Matematická štatistika

Matematická štatistika skúma určitý **štatistický znak** (zárobok, hmotnosť, doba sledovania televízie,...) pre každú **štatistickú jednotku**, t. j. pre každý objekt skúmania.
Základný súbor, množina štatistických jednotiek, ktorá je predmetom skúmania.
Náhodný výber – vybraná vzorka, na ktorej sa zistí skúmaný znak a zo získaných údajov sa urobí zovšeobecnenie vo forme štatistického záveru o celom základnom súbore.
Hladina významnosti $\alpha \in (0, 1)$ – pravdepodobnosť toho, že náš štatistický záver je chybný.
Koeficient spoľahlivosti $\gamma = 1 - \alpha$ – pravdepodobnosť správneho záveru.
Budeme sa zaoberať týmito hlavnými úlohami štatistického skúmania:

a) odhady a intervaly spoľahlivosti parametrov základného súboru;

b) testovanie štatistických hypotéz.

Náhodný výber

Vlastnosti náhodného výberu:

1. každá jednotka základného súboru by mala rovnakú šancu dostať sa do náhodného výberu;
2. náhodný výber musí mať dostatočný počet prvkov vo vzťahu k počtu prvkov základného súboru.

Rozsah náhodného výberu – počet štatistických jednotiek v náhodnom výbere.
Zber dát – zisťovanie štatistického znaku na prvkoch náhodného výberu.
Triedenie dát:

- **Prosté triedenie** – usporiadanie podľa veľkosti $x_1 \leq x_2 \leq \cdots \leq x_n$, kde x_i sú namerané hodnoty.
- **Triedenie podľa početnosti** – pre každú nameranú hodnotu x_j je n_j počet výskytov hodnoty x_j medzi nameranými hodnotami, napr.

x_j	167	170	174	175	178	180
n_j	1	3	5	6	3	2

,

- **intervalové triedenie** – máme systém intervalov, každá nameraná hodnota x_i leží práve v jednom intervale. Intervalová frekvencia n_j –počet hodnôt v I_j

I_j	$16-20$	$20-24$	$24-28$	$28-32$	$32-36$
n_j	2	7	5	4	1

Výberové charakteristiky

Definícia

Nech sú dané namerané hodnoty $x_1, x_2, \ldots, x_n$, ktoré sú získané z náhodného výberu V_n.

Výberový priemer náhodného výberu V_n *je číslo*

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ *pre prosté triedenie,*

$\overline{x} = \frac{1}{n}\sum_{j=1}^{k} n_j x_j$ *pre triedenie podľa početnosti.* $(n = \sum\limits_{j=1}^{k} n_j)$

Modifikovaný výberový rozptyl (disperzia) náhodného výberu V_n *je číslo*

$s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ *pre prosté triedenie,*

$s^{*2} = \frac{1}{n-1}\sum_{j=1}^{k}(x_j - \overline{x})^2\, n_j$ *pre triedenie podľa početnosti.*

Modifikovaná výberová odchýlka náhodného výberu V_n *je číslo* $s^* = \sqrt{s^{*2}}$

Príklad

Boli namerané nasledujúce hodnoty výšok slovenských modeliek:

x_j	167	170	174	175	178	180
n_j	1	3	5	6	3	2

Určte výberový priemer, modifikovaný výberový rozptyl a modifikovanú výberovú odchýlku.

Riešenie. Výberové charakteristiky sú:

$$\overline{x} = \frac{1}{20} \cdot (167 \cdot 1 + 170 \cdot 3 + 174 \cdot 5 + 175 \cdot 6 + 178 \cdot 3 + 180 \cdot 2) = 174,55$$

$$s^{*2} = \frac{1}{19} \cdot \Big((167 - 174,55)^2 \cdot 1 + (170 - 174,55)^2 \cdot 3 + (174 - 174,55)^2 \cdot 5 +$$

$$(175 - 174,55)^2 \cdot 6 + (178 - 174,55)^2 \cdot 3 + (180 - 174,55)^2 \cdot 2\Big) = 11,4184$$

$$s^* = \sqrt{s^{*2}} = \sqrt{11,4184} = 3,3791$$

Bodové odhady parametrov základného súboru

Parametre základného súboru sú isté veličiny, ktoré ho charakterizujú a sú určené konkrétnymi hodnotami sledovaného znaku všetkých štatistických jednotiek celého základného súbor Vo väčšine prípadov je nereálne získať ich skutočné hodnoty. Môžeme ich však prostredníctvom náhodných výberov odhadnúť. Nech Q je sledovaný parameter základného súboru (napr. stredná hodnota μ, smerodajná odchýlka σ atď.) a $\hat{Q}_n$ je výberová charakteristika náhodného výberu V_n, kde n označuje rozsah náhodného výberu (napr. $\overline{x}$, s atď.). Pod **bodovým odhadom parametra Q** rozumieme takú výberovú charakteristiku $\hat{Q}_n$, ktorá nadobúda hodnoty blízke skutočnej hodnote parametra Q. Tento bodový odhad nazývame **nevychýleným** práve vtedy, keď $E(\hat{Q}_n) = Q$. Zapisujeme prirodzeným spôsobom: $Q \approx \hat{Q}_n$.

Veta

Výberový priemer $\overline{x}$ je nevychýleným odhadom strednej hodnoty μ a modifikovaný výberový rozptyl s^{*2} je nevychýleným odhadom rozptylu σ^2 základného súboru t. j.

$$\mu \approx \overline{x} \qquad a \qquad \sigma^2 \approx s^{*2}. \tag{1}$$

Pre základný súbor z predchádzajúceho príkladu máme
$\mu \approx 174,55; \; \sigma^2 \approx 11,4184$.

Predpokladajme, že $Y \sim norm(0,1)$. Nech $\alpha \in (0,1)$ je dané reálne číslo. Hľadáme číslo k_α, pre ktoré platí

$$P(|Y| > k_\alpha) = \alpha. \tag{2}$$

Pomocou opačného javu môžeme (**??**) zapísať v tvare

$$P(-k_\alpha \leq Y \leq k_\alpha) = 1 - \alpha \tag{3}$$

a keďže ide o normovanú spojitú náhodnú premennú, dostaneme

$$P(-k_\alpha \leq Y \leq k_\alpha) = 2 \cdot \Phi(k_\alpha) - 1,$$

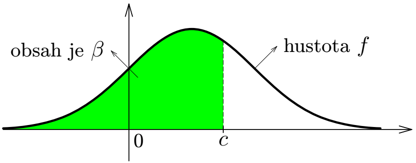
Z posledných dvoch rovností je $1 - \alpha = 2 \cdot \Phi(k_\alpha) - 1$ a odtiaľ $\Phi(k_\alpha) = 1 - \frac{\alpha}{2}$, čo znamená, že

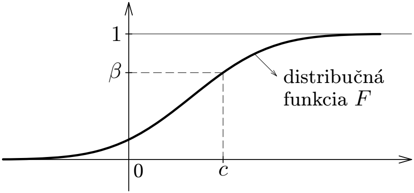
$$k_\alpha = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{4}$$

kde $\Phi^{-1}(x)$ je inverzná funkcia k distribučnej funkcii $\Phi(x)$. V matematickej štatistike sa k_α zvykne označovať takto:

$$k_\alpha = y_{1-\frac{\alpha}{2}} \tag{5}$$

Nech f je hustota a F distribučná funkcia náhodnej premennej X. Potom pre ľubovoľné $\beta \in (0,1)$ existuje také najmenšie reálne číslo c, že $F(c) \geq \beta$: t. j. c je najmenšie reálne číslo, pre ktoré je obsah vyšrafovanej časti na obrázku aspoň β.





Definícia

Najmenšie reálne číslo c, že $F(c) \geq \beta$ nazývame **β-kvantilom** *rozdelenia pravdepodobnosti F a označujeme ho $c = F^{-1}(\beta)$.*

Názov RP	NP	DF	β-kvantil
Normovaná normálna	Y	Φ	$\Phi^{-1}(\beta) = y_\beta$
Chí-kvadrát	χ^2	χ^2_n	$\chi^2_{\beta,n}$
Studentovo t	T	t_n	$t_{\beta,n}$
Fisherovo	F	$F_{n,m}$	$F_{\beta;n,m}$

Intervalové odhady parametrov základného súboru

Hlavná myšlienka intervalového odhadu parametra Q základného súboru spočíva v nájdení istého intervalu I, v ktorom leží s nami zvolenou pravdepodobnosťou p skutočná hodnota parametra Q. Potom platí, že do I na $100 \cdot p$ percent patrí skutočná hodnota Q.

Definícia

Intervalovým odhadom parametra Q základného súboru na hladine významnosti $\alpha \in (0, 1)$ *nazývame taký „číselný interval" $\langle Q_1, Q_2 \rangle$, v ktorom parameter Q leží s pravdepodobnosťou $1 - \alpha$, t. j.*

$$P(Q_1 \leq Q \leq Q_2) = 1 - \alpha, \tag{6}$$

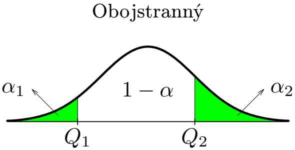
kde Q_1 a Q_2 je dvojica čísel, ktorá závisí od realizovaného náhodného výberu V_n. Interval $\langle Q_1, Q_2 \rangle$ nazývame aj **$100 \cdot (1 - \alpha)$-percentný interval spoľahlivosti pre parameter Q** *alebo skrátene* **obojstranný interval spoľahlivosti**. *Číslo $(1 - \alpha)$ nazývame* **koeficient spoľahlivosti intervalového odhadu $\langle Q_1, Q_2 \rangle$**.

Definícia

Jednostranné intervaly spoľahlivosti pre parameter Q *definujeme nasledovne:*

$$\begin{cases} \textbf{ľavostranný interval } \langle Q_1, \infty) : \; P(Q_1 \leq Q) = 1 - \alpha \\ \textbf{pravostranný interval } (-\infty, Q_2\rangle : \; P(Q \leq Q_2) = 1 - \alpha. \end{cases} \tag{7}$$

Grafická interpretácia oboch definícií je na obrázku. Pri obojstrannom intervale spoľahlivosti je $\alpha = \alpha_1 + \alpha_2$. V prípade symetrického obojstranného intervalu spoľahlivosti je $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.



V ďalšom texte bude na pozícii parametra Q vystupovať stredná hodnota μ alebo rozptyl σ^2.

Veta (Intervaly spoľahlivosti pre strednú hodnotu μ, ak σ poznáme)

Nech náhodná premenná X základného súboru má normálne rozdelenie pravdepodobnosti s parametrami μ a σ (t. j. $X \sim norm(\mu, \sigma)$). Potom symetrický obojstranný interval spoľahlivosti pre strednú hodnotu μ na hladine významnosti α má tvar

$$\mu \in \left\langle \underbrace{\overline{x} - y_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{Q_1}, \underbrace{\overline{x} + y_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{Q_2} \right\rangle \tag{8}$$

a jednostranné intervaly majú tvar

$$\mu \in \left\langle \underbrace{\overline{x} - y_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}}_{Q_1}, \infty \right), \quad \textit{resp.} \quad \mu \in \left(-\infty, \underbrace{\overline{x} + y_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}}_{Q_2} \right\rangle, \tag{9}$$

kde $\overline{x}$ je výberový priemer, n je rozsah výberu a $y_{1-\frac{\alpha}{2}}$ a $y_{1-\alpha}$ sú kvantily normovaného normálneho rozdelenia.

Veta (Intervaly spoľahlivosti pre strednú hodnotu μ, ak σ nepoznáme)

Nech náhodná premenná X základného súboru má normálne rozdelenie pravdepodobnosti s parametrami μ a σ (t. j. $X \sim norm(\mu, \sigma)$). Potom obojstranný interval spoľahlivosti pre strednú hodnotu μ na hladine významnosti α má tvar

$$\mu \in \left\langle \underbrace{\overline{x} - t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s^*}{\sqrt{n}}}_{Q_1}, \underbrace{\overline{x} + t_{1-\frac{\alpha}{2}, n-1} \cdot \frac{s^*}{\sqrt{n}}}_{Q_2} \right\rangle \tag{10}$$

a jednostranné intervaly majú tvar

$$\mu \in \left\langle \underbrace{\overline{x} - t_{1-\alpha, n-1} \cdot \frac{s^*}{\sqrt{n}}}_{Q_1}, \infty \right), \text{ resp. } \mu \in \left(-\infty, \underbrace{\overline{x} + t_{1-\alpha, n-1} \cdot \frac{s^*}{\sqrt{n}}}_{Q_2} \right\rangle, \tag{11}$$

kde $\overline{x}$ je výberový priemer, s^* je výberová modifikovaná smerodajná odchýlka, n je rozsah výberu a $t_{1-\frac{\alpha}{2}, n-1}$ a $t_{1-\alpha, n-1}$ sú kvantily t-rozdelenia pravdepodobnosti.

Veta (Intervaly spoľahlivosti pre rozptyl σ^2)

Nech náhodná premenná X základného súboru má normálne rozdelenie pravdepodobnosti s parametrami μ a σ (t. j. $X \sim norm(\mu, \sigma)$). Potom obojstranný interval spoľahlivosti pre rozptyl σ^2 na hladine významnosti α má tvar

$$\sigma^2 \in \left\langle \underbrace{\frac{(n-1)\cdot s^{*2}}{\chi^2_{1-\frac{\alpha}{2},\, n-1}}}_{Q_1}, \quad \underbrace{\frac{(n-1)\cdot s^{*2}}{\chi^2_{\frac{\alpha}{2},\, n-1}}}_{Q_2} \right\rangle \tag{12}$$

a jednostranné intervaly majú tvar

$$\sigma^2 \in \left\langle \underbrace{\frac{(n-1)\cdot s^{*2}}{\chi^2_{1-\alpha,\, n-1}}}_{Q_1}, \quad \infty \right) \quad \textit{resp.} \quad \sigma^2 \in \left(0\,, \quad \underbrace{\frac{(n-1)\cdot s^{*2}}{\chi^2_{\alpha,\, n-1}}}_{Q_2} \right\rangle \tag{13}$$

kde s^{*2} je výberový modifikovaný rozptyl, n je rozsah výberu a $\chi^2_{1-\frac{\alpha}{2},\, n-1}$ a $\chi^2_{1-\alpha,\, n-1}$ sú kvantily rozdelenia pravdepodobnosti χ^2 s $n-1$ stupňami voľnosti.

Príklad

Boli namerané nasledujúce hodnoty výšky slovenských modeliek:

x_j	167	170	174	175	178	180
n_j	1	3	5	6	3	2

.

Určte:
a) 90%- ný obojstranný interval spoľahlivosti pre strednú hodnotu;
b) 95%- ný pravostranný interval spoľahlivosti pre rozptyl a smerodajnú odchýlku.

Riešenie. Z predchádzajúceho príkladu máme
$\overline{x} = 174,55, \quad s^{*2} = 11,4184, \quad s^* = \sqrt{s^{*2}} = \sqrt{11,4184} = 3,3791.$
a) Máme $n = 20$, $1 - \alpha = 0,9 \Rightarrow \alpha = 0,1$ a σ^2 je neznáme. Dostávame

$$\mu \in \langle \bar{x} - t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s^*}{\sqrt{n}}, \bar{x} - t_{1-\frac{\alpha}{2},n-1} \cdot \frac{s^*}{\sqrt{n}} \rangle$$

Hodnotu $t_{1-\frac{\alpha}{2},n-1} = t_{0,95;19}$ nájdeme v tabuľke **Kvantily rozdelenia t (t_P)** nasledovne: Nájdeme riadok prislúchajúci hodnote $k = 19$ a stĺpec prislúchajúci hodnote $P = 0,95$, teda $t_{0,95;19} = 1,7291$. Dostávame

$$\mu \in \langle 174,55 - 1,7291 \cdot \frac{3,3791}{\sqrt{20}},\ 174,55 + 1,7291 \cdot \frac{3,3791}{\sqrt{20}} \rangle \approx \langle 173,2365; 175,8634 \rangle.$$

b) Máme $1 - \alpha = 0,95 \Rightarrow \alpha = 0,05$. Pre rozptyl dostávame:

$$\sigma^2 \in \left(0, (n-1) \cdot \frac{s^{*2}}{\chi^2_{\alpha,n-1}}\right\rangle$$

Hodnotu $\chi^2_{\alpha,n-1} = \chi^2_{0,05;19}$ nájdeme v tabuľke **Kvantily rozdelenia chi-kvadrát** nasledovne: Nájdeme riadok prislúchajúci hodnote $k = 19$ a stĺpec prislúchajúci hodnote $P = 0,05$, teda $\chi^2_{0,05;19} = 10,117$. Dostávame

$$\sigma^2 \in \left\langle 0;\ \frac{19 \cdot 11,4184}{10,117} \right\rangle \approx \langle 0;\ 21,444 \rangle.$$

Pre smerodajnú odchýlku máme

$$\sigma \in \left\langle 0;\ \sqrt{\frac{19 \cdot 11,4184}{10,117}} \right\rangle \approx \langle 0;\ 4,6307 \rangle.$$

Testovanie štatistických hypotéz

Štatistická hypotéza je istá domnienka o vlastnostiach rozdelenia pravdepodobnosti náhodnej premennej X alebo viacerých náhodných premenných. **Testovanie štatistickej hypotézy** (môžeme povedať aj overovanie pravdivosti domnienky) je postup, pri ktorom na základe náhodného výberu zo základného súboru rozhodneme, či na zvolenej hladine významnosti α (t. j. so spoľahlivosťou $1-\alpha$) danú hypotézu zamietame (neprijmeme) alebo nezamietame (prijmeme). V prípade zamietnutia prijmene **alternatívnu** („opačnú“) hypotézu.

Príklad

Chceme zistiť, či v podniku priemerný denný odpad μ istého kovu nepresiahne 26 jednotiek hmotnosti. Potom by sme testovali hypotézu $\mu \leq 26$, pričom alternatívna hypotéza by mala tvar $\mu > 26$. Na zjednodušenie testovacieho postupu môžeme hypotézu $\mu \leq 26$ zapísať v tvare rovnosti $\mu = 26$. Týmto krokom nestratíme na všeobecnosti úvah, lebo predpokladáme krajnú prípustnú hranicu odpadu. Teda na základe zistených odpadov počas istého počtu dní (náhodný výber) by sme testovali hypotézu $\mathcal{H}_0\colon\ \mu = 26$ oproti alternatívnej hypotéze $\mathcal{H}_1\colon\ \mu > 26$. Takému testovaniu budeme hovoriť **test zhody parametra so známou konštantou** *(parameter je μ a konštanta je 26).*

Základné pojmy testovania štatistických hypotéz

1. **Nulová hypotéza** $\mathcal{H}_0$ je hypotéza (domnienka), ktorej platnosť overujeme. Je napr. tvaru $\mathcal{H}_0\colon\ Q = Q_0$, kde Q je parameter základného súboru a Q_0 je konkrétna konštanta. Môžeme ju zapísať aj v tvare $\mathcal{H}_0\colon\ Q - Q_0 = 0$ (porovnávanie s nulou), a preto je štandardne používaný názov nulová hypotéza.
2. **Alternatívna hypotéza** $\mathcal{H}_1$ je hypotéza (domnienka), ktorú prijímame v prípade neprijatia nulovej hypotézy. Jej tvar závisí od samotnej formulácie testovania. Uvedieme tieto tri základné tvary:
   a) **pravostranná** alternatívna hypotéza : $\mathcal{H}_1\colon\ Q > Q_0$;
   b) **ľavostranná** alternatívna hypotéza : $\mathcal{H}_1\colon\ Q < Q_0$;
   c) **obojstranná** alternatívna hypotéza : $\mathcal{H}_1\colon\ Q \neq Q_0$.
3. **Testovacia charakteristika** G je istá konkrétna funkcia, ktorá závisí od náhodného výberu. Pre každú dvojicu $\mathcal{H}_0$ a $\mathcal{H}_1$ má špeciálny tvar a rozdelenie pravdepodobnosti.
4. **Kritická oblasť** K_α alebo **oblasť zamietnutia** je množina, ktorá je určená kvantilom testovacej charakteristiky G.
5. Hypotézu $\mathcal{H}_0$ zamietame na hladine významnosti α práve vtedy, keď hodnota testovacej charakteristiky $G \in K_\alpha$.

Etapy testovania štatistických hypotéz

1. Formulujeme predpoklady o náhodných premenných, ktorých sa testovanie týka.
2. Zvolíme hladinu významnosti α, resp. koeficient spoľahlivosti $\gamma = 1 - \alpha$.
3. Formulujeme nulovú hypotézu $\mathcal{H}_0$ a alternatívnu hypotézu $\mathcal{H}_1$.
4. Zo získaného náhodného výberu vyčíslime hodnotu príslušnej testovacej charakteristiky $G = g(x_1, x_2, \ldots, x_n)$.
5. Na základe zodpovedajúceho kvantilu určíme kritickú oblasť K_α.
6. Urobíme záver testovania, ktorý spočíva
   – buď v zamietnutí $\mathcal{H}_0$ (a teda prijatí $\mathcal{H}_1$)
   – alebo v prijatí (nezamietnutí) $\mathcal{H}_0$.

Predpokladáme, že náhodná premenná má normálne rozdelenie pravdepodobnosti, t. j. $X \sim norm(\mu, \sigma)$ a budeme skúmať zhodu parametra μ alebo σ, resp. σ^2, so známou konštantou. K tomu budeme potrebovať jeden náhodný výber rozsahu n s nameranými hodnotami x_1, x_2, ..., x_n (preto týmto testom hovoríme **jednovýberové testy**) a z neho vypočítané výberové charakteristiky $\overline{x}$, s^* atď.

Y-test zhody strednej hodnoty so známou konštantou μ_0 (σ poznáme)

Nulová hypotéza: $\mathcal{H}_0: \ \mu = \mu_0$.

Testovacia charakteristika: $Y = \dfrac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n}\,.$

Kritická oblasť K_α pre alternatívnu hypotézu

a) $\mathcal{H}_1: \ \mu > \mu_0$ je $K_\alpha = (y_{1-\alpha}\,;\,\infty)$

b) $\mathcal{H}_1: \ \mu < \mu_0$ je $K_\alpha = (-\infty\,;\,-y_{1-\alpha})$

c) $\mathcal{H}_1: \ \mu \neq \mu_0$ je $K_\alpha = (-\infty\,;\,-y_{1-\frac{\alpha}{2}}) \cup (y_{1-\frac{\alpha}{2}}\,;\,\infty)$

kde y_β je β- kvantil normovaného normálneho rozdelenia pravdepodobnosti.

Príklad

Pri tradičnom spôsobe opracovania súčiastok sa dosahovali priemerné hodnoty 4,4 istej kvalitatívnej vlastnosti so smerodajnou odchýlkou $\sigma = 0,4$. Pokusne sa zavádza nová metóda opracovania súčiastok, ktorou opracovali 20 súčiastok a dosiahli sa takéto výsledky:

$$4,5; \quad 4,3; \quad 4,1; \quad 4,9; \quad 4,6; \quad 3,6; \quad 4,7; \quad 5,1; \quad 4,8 \quad 4,0;$$

$$3,7; \quad 4,4; \quad 4,9; \quad 4,9; \quad 5,2; \quad 5,1; \quad 4,7; \quad 4,9; \quad 4,6; \quad 4,8.$$

Na hladine významnosti 0,05 testujte hypotézu $\mathcal{H}_0: \mu = 4{,}4$ oproti $\mathcal{H}_1: \mu > 4{,}4$ za predpokladu normálneho rozdelenia pravdepodobnosti hodnôt sledovanej vlastnosti.

Riešenie. Máme dané $n = 20, \ \sigma = 0,4$. Vypočítame $\overline{x} = 4,59$. Túto hodnotu dosadíme do testovacej charakteristiky. Dostávame

$$Y = \frac{\overline{x} - \mu_0}{\sigma} \cdot \sqrt{n} = \frac{4,59 - 4,4}{0,4} \cdot \sqrt{20} = 2,1243$$

Pre určenie kritickej oblasti potrebujeme hodnotu kvantilu normálneho rozdelenia $y_{1-\alpha} = y_{0,95} = 1,6449$. Kritická oblasť je $K_\alpha = (1,6449, \infty)$. Keďže $Y \in K_\alpha$, zamietame $\mathcal{H}_0$ a prijímame $\mathcal{H}_1$.

t-test zhody strednej hodnoty so známou konštantou μ_0 (σ nepoznáme)

Nulová hypotéza: $\mathcal{H}_0: \ \mu = \mu_0$.
Testovacia charakteristika:

$$t = \frac{\overline{x} - \mu_0}{s^*} \cdot \sqrt{n}\,. \tag{14}$$

Kritická oblasť K_α pre alternatívnu hypotézu
a) $\mathcal{H}_1: \ \mu > \mu_0$ je $K_\alpha = (t_{1-\alpha,\, n-1}\,;\, \infty)$
b) $\mathcal{H}_1: \ \mu < \mu_0$ je $K_\alpha = (-\infty\,;\, -t_{1-\alpha,\, n-1})$
c) $\mathcal{H}_1: \ \mu \neq \mu_0$ je $K_\alpha = (-\infty\,;\, -t_{1-\frac{\alpha}{2},\, n-1}) \cup (t_{1-\frac{\alpha}{2},\, n-1}\,;\, \infty)$
kde $t_{\beta,\, n-1}$ je β kvantil Studentovho t-rozdelenia pravdepodobnosti s $n-1$ stupňami voľnosti.

Príklad

Pri kontrole v pivárni kontrolór nameral nasledujúce hodnoty objemu načapovaných Zlatých Bažantov (v litroch):

0,49; 0,5; 0,48; 0,47; 0,505; 0,485; 0,49; 0,495; 0,5; 0,48.

Na hladine významnosti $\alpha = 0,05$ *otestujte hypotézu* $\mathcal{H}_0 : \mu = 0{,}5$ *oproti hypotéze* $\mathcal{H}_1 : \mu < 0{,}5$. *Môžeme na tejto hladine významnosti tvrdiť, že sa pivo nedolieva?*

Riešenie. Testujeme hypotézu $\mathcal{H}_0$ oproti alternatívnej hypotéze $\mathcal{H}_1 : \mu < 0,5$ na hladine významnosti $\alpha = 0,05$.

Vypočítame $\overline{x}$ a s^*. Dostávame $\overline{x} = 0,4895$, ${s^*}^2 = 1,1917 \cdot 10^{-4}$, $s^* = 0,0109$. Tieto hodnoty dosadíme do testovacej charakteristiky. Dostávame

$$t = \frac{\overline{x} - \mu_0}{s^*} \cdot \sqrt{n} = \frac{0,4895 - 0,5}{0,0109} \cdot \sqrt{10} = -3,0417$$

Kritická oblasť pre $\mathcal{H}_1 : \mu < 0,5$ je $K_\alpha = (-\infty, -t_{1-\alpha,n-1})$. V tabuľkách nájdeme hodnotu kvantilu t-rozdelenia $t_{1-\alpha,n-1} = t_{0,95;9} = 1,8331$. Platí, že $-3,0417 \in (-\infty, -1,8331)$. Keďže $t \in K_\alpha$, zamietame $\mathcal{H}_0$ a prijímame $\mathcal{H}_1$. Teda na hladine významnosti 0,05 môžeme tvrdiť, že pivo sa nedolieva.

χ^2-test zhody rozptylu σ^2 so známou konštantou σ_0^2

Nulová hypotéza: $\mathcal{H}_0 \colon \ \sigma^2 = \sigma_0^2$.
Testovacia charakteristika:

$$\chi^2 = \frac{n-1}{\sigma_0^2} \cdot s^{*2}. \tag{15}$$

Kritická oblasť K_α pre alternatívnu hypotézu
a) $\mathcal{H}_1 \colon \ \sigma^2 > \sigma_0^2$ je $K_\alpha = (\chi^2_{1-\alpha,\, n-1}\,;\, \infty)$
b) $\mathcal{H}_1 \colon \ \sigma^2 < \sigma_0^2$ je $K_\alpha = (0;\, \chi^2_{\alpha,\, n-1})$,
c) $\mathcal{H}_1 \colon \ \sigma^2 \neq \sigma_0^2$ je $K_\alpha = (0;\, \chi^2_{\frac{\alpha}{2},\, n-1}) \cup (\chi^2_{1-\frac{\alpha}{2},\, n-1}\,;\, \infty)$
kde $\chi^2_{\beta,\, n-1}$ je β kvantil χ^2 rozdelenia pravdepodobnosti s $n-1$ stupňami voľnosti.

Príklad

Náhodným výberom v dvanástich predajniach sa preverovali počty tovarov predávaných po záruke. Výsledky boli nasledovné:

$$15,\ 21,\ 17,\ 15,\ 17,\ 15,\ 21,\ 17,\ 18,\ 21,\ 18,\ 18.$$

Na hladine významnosti $\alpha = 0,05$ *otestujte hypotézu* $\mathcal{H}_0 : \sigma^2 = 6$ *proti* $\mathcal{H}_1 : \sigma^2 < 6$.

Máme $n = 12$, $\sigma_0^2 = 6$. Vypočítame $\overline{x}$ a ${s^*}^2$. Dostávame $\overline{x} = 17,75$, ${s^*}^2 = 5,113636$. Tieto hodnoty dosadíme do testovacej charakteristiky. Dostávame

$$\chi^2 = \frac{n-1}{\sigma_0^2} \cdot {s^*}^2 = \frac{11}{6} \cdot 5,113636 = 9,375$$

Pre určenie kritickej oblasti potrebujeme hodnotu kvantilu χ^2 rozdelenia: $\chi^2_{\alpha,n-1} = \chi^2_{0,05;11} = 4,57481$. Kritická oblasť je

$$K_\alpha = (0; 4,57481).$$

Keďže $\chi^2 \notin K_\alpha$, prijímame $\mathcal{H}_0$.