P6110: Statistical Computing with SAS

Final Project

Due: May 09, 2019 by 5:00 pm

- This is an open source project: You can use any resources or tools (e.g. lecture notes, SAS codes, books, internet) you have, but collaborating with others is strictly prohibited.

- Please use Canvas to submit 1) Final poster (PDF) and 2) SAS program.

- The poster may include, but not limited to, the following sections:
  1) Abstract: Summarize the whole study.
  2) Objective: Introduce the motivation and objective of the study.
  3) Method: Provide the summary of dataset and describe the statistical methodology.
  4) Result: Report the main findings (descriptive statistics, relevant hypothesis tests, final model and their interpretation).
  5) Conclusion [Discussion]: Summarize the analysis and state its values and limitations.
  6) Bibliography: If applicable, introduce some references.

- A basic poster template is provided on Canvas. You may prefer to use your own template. The final version of the poster should be converted to a PDF file for submission.

- Any comments or interpretations can be included in the SAS code.

Cardiovascular Disease

The World Health Organization has estimated that 12 million deaths occur worldwide every year due to the heart disease. In the United States, cardiovascular diseases are the number 1 cause of death in adults. Early diagnosis is the key to prevent heart related deaths. Potential cardiovascular patients are often sent for multiple tests, and these test results help the doctors to make a diagnosis. The accuracy of the diagnosis often relies on the individual doctor's knowledge and experiences. In order to improve the accuracy of the diagnosis, medical researchers collected patient data from 4 national hospitals; two in the United States, and two in Europe. The data set 'HD.xlsx' includes the patients age, gender, 11 test results, and final diagnosis. The patient data from the two US hospitals and two European hospitals are saved in the 4 tabs, labeled "US1", "US2", "EU1", "EU2", respectively. The dictionary of the variables is saved in the tab "Dictionary". Assume that no patient within is correlated with other patients within hospital.

The doctors wish to develop statistical models based on the collected info to use the medical test results to help predict the likelihood of heart disease. As described in the data dictionary provided in the data set, the heart disease outcome measure 'diag' was collected and coded as 0 = 'No heart disease', 1 = 'One major vessel with > 50 % diameter narrowing', …, 4 = 'Four major vessels with > 50 % diameter narrowing'.

In order to find the factors that significantly contribute to the accuracy of diagnosis, the researchers want to 1) conduct appropriate hypothesis tests and 2) fit a proper model. For the heart disease outcome measure 'diag', they would like to try two different approaches:

a) Build a model with the severity of heart disease (i.e. the number of narrowing vessels) as response variable.
b) Dichotomize the ordinal variable (Yes / No) and consider the presence of narrowing vessel as response variable.

Create descriptive statistics (table/plot) for all relevant variables. You may not include every step of your hypothesis testing or model selection in the poster, but make sure to check assumptions for conducting any hypothesis test and fitting any model. Interpret the final models, and discuss the findings. Interaction terms can be included if considered reasonable.

Use of ods rtf report;
limitation and contribution should be in the conclusion part.

add some literature review
Also the font size should be at least 22/26