Linear Regression and Modeling

Data Analysis Project Checklist

This checklist is intended to help you critically evaluate your Data Analysis Project. Remember that the more thoroughly you evaluate your project, the more useful this exercise will be to you and will set you up for success in future research projects.

Format

- [ ] Did you use the R Markdown to complete your project?

Data

Generalizability: Describe how the observations in the sample are collected, and the implications of this data collection method on generalizability.

- [ ] Did you describe the sampling method, mentioning that a random sample was used?
- [ ] Did you decide that the results can be generalized to movies produced and released before 2016?
- [ ] If you discussed potential sources of bias, think about how these affected your conclusions.

Causality: Describe how the observations in the sample are collected, and the implications of this data collection method on causality.

- [ ] Did you decide that no random assignment was used, and hence causality cannot be inferred?

Research Question One

- [ ] Is the research question phrased in a non-causal way?
- [ ] Is the research question well defined / not vague?

Note: "Well defined" means it is obvious from the research questions which variables will be involved in the analysis.

Is it clear why this research question is of interest to you and/or the reader? ☐

Exploratory Data Analysis

Plots

Do the plots address the research questions? ☐

Note: *There is no requirement on the minimum number of plots to be provided. A single plot can be sufficient, as long as it addresses the research question, or multiple plots may be needed.*

Are the plots properly constructed? ☐

Are the plots formatted well? (Size not too large, not too small, etc.) ☐

Summary Statistics

Do the summary statistics address the research question? ☐

Are the summary statistics calculated correctly? ☐

Are the summary statistics formatted well? (Not taking up pages and pages, etc.) ☐

Narrative

Is each plot and R output accompanied by a narrative? ☐

Does the narrative interpret the visuals and summary statistics correctly? ☐

Does the narrative address the research question? ☐

Modeling

Multiple linear regression model for predicting a numerical variable from various explanatory (numerical and categorical) variables.

Are the variables to be considered for the full model clearly specified? ☐

Is reasoning for excluding certain variables (that are in the dataset) from the full model included, and does the reasoning make sense for your research question? ☐

- [ ] Is reasoning for the choice of model selection method (adjusted R-squared vs. p-value) included, and does the reasoning make sense for the goals of the project?
- [ ] Is the model selection carried out correctly?
- [ ] Are the model diagnostics verified?

Note: *it is ok if all conditions do not check, what is important is that you performed the correct diagnostics (correct plots, etc.) and interpreted them correctly.*

- [ ] Are model coefficients (not all, but especially those relating to the research question) interpreted correctly and in context of the data and the research question?

Prediction

Pick a movie from 2016 (a new movie that is not in the sample) and do a prediction for this movie using the model you developed and the `predict` function in R. Also quantify the uncertainty around this prediction using an appropriate interval.

- [ ] Is the prediction done correctly?
- [ ] Is a prediction interval around this prediction provided?
- [ ] Are the prediction and the interval interpreted correctly?
- [ ] Are references provided for where the data for this movie come from?

Conclusion

A brief summary of your findings from the previous sections **without** repeating your statements from earlier as well as a discussion of what you have learned about the data and your research question. You should also discuss any shortcomings of your current study (either due to data collection or methodology) and include ideas for possible future research.

- [ ] Is the conclusion not repetitive of earlier statements?
- [ ] Does the conclusion provide a cohesive synthesis of findings that appropriately address the research question stated earlier?
- [ ] Are shortcomings and limitations of the study discussed?

Overall Project Details: Organization/Readability

- [ ] The document follows the organization of parts outlined in the template.
- [ ] The narrative uses correct grammar and clearly and succinctly addresses the research question.
- [ ] The code is clear, readable, well organized, and uses syntax and packages taught in the course.