Project Step 2: Simple Linear Regression

Saad Mouti

Goal

Your task for the SLR portion of the project is to apply the tools to simple linear regression in order to answer questions about the relationship between two continuous (quantitative) variables.

Due date is Sunday November 5th 2023

The analysis should include

Introduction (briefly refresh the reader's mind as to the variables of interest). Remember that you should include a reference for the original data source, and the reader should know to what population you are inferring your results.

- The hypotheses that you'll be addressing. It will probably be that the two variables are linearly related. (Positively? Negatively? Remember, R gives a two-sided p-value, but you can just as easily test that $\beta_1 > 0$ or $\beta_0 < 0$ by dividing the p-value by 2.)
- Check the assumptions for linear regression. Look at plots of explanatory vs. response and residual vs. predicted (include only what is interesting in your report). Comment on whether you think the data are linear with constant variability (some of the residuals analysis is still to be seen in the next couple of weeks but feel free to do some research if you want to do it before then. It's fun.). If not, try transforming the data. Remember, transforming X gives a different relationship between X & Y (might make the relationship more linear); however, transforming Y changes the variability around the line (might make the standard deviation more constant and the relationship different.)
- Compute the test of β_1 (or other test from above) or find a CI for β_1. Remember that if you have transformed data, you should be careful about your interpretations. Your test or CI should include an interpretation in the words of your variables.

- Plot your (transformed?) variables, try to think of one as explanatory and the other as response. Give the reader a CI for both the mean and individual response at some interesting value of the explanatory variable. (That is, at some x-value that is interesting to you.) Interpret these intervals.
- Assess the fit of your model. Discuss the R^2 value and the residual plot(s). Remember that residual plots (not R^2) determine whether a linear model is appropriate.
- A Conclusion (Summarize your results. Comment on anything of interest that occurred in doing the project. Were the data approximately what you expected or did some of the results surprise you? What other questions would you like to ask about the data?)

Format

Do:

- Use captions for every plot, e.g., in the chunk command give the caption: "'{r fig.cap = "here is the caption"}.
- Use complete sentences.
- Annotate everything that the reader sees.
- Keep the file to 4 or fewer pages (mostly graphics).
- submit in both the .Rmd and .pdf file on gradescope. Make sure the .Rmd file can knit!

DON'T:

- Print any warning or error messages. Only print code that is interesting and relevant to the reader (e.g., use echo=FALSE).
- do not print lists of data.
- no overplotting (use boxplots instead of scatterplots when appropriate; use alpha=0.1 for transparent plotting symbols).