STATISTICS II

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

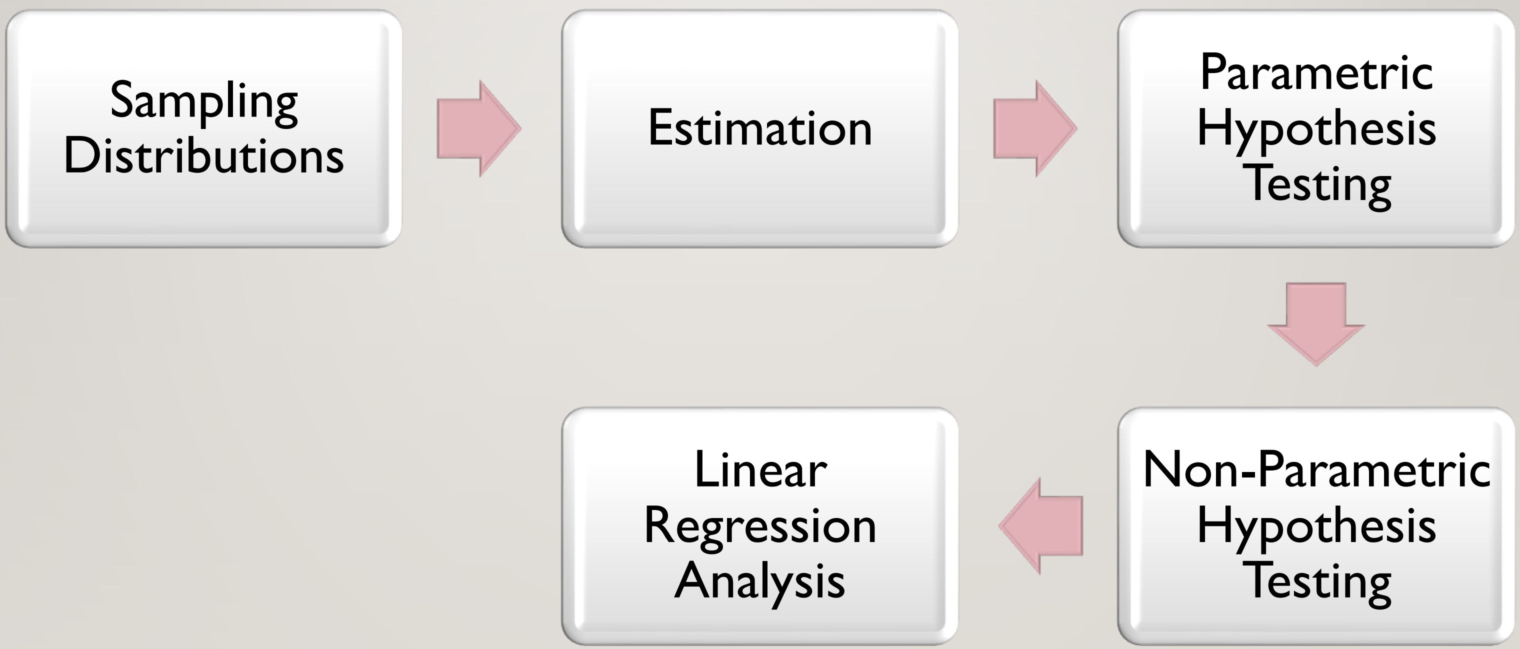
CONTACT

Professor: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

PROGRAM


Sampling Distributions
Estimation
Parametric Hypothesis Testing
Non-Parametric Hypothesis Testing
Linear Regression Analysis

BIBLIOGRAPHY

Main reference:

- Newbold, P., Carlson, W., & Thorne, B. (2013). *Statistics for business and economics* (8th ed.). Pearson.

Complementary references:

- Marôco, J. (2021), *Análise Estatística com o SPSS Statistics*. 8ª Ed. ReportNumber.
- Silvestre, A. L. (2007). *Análise de Dados: Estatística Descritiva.* Escolar Editora.
- Triola, M. (2022). *Elementary Statistics* (14th ed.). Pearson.
- Winston, W. (2016). *Microsoft Excel 2016 Data Analysis and Business Modeling*, Microsoft Press.

Packages: Excel & SPSS

ASSESSEMENT REGIME

- The **final grad**e (FG) is calculated as:

  FG = 0.70 × Written Exam + 0.10 × Attendance (theoretical and practical classes) + 0.10 × Homework in theoretical classes + 0.10 × Homework in practical classes

- In the **written exam**, students must obtain a **minimum score of 7 points** in order to be eligible for evaluation in the **regular period**.
- During the **retake period**, only the grade obtained in the final exam taken at that time will be considered.
- If the final grade exceeds 18 points, the student may be required to take an additional exam to defend the grade.
- Basic calculator may be used during the written exam.

PRESENTATION: INTRODUCE YOURSELF

Please share:

- 👤 **Name**
- 🌍 **Nationality**
- 🎓 **Academic Background**
- 📊 **Experience with Statistics** (courses, software, projects)

LECTURE 1: SAMPLING AND SAMPLING DISTRIBUTIONS

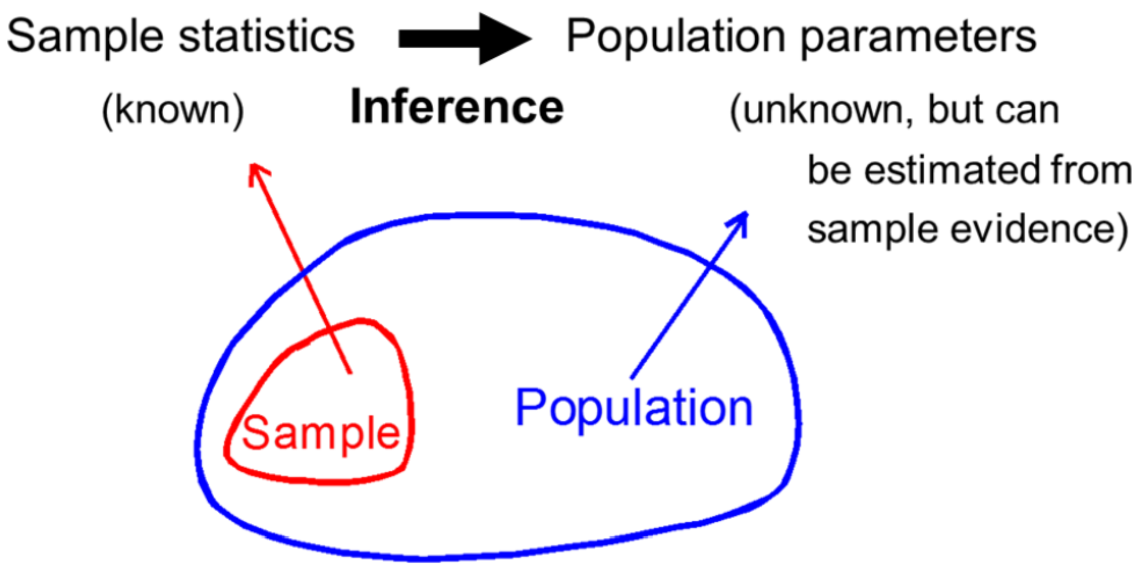
DESCRIPTIVE STATISTICS VS. INFERENTIAL STATISTICS

- **Descriptive statistics**
  - Collecting, presenting, and describing data
- **Inferential statistics**
  - Drawing conclusions and/or making decisions concerning a population based only on sample data

Newbold et al (2013)

INFERENTIAL STATISTICS

- Making statements about a population by examining sample results



Newbold et al (2013)

INFERENTIAL STATISTICS

Drawing conclusions and/or making decisions concerning a population based on sample results.

- **Estimation**
  - e.g., Estimate the population mean weight using the sample mean weight
- **Hypothesis Testing**
  - e.g., Use sample evidence to test the claim that the population mean weight is 120 pounds

Newbold et al (2013)

Example of Estimation:

- The **population mean weight** (μ) is **estimated** by the **sample mean weight** ($\bar{X}$).

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- So, the **population parameter** (the true average weight of all individuals in the population) is **estimated using** the **sample statistic** (the average weight of individuals in the sample).

Example of Hypothesis Test:

- We use the **sample data** to test whether the **population mean weight** is **120 pounds**.

$$H_0 : \mu = 120 \quad \text{(null hypothesis)}$$

$$H_1 : \mu \neq 120 \quad \text{(alternative hypothesis)}$$

- If the **p-value < α**, we **reject** H_0 and conclude that the population mean weight is **significantly different from 120 pounds**.

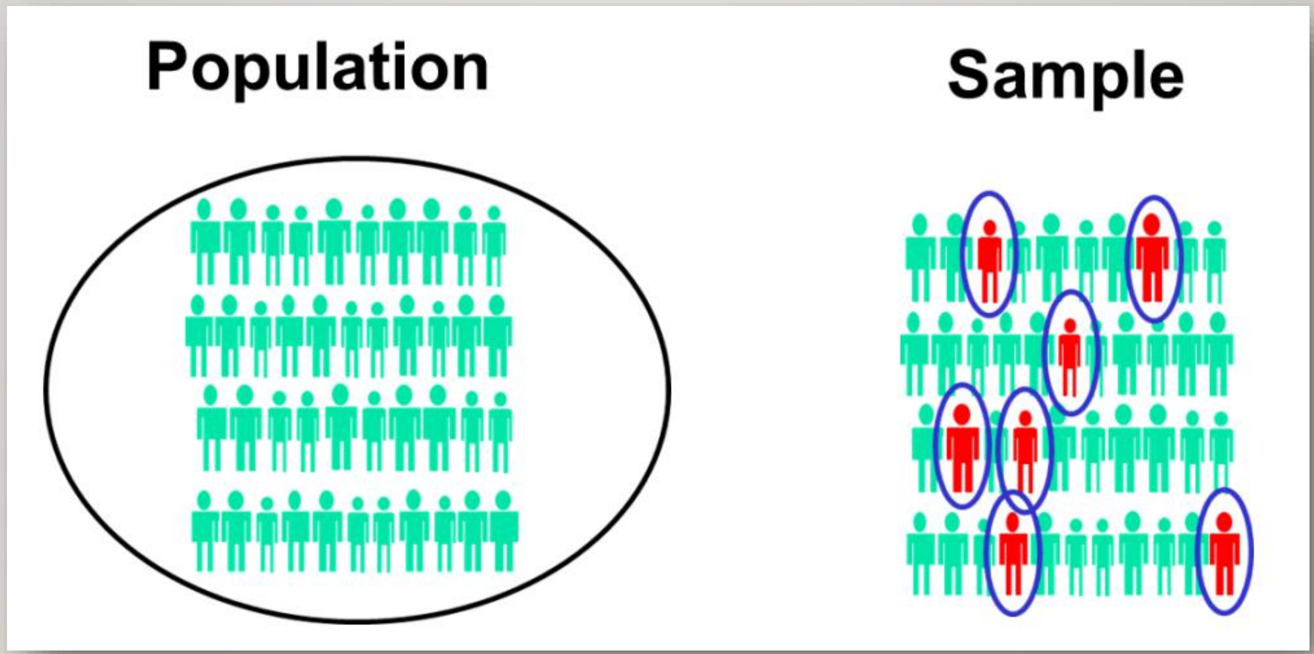
SAMPLING FROM A POPULATION

- A **Population** is the set of all items or individuals of interest
  - Examples: All likely voters in the next election
    All parts produced today
    All sales receipts for November

- A **Sample** is a subset of the population
  - Examples: 1000 voters selected at random for interview
    A few parts selected for destructive testing
    Random receipts selected for audit

Newbold et al (2013)

POPULATION VS. SAMPLE



Newbold et al (2013)

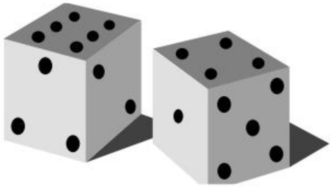
WHY SAMPLE?

- Less time consuming than a census
- Less costly to administer than a census
- It is possible to obtain statistical results of a sufficiently high precision based on samples.

Newbold et al (2013)

SIMPLE RANDOM SAMPLING

- Every object in the population has the same probability of being selected
- Objects are selected independently
- Samples can be obtained from a table of random numbers or computer random number generators



- A simple random sample is the ideal against which other sampling methods are compared

Newbold et al (2013)

Note:
- There are several sampling methods; **simple random sampling** is just one of them. The choice of method depends on the **objective of the study**.

SAMPLING DISTRIBUTIONS

- A sampling distribution is a probability distribution of all of the possible values of a statistic for a given size sample selected from a population

Newbold et al (2013)

Example of Sampling Distribution:

- Population mean weight: $\mu = 70$ kg
- Population standard deviation: $\sigma = 8$ kg
- Take random samples with sample size $n = 25$
- Compute the **sample mean ($\bar{X}$)** for each sample
- The distribution of these **sample means** is the **sampling distribution of $\bar{X}$**

Note: For large n, the sampling distribution is approximately normal.

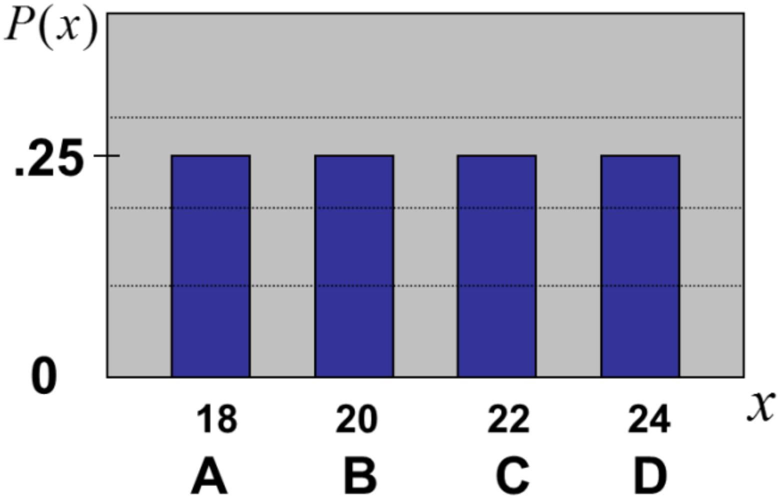
DEVELOPING A SAMPLING DISTRIBUTION

- **Assume there is a population …**
- Population size $N=4$
- Random variable, X, is age of individuals
- Values of X: 18, 20, 22, 24 (years)

Newbold et al (2013)

DEVELOPING A SAMPLING DISTRIBUTION

In this example the Population Distribution is uniform:



Uniform Distribution

Newbold et al (2013)

DEVELOPING A SAMPLING DISTRIBUTION

Now consider all possible samples of size $n = 2$

1st Obs	2nd Observation 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Newbold et al (2013)

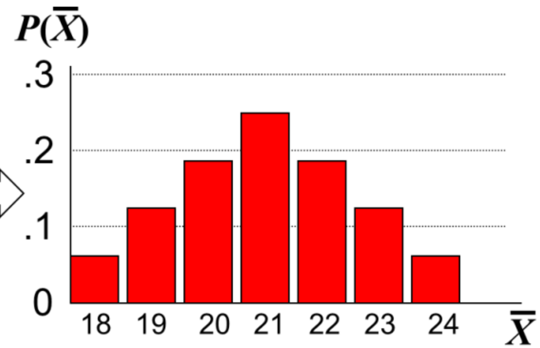
DEVELOPING A SAMPLING DISTRIBUTION

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation			
	18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Distribution of Sample Means



(no longer uniform)

Newbold et al (2013)

Sampling Distribution of the Sample Mean

The distribution of the variable "sample mean" ($\bar{X}$) over all possible samples is called the sampling distribution of $\bar{X}$.

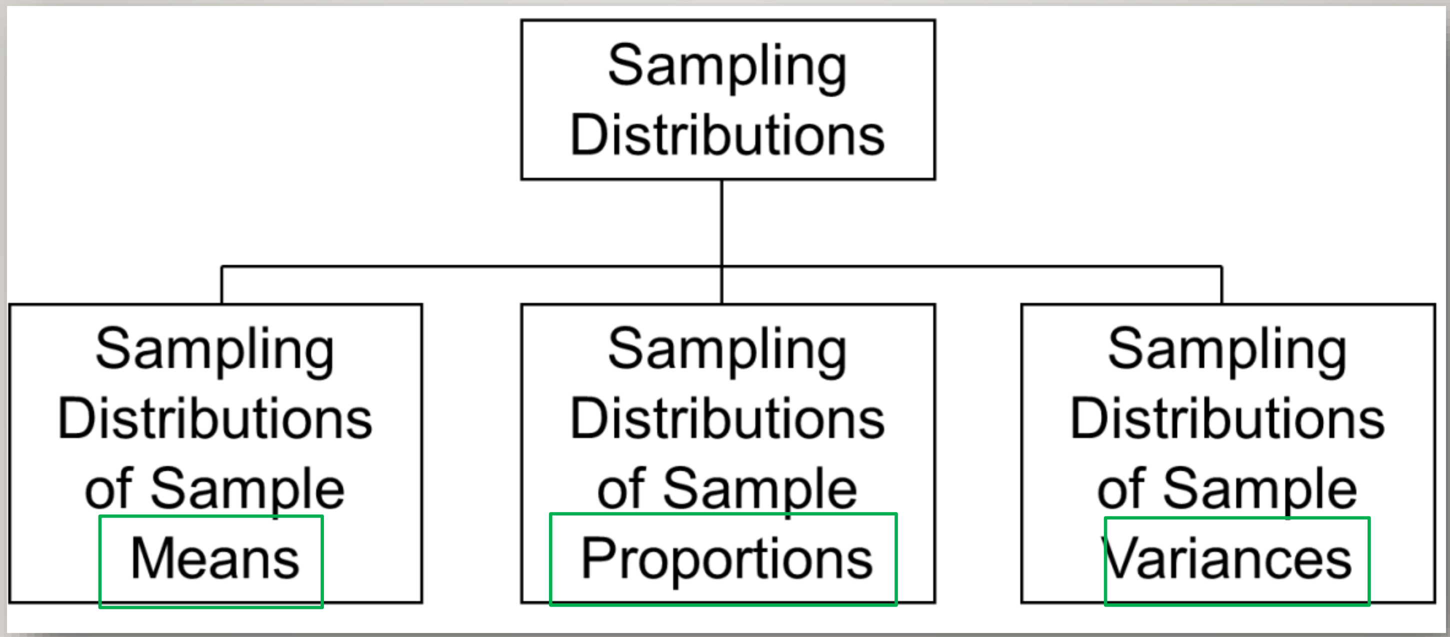
💡 **Note:** This distribution shows how $\bar{X}$ varies from sample to sample and has:

- Mean: $\mu_{\bar{X}} = \mu$
- Standard deviation (standard error): $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

Note:

- **$\bar{X}$ is a sample statistic.**
- **The sampling distribution of $\bar{X}$** is the distribution of all possible sample means from the population.

SAMPLING DISTRIBUTIONS



Newbold et al (2013)

LECTURE 1: SAMPLING DISTRIBUTIONS OF SAMPLE MEANS

SAMPLE MEAN

- Let $X_1, X_2, \ldots, X_n$ represent a random sample from a population
- The sample mean value of these observations is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Newbold et al (2013)

STANDARD ERROR OF THE MEAN

- Different samples of the same size from the same population will yield different sample means
- A measure of the variability in the mean from sample to sample is given by the Standard Error of the Mean:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

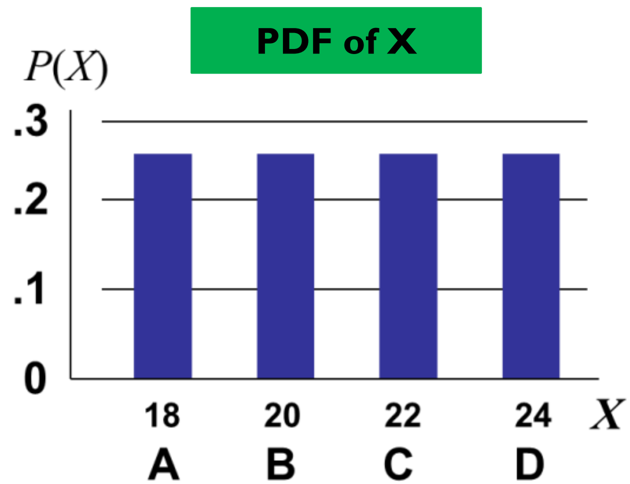
The **standard error of the mean** is the **standard deviation of the sample mean $\bar{X}$**.

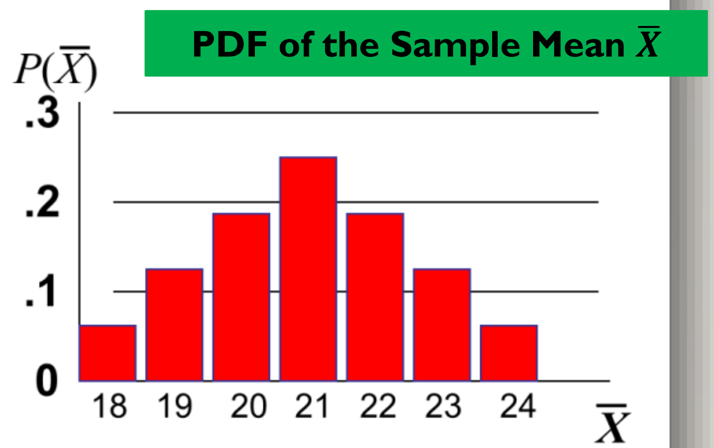
- Note that the standard error of the mean decreases as the sample size increases

Newbold et al (2013)

COMPARING THE POPULATION WITH ITS SAMPLING DISTRIBUTION

Population	Sample Means Distribution
$N = 4$	$n = 2$
$\mu = 21 \quad \sigma = 2.236$	$\mu_{\bar{X}} = 21 \quad \sigma_{\bar{X}} = 1.58$

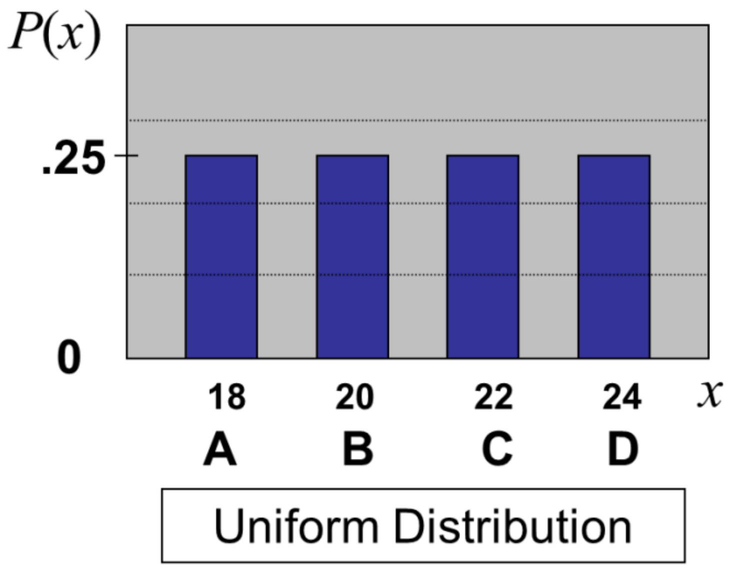




Newbold et al (2013)

MEASURES OF X

Summary Measures for the Population Distribution:



$$\mu = \frac{\sum X_i}{N}$$

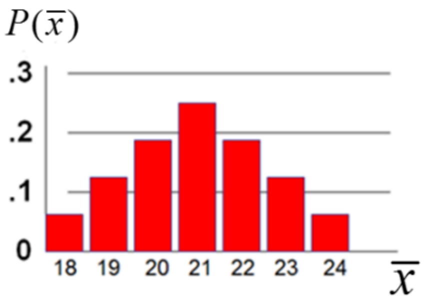
$$= \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} = 2.236$$

Newbold et al (2013)

MEASURES OF $\bar{X}$

Summary Measures of the Sampling Distribution:



$$E(\bar{X}) = \frac{\sum \bar{X}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21 = \mu$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum(\bar{X}_i - \mu)^2}{N}}$$

$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Newbold et al (2013)

IF SAMPLE VALUES ARE NOT INDEPENDENT

- If the sample size n is not a small fraction of the population size N, then individual sample members are not distributed independently of one another
- Thus, observations are not selected independently
- A finite population correction is made to account for this:

$$Var(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1} \quad \text{or} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

The term $\frac{(N-n)}{(N-1)}$ is often called a **finite population correction factor**

Newbold et al (2013)

IF THE POPULATION IS NORMAL

- If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{X}$ is also normally distributed with

$$\mu_{\bar{X}} = \mu \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- If the sample size *n* is not small relative to the population size *N*, then

$$\mu_{\bar{X}} = \mu \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Newbold et al (2013)

$X \sim$ Normal(μ, σ^2) → **$\bar{X} \sim$ Normal($\mu, \sigma^2/n$)**

STANDARD NORMAL DISTRIBUTION FOR THE SAMPLE MEANS

- Z-value for the sampling distribution of $\bar{X}$:

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

where: $\bar{X}$ = sample mean

μ = population mean

$\sigma_{\bar{X}}$ = standard error of the mean

Z is a standardized normal random variable with mean of 0 and a variance of 1

Newbold et al (2013)

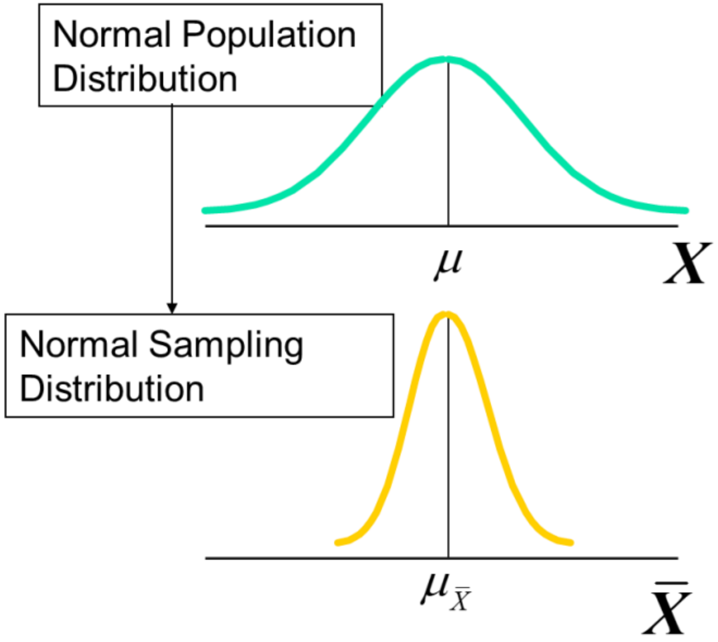
$\bar{X}$ ~ Normal(μ, σ^2/n)

$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ ~ Normal(0, 1)

SAMPLING DISTRIBUTION PROPERTIES

$$E\left[\bar{X}\right] = \mu$$

(i.e. $\bar{X}$ is unbiased)



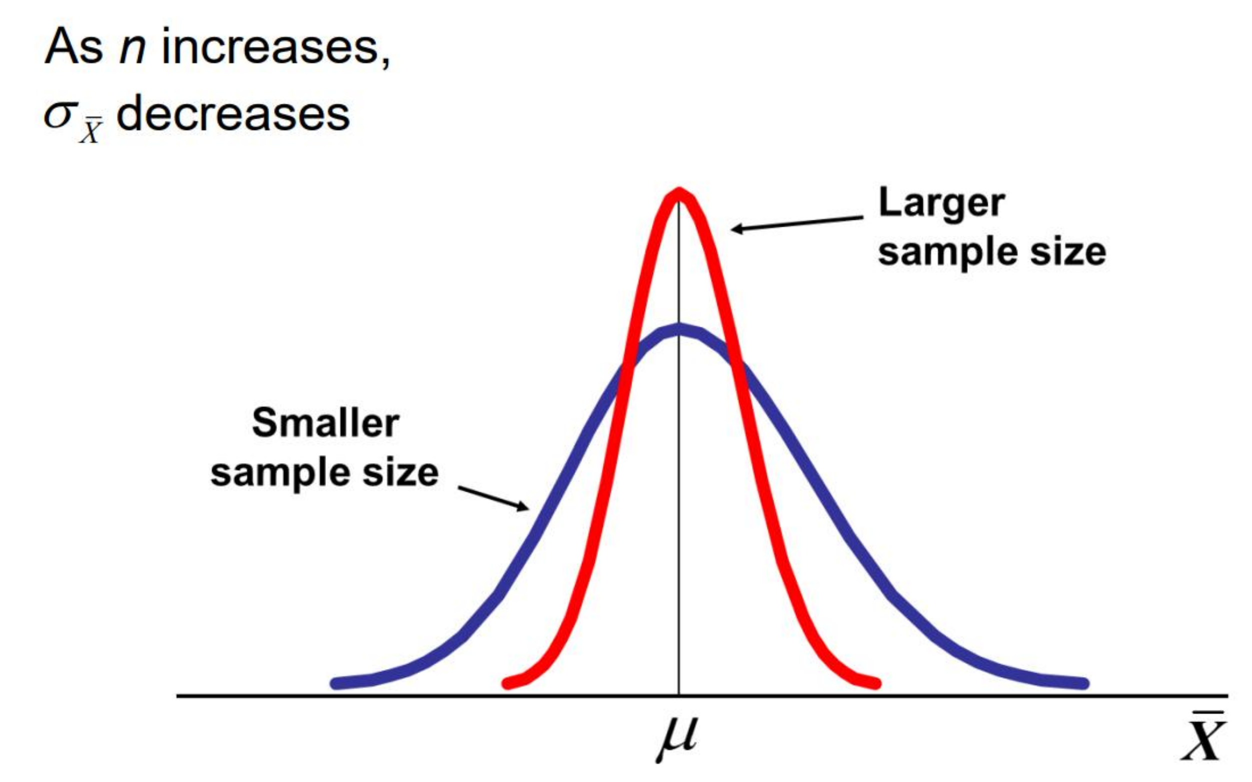
(the distribution of $\bar{X}$ has a reduced standard deviation)

(both distributions have the same mean)

Newbold et al (2013)

$X \sim$ **Normal(μ, σ^2)** → $\bar{X} \sim$ **Normal(μ, σ^2/n)**

SAMPLING DISTRIBUTION PROPERTIES



Newbold et al (2013)

CENTRAL LIMIT THEOREM (CLT)

- Even if the population is not normal,
- …sample means from the population will be approximately normal as long as the sample size is large enough.

Newbold et al (2013)

Central Limit Theorem (CLT)

If we take **samples of size n** from a population with **mean μ** and **standard deviation σ**, then the **sampling distribution of the sample mean** $\bar{X}$ will:

- Have **mean** $E(\bar{X}) = \mu$
- Have **standard deviation** $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
- And **approach a normal distribution** as n increases, **regardless of the population's shape**.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ for large } n$$

CENTRAL LIMIT THEOREM

- Let $X_1, X_2, \ldots, X_n$ be a set of *n* independent random variables having identical distributions with mean μ, variance σ^2, and $\bar{X}$ as the mean of these random variables.
- As *n* becomes large, the central limit theorem states that the distribution of

$$Z = \frac{\bar{X} - \mu_x}{\sigma_{\bar{X}}}$$

approaches the standard normal distribution

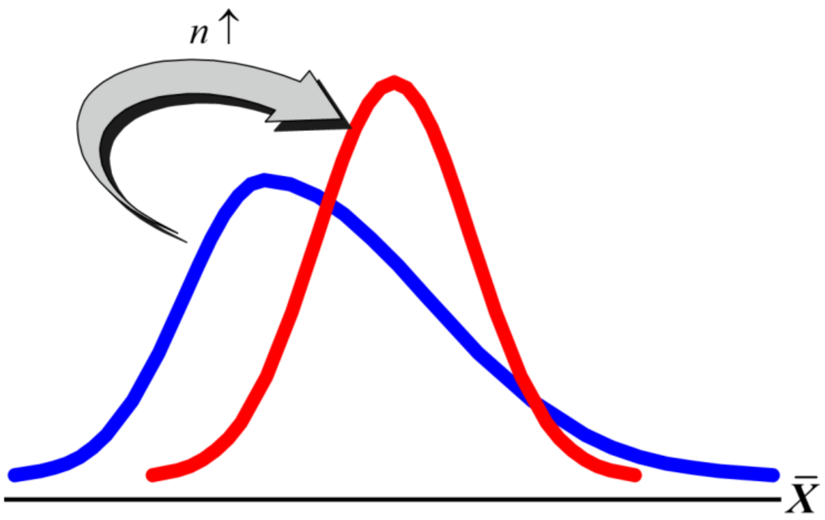
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \text{Normal}(0, 1)$$

Newbold et al (2013)

CENTRAL LIMIT THEOREM

As the sample size gets large enough…

the sampling distribution becomes almost normal regardless of shape of population



Newbold et al (2013)

IF THE POPULATION IS NOT NORMAL

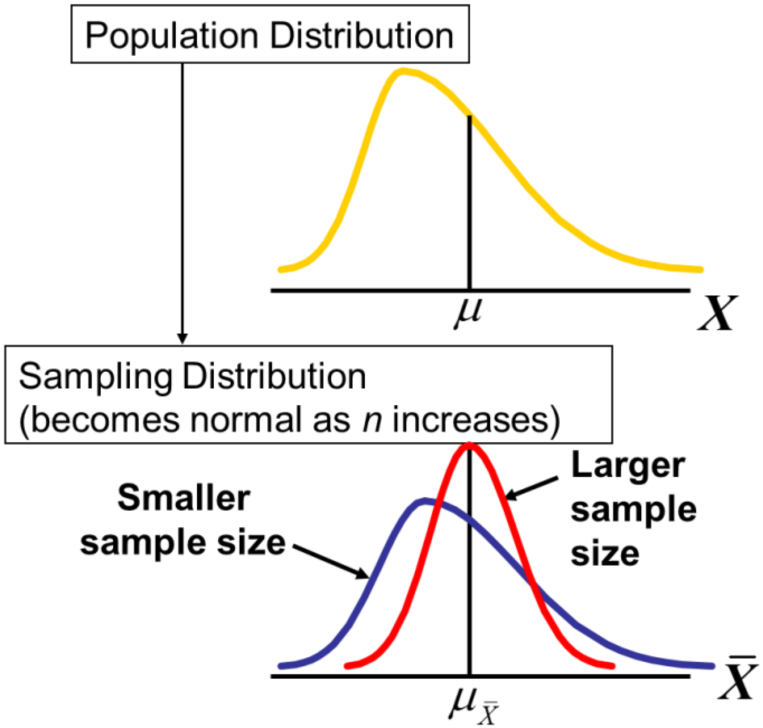
Sampling distribution properties:

Central Tendency

$$\mu_{\bar{X}} = \mu$$

Variation

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$



Newbold et al (2013)

HOW LARGE IS LARGE ENOUGH?

- For most distributions, $n > 25$ will give a sampling distribution that is nearly normal
- For normal population distributions, the sampling distribution of the mean is always normally distributed

Newbold et al (2013)

CENTRAL LIMIT THEOREM: EXAMPLE

- Suppose a large population has mean $\mu = 8$ and standard deviation $\sigma = 3$. suppose a random sample of size n = 36 is selected.
- What is the probability that the sample mean is between 7.8 and 8.2?

Newbold et al (2013)

CENTRAL LIMIT THEOREM: EXAMPLE

Solution:

- Even if the population is not normally distributed, the central limit theorem can be used (n > 25)
- … so the sampling distribution of $\bar{X}$ is approximately normal
- … with mean $\mu_{\bar{X}} = 8$
- …and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

Newbold et al (2013)

According to the CLT, the sample mean is approximately normal when n > 25.

$$\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\mathbf{Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim Normal(0, 1)}$$

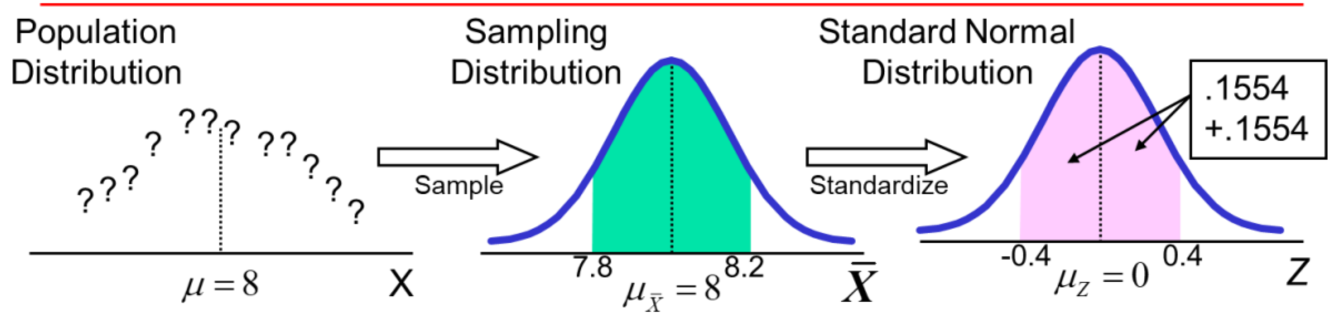
CENTRAL LIMIT THEOREM: EXAMPLE

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \text{Normal}(0, 1)$$

Solution: (continued):

$$P(7.8 < \bar{X} < 8.2) = P\left(\frac{7.8 - 8}{\frac{3}{\sqrt{36}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{8.2 - 8}{\frac{3}{\sqrt{36}}} \right)$$

$$= P(-0.4 < Z < 0.4) = \boxed{0.3108}$$



Newbold et al (2013)

EXERCISE 6.7

6.7 Given a population with a mean of $\mu = 200$ and a variance of $\sigma^2 = 625$, the central limit theorem applies when the sample size $n \geq 25$. A random sample of size $n = 25$ is obtained.

a. What are the mean and variance of the sampling distribution for the sample mean?
b. What is the probability that $\bar{x} > 209$?
c. What is the probability that $198 \leq \bar{x} \leq 211$?
d. What is the probability that $\bar{x} \leq 202$?

Newbold et al (2013)

EXERCISE 6.7 A): SOLUTION

Answer:

Given: population mean $\mu = 200$, population variance $\sigma^2 = 625$ so $\sigma = 25$. Sample size $n = 25$.

The sampling distribution of the sample mean is

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) = N\left(200,\ \frac{625}{25}\right) = N(200,\ 25).$$

Hence the standard error is $\sigma_{\bar{X}} = \sqrt{25} = 5$.

(a) Mean and variance of the sampling distribution

- Mean: $E(\bar{X}) = 200$.
- Variance: $\mathrm{Var}(\bar{X}) = 25$. (standard deviation $= 5$.)

$\bar{X}$ ~ Normal(μ, σ^2/n) → **$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ ~ Normal(0, 1)**

EXERCISE 6.7 B): SOLUTION

Answer:

Given: population mean $\mu = 200$, population variance $\sigma^2 = 625$ so $\sigma = 25$. Sample size $n = 25$.

The sampling distribution of the sample mean is

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) = N\left(200,\ \frac{625}{25}\right) = N(200,\ 25).$$

Hence the standard error is $\sigma_{\bar{X}} = \sqrt{25} = 5$.

(b) $P(\bar{X} \geq 209)$

Compute the z-score:

$$z = \frac{209 - 200}{5} = \frac{9}{5} = 1.8.$$

So

$$P(\bar{X} \geq 209) = 1 - \Phi(1.8) \approx 1 - 0.96407 = 0.03593.$$

Answer: $\boxed{0.0359\ (\approx 3.59\%)}$.

Standard Normal Distribution Table

EXERCISE 6.7 C): SOLUTION

Answer:

Given: population mean $\mu = 200$, population variance $\sigma^2 = 625$ so $\sigma = 25$. Sample size $n = 25$.

The sampling distribution of the sample mean is

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) = N\left(200,\ \frac{625}{25}\right) = N(200,\ 25).$$

Hence the standard error is $\sigma_{\bar{X}} = \sqrt{25} = 5$.

(c) $P(198 \le \bar{X} \le 211)$

Z-scores:

$$z_{\text{low}} = \frac{198 - 200}{5} = -0.4, \qquad z_{\text{up}} = \frac{211 - 200}{5} = 2.2.$$

So

$$P(198 \le \bar{X} \le 211) = \Phi(2.2) - \Phi(-0.4) \approx 0.98610 - 0.34458 = 0.64152.$$

Answer: $\boxed{0.6415\ (\approx 64.15\%)}$.

Standard Normal Distribution Table

EXERCISE 6.7 D): SOLUTION

Answer:

Given: population mean $\mu = 200$, population variance $\sigma^2 = 625$ so $\sigma = 25$. Sample size $n = 25$.

The sampling distribution of the sample mean is

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) = N\left(200,\ \frac{625}{25}\right) = N(200,\ 25).$$

Hence the standard error is $\sigma_{\bar{X}} = \sqrt{25} = 5$.

(d) $P(\bar{X} \le 202)$

Z-score:

$$z = \frac{202 - 200}{5} = 0.4.$$

So

$$P(\bar{X} \le 202) = \Phi(0.4) \approx 0.65542.$$

Answer: $\boxed{0.6554\ (\approx 65.54\%)}$.

Standard Normal Distribution Table

HOMEWORK OF LECTURE 2: QUESTIONS

EXERCISE 6.18

6.18 An industrial process produces batches of a chemical whose impurity levels follow a normal distribution with standard deviation 1.6 grams per 100 grams of chemical. A random sample of 100 batches is selected in order to estimate the population mean impurity level.

a. The probability is 0.05 that the sample mean impurity level exceeds the population mean by how much?
b. The probability is 0.10 that the sample mean impurity level is below the population mean by how much?
c. The probability is 0.15 that the sample mean impurity level differs from the population mean by how much?

Newbold et al (2013)

EXERCISE 6.18 A): SOLUTION

Answer:

Given: population standard deviation σ = 1.6 g per 100 g, sample size n = 100.

Standard error:

$$\sigma_{\bar{X}} = \frac{1.6}{\sqrt{100}} = \frac{1.6}{10} = 0.16$$

$\bar{X} \sim$ Normal(μ, 0.16^2)

(a) Probability is 0.05 that $\bar{X}$ exceeds μ by how much?

We need $P(\bar{X} - \mu > E) = 0.05.$

$$P\left(\frac{\bar{X}-\mu}{0.16} > \frac{E}{0.16}\right) = 0.05 \Leftrightarrow P(Z > E/0.16) = 0.05$$

Then, $E/0.16 = z_{0.95} = 1.645 \Leftrightarrow E = 0.16 \times 1.645 = 0.2632$

EXERCISE 6.17

6.17 The times spent studying by students in the week before final exams follows a normal distribution with standard deviation 8 hours. A random sample of four students was taken in order to estimate the mean study time for the population of all students.

a. What is the probability that the sample mean exceeds the population mean by more than 2 hours?
b. What is the probability that the sample mean is more than 3 hours below the population mean?
c. What is the probability that the sample mean differs from the population mean by more than 4 hours?
d. Suppose that a second (independent) random sample of 10 students was taken. Without doing the calculations, state whether the probabilities in parts (a), (b), and (c) would be higher, lower, or the same for the second sample.

Newbold et al (2013)

THANKS!

Questions?