

Contents

1	Data	1
1.1	Data types	1
1.2	Measure quantitative data	1
1.3	Measure of center	1
2	Notation	2
2.1	Random variables	2
2.2	Aggregation	2

1 Data

Data is defined as distinct pieces of information. It can be text, video, image, spreadsheet, ...etc.

We can utilize data to make decisions and set goals.

1.1 Data types

- Quantitative data
 - Numerical values that allows us to perform mathematical operations
 - Continuous
 - * Any real numeric value
 - * Age is an example of such a data
 - Discrete
 - * Discontinuous values
- Categorical data
 - A group or a set of items
 - Ordinal
 - * Ranked
 - Nominal
 - * Not ranked

1.2 Measure quantitative data

When describing quantitative data we generally discuss 4 main aspects: 1. Center 2. Spread - Gives you an idea of how data is different. 3. Shape 4. Outlier

1.3 Measure of center

Gives you an idea of what average is.

1. Mean
 1. When we ask how many (something) do you expect (in some circumstances), the answer for that is the **mean**.
 2. $\bar{x} = \sum_{i=1}^n x_i / n$

2. Median
 1. Sort and take the middle
 2. Depends on the parity of count
 1. If it is odd, take the middle
 2. If it is even, we don't have middle, take the mean of two the values at the middle
 3. Mode
 1. Value that occurs most often
-

2 Notation

2.1 Random variables

Each column a the dataset is associated with a random variable. A random variable is a placeholder for the possible values of some process.

- Capital letters are used to represent a random variable, e.g. X .
- A specific value is denoted by a small letter with subscript number, e.g. x_1, x_2, \dots, x_n .

2.2 Aggregation

Is just a way to turn multiple numbers in a fewer number, usually one number. Common aggregations include the measure of center, mean, median, and the mode.

Examples: - \sum for sum - \prod for product

To calculate mean value of X :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$