



# Statistical Variables

Machine Learning

- Gender, Head Size are Features or Independent Variables
- Brain Weight is Target or Dependent Variable

	Gender	Head Size	Brain Weight
0	male	4112	1590
1	female	3690	1366
2	female	4557	1588
3	male	3986	1411
4	female	3299	1237
5	male	4723	1569
7	male	4200	aprox. 1599
8	female	4200	aprox. 1580

# Types of Statistical Variables

Statistical models contain variables that can be used to explain relationships between other variables.

Variable in statistics is an attribute that can be a thing, a number or emotion etc.

We have 2 types of Statistical variables

- Numerical Variable
- Categorical Variable

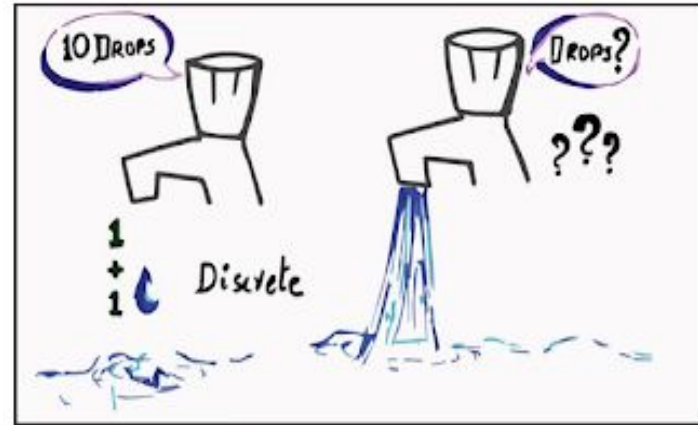


# Numerical Variable

- Numerical Variable is also known as Quantitative Variable.
- An numerical variable is a data variable that takes on any value within a finite or infinite interval.
- In numerical variable, the intervals between the values of the variable are equally spaced.
- Eg: distance, scores, age, etc..

Numerical variables again contains 2 types

1. Discrete Variable
2. Continuous Variable



# Numerical Variables Contin...

## Discrete variables:

- These Variables are countable in a finite amount of time (can classify).
- A discrete variable is a variable whose value is obtained by counting.
- And it is always numeric.

Eg:

- No.of Students
- number of heads when flipping three coins
- 0's or 1's
- etc..



# Numerical Variable Contin...



## Continuous Variables:

- A continuous variable is a variable whose value is obtained by measuring.
- It can take on an unlimited number of values between the lowest and highest points of measurement.
- These variables will take forever to count. In fact, you would get to “forever” and never finish counting them.
- Continuous variables include things as Time, speed and distance.

Eg: Age, Time, height of students, etc..

# Why 'Age' is Continuous Variable

Let's take a random person age, and it can be like  
25 years, 10 months, 2 days, 5 hours, 4 seconds, 4 milliseconds, 8 nanoseconds,  
99 picoseconds.....and so on.

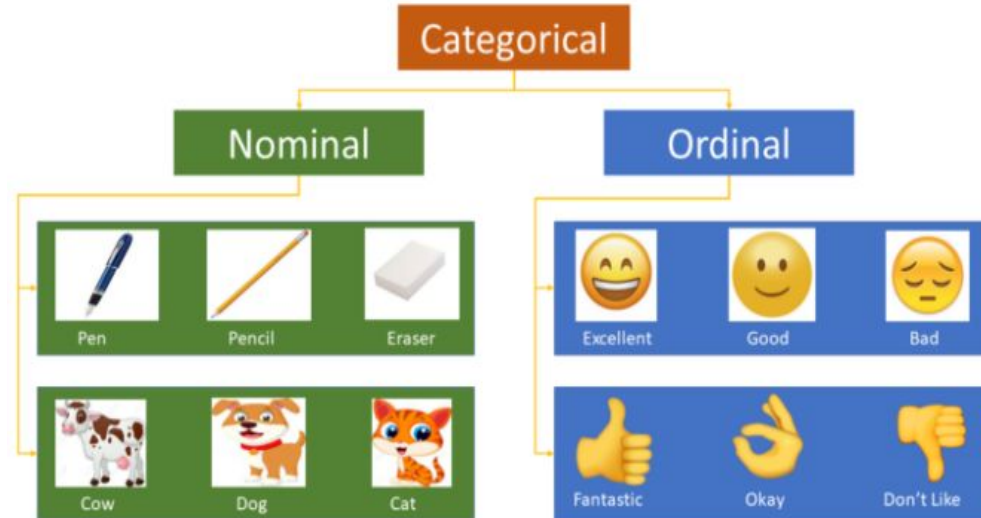
Note: Based on Dataset, Age will be either Continuous or Discrete

# Categorical Variables

- Categorical data is a collection of information that is divided into groups.
- These variables are similar to discrete variables(Classified Variables).
- Categorical Variable is also known as Qualitative Variable.

These variables can be further categorized as

- Nominal
- Ordinal





# Categorical Variable Contin..

## Nominal:

- Nominal scale is a naming scale, where variables are simply “named” or labeled, with no specific order.
- A Nominal variable is one that has two or more categories, but there is no intrinsic ordering to the categories.
- Eg: Red or Green, Male or Female, etc..

## Ordinal:

- A variable that there is a clear ordering of the categories.
- Ordinal scale has all its variables in a specific order, beyond just naming them.
- Eg:
  - Good or Bad
  - Agree or Disagree
  - Happy or Sad, etc..



# Why does it matter whether a variable is categorical or numerical?

Statistical analyses assume that the variables have a specific levels of measurement.

For example, it would not make sense to compute an average hair color. An average of a categorical variable does not make much sense because there is no intrinsic ordering of the levels of the categories.

If you tried to compute the average of hair color, you would also obtain a nonsensical result. Because the spacing between the types of educational hair colors are very uneven, the meaning of this average would be very questionable.

That is why it matter whether a variable is categorical or Numerical

# Binary Variable

A binary variable is a variable with only two values.that has the potential to take on one of two values. Some examples are:

- 1 / 0.
- Yes / No.
- Success / Failure.
- Male / Female.
- Black / White

Note: Binary Variable will be Numerical or Categorical, but have only 2 values.

