United States
Department of
Agriculture

Forest Service

Southern
Research Station

e-Research Note
SRS–17

# Testing Hypotheses for Differences Between Linear Regression Lines

**Stanley J. Zarnoch**

May 2009

**Abstract**

Five hypotheses are identified for testing differences between simple linear regression lines. The distinctions between these hypotheses are based on a priori assumptions and illustrated with full and reduced models. The contrast approach is presented as an easy and complete method for testing for overall differences between the regressions and for making pairwise comparisons. Use of the Bonferroni adjustment to ensure a desired experimentwise type I error rate is emphasized. SAS software is used to illustrate application of these concepts to an artificial simulated dataset. The SAS code is provided for each of the five hypotheses and for the contrasts for the general test and all possible specific tests.

**Keywords**: Contrasts, dummy variables, F-test, intercept, SAS, slope, test of conditional error.

## Introduction

Researchers who compare regression lines and corresponding model parameters are often not fully aware of the specific hypothesis being tested. Similarly, problems may arise in the formulation of the correct model or hypothesis test in the statistical software package being used by the researcher. Most statistical references such as Draper and Smith (1981) and Milliken and Johnson (1984) formulate the problem as a test of conditional error based on full and reduced models and specify an F-test upon which rejection of the hypothesis is based. In this approach, the populations for the regressions are considered classes of an independent variable, dummy variables are then defined, and a regression is developed with the dummy variables and their interactions with the independent variable.

The general linear model (GLM) is often used to test for differences between means, as for example when the analysis of variance is used in data analysis for designed studies. However, the GLM has a much wider application and can also be used to compare regression lines. The GLM allows for the use of both qualitative and quantitative predictor variables. Quantitative variables are continuous values such as diameter at breast height, tree height, age, or temperature. However, many predictor variables in the natural resources field are qualitative. Examples of qualitative variables include tree class (dominant, codominant); species (loblolly, slash, longleaf); and cover type (clearcut, shelterwood, seed tree).

Fraedrich and others (1994) compared linear regressions of slash pine cone specific gravity and moisture content on collection date between families of slash pine grown in a seed orchard. Multiple comparison tests were applied to the slopes to examine specific differences between the families. Murthy and others (1997) compared the intercepts and slopes of light-saturated net photosynthesis response functions over time for loblolly pine grown under various combinations of moisture, nutrient, and $CO_2$ treatments.

The exact form of the regression model used in testing depends on the specific hypothesis under consideration. There are five alternative hypotheses to be considered in this kind of work, each based on a specific formulation of the regression model (table 1). The researcher must be careful to select the hypothesis that is pertinent to the question under study. The objectives of this paper are to: (1) present the underlying statistical methodology for developing the full and reduced models for testing simple linear regressions,

Mathematical Statistician, U.S. Forest Service, Southern Research Station, Asheville, NC 28804.

(2) define the five hypotheses for testing differences between regression lines, (3) present three alternative methods for performing the analysis while controlling experimentwise type I error by means of the Bonferroni approach, and (4) provide SAS code that shows in detail how these five hypotheses are formulated.

## Full and Reduced Models

The methodology for testing hypotheses for differences between simple linear regressions will be illustrated assuming $r = 3$ regression lines are to be compared, with extensions to higher or lower dimensions of $r$ relatively straightforward. Suppose a researcher is interested in comparing regression lines between three treatments ($r = 3$). The general model can be defined as

$$y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i \quad (1)$$

where

$y_i$ = the i[th] observation for the dependent variable y, $i = 1, 2, 3, \ldots, n$,

$x_i$ = the i[th] observation for the independent variable x, $i = 1, 2, 3, \ldots, n$,

$n$ = number of ($x_i, y_i$) paired observations,

$Z_{1i}$ = 1 if $y_i$ and $x_i$ are from treatment 1
0 if $y_i$ and $x_i$ are not from treatment 1,

$Z_{2i}$ = 1 if $y_i$ and $x_i$ are from treatment 2
0 if $y_i$ and $x_i$ are not from treatment 2, and

$\varepsilon_i$ = the residual error for observation i (assumed normal with homogeneous variance for all regressions where $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$).

**Table 1—The five hypotheses used for testing differences between regression lines (all hypotheses assume the usual regression assumptions of normality, independence and homogeneity of variances)**

| Hypothesis | Assumptions | Model (full and reduced) | $H_0$ | $H_1$ |
|---|---|---|---|---|
| 1. All slopes are equal | All intercepts are equal but unknown | $y_i = \alpha_0 + \alpha_1 x_i + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$<br>$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$ | $\gamma_1 = \gamma_2 = 0$ | At least one $\gamma_1$ or $\gamma_2 \neq 0$ |
| 2. All slopes are equal | All intercepts are not necessarily equal but unknown | $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$<br>$y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \varepsilon_i$ | $\gamma_1 = \gamma_2 = 0$ | At least one $\gamma_1$ or $\gamma_2 \neq 0$ |
| 3. All intercepts are equal | All slopes are equal but unknown | $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \varepsilon_i$<br>$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$ | $\beta_1 = \beta_2 = 0$ | At least one $\beta_1$ or $\beta_2 \neq 0$ |
| 4. All intercepts are equal | All slopes are not necessarily equal but unknown | $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$<br>$y_i = \alpha_0 + \alpha_1 x_i + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$ | $\beta_1 = \beta_2 = 0$ | At least one $\beta_1$ or $\beta_2 \neq 0$ |
| 5. All intercepts are equal and all slopes are equal | None | $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$<br>$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$ | $\gamma_1 = \gamma_2 = 0$<br>$\beta_1 = \beta_2 = 0$ | At least one $\beta_1$, $\beta_2$, $\gamma_1$ or $\gamma_2 \neq 0$ |

The regression parameters to be estimated from the data are $\alpha_0$, $\alpha_1$, $\beta_1$, $\beta_2$, $\gamma_1$, and $\gamma_2$. The dummy variables are $Z_{1i}$ and $Z_{2i}$. Equation 1 defines all three treatment regression lines in terms of the third treatment regression line which serves as a baseline for the other treatment regressions. The parameters $\alpha_0$ and $\alpha_1$ are known as the intercept and slope parameters, respectively, and correspond to the third treatment regression. The parameters $\beta_1$ and $\beta_2$ correspond to deviations from the intercept ($\alpha_0$) for treatments 1 and 2, respectively. Similarly, the parameters $\gamma_1$ and $\gamma_2$ correspond to deviations from the slope ($\alpha_1$) for treatments 1 and 2, respectively. Thus, for treatment = 3, $Z_{1i} = 0$ and $Z_{2i} = 0$, and the regression equation is given as

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i \qquad (2)$$

For treatment = 1, $Z_{1i} = 1$ and $Z_{2i} = 0$, and the regression equation is given as

$$y_i = \alpha_0 + \alpha_1 x_i + \beta_1 + \gamma_1 x_i + \varepsilon_i = \left(\alpha_0 + \beta_1\right) + \left(\alpha_1 + \gamma_1\right) x_i + \varepsilon_i \qquad (3)$$

where

($\alpha_0 + \beta_1$) and ($\alpha_1 + \gamma_1$) correspond to the intercept and slope for treatment = 1. Thus, $\beta_1$ and $\gamma_1$ are the deviations (increase or decrease) in the intercept and slope, respectively, for treatment = 1 as compared to treatment = 3. Similarly, treatment = 2 is defined where $Z_{1i} = 0$ and $Z_{2i} = 1$ yielding

$$y_i = \alpha_0 + \alpha_1 x_i + \beta_2 + \gamma_2 x_i + \varepsilon_i = \left(\alpha_0 + \beta_2\right) + \left(\alpha_1 + \gamma_2\right) x_i + \varepsilon_i \qquad (4)$$

where

($\alpha_0 + \beta_2$) and ($\alpha_1 + \gamma_2$) correspond to the intercept and slope for treatment = 2. Thus, $\beta_2$ and $\gamma_2$ are the deviations (increase or decrease) in the intercept and slope, respectively, for treatment = 2 as compared to treatment = 3.

The extension to the general case of $r$ treatment regression lines results in a general model consisting of $r-1$ dummy variables structured analogously to those in equation 1. If this is extended to the multiple linear regression where there are $p$ independent $x$ variables, there would be a total of

$$\left(p+1\right) + \left(r-1\right)\left(p+1\right) = r\left(p+1\right) \qquad (5)$$

parameters in the multiple linear regression model.

## Hypotheses

The exact methodology for comparing regression lines depends on the specific hypothesis that the researcher is testing. The hypotheses under consideration in this paper can be developed from equation 1 by comparing full and reduced models. The hypotheses and a priori assumptions about the intercepts or slopes, or both, will define the full model. All appropriate parameters and associated dummy variables must be in the full model. Depending on the assumptions, some parameters of model (1) may not be in the full model. In contrast, the reduced model is formed by allowing the parameters in question to take the values specified in the null hypothesis $H_0$. Table 1 lists the five hypotheses to be tested, corresponding a priori assumptions and the full and reduced models used to test the hypotheses under consideration. A more thorough review of the hypotheses to be tested is now presented.

Hypothesis 1: all slopes are equal

Assumptions: all intercepts are equal and all are unknown (implies a single common intercept $(\alpha_0)$ and, thus, $\beta_1 = \beta_2 = 0$)

Full model: $y_i = \alpha_0 + \alpha_1 x_i + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$

Reduced model: $y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$

Statistical hypothesis: $H_0$: $\gamma_1 = \gamma_2 = 0$

$H_1$: at least one $\gamma_1$ or $\gamma_2 \neq 0$

Example: To investigate the effect of three fertilizers (qualitative variable) on the total biomass growth $(y_i)$ of young loblolly pine, 1-year-old, uniform, nursery grown seedlings were randomly assigned to the three fertilizers. Each month $(x_i)$, for a year, a tree was removed from each treatment and its biomass measured. It was desired to test the hypothesis that the three rates of height growth (slopes) associated with the fertilizers were equal, assuming that the initial biomasses (intercepts) were identical for the three fertilizers.

Hypothesis 2: all slopes are equal

Assumptions: all intercepts are not necessarily equal and all are unknown

Full model: $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$

Reduced model: $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \varepsilon_i$

Statistical hypothesis: $H_0$: $\gamma_1 = \gamma_2 = 0$

$H_1$: at least one $\gamma_1$ or $\gamma_2 \neq 0$

Example: To investigate the effect of three fertilizers on the growth rate of young loblolly pine, 1-year-old, nursery grown seedlings were randomly assigned to the three fertilizers. However, the uniformity of the seedlings was in doubt, so it was believed that at least one of the fertilizer treatments might be favored by having larger size seedlings assigned to it. Each month over the course of the growing season a tree was removed from each treatment and measured for biomass. It was desired to test the hypothesis that the three rates of height growth (slopes) were equal, assuming that the initial biomasses (intercepts) were not necessarily identical.

Hypothesis 3: all intercepts are equal

Assumptions: all slopes are equal and all are unknown (implies a common slope $(\alpha_1)$ and, thus, $\gamma_1 = \gamma_2 = 0$)

Full model: $y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \varepsilon_i$

Reduced model: $y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$

Statistical hypothesis: $H_0$: $\beta_1 = \beta_2 = 0$

$H_1$: at least one $\beta_1$ or $\beta_2 \neq 0$

Example: This is analogous to the typical analysis of covariance model (ANCOVA) where the treatment effect is analyzed after adjusting for a covariate. To study the effect of fertilization on loblolly pine growth, researchers randomly assigned seedlings to one of three fertilizers (qualitative variable). The study called for growing the seedlings for 2 years, at which point they would be harvested and weighed for final total biomass $(y_i)$. However, it was believed that initial seedling biomass might affect final seedling biomass in addition to any fertilizer effect. Before planting, the seedlings were weighed for initial biomass, which was treated as a covariate corresponding to $(x_i)$ and used to adjust the treatment estimates of final total biomass. In the classical ANCOVA it is assumed that the slopes are equal across the treatments, and this assumption leads to a test of the treatment intercepts. Conclusions about the intercepts can be extended to the treatment effects because the slopes are assumed equal. However, when the slopes are not necessarily equal, as in hypothesis 4, a much more complex situation occurs which requires caution when extending conclusions about the intercepts to the treatment effects.

Hypothesis 4: all intercepts are equal

Assumptions: all slopes are not necessarily equal and all are unknown

Full model:
$y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$

Reduced model: $y_i = \alpha_0 + \alpha_1 x_i + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$

Statistical hypothesis: $H_0$: $\beta_1 = \beta_2 = 0$

$H_1$: at least one $\beta_1$ or $\beta_2 \neq 0$

Example: Hypothesis 4 is similar to hypothesis 3 except hypothesis 4 does not assume that the slopes of the three treatments are equal. Thus, as explained previously, any conclusions about the intercepts cannot be extended to the treatment effects because the slopes are not assumed equal. This is analogous to an ANCOVA problem with unequal slopes and requires comparing the treatments at a minimum of three values of the covariate which is beyond the scope of this paper. Interested readers are referred to Milliken and Johnson (2002).

Hypothesis 5: all intercepts are equal and all slopes are equal

Assumptions: none (except for the usual normality, independence and homogeneity of variances)

Full model:
$y_i = \alpha_0 + \alpha_1 x_i + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \gamma_1 Z_{1i} x_i + \gamma_2 Z_{2i} x_i + \varepsilon_i$

Reduced model: $y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i$

Statistical hypothesis: $H_0$: $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 0$

$H_1$: at least one $\beta_1$, $\beta_2$, $\gamma_1$ or $\gamma_2 \neq 0$

Example: This hypothesis tests the entire regression line simultaneously to determine if either the intercepts or slopes, or both, are different. This is a joint test for both intercept and slope parameters as opposed to separate tests as defined in the previous four hypotheses. Forest inventory often uses a regression model to predict height $(y_i)$ from diameter at breast height $(x_i)$. For example, a height model for loblolly pine is to be developed for different regions of the Southeastern United States. It is often advantageous to determine whether regional differences exist or whether the data should be pooled, with one common model fitted across all regions. In this situation, both the intercept and slope should be tested simultaneously (the entire regression line). Note that this is not equivalent to testing the hypothesis of equal intercepts (hypothesis 3 or 4) and then testing the hypothesis of equal slopes (hypothesis 1 or 2).

## Methodology

As with most statistical analyses, there are alternative methods for testing regression lines that give identical results, although they may vary in their complexity and completeness in addressing the desired hypothesis test. Three methods that are presented here include the test of conditional error, Type III F-tests, and contrasts. The test of conditional error is very informative because it explains the foundations underlying the hypotheses tests, but it is quite cumbersome if many sets of regression lines are to be tested. The Type III tests are simple to perform using a statistical software package but give little insight into what is actually being tested and are quite limited for some hypotheses when they do not yield the full range of comparison tests. Due to the limitations of these two methods, contrasts are proposed as the method of choice for testing hypotheses about differences between regression lines. They are simple to perform, can test all possible combinations of regression lines and parameters, and make it clearer as to what is actually being tested.

### Test of Conditional Error

Textbooks about regression usually present the test of conditional error to test these hypotheses (Draper and Smith 1981, Milliken and Johnson 1984). This test is based on the full and reduced models' sum of squares error (SSE) and degrees of freedom for error (DFE). These are given as $SSE_F$ and $DFE_F$ and $SSE_R$ and $DFE_R$, for the full and reduced model, respectively. The reduced model contains fewer parameters; hence, $SSE_R$ will always be greater than $SSE_F$ and $DFE_R$ will be greater than $DFE_F$. The F-test is used to compare the sum of squares error for the full and reduced models and is given as

$$F = \frac{\left( \left( SSE_R - SSE_F \right) \big/ \left( DFE_R - DFE_F \right) \right)}{\left( SSE_F \big/ DFE_F \right)} \tag{6}$$

This F-statistic is then compared to F*, known as the upper critical F-value of the F-distribution with significance level $\alpha$ and numerator and denominator degrees of freedom of $DFE_R - DFE_F$ and $DFE_F$, respectively. Two decision rules may be used for rejecting the null hypothesis ($H_0$): (1) if F > F* then reject $H_0$ or (2) if the p-value of the test $\leq \alpha$ then reject $H_0$. All hypotheses for differences between regression lines can be tested using this approach because it is fairly general.

A limitation of this approach is that although the test of conditional error tests if there is a difference between the regression lines, it will not indicate which pairwise differences, if any, exist between the treatments. Given $r$ regression lines (treatments), the total number of pairwise comparisons is

$$s = \binom{r}{2} = \frac{r!}{2!\left(r-2\right)!} = \frac{r\left(r-1\right)\left(r-2\right)\ldots\left(1\right)}{2\left(r-2\right)\left(r-3\right)\ldots\left(1\right)} = \frac{r\left(r-1\right)}{2} \tag{7}$$

One option is to fit all possible $(s)$ full and reduced models, rejecting each null hypothesis using a Bonferroni-corrected alpha level. The Bonferroni correction is a multiple-comparison correction used when several statistical tests are being performed simultaneously. While a given significance level ($\alpha$) may be appropriate for each individual comparison, it is not appropriate for the set of all comparisons. The significance level needs to be lowered to account for the number of comparisons being performed. The Bonferroni significance level is defined as $\alpha / s$, and this controls the experimentwise type I error rate at $\alpha$. However, the size of $s$ increases rapidly as $r$ increases in equation 7, and it may become too burdensome to implement the Bonferroni correction. In addition, this may not be equivalent to testing the desired hypotheses. For instance, if the slopes are being tested for equality assuming a common intercept (hypothesis 1), then each pairwise test will consist of a common intercept for that pair, not a common intercept for the entire set of $r$ regression lines, which is not the original hypothesis.

### Type III F-Tests

Instead of specifying the full and reduced models directly and using the test of conditional error [equation 6] to test a hypothesis, it is often more convenient to simply use the typical Type III F-tests obtained by fitting the full model (SAS Institute Inc. 2004). The F-tests obtained are equivalent to the test of conditional error but caution must be used because the appropriate F-test must be selected on the basis of the hypothesis being tested. In addition, the approach is inadequate when one wants to test the entire regression line simultaneously (hypothesis 5), because it tests the intercepts and the slopes separately. Pairwise comparisons between the first $r-1$ and last regression lines can easily be obtained from the t-tests of the full regression model parameter estimates, if one bears in mind that the equivalent F-test is simply the t statistic squared. However, the other pairwise comparisons, e.g., between the first and second regression lines, are not readily available. The Bonferroni adjustment explained previously can be applied to the pairwise comparisons to control the experimentwise type I error rate at $\alpha$.

### Contrasts

To avoid the complexity and limitations of the test of conditional error and the Type III F-tests, the approach advocated in this paper is to form contrasts that will give correct F-tests for any of the five hypotheses. Multiple degrees of freedom contrasts are used to test the overall hypotheses that there are differences between the intercepts or slopes, or both. In addition, single and multiple degrees of freedom contrasts can be used to test all pairwise comparisons and an experimentwise type I error rate can be established by using a Bonferroni corrected alpha level.

If one wishes to test the overall effect for slope differences in hypotheses 1 and 2, one uses the following contrast, which has two degrees of freedom, to specify the null hypothesis ($H_0$)

$$\begin{bmatrix} 1\gamma_1 + 0\gamma_2 - 1\gamma_3 = 0 \\ 0\gamma_1 + 1\gamma_2 - 1\gamma_3 = 0 \end{bmatrix} \tag{8}$$

The parameter $\gamma_3$ in equation 8 represents the slope for treatment 3 and, when combined with $\gamma_1$ and $\gamma_2$ which were introduced in equation 1, allows one to specifically define the desired contrast in terms of all three treatments. This is of particular importance for defining contrasts in SAS, which overparameterizes the model (see next section for more details on SAS). Similarly, to test the overall effect for intercept differences in hypotheses 3 and 4, one uses the following two degrees of freedom contrast to specify the null hypotheses ($H_0$)

$$\begin{bmatrix} 1\beta_1 + 0\beta_2 - 1\beta_3 = 0 \\ 0\beta_1 + 1\beta_2 - 1\beta_3 = 0 \end{bmatrix} \tag{9}$$

Note that the parameter $\beta_3$ is included to represent the intercept for treatment 3. The testing of the entire regression line in hypothesis 5 requires that the null hypothesis ($H_0$) be tested by combining contrasts (8) and (9) into a four degrees of freedom contrast defined as

$$\begin{bmatrix} 1\gamma_1 + 0\gamma_2 - 1\gamma_3 = 0 \\ 0\gamma_1 + 1\gamma_2 - 1\gamma_3 = 0 \\ 1\beta_1 + 0\beta_2 - 1\beta_3 = 0 \\ 0\beta_1 + 1\beta_2 - 1\beta_3 = 0 \end{bmatrix} \tag{10}$$

It should be noted that contrasts (8), (9), and (10) are not unique and could be formulated differently to yield the same results. Also, it must be emphasized that contrast (10) is a simultaneous test of the slopes and intercepts and is not equivalent to the two separate contrasts (8) and (9). It is quite possible in some situations that contrast (10) may be significant, implying that the regressions are different, while neither contrast (8) nor (9) is significant. This may be due to subtle differences in the slopes and intercepts that are not large enough to be detected by the separate two degrees of freedom contrasts (8) and (9) but are detectable by the more powerful four degrees of freedom contrast (10).

If the contrast for an overall test is significant, the next logical step is to determine which regressions (treatments) are different. This is analogous to the typical analysis of variance problem where a significant treatment effect is followed by a set of multiple comparisons. To determine which regressions have different slopes for hypotheses 1 and 2, the $s = 3$ single degree of freedom contrasts are

$$\text{1 versus 2: } \begin{bmatrix} 1\gamma_1 - 1\gamma_2 + 0\gamma_3 = 0 \end{bmatrix} \tag{11a}$$
$$\text{1 versus 3: } \begin{bmatrix} 1\gamma_1 + 0\gamma_2 - 1\gamma_3 = 0 \end{bmatrix} \tag{11b}$$
$$\text{2 versus 3: } \begin{bmatrix} 0\gamma_1 + 1\gamma_2 - 1\gamma_3 = 0 \end{bmatrix} \tag{11c}$$

To determine which regressions have different intercepts for hypotheses 3 and 4, the $s = 3$ single degree of freedom contrasts are

$$\text{1 versus 2: } \begin{bmatrix} 1\beta_1 - 1\beta_2 + 0\beta_3 = 0 \end{bmatrix} \tag{12a}$$
$$\text{1 versus 3: } \begin{bmatrix} 1\beta_1 + 0\beta_2 - 1\beta_3 = 0 \end{bmatrix} \tag{12b}$$
$$\text{2 versus 3: } \begin{bmatrix} 0\beta_1 + 1\beta_2 - 1\beta_3 = 0 \end{bmatrix} \tag{12c}$$

Although hypothesis 5 appears more complex, it is merely a combination of the previous sets of contrasts (11) and (12). To determine which regression lines (intercept and slope simultaneously) are different, the $s = 3$ two degrees of freedom contrasts are

$$\text{1 versus 2: } \begin{bmatrix} 1\beta_1 - 1\beta_2 + 0\beta_3 = 0 \\ 1\gamma_1 - 1\gamma_2 + 0\gamma_3 = 0 \end{bmatrix} \tag{13a}$$

$$\text{1 versus 3: } \begin{bmatrix} 1\beta_1 + 0\beta_2 - 1\beta_3 = 0 \\ 1\gamma_1 + 0\gamma_2 - 1\gamma_3 = 0 \end{bmatrix} \tag{13b}$$

$$\text{2 versus 3: } \begin{bmatrix} 0\beta_1 + 1\beta_2 - 1\beta_3 = 0 \\ 0\gamma_1 + 1\gamma_2 - 1\gamma_3 = 0 \end{bmatrix} \tag{13c}$$

Also, it must be emphasized that contrasts (13) are simultaneous tests of the slopes and intercepts and are not equivalent to the two separate contrasts (11) and (12). It is quite possible in some situations that contrast (13) may be significant, implying that the regressions are different, while neither contrast (11) nor (12) is significant. This may be due to subtle differences in the slopes and intercepts that are not large enough to be detected by the separate one degree of freedom contrasts (11) and (12) but are detectable by the more powerful two degrees of freedom contrast (13).

Constructing contrasts may at first seem complex, but it is really very easy and gives more complete results than doing the test of conditional error or the Type III F-tests, especially when contrasts are constructed with the help of a statistical software package. Contrasts (8) through (13) can easily be extended to any number of $r$ regressions. For more specific information on how to construct contrasts, see Milliken and Johnson (1984).

### Implementation Using SAS

Although most of the above hypotheses may be tested by comparing the full and reduced models and using the Type III F-tests, the simplest and most complete method is to fit the full model and then specify contrasts that test the appropriate hypotheses. This can be done with procedures available in various statistical packages. Here, for purposes of illustration, we use the SAS PROC GLM (SAS Institute Inc. 2004). Instead of specifying the dummy variables $Z_1$ and $Z_2$, one only needs to define a class variable that contains integer or character variables that are unique to each population. For instance, one can define a variable trt

where

trt = 1 if $x_i$ is from regression 1
trt = 2 if $x_i$ is from regression 2
trt = 3 if $x_i$ is from regression 3

SAS will automatically set up the two dummy variables along with a third one, which overparameterizes the model. The variable trt is then treated like any other class variable in PROC GLM. It is imperative that one understands the parameterization used in PROC GLM, particularly the ordering of the parameters, or else erroneous results may occur. The "e" option on the contrast statement displays the coefficients used in the contrast and is quite helpful for confirming the ordering of the parameters in the contrast. Although this option is given in the following SAS code, to conserve space, the output it yields is not presented in the appendix.

To illustrate how SAS can be used to specify the models and contrasts, an artificial dataset was simulated for the three regressions (appendix). The data consisted of observations generated at the independent variable $x_i = 1,2,3,\ldots,10$ for three regression models defined as treatment 1: $y_i = 1 + 1x_i + \varepsilon_i$, treatment 2: $y_i = 2 + 2x_i + \varepsilon_i$, and treatment 3: $y_i = 2 + 2x_i + \varepsilon_i$ where $\varepsilon_i \sim N\left(0,1\right)$ and $\mathrm{cov}\left(\varepsilon_i, \varepsilon_j\right) = 0$ for $i \neq j$. Note that treatments 2 and 3 represent the identical regression model with the same intercepts and slopes while treatment 1 is substantially different with respect to both parameters. The simulated data contained stochastic variation, so all observations are realistic in that they represent data that vary from their respective models to a certain degree as shown in figure 1. The SAS model for each hypothesis and the appropriate contrasts were formulated in PROC GLM as shown in the following SAS code. Additional information on PROC GLM can be found in the "SAS/STAT[R] 9.1 User's Guide" (SAS Institute Inc. 2004) which explains the statements used in the following code as well as parameterization of the model and construction of contrasts. The first contrast for each hypothesis is a multiple degrees of freedom test for the general hypothesis. The other contrasts perform all possible pairwise comparison tests between the three regression lines and to ensure an $\alpha = 0.05$ experimentwise error, each should use a Bonferroni corrected alpha level of 0.05/3 = 0.0167.

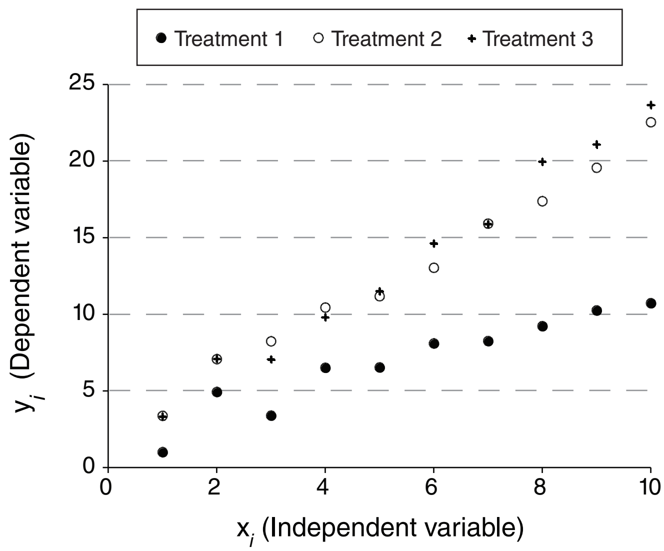


Figure 1—Plot of $y_i$ (dependent variable) versus $x_i$ (independent variable) by treatment for the simulated dataset.

**Hypothesis 1:**

```
title1 'Comparing Regression Lines';
title2 'Hypothesis 1:  Test whether slopes are equal';
title3 'Assume:  Intercepts are equal';
proc glm data=a;
class trt;
model y=x trt*x/ss3 solution;
contrast 'Slope_Only' trt*x 1 0 -1,
                      trt*x 0 1 -1/e;
contrast 'Slope1 vs. Slope2' trt*x 1 -1 0/e;
contrast 'Slope1 vs. Slope3' trt*x 1 0 -1/e;
contrast 'Slope2 vs. Slope3' trt*x 0 1 -1/e;
```

**Hypothesis 2:**

```
title1 'Comparing Regression Lines';
title2 'Hypothesis 2:  Test whether slopes are equal';
title3 'Assume:  Intercepts are not necessarily equal';
proc glm data=a;
class trt;
model y=x trt trt*x/ss3 solution;
contrast 'Slope_Only' trt*x 1 0 -1,
                      trt*x 0 1 -1/e;
contrast 'Slope1 vs. Slope2' trt*x 1 -1 0/e;
contrast 'Slope1 vs. Slope3' trt*x 1 0 -1/e;
contrast 'Slope2 vs. Slope3' trt*x 0 1 -1/e;
```

**Hypothesis 3:**

```
title1 'Comparing Regression Lines';
title2 'Hypothesis 3: Test whether intercepts are equal';
title3 'Assume: Slopes are equal';
proc glm data=a;
class trt;
model y=x trt/ss3 solution;
contrast 'Int_Only' trt 1 0 -1,
                    trt 0 1 -1/e;
contrast 'Int1 vs. Int2' trt 1 -1 0/e;
contrast 'Int1 vs. Int3' trt 1 0 -1/e;
contrast 'Int2 vs. Int3' trt 0 1 -1/e;
```

**Hypothesis 4:**

```
title1 'Comparing Regression Lines';
title2 'Hypothesis 4:  Test whether intercepts are equal';
title3 'Assume:  Slopes are not necessarily equal';
proc glm data=a;
class trt;
model y=x trt trt*x/ss3 solution;
contrast 'Int_Only' trt 1 0 -1,
                    trt 0 1 -1/e;
contrast 'Int1 vs. Int2' trt 1 -1 0/e;
contrast 'Int1 vs. Int3' trt 1 0 -1/e;
contrast 'Int2 vs. Int3' trt 0 1 -1/e;
```

**Hypothesis 5:**

```
title1 'Comparing Regression Lines';
title2 'Hypothesis 5: Test whether intercepts are equal and
slopes are equal simultaneously';
proc glm data=a;
class trt;
model y=x trt trt*x/ss3 solution;
contrast 'Int_Slope' trt 1 0 -1,
                     trt 0 1 -1,
                     trt*x 1 0 -1,
                     trt*x 0 1 -1/e;
contrast 'IntSlope1 vs. IntSlope2' trt 1 -1 0,
                                   trt*x 1 -1 0/e;
contrast 'IntSlope1 vs. IntSlope3' trt 1 0 -1,
                                   trt*x 1 0 -1/e;
contrast 'IntSlope2 vs. IntSlope3' trt 0 1 -1,
                                   trt*x 0 1 -1/e;
```

Output from each of the five hypotheses based on the traditional test of conditional error using the full and reduced model is shown in table 2. In addition, the appendix shows (in bold type) output based on the contrast approach along with the Type III F-tests and parameter t-tests. Note that the tests of conditional error are equivalent to the Type III F-tests and the first contrast of each hypothesis which is a test of the generalized hypothesis. In addition, some of the parameter t-tests are equivalent to some of the other specific contrasts, keeping in mind that the square of the t-statistic is equal to the F-statistic. Remember that these specific tests should be used with the Bonferroni adjustment, which for this example would imply significance only if the p-value is $\leq 0.05/3 = 0.0167$ for an experimentwise error rate of 0.05. It is interesting to note that all specific tests are not available from the parameter t-tests, for instance, the Slope 1 versus Slope 2 contrast for Hypothesis 1. Moreover, neither the general test nor the specific tests for Hypothesis 5 are available from the Type III F-test or the parameter t-tests. The general test can be accomplished by obtaining sequential Type I F-tests of the trt and x*trt effects, combining their sum of squares and degrees of freedom and then forming an appropriate F-test. Although this is not complex, it is labor intensive and is beyond the scope of this paper. It appears evident that the use of contrasts enables the complete set of tests to be performed in an easy and consistent manner.

**Table 2—Results from testing for differences between regression lines (artificial simulated data[a]) using the test of conditional error based on the full and reduced models**

| Components | Hypothesis 1[b] | Hypothesis 2[c] | Hypothesis 3[d] | Hypothesis 4[e] | Hypothesis 5[f] |
|---|---|---|---|---|---|
| $SSE_R$ | 349.71 | 88.42 | 349.71 | 17.69 | 349.71 |
| $SSE_F$ | 17.69 | 16.81 | 88.42 | 16.81 | 16.81 |
| $SSE_R - SSE_F$ | 332.02 | 71.61 | 261.29 | 0.88 | 332.90 |
| $DFE_R$ | 28 | 26 | 28 | 26 | 28 |
| $DFE_F$ | 26 | 24 | 26 | 24 | 24 |
| $DFE_R - DFE_F$ | 2 | 2 | 2 | 2 | 4 |
| F-test | 243.99 | 51.12 | 38.42 | 0.63 | 118.82 |
| F*(0.05) | 3.37 | 3.40 | 3.37 | 3.40 | 2.78 |
| p-value | <0.0001 | <0.0001 | <0.0001 | 0.5412 | <0.0001 |
| Conclusion | Reject $H_0$ | Reject $H_0$ | Reject $H_0$ | Fail to reject $H_0$ | Reject $H_0$ |

[a] The artificial simulated data consisted of observations generated at the independent variable $x_i = 1,2,3,...,10$ for three regression models defined as treatment 1: $y_i = 1 + 1x_i + \varepsilon_i$; treatment 2: $y_i = 2 + 2x_i + \varepsilon_i$; and treatment 3: $y_i = 2 + 2x_i + \varepsilon_i$ where $\varepsilon_i \sim N(0,1)$ and $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$.

[b] Hypothesis 1: all slopes are equal; assumptions: all intercepts are equal and all are unknown.

[c] Hypothesis 2: all slopes are equal; assumptions: all intercepts are not necessarily equal and all are unknown.

[d] Hypothesis 3: all intercepts are equal; assumptions: all slopes are equal and all are unknown.

[e] Hypothesis 4: all intercepts are equal; assumptions: all slopes are not necessarily equal and all are unknown.

[f] Hypothesis 5: all intercepts are equal and all slopes are equal; assumptions: none.

## Conclusion

Researchers who compare regression lines quite often are not aware of the hypothesis being testing. This can result in invalid inferences or less than optimal power for testing the hypothesis. The author identifies five hypotheses for testing differences between several regression lines. The distinctions between these hypotheses were based on a priori assumptions and illustrated by means of the test of conditional error based on the full and reduced model, Type III F-tests, and multiple contrasts. The contrast approach was shown to be the easiest and most complete method for testing for overall and pairwise differences between regression lines.

## Acknowledgments

I would like to express my appreciation to Lewis Jordan from the University of Georgia, and William A. Bechtold and KaDonna C. Randolph, both from the Forest Service, for their reviews of an earlier version of this manuscript.

## Literature Cited

**Draper, N.R.; Smith, H.** 1981. Applied regression analysis. 2d ed. New York: John Wiley. 709 p.

**Fraedrich, S.W.; Miller, T.; Zarnoch, S.J.** 1994. Factors affecting the incidence of black seed rot in slash pine. Canadian Journal of Forest Research. 24: 1717–1725.

**Milliken, G.A.; and Johnson, D.E.** 1984. Analysis of messy data. Volume I: Designed experiments. Van Nostrand Reinhold Co. New York. 473 p.

**Milliken, G.A.; and Johnson, D.E.** 2002. Analysis of messy data. Volume III: Analysis of covariance. Chapman & Hall/CRC. New York. 605 p.

**Murthy, R.; Zarnoch, S.J.; Dougherty, P.M.** 1997. Seasonal tends of light-saturated net photosynthesis and stomatal conductance of loblolly pine trees grown in contrasting environments of nutrition, water and carbon dioxide. Plant, Cell and Environment. 20: 558–568.

**SAS Institute Inc.** 2004. SAS/STAT® 9.1 user's guide. Cary, NC: SAS Institute Inc. 5121 p.

# Appendix[2]

## SAS Output for the Five Hypotheses

```
*  Generate data;
data a;
 seed=100109;
 do trt=1 to 3;
  do x=1 to 10;
   e=rannor(seed);
   if trt eq 1 then y=1+1*x+e;
   if trt eq 2 then y=2+2*x+e;
   if trt eq 3 then y=2+2*x+e;
   output;
  end;
 end;
proc sort data=a;
 by trt;
proc print data=a;
title2 'The Data';
run;
```

The Data

| Obs | trt | x | y |
|---|---|---|---|
| 1 | 1 | 1 | 0.9887 |
| 2 | 1 | 2 | 4.9203 |
| 3 | 1 | 3 | 3.3782 |
| 4 | 1 | 4 | 6.4947 |
| 5 | 1 | 5 | 6.5170 |
| 6 | 1 | 6 | 8.0806 |
| 7 | 1 | 7 | 8.2427 |
| 8 | 1 | 8 | 9.2123 |
| 9 | 1 | 9 | 10.2404 |
| 10 | 1 | 10 | 10.7126 |
| 11 | 2 | 1 | 3.3591 |
| 12 | 2 | 2 | 7.0620 |
| 13 | 2 | 3 | 8.2165 |
| 14 | 2 | 4 | 10.4280 |
| 15 | 2 | 5 | 11.1628 |
| 16 | 2 | 6 | 13.0148 |
| 17 | 2 | 7 | 15.9009 |
| 18 | 2 | 8 | 17.3623 |
| 19 | 2 | 9 | 19.5447 |
| 20 | 2 | 10 | 22.5197 |
| 21 | 3 | 1 | 3.3154 |
| 22 | 3 | 2 | 7.0802 |
| 23 | 3 | 3 | 7.0498 |
| 24 | 3 | 4 | 9.7932 |
| 25 | 3 | 5 | 11.4865 |
| 26 | 3 | 6 | 14.6098 |
| 27 | 3 | 7 | 15.8751 |
| 28 | 3 | 8 | 19.9446 |
| 29 | 3 | 9 | 21.0781 |
| 30 | 3 | 10 | 23.6431 |

[2] The Note at the end of each of the following Hypothesis outputs reflects the way SAS parameterizes the model in PROC GLM. This is a normal message and could be ignored by those not interested in the details of the statistical estimation process.

Comparing Regression Lines
**Hypothesis 1:** Test whether slopes are equal
Assume: Intercepts are equal

The GLM Procedure

Dependent Variable: y

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 1064.113217 | 354.704406 | 521.28 | <.0001 |
| Error | 26 | 17.691695 | 0.680450 | | |
| Corrected Total | 29 | 1081.804912 | | | |

| R-Square | Coeff Var | Root MSE | y Mean |
|---|---|---|---|
| 0.983646 | 7.471094 | 0.824894 | 11.04114 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| x | 1 | 732.0921193 | 732.0921193 | 1075.89 | <.0001 |
| x*trt | 2 | 332.0210973 | 166.0105487 | **243.97** | **<.0001** |

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Slope_Only | 2 | 332.0210973 | 166.0105487 | **243.97** | **<.0001** |
| Slope1 vs. Slope2 | 1 | 218.6880873 | 218.6880873 | **321.39** | **<.0001** |
| Slope1 vs. Slope3 | 1 | 276.0111881 | 276.0111881 | **405.63** | **<.0001** |
| Slope2 vs. Slope3 | 1 | 3.3323705 | 3.3323705 | **4.90** | **0.0359** |

| Parameter | | Estimate | | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | | 1.581860458 | | 0.32534251 | 4.86 | <.0001 |
| x | | 2.162867048 | B | 0.06267025 | 34.51 | <.0001 |
| x*trt | 1 | -1.197424049 | B | 0.05945422 | **-20.14** | **<.0001** |
| x*trt | 2 | -0.131571333 | B | 0.05945422 | **-2.21** | **0.0359** |
| x*trt | 3 | 0.000000000 | B | . | . | . |

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Comparing Regression Lines
**Hypothesis 2:** Test whether slopes are equal
Assume: Intercepts are not necessarily equal

The GLM Procedure

Dependent Variable: y

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 1064.995569 | 212.999114 | 304.12 | <.0001 |
| Error | 24 | 16.809343 | 0.700389 | | |
| Corrected Total | 29 | 1081.804912 | | | |

| R-Square | Coeff Var | Root MSE | y Mean |
|---|---|---|---|
| 0.984462 | 7.579768 | 0.836893 | 11.04114 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| x | 1 | 732.0921193 | 732.0921193 | 1045.26 | <.0001 |
| trt | 2 | 0.8823520 | 0.4411760 | 0.63 | 0.5412 |
| x*trt | 2 | 71.6138566 | 35.8069283 | **51.12** | **<.0001** |

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Slope_Only | 2 | 71.61385657 | 35.80692829 | **51.12** | **<.0001** |
| Slope1 vs. Slope2 | 1 | 40.30032723 | 40.30032723 | **57.54** | **<.0001** |
| Slope1 vs. Slope3 | 1 | 64.32480438 | 64.32480438 | **91.84** | **<.0001** |
| Slope2 vs. Slope3 | 1 | 2.79565326 | 2.79565326 | **3.99** | **0.0572** |

| Parameter | | Estimate | | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | | 1.161643580 | B | 0.57170652 | 2.03 | 0.0534 |
| x | | 2.222898031 | B | 0.09213885 | 24.13 | <.0001 |
| trt | 1 | 0.359317067 | B | 0.80851511 | 0.44 | 0.6607 |
| trt | 2 | 0.901333567 | B | 0.80851511 | 1.11 | 0.2760 |
| trt | 3 | 0.000000000 | B | . | . | . |
| x*trt | 1 | -1.248755059 | B | 0.13030401 | **-9.58** | **<.0001** |
| x*trt | 2 | -0.260333271 | B | 0.13030401 | **-2.00** | **0.0572** |
| x*trt | 3 | 0.000000000 | B | . | . | . |

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Comparing Regression Lines
**Hypothesis 3:** Test whether intercepts are equal
Assume: Slopes are equal

The GLM Procedure

Dependent Variable: y

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 3 | 993.381712 | 331.127237 | 97.36 | <.0001 |
| Error | 26 | 88.423200 | 3.400892 | | |
| Corrected Total | 29 | 1081.804912 | | | |

| R-Square | Coeff Var | Root MSE | y Mean |
|---|---|---|---|
| 0.918263 | 16.70254 | 1.844151 | 11.04114 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| x | 1 | 732.0921193 | 732.0921193 | 215.26 | <.0001 |
| trt | 2 | 261.2895928 | 130.6447964 | **38.41** | **<.0001** |

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Int_Only | 2 | 261.2895928 | 130.6447964 | **38.41** | **<.0001** |
| Int1 vs. Int2 | 1 | 178.7025264 | 178.7025264 | **52.55** | **<.0001** |
| Int1 vs. Int3 | 1 | 211.8247145 | 211.8247145 | **62.29** | **<.0001** |
| Int2 vs. Int3 | 1 | 1.4071482 | 1.4071482 | **0.41** | **0.5257** |

| Parameter | | Estimate | | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | | 3.928305519 | B | 0.86934104 | 4.52 | 0.0001 |
| x | | 1.719868587 | | 0.11722192 | 14.67 | <.0001 |
| trt | 1 | -6.508835757 | B | 0.82472933 | **-7.89** | **<.0001** |
| trt | 2 | -0.530499426 | B | 0.82472933 | **-0.64** | **0.5257** |
| trt | 3 | 0.000000000 | B | . | . | . |

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Comparing Regression Lines

**Hypothesis 4**: Test whether intercepts are equal

Assume: Slopes are not necessarily equal

The GLM Procedure

Dependent Variable: y

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 1064.995569 | 212.999114 | 304.12 | <.0001 |
| Error | 24 | 16.809343 | 0.700389 | | |
| Corrected Total | 29 | 1081.804912 | | | |

| R-Square | Coeff Var | Root MSE | y Mean |
|---|---|---|---|
| 0.984462 | 7.579768 | 0.836893 | 11.04114 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| x | 1 | 732.0921193 | 732.0921193 | 1045.26 | <.0001 |
| trt | 2 | 0.8823520 | 0.4411760 | **0.63** | **0.5412** |
| x*trt | 2 | 71.6138566 | 35.8069283 | 51.12 | <.0001 |

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Int_Only | 2 | 0.88235203 | 0.44117601 | **0.63** | **0.5412** |
| Int1 vs. Int2 | 1 | 0.31476631 | 0.31476631 | **0.45** | **0.5090** |
| Int1 vs. Int3 | 1 | 0.13833081 | 0.13833081 | **0.20** | **0.6607** |
| Int2 vs. Int3 | 1 | 0.87043093 | 0.87043093 | **1.24** | **0.2760** |

| Parameter | | Estimate | | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | | 1.161643580 | B | 0.57170652 | 2.03 | 0.0534 |
| x | | 2.222898031 | B | 0.09213885 | 24.13 | <.0001 |
| trt | 1 | 0.359317067 | B | 0.80851511 | **0.44** | **0.6607** |
| trt | 2 | 0.901333567 | B | 0.80851511 | **1.11** | **0.2760** |
| trt | 3 | 0.000000000 | B | . | . | . |
| x*trt | 1 | -1.248755059 | B | 0.13030401 | -9.58 | <.0001 |
| x*trt | 2 | -0.260333271 | B | 0.13030401 | -2.00 | 0.0572 |
| x*trt | 3 | 0.000000000 | B | . | . | . |

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.

Comparing Regression Lines

**Hypothesis 5:** Test whether intercepts are equal and slopes are equal simultaneously

The GLM Procedure

Dependent Variable: y

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 5 | 1064.995569 | 212.999114 | 304.12 | <.0001 |
| Error | 24 | 16.809343 | 0.700389 | | |
| Corrected Total | 29 | 1081.804912 | | | |

| R-Square | Coeff Var | Root MSE | y Mean |
|---|---|---|---|
| 0.984462 | 7.579768 | 0.836893 | 11.04114 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| x | 1 | 732.0921193 | 732.0921193 | 1045.26 | <.0001 |
| trt | 2 | 0.8823520 | 0.4411760 | 0.63 | 0.5412 |
| x*trt | 2 | 71.6138566 | 35.8069283 | 51.12 | <.0001 |

| Contrast | DF | Contrast SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Int_Slope | 4 | 332.9034493 | 83.2258623 | **118.83** | **<.0001** |
| IntSlope1 vs. IntSlope2 | 2 | 219.0028536 | 109.5014268 | **156.34** | **<.0001** |
| IntSlope1 vs. IntSlope3 | 2 | 276.1495189 | 138.0747595 | **197.14** | **<.0001** |
| IntSlope2 vs. IntSlope3 | 2 | 4.2028015 | 2.1014007 | **3.00** | **0.0687** |

| Parameter | | Estimate | | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | | 1.161643580 | B | 0.57170652 | 2.03 | 0.0534 |
| x | | 2.222898031 | B | 0.09213885 | 24.13 | <.0001 |
| trt | 1 | 0.359317067 | B | 0.80851511 | 0.44 | 0.6607 |
| trt | 2 | 0.901333567 | B | 0.80851511 | 1.11 | 0.2760 |
| trt | 3 | 0.000000000 | B | . | . | . |
| x*trt | 1 | -1.248755059 | B | 0.13030401 | -9.58 | <.0001 |
| x*trt | 2 | -0.260333271 | B | 0.13030401 | -2.00 | 0.0572 |
| x*trt | 3 | 0.000000000 | B | . | . | . |

NOTE: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.