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

HOMEWORK OF LECTURE 15: QUESTIONS AND SOLUTIONS

EXERCISE 9.33

9.33 Of a random sample of 199 auditors, 104 indicated some measure of agreement with this statement: *Cash flow is an important indication of profitability.* Test at the 10% significance level against a two-sided alternative the null hypothesis that one-half of the members of this population would agree with this statement. Also find and interpret the *p*-value of this test.

Newbold et al (2013)

EXERCISE 9.33: SOLUTION

Note:
Since n×$\hat{p}$×(1- $\hat{p}$) > 5, the normal approximation is valid and the one-sample proportion test can be applied.

Answer:

n×$\hat{p}$×(1- $\hat{p}$) = 199 × 0.5226 × (1-0.5226) = 49.65> 5

Given

- Sample size: $n = 199$
- Number agreeing: $x = 104$
- Sample proportion:

Two-Tailed Test

$$\hat{p} = \frac{104}{199} \approx 0.5226$$

- Null hypothesis: $H_0 : P = 0.5$
- Alternative hypothesis: $H_1 : P \neq 0.5$ (**two-sided**)
- Significance level: $\alpha = 0.10$

Step 1: Test statistic

For large samples, the **Z-test for a proportion** is:

$$Z = \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

Where $P_0 = 0.5$.

Standard error:

$$SE = \sqrt{\frac{0.5 \cdot 0.5}{199}} = \sqrt{\frac{0.25}{199}} \approx \sqrt{0.001256} \approx 0.0354$$

Test statistic:

$$\boxed{Z_0 = \frac{0.5226 - 0.5}{0.0354} \approx \frac{0.0226}{0.0354} \approx 0.64}$$

EXERCISE 9.33: SOLUTION

Answer:

Decision rule: Reject H_0 if the test statistic z_0 is in the rejection region (RR) or if p-value < α.

Step 2: Critical values

Two-sided test at $\alpha = 0.10$:

RR =] -∞; -1.645]U[1.645; +∞[

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

Decision rule:

$$\text{Reject } H_0 \text{ if } |Z_0| > 1.645$$

Step 3: Decision

$$\boxed{z_0 = 0.64 \in RR} \Leftrightarrow |Z_0| = 0.64 < 1.645$$

$\Rightarrow$ Do not reject H_0.

Step 4: p-value

Two-sided p-value:

P-value = 2 × P(T > 0.64) ~ 0.522

From standard normal table:

$$P(Z > 0.64) \approx 0.2611$$

$$\boxed{p\text{-value} = 2 \cdot 0.2611 \approx 0.522}$$

Conclusion

- p-value = 0.522 > 0.10 → do not reject H_0.
- There is **no significant evidence** at the 10% level that the proportion of auditors agreeing differs from 50%.

INFERENCE FOR A POPULATION PROPORTION

Note: The statistic used for a **confidence interval** and for a **hypothesis test** for a population proportion are **different**.

Confidence Interval for a Population Proportion

$$\boxed{\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}}$$

- Uses the **sample proportion** $\hat{p}$ to estimate the standard error.

Confidence Interval for a Population Proportion

A $100(1-\alpha)\%$ confidence interval for the population proportion p is:

$$\boxed{\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

- $\hat{p}$: sample proportion
- n: sample size
- $z_{\alpha/2}$: critical value from the standard normal distribution

Hypothesis Test for a Population Proportion

The test statistic is:

$$\boxed{z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$$

- p_0: proportion under the null hypothesis

Key difference:
Confidence interval: standard error based on $\widehat{p}$.
Hypothesis test: standard error based on p_0.

POPULATION PROPORTIONS: WHY ARE THE STATISTICS DIFFERENT FOR CONFIDENCE INTERVALS AND HYPOTHESIS TESTS?

Note: The difference comes from the **role played by the unknown population proportion** p in each procedure.

Confidence Interval

$$\boxed{\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

- The true proportion p is **unknown**.
- The standard error is therefore **estimated** using the sample proportion $\hat{p}$.
- The goal is **estimation**, not decision-making.

Hypothesis Test

$$\boxed{z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}$$

- Under the null hypothesis, the population proportion is **fixed at** p_0.
- The standard error is computed using this **assumed value**.
- The goal is to **test a claim** about p.

Key Takeaway:
Confidence interval: estimates variability using $\widehat{p}$.
Hypothesis test: evaluates evidence assuming p = p_0.
The statistics differ because **estimation** and **hypothesis testing** answer different statistical questions.

LECTURE 16: TESTS OF THE VARIANCE OF A NORMAL DISTRIBUTION

TESTS OF THE VARIANCE OF A NORMAL DISTRIBUTION

- Goal: Test hypotheses about the population variance, σ^2 (e.g., $H_0 : \sigma^2 = \sigma_0^2$)
  – If the population is normally distributed,

$$Q = \frac{(n-1)s^2}{\sigma^2}$$

has a chi-square distribution with $(n-1)$ degrees of freedom

Two-Tailed Test

$H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 \neq \sigma_0^2$

Newbold et al (2013)

TESTS OF THE VARIANCE OF A NORMAL DISTRIBUTION

The test statistic for hypothesis tests about one population variance is

$$\boxed{Q = \frac{(n-1)s^2}{\sigma_0^2}}$$

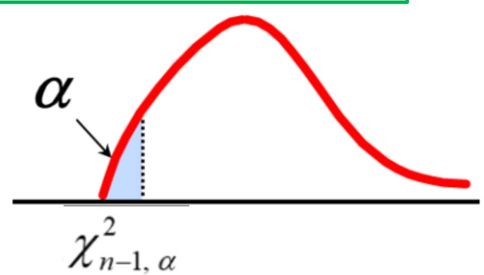
Newbold et al (2013)

TESTS OF THE VARIANCE OF A NORMAL DISTRIBUTION: DECISION USING THE RR

Population variance

Left-Tailed Test

$$H_0 : \sigma^2 \geq \sigma_0^2$$
$$H_1 : \sigma^2 < \sigma_0^2$$



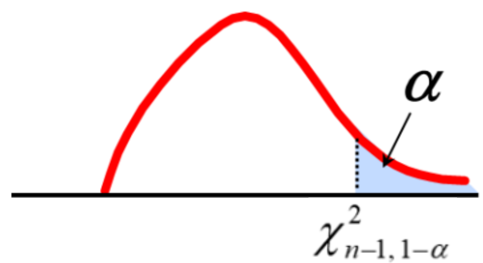
Reject H_0 if

q $< \chi^2_{n-1,\,\alpha}$

RR =]-∞; $\chi^2_{n-1,\alpha}$]

Right-Tailed Test

$$H_0 : \sigma^2 \leq \sigma_0^2$$
$$H_1 : \sigma^2 > \sigma_0^2$$



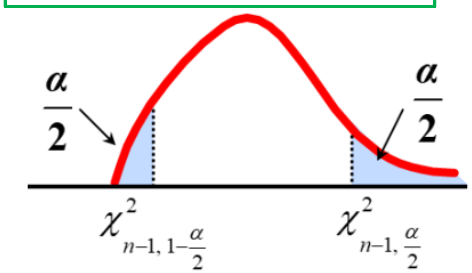
Reject H_0 if

q $> \chi^2_{n-1,\,1-\alpha}$

RR = [$\chi^2_{n-1,1-\alpha}$; +∞[

Two-Tailed Test

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 \neq \sigma_0^2$$



Reject H_0 if

q $> \chi^2_{n-1,\,1-\frac{\alpha}{2}}$

or

q $< \chi^2_{n-1,\,\frac{\alpha}{2}}$

RR =] -∞; $\chi^2_{n-1,\alpha/2}$]U[$\chi^2_{n-1,1-\alpha/2}$; +∞[

et al (2013)

TESTS OF THE VARIANCE OF A NORMAL DISTRIBUTION: DECISION USING THE P-VALUE

Test Statistic

$$Q = \frac{(n-1)s^2}{\sigma_0^2}, \qquad Q \sim \chi^2_{n-1}$$

Left-Tailed Test

- **Hypotheses:**

$$H_0 : \sigma^2 \geq \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 < \sigma_0^2$$

- **p-value:**

$$\boxed{p\text{-value} = P(Q \leq q_0)}$$

Right-Tailed Test

- **Hypotheses:**

$$H_0 : \sigma^2 \leq \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 > \sigma_0^2$$

- **p-value:**

$$\boxed{p\text{-value} = P(Q \geq q_0)}$$

Two-Tailed Test

- **Hypotheses:**

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 \neq \sigma_0^2$$

- **p-value:**

$$\boxed{p\text{-value} = 2 \times \min\left\{P(Q \leq q_0),\ P(Q \geq q_0)\right\}}$$

Note:
Q denotes the **random test statistic**, while q_0 is its **observed value**.
For a two-tailed test, the p-value is twice the smaller tail probability.

EXERCISE 9.49

9.49 Plastic sheets produced by a machine are periodically monitored for possible fluctuations in thickness. If the true variance in thicknesses exceeds 2.25 square millimeters, there is cause for concern about product quality. Thickness measurements for a random sample of 10 sheets produced in a particular shift were taken, giving the following results (in millimeters):

226 226 232 227 225 228 225 228 229 230

a. Find the sample variance.
b. Test, at the 5% significance level, the null hypothesis that the population variance is at most 2.25.

Newbold et al (2013)

EXERCISE 9.49 A): SOLUTION

Answer:

Given

- Sample size: $n = 10$
- Sample measurements (in mm): 226, 226, 232, 227, 225, 228, 225, 228, 229, 230
- Significance level: $\alpha = 0.05$
- Null hypothesis: $H_0 : \sigma^2 \leq 2.25$
- Alternative hypothesis: $H_1 : \sigma^2 > 2.25$ (**right-tailed test** for variance)

Right-Tailed Test

We will use the **chi-squared test for variance**:

$$Q = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1} \text{ under } H_0$$

Sample Variance

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

- x_i = each observation
- $\bar{x}$ = sample mean
- n = sample size

$$\bar{x} = \frac{2276}{10} = 227.6$$

$$S^2 = \frac{46.40}{n-1} = \frac{46.40}{9} \approx 5.1556$$

Sample variance: $S^2 \approx 5.156\,\mathrm{mm}^2$

EXERCISE 9.49 B): SOLUTION

Answer:

b) Test for variance

$$q_0 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9 \cdot 5.156}{2.25} \approx \frac{46.404}{2.25} \approx 20.62$$

Degrees of freedom: $df = n - 1 = 9$

Step 2: Critical value

Right-tailed test at $\alpha = 0.05$:

RR = [16.919; +∞[

$$\chi^2_{0.95,9} \approx 16.919$$

Decision Using the RR

Step 3: Decision

$$q_0 = 20.62 > 16.919$$

$\Rightarrow$ Reject H_0 at the 5% significance level.

Conclusion

There is **evidence that the population variance exceeds 2.25 mm²**, so the machine may indeed be producing **excessive variability in sheet thickness**.

HOMEWORK OF LECTURE 16: QUESTIONS

EXERCISE 9.51

9.51 A company produces electric devices operated by a thermostatic control. The standard deviation of the temperature at which these controls actually operate should not exceed 2.0°F. For a random sample of 20 of these controls, the sample standard deviation of operating temperatures was 2.36°F. Stating any assumptions you need to make, test, at the 5% level, the null hypothesis that the population standard deviation is 2.0 against the alternative that it is larger.

Newbold et al (2013)

EXERCISE 9.51: SOLUTION

Answer:

Right-Tailed Test

$$H_0 : \sigma \leq 2.0 \quad \text{(acceptable variability)}$$

$$H_1 : \sigma > 2.0 \quad \text{(too large variability)}$$

Assumption: The **population of operating temperatures is normally distributed** (required for the chi-squared test of variance).

Step 1: Compute the test statistic

For a **right-tailed chi-squared test for variance**:

$$q_0 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$s^2 = 2.36^2 = 5.5696$$

$$\sigma_0^2 = 2^2 = 4$$

$$\boxed{q_0 = \frac{(20-1)\cdot 5.5696}{4} = \frac{19 \cdot 5.5696}{4} = \frac{105.8224}{4} \approx 26.4556}$$

EXERCISE 9.51: SOLUTION

Answer:

Step 2: Critical value

Degrees of freedom: $df = n - 1 = 19$

Right-tailed test at $\alpha = 0.05$:

RR = [30.144; +∞[

$$\chi^2_{0.95,19} \approx 30.144$$

Step 3: Decision

$$\boxed{q_0 = 26.456 < 30.144}$$

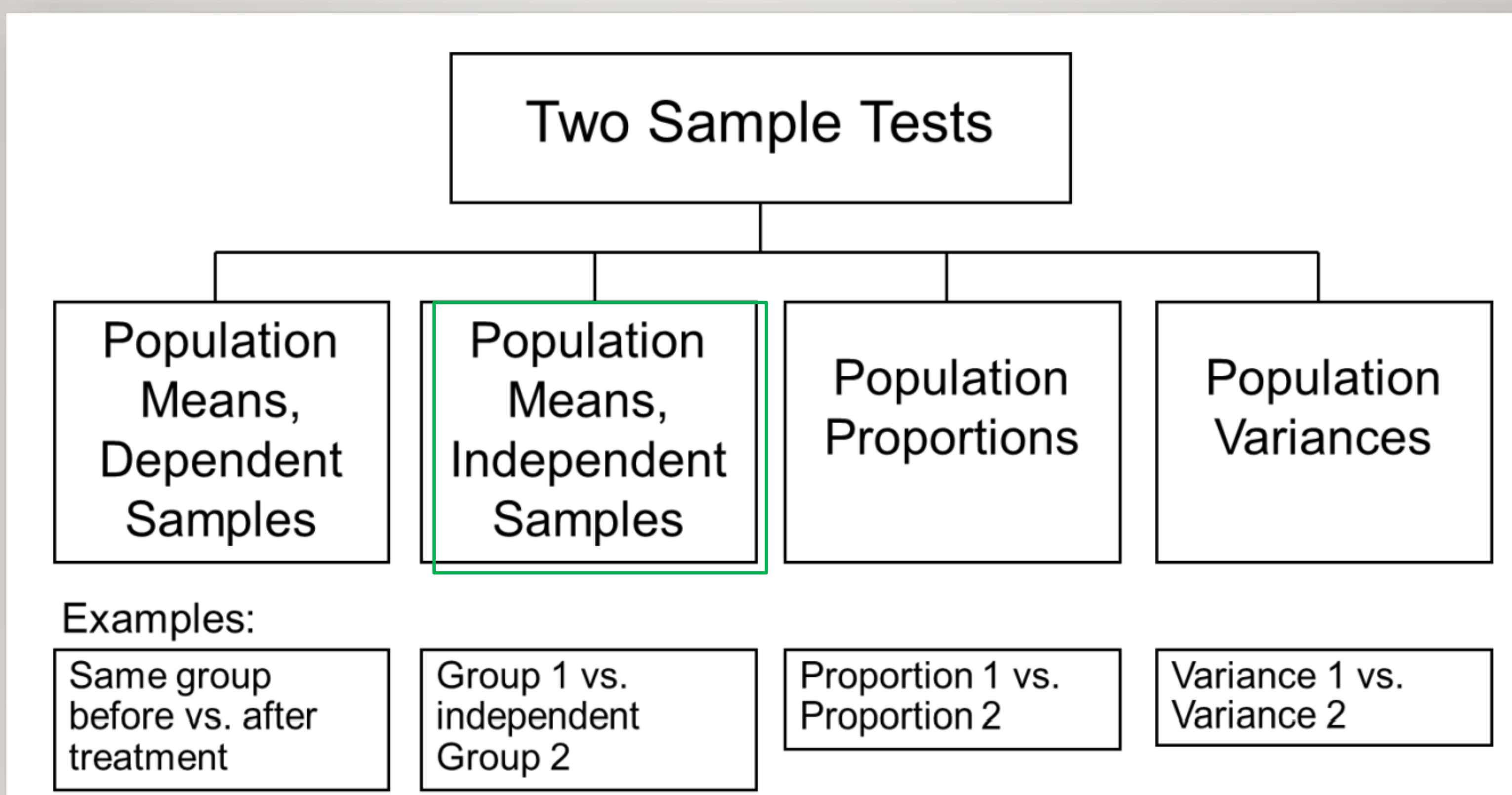
⇒ **Do not reject H_0 at the 5% level.**

Step 4: Interpretation

- Although the sample standard deviation is slightly larger than 2.0°F, there is **not enough evidence** to conclude that the population standard deviation exceeds 2.0°F at the 5% significance level.
- The machine variability is **still considered acceptable** based on this sample.

LECTURE 17: TESTS OF THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS

TWO SAMPLE TESTS



Newbold et al (2013)

INDEPENDENT SAMPLES

Population means, independent samples

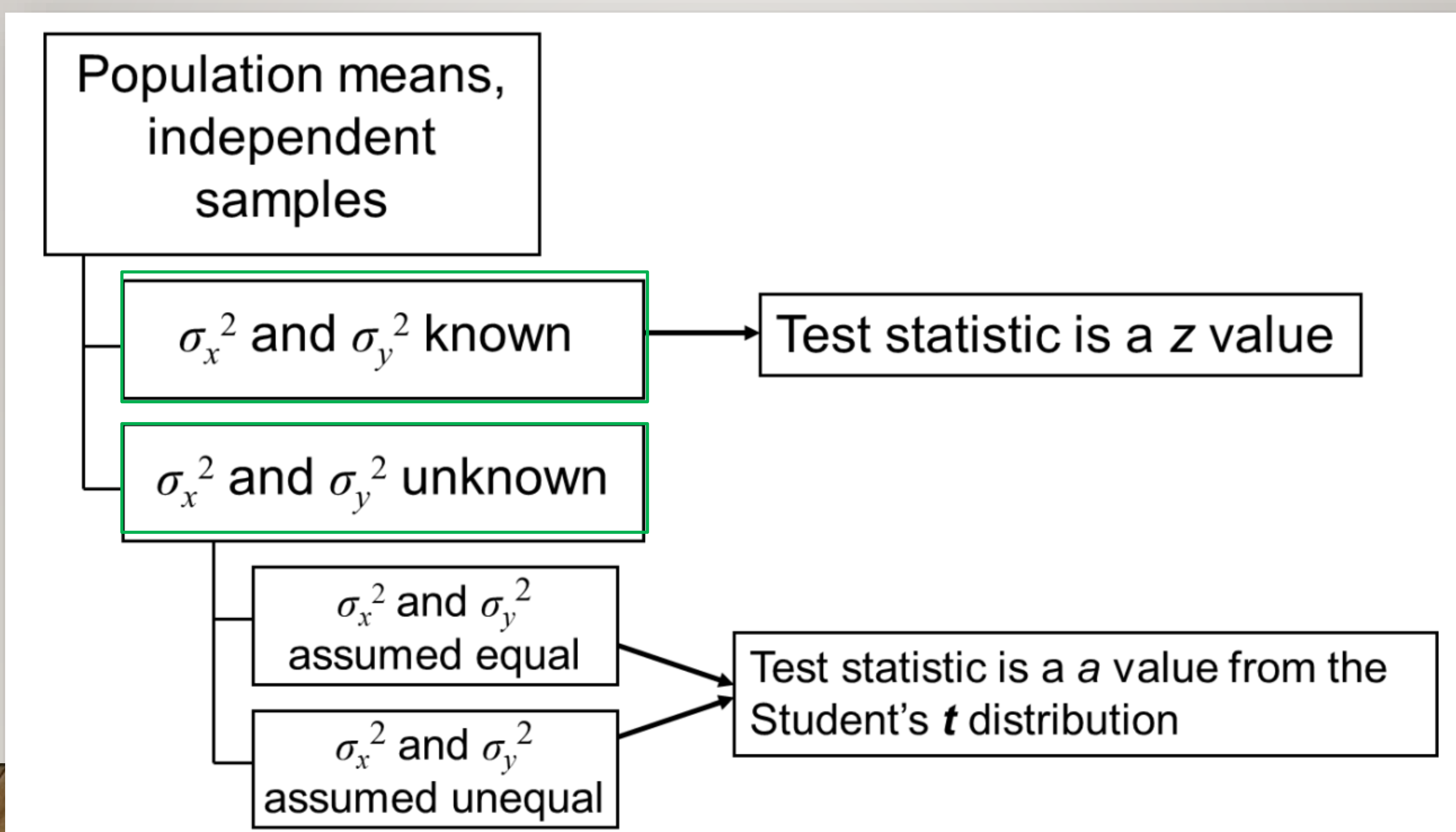
Tests of the Difference Between Two Normal Population Means: Dependent Samples

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

- Different populations
  - Unrelated
  - Independent
    - Sample selected from one population has no effect on the sample selected from the other population
  - Normally distributed

Newbold et al (2013)

DIFFERENCE BETWEEN TWO MEANS



Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (KNOWN VARIANCES)

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

Assumptions:

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

Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (KNOWN VARIANCES)

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

Formula given in the formula sheet

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

When σ_x^2 and σ_y^2 are known and both populations are normal, the variance of $\bar{X} - \bar{Y}$ is

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

…and the random variable

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma^2_x}{n_X} + \frac{\sigma^2_y}{n_Y}}}$$

has a standard normal distribution

Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (KNOWN VARIANCES)

Population means, independent samples

- σ_x^2 and σ_y^2 known *
- σ_x^2 and σ_y^2 unknown

Two-Tailed Test

$$H_0 : \mu_x - \mu_y = 0$$

The test statistic for $\mu_x - \mu_y$ is:

$$z = \frac{(\bar{x} - \bar{y})}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$$

Newbold et al (2013)

HYPOTHESIS TESTS FOR TWO POPULATION MEANS

Two Population Means, Independent Samples

Left-Tailed Test	Right-Tailed Test	Two-Tailed Test
$H_0 : \mu_x \geq \mu_y$ $H_1 : \mu_x < \mu_y$	$H_0 : \mu_x \leq \mu_y$ $H_1 : \mu_x > \mu_y$	$H_0 : \mu_x = \mu_y$ $H_1 : \mu_x \neq \mu_y$
i.e.,	i.e.,	i.e.,
$H_0 : \mu_x - \mu_y \geq 0$ $H_1 : \mu_x - \mu_y < 0$	$H_0 : \mu_x - \mu_y \leq 0$ $H_1 : \mu_x - \mu_y > 0$	$H_0 : \mu_x - \mu_y = 0$ $H_1 : \mu_x - \mu_y \neq 0$

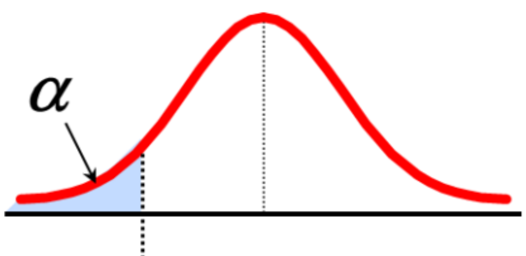
Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (KNOWN VARIANCES): DECISION USING THE RR

Two Population Means, Independent Samples, Variances Known

Left-Tailed Test

$$H_0: \mu_x - \mu_y \geq 0$$
$$H_1: \mu_x - \mu_y < 0$$



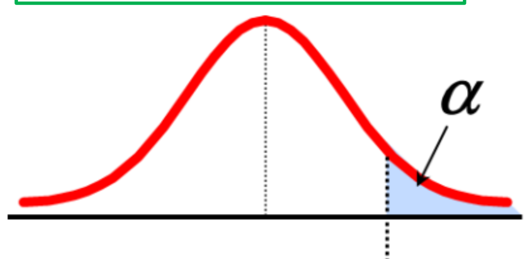
z_α

Reject H_0 if $z < z_\alpha$

RR =]-∞; z_α]

Right-Tailed Test

$$H_0: \mu_x - \mu_y \leq 0$$
$$H_1: \mu_x - \mu_y > 0$$



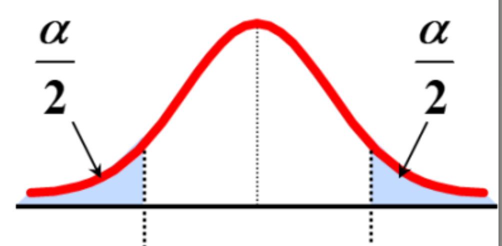
$z_{1-\alpha}$

Reject H_0 if $z > z_{1-\alpha}$

RR = [$z_{1-\alpha}$; +∞[

Two-Tailed Test

$$H_0: \mu_x - \mu_y = 0$$
$$H_1: \mu_x - \mu_y \neq 0$$



$z_{\alpha/2}$ $z_{1-\alpha/2}$

Reject H_0 if $z < z_{\alpha/2}$

or $z > z_{1-\alpha/2}$

RR =] -∞; $-z_{1-\alpha/2}$]U[$z_{1-\alpha/2}$; +∞[

Newbold et al (2013)

EXERCISE 10.6

10.6 You have been asked to determine if two different production processes have different mean numbers of units produced per hour. Process 1 has a mean defined as μ_1 and process 2 has a mean defined as μ_2. The null and alternative hypotheses are as follows:

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0$$

Use a random sample of 25 observations from process 1 and 28 observations from process 2 and the known variance for process 1 equal to 900 and the known variance for process 2 equal to 1,600. Can you reject the null hypothesis using a probability of Type I error $\alpha = 0.05$ in each case?

a. The process means are 50 and 60.

Newbold et al (2013)

EXERCISE 10.6 A): SOLUTION

Answer:

Two-Sample t-Test for the Difference of Means (Unequal and Known Variances)

Data

- Sample sizes: $n_1 = 25, n_2 = 28$
- Population variances: $\sigma_1^2 = 900, \sigma_2^2 = 1600$
- Observed (or assumed) means: $\bar{x}_1 - \bar{x}_2 = 50 - 60$
- Significance level: $\alpha = 0.05$

Step 1: Hypotheses

Two-Tailed Test

$$\boxed{H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 \neq 0}$$

EXERCISE 10.6 A): SOLUTION

Answer:

Step 2: Test Statistic

For known variances, the test statistic is:

$$\boxed{Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim \text{Normal}(0,1)}$$

Compute:

$$\bar{x}_1 - \bar{x}_2 = 50 - 60 = -10$$

$$\text{Standard error} = \sqrt{\frac{900}{25} + \frac{1600}{28}} = \sqrt{36 + 57.14} = \sqrt{93.14} \approx 9.65$$

$$z \approx \frac{-10}{9.65} \approx -1.036$$

The value of the test statistic is $z = -1.036$.

EXERCISE 10.6 A): SOLUTION

Answer:

Step 3: Compare test statistic to rejection region

$$z \approx -1.036$$

The value of the test statistic is z = -1.036.

- RR =] -∞; -1.96]U[1.96; +∞[

α = 0.05
1- α/2 = 0.975
$z_{1-\frac{\alpha}{2}}$ = $z_{0.975}$ = 1.96 (from the standard normal table)

- Two-sided test ($\alpha = 0.05$)
- Critical values: $z_{0.025} = \pm 1.96$
- **Reject H_0 if $z < -1.96$ or $z > 1.96$**

Step 4: P-value

Two-sided p-value:

P-value = 2× P(Z > |-1.036|) = 2× P(Z > 1.036) = 2×[1-Φ(1.036)] = 2×[1-0.8485] = 2×0.1515 ~ 0.303
Approximate value (from the standard normal table)

P-value = 2 × P(T > 1.036) ~ 2 × 0.150 ~ 0.30

Exact value

EXERCISE 10.6 A): SOLUTION

✅ Answer:

Step 5: Conclusion

- P-value $p = 0.30 > 0.05$
- Test statistic not in rejection region

$\boxed{\textbf{Decision:}\text{ Fail to reject } H_0}$

Interpretation (slide-ready): There is **no statistically significant evidence** at the 5% level that the mean number of units produced differs between the two processes when the means are 50 and 60.

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT EQUAL VARIANCES)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal *
    - σ_x^2 and σ_y^2 assumed unequal

Assumptions:

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

Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT EQUAL VARIANCES)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal *
    - σ_x^2 and σ_y^2 assumed unequal

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

Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT EQUAL VARIANCES)

Two-Tailed Test

$\sigma_x^{\,2}$ and $\sigma_y^{\,2}$ unknown

- $\sigma_x^{\,2}$ and $\sigma_y^{\,2}$ assumed equal *
- $\sigma_x^{\,2}$ and $\sigma_y^{\,2}$ assumed unequal

Formula given in the formula sheet

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

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

The test statistic for $H_0 : \mu_x - \mu_y = 0$ is:

$$\boxed{t = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}}}$$

Where t has $(n_1 + n_2 - 2)$ d.f., and

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

Newbold et al (2013)

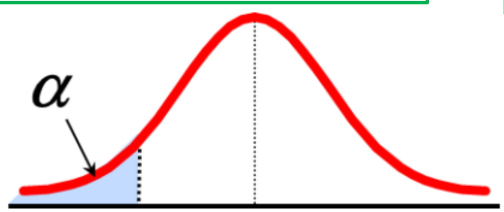
TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT EQUAL VARIANCES): DECISION USING THE RR

Two Population Means, Independent Samples, Variances Unknown

Left-Tailed Test

$H_0 : \mu_x - \mu_y \geq 0$

$H_1 : \mu_x - \mu_y < 0$



t_α

Reject H_0 if

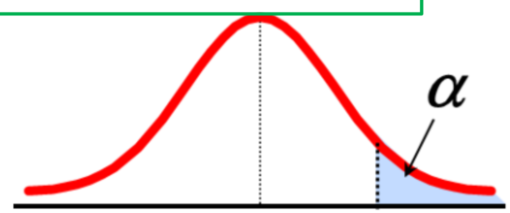
$t < t_{(n1+n2-2),\ \alpha}$

RR =]-∞; t_α]

Right-Tailed Test

$H_0 : \mu_x - \mu_y \leq 0$

$H_1 : \mu_x - \mu_y > 0$



$t_{1-\alpha}$

Reject H_0 if

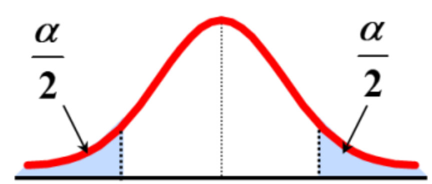
$t > t_{(n1+n2-2),\ 1-\alpha}$

RR = [$t_{1-\alpha}$; +∞[

Two-Tailed Test

$H_0 : \mu_x - \mu_y = 0$

$H_1 : \mu_x - \mu_y \neq 0$



$t_{\alpha/2}$ $t_{1-\alpha/2}$

Reject H_0 if

$t < t_{(n1+n2-2),\ \alpha/2}$

or

$t > t_{(n1+n2-2),\ 1-\alpha/2}$

RR =] -∞; -$t_{1-\alpha/2}$]U[$t_{1-\alpha/2}$; +∞[

Newbold et al (2013)

POOLED VARIANCE T TEST: EXAMPLE

You are a financial analyst for a brokerage firm. Is there a difference in dividend yield between stocks listed on the NYSE & NASDAQ? You collect the following data:

	NYSE	NASDAQ
Number	**21**	**25**
Sample mean	**3.27**	**2.53**
Sample std dev	**1.30**	**1.16**

Assuming both populations are approximately normal with equal variances, is there a difference in average yield $(\alpha = 0.05)$?

Newbold et al (2013)

CALCULATING THE TEST STATISTIC: EXAMPLE

$$H_0 : \mu_1 - \mu_2 = 0 \text{ i.e. } (\mu_1 = \mu_2)$$

$$H_1 : \mu_1 - \mu_2 \neq 0 \text{ i.e. } (\mu_1 \neq \mu_2)$$

Two-Tailed Test

The test statistic is:

$$\boxed{t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(3.27 - 2.53)}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = 2.040}$$

$$S_p^2 = \frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(21-1)1.30^2 + (25-1)1.16^2}{(21-1) + (25-1)} = 1.5021$$

Newbold et al (2013)

SOLUTION USING THE RR: EXAMPLE

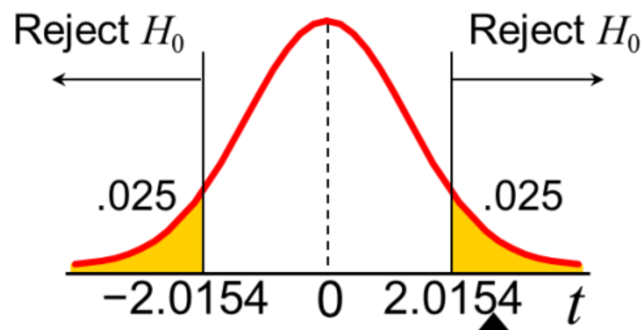
$H_0: \mu_1 - \mu_2 = 0$ i.e. $(\mu_1 = \mu_2)$

$H_1: \mu_1 - \mu_2 \neq 0$ i.e. $(\mu_1 \neq \mu_2)$

$\alpha = 0.05$

$\text{df} = 21 + 25 - 2 = 44$

Critical Values: $t = \pm 2.0154$



RR =] -∞; -2.0154]U[2.0154; +∞[

$\boxed{2.040}$

Test Statistic:

$$t = \frac{3.27 - 2.53}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

There is evidence of a difference in means.

Newbold et al (2013)

EXERCISE 10.13

10.13 A publisher is interested in the effects on sales of college texts that include more than 100 data files. The publisher plans to produce 20 texts in the business area and randomly chooses 10 to have more than 100 data files. The remaining 10 are produced with at most 100 data files. For those with more than 100, first-year sales averaged 9,254, and the sample standard deviation was 2,107. For the books with at most 100, average first-year sales were 8,167, and the sample standard deviation was 1,681. Assuming that the two population distributions are normal with the same variance, test the null hypothesis that the population means are equal against the alternative that the true mean is higher for books with more than 100 data files.

Newbold et al (2013)

EXERCISE 10.13: SOLUTION

Answer:

Two-Sample t-Test for the Difference of Means (Pooled Variance)

Data

- Books with >100 files: $n_1 = 10, \bar{x}_1 = 9{,}254, s_1 = 2{,}107$
- Books with ≤100 files: $n_2 = 10, \bar{x}_2 = 8{,}167, s_2 = 1{,}681$

Assumption: Both populations are **normal** with **equal variance** → use **pooled t-test**.

Step 1: Hypotheses

Right-Tailed Test

$$\boxed{\begin{aligned} H_0 &: \mu_1 \le \mu_2 \\ H_a &: \mu_1 > \mu_2 \end{aligned}}$$

$H_0 : \mu_1 \le \mu_2$ (no difference in mean sales)

$H_a : \mu_1 > \mu_2$ (books with >100 files have higher sales)

EXERCISE 10.13: SOLUTION

✅ Answer:

Step 2: Pooled variance

$$\boxed{s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(10-1)2107^2 + (10-1)1681^2}{10+10-2}}$$

Compute each term:

$$(9)(2107^2) = 9 \cdot 4{,}437{,}649 \approx 39{,}938{,}841$$

$$(9)(1681^2) = 9 \cdot 2{,}825{,}761 \approx 25{,}431{,}849$$

$$\text{Sum} = 39{,}938{,}841 + 25{,}431{,}849 = 65{,}370{,}690$$

$$s_p^2 = \frac{65{,}370{,}690}{18} \approx 3{,}631{,}705$$

$$s_p \approx \sqrt{3{,}631{,}705} \approx 1{,}905$$

EXERCISE 10.13: SOLUTION

Answer:

Step 3: Test statistic(pooled t)

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{9{,}254 - 8{,}167}{1{,}905\sqrt{\frac{1}{10} + \frac{1}{10}}}}$$

Compute:

$$\text{Difference of means} = 9{,}254 - 8{,}167 = 1{,}087$$

$$\sqrt{\frac{1}{10} + \frac{1}{10}} = \sqrt{0.2} \approx 0.4472$$

$$s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \approx 1{,}905 \cdot 0.4472 \approx 851.7$$

$$t \approx \frac{1{,}087}{851.7} \approx 1.277$$

The value of the test statistic is t = 1.277.

EXERCISE 10.13: SOLUTION

Answer:

Step 4: Rejection region

$$t \sim t(n_1 + n_2 - 2) = t(18)$$

- One-sided test ($H_a : \mu_1 > \mu_2$)
- Significance level: $\alpha = 0.05$
- Critical value: $t_{0.95;18}$ ~1.734

Rejection region: $t > 1.734$

α = 0.05
1- α = 0.95
$t_{1-\alpha;n-1}$ = $t_{0.95;18}$ ~ 1.734
(from the student´s t table)

RR = [1.734; +∞[

Step 5: P-value

- Test statistic: $t \approx 1.277 < 1.734$ → not in rejection region
- One-sided p-value: $p \approx 0.11$

Exact value

P-value = P(T > 1.277) ~ 0.1
Approximate value (from the t-table)

EXERCISE 10.13: SOLUTION

Answer:

Step 6: Conclusion

- Test statistic $t = 1.277 < 1.734 \Rightarrow$ **not in rejection region**
- P-value $p = 0.11 > 0.05$

Decision: Fail to reject H_0

Interpretation: There is **no statistically significant evidence** that books with more than 100 data files have higher first-year sales.

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT UNEQUAL VARIANCES)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal
    - σ_x^2 and σ_y^2 assumed unequal *

Assumptions:

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

Newbold et al (2013)

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT UNEQUAL VARIANCES)

- Population means, independent samples
  - σ_x^2 and σ_y^2 known
  - σ_x^2 and σ_y^2 unknown
    - σ_x^2 and σ_y^2 assumed equal
    - σ_x^2 and σ_y^2 assumed unequal *

Forming interval estimates:

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

$$v = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\frac{\left(\frac{s_x^2}{n_x}\right)^2}{(n_x - 1)} + \frac{\left(\frac{s_y^2}{n_y}\right)^2}{n_y - 1}}$$

Newbold et al (2013)

Note: When testing the difference between two means with unknown and unequal variances, the degrees of freedom are computed using the Welch–Satterthwaite formula and need not be an integer. When using tables, the degrees of freedom can be conservatively approximated by taking the **integer part (floor)** of the computed value.

TESTS FOR THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS (UNKNOWN BUT UNEQUAL VARIANCES)

Two-Tailed Test

σ_x^2 and σ_y^2 unknown

- σ_x^2 and σ_y^2 assumed equal
- σ_x^2 and σ_y^2 assumed unequal *

The test statistic for $H_0: \mu_x - \mu_y = 0$ is:

$$t = \frac{(\bar{x} - \bar{y})}{\sqrt{\dfrac{s_x^2}{n_X} + \dfrac{s_y^2}{n_Y}}}$$

Where t has ν degrees of freedom:

$$\nu = \frac{\left[\left(\dfrac{s_x^2}{n_x}\right) + \left(\dfrac{s_y^2}{n_y}\right)\right]^2}{\dfrac{\left(\dfrac{s_x^2}{n_x}\right)^2}{(n_x - 1)} + \dfrac{\left(\dfrac{s_y^2}{n_y}\right)^2}{n_y - 1}}$$

Formula given in the formula sheet

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_\nu$$

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

Newbold et al (2013)

EXERCISE 10.9

10.9 For a random sample of 125 British entrepreneurs, the mean number of job changes was 1.91 and the sample standard deviation was 1.32. For an independent random sample of 86 British corporate managers, the mean number of job changes was 0.21 and the sample standard deviation was 0.53. Test the null hypothesis that the population means are equal against the alternative that the mean number of job changes is higher for British entrepreneurs than for British corporate managers.

Newbold et al (2013)

EXERCISE 10.9: SOLUTION

Answer:

Two-Sample t-Test for the Difference of Means (Unequal and Unknown Variances)

In this problem, the population variances are **unknown** (we only have sample standard deviations) and the sample variances are also **clearly different**:

- Entrepreneurs: $s_1 = 1.32$
- Corporate managers: $s_2 = 0.53$

Because the variances are not assumed to be equal, the appropriate test is the **Welch t-test**, which does *not* assume equal population variances.

EXERCISE 10.9: SOLUTION

Answer:

Data

- Entrepreneurs: $n_1 = 125,\ \bar{x}_1 = 1.91,\ s_1 = 1.32$.
- Corporate managers: $n_2 = 86,\ \bar{x}_2 = 0.21,\ s_2 = 0.53$.

Hypotheses

Right-Tailed Test

- $H_0 : \mu_1 \leq \mu_2$
- $H_a : \mu_1 > \mu_2$ (mean number of job changes is higher for entrepreneurs)

EXERCISE 10.9: SOLUTION

Answer:

Test Statistic and Its Distribution

$$\boxed{T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t_\nu}$$

Where ν is the Welch–Satterthwaite degrees of freedom.

The value of the test statistic is t = 12.9603.

Compute the pieces:

- Difference of sample means: $\bar{x}_1 - \bar{x}_2 = 1.91 - 0.21 = 1.70.$
- Standard error:

$$t \approx \frac{1.70}{0.13117} \approx 12.9603.$$

$$\mathrm{SE} = \sqrt{\frac{1.32^2}{125} + \frac{0.53^2}{86}} = \sqrt{\frac{1.7424}{125} + \frac{0.2809}{86}} = \sqrt{0.0139392 + 0.0032669} = 0.13117$$

EXERCISE 10.9: SOLUTION

Answer:

Degrees of freedom (Welch–Satterthwaite)

$$\nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} \approx 174$$

α = 0.05
1- α = 0.95
$t_{1-\alpha;n-1}$ = $t_{0.95;175}$ ~ 1.645
(from the student´s t table)

Rejection region (one-sided, right tail): reject H_0 if $t > t_{\alpha,\nu}$.

- For $\alpha = 0.05$, $t_{0.95;174}$ ~1.645

RR = [1.645; +∞[

Compare: $12.9603 > 1.6536 \Rightarrow$ test statistic lies in the rejection region.

P-value = P(T > 12.9603) < 0.001
Approximate value (from the t-table)

P-value (one-sided, upper tail)

- Using $t \approx 12.9603$ with $\nu \approx$ 174 , the one-sided p-value is approximately

$$p \approx 1.16 \times 10^{-27}.$$

Exact value

EXERCISE 10.9: SOLUTION

Answer:

Conclusion $t = 12.9603 > 1.645$ and P-value < 0.05, then reject H_0

- **Reject** H_0. Both the test statistic and the p-value indicate strong evidence that British entrepreneurs have a higher mean number of job changes than British corporate managers.

95% confidence interval for $\mu_1 - \mu_2$ (two-sided)

- Critical $t_{0.975,\nu} \approx 1.9736$.
- CI:

$\mathrm{CI}_{95\%}\,(\mu_1 - \mu_2) = (1.4411, 1.9589)$

$$1.70 \pm 1.9736 \times 0.13117 \Rightarrow (1.4411,\ 1.9589).$$

Interpretation: We are 95% confident that entrepreneurs have on average between **≈1.44 and ≈1.96** more job changes than corporate managers.

SUMMARY - TESTS OF THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS: INDEPENDENT SAMPLE

The test statistic for $H_0 : \mu_x - \mu_y = 0$ is :

Note:
In a test for the difference between two independent population means, the **test statistic** differs depending on whether the population **variances are known or unknown,** and whether they are assumed **equal or unequal**.

2. Test Statistic

If population variances are known:

$$\boxed{Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim \text{Normal}(0,1)}$$

population variances are unknown (and assumed equal):

$$\boxed{T = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t_{n_1+n_2-2},} \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

population variances are unknown and unequal (Welch's t-test):

$$\boxed{T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t_\nu} \quad \nu \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}}$$

where ν is the approximate degrees of freedom.

Newbold et al (2013)

SUMMARY - TESTS OF THE DIFFERENCE BETWEEN TWO NORMAL POPULATION MEANS: INDEPENDENT SAMPLES

3. Decision Rule

- Using critical value(s): Reject H_0 if the test statistic falls in the **Rejection Region**.
- Using p-value: Reject H_0 if $p\text{-value} < \alpha$.

The test statistic follows a normal distribution.

Two-tailed test:

- Rejection Region: $|Z| \geq Z_{1-\alpha/2}$
- P-value: $2 \cdot P(Z \geq |z_0|)$

Right-tailed test:

- Rejection Region: $Z \geq Z_{1-\alpha}$
- P-value: $P(Z \geq z_0)$

Left-tailed test:

- Rejection Region: $Z \leq Z_{\alpha}$
- P-value: $P(Z \leq z_0)$

The test statistic follows a Student's t-distribution.

Two-tailed test:

- Rejection Region: $|T| \geq t_{1-\alpha/2,n-1}$
- P-value: $2 \cdot P(T_{n-1} \geq |t_0|)$

Right-tailed test:

- Rejection Region: $T \geq t_{1-\alpha,n-1}$
- P-value: $P(T_{n-1} \geq t_0)$

Left-tailed test:

- Rejection Region: $T \leq t_{\alpha,n-1}$
- P-value: $P(T_{n-1} \leq t_0)$

Note:

For tests comparing two independent sample means:

- If the population variances are known, use the Z test statistic.
- If the population variances are unknown but assumed equal, use the pooled T test statistic.
- If the population variances are unknown and unequal, use Welch's T test statistic.

HOMEWORK OF LECTURE 17: QUESTIONS

EXERCISE 10.10

10.10 A political science professor is interested in comparing the characteristics of students who do and do not vote in national elections. For a random sample of 114 students who claimed to have voted in the last presidential election, she found a mean grade point average of 2.71 and a standard deviation of 0.64. For an independent random sample of 123 students who did not vote, the mean grade point average was 2.79 and the standard deviation was 0.56. Test, against a two-sided alternative, the null hypothesis that the population means are equal.

Newbold et al (2013)

THANKS!

Questions?