

Téma 8: Parametrické úlohy o jednom náhodném výběru z normálního rozložení a dvourozměrného normálního rozložení

Úkol 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, \dots, 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.

Vytvoříme datový soubor o jedné proměnné a o jednom případě. Do Dlouhého jména této proměnné napíšeme $=1 - \text{INormal}(80;72;\text{sqrt}(8,1))$. Zjistíme, že $1 - \Phi(80) = 0,00247005$. Funkce $\text{INormal}(x;\mu;\sigma)$ počítá hodnotu distribuční funkce rozložení $N(\mu,\sigma^2)$ v bodě x .

	1
	Prom1
1	0,00247

Úkol k samostatnému řešení: 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 devíti náhodně vybraných pomerančů balených do sít'ky překročí 1,5 kg?

Výsledek: Hledaná pravděpodobnost je 0,797.

Úkol 3.: 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 mívají takové přírůstky normální rozložení, avšak střední hodnota i rozptyl se měnívají. Přírůstky v Dg: 62, 54, 55, 60, 53, 58.

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

Návod:

Vytvoříme datový soubor o 4 proměnných a 6 případech. První proměnnou nazveme hmotnost, druhou dm1, třetí dm2 a čtvrtou hm2. Do proměnné hmotnost napíšeme zjištěné údaje. Pomocí Popisných statistik zjistíme realizace výběrového průměru a výběrové směrodatné odchylky.

Proměnná	Popisné statistiky	
	Průměr	Sm. odch.
hmotnost	57,00000	3,577709

ad a) Dolní mez $100(1-\alpha)\%$ empirického jednostranného intervalu spolehlivosti pro μ při neznámém σ^2 je $m - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1)$, tedy v našem případě $57 - \frac{3,577709}{\sqrt{6}} t_{0,95}(5) = 54,06$

Do Dlouhého jména proměnné dm1 zapíšeme výraz

$$= 57 - 3,577709 * VStudent(0,95;5)/sqrt(6)$$

Funkce VStudent(x;df) počítá x-kvantil rozložení t(df).

Dostaneme výsledek 54,05682, tedy $\mu > 54,06$ Dg s pravděpodobností aspoň 0,95.

ad b) Meze $100(1-\alpha)\%$ empirického oboustranný intervalu spolehlivosti pro σ při neznámém

$$\mu \text{ jsou } \left(\frac{\sqrt{n-1} s}{\sqrt{\chi_{1-\alpha/2}^2(n-1)}} ; \frac{\sqrt{n-1} s}{\sqrt{\chi_{\alpha/2}^2(n-1)}} \right).$$

Do Dlouhého jména proměnné dm2 zapíšeme výraz

$$= 3,577709 * sqrt(5)/sqrt(VChi2(0,975;5)). \text{ Vyjde } 2,233235.$$

Podobně do Dlouhého jména proměnné hm2 zapíšeme výraz

$$= 3,577709 * sqrt(5)/sqrt(VChi2(0,025;5)) \text{ Vyjde } 8,774739.$$

Funkce VChi2(x;nu) počítá x-kvantil rozložení $\chi^2(nu)$.

Dostaneme výsledek: $2,23 \text{ g} < \sigma < 8,77 \text{ g}$ s pravděpodobností aspoň 0,95.

	1	2	3
	dm1	dm2	hm2
1	54,05683	2,233235	8,774739

Upozornění: STATISTICA verze 8 umí počítat meze $100(1-\alpha)\%$ empirického intervalu spolehlivosti pro neznámou směrodatnou odchylku při neznámé střední hodnotě: v Popisných statistikách zaškrtneme Meze sp. směr. odch. Dostaneme tabulku:

	Popisné statistiky (Tabulka1)	
	Spolehlivost Sm.Odch.	Spolehlivost Sm.Odch.
Proměnná	-95.000%	+95.000%
Prom1	2,233234	8,774739

Úkol k samostatnému řešení: Při provádění určitého pokusu bylo zapotřebí udržovat v laboratoři konstantní teplotu $26,5^\circ\text{C}$. 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: $m = 26,33^\circ\text{C}$, $s = 0,748^\circ\text{C}$. Za předpokladu, že výsledky měření teploty se řídí rozložením $N(\mu, \sigma^2)$, vypočtěte 95% empirický interval spolehlivosti

a) pro střední hodnotu μ

b) pro směrodatnou odchylku σ .

Výsledek:

ad a) $26,11^\circ\text{C} < \mu < 26,55^\circ\text{C}$ s pravděpodobností aspoň 0,95.

ad b) $0,62^\circ\text{C} < \sigma < 0,94^\circ\text{C}$ s pravděpodobností aspoň 0,95.

Úkol 4.: 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. Vytvoříme datový soubor o jedné proměnné a devíti případech, kam zapíšeme naměřené hodnoty.

1. způsob: V Základních statistikách a tabulkách vybereme t-test, samostatný vzorek. Do Referenční hodnoty 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 $H_0: \mu = 10$ ve prospěch oboustranné alternativní hypotézy $H_1: \mu \neq 10$ na hladině významnosti 0,05. V opačném případě H_0 nezamítáme. V našem případě je

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. S rizikem omylu nejvýše 5% lze tedy odchylky od hodnoty 10 vysvětlit působením náhodných vlivů.

Všimněme si ještě hodnoty testového kritéria: $t_0 = 0,942611$. Kritický obor

$$W = (-\infty, -t_{1-\alpha/2}(n-1)) \cup (t_{1-\alpha/2}(n-1), \infty) = (-\infty, -t_{0,975}(8)) \cup (t_{0,975}(8), \infty) = (-\infty, -2,306) \cup (2,306, \infty)$$

Protože $t_0 \notin W$, nezamítáme na hladině významnosti 0,05 hypotézu H_0 .

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.

Úkol k samostatnému řešení: Při kontrole balicího automatu, který má plnit cukrem balíčky o hmotnosti 1000 g, byly při přesném převážení 5 balíčků zjištěny tyto odchylky (v gramech) od požadované hodnoty: 3, -2, 2, 0, 1. Na hladině významnosti 0,05 testujte hypotézu, že automat nemá systematickou odchylku od požadované hodnoty.

Výsledek: Protože p-hodnota je 0,405023, nulovou hypotézu nezamítáme na hladině významnosti 0,05.

Úkol 5.: 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

$W = (0; \chi^2_{0,025}(24)) \cup (\chi^2_{0,975}(24); \infty) = (0; 12,4) \cup (39,4; \infty)$, nejsme oprávněni na hladině významnosti 0,05 zamítnout tvrzení výrobce.)

V systému STATISTICA otevřeme datový soubor o třech proměnných a jednom případě. 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

$=VChi2(0,975;24)$

Úkol 6.: 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). 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, sestrojte 95% interval spolehlivosti pro rozdíl středních hodnot.

Návod:

Vytvoříme datový soubor o třech proměnných a šesti případech. Do proměnných v1 a v2 zapíšeme naměřené přírůstky, do proměnné v3 uložíme rozdíly v1 - v2.

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 v3 tak, že zaškrtneme Meze spolehl. prům.

Proměnná	Popisné statistiky	
	Int. spolehl.	Int. spolehl.
	-95,000%	+95,000%
Prom3	0,626461	10,70687

Dostaneme výsledek: $0,63 \text{ Dg} < \mu < 10,71 \text{ Dg}$ s pravděpodobností aspoň 0,95.

Úkol 7.: 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, testujte na hladině významnosti 0,05 hypotézu, že obě pneumatiky se sjíždějí stejně rychle.

Návod:

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. Vytvoříme datový soubor o dvou proměnných a šesti případech. Do proměnných v1 a v2 zapíšeme naměřené přírůstky. V Základních statistikách vybereme t-test,

závislé vzorky. Zadáme názvy obou proměnných a ve výstupu se podíváme na hodnotu testového kritéria a 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.

Všimněme si ještě hodnoty testového kritéria: $t_0 = 1,051758$. Kritický obor

$$W = (-\infty, -t_{1-\alpha/2}(n-1)) \cup (t_{1-\alpha/2}(n-1), \infty) = (-\infty, -t_{0,975}(5)) \cup (t_{0,975}(5), \infty) = (-\infty, -2,5706) \cup (2,5706, \infty)$$

Protože $t_0 \notin W$, nezamítáme na hladině významnosti $0,05$ hypotézu H_0 .

Úkol k samostatnému řešení: Zkouška ze statistiky se skládá z písemné části, v níž je možno získat maximálně 20 bodů a z ústní části, kde je možno získat maximálně 10 bodů. Výsledky 20 náhodně vybraných studentů (X – počet bodů z písemné části, Y – počet bodů z ústní části):

č. st.	1	2	3	4	5	6	7	8	9	10
X	6	11	8	18	6	11	6	3	14	7
Y	4	7	6	8	3	5	6	4	9	8

č. st.	11	12	13	14	15	16	17	18	19	20
X	17	12	8	4	15	20	13	5	10	0
Y	10	9	6	5	7	10	8	6	7	3

Na hladině významnosti $0,05$ testujte hypotézu, že rozdíl středních hodnot počtu bodů v písemné a ústní části se liší o 3 body proti oboustranné alternativě.

Výsledek: Hodnota testové statistiky = $0,178431$, p-hodnota = $0,806273$, na hladině významnosti $0,05$ tedy nezamítáme nulovou hypotézu.