

UNM Statistics Qualifying Exam Take-Home

Spring 2010

Due 12:00pm Friday January 15, 2010. Return to Roxanne Littlefield in the Math/Stat Dept Office.

Directions: The answer to each problem should be presented as a summary. It should be word processed and double spaced. An appendix is allowed for each problem but will be examined only at the discretion of the graders. The better organized your appendix is, the more likely it is to get examined.

You may **not** consult any other person when working on this exam or discuss your exam with anyone else regardless of whether or not the person is taking the exam. You may use your course notes as well as any available books or web resources for the exam. Questions pertaining to clarification about these questions can be directed to Curt Storlie, storlie@stat.unm.edu.

1. The data for this problem is available at http://www.math.unm.edu/~storlie/sat_data.txt. These data consist of average SAT score for each state in the US during 1994-95. The variables measured on each state are:
 - COST: current expenditure per pupil excluding teacher salary (measured in thousands of dollars per average daily attendance in public elementary and secondary schools).
 - RATIO: average pupil/teacher ratio in public elementary and secondary schools during Fall 1994.
 - SALARY: estimated average annual salary of teachers in public elementary and secondary schools during 1994-95 (in thousands of dollars).
 - PERCENT: percentage of all eligible students taking the SAT in 1994-95.
 - TOTAL: average total score on the SAT in 1994-95.

Using the data, build an appropriate regression model. Summarize your findings paying particular attention to the following questions:

- (a) Which of the covariates seem to be most related to SAT?
 - (b) Suppose it is argued by a state representative based on this data that increasing each teachers' salary by \$5,000 would also have the benefit of increasing average SAT score. Comment on this statement.
2. The data is available at: <http://www.math.unm.edu/~storlie/adapt.txt>. The 5 columns in the data represent the variables *potato*, *regime*, *temp*, *photo*, and *leak*, respectively, which are described below. This dataset is from an experiment on how plants adapt to cold climates. The investigators decided to study this problem after observing that plants that have been conditioned to cold previously appear to suffer less damage from the cold. Two species of potato were studied (species 1 and 2) which is represented by the *potato* variable. Each plant was exposed to one of two acclimatization regimes (*regime*) for several days (1=plant was kept in cold room; 0= plant was kept at room temperature). Later, plants were subjected to one of two cold temperatures (*temp*) (-4 degrees C is coded as 1; -8 degrees C is coded as 2). Two responses were measured: damage score for photosynthesis (*photo*), and damage score for ion leakage (*leak*). In this project, you will use photosynthesis

(*photo*) to be the response variable. Some of the 80 plants originally assigned to the treatment combinations were lost during the experiment. Your task is to analyze the data from the plants that made it through, and assess the effects of the three experimental factors species, regime, and temperature on the response *photo*. Make sure that your analysis includes appropriate displays, that you mention any unusual features of the data, and that you provide conclusions about your analysis. Be sure to have your analysis account for the unequal sample sizes in the different cells. Be sure to check the model assumptions such as constant variance, independence, normality and also check for outliers.