MATHEMATICS 156/E-156, FALL 2014
MATHEMATICAL FOUNDATIONS OF STATISTICAL SOFTWARE
Final Project Guidelines

Last modified: November 24, 2014

**Required technical elements - the dataset - 4 points**

1. A dataframe.

2. At least two categorical or logical columns.

3. At least two numeric columns.

4. At least 20 rows, preferably more, but real-world data may be limited.

**Required technical elements - analysis - 6 points**    Any two of the following - 3 points each

1. Linear regression

2. Student t confidence interval

3. Bayesian prior updated by data

**Required technical elements - graphical display - 2 points**    Any two of the following

1. A scatter plot with a regression line.

2. A plot of Bayesian prior and posterior distributions.

3. A display illustrating confidence intervals.

**Required technical elements -presentation − 6 points**

1. A .csv file with the dataset, uploaded to the course website.

2. A long, well-commented script that loads the dataset, explores it, and does all the analysis.

3. A short script that presents interesting highlights in ten minutes.

4. A one-page handout (bring 19 copies) that explains the dataset and summarizes the analysis.

5. A one-paragraph abstract.

6. Meeting deadlines for posting abstract (Friday evening before), dataset and scripts(9 AM on day of presentation)

**Bonus points for creativity or complexity – maximum of 9**

1. Use of regression, Student t, or Bayesian methods (not counted above) (2 points)

2. Comparison of analysis by classical methods and simulation methods.

3. Comparision of analysis by Bayesian and frequentist approaches.

4. Calculation and display of a logistic regression curve

5. A data set with lots of columns, allowing comparison of many different variables.

6. A graphical display that is different from those in the textbook or in the class scripts.

7. Appropriate use of R functions for a distribution and its conjugate prior distribution.

8. Appropriate use of bootstrap techniques (2 points)

9. A convincing demonstration of a relationship that might not have been statistically significant but that turns out to be so.

10. A convincing demonstration of a relationship that might have been statistically signifcant but that turns out not to be so.

11. Professional-looking software engineering (e.g defining and using your own functions).

12. Nicely labeled graphics, with good use of color, line styles, etc., that tells a convincing story.

13. Appropriate use of novel statistics (e.g. trimmed mean, maximum or minimum, skewness, ratios, correlations, R squared).

14. Use of theoretical knowledge of chi-square, gamma, or beta distributions.

15. Maximum-likelihood estimation of parameters. (2 points)

16. Appropriate use of covariance or correlation.

**Subjective impression – if these folks were applying for a job that requires computerized statistical analysis, I would**

1. Immediatey disband the search committee and hire them. (3 points)

2. Add them to a short list of leading candidates. (2 points)

3. View them as acceptable if no one better turns up. (1 point)