

Independent Research Ideas for Stat 4893W

Using the Island:

1. Many of the standard deviation estimates calculated from the pilot data (blood pressure, total cholesterol, etc.) were too low resulting in sample sizes that were less than 10. Design a study on the Island to obtain a better estimate of the standard deviation and compare estimates to the literature. Based on your findings, what would you recommend to future classes on how to collect the pilot data to obtain a reasonable estimate of variability?
2. If you have not taken Stat 5421: Research how to analyze the data if the primary response was binary instead of continuous. Create a dataset from the Island in which the response is binary, add in some demographics as well as any predictors of interest, and build a multiple logistic regression model.
3. If you have not taken Stat 5421: Research how to analyze the data if the primary response was ordinal like mood. Collect ordinal data to analyze, research categorical literature and apply what you researched. The Island has a lot of variables that are ordinal.
4. Most of the groups are comparing treatments with respect to differences *before – after*. How does this relate to a repeated measures analysis? Rework the problem as a repeated measures with two timepoints, compare and contrast results from the two different approaches.
5. How would you analyze the data if there were three timepoints per Islander? Create a dataset from the Island with three (or more) timepoints per Islander. Rerun the experiment collecting three timepoints per Islander and compare treatments.
6. How would you analyze the data if the normality assumptions were violated using nonparametric techniques? Research newer nonparametric techniques and compare to both parametric and older nonparametric techniques.
7. How would you proceed if the investigators had several multiple comparisons and you were concerned about inflation of Type I error? Research recent developments (within the last 10 – 15 years) in multiple comparison procedures that are less conservative methods than Bonferroni's. Create a dataset from the Island and use your research to guide you in the analysis.
8. How would you deal with the analysis if the some of the data were missing? Examine techniques for handling the three cases of missing data: missing at random, missing completely at random, and missing not at random. Use the existing dataset from your consulting or create a new dataset from the Island and use your research to guide you in the analysis.
9. How would you redesign this study as a cross-over design? Collect data, rerun the experiment and analyze it as a cross-over. What would be the benefit of this approach? What would be the limitations?

10. If you have not already taken Stat 5401: What if the investigators had two primary responses, how would you analyze the data to test them? Collect data, rerun the experiment and analyze the responses jointly using multivariate methods.

Other ideas:

1. You learned briefly about a statistical concept in a class, and would like to study it more in depth and run an example of the method in R. Some successful past papers: Overdispersion, Markov Chain Monte Carlo, Autoregressive models (if not taking Time Series).
2. You have a data set you created for work (or from sports) that you would like to analyze. Can be the same data used for an honor's project, but you need to take a different focus. Successful past papers: analysis of survey data for an advertising firm, U of MN Quidditch team data analysis, looking at model building for an Econ honors paper (the Econ honors paper focused on the meaning of the models.), ranking sports teams and estimating home court advantage using Bradley-Terry Model.
3. You have a question you would like to analyze in a way that has not been done before, and you found a database with the data you need. Some successful past papers include: testing to see if the UN has met its Millenium Development goals, determining what factors lead to a woman not seeking health care, trying to understand the relationship between greek presence and academic success at universities.
4. You would like to learn a new statistical language or database method. You would need to compare your results to R. Some successful past papers include: integrating Hadoop into R, re running the analysis in SAS and comparing it to R. SAS has a free University Edition plus free tutorials on SAS are available.
5. Time to event data has many applications in medical research and reliability analysis. An example of this type of data in the medical field would be the time to recurrence of a particular type of cancer or medical condition after a treatment was administered. In the engineering field, it might be measuring the length of an electrical component before it fails. The interesting feature of this type of data is that a time is recorded if the event occurs and if the event does not occur during the course the study the data are censored. Find a dataset that reports this data, research the survival analysis literature and analyze the dataset.

Additional Reminders:

1. Once you have a topic for the independent research portion of your capstone project, you must discuss and get approval to proceed.
2. If you choose to use someone else's data, remember that you must cite your source or you are conveying that you collected the data. Not citing your source is a form of plagiarism.
3. The risk of using someone's data is that you must perform statistical research *beyond* what is currently reported in the literature. For example, if you performed a logistic regression using a data set that has already been used to illustrate logistic regression in a textbook, journal article or course handout, you must go beyond what is reported in that source. It is not adequate to repeat someone else's work and call it your own. This is a form of plagiarism.

4. For all topics, I do expect you to work alone and do original work. Your project needs to represent your work and not that of someone else. I do have access to paper-writing services on the web and will file an academic misconduct report for student who uses one of these services.