

ADDS cviceni

Pavlina Kuranova

Testy pro dva nezávislé výběry

- Mannův Whitneyho test
 - Založen na Wilcoxonově statistice W
 - založen na pořadí jednotlivých pozorování (oba výběry spojeny do jednoho celku) –v případě stejně velkých výběrů pro oba je pak Wilc. statistika menší hodnota ze získaných součtů.
- Test vhodný i pro ordinální data.

Mannův-Whitneyho test

- Příklad (soubor *spokojenost_v_praci*)
- Budeme zjišťovat zda je spokojenost se svou pracovní pozicí závislá na pohlaví.
- *H₀: spokojenost v práci nezávisí na pohlaví*
- *H_A: spokojenost v práci závisí na pohlaví*
- *Analyze – Nonparametric Tests a 2 Independent Samples*

Mannův-Whitneyho test

- Příklad (soubor *spokojenost_v_praci*)
- Budeme zjišťovat zda je spokojenost se svou pracovní pozicí závislá na pohlaví.
- *Analyze – Nonparametric Tests a 2 Independent Samples*

	sex	N	Mean Rank	Sum of Ranks
pozice	muž	47	55,70	2618,00
	žena	61	53,57	3268,00
	Total	108		

	pozice
Mann-Whitney U	1377,000
Wilcoxon W	3268,000
Z	-,373
Asymp. Sig. (2-tailed)	,709

a. Grouping Variable: sex

- *Na základě hodnoty Asymp. Sig. nezamítáme H_0 , tzn. Spokojenost v práci nezávisí na pohlaví.*

Testy pro více než dva nezávislé výběry

- **Kruskal-Wallisův test**
 - Všechny měření uspořádáme podle velikosti
 - Hodnoty nahradíme jejich pořadími
 - Měříme rozdíl průměrů pořadí ve skupinách
 - Vhodný i pro ordinální data
 - *Přibližně Chí-kvadrát rozdělení s $K-1$ stupni volnosti*
- **Předpoklad:**

Použití by mělo být jen když minimum $n_i \geq 5$

Testy pro více než dva nezávislé výběry

- Kruskal-Wallisův test - příklad

Na základě získaných dat budeme zjišťovat zda spokojenost se svou pracovní pozicí závisí na věku respondentů.

H_0 : spokojenost s pracovní pozicí nezávisí na věku respondenta

H_A : spokojenost s pracovní pozicí závisí na věku respondenta

Jedná se o test, který využívá pro výpočet hodnotu mediánu, tzn.

H_0 : mediány ve věkových skupinách jsou shodné,

H_A : alespoň jeden medián je různý od ostatních

Testy pro více než dva nezávislé výběry

- Kruskal-Wallisův test - příklad

Budeme zjišťovat zda spokojenost se svou pracovní pozicí závisí na věku respondentů.

- *Analyze – Nonparametric Tests – K Independent Samples –*

Zaškrtnuto Kruskal–Wallis H

- *Proměnnou věk nutno překódovat na metrickou....*

roky	N	Mean Rank
pozice do 25-ti let	14	58,39
26 -35 let	56	58,45
36 - 45let	26	48,40
nad 46 let	12	44,75
Total	108	

	pozice
Chi-Square	3,699
df	3
Asymp. Sig.	,296

a. Kruskal Wallis Test

b. Grouping Variable:
roky

- *Na základě hodnoty Asymp. Sig, vidíme, že na 5% hladině významnosti nezamítáme H_0 o shodě mediánů. Spokojenost s pracovní pozicí nezávisí na věku respondentů.*

Mediánový test pro K výběrů

- Používá se když předpokládáme ve výběrech různá rozdělení.
- Nevytváříme pořadí, ale spočteme medián ze všech hodnot vysvětlované proměnné (grand median)
- Není ovlivněn odlehlými hodnotami
- Konzervativnější - méně informací určuje pouze zda se nacházíme nad nebo pod mediánem

Mediánový test pro K výběrů

- Příklad: zadání stejné jako v předchozím případě:
- Budeme zjišťovat zda spokojenost se svou pracovní pozicí závisí na věku respondentů.

Analyze – Nonparametric Tests – K Independent Samples – zaškrtneme položku Median

		roky			
		do 25-ti let	26 -35 let	36 - 45let	nad 46 let
pozice	> Median	6	17	4	4
	<= Median	8	39	22	8

	pozice
N	108
Median	2,0000
Chi-Square	3,825 ^b
df	3
Asymp. Sig.	,281

a. Grouping Variable:
roky

b. 2 cells (25,0%) have expected frequencies less than 5. The minimum expected cell frequency is 3,4.

Na základě této tabulky bychom byli schopni spočítat očekávané četnosti...

Mediánový test pro K výběrů

- Příklad: zadání stejné jako v předchozím případě:
Zjišťujeme, zda spokojenost se svou pracovní pozicí závisí na věku respondentů.

Test Statistics^a

	pozice
N	108
Median	2,0000
Chi-Square	3,825 ^b
df	3
Asymp. Sig.	,281

a. Grouping Variable:
roky

b. 2 cells (25,0%) have
expected frequencies
less than 5. The
minimum expected
cell frequency is 3,4.

» Zde vidíme údaje: N – celkový počet pozorování

Median – určený ze všech pozorování (N hodnot vysvětlující proměnné) – gradn
Median

Chi-Square – hodnota testové statistiky

df- počet stupňů volnosti

Asymp. Sig. – minimální hladina významnosti od které zamítáme nulovou
hypotézu

Pozor! Měly by platit stejné předpoklady jako u chí-kvadrát testu!!!

Výsledek: Na 5% hladině významnosti nezamítáme H_0 , spokojenost s pracovní pozicí
nezávisí na věku respondenta

ANOVA

- ANOVA = ANalysis Of VAriance
- H_0 : proměnné jsou nezávislé (průměry ve všech skupinách jsou si rovny)
- H_A : proměnné jsou závislé (alespoň dva průměry jsou různé)

Variabilita	Součet čtverců	Stupně volnosti	Průměrný čtverec	F	p -hodnota
Meziskupinová	SS_b	$k - 1$	$\frac{SS_b}{k - 1}$	$F = \frac{MS_b}{MS_w}$	p
Vnitroskupinová	SS_w	$n - k$	$\frac{SS_w}{n - k}$		
Celková	SS_T	$n - 1$			

ANOVA - předpoklady

- Zkoumaná proměnná je metrická
- Všechna měření jsou vzájemně nezávislá uvnitř skupin i mezi skupinami
- Výběry pocházejí z normálního rozdělení
- Všechny rozptyly ve skupinách jsou stejné (minimálně musí platit $\frac{\max s_i}{\min s_i} \leq 3$, kde s_i jsou směrodatné odchylky v jednotlivých skupinách)

Levenův test

- Zkoumá rovnost rozptylů

$$H_0: \sigma_1^2 = \sigma_2^2, H_A: \sigma_1^2 \neq \sigma_2^2$$

- Předpoklady:

- Metrická proměnná
- Náhodné výběry jsou nezávislé
- Robustní test, pokud $n > 30$, nemělo by vadit porušení normality

Využití u Anovy jako : Test of Homogeneity Variance

ANOVA - příklad

- Otestujme zda příjem obyvatel je závislý na dosaženém vzdělání.
- Soubor *spokojenost_v_praci* (pro jednoduchost spojíme data do 3 skupin)
- Překódování je nutné!!! Pro ANOVU musí být proměnné metrické!!!

Proměnná vzdělání:

- Nižší (základní + vyučen)
- Střední (maturita)
- Vyšší (vysokoškolské + vyšší odborné)

Proměnná příjem:

překódujeme do řádu tisíců

ANOVA - příklad

- Otestujme zda příjem obyvatel je závislý na dosaženém vzdělání.

Descriptives

mzda2

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
nižší	16	19,0625	2,01556	,50389	17,9885	20,1365	15,00	20,00
střední	48	26,1458	8,58249	1,23878	23,6537	28,6379	15,00	45,00
vyšší	44	31,5909	10,38630	1,56579	28,4332	34,7486	15,00	45,00
Total	108	27,3148	9,72582	,93587	25,4596	29,1701	15,00	45,00

Test of Homogeneity of Variances

mzda2

Levene Statistic	df1	df2	Sig.
15,800	2	105	,000

Na základě testu vidíme že hodnota $p_value (Sig.) < \alpha (5\%)$

Na hladině významnosti 5% zamítám nulovou hypotézu o rovnosti rozptylů.

ANOVA - příklad

- Otestujme zda příjem obyvatel je závislý na dosaženém vzdělání.

ANOVA

mzda2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1959,743	2	979,872	12,606	,000
Within Groups	8161,553	105	77,729		
Total	10121,296	107			

- Na základě hodnoty Sig. (p-value) bychom zamítli nulovou hypotézu na hladině významnosti 5%.
- NEBYL splněn předpoklad testu!!! Anova by neměla být prováděna, je možno využít nějakou alternativu (K-W test, mediánový test)

ANOVA – příklad 2

- Otestujme zda příjem je závislý na věku respondentů.
- Soubor *spokojenost_v_praci*
- Zkontrolujte proměnné!!! Pro ANOVU musí být proměnné metrické!!!

Proměnná věk (4 skupiny):

- do 25-ti let

- Nad 46 let

Proměnná příjem:

překódujeme do řádu tisíců

ANOVA - příklad

- Otestujme zda příjem obyvatel je závislý na věku.

Descriptives

mzda2

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					do 25-ti let	14		
26 -35 let	56	29,5536	9,15904	1,22393	27,1008	32,0064	15,00	45,00
36 - 45let	26	27,5000	10,79352	2,11678	23,1404	31,8596	15,00	45,00
nad 46 let	12	23,7500	8,29156	2,39357	18,4818	29,0182	15,00	40,00
Total	108	27,3148	9,72582	,93587	25,4596	29,1701	15,00	45,00

Test of Homogeneity of Variances

mzda2

Levene Statistic	df1	df2	Sig.
1,555	3	104	,205

Na základě testu vidíme že hodnota $p_value (Sig.) \geq \alpha(5\%)$

Na hladině významnosti 5% nezamítám nulovou hypotézu o rovnosti rozptylů.

ANOVA - příklad

- Otestujme zda příjem obyvatel je závislý na věku.

ANOVA

mzda2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	979,778	3	326,593	3,716	,014
Within Groups	9141,518	104	87,899		
Total	10121,296	107			

- Na základě hodnoty Sig. (p-value) zamítáme nulovou hypotézu na hladině významnosti 5%. *Na hladině významnosti 1% bychom nulovou hypotézu nemohli zamítnout!!!*
- Výsledek: Příjem respondentů nezávisí na věku.

ANOVA – POST HOC testy

Multiple Comparisons

Dependent Variable: mzda2

	(I) roky	(J) roky	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
LSD	do 25-ti let	26 - 35 let	-8,48214*	2,80145	,003	-14,0375	-2,9268
		36 - 45let	-6,42857*	3,10793	,041	-12,5917	-,2654
		nad 46 let	-2,67857	3,68829	,469	-9,9926	4,6354
	26 - 35 let	do 25-ti let	8,48214*	2,80145	,003	2,9268	14,0375
		36 - 45let	2,05357	2,22494	,358	-2,3586	6,4657
		nad 46 let	5,80357	2,98238	,054	-,1106	11,7177
	36 - 45let	do 25-ti let	6,42857*	3,10793	,041	,2654	12,5917
		26 - 35 let	-2,05357	2,22494	,358	-6,4657	2,3586
		nad 46 let	3,75000	3,27195	,254	-2,7384	10,2384
	nad 46 let	do 25-ti let	2,67857	3,68829	,469	-4,6354	9,9926
		26 - 35 let	-5,80357	2,98238	,054	-11,7177	,1106
		36 - 45let	-3,75000	3,27195	,254	-10,2384	2,7384
Bonferroni	do 25-ti let	26 - 35 let	-8,48214*	2,80145	,019	-16,0171	-,9472
		36 - 45let	-6,42857	3,10793	,246	-14,7879	1,9307
		nad 46 let	-2,67857	3,68829	1,000	-12,5988	7,2417
	26 - 35 let	do 25-ti let	8,48214*	2,80145	,019	,9472	16,0171
		36 - 45let	2,05357	2,22494	1,000	-3,9308	8,0379
		nad 46 let	5,80357	2,98238	,326	-2,2180	13,8251
	36 - 45let	do 25-ti let	6,42857	3,10793	,246	-1,9307	14,7879
		26 - 35 let	-2,05357	2,22494	1,000	-8,0379	3,9308
		nad 46 let	3,75000	3,27195	1,000	-5,0504	12,5504
	nad 46 let	do 25-ti let	2,67857	3,68829	1,000	-7,2417	12,5988
		26 - 35 let	-5,80357	2,98238	,326	-13,8251	2,2180
		36 - 45let	-3,75000	3,27195	1,000	-12,5504	5,0504

*. The mean difference is significant at the 0.05 level.

ANOVA – POST HOC testy

- V případě zamítnutí nulové hypotézy!! Otestujme zda jsou některé věkové skupiny ekvivalentní.

ANOVA

mzda2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	979,778	3	326,593	3,716	,014
Within Groups	9141,518	104	87,899		
Total	10121,296	107			

- *Analyze – Compare Means – One-Way Anova – POST HOC*

- Graphs – legacy dialogs – error bar -

