

FREE EDITION · NOTES + 3 SAMPLE MCQS

CUET · ECONOMICS · CLASS XI · CODE 309

Collection of Data

CUET unit: Statistics for Economics

By UniDrill · NCERT-grounded study material

WWW.UNIDRILL.IN

UniDrill

Snapshot

- Raw economic facts (data) must be gathered before any statistical analysis can begin.
- Data are either **Primary or Secondary**; collection follows either a **Census or a Sample Survey** approach.
- The three modes of data collection are **Personal Interview, Mailed Questionnaire and Telephone Interview**, each with merits and demerits.
- Sampling may be **Random or Non-Random**; errors are **sampling or non-sampling**; **Random Number Tables** aid random selection.
- Key Indian data-collection agencies (**Census of India, NSS, CSO, RGI, DGCIS, Labour Bureau**) are a frequent CUET fact-recall zone.
- Date grid: 1881 (first Census), 1951 (first post-independence Census), 2011 (last Census, population 121.09 crore), Sarvekshana (NSS journal).

Detailed Notes

2.1 Core concepts

- **Data and variables:** "Data" are values of variables (denoted X, Y, Z); each value is an observation. Production of food grain in India varied from 108 million tonnes (1970-71) to 272 million tonnes (2016-17) — a classic textbook example of how data show variation (NCERT §1, pp. 9–10).
- **Two sources of statistical data:** **Primary data** are collected first-hand by the researcher through an enquiry; **Secondary data** have already been collected, scrutinised and tabulated by some other agency. Data are primary to the source that first collects them, and secondary for all later users (NCERT §2, p. 10).
- **Survey** is the method of gathering information from individuals; the most common instrument is the **questionnaire / interview schedule**, which may be self-administered or administered by an enumerator (NCERT §3, p. 11).
- **Good questionnaire design:** keep it short, easy to understand, arrange questions general-to-specific, make them precise and unambiguous, avoid double negatives, avoid leading questions, and do not suggest alternatives in the question itself (NCERT §3, pp. 11–12).
- **Two types of questions:**

- **Closed-ended / structured** — two-way Yes/No or multiple choice — easy to score but may not capture the true response, which is why an "Any Other" option is often provided.
- **Open-ended / unstructured** — more individualised responses but harder to score and interpret (NCERT §3, p. 12).
- **Three modes of data collection** (NCERT §3, pp. 13–14):
 - **Personal Interview** — face-to-face; highest response rate; allows clarification; but expensive and time-consuming.
 - **Mailing Questionnaire** — least expensive; reaches remote areas; maintains anonymity; no interviewer influence; but low response rate and unsuited to illiterates.
 - **Telephone Interview** — cheap and quick; allows clarification; but limited to phone-owners.
- **Pilot Survey / Pre-testing**: a try-out of the questionnaire on a small group before the actual survey, to detect shortcomings, test clarity, assess enumerator performance and estimate cost and time (NCERT §3, p. 14).
- **Census (Complete Enumeration)** covers every element of the population. The Census of India is carried out every ten years; the last Census was held in 2011, recording India's population at **121.09 crore** (102.87 crore in 2001; 23.83 crore in 1901). The annual growth rate fell from **2.2% (1971-81) to 1.97% (1991-2001) to 1.64% (2001-2011)** (NCERT §4, p. 15).
- **Population / Universe** is the totality of items under study; a **Sample** is a representative subset from which information is obtained. A good sample is smaller than the population yet gives reasonably accurate information at lower cost and shorter time (NCERT §4, p. 15).
- **Random sampling**: every individual unit of the sampling frame has an equal chance of being selected — also called the **lottery method**; modern surveys use computer programs or **Random Number Tables**. **Non-random sampling**: units are chosen on the basis of judgement, convenience, purpose or quota — not every unit has an equal chance (NCERT §4, pp. 16–17).
- **Sampling error** = difference between the sample estimate and the actual population parameter; it can be reduced by taking a larger sample. Example: true mean of 5 farmers' incomes = 600; a sample of two gives estimate 550, so sampling error = 50 (NCERT §5, pp. 17–18).
- **Non-sampling errors** are more serious than sampling errors because they cannot be reduced by enlarging the sample — even a Census can contain them. Types:
 - **Sampling Bias** — some population members had no chance of being included.
 - **Non-Response Errors** — respondent unreachable or refuses.
 - **Errors in Data Acquisition** — incorrect recording, e.g., writing 13 instead of 31 (NCERT §5, p. 18).

- **National data-collection agencies: Census of India, National Sample Survey (NSS), Central Statistics Office (CSO), Registrar General of India (RGI), Directorate General of Commercial Intelligence and Statistics (DGCI), Labour Bureau.** The Census has been conducted every ten years since 1881; the first post-independence Census was in 1951. NSS publishes the quarterly journal **Sarvekshana**; the NSS 60th round (Jan–June 2004) was on morbidity and healthcare; the NSS 68th round (2011-12) was on consumer expenditure (NCERT §6, pp. 18–19).
- **Why we need data at all:** without raw numerical inputs, all the techniques of presentation, central tendency, dispersion and correlation are inert (NCERT §1, p. 9). Data are the raw material of every statistical technique that follows.
- **The investigator-respondent relationship:** NCERT distinguishes the **investigator** (the person/body conducting the enquiry) from the **enumerator** (field worker who collects the data) and from the **respondent / informant** (person from whom information is obtained). Confusing these three roles is a classic CUET trap (NCERT §3, p. 11).
- **Schedule vs questionnaire:** although NCERT uses the two terms loosely, technically a **questionnaire** is filled in by the respondent personally (as in mail surveys) whereas a **schedule** is filled in by the enumerator after questioning the respondent (as in personal interviews). Both are listed by NCERT as instruments of survey (NCERT §3, p. 11).
- **Long form of Census:** the full title is the **Census of India**, conducted under the **Census of India Act 1948** by the Office of the Registrar General and Census Commissioner under the Ministry of Home Affairs — a frequently asked institutional-fact MCQ (NCERT §6, p. 18).
- **Population numbers at a glance:** 1901 — 23.83 crore; 2001 — 102.87 crore; 2011 — 121.09 crore. Decadal growth slowed from 21.5% (1991-2001) to 17.7% (2001-2011), and the annual exponential rate fell from 1.97% to 1.64% over the same window (NCERT §4, p. 15). These figures repeatedly anchor CUET data-interpretation questions.
- **NSS organisational note:** NSS was set up in 1950 and brings out the journal *Sarvekshana* — students should not confuse NSS (the survey organisation, now NSSO) with CSO (which compiles national-income statistics and the Index of Industrial Production) (NCERT §6, pp. 18–19).
- **DGCI** is based at Kolkata and compiles India's external-trade statistics, while the **Labour Bureau** at Chandigarh/Shimla compiles wage, employment and CPI-IW figures — the two agencies most often confused in CUET MCQs because both deal with prices and labour-market data (NCERT §6, p. 19).
- **Sample size vs accuracy:** the NCERT explicitly notes that "as the sample size increases, the sample mean approaches the population mean" — this is the intuitive statement of the Law of Large Numbers used to motivate why larger samples are

better (NCERT §5, p. 18). It also notes that beyond a point, marginal accuracy gains diminish — hence sampling is preferred to a costly census.

- **Mode-of-collection trade-offs in one line each:** personal interview = highest accuracy but highest cost; mail = lowest cost but lowest response rate; phone = quickest but excludes non-phone users — a 3×3 trade-off matrix worth memorising (NCERT §3, pp. 13–14).
- **Random Number Table — practical use:** NCERT explains that to draw, say, 10 households out of 1000 numbered 000–999, the investigator reads three-digit groups from a random number table and selects those whose numbers fall in 000–999, ignoring repeats — a deterministic and reproducible procedure (NCERT §4, p. 16). The technique removes investigator bias.
- **Worked sampling-error illustration (NCERT pp. 17–18):** income data for five farmers — 500, 550, 600, 650, 700. Population mean $\mu = (500+550+600+650+700)/5 = 2750/5 = 600$. A two-farmer sample of 500 and 600 gives sample mean $\bar{x} = (500+600)/2 = 550$. Sampling error = $\mu - \bar{x} = 600 - 550 = 50$, i.e., $(50/600) \times 100 \approx 8.33\%$ of the true mean. A larger sample of four farmers (550, 600, 650, 700) gives $\bar{x} = 625$, error = 25 — confirming that error shrinks as n rises.
- **Sampling versus complete enumeration trade-off:** a Census gives the true value (μ) but at very high cost and time; a sample gives an estimate (\bar{x}) at low cost but with sampling error. The choice depends on the trade-off between accuracy required and resources available (NCERT §4, p. 15).
- **Two NCERT illustrations of non-sampling error:** (i) recording an age of 13 instead of 31 (digit transposition), and (ii) a postal questionnaire returned unanswered because the respondent moved house (non-response) — both errors stand even if every member of the population is contacted (NCERT §5, p. 18).
- **Two-stage and stratified samples** are more sophisticated designs beyond simple random sampling. Remember random \neq unrepresentative; a small random sample may still mis-represent a heterogeneous population (NCERT §4, p. 17 box).
- **Exercise items (NCERT pp. 19–20)** include the difference between Census and Sample, advantages of sampling, types of sampling errors, and design of a questionnaire on adult literacy — a hint to which application questions CUET may pose.

2.2 Definitions to memorise

Term	Definition	Page
Variable	A characteristic whose value varies; represented by X, Y or Z	10
Observation	Each value taken by a variable	10
Primary data	Data collected first-hand by the investigator	10

Term	Definition	Page
Secondary data	Data already collected, scrutinised and tabulated by another agency	10
Survey	Method of gathering information from individuals	11
Questionnaire / Interview schedule	Instrument used to ask questions in a survey	11
Closed-ended question	Two-way (Yes/No) or multiple-choice question with fixed options	12
Open-ended question	Unstructured question allowing free response	12
Leading question	Question phrased to suggest a particular answer; to be avoided	12
Pilot Survey	Pre-testing the questionnaire on a small group before the main survey	14
Census / Complete Enumeration	Survey covering every element of the population	15
Population (Universe)	Totality of items under study	15
Sample	A representative subset of the population from which information is collected	15
Sampling frame	The list of all sampling units of the population	16
Random sampling	Sampling in which every individual unit has an equal chance of being selected (lottery method)	16
Non-random sampling	Sampling based on judgement, convenience, purpose or quota	17
Random Number Tables	Pre-generated tables of random digits used to draw a sample	16
Sampling error	Difference between sample estimate and population parameter	18
Non-sampling error	Error not reducible by enlarging the sample — bias, non-response, recording errors	18
Sampling bias	Some members of the target population have no possibility of being included	18
Non-response error	Respondent cannot be contacted or refuses to respond	18
Errors in data acquisition	Mistakes in recording or transcribing responses	18
Sarvekshana	Quarterly journal of the NSS	18
Census of India	Decennial complete enumeration of India's population, conducted since 1881	15, 18
Pilot study purpose	Detect shortcomings, test clarity, estimate cost and time	14

2.3 Diagrams / processes to remember

- **Table 2.1 — Production of Food Grain in India** (Million Tonnes): seven paired observations of X (year) and Y (output) showing variation across time (p. 10).
- **Comparative table of Advantages and Disadvantages** of the three survey modes (Personal / Mailed / Telephonic) (p. 13).
- **Schematic figure** showing a population of 20 Kuchha + 20 Pucca houses, with a Representative Sample versus a Non-representative Sample drawn from it (p. 17).
- **Worked example of sampling error** with 5 farmers' incomes (500, 550, 600, 650, 700) — population mean 600, sample mean 550, error 50 (pp. 17–18).
- **Data-collection process flow:** define objective → design questionnaire → pilot survey → fieldwork (Census or Sample) → editing → tabulation → analysis.
- **Sampling decision tree:** define population → construct sampling frame → choose random vs non-random → select sample → estimate parameters → quantify sampling error.
- **Worked sampling-error numerical (re-derived):** take the NCERT income data {500, 550, 600, 650, 700}. (i) Population mean $\mu = 600$ ✓. (ii) All possible samples of size $n=2$ number ${}^5C_2 = 10$ — namely (500,550), (500,600), (500,650), (500,700), (550,600), (550,650), (550,700), (600,650), (600,700), (650,700) — with means 525, 550, 575, 600, 575, 600, 625, 625, 650, 675 respectively. The arithmetic mean of these ten sample means is $6000/10 = 600$ — which equals μ , illustrating the **unbiasedness** of the sample mean. (iii) The maximum sampling error in this set is $600 - 525 = 75$ (best case 0 for the (500,700) sample). The exercise demonstrates that the sample mean is an unbiased but noisy estimator, and that the spread of possible errors is what later chapters quantify as **standard error**.
- **Random-number-table walk-through:** to pick 5 households from a sampling frame of 100 numbered 00–99, the enumerator opens the table at any page, reads two-digit groups (say 23, 91, 04, 47, 91, 18); the repeated 91 is skipped, giving the random sample {23, 91, 04, 47, 18}. This deterministic procedure eliminates personal judgement entirely (NCERT §4, p. 16, illustrative).
- **Pilot-survey checklist (mental table):** pilot tests check — (a) questionnaire wording, (b) length of interview, (c) enumerator clarity, (d) appropriateness of response categories, (e) sensitive-question reactions, (f) cost per schedule. Findings feed back into a revised questionnaire before the main field round (NCERT §3, p. 14).
- **Population pyramid mental sketch (linked to Census):** India's 2011 population of 121.09 crore can be visualised as a broad-based pyramid with high youth share, which the Sample-Registration-System (SRS) and NSS surveys complement with annual updates between Census years — together giving a near-continuous picture of population dynamics.

2.5 Key formulas / structural ratios

Formula / Indicator	Meaning	NCERT page
Population mean $\mu = (\sum X_i) \div N$	Average value over all N units of the population	17
Sample mean $\bar{x} = (\sum x_i) \div n$	Average value over n sample units	17
Sampling error = $\mu - \bar{x}$	Gap between population and sample mean	18
Census of India period = 10 years	Frequency of complete enumeration	15
2011 Census population = 121.09 crore	India's population at last Census	15
Annual growth rate 2001-2011 = 1.64%	Inter-censal population growth	15

2.4 Common confusions / NTA trap points

- **Primary vs Secondary** is defined **relative to the source that first collected** the data; the same dataset can be primary to one user and secondary to another.
- "**Lottery method**" is a **type** of random sampling, not non-random sampling.
- **Sampling error decreases with larger sample**, but **non-sampling error does NOT** — even a Census has non-sampling errors.
- Mailed survey is "least expensive" but has "low response rate"; students often misattribute high response rate to mail.
- **Census of India** every 10 years **since 1881**, but the **first post-independence Census was 1951** — both dates testable.
- **2011 Census population = 121.09 crore** (not 102.87 crore — that is 2001).
- **Census is a method, not the agency**: the agency is the **Office of the Registrar General and Census Commissioner (RGI)**.
- **NSS publishes Sarvekshana**, not the CSO.
- **Pilot survey is for pre-testing**, not for the main estimate.
- **Closed-ended vs open-ended**: closed-ended is easier to score, open-ended is harder. Often inverted.
- **Sampling bias is a non-sampling error** — students often classify it under sampling errors.
- **Sample need not be small**; it just needs to be smaller than the population and representative.

Practice MCQs

Q1. Which of the following best describes "Primary data"?


- A. Data collected, tabulated and published by a government agency
- B. Data based on first-hand information collected by the investigator
- C. Data obtained from newspapers and books
- D. Data already processed by another researcher

Q2. The Census of India is conducted once in every:

- A. Five years
- B. Seven years
- C. Ten years
- D. Fifteen years

Q3. Which of the following statements about random sampling is correct?

- A. The investigator selects units based on convenience
- B. Every individual unit of the population has an equal chance of being selected
- C. Only households known to the investigator are chosen
- D. Quota and judgement decide the selection of units

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

PYQ Alignment

Collection of Data is a high-frequency CUET chapter: questions on Primary vs Secondary data, Census vs Sample, types of sampling (random/non-random, lottery method) and the distinction between sampling and non-sampling errors appear almost every year, typically as direct fact-recall MCQs or statement-based "Which is correct?" items. The numerical

sampling-error example and the names of agencies (NSS, RGI, CSO, DGCIS, Census of India) are also recurring NTA pick-points. See [previous CUET PYQs on this chapter](#).

