

LEKCE 6

SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT A ZÁKLADY TESTOVÁNÍ HYPOTÉZ

TESTOVÁNÍ HYPOTÉZ

V minulé lekci jsme si řekli, že statistická generalizace se děje prostřednictvím stanovení intervalu spolehlivosti a prostřednictvím testů statistické signifikance, soustředili jsme se však pouze na intervaly spolehlivosti. Nyní přišel čas vysvětlit si problematiku testování hypotéz a statistické signifikance. (viz deVaus 166 an.)

Při analýze dat pracujeme většinou s výběrovým souborem, který jsme se snažili vybrat podle pravidel pravděpodobnostního (to je náhodného) výběru. Výsledky, které v něm zjistíme, vypovídají pouze o jeho jednotkách. Naším cílem je však vypovídat o celém základním souboru. Jelikož si ale nejsme jisti, zdali náš výsledek není způsoben výběrovou chybou, nasazujeme na naše výsledky tzv. testy signifikance neboli **testy statistické významnosti**. Ty nám řeknou, s jakou pravděpodobností výsledky ve výběrovém souboru vznikly díky výběrové chybě anebo skutečně odrážejí vlastnosti základního souboru, z něhož byl náš výběrový soubor generován.

Testy signifikance patří do sféry statistické inference a jsou součástí úloh, jimž se říká testování hypotéz. Abychom tyto statistické úlohy mohli pochopit, musíme si nejdříve objasnit koncept nulové hypotézy a odlišit jej od konceptu hypotézy výzkumné, substantivní či meritorní (všechna tři adjektiva je možné chápat jako synonyma).¹

Výzkumná hypotéza, zkráceně pouze hypotéza, je ve vědě výrokem o vztahu dvou nebo více proměnných. Tento výrok má tu vlastnost, že ještě nebyl empiricky ověřen (testován), je tedy formulací naší výzkumné domněnky, která čeká na prověrku empirickými daty. Příkladem takové hypotézy nechť je např. výrok: „Blondýnky budou dosahovat horších výsledků v testu inteligence než černo-
vlásky“. Abychom zjistili, jak to je, vybrali bychom reprezentativní soubor blondýnek a černo-
vlásek (nejlépe stejného věku, abychom kontrolovali případné působení věku), podrobili bychom je IQ testu a výsledky srovnali. Pokud by blondýnky dosáhly podobného výsledku jako černo-
vlásky, měli bychom empirickou evidenci pro to, abychom tuto hypotézu zamítli.

Vedle výzkumné hypotézy ještě existuje **hypotéza nulová**. Nulová hypotéza je statistickým pomocníkem k ověření hypotézy výzkumné. Je to poněkud bizarní koncept, neboť má podobu výroku, který se snažíme vyvrátit a který formulován v negaci k hypotéze výzkumné (viz rámeček 6.1). Nulová hypotéza je vždy výrokem o neexistenci v základním souboru: o neexistenci rozdílu mezi skupinami, o neexistenci vztahu mezi proměnnými apod. Je to tedy výrok, v němž předpokládáme, že v populaci bude nula, proto nulová hypotéza.

Nulová hypotéza k naší vědecké hypotéze, že blondýnky budou mít horší výsledky v IQ testu než černo-
vlásky, by zněla: „Blondýnky a černo-
vlásky se v IQ testu nebudou lišit, jejich rozdíl v IQ testu bude nulový“. Nyní si představme, že jsme v našem výzkumu zjistili, že blondýnky měly průměrné skóre v IQ testu o 2,22 bodu nižší než černo-
vlásky. Jenže naše nulová hypotéza předpokládala, že rozdíl bude nulový. Co to znamená z hlediska statistické hypotézy? Může to znamenat dvě věci:

1. Naše výběrová data jsou špatná, takže náš neopravňují k revizi našeho původního předpokladu, že rozdíl v IQ testu nulový.
2. Původní předpoklad je špatný a zjištěný rozdíl skutečně znamená, že v populaci se blondýnky a černo-
vlásky v IQ testu liší.

¹ Doporučuji seznámit se s textem: Řehák, Jan 1981. „Data, statistika a sociologie“. *Sociologický časopis* 17:273–290.

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: $H_a: AnmL - AnmP \neq 0$

Pokud však budeme mít o problému již nějaké znalosti, takže budeme vědět, že operacionalizovaný koncept anomie měří míru frustrace a vyvážení 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)

Pokud platí první vysvětlení, pak to znamená, že výsledek byl způsoben charakterem našeho výběrového souboru, tento soubor je zatížen výběrovou chybou, takže zjištěný rozdíl byl způsoben náhodou, což znamená, že my *nemůžeme nulovou hypotézu zamítnout*. Proto dále předpokládáme, že i v populaci je rozdíl mezi blondýnkami a černovláskami z hlediska IQ nulový. Pozor ale: fakt, že nemůžeme zamítnout nulovou hypotézu ještě neznamená, že nulová hypotéza je dokázána.

Pokud ale platí druhé vysvětlení, pak *musíme nulovou hypotézu zamítnout*. Což znamená, že můžeme očekávat, že rozdíl nalezený v našem výběrovém souboru, bude existovat i v populaci.

Na základě čeho se ale rozhodneme, zda platí první či druhé vysvětlení? Pokud máme skutečně reprezentativní soubor, je možné na základě pravděpodobnosti odhadnout s jakou pravděpodobností dostaneme výsledek, který neodráží vlastnosti v populaci. Můžeme tedy stanovit pravděpodobnost, že zís-

káme vzorek, v němž existuje rozdíl mezi IQ blondýnek a černovlásek, byť v populaci takovýto rozdíl neexistuje.

Ve statistice je již ustálenou tradicí, že onou magickou pravděpodobností pro zamítnutí nulové hypotézy je 5 % – což znamená, že jen malý počet výběrových souborů by přineslo zkreslení výsledků z důvodů výběrové chyby. Jak malý počet? Pokud bychom vybrali 100 výběrových souborů blondýnek a černovlásek z populace, v níž neexistuje rozdíl v jejich IQ, pak v 5 z nich bychom našli rozdíl minimálně 2,22 bodů. Máme-li velký výběrový soubor, je dobré tuto hranici zpřísnit a volit pravděpodobnost 1 %.

Ve statistickém žargonu hovoříme o této hranici jako o **hladině významnosti (α)**. Nazývá se tak pravděpodobnost, že náhodná odchylka (daná výběrovou chybou) překročí určitou danou hodnotu, nazývanou hranice významnosti či *kritická hodnota*. Představuje velikost rizika chyby, jež připustíme.

Zvolíme-li hladinu významnosti 0,05, pak požadujeme 95% jistotu, že nalezneme ve vzorku rozdíl, který neexistuje v populaci. Požadujeme-li ještě vyšší jistotu (99 %), volíme hladinu významnosti na úrovni 0,01. Pokud bychom chtěli mít jistotu úplně nejvyšší (99,9 %), volíme hladinu významnosti 0,001, která říká, že jen v 0,1 % případů bychom našli ve vzorku rozdíl, který neexistuje v populaci. V sociálních vědách ale s takto „přísnou“ hladinou α obvykle nepracujeme.

Výraz „nalézt rozdíl ve vzorku, který neexistuje v populaci“ znamená, že výsledek připisujeme výběrové chybě. Což dále znamená, že čím menší je pravděpodobnost, že výběrová chyba je odpovědná za výsledek zjištěný ve výběrovém souboru, tím více si můžeme být jisti, že vzorek odráží reálné vlastnosti základního souboru a že tedy také přináší skutečné a nezkrácené výsledky. Zjistíme-li tedy, že můžeme s 95% (nebo 99%) pravděpodobností zamítnout nulovou hypotézu, pak máme také 95% (nebo 99%) šanci, že naše výběrová data odrážejí skutečnost základního souboru.

Pokud bychom tedy u našich blondýnek a černovlásek zjistili, že rozdíl 2,22 bodů bychom získali pouze v 5 % výběrů (že tedy hladina významnosti je 0,05), museli bychom konstatovat, že tento rozdíl je *statisticky významný* na hladině významnosti 0,05, takže nulovou hypotézu o neexistenci rozdílu v populaci bychom museli zamítnout.

Pokud bychom ale zjistili, že rozdíl 2,22 bodů bychom získali v 6 % výběrů (že tedy hladina významnosti je 0,06), museli bychom konstatovat, že tento rozdíl *není statisticky významný*, takže nulovou hypotézu o neexistenci rozdílu v populaci bychom nemohli zamítnout.

Pokud bychom ale zjistili, že rozdíl 2,22 bodů bychom získali ve 3 % výběrů (že tedy hladina významnosti je 0,03), museli bychom konstatovat, že tento rozdíl je statisticky významný (neboť pravděpodobnost je nižší než ona magická pětiprocentní hranice), takže nulovou hypotézu o neexistenci rozdílu v populaci bychom museli zamítnout.

To vše, co jsme si v předchozích oddílech popsali, byl příklad testu statistické signifikance. Test signifikance vypočítává SPSS a my pouze kontrolujeme, zdali vypočtená hladina významnosti je nižší než 0,05 (ve velkých souborech raději 0,01). Pokud je nižší než 0,05 (nebo 0,01), nulovou hypotézu zamítáme a máme důvěru, že to, co jsme zjistili ve výběrovém souboru, můžeme očekávat také v souboru základním. Pokud je vypočtená hladina významnosti vyšší než 0,05, nemáme důvod nulovou hypotézu zamítnout a navzdory nalezenému rozdílu musíme očekávat, že v populaci bude rozdíl nulový.

Výpočet statistické signifikace může mít podobu *jednostranného*, nebo *dvoustranného* testu. Jaký je mezi nimi rozdíl? Volba jednoho či druhého testu závisí na povaze naší výzkumné (meritorní) hypotézy, která buď má nebo nemá směr. Hypotéza beze směru by byla hypotéza, že blondýnky a černovlásky se budou lišit ve výsledku IQ testu, aniž budeme specifikovat, která skupina ‘vlásek’ bude mít lepší výsledky. Máte-li takovouto hypotézu beze směru, použijte testu dvoustranného.

Vaše meritorní hypotéza může mít ovšem směr: můžeme předpokládat, že blondýnky budou mít horší výsledky, než černovlásky (pozor: pro směrovou hypotézu byste ale měli mít vždy podstatný vědecký důvod – pouhá existence vtípů o hloupých blondýnkách tímto důvodem samozřejmě není). V tom případě použijte výpočet jednostranného testu.

Testy statistické významnosti jsou poněkud kontroverzním tématem. Jejich problém spočívá v tom, že jsou často – z důvodů neznalosti – nadužívány. Byly vynalezeny proto, abychom mohli s relativní jistotou zobecňovat výsledky ze vzorku na populaci. Avšak pozor, můžeme tak činit pouze za předpokladu, že náš výběrový soubor byl vybrán na základě pravděpodobnostních postupů, že je tedy reprezentativní. Pokud by tomu tak nebylo, je použití testů signifikace nesmyslné!² Stejně tak nesmyslné je použití testů signifikace v případě, když pracujete ne s výběrem, ale s celou populací, když tedy máte k dispozici všechny jednotky, které tvoří váš základní soubor.

Dalším bodem, který je u testů významnosti kritizován je jejich rigidita v tradičně a konvenčně stanoveném limitu 95 %. Mnohé učebnice statistiky dnes již nabádají, abychom chápali statistickou významnost jako kontinuum – a abychom proto uváděli v našich publikacích skutečnou hladinu významnosti. Je pak na čtenáři, aby si udělal sám svou představu o pravděpodobnosti výběrové chyby. Alternativním přístupem je dokonce snaha používat spíše intervalů spolehlivosti než testů signifikace.

Nešťastný je samotný výraz „signifikance“, „významnost“. Jak uvidíme dále, statistická signifikace ještě vůbec nemusí znamenat, že výsledek je i věcně významný nebo dokonce důležitý. Ve velkých souborech i malé rozdíly (nebo nízké koeficienty asociace) vycházejí jako statisticky signifikantní. Je proto důležité v analýze nespolehat se pouze na údaj o signifikaci, ale zabývat se také velikostí měřeného rozdílu (či velikostí asociace).

* * *

SROVNÁVÁNÍ SKUPIN NA ZÁKLADĚ STŘEDNÍCH HODNOT

V předchozích příkladech jsme nejdnou opakovali, že informace, které pocházejí z tzv. univariační analýzy, jsou sice velmi důležité, ale jelikož popisují ve zhuštěné podobě celý soubor, který je vždy nejruznějším způsobem strukturován, nepodávají mnohdy informace adekvátní. Někdy mohou dokonce i poněkud zavádět právě z toho důvodu, že ve svém souhrnném vyjádření tuto strukturu zakrývají.³ Proto prvními skutečnými kroky v analýze dat sociologického výzkumu jsou postupy bivariační, jimiž analyzujeme údaje o dvou vlastnostech (znacích, proměnných).

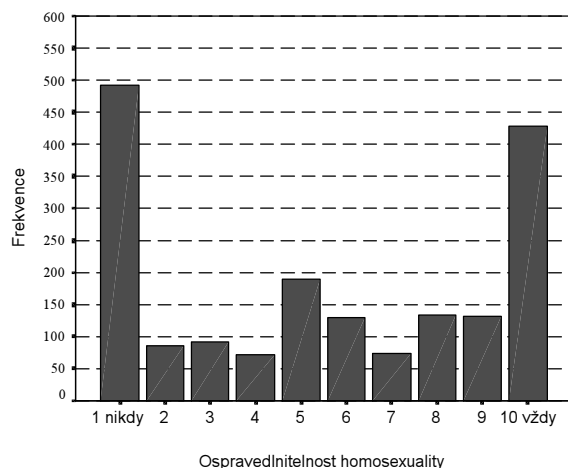
PROCEDURA MEANS

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, viz obr. 6.1) se ani zdaleka neblížil charakteru normálního rozložení, což je v sociologických datech 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)

² Výsledky psychologických výzkumů velmi často pracují s testy významnosti, byť výběrový soubor často nevznikl na základě náhodných výběrových postupů.

³ 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.



Přesvědčme se tedy, zdali tomu tak skutečně je. Můžeme totiž očekávat, že postoje k homosexualitě budou pravděpodobně záviset na věku respondentů. Formulujme proto hypotézu:

„Věk respondenta bude ovlivňovat jeho postoj k homosexualitě a dá se předpokládat, že mladší respondenti budou tolerantnější než respondenti starší.“

Z této vědecké hypotézy lze formulovat hypotézu nulovou:

H_0 : Průměrný věk respondenta se nebude u různých postojů k homosexualitě odlišovat.

Alternativní hypotéze bude znít:

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

Řešení:

V datech EVS-ČR1999 máme samozřejmě údaje o věku, které jsme pro jednoduchost kategorizovali do tří velkých věkových skupin: 18–29, 30–49, 50+. Zjistíme-li, že postoje těchto tří skupin k homosexualitě jsou odlišné, vyvrátíme naši nulovou hypotézu. Odlišnost postojů k homosexualitě zjistíme tak, že vypočítáme průměry pro jednotlivé skupiny nezávisle proměnné (procedura *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

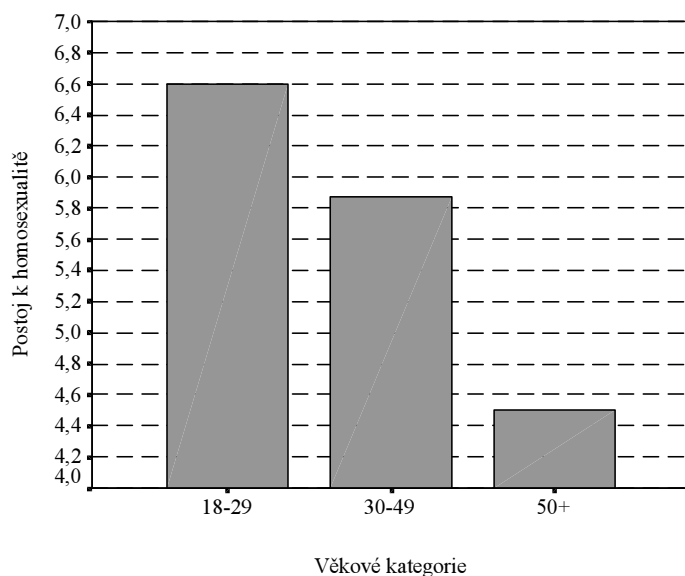
Vidíme, že rozdíly mezi věkovými skupinami skutečně existují, že mají navíc i vzestupný charakter (zde je příklad toho, jak důležité je mít při analýze dat neustále po ruce dotazník – v něm poznáme, jakým způsobem byla měřicí stupnice orientována. V tomto případě to bylo od naprosté netolerance 1 – k naprosté toleranci 10). Čím více byl respondent starší, tím více svůj postoj umísťoval do prostoru stupnice, kteří říkala, že homosexualita není ospravedlnitelná a naopak mladší respondenti byli k homosexualitě tolerantnější.

Jelikož rozdíly jsou docela zřejmé (ale to je v analýze sociologických data kardinální otázka: Kdy je rozdíl v průměrech natolik velký, abychom ho mohli považovat za meritorně význačný – blíže k tomu viz problém meritorní a statistické signifikance, o němž je pojednáno dále). Mnohé zde nasvědčuje tomu, že je možné zamítnout nulovou hypotézu a že je možné přijmout hypotézu alternativní. Důvod,

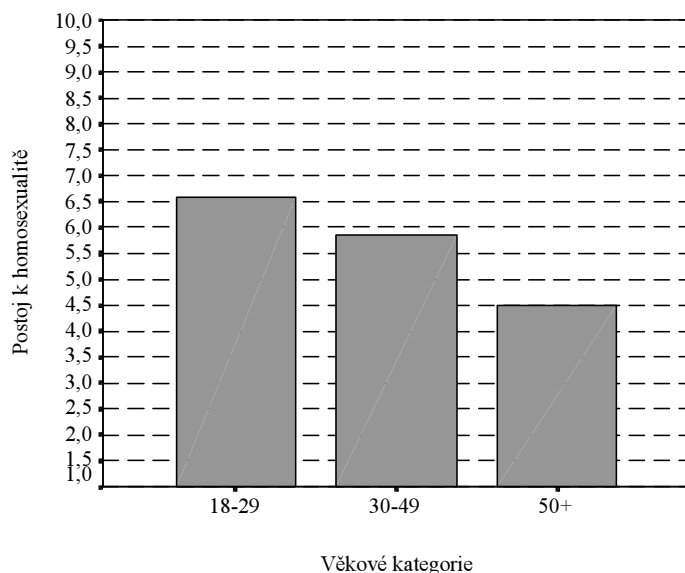
proč tento výrok formulujeme poněkud zdrženlivě, si vysvětlíme za chvíli, až budeme hovořit o t-testu a jednovýchodné analýze rozptylu.

Tento výsledek 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*).

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



Rozdíly vypadají velmi impresivně, ale pozor, je to jistý optický klam, který je způsoben měřítkem osy Y. Pokud na ni vyneseme všech deset bodů měřicí stupnice (a ne jenom body v rozmezí 4,0–7,0), bude vypadat obrázek jinak, rozdíly se značně (opticky samozřejmě) sníží:



Věková struktura 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. Vy-

počet je stejný jako v předchozím případě, pouze u *Independent List* přidáme další *Layer* (vrstvu), do níž vložíme proměnnou pohlaví (q84).

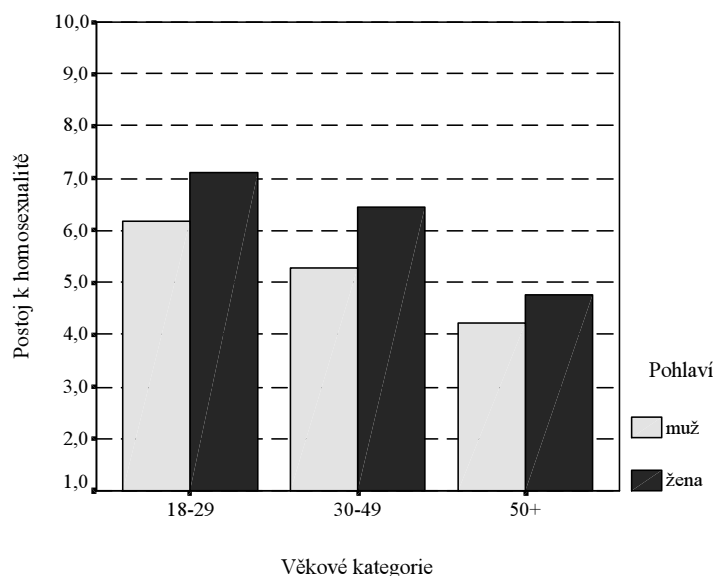
Výsledek:

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 (v řádce Total), že celkově jsou muži tolerantnější k homosexualitě než ženy 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 6.3 (získáte ho stejně jako v předchozím případě, pouze při zadávání grafu budete požadovat ne *Simple*, nýbrž *Clustered* formát).

Obr. 6.3: Postoje k homosexualitě podle věkových skupin a pohlaví (ČR 1999)



T-TEST

Příklad 6.2: 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, jací respondenti 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 výsledek z výběru lze očekávat i v základním souboru, zdali je možné jej generalizovat na populaci. V jazyce testování hypotéz říkáme, že testujeme nulovou hypotézu, že zjištěný rozdíl je způsoben výběrovou chybou a že v populaci není rozdíl v průměrech žádný.

V našem příkladu jsme viděli, že např. rozdíl v toleranci k homosexualitě mezi respondenty ve věku 18–29 let a 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 statisticky, tuto skutečnost ověříme prostřednictvím testu hypotézy o dvou nezávislých průměrech (*two independent-samples t test*), neboli **t-testu pro dva nezávislé průměry**. Nezávislých proto, že se nejedná o párovaná data, že výsledek měření u mladších respondentů nijak neovlivňuje výsledek měření respondentů starších.

Řešení:

V SPSS si zadáme proceduru *Statistics — Compare means — Independent-Samples T Test* a dosadíme do ní příslušné proměnné: naši *Test Variable* je q65a_8 a *Grouping Variable* je vek_kat1. Pak musíme definovat skupiny, které chceme srovnávat: *Define groups – Group 1* je 1 (což je kód pro věkovou skupinu 18–29 let) a *Group 2* je 3 (to je věková skupina 50+).

Výsledek výpočtu uvádějí tabulky 6.1 a 6.2 níže (pozor, tabulku jsem pro lepší čtení editoval a přehodil jsem v *Pivot – Transpose Rows and Columns* sloupce a řádky).

Tab 6.1: Deskriptivní statistiky z t-testu

		Group Statistics			
	VEK_KAT3 tri vekove	N	Mean	Std. Deviation	Std. Error Mean
Q65A_8	1 18-29	409	6,59	3,467	,171
Homosexualita	3 50+	755	4,51	3,478	,127

Tab. 6.2: Výpočet t-testu v SPSS

Independent Samples Test

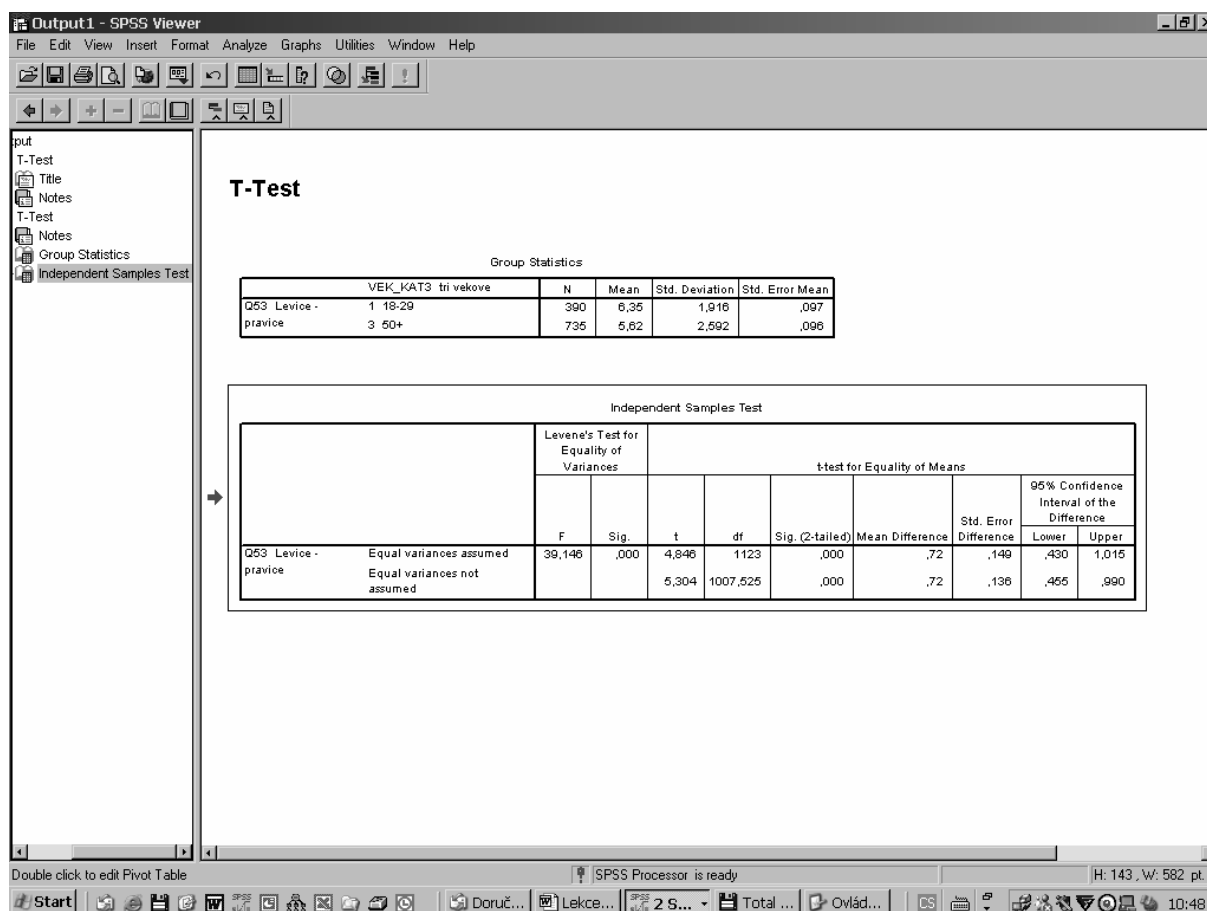
		Q65A_8 Homosexualita	
		Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F	,067	
	Sig.	,796	
t-test for Equality of Means	t	9,832	9,842
	df	1163	841,659
	Sig. (2-tailed)	,000	,000
	Mean Difference	2,10	2,10
	Std. Error Difference	,21	,21
95% Confidence Interval of the Difference	Lower	1,68	1,68
	Upper	2,51	2,51

Jak tuto tabulku číst, abychom zjistili, jak to vlastně je? Nejdříve se rozhodneme, zdali budeme hledat ve sloupci *Equal variances assumed* (předpoklad stejných rozptylů), nebo zda ve sloupci *Equal variances not assumed* (předpoklad rozdílných rozptylů). 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, takže držíme předpoklad, že rozptyly jsou shodné). Pokud by F sig. vyšlo nižší než 0,05, pokračovali bychom s údaji z 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 hladiny významnosti 0,000 je velmi nízká, což nám velí zamítnout nulovou hypotézu. Můžeme proto vyslovit závěr, že nalezený rozdíl v toleranci k homosexualitě (činil 2,1 bodu) mezi mladšími a staršími respondenty nevznikl výběrovou chybou při volbě výběrového souboru, nýbrž je produktem nějakého systematického působení. Velikost rozdílu, který je 2,10 se – jak napovídají poslední dva řádky tabulky uvádějící údaj o 95% intervalu spolehlivosti – bude v populaci s 95% jistotou pohybovat mezi 1,68–2,51.

Pokud se vám zdá složité použití F statistiky pro testování pro shodnost rozptylů, máme ku pomoci opět skript Jana Řeháka (viz soubor T-test.sbs), který udělá to, že tabulku t-testu přeloží do češtiny, vybere na základě Levenova testu tu správnou statistiku t a provede test signifikance. Navíc přidá znaménkové schéma.

Používá se tak, že tabulku z t-testu (avšak netransponovanou) označíte jedním kliknutím myši (viz obr níže) a pod lištou *Utilities* zvolíte příkaz *Run script*. V adresáři, kde máte skripty uloženy pak kliknutím myši vyberete soubor T-test.sbs a dále kliknutím na tlačítko Run necháte prográmek proběhnout.⁴ Pro ilustraci jsem zvolili test rozdílu v průměrech proměnné *levice x pravice (q53)* pro základní a VŠ vzdělání respondenta.

⁴ Pokud se tak nestane, musíte přes ovládací panely Windows a jeho *Místní a jazykové nastavení* změnit nastavený desetinný oddělovač z čárky (,) na tečku (.).



Výsledkem je počestěná tabulka, kterou je ale třeba transponovat (přehodit řádky a sloupce), aby se vůbec vešla na obrazovku. Nejdůležitějším údajem je v ní znaménkové schéma a signifikance.

Tab. 6.3: Výstup ze skriptu t-test

t-test shody středních hodnot pro nezávislé výběry

Statistiky	Proměnné
	Q53 Levice - pravice
Znaménkové schéma	+++
Sig. (Oboustranná)	1.39359140471733E-07
Rozdíl průměrů	0.722410160696344
Standardní chyba rozdílu průměrů	0.136204412341099
Dolní mez ^a	0.45513333850482
Horní mez ^a	0.989686982887869

a. 95% Interval spolehlivosti pro rozdíl středních hodnot

Znaménková schéma v této tabulce říká, že rozdíl v průměrech je signifikantní přinejmenším na hladině významnosti 0,001. Jak to? Znaménkové schéma používá symbolů plus (+) nebo mínus (-) a důležitý je jejich počet: jedno plus (+) nebo jedno mínus (-) znamená, že rozdíl je signifikantní na hladině alfa = 0,05. Dvě znaménka (++) (- -) indikují signifikanci 0,01 a tři znaménka (+++) (- - -) pak signifikanci 0,001. To, zdali se objeví kladné nebo záporné znaménko, závisí na rozdílu v průměrech mezi skupinami. Pokud je rozdíl kladný (což je náš případ, neboť rozdíl byl 0,72), tiskne skript znaménka

Hodnota 0,08 naznačuje středně silný efekt. Nicméně po vynásobení této hodnoty 100 získáme procentuální působení nezávisle proměnné na proměnnou závislou. Pouze 8 % rozptylu v toleranci homosexuality je způsobeno věkem. Zbýlých 92 % je nutno připsat působení dalších faktorů.

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 děláme, jsme si ukázali v lekci 4 – prostřednictvím Q-Q testu a Kolmogorov-Smirnovova testu. Z lekce 4. 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 *Man-Whitney test*.

Procedura:

Analyze — Nonparametric Tests — 2 Independent Samples

Výpočet

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	1 18-29	411	731,69	300725,50
Homosexualita	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 grafický test (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 případech, kdy si skutečně nejsme jisti, zdali použít parametrický nebo neparametrický test, zlaté pravidlo zní: **použijte obou!** Pokud v obou druhích testu dojdete k stejnému závěru, je vše v naprostém pořádku. Pokud výsledky v neparametrickém testu nejsou signifikantní a v 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, tak ovlivňují průměr, což může být příčina oné nejednoznačnosti.

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 mocnější.**

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. V našem případě takovýto předpoklad máme: byli jsme přesvědčeni, 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 nelze. Pokud by ale dvoustranný test signifikance měl např. hodnotu 0,095, což je výsledek, který nám velí podržet nulovou hypotézu, ale my měli důvod pro jednostranný test, jeho dvěma dělená 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.

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 (inferenční) statistiky. Při rozhodování, zdali podržet, nebo zamítnout nulovou hypotézu, platí základní pravidlo:

Zlaté pravidlo pro induktivní statistiku:

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:

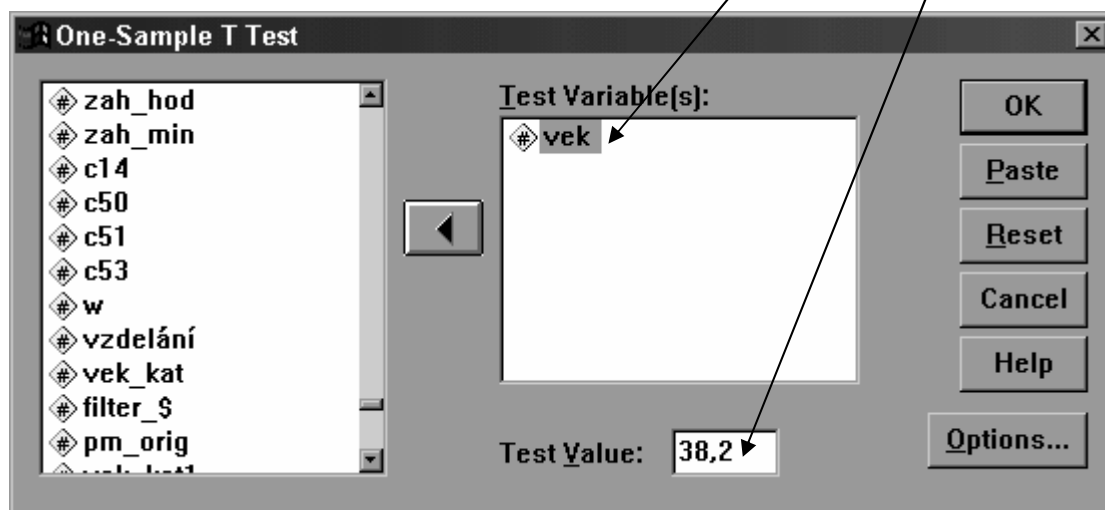
V úlohách o průměru je možno ale také testovat ještě jiný případ. Jedná se o proceduru *One-Sample T Test*. To je úloha, při níž známe průměr nějakého znaku v základním souboru (tzv. populačního parametru) a ptáme se, zdali náš výběr pochází z této populace. Srovnáváme tedy hodnotu výběrové statistiky nějaké proměnné se známou hodnotou této proměnné v populaci.

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) je mimo interval spolehlivosti je signálem, že obě hodnoty se budou statisticky významně lišit. Zjistíme to ale přesně. Úloha tedy zní: Pochází z hlediska věku náš výběrový soubor z této základní populace? Nulová hypotéza zde zní: hodnota populačního parametru a hodnota výběrové statistiky se nebudou lišit.

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.



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* je velmi nízká (0,000), což znamená, že musíme zamítnout nulovou hypotézu, že obě charakteristiky se neliší. Věkový průměr našeho výběrového souboru se statisticky významně odlišuje od věkového průměru populace. Náš výběrový soubor tedy nepochází z populace, jejíž průměr byl 38,2 roku. Je to tedy výsledek, který již naznačoval interval spolehlivosti.

Co tento výsledek ale věcně znamená? Máme snad špatně vybraný soubor? Ale vždyť náš soubor je reprezentativní pro celou populaci ČR! Nebo snad reprezentativní není? Nebojte, je reprezentativní. Nesoulad zde vznikl tím, že náš soubor EVS-ČR1999 není z hlediska věku reprezentativní pro celou populaci ČR, nýbrž pouze pro její *dospělou* populaci, to je populaci starší 18 let. A jelikož průměrný věk celé 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 adekvátními populačním průměrem si provedeme 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 deklarace o politické orientaci (levice versus pravice) podle věkových skupin nebo podle vzdělání apod. Proto si nyní ukážeme, jak testovat nulovou hypotézu, že průměry 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 rozptyl: 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 rozptylu, 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 *ANOVY*, 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. dvoufaktorovou 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?

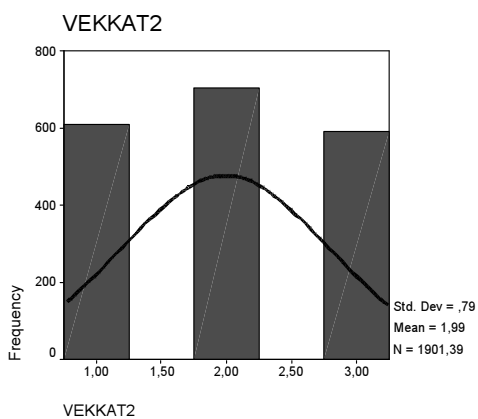
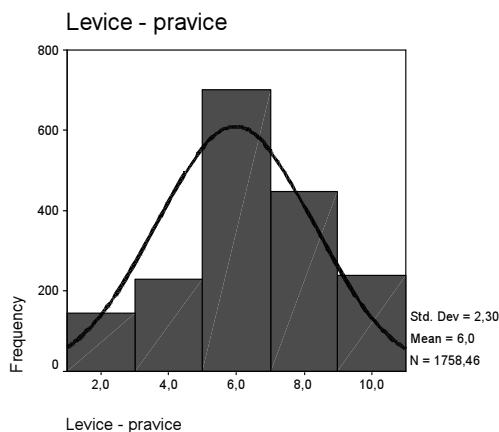
1. Jednotlivá pozorování musí být na sobě nezávislá. Tento předpoklad je v sociologických šetřeních vždy bez problémů, výzkumné designy s opakovaným měřením stejných subjektů (respondentů) nejsou příliš časté.
2. Rozložení v populaci je normální. Tento předpoklad se v praxi často zanedbává, v úvahu musí 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 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, již je v tomto případě neparametrický test Kruskal-Wallisův.

Příklad 7.1: 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*)?

Řešení

Nejdříve se podívejme, jak tvar má rozložení obou proměnných, zdali není porušen předpoklad normality rozložení.

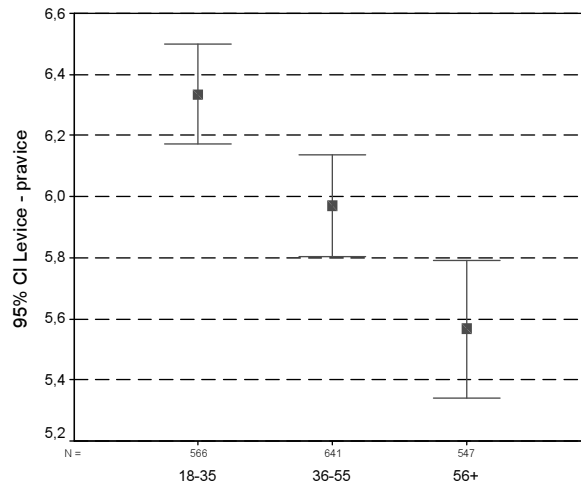


Vidíme, že obě proměnné jsou přibližně normálně rozloženy.

Nyní se podívejme, jak se průměry v těchto skupinách odlišují graficky. Použijme k tomu graf intervallů 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

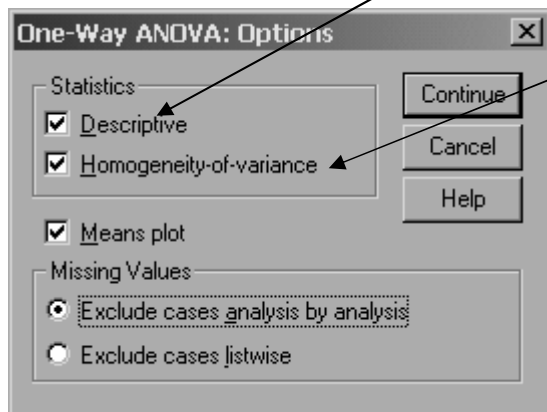
Obr. 7.3



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), a že je tudíž porušen jeden z předpokladů pro analýzu rozptylu.⁵ Tato situace by naznačovala, že v analýze není možné pokračovat. Naštěstí tento předpoklad je poměrně robustní a 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. Porušení tohoto předpokladu tedy můžeme ignorovat.

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 nulové, že tedy mezi skupinami nebude žádný rozdíl. Alternativní hypotézou je předpoklad, že průměry se liší.

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 rozložením⁶ (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 nesterčné, budou se lišit.

Pokud by F test nevyšel statisticky významný a my museli podržet nulovou hypotézu o neexistenci rozdílu v populačních průměrech, analýza by zde skončila.

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 skupin

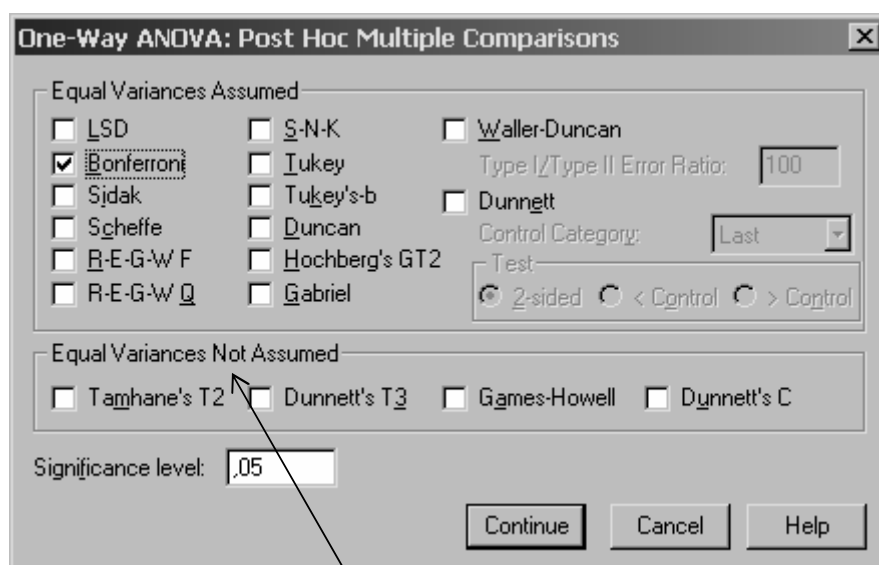
⁵ Zopakujme si, že pokud by tento předpoklad neměl být porušen, musela by vyjít signifikance Levenova testu vyšší než 0,05.

⁶ 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.

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*



Vidíte, že nabídka testů je vskutku bohatá. Volba toho patřičného závisí na mnoha okolnostech, zmíním pouze ty zásadní. Pokud je počet srovnávaných průměrů malý, je dobré použít Bonferroniho testu. Pokud je ale počet srovnávání velký, je lepší použít testu Tukeyho. Výborný je také test R-E-G-W Q (což je akronym ze jmen Ryan, Einot, Gabriel a Welsch), je dokonce považovaný za nejlepší, avšak pokud máte nesterjné početné srovnávané skupiny (což je náš případ), používat by se neměl. V případě, že předpokládáte, že populační rozptyly se budou odlišovat, nabízí SPSS čtyři testy (na řádce pod označením *Equal Variances Not Assumed*). Z nich je nejlepší Games-Howellův test. Celkově doporučuji postup v souladu s Fieldingem (2000:274-276)) k jakémukoliv zvolenému testu ještě přidat test Games-Howellův.

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*), 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: 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.

Zjistíme 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:

$$\eta^2 = \frac{\text{Sum of squares between groups}}{\text{Total sum of squares}}$$

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 nízký efekt.

V tomto příkladu jsme tedy zjistili statisticky signifikantní výsledky, ale věcně nejsou rozdíly příliš velké, což je dokumentováno také výpočtem velikosti účinku ($\eta^2 = 0,02$). Statistické signifikance jsme dosáhli proto, že náš výběrový soubor je poměrně velký. V tomto kontextu tak musím opět vyjádřit určité varování: nespolehejte příliš na statistickou signifikanci ale pracujte i s jinými ukazateli (s intervaly spolehlivosti nebo s velikostí účinku).

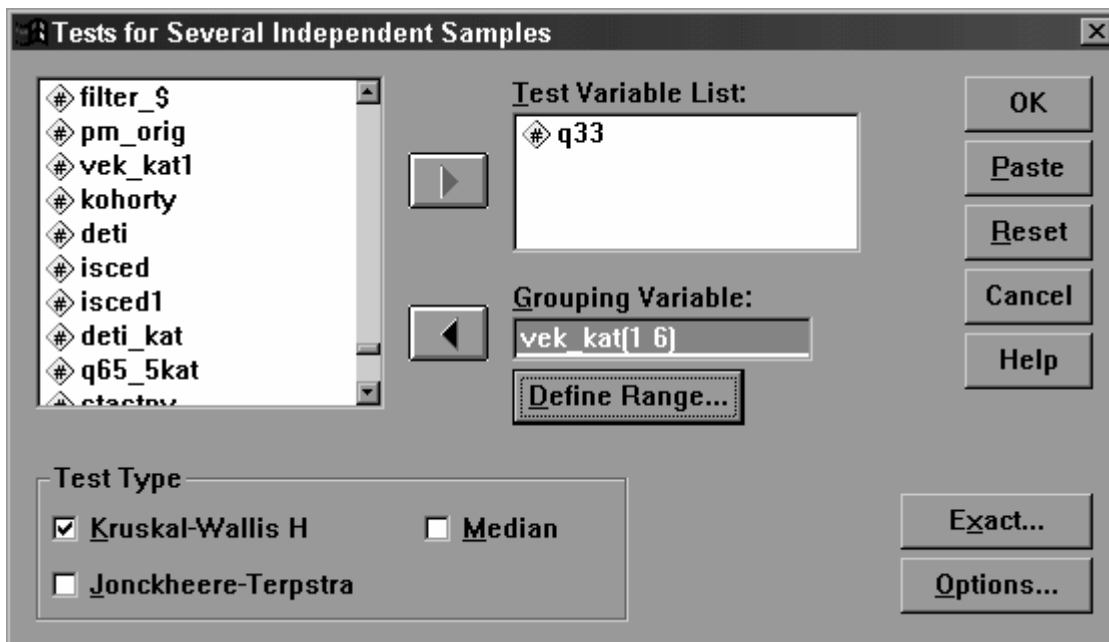
Zobecnění postupu jednotlivých kroků v ANOVĚ:

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.

Literatura:

Pallant Julie 2001. *SPSS Survival Manual*. Open University Press, Buckingham.