

HOMEWORK 4
 STA 321, Basic Statistical Theory I
 Spring Semester, 2014

Due: Feb 20, 2011

1 RECALL: The following data set lists the number of goals in 2007 FIFA Womens World Cup games (starting with the opening match that had 11 goals, and ending with the final with 2 goals):

11, 4, 1, 0, 2, 7, 4, 2, 2, 2, 1, 3, 5, 3, 4, 2, 9, 4, 5, 5, 2, 4, 2, 1, 3, 1, 3, 5, 3, 4, 5, 2.

The number of goals in the 2006 FIFA Mens World Cup games were as follows:

6, 2, 1, 3, 3, 3, 1, 0, 2, 1, 4, 2, 3, 1, 6, 3, 0, 5, 4, 1, 0, 2, 3,
 2, 2, 3, 2, 2, 2, 3, 1, 4, 2, 0, 5, 4, 0, 3, 2, 2, 2, 4, 4, 4, 4,
 1, 1, 2, 3, 1, 1, 1, 0, 3, 4, 2, 3, 0, 1, 2, 1, 4, 2.

(Goals in extra time are counted, whereas goals in penalty shoot-outs are not counted)

- (a) Create side-by-side box plots for these two data sets.
- (b) Are there any outliers?
- (c) How does the empirical rule work for these two data sets?
- (d) How does Chebycheffs inequality work here?

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

		Cigarette	Use
		Yes	No
Alcohol	Yes	1449	500
Use	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.