Orginal

# STATISTICS Ph.D. QUALIFYING /MS EXAM: TAKE HOME

## Due 2PM on Friday January 13, 2006. Return to Donna George in the Dept Office

*Directions*: The exam has two questions of equal value. Your answer to each problem should be word-processed, double spaced, and should be no longer than four pages. An appendix is allowed for each problem but will be examined only at the discretion of the graders. The better constructed your appendix, the more likely it is to get examined. Please use your ID number (last 4 digits of you Social Security Number) for identification on each page. Do not put your name on the exam.

---

1. The California Child Health and Development Study involved women on the Kaiser Health plan who received prenatal care and later gave birth in the Kaiser clinics. Approximately 19,000 live-born children were delivered in the 20,500 pregnancies. We consider the 680 live-born white male infants in the study. Data were collected on a variety of features of the child, the mother, and the father.

   The columns in the data set are, from left to right:

   1) ID
   2) child's head circumference (inches)
   3) child's length (inches)
   4) child's birth weight (pounds)
   5) gestation (weeks)
   6) maternal age (years)
   7) maternal smoking (cigarettes/day)
   8) maternal height (inches)
   9) maternal pre-pregnancy weight (pounds)
   10) paternal age (years)
   11) paternal years of education
   12) paternal smoking (cigarettes/day)
   13) paternal height (inches)

   A goal here is to build a regression model to predict child's birth weight from the data on the mother (including gestation) and father.

   Given your statistical analysis, provide a report that includes

   a. A discussion of the process you used to build the regression model, including relevant output.

   b. A discussion of which maternal and paternal features are useful for predicting child's birth weight, and whether the variables selected "make sense." Also, interpret the sign of the regression coefficients in the final model - i.e. which predictor variables are positively associated with birth weights (holding the other predictors constant), and which are negatively related.

   c. A summary table that includes the important predictors of birth weight, and other features (for example regression coefficients and standard errors, or p-values) that would be useful to report in a scientific paper.

   d. A summary of the analysis, including any potential limitations you might see with your conclusions.

   **The data are in http://math.unm.edu/~schrader/datasets/CHDS.xls**

2. This experiment studied the influence of four process variables on a decontamination process for removing radioactive isotopes from a liquid waste. The process employed scavenging agents with the ultimate object being to adjust operating conditions in such a manner that the major portion of the radioactivity became associated with the solid phases present. These could then be filtered off and disposed of by burial.

   The process variables were each studied at two different levels. They were

   B: Barium Chloride, 0.4 and 2.5g of $BaCl_2 \cdot 2H_2O$ per l.;
   A: Aluminum Sulphate, 0.4 and 2.5g of $Al_2(SO_4)_3 \cdot 18H_2O$ per l.;
   C: Carbon, 0.08 and 0.4 g of carbon per l.;
   P: Final pH of the solution (obtained by adding hydrochloric acid), 6 and 10.

   The reagents were added and the liquid stirred for a fixed period of time, then filtered and the alpha radiation of the remaining solution measured (in tens of counts per minute per ml of filtrate). The goal is to make the alpha radiation in the filtrate as small as possible.

   It was only possible to run and count eight tests in a single day. The experiment was run on four consecutive days, labeled blocks in the data set, with eight combinations in each block. Blocks 1 and 2 comprise one complete replicate of the experiment, and blocks 3 and 4 comprise the second replicate. Blocks 1 and 3 are duplicate blocks, and similarly for blocks 2 and 4. The data are presented in actual temporal order in which the runs were carried out. This order was chosen at random.

   **The data are in http://math.unm.edu/~schrader/datasets/radioact.txt**

   The experimental layout is implicit in the data file. A 0 indicates the lower level of the factor, a 1 the higher level.

   Prepare a report that includes

   a. An analysis of the confounding present in this design.
   b. Reasonable assumptions to make on interactions, in light of your answers to part a.
   c. An analysis of which factors and interactions contribute most to the remaining contamination. Include relevant contrasts.
   d. Recommendations for the client.