STA 6207 — Section 8907 — Fall 2008

Assignment 4

1. Using a calculator (*not* a computer), perform Exercise 4.11 from the textbook, Section 4.8. You should assume that the model has an intercept as its first mean-related parameter, and use notation of the form X_j to represent the j^{th} independent variable (and X_{ij} to represent its value for the i^{th} observation, if needed).

   In addition, when you answer part (e), do the following:

   (i) First write out an appropriate *scalar* model equation for the reduced model, defining any new variables and corresponding parameters you may need. How can you tell that this is a *linear* reduced model?

   (ii) Next, *compute* $(\boldsymbol{X}^*)'\boldsymbol{X}^*$ and $(\boldsymbol{X}^*)'\boldsymbol{Y}$, where $\boldsymbol{X}^*$ is the $\boldsymbol{X}$ matrix corresponding to the reduced model defined by your scalar model equation. (Hint: You should be able to compute these from $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{X}'\boldsymbol{Y}$ by considering the relationship between your new variable(s) and the original variables.)

   (iii) Finally, *compute* the residual sum of squares for your reduced model, using your results from previous parts. Use this value and your previous results to perform an F-test of the GLH from part (d) (with $\alpha = 0.05$).

   [Note: In parts (b) and (c), "standard error" refers to the *estimated* standard error. In part (d), give *both* degrees of freedom used in the test.]

2. Refer to the model and your data from Problem 2 of Assignment 3 (rental car volume of business) to answer the following parts. (You may perform the computations by hand or using SAS® or R. Submit any relevant code and output.)

   (a) Test the null hypothesis that $\beta_1 = \beta_2 = 0$ (i.e. that volume of travel, whether business-related or personal, is not linearly related to volume of car rentals). Use $\alpha = 0.05$.

   (b) Form an individual 95% two-sided confidence interval for β_1 (the unit increase in number of rentals per unit increase in number of business travelers).

   (c) Test the null hypothesis that $\beta_2 = 0$ versus the alternative hypothesis that $\beta_2 > 0$ (i.e. whether or not an increase in non-business tourism increases volume of rentals). Use $\alpha = 0.05$.

   (d) Based on the full linear regression model that you estimated, compute a prediction $\hat{Y}$ of the volume of rental business in a month in which 16 thousand business travelers and 22 thousand other tourists visit the city. Estimate the variance of $\hat{Y}$ and the mean squared error of prediction. Form a 95% two-sided prediction interval.

   (e) Test the null hypothesis that $\beta_1 = \beta_2$ versus the alternative that $\beta_1 \neq \beta_2$ (i.e. determine whether the apparent rental rate among business travelers is different than that among other tourists). Use $\alpha = 0.05$.

   (f) Obtain $R(\beta_0)$ and the following sequential sums of squares: $R(\beta_1 \mid \beta_0), R(\beta_2 \mid \beta_0\ \beta_1)$

   (g) Obtain the *partial* sums of squares for X_1 and X_2. What hypotheses are these used to test? (Give the appropriate null and alternative hypotheses in each case.)

(h) Form Bonferroni 95% simultaneous two-sided confidence intervals for β_0, β_1, and β_2.

(i) Form Scheffé 95% simultaneous two-sided confidence intervals for β_0, β_1, and β_2.

3. For data from the linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

the *coefficient of partial determination* between Y and X_1 (after controlling for X_2) is

$$r^2_{Y,X_1|X_2} = \frac{R(\beta_1 \mid \beta_0 \; \beta_2)}{\mathrm{SS}(\mathrm{Res}_{\mathrm{red}})}$$

where $\mathrm{SS}(\mathrm{Res}_{\mathrm{red}})$ is the residual sum of squares for the model after dropping X_1 (but retaining X_2).

In words, it is the ratio of the partial sum of squares for X_1 to the residual sum of squares for the model without X_1. Conceptually, it is the fraction of that variation in Y not explained by X_2 that *is* explained by X_1 (when it is added to the model already containing X_2). It represents the strength of the marginal contribution of X_1 for predicting Y after accounting for the contribution from X_2 alone.

The coefficient of partial determination between Y and X_2 (after controlling for X_1), denoted $r^2_{Y,X_2|X_1}$, is defined similarly (with the roles of X_1 and X_2 interchanged).

(a) Show algebraically that (with probability 1)

$$r^2_{Y,X_1|X_2} \leq R^2,$$

where R^2 is the usual coefficient of (multiple) determination for the full model. Also, state the condition under which this inequality becomes an equality (using only R-notation).

Identify any facts or definitions from lecture or the textbook that you use in your derivation.

(b) Compute $r^2_{Y,X_1|X_2}$ and $r^2_{Y,X_2|X_1}$ for the model and data you used in Problem 2 above (which is the same model and data as for Problem 2 on Assignment 3).

[Remark: This definition can be extended easily to models with more than two independent variables, by controlling for all other independent variables. The "`Partial R-Square`" that appears in some SAS® output is an example of this.]