Correlation

Md. Ismail Hossain Riday

Correlation

- In real life situations, especially in social sciences and in business, we often to know whether two or more variables are related, and if so, how they are related.

- The following are some questions of interest,

1. Is there a relationship between two or more variables?
2. If so, what is the relationship?

Association

Connection

Correlation

- Relationship

Correlation

- Relationship between two or more variables

Correlation

- Relationship between two or more variables
- Which gives us strength or degree

Correlation

- Relationship between two or more variables
- Which gives us strength or degree and direction of association

Correlation:

Relationship between two or more variables which gives us strength or degree and direction of association

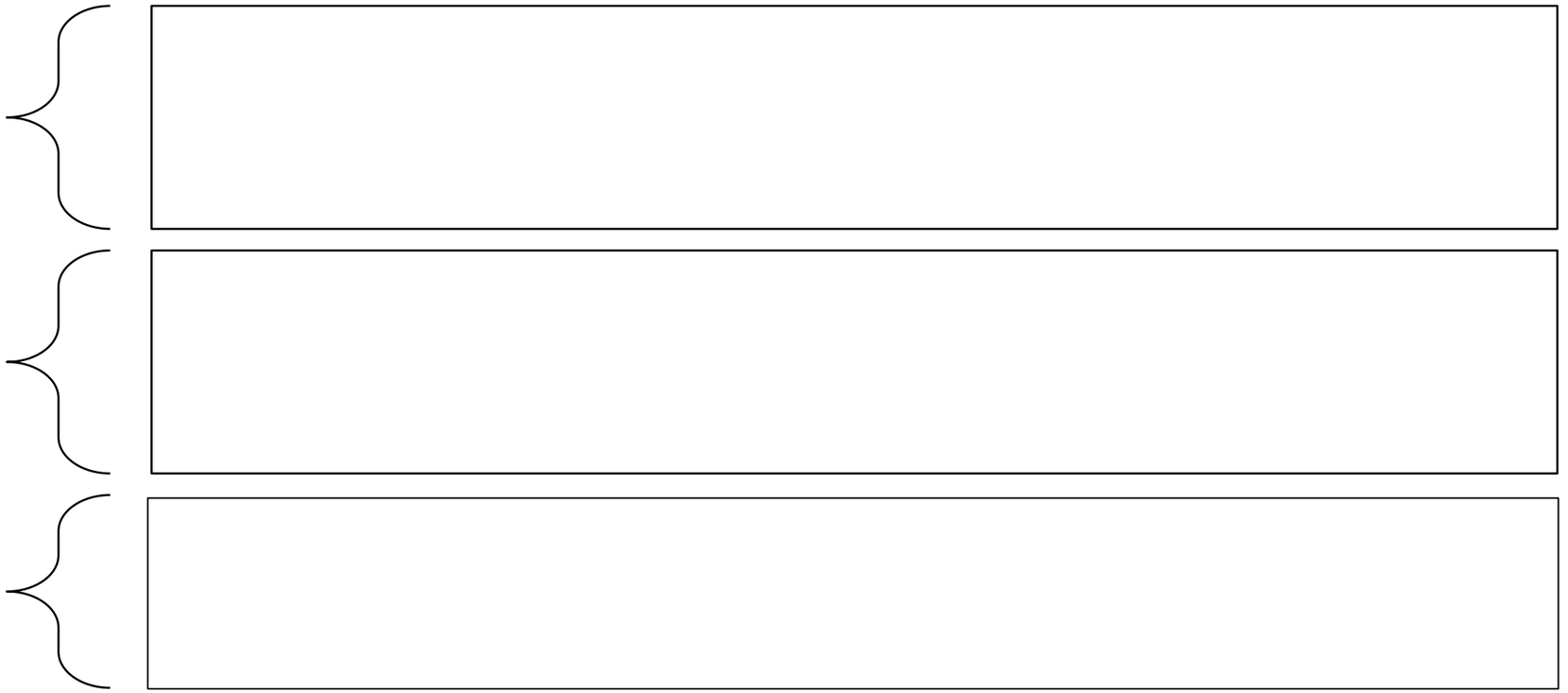
Objectives

- The primary objective of correlation analysis is to measure,

1. Degree or strength of relationships

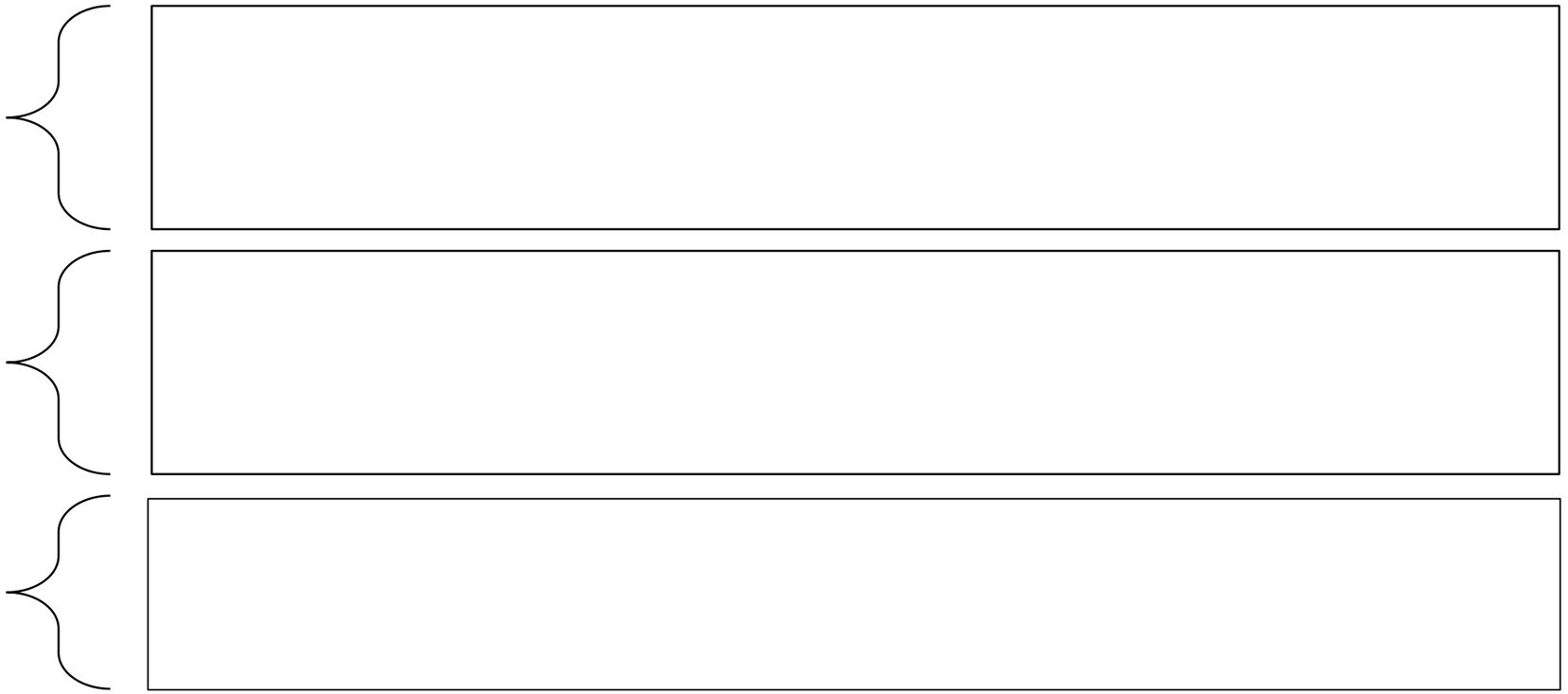
2. Direction of relationship

Types of Correlation



Types of Correlation

On the basis of direction



Types of Correlation

On the basis of direction

On the basis of variables

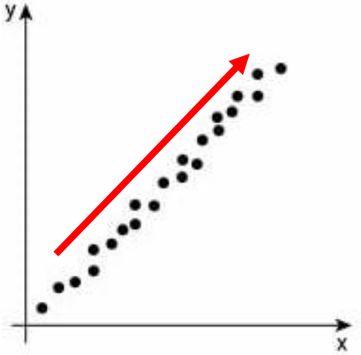
Types of Correlation

On the basis of direction

On the basis of variables

On the basis of linearity

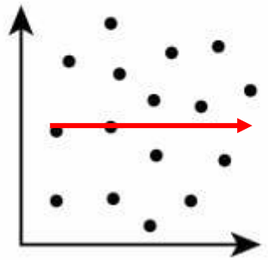
Types of Correlation



On the basis of direction

- Positive correlation
- Negative correlation

No Correlation



On the basis of variables

On the basis of linearity

Types of Correlation

On the basis of direction	• Positive correlation • Negative correlation
On the basis of variables	• Simple correlation • Multiple correlation
On the basis of linearity	

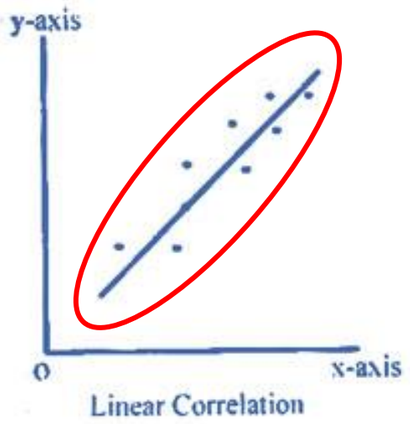
Types of Correlation

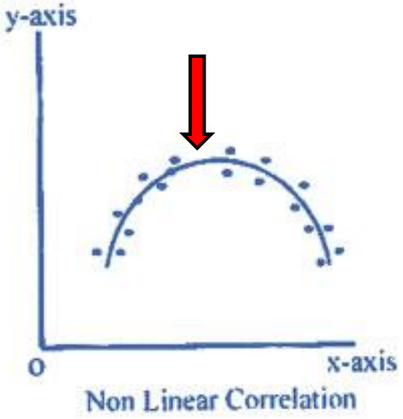
On the basis of direction	• Positive correlation • Negative correlation
On the basis of variables	• Simple correlation (Two variables) • Multiple correlation
On the basis of linearity	

Types of Correlation

On the basis of direction	• Positive correlation • Negative correlation
On the basis of variables	• Simple correlation • Multiple correlation (More than two)
On the basis of linearity	

Types of Correlation

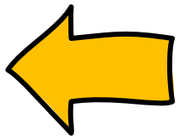
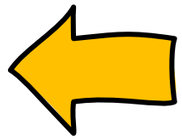




On the basis of direction	• Positive correlation • Negative correlation
On the basis of variables	• Simple correlation • Multiple correlation
On the basis of linearity	• Linear correlation • Non-Linear correlation

Measures

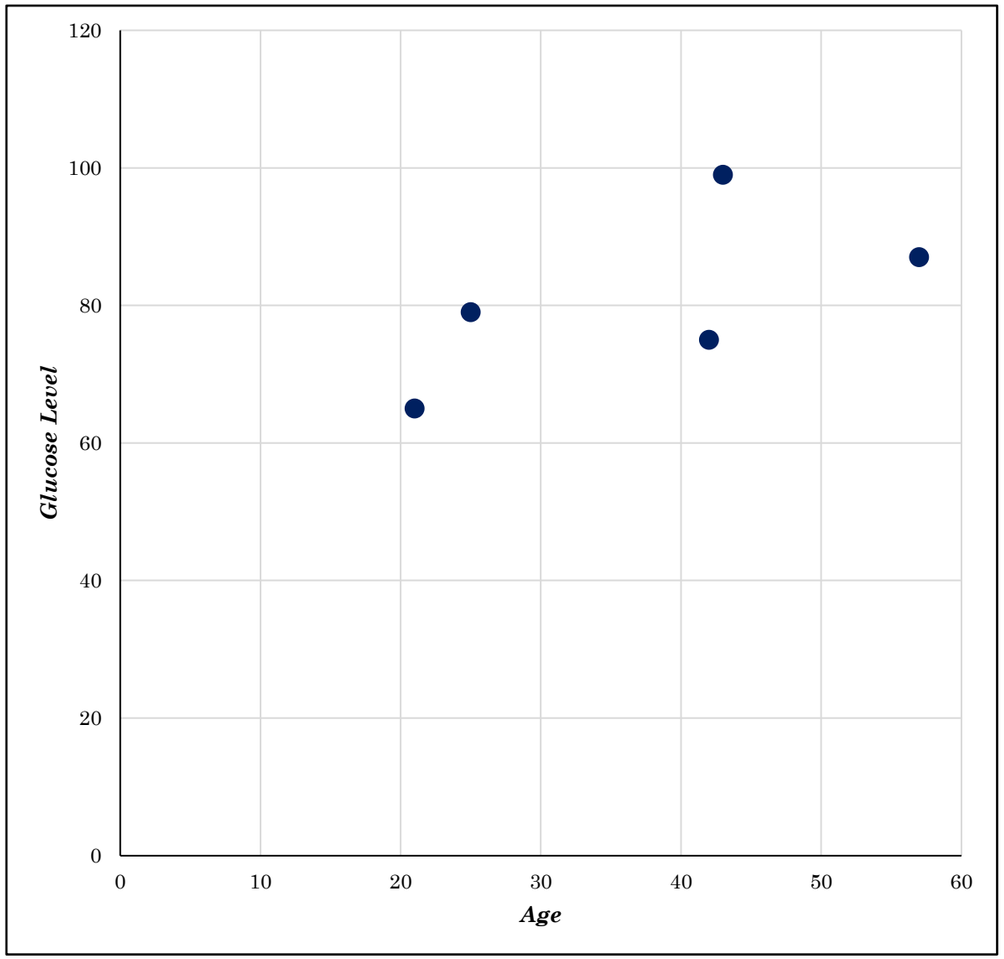
- Correlation can be measured by

1. Scatter Diagram 
2. Karl Pearson’s Correlation Coefficient 
3. Spearman Rank Correlation Coefficient
4. Method of Least Square

Scatter Diagram

Age	Glucose Level
43	99
21	65
25	79
42	75
57	87

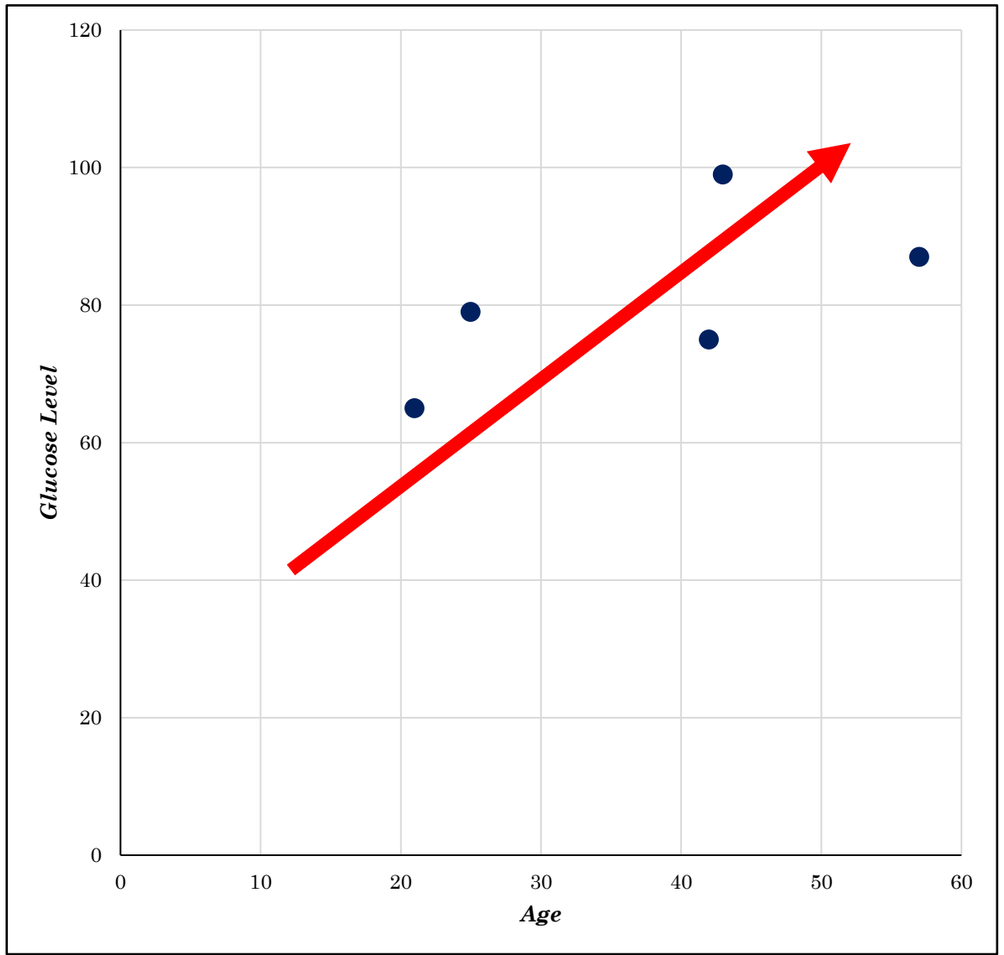
a) Draw the scatter diagram
b) Is there any relationship between “Age” and “Glucose Level”?



With the increase of age, the Glucose Level also increased. Thus, there is a positive correlation between “Age” and “Glucose Level”.

Limitations

- Accurate degree and strength of correlation can not be obtained by scatter diagram.



Karl Pearson's Correlation

- Both variables are measured in interval or ratio scales
- Relationship between variables is linear
- Denoted by ***r***

Karl Pearson's Correlation

$$r = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\ \sqrt{\sum(y_i - \bar{y})^2}}$$

Karl Pearson's Correlation

$$r = \frac{\sum (x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\ \sqrt{\sum (y_i - \bar{y})^2}}$$

$$r = \frac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}}\ \sqrt{\sum y_i^2 - \dfrac{(\sum y_i)^2}{n}}}$$

Karl Pearson's Correlation

$$r = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\ \sqrt{\sum(y_i - \bar{y})^2}}$$

$$\boldsymbol{r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}\ \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}}$$

Range of r is [-1 to +1]

If $r = +1$; Perfect positive correlation

If $r = -1$; Perfect negative correlation

If $r = 0$; No correlation

Correlation	Negative	Positive
Weak	-0.29 to -0.10	0.10 to 0.29
Medium	-0.49 to -0.30	0.30 to 0.49
Moderate	-0.50 to -0.79	0.50 to 0.79
Strong	-1.00 to -0.80	0.80 to 1.00

Karl Pearson's Correlation

Example: The monthly income and saving data for a sample of 10 garments workers are given below:

Income ($)	Savings ($)
60	5
66	7
66	8
66	9
68	11
68	12
70	14
72	16
74	21
80	27

a) Draw the scatter plot

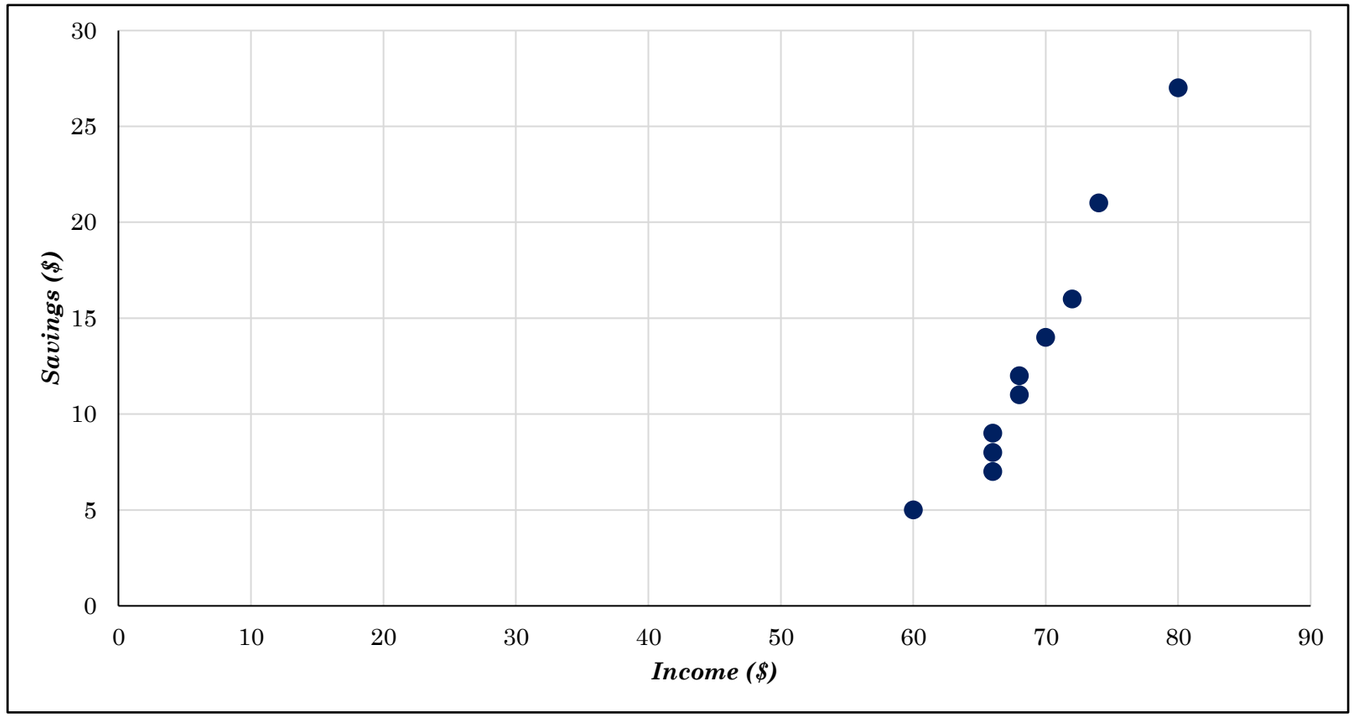
b) Compute the correlation coefficient value with proper interpretation

Karl Pearson's Correlation

Example: The monthly income and saving data for a sample of 10 garments workers are given below:

Income ($)	Savings ($)
60	5
66	7
66	8
66	9
68	11
68	12
70	14
72	16
74	21
80	27

a) Draw the scatter plot



Cont.

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

x (income)	y (Savings)	x_i^2	y_i^2	$x_i \times y_i$
60	5	3600	25	300
66	7	4356	49	462
66	8	4356	64	528
66	9	4356	81	594
68	11	4624	121	748
68	12	4624	144	816
70	14	4900	196	980
72	16	5184	256	1152
74	21	5476	441	1554
80	27	6400	729	2160
$\sum x_i = 690$	$\sum y_i = 130$	$\sum x_i^2 = 47876$	$\sum y_i^2 = 2106$	$\sum x_i y_i = 9294$

Cont.

Now, the correlation coefficient

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}}$$

Cont.

Correlation	Negative	Positive
Weak	-0.29 to -0.10	0.10 to 0.29
Medium	-0.49 to -0.30	0.30 to 0.49
Moderate	-0.50 to -0.79	0.50 to 0.79
Strong	-1.00 to -0.80	0.80 to 1.00

Now, the correlation coefficient

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}} = \frac{9294 - \frac{690 \times 130}{10}}{\sqrt{47876 - \frac{(690)^2}{10}} \sqrt{2106 - \frac{(130)^2}{10}}} = 0.97$$

The values of $r = 0.97$, suggests a strong positive correlation between income and savings of garments workers. That is, as income increases, there is a strong tendency for saving increase.

Self Practice

- Following are the heights and weights of 10 students

Height	62	72	68	58	65	70	66	63	60	72
Weight	50	65	63	50	54	60	61	55	54	65

Properties

1. Correlation coefficient has no unit.
2. The sign of correlation coefficient gives the direction of the association.
3. The correlation coefficient is between -1 and $+1$.
4. Correlation coefficient is a symmetric measure, $i.e., r_{xy} = r_{yx}$
5. Correlation is sensitive to extreme observations.

Thank You