

# Data Sampling and Errors

## Data Sampling

Sampling is the process of selecting a subset (sample) from a larger group (population) to make inferences about the whole population. Since studying the entire population is often expensive or impossible, sampling provides estimates that are quicker and cost-effective.

---

## Errors in Sampling

When we use samples, errors may occur because the sample may not perfectly represent the population.

### 1. Sampling Errors

- These errors occur because we are studying only a subset of the population, not the entire population.
- They arise by chance and can be reduced by increasing sample size and using proper random sampling methods.
- **Example:** If the average height of the population is 170 cm, a random sample might give 168 cm.

### 2. Non-Sampling Errors

- These errors occur due to human, procedural, or technical issues, not because of the sample size.
  - Causes include wrong data collection, faulty measuring instruments, data entry mistakes, or biased questionnaire design.
  - **Example:** A survey asking a leading question such as *“Don’t you agree online learning is better?”* may bias responses.
- 

## Errors in Hypothesis Testing

When using samples for hypothesis testing, two main types of errors can occur:

### Type I Error (False Positive)

- Rejecting the null hypothesis ( $H_0$ ) when it is actually true.
- The probability of committing a Type I error is represented by  $\alpha$  (significance level).

- **Example:** Concluding a new medicine works when it actually does not.

### Type II Error (False Negative)

- Failing to reject the null hypothesis when it is actually false.
- The probability of committing a Type II error is represented by  $\beta$ .
- **Example:** Concluding a new medicine does not work when it actually does.

---

### Summary Table

Error Type	Cause	Example
Sampling Error	Studying only part of the population	Sample mean = 168 cm, true mean = 170 cm
Non-Sampling Error	Human/procedural bias, faulty instruments	Wrong data entry or biased survey
Type I Error ( $\alpha$ )	Rejecting a true null hypothesis	Approving ineffective drug
Type II Error ( $\beta$ )	Accepting a false null hypothesis	Rejecting an effective drug