Data Analysis: HWA02 – Group Exercise Sheet: 2

Name: Sean Paolo (Group 8)

Wed 28 Feb 2024

Descriptive statistics – Introduction

General instructions:

Follow the **same instructions as described in HWA1 template and uploading process** whenever applicable, namely if you need to run an R script. Use this template Rmarkdown file in order to **insert the equations** which you may need/use for the analytical calculations. The purpose is to get some hands-on experience in typing simple math formulas with LaTeX. Providing the mathematical formula(s) is part of the evaluation and the learning goal. **This is an ".Rmd" file. Give you answers below using blue colored text, click "Knit" and finally submit/upload YOUR PDF file on TEAMS.**

Exercise 1: (1 Point)

Which of the following is an example of ratio scale data?

(a) Score on an accounting final

(b) temperature in Fahrenheit

(c) weight of cows

(d) occupation

(e) all of the above.

Exercise 2: (1 Point)

A researcher should not use the mean as a measure of central tendency unless the data is at least on a(n)_____ scale:

(a) nominal

(b) ordinal

(c) interval

(d) ranking

(e) ratio

Exercise 3: (1 Point)

Which of the following is a parameter?

(a) sample mean

(b) sample standard deviation

(c) population mean

(d) sample median

(e) sample mode

Exercise 4: (1 Point)

Which of the following is an example of continuous data?

(a) number of children

(b) amount of time it takes to assemble an IKEA bookcase

(c) total number of phone calls made in a week

(d) number of bathrooms in a house

(e) all of the above.

Exercise 5: (1 Point)

Which of the following is an example of discrete data?

(a) circumference of American women's wrists

(b) amount of time spent playing computer games

(c) total number of phone calls made in a week

(d) length of elephant tusks

(e) all of the above.

Exercise 6: (1 Point)

Which of the following is an example of interval scale data?

(a) religion of Americans

(b) ethnicity of Americans

(c) temperature in Centigrade (Celsius)

(d) height of Americans

(e) all of the above.

Exercise 7: (1 Point)

A manufacturer of supercomputers wants to sample 20 out of 500 that have been manufactured this year. An ID number is assigned to each of the 500 computers and then 20 random numbers are generated to see which computers to choose for the sample. This is an example of a:

(a) random data listing

(b) frequency distribution

(c) simple random sample

(d) census

(e) none of the above.

Exercise 8: (12 Points)

A firm wants to establish how long it is supposed to take to assemble a certain product. A random sample of 13 employees is selected and the company measures how long it takes each of them to assemble the product rounded to the nearest minute.

Data: 16; 8; 11; 9; 14; 22; 20; 16; 30; 25; 19; 16; 15 in order: 8; 9; 11; 14; 15; 16; 16; 16; 19; 20; 22; 25; 30

Solve the following by hand and type the formula you used in LaTeX.

1. The mean is: _____

$$\text{Mean} = \frac{8+9+11+14+15+16+16+16+19+20+22+25+30}{13} = \frac{221}{13} = 17$$

   (0.5 Point)

2. The median is: _____ Arrange the data in ascending order: $\{8, 9, 11, 14, 15, 16, 16, 16, 19, 20, 22, 25, 30\}$. The median is 16. (0.5 Point)

3. The mode is: _____16 (0.5 Point)

4. The first quartile is: _____First Quartile ($Q1$):

$$Q1 = \frac{11+14}{2} = 12.5$$

   (0.5 Point)

5. The third quartile is: _____ Third Quartile ($Q3$):

$$Q3 = \frac{20+22}{2} = 21$$

   (0.5 Point)

6. The range is: _____30-8=22 (0.5 Point)

7. The IQR is: _____IQR=Q3-Q1=21-12.5=8.5 (0.5 Point)

8. The variance is: _____ Variance:

$$\text{Variance} = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{468}{12} = 39$$

   (0.5 Point)

9. The standard deviation is: _____Standard deviation $= \sqrt{\text{Variance}} \approx 6.245$ (0.5 Point)

10. The coefficient of variation is: _____ Coefficient of Variation $= \frac{\text{Standard Deviation}}{\text{Mean}} \approx 0.367$ (0.5 Point)

11. Solve again the above using R. (7 Points) (**submit R code and result**)

Exercise 9: (1 Point)

Which of the following is not a measure of location?

(a) mean

(b) median

(c) variance

(d) percentile

(e) quartile

Exercise 10: (1 Point)

Which of the following statements is **False**?

(a) the median is a measure of central tendency

(b) quartiles do not measures of central tendency

(c) the 50th percentile is the median Q2

(d) an extreme value is likely to have a greater effect on the median than the mean false

(e) the 25th percentile is the first quartile (Q1)