

### Exercise on Correlation & Regression

#### Problem 1:

The CEO of a Micro Finance Institution (MFI) believes there is a relationship between the number of client contacts by the field officers and reimbursement of loan by the clients. To document this assertion, he gathered the information indicating the number of client contacts and amount of loan reimbursement as given below:

Number of client contacts by the field officers	Amount of loan reimbursement by the clients (Tk. thousands)	Number of client contacts by the field officers	Amount of loan reimbursement by the clients (Tk. thousands)
14	24	23	30
12	14	48	90
20	28	50	85
16	30	55	120
46	80	50	110

- Draw a scatter diagram for the given data and comment on the approximate relationship between the variables under study.
- Determine using Pearson's coefficient of correlation, whether there exist any relationship between the number of client contacts by the field officers and the amount of loan reimbursement by the clients.

#### Problem 2:

With the following data in 6 cities, determine and comment on the relation between the density of population and death rate:

City	Area (in sq Km)	Population (in '000)	Number 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

**Problem 3:**

The production department of 'CATSEYE' wants to explore the relationship between the number of employees who assemble polo shirt and the number produced. As an experiment, two employees were assigned to assemble polo shirt. They produced 15 during a one hour period. Then four employees assembled them and produced 25 during a one – hour period. The complete set of paired observations follows.

Number of assemblers	One-Hour Production (Units)
2	15
4	25
1	10
5	40
3	30

The dependent variable is production; that is assumed that the level of production depends upon the number of employees.

- Draw a scatter diagram.
- Based on scatter diagram, does there appear to be any relationship between the number of assemblers and production? Explain.
- Compute the coefficient of correlation and hence comment on the relationship of the concerned variables.
- Evaluate the strength of the relationship by computing the determination.

**Problem 4:**

The following sample observations were randomly selected

X:	5	3	6	3	4	4	6	8
Y:	-13	15	7	12	13	11	9	5

Determine the coefficient of correlation and the coefficient of determination. Interpret the association between X and Y.

**Problem 5:**

Dhaka Metropolitan Police (DMP) authority is considering increasing the number of police in an effort to reduce crime. Before making a final decision the DMP authority asks the officer in charge (OC) of different police stations to submit report on the number of police force in each of the police stations and the number of reported crimes in each of the selected area covered by different police station. The DMP collected the following sample data:

Name of Police station	Number of police force	Number of crimes reported	Name of Police station	Number of police force	Number of crimes reported
Badda	150	172	Mohammadpur	175	214
Mugdapara	177	135	Dhanmondi	125	65
Khilgaon	255	56	Mirpur	112	76
Uttara	278	78	Hazaribagh	224	85
Shahbagh	176	74	Turag	165	112

- If we want to estimate crimes on the basis of the number of police, which variable is the dependent variable and which is the independent variable?
- Draw a scatter diagram.
- Determine the coefficient of correlation.
- Determine the coefficient of determination.
- Interpret these statistical measures.

#### Problem 6:

It is assumed that sales of major home appliances vary with the standing housing market: when new home sales are good, so are sales of dishwashers, washing machines, driers and refrigerators. A trade association compiled the following data (in thousands of units) on housing sales and major home appliances sales:

Apartment Sales (thousand)	Appliance Sales (thousand)	Apartment Sales (thousand)	Appliance Sales (thousand)
2.0	5.0	4.0	7.7
2.5	5.5	4.2	8.4
3.2	6.0	4.6	9.0
3.6	7.0	4.8	9.7
3.3	7.2	5.0	10.0

- Draw a scatter diagram for the given data and comment on the approximate relationship between the variables under study.
- Using Pearson's correlation coefficient, determine whether there exists any relationship between the apartment sales and the amount of home appliance sales and also comment on the strength of the relationship.
- Develop an equation for the relationship between appliances sales (in thousand) and apartment sales (in thousands) and comment.
- Using coefficient of determination comment on how much variation in the home appliances sales is expressed by apartment sales.

**Problem 7:**

For the following data

X	33	38	24	61	52	45	65	82	29	79	50	63
Y	3	7	6	6	10	12	12	13	12	15	14	13

- Plot the scatter diagram
- Develop the estimating equation that best describe the data.
- Calculate the sample coefficient of determination and sample coefficient of correlation.

**Problem 8:**

The CEO of a Micro Finance Institution (MFI) believes there is a relationship between the number of client contacts by the field officers and reimbursement of loan by the clients. To document this assertion, he gathered the information indicating the number of client contacts and amount of loan reimbursement as given below:

Number of client contacts by the field officers	Amount of loan reimbursement by the clients (Tk. thousands)	Number of client contacts by the field officers	Amount of loan reimbursement by the clients (Tk. thousands)
14	24	23	30
12	14	48	90
20	28	50	85
16	30	55	120
46	80	50	110

- Draw a scatter diagram for the given data and comment on the approximate relationship between the variables under study.
- Determine the regression equation of loan reimbursement by the clients on the number of client contacts by the field officers.
- Determine the estimated loan reimbursement if 40 and 75 client contacts have been made by the field officers.

**Problem 9:**

The following data relate to the scores obtained by 9 salesmen of a company in an intelligence test and their weekly sales (in tk. '000s)

Salesmen:	A	B	C	D	E	F	G	H	I
Test scores	50	60	50	60	80	50	80	40	70
Weekly sales	30	60	40	50	60	30	70	50	60

- Determine whether there is any relation between the test scores and weekly sales.
- Obtain regression equation of sales on intelligence test scores of the salesmen

- iii. If the intelligence test scores of a salesman is 65, what would be his expected weekly sales?

**Problem 10:**

Cost accounts often estimate overhead based on the level of production. At the standard limiting co. they have collected information on overhead expenses and units produced at different plants.

Overhead (Y)	191	170	272	155	280	173	234	116	153
Units (X)	40	42	53	35	56	39	48	30	37

- iv. Compute the value of correlation co-efficient and hence comment.  
 v. Develop the regression equation for the cost accounts.  
 vi. Predict the overhead when 50 units are produced.

**Problem 11:**

A study by the Atlanta, Georgia, and Department of Transportation on the effect of bus-ticket prices on the number of passengers produced the following results:

Ticket prices (Cuts) (X)	25	30	35	40	45	50	55	60
Passenger per 100 miles (Y)	800	780	780	660	640	600	620	620

- i. Compute the value of correlation co-efficient and hence comment.  
 ii. Develop the regression equation that best describe the data.  
 iii. Predict the number of passengers if the ticket price is 50 cents.

**Problem 12:**

A company is introducing a job evaluation scheme in which all jobs are graded by points for skill, responsibility and so on. Monthly pay scales (tk. in '000) are then drawn up according to the number of points allocated and other factors such as experience and local conditions. To date the company has applied this scheme to jobs:

Job:	A	B	C	D	E	F	G	H	I
Points	5	25	7	19	10	12	15	28	16
Pay (in '000 tk.)	3.0	5.0	3.25	6.5	5.5	5.6	6.0	7.2	6.1

- i. Find the least squares regression line for linking pay scales to points  
 ii. Estimate the monthly pay for a job graded by 20 points.