26/06/2019
widnesday

Date: / /
Page:

ST1CMT01

Basic Statistics and Introductary Probability Theory

→ Module I

· Graphical representation 1. Histogram 2. Frequency Polygon 3. Frequency curve 4. ogives and stem and leaf chart.

· Measures of central Tendency - Mean, median, Mode, Quartile points - Quadriles, Percentiles, Deciles.

· Measures of Dispersion - Range, Quartile deviation, mean deviation, standard deviation, coefficient of variation. Box plot

→ Module II
Introduction to Bivariate data, Scatter Diagram, curve fitting by the method of Least Squares (without Proof). fitting of straight lines, exponential curve, power curve; Linear corelation, covariance method (formula only) and simple problems, Linear regression - regression equations identification of regression lines and properties

→ Module III
Probability concepts, Random expirement, sample space, events, probability measure, approaches to probability - Classical, Statistical, Axiomatic; Addition Theorem (upto 3 events), Conditional

Probability, Independence of events, multiplication Theorem (upto three events), Total probability law, Bayes Theorem and its applications

→ Module IV

Random variables and distribution functions, random variables, probability density (mass) function, distribution function, properties expectation of a discrete and continous random variables - properties (without proof) mean and standard deviation of different probability density function, moment generating function, important properties (without proof),

Textbooks

1. SP Gupta - Statistical methods (Sultan Chand and sons) Publication
2. SC Gupta and V.K Kapoor - Fundamentals of Mathematical statistics ⇒ publication (Sultan chand and sons)
3. B.L Agarwal - Basic Statistics (New Age International) ⇐ Publication

MODULE - I

Measures of central Tendency

- Arithemetic mean

It is defined as the ratio between the sum of the values and their number. It is calculated as

If $x_1, x_2, x_3 \ldots x_n$ are 'n' individual values then

$$\boxed{\text{Arithemetic mean} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n}}$$

- Arithemetic mean in individual series

* Arithemetic mean in individual series can be calculated by

1. Direct Method
2. Shortcut method

1. Direct Method

Steps:- 1. Consider the given values and call it $x$
2. Add all the values and the sum will be $\Sigma x$
3. Count the number of values and call it $n$
4. Divide $\Sigma x$ by $n$
5. This is arithemetic mean and is denoted as

$$\boxed{\bar{x} = \frac{\Sigma x}{n}}$$

Date: / /
Page:

Ques 1 Calculate Arithemetic mean of daily income of five families Rs 10, Rs 90, Rs 85, Rs 103, Rs 11.

Arithemetic mean = $\frac{\Sigma x}{n}$

$\Sigma x = 10 + 90 + 85 + 103 + 11$

$\Sigma x = 299$

Mean = $\frac{299}{5}$

mean = $\underline{\underline{59.8}}$

2. Shortcut method

Steps: - 1. Take any ~~a~~ one value from the series and calls it assumed average (a)

2. Subtract this assumed average from all the given values. These differences are denoted by d. $\therefore \boxed{d = x - a}$

3. Find the sum of all these differences and call it $\Sigma d$

4. Then divide $\frac{\Sigma d}{n}$ and add it to a. we get the actual mean. i.e. a

$\therefore \boxed{Am = a + \frac{\Sigma d}{n}}$

Date: / /
Page:

Ques 2 Find arithemetic mean of the following values
45, 48, 50, 52, 60

Let $a = 50$

| $x$ | $d = x - a$ i.e. $d = x - 50$ |
|---|---|
| 45 | -5 |
| 48 | -2 |
| 50 | 0 |
| 52 | 2 |
| 60 | 10 |

$\Sigma d = -5 + -2 + 12$
$= -7 + 12$
$\Sigma d = 5$

$Am = a + \frac{\Sigma d}{n}$
$= 50 + \frac{5}{5}$

$Am = \underline{\underline{51}}$

• Arithemetic Mean in Discrete Series

It can be calculated by
1) Direct Method 2) Shortcut method 3) Step Deviation Method

1). Direct Method
Let $x_1, x_2, x_3 \dots x_n$ be n observations

with corresponding frequencies $f_1, f_2, f_3 \ldots f_n$ respectively. Let $f_1 + f_2 + f_3 \ldots f_n = N$

Then arithmetic mean:

$$\boxed{\begin{aligned}\bar{x} &= \frac{x_1 f_1 + x_2 f_2 + \ldots x_n f_n}{f_1 + f_2 + \ldots f_n} \\ &= \frac{x_1 f_1 + x_2 f_2 + \ldots x_n f_n}{N} \\ &= \frac{\Sigma f x}{N}\end{aligned}}$$

Ques 3 Calculate mean from the following data

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| values : | 5 | 15 | 25 | 35 | 45 | 55 | 65 | 75 |
| frequency : | 15 | 20 | 25 | 24 | 12 | 31 | 71 | 52 |

| x | f | xf |
|---|---|---|
| 5 | 15 | 75 |
| 15 | 20 | 300 |
| 25 | 25 | 625 |
| 35 | 24 | 840 |
| 45 | 12 | 540 |
| 55 | 31 | 1705 |
| 65 | 71 | 4615 |
| 75 | 52 | 3900 |
| | N=250 | |

$\Sigma fx = 12600$

$\Sigma fx = 75 + 300 + 625 + 840 + 540 + 1705 + 4615 + 3900$

$$\bar{x} = \frac{\Sigma fx}{N}$$

$$= \frac{12600}{250} = \frac{1260}{25} = \frac{252}{5}$$

$$\bar{x} = 50.4$$

2) Short-cut method

- Consider the given values of the variable as $x$.
- From all these values subtract an assumed average $a$ $\therefore$ $\boxed{d = x - a}$
- Multiply $d$ with corresponding frequencies to get $fd$ values
- Find the sum of $fd$ values and call it $\Sigma fd$
- Find the sum of frequencies that is $\Sigma f$ and take it as $N$
- Divide $\Sigma fd$ by $N$
- Mean is obtained by adding $\frac{\Sigma fd}{N}$ with $a$

$\therefore$ Arithmetic mean $\Rightarrow$ ~~a + Σ~~

$$\boxed{AM = a + \frac{\Sigma fd}{N}}$$

Ques 4

$a = 35$

| x | f | d = x − a (x − 35) | fd |
|---|---|---|---|
| 5 | 15 | −30 | −450 |
| 15 | 20 | −20 | −400 |
| 25 | 25 | −10 | −250 |
| 35 | 24 | 0 | 0 |
| 45 | 12 | 10 | 120 |
| 55 | 31 | 20 | 620 |
| 65 | 71 | 30 | 2130 |
| 75 | 52 | 40 | 2080 |
| | N = 250 | | |

$\Sigma fd = 3850$

$$Am = a + \frac{\Sigma fd}{N}$$

$$= 35 + \frac{3850}{250} = 35 + 15.4$$

Am = 50·4

3) Step Deviation method

~~Let~~ If the values have a common difference c then take $\boxed{d' = \frac{x-a}{c}}$

ie $\boxed{d' = \frac{d}{c}}$ OR $\boxed{d = d'c}$

$\therefore \boxed{\bar{x} = a + \frac{\Sigma fd'}{N} \times c}$

| x | f | $d' = \frac{x-35}{10}$ | fd' |
|---|---|---|---|
| 5 | 15 | -3 | -45 |
| 15 | 20 | -2 | -40 |
| 25 | 25 | -1 | -25 |
| 35 | 24 | 0 | 0 |
| 45 | 12 | 1 | 12 |
| 55 | 31 | 2 | 62 |
| 65 | 71 | 3 | 213 |
| 75 | 52 | 4 | 208 |
| | N = 250 | | $\Sigma fd' = 385$ |

a = 35 , c = 10

$\bar{x} = a + \frac{\Sigma fd'}{N} \times c$

$\bar{x} = 35 + \frac{385}{250} \times 10$

$= 35 + 15·4$

$\bar{x} = 50·4$

Arithmetic mean in continous series (exclusive type)

In the case of continous series we write the mid value of classes as x. Therefore

$\boxed{x = \frac{l_1 + l_2}{2}}$ where $l_1$ is the lower limit and $l_2$ is the upper limit of each class

Then we find arithmetic mean by using three formulae

$\boxed{1)\ \frac{\Sigma fx}{N}}$

$\boxed{2)\ a + \frac{\Sigma fd}{N}}$

$\boxed{3)\ a + \frac{\Sigma fd'}{N} \times c}$

Ques 5 Find AM from the following data

| Age | No of person dying |
|---|---|
| 0-10 | 15 |
| 10-20 | 30 |
| 20-30 | 53 |
| 30-40 | 75 |
| 40-50 | 100 |
| 50-60 | 110 |
| 60-70 | 115 |
| 70-80 | 125 |

$x = \frac{l_1 + l_2}{2}$

| Class | f | x | fx | $d = x - a$ | fd | $d' = d/c$ | fd' |
|---|---|---|---|---|---|---|---|
| 0-10 | 15 | $\frac{0+10}{2} = 5$ | 75 | -30 | -450 | -3 | ~~-225~~ -45 |
| 10-20 | 30 | = 15 | 450 | -20 | -600 | -2 | ~~-200~~ -60 |
| 20-30 | 53 | 25 | 1325 | -10 | -530 | -1 | -53 |
| 30-40 | 75 | 35 | 2625 | 0 | 0 | 0 | 0 |
| 40-50 | 100 | 45 | 4500 | 10 | 1000 | 1 | 100 |
| 50-60 | 110 | 55 | 6050 | 20 | 2200 | 2 | 220 |
| 60-70 | 115 | 65 | 7475 | 30 | 3450 | 3 | 345 |
| 70-80 | 125 | 75 | 9375 | 40 | 5000 | 4 | 500 |
| | N = 623 | | $\Sigma fx = 31875$ | | $\Sigma fd = 10070$ | | $\Sigma fd' = 1007$ |

~~a = 35~~

(1) $\frac{\Sigma fx}{N} = \frac{31875}{623} = 51.16$ years

(2) $a + \frac{\Sigma fd}{N}$

Let $a = 35$

$= 35 + \frac{10070}{623}$

$= 51.16$ years

(3) $a + \frac{\Sigma fd'}{N} \times c$

$a = 35$

$c = 10$

$= 35 + \frac{1007}{623} \times 10$

$= 35 + \frac{10070}{623}$

$= 51.16$ years

Ques 6 Find mean from the following frequency distribution

| Class | 15-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-75 |
|---|---|---|---|---|---|---|
| frequency | 4 | 11 | 19 | 14 | 0 | 2 |

| Class | frequency | $x = \frac{l_1+l_2}{2}$ | $fx$ | $d = x - a$ | $fd$ | $d' = d/c$ ($d/10$) | $fd'$ |
|---|---|---|---|---|---|---|---|
| 15-25 | 4 | 20 | 80 | -30 | -120 | -3 | -12 |
| 25-35 | 11 | 30 | 330 | -20 | -220 | -2 | -22 |
| 35-45 | 19 | 40 | 760 | -10 | -190 | -1 | -19 |
| 45-55 | 14 | 50 | 700 | 0 | 0 | 0 | 0 |
| 55-65 | 0 | 60 | 0 | 10 | 0 | 1 | 0 |
| 65-75 | 2 | 70 | 140 | 20 | 40 | 2 | 4 |
| | $N = 50$ | | $\Sigma fx = 2010$ | | $\Sigma fd = -490$ | | $\Sigma fd' = -49$ |

① $\frac{\Sigma fx}{N} =$

$= \frac{2010}{50} = \frac{201}{5} = 40.2$

② $a + \frac{\Sigma fd}{N}$, $a = 50$. $\Sigma fd = -590$

$= 50 + \frac{-590}{50}$ $= 50 + \frac{-49}{5}$

$= 50 + -11.8$

$= 38.2$

③ $a + \frac{\Sigma fd'}{N} \times c$, $c = 10$

$a = 50$,

$= a + \frac{\Sigma fd'}{N} \times c$

$= 50 + \frac{-49 \times 10}{50}$

$= 50 + \frac{-49}{5}$

## Arithmetic mean in inclusive series

For all classes subtract 0.5 from the lower limit and add 0.5 with the upper limit. 0.5 is half the difference between upper limit of a class and the lower limit of the succeeding class

$\frac{20-19}{2} = \frac{1}{2} = 0.5$

eg:-

| Class | frequency |
|---|---|
| 10-19 | 3 |
| 20-29 | 5 |
| 30-39 | 7 |
| 40-49 | 2 |

$\Rightarrow$

| class | frequency |
|---|---|
| 9.5-19.5 | 3 |
| 19.5-29.5 | 5 |
| 29.5-39.5 | 7 |
| 39.5-49.5 | 2 |

Ques 7 ~~From~~ calculate mean from the following data

| Pocket Expenses | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 |
|---|---|---|---|---|---|
| No. of students | 10 | 8 | 6 | 4 | 2 |

| P.E | f | Class | $x = \frac{l_1+l_2}{2}$ | fx |
|---|---|---|---|---|
| 20-29 | 10 | 19.5-29.5 | 49/2 = 24.5 | 245 |
| 30-39 | 8 | 29.5-39.5 | 69/2 = 34.5 | 276 |
| 40-49 | 6 | 39.5-49.5 | 89/2 = 44.5 | 267 |
| 50-59 | 4 | 49.5-59.5 | 109/2 = 54.5 | 218 |
| 60-69 | 2 | 59.5-69.5 | 129/2 = 64.5 | 129 |
| | $\boxed{N=30}$ | | | $\boxed{1135}$ |

mid values $\frac{30-29}{2} = \frac{1}{2} = 0.5$

① $\frac{\Sigma fx}{N} = \frac{1135}{30} = 37$

## Arithmetic Mean in Cumulative (less than) series

$f_2 - f_1$

| Class | frequency |
|---|---|
| Below 10 | 3 |
| Below 20 | 10 |
| Below 30 | 25 |
| Below 40 | 30 |
| ~~Below 50~~ | |

$\Rightarrow$

| Class | frequency |
|---|---|
| 0-10 | 3 |
| 10-20 | 7 |
| 20-30 | 15 |
| 30-40 | 5 |

Ques 8 Find arithmetic mean of the following data

| marks | no. of students |
|---|---|
| Less than 10 | ~~no. of students~~ 5 |
| less than 20 | 17 |
| less than 30 | 31 |
| less than 40 | 41 |
| less than 50 | 49 |

| Class | f | $x = \frac{l_1+l_2}{2}$ | fx |
|---|---|---|---|
| 0-10 | ~~17~~ 5 | 5 | 25 |
| 10-20 | ~~24~~ 12 | 15 | 180 |
| 20-30 | ~~10~~ 14 | 25 | 350 |
| 30-40 | 10 | 35 | 350 |
| 40-50 | 8 | 45 | 360 |
| | N = 49 | | $\Sigma fx =$ ~~1015~~ 1265 |

$\bar{x} = \frac{\Sigma f x}{N}$

$= \frac{2015}{59} = \frac{1265}{49}$

$= \underline{25.81}$

Arithmetic mean in cumilative (more than) series

$f_1 - f_2$

| Class | Frequency |
|---|---|
| Above 10 | 40 |
| Above 20 | 32 |
| Above 30 | 12 |
| Above 40 | 5 |

$\Rightarrow$

| Marks | frequency |
|---|---|
| 10 - 20 | 8 |
| 20 - 30 | 20 |
| 30 - 40 | 7 |
| 40 - 50 | 5 |

Ques 9 Find the arithmetic mean

| Marks | No. of students |
|---|---|
| More than 0 | 30 |
| " 2 | 28 |
| " 4 | 24 |
| " 6 | 18 |
| " 8 | 10 |

| Marks | No. of students (frequency) | $x = \frac{l_1 + l_2}{2}$ | $fx$ |
|---|---|---|---|
| 0 - 2 | 2 | 1 | 2 |
| 2 - 4 | 4 | 3 | 12 |
| 4 - 6 | 6 | 5 | 30 |
| 6 - 8 | 8 | 7 | 56 |
| 8 - 10 | 10 | 9 | 90 |
| | N = 30 | | $\Sigma fx = 190$ |

$\bar{x} = \frac{\Sigma f x}{N} = \frac{190}{30} = \frac{19}{3} = 6.333$

## Combined Mean

If a sample has size $n_1$ and mean $\bar{x}_1$ and another sample has size $n_2$ and mean $\bar{x}_2$ then mean of the combined sample

$$\boxed{= \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}}$$

Ques 10 If a sample of size 22 items has a mean of 15 and another sample of size 18 items has a mean of 20, find mean of the combined sample

$n_1 = 22$ $\bar{x}_1 = 15$

$n_2 = 18$ $\bar{x}_2 = 20$

Combined mean $= \frac{22 \times 15 + 18 \times 20}{22 + 18} = \frac{330 + 360}{40}$

$$= \frac{690}{40} = 17.25$$

Combined mean is 17.25

Ques 11 The mean wage of 100 labourers working in a factory running two shifts of 60 and 40 workers respectively is Rs 38. The mean wage of 60 labourers working in the morning shift is Rs 40. Find the mean wage of labourers working in the evening shift.

Combined mean $\bar{x}$ = Rs 38

$n_1 + n_2 = 100$

$n_1 = 60$, $n_2 = 40$

$\bar{x}_1 = 40$, $\bar{x}_2 = ?$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} = \frac{60\times 40 + 40\times \bar{x}_2}{100}$$

$$38 = \frac{2400 + 40\bar{x}_2}{100}$$

$$3800 - 2400 = 40\bar{x}_2$$

$$1400 = 40\bar{x}_2$$

$$\bar{x}_2 = \frac{1400}{40} = 35$$

Mean wage of labourers in evening shift is Rs 35.

Ques 12 The mean annual salary payed to all employees of a company was Rs 5,000. The mean annual salaries payed to male and female employees were Rs 5,200 and Rs 4,200 respectively. Determine the percentage of males and females employed by the company

$\bar{x} = 5,000$

$\bar{x}_1 = 5200$

$\bar{x}_2 = 4200$

Assume $n_1 + n_2 = 100$

where $n_1$ = male employee

$100 - n_1$ = female employee

$$\bar{x} = \frac{n_1 \times 5200 + (100 - n_1)4200}{100}$$

$$5000 = \frac{5200n_1 + 420000 - 4200n_1}{100}$$

$$500000 = 1000n_1 + 420000$$

$$500000 - 420000 = 1000n_1$$

$$80000 = 1000n_1$$

$$n_1 = \frac{80000}{1000}$$

$$n_1 = 80$$

$$n_2 = 100 - n_1$$
$$n_2 = 100 - 80$$
$$n_2 = 20$$

The percentages are 80% and 20%.

## Corrected Mean

Ques 13 Mean marks obtained by 100 students was found to be 40. Later on it was noted that one value was read as 83 instead of 53. Find out the correct mean.

$$\bar{x} = \frac{\Sigma x}{n}$$

$n = 100$ , $\bar{x} = 40$

$$40 = \frac{\Sigma x}{100}$$

$$\Sigma x = 4000$$

$$\Sigma x = 4000 + 53 - 83$$
$$= 4000 - 30$$
$$\Sigma x = 3970$$

$$\bar{x} = \frac{\Sigma x}{n}$$

$$\bar{x} = \frac{3970}{100}$$

$$\bar{x} = 39.7$$

correct mean $\bar{x} = 39.7$

HW

Ques 14 Arithmetic mean of 100 items is 34. At the time of calculation 118, 70, and 19 were wrongly taken as 180, 17, and 90 resp. What is the correct mean.

Ques 15 Find mean of the following data

| Sl.no. | Marks |
|---|---|
| 1 | 25 |
| 2 | 32 |
| 3 | 18 |
| 4 | 20 |
| 5 | 35 |
| 6 | 40 |
| 7 | 70 |
| 8 | 49 |
| 9 | 50 |
| 10 | 24 |

HW

Ques 16 find out the mean

| Marks | No. of students |
|---|---|
| 0 - 10 | 5 |
| 10 - 20 | 7 |
| 20 - 30 | 15 |
| 30 - 40 | 25 |
| 40 - 50 | 20 |
| 50 - 60 | 15 |
| 60 - 70 | 8 |
| 70 - 80 | 5 |

HW

Ques 17 Calculate the mean

| Marks (less than) | No. of students |
|---|---|
| 10 | 5 |
| 20 | 13 |
| 30 | 20 |
| 40 | 32 |
| 50 | 60 |
| 60 | 80 |
| 70 | 90 |
| 80 | 100 |

HW

Ques 18 A sample of size 80, 40, 30 having means 12.5, 13 and 11 respectively are combined. Find the mean of the combined sample.

Ques 14

$n = 100$

$x = 34$

$\bar{x} = \frac{\Sigma x}{n}$

$\Sigma x = \bar{x} \times n$

$= 34 \times 100$

$= 3400$

real $\Sigma x$,

$\Sigma x = 3400 + (118 + 70 + 19) - (180 + 17 + 90)$

$= 3400 + 207 - 287$

$= 3400 - 80$

$\Sigma x = 3320$

$\bar{x} = \frac{\Sigma x}{n}$

$= \frac{3320}{100} = 33.2$

Ques 15

$n = 10$

$\Sigma x = 363$

$\bar{x} = \frac{363}{10}$

$\bar{x} = 36.3$

Ques 16

$\bar{x} = \frac{\Sigma fx}{n}$

| marks | f | $x = \frac{l_1+l_2}{2}$ | fx |
|---|---|---|---|
| 0-10 | 5 | 5 | 25 |
| 10-20 | 7 | 15 | 105 |
| 20-30 | 15 | 25 | 375 |
| 30-40 | 25 | 35 | 875 |
| 40-50 | 20 | 45 | 900 |
| 50-60 | 15 | 55 | 825 |
| 60-70 | 8 | 65 | 520 |
| 70-80 | 5 | 75 | 375 |
| | $\boxed{N=100}$ | | $\boxed{\Sigma fx = 4000}$ |

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{4000}{100} = \underline{\underline{40}}$$

Ques 17

| Marks/Class | No. of students (f) | $x = \frac{l_1+l_2}{2}$ | fx |
|---|---|---|---|
| 0-10 | 5 | 5 | 25 |
| 10-20 | 8 | 15 | 120 |
| 20-30 | 7 | 25 | 175 |
| 30-40 | 12 | 35 | 420 |
| 40-50 | 28 | 45 | 1260 |
| 50-60 | 20 | 55 | 1100 |
| 60-70 | 10 | 65 | 650 |
| 70-80 | 10 | 75 | 750 |
| | $\boxed{N = 100}$ | | $\Sigma fx = 4500$ |

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{4500}{100} = \underline{\underline{45}}$$

Q18 Combined mean, $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1 + n_2 + n_3}$

$n_1 = 80$, $n_2 = 40$, $n_3 = 30$

$\bar{x}_1 = 12.5$, $\bar{x}_2 = 13$, $\bar{x}_3 = 11$

$$\bar{x} = \frac{80 \times 12.5 + 40 \times 13 + 30 \times 11}{80 + 40 + 30}$$

$$= \frac{1000 + 520 + 330}{150}$$

$$= \frac{1850}{150}$$

$$= \underline{\underline{12.33}}$$

**Weighed mean**

Let $x_1, x_2 \dots x_n$ be n values with corresponding weights $w_1, w_2 \dots w_n$ then the weighed average

$$= \boxed{\frac{w_1x_1 + w_2x_2 + \dots w_nx_n}{w_1 + w_2 + \dots w_n}}$$

$$= \boxed{\frac{\Sigma wx}{\Sigma w}}$$

Q:-19 A canditade obtained the following % of marks English - 60, Hindi - 75, Maths - 63, Physics 60, chemistry - 55. Find the weighted mean if weights are 1, 1, 2, 3, 3

| Subject | Marks ($x$) | weight ($w$) | $wx$ |
|---|---|---|---|
| English | 60 | 1 | 60 |
| Hindi | 75 | 1 | 75 |
| Maths | 63 | 2 | 126 |
| Physics | 60 | 3 | 180 |
| chemistry | 55 | 3 | 165 |
| | | $\Sigma w = 10$ | $\Sigma wx = 606$ |

$$\bar{x} = \frac{\Sigma wx}{\Sigma w}$$

$$= \frac{606}{10}$$

$$\bar{x} = 60.6$$

7/2019 **Median**

Median of a series is the size of that item of the series which occupies the central position of the series when the items are arranged in the ascending or descending order of their magnitude.

OR

→ The median is that value of the variable which divides the values of the variable into two equal parts,
One part containing all values greater than the median value and the other part containing all the values smaller than the median values

**Median in individual series**

Median = $\boxed{\text{size of } \left(\frac{n+1}{2}\right) \text{th item}}$, then the items are arranged in the ascending or descending order of their magnitude.

Ques 20 Find the median for the following values:-
4, 45, 60, 20, 83, 19, 26, 11, 27, 12, 52

4, 11, 12, 19, 20, 26, 27, 45, 52, 60, 83

$n = 11$

median = size of $\left(\frac{n+1}{2}\right)$ th item

$= \frac{11+1}{2}$ th

$= 6$th

The median is 26

Ques21 Calculate median:- 35, 23, 45, 50, 80, 61, 92, 40, 52, 61

23, 35, 40, 45, 50, 52, 61, 61, 80, 92

$n = 10$

size of $\left(\frac{11}{2}\right)^{th}$ item

size of $5.5^{th}$ item

$= \frac{\text{size of } 5^{th} + \text{size of } 6^{th} \text{ item}}{2}$

$= \frac{50+52}{2} = \frac{102}{2} = \underline{\underline{51}}$

Ques22 Locate median marks from the following data of marks secured by students in statistics

(i) 17, 32, 35, 33, 15, 21, 41, 32, 11, 10, 20

(ii) 12, 6, 14, 16, 3, 11, 8, 4

(iii) 390, 380, 540, 672, 784, 176, 197, 155, 135

(i) 10, 11, 15, 17, 20, 21, 32, 32, 33, 35, 41

$n = 11$

median = size of $\left(\frac{n+1}{2}\right)^{th}$ item

= size of $6^{th}$ item

median = 21

(ii) 3, 4, 6, 8, 11, 12, 14, 16

$n = 8$

median = size of $\left(\frac{n+1}{2}\right)^{th}$ item

= size of $4.5^{th}$ item

$= \frac{\text{size of } 4^{th} + \text{size of } 5^{th} \text{ item}}{2}$

$= \frac{8+11}{2} = \frac{19}{2} = \underline{9.5}$

(iii) 110, 135, 155, 176, 197, 380, 390, 540, 672, 784

$n = 10$ median = $11/2^{th}$ item = 5.5

median = $\frac{\text{size of } 5^{th} + \text{size of } 6^{th} \text{ item}}{2} = \frac{197+380}{2}$

$= \frac{577}{2} = 288.5$

**Median in Discrete Series**

Median = $\boxed{\text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ item}}$

where $\boxed{N = \Sigma f}$

Ques23 Calculate median

size - 5, 8, 10, 15, 20, 25

freq: 3, 12, 8, 7, 5, 4

| size | freq | cumulative frequency |
|---|---|---|
| 5 | 3 | 3 |
| 8 | 12 | 3+12 = 15 |
| 10 | 8 | 15+8 = 23 |
| 15 | 7 | 23+7 = 30 |
| 20 | 5 | 30+5 = 35 |
| 25 | 5 | 35+5 = 40 |
| | 4 | 40+4 = 44 |
| | N = 39 | |

Date: / /
Page:

size of $\frac{(N+1)}{2}^{th}$ item

| size | freq | cf |
|---|---|---|
| 5 | 3 | 3 |
| 8 | 12 | 15 |
| 10 | 8 | 23 |

size of $\frac{39+1}{2}^{th}$ item

= size of $20^{th}$ item

median = 10

Ques24 Calculate median:-

wages:-(Rs): 10 12 15 18 20 25 30
No. of workers: 3 5 8 12 13 12 7

| wages | freq. | cumulative freq |
|---|---|---|
| 10 | 3 | 3 |
| 12 | 5 | 8 |
| 15 | 8 | 16 |
| 18 | 12 | 28 |
| 20 | 13 | 41 |
| 25 | 12 | 53 |
| 30 | 7 | 60 |

$\boxed{N = \Sigma f = 60}$

always lesser than is taken

Size of $\frac{(N+1)}{2}^{th}$ item

Size of $\frac{61}{2}^{th}$ item

= Size of $30.5^{th}$ item

Median = 20

Date: / /
Page:

Ques25 The following table shows age of 8 students. Find median each

| S.no. | age |
|---|---|
| 1 | 18 |
| 2 | 16 |
| 3 | 14 |
| 4 | 11 |
| 5 | 13 |
| 6 | 10 |
| 7 | 9 |
| 8 | 20 |

n = 8

⇒ 9, 10, 11, 13, 14, 16, 18, 20

Median = size of $\left(\frac{n+1}{2}\right)^{th}$ term

= size of $\left(\frac{9}{2}\right)^{th}$ term

= size of $4.5^{th}$ term

Median = $\frac{\text{size of } 4^{th} + \text{size of } 5^{th} \text{ term}}{2} = \frac{13+14}{2}$

Median = $\frac{27}{2} = 13.5$

Ques26 Find the average no. of persons per house using median.

| No. of persons | no. of houses |
|---|---|
| 1 | 26 |
| 2 | 113 |
| 3 | 120 |
| 4 | 95 |
| 5 | 60 |
| 6 | 42 |
| 7 | 21 |
| 8 | 14 |
| 9 | 54 |

## Median in continous series

Steps

- Form the Cumulative Frequency Column
- Find $\frac{N}{2}$, median class correspondes to the cumulative frequency which includes $\frac{N}{2}$
- After getting median class, find median by using interpolation formula

$$\boxed{\text{Median} = l_1 + \frac{\left(\frac{N}{2} - cf\right) \times c}{f}}$$

where $l_1$ is the lower limit of the median class

→ $cf$ is the cumulative frequency of the class just proceeding the median class

→ $f$ is the frequency of median class

→ $c$ is the interval of median class

Ques 29 Compute median

| class | frequency | cf |
|---|---|---|
| 0-10 | 8 | 8 |
| 10-20 | 12 | 20 |
| 20-30 | 20 | 40 |
| 30-40 | 23 | 63 |
| 40-50 | 18 | 81 |
| 50-60 | 7 | 88 |
| 60-70 | 2 | 90 |
| | N = 90 | |

$\frac{N}{2} = \frac{90}{2} = 45$

∴ 30 - 40 is the median class

Applying interpolation formula

class = 40-30 = 10

$l_1 = 30$, $cf = 40$, $c = 10$, $f = 23$

$$\text{Median} = l_1 + \frac{\left(N/2 - cf\right) \times c}{f}$$

$$= 30 + \frac{45 - 40}{23} \times 10$$

$$= 30 + \frac{50}{23}$$

$$= 30 + 2.173$$

$$= 32.173$$

## Median in Inclusive Series

Ques 28 Calculate median of the following data

| Marks | No. of students |
|---|---|
| 11-15 | 7 |
| 16-20 | 10 |
| 21-25 | 13 |
| 26-30 | 26 |
| 31-35 | 35 |
| 36-40 | 22 |
| 41-45 | 11 |
| 46-50 | 5 |

| Class | freq | cf |
|---|---|---|
| 10.5-15.5 | 7 | 7 |
| 15.5-20.5 | 10 | 17 |
| 20.5-25.5 | 13 | 30 |
| 25.5-30.5 | 26 | 56 |
| 30.5-35.5 | 35 | 91 |
| 35.5-40.5 | 22 | 113 |
| 40.5-45.5 | 11 | 124 |
| 45.5-50.5 | 5 | 129 |
| | $\boxed{N=129}$ | |

31.75

$\frac{N}{2} = \frac{129}{2} = 64.5$

$c = 5$, $cf = 56$, $f = 35$, $l_1 = 30.5$

$\text{median} = 30.5 + \left(\frac{64.5 - 56}{35}\right) \times 5$

$= 30.5 + \frac{8.5}{7}$

$= 30.5 + 1.25$

$= \underline{31.75}$

## Median in Cumulative Frequency distribution

Ques 29 Calculate median

| Wage (less than) Rs | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| No. of workers | 15 | 35 | 60 | 84 | 96 | 127 | 198 | 250 |

| Wage | no. of workers | cf |
|---|---|---|
| 0-10 | 15 | 15 |
| 10-20 | 20 | 35 |
| 20-30 | 25 | 60 |
| 30-40 | 24 | 84 |
| 40-50 | 12 | 96 |
| 50-60 | 31 | 127 |
| 60-70 | 71 | 198 |
| 70-80 | 52 | 250 |
| | $\boxed{N=250}$ | |

$\frac{N}{2} = \frac{250}{2} = 125$ $\quad cf = 96$

$l_1 = 50$ $\quad f = 31$ $\quad c = 10$

$\text{Median} = 50 + \left(\frac{125 - 96}{31}\right) \times 10$

$= 50 + \frac{290}{31}$ $= 50 + 9.35$

$= \underline{59.35}$

## Median in mid-value series

Ques 30 Find median of the following series

| Mid (x) | 10 | 15 | 20 | 25 | 30 | 35 |
|---|---|---|---|---|---|---|
| freq. | 8 | 12 | 10 | 9 | 4 | 3 |

10-15
diff 5

| Class | f | cf |
|---|---|---|
| 7.5 - 12.5 | 8 | 8 |
| 12.5 - 17.5 | 12 | 20 |
| 17.5 - 22.5 | 10 | 30 |
| 22.5 - 27.5 | 9 | 39 |
| 27.5 - 32.5 | 4 | 43 |
| 32.5 - 37.5 | 3 | 46 |
| | N = 46 | |

$\frac{N}{2} = \frac{46}{2} = 23$

Median $= l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$

$l_1 = 17.5$, $N/2 = 23$, $cf = 20$, $f = 10$

$c = 5$

Median $= 17.5 + \left(\frac{23 - 20}{10}\right) \times 5$

$= 17.5 + \frac{15}{10}$

$= 17.5 + 1.5$

$= \underline{19}$

## Median in Open End Class Frequency Distribution

Ques 31 Find the most appropriate average of the following distribution

| Size of the farm (in acres) | No. of occupiers | cf |
|---|---|---|
| upto 50 0-50 | 57 | 57 |
| 50 - 100 | 256 | |
| 100 - 150 | 132 | |
| 150 - 200 | 25 | |
| 200 - 250 | 10 | |
| Over 250 250-300 | 12 | |

| Size of the farm | no. of occupiers | cf |
|---|---|---|
| 0 - 50 | 57 | 57 |
| 50 - 100 | 256 | 313 |
| 100 - 150 | 132 | 445 |
| 150 - 200 | 25 | 470 |
| 200 - 250 | 10 | 480 |
| 250 - 300 | 12 | 492 |
| | N = 492 | |

$N/2 = \frac{492}{2} = 246$ $l_1 = 50$ $cf = 57$

$f = 256$, $c = 50$

Median $= l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$

$= 50 + \frac{246 - 57}{256} \times 50$

$= 50 + \frac{189 \times 50}{256}$

$= 50 + \frac{9450}{256}$

$= 50 + 36.9$

$= 86.91$

Ques 32 Find median from the following data

| Marks | 15-25 | 25-35 | 35-45 | 45-55 | 55-65 | 65-75 | 75-85 |
|---|---|---|---|---|---|---|---|
| No. of students | 3 | 5 | 12 | 15 | 9 | 9 | 7 |

| Marks | frequency / No. of students | cf |
|---|---|---|
| 15-25 | 3 | 3 |
| 25-35 | 5 | 8 |
| 35-45 | 12 | 20 |
| 45-55 | 15 | 35 |
| 55-65 | 9 | 44 |
| 65-75 | 9 | 53 |
| 75-85 | 7 | 60 |
| N = 60 | | |

$\frac{N}{2} = \frac{60}{2} = 30 \quad f = 15 \quad cf = 20$

$C = 10, \quad l_1 = 45$

$\text{Median} = l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$

$= 45 + \left(\frac{30-20}{15}\right) \times 10 = 45 + \frac{100}{15}$

$= 45 + 6.6 = \underline{51.6}$

Ques 33 Compute median from the following data

| Midvalue | 115 | 125 | 135 | 145 | 155 | 165 | 175 | 185 | 195 |
|---|---|---|---|---|---|---|---|---|---|
| frequency | 6 | 25 | 28 | 72 | 116 | 60 | 38 | 22 | 3 |

| Mid values | class | f | cf |
|---|---|---|---|
| 115 | 110 - 120 | 6 | 6 |
| 125 | 120 - 130 | 25 | 31 |
| 135 | 130 - 140 | 28 | 59 |
| 145 | 140 - 150 | 72 | 131 |
| 155 | 150 - 160 | 116 | 247 |
| 165 | 160 - 170 | 60 | 307 |
| 175 | 170 - 180 | 38 | 345 |
| 185 | 180 - 190 | 22 | 367 |
| 195 | 190 - 200 | 3 | 370 |
| | | N = 370 | |

$N = 370$

$N/2 = 370/2 = 185$

$cf = 131$

$f = 116$

$l_1 = 150$

$c = 10$

$\text{Median} = l + \left(\frac{N/2 - cf}{f}\right) \times c$

$= 150 + \frac{185 - 131}{116} \times 10$

$= 150 + \frac{540}{116} = 150 + 4.6$

$= \underline{154.6}$

Ques 34 Calculate median

| Marks (less than) | 15 | 30 | 45 | 60 | 75 | 90 |
|---|---|---|---|---|---|---|
| No. of students | 18 | 35 | 62 | 81 | 95 | 100 |

| Marks | frequency / No. of students | cf |
|---|---|---|
| 0 - 15 | 18 | 18 |
| 15-30 | 17 | 35 |
| 30-45 | 27 | 62 |

| | | |
|---|---|---|
| 45-60 | 19 | 81 |
| 60-75 | 14 | 95 |
| 75-90 | 5 | 100 |
| | $\boxed{N=100}$ | |

$N = 100$

$N/2 = 50$

$l_1 = 30$, $c = 15$, $cf = 35$, $f = 27$

$$\text{Median} = l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$$

$$= 30 + \frac{50 - 35}{27} \times 15$$

$$= 30 + \frac{15 \times 15}{27}$$

$$= 30 + \frac{225}{27}$$

$$= 30 + 8.3$$

$$= 38.3$$

Ques 35 Find Median

| Age (more than) | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 |
|---|---|---|---|---|---|---|---|---|
| No. of people | 100 | 96 | 90 | 80 | 70 | 45 | 23 | 5 |

| Age | freq. | cf |
|---|---|---|
| 0-5 | 4 | 4 |
| 5-10 | 6 | 10 |
| 10-15 | 10 | 20 |
| 15-20 | 10 | 30 |
| 20-25 | 25 | 55 |
| 25-30 | 22 | 77 |

| | | |
|---|---|---|
| 30-35 | 18 | 95 |
| 35-40 | 5 | 100 |
| | $\boxed{N=100}$ | |

$N/2 = \frac{100}{2} = 50$

$cf = 30$, $f = 25$, $c = 5$, $l_1 = 20$

$$\text{Median} = l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$$

$$= 20 + \frac{50 - 30}{25} \times 5$$

$\frac{20 \times 5}{25}$

$$= 20 + 4$$

$$= 24$$

Ques 36 Calculate Median

| Class : | 10-20 | 20-40 | 40-70 | 70-120 | 120-140 |
|---|---|---|---|---|---|
| freq : | 4 | 10 | 26 | 8 | 2 |

| Class | freq | cf |
|---|---|---|
| 10-20 | 4 | 4 |
| 20-40 | 10 | 14 |
| 40-70 | 26 | 40 |
| 70-120 | 8 | 48 |
| 120-140 | 2 | 50 |
| | $\boxed{N=50}$ | |

$N = 50$

$N/2 = 50/2 = 25$

$l_1 = 40$, $cf = 14$, $f = 26$, $c = 30$

$$\text{Median} = l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$$

$$= 40 + \frac{25 - 14}{26} \times 30$$

$$= 40 + \frac{11 \times 30}{26}$$

$$= 40 + \frac{330}{26}$$

$$= 40 + 12.6$$

$$= 52.6$$

10/02/2017

## Mode

The value which occurs more frequently than any other value in the data is called mode

Mode in Individual Series

Arrange the values in ascending order then by inspection identify the value which occurs more number of times

Ques37 Find mode : 23, 35, 28, 42, 62, 53, 35, 28, 42, 35, 23, 42, 35

23, 23, 28, 28, 35, 35, 35, 35, 42, 42, 42, 53, 62

Mode = 35

Ques 38 Find mode from the following values :- 40, 25, 60, 35, 81, 75, 90, 10.

$$\text{Mean} = \frac{\Sigma x}{n}$$

$$= \frac{40 + 25 + 60 + 35 + 81 + 75 + 90 + 10}{8}$$

$$= \frac{416}{8} = 52$$

Median =

10, 25, 35, 40, 60, 75, 81, 90

n = 8

$$\text{Median} = \text{size of } \frac{(n+1)}{2}^{th} \text{ item}$$

$$= \text{size of } \frac{9}{2}^{th} \text{ item}$$

$$= \text{size of } 4.5^{th} \text{ item}$$

$$= \frac{\text{size of } 4^{th} \text{ item} + \text{size of } 5^{th} \text{ item}}{2}$$

$$= \frac{40 + 60}{2} = \frac{100}{2}$$

$$= 50$$

Mode = 3 Median − 2 Mean

= 3 × 50 − 2 × 52

= 150 − 104

Mode = 46

## Mode in the case of discrete series

In discrete series the value having higest frequency is taken as mode.

Ques 39 find mode

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| size : | 5 | 8 | 10 | 12 | 29 | 35 | 40 | 46 |
| no. of items | 3 | 12 | 25 | 40 | 31 | 20 | 18 | 7 |

mode = 12

(∵ The value 12 has the higest frequency)

## Mode in the case of continous series (Exclusive Type)

Mode lies in the class having the highest frequency. Mode is calculated using the interpolation formula

$$\boxed{Mode = l_1 + \frac{(f_1 - f_0) \times c}{2f_1 - f_0 - f_2}}$$

where $l_1$ is the lower limit of the modal class

→ $f_0$ and $f_2$ are respectively the frequencies of classes just preceeding an succeeding model class.

→ $f_1$ is the frequency of the model class

Ques 40 Find mode,

size : 10-15 15-20 20-25 25-30 30-35 35-40 40-45, 45-50

~~frequency:~~

freq : 4, 8, 18, 30, 20, 10, 5, 2

| size | frequency |
|---|---|
| 10 - 15 | 4 |
| 15 - 20 | 8 |
| 20 - 25 | 18 |
| 25 - 30 | 30 model class |
| 30 - 35 | 20 |
| 35 - 40 | 10 |
| 40 - 45 | 5 |
| 45 - 50 | 2 |

c = 5

$l_1 = 25$ ; $f_0 = 18$ , $f_1 = 30$ , $f_2 = 20$

$$mode = l_1 + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$$

$$= 25 + \frac{(30 - 18)}{2 \times 30 - 18 - 20} \times 5$$

$$= 25 + \frac{12 \times 5}{60 - 18 - 20}$$

$$= 25 + \frac{60}{22}$$

$$= 25 + 2.72$$

Mode = 27.72

Ques 41 Compute mode

Marks (mode) : 5 10 15 20 25 30 35 40 45

no. of students : 20 44 76 104 124 140 174 184 192

| Marks | $f_x$ |
|---|---|
| 0 - 5 | 20 |
| 5 - 10 | 24 |
| 10 - 15 | 32 |
| 15 - 20 | 28 |
| 20 - 25 | 20 |
| 25 - 30 | 16 |
| 30 - 35 | 34 Model class |
| 35 - 40 | 10 |
| 40 - 45 | 8 |

$c = 5$, $l_1 = 30$, $f_0 = 16$, $f_1 = 34$, $f_2 = 10$

$$\text{Mode} = l_1 + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$$

$$= 30 + \frac{(34-16) \times 5}{2 \times 34 - 16 - 10}$$

$$= 30 + \frac{18 \times 5}{68 - 26}$$

$$= 30 + \frac{90}{42}$$

$$= 30 + 2.14$$

$$= 32.14$$

HW

Ques 42 Calculate mode for the following distribution

Class : 0-10 10-20 20-30 30-40 40-50 50-60

freq : 7 15 25 24 20 9

| Class | freq. |
|---|---|
| 0 - 10 | 7 |
| 10 - 20 | 15 |
| 20 - 30 | 25 Model Class |
| 30 - 40 | 24 |
| 40 - 50 | 20 |
| 50 - 60 | 9 |

$c = 10$, $l_1 = 20$

$f_0 = 15$, $f_1 = 25$, $f_2 = 24$

$$\text{mode} = l_1 + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$$

$$= 20 + \frac{(25-15) \times 10}{2 \times 25 - 15 - 24} = 20 + \frac{100}{50 - 39}$$

$$= 20 + \frac{100}{11} = 20 + 9.09$$

$$= 29.09$$

HW

Ques 43 Calculate ~~mean~~ mean, median and mode for the following data.

Class - 0-10 10-20 20-30 30-40 40-50

freq 8 12 20 6 4

| Class | freq | $x$ | $fx$ |
|---|---|---|---|
| 0 - 10 | 8 | 5 | 40 |
| 10 - 20 | 12 | 15 | 180 |
| 20 - 30 | 20 | 25 | 500 |

Model Class

~~Mean~~ =

| | | | |
|---|---|---|---|
| 30-40 | 6 | 35 | 210 |
| 40-50 | 4 | 45 | 180 |
| | N = 50 | | $\Sigma fx = 1110$ |

Mean $= \frac{\Sigma fx}{N} = \frac{1110}{50} = 22.2$

$l_1 = 20$, $c = 10$, $f_0 = 12$, $f_1 = 20$, $f_2 = 6$

Mode $= 20 + \frac{(20-12)}{40-12-6} \times 10 = 20 + \frac{80}{22}$

Mode $= \underline{23.6}$

Mode = 3 Median − 2 Mean
23.6 = 3 Median − 2 × 22.2
23.6 + 44.4 = 3 Median

$\frac{68}{3}$ = Median

Median = 22.6

HW

Ques44 Calculate mean, median and mode

| Marks (less than) : | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| frequency : | 3 | 13 | 18 | 12 | 5 |

| Marks | f | x | fx |
|---|---|---|---|
| 0-10 | 3 | 5 | 15 |
| 10-20 | 10 | 15 | 150 |
| 20-30 | 5 | 25 | 125 |
| 30-40 | 6 | 35 | 210 |
| 40-50 | 7 | 45 | 315 |
| | N = 28 | | $\Sigma fx = 815$ |

Mean $= \frac{\Sigma fx}{N} = \frac{815}{28} = 29.10$

$l_1 = 10$, $f_0 = 3$, $f_1 = 10$, $f_2 = 5$, $c = 10$

Mode $= l_1 + \frac{(f_1 - f_0)}{2f_1 - f_0 - f_2} \times c$

$= 10 + \frac{(10-3) \times 10}{2 \times 10 - 3 - 5}$

$= 10 + \frac{70}{20-8} = 10 + \frac{70}{12}$

Mode $= \underline{45}$

Mode = 3 Median − 2 Mean
45 = 3 × Median − 2 × 29.10
45 = 3 × Median − 58.2
103.2 = 3 × Median

Median $= \frac{103.2}{3}$

Median = 34.4

HW

Ques45 Calculate mean, median and mode

| Roll no : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks : | 23 | 32 | 25 | 38 | 56 | 12 | 32 | 38 | 48 | 70 |

12, 23, 25, 32, 32, 38, 38, 48, 56, 70

Median = size of $\left(\frac{n+1}{2}\right)^{th}$ item

$=$ size of $\frac{11}{2}^{th}$ item $=$ size of $5.5^{th}$ item

$= \frac{\text{size of } 5^{th} + \text{size of } 6^{th}}{2} = \frac{32+32}{2}$

Median = 32

Mean $= \frac{\Sigma x}{n} = \frac{23+32+25+38+56+12+32+38+40+70}{10}$

$= \frac{366}{10}$ ∴ Mean = 36.6

Mode = 3 median − 2 Mean

$= 3 \times 32 - 2 \times 36.6$

$= 96 - 73.2$

Mode = 22.8

HW

Ques46 Find mean and median for the following data

| Size | 10 | 12 | 15 | 20 | 22 | 28 | 30 |
|---|---|---|---|---|---|---|---|
| freq | 3 | 9 | 12 | 25 | 18 | 7 | 6 |

| x | f | fx |
|---|---|---|
| 10 | 3 | 30 |
| 12 | 9 | 108 |
| 15 | 12 | 180 |
| 20 | 25 | 500 |
| 22 | 18 | 396 |
| 28 | 7 | 196 |
| 30 | 6 | 180 |
| | $\boxed{N = 80}$ | $\boxed{\Sigma fx}$ = 1590 |

Mean $= \frac{\Sigma fx}{N}$

$= \frac{1590}{80}$

$= \frac{159}{8}$

$= 19.875$

Mode = 20

Mode = 3 Median − 2 Mean

$20 = 3 \times \text{Median} - 2 \times 19.875$

$20 + 39.75 = 3 \times \text{Median}$

Median $= \frac{59.75}{3}$

Median = 19.916

Ques47 In a certain frequency distribution mean is equal to 30 kg, median = 27 kg. Find mode

Mode = 3 Median − 2 Mean

$= 3 \times 27 - 2 \times 30$

$= 81 - 60$

Mode = 21 kg

Ques48 In a moderatly assymmetrical distribution the mode and mean are 32.1 and 35.4 respectively. Calculate median.

Mode = 3 Median − 2 Mean

$32.1 = 3 \times \text{Median} - 2 \times 35.4$

$32.1 = 3\,\text{Median} - 70.8$

$32.1 + 70.8 = 3\,\text{Median}$

$102.9 = 3\,\text{median}$

median $= \frac{102.9}{3} = 34.3$

Ques 9 For a freq. distribution median = 132.8, mode = 141.3. Find mean

Mode = 3 median − 2 Mean

141.3 = 3 × 132.8 − 2 Mean

141.3 = 398.4 − 2 Mean

~~398.4~~

2 Mean = 398.4 − 141.3

Mean = $\frac{257.1}{2}$

Mean = 128.55

## PARTITION VALUES (Median, Quartiles, Deciles, Percentiles)

Quartiles are values of the variable which divide the data into four equal parts. There are three quartiles known as $Q_1$, $Q_2$, $Q_3$.

Deciles divide the data into ten equal parts and percentiles divide the data into 100 equal parts

There are 9 deciles and 99 percentiles. The deciles are denoted by $D_1, D_2 \dots D_9$. The percentiles are denoted by $P_1, P_2 \dots P_{99}$. $D_5$ and $P_{50}$ are the median.

Partition values in individual series

Median = size of $\frac{(n+1)}{2}^{th}$ item

$Q_1$ = size of $\frac{(n+1)}{4}^{th}$ item

$Q_3$ = size of $\left(\frac{(n+1)}{4} \times 3\right)^{th}$ item

$K^{th}$ decile = $D_K$ = size of $\left(\frac{(n+1)}{10} \times K\right)^{th}$ item

$r^{th}$ Percentile $P_r$ = size of $\left(\frac{n+1}{100} \times r\right)^{th}$ item

When the items are arranged in the ascending order of magnitude, n stands for number of values.

Ques 1 Find median and quartile from the following values ~~44~~, 33, 37, 30, 47, 60, 87, 15, 30, 45, 43, 44

15, 30, 30, 33, 37, 43, 44, 45, 47, 60, 87

Median = size of $\left(\frac{n+1}{2}\right)^{th}$ item

= size of $\frac{(11+1)}{2}^{th}$ item

= size of $6^{th}$ item

Median = 43

$Q_1 = \text{size of } \left(\frac{n+1}{4}\right)^{th} \text{ item}$

$= \text{size of } 3rd \text{ item}$

$= 30$

$Q_3 = \text{size of } \left(\frac{n+1}{4} \times 3\right)^{th} \text{ item}$

$= \text{size of } 9^{th} \text{ item}$

$= 49$

Ques 2 Compute 6th decile and 40th percentile of the following values

52, 89, 101, 23, 95, 43, 55, 15, 65

15, 23, 43, 52, 55, 65, 89, 95, 101

$n = 9$

$6^{th} \text{ decile} = D_6 = \text{size of } \left(\frac{n+1}{10}\right) \times 6^{th} \text{ item}$

$= \text{size of } \left(\frac{10}{10} \times 6^{th}\right) \text{ item}$

$= \text{size of } 6^{th} \text{ item}$

$6^{th} \text{ decile} = 65$

$40^{th} \text{ percentile} = P_{40} = \text{size of } \left(\frac{n+1}{100} \times 40\right)^{th} \text{ item}$

$= \text{size of } \left(\frac{10}{100} \times 40\right)^{th} \text{ item}$

$= \text{size of } 4^{th} \text{ item}$

$40^{th} \text{ percentile} = 52$

Ques 3 Compute quartiles, 3rd decile and 80th percentile of the following data :- 28, 30, 13, 15, 14, 34, 50, 90, 15, 21.

13, 14, 15, 15, 21, 28, 30, 34, 50, 90

quartiles

$n = 10$

$Q_1 = \text{size of } \left(\frac{n+1}{4}\right)^{th} \text{ item}$

$= \text{size of } \left(\frac{9+1}{4}\right)^{th} \text{ item}$

$= \text{size of } \left(\frac{11}{4}\right)^{th} \text{ item}$

$= \text{size of } 2.75^{th} \text{ item}$

$= 2nd \text{ item} + .75 \,(3rd \text{ item} - 2nd \text{ item})$

$= 2nd \text{ item} + .75 \,(15 - 14)$

$= 14 + .75 \times 1$

$= 14.75$

$2 \overline{)11} \quad 4 \overline{)11}$

$20$

$2.25$

(Median) OR

$Q_2 = \text{size of } \left(\frac{n+1}{2}\right)^{th} \text{ item}$ $\quad n = 10$

$= \text{size of } \frac{11}{2}^{th} \text{ item}$

$= \text{size of } 5.5^{th} \text{ item}$

$= \frac{\text{size of } 5^{th} + 6^{th} \text{ item}}{2}$

$= \frac{21 + 28}{2}$

$= 24.5$

$Q_3$ = size of $\left(\frac{n+1}{4} \times 3\right)^{th}$ item

= size of $\left(\frac{11}{4} \times 3\right)^{th}$ item

= size of $\frac{33}{4}$ th item

= size of $8.25^{th}$ item

= $8^{th}$ item + .25 ($9^{th}$ item − $8^{th}$ item)

= 34 + .25 (50 − 34)

= 34 + .25 × 16

= 34 + 4

= $\underline{\underline{38}}$

$80^{th}$ percentile = size of $\left(\left(\frac{n+1}{100}\right) \times 80\right)$ item

= size of $\left(\frac{11}{100} \times 80\right)^{th}$ item

= size of $8.8^{th}$ item

= $8^{th}$ item + 0.8 ($9^{th}$ item − $8^{th}$ item)

= 34 + 0.8 (50 − 34)

= 34 + 0.8 × 16

= 34 + 12.8

= $\underline{\underline{46.8}}$

$D_3$ = size of $\left(\frac{n+1}{10} \times 3\right)^{th}$ item

= size of $\frac{33}{10}$ th item

= size of $3.3^{th}$ item

= 3rd item + 0.3 ($4^{th}$ item − 3rd item)

= 15 + 0.3 (15 − 15)

= $\underline{\underline{15}}$

HW

Ques Find median, $Q_1$ and $Q_3$ from the values

14, 28, 16, 35, 58, 29, 48, 60, 12, 14, 15

12, 14, 14, 15, 16, 28, 29, 35, 48, 58, 60

$n = 11$

median = size of $\left(\frac{n+1}{2}\right)^{th}$ item

= size of $6^{th}$ item = $\underline{\underline{28}}$

$Q_1$ = size of $\left(\frac{n+1}{4}\right)^{th}$ item

= size of 3rd item = $\underline{\underline{14}}$

$Q_3$ = size of $\left(\frac{n+1}{4} \times 3\right)^{th}$ item

= size of $\left(\frac{12}{4} \times 3\right)^{th}$ item

= size of $9^{th}$ item

= $\underline{\underline{48}}$

The values are
median = 28
$Q_1$ = 14
$Q_3$ = 48

HCW
Ques Find $Q_1$, $Q_3$, $P_{60}$, $D_4$ from the values
20, 52, 60, 10, 25, 72, 70, 55, 60, 70

$Q_1$ = size of $\left(\frac{n+1}{4}\right)^{th}$ item

10, 20, 25, 52, 55, 60, 60, 70, 70, 72

n = 10

$Q_1$ = size of $\left(\frac{n+1}{4}\right)^{th}$ item

= size of $\left(\frac{11}{4}\right)^{th}$ item

= $2.75^{th}$ item
= 2nd item + 0.75(3rd - 2nd item)
= 20 + 0.75 × (25 - 20)
= 20 + 3.75
$Q_1$ = 23.75

$Q_3$ = size of $\left(\frac{n+1}{4} \times 3\right)^{th}$ item

= size of $\left(\frac{11}{4} \times 3\right)^{th}$ item

= size of $\left(\frac{33}{4}\right)^{th}$ item

= size of $8.25^{th}$ item
= $8^{th}$ item + 0.25($9^{th}$ - $8^{th}$ item)
= 70 + 0.25(70 - 70)
= 70

$P_{60}$ = size of $\left(\frac{n+1}{100} \times 60\right)^{th}$ item

= size of $\left(\frac{11}{100} \times 60\right)^{th}$ item

= size of $6.6^{th}$ item

= $6^{th}$ item + 0.6($7^{th}$ - $6^{th}$ item)
= 60 + 0.6(60 - 60)
= 60

$D_4$ = size of $\left(\frac{n+1}{10} \times 4\right)^{th}$ item

= size of $\left(\frac{44}{10}\right)^{th}$ item

= size of $4.4^{th}$ item

= $4^{th}$ item + 0.4($5^{th}$ - $4^{th}$ item)

= 52 + 0.4(55 - 52)
= 52 + 0.4 × 3
= 52 + 1.2
= 53.2

## ~~Ques~~ Partition Values in Discrete Series

$Q_1$ = size of $\left(\frac{N+1}{4}\right)^{th}$ item

$Q_3$ = size of $\left(\frac{N+1}{4} \times 3\right)^{th}$ item

$D_K$ = size of $\left(\frac{N+1}{10} \times K\right)^{th}$ item

$P_r$ = size of $\left(\frac{N+1}{100} \times r\right)^{th}$ item

where $N = \Sigma f$ (total frequency)

Que 6 Find quartiles, 9th decile and 30th percentile

Size: 4 6 10 12 18 20 25

freq: 3 7 10 15 8 6 4

| size | f | cf | |
|---|---|---|---|
| 4 | 3 | 3 | |
| 6 | 7 | 10 | |
| 10 | 10 | 20 | $Q_1$ |
| 12 | 15 | 35 | |
| 18 | 8 | 43 | $Q_3$ |
| 20 | 6 | 49 | $D_9$ |
| 25 | 4 | 53 | |
| | $\boxed{N = 53}$ | | |

$N = 53$

$Q_1$ = size of $\left(\frac{N+1}{4}\right)^{th}$ item

= size of $\frac{54}{4}$th item

= size of $13.5^{th}$ item

= $\underline{\underline{10}}$

$Q_3$ = size of $\left(\frac{54}{4} \times 3\right)^{th}$ item

= size of $40.5^{th}$ item

= $\underline{\underline{18}}$

$D_9$ = size of $\left(\frac{54}{10} \times 9\right)^{th}$ item

= size of $48.6^{th}$ item

= $\underline{\underline{20}}$

$P_{30}$ = size of $\left(\frac{54}{100} \times 30\right)^{th}$ item

= size of $16.2^{th}$ item

= $\underline{\underline{10}}$

## Partition Values in Continous Series

$Q_1$ = size of $\left(\frac{N}{4}\right)^{th}$ item

$Q_3$ = size of $\left(\frac{N}{4} \times 3\right)^{th}$ item

$D_K$ = size of $\left(\frac{N}{10} \times K\right)^{th}$ item

$P_r$ = size of $\left(\frac{N}{100} \times r\right)^{th}$ item

where $N = \Sigma f$ (total frequency)

for continous series the following interpolation formula are applied

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times c$$

$$Q_3 = l_1 + \frac{\left(\left(\frac{3N}{4}\right) - cf\right)}{f} \times c$$

$$D_k = l_1 + \frac{\left(\frac{kN}{10} - cf\right)}{f} \times c$$

$$P_k = l_1 + \frac{\left(\frac{k \times N}{100} - cf\right)}{f} \times c$$

where $l_1$ = lower limit of the respective class

$cf$ = cumulative frequency of the class just preceeding

$f$ = frequency of the respective class

$c$ = class interval of the respective class

Que Obtain the quadriles, D7 and P90 for the data below

| Age | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| Frequency | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 125 |

| Size | f | cf |
|---|---|---|
| 0-10 | 15 | 15 |
| 10-20 | 30 | 45 |
| 20-30 | 53 | 98 cf |
| 30-40 | 75 | 173 $Q_1$ |
| 40-50 | 100 | 273 |
| 50-60 | 110 | 383 cf |
| 60-70 | 115 | 498 $Q_3$ $D_7$ |
| 70-80 | 125 | 623 $P_{90}$ |
| | N = 623 | |

$Q_1$ = size of $\left(\frac{N}{4}\right)^{th}$ item

= size of $\frac{623}{4}$ th item

= size of $155.75^{th}$ item

$l_1 = 30$, $\frac{N}{4} = 155.75$, $cf = 98$, $f = 75$, $c = 10$

$$Q_1 = 30 + \frac{(155.75 - 98)}{75} \times 10$$

$= 30 + \cancel{57.75}\ \frac{577.5}{75}$

$= 30 + 7.7$

$Q_1 = 37.7$

$Q_3$ = size of $\left(\frac{3N}{4}\right)$ th item

= size of $\left(\frac{623 \times 3}{4}\right)$ th item

$= \text{size of } \left(\frac{1869}{4}\right)^{th} \text{ item}$

$= \text{size of } 467.25^{th} \text{ item}$

$l_1 = 60, \; \frac{N}{4} = 155.75 \quad cf = 383 \quad f = 115, \; c = 1$

$$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times c$$

$$= 60 + \frac{\left(\frac{1869}{4} - 383\right)}{115} \times 10$$

$$= 60 + \frac{(467.25 - 383)}{115} \times 10$$

$$= 60 + \frac{\cancel{84.25}\; 842.5}{115}$$

$$= 60 + 7.326$$

$$Q_3 = 67.326$$

$$D_7 = \text{size of } \left(\frac{7N}{10}\right)^{th} \text{ item}$$

$$= \text{size of } \left(\frac{7 \times 623}{10}\right)^{th} \text{ item}$$

$$= \text{size of } 436.1^{th} \text{ item}$$

$l_1 = 60, \; \frac{7N}{10} = 436.1 \quad cf = 383, \; f = 115 \quad c = 10$

$$D_7 = l_1 + \frac{\left(\frac{KN}{10} - cf\right)}{f} \times c$$

$$= 60 + \frac{436.1 - 383}{115} \times 10$$

$$= 60 + \frac{531}{115}$$

$$= \cancel{60 + 0.461}$$

$$= \cancel{60 = 461}\; 60 + 4.617$$

$$= 64.617$$

$$P_{90} = \text{size of } \left(\frac{90N}{100}\right)^{th} \text{ item}$$

$$= \text{size of } \left(\frac{90 \times 623}{100}\right)^{th} \text{ item}$$

$$= \text{size of } \left(\frac{56070}{100}\right)^{th} \text{ item}$$

$$= \text{size of } 560.7^{th} \text{ item}$$

$l_1 = 70, \; cf = 498, \; f = 125, \; c = 10, \; \frac{90N}{100} = 560.7$

$$P_{90} = l_1 + \frac{\left(\frac{90N}{100} - cf\right)}{f} \times c$$

$$= 70 + \frac{(560.7 - 498)}{125} \times 10$$

$$= 70 + \frac{627}{125}$$

$$= 70 + 5.016$$

$$P_{90} = 75.016$$

Ques From the following distribution of marks of 199 students in an examination, find out 15th percentile and 4th decile

| Marks | 0-20 | 21-30 | 31-40 | 41-50 | 51-60 | 61-70 | 71- |
|---|---|---|---|---|---|---|---|
| No. of students | 21 | 19 | 60 | 42 | 24 | 18 | 15 |

| Marks | Freq | cf |
|---|---|---|
| -0.5-20.5 | 21 | 21 |
| 20.5-30.5 | 19 | 40 |
| 30.5-40.5 | 60 | 100 |
| 40.5-50.5 | 42 | 142 |
| 50.5-60.5 | 24 | 166 |
| 60.5-70.5 | 18 | 184 |
| 70.5-80.5 | 15 | 199 |
| | N = 199 | |

N = 199

$\frac{N}{2} = \frac{199}{2} = 99.5$

15th Percentile = size of $\left(\frac{N}{100} \times r\right)^{th}$ item

= size of $\left(\frac{199}{100} \times 15\right)^{th}$ item

= size of $29.85^{th}$ item

$l_1 = 20.5$ $c = 10$ $cf = 21$ $f = 19$

$P_{15} = l_1 + \frac{\left(\frac{r \times N}{100} - cf\right)}{f} \times c$

$= 20.5 + \left(\frac{29.85 - 21}{19}\right) \times 10$

$= 20.5 + \frac{88.5}{19}$

$= 20.5 + 4.658\cancel{5947}$

$P_{15} = 25.158$

4th decile = size of $\left(\frac{N}{10} \times k\right)^{th}$ item

= size of $\left(\frac{199}{10} \times 4\right)^{th}$ item

= size of $79.6^{th}$ item

$l_1 = 30.5$ $f = 60$, $cf = 40$, $c = 10$

$D_4 = l_1 + \frac{\left(\frac{kN}{10} - cf\right)}{f} \times c$

$= 30.5 + \left(\frac{79.6 - 40}{60}\right) \times 10$

$= 30.5 + \frac{396}{60}$

$= 30.5 + 6.6$

$D_4 = 37.1$

16/03/2024
Tuesday

## Measures of Dispersion (measures of Variability)

Measures of dispersion are the statistical devices to measure the variability in a series. They tell us the extend to which the values of a series differ between each other or from their average.

They are classified into
① Absolute measures
② Relative measures

**Absolute measures of dispersion**

It includes ① range, ② quartile deviation, ③ Mean deviation and ④ Standard deviation.

**Relative measures of dispersions**

It includes ① coefficient of range, ② coefficient of quartile deviation, ③ Coefficient of Mean deviation and ④ coefficient of variation.

**Range**

It is the difference between highest and the lowest value in a series so

Range = Highest value - Lowest value
$= H - L$

Coefficient of range $= \dfrac{H-L}{H+L}$

**Range in individual series**

Ques 1 Find the range and coefficient of range for the following values.
25, 32, 85, 32, 42, 10, 20, 18, 28

~~10, 20, 25, 32, 32, 42~~

10, 18, 20, 25, 28, 32, 32, 42, 85

Range $= H - L$
$= 85 - 10$
$= 75$

Coefficient of range $= \dfrac{H-L}{H+L} = \dfrac{75}{95}$
$= 0.789$

Ques 2 Compare the two series for their variability

Series A 5 6 8 19 12 16 19 21 25
Series B 25 30 30 35 40 45

Series A - 5, 6, 8, 12, 16, 19, 19, 21, 25

Coefficient of range for series A
$= \dfrac{25-5}{25+5} = \dfrac{20}{30} = \dfrac{2}{3} = 0.6666$

Series B – 25, 30, 30, 35, 40, 45

Coefficient of range for Series B

$= \frac{45-25}{45+25} = \frac{20}{70} = \frac{2}{7}$

$= 0.285$

Series A is more variable than Series B (since coefficient of range is more in A)

Ques
Range in Discrete Series

Ques3 Find range and its coefficient

| Weight | 5 | 8 | 10 | 12 | 25 | 30 | 38 |
|---|---|---|---|---|---|---|---|
| No. of children | 2 | 3 | 8 | 10 | 9 | 3 | 2 |

don't consider the value

Range = 38 − 5
= 33

Coefficient of range $= \frac{38-5}{38+5} = \frac{33}{43}$

$= 0.767$

$= 0.77$

Range in continuous series

Ques 4 Find range and coefficient of range

| Marks | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|
| no. of students | 5 | 10 | 12 | 8 | 4 |

Range = 60 − 10 = 50

Coefficient of range $= \frac{60-10}{60+10} = \frac{50}{70} = \frac{5}{7}$

$= 0.7142$
$= 0.714$

Quatrile Deviation (Semi Inter Quatrile Range)

Quatrile deviation is defined as half the distance between the third and first quatriles

ie Quatrile deviation $= \frac{Q_3 - Q_1}{2}$

Coefficient of Dispersion (quatrile deviation) $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Individual Series

$Q_1$ = size of $\left(\frac{n+1}{4}\right)$th item

when the values are arranged

in the ascending order of their magnitude

$Q_1$ = size of $\left(\frac{n+1}{4} \times 3\right)^{th}$ item

~~Qd = size of n~~

$Qd = \frac{Q_3 - Q_1}{2}$

Ques 5 Find quatrile deviation for the following value:

28, 32, 25, 42, 55, 82, 10, 25, 40, 38, 39

10, 25, 25, 28, 32, 38, 39, 40, 42, ~~82~~ 55, 82

$n = 11$

Quatrile deviation = $\frac{Q_3 - Q_1}{2}$

$Q_3$ = size of $\left(\frac{11+1}{4} \times 3\right)^{th}$ item

= size of 9th item

= 42

$Q_1$ = size of $\left(\frac{n+1}{4}\right)^{th}$ item

= size of 3th item

= 25

$Qd = \frac{42 - 25}{2} = \frac{17}{2}$

$Qd = 8.5$

NOTE

Inter quatrile range = $Q_3 - Q_1$

H.W

Ques 6 Compute quatrile measure of dispersion, inter quatrile range and coefficient of quatrile deviation for the following values

23, 25, 8, 10, 9, 29, 45, 85, 10, 16.

8, 9, 10, 10, 16, 23, 25, 29, 45, 85

~~Quatrile deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$~~

$n = 10$

$Q_3$ = size of $\left(\frac{n+1}{4} \times 3\right)^{th}$ item

= size of $\frac{33}{4}$ th item

= size of 8.25 th item

= ~~0.25~~ 8th item + 0.25(9th − 8th item)

= 29 + 0.25(45 − 29)

= 29 + 0.25 × 16

= 29 + 4

$Q_3 = 33$

$Q_1 = \text{size of } \left(\frac{n+1}{4}\right)^{th} \text{ item}$

$= \text{size of } \left(\frac{11}{4}\right)^{th} \text{ item}$

$= \text{size of } 2.75^{th} \text{ item}$

$Q_1 = \text{2nd item} + 0.75\,(\text{3rd} - \text{2nd item})$
$= 9 + 0.75(10 - 9)$
$= 9 + 0.75$

$Q_1 = 9.075$

Quatrile deviation $= \frac{Q_3 - Q_1}{2}$

$= \frac{33 - 9.075}{2} = \frac{23.925}{2}$

$Qd = 11.9625$

Coefficient of quatrile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{33 - 9.075}{33 + 9.075}$

$= \frac{23.925}{42.075}$

$= 0.5686274$

Interquatrile range $= Q_3 - Q_1$
$= 33 - 9.075$
$= 23.925$

Ques Using quatrile compare the following two series and state which is more variable

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| series I | 5 | 10 | 27 | 90 | 38 | 56 | 29 | 43 | 39 | 86 | 30 |
| series II | 10 | 27 | 15 | 35 | 89 | 72 | 28 | 40 | 45 | 28 | 39 |

Ser I : 5, 10, 27, 29, 30, 38, 39, 43, 56, 86, 90

$Q_1 = \text{size of } \left(\frac{n+1}{4}\right)^{th} \text{ term}$

$= \text{size of } \frac{12}{4}^{th} \text{ term} = \text{size of 3rd term}$

$Q_1 = 27$

$Q_3 = \text{size of } \left(\frac{n+1}{4} \times 3\right)^{th} \text{ term}$

$= \text{size of } \left(\frac{12}{4} \times 3\right)^{th} \text{ term}$

$Q_3 = \text{size of } 9^{th} \text{ term} = 56$

$QD = \frac{Q_3 - Q_1}{2} = \frac{56 - 27}{2} = \frac{29}{2} = 14.5$

Coefficient of Q.D. = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{56 - 27}{56 + 27} = \frac{29}{83}$

$= 0.3493$

Ser II : 10, 15, 27, 28, 28, 35, ~~38~~, 39, 40, ~~56~~, 45, ~~86~~, ~~90~~ 72, 89

$Q_1$ = size of $\left(\frac{n+1}{4}\right)^{th}$ term  $n = 11$

= size of $\left(\frac{12}{4}\right)^{th}$ term = size of 3rd term

$Q_1 = 27$

$Q_3$ = size of $\left(\frac{n+1}{4} \times 3\right)^{th}$ term

= size of $\left(\frac{12}{4} \times 3\right)^{th}$ term = size of 9th term

$Q_3 = 45$

Coefficient of QD = $\frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{45 - 27}{45 + 27} = \frac{18}{72}$

$= 0.25$

Series I is more variable than series II

Quartile deviation in Discrete Series

$Q_1$ = size of $\left(\frac{N+1}{4}\right)^{th}$ item

$Q_3$ = size of $\left(\frac{N+1}{4} \times 3\right)^{th}$ item,

where $N = \Sigma f$

$QD = \frac{Q_3 - Q_1}{2}$

Ques 8 Find quartile deviation

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| size | 5 | 8 | 10 | 12 | 19 | 20 | 32 |
| freq | 3 | 10 | 15 | 20 | 8 | 7 | 6 |

| size | freq | cf |
|---|---|---|
| 5 | 3 | 3 |
| 8 | 10 | 13 |
| 10 $Q_1$ | 15 | 28 |
| 12 | 20 | 48 |
| 19 $Q_3$ | 8 | 56 |
| 20 | 7 | 63 |
| 32 | 6 | 69 |

$\boxed{N = 69}$

$Q_1$ = size of $\left(\frac{N+1}{4}\right)^{th}$ item

= size of $\left(\frac{70}{4}\right)^{th}$ item

= size of $17.5^{th}$ item

$Q_1 = 10$

$Q_3 =$ size of $\left(\frac{N+1}{4} \times 3\right)^{th}$ item

$=$ size of $\left(\frac{30}{4} \times 3\right)^{th}$ item

$=$ size of 22.5 th item

$Q_3 = 19$

$Q_D = \frac{19-10}{2} = \frac{9}{2} = 4.5$

Quartile deviation in continous series

$Q_1 =$ size of $\left(\frac{N}{4}\right)^{th}$ item

$Q_3 =$ size of $\left(\frac{N}{4} \times 3\right)^{th}$ item

where $N = \Sigma f$

The values of the quatriles are obtained by the interpolation formulae

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times c$$

$$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times c$$

where $l_1$ is the lower limit of the quatrile class

$f$ is the frequency of the class

$cf$ is the cumulative frequency of the preceeding class

$c$ is the class interval of the quatrile class.

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

Ques 9 Obtain the quatrile measure of dispersion for the d. and its coefficient for the data given below.

| Age | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| no. of person | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 125 |

| Age | freq | cf |
|---|---|---|
| 0-10 | 15 | 15 |
| 10-20 | 30 | 45 |
| 20-30 | 53 | 98 |
| 30-40 | 75 | 173 |
| 40-50 | 100 | 273 |
| 50-60 | 110 | ~~380~~ 383 |
| 60-70 | 115 | 498 |
| 70-80 | 125 | 623 |
| | N = 623 | |

$Q_1$ = size of $\left(\frac{N}{4}\right)^{th}$ item

= size of $\left(\frac{623}{4}\right)^{th}$ item

= size of $155.75^{th}$ item

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - cf\right)}{f} \times c$$

$$= 30 + \frac{\left(\frac{623}{4} - 98\right)}{75} \times 10$$

$$= 30 + \frac{(155.75 - 98)}{75} \times 10$$

$$= 30 + \frac{577.5}{75}$$

$$= 30 + 7.7$$

$$Q_1 = 37.7$$

$Q_3$ = size of $\left(\frac{N}{4} \times 3\right)^{th}$ item

= size of $\left(\frac{623 \times 3}{4}\right)^{th}$ item

= size of $\left(\frac{1869}{4}\right)^{th}$ item

= size of $467.25^{th}$ item

$$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - cf\right)}{f} \times c$$

$$= 60 + \frac{(467.25 - 383)}{115} \times 10$$

$$= 60 + \frac{842.5}{115}$$

$$= 60 + 7.327$$

$$= 67.327$$

$$QD = \frac{Q_3 - Q_1}{2}$$

$$= \frac{67.327 - 37.7}{2}$$

$$= \frac{29.627}{2}$$

$$= 14.8135$$

$$QD = 14.8$$

Coefficient of Q.D $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{29.627}{105.027}$$

$$= 0.2820$$

Ques 10 Find Q.D from the following data

| wages (in Rs) (above) | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| No. of workers | 600 | 580 | 490 | 340 | 240 | 170 | 120 | 70 | 25 |

| Class | freq. | cf |
|---|---|---|
| 10-20 | 20 | 20 |
| 20-30 | 90 | 110 |
| 30-40 | 150 | 260 |
| 40-50 | 100 | 360 |
| 50-60 | 70 | 430 |
| 60-70 | 50 | 480 |
| 70-80 | 50 | 530 |
| 80-90 | 45 | 575 |
| 90-100 | 25 | 600 |
| | $\boxed{N=600}$ | |

$Q_1 = \text{size of } \left(\frac{N}{4}\right)^{th} \text{ item}$

$= \text{size of } \left(\frac{600}{4}\right)^{th} \text{ item}$

$= \text{size of } 150^{th} \text{ item}$

$Q_1 = l_1 + \left(\frac{N/4 - cf}{f}\right) \times c$

$= 30 + \left(\frac{150 - 110}{150}\right) \times 10$

$= 30 + \frac{400}{150}$

$= 30 + 2.66$

$= \cancel{30}\ 32.66$

$Q_3 = \text{size of } \left(\frac{3N}{4}\right)^{th} \text{ item}$

$= \text{size of } \left(\frac{3 \times 600}{4}\right)^{th} \text{ item}$

$= \text{size of } 450^{th} \text{ item}$

$Q_3 = l_1 + \left(\frac{\frac{3N}{4} - cf}{f}\right) \times c$

$= 60 + \left(\frac{450 - 430}{50}\right) \times 10$

$= 60 + \frac{200}{50}$

$= 60 + 4$

$Q_3 = 64$

$QD = \frac{Q_3 - Q_1}{2} = \frac{64 - 32.66}{2}$

$= \frac{31.34}{2}$

$QD = 15.67$

$QD = 15.7$

## Mean Deviation

It is defined as the arithmetic mean of deviations of all the values in a series from their average, counting all such deviations as positive.
The average selected maybe mean, median or mode

Mean deviation = $\frac{\Sigma \text{ absolute value of } d}{n}$

$= \frac{\Sigma |d|}{n}$

where $|d|$ represents deviation from average without sign

Coefficient of mean deviation = $\frac{\text{Mean deviation}}{\text{Average from which mean deviation is computed}}$

### Mean deviation for Individual series

Que Find mean deviation from mean and its coefficient for its following values
25, 63, 85, 75, 62, 70, 83, 28, 30, 12

Mean = $\frac{\Sigma x}{n}$

$= \frac{533}{10}$

Mean = 53.3

| | $\lvert d \rvert$ |
|---|---|
| 25 | $\lvert 53.3 - 25 \rvert = 28.3$ |
| 63 | $\lvert 53.3 - 63 \rvert = 9.7$ |
| 85 | $\lvert 53.3 - 85 \rvert = 31.7$ |
| 75 | $\lvert 53.3 - 75 \rvert = 21.7$ |
| 62 | $\lvert 53.3 - 62 \rvert = 8.7$ |
| 70 | $\lvert 53.3 - 70 \rvert = 16.7$ |
| 83 | $\lvert 53.3 - 83 \rvert = 29.7$ |
| 28 | $\lvert 53.3 - 28 \rvert = 25.3$ |
| 30 | $\lvert 53.3 - 30 \rvert = 23.3$ |
| 12 | $\lvert 53.3 - 12 \rvert = 41.3$ |

$\Sigma |d| = 236.4$

Mean deviation from mean = $\frac{\Sigma |d|}{n} = \frac{236.4}{10}$

$= 23.64$

Coefficient of mean deviation = $\frac{\text{Mean Deviation}}{\text{Mean}}$

$$= \frac{23.64}{53.3}$$

Coefficient of mean deviation $= 0.44$

Ques Calculate mean deviation from median and its coefficient for the following data.
5, 28, 33, 44, 83, 87, 96, 99, 25, 35, 82

5, 25, 28, 33, 35, 44, 82, 83, 87, 96, 99

$n = 11$

Median = size of $\left(\frac{n+1}{2}\right)$ th item

= size of $\frac{12}{2}$ th item

= size of 6th item

Median $= 44$

| freq | \|d\| |
|---|---|
| 5 | 39 |
| 28 | 16 |
| 33 | 11 |
| 44 | 0 |
| 83 | 39 |
| 87 | 43 |
| 96 | 52 |
| 99 | 55 |
| 25 | 19 |
| 35 | 9 |
| 82 | 38 |

$\Sigma d = 321$

Mean deviation $= \frac{\Sigma d}{n}$

$= \frac{321}{11}$

$\doteq 29.18$

Coefficient of mean deviation = $\frac{\text{Mean deviation}}{\text{median}}$

$= \frac{29.18}{44}$

$= 0.663$

Ques With median as the base calculate the mean deviation and compare the variability of the two series A and B

Ser A 3464 4572 4124 3682 5624 4388 3680 4308

Ser B 487 508 620 382 408 266 186 218

Ser A 3484, 3680, 3682, 4124, 4308, 4388, ~~4570~~ 4572, 5624

$n = 8$

Median = size of $\left(\frac{n+1}{2}\right)$th term

= size of $\left(\frac{9}{2}\right)$th term

= size of 4.5th term

= 4 term + 0.5 (5th − 4th term)

= 4124 + 0.5 (4308 − 4124)

= 4124 + 0.5 (184)

= 4124 + 92

= 4216

Ser B - 186, 218, 266, 382, 408, 487, 508, 620

median = size of $\left(\frac{n+1}{2}\right)$th item $n = 8$

= size of 4.5th item

= 4th item + 0.5 (5th − 4th item)

= 382 + 0.5 (408 − 382)

= 382 + 0.5 × 26

= 382 + 13

= 395

Series A

| $x$ | $\lvert d \rvert$ |
|---|---|
| 3484 | 732 |
| 3680 | 536 |
| 3682 | 534 |
| 4124 | 92 |
| 4308 | 92 |
| 4388 | 172 |
| 4572 | 356 |
| 5624 | 1408 |

$\Sigma \lvert d \rvert = 3922$

$N = 8$

Mean deviation = $\frac{\Sigma \lvert d \rvert}{N} = \frac{3922}{8}$

$= 490.25$

Coefficient of mean deviation = $\frac{\text{Mean deviation}}{\text{median}}$

$= \frac{490.25}{4216}$

$= 0.116$

Ser B

| $x$ | $\lvert d \rvert$ |
|---|---|
| 186 | 209 |
| 218 | 177 |
| 266 | 129 |
| 382 | 13 |
| 408 | 13 |

| | |
|---|---|
| 487 | 92 |
| 508 | 113 |
| 620 | 225 |

$\sum|d| = 971$  $N = 8$

Mean deviation $= \frac{\sum|d|}{N} = \frac{971}{8}$

$= 121.375$

Coefficient of mean deviation $= \frac{\text{mean deviation}}{\text{mean}}$

$= \frac{121.375}{971}$

$= 0.307$

Series B is more variable than Series A since coefficient of mean deviation is greater for B

Que Compute mean deviation about mode for the following values
5, 86, 92, 45, 36, 26, 35, 45, 36, 85, 36

~~5, 36, 52~~

mode = 36

$n = 11$

| $x$ | $\lvert d\rvert$ |
|---|---|
| 5 | 31 |
| 86 | 50 |
| 92 | 56 |
| 45 | 9 |
| 36 | 0 |
| 26 | 10 |
| 35 | 1 |
| 45 | 9 |
| 36 | 0 |
| 85 | 49 |
| 36 | 0 |

$\sum|d| = 215$

Mean deviation $= \frac{215}{11}$

$= 19.54$

Coeff of mean deviation $= \frac{\text{mean deviation}}{\text{mode}}$

$= \frac{19.54}{36}$

$= 0.5428$

22/07/2019
Monday

## Mean deviation for Discrete Series

for discrete series,

$$\text{mean deviation} = \frac{\Sigma f|d|}{N}$$

Ques Compute mean deviation about mean and its coefficient for the following data

| No of children | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| no of families | 171 | 82 | 50 | 25 | 13 | 7 | 2 |

| x | f | fx | d = x - 1.02 | Σf\|d\| |
|---|---|---|---|---|
| 0 | 171 | 0 | 1.02 | 174.42 |
| 1 | 82 | 82 | 0.02 | 1.64 |
| 2 | 50 | 100 | 0.98 | 49 |
| 3 | 25 | 75 | 1.98 | 49.5 |
| 4 | 13 | 52 | 2.98 | 38.74 |
| 5 | 7 | 35 | 3.98 | 27.86 |
| 6 | 2 | 12 | 4.98 | 9.96 |
| N = 350 | | Σfx = 356 | Σd = | 351.12 |

$$\text{Mean} = \frac{\Sigma fx}{n} = \frac{356}{350} = 1.017$$

$$= 1.02$$

$$\text{Mean deviation about mean} = \frac{\Sigma f|d|}{N}$$

$$= \frac{351.12}{350}$$

$$= 1.0032$$

$$\text{Coefficient} = \frac{\text{Mean Deviation}}{\text{Mean}} = \frac{1}{1.02}$$

$$= 0.98$$

Ques Compute mean deviation about median and the coefficient of mean deviation for the frequency distribution given below

| Size: | 5 | 8 | 13 | 20 | 25 | 30 | 40 |
|---|---|---|---|---|---|---|---|
| freq: | 2 | 10 | 20 | 35 | 18 | 7 | 5 |

median:

~~5, 8, 13, 20, 25, 30, 40~~

2, 10, 20, 35, 18, 7, 5

N = 97

$$\text{median} = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ term}$$

$$= \cancel{\frac{8^{th}\text{ term}}{2}} = \frac{98}{2}^{th} \text{ term}$$

$$= \cancel{4^{th}\text{ term}} = 49^{th} \text{ term}$$

$$= \cancel{20} \qquad = \frac{20}{2}$$

| x | f | cf | \|d\| = x - 20 | f\|d\| ~~20th term~~ |
|---|---|---|---|---|
| 5 | 2 | 2 | 15 | 30 |
| 8 | 10 | 12 | 12 | 120 |
| 13 | 20 | ~~30~~ 32 | 7 | 140 |
| 20 | 35 | 67 | 0 | 0 |
| 25 | 18 | 85 | 5 | 90 |
| 30 | 7 | 92 | 10 | 70 |
| 40 | 5 | 97 | 20 | 100 |
| N = 97 | | | | 550 |

(4)

Sol: Mean deviation about median $= \frac{\Sigma f|d|}{N} = \frac{550}{97}$

$= 5.670$

Coefficient of mean deviation

$= \frac{\text{mean deviation}}{\text{median}}$

$= \frac{5.670}{20}$

$= 0.28$

Mean deviation in continuous series

Que: Find the mean deviation about the arithmetic mean for the following frequency distribution of marks of 60 college students.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| Freq | 4 | 6 | 10 | 20 | 10 | 6 | 4 |

| $x$ | $f$ | $fx$ | $\|d\| = x-35$ | $f\|d\|$ |
|---|---|---|---|---|
| 5 | 4 | 20 | 30 | 120 |
| 15 | 6 | 90 | 20 | 120 |
| 25 | 10 | 250 | 10 | 100 |
| 35 | 20 | 700 | 0 | 0 |
| 45 | 10 | 450 | 10 | 100 |
| 55 | 6 | 330 | 20 | 120 |
| 65 | 4 | 260 | 30 | 120 |
| N = 60 | | $\Sigma fx = 2100$ | | $\Sigma f\|d\| = 680$ |

Mean $= \frac{\Sigma fx}{n} = \frac{2100}{60} = \frac{210}{6} = 35$

Mean deviation

$= \frac{\Sigma f|d|}{N} = \frac{680}{60}$

$= \frac{68}{6}$

MD $= 11.333$

Que: Calculate mean deviation about median for the following data. Also find coefficient of mean deviation.

| $x$ : | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|---|
| freq: | 18 | 16 | 15 | 12 | 10 | 5 | 2 | 2 |

| $x$ | freq | cf | $\|d\| = \|x-24\|$ | $f\|d\|$ |
|---|---|---|---|---|
| 5 | 18 | 18 | 19 | 342 |
| 15 | 16 | 34 | 9 | 144 |
| 25 | 15 | 49 | 1 | 15 |
| 35 | 12 | 61 | 11 | 132 |
| 45 | 10 | 71 | 21 | 210 |
| 55 | 5 | 76 | 31 | 155 |
| 65 | 2 | 78 | 41 | 82 |
| 75 | 2 | 80 | 51 | 102 |
| N = 80 | | | $\Sigma$ | $\Sigma f\|d\| = 1182$ |

median $= \frac{N}{2} = \frac{80}{2} = 40^{th}$ term

Median $= N/2 = \frac{80}{2} = 40^{th}$ term

median $= l_1 + \left(\frac{N/2 - cf}{f}\right) \times c$

$l_1 = 20$, $N/2 = 40$, $cf = 34$, $c = 10$, $f = 15$

median $= 20 + \left(\frac{40 - 34}{15}\right) \times 10$

$= 20 + \frac{16}{15} \times 10$

$= 20 + \frac{160}{15}$

$= 20 + 4 = 24$

Mean deviation about median $= \frac{\Sigma f|d|}{N} = \frac{1182}{80}$

$= 14.775$

Coefficient of mean deviation

$= \frac{\text{Mean deviation}}{\text{median}}$

$= \frac{14.775}{24}$

$= .6156$

03/07/2019

## Standard Deviation

It is the square root of the mean of the squares of deviations of all values of a series from their arithmetic mean.

If $x_1, x_2 \dots x_n$ are $n$ values then the standard deviation $\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots (x_n - \bar{x})^2}{n}}$

$$\sigma = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n}}$$

where $n$ is the number of items

The minimum value of standard deviation of a series is 0.

## Standard deviation in Individual series

**Ques** find standard deviation of the values : 4, 8, 10, 12, 15, 9, 7, 7

| $x$ | $x-\bar{x}$ $(x-9)$ | $(x-\bar{x})^2$ |
|---|---|---|
| 4 | -5 | 25 |
| 8 | -1 | 1 |
| 10 | 1 | 1 |
| 12 | 3 | 9 |
| 15 | 6 | 36 |
| 9 | 0 | 0 |
| 7 | -2 | 4 |
| 7 | -2 | 4 |
| $\boxed{\Sigma x = 72}$ | | $\Sigma(x-\bar{x})^2 = 80$ |

$$\bar{x} = \frac{\Sigma x}{n} = \frac{72}{8} = 9$$

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$= \sqrt{\frac{80}{8}}$$

$$= \sqrt{10}$$

$$\sigma = 3.16$$

**Ques**

**NOTE** This formula can also be written as

$$\frac{SD}{(\sigma)} \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

**Ques** For the following values find standard deviation
5, 8, 7, 11, 9, 10, 8, 2, 4, 6,

| $x$ | $x^2$ |
|---|---|
| 5 | 25 |
| 8 | 64 |
| 7 | 49 |
| 11 | 121 |
| 9 | 81 |
| 10 | 100 |
| 8 | 64 |
| 2 | 4 |
| 4 | 16 |
| 6 | 36 |
| $\Sigma x$ | $\boxed{\Sigma x^2 =}$ |
| $\Sigma x = 70$ | 560 |

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{560}{10} - \left(\frac{70}{10}\right)^2}$$

$$= \sqrt{56 - 49}$$

$$= \sqrt{7}$$

$$\sigma = 2.645$$

**Shortcut method**

$$SD = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

Coefficient of SD = $\frac{SD}{Mean}$

$$SD = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$$

(assumed average)

where d is the deviation from any origin

Coefficient of variation

$$= \frac{SD}{Mean} \times 100$$

Que Find the standard deviation and coefficient of variation of the values:- 10, 12, 80, 70, 60, 100, 0, 4

| $x$ | $x^2$ |
|---|---|
| 10 | 100 |
| 12 | 144 |
| 80 | 6400 |
| 70 | 4900 |
| 60 | 3600 |
| 100 | 10000 |
| 0 | 0 |
| 4 | 16 |
| $\Sigma x = 336$ | $\Sigma x^2 = 25160$ |

Mean, $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{336}{8}$

$= 42$

$$\sigma = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$= \sqrt{\frac{25160}{8} - \left(\frac{336}{8}\right)^2}$$

$$= \sqrt{3145 - (42)^2}$$

$$= \sqrt{3145 - 1764}$$

$$= \sqrt{1381}$$

$SD = 37.16$

Coeff. of variation $= \frac{SD}{Mean} \times 100$

$= \frac{37.16}{42} \times 100$

$= \frac{3716}{42}$

$= 88.476$

Variance

Variance = (Standard Deviation)$^2$

$$= \frac{\Sigma(x - \bar{x})^2}{n}$$

OR

$$= \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2$$

Ques find mean, standard deviation and variance of the following values : 2, 3, 5, 4, 6, 8, 10, 2

| $x$ | $x^2$ |
|---|---|
| 2 | 4 |
| 3 | 9 |
| 5 | 25 |
| 4 | 16 |
| 6 | 36 |
| 8 | 64 |
| 10 | 100 |
| 2 | 4 |
| $\boxed{\Sigma x = 40}$ | $\boxed{\Sigma x^2 = 258}$ |

Mean, $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{40}{8}$

$\bar{x} = 5$

$SD = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

$= \sqrt{\frac{258}{8} - \left(\frac{40}{8}\right)^2}$

$= \sqrt{32.25 - 25}$

$SD = \sqrt{7.25}$

$SD = 2.692$

Variance $= 7.25 \Rightarrow \left[\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2\right]$

Ques calculate the coefficient of variation of a series on the basis of the following results.

$N = 50$, $\Sigma x = -100$, $\Sigma x^2 = 1000$ where $x$ is the deviation from assumed mean 14.5 and N is the no. of items.

$a = 14.5$

$SD = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$

$= \sqrt{\frac{1000}{50} - \left(\frac{-100}{50}\right)^2}$

$= \sqrt{20 - 4}$

$= \sqrt{16}$

$SD = 4$

Coeff $= \frac{SD}{Mean} \times 100 = \frac{4}{Mean} \times 100$

Mean $= a + \frac{\Sigma x}{n}$

$= 14.5 + \frac{-100}{50}$

$= 14.5 - 2$

Mean $= 12.5$

$$\text{coeff} = \frac{SD}{Mean} \times 100$$

$$= \frac{4}{12.5} \times 100$$

$$\text{coeff} = \frac{400}{12.5}$$

$$\text{coeff of variation} = 32$$

Que For a set of 7 observations Mean = 60, sum of squared deviation from mean = 184. Find coefficient of variation

$n = 7$ $\quad \bar{x} = 60$

$\Sigma(x - \bar{x})^2 = 184$

$$SD = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{184}{7}}$$

$$= \sqrt{26.29}$$

$$= 5.127$$

$$SD = 5.13$$

$$\text{Coeff} = \frac{SD}{Mean} \times 100$$

$$= \frac{5.13}{60} \times 100$$

$$= \frac{51.3}{6}$$

$$\text{Coeff of variation} = 8.55$$

Que If $N = 10$, $\bar{x} = 12$, $\Sigma x^2 = 1530$, find coefficient of standard deviation and coefficient of variation

$N = 10$, $\bar{x} = 12$, $\Sigma x^2 = 1530$

$$\bar{x} = \frac{\Sigma x}{n}$$

$$12 = \frac{\Sigma x}{10}$$

$$\Sigma x = 120$$

$$\cancel{SD = \sqrt{\Sigma(x - \bar{x})}}$$

$$SD = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}$$

$$SD = \sqrt{\frac{1530}{10} - \left(\frac{120}{10}\right)^2}$$

$$= \sqrt{153 - 144}$$

$$= \sqrt{9}$$

$SD = 3$

Variance = 9

Coeff. of variation $= \frac{SD}{Mean} \times 100$

$= \frac{3}{12} \times 100$

$= \frac{1}{4} \times 100$

$= 0.25 \times 100$

$= 25$

Coefficient of SD $= \frac{SD}{Mean}$

$= \frac{3}{12}$

$= \frac{1}{4} = 0.25$

Standard deviation in discrete series

Direct method

$$SD = \sqrt{\frac{\Sigma f (x - \bar{x})^2}{N}}$$

Ques find standard deviation of the following data

Size : 2 3 4 5 6 7 8

freq : 1, 2, 3, 5, 3 2 1

| $x$ | $f$ | $fx$ |
|---|---|---|
| 2 | 1 | 2 |
| 3 | 2 | 6 |
| 4 | 3 | 12 |
| 5 | 5 | 25 |
| 6 | 3 | 18 |
| 7 | 2 | 14 |
| 8 | 1 | 8 |
| $\Sigma x = 35$ | $N = 17$ | $\Sigma fx = 85$ |

~~$n = 7$~~

~~$\bar{x} = \frac{\Sigma x}{n}$~~

~~$\bar{x} = \frac{35}{n}$~~

~~$\bar{x} =$~~

$\bar{x} = 5$

| $(x - \bar{x})$ | $(x - \bar{x})^2$ | $f(x - \bar{x})^2$ |
|---|---|---|
| -3 | 9 | 9 |
| -2 | 4 | 8 |
| -1 | 1 | 3 |
| 0 | 0 | 0 |
| 1 | 1 | 3 |
| 2 | 4 | 8 |
| 3 | 9 | 9 |
| | | $\Sigma = 40$ |

12
16
6
40

$\bar{x} = \frac{\Sigma f x}{n}$

$\bar{x} = \frac{85}{17}$

$\bar{x} = 5$

$$SD = \sqrt{\frac{\Sigma f (x - \bar{x})^2}{N}}$$

$$= \sqrt{\frac{40}{17}}$$

$$= \sqrt{2.35}$$

$SD = 1.53$

The formula can also be writtenas

$$SD = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

Ques For the following data calculate standard deviation.

Marks : 2 | 4 | 6 | 8 | 10

freq : 8 | 10 | 16 | 9 | 7

(no of stdnts)

| $x$ | $f$ | $fx$ | $x^2$ | $fx^2$ |
|---|---|---|---|---|
| 2 | 8 | 16 | 4 | 32 |
| 4 | 10 | 40 | 16 | 160 |
| 6 | 16 | 96 | 36 | 576 |
| 8 | 9 | 72 | 64 | 576 |
| 10 | 7 | 70 | 100 | 700 |
| | $N = 50$ | $\Sigma fx = 294$ | | $\Sigma fx^2 = 2044$ |

$(\Sigma f x)^2 = \cancel{86436}$

$$SD = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

$$= \sqrt{\frac{2044}{50} - \left(\frac{294}{50}\right)^2}$$

$$= \sqrt{40.88 - 34.57}$$

$SD = \sqrt{6.31}$

$SD = 2.51$

## Shortcut method

$$SD = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \qquad d = x - a$$

Ques find SD

| Marks | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|
| No. of students | 8 | 10 | 16 | 9 | 7 |

| $x$ | $f$ | $d = x - a$ | $fd$ | $d^2$ | $fd^2$ |
|---|---|---|---|---|---|
| 2 | 8 | -4 | -32 | 16 | 128 |
| 4 | 10 | -2 | -20 | 4 | 40 |
| 6 | 16 | 0 | 0 | 0 | 0 |
| 8 | 9 | 2 | 18 | 4 | 36 |
| 10 | 7 | 4 | 28 | 16 | 112 |
| | $\boxed{N = 50}$ | | $\boxed{\Sigma = -6}$ | | $\boxed{\Sigma = 316}$ |

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

$$= \sqrt{\frac{316}{50} - \left(\frac{-6}{50}\right)^2}$$

$$= \sqrt{6.32 - 0.014}$$

$$= \sqrt{6.31}$$

$$SD = 2.51$$

25/08/2019

## Standard deviation in continous series

In continous series, we take mid values of classes and consider it as $x$ values

1. Direct Method

$$\sqrt{\frac{\Sigma f(x-\bar{x})^2}{n}} \quad \text{OR} \quad \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

2. Shortcut method

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

Variance $= (SD)^2$

3. Step deviation method

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

where $d = x - a$ and $d' = \frac{x-a}{c}$

where $c$ = class interval

Ques Find standard deviation from the following data using direct method, shortcut method, step deviation method

| Size | 0-2 | 2-4 | 4-6 | 6-8 | 8-10 | 10-12 |
|---|---|---|---|---|---|---|
| freq | 2 | 4 | 6 | 4 | 2 | 6 |

$a = 5$

| Size | Mid value (x) | freq | fx | $x^2$ | $fx^2$ | $d = x - a$ | $d^2$ | $fd^2$ | fd | $d' = \frac{x-a}{c}$ | $d'^2$ | $fd'^2$ | $fd'$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0-2 | 1 | 2 | 2 | 1 | 2 | -4 | 16 | 32 | -8 | -2 | 4 | 8 | -4 |
| 2-4 | 3 | 4 | 12 | 9 | 36 | -2 | 4 | 16 | -8 | -1 | 1 | 4 | -4 |
| 4-6 | 5 | 6 | 30 | 25 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6-8 | 7 | 4 | 28 | 49 | 196 | 2 | 4 | 16 | 8 | 1 | 1 | 4 | 4 |
| 8-10 | 9 | 2 | 18 | 81 | 162 | 4 | 16 | 32 | 8 | 2 | 4 | 8 | 4 |
| 10-12 | 11 | 6 | 66 | 121 | 726 | 6 | 36 | 216 | 36 | 3 | 9 | 54 | 18 |
| | | N $\boxed{24}$ | $\boxed{156}$ | | $\boxed{1272}$ | | | $\boxed{312}$ | $\boxed{36}$ | (4) | | 78 | ~~26~~ 18 |

D.M, SD $= \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$

$= \sqrt{\frac{1272}{24} - \left(\frac{156}{24}\right)^2}$

$= \sqrt{53 - 42.25}$

$\sqrt{10.75}$

$= 3.27$

$= 3.3$

~~Mean $= \frac{\Sigma fx}{n} = \frac{156}{6} = 26$~~

S.M, SD $= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

$= \sqrt{\frac{312}{24} - \left(\frac{36}{24}\right)^2}$

$= \sqrt{13 - (1.5)^2}$

$= \sqrt{13 - 2.25}$

$= \sqrt{10.75}$

$= 3.3$

~~Step~~ .

S.D, SD $= \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$

$= \sqrt{\frac{78}{24} - \left(\frac{18}{24}\right)^2} \times 2$

$= \sqrt{3.25 - 0.5625} \times 2$

$\sqrt{2.686 \times 2}$

$\sqrt{2.075 \times 2}$ $= 1.63 \times 2$

$= \cancel{1.44 \times 2}$

$= 3.27$

Ques Find standard deviation and variance

| Income | above 0 | above 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|---|
| no. of persons | 100 | 97 | 89 | 77 | 57 | 27 | 12 | 5 | 0 |

| Income | f | x | fx | fx² | x² |
|---|---|---|---|---|---|
| 0-10 | 3 | 5 | 15 | 75 | 25 |
| 10-20 | 8 | 15 | 120 | 1800 | 225 |
| 20-30 | 12 | 25 | 300 | 7500 | 625 |
| 30-40 | 20 | 35 | 700 | 24500 | 1225 |
| 40-50 | 30 | 45 | 1350 | 60750 | 2025 |
| 50-60 | 15 | 55 | 825 | 45375 | 3025 |
| 60-70 | 7 | 65 | 455 | 29575 | 4225 |
| 70-80 | 5 | 75 | 375 | 28125 | 5625 |
| 80-90 | 0 | 85 | 0 | 0 | 7225 |
| | 100 | | 4140 | 197700 | |

$$SD = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

$$= \sqrt{\frac{197700}{100} - \left(\frac{4140}{100}\right)^2}$$

$$= \sqrt{1977 - 1713.96}$$

$$= \sqrt{263.04}$$

$SD = 16.218$

Variance $= (SD)^2$

$= 263.023$

SD 15.94

cof = 51.42

Ques Obtain standard deviation for the data on scores given below. Also find coefficient of variation.

| Score | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| no. of stu: | 10 | 15 | 25 | 25 | 10 | 10 | 5 |

| Score | f | x | x² | fx | fx² |
|---|---|---|---|---|---|
| 0-10 | 10 | 5 | 25 | 50 | 250 |
| 10-20 | 15 | 15 | 225 | 225 | 3375 |
| 20-30 | 25 | 25 | 625 | 625 | 15625 |
| 30-40 | 25 | 35 | 1225 | 875 | 30625 |
| 40-50 | 10 | 45 | 2025 | 450 | 20250 |
| 50-60 | 10 | 55 | 3025 | 550 | 30250 |
| 60-70 | 5 | 65 | 4225 | 325 | 21125 |
| | 100 | | | 3100 | 121500 |

$N = 100$ $\quad \Sigma fx = 3100$

$\Sigma fx = 121500$

$$SD = \sqrt{\frac{\Sigma f x^2}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

$$= \sqrt{\frac{121500}{100} - \left(\frac{3100}{100}\right)^2}$$

$$= \sqrt{1215 - 961}$$

$$SD = \sqrt{254}$$

$$SD = 15.93$$

Coefficient of variation $= \frac{SD}{Mean} \times 100$

$$Mean = \frac{\Sigma fx}{n} = \frac{3100}{100} = 31$$

Coefficient of variation $= \frac{SD}{Mean} \times 100$

$$= \frac{15.93}{31} \times 100$$

$$= 0.5138 \times 100$$

Coeff $= 51.38$

Ques The scores of two batsmen A and B in 8 innings ~~bowling~~ a certain match are as follows: Examine which of the two batsman is more consistent in scoring. Who is more efficient batsman?

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Bat A | 10 | 12 | 80 | 70 | 60 | 100 | 0 | 4 |
| Bats B | 8 | 9 | 7 | 10 | 5 | 9 | 10 | 8 |

~~Coeff~~

~~Not~~ NOTE: We can compare the variability in the scores of two batsmen using coefficient of variation. The series for which the coefficient of variation is less is considered to be more stable (or consistent)

$$\sqrt{\frac{\Sigma f x}{N} - \left(\frac{\Sigma f x}{N}\right)^2}$$

~~Batsman~~

$$SD = \sqrt{\frac{\Sigma x^2}{N} - \left(\frac{\Sigma x}{N}\right)^2}$$

Batsman A

| ~~Bats A~~ x | ~~Bats B~~ | $x^2$ |
|---|---|---|
| 10 | ~~8~~ | 100 |
| 12 | ~~9~~ | 144 |
| 80 | ~~7~~ | 6400 |
| 70 | ~~10~~ | 4900 |
| 60 | ~~5~~ | 3600 |
| 100 | ~~9~~ | 10600 |
| 0 | ~~10~~ | 0 |
| 4 | ~~8~~ | 16 |
| 336 | | 25166 |

Batsman B

| x | $x^2$ |
|---|---|
| 8 | 64 |
| 9 | 81 |
| 7 | 49 |
| 10 | 100 |
| 5 | 25 |
| 9 | 81 |
| 10 | 100 |
| 8 | 64 |
| 66 | 464 |

Batsman A

Mean $= \frac{\Sigma x}{n} = \frac{336}{8} = \underline{42}$

SD $= \sqrt{\frac{25160}{8} - \left(\frac{336}{8}\right)^2}$

$= \sqrt{3145 - 1764}$

$= \sqrt{1381} = \underline{\underline{7.16}}$

Coff of variance $= \frac{SD}{Mean} \times 100 = \frac{37.16}{42} \times 100$

$= \frac{3716}{42} = \underline{\underline{88.47}}$

Batsman B

Mean, $\bar{x} = \frac{\Sigma x}{n} = \frac{66}{8} = \underline{\underline{8.25}}$

SD $= \sqrt{\frac{564}{8} - \left(\frac{66}{8}\right)^2} = \sqrt{70.5 - 68}$

$= \sqrt{2.15} = \underline{\underline{1.58}}$

Coff of variance $= \frac{1.58}{8.15} \times 100$

$= 18.15$

Batsman A is more ~~the~~ efficient.

Ques Shares of 2 companies have the following information

| | Mean of share values | SD of share values |
|---|---|---|
| Company A : | 15 | 5 |
| Company B : | 20 | 8 |

Examine, (a) which companie's shares are better

(b) Which companies shares have greater variability?

(a) Average share value in company B is greater so shares of company B are better.

(b) Coeff of variation of company A :-

$\frac{SD}{mean} \times 100 = \frac{5}{15} \times 100$

$= \underline{\underline{33.33}}$

Coeff of variation of company B

$= \frac{8}{20} \times 100 = \underline{\underline{40}}$

Company B has greater variability!

Graphical representation

Graphs

Graphs are statistical devices which can be used for presenting frequency distribution or representing the relation between two variables.

The most commonly used graphs for representing a frequency distribution are

1. Histogram

2. Frequency ~~polgon~~ polygon

3. Line graph

4. Frequency curve

5. Ogive / Cumulative frequency curve

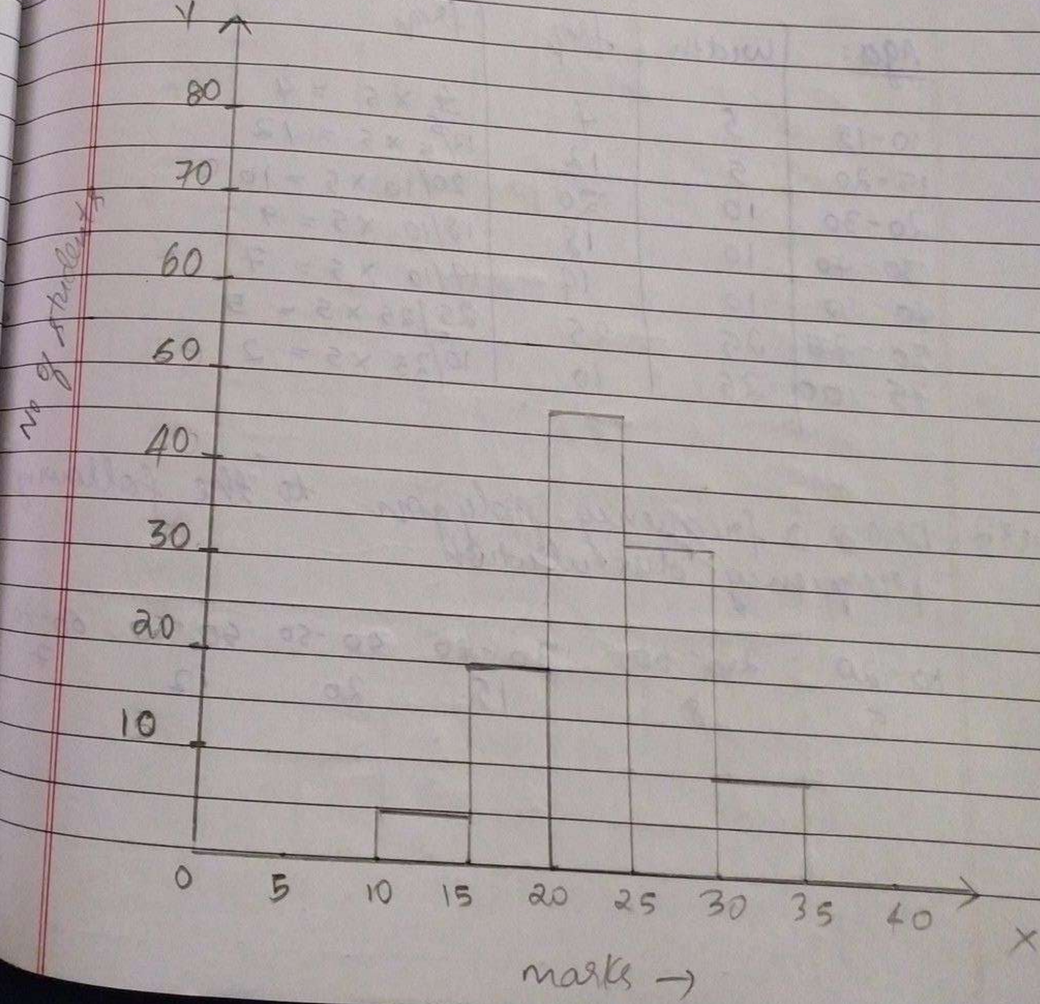
Histogram

It is a set of vertical bar whose areas are proportional to the frequencies. The variable is always taken on the x axis and the frequencies on the Y axis.

The width of the bars in the histogram will be proportional to the class interval. The bars are drawn without leaving any space between them.

Que Represent the following frequency table by histogram

| Marks | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 |
|---|---|---|---|---|---|
| no of students | 5 | 20 | 47 | 38 | 10 |



Histogram, when class intervals are not uniform

Ques1 Draw a histogram to the frequency distribution showing the ages of people.

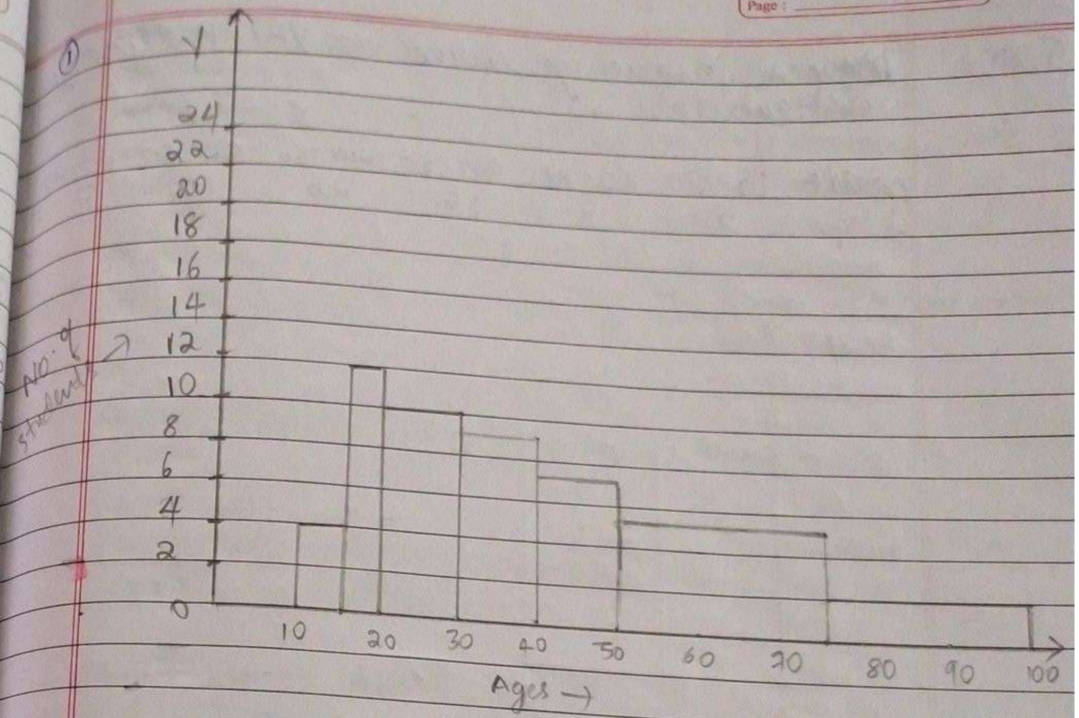
| Ages | 10-15 | 15-20 | 20-30 | 30-40 | 40-50 | 50-75 | 75-100 |
|---|---|---|---|---|---|---|---|
| freq | 4 | 12 | 20 | 18 | 14 | 25 | 10 |

| Ages: | width | freq | freq |
|---|---|---|---|
| 10-15 | 5 | 4 | $\frac{4}{5} \times 5 = 4$ |
| 15-20 | 5 | 12 | $12/5 \times 5 = 12$ |
| 20-30 | 10 | 20 | $20/10 \times 5 = 10$ |
| 30-40 | 10 | 18 | $18/10 \times 5 = 9$ |
| 40-50 | 10 | 14 | $14/10 \times 5 = 7$ |
| 50-75 | 25 | 25 | $25/25 \times 5 = 5$ |
| 75-100 | 25 | 10 | $10/25 \times 5 = 2$ |

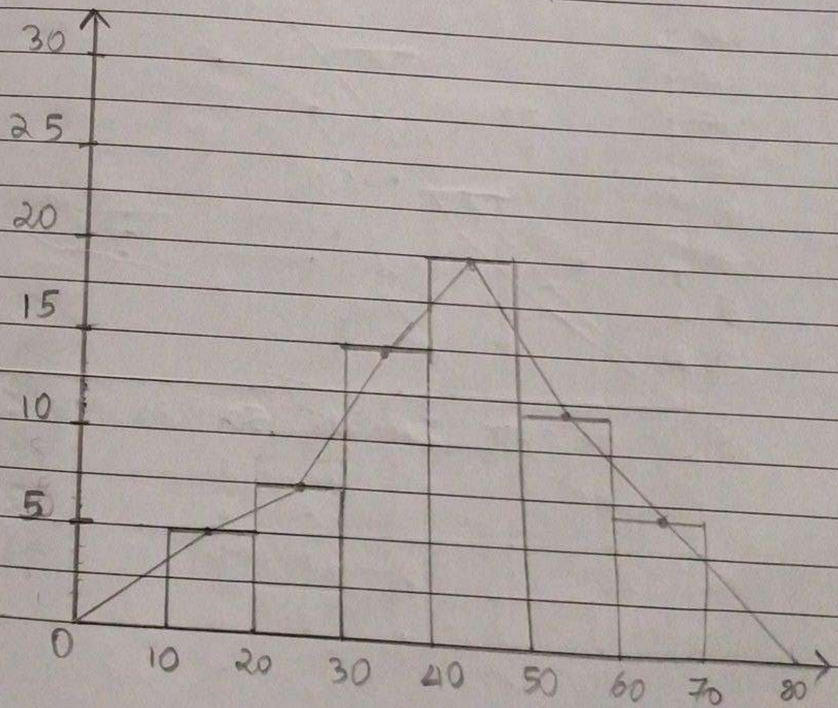
Ques2 Draw a frequency polygon to the following frequency distribution

| 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|
| 5 | 8 | 15 | 20 | 12 | 7 |

①

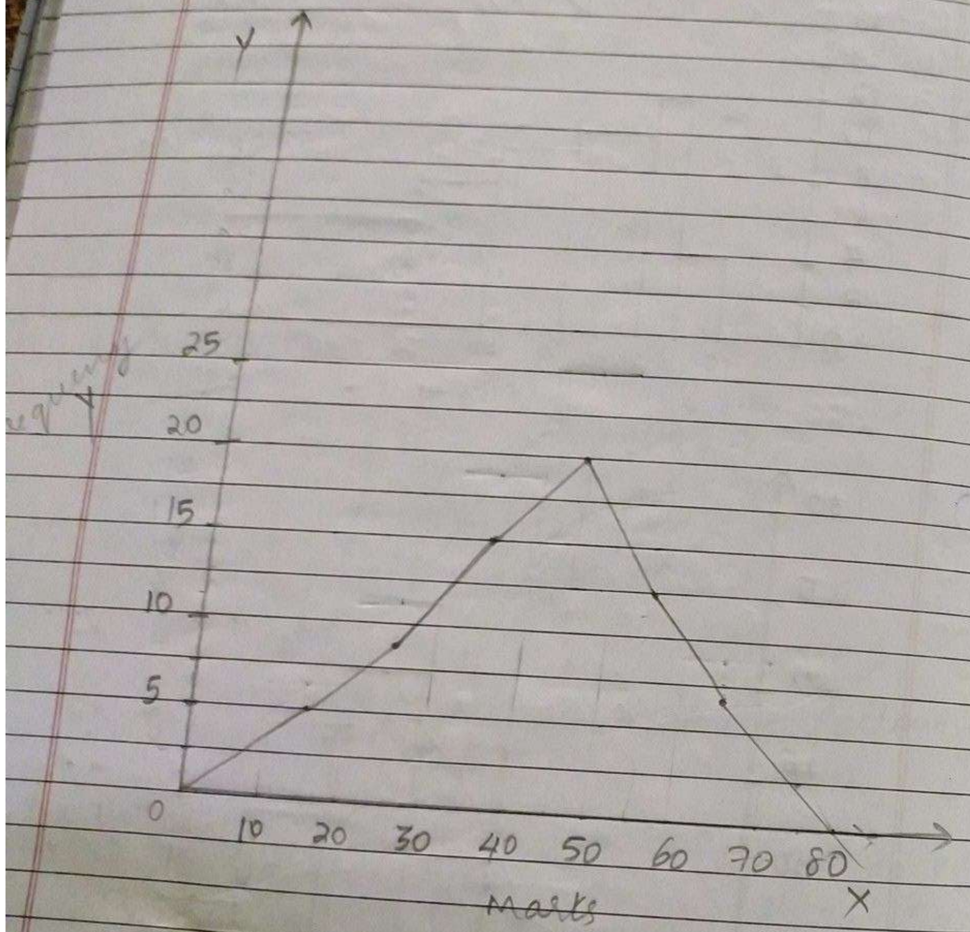



②



Ques Draw a frequency curve to the following distribution

| Marks | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|
| freq | 5 | 9 | 15 | 20 | 12 | 7 |




7/08/2019
Wednesday

## False Base Line

If the fluctuation in the values of a variable are very small as compared to the size of items & the techinque of false base line is used.
Two zig-zag horizontal lines are drawn above the x axis

## Ogive (Cumulative Frequency Curve)

Ques Form less than and more than cumulative frequency distributions for the following frequency distribution.

| Marks | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|
| No of students | 2 | 5 | 10 | 8 | 3 |

Less than c.f distribution

| Marks (Below) | c.f. |
|---|---|
| 10 | 0 |
| 20 | 2 |
| 30 | 7 |
| 40 | 17 |
| 50 | 25 |
| 60 | 28 |

More than c.f distribution

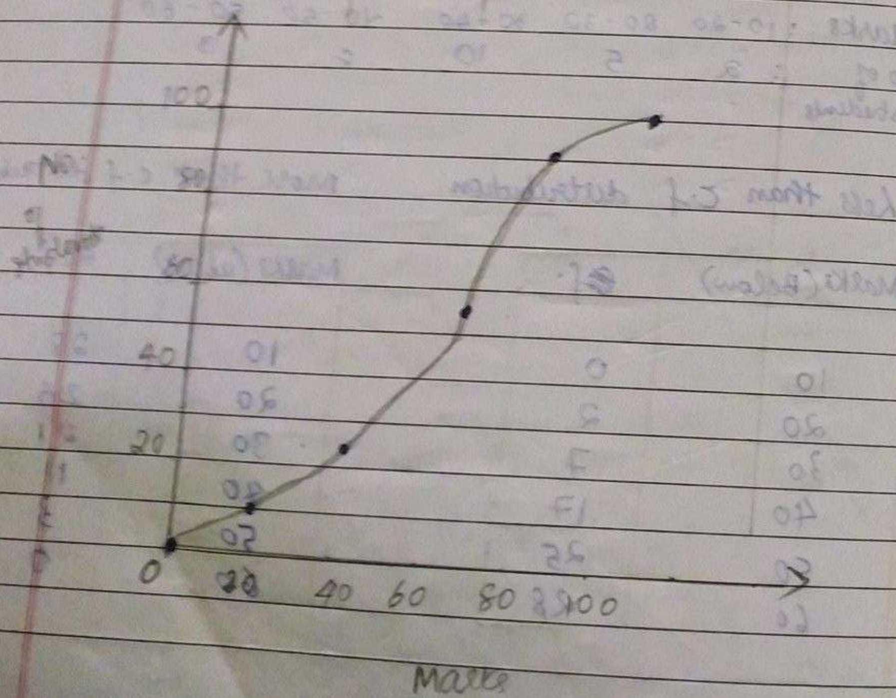
| Marks (above) | c.f |
|---|---|
| 10 | 28 |
| 20 | 26 |
| 30 | 21 |
| 40 | 11 |
| 50 | 3 |
| 60 | 0 |

Ques Draw less than ogive for the following data

| Marks | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
|---|---|---|---|---|---|
| no of students | 10 | 15 | 30 | 35 | 10 |

Less than c.f distribution

| Marks (Below) | f |
|---|---|
| 0 | 0 |
| 20 | 10 |
| 40 | 25 |
| 60 | 55 |
| 80 | 90 |
| 100 | 100 |

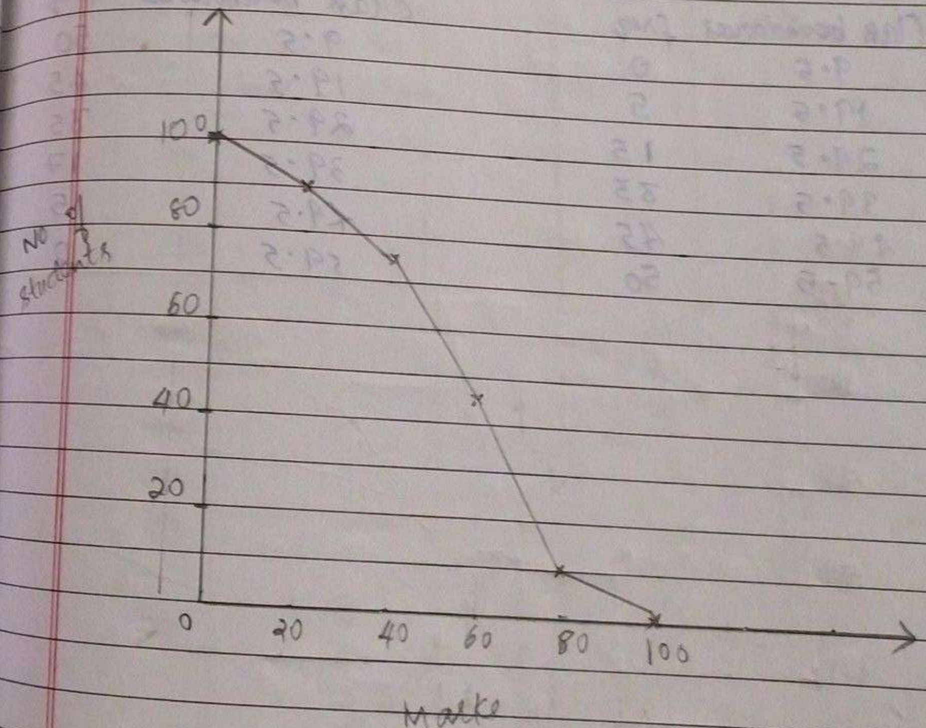



Ques Draw more than ogive for the following data

| Marks | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
|---|---|---|---|---|---|
| no of students | 10 | 15 | 30 | 35 | 10 |

More than c.f distribution

| Marks (above) | f |
|---|---|
| 0 | 100 |
| 20 | 90 |
| 40 | 75 |
| 60 | 45 |
| 80 | 10 |
| 100 | 0 |



Ques Draw the two ogives for the following data on the same graph.

| Marks | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 |
|---|---|---|---|---|---|
| no of students | 5 | 10 | 18 | 12 | 5 |

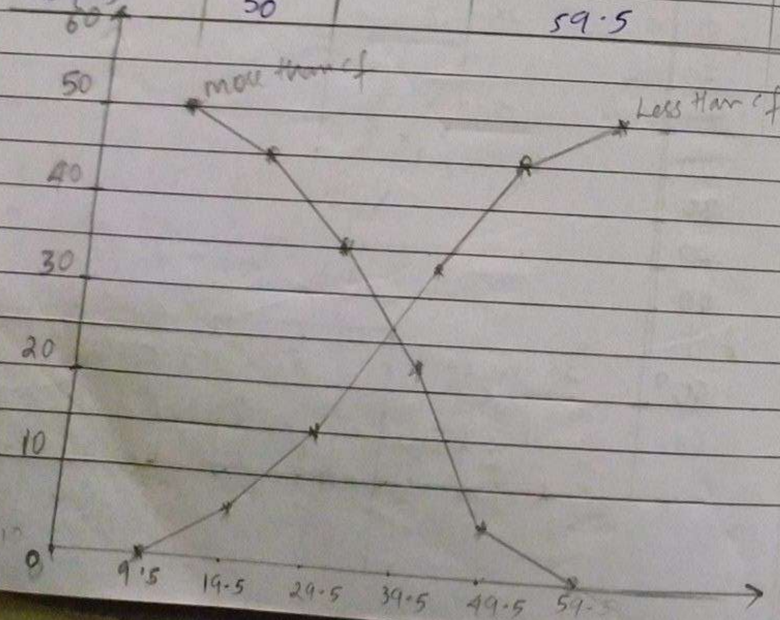
| Marks | frequency |
|---|---|
| 9.5-19.5 | 5 |
| 19.5-29.5 | 10 |
| 29.5-39.5 | 18 |
| 39.5-49.5 | 12 |
| 49.5-59.5 | 5 |

Less than c.f

| Class boundaries | freq |
|---|---|
| 9.5 | 0 |
| 19.5 | 5 |
| 29.5 | 15 |
| 39.5 | 33 |
| 49.5 | 45 |
| 59.5 | 50 |

More than c.f

| Class boundaries | freq |
|---|---|
| 9.5 | 50 |
| 19.5 | 45 |
| 29.5 | 35 |
| 39.5 | 17 |
| 49.5 | 5 |
| 59.5 | 0 |




Graphic Method of Finding Median

Ques Draw the two ogives for the following data and estimate the median value.

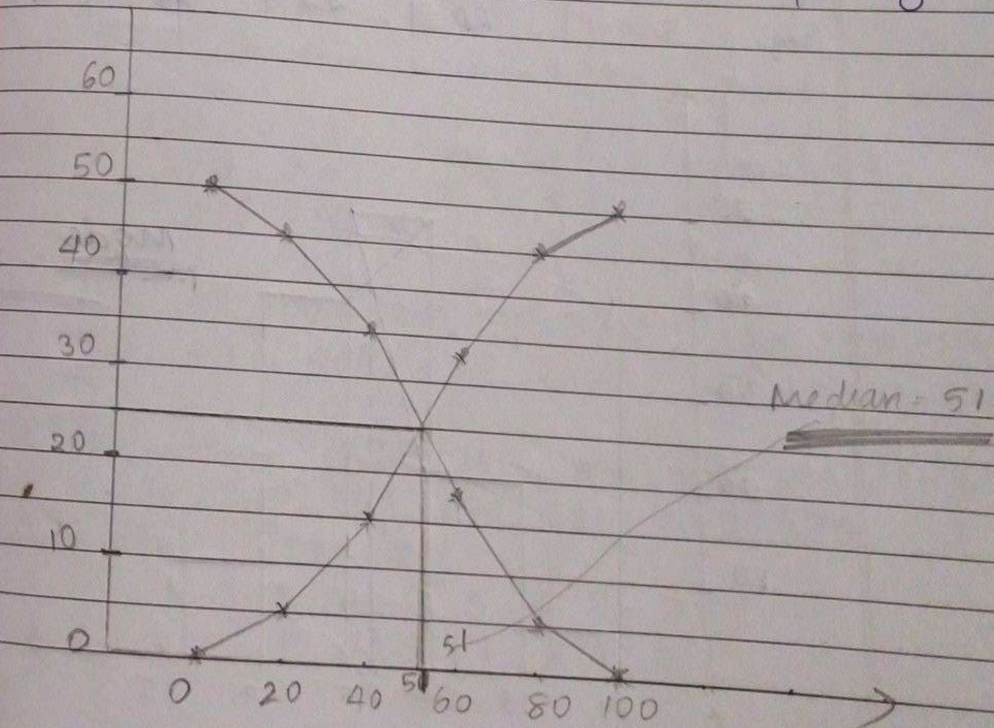
| Marks | 0-20 | 20-40 | 40-60 | 60-80 | 80-100 |
|---|---|---|---|---|---|
| no of stdn | 5 | 10 | 18 | 12 | 5 |

Less than c.f

| Class boundaries | freq |
|---|---|
| 0 | 0 |
| 20 | 5 |
| 40 | 15 |
| 60 | 33 |
| 80 | 45 |
| 100 | 50 |

More than c.f

| Class bound | freq |
|---|---|
| 20 | 50 |
| 20 | 45 |
| 40 | 35 |
| 60 | 17 |
| 80 | 5 |
| 100 | 0 |

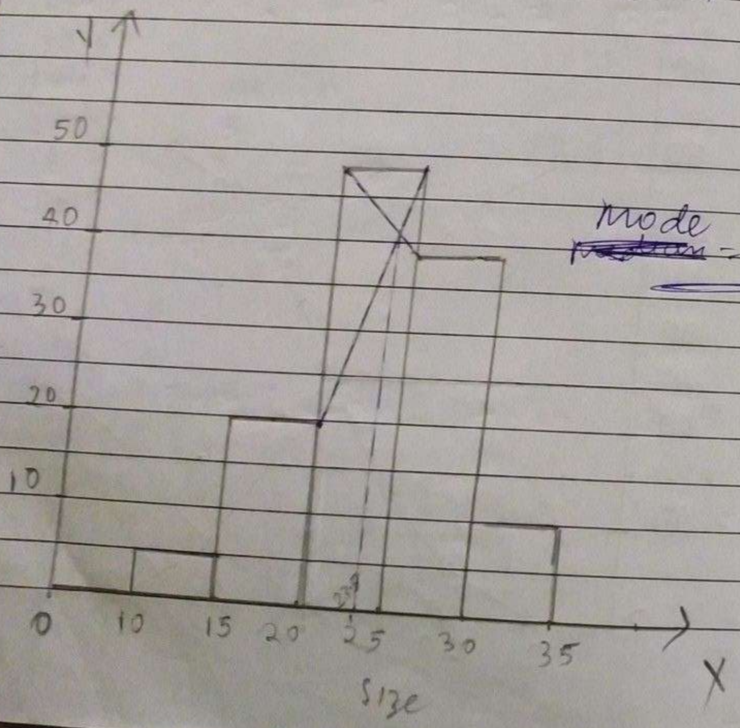


Graphic method of Finding mode

Steps
1. Present the data in the form of a histogram
2. Identify the rectangle of maximum height
3. Join the corners of this rectangle with the immediately next corners of adjacent rectangle
4. From the point of intersection of these two lines draw perpendicular down to x axis
5. The foot of the perpendicular is mode

Ques Represent the following frequency table by histogram.

| Size | 10 - 15 | 15 - 20 | 20 - 25 | 25 - 30 | 30 - 35 |
|---|---|---|---|---|---|
| freq | 5 | 20 | 47 | 38 | 10 |




mode ~~median~~ = 22.5

Stem and Leaf chart:

Stem and Leaf chart is a device for presenting quantitative data in a graphic format similar to a histogram

To construct a stem and leaf chart the observations must be sorted in ascending order and then sort the leaves.

eg:- 24, 26, 27, 29, 43, 44, 46, 48, 48, 52, 52, 55, 56, 61, 64, 68, 86

| ~~Leaf~~ Stem | ~~Stem~~ Leaf |
|---|---|
| 2 | 4 6 7 9 |
| 4 | 3 4 6 8 8 |
| 5 | 2 2 5 6 |
| 6 | 1 4 8 |
| 8 | 6 |

Key 2/4 = 24 ; 2/6 = 26 ; 2/7 = 27 ;
2/9 = 29 ; 4/3 = 43 ; 6/1 = 61 ;
8/6 = 86 . etc

eg:- 110, 120, 100, 120, 150, 160, 170, 190, 170, 210, 240, 260, 230, 270, 290, 290, 280

100, 110, 120, 120, 150, 160, ~~160~~, 170, 170, 190, 210, 230, 240, 260, 270, 280, 290, 290 .

| ~~Leaf~~ Stem | ~~Stem~~ Leaf |
|---|---|
| 1 | 0 1 2 2 5 6 7 7 9 |
| 2 | 1 3 4 6 7 8 9 9 |

key 2/1 means 210 ; 2/3 means 230 ; 2/9 means 290 ; etc (so on)

eg: 5.8, 5.9, 6.1, 6.2, 6.8, 7.3, 7.4, 7.6, 7.7, 8.1, 8.1, 8.2, 8.8, 9.2

| Stem | Leaves |
|---|---|
| 5 | 8 9 |
| 6 | 1 2 8 |
| 7 | 3 4 6 7 |
| 8 | 1 1 2 8 |
| 9 | 2 |

Key 5/8 means 5.8. 9/2 means 9.2 6/1 means 6.1 ; soon

eg 23.25, 24.13, 24.76, 24.81, 24.98, 25.31, 25.57, 25.89, 26.28, 26.34, 27.09.

23.3, 24.1, 24.8, 24.8, 25.0, 25.3, 25.6, 25.9, 26.3, 26.3, 27.1

| Stem | Leaves |
|---|---|
| 23 | 3 |
| 24 | 1 8 8 |
| 25 | 0 3 6 9 |
| 26 | 3 3 |
| 27 | 1 |

key 25/0 = 25.0, 26/3 = 26.3 23/3 = 23.3 etc

eg 80, 65, ~~78~~, 83, 79, 72, 95, 67, ~~71~~, 85, 83, 92, 86, 63, 91, 86, 75, ~~78~~, ~~73~~, 86, 91, 77, 79, 80, 91, 87, 85, ~~76~~, 68, ~~76~~, 89, 96, 98.

63, 65, 67, 68, 71, 73, 75, ~~78~~, ~~77~~, ~~79~~ 76, 76, 77, 78, 78,

07/08/2019
Wednesday

## MODULE - III

### PROBABILITY

The word probability means chance. An event which is not certain to happen is a problem of probability.

for example: when we toss a coin we can enumerate all the possible outcomes. The possible outcomes are either the coin will turn head or will turn tail. One of them will happen. But we cannot say which one it will be.
In these kind of problems we can only make predictions. while making predictions about the occurrence about the uncertain events we work out the amount of likely hood of the occurence of those events ie probability.

The probability of a given event maybe defined as the numerical value given to the likelyhood of the occurence of that event. It is a number lying between zero and one.
Zero for an event which cannot occur and one for an event certain to occur.

When the occurence of an event is uncertain, probability is a number lying between 0 and 1.

for eg:- when we toss a coin the event of getting head is uncertain, so its probability is neither 0 nor 1, but between the two.
Therefore the probability of head = 1/2

probability of tail = 1/2.

Random Experiments

An experiment that has two or more outcomes which vary in an upredictable manner from trail to trail when conducted under uniform conditions is called a random experiment.

In a random experiment all the possible outcomes are known in advance but none of the outcomes can be predicted with certainity.

Features of random experiments are
1. It has more than one outcome.
2. The outcomes are unpredictable.
3. The experiment is repeatable.

Sample point / Simple Event / Elementary outcome

Every indecomposable outcome of a random experiment is called a sample point of that random experiment

Sample Space

The sample space of a random experiment is a set of containing all the sample points of that random experiment. Therefore the sample space of a random experiment is the totality of all the elementary outcomes of that random experiment.

for eg:- when we toss a coin, the sample space [H, T].

for eg:- When two coins are tossed, the sample space is [HH, TT, HT, TH].

Ques A box contain 10 tickets each numbered 1 to 10. A ticket is drawn. What is the sample space?

Sample space = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

Ques From a lot containing good and bad items three items are chosen. Prepare the sample space. $2^n$

Let G stands for good items
B stands for bad items

Sample Space = [GGG, GGB, GBB, GBG, BBB, BBG, BGB, BGG]

Event

An event is a subset of the sample space of a random experiment.

for eg when two coins are tossed getting two heads is an event. An event maybe simple or compound. An event is said to be simple if it corresponds to a single elementary outcome of an experiment.

for example when a die is thrown getting 2 is a simple event.

The joint occurence of two or more simple event is called a compound event

for example when a die is thrown getting an even number is a compound event because even number can be ~~to~~ 2, 4 or 6.

(Sure Event, Impossible Event, Impossible/uncertain Event)

Equally Likely Events

Two events are said to be equally likely if any one of them cannot be expected to occur in preference to the other.

/08/2019 esday

Algebra of Events

If A and B are two events then

1. $A^c \Rightarrow$ not A $\quad \Rightarrow A^c = U - A$
2. $A \cup B \Rightarrow$ atleast one
3. $A \cap B \Rightarrow$ both A and B
4. $A \cap B^c \Rightarrow$ only A or exactly A
5. $(A \cap B^c) \cup (A^c \cap B) \Rightarrow A \cup B \Rightarrow$ atleast one or only one or exactly one

Different Definitions of Probability

The definition of probability can be labelled as
1. Classical
2. Empirical (statistical approach) or (frequency approach)
3. Modern (Axiomatic)

1. Classical definitions of probability

Let a random experiment produce only a finite number of outcomes, say n let all these outcomes be equally likely and mutually exclusive. Let f of these outcomes be favourable to an event A then the probability of the event A is defined as the ratio between f and n.

ie. $P(A) = \frac{f}{n}$

$= \frac{\text{no. of cases favourable to A}}{\text{total possible no. of cases}}$

Ques What is the probability of selecting a boy from a class containing four boys and three girls

No. of boys = 4 $\quad$ No. of girls = 3

Total ~~poss~~ no. of students = 7

P(selecting a boy) = $\frac{\text{no. of cases favourable}}{\text{total possible no. of cases}}$

$= \frac{4}{7}$

Ques What is the chance of getting a head when a coin is tossed?

P(getting a head) = $\frac{1}{2}$

Ques Two coins are tossed, what is the probability of getting (a) both heads (b) one head (c) atleast one head (d) no head

Outcomes = {(HH),(TH),(HT),(TT)}

(a) P(getting both heads) = $\frac{1}{4}$

(b) P(getting one head) = $\frac{2}{4} = \frac{1}{2}$

(c) P(getting one head) = $\frac{3}{4}$

(d) P(getting no head) = $\frac{1}{4}$

Ques Three unbiased coins are tossed. What is the probability of obtaining (a) all heads (b) two heads (c) one head (d) atleast one heads (e) atleast two heads (f) atmost one head

Sample Space
Outcomes = {(HHH)(HHT), (HTT), (TTT) (TTH) (THH), (HTH) (THT)}

(a) P(getting all heads) = $\frac{1}{8}$

(b) P(getting two heads) = $\frac{3}{8}$

(c) P(getting one head) = $\frac{3}{8}$

(d) P(getting atleast one head) = $\frac{7}{8}$

(e) P(getting ~~two~~ atleast two heads) = $\frac{4}{8}$

(f) P(atmost one head) = $\frac{4}{8}$

Ques A die is thrown. Find the probability of getting (a) a four (b) an even number (c) 3 or 5 (d) less than 3

Sample Space = {1, 2, 3, 4, 5, 6}

(a) P(getting a four) = $\frac{1}{6}$

(b) P(getting an even number) = $\frac{\cancel{3}\,1}{\cancel{6}\,2} = \frac{1}{2}$

(c) P(getting 3 or 5) = $\frac{2}{6} = \frac{1}{3}$

(d) P(getting less than 3) = $\frac{2}{6} = \frac{1}{3}$

Ques Two unbiased dice are thrown. Find the probability (a) both the dice show the same number. (b) one dice shows 5 (c) first die shows five (d) the total of the numbers on the dice is 8 (e) total of the numbers on the dice is greater than 8. (f) a sum of 10

Sample space = {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)
(2,1) (2,2) (2,3) (2,4) (2,5) (2,6)
(3,1) (3,2) (3,3) (3,4) (3,5) (3,6)
(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)
(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)
(6,1) (6,2) (6,3) (6,4) (6,5) (6,6)

(a) P(both the dice shows the same numbes) = $\frac{6}{36} = \frac{1}{6}$

(b) P(one die shows 5) = $\frac{11}{36}$

(c) P(first die shows 5) = $\frac{6}{36} = \frac{1}{6}$

(d) P(total no. of die is 8) = $\frac{5}{36}$

(e) P(total no. of die is greater than 8) = $\frac{\cancel{10}\ 5}{\cancel{36}\ 18} = \frac{5}{18}$

(f) P(sum is 10) = $\frac{\cancel{3}\ 1}{\cancel{36}\ 12} = \frac{1}{12}$

Ques A ball is drawn from a bag containing 4 white ~~and~~ 6 black and 5 green balls. Find the probability that a ball is drawn is (a) white (b) green (c) black (d) not green (e) green or white

no. of white balls = 4
no. of black balls = 6
no. of green balls = 5
total = 15

(a) P(white ball) = $\frac{4}{15}$

(b) P(green ball) = $\frac{5}{15} = \frac{1}{3}$

(c) P(black ball) = $\frac{6}{15}$

(d) P(not green) = $\frac{10}{15} = \frac{2}{3}$

(e) P(green or white) = $\frac{9}{15}$

Ques A card is drawn from a pack of cards. What is the probability that it is (a) black card (b) a king (c) a queen (d) a spade (e) a spade king (f) a king or queen

Total = 54

(a) P(black card) = $\frac{26}{\cancel{54}\ 52}$

(b) P(a king) = $\frac{4}{52}$

(c) P(a queen) = $\frac{4}{52}$

(d) P(a spade) = $\frac{13}{52}$

(e) P(spade king) = $\frac{1}{52}$

(f) P(king or queen) = $\frac{8}{52}$

$2\overline{)52}$ 26
12
12

(13)

NOTE:-

Suppose the favourable number of cases of an event A is m and unfavourable number of cases is n then we can say odds in favour of the event are m to n and odds against the event are n to m, then probability of the event $= \frac{m}{m+n}$

Ques If odds in favour of A solving a problem are 2 to 3 and odds against B solving the same problem are 3 to 5. Find probability for (a) A solving the problem (b) B solving the problem.

Probability $= \frac{m}{m+n}$

for A

$m = 2$

$n = 3$

for B $m = 5$

$n = 3$

(a) A solving the problem $= \frac{m}{m+n} = \frac{2}{3+2} = \frac{2}{5}$

(b) B solving the problem $= \frac{m}{m+n} = \frac{5}{5+3} = \frac{5}{8}$

Results from combination

Ques If a bag contains 6 white and 4 black balls then P(drawing a white ball) $= \frac{6C_1}{10C_1} = \frac{6 \times 5}{10 \times 5} = \frac{3}{5}$

P(drawing 2 white balls) $= \frac{6C_2}{10C_2} = \frac{6 \times 5 \times 4}{10 \times 9 \times 8} = \frac{1}{6}$

Sol

P(drawing 2 white and 2 black ball) $= \frac{6C_2 \times 4C_2}{10C_4}$

$=$

P(drawing 2 white and 1 black ball) $= \frac{6C_2 \times 4C_1}{10C_4}$

Ques What is the probability of getting 3 white balls from a box containing ~~five~~ 5 white and 4 black balls

$5C_3 = \frac{5 \times 4 \times 3}{1 \times 2 \times 3}$

P(getting 3 white balls) $= \frac{5C_3}{9C_3}$

$= \frac{5 \times 4 \times 3}{9 \times 8 \times 7} = \frac{5}{6 \times 7} = \frac{5}{42}$

Ques A committee is to be consituted by selecting two people at random from a group consisting of ~~the~~ 3 economists and 4 statistations. Find the prob that the committee will consist of (a) 2 economists (b) 2 statistations (c) 1 eco and 1 stat.

(a) P (2 economists) $= \frac{3C_2}{7C_2} = \frac{3\times2\times\cancel{1}}{7\times6}$

(b) P(2 statistician) $= \frac{\cancel{4}C_2}{7C_2} = \frac{4\times3}{7\times6}$

(c) P(1 eco and ~~2~~ 1 stat) $= \frac{3C_1 \times 4C_1}{7C_2}$

$= \frac{3\times4}{\frac{7\times6}{1\times2}} = \cancel{2}$

$= \frac{3\times4\times2}{7\times6} = \frac{4}{7}$

Ques A bag contains 7 white and 9 black balls. 3 balls are drawn together. What is the prob that (a) all are black (b) all are white (c) one white & 2 black (d) 2 white & 1 black

(a) P(all are black) $= \frac{9C_3}{16C_3} = \frac{9\times8\times7}{16\times15\times14}$

$= \frac{9}{15\times4} = \frac{3}{20}$

P(all are white) $= \frac{7C_3}{16C_3} = \frac{7\times6\times5}{16\times15\times14} = \frac{3}{6\times8}$

$= \frac{1}{16}$

(c) P(1 white and 2 black) $= \frac{7C_1 \times 9C_2}{16C_3}$

$= \frac{7\times9\times8}{16\times15\times14} \frac{\times1\times2\times3}{\times2} = \frac{9\times3}{15\times2\times2}$

$= \frac{9}{10\times2} = \frac{9}{20}$

(d) P(2 white and 1 black) $= \frac{7C_2 \times 9C_1}{16C_3}$

$= \frac{7\times6\times9}{16\times15\times14} \frac{\times1\times2\times3}{\times2}$

$= \frac{6\times9}{5\times2\times16} = \frac{27}{80}$

21/08/2019
Wednesday

Ques The letters of the word ~~statistics~~ STATISTICS are written in 10 identical cards. If 2 cards are drawn at random what is the probability that (a) one S and one I will ~~occu~~ occur. (b) two I will occur.

no. of S = 3 no. of T = 3 no. of A = 1 no. of C = 1
no. of I = 2

(a) P( one S and one I) $= \dfrac{{}^3C_1 \times {}^2C_1}{{}^{10}C_2}$

$= \dfrac{3 \times 2 \times 1 \times 2}{\cancel{10}_5 \times \cancel{9}}$

$= \dfrac{2}{15}$

(b) P( two I will occur) $= \dfrac{{}^3C_2}{{}^{10}C_2}$

$= \dfrac{\cancel{3} \times 2 \times 1}{\cancel{10}_5 \times \cancel{9}_3}$

$= \dfrac{1}{15}$

NOTE

P(getting atleast one) = 1 - P(getting none)

There are 4 men and 3 women. Find the probability of selecting 3 of which (a) exactly 2 are women (b) no women (c) atleast one woman (d) atleast ~~two~~ 2 women (e) at the most 2 women

(a) P( exactly 2 are women) $= \dfrac{{}^3C_2 \times {}^4C_1}{{}^7C_3}$

$= \dfrac{3 \times 2 \times 4}{7 \times 6 \times 5} \times 1 \times 2 \times 3 = \dfrac{8 \times 3}{7 \times 5}$

$= \dfrac{6 \times 4}{7 \times 5} = \dfrac{\left(\dfrac{3 \times 2}{2}\right) \times 4}{\left(\dfrac{7 \times 6 \times 5}{3 \times 2}\right)}$

$= \dfrac{3 \times 4}{7 \times 5} = \dfrac{12}{35}$

(b) P( no women) $= \dfrac{{}^4C_3}{{}^7C_3} = \dfrac{4 \times 3 \times 2}{7 \times 6 \times 5}$

$= \dfrac{4}{35}$

(c) P( atleast one woman) = 1 - P( no women)

$= 1 - \dfrac{4}{35}$

$= \dfrac{35-4}{35} = \dfrac{31}{35}$

(d) P( ~~atleast~~ atleast 2 women) = P( 2 women or 3 women)

= P[(2 women and 1 man) or (3 woman)]

$$= \frac{3C_2 \times 4C_1 + 3C_3}{7C_3}$$

$$= \frac{\frac{3\times 2}{1\times 2}\times 4 + \frac{3\times 2\times 1}{1\times 2\times 3}}{\frac{7\times 6\times 5}{1\times 2\times 3}}$$

$$= \frac{12+1}{35}$$

$$= \frac{13}{35}$$

(c) P(atmost 2women) = P (0woman or 1woman or 2women)

= P[(0woman and 3 men) or (1woman and 2men) or (2women and 1man)]

$$= \frac{4C_3 + 3C_1 \times 4C_2 + 3C_2 \times 4C_1}{7C_3}$$

$$= \frac{\frac{4\times 3\times 2}{1\times 2\times 3} + 3\times\frac{4\times 3}{1\times 2} + \frac{3\times 2}{1\times 2}\times 4}{\frac{7\times 6\times 5}{1\times 2\times 3}}$$

$$= \frac{4+18+12}{35}$$

$$= \frac{34}{35}$$

Ques A bag contains 8 black and 4 white balls. If 5 balls are drawn at random find the chance that 3 of them are black

$$P(\text{3 of them are black}) = \frac{8C_3 \times 4C_2}{12C_5}$$

$$= \frac{\frac{8\times 7\times 6}{1\times 2\times 3} \times \frac{4\times 3}{1\times 2}}{\frac{12\times 11\times 10\times 9\times 8}{1\times 2\times 3\times 4\times 5}}$$

$$= \frac{8\times 7\times 2\times 3}{12\times 11\times 6} = \frac{14}{33}$$

Ques A sub-committee of 6 members is to formed out of a group of 7 men and 4 ladies. obtain the probability that the sub-committee will consist of (a) exactly two ladies (b) atleast 2 ladies (c) atmost 2 ladies

(a) P(exactly 2 ladies) = $\frac{4C_2 \times 7C_4}{11C_6}$

$= \frac{\frac{4\times3}{1\times2} \times \frac{7\times6\times5\times4}{1\times2\times3\times4}}{\frac{11\times10\times9\times8\times7\times6}{1\times2\times3\times4\times5\times6}} = \frac{7\times6\times5}{5\times7\times11}$

$= \frac{5}{11}$

(b) P(atleast 2 ladies) = ~~1 - P(men)~~

P(2 lady and 4 men) or (3 lady and 3 men) or (4 lady and 2 men)

$= \frac{4C_2 \times 7C_4 + 4C_3 \times 7C_3 + 4C_4 \times 7C_2}{11C_6}$

$= \frac{\frac{4\times3}{1\times2} \times \frac{7\times6\times5\times4}{1\times2\times3\times4} + \frac{4\times3\times2}{1\times2\times3} \times \frac{7\times6\times5}{1\times2\times3} + 1 \times \frac{7\times6}{1\times2}}{\frac{11\times10\times9\times8\times7\times6}{1\times2\times3\times4\times5\times6}}$

$= \frac{7\times6\times5 + 4\times7\times5 + 7\times3}{11\times2\times3\times7}$

$= \frac{210 + 140 + 21}{462}$

$= \frac{371}{462}$ $\frac{53}{66}$

(c) P(~~at least~~ atmost 2 ladies) = P(0 lady and 6 men) or (1 lady and 5 men) or (2 lady and 4 men)

$= \frac{7C_6 + 4C_1 \times 7C_5 + 4C_2 \times 7C_4}{11C_6}$

$= \frac{\frac{7\times6\times5\times4\times3\times2}{1\times2\times3\times4\times5\times6} + 4 \times \frac{7\times6\times5\times4\times3}{1\times2\times3\times4\times5} + \frac{4\times3}{1\times2} \times \frac{7\times6\times5\times4}{1\times2\times3\times4}}{\frac{11\times10\times9\times8\times7\times6}{1\times2\times3\times4\times5\times6}}$

$= \frac{7 + 4\times7\times3 + 6\times35}{66} = \frac{84+11}{66}$

$= \frac{95}{66} = \frac{84 + 210 + 7}{66}$

$\frac{238}{462} = \frac{301}{66}$

## Frequency Ratio

Consider a random experiment. Let A be an event associated with this random experiment. Let us repeat the experiment n times. Let the event A happen f out of n repetitions of the experiment, then $\frac{f}{n}$ is called frequency ratio or relative frequency of the event A.

2. ## Statistical probability or Empirical probability

If we repeat ~~a~~ random experiment a great number of times under essentially the same conditions, the limit of the ratio of the number of times that an event happens to the total number of ~~trails~~ trials as the number of trials increase indefinitely is called the probability of the happening of the event.

$\therefore P(A) = \lim_{n \to \infty} \frac{f}{n}$ where, n is the number of repeatations of the experiment

f is the number of times A happens.

### ∴ Properties of Probability

(1) Probability of an event lies between 0 and 1

2) Let S denote the sample space therefore S is a sure event, then $P(S) = 1$

(3) If A and B are two mutually exclusive events then $P(A \cup B) = P(A) + P(B)$.

Ques The following table gives a distribution of wages of 1200 workers :-

| Wage (in Rs) | 100-120 | 120-140 | 140-160 | 160-180 | 180-200 | 200-220 | 220-240 |
|---|---|---|---|---|---|---|---|
| No. of workers | 10 | 100 | 500 | 320 | 175 | 53 | 42 |

find the probability that a worker selected has wage (a) under Rs 140 (b) above 200 (c) between 140 and 200

(a) Under 140 $= \frac{10+100}{1200} = \frac{110}{1200} = \frac{11}{120}$

$= 0.09167$

(b) above 200 $= \frac{53+42}{1200} = \frac{95}{1200} =$

$= 0.07917$

(c) b/w 140 and 200 $= \frac{500+320+175}{1200} = \frac{995}{1200}$

$= 0.82917$

3. Axiomatic Approach to probability

Let S be a sample ~~sample~~ space of a random experiment, Let A be an event of the ~~standard~~ random experiment so that A is a subset of S $A \subseteq S$, then we can associate a real number $P(A)$ to the event A.

This number $P(A)$ will be called probability of A if it satisfies the following three axioms.

Axiom 1: $P(A)$ is a real number such that $P(A) \geq 0$ for every $A \subseteq S$

Axiom 2: $P(S) = 1$ where S is a sample space

Axiom 3: $P(A \cup B) = P(A) + P(B)$ where A and B are two non-intersecting subset of S.

Addition and Multiplication rules of probability

Addition rule of probability

Additional rule for mutually exclusive events.

If A and B are two mutually exclusive events, then the probability for A or B to happen is the sum of their probabilities i.e. probability of $(A \cup B) = P(A) + P(B)$ if A and B are disjoint.

Proof:-
Let $n(A)$ be the number of elementary outcomes in A and $n(B)$ the number of elementary outcomes in B and $n(A \cup B)$ be the number of outcomes in $A \cup B$.

Let $n(S)$ be the number of elementary outcomes in the sample space, then

$$P(A \cup B) = \frac{n(A \cup B)}{n(S)}$$

$$P(A) = \frac{n(A)}{n(S)} \quad ; \quad P(B) = \frac{n(B)}{n(S)}$$

$\because$ A and B are mutually exclusive they are disjoint

$$\therefore A \cap B = \phi$$

$$n(A \cup B) = n(A) + n(B)$$

Consider $P(A \cup B) = \frac{n(A \cup B)}{n(S)} = \frac{n(A) + n(B)}{n(S)}$

$$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)}$$

$$= P(A) + P(B)$$

$$\therefore P(A \cup B) = P(A) + P(B)$$

Addition rule for any two events (not mutually exclusive)

If A and B are two events then the probability of A or B ~~to~~ happen is the sum of their probabilities - probability for both to happen

ie $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
[If A and B are not disjoint]

Proof:-

~~If A and~~:

$n(A \cup B) = n(A) + n(B) - n(A \cap B)$

$P(A \cup B) = \frac{n(A \cup B)}{n(S)}$

$= \frac{n(A) + n(B) - n(A \cap B)}{n(S)}$

$= \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$

$= P(A) + P(B) - P(A \cap B)$

ie $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

22/08/2019

Ques If A and B are two mutually exclusive events $P(A) = 0.45$ and $P(B) = 0.35$, find A or B. $(A \cup B)$

$P(A \cup B) = P(A) + P(B)$
$= 0.45 + 0.35$
$= \underline{0.8}$

Ques A card is drawn at random from an ordinary pack of 52 cards. find the prob. that the card drawn is either spade ~~of~~ or diamond.

$P(\text{card is spade})$ or $P(A)$ $= \frac{13}{52}$.

$P(\text{card is diamond})$ or $P(B)$ $= \frac{13}{52}$

$P(\text{card is spade or diamond}),\ P(A \cup B) = P(A) + P(B)$
$= \frac{13}{52} + \frac{13}{52}$

$P(A \cup B) = \frac{26}{52} = \underline{\frac{1}{2}}$

Ques Find the probability of getting a total of 7 or 11 in a single throw with 2 dice

$P(A) \Rightarrow P(\text{getting } 7) = (1,6)\ (2,5)\ (3,4)\ (4,3)\ (5,2)\ (6,1)$

$P(A) = \frac{6}{36}$

$P(B) \Rightarrow P(\text{getting } 11) = (6,5)\ (5,6)$

$P(B) = \frac{2}{36}$

$\therefore P(A \cup B) = P(A) + P(B)$

$= \frac{6}{36} + \frac{2}{36}$

$= \frac{8}{36} = \frac{2}{9}$

Ques If $P(A) = 1/5$, $P(B) = 1/4$ and $P(A \cap B)$ or $P(A \text{ and } B) = 1/10$. Find $P(A \cup B)$ or $P(A \text{ or } B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= \frac{1}{5} + \frac{1}{4} - \frac{1}{10}$

$P(A \cup B) = \frac{4+5-2}{20}$

$P(A \cup B) = \frac{7}{20}$

Ques Find the probability of drawing an ace or a spade from a pack of cards

$P(B) \Rightarrow P(\text{drawing a spade}) = \frac{13}{52}$

$P(A) \Rightarrow P(\text{drawing ace}) = \frac{4}{52}$

$P(A \cap B) = \frac{1}{52}$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$

$P(A \cup B) = \frac{16}{52} = \frac{4}{13}$

Ques If $P(A) = 1/13$, $P(B) = 1/4$ and $P(A \cup B) = 4/13$ find $P(A \cap B)$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\frac{4}{13} = \frac{1}{13} + \frac{1}{4} - P(A \cap B)$

$P(A \cap B) = \frac{1}{13} + \frac{1}{4} - \frac{4}{13}$

$= \frac{-3}{13} + \frac{1}{4}$

$= \frac{-12+13}{52} = \frac{1}{52}$

Ques The probability that a contractor will get a plumbing contract is 2/3. and the probability that he will not get an electric contract is 5/9. If the probability of getting atleast one contract is 4/5. what is the prob. that he will ~~both~~ get both the contracts.

$P(A) \Rightarrow P(\text{get a plumbing contract}) = 2/3$

$P(B) \Rightarrow P(\text{get a electric contract}) = 1 - \frac{5}{9}$

$= \frac{4}{9}$

$P(A \cup B) = 4/5$

$P(A \cap B) = ?$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\frac{4}{5} = \frac{2}{3} + \frac{4}{9} - P(A \cap B)$

$P(A \cap B) = \frac{2}{3} + \frac{4}{9} - \frac{4}{5}$

$= \frac{6+4}{9} - \frac{4}{5}$

$= \frac{10}{9} - \frac{4}{5} = \frac{50-36}{45} = \frac{14}{45}$

Ques The probability of a student passing statistics is 2/3. The probability that he passes both statistics and accountancy 14/45. The probability that he passes atleast 1 test is 4/5. what is the probability that he passes the accountancy test

$P(A) = 2/3$ $\qquad$ $P(B) = ?$

$P(A \cap B) = 14/45$ $\qquad$ $P(A \cup B) = 4/5$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\frac{4}{5} = \frac{2}{3} + P(B) - \frac{14}{45}$

$P(B) = \frac{4}{5} + \frac{14}{45} - \frac{2}{3}$

$= \frac{36+14}{45} - \frac{2}{3}$

$= \frac{50}{45} - \frac{2}{3}$

$= \frac{10-6}{9}$

$P(B) = \frac{4}{9}$

Ques The probability that a student passes maths is 2/3 and statistics is 4/9. If the probability of passing atleast one subject is 4/5, what is the probability that he passes both the subject

$P(A) = 2/3$

$P(B) = 4/9$

$\cancel{P(C)} = P(A \cup B) = 4/5$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$\frac{4}{5} = \frac{2}{3} + \frac{4}{9} - P(A \cap B)$

$P(A \cap B) = \frac{2}{3} + \frac{4}{9} - \frac{4}{5}$

$= \frac{30 + 20 - 36}{45}$

$P(A \cap B) = \frac{14}{45}$

Ques Two six face dices are drawn simultaneously if $x$ denotes the sum of numbers appearing on the upper face of the dice find the probability that $x$ is
(a) either 8 or 9
(b) neither 7 or 11

(a) $S = \{(2,6)\ (4,4)\ (6,2)\ (6,2)\ (3,6)\ (4,5)\ (5,4)\ (3,6)\ (6,3)\}$

$P(\text{either } 8 \text{ or } 9) = \frac{9}{36} = \frac{1}{4}$

(b) $S = \{(1,6)\ (6,1)\ (2,5)\ (5,2)\ (3,4)\ (4,3)\ (5,6)\ (6,5)$

$P(\text{either } 7 \text{ or } 11) = \frac{8}{36}$

$P(\text{neither } 7 \text{ or } 11) = 1 - \frac{8}{36}$

$= \frac{36-8}{36} = \frac{28}{36}$

$P(\text{neither } 7 \text{ or } 11) = \frac{7}{9}$

Conditional Property

Probability of an event A given that B has happened is called the conditional probability of A given B and is denoted by $P(A/B)$

$P(A/B) = \frac{P(A \cap B)}{P(B)}$

$P(B/A) = \frac{P(A \cap B)}{P(A)}$

eg:- consider a family with 2 children, the different outcomes are (BB) (BG) (GB) and (GG). If it is known that first is a boy the outcome are only (BB) and (BG)

So that probability for both boys $= 1/2$. This is under the condition first is a boy

$\therefore P(\text{both boys} / \text{first is a boy}) = P(A/B) = 1/2$

This is called the conditional probability as the condition is that first is boy. The condition is not given probability of both boys $= 1/4$

Proof :-

Let $n(A)$ and $n(B)$ be the elementary outcome in the events A and B respectively. Let $n(A \cap B)$ be the elementary outcomes in the event $A \cap B$. Let $n(S)$ be the elementary outcomes in the event $A \cap B$. Let $n(S)$ be the elementary outcomes in the sample space. Then,

$$P(A) = \frac{n(A)}{n(S)}$$

$$P(B) = \frac{n(B)}{n(S)}$$

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)}$$

Under the condition that B has occured the outcomes in the sample space is reduced $n(B)$ only and outcome in A is reduced to $n(A \cap B)$.

$$P(A/B) = \frac{n(A \cap B)}{n(B)}$$

divide the numerator and denomenator by $n(S)$

$$= \frac{n(A \cap B) / n(S)}{n(B) / n(S)}$$

$$= \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Ques If $P(A) = 1/13$, $P(B) = 1/4$ and $P(A \cap B) = 1/52$ find (i) $P(A/B)$ (ii) $P(B \cap A)$

(i) $P(A/B) = \frac{P(A \cap B)}{P(B)}$.

$$= \frac{\frac{1}{52}}{\frac{1}{4}}$$

$$= \frac{1}{52} \times \frac{4}{1} = \frac{1}{13}$$

(ii) $P(B/A) = \frac{P(A \cap B)}{P(A)}$

$$= \frac{\frac{1}{52}}{\frac{1}{13}}$$

$$= 1/4$$

## Independence of two events

If A and B event are such that $P(A/B) = P(B)$ and $P(B/A) = P(B)$, we say that A and B are independent. When $P(A/B) = P(A)$ we have $P(A \cap B) = P(A) \times P(B)$

## Multiplication Rule of Probability

(a) Multiplication rule for any two events

If A and B are any two events then probability for both A and B to take place together denoted by $P(A \cap B) = P(A)\ P(B/A)$ provided $P(A) \neq 0$ $P(B)\ P(A/B)$ provided $P(B) \neq 0$

proof

If A and B are any two events then $P(B/A) = \frac{P(A \cap B)}{P(A)}$ by conditional probability

$\therefore P(A \cap B) = P(A)\ P(B/A)$

Similarly $P(A/B) = \frac{P(A \cap B)}{P(B)}$

$$P(A \cap B) = P(B)\ P(A/B)$$

(b) Multiplication rule for two independent events

If A and B are two independent events then the probability for both A and B ~~to~~ to happen is the product of their probability i.e. $P(A \cap B) = P(A)\ P(B)$ if A and B are independent

As we know multiplication theorem any two events A and B is $P(A \cap B) = P(A)\ P(B/A)$. We know that when A and B are independent so $P(B/A) = P(B)$ $\therefore P(A \cap B) = P(A)\ P(B)$

Ques If $P(A) = 4/5$ $P(B) = 3/5$. Find $P(A \cap B)$ if A and B are independent

$$P(A \cap B) = P(A)\ P(B)$$

$$= \frac{4}{5} \times \frac{3}{5}$$

$$= \frac{12}{25}$$

Ques If $P(A) = 2/5$ and $P(B) = 3/8$, $P(A \cap B) = 1/20$ examine whether A and B are independent

$P(A) \times P(B) = \frac{2}{5} \times \frac{3}{8}$

$= \frac{6}{40}$

$P(A \cap B) = 1/20$

$\because P(A \cap B) \neq P(A) \times P(B)$

$\therefore$ A and B are not independent

Ques Probability that A will pass a paper 1 is ·3 and probability that he will pass paper 2 is ·7, what is the probability that he will pass both paper assuming that 2 papers are independent

$P(P_1 \cap P_2) = \frac{\cdot 3}{10} \times \frac{\cdot 7}{10} = \frac{\cdot 21}{100} = 0.21$

Ques If $P(A) = 0.4$ and $P(B) = 0.3$ $\cdot P(A \cap B) = 0.2$
Find the probability of
(i) atleast one event to occur
(ii) exactly one of the event occur
(iii) None of the events to occur

(i) $P(A \cup B) = 0.4 + 0.3 - 0.2$
$= 0.5$

(ii) $P[(A \cap B') \cup (B \cap A')] = P[(A) - P(A \cap B)] + P[(B) - P(A \cap B)]$

$= (0.4 - 0.2) + (0.3 - 0.2)$
$= 0.2 + 0.1$
$= 0.3$

(iii) $P(\text{none}) = 1 - P(A \cup B)$
$= 1 - 0.5$
$= 0.5$

Ques The probability that A solves a problem in stat. is 2/5 and the probability that B solves it is 3/8. If they try independently, find the probability that
(a) Both solves the problem
(b) None solves the problem

(a) $P(A \cap B) = \frac{2}{5} \times \frac{3}{8} = \frac{6}{40}$

(b) $P(A \cup B) = \frac{2}{5} + \frac{3}{8} - \frac{6}{40}$
$= \frac{16 + 15 - 6}{40}$
$= \frac{31 - 6}{40}$
$= \frac{25}{40} = \frac{5}{8}$

(c) $P(\text{none}) = 1 - P(A \cup B)$

$= 1 - \frac{5}{8} = \frac{8-5}{8} = \frac{3}{8}$

Ques An article manufactured by a company consists of 2 parts A and B. In the process of manufacture of part A 9 out of 100 are likely to be defective similarly 5 out of 100 are likely to be defective in the manufacture of part B. Calculate the probability that the assembled part will not be defective?

P(part A not to be defective) $= P(A')$

$= 1 - P(A)$

$= 1 - \frac{9}{100}$

$= \frac{91}{100}$

P(part B not be defective) $= P(B')$

$= 1 - \frac{5}{100}$

$= \frac{95}{100}$

$P(A') \times P(B') = \frac{91}{100} \times \frac{95}{100} = \frac{8645}{10000}$

Ques A university has to select an examiner from a list of 50 persons. 20 of them are women and 30 are men. 10 of them know hindi and 40 do not. 15 of them are teachers and remaining are not. What is the probability of the university selecting a hindi knowing women teacher?

P(selecting a hindi knowing person) $= P(A)$

$= \frac{10}{50}$

$= \frac{1}{5}$

P(selecting women) $= P(B) = \frac{20}{50} = \frac{2}{5}$

P(selecting a teacher) $= P(C)$

$= \frac{15}{50}$

$= \frac{3}{10}$

P(selecting a hindi knowing women teacher)

$= P(A) \times P(B) \times P(C)$

$= \frac{10}{50} \times \frac{20}{50} \times \frac{15}{20} = \frac{3000}{125000}$

$= \frac{3}{125}$

Ques Two persons A and B attempt independently to solve a puzzle. The probability that A will solve is 3/5 and the probability that B will solve is 1/3. Find the probability that the puzzle will be solved by atleast one of them.

$P(\text{atleast one}) = 1 - P(\text{none})$

$= 1 - [P(A') \cap P(B')]$

$P(A) = \frac{3}{5}$

$1 - \frac{3}{5} = \frac{2}{5}$

$P(B) = 1/3$

$1 - \frac{1}{3} = \frac{2}{3}$

$P(\text{atleast one}) = 1 - P(\text{none})$

$= 1 - \left[\frac{2}{5} \times \frac{2}{3}\right]$

$= 1 - \frac{4}{15}$

$= 11/15$

Ques The odds against x solving the business statistic problem are 8-6 and odds in favour of student y solving the same problem are 14-16. what is the probablity that the (i) problem is solved (ii) problem is not solved

$P(X) = 6/14$

$P(Y) = 14/30$

$P(X') = 1 - \frac{6}{14} = \frac{14-6}{14} = \frac{8}{14}$

$P(Y') = 1 - \frac{14}{30} = \frac{30-14}{30} = \frac{16}{30}$

(i) $P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$

$= \frac{6}{14} + \frac{14}{30} - \left(\frac{16}{14} \times \frac{14}{30}\right)$

$= \frac{180 + 196 - 84}{420} = \frac{292}{420} = \frac{73}{105}$

(ii) $P(X' \cap Y') = P(X') \times P(Y')$

$= \frac{8}{14} \times \frac{16}{30} = \frac{128}{420}$

$= \frac{32}{105}$

Ques A bag contains 10 items which 4 are defective, 3 items are drawn one after another without replacement. what is the probability that all the three items drawn are defective

P( all the three item are drawn defective)

$= \frac{4}{10} \times \frac{3}{9} \times \frac{2}{8}$

$= \frac{24}{720}$

$= \frac{1}{30}$

Ques Probability that a patient is correctly diagonised is 0.4. If a patient is correctly diagonised the probability to survive is 0.8. what is the probability that a patient is correctly diagonized and survived.

$P(A) = 0.4$

$P(B/A) = 0.8$

$P(A \cap B) = P(A) \times P(B/A)$

$= 0.4 \times 0.8$

$= 0.32$

$(A \cap B) = \frac{3}{4} - \frac{4}{7} = \frac{3}{7}$

Ques A and B worked independently on a problem the probability that A will solve the problem is 3/4 and B will solve is 4/7.

what is the probability that

i) The problem will be solved

ii) The problem will not be solved

$P(A) = 3/4$

$P(B) = 4/7$

$P(A \cap B) = \frac{3}{4} - \frac{4}{7} = \frac{3}{7}$

(i) Problem will be solved $= P(A \cup B)$

$= P(A) \cancel{*} P(B) - P(A \cap B)$

$= \frac{3}{4} \times \frac{4}{7} \cancel{*} \frac{3}{7}$

$= \frac{25}{28}$

(ii) problem will not be solved

$= 1 - \frac{25}{28}$

$= \frac{28 - 25}{28}$

$= \frac{3}{28}$

26/08/2019
Monday

Probability

Addition rule

(a) Mutually exclusive
$P(A \cup B) = P(A) + P(B)$

(b) Not mutually exclusive
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- P (either A or B)
- P(atleast one)

Multiplication rule

(a) Independent event
$P(A \cap B) = P(A) \times P(B)$

(b) Dependent event
$P(A \cap B) = P(A)\, P(B/A)$

- P(both)
- P( A aswellas B)

0 Problems based on both addition and multiplication

Ques A speaks truth in 70% of cases and B in 85% of cases. In what percentage of cases are they likely to contradict each other in stating the same fact?

$P(A) = \frac{70}{100}$ $\quad P(A') = \frac{30}{100}$

$P(B) = \frac{85}{100}$ $\quad P(B') = \frac{15}{100}$

P(A speaks truth and B lies or A lies and B speaks truth)

$P(A \cap B') + P(B \cap A')$

Independent

$= P(A) \times P(B') + P(B) \times P(A')$

$= \frac{70}{100} \times \frac{15}{100} + \frac{85}{100} \times \frac{30}{100}$

$= \frac{1050 + 2550}{10000}$

$= \frac{3600}{10000}$

$= \frac{36}{100}$

$= \frac{9}{25}$

Percentage in cases which they contradict $= 0.36 \times 100$ $= \cancel{0.36 \times 100}$ $= 36\%$

Ques 20% of all students in a university are graduates and 80% are undergraduates. The probability that graduate students is married is 0.50 and the probability that an undergraduate student is married is 0.10. One student is selected at random. What is the probability that the student selected is married?

Dependent

P[(graduate student and is married) or ~~B~~ (under graduate and is married)]

$P(A) = \frac{20}{100}$ $\quad P(B) =$ (married graduate)

$P(C) = \frac{80}{100}$ $\quad P(D) =$ (married under grad)

$P(B/A) = 0.50$

$P(D/C) = 0.10$

$= P[(A \cap B) \cup (C \cap D)]$

$= P(A) \times P(B/A) + P(C) \times P(D/C)$

$= \left(\frac{20}{100} \times 0.50\right) + \left(\frac{80}{100} \times 0.10\right)$

$= \frac{10}{100} + \frac{8}{100} = \frac{18}{100}$

$= 0.18$

Ques An urn A contains 4 white and 3 red marbles and B contains 2 white and 5 red marbles. One of the urn to be chosen at random and a marble is to selected from the chosen urn. What is the probability of drawing a white marble?

urn A: 4W, 3R    urn B: 2W, 5R

$P(A) = 1/2$ (urn A)    $P(B) = 1/2$ (urn B)

$P(C)$: P(drawing a white marble from urn A) $= 4/7$

$P(D)$: P(drawing a white marble from urn B) $= 2/7$

$\Rightarrow$ P[drawing a white marble from one of the urn]

= P[(urn A and white marble) or (urn B and white marble)]

$= P(A \cap C) + P(B \cap D)$

$= P(A) \cdot P(C) + P(B) \cdot P(D)$

$= \frac{1}{2} \times \frac{4}{7} + \frac{1}{2} \times \frac{2}{7}$

$= \frac{2}{7} + \frac{1}{7} = 3/7$

Ques A bag contains 8 balls identical except for colour of which 5 are red and 3 white. A man draws 2 balls at random one after another without replacement. What is the probability that one of the balls drawn is white and other is red? What would be the value of these probabilities if a ball drawn is replaced before another ball is drawn?

$P(C) = 5/8$    $P(A) = 3/8$

without replacement

P[(one of balls drawn is white and other is red) or (one is red and the other is white)]

$= P(A \cap B) \cup P(C \cap D)$

$= P(A) \times P(B/A) + P(C) \times P(D/C)$

$= \left(\frac{3}{8} \times \frac{5}{7}\right) + \left(\frac{5}{8} \times \frac{3}{7}\right)$

$= \frac{15 + 15}{56} = \frac{30}{56} = \frac{15}{28}$

with replacement

$= P(A \cap B) \cup P(C \cap D)$

$= P(A) \times P(B/A) + P(C) \times P(D/C)$

$= \frac{3}{8} \times \frac{5}{8} + \frac{5}{8} \times \frac{3}{8}$

$= \frac{15 + 15}{64} = \frac{30}{64} = \frac{15}{32}$

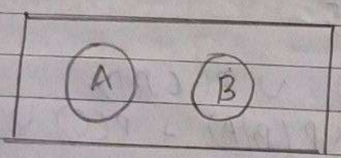
## Mutually Exclusive Events

A set of events are said to be mutually exclusive if the occurrence of one of them excludes the possibility of the occurrence of the others. Two mutually exclusive events cannot occur simultaneously in the same trail.

eg1: Getting an ace and getting a king when a card is drawn from a pack of cards are mutually exclusive

eg 2: getting head and getting tail when a coin is tossed are mutually exclusive

NOTE: If A and B are mutually exclusive, then $A \cap B = \phi$ ie the two sets are disjoint




## Exhaustive events

A group of events is said to be exhaustive when it includes all possible outcomes of the random experiment under consideration

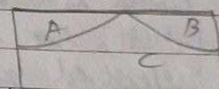
That is, if a set of events are exhaustive at least one of them will happen in any trail of the random experiment.

eg:- When a dice is thrown, the outcome 1, 2, 3, 4, 5 and 6 will together form exhaustive event and one of them will occur when a die is thrown.

## Mutually Exclusive and Exhaustive Event

A set of events are said to be mutually exclusive and exhaustive if one of them must and only one can occur

ie. A, B, and C are said to be mutually exclusive and exhaustive. If they are
(i) mutually exclusive
(ii) mutually exhaustive




eg:- In tossing two coins, getting two heads and getting 3 heads. are mutually exclusive and exhaustive.

NOTE: A, B, and C are mutually exclusive and exhaustive then, $A \cap B = \phi$, $A \cap C = \phi$, $B \cap C = \phi$ and $S = A \cup B \cup C$

## Independent events

Two or more events are said to be independent if the happening of one of them in no way affects the occurrence of the other or the others.

eg:- In the tossing of a coin twice, the result of the second tossing is not affected by the result of the first toss.

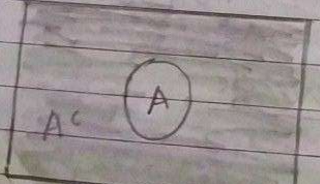
**Dependent events**

Two or more events are said to be dependent if the happening of one of them affects the happening of the other. In the case of dependent events the chance of one event depends on the happening of the other event.

eg:- From a pack of 52 cards if one card is drawn then 51 cards are left. If another card is drawn without replacing the first, the chance of the second draw is affected by the first law.

$\therefore$ drawing a king first and without replacing & drawing again a kind are dependent events.

Complement of event A ( Event 'not A')




The event A and the event 'not A' are called complementary events. If A is an event then 'not A' is the complement of A.

$A^c = U - A$ where 'U' stands for sample space. A and $A^c$ belong to the same random experiment.

eg1: In tossing a coin getting head and getting tail are complementary events ie A = getting head and AC = getting tail.

eg2: Getting at least one head and getting no head, when two coins are tossed, are complementary events. ie A = at least one head, AC = no head

The outcomes of $A^c$ will be those outcomes of the sample space which are not in A.

**Union of two events (At least one) (A or B)**

The union of two events A and B denoted by $A \cup B$ is the set of sample points in A or in B or in both.

eg:- A = getting a multiple of 5, B = getting a multiple of 3 then $A \cup B$ = getting a multiple of 5 or 3.

fig 1

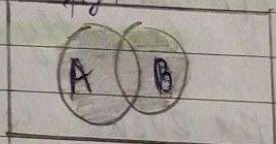
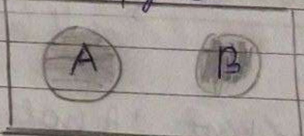



fig 2




In fig 1. A and B are intersecting

In fig 2, A and B are not intersecting

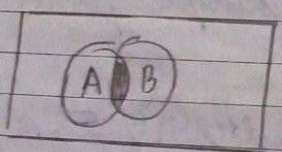
In both the cases, $A \cup B$ is shaded
$A \cup B$ stands for atleast one among the events A and B

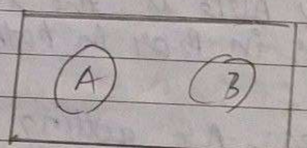
## Intersection of two events (both A and B)

The intersection of two events A and B denoted by $A \cap B$ is the set of sample space of points common to both A and B.

eg: A = getting a multiple of 5. The
$A \cap B$ = getting a multiple of 15

In fig 1, A and B are intersecting $A \cap B$ is shaded
In fig 2, A and B are not intereseciting
$\therefore A \cap B = \phi$



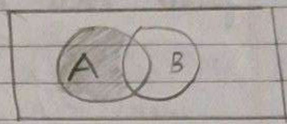




## Difference of two events (A and not B) (only A) (exactly A)

The event 'A not B' is the event whose outcomes are those belonging to A, but not B. $\therefore$ 'A not B' excludes from A outcomes common to A and B

$\therefore$ A not B = $A \cap B' = (A - B) = A - (A \cap B)$

In fig 1, it is shaded

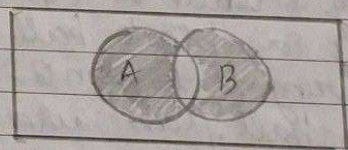



## exactly one (symmetric difference) of two events

If A and B are two events exactly one of them is the event whose outcomes are in A only or in B only

That is, the outcomes common to A and B are excluded.

$\therefore$ exactly one of A and B
$= (A \cap B') \cup (B \cap A')$
or $(A \cup B) - (A \cap B)$ - shaded portion



Ques An urn A contains 2 white and 4 black balls. Another urn B contains 5 white and 7 black balls. A ball is transferred from urn A to urn B. Then a ball is drawn from urn B. Find the prob. that it will be white

urn A: 2W, 4B  urn B: 5W, 7B

P(white ball from urn A) = 2/6  [P(A)]

P(black ball from urn A) = 4/6  [P(C)]

P(drawing a white ball from urn B if the transferred ball is white) = $\frac{6}{13}$  [P(B/A)]

P(drawing a white ball from urn B if the transferred ball is black) = 5/13  [P(D/C)]

P(drawing a white ball from urn B)
= P(drawing a white ball from urn A and transferring it to urn B then drawing a white ball from urn B or drawing a black ball from urn A transferring it to urn B then drawing a white ball from urn B)

$= P[(A \cap B) \cup (C \cap D)]$

$= P(A) \times P(B/A) + P(C) \times P(D/C)$

$= \frac{2}{6} \times \frac{6}{13} + \frac{4}{6} \times \frac{5}{13}$

$= \cancel{\frac{2}{13} + \frac{4 \times 4}{13 \times 3}}$

$= \cancel{\frac{6+16}{39} = \frac{22}{39}}$

$= \frac{2}{13} + \frac{10}{39}$

$= \frac{6+10}{39} = \frac{16}{39}$

Ques One bag contains 4 white and 2 black balls another contains 3 white and 5 black balls. One ball is drawn from each bag. find the prob. that both are of the same color.

Bag A: 4W, 2B  Bag B: 3W, 5B

P(ball drawn from bag A is white) = $\frac{4}{6}$  [P(A)]

P(ball drawn from bag A is black) = $\frac{2}{6}$  [P(C)]

P(ball drawn from bag B is white) = $\frac{3}{8}$  [P(B)]

$P(D)$

P( ball drawn from bag B is black) $= \frac{5}{8}$

P( ball drawn from each bag has same colour) = P( white ball from bag A and white ball from bag B) + P( black ball from bag A + black ball from bag B)

$= P(A \cap B) \cup P(C \cap D)$

$= P(A) \times P(B) + P(C) \times P(D)$

$= \frac{4}{6} \times \frac{3}{8} + \frac{2}{6} \times \frac{5}{8}$

$= \frac{12+10}{48}$

$= \frac{22}{48} = \frac{11}{24}$

**Ques** A candidate is selected for interview for 3 posts. For the first post, there are 3 candidates. For the second, there are 4 candidates and for the third, there are 2 candidates. What are the chances of him getting atleast one post. ?

P( getting first post) $= \frac{1}{3}$

P( getting second post) $= \frac{1}{4}$

P(getting third post) $= \frac{1}{2}$

P( not been selected for 1st post) $= 1 - \frac{1}{3} = \frac{2}{3}$

P(not been selected for 2nd post) $= 1 - \frac{1}{4} = \frac{3}{4}$

P(not been selected for 3rd post) $= 1 - \frac{1}{2} = \frac{1}{2}$

P(atleast one post) = 1 − P( none)

$= 1 - \left(\frac{2}{3} \times \frac{3}{4} \times \frac{1}{2}\right)$

$= 1 - \frac{1}{4}$

P(atleast one post) $= \frac{3}{4}$

**Ques** 3 persons A, B and C are simultaneously shooting a target. Probability of A hitting the target is 1/4. B is 1/2. C is 2/3. Find the probability that (i) exactly one of them will hit the target (ii) atleast one of them will hit the target.

(i) P( exactly one of them hit the target)

= P( A hits, B and C does not) or (B hits, A and C does not) or (C hits, A and B does not)

P =

$P(A \text{ hits}) = \frac{1}{4}$ $\quad P(A \text{ not hit}) = \frac{3}{4}$

$P(B \text{ hits}) = \frac{1}{2}$ $\quad P(B \text{ not hit}) = \frac{1}{2}$

$P(C \text{ hits}) = \frac{2}{3}$ $\quad P(C \text{ not hit}) = \frac{1}{3}$

(i) $P(\text{exactly one of them hit})$

$= P[A \cap (B' \cap C')] \cup P[B \cap (A' \cap C')] \cup P[C \cap (A' \cap B')]$

$= P[A \cap (B' \cap C')] + P[B \cap (A' \cap C')] + P[C \cap (A' \cap B')]$

$= P(A) \times P(B') \times P(C') + P(B) \times P(A') \times P(C') + P(C) \times P(A') \times P(B')$

$= \frac{1}{4} \times \frac{1}{2} \times \frac{1}{3} + \frac{1}{2} \times \frac{3}{4} \times \frac{1}{3} + \frac{3}{3} \times \frac{3}{4} \times \frac{1}{3}$

$= \frac{1+3+6}{24}$

$= \frac{10}{24} = 5/12$

(ii) $P(\text{atleast one}) = 1 - P(\text{none})$

$= 1 - (P(A') \times P(B') \times P(C'))$

$= 1 - \left[\frac{3}{4} \times \frac{1}{2} \times \frac{1}{3}\right]$

$= 1 - \frac{1}{8}$

$= \frac{7}{8}$

31/08/2019
Saturday

NOTE:
A number of events are said to be mutually exclusive and exhaustive if one of them must and only one can happen. Also sum of the probabilities of events is 1.

Ques If A, B and C are mutually exclusive and exhaustive events $P(A) = 1/2 \; P(B)$ and $P(A) = 1/3 \; P(C)$. Find $P(A)$, $P(B)$ and $P(C)$.

$P(A) = 1/2 \; P(B)$

$\therefore P(B) = 2P(A)$

$P(A) = 1/3 \; P(C)$

$\therefore P(C) = 3P(A)$

$P(A) + P(B) + P(C) = 1$

$P(A) + 2P(A) + 3P(A) = 1$

$6P(A) = 1$

$P(A) = \frac{1}{6}$

$P(B) = 2P(A)$ $\quad$ $P(C) = 3(P(A)$

$P(B) = 2 \times \frac{1}{6}$ $\quad$ $P(C) = 3 \times \frac{1}{6}$

$= \frac{1}{3}$ $\quad$ $= \frac{1}{2}$

## Probability For Three Events

- If A, B and C are any three events, then
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$$

Proof:

$$P(\underbrace{A \cup B}_{A} \cup \underbrace{C}_{B}) = P(A \cup B) + P(C) - P((A \cup B) \cap C)$$

$$= P(A) + P(B) - P(A \cap B) + P(C) - P[\underbrace{(A \cap C)}_{A} \cup \underbrace{(B \cap C)}_{B}]$$

$$= P(A) + P(B) - P(A \cap B) + P(C) - [P(A \cap C) + P(B \cap C) - P[\underbrace{(A \cap C) \cap (B \cap C)}_{A \cap B \cap C}]]$$

$$= P(A) + P(B) - P(A \cap B) + P(C) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

- If A, B, C are mutually exclusive then

$P(A \cup B \cup C) = P(A) + P(B) + P(C)$

$\Rightarrow$ $P(A \cap C)$, $P(A \cap B)$, $P(B \cap C)$ and $P(A \cap B \cap C)$ are 0.

Also

$$P(A \cup B \cup C) = 1 - P(A' \cap B' \cap C')$$

- If A, B, C are any three events then

$$P(A \cap B \cap C) = P(A) \times P(B/A) \times P(C/A \cap B)$$

- If A, B, C are mutually independent event

$$P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$$

Que A problem is A, B and C whose chances in solving it are 1/2, 1/3 and 1/7 resp. If they all try independently, what is the

probability that (i) all solve the problem.
(ii) the problem is not solved by anyone of them. (iii) The problem will be solved atleast by one

(i) $P(A \cap B \cap C) = P(A) * P(B) \times P(C)$

$= \frac{1}{2} \times \frac{1}{3} \times \frac{1}{4}$

$= \frac{1}{24}$

(ii) $P(A' \cap B' \cap C') = 1 - [P(A') \times P(B') \times P(C')]$

$P(A') = 1 - \frac{1}{2} = \frac{1}{2}$

$P(B') = 1 - \frac{1}{3} = 2/3$

$P(C') = 1 - \frac{1}{4} = 3/4$

$P(A' \cap B' \cap C') = 1 - \frac{1}{2} \times \frac{2}{3} \times \frac{3}{4}$

$= 1 - \frac{1}{4}$

$= 3/4$

(iii) $P(A \cup B \cup C) =$

Ques A speaks truth in 80% cases, B speaks truth in 70% cases while C speaks truth in 60% cases. In what percentage of cases they are similar in stating the fact.

$P(A) = \frac{80}{100}$ $P(B) = \frac{70}{100}$ $P(C) = \frac{60}{100}$

$P(A') = 1 - \frac{80}{100} = \frac{20}{100}$

$P(B') = 1 - \frac{70}{100} = \frac{30}{100}$

$P(C') = 1 - \frac{60}{100} = \frac{40}{100}$

all true $P(A \cap B \cap C) = P(A) \times P(B) * P(C)$

$= \frac{80}{100} \times \frac{70}{100} \times \frac{60}{100}$

$= \frac{336}{1000} = 0.336$

all lies $P(A' \cap B' \cap C') = P(A') \times P(B') \times P(C')$

$= \frac{20}{100} \times \frac{30}{100} \times \frac{40}{100}$

$= \frac{24}{1000} = 0.024$

Percentage in which they are similar in stating the fact
P(A speaks truth B speaks truth C speaks truth) or P(A speaks lies B speaks lies C speaks lies)

$$= P\left[(A \cap B \cap C) \cup (A' \cap B' \cap C')\right]$$

$= 0.336 + 0.024$

$= .360$

$= 0.360 \times 100$

$= 36\%$

imp

## Bayes Theorem

If an event A can happen only if one or the other of a set of mutually exclusive events $B_1, B_2, B_3 \ldots B_n$ happens, $P(B_i) \neq 0$ and $i = 1, 2, 3 \ldots n$

Then $P(B_i/A) = \dfrac{P(B_i)\, P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\, P(A/B_i)}$

**Proof:**

By the definition of conditional probability $P(A/B_i) = \dfrac{P(A \cap B_i)}{P(B)}$

$P(A \cap B_i) = P(B_i) \cdot P(A/B_i)$ —— (1)

$= P(A) \cdot P(B_i/A)$

$P(B_i/A) = \dfrac{P(A \cap B_i)}{P(A)}$

$= \dfrac{P(B_i) \cdot P(A/B_i)}{P(A)}$ —— (1)

Now A can happen if and only if (iff) or the other of the set of mutually exclusive events $B_1, B_2 \ldots B_n$ happens
So A can happen in n mutually exclusive ways $A \cap B_1, A \cap B_2 \ldots A \cap B_n$

ie $A = A \cap B_1 \cup A \cap B_2 \cup A \cap B_3 \ldots \cup A \cap B_n$

$\therefore P(A) = P(A \cap B_1) + P(B \cap B_2) + \ldots P(A \cap B_n)$

$= \sum_{i=1}^{n} P(A \cap B_i)$

ie $P(A) = \sum_{i=1}^{n} P(B_i)\, P(A/B_i)$ —— (2)

Substituting (2) in (1) we get

$$= \frac{P(B_i) \cdot P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\, P(A/B_i)}$$

Ques There are two identical boxes containing respectively 4 white and 3 ~~ball~~ red balls, 3 white and 7 red balls. A box is chosen at random and a ball is drawn from it. Find the probability that the ball is white. If the ball is white what is the probability that it is from the first box?

Let A be the event of drawing a white ball and $B_1, B_2$ denote the events of chosing first box and second box respectively.

$P(B_1) = 1/2$ $\qquad$ $P(B_2) = 1/2$

$P(A) = ?$

white ball from box 1 $P(A/B_1) = \frac{4}{7}$ $\qquad$ $P(A/B_2) = 3/10$

(i) $P(A) = \sum_{i=1}^{n} P(B_i)\, P(A/B_i)$

$= P(B_1)\, P(A/B_1) + P(B_2) \cdot P(A/B_2)$

$= \frac{1}{2} \times \frac{4}{7} + \frac{1}{2} \times \frac{3}{10}$

$= \frac{2}{7} + \frac{3}{20}$ $\qquad$ 7 | 20, 7; 20 | 20, 7

$= \frac{40 + 21}{140}$

$= \frac{61}{140}$

P(ball from box 1 (A) ball is white) $P(B_i/A) = \frac{P(B_i)\, P(A/B_i)}{\sum_{i=1}^{n} P(B_i)\, P(A/B_i)}$

$= \frac{P(B_1)\, P(A/B_1)}{P(B_1)\, P(A/B_1) + P(B_2) \cdot P(A/B_2)}$

$= \frac{\frac{1}{2} \times \frac{4}{7}}{\frac{1}{2} \times \frac{4}{7} + \frac{1}{2} \times \frac{3}{10}}$

$= \frac{\frac{2}{7}}{\frac{61}{140}}$

$= \frac{2}{7} \times \frac{140}{61}$

$= \frac{40}{61}$

Ques 3 Machines A, B and C produce respectively 60%, 30% and 10% of the total number of items of a factory. Percentage of defective pieces produced by the machines are 2%, 3% and 4%. An item is selected at random is found to be defective. Find the probability that the item was produced by the machine C

Let A be the event of selecting a defective piece

$B_1, B_2$ and $B_3$ production of item

$P(B_1) = \frac{60}{100}$, $P(B_2) = \frac{30}{100}$

$P(B_3) = \frac{10}{100}$

$P(A|B_1) = 2/100$

$P(A|B_2) = 3/100$

$P(A|B_3) = 4/100$

$P(B_3/A) = ?$

$$P(B_3/A) = \frac{P(B_3)\,P(A|B_3)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)}$$

$$= \frac{\frac{10}{100} \times \frac{4}{100}}{\frac{60}{100} \times \frac{2}{100} + \frac{30}{100} \times \frac{3}{100} + \frac{10}{100} \times \frac{4}{100}}$$

$$= \frac{\frac{40}{100 \times 100}}{\frac{120}{10000} + \frac{90}{10000} + \frac{40}{10000}}$$

$$= \frac{\frac{40}{10000}}{\frac{250}{10000}}$$

$$= \frac{40}{10000} \times \frac{10000}{250}$$

$$= \frac{40}{250}$$

02.09.2019 T

Monday Ques There are 2 urns, one containing 5 white and 4 black balls and the other containing 6 white and 5 black balls. One urn is chosen and one ball is drawn. If it is white, what is the probability that the urn selected is the first.

| 5W 4B | 6W 5B |
|---|---|
| urn A | urn B |

~~P(A) = draw~~

Let A be drawing a white ball and B1 and B2 be the urn A and urn B respectively

$P(B_1) = 1/2$

$P(B_2) = 1/2$

$P(A/B_1) = 5/9$

$P(A/B_2) = 6/11$

$$P(B_1/A) = \frac{P(B_1)\ P(A/B_1)}{P(B_1)\ P(A/B_1) + P(B_2)\ P(A/B_2)}$$

$$= \frac{1/2 \times 5/9}{\frac{1}{2} \times \frac{5}{9} + \frac{1}{2} \times \frac{\cancel{6}\,3}{11}}$$

$$= \frac{5/18}{\frac{5}{18} + \frac{3}{11}}$$

$$= \frac{\frac{5}{18}}{\frac{55+54}{198}} = \frac{5}{18} \times \frac{198}{109} = \frac{55}{109}$$

The probability that a doctor will diagonise a particular disease correctly is 0.6. The probabelity that a patient will die by his treatment after correct diagnosis is 0.4 and the probability of death by wrong diagnosis is 0.7. A patient of the doctor who had the disease died. What is the probability that this disease was not correctly diagonised

Let A be the event of patient's death

$B_1$ stands for correctly diagonised

$B_2$ stands for not correctly diagonised

$P(B_1) = 0.6$

$P(B_2) = 1 - 0.6 = 0.4$

$P(A/B_1) = 0.4$

$P(A/B_2) = 0.7$

$P(B_2/A) = ?$

$$P(B_2/A) = \frac{P(B_2)\ P(A/B_2)}{P(B_1)\ P(A/B_1) + P(B_2)\ P(A/B_2)}$$

$$= \frac{0.4 \times 0.7}{0.6 \times 0.4 + 0.4 \times 0.7}$$

$$= \frac{0.28}{0.24 + 0.28}$$

$$= \frac{0.28}{0.52}$$

$$= 0.538 = 0.54$$

Ques Box A contains 9 cards numbered 1 to 9. Box B contains 5 cards numbered 1 to 5. A box is chosen at random and a card is drawn. If the number is even find the probability that the card is drawn from A

Let A be the event of taking even numbered card

$B_1$ be the ~~card~~ event of drawing a card from Box A and $B_2$ be the event of drawing a card from Box B

$P(B_1) = 1/2$

$P(B_2) = 1/2$

$P(A/B_1) = 4/9$

$P(A/B_2) = 2/5$

$P(B_1/A) = ?$

$$P(B_1/A) = \frac{P(B_1) \times P(A/B_1)}{P(B_1) \times P(A/B_1) + P(B_2) \times P(A/B_2)}$$

$$= \frac{1/2 \times 4/9}{1/2 \times 4/9 + 1/2 \times 2/5}$$

$$= \frac{\frac{2}{9}}{\frac{2}{9} + \frac{1}{5}}$$

$$= \frac{\frac{2}{9}}{\frac{10+9}{45}} = \frac{2}{9} \times \frac{45}{19}$$

$$P(B_1/A) = \frac{10}{19}$$

Wednesday 4/09/2019

Ques There are 3 bags each containing 4 white balls and 3 black balls and there are 4 bags each containing 5 black and 2 white balls. A bag was chosen and a ball was drawn. It was found to be black. What is the probability that the bag chosen was from the first set of bags?

Let A be the event of drawing a black ball

$B_1$ be the event of chosing first set of 3 bags

$B_2$ be the event of chosing second set of 4 bags

$P(B_1) = \cancel{1/3}\ 3/7$

$P(B_2) = 5/7$

$P(A/B_1) = 3/7$

$P(A/B_2) = 5/7$

$$P(B_1/A) = \frac{P(B_1) \times P(A/B_1)}{P(B_1) \times P(A/B_1) + P(B_2) \times P(A/B_2)}$$

$$= \frac{\frac{3}{7} \times \frac{3}{7}}{3/7 \times \frac{3}{7} + 5/7 \times \frac{5}{7}}$$

$$= \frac{\frac{9}{49}}{\frac{9+25}{49}}$$

$$= \frac{9}{49} \times \frac{49}{34}$$

$\frac{25}{34}$

$$P(B_1/A) = \frac{9}{34}$$

Ques In a bolt factory machine A, ~~too~~ B and C manufacture respectively 25%, 35% and 40% of their total. Of the outcomes 5%, ~~4~~ 4% and 2% are defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the prob. that it was manufactured by machine A?

Let A be the event of defective bolts
B1 be the manufacture of machine A
B2 be the manufacture of machine B
B3 be the manufacture of machine C

$P(B_1) = \frac{25}{100}$ $\quad P(B_2) = \frac{35}{100}$

$P(B_3) = \frac{40}{100}$

$P(A/B_1) = \frac{5}{100}$ $\quad P(A/B_2) = \frac{4}{100}$

$P(A/B_3) = \frac{2}{100}$

$$P(B_1/A) = \frac{P(B_1) \times P(A/B_1)}{P(B_1) \times P(A/B_1) + P(B_2)\, P(A/B_2) + P(B_3)\, P(A/B_3)}$$

$$= \frac{\frac{25}{100} \times \frac{5}{100}}{\frac{25}{100} \times \frac{5}{100} + \frac{35}{100} \times \frac{4}{100} + \frac{40}{100} \times \frac{2}{100}}$$

$$= \frac{\frac{125}{10000}}{\frac{125+140+80}{10000}}$$

$$= \frac{125}{10000} \times \frac{10000}{345} = \frac{125}{345}$$

$$P(B_1/A) = 0.3623$$

16/09/2019
Monday

# MODULE 4

## Random Variable and Probability Distribution

### * Random Variable

always asked

A real valued function defined over the sample space of a random experiment is called the random variable.

ie the value of a random variable correspond to the outcomes of a random experiment

eg:- In case of tossing 3 coins the outcomes can be described as getting 0 head, 1 head, 2 head, 3 head.
Let us consider a variable $x$ which takes values 0, 1, 2 and 3. The value of $x$ corresponds to the outcomes 0 head, 1 head, 2 head and 3 head.

Then $x$ can be considered to be a random variable associated ~~to~~ the random experiment of tossing 3 coins.

A randome variables are of two types
1. Discrete
2. Continous

A random variable to said to be <u>discrete</u> if it assumes only a finite number of values or countably infinite number of values.

eg:- A coin is tossed twice. Let $X$ denote the number of heads in each outcome. Then
$S = \{(HH)\ (HT)\ (TH)\ TT\}$

$X(HH) = 2$
$X(HT) = 1$
$X(TH) = 1$
$X(TT) = 0$

$X$ takes the values 0, 1, 2

A random variable is said to be <u>continous</u> if it can assume any value in a given interval. When $X$ takes any value in a given interval $(a, b)$ it is a continous variable in that interval.

eg:- Height and weight of students in a class, body temperature of patients. etc

17/09/2019
Tuesday

Probability distribution OR Probability function OR Density Function of a discrete random variable

Let $X$ be a random variable assuming values $x_1, x_2, x_3 \ldots$. Let $x$ stand for any one of $x_1, x_2, x_3 \ldots$. Then probability that the random variable $X$ takes the value $x$ is defined as probability function of $X$ and is denoted by

$f(x)$ or $P(x)$

$\therefore P(x) = P(X = x)$ where $X$ is the random variable and $x$ stands for values of $X$; is the probability function of $x$.

When $X$ takes the values $x_1, x_2 \ldots$ we have the corresponding probabilities $P(x_1), P(x_2) \ldots$ such that

$$P(x_1) + P(x_2) + \ldots = 1 \quad \text{for all } P(x) \geq 0$$

eg:- In tossing 2 coins the random variables representing the number of $x$ takes the values 0, 1, 2.

$x = 0, 1, 2$

$S = \{HH, HT, TH, TT\}$

$P(\text{no head}) = 1/4$

$P(1 \text{ head}) = 2/4$

$P(2 \text{ head}) = 1/4$

$\therefore P(x = 0) = 1/4$

$P(x = 1) = 2/4$

$P(x = 2) = 1/4$

So the probability distribution of $x$ is

| $x$ | $P(x)$ |
|---|---|
| 0 | 1/4 |
| 1 | 2/4 |
| 2 | 1/4 |

**Properties of Probability Distribution**

Let $P(x)$ be the probability density function then :-

1. $P(x) \geq 0$, for all values of $x$.
2. $\Sigma P(x) = 1$

Ques Examine whether the following is a probability distribution.

$f(x) = 0.2$ for $x = -1$

$f(x) = 0.3$ for $x = 0$

$f(x) = 0.1$ for $x = 1$

$f(x) = 0.2$ for $x = 2$

$f(x) = 0.2$ for $x = 3$

$f(x) = 0$ otherwise

$\Sigma f(x) = 0.2 + 0.3 + 0.1 + 0.2 + 0.2 + 0$
$= 1$

$\Sigma f(x) = 1$

and $f(x) \geq 0$, for all values of $x$

∴ It is a probability distribution function.

Ques Evaluate K if the following is a probability density function. Also obtain $P(1 \leq x \leq 3)$

| x | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| P(x) | 1/6 | 1/2 | K/10 | 1/30 |

Sol

We know that $\Sigma P(x) = 1$

$\frac{1}{6} + \frac{1}{2} + \frac{K}{10} + \frac{1}{30} = 1$

$1 = \frac{5 + 15 + 3K + 1}{30}$

$1 = \frac{21 + 3K}{30}$

$30 = 21 + 3K$

$3K = 9$

$K = \frac{9}{3}$ $\quad K = 3$

$P(1 \leq x \leq 3)$

The probability distribution is

| x | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| P(x) | 1/6 | 1/2 | 3/10 | 1/30 |

$P(1 \leq x \leq 3)$

$= P(x=1) + P(x=2) + P(x=3)$

$= \frac{1}{2} + \frac{3}{10} + \frac{1}{30}$

$= \frac{15 + 9 + 1}{30} = \frac{25}{30}$

$= \frac{5}{6}$

Ques A random variable $x$ has the following PDF (probability density function):-

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| P(x) | K | 3K | 5K | 7K | 8K | 11K | 0 | K | 3K |

(1) Find K

(2) Evaluate $P(x < 4)$, $P(x \geq 7)$, $P(2 < x < 5)$

(i) $K + 3K + 5K + 7K + 8K + 11K + K + 3K = 1$

$1 = 39K$

$K = 1/39$

(ii)

| $X$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $P(x)$ | 1/39 | 3/39 | 5/39 | 7/39 | 8/39 | 11/39 | 0 | 1/39 | 3/39 |

$P(x < 4)$

$= P(x=1) + P(x=2) + P(x=3)$

$= \frac{1}{39} + \frac{3}{39} + \frac{5}{39} + \frac{7}{39}$

$= \frac{16}{39}$

$P(x \geq 7)$

$= P(x=7) + P(x=8)$

$= \frac{1}{39} + \frac{3}{39}$

$= \frac{4}{39}$

$P(2 < x < 5)$

$= P(x=3) + P(x=4)$

$= \frac{7}{39} + \frac{8}{39}$

$= \frac{15}{39}$

Ques Evaluate $P(A)$, $P(B)$, $P(A \cap B)$ for the distribution given by

| $x$ | -5 | -1 | 0 | 1 |
|---|---|---|---|---|
| $f(x)$ | 1/4 | 1/4 | 1/4 | 1/4 |

where $A = \{-\infty < x < -2\}$

$B = \{-3 < x < 1/2\}$

$P(A) = P\{-\infty < x < -2\}$

$= P\{x = -5\}$

$= 1/4$

$P(B) = P\{-3 < x < 1/2\}$

$= P\{x = -1 \text{ and } x = 0\}$

$= 1/4 + 1/4$

$= 2/4$

$= 1/2$

$P(A \cap B) = P\{-3 < x < -2\}$

$= \underline{\underline{0}}$

$\therefore P(A) = 1/4$

$P(B) = 1/2$

$P(A \cap B) = 0$

Ques A random variable $x$ has the following probability function. find the value of $K$ and then evaluate

$P(x < 6)$, $P(x \geq 6)$ $P(0 < x < 5)$

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $f(x)$ | 0 | $2k$ | $3k$ | $k$ | $2k$ | $k^2$ | $7k^2$ | $2k^2+k$ |

$0 + 2k + 3k + k + 2k + k^2 + 7k^2 + 2k^2 + k = 1$

$9k + 10k^2 = 1$

$9k + 10k^2 - 1 = 0$

$10k^2 + 9k - 1 = 0$

$S = 9$

$P = -10$

$-1 \times 10$

$-1 \times 10$

$S = 9$ $P = -10$

$\frac{2}{100} + \frac{1}{10}$

$= \frac{2+10}{100}$

$10k^2 + 10k - 1k - 1 = 0$

$10k(k+1) - 1(k+1) = 0$

$(10k - 1)(k+1) = 0$

$10k = 1$ $\qquad k = -1$

$k = 1/10$ $\qquad$ it is neglected

| $x$ | 0 | 1 | 2 | 4 | 5 |
|---|---|---|---|---|---|
| $f(x)$ | 0 | 2/10 | 3/10 | 2/10 | 1/10 |

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $f(x)$ | 0 | 2/10 | 3/10 | 1/10 | 2/10 | 1/100 | 7/100 |

| | 7 | | 7 |
|---|---|---|---|
| | 2/100 + 1/10 | or | 12/100 |

$P(x < 6)$

$= P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4) + P(x=5)$

$= 0 + \frac{2}{10} + \frac{3}{10} + \frac{1}{10} + \frac{2}{10} + \frac{1}{100}$

$= \frac{20+30+10+20+1}{100} = \frac{81}{100}$

$P(x \geq 6)$

$= P(x=0) + P(x=1) + P(x=2) + P(x=3) + P(x=4)$
$+ P(x=5) + P(x=6)$

$= 0 + \frac{2}{10} + \frac{3}{10} + \frac{1}{10} + \frac{2}{10} + \frac{1}{100} + \frac{7}{100}$

$= \frac{8}{10} + \frac{8}{100}$

$= \frac{80+8}{100} = \frac{88}{100}$

$P(0 < x < 5)$

$P(x=1) + P(x=2) + P(x=3) + P(x=4)$

$= \frac{2}{10} + \frac{3}{10} + \frac{1}{10} + \frac{2}{10}$

$= \frac{5+3}{10}$

$= \frac{8}{10}$

## Expectation of X

Let the random variable $x$ assume the values $x_1, x_2 \ldots$ with corresponding probabilities $P(x_1, x_2, \ldots)$

Then the expected value of random variable $x$ denoted by $E(x)$ is given by the formula

$$E(x) = x_1 P(x_1) + x_2 P(x_2) + \ldots$$

ie $E(x) = \Sigma\, x P(x)$

eg:- When a die is thrown the random variable $x$ takes the values 1, 2, 3, 4, 5, 6 with corresponding probabilities 1/6, 1/6, 1/6, 1/6, 1/6, 1/6. then :

$$E(x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$$

$$= \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}$$

$$= \frac{21}{6}$$

Ques A random variable X takes values 0, 1, 2, 3, 4 with corresponding probabilities 1/4, 1/5, 2/5, 1/8, 1/40. Find the expectation of $x$

| $(x)$ | $P(x)$ | $x(P(x))$ |
|---|---|---|
| 0 | 1/4 | 0 |
| 1 | 1/5 | 1/5 |
| 2 | 2/5 | 4/5 |
| 3 | 1/8 | 3/8 |
| 4 | 1/40 | 4/40 or 1/10 |

$E(x) = \sum x P(x)$

$= 0 + \frac{1}{5} + \frac{4}{5} + \frac{3}{8} + \frac{1}{10}$

$= \frac{8+32+15+4}{40}$

$E(x) = \frac{59}{40}$

## Variance of X

Variance of a random variable $x$ whose expectation denoted by $E(x)$ is defined as

$V(x) = E[x - E(x)]^2$

$= E(x^2) - [E(x)]^2$

where $E(x^2) = \sum x^2 P(x)$

Ques A random variable $x$ takes values 1 and 2 with corresponding probabilities 1/3 and 2/3 find $E(x)$ and $V(x)$.

| $(x)$ | $P(x)$ | $x^2$ | $x P(x)$ | $x^2 P(x)$ |
|---|---|---|---|---|
| 1 | 1/3 | 1 | 1/3 | 1/3 |
| 2 | 2/3 | 4 | 4/3 | 8/3 |

$E(x) = \sum x P(x)$

$= \frac{1}{3} + \frac{4}{3}$

$= \frac{5}{3}$

$E(x^2) = \sum x^2 P(x)$

$= \frac{1}{3} + \frac{8}{3}$

$= \frac{9}{3}$

$= 3$

$V(x) = E(x^2) - (Ex)^2$

$= \frac{9}{3} - \left(\frac{5}{3}\right)^2$

$= \frac{9}{3} - \frac{25}{9}$

$= \frac{27-25}{9}$

$V(x) = \frac{2}{9}$

Ques A random variable x follows a prob. distribution as given below. find the value of k and also find mean (E(x)) and variance of the variable

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f(x)$ | $k/2$ | $k/3$ | $\frac{k+1}{3}$ | $\frac{2k-1}{6}$ |

$1 = \frac{k}{2} + \frac{k}{3} + \frac{k+1}{3} + \frac{2k-1}{6}$

$= \frac{3k + 2k + 2(k+1) + 2k - 1}{6}$

$1 = \frac{5k + (2k+2) + 2k - 1}{6}$

$1 = \frac{9k+1}{6}$

$9k + 1 = 6$

$k = \frac{5}{9}$

Mean $= E(x) = \Sigma x P(x)$

| | | | | |
|---|---|---|---|---|
| $x^2$ | 0 | 1 | 4 | 9 |
| $x$ | 0 | 1 | 2 | 3 |
| $f(x)$ | 5/18 | 5/27 | 14/27 | 1/54 |
| $xf(x)$ | 0 | 5/27 | 28/27 | 3/54 |
| $x^2 f(x)$ | 0 | 5/27 | 56/27 | 9/54 |

$E(x) = \Sigma x P(x)$

$= \frac{5}{18} + \frac{5}{27} + \frac{14}{27} + \frac{1}{54}$

$= \frac{15 + 10 + 28 + 1}{54}$

$= \frac{15 + 29}{54}$

$= \frac{44}{54}$

$\frac{\frac{10}{9} - \frac{1}{9}}{6}$

$k/2$

$\frac{5/9}{2}$

$\frac{1}{2} \times \frac{5}{9}$

$\frac{k+1}{3}$

$\frac{k}{3} + \frac{1}{3}$

$\frac{5}{9} \times \frac{1}{2}$

$\frac{5}{9} \times \frac{1}{3}$

$\frac{\frac{5}{9} + 1}{3}$

$\frac{\frac{5+9}{9}}{3}$

$= \frac{14/9}{3}$

$= \frac{14}{9} \times \frac{1}{3}$

$$E(x) = 0 + \frac{5}{27} + \frac{28}{27} + \frac{3}{54}$$

$$= \frac{10 + 56 + 3}{54}$$

$$= \frac{66 + 3}{54}$$

$$E(x) = \frac{69}{54}$$

$$E(x^2) = \sum x^2 f(x)$$

$$= 0 + \frac{5}{27} + \frac{56}{27} + \frac{9}{54}$$

$$= \frac{10 + 112 + 9}{54}$$

$$E(x^2) = \frac{131}{54}$$

$$\text{Variance} = E(x^2) - \left(E(x)\right)^2$$

$$= \frac{131}{54} - \left(\frac{69}{54}\right)^2$$

$$= \frac{131}{54} - \frac{4761}{2916}$$

$$= \frac{7074 - 4761}{2916}$$

$$= \frac{2313}{2916}$$

$$= 0.793$$

$$\text{Variance} = 0.8$$

18/09/2019
Wednesday

## Distribution Function

Let $X$ be a random variable and $x$ be any value of it, then

$P(X \le x)$ denoted by $F(x)$ is called distribution function.

## Properties of Distribution Function

1. $F(x)$ must be greater than or equal to zero

   ie $F(x) \ge 0$

2. $F(\infty) = 1$

3. $F(x)$ is a non-decreasing function.

   ie $F(x) \ge F(y)$ when $(x \ge y)$

Ques A random variable $x$ has the following probability function

$f(0) = 1/6 \quad f(1) = 2/6 \quad f(2) = 3/6$

$f(x) = 0$ otherwise

(i) write down the density function or probability distribution and distribution function.

(ii) find $P(x < 2)$

(iii) find $P(0 < x < 2)$

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| ~~$f(x)$~~ | ~~1/6~~ | ~~2/6~~ | ~~3/6~~ |
| ~~$F(x)$~~ | ~~1/6~~ | ~~2/6 + 1/6 = 3/6~~ | |

| $x$ | $f(x)$ | $F(x)$ |
|---|---|---|
| 0 | 1/6 | 1/6 |
| 1 | 2/6 | 1/6 + 2/6 = 3/6 |
| 2 | 3/6 | 3/6 + 3/6 = 6/6 |

$f(x) = \frac{x+1}{6}$ for $x = 0, 1, 2$

$F(x) = \sum_{0}^{x} \frac{x+1}{6}$ for $x = 0, 1, 2$

(i) $P(x < 2)$

$P(x = 0) + P(x = 1)$

$= \frac{1}{6} + \frac{2}{6} = \frac{3}{6}$

(ii) $P[0 < x < 2]$

$P(x = 1)$

$= \frac{2}{6}$

old H/w

Ques Find $k$ if $f(1) = k/2 \quad f(2) = k/3$
$f(3) = k/4, \quad f(4) = \frac{k+1}{2}$. Also find mean and variance

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $f(x)$ | $k/2$ | $k/3$ | $k/4$ | $\frac{k+1}{2}$ |

$\Rightarrow \frac{k}{2} + \frac{k}{3} + \frac{k}{4} + \frac{k+1}{2} = 1$

$\Rightarrow \frac{6k + 4k + 3k + 6(k+1)}{12} = 1$

$1 = \frac{6k + 4k + 3k + 6k + 6}{12}$

$12 = 19k + 6$

$19k = 6$

$k = \frac{6}{19}$

| $x$ | $f(x)$ | $f(x)$ | $x^2$ | $x^2 f(x)$ | $x f(x)$ |
|---|---|---|---|---|---|
| 1 | $\cancel{k/2}$ | $3/19$ | 1 | $3/19$ | $3/19$ |
| 2 | $\cancel{k/3}$ | $2/19$ | 4 | $8/19$ | $4/19$ |
| 3 | $\cancel{k/4}$ | $3/38$ | 9 | $27/38$ | $9/38$ |
| 4 | $\frac{k+1}{2}$ | $25/38$ | 16 | $400/38$ | $100/38$ |

$E(x) = \Sigma x f(x)$

$= \frac{3}{19} + \frac{4}{19} + \frac{9}{38} + \frac{100}{38}$

$= \frac{6+8+9+100}{38}$

$= \frac{23+100}{38} = \frac{123}{38} = 3.23$

$E(x^2) = \Sigma x^2 f(x)$

$= \frac{3}{19} + \frac{8}{19} + \frac{27}{38} + \frac{400}{38}$

$= \frac{6+16+27+400}{38}$

$E(x^2) = \frac{449}{38} = 11.81$

Variance $\{V(x) = E(x^2) - (E(x))^2$

$= 11.81 - (3.23)^2$

$= 11.81 - 10.43$

$V(x) = \underline{1.38}$

Ques Find the expected value of the number of heads when two coins are tossed.

Ans

| $x$ | $P(x)$ | $\Sigma x P(x)$ |
|---|---|---|
| 0 | $1/4$ | 0 |
| 1 | $2/4$ | $2/4$ |
| 2 | $1/4$ | $2/4$ |

HH
HT
TH
TT $1/4$

The random variable X assumes the value 0, 1, and 2 (no head, one head and two head)

$E(x) = \Sigma x P(x)$

$= \frac{2}{4} + \frac{2}{4}$

$E(x) = \underline{1}$

Ques A box contains 6 tickets, 2 of the tickets carry a price of Rs 5 each and the other 4 carry a price of Rs 1 each.
(a) If one ticket is drawn, what is the expected value of the price?

(b) If 2 tickets are drawn, what is the expected value of the price?

$$\boxed{\begin{matrix} 5 & 5 & 1 \\ 1 & 1 & 1 \end{matrix}}$$

(a) $X = 1, 5$ (one ticket is drawn)

$P(X=1) = \frac{4C_1}{6C_1} = \frac{4}{6}$

$P(X=5) = \frac{2C_1}{6C_1} = \frac{2}{6}$

| $x$ | $P(x)$ | $x\,P(x)$ |
|---|---|---|
| 1 | 4/6 | 4/6 |
| 5 | 2/6 | 10/6 |

$E(x) = \sum x\,P(x)$

$= \frac{4}{6} + \frac{10}{6}$

$E(x) = \frac{14}{6} = \frac{7}{3}$

(b) two tickets are drawn

$X = 2, 6, 10$

1 + 5 = 6

5 + 5 = 10

1 + 1 = 2

$P(X=2) = \frac{4C_2}{6C_2} = \cancel{\frac{4\times3}{6\times5}}$

$= \frac{4\times3}{1\times2} \div \frac{6\times5}{1\times2}$

$= \frac{4\times3}{6\times5}$

$P(X=2) = \frac{4}{10}$

$P(X=6) = \frac{4C_1 \times 2C_1}{6C_2} = \cancel{\frac{4\times3}{1\times2} \times \frac{2\times1}{2}}$

$= \cancel{\frac{4\times2}{6\times5} \times 1\times2}$

$= \cancel{\frac{12\times2}{5}}$

$= \frac{\frac{4\times2}{6\times5}}{1\times2} = \frac{4\times4}{6\times5} = \frac{8}{15}$

$P(X=10) = \frac{2C_2}{6C_2} = \frac{2\times1}{1\times2} \times \frac{1\times2}{6\times5}$

$= \frac{1}{15}$

| $x$ | $P(x)$ | $x\,P(x)$ |
|---|---|---|
| 2 | 6/15 | 12/15 |
| 6 | 8/15 | 48/15 |
| 10 | 1/15 | 10/15 |

$E(x) = \Sigma x\,P(x)$

$= \frac{12}{15} + \frac{48}{15} + \frac{10}{15}$

$E(x) = \frac{70}{15}$

Ques A player is to ~~toss~~ 3 coins. He wins Rs 10 if 3 heads appear, Rs 5 if 2 head appear and Rs 1 if one head appear. He will lose Rs 12 if no head appears. What is the expected amount

$S = \{HHH, HTH, HHT, THH, TTH, TTT, THT, HTT\}$

3 heads - Rs 10
2 heads - Rs 5
1 head - Rs 1
no head - ~~R~~ -12

$X = \cancel{10}1, 5, 10, -12$

$P(X=1) = 3/8$

$P(X=5) = 3/8$

$P(X=10) = 1/8$

$P(X=-12) = 1/8$

| $x$ | $P(x)$ | $x\,P(x)$ |
|---|---|---|
| 1 | 3/8 | 3/8 |
| 5 | 3/8 | 15/8 |
| 10 | 1/8 | 10/8 |
| -12 | 1/8 | -12/8 |

$E(x) = \Sigma x\,P(x)$

$= \frac{3}{8} + \frac{15}{8} + \frac{10}{8} - \frac{12}{8}$

$= \cancel{\frac{18}{8}} \quad \frac{28}{8} - \frac{12}{8}$

$E(x) = \frac{16}{8} = 2$

Ques 3 urns contains respectively 3 green and 2 white balls; 5 green and 6 white balls and 2 green and 4 white balls. One ball is drawn from each urn. Find the expected number of white balls drawn out.

3G 2W | 5G 6W | 2G 4W

The number of white balls drawn can be 0, 1, 2 or 3

Random variable, $X = 0, 1, 2, 3$

$P(X=0)$ (drawing no white ball or drawing green ball)

$P(X=0) = \frac{3}{5} \times \frac{5}{11} \times \frac{2}{6}$

$= \frac{1}{11}$

$P(X=1)$ WGG / GWG / GGW

$= \cancel{\frac{2}{5} \times \frac{5}{11} \times \frac{2}{6} + \frac{3}{5} \times \frac{6}{11} \times \frac{2}{6} + \frac{3}{5} \times \frac{5}{11} \times \frac{4}{6}}$

$= \cancel{\frac{2}{33} + \frac{6}{55} + \frac{2}{11}}$

$= \cancel{\frac{\quad}{330}}$

$P(X=1) = \frac{2}{5} \times \frac{5}{11} \times \frac{2}{6} + \frac{3}{5} \times \frac{6}{11} \times \frac{2}{6} + \frac{3}{5} \times \frac{5}{11} \times \frac{4}{6}$

$= \frac{20 + 36 + 60}{330}$

$= \frac{116}{330}$

$P(X=2) =$ (WWG / WGW / GWW)

$= \frac{2}{5} \times \frac{6}{11} \times \frac{2}{6} + \frac{2}{5} \times \frac{5}{11} \times \frac{4}{6} + \frac{3}{5} \times \frac{6}{11} \times \frac{4}{6}$

$= \frac{24 + 40 + 72}{330}$

$= \frac{136}{330}$

$P(X=3) = \frac{2}{5} \times \frac{\cancel{6}}{11} \times \frac{4}{\cancel{6}}$

$= \frac{8}{5 \times 11} = \frac{8}{55}$ or $\frac{48}{330}$

| $x$ | $P(x)$ | $xP(x)$ |
|---|---|---|
| 0 | 7/11 | 0 |
| 1 | 116/330 | 116/330 |
| 2 | 136/330 | 272/330 |
| 3 | 48/330 | 144/330 |

Soln. $E(x) = \sum x P(x)$

$= \frac{116}{330} + \frac{272}{330} + \frac{144}{330}$

$= \frac{532}{330} = \frac{266}{165}$

25/09/2019
Wednesday

Ques 3 coins whose 2 faces are marked 1 and 2 are thrown. find the expectation of the no. obtained

Sample space = {111, 112, 121, 122, 211, 221, 212, 222}

$x = 3, 4, 5, 6$

$P(X=3) = 1/8$

$P(X=4) = 3/8$

$P(X=5) = 3/8$

$P(X=6) = 1/8$

| $x$ | $P(x)$ | $xP(x)$ |
|---|---|---|
| 3 | 1/8 | 3/8 |
| 4 | 3/8 | 12/8 |
| 5 | 3/8 | 15/8 |
| 6 | 1/8 | 6/8 |

$E(x) = \sum x P(x)$

$= \frac{3}{8} + \frac{12}{8} + \frac{15}{8} + \frac{6}{8}$

$= \frac{15}{8} + \frac{15}{8} + \frac{6}{8}$

$= \frac{\cancel{36}\ \cancel{18}\ 9}{\cancel{8}\ \cancel{4}\ 2} = \frac{9}{2}$

$E(x) = \frac{9}{2}$

Ques You have been offered a dice game in which you will recieve Rs 20 each time if the point total of a toss of 2 dice is 6. If it costs you Rs 2·50 per toss to participate, should you play or not?

(b) Will it make any difference in your decision if it costs Rs 3 per toss instead of Rs 2·50?

(a) if 6 $\longrightarrow$ Rs 20
but Rs 2·50
cost so
₹ 17·50

$= \frac{5}{36} \Rightarrow [(1,5), (2,4), (3,3), (4,2), (5,1)]$

if not 6 $\rightarrow$ $-2·50$ loss $= \frac{31}{36}$

(a)

| $x$ | $P(x)$ | $x\,P(x)$ |
|---|---|---|
| 17·50 | 5/36 | 875/360 |
| −2·50 | 31/36 | −775/360 |

$E(x) = \sum x\,P(x)$

$= \frac{875}{360} + \frac{-775}{360}$

$E(x) = \frac{100}{360} = \frac{5}{18}$

As $\frac{10}{36}$ is positive there is gain so it is advisable to participate

(b) 6 $\rightarrow$ Rs 20
but Rs 3 charge
∴ ₹ 17

$= \frac{5}{36}$

not 6 $\rightarrow$ − Rs 3 $= \frac{31}{36}$

| $x$ | $P(x)$ | $x\,P(x)$ |
|---|---|---|
| 17 | 5/36 | 85/36 |
| −3 | 31/36 | −93/36 |

$E(x) = \sum x P(x)$

$= \frac{85}{36} + \frac{-93}{36}$

$= \frac{-8}{36}$

As $-\frac{8}{36}$ is negative there is loss so it advisable not to participate.

## Properties of Expectation

1. $E(a) = a$, where a is a constant.
2. $E(ax) = aE(x)$, where a is a constant.
3. $E(x+y) = E(x) + E(y)$
4. $E(xy) = E(x) \cdot E(y)$ if x and y are independent

Ques If $U = ax + b$ find expectation of U where a and b are constant

$E(U) = ?$

$U = ax + b$

$E(U) = E(ax+b)$

$= E(ax) + E(b)$

$= aE(x) + E(b)$

$= a\underline{E(x) + b}$

Ques If $E(x) = 2.5$ ~~and~~ find $E(3x+7)$?

$E(3x+7) = E(3x) + E(7)$

$= 3E(x) + 7$

$= 3 \times 2.5 + 7$

$= 7.5 + 7$

$E(3x+7) = \underline{14.5}$

## Properties of Variance

1. $V(K) = 0$ when K is a constant.
2. $V(Kx) = K^2 V(x)$ where K is a constant.
3. $V(Kx + b) = K^2 V(x)$ where K is a constant.
4. $V(x+y) = V(x-y) = V(x) + V(y)$ if x and y are independent

Ques X is a random variable whose mean is $\mu$ and standard deviation is $\sigma$. What will be the mean and standard deviation of (a) $2x$ (b) $2x+1$

$\cancel{SD} = x$ $\quad V = (SD)^2$

$E(x) = \mu$

$V(x) = (\sigma)^2$

(a) $E(2x) = 2E(x)$
$= 2 \times \mu$

$E(2x) = 2\mu$

(b) ~~E(2x)~~
$V(2x) = 2^2 V(x)$
$= 4\sigma^2$
$= 4\sigma^2$

(b) $E(2x+1) = E(2x) + E(1)$
$= 2E(x) + 1$
$= 2 \times \mu + 1$
$= 2\mu + 1$

$V(2x+1) = \cancel{2x}\ \cancel{V}\ 2x$.
$= V(2x) + V(1)$
$= 2^2 V(x) + 0$
$= 4V(x)$

$= 4\sigma^2$

Ques If $x$ and $y$ are two independent variates and $V(x) = 2$ and $V(y) = 3$. find $V(2x + 3y)$

$V(2x+3y) = V(2x) + V(3y)$
$= 2^2 V(x) + 3^2 V(y)$
$= 4 \times 2 + 9 \times 3$
$= 8 + 27$
$V(2x+3y) = 35$

Ques Find mean and variance of $ax + b$ if $\bar{x} = 4$ and $V(x) = 8$

$E(x) = 4 \quad V(x) = 8$

$E(ax+b) = E(ax) + E(b)$
$= aE(x) + b$
$= a4 + b$

$E(ax+b) = 4a + b$

$V(ax+b) = V(ax) + V(b)$
$= a^2 V(x) + 0$
$= a^2 \times 8$
$= 8a^2$

$V(ax+b) = 8a^2$

Ques A man draws 2 balls from a bag containing 4 red and 6 green balls. If he is to recieve Rs 45 for every red ball which he draws and Rs 22.5 for every green ball he draws. What is his expectaton?

$S = \{RR, RG, GR, GG\}$ $\left.\begin{matrix} 4R \\ 6G \end{matrix}\right\}$ 10 Balls

| $x$ | $P(x)$ | $x \cdot P(x)$ |
|---|---|---|
| 45 | 4/10 | 180/10 |
| 22.5 | 6/10 | 135/10 |

$E(x) = \sum x \, P(x)$

$= \frac{180}{10} + \frac{135}{10}$

$= \frac{315}{10}$

$= 31.5$

$E(x) = 2 \times 31.5$

$= 63$

---

7/2019 Monday

## Moments of Random Variable

Let $x$ be a random variable then $E(x - \bar{x})^r$ is the $r$th moment of $x$ with mean $\bar{x}$ and is denoted by $\mu_r$. This is also called central moment of order $r$

Putting $r = 1, 2 \ldots$

$\mu_1 = E(x - \bar{x})^1$

$\mu_2 = E(x - \bar{x})^2$

and so on.

NOTE :- $E(x - a)^r$ where $a$ is any artibrary number is called $r^{th}$ raw moment about $a$ and is denoted by $\mu'_r$

when $a = 0$, we get $\mu'_r$ about the origin ie

$\mu'_r$ about the origin $= E(x^r)$

$E(x - \bar{x})^r$ is the $r^{th}$ central moment and is denoted by $\mu_r$

Recurrence relation between central moments and Row Moments

Let $\mu_r$ be the $r^{th}$ central moment and $\mu'_r$ be the $r^{th}$ Row moment

then $\mu_r = E(x - \bar{x})^r$

$$\mu_r = E(x - \mu'_1)^r$$

expanding first and then finding expectation of each term we get the following

$$\mu_r = \mu'_r - {}^rC_1 \mu'_{r-1} \mu'_1 + {}^rC_2 \mu'_{r-2} \mu'^2_1 \dots + (-1)^r (\mu'_1)^r$$

putting $r = 1, 2, 3, 4 \dots$ we get,

$$\mu_1 = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3.$$

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

Uses of Moments

Moments can be used to describe
(i) Mean
(ii) Standard deviation
(iii) Coefficient of skewness
(iv) Measure of kurtosis

(i) Mean = $\mu'_1$ (ie first moment about origin is the mean)

(ii) Standard deviation,

$$SD = \sqrt{\mu_2}$$

ie square root of second central moment is standard deviation

(iii) Coefficient of skewness $= \dfrac{\mu_3}{\sqrt{\mu_2^3}}$

(iv) Measure of kurtosis,

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

## Moment Generating Function (MGF)

Moment Generating Function of a random variable $X$ is defined as expectation of $e^{tx}$

$\therefore$ MGF denoted by $\boxed{M_x(t) = E(e^{tx})}$

when $M_x(t)$ is expanded we get a series of the form

$$M_x(t) = 1 + \frac{t}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \ldots$$

The coefficient of $t^r/r!$ in this series represents $\mu_r'$

$\therefore$ to find the moments obtain MGF and expand it

NOTE when $x$ is discrete

$$M_x(t) = E(e^{tx})$$
$$= \Sigma e^{tx} f(x)$$

where $f(x)$ is the probability function of $x$.

for continous random variable $M_x(t)$

$$M_x(t) = E(e^{tx})$$
$$= \int_{-\infty}^{\infty} f(x) \cdot e^{t} \cdot dx$$

If a moment generating function exist for a random variable $x$ then

(i) the mean of $x$ can be ~~formed~~ found by evaluating the first derative of MGF at $t=0$

~~The variance of $x$ can be found by~~

ie $\mu = E(x) = M'(0)$

(ii) the variance of $x$ can be found by evaluating the first and second derative of MGF at $t=0$

ie $\sigma^2 = E(X)^2 - (E(x))^2$

$$= M''(0) - (M'(0))^2$$

$E(x), E(x^2) \ldots E(x^r)$ are called moment about origin.

In general $r^{th}$ moment about the origin can be found by evaluating the

$r^{th}$ derivative of the MGF at $t = 0$

$M_x(t) = E(e^{tx}) = \sum e^{tx} f(x)$

$$\left.\frac{d}{dt} M_x(t)\right]_{t=0} = M'(0) = \sum x\, e^{tx} f(x) \Big|_{t=0}$$

$$= \sum x\, f(x) = E(x)$$

$$\left.\frac{d^2 M_x(t)}{dt}\right]_{t=0} = M''(0) = \sum x^2 e^{tx} f(x) \Big|_{t=0}$$

$$= \sum x^2 f(x) = E(x^2)$$

So $\mu = M'(0)$

$\sigma^2 = M''(0) - (M'(0))^2$

ie $E(x^2) - (E(x))^2$

Ques Find the moment generating function of $x$ whose $f(x) = \frac{e^{-m} m^x}{x!}$ for $x = 0, 1, 2 \ldots$

$M_x(t) = E(e)^{tx}$

$$= \sum_0^\infty e^{tx} f(x)$$

$$= \sum_0^\infty \frac{e^{tx} e^{-m} m^x}{x!}$$

$$= \sum_0^\infty \frac{(e^t m)^x \cdot e^{-m}}{x!}$$

$$= e^{-m} \sum_0^\infty \frac{(me^t)^x}{x!}$$

$$= e^{-m} \cdot E^{me^t}$$

$= \cancel{e^{-m} (1 \times e^t)}$ $e^{-m} (1 \times e^t)$

$$e^{-m}(1 - e^t) = 1 + \frac{-m(1-e^t)}{1!} + \frac{m^2(1-e^t)^2}{2!} + \ldots$$

$$1 + \frac{m}{1!}\left(\frac{t}{1!} + \frac{t^2}{2!} + \ldots\right] + \frac{m^2}{2!}\left[\frac{t}{1!} + \frac{t^2}{2!} + \ldots\right]$$

$+ \ldots$

equating coefficients of $\frac{t}{1!}, \frac{t^2}{2!} \ldots$

we get $\mu_1', \mu_2' \ldots$

$\mu_1' = $ coeff of $\frac{t}{1!} = m$

$\mu_2' = $ coeff of $\frac{t^2}{2!} = m + m^2 t \ldots$

Ques If the MGF of t $M(t) = \frac{2}{5}e^t + \frac{1}{5}e^{2t} + \frac{2}{5}e^{3t}$ find mean and variance

Mean $= E(x) = M'(t) =$

$= \frac{2}{5}e^t + 2 \times \frac{1}{5}e^{2t} + 3 \times \frac{2}{5}e^{3t}$

$= \frac{2}{5}e^t + \frac{2}{5}e^{2t} + \frac{6}{5}e^{3t}$

put $t=0$

$M'(0) = \frac{2}{5} + \frac{2}{5} + \frac{6}{5}$

$= \frac{10}{5} = \underline{\underline{2}}$

Variance $= M'(t) = \frac{2}{5}e^t + \frac{2}{5}e^{2t} + \frac{6}{5}e^{3t}$

$= \frac{2}{5}e^t + \frac{4}{5}$ ~~e~~

$M''(t) = \frac{2}{5}e^t + \frac{4}{5}e^{2t} + \frac{18}{5}e^{3t}$

$M''(0) = \frac{2}{5} + \frac{4}{5} + \frac{18}{5}$

$= \underline{\frac{24}{5}}$

Variance $= M''(0) - (M'(0))^2$

$= \frac{24}{5} - 2^2$

$= \frac{24}{5} - 4$

$= \underline{\frac{4}{5}}$

Ques If $x$ denotes the outcome when a fair die is tossed find MGF($x$) and hence find mean and variance of $x$

| | | | | | | |
|---|---|---|---|---|---|---|
| X | 1 | 2 | 3 | 4 | 5 | 6 |
| P(x) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

$M(t) = E(e^{tx})$

$= \sum P(x)\, e^{tx}$

$= \frac{1}{6}\left[e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t}\right]$

$$\text{Mean} = \mu_1' = E(x) = \frac{d}{dt} M_0(t)\Big]_{t=0}$$

$$= \frac{1}{6}\left[e^t + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}\right]_{t=0}$$

$$= \frac{1}{6}\left[1+2+3+4+5+6\right]$$

$$= \frac{21}{6} = 7/2$$

$$\mu_2' = \frac{1}{6}\Big[e^t + 2\times 2e^{2t} + 3\times 3e^{3t} + 4\times 4e^{4t} + 5\times 5e^{5t} + 6\times 5e^{6t}\Big]_{t=0}$$

$$= \frac{1}{6}\left[1+4+9+16+25+36\right]$$

$$= \frac{91}{6}$$

$$\text{Variance} = \mu_2' - (\mu_1')^2$$

$$= \frac{91}{6} - \left(\frac{7}{2}\right)^2$$

$$= \frac{91}{6} - \frac{49}{4}$$

$$= \frac{364-294}{24} = \frac{\cancel{70}\ 35}{\cancel{24}\ 12}$$

Variance $= \frac{35}{12}$

NOTE

Mean $= E(x)$

Variance $=$ var$(x)$ or $V(x)$

$V(x) = E(x^2) - (E(x))^2$

$E(x) = \sum p_i x_i \rightarrow$ discrete

$\int_{-\infty}^{\infty} x\, f(x)\, dx \rightarrow$ continous

$E(x) = \mu_1'$

$E(x^2) = \sum p_i x_i^2 \rightarrow$ discrete

$\int_{-\infty}^{\infty} x^2 f(x)\, dx \rightarrow$ continous

$E(x^2) = \mu_2'$

NOTE

$f(x) = \frac{d}{dx} F(x)$

Ques A continuous random variable $x$ has a probability distribution

$$f(x) = \begin{cases} \frac{4}{81} x(9-x^2) & \text{when } 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

find first 4 moments about the origin and mean

$$\mu_1' = E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x \frac{4}{81} x(9-x^2)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81} x^2(9-x^2)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81} (9x^2 - x^4)\, dx$$

$$x^n \quad \frac{x^{n+1}}{n+1}$$

$$= \frac{4}{81} \int_{-\infty}^{\infty} (9x^2 - x^4)\, dx$$

$$= \frac{4}{81} \left[ \frac{9x^3}{3} - \frac{x^5}{5} \right]_0^3$$

$$= \frac{4}{81} \left[ \left( \frac{9 \times 3^3}{3} - \frac{3^5}{5} \right) - (0) \right]$$

$$= \frac{4}{81} \left[ 81 - \frac{243}{5} \right]$$

$$= \frac{4}{81} \left( \frac{405 - 243}{5} \right)$$

$$= \frac{4}{81} \times \frac{162}{5}$$

18×9

$$\mu_1' = \frac{8}{5}$$

$$\begin{array}{r} 81 \\ \times 2 \\ \hline \end{array}$$

$$\mu_2' = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^2 \frac{4}{81} x(9-x^2)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81} x^3 (9-x^2)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81} (9x^3 - x^5)\, dx$$

$$= \frac{4}{81}\int_{-\infty}^{\infty}(9x^3 - x^5)\,dx$$

$$= \frac{4}{81}\left[\frac{9x^4}{4} - \frac{x^6}{6}\right]_0^3$$

$$= \frac{4}{81}\left[\left(\frac{9\times 3^4}{4} - \frac{3^6}{6}\right) - (0)\right]$$

$$= \frac{4}{81}\left(\frac{729}{4} - \frac{729}{6}\right)$$

$$= \frac{4}{81}\left(\frac{4374}{24} - \frac{2916}{24}\right)$$

$$= \frac{\cancel{4}}{\cancel{81}} \times \frac{\cancel{1458}^{18}}{\cancel{24}_{6}}$$

$$= \frac{18}{6}$$

$$\mu_2' = 3$$

$$\mu_3' = \int_{\infty}^{\infty} x^3 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^3 \frac{4}{81}\, x(9 - x^2)\,dx$$

$$= \int_{\infty}^{\infty} \frac{4}{81}\, x^4(9 - x^2)\,dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81}(9x^4 - x^6)\,dx$$

$$= \frac{4}{81}\int_{-\infty}^{\infty}(9x^4 - x^6)\,dx$$

$$= \frac{4}{81}\left(\frac{9x^5}{5} - \frac{x^7}{7}\right)_0^3$$

$$= \frac{4}{81}\left(\frac{9\times 3^5}{5} - \frac{3^7}{7}\right)$$

$$= \frac{4}{81}\left(\frac{2187}{5} - \frac{2187}{7}\right)$$

$$= \frac{4}{81}\left[\frac{15309 - 10935}{35}\right)$$

$$= \frac{4}{\cancel{81}} \times \frac{\cancel{4374}^{54}}{35} = \frac{216}{35}$$

$$\mu_4' = \int_{-\infty}^{\infty} x^4 f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^4 \frac{4}{81} x(9-x^2)\,dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81} x^5(9-x^2)\,dx$$

$$= \int_{-\infty}^{\infty} \frac{4}{81} (9x^5 - x^7)\,dx$$

$$= \frac{4}{81}\left[\frac{9x^6}{6} - \frac{x^8}{8}\right]_0^3$$

$$= \frac{4}{81}\left[\frac{9 \times 3^6}{6} - \frac{3^8}{8}\right]$$

$$= \frac{4}{81}\left[\frac{6561}{6} - \frac{6561}{8}\right]$$

$$= \frac{4}{81}\left[\frac{52488 - 39366}{48}\right]$$

$$= \frac{4}{81} \times \frac{13122}{48}$$ 162, 12

If the distribution function of a random variable is given by

$$F(x) = \begin{cases} 1 - 4/x^2 & , x > 2 \\ 0 & , x \le 2 \end{cases}$$

find (a) $P(x < 3)$
(b) Mean and Variance

$$\boxed{f(x) = \frac{d}{dx} F(x)}$$

$$\frac{d}{dx}\left(1 - 4/x^2\right) = 0 - 4\frac{d}{dx}(x^{-2})$$

$$= 0 - 4 \times -(2 \times x^{-3})$$

$$= 0 + 8x^{-3}$$

$$f(x) = 8/x^3$$

$$f(x) = \begin{cases} 8/x^3 & x > 2 \\ 0 & x \le 2 \end{cases}$$

$$= P(x < 3)$$

$$= \int_2^3 \frac{8}{x^3}\, dx$$

$$= 8 \int_2^3 (x^{-3})\, dx$$

$$x^n = \frac{x^{n+1}}{n+1}$$

$$= 8 \left[\frac{x^{-2}}{-2}\right]_2^3$$

$$= 8 \left(\frac{(3)^{-2}}{-2} - \frac{(2)^{-2}}{-2}\right)$$

$$= 8 \left(\frac{1}{9 \times 2} - \frac{1}{4 \times 2}\right) = 8 \left(\cancel{\frac{1}{18}} - \cancel{\frac{1}{8}}\right)$$

$$= \frac{8}{2} \left(\frac{1}{-9} - \frac{1}{-4}\right)$$

$$= \frac{8}{2} \left(\frac{4-9}{-36}\right) = \frac{8}{2} \times \frac{-5}{-36}$$

$$= 4 \times \frac{-5}{-36\,9} = \underline{\underline{\frac{+5}{9}}}$$

$$\text{Mean} = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= \int_2^{\infty} x \cdot \frac{8}{x^3}\, dx$$

$$= \int_2^{\infty} 8 x^{-2}\, dx$$

$$= 8 \int_2^{\infty} x^{-2}\, dx$$

$$= 8 \left[\frac{x^{-1}}{-1}\right]_2^{\infty}$$

$$= -8 \left(\frac{1}{x}\right)_2^{\infty}$$

$$= -8 \left[\frac{1}{\infty} - \frac{1}{2}\right]$$

$$= -8 \left(-\frac{1}{2}\right)$$

$$= \frac{8}{2}$$

$$\text{Mean} = \underline{\underline{4}}$$

~~Variance~~ $= E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$

$$E(x^2) = \int_2^{\infty} x^2 \cdot \frac{8}{x^3}\, \cancel{f(x)}\, dx$$

$$= \int_2^{\infty} \frac{8}{x}\, \cancel{f(x)}\, dx$$

$$= 8 \int_2^{\infty} x^{-1}\, dx$$

$E(x^2) = 8\left[\log x\right]_2^{\infty}$

$= 8\left[\log \infty - \log 2\right]$

$= \infty$

$\therefore$ Variance does not exist

A continuous random variable has a probability density function $f(x) = 6(x - x^2)$, $0 \le x \le 1$. Find mean and variance.

$$\text{Mean} = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= \int_0^1 x \cdot 6(x - x^2)\, dx$$

$$= \int_0^1 6(x^2 - x^3)\, dx$$

$$= 6 \int_0^1 (x^2 - x^3)\, dx$$

$$= 6 \left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1$$

$$= 6\left[\left(\frac{1}{3} - \frac{1}{4}\right) - (0)\right]$$

$$= \frac{6 \times (4-3)}{12} \qquad = \frac{6 \times 1}{12}$$

$$E(x) = \frac{6}{12} = \frac{1}{2}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^2\, 6(x - x^2)\, dx$$

$$= \int_0^1 6(x^3 - x^4)\,dx$$

$$= 6\int_0^1 (x^3 - x^4)\,dx$$

$$= 6\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1$$

$$= 6\left(\left[\frac{1}{4} - \frac{1}{5}\right] - [0]\right)$$

$$= \frac{6 \times (5-4)}{20}$$

$$E(x^2) = \frac{\cancel{6}^{3} \times 1}{\cancel{20}_{10}} = \frac{3}{10}$$

$$\text{variance} = E(x^2) - (E(x))^2$$

$$= \frac{3}{10} - \left(\frac{1}{2}\right)^2$$

$$= \frac{3}{10} - \frac{1}{4}$$

$$= \frac{12-10}{40} = \frac{\cancel{2}^{1}}{\cancel{40}_{20}}$$

$$V(x) = \frac{1}{20}$$

classmate
Date
Page

03/10/2019
Thursday

# MODULE - II

## Corelation and Regression

definition of corelation:

Two variables are said to be corelated if the change in one variable results in a corresponding change in other variable.

ie when two variables move together we say they are corelated.

for eg:- when the price of a commodity rises the supply for that commodity also rises. Here both variables moves together hence price and supply are corelated.

Corelation is defined as the tendency of two or more groups or series of items to vary together directly or inversly.

Boddington states that whenever some definite connection exist between the two or more groups, classes or series or data there is said to be a corelation

According to AM Tuttle, corelation is the analysis of the association between two or more variables.

04/10/2019
Friday

## Measures of studying corelation

Corelation between two variables can be measured by both graphic and algebraic method.

Scatter diagram and corelation graph are the two important graphic methods while coefficient of corelation is an algebraic method used for measuring corelation.

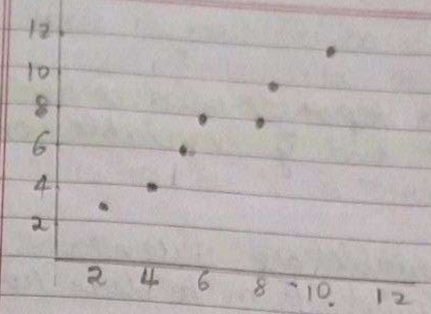
### I Scatter diagram

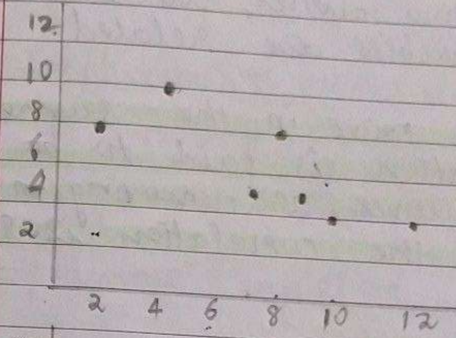
This is a graphical method of studying corelation between two variables. One of the variables is shown on the x-axis and the other on the y axis. Each pair of values is plotted on the graph by means of a dot mark.

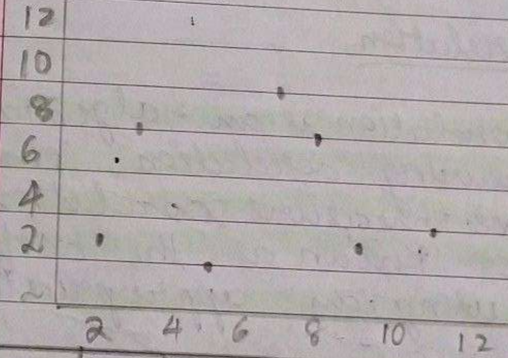
After all the items are plotted we get as many dots on the graph paper as the number of points.

If these points show some trend either upward or downward the two variables are said to be corelated.

If the plotted points donot show any trend the two variables are not corelated.

I

positive corelation

II

negative corelation

III

no corelation between x and y

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| x | 5 | 8 | 2 | 4 | 10 | 6 | 8 | I |
| y | 6 | 10 | 3 | 4 | 12 | 8 | 8 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| x | 12 | 7 | 10 | 8 | 4 | 9 | 2 | II |
| y | 4 | 5 | 4 | 8 | 10 | 5 | 8 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | 2 | 5 | 7 | 8 | 3 | 9 | 11 | III |
| y | 3 | 2 | 10 | 8 | 8 | 3 | 4 | |

I Corelation graph

Under this method separate curves are drawn for the x variable and y variable on the same graph paper.

The values of the variable are taken as ordinates of the points plotted. From the direction and closeness of the two curves, we can infer whether the variables are related.

If both the curves move in the same direction corelation is said to be positive. If the curves are moving in the opposite direction the corelation is said to be negative.

II Coefficient of corelation

Coefficient of corelation is an algebraic method of measuring corelation. Under this method we measure corelation by finding a value known as the coefficient of corelation using an appropriate formula.

Corelation Coefficient is a numerical value. It shows the degree or extend of corelation between two variables. This is a number lying between negative one and positive one (-1 and 1). When the corelation is negative it lies between -1 and 0

when the corelation is positive it lies between 0 and 1.
When the coefficient of corelation is 0, it indicates that there is no corelation between the variables.

When the corelation coefficient is 1, there is perfect corelation.

Coefficient of corelation can be computed by applying the methods below.

1) Karl pearson's method

2) Spearman's method

3) Concurrent deviation method

1) Karl pearson's coefficient of corelation

Karl pearson's method known as pearsonian coefficient of corelation is most widely used method of measuring corelation.

The pearsonian coefficient of corelation is denoted by the symbol r.
The formula for computing pearsonian coefficient of corelation is

$$r = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{n \sigma_x \sigma_y}$$

where $\sigma_x$ is equal to standard deviation of $x$ series

$\sigma_y$ is equal to standard deviation of $y$ series.

$n$ = number of pairs of observation

This is also known as product moment corelation coefficient.

The above formula can also be expressed in the form

$$r = \frac{n \Sigma xy - (\Sigma x \, \Sigma y)}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \sqrt{n \Sigma y^2 - (\Sigma y)^2}}$$

Ques Calculate coefficient of corelation

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x$: | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $y$: | 4 | 5 | 6 | 12 | 9 | 5 | 4 |

$n$ = no of

$n = 7$

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 2 | 4 | 8 | 4 | 16 |
| 3 | 5 | 15 | 9 | 25 |
| 4 | 6 | 24 | 16 | 36 |
| 5 | 12 | 60 | 25 | 144 |
| 6 | 9 | 54 | 36 | 81 |
| 7 | 5 | 35 | 49 | 25 |
| 8 | 4 | 32 | 64 | 16 |
| $\Sigma x = 35$ | $\Sigma y = 45$ | $\Sigma xy = 228$ | $\Sigma x^2 = 203$ | $\Sigma y^2 = 343$ |

$$r = \frac{n \Sigma xy - (\Sigma x \cdot \Sigma y)}{\sqrt{n \Sigma x^2 - (\Sigma x)^2} \sqrt{n \Sigma y^2 - (\Sigma y)^2}}$$

$$= \frac{7 \times 228 - (35 \cdot 45)}{\sqrt{7 \times 203 - (35)^2} \sqrt{7 \times 343 - (45)^2}}$$

$$= \frac{1596 - 1575}{\sqrt{1421 - 1225} \sqrt{2401 - 2025}}$$

$$= \frac{21}{\sqrt{196} \sqrt{376}}$$

$$= \frac{21}{14 \times 19.39} = \frac{21}{271.46}$$

$$= 0.0773$$

Ques Compute the coeff corelation between price and demand.

price : 7 8 9 6 5

demand : 8 6 7 9 10

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 7 | 8 | 56 | 49 | 64 |
| 8 | 6 | 48 | 64 | 36 |
| 9 | 7 | 63 | 81 | 49 |
| 6 | 9 | 54 | 36 | 81 |
| 5 | 10 | 50 | 25 | 100 |
| $\Sigma x = 35$ | $\Sigma y = 40$ | $\Sigma xy = 271$ | $\Sigma x^2 = 255$ | $\Sigma y^2 = 330$ |

$n = 5$

$$r = \frac{n\Sigma xy - (\Sigma x \cdot \Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

$$r = \frac{5 \times 271 - (35 \cdot 40)}{\sqrt{5 \times 225 - (35)^2}\sqrt{5 \times 330 - (35)^2}}$$

$$r = \frac{1355 - 1400}{\sqrt{1275 - 1225}\sqrt{1650 - 1600}}$$

$$r = \frac{-45}{\sqrt{50}\sqrt{50}}$$

$$= \frac{-45}{7.07 \times 7.07}$$

$$= \frac{-45}{49.98}$$

$$r = -0.9$$

Ques Compute Karl pearson's coeff of relation and comment on the result

Price : 11 12 13 14 15 16 17 18 19 20

Demand : 30 29 29 25 24 24 24 21 18 15

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 11 | 30 | 330 | 121 | 900 |
| 12 | 29 | 348 | 144 | 841 |
| 13 | 29 | 377 | 169 | 841 |
| 14 | 25 | 350 | 196 | 625 |
| 15 | 24 | 360 | 225 | 576 |
| 16 | 24 | 384 | 256 | 576 |
| 17 | 24 | 408 | 289 | 576 |
| 18 | 21 | 378 | 324 | 441 |
| 19 | 18 | 342 | 361 | 324 |
| 20 | 15 | 300 | 400 | 225 |
| $\Sigma x = 155$ | $\Sigma y = 239$ | $\Sigma xy = 3577$ | $\Sigma x^2 = 2485$ | $\Sigma y^2 = 5925$ |

$n = 10$

$$r = \frac{n\sum xy - (\sum x \cdot \sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r = \frac{10 \times 3577 - (155 \times 239)}{\sqrt{10 \times 2485 - (155)^2}\sqrt{10 \times 5925 - (239)^2}}$$

$$r = \frac{35770 - 37045}{\sqrt{24850 - 24025}\sqrt{59250 - 57121}}$$

$$r = \frac{-1275}{\sqrt{825}\sqrt{2129}}$$

$$r = \frac{-1275}{28.72 \times 46.14}$$

$$r = \frac{-1275}{1325.14}$$

$$r = -0.9621$$

very high negative corelation between price and demand.

14/10/2019
Monday

The coefficient of corelation is not affected by change of origin. So we can replace x and y in the formula by dx and dy. Hence the formula becomes

$$r = \frac{n\sum dxdy - (\sum dx \cdot \sum dy)}{\sqrt{n\sum dx^2 - (\sum dx)^2}\sqrt{n\sum dy^2 - (\sum dy)^2}}$$

where dx and dy are deviations taken from assumed averages in the two series.

Ques Calculate coefficient of corelation from the following data using 44 and 26 respectively as the origin of x and y.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 43 | 44 | 46 | 40 | 44 | 42 | 45 | 42 | 38 | 40 | 42 | 57 |
| Y | 29 | 31 | 19 | 18 | 19 | 27 | 27 | 29 | 41 | 30 | 26 | 10 |

Let $dx = x - 44$

$dy = y - 26$

-38 -5 -7 -21

| x | y | dx | dy | dxdy | dx² | dy² |
|---|---|---|---|---|---|---|
| 43 | 29 | -1 | 3 | -3 | 1 | 9 |
| 44 | 31 | 0 | 5 | 0 | 0 | 25 |
| 46 | 19 | 2 | -7 | -14 | 4 | 49 |
| 40 | 18 | -4 | -8 | 32 | 16 | 64 |
| 44 | 19 | 0 | -7 | 0 | 0 | 49 |
| 42 | 27 | -2 | 1 | -2 | 4 | 1 |
| 45 | 27 | +1 | 1 | 1 | 1 | 1 |
| 42 | 29 | -2 | 3 | -6 | 4 | 9 |
| 38 | 41 | -6 | 15 | -90 | 36 | 225 |
| 40 | 30 | -4 | 4 | -16 | 16 | 16 |
| 42 | 26 | -2 | 0 | 0 | 4 | 0 |
| 57 | 10 | 13 | -16 | -208 | 169 | 256 |
| | | -5 | -6 | -306 | 255 | 704 |

n=12

$$r = \frac{n\sum dx\,dy - (\sum dx \cdot \sum dy)}{\sqrt{n\sum dx^2 - (\sum dx)^2}\sqrt{n\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{12 \times -306 - (-5 \times -6)}{\sqrt{12 \times 255 - (-5)^2} \cdot \sqrt{12 \times 704 - (-6)^2}}$$

$$= \frac{-3672 - 30}{\sqrt{3060 - 25}\sqrt{8448 - 36}}$$

$$= \frac{-3702}{\sqrt{3035}\sqrt{8412}}$$

$$= \frac{-3702}{55.09 \times 91.71}$$

$$= \frac{-3702}{5052.39}$$

$$r = -0.7327$$

Ques In a corelation analysis of 13 pairs of observations of x and y, the following values are obtained. Sum of deviations of x and y values are -117 and -260. Sum of squares of deviations of x and y values are 1313 and 6580. Sum of products of deviations of x and y values is 2827. Find coefficient of corelation

$\sum dx = -117$
$\sum dy = -260$

n=13

$\sum dx^2 = 1313$
$\sum dy^2 = 6580$

$\sum dx\,dy = 2827$

$$r = \frac{n\sum dx\,dy - (\sum dx \cdot \sum dy)}{\sqrt{n\sum dx^2 - (\sum dx)^2}\ \sqrt{n\sum dy^2 - (\sum dy)^2}}$$

$$= \frac{13 \times 2827 - (-117 \times -260)}{\sqrt{(13 \times 1313) - (-117)^2}\ \sqrt{(13 \times 6580) - (-260)^2}}$$

$$= \frac{36751 - 30420}{\sqrt{17069 - (13689)}\ \sqrt{85540 - 67600}}$$

$$= \frac{6331}{\sqrt{3380}\ \sqrt{17940}}$$

$$= \frac{6331}{58.137 \times 133.940}$$

$$= \frac{6331}{7785.93}$$

$$r = 0.8131$$

NOTE

If we can take the deviations from the actual mean then $dx = x - \bar{x}$ and $dy = y - \bar{y}$ so that $\sum dx = 0$ and $\sum dy = 0$ then the formula becomes

$$\frac{\sum dx\,dy}{\sqrt{\sum dx^2}\ \sqrt{\sum dy^2}}$$

Pull Find the coefficient of correlation from the following data.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 12 | 20 | 15 | 22 | 18 | 24 | 20 | 12 | 15 | 22 |
| Y | 30 | 35 | 28 | 36 | 29 | 39 | 30 | 25 | 30 | 38 |

$n = 10$ $\quad \bar{x} = \frac{\sum x}{n} = \frac{180}{10} = 18$

$\bar{y} = \frac{\sum y}{n} = \frac{320}{10} = 32$

~~dx = dx~~

$dx = x - \bar{x}$
$= x - 18$

$dy = y - \bar{y}$
$= y - 32$

| $x$ | $y$ | $dx$ | $dy$ | $dxdy$ | $dx^2$ | $dy^2$ |
|---|---|---|---|---|---|---|
| 12 | 30 | -6 | -2 | 12 | 36 | 4 |
| 20 | 25 | 2 | 3 | 6 | 4 | 9 |
| 15 | 28 | -3 | -4 | 12 | 9 | 16 |
| 22 | 36 | 4 | 4 | 16 | 16 | 16 |
| 18 | 29 | 0 | -3 | 0 | 0 | 9 |
| 24 | 39 | 6 | 7 | 42 | 36 | 49 |
| 20 | 30 | 2 | -2 | -4 | 4 | 4 |
| 12 | 25 | -6 | -7 | +42 | 36 | 49 |
| 15 | 30 | -3 | -2 | +6 | 9 | 4 |
| 22 | 38 | 4 | 6 | 24 | 16 | 36 |
| | | 0 | 0 | 156 | 166 | 196 |

$$r = \frac{\sum dx\,dy}{\sqrt{\sum dx^2}\sqrt{\sum dy^2}}$$

$$= \frac{156}{\sqrt{166}\sqrt{196}}$$

$$= \frac{156}{12.88 \times 14}$$

$$= \frac{156}{180.72}$$

$$r = \underline{\underline{0.865}}$$

NOTE :-

Coefficient of corelation is not affected by change of scale

∴ If we multiply or divide all the values of a variable by a constant, the coefficient of corelation will not change.

Ques Find coefficient of corelation between x and y from the following data giving test score of 10 candidates in mathematics and statistics and interpret

| x | 98 | 70 | 40 | 20 | 85 | 75 | 95 | 80 | 10 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|
| y | 85 | 65 | 32 | 30 | 80 | 60 | 80 | 70 | 20 | 10 |

| x | y | $x^2$ | $y^2$ | xy |
|---|---|---|---|---|
| 98 | 85 | 9604 | 7225 | 8330 |
| 70 | 65 | 4900 | 4225 | 4550 |
| 40 | 32 | 1600 | 1024 | 1280 |
| 20 | 30 | 400 | 900 | 600 |
| 85 | 80 | 7225 | 6400 | 6800 |
| 75 | 60 | 5625 | 3600 | 4500 |
| 95 | 80 | 9025 | 6400 | 7600 |
| 80 | 70 | 6400 | 4900 | 5600 |
| 10 | 20 | 100 | 400 | 200 |
| 5 | 10 | 25 | 100 | 50 |
| 578 | 532 | 44904 | 35174 | 39510 |

n=10

$$r = \frac{n\sum xy - (\sum x \cdot \sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

$$= \frac{10 \times 39510 - (578 \cdot 532)}{\sqrt{10 \times 44904 - (578)^2}\sqrt{10 \times 35174 - (532)^2}}$$

$$= \frac{395100 - 307496}{\sqrt{449040 - 334084}\sqrt{351740 - 283024}}$$

$$= \frac{87604}{\sqrt{114956}\ \sqrt{68716}}$$

$$= \frac{87604}{339.05 \times 262.13}$$

$$= \frac{87604}{88875.175}$$

$r = 0.985$

positive corelation

15/10/2019
Tuesday

Ques Find coefficient of corelation between age and playing habit of the following students:–

| Age | 14.5-15.5 | 15.5-16.5 | 16.5-17.5 | 17.5-18.5 | 18.5-19.5 | 19.5-20.5 |
|---|---|---|---|---|---|---|
| No of students | 250 | 200 | 150 | 120 | 100 | 80 |
| regular players | 200 | 150 | 90 | 48 | 30 | 12 |

Playing habit in % = $\frac{\text{Regular player}}{\text{No of students}} \times 100$

$n = 6$

63

| Age(x) | Playing habit (%) y | dx x-18 | dy y-50 | dxdy | $dx^2$ | $dy^2$ |
|---|---|---|---|---|---|---|
| 15 | 80 | -3 | 30 | -90 | 9 | 900 |
| 16 | 75 | -2 | 25 | -50 | 4 | 625 |
| 17 | 60 | -1 | 10 | -10 | 1 | 100 |
| 18 | 40 | 0 | -10 | 0 | 0 | 100 |
| 19 | 30 | 1 | -20 | -20 | 1 | 400 |
| 20 | 15 | 2 | -35 | -70 | 4 | 1225 |
| $\Sigma x = 105$ | $\Sigma y = 300$ | $\Sigma dx = -3$ | $\Sigma dy = 0$ | $\Sigma dxdy = -240$ | $\Sigma dx^2 = 19$ | $\Sigma dy^2 = 3350$ |

$$r = \frac{n \Sigma dxdy - (\Sigma dx \cdot \Sigma dy)}{\sqrt{n \Sigma dx^2 - (\Sigma dx)^2}\ \sqrt{n \Sigma dy^2 - (\Sigma dy)^2}}$$

$$= \frac{6 \times -240 - (-3 \times 0)}{\sqrt{6 \times 19 - (-3)^2}\ \sqrt{6 \times 3350 - (0)^2}}$$

$$= \frac{-1440 + 0}{\sqrt{114 - 9}\ \sqrt{20100}}$$

$$= \frac{-1440}{\sqrt{105} \times \sqrt{20100}}$$

$$= \frac{-1440}{10.24 \times 141.77}$$

$$= \frac{-1440}{1451.72}$$

$$r = -0.991$$

Ques With the following data in 6 cities calculate the coff of corelation by Pearson's method b/w the density of population and the death rate.

| | Area | Population | No of deaths |
|---|---|---|---|
| A | 150 | 30 | 300 |
| B | 180 | 90 | 1440 |
| C | 100 | 40 | 560 |
| D | 60 | 42 | 840 |
| E | 120 | 72 | 1224 |
| F | 80 | 24 | 312 |

$$\text{Density} = \frac{\text{Population}}{\text{area}}$$

Death ~~rat~~ rate - $\frac{\text{No of death}}{\text{population}}$

| density $(x)_{\times 10}$ | death rate (y) | dx (x-4) | dy (y-16) | dxdy | $dx^2$ | $dy^2$ |
|---|---|---|---|---|---|---|
| 2 | 10 | -2 | -6 | 12 | 4 | 36 |
| 5 | 16 | 1 | 0 | 0 | 1 | 0 |
| 4 | 14 | 0 | -2 | 0 | 0 | 4 |
| 7 | 20 | 3 | 4 | 12 | 9 | 16 |
| 6 | 17 | 2 | 1 | 2 | 4 | 1 |
| 3 | 13 | -1 | -3 | 3 | 1 | 9 |
| 27 | 90 | 3 | -6 | 29 | 19 | 66 |

$n = 6$

$$r = \frac{n \sum dxdy - (\sum dx \cdot \sum dy)}{\sqrt{n \sum dx^2 - (\sum dx)^2}\sqrt{n \sum dy^2 - (\sum dy)^2}}$$

$$= \frac{6 \times 29 - (3 \times -6)}{\sqrt{6 \times 19 - (3)^2}\sqrt{6 \times 66 - (-6)^2}}$$

$$= \frac{174 + 18}{\sqrt{105}\sqrt{360}}$$

$$= \frac{192}{10.24 \times 18.97}$$

$$r = \frac{192}{194.25} = 0.988$$

## Properties of Corelation Coefficient

① Coefficient corelation is a pure number and is independent of the units of measurement.

② It lies between $-1$ and $+1$.

③ It does not change with reference ~~wit~~ to change of origin or change of scale

④ coefficient of corelation b/w $x$ and $y$ is same as that between $y$ and $x$.

## Covariance between $x$ and $y$

covariance b/w two variables $x$ and $y$ is defined as

$$\frac{\Sigma(x-\bar{x})\cdot(y-\bar{y})}{n}$$

which can be simplified as

$$\frac{\Sigma xy}{n} - \left[\frac{\Sigma x}{n} \times \frac{\Sigma y}{n}\right] \text{ OR } \frac{\Sigma dxdy}{n} - \left[\frac{\Sigma dx}{n} \times \frac{\Sigma dy}{n}\right]$$

When the covariance between two variables is divided by the product of standard deviations of those variables, we get the coefficient of corelation

$$\text{coefficient of correlation} = \frac{\text{Covariance b/w } x \text{ and } y}{\begin{pmatrix}\text{standard}\\ \text{deviation}\\ \text{of } x\end{pmatrix} \times \begin{pmatrix}\text{standard}\\ \text{deviation}\\ \text{of } y\end{pmatrix}}$$

Ques Find the coefficient of corelation between $x$ and $y$ using covariance method.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| $y$ | 4 | 5 | 6 | 8 | 9 | 7 | 10 |

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 2 | 4 | 8 | 4 | 16 |
| 3 | 5 | 15 | 9 | 25 |
| 4 | 6 | 24 | 16 | 36 |
| 5 | 8 | 40 | 25 | 64 |
| 6 | 9 | 54 | 36 | 81 |
| 7 | 7 | 49 | 49 | 49 |
| 8 | 10 | 80 | 64 | 100 |
| 35 | 49 | 270 | 203 | 371 |

$n = 7$

$$\text{Covariance} = \frac{\Sigma xy}{n} - \left(\frac{\Sigma x}{n} \times \frac{\Sigma y}{n}\right)$$

$$= \frac{270}{7} - \left(\frac{35}{7} \times \frac{49}{7}\right)$$

$$= 38.57 - (5 \times 7)$$

$$= 38.57 - 35$$

$$= 3.57$$

Standard deviation

$$\boxed{\sigma_x = \sqrt{\frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2}} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$= \sqrt{\frac{203}{7} - \left(\frac{35}{7}\right)^2}$$

$$= \sqrt{29 - 25}$$

$$= \sqrt{4}$$

$$\sigma_x = 2$$

$$\boxed{\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\frac{\Sigma y}{n}\right)^2}} = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n}}$$

$$= \sqrt{\frac{371}{7} - \left(\frac{49}{7}\right)^2}$$

$$= \sqrt{53 - 49}$$

$$= \sqrt{4}$$

$$\sigma_y = 2$$

Coeff of corelation $= \dfrac{3.57}{2 \times 2}$

$$= \frac{3.57}{4} = 0.8925$$

16/10/2019
Wednesday

# Principles of Least Squares and Curve Fitting

- Curve Fitting

Let $(x_1, y_1)$ $(x_2, y_2)$ .... $(x_n, y_n)$ be $n$ pairs of values of two variables $x$ and $y$

ie $x_i$ is the value of the variable $x$ and $y$ takes the value $y_i$

Plot these values on a graph and get a scatter diagram. From this scatter diagram it is possible to visualise a curve approximating the data.

The curve maybe a straight line or a non-linear curve. When the curve is a straight line it exists a linear relation between the two variables. The non-linear curves maybe parabola, exponential curve and logarithmic curve.

The equation showing the linear relation can be expressed in the form of $y = a + bx$.

Similarly equations showing non-linear relations can be expressed in the form $y = a + bx + cx$, $y = a \times b^x$, $y = a x^b$ .etc

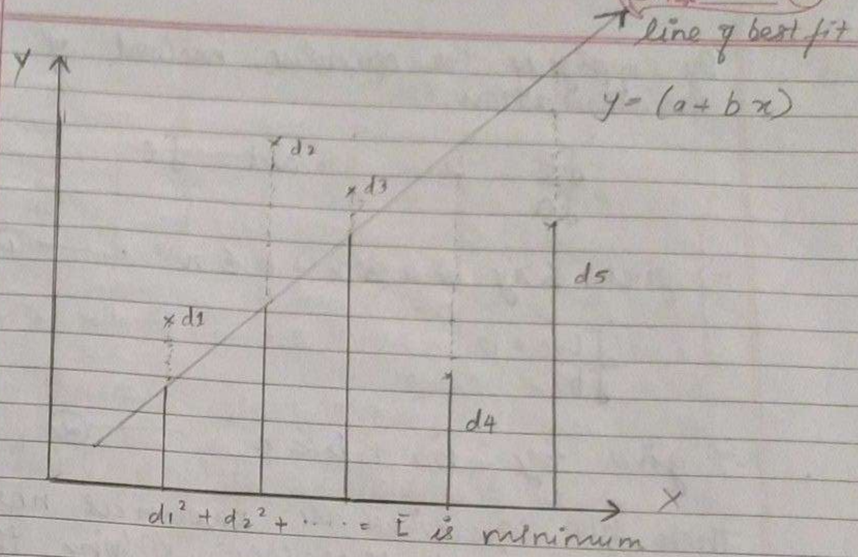
(exam) → Finding equations of the approximating curve which fit the given data is called curve fitting. The curve fitting enables us to express a relation between two variables by means of an equation.

* **Principle of Least Squares**

The scatter diagram obtained with reference to the pairs of observations of $x$ and $y$ indicates the approximating curve to the data.

Then the problem is to find the equation of the approximating curve which is best fitting. The approximating curve is known as best fitting curve if it is on the basis of the principle known as principle of least squares.

The principle of least squares states that the sum of the squares of the difference between the actual values of $y$ and the corresponding estimated values of $y$ should be minimum.
i.e. when $x$ takes $x_i$ the observed value of $y$ is $y_i$, the estimated value of $y$ from the relation is $a + bx_i$ but $y_i$ and $a + bx_i$ need not be equal.
The principle of least squares states that $a$ and $b$ should be so chosen that $\Sigma(y_i - a - bx_i)^2 = E$ should be minimum.




* **Fitting a straight line**

Let the collected data for the variable of $x$ and $y$ be $(x_1, y_1)$ $(x_2, y_2)$ ..... $(x_n, y_n)$. These when plotted gives a scatter diagram and we may find that the straight may be an approximating curve to the scatter diagram i.e. there exist a linear relation between $x$ and $y$.

The simple type of equation representing a linear relation can be written as $y = a + bx$.
The principle of least squares state that the sum of the squares of the differences between each observed value of $y$ and its estimated value from the relation should be minimum.
i.e; $E = \Sigma(y_i - a - bx_i)^2$ should be minimum

By applying the calculus method of minimisation

$\frac{\delta E}{\delta a} = 0$ and ~~$\delta E$~~

$\Rightarrow$ gives $\Sigma xy = a\Sigma x + b\Sigma x^2$ —— ①

$\frac{\delta E}{\delta b} = 0$

$\Rightarrow$ gives $\Sigma y = na + b\Sigma x$ —— ②

These two equations are known as normal equations and therefore solving these two equations we get the values of a and b.

Substituting them in the equation $y = a + bx$ we get the best fitting straight line

Ques Fit a straight line to the following data by the method of least squares.

| | | | | | |
|---|---|---|---|---|---|
| x | 3 | 5 | 8 | 4 | 10 |
| y | 2 | 3 | 7 | 5 | 8 |

Let the equation for the straight line be $y = a + bx$ ; normal equations are

$\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

| x | y | xy | $x^2$ |
|---|---|---|---|
| 3 | 2 | 6 | 9 |
| 5 | 3 | 15 | 25 |
| 8 | 7 | 56 | 64 |
| 4 | 5 | 20 | 16 |
| 10 | 8 | 80 | 100 |
| $\Sigma x = 30$ | $\Sigma y = 25$ | $\Sigma xy = 177$ | $\Sigma x^2 = 214$ |

$n = 5$

Substituting the values;

$25 = 5a + b30$

$177 = 30a + 214b$

i.e. $5a + 30b = 25$ $\times 6$ —— ③

$30a + 214b = 177$

$30a + 180b = 150$

$- \underline{30a + 214b = 177}$

$0a \quad -34b = -27$

$b = \frac{27}{34}$ $= 0.79$

$5a = 25 - 30b$

$= 25 - 30 \times \frac{27}{34}$ $0.79$

~~$5a = \frac{850 - 810}{34}$~~

$5a = \frac{40}{34}$

$a = \frac{40}{34} \times \frac{1}{5}$

$= \frac{8}{34}$ $= \frac{8}{34}$

$5a = 25 - 23.7$

$5a = 1.3$

$a = \frac{1.3}{5}$

$a = 0.26$

∴ The equation of the straight line is

$y = 0.26 + 0.79x$

Ques Fit a straight line to the following data by the method of least squares. Find y when x = 5

| | | | | | |
|---|---|---|---|---|---|
| x | 0 | 1 | 2 | 3 | 4 |
| y | 1.0 | 1.8 | 3.3 | 4.5 | 6.3 |

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$\Sigma y = na + b\Sigma x$

| x | y | xy | x2 |
|---|---|---|---|
| 0 | 10 | 10 | 0 |
| 10 | 18 | 180 | 100 |
| 20 | 33 | 660 | 400 |
| 30 | 45 | 1350 | 900 |
| 40 | 63 | 2520 | 1600 |
| 100 | 169 | 4720 | 3000 |

n = 5

$169 = 5a + 100b$ —— (1) × 20

$4720 = 100a + 3000b$ —— (2)

(1) × 20

$3380 = 100a + 2000b$

$- \quad 4720 = 100a + 3000b$

$-1340 = 0a - 1000b$

4720 = 100a + 4020

$1000b = 1340$

$b = \frac{1340}{1000}$

$b = 1.34$

$169 = 5a + 100 \times 1.340$

$169 = 5a + 134$

$5a = 169 - 134$

$a = 35/5$

$a = 7$

0.7

a + bx

∴ The eqn becomes $y = 0.7 + 1.34x$

$y = 0.7 + 1.34 \times 5 = 0.7 + 6.7$

$= 7.4$

Ques Find the straight line to the following data

| | | | | | |
|---|---|---|---|---|---|
| Sales | 60 | 65 | 62 | 68 | 70 |
| Purchase | 50 | 52 | 55 | 60 | 60 |

| $x$ | $y$ | $xy$ | $x^2$ |
|---|---|---|---|
| 60 | 50 | 3000 | 3600 |
| 65 | 52 | 3380 | 4225 |
| 62 | 55 | 3410 | 3844 |
| 68 | 60 | 4080 | 4624 |
| 70 | 60 | 4200 | 4900 |
| 325 | 277 | 18070 | 21193 |

$18070 = 325a + 21193b$

$277 = 5a + 325b \quad \text{—} \times 65$

$18070 = 325a + 21193b$

$18005 = 325a + 21125b$

$-65 = 0a + -68b$

20133.3

$b = \frac{+65}{68}$

$b = 0.95$

$5a = 277 - 325b$

$5a = 277 - 325 \times 0.95$

$5a = 277 - 308.75$

$5a = -31.75$

$a = -6.35 \quad \longrightarrow (6.9)$

$\therefore$ equation of straight line is

$$y = -6.35 + 0.95x$$

## Fitting a parabolic equation (second degree equation)

Let the collected data for the variables $x$ and $y$ be $(x_1, y_1)$ $(x_2, y_2)$ ... $(x_n, y_n)$. These pairs of values when plotted on a graph paper give a scatter diagram.

Suppose a parabola is the approximating curve of the scatter diagram. The equation of the parabola is of the form

$$y = a + bx + cx^2$$

Then by the principle of least squares

$$E = \Sigma [y_i - (a + bx_i + cx_i^2)]^2$$ should be minimum

Applying calculus method

$$\frac{\delta E}{\delta a} = 0 \qquad \frac{\delta E}{\delta b} = 0 \qquad \frac{\delta E}{\delta c} = 0$$

They intum gives the equations:-

$$\frac{\delta E}{\delta a} = 0 \Rightarrow \Sigma y = na + b\Sigma x + c\Sigma x^2$$

$$\frac{\delta E}{\delta b} = 0 \Rightarrow \Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3$$

$$\frac{\delta E}{\delta c} = 0 \Rightarrow \Sigma x^2 y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4$$

These equations are known as normal equations, solving them we get the values a, b and c.

Substituting them in the equation $y = a + bx + cx^2$ we get the equation of the best fitting parabola

Que Fit the parabola $y = a + bx + cx^2$ for the following data by the method of least squares. Estimate the value of y when $x = 10$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| y | 2 | 6 | 7 | 8 | 10 | 11 | 11 | 10 | 9 |

| x | y | xy | $x^2$ | $x^2y$ | $x^3$ | $x^4$ |
|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 1 | 2 | 1 | 1 |
| 2 | 6 | 12 | 4 | 24 | 8 | 16 |
| 3 | 7 | 21 | 9 | 63 | 27 | 81 |
| 4 | 8 | 32 | 16 | 128 | 64 | 256 |
| 5 | 10 | 50 | 25 | 250 | 125 | 625 |

a = -0.98
b = 2.55
c = -0.20

7 (7, 45
5 (7, 45

| x | y | xy | $x^2$ | $x^2y$ | $x^3$ | $x^4$ |
|---|---|---|---|---|---|---|
| 6 | 11 | 66 | 36 | 396 | 216 | 1296 |
| 7 | 11 | 77 | 49 | 539 | 343 | 2401 |
| 8 | 10 | 80 | 64 | 640 | 512 | 4096 |
| 9 | 9 | 81 | 81 | 720 | 729 | 6561 |

$\Sigma x = 45$ | $\Sigma y = 74$ | $\Sigma xy = 421$ | $\Sigma x^2 = 285$ | $\Sigma x^2 y = 2762$ | $\Sigma x^3 = 2025$ | $\Sigma x^4 = 15333$

n = 9

$74 = 7a + 45b + 285c$

$421 = 45a + 285b + 2025c$

$2762 = 285a + 2025b + 15333c$

Imont cusk for Exam

$\cancel{2} = 2 \times 5$

$3\cancel{20A} =$

(2VS)

17/10/2019
Thursday

# Regression Analysis

Regression analysis means the estimation or prediction of the unknown value of one variable from the known value of the other variable. It is a statistical device used to study the relationship between two or more variables that are related.

In the words of M.M Blair, regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data.

## Dependent and Independent variables

The variable whose value is influenced or is to be predicted is called dependent variable. and the variable which influences the value or is used for prediction is called independent variable.

Bivariate - data - x and y varia.. a value (imp)

## Line of best fit or Regression Line

When the given bivariate data are plotted on the graph, we get the scatter diagram.

If the points of the scatter diagram concentrate around a straight line that line is called the line of best fit. i.e the line of best fit is that line which is closer to the points of the scatter diagram.

This line is also known as regression line so, regression line is a graphic technique to show the functional relationship between dependent variables and independent variables.

It shows average relationship between the variables.

*. There are two lines of regression because while estimating or predicting the value of y for any given value of $x$ we take y as dependent variable and $x$ as independent variable. Then we get line of regression of y on $x$.

Similarly for estimating or predicting $x$ for any given value of y we use the regression of $x$ on y.
Here $x$ is dependent and y is independent variable.
Thus there are two regression lines.

classmate
Date
Page

## Regression Equations

- Regression equations are the equations of the regression lines. It is a mathematical relation between dependent and independent variables. There are two regression lines and therefore two regression equations

(1) Regression equation of $y$ on $x$.

(2) Regression equation of $x$ on $y$.

Derivation of regression equation on $y$ on $x$ when the relation is linear

Let $y = a + bx$ be the equation of the regression line on on $y$. To find the best value of $a$ and $b$, we apply the method of least squares

According to method of least squares, $\Sigma[y - (a+bx)]^2$ is minimum.

By applying the method of calculus we get two normal equation

$$\Sigma y = na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

Solving the two equations we get

classmate
Date
Page

$$b = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

and

$$a = \frac{\Sigma x - b\Sigma x}{n}$$

$$= \frac{\Sigma y}{n} - \frac{b\Sigma x}{n}$$

$$= \bar{y} - b\bar{x}$$

Regression of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$\bar{y} = \frac{\Sigma y}{n}$

$\bar{x} = \frac{\Sigma x}{n}$

Regression of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$b_{yx} = \frac{n\Sigma xy - (\Sigma x \cdot \Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$b_{xy} = \frac{n\Sigma xy - (\Sigma x \cdot \Sigma y)}{n\Sigma y^2 - (\Sigma y)^2}$$

where $b_{xy}$, $b_{yx}$ are regression coefficient

Ques From the following data of values of x and y find the regression equation of y on x

| | | | | | |
|---|---|---|---|---|---|
| x | 2 | 3 | 4 | 5 | 6 |
| y | 3 | 5 | 4 | 8 | 9 |

→ The equation of regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

| x | y | xy | $x^2$ |
|---|---|---|---|
| 2 | 3 | 6 | 4 |
| 3 | 5 | 15 | 9 |
| 4 | 4 | 16 | 16 |
| 5 | 8 | 40 | 25 |
| 6 | 9 | 54 | 36 |
| $\Sigma x = 20$ | $\Sigma y = 29$ | $\Sigma xy$ 131 | $\Sigma x^2 = 90$ |

$n = 5$

$$b_{yx} = \frac{n\Sigma xy - (\Sigma x \cdot \Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$

$$= \frac{5 \times 131 - (20 \times 29)}{5 \times 90 - (20)^2}$$

$$= \frac{655 - 580}{450 - 400}$$

$$= \frac{75}{50}$$

$$b_{yx} = 1.5$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{20}{5} = 4$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{29}{5} = 5.8$$

$$y - 5.8 = 1.5(x - 4)$$

Tuesday 2/10/2019

Ques From the given data find the regression equation of x on y.

| | | | | | |
|---|---|---|---|---|---|
| x | 5 | 6 | 7 | 3 | 2 |
| y | 4 | 5 | 8 | 2 | 1 |

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

| x | y | xy | $y^2$ |
|---|---|---|---|
| 5 | 4 | 20 | 16 |
| 6 | 5 | 30 | 25 |
| 7 | 8 | 56 | 64 |
| 3 | 2 | 6 | 4 |
| 2 | 1 | 2 | 1 |
| 23 | 20 | 114 | 110 |

$\bar{x} = \frac{23}{5} = 4.6$

$\bar{y} = \frac{20}{5} = 4$

$b_{xy} = \frac{5 \times 114 - (23 \cdot 20)}{5 \times 110 - (20)^2}$

$= \frac{570 - 460}{550 - 400}$

$= \frac{110}{150}$

$b_{xy} = 0.73$

eqn $(x - 4.6) = 0.73(y - 4)$

$x = 0.73y + 1.68$

Ques From the following data of the age of husband and the age of wife form a regression equation and calculate the husband's age when the wife's age is 16. Also find the age of wife when husband's age is 40.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Hus's age | 36 | 23 | 27 | 28 | 28 | 29 | 30 | 31 | 33 | 35 |
| wife's age | 29 | 18 | 20 | 22 | 27 | 21 | 29 | 27 | 29 | 28 |

Take husband's age as x and wife's age as y.

| x | y | $x^2$ | $y^2$ | xy |
|---|---|---|---|---|
| 36 | 29 | 1296 | 841 | 1044 |
| 23 | 18 | 529 | 324 | 414 |
| 27 | 20 | 729 | 400 | 540 |
| 28 | 22 | 784 | 484 | 616 |
| 28 | 27 | 784 | 729 | 756 |
| 29 | 21 | 841 | 441 | 609 |
| 30 | 29 | 900 | 841 | 870 |
| 31 | 27 | 961 | 729 | 837 |
| 33 | 29 | 1089 | 841 | 957 |
| 35 | 28 | 1225 | 784 | 980 |
| 272 | 250 | 7138 | 5930 | 7623 |

$b_{yx} = 0.89$

$b_{xy} = 0.75$

$y - 25 = 0.89(x - 30)$

$y = 0.89x - 26.7 + 25$

$y = 0.89x - 1.7$

$x - 30 = 0.75(y - 25)$

$x = 0.75y - 18.75 + 30$

$x = 0.75y + 11.25$

when $y = 16$

$x = 0.75 \times 16 + 11.25$

$x = 23.25$

when $x = 40$

$y = 0.89 \times 40 - 1.7$

$= 33.9$

<u>Relation b/w Correlation coefficient and Regression Coefficient</u>

$$b_{yx} = r \frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

## How to get $b_{xy}$ and $b_{yx}$ from regression equations

$b_{yx}$ is obtained from the equation of y on x and $b_{xy}$ from x on y.

When the regression equation of y on x is expressed in the form $y = ax + b$ then a is $b_{yx}$.

Similarly regression equation of x on y is expressed in the form $x = cy + d$ then c is $b_{xy}$

## How to identify the two regression equation

(if $b_{xy} \times b_{yx}$ is not greater than 1)

By supposing one of the equation as the regression of y on x and the other as x on y we obtain $b_{yx}$ and $b_{xy}$ If the product of these two is numerically not greater than one. then our supposition is true.

$$b_{yx} \times b_{yx} = \frac{r\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y} = r^2$$

$$\therefore r = \sqrt{b_{yx} \times b_{xy}}$$