

Section 2: ANALYZING DATA

Measures of central tendency

Measures of central tendency: Different ways we've come up with to describe the 'middle', 'center', or most typical value of data

Mean, arithmetic mean: The balancing point of the data

$$\mu = \frac{\sum_{i=1}^n x_i}{n} = \frac{\text{the sum of all data points}}{\text{the number of data points}}$$

Median: The value at the middle of the data set when we line up all the data points in order from least to greatest

Mode: The value in the data set that occurs most often

Measures of Spread

Spread, dispersion, scatter: How, and by how much, our data set is spread out around its center

Range: The difference between the largest value and smallest value

Quartiles: The values that mark the 25th, 50th, 75th, and 100th percentiles of the data, which are Q₁, Q₂, Q₃, and Q₄, respectively

Interquartile range (IQR): The difference between the first and third quartiles, Q₃ - Q₁

Changing the data and outliers

Shifting: Adding or subtracting a value from every point in data set. Shifting changes the mean, median and mode, but not the range or IQR.

Scaling: Multiplying or dividing a value from every point in data set. Scaling changes the mean, median, mode, range and IQR.

Outlier: A number on the extreme upper or extreme lower end of a data set.

Box-and-whisker plots

Box-and-whisker plots, box plots: A useful plot for representing the median and spread of the data at the same time. The box plotted in center extends



from Q₁ and Q₃, and the whiskers extend beyond the box to the lower and upper ends of the range of the data.

Five-number summary, five-figure summary: A summary table that includes the minimum and maximum values, the median, and Q₁ and Q₃ for the data set.

Min	Q ₁	Median	Q ₃	Max
2	5	11	13	14