

Tak3-notes

Data are **measurements or observations that are collected as a source of information**.

Quantitative variables are any variables where the data represent amounts (e.g. height, weight, or age).

Categorical variables are any variables where the data represent groups. This includes rankings (e.g. finishing places in a race), classifications (e.g. brands of cereal), and binary outcomes

Continuous and Discrete data

Continuous data refers to data that can be measured. This data has values that are not fixed and have an infinite number of possible values. These measurements can also be broken down into smaller individual parts.

Discrete data also referred to as discrete values, is data that only takes certain values. Commonly in the form of whole numbers or integers, this is data that can be counted and has a finite number of values. These values must be able to fall within certain classifications and are unable to be broken down into smaller parts.

Nominal and Ordinal scale

Nominal scale is a naming scale, where variables are simply “named” or labeled, with no specific order. Ordinal scale has all its variables in a specific order, beyond just naming them. Interval scale offers labels, order, as well as, a specific interval between each of its variable options.

Measures of central tendency

A measure of central tendency (also referred to as measures of centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.

There are three main measures of central tendency:

- mode
- median
- mean

Mode

The mode is the most commonly occurring value in a distribution.

Median

The median is the middle value in distribution when the values are arranged in ascending or descending order.

Mean

The mean is the sum of the value of each observation in a dataset divided by the number of observations. This is also known as the arithmetic average.

Notation

notation to simplify the description of how a statistic should be computed.

- Statistical formulas use algebraic notation, which rely on letters to designate a variable. By convention, if there is just one variable in a formula, the letter X is used to designate the variable. If there is a second variable in the formula, traditionally the letter Y is used to indicate the variable. If there is a third variable, the letter Z is traditionally used. After that, there are no universal traditions, but it is rare to have statistical formulas that involve more than three variables.
- The capital letter N traditionally refers to the total number of participants in a study.
- The single letter in statistical formulas refers to the variable. The individuals scores on that variable can be indicated by subscripts, which are numbers written below the letter to refer to a specific score. For example, X_i refers to the score for the first person on the X variable, and X_{27} refers to the score for the 27th person on the X variable. Y_{11} refers to the score on the Y variable for the 11th person.
- If there are several groups of participants, the number of participants in each group is indicated by a lower-case n with a subscript to indicate the group number. For example, n_i refers to the number of participants in the first group.
- Traditionally, the number of groups in a study are referred to by the lower-case letter k , although in complex designs, this tradition is modified. Therefore, n_k refers to the number of participants in the k^{th} group, which is the last group.
- This is a specified order in which functions are to be carried out. The order is:
 - The highest priority action should be to raise any variables to a power. For example, to compute $2X^2$, you would first square the value of X and then multiply by 2.
 - The next highest priority action is multiplication or division. For example, to compute $2X+1$, you would multiply the value of X by 2 and then add 1.
 - The lowest priority action is addition or subtraction.

Summation Notation

- Many statistical formulas involve adding a series of numbers. The notation for adding a series of numbers is the capital Greek letter sigma. The sigma stands for "add up everything that follows." Therefore, if the sigma is followed by the letter X , it means that you should add up all of the X scores.

$$\sum X$$