Bociat Daniel-Tiberiu

1. Define mean, median and range

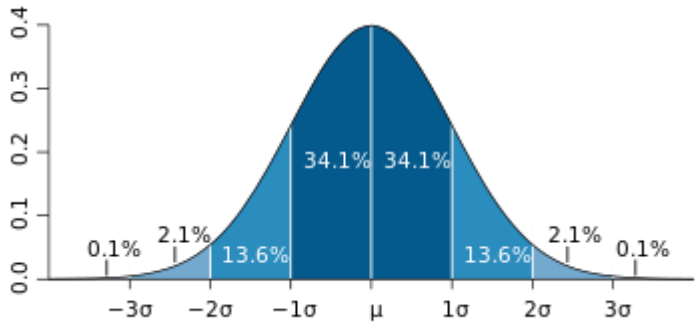
Mean = arithmetic mean of the data

Median = the **middle number** in a data set (position-wise not value-wise)

Range = the set of y-values that are output for the domain.

2. Define standard deviation and variance

Standard deviation is a measure of dispersement in statistics. “Dispersement” tells you how much your data is spread out. Specifically, it shows you how much your data is spread out around the mean or average.



$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

The variance is mathematically defined as **the average of the squared differences from the mean**.

$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1} \qquad s^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}$$

3. What is sample and what is population?

A sample is a select number of items taken from a population.
A population is a whole, it’s **every member of a group.** A population is the opposite to a sample, which is a fraction or percentage of a group

4. What is interquartile range?

The **interquartile range** is a measure of where the "middle fifty" is in a data set.

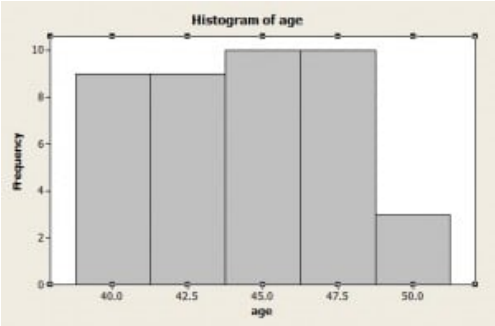
The interquartile range formula is the first quartile subtracted from the third quartile:

$$IQR = Q_3 - Q_1.$$

Think of Q1 as a median in the lower half of the data and think of Q3 as a median for the upper half of data.

5. What is histogram?

Histograms are similar to bar charts; they are a way to display counts of data. A bar graph charts actual counts against categories; The height of the bar indicates the number of items in that category. A histogram displays the same categorical variables in "bins".



6. What is bias in statistics?

Bias is the tendency of a statistic to **overestimate** or **underestimate** a parameter.

Bias can seep into your results for a slew of reasons including sampling or measurement errors, or unrepresentative samples.

7. What is a statistical distribution? What is the most common one?

The distribution of a variable is a description of the relative numbers of times each possible outcome will occur in a number of trials. The function describing the probability that a given value will occur is called the probability density function (abbreviated PDF), and the function describing the cumulative probability that a given value *or any value smaller than it* will occur is called the distribution function (or cumulative distribution function, abbreviated CDF).

The most common distribution is the normal distribution.

8. What is the difference between discrete and continous distribution?

A discrete distribution is one in which the data can only take on certain values, for example integers.

A continuous distribution is one in which data can take on any value within a specified range (which may be infinite), for example a point on a segment.