**625.661 Statistical Models and Regression**

**Module 5-6 Assignment**

**H.M. James Hung**

**Please complete all the following problems.**

1. Do Problem 7.6, (a), (b), (d), (e), page 255 of Textbook

2. Do Problem 8.9, page 281 of Textbook

$$\mathbf{y} = [y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{31}, y_{32}, y_{33}, y_{34}, y_{41}, y_{42}, y_{43}]'$$

$$\mathbf{X} = \begin{bmatrix} 1&1&1&1&1&1&1&1&1&1&1&1 \\ 1&1&1&0&0&0&0&0&0&0&0&0 \\ 0&0&0&1&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&1&1&1&0&0&0 \end{bmatrix}'$$

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 12 & 3 & 2 & 4 \\ 3 & 3 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix} \qquad (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{5}{6} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{7}{12} \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = [y_{..} \;\; y_{1.} \;\; y_{2.} \;\; y_{3.}]'$$

$$[\hat{\beta}_0 \;\; \hat{\beta}_1 \;\; \hat{\beta}_2 \;\; \hat{\beta}_3]' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad \textbf{(1)}$$

**Note:** $y_{..} = y_{1.} + y_{2.} + y_{3.} + y_{4.}$

**From (1), after algebraic manipulations, we can obtain the OLS estimators,**

$$3\hat{\beta}_0 = y_{..} - y_{1.} - y_{2.} - y_{3.} = y_{4.}$$

$$3\hat{\beta}_1 = -y_{..} + 2y_{1.} + y_{2.} + y_{3.} = y_{1.} - y_{4.}$$

$$6\hat{\beta}_2 = -2y_{..} + 2y_{1.} + 5y_{2.} + 2y_{3.} = 3y_{2.} - 2y_{4.}$$

$$12\hat{\beta}_3 = -4y_{..} + 4y_{1.} + 4y_{2.} + 7y_{3.} = 3y_{3.} - 4y_{4.}$$

**Note:** $\bar{y}_{1.} = \frac{y_{1.}}{3}$ $\bar{y}_{2.} = \frac{y_{2.}}{2}$ $\bar{y}_{3.} = \frac{y_{3.}}{4}$ $\bar{y}_{4.} = \frac{y_{4.}}{3}$

$$\hat{\beta}_0 = \bar{y}_{4.}$$

$$\hat{\beta}_1 = \bar{y}_{1.} - \bar{y}_{4.}$$

$$\hat{\beta}_2 = \bar{y}_{2.} - \bar{y}_{4.}$$

$$\hat{\beta}_3 = \bar{y}_{3.} - \bar{y}_{4.}$$

**No, there is no complication introduced by the unbalanced nature of these data.**

3. Do Problem 8.11, (a), (b), (c), (d), page 282 of Textbook

4. Problem 8.12, (a), (b), (c), page 282-283 of Textbook

(a)

**Note that in the model given in Problem 8.12, the conditions *like***

$$\sum_{i=1}^{a} \tau_i = 0, \ \sum_{j=1}^{b} \gamma_j = 0, \ \sum_{i=1}^{a} (\tau\gamma)_{ij} = 0, \ \sum_{j=1}^{b} (\tau\gamma)_{ij} = 0,$$

**are required such that the design matrix for these parameters are identifiable (i.e., of full rank). Therefore, for $a = 2$, $b = 2$, the given model allows only four free parameters.**

**Now let $x_1 = 1$ for $\tau_1$ and $0$ for $\tau_2$; let $x_2 = 1$ for $\gamma_1$ and $0$ for $\gamma_2$. We can then construct the regression model, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$, for the given model. This regression model also has four free parameters $\beta$'s. Therefore, we can express the parameters, $\tau$'s, $\gamma$'s, $(\tau\gamma)$'s,** as four linear combinations of $\beta$'s in one-to-one correspondence. In addition, testing the hypotheses for $\tau$'s, $\gamma$'s and $(\tau\gamma)$'s is equivalent to testing the hypotheses for the corresponding $\beta$'s.

(b)

**For the regression model constructed, we have**

$$Y = \begin{bmatrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

(c)

**To test $H_0$: $\tau_1 = \tau_2 = 0$, *F*-test is based on $SS_R(\beta_1 \mid \beta_0, \beta_2, \beta_3)$.**

**To test $H_0$: $\gamma_1 = \gamma_2 = 0$, *F*-test is based on $SS_R(\beta_2 \mid \beta_0, \beta_1, \beta_3)$.**

**To test $H_0$: $(\tau\gamma)_{11} = (\tau\gamma)_{12} = (\tau\gamma)_{22} = 0$, *F*-test is based on $SS_R(\beta_3 \mid \beta_0, \beta_1, \beta_2)$.**

5. Use any math/stat software (e.g., *www.***numbergenerator***.org/***randomnumbergenerator**) of your choice to find a random number generator to randomly select 12 rows of data table used in Problem 8.16 (page 283) of Textbook and then perform an analysis of your generated data and discuss you results. State the assumptions for your analysis.