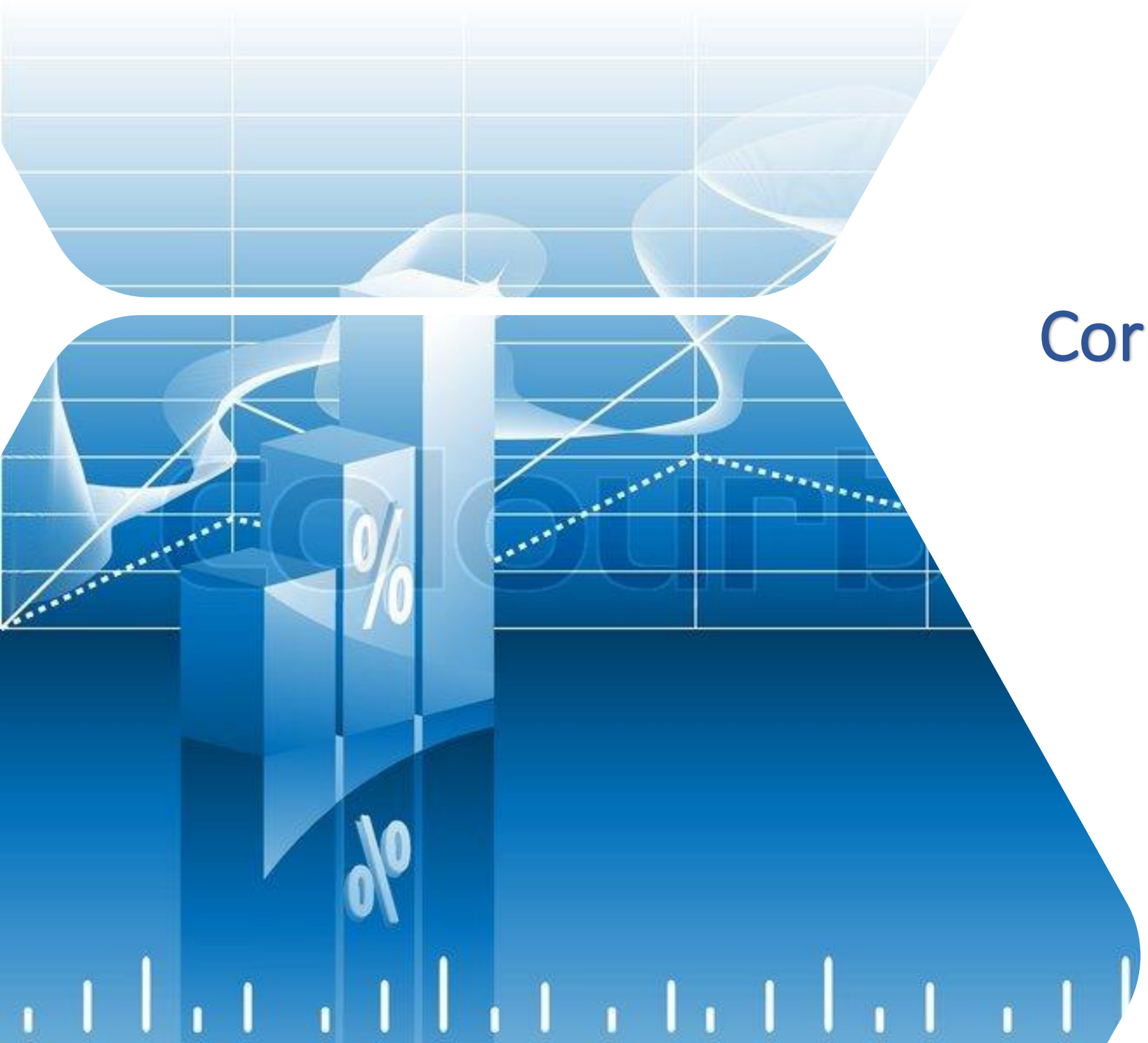


Correlation Analysis (Part-2)

Objectives

- **Find out Whether there is any relation between or among of the variables.**
- **Strength of the relationship.**

OUTLINE

- ❑ **Measure correlation by correlation co-efficient (r)**
- ❑ **Interpretation of r**
- ❑ **Properties of r**
- ❑ **Advantages, disadvantages and uses of correlation coefficient.**
- ❑ **Exercise.**

METHODS OF STUDYING CORRELATION

Correlation can be studied by the following methods:

1. **Scatter Diagram Method**
2. **Karl Pearson's Correlation Coefficient**

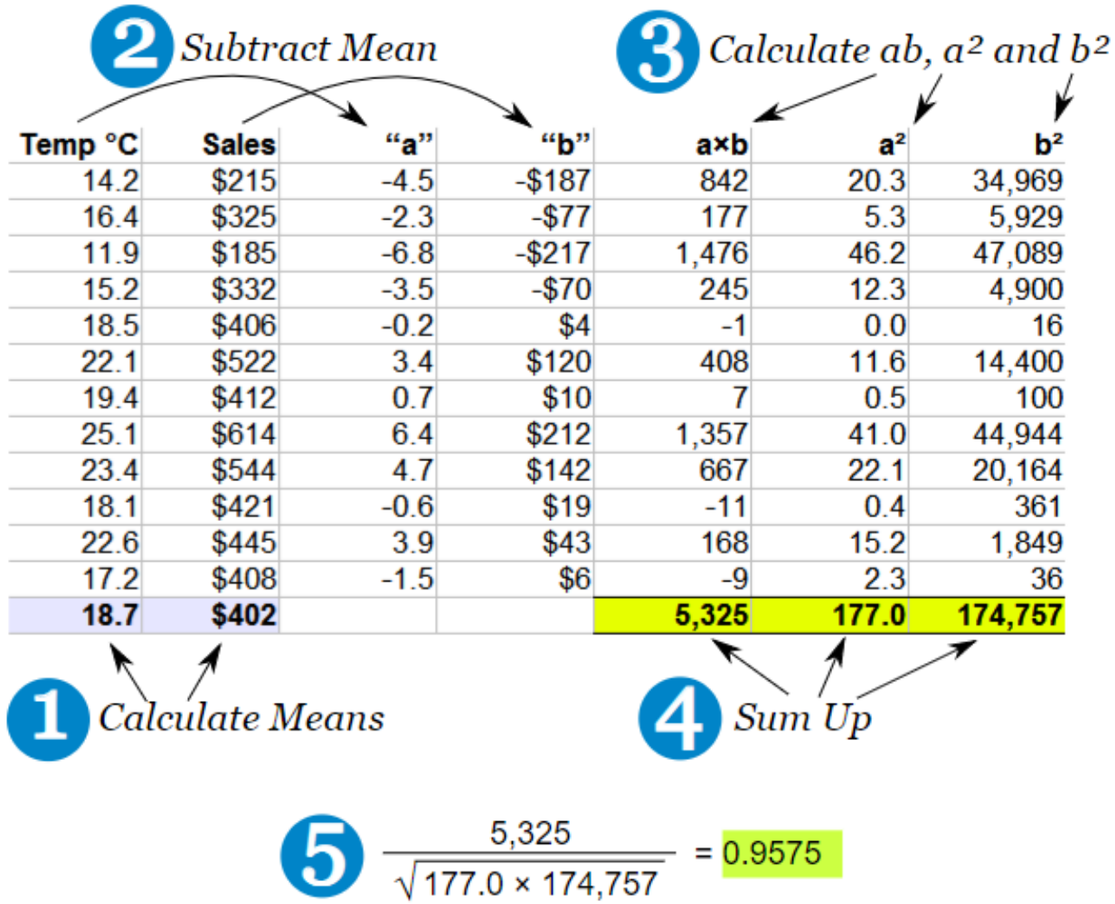
We will practice the maths of this technique.

Calculation of correlation by Karl Pearson's Correlation Coefficient

2. Karl Pearson's Correlation Coefficient

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Where, n= number of pairs in sample.



Temp °C	Sales	"a"	"b"	a×b	a²	b²
14.2	$215	-4.5	-$187	842	20.3	34,969
16.4	$325	-2.3	-$77	177	5.3	5,929
11.9	$185	-6.8	-$217	1,476	46.2	47,089
15.2	$332	-3.5	-$70	245	12.3	4,900
18.5	$406	-0.2	$4	-1	0.0	16
22.1	$522	3.4	$120	408	11.6	14,400
19.4	$412	0.7	$10	7	0.5	100
25.1	$614	6.4	$212	1,357	41.0	44,944
23.4	$544	4.7	$142	667	22.1	20,164
18.1	$421	-0.6	$19	-11	0.4	361
22.6	$445	3.9	$43	168	15.2	1,849
17.2	$408	-1.5	$6	-9	2.3	36
18.7	**$402**			**5,325**	**177.0**	**174,757**

Example of correlation calculation

Problem: Mr. Johnson is concerned about the cost to students of textbooks. He believes there is a relationship between the number of pages in the text and the selling price of the book. To provide insight into the problem he selects a sample of eight textbooks currently on sale in the bookstore. Compute the correlation coefficient.

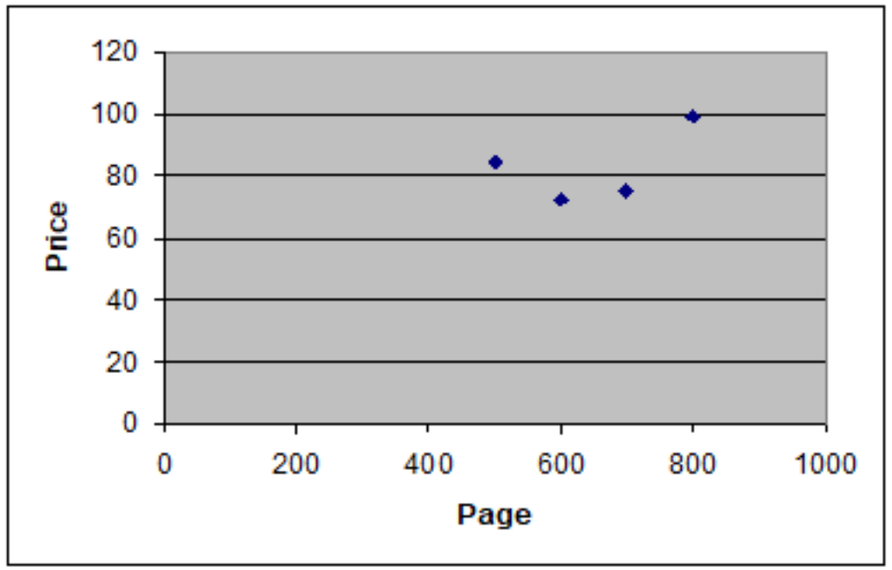
Book	Page(x)	Price($)(y)
Introduction to History	500	84
Basic Algebra	700	75
Business Management	800	99
Introduction to Sociology	600	72

a) Draw Scatter diagram
b) Determine the coefficient of correlation
c) Interpret the result.

Example of correlation calculation

Solution:

a) Scatter diagram:



b) Here,

We know, Correlation coefficient,

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Where,

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2600}{4} = 650$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{330}{4} = 82.5$$

Example of correlation calculation

X_i	Y_i	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})^* (Y_i - \bar{Y})$
500	84	-150	1.5	22500	2.25	-225
700	75	50	-7.5	2500	56.25	-375
800	99	150	16.5	22500	272.25	2475
600	72	-50	-10.5	2500	110.25	525
Total				$\sum_{i=1}^{n}(X_i - \bar{X})^2 = 50000$	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 441$	$\sum_{i=1}^{n}(X_i - \bar{X}) * (Y_i - \bar{Y}) = 2400$

Example of correlation calculation

Now,
From the correlation coefficient

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}} = \frac{2400}{\sqrt{50000 * 441}}$$

$$= 0.511$$

Example of correlation calculation

c) Interpretation: The correlation between the number of pages and the selling price of the book is 0.511. This indicates a moderate association between the variables.

Correlation Coefficient Interpretation:

Coefficient Range	Strength of Relationship
0.00 - 0.20	Very Low
0.21 - 0.40	Low
0.41- 0.60	Moderate
0.60 - 0.80	High Moderate
0.81- 1.00	Very High

Properties of Correlation Coefficient

1. Correlation coefficient lies between -1 and +1. Symbolically, $-1 \leq r \leq +1$.
2. Correlation coefficient has no unit.
3. Correlation coefficient can only measures linear relationship.
4. Correlation coefficient can only be measured for quantitative variables.

Criteria of Correlation coefficient

Advantages of Correlation Coefficient

1. It is simple to understand and easy to calculate.
2. It is very useful in the case of data which are quantitative in nature.
3. It provides direction as well as strength of linear relationship.

Disadvantages of Correlation Coefficient:

1. It cannot calculate nonlinear relationship.
2. For qualitative data correlation coefficient is not possible to measure.

Some exercise Problems to solve

The following table shows the mean weight in kilograms of members of a group of young children of various ages.

Age (x years)	1.6	2.5	3.3	4.4	5.6
Weight (y kg)	12	15	16	17	20

- Find Correlation coefficient and interpret it.
- Draw scatter diagram.

The following table shows the average weights for given heights in a population of men.

Heights (x cm)	160	165	170	175	180	185
Weights (y kg)	65.1	67.9	70.1	72.8	75.4	77.2

Some exercise Problems to solve

The following table shows the amount of diesel required by a train to travel certain distances.

Distance (x km)	90	150	230	310	390
Diesel used (y litres)	19.2	33.9	49.0	79.5	89.9

- Find Correlation coefficient and interpret it.
- Draw scatter diagram.

In a survey of insect life near a stream, a student collected data about the number of different insect species (y) that were found at different distances (x) in meters from the stream.

Distance (x)	2	5	8	11	14	17	22	33	39
Insect species (y)	26	25	19	19	14	9	5	3	2


Thank
you