

In-Class Team Project 1

ECO391

1. The AnnArbor_Rental data set shows a portion of data consisting of the rent, the number of bedrooms, the number of bathrooms, and the square footage for 40 apartments in the college town of Ann Arbor, Michigan. You are given the task of predicting the rent of an apartment on the basis of the number of bedrooms, the number of bathrooms, and the square footage for 40 apartments.
 - a. Name the response variable.
 - b. Name the explanatory variables.
 - c. Use Excel to find the sample correlation between rent and square footage and report it in the space below.
 - d. Specify the competing hypotheses in order to determine whether there is a positive population correlation between rent and square footage.
 - e. Calculate the value of the test statistic and its approximate p-value. At the 5% significance level, what is the conclusion to the test?

f. Write down the population regression model.

g. Write down the sample regression equation.

Now, estimate the model using the regression option on Excel.

h. Write down the sample regression equation using intercept and slope coefficients from the regression output.

i. Interpret the intercept and slope coefficients.

j. Do you think the actual rent is equal to predicted rent for every apartment? Why?

- k. Predict the rent for a 1500-square-foot apartment with 2 bedrooms and 1 bathroom.
 - l. Using the regression output, report the standard error, R^2 , and the Adjusted- R^2 .
 - m. In this case, is it okay to say that this model is a good fit just by looking at the standard error?
 - n. Interpret R^2 and Adjusted- R^2 .
-
- 2. Is it defense or offense that wins football games? Consider the Football data set which contains a portion of data, which includes a team's winning record (Win), the average number of yards gained, and the average number of yards allowed during the 2009 NFL season.
 - a. Write down two simple linear regression equations, where Model 1 predicts the winning percentage based on Yards Gained and Model 2 uses Yards Allowed and compare their standard errors.

- b. Note down the values of three goodness-of-fit measures from Models 1 and 2.

Estimate the multiple linear regression model, Model 3, that applies both Yards Gained and Yards Allowed to forecast the winning percentage.

- c. Can you use all three goodness-of-fit measures to compare between the three models? Why or why not?

- d. Is this model an improvement over the other two models? Explain using appropriate goodness-of-fit measures. Which is the model with the best fit?

3. Does correlation necessarily imply causation? Explain.