Stat 615/415 (Regression) Homework # 6

Multivariate Regression (chap. 6)

1. (**6.3**) A student stated: "Adding predictor variables to a regression model can never reduce R^2, so we should include all available predictor variables in the model." Comment.

2. (**6.4**) Why is it not meaningful to attach a sign to the coefficient of multiple correlation R, although we do so for the coefficient of simple correlation r_{12}?

3. (**6.27**) In a small-scale regression study, the following data were obtained,

Y	$X1$	$X2$
42.0	7.0	33.0
33.0	4.0	41.0
75.0	16.0	7.0
28.0	3.0	49.0
91.0	21.0	5.0
55.0	8.0	31.0

   Assume the standard multiple regression model with independent normal error terms. Compute **b, e, H,** SSErr, R^2, $s^2_{\mathbf{b}}$, $\hat{Y}$ for $X_1 = 10, X_2 = 30$. You can do the computations using software or by hand, although it would be lengthy to do them by hand.

4. (Computer project, **#6.5—#6.8**) Dataset "Brand preference" is available on our Blackboard, on http://statweb.lsu.edu/EXSTWeb/StatLab/DataSets/NKNWData/CH06PR05.txt, and here:

Y_i	64	73	61	76	72	80	71	83	83	89	86	93	88	95	94	100
X_{i1}	4	4	4	4	6	6	6	6	8	8	8	8	10	10	10	10
X_{i2}	2	4	2	4	2	4	2	4	2	4	2	4	2	4	2	4

   It was collected to study the relation between degree of brand liking (Y) and moisture content (X_1) and sweetness (X_2) of the product.

   (a) Fit a regression model to these data and state the estimated regression function. Interpret b_1.

   (b) Obtain residual plots. What information do they provide? Plot residuals against fitted values, against each predictor, and against the product of predictors.

   (c) Test homoscedasticity.

   (d) Conduct a formal lack of fit test.

   (e) Test whether the proposed linear regression model is significant. What do the results of the ANOVA F-test imply about the slopes?

   (f) Estimate both slopes simultaneously using the Bonferroni procedure with at least a 99 percent confidence level.

   (g) Report R^2 and adjusted R^2. How are they interpreted here?

   (h) Calculate the squared correlation coefficient between Y_i and $\hat{Y}_i$. Compare with part (g).

(i) Obtain a 99% confidence interval for $\mathbf{E}(Y)$ when $X_1 = 5$ and $X_2 = 4$. Interpret it.

(j) Obtain a 99% prediction interval for a new observation Y when $X_1 = 5$ and $X_2 = 4$. Interpret it.

5. (**# 6.26, Stat-615 only**) Show that the squared sample correlation coefficient between Y and $\hat{Y}$ equals R^2.
Remark. Now you can check if you did #3h correctly.
Hints. First, show that the sample averages of Y_i *and* $\widehat{Y}_i$ *are the same. Then, write the sample correlation coefficient between* Y *and* $\hat{Y}$ *as*

$$r_{Y\hat{Y}} = \frac{\sum(Y_i - \overline{Y})(\widehat{Y}_i - \overline{Y})}{\sqrt{\sum(Y_i - \overline{Y})^2 \sum(\widehat{Y}_i - \overline{Y})^2}} = \frac{\sum(\widehat{Y}_i - \overline{Y} + Y_i - \widehat{Y}_i)(\widehat{Y}_i - \overline{Y})}{\sqrt{\sum(Y_i - \overline{Y})^2 \sum(\widehat{Y}_i - \overline{Y})^2}}$$

and use known properties of residuals $\sum e_i = 0$, $\sum X_{ij} e_i = 0$, $\sum \widehat{Y}_i e_i = 0$.