

Neparametrické metody

Petr Soukup, Statistika II, FSV UK

Neparametrické testy

- Všeobecne testy nevyžadujú toľko predpokladov ako parametrické testy (t-testy, analýza rozptylu apod. viz predpoklady normality, homogenita premenných apod.)

Typy neparametrických testov

- 1. Obdoby parametrických t- testov a analýzy rozptylu
- 2. Neparametrické testy pre nominálne a dichotomické premenné (Binomial Test, Chi-Square test, Runs test)
- 3. Neparametrické testy pro rozdelenia premenných (Kolmogorov-Smirnov test pro 1 a 2 výbĚry)

1. Obdoby parametrických testov

- založené najmä na mediánoch a poradí teda sú určené najmä pre ordinálne premenné a malé súbory
- Pozn. predpoklady parametrických testov- kardinálna veličina s normálnym rozdelením (pri t-testoch nevadí, pokiaľ je normalita porušená - testy sú „robustné“) a aspoň 30 pozorovaní

Predpoklady pre neparametrické testy:

- malé výbery (cca do 30)
- nekardinálny charakter premenných alebo kardinálne premenné s nenormálnym rozdelením (malých výberov)

1. Obdoby parametrických testov

situácia	Parametrický test	Neparametrický test
a) Dva nezávislé výbery	Dvojvýběrový t-test	Man-Whitney test aj.
b) Dva závislé výběry	Párový t-test	Wilcoxonův test aj.
c) Viac nezávislých výberov	Jednofaktorová analýza rozptylu	Kruskal-Wallis test aj.
d) Viac závislých výberov		Friedmanov test aj.

1. Obdoby parametrických testov

a) Dva nezávislé výbery

- **Mann-Whitney test**
- pre dva nezávislé výbery
- pre malé výbery, ordinálne veličiny
- H_0 : mediány vo výberoch sa rovnajú
- H_1 : mediány sa líšia
- pre výbery menšie než 15 použitie Exact Sig.

Mann-Whitneyov U – test

- neparametrický test pre **nezávislé** súbory (nepárový)

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \Sigma P_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - \Sigma P_2$$

- každému športovcovi na základe hodnôt výkonu priradíme poradie stanovené z obidvoch výberov,
- ΣP_1 je súčet čísiel poradia jednej skupiny,
- ΣP_2 je súčet čísiel poradia druhej skupiny športovcov,
- **nižšia** vypočítaná hodnota **U** je **kritériom testu**.

Krok 1.

Zápis nameraných hodnôt

Zápis číslo:	Zistené hodnoty	
	Experiment:	
	A	B
1	5	7
2	22	27
3	11	16
4	29	34
5	22	27
6	19	25
7	16	21
8	33	38
9	29	34
10	27	32
11	11	16
12	31	36
13	18	31
14	-	25

Krok 2.

Tabuľku prepíšeme do formy:

Pri úprave si pomáhame ponukou EXCLU

A	5	B	7
A	22	B	27
A	11	B	16
A	29	B	34
A	22	B	27
A	19	B	25
A	16	B	21
A	33	B	38
A	29	B	34
A	27	B	32
A	11	B	16
A	31	B	36
A	18	B	31
-	-	B	25

$$n_1 = 13; \quad n_2 = 14$$

Krok 3

Údaje zapíšeme pod seba:

Prikazom "ZORADIŤ"
usporiadame tabuľku
vzostupne podľa nameraných
hodnôt.

A	5
B	7
A	11
A	11
A	16
B	16
B	16
A	18
A	19
B	21
A	22
A	22
B	25
B	25
A	27
B	27
B	27
A	29
A	29
A	31
B	31
B	32
A	33
B	34
B	34
B	36
B	38

Krok 4

Nameraným údajom priradíme poradové číslo podľa
tohto vzoru!

Experiment y	Namerané hodnoty	Poradie pre výpočet	Pomocná kolónka
A	5	1,0	1
B	7	2,0	2
A	11	3,5	3
A	11	3,5	4
A	16	6,0	5
B	16	6,0	6
B	16	6,0	7
A	18	8,0	8
A	19	9,0	9
B	21	10,0	10
A	22	11,5	11
A	22	11,5	12
B	25	13,5	13
B	25	13,5	14
A	27	16,0	15
B	27	16,0	16
B	27	16,0	17
A	29	18,5	18
A	29	18,5	19
A	31	20,5	20
B	31	20,5	21
B	32	22,0	22
A	33	23,0	23
B	34	24,0	24
B	34	25,0	25
B	36	26,0	26
B	38	27,0	27

SUMA poradia B; T₂ **227,5**SUMA poradia A; T₁ **150,5**

1. Obdoby parametrických testů

b) Dva závislé výběry

- **Wilcoxonov test pro dva závislé výběry**
- pro ordinálne veličiny, kedy neplatí normalita a je malý počet pozorovaní
- H_0 : mediány u dvoch závislých premenných (výberoch) sa rovnajú
- H_1 : mediány u dvoch závislých premenných sa líšia

1. Obdoby parametrických testů

b) Dva závislé výběry

- Sign (Znaménkový) – slabší síla oproti Wilcoxon,

1. Obdoby parametrických testov

c) Viac nezávislých výberov

- **Kruskal- Wallisov test**
- obdoba analýzy rozptylu pre ordinálne premenné alebo pre kardinálne premenné, kde sa rozptyly v skupinách výrazne líšia
- pri teoretickom výpočtu sa hodnota testového kritéria porovnáva s tabuľkou normálneho rozdelenia
- H_0 zamietame menej často než pri parametrickom teste (máme menej kvalitné dáta)- účinnosť testu v porovnaní s analýzou rozptylu cca 90%
- H_0 : mediány vo všetkých skupinách (výberoch) sa rovnajú
- H_1 : mediány aspoň dvoch skupín sa líšia
- Po zamietnutí H_0 je nutné skúmať ktoré skupiny sa líšia

<i>Metody výuky</i>		
A	B	C
19	27	13
25	30	22
18	22	24
18	29	20
Σ80	Σ108	Σ79

<i>Metody výuky</i>		
A	B	C
4	10	1
9	12	6,5
2,5	6,5	8
2,5	11	5
Σ18	Σ39,5	Σ20,5

$$H = \left[\frac{12}{n \cdot (n+1)} \sum \frac{R_i^2}{n_i} \right] - 3(n+1)$$

$$H = \left[\frac{12}{12 \cdot 13} \left(\frac{18^2}{4} + \frac{39,5^2}{4} + \frac{20,5^2}{4} \right) \right] - 3 \cdot 13 = 5,317$$

$$\chi_{0,05}^2(2) = 5,991$$

TK < KH

1. Obdoby parametrických testů

c) Viac nezávislých výberov

- **Mediánový test**
- H_0 : mediány v skupinách v populácií sa rovnajú
- H_1 : mediány aspoň dvoch skupín sa líšia
- Menej silný test než K-W test

1. Obdoby parametrických testů

d) Více závislých výběrů

- **Friedmanův test**
- Otázka: existuje rozdíl mezi mediány v jednotlivých závislých skupinách?
- H_0 : rozdělení ve více závislých výběrech jsou shodné
- H_1 : rozdělení v alespoň dvou závislých výběrech se liší
- po zamítnutí H_0 musí následovat další test, abychom zjistili, které skupiny se mezi sebou liší- vytvoříme dvojice a následně použijeme např. Wilcoxonův test (pro dva závislé výběry)

1. Obdoby parametrických testů

d) Více závislých výběrů

- **Kendall** (stejně jako Friedman, jen navíc počítá Kendallův koeficient konkordance W od 0 do 1 čím více shoda v jednotlivých výběrech tím větší hodnota koeficientu)
- **Cochranovo Q** (Friedman pouze pro dichotomické proměnné)
- Př. Posouzení shody hodnocení u několika porotců, shoda

Spearmanov koeficient poradovej korelácie

- Spravidla nás zaujíma, aká tesná závislosť je (aký veľký je stupeň závislosti medzi sledovanými javmi, ktoré boli zachytené pomocou ordinálneho (poradového) merania
- n je počet porovnávaných dvojíc hodnôt a d je rozdiel (diferencie) poradí pre jednu dvojicu hodnôt.

$$r_s = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)}$$

Koeficient korelace	Interpretace
$r = 1$	naprostá závislosť (funkční závislosť)
$1,00 > r \geq 0,90$	velmi vysoká závislosť
$0,90 > r \geq 0,70$	vysoká závislosť
$0,70 > r \geq 0,40$	střední (značná) závislosť
$0,40 > r \geq 0,20$	nížká závislosť
$0,20 > r \geq 0,00$	velmi slabá závislosť
$r = 0$	naprostá nezávislosť

Dítě	Hmotnost (kg)	Pořadí podle hmotnosti	Pořadí podle rychlosti v běhu		
Kristina	16,70	9	1	8	64
Vladimír	18,45	7	2	5	25
Milan	16,70	9	3	6	36
Lenka	19,50	6	4	2	4
Robert	16,70	9	5	4	16
Ivan	23,30	2	6	-4	16
Jiří	23,40	1	7	-6	36
Zuzana	19,80	4,5	8	-3,5	12,25
Blanka	19,80	4,5	9	-4,5	20,25
Věra	22,20	3	10	-7	49

Σ 278,50

$$r_s = 1 - \frac{6 \cdot 278,5}{10(10^2 - 1)} = -0,69$$