TESTOVANIE HYPOTÉZ

Úvod do testovania hypotéz

Hypotéza je výrok, alebo tvrdenie o stave sveta (o skutočnej hodnote neznámeho parametra populácie - základného súboru), napr:

- ✓ Obvinený je nevinný
- ✓ $\mu = 100$

Každá hypotéza implikuje svoj protiklad alebo alternatívu, napr:

- ✓ Obvinený je vinný
- ✓ $\mu \neq 100$

Hypotéza je pravdivá alebo nepravdivá a môžeme ju zamietnuť alebo nezamietnuť, alebo ju môžeme zamietnuť na základe určitých informácií, napr. :

- ➢ svedectvá alebo dôkazy na súdnom pojednávaní
- ➢ výberové dáta

Rozhodovanie

Jednu z hypotéz považujeme za pravdivú, pokiaľ sa nerozhodneme zamietnuť ju ako nesprávnu, napr.:

✓ Vina je dokázaná “mimo všetkých rozumných pochybností”

✓ Alternatíva je vysoko nepravdepodobná

Rozhodnutie nezamietnuť alebo zamietnuť hypotézu môže byť:

Správne

- Pravdivá hypotéza môže byť nezamietnutá
  - » Nevinný obžalovaný môže byť oslobodený
- Nesprávna hypotéza môže byť zamietnutá
  - » Vinný obžalovaný môže byť odsúdený

Nesprávne

- Pravdivá hypotéza môže byť zamietnutá
  - » Nevinný obžalovaný môže byť odsúdený
- Nesprávna hypotéza nemusí byť zamietnutá
  - » Vinný obžalovaný môže byť oslobodený

Štatistické testovanie hypotéz

Nulová hypotéza (označujeme ju H_0) je tvrdenie o jednom alebo o viacerých parametroch populácie. Je to tvrdenie, ktoré považujeme za pravdivé, pokiaľ nemáme dostatočné štatistické dôvody pre opačný záver, napr.

H_0: $\mu = 100$

Alternatívna hypotéza (označujeme ju H_1) je tvrdenie o všetkých situáciách *nepokrytých* nulovou hypotézou, napr.:

H_1: $\mu \neq 100$

H_0 a H_1 sú:

- Vzájomne disjunktné (vylučujú sa)
  - – Iba jedna môže byť pravdivá.
- Vyčerpávajúce
  - – Spolu pokrývajú *všetky* možnosti tak, že jedna alebo druhá *musí* byť pravdivá.

Nulová hypotéza H_0

Nulová hypotéza:

- Často predstavuje situáciu status quo alebo existujúce presvedčenie.
- Nezamietne sa alebo sa považuje za pravdivú, pokiaľ test nevedie k jej zamietnutiu v prospech alternatívnej hypotézy.
- Nezamietne sa ako pravdivá alebo sa zamietne ako nepravdivá na základe úvah o rozdelení ***testovacej štatistiky*** v prípade platnosti H_0 a porovnaní jej následne vypočítanej hodnoty s kritickými hodnotami tohto rozdelenia.

Základné pojmy testovania hypotéz

- **Testovacia štatistika** je (vopred) vhodne stanovená funkcia výberových dát. Následne vypočítaná hodnota tejto štatistiky sa používa na rozhodnutie, či máme zamietnuť alebo nezamietnuť nulovú hypotézu.

- **Rozhodovacie pravidlo** testu štatistickej hypotézy je pravidlo, ktoré špecifikuje podmienky, pri splnení ktorých možno nulovú hypotézu zamietnuť.

- **Rozhodovanie:**

Sú dva možné stavy:

- ✓ H_0 je *pravdivá*
- ✓ H_0 je *nepravdivá*

Existujú dve možné rozhodnutia:

- ✓ *Nezamietnuť* H_0
- ✓ *Zamietnuť* H_0

Prijatie rozhodnutia

Rozhodnutie môže byť správne dvomi spôsobmi:

- Nezamietnutie pravdivej nulovej hypotézy H_0
- Zamietnutie nepravdivej nulovej hypotézy H_0

Rozhodnutie môže byť nesprávne tiež dvomi spôsobmi:

- **Chyba typu I**: Zamietnutie pravdivej hypotézy H_0

  Pravdepodobnosť chyby typu I označíme α.
- **Chyba typu II**: Nezamietnutie nepravdivej hypotézy H_0

  Pravdepodobnosť chyby typu II označíme β.

Chyby pri testovaní hypotéz

Rozhodnutie môže byť chybné dvomi spôsobmi:

- Chyba typu I: Zamietnutie pravdivej hypotézy H_0
  - Vopred zvolenú (nízku) hornú hranicu pre akceptovateľnú hodnotu pravdepodobnosti chyby typu I označíme α.
  - α nazývame hladina významnosti testu
- Chyba typu II: Nezamietnutie nepravdivej hypotézy H_0
  - Pravdepodobnosť chyby typu II označíme β.
  - $1-\beta$ sa nazýva sila testu.

α a β sú podmienené pravdepodobnosti:

$$\alpha = P(\text{zamietnutie } H_0 \mid H_0 \text{ je pravdivá})$$

$$\beta = P(\text{nezamietnutie } H_0 \mid H_0 \text{ je nepravdivá})$$

Chyby typu I a II

Nasledovná tabuľka ilustruje možné výsledky testovania štatistickej hypotézy

	Skutočný stav	
Rozhodnutie	H_0 pravdivá	H_0 nepravdivá
Nezamietnutie H_0	Správne	Chyba typu II (β)
Zamietnutie H_0	Chyba typu I (α)	Správne

p-hodnota (p-value)

p-value je pravdepodobnosť (apriorná pred realizáciou pokusu, z ktorého získame dáta), že získame hodnotu testovacej štatistiky, ktorá je väčšia alebo rovná ako skutočne získaná hodnota za predpokladu, že je nulová hypotéza pravdivá.

p-value je najnižšia hladina významnosti, pri ktorej možno zamietnuť nulovú hypotézu.

Keď je p-value je menšie alebo rovné ako α, zamietneme H_0 na hladine významnosti α

Sila testu

Sila testu štatistickej hypotézy je pravdepodobnosť zamietnutia nulovej hypotézy, keď je táto hypotéza nepravdivá:

$$\text{sila testu} = 1 - \beta$$

Faktory ovplyvňujúce funkciu sily testu:

- Sila testu závisí od vzdialenosti medzi hodnotou parametra pri platnosti nulovej hypotézy a *skutočnou hodnotou* tohto parametra: *čím väčšia je táto vzdialenosť, tým väčšia je sila testu*.
- Sila testu závisí od štandardnej odchýlky populácie: *čím menšia je štandardná odchýlka populácie, tým väčšia je sila testu.*
- Sila testu závisí od použitého výberu: *čím väčší je výber, tým väčšia je sila testu*.
- Sila testu závisí od hladiny významnosti testu: *čím nižšia je hladina významnosti* α, *tým nižšia je sila testu.*

Jednostranný a dvojstranný test

Ak je treba prijať rozhodnutie, či hodnota parametra je väčšia ako nejaká hodnota *a*, potom alternatívna hypotéza je, že parameter je väčší ako *a,* ide o *pravostranný* test

H_0: μ ≤ 50 H_1: μ > 50

Ak je treba prijať rozhodnutie, či hodnota parametra je menšia ako nejaká hodnota *a*, potom alternatívna hypotéza je, že parameter je menší ako *a,* ide o *ľavostranný* test

H_0: μ ≥ 50 H_1: μ < 50

Ak je treba prijať rozhodnutie, či hodnota parametra je menšia alebo väčšia ako nejaká hodnota *a*, potom alternatívna hypotéza je, že parameter nie je rovný *a* a ide o *dvojstranný (two – tailed)* test

H_0: μ = 50 H_1: μ ≠ 50

Oblasť zamietnutia

Oblasť zamietnutia testu štatistickej hypotézy je rozsah možných hodnôt testovacej štatistiky, ktorý nás privedie k zamietnutiu nulovej hypotézy H_0 v prípade, ak následne vypočítaná hodnota testovacej štatistiky padne do tejto oblasti. Oblasť zamietnutia sa tiež nazýva kritickou oblasťou, a je definovaná kritickými bodmi. Oblasť zamietnutia je definovaná tak, aby predtým, než uskutočníme výber, naša testovacia štatistika s pravdepodobnosťou α padla do oblasti zamietnutia, ak je nulová hypotéza pravdivá.

Oblasť nezamietnutia

Oblasť nezamietnutia je rozsah hodnôt (tiež určený kritickými bodmi), ktorý vedie k nezamietnutiu nulovej hypotézy, ak testovacia štatistika padne do tejto oblasti. Oblasť nezamietnutia je navrhnutá tak, aby (predtým, ako sa uskutoční výber) bola pravdepodobnosť, že hodnota testovacej štatistiky padne do oblasti nezamietnutia rovná 1 - α (pri platnosti nulovej hypotézy H_0).

V *obojstrannom teste* pozostáva oblasť zamietnutia z hodnôt na obidvoch okrajoch výberového rozdelenia.

p-value a testovanie hypotéz

Tzv. hodnota *p-value* sa počíta pomocou pôvodnej distribučnej funkcie F(x) testovacej štatistiky (za platnosti H_0), do ktorej dosadíme hodnotu testovacej štatistiky (TS).

Toto číslo udáva pravdepodobnosť, s akou sa mohli hodnoty testovacej štatistiky ocitnúť v intervale, ktorý je sprava ohraničený následne zistenou hodnotou testovacej štatistiky a je rovné tzv. p-hodnote v prípade ľavostranného testu (p-value = F(TS) = P(X ≤ TS)).

V prípade pravostranného testu je p-hodnota rovná rozdielu 1 a vyššie uvedenej hodnoty pôvodnej distribučnej funkcie testovacej štatistiky v jej následne zistenej hodnote. Je to pravdepodobnosť, s akou sa mohli hodnoty testovacej štatistiky ocitnúť v intervale, ktorý je zľava ohraničený jej následne zistenou hodnotou (p-value = 1 - F(TS) = P(X > TS)).

Pri obojstrannom teste je p-hodnota 2-násobkom minima p-hodnôt pre ľavostranný a pravostranný test (p-value = 2*min(F(TS), 1 - F(TS))).

Pri danej hladine významnosti α:

zamietame nulovú hypotézu vtedy a len vtedy, ak $\alpha \geq$ p-value

Zobrazenie testovania hypotéz

Spoločnosť, ktorá doručuje balíky v oblasti veľkého mesta tvrdí, že balík od vašich dverí doručí na miesto určenia priemerne za 28 minút. Overte, či toto tvrdenie môžeme považovať za pravdivé.

Nulová a alternatívna hypotéza:

$H_0: \mu = 28$ $H_1: \mu \neq 28$

Výberové štatistiky:

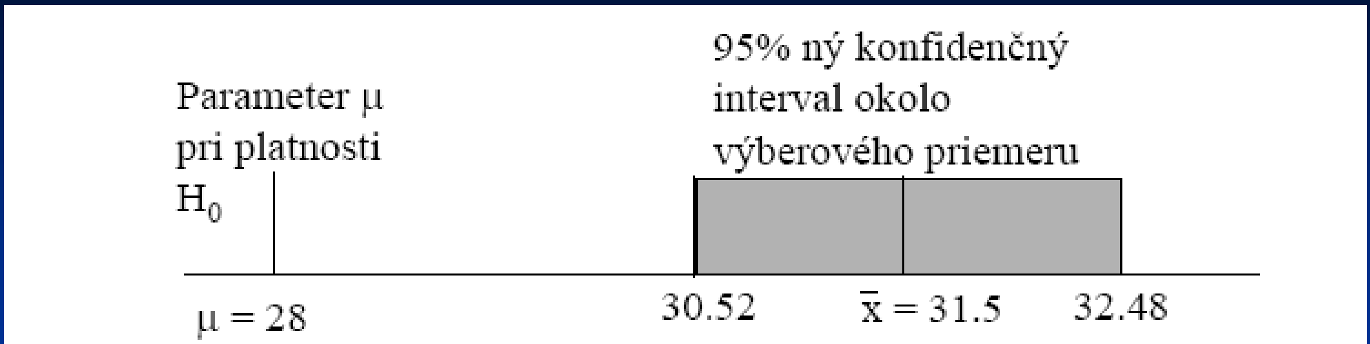
$n = 100$

$\bar{x} = 31.5$

$s = 5$

95% -ný konfidenčný interval pre priemerný čas doručenia:

$[30.52, 32.48]$



Zdá sa rozumné zamietnuť H_0: $\mu = 28$, pretože hodnota 28 leží mimo 95%-ného konfidenčného intervalu. Ak sme si na 95% istí, že stredná hodnota doručenia je medzi 30.52 a 32.58 minútami, je veľmi nepravdepodobné, že stredná hodnota doručenia balíka je 28 minút.

Stredná hodnota doručenia balíka *môže* byť 28 (nulová hypotéza môže byť pravdivá), ale potom pozorovaný výberový priemer 31.5 by bol *veľmi nepravdepodobným* javom. Existuje len malá šanca ($\alpha = 0.05$), že by sme mohli zamietnuť pravdivú nulovú hypotézu.

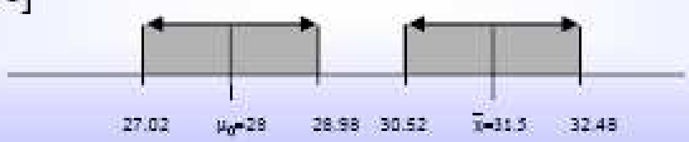
α predstavuje hladinu významnosti (level of significance) testu.

Oblasť nezamietnutia

Ak pozorovaný výberový priemer padne do *oblasti nezamietnutia*, potom nezamietneme H_0, pretože je pravdivá. Okolo predpokladanej populačnej strednej hodnoty konštruujeme 95%-ný toleračný interval a porovnáme ho s 95%-ným konfidenčným intervalom okolo pozorovaného výberového priemeru:

$$\mu_0 \pm z_{.025}\frac{s}{\sqrt{n}} = 28 \pm 1.96\frac{5}{\sqrt{100}}$$
$$= 28 \pm .98 = [27.02, 28.98]$$

$$\bar{x} \pm z_{.025}\frac{s}{\sqrt{n}} = 31.5 \pm 1.96\frac{5}{\sqrt{100}}$$
$$= 31.5 \pm .98 = [30.52, 32.48]$$

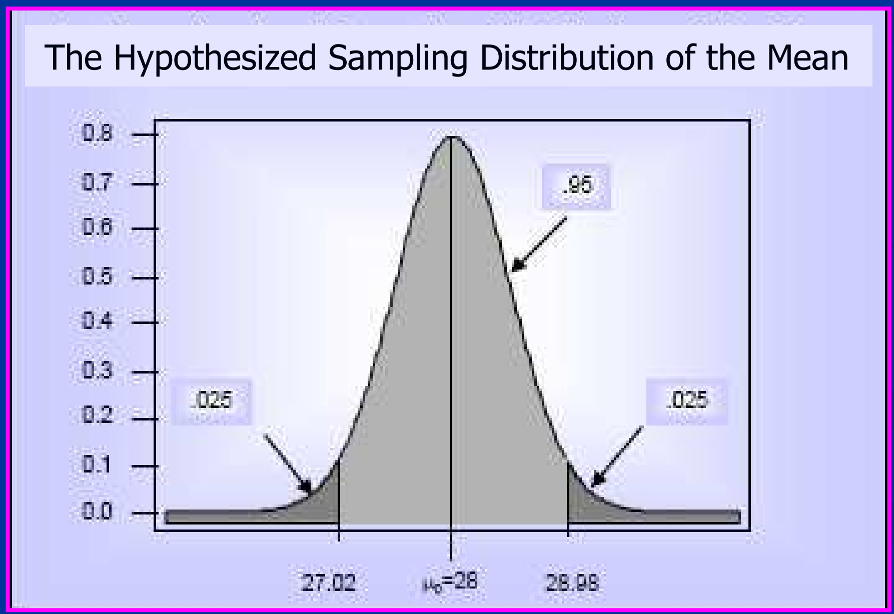


Oblasť nezamietnutia a konfidenčný interval majú tú istú šírku, ale ich stredy sú v odlišných bodoch. V tomto prípade oblasť nezamietnutia nezahrňuje pozorovaný výberový priemer a konfidenčný interval neobsahuje populačnú strednú hodnotu.

Zobrazenie oblasti nezamietnutia

Ak by bola nulová hypotéza pravdivá, potom:

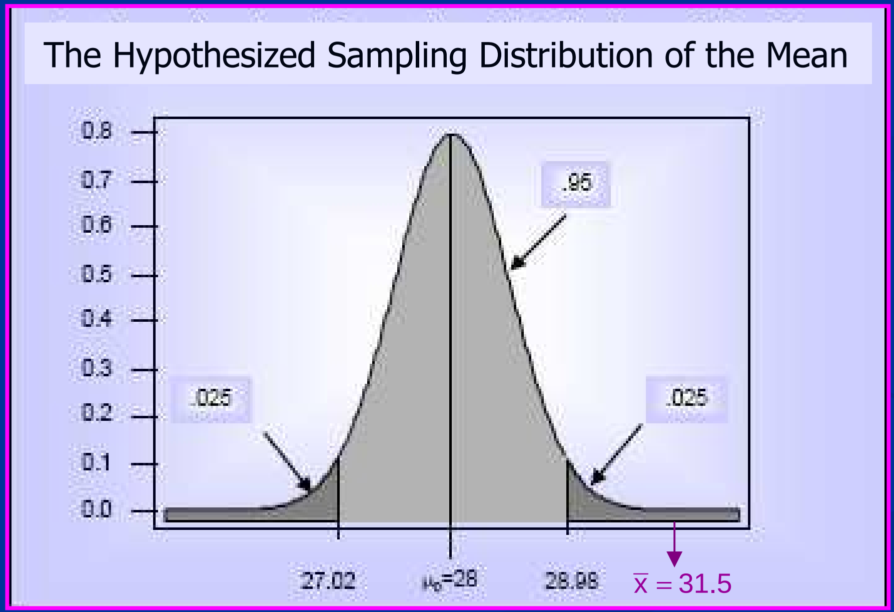
Nájdeme oblasť 95%-ného výskytu hodnôt testovacej štatistiky medzi *kritickými bodmi* 27.02 a 28.98. Ostávajúcu oblasť rozdelíme na 2.5% pod 27.02 a 2.5% nad 28.98 (dvojstranný *test*). 95%-ný interval okolo predpokladanej strednej hodnoty definuje oblasť nezamietnutia (so zvyšnými 5%-ami v dvoch oblastiach).



Rozhodovacie pravidlo

Konštruujeme (1-α) oblasť nezamietnutia okolo hypotetizovanej strednej hodnoty populácie:

- ❑ H_0 nezamietneme, ak výberový priemer padne do oblasti nezamietnutia (medzi kritické body).
- ❑ H_0 zamietneme, ak výberový priemer padne mimo oblasť nezamietnutia.



Oblasť zamietnutia	Oblasť nezamietnutia	Oblasť zamietnutia

a) Testovanie strednej hodnoty populácie

Prípady, keď testovacou štatistikou je *Z*:

- ✓ σ je známe a rozdelenie populácie je normálne;
- ✓ σ je známe a veľkosť výberu je aspoň 30 (populácia nemusí mať normálne rozdelenie).

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

Prípady, keď testovacou štatistikou je *t*:

σ je neznáme, ale výberová štandardná odchýlka je známa a populácia je normálne rozdelená.

$$t = \frac{\bar{x} - \mu}{\left(\frac{\hat{s}}{\sqrt{n}}\right)}$$

b) Testovanie rozptylu populácie

Pri testovaní hypotézy o rozptyle populácie je testovacia štatistika:

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

kde σ_0^2 je zadaný populačný rozptyl pri nulovej hypotéze. Počet stupňov voľnosti pre chi-kvadrát rozdelenie je n - 1.

<u>Poznámka</u>: Pretože chi-kvadrát tabuľky poskytujú iba kritické hodnoty, nemožno ich použiť na výpočet presných p-hodnôt. Ako v prípade t-tabuliek, možno usudzovať iba o rozsahu možných hodnôt. Štatistické programy však umožňujú aj výpočet týchto p-hodnôt.

Príklad 1

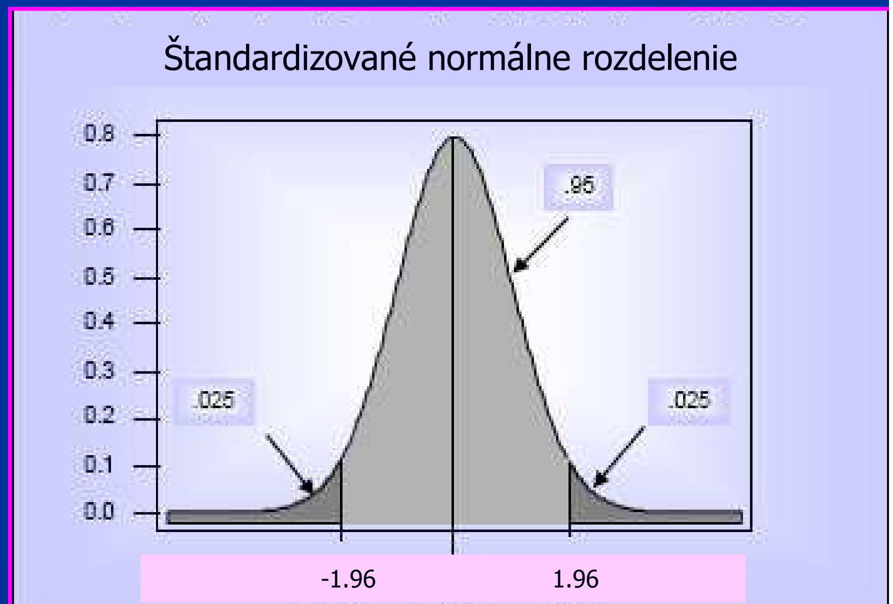
Letecká spoločnosť chce stanoviť batožinovú kapacitu v lietadle. Analytik predpokladá, že priemerná váha batožiny na osobu je $\mu_0 = 12$ libier. Aby overil svoj predpoklad, u 144 náhodne vybraných cestujúcich presne zmeral váhu ich batožiny a zistil, že priemerná váha bola 14.6 libier a štandardná odchýlka 7.8. Na základe výsledkov tohto prieskumu otestujte na hladine významnosti $\alpha = 0.05$, či je predpoklad analytika správny.

H_0: $\mu = 12$ oproti H_1: $\mu \neq 12$

Pre $\alpha = 0.05$ sú kritické hodnoty: ± 1.96

H_0 nezamietame, ak $-1.96 \leq z \leq 1.96$

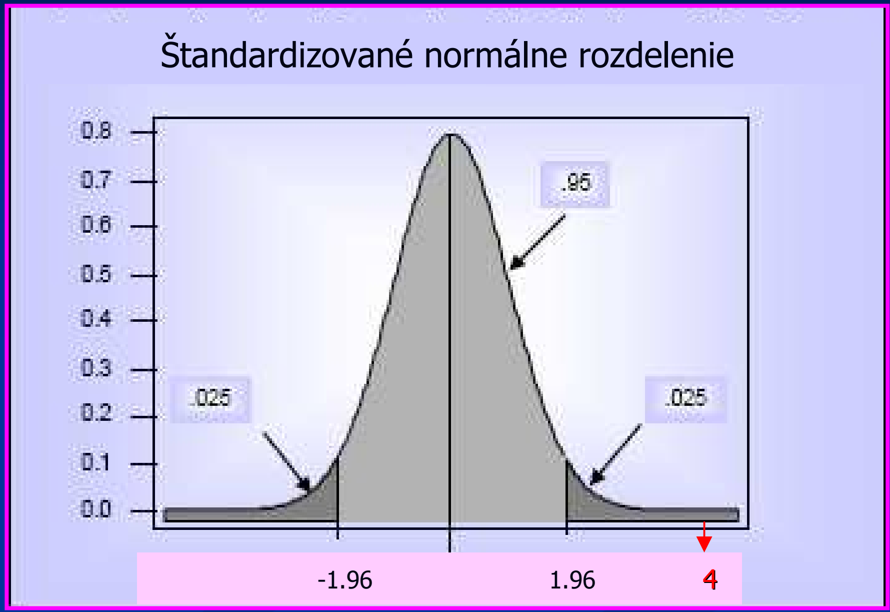
H_0 zamietame, ak $z < -1.96$ alebo $z > 1.96$.



Oblasť zamietnutia	Oblasť nezamietnutia	Oblasť zamietnutia

Riešenie príkladu 1

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{14.6 - 12}{\left(\frac{7.8}{\sqrt{144}}\right)} = \frac{2.6}{0.65} = 4$$



Oblasť zamietnutia	Oblasť nezamietnutia	Oblasť zamietnutia

Pretože testovacia štatistika padla do hornej oblasti zamietnutia, H_0 sa zamieta a uzatvárame, že priemerná váha batožiny je viac ako 12 libier.

Príklad 2

Poisťovacia spoločnosť tvrdí, že za posledných niekoľko rokov bola priemerná hodnota poistenia 2 000 dolárov. U 100 náhodne vybraných klientov zistili, že priemerné poistenie bolo 2 700 a štandardná odchýlka 947. Otestujte na hladine významnosti $\alpha = 0.01$, či je tvrdenie poisťovacej spoločnosti pravdivé.

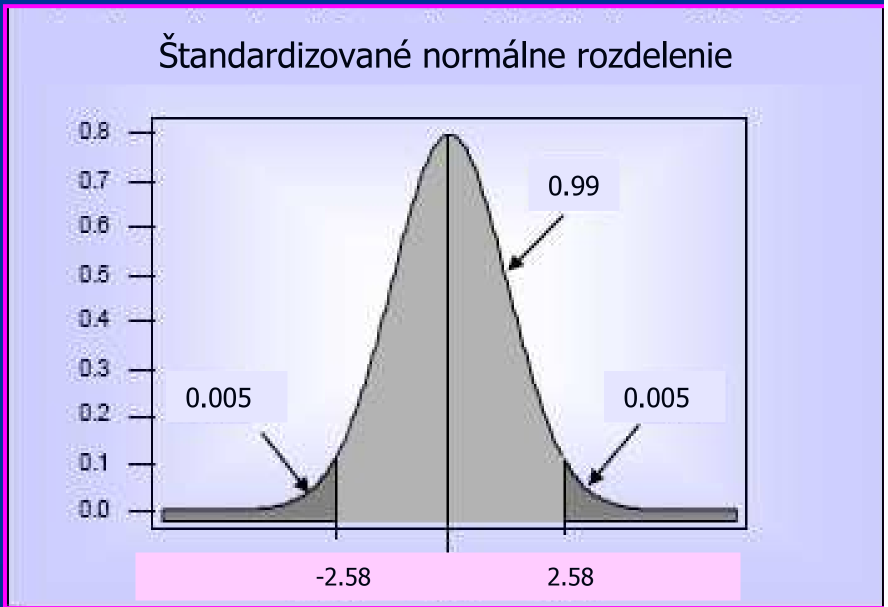
H_0: $\mu = 2000$ oproti H_1: $\mu \neq 2000$

Pre $\alpha = 0.01$ sú kritické hodnoty: ± 2.58

H_0 nezamietame, ak $-2.58 \leq z \leq 2.58$

H_0 zamietame, ak $z < -2.58$ alebo $z > 2.58$.

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{2700 - 2000}{\left(\frac{947}{\sqrt{100}}\right)} = \frac{700}{94.7} = 7.39$$



Oblasť zamietnutia	Oblasť nezamietnutia	Oblasť zamietnutia

Pretože testovacia štatistika padla do hornej oblasti zamietnutia, H_0 zamietame a môžeme uzavrieť, že priemerná hodnota poistenia bola vyššia ako 2000 dolárov.

Príklad 3

Priemerný čas, za ktorý počítač vykoná určitú prácu je 3.24 sekúnd. Testuje sa štatistická hypotéza, že priemerný čas vykonania práce pomocou nového algoritmu je rovnaký, proti alternatíve, že priemerný čas nie je rovnaký. Pri 200 opakovaniach sa zistil priemerný čas 3.48 a štandardná odchýlka 2.8. Test sa vykonáva na hladine významnosti $\alpha = 0.05$.

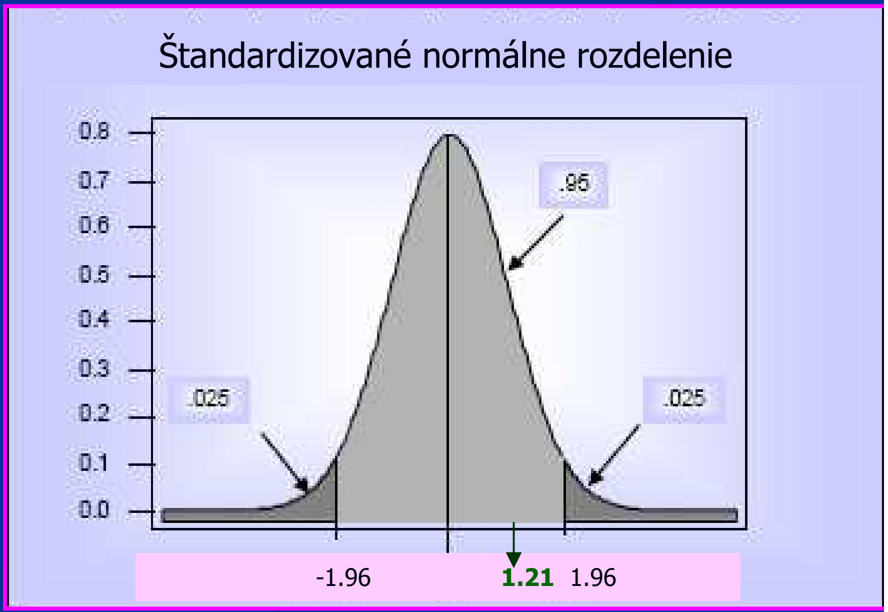
H_0: $\mu = 3.24$ oproti H_1: $\mu \neq 3.24$

Pre $\alpha = 0.05$ sú kritické hodnoty: ± 1.96

H_0 nezamietame, ak $-1.96 \leq z \leq 1.96$

H_0 zamietame, ak $z < -1.96$ alebo $z > 1.96$.

$$Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{3.48 - 3.24}{\left(\frac{2.8}{\sqrt{200}}\right)} = \frac{0.24}{0.20} = 1.21$$



Oblasť zamietnutia	Oblasť nezamietnutia	Oblasť zamietnutia

Pretože testovacia štatistika padla do hornej oblasti nezamietnutia, H_0 nezamietame a môžeme uzavrieť, že priemerný čas na vykonanie práce sa nezmenil – zostal 3.24 sekúnd.

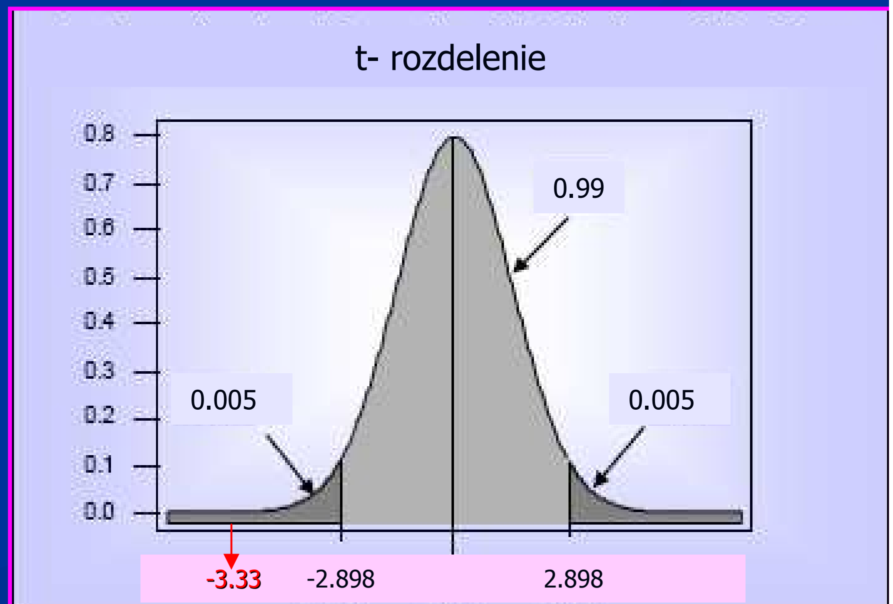
Príklad 4

Podľa Japonskej národnej agentúry pôdy, priemerná cena pôdy v centre Tokia prudko stúpla o 49% za prvých 6 mesiacov 1995. Spoločnosť pre medzinárodné investície do nehnuteľností chce testovať toto tvrdenie proti alternatíve, že priemerná cena nestúpla o 49%, na hladine významnosti 0.01. U 18 náhodne vybraných pozemkov sa zistilo zvýšenie ceny v priemere o 38% a štandardná odchýlka 14%.

H_0: $\mu = 49$ oproti H_1: $\mu \neq 49$

Pre $\alpha = 0.01$ a $(18 - 1) = 17$ df sú kritické hodnoty t-rozdelenia: ± 2.898

$$t = \frac{\bar{x} - \mu}{\left(\frac{\hat{s}}{\sqrt{n}}\right)} = \frac{38 - 49}{\left(\frac{14}{\sqrt{18}}\right)} = \frac{-11}{3.3} = -3.33$$



Oblasť zamietnutia	Oblasť nezamietnutia	Oblasť zamietnutia

Pretože testovacia štatistika je v dolnej oblasti zamietnutia, H_0 zamietame a môžeme uzavrieť, že priemerná cena stúpla o menej ako o 49%.

Príklad 5

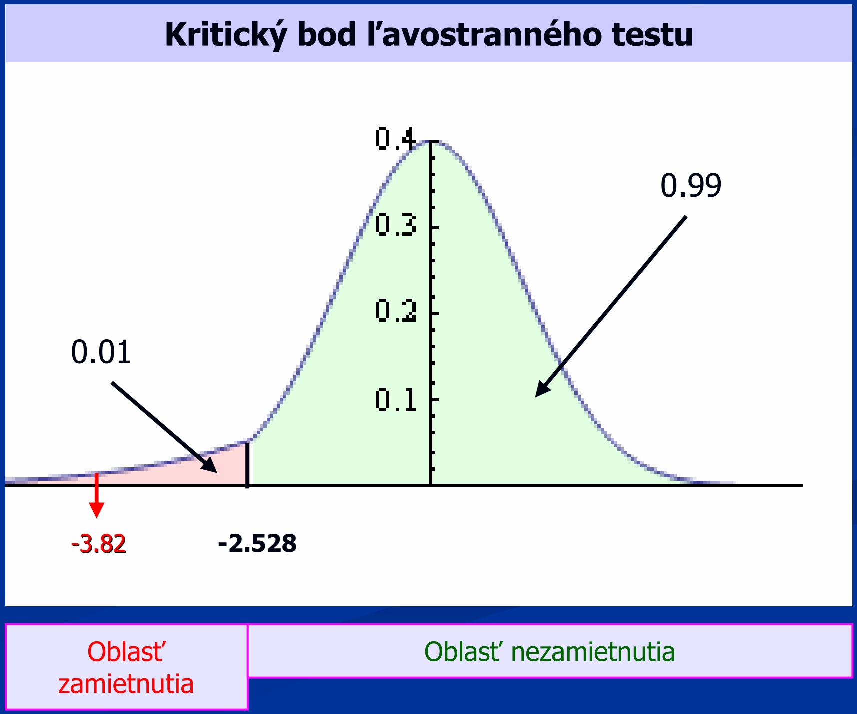
Výrobca tvrdí, že osvetľovacie teleso vydrží v priemere 65 hodín. Konkurencia predpokladá, že priemerná životnosť osvetľovacieho telesa je nižšia ako stanovuje výrobca a chce dokázať, že tvrdenie výrobcu je nesprávne. Uskutočnil sa náhodný výber 21 osvetľovacích telies. Vypočítaný výberový priemer je 62.5 hodín a výberová štandardná odchýlka je 3. Nech α = 0.01. Existujú dôvody pre záver, že tvrdenie výrobcu je nesprávne?

H_0: $\mu \geq 65$ oproti H_1: $\mu < 65$

Pre α = 0.01 a (21 – 1) = 20 df je kritická hodnota t-rozdelenia $t_{0.01}$: - 2.528

$$t = \frac{\bar{x} - \mu}{\left(\frac{\hat{s}}{\sqrt{n}}\right)} = \frac{62.5 - 65}{\left(\frac{3}{\sqrt{21}}\right)} = -3.82$$

Pretože testovacia štatistika padla do oblasti zamietnutia, H_0 zamietame a môžeme prijať záver, že tvrdenie výrobcu je nesprávne, a priemerná životnosť je menšia ako 65 hodín.



Príklad 6

Výrobca golfových palíc tvrdí, že váhy palíc sú starostlivo kontrolované, takže ich rozptyl nie je väčší ako 1 mg^2. Z náhodného výberu 31 palíc sa vypočítal výberový rozptyl 1.62 mg^2. Je to dostatočný dôvod na zamietnutie tvrdenia výrobcu pri $\alpha = 0.05$ (5%)?

Nech σ^2 je rozptyl populácie. Potom

H_0: $\sigma^2 \leq 1$ oproti H_1: $\sigma^2 > 1$

$$\chi^2 = \frac{(31-1)\cdot 1.62}{1} = 48.6$$

Pre $\alpha = 0.05$ a $(31 - 1) = 30$ df je kritická hodnota χ^2 – rozdelenia $\chi^2_{0.95}(30) = 43.773$

Pretože testovacia štatistika padla do oblasti zamietnutia (je väčšia ako kritická hodnota), H_0 zamietame a môžeme prijať záver, že tvrdenie výrobcu je nesprávne.