Homework: Boxplots

MATH 150

Due: Feb 7, 2024

Sepehr Akbari

Problem 3

Refer to the iris data set, which is built-in in R.

(a) Compute the five-number summary and interquartile range for the variable Sepal.Width using one command each (no arithmetic or sorting needed).

(b) Should any of these observations be considered outliers? Apply the standard from class.

(c) Sketch a boxplot for this data.

Answer

(a) Using R:

```
View(iris)

fivenum(iris$Sepal.width) # 2.0 2.8 3.0 3.3 4.4
IQR(iris$Sepal.width) # 0.5
```

Therefore $IQR = 3.3 - 2.8 = 0.5$

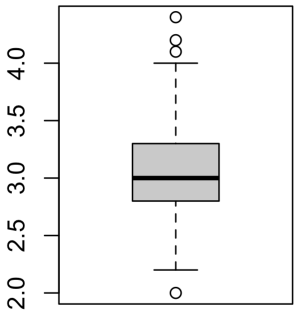
(b) Using R:

```
View(iris)

fivenum(iris$Sepal.Width)[2] - IQR(iris$Sepal.Width) * 1.5 # Lower bound = 2.05
fivenum(iris$Sepal.Width)[4] + IQR(iris$Sepal.Width) * 1.5 # Upper bound = 4.05
```

Therefore 2.0 and 4.4 are potential outliers.

(c) Using R:



Homework:
Percentiles & Quantiles

MATH 150

Due: Feb 7, 2024

Sepehr Akbari

Problem 1

The first two problems refer to the erykah data set, available on Moodle. Compute the five-number summary and inter-quartile range of the tempo variable. You should only need one R function for each.

Answer

Using R:

```
library(readxl)
erykah <- read_excel("Documents/LFC/MATH 150/DSs/erykah.xlsx")
View(erykah)

fivenum(erykah$tempo) # 0.000  82.485  92.109 129.868 215.079
IQR(erykah$tempo) # 47.383
```

Problem 2

What is the 40th percentile of the duration variable? How long is this in minutes?

Answer

Using R:

```
quantile(erykah$duration_ms, 0.4) # 40th quantile = 248309
quantile(erykah$duration_ms, 0.4) / 60000 # approx. 4.138483 minutes
```

Problem 3

The remaining problems refer to the following data, which represents the ages of 18 customers at a restaurant.

49 58 61 39 55 57 53 50 64
42 45 57 45 51 30 37 44 49

Which ages are below the 30th percentile?

Answer

Using R:

```
data <- c(49,58,61,39,55,57,53,50,64,42,45,57,45,51,30,37,44,49)

data[data < quantile(data, 0.3)] # 39 42 30 37 44
```

Problem 4

Which ages are above the 60th percentile?

Answer

Using R:

```
data[data > quantile(data, 0.6)] # 58 61 55 57 53 64 57
```

Homework:
Working with Z-scores

MATH 150

Due: Feb 7, 2024

Sepehr Akbari

Problem 1

Take a look at the faithful data set, which is pre-loaded in R. Find the mean and standard deviation of eruption lengths. What is the z-score of a five-minute eruption? Interpret your answer in ordinary human language.

Answer

Using R:

```
View(faithful)

x <- 5
mx <- mean(faithful$eruptions) # 3.487783
s <- sd(faithful$eruptions) # 1.141371

(x - mx) / s # 1.324912
```

–

The z-score of apprx. 1.3 suggests that a five-minute eruption is relatively, but not boldly unusual, longer than the average eruption length in the sample.

Problem 2

A certain kids' fun run has two age categories: 8-11 and 12-14. Finishing times in the younger group have mean 33 minutes and standard deviation 4 minutes, while finishing times in the older group have mean 29 minutes and standard deviation 5 minutes.

(a) Find and interpret the z-score of an 8-11 year old who finishes in 24 minutes.

(b) Find and interpret the z-score of a 12-14 year old who finishes in 24 minutes.

(c) Which is the more unusual of these two?

Answer

$$z = \frac{x - \overline{x}}{\sigma}$$

(a) Using R:

```
(24 - 33) / 4 # -2.25
```

–

The z-score of -2.25 suggests that the finishing time of this sample is 2.25 standard deviations below the mean. This suggests that the 8-11 year old finished the race much faster than the typical participant in their age category, making this sample a potentially unusual one.

(b) Using R:

```
(24 - 29) / 5 # -1
```

Problem 2 continued on next page...

–

The z-score of -1 suggests that the finishing time of this sample is 1 standard deviations below the mean. This suggests that the 12-14 year old finished the race faster than the typical participant in their age category. However, the difference is not as significant as the younger age-group.

(c) As the z-score of the younger-age sample has an absolute value higher than 2, and is larger than the second sample's z-score, it is more unusual.
It is also important to recognize that the z-score of the 12-14 year-olds, although signifying they are faster than the average in their age group, but cannot be considered unusual, as it is not high enough.