

STAT 516 - Spring 2002 - Homework 2

Due: Thursday February 7th, 2002

1) (7 points) A continuation of Problem 3, on Homework 1.

a) Construct an ANOVA table for this data set by hand. (You may use the by hand calculations from last time to help, and may use SAS to calculate any needed means and variances.) Use an F-table to test $H_0: \beta_1=0$ and $H_A: \beta_1 \neq 0$ at $\alpha=0.05$. Make sure to state your conclusion.

b) Comment on whether the various regression assumptions seem to be met (saying which plot you checked for each one). If an assumption doesn't appear to be met, or if you can't tell, briefly describe why.

c) One of the questions that Galton was interested in was whether or not there was a regression to the mean effect. That is, do the larger parent peas tend to produce offspring that are somewhat smaller than themselves, and do the smaller parent peas tend to produce offspring that are somewhat larger than themselves. This could be examined by testing the null hypothesis $H_0: \beta_1=1$ against the appropriate alternate hypothesis. What is the appropriate alternate hypothesis to test Galton's experimental hypothesis? Using the output contained in the "Parameter Estimates" box on the PROC INSIGHT output, construct this test of hypothesis by hand (use $\alpha=0.05$).

d) By SAS or by hand, what is the 95% confidence interval for predicting the average size of the offspring peas of a parent pea plant with peas of diameter 20.5.

2) (3 points) The data set on the web is from the first year that SAT scores were published on a state-by-state basis in the U.S. It was originally published in the *Harvard Educational Review* in 1984, and is also reported in Ramsey and Schafer, 1997. The variables included are:

sat = average total SAT score for the state

takers = percent of eligible students in the state who took the exam

income = the median family income of students in the state who took the exam

years = the average number of years that the test-takers had for studies in the core subjects

expend = the states expenditures on education in hundreds of dollars per student

rank = the median percentile ranking of the test-takers in their high-school class

Perform a multiple regression to predict the average SAT score in the state from the other variables, and answer the following questions. (Present copies of the relevant portions of the SAS output.)

a) Test whether the other five variables as a group are statistically significant predictors of the states average SAT scores. (Report the p-value and your conclusion)

b) Check the assumptions, and state whether you feel comfortable trusting the results of the regression.

c) What percentage of the variation in state average SAT score do the five variables explain?