Name KEY

11.3 Data Distributions – Guided Notes

Q. How can you represent the distribution of data in a given set of values?

Review: Method 1

Mean (average): measure of center. Add all values & divide by the number of values

Standard Deviation: measure of spread. how much each value deviates from the mean

Q. Do you need to arrange the data in numerical order when finding the mean and standard deviation?

A. No

Ex. 1a. Find the mean and standard deviation of the following data set:

4, 12, 15, 9, 14, 16, 13, 6, 7, 6, 25, 3, 13, 17, 22, 4

Mean: $\bar{x} = \frac{\text{sum of values}}{16} = 11.625$

~~*Standard Deviation:*~~

Method 2: The Five-Number Summary

We can look at the same data differently…

① *Minimum Value:* lowest value

④ *First Quartile* median of the first 1/2 of data

③ *Median:* value halfway btwn max & min

⑤ *Third Quartile:* median of the second 1/2 of data

② *Maximum Value:* highest value

***NOTE If there is an even number of data values, average the two middle values to get median

Range: Distance btwn max & min → Max - min

Interquartile Range: distance btwn first quartile & third quartile

3rd quartile - 1st quartile.

Q. Do you need to arrange the data in numerical order when finding the median and quartiles?

A. Yes

Ex1b. Use the same data set from Example 1a to find the five-number summary.

~~4, 12, 13, 8, 14, 16, 13, 6, 7, 6, 25, 6, 13, 17, 22, 4~~

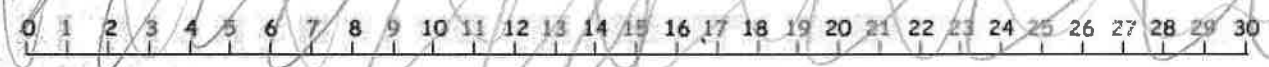
Step 1: Rearrange the data in ascending numerical order.

3, 4, 4, 6, 6, 7, 9, 12, 13, 13, 14, 15, 16, 17, 22, 25

Step 2: Calculate results.

Minimum	First Quartile	Median	Third Quartile	Maximum
3	6	12.5	15.5	25

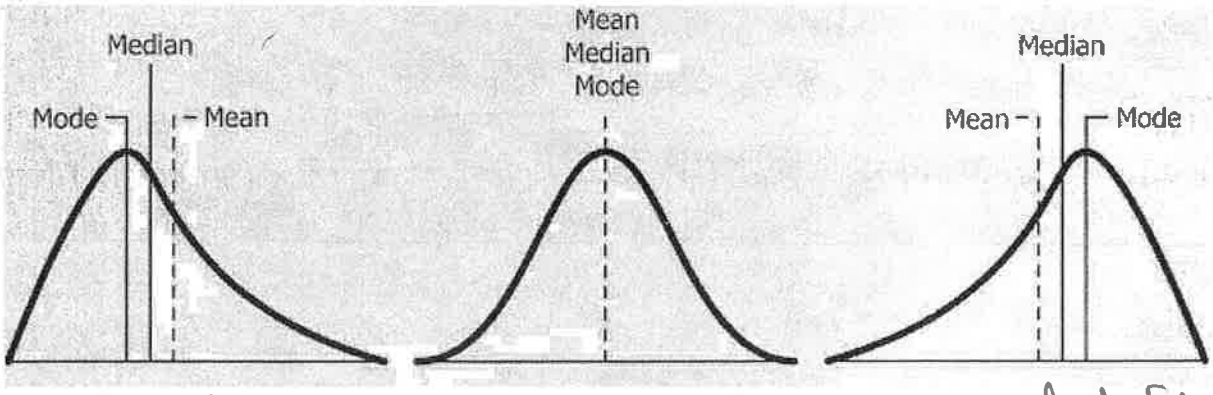
Step 3: Represent the data graphically with a box and whisker plot.



Q. Which method (mean/standard deviation vs median/quartiles) is a better representation of the data?

To decide look at shape of histogram

Skewed versus Symmetrical Distributions



Skewed right symmetric Skewed left

Measures of Center and Spread

Mean & Standard Deviation	Median & Interquartile Range
- Data symmetric - Good for representing normal distribution - mean, median, mode are about the same	- Not symmetric - Good when data is skewed to left or right - median & IQR are less affected by outliers

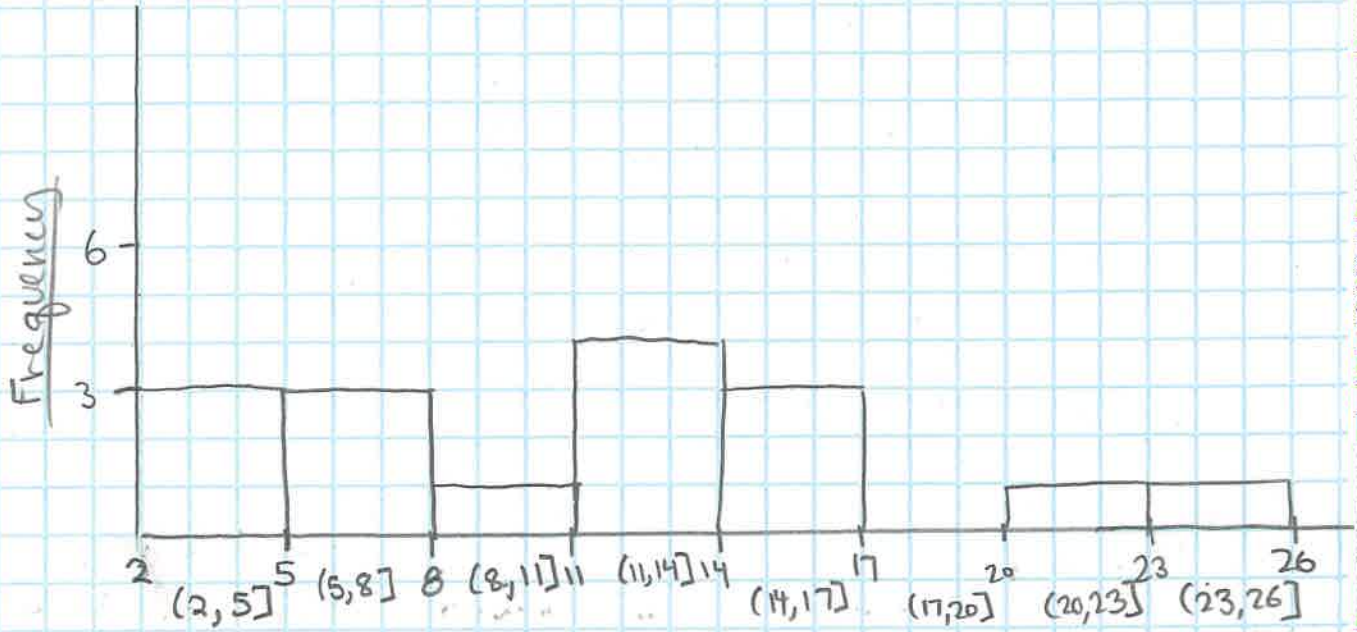
A. The data in **Example 1** can be described as a skewed right distribution, therefore the median & IQR are a better measure of center and spread.

Histogram

- Need 6 or more bins of data
- To find bin width, divide range by 6
- Include the right value in the bin

  Ex: 10-20 the interval (10, 20]
- Bins with whole numbers or categories, write # of category in the middle.

Ex 1



Bin width = $\frac{\text{Range}}{6} = \frac{22}{6} = 3\frac{2}{3} \rightarrow 4$

Skewed right

Best measure of center/spread: Median & IQR