Date:_____

Introduction to Statistics in Data Science

Class #1

Roadmap :-

Basic Section :-

(1) Descriptive Stats (2) Inferential Stats.

(1) Descriptive Stats

i, Measure of Central tendency (mean, median, mode)
ii, Measure of Dispersion (Variance, Standered Deviation)
iii, Histograms, Pdf, CDF, Probability, Permutation
iv, Gaussian Distribution
v, log normal "
vi, Binomal "
vii, Bernoli "
viii, Power Law "
ix, Standard normal "
x, Transformation and Standarization

Date:

② Inferential test

i, 2 test
ii, T- Test
iii, Anova - Test
iv, Chisquare- Test
v, Hypothesis Testing (P, value, Confidence Interval)
vi, 2-table

Statistics:

Statistics is the Science of Collecting, Organizing and analyzing Data

Data is used for better Desision Making

Data: Facts, pieces of information that can be measured

Eg: The I.Q of the students in class
{98, 75, 65, 78, 76} → Data

The most thing in Data is "It can be measured"

Date:

Types of Statistics

Descriptive Stats: It consist of Organizing & Summarizing Data

(2) Inferential Stats: It is a technique where we use the data that we have measured to form Conclusion

Eg of D. Stats:

Classroom of maths Students (20 students marks with respect to Percentage

{84, 95, 96, 97, ...}

Q) what kind of Question may come with descriptive stats

A) ~~the~~ what Average marks of the Student in the class

(2) what is the passing Percentage of the class?

These type of Questions may come in Descriptive stats

Date:

Now,
what may the question with respect to Inferrial Stats?

A) Are the marks of the students of this classroom similar to the marks of the classroom in the college?

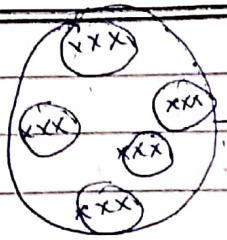
Now, ~~the~~ ~~the~~ maths classroom in the entire college is my ~~popul~~ "population" & this specific classroom is my "sample"

Suppose there are 5 different maths classroom in the entire college and I have taken the data from any one classroom, so 1 is my sample and 5 is my population

(2) Elections:

Suppose, we have an election in Karachi & we need to find out ~~the~~ who will win

Date:



→ Karachi (10, million) Population

Suppose this is entire Karachi population & it is not possible to go to every person & asked whoom will you ~~voted~~ ~~They~~ Reporters they take samples from differet areas of the city & they asked whoom did you vote. So, the entire population of Karachi is my population & the group of people from different ~~types~~ ~~of~~ areas of karachi is my sample. Population is given by (N) & Sample is given by (n)
We have different types of taking samples.

Sampling Techniques:

So, Now the questions arises why you are taking sample randomly

(i) Simple Random Sampling: Every member of the population has an equal chance of being selected for the sample. This technique is used for simple testing like Elections.

Example:

for the exit poll you can use this technique but test like check of some medicines

Date:

you can't use this technique because for medicine test you have to check person's medical history etc. so, we have condition to use sampling in that condition.

2. Stratified Sampling: It is a technique where the population is split into non-overlaping groups.

Example:-

lets consider Genders (male, female) and I want to do a survey and for a survey I will require some people. So, based on Genders my sample will be divided into male & female because male may give different kind of survey & female may give different kind of survey. so, on that basis we have groups.

Example 2:

On the basis of age groups we can also do stratified sampling

Ages

(0-9) (10-20) (21-30) (30-60) (61-100)

Q) Based on Profession can we do stratified sampling?

Date:

Professions: Dot net developer Php developer

Phython developer

Ans) No, we can't do, because It may be possible a php person may know Don't net, a dot net person may know python so, there is some possibility of overlapping.

But Stratified Sampling can apply to engineers, Doctors

3. Systematic Sampling: From the population

(N) we pick every n^{th} individual (N) ⟶ n^{th} individu

Example: let's consider we are outside the mall & we are taking survey regarding covid, we are choosing any 8^{th} person & take a survey from that. 8^{th} person is an example we can take any person

Question: If we snap the finger, what kind of sampling technique we can use?

Answer: Simple random Sampling

Date:

4. Convienience Sampling:

Domain Expert survey means who are expert in particular Domain ~~to~~ that can participate on this survey

Example:

Suppose we are doing a survey related to Data Science and we found the people those people who are intested or expert in data Science can do this survey then it will ~~beco~~ become "convienience Sampling". Because with the help of survey we will find out some data.

NOTE :-

#1. Based on the use-case you will decide which sampling technique will you use.

Q) A drug needs to be tested so, what kind of sampling technique will you ~~use~~ use.

A. Stratified sampling. Because it depends on the use case means if drug is for every one we can use random sampling but drug is for some age groups so we will use stratified sampling.

Date:

Variables:

A variable is a property that can take on any value

Eg: weight, height etc. It can have any value

Two kinds of Variable:

1- Qualitative variable / catogerical variable
2- Quantitative variable

1- Quantitative Variable: It has some proper

1- It can be measured numerically (Add.
eg: Age, Height, weight etc.

2. Qualitative or Catogerical variable:-
We have catogeries in Catogerical vari
Here we can not do any measurment
Add. Subtract etc.

Eg: Gender, Blood groups etc

Date:

Example Scenario:-

let's we have some IQ classes like.

0-10	10-50	50-100
↓	↓	↓
less IQ	Medium IQ	Good IQ

Based on some characteristics we have some catogories ~~which~~ which is called IQ.

Quantitative variable also have two different kinds of catogories

Quantitative

- Discrete Var
  Eg: whole numbers,
  Bank account,
  No of children
  {2,3,4,5}
- Continious Variable
  Eg: height, weight
  {172.5, 163.5}
  Amount of rainfall in inches {7.6, 5.2}

Date:

Questions:

1. what kind of variable Genders is?
   Catogerical variable
2. " Rivers length?
   Continious Variable
3. " population in city
   Quantative (Discrete)
   "

Variable Measurment Scale:

There are 4 types of variable measurment Scale

1- Nominal
2- Ordinal
3- Intervual
4. Ratio

Variable measurment types is very important bec
in Datasets we have these types of variable s
we have to learn this so, we are able to
do good EDA.