Linear Statistical Models

SYS 4021

Project 1

Analysis of Train Accidents in the U.S. During 2001-2015

Laura E. Barnes
lbarnes@virginia.edu

Summary

Insert the summary of your recommendation(s) supported by evidence. Your summary must include a statement of statistical confidence in the evidence. The total length of your summary should be less than 500 words.

Honor Pledge: On my honor, I pledge that I am the sole author of this paper and I have accurately cited all help and references used in its completion.
You must include this honor pledge with your submission.

1 Problem Description

1.1 Situation

Provide a summary of the current safety situation as the context for your analysis. Include graphics and tables that provide information about the extent and characteristics of the problem. At a minimum you should reference the assignment [1] and this template [2].

1.2 Goal

Write the goal of this study. If you want you can decompose the goal into subgoals and objectives.

1.3 Metrics

Give the quantitative measures of performance against the goals.

1.4 Hypothesis(es)

Write at least one well-formed hypothesis for each goal.

2 Approach

2.1 Data

Follow the guidance given in class about documenting the data you used for this analysis. Be sure to cite the sources of the data you used in this study [3]. Also indicate any missing data or biases you observed. Explain your actions to correct for missing data or bias.

2.2   Analysis

Describe the analysis techniques and modeling approaches you applied to this problem. At a minimum this should include the following:

- The linear models you used to provide evidence;

- The feature and model selection techniques you used to find appropriate models for this problem;

- Your treatment of ordinal and categorical variables (i.e., how were they coded);

- How you assessed your models (e.g., adjusted R^2, AIC, etc.);

- How you diagnosed problems with the models; and

- How you adjusted the models based on these assessments.

3   Evidence

How do you answer the questions posed your goal and hypotheses? What evidence do you have that support your answer? In particular, describe your results from the application of the methods in 2.2 to the data in 2.1 to produce evidence for testing the hypothesis(es) in 1.4. These results should answer the question posed by the problem in 1.2. You should also formally describe your confidence in the results using confidence intervals.

4   Recommendation

Summarize your finding and your recommendations for safety improvements based on the evidence you have discovered. Be sure to back-up your recommendations with formal measures of confidence (i.e., confidence intervals), model validity (e.g., adjusted R^2) and possibly visual displays. You do need to repeat your results or graphics in this section. Instead you can summarize them and refer the reader to the appropriate sections, tables, or figures.)

5   References

[1] L. E. Barnes, "Project 1: Train accidents," *Class project in SYS 4021*, 2015.

[2] ——, "Project 1 template," *Class template in SYS 4021*, 2015.

[3] F. R. Administration. (2015) Federal railroad administration office of safety analysis. http://safetydata.fra.dot.gov/.

A Optional Appendices