

NYC Data Science Bootcamp

Simple Linear Regression

Question #1: Anatomical Data from Domestic Cats

Load the cats dataset from the MASS library. This dataset includes the body and heart weights of both male and female adult domestic cats.

- 1. Create a scatterplot of heart weight versus body weight. From this plot alone, do you think simple linear regression would be a good fit for the data? Why?
- 2. Regress heart weight onto body weight. For this model:
 - a. Write out the regression equation.
 - b. Interpret the meaning of the coefficients in context of the problem.
 - c. Are the coefficients significant? How can you tell?
 - d. Is the overall regression significant? How can you tell? How does the answer to part c. relate?
 - e. Find and interpret the RSE.
 - f. Find and interpret the coefficient of determination.
- 3. Add the regression line to your plot from part 1.
- 4. Add a visualization of the residuals to your plot from part 3. Do any of the residuals seem abnormally large?
- 5. Construct 95% confidence intervals for the model coefficients. Interpret the intervals in context of the problem.
- 6. Assess each of the assumptions of the model.
- 7. Redraw the scatterplot and regression line from part 3 and add both confidence and prediction bands.
 - a. Why is the prediction band wider than the confidence band?
 - b. Why does the confidence band widen as it travels away from the center of the regression line?
- 8. Construct confidence and prediction intervals for body weights of 2.8 kg, 5 kg, and 10 kg. Do you foresee any issues with reporting any of these intervals?

Question #2: Considering Transformations

Continue with the cats dataset from the MASS library.

- 1. Create a Box-Cox plot for transforming the cats regression you created in question 1 above.
- 2. Determine the best value of lambda for a Box-Cox transformation on the cats regression. (Hint: Try to balance interpretability and accuracy; when taking this perspective, there isn't really a completely "correct" answer.)
- 3. Transform your data based on your answer to part 2.
- 4. Construct a new regression now using the transformed data.
- 5. Create a scatterplot of the transformed data and overlay the new regression line.
- 6. Inspect the summary information and validate the assumptions of the linear regression model. Is there anything to be concerned about in the new model?
- 7. Compare the models you created:
 - a. Give one reason why you might use the original model instead of the Box-Cox transformed model.
 - b. Give one reason why you might use the Box-Cox transformed model instead of the original model.
- 8. Attempt to apply a Box-Cox transformation on the model on which you already applied a Box-Cox transformation. What happens?