Slovenská technická univerzita
v Bratislave
Stavebná fakulta

Študentská vedecká konferencia
Akademický rok 2008/2009

Neparametrické štatistické metódy a ich ekonomické aplikácie

Meno Priezvisko študenta:	Martin Štiglic
Ročník a program/odbor štúdia:	3. ročník/ Matematicko - počítačové modelovanie
Vedúci práce:	Prof. RNDr. Magda Komorníková CSc.
Katedra:	matematiky a deskriptívnej geometrie

Košice 28. máj 2009

Anotácia:

Práca sa zaoberá neparametrickými štatistickými metódami na testovanie hypotéz a výpočet neparametrických korelačných koeficientov. Jedným z cieľov práce je vysvetliť rozdiel medzi parametrickými a neparametrickými štatistickými metódami a popísať vybrané neparametrické metódy. Dôraz je kladený na počítačovú realizáciu týchto metód v systéme Mathematica a ich aplikáciu na príkladoch vstupných dát z ekonomickej oblasti.

Annotation:

The aim of this work is to explain the difference between parametric and nonparametric statistical methods and to describe some nonparametric statistical methods, including their use for testing of hypotheses and calculating of (nonparametric) coefficients of correlation. The emphasis is put on creation of computational procedures for realizations of nonparametric statistical methods (using the system Mathematica) and their applications to various economic input data.

Úvod

Témou práce sú neparametrické štatistické metódy a ich aplikácie v ekonomickej oblasti. Hlavným cieľom práce je pre tieto metódy vytvoriť procedúry v systéme Mathematica. Nadväzujem na minuloročnú prácu Pavla Hlavatého, ktorú dopĺňam o nové neparametrické štatistické metódy.

Teoretická časť bola čerpaná z kníh uvedených v časti Použitá literatúra a praktická časť sa nachádza v prílohách vo forme zdrojových kódov.

Každá metóda je najskôr stručne popísaná (postup výpočtu, testovacia štatistika, oblasť zamietnutia) a neskôr vysvetlená na konkrétnom príklade z ekonomickej oblasti, ktorý je potom tiež vypočítaný pomocou naprogramovanej procedúry v systéme Mathematica.

Na začiatku si vysvetlíme rozdiely medzi parametrickými a neparametrickými metódami, povieme si, podľa akých kritérií zvoliť test. Zoznámime sa so základným princípom neparametrických metód a uvedieme ich výhody a nevýhody. Ďalej si postupne prejdeme vybrané testy. Začneme jednovýberovým znamienkovým testom, ktorý testuje hypotézu o predpokladanom mediáne rozdelenia náhodného výberu. Pri dvojvýberových testoch spomenieme Kolmogorov-Smirnovov test a Wald-Wolfowitzov test. Zameriame sa aj na neparametrickú analýzu rozptylu pre závislé výbery (Friedmanov test) a nezávislé výbery (Kruskall-Walisov test), ktorá testuje hypotézu rovnosti distribučných funkcií. Závislosť (koreláciu) medzi usporiadanými pármi overíme pomocou Spearmanovho a Kendallovho testu.

Parametrické versus neparametrické štatistické metódy

V tejto časti porovnáme parametrické a neparametrické metódy, vysvetlíme rozdiel medzi nimi a podľa akého kritéria vybrať jednotlivú metódu. Nakoniec uvedieme výhody a nevýhody neparametrických metód.

a) **Parametrické** štatistické metódy narábajú s parametrami základných súborov, pričom vychádzajú zo špecifických predpokladov o rozdeleniach pravdepodobnosti základných súborov a výberových štatistík. Pri ich použití sa obyčajne predpokladá, že rozdelenie základného súboru je normálne (Gaussovo). Príkladom parametrickej metódy je Fisherova analýza rozptylu (ANOVA). V prípade porušenia normality možno skúsiť transformáciu (najčastejšie sa používa logaritmovanie) alebo použiť neparametrické metódy, ktoré nepredpokladajú konkrétne rozdelenie premennej. Dôsledkom centrálnej limitnej vety je, že parametrické metódy možno použiť bez ohľadu na rozdelenie premennej v základnom súbore, ak sú vzorky dostatočne veľké ($n > 50$). Veľký pozor si však vždy treba dať na extrémne hodnoty premenných, ktoré často spôsobujú nesprávnosť výsledkov. Vyradenie jediného extrémneho pozorovania môže úplne zmeniť výsledok parametrického testu.

b) **Neparametrické** metódy sa nespoliehajú na odhad parametrov charakterizujúcich rozdelenie premennej v základnom súbore. Preto sa tieto metódy niekedy (a správnejšie) označujú ako metódy s voľnými rozdeleniami. Neparametrické metódy pracujú s početnosťami (napr. Chi-kvadrát test nezávislosti) alebo s poradovými číslami, ktoré boli pridelené pôvodným údajom (napr. Kruskal-Wallisov test).

Prečo sa jednoducho vždy nepoužijú priamo neparametrické metódy? Pri použití malých vzoriek zo základných súborov s normálnym rozdelením vykazujú neparametrické metódy v porovnaní s parametrickými menšiu silu testu. Na druhej strane v prípade veľkých vzoriek aj pri zjavnom porušení normality možno pokojne použiť parametrické metódy, ktoré sa počítajú ľahšie ako neparametrické.

Kedy a aký test použiť

Parametrické testy je vhodné použiť, ak sú splnené predpoklady:
- normality rozloženia základného súboru,
- homogenity rozptylu,
- rovnosti intervalov (meranie na intervalovej a pomerovej škále).

Neparametrické testy používame:
- v situáciách, keď sú k dispozícií iba klasifikačné alebo poradové dáta,
- ak sú porušené základné predpoklady parametrických testov (parametrický test stráca na vypovedacej schopnosti a je rozumné zmeniť dáta do formátu kompatibilného pre analýzu s vhodným neparametrickým testom),
- ak nevieme overiť, že výsledky meraní sú normálne rozdelené.

Vo všeobecnosti sú parametrické testy robustnejšie (silnejšie). Robustnosťou alebo silou štatistického testu rozumieme pravdepodobnosť, s ktorou bude nulová hypotéza zamietnutá, ak je skutočne nesprávna. Ak však vznikne podozrenie, že sa rozdelenia v niektorom smere rozchádzajú s požadovanými predpokladmi, platnosť alebo preukázateľnosť záverov, ku ktorým sme sa dopracovali pomocou parametrických testov sa znižuje alebo vytráca a mali by sme dať prednosť neparametrickým testom.

Neparametrické testy sa dajú použiť aj pre údaje z intervalovej a pomerovej škály, ak ich prevedieme na poradové dáta zavedením vhodnej kategorizácie alebo vhodného usporiadania. Pri tomto prevedení môže však dôjsť k citeľnému informačnému úbytku.

Základný princíp neparametrických testov

Test je založený na poradiach, to znamená, že môžeme pracovať s poradovými alebo kvantitatívnymi premennými. Kvantitatívne dáta prevedieme na poradia tak, že najmenšie číslo dostane poradie 1, druhé najmenšie dostane poradie 2,... a najväčšie dostane poradie n, kde n zodpovedá veľkosti súboru. Ak pracujeme s poradovými dátami, občas sa objaví problém s rovnakými hodnotami. Napr. ako by mali byť zoradené merania 2 5 7 7 10. Obyčajne sa odporúča použiť metódu priemerného poradia. Táto metóda priradí rovnaké poradie každému rovnakému meraniu, toto poradie je jednoducho priemer z poradí, ktoré obsadzujú rovnaké merania.

Prvok	2	5	7	7	10
Poradie	**1**	**2**	$\frac{3+4}{2} = \mathbf{3,5}$	$\frac{3+4}{2} = \mathbf{3,5}$	**5**

Výhody a nevýhody neparametrických testov

Výhody neparametrických testov:
- nie sú závislé od tvaru rozdelenia základného súboru,
- nevyžadujú výpočet parametrov distribučnej funkcie,
- nevyžadujú typ jednovrcholového rozdelenia,
- sú nenáročné na informácie (jednoduché výpočty testovacích charakteristík),
- sú použiteľné aj pre malé rozsahy výberov.

Nevýhody neparametrických testov
- menšia sila testu v porovnaní s parametrickými testami (kompenzuje sa väčším počtom meraní),
- v prípade kvantitatívnych znakov nevyužívajú celú informáciu o údajoch, ale len poradie údajov.

Znamienkový test

Znamienkový test je jeden z najjednoduchších neparametrických testov. Je špeciálne vytvorený na testovanie hypotézy s mediánom akéhokoľvek výberu. Medián je mierou stredu alebo umiestnenia rozdelenia, preto sa niekedy označuje ako test určenia polohy – mediánový test. V porovnaní s inými neparametrickými testami má malú silu. Pre čo najlepší výsledok je treba veľký počet meraní.

Nech X_1 , ... , X_n je náhodný výber zo spojitého rozdelenia. Označme M medián tohto rozdelenia. Pravdepodobnosť, že vybraná hodnota X_i z výberu je väčšia alebo menšia ako M je 0.5, t.j. $P(X_i > M) = P(X_i < M) = 0.5$. Testujeme hypotézu, že sa medián rovná zvolenému číslu $H_0 \colon M = X_0$ oproti alternatívnej hypotéze, že sa mu nerovná $H_1 \colon M \neq X_0$ (pre obojstranný test), pre jednostranný test $H_1 \colon M > X_0$ (alebo $H_1 \colon M < X_0$), kde X_0 je zvolené číslo. Znamienkový test používa testovaciu štatistiku S, kde S je počet meraní, ktoré presiahnu hodnotu X_0. Vytvoríme najskôr rozdiely $X_1 - X_0, \dots, X_n - X_0$, a počet kladný rozdielov označíme symbolom S. Ak sú rozdiely nulové, tak ich vynecháme a počet meraní znížime o počet nulových rozdielov. Pre $n < 30$ H_0 nezamietame, ak $k_1 < S < k_2$ (na danej hladine významnosti α pre tabuľkové kritické hodnoty k_1, k_2). Pre $n \geq 30$ štatistiku S aproximujeme testovacou štatistikou z s normálnym rozdelením $N(0,1)$ podľa vzorca:

$$z = \frac{2S - n}{\sqrt{n}} \approx N(0,1)$$

Pre jednostranný test zamietame H_0 na hladine významnosti α, ak $z > u_{1-\alpha}$ $(alebo\ z < u_\alpha)$ a pre obojstranný test, ak $|z| > u_{1-\alpha/2}$, kde $u_{1-\alpha/2}, u_{1-\alpha}, u_\alpha$ sú príslušné kvantily normálneho rozdelenia.

Príklad:
V tabuľke sú uvedené priemerné platy a dôchodky niektorých krajín európskej únie (v €).

Luxembursko	Plat:	3 880,-	**Taliansko**	Plat:	2 346,-
	Dôchodok :	3 426,-		Dôchodok :	1 593,-
Holandsko	Plat:	3 222,-	**Cyprus**	Plat:	1 775,-
	Dôchodok :	2 693,-		Dôchodok :	853,-
Nemecko	Plat:	2 951,-	**Grécko**	Plat:	1 710,-
	Dôchodok :	1 177,-		Dôchodok:	1 636,-
Francúzsko	Plat:	2 935,-	**Španielsko**	Plat:	1 617,-
	Dôchodok :	1 502,-		Dôchodok	1 318,-
Fínsko	Plat:	2 873,-	**Slovinsko**	Plat:	1381,-
	Dôchodok :	2 666,-		Dôchodok :	341,-
Rakúsko	Plat:	2 801,-	**Malta**	Plat:	1344,-
	Dôchodok :	1 288,-		Dôchodok	416,-
Belgicko	Plat:	2 799,-	**Portugalsko**	Plat:	1 253,-
	Dôchodok :	2 099,-		Dôchodok:	451,-
Írsko	Plat:	2 412,-	**Slovensko**	Plat:	678,-
	Dôchodok:	783,-		Dôchodok:	311,-

(Zdroj dát www.ksskvke.sk)

Na hladine významnosti $\alpha = 0.05$ testujte hypotézu, že medián priemerných platov je väčší ako priemerný plat Slovenska.

Riešenie:

Testujeme H_0: $M \leq 678$ oproti H_1: $M > 678$. Počet kladných rozdielov zistíme tak, že od každej hodnoty odčítame priemerný plat Slovenska.

PP	PP-678	Rozdiely
3880	3202	+
3222	2544	+
2951	2273	+
2935	2257	+
2873	2195	+
2801	2123	+
2799	2121	+
2412	1734	+
2346	1668	+
1775	1097	+
1710	1032	+
1617	939	+
1381	703	+
1344	666	+
1253	575	+
678	0	0
Počet kladných rozdielov(Sk)		**15**

Počet kladných rozdielov: 15

Vypočítame medián výberu: 2379

Podľa vzorca vypočítame testovaciu štatistiku:

$$z = \frac{2.15 - 16}{\sqrt{16}} = 3{,}5$$

Pretože ide o jednostranný test, porovnáme hodnotu testovacej štatistiky z s kvantilom $u_{1-\alpha}$ normálneho rozdelenia N(0,1) pre $\alpha = 0{,}05$: $3{,}5 > 1.65$.

Vypočítaná testovacia štatistika je väčšia ako kvantil N(0,1) čo znamená, že **zamietame** H_0 na hladine významnosti $\alpha = 0{,}05$. Priemerný plat Slovenska je menší ako medián daného výberu.

Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:

```
ZnamienkovyTest[data,α,X0]
data – dáta súboru (zadáva sa ako list)
α – hladina významnosti
X0 – testovaná hodnota
```

Výstup procedúry na počítanom príklade:

Median je :	2379	
	Obojstranny test	Jednostranny test
Hladina vyznamnosti α	0.05	
p value	0.000465258	0.000232629
Vyhodnotenie na zaklade p value	Zamietam H_0	Zamietam H_0
Vypocitana statistika z	3.5	
Kvantil N(0,1)	1.95996	1.64485
Vyhodnotenie na zaklade kvantilu	Zamietam H_0	Zamietam H_0

Dvojvýberové poradové testy

Sú neparametrickou obdobou parametrického dvojvýberového t-testu. Ich podstatou je, že máme k dispozícii dva nezávislé výbery s distribučnými funkciami. K štatistickému overeniu rozdielov medzi dvoma nezávislými výbermi používame najčastejšie Wald-Wolfowitzov test iterácií a Kolmogorov-Smirnovov test.

Pre obidva testy použijeme rovnaký príklad.

Príklad:
V tabuľke sú (v %) uvedené dôvody, prečo ľudia nechcú mať internet v domácnosti. Na hladine významnosti $\alpha = 0{,}05$ testujeme hypotézu, že výbery pochádzajú z rovnakého rozdelenia.

Štát	Majú prístup inde	Nechcú internet	Nepotrebujú internet
	1. dôvod	2. dôvod	3. dôvod
Česká republika	16	5	37
Dánsko	6	1	57
Nemecko	32	7	51
Estónsko	31	46	48
Írsko	15	15	28
Grécko	12	20	45
Španielsko	14	20	56
Francúzsko	20	33	45
Cyprus	14	17	52
Lotyšsko	33	23	49
Litva	8	1	69
Luxembursko	9	17	53
Maďarsko	26	25	49
Malta	3	15	33
Holandsko	12	27	40
Rakúsko	20	19	49
Polsko	8	5	45
Portugalsko	8	45	72
Rumunsko	9	11	25
Slovinsko	13	15	61
Slovensko	24	6	40
Švédsko	29	42	42
Anglicko	13	21	30

(Dáta-zdroj Eurostat)

Pomocou Wald-Wolfowitzovho testu a Kolmogorovho-Smirnovovho testu overte, či výbery pochádzajú z rovnakého rozdelenia. Keďže tento príklad použijeme aj neskôr, vybrali sme tri dôvody, pričom najskôr otestujeme 1. vs. 2. dôvod a potom 1. vs 3. dôvod.

Kolmogorov – Smirnovov test

Je navrhnutý na testovanie nulovej hypotézy $H_0: F(x) = G(x)$ pre všetky $x \in \mathcal{R}$ oproti alternatívnej hypotéze $H_1: F(x) \neq G(x)$ aspoň pre jedno x. F a G sú distribučné funkcie dvoch nezávislých výberov.

Nech $X = (X_1, \dots, X_m), Y = (Y_1, \dots, Y_n)$ sú dva nezávislé náhodné výbery rôznej veľkosti m a n. Spojíme ich do jedného výberu Z dĺžky $n + m$. Pre reálne číslo x definujme náhodné veličiny $\xi_i(x) = \begin{cases} 1 & pre\ Z_i < x \\ 0 & pre\ Z_i \geq x \end{cases}$, $i = 1, \dots, m + n$. Označme ako $\hat{F}(x)$ empirickú distribučnú funkciu 1. výberu a $\hat{G}(x)$ empirickú distribučnú funkciu 2. výberu. Potom $\hat{F}(x) = \dfrac{1}{m}\sum_{i=1}^{m}\xi_i(x)$ a $\hat{G}(x) = \dfrac{1}{n}\sum_{i=m+1}^{m+n}\xi_i(x)$. Testovaciu štatistiku D získame podľa vzťahu $D = \max\limits_{-\infty<x<\infty} \left| \hat{F}(x) - \hat{G}(x) \right|$.

Nulovú hypotézu zamietame na hladine významnosti α, ak $D \geq D_{m,n}$, kde $D_{m,n}$ je tabuľková kritická hodnota.

V prípade väčších výberov m, n aproximujeme $D_{m,n}$ číslom D^*

$$D^* = \sqrt{\frac{m+n}{2\,m\,n}\, ln\frac{2}{\alpha}}$$

H_0 zamietame, ak $D > D^*$.

Riešenie:

1. vs. 2. dôvod

	$\hat{F}(x)$	$\hat{G}(x)$	$\hat{F}(x) - \hat{G}(x)$
1	0	2/23	-2/23
3	1/23	2/23	-1/23
5	1/23	4/23	-3/23
5	1/23	4/23	-3/23
6	2/23	5/23	-3/23
6	2/23	5/23	-3/23
7	2/23	6/23	-4/23
8	5/23	6/23	-1/23
8	5/23	6/23	-1/23
8	5/23	6/23	-1/23
9	7/23	6/23	1/23
... pokračujeme po koniec			

$$D = 0.261$$

$$D^* = \sqrt{\frac{23+23}{2*23*23}\, ln\frac{2}{0{,}05}} = 0.400$$

$0{,}261 < 0{,}400$ => **nezamietam** H_0 na hladine významnosti $\alpha = 0{,}05$ (obidva výbery sú z rovnakého rozdelenia)

1. vs. 3. dôvod

	$\hat{F}(x)$	$\hat{G}(x)$	$\hat{F}(x) - \hat{G}(x)$
3	1/23	0	1/23
6	2/23	0	2/23
8	5/23	0	5/23
8	5/23	0	5/23
8	5/23	0	5/23
9	7/23	0	7/23
9	7/23	0	7/23
12	9/23	0	9/23
12	9/23	0	9/23
13	11/23	0	11/23
13	11/23	0	11/23
... pokračujeme po koniec			

$$D = 0.826$$

$$D^* = \sqrt{\frac{23+23}{2*23*23}\,ln\frac{2}{0{,}05}} = 0.400$$

$0{,}826 > 0{,}400$ => **zamietam** H_0 na hladine významnosti $\alpha = 0{,}05$ (výbery nie sú z rovnakého rozdelenia)

Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:

```
KolmogorovSmirnovTest[data1,data2,α]
data1 – dáta prvého súboru (zadáva sa ako list)
data2 – dáta druhého súboru (zadáva sa ako list)
α – hladina významnosti
```

Výstup procedúry na počítanom príklade:

1. vs. 2. Dôvod

```
n1 : 23
n2 : 23
Priemer 1 : 16.3043
Priemer 2 : 18.9565
```

	Obojstranny test
Hladina vyznamnosti α	0.05
Vypocitane D	0.26087
Kriticka hodnota D*	0.400482
Vyhodnotenie na zaklade KH	Nezamietam H_0

1. vs. 3. dôvod

```
n1 : 23
n2 : 23
Priemer 1 : 16.3043
Priemer 2 : 46.7826
```

	Obojstranny test
Hladina vyznamnosti α	0.05
Vypocitane D	0.826087
Kriticka hodnota D*	0.400482
Vyhodnotenie na zaklade KH	Zamietam H_0

Wald – Wolfowitzov test

Je test nepretržitých iterácií (opakovaní). Overujeme ním nulovú hypotézu H_0, že obidva výbery pochádzajú z rovnakého rozdelenia, proti alternatívnej hypotéze H_1, že sa v niektorom aspekte rozchádzajú, ako napr. medián, variabilita alebo tvar rozloženia.

Pri aplikácií testu na dáta z dvoch nezávislých výberov $X = (X_1, \dots, X_m), Y = (Y_1, \dots, Y_n)$ rôznej veľkosti m a n všetkých $m + n$ hodnôt najprv usporiadame do neklesajúcej postupnosti podľa veľkosti. Identitu každého prvku určíme tým, že ho označíme značkou skupiny (X alebo Y), do ktorej patrí. Potom sledujeme poradie znakov X a Y, spočítame ich a dostaneme počet iterácií r (t.j. počet postupností po sebe nasledujúcich hodnôt patriacich do toho istého výberu).

Pre výbery malých rozsahov $(m, n < 20)$ H_0 zamietame, ak $r \leq r_\alpha(m, n)$, pričom $r_\alpha(m, n)$ je tabuľková kritická hodnota.

Ak ide o veľké výbery, potom je rozdelenie r (za platnosti H_0) aproximované normálnym rozložením s priemerom Mr a štandardnou odchýlkou Sr:

$$Mr = \frac{2\,m\,n}{m+n} + 1 \qquad\qquad Sr = \sqrt{\frac{2\,m\,n\,(2\,m\,n - m - n)}{(m+\,n)^2\,(m+\,n-1)}}$$

Testovaciu štatistiku vypočítame podľa vzorca:

$$z = \frac{\left|r - \left(\frac{2\,m\,n}{m+n} + 1\right)\right| - 0{,}5}{\sqrt{\frac{2\,m\,n\,(2\,m\,n - m - n)}{(m+\,n)^2\,(m+\,n-1)}}} \approx N(0{,}1)$$

H_0 zamietame na hladine významnosti α, ak $|z| \geq u_{1-\alpha/2}$, kde $u_{1-\alpha/2}$ je kvantil normálneho rozdelenia N(0,1)

Riešenie:
Hodnoty oboch výberov usporiadame vzostupne a priradíme im skupinu 1 alebo 2 (resp. 1 alebo 3). V krútených zátvorkách sú uvedené iterácie.

1. vs. 2. dôvod

{2,2},{1},{2,2},{1},{2,2},{1,1,1,1,1},{2},{1,1,1,1,1,1,1},{2,2,2},{1},{2,2,2},{1,1},{2,2,2,2},{1},{2},{1},{2},{1,1,1,1},{2,2,2,2}

Počet iterácií je 19 => $r = 19$

$$z = \frac{\left|19 - \left(\frac{2 * 23 * 23}{23 + 23} + 1\right)\right| - 0,5}{\sqrt{\frac{2 * 23 * 23\,(2 * 23 * 23 - 23 - 23)}{(23 + 23)^2\,(23 + 23 - 1)}}} = 1.342$$

Keďže platí $1,342 < 1,96 = u_{1-\alpha/2}$ **nezamietam** H_0 na hladine významnosti $\alpha = 0,05$, t.j. obidva výbery pochádzajú z rovnakého rozdelenia.

1. vs. 3. dôvod

{1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1},{3},{1},{3},{1},{3},{1,1,1},{3,3,3,3,3,3,3,3,3,3,3,3,3, 3,3,3,3,3,3,3}

Počet iterácií je 8 => $r = 8$

$$z = \frac{\left|8 - \left(\frac{2 * 23 * 23}{23 + 23} + 1\right)\right| - 0,5}{\sqrt{\frac{2 * 23 * 23\,(2 * 23 * 23 - 23 - 23)}{(23 + 23)^2\,(23 + 23 - 1)}}} = 4.622$$

Keďže platí $4,622 > 1,96 = u_{1-\alpha/2}$ **zamietam** H_0 na hladine významnosti $\alpha = 0,05$, t.j. náhodné výbery nie sú z rovnakého rozdelenia.

Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:

```
WaldWolfowitzovTest[data1,data2,α]
data1 – dáta prvého súboru (zadáva sa ako list)
data2 – dáta druhého súboru (zadáva sa ako list)
α – hladina významnosti
```

Výstup procedúry na počítanom príklade:

1. vs. 2. dôvod

```
n1 : 23
n2 : 23
Priemer 1 : 16.3043
Priemer 2 : 18.9565
```

Pocet iteracii :	19
	Obojstranny test
Hladina vyznamnosti α	0.05
p value	0.179605
Vyhodnotenie na zaklade p value	Nezamietam H_0
Vypocitana z statistika	1.34197
Kvantil N(0,1)	1.95996
Vyhodnotenie na zaklade kvantilu	Nezamietam H_0

1. vs. 3. dôvod

n1 : 23

n2 : 23

Priemer 1 : 16.3043

Priemer 2 : 46.7826

Pocet iteracii :	8
	Obojstranny test
Hladina vyznamnosti α	0.05
p value	3.7942×10^{-6}
Vyhodnotenie na zaklade p value	Zamietam H_0
Vypocitana z statistika	4.62235
Kvantil N(0,1)	1.95996
Vyhodnotenie na zaklade kvantilu	Zamietam H_0

Neparametrická analýza rozptylu

Výhodou poradovej analýzy rozptylu je predovšetkým to, že nemusíme skúmať normalitu rozdelenia, čo je potrebné brať do úvahy pri jej parametrickej obdobe. Pretože neparametrický variant tejto štatistiky vychádza z poradových hodnôt, je aplikovateľný na omnoho širší okruh výskumných problémov.

Kruskal – Walisova analýza rozptylu pre k nezávislých výberov

Predpokladom pre použitie testu je, aby všetky pozorovania boli na sebe nezávislé, aby uvažovaná premenná bola meraná na ordinálnej škále, a aby všetkých k výberových distribučných funkcií malo približne rovnaký tvar.

Majme k nezávislých náhodných výberov $n_1, \dots, n_k$,kde n_i je veľkosť i-teho výberu, ktorý pochádza z rozdelenia so spojitou distribučnou funkciou $F_i(x), i = 1, \dots, k$. Označme $N = \sum_{i=1}^{k} n_i$. Testujeme hypotézu, že všetky výbery pochádzajú z rovnakého rozdelenia, poprípade z rozdelení s identickými distribučnými funkciami: $H_0\colon F_1(x) = F_2(x) = \cdots = F_k(x)$ oproti alternatívnej hypotéze, že nie všetky distribučné funkcie sa rovnajú $H_1\colon neplatí\, H_0$. Všetky namerané hodnoty zoradíme do jednej vzostupnej postupnosti a každej hodnote priradíme poradie, ktoré jej prislúcha v celkovom usporiadaní. Pre každý výber spočítame poradové hodnoty všetkých jeho členov a ich súčet v i-tom výbere označíme R_i. Za predpokladu platnosti H_0 by sa priemery poradí jednotlivých výberov $\bar{R} = R_i/n_i$ nemali výrazne odlišovať a pozorované odchýlky by mali vyjadrovať náhodné kolísanie. Na zistenie platnosti H_0 použijeme testovaciu štatistiku H podľa vzorca:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

ktorá má približne χ^2 rozdelenie s $(k-1)$ stupňami voľnosti, ak v žiadnom výbere nie je počet pozorovaní menší ako 5. H_0 zamietame na hladine významnosti α v prípade, že $H \geq \chi^2_{1-\alpha}(k-1)$, kde $\chi^2_{1-\alpha}(k-1)$ je kritická hodnota χ^2 rozdelenia s $(k-1)$ stupňami voľnosti.

Ak je rovnakých hodnôt v postupnosti značný počet, priradíme im priemerné poradie a H štatistika musí byť upravená korekciou, ktorú urobíme tak, že výraz H vydelíme korekčným indexom:

$$H_{kor} = \frac{H}{1 - \sum_{j=1}^{s} \left(t_j^3 - t_j\right) / (N^3 - N)}$$

$s - počet\ hodnôt, ktoré\ sa\ opakujú$

$t_j - koľkokrát\ sa\ opakuje\ j - ta\ hodnota\ v\ usporiadanej\ postupnosti$

Korekcia zvyšuje hodnotu H. Pokiaľ nie je rovnakých pozorovaní príliš mnoho a nie sú príliš rozsiahle je účinok korekcie nepatrný.

Príklad:

V tabuľke sú (v %) uvedené dôvody, prečo ľudia nechcú mať internet v domácnosti. Testujeme hypotézu, že všetky výbery pochádzajú z rovnakého rozdelenia.

Štát	Majú prístup inde		Nechcú internet		Nepotrebujú internet		Vybavenie je drahé		Pripojenie je drahé		Málo skúseností	
	Data	R	Data	R	Data	R	Data	R	Data	R	Data	R
Česká republika	16	**47,5**	5	**7,0**	37	**102,0**	35	**99,0**	27	**78,5**	32	**89,5**
Dánsko	6	**9,5**	1	**1,5**	57	**132,0**	5	**7,0**	9	**21,5**	10	**24,5**
Nemecko	32	**89,5**	7	**13,0**	51	**126,0**	34	**96,0**	32	**89,5**	31	**86,5**
Estónsko	31	**86,5**	46	**117,0**	48	**119,5**	58	**133,0**	51	**126,0**	62	**135,5**
Írsko	15	**43,0**	15	**43,0**	28	**80,5**	15	**43,0**	7	**13,0**	20	**61,5**
Grécko	12	**28,5**	20	**61,5**	45	**114,5**	14	**37,5**	9	**21,5**	30	**84,5**
Španielsko	14	**37,5**	20	**61,5**	56	**131,0**	20	**61,5**	20	**61,5**	26	**75,0**
Francúzsko	20	**61,5**	33	**93,0**	45	**114,5**	41	**108,5**	37	**102,0**	37	**102,0**
Cyprus	14	**37,5**	17	**50,5**	52	**128,5**	13	**33,0**	12	**28,5**	41	**108,5**
Lotyšsko	33	**93,0**	23	**67,5**	49	**122,5**	52	**128,5**	43	**112,0**	48	**119,5**
Litva	8	**17,5**	1	**1,5**	69	**137,0**	18	**53,5**	12	**28,5**	7	**13,0**
Luxembursko	9	**21,5**	17	**50,5**	53	**130,0**	7	**13,0**	3	**3,5**	16	**47,5**
Maďarsko	26	**75,0**	25	**71,0**	49	**122,5**	49	**122,5**	34	**96,0**	28	**80,5**
Malta	3	**3,5**	15	**43,0**	33	**93,0**	7	**13,0**	4	**5,0**	34	**96,0**
Holandsko	12	**28,5**	27	**78,5**	40	**106,0**	10	**24,5**	8	**17,5**	17	**50,5**
Rakúsko	20	**61,5**	19	**56,0**	49	**122,5**	15	**43,0**	14	**37,5**	17	**50,5**
Polsko	8	**17,5**	5	**7,0**	45	**114,5**	29	**82,5**	26	**75,0**	23	**67,5**
Portugalsko	8	**17,5**	45	**114,5**	72	**138,0**	51	**126,0**	47	**118,0**	62	**135,5**
Rumunsko	9	**21,5**	11	**26,0**	25	**71,0**	35	**99,0**	26	**75,0**	20	**61,5**
Slovinsko	13	**33,0**	15	**43,0**	61	**134,0**	35	**99,0**	32	**89,5**	39	**104,0**
Slovensko	24	**69,0**	6	**9,5**	40	**106,0**	18	**53,5**	15	**43,0**	19	**56,0**
Švédsko	29	**82,5**	42	**110,5**	42	**110,5**	25	**71,0**	26	**75,0**	40	**106,0**
Anglicko	13	**33,0**	21	**66,0**	30	**84,5**	19	**56,0**	13	**33,0**	13	**33,0**
Suma R		1015,5		1192,5		2640,5		1603,5		1350,5		1788,5
R^2/n	44836,53		61828,53		303140,88		111791,84		79297,84		139075,32	

(Dáta-zdroj Eurostat, R-dopočítané poradia)

Riešenie:

$$H = 12/(138 * (138+1))(44836{,}53 + 61828{,}53 + 303140{,}88 + 111791{,}84 + 79297{,}84 + 139075{,}32) - 3 * (138+1) = 45{,}916$$

Keďže niektoré hodnoty sa opakujú mali by sme použiť korekciu

Hodnota 1 sa opakuje dvakrát	$2^3 - 2 = 6$	
Hodnota 3 sa opakuje dvakrát	$2^3 - 2 = 6$	
Hodnota 5 sa opakuje trikrát	$3^3 - 3 = 24$	$C = 1 - \frac{1998}{23*(6^3-6)} = 0{,}9992$
Hodnota 6 sa opakuje dvakrát	$2^3 - 2 = 6$	
Hodnota 7 sa opakuje päťkrát	$5^3 - 5 = 120$	
Hodnota 8 sa opakuje štyrikrát	$4^3 - 4 = 60$	$H_{kor} = \frac{45{,}916}{0{,}9993} = 45{,}951$
Hodnota 9 sa opakuje štyrikrát	$4^3 - 4 = 60$	
	...	
Rovnako pokračujeme pre ostávajúce opakujúce sa prvky		$45{,}951 > 11{,}071 = \chi^2_{0{,}95}(5)$
	...	
	$T = 1998$	

Zamietame H_0 na danej hladine významnosti $\alpha = 0{,}05$, t.j. je štatisticky významný rozdiel aspoň medzi dvomi dôvodmi, prečo ľudia nechcú mať internet doma. Zamietnutie H_0 sme dokázali aj pomocou dvojvýberových testov (Kolmogorov-Smirnovov test, Wald-Wolfowitzov test), keď sme pri porovnávaní prvého a tretieho dôvodu zamietli H_0.

Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:

```
KruskalWallisovTest[data,α]
data -dáta súborov (zadáva sa ako list)
α - hladina významnosti
```

Výstup procedúry na počítanom príklade:

```
Sucet poradi v stlpci 1 : 1015.5
Sucet poradi v stlpci 2 : 1192.5
Sucet poradi v stlpci 3 : 2640.5
Sucet poradi v stlpci 4 : 1603.5
Sucet poradi v stlpci 5 : 1350.5
Sucet poradi v stlpci 6 : 1788.5
```

Hladina vyznamnosti α	0.05
H	45.9158
Korekcny koeficient	0.99924
H korekcia	45.9508
p value	9.2942×10^{-9}
Vyhodnotenie na zaklade p value	Zamietam H_0
Stupne volnosti	5
Kriticka hodnota χ^2 rozdelenia	11.0705
Vyhodnotenie na zaklade KH	Zamietam H_0

Friedmanova analýza rozptylu pre k závislých výberov

Je neparametrickým variantom parametrickej dvojrozmernej variačnej analýzy pre dáta z ordinálnej škály.

Základom Friedmanovho testu je tabuľka dvojrozmerného triedenia X_{ij} $(i = 1, \dots, n; j = 1, \dots, k)$, ktorá má n riadkov a k stĺpcov, kde n označuje rozsah výberu, ktorý je rovnaký pre všetky výbery a k počet testovaných výberov. Testujeme nulovú hypotézu, že všetky výbery pochádzajú z jedného rozdelenia, teda ich distribučné funkcie sú totožné H_0: $F_1(x) = F_2(x) = \cdots = F_k(x)$. Za platnosti nulovej hypotézy by mali byť súčty poradí pre každý stĺpec či výber približne rovnaké. Najskôr priradíme každej hodnote v riadku i poradie $R_{X_{ij}}$, od 1 po k podľa veľkosti. Hodnotu R dostaneme tak, že spočítame výsledné súčty poradí v stĺpcoch umocnené na druhú $R = \sum_{j=1}^{k} \left(\sum_{i=1}^{n} R_{X_{ij}} \right)^2$ a testovaciu štatistiku vypočítame podľa vzorca:

$$F = \frac{12}{nk(k+1)} R - 3n(k+1)$$

H_0 zamietame na hladine významnosti α v prípade, že $F \geq \chi^2_{1-\alpha}(k-1)$, kde $\chi^2_{1-\alpha}(k-1)$ je kritická hodnota χ^2 rozdelenia s $(k-1)$ stupňami voľnosti.

Ak je rovnakých hodnôt v postupnosti značný počet, priradíme im priemerné poradie a F štatistika musí byť upravená korekciou, ktorú urobíme tak, že výraz F vydelíme korekčným indexom:

$$F_{kor} = \frac{F}{1 - \sum_{j=1}^{s} \left(t_j^3 - t_j \right) / n(k^3 - k)}$$

$s - \text{počet hodnôt, ktoré sa opakujú}$

$t_j - \text{koľkokrát sa opakuje } j - \text{ta hodnota v usporiadanej postupnosti}$

Korekcia zvyšuje hodnotu F. Pokiaľ nie je rovnakých pozorovaní príliš mnoho a nie sú príliš rozsiahle je účinok korekcie nepatrný.

Príklad:

V tabuľke je uvedená priemerná ročná miera inflácie (v %) v rokoch 2004-2007. Testujeme hypotézu, že všetky výbery pochádzajú z jedného rozdelenia.

Štát	2004		2005		2006		2007	
	Data	R	Data	R	Data	R	Data	R
Česká republika	2,6	**3**	1,6	**1**	2,1	**2**	3	**4**
Dánsko	0,9	**1**	1,7	**2,5**	1,9	**4**	1,7	**2,5**
Nemecko	1,8	**1,5**	1,9	**3**	1,8	**1,5**	2,3	**4**
Estónsko	3	**1**	4,1	**2**	4,4	**3**	6,7	**4**
Írsko	2,3	**2**	2,2	**1**	2,7	**3**	2,9	**4**
Grécko	3	**1,5**	3,5	**4**	3,3	**3**	3	**1,5**
Španielsko	3,1	**2**	3,4	**3**	3,6	**4**	2,8	**1**
Francúzsko	2,3	**4**	1,9	**2,5**	1,9	**2,5**	1,6	**1**
Cyprus	1,9	**1**	2	**2**	2,2	**3,5**	2,2	**3,5**
Lotyšsko	6,2	**1**	6,9	**3**	6,6	**2**	10,1	**4**
Litva	1,2	**1**	2,7	**2**	3,8	**3**	5,8	**4**
Luxembursko	3,2	**3**	3,8	**4**	3	**2**	2,7	**1**
Maďarsko	6,8	**3**	3,5	**1**	4	**2**	7,9	**4**
Malta	2,7	**4**	2,5	**2**	2,6	**3**	0,7	**1**
Holandsko	1,4	**1**	1,5	**2**	1,7	**4**	1,6	**3**
Rakúsko	2	**2**	2,1	**3**	1,7	**1**	2,2	**4**
Polsko	3,6	**4**	2,2	**2**	1,3	**1**	2,6	**3**
Portugalsko	2,5	**3**	2,1	**1**	3	**4**	2,4	**2**
Rumunsko	11,9	**4**	9,1	**3**	6,6	**2**	4,9	**1**
Slovinsko	3,7	**3**	2,5	**1,5**	2,5	**1,5**	3,8	**4**
Slovensko	7,5	**4**	2,8	**2**	4,3	**3**	1,9	**1**
Švédsko	1	**2**	0,8	**1**	1,5	**3**	1,7	**4**
Anglicko	1,3	**1**	2,1	**2**	2,3	**3,5**	2,3	**3,5**
Suma poradi		53		50,5		61,5		65

(Dáta-zdroj Eurostat, R-dopočítané poradia)

<u>Riešenie:</u>

$$F = \frac{12}{23*4*(4+1)}\left(51{,}9^2 + 49{,}5^2 + 61{,}5^2 + 65^2\right) - 3*23*(4+1) = 3{,}6913$$

Keďže niektoré hodnoty sa opakujú mali by sme použiť korekciu

Hodnota 1,7 sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota 1,8 sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota 3 sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota 1,9 sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota 2,2 sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota 2,5 sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota 2,3 sa opakuje dvakrát $2^3 - 2 = 6$

$T = 42$

$$C = 1 - \frac{42}{23*(4^3 - 4)} = 0{,}9696$$

$$F_{kor} = \frac{3{,}6913}{0{,}9696} = 3{,}8072$$

$$3{,}8072 < 7{,}815 = \chi^2_{0{,}95}(3)$$

Nezamietame H_0 na danej hladine významnosti $\alpha = 0{,}05$, t.j. nie je štatisticky významný rozdiel medzi jednotlivými rokmi.

<u>Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:</u>

```
FriedmanovTest[data, α]
data – dáta súborov (zadáva sa ako list)
α – hladina významnosti
```

Výstup procedúry na počítanom príklade:

```
Sucet poradi v stlpci 1 : 53.
Sucet poradi v stlpci 2 : 50.5
Sucet poradi v stlpci 3 : 61.5
Sucet poradi v stlpci 4 : 65.
```

Hladina vyznamnosti α	0.05
Fr	3.6913
Korekcny koeficient	0.969565
Fr korekcia	3.80717
p value	0.283053
Vyhodnotenie na zaklade p value	Nezamietam H_0
Stupne volnosti	3
Kriticka hodnota χ^2 rozdelenia	7.81473
Vyhodnotenie na zaklade KH	Nezamietam H_0

Testy nezávislosti

Korelačný koeficient meria silu štatistickej závislosti medzi dvoma číselnými premennými. Pod pojmom korelačný koeficient sa najčastejšie myslí **Pearsonov korelačný koeficient** (Pearson's product moment) z roku 1896, ktorý je mierou lineárnej závislosti dvoch premenných. Pearsonov korelačný koeficient sa vypočíta:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{s_x s_y}$$

Čitateľ zlomku sa nazýva **kovariancia** a vyjadruje, ako sa súčasne menia hodnoty dvoch premenných. Kladná hodnota znamená, že sa menia spoločne jedným smerom, záporná hodnota znamená, že sa menia opačným smerom a nula, že sa menia nezávisle. Vydelením kovariancie štandardnými odchýlkami sa vypočíta Pearsonov korelačný koeficient. Hodnota všetkých korelačných koeficientov sa nachádza v intervale od −1 do 1, pričom −1 znamená dokonalú negatívnu závislosť, 1 znamená dokonalú pozitívnu závislosť a 0 nezávislosť X a Y. V prípade Pearsonovho korelačného koeficientu hodnota −1 znamená, že všetky body v X-Y grafe ležia na klesajúcej priamke, hodnota 1, že ležia na stúpajúcej priamke a hodnota 0, že sú rozptýlené.

Pearsonov korelačný koeficient je silne ovplyvniteľný extrémnymi hodnotami (outliers) a to v oboch smeroch. Jediná extrémna hodnota vo veľkom súbore môže významne znížiť silnú závislosť, ale aj vyrobiť silnú závislosť tam, kde žiadna nie je. Touto citlivosťou na extrémne hodnoty sa nevyznačujú neparametrické poradové korelačné koeficienty. Najčastejšie používaným poradovým korelačným koeficientom je **Spearmanov koeficient**, ktorý možno vypočítať zo vzorca na výpočet Pearsonovho koeficientu, pričom namiesto originálnych hodnôt sa použijú ich poradové čísla. Spearmanov korelačný koeficient je preto veľmi ťažko interpretovateľný. **Kendallov koeficient** (1948) na rozdiel od Spearmanovho vyjadruje silu závislosti medzi dvoma poradovými premennými, čiže vyjadruje rozdiel medzi pravdepodobnosťou, že hodnoty dvoch premenných sú v rovnakom poradí oproti pravdepodobnosti, že hodnoty nie sú v rovnakom poradí.

Obidva testy overujú platnosť nulovej hypotézy, že medzi usporiadanými pármi neexistuje žiadna korelácia (závislosť) $H_0: c = 0$ proti alternatívnej hypotéze , že usporiadané páry sú korelované $H_0 \neq c$ (pre obojstranný test) alebo že usporiadané páry sú pozitívne korelované, $H_1: c > 0$ (alebo usporiadané páry sú negatívne korelované, $H_1: c < 0$) pre jednostranný test. Symbol c je buď Spearmanov koeficient ρ alebo Kendallov koeficient τ.

Hodnoty oboch neparametrických korelačných koeficientov (aj keď sú vypočítané z rovnako zoradených nameraných údajov) sa nerovnajú a nie sú navzájom priamo porovnateľné. Kendallovo τ dáva hodnoty nižšie než Spearmanovo ρ. Ale v oboch koeficientoch sú v podstate využité rovnaké množstvá informácií zahrnuté do analyzovaných hodnôt, a preto majú rovnakú silu, pokiaľ ide o zistenie vzťahu medzi premennými uvažovanej populácie.

Výpočet Spearmanovho ρ je síce menej náročný (ak ide o väčší počet pozorovaní), ale prednosťou Kendallovho τ je, že je s ním spojená rýchlejšia konvergencia výberovej distribúcie k normálnej. Z toho sa ponúka možnosť použiť aproximáciu normálnym rozdelením aj pre menej rozsiahle výbery a tiež stanoviť interval spoľahlivosti pre signifikantné τ.

Obidva korelačné koeficienty budeme počítať na rovnakom príklade.

<u>Príklad:</u>
V tabuľke sú zaznamenané percentá ľudí s pracovnou zmluvou na dobu určitú (výber X) a percentá nezamestnaných ľudí (výber Y) v krajinách EÚ.

Štát	X	RX	Y	RY
Belgicko	8,60	12,5	7,50	20,0
Bulharsko	5,20	6,5	6,90	17,5
Česká republika	8,60	12,5	5,30	10,5
Dánsko	8,70	14,0	3,70	2,0
Nemecko	14,60	20,0	8,40	25,0
Estónsko	2,10	2,0	4,70	7,5
Írsko	7,30	10,5	4,50	6,0
Grécko	10,90	16,0	8,30	23,0
Španielsko	31,70	27,0	8,30	23,0
Francúzsko	14,40	19,0	8,30	23,0
Taliansko	13,20	17,5	6,10	13,5
Cyprus	13,20	17,5	3,90	3,0
Lotyšsko	4,20	4,0	6,00	12,0
Litva	3,50	3,0	4,30	4,0
Luxembursko	6,80	9,0	4,70	7,5
Maďarsko	7,30	10,5	7,40	19,0
Malta	5,20	6,5	6,40	15,5
Holandsko	18,10	23,0	3,20	1,0
Rakúsko	8,90	15,0	4,40	5,0
Polsko	28,20	26,0	9,60	26,0
Portugalsko	22,40	25,0	8,00	21,0
Rumunsko	1,60	1,0	6,40	15,5
Slovinsko	18,50	24,0	4,80	9,0
Slovensko	5,10	5,0	11,10	27,0
Fínsko	15,90	21,0	6,90	17,5
Švédsko	17,50	22,0	6,10	13,5
Anglicko	5,80	8,0	5,30	10,5

(Dáta-zdroj Eurostat, RX,RY-dopočítané poradia výberov)

Pomocou Spearmanovho a Kendallovho koeficientu zistite či existuje závislosť medzi výbermi. Testujte na hladine významnosti $\alpha = 0{,}01$.

Spearmanov koeficient (ρ, r_s)

Spearmanov koeficient používame na zistenie, či sú dve premenný X a Y korelované.

Pri výpočte Spearmanovho ρ dvoch nezávislých výberov $X = (X_1, \dots, X_n), Y = (Y_1, \dots, Y_n)$ rovnakej veľkosti n najskôr priradíme poradie každému prvku v rámci výberu $R_{X_i}(resp. R_{Y_i})$. Potom urobíme rozdiel poradových hodnôt $d_i = R_{X_i} - R_{Y_i}$, kde $i = 1, \dots, n$ a tento rozdiel umocníme na druhú d_i^2 (čiže poradie prvého prvku skupiny X odčítame od poradia prvého prvku skupiny Y => $d_1 = R_{X_1} - R_{Y_1}$ a umocníme na druhú d_1^2), čo opakujeme až po posledný prvok v skupinách.

Ak sa v skupinách neopakujú rovnaké hodnoty, nemusíme robiť priemer poradí a na výpočet Spearmanovho ρ použijeme vzorec:

$$\rho = r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

$$\sum_{i=1}^{n} d_i^2 - \mathit{je\ súčet\ všetkých\ rozdielov\ umocnených\ na\ druhú}$$

V praxi vždy nemôžeme použiť tento spôsob výpočtu. Prekážkou býva častý výskyt rovnakých dát alebo poradových hodnôt. Ak neberieme ohľad na rovnaké poradia, koeficient r_s môže byť neúmerne vysoký. To platí predovšetkým v situáciách, kde sú početnosti rovnakých poradí rozsiahle. V takých prípadoch použijeme upravený vzorec:

$$r_s = \frac{\frac{1}{6}(n^3 - n) - \sum_{i=1}^{n} d_i^2 - T_X - T_Y}{\sqrt{\left[\frac{1}{6}(n^3 - n - 2T_X)\right]\left[\frac{1}{6}(n^3 - n - 2T_Y)\right]}}$$

T_X $(resp.\ T_Y)$ – pre každú opakujúcu sa hodnotu v skupine X $(resp. Y)$ je jej počet opakovaní odčítaný od tretej mocniny jej počtu opakovaní a výsledky sú spočítané

$$T_X = \sum_{i=1}^{s_X} (t_{i(X)}^3 - t_{i(X)})$$

s_X $(resp.\ s_Y)$ $-$ *počet hodnôt, ktoré sa opakujú v skupine* $X(\ Y)$

$t_{i_{(X)}}$ $\left(resp. t_{i_{(Y)}}\right)$ $-$ *reprezentuje počet opakovaní hodnoty* X_j (Y_j) *v skupine* $X(Y)$

Keď má výber malý rozsah ($n \leq 30$), môžeme použiť štatistické tabuľky, v ktorých sú uvedené kritické hodnoty pre $4 \leq n \leq 30$ pri 5% a 1% hladine významnosti. Ide o štatistiku pre jednostranný test, keď môžeme vopred stanoviť, či bude korelačný koeficient medzi X a Y kladný alebo záporný. Nulovú hypotézu H_0 zamietame na danej hladine významnosti α, ak je absolútna hodnota vypočítaného ρ väčšia alebo rovná tabuľkovej kritickej hodnote.

Pre väčšie výbery $n > 30$ je možné rozdelenie pravdepodobnosti aproximovať t-rozdelením s $(n - 2)$ stupňami voľnosti. Potom má testovacia štatistika tvar:

$$t = r_s \sqrt{\frac{(n - 2)}{(1 - r_s^2)}}$$

Túto náhodnú veličinu používame aj vtedy, keď sa v premenných X a Y vyskytne väčší počet rovnakých meraní.

Nulovú hypotézu H_0 zamietame na hladine významnosti α, ak $|t| \geq t_{1-\alpha/2}$ (pre dvojstranný test), resp. $|t| \geq t_{1-\alpha}$ (pre jednostranný test), kde $t_{1-\alpha/2}$ (resp. $t_{1-\alpha}$) je kritická hodnota Studentovho t rozdelenia s $(n-2)$ stupňami voľnosti.

Pre veľmi veľké výbery $n > 100$ aproximujeme Spearmanov koeficient r_s normálnym rozdelením N(0,1):

$$z = r_s\sqrt{(n-1)}$$

Riešenie:

	Štát	RX	RY	d=RX-RY	d^2
1.	Belgicko	12,5	20,0	-7,5	56,3
2.	Bulharsko	6,5	17,5	-11,0	121,0
3.	Česká republika	12,5	10,5	2,0	4,0
4.	Dánsko	14,0	2,0	12,0	144,0
5.	Nemecko	20,0	25,0	-5,0	25,0
6.	Estónsko	2,0	7,5	-5,5	30,3
7.	Írsko	10,5	6,0	4,5	20,3
8.	Grécko	16,0	23,0	-7,0	49,0
9.	Španielsko	27,0	23,0	4,0	16,0
10.	Francúzsko	19,0	23,0	-4,0	16,0
11.	Taliansko	17,5	13,5	4,0	16,0
12.	Cyprus	17,5	3,0	14,5	210,3
13.	Lotyšsko	4,0	12,0	-8,0	64,0
14.	Litva	3,0	4,0	-1,0	1,0
15.	Luxembursko	9,0	7,5	1,5	2,3
16.	Maďarsko	10,5	19,0	-8,5	72,3
17.	Malta	6,5	15,5	-9,0	81,0
18.	Holandsko	23,0	1,0	22,0	484,0
19.	Rakúsko	15,0	5,0	10,0	100,0
20.	Polsko	26,0	26,0	0,0	0,0
21.	Portugalsko	25,0	21,0	4,0	16,0
22.	Rumunsko	1,0	15,5	-14,5	210,3
23.	Slovinsko	24,0	9,0	15,0	225,0
24.	Slovensko	5,0	27,0	-22,0	484,0
25.	Fínsko	21,0	17,5	3,5	12,3
26.	Švédsko	22,0	13,5	8,5	72,3
27.	Anglicko	8,0	10,5	-2,5	6,3
					2538,5

Pretože niektoré hodnoty sa opakujú, použijeme korekciu
Výber X:

Hodnota $8{,}6$ sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota $5{,}2$ sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota $7{,}3$ sa opakuje dvakrát $2^3 - 2 = 6$
Hodnota $13{,}2$ sa opakuje dvakrát $2^3 - 2 = 6$

$T_X = 24$

Výber Y:

Hodnota $4{,}7$ sa opakuje dvakrát	$2^3 - 2 = 6$
Hodnota $5{,}3$ sa opakuje dvakrát	$2^3 - 2 = 6$
Hodnota $6{,}1$ sa opakuje dvakrát	$2^3 - 2 = 6$
Hodnota $6{,}4$ sa opakuje dvakrát	$2^3 - 2 = 6$
Hodnota $6{,}9$ sa opakuje dvakrát	$2^3 - 2 = 6$
Hodnota $8{,}3$ sa opakuje trikrát	$3^3 - 3 = 24$
	$T_Y = 54$

$$r_s = \frac{\frac{1}{6}(26^3 - 26) - 2538{,}5 - 24 - 54}{\sqrt{\left[\frac{1}{6}(26^3 - 26 - 48)\right]\left[\frac{1}{6}(26^3 - 26 - 108)\right]}} = 0.224$$

$$t = 0{,}224\sqrt{\frac{(26-2)}{(1 - 0.050176)}} = 1.14694$$

Keďže $1.14694 < 2.48511 = t_{0{,}99}(25)$ **nezamietame** H_0 na danej hladine významnosti $\alpha = 0{,}01$, t.j. medzi percentami ľudí s pracovnou zmluvou na dobu určitú a nezamestnaných ľudí v krajinách EÚ nie je štatisticky významná závislosť.

Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:

```
SpearmanovKoeficient[X,Y,α]
X – prvý výber (zadáva sa ako list)
Y – druhý výber (zadáva sa ako list)
α – hladina významnosti
```

Výstup procedúry na počítanom príklade:

Spearmanov koeficient :	0.223582	
Pocet merani vo vybere :	27	
	Obojstranny test	Jednostranny test
Hladina vyznamnosti α	0.01	
p value	0.262264	0.131132
Vyhodnotenie na zaklade p value	Nezamietam H_0	Nezamietam H_0
Vypocitana statistika t	1.14694	
Kriticka hodnota Studentovho t roz.	2.78744	2.48511
Vyhodnotenie na zaklade KH	Nezamietam H_0	Nezamietam H_0

Kendallov koeficient (τ)

Meria silu závislosti medzi dvoma poradovými premennými.

Pri výpočte Kendallovho τ dvoch nezávislých výberov $X = (X_1, \dots, X_n), Y = (Y_1, \dots, Y_n)$ rovnakej veľkosti n najskôr každému prvku v rámci skupiny priradíme poradie (R_{X_i}, R_{Y_i}). Poradie prvkov v skupinách je ľubovoľné a preto je nutné preorganizovať prvky a to tak, že zoradíme poradia prvkov jednej skupiny vzostupne (napr. podľa skupiny X) a k nim priradíme prislúchajúce poradia skupiny Y .

Prislúchajúce poradie znamená, že ak k-ty prvok skupiny X má poradie R_{X_k} priradíme k nemu poradie R_{Y_k} k-teho prvku skupiny Y. Ďalej vypočítame počet súhlasných a nesúhlasných párov nasledovne:
ak platí $R_{X_i}>R_{X_j}$ a $R_{Y_i}>R_{Y_j}$ alebo $R_{X_i}<R_{X_j}$ a $R_{Y_i}<R_{Y_j}$, potom sa pár nazýva súhlasný (concordant)
ak platí $R_{X_i}>R_{X_j}$ a $R_{Y_i}<R_{Y_j}$ alebo $R_{X_i}<R_{X_j}$ a $R_{Y_i}>R_{Y_j}$, potom sa pár nazýva nesúhlasný (discordant).
ak platí $R_{X_i}=R_{X_j}$ a/alebo $R_{Y_i}=R_{Y_j}$, potom sa pár nazýva nerozhodný (tie),
kde (R_{X_i}, R_{Y_i}) je dvojica poradí a (R_{X_j} , R_{Y_j}) je dvojica nasledujúcich poradí, pričom i, j = 1, ..., n a i ≠ j.

Ak sa vo výbere nevyskytujú zhodné údaje, tak pre výpočet Kendallovho koeficientu použijeme vzorec:

$$\tau = \frac{n_C - n_D}{\left[\frac{n(n-1)}{2}\right]}$$

n_C – *je počet súhlasných párov (concordant pair)*
n_D – *je počet nesúhlasných párov (disconcordant pair)*
$n(n-1)/2$ – *počet všetkých možných párov*

Kendallovo τ vyjadruje rozdiel medzi pravdepodobnosťou, že hodnoty dvoch premenných sú v rovnakom poradí oproti pravdepodobnosti, že hodnoty nie sú v rovnakom poradí. Ak sa vyskytnú v skupinách X a Y totožné údaje, dostanú priemerné poradové čísla a pre výpočet koeficientu τ použijeme upravený vzorec. Upravený vzorec pre výpočet τ použijeme v prípade väčšieho výskytu nerozhodných párov.

$$\tau = \frac{2(n_C - n_D)}{\sqrt{n(n-1)T_X}\sqrt{n(n-1)T_Y}}$$

T_X $(resp.\ T_Y)$ – pre každú opakujúcu sa hodnotu v skupine X $(resp. Y)$ je jej počet opakovaní odčítaný od druhej mocniny jej počtu opakovaní a výsledky sú sčítané

$$T_X = \sum_{i=1}^{s_X} (t_{i(X)}^2 - t_{i(X)})$$

s_X $(resp.\ s_Y)$ – *počet hodnôt, ktoré sa opakujú v skupine* $X(Y)$
$t_{i(X)}\left(resp.\, t_{i(Y)}\right)$ – *reprezentuje počet opakovaní hodnoty* X_j (Y_j) *v skupine* $X(Y)$

Pre výbery malých rozsahov ($n \leq 50$) bolo odvodené presné rozdelenie pravdepodobnosti s $E(\tau) = 0, D(\tau) = \frac{2(2n+5)}{9n(n-1)}$. H_0 zamietame na danej hladine významnosti α, ak je absolútna hodnota vypočítaného τ väčšia alebo rovná tabuľkovej kritickej hodnote.

S narastajúcim rozsahom výberu (n > 50) koeficient τ má za platnosti nulovej hypotézy asymptoticky normálne rozdelenie, a preto je možné k testovaniu jeho významnosti použiť štatistiku z:

$$z = \frac{3\tau\sqrt{n(n-1)}}{\sqrt{2(2n+5)}}$$

Pre veľké výbery nulovú hypotézu H_0 zamietame na hladine významnosti α, ak $|z| \geq u_{1-\alpha/2}$ (pre dvojstranný test), resp. $|z| \geq u_{1-\alpha}$ (pre jednostranný test), kde $u_{1-\alpha/2}$ (resp. $u_{1-\alpha}$) je kvantil N(0, 1).

Riešenie:

RX	RY																	
1,0	15,5	15,5																
2,0	7,5	D	7,5															
3,0	4,0	D	D	4,0														
4,0	12,0	D	C	C	12,0													
5,0	27,0	C	C	C	C	27,0												
6,5	17,5	C	C	C	C	D	17,5											
6,5	15,5	T	C	C	C	D	T	15,5										
8,0	10,5	D	C	C	D	D	D	D	10,5									
9,0	7,5	D	T	C	D	D	D	D	D	7,5								
10,5	6,0	D	D	C	D	D	D	D	D	D	6,0							
10,5	19,0	C	C	C	C	D	C	C	C	C	T	19,0						
12,5	20,0	C	C	C	C	D	C	C	C	C	C	C						
12,5	10,5	D	C	C	D	D	D	D	T	C	C	D						
14,0	2,0	D	D	D	D	D	D	D	D	D	D	D						
15,0	5,0	D	D	C	D	D	D	D	D	D	D	D						
16,0	23,0	C	C	C	C	D	C	C	C	C	C	C						
17,5	13,5	D	C	C	C	D	D	D	C	C	C	D						
17,5	3,0	D	D	D	D	D	D	D	D	D	D	D						
19,0	23,0	C	C	C	C	D	C	C	C	C	C	C	Analogicky postupujeme až po					
20,0	25,0	C	C	C	C	D	C	C	C	C	C	C	predposledný prvok					
21,0	17,5	C	C	C	C	D	T	C	C	C	C	D						
22,0	13,5	D	C	C	C	D	D	D	C	C	C	D						
23,0	1,0	D	D	D	D	D	D	D	D	D	D	D						
24,0	9,0	D	C	C	D	D	D	D	D	C	C	D						
25,0	21,0	C	C	C	C	D	C	C	C	C	C	C						
26,0	26,0	C	C	C	C	D	C	C	C	C	C	C						
27,0	23,0	C	C	C	C	D	C	C	C	C	C	C						
Pocet c parov		11	18	21	14	0	8	9	11	13	12	7	...				nc=	195
Pocet d parov		14	6	3	9	22	11	11	7	5	4	9	...				nd=	144

Pretože niektoré hodnoty sa opakujú, použijeme korekciu

Výber X:

Hodnota $8{,}6$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $5{,}2$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $7{,}3$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $13{,}2$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

$$T_X = 8$$

Výber Y:

Hodnota $4{,}7$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $5{,}3$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $6{,}1$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $6{,}4$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $6{,}9$ sa opakuje dvakrát $\quad 2^2 - 2 = 2$

Hodnota $8{,}3$ sa opakuje trikrát $\quad 3^2 - 3 = 6$

$$T_Y = 16$$

$$\tau = \frac{2 * (195 - 144)}{\sqrt{27 * (27 - 1) * 8}\sqrt{27 * (27 - 1) * 16}} = 0{,}1478$$

$$z = \frac{3 * 0{,}1478 * \sqrt{27 * (27 - 1)}}{\sqrt{2 * (2 * 27 + 5)}} = 1{,}0817$$

Keďže $1{,}0817 < 2{,}57583 = z_{0,99}$ **nezamietame** H_0 na danej hladine významnosti $\alpha = 0{,}01$, t.j. medzi percentami ľudí s pracovnou zmluvou na dobu určitú a nezamestnaných ľudí v krajinách EÚ nie je štatisticky významná závislosť.

<u>Volanie naprogramovanej procedúry v Mathematice 6 a jej vstupné parametre:</u>

```
KendallovKoeficient[X,Y,α]
X – prvý výber (zadáva sa ako list)
Y – druhý výber (zadáva sa ako list)
α – hladina významnosti
```

<u>Výstup procedúry na počítanom príklade:</u>

Kendallov koeficient :	0.147829	
Pocet merani vo vybere :	27	
	Obojstranny test	Jednostranny test
Hladina vyznamnosti α	0.01	
p value	0.279385	0.139693
Vyhodnotenie na zaklade p value	Nezamietam H_0	Nezamietam H_0
Vypocitana z statistika	1.0817	
Kvantil N(0,1)	2.57583	2.32635
Vyhodnotenie na zaklade kvantilu	Nezamietam H_0	Nezamietam H_0

Kendallov koeficient je po všetkých stránkach lepší ako často používaný **Spearmanov koeficient**. Spearmanov koeficient je postačujúci na testovanie nulovej hypotézy nezávislosti dvoch premenných. V prípade zamietnutia nulovej hypotézy sa však nedá interpretovať. Kendallov koeficient je lepší aj ako test nezávislosti, pretože je citlivý na niektoré druhy závislostí, ktoré nemôžu byť zachytené Spearmanovým koeficientom. Kendallov koeficient na rozdiel od Spearmanovho vyjadruje silu závislosti skúmaných premenných.

Potvrdilo sa nám, že Kendallov koeficient dáva menšie hodnoty ako Spearmanov koeficient: $0{,}147829 < 0{,}223582$.

Záver

Výpočtový systém Mathematica je veľmi rozsiahly matematický systém. Dokáže vypočítať veľké množstvo funkcií od elementárnych až po špeciálne, kresliť grafy a taktiež obsahuje veľké množstvo knižníc pre rôzne druhy výpočtov. Štatistická knižnica neobsahuje funkcie pre testovanie hypotéz pomocou neparametrických štatistických metód a preto bolo hlavným cieľom práce doprogramovať tieto procedúry a doplniť ich do systému Mathematica 6.

Správne fungovanie naprogramovaných procedúr sa potvrdilo na ručne riešených príkladoch, keď nám vyšli rovnaké výsledky. Každá procedúra bola testovaná aj na iných príkladoch z rôznych oblastí a porovnaná s riešením v komerčných programoch StatsDirect a Statistica 8. Štatistická knižnica systému Mathematica môže byť doplnená o tieto metódy. Ich zdrojové kódy sa nachádzajú v prílohách.

Procedúry môžu byť taktiež použité pri výučbe štatistiky v prostredí systému Mathematica.

Použitá literatúra

[1] David J. Sheskin: Handbook of parametric and nonparametric statistical procedures 3. ed., New York 2003

[2] Papica J. : Neparametrická statistika, Univ. Palackého, Olomouc 1992

[3] Terry Sincich: Business statistics by example, Dellen Publ., San Francisco 1989

[4] Richterová K. a kolektív: Marketingový výskum, Ekonóm, Bratislava 2007

[5] http://rimarcik.com/

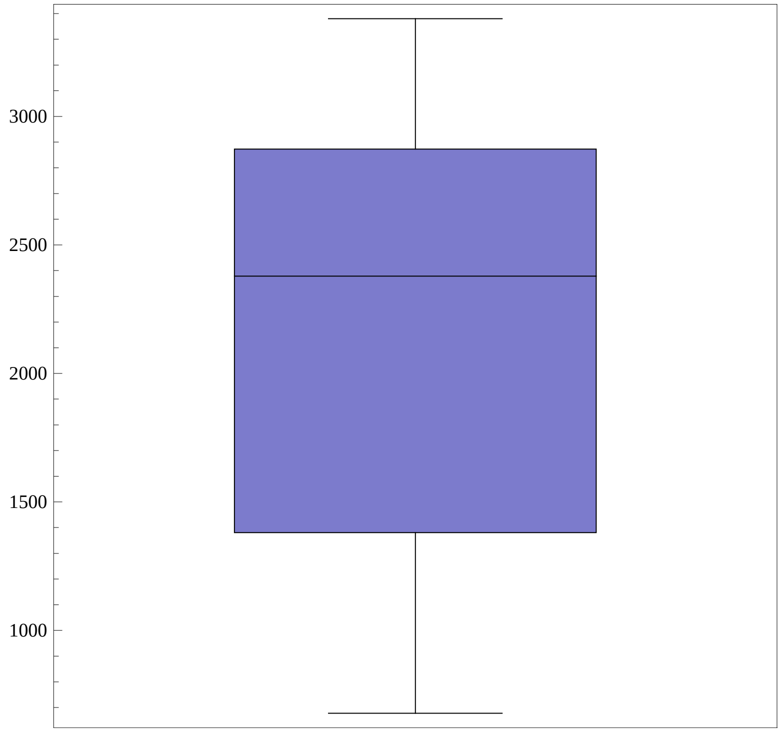
[6] Anne E. Magurran: Measuring biological diversity, Wiley-Blackwell, 2004

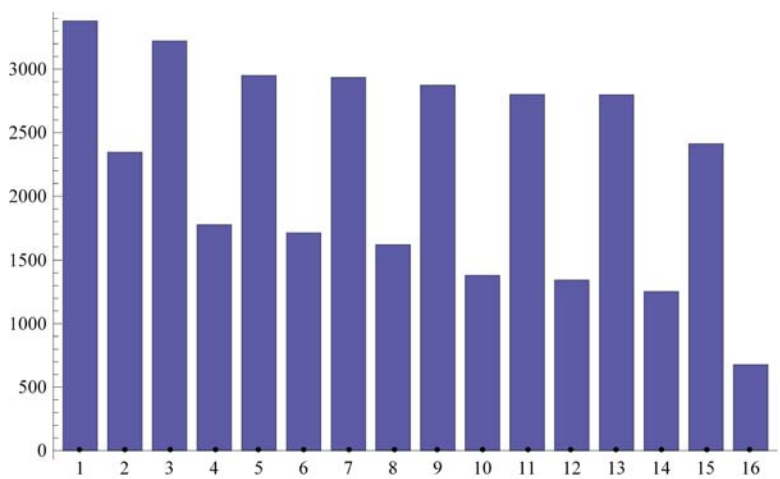
Prílohy

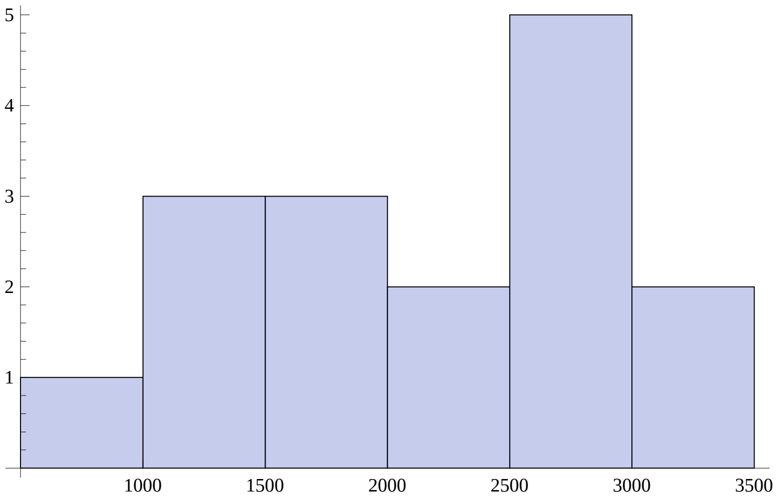
Zdrojový kód Znamienkového testu

```
<< "StatisticalPlots`"
<< "Histograms`"
<< "HypothesisTesting`"
ZnamienkovyTest[data_List, α_, X0_] :=
 Module[{median, rozdiel, pocet, Sz = 0, Sk = 0, Sn = 0, kv, pvalJS, pvalOS, kvOS, kvJS},
  pocet = Length[data];
  median = Median[data];
  rozdiel = data - X0;
  Table[
   If[rozdiel[[i]] > 0, Sk = Sk + 1, If[rozdiel[[i]] < 0, Sz = Sz + 1, Sn = Sn + 1]], {i, 1, pocet}];
  Print[BoxWhiskerPlot[data]];
  Print[BarChart[data]];
  Print[Histogram[data]];
  z[t_] := (2 t - pocet)/Sqrt[pocet];
  pvalJS = NormalPValue[z[Sk]];
  pvalJS = OneSidedPValue /. pvalJS;
  pvalOS = NormalPValue[z[Sk], TwoSided → True];
  pvalOS = TwoSidedPValue /. pvalOS;
  kvOS = Quantile[NormalDistribution[0, 1], 1 - α/2];
  kvJS = Quantile[NormalDistribution[0, 1], 1 - α];
  Print["Dlzka vyberu je: ", pocet];
  Print["Pocet kladnych hodnot je : ", Sk];
  Print[Grid[{{"Median je : ", median, SpanFromLeft},
     {"", "Obojstranny test", "Jednostranny test"}, {"Hladina vyznamnosti α ", α,
      SpanFromLeft}, {"p value ", pvalOS, pvalJS}, {"Vyhodnotenie na zaklade p value ",
      If[pvalOS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]],
      If[pvalJS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]},
     {"Vypocitana statistika z ", z[Sk] // N, SpanFromLeft}, {"Kvantil N(0,1) ", kvOS, kvJS},
     {"Vyhodnotenie na zaklade kvantilu ", If[z[Sk] > kvOS, Text[Style["Zamietam H0", Red]],
       Text[Style["Nezamietam H0", Green]]], If[z[Sk] > kvJS, Text[Style["Zamietam H0", Red]],
       Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
 ]

PriemernePlaty = {3380, 2346, 3222, 1775, 2951,
   1710, 2935, 1617, 2873, 1381, 2801, 1344, 2799, 1253, 2412, 678};
ZnamienkovyTest[PriemernePlaty, 0.05, 678]
```


3000
2500
2000
1500
1000


3000
2500
2000
1500
1000
500
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16



Dlzka vyberu je: 16

Pocet kladnych hodnot je : 15

Median je :	2379	
	Obojstranny test	Jednostranny test
Hladina vyznamnosti α	0.05	
p value	0.000465258	0.000232629
Vyhodnotenie na zaklade p value	Zamietam H_0	Zamietam H_0
Vypocitana statistika z	3.5	
Kvantil N(0,1)	1.95996	1.64485
Vyhodnotenie na zaklade kvantilu	Zamietam H_0	Zamietam H_0

Zdrojový kód Kolmogorovho-Smirnovovho testu

```
<< "HypothesisTesting`"
KolmogorovSmirnovTest[data1_List, data2_List, α_] :=
 Module[{X, Y},
  X = Sort[data1];
  Y = Sort[data2];
  pocdata1 = Length[data1];
  pocdata2 = Length[data2];
  P = Sort[Join[X, Y]];
  GX[0] = FX[0] = 0;
  For[i = 1, i <= Length[P], i++,
   {
    f = Flatten[Position[X, P[[i]]]];
    If[Length[f] ≠ 0, FX[i] = Take[f, -1] / Length[X], FX[i] = FX[i - 1]];
   }
  ];
  For[i = 1, i <= Length[P], i++,
   {
    g = Flatten[Position[Y, P[[i]]]];
    If[Length[g] ≠ 0, GX[i] = Take[g, -1] / Length[Y], GX[i] = GX[i - 1]];
   }
  ];
  MO = Table[FX[i] - GX[i], {i, 1, Length[P]}];
  ot = Max[Abs[MO]] // N;
  khJS = 4 * (ot^2) * (Length[X] * Length[Y] / (Length[X] + Length[Y]));
  chi = Quantile[ChiSquareDistribution[2], 1 - α/2];
  khOS = Sqrt[(Length[X] + Length[Y]) / (2 * Length[X] * Length[Y]) * (Log[2 / α])] ;
  Print["n1 : ", pocdata1];
  Print["n2 : ", pocdata2];
  Print["Priemer 1 : ", Mean[data1] // N];
  Print["Priemer 2 : ", Mean[data2] // N];
  Print[
   Grid[{{"", "Obojstranny test"}, {"Hladina vyznamnosti α ", α}, {"Vypocitane D ", ot // N},
     {"Kriticka hodnota D* ", khOS}, {"Vyhodnotenie na zaklade KH ", If[ot > khOS,
       Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
 ]

(*1 vs. 2 dovod*)
KolmogorovSmirnovTest[
 {16, 6, 32, 31, 15, 12, 14, 20, 14, 33, 8, 9, 26, 3, 12, 20, 8, 8, 9, 13, 24, 29, 13},
 {5, 1, 7, 46, 15, 20, 20, 33, 17, 23, 1, 17, 25, 15, 27, 19, 5, 45, 11, 15, 6, 42, 21}, 0.05]
```

n1 : 23

n2 : 23

Priemer 1 : 16.3043

Priemer 2 : 18.9565

	Obojstranny test
Hladina vyznamnosti α	0.05
Vypocitane D	0.26087
Kriticka hodnota D^*	0.400482
Vyhodnotenie na zaklade KH	Nezamietam H_0

```
(*1 vs. 3 dovod*)
KolmogorovSmirnovTest[
 {16, 6, 32, 31, 15, 12, 14, 20, 14, 33, 8, 9, 26, 3, 12, 20, 8, 8, 9, 13, 24, 29, 13},
 {37, 57, 51, 48, 28, 45, 56, 45, 52, 49, 69, 53, 49, 33, 40, 49, 45, 72, 25, 61, 40, 42, 30}, 0.05]
```

n1 : 23

n2 : 23

Priemer 1 : 16.3043

Priemer 2 : 46.7826

	Obojstranny test
Hladina vyznamnosti α	0.05
Vypocitane D	0.826087
Kriticka hodnota D^*	0.400482
Vyhodnotenie na zaklade KH	Zamietam H_0

Zdrojový kód Wald-Wolfowitzovho testu

```
<< "HypothesisTesting`"
WaldWolfowitzovTest[data1_List, data2_List, α_] :=
 Module[{pocdata1, pocdata2, pomdata1, pomdata2, pomdata, iteracie, Mr, Sr, p1, p, z},
  pocdata1 = Length[data1];
  pocdata2 = Length[data2];
  pomdata1 = Table[Tuples[{data1[[i]], 1}, 1], {i, 1, Length[data1]}];
  pomdata2 = Table[Tuples[{data2[[i]], 2}, 1], {i, 1, Length[data2]}];
  pomdata = Flatten[Sort[Partition[Flatten[Join[pomdata1, pomdata2]], 2]]];
  iteracie = Length[Split[Drop[pomdata, {1, -1, 2}]]];
  Mr = (2 pocdata1 pocdata2)/(pocdata1 + pocdata2) + 1;
  Sr = Sqrt[(2 pocdata1 pocdata2 (2 pocdata1 pocdata2 - pocdata1 - pocdata2))/((pocdata1 + pocdata2)^2 (pocdata1 + pocdata2 - 1))];
  z = (Abs[iteracie - Mr] - 0.5)/Sr // N;
  Print["n1 : ", pocdata1];
  Print["n2 : ", pocdata2];
  Print["Priemer 1 : ", Mean[data1] // N];
  Print["Priemer 2 : ", Mean[data2] // N];
  pvalJS = NormalPValue[z];
  pvalJS = OneSidedPValue /. pvalJS;
  pvalOS = NormalPValue[z, TwoSided → True];
  pvalOS = TwoSidedPValue /. pvalOS;
  kvOS = Quantile[NormalDistribution[0, 1], 1 - α/2];
  kvJS = Quantile[NormalDistribution[0, 1], 1 - α];
  Print[Grid[{{"Pocet iteracii : ", iteracie}, {"", "Obojstranny test"},
     {"Hladina vyznamnosti α ", α}, {"p value ", pvalOS}, {"Vyhodnotenie na zaklade p value ",
      If[pvalOS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]},
     {"Vypocitana z statistika ", z // N}, {"Kvantil N(0,1) ", kvOS},
     {"Vyhodnotenie na zaklade kvantilu ", If[z > kvOS, Text[Style["Zamietam H0", Red]],
       Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
 ]

(*1 vs. 2 dovod*)
WaldWolfowitzovTest[
 {16, 6, 32, 31, 15, 12, 14, 20, 14, 33, 8, 9, 26, 3, 12, 20, 8, 8, 9, 13, 24, 29, 13},
 {5, 1, 7, 46, 15, 20, 20, 33, 17, 23, 1, 17, 25, 15, 27, 19, 5, 45, 11, 15, 6, 42, 21}, 0.05]
```

n1 : 23

n2 : 23

Priemer 1 : 16.3043

Priemer 2 : 18.9565

Pocet iteracii :	19
	Obojstranny test
Hladina vyznamnosti α	0.05
p value	0.179605
Vyhodnotenie na zaklade p value	Nezamietam H_0
Vypocitana z statistika	1.34197
Kvantil N(0,1)	1.95996
Vyhodnotenie na zaklade kvantilu	Nezamietam H_0

```
(*1 vs. 3 dovod*)
WaldWolfowitzovTest[
 {16, 6, 32, 31, 15, 12, 14, 20, 14, 33, 8, 9, 26, 3, 12, 20, 8, 8, 9, 13, 24, 29, 13},
 {37, 57, 51, 48, 28, 45, 56, 45, 52, 49, 69, 53, 49, 33, 40, 49, 45, 72, 25, 61, 40, 42, 30}, 0.05]
```

n1 : 23

n2 : 23

Priemer 1 : 16.3043

Priemer 2 : 46.7826

Pocet iteracii :	8
	Obojstranny test
Hladina vyznamnosti α	0.05
p value	3.7942×10^{-6}
Vyhodnotenie na zaklade p value	Zamietam H_0
Vypocitana z statistika	4.62235
Kvantil N(0,1)	1.95996
Vyhodnotenie na zaklade kvantilu	Zamietam H_0

Zdrojový kód Kruskal-Wallisovho testu

```
<< "HypothesisTesting`"
KruskalWallisovTest[data_List, α_] :=
 Module[{pocsub, pocdat, pomdata, pocetnost,
   sumkor = 0, korekcia, poradie, poradiedata, H, Hkor, chi},
  pocsub = Length[data];
  Table[Print["Pocet prvkov v skupine ",
    i, " : ", pocdata[i] = Length[data[[i]]]], {i, 1, pocsub}];
```

$$\texttt{pocdat} = \sum_{i=1}^{\texttt{pocsub}} \texttt{pocdata[i]};$$

```
  pomdata = Sort[Flatten[Table[data[[j, i]], {j, 1, pocsub}, {i, 1, pocdata[j]}]]];
  pocetnost = Tally[pomdata];
```

$$\texttt{Table}\left[\texttt{If}\left[\texttt{pocetnost[[i, 2]]} \neq 1,\ \texttt{sumkor} = \texttt{sumkor} + \left(\texttt{pocetnost[[i, 2]]}^3 - \texttt{pocetnost[[i, 2]]}\right)\right],\ \{i, 1, \texttt{Length[pocetnost]}\}\right];$$

$$\texttt{korekcia} = 1 - \left(\frac{\texttt{sumkor}}{\texttt{pocdat}^3 - \texttt{pocdat}}\right);$$

```
  poradie = Flatten[Table[Total[Position[pomdata, pocetnost[[i, 1]]]] / pocetnost[[i, 2]],
     {i, 1, Length[pocetnost]}]];
  poradiedata = Partition[Flatten[Table[Tuples[{pocetnost[[i, 1]], poradie[[i]]}, 1],
      {i, 1, Length[poradie]}]], 2];
  Table[pocporadiadata[i] = 0, {i, 1, pocsub}]
   For[k = 1, k ≤ Length[poradie], k++,
    Table[If[data[[i, j]] == poradiedata[[k, 1]], pocporadiadata[i] =
        pocporadiadata[i] + poradiedata[[k, 2]]], {i, 1, pocsub}, {j, 1, Length[data[[i]]]}];
   ];
  Table[Print["Sucet poradi v stlpci ", i, " : ", pocporadiadata[i] // N], {i, 1, pocsub}];
```

$$H = \frac{12}{\texttt{pocdat} * (\texttt{pocdat} + 1)} * \left(\sum_{i=1}^{\texttt{pocsub}} \frac{\texttt{pocporadiadata[i]}^2}{\texttt{pocdata[i]}}\right) - 3 * (\texttt{pocdat} + 1);$$

$$\texttt{Hkor} = \frac{H}{\texttt{korekcia}};$$

```
  pvalJS = ChiSquarePValue[Hkor, (pocsub - 1)];
  pvalJS = OneSidedPValue /. pvalJS;
  chi = Quantile[ChiSquareDistribution[(pocsub - 1)], 1 - α];
  Print[
   Grid[{{"Hladina vyznamnosti α ", α}, {"H", H // N}, {"Korekcny koeficient", korekcia // N},
     {"H korekcia", Hkor // N}, {"p value", pvalJS}, {"Vyhodnotenie na zaklade p value",
      If[pvalJS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]},
     {"Stupne volnosti", (pocsub - 1)}, {"Kriticka hodnota χ2 rozdelenia", chi},
     {"Vyhodnotenie na zaklade KH", If[H > chi, Text[Style["Zamietam H0", Red]],
       Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
 ]
```

```
data = {{16, 6, 32, 31, 15, 12, 14, 20, 14, 33, 8, 9, 26, 3, 12, 20, 8, 8, 9, 13, 24, 29, 13},
   {5, 1, 7, 46, 15, 20, 20, 33, 17, 23, 1, 17, 25, 15, 27, 19, 5, 45, 11, 15, 6, 42, 21},
   {37, 57, 51, 48, 28, 45, 56, 45, 52, 49, 69, 53, 49, 33, 40, 49, 45, 72, 25, 61, 40, 42, 30},
   {35, 5, 34, 58, 15, 14, 20, 41, 13, 52, 18, 7, 49, 7, 10, 15, 29, 51, 35, 35, 18, 25, 19},
   {27, 9, 32, 51, 7, 9, 20, 37, 12, 43, 12, 3, 34, 4, 8, 14, 26, 47, 26, 32, 15, 26, 13},
   {32, 10, 31, 62, 20, 30, 26, 37, 41, 48, 7, 16, 28, 34, 17, 17, 23, 62, 20, 39, 19, 40, 13}};
KruskalWallisovTest[data, 0.05]
```

Pocet prvkov v skupine 1 : 23

Pocet prvkov v skupine 2 : 23

Pocet prvkov v skupine 3 : 23

Pocet prvkov v skupine 4 : 23

Pocet prvkov v skupine 5 : 23

Pocet prvkov v skupine 6 : 23

Sucet poradi v stlpci 1 : 1015.5

Sucet poradi v stlpci 2 : 1192.5

Sucet poradi v stlpci 3 : 2640.5

Sucet poradi v stlpci 4 : 1603.5

Sucet poradi v stlpci 5 : 1350.5

Sucet poradi v stlpci 6 : 1788.5

Hladina vyznamnosti α	0.05
H	45.9158
Korekcny koeficient	0.99924
H korekcia	45.9508
p value	9.2942×10^{-9}
Vyhodnotenie na zaklade p value	Zamietam H_0
Stupne volnosti	5
Kriticka hodnota χ^2 rozdelenia	11.0705
Vyhodnotenie na zaklade KH	Zamietam H_0

Zdrojový kód Friedmanovho testu

```
<< "HypothesisTesting`"
FriedmanovTest[data_List, α_] :=
 Module[{pocsub, dlzka, pocdata, Fr, sumkor = 0, korekcia, Frkor, chi},
  pocsub = Length[data];
  dlzka = Length[data[[1]]];
  Table[pomdata[i] = Sort[data[[i]]], {i, 1, Length[data]}];
  Table[pocetnost[i] = Tally[pomdata[i]], {i, 1, Length[data]}];
  For[j = 1, j ≤ pocsub, j ++,
   Table[If[pocetnost[j][[i, 2]] ≠ 1, sumkor =
       sumkor + (pocetnost[j][[i, 2]]^3 - pocetnost[j][[i, 2]])], {i, 1, Length[pocetnost[j]]}];
  ];
  korekcia = 1 - (sumkor / (pocsub (dlzka^3 - dlzka)));
  For[i = 1, i <= Length[data], i ++,
   poradie[i] =
     Partition[Flatten[Table[Tuples[{pomdata[i][[j]], j}, 1], {j, 1, Length[data[[i]]]}]], 2];
  ];
  For[j = 1, j <= Length[data], j ++,
   poradie[j] = Flatten[Table[Total[Position[pomdata[j], pocetnost[j][[i, 1]]]] /
       pocetnost[j][[i, 2]], {i, 1, Length[pocetnost[j]]}]];
  ];
  For[j = 1, j <= Length[data], j ++,
   poradiedata[j] = Partition[Flatten[Table[
       Tuples[{pocetnost[j][[i, 1]], poradie[j][[i]]}, 1], {i, 1, Length[poradie[j]]}]], 2];
  ];
  For[k = 1, k <= Length[data], k ++,
   {For[j = 1, j <= Length[poradiedata[k]], j ++,
     Table[If[poradiedata[k][[j, 1]] == data[[k, i]],
        uspordata[i] = poradiedata[k][[j]]], {i, 1, Length[data[[k]]]}];
    ];
    poradovedata[k] = Table[uspordata[i], {i, 1, Length[data[[k]]]}];}
  ];
  For[j = 1, j ≤ dlzka, j ++,
   sucetstlpcov[j] = Sum[poradovedata[i][[j, 2]], {i, 1, pocsub}];
  ];
  Table[Print["Sucet poradi v stlpci ", i, " : ", sucetstlpcov[i] // N], {i, 1, dlzka}];
  Fr = 12 / (pocsub * dlzka * (dlzka + 1)) * (Sum[sucetstlpcov[i]^2, {i, 1, dlzka}]) - (3 * pocsub * (dlzka + 1));
  Frkor = Fr / korekcia;
  pvalJS = ChiSquarePValue[Frkor, (dlzka - 1)];
  pvalJS = OneSidedPValue /. pvalJS;
  chi = Quantile[ChiSquareDistribution[(dlzka - 1)], 1 - α];
  Print[Grid[{{"Hladina vyznamnosti α ", α},
    {"Fr", Fr // N}, {"Korekcny koeficient", korekcia // N}, {"Fr korekcia",
```

```
        Frkor // N}, {"p value", pvalJS}, {"Vyhodnotenie na zaklade p value",
        If[pvalJS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]},
       {"Stupne volnosti", (dlzka - 1)}, {"Kriticka hodnota χ^2 rozdelenia", chi},
       {"Vyhodnotenie na zaklade KH", If[Frkor > chi, Text[Style["Zamietam H0", Red]],
         Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
  ]

data = {{2.6, 1.6, 2.1, 3}, {0.9 , 1.7,  1.9, 1.7}, {1.8, 1.9, 1.8, 2.3},
   {3, 4.1, 4.4, 6.7}, {2.3, 2.2, 2.7, 2.9}, {3, 3.5, 3.3, 3}, {3.1, 3.4, 3.6, 2.8},
   {2.3, 1.9, 1.9, 1.6}, {1.9, 2, 2.2, 2.2}, {6.2, 6.9, 6.6, 10.1}, {1.2, 2.7, 3.8, 5.8},
   {3.2, 3.8, 3, 2.7}, {6.8, 3.5, 4, 7.9}, {2.7, 2.5, 2.6, 0.7}, {1.4, 1.5, 1.7, 1.6},
   {2, 2.1, 1.7, 2.2}, {3.6, 2.2, 1.3, 2.6}, {2.5, 2.1, 3, 2.4}, {11.9, 9.1, 6.6, 4.9},
   {3.7, 2.5, 2.5, 3.8}, {7.5, 2.8, 4.3, 1.9}, {1, 0.8, 1.5, 1.7}, {1.3, 2.1, 2.3, 2.3}};
FriedmanovTest[
 data,
 0.05]
```

Sucet poradi v stlpci 1 : 53.

Sucet poradi v stlpci 2 : 50.5

Sucet poradi v stlpci 3 : 61.5

Sucet poradi v stlpci 4 : 65.

Hladina vyznamnosti α	0.05
Fr	3.6913
Korekcny koeficient	0.969565
Fr korekcia	3.80717
p value	0.283053
Vyhodnotenie na zaklade p value	Nezamietam H_0
Stupne volnosti	3
Kriticka hodnota χ^2 rozdelenia	7.81473
Vyhodnotenie na zaklade KH	Nezamietam H_0

Zdrojový kód testu významnosti Spearmanovho koeficientu

```
<< "HypothesisTesting`"
SpearmanovKoeficient[X_List, Y_List, α_] :=
 Module[{dlzka, rozdiel, mocnina, sucet, rs, t, z,
   rozporad, sumpor, pomU = 0, pomT = 0, U, T, pvalJS, pvalOS, kvOS, kvJS},
  dlzka = Length[X];
  data[1] = X; data[2] = Y;
  Table[pomdata[i] = Sort[data[i]], {i, 1, 2}];
  Table[pocetnost[i] = Tally[pomdata[i]], {i, 1, 2}];
  For[i = 1, i ≤ 2, i++,
   poradie[i] = Partition[Flatten[Table[Tuples[{pomdata[i][[j]], j}, 1], {j, 1, dlzka}]], 2];
  ];
  For[j = 1, j ≤ 2, j++,
   poradie[j] = Flatten[Table[Total[Position[pomdata[j], pocetnost[j][[i, 1]]]] /
        pocetnost[j][[i, 2]], {i, 1, Length[pocetnost[j]]}]];
  ];
  For[j = 1, j ≤ 2, j++,
   poradiedata[j] = Partition[Flatten[Table[
        Tuples[{pocetnost[j][[i, 1]], poradie[j][[i]]}, 1], {i, 1, Length[poradie[j]]}]], 2];
  ];
  For[k = 1, k ≤ 2, k++,
   {For[j = 1, j <= Length[poradiedata[k]], j++,
     Table[If[poradiedata[k][[j, 1]] == data[k][[i]],
        uspordata[i] = poradiedata[k][[j]]], {i, 1, dlzka}];
    ];
    poradovedata[k] = Table[uspordata[i], {i, 1, dlzka}];}
  ];
  rozporad = poradovedata[1] - poradovedata[2];
  sumpor = Sum[rozporad[[i, 2]]^2, {i, 1, Length[rozporad]}];
  Table[
   If[pocetnost[1][[i, 2]] ≠ 1, pomT = pomT + (pocetnost[1][[i, 2]]^3 - pocetnost[1][[i, 2]])],
   {i, 1, Length[pocetnost[1]]}];
  T = 1/12 * pomT;
  Table[
   If[pocetnost[2][[i, 2]] ≠ 1, pomU = pomU + (pocetnost[2][[i, 2]]^3 - pocetnost[2][[i, 2]])],
   {i, 1, Length[pocetnost[2]]}];
  U = 1/12 * pomU;

  rs = (1/6 * (dlzka^3 - dlzka) - sumpor - T - U) /
     Sqrt[(1/6 * (dlzka^3 - dlzka) - 2 * T) * (1/6 * (dlzka^3 - dlzka) - 2 * U)];

  t = (rs * Sqrt[(dlzka - 2)]) / Sqrt[1 - rs^2];
```

```
z = rs * √(dlzka - 1) ;
pvalJS = StudentTPValue[t, (dlzka - 2)];
pvalJS = OneSidedPValue /. pvalJS;
pvalOS = StudentTPValue[t, (dlzka - 2), TwoSided → True];
pvalOS = TwoSidedPValue /. pvalOS;
kvOS = Quantile[StudentTDistribution[dlzka - 2], 1 - α/2];
kvJS = Quantile[StudentTDistribution[dlzka - 2], 1 - α];
Print[Grid[{{"Spearmanov koeficient : ", rs // N, SpanFromLeft},
   {"Pocet merani vo vybere : ", dlzka, SpanFromLeft}, {"", "Obojstranny test",
    "Jednostranny test"}, {"Hladina vyznamnosti α ", α, SpanFromLeft},
   {"p value ", pvalOS, pvalJS}, {"Vyhodnotenie na zaklade p value ",
    If[pvalOS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]],
    If[pvalJS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]},
   {"Vypocitana statistika t ", t // N, SpanFromLeft},
   {"Kriticka hodnota Studentovho t roz. ", kvOS, kvJS},
   {"Vyhodnotenie na zaklade KH ", If[t > kvOS, Text[Style["Zamietam H0", Red]],
     Text[Style["Nezamietam H0", Green]]], If[t > kvJS, Text[Style["Zamietam H0", Red]],
     Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
]

X = {8.6, 5.2, 8.6, 8.7, 14.6, 2.1, 7.3, 10.9, 31.7, 14.4, 13.2, 13.2, 4.2,
   3.5, 6.8, 7.3, 5.2, 18.1, 8.9, 28.2, 22.4, 1.6, 18.5, 5.1, 15.9, 17.5, 5.8};
Y = {7.5, 6.9, 5.3, 3.7, 8.4, 4.7, 4.5, 8.3, 8.3, 8.3, 6.1, 3.9, 6, 4.3,
   4.7, 7.4, 6.4, 3.2, 4.4, 9.6, 8, 6.4, 4.8, 11.1, 6.9, 6.1, 5.3};
SpearmanovKoeficient[X, Y, 0.01]
```

Spearmanov koeficient :	0.223582	
Pocet merani vo vybere :	27	
	Obojstranny test	Jednostranny test
Hladina vyznamnosti α	0.01	
p value	0.262264	0.131132
Vyhodnotenie na zaklade p value	Nezamietam H_0	Nezamietam H_0
Vypocitana statistika t	1.14694	
Kriticka hodnota Studentovho t roz.	2.78744	2.48511
Vyhodnotenie na zaklade KH	Nezamietam H_0	Nezamietam H_0

Zdrojový kód testu významnosti Kendallovho koeficientu

```
<< "HypothesisTesting`"
KendallovKoeficient[X_List, Y_List, α_] :=
 Module[{dlzka, pomoc1, pomoc2, pomoc, pomocna, pocinverz = 0,
   pocneinverz = 0, pomT = 0, pomU = 0, U, T, S, z, tau, pvalJS, pvalOS, kvJS, kvOS},
  dlzka = Length[X];
  data[1] = X; data[2] = Y;
  Table[pomdata[i] = Sort[data[i]], {i, 1, dlzka}];
  Table[pocetnost[i] = Tally[pomdata[i]], {i, 1, 2}];
  For[i = 1, i ≤ 2, i++,
   skporadie[i] = Partition[Flatten[Table[Tuples[{pomdata[i][[j]], j}, 1], {j, 1, dlzka}]], 2];
  ];
  For[j = 1, j ≤ 2, j++,
   poradie[j] = Flatten[Table[Total[Position[pomdata[j], pocetnost[j][[i, 1]]]] /
        pocetnost[j][[i, 2]], {i, 1, Length[pocetnost[j]]}]];
  ];
  For[j = 1, j ≤ 2, j++,
   poradiedata[j] = Partition[Flatten[Table[
        Tuples[{pocetnost[j][[i, 1]], poradie[j][[i]]}, 1], {i, 1, Length[poradie[j]]}]], 2];
  ];
  Table[If[data[1][[j]] == poradiedata[1][[i, 1]], pdata1[j] = poradiedata[1][[i, 2]]],
   {i, 1, Length[poradiedata[1]]}, {j, 1, dlzka}];
  pomoc1 = Table[pdata1[i], {i, 1, dlzka}];
  Table[If[data[2][[j]] == poradiedata[2][[i, 1]], pdata2[j] = poradiedata[2][[i, 2]]],
   {i, 1, Length[poradiedata[2]]}, {j, 1, dlzka}];
  pomoc2 = Table[pdata2[i], {i, 1, dlzka}];
  pomoc = Partition[Flatten[Table[Tuples[{pomoc1[[i]], pomoc2[[i]]}, 1], {i, 1, dlzka}]], 2];
  pomocna = Sort[pomoc];
  For[j = 1, j ≤ dlzka, j++,
   Table[If[pomocna[[j, 2]] > pomocna[[i, 2]] && pomocna[[j, 1]] ≠ pomocna[[i, 1]],
     pocneinverz++], {i, j, dlzka}];
   Table[If[pomocna[[j, 2]] < pomocna[[i, 2]] && pomocna[[j, 1]] ≠ pomocna[[i, 1]],
     pocinverz++], {i, j, dlzka}];
  ];
  S = pocinverz - pocneinverz;
  Table[
   If[pocetnost[1][[i, 2]] ≠ 1, pomT = pomT + (pocetnost[1][[i, 2]]^2 - pocetnost[1][[i, 2]])],
   {i, 1, Length[pocetnost[1]]}];
  T = pomT;
  Table[
   If[pocetnost[2][[i, 2]] ≠ 1, pomU = pomU + (pocetnost[2][[i, 2]]^2 - pocetnost[2][[i, 2]])],
   {i, 1, Length[pocetnost[2]]}];
  U = pomU;
  tau = (2 * S) / (Sqrt[dlzka * (dlzka - 1) - T] * Sqrt[dlzka * (dlzka - 1) - U]);
  z = (3 * tau * Sqrt[dlzka * (dlzka - 1)]) / Sqrt[2 * (2 * dlzka + 5)];
  pvalJS = NormalPValue[z];
```

```
    pvalJS = OneSidedPValue /. pvalJS;
    pvalOS = NormalPValue[z, TwoSided → True];
    pvalOS = TwoSidedPValue /. pvalOS;
    kvOS = Quantile[NormalDistribution[0, 1], 1 - α/2];
    kvJS = Quantile[NormalDistribution[0, 1], 1 - α];
    Print[Grid[{{"Kendallov koeficient : ", tau // N, SpanFromLeft},
       {"Pocet merani vo vybere : ", dlzka, SpanFromLeft}, {"", "Obojstranny test",
        "Jednostranny test"}, {"Hladina vyznamnosti α ", α, SpanFromLeft},
       {"p value ", pvalOS, pvalJS}, {"Vyhodnotenie na zaklade p value ",
        If[pvalOS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]],
        If[pvalJS < α, Text[Style["Zamietam H0", Red]], Text[Style["Nezamietam H0", Green]]]},
       {"Vypocitana z statistika ", z // N, SpanFromLeft}, {"Kvantil N(0,1) ", kvOS, kvJS},
       {"Vyhodnotenie na zaklade kvantilu ", If[z > kvOS, Text[Style["Zamietam H0", Red]],
         Text[Style["Nezamietam H0", Green]]], If[z > kvJS, Text[Style["Zamietam H0", Red]],
         Text[Style["Nezamietam H0", Green]]]}}, Frame → All]];
   ]

  X = {8.6, 5.2, 8.6, 8.7, 14.6, 2.1, 7.3, 10.9, 31.7, 14.4, 13.2, 13.2, 4.2,
     3.5, 6.8, 7.3, 5.2, 18.1, 8.9, 28.2, 22.4, 1.6, 18.5, 5.1, 15.9, 17.5, 5.8};
  Y = {7.5, 6.9, 5.3, 3.7, 8.4, 4.7, 4.5, 8.3, 8.3, 8.3, 6.1, 3.9, 6, 4.3,
     4.7, 7.4, 6.4, 3.2, 4.4, 9.6, 8, 6.4, 4.8, 11.1, 6.9, 6.1, 5.3};
  KendallovKoeficient[X, Y, 0.01]
```

Kendallov koeficient :	0.147829	
Pocet merani vo vybere :	27	
	Obojstranny test	Jednostranny test
Hladina vyznamnosti α	0.01	
p value	0.279385	0.139693
Vyhodnotenie na zaklade p value	Nezamietam H_0	Nezamietam H_0
Vypocitana z statistika	1.0817	
Kvantil N(0,1)	2.57583	2.32635
Vyhodnotenie na zaklade kvantilu	Nezamietam H_0	Nezamietam H_0