BIOS -900 Homework #4

Date assigned: 11/20/2017
Due Date: 12/04/2017 by 11:59 pm (Blackboard clock time);

Instructions:

1. To receive full credit, show all work. Please make your work legible.
2. Total points for this homework are 100.
3. Do not forget to write your name on the homework.
4. Insert page numbers on all pages and total # of pages submitted.
5. Homework can be typed or hand-written. Provide SAS code wherever necessary.
6. Use the BLACKBOARD drop box to turn in the homework (preferably as pdf) or bring it to class on 12/04/2017.

Question # 1: **25 points**

Two varieties of corn (variety A and variety B) were compared in a field trial. In addition to the varieties, three levels of nitrogen were used (100, 150, 200 pounds per acre). Six different fields were used and the six combinations of varieties and nitrogen levels were randomly assigned to the fields. Let Y_{ij} denote the yield (in bushels per acre) of the i^{th} variety of corn when the j^{th} level of nitrogen is applied. Let ε_{ij}, $i = 1, 2, j = 1, 2, 3$ denote independent $N(0, \sigma^2)$ random variables where σ^2 is an unknown variance. The following two models were proposed:

Model #1:

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -50 & 2500 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 50 & 2500 \\ 0 & 1 & -50 & 2500 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 50 & 2500 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix}$$

Model #2:

$$\begin{pmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{21} \\ Y_{22} \\ Y_{23} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 0 & -2 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & 0 & -2 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix}$$

In model #1, note that the third column of the model matrix are the values of the nitrogen level minus 150 lb/acre and the fourth column contains squares of the elements in the third column.

a) For model #1, determine if $\gamma_1 - 10\delta_1 + 100\delta_2$ is estimable. Give a brief explanation to support your conclusions.

b) For model #2, determine if $\mu + \alpha_1$is estimable. Give a brief explanation to support your conclusion.

c) Expressing model #2 as $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ a solution to the normal equations is $\boldsymbol{b} = (X^T X)^{-} X^T \boldsymbol{Y}$**.** Explain how a generalized inverse $(X^T X)^{-}$ can be computed. You are not expected to obtain a numerical value for the generalized inverse, just briefly outline a procedure for how it can be computed.

d) Using $\boldsymbol{b} = (X^T X)^{-} X^T \boldsymbol{Y}$ from part (c), define the estimator $\widehat{\alpha_1} - \widehat{\alpha_2} = [\,0\ \ 1\ \ -1\ \ 0\ \ 0\,]\,\boldsymbol{b}$**.** What are the properties of this estimator?

e) Would the residual sum of squares from fitting model 1 and model 2 be the same? Explain.

f) Using $\boldsymbol{Y} = W\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ to express model #1 in matrix notation, show how to construct a 95% prediction interval for the corn yield that would be observed if variety 1 is grown with a nitrogen application of 120 lb/acre.

g) With respect to model #1, derive the test statistic to test the null hypothesis $Ho\colon \gamma_1 = \gamma_2$ and $\delta_2 = 0$ against the alternative that Ho is not true. This null hypothesis is equivalent to the statement that the same strait line model for the effect of nitrogen on the mean corn yield is appropriate for both varieties. State how you would then use this test statistic.

Question # 2: **20 points**

For the linear model from Q#4 in HW#3 (i.e., $\boldsymbol{Y} = \boldsymbol{XB} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{I})$).

a) Determine which of the following hypotheses are testable. In each case, give a justification for your answer.

   i. $H_o\colon \alpha_1 = \alpha_2$

   ii. $H_o\colon \alpha_3 = 0$

   iii. $H_o\colon \mu = 0$

   iv. $H_o\colon \alpha_1 = \alpha_3$ and $\alpha_1 - 2\alpha_2 + \alpha_3 = 0$

b) Find the distribution of the quadratic form

$$SSE = \frac{1}{\sigma^2}\boldsymbol{Y^T}(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{Y}, \text{ where } P_X = \boldsymbol{X}(\boldsymbol{X^T X})^{-}\boldsymbol{X^T}.$$

c) Show how to construct an F-test of the null hypothesis $H_o\colon \alpha_1 = \alpha_2$ and $\alpha_1 - 2\alpha_2 + \alpha_3 = 0$ against the alternative $H_a\colon \alpha_1 \neq \alpha_2$ or $\alpha_1 - 2\alpha_2 + \alpha_3 \neq 0$.

d) Express the non-centrality parameter for F-test in part (c) as a function of $\alpha_1, \alpha_2, \alpha_3$.

Question # 3: **20 points**

Using workout similar to Page #365-369 show how the orthogonal polynomial coefficients can be obtained for k = 5 as shown in Table 13.5.

Question # 4: **20 points**

Derive the expression for the expected mean square for the interaction term A*B as shown in Table 14.5 using:
{i} Sums of Squares approach
{ii} Quadratic form approach

Question # 5: **15 points**

{i}
Figure 14.1 on Page #386 of the textbook shows the cell means for a model with a=3 and b=2. Equation 14.35 on Page #387 shows the null hypothesis for a 'test of interaction.
Develop similar hypothesis for a 'test of interaction' in the case where a = 4 and b = 3. Discuss how Theorem 12.7[b] can be used to conduct this hypothesis test.

{ii}
In class, we discussed solving Normal equations for a balanced one-way ANOVA. By imposing suitable 'side conditions' we were able to obtain estimates of the parameter vector (see Page #342 of textbook).This allowed us to construct the 'Sum of Squares' Table 13.1 (See Page # 346 of textbook). Adopt a similar approach and develop a table similar to Table 13.1 in the case of an unbalanced design i.e. where n_i observations are available for treatment level i for $i = 1,2, \ldots, a$.

Question # 6 (Bonus) **5 points**

Read the pdf related to Chapter #12 "Coding schemes for Regression and Chapter #13 "One-way ANOVA" from the book "Regression and ANOVA – An Integrated Approach Using SAS Software" posted on Blackboard in the folder named 'ExtraReadingMaterial'.
Confirm that you have studied the different coding schemes discussed in these two book chapters.

GOOD LUCK ☺☺