1 Štatistika (O.Nánásiová)

1.1 Bodové odhady .

Odhady momentov:

$$\mu \approx \frac{\sum_i X_i}{n} = \overline{X}$$

$$\sigma^2 \approx \begin{cases} \frac{\sum_i (X_i-\mu)^2}{n} = S_0^2 & \mu \text{ je známe} \\ \\ \frac{\sum_i (X_i-\overline{X})^2}{n-1} = S^2 & \mu \text{ je neznáme.} \end{cases}$$

1.2 Intervalové odhady

Bodový odhad, stanovený podľa niektorej z metód, je vlastne tiež náhodnou premennou. Pomocou intervalového odhadu stanovujeme interval (interval spoľahlivosti), do ktorého skutočná hodnota odhadovaného parametra padne s dopredu stanovenou pravdepodobnosťou (spoľahlivosťou). Nech T je nejaký bodový odhad odhadovaného parametra θ. Potom intevral spoľahlivosti pre parameter θ vypočítame pomocou pravdepodobnostného rozdelenia bodového odhadu T.

Nech $(1-\alpha)\cdot 100\,\%, \quad (\alpha \in (0,1))$ je dopredu daná spoľahlivosť. Rozoznávame tri základné druhy intervalov spoľahlivosti:

(OIS) Pod obojstranným intervalom spoľahlivosti rozumieme interval $(a,b]$, pre ktorý platí

$$P(\theta \in (a,b]) = 1-\alpha \quad \text{a} \quad P(\theta \le a) = P(\theta > b) = \frac{\alpha}{2}. \tag{1}$$

(LIS) Pod ľavostranným intervalom spoľahlivosti rozumieme interval (a,∞), pre ktorý platí

$$P(\theta \in (a,\infty)) = 1-\alpha \quad \text{a} \quad P(\theta \le a) = \alpha. \tag{2}$$

(PIS) Pod pravostranným intervalom spoľahlivosti rozumieme interval $(-\infty,b]$, pre ktorý platí

$$P(\theta \in (-\infty,b]) = 1-\alpha \quad \text{a} \quad P(\theta \ge b) = \alpha. \tag{3}$$

V praktických aplikáciách sa najčastejšie používa

- $\alpha = 0,05$ (spoľahlivosť 95 %),
- $\alpha = 0,1$ (spoľahlivosť 90 %),
- $\alpha = 0,01$ (spoľahlivosť 99 %).

1.3 Rozdelenia odhadov μ a σ^2 pre náhodný výber z $N(\mu, \sigma^2)$

Nech $X_1, \dots, X_n$ je náhodný výber z $N(\mu, \sigma^2)$. Z vlastností normálneho rozdelenia vieme, že

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right), \tag{4}$$

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \qquad \text{a} \qquad \frac{(\overline{X}-\mu)^2}{\frac{\sigma^2}{n}} \sim \chi^2(1). \tag{5}$$

Pretože $\{X_i\}_i$ je náhodný výber a pre každé $i = 1, \dots, n$

$$Y_i = \frac{X_i - \mu}{\sigma} \sim N(0,1),$$

tak

$$\sum_{i=1}^{n} Y_i^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i - \mu)^2 \sim \chi^2(n). \tag{6}$$

Z Vety **??**, (4) a (6) a vlastností rozdelenia χ^2 vyplýva, že

$$\underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}_{\chi^2(n)} = \underbrace{n\frac{(\overline{X}-\mu)^2}{\sigma^2}}_{\chi^2(1)} + \underbrace{\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}_{\chi^2(n-1)}. \tag{7}$$

Teraz môžeme sformulovať vetu o pravdepodobnostných rozdeleniach odhadov parametrov $N(\mu, \sigma^2)$.

Veta 1:
Nech $\{X_i\}_{i=1}^n$ je náhodný výber z $N(\mu, \sigma^2)$ a

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2, \qquad S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

Potom platí

1. $$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1); \tag{8}$$

2. $$\frac{nS_0^2}{\sigma^2} \sim \chi^2(n); \tag{9}$$

3. $$\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2(n-1); \tag{10}$$

4. $$\frac{\overline{X}-\mu}{S_1}\sqrt{n} \sim t(n-1). \tag{11}$$

Dôkaz. Napríklad (9) vyplýva v faktu, že

$$\frac{n}{\sigma^2}S_0^2 = \frac{n}{\sigma^2}\cdot\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 .$$

Aj zvyšné tvrdenia (8), (10) a (11) vyplývajú priamo z (4) až (7).
Úplne dôkazy tvrdení Vety 1 možno nájsť napr. v [?, ?].

Poznámka 1:
Veta 1 je fundametálna veta nielen pre konštrukciu intervalových odhadov parametrov $N(\mu, \sigma^2)$. Vzhľadom na centrálnu limitnú vetu sa jej tvrdenia používajú v praxi aj v prípadoch, v ktorých nejde o náhodný výber z normálneho rozdelenia.

1.4 Intervalové odhady pre μ v prípade náhodného výberu z $N(\mu, \sigma^2)$

V celej tejto časti budeme predpokladat', že $\{X_i\}_{i=1}^n$ je náhodný výber z $N(\mu, \sigma^2)$. Konštrukciu intervalových odhadov musíme vzhľadom na Vetu 1 rozdeliť na dva prípady:

a) menej častý prípad: σ^2 poznáme.

b) častejší prípad: σ^2 nepoznáme.

a) Interval spoľahlivosti pre μ, ak poznáme σ^2

Pre (OIS) platí vzt'ah (1)

$$P(\theta \in (a, b]) = 1 - \alpha \qquad \text{a} \qquad P(\theta \leq a) = P(\theta > b) = \frac{\alpha}{2}.$$

Neznámy parameter θ je μ. Použijeme (8) z Vety 1. Pripomíname, že funkcia hustoty pre $N(0,1)$ je párna funkcia a pre každú spojitú distribučnú funkciu F platí

$$F(t) = P(X \leq t) = P(X < t).$$

Pre $\alpha \in (0,1)$ $(\alpha < 0,5)$ hľadáme číslo $K > 0$ s vlastnost'ou:

$$P(|Y| < K) = 1 - \alpha \qquad \text{a} \qquad P(Y \leq -K) = P(Y \geq K) = \frac{\alpha}{2}.$$

Teda $P(Y < K) = \Phi(K) = 1 - \frac{\alpha}{2}$. To znamená, že K je $\left(1 - \frac{\alpha}{2}\right)$-kvantil pre $N(0,1)$:

$$K = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Ak dosadíme za Y

$$P\left(-K < \frac{\overline{X} - \mu}{\sigma}\sqrt{n} < K\right) = 1 - \alpha$$

a výraz upravíme dostaneme:

$$P\left(\overline{X} - \frac{K \cdot \sigma}{\sqrt{n}} < \mu < \overline{X} + \frac{K \cdot \sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Pre interval

$$\left(\overline{X} - \frac{\Phi^{-1}(1-\frac{\alpha}{2})\cdot\sigma}{\sqrt{n}}, \quad \overline{X} + \frac{\Phi^{-1}(1-\frac{\alpha}{2})\cdot\sigma}{\sqrt{n}}\right),$$

ktorý nazývame OIS pre μ, ak poznáme σ^2 platí

$$P\left(\mu \in \left(\overline{X} - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right)\right) = 1-\alpha.$$

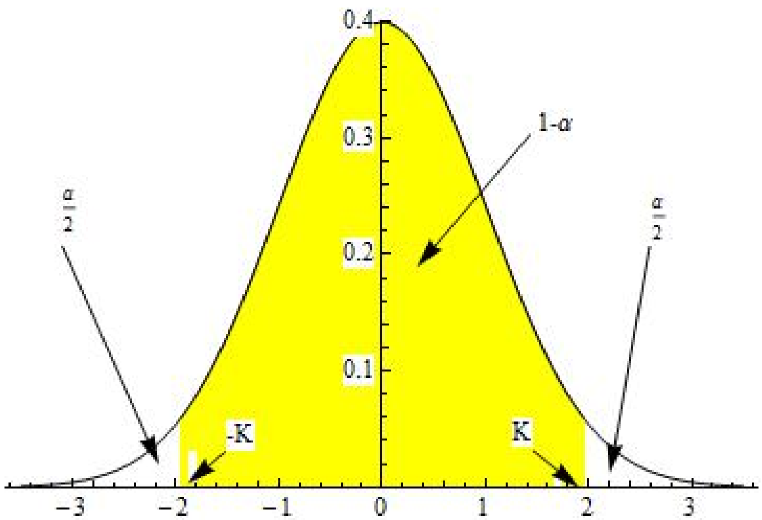

Fig. 1: $P(|Y| < K) = 1-\alpha$, pre $\alpha = 0,05$

Jednostranné intervaly spoľahlivosti (LIS, PIS).
Pretože funkcia hustoty pre $N(0,1)$ je párna funkcia a platí $\Phi(-K) = 1 - \Phi(K)$, v oboch prípadoch stačí nájsť v tabuľke pre $N(0,1)$ hodnotu pre $1-\alpha$-kvantil $(K > 0)$

$$K = \Phi^{-1}(1-\alpha).$$

Pre n.p. Y a číslo K platí

$$\alpha = P\left(Y \geq K\right) = P\left(Y < -K\right). \tag{12}$$

Pre (LIS) sa použijeme vzťah (2) a (8) a pre (PIS) (3), (8). Postupujeme analogicky ako v predošlom prípade. Pretože $Y \sim N(0,1)$ a $K = \Phi^{-1}(1-\alpha)$, tak stačí do (12) za Y a K dosadiť

$$Y = \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \qquad \text{a} \qquad K = \Phi^{-1}(1-\alpha)$$

a upraviť nerovnosti.

Zhrnutie:

1. $(1-\alpha)\cdot 100\,\%$-ný OIS pre μ, ak poznáme σ^2:

$$\left(\overline{X} - \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}, \quad \overline{X} + \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right);$$

2. $(1-\alpha)\cdot 100\,\%$-ný LIS pre μ, ak poznáme σ^2

$$\left(\overline{X} - \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}, \quad \infty\right);$$

3. $(1-\alpha)\cdot 100\,\%$-ný PIS pre μ, ak poznáme σ^2

$$\left(-\infty, \quad \overline{X} + \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}\right).$$

b) Interval spoľahlivosti pre μ, ak nepoznáme σ^2

V tomto prípade nahradíme σ^2 jeho odhadom

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})$$

a namiesto $\Phi(K)$ použijeme distribučnú funkciu pre $t(n-1)$, ktorú označíme $F_{t(n-1)}$ a použijeme Studentovo rozdelenie s $n-1$ stupňami voľnosti. Vzhľadom na vlastnosti Studentovho rozdelenia podobné vlastnostiam štandardizovaného normálneho rozdelenia $N(0,1)$ budeme postupovať analogicky ako v predošlom prípade, pričom $(1-\alpha)$-kvantil sa označuje $t_{n-1}(1-\alpha)$

$$t_{n-1}(1-\alpha) = F^{-1}_{t(n-1)}(1-\alpha)$$

Zhrnutie:

1. $(1-\alpha)\cdot 100\%$-ný OIS pre μ, ak nepoznáme σ^2:

$$\left(\overline{X} - t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S_1}{\sqrt{n}}, \quad \overline{X} + t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S_1}{\sqrt{n}}\right);$$

2. $(1-\alpha)\cdot 100\,\%$-ný $(1-\alpha)\cdot 100\,\%$-ný LIS pre μ, ak nepoznáme σ^2:

$$\left(\overline{X} - t_{n-1}(1-\alpha)\frac{S_1}{\sqrt{n}}, \quad \infty\right).$$

3. $(1-\alpha)\cdot 100\,\%$-ný PIS pre μ, ak nepoznáme σ^2:

$$\left(-\infty, \quad \overline{X} + t_{n-1}(1-\alpha)\frac{S_1}{\sqrt{n}}\right);$$

Postup pri riešení:

1. Z nameraných $x_1, \ldots, x_n$ hodnôt, ktoré sú jednou realizáciou náhodného výberu z $N(\mu, \sigma^2)$, vypočítame aritmetický priemer.

2. Stanovíme si výšku spoľahlivosti $(1-\alpha)$ a nájdeme príslušné hodnoty kvantilov v štatistických tabuľkách, napr. EXCELu.

3. Dosadíme do vypočítaných vzorcov podľa toho, aký interval spoľahlivosti potrebujeme.

Príklad 1:
Predpokladajme, že firma, ktorá sa zaoberá výrobou výliskov z plastických hmôt, chce poznať účinok technologického procesu na kmitanie podlahy vo výrobnej hale. Ohrozenie podlahy sa hodnotí pomocou vibrácií, ktorými sa

podlaha rozkmitá v dôsledku rázov pri lisovaní. Meranie sme opakovali 20-krát. Z nameraných hodnôt vibrácií sme vypočítali bodový odhad strednej hodnoty μ ako aritmetický priemer nameraných údajov

$$\overline{X} = \frac{1}{20}\sum_{i=1}^{20} x_i = 1,25$$

a vieme, že $\sigma^2 = 0,81$.

To znamená, že náhodný výber, ktorého realizáciou sú namerané hodnoty

$$x_1, \ldots, x_{20},$$

pochádza z normálneho rozdelenia $N(\mu, 0,81)$. Z toho vyplýva, že náhodná premenná $\overline{X} \sim N(\mu, \frac{0,81}{20})$. Potrebujeme nájsť OIS pre $\alpha = 0,05$. V tabuľke pre $N(0,1)$ nájdeme $(1-\frac{\alpha}{2})$-kvantil $\Phi^{-1}(0,975) \doteq 1,96$ a dosadíme do vzorca pre OIS

$$\left(1,25 - 1,96 \cdot \frac{0,9}{\sqrt{20}}, \quad 1,25 + 1,96 \cdot \frac{0,9}{\sqrt{20}}\right) = (0,9984; \quad 1,5016).$$

To znamená, že

$$P\left(\mu \in (0,9984; \quad 1,5016)\right) = 0,95.$$

Záver. Všetky hodnoty vibrácií v danej výrobnej hale sa nachádzajú s 95 %-tnou spoľahlivosťou v intervale $(0,9984\, 1,5016)$.

Príklad 2:
Potrebujeme riešiť problém z Príkladu 1, ale teraz nepoznáme σ^2 a hodnota $0,81$ je len odhad σ^2. To znamená, že $S_1^2 = 0,81$. V takomto prípade namiesto $\Phi^{-1}(0,975)$ použijeme $t_{19}(0,975)$, ktorú nájdeme v tabuľke pre $t(19)$. Pretože

$$t_{19}(0,975) \doteq 2,093,$$

potom OIS je

$$\left(1,25 - 2,093 \cdot \frac{0,9}{\sqrt{20}}, \quad 1,25 + 2,093 \cdot \frac{0,9}{\sqrt{20}}\right) = (0,8288; \quad 1,6712).$$

Vidíme, že tento interval je o niečo širší, ako v Príklade 1.
Záver: Všetky hodnoty vibrácií v danej výrobnej hale sa nachádzajú s 95 %-tnou spoľahlivosťou v intervale $(0,8288; \quad 1,6712)$.

1.5 Intervalové odhady pre σ^2 v prípade náhodného výberu z $N(\mu, \sigma^2)$

V celej tejto časti budeme predpokladať, že $\{X_i\}_{i=1}^n$ je náhodný výber z $N(\mu, \sigma^2)$. Konštrukciu intervalových odhadov musíme vzhľadom na Vetu 1 rozdeliť na dva prípady:

a) menej častý prípad: μ poznáme,

b) v castejší prípad: μ nepoznáme.

Budeme postupovať podobne ako pri intervaloch pre μ, len v tomto prípade použijeme iné pravdepodobnostné rozdelenia, konkrétne $\chi^2(n)$ v prípade (a) a $\chi^2(n-1)$ v prípade (b). Teraz budeme používať vzťahy (9) a (10) z Vety 1. To znamená, že budeme v tabuľkách pre pravdepodobnostné rozdelenia hľadať p-kvantily pre $\chi^2(n)$ a $\chi^2(n-1)$. Tieto kvantily sa zvyknú označovať nasledovne: ak k sú stupne voľnosti, tak p-kvantil pre $p \in (0,1)$ je

$$\chi_k^2(p).$$

a) Interval spoľahlivosti pre σ, ak poznáme μ

Z (9) vyplýva, že n.p.

$$Q = \frac{nS_0^2}{\sigma^2} \sim \chi^2(n).$$

To znamená, že pre $p \in (0,1)$ platí

$$P\left(Q < \chi_n^2(p)\right) = p$$

a

$$P\left(Q < \chi_n^2(1-p)\right) = 1 - p$$

čo je to isté, ako

$$P\left(\frac{1}{Q} > \frac{1}{\chi_n^2(1-p)}\right) = 1 - p.$$

Pripomíname, že pre $p_1 < p_2$ platí $\chi_n^2(p_1) < \chi_n^2(p_2)$.

Teraz odvodíme OIS ($\alpha < 0,5$). Ak

$$p_1 = \frac{\alpha}{2} \qquad \text{a} \qquad p_2 = 1 - \frac{\alpha}{2},$$

potom dostaneme

$$P\left(\chi_n^2\left(\frac{\alpha}{2}\right) < Q < \chi_n^2\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha$$

Ak dosadíme za za n.p. Q výraz nS_0^2/σ^2 dostaneme

$$P\left(\chi_n^2\left(\frac{\alpha}{2}\right) < \frac{nS_0^2}{\sigma^2} < \chi_n^2\left(1-\frac{\alpha}{2}\right)\right) = 1-\alpha$$

$$P\left(\frac{nS_0^2}{\chi_n^2\left(1-\frac{\alpha}{2}\right)} < \sigma^2 < \frac{nS_0^2}{\chi_n^2\left(\frac{\alpha}{2}\right)}\right) = 1-\alpha.$$

Intervaly LIS a PIS odvodíme podobne, a teda

$$P\left(\frac{nS_0^2}{\chi_n^2\left(1-\alpha\right)} < \sigma^2 < \infty\right) = 1-\alpha.$$

a

$$P\left(0 < \sigma^2 < \frac{nS_0^2}{\chi_n^2\left(\alpha\right)}\right) = 1-\alpha.$$

Zhrnutie:

1. $(1-\alpha)\cdot 100\,\%$-ný OIS pre σ^2, ak poznáme μ:

$$\left(\frac{nS_0^2}{\chi_n^2\left(1-\frac{\alpha}{2}\right)}, \quad \frac{nS_0^2}{\chi_n^2\left(\frac{\alpha}{2}\right)}\right);$$

2. $(1-\alpha)\cdot 100\,\%$-ný LIS pre σ^2, ak poznáme μ:

$$\left(\frac{nS_0^2}{\chi_n^2\left(1-\alpha\right)}, \quad <\infty\right);$$

3. $(1-\alpha)\cdot 100\,\%$-ný PIS pre σ^2, ak poznáme μ:

$$\left(0; \quad \frac{nS_0^2}{\chi_n^2\left(\alpha\right)}\right).$$

b) Interval spoľahlivosti pre σ^2, ak nepoznáme μ

V podstate ide o ten istý problém, lenže v tomto prípade použijeme p-kvantily z $\chi^2(n-1)$ a odhad S_1^2.

Zhrnutie:

1. $(1-\alpha)\cdot 100\,\%$-ný OIS pre σ^2, ak nepoznáme μ:
$$\left(\frac{(n-1)S_1^2}{\chi^2_{n-1}\left(1-\frac{\alpha}{2}\right)}, \quad \frac{(n-1)S_1^2}{\chi^2_{n-1}\left(\frac{\alpha}{2}\right)}\right);$$

2. $(1-\alpha)\cdot 100\,\%$-ný LIS pre σ^2, ak nepoznáme μ:
$$\left(\frac{(n-1)S_1^2}{\chi^2_{n-1}\left(1-\alpha\right)}, \quad \infty\right);$$

3. $(1-\alpha)\cdot 100\,\%$-ný PIS pre σ^2, ak nepoznáme μ:
$$\left(0; \quad \frac{(n-1)S_1^2}{\chi^2_{n-1}\left(\alpha\right)}\right).$$

Príklad 3:
V laboratóriu sme testovali prístroj na meranie teploty vody. Voda mala presne $10^\circ\ C$. Urobili sme 20 nezávislých meraní a vypočítali sme, že $S_0^2 = 1,21$. Vypočítajte $95\,\%$-ný interval PIS pre rozptyl meraní používaného meracieho prístroja.

Riešenie pr. 3: Pretože $S_0^2 = 1,21$, $\alpha = 0,05$ a $\chi^2_{20}(0,05) = 10,851$ (hodnota z tabuľky pre χ^2_{20}), tak PIS je
$$\left(0; \quad \frac{20\cdot 1,21}{10,851}\right) = (0; \quad 2,2302)\,.$$

To znamená, že
$$P\left(\sigma^2 \in (0; \quad 2,2302)\right) = 0,95,$$
teda na $95\,\%$ rozptyly meraní daným prístrojom neprekročia hodnotu $2,2302$.

Príklad 4:
Potrebovali sme zistiť rozptyl teploty vody. Urobili sme 20 nezávislých meraní a vypočítali sme, že priemerná teplota je $\overline{x} = 20,5^\circ\ C$ a $S_1^2 = 2,24$. Vypočítajte $95\,\%$-ný interval PIS pre rozptyl teploty vody.

Riešenie pr. 4: Pretože $S_1^2 = 2,24$, $\overline{x} = 20,5\,^\circ\ C$, $\alpha = 0,05$ a $\chi^2_{19}(0,05) = 10,117$ (hodnota tabuľky pre χ^2_{19}), tak PIS je

$$\left(0; \quad \frac{19 \cdot 2,24}{10,117}\right) = (0; \quad 4,2068)\,.$$

To znamená, že

$$P\left(\sigma^2 \in (0; \quad 4,2068)\right) = 0,95.$$

Na 95 % rozptyly meraní teploty vody daným prístrojom neprekročia hodnotu 4,2068

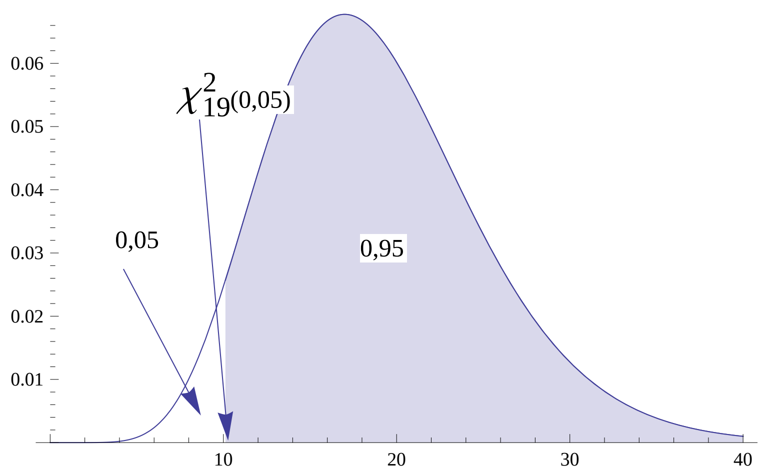

Fig. 2: 0,05-kvantil rozdelenia $\chi^2(19)$

1.6 Testovanie štatistických hypotéz

Jedným z cieľov matematickej štatistiky je overenie tvrdenia na základe nameraných resp. napozorvaných údajov. Táto časť matematickej štatistiky sa nazýva testovanie štatistických hypotéz. Testovanie štatistických hypotéz patrí medzi základné disciplíny matematickej štatistiky. Cieľom tejto časti je ukázať niektoré základné štatistické testy.

Najskôr sa budeme venovať testovaniu všeobecne.

Definícia 1:
Pod štatistickou hypotézou rozumieme výrok alebo tvrdenie, ktorého platnosť budeme overovať na základe postupov, ktoré boli odvodené pomocou matematickej štatistiky.

Štatistické hypotézy rozdeľujeme na:

(H0) *Nulová hypotéza* – tvrdenie, ktorého platnosť overujeme – označujeme H_0.

(H1) *Alternatívna hypotéza* – opak tvrdenia H_0 – označujeme H_1.

Rozoznávame dva základné druhy alternatívnej hypotézy:

- *Jednoduchá alternatíva* – rozhodujeme sa medzi dvomi alternatívami. Napríklad:
  H_0: Teplota je presne $10^\circ\ C$.
  H_1: Teplota je presne $14^\circ\ C$.

- *Zložená alternatíva* Napríklad:

(a) H_0: Teplota je presne $10^\circ\ C$.
H_1: Teplota nie je $10^\circ\ C$. – Obojstranná alternatíva.

(b1) H_0: Teplota je presne $10^\circ\ C$.
H_1: Teplota je väčšia ako $10^\circ\ C$. – Jednostranná alternatíva.

(b2) H_0: Teplota je presne $10^\circ\ C$.
H_1: Teplota je menšia ako $10^\circ\ C$. – Jednostranná alternatíva.

Cieľ testovania: Cieľom štatistického testovania je rozhodnúť o platnosti resp. neplatnosti nulovej hypotézy na základe nameraných resp. napozorovaných údajov.

Voľba štatistickej metódy: Voľba metódy pomocou ktorej budeme postupovať závisí od charakteru dát, od postupu pri získavaní údajov, od formulácie nulovej a alternatívnej hypotézy a pod. Na riešenie daného problému možno použiť viacero štatistických testov. Napriek dodržaniu všetkých predpokladov štatistického testu sa môže stať, že dostaneme rôzne výsledky.

Rozhodovanie na základe výsledkov testu: Správna interpretácia výsledkov štatistického testu je veľmi dôležitá. Pri rozhodovaní sa postupuje nasledovne:

- Zvolíme si číslo $\alpha \in (0, 1)$, hladinu významnosti testu, pravdepodobnosť, že sme zamietli platnú nulovú hypotézu, teda rozhodli sme sa nesprávne;

- Množinu všetkých možných výsledkov testovacieho kritéria za predpokladu, že platí hypotéza H_0, rozdelíme na dve časti H_0 a H_1 tak, aby v prípade platnosti nulovej hypotézy výsledok testu s pravdepodobnosťou $1 - \alpha$ padol do množiny H_0. Množina H_1 sa nazýva *kritická oblasť*. (Množina a hypotéza sa označujú rovnako);

- Ak hodnota testovacej štatistiky padne do množiny $H0$, tak nulovú hypotézu H_0 nezamietame, v opačnom prípade H_0 zamietame. Treba si uvedomiť, že nerozhodujeme o platnosti hypotézy H_1 a ani nerozhodujeme o prijatí H_0. Záver je, že zamietame, alebo nezamietame H_0. V konečnom dôsledku, ak nezamietame H_0, správame sa tak, akoby H_0 platila.

Chyby pri štatistickom rozhodovaní: Pri štastickom rozhodovaní sa dopúšť ame dvoch druhov chýb:

- *Chyba prvého druhu* (*hladina významnosti testu*) (α). Zamietneme H_0 napriek tomu, že H_0 platí. Túto chybu si volíme.

- *Chyba druhého druhu* (β) je chyba, ktorej sa dopúšťame, ak nezamietneme H_0 v prípade, že H_0 neplatí. Hodnota $1 - \beta$ sa nazýva sila testu.

Medzi oboma chybami je nasledujúci vzťah:

ak zmenšíme α, tak sa zväčší β

ak zväčšíme α, tak sa zmenší β.

Na obrázku 1.6.1 je znázornená situácia, keď rozhodujeme na základe experimentu o strednej hodnote n.p. $X \sim N(\mu, 1)$. Predpokladáme, že stredná

Rozhodnutie	Skutočnosť H_0 platí	Skutočnosť H_0 neplatí
Zamietame H_0	α	nedopúšťame sa chyby s pravdepodobnosťou (istotou) $1-\alpha$
Nezamietame H_0	nedopúšťame sa chyby s pravdepodobnosťou (istotou) $1-\beta$	β

Tab. 1: α – chyba prvého druhu, β– chyba druhého druhu

hodnota nie je menšia ako 10. Postavíme nulovú a alternatívnu hypotézu:

$$H_0: \quad \mu = 10 \qquad \text{a} \qquad H_1: \quad \mu < 10.$$

Na základe výsledku testu nezamietame H_0, ale H_0 neplatí (v skutočnosti $\mu = 8$).

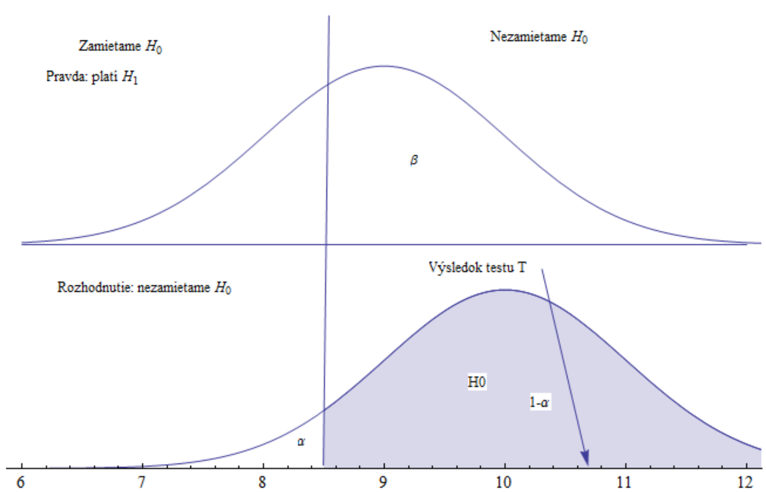

Fig. 3: Chyby pri štatistickom rozhodovaní.

p-hodnota je pravdepodobnosť, že na základe napozorovaných hodnôt hypotézu H_0 zamietame, napriek tomu, že platí. Nech S je testovacia štatistika a s^* je hodnota štatistiky S, vypočítaná na základe napozorovaných hodnôt. Potom p-hodnota p je definovaná
pre pravostranný test:

$$p = P(S \leq s^* | H_0)$$

pre ľavostranný test:

$$p = P(S \geq s^* | H_0)$$

pre obojstranný test:

$$p = 2 \cdot \min\{P(S \leq s^* | H_0), P(S \geq s^* | H_0)\}.$$

To znamená, že ak nami zvolená hladina významnosti testu α je menšia, alebo sa rovná p ($p \leq \alpha$), tak hypotézu H_0 na hladine významnosti α zamietame, v opačnom prípade H_0 na hladine významnosti α nezamietame. To znamená, že p-hodnota je vlastne odhad chyby prvého druhu, ktorá bola vypočítaná na základe napozorovaných údajov. V dnešnej dobe sa uprednostňuje na rozhodovanie výpočet p-hodnoty pred rozhodovaním na základe subjektívne zvolenej hladiny významnosti testu.

Druhy štatistických testov: V matematickej štatistike bolo odvodených veľmi veľa štatistických testov pre rôzne situácie, keď musíme prijať nejaké rozhodnutie na základe pozorovania resp. merania. Rozdeľujeme ich do dvoch základných skupín: parametrické a neparametrické testy. Parametrickými testami rozhodujeme o základných parametroch (strednej hodnote, rozptyle, štandardnej odchýlke kvantitatívneho znaku, resp. podiele kvalitatívneho znaku). Okrem toho sú testy, ktoré sa zaoberajú len jedným dátovým súborom (jednovýberové), alebo viacerými dátovými súbormi (dvoj a viac výberové testy). Iná klasifikácia testov je napr. podľa metódy, ktorá sa použila pri stanovení testovacieho kritéria atď. Existuje aj množstvo testov, ktoré sa zaoberajú vzťahmi medzi sledovanými objektmi. Dôležitými testami sú aj testy dobrej zhody, ktoré testujú, či rozdelenie nameraných údajov zodpovedá vhodnému teoretickému pravdepodobnostnému rozdeleniu.

1.7 Testy pre parametre normálneho rozdelenia

Pre testy o parametroch platí, že množina H_0 sa zhoduje s intervalom spoľahlivosti pre odhadovaný parameter.

Testy pre strednú hodnotu μ

Nech $X_1, ..., X_n$ je náhodný výber z $N(\mu, \sigma^2)$. Chceme si overiť predpoklad o stredných hodnotách: $\mu = \mu_0$, kde μ_0 je teoretická stredná hodnota. Postavíme si nulovú (H_0) a alternatívnu hypotézu (H_1):

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0 \quad (\mu > \mu_0) \quad (\mu < \mu_0)$$

a zvolíme si hladinu významosti α. Ak platí H_0, potom

a) v prípade, že σ^2 poznáme

$$U = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \sim N(0,1);$$

b) v prípade, že σ^2 nepoznáme

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S_1} \sim t(n-1),$$

kde n je počet meraní a

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

je nevychýlený odhad rozptylu. V oboch prípadoch (a) a (b) ak μ_0 padne do príslušného intervalu spoľahlivosti, hypotézu H_0 nezamietame, v opačnom prípade zamietame.

Testy pre rozptyl

Nech $X_1, ..., X_n$ je náhodný výber z $N(\mu, \sigma^2)$. Chceme si overiť' predpoklad o rozptyloch: $\sigma^2 = \sigma_0^2$, kde σ_0^2 je teoretický rozptyl. Postavíme si nulovú (H_0) a alternatívnu hypotézu (H_1):

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 \neq \sigma_0^2 \quad (\sigma^2 > \sigma_0^2) \quad (\sigma^2 < \sigma_0^2)$$

a zvolíme si hladinu významosti α. Ak platí H_0, potom a2) v prípade, že μ poznáme, tak štatistika

$$Q_0 = \frac{nS_0^2}{\sigma_0^2} \sim \chi^2(n),$$

kde n je počet meraní, S_0^2 je odhad rozptylu vypočítaný podľa vzorca:

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

b2) v prípade, že μ nepoznáme, tak štatistika

$$Q = \frac{(n-1)S_1^2}{\sigma_0^2} \sim \chi^2(n-1),$$

kde n je počet meraní, S_1^2 je odhad rozptylu vypočítaný podľa vzorca:

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Obojstranný test

Postup:

a2) Vypočítame S_0^2 a hodnotu štatistiky Q_0. V tabuľkách pre $\chi^2(n)$ nájdeme kvantily $A = \chi^2_{(n)}\left(\frac{\alpha}{2}\right)$ a $B = \chi^2_{(n)}\left(1 - \frac{\alpha}{2}\right)$. Ak $Q_0 \in (A, B)$, potom H_0 nezamietame na hladine významnosti α.

b2) Vypočítame S_1^2 a hodnotu štatistiky Q. V tabuľkách pre $\chi^2(n-1)$ nájdeme kvantily $A = \chi^2_{(n-1)}\left(\frac{\alpha}{2}\right)$ a $B = \chi^2_{(n-1)}\left(1 - \frac{\alpha}{2}\right)$. Ak $Q \in (A, B)$, potom H_0 nezamietame na hladine významnosti α.

V oboch prípadoch dostaneme ten istý výsledok, ak použijeme obojstranné intervaly spoľahlivosti pre σ^2 (OIS). Rovnako budeme postupovať aj pri jednostranných hypotézach.

Jednostranné testy

Postup:

a2) Vypočítame S_0^2. V prípade

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 > \sigma_0^2,$$

vlastne testujeme, či teoretická hodnota σ_0^2 je výrazne menšia ako odhadnutá hodnota S_0^2. Vypočítame

$$LIS = \left(\frac{nS_0^2}{\chi_n^2(1-\alpha)}, \infty \right).$$

Ak $\sigma_0^2 \notin LIS$, potom H_0 zamietame.

Analogicky v prípade

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 < \sigma_0^2,$$

vlastne testujeme, či teoretická hodnota σ_0^2 je výrazne väčšia, ako odhadnutá hodnota S_0^2. Vypočítame

$$PIS = \left(0, \frac{nS_0^2}{\chi_n^2(\alpha)} \right).$$

Ak $\sigma_0^2 \notin PIS$, potom H_0 zamietame.

b2) Vypočítame S_1^2. V prípade

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 > \sigma_0^2,$$

vlastne testujeme, či teoretická hodnota σ_0^2 je výrazne menšia, ako odhadnutá hodnota S_1^2. Vypočítame

$$LIS = \left(\frac{nS_1^2}{\chi_{n-1}^2(1-\alpha)}, \infty \right).$$

Ak $\sigma_0^2 \notin LIS$, potom H_0 zamietame.

Analogicky v prípade

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 < \sigma_0^2,$$

vlastne testujeme, či teoretická hodnota σ_0^2 je výrazne väčšia, ako odhadnutá hodnota S_1^2. Vypočítame

$$PIS = \left(0, \frac{nS_1^2}{\chi_{n-1}^2(\alpha)} \right).$$

Ak $\sigma_0^2 \notin PIS$, potom H_0 zamietame.

Jednovýberový Wilcoxonov test

Nech

$$X_1, \ldots, X_n$$

je náhodný výber z nejakého spojitého rozdelenia, ktorého funkcia hustoty je symetrická okolo mediánu

$$f(t-a) = f(t+a)$$

a existuje konečná stredná hodnota $E(X_i) = a$. Nech

$$X = \{x_1, \ldots, x_n\}$$

je jeho realizácia. Označme symbolom $\widetilde{X}$ výberový medián.

Obojstranný test:

$$H_0 : \widetilde{X} = x_0 \qquad H_1 : \widetilde{X} \neq x_0$$

Jednostranné testy:

$$\text{(a)} \quad H_0 : \widetilde{X} \leq x_0 \qquad H_1 : \widetilde{X} > x_0$$

$$\text{(b)} \quad H_0 : \widetilde{X} \geq x_0 \qquad H_1 : \widetilde{X} < x_0$$

Predpoklad: Pre všetky hodnoty $x \in X$ platí, že $x \neq x_0$. Ak by sa stalo, že existuje hodnota $x \in X$ s vlastnosťou, že $x = x_0$, potom hodnotu x zo súboru vyúčime a budeme pracovať so súborom o rozsahu $n-1$.

Postup:

1. Vytvoríme novú náhodnú premennú $Y_i = X_i - x_0$ dostaneme ich realizáciu $y_i = x_i - x_0$.

2. Usporiadame podľa veľkosti hodnoty $|y_i|$ a jednotlivým realizáciam priradíme poradia premennej $|y_i|$ nasledovne: R_i^+ ak $y_i > 0$ a R_i^- ak $y_i < 0$.

3. Vypočítame
$$S^+ = \sum_{y_i > 0} R_i^+, \qquad S^- = \sum_{y_i < 0} R_i^-.$$

ľahko sa dá overiť, že

$$S^+ + S^- = \frac{n(n+1)}{2} = \sum_{i=1}^{n} i.$$

Ideálny stav je, ak

$$S^+ = S^- = \frac{n(n+1)}{4}.$$

4. V tabuľkách pre Wilcoxonov test (Tab. **??**) nájdeme kvantil $w_n(\alpha)$. Ak

$$min(S^+, S^-) \leq w_n(\alpha),$$

potom hypotézu H_0 na hladine významnosti α zamietame.

5. V roku 1967 Hájek a Šidák dokázali, že S^+ má asymptoticky normálne rozdelenie, a teda pre veľké n môžeme namiesto špeciálnych tabuliek použiť tabuľky pre $N(0,1)$. Pretože

$$E(S^+) = \frac{n(n+1)}{4} \qquad D(S^+) = \frac{n(n+1)(2n+1)}{24}$$

potom náhodná premenná

$$U = \frac{S^+ - E(S^+)}{\sqrt{D(S^+)}} \sim N(0,1).$$

Nech $\Phi^{-1}(\alpha)$ je kvantil z $N(0,1)$. V prípade obojstranného testu, ak

$$|U| \leq \phi(\alpha)$$

potom H_0 zamietame na hladine významnosti α. V prípade jednostranných testov, ak

(a) $U \leq \phi(1-\alpha)$

(b) $U \leq \phi(\alpha)$

potom H_0 zamietame na hladine významnosti α.

Poznámka 2:
Hypotéza H_0 sa zamietne aj vtedy, ak medián je naozaj x_0, ale funkcia hustoty je výrazne nesymetrická.

Príklad 5:
Chceli sme zistiť, ako ľudia odhadujú jednu minútu. Pokus sme urobili na desiatich respodentoch $n = 10$. Ideme riešiť úlohu:

$$H_0 : \widetilde{X} = 60 \qquad H_1 : \widetilde{X} \neq 60$$

Výsledky sú v nasledujúcej tabuľke (v sekundách). Výsleky sú usporiadané podľa veľkosti.

Respondent	1	2	3	4	5	6	7	8	9	10
odhad x_i	45	48	50	51	53	55	56	58	63	68
$y_i = x_1 - 60$	-15	-12	-10	-9	-7	-5	-4	-2	3	8
$\|y_i\|$	15	12	10	9	7	5	4	2	3	8
Poradie	10	9	8	7	5	4	3	1	2	6
R_i^+									2	6

Teda $S^+ = 2 + 6 = 8$. ľahko sa dá overiť, že $S^- = 48$. Tabuľková hodnota $w_{10}(0,05) = 8$. Pretože

$$min(7, 48) = 7 \leq 8,$$

tak hypotézu H_0 zamietame.

Aproximácia normálnym rozdelením: Pretože

$$E(S^+) = \frac{10.11}{4} = 27,5 \qquad D(S^+) = \frac{10.11.21}{24} = 96,25,$$

tak štatistika

$$U = \frac{7 - 27,5}{\sqrt{96,25}} \cong -2,08955.$$

V tabuľkách pre $N(0, 1)$ nájdeme $\phi(0,975) \cong 1,96$.
Záver:

$$|U| = 2,08955 \geq 1,96,$$

a teda H_0 zamietame.

Znamienkový test pre medián

Nech

$$X_1, \ldots, X_n$$

je náhodný výber z nejakého spojitého rozdelenia a

$$X = \{x_1, \ldots, x_n\}$$

je jeho realizácia. Označme symbolom $\widetilde{X}$ označme teoretický medián. To znamená, že

$$P(X_i < \widetilde{X}) = P(X_i > \widetilde{X}) = 0,5$$

pre každé $i = 1, \ldots, n$. Predpoklad: Všetky hodnoty $x \in X$ platí, že $x \neq x_0$. Ak by sa stalo, že existuje hodnota $x \in X$, potom hodnotu x zo súboru vyúčime a budeme pracovať so súborom o rozsahu $n - 1$.

Ideme riešiť úlohu:

$$H_0 : \widetilde{X} = x_0 \qquad H_1 : \widetilde{X} \neq x_0$$

Postup:

1. Vytvoríme novú náhodnú premennú

$$Y_i \begin{cases} = 0 & x_i < x_0 \\ & \\ = 1 & x_i > x_0. \end{cases}$$

2. Označíme $Y = \sum_i Y_i$. Pretože náhodná premenná $Y \sim Bi(n, \frac{1}{2})$, tak $E(Y) = \frac{n}{2}$ a $D(Y) = \frac{n}{4}$. Označme

$$U = \frac{Y - E(Y)}{\sqrt{D(Y)}}.$$

Nech $\Phi^{-1}(\alpha)$ je kvantil z $N(0,1)$. Ak

$$|U| \geq \phi(\alpha),$$

potom H_0 zamietame.

Príklad 6:
Chceli sme zistiť, ako ľudia odhadujú jednu minútu. Pokus sme urobili na desiatich respodentoch $n = 10$. Ideme riešiť úlohu:

$$H_0 : \widetilde{X} = 60 \qquad H_1 : \widetilde{X} \neq 60$$

Výsledky sú v nasledujúcej tabuľke (v sekundách). Výsleky sú usporiadané podľa veľkosti.

Respondent	1	2	3	4	5	6	7	8	9	10
x_i	45	48	50	51	53	55	56	58	63	68
y_i	0	0	0	0	0	0	0	0	1	1

Teda $Y = 2$.

$$E(Y) = \frac{10}{2} = 5 \qquad D(Y) = \frac{10}{4} = 2,5$$

tak štatistika

$$U = \frac{2-5}{\sqrt{2,5}} \cong -1,89737.$$

V tabuľkách pre $N(0,1)$ nájdeme $\phi(0,975) \cong 1,96$.
Záver:

$$|U| = 1,89737 \leq 1,96$$

a teda H_0 nezamietame. Aj keď sa to nezdá, ale táto vzorka nezamieta, že medián by mohol byť 60 sekúnd. Ak však počítame presne podľa binomického rozdelenia, tak rýchlo zistíme, že

$$P(Y \leq 2) = 0,0546875 > 0,025$$

a

$$P(Y \leq 1) = 0,010742188 < 0,025.$$

Teda hypotézu by sme zamietli na hladine významnosti $\alpha = 0,05$ ak $Y \leq 1$.

Test pre binomické rozdelenie

Nech $Y \sim Bi(n,p)$, pričom p nepoznáme. Na hladine významnosti α budeme testovať:

$$H_0 : p = p_0 \qquad H_1 : p \neq p_0,$$

kde p_0 je predpokladaná hodnota.

Pre dané α vypočítame najväčšie celé čísla k_i, kde $i = 1,2$ aby platilo

$$P(Y \leq k_1) = \sum_{i=0}^{k_1} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \frac{\alpha}{2}$$

$$P(Y \geq k_2) = \sum_{i=k_2}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \frac{\alpha}{2}.$$

Hypotézu H_0 zamietneme, ak $Y \leq k_1$ alebo $Y \geq k_2$.

Príklad 7:
Podľa výsledkov celoslovenského prieskumu sa zistilo, že $69,2\,\%$ obyvateľov pije na raňajky kávu. Náhodne sme vybrali 62 vysokoškolských učiteľov a z nich 37 uviedlo, že na raňajky pije kávu. Na hladine významnosti $\alpha = 0,05$ chceme zistiť, či podiel učiteľov pijúcich na raňajke kávu je taký istý, ako v celej dospelej populácii.

Ideme teda na hladine významnosti $0,05$ testovať:

$$H_0 : p = 0,692 \qquad H_1 : p \neq 0,692,$$

Ak platí H_0, potom náhodná premenná Y počet učiteľov pijúcich kávu, má binomické rozdelenie $Y \sim Bi(n,p)$, kde $n = 62$ a $p = 0,692$.

$$P(Y \leq k_1) = \sum_{i=0}^{k_1} \binom{62}{i} 0,692^i (1-0,692)^{n-i} \leq 0,025$$

Z binomického rozdelenia ľahko zistíme. že

$$P(Y \leq 35) = 0,023096247 < 0,025$$

a

$$P(Y \leq 36) = 0,041626452 > 0,025.$$

Z toho vyplýva, že $k_1 = 35$. Riešime úlohu: nájsť také najmenšie k_2, aby platilo

$$P(Y \geq k_2) = \sum_{i=k_2}^{n} \binom{62}{i} 0,692^i (1-0,692)^{n-i} \leq 0,025.$$

Pretože

$$P(Y \geq k_2) = 1 - P(Y < k_2) = 1 - P(Y \leq k_2 - 1)$$

a z binomického rozdelenia zistíme, že

$$1 - P(Y \leq 49) = 1 - 0,969018281 = 0,030981719 > 0,025$$

$$1 - P(Y \leq 50) = 1 - 0,0,984954961 = 0,015045039 < 0,025,$$

tak $k_2 = 51$.
Záver:
Ak $Y \in \{36, 37, \ldots, 50\}$ tak H_0 nezamietame. V našom prípade teda môžeme skonštatovať, že na základe prieskumu učitelia v prípade pitia kávy na raňajky sa správajú rovnako, ako celá slovenská populácia.

1.8 Test nezávislosti pre kategoriálne dáta

Sledujeme dva znaky **A** a **B**. Znak **A** má k úrovní a znak **B** má n úrovní. Pričom máme N pozorovaní. Teda dáta sú (x_t, y_t) $t = 1, \ldots, N$. Označme početnosti $n_{i.}$ početnosť i-tej úrovne znaku **A**, $n_{.j}$ početnosť j-tej úrovne znaku **B** a n_{ij} početnosť i-tej úrovne znaku **A** a zároveň početnosť j-tej úrovne znaku **B**. Chceme zistiť či sú **A** a **B** nezávislé.

$$H_0 : \text{znaky } \mathbf{A} \text{ a } \mathbf{B} \text{ sú nezávislé} \qquad H_1 : H_0 \text{ neplatí}$$

$$G = \sum_{i-1}^{k} \sum_{j=1}^{n} \frac{(n_{ij} - n'_{ij})^2}{N} \sim \chi^2_{(n-1)(k-1)},$$

kde

$$n'_{ij} = \frac{n_{i.}.n_{.j}}{N}.$$

Príklad 8:
Prieskum pozastával z dvoch otázok respondentom:

A: Súhlasíte, aby olympijské hry boli vo Vysokých Tatrách?

- A_1 áno
- A_2 neviem
- A_3 nie

B: Porovnajte úroveň cestovania vlakom v SR a ČR

- B_1 oveľa lepšie v SR
- B_2 oveľa lepšie v ČR
- B_3 je to rovnaké
- B_4 neviem porovnať

A/**B**	$\mathbf{B}_1$	$\mathbf{B}_2$	$\mathbf{B}_3$	$\mathbf{B}_4$	Marginálne **A**
$\mathbf{A}_1$	10	10	4	5	29
$\mathbf{A}_2$	12	5	8	11	36
$\mathbf{A}_3$	10	7	8	10	35
Marginálne **B**	32	22	20	26	100

Tab. 2: Výsledky prieskumu

Opýtali sme sa 100 respondentov. Výsledky sú v tabuľke 2:

Teda $k = 3$, $n = 4$ a $N = 100$.

A	$n_{1.}$	$n_{2.}$	$n_{3.}$
	29	36	35

B	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$
	32	22	20	26

Očakávané pri nezávislosti:

n'_{ij}	32	22	20	26
29	$9,28$	$6,38$	$5,8$	$7,54$
36	$11,52$	$7,92$	$7,2$	$9,36$
35	$11,2$	$7,7$	7	$9,1$

Napríklad $n'_{11} = \frac{29.32}{100} = 9,28$

$$\begin{aligned} G =& \frac{(n_{11} - n'_{11})^2 + \cdots + (n_{34} - n'_{34})^2}{100} \\ =& \frac{(10 - 9,28)^2 + \cdots + (10 - 9,1)^2}{100} \\ \doteq& 0,39141 \end{aligned}$$

Nájdeme kvantil $\chi^2_{2.3}(0,95)$ (ak $\alpha = 0,05$). Pretože

$$0,39141 \doteq G < \chi^2_6(0,95)^{-1} \doteq 12,5916$$

tak H_0 nezamietame. V tabuľkách nájdeme p-hodnotu:

$$1 - \chi^2_6(0,391408) \doteq 0,9989.$$

Záver: Na hladine významnosti $0,95$ môzeme povedať, že odpovede na otázky **A** a **B** sú nezávislé. Pretože p hodnota je veľmi vysoká, skoro so $100\,\%$-tnou istotou môžeme povedať, že nie je žiaden vzťah medzi tým, či respondent súhlasí s OH vo VT a jeho hodnotením porovnania cestovania po železnici v SR a ČR.

2 Testy dobrej zhody

Pri riešení úloh často predpokladáme určitý typ pravdepodobnostného rozdelenia základného súboru. Testami dobrej zhody je možné overiť do akej miery zodpovedá empirické rozdelenie pravdepodobnosti teoretickému pravdepodobnostnému rozdeleniu. V tejto časti sú uvedené dva základné testy: χ^2-test dobrej zhody a Kolmogorov-Smirnov test dobrej zhody.

2.1 χ^2-test dobrej zhody

Nech X je náhodná premenná a $X_1, \ldots, X_n$ je jej náhodný výber a $C_1, \ldots, C_k$ sú disjunktné množiny s vlastnosťou, že ľubovoľná realizácia náhodnej premennej X padne do niektorej množiny C_j, $j = 1, \ldots, k$. Označme

$$p_j = P(X \in C_j),$$

teoretickú pravdepodobnosť a n_j počet meraní (realizácií), ktoré padnú do množiny C_j, $j = 1, \ldots, k$. Našim cieľom je overiť predpoklad, že náhodná premenná X má dané teoretické rozdelenie pravdepodobnosti. To znamená

$$\begin{aligned} H_0 &: P(X \in C_j) = p_j, \text{ pre každé } j = 1, \ldots, k \\ H_1 &: \text{existuje aspoň jedno } j \in \{1, \ldots k\} \text{ s vlastnosťou, že } P(X_i \in C_j) \neq p_j \end{aligned}$$

Ak označíme Y_j náhodnú premennú, ktorá predstavuje počet meraní, ktoré pri n-násobnom opakovaní pokusu padnú do C_j, potom $Y_j \sim Bi(n, p_j)$, pre $j = 1, \ldots, k$. Z vlastností binomického rozdelenia vieme, že

$$E(Y_j) = np_j \text{ a } D(Y_j) = np_j(1 - p_j).$$

Označme

$$U_j = \frac{Y_j - np_j}{\sqrt{np_j(1 - p_j)}},$$

potom $E(U_j) = 0$ a $D(U_j) = 1$. Pre veľké n náhodná premenná

$$U = \sum_{j=1}^{k} U_j^2 = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j(1-p_j)}$$

má asymptoticky $\chi^2(k)$ [**?**, **?**]. Pretože

$$U = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j(1-p_j)} = \sum_{j-1}^{k} \frac{(Y_j - np_j)^2((1-p_j)+p_j)}{np_j(1-p_j)}$$

$$U = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{n(1-p_j)} + \sum_{j-1}^{k} \frac{(Y_j - np_j)^2}{np_j},$$

Potom platí [**?**, **?**]

$$U = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j(1-p_j)} \sim \chi^2(k) \tag{13}$$

$$Q = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{n(1-p_j)} \sim \chi^2(1) \tag{14}$$

$$C = \sum_{j-1}^{k} \frac{(Y_j - np_j)^2}{np_j} \sim \chi^2(k-1) \tag{15}$$

Ako testovacia štatistika v χ^2-teste sa používa náhodná premenná C. Všimnime si, že v prípade ideálnej zhody výraz v čitateli sa rovná 0. To znamená, že čím je štatistika C bližšie k nule, tým viac dáta zodpovedajú teoretickému rozdeleniu.

Postup – diskrétne dáta

1. Urobíme frekvenčnú tabuľku z napozorovaných hodnôt.

2. Stanovíme hypotézy H_0, H_1 a hladinu významnosti α. Vypočítame teoretické pravdepodobnosti p_j, ktoré zodpovedajú teoretickému rozdeleniu pravdepodobnosti z hypotézy H_0.

3. Vypočítame testovaciu štatistiku C.
4. V štatistických tabuľkách si nájdeme kvantil z χ^2 rozdelenia pre zodpovedajúci počet stupňov voľnosti $(\chi^2_{k-1})(1-\alpha)$ resp. p-hodnotu $p = \chi^2_{k-1}(C)$.
5. Rozhodnutie: Ak $C \geq (\chi^2_{k-1})(1-\alpha)$ resp. $p \leq \alpha$, potom hypotézu H_0 na hladine významnosti α zamietame.

Príklad 9:
Testujeme, či hracia kocka je regulárna. Urobili sme 600 pokusov. Hypotéza H_0 je, že kocka je v poriadku, oproti alternatíve H_1 – kocka je falošná. To znamená, že predpokladáme

$$P(1) = P(2) = \dots P(6) = \frac{1}{6}.$$

Pretože $n = 600$ a $p_j = 1/6$, potom $np_j = 100$ pre $j = 1, 2, \dots 6$. V nasledujúcej tabuľke sú uvedené početnosti.

j	1	2	3	4	5	6
n_j	104	92	109	93	101	101
Δ_j	4	-8	9	-7	1	1
$\frac{\Delta_j^2}{100}$	0,16	0,64	0,81	0,49	0,01	0,01

kde $\Delta_j = n_j - np_j = n_j - 100$. Štatistika C je súčet hodnôt v poslednom riadku tabuľky:

$$C = 2,12 < 11,07049775 = (\chi^2_5)^{-1}(0,95),$$

preto H_0 na hladine $\alpha = 0,05$ nezamietame. Všimnime si, že p-hodnota je pomerne vysoká $p = \chi^2_5(2,12) = 0,832$.

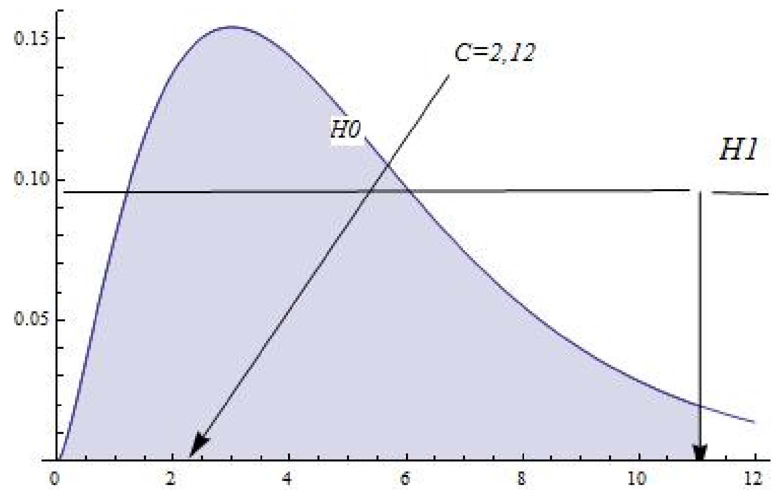

Fig. 4: Kritická oblasť

Príklad 10:
Pokus pozostával z piatich nezávislých hodov mincou a náhodná premenná X je počet padnutých znakov. Urobili sme 30 pokusov. Výsledky sú uvedené vo frekvenčnej tabuľke. Ak predpokladáme, že $X \sim Bi(5; 0,5)$, to znamená, že

$$\mathbf{H_0} : X \sim Bi(5; 0,5) \qquad \mathbf{H_1} : H_0 \textit{ neplatí}$$

Ak H_0 platí, potom teoretickú pravdepodobnosť vypočítame podľa vzorca pre binomické rozdelenie pravdepodobnosti:

$$p_j = p_{i+1} = \binom{5}{i} 0,5^5$$

pre $i = 0, 1, \ldots, 5$. Hodnoty sú v nasledujúcej tabuľke:

j	1	2	3	4	5	6
X	0	1	2	3	4	5
n_j	1	5	5	10	8	1
p_j	0,03125	0,15625	0,3125	0,3125	0,15625	0,03125
$30.p_j$	0,9375	4,6875	9,375	9,375	4,6875	0,9375
$\Delta(x_j)$	-0,0625	-0,3125	4,375	-0,625	-3,3125	-0,0625
$\frac{\Delta(x_j)^2}{30.p_j}$	0,0042	0,02083	2,0417	0,0417	2,3408	0,00417

Štatistika C je súčet hodnôt v poslednom riadku tabuľky $C = 4,4533$. Kvantil (kritickú hodnotu) hľadáme v tabuľkách pre χ^2 rozdelenie s 5 stupňami voľnosti pre $\alpha = 0,05$: $(\chi^2_5)^{-1}(0,95) = 11,07$. Pretože $C = 4,4533 < 11,07$ tak H_0 na hladine $\alpha = 0,05$ nezamietame.

Ak náhodná premenná je spojitá, tak najskôr musíme stanoviť triedne intervaly, do ktorých budeme namerané, resp. napozorované hodnoty zadeľovať. Väčšinou volíme ekvidistatnú dĺžku intervalov $(c_j, c_j+d]$, $j = 1, \dots, k-2$, $d > 0$.

Postup – spojité dáta

1. Urobíme základnú popisnú štatistiku, roztriedime dáta do triednych intervalov, urobíme frekvenčnú tabuľku.
2. Stanovíme hypotézy H_0, H_1 a hladinu významnosti α.
3. Vypočítame teoretické pravdepodobnosti p_j, ktoré zodpovedajú teoretickému rozdeleniu pravdepodobnosti z hypotézy H_0.
4. Vypočítame testovaciu štatistiku C.
5. V štatistických tabuľkách si nájdeme kvantil z χ^2 rozdelenia pre zodpovedajúci počet stupňov voľnosti ν resp. p-hodnotu p.
6. Rozhodnutie: Ak $C \geq \nu$ resp. $p \leq \alpha$ hypotézu H_0 na hladine významnosti α zamietame.

Príklad 11:
Merali sme teplotu vody v podzemnom prameni. Merania sme nezávisle po sebe zopakovali 100-krát. Predpokladáme, že údaje pochádzajú z $N(12; 1)$.

Riešenie pr. 11: Sformulujeme štatistické hypotézy

$$\mathbf{H_0} : X \sim N(12;1) \qquad \mathbf{H_1} : H_0 \text{ neplatí}$$

V naledujúcej tabuľke je uvedený výsledok intervalového triedenia:

Štatistika C je súčet posledného riadku v tabuľke 3

$$C = 3,448 < 12,5916 = (\chi^2_6)^{-1}(0.95),$$

preto nulovú hypotézu na hladine významnosti $\alpha = 0,05$ nezamietame .

j	1	2	3	4	5	6	7
tr. int.	≤ 10	$(10]$	$(10,5]$	$(11]$	$(11,5]$	$(12]$	$> 12,5$
n_j	3	7	9	16	18	14	33
p_j	0,023	0,044	0,092	0,15	0,191	0,192	0,308
$100.p_j$	$2,3$	$4,4$	$9,2$	15	$19,1$	$19,2$	$30,8$
Δ_j	0,7	2,6	-0,2	1	-1,1	-5,2	2,2
Δ_j^2	0,49	6,76	0,04	1	1,21	27,04	4,84
$\frac{\Delta_j^2}{100p_j}$	0,213	1,536	0,004	0,0667	0,0634	1,4083	0,1571

Tab. 3: $(a] = (a,\, a + 0,5]$ a $\Delta_j = n_j - 100.p_j$.

Poznámka 3:
V prípade, že niektorý z parametrov teoretického rozdelenia nahrádzame jeho odhadom, potom znižujeme stupne voľnosti o jeden pre každý parameter. Napríklad, ak v Príklade 11 odhadneme strednú hodnotu a smerodajnú odchýlku z nameraných hodnôt, tak kvantil hľadáme v tabuľkách $\chi^2_{k-1-2} = \chi^2_4$.

2.2 K-S test dobrej zhody

Kolmogorov-Smirnov (K-S) test [?] je na výpočet veľmi jednoduchý a dáva dobré výsledky. Tento test je založený na testovaní najväčšej odchýlky teoretickej distribučnej funkcie a empirickej distribučnej funkcie. Označme si $F_n(x)$ empirickú distribučnú funkciu pre náhodný výber rozsahu n a $F(x)$ príslušnú teoretickú distribučnú funkciu. Rozoznávame tri testy:
Obojstranný test:

a) $H_0: \quad F_n(x) = F(x)$ oproti alternatíve $H_1: \quad F_n(x) \neq F(x)$.

Jednostranné testy:

b) $H_0: \quad F_n(x) \leq F(x)$ oproti alternatíve $H_1: \quad F_n(x) > F(x)$.

c) $H_0: \quad F_n(x) \geq F(x)$ oproti alternatíve $H_1: \quad F_n(x) < F(x)$.

V prípade jednotlivých testov budeme používať nasledujúce štatistiky:

a) $T = \sup_x |F(x) - F_n(x)|$.

b) $T^+ = \sup_x (F(x) - F_n(x))$.

c) $T^- = \sup_x (F_n(x) - F(x))$.

Je zrejmé, že štatistiky T, T^+, T^- môžu nadobúdať hodnoty len z intervalu $\langle 0,\, 1\rangle$.

Podľa charakteru funkcie $F(x)$, máme dva prípady: $F(x)$ je spojiá, alebo $F(x)$ je diskrétna. Budeme sa zaoberať prípadom, keď $F(x)$ je spojitá distibučná funkcia. Korekcie pre p-hodnotu v prípade diskrétnej náhodnej premennej je možné nájsť napríklad v [**?**]

Nech $F(x)$ je spojitá. Potom v prípadoch (b), (c), ak platí H_0, tak T^+ a T^- májú rovnakú distribučnú

funkciu

$$F_n(x) = 1 - x \sum_{j=0}^{[n(1-x)]} \binom{n}{j} \left(1 - x - \frac{j}{n}\right)^{n-j} \left(x + \frac{j}{n}\right)^{j-1},$$

kde $[n(1-x)]$ je najväčšie celé číslo menšie, alebo rovné $n(1-x)$ ($[5,78] = 5$). Napríklad

$$\begin{aligned}
F_{10}(0,81) =& 1 - 0,81 \sum_{j=0}^{[10(1-0,81)]} \binom{10}{j} \left(1 - 0,81 - \frac{j}{10}\right)^{10-j} \left(0,81 + \frac{j}{10}\right)^{j-1} \\
=& 1 - 0,81 \sum_{j=0}^{[10(0,19)]} \binom{10}{j} \left(0,19 - \frac{j}{10}\right)^{10-j} \left(0,81 + \frac{j}{10}\right)^{j-1} \\
=& 1 - 0,81 \sum_{j=0}^{[1,9]} \binom{10}{j} \left(0,19 - \frac{j}{10}\right)^{10-j} \left(0,81 + \frac{j}{10}\right)^{j-1} \\
=& 1 - 0,81 \sum_{j=0}^{1} \binom{10}{j} \left(0,19 - \frac{j}{10}\right)^{10-j} \left(0,81 + \frac{j}{10}\right)^{j-1} \\
=& 1 - 0,81 \left(\frac{0,19^{10}}{0,81} + 10 \cdot 0,09^9\right)
\end{aligned}$$

Pre $n \to \infty$ distribučná funkcia pre $\sqrt{n} \cdot T^+$ a $\sqrt{n} \cdot T^-$) má tvar

$$F(x) = \lim_{n\to\infty} F_n\left(\frac{x}{\sqrt{n}}\right) = 1 - e^{-2x^2}.$$

Aproximatívna distribučná funkcia pre T je daná vzťahom

$$P(T \leq x) \doteq (F_n(x))^2 .$$

Používa sa pre $n > 40$.

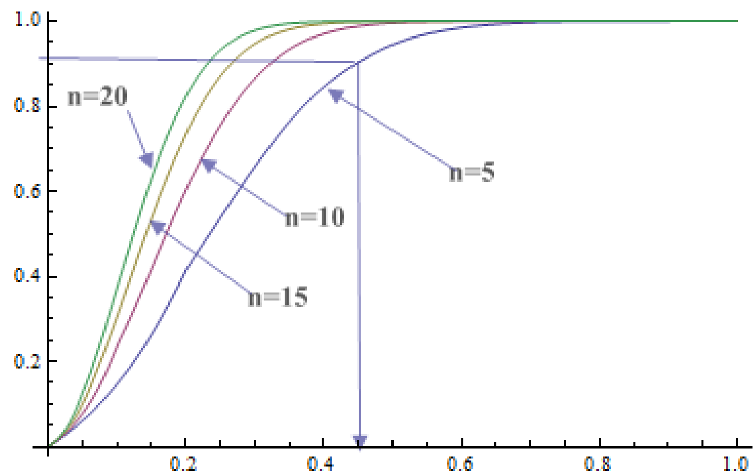

Fig. 5: $F_n(x) = P(X \leq t)$, pre $n \in \{5, 10, 15, 20\}$

Ak platí H_0, potom

$$P(max|F_n(x) - F(x)| > D_n(\alpha)) = \alpha,$$

kde $D_\alpha(n)$ je kvantil (kritická hodnota), vypočítaná z teoretického rozdelenia pravdepodobnosti Kolmogorovom [?]. Hodnoty $D_n(\alpha)$ pre niektoré

$$\alpha = 0,05; 0,01; 0,1$$

sú uvedené napr. v [?].
Ak platí H_0, potom

$$P(max(F_n(x) - F(x)) < D_n(\alpha)) = \alpha,$$

kde $D_\alpha(n)$ je kvantil (kritická hodnota), vypočítaná z teoretického rozdelenia pravdepodobnosti Kolmogorovom [?, ?]. Tabuľka pre K-S test nie je uvedená, ale odporúăme radšej vypočítať p-hodnotu p. V prípade obostranného testu $p = 2 \cdot F_n(KS)$, v prípade (a)

Postup:

1. Namerané hodnoty usporiadame podľa veľkosti

   $$x_1 \leq x_2 \leq \cdots \leq x_n.$$

   a stanovíme hodnoty empirickej distribučnej funkcie $F_n(x)$.

2. Pomocou nameraných údajov odhadneme parametre pre predpokladanú teoretickú distribučnú funkciu a vypočítame hodnoty teoretickej distribučnej funkcie $F(x)$.

3. Zistíme maximálnu hodnotu absolútneho rozdielu medzi teoretickou a empirickou distribučnou funkciou

   $$KS = max_i|F(x_i) - F_n(x_i)|.$$

4. V štatistikej tabuľke pre K-S test pre zvolenú hladinu významnosti nájdeme a pre daný rozsah súboru n si nájdeme kvantil $D_n(\alpha)$.

5. Ak $KS \geq D_n(\alpha)$, potom hypotézu H_0 o zhodnosti distribučných funkcií zamietame na hladine významnosti α . V opačnom prípade nemáme k zamiestnutiu hypotézy H_0 dostatočné dôvody.

Príklad 12: (Pr. 11
Namerané údaje z príkladu 11 teraz spracujeme pomocou K-S testu. Štatistické Hypotézy sú také isté:

$$\mathbf{H_0} : X \sim N(12; 1) \qquad \mathbf{H_1} : H_0 \text{ neplatí}$$

Merali sme 10 krát teplotu a namerali nasledujúce teploty (údaje sme usporiadali podľa veľkosti:

j	1	2	3	4	5
x_j	10,49	10,85	10,92	11,05	11,96
$F_{10}(x_j)$	0,1	0,2	0,3	0,4	0,5
$F(x_j)$	0,066	0,125	0,14	0,171	0,484
$\Delta(x_j)$	0,034	0,075	0,16	**0,229**	0,016

j	6	7	8	9	10
x_j	12,06	12,61	13,1	13,55	13,58
$F_{10}(x_j)$	0,6	0,7	0,8	0,9	1
$F(x_j)$	0,524	0,729	0,864	0,939	0,943
$\Delta(x_j)$	0,076	0,029	0,064	0,039	0,057

kde $\Delta(x_j) = |F_{10}(x_j) - F(x_j)|$. Štatistika

$$KS = max_j(|F_{10}(x_j) - F(x_j)|) = 0,229$$

a v tabuľkách pre K-S $D_{10}(0,05) = 0,409$. Pretože $KS < D_{10}(0,05)$ tak H_0 teda na hladine významnosti $\alpha = 0,05$ nezamietame.

3 Príklady

Príklad 13:
Nech množina $A = \{1, 2, 3\}$ a Ω je množina všetkých trojciferných čísel.
a) Koľko prkov má množina Ω?
b) Koľko prkov má množina Ω_1: každé číslo môžeme použiť len raz?
c) Koľko prkov má množina Ω_2: každé číslo, okrem 1, môžeme použiť len raz?

Riešenie pr. 13: a) 27. b) 6. c) 13.

Príklad 14:
Vieme že prístupový kód sa skladá, z desiatich rôznych znakov, ktoré poznáme. Aká je pravdepodobnosť, že na prvý krát uhádneme prístupový kód?

Riešenie pr. 14: $P(\text{uhádne kód na prvý krát }) = \frac{1}{10!} = 2,75573 \cdot 10^{-7}$.

Príklad 15:
V urne máme 10 loptičiek označených $1, 2, 3, \ldots, 10$. Ťaháme postupne po jednej loptičke. Vytiahnutú loptičku nevraciame do urny. Toto opakujem dovtedy, kým urna nebude prázdna. Aká je pravdepodobnosť, že ich vytiahneme v poradí $1, 3, 2, 4, 5, 6, 7, 8, 10, 9$?

Riešenie pr. 15: $P(1, 3, 2, 4, 5, 6, 7, 8, 10, 9)) = \frac{1}{10!} = 2,75573 \cdot 10^{-7}$

Príklad 16:
V urne máme 10 loptičiek označených tak, že 5 z nich má znak 1 a zvyšné $2, 3, 4, 5, 6$. ťaháme postupne po jednej loptičke. Vytiahnutú loptičku nevraciame do urny. Toto opakujeme dovtedy, kým urna nebude prázdna. Aká je pravdepodobnosť, že ich vytiahneme v poradí $1, 3, 2, 4, 5, 6, 1, 1, 1, 1$?

Riešenie pr. 16: $P(1, 3, 2, 4, 5, 6, 1, 1, 1, 1) = \frac{1}{30240} = 3,30688 \cdot 10^{-5}$

Príklad 17:
V urne máme 12 loptičiek označených $1, 2, 3$, tak, že 5 z nich je označených číslom 1, číslom 2 sú označené 4 a zvyšné číslom 3. Ťaháme postupne po jednej loptičke. Vytiahnutú loptičku nevraciame do urny. Toto opakujeme dovtedy, kým urna nebude prázdna. Aká je pravdepodobnosť, že ich vytiahneme v poradí $1, 2, 2, 2, 2, 3, 1, 1, 1, 1, 3, 3$?

Riešenie pr. 17: $P(1, 2, 2, 2, 2, 3, 1, 1, 1, 1, 3, 3) = \frac{1}{27720} = 3,6075 \cdot 10^{-5}$

Príklad 18:
V urne máme 12 loptičiek označených $1, 2, 3$, tak, že 5 z nich je označezných číslom 1, 4 sú označené číslom 2 a zvyšné číslom 3. Ťaháme postupne po jednej loptičke. Vytiahnutú loptičku nevraciame do urny. Toto opakujeme 4 krát. Aká je pravdepodobnosť, že všetky budú označené číslom 1?

Riešenie pr. 18: $= \frac{5!}{4! \cdot 1!} = 5. Teda P(1, 1, 1, 1) = \frac{5}{495} = 0,01\overline{01}$

Príklad 19:
Máme udalosti A a B. Nech $P(A) = 0,3$ a $P(B) = 0,6$. Nech $0 \leq P(A \cap B) \leq 0,5$.
a) Nájdite intervaly, do ktorých padnú hodnoty $P(A|B)$, $P(B|A)$.
b) Môžu byť A, B navzájom nezávislé?

Riešenie pr. 19: a) $0 \leq P(A|B) \leq 0,5$ a $0 \leq P(B|A) \leq 1$. b) Áno.

Príklad 20:
Bayesová formula: Priestor náhodných udalostí je rozdelený na 5 častí $H_1, \dots, H_5$ (navzájom sa vylučujúcich) a je daná náhodná udalosť A. Ak poznáme $P(A|H_i)$ a $P(H_i)$ pre $i = 1, \dots, 5$ ako vypočítame $P(A)$ a $P(H_3|A)$?

Riešenie pr. 20: $P(A) = \sum_i P(H_i)P(A|H_i)$ a $P(H_3|A) = \frac{P(H_3)P(A|H_3)}{P(A)}$.

Príklad 21:
V ročníku je 100 študentov, ktorí sú rozdelení do piatich skupín. Vypočítajte koľko dievčat je v ročníku. Ak náhodne vyberieme jedno dievča, aká je pravdepodobnosť, že patrí do tretej skupiny? Rozdelenie študentov je v tabuľke 4:

skupiny	H_1	H_2	H_3	H_4	H_5
percento študentov v skupine	10%	30%	10%	20%	30%
percento dievčat v skupine	50%	10%	10%	0%	10%

Tab. 4: Študenti (Pr. 21)

Riešenie pr. 21: V ročníku je 12 dievčat a a $P(H_3|\text{ dievča}) = \frac{1}{12}$.

Príklad 22:
V jednej krabici je 10 modrých a 5 červených balónov. V druhej krabici je osem bielych a 12 modrých balónov.
a) Náhodne si zvolíme jednu krabicu a vyberieme z nej 1 balón. Aká je pravdepodobnosť, že nebude modrý?
b) Náhodne vyberieme jednu krabicu. Aká je pravdepodobnosť, že z nej vytiahnutý balón bude biely?
c) Z oboch krabíc vyberieme po jednom balóne. Aká je pravdepodobnosť, že oba budú modré?

Riešenie pr. 22: a) $0,3\overline{6}$, b) $0,2$, c) $0,4$.

Príklad 23:
V lietadle je 20% cestujúcich zo SR. Je známe, že 60% obyvateľ ov SR pije po obede pivo, kým obyvatelia iných štátov vypijú pivo po obede len v 20%.
a) Aké percento cestujúcich v lietadle neuprednostňuje pivo po obede?
b) Cestujúci si po obede vypýta pivo. S akou pravdepodobnosťou je to občan SR?

Riešenie pr. 23: a) 72%, b) $0,42857$.

Príklad 24:
Je známe, že 25% obyvateľstva je ľavákov. Aká je pravdepodobnosť, že na seminári kde je 30 účastníkov sú maximálne traja ľaváci?

Riešenie pr. 24: $\doteq 0,03745$.

Príklad 25:
Nikto nie je neomylný: obvodný lekár určí v 50% prípadoch správnu diagnózu, v 20% prípadoch nesprávnu diagnózu a v 30% prípadoch odporučí pacienta na vyšetrenie k špecialistovi na polikliniku. Špecialista určí v 60% správnu diagnózu, v 15% nesprávnu a pri 25% pošle pacienta na konziliárne vyšetrenie k primárovi. Primár určí správnu diagnózu v 85% a nesprávnu v 15%. a) Aká je pravdepodobnosť, že obvodný lekár určí diagnózu správne?
b) Aká je pravdepodobnosť, že pacient bude mať nesprávne určenú diagnózu?

Riešenie pr. 25: a) $0,5$, b) $0,25625$.

Príklad 26:
Máme tri košíky. Každý z troch košíkov obsahuje jednu bielu a dve čierne guľôčky. Z prvého košíka náhodne vyberieme guličku a vložíme do druhého. Z druhého košíka vybierieme náhodne jednu guličku a vložíme do tretieho. Aká je pravdepodobnosť, že z tretieho košíka náhodne vytiahneme čiernu guličku?

Riešenie pr. 26: $0,\overline{6}$

Príklad 27:
Dva závody vyrábajú okenné rámy. Prvý závod vyrába 45% celkovej produkcie, druhý 55%. Z produkcie prvého závodu: 90% I.kategórie, 10% II. kategórie; druhého závodu: 95% I. kategórie, 5% II. kategórie, Určte pravdepodobnosť toho, že náhodne vybraný rám je I. kategórie.

Riešenie pr. 27: Podmienená pravdepodobnosť, veta o úplnej pravdepodobnosti, Bayesova veta.

Príklad 28:
Päť krát hádžem mincou. N.p. X je „počet hláv mínus počet znakov".
a) Aké hodnoty môže nadobúdať n.p. X.
b) Napíšte všetky elementárne udalosti pre prípad $X(\omega) = 3$.

Riešenie pr. 28: a) $X \in \{-5, -3, -1, 1, 3, 5\}$. b) (H, H, H, H, Z), (H, H, H, Z, H), (H, H, Z, H, H), (H, Z, H, H, H), (Z, H, H, H, H).

Príklad 29:
N.p. X je rozdiel "počet hláv mínus počet znakov"pri trojnásobnom hode mincou.
a) Aké hodnoty môže n.p. X nadobúdať?
b) Nájdite pre n.p. X príslušné elementárne udalosti.
c) Určite rozdelenie pravdepodobnosti a distribučnú funkciu pre n.p. X.

Riešenie pr. 29: a) $X \in \{-3, -1, 1, 3\}$;
b) $X^{-1}(-3) = \{(Z, Z, Z)\}$, $X^{-1}(-1) = \{(H, Z, Z), (Z, H, Z), (Z, Z, H)\}$,
$X^{-1}(-1) = \{(H, H, Z), (H, Z, H), (Z, H, H)\}$, $X^{-1}(3) = \{(H, H, H)\}$;
c) $P_X = \{(-3,\ 0,125), (-1\,0,375), (1,\ 0,375), (3\,0,125)\}$,
$F_X = \{(-3,\ 0,125), (-1,\ 0,5), (1,\ 0,875), (3,\ 1)\}$.

Príklad 30:
N.p. X nadobúda hodnoty z množiny $\{1, 2, \ldots, 10\}$ a jej pravdepodobnostné rozdelenie je dané v tabuľke:

x_i	1	2	3	4	5	6	7	8	9	10
p_i	$0,05$	$0,05$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,2$	$0,1$

Vypočítajte strednú hodnotu $E(X)$ a disperziu $D(X)$.

Riešenie pr. 30: $E(X) = 6,25$ a $D(X) = 7,2875$.

Príklad 31:
Advokátova obhajoba je úspešná s pravdepodobnosťou $p = 0,8$. Pri úspešnej obhajobe dostane 2000EUR, pri neúspešnej obhajobe musí zaplatiť súdne trovy vo výške 500EUR.
a) Aký vysoký je priemerný zisk obhájcu?
b) Keby obhajca vydal 800EUR navyše na prípravu obhajoby, aký by bol jeho očakávaný zisk?

Riešenie pr. 31: a) 1500EUR b) 700EUR.

Príklad 32:
Hráč si zvolí číslo medzi $1-6$ a hodí tromi kockami. Ak všetky 3 kocky ukážu to isté zvolené číslo, vyhrá 3EUR. Ak toto číslo ukážu 2 kocky, vyhráva 2EUR a ak iba 1 kocka vyhráva 1EUR. Ak ani jedna kocka neukáže zvolené číslo, potom musí zaplatiť 1EUR. Náhodná premenná X znamená výšku výhry.
a) Aké hodnoty nadobúda n.p. X?
b) Zistite pravdepodobnostné rozdelenie n.p. X a vypočítajte jej strednú hodnotu a disperziu.

Riešenie pr. 32: a) $X \in \{-1, 1, 2, 3\}$, b) $P(X = -1) = 125/216$, $P(X = 1) = 75/216$, $P(X = 2) = 15/216$, $P(X = 3) = 1/216$, $E(X) \doteq -0,079$, $D(X) \doteq 1,24$.

Príklad 33: Semafór na križovatke ukazuje 25% času červenú. Aká veľká je pravdepodobnosť, že z 5 náhodne prechádzajúcich áut
a) nemusí čakať ani jedno auto;
b) musí čakať nanajvýš jedno auto;
c) musia čakať práve tri autá;
d) musia čakať maximálne dve autá?

Riešenie pr. 33: a) $0,237305$, b) $0,633$, c) $0,088$, d) $0,896$.

Príklad 34:
Pri majstrovstvách v tenise hrá hráč A proti hráčovi B toľko po sebe idúcich

setov, kým jeden z hráčov nevyhrá tri sety. Výsledky setov sú od seba nezávislé a hráč A vyhrá set s pravdepodobnosťou $0,6$. N.p. X označuje počet hraných setov.
a) Určte hodnoty n.p. X a nájdite elementárne udalosti prislúchajúce $X = 4$.
b) Určte rozdelenie pravdepodobnosti n.p. X.
c) S akou pravdepodobnosťou sa zápas skončí po piatom sete?
d) Aká je pravdepodobnosť, že na víťazstvo jedného z nich sú potrebné maximálne 4 sety? e) Aká je pravdepodobnosť, že na víťaztsvo hráča A sú potrebné maximálne 4 sety?

Riešenie pr. 34: a) $X \in \{3, 4, 5\}$, $AABA$, $ABAA$, $BAAA$, $BBAB$, $BABB$, $ABBB$; b) $P_X = \{(3,\ 0,28), (4,\ 0,3744), (5,\ 0,3456),\}$; c)$0,3456$; d) $0,6544$; e) $0,4752$.

Príklad 35:
Máme dané rozdelenie pravdepodobnosit n.p. X:

$$P_X = \{(0,\ 0,1), (1,\ 0,2), (2,\ 0,3), (3,\ 0,4)\}.$$

a) Nakreslite graf distribučnej funkcie n.p. X. b) Zistite modus. c) Určte strednú hodnotu a disperziu n.p. X.

Riešenie pr. 35: b) $Mo_X = 3$; c) $E(X) = 2$, $D(X) = 1$.

Príklad 36:
Udalosť A sa vyskytuje pri experimente s pravdepodobnosťou $P(A) = 0,4$. Pokus sa opakuje dovtedy, kým sledovaná udalosť A nastane. Aká je pravdepodobnosť, že budeme potrebovať zopakovať 3?

Riešenie pr. 36: $0,144$.

Príklad 37:
Infekcia sa prenáša kontaktom. Pravdepodobnosť prenosu infekcie na zdravého človeka pri prvom kontakte je $0,4$.
a) Jeden infikovaný má kontakt s piatimi zdravými ľuďmi. Určite rozdelenie n.p. X, pričom n.p. X znamená počet osôb, ktoré ochorejú.
b) Vypočítajte $E(X)$, $D(X)$, $P(X = 0)$, $P(X < 3)$.

Riešenie pr. 37: a) $Bi(5; 0,4)$, b) 2; $1,2$; $0,07776$; $0,68256$.

Príklad 38:
Prieskum verejnej mienky ukázal, že 80% obyvateľov podporuje zákaz výstavby v lokalite Včelín. Aká je pravdepodobnosť, že z dvadsatich náhodne oslovených obyvateľov je maximálne 12 za zákaz výstavby v lokalite Včelín?

Riešenie pr. 38: $0,03214$.

Príklad 39:
Len 30% ľudí vo veľkom meste si myslí, že verejná doprava v meste (MHD) je uspokojivá. Náhodne vyberiem desať osôb.
a) Aká je pravdepodobnosť, že najviac 5 z nich si myslí, že MHD je uspokojivá?
b) Aká je pravdepodobnosť, že práve 6 si myslí, že MHD je uspokojivá?

Riešenie pr. 39: a) $0,4164$; b) $0,1916$.

Príklad 40:
Basketbalový hráč má osobnú štatistiku úspešných hodov 70%.
a) Aká je pravdepodobnosť, že pri desiatich hodoch trafí 8 krát?
b) Aká je pravdepodobnosť, že pri sto pokusoch trafí menej ako 60 krát?

Riešenie pr. 40: a) $X \sim Bi(10; 0,7)$ a riešime $P(X = 8)$; b)$X \sim Bi(100; 0,7)$ a riešime $P(X < 60)$.

Príklad 41:
V krabici s 20-timi fixkami na bielu tabuľu sa nachádzajú 3 vadné. Učiteľ si vyberie 5. N.p. X predstavuje počet vadných fixiek.
a) Aké je rozdelenie pravdepodobnosti n.p. X?
b) Aká je pravdepodobnosť, že minimálne jedna z jeho piatich fixiek je vadná?

Riešenie pr. 41: a) $P_X = \{(0,\ 0,3991), (1,\ 0,4605), (2,\ 0,1316), (3,,\ 0,0088)\}$; b) $0,6009$.

Príklad 42:
Vypočítajte $E(X)$ a $D(X)$ n.p. X a nakreslite graf rozdelenia pravdepobodnosti a graf distribučnej funkcie, ak má nasledujúce rozdelenie pravdepodobnosti:

$$P_X = \{(2,\ 0,1), (3,\ 0,3), (4,\ 0,3), (5,\ 0,2), (6,\ 0,1)\}.$$

Riešenie pr. 42: $E(X) = 3,9$ a $D(X) = 1,29$.

Príklad 43:
Nech X je n.p. a

$$P_X = \{(-1,\ 0,1), (-1/2,\ 0,2), (0,\ 0,4), (1,\ 0,2), (2,\ 0,1)\}\,.$$

Vypočítajte $\rho(X, X^2)$.

Riešenie pr. 43: $\rho(X, X^2) = 0,74053$.

Príklad 44:
Vedenie obchodnej firmy zaujímalo, či je rozdiel medzi výsledkami ich zamestnancov (predajcov) v jednotlivých skupinách štatisticky významný. Zamestanci boli rozdelení do nasledujúcich skupín podľa oblastí, v ktorých pôsobia: A (Staré mesto); B (Petržálka); C (Dúbravka); D (Ružinov). V každej časti pôsobilo 20 dílerov. $\overline{X}_i$ je priemerný zisk a s_i je odhad smerodajnej odchylky v skupine za týždeň, $i = 1, 2, 3, 4$.

mestksá časť	$\overline{X}_i$	s_i
A (Staré mesto)	$1253,2$ EUR	150 EUR
B (Petržálka)	$1085,4$ EUR	148 EUR
C (Dúbravka)	$1120,2$ EUR	152 EUR
D (Ružinov)	$1990,2$ EUR	170 EUR

Zistite, či je štatisticky významný rozdiel medzi jednotlivými skupinami dílerov.

Riešenie pr. 44: (ANOVA). Otestujte najskôr rovnosť rozptylov.

Príklad 45:
Zisťoval sa počet detí v rodine. Prieskum bol urobený v 200 rodinách. Výsledky sú v nasledujúcej frekvenčnej tabuľke:

Počet detí v rodine	0	1	2	3	4
Počet rodín	50	67	23	15	45

a) Zistite modus., medián, dolný a horný kvartil. b) Načrtnite krabicový graf. c) Vypočítajte odhad strednej hodnoty a disperzie.

Riešenie pr. 45: a) Modus $= 0$, medián $= 1$, dolný kvartil $= 0,5$ a horný kvartil $= 4$; c) $\mu \doteq \overline{x} = 1,69$, $\sigma^2 \doteq S_1^2 = 0,056$.

Príklad 46:
Urobilo sa 20 meraní obsahu dusíka vo vodnej nádrži. Vypočítali sme obojstranné konfidenčné intervaly pre stredú hodnotu a rozptyl na rôznych hladinách významnosti $\alpha \in \{0,1;\ 0,05;\ 0,01\}$. Priraďte jednotlivým intervalom ich hladiny významnosti a rozhodnite či stredná hodnota sa rovná $0,253$ oproti alternatíve, že sa nerovná $0,253$ a či rozptyl sa rovná $0,025$ oproti alternatíve, že sa nerovná $0,024$ na jednotlivých hladinách významnosti. Aký aritmetický priemer pre obsah dusíka bol vypočítaný?

IS pre μ	α_μ	$H_0: \mu = \mu_0$	IS pre σ^2	α_{σ^2}	$H_0: \sigma^2 = \sigma_0^2$
$(0,22;\ 0,26)$			$(0,018;\ 0,023)$		
$(0,21;\ 0,27)$			$(0,017;\ 0,025)$		
$(0,23;\ 0,25)$			$(0,009;\ 0,032)$		

Riešenie pr. 46: $\overline{x} = 0,24$ a IS v tabuľke 5

IS pre μ	α_μ	$H_0: \mu = \mu_0$	IS pre σ^2	α_{σ^2}	$H_0: \sigma^2 = \sigma_0^2$
$(0,22;\ 0,26)$	$0,05$	NH0	$(0,018;\ 0,023)$	$0,1$	ZH0
$(0,21;\ 0,27)$	$0,01$	NH0	$(0,017;\ 0,025)$	$0,05$	ZH0
$(0,23;\ 0,25)$	$0,1$	ZH0	$(0,009;\ 0,032)$	$0,01$	NH0

Tab. 5: NH0 – nezamietame H_0 a ZH0 – zamietamme H_0

Príklad 47:
Vedenie fakulty sa rozhodlo, že odmení študentov za výsledok zo skúsky z Matematiky 1 (tabuľka 6).

Výsledok skúsky	A	B	C	D	E
X: Odmena v EUR	25	15	10	5	0
P_X	$0,1$	$0,15$	$0,25$	$0,2$	$0,3$

Tab. 6: P_X je pravdepodobnosť s akou známkou študent ukončí skúšku. Pravdepodobnosť sa zistila dlhodobým sledovaním výsledkov skúšky z Matematiky. Do úvahy sa nebrala známka FX.

V ročníku je 100 študentov. Pre plánovanie financií je potrebné vedieť, koľko EUR budú asi potrebovať, ak predpokladáme, že výsledky zo skúšky budú zodpovedať predchádzajúcim ročníkom.

Riešenie pr. 47: Odhad strednej hodnoty n.p. X je očakávaná odmena v priemere na jedného študenta a smerodajná odchýlka predstavuje variabilitu.

4 Základné štatistické tabuľky

1. N.p. $X \sim N(0,\,1)$: $F(t) = P(X \leq t) = (\Phi(t))$, pre $t \in R$
2. N.p. $X \sim N(0,\,1)$: $F^{-1}(\alpha) = P(X \leq t) = (\Phi(t))$, pre $t \in R$
3. N.p. $T \sim t(n)$: $t_n^{-1}(\alpha)$ je α-kvantil pre n.p. T. To znamená

$$F^{-1}(t_n^{-1}(\alpha)) = P(T \leq t_n^{-1}(\alpha)) = \alpha,$$

kde $\alpha \in \{0,75;\, 0,9;\, 0,95;\, 0,975;\, 0,99;\, 0,995\}$ a $n \in \{1,\, 2, \ldots, 35,\, 40,\, 60\}$.

$X \sim N(\mu, \sigma^2)$
Pravidlo 3σ: $a_k = \mu - k \cdot \sigma, \quad b_k = \mu + k \cdot \sigma$

k	$P(X \leq a_k)$	$P(X \leq b_k)$	$P(a_k < X \leq b_k)$
1	0,159	0,841	0,683
2	0,023	0,977	0,954
3	0,001	0,999	0,997

$X \sim N(0,1)$

Vybrané hodnoty kvantilu: $P(X \leq Q_\alpha) = \alpha$:

α	0,500	0,900	0,950	0,975	0,990	0,995	0,999
Q_α	0,000	1,282	1,645	1,960	2,326	2,576	3,091

Tab. 7: Ak $X \sim N(0,1)$ a $P(X \leq t + r)$, potom $P(X \leq 1,55) = 0,93943$, pre $t + k = 1,55$.

t/k	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861
3	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983

Tab. 8: Ak $T \sim t(n)$ a $P(T \leq r) = p$, potom $P(T \leq 1,53321) = 0,9$, pre $T \sim t(4)$.

n/p	0,75	0,9	0,95	0,975	0,99	0,995	0,999
1	1,00000	3,07768	6,31375	12,70620	31,82052	63,65674	318,30884
2	0,81650	1,88562	2,91999	4,30265	6,96456	9,92484	22,32712
3	0,76489	1,63774	2,35336	3,18245	4,54070	5,84091	10,21453
4	0,74070	1,53321	2,13185	2,77645	3,74695	4,60409	7,17318
5	0,72669	1,47588	2,01505	2,57058	3,36493	4,03214	5,89343
6	0,71756	1,43976	1,94318	2,44691	3,14267	3,70743	5,20763
7	0,71114	1,41492	1,89458	2,36462	2,99795	3,49948	4,78529
8	0,70639	1,39682	1,85955	2,30600	2,89646	3,35539	4,50079
9	0,70272	1,38303	1,83311	2,26216	2,82144	3,24984	4,29681
10	0,69981	1,37218	1,81246	2,22814	2,76377	3,16927	4,14370
11	0,69745	1,36343	1,79588	2,20099	2,71808	3,10581	4,02470
12	0,69548	1,35622	1,78229	2,17881	2,68100	3,05454	3,92963
13	0,69383	1,35017	1,77093	2,16037	2,65031	3,01228	3,85198
14	0,69242	1,34503	1,76131	2,14479	2,62449	2,97684	3,78739
15	0,69120	1,34061	1,75305	2,13145	2,60248	2,94671	3,73283
16	0,69013	1,33676	1,74588	2,11991	2,58349	2,92078	3,68615
17	0,68920	1,33338	1,73961	2,10982	2,56693	2,89823	3,64577
18	0,68836	1,33039	1,73406	2,10092	2,55238	2,87844	3,61048
19	0,68762	1,32773	1,72913	2,09302	2,53948	2,86093	3,57940
20	0,68695	1,32534	1,72472	2,08596	2,52798	2,84534	3,55181
21	0,68635	1,32319	1,72074	2,07961	2,51765	2,83136	3,52715
22	0,68581	1,32124	1,71714	2,07387	2,50832	2,81876	3,50499
23	0,68531	1,31946	1,71387	2,06866	2,49987	2,80734	3,48496
24	0,68485	1,31784	1,71088	2,06390	2,49216	2,79694	3,46678
25	0,68443	1,31635	1,70814	2,05954	2,48511	2,78744	3,45019
26	0,68404	1,31497	1,70562	2,05553	2,47863	2,77871	3,43500
27	0,68368	1,31370	1,70329	2,05183	2,47266	2,77068	3,42103
28	0,68335	1,31253	1,70113	2,04841	2,46714	2,76326	3,40816
29	0,68304	1,31143	1,69913	2,04523	2,46202	2,75639	3,39624
30	0,68276	1,31042	1,69726	2,04227	2,45726	2,75000	3,38518
31	0,68249	1,30946	1,69552	2,03951	2,45282	2,74404	3,37490
32	0,68223	1,30857	1,69389	2,03693	2,44868	2,73848	3,36531
33	0,68200	1,30774	1,69236	2,03452	2,44479	2,73328	3,35634
35	0,68156	1,30621	1,68957	2,03011	2,43772	2,72381	3,34005
60	0,67860	1,29582	1,67065	2,00030	2,39012	2,66028	3,23171

Tab. 9: $Q \sim \chi^2(n)$ a $P(Q \leq r) = p$. Napríklad $P(Q \leq 9,48773) = 0,95$, pre $Q \sim \chi^2(4)$.

n/p	0,25	0,75	0,5	0,1	0,9	0,05	0,95
1	0,10153	1,32330	0,45494	0,01579	2,70554	0,00393	3,84146
2	0,57536	2,77259	1,38629	0,21072	4,60517	0,10259	5,99146
3	1,21253	4,10834	2,36597	0,58437	6,25139	0,35185	7,81473
4	1,92256	5,38527	3,35669	1,06362	7,77944	0,71072	9,48773
5	2,67460	6,62568	4,35146	1,61031	9,23636	1,14548	11,07050
6	3,45460	7,84080	5,34812	2,20413	10,64464	1,63538	12,59159
7	4,25485	9,03715	6,34581	2,83311	12,01704	2,16735	14,06714
8	5,07064	10,21885	7,34412	3,48954	13,36157	2,73264	15,50731
9	5,89883	11,38875	8,34283	4,16816	14,68366	3,32511	16,91898
10	6,73720	12,54886	9,34182	4,86518	15,98718	3,94030	18,30704
11	7,58414	13,70069	10,34100	5,57778	17,27501	4,57481	19,67514
12	8,43842	14,84540	11,34032	6,30380	18,54935	5,22603	21,02607
13	9,29907	15,98391	12,33976	7,04150	19,81193	5,89186	22,36203
14	10,16531	17,11693	13,33927	7,78953	21,06414	6,57063	23,68479
15	11,03654	18,24509	14,33886	8,54676	22,30713	7,26094	24,99579
16	11,91222	19,36886	15,33850	9,31224	23,54183	7,96165	26,29623
17	12,79193	20,48868	16,33818	10,08519	24,76904	8,67176	27,58711
18	13,67529	21,60489	17,33790	10,86494	25,98942	9,39046	28,86930
19	14,56200	22,71781	18,33765	11,65091	27,20357	10,11701	30,14353
20	15,45177	23,82769	19,33743	12,44261	28,41198	10,85081	31,41043
21	16,34438	24,93478	20,33723	13,23960	29,61509	11,59131	32,67057
22	17,23962	26,03927	21,33704	14,04149	30,81328	12,33801	33,92444
23	18,13730	27,14134	22,33688	14,84796	32,00690	13,09051	35,17246
24	19,03725	28,24115	23,33673	15,65868	33,19624	13,84843	36,41503
25	19,93934	29,33885	24,33659	16,47341	34,38159	14,61141	37,65248
26	20,84343	30,43457	25,33646	17,29188	35,56317	15,37916	38,88514
27	21,74940	31,52841	26,33634	18,11390	36,74122	16,15140	40,11327
28	22,65716	32,62049	27,33623	18,93924	37,91592	16,92788	41,33714
29	23,56659	33,71091	28,33613	19,76774	39,08747	17,70837	42,55697
30	24,47761	34,79974	29,33603	20,59923	40,25602	18,49266	43,77297

Tab. 10: $Q \sim \chi^2(n)$ a $P(Q \leq r) = p$. Napr. $P(Q \leq 2,55821) = 0,01$, pre $Q \sim \chi^2(10)$.

n/p	0,025	0,975	0,01	0,99	0,005	0,995	0,001	0,999
1	0,00098	5,02389	0,00016	6,63490	0,00004	7,87944	0,00000	10,82757
2	0,05064	7,37776	0,02010	9,21034	0,01003	10,59663	0,00200	13,81551
3	0,21580	9,34840	0,11483	11,34487	0,07172	12,83816	0,02430	16,26624
4	0,48442	11,14329	0,29711	13,27670	0,20699	14,86026	0,09080	18,46683
5	0,83121	12,83250	0,55430	15,08627	0,41174	16,74960	0,21021	20,51501
6	1,23734	14,44938	0,87209	16,81189	0,67573	18,54758	0,38107	22,45774
7	1,68987	16,01276	1,23904	18,47531	0,98926	20,27774	0,59849	24,32189
8	2,17973	17,53455	1,64650	20,09024	1,34441	21,95495	0,85710	26,12448
9	2,70039	19,02277	2,08790	21,66599	1,73493	23,58935	1,15195	27,87716
10	3,24697	20,48318	2,55821	23,20925	2,15586	25,18818	1,47874	29,58830
11	3,81575	21,92005	3,05348	24,72497	2,60322	26,75685	1,83385	31,26413
12	4,40379	23,33666	3,57057	26,21697	3,07382	28,29952	2,21421	32,90949
13	5,00875	24,73560	4,10692	27,68825	3,56503	29,81947	2,61722	34,52818
14	5,62873	26,11895	4,66043	29,14124	4,07467	31,31935	3,04067	36,12327
15	6,26214	27,48839	5,22935	30,57791	4,60092	32,80132	3,48268	37,69730
16	6,90766	28,84535	5,81221	31,99993	5,14221	34,26719	3,94163	39,25235
17	7,56419	30,19101	6,40776	33,40866	5,69722	35,71847	4,41609	40,79022
18	8,23075	31,52638	7,01491	34,80531	6,26480	37,15645	4,90485	42,31240
19	8,90652	32,85233	7,63273	36,19087	6,84397	38,58226	5,40682	43,82020
20	9,59078	34,16961	8,26040	37,56623	7,43384	39,99685	5,92104	45,31475
21	10,28290	35,47888	8,89720	38,93217	8,03365	41,40106	6,44668	46,79704
22	10,98232	36,78071	9,54249	40,28936	8,64272	42,79565	6,98297	48,26794
23	11,68855	38,07563	10,19572	41,63840	9,26042	44,18128	7,52924	49,72823
24	12,40115	39,36408	10,85636	42,97982	9,88623	45,55851	8,08488	51,17860
25	13,11972	40,64647	11,52398	44,31410	10,51965	46,92789	8,64934	52,61966
26	13,84390	41,92317	12,19815	45,64168	11,16024	48,28988	9,22213	54,05196
27	14,57338	43,19451	12,87850	46,96294	11,80759	49,64492	9,80278	55,47602
28	15,30786	44,46079	13,56471	48,27824	12,46134	50,99338	10,39088	56,89229
29	16,04707	45,72229	14,25645	49,58788	13,12115	52,33562	10,98605	58,30117
30	16,79077	46,97924	14,95346	50,89218	13,78672	53,67196	11,58795	59,70306