

DATA COLLECTION

CCDATSCL | COM221-ML

INTRODUCTION

Modern businesses run on data.

They regularly capture, store and analyze large amounts of data to which they can apply analytics to make strategic decisions and make some conclusions about their performance.

For example, in a coffee shop business, they may capture data about what selections the menu sells the most at a particular point in time.

Or they may capture information about the purpose of each customer's visit or their demographic.

By gathering these data, a coffee shop can analyze them and know amount of stock they need to purchase based on customer demand.

Or add more selections in the menu based on the purpose of visit of each customers.

Thus, data collection is essential to analyze the performance of any business, whether to solve a problem and making assumptions about specific things when required.

What is Data Collection?

In statistics, data collection is a process of gathering information from all relevant sources to find a solution to a research problem.

The data collection methods allow a person to conclude an answer to the relevant question.

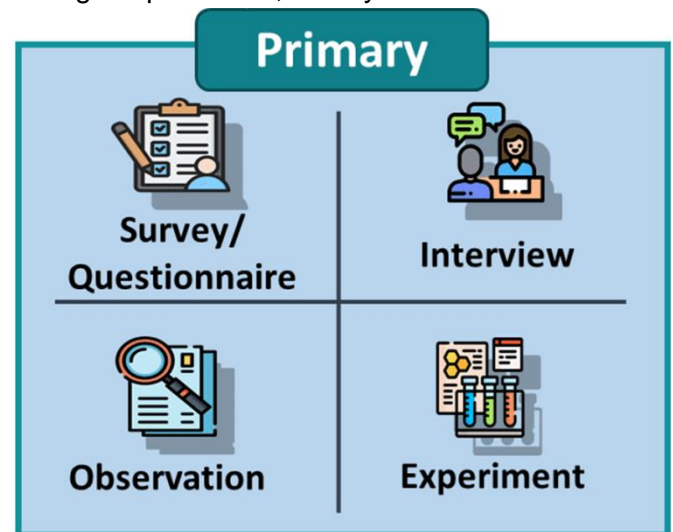
Most of the organizations use data collection methods to make assumptions about future probabilities and trends.

Depending on the type of data, the data collection method is divided into two categories namely,

- Primary Data Collection
- Secondary Data Collection

Primary Data Collection

Primary data or raw data is a type of information that is obtained directly from the first-hand source through experiments, surveys or observations.

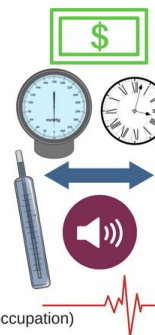


The primary data collection method is further classified into two types. They are

- Quantitative Data Collection Methods
- Qualitative Data Collection Methods

Quantitative Data

- money
- time
- speed
- movement
- height
- length
- area
- volume
- weight
- temperature
- humidity
- pressure
- sound level
- categories (age, gender, occupation)
- positioning
- status



Qualitative Data

- verbal and written feedback
- narrative story
 - first-hand (direct experience)
 - second-hand (telling someone else)
 - third-hand (outside story-teller)
- visual images, drawings, or models
- experiential sensations
- descriptions of
 - colors
 - textures
 - smells
 - tastes
 - appearance
 - beauty
 - feelings
 - intuition
 - sensations
 - choices
 - values
 - beliefs



1. Quantitative Data Collection

- Quantitative Data is information that can be **measured, counted, and expressed in numerical values**, answering questions like "how much," "how many," or "how often".
- It can be divided into two main types:
 - **Discrete** (countable, like the number of students).

- **Continuous** (measurable within a range, like height or temperature).
- 2. **Qualitative Data Collection**
 - Qualitative data is non-numerical information used to understand concepts, opinions, and experiences, collected through methods like interviews, observations, and text analysis.

Data Collection Tools

1. Surveys

- *Surveys allow you to collect data about opinions, behaviours, experiences, and demographic characteristics by asking people directly.*
- *Surveys are ideal for documenting perceptions, attitudes, beliefs, or knowledge within a clear, predetermined sample of individuals.*
- *Surveys can be done using questionnaires where you ask people to fill out questions themselves*
- *Or conduct interviews where you ask questions and record the answers.*

- **Questionnaires**

Questionnaires are more common in quantitative research.

They usually include closed questions with multiple-choice answers or rating scales.

They can also be used to obtain qualitative data by asking open-ended questions with a free-text format.

Open-ended questions should be specific enough to yield coherent responses across respondents, yet broad enough to invite a spectrum of answers.

This allows you to collect consistent data from many people and analyze the responses statistically.

For example, if you want to measure student satisfaction with their teacher, you could use a questionnaire to collect responses from many students.

For example, if you want to measure alcoholic consumption of a patient, you could use a questionnaire to collect responses from many patients.

- **Interviews**

Interviews are more common in qualitative research.

They usually allow participants to answer questions in their own words.

Interviews are ideal when used to document participants' accounts, perceptions of, or stories about attitudes toward and responses to certain situations or phenomena.

TYPES OF INTERVIEWS

They can be structured or unstructured

Interviews can either follow a tightly written script that mimics a survey or be inspired by a loose set of questions that invite interviewees to express themselves more freely.

Interviews vs Surveys

Many research questions that can be answered with surveys can also be answered through interviews, but interviews will generally yield richer, more in-depth data than surveys.

Unlike questionnaires, you can ask a follow-up question and explore ideas in more depth.

Disadvantage of Interviews

- 1) Because interviewers are the instruments of data collection, interviewers should be trained to collect comparable data.
- 2) The number of interviews required depends on the research question and the overarching methodology used.

When to use Interviews

- a) You wish to gather very detailed information

- b) You anticipate wanting to ask respondents follow-up questions based on their responses
- c) You plan to ask questions that require lengthy explanation
- d) You are studying a complex or potentially confusing topic to respondents.
- e) You are studying processes, such as how people make decisions

2. Observations

- *Observations allow you to collect data unobtrusively, observing characteristics, behaviours or social interactions without relying on self-reporting.*
- *Observations may be conducted in real time, taking notes as you observe, or you might make audio visual recordings for later analysis.*
- *In theory, observations allow you to collect data on how people really behave and not just how they say they do.*
- *But keep in mind that being observed may make people behave differently than they normally would.*

Quantitative Observation

- Quantitative observation involves systematically measuring or counting specific events, behaviours, characteristics etc.
- With this method, it is important to have a clear objective rules to ensure you count the same thing consistently.
- You also need to clearly define the categories and criteria of your observation in advance.
- For example, to research your own improvement in running, you could observe your performance directly using a running application.

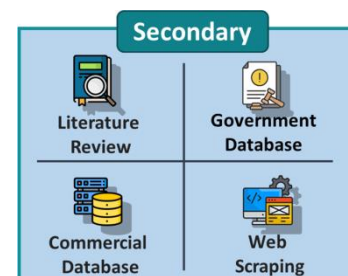
Qualitative Observation

- Qualitative observation involves taking detailed notes and writing rich descriptions of what is observed.

- This means that you do not have to decide in advance how to categorize your observations.

Secondary Data Collection

- In media and communication, you might collect speeches samples of texts to be analyzed. (e.g speeches, articles or social media posts)
- In psychology, you might use technologies like neuroimaging, eye-tracking to measure things such as attention or reaction time.
- In education, you might use tests or assignments to collect data about knowledge and skills.
- In physical sciences, you might use specific instruments to measure things such as weight, blood pressure or chemical composition.
- What If you do not have the time or resources to collect data from the population you to study?
- Instead of collecting your own data, you can use secondary data that other researchers already collected.
- For example, datasets from government surveys or previous studies on your topic.
- With this raw data, you can do your own analysis to answer new research questions that were not addressed by the original study.



Advantage of using Secondary Data

- Using secondary data gives you access to much larger and more varied samples than you could collect yourself.

Limitations of using Secondary Data

- However, it also means that you do not have any control over which variables to measure or how to measure them, so the conclusions you can draw may be limited.