

MASARYKOVA UNIVERZITA

Přírodovědecká fakulta



**Testovanie hypotéz o parametroch  
polohy a variability v systéme  
STATISTICA**

Bakalárska práca

Brno 2008

Stanislav Abaffy

## **Prehlásenie**

Prehlasujem, že som bakalársku prácu vypracoval samostatne podľa pokynov vedúcej bakalárskej práce a všetky použité materiály sú uvedené v zozname literatúry.

V Brne dňa .....

.....  
Stanislav Abaffy

## **Pod'akovanie**

Ďakujem vedúcej bakalárskej práce RNDr. Marii Budíkovej, Dr. za ústretovosť, cenné rady a odborné vedenie, ktoré mi poskytla pri spracovaní témy bakalárskej práce.

# Obsah

Úvod	6
<b>1 Testy hypotéz o parametroch jedného náhodného výberu</b>	<b>7</b>
1.1 Testy s náhodným výberom pochádzajúcim z normálneho rozloženia . . . .	7
1.1.1 Jednovýberový t-test . . . . .	7
1.1.2 Test o rozptyle . . . . .	8
1.2 Testy s náhodným výberom nepochádzajúcim z normálneho rozloženia . .	9
1.2.1 Znamienkový test . . . . .	9
1.2.2 Jednovýberový Wilcoxonov test . . . . .	11
<b>2 Testy hypotéz o parametroch dvojrozmerného náhodného výberu</b>	<b>14</b>
2.1 Testy s náhodným výberom pochádzajúcim z dvojrozmerného normálneho rozloženia . . . . .	14
2.1.1 Párový t-test . . . . .	14
2.1.2 Morgan-Pitmanov test . . . . .	15
2.2 Testy s náhodným výberom nepochádzajúcim z dvojrozmerného normálneho rozloženia . . . . .	17
2.2.1 Párový znamienkový test . . . . .	17
2.2.2 Párový Wilcoxonov test . . . . .	18
<b>3 Testy hypotéz o parametroch 2 nezávislých náhodných výberov</b>	<b>20</b>
3.1 Testy s náhodnými výbermi pochádzajúcimi z normálneho rozloženia . . .	20
3.1.1 Dvojvýberový t-test . . . . .	20
3.1.2 F-test . . . . .	22
3.2 Testy s náhodnými výbermi nepochádzajúcimi z normálneho rozloženia . .	23
3.2.1 Dvojvýberový Wilcoxonov test . . . . .	23
3.2.2 Wald-Wolfowitzov test . . . . .	25
3.2.3 Dvojvýberový Kolmogorov-Šmirnovov test . . . . .	26
<b>4 Testy homogenity rozptylov viacerých nezávislých náhodných výberov</b>	<b>29</b>
4.1 Levenov test . . . . .	29
4.2 Brown-Forsythov test . . . . .	31

---

Záver	33
Použitá literatúra	34
Prílohy	36

# Úvod

Testovanie hypotéz o parametroch polohy (stredná hodnota) a variability (rozptyl) je dôležitou súčasťou štatistiky, má široké využitie pri riešení praktických úloh v ľubovoľnej oblasti. Spočíva v zamietnutí, resp. nezamietnutí nulovej hypotézy, t.j. určitého predpokladu, ktorý si stanovíme dopredu bez akéhokoľvek ovplyvnenia základným súborom, teda nameranými hodnotami.

V tejto práci sa čitateľ oboznámi s najdôležitejšími testami s jedným, dvoma a viacerými náhodnými výbermi a ich vykonaním v štatistickom systéme STATISTICA s využitím rôznorodých príkladov. Vo výsledkoch testov vyriešených v tomto systéme je najpodstatnejšou hodnotou dosiahnutá hladina zodpovedajúcej testovej štatistiky, tzv. p-hodnota (anglicky P-value, significance value), čo je najmenšia hladina testu, pri ktorej by sme ešte nulovú hypotézu zamietli [7, str.6]. Dáta sa ešte môžu testovať pomocou diagnostických grafov a testov normality, avšak vzhľadom na rozsah práce som tieto testy vynechal.

Práca je rozdelená do 4 kapitol, na konci sa nachádzajú ešte prílohy s 2 testami. Prvá kapitola sa venuje testom hypotéz s jedným jednorozmerným náhodným výberom vrátane najpoužívanejšieho z týchto testov - jednovýberového t-testu. Druhá sa zaoberá dvojrozmerným náhodným výberom, obsahuje okrem iného aj Morgan-Pitmanov test, ktorý je veľmi náročné nájsť, existuje malé množstvo publikácií, ktoré ho opisujú, z nich uvediem napríklad [1]. Tretia kapitola popisuje testy s dvoma náhodnými výbermi. Všetky tieto kapitoly sa členia na 2 časti podľa toho, či náhodný výber pochádza z normálneho rozloženia alebo nie. Vo štvrtej kapitole sú uvedené 2 testy s viacerými náhodnými výbermi - Levenov a Brown-Forsythov. V prílohách uvádzam mnou vytvorené zdrojové kódy testu o rozptyle a Morgan-Pitmanovho testu.

# Kapitola 1

## Testy hypotéz o parametroch jedného náhodného výberu

### 1.1 Testy s náhodným výberom pochádzajúcim z normálneho rozloženia

#### 1.1.1 Jednovýberový t-test

**Definícia.** Nech  $X_1, \dots, X_n$  je náhodný výber z  $N(\mu, \sigma^2)$ , kde  $\sigma^2$  nepoznáme. Nech  $n \geq 2$  a  $c$  je konštanta. Test  $H_0 : \mu = c$  proti  $H_1 : \mu \neq c$  (v prípade jednostranných alternatív  $H_1 : \mu < c$  alebo  $H_1 : \mu > c$ ) sa nazýva **jednovýberový t-test**.

**Návod.**

Nulovú hypotézu  $H_0 : \mu = c$  proti  $H_1 : \mu \neq c$  ( $H_1 : \mu < c$ ,  $H_1 : \mu > c$ ) zamietame na hladine významnosti  $\alpha$ , ak:

$$\left| \frac{m - c}{\frac{s}{\sqrt{n}}} \right| \geq t_{1-\alpha/2}(n-1) \quad (1.1)$$

$$\left( \text{resp. } \frac{m - c}{\frac{s}{\sqrt{n}}} \leq -t_{1-\alpha}(n-1), \text{ resp. } \frac{m - c}{\frac{s}{\sqrt{n}}} \geq t_{1-\alpha}(n-1) \right). \quad (1.2)$$

**Príklad.** Systematická chyba meracieho prístroja sa eliminuje jeho nastavením a meraním etalónu, ktorého správnou nameranou hodnotou je  $\mu = 10,00$ . Nezávislými meraniami za rovnakých podmienok boli získané hodnoty 10,24; 10,12; 9,91; 10,19; 9,78; 10,14; 9,86; 10,17 a 10,05, ktoré považujeme za realizácie náhodného výberu z  $N(\mu, \sigma^2)$ . Je možné pri riziku  $\alpha = 0,05$  vysvetliť odchýlky od hodnoty 10,00 náhodnými vplyvmi?

*Riešenie.*

$m = 10,0511$ ,  $s = 0,1627$ , na hladine významnosti  $\alpha = 0,05$  testujeme hypotézu

$H_0 : \mu = 10,00$  proti obojstrannej alternatíve  $H_1 : \mu \neq 10,00$ . Testové kritérium

$$\frac{m - c}{\frac{s}{\sqrt{n}}} = \frac{10,0511 - 10}{\frac{0,1627}{\sqrt{9}}} = 0,9426.$$

Absolútnu hodnotu testového kritéria porovnáme s kvantilom  $t_{0,975}(8) = 2,3040$ . Keďže  $0,9426 \leq 2,3040$ , nezamietame nulovú hypotézu na hladine významnosti 0,05 a teda odchýlky je možné vysvetliť iba náhodnými vplyvmi.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o jednej premennej a 9 prípadoch a vložíme namerané hodnoty. V *Basic Statistics/Tables* vyberieme možnosť *t-test, single sample*, do Reference values zadáme hodnotu 10, danú zo zadania. V tabuľke, ktorá vznikne po otestovaní, nás zaujíma najmä p-hodnota testu, prípadne hodnota testovej štatistiky.

Variable	Test of means against reference constant (value) (ttest)							
	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
Var1	10,05111	0,162669	9	0,054223	10,00000	0,942611	8	0,373470

Vidíme, že hodnota testovej štatistiky je  $t = 0,942611$  a p-hodnota testu  $p = 0,37347$ . Pretože p-hodnota je väčšia ako hladina významnosti 0,05, nulovú hypotézu nemôžeme na tejto hladine zamietnuť.

### 1.1.2 Test o rozptyle

**Definícia.** Nech  $X_1, \dots, X_n$  je náhodný výber z  $N(\mu, \sigma^2)$ , kde  $\mu$  nepoznáme. Nech  $n \geq 2$  a  $c$  je konštanta. Test  $H_0 : \sigma^2 = c$  proti  $H_1 : \sigma^2 \neq c$  (príp.  $H_1 : \sigma^2 < c$  alebo  $H_1 : \sigma^2 > c$ ) sa nazýva **test o rozptyle**.

Nulovú hypotézu  $H_0 : \sigma^2 = c$  proti  $H_1 : \sigma^2 \neq c$  ( $H_1 : \sigma^2 < c$ ,  $H_1 : \sigma^2 > c$ ) zamietame na hladine významnosti  $\alpha$ , ak:

$$\frac{(n-1)s^2}{c} \in (-\infty, \chi_{\alpha/2}^2(n-1)) \cup (\chi_{1-\alpha/2}^2(n-1), \infty) \quad (1.3)$$

$$\left( \text{resp. } \frac{(n-1)s^2}{c} \leq \chi_{\alpha}^2(n-1) \text{ alebo } \frac{(n-1)s^2}{c} \geq \chi_{1-\alpha}^2(n-1) \right). \quad (1.4)$$

**Príklad.** Do obchodu sú dodávané balíčky cukríkov, ktoré sú plnené automaticky. Automat bol skonštruovaný tak, aby smerodatná odchýlka hmotnosti balíčka činila 10 g. Predpokladáme, že hmotnosť automaticky plnených balíčkov je náhodná veličina s normálnym rozdelením. Chceme zistiť, či v priebehu času nedošlo k zhoršeniu presnosti pri plnení balíčkov (teda k zväčšeniu smerodatnej odchýlky  $\sigma$ ), ak bolo nameraných týchto 10 hodnôt: 489 473 507 498 492 477 488 503 482 491.



*Riešenie.*

Testujeme hypotézu  $H_0 : \sigma^2 = 100$  proti pravostrannej alternatíve  $H_1 : \sigma^2 > 100$ .  $m = 490$ ,  $s = 10,8218$ , testovým kritériom je  $\frac{(n-1)s^2}{c} = \frac{9 \cdot 10,8218^2}{100} = 10,54$ , kvantil Pearsonovho rozloženia zistíme z tabuliek  $\chi_{1-\alpha}^2(n-1) = \chi_{0,95}^2(9) = 16,919$ . Pretože  $10,54 < 16,919$ , nulovú hypotézu nezamietame na hladine významnosti 0,05. Test nepreukázal zníženu presnosť hmotnosti pripravovaných balíčkov.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 1 premennej a 10 prípadoch, načítame namerané hodnoty. Tento test nie je implementovaný v systéme STATISTICA, preto som vytvoril makro *testorozptyle.svb*, ktoré nám tento test pomôže vyriešiť. Jeho zdrojový kód sa nachádza v prílohách na konci práce. Otvoríme ho klasickým spôsobom ako akýkoľvek dátový súbor a pomocou klávesy F5 ho spustíme. Postupne sme vyzvaní k výberu premennej, zvoleniu hladiny významnosti a výberu typu alternatívy. Vznikne nám nasledujúca tabuľka:

	Test of Variance Against Reference Constant				
	1	2	3	4	5
	Mean	St. Dev.	K	Chi 1-alpha (n-1)	p-value
Summary	490	10,821789	10,54	16,9189776	0,308560758

Nájdeme v nej výberový priemer a výberovú smerodatnú odchýlku daného výberu, hodnotu testovej štatistiky  $K$ ,  $100(1-\alpha)\%$  kvantil Pearsonovho rozloženia o  $n-1$  stupňoch voľnosti a prislúchajúcu  $p$ -hodnotu. Keďže hodnota testovej štatistiky  $K$  je menšia ako príslušný kvantil, nulovú hypotézu o zhode rozptylu a danej konštanty nezamietame na danej hladine významnosti. To nám potvrdzuje aj  $p$ -hodnota  $p = 0,30856 >$  hladina významnosti  $\alpha = 0,05$ .

## 1.2 Testy s náhodným výberom nepochádzajúcim z normálneho rozloženia

### 1.2.1 Znamienkový test

Nech  $X_1, \dots, X_n$  je náhodný výber zo spojitého rozloženia so spojitou distribučnou funkciou  $\Phi(x)$ . Nech  $x_{0,50}$  je medián tohto rozloženia, t.j.  $\Phi(x_{0,50}) = 0,5$ . Nech  $c$  je reálna konštanta. Testujeme hypotézu  $H_0 : x_{0,50} = c$  proti obojstrannej alternatíve  $H_1 : x_{0,50} \neq c$  (resp. proti ľavostrannej alternatíve  $H_1 : x_{0,50} < c$  alebo pravostrannej alternatíve  $H_1 : x_{0,50} > c$ ).

#### Návod.

Vytvoríme rozdiely  $Y_i = X_i - c, i = 1, \dots, n$ . Ak sú niektoré rozdiely nulové, potom za  $n$  berieme len počet nenulových hodnôt.

Zavedieme štatistiku  $S_Z^+$ , ktorá udáva počet kladných rozdielov. Ak platí  $H_0$ , potom  $S_Z^+ \sim Bi(n, 1/2)$ , teda  $E(S_Z^+) = n/2$ ,  $D(S_Z^+) = n/4$ . Kritický obor budú tvoriť hodnoty

testovej štatistiky  $S_Z^+$ , ktoré sú blízke 0 alebo  $n$ , teda  $W = \langle 0, k_1 \rangle \cup \langle k_2, n \rangle$ . Pre  $n \leq 20$  a  $\alpha = 0,05$  resp.  $0,01$  sú tabelované kritické hodnoty  $k_1, k_2$ .  $H_0$  zamietame na hladine významnosti  $\alpha$ , keď  $S_Z^+ \in W$ .

Pre veľké  $n$  (v tomto prípade  $n > 20$ ) možno využiť asymptotickú normalitu štatistiky  $S_Z^+$ . Ak platí nulová hypotéza  $H_0$ , potom testová štatistika

$$U_0 = \frac{S_Z^+ - E(S_Z^+)}{\sqrt{D(S_Z^+)}} = \frac{S_Z^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \approx N(0, 1). \quad (1.5)$$

Kritickým oborom pre obojstrannú alternatívu je

$$W = (-\infty, -u_{1-\alpha/2}) \cup \langle u_{1-\alpha/2}, \infty \rangle. \quad (1.6)$$

Analogicky pre jednostranné alternatívy.  $H_0$  zamietame na asymptotickej hladine významnosti  $\alpha$ , ak  $U_0 \in W$ .

*Poznámka.* Aproximácia normálnym rozložením  $N(0, 1)$  sa zlepšuje, ak použijeme tzv. korekciu na nespojitosť. Testová štatistika  $U_0$  má potom tvar

$$U_0 = \frac{S_Z^+ - \frac{n}{2} \pm \frac{1}{2}}{\sqrt{\frac{n}{4}}}, \quad (1.7)$$

pričom  $\frac{1}{2}$  pričítame, keď  $S_Z^+ < \frac{n}{2}$ , odpočítame v opačnom prípade.

**Príklad.** Trinásť laboratórnych zvierat bolo od narodenia do 12 mesiacov veku kŕmených špeciálnou stravou. Prírastky ich hmotnosti (v gramoch) boli: 64, 69, 80, 66, 65, 77, 75, 67, 67, 68, 74, 70, 77. Môžeme z týchto hodnôt na hladine významnosti 0,05 usúdiť, že výsledky kŕmenia znamenajú v priemere prírastok váhy 70 g?

*Riešenie.*

Testujeme hypotézu  $H_0 : x_{0,50} = 70$  proti obojstrannej alternatíve  $H_1 : x_{0,50} \neq 70$ .

$x_i$	64	69	80	66	65	77	75	67	67	68	74	70	77
$x_i - c$	-6	-1	10	-4	-5	7	5	-3	-3	-2	4	0	7

$S_Z^+ = 5$ , počet nenulových rozdielov  $n = 12$ . V tabuľkách nájdeme pre  $n = 12$ ,  $\alpha = 0,05$  kritické hodnoty 2 a 10  $\Rightarrow$  kritický obor  $W = \langle 0, 2 \rangle \cup \langle 10, 12 \rangle$ . Pretože  $S_Z^+ \notin W$ , nulovú hypotézu nezamietame na hladine významnosti 0,05. Preto môžeme usúdiť, že priemerný prírastok váhy bol 70 g.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 13 prípadoch, prvá premenná obsahuje namerané hodnoty, druhá konštantu zo zadania. Postupujeme cez *Statistics*, v *Nonparametrics* vyberieme možnosť *Comparing two dependent samples*, zvolíme premenné a otestujeme pomocou *Sign test*. Vo výslednej tabuľke nás zaujíma najmä p-hodnota testu.

Pair of Variables	Sign Test (Spreadsheet2)			
	Marked tests are significant at p < ,05000			
	No. of Non-ties	Percent v < V	Z	p-level
X & Y	12	58,33333	0,288675	0,772830

V tabuľke máme uvedené 4 hodnoty, a to počet nenulových rozdielov, ďalej percentuálne vyjadrenie počtu kladných rozdielov ( $58,3\%$  z  $12 = 7$ ), hodnotu realizácie testovej štatistiky  $U_0$  (v tabuľke označená ako  $Z$ ) a p-hodnotu. Vidíme, že p-hodnota =  $0,77283$  je väčšia ako hladina významnosti  $0,05$ , čo znamená, že nulovú hypotézu nemôžeme na tejto hladine zamietnuť.

Druhou možnosťou by bolo porovnanie hodnoty testového kritéria, ktorá sa tiež v tabuľke nachádza, s príslušným kvantilom štandardizovaného normálneho rozloženia. Avšak v tomto prípade je vhodnejšie použiť kritické hodnoty z tabuliek, pretože rozsah výberu nespĺňa podmienku asymptotickej normality štatistiky  $S_Z^+$ , t.j.  $n > 20$ .

### 1.2.2 Jednovýberový Wilcoxonov test

Nech  $X_1, \dots, X_n$  je náhodný výber zo spojitého rozloženia s hustotou  $\varphi(x)$ , ktorá je symetrická okolo mediánu  $x_{0,50}$ , t.j.  $\varphi(x_{0,50} + x) = \varphi(x_{0,50} - x)$ . Nech  $c$  je reálna konštanta. Testujeme hypotézu  $H_0 : x_{0,50} = c$  proti obojstrannej alternatíve  $H_1 : x_{0,50} \neq c$  (resp. proti ľavostrannej alternatíve  $H_1 : x_{0,50} < c$  alebo pravostrannej alternatíve  $H_1 : x_{0,50} > c$ ).

#### Návod.

Vytvoríme rozdiely  $Y_i = X_i - c, i = 1, \dots, n$ . Ak sú niektoré rozdiely nulové, potom za  $n$  berieme len počet nenulových hodnôt. Absolútne hodnoty  $|Y_i|$  usporiadame vzostupne podľa veľkosti a spočítame poradie  $R_i$ .

Zavedieme štatistiku  $S_W^+ = \sum_{Y_i > 0} R_i^+$ , čo je súčet poradí cez kladné hodnoty  $Y_i$ , analogicky

$S_W^- = \sum_{Y_i < 0} R_i^-$  je súčet poradí cez záporné hodnoty  $Y_i$ . Zároveň platí  $S_W^+ + S_W^- = n(n+1)/2$ .

Za platnosti nulovej hypotézy  $H_0$  má štatistika  $S_W^+$  strednú hodnotu  $E(S_W^+) = n(n+1)/4$  a rozptyl  $D(S_W^+) = n(n+1)(2n+1)/24$ .

$H_0$  zamietame na hladine významnosti  $\alpha$ , ak je testová štatistika  $\min(S_W^+, S_W^-)$  (pre obojstrannú alternatívu),  $S_W^+$  (pre ľavostrannú alternatívu), resp.  $S_W^-$  (pre pravostrannú alternatívu) menšia alebo rovná tabelovanej kritickkej hodnote.

Pre veľké  $n$  (v tomto prípade  $n > 30$ ) možno využiť asymptotickú normalitu štatistiky  $S_W^+$ . V prípade platnosti  $H_0$

$$U_0 = \frac{S_W^+ - E(S_W^+)}{\sqrt{D(S_W^+)}} = \frac{S_Z^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \approx N(0, 1) \quad (1.8)$$

Kritickým oborom pre obojstrannú alternatívu je

$$W = (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty). \quad (1.9)$$

Analogicky pre jednostranné alternatívy.  $H_0$  zamietame na asymptotickej hladine významnosti  $\alpha$ , ak  $U_0 \in W$ .

*Poznámka.* Jednovýberový Wilcoxonov test je silnejší ako znamienkový test, avšak je vhodný len pre náhodný výber zo symetrického rozloženia.

**Príklad.** Americkí vedci skúmali mladých ľudí, ktorí opustili školu v 16 rokoch a začali pracovať. Zistili, že ich priemerný výsledok určitého testu bol 60 bodov. Potom zadali rovnaký test náhodnému výberu ľudí, ktorí študujú i po svojich 16 rokoch a zistili nasledujúce hodnoty: 72 62 52 57 91 78 74 67 51 62 84 59 51 57 89 64 80 72 92 64 57. Môžeme na základe tohto výberu a na hladine významnosti  $\alpha = 0,05$  tvrdiť, že ľudia študujúci po svojich šestnástich narodeninách majú lepší výsledok v tomto teste?

*Riešenie.*

Testujeme hypotézu  $H_0 : x_{0,50} = 60$  proti pravostrannej alternatíve  $H_1 : x_{0,50} > 60$ .

$x_i$	72	62	52	57	78	74	67	51	62	84	59	51	57	89
$x_i - c$	12	2	-8	-3	18	14	7	-9	2	24	-1	-9	-3	29
$ x_i - c $	<b>12</b>	<b>2</b>	8	3	<b>18</b>	<b>14</b>	<b>7</b>	9	<b>2</b>	<b>24</b>	1	9	3	<b>29</b>
Poradie	<b>13,5</b>	<b>2,5</b>	10	5	<b>16</b>	<b>15</b>	<b>9</b>	11,5	<b>2,5</b>	<b>18</b>	1	11,5	5	<b>19</b>

$x_i$	64	80	72	92	64	57
$x_i - c$	4	20	12	32	4	-3
$ x_i - c $	4	<b>20</b>	<b>12</b>	<b>32</b>	4	3
poradie	<b>7,5</b>	<b>17</b>	<b>13,5</b>	<b>20</b>	<b>7,5</b>	5

$S_W^+ = 161$ ,  $S_W^- = 49$ , počet nenulových rozdielov  $n = 20$ . V štatistických tabuľkách pre  $n = 20$ ,  $\alpha = 0,05$  je kritická hodnota rovná 149. Hodnota testovej štatistiky pre pravostrannú alternatívu  $S_W^- = 49 \leq 61$ , preto nulovú hypotézu zamietame na hladine významnosti 0,05, a teda z toho môžeme vyvodiť záver, že ľudia študujúci aj po svojich 16. narodeninách majú lepší výsledok v teste.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 20 prípadoch, do prvej premennej zadáme namerané hodnoty, do druhej konštantu 60 zo zadania. Postupujeme cez *Statistics - Non-parametrics - Comparing two dependent samples*. Zvolíme premenné a pomocou *Wilcoxon matched pairs test* tento test vykonáme. Vznikne nám nasledujúca tabuľka:

		Wilcoxon Matched Pairs Test (jednWilctest)			
		Marked tests are significant at p < ,05000			
Pair of Variables		Valid N	T	Z	p-level
Var1	& Var2	20	49,00000	2,090630	0,036562

V tomto príklade počítame s jednostrannou alternatívou, systém STATISTICA udáva vo výslednej tabuľke p-hodnotu pre obojstranný test, a teda  $\min(S_W^+, S_W^-)$  ako hodnotu testovej štatistiky. Jednostranné testy nie sú implementované, preto môže byť p-hodnota aj hodnota testovej štatistiky odlišná od tej, ktorú očakávame. V našom prípade je teda

vhodnejšie použiť ručný výpočet. Pre obojstrannú alternatívu by sme nulovú hypotézu zamietli na hladine významnosti 0,05.

# Kapitola 2

## Testy hypotéz o parametroch dvojrozmerného náhodného výberu

### 2.1 Testy s náhodným výberom pochádzajúcim z dvojrozmerného normálneho rozloženia

#### 2.1.1 Párový t-test

**Definícia.** Nech  $(\begin{smallmatrix} X_1 \\ Y_1 \end{smallmatrix}), \dots, (\begin{smallmatrix} X_n \\ Y_n \end{smallmatrix})$  je náhodný výber z rozloženia  $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$ , pričom  $n \geq 2$ . Označíme  $\mu = \mu_1 - \mu_2$  a zavedieme rozdielový náhodný výber

$$Z_1 = X_1 - Y_1, \dots, Z_n = X_n - Y_n. \quad (2.1)$$

Odtiaľ  $M = \frac{1}{n} \sum_{i=1}^n Z_i$ ,  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - M)^2$ .

Testujeme  $H_0 : \mu_1 - \mu_2 = 0$  (čiže  $\mu = 0$ ) proti  $H_1 : \mu_1 - \mu_2 \neq 0$  (čiže  $\mu \neq 0$ ), prípadne proti jednostranným alternatívam  $H_1 : \mu_1 - \mu_2 < 0$  alebo  $H_1 : \mu_1 - \mu_2 > 0$ . Ďalej už postupujeme ako pri jednovýberovom t-teste.

**Príklad.** Nakoľko ovplyvňuje zaujímavé prostredie skutočný fyzický rozvoj mozgu? Pre rozriešenie tejto otázky, zatiaľ aspoň pre pokusné krysy, podnikol v roku 1964 Rosenzweig pokusy s 10 vrhmi čistokrvných laboratorných krýs. Z každého vrhu bola vždy jedna krysa vybraná do pokusnej skupiny a jedna do kontrolnej skupiny. Obe skupiny boli potom chované v úplne rovnakých podmienkach iba s výnimkou, že pokusné krysy žili pohromade v terárii s mnohými zaujímavými hračkami, kdežto krysy z kontrolnej skupiny žili v totálnej izolácii. Po mesiaci boli krysy zabitú a ich mozgová kôra (najvyvinutejšia časť mozgu) zvážená s nasledujúcimi výsledkami (v centigramoch) pre 10 párov:

Vrh	1	2	3	4	5	6	7	8	9	10
Pokus	68	65	66	66	67	66	66	64	69	63
Kontrola	65	62	64	65	65	64	59	63	65	58

*Riešenie.*

Označme  $\mu = \mu_1 - \mu_2$ , testujeme hypotézu  $H_0 : \mu = 0$  proti obojstrannej alternatíve  $H_1 : \mu \neq 0$ . Vypočítame  $m = 2,7$ ,  $s = 1,3375$  a testové kritérium  $t_0 = \frac{m-\mu}{s}\sqrt{n} = 6,38$ . Absolútnu hodnotu testového kritéria porovnáme s kvantilom  $t_{0,975}(9) = 2,2622$ . Pretože  $6,38 \geq 2,2622$ , zamietame nulovú hypotézu na hladine významnosti 0,05, teda s rizikom omylu najviac 5 % môžeme tvrdiť, že zaujímavé prostredie ovplyvňuje fyzický vývoj mozgu.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 10 prípadoch, do prvej premennej zadáme namerané hodnoty pokusnej skupiny, do druhej hodnoty kontrolnej skupiny. Postupujeme cez *Statistics - Basic Statistics/Tables - t-test, dependent samples*. Zvolíme premenné a pomocou *Summary: T-tests* tento test vykonáme. Vznikne nám nasledujúca tabuľka:

T-test for Dependent Samples (parovyttest)								
Marked differences are significant at p < ,05000								
Variable	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
Var1	66,00000	1,763834						
Var2	63,00000	2,581989	10	3,000000	1,885618	5,031153	9	0,000708

V tabuľke nás zaujíma predovšetkým p-hodnota, ktorá je veľmi blízka nule a menšia než 0,05, preto zamietame nulovú hypotézu. Vidíme, že by sme ju zamietli aj pri oveľa nižšej hladine významnosti.

### 2.1.2 Morgan-Pitmanov test

**Definícia.** Nech  $(X_1), \dots, (X_n)$  je náhodný výber z rozloženia  $N_2 \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right)$ . Nech  $\sigma_1^2 > 0$ ,  $\sigma_2^2 > 0$ ,  $\rho \in (-1, 1)$  ( $\sigma_{12} = \rho\sigma_1\sigma_2$ ) a  $n \geq 3$ . Test hypotézy  $H_0 : \sigma_1^2 = \sigma_2^2$  proti  $H_1 : \sigma_1^2 \neq \sigma_2^2$  (resp. proti jednostranným alternatívam  $H_1 : \sigma_1^2 < \sigma_2^2$  či  $H_1 : \sigma_1^2 > \sigma_2^2$ ) sa nazýva Morgan-Pitmanov test.

**Návod.**

Postupne označíme

$$M_1 = \frac{1}{n} \sum_{i=1}^n X_i, \quad M_2 = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$S_1^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_1)^2, \quad S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - M_2)^2.$$

Vypočítame výberový korelačný koeficient

$$R = \frac{\sum_{i=1}^n X_i Y_i - n M_1 M_2}{\sqrt{(\sum_{i=1}^n X_i^2 - n M_1^2)(\sum_{i=1}^n Y_i^2 - n M_2^2)}} \quad (2.2)$$

a následne testovú štatistiku  $T_0$  položíme rovnú

$$T_0 = \frac{S_1^2 - S_2^2}{2S_1S_2} \sqrt{\frac{n-2}{1-R^2}} \quad (2.3)$$

Nulovú hypotézu  $H_0$  zamietame v prípade, že  $|T_0| \geq t_{1-\alpha/2}(n-2)$ , resp.  $T_0 \leq -t_{1-\alpha}(n-2)$  a  $T_0 \geq t_{1-\alpha}(n-2)$  pri jednostranných alternatívach.

**Príklad.** U 10 vzorkov železnej rudy bol stanovovaný obsah železa. Použili sa pritom 2 metódy. Jedna z nich je klasická a jej výsledky sú označené  $X_i$ . Druhá z nich je nová, jej výsledkami sú  $Y_i$ . Máme porovnať rozptyly oboch metód na základe uvedených dát.

$X_i$	36,1	40,6	35,0	39,3	31,2	38,6	31,8	36,1	36,9	35,2
$Y_i$	35,2	49,6	38,3	48,6	27,6	39,9	28,5	37,3	35,8	34,3

*Riešenie.*

Najskôr si vypočítame výberové priemery a smerodatné odchýlky dvojrozmerného výberu. Zistíme, že  $m_1 = 36,08$ ,  $m_2 = 37,51$ ,  $s_1 = 3,0165$ ,  $s_2 = 7,2533$ . Hodnota výberového korelačného koeficientu činí  $r = 0,9335$ , hodnota testovej štatistiky  $t_0 = -7,8426$ , kvantil Studentovho rozloženia  $t_{0,975}(8) = 2,306$ . Keď porovnáme absolútnu hodnotu testovej štatistiky s kvantilom, vidíme, že  $7,8426 \geq 2,306$ , teda zamietame nulovú hypotézu o zhode rozptylov daného dvojrozmerného výberu.

*Poznámka.* Tento príklad je riešený v [1, str.79], vyskytla sa tam však chyba pri výpočte hodnoty výberového korelačného koeficientu, preto na tomto mieste uvádzam správne riešenie.

### Postup v programe STATISTICA.

Tento test nie je implementovaný v systéme STATISTICA, preto je na riešenie použité makro *MorganPitman.sub*, zdrojový kód je uvedený v prílohách na konci práce.

Vytvoríme dátový súbor o 2 premenných a 10 prípadoch. Do prvej premennej načítame výsledky klasickej metódy, do druhej výsledky novej metódy. Otvoríme makro, spustíme ho pomocou klávesy F5. Po zadaní premenných, hladiny významnosti a typu alternatívy nám vznikne nasledujúca tabuľka:

	Morgan-Pitman Test of Homogeneity of Variances							
	1	2	3	4	5	6	7	8
	Mean1	Mean2	St. Dev.1	St. Dev.2	Coef. of corr.	T	t 1-alpha/2 (n-2)	p-value
Summary	36,08	37,51	3,016547	7,2532675	0,933493278	-7,84259	2,30600414	0,00005037

V tabuľke máme možnosť vidieť výberové priemery a výberové smerodatné odchýlky oboch výberov, hodnotu výberového korelačného koeficientu, hodnotu testovej štatistiky  $t_0 = -7,84259$ , kvantil Studentovho rozloženia  $t_{0,975}(8) = 2,306$  a prislúchajúcu p-hodnotu. Absolútna hodnota testového kritéria je väčšia ako kvantil, preto nulovú hypotézu o zhode rozptylov zamietame na hladine významnosti 0,05. Zamietnutie nulovej hypotézy nám potvrdzuje aj p-hodnota  $p = 0,00005$ , ktorá je oveľa menšia ako zvolená hladina významnosti 0,05.



## 2.2 Testy s náhodným výberom nepochádzajúcim z dvojrozmerného normálneho rozloženia

### 2.2.1 Párový znamienkový test

Nech  $(X_1, Y_1), \dots, (X_n, Y_n)$  je náhodný výber zo spojitého dvojrozmerného rozloženia. Nech  $x_{0,50}$  a  $y_{0,50}$  sú mediány tohto rozloženia, t.j.  $\Phi(x_{0,50}) = 0,5$ , resp.  $\Phi(y_{0,50}) = 0,5$ . Nech  $c$  je reálna konštanta. Testujeme hypotézu  $H_0 : x_{0,50} - y_{0,50} = c$  proti  $H_1 : x_{0,50} - y_{0,50} \neq c$  (resp. proti ľavostrannej alternatíve  $H_1 : x_{0,50} - y_{0,50} < c$  alebo pravostrannej alternatíve  $H_1 : x_{0,50} - y_{0,50} > c$ ).

#### Návod.

Vytvoríme rozdiely  $Z_i = X_i - Y_i, i = 1, \dots, n$  a testujeme hypotézu o mediáne  $H_0 : z_{0,50} = c$  proti  $H_1 : z_{0,50} \neq c$ , resp.  $H_1 : z_{0,50} < c$  alebo  $H_1 : z_{0,50} > c$ .

Zavedieme štatistiku  $S_Z^+$ , ktorá udáva počet kladných rozdielov. Kritický obor budú tvoriť hodnoty testovej štatistiky  $S_Z^+$ , ktoré sú blízke 0 alebo  $n$ , teda  $W = \langle 0, k_1 \rangle \cup \langle k_2, n \rangle$ . Pre  $n \leq 20$  a  $\alpha = 0,05$  resp.  $0,01$  sú tabelované kritické hodnoty  $k_1, k_2$ .  $H_0$  zamietame na hladine významnosti  $\alpha$ , keď  $S_Z^+ \in W$ .

Pre veľké  $n > 20$  postupujeme podobne ako pri znamienkovom teste.

**Príklad.** Pre overenie účinnosti dvoch prací prostriedkov bol vykonaný nasledujúci experiment. 12 rôzne špinavých kusov látky bolo rozpolených. Pre prvú polovicu látok bol vyskúšaný jeden prací prostriedok a pre druhú polovicu druhý. Po usušení boli vyprané vzorky ohodnotené stupnicou do 10 bodov podľa kvality vyprania. Výsledky testu sú v nasledujúcej tabuľke:

Číslo kusu	1	2	3	4	5	6	7	8	9	10	11	12
Body po použití 1. prostr.	9	8	7	9	7	7	7	8	7	9	7	8
Body po použití 2. prostr.	8	10	8	8	9	9	8	10	9	9	8	9

Môžeme tvrdiť na 5% hladine významnosti, že medián rozdielov bodov nie je nulový?

*Riešenie.*

Všetky rozdiely si prehľadne zapíšeme do tabuľky:

Číslo kusu	1	2	3	4	5	6	7	8	9	10	11	12
Body po použití 1. prostr.	9	8	7	9	7	7	7	8	7	9	7	8
Body po použití 2. prostr.	8	10	8	8	9	9	8	10	9	9	8	9
2. - 1.	-1	2	1	-1	2	2	1	2	2	0	1	1
2. - 1. - c	-1	2	1	-1	2	2	1	2	2	0	1	1

Testová štatistika  $S_Z^+$  nadobúda hodnotu 9, počet nenulových rozdielov je 11. V štatistických tabuľkách pre  $n = 11$ ,  $\alpha = 0,05$  nájdeme kritické hodnoty  $k_1 = 1$ ,  $k_2 = 10$ . Kritickým oborom je teda  $W = \langle 0, 1 \rangle \cup \langle 10, 11 \rangle$ .  $S_Z^+ \notin W$ , nezamietame nulovú hypotézu, nepreukázali sme výrazné rozdiely medzi mediánmi rozdielov oboch metód.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 12 prípadoch. Prvá premenná obsahuje body po použití 1. prostriedku, druhá premenná body po použití 2. prostriedku. Ďalej postupujeme cez *Statistics - Nonparametrics - Comparing two dependent samples*, zvolíme premenné a otestujeme pomocou *Sign test*. Vo výslednej tabuľke nás zaujíma predovšetkým p-hodnota testu.

Pair of Variables		Sign Test (parsigntest)			
		No. of Non-ties	Percent v < V	Z	p-level
Var1	& Var2	11	81,81818	1,809068	0,070440

V tabuľke máme uvedené 4 hodnoty: počet nenulových rozdielov, percentuálne vyjadrenie počtu kladných rozdielov ( $81,81\%$  z  $11 = 9$ ), hodnotu realizácie testovej štatistiky  $U_0$  (v tabuľke označená ako  $Z$ ) a p-hodnotu. Vidíme, že p-hodnota =  $0,070440$  je väčšia ako hladina významnosti  $0,05$ , čo znamená, že nulovú hypotézu nemôžeme na tejto hladine zamietnuť.

Druhou možnosťou by bolo porovnanie hodnoty testovej štatistiky s príslušným kvantilom štandardizovaného normálneho rozloženia. Podobne ako pri znamienkovom teste je vhodnejšie použiť kritické hodnoty z tabuliek, pretože rozsah výberu nespĺňa podmienku asymptotickej normality štatistiky  $S_Z^+$ , t.j.  $n > 20$ .

### 2.2.2 Párový Wilcoxonov test

Nech  $(X_1, Y_1), \dots, (X_n, Y_n)$  je náhodný výber zo spojitého rozloženia. Nech  $c$  je reálna konštanta. Testujeme hypotézu  $H_0 : x_{0,50} - y_{0,50} = c$  proti obojstrannej alternatíve  $H_1 : x_{0,50} - y_{0,50} \neq c$  (resp. proti ľavostrannej alternatíve  $H_1 : x_{0,50} - y_{0,50} < c$  alebo pravostrannej alternatíve  $H_1 : x_{0,50} - y_{0,50} > c$ ).

#### Návod.

Vytvoríme rozdiely  $Z_i = X_i - Y_i, i = 1, \dots, n$ . Za  $n$  berieme len počet nenulových rozdielov. Absolútne hodnoty  $|Z_i|$  usporiadame vzostupne podľa veľkosti a spočítame poradie  $R_i$ . Ďalej postupujeme podobne ako pri jednovýberovom Wilcoxonovom teste.

**Príklad.** Zoberme si rovnaký príklad ako pri párovom znamienkovom teste, aby sme dokázali, že párový Wilcoxonov test je silnejší.

#### Riešenie.

Číslo kusu	1	2	3	4	5	6	7	8	9	10	11	12
Body po použití 1. prostr.	9	8	7	9	7	7	7	8	7	9	7	8
Body po použití 2. prostr.	8	10	8	8	9	9	8	10	9	9	8	9
2. - 1.	-1	2	1	-1	2	2	1	2	2	0	1	1
$ 2. - 1. - c $	1	2	1	1	2	2	1	2	2	0	1	1
Poradie	3,5	<b>9</b>	<b>3,5</b>	3,5	<b>9</b>	<b>9</b>	<b>3,5</b>	<b>9</b>	<b>9</b>	-	<b>3,5</b>	<b>3,5</b>

$S_W^+ = 59$ ,  $S_W^- = 7$ ,  $n = 11$ ,  $\alpha = 0,05$ . Kritická hodnota z tabuliek = 10, testová štatistika je rovná  $\min(59, 7) = 7$ . Pretože hodnota testovej štatistiky je menšia ako kritická hodnota, zamietame nulovú hypotézu na hladine významnosti 0,05. Vidíme, že Wilcoxonov test nám zamietol hypotézu o rovnosti mediánov, je preto silnejší ako znamienkový.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 12 prípadoch, do prvej premennej zadáme body po použití 1. prostriedku, do druhej body po použití druhého prostriedku. Postupujeme cez *Statistics - Nonparametrics - Comparing two dependent samples*. Zvolíme premenné a pomocou *Wilcoxon matched pairs test* tento test vykonáme. Vznikne nám nasledujúca tabuľka:

Pair of Variables		Wilcoxon Matched Pairs Test (parwilctest)			
		Marked tests are significant at p <,05000			
Valid	N	T	Z	p-level	
Var1	& Var2	12	7,000000	2,311682	0,020796

Je to ten istý typ tabuľky ako pri jednovýberovom Wilcoxonovom teste, premenná  $T$  udáva hodnotu testovej štatistiky, teda  $\min(S_W^+, S_W^-)$ , premenná  $Z$  reprezentuje asymptotickú testovú štatistiku  $U_0$  (použiteľná pri väčších výberoch) a ako posledná je daná zodpovedajúca p-hodnota. Keďže je menšia ako hladina významnosti 0,05, nulovú hypotézu o rovnosti mediánov zamietame.

# Kapitola 3

## Testy hypotéz o parametroch 2 nezávislých náhodných výberov

### 3.1 Testy s náhodnými výbermi pochádzajúcimi z normálneho rozloženia

#### 3.1.1 Dvojvýberový t-test

**Definícia.** Nech  $X_{11}, \dots, X_{1n_1}$  je náhodný výber pochádzajúci z rozloženia  $N(\mu_1, \sigma^2)$ ,  $X_{21}, \dots, X_{2n_2}$  je na ňom nezávislý náhodný výber z rozloženia  $N(\mu_2, \sigma^2)$ , pričom  $n_1 \geq 2$  a  $n_2 \geq 2$ . Nech  $c$  je konštanta. Test  $H_0 : \mu_1 - \mu_2 = c$  proti  $H_1 : \mu_1 - \mu_2 \neq c$ , resp. proti jednostranným alternatívam  $H_1 : \mu_1 - \mu_2 < c$  alebo  $H_1 : \mu_1 - \mu_2 > c$ , sa nazýva **dvojvýberový t-test**.

**Návod.**

Nulovú hypotézu  $H_0 : \mu_1 - \mu_2 = c$  proti  $H_1 : \mu_1 - \mu_2 \neq c$  zamietame na hladine významnosti  $\alpha$ , ak:

$$\left| \frac{m_1 - m_2 - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \geq t_{1-\alpha/2}(n_1 + n_2 - 2) \quad (3.1)$$

Pri jednostranných alternatívach

$$\frac{m_1 - m_2 - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq -t_{1-\alpha}(n_1 + n_2 - 2), \text{ resp. } \frac{m_1 - m_2 - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \geq t_{1-\alpha}(n_1 + n_2 - 2). \quad (3.2)$$

*Poznámka.* V prípade, že nie je splnený predpoklad o rovnosti rozptylov, možno zostrojiť aspoň  $100(1 - \alpha)\%$  interval spoľahlivosti pre  $\mu_1 - \mu_2$ . V tomto prípade má štatistika

$$T = \frac{\mu_1 - \mu_2 - c}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

približne rozloženie  $t(\nu)$ , kde počet stupňov voľnosti  $\nu$  je daný vzťahom<sup>1</sup>

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Ak  $\nu$  nie je celé číslo, použijeme v tabuľkách kvantilov Studentovho rozloženia lineárnu interpoláciu.

*Poznámka.* Možno sa stretnúť ešte s inou variantou, napr. v [10, str.392] alebo [12, str.84]:

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1}} - 2$$

**Príklad.** Sú analyzované 2 katalyzátory, aby sa zistilo, ako ovplyvňujú priemerný výnos chemického procesu. Katalyzátor 1 je momentálne v prevádzke, katalyzátor 2 je k dispozícii. Vzhľadom na to, že je lacnejší, mal by byť zavedený, za predpokladu, že nezmení výnos procesu. Výsledky testu sú uvedené v nasledujúcej tabuľke.

Rastlina	1	2	3	4	5	6	7	8
Katalyzátor 1	91,5	94,18	92,18	95,39	91,79	89,07	94,72	89,21
Katalyzátor 2	89,19	90,95	90,46	93,21	97,19	97,04	91,07	92,75

Je nejaký rozdiel medzi priemernými výnosmi po použití katalyzátorov? Predpokladajme rovnosť rozptylov a hladinu významnosti  $\alpha = 0,05$ .

*Riešenie.*

Testujeme  $H_0 : \mu_1 - \mu_2 = 0$  proti  $H_1 : \mu_1 - \mu_2 \neq 0$ . Vypočítame  $m_1 = 92,255$ ,  $m_2 = 92,7325$ ,  $s_1 = 2,385$ ,  $s_2 = 2,98345$ . Pretože predpokladáme rovnosť rozptylov  $\sigma_1^2 = \sigma_2^2$ , za ich odhad vezmeme vážený priemer výberových rozptylov  $s_*^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 7,2947$ , teda  $s_* = 2,70086$ . Testové kritérium

$$t_0 = \frac{m_1 - m_2 - c}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{92,255 - 92,7325 - 0}{2,70086 \sqrt{\frac{1}{8} + \frac{1}{8}}} = -0,3536$$

Absolútnu hodnotu testového kritéria porovnáme s kvantilom  $t_{0,975}(14) = 2,1448$ . Pretože  $0,3536 < 2,1448$ , nulovú hypotézu nezamietame na hladine významnosti 0,05. Teda medzi priemernými výnosmi po použití katalyzátorov nie je významný rozdiel, druhý katalyzátor môže byť zavedený.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 8 prípadoch. Hodnoty namerané po použití prvého katalyzátora tvoria prvky prvej premennej, hodnoty namerané po použití druhého katalyzátora sú obsiahnuté v druhej premennej. Postupujeme cez *Statistics - Basic Statistics/Tables - t-test, independent, by variables*. Zvolíme premenné a pomocou *Summary: T-tests* vykonáme test. Vznikne nasledujúca tabuľka:

<sup>1</sup>napr. podľa [3, str.86] alebo [6, str.87]

		T-test for Independent Samples (dvojtest)										
		Note: Variables were treated as independent samples										
Group 1 vs. Group 2		Mean	Mean	t-value	df	p	Valid N	Valid N	Std.Dev.	Std.Dev.	F-ratio	p
Var1 vs. Var2		Group 1	Group 2				Group 1	Group 2	Group 1	Group 2	Variations	Variations
		92,25500	92,73250	-0,353591	14	0,728914	8	8	2,385019	2,983453	1,564786	0,569131

V nej nás zaujíma hodnota testového kritéria  $t_0 = -0,353591$ , prislúchajúca p-hodnota  $p = 0,728914$  je väčšia než hladina významnosti  $0,05$ , nulovú hypotézu nezamietame.

### 3.1.2 F-test

**Definícia.** Nech  $X_{11}, \dots, X_{1n_1}$  je náhodný výber pochádzajúci z rozloženia  $N(\mu_1, \sigma_1^2)$ ,  $X_{21}, \dots, X_{2n_2}$  je na ňom nezávislý náhodný výber z rozloženia  $N(\mu_2, \sigma_2^2)$ , pričom  $n_1 \geq 2$ ,  $n_2 \geq 2$ . Nech  $c$  je konštanta. Test  $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$  proti  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ , resp. proti jednostranným alternatívam  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1$  alebo  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$  sa nazýva **F-test**.

**Návod.**

Hypotézu  $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$  proti  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$  (resp.  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1$  alebo  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$ ) zamietame na hladine významnosti  $\alpha$ , ak

$$\frac{s_1^2}{s_2^2} \in (0, F_{\alpha/2}(n_1 - 1, n_2 - 1)) \cup \langle F_{1-\alpha/2}(n_1 - 1, n_2 - 1), \infty \rangle \quad (3.3)$$

$$\left( \text{resp. } \frac{s_1^2}{s_2^2} \in (0, F_{\alpha}(n_1 - 1, n_2 - 1)) \text{ alebo } \frac{s_1^2}{s_2^2} \in \langle F_{1-\alpha}(n_1 - 1, n_2 - 1), \infty \rangle \right). \quad (3.4)$$

**Príklad.** Máme k dispozícii dva výberové súbory nameraných údajov rozmeru odliatku v mm.

Výberový súbor č.1: 3,7 3,6 4,0 3,8 4,0 3,2 3,7 4,0 3,5 4,1.

Výberový súbor č.2: 3,9 4,6 3,9 4,9 4,7 3,7 4,8 5,5.

Máme otestovať, či tieto výberové súbory pochádzajú z rovnakého základného súboru.

*Riešenie.*

Testujeme teda hypotézu o zhode rozptylov  $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$  proti alternatíve  $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ . Určíme si výberové smerodatné odchýlky  $s_1 = 0,2797$ ,  $s_2 = 0,6164$ . Testovým kritériom je podiel  $\frac{s_1^2}{s_2^2} = \frac{0,2797^2}{0,6164^2} = \frac{0,0782}{0,38} = 0,2058$ . K riešeniu úlohy potrebujeme ešte kvantily Fisher-Snedecorovho rozloženia  $F_{0,025}(9, 7) = 0,2383$  a  $F_{0,975}(9, 7) = 4,8232$ . Pretože hodnota testového kritéria  $0,2058 \in (0; 0,2383) \cup \langle 4,8232, \infty \rangle$ , zamietame nulovú hypotézu o zhode rozptylov.

#### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 10 prípadoch. Hodnoty výberového súboru č.1 budú zahrnuté v prvej premennej, hodnoty súboru č.2 v druhej premennej. Postupujeme cez *Statistics - Basic Statistics/Tables - t-test, independent, by variables*. Zvolíme premenné a pomocou *Summary: T-tests* vykonáme test. Vznikne nasledujúca tabuľka:

		T-test for Independent Samples (ftest)										
		Note: Variables were treated as independent samples										
Group 1 vs. Group 2		Mean Group 1	Mean Group 2	t-value	df	p	Valid N Group 1	Valid N Group 2	Std.Dev. Group 1	Std.Dev. Group 2	F-ratio Variances	p Variances
Var1 vs. Var2		3,760000	4,500000	-3,40230	16	0,003642	10	8	0,279682	0,616441	4,857955	0,031882

Zaujímá nás hlavne hodnota testovej štatistiky F-testu, v tomto prípade 4,857955, a príslušná p-hodnota  $0,031882 \leq 0,05$ , nulovú hypotézu zamietame na hladine významnosti 0,05. Hodnota testovej štatistiky sa líši oproti ručnému výpočtu z dôvodu, že systém STATISTICA berie pri podiele  $\frac{s_1^2}{s_2^2}$  za  $s_1^2$  vždy väčšiu z hodnôt výberových smerodatných odchýlok, teda tieto výsledky sú si navzájom prevrátené hodnoty.

Pri prevrátení smerodatných odchýlok prichádza aj k menšej korekcii kritického oboru, v kvantile dochádza k prevráteniu stupňov voľnosti, t.j. namiesto  $F_{0,975}(9, 7)$  počítame s  $F_{0,975}(7, 9)$  a namiesto  $F_{0,025}(9, 7)$  počítame s  $F_{0,025}(7, 9)$ .

## 3.2 Testy s náhodnými výbermi nepochádzajúcimi z normálneho rozloženia

### 3.2.1 Dvojjvýberový Wilcoxonov test

Nech  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  sú 2 nezávislé náhodné výbery z dvoch spojitých rozložení, ktorých distribučné funkcie sa môžu líšiť len posunutím. Označme  $x_{0,50}$  medián prvého rozloženia a  $y_{0,50}$  medián druhého rozloženia. Testujeme hypotézu o rovnosti distribučných funkcií oboch rozložení  $H_0 : \Phi(x) = \Phi(y)$  alebo o rovnosti mediánov  $H_0 : x_{0,50} = y_{0,50}$  proti alternatíve ich nerovnosti  $H_1 : \Phi(x) \neq \Phi(y)$ , resp.  $H_1 : x_{0,50} \neq y_{0,50}$ .

#### Návod.

Všetkých  $n+m$  hodnôt  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  usporiadame vzostupne podľa veľkosti. Súčet poradí  $X_1, \dots, X_n$  označme  $T_1$ , súčet poradí  $Y_1, \dots, Y_m$  označme  $T_2$ . Určíme štatistiky  $U_1 = mn + n(n+1)/2 - T_1$  a  $U_2 = mn + m(m+1)/2 - T_2$ . Platí rovnosť  $U_1 + U_2 = mn$ , ktorá sa dá jednoducho dokázať.

Nulovú hypotézu zamietame na hladine významnosti  $\alpha$ , ak  $\min(U_1, U_2) \leq$  tabelovaná kritická hodnota pre dané  $m, n$  a  $\alpha$ .

Pre veľké hodnoty  $n, m$  ( $n, m > 30$ ) sa využíva asymptotická normalita štatistiky  $U_1$ . V prípade platnosti nulovej hypotézy platí

$$U_0 = \frac{U_1 - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}} \quad (3.5)$$

Kritickým oborom pre obojstrannú alternatívu je

$$W = (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty). \quad (3.6)$$

Analogicky pre jednostranné alternatívy.  $H_0$  zamietame na asymptotickej hladine významnosti  $\alpha$ , ak  $U_0 \in W$ .

**Príklad.** Majiteľ obchodu chcel zistiť, či veľkosti nákupov v USD platených kreditnými kartami MasterCard a Visa sú v priemere rovnaké. Náhodne vybral 7 nákupov platených MasterCard a 9 platených Visou.

MasterCard	42	77	46	73	78	33	37		
Visa	39	10	119	68	76	126	53	79	102

Môžeme na hladine významnosti 5% tvrdiť, že mediány nákupov platených týmito dvoma kartami sa zhodujú?

*Riešenie.*

Všetky hodnoty si usporiadame do tabuľky vzostupne podľa veľkosti a určíme ich poradie vzhľadom na zjednotenie oboch výberov.

usporiadané hodnoty	10	33	37	39	42	46	53	68	73	76	77	78	79	102
poradie MasterCard	-	2	3	-	5	6	-	-	9	-	11	12	-	-
poradie Visa	1	-	-	4	-	-	7	8	-	10	-	-	13	14

usporiadané hodnoty	119	126
poradie MasterCard	-	-
poradie Visa	15	16

$$T_1 = 2 + 3 + 5 + 6 + 9 + 11 + 12 = 48, T_2 = 1 + 4 + 7 + 8 + 10 + 13 + 14 + 15 + 16 = 88$$

$$U_1 = 7.9 + 7.8/2 - 48 = 43, U_2 = 7.9 + 9.10/2 - 88 = 20$$

Kritická hodnota pre  $\alpha = 0,05$ ,  $n = 7$  a  $m = 9$  je 12. Pretože  $\min(43, 20) = 20 > 12$ , môžeme s 5% rizikom omylu tvrdiť, že mediány nákupov platených kartami MasterCard a Visa sa zhodujú.

**Postup v programe STATISTICA.**

Vytvoríme dátový súbor o 2 premenných a 16 prípadoch, do prvej premennej načítame v ľubovoľnom poradí prvky oboch výberov, druhá premenná obsahuje hodnoty 1 alebo 2 v závislosti od príslušnosti daného prvku k prvému alebo druhému náhodnému výberu. K testu sa dostaneme cez *Statistics - Nonparametrics - Comparing two independent samples (groups)*, zadáme premenné, ako závislú zvolíme premennú 1 s nameranými hodnotami, ako grupujúcu premennú 2 s indexami oboch výberov. Test vykonáme pomocou *Mann-Whitney U Test*.

Mann-Whitney U Test (dvojitest)										
By variable Var2										
Marked tests are significant at p < .05000										
variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-level	Z adjusted	p-level	Valid N Group 1	Valid N Group 2	2*1 sided exact p
Var1	48,00000	88,00000	20,00000	-1,21729	0,223495	-1,21729	0,223495	7	9	0,252273

V tabuľke nájdeme súčty  $T_1$  a  $T_2$ , hodnotu testovej štatistiky  $\min(U_1, U_2)$  označenú ako  $U$ , hodnotu asymptotickej testovej štatistiky  $U_0$  (označená  $Z$ ), p-hodnotu príslušnú tejto štatistike, ale nás zaujíma predovšetkým p-hodnota označená *2\*1 one sided exact p*, ktorá sa používa pre rozsahy náhodných výberov menšie ako 30. Keďže je väčšia ako hladina významnosti  $\alpha = 0,05$ , nulovú hypotézu nezamietame.



### 3.2.2 Wald-Wolfowitzov test

Nech  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  sú 2 nezávislé náhodné výbery z dvoch spojitých rozložení. Testujeme hypotézu, že oba výbery pochádzajú z rovnakého rozloženia proti alternatíve, že pochádzajú z 2 rôznych rozložení.

#### Návod.

Všetkých  $n+m$  hodnôt  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  usporiadame vzostupne podľa veľkosti. Testovou hypotézou je počet iterácií  $R$ , tzn. počet postupností za sebou nasledujúcich hodnôt patriacich do toho istého výberu. Ak je  $R \leq$  tabelovaná kritická hodnota pre dané  $n, m$  a  $\alpha$ ,  $H_0$  zamietame na hladine významnosti  $\alpha$ .

Pre rozsahy výberov  $n, m > 20$  možno využiť asymptotickú normalitu štatistiky  $R$ . V tomto prípade sa platnosť nulovej hypotézy overuje pomocou testovej štatistiky

$$U_0 = \frac{R - E(R)}{\sqrt{D(R)}}, \quad (3.7)$$

ktorá má pri platnosti  $H_0$  asymptotické rozloženie  $N(0, 1)$ , t.j.  $U_0 \approx N(0, 1)$ . Pritom platí

$$E(R) = \frac{2nm}{n+m} + 1 \quad (3.8)$$

$$D(R) = \frac{2nm(2nm - n - m)}{(n+m)^2(n+m-1)} \quad (3.9)$$

Nulovú hypotézu zamietame na asymptotickej hladine významnosti  $\alpha$ , ak  $|U_0| \geq u_{1-\alpha/2}$ .

**Príklad.** Pre nasledujúce náhodné výbery z dvoch populácií použijeme 5% hladinu významnosti a testujeme hypotézu  $H_0 : x_{0,50} = y_{0,50}$  proti  $H_1 : x_{0,50} \neq y_{0,50}$ .

Výber 1	40	34	53	28	41		
Výber 2	29	31	52	29	20	31	26

*Riešenie.*

Všetky hodnoty si vzostupne usporiadame podľa veľkosti, ku každej určíme výber, z ktorého pochádza a zistíme počet iterácií. Prehľadnejšie je zapísať všetko do tabuľky:

Usp.hodn.	20	26	28	29	29	31	31	34	40	41	52	53
Č.výb.	2	2	1	2	2	2	2	1	1	1	2	1
Č.iter.	1		2	3			4		5	6		

Počet iterácií:  $R = 6$ ,  $n = 5$ ,  $m = 7$ ,  $\alpha = 0,05$ .

Vzhľadom na to, že nebolo možné zo žiadneho zdroja získať kritické hodnoty pre tento test, ďalej pokračovať vo výpočte nemôžem, výsledok testu je preto určený iba pomocou systému STATISTICA.

#### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 16 prípadoch, do prvej premennej načítame

v ľubovoľnom poradí prvky oboch výberov, druhá premenná obsahuje hodnoty 1 alebo 2 v závislosti od príslušnosti daného prvku k prvému alebo druhému náhodnému výberu. K testu sa dostaneme cez *Statistics - Nonparametrics - Comparing two independent samples (groups)*, zadáme premenné, ako závislú zvolíme premennú 1 s nameranými hodnotami, ako grupujúcu premennú 2 s indexami oboch výberov. Test vykonáme pomocou *Wald-Wolfowitz runs test*.

Wald-Wolfowitz Runs Test (waldwolftest)										
By variable Var2										
Marked tests are significant at p <,05000										
Variable	Valid N	Valid N	Mean	Mean	Z	p-level	Z adjstd	p-level	No. of	No. of
	Group 1	Group 2	Group 1	Group 2					Runs	ties
Var1	5	7	39,20000	31,14286	-0,520515	0,602705	0,208206	0,835068	6	0

Tabuľka obsahuje rozsahy a priemery oboch výberov, hodnotu asymptotickej testovej štatistiky  $U_0$  (označenú ako Z) a príslušnú p-hodnotu, hodnotu asymptotickej testovej štatistiky s opravou na spojitosť (Z adjstd) a tiež jej príslušnú p-hodnotu, ako posledné dve hodnoty sú uvedené počet iterácií a počet zhodných pozorovaní. Vidíme, že p-hodnota je väčšia ako 0,05, preto nulovú hypotézu o zhode mediánov nezamietame.

### 3.2.3 Dvojvýberový Kolmogorov-Smirnovov test

Nech  $X_1, \dots, X_n$  a  $Y_1, \dots, Y_m$  sú 2 nezávislé náhodné výbery z dvoch spojitých rozlození, ktorých distribučné funkcie sa môžu líšiť nielen posunutím, ale aj tvarom. Testujeme hypotézu, že tieto distribučné funkcie sú zhodné, tzn. všetky náhodné veličiny pochádzajú z rovnakého rozloženia, proti alternatíve, že sú rozdielne.

#### Návod.

Nech  $F_1(x) = \frac{1}{n} \text{card} \{i; X_i \leq x\}$  je výberová distribučná funkcia prvého náhodného výberu a  $F_2(x) = \frac{1}{m} \text{card} \{i; Y_i \leq y\}$  je výberová distribučná funkcia druhého náhodného výberu. Testovou štatistikou je pri tomto teste

$$D = \max_{-\infty < x < \infty} |F_1(x) - F_2(x)|. \tag{3.10}$$

a kritickou hodnotou  $D_{n,m}(\alpha)$ , je udávaná v tabuľkách. Nulovú hypotézu zamietame na hladine významnosti  $\alpha$  v prípade, že  $D \geq D_{n,m}(\alpha)$ .

Pre veľké rozsahy  $n, m$  je možné kritickú hodnotu aproximovať podľa vzorca

$$\sqrt{\frac{n+m}{2nm} \ln \frac{2}{\alpha}} \tag{3.11}$$

**Príklad.** Testujme na 5% hladine významnosti hypotézu, že predpovede rastu životných nákladov (v %) v budúcom roku oproti tomuto roku univerzitných ekonómov  $X_2$  sú rovnaké ako predpovede  $X_1$  vládných ekonómov.

$X_1$	4,4	5,8	3,9	6,7	1,3	10,5	4,9
$X_2$	3,1	4,8	2,3	5,6	0,0	2,9	

*Riešenie.*

Hodnoty výberovej distribučnej funkcie 1. výberu:

1.  $x < 1,3 : F_1(x) = 0$
2.  $1,3 \leq x < 3,9 : F_1(x) = \frac{1}{7}$
3.  $3,9 \leq x < 4,4 : F_1(x) = \frac{2}{7}$
4.  $4,4 \leq x < 4,9 : F_1(x) = \frac{3}{7}$
5.  $4,9 \leq x < 5,8 : F_1(x) = \frac{4}{7}$
6.  $5,8 \leq x < 6,7 : F_1(x) = \frac{5}{7}$
7.  $6,7 \leq x < 10,5 : F_1(x) = \frac{6}{7}$
8.  $x > 10,5 : F_1(x) = 1$

Hodnoty výberovej distribučnej funkcie 2. výberu:

1.  $x < 0,0 : F_2(x) = 0$
2.  $0,0 \leq x < 2,3 : F_2(x) = \frac{1}{6}$
3.  $2,3 \leq x < 2,9 : F_2(x) = \frac{2}{6}$
4.  $2,9 \leq x < 3,1 : F_2(x) = \frac{3}{6}$
5.  $3,1 \leq x < 4,8 : F_2(x) = \frac{4}{6}$
6.  $4,8 \leq x < 5,6 : F_2(x) = \frac{5}{6}$
7.  $x > 5,6 : F_2(x) = 1$

Hodnota testovej štatistiky  $D = \max_{-\infty < x < \infty} |F_1(x) - F_2(x)| = \left| \frac{1}{7} - \frac{4}{6} \right| = 0,52381$ , kritická hodnota z tabuliek pre  $n = 7$ ,  $m = 6$  a  $\alpha = 0,05$  je  $D_{7,6}(0,05) = 0,714$ ; keďže  $D < D_{n,m}(\alpha)$ , nulovú hypotézu o rovnosti predpovedí vládnych a univerzitných ekonómov nezamietame na hladine významnosti 0,05.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 13 prípadoch, v prvej premennej sú načítané všetky predpovede, druhá premenná nadobúda hodnotu 1 pre predpoveď vládnych ekonómov, resp. 2 pre predpoveď univerzitných ekonómov. Testujeme cez *Statistics - Nonparametrics - Comparing two independent samples (groups)*, zadáme premenné, ako závislú zvolíme premennú 1 s hodnotami predpovedí, ako grupujúcu premennú 2 s indexami oboch výberov. Test vykonáme pomocou *Kolmogorov-Smirnov two-sample test*.

Kolmogorov-Smirnov Test (dvojkolmsmirntest)									
By variable Var2									
Marked tests are significant at p <,05000									
variable	Max Neg Differnc	Max Pos Differnc	p-level	Mean Group 1	Mean Group 2	Std.Dev. Group 1	Std.Dev. Group 2	Valid N Group 1	Valid N Group 2
Var1	0,00	0,523810	p > .10	5,357143	3,116667	2,833053	1,971209	7	6

V tabuľke máme možnosť vidieť maximálny záporný a kladný rozdiel hodnôt oboch výberových distribučných funkcií, výberové priemery, smerodatné odchýlky, rozsahy a hlavne p-hodnotu, ktorá je v tomto prípade  $> 0,05$ , nulovú hypotézu nezamietame na hladine významnosti  $\alpha = 0,05$ .

# Kapitola 4

## Testy homogenity rozptylov viacerých nezávislých náhodných výberov

Predpokladajme, že faktor  $A$ , t.j. náhodná veličina nominálneho charakteru, má  $r \geq 3$  úrovní a  $i$ -tej úrovni zodpovedá  $n_i$  výsledkov  $X_{i1}, \dots, X_{in_i}$  tvoriacich náhodný výber z rozloženia  $N(\mu_i, \sigma^2)$ ,  $i = 1, \dots, r$ , pričom tieto náhodné výbery sú stochasticky nezávislé, teda  $X_{ij} = \mu_i + \varepsilon_{ij}$ , kde  $\varepsilon_{ij}$  sú stochasticky nezávislé náhodné veličiny s rozložením  $N(0, \sigma^2)$ , kde  $i = 1, \dots, r$  a  $j = 1, \dots, n_i$ .

### 4.1 Levenov test

Položme  $Z_{ij} = |X_{ij} - M_{i\cdot}|$ , kde  $M_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$  a označme

$$\begin{aligned} M_{Zi\cdot} &= \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij} & M_{Z\cdot\cdot} &= \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} Z_{ij} \\ S_{ZE} &= \sum_{i=1}^r \sum_{j=1}^{n_i} (Z_{ij} - M_{Zi\cdot})^2 & S_{ZA} &= \sum_{i=1}^r n_i (M_{Zi\cdot} - M_{Z\cdot\cdot})^2 \end{aligned}$$

Ak platí nulová hypotéza o zhode rozptylov, potom

$$F_Z = \frac{\frac{S_{ZA}}{(r-1)}}{\frac{S_{ZE}}{(n-r)}} \sim F(r-1, n-r) \quad (4.1)$$

$H_0$  potom zamietame na hladine významnosti  $\alpha$  v prípade, že  $F_Z \geq F_{1-\alpha}(r-1, n-r)$ . Jednotlivé výsledky môžeme zosumarizovať v modifikovanej ANOVA tabuľke:

Zdroj variability	Súčet štvorcov	Stupne voľnosti	Podiel	$F_Z$
skupiny	$S_{ZA}$	$f_{ZA} = r - 1$	$\frac{S_{ZA}}{f_{ZA}}$	$\frac{S_{ZA}/f_{ZA}}{S_{ZE}/f_{ZE}}$
reziduálny	$S_{ZE}$	$f_{ZE} = n - r$	$\frac{S_{ZE}}{f_{ZE}}$	—
celkový	$S_{ZT}$	$f_{ZT} = n - 1$	—	—

**Príklad.** Na 50 pozemkoch rozmiestnených v skúmanej poľnohospodárskej oblasti boli vykonané pokusy so štyrmi druhmi olejnatých rastlín. Výsledky, udávajúce množstvo získaného oleja v tonách na hektár, sú uvedené podľa jednotlivých druhov rastlín.

Horčica:

0,188 0,067 0,232 0,124 0,285 0,300 0,387 0,184 0,155 0,031

Repka olejka:

0,415 0,291 0,113 0,114 0,062 0,270 0,068 0,196 0,308 0,365 0,230 0,262 0,050 0,127 0,078

Ľaničník maloplodý (Ľnička maloplodá):

0,382 0,199 0,473 0,262 0,152 0,293 0,428 0,241 0,390 0,195

Sója:

0,227 0,357 0,402 0,267 0,017 0,240 0,167 0,321 0,179 0,086 0,020 0,280 0,384 0,214 0,168

Máme overiť na 5% hladine významnosti hypotézu o rovnosti rozptylov množstva získaného oleja u týchto 4 druhov rastlín.

*Riešenie.*

Výberové priemery:  $m_1 = 0,1953$ ,  $m_2 = 0,1966$ ,  $m_3 = 0,3015$ ,  $m_4 = 0,221933$ . Tabuľku hodnôt  $Z_{ij}$  neuvádzam, podľa vzorca sa dá ľahko zostaviť z vyššie uvedených nameraných hodnôt. Vypočítame:  $m_{Z1.} = 0,08456$ ,  $m_{Z2.} = 0,101973$ ,  $m_{Z3.} = 0,0934$ ,  $m_{Z4.} = 0,0937$ ,  $m_{Z..} = 0,0943$ ,  $S_{ZA} = 0,001846$ ,  $S_{ZE} = 0,16895$ . Hodnota testovej štatistiky  $F_Z = 0,1675$  a kvantil  $F_{0,95}(3, 46) = 2,8068$ . Pretože  $0,167528 < 2,8068$ , nulovú hypotézu o zhode rozptylov množstva získaného oleja u 4 druhov rastlín nezamietame na hladine významnosti 0,05.

### Postup v programe STATISTICA.

Vytvoríme dátový súbor o 2 premenných a 50 prípadoch. Prvá premenná obsahuje všetky namerané hodnoty, druhá indexy 1-4 podľa príslušnosti výsledku k danému druhu rastliny. Postupujeme cez *Statistics - Basic Statistics/Tables - Breakdown & one-way ANOVA*, v záložke *Lists of tables* zvolíme premennú 1 ako závislú, premennú 2 ako grupujúcu a zaškrtneme *Levene test*. Vznikne nasledovná tabuľka:

Levene Test of Homogeneity of Variances (levene)								
Marked effects are significant at p < ,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Var1	0,001846	3	0,000615	0,168949	46	0,003673	0,167528	0,917739

V nej vidíme hodnotu skupinového ( $S_{ZA} = 0,001846$ ) a reziduálneho ( $S_{ZE} = 0,168949$ ) súčtu štvorcov, počet stupňov voľnosti  $f_{ZA} = 3$  a  $f_{ZE} = 46$ , testová štatistika nadobúda hodnotu  $F_Z = 0,167528$  a zodpovedajúca p-hodnota = 0,917739. Keďže p-hodnota je väčšia ako hladina významnosti 0,05, hypotézu o rovnosti rozptylov nemôžeme na tejto hladine zamietnuť.

*Poznámka.* Nevýhodou Levenovho testu je, že jednotlivé náhodné výbery môžu byť ovplyvnené odchýlkami od normality. Tento problém rieši nasledujúci test.

## 4.2 Brown-Forsythov test

Položme  $Z'_{ij} = |X_{ij} - \widetilde{M}_i|$ , kde  $\widetilde{M}_i$  je medián  $i$ -teho výberu a označme

$$M_{Z'i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z'_{ij} \qquad M_{Z'..} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} Z'_{ij}$$

$$S_{Z'E} = \sum_{i=1}^r \sum_{j=1}^{n_i} (Z'_{ij} - M_{Z'i.})^2 \qquad S_{Z'A} = \sum_{i=1}^r n_i (M_{Z'i.} - M_{Z'..})^2$$

Ak platí nulová hypotéza o zhode rozptylov, potom

$$F_{Z'} = \frac{\frac{S_{Z'A}}{(r-1)}}{\frac{S_{Z'E}}{(n-r)}} \sim F(r-1, n-r) \qquad (4.2)$$

$H_0$  zamietame na hladine významnosti  $\alpha$ , ak  $F_{Z'} \geq F_{1-\alpha}(r-1, n-r)$ . Jednotlivé výsledky môžeme zosumarizovať v modifikovanej ANOVA tabuľke:

Zdroj variability	Súčet štvorcov	Stupne voľnosti	podiel	$F_{Z'}$
skupiny	$S_{Z'A}$	$f_{Z'A} = r - 1$	$\frac{S_{Z'A}}{f_{Z'A}}$	$\frac{S_{Z'A}/f_{Z'A}}{S_{Z'E}/f_{Z'E}}$
reziduálny	$S_{Z'E}$	$f_{Z'E} = n - r$	$\frac{S_{Z'E}}{f_{Z'E}}$	—
celkový	$S_{Z'T}$	$f_{Z'T} = n - 1$	—	—

**Príklad.** Prebehol experiment, aby sa zistilo, či 4 špecifické teploty ohňa ovplyvnia hustotu určitého typu tehly. Experiment viedol k týmto výsledkom:

Teplota (°C)	Hustota						
37,78	21,8	21,9	21,7	21,6	21,7	21,5	21,8
51,67	21,7	21,4	21,5	21,5	-	-	-
65,56	21,9	21,8	21,8	21,6	21,5	-	-
79,44	21,9	21,7	21,8	21,9	21,6	21,8	-

Máme overiť zhodu rozptylov meraní pri 4 rôznych teplotách.

*Riešenie.*

Mediány jednotlivých výberov:  $\widetilde{m}_1 = 0,1953$ ,  $\widetilde{m}_2 = 0,1966$ ,  $\widetilde{m}_3 = 0,3015$ ,  $\widetilde{m}_4 = 0,221933$ . Opäť tabuľku hodnôt  $Z'_{ij}$  neuvádzam, podľa vzorca sa dá ľahko zostaviť. Ďalej výpočtom získame:  $m_{Z'1.} = 0,1$ ,  $m_{Z'2.} = 0,075$ ,  $m_{Z'3.} = 0,12$ ,  $m_{Z'4.} = 0,0833$ ,  $m_{Z'..} = 0,09545$ ,  $S_{Z'A} = 0,005712$ ,  $S_{Z'E} = 0,16383$ . Hodnota testovej štatistiky je  $F_{Z'} = 0,209193$  a kvantil Fisher-Snedecorovho rozloženia  $F_{0,95}(3, 18) = 3,1599$ . Pretože  $0,209193 < 3,1599$ , nulovú hypotézu o zhode rozptylov meraní pri 4 rôznych teplotách nezamietame na hladine významnosti 0,05.

**Postup v programe STATISTICA.**

Vytvoríme dátový súbor o 2 premenných a 22 prípadoch. Prvá premenná obsahuje všetky namerané hodnoty, druhá indexy 1-4 podľa príslušnosti výsledku k danej teplote. Postupujeme cez *Statistics - Basic Statistics/Tables - Breakdown & one-way ANOVA*, v záložke *Lists of tables* zvolíme premennú 1 ako závislú, premennú 2 ako grupujúcu a zaškrtneme *Brown & Forsythe*. Vznikne nasledujúca tabuľka:

Brown-Forsythe Test of Homog. of Variances (brownforysytetest)								
Marked effects are significant at p < ,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Var1	0,005712	3	0,001904	0,163833	18	0,009102	0,209193	0,888702

Vyčítame hodnotu skupinového ( $S_{Z'A} = 0,005712$ ) a reziduálneho ( $S_{Z'E} = 0,163833$ ) súčtu štvorcov, počet stupňov voľnosti  $f_{Z'A} = 3$  a  $f_{Z'E} = 18$ , testová štatistika nadobúda hodnotu  $F_{Z'} = 0,209193$  a zodpovedajúca p-hodnota = 0,888702. Keďže p-hodnota je väčšia ako hladina významnosti 0,05, hypotézu o rovnosti rozptylov nemôžeme na tejto hladine zamietnuť.

*Poznámka.* Ak by sme Brown-Forsythovým testom riešili príklad z časti 4.1, p-hodnota by nám vyšla trochu nižšia, z čoho môžeme usudzovať, že B-F test je o čosi silnejší ako Levenov, lebo berie do úvahy aj odchýlky náhodného výberu od normality.



# Záver

Testovanie hypotéz má a vždy aj bude mať veľké využitie v praxi. Stále sa vyskytujú otázky a problémy v ktorejkoľvek oblasti, ktoré nám práve táto časť štatistiky pomôže vyriešiť.

Cieľom tejto práce bolo popísať najdôležitejšie testy, ukázať ich riešenie v systéme STATISTICA, aby čitateľ nemusel všetko počítať ručne, čo v prípade niektorých testov veľmi ocení. V prvej kapitole som sa sústredil na testy s jedným jednorozmerným náhodným výberom, teda aj na jednovýberový t-test, jeden z najpoužívanejších v praxi. Druhá kapitola bola zameraná na párové testy, kde existoval vzťah medzi náhodnými výbermi v tom zmysle, že každá náhodná veličina bola testovaná v 2 rôznych podmienkach, a teda v konečnom dôsledku boli tieto 2 výsledky na sebe závislé. Tretia kapitola sa venovala testom dvoch nezávislých náhodných výberov. Vo štvrtej som uviedol 2 testy viacerých nezávislých náhodných výberov, keďže nemusíme mať vždy výber len jeden alebo dva. Verím aj, že vytvorené makrá prispedia k rýchlejšiemu vyriešeniu príkladov na testy, ktoré zatiaľ v tomto systéme implementované nie sú.

# Použitá literatura

- [1] ANDĚL, J. *Statistické metody*. 1. vyd. Praha: Matfyzpress, 1993.
- [2] ARTLOVÁ, M.; BÍLKOVÁ, D.; JAROŠOVÁ, E.; POUROVÁ, Z. *Sbírka příkladů ze statistiky (Statistika A)*. 1. dotlač. Praha: Vysoká škola ekonomická, 1996. ISBN 80-7079-727-4.
- [3] BLATNÁ, D. *Statistika a pravděpodobnost*. 1. vyd. Praha: Bankovní institut vysoká škola, 2003. ISBN 80-7265-059-0.
- [4] BUDÍKOVÁ M. *Interné materiály k predmetu Základní statistické metody*.
- [5] BUDÍKOVÁ, M.; MIKOLÁŠ, Š.; OSECKÝ, P. *Teorie pravděpodobnosti a matematická statistika - sbírka příkladů*. 3. vyd. Brno: Masarykova univerzita, 2004. ISBN 80-210-3313-4.
- [6] BUDÍKOVÁ M.; LERCH T.; MIKOLÁŠ Š. *Základní statistické metody*. 1. vyd. Brno: Masarykova univerzita, 2005. ISBN 80-210-3886-1.
- [7] FORBELSKÁ M. *Interné materiály k predmetu Lineární statistické modely I: M5120 - 9. cvičení: Normální náhodné výběry*.
- [8] HEBÁK P.; BÍLKOVÁ D.; SVOBODOVÁ A. *Praktikum k výuce matematické statistiky: Testování hypotéz*. 1. vyd. Praha: Vysoká škola ekonomická, 2000. ISBN 80-245-0082-5.
- [9] KOŽÍŠEK J. *Statistická analýza: Příklady*. 3., prepracované vyd. Praha: Vydavatelství ČVUT, 1997. ISBN 80-01-01617-X.
- [10] MONTGOMERY D.C.; RUNGER G.C. *Applied statistics and probability for engineers*. 2nd ed. New York: John Wiley & Sons, 1999. ISBN 0471170275.
- [11] Statsoft, Inc. *STATISTICA Electronic Manual (Uživatelská příručka systému STATISTICA)*. 2007.
- [12] SVATOŠOVÁ L.; KÁBA B. *Statistické metody I*. 1. vyd. Praha: Česká zemědělská univerzita, 2007. ISBN 978-80-213-1672-0.

- 
- [13] UPTON G.; COOK I. *A Dictionary of Statistics*. 2nd ed. Oxford University Press, 2004. ISBN 0198614314.

# Prílohy

## Test o rozptyle - zdrojový kód

```
Option Base 1
```

```
Sub Main
```

```
Dim alpha As Double 'hladina vyznamnosti
Dim mean As Double 'priemer
Dim stdev As Double 'vyberova smerodatna odchylka
Dim C As Double 'ocakavana hodnota rozptylu
Dim alt As Double 'typ alternativnej hypotezy
Dim Matrix() As Double 'matica nameranych hodnot
Dim i As Double 'forcyklovske premenne
Dim pom1 As Double 'pomocna premenna vo for cykle vypoctu vyberoveho
rozptylu
Dim VarList () As Long 'zoznam vybranych premennych vyberu
Dim pmen As Double 'P(T0 <= t0)
Dim pvac As Double 'P(T0 >= t0)
Dim K~As Double 'hodnota testovej statistiky

Const SelOneVar As String = "Zvoľte jednu premennú"
Const SelVar As String = "Výber premennej"
Const SelTypeOfAltHyp As String = "Zvoľte typ alternatívnej hypotézy"
Const TypeOfAltHyp As String = "Obojstranná alternatíva|Ľavostranná
alternatíva|Pravostranná alternatíva"
Const SprHead As String = "Test of Variance Against Reference Constant"
Const SprC1 As String = "Summary"
Const SprV1 As String = "Mean"
Const SprV2 As String = "St. Dev."
Const SprV3 As String = "K"
Const SprV4 As String = "Chi_alpha/2 (n-1)"
Const SprV5 As String = "Chi_1-alpha/2 (n-1)"
Const SprV6 As String = "Chi_alpha (n-1)"
```

```
Const SprV7 As String = "Chi_1-alpha (n-1)"
Const SprV8 As String = "p-value"

numvar=ActiveSheet.NumberOfVariables
numcas=ActiveSheet.NumberOfCases

ReDim VarList(1 To numvar)
ReDim Preserve Matrix(numcas,numvar) As Double
Matrix = ActiveSpreadsheet.Data

'Volba premennej
If 0=SelectVariables1(ActiveDataSet,SelVar,1,1,VarList,Count,SelOneVar)
Then
End
End If

LevOfSign=InputBox("Zadajte hladinu významnosti, na ktorej bude testovaná
hypotéza:","Hladina významnosti")
alpha=Cdbl(LevOfSign) 'prevod stringu na Double

Konst=InputBox("Zadajte očakávanú hodnotu rozptylu:","Zadanie rozptylu")
C = Cdbl(Konst)

'Vyber alternativy
alt = DisplayListBox(SelTypeOfAltHyp,TypeOfAltHyp,1)

'do stl priradime index premennej, ktoru sme si vybrali
stl = VarList(1)

mean = 0
For i = 1 To numcas
mean = mean + Matrix (i,stl)
Next i
mean = mean / numcas

pom1 = 0
For i = 1 To numcas
pom1 = pom1 + (Matrix (i,stl) - mean)^2
Next i
stdev = Sqrt(1/(numcas-1) * pom1)

K~= (numcas - 1) * stdev^2 / C
```

```
If alt = 0 Then End
If alt = 1 Then
ChiAlphaPol = VChi2(alpha/2,numcas-1)
Chi_1 = ChiAlphaPol
Chi1MinusAlphaPol = VChi2(1-alpha/2,numcas-1)
Chi_2 = Chi1MinusAlphaPol

pmen = IChi2(K,numcas-1) 'vypocet p-hodnoty
pvac = 1-IChi2(K,numcas-1)
If (pmen <= pvac) Then
pv = 2 * pmen
Else
pv = 2 * pvac
End If

ReDim Preserve A(6) As Double
A(1) = mean
A(2) = stdev
A(3) =
K~A(4) = Chi_1
A(5) = Chi_2
A(6) = pv
'Vysledna tabulka
Set Summary = Spreadsheets.New
Summary.SetSize(1,6)
Summary.Header=SprHead
Summary.CaseName(1) = SprC1
Summary.AutoFitCase
Summary.VariableName(1) = SprV1
Summary.VariableName(2) = SprV2
Summary.VariableName(3) = SprV3
Summary.VariableName(4) = SprV4
Summary.VariableName(5) = SprV5
Summary.VariableName(6) = SprV8

Summary.Variable(1).ColumnWidth=0.6
Summary.Value(1,1) = A(1)
Summary.Variable(2).ColumnWidth=0.7
Summary.Value(1,2) = A(2)
Summary.Variable(3).ColumnWidth=0.6
Summary.Value(1,3) = A(3)
Summary.Variable(4).ColumnWidth=1.2
Summary.Value(1,4) = A(4)
```

```
Summary.Variable(5).ColumnWidth=1.3
Summary.Value(1,5) = A(5)
Summary.Variable(6).ColumnWidth=1
Summary.Value(1,6) = A(6)
```

'Ak je p-hodnota mensia alebo rovna ako zadana hladina vyznamnosti,  
zamietame nulovu hypotezu, vsetko sa zobrazi na cerveno.

```
If (pv <= alpha) Then
For i = 1 To 6
Summary.Cells(1,i).Font.Color = RGB(255,0,0)
Next i
End If
Summary.Visible=True
End If
```

```
If alt = 2 Then
Chialpha = VChi2(alpha,numcas-1)
pv = IChi2(K,numcas-1) 'p-hodnota
ReDim Preserve A(5) As Double
A(1) = mean
A(2) = stdev
A(3) =
K~A(4) = Chialpha
A(5) = pv
'Vysledna tabulka
Set Summary = Spreadsheets.New
Summary.SetSize(1,5)
Summary.Header=SprHead
Summary.CaseName(1) = SprC1
Summary.AutoFitCase
Summary.VariableName(1) = SprV1
Summary.VariableName(2) = SprV2
Summary.VariableName(3) = SprV3
Summary.VariableName(4) = SprV6
Summary.VariableName(5) = SprV8
```

```
Summary.Variable(1).ColumnWidth=0.6
Summary.Value(1,1) = A(1)
Summary.Variable(2).ColumnWidth=0.7
Summary.Value(1,2) = A(2)
Summary.Variable(3).ColumnWidth=0.6
Summary.Value(1,3) = A(3)
Summary.Variable(4).ColumnWidth=1.2
```

```
Summary.Value(1,4) = A(4)
Summary.Variable(5).ColumnWidth=1
Summary.Value(1,5) = A(5)

'Ak je p-hodnota mensia alebo rovna ako alpha, vsetko sa zobrazi na
cerveno (zamietame H0).
If (pv <= alpha) Then
For i = 1 To 5
Summary.Cells(1,i).Font.Color = RGB(255,0,0)
Next i
End If
Summary.Visible=True
End If

If alt = 3 Then
Chi1Minusalpha = VChi2(1-alpha,numcas-1)
pv = 1 - IChi2(K,numcas-1) 'p-hodnota
ReDim Preserve A(5) As Double
A(1) = mean
A(2) = stdev
A(3) =
K^A(4) = Chi1Minusalpha
A(5) = pv
'Vysledna tabulka
Set Summary = Spreadsheets.New
Summary.SetSize(1,5)
Summary.Header=SprHead
Summary.CaseName(1) = SprC1
Summary.AutoFitCase
Summary.VariableName(1) = SprV1
Summary.VariableName(2) = SprV2
Summary.VariableName(3) = SprV3
Summary.VariableName(4) = SprV7
Summary.VariableName(5) = SprV8

Summary.Variable(1).ColumnWidth=0.6
Summary.Value(1,1) = A(1)
Summary.Variable(2).ColumnWidth=0.7
Summary.Value(1,2) = A(2)
Summary.Variable(3).ColumnWidth=0.6
Summary.Value(1,3) = A(3)
Summary.Variable(4).ColumnWidth=1.2
Summary.Value(1,4) = A(4)
```



```
Summary.Variable(5).ColumnWidth=1
Summary.Value(1,5) = A(5)
```

```
'Ak je p-hodnota mensia alebo ako alpha, vsetko sa zobrazi na cerveno
(zamietame H0).
```

```
If (pv <= alpha) Then
For i = 1 To 5
Summary.Cells(1,i).Font.Color = RGB(255,0,0)
Next i
End If
Summary.Visible=True
End If
```

```
End Sub
```

## Morgan-Pitmanov test - zdrojový kód

```
Option Base 1
```

```
Sub Main
```

```
Dim alpha As Double 'hladina vyznamnosti
Dim mean1 As Double 'priemer Xn
Dim mean2 As Double 'priemer Yn
Dim stdev1 As Double 'vyberova smerodatna odchylka Xn
Dim stdev2 As Double 'vyberova smerodatna odchylka Yn
Dim alt As Double
Dim Matrix() As Double 'matica nameranych hodnot
Dim i As Integer 'forcyklovske premenne
Dim pom1 As Double 'pomocna premenna vo for cykle vypoctu vyberovej
sm.odch. Xn
Dim pom2 As Double 'pomocna premenna vo for cykle vypoctu vyberovej
sm.odch. Yn
Dim pom3 As Double 'pomocna premenna vo for cykle pomocneho vyrazu pre
korel.koef.
Dim pom4 As Double ' ---||---
Dim pom5 As Double ' ---||---
Dim t1MinusAlphaPol As Double 'kvantil Studentovho rozlozenia
t_1-alpha/2 (n-2)
Dim t_1 As Double ' ---||---
Dim t1MinusAlpha As Double 'kvantil Studentovho rozlozenia t_1-alpha (n-2)
```

```
Dim t_2 As Double '---||---
Dim pmen As Double 'P(TO <= t0)
Dim pvac As Double 'P(TO >= t0)
Dim VarList1 () As Long 'zoznam vybranych premennych prveho vyberu
Dim VarList2 () As Long 'zoznam vybranych premennych druhého vyberu

Const SelTwoVar1 As String = "Zvoľte premennú"
Const SelTwoVar2 As String = "Zvoľte závislú premennú"
Const SelVar As String = "Výber premenných"
Const SelTypeOfAltHyp As String = "Zvoľte typ alternatívnej hypotézy"
Const TypeOfAltHyp As String = "Obojstranná alternatíva|Lavostranná
alternatíva|Pravostranná alternatíva"
Const SprHead As String = "Morgan-Pitman Test of Homogeneity of Variances"
Const SprC1 As String = "Summary"
Const SprV1 As String = "Mean1"
Const SprV2 As String = "Mean2"
Const SprV3 As String = "St. Dev.1"
Const SprV4 As String = "St. Dev.2"
Const SprV5 As String = "Coef. of corr."
Const SprV6 As String = "T"
Const SprV7 As String = "t_1-alpha/2 (n-2)"
Const SprV8 As String = "-t_1-alpha (n-2)"
Const SprV9 As String = "t_1-alpha (n-2)"
Const SprV10 As String = "p-value"

numvar=ActiveSheet.NumberOfVariables
numcas=ActiveSheet.NumberOfCases

ReDim VarList1(1 To numvar)
ReDim VarList2(1 To numvar)
ReDim Preserve Matrix(numcas,numvar) As Double
Matrix = ActiveSpreadsheet.Data

'Volba premennych
If 0=SelectVariables2(ActiveDataSet,SelVar,1,1,VarList1,Count1,SelTwoVar1,
1,1,VarList2,Count2,SelTwoVar2) Then
End
End If

LevOfSign=InputBox("Zadajte hladinu významnosti, na ktorej bude testovaná
hypotéza:","Hladina významnosti")
alpha=Cdbl(LevOfSign) 'prevod stringu na Double
```

```
'Vyber alternativy
alt = DisplayListBox(SelTypeOfAltHyp,TypeOfAltHyp,1)

'do stl1 priradime index premennej, ktoru sme si vybrali z~prveho vyberu
stl1 = VarList1(1)

'do stl2 priradime index premennej, ktoru sme si vybrali z~druheho vyberu
stl2 = VarList2(1)

mean1 = 0
For i = 1 To numcas
mean1 = mean1 + Matrix (i,stl1)
Next i
mean1 = mean1 / numcas

mean2 = 0
For i = 1 To numcas
mean2 = mean2 + Matrix (i,stl2)
Next i
mean2 = mean2 / numcas

'Vypocet vyb.sm.odch.Xn
pom1 = 0
For i = 1 To numcas
pom1 = pom1 + (Matrix (i,stl1) - mean1)^2
Next i
stdev1 = Sqrt(1/(numcas-1) * pom1)

'Vypocet vyb.sm.odch.Yn
pom2 = 0
For i = 1 To numcas
pom2 = pom2 + (Matrix (i,stl2) - mean2)^2
Next i
stdev2 = Sqrt(1/(numcas-1) * pom2)

pom3 = 0
For i = 1 To numcas
pom3 = pom3 + (Matrix (i,stl1) * Matrix (i,stl2))
Next i

pom4 = 0
For i = 1 To numcas
pom4 = pom4 + (Matrix (i,stl1))^2
```

```
Next i

pom5 = 0
For i = 1 To numcas
pom5 = pom5 + (Matrix (i,st12))^2
Next i

If (Sqrt((pom4 - numcas*mean1^2)*(pom5 - numcas*mean2^2))) = 0 Then
MsgBox("Výberový korelačný koeficient nemožno spočítat', menovateľ je
rovný 0", "Chyba")
End

Else
r = (pom3 - numcas*mean1*mean2) / Sqrt((pom4 - numcas*mean1^2)*(pom5 -
numcas*mean2^2))
End If

T_0 = (stdev1^2-stdev2^2) / (2*stdev1*stdev2) * Sqrt((numcas-2)/(1-r^2))

If alt = 0 Then End
If alt = 1 Then
t1MinusAlphaPol = VStudent(1-alpha/2,numcas-2)
t_1 = t1MinusAlphaPol

pmen = IStudent(T_0,numcas-2) 'vypocet p-hodnoty
pvac = 1-IStudent(T_0,numcas-2)
If (pmen <= pvac) Then
pv = 2 * pmen
Else
pv = 2 * pvac
End If

ReDim Preserve A(8) As Double
A(1) = mean1
A(2) = mean2
A(3) = stdev1
A(4) = stdev2
A(5) = r
A(6) = T_0
A(7) = t_1
A(8) = pv
'Vysledna tabulka
Set Summary = Spreadsheets.New
```

```
Summary.SetSize(1,8)
Summary.Header=SprHead
Summary.CaseName(1) = SprC1
Summary.AutoFitCase
Summary.VariableName(1) = SprV1
Summary.VariableName(2) = SprV2
Summary.VariableName(3) = SprV3
Summary.VariableName(4) = SprV4
Summary.VariableName(5) = SprV5
Summary.VariableName(6) = SprV6
Summary.VariableName(7) = SprV7
Summary.VariableName(8) = SprV10
```

```
Summary.Variable(1).ColumnWidth=0.6
Summary.Value(1,1) = A(1)
Summary.Variable(2).ColumnWidth=0.6
Summary.Value(1,2) = A(2)
Summary.Variable(3).ColumnWidth=0.7
Summary.Value(1,3) = A(3)
Summary.Variable(4).ColumnWidth=0.7
Summary.Value(1,4) = A(4)
Summary.Variable(5).ColumnWidth=0.95
Summary.Value(1,5) = A(5)
Summary.Variable(6).ColumnWidth=0.6
Summary.Value(1,6) = A(6)
Summary.Variable(7).ColumnWidth=1.2
Summary.Value(1,7) = A(7)
Summary.Variable(8).ColumnWidth=0.8
Summary.Value(1,8) = A(8)
```

'Ak je p-hodnota mensia alebo rovna ako zadana hladina vyznamnosti,  
zamietame hypotézu o $\tilde{z}$ hode rozptylov a vsetko sa zobrazi na cerveno.

```
If (pv <= alpha) Then
For i = 1 To 8
Summary.Cells(1,i).Font.Color = RGB(255,0,0)
Next i
End If
Summary.Visible=True
End If
```

```
If alt = 2 Then
t1MinusAlpha = VStudent(1-alpha,numcas-2)
t_2 = t1MinusAlpha
```

```
pv = IStudent(T_0,numcas-2) 'vypocet p-hodnoty
ReDim Preserve A(8) As Double
A(1) = mean1
A(2) = mean2
A(3) = stdev1
A(4) = stdev2
A(5) = r
A(6) = T_0
A(7) = -t_2
A(8) = pv
'Vysledna tabulka
Set Summary = Spreadsheets.New
Summary.SetSize(1,8)
Summary.Header=SprHead
Summary.CaseName(1) = SprC1
Summary.AutoFitCase
Summary.VariableName(1) = SprV1
Summary.VariableName(2) = SprV2
Summary.VariableName(3) = SprV3
Summary.VariableName(4) = SprV4
Summary.VariableName(5) = SprV5
Summary.VariableName(6) = SprV6
Summary.VariableName(7) = SprV8
Summary.VariableName(8) = SprV10

Summary.Variable(1).ColumnWidth=0.6
Summary.Value(1,1) = A(1)
Summary.Variable(2).ColumnWidth=0.6
Summary.Value(1,2) = A(2)
Summary.Variable(3).ColumnWidth=0.7
Summary.Value(1,3) = A(3)
Summary.Variable(4).ColumnWidth=0.7
Summary.Value(1,4) = A(4)
Summary.Variable(5).ColumnWidth=0.95
Summary.Value(1,5) = A(5)
Summary.Variable(6).ColumnWidth=0.6
Summary.Value(1,6) = A(6)
Summary.Variable(7).ColumnWidth=1.2
Summary.Value(1,7) = A(7)
Summary.Variable(8).ColumnWidth=0.8
Summary.Value(1,8) = A(8)
```

'Ak je p-hodnota mensia alebo rovna ako zadana hladina vyznamnosti,

zamietame hypotézu o zhode rozptylov a vsetko sa zobrazí na cerveno.

```
If (pv <= alpha) Then
```

```
For i = 1 To 8
```

```
Summary.Cells(1,i).Font.Color = RGB(255,0,0)
```

```
Next i
```

```
End If
```

```
Summary.Visible=True
```

```
End If
```

```
If alt = 3 Then
```

```
t1MinusAlpha = VStudent(1-alpha,numcas-2)
```

```
t_2 = t1MinusAlpha
```

```
pv = 1-IStudent(T_0,numcas-2) 'vypocet p-hodnoty
```

```
ReDim Preserve A(8) As Double
```

```
A(1) = mean1
```

```
A(2) = mean2
```

```
A(3) = stdev1
```

```
A(4) = stdev2
```

```
A(5) = r
```

```
A(6) = T_0
```

```
A(7) = t_2
```

```
A(8) = pv
```

```
'Vysledna tabulka
```

```
Set Summary = Spreadsheets.New
```

```
Summary.SetSize(1,8)
```

```
Summary.Header=SprHead
```

```
Summary.CaseName(1) = SprC1
```

```
Summary.AutoFitCase
```

```
Summary.VariableName(1) = SprV1
```

```
Summary.VariableName(2) = SprV2
```

```
Summary.VariableName(3) = SprV3
```

```
Summary.VariableName(4) = SprV4
```

```
Summary.VariableName(5) = SprV5
```

```
Summary.VariableName(6) = SprV6
```

```
Summary.VariableName(7) = SprV9
```

```
Summary.VariableName(8) = SprV10
```

```
Summary.Variable(1).ColumnWidth=0.6
```

```
Summary.Value(1,1) = A(1)
```

```
Summary.Variable(2).ColumnWidth=0.6
```

```
Summary.Value(1,2) = A(2)
```

```
Summary.Variable(3).ColumnWidth=0.7
```

```
Summary.Value(1,3) = A(3)
```

```
Summary.Variable(4).ColumnWidth=0.7
Summary.Value(1,4) = A(4)
Summary.Variable(5).ColumnWidth=0.95
Summary.Value(1,5) = A(5)
Summary.Variable(6).ColumnWidth=0.6
Summary.Value(1,6) = A(6)
Summary.Variable(7).ColumnWidth=1.2
Summary.Value(1,7) = A(7)
Summary.Variable(8).ColumnWidth=0.8
Summary.Value(1,8) = A(8)
```

'Ak je p-hodnota mensia alebo rovna ako zadana hladina vyznamnosti,  
zamietame hypotézu o $\tilde{z}$ hode rozptylov a vsetko sa zobrazi na cerveno.

```
If (pv <= alpha) Then
For i = 1 To 8
Summary.Cells(1,i).Font.Color = RGB(255,0,0)
Next i
End If
Summary.Visible=True
End If

End Sub
```