→ Measures of central tendency

(1) Measure of central tendency is a summary statistic that represents central point or typical value of a dataset.

(2) This measures indicates where most of the value in the distribution falls and are also known as central location of distribution.

(3) You can think of it as a tendency of data to cluster around a middle value.

uses of measure of central tendency

(1) A measure of central tendency can be used as a standard for judging the relative positions of other items in the same set of data. (whether a number falls above or below the average and how far it is from average)

(2) A measure of central tendency can be used to compare the relative sizes of two different sets of data (for example: compare the average of two sets of data).

(3) We get a picture for the variablity (spread) of the data by looking at the dispersion (grouping of individual observations around the average). This helps us to determine the consistency among the observations.

→ Mean

The mean is arithmetic average. The value where the set of data balances also called as "point of balance" when each data value is stacked on a dataline.

Let's have a data set with values $\{x_1, x_2, x_3 \dots x_n\}$

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

where $\bar{x}$ = sample mean

n = sample size

$$\boxed{\mu = \frac{\sum_{i=1}^{N} x_i}{N}}$$

where μ = population mean
N = population ~~sample~~ size.

"The mean is the most common measure of central tendenc but has a downside as it is easily affected by outliers."

→ Median

(1) The median is the middle value that splits the data set into half. It is the middle value in distribution when values are arranged according to the size.

(2) It is the value of the variable that divides a set of data into two equal groups so that half the observations have values smaller than the median, and half the values larger than the medium.

(3) The median is measure of choice when a numerical variable has some few unusually low or high values in the data set. If this occurs then mean will be ~~pooled~~ pulled away from the center and not be representative in majority of cases.

We need to arrange elements in ascending order. The method of finding median varies whether your data has odd or even no. of elements.

→ If the number of obs(n) is odd: the median is value of the position $\boxed{\left(\frac{n+1}{2}\right)}$

→ If no. of observation (n) is even

1. Find value of position $\left(\frac{n}{2}\right)$
2. Find value of position $\left(\frac{n+1}{2}\right)$
3. Find avg of two values to get median.

→ Mode

(1) The mode is the value that occurs the largest number of times in a data set or the response category of variable that is most frequently choosen by respondents.

(2) In bar chart or histogram, mode is the tallest bar.

(3) when a distribution has one mode, we say it as uni-modal. If it has two modes we say it as bi-modal. If there are several modes we say it as multi-modal.

(4) If no values repeats data has no mode.

→ Measures of central tendency with level of measurement.

(1) Nominal: Measures of central tendency are applied to the frequencies found in different categories of a nominal variable.

In nominal data mode is only measure of central tendency that can be used.

(2) Ordinal: Either median or mode is the measure of central tendency that can be used. Here, median is preferred measure of central tendency.

(3) Interval or Ratio: The mean, median and mode may be used as measures of central tendency.

- For normal distribution, mean is the most preferred measure of central tendency.
- For skewed distribution, median is preferred over mean and mode.

The mode is only measure of central tendency that can be used for all levels of measurements.

Eg: Given a set of marks of students in a class. Find mean, median and mode.

Marks = {40, 32, 42, 40, 15, 25, 40, 10, 32, 40, 37, 23, 18, 29, 41}

Soln: No. of observations = 15

$$\text{Mean} = \frac{\sum_{x=1}^{n} x_i}{n} = \frac{464}{15} = \underline{\underline{30.93}}$$

Marks in asc order = {10, 15, 18, 23, 25, 29, 32, 32, 37, 40, 40, 40, 40, 41, 42}

No. of obs = 15 (odd)

$$\text{Median} = \left(\frac{n+1}{2}\right)^{th} \text{element} = \left(\frac{15+1}{2}\right)^{th} \text{element} = \left(\frac{16}{2}\right)^{th} = 8^{th} \text{ element}$$

$\therefore$ 8^{th} element = $\underline{\underline{32}}$. ie. median.

Mode is element with highest frequency

Elements	Frequency
10	1
15	1
18	1
23	1
25	1
29	1
32	2
37	1
40	4
41	1
42	1

Highest frequency is of 40

$\therefore$ Mode = $\underline{40}$

Let's take the above example and change some values and observe

marks = { 40, 32, 42, 65, 15, 25, 40, 10, 32, 40, 37, 23, 18, 29, 83 }

$$\text{mean}\ (\bar{x}) = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{531}{15} = 35.4$$

Asc order = { 10, 15, 18, 23, 25, 29, 32, 32, 37, 40, 40, 40, 42, 65, 83 }

$n = 15$ (odd)

$$\text{median} = \left(\frac{n+1}{2}\right)^{th} = \left(\frac{15+1}{2}\right)^{th} = \frac{16^{th}}{2} = 8^{th} \text{ element}$$

$$= 32.$$

→ We can see significant change in the mean whereas the median has no changes.

→ This is because the calculation of mean incorporates all values in data. If you change any value mean changes unlike mean, median values does not depend on all values in data set.

→ Consequently, when some of the values are more extreme, the effect of median is smaller. Of course, with other type of changes median can change.

→ One major disadvantage of mean is that it is particularly susceptible to extreme values or outliers in data.