What is a variable?

A variable is any characteristic of the population being studied or observed.
For example, if a statistician surveys 953 Canadian males who play recreational sports, asking their age, what sport they play, and how often they play, the example includes 3 variables.

What is the population?

The population is the entire group you want to know about in a study.
For example, if a researcher wants to know the attitudes of North American women on violence against women, and surveys 450 college and university students, the population of the study is North American women.

Types of Sampling

Simple Random	Every selection is likely to be selected and every combination of selections is likely to be selected
Systematic Random	Has a starting point, and then every *n*th number is selected
Stratified Random	The data is divided into groups, and a sample from each group is selected
Cluster Random	Data is divided into groups, and a random sample of the groups is selected. All members of the selected groups are surveyed.
Multi-stage Random	Data is divided into groups, a random sample of groups is chosen, and a random sample of members from the chosen groups are selected
Convenience Random	Used simply because it is easily accessible
Voluntary Random	Obtained by making a general appeal for responses

Correlation

Correlation indicates the strength and direction of a linear relationship between two variables.
Correlation analysis involves generating a single number from the data, which is called the **correlation coefficient** and is represented by *r*.
The value of *r* is always between -1 and 1.
A coefficient of 0 indicates no linear correlation.
A coefficient between 0 and 1 indicates a positive correlation. A number closer to 1 indicates a stronger positive correlation, and a number closer to 0 indicates a weaker positive correlation.
A number between -1 and 0 indicates a negative correlation. A number closer to -1 indicates a stronger negative correlation, and a number closer to 0 indicates a weaker negative correlation.
For example, if r^2 = 0.9875878, a strong positive correlation is present between two variables.

Skew

If data is bunched towards the right, it is a positive skew.
If data is bunched towards the left, it is a negative skew.
If data is distributed evenly, there is no skew.

Calculating the Mean

To calculate the mean of a group of data, add up all the values, then divide by the number of values.

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com

Median

The median is the middle number when data is arranged in ascending order.
If there is an even number of data, there will be two middle numbers. If there are two middle numbers, the median is the mean of the two numbers.
If there are *n* data points and *n* is odd, the middle data point is *n* + 1 / 2
If *n* is even, the middle two numbers are the *n* / 2 and *n* / 2 + 1 numbers

Measure of Spread

The measure of spread is the degree to which data differ from, or are spread out from, the centre.

Range

Range is the difference between minimum and maximum values.
It shows the overall spread of data.

Interquartile Range

Quartiles are a measure of spread.
A quartile is any of the three values that divide the sorted data set into four equal parts, so that each part represents one-fourth of the data set.
To find quartiles, first use the median to split the data in half.
Q1 is found by finding the median of the half of data below the median of the entire set of data.
Q2 is the median of the entire set of data.
Q3 is found by finding the median of the half of data above the median of the entire data.

Percentiles

Percentile is the value of a variable below which a certain percentage of observations fall.
Percentile is used to rank an individual's position within a set of data.
For example, the 20th percentile is the value below which 20% of the observations are found

Variance & Standard Deviation

Variance and standard deviation are measures of spread.
Deviation is the distance between a particular value and the mean.
Variance is determined by calcul ating the average of the squares of all deviations for a particular set of data

Sponsored by **Readable.com**
Measure your website readability!
https://readable.com