STAT 578: Homework #1 **Due: Friday, September 5, 2014**

For problems 1 and 2, you can cite any of the numbered items in the handout *Useful Background Information.*

1. (1.5pt) Refer to (73.) on page 12 of your notes. Verify $E(\boldsymbol{b}) = \boldsymbol{\beta}$ (assuming the proposed linear model is correct).

2. (1.5pt) Refer to (75.) on page 12 of your notes, and assume responses are independent. Thus, $\sigma_{ij} = 0$ for all $i \neq j$. Verify

$$E[(\boldsymbol{b} - \boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta})'] \;=\; \sigma^2(\boldsymbol{X'X})^{-1}.$$

3. (2pt) In (73.) on page 12 of your notes, the claim is that $E(\boldsymbol{b}) = \boldsymbol{\beta}$ if the model true model is $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ and $\boldsymbol{b} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y}$ is the least squares estimator of $\boldsymbol{\beta}$. Suppose, however, that the true model contains additional parameters given by vector $\boldsymbol{\beta}_2$. Therefore, the "true" linear model should be

$$\boldsymbol{y} \;=\; \boldsymbol{X\beta} \;+\; \boldsymbol{X}_2\boldsymbol{\beta}_2 \;+\; \boldsymbol{\epsilon}$$

where $E(\boldsymbol{\epsilon}) = 0$ and $\text{var}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I}_n$.

Suppose $\boldsymbol{X}$ is $n \times p$ and $\boldsymbol{X}_2$ is $n \times p_2$. Show $E(\boldsymbol{b}) = \boldsymbol{\beta} + \boldsymbol{A\beta}_2$. where $\boldsymbol{A} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'X}_2$ and $\mathbf{b}$ is the least squares estimator of $\boldsymbol{\beta}$.

4. (2pt) Suppose $\eta(x_1, x_2) \;=\; 2(x_1 + 1)e^{-4x_2}$. What is the second-order Taylor series expansion of η?

5. (6pt) Problem 2.6, page 65 (The data is on the Stat 578 course web page). You do not need to center and scale the data. In part (a), you should fit the first-order model. In addition to parts (a) and (b),

   - Check for model lack-of-fit and use VIFs to check for any serious collinearity problems.

6. (4.5pt) Problem 2.12, page 67. Replace (d) with

   (d) Fit a new model to the response PITCH by adding new regressors SOAKTIME×SOAKPCT and DIFFTIME×DIFFPCT to the first-order model.

7. (1.5pt) Check the model you fit in Problem 2.12 (d) for collinearity problems. Comment on what you found.

8. (4pt) Remove the last two observations in the data set for Problem 2.6. Refit the first-order model and compare these results to your previous analysis using the full data set. Do the results seem consistent? Justify your answer.

Extra Credit (3pt) Partial PROC GLM output for the interaction model using all five variables X1, X2, X3, X4, and X5 is given on the back of this page. In the ANOVA table, you see that several terms have 0 degrees of freedom and 0 sum of squares. However, in the ANOVA results generated earlier for the first-order model, the sum of squares for X1 is positive.

Why did this happen? Be as specific as possible. Hint: You may want to run this model yourself and look at all of the output.

Dependent Variable: PITCH

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	14	0.00427555	0.00030540	101.95	<.0001
Error	17	0.00005092	0.00000300		
Corrected Total	31	0.00432647			

R-Square	Coeff Var	Root MSE	PITCH Mean
0.988230	6.585450	0.001731	0.026281

Source	DF	Type III SS	Mean Square	F Value	Pr > F
X1	0	0.00000000	.	.	.
X2	1	0.00003258	0.00003258	10.88	0.0042
X1*X2	1	0.00003326	0.00003326	11.10	0.0039
X3	1	0.00000801	0.00000801	2.67	0.1205
X1*X3	1	0.00000817	0.00000817	2.73	0.1170
X2*X3	1	0.00001857	0.00001857	6.20	0.0234
X4	0	0.00000000	.	.	.
X1*X4	0	0.00000000	.	.	.
X2*X4	1	0.00000201	0.00000201	0.67	0.4236
X3*X4	1	0.00001406	0.00001406	4.69	0.0448
X5	0	0.00000000	.	.	.
X1*X5	0	0.00000000	.	.	.
X2*X5	1	0.00000316	0.00000316	1.05	0.3190
X3*X5	1	0.00000215	0.00000215	0.72	0.4086
X4*X5	1	0.00000258	0.00000258	0.86	0.3667