BA222 Pset 4

Prof Leder-Luis

Due: Wed Nov 22 2023 12PM ET
No Late Psets

1 Regress and Analyze

This problem set has 1 long question. Please do the following, and submit your answers as a PDF. You do not need to include your code.

Load the "King County Housing.csv" data frame into Pandas. We will analyze the relationship between price and the number of bedrooms in a home.

a) Where is King County? Use the zip codes if you are unsure.

b) How many observations are in the dataset? What does 1 row correspond to?

c) What are the median statistics for price, bedrooms, bathrooms, square foot of living space, and year built?

d) Run the regression:

$$Price = a + b * Bedrooms$$

- Write a full sentence explaining the coefficient on bedrooms.
- Is the coefficient statistically significant? What is the 95% confidence interval on the coefficient on bedrooms? Interpret the interval.
- If a house has 2 bedrooms, what does the one variable model predict the price will be?
- Is the relationship between bedrooms and price necessarily causal?
- Interpert the R^2 value of this model.

e) Run the regression of price on bedrooms and living square footage

$$Price = a + b * Bedrooms + c * Sqft\_living$$

- Write a full sentence explaining the coefficient on bedrooms. How has it changed? Why might it have changed?
- How has the R^2 changed from the first model?

- What does the model predict for the price of a 2 bedroom, 1000 square foot apartment?

  What does the model predict for the price of a 3 bedroom, 1000 square foot apartment?

f) Add dummies for zip code to your second model and run the regression:

$$Price = a + b * Bedrooms + c * Sqft\_living + d * Zip$$

You should have 70 zip dummies. You do not need to interpret them, just include them.

- What is the R^2 of this model? Write a full sentence.
- What is the coefficient on bedrooms? How does it compare to the other models? Is it statistically significant?
- Suppose we wanted to use this model to make a casual statement about the effect of bedrooms. Write a full sentence about the assumption we would have to make.

g) Run one more model to evaluate the effect of bedrooms on price, picking some other variable(s) for controls. What variables did you include? Write the full estimating equation, and include a screenshot of your results. What coefficient for bedrooms do you find?