

**STAT 201: Elementary Statistics Session 13 & 14, Spring 2015**

**Exam 1**

Instructor: Haigang Liu

1. (16 points) University of South Carolina wants to make an inference study about the percentage of students at the school who prefer having a several-day period between the end of classes and the start of final exams, to help reduce the level of stress as students prepare for exams. A survey is taken of 100 students from STAT 201 classes, and 34 of them believe that the period is beneficial while others do not. Based on a Z-score table, we conclude that about 29%-43% students agree with this policy across the university.

- Is the sample of 100 STAT 201 students a random one?
- Identify the population of this study.
- Which part of the example draws an inference about the population?
- The proportion of students who agree with this proposal is 34% from the sample. Is this quantity a statistic or a parameter? Why?

2. (16 points) Identify each of the following variables by labeling them as quantitative or categorical. If you identify one variable as quantitative, also write down whether it is continuous or discrete.

- Number of pets in your family
- Distance to commute when you go to work
- Choice of diet (vegetarian, non-vegetarian)
- Postal code in South Carolina (29201, 29202, . . . , 29210)

3. (14 points) Cities waxed and waned in the history. The largest 10 cities in United States in 1790, which is given by Table 1, are quite different from those in 2000. For instance, Southwark even disappeared and was merged into South Philadelphia. Read the table and answer the following questions.

New York City	33
Philadelphia	29
Boston	18
Charleston	16
Baltimore	13
Northern Liberties	10
Salem	8
Newport	7
Providence	6
Southwark	6

Table 1: The population of cities in 1790. Population in Thousand.

- Find the mean, median and mode for these data. *Hint*: The formula to find the mid-point is  $\frac{1}{2}(n + 1)$ , where  $n$  is the number of observations.
- Which is of the numerical summary is more sensitive to outlier? Mean, or median?

4. (20 points) The stock price, at January 28, 2015, of Amazon, Apple, Google, Microsoft, and Facebook are given by

307, 522, 117, 42, 77.

Note that  $\sum_{i=1}^5 x_i = 1065$  and  $\sum_{i=1}^5 (x_i - \bar{x})^2 = 161270$ , where  $\bar{x}$  is mean of  $x$ . You can use the formula as follows.

$$\text{s.e.}(X) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2},$$

where  $\text{s.e.}(X)$  is the standard deviation of  $X$ , and  $n$  is the number of observations.

- Find and *interpret* standard deviation of the five stock prices.
- According to empirical rule, find the price interval that contains approximately 68% of all stock.
- What assumption is required when applying the empirical rule?
- Consider Alibaba, another high-tech company, with price 100 USD on January 28, 2015. Would you consider it as an outlier based on Z-score? *Hint:* You can use the standard deviation from a) directly, without calculating it again by including Alibaba. The formula of Z-score is given by

$$Z = \frac{\text{An observation} - \text{Mean}}{\text{Standard deviation}}$$

5. (20 points) Researchers in a study gathered 1000 observations and plotted them in a histogram, which is displayed in Figure 1. Answer the following questions.

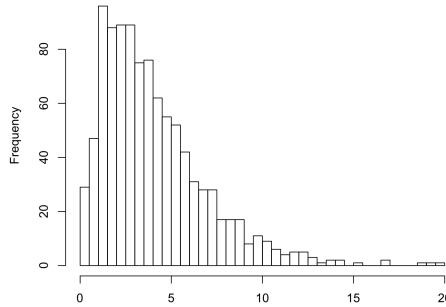


Figure 1: Histogram (Question 5)

- What is the difference between a histogram and a bar graph?
- Is the distribution in Figure 1 left skewed or right skewed?
- For a right-skewed distribution, which quantity is larger, mean or median? What about a left-skewed distribution?
- Can you come up with a real world example with a distribution skewed to right?

6. (14 points) 100 random numbers are generated based a Chi-squared distribution. The summary of data is given below, where 1st Qu. is the first quartile, and similarly, 3rd Qu. is the third quartile.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
34.36	298.60	420.40	495.90	633.20	1709.00

- Find the interquartile based on given numerical summary.
- Explain why the largest observation, 1709, is marked as outlier in the following box plot, based on interquartile.
- Based on Figure 2, comment on the shape of distribution of the 100 random numbers. Is it skewed? If so, left skewed or right skewed?

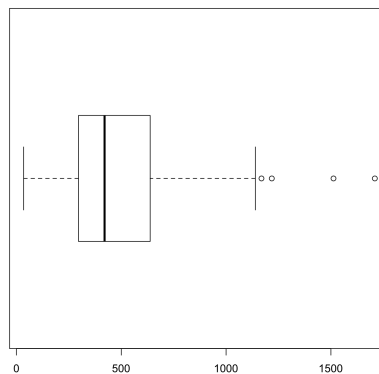


Figure 2: Box plot (Question 6)

# ANSWER SHEET PAGE ONE

First Name:

Last Name:

# ANSWER SHEET PAGE TWO

First Name:

Last Name:

# ANSWER SHEET PAGE THREE

First Name:

Last Name:

# ANSWER SHEET PAGE FOUR

First Name:

Last Name:

# ANSWER SHEET PAGE FIVE

First Name:

Last Name: