10 Zápočtová písemka

Příklad 10.1. Náhodné veličiny U_1, U_2, U_3 se řídí normálním rozdělením se střední hodnotou 0 a rozptylem 1. Určete rozdělení následujících náhodných veličin:

$$Y = U_1 + U_2 + U_3 - 2 \qquad V = \frac{\sqrt{2} \cdot U_2}{\sqrt{U_1^2 + U_3^2}}$$
$$Z = \frac{(U_1 + U_2 - U_3)^2}{3} \qquad W = \frac{U_1^2}{U_2^2}$$

$[Y \sim N(-2, 3), Z \sim \chi^2(1), V \sim t(2), W \sim F(1, 1)]$

Příklad 10.2. Určete střední hodnotu náhodné veličiny $V = X^2 + 3XY - 0.5Y^2$, přičemž známe $EX = 2$, $DX = 4$, $EY = -1$, $DY = 9$ a $R(X, Y) = 0.1$.

$[EV = -1.2]$

Příklad 10.3.

- Co musí platit, aby odhad T byl nestranným odhadem parametru ϑ?
- Máme dva nestranné odhady T a S parametru ϑ, který z odhadů je lepší a proč?

[$ET = \vartheta$, lepší odhad má menší rozptyl]

Příklad 10.4. Stroj vyrábí olověné broky. Průměr broku je náhodnou veličinou měřenou v milimetrech, o níž předpokládáme, že se řídí rozdělením $N(5, 0.05^2)$. Kolik procent broků je při kontrole vyřazeno, jestliže broky, lišící se o více než 0.1 mm od střední hodnoty, jsou vyřazovány?

[4.55 %]

Příklad 10.5. Vypočtěte koeficient korelace mezi počtem shybů X_i a počtem kliků Y_i naměřených u skupinky deseti chlapců

shyby x_i	1	3	2	0	5	6	1	4	3	5
kliky y_i	2	3	3	0	8	5	1	6	7	5

$[r_{xy} = 0.8043]$

Příklad 10.6. Pomocí Čebyševovy nerovnosti odhadněte pravděpodobnost $P(|X - EX| < 3DX)$

$[P \geq 1 - \frac{1}{9DX}]$

u	0.05	0.1	0.5	1	1.5	2	2.5	3
$F(u)$	0.5200	0.5398	0.6915	0.8413	0.9332	0.9773	0.9938	0.9986

11 Intervalové odhady parametrů

Příklad 11.1. Za předpokladu, že čas příchodu studentů na přednášku se řídí normálním rozdělením pravděpodobnosti s naměřenými hodnotami (viz příklad 8.1), určete

a) 99.5% oboustranný interval spolehlivosti pro střední hodnotu času příchodu na přednášku,

b) 99% pravostranný interval spolehlivosti pro rozptyl času příchodu na přednášku.

$[(3.742, 5.658)]$

Příklad 11.2. Mějme 2 nezávislé náhodné výběry

$$(X_1, \ldots, X_{n_1}) \sim N(\mu_1, \sigma^2), \qquad n_1 > 1,$$
$$(Y_1, \ldots, Y_{n_2}) \sim N(\mu_2, \sigma^2), \qquad n_2 = 1.$$

Zkonstruujte intervaly spolehlivosti pro rozdíl $\mu_1 - \mu_2$ a pro rozptyl σ^2.

Příklad 11.3. Odvoďte interval spolehlivosti pro parametrickou funkci $2\mu_1 - \mu_2 - \mu_3$, máme-li k dispozici 3 nezávislé náhodné výběry

$$
\begin{aligned}
(X_1, \dots, X_{n_1}) &\sim N(\mu_1, \sigma^2), \qquad n_1 > 1,\\
(Y_1, \dots, Y_{n_2}) &\sim N(\mu_2, \sigma^2), \qquad n_2 > 1,\\
(Z_1, \dots, Z_{n_3}) &\sim N(\mu_3, \sigma^2), \qquad n_3 > 1.
\end{aligned}
$$

$[2\overline{X} - \overline{Y} - \overline{Z} \pm \frac{(n_1-1)S_1^2+(n_2-1)S_2^2+(n_3-1)S_3^2}{n_1+n_2+n_3-3} \cdot \sqrt{\frac{4}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}} \cdot t_{1-\alpha/2}(m)]$

Příklad 11.4. Mějme tři nezávislé náhodné výběry

$$\begin{aligned}(X_1,\dots,X_{n_1}) &\sim N(\mu_1,\sigma_1^2), & n_1 > 1,\\ (Y_1,\dots,Y_{n_2}) &\sim N(\mu_2,\sigma_2^2), & n_2 = 1,\\ (Z_1,\dots,Z_{n_3}) &\sim N(\mu_3,\sigma_3^2), & n_3 > 1.\end{aligned}$$

Zkonstruujte dolní interval spolehlivosti pro neznámý parametr σ/σ_3, za předpokladu $\sigma_1 = \sigma_2$.

$$\left[\sigma/\sigma_3 \in \left(\frac{S_X}{S_Z}\frac{1}{\sqrt{F_{1-\alpha}(n_1-1,n_3-1)}},\infty\right)\right]$$

Příklad 11.5. Odvoďte 95% interval spolehlivosti pro rozdíl středních hodnot a podíl rozptylů doby montáže určité součásti pro dva dělníky. Dobu montáže považujeme za náhodnou veličinu s normálním rozdělením.

Příklad 11.6. Firma v Iowě má 500 lanek. Testem 40 náhodně vybraných lanek se zjistilo, že pevnost v tahu má střední hodnotu 2400 lb a směrodatnou odchylku 150 lb. Jaké jsou 95% a 99% intervaly spolehlivosti pro odhad průměrné pevnosti v tahu pro zbývajících 460 lanek? S jakou pstí můžeme říct, že střední hodnota průměrné pevnosti v tahu pro zbývajících 460 lanek je 2400 ± 35 lb?

$[2400 \pm 45, 2400 \pm 59, 87.6\%]$

Příklad 11.7. Ze základního souboru byl proveden náhodný výběr s naměřenými intervalovými hodnotami a jejich četnostmi sledovaného znaku

x_i	(15,17]	(17,19]	(19,21]	(21,23]	(23,25]	(25,27]
n_i	10	30	50	70	60	30

Určete

a) interval, ve kterém se nachází střední hodnota μ s pstí 0.95,

b) hodnotu, pod kterou se s pstí 0.95 střední hodnota nedostane,

c) hranice, v nichž se nachází směrodatná odchylka s pstí 0.95,

d) hodnotu, kterou s pstí 0.95 rozptyl nepřekročí.

$[(21.5094, 22.1706), 21.5629, (2.4398, 2.9093), 8.2149]$

Příklad 11.8. Nechť $Y \sim Bi(300, p)$, $y = 75$ je realizace Y. Stanovte přibližný 90% interval spolehlivosti pro p.

$[(0.209, 0.291)]$

Domácí úkol 11.1. Opakovanými měřeními byla zjištěna tloušťka vlákna: 210, 217, 209, 216, 216, 215, 220, 214, 213 $(10^{-3}$ mm). Je známo, že měření mají rozdělení $N(\mu, 25)$. Nalezněte 95% interval spolehlivosti pro μ.

$[(211.1, 217.7)]$

Domácí úkol 11.2. Odvoďte $(1-\alpha)\%$ interval spolehlivosti pro σ^2, máte-li náhodné výběry rozsahů n_X a n_Y ze dvou nezávislých rozdělení $N(\mu_X, \sigma^2)$ a $N(\mu_Y, \sigma^2)$. Využijte: $(n_X S_X^2 + n_Y S_Y^2)/\sigma^2 \sim \chi^2(n_X + n_Y - 2)$.

12 Testování hypotéz

Příklad 12.1. Sestrojte oboustranný 95% interval spolehlivosti pro rozptyl normálně rozdělené náhodné veličiny na základě údajů z náhodného výběru: 42, 48, 60, 43, 36, 50, 52, 38, 56, 45. Jak dopadne test hypotézy $\sigma^2 = 60$ na hladině významnosti $\alpha = 0.05$?

[$IS = (27.966, 197.037)$, hypotézu nezamítáme]

Příklad 12.2. Do obchodu jsou dodávány balíčky kávy o předepsané hmotnosti 500 g. Balíčky jsou plněny automaticky. Automat byl seřízen tak, aby směrodatná odchylka hmotnosti balíčku činila 10 g. Předpokládejme, že hmotnost plněných balíčků je náhodná veličina s normálním rozdělením. Chceme zjistit, zda během času nedošlo ke zhoršení přesnosti při plnění balíčků, tj. ke zvětšení směrodatné odchylky σ. Rozlišme dva případy:

a) Nemáme důvod předpokládat, že by nebyla dodržována předepsaná hmotnost balíčků.

b) Vezmeme v úvahu, že se mohla změnit i střední hodnota hmotnosti balíčků.

Proveďme rozhodnutí na základě 10 náhodně vybraných balíčků, u nichž jsme převážením zjistili hmotnosti (v gramech): 473, 477, 482, 488, 489, 491, 492, 498, 503, 507.

[pro $\alpha = 0.05$ H_0 zamítáme; pro $\alpha = 0.05$ H_0 nezamítáme]

Příklad 12.3. Velkoobchod dodává balíčky sušenek, o jejichž hmotnosti (v gramech) lze předpokládat, že má rozdělení $N(\mu, 25)$. Z dodávky bylo náhodně vybráno 20 balíčků a bylo zjištěno, že průměrná hmotnost je $61,2$ g se směrodatnou odchylkou 2 g. Na hladině významnosti $\alpha = 0.05$ testujte hypotézu, že střední hodnota hmotnosti balíčku je rovna 60 g.

[H_0 zamítáme]

Příklad 12.4. Na základě testu máme na 5% hladině významnosti rozhodnout, zda produkce vajec plemene kornyšek černých je nižší než plemene leghornek bílých. Náhodně jsme vybrali 50 kornyšek a 40 leghornek, u nichž byla zjištěna roční produkce vajec. Byl vypočten roční průměr produkce na slepici – kornyška 275, leghornka 280, z dřívějška jsou známy rozptyly $\sigma^2_{kor} = 48$, $\sigma^2_{leg} = 41$.

[H_0 zamítáme, kornyšky mají horší produkci vajec než leghornky]

Příklad 12.5. Máme k dispozici realizace dvou nezávislých náhodných výběrů z $N(\mu_X, \sigma^2)$ a $N(\mu_Y, \sigma^2)$ s rozsahy $n_X = 10$ a $n_Y = 15$. Byly vypočteny realizace $\overline{X} = 120.56$, $\overline{Y} = 124.13$, $S_X^2 = 9.14$, $S_Y^2 = 8.95$. Můžeme na základě těchto výsledků zamítnout hypotézu $H_0 : \mu_X = \mu_Y$ proti $H_1 : \mu_X \neq \mu_Y$ při riziku $\alpha = 0.1$?

[ano]

Příklad 12.6. Bylo provedeno měření obsahu SiO_2 ve strusce dvěma metodami.
analyticky: $20.1, 19.6, 20.0, 19.9, 20.1$
a *fotokolorometricky*: $20.9, 20.1, 20.6, 20.5, 20.7, 20.5$.
Je mezi rozptyly výsledků jednotlivých metod podstatný rozdíl?

[není]

Příklad 12.7. Pro kontrolu správného objemu pivních sklenic vyráběných v jisté firmě bylo provedeno měření pomocí etalonu s přesnou hodnotou $\mu_0 = 0.50$ l. Byly získány tyto výsledky: 0.48, 0.49, 0.51, 0.46, 0.52, 0.49, 0.50, 0.47, 0.48 (předpokládáme jejich normalitu). Lze považovat pozorované odchylky za náhodné chyby nebo je důvod k podezření na přítomnost systematické chyby?

[zřejmě jde o náhodné chyby]

Příklad 12.8. Na 9 pokusných polích byly zjišťovány výnosy odrůdy kukuřice. Jsou srovnatelné výnosy z let 1980 a 1982?

rok \ pole	A	B	C	D	E	F	G	H	I
1980	4.23	5.09	4.55	5.31	5.04	5.45	4.76	5.57	4.97
1982	5.83	7.05	6.00	4.86	6.31	5.22	6.44	4.71	6.09
rozdíl	1.60	1.96	1.45	-0.45	1.27	-0.23	1.68	-0.86	1.12

[výnosy nejsou srovnatelné]

Domácí úkol 12.1. Ověřte, zda metody z příkladu 12.6 dávají v průměru stejné výsledky.

[nedávají]

Domácí úkol 12.2. Bylo vybráno 6 nových vozů téže značky a po určité době bylo zjištěno o kolik milimetrů se sjely jejich levé a pravé přední pneumatiky. Naměřené rozdíly: 0.3, -0.1, 0.2, -0.2, 0.2, 0.1. Za předpokladu normality náhodných veličin, které udávají velikost opotřebení pravé a levé přední pneumatiky, rozhodněte, zda se obě přední pneumatiky sjíždí stejnoměrně. Hladinu významnosti volte $90\,\%$.

[nezam.]

13 Testování hypotéz

Příklad 13.1. Dobrovolníci při pokusu odhadovali dobu trvání dvou hodin. Byly získány tyto výsledky (v minutách)

$$135, 108, 152, 139, 145, 126, 95, 176, 138, 149, 110, 166, 152, 143.$$

Na 5% hladině významnosti testujte hypotézu, že odhad délky časového intervalu je v těchto podmínkách nadhodnocený.

[zam.]

Příklad 13.2. Je známo, že rozptyl věku, ve kterém se dívky vdávají je 16. Kolegu sociologa zajímá, zda dívky ze Seattlu mají tendenci se vdát dříve než dívky z Puyallupu. Bylo vybráno po deseti dívkách a zjištěny následující hodnoty průměrů $M_S = 22.2$ let, $M_P = 25.8$. Může sociolog zamítnout hypotézu, že se dívky z obou měst vdávají ve stejném věku? Použijte hladinu významnosti 0.1.

[ano, může zamítnout]

Příklad 13.3. U 15 jabloní bylo zjištěno stáří stromu v letech X a sklizeň Y v roce 1983 (kg)

x_i	2	3	3	4	4	4	5	5	5	6	6	6	6	6	10
y_i	4	4	4	5	5	5	5	6	6	6	6	6	9	10	9

Na 5% hladině významnosti rozhodněte, zda existuje korelace mezi stářím stromu a sklizní jablek.

[ano]

Testová statistika

$$T = \frac{R_{XY} \cdot \sqrt{n-2}}{\sqrt{1 - R_{XY}^2}} \sim t(n-2), \qquad R_{XY} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2 \cdot \sum (y_i - \overline{y})^2}}$$

Příklad 13.4. Linka městské autobusové dopravy má v době dopravní špičky průměrnou rychlost v centru města 8 km/h. Zjišťovalo se, zda by změna trasy vedla ke zvýšení průměrné rychlosti. Nová trasa byla projeta v desíti náhodně vybraných dnech a v době dopravní špičky byly zjištěny tyto průměrné rychlosti: 8.5, 9.5, 7.8, 8.2, 9.0, 7.5, 8.2, 7.8, 9.0, 8.5. Na hladině významnosti a) 0.01, b) 0.05 zjistěte, jestli změna trasy vede ke zvýšení průměrné rychlosti za předpokladu normálního rozdělení psti průměrné rychlosti.

$[a = nez., b = zam.]$

Příklad 13.5. Výrobce nového léku prohlašuje, že v 90 % případů uleví alergikovi po dobu 8 hodin. V náhodném výběru 200 alergiků, kterým byl aplikován tento lék, se ulevilo 160 z nich. Rozhodněte, zda prohlášení výrobce o účinnosti léku je oprávněné na hladině významnosti 0.01.

[není oprávněné]

Příklad 13.6. Dva výběry obsahují 21, resp. 9 pozorování s rozptylem 16, resp. 8. Testujte hypotézu, že rozptyl první skupiny je větší než rozptyl druhé skupiny na hladině významnosti 0.05.

[nez.]

Příklad 13.7. Roční teplota města je získávána jako průměr průměrných hodnot teploty každý 15. den měsíce. Směrodatná odchylka roční teploty za 100 let měření byla 16 °F. Během posledních 15 let byla přepočítána hodnota směrodatné odchylky na hodnotu 10 °F. Testujte hypotézu, že teplota ve městě je méně proměnlivá než v minulosti, použijte hladinu významnosti 0.05, resp. 0.01.

[0.05: pokles je signifikantni, 0.01: neni sign.]

Příklad 13.8. Pro bavlněnou přízi je předepsána horní mez variability pevnosti vlákna: rozptyl pevnosti (která má $N(\mu, \sigma^2)$) nemá překročit $\sigma_0^2 = 0.36$. Při zkoušce 16 vzorků byly zjišteny výsledky 2.22, 3.54, 2.37, 1.66, 4.74, 4.82, 3.21, 5.44, 3.23, 4.79, 4.85, 4.05, 3.48, 3.89, 4.90, 5.37. Je důvod k podezření na vyšší nestejnoměrnost než je stanoveno?

[zamítáme $H_0 : s = s_0 = 0.36$, proti $H_1 : s > s_0$ na 5% hlad. význ.]

Domácí úkol 13.1. V pokusném testu zadaném v mateřských školkách byl průměrný počet získaných bodů 74.5 se směrodatnou odchylkou 8 bodů. V jedné konkrétní "superškolce", kde se testu zúčastnilo 200 dětí, byl průměrný počet získaných bodů 75.9. Diskutujte signifikanci tohoto výsledku na hladině významnosti 0.05, když použijete a) jednostrannou, b) oboustrannou alternativní hypotézu. Vysvětlete své závěry na základě těchto testů.

[je signif. v obou testech]

Domácí úkol 13.2. Denní přírůstky váhy selat byly při krmení směsí A : 62, 54, 55, 60, 53, 58, u směsi B : 52, 56, 50, 49, 51. Je mezi nimi rozdíl?

[ano, je mezi nimi rozdil]

14 Opakování před zkouškou

Písemka z června 2003

Příklad 14.1. Urna obsahuje n koulí, bílé a černé. Byla naplněna tak, že někdo hodil $n$krát kostkou, když padla ″6″ vložil do urny bílou kouli, když ″6″ nepadla, vložil do urny černou kouli. Z takto naplněné urny byly postupně náhodně vybrány dvě koule, přičemž po prvním tahu byla vytažená koule vrácena zpět.

a) Jaká je pravděpodobnost, že obě vytažené koule byly černé?

b) Ukázalo se, že obě vytažené kule byly bílé. Jaká je pravděpodobnost, že urna před tímto tahem obsahovala právě jednu černou kouli a ostatní bílé?

Příklad 14.2. Nechť (X_1, X_2) je absolutně spojitý náhodný vektor s hustotou

$$f(x_1, x_2) = \begin{cases} x_1\,\mathrm{e}^{-x_1(1+x_2)} & x_1 > 0, x_2 > 0 \\ 0 & \text{jinak.} \end{cases}$$

a) Určete podmíněnou hustotu $f(x_1|x_2)$.

b) Určete podmíněnou střední hodnotu $E\left(\frac{1}{X_1}|X_2\right)$.

c) Odvoďte distribuční funkci transformované náhodné veličiny $Y = X_1 \cdot X_2$.

Příklad 14.3. Jestliže X_1, X_2, X_3 jsou nezávislé náhodné veličiny se stejným normálním rozdělením $N(0,1)$,

a) Vypočtěte korelační koeficient $R(1 - 2X_1 + X_2,\ 3 + 2X_1 - X_3 + X_2)$.

b) Stanovte čtvrtý centrální moment $E(X_1 + 2X_2 + X_3)^4$. [Využijte skutečnosti, že 3. centrální moment, resp. 4. centrální moment rozdělení $N(0,1)$ je roven 0, resp. 3.]

c) Určete pravděpodobnost $P(-2 \leq (4X_1 - 3X_2) \leq 5)$.

Příklad 14.4. Nechť $X_1, X_2, \ldots, X_n$ je náhodný výběr z Poissonova rozdělení $Po(\lambda)$. Pravděpodobnostní funkce Poissonova rozdělení je pro $i = 1, \ldots, n$ tvaru

$$p(x_i) = \begin{cases} \lambda^{x_i}\,\mathrm{e}^{-\lambda} \big/ x_i! & x_i = 0, 1, 2, \ldots \\ 0 & \text{jinak.} \end{cases}$$

a) Určete charakteristickou funkci $\psi_{X_i}(t)$.

b) Odvoďte charakteristickou funkci výběrového průměru $\overline{X}$.

c) Určete limitu $\psi_{\overline{X}}(t)$ pro $n \to \infty$ a dále odvoďte rozptyl tohoto rozdělení.

d) Dokažte, že výběrový průměr je nestranným odhadem rozptylu Poissonova rozdělení.

Něco navíc

Regresní přímka

Příklad 15.1. Body proložte lineární funkci a určete reziduální součet čtverců.

x_i	40	64	34	15	57	45
y_i	33	46	23	12	56	40