W MŠ 22.5.2017

Pr. 1 *a) Aké chyby poznáme pri testovaní štatistických hypotéz a aký je vzťah medzi nimi? Ako súvisí hladina významnosti testu s niektorou z chýb? (5b)*
b) Pri testovaní štatistických hypotéz pre rôzne hladiny významnosti $\alpha \in \{0,1,\ 0,05,\ 0,01\}$ boli vypočítané nasledujúce kritické oblasti (oblasti zamietania H_0): $(-\infty,\ 1,2)$, $(-\infty,\ 1,6)$, $(-\infty,\ 0,8)\}$ Priraďte oblastiam správne hladiny významnosti. (Bez zdôvodnenia rozhodnutia 1b.)

Riešenie 1 *:*
a) hladina významnosti testu je α. Ak zmenšíme α zväčšíme β a naopak.

výsledok testu/skutočnosť	H_0 platí	H_0 neplatí
H_0 zamietame	I. druh $= \alpha$	$\times$
H_0 nezamietame	$\times$	II. druh$= \beta$

b) Ak K_α je oblasť zamietania H_0 na hladine významnosti α, tak platí $P(T \in K_\alpha) = \alpha$, kde T je štatistika, tak pre $K_{\alpha_1} \subseteq K_{\alpha_2}$ platí $\alpha_1 \leqslant \alpha_2$. Stačí obrázok.

$$\alpha = 0,1 \ \Rightarrow (-\infty,\ 1,6)$$

$$\alpha = 0,05 \ \Rightarrow (-\infty,\ 1,2)$$

$$\alpha = 0,01 \ \Rightarrow (-\infty,\ 0,8)$$

Pr. 2 *Nech X_1, X_2, X_3, X_4 je náhodný výber, $E(X_i) = \mu$ a $D(X_i) = \sigma^2$.*
a) Nech $Y = 0,2X_1 + 0,1X_2 + 0,6X_3 + 0,1X_4$. Vypočítajte $E(Y), E(\overline{X})$ a zistite, ktorá z n.p. $Y, \overline{X}$ má menšiu disperziu. (7b).)
b) Nech $X_i \sim Bi(n,\ p)$. Pomocou momentovej metódy odhadnite parameter p. (3b)

Riešenie 2 *:*
Pretože sa jedná o náhodný výber, tak n.p. sú navzájom nezávislé a teda:
a)

$$\begin{aligned} E(Y) =& E(0,2X_1 + 0,1X_2 + 0,6X_3 + 0,1X_4) \\ =& 0,2E(X_1) + 0,1E(X_2) + 0,6E(X_3) + 0,1E(X_4) \\ =& \mu(0,2 + 0,1 + 0,6 + 0,1) = \mu \end{aligned}$$

$$E(\overline{X})) = \frac{1}{4}\sum_i E(X_i) = \frac{1}{4}\sum_i \mu = \mu$$

$$\begin{aligned} D(Y) =& D(0,2X_1 + 0,1X_2 + 0,6X_3 + 0,1X_4) \\ =& 0,2^2D(X_1) + 0,1^2D(X_2) + 0,6^2D(X_3) + 0,1^2D(X_4) \\ =& \sigma^2(0,04 + 0,01 + 0,36 + 0,01) = 0,42\sigma^2 \end{aligned}$$

$$D(\overline{X}) = \frac{\sigma^2}{4} = 0,25\sigma^2 < 0,42\sigma^2 = D(Y).$$

Platí tvrdenie (bolo na prednáške): Ak $X_1, \cdots, X_n$ je náhodný výber, $Y = \sum_i a_iX_i$, kde $\sum_i a_i = 1$, $a_i \in< 0; 1 >$, potom $D(Y)$ je minimálna ak $a_i = \frac{1}{n}$ pre $\forall i$.
(Aritmetický priemer je nevychýlený odhad μ s najmenšou disperziou.)
b) Ak $X_i \sim Bi(n,\ p)$, $E(X_i) = np = \mu \approx \overline{X}$.

$$np = \mu \quad \Rightarrow \quad p \approx \frac{\overline{X}}{n}$$

Pr. 3 *Firma sledovala na webe počet kliknutí na svoju reklamu počas dňa u 100 náhodne vybraných zákazníkov. Nájdite dolný a horný kvartil, medián, modus a zostrojte krabicový (box) graf. Nájdite extremálne hodnoty (outlayers), ak existujú. Výsledky štatistického prieskumu sú v nasledujúcej tabuľke:*

počet kliknutí	1	2	3	4	5	6	15
počet zákazníkov	1	26	24	10	15	20	4

Riešenie 3 *:*
$\overline{x} = 4,2$, *Modus* $Mo = 2$ *Median* $Me = 3$, $Q_L = 2$, $Q_U = 5$,
$R = 5 - 2 = 3$, $1,5R = 4,5$, $Q_L - 4,5 = -2,5 < \min\{1;2;3;4;5;6;15\} = 1$,
$Q_U + 1,5R = 5 + 4,5 = 9,5 < 15$ *vybočujća hodnota je 15.*

počet kliknutí	1	2	3	4	5	6	15	$\sum$
počet zákazníkov	1	26	24	10	15	20	4	100
	$0,01$	$2.0,26$	$3.0,24$	$4.0,1$	$5.0,15$	$6.0,2$	$15.0,04$	$4,2$

Pr. 4 *Generátor náhodných čísel vygeneroval* 100 *čísel z množiny* $\{0,1,2,3\}$. *Otestujte na hladine významnosti* $\alpha = 0,05$, *či ide o náhodný výber z* $Bi\left(3,\ \frac{1}{2}\right)$. *Výsledky sú v nasledujúcej tabuľke:*

číslo	0	1	2	3
početnosť	20	20	50	10

Riešenie 4 *:*

číslo	0	1	2	3
počet.ť f_i	20	20	50	10
očakávaná np_i	$100\frac{1}{8} = 12,5$	$100\frac{3}{8} = 37,5$	$100\frac{3}{8} = 37,5$	$100\frac{1}{8} = 12,5$
$(f_i - np_i)^2$	$7,5^2$	$(-17,5)^2$	$(-12,5)^2$	$(-2,5)^2$
$\frac{(f_i-np_i)^2}{np_i}$	$7,5^2/12.5$	$17,5^2/37,5$	$12,5^2/37,5$	$2,5^2/12,5$
$\frac{(f_i-np_i)^2}{np_i}$	$4,5$	$8,167$	$4,167$	$0,5$

H_0 *dáta majú* $Bi(3,\ 1/2)$ H_1 *dáta nemajú* $Bi(3,\ 1/2)$
dosadíme do vzorca:

$$\chi^2 = \sum_i \frac{(f_i - np_i)^2}{np_i} = 17,334$$

tab. hod. $\chi^2(0,05, \quad ,3) = 7,815$ *Pretože* $17,334 > 7,815$ H_0 *zamietame.*

Pr. 5 *V meste LM sa vyrába pivo Uf. V náhodnom výbere 100 domácností mesta LM sa sledovala spotreba Uf v priebehu roka. Z náhodného výberu sa zistil výberový priemer* $\overline{x} = 18l$ *a smerodajná odchýlka* $s = 8,5\ l$*. Priemerná ročná spotreba Uf na jednu domácnosť v SR je 17,8 l a smerodajná odchýlka* $\sigma = 6l$*. Na hladine vznamnosti* $\alpha = 0,05$*:*
a) otestujte či rozptyl je v LM je taký ako v SR alebo väčší.
b) otestujte či priemerná spotreba Uf v LM je taká istá ako v SR, alebo rôzna.
V oboch prípadoch naformulujte H_0 *a* H_1*.*

Riešenie 5 *:*
a) $H_0\ \sigma^2 = 6^2$ *oproti* $H_1\ \sigma^2 > 6^2$
Dosadíme do vzorca

$$U = \frac{(n-1).S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

Pretože $U = \frac{99.8,5^2}{36} = \frac{7152,75}{36} = 198,688$ *Tab.hodn.* $\chi^2(0,95;99) = 123,225$*. Pretože* $198,688 > 123,225$*,* H_0 *zamietame.*
b) $H_0\ \mu = 17,8$ *oproti* $H_1\ \mu \neq 17,8$
Štatistika

$$T = \frac{\sqrt{100}(\overline{x} - 17,8)}{8,5} \sim t(99).$$

$$T = \frac{10(18 - 17,8)}{8,5} = 0,235.$$

tab.hod. $T(0,975;99) = 1,984$*. Pretože* $|0,235| = 0,235 < 1,984$*,* H_0 *nezamietame.*

Pr. 6 *Máme 2 krabice. V prvej* (K_1) *sú lístky s číslami:* $1,1,2,3$ *a v druhej* (K_2) *s číslami : 2,2,3 Náhodne vytiahneme po jednom lístoku z každej krabice. X hodnota na lístku z krabice* K_1 *a Y hodnota hodnota na na lístku z krabice* K_2*.*
a) Nájdete združené a marginálne rozdelenie pravdepodobnosti náh. vek. (X,Y). b) Vypočítajte $cov(X,Y)$*.*

Riešenie 6 *:*

Združené rozd., pretože X, Y *sú nezávislé, tak* $cov(X,Y) = 0$ *a*

X/Y	2	3	marg. X
1	1/3	1/6	1/2
2	1/6	1/12	1/4
3	1/6	1/12	1/4
marg. Y	2/3	1/3	

Pr. 7 *N.p. X má funkciu hustoty* $f_X(t)$

$$f_X(t) = \begin{cases} \frac{t}{2} & 0 \leqslant t \leqslant a \\ \frac{1}{6}(4-t) & a < t \leqslant 4 \\ 0 & t \notin <0,4> . \end{cases}$$

Vypočítajte čísla a, ak $F_X(a) = 0,25$*. Načrtnite graf funkcie hustoty n.p. X, vypočítajte* $P(X \geqslant 2)$ *a vyznačte na grafe funkcie hustoty.*

Riešenie 7

$$0,25 = F_X(a) = \int_0^a \frac{t}{2}\,dt = \left[\frac{t^2}{4}\right]_0^a = \frac{a^2}{4} \quad \Rightarrow \quad a = 1$$

$$P(X \geqslant 2) = \frac{1}{6}\int_2^4 (4-t)\,dt = \frac{1}{6}\left[4t - \frac{t^2}{2}\right]_2^4 = \frac{1}{6}\left[16 - \frac{16}{2}\right] - \frac{1}{6}\left[8 - \frac{4}{2}\right] = \frac{1}{3}$$