3 STATISTICS

MEASURES OF DISPERSION

3.1 INTRODUCTION

Statistics is a set of concepts, rules, and procedures that help us to

- Organize numerical information in the form of tables, graphs, and charts
- Understand statistical techniques underlying decisions that affect our lives and well-being and make informed decisions.

It is the mathematical science involving the collection, analysis and interpretation of data. In practice, data often contain some randomness or uncertainty. Statistics handles such data using methods of probability theory. A number of specialties have evolved to apply statistical theory and methods to various disciplines such as bio statistics, insurance, finance, econometrics, operations research etc.

Researchers are often interested in defining a value that best describes some attribute of the population. Often this attribute is a measure of central tendency or a proportion. Several different measures of central tendency are defined below.

Measures of central tendency

Measures of central tendency are statistical constants which enable us to comprehend in a single effort, the significance of the whole. Thus a measure of central tendency is a representative of entire distribution. The following are the five measures of central tendency that are in common use.

1. Arithmetic mean
2. Median
3. Mode
4. Geometric Mean
5. Harmonic Mean

Arithmetic Mean

Arithmetic mean of a set n of observations $x_1, x_2, \ldots, x_n$ is given by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

In case of frequency distribution, x_i / f_i where i=1,2,…,n where f_i is the frequency of the variable x_i, Arithmetic mean is given by

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_n x_n}{N} = \frac{\sum f_i x_i}{N}, \text{ where } N = \sum f_i$$

In case of grouped or continuous frequency distribution the arithmetic mean is given by,

Mean $\bar{x} = A + \frac{\Sigma fd}{N} \times c$, Where $d = \frac{x - A}{c}$, A- assumed mean, $N = \Sigma f$

Median

For a given distribution, median is the value of the variable which divides the distribution into two equal parts.

In the case of frequency distribution, x_i / f_i ,the median is obtained by considering the cumulative frequencies.

The steps for calculating median are

(i) Find N=$\sum f_i$ =Total number of frequencies

(ii) See the cumulative frequency just greater than $\frac{N}{2}$

(iii) The corresponding value of x is called Median.

In case of continuous frequency distribution, the class corresponding to the cumulative frequency just greater than $\frac{N}{2}$ is called Median class and the value of median is obtained by the following formula.

Median = $l + \dfrac{\frac{N}{2} - m}{f} \times c$

Where l – Lower limit of the median class

c – Width of the median class

f – Frequency of the median class

m – Cumulative frequency of the class preceding the median class.

$N = \Sigma f$ = Total frequency

Mode

Mode is the value which occurs most frequently in a set of observations and around which the other items of the set cluster closely.

In the case of discrete frequency distribution, the mode is the value of x corresponding to the maximum frequency and it is given by,

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

Where l – Lower limit of the modal class

f_1 – Frequency of the modal class

f_0 – Frequency of the class preceding the modal class

f_2 – Frequency of the class succeeding the modal class

c – Length of the modal class

Note

(i) The class corresponding to the maximum frequency is called modal class.

(ii) If the maximum frequency is repeated the mode is ill defined.

(iii) Also if the maximum frequency occurs in the beginning and at the last then the mode is ill defined.

3.2 MEASURES OF DISPERSION

The measures of central tendency viz, mean, median and mode give us an idea of the concentration of the observations about the central part of the distribution. Hence, if we want to know the complete idea of a distribution, we need some other measures. One such measure is Dispersion.

The following are some important measures of dispersion

1. Quartile deviation
2. Mean deviation
3. Standard deviation

Standard deviation

Standard deviation denoted by σ is defined as the positive square root of the mean of the square of the deviation from the mean $\bar{x}$.

$$\sigma = \sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2} \times c$$

3.3 SKEWNESS

The term skewness refers to lack of symmetry or departure from symmetry. When a distribution is not symmetrical, it is called a skewed distribution. The objective of studying skewness is to have an idea about the shape of the curve which we can draw with the help of the given data.

A distribution is said to be skewed if

(i) Mean, Median and Mode are not equal.

(i.e.) Mean $\neq$ Median $\neq$ Mode

(ii) The distribution curve drawn with the help of given data is not symmetrical but stretched more to one side than to the other.

3.3.1. Measures of Skeweness

Measures of skewness gives us an idea about the extent of symmetry in a series and various measures of skewness are

(i) S_k= Mean – Median

(ii) S_k= Mean – Mode

These measures are the absolute measures of skewness. For comparing two series, we calculate the relative measures called the coefficient of skewness which are pure numbers.

3.3.2. Pearson's coefficient of skewness

The Pearson's coefficient of skewness is defined by $S_k = \dfrac{\text{Mean} - \text{Mode}}{\text{standard Deviation}}$.

When the mode is not well defined, the Pearson's coefficient of skewness is defined by $S_k = \dfrac{3(\text{Mean - Median})}{\text{Standard Deviation}}$

Example 1

Calculate the coefficient of skewness for the following data ***(November- 2001)***

x	0–10	10–20	20–30	30–40	40–50	50–60	60–70
y	15	18	25	22	17	14	10

Solution

Class interval	Mid value X	Frequency f	$d = \frac{x-A}{c}$	cf	d^2	fd	fd^2
0 – 10	5	15	−3	15	9	−45	135
10 – 20	15	18	−2	33	4	−36	72
20 – 30	25	25	−1	58	1	−25	25
30 – 40	35	22	0	80	0	0	0
40 – 50	45	17	1	97	1	17	17
50 – 60	55	14	2	111	4	28	56
60 – 70	65	10	3	121	9	30	90
		$\Sigma f = 121$				$\Sigma fd = -31$	$\Sigma fd^2 = 395$

Width of the class interval $C = 10$, $N = \Sigma f = 121$

Arithmetic Mean $\bar{x} = A + \frac{\Sigma fd}{N} \times c = 35 + \frac{(-31)}{121} \times 10 = 32.438$

$\bar{x} = 32.44$

$\frac{N}{2} = \frac{121}{2} = 60.5$

$\therefore$ 30 – 40 is the median class (Just above the value of $\frac{N}{2}$ in cf is 80).

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c = 30 + \frac{60.5 - 58}{22} \times 10 = 30 + 1.136$$

$= 31.136$

Here maximum frequency is 25, therefore the modal class is 20-30.

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c = 20 + \frac{25 - 18}{2(25) - 18 - 22} \times 10$$

$$= 20 + \left(\frac{7}{10} \times 10\right) = 27$$

$$\text{Standard Deviation S.D} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times c$$

$$= \sqrt{\frac{395}{121} - \left(\frac{-31}{121}\right)^2} \times c = \sqrt{3.264 - 0.0656} \times 10$$

$$= \sqrt{3.1984} \times 10 = 1.788 \times 10$$

$$S.D = 17.88$$

Pearson's Coefficient of skewness,

$$S_k = \frac{\text{Mean-Mode}}{\text{S.D}} = \frac{31.136-27}{17.88} = 0.231$$

Example 2

Calculate the Pearson's coefficient of skewness for the following data

Class interval	***130–134***	***135–139***	***140–144***	***145–149***	***150–154***	***155–159***	***160–164***
Frequency	***3***	***12***	***21***	***28***	***19***	***12***	***5***

Solution

Class Interval	Midvalue	f	$d = \frac{x-A}{c}$	d^2	cf	fd	fd^2
129.5 – 134.5	132	3	−3	9	3	−9	27
134.5 – 139.5	137	12	−2	4	15	−24	48
139.5 – 144.5	142	21	−1	1	36	−21	21
144.5 – 149.5	147	28	0	0	64	0	0
149.5 – 154.5	152	19	1	1	83	19	19
154.5 – 159.5	157	12	2	4	95	24	48
159.5 – 164.5	162	5	3	9	100	15	45
		$\Sigma f = 100$				$\Sigma fd = 4$	$\Sigma fd^2 = 208$

Arithmetic Mean $\bar{x} = A + \frac{\Sigma fd}{N} \times c = 147 + \frac{4}{100} \times 5$

$$\bar{x} = 147.2$$

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c = 144.5 + \left(\frac{50-36}{28} \times 5\right)$$

$$= 147$$

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c = 144.5 + \left(\frac{28-21}{2(28)-21-19} \times 5\right)$$

$$= 146.69$$

Standard Deviation $= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times c = \sqrt{\frac{208}{100} - \left(\frac{4}{100}\right)^2} \times 5$

S.D. $= 7.208$

Karl Pearson's coefficients of skewness is,

$$S_k = \frac{\text{Mean- Mode}}{\text{S.D}} = \frac{147.2 - 146.69}{7.208} = 0.070$$

Exercise

1. Calculate the Karl Pearson's coefficient of skewness for the following data

Class interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Frequency	10	40	20	0	10	40	16

(Ans.S_k = -0.753)

2. Calculate the Karl Pearson's coefficient of skewness for the following data

Marks above	0	10	20	30	40	50	60	70	80
Frequency	150	140	100	80	80	70	30	14	0

(Ans. -0.754)

3. Calculate the Karl Pearson's coefficient of skewness for the following data

Class interval	0 – 9	10 – 19	20 – 29	30 – 39	40 – 49
Frequency	10	40	20	0	10

(Ans. S_k = 0.006)

4. Calculate the Karl Pearson's coefficient of skewness for the following data

Class interval	3 – 7	8 – 12	13 – 17	18 – 22	23 – 27	28 – 32	33 – 37	38 – 42
Frequency	2	108	580	175	80	32	18	5

(Ans. S_k = 0.356)

5. Calculate Karl Pearson's coefficient of skewness for the following data ***(May-2011)***

Size	1	2	3	4	5	6	7
Frequency	10	18	30	25	12	3	2

3.4. MOMENTS

Let the symbol 'x' be used to represent the deviation of any item in a distribution from the arithmetic mean of that distribution. The arithmetic mean of the various powers of these deviations in any distribution is called the moment of the distribution.

If we take the mean of the first power of the deviation, we get the first moment about the mean. The mean of the squares of the deviations gives us the second moment about the mean. The mean of the cubes of the deviations gives us the third moment about the mean, and so on. The moments about the mean are called the 'central moment' and denoted by Greek letter μ,thus μ_1 stands for first moment about mean, μ_2 stands for second moment about mean, etc. Symbolically,

$$\mu_1 = \frac{\sum d}{N}, \mu_2 = \frac{\sum d^2}{N}, \ \mu_3 = \frac{\sum d^3}{N}, \ \mu_4 = \frac{\sum d^4}{N}, \text{ where d=}\left(X - \overline{X}\right)$$

For frequency distribution,

$$\mu_1 = \frac{\sum fd}{N}, \ \mu_2 = \frac{\sum fd^2}{N}, \ \mu_3 = \frac{\sum fd^3}{N}, \ \mu_4 = \frac{\sum fd^4}{N}, \text{ where d=}\left(X - \overline{X}\right)$$

Moments can be extended to higher powers in a similar way. Two important constants of a distribution are calculated from μ_2, μ_3, μ_4 namely,

Measure of Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$ and Measure of kurtosis $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

3.4.1. Moments about the arbitrary origin

It is difficult to calculate moments by applying the above formulae where the actual mean is a fraction. In such case, we can first compute moments about the actual mean. Moments about arbitrary origin are called 'raw moments', and denoted by the symbol μ'. Thus μ_1' would stand for the first moment about an arbitrary A. μ_2' would stand for the second moment about an arbitrary A and so on.
The calculation of raw moments shall be done as follows.

$$\mu_1' = \frac{\sum d}{N}, \mu_2' = \frac{\sum d^2}{N}, \mu_3' = \frac{\sum d^3}{N}, \ \mu_4' = \frac{\sum d^4}{N}, \text{ where } d = (X - A)$$

For frequency distribution,

$$\mu_1' = \frac{\sum fd}{N}, \ \mu_2' = \frac{\sum fd^2}{N}, \ \mu_3' = \frac{\sum fd^3}{N}, \ \mu_4' = \frac{\sum fd^4}{N}, \text{ where } d = (X - A)$$

Relation between the first four moments about an arbitrary origin and moments about mean or central moments.

$$\mu_1 = 0$$

$$\mu_2 = \mu_2^1 - \mu_1^{1^2}$$

$$\mu_3 = \mu_3^1 - 3\mu_1^1\mu_2^1 + 2\mu_1^{1^3}$$

$$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1^1 + 6\mu_2^1\mu_1^{1^2} - 3\mu_1^{1^4}$$

Also we have the following relations between the raw and central moments.

$$\mu_1^1 = \bar{x} - A$$

$$\mu_2^1 = \mu_2 + \mu_1^{1^2}$$

$$\mu_3^1 = \mu_3 + 3\mu_2\mu_1^1 + \mu_1^3$$

$$\mu_4^1 = \mu_4 + 4\mu_3\mu_1^1 + 6\mu_2\mu_1^{1^2} + \mu_1^{1^4}$$

3.4.2. Moments Based Measures of skewness and Kurtosis

Karl Pearson's defined the following four coefficients based upon the first four moments about mean.

Measure of Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$ or $\gamma_1 = \sqrt{\beta_1}$

Measure of kurtosis $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$ or $\gamma_2 = \beta_2 - 3$

Note 1

If $\beta_2 = 3$, the curve is mesokurtic

If $\beta_2 < 3$, the curve is platykurtic

If $\beta_2 < 3$, the curve is leptokurtic

Note 2

If skewness $\beta_1 = 0$, then there is no skewness. Theoretically its value lies between ± 3. However, in practical it rarely exceeds ± 1.

Example 1

Calculate the first three central moments for the following frequency distribution.

Marks less than	80	70	60	50	40	30	20	10
Frequency	100	90	80	60	32	20	13	5

Solution

Moments about A=35

$$\mu_1' = \frac{\Sigma fd}{N}\times c = \frac{100}{100}\times 10 = 10$$

$$\mu_2' = \frac{\Sigma fd^2}{N}\times c^2 = \frac{442}{100}\times 100 = 442$$

$$\mu_3' = \frac{\Sigma fd^3}{N}\times c^3 = \frac{892}{100}\times 1000 = 8920$$

Class Interval	Mid value	f	$d = \frac{x-A}{c}$	fd	fd^2	fd^3
0 – 10	5	5	−3	−15	45	−135
10 – 20	15	8	−-2	−16	32	−64
20 – 30	25	7	−-1	−7	7	−7
30 – 40	35	12	0	0	0	0
40 – 50	45	28	1	28	28	28
50 – 60	55	20	2	40	80	100
60 – 70	65	10	3	30	90	270
70 – 80	75	10	4	40	160	640
		$\sum f = 100$		$\sum fd = 100$	$\sum fd^2 = 442$	$\sum fd^3 = 892$

Central moments about mean are defined by,

$$\mu_2 = \mu_2^1 - (\mu_1^1)^2 = (442) - (10)^2 = 332$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2(\mu_1^1)^3 = 8920 - 3(442)(10) + 2(10)^3 \quad = -2340.$$

Example 2

Find the first four moments about the mean of the distribution and also find β_1 and β_2, γ_1 and γ_2. ***(May-2011)***

X	2.0	2.5	3.0	3.5	4.0	4.5	5.0
f	5	38	65	92	70	40	10

Solution

Raw moments about A=3.5 are

$$\mu_1^1 = \frac{\Sigma fd}{N} = \frac{12}{320} = 0.0375 \quad (d = X - A)$$

$$\mu_2^1 = \frac{\Sigma fd^2}{N} = \frac{145.5}{320} = 0.4546$$

$$\mu_3^1 = \frac{\Sigma fd^3}{N} = \frac{19.5}{320} = 0.0609$$

$$\mu_4^1 = \frac{\Sigma fd^4}{N} = \frac{162.375}{320} = 0.5074$$

x	f	$d = x-A$ (A=3.5)	fd	fd^2	fd^3	fd^4
2.0	5	– 1.5	– 7.5	11.25	– 16.875	25.31
2.5	38	– 1	– 38	38	– 38	38
3.0	65	–.5	– 32.5	16.25	– 8.125	4.0625
3.5	92	0	0	0	0	0
4.0	70	.5	35	17.5	8.75	4.375
4.5	40	1	40	40	40	40
5.0	10	1.5	15	22.5	33.75	50.625
	Σf=320		$\sum fd = 12$	$\sum fd^2 = 145.5$	$\sum fd^3 = 19.5$	$\sum fd^4 = 162.375$

Central moments (moments about mean) are

$$\mu_1 = 0 \text{ (always)}$$

$$\mu_2 = \mu_2^1 - \left(\mu_1^1\right)^2 = (0.4547) - (0.0375)^2 = 0.453$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2(\mu_1^1)^3$$

$$= 0.0609 - 3(0.4546)(0.0375) + 2(0.0375)^3 = 0.00985$$

$$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1^1 + 6\mu_2^1(\mu_1^1)^2 - 3(\mu_1^1)^4$$

$$= 0.5074 - 4(0.0609)(0.0375) + 6(0.4546)(0.0375)^2 - 3(0.0375)^4 = 0.502$$

Measure of Skewness

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0.000097}{0.0931} = 0.00104$$

(or) $\gamma_1 = \sqrt{\beta_1} = \sqrt{0.00104} = 0.0322$.

Example 3

Calculate β_1 and β_2 (measure of skewness and Kurtosis) of the following distribution and hence comment on the type of the frequency distribution.

X	2	3	4	5	6
Y	1	3	7	2	1

Solution

X	*f*	*d=x–A* *(A=4)*	*fd*	*fd²*	*fd³*	*fd⁴*
2	1	– 2	– 2	4	– 8	16
3	3	– 1	– 3	3	– 3	3
4	7	0	0	0	0	0
5	2	1	2	2	2	2
6	1	2	2	4	8	16
	14		$\Sigma fd = -1$	$\Sigma fd^2 = 13$	$\Sigma fd^3 = -1$	$\Sigma fd^4 = 37$

Raw Moments about $x = 4$

$$\mu_1^1 = \frac{\Sigma fd}{N} = \frac{-1}{14} = -0.07$$

$$\mu_2^1 = \frac{\Sigma fd^2}{N} = \frac{13}{14} = 0.93$$

$$\mu_3^1 = \frac{\Sigma fd^3}{N} = -\frac{1}{14} = -0.07$$

$$\mu_4^1 = \frac{\Sigma fd^4}{N} = \frac{37}{14} = 2.64$$

Central Moments

$$\mu_2 = \mu_2^1 - \mu_1^1 = 0.93 - (-0.07)^2 = 0.9251$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2(\mu_1^1)^3 = -0.07 + 3(0.07)(0.93) + 2(-0.07)^3 = 0.1246$$

$$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1 + 6\mu_2^1(\mu_1^1)^2 - 3(\mu_1^1)^4$$

$$= (2.64) - 4(-0.07)(-0.07) + 6(0.93)(-0.07)^2 - 3(-0.07)^4 = 2.647$$

Measure of Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(0.1246)^2}{(0.9251)^3} = 0.0196$

Measure of Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{2.647}{(0.9251)^2} = 2.443$,

since $\beta_2 < 3$, the curve is platykurtic.

Example 4

Calculate the coefficient of skewness for the following data. ***(November- 2001)***

X	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
Y	15	18	25	22	17	14	10

Solution

Class interval	Mid value	Frequency f	$d = \frac{x-A}{c}$ $A=35$	fd	fd^2	fd^3	fd^4
0–10	5	15	–3	–45	135	–405	1215
10–20	15	18	–2	–36	72	–144	288
20–30	25	25	–1	–25	25	–25	25
30–40	35	22	0	0	0	0	0
40–50	45	17	1	17	17	17	17
50–60	55	14	2	28	56	112	224
60–70	65	10	3	30	90	270	810
		$\Sigma f = 121$		$\sum fd =$ –31	$\sum fd^2 = 395$	$\sum fd^3 =$ –175	$\sum fd^4$ =2579

Moments about A=35 are defined by,

$$\mu_1^1 = \frac{\Sigma fd}{\Sigma f} \times c = \frac{-31}{121} \times 10 = -2.56$$

$$\mu_2^1 = \frac{\Sigma fd^2}{\Sigma f} \times c^2 = \frac{395}{121} \times 100 = 326.44$$

$$\mu_3^1 = \frac{\Sigma fd^3}{\Sigma f} \times c^3 = \frac{-175}{121} \times 1000 = -1446.28$$

$$\mu_4^1 = \frac{\Sigma fd^4}{\Sigma f} \times c^4 = \frac{2579}{121} \times 10000 = 213140.4959$$

Central moments

$$\mu_1 = 0$$

$$\mu_2 = (\mu_2^1) - (\mu_1^1)^2 = 326.44$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2(\mu_1^1)^3 = -1446.28 - 3(32664)(-2.56) + 2(-2.56)^3$$

$$= -1446.28 + 2506.75 - 33.554 = 1026.918$$

$$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1^1 + 6\mu_2^1(\mu_1^1)^2 - 3(\mu_1^1)^4$$

$$= (213140.49) - 4(-1446.28)(-2.56) + 6(326.44)(-2.56)^2 - 3(-2.56)^4$$

$$= 211037.8$$

Measure of Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(1026.918)^2}{(326.44)^3} = 0.0303$

Measure of Kurtosis $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{213140.49}{(326.44)^2} = 2.0$

Since $\beta_2 < 3$, the curve is platykurtic.

Example 5

Find the coefficient of skewness and coefficient of kurtosis from the distribution of marks of 100 students.

Marks	20–25	25– 30	30– 35	35– 40	40– 45	45– 50	50– 55	55– 60	60– 65
No.of students	6	8	11	14	21	15	11	9	5

Solution

Moments about $x = 42.5$

$$\mu_1^1 = \frac{\Sigma fd}{N} \times c = 0$$

$$\mu_2^1 = \frac{\Sigma fd^2}{N} \times c^2 = (5)^2 \frac{446}{100} = 111.5$$

$$\mu_3^1 = \frac{\Sigma fd^3}{N} \times c^3 = 125\left(\frac{-36}{100}\right) = -45$$

$$\mu_4^1 = \frac{\Sigma fd^4}{N} \times c^4 = 625\left(\frac{4574}{100}\right) = 28587.5$$

Class intervals	Mid value x	Frequency f	$d = X - 42.5$	fd	fd^2	fd^3	fd^4
20-25	22.5	6	–4	–24	96	–384	1536
25-30	27.5	8	–3	–24	72	–216	648
30-35	32.5	11	–2	–22	44	–88	176
35-40	37.5	14	–1	–14	14	–14	14
40-45	42.5	21	0	0	0	0	0
45-50	47.5	15	1	15	15	15	15
50-55	52.5	11	2	22	44	88	176
55-60	57.5	9	3	27	81	243	729
60-65	62.5	5	4	20	80	320	1280
		N=100		$\sum fd = 0$	$\sum fd^2 = 446$	$\sum fd^3 = -36$	$\sum fd^4 = 4574$

Central moments

$$\mu_1 = 0$$

$$\mu_2 = (\mu_2^1) - (\mu_1^1)^2 = 111.5$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2(\mu_1^1)^3 = -45 - 3(111.5)(0) + 2(0)^3 = -45$$

$$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1 + 6\mu_2^1(\mu_1^1)^2 - 3(\mu_1^1)^4$$

$$= (28587.5) - 4(-45)(0) + 6(111.5)(0)^2 - 3(0)^4$$

$$= 28587.5$$

Measure of Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(-45)^2}{(111.5)^3} = 0.00146$

Measure of Kurtosis $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{28587.5}{(111.5)^2} = 2.299$

Example 6

The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50. Obtain μ_2, μ_3, μ_4.

Solution

Given that the moments about the value A = 5 are $\mu_1^1 = 2$, $\mu_2^1 = 20$, $\mu_3^1 = 40$ and $\mu_4^1 = 50$

Arithmetic Mean $\bar{x} = A + \mu_1^1 = 5 + 2 = 7$

$$\mu_1 = 0$$

$\mu_2 = \mu_2^1 - \mu_1^{1^2}$ =20- $(2)^2$ =16

$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2\mu_1^{1^3}$

= 40-3(20)(2)+2$(2)^3$ =40-120+16 = −64

$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1^1 + 6\mu_2^1\mu_1^{1^2} - 3\mu_1^{1^4}$

=50−4(40)(2)+6(20)$(2)^2$−3$(2)^4$ =162.

Example 7

The first three moments of a distribution about the value 2 of the variable are 1, 16, – 40. Show that the mean is 3 and variance is 15. Also find the first three moments about x = 0

Solution

Given that the moments about the value A = 2 are $\mu_1^1 = 1$, $\mu_2^1 = 16$ and $\mu_3^1 = -40$

Mean $\bar{x}$ = $\mu_1^1 + A$ = 1 + 2 =3(Where A = 2)

$\therefore$ Mean $\bar{x}$ = 3

Variance μ_2 = $\mu_2^1 - (\mu_1^1)^2$ = 16 – = 15

$\therefore$ Variance μ_2 = 15

$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1 + 2(\mu_1^1)^3$

= – 40 – 3 × 16 (1) + 2 $(1)^3$ = – 86

To find the first three moments about x = 0

Let μ_r' denote the r[th] moment about x = 0

Here A = 0

$\mu_1^1 = \bar{x} - A$ = 3 – 0 = 3

$\mu_2^1 = \mu_2 + \mu_1^{1^2}$ = 15 + 3^2= 24

$\mu_3^1 = \mu_3 + 3\mu_2\mu_1^1 + (\mu_1^1)^3$ = -86 + 135 +27= 76.

Example 8

The first three moments of a distribution about the value 3 of the variable are 2, 10, –30. Obtain the corresponding moments about x = 0. ***(May- 2001)***

Solution

Given that A = 3

$\mu_1^1 = 2, \mu_2^1 = 10, \ \mu_3^1 = -30$

Moments about $x = 0$

Let μ_r^1 denote the r[th] moment about $x = 0$

Let $A = 0$

$$\mu_1^1 = \bar{x} - A = 5 - 0 = 5$$

$$\mu_2^1 = \mu_2 + \mu_1^{1^2} = 6 + 2^2 = 10$$

$$\mu_3^1 = \mu_3 + 3\mu_2\mu_1^1 + (\mu_1^1)^3 = -74 + 3(6)(5) + (5)^3 = -74 + 90 + 125$$

$$= 141$$

$$\text{Mean } \bar{x} = \mu_1^1 + A = 2 + 3 = 5 \ (A = 3)$$

$$\bar{x} = 5$$

$$\text{Variance } \mu_2 = \mu_2^1 - (\mu_1^1)^2 = 10 - 2^2 = 10 - 4 = 6$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1^1 + 2\mu_1^{1^3} = -30 - 3(10)(2) + 2(2)^3 = -74.$$

Example 9

The first four moments of a distribution about the value 4 of the variable are –1.5, 17, 30 and 108. Find the moments about mean, β_1 and β_2. Find the moments about the point x = 2 (April- 2003, May -2007)

Solution

Let μ_r^1 denote the r[th] moment about $A = 4$ are $\mu_1^1 = -1.5$, $\mu_2^1 = 17$, $\mu_3^1 = -30$, $\mu_4^1 = 108$

$$\text{Mean } \bar{x} = \mu_1^1 + A$$

$$= 4 - 1.5 = 2.5$$

Mean $\bar{x} = 2.5$

Moments about Mean.

$$\mu_1 = 0 \text{ (always)}$$

$$\mu_2 = \mu_2^1 - \left(\mu_1^1\right)^2 = 17 - (-1.5)^2 = 14.75$$

$$\mu_3 = \mu_3^1 - 3\mu_2^1\mu_1 + 2(\mu_1^1)^3 = -30 - 3(17)(-1.5) + 2(-1.5)^3 = 39.75$$

$$\mu_4 = \mu_4^1 - 4\mu_3^1\mu_1 + 6\mu_2^1(\mu_1^1)^2 - 3(\mu_1^1)^4$$

$$= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4 = 142.3125$$

Also Measure of Skewness $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(39.75)^2}{(14.75)^3} = 0.4926$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(4.75)^2} = 0.6543$$

Moments about the point $x = 2$

Let $A = 2$

Let μ_r^1 denote the r[th] moment about $x = 2$

$$\mu_1^1 = \bar{x} - A = 2.5 - 2$$

$$\mu_1^1 = 0.5$$

$$\mu_2^1 = \mu_2 + (\mu_1^1)^2 = 14.75 + (0.5)^2 = 15$$

$$\mu_3^1 = \mu_3 + 3\mu_2\mu_1^1 + (\mu_1^1)^3$$

$$= 39.75 + 3(14.75)(0.5) + (0.5)^3 = 62$$

$$\mu_4^1 = \mu_4 + 4\mu_3\mu_1^1 + 6\mu_2(\mu_1^1)^2 + (\mu_1^1)^4$$

$$= 142.3125 + 4(39.75)(0.5) + 6(14.75)(10.5)^2 + (0.5)^4 = 244.$$

Exercise

1. Find the first four moments about the mean of the distribution and also find β_1 and β_2 and hence comment on the nature of the distribution.

x	0	1	2	3	4	5	6	7	8
f	1	8	28	56	70	56	28	8	1

(Ans. $\mu_2 = 2$, $\mu_3 = 0$, $\mu_4 = 11$ and $\beta_1 = 0$, $\beta_2 = 2.75$)

2. Find the first four moments about the mean of the distribution and also find β_1 and β_2 And hence comment on the nature of the distribution.

x	0	1	2	3	4	5	6	7	8
f	5	10	15	20	25	20	15	10	5

(Ans. $\mu_2 = 4$, $\mu_3 = 0$, $\mu_4 = 37.6$ and $\beta_1 = 0$, $\beta_2 = 2.35$, the curve is platykurtic)

3. From the following data of the wages of 50 workers of a factory compute the first four moments about the mean of the distribution and also find β_1 and β_2 and hence comment on the results.

Daily wages	100–110	110–120	120–130	130–140	140–150	150–160	160–170
No. of Workers	1	3	7	20	12	4	3

(Ans. $\mu_2 = 159.24$, $\mu_3 = 80.35$, $\mu_4 = 83659.87$ and $\beta_1 = 0.0016$, $\beta_2 = 3.3$)

4. The first four moments of a distribution about $x = 5$ are 2, 20, 40 and 50. Show that the mean $= 16$, variance $= 16$, $\mu_3 = -64$, $\mu_4 = 162$ and $\beta_1 = 1$, $\beta_2 = 0.63$

5. The first four moments of a distribution about $x = 4$ are respectively 1, 4, 10, 45. Calculate the moments about the mean.

(Ans. Mean=54, $\mu_2 = 2.97$, $\mu_3 = -2.18$, $\mu_4 = 218.42$ and $\beta_1 = 0.007$, $\beta_2 = 2.81$)

PART-A

Questions with solutions

1. Define measures of central tendency. How many measures of central tendency are in use state?

Solution

Averages are called measures of central tendency. They are 3forms of average in common use. They are arithmetic mean, the median, and the mode.

2. Define Arithmetic Mean.

Solution

(i) Arithmetic mean of a set of observation is their sum divided by the number of observations.

$$\overline{x} = \frac{\sum x_i}{n}$$

(ii) In case of frequency distribution, the arithmetic mean is given by

Mean $\overline{x} = A + \frac{\Sigma fd}{N} \times c$, Where $d = \frac{x - A}{c}$, $N = \Sigma f$, A- Assumed mean, c- width of the interval

3. Define Median.

Solution

In the case of frequency distribution, the class corresponding to the cumulative frequency just greater than $\frac{N}{2}$ is called Median class.

The value of the median is obtained by the formula,

$$\text{Median} = l + \frac{\frac{N}{2} - m}{f} \times c$$

Where l – Lower limit of the median class

c – Width of the median class

f – Frequency of the median class

m – Cumulative frequency of the class preceding the median class.

$N = \Sigma f$ = total frequency

4. Define Mode.

Solution

In the case of discrete frequency distribution the mode is the value of x corresponding to the maximum frequency. For continuous frequency distribution mode is given by,

$$\text{Mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times c$$

Where l – lower limit of the modal class

f_1 – Frequency of the modal class

f_0 – Frequency of the class preceding the modal class

f_2 – Frequency of the class succeeding the modal class

c – Length of the modal class

5. Define Standard deviation.

Solution

Standard deviation denoted by σ is defined as the positive square root of the mean of the square of the deviation from the mean $\bar{x}$.

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times c$$

6. *Write short notes on "skewness".*

Solution

A distribution is said to be skewed or lacking in symmetry when the variables tend to be dispersed more on one side of the central value than the other.

The frequency curve is said to be positively skewed if more items are found to the right of the ordinate. In this case, the curve will have a tail to the right. In positively skewed distribution, the mode, median and arithmetic mean are in ascending order of magnitude.

In the negatively skewed distribution, the mode, median and arithmetic mean are in descending order of magnitude and the curve will have a longer tail to the left.

7. *Write short notes on" kurtosis".* ***(May-2011)***

Solution

The word Kurtosis is derived from the Greek word meaning "humped". The peakedness or flatness of the frequency distribution is known as kurtosis.

If μ_2, μ_4 are the second and fourth moments about mean, then the coefficient of kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2}$ is the measures of kurtosis. If β_2=3 the curve is mesokurtic. If β_2<3, it is platykurtic and if β_2>3, it is leptokurtic.

8. *Define Moment.*

Solution

Let the symbol x be used to represent the deviation of any item in a distribution from the arithmetic mean of that distribution. The arithmetic mean of the various powers of these deviations in any distribution is called the moment of the distribution.

9. *Write down the relation between the first four central moments and moments about the arbitrary origins.* ***(May-2011)***

Solution

$\mu_1 = 0$

$\mu_2 = \mu_2' - \mu_1'^2$

$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$

$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

10. The first three moments of a distribution about the value 2 of the variable are 1, 16, - 40. Show that the mean is 3 and variance is 15.

Solution

Given that the moments about the value A = 2 are $\mu_1^1 = 1$, $\mu_2^1 = 16$ and $\mu_3^1 = -40$

Mean $\bar{x} = \mu_1^1 + A$

$\bar{x}$ = 1 + 2 (Where A = 2) = 3

Variance $\mu_2 = \mu_2^1 - (\mu_1^1)^2$ = 16 – 1= 15

11. In a distribution if μ_4 = 200, μ_2 = 8.5. What type of distribution is this?

Solution

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{200}{(8.5)^2} = 2.768<3$$

$\therefore$ The distribution is called platykurtic.

12. Find β_2, if μ_2 =28.48, μ_3 =229.3, μ_4 =4780.31 where β_2 is coefficient of skewness.

Solution

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{4780.31}{(28.48)^2} = 5.89$$