1/

Module 4: Statistical methods

Introduction:

A statistical data may sometimes be described as distributed around some value called central value $\boxed{OR}$ average. It gives a most representative value of the entire data. Different methods give different values which are known as Measures of central tendency. The commonly used measures of central value are:

(i) Mean (ii) Median (iii) Mode.

As they do not exhibit any idea as to how the individual values differ from the central value i.e., whether they are closely packed around the central value $\boxed{OR}$ widely scattered away from it. To know the extent of variation $\boxed{OR}$ deviation of the values in comparision with the central values, we use another ...

The Important measures of dispersion are
(i) Range (ii) Quartile Deviation (iii) Mean deviation (iv) Standard deviation

(i) Range: Range is the difference between the greatest and the least values in distribution.

(ii) Quartile deviation: Quartile deviation is the average difference between the first and third Quartile

ie $Q.D = \frac{Q_3 - Q_1}{2}$.

(Note: Quartiles divide total frequency into 4 equal parts).

(iii) Mean Deviation: Mean deviation is the mean of the absolute difference of the values from the mean $\boxed{OR}$ median $\boxed{OR}$ mode

ie. Mean Deviation $= \frac{1}{n}\sum f_i |x_i - A|$ where 'A' is either mean $\boxed{OR}$ median $\boxed{OR}$ mode.

Standard deviation: Square root of the mean of squares of the difference of the variate value from their mean.

ie $\sigma = \sqrt{\dfrac{\sum f_i \cdot (x_i - \bar{x})^2}{N}}$ where N is total frequency $N = \sum f_i$

Moments:

Moments are used to describe the various characteristics of frequency distribution ie., central tendency, dispersion, skewness and Kurtosis. They are helpful in understanding how given distribution looks like. We can obtain an understanding of Spread, asymmetry, shape of any given set.

Moments about mean:

The r^{th} moment of X about the mean $\bar{x}$, is usually denoted by μ_r which is defined as

$$\mu_r = \frac{1}{N}\sum f_i (x_i - \bar{x})^r \quad \text{where } r = 0, 1, 2, 3 \ldots$$

Note: $\boxed{\mu_0 = 1}$; $\boxed{\mu_1 = 0}$; $\boxed{\mu_2 = \sigma^2 = \text{Variance}}$

$$\mu_3 = \frac{1}{N}\sum f_i (x_i - \bar{x})^3 \text{ and so on.}$$

$$\left\{ \mu_0 = \frac{1}{N}\sum f_i (x_i - \bar{x})^0 = \frac{1}{N}\sum f_i = \frac{N}{N} = 1 \right.$$

$$\mu_1 = \frac{1}{N}\sum f_i (x_i - \bar{x}) = \frac{1}{N}\left[\sum f_i x_i - \sum f_i \bar{x}\right]$$

$$= \frac{\sum f_i x_i}{N} - \frac{\sum f_i \bar{x}}{N} = \bar{x} - \bar{x}\frac{\sum f_i}{N}$$

$$\left. = \bar{x} - \bar{x} = 0. \right\}$$

Moments about arbitrary point say 'A':

The r^{th} moment about any point say 'A' is denoted by μ_r' and it is defined by

$$\mu_r' = \frac{1}{N}\sum f_i (x_i - A)^r \quad r = 0, 1, 2, 3 \ldots$$

Note $\mu_0' = 1$; $\mu_1' = \bar{x} - A$.

$$\mu_2' = \frac{1}{N}\sum f_i \cdot (x_i - A)^2 \text{ and so on.}$$

Note: μ_r, the r^{th} moment about mean, $r = 1, 2, 3 \ldots$ are also called central moments and μ_r', the r^{th} moments about any point 'A' are also called as Raw moments.

Relationship between μ_r and μ_r'

$\mu_1 = 0$ (always)

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

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

$$\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2[\mu'_1]^2 - 3[\mu'_1]^4$$

in general,

$$\mu_r = \mu'_r - {}^rC_1\,\mu'_{r-1}\,\mu'_1 + {}^rC_2\,\mu'_{r-2}\,\mu'^2_1 - {}^rC_3\,\mu'_{r-3}\,\mu'^3_1 + \cdots + (-1)^r[\mu'_1]^r.$$

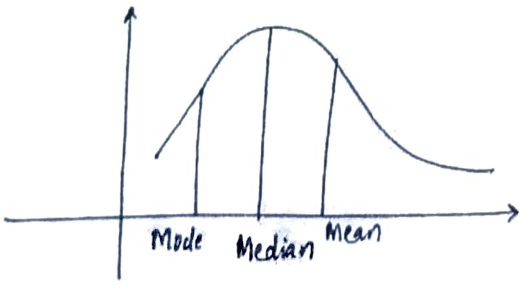
Note: (1) If we know the first four moments about any arbitrary point say 'A', we can obtain the measures of Central tendencies.

(2) $\bar{x} = \mu'_1$ (about origin ie $A=0$)

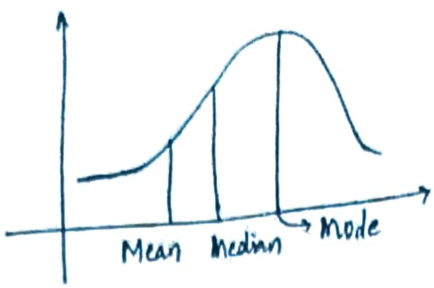
(3) $\mu_2 = \sigma^2$.

Skewness:

It measures the degree of asymmetry of the frequency distribution. If the frequency has longer tail to the right then mean is to the right of the mode, then the distribution is said to have positive skewness.



If the curve is more elongated to the left, then it is said to have Negative skewness.



The following are the measure of skewness:

1. Pearson's coefficient of skewness $= \frac{\text{Mean} - \text{Mode}}{\sigma}$

2. Quartile coefficient of skewness $= \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

{Note: its value lies between -1 & 1}.

3. Coefficient of skewness based on third moment is given by
$\gamma_1 = \sqrt{\beta_1}$ where $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$.

Note: All these measures of skewness will be +ve OR −ve according as the distribution is skewed to left OR Right. They are zero for symmetric distribution.

Measure of skewness give the direction and magnitude of ~~the~~ lack of symmetry.

Kurtosis give the idea of flatness.

Kurtosis:

Kurtosis is measure the degree of convexity $\boxed{OR}$ peakedness of the frequency distribution and it is denoted by

$$\boxed{\beta_2 = \frac{\mu_4}{\mu_2^2}}$$

The coefficient of kurtosis is given by $\boxed{\gamma_2 = \beta_2 - 3}$

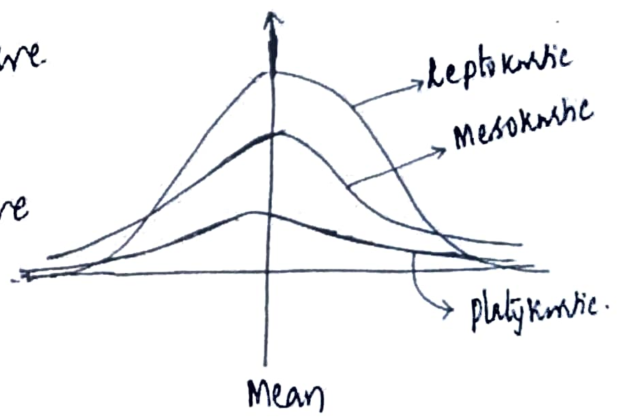
for a normal curve $\beta_2 = 3 \Rightarrow \gamma_2 = 0$ known as MESOKURTIC.

for a LEPTOKURTIC curve

$\beta_2 > 3 \Rightarrow \gamma_2 > 0$

for a PLATYKURTIC curve

$\beta_2 < 3 \Rightarrow \gamma_2 < 0$.



Applications of Moments :

- Bussiness Management system
- Data Science
- Foresenic analysis
- Digital Image Processing

Problem 1:

Calculate the first four moments about the mean and the point $A=5$. Also comment on nature of distribution.

x	1	2	3	4	5	6	7	8	9
f	1	6	13	25	30	22	9	5	2

Solution: WKT the moments about the point 'A' is

$$\mu_1' = \frac{\sum f_i\cdot(x_i - A)}{N} = \frac{\sum f_i\cdot(x_i-5)}{N} = \frac{\sum f_i\cdot d_i}{N}$$

$$\mu_2' = \frac{\sum f_i\cdot(x_i - A)^2}{N} = \frac{\sum f_i\cdot(x_i-5)^2}{N} = \frac{\sum f_i\cdot d_i^2}{N}$$

$$\mu_3' = \frac{\sum f_i\cdot(x_i - A)^3}{N} = \frac{\sum f_i\cdot(x_i-5)^3}{N} = \frac{\sum f_i\cdot d_i^3}{N}$$

$$\mu_4' = \frac{\sum f_i\cdot(x_i - A)^4}{N} = \frac{\sum f_i\cdot(x_i-5)^4}{N} = \frac{\sum f_i\cdot d_i^4}{N}$$

WKT $\mu_1' = \bar{x} - A$

$\Rightarrow \bar{x} = \mu_1' + A$

$\therefore$ mean $= \mu_1' + 5$. $(\because A = 5)$

Moments about the mean:

$$\mu_1 = 0 ; \quad \mu_2 = \mu_2' - (\mu_1')^2$$

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

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

x	f	$d = x - A$ $d = x - 5$	fd	fd^2	fd^3	fd^4
1	1	-4	-4	16	-64	256
2	6	-3	-18	54	-162	486
3	13	-2	-26	52	-104	208
4	25	-1	-25	25	-25	25
5	30	0	0	0	0	0
6	22	1	22	22	22	22
7	9	2	18	36	72	144
8	5	3	15	45	135	405
9	2	4	8	32	128	512
-	113 $= N$	0	-10 $= \Sigma fd$	282 $= \Sigma f d^2$	2 $= \Sigma f \cdot d^3$	2058 $\Sigma f \cdot d^4$

Hence, $\mu_1' = \frac{\sum f_i \cdot d_i}{N} = \frac{-10}{113} = -0.0885$

$$\mu_2' = \frac{\sum f_i \cdot d_i^2}{N} = \frac{282}{113} = 2.4956$$

$$\mu_3' = \frac{\sum f_i \cdot d_i^3}{N} = \frac{2}{113} = 0.0177$$

$$\mu_4' = \frac{\sum f_i \cdot d_i^4}{N} = \frac{2058}{113} = 18.2124.$$

Hence moments about mean are :

$\mu_1 = 0$; $\mu_2 = \mu_2' - [\mu_1']^2 = 2.4956 - (-0.0885)^2$

$$\boxed{\mu_2 = 2.4878}$$

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

$$\mu_3 = 0.0177 - 3(2.4956) \cdot (-0.0885) + 2(-0.0885)^3$$

$$\boxed{\mu_3 = 0.6789}$$

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

$$\mu_4 = 18.2124 - 4(0.0177) \cdot (-0.0885) + 6(2.4956)(-0.0885)^2 - 3(-0.0885)^3$$

$$\boxed{\mu_4 = 18.3357}$$

To find coefficient of skewness:

W.K.T $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(0.6789)^2}{(2.4878)^3} = 0.0299$

$\gamma_1 = \sqrt{\beta_1} = \sqrt{0.0299} = 0.173$ (+ve)

$\beta_2 = \dfrac{\mu_4}{(\mu_2)^2} = \dfrac{18.3357}{(2.4878)^2} = 2.9626$

$\gamma_2 = \beta_2 - 3 = -0.0373$

Interpretation: (i) Since $\gamma_1 = 0.17370 \Rightarrow$ distribution is +vely skewed
(ii) Since $\beta_2 = 2.9626 < 3 \Rightarrow$ frequency curve is slightly platy kurtic.

Session 2 (Lecture)

Sl. No.	Time (in Minutes)	Content	Learning Aid and Methodology	Faculty approach	Typical student Activity	Skill and competency developed
1	10	Review of session 1	-	Discussion	Interaction	Students able to recall definition and formulas
2	40	Problems on moments, Skewness and Kurtosis	Written material	Explain	Students comprehend the content	Students able to solve the problems on similar type
3	5	Conclusion	Interaction	summarizing	Students recall the contents	Students able to get the familiarity about the session

Problem No.1:

The following table gives the distribution of marks in Mathematics of 100 students in an examination. find the first four moments and the coefficient of Skewness and Kurtosis.

Marks	60-62	63-65	66-68	69-71	72-74
No. of students	05	18	42	27	08

Solution: for the given distribution, consider the midpoint of the class intervals. Let us take $x_1 = 61$, $x_2 = 64$, $x_3 = 67$, $x_4 = 70$, $x_5 = 73$.

WKT Mean $= \bar{x} = \frac{\sum x_i f_i}{N} = \frac{5(61) + 18(64) + 42(67) + 27(70) + 8(73)}{100}$

$$\boxed{\bar{x} = 67.45}$$

x	f	$d = x-\bar{x}$ $d = x - 67.45$	$f \cdot d$	$f \cdot d^2$	$f \cdot d^3$	$f \cdot d^4$
61	5	-6.45	-32.25	208.012	-1341.6806	8653.840
64	8	-3.45	-62.10	214.245	-739.145	2550.501
67	42	-0.45	-18.90	8.505	-3.8272	1.7222
70	27	2.55	68.85	175.567	447.697	1141.627
73	8	5.55	44.40	246.420	1367.631	7590.352
-	100	-	0	852.75	-269.325	19937.593

Moments about mean:

$$\mu_1 = \frac{\Sigma f_i(x_i - \bar{x})}{N} = \frac{\Sigma f_i \cdot d_i}{N} = \frac{0}{100} = 0$$

$$\mu_2 = \mu_2' - \mu_1' = \frac{\Sigma f d^2}{N} - \frac{\Sigma f d}{N}$$

$$\mu_2 = 8.5275 - 0 = 8.5275.$$

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

$$\mu_3' = \frac{\Sigma f d^3}{N} = \frac{-269.325}{100}$$

$$\mu_3 = -2.6932 - 3(8.5275)(0) + 2(0)^3$$

$$\boxed{\mu_3 = -2.6932}$$

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

$$\mu_4' = \frac{\Sigma f d^4}{N} = 199.375$$

$$\boxed{\mu_4 = 199.375}$$

Coefficient of Skewness:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-2.6932)^2}{(8.5273)^3} = 0.10815$$

$$\gamma_1 = \sqrt{\beta_1} = \sqrt{0.10815} = 0.3289$$

Coefficient of Kurtosis $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{199.375}{(8.5273)^2}$

$$\boxed{\beta_2 = 2.7417}$$

Problem No 2:

The first four moments about the working mean is 28.5 of a distribution are 0.294, 7.144, 42.409 and 454.98. Calculate the moments about mean. Also evaluate β_1, β_2 and comment upon the skewness and kurtosis of the distribution.

Solution: Given moments about the working mean as

$$\mu_1' = 0.294 ; \quad \mu_2' = 7.144 .$$

$$\mu_3' = 42.409 ; \quad \mu_4' = 454.98$$

WKT $\quad \mu_1' = \bar{x} - A \Rightarrow 0.294 = \bar{x} - 28.5$

$$\therefore \boxed{\bar{x} = 28.794}$$

Since, $\mu_1 = 0$ (always) $\therefore \boxed{\mu_1 = 0}$

Since, $\mu_2 = \mu_2' - [\mu_1']^2 = 7.144 - (0.294)^2 = 7.058$

Since, $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$

$\mu_3 = 42.409 - 3(7.144)(0.294) + 2(0.294)^3$

$\boxed{\mu_3 = 36.151}$

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

$\mu_4 = 454.98 - 4[42.409][0.294] + 6[7.144][0.294]^2 - 3(0.294)$

$\boxed{\mu_4 = 408.738}$

Since $\beta_1 = \frac{[\mu_3]^2}{\mu_2^2} = \frac{(36.151)^2}{(7.058)^2} = 3.717$ & $\beta_2 = \frac{\mu_4}{(\mu_2)^2} = \frac{408.734}{(7.058)^2} = 8.205.$

& $\gamma_1 = \sqrt{\beta_1} = 1.928$ which shows that highly +vely skewed.

$\gamma_2 = \beta_2 - 3 = 5.205$ which shows that the distribution is Leptokurtic

Problem 3:

The first moments of a distⁿ about the value 3 are 2, 10, -30. Show that the moments about $x=0$ are 5, 31, 141. Find mean & variance.

Solⁿ:- By data if $a=3$, $\mu_1' = 2$, $\mu_2' = 10$, $\mu_3' = -30$

Find the moments about the mean

μ_2 & μ_3 as $\boxed{\mu_1 = 0}$ always

WKT $\bar{x} = \mu_1' + a = 2 + 3 = 5$ $\therefore \boxed{\text{Mean} = 5}$

$\sigma^2 = \mu_2 = \mu_2' - (\mu_1')^2 = 10 - (2)^2 = 6$

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

$= -30 - 3(10)(2) + 2(2)^3 = -74.$

$\Rightarrow \boxed{\mu_3 = -74}$

To find moments about $x=0$, Take $\boxed{a=0}$

$\bar{x} = \mu_1' + a$ i.e $5 = \mu_1' + 0 \Rightarrow \boxed{\mu_1' = 5}$

$\mu_2 = \mu_2' - (\mu_1')^2 \Rightarrow 6 = \mu_2' - (5)^2 \Rightarrow \boxed{\mu_2' = 31}$

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

$-74 = \mu_3' - 3(31)(5) + 2(5)^3 \Rightarrow \boxed{\mu_3' = 141}$

$\therefore$ The first three moments about $x=0$ are 5, 31, 141.

(4) The first 3 moments about the point 2 are 1, 16 + -40, resp'ly. S.T mean is 3, $\sigma^2 = 15$ & $\mu_3 = -86$

$\mu_1' = \bar{x} - A \Rightarrow \bar{x} = 3,$ $\quad \sigma^2 = \mu_2 = \mu_2' - (\mu_1')^2$

$= 16 - 1$

$= 15$

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

$= \underline{\underline{-86}}$