UNIT 1

Data Levels of Measurement, Summary Statistics

Presented By :

Aditya Joshi

Assistant Professor

Graphic Era Deemed to be University, Dehradun

Data Levels of Measurement

- It is a way that a variable is measured

A variable has four different levels of measurements

1. Nominal
2. Ordinal
3. Interval and
4. Ratio

Nominal level of measurement(Non Numeric or where the numbers have no values)

- The numbers in the variable are used only to classify the data.
- In this level of measurement, words, letters, alphanumeric characters are used.

E.g. There are three gender categories

Female is classified as F

Male is classified as M

Transgender is classified as T

Jersey number of players way to identity them

This type of assigning classification is called nominal

no particular order

Ordinal Level of Measurement(order matters but the differences does not matter)

- This level of measurement depicts some ordered relationships among the variables observations.

e.g.

a student score highest CGPA 10 in MCA assigned 1st rank

Second highest 9.8 2^{nd} rank

Third highest 9.5 3^{rd} rank

Rating

Order of finishing a race

This indicates the ordering of measurements

it is in a ranked order

Interval level of measurement(order and differences between levels of a variable matter but not ratio)

- It not only classifies and orders the measurement but also specifies distance between each interval on the scale are equivalent along the scale from low interval to high interval.

e.g.

Measurement of anxiety in student between the score of 10 and 11

This interval is the same as that of student who scores between 40 and 41

Temp in centigrade : the temperature between 94 deg C and 96 deg C

Is same as between 100 deg C and 102 deg C

Ratio Level of Measurement(how much have one thing we have in comparison to another thing order, differences and ratio are all meaning full)

- In addition to having equal intervals can have value of zero as well

Absolute zero 0 deg kelvin

Zero starting point

Meaningful zero point

E.g. no of student in class

No of patients seen by doctor

Summary Statistics

In Descriptive statistics, summary statistics summarize and provide information about your sample data. It tells you something about the values in your dataset. This includes where the average lies and whether your data is skewed.

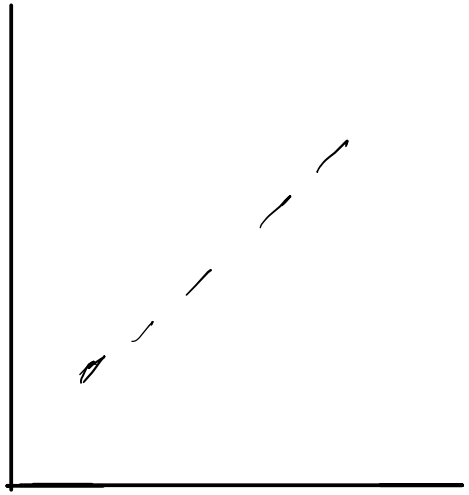
It falls into three main categories

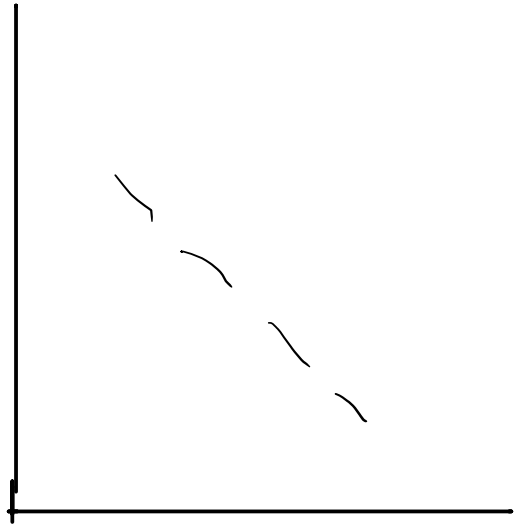
1. Measures of location(also called central tendency)
2. Measures of spread
3. Graphs and chart

Measures of location

- It tells you where your data is centered at or where a trend lies (upward or down wards).

There is also shift in data over time





1. Mean(also called arithmetic mean or average)
2. Geometric mean(used in interest rate and other type of growth)
3. Trimmed mean(the mean with outliers excluded)
4. Median(the middle of dataset)
5. Mode(the most frequent data in dataset)

Mean vs median

Both are measures where the center of the dataset lies, but they are usually different number

Question : 10, 10, 20, 40 70

- The mean is found by adding numbers together and dividing by the numbers of items in the set.
- The median is exact middle from ordering the set from lowest to highest.
- Some time it will be a same number (mean and median)

Question: 4, 2, 6, 7, 1

Mean vs Average

In Maths we call it Avg.

In Stats we call it mean

Specific mean commonly used in stats

- Mean of sampling distribution: used with probability distribution specially with the central limit theorem, its an avg of a set of distributions.
- Sample mean: The avg value in a sample
- Population mean: The avg value in a population

Other types of mean in Maths

1. Weighted mean
2. Harmonic mean
3. Geometric mean
4. Arithmetic geometric mean
5. Root-mean square mean
6. Heronian mean

Mode: it is the most common number in the set

Question: 21, 21, 21, 23, 24, 26, 26, 28, 29, 30, 31, 33

Median: The median is the middle number in a dataset to find the median, list your data point in ascending order and find the middle number.

Question: 24, 23, 28, 26, 29, 26, 30, 33, 31

If you have even set of numbers, average the middle two to find the mean.

Question: 23, 24, 26, 26, 28, 29, 30, 31, 33, 34

Question: Find mean mode and median for:
21, 23, 23, 54, 67, 21, 25, 21, 54, 72, 75

If there are two numbers that appear most often (and the same number of times) then the data has two modes. This is called bimodal. If there are more than then the data would be called multimodal. If all the numbers appear the same number of times, then the data set has no modes.

$$\text{Mean} = \frac{\text{Sum of the observations}}{\text{Total number of observations}}$$

$$\text{Median} = \frac{n+1}{2} \text{ th term if n is Odd}$$

$$\text{Median} = \frac{(\frac{n}{2})\text{th term} + (\frac{n}{2}+1)\text{ th term}}{2}$$

Mode is the most frequent items