6. Parametrické úlohy o jednom náhodném výběru

6.1. Náhodný výběr z normálního rozložení

Mnoho náhodných veličin, s nimiž se setkáváme ve výzkumu i praxi, se řídí normálním rozložením. Za jistých předpokladů obsažených v centrální limitní větě se dá rozložení jiných náhodných veličin aproximovat normálním rozložením. Proto je zapotřebí věnovat velkou pozornost právě náhodným výběrům z normálního rozložení.

6.1.1. Rozložení statistik odvozených z výběrového průměru a výběrového rozptylu

Nechť $X_1, ..., X_n$ je náhodný výběr z rozložení $N(\mu, \sigma^2)$. Pak platí

a) Výběrový průměr M a výběrový rozptyl S^2 jsou stochasticky nezávislé.

b) $M \sim N(\mu, \frac{\sigma^2}{n})$, tedy $U = \frac{M-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$.

(Pivotová statistika U slouží k řešení úloh o μ, když σ^2 známe.)

c) $K = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n\text{-}1)$.

(Pivotová statistika K slouží k řešení úloh o σ^2, když μ neznáme.)

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

(Tato pivotová statistika slouží k řešení úloh o σ^2, když μ známe.)

e) $T = \frac{M-\mu}{\frac{S}{\sqrt{n}}} \sim t(n\text{-}1)$.

(Pivotová statistika T slouží k řešení úloh o μ, když σ^2 neznáme.)

Důkaz:
ad a) Nebudeme provádět.
ad b) M je lineární kombinace náhodných veličin s normálním rozložením, má tedy normální rozložení s parametry $E(M) = \mu$, $D(M) = \sigma^2/n$. U se získá standardizací M.
ad c) Vhodnou úpravou S^2, kde použijeme obrat $X_i\text{-}M = (X_i\text{-}\mu) - (M\text{-}\mu)$, lze statistiku K vyjádřit jako součet kvadrátů n-1 stochasticky nezávislých náhodných veličin se standardizovaným normálním rozložením. Tento součet se řídí rozložením $\chi^2(n\text{-}1)$.
ad d) Tato statistika je součet kvadrátů n stochasticky nezávislých náhodných veličin se standardizovaným normálním rozložením, řídí se tedy rozložením $\chi^2(n)$.
ad e) $U \sim N(0, 1)$, $K \sim \chi^2(n\text{-}1)$ jsou stochasticky nezávislé, protože M a S^2 jsou stochasticky nezávislé, tudíž statistika $T = \frac{U}{\sqrt{\frac{K}{n-1}}} = \frac{M-\mu}{\frac{S}{\sqrt{n}}} \sim t(n\text{-}1)$.

Příklad: Hmotnost jedné porce kávy považujeme za náhodnou veličinu $X \sim N(7g, 0{,}25g^2)$. Jaká je pravděpodobnost, že k přípravě 28 porcí kávy postačí dva 100 g balíčky?

Řešení: $X_1, ..., X_{28}$ je náhodný výběr z N(7; 0,25). Počítáme

$$P\left(\sum_{i=1}^{28} X_i \le 200\right) = P\left(\frac{1}{28}\sum_{i=1}^{28} X_i \le \frac{200}{28}\right) = P\left(M \le \frac{200}{28}\right) = P\left(\frac{M-7}{\frac{0,5}{\sqrt{28}}} \le \frac{\frac{200}{28}-7}{\frac{0,5}{\sqrt{28}}}\right) =$$

$$= P(U \le 1{,}51) = \Phi(1{,}51) = 0{,}9345.$$

S pravděpodobností 93,45% můžeme předpokládat, že k přípravě 28 porcí kávy postačí dva 100 g balíčky.

Příklad: Odběratel provede kontrolu stejnorodosti dodávky výrobků tak, že změří sledovaný rozměr u 25 náhodně vybraných výrobků. Dodávku přijme, jestliže výběrová směrodatná odchylka se bude realizovat hodnotou menší nebo rovnou 0,2 mm. Je známo, že sledovaný rozměr výrobku má rozložení N(50 mm; $0{,}263^2 mm^2$). Jaká je pravděpodobnost přijetí dodávky?

Řešení: $X_1, ..., X_{25}$ je náhodný výběr z N(50; $0{,}263^2$). Počítáme $P(S \le 0{,}2) = P(S^2 \le 0{,}04) =$ $P\left(\frac{(n-1)S^2}{\sigma^2} \le \frac{(n-1)0{,}04}{\sigma^2}\right) = P\left(K \le \frac{24 \cdot 0{,}04}{0{,}263^2}\right) = P(K \le 13{,}879)$, tedy číslo 13,879 je α-kvantil Pearsonova rozložení $\chi^2(24)$. V tabulkách kvantilů Pearsonova rozložení najdeme, že α =0,05. S pravděpodobností pouhých 5% lze očekávat, že odběratel přijme dodávku.

6.1.2. Intervaly spolehlivosti pro parametry μ, σ^2

Uvedeme přehled vzorců pro meze 100(1-α)% empirických intervalů spolehlivosti pro parametry jednoho normálního rozložení

a) Interval spolehlivosti pro μ, když σ^2 známe

Oboustranný: $(d, h) = (m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2})$

Levostranný: $(d, \infty) = (m - \frac{\sigma}{\sqrt{n}} u_{1-\alpha}, \infty)$

Pravostranný: $(-\infty, h) = (-\infty, m + \frac{\sigma}{\sqrt{n}} u_{1-\alpha})$

b) Interval spolehlivosti pro μ, když σ^2 neznáme

Oboustranný: $(d, h) = (m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1))$

Levostranný: $(d, \infty) = (m - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1), \infty)$

Pravostranný: $(-\infty, h) = (-\infty, m + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1))$

c) Interval spolehlivosti pro σ^2, když μ neznáme

Oboustranný: $(d, h) = \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right)$

Levostranný: (d, ∞) = $\left(\frac{(n-1)s^2}{\chi^2_{1-\alpha}(n-1)}, \infty \right)$

Pravostranný: (-∞, h) = $\left(-\infty, \frac{(n-1)s^2}{\chi^2_{\alpha}(n-1)} \right)$

d) Interval spolehlivosti pro σ^2, když μ známe

Oboustranný: (d, h) = $\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\alpha/2}(n)}, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\alpha/2}(n)} \right)$

Levostranný: (d, ∞) = $\left(\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{1-\alpha}(n)}, \infty \right)$

Pravostranný: (-∞, h) = $\left(-\infty, \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\chi^2_{\alpha}(n)} \right)$

Příklad: 10 krát nezávisle na sobě byla změřena jistá konstanta μ. Výsledky měření byly: 2 1,8 2,1 2,4 1,9 2,1 2 1,8 2,3 2,2. Tyto výsledky považujeme za číselné realizace náhodného výběru $X_1, ..., X_{10}$ z rozložení $N(\mu, \sigma^2)$, kde parametry μ, σ^2 neznáme. Najděte 95% empirický interval spolehlivosti pro μ, a to

a) oboustranný,
b) levostranný,
c) pravostranný.

Řešení: m = 2,06, s^2 = 0,0404, s = 0,2011, α = 0,05, $t_{0,975}(9)$ = 2,2622, $t_{0,95}(9)$ = 1,8331.

ad a) $d = m - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 2,06 - \frac{0,2011}{\sqrt{10}} 2,2622 = 1,92$

$h = m + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) = 2,06 + \frac{0,2011}{\sqrt{10}} 2,2622 = 2,20$

1,92 < μ < 2,20 s pravděpodobností aspoň 0,95.

ad b) $d = m - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1) = 2,06 - \frac{0,2011}{\sqrt{10}} 1,8331 = 1,94$

1,94 < μ s pravděpodobností aspoň 0,95.

ad c) $h = m + \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1) = 2,06 + \frac{0,2011}{\sqrt{10}} 1,8331 = 2,18$

μ < 2,18 s pravděpodobností aspoň 0,95.

6.1.3. Testování hypotéz o parametrech μ, σ^2

a) Nechť $X_1, ..., X_n$ je náhodný výběr $N(\mu, \sigma^2)$, kde σ^2 známe. Nechť $n \geq 2$ a c je konstanta. Test H_0: $\mu = c$ proti H_1: $\mu \neq c$ se nazývá z-test.

b) Nechť $X_1, ..., X_n$ je náhodný výběr $N(\mu, \sigma^2)$, kde σ^2 neznáme. Nechť $n \geq 2$ a c je konstanta. Test H_0: $\mu = c$ proti H_1: $\mu \neq c$ se nazývá jednovýběrový t-test.

c) Nechť $X_1, ..., X_n$ je náhodný výběr $N(\mu, \sigma^2)$, kde μ neznáme. Nechť $n \geq 2$ a c je konstanta. Test H_0: $\sigma^2 = c$ proti H_1: $\sigma^2 \neq c$ se nazývá test o rozptylu.

6.1.4. Provedení testů o parametrech μ, σ^2 pomocí kritického oboru

a) Provedení z-testu

H_0: $\mu = c$ proti H_1: $\mu \neq c$ (resp. H_1: $\mu < c$ resp. H_1: $\mu > c$) zamítáme na hladině významnosti α,

jestliže $\left|\dfrac{m-c}{\frac{\sigma}{\sqrt{n}}}\right| \geq u_{1-\alpha/2}$ (resp. $\dfrac{m-c}{\frac{\sigma}{\sqrt{n}}} \leq -u_{1-\alpha}$ resp. $\dfrac{m-c}{\frac{\sigma}{\sqrt{n}}} \geq u_{1-\alpha}$).

b) Provedení jednovýběrového t-testu

H_0: $\mu = c$ proti H_1: $\mu \neq c$ (resp. H_1: $\mu < c$ resp. H_1: $\mu > c$) zamítáme na hladině významnosti α,

jestliže $\left|\dfrac{m-c}{\frac{s}{\sqrt{n}}}\right| \geq t_{1-\alpha/2}(n-1)$ (resp. $\dfrac{m-c}{\frac{s}{\sqrt{n}}} \leq -t_{1-\alpha}(n-1)$ resp. $\dfrac{m-c}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha}(n-1)$).

c) Provedení testu o rozptylu

H_0: $\sigma^2 = c$ proti H_1: $\sigma^2 \neq c$ (resp. H_1: $\sigma^2 < c$ resp. H_1: $\sigma^2 > c$) zamítáme na hladině významnosti α, jestliže $\dfrac{(n-1)s^2}{c} \leq \chi^2_{\alpha/2}(n\text{-}1)$ nebo $\dfrac{(n-1)s^2}{c} \geq \chi^2_{1-\alpha/2}(n\text{-}1)$ (resp. $\dfrac{(n-1)s^2}{c} \leq \chi^2_{\alpha}(n\text{-}1)$ nebo $\dfrac{(n-1)s^2}{c} \geq \chi^2_{1-\alpha}(n\text{-}1)$).

Příklad: Podle údajů na obalu čokolády by její čistá hmotnost měla být 125g. Výrobce dostal několik stížností od kupujících, ve kterých tvrdili, že hmotnost čokolád je nižší než deklarovaných 125g. Z tohoto důvodu oddělení kontroly náhodně vybralo 50 čokolád a zjistilo, že jejich průměrná hmotnost je 122g a směrodatná odchylka 8,6g. Za předpokladu, že hmotnost čokolád se řídí normálním rozložením, můžeme na hladině významnosti 0,01 považovat stížnosti kupujících za oprávněné?

Řešení: $X_1, ..., X_{50}$ je náhodný výběr z $N(\mu, \sigma^2)$. Testujeme hypotézu H_0: $\mu = 125$ proti levostranné alternativě H_1: $\mu < 125$. Protože neznáme rozptyl σ^2, použijeme jednovýběrový t-test. Testové kritérium $\dfrac{m-c}{\frac{s}{\sqrt{n}}} = \dfrac{122-125}{\frac{8,6}{\sqrt{50}}} = -2,4667$. Absolutní hodnotu testového kritéria porovnáme s kvantilem $t_{0,99}(49) = 2,4049$. Jelikož $2,4667 \geq 2,4049$, zamítáme nulovou hypotézu na hladině významnosti 0,01. Stížnosti kupujících tedy lze považovat za oprávněné.

6.2. Náhodný výběr z dvourozměrného normálního rozložení

Nechť $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ je náhodný výběr z rozložení $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$, přičemž $n \geq 2$. Označíme $\mu = \mu_1 - \mu_2$ a zavedeme rozdílový náhodný výběr $Z_1 = X_1 - Y_1, \ldots, Z_n = X_n - Y_n$. Vypočteme $M = \frac{1}{n}\sum_{i=1}^{n} Z_i$, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(Z_i - M)^2$.

6.2.1. Interval spolehlivosti pro parametr μ

Viz 6.1.2.(b)

6.2.2. Párový t-test

Testujeme H_0: $\mu_1 - \mu_2 = 0$ (tj. $\mu = 0$) proti H_1: $\mu_1 - \mu_2 \neq 0$ (tj $\mu \neq 0$). Viz 7.1.3.(b)

Příklad: Na 10 automobilech stejného typu se testovaly dva druhy benzínu lišící se oktanovým číslem. U každého automobilu se při průměrné rychlosti 90 km/h měřil dojezd (tj. dráha, kterou ujede na dané množství benzínu) při použití každého z obou druhů benzínu. Výsledky:

Číslo auta	1	2	3	4	5	6	7	8	9	10
benzín A	17,5	20,0	18,9	17,9	16,4	18,9	17,2	17,5	18,5	18,2
benzín B	17,8	20,8	19,5	18,3	16,6	19,5	17,5	17,9	19,1	18,6

Za předpokladu, že dojezd se řídí normálním rozložením, testujte na hladině významnosti 0,05 hypotézu, že rozdíl středních hodnot dojezdu při dvou druzích benzínu se neliší.

Řešení: Přejdeme k rozdílovému náhodnému výběru. Označíme $\mu = \mu_1 - \mu_2$. Testujeme hypotézu H_0: $\mu = 0$ proti H_1: $\mu \neq 0$ na hladině významnosti 0,05. Vypočteme m = -0,46, s = 0,1838 a testové kritérium $t_0 = -7{,}9148$. Absolutní hodnotu testového kritéria porovnáme s kvantilem $t_{0,975}(9) = 2{,}2622$. Protože $7{,}9148 \geq 2{,}2622$, zamítáme nulovou hypotézu na hladině významnosti 0,05.

6.3. Náhodný výběr z alternativního rozložení

6.3.1. Rozložení statistiky odvozené z výběrového průměru

Nechť $X_1, \ldots, X_n$ je náhodný výběr z rozložení $A(\vartheta)$ a nechť je splněna podmínka $n\vartheta(1-\vartheta) > 9$. Pak statistika $U = \frac{M - \vartheta}{\sqrt{\frac{\vartheta(1-\vartheta)}{n}}}$ konverguje v distribuci k náhodné veličině se standardizovaným normálním rozložením. (Říkáme, že U má asymptoticky rozložení N(0,1) a píšeme $U \approx N(0,1)$.)

Důkaz: Protože $X_1, \ldots, X_n$ je náhodný výběr z rozložení $A(\vartheta)$, bude mít statistika $Y_n = \sum_{i=1}^{n} X_i$ rozložení $Bi(n, \vartheta)$. Podle Moivre-Laplaceovy věty má standardizovaná náhodná veličina $\frac{Y_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}$ asymptoticky rozložení N(0,1). Výběrový průměr $M = \frac{1}{n}\sum_{i=1}^{n} X_i$ má střední hodnotu ϑ a rozptyl ϑ (1-ϑ)/n. Jelikož M je lineární kombinací statistiky Y_n, bude standardizovaná veličina $U = \frac{M-\vartheta}{\sqrt{\frac{\vartheta(1-\vartheta)}{n}}} \approx N(0,1)$.

6.3.2. Asymptotický interval spolehlivosti pro parametr ϑ

Pokud rozptyl D(M) = ϑ (1- ϑ)/n nahradíme odhadem M(1-M)/n, konvergence náhodné veličiny U k veličině s rozložením N(0,1) se neporuší. Tedy

$$\forall \vartheta \in \Xi : 1-\alpha \le P\left(-u_{1-\alpha/2} < \frac{M-\vartheta}{\sqrt{\frac{M(1-M)}{n}}} < u_{1-\alpha/2} \right) =$$

$$= P\left(M - \sqrt{\frac{M(1-M)}{n}} u_{1-\alpha/2} < \vartheta < M + \sqrt{\frac{M(1-M)}{n}} u_{1-\alpha/2} \right)$$

Meze 100(1-α)% asymptotického empirického intervalu spolehlivosti pro parametr ϑ jsou: $d = m - \sqrt{\frac{m(1-m)}{n}} u_{1-\alpha/2}$, $h = m + \sqrt{\frac{m(1-m)}{n}} u_{1-\alpha/2}$.

Příklad: Náhodně bylo vybráno 100 osob a zjištěno, že 34 z nich používá zubní kartáček zahraniční výroby. Najděte 95% asymptotický interval spolehlivosti pro podíl osob používajících zahraniční zubní kartáčky.

Řešení: Zavedeme náhodné veličiny $X_1, ..., X_{100}$, přičemž $X_i = 1$, když i-tá osoba používá zahraniční zubní kartáček a $X_i = 0$ jinak, i = 1, ..., 100. Tyto náhodné veličiny tvoří náhodný výběr z rozložení A(ϑ).
n = 100, m = 34/100, α = 0,05, $u_{1-\alpha/2} = u_{0,975} = 1,96$.
Ověření podmínky nϑ (1- ϑ) > 9: parametr ϑ neznáme, musíme ho nahradit výběrovým průměrem. Pak 100.0,34.0,66 = 22,44 > 9.

$$d = 0,34 - \sqrt{\frac{0,34(1-0,34)}{100}} 1,96 = 0,2472,\ h = 0,34 + \sqrt{\frac{0,34(1-0,34)}{100}} 1,96 = 0,4328.$$

S pravděpodobností přibližně 0,95 tedy 0,2472 < ϑ < 0,4328.

6.3.3. Testování hypotézy o parametru ϑ

Nechť $X_1, ..., X_n$ je náhodný výběr z rozložení A(ϑ) a nechť je splněna podmínka nϑ (1- ϑ) > 9. Na asymptotické hladině významnosti α testujeme hypotézu H_0: ϑ = c proti alternativě H_1: $\vartheta \ne c$ (resp. H_1: ϑ < c resp. H_1: ϑ > c). Testovým kritériem je statistika $T_0 = \frac{M-c}{\sqrt{\frac{c(1-c)}{n}}}$, která v případě platnosti nulové hypotézy má asymptoticky rozložení N(0,1).

Kritický obor má tvar $W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$ (resp. $W = \left(-\infty, -u_{1-\alpha}\right\rangle$ resp. $W = \left\langle u_{1-\alpha}, \infty\right)$).

Testování hypotézy o parametru ϑ lze samozřejmě provést i pomocí 100(1-α)% asymptotického intervalu spolehlivosti nebo pomocí p-hodnoty.

Příklad: Podíl zmetků při výrobě určité součástky činí $\vartheta = 0,01$. Bylo náhodně vybráno 1000 výrobků a zjistilo se, že mezi nimi je 16 zmetků. Na asymptotické hladině významnosti 0,05 testujte hypotézu H_0: $\vartheta = 0,01$ proti oboustranné alternativě H_1: $\vartheta \neq 0,01$.

Řešení: Ověření podmínky $n\vartheta(1-\vartheta) > 9$: $1000.0,01.0,99 = 9,9 > 9$. Realizace testového kritéria: $t_0 = \dfrac{0,016 - 0,01}{\sqrt{\dfrac{0,01 \cdot 0,99}{1000}}} = 1,907$. Kritický obor: $W = \left(-\infty, -u_{0,975}\right\rangle \cup \left\langle u_{0,975}, \infty\right) =$ $= \left(-\infty, -1,96\right\rangle \cup \left\langle 1,96, \infty\right)$. Protože $1,907 \notin W$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Příklady k 6. kapitole

Příklad 1.: Lze předpokládat, že hmotnost pomerančů dodávaných do obchodní sítě se řídí normálním rozložením se střední hodnotou 170 g a směrodatnou odchylkou 12 g. Jaká je pravděpodobnost, že celková hmotnost 9 náhodně vybraných pomerančů balených do síťky překročí 1,5 kg?
(Hledaná pravděpodobnost je 0,797.)

Příklad 2.: Počet bodů v testu inteligence je náhodná veličina, která se řídí rozložením N(100,225). Jaká je pravděpodobnost, že průměr v náhodně vybrané skupině 20 osob bude větší než 105 bodů?
(Hledaná pravděpodobnost je 0,06811.)

Příklad 3.: V prodejně mají zjištěno, že 25% nákupů přesahuje 300 Kč. Při nákupu nad 300 Kč obdrží zákazník slosovatelný kupón. Jaká je pravděpodobnost, že mezi 100 zákazníků bude rozdáno 20 až 25 kupónů? Výpočet proveďte jak přesně, tak pomocí aproximace normálním rozložením.
(Přesný výpočet: 0,4539, aproximativní výpočet: 0,4171.)

Příklad 4.: Při provádění určitého pokusu bylo zapotřebí udržovat v laboratoři konstantní teplotu 26,5°C. Teplota byla v jednom pracovním týdnu 46x namátkově kontrolována v různých denních a nočních hodinách. Z výsledků měření byly vypočteny realizace výběrového průměru a výběrové směrodatné odchylky: m = 26,33°C, s = 0,748°C. Za předpokladu, že výsledky měření teploty se řídí rozložením $N(\mu,\sigma^2)$, vypočtěte 95% empirický interval spolehlivosti pro střední hodnotu μ i pro směrodatnou odchylku σ.
(26,11°C < μ < 26,55°C s pravděpodobností aspoň 0,95, 0,62°C < σ < 0,94°C s pravděpodobností aspoň 0,95.)

Příklad 5.: U elektrického spotřebiče jisté značky byl sledován počet reklamací během záruční doby. Z 500 prodaných spotřebičů jich bylo reklamováno 25. Najděte 95% asymptotický interval spolehlivosti pro podíl reklamovaných spotřebičů.
(0,0309 < ϑ < 0,0691 s pravděpodobností přibližně 0,95.)

Příklad 6.: U 25 náhodně vybraných dvoulitrových lahví s nealkoholickým nápojem byl zjištěn přesný objem nápoje. Výběrový průměr činil m = 1,99 l a výběrová směrodatná odchylka s = 0,1 l. Předpokládejme, že objem nápoje v láhvi je náhodná veličina s normálním rozložením.
a) Na hladině významnosti 0,05 ověřte tvrzení výrobce, že zákazník není znevýhodněn.
b) Na hladině významnosti 0,05 ověřte tvrzení výrobce, že směrodatná odchylka je 0,08 l.
(ad a) Jelikož hodnota testového kritéria -0,5 neleží v kritickém oboru $\left(-\infty;-2{,}064\right\rangle$, nezamítáme nulovou hypotézu na hladině významnosti 0,05.
ad b) Jelikož hodnota testového kritéria 37,5 neleží v kritickém oboru $\left(0;12{,}4\right\rangle \cup \left\langle 39{,}4;\infty\right)$, nejsme oprávněni na hladině významnosti 0,05 zamítnout tvrzení výrobce.)

Příklad 7.: Bylo vybráno šest nových vozů téže značky a po určité době bylo zjištěno, o kolik mm se sjely jejich levé a pravé přední pneumatiky. Výsledky: (1,8; 1,5), (1,0; 1,1), (2,2; 2,0), (0,9; 1,1), (1,5; 1,4), (1,6; 1,4). Za předpokladu, že uvedené dvojice tvoří náhodný výběr z dvourozměrného normálního rozložení s vektorem středních hodnot (μ_1, μ_2), testujte na hladině významnosti 0,05 hypotézu, že obě pneumatiky se sjíždějí stejně rychle.

(Jelikož 1,0512 < 2,571, nelze na hladině významnosti 0,05 zamítnout nulovou hypotézu.)

Příklad 8.: Při zavádění kabelové televize na jednom velkém sídlišti se předpokládá zájem 40% domácností. Ze 70 náhodně vybraných domácností projevilo o kabelovou televizi zájem 25 domácností. Na 5% hladině významnosti ověřte hypotézu, že odchylka zjištěného zájmu od předpokládaného je způsobena jenom náhodnými vlivy.
(Testové kritérium -0,734 nepatří do kritického oboru $(-\infty, -1{,}96\rangle \cup \langle 1{,}96, \infty)$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.)

Práce se systémem STATISTICA

Téma: Parametrické úlohy o jednom náhodném výběru

Příklad 1.: **Vlastnosti výběrového průměru z normálního rozložení**
Předpokládejme, že velký ročník na vysoké škole má výsledky ze statistiky normálně rozloženy kolem střední hodnoty 72 bodů se směrodatnou odchylkou 9 bodů. Najděte pravděpodobnost, že průměr výsledků náhodného výběru 10 studentů bude větší než 80 bodů.
Návod: $X_1, ..., X_{10}$ je náhodný výběr z N(72, 81). Počítáme P(M > 80), přičemž výběrový průměr M má normální rozložení se střední hodnotou 72 a rozptylem 81/10. Tedy $P(M > 80) = 1 - P(M \le 80) = 1 - \Phi(80)$, kde $\Phi(80)$ je hodnota distribuční funkce rozložení N(72; 8,1) v bodě 80. Vytvoříme datový soubor o jedné proměnné a o jednom případu. Do Long Name této proměnné napíšeme =1 – INormal(80;72;sqrt(8,1)). Zjistíme, že $1 - \Phi(80) = 0{,}00248$. Funkce INormal($x;\mu;\sigma$) počítá hodnotu distribuční funkce rozložení $N(\mu,\sigma^2)$ v bodě x.

Příklad 2.: **Vlastnosti výběrového průměru z alternativního rozložení**
Mezi americkými voliči 60% osob volí republikány a 40% demokraty. Jaká je pravděpodobnost, že v náhodném výběru 100 amerických voličů budou voliči republikánů v menšině? Výpočet proveďte jak přesně, tak pomocí aproximace normálním rozložením.
Návod: $X_1, ..., X_{100}$ je náhodný výběr z A(0,6), $X_i = 1$, když i-tá osoba volí republikány, $X_i = 0$ jinak, i = 1, ..., 100. Zavedeme statistiku $Y_{100} = X_1 + ... + X_{100}$, $Y_{100} \sim Bi(100; 0,6)$, $E(Y_{100}) = 60$, $D(Y_{100}) = 24$. Označme $\Phi_{100}(y)$ distribuční funkci náhodné veličiny Y_{100},

$$\Phi_{100}(y) = \sum_{t=0}^{y} \binom{100}{t} 0{,}6^t 0{,}4^{100-t}.$$

Přesný výpočet: $P(Y_{100} < 50) = P(Y_{100} \le 49) = \Phi_{100}(49) = 0{,}0168$. Vytvoříme datový soubor o jedné proměnné a o jednom případu. Do Long Name této proměnné napíšeme =IBinom(49;0,6;100). Funkce IBinom(x;p;n) počítá hodnotu distribuční funkce rozložení Bi(n,p) v bodě x.
Přibližný výpočet: nejdříve ověříme splnění podmínky dobré aproximace. $n\vartheta(1-\vartheta) = 100.0{,}6.0{,}4 = 24 > 9$. Podmínka je splněna. $P(Y_{100} < 50) = P(Y_{100} \le 49) = \Phi(49)$, kde $\Phi(49)$ je hodnota distribuční funkce rozložení N(60; 24) v bodě 80. Vytvoříme datový soubor o jedné proměnné a o jednom případu. Do Long Name této proměnné napíšeme =INormal(49;60;sqrt(24)). Zjistíme, že $\Phi(49) = 0{,}0124$.

Příklad 3.: **Intervaly spolehlivosti pro parametry μ,σ^2 normálního rozložení**
Z populace stejně starých selat téhož plemene bylo vylosováno 6 selat a po dobu půl roku jim byla podávána táž výkrmná dieta. Byly zaznamenávány průměrné denní přírůstky hmotnosti v Dg. Z dřívějších pokusů je známo , že v populaci mívají takové přírůstky normální rozložení, avšak střední hodnota i rozptyl se měnívají. Přírůstky v Dg: 62, 54, 55, 60, 53, 58. Při riziku $\alpha = 0{,}05$ odvoďte:

a) levostranný interval spolehlivosti pro neznámou střední hodnotu μ při neznámé směrodatné odchylce σ.
b) interval spolehlivosti pro směrodatnou odchylku σ.

Návod: Vytvoříme datový soubor o 5 proměnných a 6 případech. První proměnnou nazveme hmotnost, druhou dm1, třetí dm2 a čtvrtou hm2. Do proměnné hmotnost zapíšeme zjištěné údaje. Pomocí Descriptive Statistics zjistíme výběrový průměr a výběrovou směrodatnou odchylku. Do LongName proměnné dm1 zapíšeme výraz = m – s* VStudent(0,95;5)/$\sqrt{6}$, kde za m a s dosadíme vypočtené hodnoty. Funkce VStudent(x;df) počítá x-kvantil rozložení t(df). Dostaneme výsledek 54,06, tedy $\mu > 54{,}06$ Dg s pravděpodobností aspoň 0,95.

Do Long Name proměnné dm2 zapíšeme výraz =s*sqrt(5)/sqrt(VChi2(0,975;5)), kde za s dosadíme vypočtenou směrodatnou odchylku. Podobně do LongName proměnné hm2 zapíšeme výraz =s*sqrt(5)/sqrt(VChi2(0,025;5)), kde za s dosadíme vypočtenou směrodatnou odchylku. Funkce VChi2(x;nu) počítá x-kvantil rozložení χ^2(nu).
Výsledek: 2,23 g < σ < 8,77 g s pravděpodobností aspoň 0,95.

Příklad 4.: **Interval spolehlivosti pro rozdíl parametrů μ_1-μ_2 dvourozměrného normálního rozložení**
Bylo vylosováno 6 vrhů selat a z nich vždy dva sourozenci. Jeden z nich vždy dostal náhodně dietu č. 1 a druhý dietu č. 2. Přírůstky v Dg jsou následující: (62,52), (54,56), (55,49), (60,50), (53,51), (58,50). Za předpokladu, že uvedené dvojice tvoří náhodný výběr z dvourozměrného normálního rozložení s vektorem středních hodnot (μ_1, μ_2), sestrojte 95% interval spolehlivosti pro rozdíl středních hodnot.
Návod: Vytvoříme datový soubor o třech proměnných a 6 případech. Do proměnných v1 a v21 zapíšeme naměřené přírůstky, do proměnné v3 uložíme rozdíly v1-v2.Pomocí Descriptive Statistics zjistíme meze 95% intervalu spolehlivosti pro střední hodnotu proměnné v3 tak, že zaškrtneme Conf. limits for mean. Dostaneme výsledek: 0,63 Dg < μ < 10,71 Dg s pravděpodobností aspoň 0,95.

Příklad 5.: **Asymptotický interval spolehlivosti pro parametr ϑ alternativního rozložení**
Může politická strana, pro niž se v předvolebním průzkumu vyslovilo 60 z 1000 dotázaných osob, očekávat se spolehlivostí 0,95, že by v této době ve volbách překročila 5% hranici pro vstup do parlamentu?
Návod: Zavedeme náhodné veličiny $X_1, ..., X_{1000}$, přičemž $X_i = 1$, když i-tá osoba se vysloví pro danou politickou stranu a $X_i = 0$ jinak, i = 1, ..., 1000. Tyto náhodné veličiny tvoří náhodný výběr z rozložení A(ϑ). V tomto případě n = 1000, m = 60/1000 = 0,06, α = 0,05, $u_{1-\alpha} = u_{0,95}$ = 1,645.
Ověření podmínky $n\vartheta(1-\vartheta) > 9$: parametr ϑ neznáme, musíme ho nahradit výběrovým průměrem. Pak 1000.0,06.0,94 = 56,4 > 9. Zajímá nás 95% levostranný interval spolehlivosti pro ϑ. Použijeme vzorec

$$d = 0{,}06 - \sqrt{\frac{0{,}06(1-0{,}06)}{1000}} \cdot 1{,}645 = 0{,}048.$$

S pravděpodobností přibližně 0,95 tedy ϑ > 0,048. Protože tento interval zahrnuje i hodnoty nižší než 0,05, nelze vyloučit, že strana získá méně než 5% hlasů.
Vytvoříme datový soubor o jedné proměnné a o jednom případu. Do Long Name této proměnné napíšeme =0,06-sqrt(0,06*0,94/1000)*VNormal(0,95;0;1).

Příklad 6.: **Testování hypotézy o parametru μ normálního rozložení**
Systematická chyba měřicího přístroje se eliminuje nastavením přístroje a měřením etalonu, jehož správná hodnota je μ = 10,00. Nezávislými měřeními za stejných podmínek byly získány hodnoty: 10,24 10,12 9,91 10,19 9,78 10,14 9,86 10,17 10,05, které považujeme za realizace náhodného výběru rozsahu 9 z rozložení N(μ, σ^2). Je možné při riziku 0,05 vysvětlit odchylky od hodnoty 10,00 působením náhodných vlivů?
Návod: Jde o úlohu na jednovýběrový t-test. Vytvoříme datový soubor o jedné proměnné a 9 případech, kam zapíšeme naměřené hodnoty. V Descriptive Statistics vybereme t-test, single sample. Do Reference values zapíšeme 10. Ve výstupu se podíváme na hodnotu testového kritéria a na p-hodnotu. Pokud p-hodnota bude menší nebo rovna 0,05, zamítneme hypotézu H_0: $\mu = 0$ ve prospěch alternativní hypotézy H_1: $\mu \neq 0$ na hladině významnosti 0,05. V opačném případě H_0 nezamítáme. V našem případě nulovou hypotézu nezamítáme na hladině významnosti 0,05.

Příklad 7.: **Testování hypotézy o rozdíl parametrů μ_1-μ_2 dvourozměrného normálního rozložení**

Pro data z příkladu 4. testujte na hladině významnosti 0,05 hypotézu, že obě výkrmné diety mají stejný vliv.

Návod: Jde o úlohu na párový t-test. Vytvoříme datový soubor o dvou proměnných a šesti případech. V Descriptive Statistics vybereme t-test, dependent samples. Zadáme názvy obou proměnných a ve výstupu se podíváme na hodnotu testového kritéria a na p-hodnotu. Pokud p-hodnota bude menší nebo rovna 0,05, zamítneme hypotézu H_0: $\mu = 0$ ve prospěch alternativní hypotézy H_1: $\mu \neq 0$ na hladině významnosti 0,05. V opačném případě H_0 nezamítáme. V našem případě nulovou hypotézu zamítáme na hladině významnosti 0,05.

Příklad 8: **Testování hypotézy o parametru ϑ alternativního rozložení**

Určitá cestovní kancelář organizuje zahraniční zájezdy podle individuálních přání zákazníků. Z několika minulých let ví, že 30% všech takto organizovaných zájezdů má za cíl zemi X. Po zhoršení politických podmínek v této zemi se cestovní kancelář obává, že se zájem o tuto zemi mezi zákazníky sníží. Ze 150 náhodně vybraných zákazníků v tomto roce má 38 za cíl právě zemi X. Potvrzují nejnovější data pokles zájmu o tuto zemi? Volte hladinu významnosti 0,05.

Návod: Máme náhodný výběr $X_1, ..., X_{150}$ z rozložení A(0,3). Testujeme H_0: $\vartheta = 0,3$ proti levostranné alternativě H_1: $\vartheta < 0,3$. V tomto případě je testovým kritériem statistika $T_0 = \frac{M - c}{\sqrt{\frac{c(1-c)}{n}}}$, která v případě platnosti nulové hypotézy má asymptoticky rozložení N(0,1).

Musíme ověřit splnění podmínky $n\vartheta(1-\vartheta) > 9$: 150.0,3.0,7 = 31,5 > 9 Vypočteme realizaci testového kritéria: $\frac{m - c}{\sqrt{\frac{c(1-c)}{n}}} = \frac{\frac{38}{150} - 0,3}{\sqrt{\frac{0,3(1-0,3)}{150}}} = -1,249$. Kritický obor: $W = \left(-\infty, -u_{1-\alpha}\right\rangle = \left(-\infty, -1,645\right\rangle$. Protože testové kritérium nepatří do kritického oboru, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Vytvoříme datový soubor o dvou proměnných a jednom případu. Vypočteme realizaci testového kritéria tak, že do LongName první proměnné zapíšeme odpovídající vzorec. Do LongName druhé proměnné napíšeme =VNormal(0,95;0;1), čímž získáme kvantil $u_{0,95}$ a testové kritérium porovnáme s opačnou hodnotou tohoto kvantilu.