STATISTICS II

Lisbon School of Economics & Management
Universidade de Lisboa

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

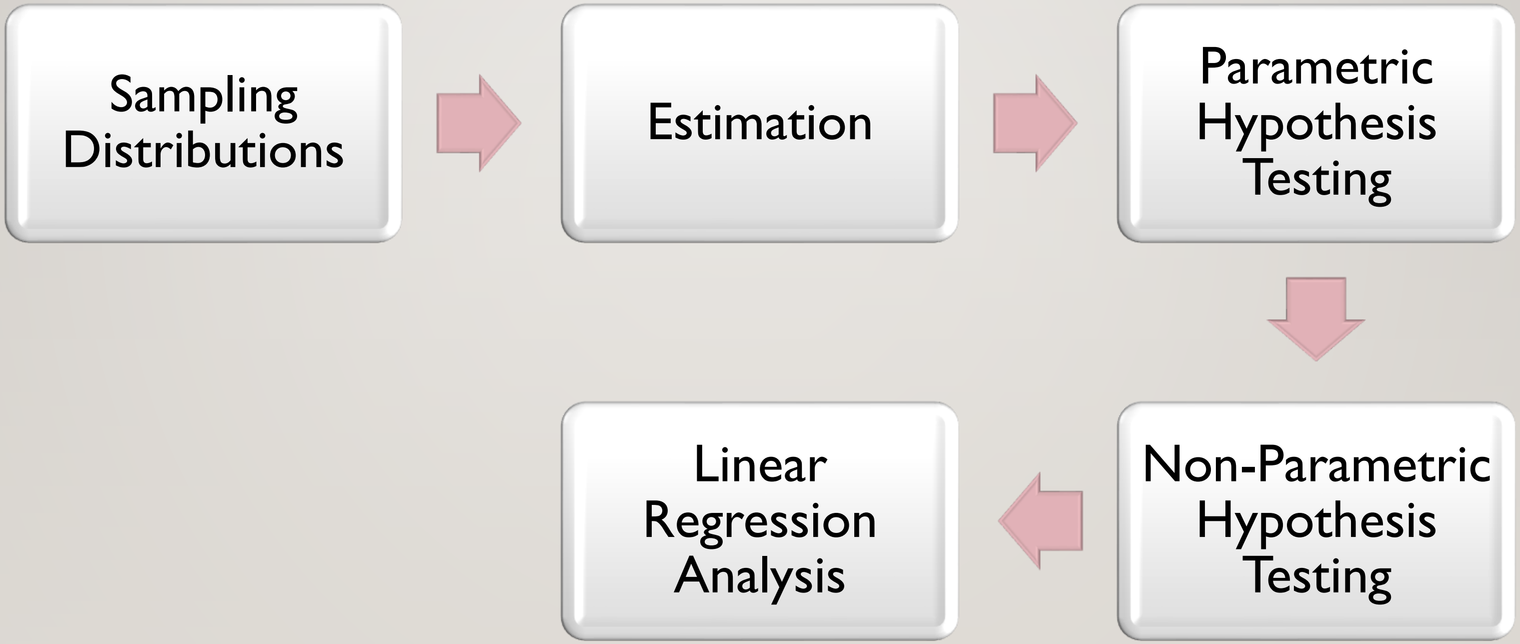
CONTACT

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

LECTURE 7 HOMEWORK: QUESTIONS AND SOLUTIONS

EXERCISE 7.70

7.70 The student government association at a university wants to estimate the percentage of the student body that supports a change being considered in the academic calendar of the university for the next academic year. How many students should be surveyed if a 90% confidence interval is desired and the margin of error is to be only 3%?

Newbold et al (2013)

EXERCISE 7.70: SOLUTION

Answer:

To determine the required sample size for estimating a **population proportion**, we use the formula

$$n = \left(\frac{z_{1-\alpha/2}}{\text{ME}}\right)^2 p(1-p)$$

Given information

- Confidence level: **90%** $\Rightarrow \alpha = 0.10$
- Margin of error: **ME = 0.03**
- Population proportion p: **unknown**

When p is unknown, we use the **conservative value**

$$p = 0.5$$

because it maximizes $p(1-p)$ and ensures a sufficiently large sample size.

For a 90% confidence level:

$$z_{1-\alpha/2} = z_{0.95} \approx 1.645$$

Sample size calculation

$$n = \left(\frac{1.645}{0.03}\right)^2 (0.5)(0.5)$$

$$n = (54.83)^2 \times 0.25 \approx 751.7$$

Final answer

$$\boxed{n = 752}$$

(The sample size is rounded **up** to ensure the desired margin of error.)

Interpretation

At least **752 students** should be surveyed to estimate the proportion of students who support the change with **90% confidence** and a **margin of error of 3%**.

EXERCISE 41

A security card issuing center has **two personalization machines**, operating independently. The processing time (in seconds) for each machine is normally distributed with the **same mean**, with a standard deviation of 10 seconds for the first machine and 15 seconds for the second.

Samples of 16 cards are taken from each machine.

Questions:

a) Calculate the probability that the **absolute difference between the sample means** of the two machines exceeds 5 seconds.

b) What is the probability that the **sample standard deviation** of the first machine is greater than that of the second machine?

Murteira et al (2015), Chapter 6

Note: Only complete part (a) of Exercise 41, as part (b) involves topics not yet covered.

EXERCISE 41 A): SOLUTION

✅ Answer:

Given:

- Machine 1: $\sigma_1 = 10$, $n_1 = 16$
- Machine 2: $\sigma_2 = 15$, $n_2 = 16$
- Both populations are normal with the **same mean** μ.

a) Probability that $|\bar{X}_1 - \bar{X}_2| > 5$

1. **Distribution of the difference of means**

$$D = \bar{X}_1 - \bar{X}_2 \sim N\left(0, \sigma_D^2\right)$$

$$\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{10^2}{16} + \frac{15^2}{16}} = \sqrt{\frac{100}{16} + \frac{225}{16}} = \sqrt{\frac{325}{16}}$$

$$\sigma_D = \sqrt{20.3125} \approx 4.51$$

2. **Standardize for Z**

$$P(|\bar{X}_1 - \bar{X}_2| > 5) = P(D > 5) + P(D < -5) = 2 \cdot P(D > 5)$$

$$Z = \frac{5 - 0}{4.51} \approx 1.11$$

$$P(D > 5) = P(Z > 1.11) \approx 0.1335$$

$$\boxed{P(|\bar{X}_1 - \bar{X}_2| > 5) = 2 \cdot 0.1335 \approx 0.267}$$

✅ Answer (a): $\mathbf{P} \approx \mathbf{0.267}$ (~26.7%)

EXERCISE 41 A): SOLUTION

Answer:

Alternative Solution:

Given:

- Machine 1: $\sigma_1 = 10$, $n_1 = 16$
- Machine 2: $\sigma_2 = 15$, $n_2 = 16$
- Both populations are normal with the **same mean** μ.

$X_1 \sim N(\mu_1, 10^2) \rightarrow$ Amostra casual : $m = 16$

$X_2 \sim N(\mu_2, 15^2) \rightarrow$ Amostra casual : $n = 16$

, onde $\mu_1 = \mu_2$

(a)

Quer-se : $P(|\bar{X}_1 - \bar{X}_2| > 5) = 1 - P(|\bar{X}_1 - \bar{X}_2| \leq 5) = 1 - P(-5 \leq \bar{X}_1 - \bar{X}_2 \leq 5)$

Sabe-se que : $Z = \dfrac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0,1)$. Logo,

EXERCISE 41 A): SOLUTION

Answer:

Alternative Solution:

$$1-P\left(-5\le \bar{X}_1-\bar{X}_2\le 5\right)=1-P\left(\frac{-5-0}{\sqrt{\frac{10^2}{16}+\frac{15^2}{16}}}\le \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\le \frac{5-0}{\sqrt{\frac{10^2}{16}+\frac{15^2}{16}}}\right)=$$

$$=1-P\left(\underset{(-1.11)}{-\frac{5}{\sqrt{20.3125}}}\le Z\le \underset{(1.11)}{\frac{5}{\sqrt{20.3125}}}\right)=1-\left[\Phi(1.11)-\Phi(-1.11)\right]=$$

$$=1-\left[2\Phi(1.11)-1\right]=2-2\Phi(1.11)=2-2\times 0.8665=\boxed{0.267}$$

LECTURE 8: SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS

DIFFERENCE OF TWO SAMPLE PROPORTIONS

Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be two independent random samples of sizes n_1 and n_2 drawn from **two Bernoulli populations** with parameters p_1 and p_2, respectively, where X_{ij} takes the value 1 for a success and 0 for a failure. Define the **sample proportions**:

$$\hat{p}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \quad \hat{p}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}.$$

If the sample sizes are large, then by the **De Moivre–Laplace theorem**:

$$\hat{p}_1 \stackrel{\cdot}{\sim} N\left(p_1, \sqrt{\frac{p_1(1-p_1)}{n_1}}\right), \quad \hat{p}_2 \stackrel{\cdot}{\sim} N\left(p_2, \sqrt{\frac{p_2(1-p_2)}{n_2}}\right),$$

and by the **additivity property of the normal distribution**:

$$\hat{p}_1 - \hat{p}_2 \stackrel{\cdot}{\sim} N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right),$$

that is,

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \stackrel{\cdot}{\sim} N(0,1).$$

Standard Deviation

Note: For large samples (n_1 > 25 and n_2 > 25), the sample proportions $\hat{p}_1$ and $\hat{p}_2$ have an approximately **normal distribution**, by the Central Limit Theorem (CLT).

Note: The dot (or the letter *"a"*) above the distribution symbol denotes that the distribution is **approximately normal**.

Note: The normal distribution can be denoted as $N(\mu, \sigma^2)$ or $N(\mu, \sigma)$,depending on whether the second parameter represents the **variance** or the **standard deviation**.

EXERCISE 1: DIFFERENCE OF TWO SAMPLE PROPORTIONS

The proportion of customers who chose the Noko brand at the **TeleMN** store was 0.35, while at the **Optcel** store it was 0.29.

Question: If a sample of 200 customers is taken from TeleMN and a sample of 150 customers from Optcel, what is the probability that the **sample proportion of Noko customers at TeleMN exceeds the sample proportion at Optcel**?

ProbabilidadesEstatistica2019.pdf

EXERCISE 1: SOLUTION

Answer:

Step 1: Define the sample proportions

Let

$$\hat{p}_1 = \text{sample proportion of Noko customers at TeleMN}, \quad n_1 = 200$$

$$\hat{p}_2 = \text{sample proportion of Noko customers at Optcel}, \quad n_2 = 150$$

We are asked to find

$$P(\hat{p}_1 > \hat{p}_2) = P(\hat{p}_1 - \hat{p}_2 > 0)$$

EXERCISE 1: SOLUTION

Answer:

Step 2: Compute the mean and standard deviation of the difference

The difference of sample proportions:

$$D = \hat{p}_1 - \hat{p}_2$$

$$E(D) = p_1 - p_2 = 0.35 - 0.29 = 0.06$$

$$\mathrm{Var}(D) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$\frac{p_1(1-p_1)}{n_1} = \frac{0.35 \cdot 0.65}{200} = \frac{0.2275}{200} \approx 0.0011375$$

$$\frac{p_2(1-p_2)}{n_2} = \frac{0.29 \cdot 0.71}{150} = \frac{0.2059}{150} \approx 0.0013727$$

$$\mathrm{Var}(D) = 0.0011375 + 0.0013727 \approx 0.0025102$$

$$\sigma_D = \sqrt{0.0025102} \approx 0.0501$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \mathrel{\dot\sim} N(0,1).$$

Standard Deviation of the Difference of Two Sample Proportions

EXERCISE 1: SOLUTION

✅ Answer:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \stackrel{\cdot}{\sim} N(0,1).$$

Step 3: Standardize to Z

$$Z = \frac{0 - E(D)}{\sigma_D} = \frac{0 - 0.06}{0.0501} \approx -1.1976$$

Step 4: Compute probability

$$P(\hat{p}_1 - \hat{p}_2 > 0) = P(Z > -1.198)$$

From standard normal tables:

P(Z > -1.198) = 1- P(Z < -1.198) = 1- $\Phi(-1.198) = 1 - [1 - \Phi(1.198)]$
$= \Phi(1.198) \sim 0.883$

Standard Normal Distribution Table

Answer

$$P(\hat{p}_1 > \hat{p}_2) \approx 0.883 \ (or\ 88.3\%)$$

Interpretation:

There is approximately an **88% chance** that the sample proportion of Noko customers at TeleMN will be higher than at Optcel, given the sample sizes and population proportions.

EXERCISE 1: SOLUTION

Answer:

Alternative Solution:

$p_1 = 0{,}35$ e $p_2 = 0{,}29$.

X_1 e X_2 dist. Bernoulli
n_1 (= 200) e n_2 (= 150) grandes

$$\Rightarrow Z = \frac{(\overline{P}_1 - \overline{P}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \overset{\circ}{\sim} N(0;1).$$

$$P(\overline{P}_1 > \overline{P}_2) = P(\overline{P}_1 - \overline{P}_2 > 0) = 1 - P(\overline{P}_1 - \overline{P}_2 \leq 0) = 1 - P\left(Z \leq \frac{0 - (0{,}35 - 0{,}29)}{\sqrt{\frac{0{,}35(1-0{,}35)}{200} + \frac{0{,}29(1-0{,}29)}{150}}}\right)$$

$$= 1 - \Phi(-1{,}2) = 1 - (1 - \Phi(1{,}2)) = \Phi(1{,}2) = 0{,}8849.$$

ProbabilidadesEstatistica2019.pdf

LECTURE 8: SAMPLING DISTRIBUTION OF THE RATIO OF TWO SAMPLE VARIANCES

RATIO OF TWO SAMPLE VARIANCES

Let

$$X_{11}, X_{12}, \ldots, X_{1n_1} \quad \text{and} \quad X_{21}, X_{22}, \ldots, X_{2n_2}$$

be **two independent random samples** of sizes n_1 and n_2, drawn from **two normal populations**, with

$$X_1 \sim N(\mu_1, \sigma_1^2) \quad \text{and} \quad X_2 \sim N(\mu_2, \sigma_2^2),$$

respectively.

Let the **corrected sample variances** be defined by

Note: The corrected sample variances are those computed by dividing by $n-1$, rather than by n.

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_{1i} - \bar{X}_1)^2, \qquad S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (X_{2i} - \bar{X}_2)^2.$$

Then, by a theorem concerning the **F distribution**, it follows that

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,\, n_2-1}.$$

Note: The variable F has a Snedecor's F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

ProbabilidadesEstatistica2019.pdf

EXERCISE 1: RATIO OF TWO SAMPLE VARIANCES

A pharmaceutical company launched a new sleep medication that has been used in hospitals. It was observed that:

- Patients **not taking the medication** slept on average 7.5 hours with a standard deviation of 1.4 hours
- Patients **taking the medication** slept on average 8 hours with a standard deviation of 2 hours

At a particular hospital, **samples** were taken:

- $n_1 = 31$ patients not taking the medication
- $n_2 = 61$ patients taking the medication

Question: Assuming normality of the data, what is the probability that the **sample variance of the first group is smaller than that of the second group**?

ProbabilidadesEstatistica2019.pdf

EXERCISE 1: SOLUTION

Note: The F table provides **right-tail probabilities**, that is, values of the form $P(F > a)$.
Therefore, it is convenient to express the probability in this form.
In this case, the numerator degrees of freedom are 30 and the denominator degrees of freedom are 60.
The value in the table closest to 2.04 is 2.03, which corresponds to a right-tail probability of 0.01.
Hence, this value is used.

Answer:

Step 1: Define the random variable

Let S_1^2 and S_2^2 be the **sample variances** of the two groups.

For independent samples from normal populations, the ratio

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1,\, n_2-1}.$$

where F_{df_1, df_2} is the **F-distribution** with $df_1 = n_2 - 1$ and $df_2 = n_1 - 1$ degrees of freedom.

We are asked:

$$P(S_1^2 < S_2^2) = P\left(\frac{S_1^2}{S_2^2} < 1\right)$$

X_1 e X_2 dist. Normal
$n_1 = 31$ e $n_2 = 61$

$$\Rightarrow F = \frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2} \sim F_{n_1-1;\, n_2-1 = 30;60}$$

F Distribution Table

$$P(S_1^2 < S_2^2) = P\left(\frac{S_1^2}{S_2^2} < 1\right) = P\left(F < 1 \times \frac{2^2}{1.4^2}\right) = P(F < 2{,}04) = 1 - P(F > 2.04) \sim 1 - 0.01 = 0.99$$

EXERCISE 1: SOLUTION

Answer:

Answer

$$P(S_1^2 < S_2^2) \approx 0.99 \text{ (or 99\%)}$$

Interpretation:

There is a **very high probability** that the sample variance of patients **not taking the medication** is smaller than that of patients **taking the medication**, which aligns with the population variances (1.96 vs 4).

LECTURE 8 HOMEWORK: QUESTIONS

EXERCISE 27

After an intensive advertising campaign, the market share of "Crispy Chips" increased from 8% to 10%. Suppose two independent surveys were conducted:

- **Before the campaign:** sample of size $n_1 = 100$
- **After the campaign:** sample of size $n_2 = 300$

Questions:

a) What is the probability that the surveys would conclude that the **market share gain exceeded 5 percentage points**?

b) What is the probability that the surveys would conclude a **loss in market share**?

Murteira et al (2015), Chapter 6

EXERCISE 27 A): SOLUTION

Answer:

Step 1: Define sample proportions

Let

$$\hat{p}_1 = \text{sample proportion before campaign}, \quad \hat{p}_2 = \text{sample proportion after campaign}$$

- Before: $p_1 = 0.08, n_1 = 100$
- After: $p_2 = 0.10, n_2 = 300$

We want probabilities involving

P($\hat{p}_2$ - $\hat{p}_1$ > 0.05)

$$D = \hat{p}_2 - \hat{p}_1$$

Step 2: Approximate distribution of difference

For large samples,

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right), \quad \hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \dot{\sim} N(0,1).$$

$$D = \hat{p}_2 - \hat{p}_1 \sim N\left(p_2 - p_1, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

EXERCISE 27 A): SOLUTION

Answer:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \stackrel{.}{\sim} N(0,1).$$

Step 3: Compute mean and standard deviation of difference

Mean of the Difference of Two Sample Proportions

$$E(D) = p_2 - p_1 = 0.10 - 0.08 = 0.02$$

$$\frac{p_1(1-p_1)}{n_1} = \frac{0.08 \cdot 0.92}{100} = 0.000736$$

$$\frac{p_2(1-p_2)}{n_2} = \frac{0.10 \cdot 0.90}{300} = 0.0003$$

Standard Deviation of the Difference of Two Sample Proportions

$$\sigma_D = \sqrt{0.000736 + 0.0003} = \sqrt{0.001036} \approx 0.0322$$

Step 4: Part a — Probability that gain > 5 points

We want $P(D > 0.05)$:

$$Z = \frac{0.05 - E(D)}{\sigma_D} = \frac{0.05 - 0.02}{0.0322} \approx \frac{0.03}{0.0322} \approx 0.932$$

Standard Normal Distribution Table

$P(\hat{p}_2 - \hat{p}_1 > 0.05) = P(D > 0.05) =$ P(Z > 0.932) = 1- P(Z < 0.932) = 1-$\Phi(0.932) \sim 1 - 0.8238 = 0.176$

Answer (a): 17.6%

> Interpretation: There is about a 17.6% chance that the surveys would conclude a market share gain larger than 5 points.

EXERCISE 27 B): SOLUTION

Answer:

Step 5: Part b — Probability of concluding a loss

$P(\hat{p}_2 - \hat{p}_1 < 0)$

We want $P(D < 0)$:

$$D = \hat{p}_2 - \hat{p}_1$$

$$Z = \frac{0 - E(D)}{\sigma_D} = \frac{0 - 0.02}{0.0322} \approx -0.621$$

$$P(\hat{p}_2 - \hat{p}_1 < 0) = P(D < 0) = P(Z < -0.621) \approx 0.267$$

Answer (b): 26.7%

> Interpretation: There is about a 27% chance that the surveys would erroneously conclude a **loss in market share**, even though the true gain was 2 points.

EXERCISE 41

A security card issuing center has **two independent personalization machines**. The processing time (in seconds) for each machine is **normally distributed** with the **same mean**, but different standard deviations:

- Machine 1: $\sigma_1 = 10$
- Machine 2: $\sigma_2 = 15$

Samples of **16 cards** are taken from each machine.

Questions:

a) What is the probability that the **absolute difference between the two sample means** exceeds 5 seconds?

b) What is the probability that the **sample standard deviation of the processing times for machine 1** is greater than that of machine 2?

Murteira et al (2015), Chapter 6

Note: Part (a) was solved earlier in a similar context and will not be repeated here.

EXERCISE 41 B): SOLUTION

Answer:

Given:

$$n_1 = n_2 = 16, \quad \sigma_1 = 10, \ \sigma_2 = 15$$

Let $\bar{X}_1$ and $\bar{X}_2$ be the **sample means**.

Part (b) — Probability that $S_1 > S_2$

Step 1: Recall the F-distribution for sample variances

For independent normal samples:

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_1-1, n_2-1}$$

- Here, $df_1 = n_1 - 1 = 15$, $df_2 = n_2 - 1 = 15$
- Population variances: $\sigma_1^2 = 100$, $\sigma_2^2 = 225$

We want:

$$P(S_1 > S_2) = P\left(\frac{S_1^2}{S_2^2} > 1\right)$$

EXERCISE 41 B): SOLUTION

✅ Answer:

Step 2: Convert to standard F-distribution

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} = \frac{S_1^2}{S_2^2} \cdot \frac{225}{100} = \frac{S_1^2}{S_2^2} \cdot 2.25$$

$$P(S_1^2 > S_2^2) = P(F_{15,15} > 2.25)$$

From **F-distribution tables or software**:

$$P(F_{15,15} > 2.25) \approx 0.044$$

✅ **Answer (b):** 0.044 (or 4.4%)

> Interpretation: It is **unlikely (4.4%)** that the sample standard deviation of machine 1 exceeds that of machine 2, which aligns with the population variances ($\sigma_1^2 < \sigma_2^2$).

LECTURE 9: CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATIONS MEANS

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS: INDEPENDENT SAMPLES

Population means, independent samples

Confidence Interval Estimation of the Difference Between Two Normal Population Means: Independent Samples

Goal: Form a confidence interval for the difference between two population means, $\mu_1 - \mu_2$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS: INDEPENDENT SAMPLES

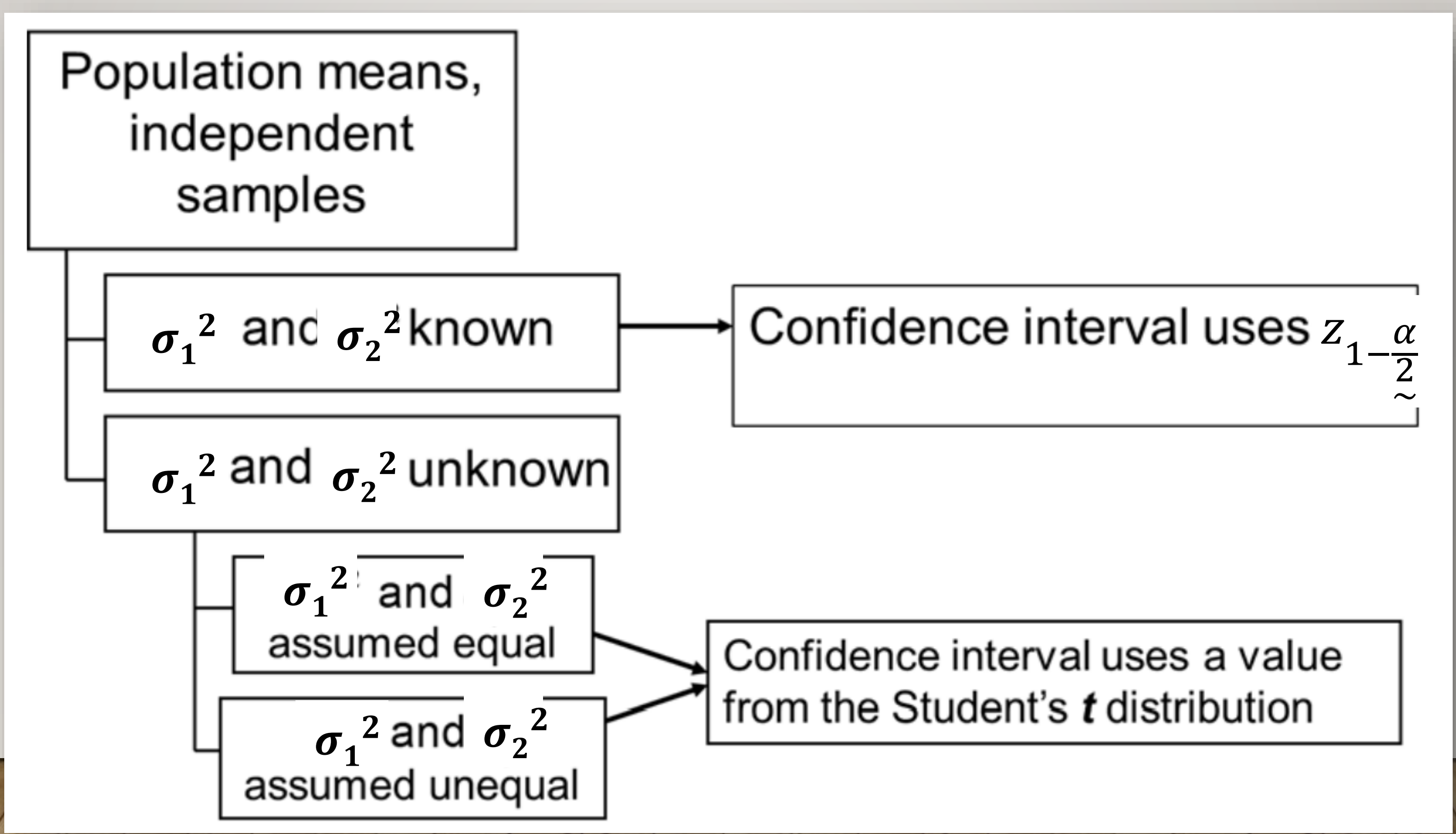
Population means, independent samples

Goal: Form a confidence interval for the difference between two population means, $\mu_1 - \mu_2$

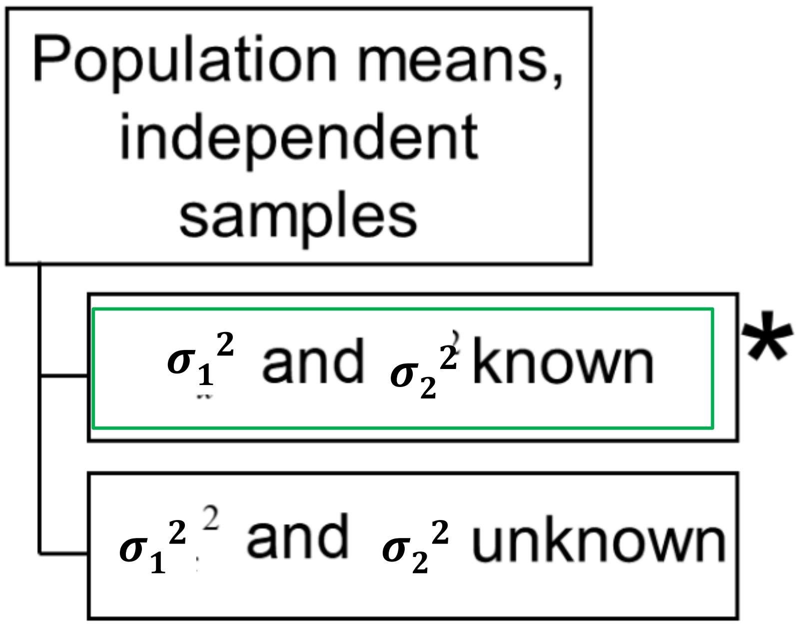
- Different data sources
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
- The point estimate is the difference between the two sample means:

$$\bar{x} - \bar{y}$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS: INDEPENDENT SAMPLES



CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 KNOWN)



Assumptions:

- Samples are randomly and independently drawn
- both population distributions are normal
- Population variances are known

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 KNOWN)

Population means, independent samples

- σ_1^2 and σ_2^2 known *
- σ_1^2 and σ_2^2 unknown

When σ_1^2 and σ_2^2 are known and both populations are normal, the variance of $(\bar{X}_1 - \bar{X}_2)$ is

$$\sigma^2_{(\bar{X}_1 - \bar{X}_2)} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

…and the random variable

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

has a standard normal distribution

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 KNOWN)

Population means, independent samples

- σ_1^2 and σ_2^2 known *
- σ_1^2 and σ_2^2 unknown

The confidence interval for $\mu_1 - \mu_2$ is :

$$(\overline{X}_1 - \overline{X}_2) \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 KNOWN)

Assuming that both populations are **normally distributed**, a 100(1-α)% confidence interval for **the difference between two means $\mu_1 - \mu_2$**, independent samples, and **known population variances σ_1^2 and σ_2^2** is given by

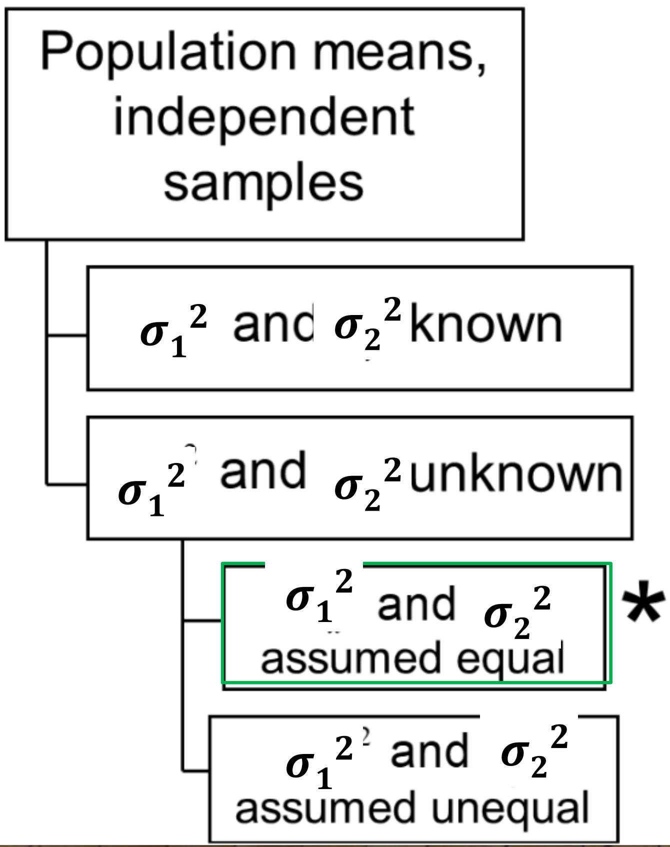
$$IC_{100(1-\alpha)\%}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

⇕

$$\left](\overline{X}_1 - \overline{X}_2) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}};\ (\overline{X}_1 - \overline{X}_2) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.\right[$$

If the populations are not normally distributed, large samples can be used due to the Central Limit Theorem (CLT).

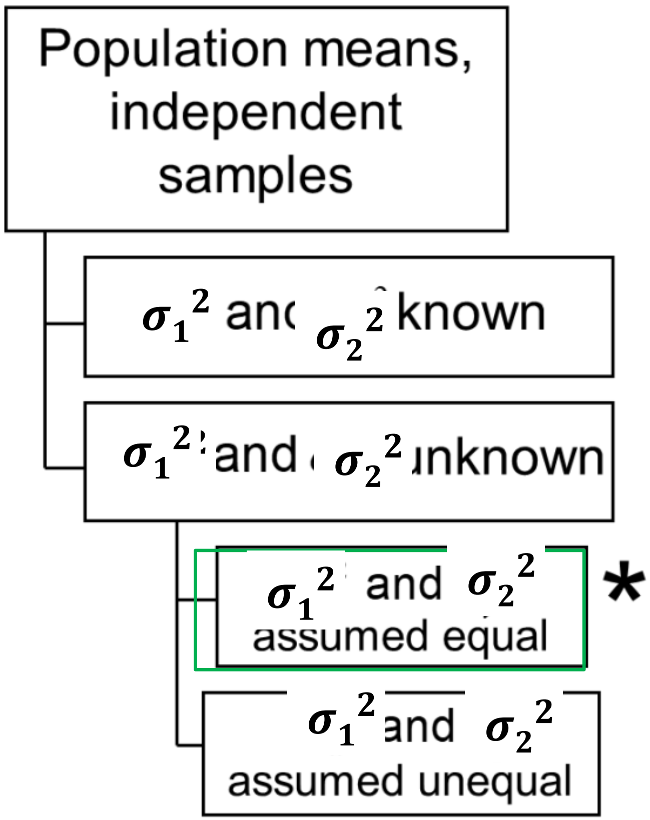
CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)



Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown but assumed equal

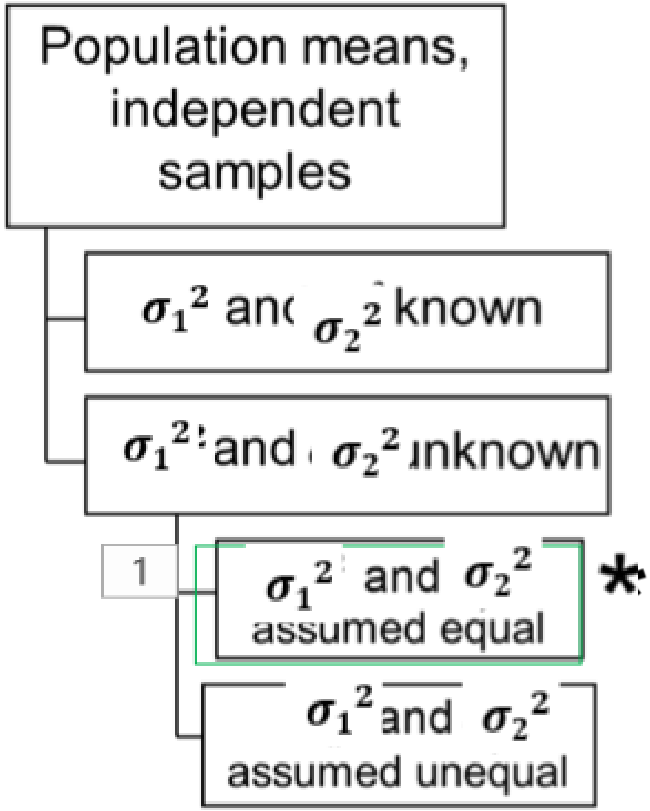
CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)



Forming interval estimates:

- The population variances are assumed equal, so use the two sample standard deviations and pool them to estimate σ
- use a t value with $n_1 + n_2 - 2$ degrees of freedom

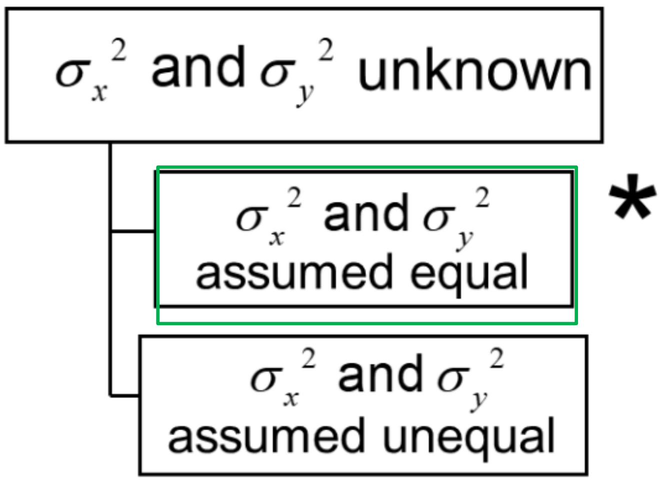
CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)



The pooled variance is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)



The confidence interval for $\mu_1 - \mu_2$ is :

$$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2,\, n_1+n_2-2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

$$IC_{100(1-\alpha)\%}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2,\, n_1+n_2-2}\; s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)

Assuming that both populations are **normally distributed** and have **equal but unknown variances**, a $100(1-\alpha)\%$ confidence interval for the **difference between two means** $\boldsymbol{\mu_1 - \mu_2}$, based on independent samples, is given by

$$IC_{100(1-\alpha)\%}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2,\, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where the **pooled sample variance** is

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}.$$

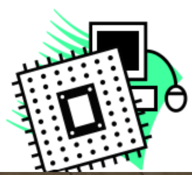
If the populations are not normally distributed, large samples can be used due to the Central Limit Theorem (CLT).

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL): EXAMPLE

You are testing two computer processors for speed. Form a confidence interval for the difference in CPU speed. You collect the following speed data (in Mhz):

	CPU_x	CPU_y
Number Tested	17	14
Sample mean	3004	2538
Sample std dev	74	56

Assume both populations are normal with equal variances, and use 95% confidence



CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL): EXAMPLE

The pooled variance is:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(17-1)74^2 + (14-1)56^2}{(17-1)+(14-1)} = 4427.03$$

The *t* value for a 95% confidence interval is:

$$t_{1-\frac{\alpha}{2};\, n_1+n_2-2} = t_{0.975;\, 29} = 2.045$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL): EXAMPLE

- The 95% confidence interval is

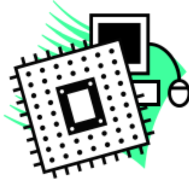
$$(\bar{x}-\bar{y}) \pm t_{1-\frac{\alpha}{2};\, n_1+n_2-2}\sqrt{\frac{s_p^2}{n_x}+\frac{s_p^2}{n_y}}$$

$$(3004-2538) \pm (2.054)\sqrt{\frac{4427.03}{17}+\frac{4427.03}{14}}$$

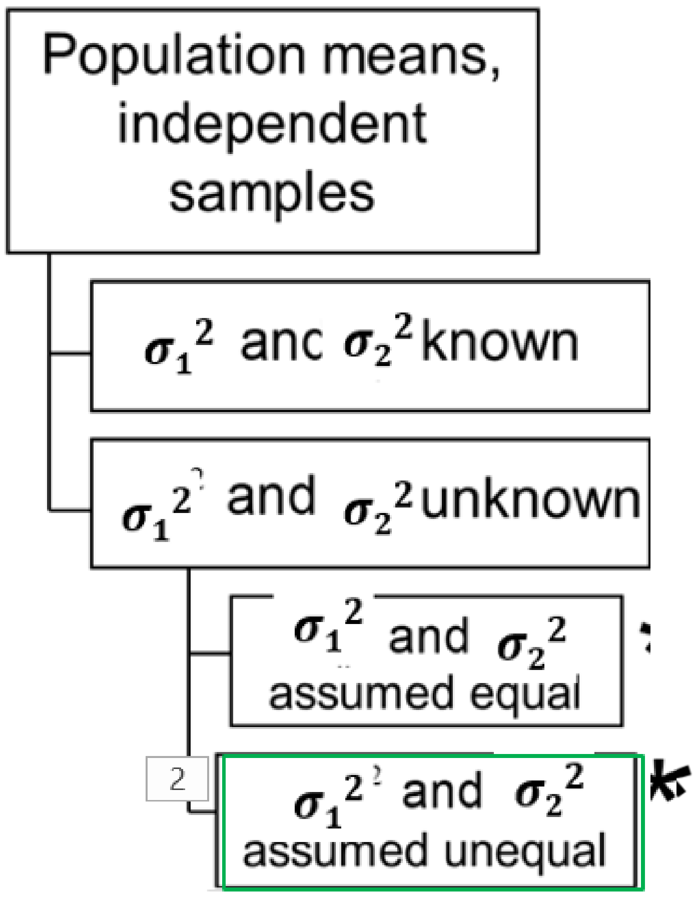
$$416.69 < \mu_1 - \mu_2 < 515.31$$

$CI_{90\%}(\mu_1 - \mu_2) = (416.69, 515.31)$

We are 95% confident that the mean difference in CPU speed is between 416.69 and 515.31 Mhz.



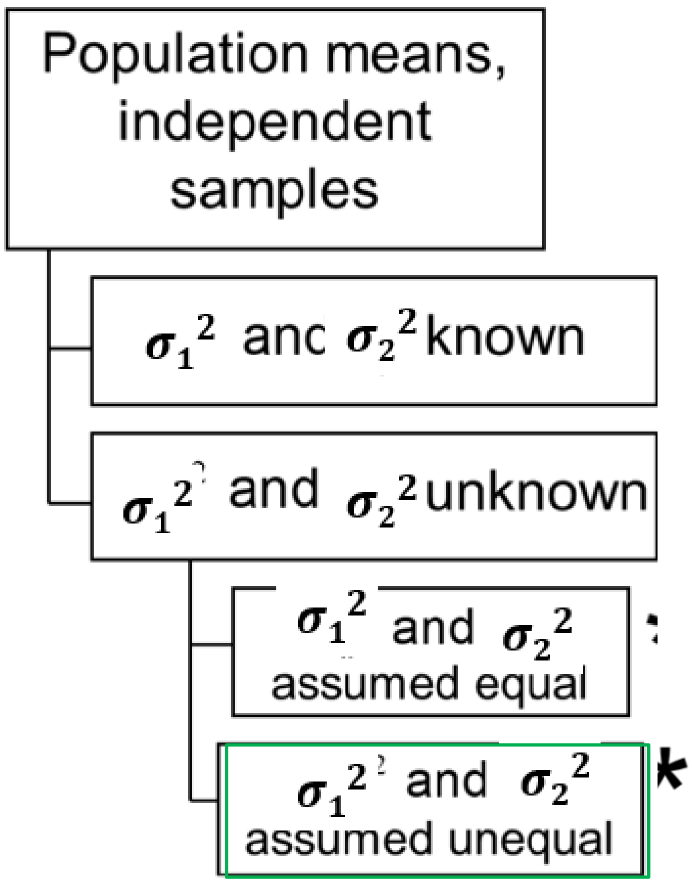
CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)



Assumptions:

- Samples are randomly and independently drawn
- Populations are normally distributed
- Population variances are unknown and assumed unequal

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)



Forming interval estimates:

- The population variances are assumed unequal, so a pooled variance is not appropriate
- use a *t* value with ν degrees of freedom, where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}.$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN BUT ASSUMED EQUAL)

- σ_1^2 and σ_2^2 unknown
  - σ_1^2 and σ_2^2 assumed equal
  - σ_1^2 and σ_2^2 assumed unequal *

The confidence interval for $\mu_1 - \mu_2$ is :

$$(\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2,\,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$$

$$IC_{100(1-\alpha)\%}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2,\,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

CONFIDENCE INTERVALS FOR THE DIFFERENCE BETWEEN TWO MEANS (σ_1^2 AND σ_2^2 UNKNOWN AND UNEQUAL)

Assuming that both populations are **normally distributed** and have **unknown and unequal variances**, a $100(1-\alpha)\%$ confidence interval for the **difference between two means $\mu_1 - \mu_2$**, based on independent samples, is given by

$$IC_{100(1-\alpha)\%}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{1-\alpha/2,\,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where the **degrees of freedom** ν are approximated by the **Welch–Satterthwaite formula**:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}.$$

If the populations are not normally distributed, large samples can be used due to the Central Limit Theorem (CLT).

EXERCISE 8.6

8.6 Independent random sampling from two normally distributed populations gives the following results:

$$n_x = 64; \bar{x} = 400; \sigma_x = 20$$

$$n_y = 36; \bar{y} = 360; \sigma_y = 25$$

Find a 90% confidence interval estimate of the difference between the means of the two populations.

Newbold et al (2013)

EXERCISE 8.6: SOLUTION

$$CI_{(1-\alpha)}(\mu_1 - \mu_2) = \left((\bar{x}_1 - \bar{x}_2) - z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Answer:

Given:

$$n_1 = 64, \quad \bar{x}_1 = 400, \quad \sigma_1 = 20$$

$$n_2 = 36, \quad \bar{x}_2 = 360, \quad \sigma_2 = 25$$

Confidence level: 90% → $\alpha = 0.10$ → $1 - \alpha/2 = 0.95$

Step 1: Confidence interval formula

For independent samples from normally distributed populations:

$$\mathrm{CI}_{90\%}(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Step 2: Standard error

$$SE = \sqrt{\frac{20^2}{64} + \frac{25^2}{36}} = \sqrt{6.25 + 17.3611} = \sqrt{23.6111} \approx 4.86$$

$1 - \alpha = 0.90$ (confidence level), then $z_{1-\frac{\alpha}{2}} = z_{0.95} = 1.645$ (see Standard Normal Distribution Table)

Step 3: Margin of error

$$ME = z_{0.95} \cdot SE = 1.645 \cdot 4.86 \approx 7.99 \approx 8$$

Step 4: Compute the interval

$$\bar{x}_1 - \bar{x}_2 = 400 - 360 = 40$$

$$\mathrm{CI}_{90\%}(\mu_1 - \mu_2) = 40 \pm 8 = (32, 48)$$

$\mathrm{CI}_{90\%}(\mu_1 - \mu_2) = (32, 48)$

EXERCISE 8.13

8.13 From a random sample of six students in an introductory finance class that uses group-learning techniques, the mean examination score was found to be 76.12 and the sample standard deviation was 2.53. For an independent random sample of nine students in another introductory finance class that does not use group-learning techniques, the sample mean and standard deviation of exam scores were 74.61 and 8.61, respectively. Estimate with 95% confidence the difference between the two population mean scores; do not assume equal population variances.

Newbold et al (2013)

EXERCISE 8.13: SOLUTION

$$CI_{(1-\alpha)}(\mu_1 - \mu_2) = \left((\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

Answer:

Given:

- Group-learning class:

$$n_1 = 6, \quad \bar{x}_1 = 76.12, \quad s_1 = 2.53$$

- Non-group-learning class:

$$n_2 = 9, \quad \bar{x}_2 = 74.61, \quad s_2 = 8.61$$

Confidence level: 95% → $\alpha = 0.05$

Step 1: Confidence interval formula (Welch's t-interval)

$$\mathrm{CI}_{95\%}(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2,\, df} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where the degrees of freedom df are approximated using the **Welch–Satterthwaite formula**:

$$df \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Step 2: Compute the standard error

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{2.53^2}{6} + \frac{8.61^2}{9}}$$

Step by step:

1. $\frac{2.53^2}{6} = \frac{6.4009}{6} \approx 1.067$
2. $\frac{8.61^2}{9} = \frac{74.1521}{9} \approx 8.239$

$$SE = \sqrt{1.067 + 8.239} = \sqrt{9.306} \approx 3.051$$

Step 3: Compute the degrees of freedom

$$df \approx \frac{(1.067 + 8.239)^2}{\frac{1.067^2}{6-1} + \frac{8.239^2}{9-1}} = \frac{9.306^2}{\frac{1.138}{5} + \frac{67.87}{8}}$$

Step by step:

- Numerator: $9.306^2 \approx 86.61$
- Denominator: $\frac{1.138}{5} \approx 0.228$ and $\frac{67.87}{8} \approx 8.484$ → Sum ≈ 8.712

$$df \approx \frac{86.61}{8.712} \approx 9.94 \approx 10$$

EXERCISE 8.13: SOLUTION

$$CI_{(1-\alpha)}(\mu_1 - \mu_2) = \left((\bar{x}_1 - \bar{x}_2) - t_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, (\bar{x}_1 - \bar{x}_2) + t_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \right)$$

Answer:

Step 4: Find $t_{1-\alpha/2,df}$ for 95% confidence

- $\alpha/2 = 0.025, df \approx 10$
- From t-tables: $t_{0.975,10} \approx 2.228$

1 – α = 0.95 (confidence level), then $t_{1-\frac{\alpha}{2}} = t_{0.975;10} = 2.228$ (see t-Student Distribution Table)

Step 5: Compute the margin of error

$$ME = t \cdot SE = 2.228 \cdot 3.051 \approx 6.79$$

Step 6: Compute the difference in sample means

$$\bar{x}_1 - \bar{x}_2 = 76.12 - 74.61 = 1.51$$

Step 7: Construct the confidence interval

$$\mathrm{CI}_{95\%}(\mu_1 - \mu_2) = 1.51 \pm 6.79$$

$$\mathrm{CI}_{95\%}(\mu_1 - \mu_2) \approx (-5.28, 8.30)$$

$CI_{95\%}(\mu_1 - \mu_2) = (-5.28, 8.30)$

Interpretation:

- We are 95% confident that the true difference in mean exam scores between the group-learning and non-group-learning classes lies between **-5.28 and 8.30**. Since the interval includes 0, we do **not have strong evidence of a difference** between the two class types.

EXERCISE 8.17

8.17 A researcher intends to estimate the effect of a drug on the scores of human subjects performing a task of psychomotor coordination. The members of a random sample of 9 subjects were given the drug prior to testing. The mean score in this group was 9.78, and the sample variance was 17.64. An independent random sample of 10 subjects was used as a control group and given a placebo prior to testing. The mean score in this control group was 15.10, and the sample variance was 27.01. Assuming that the population distributions are normal with equal variances, find a 90% confidence interval for the difference between the population mean scores.

Newbold et al (2013)

EXERCISE 8.17: SOLUTION

Answer:

Given:

- **Drug group:**

$$n_1 = 9, \quad \bar{x}_1 = 9.78, \quad s_1^2 = 17.64$$

- **Control group:**

$$n_2 = 10, \quad \bar{x}_2 = 15.10, \quad s_2^2 = 27.01$$

- Confidence level: 90% → $\alpha = 0.10$

We want **CI**$_{90\%}$ **for** $\mu_1 - \mu_2$.

Step 1: Compute the pooled variance

For **equal population variances**, the pooled variance is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Substitute the numbers:

$$s_p^2 = \frac{(9-1) \cdot 17.64 + (10-1) \cdot 27.01}{9 + 10 - 2} = \frac{8 \cdot 17.64 + 9 \cdot 27.01}{17}$$

Step by step:

- $8 \cdot 17.64 = 141.12$
- $9 \cdot 27.01 = 243.09$
- Sum = 141.12 + 243.09 = 384.21

$$s_p^2 = \frac{384.21}{17} \approx 22.601$$

$$s_p \approx \sqrt{22.601} \approx 4.756$$

EXERCISE 8.17: SOLUTION

✅ Answer:

Step 2: Compute the standard error of the difference

$$SE = s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 4.756 \cdot \sqrt{\frac{1}{9} + \frac{1}{10}}$$

Step by step:

$$\frac{1}{9} + \frac{1}{10} \approx 0.1111 + 0.1 = 0.2111$$

$$SE = 4.756 \cdot \sqrt{0.2111} = 4.756 \cdot 0.4595 \approx 2.185$$

Step 3: Determine the t-value

- Degrees of freedom: $df = n_1 + n_2 - 2 = 9 + 10 - 2 = 17$
- 90% confidence → $\alpha/2 = 0.05$
- From t-tables: $t_{0.95,17} \approx 1.740$

Step 4: Compute the margin of error

$$ME = t \cdot SE = 1.740 \cdot 2.185 \approx 3.80$$

Step 5: Compute the difference in sample means

$$\bar{x}_1 - \bar{x}_2 = 9.78 - 15.10 = -5.32$$

Step 6: Construct the confidence interval

$$\text{CI}_{90\%}(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm ME = -5.32 \pm 3.80$$

$$\text{CI}_{90\%}(\mu_1 - \mu_2) \approx (-9.12, -1.52)$$

$\text{CI}_{90\%}$ $(\mu_1 - \mu_2)$ = (-9.12, -1.52)

THANKS!

Questions?