PHYSICAL EDUCATION

Subject : Physical Education
Semester : 4th
Paper No. and Title : (401) Test, Measurement and Evaluation in Physical Education

STUDYING DEGREE OF RELATIONSHIP BETWEEN TWO VARIABLES: CORRELATION

1.0 Objectives:

This unit introduces the correlation. After reading this unit we will able to:

- define correlation
- know Karl Pearson's method and rank method for calculating correlation

1.1 Introduction:

The relationship between two variables such that a change in one variable results in a positive or negative change in other variable and also a greater change in one variable results in corresponding greater or smaller change in other variable is called correlation.

The amount of correlation(degree of relationship) between variables is measured by coefficient of correlation. It is denoted by "r" or by "ρ".

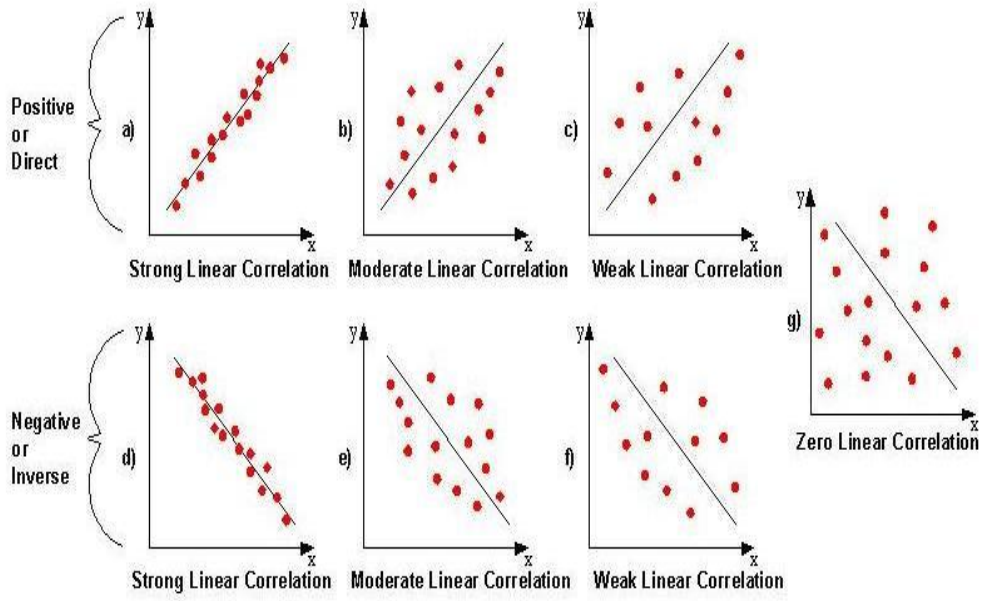
1.2 Types of correlation

- **Positive Correlation-** Here, the values of the two variables deviate in the same direction, i.e., the increase in the values of one variable results in a corresponding increase in the values of the other variable. Ex. Heights and weights.

- **Negative Correlation-** Here, the decrease in the values of one variable results in a corresponding increase in the values of the other variable. Ex. Price and demand of a commodity.

- **Linear Correlation-** Corresponding to a unit change in one variable, there is a constant change in the other variable over the entire range of the values.

- **Non-Linear Correlation-** Here, corresponding to a unit change in one variable, the other variable does not change at a constant rate but at fluctuating rate



1.3 Correlation & Causation:

- Correlation analysis enables us to have an idea about the degree and direction of relationship between two variables but fails to reflect upon the cause and effect relationship between the variables.
- Causation always implies correlation, however the converse is not true.
- The high degree of correlation between variables may be due to:
- **Mutual Dependence**: The phenomena under study may inter-influence each other. It is observed in data relating to economic and business situations. Ex. Prices of a commodity are influenced by the forces of supply and demand.

1.4 Methods of Studying Correlation:

Some commonly used methods for studying correlation(linear) between two variables are-

- Karl Pearson coefficient of correlation (covariance method)
- Spearman's rank correlation coefficient (Rank method)
- Scatter diagram method
- Two way frequency table
- Concurrent deviation method etc.

1.4.1 Karl Pearson's method:

This method of studying linear relationship between two variables was suggested by Karl Pearson (1867-1936)

It is denoted by r(x, y), r_{xy} or simply r, where x and y are the two variables. Thus,

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y} \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{\sum X^2 - n\bar{X}^2}\sqrt{\sum Y^2 - n\bar{Y}^2}}$$

Where, $Cov(x,y) = \frac{1}{n}\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)$

$$\sigma_x = \sqrt{\frac{1}{n}\sum\left(x-\bar{x}\right)^2} \quad \text{and} \quad \sigma_y = \sqrt{\frac{1}{n}\sum\left(y-\bar{y}\right)^2}$$

The Karl Pearson's method of studying linear relationship between two variables x and y can also be written as

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x{}^2 . \sum d_y^2}}$$

Where, $d_x = x - \bar{x}$ and $d_y = y - \bar{y}$

- If the variables are not capable of quantitative measurements but can be arranged in serial order (qualitative attributes like beauty, honesty, character, morality etc.) then, the Karl Pearson's coefficient of correlation can not be used.
- The Spearman rank correlation is the alternative method, which consists in obtaining the correlation coefficient between the ranks of n individuals in the two attributes.

1.4.2 Rank Method:

The Spearman's rank correlation coefficient denoted by ρ (rho) is given as $\rho = 1 - \frac{6\sum d^2}{n(n^2-1)}$

Where, d= the difference between the pair of ranks of the same individual in the two characteristics and

n=number of pairs

- ***When actual ranks are given-*** ρ is calculated directly using the above formula

- ***When actual ranks are not given-***

  - Covert the data into ranks by giving rank 1 to the highest observation, rank 2 to the next highest (next lowest) observation and so on.(ascending/ descending)

  - Then ρ is calculated in the same manner.

- ***When ranks are repeated*** –

  If there is a tie in the variable data, i.e., there is more than one item with the same value in either or both the series, then common ranks are assigned to the repeated items. These common ranks are the arithmetic mean of the ranks which these items would have got if they were different from each other. The next item will be given the rank next to the rank used in computing the common rank.

  Then ρ is calculated as $\rho = 1 - \frac{6\left[\sum D^2 + \Sigma \frac{m(m^2-1)}{12}\right]}{n(n^2-1)}$

  Where, m is the number of times an items is repeated. This factor is to be added for each repeated value in both the series.

Interpretation / Properties of r or ρ:

- If r = +1, then there is a perfect positive correlation between the variables.

- If r = -1, then there is a perfect negative correlation between the variables.

- If r = 0, then the variables are uncorrelated, i.e., there is no linear (straight line) relationship between the variables. however., if r = 0, it does not imply that the variables are independent.

- For other values of r, $-1 \le r \le 1$ the nearer the value of r to 1, the closer the relation between the variables and nearer the value of r to 0, the less close the relationship between them.

Example 1: Calculate the Pearson's product moment correlation coefficient for the given data of reading scores and arithmetic scores

Reading Scores: 95 90 85 80 75 70 65 60 55

Arithmetic Scores: 76 78 77 71 75 76 73 75 74

X	Y	X-square	Y-square	XY

95	76	9025	5776	7220
90	78	8100	6084	7020
85	77	7225	5929	6545
80	71	6400	5041	5680
75	75	5625	5625	5625
70	76	4900	5776	5320
65	73	4225	5329	4745
60	75	3600	5625	4500
55	74	3025	5476	4070
675	675	52125	50661	50725

$$r = \frac{\sum XY - n\bar{X}\bar{Y}}{\sqrt{\sum X^2 - n\bar{X}^2}\sqrt{\sum Y^2 - n\bar{Y}^2}}$$

$$= 0.43$$

Example 2: Calculate Spearman's rank correlation coefficient between advertisement cost and sales from the following data:

Adv. Cost (Rs. 1000) : 39 65 62 90 82 75 25 98 36 78

Sales (Rs. 100000) : 47 53 58 86 62 68 60 91 51 84

Solution: The formula for rank correlation coefficient is:

Adv. Cost	Sales	R1	R2	d= R1-R2	d-square
39	47	8	10	-2	4
65	53	6	8	-2	4
62	58	7	7	0	0
90	86	2	2	0	0
82	62	6	5	-2	4
75	68	5	4	1	1
25	60	10	6	4	16
98	91	1	1	0	0
36	51	9	9	0	0
78	84	4	3	1	1
Total					30

$$\rho = 1 - \frac{6 \sum d^2}{n\,(n^2 - 1)}$$

= 0.82

Note: Ranks started from 1, 2, …, 10 are to be considered either from the highest value or from the lowest value in order. In this problem ranks are given starting from the highest value in a series i.e. rank 1 to the highest value, rank 2 to the next one and so on and so forth