



Date: 27TH MARCH, 2021

**VIRTUAL COACHING CLASSES
ORGANISED BY BOS (ACADEMIC), ICAI**

**FOUNDATION LEVEL
PAPER 3: BUSINESS MATHEMATICS AND LOGICAL
REASONING & STATISTICS**

Faculty: CA Rashmi Lonikar, M.Sc., FCA, DISA



UNIVARIATE & BIVARIATE DISTRIBUTION

- **Univariate distribution** is the simplest distribution.
- It takes data and summarises it. Then it tries to find patterns.
- Eg Investment needed to yield a particular level of profit.
- WE use mean, median, mode, quartiles, dispersion , range , standard deviation for it.
- Can be easily represented as frequency charts, bar charts, histograms, pie charts etc
- **Bivariate data** is when data is collected about two variables



MARGINAL DISTRIBUTION

Marginal distributions are the totals for the probabilities. They are found in the **margins**.

$Y \backslash X$	x_1	x_2	x_3	x_4	$P_Y(Y) \downarrow$
y_1	4/32	2/32	1/32	1/32	8/32
y_2	3/32	6/32	3/32	3/32	15/32
y_3	9/32	0	0	0	9/32
$P_X(X) \rightarrow$	16/32	8/32	4/32	4/32	32/32

Joint and marginal distributions of a pair of discrete random variables, X and Y , having nonzero mutual information $I(X; Y)$. The values of the joint distribution are in the 3×4 rectangle; the values of the marginal distributions are along the right and bottom margins.



CONDITIONAL DISTRIBUTION

A conditional distribution is a probability distribution for a **sub-population**.

it shows the probability that a randomly selected item in a sub-population has a characteristic you're interested in

$i \setminus j$	1	2	3	4	5	6	$p_X(i)$
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
$p_Y(j)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	

CORRELATION ANALYSIS

Let us consider the example of investment and profit, correlation helps us find

- A) whether there is a relation between the two
- B) The extent of relation between the two
- Is correlation – **CO + RELATION**
- **CAUSE- EFFECT RELATIONSHIP**

The measures of are

- Product moment correlation
- Rank correlation coefficient
- Coefficient of concurrent deviations

Correlation analysis is a **statistical method** used to evaluate the **strength of relationship** between two quantitative variable



- .

It is represented by ' r '

It ranges between -1 and +1

It quantifies the direction and strength of the linear association between the two variables.

Eg $r = 0.9$ - a strong, positive association

- $r = -0.2$ suggest a weak, negative association.
- $r=0$ no linear association between two continuous variables.
- GRAPHS



PROPERTIES

- 1. R IS UNIT-FREE

2. Change of origin and scale. If a, c are origins of x and y and b and d are scales such that

$U = x-a/b$ $v = y-c/d$. If b and d have opposite sign, r_{xy} and r_{uv} will **have different signs and** magnitude are **equal**.



FORMULAE

PEARSON'S COEFFICIENT

Complete formula is

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

In terms of covariance and Standard deviation- by dividing above formula by n^2

Covariance and Std deviations are

$$r = r_{xy} = \frac{\text{Cov}(x, y)}{S_x \times S_y}$$

OTHER FORMULAE

$$\text{cov}(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y}$$

$$S_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

$$S_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum y_i^2}{n} - \bar{y}^2}$$

Question Time! –

Remember we must know Σx , Σy , Σxy , Σy^2 , Σx^2
18.3,

- Q1.
 - $\Sigma x = 247$
 - $\Sigma y = 486$
 - $\Sigma xy = 20,485$
 - $\Sigma x^2 = 11,409$
 - $\Sigma y^2 = 40,022$
 - n is the sample size, in our case = 6

• Ans : $r = 0.5298$

• Q

x	2	3	5	5	6	8
y	9	8	8	6	5	3

• Ans $r = -0.93$

• Q .

x	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
y	11	12	13	20	19	20	25	30

Ans $r = 0.96$



Q $r =$, $n = 20$, AM and SD of x and y are 12 and 15 and 3 and 4 respectively. The pair(20,15) was wrongly taken as (15,20). Find correct r .

(hint: First find cov, then sum xy , corrected sum x and sum y , corrected sum $sr x$ sum $sr y$)

Answer $r = 0.31$

Q1 PG. 18.47 If two variables x and y covariance, variance of x , variance of y are 40, 16,256 respectively, find correlation coefficient. Ans : 0.625



SPEARMAN'S RANK CORRELATION COEFFICIENT

For two **qualitative characteristics**

Finds level of agreement

Simpler, used to get first hand impression

r lies between -1 to +1

$d_i = x_i - y_i$ represents difference in ranks for the i th individual

$$r_R = 1 - \frac{6 \left[\sum_i d_i^2 + \sum_j \frac{(t_j^3 - t_j)}{12} \right]}{n(n^2 - 1)}$$

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$



- Always remember: **The numbering/ranking will be from 1 to n**
- **If a number is repeated twice $t^3-t^1 = 2^3-2 = 6$, if number is repeated thrice $t^3-t = 3^3-3 = 24$.**
- Calculate Rank Correlation coefficient

Marks in Eco	80	56	50	48	50	62	60
Marks in Stats	90	75	75	65	65	50	65

- Ans: 0.15

Q: The coefficient of rank correlation is:

Rank in botany: 1 2 3 4 5

Rank in Chemistry : 2 3 1 5 4

Ans: 0.60

Q: The coefficient of rank correlation of marks obtained by 10 students in English and Economics was 0.50. It was found that one difference in ranks was wrongly taken as 3 instead of 7. The correct coefficient of rank correlation is

Ans: 0.26

1. The coefficient of correlation r between x and y when $\text{cov}(x,y) = -16.5$, $\text{var}(x) = 2.89$ $\text{var}(y) = 100$

Ans: -0.97

2. The covariance between 2 variables x and y is 8.4 and their variances are 25 and 36 respectively. Calculate Karl Pearson's coefficient of correlation between them is

Ans: 0.28

3. If $r = 0.28$, $\text{cov}(x,y) = 7.6$, $\text{var}(x) = 9$. Then σ_y is

Ans: 9.04

COEFFICIENT OF CONCURRENT DEVIATIONS



Simple and casual method

Add a + sign if the number is more than previous number in the column

Add a -ive sign if the number is less than previous number in the column

Do this for both x and y column. Multiply the columns

If sign in both x and y are same, it is said to be concurrent

M = number of pairs -1

C = number of positives(concurrent) in the multiplication columns

$$r_c = \pm \sqrt{\pm \frac{(2c - m)}{m}}$$

Q 18.15 on pg. 18.24, Find concurrent deviation of

Example 18.15: Find the coefficient of concurrent deviations from the following data.

Year :	1990	1991	1992	1993	1994	1995	1996	1997
Price :	25	28	30	23	35	38	39	42
Demand :	35	34	35	30	29	28	26	23

Ans : -0.65

Q: For 10 pairs of observations, number of concurrent deviations was found to be 4. What is the value of the coefficient of concurrent deviations ?

Ans: - 1/3

18.6 on pg. 18.15

~~Coefficient of correlation between x and y for 20 items is 0.4. The AM and SD of x and y are known to be 12 and 15 and 3 and 4 respectively. Later on it was found that a pair (20,15) was wrongly taken as (15,20). Find the correct value of correlation coefficient.~~

Ans: 0.31

PROPERTIES

Q: The Coefficient of correlation between x and y is 0.60. u and v are two variables defined as $u = (x-3)/2$ and $v = (y-2)/3$, then the coefficient of correlation between u and v is

Ans: 0.60

DETERMINATION/ NON-DETERMINATION

Coefficient of determination measures relationship between 2 variables indicating the amount of variation of one variable accounted for by the other variable

$$R^2 = \text{Explained variance} / \text{Total variance}$$

$$\text{Coefficient of non-determination} = (1 - R^2)$$

Q: If $r = 0.60$ then the coefficient of non-determination is

Ans: 0.64

Q: If the coefficient of correlation between x and y variables is -0.90 then what will be the coefficient of determination

Ans: 0.81

PROBABLE ERROR

Probable error defines the half-range of an interval about a central point for the distribution, such that half of the values from the distribution will lie within the interval and half outside.

IMPORTANT:

- (1) If $r < PE$, there is no evidence of correlation.
- (2) If $r \geq 6PE$, then presence of correlation is certain
- (3) PE is never negative.

Probable Error is a method of obtaining correlation coefficient of population. It is defined as:

$$P.E = 0.674 \times \frac{1 - r^2}{\sqrt{N}}$$

Where r = Correlation coefficient fromn pairs of sample observations

$$PE = \frac{2}{3} SE$$

When SE = Standard Error of correlation coefficient

$$S.E = \times \frac{1 - r^2}{\sqrt{N}}$$

The limit of the correlation coefficient is given by $p = r \pm P.E$

Where p = Correlation coefficient of the population

Q: 18.21/pg. 18.38

Compute probable error assuming coefficient of correlation to be 0.80 and a sample of 25 pairs of items.

Q: find probable error if $r = 2/\sqrt{10}$ and $n = 36$

Ans: 0.067

Q: 18.22

If $r = 0.7$ and $n = 64$, find PE and determine limits for the population coefficient of correlation

Ans: $(0.7 + 0.043) (0.7 - 0.043) = (0.743, 0.657)$

THEORY

Q: In rank correlation the association need not be linear

(a) True (b) False (c) Partly true (d) Partly false

Q: If coefficient of correlation between a and y is 0.46. Find coefficient of correlation between x and $y/2$

Ans: 0.46

Q: If ranks of 2 characteristics by 2 judges are in reverse order then the value of Spearman's rank correlation coefficient is

Ans: -1.0

Q: Correlation coefficient between x and y will be negative when:

(a) x and y are decreasing (b) x is increasing , y is decreasing

(c) x and y are increasing (d) NOT

Ans : b

Q: If r is simple correlation coefficient, then r^2 is called as

Ans: Coefficient of determination

Q: If the sum of products of x and y series from their mean is zero, then the correlation will be

Ans: 0

Q: Two variables x and y are relate as $4x + 3y = 7$, then correlation between x and y is

(a) perfectly positive (b) Perfectly negative (c) Zero (d) NOT

Ans: Perfectly negative

Q: The speed of an automobile and the distance required to stop the car after applying brakes , the correlation is

(a) positive (b) Negative (c) zero (d) NOT



THANK YOU