

Chapter 6: Correlation

NCERT · Class 11 · Statistics

Textbook exercises

Question 1

The unit of correlation coefficient between height in feet and weight in kgs is (i) kg/feet (ii) percentage (iii) non-existent

Solution

ANSWER **(iii) non-existent.** The correlation coefficient (r) is a pure number with no unit, because it is a ratio in which the units of the two variables cancel out. So r between height in feet and weight in kg has no unit at all.

Question 2

The range of simple correlation coefficient is (i) 0 to infinity (ii) minus one to plus one (iii) minus infinity to infinity

Solution

ANSWER **(ii) minus one to plus one.** The value of r always lies between -1 and $+1$, i.e. $-1 \leq r \leq 1$. If a calculation gives a value outside this range, it indicates an error.

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

Solution

ANSWER **(i) When Y increases X increases.** A positive r means the two variables move in the same direction — both rise together and both fall together.

Question 4

If $r_{xy} = 0$ the variable X and Y are (i) linearly related (ii) not linearly related (iii) independent

Solution

ANSWER **(ii) not linearly related.** $r = 0$ means there is no *linear* relation between X and Y. It does not necessarily mean they are independent, because a non-linear relationship may still exist.

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

Solution

ANSWER **(iii) Scatter diagram.** A scatter diagram can show any type of relationship — linear or non-linear — because it simply plots the points; the two coefficients measure only linear (or rank-linear) relationships.

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

Solution

ANSWER **(i) more accurate than rank correlation coefficient.** When data are precisely measured, Karl Pearson's coefficient uses all the information in the data, whereas rank correlation uses only the ranks and so loses some information.

Question 7

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

Solution

ANSWER The correlation coefficient r is preferred to covariance because covariance depends on the units of measurement of the two variables, so its value can be very large or very small simply because of the units chosen, and it has no fixed limits. The coefficient r , on the other hand, is a **pure number** with no unit; it is obtained by dividing the covariance by the product of the two standard deviations, so the units cancel out. Because r always lies between -1 and $+1$, two correlation coefficients can be directly compared to judge which relationship is stronger. Covariance cannot be compared in this way. Hence r is a far more convenient and meaningful measure of the degree of association.

Question 8

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

Solution

ANSWER No. The value of the correlation coefficient can never lie outside the range -1 to $+1$, whatever the type of data. By its very construction (covariance divided by the product of the standard deviations), r is mathematically bound by $-1 \leq r \leq 1$. If, in any exercise, you obtain a value such as 1.2 or -1.5 , it does not mean the data are unusual — it means there is a **mistake in the calculation**, and the working must be checked.

Question 9

Does correlation imply causation?

Solution

ANSWER No, correlation does not imply causation. Correlation only measures **covariation** — the fact that two variables tend to change together — not that one causes the other. A high correlation may arise for several reasons: it may be a genuine cause-and-effect relation, a pure coincidence (e.g. the arrival of migratory birds and the local birth rate), or the effect of a third variable on both (e.g. rising temperature causes both brisk ice-cream sales and more deaths by drowning, so these two appear correlated without one causing the other). Therefore correlation must never be interpreted as cause and effect.

Question 10

When is rank correlation more precise than simple correlation coefficient?

Solution

ANSWER Rank correlation is more useful (and effectively more precise) than the simple correlation coefficient in the following situations: (i) when the variables cannot be measured numerically but can only be ranked — such as honesty, beauty or intelligence; (ii) when the data contain **extreme values (outliers)**, because rank correlation is not affected by them while Karl Pearson's r is distorted; and (iii) when the relationship is non-linear but its direction is clear. In such cases ranks give a more reliable picture of association.

Question 11

Does zero correlation mean independence?

Solution

ANSWER No. Zero correlation ($r = 0$) means only that there is **no linear relationship** between the two variables. It does not necessarily mean that the variables are independent. Two variables can have $r = 0$ and yet be related in a non-linear way. For example, for $X = -3, -2, -1, 1, 2, 3$ and $Y = 9, 4, 1, 1, 4, 9$ (where $Y = X^2$), the correlation coefficient is 0, but Y clearly depends on X . So zero correlation rules out a linear relation, not every relation.

Question 12

Can simple correlation coefficient measure any type of relationship?

Solution

ANSWER No. The simple (Karl Pearson's) correlation coefficient can measure only a **linear** relationship between two variables — one that can be represented by a straight line. If the true relationship between the variables is non-linear (curved), Karl Pearson's coefficient is misleading and should not be used. To examine such relationships, a scatter diagram should be drawn first. Hence the simple correlation coefficient cannot measure every type of relationship.

Question 13

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

Solution

ANSWER This is a data-collection activity, so your figures will vary. Record the daily price (Rs/kg) of any two vegetables over seven days, then apply Karl Pearson's formula $r = \frac{[N\sum XY - (\sum X)(\sum Y)]}{\sqrt{[(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)]}}$. **Model interpretation:** if the prices of two vegetables rise and fall together over the week (for example because of common weather or transport costs), r will be positive; if one rises while the other falls, r will be negative; and if there is no clear pattern, r will be close to zero. State which pair you found and whether the relationship was strong (r near ± 1) or weak (r near 0). (Use your own collected data.)

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.

Solution

ANSWER This too is an activity with your own data. Take X = a student's height and Y = that student's benchmate's height for, say, ten pairs, and compute r using Karl Pearson's formula. **Model interpretation:** heights of randomly seated benchmates are usually not related, so r will normally be close to zero, showing no linear relationship. If, however, students were seated by height, you might find a positive r . Report your value and interpret its sign and strength accordingly. (*Use your own measured data.*)

Question 15

List some variables where accurate measurement is difficult.

Solution

ANSWER Variables that are qualitative attributes are difficult to measure accurately and can usually only be ranked. Examples include: intelligence, honesty, beauty, physical appearance, bravery, kindness, leadership ability, friendliness, efficiency of workers, employee morale and consumer satisfaction. For such variables, Spearman's rank correlation is used instead of Karl Pearson's coefficient.

Question 16

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

Solution

ANSWER $r = 1$: perfect positive correlation. The two variables move in exactly the same direction in a fixed proportion, and on a scatter diagram all the points lie on a straight upward-sloping line. $r = -1$: perfect negative correlation. The two variables move in exactly opposite directions in a fixed proportion, and all points lie on a straight downward-sloping line. $r = 0$: no linear correlation. There is no linear relationship between the variables; the points are scattered with no upward or downward straight-line trend (though a non-linear relation may still exist).

Question 17

Why does rank correlation coefficient differ from Pearsonian correlation coefficient?

Solution

ANSWER The rank correlation coefficient differs from the Pearsonian coefficient because the two are computed from different information. Karl Pearson's coefficient uses the **actual measured values** of the variables and therefore all the information in the data, whereas Spearman's coefficient uses only the **ranks** of the items. Because ranking discards the exact magnitudes (and the gaps between consecutive values are usually not equal), rank correlation generally gives a less accurate value, and in general $r_s \leq r$. The two would give identical results only if the first differences of the ordered values were constant. Also, rank correlation is unaffected by extreme values, while Pearson's r is, so the two diverge whenever outliers are present.

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

Solution

ANSWER We use Karl Pearson's direct (raw-score) formula with $N = 8$. First build the working table.

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

Numerator: $N\Sigma XY - (\Sigma X)(\Sigma Y) = 8 \times 36118 - 534 \times 540 = 288944 - 288360 = 584$. **First bracket:**

$N\Sigma X^2 - (\Sigma X)^2 = 8 \times 35788 - (534)^2 = 286304 - 285156 = 1148$. **Second bracket:** $N\Sigma Y^2 - (\Sigma Y)^2 = 8 \times 36644 - (540)^2 = 293152 - 291600 = 1552$. **Denominator:** $\sqrt{(1148 \times 1552)} = \sqrt{1781696} \approx 1334.80$. **$r =$**

$584 \div 1334.80 \approx 0.44$. The heights of fathers and sons are **positively correlated**, but the relationship is only moderate (the value is well short of +1). *Note:* some older NCERT printings show the answer as $r = 0.603$. Re-working the given data carefully by both the direct method and the deviation method ($\bar{X} = 66.75$, $\bar{Y} = 67.5$) gives the verified value $r \approx 0.44$, so the printed 0.603 is a textbook misprint for this data set.

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

Solution

ANSWER Here $N = 6$. Notice $\Sigma X = 0$, which simplifies the direct formula. Build the table.

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

Numerator: $N\Sigma XY - (\Sigma X)(\Sigma Y) = 6 \times 0 - 0 \times 28 = 0$. Since the numerator is 0, **$r = 0$** . **Comment:** the correlation coefficient is zero, so there is **no linear relationship** between X and Y. However, the data are not independent — clearly $Y = X^2$, a perfect *non-linear* (parabolic) relation. This shows that $r = 0$ rules out only a linear relation, not every relation.

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

Solution

ANSWER Here $N = 6$. Build the working table.

X	Y	X^2	Y^2	XY
1	2	1	4	2
3	6	9	36	18
4	8	16	64	32
5	10	25	100	50
7	14	49	196	98
8	16	64	256	128
$\Sigma X = 28$	$\Sigma Y = 56$	$\Sigma X^2 = 164$	$\Sigma Y^2 = 656$	$\Sigma XY = 328$

Numerator: $N\Sigma XY - (\Sigma X)(\Sigma Y) = 6 \times 328 - 28 \times 56 = 1968 - 1568 = 400$. **First bracket:** $N\Sigma X^2 - (\Sigma X)^2 = 6 \times 164 - (28)^2 = 984 - 784 = 200$. **Second bracket:** $N\Sigma Y^2 - (\Sigma Y)^2 = 6 \times 656 - (56)^2 = 3936 - 3136 = 800$. **Denominator:** $\sqrt{(200 \times 800)} = \sqrt{160000} = 400$. **r = $400 \div 400 = 1$.** **Comment:** $r = +1$ shows **perfect positive correlation**. X and Y move in exactly the same direction in a fixed proportion (here $Y = 2X$); on a scatter diagram all points lie on a single upward-sloping straight line.