BIOS -900 Homework #3

Date assigned: 10/30/2017
Due Date: 11/13/2017 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 also 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 11/13/2017.

Question # 1: **15 points**

Solve the following exercises from your textbook
{i} Exercise 8.5
{ii} Exercise 8.9
{iii} Exercise 8.11
{iv} Equation 8.28 on Page #200.

Question # 2: **35 points**

Using the dataset SENIC.csv that you used in Q#4 of HW#2, answer the following questions related to study material covered in Chapter 8.

Note: All problems should be solved using matrix algebra using PROC IML in SAS (or equivalent R functions). You cannot use procedures like PROC REG or PROC GLM (except for verifying your answers found using matrix operations). Show all relevant SAS code.

[A] Test the following hypotheses at the 5% level of significance:

{i} $H_0: \beta_1 = 0$ {ii} $H_0: \begin{bmatrix} \beta_1 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ {iii} $H_0: \begin{bmatrix} \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

{iv} $H_0: \beta_1 = 2(\beta_3 + \beta_4)$ {v} $H_0: \beta_2 - 3(\beta_1 + \beta_3 + \beta_4) = 0.25$

{vi} Conduct a hypothesis test to separately test whether or not each β_j for $j = 1,2,3,4$ is equal to 0 using [1] Bonferroni method [2] Post-hoc tests using Scheffe’s method.

[B] Construct 95% confidence intervals for:

{i} Each individual β_j for $j = 1,2,3,4$ {ii} $\beta_1 - 2(\beta_3 + \beta_4)$ {iii} σ^2

[C] Consider a linear model with an intercept with only 'age' as the predictor. Generate a 95% prediction interval for a 67 year old patient from another hospital.

[D] Test the hypothesis given in Part [A] – {iv}, using a likelihood ratio test by obtaining MLEs of the parameters of interest separately under the null and alternate hypothesis. Show equivalency of your results with the answer you obtained by using an F-test.

Question # 3: **15 points**

Using techniques that we studied in class from Chapter #9, identify outliers and highly influential observations for the dataset given in Table 7.5 (Page #184).

Note: All problems should be solved using matrix algebra using PROC IML in SAS (or equivalent R functions). You cannot use procedures like PROC REG or PROC GLM (except for verifying your answers found using matrix operations). Show all relevant SAS code.

Question # 4: **25 points**

Consider the linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ with

$\boldsymbol{Y}^T = (Y11, Y12, Y21, Y22, Y23, Y24, Y31, Y32)$ $\boldsymbol{\beta}^T = (\mu, \alpha_1, \alpha_2, \alpha_3)$

$$
\mathrm{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}
$$

and $E(\boldsymbol{\epsilon}) = \boldsymbol{0}$ and $\mathrm{V}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$.

i. Show that **G** is the generalized inverse for $\boldsymbol{X}^T\boldsymbol{X}$

$$
\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0.50 \end{bmatrix}
$$

ii. Use **G** from (a) to obtain a solution to the normal equations $(\boldsymbol{X^T X})\boldsymbol{b} = \boldsymbol{X^T Y}$

iii. Show that $c_1^T = (1, 1, 0, 0)$, $c_2^T = c(1, 0, 1, 0)$, and $c_3^T = c(1, 0, 0, 1)$ are linearly independent.

iv. Show that $\boldsymbol{c_1^T}, \boldsymbol{c_2^T}, \boldsymbol{c_3^T}$ are a basis for the vector space spanned by the rows of **X**

v. Show that $\boldsymbol{c}_1^T\boldsymbol{\beta}, \boldsymbol{c}_2^T\boldsymbol{\beta}, \boldsymbol{c}_3^T\boldsymbol{\beta}$ are estimable functions of $\boldsymbol{\beta}$

vi. Show that $\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3$ characterize the set of all possible estimable functions in the sense that for any estimable $\boldsymbol{c}^T\boldsymbol{\beta}, \boldsymbol{c} = \boldsymbol{k}_1\boldsymbol{c}_1 + \boldsymbol{k}_2\boldsymbol{c}_2 + \boldsymbol{k}_3\boldsymbol{c}_3$ for some $(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$

vii. Comment on the over-parameterized nature of the ANOVA model that you used in parts {i} through {vii}. Choose a suitable 'reparameterization' approach and obtain estimates for you chosen model.

viii. Choose suitable 'side conditions' in this over-parametrized model and obtain parameter estimates. Show that your choice of 'side conditions' are non-estimable functions of $\boldsymbol{\beta}$

Question # 5: **10 points**

Data were collected to study the effect of a medication on blood pressure. Two different drugs A and B were used in the study. Blood pressures were measured for patients at different ages. The data are as follows:

Subject	Blood Pressure (Y)	Drug
1	120	A
2	140	B
3	135	A
4	110	B
5	160	A
6	164	B
7	155	A
8	175	B
9	170	A
10	165	B

Each subject can be considered as an independent observation. The drugs were randomized to subjects. Consider the model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \text{ i = 1, 2 and j = 1, 2, 3, 4, 5,}$$

where Y_{ij} = the observed blood pressure for the j^{th} subject on the i^{th} drug, and $\alpha_i = 1$ if the i^{th} drug was used.

Determine which, if any, of the following quantities are estimable. For each estimable quantity, report the value of a vector **a** such that $\mathbf{a}^T\mathbf{Y}$ satisfies the definition of an estimable function of $\boldsymbol{\beta}$.

{i} μ {ii} $\mu + \alpha_1$ {iii} $\alpha_1 - \alpha_2$ {iv} $\alpha_1 + \alpha_2$

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

{i} Study solutions to Exercises 8.19 to 8.23. After understanding these proofs, study Theorems 8.4 (a) through 8.4 (g).

{ii} Study solutions to Exercise 9.4 to 9.8.

Confirm that you have studied these exercise problems.

GOOD LUCK ☺☺