Linear Regression Final Project

Tom Wilson

October 13, 2018

Literature Review (completed previously)

Introduction – briefly refresh the reader's mind as to the variables of interest

The regression model

Use at least 4 explanatory variables

create a pairs plot on the explanatory and response variables.

Are any of the explanatory variables highly correlated?

Is there any reason to fit a quadratic term?

do a log transformation?

Fit your model. Include quadratic, log, or interaction terms as you see fit.

Interpret your beta coefficients to the best of your ability.

Are your coefficients significant?

Use the F test to compare two nested models.

The smaller model should have at least 2 fewer variables than the larger model.

Report the R2 and Adjusted-R2 values.

Comment on the fit of the model as determined by how much variability is explained.

Is this a guarantee that the model will accurately describe the population?

A complete analysis of the residuals.

Use plots to get an idea of which points may be contributing to the fit.

Consider re-fitting a model with and without certain data that have both high leverage and large residuals.

** Do not include every plot, but consider including plots that give the reader an idea of your analysis. **

interpretation of the model that makes sense.

Why do you think some variables stayed significant and others dropped out?

Are any of your variables highly correlated

Give CIs for a mean predicted value and a future predicted value for at least one combination of X's (from your final generalized linear model).

Summarize your report.

Count the number of total hypothesis tests that you ran (including all the ones you didn't include in the report). Call that number m. If you multiplied every single p-value in this report by that number, would any of your conclusions / analyses have fundamentally changed? Which ones? How?

Format

- Word or pdf
- Only include code that is interesting to the reader.
- Do not print lists of data.
- If you are commenting on the significance of a variable in your text, you should report the p-value.
- Residuals determine model appropriateness, not p-values or R2.
- Summarize any output from R; do not include technical calculations.
- Use complete sentences.
- Make sure sections flow nicely into one another.
- Remember to label all graphs.