PROJECT 2:
SAMPLING EXPERIMENTAL DESIGN

Statistics 401: Fall 2016

Due Tuesday, September 20

This project is worth 30 points. Your grade will be determined by how well you answer the questions and by the professionalism and clarity of your write-up.

1. Read the article *Sharp Racial Divisions in Reactions to Brown, Garner Decisions* reported by the LA Times, the Washington Post, and CNN in November 2014 (available through the STAT401 web site).

   (a) Does the article report on an observational study or an experiment? Explain why your answer is correct. If you answer "observational study", explain why it is prospective or retrospective.

   (b) Identify the population of interest.

   (c) Identify two potential types of sampling bias. Explain in terms of the problem.

   (d) Identify three explanatory variables and two response variables. Give the categories of the categorical variables, or the possible numerical values of the response variables.

   (e) The last section of the article, states that the article's results are based on "telephone interviews ... of 1,507 adults ... living in all 50 U.S. states and the District of Columbia." What sampling method was likely used in order to be assured that adults from all 50 states and from DC were included? Give details about how this sampling plan may likely have been carried out.

   (f) Use R to generate a pie chart exactly like one of the ones in the article. Use proper labeling of the slices.

   (g) For this same pie chart, use R to display the same results as a bar chart. Use proper labeling.

   (h) Use R to generate a segmented bar chart like one of the ones in the article that compares opinions about the Brown case to opinions about the Garner case.

   (i) An investigator wants to determine how Montanans feel about the grand jury decisions in the Brown and Garner cases and suggests taking a simple random sample (SRS) from Montanans via the "landline and cell phone random digit dial" technique that the Pew study used. Due to the "wide age divisions" in opinions about these cases (that is, the article reports on opinions of 18-29 year olds, 30-49 year olds, 50-64 year olds, and 65+ year olds), give a detailed explanation describing how the sampling design can be modified to ensure that the sample from the Montana population is representative of all age groups.

2. Read the article *In Praise of Gratitude* available at the STAT401 web site.

   (a) This article describes the results of an experiment conducted by Emmons and McCullough. How do we know that an experiment was conducted?

   (b) The article makes broad claims about all humans. Although the article does not specify the sampling plan, identify what the likely sampling method was, and identify at least two potential types of bias in terms of the problem.

   (c) Identify the treatment and the treatment groups in the experiment conducted by Emmons and McCullough.

   (d) In the last question, identify which group is the control group. Was a placebo and/or blinding used? Explain.

   (e) In the context of this problem, explain why it is important to include a control group.

   (f) Comment on the articles statement "Of course, studies such as this one cannot prove cause and effect." Do you agree? Explain.

3. An investigator wants to perform an experiment to evaluate gratitude on the population of young people in Bozeman's public schools, K-12.

   (a) Identify how to sample from Bozeman public schools using strata AND clusters.

   (b) The article identifies an extraneous factor that is a potential confounding variable. How should blocking be used in this experiment? Identify the blocks, and be explicit how treatments will be applied.

   (c) Identify at least other two extraneous factors and indicate how to directly control them in the experiment.

   (d) What role does replication play in the design?

   (e) It is not possible to blind the students. Why not? Is it possible or even desirable to blind teachers?

4. If one is interested in making an inference from a sample to a population, then which of the following is most important? (a) random sampling (b) random assignment

5. If one wishes to make a causal inference (i.e., the explanatory variable causes changes in the response variable), then which of the following is most important? (a) random sampling (b) random assignment

6. Describe the conditions under which a cluster random sample is more efficient than a simple random sample.

7. Describe the conditions under which a systematic random sample is more efficient than a simple random sample.