☺ 2.1 – Box Plots ☺

Objectives:

1. Create box plots from 5-number summaries; conversely, read 5-numbers summaries from box plots.
2. Describe the shape and spread of a data set from box plots.
3. Use box plots to compare centers, shapes and spreads of data sets.
4. Define and apply interquartile range.
5. Observe the effects of outliers on statistical summaries.

Five-Number Summary:

Example 1: Owen is a member of the student council and wants to present data about backpack safety to the school board. He collects these data on the weights of backpacks of 30 randomly chosen students. Owen wants to present a graph that shows the distribution and shape of the backpack data.

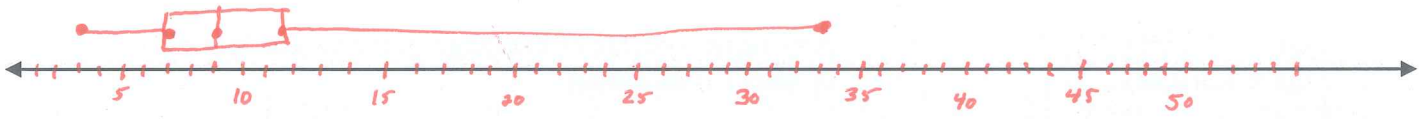
a. Below is the data Owen collected. Find the five-number summary:

3, 4, 4, 4, 6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 13, 15, 15, 16, 17, 20, 33

7 (Q1) 11.5 (Q3)

Min: 3 Q1: 7 Median: 9 Q3: 11.5 Max: 33

b. Create a box plot using the 5-number summary:



Biased: A statistical sample where Some members of the population are more likely to be included in the sample than other.

Simple Random Sample: A sample in which not only is each person/thing equally likely, but all groups of persons/things are equally likely.

Unbiased Estimates: Produced in a simple random sample because the data is collected from a sample where every member of the population is equally likely

Range: DIFFERENCE BETWEEN MAXIMUM AND MINIMUM

Interquartile Range (IQR): DIFFERENCE BETWEEN Q3 AND Q1

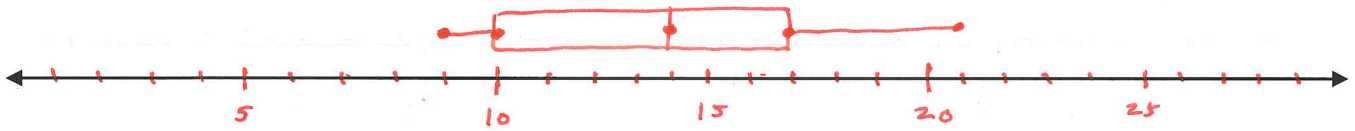
Example 2: In a classroom experiment, 11 bean plants were grown from seeds. After two weeks, the heights in centimeters of the plants were:

9, 10, 10, 13, 13, 14, 15, 16, 17, 19, 21
Q1 M Q3

a. Find the 5-number summary.

9, 10, 14, 17, 21

b. Create a box plot for this data.



Outlier: EXTREME data values.

How to determine if a value is an outlier?

Step 1:	Step 2:
Q1 – 1.5(IQR) Q3 + 1.5(IQR)	Any values outside of the range of the values found in Step 1 is an outlier.

Modified Box Plot: Box plot where OUTLIERS are graphed as separate points.

Example 3: Use the backpack data from Example 1 to answer the questions below:

a. Find the range and IQR.

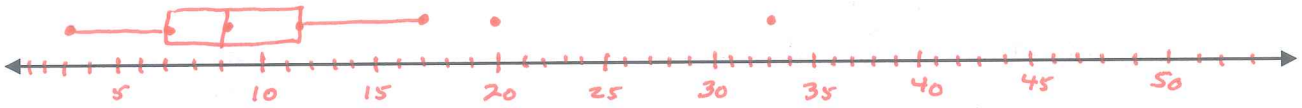
RANGE: 33 − 3 = 30 IQR: 11.5 − 7 = 4.5

b. Determine if there are any outliers.

7 − 1.5(4.5) = .25 NO

11.5 + 1.5(4.5) = 18.25 (20, 33)

c. Create a modified box plot showing the outliers.



Example 4: Determine if there are any outliers in the data below:

1, 3, 5, 9, 10, 14, 19, 27

4 9.5 16.5

IQR = 16.5 − 4 = 12.5

4 − 1.5(12.5) = 0 NO

16.5 + 1.5(12.5) = 35.25 NO

Shape: Describes how the data are distributed relative to the position of a measure of central tendency.

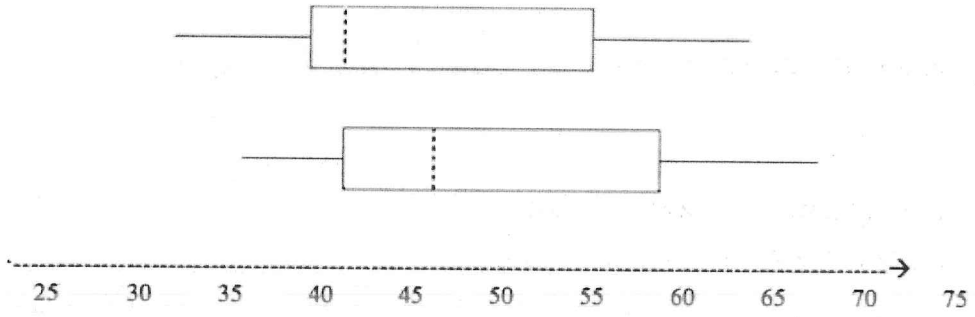
Symmetric: Data that is EVENLY DIVIDED, or nearly so, about the center.

Skewed: Data that are spread out MORE on one side of the center than the other.

Skewed Right

Skewed Left

Example 5: Use the box plot below to answer



a. What is the five-number summary for each box plot?

32-39-41-55-64

35-41-46-58-67

b. Make three comparisons about the box plots above.

The range of each is the same.

The second containg greater values than the first (all five points are higher)

Both are skewed right.

The median of the first is equal to Q1 of the second.

The IQR of the second is only slightly larger than the first.