

1. Meaning of Data in Economics

Data are recorded facts, observations, measurements or information about economic units and activities. Economic units may include individuals, households, firms, farms, banks, markets, industries, governments and entire economies. Economic data can describe quantities, prices, income, expenditure, employment, production, consumption, savings, investment, imports, exports, poverty, education, health and many other aspects of economic life.

For example, the monthly income of a household, the price of rice in a market, the annual production of tea in Assam, the unemployment rate of a state, and India's gross domestic product (GDP) are all examples of economic data. Data may be numerical, such as income of ₹25,000 per month, or categorical, such as employment status (employed, unemployed or outside the labour force).

1.1 Data, Information and Evidence

Data are raw facts, observations or measurements collected about people, households, firms, markets or the economy. By themselves, data may not communicate a clear meaning until they are organised and interpreted. Economic data may include the daily price of rice, the monthly income of a household, the number of workers employed by a firm, or the annual production of tea in Assam.

Information is data that has been processed, organised, classified, summarised or presented in a meaningful context. Information helps people understand what the data show. For example, a list of daily rice prices from several markets is data. A table showing the average price in each market, or a chart comparing prices across markets, converts those data into information. Information answers questions such as “What is the average price?”, “Which market has the highest price?” or “How much did the price change?”

Knowledge is the understanding developed by interpreting information together with experience, subject knowledge, reasoning and an understanding of the relevant context. Knowledge helps people explain why a pattern may have occurred, anticipate possible consequences and make decisions. For example, if information shows that rice prices are consistently higher in a remote market, knowledge of transport costs, road conditions, supply chains and local demand may help explain the difference and guide a decision about improving distribution. Such an explanation should still be checked against suitable evidence rather than accepted automatically.

The relationship between data, information and knowledge can therefore be understood as progression. Data provides the basic observations; information gives those observations structure and meaning; and knowledge develops when information is interpreted and connected to explanations, experience and sound judgement. The process is not always automatic: inaccurate data can produce misleading information, and information can be interpreted incorrectly if important context is ignored.

Economic application: suppose a researcher records the prices of rice in five markets for seven days. The individual price records are data. Calculating daily averages and comparing prices between markets produces information. Using that information, together with knowledge of transport costs, seasonal supply, market access and local demand, the researcher may develop an informed explanation of price differences and consider suitable policy responses. Further data collection and analysis may be needed before concluding that any one factor caused the differences.

In short, data tells us what has been recorded; information helps us understand what the records show; and knowledge helps us interpret their significance and decide what to do. In economics, all three are important for research, business decisions, planning and public policy.

1.2 Common Forms of Economic Data

- Quantitative data: expressed numerically, such as wages, prices, output, age, household expenditure and number of workers.
- Qualitative or categorical data: classify observations into groups, such as rural/urban residence, type of employment, education level or ownership of a bank account.
- Microeconomic data: concern individual units, such as a household, worker or firm.
- Macroeconomic data: concern aggregates for a region or economy, such as inflation, GDP, fiscal deficit and national unemployment.

A variable is a characteristic that can take different values across units or over time. Income, price, production and years of schooling are examples of variables. An observation is a recorded value for a particular unit, place or time. A dataset usually organises observations and variables in a systematic form.

2. Meaning and Importance of Data in Economics

Economics studies how individuals, firms and governments make choices under conditions of scarcity and how resources are produced, distributed and used. Data helps economists describe economic conditions, test theories, identify relationships, evaluate policies and make informed decisions. Economic arguments become more reliable when they are supported by relevant and properly collected evidence.

2.1 Description of the Economy

Data shows the scale and composition of economic activity. GDP data describes the value of final goods and services produced; employment data describes the labour market; price data help measure inflation; and household expenditure data reveal patterns of consumption. Without data, it is difficult to determine whether the economy is expanding, stagnating or contracting.

2.2 Testing Economic Theories

Economic theories often propose relationships between variables. The law of demand, for example, generally predicts that quantity demanded falls when price rises, other things remaining constant. Data on prices and quantities can help economists examine whether the predicted relationship is observed in a particular market. Empirical findings may support a theory, qualify it or show that further explanation is required.

2.3 Formulation and Evaluation of Public Policy

Governments use data to identify problems, design interventions and assess results. Poverty estimates can help identify disadvantaged groups; employment data can inform labour-market programmes; agricultural data can guide procurement and irrigation policy; and health and education data can support public-service planning. Evaluation compares intended outcomes with observed outcomes and, where possible, considers what might have happened without the policy.

2.4 Planning and Resource Allocation

Public authorities, businesses and institutions use data to allocate limited resources. A district with a high school-age population may require additional classrooms; a business may use sales data to decide inventory levels; and a bank may use repayment data to assess lending risk. Reliable data can reduce waste and improve the match between needs and resources.

2.5 Forecasting and Decision-Making

Time-based data help economists identify trends and recurring patterns. Governments may forecast tax revenue, inflation or demand for public services. Firms may forecast sales and production. Forecasts are not certainties: their quality depends on the data, assumptions, methods and stability of the conditions being forecast.

2.6 Comparison Across Groups and Regions

Data allows comparisons between rural and urban households, states, industries, income groups or countries. Comparisons are meaningful only when definitions, units, coverage and periods are sufficiently consistent. For example, comparing wages requires attention to hours worked, occupation, skill, price differences and whether wages are nominal or adjusted for inflation.

2.7 Measuring Inequality and Welfare

Income, consumption, wealth, access to education, health outcomes and employment indicators help assess living standards and inequality. A national average may conceal large differences among groups. Disaggregated data separated by relevant characteristics such as region, gender, age or income group—can reveal who benefits and who is left behind.

2.8 Limitations and Responsible Use of Data

- Data may contain measurement errors, omissions, reporting errors or outdated values.
- Definitions and methods may differ across sources, making direct comparison misleading.

- Correlation does not prove causation by itself; two variables may move together for other reasons.
- Averages may hide differences among individuals or groups.
- Small samples or unrepresentative samples may produce misleading conclusions.
- Personal information must be handled responsibly, with attention to privacy, confidentiality and ethical use.

Good economic analysis therefore requires not merely collecting data, but checking their source, coverage, accuracy, definitions, limitations and relevance to the question being studied.

3. Types of Data by Structure

A common way to classify economic data is by how observations are arranged across units and time. The three important types are cross-sectional data, time-series data and panel data.

3.1 Cross-Sectional Data

Cross-sectional data are collected for many individuals, households, firms, regions or other units at a particular point in time or during a short, specified period. The central feature is that the units differ, while the time reference is broadly the same.

Example: a survey records the monthly income, education and household size of 500 households in Barpeta district during a survey conducted in 2026. Each row represents a different household; the data can be used to compare income across households.

Characteristics

- Many units are observed, while the time dimension is limited or fixed.
- Useful for studying differences between units and relationships among variables.
- Often collected through household surveys, firm surveys, censuses or administrative records.
- Can provide a snapshot of economic conditions but cannot, on its own, show how the same unit changes over several years.

Applications

- Comparing household consumption across income groups.
- Examining the relationship between education and wages among workers in one year.
- Comparing productivity among firms or agricultural output across districts.

Advantages and Limitations

Cross-sectional data can cover a large number of units and are useful for studying variation across people or places. However, they may not capture changes over time. If the sample is not representative, results may not reflect the wider population. Relationships observed in cross-sectional data may also be influenced by unobserved differences between units.

3.2 Time-Series Data

Time-series data consist of observations on one unit, variable or aggregate recorded sequentially over time, usually at regular intervals. The time interval may be daily, monthly, quarterly or annually.

Example: India's annual inflation rate from 2000 to 2025, Assam's annual tea production over twenty years, or monthly retail prices of rice in a market.

Characteristics

- Time is the main organising dimension.
- Observations are arranged chronologically.
- The frequency may be daily, weekly, monthly, quarterly or annual.
- Values may be influenced by long-term trends, seasonal patterns, business cycles, policy changes and unexpected shocks.

Applications

- Studying the trend in GDP, inflation, unemployment or exports.
- Monitoring seasonal changes in agricultural prices or tourism demand.
- Forecasting sales, revenue, production or public expenditure.

Important Concepts in Time-Series Data

- Trend: the long-term direction of a series, such as a gradual increase in output.
- Seasonality: a pattern that repeats at a known interval, such as seasonal demand for a product.
- Cyclical movement: fluctuations associated with broader expansions and contractions in economic activity.
- Irregular movement: unexpected variation caused by unusual events, disasters or shocks.

Advantages include the ability to study change and support forecasting. Limitations include the possibility of structural breaks, changing definitions, missing observations and the influence of unusual events. A trend does not automatically establish the cause of a change.

3.3 Panel Data

Panel data—also called longitudinal data in many contexts—combine cross-sectional and time-series dimensions. The same units, such as households, firms, districts or countries, are observed repeatedly over two or more time periods.

Example: the income and employment status of the same 300 households recorded in 2022, 2024 and 2026. The dataset contains both differences between households and changes within each household over time.

Characteristics

- The dataset has a unit dimension and a time dimension.
- The same units are followed across multiple periods, although some units may be missing in some periods.
- Panel data can distinguish changes within a unit from differences between units.
- A balanced panel observes every unit in every period; an unbalanced panel has missing unit-period observations.

Applications

- Studying how household income changes after a training programme.
- Examining productivity changes among the same firms over several years.
- Analysing differences in economic growth among states across time.

Advantages and Limitations

Panel data can reveal dynamics that a single cross-section cannot show and can help control for some unobserved characteristics that remain stable over time. They may be expensive and difficult to collect, and respondents or firms may leave the study. Changes in measurement or missing observations can complicate analysis. Panel data do not automatically prove causation; the research design still matters.

3.4 Comparison of the Three Types

Feature	Cross-sectional	Time-series	Panel
Main dimension	Different units	Time	Different units and time
Units observed	Many units, usually once	One unit or aggregate followed over time	Same units followed repeatedly
Example	Income of 500 households in 2026	India's annual inflation, 2000–2025	Income of the same households in 2022, 2024 and 2026
Main use	Compare units	Study trends and change	Study both unit differences and changes
Key limitation	Limited view of change over time	May be affected by trends, seasonality and shocks	Cost, attrition and missing observations

A simple way to remember the distinction is: cross-sectional data compare units; time-series data follow time; panel data follow the same units through time.

4. Sources of Economic Data

A source of data is the person, institution, record or system from which observations are obtained. Economic data may be collected directly for a particular study or may already exist because another organisation collected or published them. This distinction leads to the classification of primary and secondary data.

4.1 Primary Data

Primary data are collected first-hand by a researcher or organisation for the specific purpose of the study being undertaken. The researcher decides what information is needed, whom to approach, which questions to ask and how to record the responses, subject to practical and ethical constraints.

Methods of Collecting Primary Data

Primary data are collected directly by a researcher from individuals, households, firms or other relevant sources for a specific research purpose. The major methods of collecting primary data are explained below.

1. Questionnaire: A questionnaire is a structured set of questions prepared by a researcher to collect information from respondents. Respondents read the questions and record their answers on paper or through an electronic form. Questions may be closed-ended, with fixed answer options, or open-ended, allowing respondents to express their views in their own words.

Use: Questionnaires are useful for collecting standardised information from a relatively large number of respondents and comparing their answers systematically.

2. Interview: An interview involves asking questions directly to respondents and recording their answers. It may be conducted face-to-face, by telephone or through an online meeting. Interviews may follow a fixed set of questions or allow additional questions to explore a respondent's answers in greater detail.

Use: Interviews are particularly useful when detailed explanations, personal experiences or reasons behind economic decisions are required. However, they may require more time and resources than questionnaires.

3. Observation: Observation is a method of collecting primary data by systematically watching and recording activities, behaviour or events. Instead of relying entirely on what people report, the researcher directly observes what happens in a particular situation. Observation may be structured, using a predetermined checklist, or relatively unstructured, depending on the research objective.

Use: Observation is useful for studying actual behaviour, market activities and economic transactions. However, it may not reveal people's motives, opinions or reasons for their behaviour unless combined with interviews or other methods.

4. Schedule Administered by an Investigator: A schedule is a list of questions prepared by the researcher, but unlike a self-completed questionnaire, the questions are asked by a trained investigator who records the respondents' answers. This method is particularly useful when respondents have difficulty reading or completing a written form.

Use: This method can improve the completeness of responses because the investigator can explain questions when necessary. However, it requires trained investigators and may involve higher costs. The investigator must also avoid influencing respondents' answers.

5. Experiment or Pilot Test: An experiment is a planned method of collecting data by introducing or varying a particular factor and observing the resulting response under specified conditions. It may help researchers examine whether a change produces a particular outcome. A pilot test is a small-scale preliminary study conducted before a larger survey, programme or business activity to assess its practicality and identify possible problems.

Use: Experiments can help investigate cause-and-effect relationships when properly designed, while pilot tests help assess feasibility and improve research instruments or business plans. However, experiments require careful planning, and their findings may not apply to every market or population. Ethical considerations and practical constraints must also be taken into account.

Advantages of Primary Data

- **Specific to the Research Objective:** Primary data can be collected according to the specific objectives and requirements of a research study.
- **Greater Control over Data Collection:** The researcher can decide which variables to study, the period of data collection, the geographical area and the target population.
- **Collection of First-hand Information:** Primary data provide information directly from respondents or through direct observation of economic activities.
- **Access to Unavailable Information:** Researchers can collect information that may not be available in published reports, official statistics or existing databases.
- **Use of Appropriate Methods:** The researcher can select suitable methods, such as questionnaires, interviews, observation or experiments, according to the nature of the study.
- **Clarification of Responses:** During interviews or investigator-administered surveys, researchers can explain questions and clarify respondents' answers when necessary.
- **Monitoring of Data Quality:** The researcher can supervise the collection process, check incomplete responses and identify errors during fieldwork.
- **Collection of Updated Information:** Primary data can provide current information about economic conditions, consumer preferences, market demand and employment situations.
- **Flexibility in Data Collection:** The researcher can modify questions or procedures when necessary, particularly during pilot studies, subject to the research design.
- **Greater Relevance to Local Problems:** Primary data are useful for investigating specific local economic issues, such as rural unemployment, agricultural marketing, household indebtedness and consumer behaviour.

Limitations of Primary Data

- **High Cost of Collection:** Collecting primary data may require expenditure on questionnaires, travel, field investigators, communication and data processing.
- **Time-consuming Process:** Designing research instruments, contacting respondents, conducting interviews and analysing responses may require considerable time.
- **Difficulty in Reaching Respondents:** Researchers may face problems contacting people in remote areas or obtaining information from busy, unavailable or unwilling respondents.

- **Non-response Problems:** Some respondents may refuse to participate or leave questions unanswered, reducing the completeness of the collected data.
- **Inaccurate Responses:** Respondents may forget information, misunderstand questions or deliberately provide incorrect answers.
- **Risk of Researcher Bias:** An investigator's behaviour, wording of questions or personal expectations may influence respondents' answers and affect the reliability of the findings.
- **Poor Questionnaire Design:** Ambiguous, leading or complicated questions may produce misleading or inconsistent responses.
- **Sampling Bias:** If the selected respondents do not adequately represent the target population, the findings may not be generalisable to the wider population.
- **Limited Geographical Coverage:** Due to financial and practical constraints, primary data collection may be restricted to a particular village, town, district or group of respondents.
- **Need for Trained Personnel:** Certain methods, especially interviews and investigator-administered schedules, require trained investigators to ensure accuracy and consistency.
- **Difficulty in Comparing Results:** Comparisons across studies may be difficult when researchers use different questions, definitions, survey periods or collection methods.
- **Ethical and Privacy Concerns:** Researchers must protect confidential information and obtain appropriate consent, particularly when collecting sensitive information about income, debt or household finances.

4.2 Secondary Data

Secondary data are data that have already been collected, compiled or published by another person or institution, often for a different purpose, and are subsequently used by a researcher. They may be available in printed reports, official websites, databases, books, research articles or administrative records.

Examples

- Population and housing information published by a census authority.
- National income, inflation, employment and trade statistics published by government agencies.
- Reports of the Reserve Bank of India, ministries, state departments and other public institutions.

Sources of Secondary Data:

Secondary data is obtained from sources where information has already been collected, compiled, processed or published by other individuals, institutions or organisations. These data are widely used in economic research because they help researchers study economic conditions, identify trends and compare regions or countries without collecting all the information directly. The major sources of secondary data include government publications, official statistical agencies, international organisations, research institutions, academic journals, books, newspapers, business reports and administrative records. In India, important sources include the Ministry of Statistics and Programme Implementation (MoSPI), the National Statistical Office (NSO), the Office of the Registrar General and Census Commissioner, the Reserve Bank of India (RBI), the Economic Survey and state government departments. International organisations such as the World Bank, International Monetary Fund (IMF), International Labour Organization (ILO) and United Nations also provide economic and social statistics. Furthermore, universities, research centres, market research agencies, company annual reports and industry associations publish information on production, employment, prices, income, consumption and business activities. Secondary data may be available in printed publications, official websites, online databases and research reports. However, before using such data, researchers should examine the reliability of the source, accuracy, relevance, reference period, geographical coverage, definitions and methods of data collection to ensure that the information is suitable for the research objective.

Advantages of Secondary Data

- **Economical:** Secondary data are generally less expensive to obtain than primary data because the researcher does not need to conduct a new survey or undertake extensive fieldwork.
- **Time-saving:** These data are often readily available in published reports, official websites, research articles and databases, enabling researchers to begin their analysis quickly.
- **Wide Coverage:** Secondary data may cover large populations, multiple geographical regions, different economic sectors and long periods, making them useful for comprehensive economic studies.
- **Useful for Historical Analysis:** Data collected over several years help researchers identify long-term trends and changes in variables such as income, employment, prices, production and economic growth.
- **Facilitates Comparison:** Secondary data allow comparisons across states, districts, countries, income groups and different periods, provided the definitions and measurement methods are sufficiently consistent.
- **Supports Background Research:** Existing publications help researchers understand the research problem, review previous studies and identify gaps in the available literature.
- **Helps Formulate Research Objectives:** By examining existing data, researchers can identify important economic issues, develop research questions and formulate suitable hypotheses.
- **Assists in Primary Data Collection:** Secondary data help researchers identify relevant variables, understand the characteristics of the target population and design appropriate questionnaires for primary surveys.

- Provides a Basis for Policy Analysis: Official statistics and published reports help researchers examine government policies, assess economic conditions and study changes in social and economic indicators.
- Enables Large-scale Research: Researchers with limited financial resources can undertake studies covering wider geographical areas or longer periods by using existing datasets.

Limitations of Secondary Data

- Lack of Specificity: Secondary data may not fully satisfy the requirements of a particular research study because they were originally collected for a different purpose.
- Differences in Definitions: Different organisations may define economic concepts, variables and indicators differently, making direct comparisons difficult.
- Changes in Measurement Methods: Changes in data collection procedures, classifications, survey designs or statistical methods over time may affect comparability.
- Possibility of Outdated Information: Some publications and datasets may not reflect current economic conditions, particularly in rapidly changing markets or sectors.
- Incomplete Data: Certain variables, geographical areas or population groups may be missing, limiting the scope of the analysis.
- Limited Geographical or Sectoral Detail: Data may be available only at the national or state level rather than at the district, village, household or individual level required for a study.
- Limited Control over Data Collection: Researchers generally cannot control how the original data were collected, which questions were asked, or how observations were recorded.
- Questions about Reliability: The accuracy and credibility of secondary data depend on the quality of the original source, its collection methods and its reporting practices.
- Possibility of Errors: Published datasets may contain measurement errors, omissions, transcription mistakes or classification problems that affect the results of the research.
- Difficulty in Accessing Detailed Data: Some datasets may require registration, payment, special permission or technical expertise to access and analyse.
- Limited Information on Methodology: Some reports may not provide sufficient details about sampling procedures, variable definitions, estimation methods or data limitations.
- Risk of Misinterpretation: Researchers may draw misleading conclusions if they use secondary data without understanding their context, coverage, limitations and original purpose.

4.3 Primary Data and Secondary Data: Comparison

1. **Meaning:** Primary data are collected first-hand by the researcher for the current study, whereas secondary data have already been collected, compiled or published by another person or organisation.
2. **Purpose:** Primary data are generally collected to address a specific research question, whereas secondary data were originally collected for another or broader purpose.
3. **Cost and Time:** Primary data collection usually requires more time and money because it involves surveys, interviews or fieldwork, whereas secondary data are often quicker and less costly to obtain.
4. **Control over Data Collection:** In primary data collection, the researcher has greater control over the questions, methods and procedures, whereas in secondary data analysis, control over the original collection process is limited.
5. **Coverage:** Primary data may cover a limited geographical area or a small group of respondents when resources are constrained, whereas secondary data may cover large populations, several regions and long periods.
6. **Specificity and Relevance:** Primary data can be tailored to the exact requirements of a research study, whereas secondary data may not contain all the variables or information required by the researcher.
7. **Availability:** Primary data become available after the researcher completes the collection process, whereas secondary data may already be available in published reports, official websites and databases.
8. **Accuracy and Reliability:** The researcher can monitor the collection and recording of primary data, although errors may still occur; the reliability of secondary data depends substantially on the quality of the original source and its methods.
9. **Flexibility:** Primary data collection methods and questions may be adjusted when appropriate, whereas secondary data are generally restricted to the information and definitions already available.

5. Sources of Secondary Economic Data

Secondary economic data come from a wide range of official and non-official sources. Researchers should identify the original producer of the data, rather than relying only on a website or publication that reproduces figures.

5.1 Government and Official Statistical Sources

Government agencies collect and publish data for administration, planning, public accountability and policy. Such sources are particularly important for national and state-level economic indicators.

- **Ministry of Statistics and Programme Implementation (MoSPI):** national statistical publications, national accounts and other official statistical information.

- National Statistical Office (NSO): large-scale surveys and official statistics, including household and enterprise-related information. Specific survey availability and terminology should be checked in the relevant publication.
- Office of the Registrar General and Census Commissioner, India: population census information and related demographic statistics.
- Ministry of Finance and Economic Survey: information and analysis on public finance and the broader economy.
- State government departments and Directorate of Economics and Statistics: state-level economic reviews, statistical handbooks, agriculture, industry and district-level statistics.
- Local self-government institutions: selected administrative data on local services, budgets and development activities.

5.2 Reserve Bank of India (RBI)

The RBI publishes information related to banking, money and credit, interest rates, inflation-related indicators, external-sector statistics and financial markets. Its publications and databases are useful for research on monetary economics, banking, finance and macroeconomic conditions.

5.3 Other Public Institutions and Regulators

Institutions responsible for specific sectors may publish useful data. Examples include regulators and public bodies concerned with securities markets, insurance, energy, transport, telecommunications, education, health and labour. Researchers should verify the scope and methodology of each dataset before use.

5.4 International Organisations

International organisations compile data that help compare countries and study global economic issues. Examples include:

- World Bank: development indicators, poverty, population, infrastructure and related topics.
- International Monetary Fund (IMF): macroeconomic, fiscal, monetary and external-sector data.
- International Labour Organization (ILO): labour-market indicators and labour statistics.
- United Nations agencies: demographic, social, development and sectoral indicators.
- Food and Agriculture Organization (FAO): agriculture, food and related statistics.

International databases can improve comparability, but definitions, country coverage and update schedules still need to be checked.

5.5 Academic and Research Sources

Books, peer-reviewed journal articles, university repositories, research reports, working papers and theses can provide data, explanations and references to original datasets. When a study reproduces a table or result from another source, researchers should trace the information back to the original dataset whenever possible.

5.6 Business and Industry Sources

Trade associations, market-research organisations, company annual reports, industry surveys and business publications can provide information on sales, production, prices, market shares, investment and consumer behaviour. Some data are proprietary or based on commercial estimates, so their coverage and methods may not be fully transparent.

5.7 Newspapers, Magazines and Online Sources

Newspapers and magazines are useful for identifying events, policy announcements, business developments and leads to published statistics. However, a news report is not always the best source for a precise statistical figure. Wherever possible, verify the number against the original report, official database or primary document.

5.8 Administrative Records

Administrative records are generated when institutions carry out routine activities. Examples include school enrolment records, tax filings, hospital records, employment registrations, land records and business registrations. They may offer detailed and frequent information, but they were not always designed for research. Coverage may exclude people or activities outside the relevant administrative system.

5.9 How to Evaluate a Secondary Source

1. Authority: identify the institution or author responsible for producing the data.
2. Originality: determine whether the source is the original producer or merely reproduces another source.
3. Purpose: understand why the data were collected and whether that purpose affects what was measured.
4. Definitions: check how key terms and variables are defined.
5. Coverage: examine the geographical area, population, sector and period included.
6. Methodology: look for information on sampling, measurement, estimation, revisions and missing data.
7. Timeliness: check the reference period, publication date and update frequency.
8. Comparability: confirm that definitions and units are consistent across years, regions or groups.
9. Reliability and limitations: read technical notes, caveats and revision statements.
10. Citation: record the title, institution or author, year, table or dataset name, URL where relevant, and date accessed when appropriate.

Example: before using a state-level unemployment figure, a researcher should check the survey period, age group, definition of unemployment, estimation method and whether the figure refers to a particular state or the whole country.

6. Census: Concept Only

A census is a systematic process of collecting information from every unit in a defined population. In a population census, the aim is to enumerate all people or households within the defined geographical area according to the census rules. The term census can also be used in other settings, such as an establishment census or an agricultural census, where the units are establishments or agricultural holdings.

6.1 Population and Census Unit

The population is the complete set of units about which information is required. A unit is the individual element being counted or measured, such as a person, household, firm or farm. A census must define its target population, geographical boundaries, reference date or period, and rules for inclusion to avoid omissions and double counting.

6.2 Basic Features

- Attempts to cover every unit in the defined population.
- Uses standard definitions and procedures to make records comparable.
- Can provide detailed information for small geographical areas or population groups.
- Requires substantial planning, staff, time, funding and data processing.

6.3 Advantages

- Provides broad coverage of the defined population.
- Can generate detailed counts and characteristics for small areas and subgroups.
- Does not rely on selecting only a sample, although non-sampling errors can still occur.
- Can provide a benchmark for planning and for designing or evaluating sample surveys.

6.4 Limitations

- Can be expensive and time-consuming.
- Information may become outdated as population and economic conditions change.
- People or units may be missed, counted more than once or provide inaccurate information.
- Processing and publishing results can take time.

A census is not automatically error-free. Errors can arise from incomplete coverage, misunderstanding, incorrect responses, data entry and classification. A census may also collect less detailed information than a specialised survey because it must be feasible for a very large population.

7. Sampling: Concept Only

Sampling is the process of selecting a subset of units from a defined population in order to collect information and draw conclusions about that population. The selected subset is called a sample. Sampling is widely used because studying every unit may be too costly, too slow or impractical.

7.1 Key Terms

- Population: the complete set of units of interest.
- Sample: the units selected for observation.
- Sampling frame: the list or operational source from which units are selected.
- Sample size: the number of units included in the sample.
- Representative sample: a sample that adequately reflects the important characteristics of the target population for the intended analysis.
- Sampling error: the difference between a sample estimate and the corresponding population value that occurs because only a sample is observed.
- Non-sampling error: error caused by factors such as non-response, poor measurement, data-entry mistakes or coverage problems; it can occur in both censuses and sample surveys.

7.2 Broad Types of Sampling

At the introductory level, sampling methods can be divided into probability and non-probability methods.

- Probability sampling: each population unit has a known, non-zero chance of selection. Examples include simple random sampling, systematic sampling, stratified sampling and cluster sampling.
- Non-probability sampling: selection probabilities are not known. Examples include convenience sampling, purposive sampling and quota sampling. These methods may be useful in exploratory work but generally do not support the same design-based generalisation as probability samples.

7.3 Advantages of Sampling

- Usually requires less time and money than a census.
- Can allow more detailed questioning and closer supervision of fieldwork.
- Can provide timely estimates when the sample is well designed.
- Is suitable when measuring every unit is impossible or destructive testing is involved.

7.4 Limitations of Sampling

- A sample may not represent the population if selection is biased.
- Estimates are subject to sampling uncertainty.
- Small subgroups may be difficult to estimate precisely unless the design accounts for them.
- Non-response and poor measurement can distort results even if the sample was selected properly.

7.5 Comparison between Census and Sampling

- **Coverage:** A census aims to collect information from every unit in the defined population, whereas sampling involves collecting information from only a selected subset of the population.
- **Cost and Time:** A census generally requires more money, manpower and time because all units must be covered, whereas sampling is usually less expensive and quicker.
- **Detail for Small Groups:** A census can provide detailed information about small geographical areas or population groups, whereas sampling may provide limited information about such groups if the sample size is insufficient.
- **Sampling Error:** A census does not involve sampling error arising from the selection of a subset because every unit is intended to be covered, whereas sampling estimates are subject to sampling error because they are based on only part of the population.
- **Other Errors:** A census may suffer from coverage errors, incorrect responses, non-response and data-processing mistakes, whereas sampling surveys can experience these same non-sampling errors in addition to sampling error.
- **Accuracy:** A properly conducted census can provide comprehensive information about the defined population, whereas the accuracy of sampling estimates depends on the sampling design, sample size, representativeness and quality of data collection.
- **Suitability:** A census is suitable when information about every unit is required and sufficient resources are available, whereas sampling is suitable when reliable estimates from a carefully selected subset are sufficient for the research objective.
- **Frequency of Study:** Conducting a census frequently may be difficult because of its cost and time requirements, whereas sampling surveys can often be conducted more frequently to monitor changes in economic conditions.
- **Management and Organisation:** A census generally requires extensive administrative arrangements and coordination, whereas sampling usually requires fewer field investigators and comparatively simpler field operations, depending on the survey design.
- **Example:** Collecting information about every household in a village is an example of a census, whereas collecting information from 100 carefully selected households in the same village to estimate average household expenditure is an example of sampling.

Key Terms

Term	Meaning
Data	Recorded facts, observations or measurements.
Variable	A characteristic that can take different values.
Observation	A recorded value for a unit, place or time.
Cross-sectional data	Data on multiple units at a particular time or short period.
Time-series data	Sequential observations on a variable or aggregate overtime.
Panel data	Repeated observations on the same units across time.
Primary data	First-hand data collected for the current study.
Secondary data	Data has already been collected or compiled by another source.
Population	The complete set of units of interest.
Census	Enumeration intended to cover every unit in the defined population.
Sampling	Selecting a subset of units to learn about the population.
Sampling frame	A list or operational source used to select sample units.
Sampling error	Uncertainty arises because only a sample is observed.
Non-sampling error	Errors such as non-response, measurement mistakes or data processing errors.
Administrative data	Information generated through routine institutional activities.