

## *statistics*

### *Python – part1 and part – 2*

*What comes to your mind*

*calculations*

*mean – median – mode*

*analysis*

*graphs*

*hypothesis*

*variance sd*

*You want to choose an option, suppose Virat Kohli is an opening batsman*

- 1) Gather the data*
- 2) Organize the data*
- 3) analyse the data*
- 4) interpretations*
- 5) Present the data*
- 6) Draw the conclusions*

➤ *Data divided into two parts*

- *Categorical data*
  - *English*
  - *Qualitative data*
- *Numerical data*
  - *Quantitative data*

➤ *Numerical data*

- *Continous data*
  - *Float type data*
  - *Weight : 60.5 kgs*
  - *Height: 152.5 cm*

- *Discrete data*
  - *int*
  - *countable*
  - *Roll numbers*
  - *Backlogs*

<i>Categorical data</i>	<i>Numerical data</i>
○ <i>Qualitative data</i>	○ <i>Quantitative data</i>
	• <i>Continoues data : Float</i>
	• <i>Discrete data : Int type</i>

### *1) Nominal level*

- *Categorical type*
- *It is just a name*
- *There is no relation between the names*
- *Any names*

### *2) Ordinal level*

- *Categorical type*
- *It maintains some order*
- *if you see the words, we feel some order*
- *Flop average hit superhit blockbuster*
- *Bronze silver gold*
- *Promary seconday ter*

### *3) Interval level*

- *Numerical data*
- *it does not have zero scale*
- *it has negative values*
- *ex: Temperature*

### *4) Ratio level*

- Numerical data
- Has zero point
- it does not have any negative values
- length mass

Q) What is the example of Interval level

Ans: Temperature

Q) why?

Ans: It does not have zero scale

It can have negative values

Q) Suppose the temperature in Hyderabad is = 50°C,      BLR = 25°C

can I say:  $Hyd_{temp}$  is twice of  $BLR_{temp}$

Cand: Yes

Inter:  $Hyd = 2 * BLR$

$$50^{\circ}C \equiv 122^{\circ}F$$

$$25^{\circ}C \equiv 77^{\circ}F$$

$$\frac{25}{50} = \frac{77}{122}$$

Father's weight = 100kgs    son's weight = 50kgs

1kg = 2.5pounds

$$\frac{100}{50} = \frac{250}{125}$$

Note: Ratio level means,    If you take the values in any measurement

Ratio should be equal

<i>Categorical data</i>	<i>Numerical data</i>
○ <i>Qualitative data</i>	○ <i>Quantitative data</i>
	• <i>Continoues data : Float</i>
	• <i>Discrete data : Int type</i>
<i>Nominal</i>	<i>Interval</i>
<i>Ordinal</i>	<i>Ratio</i>

*Population vs sample*

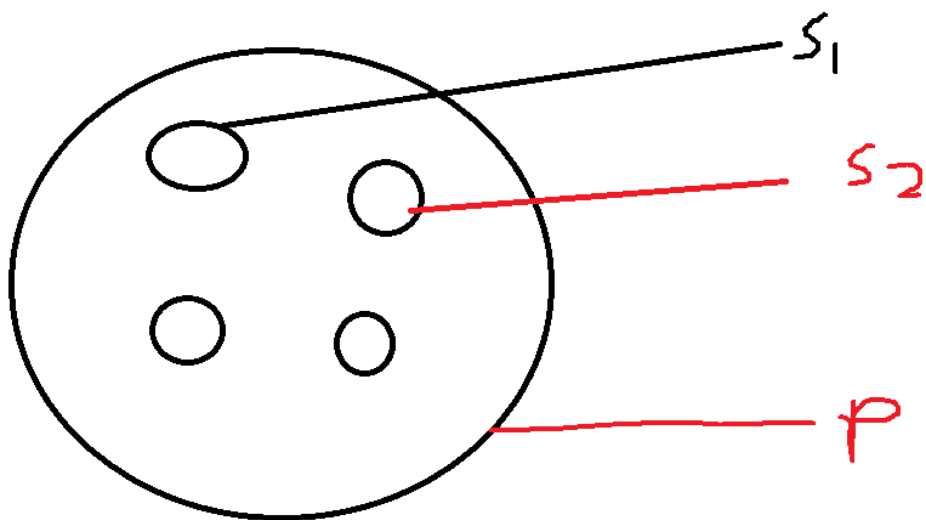
*Population: collection of all th data or objects or items*

*Each and every person in the world population*

*Sample: sub set of population*

*Consider entire world as population*

*Asia becomes sample*



*Real population never exist*

*Client is 1lakh customer data ===== this is a population (assume)*

*Working on population is complex*

*You cant got each and every home and take the conclusion*

- *Resource*
- *Money*
- *time*

*For examples you want identify the average package for NareshIT students*

*You will go to each batch, in each batch you will pick 10 students*

*ask the package of those 10 students*

*calculate the average*

*12LPA*

*Conclusion: The average package of NareshIT*

*Inference*

***Inferntial statistics: Analyse on the sample, draw the conclusion on population***

*Descriptive statistics: Analyse on the population, draw the conclusion on population*