ST503: Homework 01

Arnab Maity

Problem 1 (20 pts)

(A) Consider the *analysis of covariance (ANCOVA)* model:

$$y_{ij} = \mu + \alpha_i + x_{ij}\beta + e_{ij},$$

for $i = 1, 2, 3$ and $j = 1, \ldots, n$. Write the model in matrix form, clearly specifying all model components.

(B) Is the model matrix X full column rank? Explain.

Problem 2 (60 pts)

Consider the teen gambling data, `teengamb`, in the R package `faraway`.

(A) Write a brief description of the dataset. Produce some numerical and graphical summaries of the dataset.

(B) Fit a linear model using the `lm()` function with `gample` variable as response, and the `income` variable as predictors, and report the regression coefficients.

(C) Write the mathematical form of the model you fit in part (B). Clearly define each component in your model.

(D) Further numerical investigation: compute the mean and standard deviation of `gamble` and `income` for males (`sex=0`) and females (`sex = 1`) separately. Comment on the results.

(E) Fit the same linear regression as in part (B), but separately for male and females. Report the regression coefficients.

(F) Create a scatterplot between `gamble` (in y axis) and `income` (x axis), and color the points by `sex`. Then add two fitted regression lines from part (E) to the plot. Comment on the results.

Problem 3 (20 pts)

Consider the simple linear regression model

$$y_i = \beta_0 + x_i\beta_1 + e_i,\ i = 1, \ldots, n,$$

where the x variable has been centered and scaled so that $\sum x_i = 0, \sum x_i^2 = 1$.

(A) Write the model matrix, X

(B) Write the expression for $X^T X$, and solve the normal equations.