

STATISTICS 500 – Fall 2009 – Homework 1

DUE DATE:

on campus Friday, 4 Sept 2009, in lecture (11 am)

off campus Tuesday, 8 Sept 2009, by 4 pm to Nicole Rembert, email: rembeall@iastate.edu or
FAX: 515-294-4040 (please include cover page with Stat 500 / Nicole Rembert).

1. Design-based p-values and confidence intervals: I had the computer repeat the sampling exercises from lab a total of 500 times. The values are in lab1bigex1.txt and lab1bigex2.txt in the data files section of the class web site. The first file has results from the randomization (exercise 1) and the second has results from the bootstrap (exercise 2). The data values have been sorted for your convenience. Remember the “real” data had an observed difference of 4.49 and an observed ratio of 3.84.
 - (a) Using the appropriate sets of 500 computer-generated values, calculate the one-sided p-values for testing H_0 : difference = 0 and H_0 : ratio = 1. In both cases, “more extreme events” are larger values.
 - (b) The lab notes for week 1 describe how to calculate the two-sided p-value, but we did not do it during lab. Use the appropriate sets of 500 computer-generated values to compute the two-sided p-values for testing H_0 : difference = 0 and H_0 : ratio = 1.
Reminder: tradition is that two-sided “more extreme events” for differences are ≥ 4.49 or ≤ -4.49 . The ratio is less clear. Please use ≥ 3.84 or $\leq 1/3.84$.
 - (c) Using the appropriate sets of computer-generated values, estimate 95% confidence intervals for the difference and the ratio.
 - (d) Using the appropriate set of computer-generated values, estimate a 99% confidence interval for the difference and the ratio.
2. Experiment or observational study?

A study recorded children’s cholesterol level and the number of hours of tv watched per day. Children who watched more than two hours of television each day tend to have higher cholesterol levels than children who watched less than two hours of television daily.

 - (a) Is this an observational or experimental study?
 - (b) A newspaper report on this study has the headline: “Television watching bad for your health”. Is this headline an appropriate summary of the study? Explain why or why not.

The next six problems give you a short description of a study. For each of them:

- a) say whether this is an observational study, an experimental study with non-random assignment of treatments, or a randomized experiment.
 - b) For randomized experiments, identify the observational unit and the experimental unit. For non-randomized studies, identify what you believe are the most reasonable analogs of the e.u. and o.u., and explain (briefly) your choice.
3. An evaluation of anti-smoking treatment effectiveness. 52 smokers in Los Angeles County volunteered for the study. Each was randomly assigned to one of two treatments: the “Spiegel” method and a control. Each smoker was asked to record the number of cigarettes smoked per week. This information was recorded for the first week after initiation of treatments, the second week, and the third week.

4. An evaluation of the precision of methods to evaluate fat content of pieces of meat. A cut of meat is divided into 20 small pieces. 10 are randomly selected to be measured by method A; the other 10 are measured by method B. This is repeated for 6 pieces of meat.
5. An evaluation of the effect of paper colour (blue, green, orange) on response rates for questionnaires distributed on car windshields in parking lots. 30 cars were randomly chosen in a parking. From that list of 30, 10 cars were randomly chosen to get a blue paper on their windshield, 10 were randomly chosen to get a green paper. The last 10 got an orange paper. This was repeated in three additional parking lots (30 cars per lot, a total of 120 cars).
6. An evaluation of the effect of neighborhood poverty on reading proficiency. Census data are used to identify 10 neighborhoods with average income near the poverty line (poor) and 10 neighborhoods with average income more than twice the national average (rich). The school closest to the geographic center of each neighborhood is identified. The reading proficiency of each 8th grade students at that school is measured using a standardized test. The data are the reading scores for each student.
7. Exactly the same as the previous study, except you are not given data on individual students. You are only given the average reading score for all 8th grade students at the school.
8. A study of the longevity of three different compositions of concrete (A, B, C) in freeways. During reconstruction of a 15 mile stretch of freeway, concrete A was used in mile 1, concrete B was used in mile 2, C in mile 3, A in mile 4, B in mile 5, C in mile 6, repeatingly cyclically through mile 15. After 10 years of traffic, the strength of the concrete will be measured in the middle of each mile section.