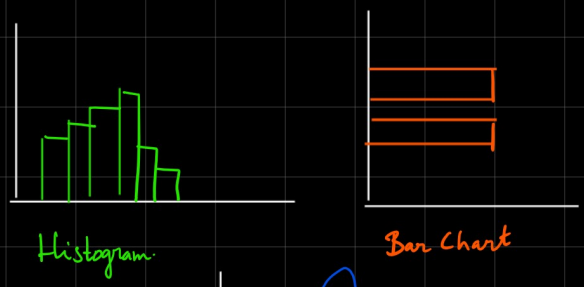
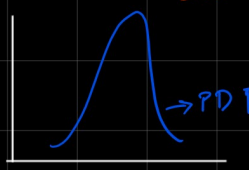
Statistics

Descriptive Stats.

* It consists of organizing and summarizing of data.

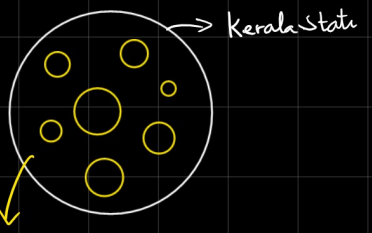




In all these diagrams we organize and summarize data (descriptive stats).

Inferential Stats.

* It consists of a technique by creating some measurements to form some conclusions.



1000 Sample data collected from people about who they voted for.

⇓

Guessing which party would win?

Conclusion

Sample → population data

⇓

hypothesis testing

Ex:- p-value, t-test, chi-square, ANNOVA

Descriptive Question: Average, Minimum, Maximum, Outlier, Distribution, Transformation.

Inferential Question: Average age of classroom is given, will it be equal to age of students in the university?

→ Sample data

→ population data.

involves some kind of hypothesis.

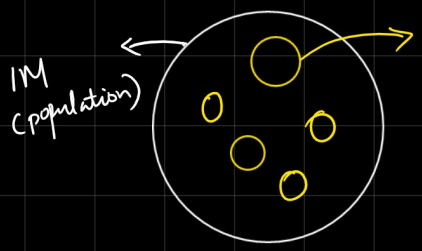
Interview Question (AMZN)

What is average size of all sharks in the world! What type of statistics is this!

ans:- (Inferential Statistics) → we consider a hypothesis with sample and population data.

Population and Sample

(finding avg. height of a city)



taking 1000 samples

⇓

(how do we pick up important samples) → we use sampling techniques

Sampling Techniques

(1) Simple Random Sampling

N → population n → sample.

:) Every member of the population (N) has an equal chance of being selected for your sample (n)

probability = $\frac{1}{1000000}$

(2) Stratified Sampling

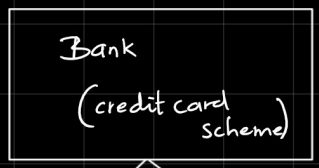
Strata → Layers ⇒ Non-overlapping.

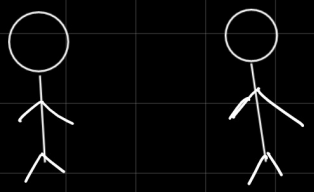
Gender → Male, Female

Blood Group → O+ve, A+ve

:) They are well seperated, they will never overlap.

(3) Systematic Sampling





Guy tries to sell scheme for every 5th person he sees.

This one tries to sell scheme for every 7th person he sees.

{ Select every n^{th} sample is selected
'n' might be any value. }

(4) Convenience Sampling :-

:) Only those people who are interested will be participating.

Data Science Survey → mail → Arts, CS, Medicine

→ only interested will participate.

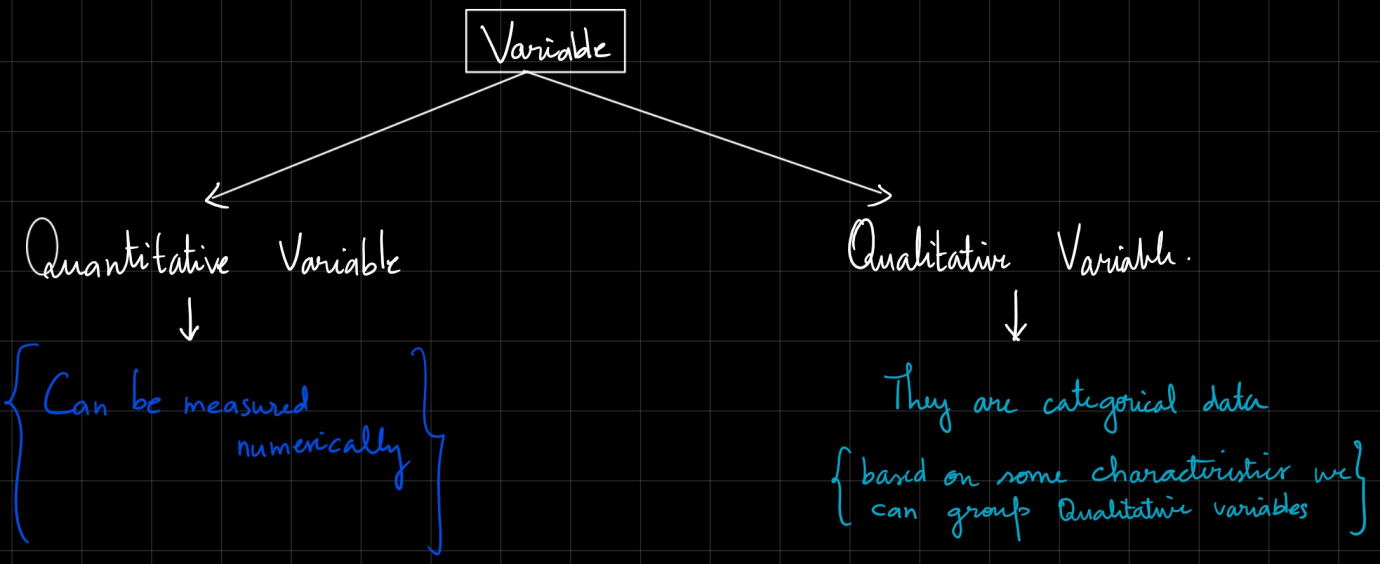
⇓

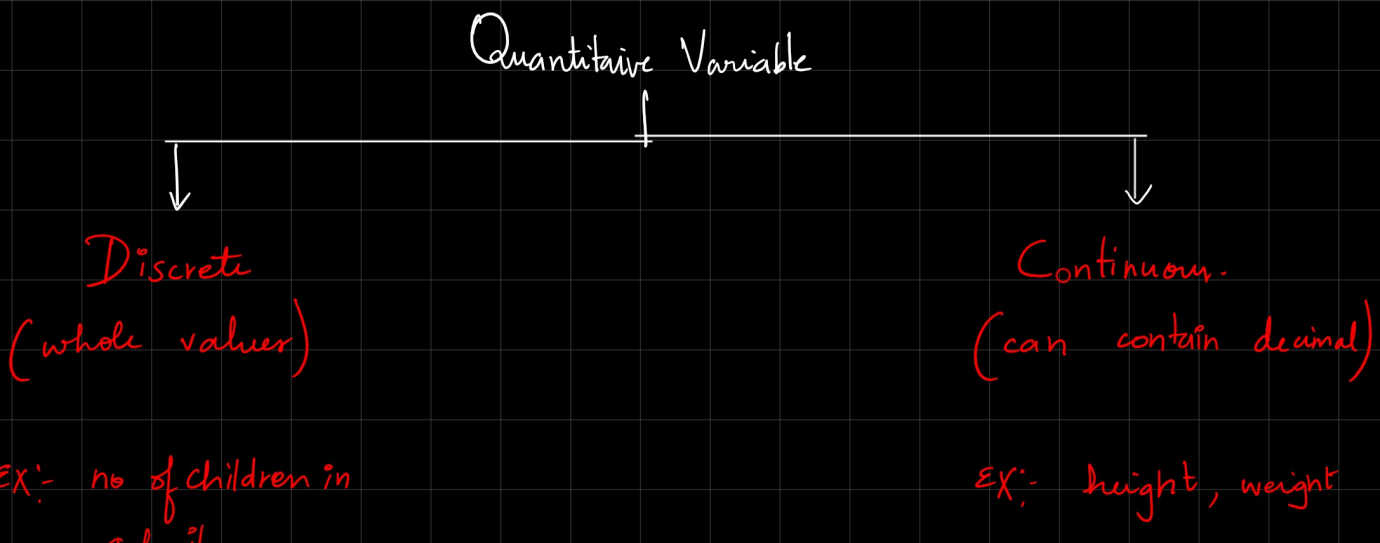
Which means we get accurate data.

has a target audience

Variable

:) A property that can take any value. Eg: Age = 24





Histograms {construction}

Sample data

Ages = { 10, 12, 17, 13, 25, 22, 23, 18, 16, 20, 13, 14 }

Step (1) sort the data.

Ages = { 10, 12, 13, 13, 14, 16, 17, 18, 20, 22, 23, 25 }

Step (2) Set bins

bin → No. of groups

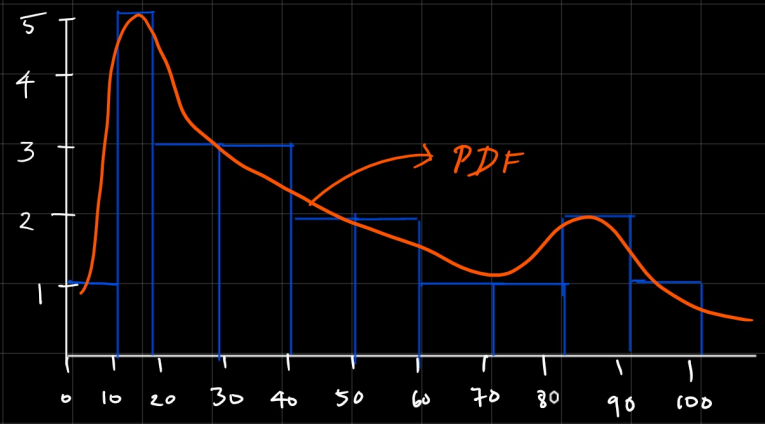
Ex: {0 – 100} bins = 10.

$\frac{100}{10}$ => Data divided to 10 groups.

Step (3) Bin Size → Size of bins.

Ex: data ranges from [0 – 100], Bins = 10

Ages = {10, 10.5, 12, 14, 18, 22, 24, 26, 37, 42, 47, 49, 51, 52, 61, 67, 77, 84, 96, 99, 100}



Bin size = $\frac{100}{10}$ = 10

PDF => {Smoothening histogram}

⇓

KDE plot

Ex: [0 – 100] Bin = 5 Bin size = $\frac{100}{5}$ = 20.

Smith

31/03/23