UNIT-1

Q1 Explain types of data and methods of collecting primary data?

Ans1 Data is a set of values that represents measurements or observations that are collected as a source of information.

Types of data is as follows:

1. Primary Data – The data that are collected for the first time by an investigator for a specific purpose is known as primary data. for ex: Census of India.

2. Secondary Data – The data that are sourced from someplace that has originally collected it. this means that this kind of data has already been collected by some researchers or investigators in the past and is available either in published or unpublished form.

Methods of collecting primary data is as follows:

1. Quantitative Data Collection Method

It is based on the mathematical calculations using various formats like close-ended questions, correlation and regression methods, mean, median or mode measures. this method is cheaper than qualitative data collection methods and it can be applied in a short duration of time.

② Qualitative Data Collection Method

It does not involve any mathematical calculations. this method is closely associated with elements that are not quantifiable. this qualitative data collection method includes interviews, questionnaries, observations, case studies, etc..

Q② What do you understand by sampling. Also explain methods of sampling and it's merits and limitations?

Ans② The sampling method or sampling technique is the process of studying the population by gathering information and analyzing that data. It is basis of the data where the sample space is enormous.

The different types of sampling methods is as follows:

① Probability Sampling – It utilizes some form of random selection. In this method, all the eligible individuals have a chance of selecting the sample from the whole sample space.

② Non-probability Sampling – It is a technique in which the researcher selects the sample based on subjective judgement rather than the random selection.

The Merits of sampling is as follows:

① Cost-effective – sampling is less expensive than studying the entire population.

② Time-Saving – It requires less time to collect and analyze data.

③ Manageability – Smaller data sets are easier to handle and process.

④ Feasibility – Sampling is practical when studying the entire population is impossible.

Limitations of sampling is as follows:

① Sampling error – Samples may not accurately represent the population.

② Bias – Improper techniques can introduce bias and skew results.

③ Limited Scope – Results may not be generalizable to the entire population.

Date
Page

Q③ The pie graph given below shows the breakup of the cost of construction of a house. Assuming that the total cost of construction is Rs.600000, answer the question given below:

(a) The sum spent on cement is
(b) The sum spent on work surpasses the sum spent on steel by
(c) The sum spent on cement, steel and supervision is the thing that percent of the aggregate expense of development.
(d) The sum spent on work surpasses the sum spent on supervision by:

Sol ③ (a) The sum spent on cement is:

Cement accounts for 20% of the total cost of construction, which is Rs 600000.

Sum spent on cement $= \frac{20}{100} \times 600000$

$= \cancel{120000 \text{ Rs}}$ Rs. 1,20,000

(b) The cost of labour = 25%
The cost of steel = 15%

Difference $= (0.25 - 0.15) * 600000 - 0.10 * 600000$
$=$ Rs. 60,000

(c) The cost of labour (work) is 25%
The cost of supervision is 15%

Difference = (25% - 15%) * 600000
= 0.1 × 600000
= Rs. 60000

(d) The sum spent on cement = Rs. 120000
The sum spent on steel = Rs. 90000
The sum spent on supervision = Rs. (0.15 × 600000)

Total = 120000 + 90000 + 90000
= 300000

Now, % of the total construction cost =
(300000/600000) × 100%
= 50%

Q(4) Write short notes on the following:

(1) Histogram

(2) Ogive

(3) Frequency polygon

(4) Frequency curve

Date

Page

Ans ④ ① Histogram — A graphical representation of continuous data using bars of varying heights.

the characteristics is as follows:

① Rectangular bars represent classes or intervals.

② No gaps between bars.

② Ogive — A graphical representation of commulative frequencies.

the characteristics is as follows:

① Plot of commulative frequencies against upper class limits.

② "S" shaped curve.

③ Frequency polygon — A graphical representation of frequency distribution using line segments.

the characteristics is as follows:

① plot of frequencies against class mid-points.

② line segments connect frequency points.

④ frequency curve — A smooth curve representing frequency distribution.

the characteristics is as follows:

① smooth curve passes through frequency points.

② used for large datasets.

Q⑤ In a city, the weekly observations made in a study on the cost of a living index are given in the following table:
Draw a frequency polygon for the data above

| cost of living index | No. of weeks |
|---|---|
| 140-150 | 5 |
| 150-160 | 10 |
| 160-170 | 20 |
| 170-180 | 9 |
| 180-190 | 6 |
| 190-200 | 2 |
| Total | 52 |

Date
Page

Sol:

| Class | Mid-Value | frequency |
|---|---|---|
| 140-150 | 145 | 5 |
| 150-160 | 155 | 10 |
| 160-170 | 165 | 20 |
| 170-180 | 175 | 9 |
| 180-190 | 185 | 6 |
| 190-200 | 195 | 2 |

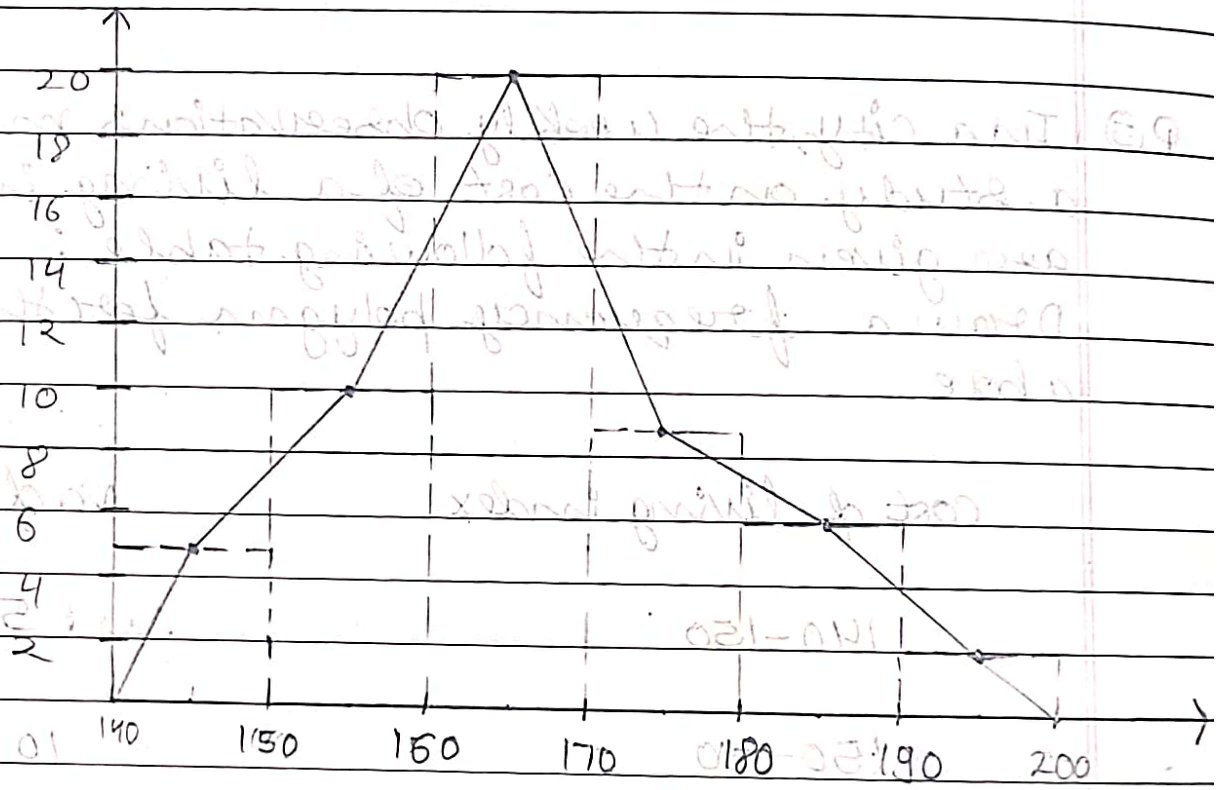


fig: frequency polygon

X

UNIT-II

Q1 Calculate Mean, Median and Mode for the following distribution:

| Marks | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 |
|---|---|---|---|---|---|---|---|---|
| frequency | 3 | 18 | 20 | 22 | 16 | 13 | 12 | 4 |

Sol:

| Marks | f | x | fx |
|---|---|---|---|
| 10-20 | 3 | 15 | 45 |
| 20-30 | 18 | 25 | 450 |
| 30-40 | 20 | 35 | 700 |
| 40-50 | 22 | 45 | 990 |
| 50-60 | 16 | 55 | 880 |
| 60-70 | 13 | 65 | 845 |
| 70-80 | 12 | 75 | 900 |
| 80-90 | 4 | 85 | 340 |
| | $\Sigma f = 108$ | | $\Sigma fx = 5150$ |

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\text{Mean} = \frac{5150}{108}$$

$$\text{Mean} = 47.68$$

$$\text{Median} = \frac{N}{2}^{th} = \frac{108}{2} = 54^{th} \text{ term}$$

$$\text{Median} = 45$$

$$\text{Mode} = 45$$

Q② find first, second and third quartile for the following data:

| Class | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| freq. | 14 | 25 | 27 | 24 | 15 |

Sol:

| Class | frequency | CF |
|---|---|---|
| 0-10 | 14 | 14 |
| 10-20 | 25 | 39 → $Q_1$ |
| 20-30 | 27 | 66 |
| 30-40 | 24 | 90 |
| 40-50 | 15 | 105 |
| | $\Sigma f = 105$ | |

① for first quartile

$$Q_1 = \frac{N}{4}$$

$$Q_1 = \frac{105}{4}$$

$$Q_1 = 26.25$$

So, the modal class is 10-20

$$Q_1 = l + \frac{\frac{N}{4} - F}{f} \times i$$

$$Q_1 = 10 + \frac{26.25 - 14}{25} \times 10$$

$$Q_1 = 10 + \frac{12.25}{25} \times 10$$

$$Q_1 = 10 + \frac{122.5}{25}$$

$$Q_1 = 14.9$$

② for second quartile

$$Q_2 = \frac{3N}{4} = \frac{3 \times 105}{4}$$

$$Q_2 = 78.75$$

$$Q_2 = l + \frac{\frac{3}{4}N - F}{f} \times i$$

$$Q_2 = 30 + \frac{78.75 - 66}{24} \times 10$$

$$Q_2 = 30 + \frac{12.75}{24} \times 10$$

$$Q_2 = 30 + \frac{127.5}{24}$$

$$Q_2 = 35.3$$

③ for third quartile

$$Q_3 = \frac{1}{2}(Q_2 - Q_1)$$

$$Q_3 = \frac{1}{2}(35.3 - 14.9)$$

$$Q_3 = 10.2$$

Date
Page

Q③ find Mean and standard deviation for the following data:

| Life (in hours) | 0-8 | 8-16 | 16-24 | 24-32 | 32-40 |
|---|---|---|---|---|---|
| No. of bulbs | 3 | 5 | 10 | 12 | 2 |

Sol:

| class | freq. (f) | $x$ | $fx$ | $(x-M)$ $M=30.9$ | $(x-M)^2$ | $f(x-M)^2$ |
|---|---|---|---|---|---|---|
| 0-8 | 3 | 4 | 12 | -26.9 | 723.61 | 2170.83 |
| 8-16 | 5 | 12 | 60 | -18.9 | 357.21 | 1786.05 |
| 16-24 | 10 | 20 | 200 | -10.9 | 118.81 | 1188.1 |
| 24-32 | 12 | 28 | 336 | -2.9 | 8.41 | 100.92 |
| 32-40 | 2 | 36 | 72 | 5.1 | 26.01 | 52.02 |
| | $\Sigma f=22$ | | $\Sigma fx=680$ | | | $\Sigma f(x-M)^2=5297.92$ |

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\text{Mean} = \frac{680}{22}$$

$$\text{Mean} = 30.9$$

$$\sigma = \sqrt{\frac{\Sigma f(x-M)^2}{\Sigma f}}$$

$$\sigma = \sqrt{\frac{5297.92}{22}}$$

$$\sigma = \sqrt{240.81}$$

$$\sigma = 15.5$$

Q4 Calculate Coefficient of skewness by Bowley's and Karl pearson's.

class 0-10 10-20 20-30 30-40 40-50 50-60 60-70 70-80

frequency 30 40 50 48 24 162 132 14

Sol:

| class | f | x | fx | $x^2$ | $fx^2$ | cf |
|---|---|---|---|---|---|---|
| 0-10 | 30 | 5 | 150 | 25 | 750 | 30 |
| 10-20 | 40 | 15 | 600 | 225 | 9000 | 70 |
| 20-30 | 50 | 25 | 1250 | 625 | 31250 | 120 |
| 30-40 | 48 | 35 | 1680 | 1225 | 58800 | 168 |
| 40-50 | 24 | 45 | 1080 | 2025 | 48600 | 192 |
| 50-60 | 162 | 55 | 8910 | 3025 | 490050 | 354 |
| 60-70 | 132 | 65 | 8580 | 4225 | 557700 | 486 |
| 70-80 | 14 | 75 | 1050 | 5625 | 78750 | 500 |

So, $N = 500$ $\Sigma fx = 23300$ $\Sigma fx^2 = 1274900$

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{23300}{500}$$

$\bar{x} = 46.6$

Mode:-

$f_1 = 162$, $l = 50$, $h = 10$; $f_0 = 24$, $f_2 = 132$

$$\text{mode} = l + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h$$

$$= 50 + \left(\frac{162-24}{324-24-132}\right) \times 10$$

$$= 50 + \frac{138 \times 10}{168}$$

$= 50 + 8.214$

$= 58.214$

$$\text{Standard deviation} = \sqrt{\frac{\Sigma fx^2}{N} - \left(\frac{\Sigma fx}{N}\right)^2}$$

$$= \sqrt{\frac{1274900}{500} - \frac{542890000}{250000}}$$

$$= \sqrt{2549.8 - 2171.56}$$

$$= \sqrt{378.24}$$

$= 19.45$

Now, Karl pearson Coefficient of skewness is

$$= \frac{\text{Mean} - \text{Mode}}{S.D}$$

$$= \frac{46.6 - 56.214}{19.45}$$

$$= -\frac{11.614}{19.45}$$

$$= -0.59$$

Bowley Coefficient of skewness is

$$\text{Median}(M) = l + \left(\frac{\frac{N}{2} - Cf}{f}\right) \times h$$

Where, $\frac{N}{2} = 250$, $f = 162$, $Cf = 192$, $l = 50$, $h = 10$

So,

$$M = 50 + \left(\frac{250 - 192}{162}\right) \times 10$$

$$M = 50 + \frac{580}{162}$$

$$M = 50 + 3.58$$

$$M = 53.58$$

for $Q_1$,

$\frac{N}{4} = \frac{500}{4} = 125$, $f = 48$, $Cf = 120$, $l = 30$,

$h = 10$

So, $Q_1 = l + \left(\frac{\frac{N}{4} - Cf}{f}\right) \times h$

$$= 30 + \left(\frac{125-120}{48}\right) \times 10$$

$$= 30 + \frac{50}{48}$$

$$= 30 + 1.04167$$

$$Q_1 = 31.04$$

for $Q_3$,

$\frac{3N}{4} = \frac{3 \times 500}{4} = 375$, $Cf = 354$, $f = 152$

$h = 10$, $l = 60$

Now,

$$Q_3 = l + \left(\frac{\frac{3N}{4} - Cf}{f}\right) \times h$$

$$= 60 + \left(\frac{375-354}{132}\right) \times 10$$

$$= 60 + \frac{210}{132}$$

$$= 60 + 1.59$$

$$= 61.59$$

So, Now, we can calculate bowley's coefficient of skewness

$$Sk_2 = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$SK_2 = \frac{61.59 + 31.04 - 2 \times 53.38}{61.59 - 31.04}$$

$$= - \frac{14.13}{30.55}$$

$$= -0.4625$$

Q⑤ Find Kurtosis for the data given below:

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $f$ | 15 | 38 | 55 | 82 | 60 | 40 | 10 |

$M = 2.98$

Sol:

| $x$ | $f$ | $fx$ | $(x-M)^2$ | $(x-M)^4$ | $f(x-M)^4$ |
|---|---|---|---|---|---|
| 0 | 15 | 0 | 8.8804 | 78.8615 | 1182.9225 |
| 1 | 38 | 38 | 3.9204 | 15.3695 | 584.041 |
| 2 | 55 | 110 | 0.9604 | 0.9223 | 50.7265 |
| 3 | 82 | 246 | 0.0004 | 1.6000 | 131.2 |
| 4 | 60 | 240 | 1.0404 | 1.0824 | 64.944 |
| 5 | 40 | 200 | 4.0804 | 16.6496 | 665.984 |
| 6 | 10 | 60 | 9.1204 | 83.1816 | 831.816 |
| | $\Sigma f = 300$ | $\Sigma fx = 894$ | $\Sigma (x-M)^2 = 28.0028$ | $\Sigma (x-M)^4 = 197.6669$ | $\Sigma f(x-M)^4 = 3511.634$ |

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\text{Mean} = \frac{894}{300}$$

$$\text{Mean} = 2.98$$

$$\text{fourth moment} = \frac{\Sigma f(x-M)^4}{\Sigma f}$$

$$= \frac{3511.634}{300}$$

$$= 11.705$$

second moment squared is

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

$$= \left(\frac{686.97}{300}\right)^2$$

$$= 5.243$$

$$\text{Kurtosis} = \frac{\frac{\Sigma (x-M)^4}{n}}{\left(\frac{\Sigma (x-M)^2}{n}\right)^2}$$

$$= \frac{\frac{197.6669}{300}}{\left(\frac{28.0028}{300}\right)^2} = \boxed{\frac{0.65888}{\cancel{0.00870}}}$$

$$= \frac{0.6588}{0.0872} = 7.55$$

## UNIT-III

Q1) Find Karl Pearson's coefficient of correlation between Sales and expenses of the following ten firms:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales ('000) | 50 | 50 | 55 | 60 | 65 | 65 | 65 | 60 | 60 | 50 |
| expenses ('000) | 11 | 13 | 14 | 16 | 16 | 15 | 15 | 14 | 13 | 13 |

Sol:

| Sales ($x$) | expenses ($y$) | $x^2$ | $y^2$ | $xy$ |
|---|---|---|---|---|
| 50 | 11 | 2500 | 121 | 550 |
| 50 | 13 | 2500 | 169 | 650 |
| 55 | 14 | 3025 | 196 | 770 |
| 60 | 16 | 3600 | 256 | 960 |
| 65 | 16 | 4225 | 256 | 1040 |
| 65 | 15 | 4225 | 225 | 975 |
| 65 | 15 | 4225 | 225 | 975 |
| 60 | 14 | 3600 | 196 | 840 |
| 60 | 13 | 3600 | 169 | 780 |
| 50 | 13 | 2500 | 169 | 650 |
| $\Sigma x = 580$ | $\Sigma y = 140$ | $\Sigma x^2 = 34000$ | $\Sigma y^2 = 1982$ | $\Sigma xy = 8190$ |

by Karl pearson's coefficient of correlation, we get:

$$r = \frac{\frac{\Sigma xy}{n} - \frac{\Sigma x . \Sigma y}{n}}{\sqrt{\frac{\Sigma x^2}{n} \times \frac{(\Sigma x)^2}{n^2}} \sqrt{\frac{\Sigma y^2}{n} - \frac{(\Sigma y)^2}{n^2}}}$$

$$r = \frac{\frac{8190}{10} - \frac{580 \times 140}{10}}{\sqrt{\frac{34000}{10} - \frac{336400}{1000}} \sqrt{\frac{1982}{10} - \frac{19600}{100}}}$$

$$r = \frac{\frac{8190}{10} - \frac{81200}{10}}{\sqrt{3400 - 3364} \sqrt{198.2 - 196}}$$

$$r = \frac{819 - 8120}{\sqrt{3400 - 3364} \sqrt{198.2 - 196}}$$

$$r = \frac{7}{\sqrt{36} \sqrt{2.2}}$$

$$r = 0.7865$$

Q ② The coefficient of correlation between the ages of husband's and wife's in a community was found to be 0.8 the average of husbands age was 27 years and that of wives age 22 years their standard deviation were 4 and 5 years respectively, find with the help of line of regression equations.

(a) expected age of husband's when wife age is 22 years.

(b) expected age of wife's when husband age is 33 years.

Sol: get the expected age by multiplying age with correlation coefficient.

(a) When wife's age is 22.
Since, there is 0.8 Correlation coefficient, this gives $22 \times 0.8 = 17.6$
This leaves 0.2 that is not in correlation and that gives $27 \times 0.2 = 5.4$ and an average, the husband is 5 years older.
Hence,
Husband's age $= 17.6 + 5.4 + 5 = 28$ years old.

(b) When husbands age is 33 years.
Similarly,

$$\text{wife's age} = (0.8 \times 33) + (0.2 \times 4)$$
$$= 26.8 + 0.8$$
$$= 27.2 \text{ years}$$

Q ③ find the regression equation of x on y and y on x from the following data, and estimate x when $y = 26$

| x | 10 | 12 | 13 | 17 | 18 | 20 | 24 | 30 |
|---|---|---|---|---|---|---|---|---|
| y | 5 | 6 | 7 | 9 | 13 | 15 | 20 | 21 |

Sol:

| $x$ | $y$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|
| 10 | 5 | 50 | 100 | 25 |
| 12 | 6 | 72 | 144 | 36 |
| 13 | 7 | 91 | 169 | 49 |
| 17 | 9 | 153 | 289 | 81 |
| 18 | 13 | 234 | 324 | 169 |
| 20 | 15 | 300 | 400 | 225 |
| 24 | 20 | 480 | 576 | 400 |
| 30 | 21 | 630 | 900 | 441 |
| $\Sigma x =$ | $\Sigma y =$ | $\Sigma xy =$ | $\Sigma x^2 =$ | $\Sigma y^2 =$ |
| 144 | 96 | 2010 | 2902 | 1426 |

Coefficient of Correlation, $r = \dfrac{\dfrac{\Sigma xy}{n} - \dfrac{\Sigma x}{n}\dfrac{\Sigma y}{n}}{\sqrt{\dfrac{\Sigma x^2}{n} - \left(\dfrac{\Sigma x}{n}\right)^2}\sqrt{\dfrac{\Sigma y^2}{n} - \left(\dfrac{\Sigma y}{n}\right)^2}}$

$$r = \frac{\dfrac{2010}{8} - \dfrac{144 \times 96}{8 \quad 8}}{\sqrt{\dfrac{2902}{8} - \dfrac{20736}{64}}\sqrt{\dfrac{1426}{8} - \dfrac{9016}{64}}}$$

$$r = \frac{251.25 - 216}{6.224 \times 5.852}$$

$$r = \frac{35.25}{36.42}$$

$$r = 0.96$$

Now,
Coefficient of regression of x and y,

$$R_{xy} = \frac{r\,\sigma x}{\sigma y}$$

$$= 0.96 \times \frac{6.224}{5.852}$$

$$= \frac{5.97504}{5.852}$$

$$= 1.021$$

Coefficient of regression of y on x,

$$R_{yx} = \frac{r\,\sigma y}{\sigma x}$$

$$= 0.96 \times \frac{5.852}{6.224}$$

$$= 0.902$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{144}{8} = 18$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{96}{8} = 12$$

Now,
the regression equation of x and y,

$(x-10) = 1.021\,(y-12)$

$x - (1.021)y = 18 - 12.252$

$x - (1.021)y = 5.748$

When

$y = 26,$

then,

$$x = 32.294$$

regression equation of y on x

$$(y-\overline{y}) = R_{yx}\,(x-\overline{x})$$

$$(y-12) = 0.902\,(x-18)$$

Q④ find rank correlation coefficient for the following data:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| x | 1 | 2 | 3 | 4 | 2 | 6 | 7 | 8 | 9 |
| y | 2 | 6 | 7 | 8 | 10 | 11 | 11 | 10 | 9 |

Sol: Here, the rank are repeated.
So,

$\rho$ = Rank correlation coefficient

$$= 1 - \frac{6(\Sigma d^2 + F)}{n(n^2-1)}$$

$$F = f_2 + f_{11} + f_{10} = \frac{1}{12}\left[2(4-1) + 2(4-1) + 2(4-1)\right]$$

$$F = 1.5$$

| $x$ | $y$ | $R_x$ | $R_y$ | $d = R_x - R_y$ | $d^2$ |
|---|---|---|---|---|---|
| 1 | 2 | 9 | 9 | 0 | 0 |
| 2 | 6 | 7.5 | 8 | -0.5 | 0.25 |
| 3 | 7 | 6 | 7 | -1 | 1 |
| 4 | 8 | 5 | 6 | -1 | 1 |
| 2 | 10 | 7.5 | 3.5 | 4 | 16 |
| 6 | 11 | 4 | 1.5 | 2.5 | 6.25 |
| 7 | 11 | 3 | 1.5 | 1.5 | 2.25 |
| 8 | 10 | 2 | 3.5 | -1.5 | 2.25 |
| 9 | 9 | 1 | 5 | -4 | 16 |
| | | | | | $\Sigma d^2 =$ 45 |

$$\rho = 1 - \frac{6(45+1.5)}{9(81-1)}$$

$$\rho = 1 - \frac{6 \times 46.5}{9 \times 80}$$

$$\rho = 1 - \frac{279}{720}$$

$$\rho = 1 - 0.3875$$

$$\rho = 0.6125$$

# UNIT- IV

Q① In a college 25% students in mathematics, 15% in physics and 10% students in mathematics and physics both are failed. A student is selected at random

① If he failed in physics then find the chance of his failure in mathematics.

② If he failed in mathematics then find the chance of his failure in physics.

③ find the chance of his failure in mathematics or physics.

Sol: Let us denote the event of failing of student in mathematics is M and the event of failing students in physics is PH.

By given problem,

$$P(M) = 25\% \quad P(PH) = 15\%$$

$$P(M \cap PH) = 10\%$$

① We need to find $P(M/PH)$

$$P(M/PH) = \frac{P(M \cap PH)}{P(PH)}$$

$$= \frac{10\%}{15\%} = \frac{2}{3} = 66.67\%$$

② We need to find $P(PH/M)$
So

$$P(PH/M) = \frac{P(PH \cap M)}{P(M)}$$

$$= \frac{10\%}{15\%}$$

$$= \frac{2}{5} = 40\%$$

③ We need to find $(M \cup PH)$
So,

$$P(M \cup PH) = P(M) + P(PH) - P(M \cap PH)$$

$$= 25\% + 15\% - 10\%$$

$$= 30\%$$

Q② Explain the following term:

① Random Variable
② Mutually exclusive event
③ Conditional probability

Ans: ① Random Variable

It is a mathematical representation of a variable whose possible possible values are determined by chance events. It's a function that assigns a numerical value to each outcome of a random experiment.

Characteristics - ① Assigns numerical values to outcomes.
② outcome values are uncertain.
③ follows probability distribution.

for ex: rolling a dice, measuring height, stock prices.

② Mutually exclusive events

It is the events which cannot occur simultaneously. If one event occurs, the other cannot.

Characteristics - ① non-overlapping events.
② probability of both events occurring together is zero.

for ex: fipping a coin, rolling a die, drawing a card.

③ Conditional probability

It measures the probability of an event occurring given that another event has occurred.

Date
Page

formula: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

for ex: probability of rain given cloudy skies, probability of a person having a disease given symptoms.

Q③ from a pack of 52 cards six cards are drawn at random, find the probability of the following events:

① 3 are red and 3 are black cards.

② 3 are Kings and 3 are queens.

sol: ① given that,

Total number of cards = 52 cards

26 black, 26 red cards.

Now,

Total no of ways of selecting 6 cards out of 52 cards is $^{52}C_6$

$$= \frac{52 \times 51 \times 50 \times 49 \times 48 \times 47}{6 \times 5 \times 4 \times 3 \times 2 \times 1}$$

$$= 20358520$$

Total ways of selecting 3 cards out of 26 black card is $^{26}C_3 = 2600$

Total ways of selecting 26 cards out of 26 red card is $^{26}C_3 = 2600$

So,

$$\text{probability} = \frac{2600 \times 2600}{2035820}$$

$$\text{probability} = 0.33$$

② Total ways of selecting 6 cards out of 52 cards is $^{52}C_6 = 20358520$

Total ways of selecting 3 King of 4 cards is $^4C_3 = 4$.

Total ways of selecting 3 queen of 4 cards is $^4C_3 = 4$.

So,

$$\text{probability} = \frac{4 \times 4}{20358520}$$

$$= 0.00000785$$

Q④ One bag contains 4 white, 6 red and 15 black balls. And second bag contains 11 white, 5 red and 9 black balls. One ball from each bag is drawn. Find the probabilities of the following events:

① both balls are red
② both balls are white
③ both balls are black
④ both balls are of same color.

Sol; $1^{st}$ bag : $4+6+15=25$ balls
$2^{nd}$ bag : $11+5+9=25$ balls

both bag contains 25 balls.

① probability of getting red from $1^{st}$ bag $= \frac{6}{25}$

probability of getting red from $2^{nd}$ bag $= \frac{5}{25}$

probability of getting red from both bags $= \frac{6}{25} \times \frac{5}{25}$

$= \frac{30}{625}$

② Similarly,

probability of getting white ball in both bags $= \frac{4}{25} \times \frac{11}{25}$

$= \frac{44}{625}$

③ Similarly,
probability of getting black ball in both bags
$= \frac{15}{25} \times \frac{9}{25} = \frac{135}{625}$

(4) Now, probability of getting both balls of same colour $= \frac{30}{625} + \frac{44}{625} + \frac{135}{625}$

$$= \frac{209}{625}$$

# UNIT-V

Q ①

① Describe Mean and variance for binomial distribution.

② Six dice are thrown 729 times. How many times do you expect at least three dice to show a five or six.

Ans ① The Binomial distribution is a discrete probability distribution that models the number of successes in a fixed number of independent trails, each with a constant probability of success.

mean ($\mu$) - the mean of the binomial distribution is:

$$\mu = np$$

where,

$n$ = no. of trails

$p$ = probability of success in each trail.

variance ($\sigma^2$) - the variance of the binomial distribution is:

$$\sigma^2 = np(1-p)$$

② Total outcomes of dice = $\{1,2,3,4,5,6\} = 6$

favourable outcomes = 2 (5 or 6)

$$\text{probability } (P) = \frac{2}{6} = \frac{1}{3}$$

$$q = 1 - p$$

$$q = 1 - \frac{1}{3}$$

$$q = \frac{2}{3}$$

$n = 6,\ r = 3$

Now,

using complement theorem, we get:

$$P(x = r) = 1 - \left[P(x=0) + P(x=1) + P(x=2)\right]$$

$$= 1 - \left[{}^6C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 + {}^6C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{5}\right)^5 + {}^6C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^4\right]$$

$$= 1 - \left[\frac{4}{9} + \frac{12}{9} + \frac{15}{9}\right] \left(\frac{2}{3}\right)^4$$

$$= 1 - \frac{16}{81} \times \frac{31}{9}$$

$$= 1 - \frac{498}{729}$$

$$= \frac{233}{729}$$

$\therefore$ in 729 trails, the expression is

$$= \cancel{729} \times \frac{233}{\cancel{729}}$$

$$= 233$$

Q② Fit poisson's distribution to the following and evaluate theoretical frequencies ($e^{-0.5} = 0.61$):

| | | | | | |
|---|---|---|---|---|---|
| Deaths: | 0 | 1 | 2 | 3 | 4 |
| frequency: | 122 | 60 | 15 | 2 | 1 |

Sol:

$$\lambda = \frac{\Sigma f(x)}{N}$$

$$\lambda = \left[ \frac{0(122)+1(60)+2(15)+3(2)+4(1)}{122+60+15+2+1} \right]$$

$$\lambda = 0.5$$

$x$ = no. of deaths = $(0, 1, 2, 3, 4)$

Now,

using poisson's probability formula

$$P(x=\lambda) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

Now, lets calculate theoretical frequencies:

$$P(x=0) = \left(\frac{e^{-0.5} \times (0.5)^0}{0!}\right)$$

$$= 0.61$$

$$P(x=1) = \left(\frac{e^{-0.5} \times 0.5}{\underline{1}}\right)$$

$$= 0.3$$

$$P(x=2) = \left(\frac{e^{-0.5} \times (0.5)^2}{2!}\right)$$

$$= 0.15$$

$$P(x=3) = \left(\frac{e^{-0.5} \times (0.5)^3}{3!}\right)$$

$$= 0.08$$

$$P(x=4) = \left(\frac{e^{-0.5} \times (0.5)^4}{4!}\right)$$

$$= 0.04$$

Q③ In a test of on 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and Standard deviation of 60 hrs. estimate the number of bulbs likely to burn for

① more than 2150 hrs.

② less than 1950 hrs.

Sol: let X be the number of hrs which bulb are used.

given: $X \sim N(2040, 60^2)$

① $P(X > 2150)$
The total area of right of $Z=0$ is 0.5.
The area between $Z=0$ and 1.833 is 0.4664.
So, $P(Z > 1.833) = 0.5 - 0.4664$
$= 0.0336$
Now, the no. of bulbs likely to burn for marks than 2150 hrs is
$= 2000 \times 0.0336$
$= 67.2$

② $P(\cancel{x}\ 21950)$
$= P(X - \cancel{\mu/\sigma})$
$= P(Z < 1.5)$

Area b/w $Z=-1.5$, $Z=0$ is same as area b/w $Z=0$, $Z=1.5$, so
area b/w $Z=0$, $Z=1.5$ is 0.4332
total $P(Z < 1.5) = 0.5 - 0.4332$
$= 0.0668$

Hence, no. of bulbs likely to burn for less than 1950 hrs is
$= 2000 \times 0.0668$
$= 133.6$

Date
Page

Q⑤ explain Beta and gamma distribution with their properties?

Ans⑤ Beta distribution

A continuous probability distribution that defined on interval [0,1]. It is often use to model random variables that represent probability.

$$F(x; \alpha, B) = \frac{(x)^{\alpha-1} \cdot (1-x)^{B-1}}{B(\alpha, B)}$$

properties — ① It is flexible and can take on a variety of shapes depending on the values of α and B.

② when $\alpha = B = 1$, the beta distribution is uniform over interval [0,1].

③ The mean of it, is $\frac{\alpha}{(\alpha+B)}$ and variance is $\left(\frac{\alpha \cdot B}{(\alpha+B)^2 (\alpha+B+1)}\right)$

④ It is commonly used in Bayesian statistics, modeling proportions and its application such as A/B testing.

# Gamma function

It is a continuous probability distribution that is defined for positive values. It is often used to model the time until an event occurs, such as time waiting between the events.

$$F(x;\alpha,\beta) = \left[ \frac{(1) * \beta^{\alpha}}{\Gamma(\alpha)} \left( x^{\alpha-1} \times e^{-x/\beta} \right) \right]$$

properties – ① The shape of it is determined by parameters $\alpha$ and $\beta$.

② The mean of gamma distribution is $\alpha * \beta$ and variance of gamma distribution is $\alpha * \beta^2$.

③ It is widely used in reliability analysis, queuing theory, survival analysis.