Téma 9.: Parametrické úlohy o jednom náhodném výběru z normálního rozložení a dvourozměrného rozložení

Upozornění: Pokud to povaha úlohy vyžaduje, proveďte test normality dat:

Příklad 1.: Vlastnosti výběrového průměru z normálního rozložení
Předpokládejme, že velký ročník na vysoké škole má výsledky zkoušky 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 $E(M) = \mu = 72$ a rozptylem $D(M) = \frac{\sigma^2}{n} = \frac{81}{10} = 8{,}1$.
Tedy $P(M > 80) = 1 - P(M \leq 80) = 1 - \Phi(80)$, kde $\Phi(80)$ je hodnota distribuční funkce rozložení N(72; 8,1) v bodě 80.
Výpočet pomocí systému STATISTICA: Vytvoříme datový soubor o jedné proměnné a o jednom případu. Do Dlouhého jména této proměnné napíšeme =1 – INormal(80;72;sqrt(8,1)). Zjistíme, že $1 - \Phi(80) = 0{,}00247005$. Funkce INormal(x;μ;σ) počítá hodnotu distribuční funkce rozložení $N(\mu,\sigma^2)$ v bodě x.

Příklad k samostatnému řešení:
Je známo, že týdenní výdaje domácností na určité potravinářské zboží se řídí normálním rozložením se střední hodnotou 90 Kč a směrodatnou odchylkou 14 Kč. Jaká je pravděpodobnost překročení hranice 100 Kč pro průměrné výdaje pěti náhodně vybraných domácností?
Výsledek: 0,0548

Příklad 2.: Intervaly spolehlivosti pro parametry μ, σ^2 normálního rozložení
Z populace stejně starých selat téhož plemene bylo vylosováno šest 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 mají takové přírůstky normální rozložení, avšak střední hodnota i rozptyl se mění. Přírůstky v Dg: 62, 54, 55, 60, 53, 58 jsou uloženy v souboru jedna_dieta.sta

a) Najděte 95% empirický levostranný interval spolehlivosti pro neznámou střední hodnotu μ při neznámé směrodatné odchylce σ.
b) Najděte 95% empirický interval spolehlivosti pro směrodatnou odchylku σ.

Návod:
Ověříme normalitu pomocí S-W testu a zjistíme, že p-hodnota je 0,7374, tedy na 5% hladině významnosti hypotézu o normalitě nezamítáme.

Statistiky – Základní statistiky/tabulky – Popisné statistiky – OK – Proměnná hmotnost – OK – na záložce Detailní výsledky zaškrtneme Meze spolehl. prům., 95 % změníme na 90 %, dále zaškrtneme Meze sp. směr. odch. a všechny ostatní volby odškrtneme – Výpočet.

Proměnná	Popisné statistiky (Tabulka4) Int. spolehl. -90,000%	Int. spolehl. 90,000	Spolehlivost Sm.Odch. -95,000%	Spolehlivost Sm.Odch. +95,000%
hmotnost	54,05683	59,94317	2,233234	8,774739

ad a) Protože mez 95% levostranného intervalu spolehlivosti pro střední hodnotu je stejná jako dolní mez 90% oboustranného intervalu spolehlivosti pro střední hodnotu, vidíme, že $\mu > 54{,}06$ Dg s pravděpodobností 0,95.

ad b) 2,23 g < σ < 8,77 g s pravděpodobností 0,95.

Příklad k samostatnému řešení:
Při měření určitého objektu byly získány tyto hodnoty (v mm): 6,42 6,44 6,38 6,60 6,50 6,51. Považujeme je za realizace náhodného výběru z rozložení $N(\mu, \sigma^2)$, kde parametry μ a σ^2 neznáme. Sestrojte 95% empirický interval spolehlivosti pro směrodatnou odchylku σ. Ověřte normalitu dat.
Výsledek: Ověříme normalitu pomocí S-W testu a zjistíme, že p-hodnota je 0,8367, tedy na 5% hladině významnosti hypotézu o normalitě nezamítáme.
$0{,}04885 < \sigma < 0{,}19700$ s pravděpodobností aspoň 0,95

Příklad 3.: Testování hypotézy o střední hodnotě μ
Systematická chyba měřicího přístroje se eliminuje nastavením přístroje a měřením etalonu, jehož správná hodnota je $\mu = 10{,}00$. 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 $N(\mu, \sigma^2)$. Je možné při riziku 0,05 vysvětlit odchylky od hodnoty 10,00 působením náhodných vlivů?

Návod:
Na hladině významnosti 0,05 testujeme hypotézu H_0: $\mu = 10$ proti oboustranné alternativě H_1: $\mu \neq 10$. Jde o úlohu na jednovýběrový t-test. Ten je ve STATISTICE implementován. Načteme datový soubor mereni_etalonu.sta.
Ověříme normalitu pomocí S-W testu a zjistíme, že p-hodnota je 0,2873, tedy na 5% hladině významnosti hypotézu o normalitě nezamítáme.
1. způsob: V Základních statistikách a tabulkách vybereme t-test, samostatný vzorek. Do Referenční hodnoty zapíšeme 10. Dostaneme výstupní tabulku:

Proměnná	Test průměrů vůči referenční konstantě (hodnotě) Průměr	Sm.odch.	N	Sm.chyba	Referenční konstanta	t	SV	p
Prom1	10,05111	0,162669	9	0,054223	10,00000	0,942611	8	0,373470

Protože p-hodnota 0,373470 > 0,05 nulovou hypotézu nezamítáme na hladině významnosti 0,05. Odchylky od hodnoty 10 lze vysvětlit působením náhodných vlivů.

2. způsob: V Základních statistikách a tabulkách vypočteme průměr a směrodatnou odchylku. Pak použijeme Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma průměry (normální rozdělení) – zaškrtneme Výběrový průměr vs. Střední hodnota – do políčka Pr1 napíšeme 10,05111, do políčka SmOd1 napíšeme 0,162669, do políčka N1 napíšeme 9, do políčka Pr2 napíšeme 10 - Výpočet. Dostaneme p-hodnotu 0,3735, tedy nezamítáme nulovou hypotézu na hladině významnosti 0,05.

Příklad k samostatnému řešení:
Nechť $X_1, ..., X_{400}$ je náhodný výběr z $N(\mu,0,01)$. Je známo, že výběrový průměr se realizoval hodnotou 0,01. Na hladině významnosti 0,05 testujte hypotézu H_0: $\mu = 0$ proti pravostranné alternativě H_1: $\mu > 0$ pomocí p-hodnoty.
Výsledek: p-hodnota = 0,02275, H_0 tedy zamítáme na hladině významnosti 0,05.

Příklad 4.: Testování hypotézy o směrodatné odchylce σ
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.
Na hladině významnosti 0,05 ověřte tvrzení výrobce, že směrodatná odchylka je 0,08 l.

Návod:
Na hladině významnosti 0,05 testujeme hypotézu H_0: $\sigma = 0,08$ proti oboustranné alternativě H_1: $\sigma \neq 0,08$ neboli H_0: $\sigma^2 = 0,0064$ proti oboustranné alternativě H_1: $\sigma^2 \neq 0,0064$. Jde o úlohu na test o rozptylu. Vypočteme realizaci testového kritéria $t_0 = \frac{(n-1)s^2}{c} = \frac{24 \cdot 0,1^2}{0,08^2} = 37,5$.
Jelikož hodnota testového kritéria 37,5 neleží v kritickém oboru $W = \left(0; \chi^2_{0,025}(24)\right\rangle \cup \left\langle \chi^2_{0,975}(24); \infty\right) = \left(0; 12,4\right\rangle \cup \left\langle 39,4; \infty\right)$, nejsme oprávněni na hladině významnosti 0,05 zamítnout tvrzení výrobce.)
V systému STATISTICA otevřeme datový soubor o třech proměnných a jednom případu. Do Dlouhého jména první proměnné napíšeme vzorec pro výpočet testového kritéria:
=24*0,1^2/0,08^2
Další dvě proměnné nám poslouží k výpočtu kvantilů Pearsonova χ^2 – rozložení.
Do Dlouhého jména druhé proměnné napíšeme
=VChi2(0,025;24)
a do Dlouhého jména třetí proměnné napíšeme
=VChi2(0,975;24)

Příklad k samostatnému řešení:
Rozptyl obsahu určité látky v tabletách, které vyrábí farmaceutická firma, nesmí překročit 0,09 mg^2. Když je tato hodnota překročena, musí se provést korekce a nastavení výrobní linky. Kontrolor náhodně vybral 25 tablet a zjistil obsah účinné látky (v mg). Údaje jsou uloženy v souboru tablety.sta. Na hladině významnosti 0,05 testujte nulovou hypotézu, že směrodatná odchylka obsahu sledované látky vyhovuje podmínce proti pravostranné alternativě. Test proveďte pomocí intervalu spolehlivosti. Nezapomeňte ověřit normalitu dat.
Výsledky:
p-hodnota S-W testu normality = 0,8588, hypotézu o normalitě nezamítáme na hladině významnosti 0,05.
Hypotézu H_0: $\sigma \leq 0,3$ nezamítáme na hladině významnosti 0,05 ve prospěch pravostranné alternativy H_1: $\sigma > 0,3$, protože $0,3 \in (0,268; \infty)$.

Příklad 5.: Interval spolehlivosti pro rozdíl parametrů $\mu_1 - \mu_2$ dvourozměrného 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). Data jsou v souboru dve_diety.sta. Za předpokladu, že uvedené dvojice tvoří náhodný výběr z dvourozměrného rozložení s vektorem středních hodnot (μ_1, μ_2) a jejich rozdíly se řídí normálním rozložením (ověřte!), sestrojte 95% interval spolehlivosti pro rozdíl středních hodnot.

Návod:
K datovému souboru přidáme proměnnou Z, do níž uložíme rozdíly X - Y.
p-hodnota S-W testu normality = 0,3241, hypotézu o normalitě proměnné Z nezamítáme na hladině významnosti 0,05.

Ve STATISTICE je implementován výpočet oboustranného intervalu spolehlivosti pro μ, když σ^2 neznáme. Pomocí Popisných statistik zjistíme meze 95% intervalu spolehlivosti pro střední hodnotu proměnné Z tak, že zaškrtneme Meze spolehl. prům.

Proměnná	Popisné statistiky (dve_diety.sta) Int. spolehl. -95,000%	Int. spolehl. 95,000%
Z	0,626461	10,70687

Dostaneme výsledek: 0,63 Dg < μ < 10,71 Dg s pravděpodobností 0,95.

Příklad 6.: Testování hypotézy o rozdílu parametrů $\mu_1 - \mu_2$ dvourozměrného rozložení
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 rozložení s vektorem středních hodnot (μ_1, μ_2) a jejich rozdíly se řídí normálním rozložením (ověřte!), testujte na hladině významnosti 0,05 hypotézu, že obě pneumatiky se sjíždějí stejně rychle. Data jsou uložena v souboru pneumatiky.sta.

Návod:
K datovému souboru přidáme proměnnou Z, do níž uložíme rozdíly X - Y.
p-hodnota S-W testu normality = 0,4522, hypotézu o normalitě nezamítáme na hladině významnosti 0,05.

Označme $\mu = \mu_1 - \mu_2$. Na hladině významnosti 0,05 testujeme hypotézu H_0: $\mu = 0$ proti oboustranné alternativě H_1: $\mu \neq 0$. Jde o úlohu na párový t-test.Ten je ve STATISTICE implementován. V Základních statistikách vybereme t-test, závislé vzorky. Zadáme názvy obou proměnných a ve výstupu se podíváme na p-hodnotu.

Proměnná	t-test pro závislé vzorky (Tabulka1) Označ. rozdíly jsou významné na hlad. p < ,05000 Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
X	1,500000	0,489898						
Y	1,416667	0,331160	6	0,083333	0,194079	1,051758	5	0,341062

Protože p-hodnota 0,341062 > 0,05, nezamítáme na hladině významnosti 0,05 hypotézu, že obě přední pneumatiky se sjíždějí stejně rychle.

Příklad k samostatnému řešení:
Dvanácti pacientům byl změřen systolický krevní tlak vždy před podáním léku a dvě hodiny po podání léku. Výsledky jsou uloženy v souboru krevni_tlak.sta. Veličina X udává tlak před podáním léku a veličina Y tlak po podání léku.

X	124	126	138	117	143	128	146	133	127	135	126	131
Y	120	124	130	118	140	128	140	135	126	130	126	127

a) Na hladině významnosti 0,05 testujte třemi způsoby hypotézu, že rozdíl tlaků před a po podání léku se řídí normálním rozložením.
Lilieforsova varianta Kolmogorovova - Smirnovova testu:
Hodnota testové statistiky = 0,1287
p-hodnota > 0,2
Rozhodnutí o nulové hypotéze: nezamítáme na hladině významnosti 0,05
Shapirův – Wilkův test:
Hodnota testové statistiky = 0,974
p-hodnota = 0,9475
Rozhodnutí o nulové hypotéze: nezamítáme na hladině významnosti 0,05
Andersonův – Darlingův test:
Hodnota testové statistiky = 0,239
p-hodnota = 0,9759
Rozhodnutí o nulové hypotéze: nezamítáme na hladině významnosti 0,05

b) Najděte meze 95% intervalu spolehlivosti
pro střední hodnotu systolického krevního tlaku před podáním léku:
dolní mez = 125,9
horní mez = 136,4
pro směrodatnou odchylku systolického krevního tlaku po podáním léku:
dolní mez = 4,9
horní mez = 11,8

c) Na hladině významnosti 0,05 testujte pomocí párového t-testu hypotézu, že podaný lék nemá vliv na systolický krevní tlak.
Zápis nulové hypotézy: H_0: $\mu_1 - \mu_2 = 0$

Zápis alternativní hypotézy: H_1: $\mu_1 - \mu_2 \neq 0$

p-hodnota = 0,0156 < 0,05, tedy H_0 zamítáme na hladině významnosti 0,05