Stat 401 Homework 1: Due Thursday, Sep. 7 by 1:40pm

General Instructions: Most homeworks will consist of a set of problems from the textbook, plus additional problems or data analyses. Turn in a physical copy of your homework at the beginning of class (or earlier) on the due date. For problems that require the use of R, include relevant and well-labeled R code and output with the problem. Additional R code and output may be included in an appendix.

Open Intro Exercises

- 1.4. Additionally, identify (d) which variables are explanatory variables and which are response variables, (e) the type of study (observational or experiment), (f) the type of sampling method used to select the sample, and (g) if selection, response, and/or non-response bias may be present and why.

- 1.11
- 1.14
- 1.20
- 1.34. Use blocking in your design.

Additional Problems

1. In each situation, identify the explanatory variable and the response variable. Indicate whether you think an observational study or a randomized experiment would be used. Explain why in each case.

(a) A teacher wants to compare the grade point averages of female students who are in sororities to grade point averages of female students who are not in sororities.

(b) A social psychologist wants to determine whether restaurant servers will get better tips if they introduce themselves by name to the people they serve.

2. A fast-food chain sells its burgers alone or as part of a “value meal” that includes fries and a drink. They know that some customers are health-conscious. They want to do an experiment to determine whether the proportion of customers choosing the meal would increase if they offered baby carrots with the meal as an alternative to fries. They will compare three treatments: (1) status quo, (2) offering carrots as an alternative but with no advertising, and (3) offering carrots as an alternative and advertising this option in the local area. They have restaurants in three types of areas: cities, suburban areas, and along major highways.

(a) Explain how they would conduct this experiment using a completely randomized design.
(b) Explain how they would conduct this experiment using a randomized block design.

(c) Explain why a matched-pair design could not be used for this experiment.

3. Students who had meningitis were matched with students without meningitis using sex, undergraduate (or graduate) status, and college. The students’ recent activities were examined to discover risk factors for meningitis. Explain whether each of the following terms applies to this study.

(a) Randomized experiment
(b) Prospective study
(c) Retrospective study
(d) Case-control study

4. Load the babies data set from Lab 1 into your R session and use R to answer the following questions about these data. Turn in relevant R code, output, and plots used to answer these questions.

(a) What proportion of mothers in the study smoke?

(b) Babies are considered preterm if they are born before 37 completed weeks of gestation. How many babies in the data set were preterm?

(c) Does smoking status seem to be associated with whether or not a baby is preterm in these data? Justify your answer using both summary statistics and a plot.