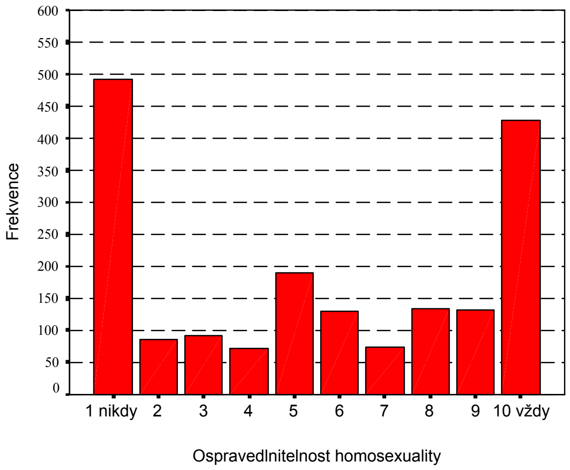
LEKCE 6
SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT

V předchozích příkladech jsme nejednou opakovali, že informace, které pocházejí z tzv. jednorozměrné (univariační, *univariate*) analýzy, jsou sice velmi důležité, ale jelikož popisují ve zhuštěné podobě celý soubor, který je vždy nejrůznějším způsobem strukturován, nepodávají mnohdy informace adekvátní. Někdy mohou dokonce i poněkud zavádět právě z toho důvodu, že ve svém souhrnném vyjádření tuto strukturu zakrývají.[1] Proto prvními skutečnými kroky v analýze dat sociologického výzkumu jsou postupy dvourozměrné (bivariační, *bivariate*), jimiž analyzujeme údaje o dvou vlastnostech (proměnných, znacích).

Příklad P6.1: Vraťme se nyní k našemu příkladu z lekce 4 o tom, zdali se česká populace domnívá, že homosexualita je ospravedlnitelná či nikoliv (proměnná q65a_8). Viděli jsme, že tvar distribuce (obrázek uvádíme ještě jednou) se ani zdaleka neblížil charakteru normálního rozložení, což v sociologických datech je často určitým signálem toho, že v souboru mohou být některé skupiny, které se od ostatních svými názory podstatně liší.

Obr. 6.1: Ospravedlnitelnost homosexuality (ČR 1999)



Přesvědčme se tedy, zdali to tak je. Můžeme totiž očekávat, že postoje k homosexualitě budou pravděpodobně záviset na věku respondentů. Formulujme proto výzkumnou hypotézu (o hypotézách viz rámeček 1): Věk respondenta bude ovlivňovat jeho postoj k homosexualitě a dá se předpokládat, že mladší respondenti budou tolerantnější než respondenti starší. (Zopakujme si, že v jazyku analýzy dat je věk respondenta proměnná nezávislá, zatímco postoj k homosexualitě je proměnná závislá.)

Z této výzkumné hypotézy lze formulovat hypotézu statistickou, která má obvykle podobu hypotézy nulové:

H_0: Věk respondenta nebude mít vliv na postoj k homosexualitě.

Alternativní hypotéze bude znít:
H_a: Postoj mladších respondentů k homosexualitě bude méně tolerantnější než postoj respondentů věkově starších, kteří budou naopak méně tolerantní.

[1] S trochou nadsázky bychom mohli např. konstatovat, že vzhledem k tomu, že někteří obratlovci mají dvě nohy a někteří nohy čtyři, mají obratlovci v průměru tři nohy.

Rámeček 6.1

Pojem *nulové hypotézy* (a *testu významnosti*) zavedl R. A. Fischer (1925). Nulovou hypotézu můžeme buď vyvrátit, nebo ji nemůžeme vyvrátit. Ale to, že nemůžeme vyvrátit nulovou hypotézu, ještě neznamená, že ji můžeme přijmout. Pokud nám data nedovolují vyvrátit nulovou hypotézu, znamená to, že věrohodnost hypotézy se zvyšuje. Vědecké poznání se tedy produkuje ne tím způsobem, že potvrzujeme hypotézy, ale tím, že zjišťujeme, že je nemůžeme vyvrátit. Z tohoto úhlu pohledu je poznání poněkud problematické – to, co považujeme za vědecké pravdy, jsou pouze výroky, které mají nízkou pravděpodobnost, že budou v budoucnu shledány jako nepravdivé. („Nastane-li málo pravděpodobný jev, pak buď hypotéza neplatí, nebo nastal zázrak“ – R. A. Fischer).

Samotný výraz „nulová hypotéza” je pro studenty často nejasný. Užívá se totiž minimálně ve dvou různých významech, které se navíc překrývají. První význam je ten, jak ho zavedl Fischer, tedy, že nulová hypotéza je negací vědecké hypotézy, o níž se domníváme, že je pravdivá. V tomto významu formulujeme hypotézu tak, aby mohla být zamítnuta, čímž se zvyšuje pravděpodobnost pravdivosti její alternativní formy. Druhý význam je význam statistický a ten říká, že výskyt nějakého parametru je nulový. Tento parametr musí být specifikován.

Příklad prvního významu: Předpokládáme, že výskyt delikvence mladistvích závisí na míře, v níž se rozpadávají rodiny. Nulová hypotéza by tedy zněla, že mezi těmito dvěma jevy není žádný vztah. Logika za tímto uvažováním je taková, že pokud hypotéza o neexistujícím vztahu může být zamítnuta, je velmi pravděpodobné, že mezi těmito dvěma jevy nějaký vztah je.

Příklad druhého významu: Hypotézu postavíme tak, že rozdíl mezi mírou delikvence mladistvích z rozpadlých rodin a mírou delikvence mladistvích z rodin úplných je nulový. Zde je parametrem rozdíl mezi dvěma mírami a předpokládáme, že tento parametr je roven nule.

V analýze dat se budeme pohybovat v oblasti nulových hypotéz ve druhém významu, tedy ve statistickém smyslu. Součástí statistického testování hypotéz je také to, že k hypotéze nulové formulujeme hypotézu alternativní (to je přístup *Neyman-Pearsonův*). To, jakým způsobem je nulová a alternativní hypotéza formulována, závisí na povaze vědecké hypotézy a na míře našich vědomostí o zkoumaném fenoménu. Pokud jsou naše znalosti minimální, alternativní hypotéza obvykle nebude mít ani směr ani přesnost. Např. nulová hypotéza o míře anomie mezi voliči stran levicových a voliči stran pravicových bude říkat, že tento rozdíl bude nulový: H_0: AnmL – AnmP = 0

Alternativní hypotéza by pak pravila, že tento rozdíl bude různý od nuly, že tedy míra anomie bude odlišná u obou typů voličů: H_a: AnmL – AnmP ≠ 0

Pokud však budeme mít o problému již nějaké znalosti, takže budeme vědět, že operacionalizovaný koncept anomie měří míru frustrace a vyvázání se ze společnosti, pak naše alternativní hypotéza už může mít směr, neboť budeme předpokládat, že míra anomie bude vyšší u voličů stran levicových než u voličů stran pravicových: H_a: AnmL – AnmP > 0

A pokud bychom byli experty na tento problém, mohli bychom dokonce alternativní hypotézu postavit jako směrovanou a přesnou: H_a: AnmL – AnmP = 2,3

(Zpracováno podle: Henkel, R. A. 1976. *Tests of Significance*. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-400. Beverly Hills and London: Sage Publications)

<u>Řešení:</u>

V datech EVS-ČR1999 máme samozřejmě údaje o věku, které jsme kategorizovali pro jednoduchost do tří velkých věkových skupin: 18–29 let, 30–49 let, 50 a více let. Zjistíme-li, že postoje těchto tří skupin k homosexualitě jsou odlišné, vyvrátíme naši nulovou hypotézu. Odlišnost postojů k homosexualitě zjistíme tak, že vypočítáme průměry (*means*) pro jednotlivé skupiny nezávisle proměnné. K tomu použijeme proceduru *Compare Means–- Means*: (*Analyze* – ***Compare Means*** – ***Means*** – *Dependent List* (q65a_8), *Independent list* (vek_kat1).

Zde je výsledek:

Q65A_8 Homosexualita

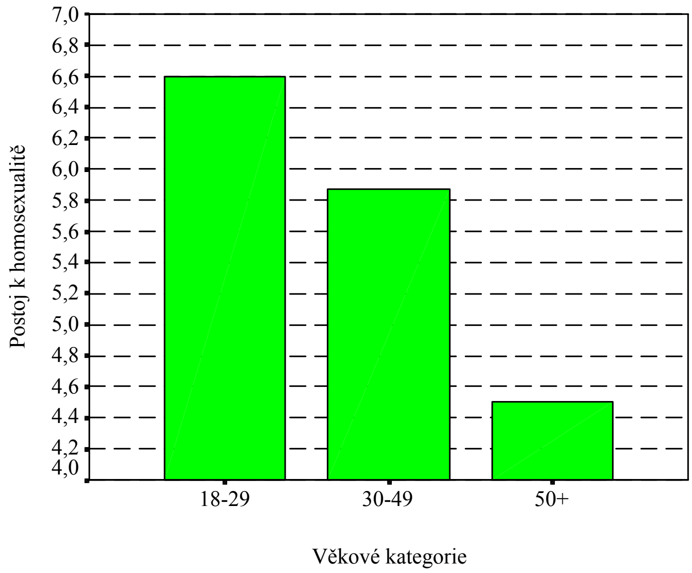
VEK_KAT1 Věkové kategorie	Mean	N	Std. Deviation
1 18-29	6,60	410	3,47
2 30-49	5,88	660	3,42
3 50+	4,51	755	3,48
Total	5,47	1824	3,56

Vidíme, že rozdíly v průměrech zde skutečně existují (nejmladší skupina má průměr 6,6, nejstarší 4,5) a že navíc mají i vzestupný charakter naznačující lineární vztah. Jelikož podle dotazníku víme, že měřicí stupnice byla orientována od naprosté netolerance (hodnota 1) k naprosté toleranci (hodnota 10), pak můžeme konstatovat, že starší věkové skupiny se ve svých odpovědích v průměru umísťovaly do prostoru stupnice, kteří říkala, že homosexualita není ospravedlnitelná. Naopak mladší respondenti byli k homosexualitě tolerantnější. (Zde je příklad toho, jak důležité je mít při analýze dat neustále po ruce dotazník, abychom mohli kontrolovat, jakým způsobem jsou stupnice orientovány, jestli vzestupně nebo sestupně.)

Náš výpočet tedy zjistil rozdíly v průměrech, které jsou docela zřejmé. Ovšem kdy je už rozdíl zřejmý', a tudíž věcně význačný a hodný úvah, a kdy je naopak natolik nízký, takže je v podstatě zanedbatelný a my můžeme tvrdit, že neexistuje, to je v analýze sociologických data kardinální otázka, na niž není lehké odpovědět. (blíže k tomu viz problém meritorní a statistické signifikance, o němž je pojednáno v čítance). Nicméně v našem příkladu rozdíly nasvědčují tomu, že je možné nulovou hypotézu zamítnout a že je možné přijmout hypotézu alternativní. Důvod, proč tento výrok formulujeme poněkud zdrženlivě, si vysvětlíme za chvíli, až budeme hovořit o t-testu a jednofaktorové analýze rozptylu.

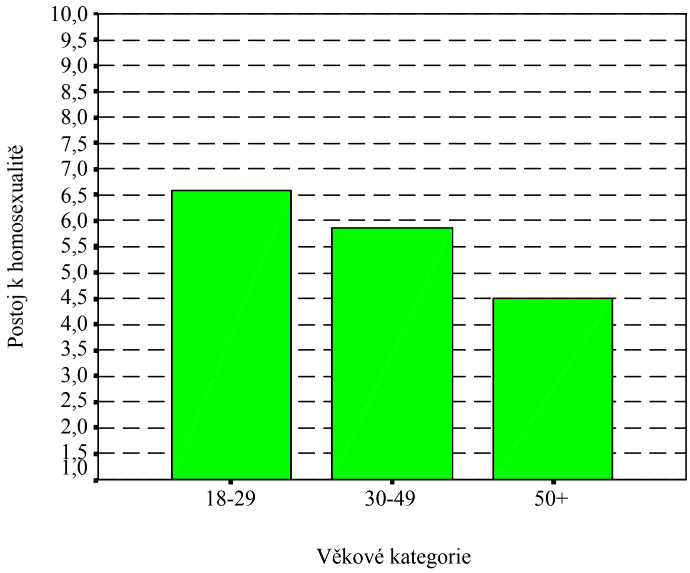
Výsledek našeho výpočtu lze dokumentovat i graficky (dostaneme jej v proceduře *Graphs – Bar – Simple – Define* – zde musíme zakliknout *Other summary function.* A nezapomeňte zrušit požadavek na to, aby SPSS počítal průměry i pro ty, kdo na otázku o náboženském přesvědčení neodpověděli – klikněte na tlačítko *Options* a zrušte požadavek na *Display groups defined by missing values*). Vidíme jej na obr. 6.1a.

Obr. 6.1a: Postoje k homosexualitě podle věkových skupin (ČR 1999)



Rozdíly, které vykreslila grafická procedura programu SPSS, vypadají velmi impresivně. Avšak pozor, je to jistý optický klam, který je způsoben měřítkem osy Y. Všimněme si, že SPSS na ni vložil hodnoty od 4 do 7, avšak naše měřicí stupnice se pohybuje od 1 do 10. Pokud na ni vyneseme všech deset bodů měřicí stupnice (a ne jenom body v rozmezí 4,0–7,0), bude vypadat obrázek jinak, optické rozdíly se značně sníží:

Obr. 6.1b: Upravený obrázek postojů k homosexualitě podle věkových skupin (ČR 1999)



Věková struktura výběrového souboru je tedy pravděpodobně jednou z příčin, proč rozložení postoje k homosexualitě nezískalo tvar zvonovité křivky.

Analyticky by nás v této souvislosti mohlo ještě zajímat, zdali vztah mezi věkem a postojem k homosexualitě bude nalezen i mezi muži a ženami. Výpočet je stejný jako v předchozím případě, pouze u *Independent List* přidáme další *Layer* (vrstvu), do níž vložíme proměnnou pohlaví (q84).

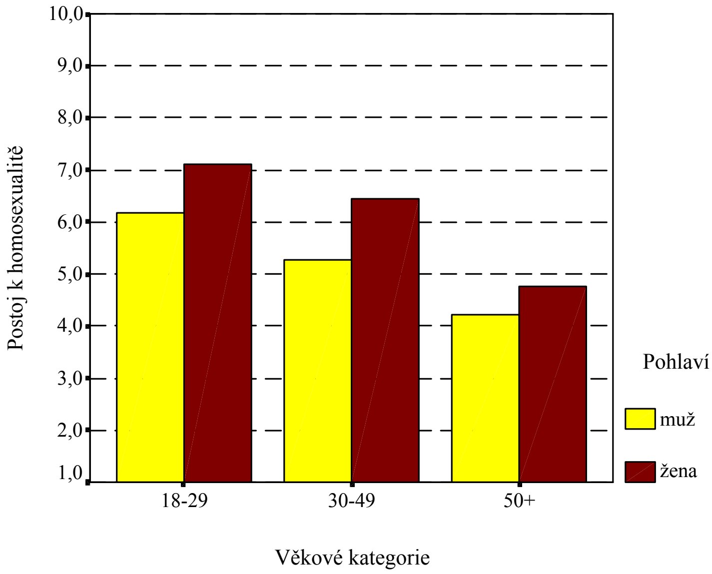
Výsledek:

Q65A_8 Homosexualita

VEK_KAT1 Vekové kategorie	Q84 Pohlaví	Mean	N	Std. Deviation
1 18-29	1 muž	6,18	226	3,56
	2 žena	7,12	184	3,29
	Total	6,60	410	3,47
2 30-49	1 muž	5,28	322	3,43
	2 žena	6,44	338	3,31
	Total	5,88	660	3,42
3 50+	1 muž	4,21	340	3,47
	2 žena	4,75	415	3,47
	Total	4,51	755	3,48
Total	1 muž	5,10	887	3,56
	2 žena	5,82	937	3,52
	Total	5,47	1824	3,56

Vidíme, že celkově jsou ženy tolerantnější k homosexualitě než muži a že tento rozdíl existuje ve všech třech věkových kategoriích. Zajímavý je ale fakt, že největší rozdíl v postojích je u střední věkové kategorie (1,16 bodu), a nejmenší není u nejmladší věkové kategorie, jak by se dalo předpokládat, ale u kategorie nejstarší (0,54 bodu). Ilustruje to i graf 6.2 – získáte ho stejně jako v předchozím případě, pouze při zadávání grafu budete požadovat ne *Simple*, nýbrž *Clustered* formát.

Obr. 6.2: Postoje k homosexualitě podle věkových skupin a pohlaví (ČR 1999)



Statistické usuzování o shodě dvou nezávislých průměrů: t-test

V předchozím příkladu jsme zjistili, že tolerance k homosexualitě závisí na věku, že podle našich výběrových dat byli mladší respondenti tolerantnější než ti starší. V tomto kontextu si musíme položit důležitou otázku, zdali rozdíl zjištěný ve výběrovém souboru, je možné očekávat také v souboru základním. Jak totiž už dobře víme, je variabilita mnohdy způsobena prostě tím, kteří respondenti a s jakými vlastnostmi se nám dostali do výběru a že kdybychom těchto výběrů učinili mnoho, nacházeli bychom různé hodnoty průměrů.

Otázky, které si při těchto úlohách klademe, je možno formulovat následovně: **Odráží pozorovaná diference v průměrech stav, který existuje i v základní populaci, nebo je to jen důsledek výběru, který způsobil, že se do souboru dostalo příliš mnoho respondentů s určitými charakteristikami? Přinesl by jiný výběr jiné průměrné hodnoty nebo by dokonce způsobil, že rozdíl v průměrech nebude žádný?**

Proto při statistické analýze dat – pokud jsou naše data reprezentativní, samozřejmě – se vždy snažíme zjistit, zdali výsledky z výběru lze očekávat i v základním souboru. V jazyce testování hypotéz říkáme, že testujeme hypotézu, že zjištěný rozdíl je způsoben výběrovou chybou a že v populaci není rozdíl v průměrech žádný. Testujeme tedy nulovou hypotézu o neexistenci rozdílu.

Příklad 6.2: V našem příkladě jsme viděli, že rozdíl v toleranci k homosexualitě mezi nejmladšími respondenty ve věku 18–29 let a nejstaršími respondenty ve věku 50+ byl relativně velký (6,60 oproti 4,51). Otestujme tedy nulovou hypotézu, že tyto dva průměry budou v populaci shodné, že tedy v populaci mezi těmito průměry nebude žádný rozdíl. Řečeno jinými slovy, tuto skutečnost ověříme prostřednictvím testu hypotézy o dvou nezávislých výběrech (*two independent-samples t test*), neboli t-testu pro dva nezávislé průměry.

Řešení:

V SPSS si zadáme proceduru *Statistics — Compare means —* ***Independent-Samples T Test*** a dosadíme do ní příslušné proměnné: naší *Test Variable* je q65a_8 a *Grouping Variable* je vek_kat1, *Define groups – Group 1* je 1 (tedy věková skupina 18–29 let) a *Group 2* je 3 (to je věková skupina 50+).

Výsledek výpočtu uvádí tabulka níže (pozor, tabulku jsme pro lepší čtení editovali a přehodili jsme v *Pivot – Transpose Rows and Columns* sloupce a řádky).

Independent Samples Test

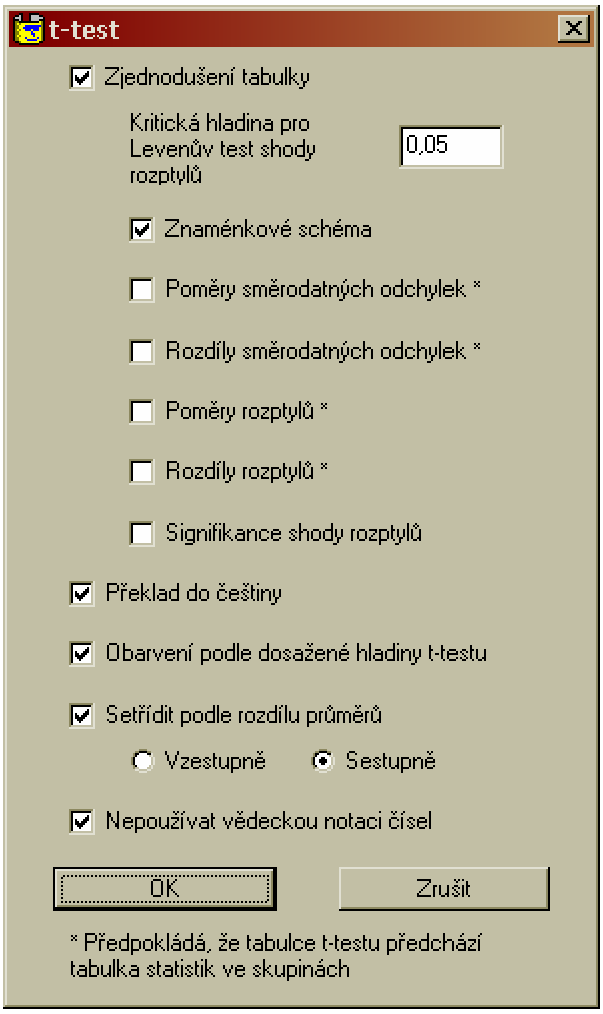
			q65a_8 Homosexualita	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		,065	
	Sig.		,798	
t-test for Equality of Means	t		9,787	9,796
	df		1162	839,091
	Sig. (2-tailed)		,000	,000
	Mean Difference		2,088	2,088
	Std. Error Difference		,213	,213
	95% Confidence Interval of the Difference	Lower	1,669	1,669
		Upper	2,506	2,506

Jak tuto tabulku číst, abychom zjistili, jak to vlastně je? Nejdříve se rozhodneme, zdali budeme hledat ve sloupci *Equal variances assumed*, nebo zda ve sloupci *Equal variances not assumed*. Poznáme to podle signifikance charakteristiky F. Je-li rovna nebo vyšší než 0,05, pohybujeme se v prvním sloupci (To proto, že podle statistické konvence je 0,05 kritická hodnota a vyšší číslo prostě říká, že držíme nulovou hypotézu o neexistenci rozdílu v rozptylech). Pokud by F sig. vyšlo nižší než 0,05, pokračovali bychom s údaji ze druhého sloupce. V našem případě je F sig. 0,796 (tedy mnohem vyšší než 0,05), pracujeme proto s údaji ze sloupce *equal variance assumed*. Zde hledáme signifikanci statistiky t. Její vypočtená hodnota 0,000 je velmi nízká, což nám velí zamítnout nulovou hypotézu. Můžeme proto vyslovit závěr, že nalezený rozdíl v toleranci k homosexualitě mezi mladšími a staršími respondenty, jenž činil 2,1 bodu, nevznikl náhodnou variabilitou při volbě výběrového souboru, nýbrž je produktem nějakého systematického působení. Velikost rozdílu (jak napovídají poslední dva řádky tabulky uvádějící údaj o 95% intervalu spolehlivosti) by se s 95% jistotou pohybovala v populaci mezi 1,68–2,51 bodů.

I pro t-test existuje skript. Spustíme jej tak, že tabulku, v níž SPSS zobrazuje t-test (pozor, tabulka musí být v původní podobě, řádky a sloupce nesmí být přehozeny), označíme kliknutím a pustíme na ni skript *ttest*. Skript slouží především ke zjednodušení tabulky Studentova T-testu pro dva nezávislé výběry. Podle zadaného kriteria vybere do nové tabulky pouze relevantní řádky na základě výsledku Levenova testu shody rozptylů ve skupinách. Kriteriem je hladina významnosti Levenova testu, které si uživatel volí. Dále skript umí přeložit tabulku do češtiny, obarvit významné řádky, či setřídit výsledky podle rozdílu průměrů. Všechny funkce skriptu uživatel ovládá z úvodního dialogu, který se zobrazí při spuštění skriptu (viz obrázek 6.3). Zde je možno vybrat z nabízených funkcí. Doporučujeme pracovat s implicitním nastavením, jak je uvedeno na obr. 6.3. Zaškrtnutá políčka znamenají následující:

- **Zjednodušení tabulky**: Skript převede výsledky do nové tabulky, která obsahuje pouze řádky vybrané na základě signifikance Levenova testu shody rozptylů.
- **Kritická hladina pro Levenův test shody rozptylů**: Kriterium pro výběr relevantních řádků. Je-li spočtená signifikance Levenova Testu menší než zadaná mez, do nové tabulky se zahrne řádek s T-testem, který nepředpokládá stejné rozptyly. V opačném případě se do zjednodušené tabulky dostane řádek s variantou T-testu pro shodné rozptyly ve skupinách.
- **Znaménkové schéma**: Názorné zobrazení dosažené hladiny významnosti pomocí znaménkového schématu. Je-li signifikance větší než 0,05 vyskytuje se v daném řádku nula. Při signifikanci od 0,05 do 0,01 se zobrazí jedno znaménko, při signifikanci od 0,01 do 0,001 dvě znaménka a je-li signifikance nižší než 0,001 zobrazí se znaménka tři. Typ znaménka je shodný se znaménkem rozdílu průměrů.

Obr. 6.3: Dialogové okno pro formát výstupu skriptu t-testu



- **Překlad do češtiny**: Převedení sloupcových a řádkových popisků (mimo popisků proměnných) do českého jazyka.
- **Obarvení podle dosažené hladiny testu**: Barevné zvýraznění signifikantních řádků. K obarvení řádků skript používá celkem tři barvy (zelená, žlutá, červená) podle dosažené signifikance testu (5%, 1% a 0.1%), podobně jako znaménkové schéma.
- **Setřídit podle rozdílu průměrů**: Setřídění řádků v tabulce **sestupně** či **vzestupně** podle rozdílu průměrů u jednotlivých testovaných proměnných.
- **Nepoužívat vědeckou notaci čísel**: Potlačení zobrazování číselných údajů pomocí mantisy a exponentu.

Výsledek je tento:

t-test shody středních hodnot pro nezávislé výběry

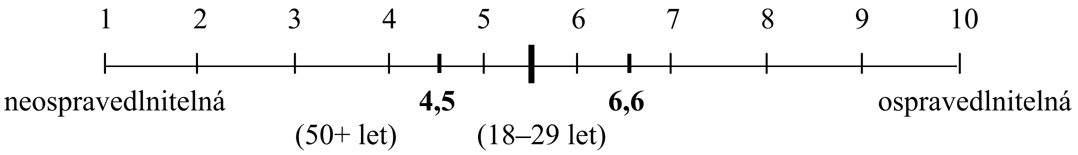
:

Proměnné	Statistiky					
	Znaménkové schéma	Sig. (Obous tranná)	Rozdíl průměrů	Standardní chyba rozdílu průměrů	Dolní mez[a]	Horní mez[a]
q65a_8 Homosexualita	+++	,000	2,088	,213	1,669	2,506

a. 95% Interval spolehlivosti pro rozdíl středních hodnot

Varování
Při relativně velkých vzorcích s nimiž pracují reprezentativní sociologické výzkumu (pohybují se kolem tisícovky respondentů) vycházejí i poměrně malé rozdíly v průměrech jako statisticky významné (signifikantní) – proč asi? Z tohoto důvodu nemůžeme dělat ze statistické signifikance jakýsi fetiš. To, že můžeme zamítnout nulovou hypotézu ještě neznamená, že jsme objevili velký nebo důležitý rozdíl. **Statistická významnost totiž ještě neznamená významnost věcnou, meritorní.** Kdykoliv jsme nadšeni z faktu, že nám t-test vyšel jako statisticky singnifikantní, zabrzděme a hodnoťme zjištěný rozdíl z jeho faktické, věcné stránky. Je rozdíl dvou bodů v toleranci na desetibodové stupnici skutečně tak velký, jak to statisticky vypadá? Zobrazme si tuto otázku graficky.

Homosexualita



Vidíme, že rozdíl de facto není nijak dramaticky velký. Je ale pravda, že skupina starších respondentů se umístila do „netolerantní“ poloviny stupnice, zatímco ta mladší do její „tolerantní“ části (a zajímavé je, že skóre obou skupin je téměř symetricky rozloženo od poloviny stupnice).

Pokud pracujeme s malým výběrovým souborem, měli bychom se mít při formulaci závěrů rovněž na pozoru, zvláště když nemůžeme podle výsledků t-testu zamítnout nulovou hypotézu ačkoliv rozdíl v průměrech vychází relativně velký. Z toho vyplývá závěr, že pokud uvažujete o provedení vlastní výzkumu, např. pro svou diplomovou práci, je lepší plánovat větší výběrový soubor. Výmluvným faktem nechť je pro vás vztah o velikosti výběrové chyby a velikosti výběrového souboru, který je uveden na konci vaší čítanky. Proto ve studentské sociologické práci by měla být velikosti 100 respondentů minimem, ale uvažujte raději o dvou až třech stovkách.

PARAMETRICKÉ A NEPARAMETRICKÉ TESTY

Upozornění 1: T-test je tzv. **parametrický test**, to znamená, že předpokládá, že proměnná, s níž pracujeme, pochází z populace, v níž je normálně rozložena. Měli bychom tedy, před tím, než t-test použijeme, testovat, zdali proměnná, jejichž průměr budeme v dané úloze počítat, má v populaci normální rozložení. Způsob, jak to uděláme, jsme si ukázali v lekci 3 – prostřednictvím Q-Q grafu a Kolmogorov-Smirnovova testu. Z lekce 3. také víme, že tyto testy nás varovaly, že tato proměnná není s velkou pravděpodobností v populaci normálně rozložena.

V případě, že máme vážné pochybnosti o tom, že rozložení naší proměnné nesplňuje podmínku normálního rozložení, musíme pro test významnosti dvou nezávislých průměrů použít tzv. *distribution-free* test, tedy test nezávislý na rozložení proměnné, **neparametrický test**. Alternativou pro parametrický t-test je tzv. ***Mann-Whitney test***.

Procedura:

Analyze — Nonparametric Tests — 2 Independent Samples

<u>Výpočet</u>

Při zadávání výpočtu postupujeme stejně jako v t-testu: do okna *Test Variable List* vložíme závisle proměnnou, jejíž průměry srovnáváme (q65a_8), do okna *Grouping Variable* vložíme nezávisle proměnnou (vek_kat1), do *Define Groups* definujeme dvě srovnávané skupiny (v našem případě 1 a 3).

Ranks

	VEK_KAT1 Vekové kategorie	N	Mean Rank	Sum of Ranks
Q65A_8 Homosexualita	1 18-29	411	731,69	300725,50
	3 50+	788	531,31	418674,50
	Total	1199		

Všimněte si, že co zde srovnáváme, nejsou průměry, nýbrž průměrné pořadí (*Mean Rank*). Co to znamená? SPSS vytvoří pořadí respondentů podle jejich hodnoty v proměnné q65a_8 a pak v jednotlivých kategoriích vypočítá z těchto pořadí průměr.

Test Statistics[a]

	Q65A_8 Homosexualita
Mann-Whitney U	103859,500
Wilcoxon W	383237,500
Z	-9,208
Asymp. Sig. (2-tailed)	**,000**

a. Grouping Variable: VEK_KAT1 Vekové kategorie

Signifikance je menší než 0,05, proto zamítáme nulovou hypotézu o shodnosti průměrů a přijímáme hypotézu alternativní o tom, že rozdíly v průměrech mezi věkovými kategoriemi nevznikly náhodou, výběrovou chybou, ale že je můžeme prakticky s jistotou očekávat i v základním souboru.

Jak vidět, výsledky parametrického i neparametrického testu vyšly shodné, což nám dává velkou důvěru v jejich spolehlivost.

V některých případech si nemusíme být zcela jisti, zdali použít parametrický či neparametrický test. Například Q-Q graf naznačuje normalitu rozložení, statistický nikoli. Co dělat, jaký test použít? V takové situaci záleží na velikosti vzorku. Pokud máme velký vzorek, i velmi malé odchylky od normality dávají statisticky nízké hodnoty signifikance v Kolmogorov-Smirnovově testu. To je jen další ukázka situace, kdy v případě velkých vzorků vycházejí i malé rozdíly jako statisticky signifikantní (významné), byť s nepříliš praktickým významem. Takže máme-li velký vzorek a distribuce (rozložení) hodnot proměnné není příliš vzdáleno normalitě, není třeba si dělat starosti — v takovém případě je možno t-testu použít. T-test je totiž dost robustní technika, aby malé odchylky od normality nějak podstatně ovlivnily výsledek. V případě velkých výběrů navíc platí, že nevadí ani poměrně velké odchylky od normality.

V případech, kdy si skutečně nejsme jisti, zdali použít parametrický nebo neparametrický test, zlaté pravidlo zní: **použijte obou!** Pokud v obou druzích testu dojdete ke stejnému závěru, je vše v naprostém pořádku. Pokud výsledky v neparametrickém testu nejsou signifikantní, avšak v testu parametrickém jsou, musíme se snažit zjistit důvod: Máme v souboru několik dat, která mají výrazně nižší či vyšší hodnotu než ostatní (tedy outliers)? Pokud ano, pak samozřejmě ovlivňují hodnotu průměru, což může být příčina oné nejednoznačnosti. A tudíž je v tomto případě dobré užít neparametrického testu.

Pokud problém spočívá v nenormalitě rozložení, data transformujte – např. tak, že je všechna odmocníte nebo tak, že je zlogaritmujete (procedura *Compute*). Pokud se taková transformace povede a rozložení se změní na přibližně normální, **použijte raději parametrický test, který je silnější,** což znamená, že vede častěji k zamítnutí nulové hypotézy**.** Obecně totiž platí, že neparametrické testy mají ve srovnání s testy parametrickými menší sílu testu, jsou tedy „konzervativní“ (Hendl 2003:196).

JEDNOSTRANNÁ A DVOUSTRANNÁ SIGNIFIKANCE

Upozornění 2: Všimněte si, že Mann-Whitneyův test počítá signifikanci dvoustrannou (*2-tailed*). To znamená, že kritická hodnota oboru přijetí či zamítnutí hypotézy je rozložena po jeho obou stranách. Dvoustrannou signifikanci používáme tehdy, pokud nevíme (nebo pokud nepředpokládáme), zdali rozdíl v populaci by měl být v obou skupinách odlišný. Dvoustranné testy se používají častěji než jednostranné a jsou v mnoha testech nastaveny jako default, tedy implicitně.

V našem případě ovšem předpoklad o odlišnosti sledovaných skupin máme: měli jsme hypotézu, že mladší věková skupina by měla být k homosexualitě tolerantnější než skupina věkově mnohem starší. Z tohoto důvodu je třeba tento dvoustranný test „překlopit“ do testu jednostranného. Uděláme to jednoduše: spočítanou *2-tailed* signifikanci podělíme 2. V našem příkladu tento krok nemá smysl, neboť výsledek dvoustranného testu je již sám o sobě tak nízký (0,00), že dělit jej ještě dvěma nemá význam [2]. Pokud by ale dvoustranný test signifikance měl hodnotu 0,095, což je výsledek, který nám velí podržet nulovou hypotézu, ale my měli důvod pro jednostranný test, jeho dvěma podělená hodnota by byla 0,095/2 = 0,047. A to je již výsledek, který je podkladem pro zamítnutí nulové hypotézy.

Rozvaha o jednostranném či dvoustranném testu statistické významnosti platí samozřejmě i pro parametrický t-test. I tam se musíme na základě našich znalostí analyzované problematiky rozhodnout, zdali vypočtenou signifikanci budeme či nebudeme dělit dvěma. Jelikož při analýze máme většinou radost, když můžeme zamítnout nulovou hypotézu, měli bychom se snažit formulovat hypotézy jednostranné, které jsou silnější. Nicméně formulace jednostranných hypotéz vyžaduje, abychom zkoumaný problém měli již před výzkumem pokud možno co nejlépe teoreticky nastudovaný a byli tak schopni smysluplné jednostranné testy formulovat.

OBECNÝ ZÁVĚR

V této lekci jsme řešili úlohu, v níž jsme zjišťovali, zdali rozdíly mezi průměry dvou podsouborů. jsou statisticky významné, to je ptali jsme, zdali rozdíl zjištěný v průměrech našeho výběrového souboru je možné také očekávat v souboru základním. Řešili jsme tedy úlohu tzv. induktivní statistiky. Při rozhodování, zdali podržet, nebo zamítnout nulovou hypotézu, platí základní pravidlo:

<u>Zlaté pravidlo pro induktivní statistiku:</u>

vysoká hodnota testu signifikance (tj. $\alpha > 0,05$) — držíme nulovou hypotézu
nízká hodnota testu signifikance (tj. $\alpha \leq 0,05$) — zamítáme nulovou hypotézu

* * *

ONE-SAMPLE T-TEST:

V úlohách o průměru je možno ale také testovat ještě jiný případ. Jedná se o proceduru *One-Sample T Test*. To je úloha, při níž známe průměr nějakého znaku v základním souboru (tzv. populačního para-

[2] Aby nedošlo k mýlce. Víme, že nulu nelze dělit. Avšak výsledek výpočtu 0,00 je zaokrouhlen. Kdybychom povolili větší počet desetinných míst, možná bychom zjistili, že jeho hodnota 0,0024, která už samozřejmě technicky dělit lze. Věcný význam ovšem takové dělení nemá.

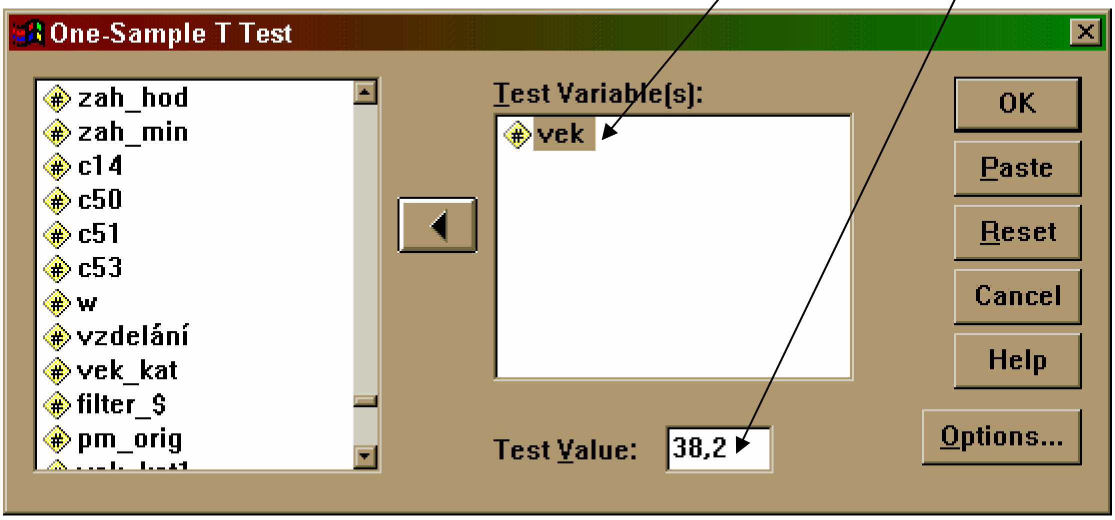
metru) a ptáme se, zdali náš výběr pochází z této populace. Srovnáváme tedy hodnotu výběrového průměru nějaké proměnné se známou hodnotou této proměnné v populaci.[3]

Příklad:
Průměrný věk našeho výběru EVS-ČR1999 je 45,69, směrodatná chyba je 0,39 a 95% interval spolehlivosti je 44,93–46,45 (zkontrolujte si tyto údaje vlastním výpočtem pro proměnnou *vek*). S 95% jistotou tedy můžeme předpokládat, že v základním souboru se bude podle našich dat průměr pohybovat v rozmezí 44,9 – 46,5. Jenže my z dat demografické statistiky víme, že průměrný věk populace ČR byl v době konání výzkumu EVS roce 1999 38,2 roku. To, že průměr základního souboru (38,2) leží mimo vypočtený interval spolehlivosti je signálem, že obě hodnoty se budou statisticky významně lišit. Zjistěme to ale přesně. Úloha tedy zní: Pochází z hlediska věku náš výběrový soubor z této základní populace?

Procedura:

Analyze — Compare means — ***One-Sample T Test***

V dialogových oknech vložíme do *Test variable(s)* proměnnou *vek* a do okna *Test Value* známou hodnotu znaku v populaci, v našem případě průměrný věk populace 38,2.



Výsledek:

One-Sample Test

	Test Value = 38.2					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference: Lower	Upper
VEK	19,406	1900	,000	7,49	6,73	8,25

Hodnota vypočtené významnosti statistiky *t* (Sig. 2-tailed) je velmi nízká (0,000), což znamená, že musíme zamítnout nulovou hypotézu, že náš výběrový soubor pochází z populace o průměru 38,2 roku. Je to tedy výsledek, který již naznačoval interval spolehlivosti. Řečeno jinými slovy, průměrný věk našeho výběrového souboru je natolik odlišný od průměru populace, že je zřejmé, že náš výběr

[3] Popřípadě známe hodnotu z minula a chceme zjistit, zda nedošlo nyní ke změně.

z této populace nepochází. Co to znamená? Máme snad špatně vybraný soubor? Ale vždyť náš soubor je reprezentativní pro celou populaci ČR! Nebo snad reprezentativní není?

Nebojte, reprezentativní je. Nesrovnalost zde vznikla tím, že náš soubor EVS-ČR1999 není z hlediska věku reprezentativní pro celou populaci ČR, nýbrž pouze pro její *dospělou* populaci, to je starší 18 let. A jelikož průměrný věk populace se počítá ze všech žijících osob, musí být nutně náš výběrový soubor v průměru starší než populace ČR. To, že test nám řekl, že náš výběrový soubor nepochází z populace ČR je tedy správně. Test s adekvátním populačním průměrem si proveďte za domácí úlohu.

JAK TESTOVAT NULOVOU HYPOTÉZU O SHODĚ NĚKOLIKA POPULAČNÍCH PRŮMĚRŮ: (ANALÝZA ROZPTYLU)

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 deklarace o politické orientaci (měřené na kontinuu od levice k pravici) podle věkových skupin nebo podle vzdělání apod. Proto si nyní ukážeme, jak testovat nulovou hypotézu, že průměry různých skupin jsou v populaci stejné. Procedura, 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* – zde se předpokládá, že odlišnost vzniká působením náhody), jednak analyzuje, jak mnoho se liší průměry **mezi** skupinami (*between-groups variability* – zde předpokládáme, že tyto rozdíly vznikají díky působení nezávisle proměnné). 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.

Při hledání statistické významnosti v rozdílech v politické orientaci u jednotlivých vzdělanostních kategoriích použijeme nejjednodušší variantu *ANOVY*, tzv. jednofaktorovou (*one-way analysis of variance*), neboť srovnáváme, jak se liší hodnoty závisle proměnné u skupin (kategorií) jedné nezávisle proměnné. Této proměnné se říká faktor – proto jednofaktorová analýza 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 požít tzv. dvoufaktorovou analýzu rozptylu (*two-way analysis of variance*), neboť nezávisle proměnné (nebo též faktory) máme dvě, to je věkové skupiny a vzdělání. Jaké předpoklady musí být splněny, abychom mohli *ANOVu* použít?

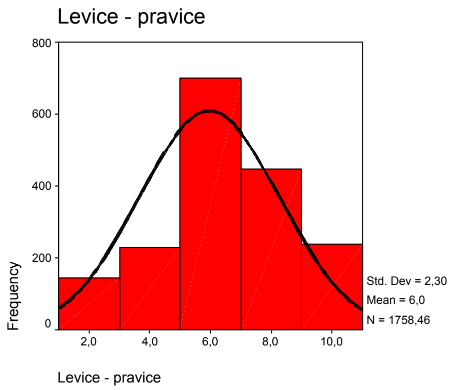
1. Jednotlivá pozorování musí být na sobě nezávislá. Tento předpoklad je v sociologických šetřeních vždy bez problémů, výzkumné designy s opakovaným měřením stejných subjektů (respondentů) nejsou příliš časté.
2. Rozložení závisle proměnné v populaci je normální. Tento předpoklad se v praxi často zanedbává, v úvahu musí být brán pouze tehdy, pokud jsou naše data rozložena extrémně nenormálně. Rozložení v populaci navíc často neznáme, takže je odhadujeme z rozložení výběrového. U velkých souborů by nenormalita rozložení neměla působit problémy.
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 homogenitu rozptylů.

Pokud data tyto předpoklady nesplňují, 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, jíž je v tomto případě neparametrický test Kruskal-Wallisův.

Příklad 6.3: Je v souboru EVS-ČR1999 přihlášení se k levici, středu či pravici (*q53*) vnímána rozdílně věkovými skupinami mladých respondentů, respondentů středního věku a staršího věku (*vekkat2*)? A je možné tuto odlišnost také očekávat v základním souboru?

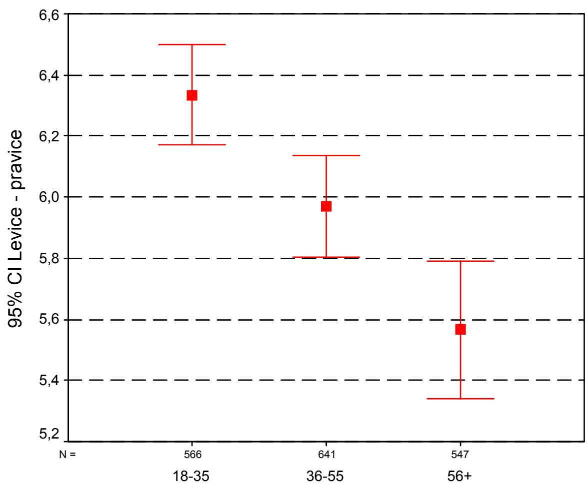
<u>Řešení</u>

Nejdříve se podívejme, jaký tvar má rozložení závisle proměnné, zdali není porušen předpoklad normality rozložení. U nezávisle proměnné (faktoru) nás rozložení nezajímá, statisticky to nehraje žádnou roli, navíc faktor je často proměnná nominální.



Jak vidíme, závisle proměnná je přibližně normálně rozložena. Nyní se podívejme, jak se průměry ve věkových skupinách odlišují graficky. Použijme k tomu graf intervalů spolehlivosti pro jednotlivé skupiny:
(*Graphs – Error Bar – Simple – Define – Variable (q53) – Category Axis (vekkat2 – Confidence interval for Mean 95 %* – v *Options* nezapomeňte zrušit kategorii *Missing values*)

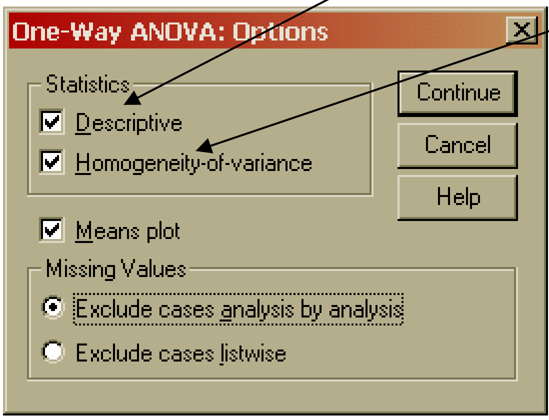
Obr. 6.3



Už tento obrázek mnohé naznačuje o signifikanci rozdílů. Průměry těch skupiny, jejichž „vousy“ se v grafu nepřekrývají budou statisticky významně odlišné. Přesvědčme se o tom i výpočtem pro analýzu rozptylu:

Compare Means – ***One-Way Anova*** *– Dependent list(q53), Factor (vekkat2) – Options*

Nezapomeňte v *Options* zvolit deskriptivní statistiky a test homogenity variance.



Výsledky vypadají takto:

Descriptives

Q53 Levice - pravice

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
					Lower Bound	Upper Bound
1,00 18-35	566	6,33	1,991	,084	6,17	6,50
2,00 36-55	641	5,97	2,151	,085	5,80	6,14
3,00 56+	547	5,57	2,685	,115	5,34	5,79
Total	1754	5,96	2,304	,055	5,85	6,07

Test of Homogeneity of Variances

Q53 Levice - pravice

Levene Statistic	df1	df2	**Sig.**
29,765	2	1751	**,000**

Test homogenity rozptylu (to je Levenův test) vychází statisticky signifikantní, což znamená, že rozptyly jednotlivých věkových skupin nejsou stejné (homogenní), takže je porušen jeden z předpokladů pro analýzy rozptylu.[4] Tato situace by naznačovala, že v analýze není možné pokračovat. Naštěstí analýza rozptylu je i při nesplnění předpokladu poměrně robustní, takže i když signifikance vyjde menší než 0,05, není ještě třeba zoufat. Druhé pravidlo totiž zní, že pokud poměr mezi nejpočetnější kategorií a kategorií nejméně početnou je menší než 1,5, je vše v pořádku. Naše nejpočetnější kategorie má N=641, nejméně početná N= 547, poměr je tedy 1,17. Třetí pravidlo říká, že pokud je poměr mezi největším a nejmenším rozptylem ve skupinách maximálně 4:1, lze analýzu rozptylu použít. I toto pravidlo je neporušeno.

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

[4] Zopakujme si, že pokud by tento předpoklad neměl být porušen, musela by vyjít signifikance Levenova testu vyšší než 0,05.

ANOVA

Q53 Levice - pravice

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	164,001	2	82,001	**15,705**	**,000**
Within Groups	9137,273	1750	5,221		
Total	9301,274	1752			

Důležitými údaji v tabulce ANOVY 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 82,001 / 5,221 = 15,705.

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 rozložením[5] (to za nás udělá 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 nestejné, 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 přece zjistit, mezi kterými konkrétními skupinami se tento rozdíl objevuje. Možná se odlišují všechny tři skupiny 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 — ∨Bonferroni – Significance level: ,05*

(Pozn.: Z nabídky různých testů se doporučuje užívat buď testu Bonferroniho nebo Tukeyho. LSD test je příliš liberální a Scheffeho test je naopak příliš konzervativní – dlouho mu trvá, než dovolí zamítnout nulovou hypotézu, což se nám samozřejmě nelíbí.)

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 se srovnávají průměrné hodnoty respondentů ve věku 18-35 let se zbylými dvěma věkovými kategoriemi. Ze všech údajů, které jsou v tabulce mnohonásobného srovnání uvedeny, jsou pro interpretaci nejdůležitější hodnoty v druhém sloupci (označeném *Mean Difference*), to je 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.

[5] 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.

Multiple Comparisons

Dependent Variable: Q53 Levice - pravice
Bonferroni

(I) VEKKAT2	(J) VEKKAT2	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1,00 18-35	2,00 36-55	,36*	,132	**,017**	,05	,68
	3,00 56+	,77*	,137	**,000**	,44	1,10
2,00 36-55	1,00 18-35	-,36*	,132	**,017**	-,68	-,05
	3,00 56+	,40*	,133	**,008**	,08	,72
3,00 56+	1,00 18-35	-,77*	,137	**,000**	-1,10	-,44
	2,00 36-55	-,40*	,133	**,008**	-,72	-,08

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

Výsledky analýzy rozptylu tak v našem příkladě říkají, že rozdíly v politické orientaci 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. Meritorně se ovšem o příliš výrazné rozdíly nejedná, hodnoty námi srovnávaných skupin se totiž pohybují v poměrně úzkém intervalu 5,6-7,3.

Zjistěme ještě, jaká je síla efektu nezávisle proměnné na proměnnou závislou. Vypočítáme ji, stejně jako v případě t-testu, prostřednictvím umocněné ety. Vzorec pro výpočet zní:

$$\text{umocněná eta} = \frac{\text{Sum of squares between groups}}{\text{Total sum of squares}}$$

Pro tyto hodnoty musíme jít do tabulky anovy: Sum of squares between groups je 164, Total sum of squares je 9301,3. Poměr tedy je 164/9301,3 = 0,02, což je podle přijímaného pravidla velmi nízký (de facto nulový) efekt.

Zobecnění postupu jednotlivých kroků v ANOVĚ:

1. Nejdříve zjistíme, zdali jsou v rozptylech signifikantní rozdíly. Není-li F test signifikatní (to když je jeho sig. větší než 0,05), končíme analýzu a dále nepokračujeme. Konstatujeme pak, že s 95% pravděpodobní nelze v populaci očekávat, že mezi sledovanými skupinami bude v měřeném průměru rozdíl.
2. Pokud F signifikantní je (menší 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.[6]

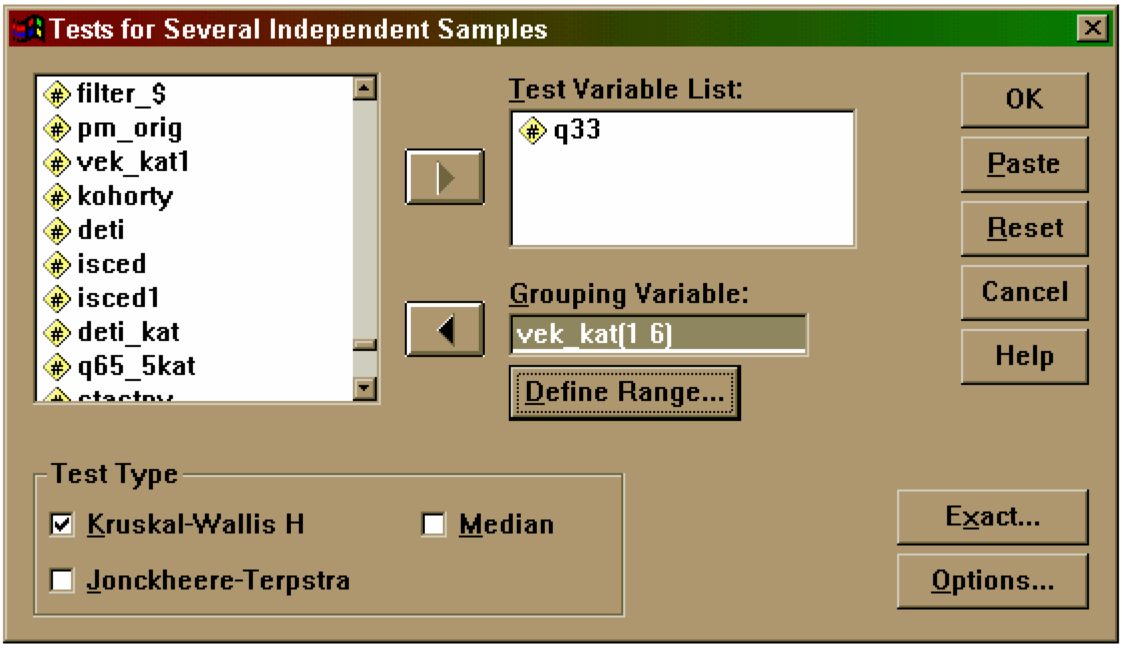
* * *

[6] Pro zvídavé ještě poznamenejme, že dle výsledku Leveneho testu se rozhodujeme, zda použít vícenásobná porovnání pro stejné či nestejné rozptyly. V našem příkladu jsme pro jednoduchost použili porovnání pro stejné rozptyly, i když Leveneho test nám radí (Sig.<0,05) použít vícenásobná porovnání pro nestejné rozptyly. Za domácí úkol si můžete toto zkusit s metodou Dunnet T3. Jsou výsledky srovnatelné?

KRUSKAL-WALLISŮV TEST
aneb
NEPARAMETRICKÝ BRATRANEC JEDNOFAKTOROVÉ ANALÝZY ROZPTYLU

Pokud jsou předpoklady pro použití ANOVY výrazně porušeny[7], 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	

[7] Tedy pokud jsou výrazně odlišné rozptyly ve skupinách a pokud v případě malých výběrů mají data výrazná odlehlá pozorování.

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

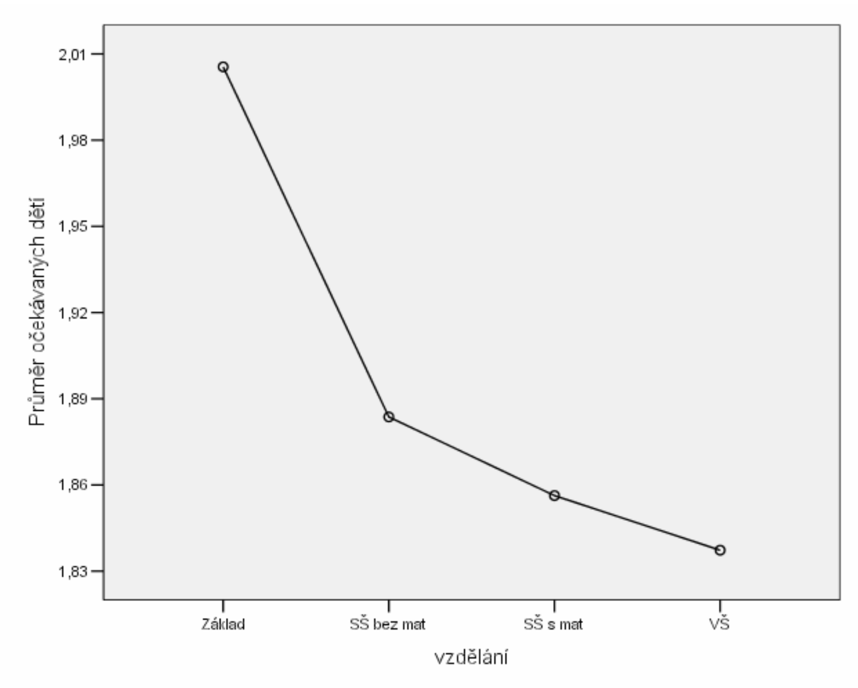
I v tomto testu vycházejí rozdíly mezi jednotlivými skupinami jako statisticky vysoce významné (Signifikance, že můžeme podržet nulovou hypotézu je 0,000). Nulovou hypotézu zamítáme a očekáváme rozdíly v průměrech i v základním souboru. Kruskal-Wallisův test bohužel neumí testovat signifikanci rozdílů mezi jednotlivými skupinami nezávisle proměnné.[8]

Ještě jedno upozornění – jak nelhat s grafy:

Jedním z výstupů procedury ANOVA jsou jednak deskriptivní statistiky a také graf průměrů (*means plot*). Pozor ale na něj, SPSS jej často ‚namaluje'tak, že opticky zvětšuje nalezené rozdíly mezi skupinami.

Příklad:
V našem reprezentativním výzkumu na téma manželství–práce–rodina jsme také sledovali, jak se budou odlišovat jednotlivé,vzdělanostní kategorie žen z hlediska průměrného počtu dětí, které během svého života porodí. V rámci testování rozdílů jsme si nechali také v rámci ANOVA udělat příslušný graf. Získali jsme tento výsledek:

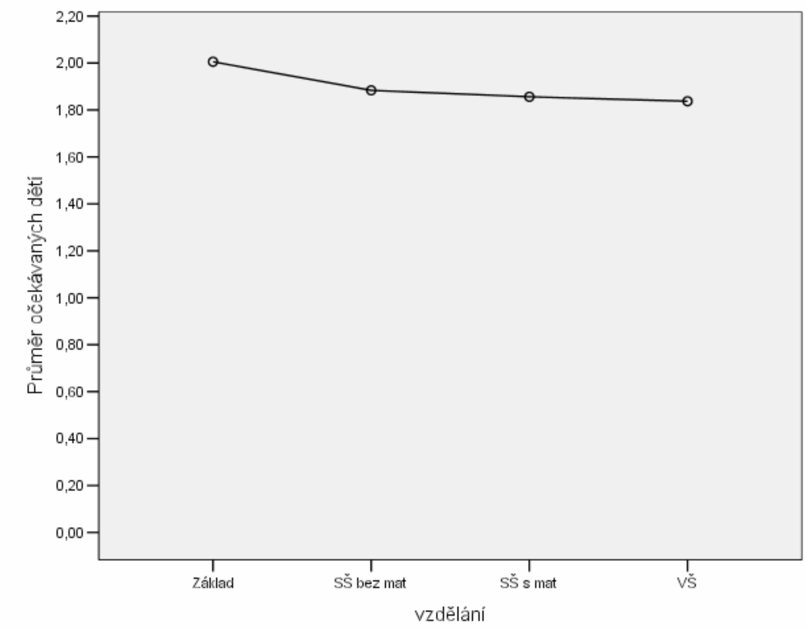
Obr. 6.4a: Očekávaný počet dětí podle vzdělání žen



[8] Jediná možnost je toto udělat sami pomocí párových srovnání Mann-Whitneyho testy (viz výše). Pro zvídavé čtenáře poznamenejme, že u těchto párových srovnání musíme snížit hladinu významnosti (klasickou 0,05) tak, že ji podělíme počtem párových srovnání (totéž dělá i Post Hoc – Bonferroni).

Graf naznačuje existenci značných rozdílů v závislosti na vzdělání. Když si ale všimnete měřítka na svislé ose Y, zjistíte, že se pohybuje v rozmezí 1,83–2,01 dětí. A to je rozmezí příliš malé. Upravme tedy stupnici, aby byla více realistická, to je aby se pohybovala od 0 do 2,2 dětí. Graf se výrazně promění (viz obr. 6.4b) a získá kontury, které mnohem lépe odrážejí realitu. A kdybychom stupnici Y prodloužili až do hodnoty 3, rozdíly by se opticky ještě více snížily.

Obr. 6.4b: Očekávaný počet dětí podle vzdělání žen upravený podle realistické stupnice Y



Jaké z toho plyne poučení? Na rozdíly, které v grafické podobě vypadají větší, než ve skutečnosti jsou, si dávejte velký pozor. Mohli byste být obviněni ze snahy lhát prostřednictvím statistiky. Zvláště grafy, které produkuje SPSS, mají tendenci obsahovat krátké rozsahy hodnot na osách, což vede k optické deformaci výsledků – to je také jeden z důvodů, proč nepovažujeme grafy z SPSS za dobrý výstup.

Literatura:

Hendl, J. 2004. *Přehled statistických metod zpracování dat*. Portál, Praha.