Homework 4

John Kubale

2024-09-23

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

This exercise involves the Boston housing data set in ISLR2. Assume that we are interested in median home values, medv.

1. Examine `medv` as a function of `crim`, `zn` and `indus` in a multiple linear regression.

A. Identify the predictors which are "statistically significant" at $\alpha = 0.05$.

B. List the null and alternative hypotheses tested in 1A and your conclusions.

C. Interpret each of the regression coefficients as if it were the primary exposure of interest. Do they make sense?

D. It's generally not good practice to interpret all predictors as if they were the exposure of interest. Why do you think doing so could be problematic?

E. Construct and interpret 95% confidence intervals for $\hat{\beta}_{crim}$, $\hat{\beta}_{zn}$, *and* $\hat{\beta}_{indus}$ (you do not need to calculate them "by hand"). How does the confidence intervals correspond to the hypotheses tested in 1A and 1B?

F. Calculate R^2 and R^2_{adj} "by hand" (you can use helper functions from R to get the components needed for the formula, but do not simply extract it from the model object). What do they mean?

2. Fit a simple linear regression model with `medv` as a function of `zn` and compare it to the model from question 1 using the global F test and one other method. Which model do you prefer based on the results of the comparison?