Statistical Analysis

Hypothesis testing

Ana Cristina Costa

ccosta@novaims.unl.pt

Master in Statistics and Information Management

2021

Topics

- **LU6 – Hypothesis testing**
  - Concepts and methodology
  - Hypothesis testing for the mean
  - Hypothesis testing for the difference between means
  - Hypothesis testing for the proportion
  - Hypothesis testing for the difference between proportions
  - Hypothesis testing for the variance
  - Hypothesis testing for the ratio between variances
  - Correlation coefficient

Objectives

- **At the end of this learning unit students should be able to**
  - Explain the two types of error in statistical tests
  - Explain the power of a statistical test
  - Formulate the hypotheses of the statistical test and decide based on an appropriate test
  - Calculate and interpret the p-value
  - Apply one sample tests for means, proportions and variance
  - Apply tests for the difference between means
  - Apply tests for the difference between proportions
  - Apply the test for the ratio between variances
  - Apply the test for the correlation coefficient

Suggested reading

- Newbold, P., Carlson, W. L., Thorne, B. (2013). Statistics for Business and Economics. 8th Edition, Boston: Pearson, chapters 9 & 10.
- Mariappan, P. (2019). Statistics for Business. New York: Chapman and Hall/CRC, pages 225-300 (ch. 13).
- Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2013). *Introduction to Probability and Statistics*. 14th Edition, Boston: Brooks/Cole, Cengage Learning, pages 324-424.
- Holmes, A., Illowsky, B. & Dean, S. (2019) Introductory Business Statistics, chapters IX & X (accessed: August 2021)
- Oakley, J. (2021) MAS113 Part 2: Data Science, chapters 6-8. (accessed: August 2021)
- The Pennsylvania State University (2021) "Section 2: Hypothesis Testing". In STAT 415 Introduction to Mathematical Statistics. (accessed: August 2021)

Resources on the Internet

- Marshall, E., & Boggis, E. (2016). The statistics tutor's quick guide to commonly used statistical tests. Statstutor Community Project, 1-57, http://www.statstutor.ac.uk/resources/uploaded/tutorsquickguidetostatistics.pdf. (accessed: August, 2021)

- Koji Yatani (n.a.) "What statistical test should I use?", in *Statistical Methods for HCI Research*. (wiki with R examples): https://yatani.jp/teaching/doku.php?id=hcistats:start. (accessed: August, 2021)

- Addinsoft (2019) "Which statistical test should you use?", XLSTAT Support Center: https://help.xlstat.com/s/article/which-statistical-test-should-you-use?language=en_US. (accessed: August, 2021)

- Jagandeep Singh (n.a.) Statistical Tests with Python., https://medium.com/python-in-plain-english/statistical-tests-with-python-880251e9b572. (accessed: August, 2021)

Concepts and methodology

- **Introduction**
  - In this learning unit, a hypothesis test or a statistical test is a process used to decide between two complementary hypotheses based on
    - Formulating statistical hypotheses
      - ❑ “The batteries of brand A last the same as brand B batteries” *versus* “The brand A batteries last more than brand B batteries”
    - Using a statistical method to decide between the hypotheses
      - ❑ If the sample observations disagree with the statistical theory, the solution is to reject the assumptions made by the theory
      - ❑ If there is no disagreement, either the theory is true, or the sample did not identify differences between the actual values and the values of the population parameters that were assumed by the theory

Concepts and methodology

- Steps of hypothesis testing

1. **It is made an assumption about the value of the population parameter**
   - This assumption is called null hypothesis and is denoted by H_0

   Another assumption is stated
   - This assumption is called alternative hypothesis and is denoted by H_1
   - It is the opposite assumption defined by the null hypothesis

2. **A sampling statistic is selected to perform the test**

3. **The critical values of the test are determined**
   - They define the rejection region of the null hypothesis

4. **A statistical decision is made**
   - If the sample value of the test statistic falls in the rejection region, the null hypothesis is rejected

Concepts and methodology

Statistical hypotheses

- **H_0: Null hypothesis** is a maintained hypothesis that is held true unless sufficient evidence to the contrary is obtained
  - ☐ Always contains an equality
  - ☐ **It is the hypothesis that is accepted by default, without proof**

- **H_1: Alternative hypothesis** is the hypothesis against which the null hypothesis is tested, and which will be held to be true if the null hypothesis is held false based on sample data
  - ☐ Always contains an inequality ($>$, $<$ or $\neq$)
  - ☐ **It is the hypothesis stated by the researcher** (i.e., the one we believe is likely)

Concepts and methodology

- **Parametric tests considered**

1. **Two-sided test**
   - $H_0: \theta = \theta_0$
   - $H_1: \theta \neq \theta_0$

2. **One-sided test (to the right)**
   - $H_0: \theta \leq \theta_0$
   - $H_1: \theta > \theta_0$

3. **One-sided test (to the left)**
   - $H_0: \theta \geq \theta_0$
   - $H_1: \theta < \theta_0$

Concepts and methodology

- **Example: specify the null and alternative hypothesis**
  - Consider a car model that currently takes on average 40 km with 1 litre of fuel
  - A research team developed a new fuel injection system designed especially to increase the rate km / litter. To evaluate the new system, new vehicles will be built and road tests will be conducted, properly controlled.
  - In this case, the hypothesis posed by the researchers is that the new system will allow the vehicle to travel on average, more than 40 kilometres with 1 litter of fuel
  - In other words, *is* $\mu > 40$ *?*

H_0: $\mu \leq 40$

H_1: $\mu > 40$

Concepts and methodology

- **Example:** decision
  - If sampling results indicate that H_0 can not be rejected, researchers can not conclude that the new injection system of gasoline is best
    - ☐ In this case, they may need to investigate further and subsequently conduct new tests

  - However, **if the sampling results indicate that H_0 can be rejected**, researchers can infer that **H_1: $\mu > 40$ is true**
    - ☐ With this finding, researchers have statistical evidence to say that the new system increases, on average, the rate km/litter. The new cars are then fabricated with the developed system.

Concepts and methodology

- **Decision**
  - Ideally, the hypotheses testing procedure should allow to accept H_0 when H_0 is true, and reject H_0 when H_0 is false
  - Unfortunately, the correct conclusion is not always possible
  - As the findings are based on information provided by a sample, there is always the possibility of errors to occur

Concepts and methodology

- Type I and Type II errors
  - **Type I error** → reject H_0 when H_0 is true
  - **Type II error** → accept H_0 when H_0 is false

		Condition in the population	
		H_0 True	H_0 False
Decision	Reject H_0	Type I error	Correct decision
	Not reject H_0	Correct decision	Type II error

Concepts and methodology

- **Example:** decision

  Imagine a court and consider the differences between

  - **H_0**: Innocent *versus* **H_1**: Guilty
    - It means that we want to test the guilt of the individual
    - If there is not strong evidence, he will always be considered innocent. He will only be arrested if there is strong evidence of a crime.
  - **H_0**: Guilty *versus* **H_1**: Innocent
    - It means that we want to test the innocence of the individual
    - If there is not strong evidence, he will always be found guilty. In case of doubt, he is arrested.

Concepts and methodology

- **Example:** type I and type II errors

  Imagine a court and consider the differences between

  - **H_0**: Innocent *versus* **H_1**: Guilty
    - ❑ Type I error: the person is convicted, but is innocent
    - ❑ Type II error: the person is absolved, but is guilty

  - It is attributed more importance to the Type I error because the possibility of incorrectly rejecting H_0 is considered serious
    - ❑ This is the hypothesis to be defended, unless there is convincing evidence pointing in the opposite direction

Concepts and methodology

- **Observation**
  - Since, generally, the probability of Type II error occurring is not controlled, it is recommended to use the terms "**the null hypothesis is not rejected**" instead of "**the null hypothesis is accepted**"
  - If we state "we accept the null hypothesis", we may be committing a Type II error
  - But, if we claim "we can not reject the null hypothesis," there is an implicit recommendation to restrain judgment and action

Concepts and methodology

- **Significance level and power of the test**
  - α → is named **significance level of the test** and corresponds to the probability of committing a Type I error
  - **$1 – \beta$** → is named **power of the test** and corresponds to the probability of not committing a Type II error

		Condition in the population	
		H_0 True	**H_0 False**
Decision	**Reject H_0**	Type I error α	Correct decision **$1 – \beta$**
	Not reject H_0	Correct decision **$1 – \alpha$**	Type II error β

Concepts and methodology

- **Significance level and power of the test**
  - **Increasing the power of the test**
    - Increasing the probability of not committing a type II error
      - Increasing P(reject H_0 | H_0 false)
    - Decreasing the probability of committing a type II error β
      - Decreasing P(not reject H_0 | H_0 false)
    - Increasing the rejection region of H_0
      - Increasing P(reject H_0 | H_0 true)
    - Increasing the significance level α
    - **Increasing the probability of committing a type I error**

Magnusson, K. (n.a.) Understanding Statistical Power and Significance Testing: an Interactive Visualization. http://rpsychologist.com/d3/NHST/ (accessed 19 April 2018)

Concepts and methodology

- **Significance level and power of the test**
  - The only way to simultaneously minimize the two types of errors is increasing the sample size
  - **Neyman-Pearson approach to control the errors**
    - **Set as a fix value the probability** α associated with Type I error
    - Minimize the probability β associated with Type II error
      - It is equivalent to **maximize the power of the test: 1–β**
  - If one performs a large number of repeated samplings and decisions
    - If H_0 is true, H_0 is rejected in $\alpha \times 100\%$ of the cases
    - If H_0 is false, H_0 is rejected in $(1-\beta) \times 100\%$ of the cases

Concepts and methodology

- **Test statistic**
  - A **test statistic** is a function of the sample observations whose value will determine the conclusion from the statistical test
    - When we want to test the value of a parameter, the test statistic is usually an estimator of that parameter
    - The test statistics are typically those that are used for confidence interval estimation
  - The **observed test statistic** is a single number calculated from the sample data

Concepts and methodology

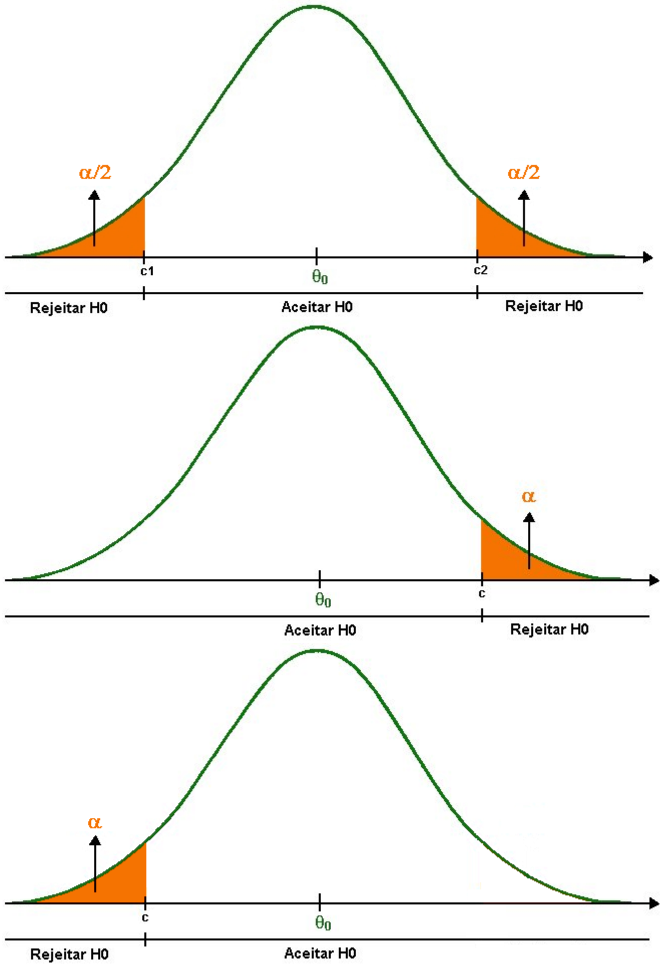
- **Critical values**
  - The critical value approach involves determining whether or not the observed test statistic is more extreme than would be expected if the null hypothesis were true. If the test statistic is more extreme than the critical value, then the null hypothesis is rejected in favour of the alternative hypothesis. If the test statistic is not as extreme as the critical value, then the null hypothesis is not rejected.
  - **Critical values** determine the set of values of the test statistic that leads to rejection of the null hypothesis. This set of values is called the critical region or **rejection region** of the null hypothesis

- **Rejection region of the null hypothesis**
  - The rejection region of the null hypothesis is the **region that**, if the null hypothesis is true, **contains the test statistic with probability** α

Concepts and methodology

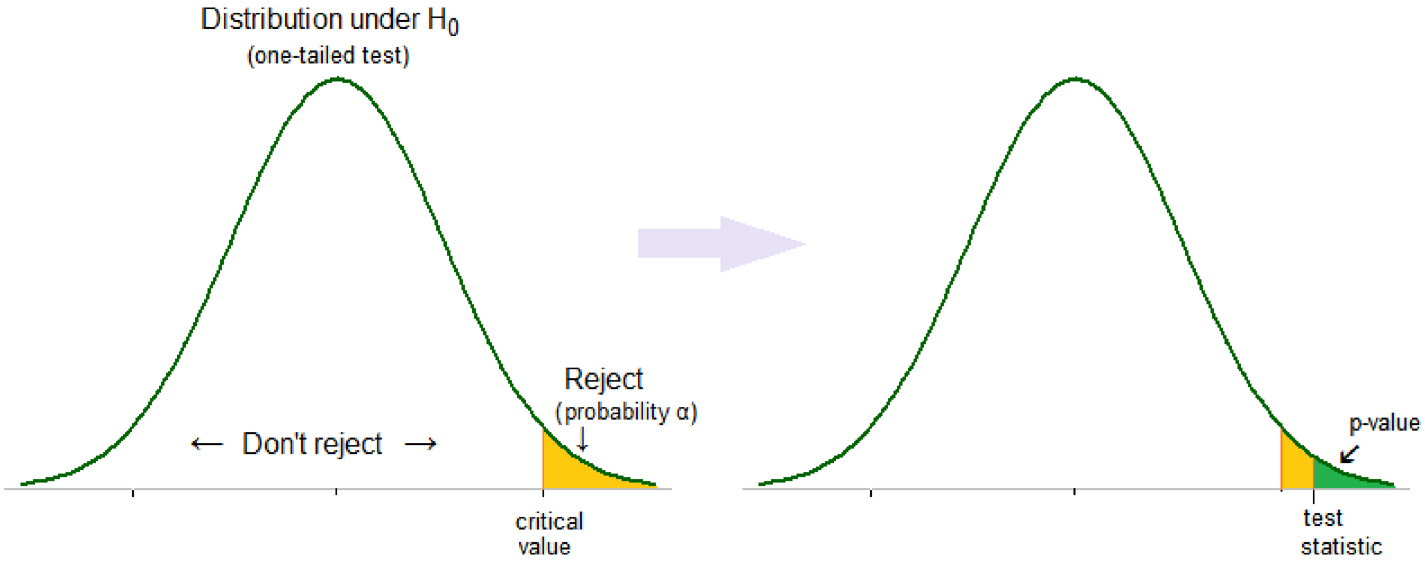
- **Statistical decision rule**
  - One rejects H_0 if the observed value of the test statistic falls in the rejection region

1. **Two-sided test**
   - $H_0: \theta = \theta_0$
   - $H_1: \theta \neq \theta_0$
2. **Right-sided test**
   - $H_0: \theta \leq \theta_0$
   - $H_1: \theta > \theta_0$
3. **Left-sided test**
   - $H_0: \theta \geq \theta_0$
   - $H_1: \theta < \theta_0$



Concepts and methodology

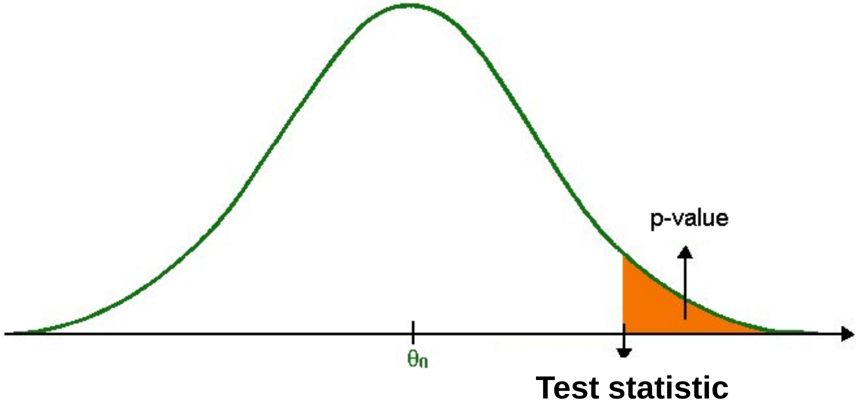
- **p-value**
  - The ***p-value*** is the smallest significance level at which H_0 can be rejected with the observed sample
  - For any test with significance level equal to α,
    - ☐ If ***p-value*** $\leq \alpha$ then **H_0 is rejected**



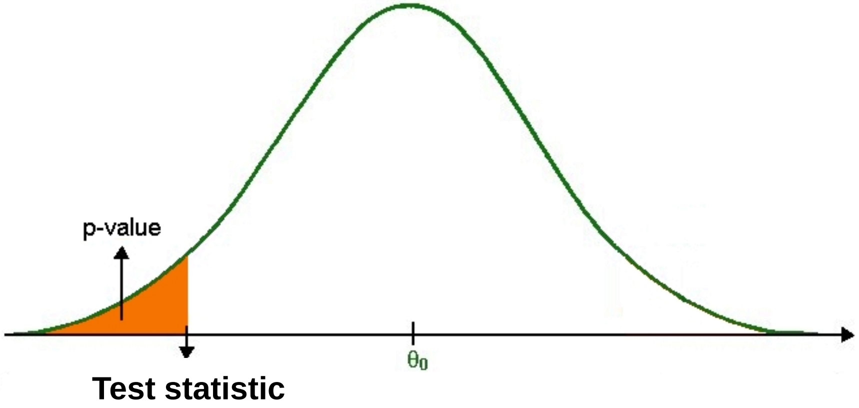
Concepts and methodology

- **p-value**

1. **Right-sided test**



2. **Left-sided test**

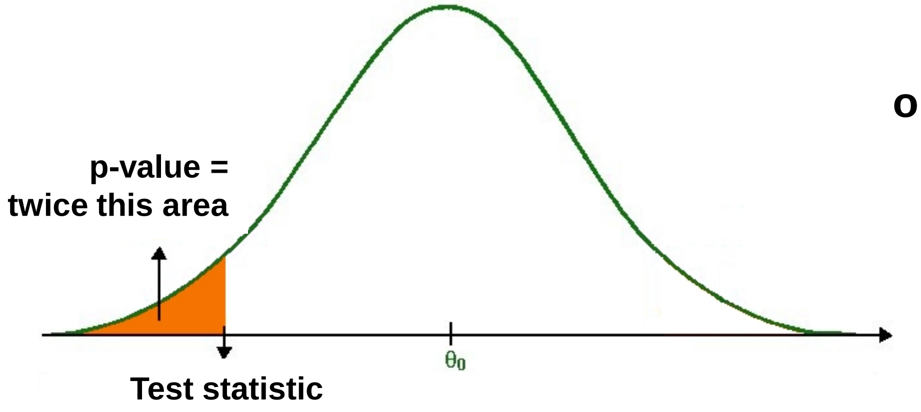


Concepts and methodology

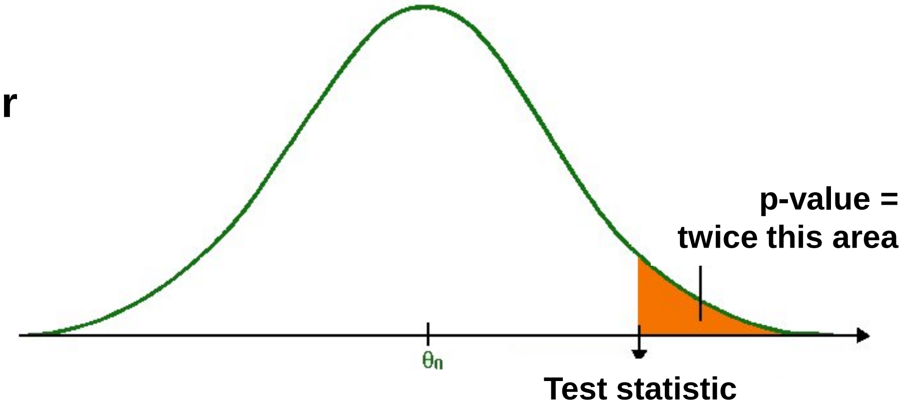
- **p-value**

  3. **Two-sided test**

The p-value for a two-sided test is always the double of a one-sided test



or



Hypothesis testing for the mean

Hypotheses test for μ – case I

- Let $X_1, X_2, \dots, X_n$ be a random sample of iid random variables from a Normal population with mean μ and known variance σ^2

❖ Normal population

❖ σ^2 known

❖ Any sample size

H_0: μ = μ_0

Test statistic

$$\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

- Let $X_1, X_2, \dots, X_n$ be a random sample of iid random variables with mean μ and finite variance σ^2

❖ Any population

❖ σ^2 known

❖ Large sample size

H_0: μ = μ_0

Test statistic

$$\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \overset{a}{\sim} N(0,1)$$

Hypothesis testing for the mean

- **Example 1**
  - A marketing company usually does surveys to determine the degree of satisfaction of car buyers. These surveys usually take an average of 12 minutes with a standard deviation of 3 minutes. To make it faster, it was decided to restructure it and test if the new survey took less time. Thus, 36 car buyers were randomly chosen, and the obtained average response time was 11.3 minutes.
    a) Can one conclude, with a significance level of 5%, that the new survey is more efficient?
    b) What is the lowest significance level for which you reject the null hypothesis previously tested?
  - Note: the marketing company believes that the standard deviation of the response time to the new survey is still three minutes.

Solved in the **Example1** sheet of the LU6_Examples Excel file

Hypothesis testing for the mean

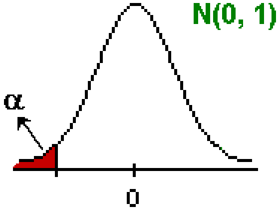
- **Example 1**

H_0: $\mu \geq 12$

H_1: $\mu < 12$

- $\sigma = 3,\ n = 36,\ \bar{x} = 11.3,\ \alpha = 5\%,\ z_{0.95} = 1.645$

- **Test statistic**: $\dfrac{\overline{X}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{11.3-12}{\dfrac{3}{\sqrt{36}}} = -1.4 > -1.645$



  - The observed test statistic does not belong to the rejection region of H_0

- **Decision**: do not reject H_0 at the 5% significance level. There is no evidence that the new survey is more efficient.

- *p-value* = P(Z < –1.4) = 0.0808

Hypothesis testing for the mean

- **Hypotheses test for μ – case II**
  - Let $X_1, X_2, \ldots, X_n$ be a random sample of iid random variables with mean μ and unknown variance σ^2

- Any population
- σ^2 unknown
- Large sample size

H_0: $\mu = \mu_0$

Test statistic

$$\frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \overset{a}{\sim} N(0, 1)$$

Hypothesis testing for the mean

- **Hypotheses test for μ – case III**
  - Let X_1, X_2, …, X_n be a random sample of iid random variables from a Normal population with mean μ and unknown variance σ^2

❖ Normal population

❖ σ^2 unknown

❖ Any sample size

H_0: μ = μ_0

Test statistic

$$\frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} \sim t_{(n-1)}$$

Hypothesis testing for the difference between means

- **Independent samples – case I: tests for $\mu_1 - \mu_2$**
  - Let $\widehat{X}_1$ and $\widehat{X}_2$ be the means of two mutually independent random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from two populations with means $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ and known variances $\boldsymbol{\sigma_1^2}$ and $\boldsymbol{\sigma_2^2}$, respectively.

❖ Normal populations
❖ σ_1 and σ_2 known
❖ Any sample sizes

H_0: $\mu_1 - \mu_2 = D_0$

Test statistic

$$\frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

❖ Any population
❖ σ_1 and σ_2 known
❖ Large sample sizes

H_0: $\mu_1 - \mu_2 = D_0$

Test statistic

$$\frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \overset{a}{\sim} N(0, 1)$$

Hypothesis testing for the difference between means

- **Independent samples – case II: Student's t-test for $\mu_1 – \mu_2$**
  - Let $\widehat{X}_1$ and $\widehat{X}_2$ be the means, and S_1^2 and S_2^2 the variances, of two mutually independent random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from two populations $N(\mu_1,\sigma)$ and $N(\mu_2,\sigma)$, respectively, where σ is unknown.

❖ Normal populations

❖ $\sigma_1 = \sigma_2$ unknown

❖ Any sample sizes

H_0: $\mu_1 – \mu_2 = D_0$

Test statistic

$$\frac{(\bar{X}_1-\bar{X}_2)-D_0}{S'\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(n_1+n_2-2)}$$

$$S'^2=\frac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{n_1+n_2-2}$$

Analysis Toolpak in Excel:
t-Test: Two-Sample Assuming Equal Variances

Hypothesis testing for the difference between means

- **Independent samples – case III: test for μ_1–μ_2**
  - Let $\widehat{X}_1$ and $\widehat{X}_2$ be the means, and S_1^2 and S_2^2 the variances, of two mutually independent random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from two populations with means $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$, respectively, and unknown variance $\boldsymbol{\sigma^2}$.

- Any populations
- $\sigma_1 = \sigma_2$ unknown
- Large sample sizes

H_0: $\mu_1 - \mu_2 = D_0$

Test statistic

$$\frac{(\bar{X}_1 - \bar{X}_2) - D_0}{S'\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \overset{a}{\sim} N(0,1)$$

$$S'^2 = \frac{(n_1 - 1)S_1^{\ 2} + (n_2 - 1)S_2^{\ 2}}{n_1 + n_2 - 2}$$

Hypothesis testing for the difference between means

- **Independent samples – case IV: Welch's test for $\mu_1 - \mu_2$**
  - Let $\widehat{X}_1$ and $\widehat{X}_2$ be the means, and S_1^2 and S_2^2 the variances, of two mutually independent random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from two populations $N(\boldsymbol{\mu_1},\boldsymbol{\sigma_1})$ and $N(\boldsymbol{\mu_2},\boldsymbol{\sigma_2})$, respectively, where the variances are unknown.

- ❖ Normal populations
- ❖ $\sigma_1 \neq \sigma_2$ unknown
- ❖ Any sample sizes

H_0: $\mu_1 - \mu_2 = D_0$

Test statistic

$$\frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(r)}$$

r is the **integer part** of: $r^* = \dfrac{\left(\dfrac{S_1^{\,2}}{n_1} + \dfrac{S_2^{\,2}}{n_2}\right)^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{S_1^{\,2}}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{S_2^{\,2}}{n_2}\right)^2}$

Analysis Toolpak in Excel:
t-Test: Two-Sample Assuming Unequal Variances

Hypothesis testing for the difference between means

- **Independent samples – case V: test for $\mu_1 – \mu_2$**
  - Let $\widehat{X}_1$ and $\widehat{X}_2$ be the means, and S_1^2 and S_2^2 the variances, of two mutually independent random samples of sizes $\mathbf{n_1}$ and $\mathbf{n_2}$ drawn from two populations with means $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ and unknown variances $\boldsymbol{\sigma_1^2}$ and $\boldsymbol{\sigma_2^2}$, respectively.

- Any populations
- $\sigma_1 \neq \sigma_2$ unknown
- Large sample sizes

H_0: $\mu_1 – \mu_2 = D_0$

Test statistic

$$\frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim} N(0,1)$$

Hypothesis testing for the difference between means

- **Student's t-test** (case II) ***versus*** **Welch's test** (case IV)
  - Both tests assume Normality
  - Student's t-test assumes equal population variances (Student, 1908)
  - Welch's test does not have the assumption of equal variances (Welch, 1947)
    - ☐ Delacre, Lakens, & Leys (2017) show that the Welch's test provides a better control of Type I error rates when the assumption of homogeneity of variance is not met, and loses little robustness compared to Student's t-test when the assumptions are met. Therefore, Welch's test should be used as a default strategy, regardless of the results of tests to compare variances (e.g., F-test or Levene test)
      - ☐ Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology*, *30*(1), in press. DOI: 10.17605/OSF.IO/SBP6K
      - ☐ Student (1908). The probable error of a mean. *Biometrika*, 1–25.
      - ☐ Welch, B. L. (1947). The generalization of Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28-35.

Welch's t-test is also known as the Satterwaite's test, or the Smith/Welch/Satterwaite test, or the Aspin-Welch test, or the unequal variances t-test

Hypothesis testing for the difference between means

- **Example 2**
  - In an experiment to compare two new analgesics, 65 volunteers were divided into two groups of n_1=35 and n_2=30 patients to whom was administered an equivalent dose of analgesics *1* and *2*, respectively
  - In the first group, the absence of pain lasted on average 6.3 hours with a standard deviation of 1.2 hours, while in the second group, the absence of pain lasted on average 5.2 hours with a standard deviation of 1.4 hours
  - Can one conclude, with a significance level of 5%, the analgesic 1 has a more lasting effect than the 2?
  - Note: doctors believe that the standard deviation of the time of absence of pain is the same for both painkillers

Solved in the **Example2** sheet of the LU6_Examples Excel file

Hypothesis testing for the difference between means

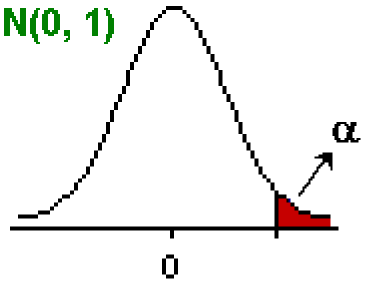
Example 2

H_0: $\mu_1 \leq \mu_2 \quad \Leftrightarrow \quad H_0$: $\mu_1 - \mu_2 \leq 0$

H_1: $\mu_1 > \mu_2 \quad \Leftrightarrow \quad H_1$: $\mu_1 - \mu_2 > 0$

- $n_1 = 35,\ n_2 = 30,\ \bar{x}_1 = 6.3,\ \bar{x}_2 = 5.2,\ s_1 = 1.2,\ s_2 = 1.4$
- **Test statistic**: $s'^2 = \dfrac{(35-1)(1.2)^2 + (30-1)(1.4)^2}{35+30-2} \approx 1.68 = 1.2959^2$

$$\frac{6.3-5.2-0}{1.2959\sqrt{\frac{1}{35}+\frac{1}{30}}} \approx 3.41 > 1.645$$



  - The observed test statistic falls in the rejection region of H_0

- **Conclusion**: there is evidence that the painkiller 1 has a more lasting effect than the painkiller 2, for a significance level of 5%

Solved in the **Example2** sheet of the LU6_Examples Excel file

Hypothesis testing in Excel

- **Example 3**
  - The placement director at a major MBA program wants to know, do marketing graduates and finance graduates get different starting salaries? Consider the data of marketing and finance students' starting salaries in the **Example3** sheet of the LU6_Examples Excel file.
    - ❑ Assume the populations' distributions are unknown (*even though normality of each sample should be tested before continuing*)
    - ❑ Assume the population variances are different *(if populations are not normal, this should be tested before continuing*)
  - ❑ Compute the samples variances using the VAR.S function. You will need these values as an input for the **Analysis Toolpak** add-in, and it will not take cell references.
    - ❑ Select "z-Test: Two Sample for Means" from the Data Analysis button of the DATA menu

Hypothesis testing for the difference between means

- **Paired samples & normality: tests for $\mu_1 - \mu_2$**
  - Let **$D=X_1-X_2$** be a Normal population with mean **$\mu_D=\mu_1-\mu_2$** and standard deviation **σ_D**, from which it is drawn a random sample of paired values $D_i=X_{1i}-X_{2i}$, $i=1,2,\dots,n$, with mean $\bar{D}$ and variance **S_D^2**.

❖ Normal population of differences

❖ σ_D known

❖ Any sample size

H_0: $\mu_1 - \mu_2 = D_0$

Test statistic

$$\frac{\bar{D}-D_0}{\sigma_D/\sqrt{n}} \sim N(0,1)$$

❖ Normal population of differences

❖ σ_D unknown

❖ Any sample size

H_0: $\mu_1 - \mu_2 = D_0$

Test statistic

$$\frac{\bar{D}-D_0}{S_D/\sqrt{n}} \sim t_{(n-1)}$$

Analysis Toolpak in Excel:

t-Test: Paired Two Sample for Means

Hypothesis testing for the difference between means

- **Paired samples & non-normality: tests for $\mu_1 – \mu_2$**

  - Any population of differences
  - σ_D known
  - Large sample size

H_0: $\mu_1 – \mu_2 = D_0$

Test statistic

$$\frac{\bar{D}-D_0}{\sigma_D/\sqrt{n}} \overset{a}{\sim} N(0,1)$$

  - Any population of differences
  - σ_D unknown
  - Large sample size

H_0: $\mu_1 – \mu_2 = D_0$

Test statistic

$$\frac{\bar{D}-D_0}{S_D/\sqrt{n}} \overset{a}{\sim} N(0,1)$$

Hypothesis testing in Excel

- **Example 4**
  - We want to test if a new type of insulation reduces heating bills. A sample of ten pairs of two houses with roughly the same size and same design that had the same heating bill last winter was selected. One member of the pair kept their old insulation while the other one installed a new insulation. Does the new insulation improve heating bills over the old insulation?
    - Note that the only difference in each pair of houses on the heating bill is the new type of insulation versus the old one. The data was collected as a paired-sample.
    - ☐ Assume the population of differences is Normal
    - ☐ Consider the data on changes in the heating bills ($) in the **Example4** sheet of the LU6_Examples Excel file
    - ☐ Select "t-Test: Paired Two Sample for Means" from the Analysis Toolpak add-in menu

Hypothesis testing for the proportion

- **Hypotheses test for *p***
  - Let $\widehat{\mathrm{P}}$ be the proportion of successes in an iid random sample of size *n*

❖ Bernoulli population

❖ Large sample size

H_0: $p = p_0$

Test statistic

$$\frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \overset{a}{\sim} N(0, 1)$$

Hypothesis testing for the difference between proportions

Hypotheses tests for $p_1 - p_2$

- Let $\widehat{P}_1$ and $\widehat{P}_2$ be the proportions of successes in two independent random samples with *large* sizes n_1 and n_2

- ❖ Bernoulli populations
- ❖ Large samples sizes

H_0: $p_1 - p_2 = 0$

Test statistic

$$\frac{\widehat{P}_1 - \widehat{P}_2}{\sqrt{\frac{\widehat{P}(1-\widehat{P})}{n_1} + \frac{\widehat{P}(1-\widehat{P})}{n_2}}} \overset{a}{\sim} N(0,1)$$

$$\widehat{P} = \frac{n_1\widehat{P}_1 + n_2\widehat{P}_2}{n_1 + n_2}$$

Under H_0, $p_1=p_2=p$. Hence, an estimate of p is the weighted average of p_1 and p_2.

H_0: $p_1 - p_2 = D_0$

Test statistic

$$\frac{\widehat{P}_1 - \widehat{P}_2 - D_0}{\sqrt{\frac{\widehat{P}_1\,(1-\widehat{P}_1)}{n_1} + \frac{\widehat{P}_2\,(1-\widehat{P}_2)}{n_2}}} \overset{a}{\sim} N(0,1)$$

Analysis Toolpak in Excel:
z-Test: Two-Sample for Means, using variance = p(1–p) for each sample

Hypothesis testing for the difference between proportions

- **Example 5**
  - It is suspected that the difference between the proportion of workers (p_1) and industrials (p_2) that are favourable to the tax reform is large (over 50%). There were two surveys.
    - ☐ The first involved 50 workers, chosen at random, in which 84% of respondents expressed themselves in favour of tax reform.
    - ☐ The second was carried out with 30 industrials, chosen at random, and 90% of them were against.
  - Determine if there is statistical evidence to support the existing assumption, with a significance level of 5%.

Solved in the **Example5** sheet of the LU6_Examples Excel file

Hypothesis testing for the difference between proportions

- **Example 5**

  H_0: $p_1 - p_2 \leq 0.5$

  H_1: $p_1 - p_2 > 0.5$

  - $n_1 = 50,\ n_2 = 30,\ \hat{p}_1 = 0.84,\ \hat{p}_2 = 0.10,\ z_{0.95} = 1.645$

  - **Test statistic**:

$$\frac{\left(\hat{P}_1 - \hat{P}_2\right) - 0.5}{\sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}} = \frac{(0.84 - 0.10) - 0.5}{\sqrt{\frac{0.84 \times 0.16}{50} + \frac{0.10 \times 0.9}{30}}} = 3.182 > 1.645$$

  - The observed value of the test statistic falls in the rejection region of H_0

  - **Decision**: There is evidence that the difference between the proportion of workers and industrials that are favourable to the tax reform is greater than 50%, at the 5% significance level

Hypothesis testing in Excel

- **Example 6**
  - Northern States Marketing Research has been asked to determine if an advertising campaign for a new cell phone increased customer recognition of the new World A phone. A random sample of 270 residents of a major city were asked if they knew about the World A phone before the advertising campaign. After the advertising campaign, a second random sample of 203 residents were asked exactly the same question using the same protocol. Do these results provide evidence that customer recognition increased after the advertising campaign?
    - Consider the data on changes in the heating bills ($) in the **Example6** sheet of the LU6_Examples Excel file
    - Select “z-Test: Two-Sample for Means” from the Analysis Toolpak add-in menu, and input the samples variances equal to
      - ✓ $s^2{}_1 = \hat{p}_1(1-\hat{p}_1) = 0.151$ $s^2{}_2 = \hat{p}_2(1-\hat{p}_2) = 0.24$

Hypothesis testing for the variance

- **Hypotheses test for σ^2**
  - Let S^2 be the variance of an iid random sample of size *n* drawn from a Normal population with variance σ^2

❖ Normal population

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

Test statistic

$$\frac{(n-1)S^2}{{\sigma_0}^2} \sim {\chi^2}_{(n-1)}$$

Hypothesis testing for the variance

- **Example 7**
  - In a marketing campaign to attract investors, the administration claims that investment in its shares on the stock exchange is safe and that the standard deviation of the value of the shares is less than 2 €.
  - Suppose you are a potential investor and that, before applying your capital, you randomly chose 30 days of the last three years and noted the value of the shares, getting s = 1,70 €.
  - Does this value suggest, with a significance level of 5%, that the administration is correct?

Solved in the **Example7** sheet of the LU6_Examples Excel file

Hypothesis testing for the variance

- **Example 7**

$H_0: \sigma \geq 2 \quad \Leftrightarrow \quad H_0: \sigma^2 \geq 4$

$H_1: \sigma < 2 \quad \Leftrightarrow \quad H_1: \sigma^2 < 4$

- $n = 30,\ s^2 = (1.70)^2 = 2.89,\ \chi^2_{(29)0.05} = 17.708$

- **Test statistic**:

$$\frac{(n-1)s^2}{\sigma_0{}^2} = \frac{29 \times 2.89}{4} = 20.953 > 17.708$$

- **Decision**: Do not reject H_0 at the 5% significance level. There is no evidence that the administration is correct.

Solved in the **Example7** sheet of the LU6_Examples Excel file

Hypothesis testing for the ratio between variances

- **Hypotheses test for σ_1^2/ σ_2^2**
  - Let S_1^2 and S_2^2 be the variances of two iid random samples, mutually independent, of sizes n_1 and n_2, from two Normal populations

❖ Normal populations

H$_0$: $\boldsymbol{\frac{\sigma_1^2}{\sigma_2^2} = \left(\frac{\sigma_1^2}{\sigma_2^2}\right)_0}$

Test statistic

$$\frac{S_1^2}{S_2^2}\left(\frac{\sigma_2^2}{\sigma_1^2}\right)_0 \sim F_{(n_1-1;\, n_2-1)}$$

Hypothesis testing for the ratio between variances

- **Example 8**

  Consider the problem of Example 2

  - In an experiment to compare two new analgesics, 65 volunteers were divided into two groups of n_1=35 and n_2=30 patients to whom was administered an equivalent dose of analgesics *1* and *2*, respectively

  - In the first group, the absence of pain lasted on average 6.3 hours with a standard deviation of 1.2 hours, while in the second group, the absence of pain lasted on average 5.2 hours with a standard deviation of 1.4 hours

  - Can the doctors conclude that the standard deviation of the time of absence of pain is the same for both painkillers?

Solved in the **Example8** sheet of the LU6_Examples Excel file

Hypothesis testing for the ratio between variances

- **Example 8**

H$_0$: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ H$_1$: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$

- $n_1 = 35,\ n_2 = 30,\ \ s_1 = 1.2,\ s_2 = 1.4$
- **Test statistic**:

$$\frac{S_1^2}{S_2^2} \times 1 = \frac{1.2^2}{1.4^2} = 0.7347$$

- Critical values: $F_{(34,\ 29;\ 0.025)} = 0.4949$; $F_{(34,\ 29;\ 0.975)} = 2.0623$
- **Decision**: Do not reject H$_0$ at the 5% significance level. There is not enough evidence to conclude that the standard deviation of the time of absence of pain is not the same for both painkillers.

Solved in the **Example8** sheet of the LU6_Examples Excel file

Hypothesis testing in Excel

- **Example 9**
  - A sample of students took Statistics in a hybrid class, which means they had online lectures. Another sample of students took classes in person. They all took the same test. Is there a significant mean difference between the tests' scores?
    - Consider the tests' scores of the "hybrid" and "in person" students in the **Example9** sheet of the LU6_Examples Excel file
    - Use the Analysis Toolpak to
      a) Test if the population variances are different
        - Assume the populations' distributions are Normal (but, normality of each sample should be tested before applying this test)

      b) Test if the population means are different

Correlation coefficient

- **Correlation coefficient**
  - Consider the population associated with the random pair (X, Y) with bivariate normal distribution, i.e. **$(X, Y) \sim N(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$**. Let (X_1,Y_1), (X_2,Y_2), …, (X_n,Y_n) be a random sample drawn from this population.
  - ρ describes the degree of linear association between the r.v. X and Y
  - The sample correlation coefficient R is the maximum likelihood estimator of ρ and is given by

$$R = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\left(\sum (X_i - \overline{X})^2\right)\left(\sum (Y_i - \overline{Y})^2\right)}} = \frac{n\sum X_i Y_i - \left(\sum X_i\right)\left(\sum Y_i\right)}{\sqrt{\left(n\sum X_i^2 - \left(\sum X_i\right)^2\right)\left(n\sum Y_i^2 - \left(\sum Y_i\right)^2\right)}}$$

Correlation coefficient

- **Hypotheses test for independence**
  - **H_0: ρ = 0**
    - Assuming that the null hypothesis is true (**ρ=0**), the sampling distribution of R is symmetric with respect to the line R = 0 in the interval [–1, 1]

Test statistic

$$\frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \sim t_{(n-2)}$$

Excel: Real Statistics Resource Pack

Add-Ins menu > Real Statistics > Data Analysis tools > CORR tab + Correlation tests

Correlation coefficient

- **Hypotheses test for ρ**
  - **H_0: $\rho = \rho_0 \neq 0$**
    - Assuming that the null hypothesis is true ($\rho\neq 0$), the sampling distribution of R is quite complex. The following result is valid even when *n* has a value as low as ten.

Test statistic

$$\frac{\frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)-\frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{n-3}}}\overset{a}{\sim} N(0,1)$$

Correlation coefficient

- **Example 10a**
  - A sample of 10 students was taken at an university, and the classifications for measuring math skills upon entry into the first year were recorded, as well as the final grades of the Calculus course at the end of the first semester.
    - ☐ Consider the grades of the math skills (X) and of the Calculus course (Y) in the **Example10a** sheet of the LU6_Examples Excel file
    - ☐ Is there sufficient evidence that the classifications for measuring math skills (X) and the final grades of the Calculus course (Y) are correlated?

H_0: X and Y independent. So, assuming that (X, Y)~Normal:

$H_0: \rho = 0$

$H_1: \rho \neq 0$

Correlation coefficient

- **Example 10a**
  - $n = 10$, $r = 0.8398$, $t_{(10-2)0.975} = 2.306$, $t_{(10-2)0.025} = -2.306$
  - **Test statistic**

$$\frac{0.8398\sqrt{8}}{\sqrt{1-0.8398^2}} = 4.375 > 2.306$$

  - **Decision**: Reject H_0 at the 5% significance level.
  - **Conclusion**: There is evidence that X and Y are not independent, at the 5% significance level
  - *p-value* $= 2 \times P(t_{(8)} > 4.375) = 2 \times 0.0012 = 0.0024$

Correlation coefficient

- **Example 10b**
  - Is there sufficient evidence that the classifications for measuring math skills (X) and the final grades of the Calculus course (Y) have a correlation coefficient greater than 0.5?
    - ☐ Consider the grades of the math skills (X) and of the Calculus course (Y) in the **Example10b** sheet of the LU6_Examples Excel file

Assuming that (X, Y)~Normal

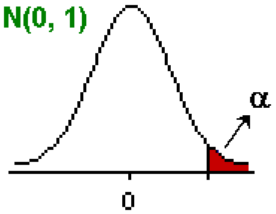
H_0: $\rho \leq 0.5$

H_1: $\rho > 0.5$

Correlation coefficient

- **Example 10b**
  - n = 10, r = 0.8398, $z_{(0.95)}$ = 1.645
  - **Test statistic**:

$$\frac{{}^{1}\!/_{2}\ln\left(\frac{1+0.8398}{1-0.8398}\right)-{}^{1}\!/_{2}\ln\left(\frac{1+0.5}{1-0.5}\right)}{{}^{1}\!/\sqrt{10-3}}=1.7758>1.645$$



  - The observed value of the test statistic falls in the rejection region of H_0
  - **Decision**: The sample provides sufficient evidence that ρ > 0.5, at the 5% significance level
  - *p-value* = P(Z > 1.7758) = 1 – 0.9621 = 0.0379

Hypothesis testing

Do the homework*!*

Ana Cristina Costa, ccosta@novaims.unl.pt