Equitable Equations: *The 5-number summary*

Problem 1

Use the following data for this problem.

$$3, 3, 4, 5, 7, 7, 7, 8, 9, 11, 12, 12, 14$$

(a) Compute the five-number summary and IQR.

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

Problem 2

Using a single R command, find the 5-number summary for the variable `Sepal.Width` in the built-in `iris` data set. What is the interquartile range for this variable? Identify any outliers using the 1.5× IQR test. The `sort` command may be helpful for this last part.

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.

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

(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).

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

(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).

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

(c) Sketch a boxplot for this data.

Equitable Equations: *Percentiles and quantiles*

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

Problem 2

The live version of the song, "Tyrone," has duration 221866 ms. What is the percentile of this observation in the `erykah` set. Round your answer to the nearest percentage.

Problem 3

The remaining problems refer to the following data, which represents the ages of 18 customers at a restaurant. You should code these values as a vector in R before proceeding.

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

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

Problem 4

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

Problem 5

Find the percentile that corresponds to an age of 43.