# Introduction online course on Mathematics and Statistics

Preparatory Course for M.Sc. Integrated Natural Resource Management

## 4 Hypothesis Testing

Department for Agricultural Economics  |  Resource Economics Group

**Syllabus**

4 Hypothesis Testing

Example:

- Those who are pursuing an academic career, obtain their PhD degree on the average at the age of 31.7
- A sample of doctoral graduates from Germany shows that the average age of obtaining the PhD degree is 33.
- 31.7 < 33 → doctoral graduates in Germany are older than doctoral candidates elsewhere?

A *hypothesis* is a statement about some characteristic of a variable or a collection of variables [1].

The *null hypothesis* ($H_0$) is a statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters [1].

The *alternative hypothesis* ($H_1$) is a statistical hypothesis that states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters [1].

→ One expects to find out whether the age of obtaining the PhD degree has increased/decreased during the last decade. In this case, since one knows that the average age of becoming a doctoral graduate overall is 31.7, the hypotheses would be:

$H_0$: $\mu = 31.7$ and $H_1$: $\mu \neq 31.7$ (or $\mu > 31.7$; or $\mu < 31.7$).

- The means of samples of a specific size, which are selected from a population, vary about the population mean [1]:
→ the null hypothesis is true and the difference between the sample and population means is due to chance and is insignificant;
→ there is an evidence for rejecting the null hypothesis as the sample mean is getting farther away from the population mean.

→ How to define the point from which we decide to reject the null hypothesis?

A ***statistical test*** uses the data obtained from a sample to make a decision about whether the null hypothesis should be rejected [1].

The ***numerical value*** obtained from a statistical test is called the test value [1].

- Possible outcomes of a hypothesis test [1]:
  - ➢ two possibilities for a correct decision
    - the null hypothesis is true and is not rejected
    - the null hypothesis is false and is rejected

  - ➢ two possibilities for an incorrect decision
    - the null hypothesis is false and is not rejected
    - the null hypothesis is true and is rejected

A *type I error* occurs when the null hypothesis is rejected though being true [1].

A *type II error* occurs when the false hypothesis is not rejected [1].

→ The decision is made based on probability and not on the statistical analysis of the whole population

- The difference between the sample and population means: how much the difference should be in order to reject the null hypothesis. To find this, the level of significance matters [1].

The *level of significance* is the maximum probability of committing a type I error. This probability is symbolized by α → P(type I error) = α [1].

- The probability of a type II error is symbolized by β, the Greek letter beta. → P(type II error) = β.
- Three significance levels [1]:
  - 0.10 → 10% of rejecting the null hypothesis or P(type I error) = 0.10 (10%)
  - 0.05 → 5% of rejecting the null hypothesis or P(type I error) = 0.05 (5%)
  - 0.01 → 1% of rejecting the null hypothesis or P(type I error) = 0.01 (1%)

The *critical value* (C.V.) is the value that separates the critical region from the noncritical region [1].

The *critical or rejection region* is the range of values of the test value that indicates that there is a significant difference and that the null hypothesis should be rejected [1].

The *noncritical or nonrejection region* is the range of values of the test value that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected [1].

HUMBOLDT-UNIVERSITÄT ZU BERLIN

HUMBOLDT-UNIVERSITÄT ZU BERLIN

**Resource Economics**

A **one-tailed test** indicates that the null hypothesis should be rejected when the test value is in the critical region on one side of the mean. A one-tailed test is either a right-tailed test or left-tailed test, depending on the direction of the inequality (> or <) of the alternative hypothesis [1].

To define the critical value[1]:

▪ **A right-sided test**: $\mu > 6,463$ and P(type I error) = 0.10 (10%)

→ To find the corresponding z value: 10% of the area falls to the right of the z value and 90% - to the left.

→ The critical value would be then 1.28, since it corresponds to the area closest to 0.9000 (0.8997).

Standard Normal Distribution Table [10]
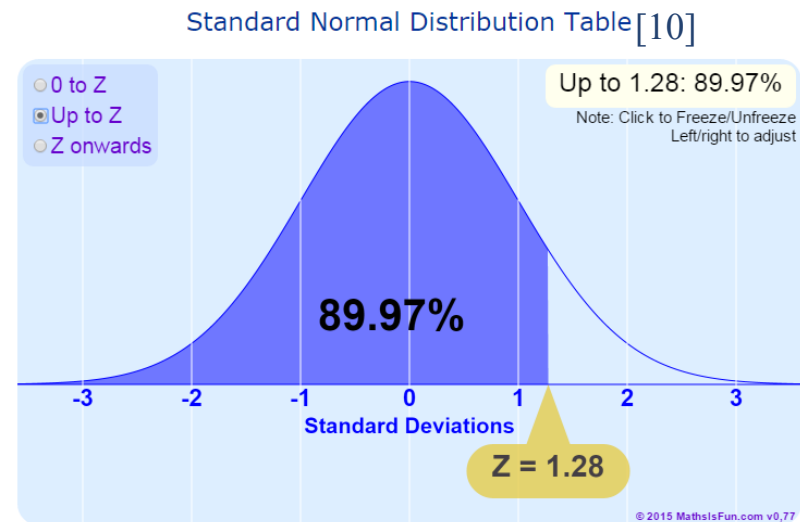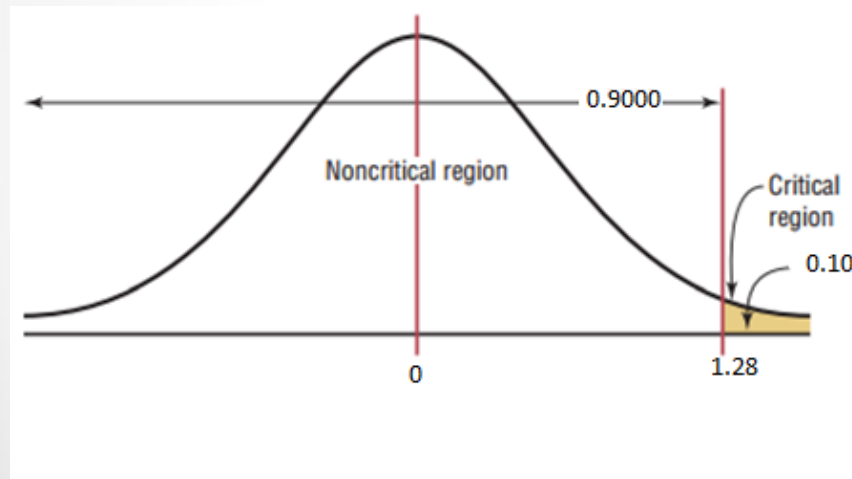


*Finding the Critical Value for α = 0.10 (a right-sided test) [1].*

➢ The critical region, with an area equal to α, is on the right side of the mean.
➢ To find the critical value, use the z value that corresponds to the area equivalent to 1 - α.

- **A left-sided test** [1]: $\mu < 6{,}463$ and P(type I error) = 0.10 (10%)
→ To find the corresponding z value: 90% of the area falls to the right of the z value and 10% - to the left.
→ The critical value would be then -1.28, since it corresponds to the area closest to 0.1000 (0.1003).
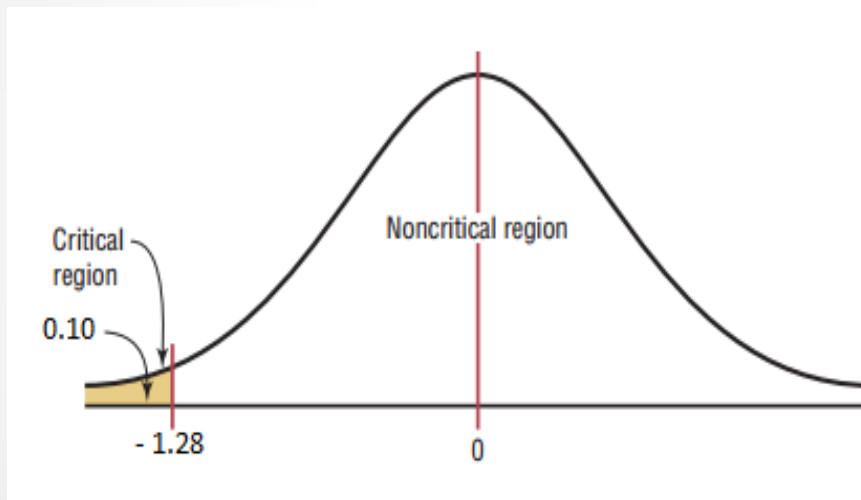


*Finding the Critical Value for α = 0.10 (a left-sided test) [1].*

➢ The critical region, with an area equal to α, is on the left side of the mean.
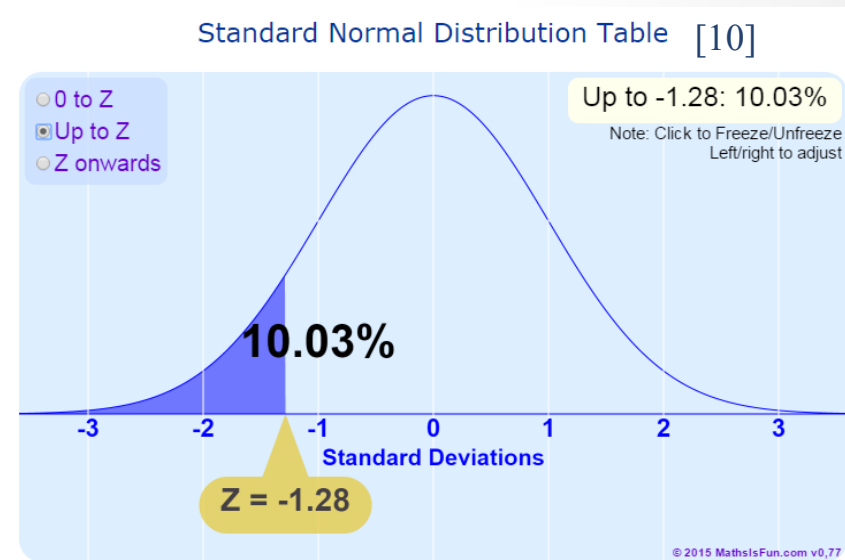➢ To find the critical value, use the z value that corresponds to the area equivalent to α in the standard normal distribution table.
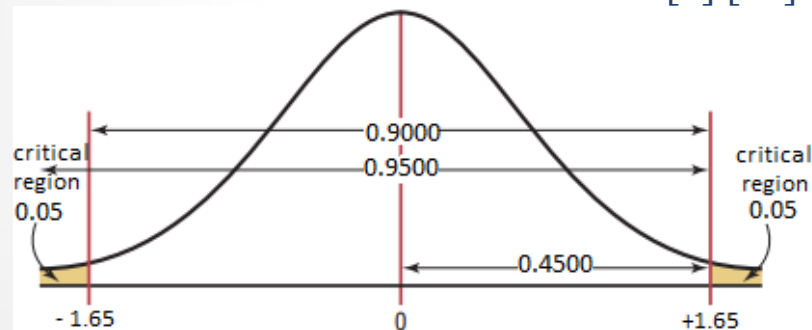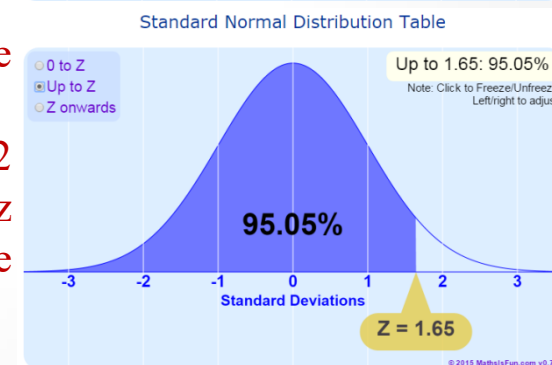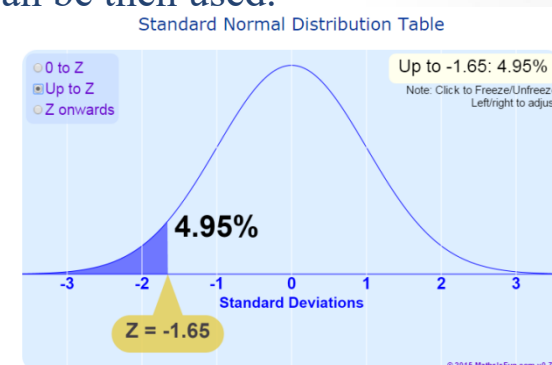
A **_two-tailed test_** indicates that the null hypothesis should be rejected when the test value is in either of the two critical regions [1].

- **A two-tailed test**: $\mu \neq 6{,}463$ and P(type I error) = 0.10 (10%)
→ To find the corresponding z value: define the z value on the left side of the mean, which corresponds to an area of 0.0500, and the z value on the right side of the mean, which corresponds to an area of 0.9500.
→ The z value on the left side of the mean is about halfway between -1.64 and -1.65 as there are two values corresponding to the area closest to 0.0500 (0.0505 and 0.0495). In this case the average value can be used (-1.645), or if needed the z value can be rounded to 2 decimal places, and then -1.65 is used. On the right side the value also falls between 1.64 and 1.65, so 1.645 or 1.65 can be then used.
→ The critical values are -1.65 and +1.65 [1] [10].





Standard Normal Distribution Table

○ 0 to Z
☒ Up to Z
○ Z onwards

Up to -1.65: 4.95%
Note: Click to Freeze/Unfreeze
Left/right to adjust

**4.95%**

Z = -1.65

© 2015 MathsIsFun.com v0.77

> The critical region is divided by 2; one-half of the area is to the right, and one-half is to the left of the mean [1].
> To find the critical value, use the z value that corresponds to α/2 for the left value. It will be negative. For the right value, use the z value that corresponds to the area equivalent to 1 - α/2. It will be positive [1].

Standard Normal Distribution Table

○ 0 to Z
☒ Up to Z
○ Z onwards

Up to 1.65: 95.05%
Note: Click to Freeze/Unfreeze
Left/right to adjust

**95.05%**

Z = 1.65

© 2015 MathsIsFun.com v0.77

**Traditional method of hypothesis testing [1] [5]:**

➢ state the null and alternative hypotheses;

➢ design the study:
- ✓ choosing the statistical test;
- ✓ choosing the level of significance (the α-level);
- ✓ defining the population;
- ✓ selecting the sample;
- ✓ deciding on methods;

➢ collect the data;

➢ evaluate the data:
- ✓ defining the critical value(s);
- ✓ computing the test value.

➢ decide on rejecting or not rejecting the null hypothesis;

➢ summarize the results.

- Common formula for a statistical test [1]:

$$\text{Test value} = \frac{\text{observed value} - \text{expected value}}{\text{standard error}}$$

- Formula for a z-test [1]:

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

where $\overline{X}$ is a sample mean;

μ is a population mean;

σ is a standard deviation;

n is a sample size.

The **z test** is a statistical test for the mean of a population that can be used when $n \geq 30$, or when the population is normally distributed and the standard deviation is known [1].

<u>Example:</u> *It has been reported that the monthly food expenditure for families of two (19-50 years) in the US is on the average higher than $618.60 [4]. A sample of 35 American families has a mean monthly food expenditure of $630. At a α = 0.10 test the claim that the monthly food expenditure for families of two (19-50 years) in the US is on the average higher than $618.60. The standard deviation of the population is $15.*

***Solution:***
- *State the hypotheses and identify the claim: $H_0$: μ = $618.60*
$$H_1: μ > \$618.60$$

- *Find the critical value: since α = 0.10, the critical value is then z = +1.28.*
- *Compute the test value:*

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{630 - 618{,}60}{15/\sqrt{35}} \approx 0{,}129$$

- *Make the decision: since the test value, 0,129, is less than the critical value, 1.28, and subsequently is not in the critical region, the decision would be not reject the null hypothesis.*

- *Summarize the results: there is not enough evidence to support the claim that the monthly food expenditure for families of two (19-50 years) in the US is on the average higher than $618.60.*

**Claim is $H_0$**

*Reject $H_0$*

*Do not reject $H_0$*

There is enough evidence to reject the claim.

There is not enough evidence to reject the claim.

**Claim is $H_1$**

*Reject $H_0$*

*Do not reject $H_0$*

There is enough evidence to support the claim.

There is not enough evidence to support the claim.
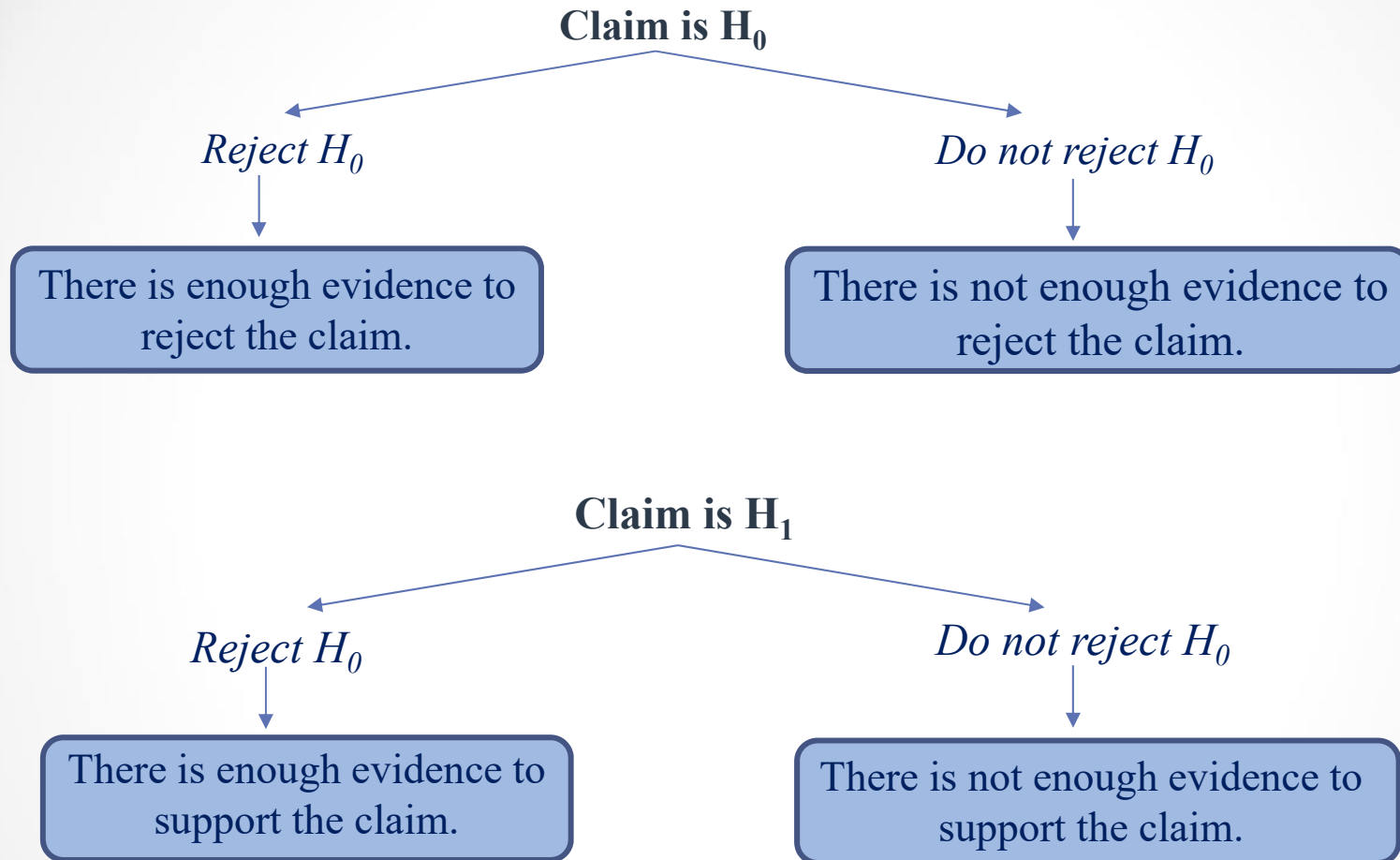
Exercise #1: One claims that the average income in Germany is less than 3200 €. Suppose you conducted a survey and gathered the following data (average income) from a random sample of 30 people (see below). Assuming that $\alpha = 0.10$ and $\sigma = 250$ €, is there enough evidence to support the initial claim?

| | | | | |
|------|------|------|------|------|
| 3340 | 3050 | 3360 | 3100 | 3050 |
| 3750 | 3120 | 3420 | 3130 | 3300 |
| 2890 | 3670 | 3610 | 3060 | 2950 |
| 3520 | 3240 | 2980 | 2900 | 3400 |
| 3180 | 2970 | 2840 | 2750 | 3500 |
| 2910 | 2840 | 3000 | 2800 | 3250 |

HUMBOLDT-UNIVERSITÄT ZU BERLIN

**Resource Economics**

---

The ***P-value (or probability value)*** is the probability of getting a sample statistic (such as the mean) or a more extreme sample statistic in the direction of the alternative hypothesis when the null hypothesis is true [1].

- Traditional method: comparison of a test statistic and a critical value;
- P-value approach: comparison of probabilities, or areas [1].

---

Example: $H_1$: $\mu > 30$, $\overline{X} = 34$, and $P = 0.0256$ → *This means that the probability of getting a sample mean of 34 or greater is 0.0256 if the true population mean is 30. And for the P = 0.0256, the null hypothesis would be rejected at α = 0.05 but not at α = 0.01 as 0.01 ≤ 0.0256 ≤ 0.05.*

*When the hypothesis test is two-tailed, the area in one tail must be doubled.* → *This means that if α = 0.05 and the area in one tail is 0.0256, P = 0.0512. Since 0.0512 > 0.05, the null hypothesis would not be rejected [1].*

---

- When the P < α → the null hypothesis would be rejected;
- If the P > α → the null hypothesis should not be rejected [1].

HUMBOLDT-UNIVERSITÄT ZU BERLIN

**Resource Economics**

HUMBOLDT-UNIVERSITÄT ZU BERLIN

---

<u>Example:</u> *It has been reported that the monthly food expenditure for families of two (19-50 years) in the US is on the average $618.60 [4]. A sample of 35 American families has a mean monthly food expenditure of $630. Use the P-value approach to identify whether there is enough evidence to reject this claim, given that the level of significance equals 0.05 and the standard deviation is $15.*

*Solution:*

- *State the hypotheses and identify the claim: $H_0$: $\mu$ = $618.60 (claim);*
  *$H_1$: $\mu \neq$ $618.60.*

- *Compute the test value:* $z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{630 - 618,60}{15/\sqrt{35}} \approx 0,13$

- *Compute the P-value: using the table of the standard normal distribution we find the corresponding area for z-value that equals to 0.13. It is 0.5517.*
  *1 − 0.5517 = 0.4483*
  *It is a two-tailed test → 2 × 0.4483 = 0.8966*

- *Compare this p-value to the original α: 0.8966 > 0.05 → do not reject the null hypothesis.*
- *Draw a final conclusion: there is not enough evidence to reject the claim that the monthly food expenditure for families of two (19-50 years) in the US is on the average $618.60.*

If P-value $\leq \alpha$ → the decision should be to reject the null hypothesis [1].
If P-value $> \alpha$ → the decision should be to not reject the null hypothesis [1].

The use of P- and $\alpha$ values [1]:
- $\alpha$-value → before the statistical test is conducted;
- P-value → after the sample mean has been found.

**Guidelines for P-values** [1]:

- If $P \leq 0.01$ → the null hypothesis should be rejected (the difference in this case is highly significant);

- If $0.01 < P \leq 0.05$ → the null hypothesis would be rejected as well (the difference is significant);

- If $0.05 < P \leq 0.10$ → the consequences of type I error should be considered before taking the decision on rejecting the null hypothesis;

- If $P > 0.10$ → the null hypothesis is not to be rejected (the difference is not significant).

Exercise #2: Here is some data on the consumption of fertilizers (tonnes of nutrients) per 1000 hectares of agricultural land area in Africa in 2013 [7]. Suppose you want to test the claim that the average consumption of fertilizers is lower than 15 tonnes/ha. For this purpose, you select a sample of 30 African countries (except for North and Central Africa) with the mean equal to 6,3 tonnes/ha. Given $\alpha = 0.05$ and using the P-value method, is there evidence to support the claim?

| Southern Africa | |
|---|---|
| Angola | 0,33 |
| Botswana | 0,85 |
| Comoros | - |
| Lesotho | - |
| Madagascar | 0,20 |
| Malawi | 20,50 |
| Mauritius | 71,60 |
| Mozambique | 0,90 |
| Namibia | 0,06 |
| Seychelles | 35,81 |
| South Africa | 4,30 |
| Swaziland | - |
| Zambia | 5,08 |
| Zimbabwe | 4,82 |

| West Africa | |
|---|---|
| Benin | 1,68 |
| Burkina Faso | 3,55 |
| Cabo Verde | - |
| Côte d'Ivoire | 1,70 |
| Gambia | 0,09 |
| Ghana | 3,94 |
| Guinea | 0,25 |
| Guinea-Bissau | - |
| Liberia | - |
| Mali | 3,08 |
| Mauritania | - |
| Niger | 0,16 |
| Nigeria | 6,66 |
| Senegal | 3,40 |
| Sierra Leone | - |

| North Africa | |
|---|---|
| Algeria | 1,08 |
| Egypt | 373,91 |
| Libya | 0,00 |
| Morocco | 5,94 |
| Sudan | 1,38 |
| Tunisia | 7,84 |

| Central Africa | |
|---|---|
| Cameroon | 2,00 |
| Central African Republic | - |
| Chad | - |
| Congo | 0,05 |
| Democratic Republic of the Congo | 0,20 |
| Equatorial Guinea | - |
| Gabon | 0,15 |
| Sao Tome and Principe | - |

| East Africa | |
|---|---|
| Burundi | 2,04 |
| Djibouti | - |
| Eritrea | 0,07 |
| Ethiopia | 4,71 |
| Kenya | 4,72 |
| Rwanda | 3,64 |
| Somalia | - |
| Sudan | 1,38 |
| Uganda | 0,63 |
| United Republic of Tanzania | 1,08 |

- Many hypothesis testing situations may imply proportions:
  - *Almost 30% of German population lives in rural areas [3].*

- Common formula for a statistical test [1]:

$$\text{Test value} = \frac{\text{observed value} - \text{expected value}}{\text{standard error}}$$

- Formula for the z-test for a proportion [1]:

$$z = \frac{\hat{p} - p}{\sqrt{pq/n}} \quad \leftarrow \quad z = \frac{X - \mu}{\sigma}$$

where $\hat{p} = \dfrac{X}{n}$ - sample proportion

p – population proportion
n – sample size
q – numerical probability of a failure

HUMBOLDT-UNIVERSITÄT ZU BERLIN

HUMBOLDT-UNIVERSITÄT ZU BERLIN

**Resource Economics**

<u>Example:</u> *Suppose a researcher is conducting his study on sustainable lifestyles and claims that 30% of people are concerned about environment but still have some difficulties to translate this into greener everyday life choices [6]. The researcher randomly selects 100 people and finds out that 35 people proved that they were trying to change their life patterns towards more sustainable ones. Is there enough evidence to support the researcher's claim, when α = 0.01?*

***Solution:***

- *State the hypotheses and identify the claim: $H_0$: p = 0.30 (claim) and $H_1$: p ≠ 0.30.*
- *Find the critical values: since α = 0.01 and the test is two-tailed → the critical values are ± 2.58.*
- *Compute the test value:*

  - $\hat{p} = \dfrac{X}{n} = \dfrac{35}{100} = 0.35\ (35\%)$

  - *p = 0.30*

  - *q = 1 – 0.30 = 0.70*

  → $z = \dfrac{\hat{p} - p}{\sqrt{pq/n}} = \dfrac{0.35 - 0.30}{\sqrt{(0.30)(0.70)/100}} \approx 1.09$

- *Take a decision: since the test value is outside the critical region → do not reject the null hypothesis.*
- *Make a conclusion: there is not enough evidence to reject the claim that 30% of people are concerned about environment but still have some difficulties to translate this into greener everyday life choices.*

**4.5 Hypothesis testing: z-test for a proportion**

HUMBOLDT-UNIVERSITÄT ZU BERLIN

**HUMBOLDT-UNIVERSITÄT**
ZU BERLIN

HUMBOLDT-UNIVERSITÄT ZU BERLIN

**Resource Economics**

<u>Example:</u> *The results of a survey conducted in Brazil, Canada, China, France, Germany, India, UK and the USA in 2007 show that at least 50% of consumers were concerned about environmental and social issues, but were not willing to pay for products with environmental and social benefits [6]. Suppose that a sample of 80 people showed that 50 had some concerns about environmental and social issues, but still did not demonstrated any willingness to pay for this. At α = 0.05, is there enough evidence to support the claim that 50% of consumers are concerned about environment, but are not willing to pay for more environmental and socially beneficial products? Use the P-value approach.*

***Solution:***
- *State the hypotheses and identify the claim: $H_0$: p = 0.50 and $H_1$: p > 0.50 (claim).*
- *Compute the test value:*
  - $\hat{p} = \dfrac{X}{n} = \dfrac{50}{80} = 0.625$
  - *p = 0.5*
  - *q = 1 – 0.5 = 0.5* → $z = \dfrac{\hat{p} - p}{\sqrt{pq/n}} = \dfrac{0.625 - 0.5}{\sqrt{(0.5)(0.5)/80}} \approx 2.23$
- *Find the P-value: using the table of the standard normal distribution we find the corresponding area for z-value that equals to 2.23. It is 0.9871.*
  *1 – 0.9871 = 0.0129* → *The P-value is 0.0129.*
- *Compare this p-value to the original α: 0.0129 < 0.05* → *reject the null hypothesis.*
- *Make a conclusion: there is enough evidence to support the claim.*

Exercise #3: According to the OECD Data 68% of households in Greece have access to the Internet [3]. Suppose that a researcher selects a sample of 50 households to check whether the claim is valid and finds that 30 households have access to Internet.

a) At $\alpha = 0.10$, test the claim that 68% of Greek households have access to the Internet.

b) At $\alpha = 0.05$, test the claim that less than 68% of Greek households have access to the Internet. Use the P-value method.

|  | $H_0$ true | $H_0$ false |
|---|---|---|
| **Reject $H_0$** | Type I error $\alpha$ | Correct decision $1 - \beta$ |
| **Do not reject $H_0$** | Correct decision $1 - \alpha$ | Type II error $\beta$ |

- Type I error: when the null hypothesis is rejected.
- Type II error: when the null hypothesis is not rejected, though being false.
        → the probability of committing a type II error is symbolized by $\beta$
- $\beta$ depends on:
    - the value of $\alpha$
    - the standard deviation of population
    - the sample size
    - difference between the hypothesized value of the parameter and the true one
- It is difficult to compute the value of $\beta$ → However, the researcher can try to minimize the size of $\beta$ or to maximize the size of $1-\beta$ (1 minus the probability of committing a type II error) [1].

The ***power of a statistical test*** evaluates the sensitivity of the test to identify an actual difference in parameters if it actually exists [1].

- The power of a test is a probability and, hence, can be ranked from 0 to 1 → the higher is the power of a statistical test, the more sensitive to identifying the difference between parameters the test is [1].

- Since the value of β is difficult to compute → the power of a test is also difficult to compute as it depends on the probability of committing a type II error [1].

- However, the researcher can control several factors that could help to increase the power of a test [1]:
  - ➢ the value of the level of significance can be increased
  - ➢ selecting a larger sample size → results in a smaller standard error of the mean that will also reduce the value of β.

> If the null hypothesis fails to be rejected though being false, this does not mean that there is not enough evidence to support alternative hypotheses. It just can be that the statistical test has a low power to identify the difference between the tested and real parameters [1].

- What is the difference between the null and alternative hypothesis? Please give an example.
- What is a statistical test? Which methods/approaches can be used for a statistical test?
- Which types of error may occur while running a statistical test? How do they differ and how are they related?
- Which types of a statistical test do you know? How do they differ in their use?
- What is meant by the critical and noncritical region?
- What are the main steps for hypothesis testing?
- What is meant by the z-test for a mean?
- Is it important to distinguish which hypothesis has been claimed: the null or alternative hypothesis? Why?
- What does the P-value method show? How can one define the P-value?
- How does the level of significance influence the result of a statistical test?
- What is the general formula for a statistical test? Is the formula for the z-test for a proportion based on the general formula?
- Is it possible to use the P-value method while running the z-test for a proportion?
- What are the necessary requirements when testing hypotheses by using proportions?
- Why is it difficult to compute the value of the probability of committing a type II error?
- Can one be absolutely sure about the results obtained and the decisions taken after having conducted a statistical test?

**Exercise #1:**

1) $H_0 : \mu = 3200\ €$

   $H_1 : \mu < 3200\ €$

2) the critical value is $-1,28$ since the test is a left-tailed test and $\alpha = 0.10$

3) Sample mean $\approx 3163$

4) $z \approx -0,814 \rightarrow$ do not reject $H_0 \rightarrow$ there is not enough evidence to support the claim

**Exercise #2:**

One can only do these kinds of test if one deals with a normal distribution and African countries are not a homogeneous population.

## Exercise #3:

a) 1) $H_0 : \mu = 0.68$ (claim)

   $H_1 : \mu \neq 0.68$

   2) Critical values: since $\alpha = 0.10$ and the test is two-tailed → the critical values are $\pm 1.65$.

   3) Test value:

   - Sample proportion: 0.6
   - $p = 0.68$
   - $q = 0.32$
   - → $z \approx -0{,}02$

   4) Since the test value is outside the critical region → do not reject the null hypothesis.

   5) There is not enough evidence to reject the claim.

b) 1) $H_0 : \mu = 0.68$

   $H_1 : \mu < 0.68$ (claim)

   2) Test value:

   - Sample proportion: 0.6
   - $p = 0.68$
   - $q = 0.32$
   - → $z \approx -0{,}02$

   3) P-value: $1 - 0.4920 = 0.508$

   4) $0.508 > 0.05$ → do not reject the null hypothesis

   5) There is not enough evidence to support the claim.

[1] Allan G. Bluman, 2009. *Elementary Statistics: A Brief Version*, 7th Edition, New York: McGraw-Hill.

[2] Jarkko Isolato, Basics of Statistics. Retrieved from http://www.mv.helsinki.fi/home/jmisotal/BoS.pdf

[3] OECD Data. Retrieved from: https://data.oecd.org/

[4] Official USDA Food Plans: Cost of Food at Home at Four Levels, U.S. Average, July 2014. Center for Nutrition Policy and Promotion. US Department of Agriculture. Retrieved from: http://www.cnpp.usda.gov/sites/default/files/usda_food_plans_cost_of_food/CostofFoodJul2014.pdf

[5] Online Statistics Education: A Multimedia Course of Study. Project Leader: David M. Lane, Rice University. Retrieved from http://onlinestatbook.com/

[6] Partidario M.R., Vicente G., Belchior C. *Can New Perspectives on Sustainability Drive Lifestyles?* Sustainability 2: 2849-2872 (2010).

[7] United Nations Statistics Data. Retrieved from: http://unstats.un.org/unsd/environment/qindicators.htm

[8] https://www.mathsisfun.com/data/standard-normal-distribution-table.html