

Part I. (85 points) Do all calculations in SAS. Use a word processor of your choice to write a report. Insert computer text output and graphics to support what you are saying, but you need to write something that looks like an academic paper — not a pile of computer output. Turn in a hard copy of your HW in class (i.e., don't email me your HW).

(85^{pts})

1. CCHD birth weight: 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 subset of 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), y response
- 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 education (years)
- 12) paternal smoking (cigarettes/day)
- 13) paternal height (inches)

A goal here is to build a multiple regression model to predict child's birth weight (4) from the data on the mother and father (6–13). A reasonable strategy would be to:

1. Examine the relationship between birth weight and the potential predictors.
2. Decide whether any of the variables should be transformed.
3. Perform a backward elimination using the desired response and predictors.
4. Given the selected model, examine the residuals and check for influential cases.
5. Repeat the process, if necessary.

Given your statistical analysis, provide a writeup that includes the following parts.

- (a) (40 pts) A discussion of the process you used to build the regression model, including relevant output (e.g., make sure to do a residual analysis to check model assumptions).
- (b) (20 pts) A discussion of which maternal and paternal features are useful for predicting child's birth weight, and whether the variables selected by the backwards elimination procedure "makes sense." Also, interpret the sign of the regression coefficients in the final model, that is, which predictor variables are positively associated with birth weights (holding the other predictors constant), and which are negatively related.
- (c) (10 pts) 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) (15 pts) A summary of the analysis, including any potential limitations you might see with your conclusions.

85 pts

85 pts