STATISTICS II

Bachelor's degrees in Economics, Finance and Management

2nd year/2nd Semester
2025/2026

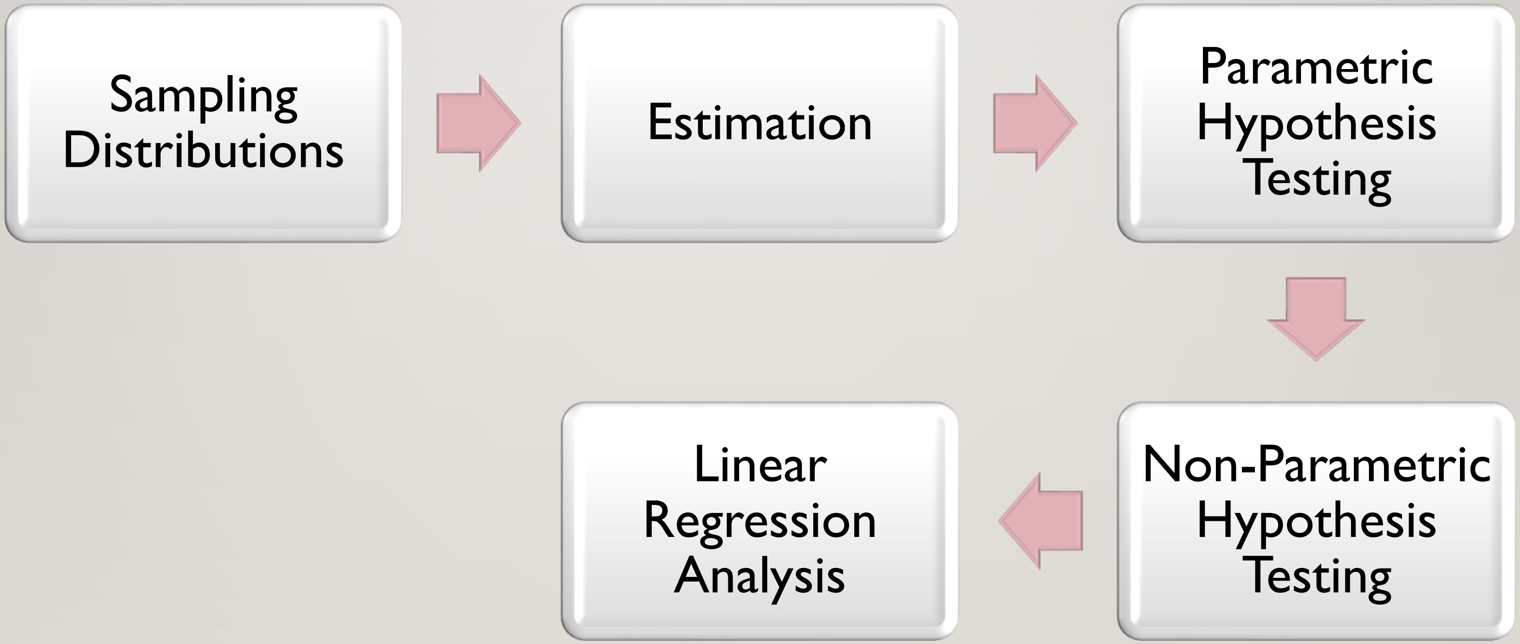
CONTACT

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 17: QUESTIONS AND SOLUTIONS

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)

EXERCISE 10.10: SOLUTION

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

Data

- Voters: $n_1 = 114,\ \bar{x}_1 = 2.71,\ s_1 = 0.64$.
- Non-voters: $n_2 = 123,\ \bar{x}_2 = 2.79,\ s_2 = 0.56$.

Hypotheses (two-sided)

Two-tailed Test

$$H_0 : \ \mu_1 = \mu_2 \qquad \text{vs.} \qquad H_a : \ \mu_1 \neq \mu_2$$

Test used

Welch two-sample t-test (variances unknown and not assumed equal).

EXERCISE 10.10: SOLUTION

✅ Answer:

Test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(\nu)$$

ν is given by the Welch–Satterthwaite formula. Calculated value:

$$\nu \approx 225.23.$$

Compute pieces:

- Difference of means: $\bar{x}_1 - \bar{x}_2 = 2.71 - 2.79 = -0.08.$
- Standard error:

$$\mathrm{SE} = \sqrt{\frac{0.64^2}{114} + \frac{0.56^2}{123}} = \sqrt{\frac{0.4096}{114} + \frac{0.3136}{123}} = 0.07837$$

Test statistic:

$$t \approx \frac{-0.08}{0.07837} \approx -1.0207.$$

EXERCISE 10.10: SOLUTION

✅ Answer:

P-value (two-sided)

Using $t \approx -1.0207$ with $\nu \approx 225.23$ (large df), the two-sided p-value is approximately

$$p \approx 0.307 \text{ (about 0.31).}$$

Rejection region (two-sided): reject H_0 if $|t| > t_{\alpha/2,\nu}$.

- For $\alpha = 0.05$, $t_{0.025,\nu} \approx 1.9706$.

Compare: $|-1.0207| = 1.0207 < 1.9706$ ⇒ test statistic **does not** lie in the rejection region.

EXERCISE 10.10: SOLUTION

✅ Answer:

95% confidence interval for $\mu_1 - \mu_2$

Critical value (approx.) $t_{0.975,\nu} \approx 1.971$.

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.975,\nu} \cdot \mathrm{SE} = -0.08 \pm 1.971(0.07837)$$

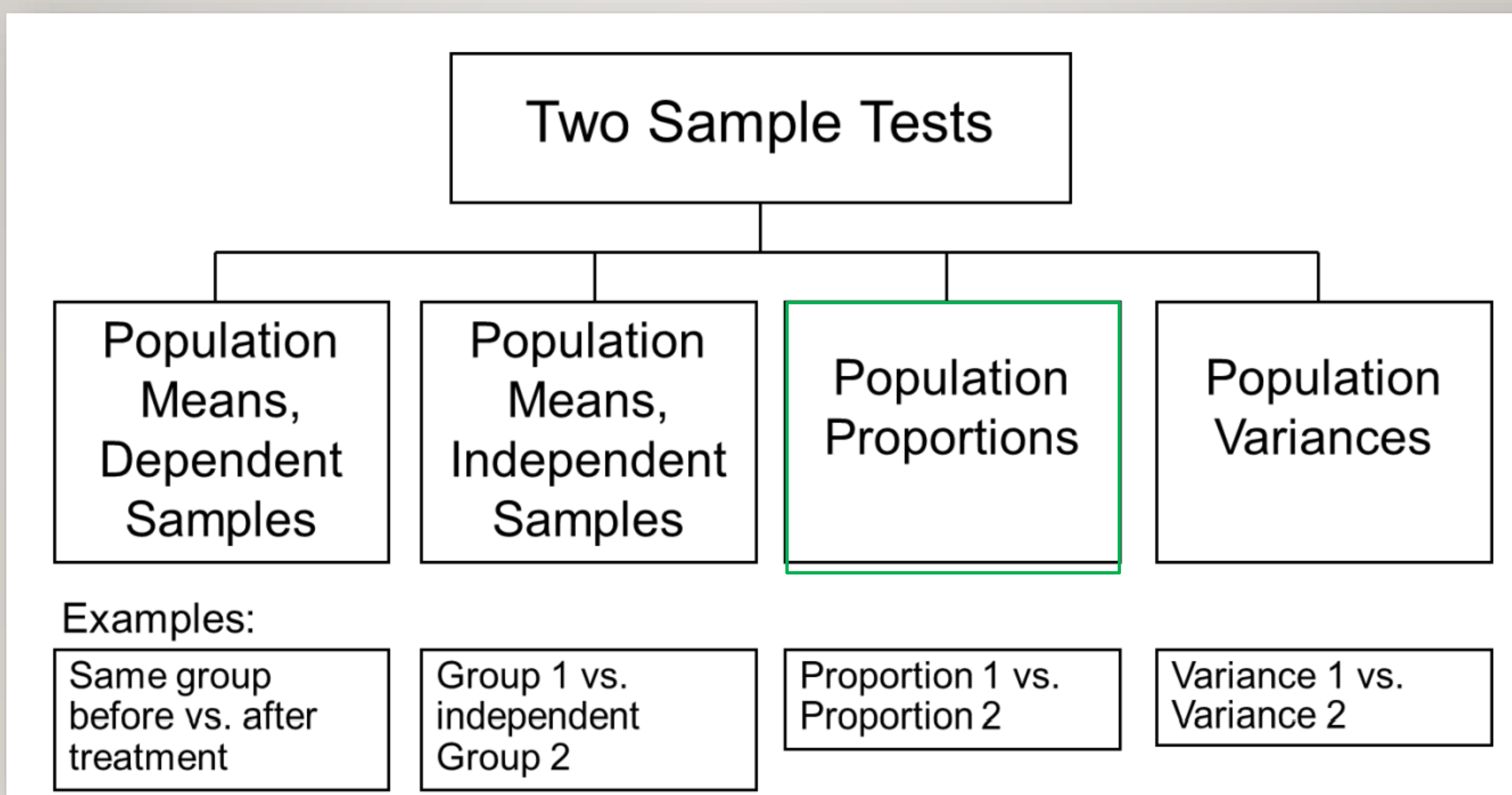
$$\Rightarrow \ \mathrm{CI} \approx (-0.234,\ 0.074).$$

Conclusion

- The two-sided p-value is about **0.307**, which is greater than common significance levels (e.g. 0.05).
- **Fail to reject H_0**. Both the test statistic and the p-value indicate no statistically significant difference in mean GPA between students who voted and those who did not.
- The 95% confidence interval for $\mu_1 - \mu_2$ contains 0 $(\approx -0.234 \text{ to } 0.074)$, consistent with the same conclusion.

LECTURE 18: TESTS OF THE DIFFERENCE BETWEEN TWO NORMAL POPULATION PROPORTIONS

TWO SAMPLE TESTS



Newbold et al (2013)

TWO POPULATION PROPORTIONS

Population proportions

Tests of the Difference Between Two Population Proportions (Large Samples)

Goal: Test hypotheses for the difference between two population proportions, $P_x - P_y$

Assumptions:

Both sample sizes are large,

$$nP(1-P) > 5$$

Newbold et al (2013)

TWO POPULATION PROPORTIONS

Population proportions

- The random variable

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (P_x - P_y)}{\sqrt{\frac{P_x(1-P_x)}{n_x} + \frac{P_y(1-P_y)}{n_y}}}$$

has a standard normal distribution

Newbold et al (2013)

TEST STATISTIC FOR TWO POPULATION PROPORTIONS

Population proportions

Two-tailed Test

$H_0: P_x - P_y = 0$

$H_1: P_x - P_y \neq 0$

The test statistic for

$$H_0 : P_x - P_y = 0$$

is a z value:

$$z = \frac{\left(\hat{p}_x - \hat{p}_y\right)}{\sqrt{\frac{\hat{p}_0\left(1-\hat{p}_0\right)}{n_x} + \frac{\hat{p}_0\left(1-\hat{p}_0\right)}{n_y}}}$$

Where $\hat{p}_0 = \dfrac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}$

Newbold et al (2013)

GENERAL HYPOTHESIS TEST FOR DIFFERENCE OF TWO PROPORTIONS

Hypothesis

$$H_0 : p_1 - p_2 = a$$

$$H_1 : p_1 - p_2 \neq a \quad \text{(or one-sided alternative)}$$

Two-tailed Test

H_0: $P_1 - P_2 = a$

H_1: $P_1 - P_2 \neq a$

Test Statistic

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - a}{\sqrt{\hat{p}_0(1 - \hat{p}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where the **pooled proportion under** H_0 is:

$$\hat{p}_0 = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} \iff \hat{p}_0 = \frac{x_1 + x_2}{n_1 + n_2}$$

- x_1, x_2: number of successes in each sample
- n_1, n_2: sample sizes
- Note: if $a \neq 0$, we still **pool under the null** assumption of equality shifted by a.

Note:
When $H_0: p_1 - p_2 = 0$,we assume the two population proportions are equal ($p_1 = p_2$) and use the **pooled variance** for the test.
When $H_0: p_1 - p_2 = a \neq 0$,the proportions are **not equal**, but we assume a **fixed difference** $a$under the null. The **pooled variance** is still used because it reflects the **expected variability of the difference under** H_0.

For **confidence intervals**, we do **not assume any fixed difference**, so we use the **separate variances** of each sample.
Key idea:
Pooled variance → for hypothesis testing under the null
Separate variances → for estimation (confidence intervals)

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

Confidence Interval for $p_1 - p_2$

A $100(1-\alpha)\%$ confidence interval for the difference of two population proportions is:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- Both population proportions p_1 and p_2 are **unknown**.
- The standard error is **estimated from the sample proportions** $\hat{p}_1$ and $\hat{p}_2$.
- The goal is **estimation** of $p_1 - p_2$.

Hypothesis Test for $p_1 - p_2$

To test:

$$H_0 : p_1 - p_2 = 0$$

the test statistic is:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}_0(1-\hat{p}_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where the **pooled proportion** is:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

- Under H_0, the two population proportions are assumed to be **equal**.
- The standard error is computed using the **pooled estimate** $\hat{p}$.
- The goal is **hypothesis testing**, not estimation.

Key Takeaway:

Confidence interval: uses separate sample proportions $\widehat{p_1}$ and $\widehat{p_2}$.

Hypothesis test: uses a **pooled proportion** based on H_0.

The statistics differ because the hypothesis test assumes equality under H_0, while the confidence interval does not.

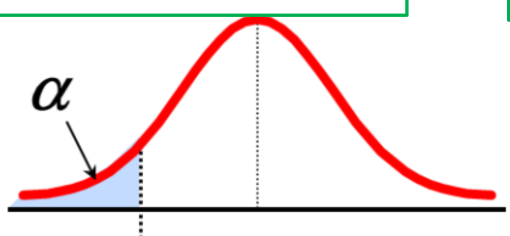
DECISION RULES: TEST STATISTIC FOR TWO POPULATION PROPORTIONS

Population proportions

Left-Tailed Test

$H_0: P_x - P_y \geq 0$

$H_1: P_x - P_y < 0$



z_{α}

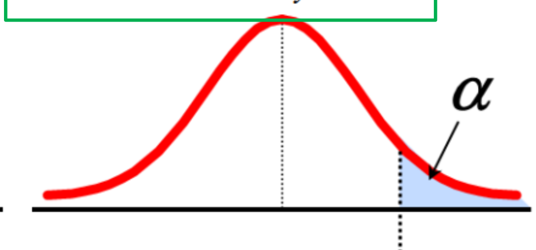
Reject H_0 if $z < z_{\alpha}$

RR = $]-\infty; z_{\alpha}]$

Right-Tailed Test

$H_0: P_x - P_y \leq 0$

$H_1: P_x - P_y > 0$



$z_{1-\alpha}$

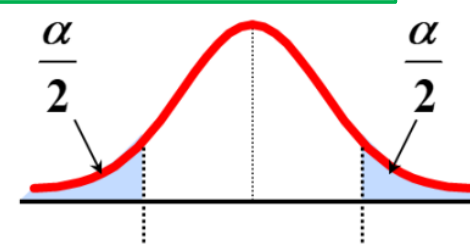
Reject H_0 if $z > z_{1-\alpha}$

RR = $[z_{1-\alpha}; +\infty[$

Two-Tailed Test

$H_0: P_x - P_y = 0$

$H_1: P_x - P_y \neq 0$



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

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

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

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

Newbold et al (2013)

TEST STATISTIC FOR TWO POPULATION PROPORTIONS: EXAMPLE

Is there a significant difference between the proportion of men and the proportion of women who will vote Yes on Proposition *A*?

- In a random sample, 36 of 72 men and 31 of 50 women indicated they would vote Yes
- Test at the .05 level of significance

Newbold et al (2013)

TEST STATISTIC FOR TWO POPULATION PROPORTIONS: EXAMPLE

Two-tailed Test

H_0: P_x - P_y = 0

H_1: P_x - P_y ≠ 0

- The hypothesis test is:

  $H_0 : P_M - P_W = 0$ (the two proportions are equal)

  $H_1 : P_M - P_W \neq 0$ (there is a significant difference between proportions)

- The sample proportions are:
  - Men: $\hat{p}_M = \frac{36}{72} = .50$
  - Women: $\hat{p}_W = \frac{31}{50} = .62$
- The estimate for the common overall proportion is:

$$\hat{p}_0 = \frac{n_M \hat{p}_M + n_W \hat{p}_W}{n_M + n_W} = \frac{72\left(\frac{36}{72}\right) + 50\left(\frac{31}{50}\right)}{72 + 50} = \frac{67}{122} = .549$$

Newbold et al (2013)

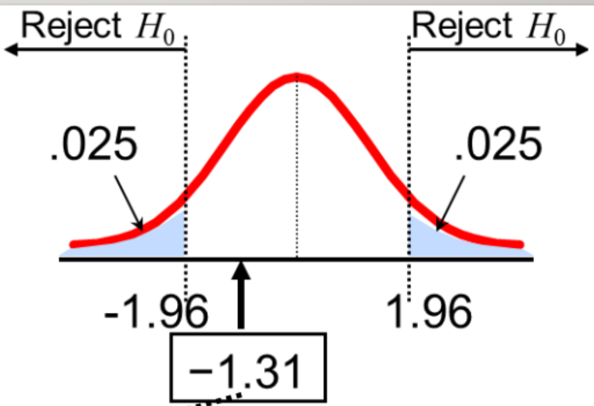
TEST STATISTIC FOR TWO POPULATION PROPORTIONS: EXAMPLE

The test statistic for $P_M - P_W = 0$ is:

$$z = \frac{(\hat{p}_M - \hat{p}_W)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_1} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_2}}}$$

$$= \frac{(.50 - .62)}{\sqrt{\left(\frac{.549(1-.549)}{72} + \frac{.549(1-.549)}{50}\right)}}$$

$$= \boxed{-1.31}$$



Decision: Do not reject H_0

Conclusion: There is not significant evidence of a difference between men and women in proportions who will vote yes.

Critical Values = ± 1.96

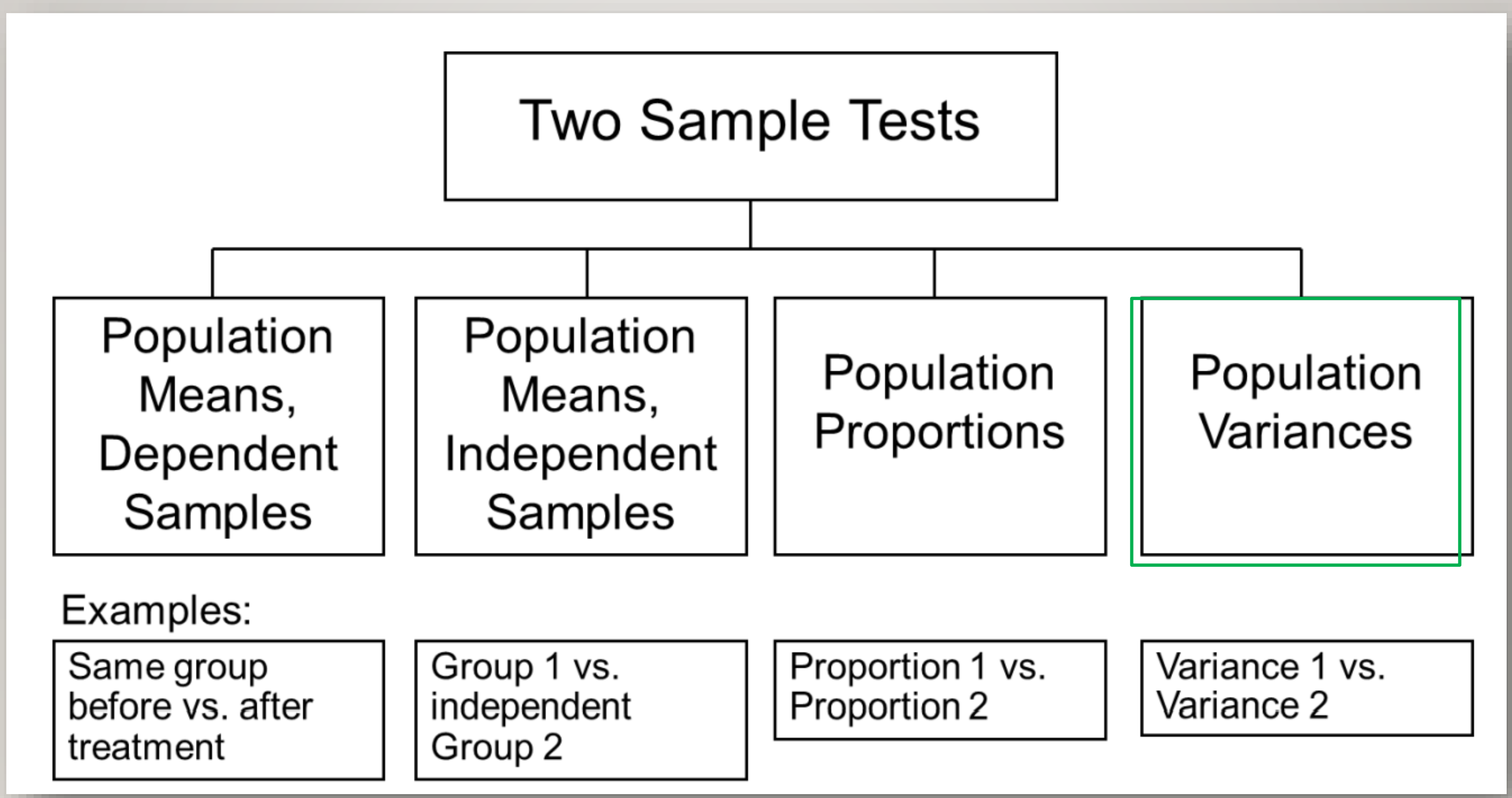
For $\alpha = .05$

RR =] -∞; -1.96]U[1.96; +∞[

Newbold et al (2013)

LECTURE 18: TESTS OF EQUALITY OF TWO VARIANCES

TWO SAMPLE TESTS



Newbold et al (2013)

TESTS OF EQUALITY OF TWO VARIANCES

Tests for Two Population Variances

F test statistic

- Goal: Test hypotheses about two population variances

$$H_0: \sigma_x^2 \geq \sigma_y^2$$
$$H_1: \sigma_x^2 < \sigma_y^2$$

Left-Tailed Test

$$H_0: \sigma_x^2 \leq \sigma_y^2$$
$$H_1: \sigma_x^2 > \sigma_y^2$$

Right-Tailed Test

$$H_0: \sigma_x^2 = \sigma_y^2$$
$$H_1: \sigma_x^2 \neq \sigma_y^2$$

Two-Tailed Test

The two populations are assumed to be independent and normally distributed

Newbold et al (2013)

TESTS OF EQUALITY OF TWO VARIANCES

Tests for Two Population Variances

↓

F test statistic

The random variable

$$F = \frac{\frac{s_x^2}{\sigma_x^2}}{\frac{s_y^2}{\sigma_y^2}}$$

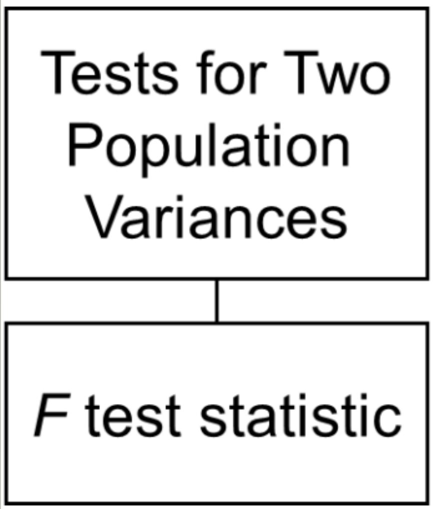
$$F = \frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2} = \frac{s_1^2}{s_2^2} \times \frac{\sigma_2^2}{\sigma_1^2}$$

Has an *F* distribution with $(n_x - 1)$ numerator degrees of freedom and $(n_y - 1)$ denominator degrees of freedom

Denote an *F* value with v_1 numerator and v_2 denominator degrees of freedom by F_{v_1, v_2}

Newbold et al (2013)

TESTS OF EQUALITY OF TWO VARIANCES: TEST STATISTIC



The critical value for a hypothesis test about two population variances is

$$F = \frac{s_x^2}{s_y^2}$$

where F has $(n_x - 1)$ numerator degrees of freedom and $(n_y - 1)$ denominator degrees of freedom

Newbold et al (2013)

F-TEST FOR TWO VARIANCES: REJECTION REGIONS

1. **Right-Tailed Test**

- Hypotheses:

$$H_0 : \sigma_1^2 \leq \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 > \sigma_2^2$$

- Statistic: $F = s_1^2 / s_2^2$
- Rejection Region: $F \geq F_{1-\alpha, df_1, df_2}$

RR = $[F_{1-\alpha}; +\infty[$

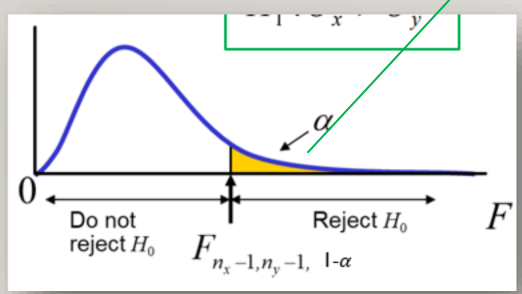
2. **Left-Tailed Test**

- Hypotheses:

$$H_0 : \sigma_1^2 \geq \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 < \sigma_2^2$$

- Statistic: $F = s_1^2 / s_2^2$
- Rejection Region: $F \leq F_{\alpha; df_1, df_2}$

RR = $[0; F_{\alpha}]$



3. **Two-Tailed Test**

- Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{vs} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

- Statistic: $F = s_1^2 / s_2^2$
- Rejection Region:

$$F \leq F_{\alpha/2, df_1, df_2} \quad \text{or} \quad F \geq F_{1-\alpha/2, df_1, df_2}$$

RR = $[0; F_{\alpha/2}] \cup [F_{1-\alpha/2}; +\infty[$

F-TEST FOR TWO VARIANCES: P-VALUES

Test statistic

$$F = \frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2} = \frac{s_1^2}{s_2^2} \times \frac{\sigma_2^2}{\sigma_1^2}$$

Under H_0, the test statistic follows an

$F(n_1 - 1,\ n_2 - 1)$ distribution.

1. **Left-Tailed Test**

Hypotheses:

$$H_0 : \sigma_1^2 \geq \sigma_2^2$$
$$H_1 : \sigma_1^2 < \sigma_2^2$$

p-value:

$$p\text{-value} = P(F_{n_1-1,\, n_2-1} \leq F_{\text{obs}})$$

2. **Right-Tailed Test**

Hypotheses:

$$H_0 : \sigma_1^2 \leq \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

p-value:

$$p\text{-value} = P(F_{n_1-1,\, n_2-1} \geq F_{\text{obs}})$$

3. **Two-Tailed Test**

Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

p-value:

$$p\text{-value} = 2 \times \min\left\{P(F_{n_1-1,\, n_2-1} \leq F_{\text{obs}}),\ P(F_{n_1-1,\, n_2-1} \geq F_{\text{obs}})\right\}$$

Note: Due to the asymmetry of the F distribution, the two-tailed p-value is obtained by doubling the probability corresponding to the more extreme tail.

SOME COMMENTS ON HYPOTHESIS TESTING

- A test with low power can result from:
  - Small sample size
  - Large variances in the underlying populations
  - Poor measurement procedures
- If sample sizes are large it is possible to find significant differences that are not practically important
- Researchers should select the appropriate level of significance before computing *p*-values

Newbold et al (2013)

HOMEWORK OF LECTURE 18: QUESTIONS

EXERCISE 10.15

10.15 Random samples of 900 people in the United States and in Great Britain indicated that 60% of the people in the United States were positive about the future economy, whereas 66% of the people in Great Britain were positive about the future economy. Does this provide strong evidence that the people in Great Britain are more optimistic about the economy?

Newbold et al (2013)

EXERCISE 10.16

10.16 A random sample of 1,556 people in country A were asked to respond to this statement: *Increased world trade can increase our per capita prosperity*. Of these sample members, 38.4% agreed with the statement. When the same statement was presented to a random sample of 1,108 people in country B, 52.0% agreed. Test the null hypothesis that the population proportions agreeing with this statement were the same in the two countries against the alternative that a higher proportion agreed in country B.

Newbold et al (2013)

EXERCISE 10.24

10.24 It is hypothesized that the total sales of a corporation should vary more in an industry with active price competition than in one with duopoly and tacit collusion. In a study of the merchant ship production industry it was found that in 4 years of active price competition, the variance of company A's total sales was 114.09. In the following 7 years, during which there was duopoly and tacit collusion, this variance was 16.08. Assume that the data can be regarded as an independent random sample from two normal distributions. Test, at the 5% level, the null hypothesis that the two population variances are equal against the alternative that the variance of total sales is higher in years of active price competition.

Newbold et al (2013)

EXERCISE 10.25

10.25 In light of a number of recent large-corporation bankruptcies, auditors are becoming increasingly concerned about the possibility of fraud. Auditors might be helped in determining the chances of fraud if they carefully measure cash flow. To evaluate this possibility, samples of midlevel auditors from CPA firms were presented with cash-flow information from a fraud case, and they were asked to indicate the chance of material fraud on a scale from 0 to 100. A random sample of 36 auditors used the cash-flow information. Their mean assessment was 36.21, and the sample standard deviation was 22.93. For an independent random sample of 36 auditors not using the cash-flow information, the sample mean and standard deviation were respectively 47.56 and 27.56.

Test the assumption that population variances for assessments of the chance of material fraud were the same for auditors using cash-flow information as for auditors not using cash-flow information against a two-sided alternative hypothesis.

Newbold et al (2013)

THANKS!

Questions?