Simple Linear Regression

Goal (+18 of 20)

Your task for the SLR project is to apply the tools tofsimple linear regression in order to answer questions about the relationship between two continuous (quantitative) variables. After the report is turned in, your pair (or you solo) will assess a different project based on the questions below. It is in your best interest to read those questions before writing up your analysis! The authors will be asked for a one-sentence reflection on how helpful the review was.

The report 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 $\beta_1 < 0$ 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. 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.

- Asses 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

There are a series of tasks above, make sure the sections flow nicely into one another. You should create a report on the data not a homework assignment. (Try to tell a good story.) You do not need to answer the questions above in any order, and certainly not with bullet points or enumeration.

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.

- be succinct, report shouldn't be very long (maybe 4-5 pdf pages?).

- remember things we've learned: e.g., provide the reader with residual plots which are most informative.

- be very careful with the difference between individual prediction intervals and mean (average) intervals.

- use appropriate wording. E.g., a p-value is a probability of the *data...* the relationships you are testing are *linear...*

- push both the .Rmd and .pdf file to Git. Your .Rmd file must compile to .pdf. **In order for the file to compile, the data must live in the GitHub repository!**

Don't:

- do not print any warning or error messages. Only print code that is interesting and relevant to the reader (e.g., use `echo=FALSE`); maybe don't include any code at all?

- do not print lists of data.

- no overplotting (use boxplots instead of scatterplots when appropriate; use `alpha=0.1` for transparent plotting symbols).

- do not include any tables, output, or graphs which are unannotated.

- do not be tempted to turn in everything you do. Only turn in the interesting parts of the analysis. One of the hardest parts of being a consultant is figuring out what to tell the client.

**Peer Assessment**[1]

**Due: Thursday, March 3**

Critically reviewing other's work is an important part of the scientific process, and we will practice that evaluation in Math 158. Each pair has been given read access to the GitHub repository of a different project.

**Reviewing the draft**

Carefully read the SLR project. Consider the questions below as you read it. You will submit your review by creating new Issues in the team's GitHub repo. You may choose to do the assessment together (as your pair), or you may choose the divvy up the assessment (e.g., one person respond to Issues 1 & 2, the other person respond to Issues 3 & 4). To respond:

1. Go to the team's repo and click Issues.

2. Click New issue.

3. You will see several options that begin with "Peer review". Click Get started and it will open a new issue.

4. Type your response under each question header.

   - If you responded Yes, briefly summarize the answer from the draft. For example, if you answer yes that the draft includes citations for outside research, briefly summarize what that outside research is.
   - If you responded Somewhat or No, briefly summarize what is incomplete or inaccurate. In other words, briefly summarize why you did not respond Yes to that item.

**Issue 1: Introduction + Data**

- Is the research question and goal of the report clearly stated?

- Does the introduction provide appropriate background context and motivation for a general reader? This includes citations for any claims or previous research mentioned.

- Is the original source of the data stated and cited?

- Is it clear when and how the data were originally collected?

- Are the observations and variables that are relevant to the analysis clearly described? At a minimum, the observations, response variable, and predictor variables in the final model should be clearly described.

- Include any additional comments or suggestions on the introduction and data description.

---

[1]Thanks to Maria Tackett at Duke University for much of the structure and content ideas for peer reviewing.

**Issue 2: Exploratory data analysis**

- Is the data cleaning and data wrangling process clearly described? This includes how the group handled missing data, created new variables, reduced the number of levels for categorical variables, etc.

- Do the visualizations follow the guidelines above? This includes using plots that are appropriate for the data, having proper axis labels, titles, captions, etc.

- Are any tables and figures clear, effective, and informative? Are they neatly printed with a reasonable number of digits displayed?

- Should any visualizations, figures, or tables be eliminated, or are there any new visualizations, tables, or figures that should be added?

- Include any additional comments or suggestions for the exploratory data analysis.

**Issue 3: Methodology + Results**

- Are the methods described in enough detail that the work could be replicated by someone else? Is it clear what approach and model were used to evaluate hypotheses of interest? If not, point out areas for further work.

- What type of diagnostic methods were used to check any modeling conditions, and are you satisfied the conditions of the model are valid? Should any additional analyses be performed?

- Does the report contain a correct and effective interpretation of the results provided? Is all information needed to substantiate the results and conclusions included? If not, point out areas for further work.

- Are the conclusions valid for the data at hand? Is it clear to whom the results generalize?

**Issue 4: Presentation + general comments**

- Is the paper professionally presented and generally free of distracting errors or other issues, including (but not limited to) insufficient organization or formatting; poor grammar, spelling, or punctuation? Is the overall paper easily readable for someone with your expert level of knowledge? Note any concerns here.

- What is one question you have about the data and/or analysis that isn't yet addressed in the report?

**Applying to your project**

Discuss the following as a group. You **do not** need to submit a response to this question.

- After giving feedback to this group, what is one thing you want to change or continue working on for your report?

**Peer Review Grade (+2 of 20)**

The peer review will be graded on the extent to which it comprehensively and constructively addresses the components of the partner team's report: the research context and motivation, exploratory data analysis, reproducibility, and any inference, modeling, or conclusions. The authors will be asked whether or not the review was constructive for their project. You will be graded based on the submitted issues on GitHub.