Table of Contents

# 1.5 Sampling Funnel, CLT & Confidence Interval

## Sampling funnel

When analysing a large population, it is often impractical to measure every member. A **sampling funnel** narrows the population down to a manageable sample:

1. **Population** – All individuals satisfying the study’s criteria.
2. **Sampling frame** – A list or source containing the accessible members of the population. Bias can be introduced if the frame excludes certain groups【440273445372831†L292-L321】.
3. **Simple random sampling** – Each member has an equal probability of selection, producing an unbiased sample【440273445372831†L292-L321】.
4. **Sample** – The subset of individuals selected for analysis.

## Central Limit Theorem (CLT)

The CLT states that, for a large sample size (for example, n ≥ 30), the sampling distribution of the sample mean is approximately normal, regardless of the population’s distribution【353193119138672†L306-L349】. Its key implications are:

* The mean of the sampling distribution equals the population mean【353193119138672†L364-L390】.
* The standard deviation of the sampling distribution (standard error) equals the population standard deviation divided by the square root of the sample size【353193119138672†L398-L427】.
* As sample size increases, the sampling distribution becomes more concentrated around the mean, reducing variability【353193119138672†L398-L427】.

## Sampling variability and confidence intervals

* **Sampling variability** – Different random samples yield different sample means. The variability decreases with larger sample sizes【973746162079337†L58-L110】.
* **Confidence interval** – Provides a range of plausible values for a population parameter. It is computed as “estimate ± critical value × standard error.” For a 95 % confidence level, the critical value is approximately 1.96【441469260235965†L331-L349】【441469260235965†L350-L373】.

## Example

Suppose we wish to estimate the average daily time people spend on a social media platform. We randomly sample 100 users and compute a sample mean of 90 minutes with a standard deviation of 30 minutes. The standard error is 30 divided by the square root of 100, which equals 3. A 95 % confidence interval is 90 ± 1.96 × 3 = [84.1, 95.9] minutes.

## Summary

Sampling allows us to learn about a population without measuring every member. The CLT justifies using the normal distribution when analysing sample means, and confidence intervals quantify the uncertainty in our estimates【353193119138672†L306-L349】【353193119138672†L364-L390】【441469260235965†L331-L349】【441469260235965†L350-L373】.

## Reflection questions

1. Why is random sampling important when selecting a sample?
2. How does increasing the sample size affect the width of a confidence interval?
3. Explain the meaning of a 95 % confidence level.

## References