

# Distributions

## Frequency distribution

\* qualitative data

Summary of the number of occurrences of a **given value**

Bar chart

\* quantitative data

Summary of the number of occurrences in a **given bin**

Histogram

## Relative frequency

Ratio of frequency of a value  
to that of the total

Car
Tesla
Ford
Toyota
Toyota
Tesla
Toyota

Car	Frequency	Relative frequency
Tesla	2	$2/6 = 0.33$
Ford	1	$1/6 = 0.16$
Toyota	3	$3/6 = 0.5$



# Analyzing Distributions

## Percentile

**value of a variable** at which a specified [approximate] percentage of observations are **below that value**.

Location of the  $p$ th percentile, 
$$L_p = \frac{p}{100} (n + 1)$$

## Quartile

Data divided into 4 parts each containing **25% of observations**.

Box plot

**Q1** → [Min, 25% percentile] ; **Q2** → [25% percentile, 50% percentile]; **Q3** → [50% percentile, 75% percentile]  
**Q4** → [75% percentile, max]



# Methods of central tendency/location

Mean ( $\bar{x}$ )

Average value for a variable

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Median

Middle value of data when arranged in ascending order

Mode

Most frequently occurring value

Geometric Mean

$n$ th root of the product of  $n$  values

$$\sqrt[n]{(x_1)(x_2) \dots (x_n)}$$

Sales

\$20

\$10

\$10

\$10

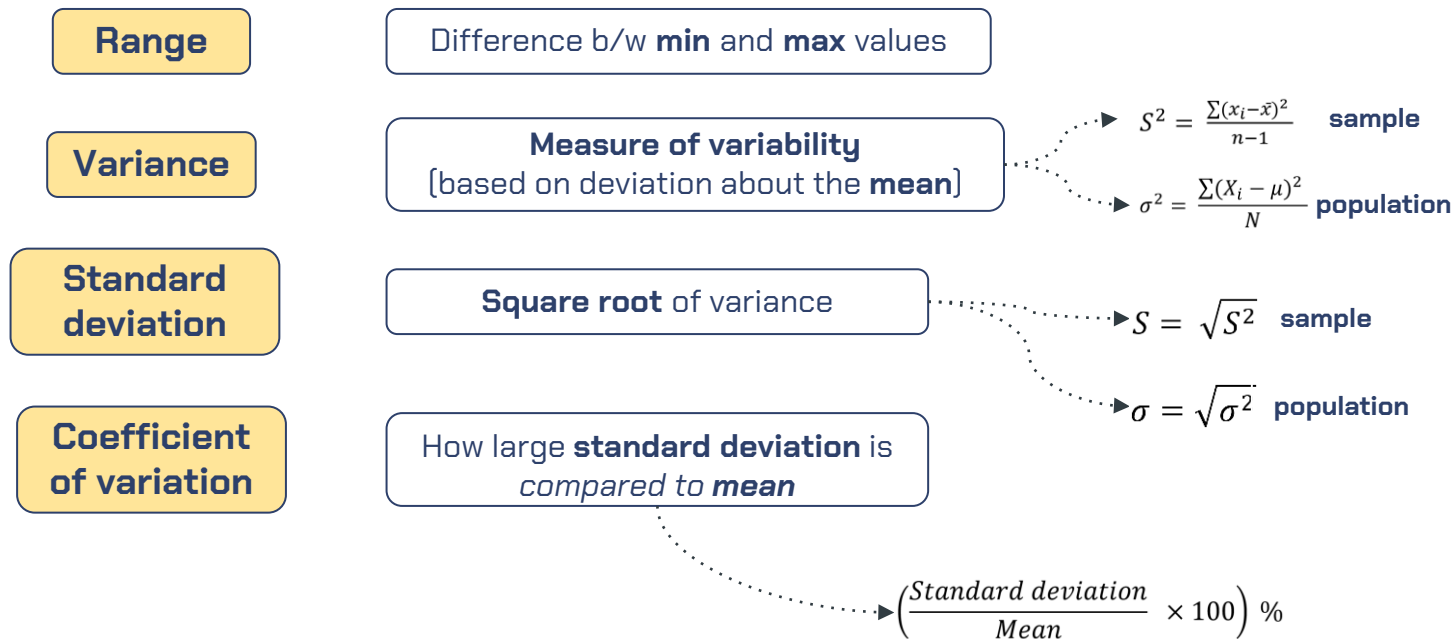
\$100

Mean = 30, median = 10, mode = 10, geometric mean = 18

Mean is susceptible to **outliers\***



# Methods of dispersion/varaiability



# Analyzing Distributions - Z-score

Z-score

Distance of a value away from mean in "standard deviation"

$$z_i = \frac{x_i - \bar{x}}{s}$$

