Stat 615/415 (Regression) Homework # 8

Multivariate Regression – inference, tests, model comparison, categorical predictors
chap. 7–8 (7.1–7.3, 8.2–8.5)

1. (**7.1**) State the number of degrees of freedom that are associated with each of the following extra sums of squares: $\text{SSReg}(X_1 \mid X_2)$, $\text{SSReg}(X_2 \mid X_1, X_3)$, $\text{SSReg}(X_1, X_2 \mid X_3, X_4)$, $\text{SSReg}(X_1, X_2, X_3 \mid X_4, X_5)$.

   A note about the notation. SSReg($A \mid B$) is the extra sum of squares that appeared as a result of including variables A into the regression model that already had variables B in it. Thus, it is used to compare the full model with both A and B in it against the reduced model with only B.

2. (**7.2**) Explain in what sense the regression sum of squares $\text{SSReg}(X_1)$ is an extra sum of squares.

3. (**7.28b**) For a multiple regression model with five X variables, what is the relevant extra sum of squares for testing

   (a) whether or not $\beta_5 = 0$?

   (b) whether or not $\beta_2 = \beta_4 = 0$?

4. (**7.28b, Stat-615 only**)
   Show that $\text{SSReg}(X_1, X_2, X_3, X_4) = \text{SSReg}(X_2, X_3) + \text{SSReg}(X_1 \mid X_2, X_3) + \text{SSReg}(X_4 \mid X_1, X_2, X_3)$.

5. (**7.3, 7.24, 7.30**) Continue working with the *Brand Preference* data, which are available on our Blackboard, on http://statweb.lsu.edu/EXSTWeb/StatLab/DataSets/NKNWData/CH06PR05.txt, and in the previous homework.

   (a) Obtain the ANOVA table that decomposes the regression sum of squares into extra sum of squares associated with X_1 and with X_2, given X_1.

   (b) Test whether X_2 can be dropped from the model while X_1 is retained.

   (c) Fit first-order simple linear regression for relating brand liking (Y) to moisture content (X_1).

   (d) Compare the estimated regression coefficient for X_1 with the corresponding coefficient obtained in (a).

   (e) Does $SS_{reg}(X_1)$ equal $SS_{reg}(X_1|X_2)$ here? Is the difference substantial?

   (f) Regress Y on X_2 and obtain the residuals.
   Regress X_1 on X_2 and obtain the residuals.
   Regress residuals from the model "Y on X_2" on residuals from the model "X_1 on X_2"; compare the estimated slope, error sum of squares with #1. What about R^2?

6. **(8.13)** Consider a regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where X_1 is a numerical variable, and X_2 is a dummy variable. Sketch the response curves (the graphs of $\mathbf{E}(Y)$ as a function of X_1 for different values of X_2), if $\eta_0 = 25$, $\beta_1 = 0.2$, and $\beta_2 = -12$.

7. Continue the previous exercise. Sketch the response curves for the model with interaction, $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$, given that $\beta_3 = -0.2$.

8. **(8.34)** In a regression study, three types of banks were involved, namely, (1) commercial, (2) mutual savings, and (3) savings and loan. Consider the following dummy variables for the type of bank:

Type of Bank	X_1	X_2
Commercial	1	0
Mutual saving	0	1
Saving and loan	0	0

   (a) Develop the first-order linear regression model (no interactions) for relating last year's profit or loss (Y) to size of bank (X_1) and type of bank (X_2, X_3).

   (b) State the response function for the three types of banks.

   (c) Interpret each of the following quantities: (1) β_2, (2) β_3, (3) $\beta_2 - \beta_3$.

9. **(8.16, 8.20)** Refer to our old **GPA data**, available on Blackboard and on `http://statweb.lsu.edu/EXSTWeb/StatLab/DataSets/NKNWData/CH01PR19.txt`.
An assistant to the director of admissions conjectured that the predictive power of the model could be improved by adding information on whether the student had chosen a major field of concentration at the time the application was submitted. Suppose that the first 10 students chose their major when they applied.

   (a) Fit the regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \varepsilon$, where X_1 is the entrance test score and $X_2 = 1$ if a student has indicated a major at the time of application, otherwise $X_2 = 0$. State the estimated regression function.

   (b) Test whether X_2 can be dropped from the model, using $\alpha = 0.05$.

   (c) Fit the regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$ and state the estimated regression function. Interpret β_3. Test significance of the interaction term.