**Discussion 8: Regression**

1. What does the dependent (y) variable measure? What do the independent (x) variables measure? What does the regression constant measure?
2. What does the R^2 (r-squared) coefficient measure? Why is it important?
3. What is a p-value, the most common p-value of interest, and why does it matter if a variable’s p-value is below that common value of interest?
4. How do we know when variables are insignificant? If insignificant, what does that mean for our interpretation of the variable’s coefficients?
5. Interpret the coefficients (marked Coef.) being careful to consider their p-values as well. If you are unfamiliar with this regression output; the y-variable is “lprice”, the x-variables are “y81, ldist, y81ldist”, and the regression coefficient is “\_cons”.

Table

Description automatically generated

*(source: https://i.stack.imgur.com/phgkm.png)*

1. What is overfitting? How do we know if our model is overfitting (hint: think of training and testing accuracy)?
2. Which of the following models would you feel the most comfortable using and why?

|  |  |  |
| --- | --- | --- |
| Model | Training Accuracy | Testing Accuracy |
| A | 55% | 85% |
| B | 70% | 65% |
| C | 95% | 55% |
| D | 50% | 50% |

1. What is the difference between Lasso Regression and Linear Regression?
2. Interpret the regression output below including the R^2 coefficient as well as the dependent variable’s and regression constant’s coefficients and p-values.

Table

Description automatically generated

*(source: https://www.academicianhelp.com/ckfinder/userfiles/files/fig1(2).png)*