

## Téma 8: Parametrické úlohy o dvou nezávislých náhodných výběrech z normálních rozložení a jednom náhodném výběru z alternativního rozložení

### Úkol 1.: Intervaly spolehlivosti pro parametrické funkce $\mu_1 - \mu_2$ , $\sigma_1^2 / \sigma_2^2$

Bylo vylosováno 11 stejně starých selat téhož plemene. Šesti z nich byla předepsána výkrmná dieta č. 1 a zbylým pěti výkrmná dieta č. 2. Průměrné denní přírůstky v Dg za dobu půl roku jsou následující:

dieta č. 1: 62, 54, 55, 60, 53, 58

dieta č. 2: 52, 56, 49, 50, 51.

Zjištěné hodnoty považujeme za realizace dvou nezávislých náhodných výběrů pocházejících z rozložení  $N(\mu_1, \sigma_1^2)$  a  $N(\mu_2, \sigma_2^2)$ .

a) Sestrojte 95% empirický interval spolehlivosti pro podíl rozptylů.

b) Za předpokladu, že data pocházejí z rozložení  $N(\mu_1, \sigma^2)$  a  $N(\mu_2, \sigma^2)$ , sestrojte 95% empirický interval spolehlivosti pro rozdíl středních hodnot  $\mu_1 - \mu_2$ .

### Návod:

Načteme datový soubor `dve_diety.sta` o 2 proměnných hmotnost a dieta a 11 případech. Pomocí Popisných statistik zjistíme realizace výběrových průměrů, výběrových rozptylů a výběrových směrodatných odchylek.

Pro první dietu:

Popisné statistiky (Tabulka1)				
Zhrnout podmínku: v2=1				
Proměnná	N platných	Průměr	Rozptyl	Sm.odch.
hmotnost	6	57,00000	12,80000	3,577709

Pro druhou dietu:

Popisné statistiky (Tabulka1)				
Zhrnout podmínku: v2=2				
Proměnná	N platných	Průměr	Rozptyl	Sm.odch.
hmotnost	5	51,60000	7,300000	2,701851

ad a)

Meze 100(1- $\alpha$ )% empirického intervalu spolehlivosti pro podíl rozptylů jsou:

$$(d, h) = \left( \frac{s_1^2 / s_2^2}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{s_1^2 / s_2^2}{F_{\alpha/2}(n_1 - 1, n_2 - 1)} \right).$$

Otevřeme nový datový soubor o dvou proměnných d a h a jednom případě.

Do Dlouhého jména proměnné d napíšeme

`=(12,8/7,3)/VF(0,975;5;4)`

(Funkce VF(x;ný;omega) počítá x-quantil Fisherova – Snedecorova rozložení F(ný, omega).)

Do Dlouhého jména proměnné h napíšeme

`=(12,8/7,3)/VF(0,025;5;4)`

	1	2
	d	h
1	0,187242	12,9541

S pravděpodobností aspoň 0,95 tedy platí:  $0,1872 < \sigma_1^2 / \sigma_2^2 < 12,954$ .

ad b) Meze 100(1- $\alpha$ )% empirického intervalu spolehlivosti pro rozdíl středních hodnot (v případě, že rozptyly neznáme, ale víme, že jsou shodné) můžeme ve STATISTICE vypočítat pomocí dvouvýběrového t-testu:

Statistika – Základní statistiky a tabulky – t-test, nezávislé, dle skupin – OK, Proměnné – Závislé proměnné hmotnost, Grupovací proměnná dieta – OK. Na záložce Možnosti zaškrtneme Meze spol. pro odhady – Výpočet. Zajímají nás poslední dva sloupce ve výstupní tabulce.

	Int. spolehl.	Int. spolehl.
Proměnná	-95,000%	+95,000%
hmotnost	0,991963	9,808037

S pravděpodobností aspoň 0,95 tedy  $0,99 \text{ Dg} < \mu_1 - \mu_2 < 9,81 \text{ Dg}$ .

**Úkol k samostatnému řešení:** Jsou dány dva nezávislé náhodné výběry o rozsazích  $n_1 = 25$ ,  $n_2 = 10$ , první pochází z rozložení  $N(\mu_1, \sigma_1^2)$ , druhý z rozložení  $N(\mu_2, \sigma_2^2)$ , kde parametry  $\mu_1$ ,  $\mu_2$ ,  $\sigma_1^2$ ,  $\sigma_2^2$  neznáme. Byly vypočteny realizace výběrových rozptylů:  $s_1^2 = 1,7482$ ,  $s_2^2 = 1,7121$ . Sestrojte 95% empirický interval spolehlivosti pro podíl rozptylů.

**Výsledek:**

$0,28 < \sigma_1^2 / \sigma_2^2 < 2,76$  s pravděpodobností aspoň 0,95.

**Úkol 2.: Testování hypotéz o parametrických funkcích  $\mu_1 - \mu_2$ ,  $\sigma_1^2 / \sigma_2^2$**

Pro datový soubor z úkolu 2 testujte na hladině významnosti 0,05 hypotézu, že

a) rozptyly hmotnostních přírůstků selat při obou výkrmných dietách jsou shodné

b) obě výkrmné diety mají stejný vliv na hmotnostní přírůstky selat.

**Návod:**

Provedeme dvouvýběrový t-test současně s testem o shodě rozptylů:

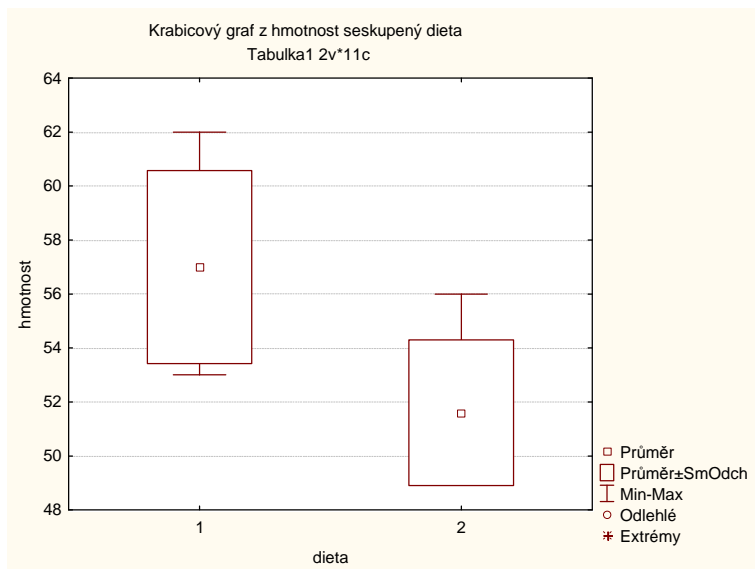
Statistika – Základní statistiky a tabulky – t-test, nezávislé, dle skupin – OK, Proměnné – Závislé proměnné hmotnost, Grupovací proměnná dieta – OK.

	t-testy; grupováno: dieta (Tabulka1)										
	Skup. 1: 1										
	Skup. 2: 2										
Proměnná	Průměr	Průměr	t	sv	p	Poč.plat	Poč.plat.	Sm.odch.	Sm.odch.	F-poměr	p
	1	2				1	2	1	2	Rozptyly	Rozptyly
hmotnost	57,00000	51,60000	2,771222	9	0,021710	6	5	3,577709	2,701851	1,753425	0,606345

Testová statistika pro test shody rozptylů se realizuje hodnotou 1,7534, odpovídající p-hodnota je 0,6063, tedy na hladině významnosti 0,05 nezamítáme hypotézu o shodě rozptylů. (Upozornění: v případě zamítnutí hypotézy o shodě rozptylů je zapotřebí v tabulce t-testu pro nezávislé vzorky dle skupin zaškrtnout volbu Test se samostatnými odhady rozptylu.)

Dále z tabulky plyne, že testová statistika pro test shody středních hodnot se realizuje hodnotou 2,7712, počet stupňů volnosti je 9, odpovídající p-hodnota 0,0217, tedy hypotézu o shodě středních hodnot zamítáme na hladině významnosti 0,05. Znamená to, že s rizikem omylu nejvýše 5% se prokázalo, že obě výkrmné diety se liší účinností.

Tabulku ještě doplníme krabicovými diagramy. Na záložce Details zaškrtneme krabicový graf a vybereme volbu Průměr/SmOdch/Min-Max.



**Upozornění:** Dvouvýběrový t-test lze v systému STATISTICA provést ještě jiným způsobem, který je vhodný zvláště tehdy, známe-li realizace výběrových průměrů a výběrových směrodatných odchylek.

Statistiky – Základní statistiky a tabulky – Testy rozdílů: r, %, průměry – OK – vybereme Rozdíl mezi dvěma průměry (normální rozdělení) – do políčka Pr1 napíšeme 57, do políčka SmOd1 napíšeme 3,5777, do políčka N1 napíšeme 6, do políčka Pr2 napíšeme 51,6, do políčka SmOd2 napíšeme 2,7019, do políčka N2 napíšeme 5 - Výpočet.

Testy rozdílů: r, %, průměry: Tabulka1

Poslat/tisknout výsledky každ. výpočtu do okna protokolu Storno

Rozdíl mezi dvěma korelačními koeficienty

r1: 0,00 N1: 10 r2: 0,00 N2: 10 p: 1,0000 Jednostr. Výpočet  
 Jednostr.  Oboustr.

Rozdíl mezi dvěma průměry (normální rozdělení)

Pr1: 57 SmOd1: 3,5777 N1: 6 p: .0217 Výpočet  
 Pr2: 51,6 SmOd2: 2,7019 N2: 5 Jednostr.  Oboustr.

Výběrový průměr vs. střední hodnota

Rozdíl mezi dvěma poměry

P 1: .50000 N1: 10 P 2: .50000 N2: 10 p: 1,0000 Jednostr. Výpočet  
 Jednostr.  Oboustr.

Dostaneme p-hodnotu 0,0217, tedy zamítáme nulovou hypotézu na hladině významnosti 0,05.

**Úkol k samostatnému řešení:** Do systému STATISTICA načtete datový soubor studentky.sta, který obsahuje údaje o výšce 48 studentek VŠE v Praze (proměnná vyska) a obor jejich studia (1 – národní hospodářství, 2 – informatika).

- a) Pomocí S-W testu ověřte na hladině významnosti 0,1 předpoklad o normalitě výšek v obou skupinách studentek
- b) Na hladině významnosti 0,1 testujte hypotézu o shodě rozptylů výšek studentek v daných dvou oborech studia.
- c) Na hladině významnosti 0,1 testujte hypotézu o shodě středních hodnot výšek studentek v daných dvou oborech studia.

Výpočet doplňte krabicovými diagramy.

**Výsledek:**

ad a) p-hodnota S-W testu pro studentky oboru nh je 0,6068 a pro studentky oboru informatika je 0,1119, tedy na hladině významnosti 0,1 hypotézu o normalitě nezamítáme ani v jednom případě.

ad b) Protože p-hodnota F-testu je 0,1249, což je větší než hladina významnosti 0,1, nulovou hypotézu o shodě rozptylů nezamítáme na hladině významnosti 0,1.

ad c) Protože p-hodnota dvouvýběrového t-testu je 0,0878, což je menší než hladina významnosti 0,1, nulovou hypotézu o shodě středních hodnot zamítáme na hladině významnosti 0,1.

**Úkol 3.: Asymptotický interval spolehlivosti pro parametr  $\vartheta$  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  $X_1, \dots, X_{1000}$ , přičemž  $X_i = 1$ , když i-tá osoba se vysloví pro danou politickou stranu a  $X_i = 0$  jinak,  $i = 1, \dots, 1000$ . Tyto náhodné veličiny tvoří náhodný výběr z rozložení  $A(\vartheta)$ . V tomto případě  $n = 1000$ ,  $m = 60/1000 = 0,06$ ,  $\alpha = 0,05$ ,  $u_{1-\alpha} = u_{0,95} = 1,645$ .

Ověření podmínky  $n\vartheta(1-\vartheta) > 9$ : parametr  $\vartheta$  neznáme, musíme ho nahradit výběrovým průměrem. Pak  $1000 \cdot 0,06 \cdot 0,94 = 56,4 > 9$ .

95% levostranný interval spolehlivosti pro  $\vartheta$  je

$$\left( m - \sqrt{\frac{m(1-m)}{n}} u_{1-\alpha} ; \infty \right) = \left( 0,06 - \sqrt{\frac{0,06(1-0,06)}{1000}} u_{0,95} ; \infty \right).$$

V našem případě

$$d = 0,06 - \sqrt{\frac{0,06 \cdot 0,94}{1000}} \cdot 1,645 = 0,0476$$

S pravděpodobností přibližně 0,95 tedy  $\vartheta > 0,048$ . Protože tento interval zahrnuje i hodnoty nižší než 0,05, nelze vyloučit, že strana získá méně než 5% hlasů.

**Postup ve STATISTICE:**

Statistiky – Analýza síly testu – Odhad intervalu – Jeden podíl, Z, Chí-kvadrát test – OK – Pozorovaný podíl p: 0,06, Velikost vzorku: 1000, Spolehlivost: 0,9 – Vypočítat.

Dostaneme tabulku:

	Hodnota
Podíl vzorku p	0,0600
Velikost vz. ve skup. (N)	1000,0000
Interval spolehlivosti	0,9000
Meze spolehlivosti:	
Pí (přesně):	
Dolní mez	0,0481
Horní mez	0,0738
Pí (přibližně):	
Dolní mez	0,0483
Horní mez	0,0741
Pí (původ.):	
Dolní mez	0,0476
Horní mez	0,0724

Zajímá nás výsledek uvedený v dolní části tabulky, tj. Pí (původ.).

Protože dolní mez oboustranného 90% intervalu spolehlivosti pro parametr  $\vartheta$  je shodná s dolní mezí 95% jednostranného intervalu spolehlivosti, můžeme konstatovat, že voliči budou volit danou politickou stranu s pravděpodobností aspoň 4,76%. Na základě uvedených dat strana tedy nemá zaručeno, že překročí 5% hranici pro vstup do parlamentu.

**Úkol k samostatnému řešení:** Přírůstky cen akcií na burze (v %) u 10 náhodně vybraných společností dosáhly těchto hodnot: 10, 16, 5, 10, 12, 8, 4, 6, 5, 4. Sestrojte 95% asymptotický empirický interval spolehlivosti pro pravděpodobnost, že přírůstek ceny akcie překročí 8,5%.

**Výsledek:**  $0,096 < \vartheta < 0,704$  s pravděpodobností aspoň 0,95.

Znamená to, že pravděpodobnost, že přírůstek ceny akcie překročí 8,5%, je aspoň 9,6% a nanejvýš 70,4% (při spolehlivosti 95%).

#### Úkol 4.: Testování hypotézy o parametru $\vartheta$ 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  $X_1, \dots, X_{150}$  z rozložení  $A(0,3)$ . Testujeme  $H_0: \vartheta = 0,3$  proti jednostranné alternativě  $H_1: \vartheta < 0,3$ . V tomto případě je testovým kritériem statistika

$$T_0 = \frac{M - c}{\sqrt{\frac{c(1-c)}{n}}}, \text{ která v případě platnosti nulové hypotézy má asymptoticky rozložení } N(0,1).$$

Musíme ověřit splnění podmínky  $n\vartheta(1-\vartheta) > 9$ :  $150 \cdot 0,3 \cdot 0,7 = 31,5 > 9$ . Vypočteme realizaci

$$\text{testového kritéria: } t_0 = \frac{m - c}{\sqrt{\frac{c(1-c)}{n}}} = \frac{\frac{38}{150} - 0,3}{\sqrt{\frac{0,3(1-0,3)}{150}}} = -1,24722. \text{ Kritický obor:}$$

$W = (-\infty, -u_{1-\alpha}) = (-\infty, -1,645)$ . Protože testové kritérium nepatří do kritického oboru,  $H_0$  nezamítáme na asymptotické hladině významnosti 0,05.

**Postup ve STATISTICE:**

Test provedeme pomocí 95% pravostranného intervalu spolehlivosti, který vypočítáme v modulu Analýza síly testu.

Statistiky – Analýza síly testu – Odhad intervalu – Jeden podíl, Z, Chí-kvadrát test – OK – Pozorovaný podíl p: 0,2533, Velikost vzorku: 150, Spolehlivost: 0,9 – Vypočítat.

Dostaneme tabulku:

	Hodnota
Podíl vzorku p	0,2533
Velikost vz. ve skup. (N)	150,0000
Interval spolehlivosti	0,9000
Meze spolehlivosti:	
Pí (přesně):	
Dolní mez	0,1957
Horní mez	0,3185
Pí (přibližně):	
Dolní mez	0,1966
Horní mez	0,3193
Pí (původ.):	
Dolní mez	0,1949
Horní mez	0,3117

Zajímá nás výsledek uvedený v dolní části tabulky, tj. Pí (původ.).

Protože horní mez oboustranného 90% intervalu spolehlivosti pro parametr  $\vartheta$  je shodná s horní mezí 95% pravostranného intervalu spolehlivosti, vidíme, že  $0,3 \in (0; 0,3117)$ , tudíž nelze na asymptotické hladině významnosti 0,05 zamítnout hypotézu, že zájem o danou zemi se nezměnil.