

Chapter

1

Correlation Analysis

In the previous class the students familiarises the Bivariate Data which consists of two variables with certain relationship. They also studied to represent the bivariate data using scatter plots. This chapter helps the students to identify the meaning of correlation, different types of correlation, methods of studying correlation, rank correlation coefficient. Two or more variables may have some relationships either linear or non linear (curvy linear). In this chapter we discuss the linear relationship between two variables only.

Significant Learning Outcomes:

After the completion of this chapter, the learner:

- Identifies the meaning of correlation.
- Recognises different types of correlation.
- Identifies the methods of studying correlation.
- Identifies rank correlation coefficient.
- Familiarises and uses the formulae for correlation coefficient.
- Applies the idea of correlation coefficient in practical situations.

Unit Frame

Concept/process skills	Teaching-learning process/ activities with assessments	Learning outcomes
Meaning of correlation. <ul style="list-style-type: none"> • Observing. • Communicating. • Identifying 	Group activity.	Identifies the meaning of correlation.
Types of correlation. <ul style="list-style-type: none"> - Positive correlation. - Negative correlation. - Zero correlation. Perfect correlation <ul style="list-style-type: none"> • categorising • Illustrating • Identifying • Discriminating 	Debate, Group discussion	Recognises different types of correlation.
Methods of Studying Correlation <ul style="list-style-type: none"> - Scatter diagram - Coefficients of Correlation - Karl Pearson's coefficient of correlation - Interpretation of Karl Pearson's coefficient of correlation - Properties of coefficient of correlation <ul style="list-style-type: none"> • Finding • carrying out • using 	Group discussion, Problem solving	Identifies the methods of studying correlation.

Spearman's rank correlation coefficient <ul style="list-style-type: none">• Finding• carrying out• using	Group discussion, Problem solving	Identifies rank correlation coefficient.
Solving the problems <ul style="list-style-type: none">• distinguishing• selecting• developing• using	Group activity, Problem solving	Familiarises and uses the formulae for correlation coefficient.
Solving the problems <ul style="list-style-type: none">• distinguishing• selecting• developing• using	Group activity, Problem solving	Applies the idea of correlation coefficient in practical situations.

Values and attitudes

The learner:

- Identifies the relationships between the objects in nature.
- Develops the ability to observe the correlation among variables.
- Develops the habit of proper observation and ranking.

1.1 Concept: Meaning of correlation

Suggested activity: group activity

Students asked to draw the scatter plot for the cases given in textbook and make a discussion on the relation between each pair of variables. They identify the variables which are correlated and not correlated. Teacher helps the students to identify the meaning of correlation between two variables using more illustrations. Students list correlated variables and not correlated variables.

Correlation is the study of the degree of relationship between two variables in a bivariate distribution.

CE Activity

List any two pair of correlated variables / not correlated variables.

1.2 Concept: Types of correlation.

Suggested activity: debate, group discussion

Students directed to collect some bivariate data and argue on they are correlated or not correlated. By drawingscatter diagram they can classify the variables into correlated or not correlated. The students recognises the different types of correlation (positive, negative, zero) with the help of illustrations by the teacher. The teacher also gives the examples for variables with perfect correlation (positive and negative) and the students recognise the perfect positive and perfect negative correlation.

If the two variables are moving together in the same direction then the correlation is called positive correlation.

If the two variables are moving in the opposite direction then the correlation is called negative correlation.

If there is no association between the two variables, we say that there is no correlation or zero correlation.

If the change in the value of one variable is proportional to the changes in the value of the other variable, then the correlation is said to be perfect. If the proportion is direct then the correlation is perfect positive and if it is inverse proportion, then the correlation is perfect negative.

CE Activity

Collect two cases for pair of correlated variables (positive, negative, perfect positive, perfect negative).

1.3 Concept: Scatter diagram

Suggested activity: group work

Students asked to collect the data on following.

- CE score and TE score of 10 students in different subjects.
- TE score in statistics and Economics of 10 students.
- Height score in English of 10 students.
- Speed of a vehicle and time taken to reach a certain place.

The students asked to draw scatter plot. They recognise the different types of correlation using scatter plots.

1.4 Concept: Coefficient of Correlation

Concept: Karl Pearson's coefficient of correlation, Interpretation & its properties

Suggested activity: group discussion, problem solving

Teacher leads a discussion on merits and demerits of scatter diagram. Students recognise the need of a tool which gives the exact measurement of correlation between two variables. Teacher illustrates the Karl Pearson's coefficient of correlation formula. Students use Karl Pearson's coefficient of correlation formula to solve the problems.

1.5 Concept: Spearman's rank correlation coefficient

Suggested activity: group discussion, problem solving

Teacher illustrates that a bivariate distribution can be both numerical, both categorical or one numerical and one categorical. Students recognise the need of another tool to find the coefficient of correlation when one or both variables

are categorical. Teacher illustrates the Spearman's rank correlation coefficient. Students use the formula of Spearman's rank correlation coefficient when the ranks are given. Teacher gives focus on method of giving ranks when values are repeated.

Activities for CE items

- Debate
- Group discussion
- Collection

Port folio assessment

- Reports
- Data collected

PE activity

1. The following scores represent a nurse's assessment (X) and physician's assessment (Y) of the condition of 10 patients at time of admission to a trauma centre

X	18	13	18	15	10	12	8	4	7	3
Y	23	20	18	16	14	11	10	7	6	4

- (a) Construct a scatter diagram for the above data
 - (b) Compute variance of Nurses assessment
 - (c) Compute the correlation between two assessments using CORREL function.
2. Height is frequently named as good predictor variable for weight among people of the same age & gender. The following are the heights and weights of 14 males between the age of 19 and 26 years who were participated in a study.

Weight(Kgs)	83	99	63	71	65	79	70	69	56
	66	82	59	64	78				
Height(cms)	185	180	173	168	175	183	184	174	164
	169	205	161	177	174				

Compute the correlation coefficient between weight and height using data analysis.

TE items

1. Classify the following into positive, negative and no correlation.
 - a. The ages of husbands and wives.
 - b. Shoe size and intelligence.
 - c. Years of education and income.
 - d. Insurance companies' profits and the number of claims they have today.
 - e. Amount of rainfall and yield of crop.
 - f. Distance and intensity of light.
2. Indicate the type of correlation (positive, negative, zero) likely to exist in the following pairs of variables.
 - a. Height and Weight
 - b. Price and Demand
 - c. Intelligence and beauty
 - d. Distance from the school and age of father
3. The scores obtained by 10 students in Economic and statistics are given below.
 - a. Draw the scatter diagram
 - b. Interpret the diagram

Score in Economics	:	59	65	45	52	60	62	70	55	45
Score in Statistics	:	75	70	55	65	60	69	80	65	50

4. The following table gives the number of blind per lakh of population in different ages

Age in years	:	15	16	17	18	19	20
No of blind per lakh	:	75	60	50	50	45	40

- a) Find the correlation coefficient between age and blindness
 - b) What is your opinion about the correlation?
5. The number of rooms and number of lamps in 7 houses are given below. Draw the scatter diagram and write your conclusion on the type and degree correlation exists.
(2, 1), (4, 5), (5, 8), (7, 14), (9, 20), (10, 23), (11, 28)

STATISTICS

6. Price and demand of different brands of vegetable oil are given below:
Calculate the correlation coefficient

Price in Rs.	42	38	42	45	42	44	40	46
Demand in kg.	26	40	29	27	30	27	35	25

7. The scores given by two judges in a Mohiniyattam competition for 8 participants are as follows:

Judge I	70	65	71	62	58	69	74	64
Judge II	91	76	65	83	90	64	55	48

- a) rank the given data
b) Find the rank correlation coefficient
8. Find spearman rank correlation coefficient between x and y .

x	12	13	15	16	18	20	24	26	28	29
y	25	24	26	28	29	30	32	23	56	44

Chapter

2

Regression Analysis

In the previous chapter we have discussed about scatter diagram and correlation. In scatter diagram we can fit the equation of a straight line called regression line or regression is a functional form of relationship between the variables. It is used for scientific prediction

Significant Learning outcomes

After completion of this chapter, the learner:

- Identifies the meaning and concept of regression.
- Explains estimation of dependent variable from the known values of independent variable.
- Identifies regression lines and their point of intersection.
- Understands properties of regression coefficients.
- Compares correlation and regression

Unit Frame

Concept/process skills	Teaching-learning process/ activities with assessments	Learning outcomes
Regression <ul style="list-style-type: none"> • Observing • Charting • Interpreting 	Group activity based on suitable examples	Identifies the meaning and concept of regression
Estimation <ul style="list-style-type: none"> • Understanding • Problem solving • Interpreting 	Group discussion on previous example ,then illustrate formulae and solve problems	Explains estimation of dependent variable from the known values of independent variable.
Identify regression lines and their point of intersection <ul style="list-style-type: none"> • Illustrating • Observing • Comparing • Problem solving • Interpreting 	Group discussion and comparing examples with general form and solve problems	Understands properties of regression coefficients. Identifies regression lines and their point of intersection.
Compares correlation and regression <ul style="list-style-type: none"> • Observing • Understanding • Classifying • Discriminating 	Group discussion based on the findings of the previous activities	Comparing correlation and regression

Values and attitudes

The learner

- Develops the ability to scientific planning through proper observation
- Develops the proper understanding of the relationship between the variables
- Create new ideas and designs
- Creates an investigative approach to a situation.

2.1 concept :Regression

Suggested activity: group work, group discussion, charting

Teacher starts the discussion related to the examples we have already discussed in the previous chapter. Teacher divides the whole students in to 5 different groups

Income	200	500	700	850	1000	1250	1500
Expenditure	150	400	650	700	800	1000	1200

Teacher directs the students to draw a scatter diagram for the above data and then draw a straight line which is most close to all points in the diagram.

Let each group present their findings but they noticed that all the answers are different. Why this happened? Discussion continues and teacher concludes the discussion to the following concepts.

1. A line of best fit can be drawn to any bivariate data by the method of least squares
2. This line of best fit is called regression line
3. This line is a first degree equation in X and Y

CE Activity

Assign the groups to collect a bivariate data and then draw a least square line for the collected data

2.2 Concept: Estimation of one value from other value

Suggested activity: Group discussion, Illustration and Problem Solving

Teacher starts the discussion by asking a question. For the above discussed example can you tell the expenditure if the income is 2000? Students responds

that we can able to fit regression equation containing X and Y. From this line of regression, if we know X we can predict Y and vice versa. Teacher leads the discussion to the following points

- If we know the equation of regression line we can predict one value of the variable from the other value.
- The variable whose value is to be predicted is called dependent variable and the variable whose value is used for prediction is called independent variable.
- Since X and Y is not reversible we have two regression lines. Regression line of X on Y is used for predicting X from known Y and regression line of Y on X is used for predicting Y for given X.
- Illustrating the formulae and more problems can be worked out.

CE Activities

Assign the students to find which is independent and which is dependent variable for the data collected above and estimate dependent variable from the known value of independent variable.

2.3 Concept: Identify regression lines and their point of intersection

Suggested Activity: Observing, Comparison, Problem Solving

Teacher give some examples of regression lines

Eg: $y-2 = 1.2(x-6), b_{yx} = 1.2$

$$x-1 = 0.7(y-5), b_{xy} = 0.7$$

ask the students to compare this with the general form of regression equations. students respond that

In regression equation of Y on X, b_{yx} is the coefficient of X.

In regression equation of X on Y, b_{xy} is the coefficient of Y.

Again the discussion continued concentrating with regression coefficients. Teacher give an example of two regression lines . Out of the two lines of regression given by $X+ 2Y-5=0$ and $2X+3Y-8=0$ which one is the regression line of X on Y and which one is regression line of Y on X? Discussion continues and students respond that let us suppose one equation as regression line of Y on

X and other as regression line of X on Y we can able to identify regression lines. Teacher illustrate this point with the help of the above example.

Consolidation points

- The signs of both regression coefficients are same.
- the product of both regression coefficients should be below one.
- the geometric mean of the regression coefficients is Coefficient of correlation.
- $b_{yx} = r \times \frac{\dagger_y}{\dagger_x}$ and $b_{xy} = r \times \frac{\dagger_x}{\dagger_y}$ where \dagger_y is standard deviation of Y and \dagger_x is standard deviation of X.
- $b_{yx} \neq b_{xy}$ in general.

Teacher starts the discussion by asking a question, can you find the point of intersection of two regression lines? Students respond that two regression lines pass through the point of intersection and it can be get by solving equation of two regression lines. More problems can be worked out

Consolidation point

- The x co-ordinate of the point of intersection is the mean value X and y co-ordinate of the point of intersection is the mean value of Y

2.4 Concept: comparing correlation and regression

Suggested Activity: Group Discussion

Teacher divides the students in to 5 groups and asks them a group work on comparing correlation with regression. Let each group present their findings. Then teacher consolidate the important points.

CE Activities

- seminar
- Debate
- Presentation

Portfolio Assessment

- Collected data
- Reports

PE Activities

1. The following data relate to marketing expenditure (in Rs.Lakh) and the corresponding Sales (in Rs.Crores)

Marketing Expenditure(X) :	10	12	15	20	23
Sales (Y) :	14	17	23	21	35

- (a) Construct the regression equation Y on X
 (b) Estimate the market expenditure to obtain sales target of Rs.40 Crores
2. Height is frequently named as good predictor variable for weight among people of the same age & gender. The following are the heights and weights of 9 males between the age of 19 and 26 years who were participated in a study.

Weight (Kgs)	83	99	63	71	65	79	70	69	56
Height (Cms)	185	180	173	168	175	183	184	174	164

- (a) Obtain the regression equation of Weight on Height
 (b) Predict weight when Height is 170cms.

TE Items

1. If $b_{xy} > 1$, then b_{yx} is
- (a) less than 1 (b) greater than 1 (c) equal to one (d) any value
2. Regression equation is also named as
- (a) predicting equation
 (b) estimating equation
 (c) equation of average relationship (d) all of the above
3. The two regression lines intersect at the point
- (a) (x,y) (b) (\bar{X}, \bar{Y}) (c) (0,0) (d) (1,1)
4. If two variables X and Y are independent, the value of regression coefficient b_{yx} is
- (a) 0 (b) 1 (c) 2 (d) any value
5. If the regression line of Y on X is $3X=2Y+6$, the estimated value of Y for a given value of X=10 is
- (a) 10 (b) 12 (c) 8 (d) 4

6. If the correlation coefficient is ± 1 , the two lines of regression are
 (a) coincident (b) parallel (c) perpendicular (d) none
7. Given the regression lines $X + 2Y - 5 = 0$, $2X + 3Y - 8 = 0$ and $\text{var}(X) = 16$ find $\text{var}(Y)$.
8. Given the following values $\bar{X} = 9, \bar{Y} = 16, \dagger_x = 2, \dagger_y = 1.5$ and $\text{cov}(x, y) = 0.84$ find the regression line of Y on X.
9. If two regression equation of data on advertising expenditure(x lakhs) and sales(y lakhs) are as follows
 $(y-90) = 3.2(x-10)$ and $(x-10) = 0.2(y-90)$
 (a) Write the value of regression coefficient of x on y.
 (b) Find the value of correlation coefficient.
 (c) Estimate the advertising expenditure required to attain a sales target of Rs 120 lakhs.
10. From the following results, obtain the two regression equations and estimate the yield of crops when rainfall is 29 cms. Coefficient of correlation between yield and rainfall is 0.52.

	Y(Yeild in kgs)	X(Rainfall in cms)
Mean	508.4	26.7
S.D	36.7	4.6