

Basic stats

1) Statistics: It is the science of collecting, organising and analysing data.

→ stats are used in Research interpretations and drawing conclusions, Designing surveys, statistical modeling etc.

2) Data: Data is facts or pieces of information that can be measured.

Eg: 1) The personal info and salary of employees.
2) patients records in a hospital.
3) Voter's information (Election commission).

Types of statistics.

1) Descriptive stats.

It consists of organising and summarising data.

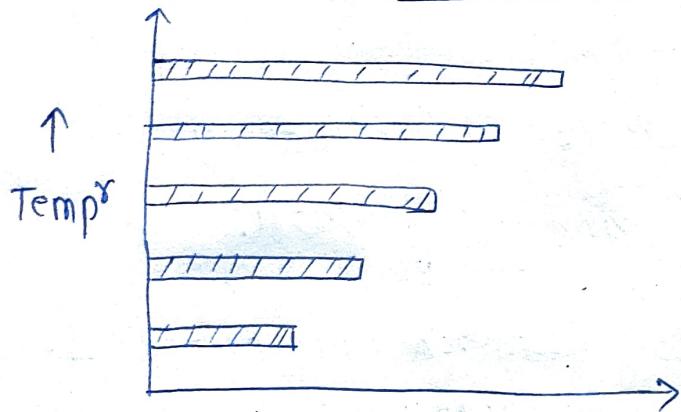
Eg: Better understanding data with graphs.



← Scatter plot

Here as Height increases weight increases.

Summertime

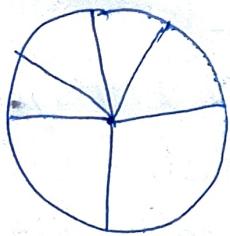


← Bar graph

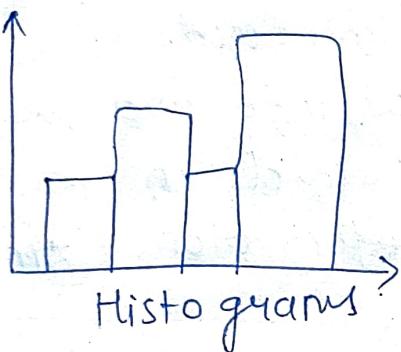
As temperature in an area increases the consumption of electricity also increases.

Electricity consumption →

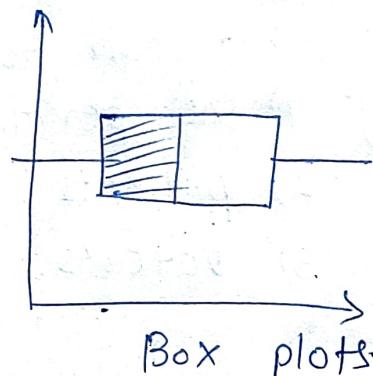
* other graphs can be



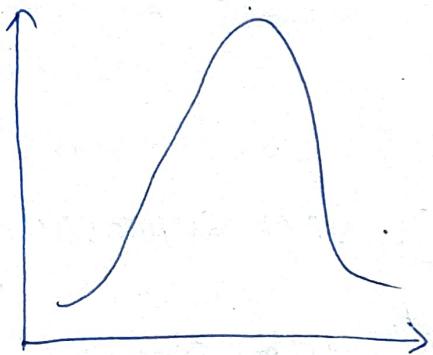
pie charts



Histo grams →



Box plots →



Distributions.

② Inferential stats

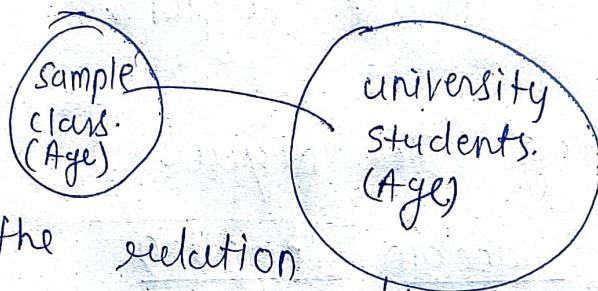
It is technique where we use data to draw conclusions.

Eg: class room (maths students) (10 students)
1st sem marks.

marks: { 84, 76, 29, 47, 56, 76, 87, 94, 99, 100 }
questions (Descriptive stats)

- 1) what is avg. marks, Avg No. of students passed., etc
- 2) percentage of students failed.

questions (Inferential stats).



- 1) what is the relation students in class with blue ages of university students.
- 2) what % of students passed in sample class. w.r.t university as a whole etc..

population (N): It is the entire data set from which we want to draw conclusion.

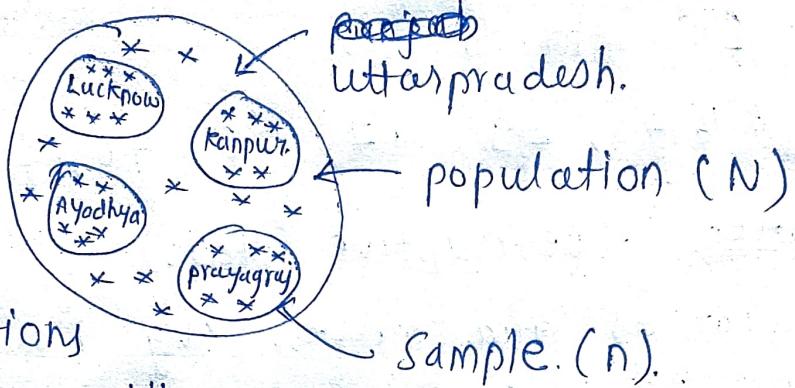
sample (n): It is a small part of population. Its purpose is that it is used for testing.

So in above example university student's age is population whereas class Age is sample.

Example : Elections in ~~poor~~ uttar pradesh

Exit polls.

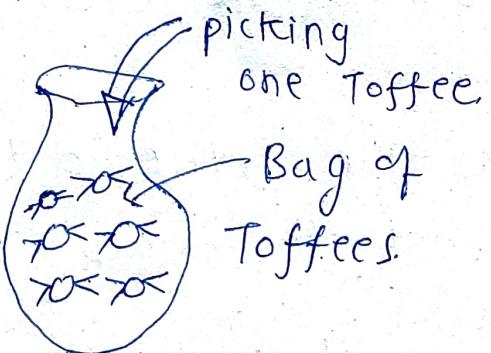
The agency conducting exit polls for elections in uttar pradesh will collect sample in every region. And using some ~~standard~~ techniques they can predict the results of elections in these regions.



* Sampling Techniques

1) Simple Random Sampling

When every one in population(N) has equal probability to get selected for sample(n). It is called simple random sampling.

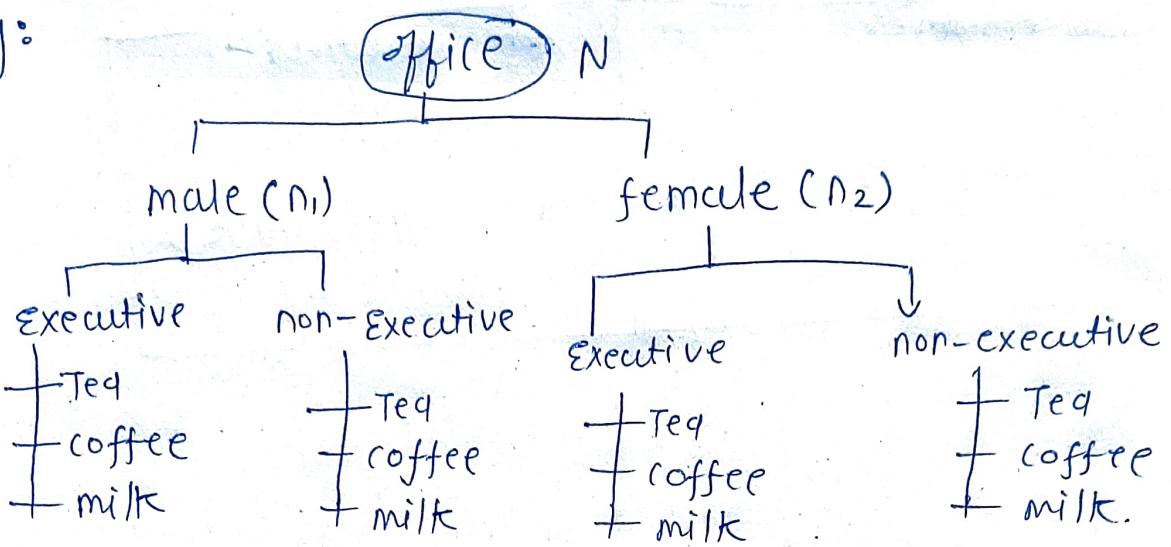


Here if a person wants to choose a toffee from a bag of orange and chocolate toffee the probability of selecting an orange toffee is equal to that of a chocolate toffee.

② stratified Sampling

It is a technique where population (N) is split into non-overlapping groups (strata) (strata \rightarrow layers \rightarrow group).

Eg:



→ In above example the population (office) is sub divided into two groups / strata Male (n_1) and female (n_2).

→ Also it can further be divided / stratified into types executives and non-executive staff.

→ Further stratification can be based on their preference of drinks in office i.e Tea, coffee or milk.

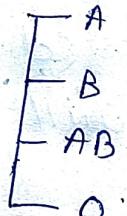
→ One thing can be noted that all above groupings are non overlapping, which is shown below:

	male			Female		
Executive →	Tea	coffee	milk	Tea	coffee	milk
non executive →	Tea	coffee	milk	Tea	coffee	milk

Office

* other examples can be

1) Blood group



2) Age

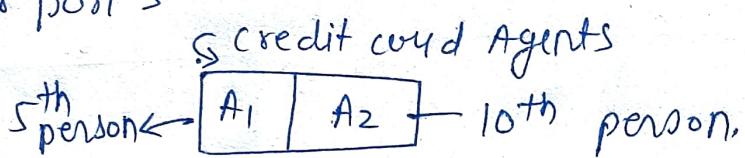
- Infant (0 - 3)
- Kid (4 - 10)
- Teen (11 - 18)
- Adult (18 - 50)
- Senior citizen (50 -).

③ Systematic Sampling:

when from population (N) every n th person is selected as a sample.

Eg: In a chocolate factory; during production every 100th chocolate toffee is selected from Batch for quality testing.

Eg: At air port's



Two credit card agents at air port decide that they will ask every 5th and 10th person for a credit card registration.

④ convenience sampling

* people having interest in a particular Survey will only participate in Sampling.

Eg: An Engineering college recently conducted on Recent trends in Automobile Technologies. So people selected to automobile Engg or mechanical Engg will only participate in this survey.

Ques: What sampling technique will be used for the following.

stratified — 18 < (Age) — 18 >

1) Exit polls : Random Sampling

2) RBI : Stratified + Random.
(Women's Survey)

3) Drug Testing ⇒ for age > 15 years.
(Stratified + Random)

* Variables

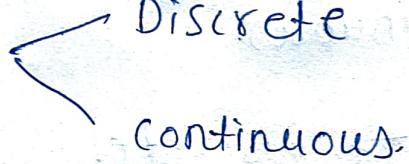
A variable or property that can take any value.

Eg: Height = { 189, 197, 165, 140 } cm.

Age = { 10, 12, 50, 65, 45 } years.

Types of Variables

1) Quantitative



2) Qualitative/ variable
categorical.

Eg: male, female, etc.

A) Quantitative variables

↳ It can be measured numerically.

Eg: Height, weight, Age, etc

① Discrete variable

→ It will contain
whole numbers.

1) No. of children in
a family

2) No. of employees
in organisations.

continuous variables

→ It can have decimal
values.

Eg: 1) Height, weight,
Age, etc.

2) Rainfall

3) Humidity, Temperature,
etc.

Ques: mention the type of variable for the following.

- 1) Gender → categorical
- 2) Ganga river length → continuous.
- 3) Marital status → categorical
- 4) movie duration → continuous.
- 5) pincode → discrete
- 6) IQ → continuous.
- 7) population of state → Discrete.