Chapter 7 - Correlation

Question 1:

The unit of correlation coefficient between height in feet and weight in kgs is

(i) kg/feet

(ii) percentage

(iii) non-existent

Answer:

As there is non-existent of correlation between the height in feet and weight in kilograms, so the unit of correlation between the two is zero.

Question 2:

The range of simple correlation coefficient is

(i) 0 to infinity

(ii) minus one to plus one

(iii) minus infinity to infinity

Answer:

The range of simple correlation coefficient is from (−) 1 to (+) 1

Question 3:

If r_{xy} is positive the relation between X and Y is of the type

(i) When Y increases X increases

(ii) When Y decreases X increases

(iii) When Y increases X does not change

Answer:

When the variables Y and X share positive relationship (i.e. when Y and X both increases simultaneously), then the value of r_{xy} is positive.

Question 4:

If r_{xy} = 0 the variable X and Y are

(i) linearly related

(ii) not linearly related

(iii) independent

Answer:

The value of r_{xy} becomes 0 when the two variables are not linearly related to each other. It may happen that both the variables may be non-linearly related to each other. It does not necessarily imply that both are independent of each other.

Question 5:

Of the following three measures which can measure any type of relationship

(i) Karl Pearson's coefficient of correlation

(ii) Spearman's rank correlation

(iii) Scatter diagram

Answer:

Scatter diagram can measure any type of relationship whether the variables are highly related or not at all related. Just by looking at the diagram, the viewer can easily conclude the relationship between the two variables involved. On the other hand, Karl Pearson's coefficient of correlation is not suitable for the series where deviations are calculated from assumed mean. Likewise, Spearman's rank correlation also disqualifies to measure any kind of relationship as its domain is restricted only to the qualitative variables (leaving quantitative variables).

Question 6:

If precisely measured data are available the simple correlation coefficient is

(i) more accurate than rank correlation coefficient

(ii) less accurate than rank correlation coefficient

(iii) as accurate as the rank correlation coefficient

Answer:

Generally, all the properties of Karl Pearson's coefficient of correlation are similar to that of the rank correlation coefficient. However, rank correlation coefficient is generally lower or equal to Karl Pearson's coefficient. The reason for this difference between the two coefficients is because the rank correlation coefficient uses ranks instead of the full set of observations that leads to some loss of information. If the precisely measured data are available, then both the coefficients will be identical.

Question 7:

Why is *r* preferred to covariance as a measure of association?

Answer:

Although correlation coefficient is similar to the covariance in a manner that both measure the degree of linear relationship between two variables, but the former is generally preferred to covariance due to the following reasons.

1. The value of the correlation coefficient (*r*) lies between 0 and 1. Symbolically $-1 \leq r \leq +1$

2. The correlation coefficient is scale free.

Question 8:

Can *r* lie outside the –1 and 1 range depending on the type of data?

Answer:

No, the value of *r* cannot lie outside the range of –1 to 1. If $r = -1$, then there exists perfect negative correlation and if $r = 1$, then there exists perfect positive correlation between the two variables. If at any point of time the calculated value of *r* is outside this range, then there must be some mistake committed in the calculation.

Question 9:

Does correlation imply causation?

Answer:

No, correlation does not imply causation. The correlation between the two variables does not imply that one variable causes the other. In other words, cause and effect relationship is not a prerequisite for the correlation. Correlation only measures the degree and intensity of the relationship between the two variables, but surely not the cause and effect relationship between them.

Question 10:

When is rank correlation more precise than simple correlation coefficient?

Answer:

Rank Correlation method is more precise than simple correlation coefficient when the variables cannot be measured quantitatively. In other words, rank correlation method measures the correlation between the two qualitative variables. These variable attributes are given the ranks on the basis of preferences. For example, selecting the best candidate in a dance competition depends on the ranks and preferences awarded to him/her by the judges. Secondly, the rank correlation method is preferred over the simple correlation coefficient when extreme values are present in the data. In such case using simple correlation coefficient may be misleading.

Question 11:

Does zero correlation mean independence?

Answer:

Correlation measures the linear relationship between the two variables. So, r being 0 implies the absence of linear relationship. But they may be non-linearly related. Hence, if two variables are not correlated, it does not necessarily follow that they are independent.

Question 12:

Can simple correlation coefficient measure any type of relationship?

Answer:

No, the simple correlation coefficient cannot measure any type of relationship. The simple correlation coefficient can measure only the direction and magnitude of linear relationship between the two variables. It cannot measure non-linear relationship like quadratic, trigonometric, cubic, etc. Therefore, in such cases, the purview of simple correlation coefficient falls short. For example, the simple correlation coefficient may depict that X and Y are not correlated in the equation $X = Y^2$, hence it may be concluded that both the variables are independent, but such conclusion may be wrong.

Question 13:

Collect the price of five vegetables from your local market every day for a week. Calculate their correlation coefficients. Interpret the result.

Answer:

This question is about multivariate correlation that is out of syllabus

Question 14:

Measure the height of your classmates. Ask them the height of their benchmate. Calculate the correlation coefficient of these two variables. Interpret the result.

Answer:

Height of Classmate X	Height of Benchmate Y
67	65
56	66
65	57
68	67

	72		68	
	72		69	
	69		70	
	71		72	

X	Y	XY	X^2	Y^2
67	65	4355	4489	4225
56	66	3696	3136	4356
65	57	3705	4225	3249
68	67	4556	4624	4489
72	68	4896	5184	4624
72	69	4968	5184	4761
69	70	4830	4761	4900
71	72	5112	5041	5184
$\sum X = 540$	$\sum Y = 534$	$\sum XY = 36118$	$\sum X^2 = 36644$	$\sum Y^2 = 35788$

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}} \sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$$

$$= \frac{36118 - \frac{540 \times 534}{8}}{\sqrt{36644 - \frac{(540)^2}{8}} \sqrt{35788 - \frac{(534)^2}{8}}}$$

$$= \frac{36118 - 36045}{\sqrt{36644 - 36450} \sqrt{35788 - 35644.5}}$$

$$= \frac{73}{\sqrt{194} \sqrt{143.5}}$$

$$= \frac{73}{13.93 \times 11.98}$$

$$= \frac{73}{166.88}$$

$$= 0.44$$

Question 15:

List some variables where accurate measurement is difficult.

Answer:

The following are the some variables where the accurate measurement is difficult.

1. Temperature and number of people falling ill.

2. Change in temperature with the height of mountain.

3. Low rainfall and agricultural productivity

4. High population growth and degree of poverty

5. Number of tourists and change in the political atmosphere in India.

Question 16:

Interpret the values of r as 1, –1 and 0.

Answer:

The value of r being 1 implies that there is a perfect positive correlation between the two variables involved. A high value of r (i.e. close to 1) represents a strong positive linear relationship between the two variables.

If $r = –1$, then the correlation is perfectly negative. A negative value of r indicates an inverse relation. A low value of r (i.e. close to –1) represents a strong negative linear relationship between the variables. On the other hand, if the value of $r = 0$, then it implies that the two variables are uncorrelated to each other. But this should not be misunderstood as the variables are independent of each other. The value of r equals zero confirms only the non-existence of any linear relation but the variables may be non-linearly related to each other.

Question 17:

Why does rank correlation coefficient differ from Pearsonian correlation coefficient?

Answer:

Generally, all the properties of Karl Pearson's coefficient of correlation are similar to that of the rank correlation coefficient. However, rank correlation coefficient is generally lower or equal to Karl Pearson's coefficient. Rank correlation coefficient is generally preferred to measure the correlation between the two qualitative variables. These variable attributes are given the ranks on the basis of preferences. The difference between the two coefficients is due to the fact that the rank correlation coefficient uses ranks instead of the full set of observations that leads to some loss of information. If the precisely measured data are available, then both the coefficients will be identical. Secondly, if extreme values are present in the data, then the rank correlation coefficient is more precise and reliable and consequently its value differs from that of the Karl Pearson's coefficient.

Question 18:

Calculate the correlation coefficient between the heights of fathers in inches (X) and their sons (Y)

X	65	66	57	67	68	69	70	72
Y	67	56	65	68	72	72	69	71

Answer:

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	56	3696	4356	3136
57	65	3705	3249	4225
67	68	4556	4489	4624
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
$\sum X = 534$	$\sum Y = 540$	$\sum XY = 36118$	$\sum X^2 = 35788$	$\sum Y^2 = 36644$

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$$

$$= \frac{36118 - \frac{534 \times 540}{8}}{\sqrt{35788 - \frac{(534)^2}{8}}\sqrt{36644 - \frac{(540)^2}{8}}}$$

$$= \frac{36118 - \frac{288360}{8}}{\sqrt{35788 - \frac{285156}{8}}\sqrt{36644 - \frac{291600}{8}}}$$

$$= \frac{36118 - 36045}{\sqrt{35788 - 35644.5}\sqrt{36644 - 36450}}$$

$$= \frac{73}{\sqrt{143.5}\sqrt{194}}$$

$$= \frac{73}{11.98 \times 13.93}$$

$$= \frac{73}{166.88}$$

$$= 0.44$$

Note: As per textbook, correlation coefficient is 0.603. However, as per the above solution, correlation coefficient should be 0.44.

Question 19:

Calculate the correlation coefficient between X and Y and comment on their relationship:

X	−3	−2	−1	1	2	3
Y	9	4	1	1	4	9

Answer:

X	Y	XY	X^2	Y^2
−3	9	−27	9	81
−2	4	−8	4	16
−1	1	−1	1	1
1	1	1	1	1
2	4	8	4	16
3	9	27	9	81
$\sum X = 0$	$\sum Y = 28$	$\sum XY = 0$	$\sum X^2 = 28$	$\sum Y^2 = 196$

$$r = \frac{\sum XY - \dfrac{(\sum X)(\sum Y)}{N}}{\sqrt{\sum X^2 - \dfrac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \dfrac{(\sum Y)^2}{N}}}$$

$$= \frac{0 - \dfrac{0 \times 28}{6}}{\sqrt{28 - \dfrac{(28)^2}{6}}\sqrt{196 - \dfrac{(196)^2}{6}}}$$

$$= \frac{0}{\sqrt{28 - \dfrac{(28)^2}{6}}\sqrt{196 - \dfrac{(196)^2}{6}}}$$

$r = 0$

As the value of r is zero, so there is no linear correlation between X and Y

Question 20:

Calculate the correlation coefficient between X and Y and comment on their relationship

X	1	3	4	5	7	8
Y	2	6	8	10	14	16

Answer:

X	Y	XY	X^2	Y^2
1	2	2	1	4

3	6	18	9	36
4	8	32	16	64
5	10	50	25	100
7	14	98	49	196
8	16	128	64	256
$\sum X = 28$	$\sum Y = 56$	$\sum XY = 328$	$\sum X^2 = 164$	$\sum Y^2 = 656$

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{N}}\sqrt{\sum Y^2 - \frac{(\sum Y)^2}{N}}}$$

$$= \frac{328 - \frac{28 \times 56}{6}}{\sqrt{164 - \frac{(28)^2}{6}}\sqrt{656 - \frac{(56)^2}{6}}}$$

$$= \frac{328 - \frac{1568}{6}}{\sqrt{164 - \frac{784}{6}}\sqrt{656 - \frac{3136}{6}}}$$

$$= \frac{328 - 261.33}{\sqrt{164 - 130.67}\sqrt{656 - 522.67}}$$

$$= \frac{66.67}{\sqrt{33.33}\sqrt{133.33}}$$

$$= \frac{66.67}{5.77 \times 11.55}$$

$$= \frac{66.6}{66.6}$$

$$= 1.00$$

$$= 1$$

As the correlation coefficient between the two variables is +1, so the two variables are perfectly positive correlated.