

CUET · ECONOMICS · CLASS XI · CODE 309

Organisation of Data

CUET unit: Statistics for Economics

By UniDrill · NCERT-grounded study material

WWW.UNIDRILL.IN

UniDrill

Snapshot

- Raw, unorganised data are arranged into meaningful classes so that further statistical analysis becomes possible.
- **Classification** may be chronological, spatial, qualitative or quantitative; **variables** are continuous or discrete; data are organised in a **frequency distribution table**.
- Technical terms — class limits, class interval, class mark, class frequency, range, tally marks, inclusive vs exclusive method — must be defined precisely.
- A **frequency array** is for discrete data; a **frequency distribution** is for continuous data; a **bivariate frequency distribution** records two variables together.
- CUET regularly tests definitions, formulas (especially class mid-point and range), and the inclusive/exclusive distinction, making this a high-yield chapter.
- This data structure underlies all later chapters (measures of central tendency, dispersion, correlation, index numbers).

Detailed Notes

2.1 Core concepts

- **Purpose of classification:** raw data are highly disorganised and cumbersome; classification is the arranging or organising of things into groups or classes based on some criteria, bringing order and making statistical analysis possible (NCERT §1, pp. 22–24).
- **Raw data:** unclassified observations on a variable (e.g., marks of 100 students in Table 3.1; monthly food expenditure of 50 households in Table 3.2) that do not yield to statistical methods until summarised (NCERT §2, pp. 23–24).
- **Chronological classification:** data classified with reference to time (years, quarters, months, weeks) in ascending or descending order; e.g., population of India from 1951–2011, which constitutes a **Time Series** (NCERT §3, p. 25, Example 1).
- **Spatial classification:** data classified with reference to geographical locations such as countries, states, cities or districts; e.g., yield of wheat in different countries in 2013 (NCERT §3, p. 26, Example 2).

- **Qualitative classification:** based on characteristics (attributes) that cannot be measured — e.g., nationality, literacy, religion, gender, marital status — classified by presence or absence of the attribute (NCERT §3, p. 26, Example 3).
- **Quantitative classification:** based on measurable characteristics such as height, weight, age, income, marks — grouped into classes (NCERT §3, pp. 26–27, Example 4).
- **Continuous variable:** can take any numerical value — integral, fractional or irrational (e.g., height, weight, time, distance); values can be broken into infinite gradations (NCERT §4, p. 27).
- **Discrete variable:** takes only certain values; changes by finite "jumps" with no intermediate values (e.g., number of students in a class) (NCERT §4, pp. 27–28). A discrete variable can still take fractional values like $1/8$ or $1/16$ — what matters is the jump, not whether the value is a whole number.
- **Frequency distribution:** a comprehensive way to classify raw data of a quantitative variable, showing how different values are distributed across classes with their class frequencies (NCERT §5, p. 28).
- **Class limits:** the two ends of a class; the lowest value is the Lower Class Limit and the highest the Upper Class Limit (NCERT §5, p. 28).
- **Class interval / class width:** difference between the upper class limit and the lower class limit (e.g., 10 for class 60–70) (NCERT §5, p. 28).
- **Class mid-point / class mark:** middle value of a class = $(\text{Upper Class Limit} + \text{Lower Class Limit}) \div 2$; used to represent the class once data are grouped (NCERT §5, p. 29). All subsequent statistical calculations (mean, variance) are based on class marks.
- **Number of classes:** usually between 6 and 15; for equal class intervals, number of classes = $\text{range} \div \text{class interval}$ (NCERT §5, pp. 29–30).
- **Range:** difference between the largest and the smallest values of the variable (NCERT §5, p. 30).
- **Inclusive method:** values equal to both the lower and upper limits of a class are included in that class (e.g., 0–10, 11–20, 21–30...) (NCERT §5, pp. 30–31).
- **Exclusive method:** an item equal to either the upper or the lower class limit is excluded from that class (e.g., 0–10, 10–20, 20–30...) (NCERT §5, p. 31).
- **Adjustment in class intervals:** to restore continuity in an inclusive distribution, subtract 0.5 from each lower limit and add 0.5 to each upper limit (e.g., 800–899 becomes 799.5–899.5) (NCERT §5, p. 32, Tables 3.4 & 3.5).
- **Tally marks:** class frequency is counted by putting a tally (/) against the class for each observation; tallies are grouped in fives for ease of counting (NCERT §5, pp. 32–33, Table 3.6).

- **Loss of information:** once data are grouped, individual observations are replaced by the class mark in calculations, leading to loss of information; but this is more than offset by the gain in comprehension (NCERT §5, pp. 33–34).
- **Unequal class intervals:** used when a large number of values are concentrated in a small portion of the range, or when the range itself is very wide (e.g., income) (NCERT §5, pp. 29 & 34, Table 3.7).
- **Frequency array:** classification for a discrete variable, where frequencies correspond to each integral value of the variable (e.g., size of households in Table 3.8) (NCERT §5, p. 35).
- **Bivariate frequency distribution:** frequency distribution of two variables together (e.g., sales and advertisement expenditure of 20 firms in Table 3.9); used later in correlation (NCERT §6, p. 36).
- **Why organise data at all:** raw collected data are like a stack of unsorted laundry — useless until folded and put in drawers. Classification is the folding-and-sorting step that prepares data for tabulation, diagrammatic presentation and statistical computation (NCERT §1, p. 22).
- **Variable vs attribute:** a **variable** is a measurable characteristic (income, weight); an **attribute** is a qualitative characteristic (literacy status, religion). Attributes are classified by counts; variables can be classified by counts AND by magnitude (NCERT §3, p. 26). CUET often tests this distinction with examples like "gender" (attribute) vs "height" (variable).
- **Class limit terminology drill:** in the class "40–50", 40 is the **lower class limit** (LCL), 50 is the **upper class limit** (UCL), $50 - 40 = 10$ is the **class interval (h)**, $(40+50)/2 = 45$ is the **class mark (m)** (NCERT §5, p. 28). All four terms can appear in a single MCQ.
- **Equal vs unequal class intervals:** NCERT recommends equal widths when the data are evenly spread, but allows unequal widths when (i) values are concentrated in a narrow range of the variable, or (ii) the overall range is so wide that equal classes would produce many empty cells (NCERT §5, pp. 29 & 34). Income distributions typically use unequal classes (e.g., ₹0–10k, ₹10–25k, ₹25–50k, ₹50k+).
- **Continuity correction logic:** in the inclusive series 0–9, 10–19, 20–29, the value 9.5 has no home — it falls in the gap. Subtracting 0.5 from each lower limit and adding 0.5 to each upper limit produces 0.5–9.5, 9.5–19.5, 19.5–29.5... closing every gap so that every possible value sits in exactly one class (NCERT §5, p. 32).
- **Open-end classes:** NCERT briefly mentions classes such as "below 100" or "1000 and above" — used at the extremes of an income or expenditure distribution. For computational purposes, the missing limit is usually assumed by symmetry with the adjacent class (NCERT §5, p. 34).
- **Cumulative frequency** is introduced implicitly: by adding successive class frequencies, one can answer "how many observations are below value X?" — a

stepping stone to the median and percentile calculations in kest105 (NCERT §5, p. 33, implicit).

- **Bivariate setup detail:** in a bivariate frequency table, one variable is laid out along the rows (say, sales) and the other along the columns (say, advertisement expenditure); each cell records the joint frequency. Row and column totals are called **marginal frequencies**, and the grand total equals the total number of observations (NCERT §6, p. 36).
- **Time-series example:** NCERT references India's population growth over 1951–2011 to illustrate chronological classification — the year is the criterion and the population the variable. CUET sometimes asks students to identify whether a given table is time-series or cross-section; the rule is "time-series = many time points, one unit; cross-section = one time point, many units" (NCERT §3, p. 25).
- **Spatial example:** yields of wheat across countries in 2013 (Example 2, p. 26) — the criterion is location, the variable is yield, the time is held constant. A spatial table is a one-time-period geographic cross-section.
- **Loss of information caveat:** once class 60–70 has frequency 8, NCERT warns we no longer know whether those 8 students scored 60, 61, ..., 69 or all clustered at 69 — we lose that detail forever. Statisticians accept this loss because the compactness gain (8 numbers instead of a list of 100) is far larger (NCERT §5, p. 33).
- **Tally-marks five rule:** tallies are bundled into groups of five — four vertical strokes plus a diagonal — because the human eye can count five-bundles much faster than individual strokes; this is purely an ergonomic convention (NCERT §5, p. 33).
- **Discrete fraction example:** shoe sizes in India jump in halves — 7, 7.5, 8, 8.5 — so "shoe size" is a discrete variable that can take half-integer values. The defining feature of discreteness is the gap between admissible values, NOT integer-ness (NCERT §4, p. 28).
- **Number of classes — Sturges' guideline (background):** while NCERT only says "6 to 15 classes", a common rough rule is $k = 1 + 3.322 \log_{10} N$, giving about 7 classes for $N=100$ — consistent with NCERT's range (NCERT §5, p. 29, contextual).

2.2 Definitions to memorise

Term	Definition	Page
Classification	Arranging or organising things into groups or classes based on some criteria	23
Raw data	Unclassified data; observations not arranged in any order	23–24
Chronological classification	Data classified with reference to time	25
Time series	Sequence of observations on a variable taken over successive time periods	25

Term	Definition	Page
Spatial classification	Data classified with reference to geographical locations	25–26
Qualitative classification	Classification on the basis of presence or absence of an attribute	26
Quantitative classification	Classification of data on measurable characteristics into classes	26
Continuous variable	Variable that can take any numerical value, including fractional and irrational	27
Discrete variable	Variable that takes only certain values and changes by finite "jumps"	27–28
Frequency distribution	Classification showing values of a quantitative variable distributed across classes with frequencies	28
Class	A group of values into which observations are grouped	28
Class frequency	Number of values in a particular class	28
Class limits	Two ends (lower and upper) of a class	28
Class interval / class width	Upper class limit – Lower class limit	28
Class mid-point / class mark	$(\text{Upper class limit} + \text{Lower class limit}) \div 2$	29
Range	Largest value – Smallest value of the variable	30
Inclusive method	Both lower and upper class limits are included in that class	30
Exclusive method	An item equal to the upper or lower class limit is excluded from that class	31
Adjustment factor	± 0.5 added/subtracted to restore continuity when converting inclusive to exclusive	32
Tally mark	Stroke (/) used to count frequency, grouped in fives for ease of counting	33
Frequency array	Classification of a discrete variable, with frequency for each integral value	35
Bivariate frequency distribution	Frequency distribution of two variables together	36
Loss of information	Trade-off in classification — individual values are replaced by the class mark	33–34
Equal class intervals	Class width constant across all classes	29
Unequal class intervals	Class widths that differ across classes, used when data are concentrated or range wide	34

2.3 Diagrams / processes to remember

- **Fig. 3.1 (p. 29):** frequency curve — diagrammatic presentation of frequency distribution of marks of 100 students; class marks on X-axis, frequency on Y-axis.
- **Fig. 3.2 (p. 35):** frequency curve of distribution in Table 3.7 (unequal class intervals).
- **Table 3.1 (p. 24):** raw marks of 100 students in mathematics — base data used throughout this topic.
- **Table 3.2 (p. 24):** raw monthly household food expenditure of 50 households.
- **Table 3.3 (p. 29):** lower class limit, upper class limit, and class mark for ten classes 0–10, ..., 90–100.
- **Table 3.4 (p. 31) and Table 3.5 (p. 32):** frequency distribution of incomes of 550 employees — before and after adjustment of class intervals (inclusive → exclusive form with ± 0.5).
- **Table 3.6 (p. 33):** tally marking process for marks of 100 students; tallies grouped as //// with the fifth placed across.
- **Table 3.7 (p. 34):** frequency distribution with unequal class intervals.
- **Table 3.8 (p. 35):** frequency array of household size (discrete variable, values 1 to 8).
- **Table 3.9 (p. 36):** bivariate frequency distribution of sales and advertisement expenditure of 20 firms.
- **Tree diagram (Example 3, p. 26):** population branching first into Male/Female and then into Married/Unmarried — illustration of qualitative classification.
- **Classification flow:** raw data → choose criterion (time/space/attribute/measurement) → if measurable, choose continuous vs discrete → choose number and width of classes → use inclusive or exclusive method → tally → frequency distribution.
- **Worked example — building a frequency table from raw marks:** suppose 20 students score 12, 25, 33, 47, 51, 9, 18, 27, 36, 44, 55, 8, 22, 30, 41, 49, 17, 28, 39, 46. (i) Range = $55 - 8 = 47$. (ii) With class width 10 we need $k = 47/10 \approx 5 \rightarrow$ use 6 classes 0–10, 10–20, 20–30, 30–40, 40–50, 50–60 (exclusive method, upper limit excluded). (iii) Tallying: 0–10 → 2 (8, 9); 10–20 → 3 (12, 17, 18); 20–30 → 4 (22, 25, 27, 28); 30–40 → 3 (30, 33, 36, 39 — wait, that is 4) — actually 4; recount → 0–10:2, 10–20:3, 20–30:4, 30–40:4, 40–50:5 (41, 44, 46, 47, 49), 50–60:2 (51, 55), total = 20 ✓. (iv) Class marks are 5, 15, 25, 35, 45, 55 — these become the X-values for any subsequent mean calculation in *kest105*.
- **Worked example — inclusive-to-exclusive conversion:** take classes 0–9, 10–19, 20–29, 30–39 (inclusive). Apply ± 0.5 correction → $-0.5-9.5$, $9.5-19.5$, $19.5-29.5$, $29.5-39.5$ (exclusive, continuous). The class widths are now 10 each, no gaps, and a value of 9.5 lies exactly on the boundary of class 1 and class 2 — by exclusive-method convention it is placed in the next class (NCERT §5, p. 32 worked-out).

- **Worked example — class mark for unequal classes:** for classes 0–10, 10–25, 25–50, 50–100, the class marks are 5, 17.5, 37.5, 75. Note that for unequal widths, class marks are still mid-points of each class, but they are no longer equally spaced — a fact used when plotting histograms with variable-width rectangles in kest104.
- **Frequency curve sketch (Fig. 3.1, p. 29):** class marks 5, 15, ..., 95 on the X-axis; frequencies (e.g., 1, 8, 6, 7, 21, 23, 19, 6, 5, 4) on the Y-axis; points joined by a smooth curve passing through them. The curve typically rises to a peak (mode region) and tails off on either side — anticipating the unimodal distributions of kest105.

2.5 Key formulas / structural ratios

Formula	Meaning	NCERT page
Class mark $m = (UCL + LCL) \div 2$	Mid-point representing the class	29
Class interval $h = UCL - LCL$	Width of one class	28
Range $R = X_{max} - X_{min}$	Spread of the entire data set	30
Number of classes $k \approx R \div h$	Approximate number of equal-width classes	29–30
Adjustment for continuity = ± 0.5	Conversion of inclusive to exclusive form	32
$\sum f_i = N$	Sum of class frequencies equals total observations	33
Frequency relative = $f_i \div N$	Share of total observations falling in a class	35

2.4 Common confusions / NTA trap points

- **Range vs class interval:** range is across the whole data set (largest – smallest); class interval is for a single class (upper – lower limit of that class). Distractors swap them.
- **Inclusive vs exclusive method:** in inclusive, both limits belong to the same class; in exclusive, one limit (usually the upper) is excluded. NTA tests this with statements like "10 is in 10–20" (true in exclusive form when upper is excluded; true in inclusive form when 10–19, 20–29 etc. are used).
- **Class mark formula:** it is the **average** of the two limits, not the difference, product or ratio.
- **Discrete variable can still take fractions:** e.g., 1/8, 1/16, 1/32 — what matters is that it "jumps" with no values in between.
- **Statistical calculations after classification** are based on class marks, NOT on the actual observed values.

- **Frequency array** is for discrete data only; **frequency distribution with class intervals** is for continuous data.
- **Univariate vs bivariate**: a bivariate distribution involves **two variables**, not multiple categories of one variable.
- **Number of classes** is typically between 6 and 15, not "as many as observations".
- **Adjustment factor = 0.5**, not 1 or 1.5.
- **Class frequencies sum to N**, not to 100.
- **Time-series ≠ cross-section**: time-series is over time (chronological); cross-section is across units at a point in time (spatial).
- **Qualitative ≠ quantitative**: qualitative classifies by attribute; quantitative classifies by measurement.

Practice MCQs

Q1. The class midpoint is equal to:


- A. The product of the upper class limit and the lower class limit
- B. The ratio of the upper class limit to the lower class limit
- C. The average of the upper class limit and the lower class limit
- D. The difference between the upper class limit and the lower class limit

Q2. The frequency distribution of two variables together is known as:

- A. Univariate distribution
- B. Bivariate distribution
- C. Multivariate distribution
- D. Frequency array

Q3. Once raw data are grouped into classes, further statistical calculations are based on:

- A. The actual values of the observations
- B. The upper class limits
- C. The lower class limits
- D. The class midpoints

 **9 more MCQs + answer key**
Get UniDrill Pro · ₹199/year · unidrill.in/pricing

PYQ Alignment

This chapter is a CUET staple: definitions (class mark, range, class interval), the inclusive vs exclusive distinction, and the difference between continuous/discrete variables and univariate/bivariate distributions appear almost every year — typically as direct one-line factual MCQs or short statement-based questions. Expect around 6–8 MCQs annually drawn from this chapter, with a particular focus on the formula for class mid-point and on identifying classifications (chronological/spatial/qualitative/quantitative) from short descriptions. See [previous CUET PYQs on this chapter](#).

