

MEASUREMENT OF MEAN OF DATA HAVING EXTREAME BOUNDORY VALUES

Nirajkumar Kantilal Patil.

(BE Mechanicle , MSc Mathematics)

(Data Scientist)

ABSTRACT:- Apart from calculate the central tendency of scatter data, if data contain few extreme frequency in given data; then Central tendency might have some variation as compare to practical central tendency. This article describes how to overcome error and we can find it by other some means.

MEAN

Mean is most common central tendency used to analyse the data by *DATA SCIENTISTS*. Mean is nothing but average of the given Data. The average or mean is to be calculating by adding all frequency and divide it by number of frequency.

PROBLEM IN MEAN OF SCATTER DATA

Mean is on of the simple central tendency and most commonly used. According to Three central tendencies (Mean, Median, and Mode). Mean is

Shifted towards the tail of distribution. In case the data is not evenly spread then the mean is shifted towards the extreme uneven frequency .If data is not evenly spread and we calculate it by simple way (average) then it gives the error between calculated value and practical value.

According to *DATA SCIENTISTS*, Most of the firm follow the path suggested by them. But what happen when the analysis is have some error. Most of *DATA SCIENTISTS* are not aware about mathematics term, they just known the basic mathematics.

So here we will take one example and learn how to fix the error arises in the case of scatter data.

EXAMPLE:- Suppose we have to find the mean of

The given data :

1.5, 1.7, 1.9, 2.0

ANS:-

$$\text{Mean} = \frac{\sum x}{n}$$
$$\text{Mean} = \frac{1.5 + 1.7 + 1.9 + 2.0}{4}$$
$$\text{Mean} = 1.775$$

It is correct for evenly spread data. But not sufficient for the data which is not evenly spread. Like (1.3, 1.7, 1.9, 25.8). In this case 25.8 is not evenly spread. So to find the correct value around which the data is spread.

HOW TO FIX THE ERROR OF MEAN

It is Find by foolowing the following steps

1. Find the mean of the given data.
2. Find the mean by dropping extreame value.
3. Now we have to two means.
4. Find the mean of two means by 1,2

EXAMPLE:- Suppose we have to find the mean of

The given data :

2, 14, 13, 16, 15, 52

ANS:- Drop the extream value 0 first

$$\text{Mean1} = \frac{\sum x}{n}$$
$$\text{Mean1} = \frac{14 + 13 + 16 + 15 + 52}{5}$$
$$\text{Mean1} = 22$$

Now, drop the second extreame value

$$\text{Mean2} = \frac{\sum x}{n}$$
$$\text{Mean2} = \frac{2 + 14 + 13 + 16 + 15}{5}$$
$$\text{Mean2} = 10$$

Now, take all values

$$\text{Mean} = \frac{\sum x}{n}$$
$$\text{Mean} = \frac{2 + 14 + 13 + 16 + 15 + 52}{5}$$
$$\text{Mean2} = 18.67$$

Now find the mean of the three (Mean, Mean1, Mean2) that is our value around which our data spread.

$$\bar{X} = \frac{\text{Mean1} + \text{Mean2} + \text{Mean}}{3}$$

$$\text{MEAN} = \frac{10 + 22 + 18.67}{3}$$

$$\text{MEAN} = 16.89$$

CONCLUSION:-

When we have to analyse the data on the basis of Mean, then first check whether the data is evenly spread or not. Then use the concept of this article. This is basically used by DATA SCIENTISTS. DATA SCIENTISTS is the one of the branch which mostly impact by MATHEMATICS

REFERENCES:-

1. Atkinson, A. (1994). Fast very robust method for the detection of multiple outliers. Journal of the American Statistical Association. 89:1329-1339
2. Hawkins, D. (1980). Identification of Outliers. Chapman and Hall. London.
3. Atkinson, A.C, Riani, M. And Cerioli, A. (2004). Exploring Multivariate Data with the Forward Search. Springer-Verlang. New York.
4. Barnett, V. And Lewis, T. (1994). Outliers in Statistical Data. John Wiley. New York.

<https://www.canva.com/graphs/>

