

Final Data Analysis Project

Title and abstract Due: First Tuesday of November, in class.

Final Project Due: Tuesday of Finals Week, 3 pm to my mailbox.
In my mailbox in 51-254 CHS.

ABSTRACT: Should be typed 1 paragraph max. Describe the data, your tentative model(s), source of prior information and purpose of your analysis.

Please supply the following information about your project:

1. Project name. (< 1 sentence)
2. Response variable name.
3. Predictor variable names.
4. Sampling density of response.
5. Parameter vector names and descriptions. (not θ , but β - a vector of regression coefficients ... is ok).
6. Prior information source. (previous data analysis, principal investigator, etc.)
7. Prior density. (product of independent normals for the regression coefficients and inverse gamma for σ^2 , etc.)

Please type. An informative sentence fragment in response to each item is eminently satisfactory. Be very brief!

The purpose of this update is to ensure that you are on track with your project.

You will need to find a data set to analyze. Identify an appropriate data set and purpose for analyzing it early in the quarter. Places to search for data sets are a) assembling or collecting your own data; b) your current or previous work; c) occasionally your advisor may have a suggestion; d) Doctoral thesis or master's paper data set; e) consulting client. Getting a 'real' data set as opposed to a book/web data set is strongly encouraged. If you get a set from the web or from a book you will need to do additional research to create a purpose for the analysis and to find a source of prior information. You may not take a data set from the notes, from a previous statistics course like 200A or from Hoff's, Gelman's or Congdon's books.

Your report must

- be at most 5 typed double spaced pages,
- be appropriately supplemented with graphs and tables as needed in the appendix. The appendix is not part of the 5 page limit.
- include a sensitivity analysis of the main results to (some) key assumptions.
- include proper priors for all parameters. The justification for your prior parameters will form the basis for part of the grade.
- report a plot or plots of the posteriors of interest.
- include your WinBUGS code.
- *interpret the numerical results in terms of the underlying problem.* This is vital for any data analysis.
- label appropriately all tables and graphs and must refer to all tables and graphs in the main text.

The outline of your paper might include many of the following sections:

1. Problem motivation & goal(s) of the analysis.
2. Description of the data set.
3. Description of prior information. What it is and how you translate that to a prior density.
4. Choice of model(s) and prior(s).
5. Mathematical specification of the model, that is, using “twiddle notation”, for example $y_i \sim N(\mu_i, \sigma_i^2)$.
6. Numerical and graphical results.
7. Convergence issues.
8. Problems encountered.
9. Sensitivity Analysis.
10. Conclusion/discussion.
11. References.
12. Appendix: WinBugs Code, Necessary Figures and Tables.

A regression-type model or hierarchical model is recommended.

The project will be graded on the writing, including English, organization, neatness and flow; and content, including justification and sensibility of the analysis. Allowance will be made for the complexity of the problem. Simple data sets will need to have substantially more thorough sensitivity analyses as compared with complicated data sets and models where getting a single posterior requires a substantial amount of work.

Please see me if you aren't sure whether a particular analysis/data set/problem is appropriate or doable.

Biostat grad students' projects will be returned to your Biostat mailbox. Other folks are encouraged to stop by early next quarter to pick up their graded data analysis project.

Each person must analyze a different data set. Joint custody of data is not allowed.