

LEKCE 06 PŘÍKLAD

ANOVA – ONE-WAY

Při t testu jsme testovali nulovou hypotézu, že průměry dvou skupin jsou v populaci stejné. Není to příliš obvyklá situace, neboť při srovnávání průměrů máme často skupin více: např. nás zajímá, jak se liší důležitost Boha v životě jedince podle věkových skupin nebo podle vzdělání apod. Proto si nyní ukážeme, jak testovat nulovou hypotézu, že několik průměrů různých skupin jsou v populaci stejné. Procesura, která se k tomu používá, se nazývá analýza rozptylu (*analysis of variance*), často zkracovaná jako *ANOVA*.

Test analýzy rozptylu získal své jméno od způsobu analýzy, zkoumá totiž variabilitu (rozptyl) v datech výběrového souboru. Sleduje přitom dvojí variabilitu: jednak zkoumá, jak mnoho se liší hodnoty **uvnitř** jednotlivých skupin (*within-groups variability*), jednak analyzuje, jak mnoho se liší průměry **mezi** skupinami (*between-groups variability*). To, že pro testování rozdílů v populačních průměrech používáme analýzu variability, je naprosto v pořádku, neboť závěry o populačních průměrech se vždy dělají na základě analýzy variability výběrových dat.

V daném příkladě jsme použili tu nejjednodušší variantu *ANOVA*, tzv. jednovýchodnou (*one-way analysis of variance*), neboť jsme srovnávali, jak se liší hodnoty závisle proměnné u skupin jedné nezávisle proměnné. Této proměnné se říká faktor a z tohoto důvodu se tato technika také někdy nazývá jednofaktorovou analýzou rozptylu. Pokud bychom řešili úlohu, jak se např. liší průměrné postoje k důležitosti Boha pro věkové kategorie a uvnitř nich ještě podle vzdělanostních skupin, museli bychom požit tzv. dvojevýchodnou analýzu rozptylu (*two-way analysis of variance*), neboť nezávisle proměnné (nebo též faktory) máme dvě, to je věk a vzdělání.

Jaké předpoklady musí být splněny, abychom mohli *ANOVu* použít?

1. Jednotlivá pozorování musí být na sobě nezávislá. Tento předpoklad je v sociologických šetřeních vždy bez problémů, výzkumné designy s opakovaným měřením stejných subjektů (respondentů) nejsou příliš časté.
2. Rozložení v populaci je normální. Tento předpoklad se v praxi často zanedbává, v úvahu musí být brán pouze tehdy, pokud jsou naše data rozložena extrémně nenormálně.
3. Rozptyly v populaci jsou stejné. V praxi je tento předpoklad naplněn tehdy, když jednotlivé skupiny mají přibližně stejný počet jednotek. Jinak je možné tento předpoklad testovat pomocí Levenova testu pro shodnost rozptylů.

Pokud data tyto předpoklady nespĺňují, je třeba učinit následující kroky. V případě nenormálního rozložení je možno data transformovat (např. logaritmováním nebo druhou odmocninou). Pokud ani tato transformace nepomůže, je třeba použít adekvátní náhrady, již je v tomto případě neparametrický test Kruskal-Wallisův.

Příklad 7.1: Je v souboru EVS_ČR1999.sav důležitost Boha (q33) vnímána rozdílně různými věkovými skupinami (vek-kat)?

Řešení

Nejdříve se podívejme, jak tyto průměry v těchto skupinách vůbec vypadají. Použijme procedury *Means* a *Graphs – Line – Simple – Define – Other summary function*.

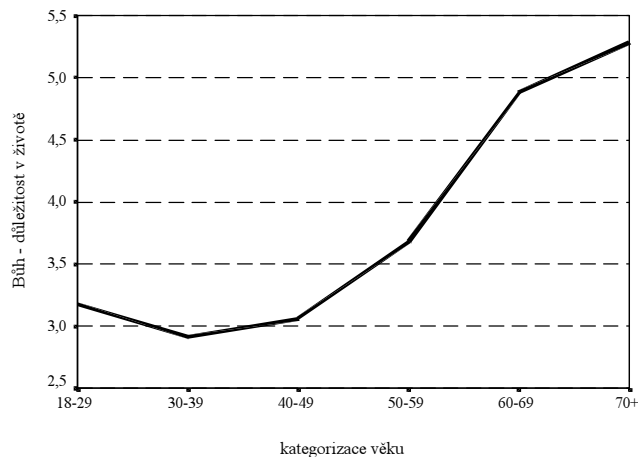
Report

Q33 Bůh - důležitost v životě

VEK_KAT	kategorizace věku	Mean	N	Std. Deviation
1	18-29	3,18	413	2,79
2	30-39	2,92	301	2,62
3	40-49	3,06	369	2,66
4	50-59	3,68	329	2,99
5	60-69	4,88	260	3,42
6	70+	5,28	170	3,58
Total		3,64	1842	3,06

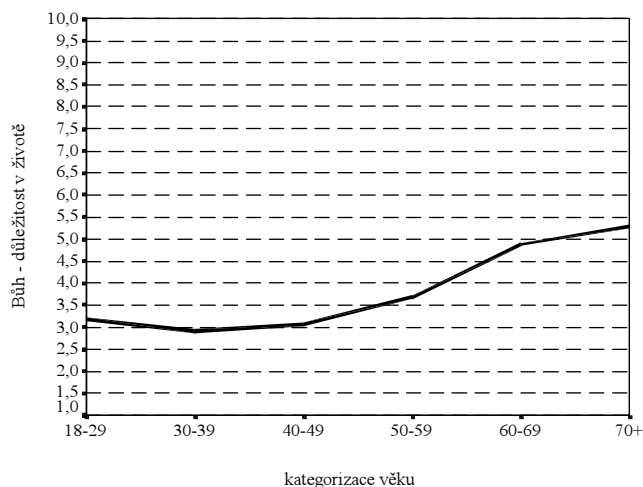
Rozdíly zde, jak vidíme, skutečně jsou. Jelikož desetibodová měřicí stupnice byla orientována od nedůležitosti k vysoké důležitosti, vidíme, že nejvyšší důležitost přisuzují Bohu nejstarší respondenti, nejméně naopak respondenti mezi 30–39 lety (Naleznete sociologické vysvětlení pro tento výsledek?).

Obr. 7.1:



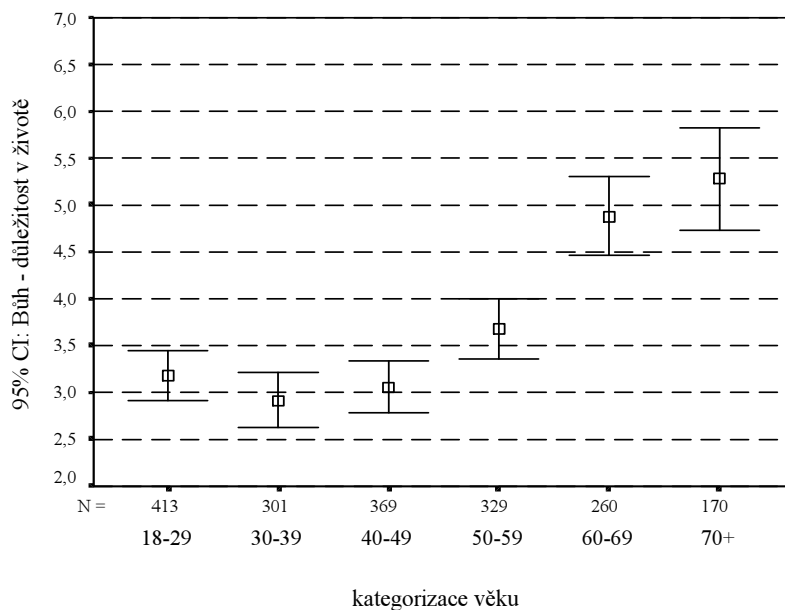
Opět pozor, impresivnost rozdílů opět zmizí, když na vertikální osu necháme vynést celou desetibodovou stupnici – graf tedy musíme ve výstupu z SPSS editovat.

Obr. 7.2



Nyní si nechejme udělat graf intervalů spolehlivosti pro jednotlivé skupiny:
 (Graphs – Error Bar – Simple – Define – Variable (q33) – Category Axis (vek-kat – Confidence interval for Mean 95 % – v Options nezapomeňte zrušit kategorii Missing values)

Obr. 7.3



Už tento obrázek mnohé naznačuje o signifikanci rozdílů. Průměry těch skupiny, jejichž „vousy“ se v grafu nepřekrývají budou statisticky významně odlišné. V našem obrázku to jsou všechny mladší věkové skupiny oproti skupinám nad 60 let. Přesvědčeme se o tom i výpočtem pro analýzu rozptylu (Compare Means – One-Way Anova).

Když si v Options zaškrtnete Descriptives, získáte nejdříve tuto tabulku (viz níže), v níž jsou základní důležité údaje popisné statistiky.

Descriptives

Q33 Bůh - důležitost v životě

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1 18-29	413	3,18	2,79	,14	2,91	3,45
2 30-39	301	2,92	2,62	,15	2,62	3,22
3 40-49	369	3,06	2,66	,14	2,79	3,33
4 50-59	329	3,68	2,99	,17	3,36	4,01
5 60-69	260	4,88	3,42	,21	4,47	5,30
6 70+	170	5,28	3,58	,27	4,74	5,82
Total	1842	3,64	3,06	,07	3,50	3,78

Tabulka analýzy rozptylu ANOVA (viz níže) již skýtá nezbytné údaje pro zodpovězení testu, zdali se průměry od sebe statisticky odlišují. Testujeme nulovou hypotézu, že rozdíly v průměrech mezi všemi šesti věkovými skupinami budou nulové, že tedy mezi skupinami nebude žádný rozdíl. Alternativní hypotézou je předpoklad, že průměry se liší.

ANOVA

Q33 Bůh - důležitost v životě

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1229,066	5	245,813	28,206	,000
Within Groups	15991,689	1835	8,715		
Total	17220,755	1840			

Důležitými údaji v tabulce jsou statistika F a její signifikance. Hodnota F vzniká jako podíl variability mezi skupinami a variability uvnitř skupin, konkrétně jejich průměrů součtu druhých mocnin směrodatných odchylek (v tabulce sloupec *Mean Square*). V našem případě je tedy hodnota F rovna podílu $245,813 / 8,715 = 28,206$.

Pokud platí nulová hypotéza, že rozdíly mezi průměry jsou nulové, musí být obě průměrné hodnoty druhých mocnin podobné a jejich vzájemný poměr (F) tedy musí být blízko 1. Náš poměr se od jedné hodně liší. Srovnáme-li vypočtenou F hodnotu s F rozložením¹ (to dělá samozřejmě SPSS), zjistíme, zdali je možno nulovou hypotézu podržet, či nikoliv. Podíváme-li se na signifikanci tohoto rozdílu, vidíme, že pravděpodobnost podržet nulovou hypotézu je velmi nízká (0,000), takže nulovou hypotézu zamítáme a můžeme si být jisti, že průměry budou v základním souboru nestejně, budou se lišit.

Statisticky signifikantní F nám ovšem říká pouze to, že je velmi nepravděpodobné, že populační průměry jsou shodné. To ale není výsledek, který by nás plně uspokojil. Cílem naší analýzy je totiž zjistit, mezi kterými konkrétními skupinami se tento rozdíl objevuje. Možná se odlišuje všech pět

¹ F rozložení je matematický model rozložení. V případě, že pracujeme s rozptyly, nemůžeme při testování hypotéz použít ani modelu normálního rozložení, ani t-rozložení, neboť rozložení rozptylů není normální.

skupin navzájem, ale možná také se liší jen některé z nich. Proto v testování pokračujeme dále. A jelikož tento test je aplikován až poté, kdy data už byla částečně analyzována, uplatníme tzv. post-hoc (následnou) proceduru tzv. **mnohonásobného srovnání** (*Post Hoc Multiple Comparison*).

Zadáme ji tak, že v dialogovém okně ANOVY klikneme myší na tlačítko *Post Hoc*. Celý postup tedy vypadá takto:

Procedura: *STATISTICS — COMPARE MEANS — ONE-WAY ANOVA — Dependent list (g33), factor (vek-kat) — Post Hoc — \checkmark Bonferroni – Signifikance level: ,05*

(Pozn.: Doporučuje se užívat buď Bonferroniho testu nebo testu Tukeyho. LSD test je příliš liberální a Scheffeho test je naopak příliš konzervativní (přísný).)

V tabulce *Multiple Comparisons* (viz níže) jsou uvedeny výsledky všech kombinací párového srovnání průměrů, na každém řádku jsou vždy porovnávány dvě skupiny. Např. v prvním bloku šesti řádků se srovnávají průměrné hodnoty respondentů ve věku 18-29 let se základním vzděláním se všemi ostatními věkovými kategoriemi. Ze všech údajů, které jsou v tabulce mnohonásobného srovnání uvedeny, jsou pro interpretaci nejdůležitější hodnoty v druhém sloupci (označeném *Mean Difference*), kde jsou uvedeny rozdíly v jednotlivých dvojicích průměrů. Zajímají nás především ty hodnoty, které jsou označeny hvězdičkou. Ta signalizuje, že daný rozdíl je statisticky významný s 95% pravděpodobností. Dokazuje to sloupec čtvrtý (*Sig.*), v němž je uvedena přesná hodnota signifikance - všude, kde je druhém sloupci hvězdička, má vypočtená signifikance hodnotu nižší než 0,05.

Multiple Comparisons

Dependent Variable: Q33 Bůh - důležitost v životě

Bonferroni

(I) VEK_KAT kategorizace věku	(J) VEK_KAT kategorizace věku	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 18-29	2 30-39	,26	,22	1,000	-,40	,92
	3 40-49	,12	,21	1,000	-,50	,74
	4 50-59	-,50	,22	,319	-1,14	,14
	5 60-69	-1,70*	,23	,000	-2,39	-1,02
	6 70+	-2,10*	,27	,000	-2,89	-1,31
2 30-39	1 18-29	-,26	,22	1,000	-,92	,40
	3 40-49	-,14	,23	1,000	-,82	,53
	4 50-59	-,76*	,24	,018	-1,46	-,07
	5 60-69	-1,96*	,25	,000	-2,70	-1,23
	6 70+	-2,36*	,28	,000	-3,20	-1,53
3 40-49	1 18-29	-,12	,21	1,000	-,74	,50
	2 30-39	,14	,23	1,000	-,53	,82
	4 50-59	-,62	,22	,081	-1,28	,03
	5 60-69	-1,82*	,24	,000	-2,53	-1,12
	6 70+	-2,22*	,27	,000	-3,03	-1,42
4 50-59	1 18-29	,50	,22	,319	-,14	1,14
	2 30-39	,76*	,24	,018	,07	1,46
	3 40-49	,62	,22	,081	-,03	1,28
	5 60-69	-1,20*	,25	,000	-1,92	-,48
	6 70+	-1,60*	,28	,000	-2,42	-,78
5 60-69	1 18-29	1,70*	,23	,000	1,02	2,39
	2 30-39	1,96*	,25	,000	1,23	2,70
	3 40-49	1,82*	,24	,000	1,12	2,53
	4 50-59	1,20*	,25	,000	,48	1,92
	6 70+	-,40	,29	1,000	-1,25	,46
6 70+	1 18-29	2,10*	,27	,000	1,31	2,89
	2 30-39	2,36*	,28	,000	1,53	3,20
	3 40-49	2,22*	,27	,000	1,42	3,03
	4 50-59	1,60*	,28	,000	,78	2,42
	5 60-69	,40	,29	1,000	-,46	1,25

*. The mean difference is significant at the .05 level.

Výsledky analýzy rozptylu tak říkají, že rozdíly v postojích k důležitosti Boha u různých věkových kategorií, který jsme zjistili v datech našeho výběrového souboru, nevznikl díky náhodné výběrové chybě, takže je možné je očekávat i v základním souboru, v daném případě tedy mezi obyvateli ČR, neboť z něj byl vzorek pořízen. Statisticky významné rozdíly však nejsou mezi všemi vzdělanostními kategoriemi navzájem – signifikantně se odlišují respondenti věkových skupin 60-69 a 70-79 od všech skupin věkově mladších. Lidé těchto starších věkových skupin jsou v průměru více přesvědčeni o důležitosti Boha než lidé mladší. Signifikantně se také odlišují skupina 30-39letá od skupiny 50-59 let. Meritorně se ovšem o příliš výrazné rozdíly nejedná, jak jsme viděli v grafu 7.2. Hodnoty všech skupin jsou totiž umístěny v první polovině stupnice, to je v oblasti, která vnímá Boha spíše jako nedůležitého.

Zobecnění postupu jednotlivých kroků v ANOVĚ:

1. Nejdříve zjistíme, zdali jsou v rozptylech signifikantní rozdíly. Není-li F test signifikantní (to když je jeho sig. větší než 0,05), končíme analýzu a dále nepokračujeme.
2. Pokud F signifikantní je (meší než 0,05), provedeme tzv. *post hoc* srovnání, abychom zjistili, které skupiny (kategorie) se od sebe odlišují z hlediska průměrných hodnot. K tomu využijeme buď testu Bonferroniho nebo Tukeyho.

* * *

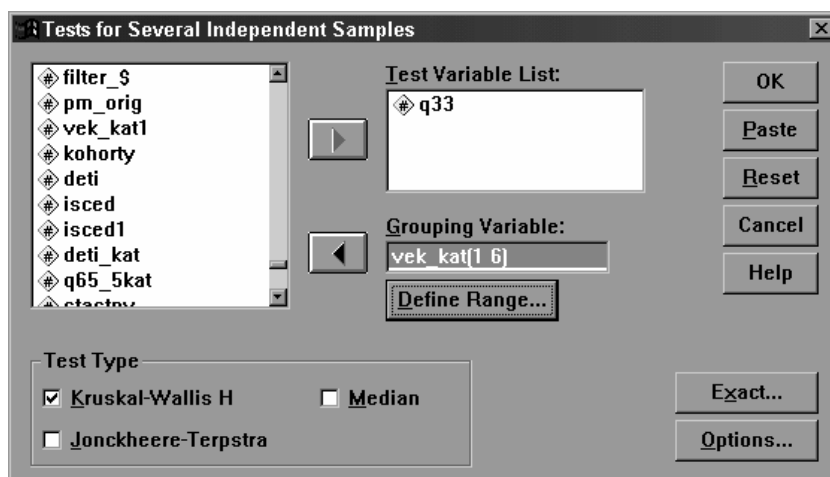
KRUSKAL-WALLISŮV TEST

aneb

NEPARAMETRICKÝ BRATRANEC JEDNOVÝCHODNÉ ANALÝZY ROZPTYLU

Pokud jsou předpoklady pro použití ANOVY výrazně porušeny, měli bychom použít neparametrického ekvivalentu ANOVY, jímž je **Kruskal-Wallisův test**. Ten srovnává ne průměry, nýbrž pořadí (ranks).

Procedura: *Analyze – Nonparametric tests – K Independent Samples – Test variable List: q33 – Grouping Variable: vek_kat – Define Range – Minimum: 1 – Maximum: 6*



Výstup Kruskal-Wallisova testu

Ranks

	VEK_KAT kategorizace věku	N	Mean Rank
Q33 Bůh - důležitost v životě	1 18-29	425	876,87
	2 30-39	317	828,23
	3 40-49	385	865,90
	4 50-59	333	963,57
	5 60-69	263	1127,38
	6 70+	152	1136,78
	Total	1875	

Test Statistics^{a,b}

	Q33 Bůh - důležitost v životě
Chi-Square	85,761
df	5
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: VEK_KAT kategorizace věku

I v tomto testu vycházejí rozdíly mezi jednotlivými skupinami jako statisticky vysoce významné (Signifikance, že můžeme podržet nulovou hypotézu je 0,000). Nulovou hypotézu zamítáme a očekáváme rozdíly v průměrech i v základním souboru. Kruskal-Wallisův test bohužel neumí testovat signifikanci rozdílů mezi jednotlivými skupinami nezávisle proměnné.