lecture 7. Measures of Dispersion and Variability

Coefficient of variation (CV):

The coefficient of variation (CV) is a statistical measure of the dispersion of data points in a data series around the mean. The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from one another.

The coefficient of variation is helpful when using the risk/reward ratio to select investments. For example, in finance, the coefficient of variation allows investors to determine how much volatility, or risk, is assumed in comparison to the amount of return expected from investments.

Ideally, the coefficient of variation formula should result in a lower ratio of the standard deviation to mean return, meaning the better risk-return trade-off. Note that if the expected return in the denominator is negative or zero, the coefficient of variation could be misleading.

$$\textbf{Coefficient of variation (CV)} = \frac{\textbf{standard deviation}}{\boldsymbol{Mean}}$$

Example: Find CV of {13,35,56,35,77}

Solution:

Number of terms (N) = 5

Mean:

$$\text{Mean} = \frac{13 + 35 + 56 + 35 + 77}{5} = 43.2$$

$$S.D = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \bar{X})^2}{n}}$$

$$S.D = \sqrt{\frac{(13-43.2)^2 + (35-43.2)^2 + (56-43.2)^2 + (35-43.2)^2 + (77-43.2)^2}{5}}$$

S.D=24.25

$$\text{Coefficient of variation (CV)} = \frac{\text{standard deviation}}{Mean}$$

$$\text{Coefficient of variation (CV)} = \frac{24.25}{43.2}$$

$$\text{Coefficient of variation (CV)} = 0.5614$$

Standard Error:

The standard error is a statistical term that measures the accuracy with which a sample distribution represents a population by using standard deviation. In statistics, a sample mean deviates from the actual mean of a population—this deviation is the standard error of the mean.

It is used to measure the amount of accuracy by which the given sample represents its population.

When you take measurements of some quantity in a population, it is good to know how well your measurements will approximate the entire population.

A large standard error would mean that there is a lot of variability in the population, so different samples would give you different mean values.

A small standard error would mean that the population is more uniform, so your sample mean is likely to be close to the population mean.

$$\text{Standard Error (SE)} = \frac{Standard\ Deviation}{\sqrt{N}}$$

Where: N is the number of observation.

Example

Calculate the standard error of the given data:

(5, 10, 12, 15, 20)

Solution: First we have to find the mean of the given data;

Mean = (5+10+12+15+20)/5 = 62/5 = 10.5

Now, the standard deviation can be calculated as;

$$s = \sqrt{\frac{(5-10.5)^2 + (10-10.5)^2 + (12-10.5)^2 + (15-10.5)^2 + (20-10.5)^2}{5}}$$

After solving the above equation, we get;

S = 5.35

Therefore, SE can be estimated with the formula;

$$\text{Standard Error (SE)} = \frac{Standard\ Deviation}{\sqrt{N}}$$

$$\text{SE} = \frac{5.35}{\sqrt{5}} = 2.39$$

Mean Deviation

In statistics and mathematics, the deviation is a measure that is used to find the difference between the observed value and the expected value of a variable. In simple words, the deviation is the distance from the center point. Similarly, the mean deviation is used to calculate how far the values fall from the middle of the data set. In this article, let us discuss the definition, formula, and examples in detail.

The mean deviation or average deviation is defined as a statistical measure that is used to calculate the average deviation from the mean value of the given data set. The mean deviation of the data values can be easily calculated using the below procedure.

$$M.D = \frac{\sum |Xi - \bar{X}|}{N}$$

Σ: represents the addition of values

X: represents each value in the data set

$\bar{X}$: represents the mean value of the data set

N: represents the number of data values

| | : represents the absolute value, which ignores the “-” symbol

$$M.D = \frac{\sum |Xi - \bar{X}| * fi}{\sum fi}$$

Example:

Determine the mean deviation for the data values:

13 ،7 ،15 ،6 ،9 ،12 ،8

$$\bar{x} = \frac{\sum x_i}{n} = \frac{70}{7} = 10$$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
8	-2	2
12	2	2
9	-1	1
6	-4	4
15	5	5
7	-3	3
13	3	3
$\sum 70$	0	20

$$\therefore M.D = \frac{\sum \lvert x_i - \bar{x} \rvert}{n} = \frac{20}{7} = 2.85$$

Calculate the mean deviation about the mean for the given data.

16−14	14−12	12−10	10−8	8−6	6−4	**Classes**
2	7	6	8	5	3	**Frequency**

Solution

Classes	Frequencies F_i	x_i	$F_i x_i$	$\bar{x} - x_i$	$\lvert \bar{x} - x_i \rvert$	$F_i \lvert \bar{x} - x_i \rvert$
6−4	3	5	15	5−	5	15
8−6	5	7	35	3−	3	15
10−8	8	9	72	1−	1	8
12−10	6	11	66	1	1	6
14−12	7	13	91	3	3	21
16−14	2	15	30	5	5	10
	31		309	0		75

$$\overline{X} = \frac{\sum f_i x_i}{\sum f_i} = \frac{309}{31} = 10$$

$$M.D = \frac{\sum f_i \left| x_i - \overline{x} \right|}{\sum f_i} = \frac{75}{31} = 2.4$$

Advantages and disadvantages of measures of dispersion

Measures of Variability	Advantages	Disadvantages
Range	It is easier to compute	The value of range is affected by only two extreme scores
	It can be used as a measure of variability where precision is not required	It is not very stable from sample to sample
		It is not sensitive to total condition of the distribution
		It is dependent on sample size, being greater when sample size is greater
Inter quartile Range	It is less sensitive to the presence of a few very extreme scores than is standard deviation	The sampling stability of IQR is good but it is not up to that of standard deviation
	If the distribution is skewed, IQR is a good measure of variation.	
Standard Deviation	It is resistant to sampling variation It is of high use both in descriptive and inferential statistics	It is responsive to exact position of each score in the distribution
		It is more sensitive than IQR to the presence of few extreme scores in the distribution
Variance	provide a summary of individual observations around the mean	sensitive to outliers
Coefficient of variation	used to compare two or more distribution that have different means	Does not vary with the magnitude of the mean