

Department of Computer Technology and Information Systems

CTIS 365: Applied Data Analysis

Semester: Fall 2019-2020

Lab Guide #4 – Week 7

OBJECTIVES: Central Tendency, Variability, Z-Score

Instructor: Dr. Seyid Amjad ALI

Assistant: Burcu ALPER

Q1. The below given vector contains data for the number of offensive drives by a certain football team.

```
offensiveDrives <- c(43, 42, 45, 44, 45, 48)
```

- Find the mean offensive drive for the team.
- Add 12 to all the data points in the offensiveDrives vector and see how the mean has changed.
- Subtract 9 to all the data points in the offensiveDrives vector and see how the mean has altered.
- Remove last 2 data points and see how the mean has changed.
- Concatenate to the truncated data points the vector `c(26, 23, 24)` and observe the change in mean.
- After finding the above mean, add this mean as a new data to the vector and find the new mean. Notice how the new mean has changed.

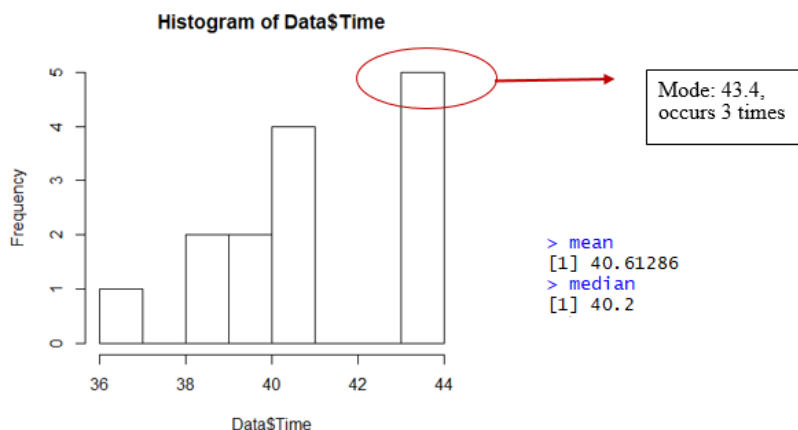
Q2. Below vector includes high temperature values for two weeks during the month of December in Chicago. Find the mean, median and mode. Decide which measurement for central tendency is more appropriate.

```
temperatures <- c(26, 23, 24, 29, 31, 28, 32, 21, 32, 29, 25, 33, 27, 34)
```

Example Run:

28.14286

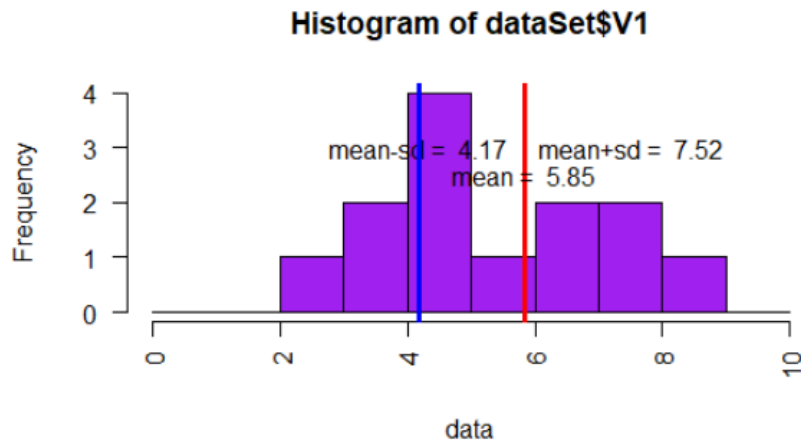
Q3. One of the events in the Winter Olympics is the Men's 500-meter Speed Skating. The best Speed Skating time by an athlete for this event are show to the right. Sketch a histogram like below and find the mean, median and mode.



Year	Time
1928	43.4
1932	43.4
1936	43.4
1948	43.1
1952	43.2
1956	40.2
1960	40.2
1964	40.1
1968	40.3
1972	39.44
1976	39.17
1980	38.03
1984	38.19
1988	36.45

Q4. A **population** dataset named "q4data.txt" is given to you

- Sketch a histogram for the given population data.
- Find the mean.
- Find variance and standard deviation by using 3 different approaches and verify that they all lead to the same result (definition formula, computation formula and R's built-in function).
- Mark the value of the sample mean and $\mu \pm \sigma$ (standard deviation) in your sketch.
- Add 2 points to each score and compute the new mean, variance and standard deviation to see the changes.



Q5. For a **population** with $\mu = 65.3$ and $\sigma = 15.4$,

- Find the z-score for each of the following X values (you can store these values in a vector and write a custom function to compute the z-score).

X = 75	X = 48	X = 84	X = 59
X = 54	X = 78	X = 51	
X = 45	X = 85	X = 74	

- Find the raw data (X value) that corresponds to each of the following z-scores.

z = 1.00	z = 0.25	z = 1.50
z = -0.50	z = -1.25	z = -2.50

- Repeat part a) by assuming that the above given data (X) **represents our sample**. Compute the sample mean and standard deviation and then convert the raw score to their equivalent z-scores.

Example Run a:

	x	freq	Z
1	45	1	-1.38
2	48	1	-1.18
3	51	1	-0.97
4	54	1	-0.77
5	59	1	-0.43
6	74	1	0.59
7	75	1	0.66
8	78	1	0.87
9	84	1	1.27
10	85	1	1.34

Example Run b:

	X	Z
1	26.8	1.00
2	46.0	0.25
3	57.6	1.50
4	69.1	-0.50
5	80.7	-1.25
6	88.4	-2.50

Example Run c:

	x	freq	Z
1	45	1	-1.31
2	48	1	-1.12
3	51	1	-0.92
4	54	1	-0.73
5	59	1	-0.41
6	74	1	0.56
7	75	1	0.63
8	78	1	0.82
9	84	1	1.21
10	85	1	1.27

- Q6.** Listed below are the first six rows of raw data from a trial IQ test. Calculate the z-score for each raw score and find the IQ score for every participant.

```
raw.test.data <- c(98,89,107.5,110,115,114,105,109,99,89,110.5,100.8)
```

Hint: To calculate the IQ score for one participant use the below formula

$$\text{IQ.score} <- 100 + (\text{zscore} * 15)$$

Example Run:

	test	Zscores	IQ
1	98.0	-0.70	89.50
2	89.0	-1.76	73.60
3	107.5	0.43	106.45
4	110.0	0.72	110.80
5	115.0	1.31	119.65
6	114.0	1.19	117.85

- Q7.**
- You are supplied with "**happiness.csv**" file that includes happiness Level by Provinces and Gender. Construct a bar graph showing the 3 big cities (Ankara, Istanbul, Izmir) happiness level according to the genders.
 - Find the mean, median and mode for females and males happiness Level.

