FREE EDITION · NOTES + 3 SAMPLE MCQS

CUET · ECONOMICS · CLASS XI · CODE 309

Correlation

CUET unit: Statistics for Economics

By **UniDrill** · NCERT-grounded study material

📌 Snapshot

- **Correlation analysis** is the statistical study of the direction and intensity of relationship between two variables — moving beyond single-variable summary measures studied earlier.
- Three tools measure correlation — the **scatter diagram** (visual), **Karl Pearson's coefficient r** (numerical, for cardinal data), and **Spearman's rank correlation r?** (for ranked or qualitative data).
- A central caution: **correlation measures covariation, NOT causation**; a third variable, coincidence, or non-linear relations can mislead interpretation.
- CUET tests definitions, **properties of r**, **range $-1 \leqslant r \leqslant +1$**, formula-recognition, distinction between Pearson and Spearman methods, and conceptual traps (**zero correlation ≠ independence**; correlation ≠ causation).
- This chapter is the bridge between descriptive statistics (central tendency, dispersion) and Index Numbers (kest108), since both require pairwise data analysis.

📖 Detailed Notes

2.1 Core concepts

- **Correlation analysis** examines whether two variables are related, whether they move together, the direction of movement and the strength of the relationship (NCERT §1, pp. 74–75). It looks at pairs of observations like price-quantity, height-weight, income-consumption.
- **Types of underlying relationships**: cause-and-effect (low rainfall causing low agricultural productivity), pure coincidence (arrival of migratory birds and birth rates), and spurious — driven by a hidden third variable (ice-cream sales and drowning deaths, both driven by temperature) (NCERT §2, p. 75).
- Correlation measures **covariation, not causation**; the presence of correlation only means that when one variable changes, the other changes in the same or opposite direction in a definite way (NCERT §2, p. 75).
- **Positive vs negative correlation**: positive — variables move in the same direction (income and consumption; ice-cream sale and temperature); negative — variables

move in opposite directions (price of apples and demand for apples) (NCERT §2, p. 76).

- For simplicity, correlation is assumed to be **linear** — i.e., relative movement can be represented by a straight line on a graph (NCERT §2, p. 76). Non-linear relations are real but not measured by Pearson's r.
- **Three techniques** are used: scatter diagrams, Karl Pearson's coefficient of correlation, and Spearman's rank correlation (NCERT §3, p. 76).
- A **scatter diagram** visually presents the nature of association without giving any specific numerical value; closeness and direction of plotted points indicate strength and type of correlation (NCERT §3, p. 76). Figures 6.1–6.5 illustrate positive, negative, no, perfect positive and perfect negative correlation; Figures 6.6–6.7 show non-linear relations.
- **Karl Pearson's coefficient** (also called product moment correlation coefficient) gives a precise numerical value of the degree of **linear** relationship between X and Y; it must NOT be used when the relation is non-linear (NCERT §3, p. 77).
- **Formulas for r** (NCERT §3, p. 79):
- **$r = \Sigma xy \div (N \cdot \sigma x \cdot \sigma y)$** — using covariance and standard deviations.
- **$r = \Sigma(X - \bar{X})(Y - \bar{Y}) \div \sqrt{}[\Sigma(X - \bar{X})^2 \cdot \Sigma(Y - \bar{Y})^2]$** — direct deviation form.
- Actual-values form using ΣXY, ΣX^2, ΣY^2.
- **Properties of r** (NCERT §3, pp. 79–80):
- r has **no unit**; it is a **pure number**.
- **Negative r** indicates inverse relation; **positive r** indicates same-direction movement.
- **r lies between −1 and +1**; a value outside this range indicates calculation error.
- r is **unaffected by change of origin and change of scale** (basis of step-deviation method).
- **r = 0** means no linear relation, but non-linear relation may still exist (zero correlation ≠ independence).
- **r = ±1** indicates perfect linear correlation; values near ±1 are "high"; values near 0 are "weak".
- **Step-deviation method**: transform variables as **$U = (X - A) \div B$** and **$V = (Y - C) \div D$**, where A, C are assumed means and B, D are common factors of the same sign; then **rUV = rXY** (NCERT §3, pp. 80, 82–83). This makes computation easier when raw values are large.
- **Spearman's rank correlation** was developed by C.E. Spearman. It is used when variables cannot be precisely measured (beauty, honesty), when only ranks are available, when relations are non-linear in direction-defined ways, or when data contain extreme values (NCERT §3, pp. 83–84).
- **Spearman's formula**: **r⍰ = $1 - [6\ \Sigma D^2 \div (n^3 - n)]$**, where D is the difference in ranks and n the number of observations (NCERT §3, p. 84).

- **Correction for ties**: when ranks are repeated, a correction factor **$(m^3 - m) \div 12$** is added for each tied group inside the bracket of the formula's numerator (NCERT §3, p. 86).
- **Properties of r_s**: it lies between −1 and +1; generally **$r_s \leqslant r$** because some information is lost when individual values are replaced by ranks; when first differences are constant, r and r_s are identical (NCERT §3, p. 84).
- **Conclusion**: the scatter diagram is the only one of the three tools **not confined to linear relations**; Pearson and Spearman both measure linear relationship; none implies causation (NCERT §4, p. 87).
- **Why study correlation**: everyday questions — does demand really fall when price rises? does smoking really raise the risk of cancer? — require a tool to measure pairwise variation rather than single-variable summary (NCERT §1, pp. 74–75).
- **Two-variable framing**: correlation always involves **paired observations** (X_i, Y_i) on the same unit (the same household, the same year, the same firm). Without pairing the data lose meaning — a common CUET trap is to give two unrelated lists and ask whether r can be computed (it cannot) (NCERT §1, p. 75).
- **Spurious correlation example (NCERT)**: number of storks counted in a Danish village and number of human births in the same village rose together for years — pure coincidence reflecting common demographic trends. NCERT uses such examples to warn against reading causation into correlation (NCERT §2, p. 75).
- **Negative-correlation classic examples**: price and quantity demanded (Law of Demand), study hours and exam errors, alcohol consumption and motor coordination — each shows $X\uparrow \Rightarrow Y\downarrow$ in a roughly linear fashion (NCERT §2, p. 76).
- **Positive-correlation classic examples**: height and weight of children, household income and expenditure, advertisement and sales, temperature and ice-cream demand — each shows $X\uparrow \Rightarrow Y\uparrow$ (NCERT §2, p. 76).
- **Why linearity is assumed**: the algebra of Pearson's r (variance, covariance, square root) implicitly fits a **best straight line** through the scatter; a curved relation would give a misleadingly low r. NCERT cautions that one must first sketch the scatter before applying r (NCERT §2, p. 76; §3, p. 77).
- **Three reasons to prefer r_s**: (i) attributes cannot be measured numerically (beauty, honesty) — only ranked; (ii) extreme values would distort Pearson's r — but ranking caps the influence of any single observation at ±1; (iii) only ranks are reported in the data (e.g., contest standings) and original scores are unavailable (NCERT §3, pp. 83–84).
- **Why $r_s \leqslant r$ generally**: converting raw numbers to ranks throws away magnitude information; what remains is only ordinal information. Pearson's r exploits magnitudes, so for well-behaved cardinal data Pearson's r captures slightly more information and is at least as large as Spearman's r_s (NCERT §3, p. 84).

- **Step-deviation legality**: r is **unchanged** by U = (X − A)/B, V = (Y − C)/D so long as B and D are of the same sign. If B and D are of opposite signs, the sign of r flips — students often miss this subtlety (NCERT §3, p. 80).
- **r interpretation bands** (informal): $|r| < 0.3$ — weak; $0.3 \leqslant |r| < 0.7$ — moderate; $|r| \geqslant 0.7$ — strong; $|r| = 1$ — perfect. NCERT does not codify these cut-offs but the bands are widely used in CUET context items.
- **Covariance and units**: covariance $\Sigma(X - \bar{X})(Y - \bar{Y})/N$ carries units (e.g., kg·cm for weight-height) — that is exactly why dividing by $\sigma x \cdot \sigma y$ in Pearson's formula makes r dimensionless. Without the standardisation, covariance values across different unit systems cannot be compared (NCERT §3, p. 79).
- **Why r^2 is useful (extension)**: r^2 (the coefficient of determination, implicit in NCERT) gives the proportion of total variation in Y explained by linear movement in X. An r of 0.8 means $r^2 = 0.64$, i.e., 64% of Y's variation is linearly accounted for by X — a more interpretable number than r itself.
- **Tied-rank logic**: when m observations tie at, say, ranks 7, 8, 9, the average rank 8 is assigned to all three. The correction factor $(m^3 - m)/12$ added inside Spearman's bracket — once for each tied group — compensates for the deflated ΣD^2 that results from artificial ties (NCERT §3, p. 86).
- **Perfect-positive vs perfect-negative diagrams**: in Fig. 6.4 every point lies on an upward line at slope > 0 (r = +1); in Fig. 6.5 every point lies on a downward line at slope < 0 (r = −1). The numerical value of the slope is **not** the same as r — slope depends on units, r does not. CUET sometimes tests this slope-vs-r distinction.
- **Non-linear examples that defeat r**: a U-shaped relation (e.g., income vs age) or an inverted-U relation (e.g., productivity vs hours of sleep) can have $r \approx 0$ even though the variables are strongly related — illustrating why $r = 0 \neq$ independence (NCERT §3, p. 80; Fig. 6.6, 6.7).

2.2 Definitions to memorise

Term	Definition	Page
Correlation	Statistical study of the direction and intensity of relationship between two variables	75
Positive correlation	Variables move in the same direction ($X\uparrow \Rightarrow Y\uparrow$)	76
Negative correlation	Variables move in opposite directions ($X\uparrow \Rightarrow Y\downarrow$)	76
Linear relationship	Relationship representable by a straight line on graph paper	76, 77
Non-linear relationship	Relationship that cannot be described by a single straight line	78
Scatter diagram	Graph plotting paired values of two variables to visually examine the form of relationship	76

Term	Definition	Page
Karl Pearson's r	Product moment correlation coefficient measuring numerical degree of linear relation	77
Covariance	$Cov(X, Y) = \Sigma (X - \bar{X})(Y - \bar{Y})/N$; its sign determines the sign of r	79
Attribute	Variable that cannot be numerically measured (intelligence, honesty, beauty)	77
Step-deviation method	Calculation shortcut using $U = (X - A)/B$, $V = (Y - C)/D$ since $r_{UV} = r_{XY}$	80, 82
Spearman's r?	Rank correlation coefficient: $1 - 6\Sigma D^2/(n^3 - n)$ using ranks instead of raw values	84
Perfect correlation	$r = +1$ or $r = -1$; exact linear relation with all points on a line	80
Tied ranks	Equal ranks awarded to observations with identical values	86
Correction factor for ties	$(m^3 - m)/12$ added for each tied group in Spearman's formula	86
Causation	A causes change in B — distinct from mere co-movement	75
Spurious correlation	Correlation arising due to a third variable, not direct linkage	75
Pure number	A quantity without measurement units	79
Independence	No statistical relation of any form — stronger than $r = 0$	80
Change of origin	Subtracting a constant from each value	80
Change of scale	Dividing each value by a constant	80
Product-moment correlation	Karl Pearson's r; another name emphasising its formula	77
Direction of correlation	Sign of r (positive or negative)	76
Intensity of correlation	Magnitude (closeness to 0 or ±1)	75
Linear scatter	Points clustered around a straight line in a scatter diagram	78
Curvilinear scatter	Points clustered around a curve, indicating non-linear relation	78
Coefficient of determination (r^2)	Square of the correlation coefficient — not formally introduced in NCERT but a natural extension	79

2.3 Diagrams / processes to remember

- **Fig. 6.1 — Positive Correlation**: points scattered around an upward-rising line (p. 78).
- **Fig. 6.2 — Negative Correlation**: points scattered around a downward-sloping line (p. 78).
- **Fig. 6.3 — No Correlation**: no rising or falling pattern; random scatter (p. 78).

- **Fig. 6.4 — Perfect Positive Correlation**: all points lie ON an upward line (p. 78).
- **Fig. 6.5 — Perfect Negative Correlation**: all points lie ON a downward line (p. 78).
- **Fig. 6.6 — Positive non-linear relation** & **Fig. 6.7 — Negative non-linear relation**: curved patterns; Pearson's r should NOT be used here (p. 78).
- **Table 6.1**: worked example computing r = 0.644 between years of schooling of farmers and annual yield per acre (p. 81).
- **Table 6.3**: step-deviation example yielding r = 0.98 between price index and money supply (p. 83).
- **Example 5**: worked Spearman calculation with repeated ranks (Y = 50 at ranks 9, 10, 11 averaged to 10; $(m^3 – m)/12$ correction applied), giving r⍰ = 0.30 (pp. 86–87).
- **Correlation decision flow**: data type (cardinal vs ordinal/attribute) → if cardinal and linear use Pearson's r; if ordinal or non-linear monotonic use Spearman's r⍰; always sketch a scatter diagram first.
- **Worked Pearson's r (small example)**: take 5 pairs — (X, Y) = (1, 2), (2, 4), (3, 5), (4, 4), (5, 5). Means $\bar{X}$ = 3, $\bar{Y}$ = 4. Deviations x = X – $\bar{X}$: –2, –1, 0, 1, 2; y = Y – $\bar{Y}$: –2, 0, 1, 0, 1. xy products: 4, 0, 0, 0, 2 → Σxy = 6. x^2: 4, 1, 0, 1, 4 → Σx^2 = 10. y^2: 4, 0, 1, 0, 1 → Σy^2 = 6. r = 6 / √(10 × 6) = 6/√60 = 6/7.746 ≈ 0.775. Interpretation: strong positive linear correlation between X and Y.
- **Worked Spearman's r⍰ (no ties)**: ranks of two judges for 5 contestants — Judge1: 1, 2, 3, 4, 5; Judge2: 2, 1, 4, 3, 5. D = R1 – R2: –1, 1, –1, 1, 0; D^2: 1, 1, 1, 1, 0; ΣD^2 = 4. r⍰ = 1 – [6 × 4 / (5^3 – 5)] = 1 – [24/120] = 1 – 0.2 = 0.8. Strong agreement between the two judges.
- **Worked Spearman with ties**: Y-values 50, 60, 50, 70, 50, 80 → ranks: three 50s tie at ranks 1, 2, 3, averaged to (1+2+3)/3 = 2; so awarded ranks are 2, 4, 2, 5, 2, 6. Correction factor for one tied group of size m=3 is (3^3 – 3)/12 = (27 – 3)/12 = 2. The factor 2 is added inside Spearman's bracket numerator before dividing — illustrating mechanically how repeated-rank cases differ from the no-tie case (NCERT §3, p. 86 logic).
- **Scatter diagram reading drill**: in a 10-point scatter that slopes upward and clusters tightly around an imaginary line, r is high positive (e.g., 0.9); if the same 10 points are scattered widely with a faint upward tendency, r is low positive (e.g., 0.3); if they form a random cloud with no slope, r ≈ 0. The visual feel of "tightness" is the qualitative analogue of |r|, and "tilt" is the analogue of sign(r) (NCERT §3, p. 78 figures).

2.5 Key formulas

Formula	Meaning	NCERT page
r = Σxy ÷ (N · σx · σy)	Pearson's r using covariance and SDs	79

Formula	Meaning	NCERT page
$r = \Sigma(X-\bar{X})(Y-\bar{Y}) \div \sqrt{[\Sigma(X-\bar{X})^2 \cdot \Sigma(Y-\bar{Y})^2]}$	Direct deviation form	79
U = (X−A)/B; V = (Y−C)/D	Step-deviation transformation	80
rUV = rXY	r is unaffected by change of origin and scale	80
r⍰ = $1 - [6\ \Sigma D^2 \div (n^3 - n)]$	Spearman's rank correlation	84
Tie correction = $(m^3 - m)/12$	Added for each tied group	86
Range of r and r⍰	$-1 \leq$ r, r⍰ $\leq +1$	80, 84
r = 0 ⇒ no linear relation	But non-linear relation may exist	80

2.4 Common confusions / NTA trap points

- **Correlation vs causation**: r measures covariation only — high r does not prove cause-and-effect.
- **Zero correlation is NOT independence**: r = 0 means no LINEAR relation, but a non-linear relation may still exist.
- **Unit of r**: r is a **pure number** — has no unit (not kg/feet or %).
- **Range of r**: strictly $-1 \leq r \leq 1$. A value outside this range means calculation error.
- **Pearson vs Spearman applicability**: Pearson's r is valid only for linear relations between precisely-measured variables; for qualitative attributes (honesty, beauty) or extreme values, use Spearman's.
- **Scatter diagram is the only tool that works for any (including non-linear) relationship** — both r and r⍰ measure only linear relationships.
- **Repeated ranks need a correction factor $(m^3 - m)/12$** for each tied group.
- **r is unaffected by change of origin and scale** — basis of the step-deviation method.
- **Generally r⍰ ⩽ r** because rank reduction loses information.
- **Sign of r matches the sign of Cov(X, Y)** — both denominators are positive.
- **Perfect correlation (±1) means all points on a line**, not "near" a line.
- **Spearman's formula uses $n^3 - n$ in the denominator**, not n^2 or n + 1.

🎯 Practice MCQs

Q1. The unit of correlation coefficient between height in feet and weight in kilograms is:

A. kg/feet

B. feet/kg

C. percentage

D. non-existent (pure number)

Q2. The range within which the simple correlation coefficient r must lie is:

A. 0 to ∞

B. $-\infty$ to $+\infty$

C. $-1 \leqslant r \leqslant +1$

D. $0 \leqslant r \leqslant 1$

Q3. Which of the following can examine ANY type of relationship between two variables (including non-linear)?

A. Karl Pearson's coefficient of correlation

B. Spearman's rank correlation

C. Scatter diagram

D. Covariance

📊 PYQ Alignment

Correlation is one of the most frequently tested chapters from the Statistics for Economics unit in CUET, typically yielding 5–7 MCQs per year. Questions emphasise (a) properties and range of r, (b) the unit-less nature of r, (c) the distinction between Pearson and Spearman methods, (d) the correlation-causation distinction, and (e) the meaning of r

= 0 (no linear relation, not independence). Numerical computation is rare in CUET; conceptual statement-based and assertion–reason items dominate. See previous CUET PYQs on this chapter.