Chapter 7:
Point Estimation of Parameters and Sampling Distributions

MAS291 - STATISTICS & PROBABILITY

Ly Anh Duong

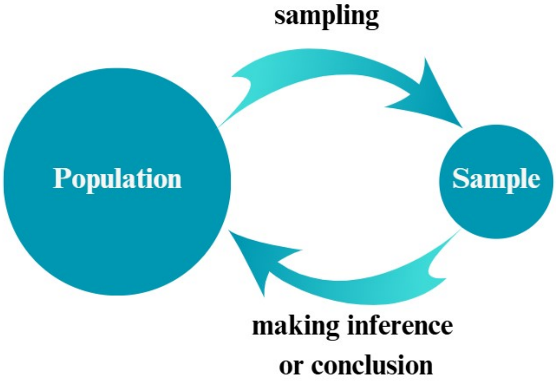
duongla3@fe.edu.vn

Table of Contents

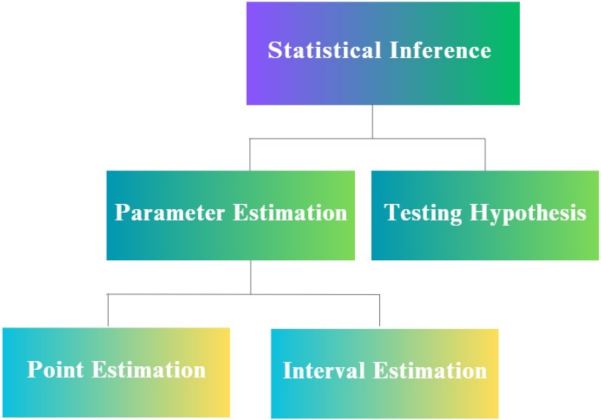
Introduction

1 Introduction

The field of statistical inference consists of those methods used to make decisions or to draw conclusions about a population.

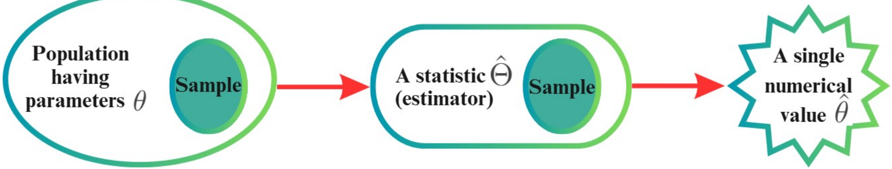


Statistical inference



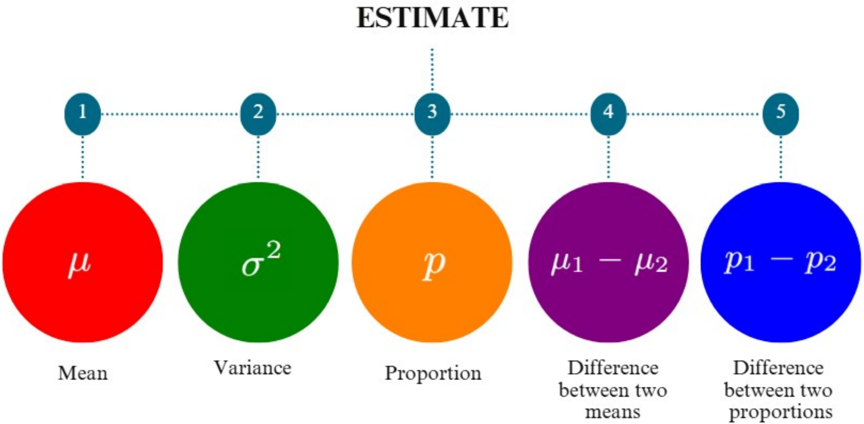
Point estimate

Point Estimator: A point estimate of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the point estimator.



A statistic $\hat{\Theta}$ is a function of the observations in a random sample $X_1, X_2, \ldots, X_n$.
Distribution of a statistic $\hat{\Theta}$ is called sampling distribution.
For example, the probability distribution of $\bar{X}$ is the sampling distribution of the mean.

(Point) Estimation problems



Reasonable estimates

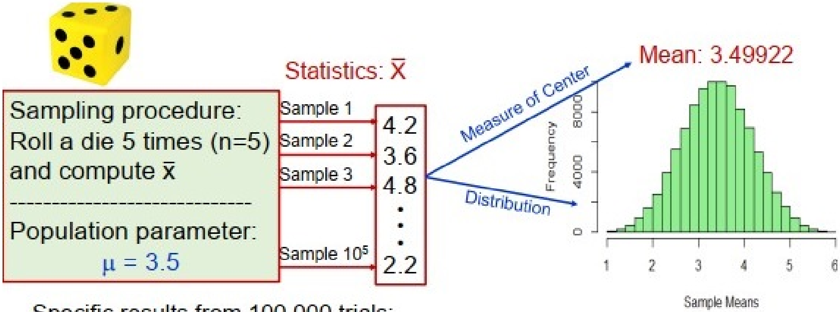
Population Parameter	Symbol	Point Estimate	Symbol
Population Mean	μ	Sample Mean	$\overline{\mathrm{x}}$
Population Standard Deviation	σ	Sample Standard Deviation	s
Population Proportion	p	Sample Proportion	$\hat{\mathrm{p}} = x/n$
Population Variance	σ^2	Sample Variance	s^2
Difference in Means of Two Populations	$\mu_1 - \mu_2$	Difference in Sample Means	$\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2$
Difference in Proportions of Two Populations	$\mathrm{p}_1 - \mathrm{p}_2$	Difference in Sample	$\hat{\mathrm{p}}_1 - \hat{\mathrm{p}}_2$

Table of Contents

Sampling distribution

Statistic: is any value calculated from the sample data.

Sampling Distribution: The probability distribution of a statistic.



Specific results from 100,000 trials:

The population mean is 3.5; the mean of the 100,000 trials is 3.49922.

Notice the distribution is "normal."

Central Limit Theorem (CLT)

CLT states that when we take a large random sample from a population with any distribution and has mean μ, variance σ^2 then the sample mean $\overline{X} \approx N(\mu, \sigma^2/n)$ as long as the sample size is sufficiently large.

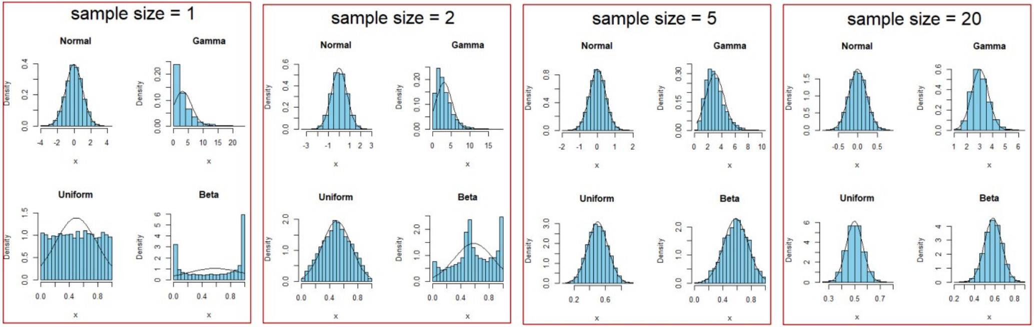
Theorem

If $X_1, ..., X_n$ is a random sample of size n taken from a population with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

is the standard normal distribution as $n \to \infty$

CLT – different sample size



Sampling distribution tends to standard normal distribution as sample size gets larger

CLC - Exercise

2 Sampling Distributions and the Central Limit Theorem

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance of fewer than 95 ohms.

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of n = 25 resistors will have an average resistance of fewer than 95 ohms.

Solution

Note that the sampling distribution of $\bar{X}$ is normal with mean $\mu_{\bar{X}} = 100$ ohms and a standard deviation of $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{10}{\sqrt{25}} = 2$

Therefore,

$$\begin{aligned} P(\bar{X} < 95) &= P\left(\frac{\bar{X} - 100}{2} < \frac{95 - 100}{2}\right) \\ &= P(Z < -2.5) \\ &= 0.0062 \end{aligned}$$

Table of Contents

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\bar{X}_1$ and $\bar{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

is approximately standard normal if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

Exercise

The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a normal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Suppose that a random sample of $n_1 = 16$ components is selected from the "old" process and a random sample of $n_2 = 25$ components is selected from the "improved" process. What is the probability that the difference in the two samples means $\overline{X}_2 - \overline{X}_1$ is at least 25 hours? Assume that the old and improved processes can be regarded as independent populations.

Solution

The distribution of $\bar{X}_1$ is normal with mean $\mu_1 = 5000$ hours and standard deviation $\sigma_1/\sqrt{n_1} = 40/\sqrt{16} = 10$ hours.
The distribution of $\bar{X}_2$ is normal with mean $\mu_2 = 5050$ hours and standard deviation $\sigma_2/\sqrt{n_2} = 30/\sqrt{25} = 6$ hours.
The distribution of $\bar{X}_2 - \bar{X}_1$ is normal with mean $\mu_2 - \mu_1 = 5050 - 5000 = 50$ hours and variance $\sigma_2^2/n_2 + \sigma_1^2/n_1 = (6)^2 + (10)^2 = 136$ hours^2.

$$P(\bar{X}_2 - \bar{X}_1 \geq 25) = P(Z \geq \frac{25 - 50}{\sqrt{136}}) = 0.9838$$

Therefore, there is a high probability (0.9838) that the difference in sample means between the new and the old process will be at least 25 hours if the sample sizes are $n_1 = 16$ and $n_2 = 25$.

Q&A

Thank you for listening!