Calculating Measures of Central Tendency & Spread

Measures of Central Tendency are used to generalize data sets and identify common values.

*Mean	**Definition:** Average of a numerical data set, denoted as $\bar{x}$ **Calculation:** Add up all the data values and divide by the number of data values **Useful When:** - Data values do not vary greatly - No outliers - Distribution is symmetric

Example: Find the mean of the following numbers. / Find average.
a. 76 77 79 80 82 88 90 92

$$\frac{\text{sum of all terms}}{\text{\# of terms}} = \frac{76+77+79+80+82+88+90+92}{8} = \boxed{83}$$

Median (Q2)	**Definition:** The middle number when the values are written in numerical order **Calculation:** Rewrite your data values in numerical order to find the middle number. o If your data set is **ODD**, then the median will be the number that falls directly in the middle. o If your data set is **EVEN**, then the median is the average of the two middle numbers. **Useful When:** - Distribution is skewed - Data values contain an outlier

Example: Find the median of the following numbers.
a. 76 77 79 80 82 88 90 92
rewrite in low to high order:

mean/average: $\frac{80+82}{2} = \boxed{81}$

Mode	**Definition:** Value that occurs most frequently. There can be no, one, or several modes **Calculation:** Find the numbers that are repeated o *NO MODE* (No numbers repeat) ▪ Say "no mode" o *ONE MODE* (One number repeats) ▪ State the number that repeats o *MORE THAN ONE MODE* (Several numbers repeat the same amount of times) ▪ State the numbers that repeat. **Useful When:** - Data set contains categorical data

Example: Find the mode of the following numbers.
a. 76 77 79 80 82 88 90 92

no #'s repeat → **No Mode**

Q1 & Q3	**Definition:** Quartiles are values that divide a list of numbers into quarters • **First (Q1) Quartile:** Median of the lower half of a data set ○ **Calculation:** Find the middle number of the values to the left of the median • **Third (Q3) Quartile:** Median of the upper half of a data set ○ **Calculation:** Find the middle number of the values to the right of the median

Example: Find the lower and upper quartiles of the following numbers.

a. 76 77 79 80 | 82 88 90 92

lower half upper half

median = middle

76 (77, 79) 80 82 (88, 90) 92

$\frac{77+79}{2} = 78 = Q1$ $\frac{88+90}{2} = 89 = Q3$

Outliers stands "OUT"	Data value that is much greater than or much less than the rest of the data in a data set If an outlier is present, you would use the median to describe the data, NOT the mean!

Example: Identify any outliers in the data set. Then determine if the median or mean best represents the data sets.

a. 15, 10, 12, 18, 10, 22

none - mean

b. 50, 15, 10, 12, 18, 10, 22

50 - median

Measures of Spread

Measures of Spread describe the "diversity" of the values in a data set. Measures of spread are used to help explain whether data values are very similar or very different.

Range	**Definition:** Difference between the greatest and least values in the set **Calculation:** Subtract the smallest data value from the biggest data value Range = Biggest # - Smallest #

Example: Find the range of the following numbers.

a. 76 77 79 80 82 88 90 92

smallest # biggest #

92 - 76 = 16

* on formula sheet

* **Interquartile Range (IQR)**	**Definition:** The difference between the third and first quartiles ($Q_3 - Q_1$). It finds the distance between two data values that represent the middle 50% of the data. **Calculation:** Subtract the first quartile value from the third quartile value ($Q_3 - Q_1$).

Example: Find the interquartile range of the following numbers.
a. 76 77 79 80 82 88 90 92

$Q_3 = 89$ $Q_1 = 78$

$Q_3 - Q_1 = 89 - 78 = \boxed{11}$

Mean Absolute Deviation (MAD)	**Definition:** Average absolute value of the difference between each data point and the mean. It essentially takes the average distance of the data points from the mean. A data set with a smaller mean absolute deviation has data values that are closer to the mean than a data set with a great mean absolute deviation. The greater the mean absolute deviation, the more the data is spread out. * (+) The formula for mean absolute deviation is: $\frac{\sum_{i=1}^{N} \lvert x_i - \bar{x} \rvert}{N}$ x_i = data value $\bar{x}$ = mean $\sum$ = sum N = number of data values ↓ positive answer/result **Calculation:** - Find the mean of the set of numbers - Subtract each number in the set by the mean and take the absolute value of each new number (new number will be positive) - Find the sum of the new numbers and divide by the number of data values

Example: Find the MAD of the following numbers.
a. 76 77 79 80 82 88 90 92

$N = 8$ $\bar{x} = \text{mean} = 83$

$$MAD = \frac{|76-83| + |77-83| + |79-83| + |80-83| + |82-83| + |88-83| + |90+83| + |92-83|}{8}$$

$$\frac{7+6+4+3+1+5+7+9}{8} = \frac{42}{8} = \boxed{5.25}$$