Stat/HRP 262, Problem Set 3

Due Wednesday May 15, 2002

Please bring your problem sets to the class. If they are late, some points will be taken off. Late work must be handed to somebody, me (Art Owen), or a TA, or a secretary (Judi Davis or Angie Martinez). If nobody can be found, then put your late work under my door (room 130). Don’t place any work in a mailbox or other unsecure place.

Problems 1 and 2 must be done individually. The nzbp problems, 3 to 7 may be done in groups of up to 3 students.

The class web page contains the needed data and some documentation. The file patientsat.data contains survey results from a hospital, relating patient satisfaction to patients’ age, illness severity, and anxiety level. The file nzbp.data contains blood pressure measurements from two large surveys conducted in New Zealand.

1. For the patientsat data run 4 regressions to predict patient satisfaction. In one of the regressions, use all three predictor variables (Age, Anxiety, and Severity). The other three regressions are each on just one of the predictor variables. Report a $3 \times 2$ table of $p$-values. There should be a row for each of the three predictor variables. The first column is for the $p$ value of that predictor’s regression coefficient in a 1 predictor model. The second column is for the $p$ value of that predictor’s regression coefficient in a 3 predictor model. Are any of the variables significant at the 0.05 level both times? Are any of the variables insignificant both times?

2. For the kidneyfunction data report the $R^2$ for predictions of creatinine clearance based on a linear regression on age, weight, and serum creatinine. Report the $R^2$ for a model that only uses serum creatinine, ignoring age and weight.

Report the $R^2$ for a regression including the above as well as squares and cross-products of the above three variables. That is report the $R^2$ from a “full quadratic” model relating creatinine clearance to age, weight and serum creatinine.

Adding parameters to a model can never decrease $R^2$. Report the adjusted values $R_a^2$ for all three models. Recall that

$$R_a^2 = 1 - \frac{n - 1 \, SSE}{n - p \, SST},$$

1
where \( p \) is the number of regression coefficients in a model including the intercept, \( n \) is the number of cases observed, SSE is the sum of squared errors in the \( p \) coefficient model, and SST is the total sum of squares \( \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \).

3. For the \texttt{nzbp} data some of the data are missing. Say how many cases are missing, and what variables are missing. (There is so little missing data here, that it is safe to simply remove incomplete cases for the analyses below.)

4. For the \texttt{nzbp} data fit a model relating SBP to Age, Sex, and Src. Specifically, the model should allow different slopes and intercepts for men and women. There should be a different intercept for the two data sources. Report the regression coefficients, and which are statistically significant. Report the \( R^2 \).

5. Taking the previous model at face value: report the mean SBP for men and women at ages 25, 50, and 75. What is the yearly change in SBP for men, and what is it for women?

6. Run separate regression models, one for men, and one for women, on the \texttt{nzbp} data. Each model should have SBP as the response, and Age and Src as predictors. Do you learn anything from these two separate models that is not evident in the model from part 4?

7. For DBP run two separate regression models, one for men and one for women. Each model should be a quadratic in Age with a coefficient for Src. Report on differences between DPB for men and for women.