Two Group Straight Line ANCOVA

W. Gregory Alvord

1 Introduction

The simplest case of ANalysis of COVAriance (ANCOVA) is that in which there are two nominal groups and the data within each group can be fit with a straight line. Let y_{ij} be the response for the i^{th} individual ($i = 1, ..., N$) in the j^{th} group ($j = 1, 2$). Then

$$y_{ij} = \alpha_j + \beta_j x_{ij} + e_{ij} \tag{1}$$

where x_{ij} is the covariate for the i^{th} subject ($i = 1, ..., N$) in the j^{th} group ($j = 1, 2$) and e_{ij} is the error (residual) term. The expected value for y_{ij} is

$$E(y_{ij}) = \alpha_j + \beta_j x_{ij}. \tag{2}$$

Note that in this formulation, there is no term for the grand mean, μ.

2 Four Models of Interest

For the two-group straight-line ANCOVA case, four models are of interest: (A) Model A, the full model, which incorporates individual intercepts and individual slopes; (B) Model B, a completely reduced model that incorporates only a single intercept and slope; (C) Model C, a reduced model that incorporates individual intercepts and a common slope; and (D) Model D, a reduced model that incorporates a common intercept, but allows for individual slopes. Therefore, for the simple ANCOVA case in which data arise from two nominal groups, and through which straight lines can be fit, the expected values for models A, B, C, and D, respectively, can be written as:

$$E(y_{ij}) = \alpha_1 + \beta_1 x_{i1} + \alpha_2 + \beta_2 x_{i2}, \tag{3}$$

$$E(y_{ij}) = \alpha + \beta x_{ij}, \tag{4}$$

$$E(y_{ij}) = \alpha_1 + \alpha_2 + \beta x_{ij}, \text{and} \tag{5}$$

$$E(y_{ij}) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2}. \tag{6}$$

In equation (3), it is understood that if the i^{th} subject (case) is in group 1, then $\alpha_2 = 0$ and $\beta_2 = 0$. Alternatively, if the i^{th} subject (case) lies in group 2, then $\alpha_1 = 0$ and $\beta_1 = 0$. Similarly, in equation (5), if the i^{th} case is in group 1, then $\alpha_2 = 0$, and $j = 1$ in the subscript for βx_{ij}. The reverse is also true. In equation (6), if the i^{th} case lies in group 1, then $\beta_2 = 0$, and vice versa.

Figure 1 depicts the four models of interest for this general two-group, straight-line ANCOVA situation.

3 Derivation of Sums of Squares for ANOVA Table

Following Searle (1971, Chapter 3) let $\mathbf{X}$ be the design matrix for a regression model of full rank. The coefficient vector $\mathbf{b}$ has length equal to the column rank of $\mathbf{X}$. The response vector $\mathbf{y}$ has elements y_{ij}, $i = 1, ..., N$. The error vector $\mathbf{e}$ has elements e_{ij}, $i = 1, ..., N$. Then, in matrix notation,

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}, \text{ with } E(\mathbf{y}) = \mathbf{Xb}. \tag{7}$$

The solution for the estimation of parameters $\mathbf{b}$ is

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{y}. \tag{8}$$

In this paper, all design matrices are of full column rank. Therefore, $\mathbf{X}'\mathbf{X}$ is of full rank and $(\mathbf{X}'\mathbf{X})^{-1}$ exists.

The total sum of squares, SST, uncorrected for the mean, is the sum of: (1) the sum of squares due to regression, SSR, and (2) the sum of squares due to error (to the residuals), SSE (Searle, 1971, pp 93-4). That is,

$$\text{SST} = \text{SSR} + \text{SSE} \tag{9}$$

In matrix notation, this identity is written

$$\mathbf{y}'\mathbf{y} = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y} + \mathbf{e}'\mathbf{e} = \hat{\mathbf{b}}'\mathbf{X}'\mathbf{X}\hat{\mathbf{b}} + \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y} \tag{10}$$

where $\mathbf{I}$ is the identity matrix with rank N. This, in turn, yields

$$\text{SSE} = \text{SST - SSR} = \mathbf{e}'\mathbf{e} = \mathbf{y}'\mathbf{y} - \hat{\mathbf{b}}'\mathbf{X}'\mathbf{y}. \tag{11}$$

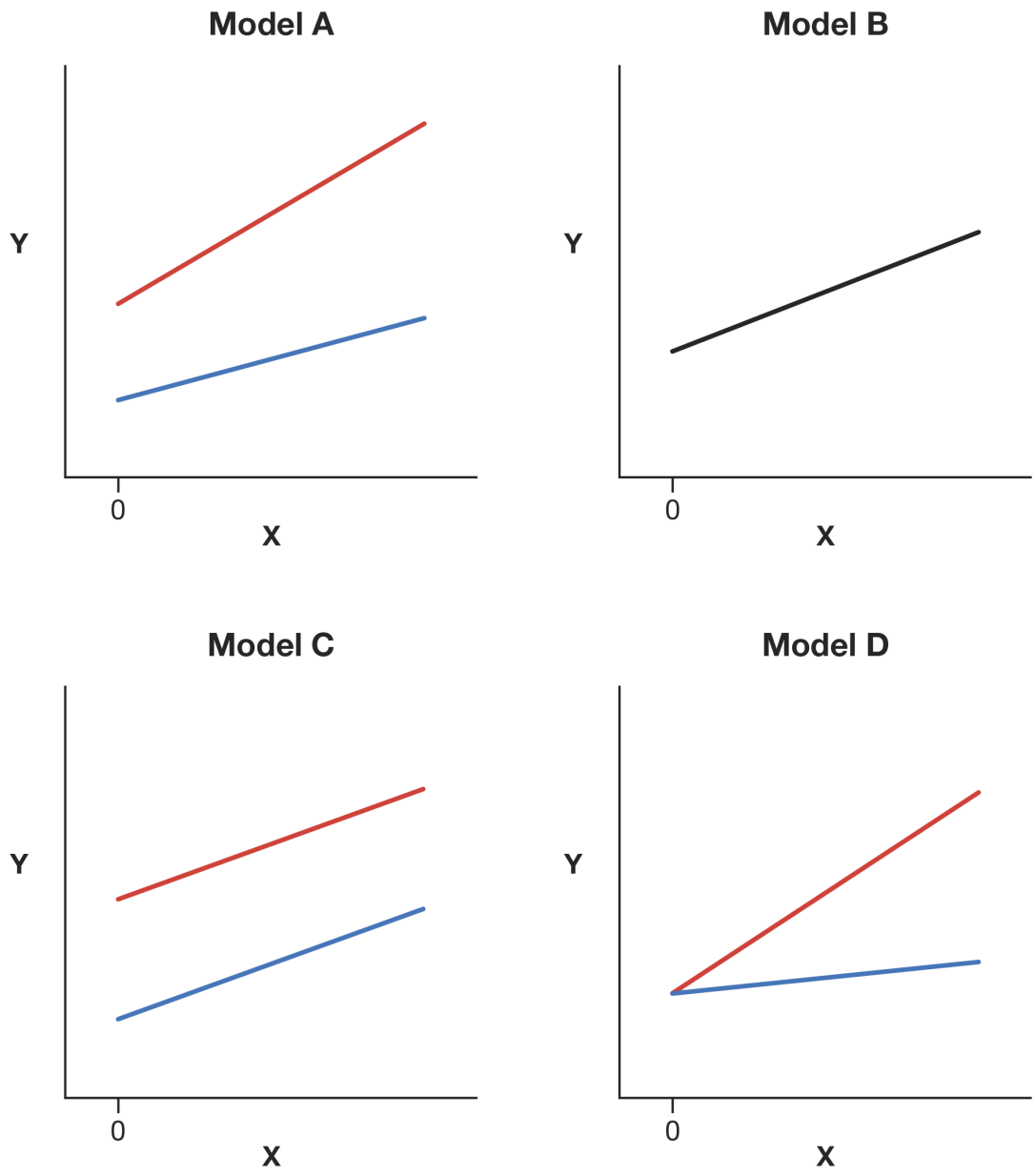

Figure 1: Four Models of Interest: (A) Individual intercepts and individual slopes, (B) Single intercept and single slope, (C) Individual intercepts and single slope, (D) Single intercept and individual slopes.

For any given set of observed data, $\mathbf{y}$, for which we want to fit models A, B, C and D, the total (uncorrected) sum of squares SST, or $\mathbf{y}'\mathbf{y}$, will be the same. However, the design matrices, $\mathbf{X}$, the coefficient vectors, $\mathbf{b}$, and the residuals $\mathbf{e}$ will, in general, differ among models A, B, C and D.

Let $\mathbf{X_A}$, $\mathbf{X_B}$, $\mathbf{X_C}$ and $\mathbf{X_D}$ be the design matrices for models A, B, C and D, respectively. Then, for model A with design matrix $\mathbf{X_A}$, error vector $\mathbf{e_A}$, and coefficient vector $\mathbf{b_A}$, we have

$$\mathbf{X_A} = \begin{bmatrix} 1 & x_{11} & 0 & 0 \\ 1 & x_{21} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n_1 1} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & x_{12} \\ 0 & 0 & 1 & x_{22} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{n_2 2} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ \dots \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_2 2} \end{bmatrix}, \mathbf{e_A} = \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n_1 1} \\ \dots \\ e_{12} \\ e_{22} \\ \vdots \\ e_{n_2 2} \end{bmatrix}, \text{ and } \mathbf{b_A} = \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \end{bmatrix}.$$

where $n_1 + n_2 = N$.

Similarly, for model B, with design matrix $\mathbf{X_B}$, we have

$$\mathbf{X_B} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{n_1 1} \\ \dots & \dots \\ 1 & x_{12} \\ 1 & x_{22} \\ \vdots & \vdots \\ 1 & x_{n_2 2} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ \dots \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_2 2} \end{bmatrix}, \mathbf{e_B} = \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n_1 1} \\ \dots \\ e_{12} \\ e_{22} \\ \vdots \\ e_{n_2 2} \end{bmatrix}, \text{ and } \mathbf{b_B} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

where, again, $n_1 + n_2 = N$.

For model C, with design matrix $\mathbf{X_C}$, we have

$$\mathbf{X_C} = \begin{bmatrix} 1 & 0 & x_{11} \\ 1 & 0 & x_{21} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{n_1 1} \\ \dots & \dots & \dots \\ 0 & 1 & x_{12} \\ 0 & 1 & x_{22} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_{n_2 2} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ \dots \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_2 2} \end{bmatrix}, \mathbf{e_C} = \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n_1 1} \\ \dots \\ e_{12} \\ e_{22} \\ \vdots \\ e_{n_2 2} \end{bmatrix}, \text{ and } \mathbf{b_C} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{bmatrix}.$$

And for model D, with design matrix $\mathbf{X_D}$, we have

$$\mathbf{X_D} = \begin{bmatrix} 1 & x_{11} & 0 \\ 1 & x_{21} & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_{n_1 1} & 0 \\ \dots & \dots & \dots \\ 1 & 0 & x_{12} \\ 1 & 0 & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{n_2 2} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n_1 1} \\ \dots \\ y_{12} \\ y_{22} \\ \vdots \\ y_{n_2 2} \end{bmatrix}, \mathbf{e_D} = \begin{bmatrix} e_{11} \\ e_{21} \\ \vdots \\ e_{n_1 1} \\ \dots \\ e_{12} \\ e_{22} \\ \vdots \\ e_{n_2 2} \end{bmatrix}, \text{ and } \mathbf{b_D} = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

We have explicitly displayed the design matrices, error (residual) vectors, and coefficient vectors, separately, for models A, B, C and D to drive home the fact that they will differ under each model. Model A, which fits four parameters, is the full model. Models B, C, and D are reduced models, which are nested within model A. Taking a_1, b_1, a_2 and b_2 as the least squares estimates of α_1, β_1, α_2 and β_2, respectively, the sum of squares due to regression for model A is

$$\mathbf{SSR_A} = \mathrm{SS}(a_1, b_1, a_2, b_2) = \hat{\mathbf{b_A}}'\mathbf{X_A}'\mathbf{X_A}\hat{\mathbf{b_A}} = \hat{\mathbf{b_A}}'\mathbf{X_A}'\mathbf{y}. \tag{12}$$

Letting a and b represent the estimates for α and β, the sum of squares due to regression for model B is

$$\mathbf{SSR_B} = \mathrm{SS}(a, b) = \hat{\mathbf{b_B}}'\mathbf{X_B}'\mathbf{X_B}\hat{\mathbf{b_B}} = \hat{\mathbf{b_B}}'\mathbf{X_B}'\mathbf{y}. \tag{13}$$

Taking a_1, a_2 and b as the least squares estimates for α_1, α_2 and β, the sum of squares due to regression for model C is

$$\mathbf{SSR_C} = \mathrm{SS}(a_1, a_2, b) = \hat{\mathbf{b_C}}'\mathbf{X_C}'\mathbf{X_C}\hat{\mathbf{b_C}} = \hat{\mathbf{b_C}}'\mathbf{X_C}'\mathbf{y}. \tag{14}$$

Finally, letting a, b_1 and b_2 be the least squares estimates of α, β_1 and β_2, the sum of squares due to regression for model D is

$$\mathbf{SSR_D} = \mathrm{SS}(a, b_1, b_2) = \hat{\mathbf{b_D}}'\mathbf{X_D}'\mathbf{X_D}\hat{\mathbf{b_D}} = \hat{\mathbf{b_D}}'\mathbf{X_D}'\mathbf{y}. \tag{15}$$

As stated above, for any given set of criterion data, $\mathbf{y}$, for which we want to fit models A, B, C and D, the total (uncorrected) sum of squares SST, or $\mathbf{y}'\mathbf{y}$, will remain the same. It is also true that the predictor (regressor) variables, x_{ij}'s, will remain the same, though they are grouped differently in the design matrices for each model. The design matrices, $\mathbf{X}$, the estimated coefficient vectors, $\hat{\mathbf{b}}$, and the residuals $\mathbf{e}$ will, in general, differ among models A, B, C and D. For example, the values of a_1 and a_2 in (12) will *not*, in general, be equal to the values of a_1 and a_2 in (14). The value of a in (13) will not be equal (in general) to a in (15), nor will the value of b in (13) be equal to b in (14).

From (9), we have SST = SSR + SSE. Table 1 shows a skeleton source table that partitions the SST, $\mathbf{y}'\mathbf{y}$, into SSR and SSE for models A, B, C and D. The number of parameters estimated in model A is 4. The numbers of parameters estimated in models B, C and D, respectively, are 2, 3, and 3. The SST = $\mathbf{y}'\mathbf{y}$ does not change for a particular data set being analyzed. Since 4 parameters are estimated in model A, the sum of squares for regression for model A, $\mathbf{SSR_A}$, necessarily (of necessity) must be greater than or equal to the sum of squares due to regression for models B, C, and/or D. That is $\mathbf{SSR_A} \geq \mathbf{SSR_B}$, $\mathbf{SSR_A} \geq \mathbf{SSR_C}$, and $\mathbf{SSR_A} \geq \mathbf{SSR_D}$. At the same time, it must be true that the residual sum of squares for model A, $\mathbf{SSE_A}$, must be less than or equal to the residual sums of squares for models B, C, or D. That is $\mathbf{SSE_A} \leq \mathbf{SSE_B}$, $\mathbf{SSE_A} \leq \mathbf{SSE_C}$, and $\mathbf{SSE_A} \leq \mathbf{SSE_D}$.

Table 1: Partitioning of SST for Models A, B, C and D

Model	No. Parms Estimated	Resid df	SST	SSR	SSE
A	4	$N-4$	$\mathbf{y}'\mathbf{y}$	$\mathbf{SSR_A} = \mathbf{b_A}'\mathbf{X_A}'\mathbf{y}$	$\mathbf{SSE_A} = \mathbf{y}'\mathbf{y} - \mathbf{b_A}'\mathbf{X_A}'\mathbf{y}$
B	2	$N-2$	$\mathbf{y}'\mathbf{y}$	$\mathbf{SSR_B} = \mathbf{b_B}'\mathbf{X_B}'\mathbf{y}$	$\mathbf{SSE_B} = \mathbf{y}'\mathbf{y} - \mathbf{b_B}'\mathbf{X_B}'\mathbf{y}$
C	3	$N-3$	$\mathbf{y}'\mathbf{y}$	$\mathbf{SSR_C} = \mathbf{b_C}'\mathbf{X_C}'\mathbf{y}$	$\mathbf{SSE_C} = \mathbf{y}'\mathbf{y} - \mathbf{b_C}'\mathbf{X_C}'\mathbf{y}$
D	3	$N-3$	$\mathbf{y}'\mathbf{y}$	$\mathbf{SSR_D} = \mathbf{b_D}'\mathbf{X_D}'\mathbf{y}$	$\mathbf{SSE_D} = \mathbf{y}'\mathbf{y} - \mathbf{b_D}'\mathbf{X_D}'\mathbf{y}$

4 The Extra Sum of Squares Principle and Tests for Parameters Being Zero

Recall that $\mathbf{SSR_A} = \text{SS}(a_1, b_1, a_2, b_2)$ to denote the regression sum of squares for the full model, model A. Similarly, for the completely reduced model B, we write $\mathbf{SSR_B} = \text{SS}(a, b)$. The regression sum of squares for the reduced model C is written $\mathbf{SSR_C} = \text{SS}(a_1, a_2, b)$ and the regression sum of squares for the reduced model D is written $\mathbf{SSR_D} = \text{SS}(a, b_1, b_2)$. As stated previously, none of the estimated values in model A, $(a_1, b_1, ...)$, etc., will, in general, be equal to any of the estimates of a_1, b_1, etc., in models B, C or D.

In the regression analysis of the two-group, straight-line ANCOVA setup, there are typically three null hypotheses of interest that arise in connection with comparisons among these four models. These are referred to as the null hypotheses for: (1) *equivalent data sets*, (2) *equivalent slopes*, and (3) *equivalent intercepts*. Under standard assumptions, tests for acceptance or rejection of these null hypotheses can be accomplished using the *reduction in sums of squares* principle (Searle, 1971, pp 99-105, pp 116-20, pp 246-9), the *extra sum of squares* principle (Draper and Smith, 1998, pp 149-51), or the *incremental sum of squares* principle (Fox, 2008, pp 158-9). It can be shown (e.g., Searle, 1971, pp 99-105, pp 246-9; Draper and Smith, pp 38-39) that under assumptions of the equivalence of specific parameters in the models being compared, the expected value of the difference in the regression sums of squares between the full model (estimating p parameters) and a reduced model (estimating q parameters), appropriately modified by their degrees of freedom $(p - q)$ to yield mean squares, are equal to σ^2. In addition, if the errors are normally distributed, then their difference is distributed as $\sigma^2\chi^2_{(p-q)}$ independently of s^2. This means that the mean squares can be compared to σ^2 with an $F(p - q, \nu)$ test, where ν is the number of degrees of freedom on which σ^2 is based, i.e., on the Mean Square Error (MSE), or s^2, for model A.

We discuss the tests for each of these null hypotheses in turn.

4.1 Test for Equivalence of Data Sets

Model A represents the full model. It requires the fitting of four parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$. Model B is a reduced model; it requires the fitting of only two parameters (α, β). The null hypothesis of *equivalent data sets*, is (states) that the data under consideration, which are designated by two nominally distinct groups and which require two intercepts and two slopes in model A, can be more parsimoniously explained by a model containing a single intercept and slope in model B. A test for acceptance or rejection of the null hypothesis of equivalent data sets can be accomplished using the *extra sum of squares* principle. Let the regression sum of squares for model A be denoted as $\mathbf{SSR_A}$ and regression sum of squares for model B be denoted as $\mathbf{SSR_B}$. Then $\mathbf{SSR_A} - \mathbf{SSR_B}$ is the extra sum of squares due to the inclusion of two intercepts and slopes in model A over the inclusion of a

single intercept and slope in model B. $\mathbf{SSR_A}$ is obtained by estimating $p = 4$ parameters, yielding $(N - p) = (N - 4)$ residual degrees of freedom. $\mathbf{SSR_B}$ is obtained by estimating $q = 2$ parameters, thus yielding $(N - q) = (N - 2)$ residual degrees of freedom. The test for $\mathbf{SSR_A} - \mathbf{SSR_B}$, therefore, has $p - q = 4 - 2 = 2$ degrees of freedom. It can be shown (Searle, pp 99-105; Draper and Smith, pp 149-51) that if $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$ then $E\{(\mathbf{SSR_A} - \mathbf{SSR_B})/(p - q = 2)\} = \sigma^2$. In addition, if the errors are normally distributed, then $\mathbf{SSR_A} - \mathbf{SSR_B}$ is distributed as $\sigma^2\chi^2_{(p-q=2)}$ independently of s^2. Hence, the mean square $\mathbf{MSR_{(A-B)}} = \{(\mathbf{SSR_A} - \mathbf{SSR_B})/(p - q = 2)\}$, can be compared to σ^2 with an $F(p - q = 2, \nu)$ test, where ν is the number of degrees of freedom on which σ^2 is based, i.e., on the Mean Square Error (MSE) for model A, designated as $\mathbf{MSE_A}$, or s^2.

In an alternative formulation (Draper and Smith, p. 151), we arrive at the same conclusion by considering the residual sums of squares for each model instead of the regression sums of squares. The quantity $\mathbf{SSE_B} - \mathbf{SSE_A} = \mathbf{SSR_A} - \mathbf{SSR_B}$, and thus, in an obvious notation, $\mathbf{MSE_{(B-A)}} = \mathbf{MSR_{(A-B)}}$, and the F statistic is computed exactly as described above.

4.2 Test for Equivalence of Slopes

Again, model A represents the full model; it requires that four parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ be fit to the data. Model C is a reduced model that requires fitting the data with only three parameters, i.e., with two intercepts and a single slope, which is common to both groups $(\alpha_1, \alpha_2, \beta)$. The null hypothesis of interest here, the hypothesis of *equivalent slopes*, is that the data under consideration, which are designated by two nominally distinct groups, and which require two intercepts and two slopes in model A, can be more parsimoniously explained by a model that requires two distinct intercepts, but only a single slope common to both groups, i.e., model C. A test for acceptance or rejection of the null hypothesis of equivalent slopes can be accomplished using the *extra sum of squares* principle. As before, let the regression sum of squares for model A be denoted as $\mathbf{SSR_A}$ and regression sum of squares for model C be denoted as $\mathbf{SSR_C}$. Then $\mathbf{SSR_A} - \mathbf{SSR_C}$ is the extra sum of squares due to the inclusion of two separate intercepts and slopes in model A over the inclusion of two intercepts and a single slope in model C. Since $\mathbf{SSR_A}$ is fit with $p = 4$ parameters (yielding $(N - p) = (N - 4)$ residual degrees of freedom) and $\mathbf{SSR_C}$ is fit with $q = 3$ parameters (yielding $(N - q) = (N - 3)$ residual degrees of freedom), then the test for $\mathbf{SSR_A} - \mathbf{SSR_C}$ has $p - q = 4 - 3 = 1$ degree of freedom. If $\beta_1 = \beta_2$ then $E\{\mathbf{SSR_A} - \mathbf{SSR_C}/(p - q = 1)\} = \sigma^2$ (Searle, 1971). If the errors are normally distributed, then $\mathbf{SSR_A} - \mathbf{SSR_C}$ is distributed as $\sigma^2\chi^2_{(p-q=1)}$ independently of s^2. Thus, $\{\mathbf{SSR_A} - \mathbf{SSR_C}/(p - q = 1)\}$ can be compared to σ^2 with an $F(p - q = 1, \nu)$ test, where ν is the number of degrees of freedom on which σ^2 is based, i.e., the Mean Square Error (MSE) for model A, designated as $\mathbf{MSE_A}$, or s^2.

Alternatively, we arrive at the same conclusion by considering the residual sums of squares for each model instead of the regression sums of squares (Draper and Smith, p. 151). The quantity $\mathbf{SSE_C} - \mathbf{SSE_A} = \mathbf{SSR_A} - \mathbf{SSR_C}$; thus, $\mathbf{MSE_{(C-A)}} = \mathbf{MSR_{(A-C)}}$, and the F statistic is computed exactly as described above.

4.3 Test for Equivalence of Intercepts

Again, model A represents the full model; four parameters $(\alpha_1, \beta_1, \alpha_2, \beta_2)$ are fit to the data. Model D is a reduced model that fits the data with a single, common intercept and two distinct slopes. It requires a fit using only three parameters $(\alpha, \beta_1, \beta_2)$. The null hypothesis of interest here, the hypothesis of *equivalent intercepts*, is that the data which are designated by two nominally distinct groups, and which require two distinct intercepts and two distinct slopes, can be more parsimoniously explained by a model that requires only a single intercept, but with two distinct slopes. A test for acceptance or rejection of the null hypothesis of equivalent intercepts can again be accomplished using the *extra sum of squares* principle. Let the regression sum of squares for model A be denoted as $\mathbf{SSR_A}$ and the regression sum of squares for model D be denoted as $\mathbf{SSR_D}$. Then $\mathbf{SSR_A} - \mathbf{SSR_D}$ is the extra sum of squares due to the inclusion of two separate intercepts and slopes in model A over the inclusion of a single intercept and two distinct slopes in model D. Since $\mathbf{SSR_A}$ is fit with $p = 4$ parameters (and yields $(N - P) = (N - 4)$ residual degrees of freedom) and $\mathbf{SSR_D}$ is fit with $q = 3$ parameters (and yields $(N - p) = (N - 3)$ residual degrees of freedom), then the test for $\mathbf{SSR_A} - \mathbf{SSR_D}$ has $p - q = 4 - 3 = 1$ degree of freedom. If $\alpha_1 = \alpha_2$ then $E\{\mathbf{SSR_A} - \mathbf{SSR_D}/(p - q = 1)\} = \sigma^2$, and if the errors are normally distributed, then $\mathbf{SSR_A} - \mathbf{SSR_C}$ is distributed as $\sigma^2\chi^2_{(p-q=1)}$ independently of s^2. Hence, $\{\mathbf{SSR_A} - \mathbf{SSR_D}/(p - q = 1)\}$ can be compared to σ^2 with an $F(p - q = 1, \nu)$ test, where ν is the number of degrees of freedom on which σ^2 is based, i.e., the Mean Square Error (MSE) for Model A, designated as $\mathbf{MSE_A}$, or s^2.

Again, we arrive at the same conclusion by considering the residual sums of squares for each model instead of the regression sums of squares. The quantity $\mathbf{SSE_D} - \mathbf{SSE_A} = \mathbf{SSR_A} - \mathbf{SSR_D}$, and thus, $\mathbf{MSE_{(D-A)}} = \mathbf{MSR_{(A-D)}}$, with the F statistic being computed exactly as described above.

4.4 Analyses of Variance

Table 2 shows (displays) a composite ANOVA source table for the tests of differences between model A and reduced models B, C, and D, respectively. The table is divided into four major 'rows', separated by double lines, the first three being further subdivided into two 'sub-rows'. Under the "Source of Variation" column, in the first three major rows, we

list the source of variation as being comprised of two quantities: (1) the variation due to the estimated parameters in the full model, i.e., a_1, b_1, a_2, and b_2, 'over' (2) the variation due to the estimated parameters in the reduced model of interest. The term 'over' in this context means that we determine the variation due to the estimated parameters in the full model A, 'over and above' or 'after' determining the variation due to the estimated parameters in the reduced model of interest, B, C or D. For example, in the second major 'row', which deals with the comparison of models A and C, we describe the source of variation as (a_1, b_1, a_2, b_2) over (a_1, a_2, b). Under the "Sum of Squares" column, in the first three major rows, the *difference* in the sums of squares can be formulated in two different, but equivalent, ways: (1) as the difference in the sums of squares due to *regression*, or (2) as the difference in the sums of squares due to *error* (*residual*). Thus, for example, in the first 'sub-row' of the second major 'row' under the "Sum of Squares" column, we have the $\mathbf{SSR_A} - \mathbf{SSR_C}$. In our ANCOVA setup, the $\mathbf{SSR_A}$ is obtained by fitting four parameters to the data. The $\mathbf{SSR_C}$ is obtained by fitting three parameters to the same data. Of necessity, the $\mathbf{SSR_A} \geq \mathbf{SSR_C}$. In the second sub-row of the second major 'row', we have in the "Sum of Squares" column $\mathbf{SSE_C} - \mathbf{SSE_A}$. Necessarily, the $\mathbf{SSE_C} \geq \mathbf{SSE_A}$, and this value is identically equivalent to $\mathbf{SSR_A} - \mathbf{SSR_C}$. That is, $\mathbf{SSR_A} - \mathbf{SSR_C} \equiv \mathbf{SSE_C} - \mathbf{SSE_A}$. This is due to the fact that both models A and C are of full rank and are being fit to the same data and that SST = SSR + SSE, regardless of the model being fit. (The same reasoning applies when considering the first and third major rows of Table 2.) The final row under "Source of Variation" designates the Residual (or Error) variation for our ANCOVA setup. The residual sum of squares, designated as $\mathbf{SSE_A}$, is equal to the total sum of squares, $\mathbf{y}'\mathbf{y}$, minus the sum of squares due to regression from model A, $\mathbf{SSR_A}$.

The "Degrees of Freedom" column displays the degrees of freedom, *df*, associated with the hypothesis of interest, in addition to the *df* associated with the residuals. For the hypotheses of interest, these are equal to the number of parameters estimated under model A, $Np(A)$, minus the number of parameters estimated under reduced models B, C, and D, respectively. The *df* associated with the difference between model A and model B is $Np(A) - Np(B) = 2$, between model A and model C is $Np(A) - Np(C) = 1$, and between model A and model D is $Np(A) - Np(D) = 1$. The degrees of freedom due to the residuals, ν, is equal the to number of observations, N, minus the number of parameters fit in model A, which is always 4. Hence, $\nu = N - 4$.

The "Mean Square" and "F Statistic" columns complete Table 2. The "Mean Square" column displays the sums of squares divided by their respective degrees of freedom. Searle (1971, pp 99-100, p.174) shows that $\text{SSE}/\sigma^2 \sim \chi^2_{N-r}$, where $r = rank(\mathbf{X})$ for design matrix $\mathbf{X}$. In our application $\mathbf{X} \equiv \mathbf{X_A}$, and it is always true that $r = 4$; hence $\mathbf{SSE_A}/\sigma^2 \sim \chi^2_{N-4}$, specifically, that is, as a *central* χ^2 distribution with $N - 4$ degrees of freedom. This, in turn, implies that $(N-4)\mathbf{MSE_A}/\sigma^2 \sim \chi^2_{N-4}$. The $\mathbf{MSE_A}$ acts as the denominator in the F tests for the three comparisons of models B, C and D, to model A. Searle (1971, p.100, p.111, p.175) further shows how the (remaining) regression sums of squares, those

associated with the hypotheses tests of interest, are distributed as *non-central* χ^2 distributions with their appropriate degrees of freedom (df) and non-centrality parameters λ, denoted as $\chi^{2'}(df, \lambda)$. Having established that the distributions of the *differences* in sums of squares due to regression, i.e., $\mathbf{SSR_A} - \mathbf{SSR_B}$, $\mathbf{SSR_A} - \mathbf{SSR_C}$, and $\mathbf{SSR_A} - \mathbf{SSR_D}$, are distributed as *non-central* $\chi^{2'}(df, \lambda)$ distributions and that the sum of squares due to the residuals, $\mathbf{SSE_A}$, is distributed as a *central* χ^2 distribution, with $N-4$ degrees of freedom, we now have the components for the mean squares to be used in the F statistics. Searle (pp 104-20) shows how the F statistics are distributed as *non-central* F distributions with their appropriate numerator and denominator degrees of freedom, df_1 and df_2, and with their non-centrality parameters, λ, which vary accordingly for the hypothesis being tested, $F'(df_1, df_2, \lambda)$. Under certain null hypotheses, the λ's vanish (i.e., $\lambda = 0$). Hence, these *non-central* F statistics become *central* F statistics, thus providing us with tests for the hypotheses of (1) equivalence of data sets, (2) equivalence of slopes, and (3) equivalence of intercepts (Searle, p. 104). Thus, in our ANCOVA setup, if the non-centrality parameter, $\lambda = 0$, for a particular hypothesis, then $F \sim F_{df_1, N-4}$ and the probability value for the F distribution can be obtained.

Table 2: ANOVA Source Table for Tests of Differences Between Reduced Models B, C, and D Against the Full Model A

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Statistic
(a_1, b_1, a_2, b_2)	$Np(A) - Np(B)$	$\mathbf{SSR_A} - \mathbf{SSR_B}$	$(\mathbf{SSR_A} - \mathbf{SSR_B})/2$	$\mathbf{MSR_{(A-B)}}/\mathbf{MSE_A}$
over (a, b)	$4 - 2 = 2$	$\mathbf{SSE_B} - \mathbf{SSE_A}$	$(\mathbf{SSE_B} - \mathbf{SSE_A})/2$	$\mathbf{MSE_{(B-A)}}/\mathbf{MSE_A}$
(a_1, b_1, a_2, b_2)	$Np(A) - Np(C)$	$\mathbf{SSR_A} - \mathbf{SSR_C}$	$(\mathbf{SSR_A} - \mathbf{SSR_C})/1$	$\mathbf{MSR_{(A-C)}}/\mathbf{MSE_A}$
over (a_1, a_2, b)	$4 - 3 = 1$	$\mathbf{SSE_C} - \mathbf{SSE_A}$	$(\mathbf{SSE_C} - \mathbf{SSE_A})/1$	$\mathbf{MSE_{(C-A)}}/\mathbf{MSE_A}$
(a_1, b_1, a_2, b_2)	$Np(A) - Np(D)$	$\mathbf{SSR_A} - \mathbf{SSR_D}$	$(\mathbf{SSR_A} - \mathbf{SSR_D})/1$	$\mathbf{MSR_{(A-D)}}/\mathbf{MSE_A}$
over (a, b_1, b_2)	$4 - 3 = 1$	$\mathbf{SSE_D} - \mathbf{SSE_A}$	$(\mathbf{SSE_D} - \mathbf{SSE_A})/1$	$\mathbf{MSE_{(D-A)}}/\mathbf{MSE_A}$
Residual	$\nu = N - 4$	$\mathbf{SSE_A} = \mathbf{y'y} - \mathbf{SSR_A}$	$\mathbf{MSE_A} = \mathbf{SSE_A}/\nu$	—

5 Example

Suppose we have 10 observations, two sets of five observations in two groups. Hence, $\mathbf{y}' = \begin{bmatrix} 4 & 7 & 8 & 11 & 15 & | & 14 & 17 & 18 & 21 & 24 \end{bmatrix}$. The | symbol represents the distinction between the two groups. Let let the regressors be $\mathbf{x}' = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & | & 1 & 2 & 3 & 4 & 5 \end{bmatrix}$. When we fit models A, B, C and D to these data, we produce the outcomes reported in Table 3. The total sum of squares, uncorrected for the mean, SST, is $\mathbf{y}'\mathbf{y} = 2301$. This value does not change for the four models in this example. However, the sums of squares due to regression, SSR, and the sums of squares to error (residual), SSE, *do* change in each of the models.

For model A, the sum of squares due to regression (SSR) is $\mathbf{SSR_A} = 2297.40$, while the sum of squares due to error (SSE) is $\mathbf{SSE_A} = 3.60$. These two sums of squares add up to the total sum of squares (SST), which is $\mathbf{y}'\mathbf{y} = 2301$. For model B, the sum of squares due to regression (SSR) is $\mathbf{SSR_B} = 2057.10$, while the sum of squares due to error (SSE) is $\mathbf{SSE_B} = 243.90$. In model C, the sum of squares due to regression (SSR) is $\mathbf{SSR_C} = 2297.20$, while the sum of squares due to error (SSE) is $\mathbf{SSE_C} = 3.80$. And, for model D, the sum of squares due to regression (SSR) is $\mathbf{SSR_D} = 2248.24$, while the sum of squares due to error (SSE) is $\mathbf{SSE_D} = 52.76$. In all four cases, the sums of squares due to regression, SSR, plus the sums of squares due to error, SSE, add up to the total (uncorrected) sum of squares, SST, i.e., $\mathbf{y}'\mathbf{y} = 2301$.

Table 3: ANOVA Source Table for Example Data

Model	No. Parms Estimated	Resid df	SST	SSR	SSE
A	4	6	$\mathbf{y}'\mathbf{y} = 2301$	$\mathbf{SSR_A} = 2297.40$	$\mathbf{SSE_A} = 3.60$
B	2	8	$\mathbf{y}'\mathbf{y} = 2301$	$\mathbf{SSR_B} = 2057.10$	$\mathbf{SSE_B} = 243.90$
C	3	7	$\mathbf{y}'\mathbf{y} = 2301$	$\mathbf{SSR_C} = 2297.20$	$\mathbf{SSE_C} = 3.80$
D	3	7	$\mathbf{y}'\mathbf{y} = 2301$	$\mathbf{SSR_D} = 2248.24$	$\mathbf{SSE_D} = 52.76$

Next, we perform the tests for the three hypotheses of interest. In Table 4, the difference in the sum of squares due to regression for model A and the sum of squares due to regression for model B, $\mathbf{SSR_A} - \mathbf{SSR_B}$, is denoted as $\mathbf{SSR_{(A-B)}}$. Similarly, the difference in the sum of squares due to error for model B and the sum of squares due to error for model A, $\mathbf{SSE_B} - \mathbf{SSE_A}$, and is denoted as $\mathbf{SSE_{(B-A)}}$. These quantities are equal. That is $\mathbf{SSR_{(A-B)}} = \mathbf{SSE_{(B-A)}}$. The degrees of freedom associated with the difference between

model A and model B is $Np(A) - Np(B)$. In this paper it is always true that $Np(A) = 4$ and $Np(B) = 2$. Hence, for the problem considered here, $Np(A) - Np(B) = (4 - 2) = 2$. The difference in the sums of squares for models A and B due to, say, regression is $\mathbf{SSR_{(A-B)}} = 240.30$. This is identical to the the difference in the sums of squares due to error for models A and B, $\mathbf{SSE_{(B-A)}}$. The mean square for the difference between models A and B, $\mathbf{MSR_{(A-B)}}$, is $\mathbf{SSR_{(A-B)}}/2 = 240.30/2 = 120.15$. The F statistic is obtained by dividing the mean square due to regression by the mean square error for model A, $\mathbf{MSR_{(A-B)}}/\mathbf{MSE_A} = 120.15/0.6 = 200.25$.

The difference in the sum of squares due to regression for model A and the sum of squares due to regression for model C, $\mathbf{SSR_A} - \mathbf{SSR_C}$, is denoted as $\mathbf{SSR_{(A-C)}}$. Similarly, the difference in the sum of squares due to error for model C and the sum of squares due to error for model A, $\mathbf{SSE_C} - \mathbf{SSE_A}$, and is denoted as $\mathbf{SSE_{(C-A)}}$. These are equivalent; hence $\mathbf{SSR_{(A-C)}} = \mathbf{SSE_{(C-A)}}$. The degrees of freedom associated with the difference between model A and model C is $Np(A) - Np(C)$. In this paper $Np(A) = 4$ and $Np(C) = 3$. Hence, $Np(A) - Np(C) = (4 - 3) = 1$. The difference in the sums of squares for models A and C due to regression is $\mathbf{SSR_{(A-C)}} = 0.20$, which is identical to the the difference in the sums of squares due to error for models A and C, $\mathbf{SSE_{(C-A)}}$. The mean square for the difference between models A and C, $\mathbf{MSR_{(A-C)}}$, is $\mathbf{SSR_{(A-C)}}/1 = 0.20/1 = 0.20$. The F statistic is obtained by dividing the mean square due to regression by the mean square error for model A, $\mathbf{MSR_{(A-C)}}/\mathbf{MSE_A} = 0.20/0.6 = 0.33$.

A similar process of reasoning takes place in considering the difference in the sum of squares due to regression for model A and the sum of squares due to regression for model D, $\mathbf{SSR_A} - \mathbf{SSR_D}$, denoted by $\mathbf{SSR_{(A-D)}}$. The F statistic is obtained by dividing the mean square due to regression by the mean square error for model A, $\mathbf{MSR_{(A-D)}}/\mathbf{MSE_A} = 49.16/0.6 = 81.94$.

Table 5 simplifies Table 4 and adds probability values (p values) for the hypothesis tests of interest. In Table 5 the hypothesis of *equivalent data sets* is rejected, $p < 0.0001$. The hypothesis of *equivalent slopes* is not rejected, $p = 0.5847$. The hypothesis of *equivalent intercepts* is rejected, $p = 0.0001$. Therefore, as the p value for *equivalent slopes* exceeds 0.05, we conclude that the `eqslo` data are best explained by the model that allows for two statistically distinct intercepts and a common slope, i.e., model C.

Table 4: ANOVA Source Table for Tests of Differences Between
Reduced Models B, C, and D Against the Full Model A with Example Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Statistic
(a_1, b_1, a_2, b_2)	$Np(A) - Np(B)$	$\mathbf{SSR_{(A-B)}} = 240.30$	$\mathbf{MSR_{(A-B)}} = 120.15$	$\mathbf{F_{reg_{(A-B)}}} = 200.25$
over (a, b)	2	$\mathbf{SSE_{(B-A)}} = 240.30$	$\mathbf{MSE_{(B-A)}} = 120.15$	$\mathbf{F_{err_{(B-A)}}} = 200.25$
(a_1, b_1, a_2, b_2)	$Np(A) - Np(C)$	$\mathbf{SSR_{(A-C)}} = 0.20$	$\mathbf{MSR_{(A-C)}} = 0.20$	$\mathbf{F_{reg_{(A-C)}}} = 0.33$
over (a_1, a_2, b)	1	$\mathbf{SSE_{(C-A)}} = 0.20$	$\mathbf{MSE_{(C-A)}} = 0.20$	$\mathbf{F_{err_{(C-A)}}} = 0.33$
(a_1, b_1, a_2, b_2)	$Np(A) - Np(D)$	$\mathbf{SSR_{(A-D)}} = 49.16$	$\mathbf{MSR_{(A-D)}} = 49.16$	$\mathbf{F_{reg_{(A-D)}}} = 81.94$
over (a, b_1, b_2)	1	$\mathbf{SSE_{(D-A)}} = 49.16$	$\mathbf{MSE_{(D-A)}} = 49.16$	$\mathbf{F_{err_{(D-A)}}} = 81.94$
Residual	$\nu = [N - Np(A)]$	$\mathbf{y'y} - \mathbf{SSR_A}$	$\mathbf{MSE_A}$	—
	(10 - 4) = 6	(2301 - 2297.4) = 3.6	= 0.6	—

Table 5: ANOVA Source Table for Tests of Differences Between
Reduced Models B, C, and D Against the Full Model A with Example Data

H_0	df	SS	MS	F	Prob
Equivalent Data Sets	2	240.30	120.15	200.25	< 0.0001
Equivalent Slopes	1	0.20	0.20	0.33	0.5847
Equivalent Intercepts	1	49.16	49.16	81.94	0.0001
Residual	6	3.6	0.6		

6 The sla Package

6.1 Application to the `eqslo` data frame

We examine the example problem with the sla package. The `eqslo` data frame consists of three columns. The first column has class `factor` with two levels. The second column contains the x, or predictor (regressor) variable, and the third column contains the y, or criterion variable. The data in the `eqslo` data frame are contrived to provide statistically different intercepts with a single slope.

```
library(sla)
```

Call the `eqslo` data frame into the workspace and display it.

```
data(eqslo)
eqslo
```

```
##    group x  y
## 1    one 1  4
## 2    one 2  7
## 3    one 3  8
## 4    one 4 11
## 5    one 5 15
## 6    two 1 14
## 7    two 2 17
## 8    two 3 18
## 9    two 4 21
## 10   two 5 24
```

Execute the `sla` function on the `eqslo` data frame. Call this object `obj`.

```
obj <- sla(eqslo)
```

The print function for the object yields

```
obj
```

```
##
## Call:  sla.default(facxy = eqslo)
##
## Coefficients for Models A, B, C and D:
##
## Model A:
## Int_1 Slo_1 Int_2 Slo_2
##   1.2   2.6  11.6   2.4
##
## Model B:
## Com_Int Com_Slo
##     6.4     2.5
##
## Model C:
##   Int_1   Int_2 Com_Slo
##     1.5    11.3     2.5
##
## Model D:
```

```
## Com_Int   Slo_1   Slo_2
##   6.400   1.182   3.818
```

The `print` function for the object `obj` displays the call and the coefficients for models A, B, C, and D, respectively. Under model A, two lines are fit through the data. The estimated intercept and slope for the first group are $a_1 = 1.2$ and $b_1 = 2.6$. The estimated intercept and slope for the second group are $a_2 = 11.6$ and $b_2 = 2.4$. For model B, the estimates for the intercept and slope (for both groups combined) are $a = 6.4$ and $b = 2.5$. For model C, the estimates for the two intercepts are $a_1 = 1.5$ and $a_2 = 11.3$. The estimate for the common slope is $b = 2.5$. Finally, for model D, the estimate for the common intercept is $a = 6.4$, while the estimates for the slopes are, respectively, $b_1 = 1.18$ and $b_2 = 3.82$.

The `summary` function displays the call and two tables. The top table provides a brief description of each model, the number of parameters fit, the residual degrees of freedom, the residual sum of squares and the residual mean square for models A, B, C and D. The bottom table provides descriptions for the three tests: (1) equivalent data sets, (2) equivalent slopes, and (3) equivalent intercepts. It includes the reductions (differences) in the sums of squares between, respectively, models B, C, and D and model A, the F statistics, and the probabilities associated with hypotheses.

```
summary(obj)

##
## Call:  sla.default(facxy = eqslo)
##
## Summary of ANCOVA Tests. . .
##
##   Description of Fits for 4 ANCOVA Models
##
##   Description of Fit Np Res df Res SS Res MS
## 1 Mod A: Ind I,Ind S  4      6   3.60   0.60
## 2 Mod B: Com I,Com S  2      8 243.90  30.49
## 3 Mod C: Ind I,Com S  3      7   3.80   0.54
## 4 Mod D: Com I,Ind S  3      7  52.76   7.54
##
##   ANCOVA Tests: Two Groups/Straight Line Fits
##
##               Test df     SS F Stat   prob
## 1 Ho: Equiv D.Sets  2 240.30 200.25 0.0000
## 2 Ho: Equiv Slopes  1   0.20   0.33 0.5847
## 3 Ho: Equiv Inters  1  49.16  81.94 0.0001
```

In this problem, the data were contrived to yield statistically equivalent slopes with statistically different intercepts. The hypothesis for equivalent data sets is rejected (at the $\alpha = 0.05$ level), p = 0.0000. The hypothesis for equivalent slopes is *not* rejected, p = 0.5847. The hypothesis for equivalent slopes is rejected, p = 0.0001.

The `plot` function plots the data evaluated in the `sla` object, along with the fitted lines for the model specified in the `modelType2Plot` argument. The default selection is `'A'`. For the current problem, we view all four plots with the following code chunk (Figure 2):

```
par(mfrow = c(2, 2))
plot(obj, modelType2Plot = 'A')
plot(obj, modelType2Plot = 'B')
plot(obj, modelType2Plot = 'C')
plot(obj, modelType2Plot = 'D')
```

6.2 Application to the `whiteside` data frame

We next apply the sla package to the `whiteside` data frame discussed in Section 6.1 of *Modern Applied Statistics with S* (Venables and Ripley, 2002, pp 139-44). The `whiteside` data frame consists of three columns. The first column is a `factor` variable, `Insul`, at two levels, `Before` and `After`. The second column contains the predictor (regressor) variable, `Temp`, and the third column contains the criterion variable, `Gas`.

```
library(sla)
library(MASS)
data(whiteside)
```

Apply the `sla` function to the `whiteside` data frame; call this object `wsobj`.

```
wsobj <- sla(whiteside)
```

The print function for the object yields

```
wsobj

##
## Call:  sla.default(facxy = whiteside)
##
## Coefficients for Models A, B, C and D:
##
## Model A:
```

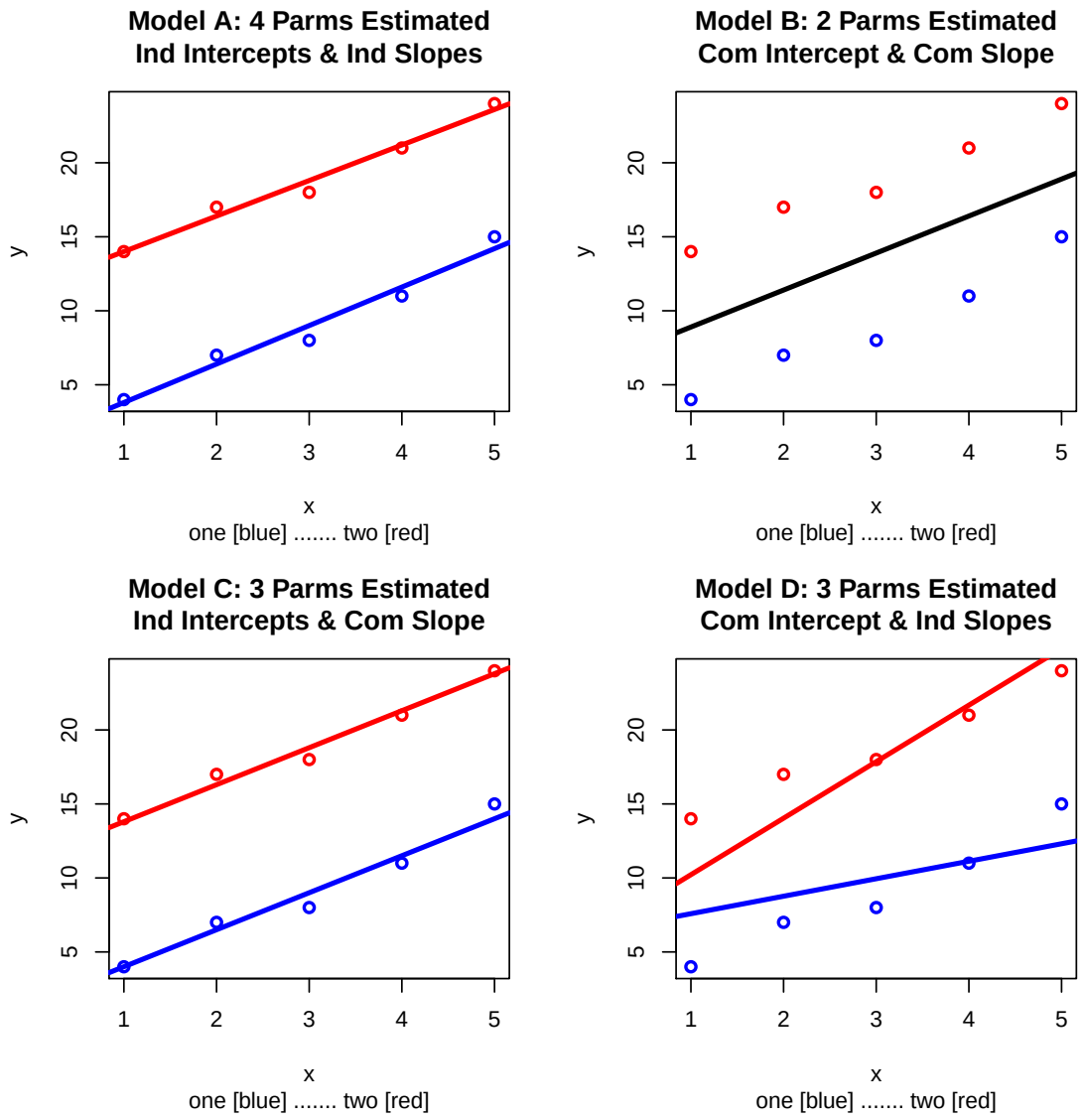

Figure 2: Four models of interest fit to the eqslo data: (A) Individual intercepts and individual slopes, (B) Single intercept and single slope, (C) Individual intercepts and single slope, (D) Single intercept and individual slopes.

```
##   Int_1   Slo_1   Int_2   Slo_2
##  6.8538 -0.3932  4.7238 -0.2779
##
## Model B:
## Com_Int Com_Slo
##  5.4862 -0.2902
##
## Model C:
##   Int_1   Int_2 Com_Slo
##  6.5513  4.9861 -0.3367
##
## Model D:
## Com_Int   Slo_1   Slo_2
##  5.6398 -0.2156 -0.4320
```

The `print` function for the object `wsobj` displays the call and the coefficients for models A, B, C, and D, respectively. Under model A, two unconstrained lines are fit through the data. The estimated intercept and slope for the `Before` group are $a_1 = 6.854$ and $b_1 = -0.393$, while the estimated intercept and slope for the `After` group are $a_2 = 4.724$ and $b_2 = -0.278$.

Greater detail regarding various aspects of the `wsobj` object can be obtained by interogating its attributes.

```
attributes(wsobj)
```

```
## $names
##  [1] "Call"                          "INPUT.df"
##  [3] "Summary of Input Data Frame" "Mod.A"
##  [5] "Mod.B"                         "Mod.C"
##  [7] "Mod.D"                         "Fit.Table"
##  [9] "Test.Table"                    "Fit.Table.Pretty"
## [11] "Test.Table.Pretty"
##
## $class
## [1] "sla"
```

For example, suppose we are interested in further exploring a summary of the full model, i.e., model A. The following command produces essentially the same output as that obtained by Venables and Ripley (p. 142) for their `gasBA` object.

```
summary(wsobj$Mod.A)

##
## Call:
## lm(formula = y ~ X.A - 1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9780 -0.1801  0.0376  0.2093  0.6380
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## X.Ai1    6.8538     0.1360    50.4   <2e-16
## X.Ai1x  -0.3932     0.0225   -17.5   <2e-16
## X.Ai2    4.7238     0.1181    40.0   <2e-16
## X.Ai2x  -0.2779     0.0229   -12.1   <2e-16
##
## Residual standard error: 0.323 on 52 degrees of freedom
## Multiple R-squared:  0.995,Adjusted R-squared:  0.994
## F-statistic: 2.39e+03 on 4 and 52 DF,  p-value: <2e-16
```

The `summary` function for this object once again displays the call and two tables. The top table (essentially `wsobj$Fit.Table.Pretty`) provides a brief description of each model, the number of parameters fit, the residual degrees of freedom, the residual sum of squares and the residual mean square for models A, B, C and D. The bottom table (essentially `wsobj$Test.Table.Pretty`) provides description for the three tests: (1) equivalent data sets, (2) equivalent slopes, and (3) equivalent intercepts. It includes the reductions (differences) in the sums of squares between, respectively, models B, C, and D and model A, the F statistics, and the probabilities associated with hypotheses.

```
summary(wsobj)

##
## Call:  sla.default(facxy = whiteside)
##
## Summary of ANCOVA Tests. . .
##
##   Description of Fits for 4 ANCOVA Models
##
##   Description of Fit Np Res df Res SS Res MS
## 1 Mod A: Ind I,Ind S  4     52   5.43   0.10
```

```
## 2 Mod B: Com I,Com S  2     54  39.99   0.74
## 3 Mod C: Ind I,Com S  3     53   6.77   0.13
## 4 Mod D: Com I,Ind S  3     53  20.02   0.38
##
##   ANCOVA Tests: Two Groups/Straight Line Fits
##
##               Test df    SS F Stat  prob
## 1 Ho: Equiv D.Sets  2 34.57 165.67 0e+00
## 2 Ho: Equiv Slopes  1  1.35  12.89 7e-04
## 3 Ho: Equiv Inters  1 14.59 139.88 0e+00
```

For the `whiteside` data, the null hypotheses for equivalent (1) data sets, (2) slopes, and (3) intercepts are all rejected. The hypothesis for equivalent data sets is rejected, $p << 0.0000$. The hypothesis for equivalent slopes is also rejected, $p = 0.0007307$. The following command, which makes use of two of the attributes of `wsobj`, reproduces Venables and Ripley's (2002) results on page 143.

```
anova(wsobj$Mod.C, wsobj$Mod.A)

## Analysis of Variance Table
##
## Model 1: y ~ X.C - 1
## Model 2: y ~ X.A - 1
##   Res.Df  RSS Df Sum of Sq    F  Pr(>F)
## 1     53 6.77
## 2     52 5.43  1      1.34 12.9 0.00073
```

Finally, the hypothesis of equivalent intercepts is rejected, $p << 0.0000$. In summary, we conclude that model A, the full model, is the best-fitting model for the `whiteside` data. We agree with the conclusion of Venables and Ripley (p. 143) that ". . . separate slopes are indeed necessary." A plot of model A is obtained with command:

```
plot(wsobj, mod = 'A')
```

Figure 3 displays the `whiteside` data with the fitted regression lines for model A.

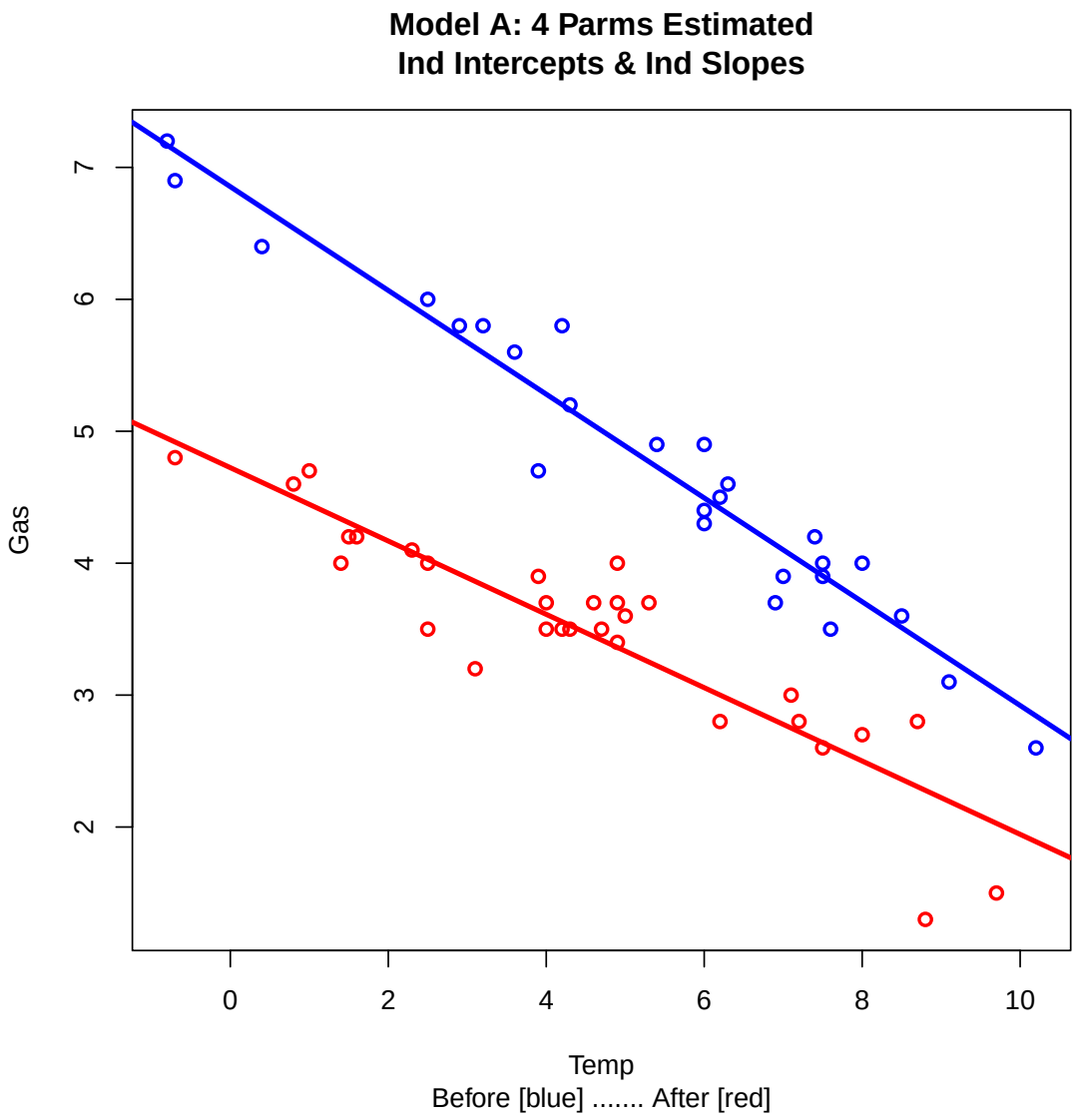

Figure 3: Model A: Individual intercepts and individual slopes for the whiteside data.

References

[1] Dalgaard P (2002) *Introductory Statistics with R*, Springer.

[2] Draper NR and Smith H (1998) *Applied Regression Analysis*, 3rd ed. Wiley.

[3] Fox J (2008) *Applied Regression Analysis and General Linear Models*, 2nd ed. Sage.

[4] Fox J and Weisberg S (2011) *An R Companion to Applied Regression*, 2nd ed. Sage.

[5] Searle SR (1971) *Linear Models*, Wiley.

[6] Venables WN and Ripley BD (2002) *Modern Applied Statistics with S*, 4th ed. Springer