Outline for Data Analysis Project

The purpose of this project is to apply what you’ve learned about descriptive statistics, graphical methods, normal approximations, correlation and regression, and hypothesis testing to your dataset of interest.

Step I

- Begin searching the net and other sources for a dataset that captivates your interest. See the course website for dataset ideas, however if you still have trouble finding a data source that tickles your fancy, ask me for assistance. You may also collect your own data using a survey or by conducting an experiment. Your dataset should contain at least 30 observations and 10 to 20 variables. Variables should include qualitative and quantitative (both discrete and continuous).

- Decide if you’re going to work individually or in a group (no more than 4 students to a group). Then, confirm your choice of dataset and individual/group work with me no later than Monday, June 3. Explicitly, you will hand in a paper stating the dataset (including where you found it and a list of variables contained in it) by the above date.

Step II

- You’ll want to begin your project with a general research question, ie an underlying theme of the project—something you’d like to set out to find: relationships between variables, for example. You should have an introduction to your project which includes your research question (including any reasons for being interested in your topic) and a list of the variables and what they represent.

Step III

- The next step is to construct a separate page for each topic to be covered. These topics include:
  - graphs (histogram, box-plot, stem-and-leaf),
  - normal approximation,
  - measures of location and variation (means, modes, medians, ranges, and standard deviations),
  - standard error, confidence intervals, and H₀: Mean=x,
  - scatter plot,
  - contingency table and H₀: Independence,
  - correlation scatterplot matrix and H₀: Correlation=0,
  - regression and residual plots, and
  - two-sample tests.

- Additionally, each page should consist of the overall research question at the top, the research question for each topic, a list of statistical methods used to obtain results, statistical output, an explanation of the output analysis, and conclusions. This does not mean submitting formulas, but rather an interpretation of your results. You must use JMP IN for this project.
Step IV

- Finally, a 1-2 page conclusion and discussion is required. This should consist of a summary of what you’ve learned about your research question along with statistical arguments supporting your conclusions. Furthermore, critique your own methods and provide suggestions for improving your analysis, including a paragraph explaining what you’d do differently if you had to start over or what you’d do next if you continue your study in the future.

Step V

- Pay close attention to your presentation. Neatness, clarity, and coherency will count! Be creative, and utilize your critical thinking skills!

Step VI

- Attach your dataset to the project and submit it in class on Monday, June 24. Presentations will be during the class and lab periods on this day (the last day of class); and, depending on how many students we have working in groups will determine how long each “group” will have to present his/her/their findings! Presentations will be graded on clarity (this includes volume..speak loudly), visual aids, and overall presentation skills.

Step VII and Most Important

- This should be a fun project! The key is to choose a dataset that interests you: you’ll be amazed at what you learn!