STAT 251 Project Specifications

Hard Copy Due Monday, April 11th, 5 PM

For this project, you will be allowed to either work alone or with one other class member. However, besides your team member, the only other mortals you may discuss your project with are the course TAs (Tom, Matt, and Josh) and me.

You will need to identify your own data for this project. There are several important requirements for the data, and they must all be met!

1. The data must be publicly available, and they must be already available (note this precludes you from soliciting responses).
2. You must include the source of the data in your project writeup.
3. You must be able to explain why you chose the research question and data that you did.
4. The analysis that you do MUST BE ORIGINAL, and therefore you should choose data that have not already been analyzed in a similar fashion by anyone, anywhere. If your analysis resembles too closely an existing analysis, that is grounds for an automatic E in the course and being reported to the university for a violation of academic integrity. This is a serious matter!
5. Although not required, I highly encourage you to send me an email ([page@stat.byu.edu](mailto:page@stat.byu.edu)) as soon as possible containing the following information: the proposed research question, the proposed data source and a short description of its contents (including a URL if possible), and either the name of your team partner or a statement that you will be working alone. I will then be able to give you feedback and, if necessary, redirection.

For this project, you should have one principal research question in mind, and this research question must be such that it requires a direct comparison of **two** populations. You will determine and justify your prior beliefs, report data summaries, conduct posterior inference on each population separately, and answer your primary research question using the most appropriate methodology.

Below is an outline of what your project writeup must entail:

* Introduction: Explain the question you decided to research as well as the value of answering this question.
* Literature Review: This part may not be applicable, but if possible briefly cite one or more sources that have looked at relevant (but not identical) questions and what their conclusions have been.
* Methods: Explain how the data were obtained, as well as providing a citation of their source. Justify your choice of a likelihood (distribution) for the data. Explain what the relevant parameters are for the likelihood. Choose and justify your prior distributions. Plot the prior distributions. Explain how you will use the posterior distribution(s) to answer your research question.
* Results: Report summary statistics for the data. Perform posterior inference on the parameters for each population and include two sets of properly formatted graphics: one that compares the parameters from the two populations to each other; another that compares the priors to the posteriors. Include and interpret credible intervals. Then perform the posterior inference on the function of the parameters that directly answers your primary research question (for example, the difference in means, or the ratio of variances). Report and properly interpret an associated credible interval for this function of parameters. Include properly formatted and informative graphic(s) for this function of parameters.
* Discussion: Summarize your results, identify shortcomings of the work, and explain what could be done in the future to either better answer your question or to address a natural follow-up question.
* Appendix: include all R code necessary to conduct your analysis.