Biostatistics 5300

Data analysis project

Due Nov 27, 2017

This project is a full analysis of a dataset using the tools we have learnt throughout the course. The project should have the following parts.

1. Introduction:

In this section, you should give a description of the study underlying their dataset. Possible questions to be answered are the following:

(a) What field does the data come from?

(b) What are the goals of the study? Are there any effects of particular interest?

(c) How might these goals be answered, i.e. tests / confidence intervals?

2. The data:

In this section, you should describe the data set.

3. The methods:

In this section describe the descriptive and inferential statistical methods you intend to use and why.

4. The results:

In this section, you should report their results obtained by performing the proposed methods in the previous section. Emphasis should be placed on clarity, as if the report were a statistical consultant's report for a non-statistician. For instance, loads of SAS output would, in general, not be acceptable. Plots and well-organized tables are good things to have in this section.

After running analyses, create tables and/or graphs that show the results to the reader. Place a title that begins with a unique table number over each table body; for example, “Table 5. Hypothesis Testing Results Comparing Education with Success.” Tables should be numbered sequentially and appear in numerical order. If you are producing a single table for each variable’s descriptive statistics (using four variables) and a single table for each hypothesis test’s inferential statistics (testing three hypotheses), the project will have 7 tables.

After producing tables that show the outcomes of the statistical procedures, write a narrative about the results. The narrative describes the sample (interprets the descriptive statistics for each variable) and explains the outcome of each hypothesis test (whether or not the hypotheses are supported). The narrative interprets the statistics and provides logic for determining whether or not each hypothesis is supported. The narrative explicitly refers to each table by number when its statistics are discussed.

To summarize, the results section includes tables and a brief narrative. These document the outcomes of the statistical procedures that (1) describe the sample’s generalizability and attitudes and behaviors (descriptive statistics) and (2) test the hypotheses (inferential statistics).

5. Conclusions and Discussion:

In this section, discuss the general conclusions you can draw from your results. It summarizes major findings in plain terms, tells the reader how knowledge about the topic has changed or whether it has not changed, states social implications of the findings, assesses weaknesses of the study (such as limitations of the sample or problems with question phrasing), and gives advice and helpful comments to researchers of the topic about how to proceed with their future research.

6. Appendix (SAS Code):

In this section, you should attach a final, edited copy of the SAS code used in the analysis.

7. Acknowledgements/References:

If you consult outside sources that refer to this data set, you should cite these as references, and describe what you used from each source. Sources include material found on the internet, journal articles and books.