SMML Project 3

John Kubale

9/17/2024

- Note: For each question, include R code and output *pertinent* to your answers.

1. JWHT Chapter 2. Modified Exercise 10.
This exercise involves the `Boston` housing data set. Assume that we are interested in per capita crime rate, `crim`.

A. Examine `crim` with `summary()` and in a histogram.

B. Focus on suburbs with the crime rate above 25. How many suburbs fall into this group? What are the pupil-teacher ratios like in those suburbs? How about property tax rates? How about median home values? How do the pupil-teacher ratios, property tax rates and median home values compare between these suburbs and the remaining suburbs?

C. Create a scatter plot of the crime rates and the median home values for 1) all suburbs, 2) suburbs bounding Charles River, and 3) suburbs not bounding Charles River. What do you observe?

D. Analyze the crime rates as a function of median home values in a simple linear regression with an intercept. Report what the regression coefficients mean in lay terms.

E. Calculate the coefficients reported in D as well as their standard errors by hand.

F. Create a scatter plot of the crime rates and the median home values with a regression line. Is the regression line a good summary of the crime rates? Examine residuals to assess this.

G. Create a scatter plot of predicted `crim` (**x-axis**) and residuals (**y-axis**). What do you observe?