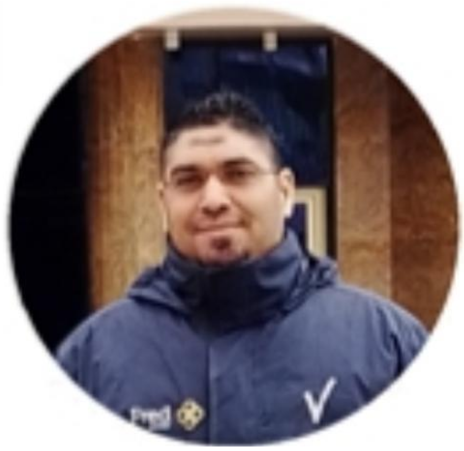
Data analysis

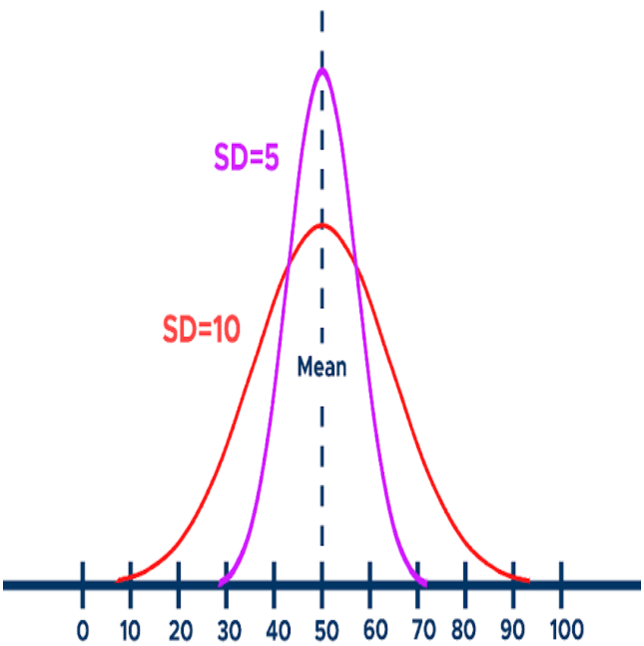


View statistical data

Measures of the Spread of Data

Measures of Shape _ Moments

Part 10



In this part, we present a vital measure of the shape used In data analysis. The measure of moments is very accurate In describing data dispersion, as It is indispensable in physics and medicine because of Its consistent effect in describing data .dispersion

Measures of the Spread of Data

Measures of Shape

Introduction

If measures of central tendency and measures of dispersion allow summarizing the data of any phenomenon in the form of numbers that give an idea of the characteristics of the distribution of these data and the degree of their homogeneity or difference, then this description still lacks the accuracy required to know the properties of the distribution, especially with regard to the spread of The data is on the represented curve in terms of its distortion and its flattening from the normal position. The need to know how the data is spread and distributed prompts us to calculate coefficients that give us quantitative estimates for either torsion (torsion coefficient) or kurtosis (kurtosis coefficient). This, in turn, requires knowing how to calculate what is known as moments.

Moments

Moments are statistical values that are around the starting point (zero) or the arithmetic mean ($\bar{x}$), while the rank of Torque is determined by the degree of force (the exponent) to which the values are raised or their deviations from zero or the arithmetic mean, and on this basis we distinguish There are two types of moments: eccentric moments and central moments , The general expression for moments is:

$$\frac{\sum f_i(x_i - b)^r}{\sum f_i} = \sum Fr(x_i - b)^r \quad ; b = \begin{cases} 0 \\ \bar{x} \end{cases}$$

Eccentric moments (m_r):

Is a statistical value of order r where $r \in N$ is centered around zero (b = 0) The methods for calculating the eccentric moment differ according to the nature of the data.

- **Non-recurring discrete data**:

$$m_r = \frac{\sum x_i^r}{n}$$

The first four decentralized moments in this case are as follows:

$$m_2 = \frac{\sum x_i^2}{n} = (MQ)^2 \qquad m_3 = \frac{\sum x_i^3}{n} \qquad m_4 = \frac{\sum x_i^4}{n}$$

Example

Calculate the first four decentralized moments for the following data: 10 9 8 7 5 3 1

Solution

x_i	x_i^2	x_i^3	x_i^4
1	1	1	1
3	9	27	81
5	25	125	625
7	49	343	2401
8	64	512	4096
9	81	729	6561
10	100	1000	10000
43	**329**	**2737**	**23765**

$$m_1 = \frac{\sum x_i^1}{n} = \frac{43}{7} = \mathbf{6,14}$$

$$m_2 = \frac{\sum x_i^2}{n} = \frac{329}{7} = \mathbf{47}$$

$$m_3 = \frac{\sum x_i^3}{n} = \frac{2737}{7} = \mathbf{391}$$

$$m_4 = \frac{\sum x_i^4}{n} = \frac{23765}{7} = \mathbf{3395}$$

- **Recurring discrete data:**

$$m_r = \frac{\sum f_i x_i^r}{\sum f_i} = \sum Fr \times x_i^r$$

The first four decentralized moments in this case are as follows:

$$m_1 = \frac{\sum f_i x_i}{\sum f_i} = \bar{x} \qquad m_2 = \frac{\sum f_i x_i^2}{\sum f_i} = (MQ)^2 \qquad m_3 = \frac{\sum f_i x_i^3}{\sum f_i} \qquad m_4 = \frac{\sum f_i x_i^4}{\sum f_i}$$

- **Connected data**

$$m_r = \frac{\sum f_i c_i^r}{\sum f_i} = \sum Fr \times c_i^r$$

The first four decentralized moments in this case are as follows:

$$m_1 = \frac{\sum f_i c_i}{\sum f_i} = \bar{x} \qquad m_2 = \frac{\sum f_i c_i^2}{\sum f_i} = (MQ)^2 \; m_3 = \frac{\sum f_i c_i^3}{\sum f_i} \qquad m_4 = \frac{\sum f_i c_i^4}{\sum f_i}$$

Example

Calculate the first four decentralized moments for the following data:

Classes	1 - 3	3 – 5	5 - 7	7 – 9	9 – 11
f_i	5	4	2	3	6

Solution

Classes	f_i	c_i	$f_i \times c_i$	c_i^2	$f_i \times c_i^2$	c_i^3	$f_i \times c_i^3$	c_i^4	$f_i \times c_i^4$
1 - 3	*5*	*2*	*10*	*4*	*20*	*8*	*40*	*16*	*80*
3 - 5	*4*	*4*	*16*	*16*	*64*	*64*	*256*	*256*	*1024*
5 - 7	*2*	*6*	*12*	*36*	*72*	*216*	*432*	*1296*	*2592*
7 - 9	*3*	*8*	*24*	*64*	*192*	*512*	*1536*	*4096*	*12288*
9 - 11	*6*	*10*	*60*	*100*	*600*	*1000*	*6000*	*10000*	*60000*
$\sum$	***20***	-	***122***	-	***948***	-	***8264***	-	***75984***

$$m_1 = \frac{\sum f_i c_i}{\sum f_i} = \bar{x} = \frac{122}{20} = \mathbf{6,1} \qquad m_2 = \frac{\sum f_i c_i^2}{\sum f_i} = \frac{948}{20} = \mathbf{47,4}$$

$$m_3 = \frac{\sum f_i c_i^3}{\sum f_i} = \frac{8264}{20} = \mathbf{413,2} \qquad m_4 = \frac{\sum f_i c_i^4}{\sum f_i} = \frac{75984}{20} = \mathbf{3799,2}$$

Central moments (μ_r)

Is a statistical value of rank r where $r \in N$ is centered around the mean Arithmetic (b = $\bar{x}$) The methods for calculating eccentric moment vary depending on the nature of the data.

- ## Non-repeating discrete data:

$$\mu_r = \frac{\sum(x_i - \bar{x})^r}{n}$$

The first four decentralized moments in this case are as follows:

$$\frac{\sum(x_i - \bar{x})^1}{n} = 0 \quad \mu_2 = \frac{\sum(x_i - \bar{x})^2}{n} = \delta_x^2 \quad \mu_2 = \frac{\sum(x_i - \bar{x})^2}{n} \quad \mu_2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

- ## Recurring discrete data:

$$\mu_r = \frac{\sum f_i(x_i - \bar{x})^r}{\sum f_i} = \sum Fr(x_i - \bar{x})^r$$

The first four decentralized moments in this case are as follows:

$$\mu_2 = \frac{\sum f_i(x_i - \bar{x})^2}{\sum f_i} = \delta_x^2 \qquad \mu_3 = \frac{\sum f_i(x_i - \bar{x})^3}{\sum f_i} \qquad \mu_4 = \frac{\sum f_i(x_i - \bar{x})^4}{\sum f_i}$$

- ## Connected data :

$$\mu_r = \frac{\sum f_i(c_i - \bar{x})^r}{\sum f_i} = \sum Fr(c_i - \bar{x})^r$$

The first four decentralized moments in this case are as follows:

$$\mu_1 = 0 \qquad \mu_2 = \frac{\sum f_i(c_i - \bar{x})^2}{\sum f_i} = \delta_x^2 \qquad \mu_3 = \frac{\sum f_i(c_i - \bar{x})^3}{\sum f_i} \qquad \mu_4 = \frac{\sum f_i(c_i - \bar{x})^4}{\sum f_i}$$

Example

Going back to the previous example, I calculate the first four central moments:

Solution:

We found that $\bar{x}$ = 6,1

$Classes$	f_i	c_i	$c_i - \bar{x}$	$(c_i - \bar{x})^2$	$f_i(c_i - \bar{x})^2$	$(c_i - \bar{x})^3$	$f_i(c_i - \bar{x})^3$	$(c_i - \bar{x})^4$	$f_i(c_i - \bar{x})^4$
1 - 3	5	2	-4,1	16,81	84,05	-68,921	-344,61	282,576	1412,881
3 - 5	4	4	-2,1	4,41	17,64	-9,261	-37,04	19,448	77,792
5 - 7	2	6	-0,1	0,01	0,02	-0,001	-0,002	0,0001	0,0002
7 - 9	3	8	1,9	3,61	10,83	6,859	20,58	13,032	39,096
9 - 11	6	10	3,9	15,21	91,26	59,319	355,91	231,344	1388,065
$\sum$	**20**	-	-	-	203,8	-	-5,16	-	2917,834

$$\mu_1 = \mathbf{0} \qquad \mu_2 = \frac{\sum f_i(c_i - \bar{x})^2}{\sum f_i} = \delta_x^2 = \frac{203{,}8}{20} = \mathbf{10{,}19} \Rightarrow \delta_x = \sqrt{10{,}19} = 3{,}19$$

$$\mu_3 = \frac{\sum f_i(c_i - \bar{x})^3}{\sum f_i} = \frac{-5{,}16}{20} = \mathbf{0{,}258} \qquad \mu_4 = \frac{\sum f_i(c_i - \bar{x})^4}{\sum f_i} = \frac{2917{,}834}{20} = \mathbf{145{,}89}$$

- ## The relationship between central and eccentric moments:

The values of central moments can be calculated based on eccentric moments through the following relationships:

$$\mu_1 = 0$$

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3$$

$$\mu_4 = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4$$

Example

Returning to the data of the previous example, calculate the central moments based on the eccentric moments In the previous example, we obtained the following results:

$$m_1 = 6{,}1 \qquad m_2 = 47{,}4 \qquad m_3 = 413{,}2 \qquad m_4 = 3799{,}2$$

$$\mu_1 = \mathbf{0}$$

$$\mu_2 = m_2 - m_1^2 = 47{,}4 - (6{,}1)^2 = \mathbf{10,19}$$

$$\mu_3 = m_3 - 3m_2m_1 + 2m_1^3 = 413{,}2 - 3(47{,}4)(6{,}1) + 2(6{,}1)^3 = -\mathbf{0,258}$$

$$\begin{aligned}\mu_4 &= m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4 \\ &= 3799{,}2 - 4(413{,}2)(6{,}1) + 6(47{,}4)(6{,}1)^2 - 3(6{,}1)^4 \\ &= 3799{,}2 - 10082{,}08 + 10582{,}524 - 4153{,}7523 = \mathbf{145,89}\end{aligned}$$

It is noticeable that the same results were obtained in the example when we calculated the central moments based on the direct law

Mohamed abulibdah

LinkedIn

If you like the scientific material, share it with others so that the benefit spreads, and if you have questions, I am happy to communicate with you. Mohamed Abu Libdah, a statistician and data analyst at the National Cancer Institute, Cairo University, Egypt