Equitable Equations: *Boxplots*

Problem 1

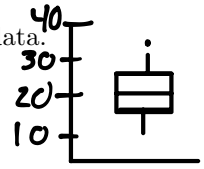
The following table shows 20 observations of gas mileages of 20 cars from model year 1974.

10.4	13.3	15.0	15.2	15.2	15.8	16.4	18.1	18.7	19.2
19.2	21.0	21.0	21.4	22.8	22.8	27.3	30.4	32.4	33.9

(a) Compute the five-number summary and IQR. Min: 10.4, Q1: 15.5, M: 19.2, Q3: 22.8 IQR: 7.3

(b) Should any of these observations be considered outliers? Apply the standard from class. 1.5 x IQR = 11.0
33.9 is an outlier

(c) Sketch a boxplot for this data.



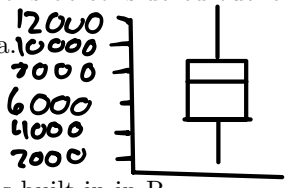
Problem 2

Refer to the `rock_sample` data set, available on Moodle.

(a) Compute the five-number summary and IQR for the `area` variable. The `sort` command may be helpful. Do NOT use more advanced tools (even the `median` function). Min: 1016 Q1: 5266.5, M: 7416, Q3: 8868 Max: 12212 IQR: 3,601.5

(b) Should any of these observations be considered outliers? Apply the standard from class. 1.5 x IQR = 5402
No outliers

(c) Sketch a boxplot for this data.



Problem 3

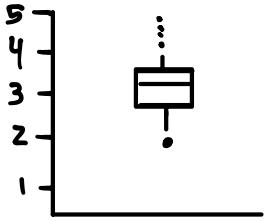
Refer to the `iris` data set, which is built-in in R.

(a) Compute the five-number summary and interquartile range for the variable `Sepal.Width` using one command each (no arithmetic or sorting needed). Min: 2.0, Q1: 2.8, M: 3.0, Q3: 3.3 Max: 4.4 IQR: 0.5

(b) Should any of these observations be considered outliers? Apply the standard from class. 1.5 x IQR = 0.75

(c) Sketch a boxplot for this data.

2.0 & 4.0 & above are outliers



Equitable Equations: *Percentiles and quantiles*

Instructions

Use R for all calculations. Include both answers and the code used to generate them.

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. quantile(erykah$tempo)

Min: 0 Q1: 82.485 M: 92.109 Q3: 129.868 Max: 215.079 IQR: 47.383

129.868 − 87.485

Problem 2

What is the 40^{th} percentile of the `duration` variable? How long is this in minutes?

quantile(erykah$duration, .4) = 248309 ms

Problem 3

quantile(erykah$duration, .4)/60,000 = 4.14 min

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

x ← c(values)

Which ages are below the 30^{th} percentile?

30-44 Below 45 - 30, 37, 39, 42, 44

Problem 4

Which ages are above the 60^{th} percentile?

51-64

above 51.4 - 53, 55, 57, 57, 58, 61, 64

Equitable Equations: *Working with z-scores*

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.

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?

1) Old Faithful erupts for an average of 3.49 minutes. This time generally varies by up to 1.14 minutes.

$$Z = \frac{5 - 3.49}{1.14} = 1.32$$

A 5 minute eruption varies is 1.32 times longer than the usual deviation from the average but it is not an unusual eruption length.

Problem 2)

a) $\frac{24-33}{4} = -2.25$

2.25 standard deviations faster than the average. An unusual time that probably placed them pretty high in the race.

b) $\frac{24-29}{5} = -1$

1 standard deviation faster than average. Not unusual.

c) The 8-11 year old is more unusual. The z-score was further from 0.