STAT 2600 Recitation
10/27/2020

Recitation Week 10 Problems

Today we will be drinking wine, feeling fine. Just kidding, we're doing price hedonics on wine!! Thats long for "regressions on price." What characteristics about a bottle of wine do you think are most indicative of how expensive that bottle is? Is it the country its from, the region the grapes were grown, the type of grape, the critics reviews, the lay person reviews? Or some combination of them all? To begin, download the Wine_Hedonics_Data.csv file and import it into an R markdown file.

1) 1 MINUTE QUESTION: Get to know your data (i.e. run data diagnostics) - what variables do you have? what levels do they have (ie what countries, regions, etc.. are in the data)? are there missing values we need to worry about? (hint: use str( ), count ( ), summary( ), levelsof( ), or just look at it in excel

2) Make a prediction - based only on reasoning, which variable do you think will best predict the price of the wine?.

3) Run a single variable linear regression using your answer for (2) as your independent variable. This will have the form $\hat{y}_i = \beta_0 + \beta_1 * x_{1_i} + \epsilon_i$. Interpret your results: what is the coefficient of regression, β_1? Is it significant? how good was the fit of your regression (the R^2)? Challenge: what does the intercept term, β_0 represent, what does the R^2 term represent?

4) Now lets see if you were right: run a multivariate linear regression on price. This will have the form $\hat{y}_i = \beta_1 * x_{1_i} + \beta_2 * x_{2_i} ... + \beta_n * x_{n_i} + \epsilon_i$. Interpret your results: what are each of the coefficients of regression? are they significant? how good was the fit of your regression?

5) Were you right?

6) Bonus: what are wine critics good for anyways? are they better predictors of "expensive" bottles of wine than average reviewers?