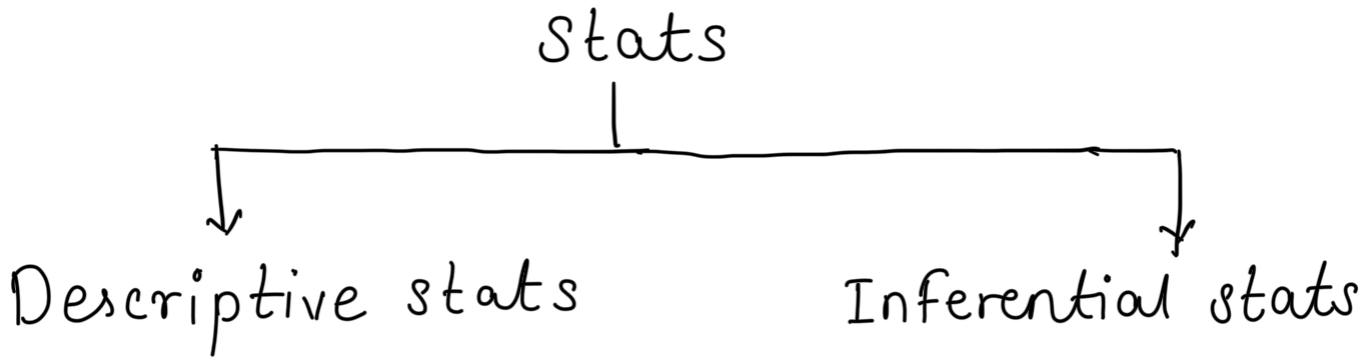
Statistics



It consists of organizing & summarizing data

1) measure of central tendency (mean, median, mode)
2) measure of dispersion (variance, std deviation)
3) Diff types distribution of data
4) Histograms

It consists of using data you have measured to form a conclusion

1) Z-test
2) t-test
3) CHI-Square test
4) ANOVA

Hypothesis Testing

Ho, H1

P-value, significance value

Part 1

* Population & Sample

Population : The group that we are interested in studying

Sample : A subset of population

Sampling Technique

The goal of sampling is to create a sample that is representative of the entire population.

Population denoted as N

Sample denoted as n

Types of sampling

1) Simple Random Sampling

When performing simple random sampling, every member of population (N) has an equal chance of being selected for your sampling (n) Random method = Lottery method

2) Stratified sampling

Stratified → strata

⇓

Layering

⇓

Non-overlapping group

It is a probability sampling method where a population is divided into distinct, non-overleeping subgroups based on shared characteristics (like gender, age, income) to ensure each subgroup is represented proportionally in Final sample leading to more accurate & less biased results.

Stratified sampling = group First, random later

eg

* A class with 100 student: 60 boys 40 girls
you want to select 10 students for survey
1) Create a strata (Group)
Gr 1 - Boys Gr. 2 - Girls
2) Sample From each grp
Pick 6 boys randomly. Pick 4 girls randomly
* This keeps the same proportion as the original class

3) Systematic sampling

A probability sampling method where researchers select items from an ordered list at fixed, regular interval (K) after choosing a random starting point, ensuring each population member has a known chance of selection for an efficient, often representative sample, useful when dealing with large populations & tight timelines.

systematic sampling = every k^{th} item
K = Population size / sample size

eg 100 students list & you want 10 students

1) Find the interval K ⇒ k = 100/10
2) Choose a random start - Pick no- beth 1 to 10 → say 4
3) Select every 10th student -

4, 14, 24, 34, 44, 54, 64, 74, 84, 94

There are your sample

eg factory checks every 20th product For quality

Bad if data has pattern. Results can be biased
eg every 10th item is defective

Convenience sampling

A non-probability research method where participants are chosen because they are easily accessible & available to the researcher, rather than through random selection, making it quick & affordable but potentially biased, it is useful for pilot studies or initial exploration but not

ideal for generalizing results, often used for quick feedback on new products.

eg. A shop owner surveys customers currently in the store

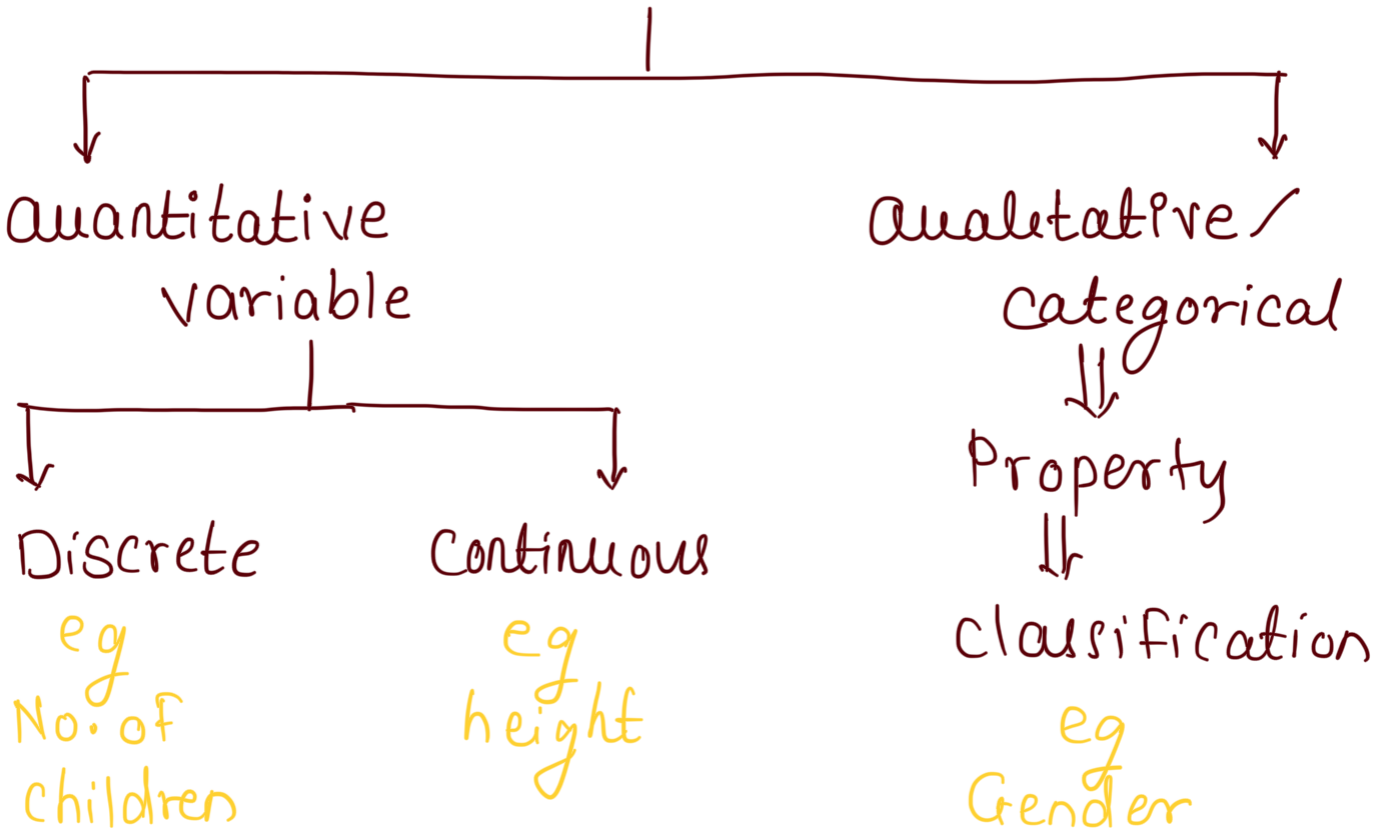
High bids
Results may not represent the whole population
Not reliable for serious research

Exit Pole	Random sampling
Disecue information	convinience / Random
Household expenses	stratified

* What are variables & it's types

variable is a property that can take on many values.

Variable



* Measure of central Tendency