ST505: Fall 2012 Homework 6. Due: Wed. Nov. 7

1. Consider the patient satisfaction data in problem 6.15 in the text.

   (a) For each of the models coming later the vector of outcomes $\mathbf{Y}$ and the vector of error terms $\boldsymbol{\epsilon}$ will be of the same form (although what is in $\boldsymbol{\epsilon}$ and its meaning depends on the model).

   - What are the dimensions of $\mathbf{Y}$? Describe $\mathbf{Y}$, showing explicitly what is in the first three and last three elements. You can just put ... to represent what is in between.

   - What are the dimensions of $\boldsymbol{\epsilon}$? Give a general representation of it (you don't need to list all 46 values, just describe its general form).

   (b) For each of the models below:

   i) give the dimensions of $\mathbf{X}$ and write out the first two and last two rows of $\mathbf{X}$
   ii) give the dimensions of $\boldsymbol{\beta}$ and write out $\boldsymbol{\beta}$ explicitly.
   iii) state what the dimensions of $\mathbf{b}$ (the least squares estimator of $\boldsymbol{\beta}$) are.
   iv) Write out what the components of $\sigma^2\{\mathbf{b}\}$ are (and in so doing you will specify the dimensions of $\sigma^2\{\mathbf{b}\}$). Note I'm not asking you to compute $(\mathbf{X}'\mathbf{X})^{-1}$ in doing this, just to state describe the various components as the variance or covariances of such and such.

   The models are:
   MODEL 1: A simple linear regression model of Y on X_1 only; that is, $Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$.
   Model 2: A simple linear regression model of Y on all three X's; $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$.
   Model 3: A model that is linear in X_1 and X_2 and allows an interaction between them; that is, $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \epsilon_i$.

2. Fitting the Patient satisfaction data. This will do various parts of what are being asked for in problems 6.15 and 6.16 (but we'll put off other parts of these questions until later). All calculations should be done out in R or SAS. Turn in your code. Extract the parts of the output needed to answer the questions so I know exactly what is answering what.

   Assume to start the linear regression model in the three predictors is adequate. That is, $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$ with the usual assumptions on the ϵ_i to start.

   (a) Fit the regression model Show the associated analysis of variance table, the estimated coefficients, standard errors, associated t-test, 95% confidence intervals for the individual coefficients along with the estimate of $\sigma^2\{\mathbf{b}\}$ (the variance- covariance matrix of $\mathbf{b}$) computed two ways, one assuming that the usual assumptions on ϵ_i are correct and ii) allowing that the variances may be changing over i in some unspecified way.

   (b) Give verbal interpretations for your fitted values for b_1, b_2 and b_3.

   (c) Interpret the three t-tests associated with b_1, b_2 and b_3. What are they testing and what are the conclusions?

   (d) What if any connection do the t-tests for the individual coefficients have to do with the 95% confidence intervals for the coefficients?

   (e) Interpret the F-test in the analysis of variance table. What is it testing and what is the conclusions?

   (f) Get the residuals from the fit above. (We could assess normality via a histogram and a test for normality as done before but this requires we have the rest of the model correct, so we'll skip that for now but you should know how to do it.)

   - Plot the residuals versus each predictor and versus the fitted value. Describe what you learn?

   - Plot the residuals versus each of the three products, X_1X_2, X_1X_3 and X_2X_3. A pattern in these residuals suggest an interaction. What do you conclude.

Remainder are for ST697R students to do and hand in. (ST505 students should take a look and are responsible for knowing how to do them, but I'm not asking you to do them out.)

(g) Now fit a model that has the three additional terms and the three products. What do you conclude about the product terms? Do the three individual tests for the products test that null hypothesis that none of the three products are needed (simultaneously)? We'll explore how to test that with the full-reduced/general linear test approach later.

3. Problem 6.22

4. Problem 6.16 b). Hint: Use Bonferroni's method.