

HW01 - Statistical Modeling

MATH 456 - Spring 2020

Submission Instructions.

- [Right click this link to download the R Markdown template for this homework.] This template is required.
 - Rename this file with your username. I.e. `hw01-modeling-template.Rmd` becomes `hw01-modeling-rdonatello.Rmd`
- Knit to PDF and submit to Google Drive: HW01: Statistical Modeling/Draft folder by the due date.
- This is a peer reviewed assignment. After the PR period, you should update your assignment with corrections by your peers, and submit a final version into the HW01: Statistical Modeling/Final Google folder.

[Click here to download this file as a PDF]

Part I: Statistical Modeling

1. From the depression data set, fit a linear regression model on depression as given by `cesd`, using `income` and `age` as independent variables. (PMA6 problem 8.5, modified)
 - a. Analyze the residuals and decide whether or not it is reasonable to assume that they follow a normal distribution.
 - b. Interpret each coefficient (including the intercept) in full context of the problem. You must include the point estimate ($\hat{\beta}$), a 95% confidence interval for that estimate, and a p-value in your conclusion.
2. Does gender (`sex`) modify the relationship between age and depression? Don't forget to recode `sex` into a categorical variable before you begin. Answer this question in two ways: a) using a stratified model b) using an interaction model.
3. Which of the two models in question 2 assumes that the affect of income on depression is constant (does not change) between males and females?
4. For adult males it has been demonstrated that `age` and `height` are useful in predicting `FEV1`. Using the lung function data set, determine whether the regression plane can be improved by also including `weight`. Use two measures of model fit to justify your answer to this question (PMA6 problem 9.4, Ref PMA6 section 9.4). Note: all variables for adult males start with the letter `F` (`FACE`, `FEV1`, `FHEIGHT`, etc.)
5. For adult males, does weight *confound* the relationship between age or height and FEV1? You must show relevant model output and explain how you came to your conclusion.
6. Using the depression data set, fit a model of income using age, sex, educational level and religion as predictors. (PMA6 problem 10.3 modified)
 - a. Use a general F test to determine whether religion has an effect on income.
 - b. State the reference categories for both religion and educational level.
 - c. Interpret the coefficient for each level of educational level in full context of the problem, being sure to include the comparison group. You must include the point estimate ($\hat{\beta}$), a 95% confidence interval for that estimate, and a p-value in your conclusion.

Part II: Variable Selection

PMA6 problems 9.9, 9.11, 9.12, 9.13, 9.14