

Chase Joyner

801 Homework 5

November 10, 2015

Problem 1:

In testing a reduced model $Y = X_0\gamma + e$ against a full model $Y = X\beta + e$, what linear parametric function of the parameters is being tested?

Solution: Since we are testing a reduced versus full model, we have $C(X_0) \subset C(X)$. Therefore, there exists a matrix U such that $X_0 = XU$. Then, it follows that $X_0\gamma = XU\gamma$, i.e. $X_0\gamma = X(U\gamma)$ and so $\beta = U\gamma$. This implies that $\beta \in C(U)$. If we let Λ be the matrix such that $C(\Lambda)^\perp = C(U)$, then it follows that we are simply testing $\Lambda^T\beta = 0$.

Problem 2:

Show that $Y'(A - A_0)'V^{-1}(A - A_0)Y/[r(X) - r(X_0)]$ is the appropriate numerator mean square for testing model (4) against model (2), where model (4) is

$$Q^{-1}Y = Q^{-1}X_0\beta_0 + Q^{-1}e, \quad Q^{-1}e \sim N(0, \sigma^2 I)$$

and model (2) is

$$Q^{-1}Y = Q^{-1}X\beta + Q^{-1}e, \quad Q^{-1}e \sim N(0, \sigma^2 I).$$

Solution: Define $Y^* = Q^{-1}Y$ and the following two o.p.ms

$$\begin{aligned} M^* &= Q^{-1}X(X'Q^{-1'}Q^{-1}X)^-X'Q^{-1'} \\ M_0^* &= Q^{-1}X_0(X_0'Q^{-1'}Q^{-1}X_0)^-X_0'Q^{-1'}. \end{aligned}$$

Recall that A and A_0 are defined to be

$$A = X(X'V^{-1}X)^-X'V^{-1}$$

and

$$A_0 = X_0(X_0'V^{-1}X_0)^-X_0'V^{-1}.$$

Note that since $V = QQ'$, then $V^{-1} = Q^{-1'}Q^{-1}$. We know the numerator mean square is

$$Y^{*'}(M^* - M_0^*)Y^*/r(M^* - M_0^*),$$

and we easily see that since M^* and M_0^* are o.p.ms and Q is invertible,

$$r(M^* - M_0^*) = r(M^*) - r(M_0^*) = r(Q^{-1}X) - r(Q^{-1}X_0) = r(X) - r(X_0).$$

The last thing to show is the that

$$Y^{\star'}(M^{\star} - M_0^{\star})Y^{\star} = Y'(A - A_0)V^{-1}(A - A_0)Y.$$

First note that each side above can be written as

$$Y^{\star'}(M^{\star} - M_0^{\star})Y^{\star} = [(M^{\star} - M_0^{\star})Y^{\star}]'[(M^{\star} - M_0^{\star})Y^{\star}]$$

and

$$\begin{aligned} Y'(A - A_0)V^{-1}(A - A_0)Y &= Y'(A - A_0)Q^{-1'}Q^{-1}(A - A_0)Y \\ &= [Q^{-1}(A - A_0)Y]'[Q^{-1}(A - A_0)Y]. \end{aligned}$$

Therefore, it suffices to show that $Q^{-1}(A - A_0)Y = (M^{\star} - M_0^{\star})Y^{\star}$. Indeed,

$$\begin{aligned} (M^{\star} - M_0^{\star})Y^{\star} &= [Q^{-1}X(X'V^{-1}X)^{-}X'Q^{-1'} - Q^{-1}X_0(X_0'V^{-1}X_0)^{-}X_0'Q^{-1'}]Q^{-1}Y \\ &= Q^{-1}[X(X'V^{-1}X)^{-}X'V^{-1} - X_0(X_0'V^{-1}X_0)^{-}X_0'V^{-1}]Y \\ &= Q^{-1}(A - A_0)Y. \end{aligned}$$

Therefore, this is the correct numerator mean square for testing model (4) against model (2).

Problem 3:

Verify that the estimate of $\mu + \alpha_i$ is $\bar{y}_{i.}$ and that the algebraic formulas for the sums of squares in the ANOVA table are correct. Hint: To find, for example, $Y(M - \frac{1}{n}J_n^n)Y = Y'M_{\alpha}Y$, use Exercise 4.2 to get $M_{\alpha}Y$ and recall that $Y'M_{\alpha}Y = (M_{\alpha}Y)'(M_{\alpha}Y)$.

Solution: Recall that the LSE of $\rho'X\beta$ is $\rho'MY$. Let $\rho = e_{N_i}$, i.e. the vector of all zeroes with a 1 in the N_i th position. Then, we have

$$\rho'X\beta = \mu + \alpha_i.$$

Also, notice that

$$MY = (\text{rep}(\bar{y}_{1.}, N_1), \dots, \text{rep}(\bar{y}_{t.}, N_t))',$$

where $\text{rep}(\bar{y}_{i.}, N_i)$ denotes the vector of $\bar{y}_{i.}$ of length N_i . It now follows that

$$\rho'MY = \bar{y}_{i.},$$

and hence the estimate of $\mu + \alpha_i$ is $\bar{y}_{i.}$. Next, we verify the algebraic forms of the four sums of squares given in the ANOVA table. Let $n = \sum_{i=1}^t N_i$. The SS of the grand mean can be written as

$$Y' \left(\frac{1}{n} J_n^n \right) Y = Y' \frac{1}{n} \begin{bmatrix} \sum y_{..} \\ \vdots \\ \sum y_{..} \end{bmatrix} = Y' \begin{bmatrix} \bar{y}_{..} \\ \vdots \\ \bar{y}_{..} \end{bmatrix} = \bar{y}_{..} \sum y_{..} = n\bar{y}_{..}^2.$$

This verifies the algebraic form of the SS of the grand mean. Next, the SS of the treatments can be written as

$$\begin{aligned} Y' \left(M - \frac{1}{n} J_n^n \right) Y &= Y' M Y - Y' \left(\frac{1}{n} J_n^n \right) Y = (M Y)' (M Y) - n \bar{y}_{..}^2 \\ &= \sum_{i=1}^t N_i \bar{y}_{i.}^2 - n \bar{y}_{..}^2 = \sum_{i=1}^t N_i (\bar{y}_{i.} - \bar{y}_{..})^2. \end{aligned}$$

The last equality can be seen by

$$\begin{aligned} \sum_{i=1}^t N_i (\bar{y}_{i.} - \bar{y}_{..})^2 &= \sum_{i=1}^t N_i (\bar{y}_{i.}^2 - 2 \bar{y}_{..} \bar{y}_{i.} + \bar{y}_{..}^2) \\ &= \sum_{i=1}^t N_i \bar{y}_{i.}^2 - 2 \bar{y}_{..} \sum_{i=1}^t N_i \bar{y}_{i.} + n \bar{y}_{..}^2 \\ &= \sum_{i=1}^t N_i \bar{y}_{i.}^2 - 2 \bar{y}_{..} \sum_{i=1}^t y_{i.} + n \bar{y}_{..}^2 \\ &= \sum_{i=1}^t N_i \bar{y}_{i.}^2 - 2 n \bar{y}_{..}^2 + n \bar{y}_{..}^2 \\ &= \sum_{i=1}^t N_i \bar{y}_{i.}^2 - n \bar{y}_{..}^2. \end{aligned}$$

This verifies the algebraic form of the SS of the treatments. Next, we verify the algebraic form of the SS of the error. The SS of the error can be written as

$$Y'(I - M)Y = Y'Y - Y'MY = \sum y_{..}^2 - \sum_{i=1}^t N_i \bar{y}_{i.}^2 = \sum_{i=1}^t \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{i.})^2.$$

The last equality can be seen by

$$\begin{aligned} \sum_{i=1}^t \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{i.})^2 &= \sum_{i=1}^t \sum_{j=1}^{N_i} (y_{ij}^2 - 2 y_{ij} \bar{y}_{i.} + \bar{y}_{i.}^2) \\ &= \sum_{i=1}^t \sum_{j=1}^{N_i} y_{ij}^2 - 2 \sum_{i=1}^t \bar{y}_{i.} \sum_{j=1}^{N_i} y_{ij} + \sum_{i=1}^t N_i \bar{y}_{i.}^2 \\ &= \sum_{i=1}^t \sum_{j=1}^{N_i} y_{ij}^2 - 2 \sum_{i=1}^t \bar{y}_{i.} y_{i.} + \sum_{i=1}^t N_i \bar{y}_{i.}^2 \\ &= \sum_{i=1}^t \sum_{j=1}^{N_i} y_{ij}^2 - 2 \sum_{i=1}^t \bar{y}_{i.} N_i \bar{y}_{i.} + \sum_{i=1}^t N_i \bar{y}_{i.}^2 \\ &= \sum_{i=1}^t \sum_{j=1}^{N_i} y_{ij}^2 - \sum_{i=1}^t N_i \bar{y}_{i.}^2. \end{aligned}$$

This verifies the algebraic form of the SS of the errors. The algebraic form of the total SS is trivial. Therefore, we have confirmed the algebraic form of all of the sums of squares in the ANOVA table.

Problem 4:

Show that $\alpha_1 = \alpha_2 = \dots = \alpha_t$ if and only if all contrasts are zero.

Solution: Assume that $\alpha_1 = \alpha_2 = \dots = \alpha_t$ and that λ be any vector satisfying $\lambda'J_t = 0$, i.e. $\sum_{i=1}^t \lambda_i = 0$. We want to show that

$$\sum_{i=1}^t \lambda_i \alpha_i = 0.$$

By assumption, $\alpha_i = \alpha_1$ for all $i = 1, \dots, t$. Therefore, we have

$$\sum_{i=1}^t \lambda_i \alpha_i = \alpha_1 \sum_{i=1}^t \lambda_i = 0.$$

Thus, since λ was any vector with $\lambda'J_t = 0$, all contrasts are zero.

Now for the converse, assume that all contrasts are zero, i.e. for any vector λ satisfying

$$\sum_{i=1}^t \lambda_i = 0, \tag{*}$$

we have

$$\sum_{i=1}^t \lambda_i \alpha_i = 0. \tag{**}$$

Take $\lambda = (1, -1, 0, \dots, 0)'$. Then, clearly this λ satisfies (*) and implies $\alpha_1 = \alpha_2$ by (**). Next, take $\lambda = (0, 1, -1, 0, \dots, 0)'$. Then, by the same logic, we obtain $\alpha_2 = \alpha_3$. Repeating this process will show that $\alpha_1 = \alpha_2 = \dots = \alpha_t$.

Problem 5:

After the final exam of spring quarter, 30 of the subjects of the previous experiment decided to test the sturdiness of 3 brands of sport coats and 2 brands of shirts. In this study, sturdiness was measured as the length of time before tearing when the instructor was hung by his collar out of his second-story office window. Each brand was randomly assigned to 6 students, but the instructor was occasionally dropped before his collar tore, resulting in some missing data. The data are listed below.

Coat 1: 2.34 2.46 2.83 2.04 2.69
 Coat 2: 2.64 3.00 3.19 3.83
 Coat 3: 2.61 2.07 2.80 2.58 2.98 2.30
 Shirt 1: 1.32 1.62 1.92 0.88 1.50 1.30
 Shirt 2: 0.41 0.83 0.53 0.32 1.62

- Give an ANOVA table for these data, and perform and interpret the F test for the differences between brands.
- Test whether, on average, these brands of coats are sturdier than these brands of shirts.
- Give three contrasts that are mutually orthogonal and orthogonal to the contrast in (b). Compute the sums of squares for all four contrasts.
- Give a 95% confidence interval for the difference in sturdiness between shirt brands 1 and 2. Is one brand significantly sturdier than the other?

Solution:

- Below is the ANOVA table for these data:

Source	df	SS	MS	F
treatment	4	18.2959	4.5739	28.29818
error	21	3.3943	0.16163	-
total	25	21.69019	-	-

These values were calculated in R and obtained by the following formulas:

$$df_R = t - 1 = 4;$$

$$df_E = n - t = 21$$

$$SSR = \sum_{i=1}^5 N_i (\bar{y}_{i.} - \bar{y}_{..})^2 = 18.2959$$

$$SSE = \sum_{i=1}^5 \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{i.})^2 = 3.3943$$

$$MSR = SSR/df_R = 4.5739$$

$$MSE = SSE/df_E = 0.16163$$

$$F = MSR/MSE = 28.29818,$$

where N_i is the i th element of $N = (5, 4, 6, 6, 5)$. The R code can be found attached at the end of the homework. Note that $f(.95, 4, 21) = 2.8401$. Since $F > 2.8401$, we reject $H_0: \alpha_1 = \dots = \alpha_5$ and conclude there is a significant difference between at least two brands.

- (b) Recall from section 4.2 in the textbook that we reject $H_0: \frac{\alpha_1 + \alpha_2 + \alpha_3}{3} = \frac{\alpha_4 + \alpha_5}{2}$ if

$$\frac{(\sum_{i=1}^t \lambda_i \bar{y}_i)^2 / (\sum_{i=1}^t \lambda_i^2 / N_i)}{MSE} > F(1 - \alpha, 1, df_E),$$

where $\lambda = (\lambda_1, \dots, \lambda_t)$ satisfies $\sum_{i=1}^t \lambda_i = 0$. Take λ to be the vector

$$\lambda = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{2}, -\frac{1}{2} \right).$$

Then, we calculate the F stat by the above formula

$$F = \frac{(\frac{1}{3}\bar{y}_1 + \frac{1}{3}\bar{y}_2 + \frac{1}{3}\bar{y}_3 - \frac{1}{2}\bar{y}_4 - \frac{1}{2}\bar{y}_5)^2}{\frac{1}{9} \cdot \frac{1}{5} + \frac{1}{9} \cdot \frac{1}{4} + \frac{1}{9} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{5}} \cdot \frac{1}{0.16163} = 104.9664.$$

Again, this calculation was done in R and in the code section. We compare this to $f(.95, 1, 21) = 4.325$ and since $104.9664 > 4.325$, we reject H_0 and conclude a difference in sturdiness between coats and shirts, on average.

- (c) The contrast used in part (b) was $\lambda = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{2}, -\frac{1}{2})$. Notice that three contrasts orthogonal to this contrast and between themselves are

$$\lambda_1 = (0, 1, -1, 0, 0)$$

$$\lambda_2 = (0, 0, 0, 1, -1)$$

and

$$\lambda_3 = \left(1, -\frac{1}{2}, -\frac{1}{2}, 0, 0 \right).$$

Their associated sums of squares are

$$SS = \left(\sum_{i=1}^5 \lambda_i \bar{y}_i \right)^2 / \left(\sum_{i=1}^5 \lambda_i^2 / N_i \right) = 16.96621$$

$$SS_1 = \left(\sum_{i=1}^5 \lambda_{1i} \bar{y}_i \right)^2 / \left(\sum_{i=1}^5 \lambda_{1i}^2 / N_i \right) = 0.88817$$

$$SS_2 = \left(\sum_{i=1}^5 \lambda_{2i} \bar{y}_i \right)^2 / \left(\sum_{i=1}^5 \lambda_{2i}^2 / N_i \right) = 1.26604$$

$$SS_3 = \left(\sum_{i=1}^5 \lambda_{3i} \bar{y}_i \right)^2 / \left(\sum_{i=1}^5 \lambda_{3i}^2 / N_i \right) = 0.49707.$$

These values were calculated using R and the code can be found at the end.

(d) Recall from section 4.2 of the textbook that we reject $H_0: \alpha_4 = \alpha_5$ if

$$\frac{|\sum_{i=1}^t \lambda_i \bar{y}_i|}{\sqrt{MSE \cdot \sum_{i=1}^t \lambda_i^2 / N_i}} > t\left(1 - \frac{\alpha}{2}, df_E\right).$$

Then, taking $\lambda = (1, -1)$, a 95% confidence interval for the difference between shirt brand 1 and shirt brand 2 is

$$\left[\bar{y}_4 - \bar{y}_5 \pm t(.975, 21) \sqrt{MSE \cdot \sum_{i=1}^2 \lambda_i^2 / N_i} \right] = [0.17506, 1.18761].$$

Since 0 is not in this interval, we conclude there is a difference between shirt brands 1 and 2.