Null hypothesis (H_0) – nulová hypotéza

Definition:
The hypothesis tentatively assumed true in the hypothesis-testing procedure.

Note: The statement do not reject H_0 is often used to avoid accepting that H_0 is true thus this eliminates the possibility of making a Type II error.

Alternate hypothesis (H_1/H_a) – alternatívna hypotéza

Definition:
The hypothesis conducted to be true if the null hypothesis is rejected.

Note:
In general, hypothesis testing can be divided into two main parts:

- *testing the validity of a claim* – the claims is either correct or not (H_0 can either be accepted (i.e. not rejected) or rejected, action is only taken if H_0 is rejected
- *testing in decision making situations* – (H_0 can either be accepted (i.e. not rejected) or rejected, action must be taken if H_0 cannot be rejected or H_0 can be rejected. In this case, the decision maker must choose between two courses of action – one associated with the null hypothesis and another one associated with the alternative hypothesis.

Type I error – chyba prvého druhu

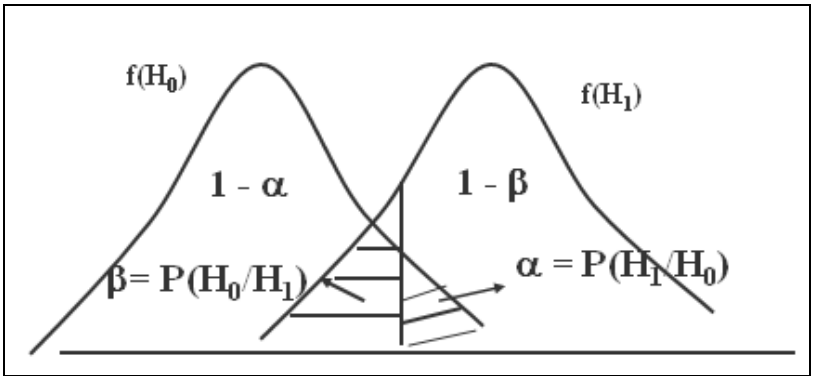
Definition:
The error of rejecting H_0 when it is true.

Note: Alpha is the probability of making a Type I error. Expression (1-alpha) is the probability of accepting the correct hypothesis.

Type II error – chyba druhého druhu

Definition:
The error of accepting H_0 when it is false.

Note: Beta is the probability of making a Type II error. Expression (1-beta) is the power of the test.



Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra
Reference: JAISINGH, L.: Statistics for the Utterly Confused

Power – sila testu

Definition:
The probability of correctly rejecting H_0 when it is false.

Critical value – kritická hodnota (tabuľková hodnota)

Definition:
A value that is compared with the test statistic to determine whether or not H_0 should be rejected.

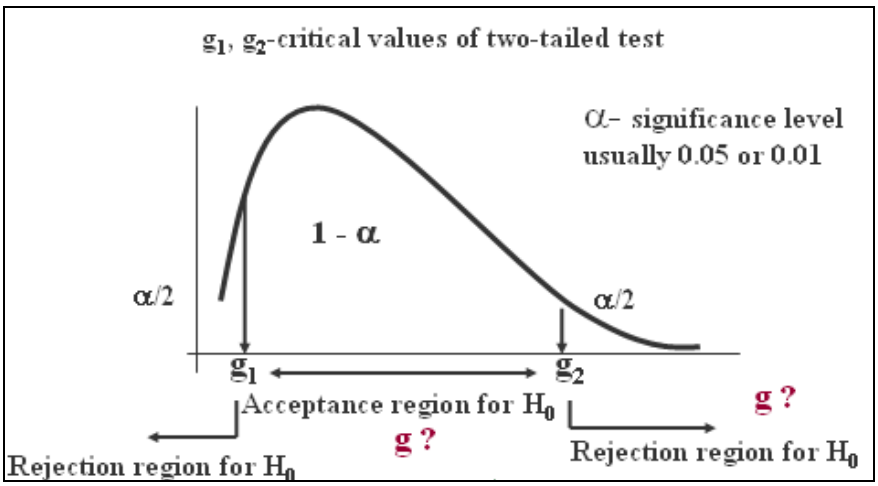
Level of significance – hladina významnosti

Definition:
The maximum probability of a Type I error.

Test statistic – testovacie kritérium (testovacia štatistika)

Definition:
The value of the formula of the test statistic that is compared with the critical value to determine whether or not H_0 should be rejected.

Rejection region – obor zamietnutia
Acceptance region – obor prijatia (nezamietnutia)



One tailed test – jednostranný test (**upper-tailed test; right-tailed test** (pravostranný test) or **lower-tailed test; left-tailed test** (ľavostranný test))

Definition:
A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in one tail of the sampling distribution.

Two-tailed test – obojstranný test

Definition:
A hypothesis test in which rejection of the null hypothesis occurs for values of the test statistic in either tail of the sampling distribution.

Hypothesis test about a population mean – test zhody strednej hodnoty so známou konštantou

$$u = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}}$$ (normal distribution) $$t = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}}$$ (Student t distribution)

μ_0 denotes the specific numerical value being considered in the null and alternative hypotheses.

There can be three possibilities:

- population variance is known and the random variable can be approximated by normal distribution (theoretical point),
- population variance is not known (we use the point estimation of population variance=sample variance) and sample size is greater than 30; then the random variable can be approximated by normal distribution N (0,1),
- population variance is not known (we use the point estimation of population variance=sample variance) and sample size is less than 30; then the random variable can be approximated by Student t distribution TINV(alpha; (n-1)).

Solution to hypothesis testing:

- *if critical value is greater than test statistic, then the null hypothesis cannot be rejected*
- *if critical value is smaller than test statistic, then the null hypothesis is rejected and the alternative hypothesis is accepted*

Independent samples – nezávislé súbory

Definition:
Samples selected from two (or more) populations where the elements making up one sample are chosen independently of the elements making up the other sample(s).

Example: average salary within the Bratislava region and Nitra region

---

Matched samples – závislé súbory

Definition:
Samples where each data value in one sample is matched with a corresponding data value in the other sample.

Example: impact of advertisement on the height of company's turnover

---

Hypothesis test about the difference between means of two populations (independent samples) – test zhody dvoch stredných hodnôt nezávislých súborov

There can be four possibilities:

- population variance of all data sets is known and the random variable can be approximated by normal distribution (theoretical point),
- population variance is not known (we use the point estimation of population variance=sample variance) and sample size of all data sets is greater than 30; then the random variable can be approximated by normal distribution N (0,1),
- population variance is not known (we use the point estimation of population variance=sample variance) and sample size of at least one of the data sets is less than 30; then the random variable can be approximated by Student t distribution TINV(alpha; (n-1)) – on condition that population variances are equal
- population variance is not known (we use the point estimation of population variance=sample variance) and sample size of at least one of the data sets is less than 30; then the random variable can be approximated by Student t distribution TINV(alpha; (n-1)) – on condition that population variances are no equal (Behrens-Fisher test)

$$u = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{n_2 s_{11}^2 + n_1 s_{12}^2}{n_1.n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_2 s_{11}^2 + n_1 s_{12}^2}{n_1.n_2}}} \qquad \text{(normal distribution)}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_{11}^2 + (n_2-1)s_{12}^2}{n_1+n_2-2}}} \cdot \sqrt{\frac{n_1.n_2}{n_1+n_2}} \qquad \text{(Student t distribution)}$$

Hypothesis test about the difference between means of two populations (matched samples) – test zhody dvoch stredných hodnôt závislých súborov

$$t = \frac{\bar{d}}{\sqrt{\frac{\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n \cdot (n-1)}}} \qquad \text{(only Student t distribution)}$$

Tools/Data Analysis:

t-test: Paired two-sample for means – test zhody dvoch stredných hodnôt závislých súborov
t-test: Two-sample assuming equal variances – Studentov t test pri homogénnej variancii (test zhody dvoch stredných hodnôt nezávislých súborov pri homogénnej variancii)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1-1)s_{11}^2 + (n_2-1)s_{12}^2}{n_1+n_2-2}}} \cdot \sqrt{\frac{n_1.n_2}{n_1+n_2}}$$

Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra
Reference: JAISINGH, L.: Statistics for the Utterly Confused

t-test: Two-sample assuming unequal variances – Studentov t test pri nehomogénnej variancii (Behrens-Fisher test)
z-test: Two-sample for means – test zhody dvoch stredných hodnôt, ak poznáme rozptyly základných súborov