\theta(C)} & \text{Cov}_{\theta(C)\theta(A)} \\ \text{Cov}_{\theta(A)\theta(C)} & V_{\theta(A)} \end{pmatrix} \right] \begin{pmatrix} 1 \\ X \end{pmatrix} > 0$$

or $Ax^2 + Bx + C > 0$ where $A = \theta_A^2 - sF_{1-\alpha, s, v} V_{\theta(A)}$, $B = 2(\theta_A \theta_C - sF_{1-\alpha, s, v} \text{Cov}_{\theta(A)\theta(C)})$, $C = \theta_C^2 - sF_{1-\alpha, s, v} V_{\theta(C)}$, and the discriminant, $D = B^2 - 4AC$. Note, $s = 2$ for comparing two groups and the Scheffé criterion is determined from the number of rows in the C matrix, which is 2. When comparing more than two groups, s and h will change (for example, for three groups, when $C_2 = 0$, a researcher might choose

$$\theta = C_1 \quad \beta = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 - \alpha_3 \\ \beta_1 - \beta_2 \\ \beta_1 - \beta_3 \end{pmatrix},$$

with $s = 4$ and $h_X = (1 \ 1 \ X \ X)$.

After estimating A , B , C , and D , one of three possible cases determines the significance region, where the vertical difference between the lines is large enough to be statistically significant. If the discriminant is positive, $D > 0$, the solution is real and the significance region is bounded

$$\text{by } X_1 = \left[\frac{-B - \sqrt{D}}{2A} \right] \text{ and } X_2 = \left[\frac{-B + \sqrt{D}}{2A} \right].$$

Case I occurs if $A > 0$, which implies $D > 0$, then the test is significant for the X region satisfying:

$$X < X_1 \text{ or } X > X_2.$$

The region of nonsignificance lies between X_1 and X_2 and the region of significance lies outside X_1 and X_2 . Case I occurs if, and only if, the slopes differ by Scheffé's criterion, $sF_{1-\alpha, s, v}$, such that $A > 0$ occurs if and only if the square of the t -statistic for the equality of the slopes exceeds $sF_{1-\alpha, s, v}$.

Case II occurs if $A < 0$ and $D > 0$, then the test is significant for the X region satisfying:

$$X_2 < X < X_1.$$

Case II occurs when the slopes do not differ by Scheffé's criterion, and the simultaneous test of equality of the intercepts and slopes are significant by the F -test.

Case III occurs when $D < 0$ and $A < 0$, and the quadratic equation does not touch the X -axis. This occurs when the joint test is not statistically significant, where the vertical difference between the curves is never large enough to be statistically different. Additional clarification of the cases is provided in Appendix 1.0.

2.0 Application to High School and Beyond Survey

2.1 Continuous and Normally Distributed Outcome

Using the approach from Section 1.0, we can study the relationship between mathematics achievement score (MA) and socioeconomic status score (SES), a composite score of parental income, education, occupation, and achievement, between sectors: Catholic ($n = 70$) and public ($n = 90$) high schools. To do so, we use a subset of the data from the High School and Beyond Survey (HSBS) of 7,185 students nested within 160 high schools, which averaged 45 students per school (Coleman et al., 1982). This study can address important questions: (a) For what range of SES do the mathematics achievement scores statistically differ between private and public high schools; and (b) Which school sector does better, Catholic or public high schools?

Before addressing these questions, we fit GLMM, as described in Raudenbush and Bryk (2002, pp. 80–82), of mathematics achievement score, y_{ij} , for the i th student in the j th cluster (school):

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j} (\text{SES}_{ij} - \overline{\text{SES}}_j) + e_{ij}, \\ \beta_{0j} &= \gamma_{00} + \gamma_{01} \overline{\text{SES}}_j + \gamma_{02} \text{sector}_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} \overline{\text{SES}}_j + \gamma_{12} \text{sector}_j + u_{1j}, \end{aligned} \quad (2)$$

where $\overline{\text{SES}}_j$ denotes the school-averaged student SES, sector equals 1 for Catholic (C) schools or 0 for public (P) schools, and e_{ij} represents the unobservable random vector of normally distributed errors. The e_{ij} are assumed uncorrelated with the multivariate normally distributed random effects, u_{ij} , of the school mean (u_{0j}), and SES-achievement slope (u_{1j}).

Addressing these important questions (described above) requires comparing $E_C(Y_{ij})$ with $E_P(Y_{ij})$, as described in Miyazaki and Maier (M-M) (2005, p. 244). Then,

$E_C(Y_{ij}) - E_P(Y_{ij}) = \gamma_{02} + \gamma_{12} (\text{SES}_{ij} - \overline{\text{SES}}_j)$, where

$$\underbrace{\begin{pmatrix} 1 & x \end{pmatrix}}_{h_x} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_C \underbrace{\begin{pmatrix} \gamma_{02} \\ \gamma_{12} \end{pmatrix}}_{\beta} = \begin{pmatrix} 1 & x \end{pmatrix} \begin{pmatrix} \gamma_{02} \\ \gamma_{12} \end{pmatrix} = h_x \theta, \quad \text{results in the null hypotheses of no difference in mathematical achievement scores between Catholic and public schools for values of relative SES, } (\text{SES}_{ij} - \overline{\text{SES}}_j), \text{ or } x.$$

However, before determining the significance region, we first evaluate the joint test, with $\alpha = .05$, to determine whether the population average curves of Catholic and public schools differ. Since the joint test is statistically significant ($p < .0001$), we then determine the range of relative SES where the vertical difference between the fitted population average curves of MA is large enough to be statistically significant. Figure 1 displays the estimated difference (from Model 2) in mathematical achievement scores between sectors as a function of relative student SES. The results presented in Table 1 are used to calculate A , B , C , and D (below) to determine the significance region:

$$\begin{aligned} \widehat{A} &= (\widehat{\gamma}_{12})^2 - 2F_{(0.95, 2, 7022)} \widehat{V}_{\widehat{\gamma}_{12}} = (-1.6426)^2 - (2 \times 2.997 \times 0.2398^2) = 2.353; \\ \widehat{B} &= 2 \times \{ (\widehat{\gamma}_{02} \widehat{\gamma}_{12}) - 2F_{(0.95, 2, 7022)} \widehat{\text{Cov}}_{\widehat{\gamma}_{02} \widehat{\gamma}_{12}} \} \\ &= 2 \times \{ (1.2167 \times -1.6426) - (2 \times 2.997 \times 0.0060) \} = -4.0690; \\ \widehat{C} &= (\widehat{\gamma}_{02})^2 - 2F_{(0.95, 2, 7022)} \widehat{V}_{\widehat{\gamma}_{02}} = 1.2167^2 - (2 \times 2.997 \times 0.3064^2) = 0.9176; \\ \widehat{D} &= (\widehat{B}^2 - 4\widehat{A}\widehat{C}) = (-4.0690)^2 - 4 \times 2.353 \times 0.9176 = 7.920. \end{aligned}$$

Since $\hat{A} > 0$ (implies $\hat{D} > 0$) then we select Case I. Case I is verified by checking whether the slopes, γ_{12} , significantly differ ($p < .05$) using Scheffé's criterion, where the $|t\text{-value}| >$ Scheffé's criterion: ($|-6.85| > 2.448 = \sqrt{2 \times 2.997}$). After plugging in $\hat{A}$, $\hat{B}$, $\hat{C}$, and $\hat{D}$ for Case I, as described in Section 1, the results are $x < 0.266$ or $x > 1.462$. When the relative SES is between -3.657 and 0.266 , the students have higher mathematics achievement scores in Catholic schools than public schools after adjustment for multiple comparisons. When relative SES is between 1.462 and 2.850 , students have higher mathematics scores in public high schools compared with Catholic high schools. However, when relative SES is between 0.266 and 1.462 , there is no statistical difference in mathematical achievement between sectors. The SAS program that generated the fitted models and the significance region are provided in Appendix 2.0. The results from this example are consistent regardless of the chosen estimation method for the degrees of freedom, such as containment, residual, Satterthwaite-type, or Kenward and Roger (1997).

By using the new methodology presented in Section 1.0, we can also attain the solution presented in Miyazaki and Maier (M-M) (2005) by replacing Scheffé's criterion,

$2F_{(0.95, 2, 7022)} = 5.994$, with $\chi^2_{(df=1, \alpha=.05)} = 3.84$ ($x \leq 0.36$ or $x \geq 1.29$). While the M-M solution was obtained using symbolic processing capabilities from the Mathematica software package, the same solution presented in this article was obtained by using the newly developed explicit solution methodology described in Section 1.0, where the model can be adjusted for multiple covariates. In addition, M-M made inference beyond the observed data range, where we restricted our conclusions to the observed data. Furthermore, the M-M solution did not account for multiple testing, thus significantly inflating the false positive rate, which led to some erroneous conclusions. By using Scheffé's criterion, the inflated false positive rate is controlled, as described in Section 1.0.

2.2 Dichotomous Outcome

Other important questions can be addressed by using the High School and Beyond Survey including: (a) What range of SES does the mathematics achievement scores statistically differ between the Catholic high school sector and a particular public high school (for example, school 1,296); and (b) Does the Catholic high school sector have higher math scores than a particular public high school? These questions are especially important for parents who have their child in a particular high school and need to compare the particular high school with the Catholic school sector as a group.

We apply the approach described in Section 1.0 to address these questions by first fitting a similar model as (2) except we assumed a logistic link and Bernoulli distribution for the outcomes conditional on normally distributed random effects. Table 2 displays the estimates and standard errors from this fitted model.

To determine whether there is a difference in math achievement between the Catholic sector

and public school 1,296, we first test the joint hypothesis $\theta = \begin{pmatrix} \gamma_{02} - u_{0(1296)} \\ \gamma_{12} - u_{1(1296)} \end{pmatrix}$. The joint test provides evidence for a significant difference between the Catholic group curve and high school 1,296 curve ($p < .0001$). Figure 2 displays the estimated odds ratio of high mathematics achievement for the Catholic sector compared to public school 1,296. To determine the range of SES where the mathematics achievement scores statistically differ, we then estimate A , B , C and D :

$$\begin{aligned}
\widehat{A} &= (\widehat{\gamma}_{12} - \widehat{u}_{1(1296)})^2 - 2F_{(0.95, 2, 7023)} \widehat{V}_{\widehat{\gamma}_{12} - \widehat{u}_{1(1296)}} = (-0.4572)^2 - (2 \times 2.997 \times 0.0809^2) = 0.1698; \\
\widehat{B} &= 2 \times \left[\{(\widehat{\gamma}_{02} - \widehat{\mu}_{0(1296)}) \times (\widehat{\gamma}_{12} - \widehat{u}_{1(1296)})\} - 2F_{(0.95, 2, 7023)} \widehat{\text{Cov}}_{(\widehat{\gamma}_{02} - \widehat{\mu}_{0(1296)})}(\widehat{\gamma}_{12} - \widehat{u}_{1(1296)}) \right] \\
&= 2 \times [1.0777 \times -0.4572 - (2 \times 2.997 \times -0.0041)] = -0.9363; \\
\widehat{C} &= (\widehat{\gamma}_{02} - \widehat{u}_{0(1296)})^2 - 2F_{(0.95, 2, 7023)} \widehat{V}_{\widehat{\gamma}_{02} - \widehat{u}_{0(1296)}} = 1.0777^2 - (2 \times 2.997 \times 0.3066^2) = 0.5980; \\
\widehat{D} &= (\widehat{B}^2 - 4\widehat{A}\widehat{C}) = ((-0.9363)^2 - 4 \times 0.1698 \times 0.5980) = 0.4705.
\end{aligned}$$

Since $\widehat{A} > 0$, which implies $\widehat{D} > 0$, then we select Case I, which can be verified by checking whether the slopes, γ_{12} , significantly differ ($p < .05$) by Scheffé's criterion, where the $|t\text{-value}| > \text{Scheffé's criterion}$: ($|-5.65| > 2.448$). Then, $x < 0.737$ or > 4.777 .

We restrict our inference to the SES range where the data were collected. Therefore, we conclude that when relative SES is between -3.657 and 0.737 , students from the Catholic school sector have greater odds of higher mathematics achievement scores than the public school 1,296 after adjustment for multiple testing. When relative SES is between 0.737 and 2.850 , there is no statistical difference in mathematical achievement between the Catholic sector and public school 1,296. The statistical program provided in Appendix 2.0 was used to generate the above solutions. The results from this example are consistent regardless of the chosen estimation method for the denominator degrees of freedom. We could have also addressed whether there was a difference between this same public school, case 1,296, with

all public schools, where $\theta = \begin{pmatrix} -u_{0(1296)} \\ -u_{1(1296)} \end{pmatrix}$. Additional comparisons can also be considered using the methodology from Section 1.0.

3.0 Concluding Remarks

In this article, we proposed solutions for determining the significance region for comparing fitted curves generated from the GLM and GLMM with independent and correlated data, respectively. These solutions incorporate the F -statistic and control the Type I error rate due to multiple testing. We presented two solutions for determining the significance region. The first is obtained by comparing the t -statistic to Scheffé's criterion at each X . This solution can incorporate multiple covariates of interest in the model. The second solution, or explicit solution for a single covariate, eliminates testing at every value of the covariate. Before performing any calculations to determine the significance region, we suggested using the joint test, the simultaneous test of the differences in the intercepts and the difference in the slopes, which determines whether the curves differ significantly. While the explicit solution requires that η must be linear in the regression parameters, the joint test is applicable to both solutions whether comparing two or more groups or a model with potential confounders.

The High School and Beyond Survey (HSBS) illustrated how this new facile methodology addresses important research questions for parents and schools alike. Using the new GLMM methodology, where we assumed normally and Bernoulli distributed data, we could compare sectors, such as private and public schools, and a specific public school to the Catholic sector to determine the range of relative SES that had a significance difference in terms of student mathematics achievement. While we highlighted the use of the explicit solution in the HSBS example, where the model was adjusted for other covariates, the same solution could have been obtained by comparing the t -statistic with Scheffé's criterion at particular SES values.

For studies with large sample sizes, such as HSBS, infinite denominator degrees of freedom can be assumed. However, for small sample sizes, statistical software provides options for v ,

along with suggestions as to which ones are most appropriate (Littell et al., 2006). Once v is determined from the joint test, this same v must also be used to calculate A , B , C , and D to determine the significance region and, if desired, to compare curves at each X . Otherwise, the relationship between the confidence ellipse for the differences in intercepts and differences in slopes and the projections onto the axes for each X may change in a way that negates Scheffé's method (Schaalje, McBride, & Fellingham, 2001).

Accounting for the correlation may also affect the interpretation of the regression coefficients. If $g\{E(Y|u) = X\beta + Zu\}$, where $g(\cdot)$ is a nonidentity link, then $g\{E(Y) = X\beta\}$ for all β , when averaged over the distribution of random effects. Therefore, setting $u = 0$ results in average subject curves rather than population average curves, except for linear regression models with identity link (Molenberghs & Verbeke, 2004). Depending on the scientific question, the marginal model based on the GEE can be used if desired to make inference on the population.

After exchanging Scheffé's constant, $sF_{1-\alpha, s, v}$, for $t_{1-\alpha}^2$ the significance region becomes the complement of Fieller's theorem confidence interval for the point of intersection of two regression lines, where Fieller's theorem is useful for calculating confidence intervals for the ratio of two means (Steffens, 1968, 1971; Zerbe, 1978; Zerbe et al., 1982). Additional research is needed to confirm whether this result continues to hold for GLM and GLMM.

The focus of this article has been on educational research problems but this research area has wide applicability for other fields including psychometrics, medicine, and economics (Potthoff, 1964).

Acknowledgments

The authors thank Richard D. Gelber, Robert J. Gray, and Donald Halstead of Harvard School of Public Health and Dana-Farber Cancer Institute for their helpful comments on the manuscript. The authors also thank Jill M. Norris, Susan Mikulich-Gilbertson, Richard Jones, and Diane Fairclough of University of Colorado Denver for their useful comments.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: Partial grant support from the National Institutes of Health (T32-CA-09337 and P30-DE-020752).

References

- Coleman, J.; Hoffer, T.; Kilgore, S. High school achievement: Public, Catholic, and other schools compared. New York, NY: Basic Books; 1982.
- Dempster AP, Rubin DB, Tsutakawa RK. Estimation in covariance components model. *Journal of the American Statistical Association*. 1981; 76:341–353.
- Hunka S. Identifying regions of significance in ANCOVA problems having non-homogeneous regressions. *British Journal of Mathematical and Statistical Psychology*. 1995; 48:161–188.
- Hunka S, Leighton J. Defining Johnson-Neyman regions of significance in the three-covariate ANCOVA using Mathematica. *Journal of Education and Behavioral Statistics*. 1997; 22:361–387.
- Johnson PO, Fay LC. The Johnson-Neyman technique, its theory and application. *Psychometrika*. 1950; 15:349–367. [PubMed: 14797902]
- Johnson PO, Neyman J. Tests of certain linear hypothesis and their application to some educational problems. *Statistical Research Memoirs*. 1936; 1:57–93.
- Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997; 53:993–997.
- Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. SAS for mixed models second edition. Cary, NC: SAS Institute; 2006. p. 756
- McCullagh, P.; Nelder, JA. Generalized linear models. 2. London: Chapman and Hall; 1989.

- McCulloch, CE.; Searle, SR. Generalized, linear and mixed models. New York, NY: John Wiley; 2001.
- Miyazaki Y, Maier KS. Johnson-Neyman type technique in hierarchical linear models. *Journal of Education and Behavioral Statistics*. 2005; 30:233–259.
- Molenberghs G, Verbeke G. Meaningful statistical model formulations for repeated measures. *Statistica Sinica*. 2004; 14:989–1020.
- Molenberghs, G.; Verbeke, G. Models for discrete longitudinal data. New York: Springer; 2005. p. 41
- Potthoff RF. On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*. 1964; 29:241–256.
- Raudenbush, SW.; Bryk, AS.; Cheong, YF.; Congdon, R. HLM5—Hierarchical linear and nonlinear modeling. Chicago, IL: Scientific Software International; 2000.
- Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2. Newbury Park, CA: Sage; 2002.
- Schaalje, BG.; McBride, JB.; Fellingham, GW. Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference. SAS Institute; 2001. Approximation to distributions of test statistics in complex mixed linear models using SAS Proc MIXED.
- Scheffé, H. The analysis of variance. New York, NY: Wiley; 1959.
- Snedecor, WG.; Cochran, WG. Statistical methods. 8. Ames: Iowa State University Press; 1989.
- Steffens FE. On comparing two simple linear regression lines. *South African Statistical Journal*. 1968; 2:33–53.
- Steffens FE. On confidence sets for the ratio of two normal means. *South African Statistical Journal*. 1971; 5:105–113.
- Zerbe GO. On Fieller's theorem and the general linear model. *The American Statistician*. 1978; 32:103–105.
- Zerbe GO, Archer PG, Banchero N, Lechner AJ. On comparing regression lines with unequal slopes. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*. 1982; 242:178–180.

Biographies

ANN LAZAR is Assistant Professor at the University of California, San Francisco, 3333 California Street, Ste 495, San Francisco, CA 94118; ann.lazar@ucsf.edu. Her research interests include linear and nonlinear mixed modeling including hierarchical modeling, nonparametric procedures, and subgroup approaches.

GARY ZERBE is Professor Emeritus at the University of Colorado, Denver, 1300 East 17th Place, Box B119, Aurora, CO 80045; gary.zerbe@ucdenver.edu. His research interests include applications of dose response and time response curve methodology, multivariate linear and nonlinear mixed models, and randomization and other nonparametric procedures.

Appendix 1.0

Additional Clarification of the Cases

Let a , b , c , and d denote observed values of the random variables A , B , C , and D , respectively.

The roots of $ax^2 + bx + c > 0$ are of the form $x = \frac{-b \pm \sqrt{d}}{2a}$, where $d = b^2 - 4ac$.

Note that:

$$\begin{aligned} \left[x - \frac{-b-\sqrt{d}}{2a} \right] \left[x - \frac{-b+\sqrt{d}}{2a} \right] &= \left[\frac{2ax+b+\sqrt{d}}{2a} \right] \left[\frac{2ax+b-\sqrt{d}}{2a} \right] \\ &= \frac{[(2ax+b)+\sqrt{d}][(2ax+b)-\sqrt{d}]}{4a^2} = \frac{(2ax+b)^2 - (b^2-4ac)}{4a^2} \\ &= \frac{4a^2x^2 + 4abx + b^2 - b^2 + 4ac}{4a^2} = \frac{ax^2 + bx + c}{a} \end{aligned}$$

Note a in the denominator.

Case I

If $a > 0$ (implies $d > 0$), then $ax^2 + bx + c > 0$ when $\frac{ax^2+bx+c}{a} > 0$, then

$$\frac{ax^2+bx+c}{a} = \left[x - \frac{-b-\sqrt{d}}{2a} \right] \left[x - \frac{-b+\sqrt{d}}{2a} \right] > 0$$

$$\text{when } \left[x - \frac{-b-\sqrt{d}}{2a} \right] > 0 \text{ and } \left[x - \frac{-b+\sqrt{d}}{2a} \right] > 0$$

Significance Regions for GLMM

$$\text{or when } \left[x - \frac{-b-\sqrt{d}}{2a} \right] < 0 \text{ and } \left[x - \frac{-b+\sqrt{d}}{2a} \right] < 0.$$

$$\text{That is when } \left[x > \frac{-b-\sqrt{d}}{2a} \text{ and } x > \frac{-b+\sqrt{d}}{2a} \right] \text{ or } \left[x < \frac{-b+\sqrt{d}}{2a} \text{ and } x < \frac{-b-\sqrt{d}}{2a} \right].$$

That is when $x > \frac{-b+\sqrt{d}}{2a}$, where $\frac{-b+\sqrt{d}}{2a} = x_2$ or $x < \frac{-b-\sqrt{d}}{2a}$, where $\frac{-b-\sqrt{d}}{2a} = x_1$, such that $x < x_1$ or $x > x_2$.

Case II

If $a < 0$ and $d > 0$, then $ax^2 + bx + c > 0$ when $\frac{ax^2+bx+c}{a} < 0$.

$$\text{Then, } \frac{ax^2+bx+c}{a} = \left[x - \frac{-b-\sqrt{d}}{2a} \right] \left[x - \frac{-b+\sqrt{d}}{2a} \right] < 0 \text{ when } \left[x - \frac{-b-\sqrt{d}}{2a} \right] > 0 \text{ and } \left[x - \frac{-b+\sqrt{d}}{2a} \right] < 0$$

$$\text{or when } \left[x - \frac{-b-\sqrt{d}}{2a} \right] < 0 \text{ and } \left[x - \frac{-b+\sqrt{d}}{2a} \right] > 0. \text{ That is when } \left[x > \frac{-b-\sqrt{d}}{2a} \text{ and } x < \frac{-b+\sqrt{d}}{2a} \right] \text{ or } \left[x < \frac{-b-\sqrt{d}}{2a} \text{ and } x > \frac{-b+\sqrt{d}}{2a} \right].$$

$$\text{That is when } \left[\frac{-b-\sqrt{d}}{2a} < x < \frac{-b+\sqrt{d}}{2a} \right] \text{ or } \left[\frac{-b+\sqrt{d}}{2a} < x < \frac{-b-\sqrt{d}}{2a} \right].$$

$$\text{Let } x_1 = \frac{-b-\sqrt{d}}{2a} \text{ and } x_2 = \frac{-b+\sqrt{d}}{2a}$$

That is when $x_2 < x < x_1$ since $a < 0$.

Case III

If $a < 0$ and $d < 0$, there is no real solution $x = \frac{-b \pm \sqrt{d}}{2a}$ satisfying the inequality $ax^2 + bx + c > 0$. The vertical difference between the curves is never large enough to be statistically different.

Case where $a > 0$ and $d < 0$

The case where $a > 0$ and $d < 0$ cannot occur, that is, $a > 0$ implies $d > 0$.

Proof—For brevity, let $S^2 = sF_{1-\alpha, s, v}$, $V_A = V_{\theta(A)}$, $V_C = V_{\theta(C)}$, and $COV_{AC} = COV_{\theta(A)\theta(C)}$. Then

$$\begin{aligned} a &= \theta_A^2 - S^2 V_A, \\ b &= 2(\theta_A \theta_C - S^2 COV_{AC}), \\ c &= \theta_C^2 - S^2 V_C, \\ d &= b^2 - 4ac = [2(\theta_A \theta_C - S^2 COV_{AC})]^2 - 4[\theta_A^2 - S^2 V_A][\theta_C^2 - S^2 V_C], \text{ and} \end{aligned}$$

$$\begin{aligned} d/4 &= [\theta_A^2 \theta_C^2 + S^4 COV_{AC}^2 - 2\theta_A \theta_C S^2 COV_{AC}] - [\theta_A^2 \theta_C^2 - \theta_A^2 S^2 V_C - \theta_C^2 S^2 V_A + S^4 V_A V_C] \\ &= S^4 COV_{AC}^2 - 2\theta_A \theta_C S^2 COV_{AC} + \theta_A^2 S^2 V_C + \theta_C^2 S^2 V_A - S^4 V_A V_C \\ &= S^2 [S^2 COV_{AC}^2 - 2\theta_A \theta_C COV_{AC} + \theta_A^2 V_C + \theta_C^2 V_A - S^2 V_A V_C] \\ &= S^2 V_A V_C \left[\left(\frac{\theta_C^2}{V_C} - \frac{2\theta_A \theta_C COV_{AC}}{V_A V_C} + \frac{\theta_A^2}{V_A} - S^2 + \frac{S^2 COV_{AC}^2}{V_A V_C} \right) \right] \\ &= S^2 V_A V_C \left[\left(\frac{\theta_C^2}{V_C} - \frac{2\theta_A \theta_C COV_{AC}}{V_A V_C} + \frac{\theta_A^2 COV_{AC}^2}{V_A^2 V_C} \right) + \left(\frac{\theta_A^2}{V_A} - S^2 - \frac{\theta_A^2 COV_{AC}^2}{V_A^2 V_C} + \frac{S^2 COV_{AC}^2}{V_A V_C} \right) \right] \\ &= S^2 V_A V_C \left[\left(\frac{\theta_C}{\sqrt{V_C}} - \frac{\theta_A COV_{AC}}{\sqrt{V_A V_C}} \right)^2 + \left(1 - \frac{COV_{AC}^2}{V_A V_C} \right) \left(\frac{\theta_A^2}{V_A} - S^2 \right) \right] \\ &= S^2 \left(1 - \frac{COV_{AC}^2}{V_A V_C} \right) V_A V_C \left[\left(\frac{\frac{\theta_C}{\sqrt{V_C}} - \left(\frac{COV_{AC}}{\sqrt{V_A V_C}} \right) \frac{\theta_A}{\sqrt{V_A}}}{\sqrt{1 - \frac{COV_{AC}^2}{V_A V_C}}} \right)^2 + \left(\frac{\theta_A^2}{V_A} - S^2 \right) \right]. \end{aligned}$$

Then if $a = \theta_A^2 - S^2 V_A > 0$, it follows that $\frac{\theta_A^2}{V_A} - S^2 > 0$, $d/4 > 0$, and hence $d > 0$, since all other terms

in $d/4$ are positive. $\frac{COV_{AC}^2}{V_A V_C}$ is the square of the correlation coefficient between the difference in the intercepts and the difference in the slopes. A similar proof was presented by Steffens (1968).

Appendix 2.0

```

/****Note Any Statistical Program can be used such as STATA, SPSS, R,
SAS among others; Here, we use SAS to fit the models and generate the
significance region using the methodology from Section 1.0*/
/*****1982 High School and Beyond Survey*****/
libname hsb 'G://';
/*SAS Code for Section 2.0

```

```

Application to High School and Beyond Survey*/
data hsb;set hsb.hsb12;
cses = ses - meanses;
newsector=sector;
/*Create Dichotomous Outcome Based on the Median Math
Achievement*/
if mathach ge 13.131 then math = 1;
else math = 0;
if mathach =. then math=.;
run;
/*****Section 2.0 Continuous and Normally Distributed
Outcome*****/
proc sort data=hsb;by school;
proc mixed data = hsb noclprint covtest noitprint;
class school newsector;
model mathach = meanses sector cses meanses*cses sector*cses
/ solution ddf=157 ddfm=bw notest covb;
random intercept cses / subject = school(newsector) type=un;
/*Joint Test*/
contrast 'contrast Fixed' sector 1, sector*cses 1;
/****NEED THESE COMPONENTS TO CALCULATE THE EXPLICIT SOLUTION*/
/*Difference in Slopes to Calculate A*/
estimate 'slope' sector*cses 1;
/*Difference in the Intercepts to Calculate C*/
estimate 'int' sector 1;
/*Used to Calculate the Covariance for B*/
estimate 'intslope' sector 1 sector*cses 1;
/* Outputs the above information using the output delivery system*/
ods output estimates=estimates;
/*Alternative way to determine the Significance Region is to compare
Scheffe's Constant calculated below to the t-value
Here we tried the lowest, midvalue and highest value*/
estimate 'check -3.657' sector 1 sector*cses -3.657;
estimate 'check -0.404' sector 1 sector*cses -0.404;
estimate 'check 2.850' sector 1 sector*cses 2.850;
run;
/*Note: When random effects are involved, we recommend using SAS Proc
GLIMMIX since Proc MIXED does not appear to specify the correct
contrast, which can lead to incorrect results.*/
/*****SECTION 2.0 Dichotomous and Bernoulli Distributed
Outcome*/
proc sort data=hsb;by school newsector;
PROC GLIMMIX data=hsb oddsratio;
class school newsector;
model math(event='1') = meanses sector cses sector*cses/ s ddfm=bw
dist=bin covb;
random int cses /subject=school(newsector) type=chol;
parms /lowerb=1e-4,.,1e-4;
/*Joint Test Needed to Determine the Denominator Degrees of
Freedom*/
contrast 'contrast Public School 1296 to Cath Sector' sector 1 |int -1,
sector*cses 1|cses -1/subject 0 0 1;
/*Components Needed to Calculate the Explicit Solution*/

```

```

/* Difference in Intercepts for C*/
estimate 'int' sector 1 |int -1 /subject 0 0 1;
/*Difference in Slopes for A*
/ estimate 'slope' sector*cses 1 |cses -1 /subject 0 0 1;
/*Form 1:  $V[(a_1-a_2)+(b_1-b_2)]=V(a_1-a_2)+V(b_1-b_2)+2Cov[(a_1-a_2),(b_1-b_2)]$ 
Calc Var not Cov to get the cov for B*/
estimate 'Cov' sector 1 sector*cses 1 |int -1 cses -1 /subject 0 0 1;
/*ODS output Estimates to Calculate the Significance Regions using A,
B, C and D*/
ods output estimates=estimates GLIM;
/*The Significance Regions can be obtained by comparing the t-value to
Scheffes Constant*/
estimate 'Catholic School vs Public 1296 at -3.657' sector 1
sector*cses -3.657 |int -1 cses 3.657 /subject 0 0 1;
estimate 'Catholic School vs Public 1296 at -0.404' sector 1
sector*cses -0.404 |int -1 cses 0.404 /subject 0 0 1;
estimate 'Subject 1296 at 2.850' sector 1 sector*cses 2.850 |int -1 cses
-2.850 /subject 0 0 1;
run;
/*Comparing Catholic and Public Schools*/
/*Used to calculate Scheffes criterion called ff for */
data f;
/*Denominator degrees of freedom determined from the joint test
specified in the model*/
ff=finv(0.95,2,7022);
/*Use Scheffe to compare to the t-statistic*/
Scheffe = sqrt(2*ff);
match=1;
run;
data estimates2;set estimates;match=1;run;
data estimates3;merge f estimates2;by match;
data intestimates;set estimates3;
if label='int' then do;InterceptEst=estimate;IntVar=StdErr**2;
C = (InterceptEst**2)-(2*ff*IntVar);output;end;
keep C InterceptEst IntVar ff;
run;
data slopeestimates;set estimates3;
if label='slope' then do;SlopeEst = estimate;SlopeVar=StdErr**2;
slopesq=SlopeEst**2;
A = (slopesq)-(2*ff*SlopeVar);output;end;
keep A SlopeEst SlopeVar;
run;
data intslope;set estimates3;if label='intslope' then
do;IntSlopeEst = estimate;IntSlopeVar=StdErr**2;output;end;
keep ff IntSlopeEst IntSlopeVar;
run;
data all;merge intestimates slopeestimates intslope;
/* could calculate cov using formula
Form 1:  $V[(a_1-a_2)+(b_1-b_2)]=V(a_1-a_2)+V(b_1-b_2)+2Cov[(a_1-a_2),(b_1-b_2)]$ 
So  $Cov[(a_1-a_2),(b_1-b_2)] = (V[(a_1-a_2)+(b_1-b_2)]-V(a_1-a_2)-V(b_1-b_2))/2$ 
Or use this formula

```



```

Form2:V[(a1-a2)-(b1-b2)]=V(a1-a2)+V(b1-b2)-2Cov[(a1-a2),(b1-b2)]
So Cov[(a1-a2),(b1-b2)]=[V(a1-a2)+V(b1-b2)-V[(a1-a2)-(b1-b2)]]/
2*/
covariance=(IntSlopeVar-intvar-slopevar)/2; /*using form 1*/
run;
data all1;set all;
B=2*((InterceptEst*SlopeEst)-(2*ff*Covariance));
D=(B**2)-(4*A*C);
f2sqrt=sqrt(ff*2);
if A>0 then do; Caselower=(-B-sqrt(D))/(2*A);Caselupper =
((-B+sqrt(D))/(2*A));end;
if A<0 and D>0 then do; Case2lower=(-B+sqrt(D))/(2*A);
Case2upper=(-B-sqrt(D))/(2*A);end;
label CaselLower= 'Case I X< ';
label CaselUpper = 'Or X> ';
label Case2Lower = 'Case II';
label Case2Upper = '<X< ';
F_To_Compare_to_T=sqrt(ff*2);
label F_To_Compare_to_T = 'Scheffe Constant For Comparing to
T-Value';;
run;
proc print data=all1 noobs label;var A B C D CaselLower CaselUpper
Case2Lower Case2Upper F_To_Compare_to_T;run;
/*COMPARING CATHOLIC GROUP CURVE TO A PUBLIC SCHOOL 1296*/
data fcaseGLIM;
ff=finv(0.95,2,7023);
match=1;
run;
data estimates2;set estimatesGLIM; match=1;
data estimates3;merge estimates2 fcaseGLIM;by match;run;
data covsub;set estimates3;
if label='Cov' then do;
intslopepred=estimate;
intslopese=stderr;
intslopevar=stderr**2;output;end;
a=1;
keep intslopepred intslopese intslopevar;
data intsubestimates;set estimates3;
if label='int' then do;InterceptEst=estimate;IntVar=StdErr**2;
C = (InterceptEst**2)-(2*ff*IntVar);output;end;
keep C InterceptEst ff IntVar;
run;
data slopesubestimates;set estimates3;
if label='slope' then do;SlopeEst = estimate;SlopeVar=StdErr**2;
slopesq=SlopeEst**2;
A = (slopesq)-(2*ff*SlopeVar);output;end;
keep A SlopeEst ff SlopeVar;
data allsub3;merge intsubestimates slopesubestimates covsub;
run;
proc print data=allsub3;run;
data allsub3_1;set allsub3;
/* could calcuate cov using formula
Form 1: V[(a1-a2)+(b1-b2)]=V(a1-a2)+V(b1-b2)+2Cov[(a1-a2),(b1-

```

```

b2)]
So Cov[(a1-a2),(b1-b2)] =(V[(a1-a2)+(b1-b2)]-V(a1-a2)-V(b1-b2))/2
Or use this formula
Form 2: V[(a1-a2)-(b1-b2)]=V(a1-a2)+V(b1-b2)-2Cov[(a1-a2),(b1-
b2)]
So Cov[(a1-a2),(b1-b2)]=[V(a1-a2)+V(b1-b2)-V[(a1-a2)-(b1-b2)]]/
2*/
covariance=(intslopevar-intvar-slopevar)/2; /*using form 1*/
Bb=(InterceptEst*SlopeEst)-(2*ff*Covariance);
B=2*Bb;
D=(B**2)-(4*A*C);
if A>0 then do; Caselower=(-B-sqrt(D))/(2*A);Caselupper = ((-
B+sqrt(D))/(2*A));end;
if A<0 and D>0 then do; Case2lower=(-B+sqrt(D))/(2*A);
Case2upper=(-B-sqrt(D))/(2*A);end;
F_To_Compare_to_T=sqrt(ff*2);
label CaselLower= 'Case I X< ';
label CaselUpper = 'Or X>';
label Case2Lower = 'Case II';
label Case2Upper = '<X< ';
label F_To_Compare_to_T = 'Scheffe Constant For Comparing to T-
Value';
run;
proc print data=allsub3_1 noobs label;var A B C D CaselLower CaselUpper
Case2Lower Case2Upper F_To_Compare_to_T;run;

```

FIGURE 1.

The estimated difference (Catholic minus Public high schools) in mathematic (Math) achievement as a function of relative student SES. Differences above 0 indicate Catholic sector better, otherwise public schools. Broken lines represent 95% confidence band using Scheffé's method.

FIGURE 2.

The estimated odds ratio (solid line) of high mathematics achievement for the Catholic sector compared to public school 1296 as a function of relative student SES. Odds ratio above 1 indicates (the odds of) higher mathematics achievement score in Catholic schools compared with public school 1296, otherwise public. Broken lines represent the 95% confidence band using Scheffé's method.

Intercepts and Slopes Outcome Model of Math Achievement Score: Continuous and Normally Distributed Outcome^a

TABLE 1

Fixed Effects	Coefficient	Standard Error	T Value	Degrees of Freedom (v)	P Value
Model for school means (β_0)					
Intercept (γ_{00})	12.113	.1988	60.93	157	<.0001
Mean SES (γ_{01})	5.3391	.3693	14.46	157	<.0001
Sector (γ_{02})	1.2167	.3064	3.97	157	<.0001
Model for SES-achievement (β_{1j})					
Intercept (γ_{10})	2.9388	.1551	18.95	7022	<.0001
Mean SES (γ_{11})	1.0389	.2989	3.48	7022	.0005
Sector (γ_{12})	-1.6426	.2398	-6.85	7022	<.0001

^aBased on restricted maximum likelihood from SAS Procedure MIXED version 9.1.3.

TABLE 2
Intercepts and Slopes Outcome Model of High Math Achievement Score: Dichotomous^b and Bernoulli Distributed Outcome^a

Fixed Effects	Coefficient	Standard Error	T Value	Degrees of Freedom (v)	P Value
Model for School Means (β_0)					
Intercept (γ_{00})	-0.1800	.06233	-2.89	157	.0044
Mean SES (γ_{01})	1.5195	.1186	12.81	157	<.0001
Sector (γ_{02})	0.3194	.09516	3.36	157	<.0001
Model for SES-Achievement ^c (β_{1j})					
Intercept (γ_{10})	0.8008	.05496	14.57	7023	<.0001
Sector (γ_{12})	-0.425	.07981	-5.32	7023	<.0001

^aBased on pseudolikelihood method from SAS procedure GLIMIX version 9.1.3.
^bLevels of the mathematics score were dichotomized as high (≥ 13.1) and low (<13.1). The cutoff was determined by the median of the distribution of mathematics scores among all 7,185 students.
^cMean SES $\hat{\gamma}_{11}$ was no longer statistically significant and it was subsequently removed from the model.