BIOS -900 Fall 2017 Take-Home Final Exam

Name:____

Date assigned: 12/07/2017;   Due Date: 12/14/2017 (by 11:59 pm Blackboard clock time)

Instructions:

1. To receive full credit, show all work. Please make your work readable.
2. Use $\alpha = 0.05$ unless otherwise mentioned.
3. Total points for this exam are **50**.
4. Do not forget to write your name and page numbers on the homework.
5. Do not turn in irrelevant SAS output.
6. You cannot discuss this exam with anyone other than the instructor.

Question # 1: **(12 points)**

When gasoline is pumped into the tank of a car, vapors are vented into the atmosphere. A company has developed a device that can be installed in the gas tank of a car to prevent vapors from escaping when the gas tank is filled. A small study was performed to examine the effectiveness of this device. Four cars were used in the study, and the device was installed in the gas tank of two of the cars. Gasoline was pumped into the tank of each car and the amount of gas vapor that escaped (Y) was measured. Since the temperature of the gasoline (X_i) can affect the outcome, two gasoline temperatures were used. The data are shown below. In this table $X_2 = 1$ if the device was used and $X_2 = -1$ if the device was not installed in the gas tank.

Amount of vapor that escape (Y)	Gasoline temperature (°C) (X_1)	Use of device (X_2)
Y_1	0	-1
Y_2	30	-1
Y_3	0	1
Y_4	30	1

Consider the model $Y_i = \beta_1(X_{1i}) + \beta_2(X_{2i}) + \varepsilon_i$ where ε_i is random error with E(ε_i) = 0 for i = 1, 2, 3, 4.

Answer the following questions:

a. Under what conditions would this be a normal theory Gauss-Markov model?

b. Interpret $\beta_2$in this model. What does it represent in the context of the gasoline vapor study?

c. Let $\boldsymbol{b}$ be a solution to the normal equations. What are the properties of $\boldsymbol{b}$? Assume the conditions you listed in part (a) are satisfied.

d. To test for a device effect, the researchers propose the following test statistics

$$F = \frac{(Y_3+Y_4-Y_1-Y_2)^2}{2*SSE}.$$

Assuming the assumptions you listed in part (a) are satisfied, show that this statistic has an F-distribution. Report the degrees of freedom for this statistic. Hint: Start with fact that $c^T Y \sim N(c^T X\beta, \sigma^2 c^T c)$ for $c^T = (-1 - 1\ \ 1\ \ 1)$ and $c^T Y = Y_3 + Y_4 - Y_1 - Y_2$.

e. With respect to $\beta = (\beta_1, \beta_2)^T$, describe the null hypothesis that can be tested with the F-test in part (e). What is the alternative hypothesis?

f. Is it possible to test for an interaction effect between Temperature and Device? If yes, why? If no, why not?

Question # 2: **(25 points)**

The data in the following table are breaking strengths of beams made from combinations of types of cement and mixtures of aggregate. Four beams were made from each combination, but some of the beams were of poor quality due to the fabrication process and were not tested for strength.

Cement	Aggregate A	Aggregate B	Aggregate C	Aggregate D
Type 1	21	19	19	23
Type 1	27	19	16	24
Type 1	19	22	—	23
Type 1	—	—	—	—
Type 2	25	23	19	28
Type 2	23	20	18	27
Type 2	24	24	—	25
Type 2	—	18	—	—
Type 3	20	28	14	23
Type 3	24	—	16	25
Type 3	—	—	12	22
Type 3	—	—	—	22

Answer the following questions:

a. Use an "unconstrained" cell means model to determine if there is an interaction between cement and aggregate. Also, conduct hypotheses tests for the main effect of cement and the main effect of aggregate. How will you interpret these main effects?

b. Use a "constrained (additive only)" cell means model to test for main effect of cement and the main effect of aggregate.

c. Assume that the breaking strength value of 28 for 'Cement = Type 3' and 'Aggregate = B' was not available (i.e. empty cell) due to that particular beam being of poor quality. Conduct a hypothesis tests for the main effect of cement, main effect of aggregate, and the interaction effect using the approach we took for the "Bakery Study Data" discussed in the last lecture of the course. Compare your results with the Type IV sum of squares given by PROC GLM in SAS.

d. Using the same assumption as in part [c] above, test for the interaction effect using:
{i} Full-Reduced model approach (See equation 15.50 from textbook)
{ii} Side conditions (See equation 15.54 from textbook)

e. Ignore the factor 'Cement' and consider this problem as a one-way ANOVA with 'Aggregate' as the only factor. Test for the main effect of this factor using:
{i} Cell means coding {ii} Reference cell coding {iii} Effect cell coding.

Question # 3: **(13 points)**

Answer the following questions related to the data given in Table 7.4 (Page #183) of your textbook:

i. Solve Exercise 7.54 from your textbook.
ii. Part [d] of Exercise 7.54 mentions using the second order model. Discuss the effect of "underfitting" when the second order terms are ignored.
iii. Using techniques that we studied in class from Chapter #9, identify outliers and highly influential observations (if any) for this dataset.
iv. Construct 95% confidence intervals for any three $\beta_j's$ and for σ^2 using the model in part {ii} above.

Note: All problems (except for Q#2 part [c]) 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.

☺ GOOD LUCK ☺