Takehome final

STA 321, Basic Statistical Theory I
Spring Semester, 2014

Due: Tuesday May 8th at 3:00PM (bring it to 325D MDS)

More data analysis. This time, descriptive and inferential. Wherever possible, display the data using an appropriate graphical method. Carefully choose the correct inference method (and appropriate hypotheses), analyze the data set, and formulate an interpretation/conclusion. You may analyze "by hand" (using a calculator), using the online tools, or using statistical software such as R or SAS. Check the assumptions that you can check. Otherwise, assume that they are suffciently satisfied.

1 Consider the data on survival of passengers on the Titanic that we have used in the lab. If you know where to find good and reliable data, let us know. There are several sites with breakdown by sex, class, and age, one of them is `http://www.amstat.org/publications/jse/v3n3/datasets.dawson.html` If you are using R, you can download using Titanic datasets `https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/Titanic.html`

(a) Compare male to female passenger survival, using the following statistical tools: contingency table, row percentages, expected survival numbers if there was no association, risk difference, relative risk, odds ratio. Which of the statistical tools is most intuitive and useful in this situation?

(b) Compare first, second, and third class passenger survival using a contingency table and appropriate measures.

2 The following table refers to a survey of senior high school students in Dayton, Ohio.

		Cigarette Use	
		Yes	No
Alcohol Use	Yes	1449	500
	No	46	281

(a) Construct a row percentage table. That is, cigarette smoking is treated as the response variable, and alcohol is treated as the explanatory variable.

(b) Construct a column percentage table. The role of response and explanatory variable is now reversed.

(c) If there was no association, how many people would you expect in the four different groups?

(d) Do the two variables seem independent? Or associated (dependent)? Calculate appropriate measures of association.

3 Consider the study "Oral contraceptives and myocardial infarction: results of the MICA case-control study" from British Medical Journal, June 12, 1999 by Dunn et al. Parts of the abstract are copied below, the full article can be obtained from the BMJ webpage.

Objectives: To determine the association between myocardial infarction and use of different types of oral contraception in young women.

Design: Community based case-control study. Data from interviews and general practice records.

Setting: England, Scotland, and Wales.

Participants: Cases ($n = 448$) were recruited from women aged between 16 and 44 who had suffered an incident myocardial infarction between 1 October 1993 and 16 October 1995. Controls ($n = 1728$) were women without a diagnosis of myocardial infarction matched for age and general practice.

The study measures several variables in order to get a more precise picture. To simplify issues for this homework, let us focus on the fact that 62 of the cases and 261 of the controls used oral contraception.

Calculate the appropriate measures of association, and explain why other measures may not be appropriate.

4. Egyptian Skulls. Archeologists think that ancient Egyptians interbred with several different neighboring and immigrant populations for thousands of years. This claim would be supported by any indication of changes in body structure between 4000 B.C. and 200 B.C. [Thomson and Randall-Maciver (1905), Ancient Races of the Thebaid, Oxford University Press.] The following data contain the maximum skull breadth [mm] for 30 skulls of male Egyptians from each of these two time periods.

4000 B.C.: 131 131 125 135 131 132 119 139 136 132 138 126 139 135 125 134 131 128 134 130 129 138 134 128 126 127 132 131 141 124

200 B.C.: 141 131 141 129 135 136 133 131 131 139 140 144 139 141 140 130 138 133 132 138 134 131 135 136 133 132 136 135 134 141

5. Shape of Glass. Researchers randomly assigned participants either a tall and thin "highball" or a short and wide "tumbler" glass, each of which held 355ml. Subjects were asked to pour a shot (44.3ml) into the glass. Does the shape of the glass make a difference in how much alcohol they poured?

[Warsink and van Ittersum (2005), Shape of Glass and Amount of Alcohol Poured, British Medical Journal 331, 15121514.] The summary statistics from this experiment were as follows.

"Highball" glasses: sample size $n = 99$, sample mean $\bar{Y} = 42.2$ml, sample standard deviation $s = 16.2$ml

"Tumbler" glasses: $n = 99$, $\bar{Y} = 60.9$ml, $s = 17.9$ml

6. Vitality of Treetops. The effects of acid rain on treetops were investigated by comparing the vitality of trees under a "de-acidification roof" to those exposed to normal rain. The vitality of the tree is rated on a grading scale from 1 (vital) to 10 (dead).

[Bredemeier et al. (1998), The Solling Roof Project Site Characteristics, Experiments and Results, Forest Ecology and Management 101, 281293.]

De-acidification: 1 5 2 1 2 3 5 2 2 5 3 2 2 4 7 3 2 2 4 3 4 2 1 4 2 1 4

Normal: 2 1 2 3 4 2 4 4 3 2 3 4 2 2 3 1 5 1 3 2 4 1

7. Watching TV and Aggressive Behavior. Various studies have investigated whether there is a link between television violence and aggressive behavior by those who watch a lot of TV. Researchers have randomly sampled 707 families in northern New York State. They asked how much TV the sampled teenagers watched. Then, they made follow-up observations over a period of 17 years to see whether the same teenager conducted any aggressive act against another person, as reported by the teenager or by their mother.

[Johnson et al. (2002), Science 295, March 29, 2002.]

Out of 88 teenagers who watched less than one hour per day, 5 committed an aggressive act. Out of 619 teenagers who watched at least one hour per day, 154 conducted an aggressive act.

8. Cell Phones and Driving. Most scientists think that using cell phones while driving poses potential harm. An experiment investigated whether cell phone use impairs drivers' reaction times. 64 students from the University of Utah were randomly assigned to a cell phone group ($n_1 = 32$) or a control group ($n_2 = 32$). On a machine that simulated driving situations, at irregular periods a target ashed red or green. Participants were instructed to press a "brake button" as soon as possible when they noticed a red light. The control group listened to a radio broadcast or to a book on tape, while the cell phone group carried out a conversation about a political issue on the cell phone with someone in a separate room.

[Strayer and Johnston (2001), Psychological Science 21.]

The response times in the cell phone group had a mean of 585.2 milliseconds and a standard deviation of 89.6 milliseconds. In the control group, the mean was 533.7 milliseconds, and the standard deviation was 65.3 milliseconds.