

Ekonomická univerzita v Bratislave

Fakulta podnikového manažmentu

Róbert Hanák

DÁTOVÁ ANALÝZA PRE SOCIÁLNE VEDY

Vydavateľstvo
Ekonom
2016

Autor©

Ing. Mgr. Róbert Hanák, PhD.

Recenzenti

Doc. Ing. Alojz Ritomský, CSc.

Doc. Ing. Anita Romanová, PhD.

ISBN 978-80-225-4345-3

Za odbornú stránku textu, jazykovú korektúru a konečnú verziu publikácie zodpovedá autor.

Publikácia neprešla jazykovou korektúrou.

Podakovanie

V prvom rade by som rád poďakoval vynikajúcemu metodológovi doc. Ing. Alojzovi Ritomskému, CSc. ako aj skvelému vedcovi Mgr. Miroslavovi Sirotovi PhD., ktorý bol vedúcim mojej diplomovej práce, za to, že upevnili moje rozhodnutie viac sa venovať vede a výskumu. Doc. Ing. Anite Romanovej, ktorá je inšpiráciou toho, ako má vyzerat' schopnosť bezchybne, rýchlo a efektívne ovládať osobný počítač a manažovať svoj čas. Mojím dlhoročným vedúcim doc. Ing. Mojmírovi Koklesovi a prof. PhDr. Viere Bačovej DrSc. za ich láskavú zhovievavosť a múdry pohľad na svet okolo nás. Týmto ďakujem aj mojim rodičom, ktorí svojou celoživotnou prácou a výchovou vytvorili základ, na ktorom sme s bratom mohli bezpečne stavať.

Venované Alexandre, Kristíne a Richardovi

Obsah

1	ÚVOD.....	6
2	SOFTVÉR PSPP.....	7
2.1	Získanie a inštalovanie softvéru PSPP	7
3	ZOZNÁMENIE SA ZO SOFTVÉROM PSPP	11
4	PONUKA SÚBOR (FILE).....	17
4.1	Ponuka Úpravy (Edit).....	24
4.2	Ponuka Pohľad (View)	26
4.3	Ponuka Dáta (Data).....	29
4.4	Ponuka Transformácia (Transform)	36
4.5	Ponuka Analýza (Analyze).....	43
4.6	Ponuka Pomocné funkcie (Utilities).....	54
4.7	Ponuka Okná (Windows).....	56
4.8	Ponuka Pomoc (Help).....	57
5	VEDECKÝ VÝSKUM.....	61
5.1	Populácia.....	61
5.2	Správnosť merania.....	62
5.3	Premenné a ich typy.....	64
5.4	Výskum a jeho typy	68
5.5	Testovanie hypotéz	69
6	ZÁKLADNÁ ANALÝZA DÁT	74
6.1	Frekvenčná analýza	74
6.2	Opisné charakteristiky	75
6.3	Ukazovatele centrálnej tendencie	76
6.4	Kvantily	77
6.5	Ukazovatele variability	78
7	ANALÝZA RELIABILITY	83
8	SKÚMANIE VZÁJOMNÝCH VZŤAHOV MEDZI PREMENNÝMI	87
8.1	Skúmanie vzájomných vzťahov u nominálnych premenných	87
8.2	Skúmanie vzájomných vzťahov u ordinálnych premenných	91
8.3	Skúmanie vzájomných vzťahov u intervalových premenných	95
8.3.1	Korelácia.....	95

8.3.2	Výber konkrétnej štatistickej metódy.....	102
8.4	Lineárna regresná analýza	103
8.5	Binárna logistická regresia	109
9	FAKTOROVÁ ANALÝZA	112
9.1	Exploračná faktorová analýza.....	113
10	POROVNÁVANIE STREDOVÝCH HODNÔT.....	123
10.1	Parametrické testy a podmienky ich použitia.....	123
10.1.1	Veľkosť účinku (effect size)	125
10.1.2	Dvojvýberový t – test (Independent sample t -test)	127
10.1.3	Analýza rozptylu -variability (ANOVA: Analysis of variance). 131	
10.1.4	Párový t test.....	135
10.1.5	Jednovýberový t – test (One sample T - Test)	137
10.2	Neparametrické testy	140
10.2.1	Wilcoxonov test.....	140
11	ZÁVER	145
12	POUŽITÁ LITERATÚRA	146
13	Register	149

1 ÚVOD

Na Slovensku ako aj v Českej republike existuje množstvo kvalitných publikácií o štatistike na vysokej odbornej úrovni. Nakoľko štatistika je súčasťou matematiky tieto knihy obsahujú desiatky až stovky, neraz veľmi komplikovaných, vzorcov, matematických znakov, rovníc a odborných výrazov niekedy i slangu. Tým sa niektoré z nich stávajú pre čitateľa bez matematického vzdelania ťažko čitateľnými.

Študenti a absolventi sociálnych vied ako psychológia, sociológia, sociálna práca, pedagogika, ekonómia, masmediálna komunikácia zvyčajne matematické vzdelanie nemajú vôbec (niektorí ani maturitu z matematiky) alebo len vo veľmi obmedzenej miere. Prítom však títo študenti a absolventi aktívne pracujú s dátami a robia výskum nielen pre potreby svojich záverečných prác, ale aj pre výkon vlastných povolání. Táto kniha je určená práve im.

Pri písaní tejto knihy som sa snažil držať pravidla, že na navrhnutie vlastného výskumu a jeho následnú analýzu dnes nepotrebujeme vedieť ručne zapisovať, odvodzovať a chápať rozsiahle matematicko-štatistické vzorce, ale ak pochopíme základné princípy konkrétneho problému dokážeme túto úlohu zvládnuť aj bez podrobných deklaratívnych znalostí matematiky, resp. štatistiky vo forme vzorcov a práce s nimi.

V súčasnosti viac než dve desaťročia, prebieha užívateľská štatistická analýza na osobných počítačoch, bez akéhokoľvek manuálneho počítania. Tento trend je nezvratný a bude sa len posilňovať. Práca s dátami sa tým úžasne uľahčila, ale problém pri analýze dát sa presunul z manuálneho počítania na dobré porozumenie problému, výber správnych štatistických metód, nastavenie softvéru na analýzu a následnú interpretáciu dát.

Táto učebnica je zameraná na analýzu pomocou štatistického softvéru PSPP. Tento softvér je bezplatný opensource softvér, ktorý je dostupný zdarma. Je pravidelne updatovaný a zlepšovaný a jeho funkcionality rastie. Pre bežného používateľa ako aj pre vysokoškolského študenta, či doktoranda v sociálnych vedách je dostačujúci.

Táto publikácia bola vydaná na základe finančnej podpory z: výskumného projektu KEGA č. 029EU-4/2015 s názvom „Vytvorenie elektronického portálu a voľne dostupného systému výučby analyticko - štatistických metód v sociálnych vedách pomocou opensource softvéru PSPP“ v rozsahu 100%.

2 SOFTVÉR PSPP

Softvér PSPP vznikol ako opensource freeware softvér, teda softvér, ktorý je voľne dostupný, je zdarma, nie je nikým vlastnený a súčasne je možnosť vstupovať do jeho zdrojového kódu a ten ľubovoľne upravovať. Vznik softvéru bol podmienený vysokými cenami komerčných softvérov zameranými na analýzu dát, ktoré sa v súčasnosti pohybujú od tisíc eur vyššie za jednu licenciu.

Softvér PSPP je z grafického hľadiska ako aj z hľadiska umiestnení jednotlivých funkcií veľmi podobný, až identický spoplatnenému softvéru SPSS od spoločnosti IBM, aj keď ich zdrojové kódy nemajú spoločné úplne nič a sú postavené na iných základoch. Preto tí, ktorí zvládnu používanie softvéru PSPP bez akýchkoľvek problémov či školení môžu ihneď začať pracovať s programom SPSS. Taktiež tým užívateľom, ktorí majú skúsenosti s programom MS Excel, bude užívateľské prostredie čiastočne známe a dokážu sa v ňom primerane orientovať.

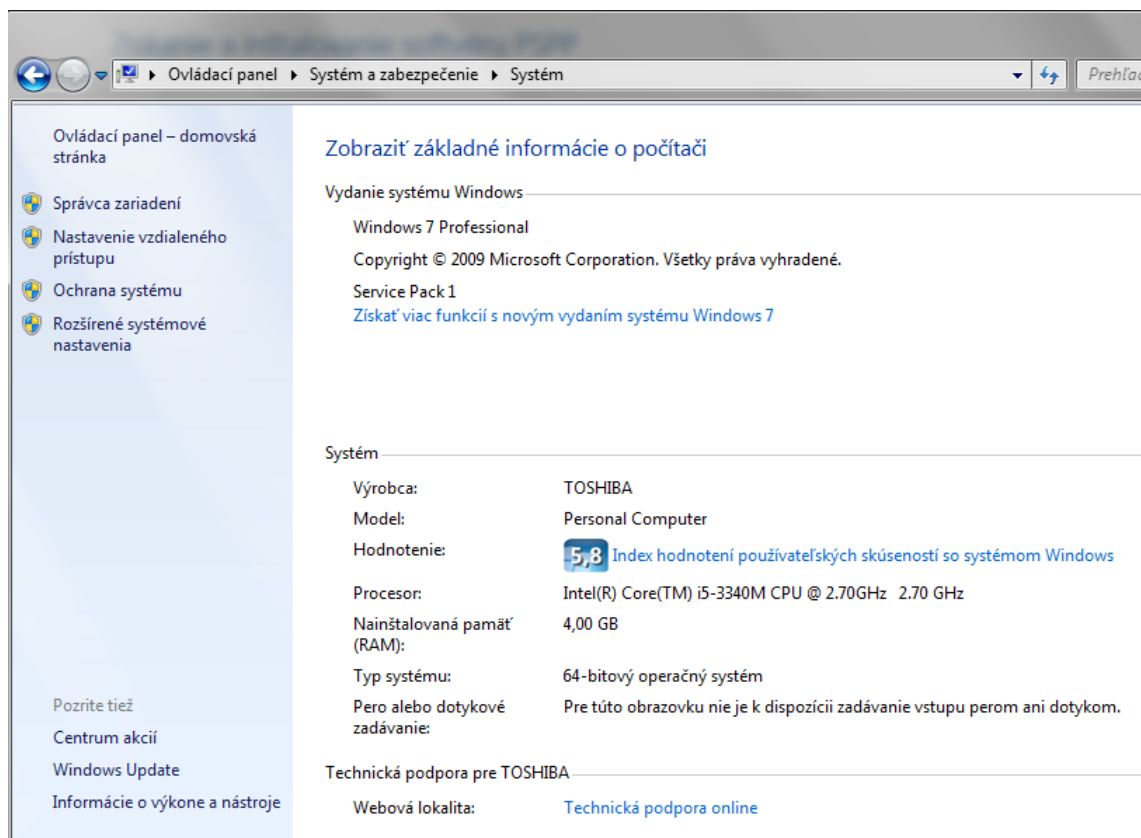
Program PSPP sa neustále vyvíja a rozširuje. V marci roku 2017 je dostupná verzia 10.2. Užívateľ nemusí mať obavy z častých aktualizácií, ktoré prebiehajú zvyčajne niekoľkokrát ročne. Zmeny sú prevažne malé a drobného charakteru, a rozdiel medzi konkrétnymi verziami si bežne nevšimne ani skúsený užívateľ.

2.1 Získanie a inštalovanie softvéru PSPP

Stiahnutie programu je možné bezplatne zo stránky : <http://pspp.awardspace.com/>, prípadne stačí dať do webového vyhľadávacieho hľadane slovné spojenia: pspp software download for windows. Na tejto stránke vidíme všetky verzie softvéru pspp aj s dátumami kedy boli zverejnené. Softvér, jeho opis, stiahnutie a prácu s ním podrobne opisuje aj naša webstránka www.statistikapspp.sk

Postup inštalácie je nasledovný:

1. Vyberieme si poslednú verziu softvéru. Pri inštalácii musíme poznať aký typ operačného systému máme na svojom počítači a podľa toho si aj vyberáme verziu softvéru PSPP. Overenie vlastného operačného systému urobíme nasledovne. V ponuke Štart si otvoríme Ovládací panel, potom si vyberieme možnosť Systém a zabezpečenie a znova si vyberieme Systém. Zobrazí sa okno, ktoré opisuje operačný systém a jeho verzie (napr. obrázok č.). V našom príklade sme identifikovali operačný systém Windows 7 profesional a skutočnosť, že je to 64 bitový operačný systém.



Obr. 2.1.1 Identifikovanie verzie operačného systému na vlastnom počítači.

2. Ak napíšeme do webového vyhľadávača spojenie slov: „PSPP software download for Windows“, potom môžeme otvoriť prvú vyhladanú webovú stránku. Zobrazí nám okrem iného dve tabuľky. V prvej tabuľke nás informuje o aktuálnej verzii softvéru, veľkosti inštalačného balíka a verzii operačného systému. V spodnej tabuľke si vyberieme poslednú verziu softvéru PSPP a podľa verzie operačného systému (32 vs 64 bitový). Následne klikneme na inštalačný balík.

PSPP is Free Software. If you want to study Free Software visit [The Free Technology Academy \(FTA\)](#).

Highlights of the current PSPP-for-MSWindows setup

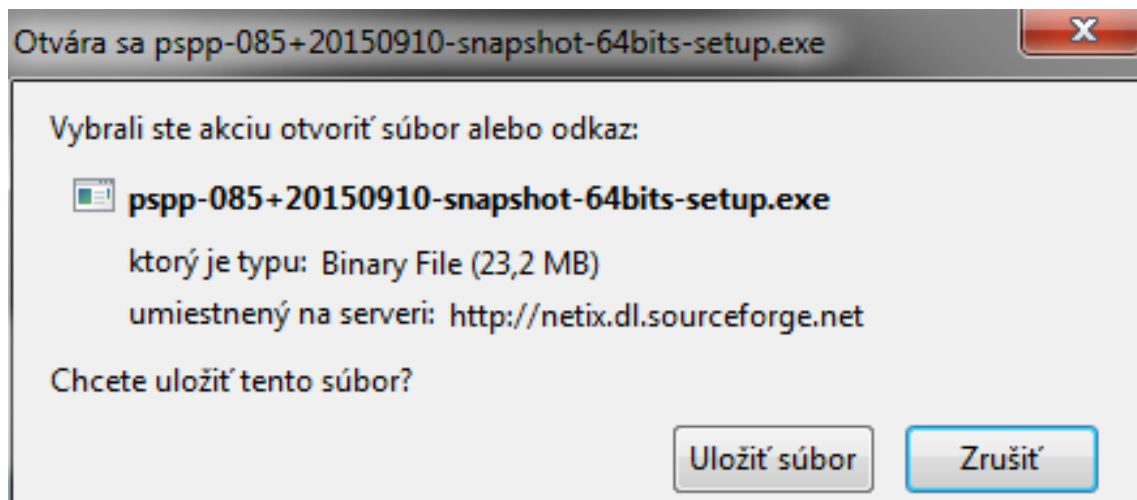
PSPP info:		Package info:	
Current version:	Master version = 0.8.5 interim version with updates, but not fully tested	MSWindows version:	MSWindows XP and newer
Questions/Suggestions:	pspp-users@gnu.org	Package Size:	32 Mb
Information about PSPP:	http://www.gnu.org/software/pspp	Size on disk:	72 Mb
PSPP Manual:	PDF or HTML (current version will be installed on your PC by the installer package)	Technical:	MinGW based Cross-compiled on openSUSE 13.2

Downloads:

Version	Installer for 32bits version Will work on 32 and 64bits MSWindows	Installer for 64bits version Works only on a 64bits MSWindows version
0.8.5-gdaa1fe	PSPP 0.8.5 2015-09-10 32bits	PSPP 0.8.5 2015-09-10 64bits
0.8.5-g2d71ac	PSPP 0.8.5 2015-06-21 32bits	PSPP 0.8.5 2015-06-21 64bits
0.8.4-g267362	PSPP 0.8.4 2015-04-27 32bits	PSPP 0.8.4 2015-04-27 64bits
0.8.4-g6036c7	PSPP 0.8.4 2015-03-27 32bits	PSPP 0.8.4 2015-03-27 64bits

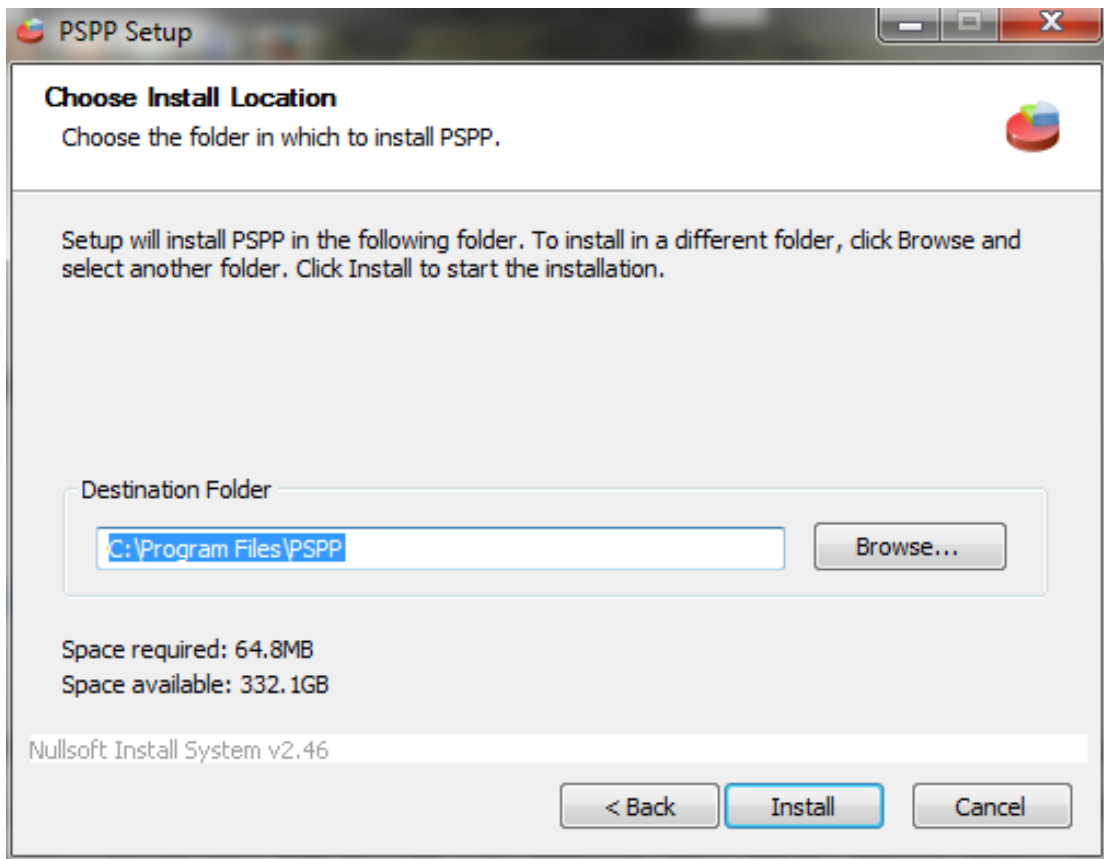
Obr. 2.1.2 Stiahnutie inštalačného balíka z internetu.

3. Po kliknutí na inštalačný balík sa nám tento stiahne a po stiahnutí, ktoré trvá len okolo minúty sa nám otvorí sprievodca inštaláciou.



Obr. 2.1.3 Spustenie inštalačného balíka

4. Pri inštalácii sa nás opýta na umiestnenie softvéru a ponúkne nám už prednastavené uloženie v Zložke súborov (Program files). Keď klikneme na tlačidlo Inštalovať (Install) inštalácia prebehne v priebehu minúty.



Obr. 2.1.4 Umiestnenie inštalácie

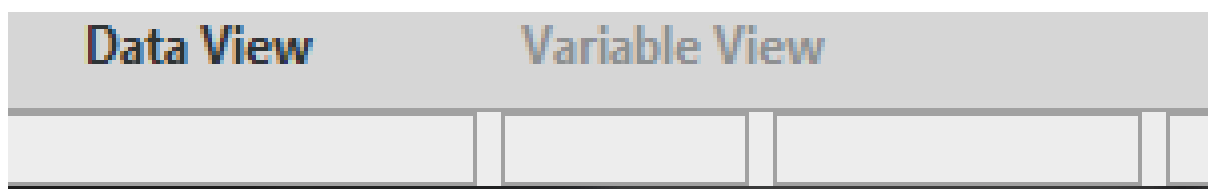
5. Po úspešnej inštalácii sa nám objaví ikona softvéru na ploche ako aj v ponuke panelu Štart, alebo ju nájdeme v nainštalovaných programoch.



Obr. 2.1.5 Ikona softvéru PSPP

3 ZOZNÁMENIE SA ZO SOFTVÉROM PSPP

Potom čo sme si PSPP softvér úspešne nainštalovali, môžeme s ním začať pracovať. Každý súbor v softvéri PSPP je zobrazovaný vždy dvojakým spôsobom. Na spodnej lište vľavo preto vidíme dva háčky a to: *Dátové Zobrazenie (Data View)*, kde sú uvedené konkrétne hodnoty premenných, zjednodušene čísla. Alebo si môžeme prepnúť hárok *Zobrazenie Premenných (Variable View)* s ich charakteristikami (názov, číslo alebo text, zaokrúhlenie, typ premennej a ostatné) pozri obrázok č. To okno, ktoré je otvorené, aktívne je označené tučným písmom. Predtým, než by sme sa pokúsili vložiť akékoľvek dáta do prázdneho súboru, musíme si najprv pripraviť maticu, do ktorej to budeme vkladať. Maticou rozumieme zadefinovanie premenných a ich charakteristík v okne premenných (Variable View). Súbor PSPP po štatistickej analýze dát ešte vytvorí tretí typ okna, v ktorom uvádza výsledky, tzv. *Výstup (Output)*. Samozrejme na začiatok musíme dáta do softvéru buď vložiť alebo skopírovať.

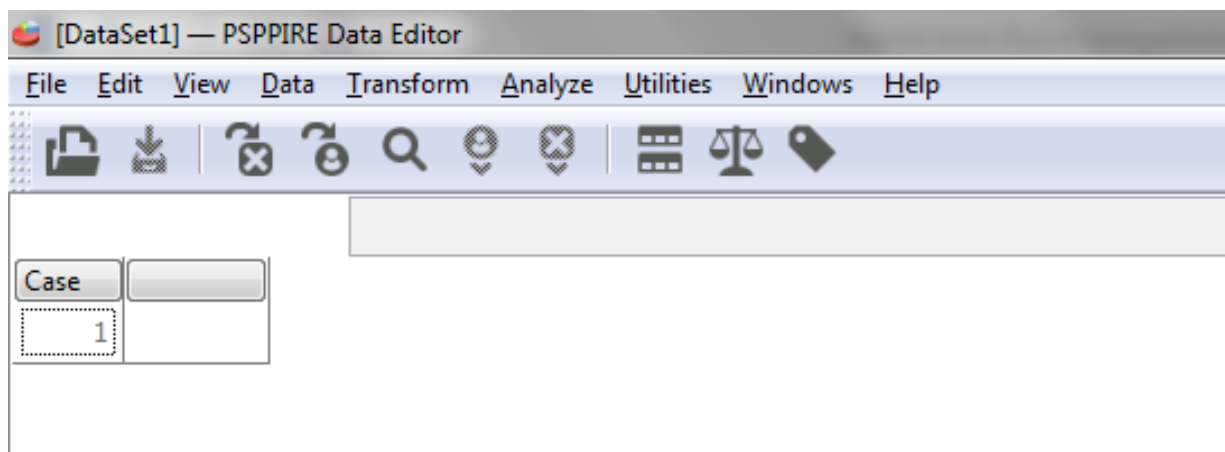


Obr. 2.1.1 Prepínanie sa medzi dátovým rozhraním alebo oknom premenných.

Prázdny súbor

Dátové Zobrazenie (Data view)

Prázdny súbor (bez dát) vyzerá odlišne než súbor v ktorom sú už vložené dáta pripravené na analýzu. Na porovnanie sme pre čitateľa pripravili dva obrázky užívateľského rozhrania pri prázdnom súbore a pri súbore s dátami. Akákoľvek analýza začína najprv prípravou súboru na to, aby sme mohli analýzu vôbec spustiť. Ak máme prázdny súbor, tak sa nám *Dátové Zobrazenie (Data view)* ukazuje nasledovne (Obr. 2.1.2). Je to prakticky prázdna plocha bez riadkov a bez stĺpcov a obsahuje len dve bunky.



Obr. 2.1.3 Prázdny súbor, *Dátové zobrazenie (Data view)*

Pri už naplnenom súbore dátami je vzhľad *Dátového Zobrazenia (Data view)* ako na obrázku Obr. 2.1.2. V prvom riadku sú uvedené názvy premenných (stĺpcov) a v nasledujúcich riadkoch sú uvedené už konkrétne hodnoty každého jednotlivého prípadu, jednotky v súbore (cases).

celkové_skr_II	PID_del	PID_intu	PID_typ	Body_jednod_u	Body_podvojn_	Body_dot	ERP_body
8,00	15,00	18,00	3,00	4,00	6,50	6,00	6,00
6,00	29,00	27,00	3,00	4,50	7,00	8,00	7,00
6,00	30,73	29,18	.	5,00	7,00	8,00	6,00
5,00	34,00	24,00	1,00	5,00	6,50	6,00	8,00
9,50	13,00	26,00	3,00	5,00	6,00	7,00	5,00
5,00	28,00	25,00	3,00	4,50	7,00	8,00	5,00
7,00	27,00	27,00	3,00	5,00	7,00	9,00	7,00

Obr. 2.1.4 Dátové zobrazenie (Data View) už naplneného súboru

Každý prípad, *Jednotka v súbore (Case)*, napríklad osoba, podnik, konkrétny prípad, udalosť a podobne, sa uvádza v jednom riadku. V stĺpcoch sú vždy jeho charakteristiky. Teda ak napríklad skúmame študentov, tak každý študent je v jednom riadku. V stĺpcoch sú jeho premenné ako napr. jeho vek, dosiahnuté body, reakčný čas, skupina a všetky ostatné, ktorými študenta charakterizujeme. Ak máme napr. 200 študentov, tak máme aj 200 riadkov. A ak tých študentov opíšeme 15-timi premennými, teda jeho charakteristikami (napr. vek, body, atď.) tak tabuľka má 15 stĺpcov.

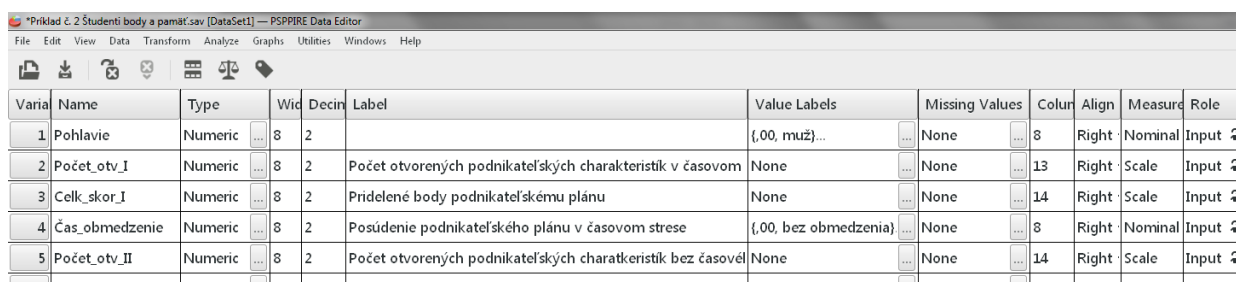
Zobrazenie Premenných (Variable View)

Keď sa preklikneme vľavo dole do *Zobrazenie Premenných (Variable View)*, tak pred vložením premenných nám softvér PSPP tiež zobrazuje prázdnu tabuľku. Preddefinovaná štruktúra je taká, že každá premenná sa nachádza v jednom riadku a je definovaná 11 charakteristikami, Obr. 2.1.4. V *Zobrazení Premenných (Variable View)* sú jednotlivé premenné tiež uvedené v riadkoch, no pri prepnutí sa do *Dátového zobrazenia (Data View)* z premenných zobrazí len *Názov (Name)* napr. Vek a tento bude uvedený ako hlavička stĺpca. Pod ním už sa budú nachádzať konkrétne čísla. Čiže v *Zobrazení Premenných (Variable View)* premennú Vek presnejšie definujeme a opisujeme napríklad na koľko desatinných miest ju budeme merať, ako budeme čísla zarovnávať, aký typ premennej to bude a ostatné.



Variable	Name	Type	Width	Decimals	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1					

Obr. 2.1.5 Uživatelské rozhranie pre prázdny súbor Zobrazenie Premenných (Variable View)



Variable	Name	Type	Width	Decimals	Label	Value Labels	Missing Values	Columns	Align	Measure	Role
1	Pohlavie	Numeric	8	2		{,00, muž}...	None	8	Right	Nominal	Input
2	Počet_otv_I	Numeric	8	2	Počet otvorených podnikateľských charakteristik v časovom	None	None	13	Right	Scale	Input
3	Celk_skor_I	Numeric	8	2	Pridelené body podnikateľskému plánu	None	None	14	Right	Scale	Input
4	Čas_obmedzenie	Numeric	8	2	Posúdenie podnikateľského plánu v časovom strese	{,00, bez obmedzenia}	None	8	Right	Nominal	Input
5	Počet_otv_II	Numeric	8	2	Počet otvorených podnikateľských charakteristik bez časové	None	None	14	Right	Scale	Input

Obr. 2.1.6 Zobrazenie Premenných (Variable View) s už vyplnenými premennými

Na to aby akýkoľvek štatistický softvér dokázal úspešne analyzovať dáta je nevyhnutné ručne nastaviť, aký typ je jednotlivá premenná, aké hodnoty nadobúda, na koľko desatinných miest je nastavená, do akých kategórií sa ešte delí a podobne. Preto je potrebné každú premennú takto vopred nastaviť a upraviť. V nasledujúcich odsekoch si opíšeme čo ktoré charakteristiky znamenajú a ako ich máme správne do softvéru zadávať. Okrem *Mena (Name)* a *Označenia (Label)*, ktoré musíme vždy zadať ručne, pre ostatné charakteristiky premennej nám softvér PSPP ponúka skupinu volieb. Z týchto zvyčajne len vyberáme, prípadne ponechávame už prednastavené možnosti. Preto je nastavovanie premenných rýchlejšie než sa zdá na prvý pohľad. Konkrétne v

Zobrazení Premenných (Variable View) nastavujeme nasledovné charakteristiky (Řehák, Brom, 2015).

Premenná (Variable) je poradové číslo premennej. Softvér PSPP si ho zadáva sám od čísla 1 vyššie.

Meno (Name) Je názov premennej, ktorý tam musíme napísať ručne. Zvyčajne sa vyjadruje ako skratka. Napríklad ak máme premennú Vek mužov, tak sem napíšeme : vek_muzov. Staršie verzie programu vyžadovali používať písmená bez diakritiky a bez medzier, preto som tam uviedol podtrhovník. Každé Meno premennej môže byť v súbore uvedené len raz, dve rovnaké mená program PSPP nepovolí.

Typ (Type) je kategória charakteru premennej. Rehák a Brom (2015, str. 38) ju definujú ako: „typ informácie“ Na obrázku č. x vidíme, že kategórie premennej môžeme nastaviť rôzne. Môže byť:

Numerické číslo (Numeric), u ktorého máme počet Desatinných miest (Decimal places) ako aj **Počet znakov (Width)**

Čiarka (Comma) alebo **Bodka (Dot)**

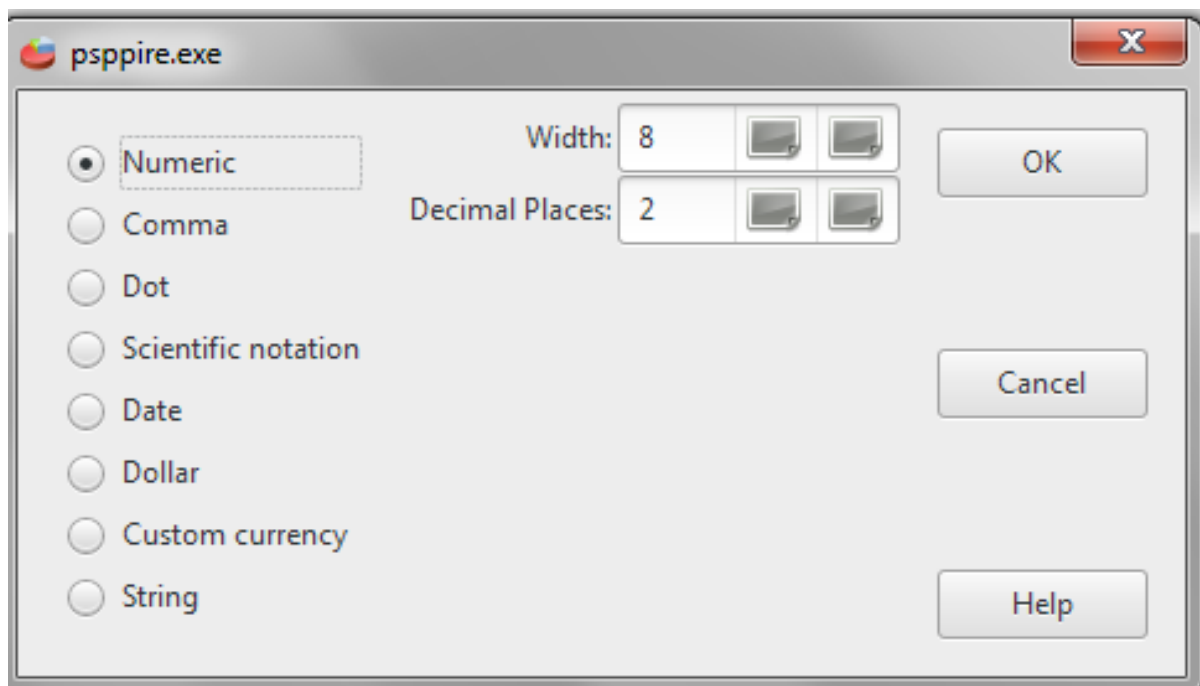
Vedecké označenie (Scientific Notation)

Dátum (Date)

Dolár (Dollar)

Iná mena (Custom Currency)

Text (String)



Obr. 2.1.7 Kategórie formátovania pri nastavení *Typu (Type)* u každej premennej.

Šírka (Width) počet znakov (písmen, čísel) premennej pri zobrazovaní dát. Softvér PSPP ponúka prednastavenú šírku na 8 znakov, avšak umožňuje počet znakov zvyšovať aj znižovať.

Desatinné miesta (Decimals) Označuje počet zobrazovaných desatinných miest.
Označenie (Label) alebo aj Rozšírený Názov premennej. V prípade ak je meno premennej dlhé, čo sa v praktickom výskume stáva často (napr. Zamestnanci po školení), tak kategória Label nám umožňuje napísať celý názov. Pri Výstupoch analýzy (Output) tieto budú obsahovať celý názov nielen Meno premennej (Name) teda skratku. Pokým mnohé iné charakteristiky premennej nemusíme, resp. zvyčajne nepotrebujeme nastavovať, tak Označenie (Label) vždy odporúčam vyplniť kvôli prehľadnosti vo výstupoch.

Označenie jednotlivých hodnôt (Value Labels) V prípade ak by premenná bola vo viacerých kategóriách (napr. pohlavie môže byť muž, ktorý by bol kódovaný 0 a žena, ktorá by bola kódovaná 1) tak v tomto stĺpci pomenujeme presne jednotlivé skupiny a priradíme im konkrétne kódy v podobe čísel. V okne *Dátové zobrazenie (Data view)* sa nám potom môžu zobrazovať ako hodnoty (teda čísla 0 alebo 1) alebo ako označenia (muž alebo žena). Pri štatistických analýzach potom zadefinujeme skupiny pomocou týchto kódov. Takýchto skupín môže byť z praktického hľadiska aj niekoľko desiatok podľa veľkosti súboru. Štatistické usudzovanie a pri mnohých analýzach pozostáva z porovnávania rozdielnych skupín. Napríklad ako sa odlišujú zamestnanci vo výkonnosti pred a po absolvovaní školenia. Ako úspešne riešia problém študenti prvého ročníka, druhého atď až po profesionálov. Mnohé štatistické testy (T testy, Anovy a ostatné) priamo vyžadujú aby ste označili konkrétne porovnávané skupiny a bez ich označenia analýza nie je možná. Preto je vždy veľmi dôležité správne vyplniť označenie jednotlivých hodnôt (Value Labels).

Chýbajúce hodnoty (Missing values) Softvér PSPP v prípade ak chýba hodnota premennej uvedie v hárku *Dátového zobrazenia (Data View)* bodku a nenechá okno prázdne. Tieto *Chýbajúce hodnoty (Missing values)* sú z analýzy vynechávané a softvér počíta len s konkrétne uvedenými hodnotami (číslami). Pri výstupoch v tabuľke potom uvedie *Počet chýbajúcich prípadov (Missing cases)*.

Stĺpce (Columns) Udávajú šírku stĺpcov v dátovej matici.

Zarovnanie (Align) Označuje zarovnanie v rámci bunky.

Premenná (Measure) Nastavenie premennej je dôležité a preto jej budeme venovať dostatok priestoru. Viac ohľadom premenných nájde čitateľ v podkapitole *Premenné a ich typy*. Je to typ premennej, ktorá v rámci softvéru PSPP môže byť definovaná ako Nominálna (Nominal), Ordinálna (Ordinal) a Škálová (Scale). Nominálna premenná je napr. muž, alebo žena. Nie je vyjadrená číslom a každý jednotlivец je buď muž, alebo žena. Čo sa týka možnosti štatistického spracovania takejto premennej naše možnosti sú malé. Pozostávajú prevažne z vyjadrenia frekvencií, prípadne analýzy pomocou Chí kvadrátu. Zvyšné dve škály merania (ordinálna a škálová) poskytujú lepšie a jemnejšie

možnosti analýzy. Ordinálna premenná zoraďuje (order) premennú do kategórií, alebo stupňov. Neplatí však že medzi kategóriami sú rovnaké vzťahy. Príkladom sú od škály rôznych dotazníkov, kde respondent odpovedá napr. nasledovne :Veľmi sa mi páči, páči sa mi, páči sa mi trochu, trochu sa mi nepáči, nepáči sa mi, veľmi sa mi nepáči. Nemôžeme síce povedať, že napr. veľmi sa mi páči je 2 krát viac než len páči sa mi, no vieme, že veľmi sa mi páči je silnejšie než páči sa mi. Posledná z možností, ktoré softvér PSPP ponúka je škálová premenná (scale). Teda odpovede sú na určitej škále, kde vieme presne povedať, že 4 je dva krát viac než 2 a presne polovica z 8, teda medzi premennými sú presné matematické vzťahy. Takto zadefinovaná premenná poskytuje najširšie a najjemnejšie možnosti štatistickej analýzy. Do kategórie škálová premenná zaradujeme dve škály merania: Pomerovú a Intervalovú.

Rola (Role) len pri niektorých špecifických procedúrach modelovania sa určuje Rola.

Imput premenná je v modeli ako prediktor, vstup, teda nezávislá premenná

Output premenná je v modeli výstupná, cieľová premenná, teda závislá

Both premenná vystupuje ako a vstupná (Imput) aj výstupná (Output)

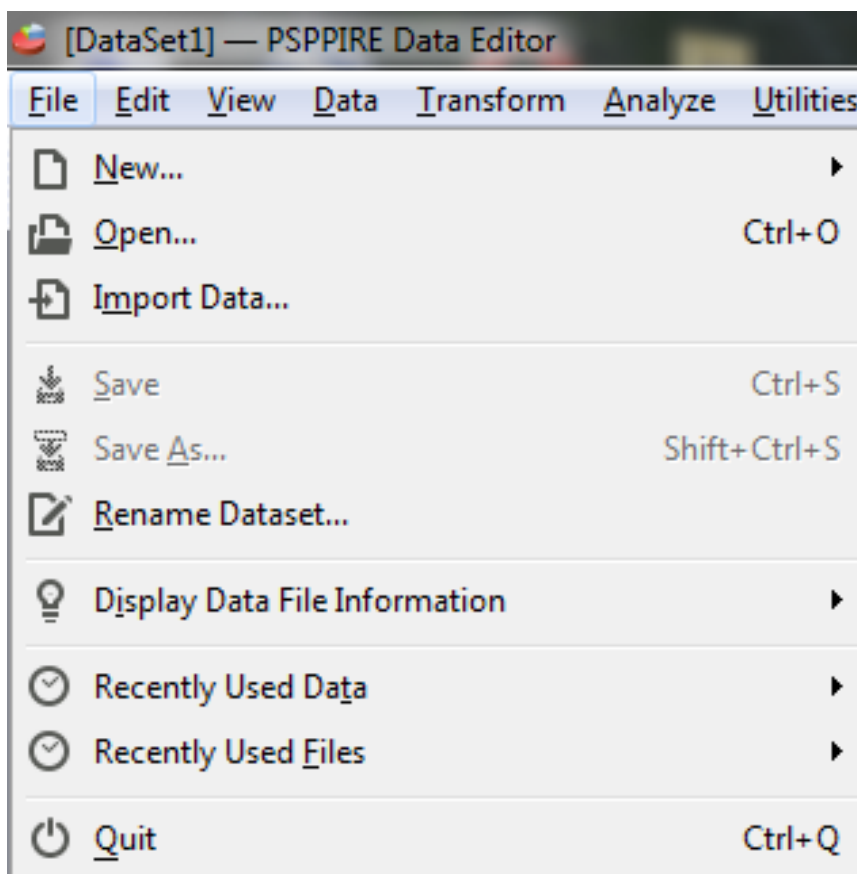
None rola premennej nie je priradená a táto je programe prednastavená

Partition premenná sa použije na rozdeľovanie súboru na tri časti (trénovacia, testovacia, validizačná)

Split podporná rola

4 PONUKA SÚBOR (FILE)

Ponuka File slúži na prácu zo súborom. V ponuke sa nachádza otvorenie nového a prázdneho súboru *Nový (New)* alebo už existujúceho *Otvoriť (Open)*. Pri otvorení nového súboru sa nám zobrazí dialógové okno na Obr. 2.1.1. Ďalej v ponuke File nájdeme *Importovanie delimitovaných textových dát (Import Delimited Text Data)*. V tejto ponuke sa taktiež nachádzajú operácie na uloženie aktualizovaného avšak už predtým uloženého súboru *Uložiť (Save)* ako aj uloženie existujúceho súboru pod novým názvom, prípadne ešte doteraz neuloženého súboru *Uložiť Ako (Save As)*. Príkaz *Premenovať dátový súbor (Rename dataset)* premenuje dátové okno. Ďalšie príkazy otvoria *Naposledy použité dáta, alebo súbory (Recently Used Data, Recently Used Files)*. Posledná operácia je zatvorenie aktuálne otvoreného súboru *Zatvoriť (Quit)*. Prakticky všetky príkazy v tejto ponuke sú určené na správu súborov. Do dát nezasahujú. V nasledujúcich podkapitolách sa zameriame na každú jednu časť ponuky.



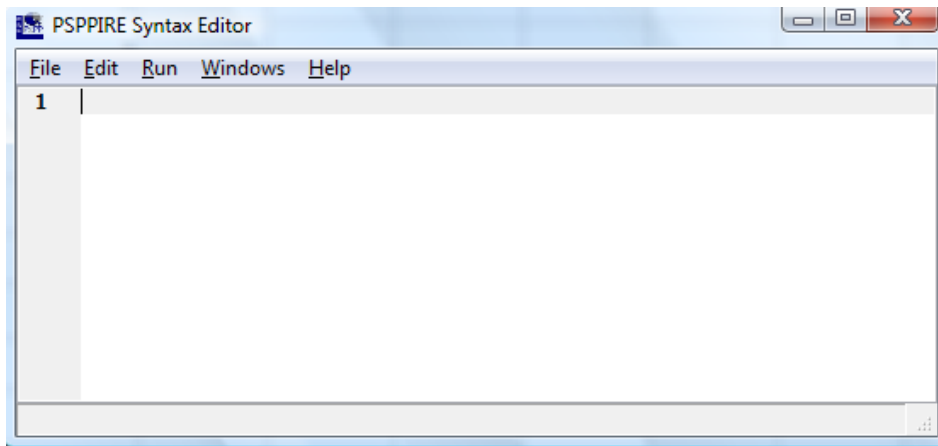
Obr. 2.1.2 Ponuka Súbor (*File*)

Založenie nového súboru (New)

Otvorí nový dátový súbor s dočasným názvom Dataset1, ktorý je prázdny a pripravený na vkladanie dát. Postup je nasledovný. Klikneme na:

Súbor (File) potom na *Nový (New)* a ak chceme vkladať dáta, tak *Dáta (Data)*

Na spustenie dialógového okna na vkladanie syntaxu použijeme nasledovný postup krokov. Klikneme na: *Súbor (File)*, potom na *Nový (New)* a *Syntax (Syntax)*.



Obr. 2.1.3 Dialógové okno syntax (*Syntax*)

Otvorenie súboru (Open):

Otvorí zvolený dátový súbor a načíta jeho obsah do okna *Úprava dát (Data Editor)*. Klikneme na *Súbor (File)*, potom *Otvoriť (Open)* a na záver Vyberieme požadovaný súbor.

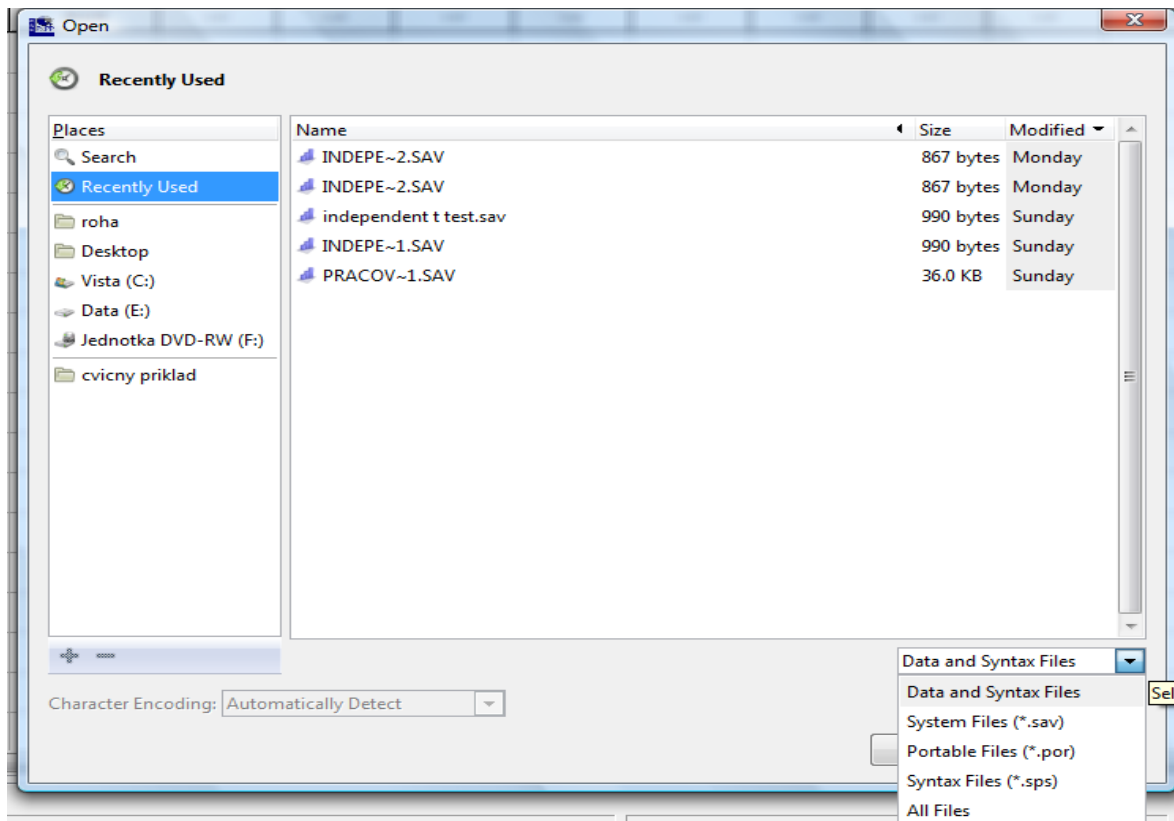
Príkaz umožňuje načítať zvolený dátový súbor určitého typu z konkrétneho disku, pracovného umiestnenia a adresára, ktoré sú zobrazené hore v dialógovom okne, obr. 4. Program PSPP umožňuje pracovať s dátovými súbormi 3 typov: systémové súbory - system files (*.sav), prenosné súbory - portable files (*.por), a syntaxové súbory – syntax files (*.sps). Tieto sú umiestnené v pravej časti dole v poli zo zoznamom. V dialógovom okne sa nachádza:

Umiestnenie (Places) je zoznam dostupných miest. Prvé dva príkazy, čo sa týka možnosti umiestnenie sú *Vyhľadávanie (Search)* a *Naposledy použité položky (Recently Used)*, ktoré zobrazujú posledné použité súbory programom PSPP. Okrem týchto umiestnení obsahuje aj všetky disky v počítači, DVD mechaniku, pracovnú plochu a konto užívateľa. Pod nimi, v ľavej dolnej časti užívateľ nájde príkazové tlačidlá + -, ktoré pridávajú, resp. uberajú dostupné miesta, odkiaľ sa môžu súbory otvárať.

Meno (Name) je konkrétny názov súboru.

Veľkosť (Size) je veľkosť súboru.

Upravené (Modified) je dátum poslednej úpravy.

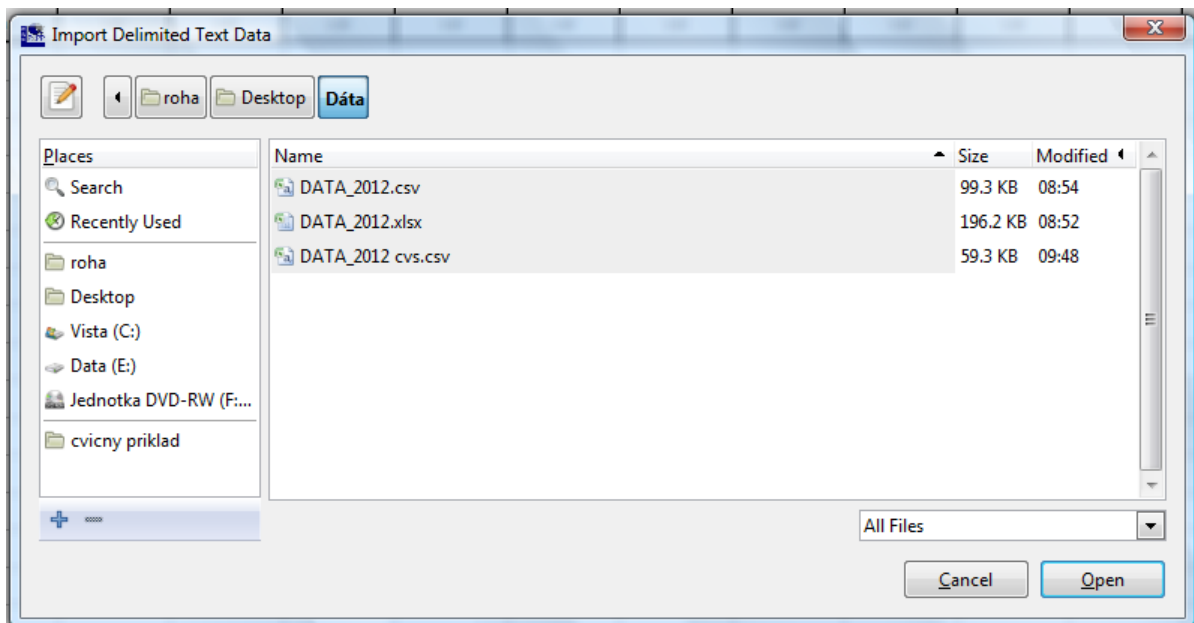


Obr. 2.1.4 Dialógové okno pre otvorenie súboru (*Open*)

Importovanie delimitovaných textových dát (Import Delimited Text Data)

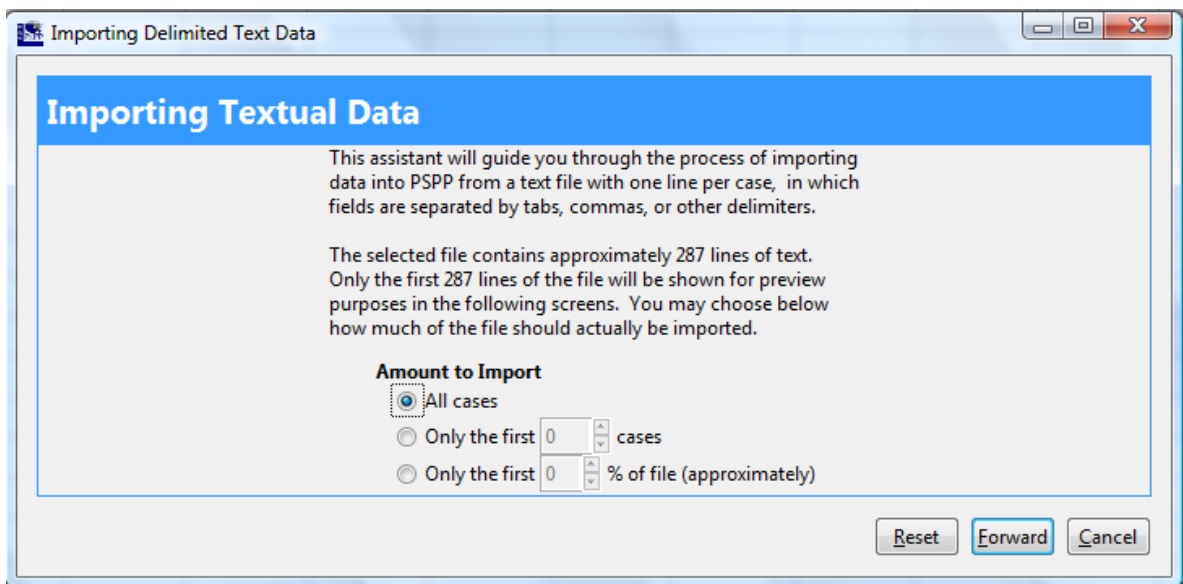
Importuje dáta z iných programov do programu PSPP.

Najčastejším zdrojom dát bežne používaných pri analýzach, ktoré nemajú formát *.sav je typ dát MS Excel s koncovkou .xls, a preto si vysvetlíme import z tohto programu. Pre import dát do programu PSPP z MS Excelu najprv musíme dáta v MS Exceli uložiť vo formáte .csv. Dáta môžeme importovať pomocou príkazu *Importovať oddelené textové dáta (Import Delimited Text Data)*, teda až vtedy, ak máme takýto súbor v MS Exceli už vytvorený. Vzhľad dialógového okna na obr. Obr. 2.1.5 je veľmi podobný vzhľadu pri otváraní súboru, ktorý sme podrobne opísali hore.



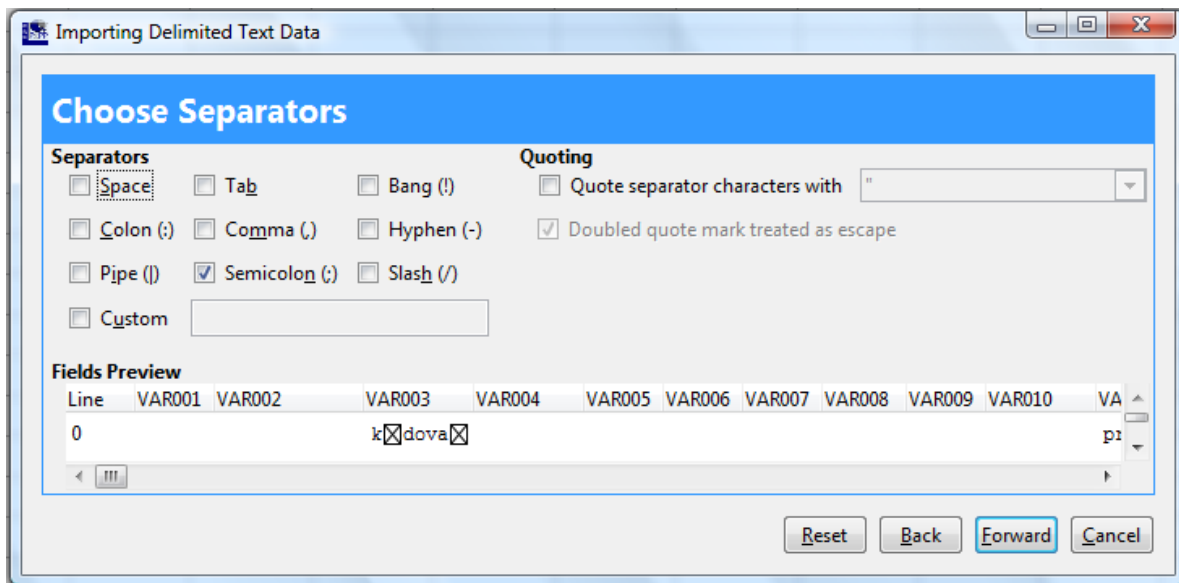
Obr. 2.1.6 Dialógové okno pre import dát (Import Delimited Text Data)

Sprievodca importom dát najprv overí, či chceme importovať *Všetky prípady (All cases)*. Ak áno pomocou príkazu *Ďalej (Forward)* sa presunieme do ďalšieho kroku.



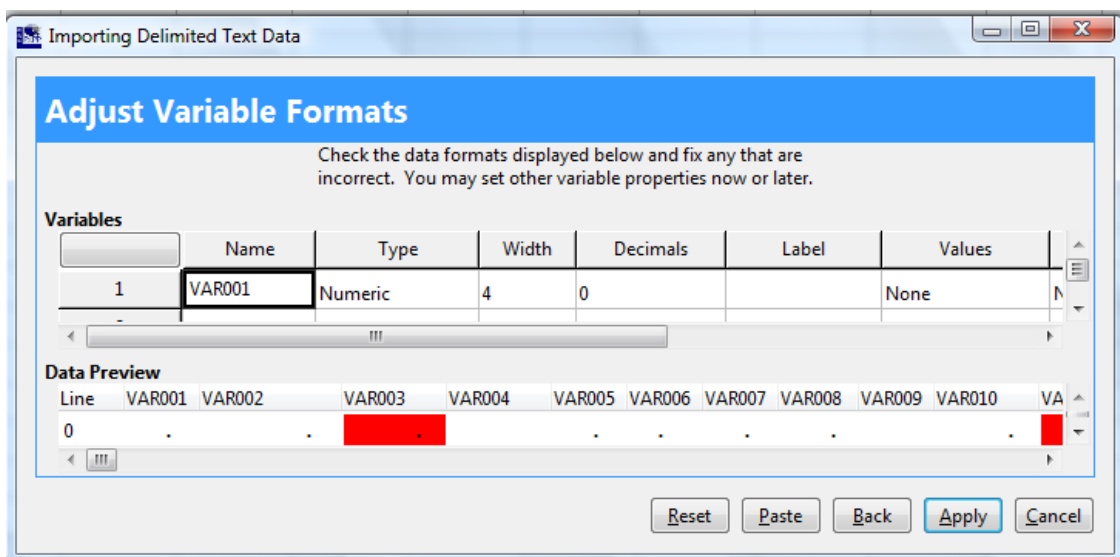
Obr. 2.1.7 Dialógové okno pre sprievodcu importom dát, výber počtu prípadov.

Následne je potrebné určiť ako sú oddelené dáta medzi sebou. Pre súbory s koncovkou *.cvs* je to *Bodkočiarka (Semicolon)*. Po zaškrtnutí príkazu bodkočiarka (Semicolon) pokračujeme príkazom *Ďalej (Forward)*.



Obr. 2.1.8 Dialógové okno pre sprievodcu importom dát, výber spôsobu rozdeľovania dát

Po príkaze Forward môžeme ešte dáta v sprievodcovi upravovať. Na záver spustíme príkaz *Vykonať (Apply)*. Je vhodnejšie dáta importovať a upravovať ich už vo finálnom súbore než upravovať ich v sprievodcovi.



Obr. 2.1.9 Dialógové okno pre sprievodcu importom dát, upravenie formátu.

Uloženie súboru (Save):

Tento príkaz uloží zvolený už existujúci dátový súbor do vybraného adresára a na vybraný zdroj. Klikneme na Súbor (File) potom Uložiť (Save) a Vyberieme súbor, do ktorého to chceme uložiť.

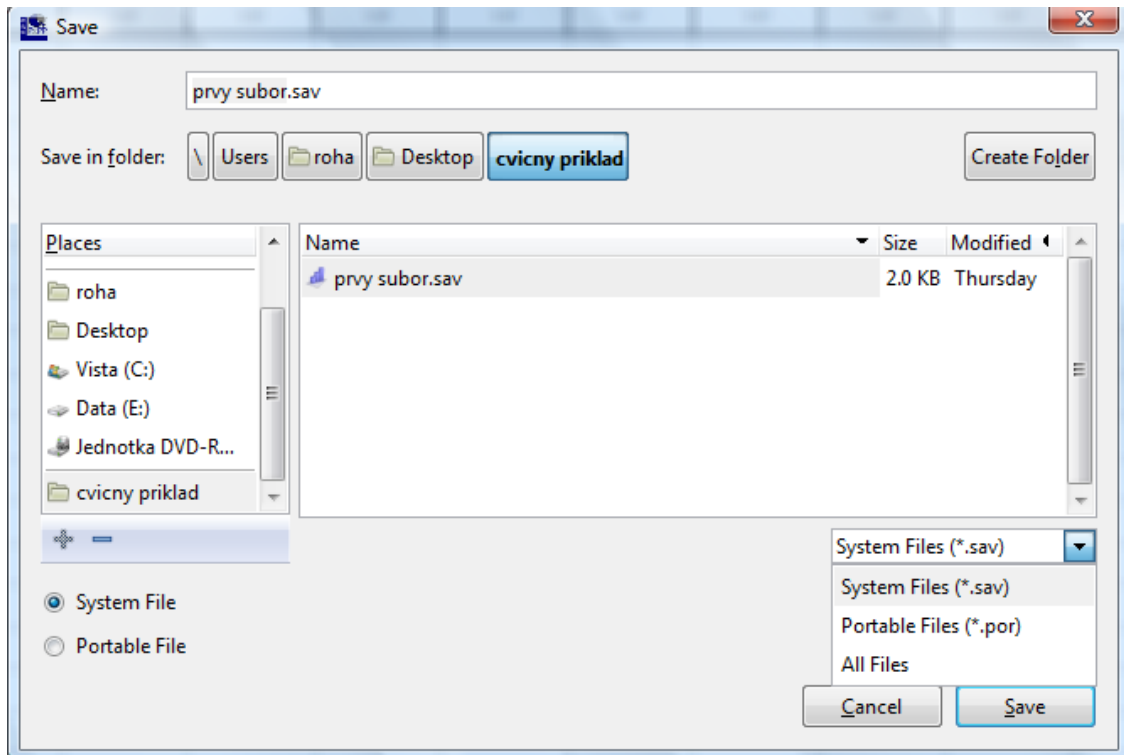
Dole uvedené základné charakteristiky vidíme na obr. č. 7

Meno (Name) je vstupné pole, ktorý slúži ako názov súboru, pre aký sme sa rozhodli. Meno je predefinované maskou *.sav.

Umiestnenie (Places) je zoznam dostupných miest na uloženie súboru. Obsahuje všetky disky v počítači, DVD mechaniku, pracovnú plochu a konto užívateľa.

Uložiť v adresári (Save in folder) zobrazuje aktuálne umiestnenie.

Formát súboru: Uloženie súboru umožňuje uložiť súbor v dvoch formátoch: system files (*.sav) a portable files (*.por)



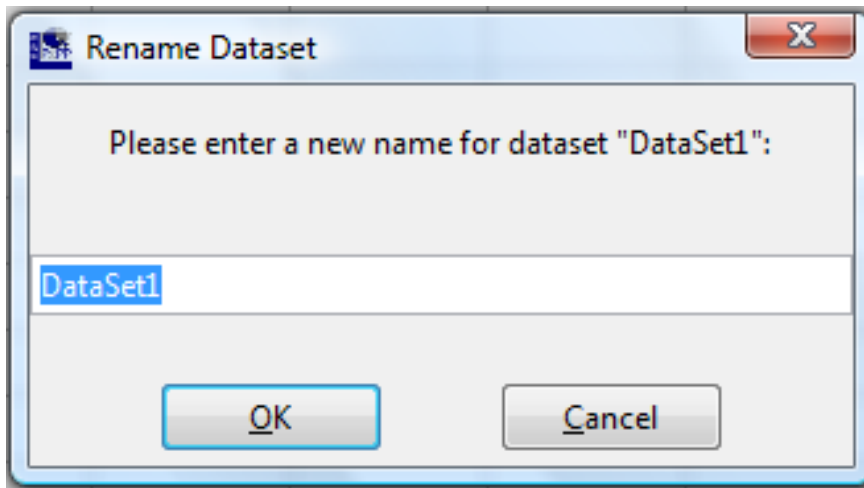
Obr. 2.1.10 Dialógové okno pre uloženie súboru (*Save*)

Uloženie súboru Ako (Save As):

Príkaz *Uložiť Ako (Save As)* uloží súbor pod iným názvom do nového súboru. Postup je rovnaký ako prípade príkazu *Uložiť (Save)*, kde klikneme na: Súbor (File) a potom na *Uložiť Ako (Save As)*.

Premenovanie dátového okna (Rename Dataset):

Premenuje dátové okno. Klikneme na Súbor (File) a potom na *Premenovať Dátové okno (Rename Dataset)*.



Obr. 2.1.11 Dialógové okno pre premenovanie dátového okna (*Rename Dataset*)

Zobrazenie informácií o dátovom súbore (Display data file information)

Program zobrazí informácie o dátovom súbore a informácie môžu byť zobrazené o práve *Otvorenom súbore (Working file)* alebo o *Externom súbore (External file)*. Informácie sú zobrazované v outputovom okne, kde je zobrazená *Premenná (Variable)*, jej *Opis (Description)* a jej *Umiestnenie (Position)*.

Naposledy používané dáta (Recently Used Data)

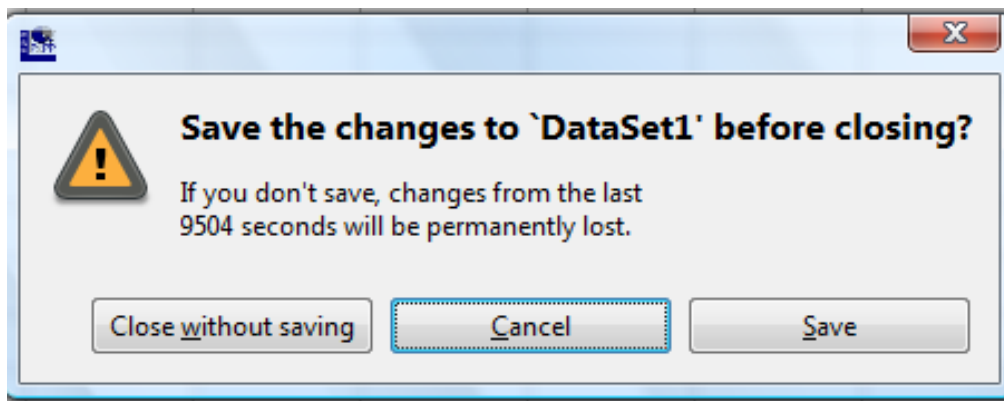
Tento príkaz zobrazí naposledy používané dátové súbory. Klikneme na *Súbor (File)* a potom na *Naposledy použité dáta (Recently Used Data)*.

Naposledy používané súbory (Recently Used Files)

Zobrazí naposledy používané nielen dátové súbory, ale aj výstupy alebo syntax. Klikneme na *Súbor (File)* a potom na *Naposledy použité súbory (Recently Used Files)*.

Ukončenie programu (Quit)

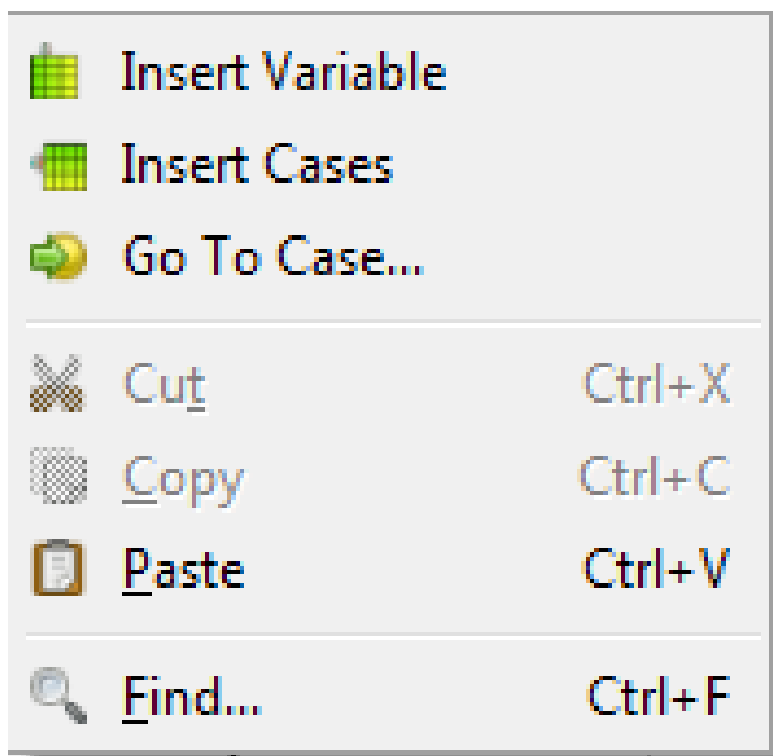
Tento príkaz ukončí činnosť programu PSPP, zatvorí všetky otvorené okná, zatvorí otvorené súbory a vypne celý program PSPP. V prípade, že neboli v niektorom zo súborov uložené zmeny, objaví sa okno v ktorom sa program PSPP pýta na uloženie zmien, predtým než sa sám zavrie (obr. 7). Klikneme na *Súbor (File)* a potom na *Ukončiť (Quit)*.



Obr. 2.1.12 Dialógové okno uloženie zmien pred zatvorením programu PSPP (*Exit*)

4.1 Ponuka Úpravy (*Edit*)

Ponuka *Úpravy (Edit)* umožňuje *Vložiť premennú (Insert Variable)* alebo konkrétne *Ďalšie prípady (Insert Cases)*. Taktiež umožňuje ísť na *Konkrétny prípad (Go To Case)*. A pri práci zo schránkou dát umožňuje dáta *Vyňať (Cut)*, *Kopírovať (Copy)* a *Prilepiť (Paste)*. Na vyhľadávanie použijeme príkaz *Hľadať (Find)*.



Obr. 4.1.1 Ponuka *Úpravy (Edit)*

Vloženie novej premennej (Insert Variable).

Vloží novú, práve jednu premennú do dátovej matice. Vloží ju vľavo od miesta, kde sa momentálne nachádza kurzor v karte Data view a nahor v karte Variable view. Klikneme na *Upraviť (Edit)* a následne na *Vložiť premennú (Insert Variable)*.

Príkaz vložiť nový prípad (Insert Cases)

Vloží prázdny riadok. Tento funguje v iba v karte Data view a vloží prázdny riadok nad miesto, kde sa momentálne nachádza kurzor. Klikneme na *Upraviť (Edit)* a potom na *Vložiť prípady (Insert Cases)*.

Príkazy: *Vyňať (Cut)*, *Kopírovať (Copy)*, *Prilepiť (Paste)* slúžia na prácu s konkrétnymi dátami, blokmi dát, textovými poľami alebo syntaxom. Sú totožné s príkazmi, ktoré čitateľ pozná z práce s MS Office a z iných programov. Všetky hore uvedené príkazy sú prístupné aj pri kliknutí na pravé tlačidlo myši.

Príkaz Vyňať (Cut)

Vyjme dáta, premennú, časť dát alebo syntax. Tento príkaz vyjme označené konkrétne dáta, premennú, časť dát, textového poľa alebo syntaxe a presunie ich do schránky. Na pôvodnom mieste označené a vyňaté dáta alebo syntax vymaže. Dáta označíme, a potom klikneme na *Upraviť (Edit)*, potom na *Vyňať (Cut)*.

Príkaz kopírovať (Copy)

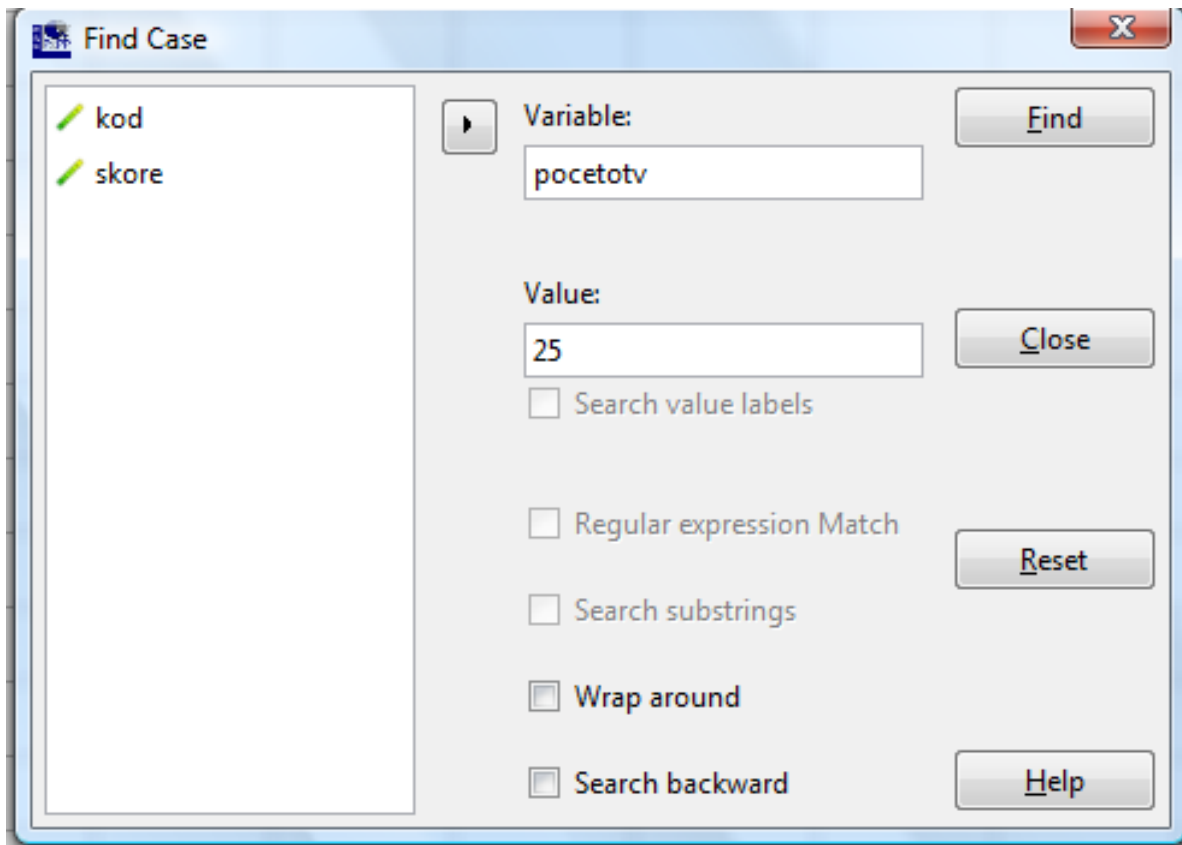
Skopíruje dáta, premennú, časť dát alebo syntax. Skopíruje označené dáta, časť dát, textového poľa, premennú alebo syntaxe do schránky, kde sú dáta pripravené na vloženie. Na rozdiel od príkazu *Vyňať (Cut)* ponechá skopírované dáta na pôvodnom mieste. Dáta označíme, a potom klikneme na *Upraviť (Edit)* a potom na *Kopírovať (Copy)*.

Príkaz Prilepiť (Paste)

Vloží obsah schránky na miesto označené nastavením kurzora. Klikneme na *Upraviť (Edit)*, následne na *Prilepiť (Paste)*.

Vyhľadanie konkrétneho prípadu (Find Case)

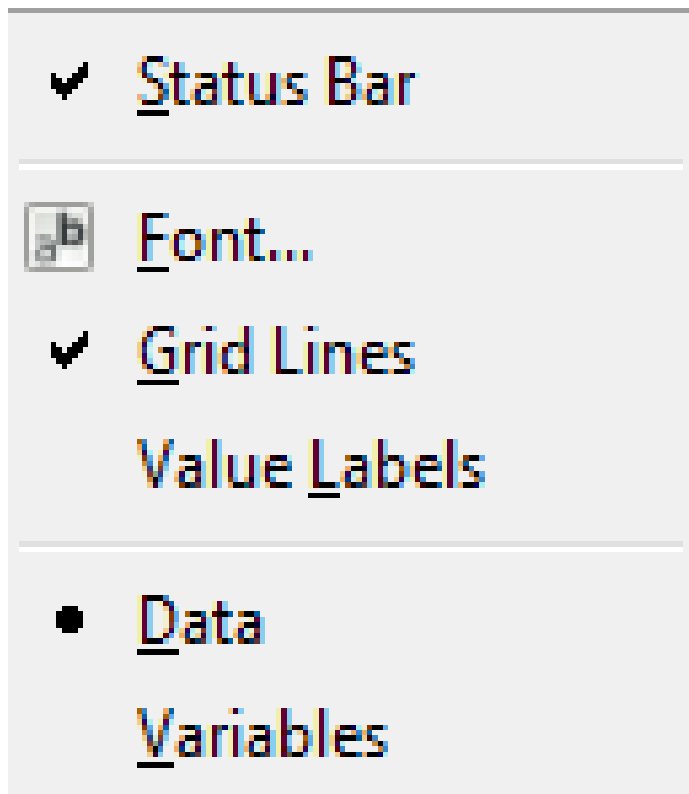
Vyhľadá konkrétny prípad. Pri vyhľadávaní konkrétneho prípadu musíme definovať premennú (Variable) a hodnotu hľadanej premennej (Value). Zaškrtávacie políčko *Hľadať naspäť (Search backward)* hľadá od miesta kurzora späť (nahor), inak program hľadá vpred (nadol). Klikneme na *Upraviť (Edit)* a následne na *Hľadať (Find)*.



Obr. 4.1.2 Vyhľadavanie konkrétnych prípadov

4.2 Ponuka Pohľad (View)

Ponuka *Pohľad (View)* umožňuje zobrazovať a skrývať vzhľad prvkov dátového okna programu PSPP. Príkaz *Stavový riadok (Status Bar)* skrýva alebo odkrýva stavový riadok. Príkaz *Font (Font)* nastavuje typ písma, jeho veľkosť a štýl. Mriežka a jej zobrazenie, resp. skrytie sa upravuje cez príkaz *Mriežka (Grid Lines)*. Prepínanie medzi hodnotami a ich konkrétnymi popismi umožňuje príkaz *Označenie Hodnôt (Value Labels)*. Posledné dva príkazy umožňujú prepínať medzi záložkami, a to *Variable View* a *Data View*. Ponuka *Pohľad (View)* je zameraná na vzhľadovú, resp. formátovaciu časť programu PSPP.



Obr. 4.2.1 Ponuka Pohľad (View)

Príkaz Stavový riadok (Status Bar)

Zobrazí stavový riadok. Tento príkaz zapne alebo vypne zobrazenie Stavového riadka v spodnej časti okna. Klikneme na *Pohľad (View)* a následne na *Stavový riadok (Status Bar)*.

Príkaz Font (Font)

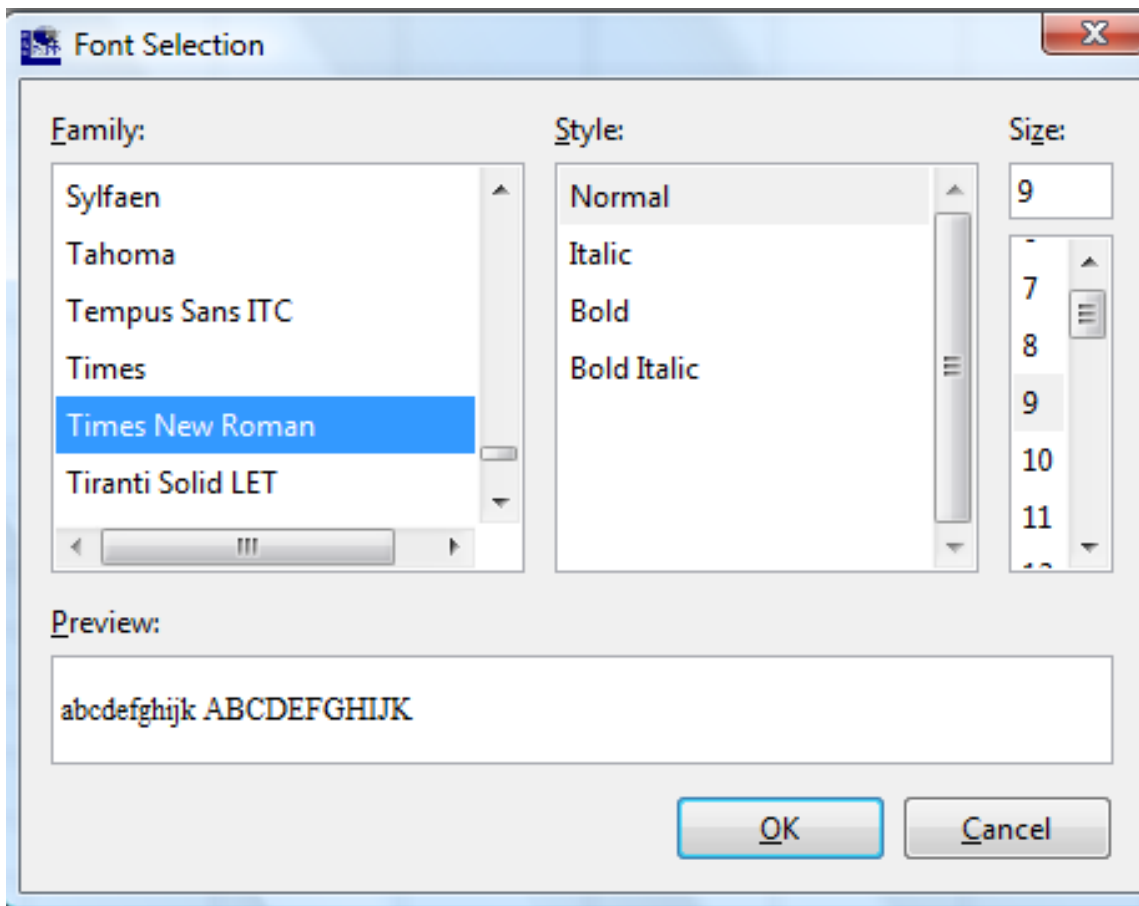
Nastaví formát písma pre dátovú maticu. Klikneme na *Pohľad (View)* a potom na *Fond (Font)*. V dialógovom boxe sa nachádza :

Typ písma (Family) znamená konkrétne typy písma.

Štýl (Style) obsahuje štyri skupiny štýlov: Normal, Italic, Bold a Bold Italic.

Veľkosť (Size) upravuje veľkosť písma.

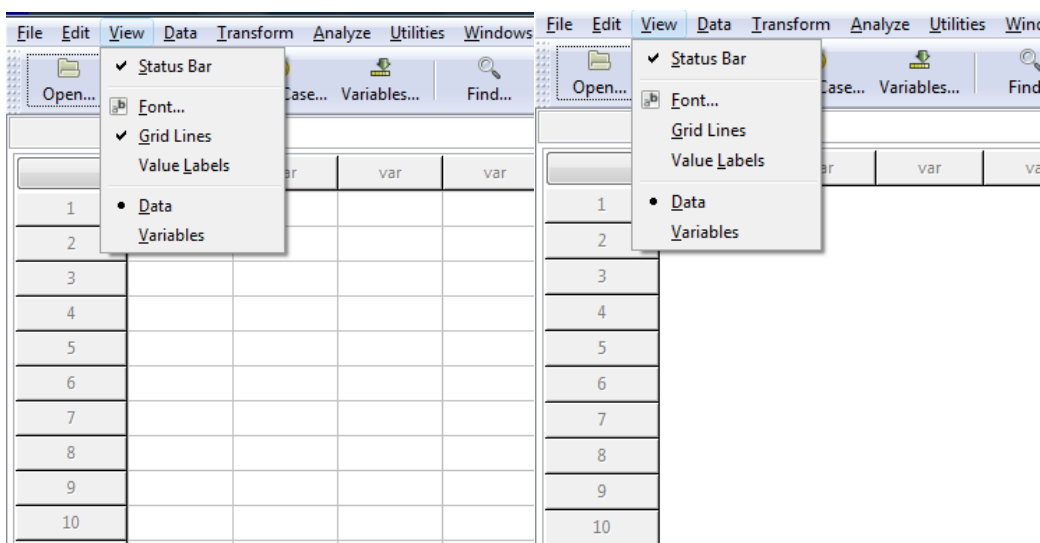
Náhľad (Preview) Zobrazí ako kontrétne bude vyzerat' zvolená kombinácia Typu písma, štýlu a jeho veľkosti.



Obr. 4.2.2 Dialógové okno Font (*Font*)

Príkaz Zobrazenie Mriežky (Grid lines)

Zobrazí, resp. skryje mriežku. Klikneme na *Zobrazíť* (*View*) a následne na *Mriežku* (*Grid lines*).



Obr. 4.2.3 Zapnutá a vypnutá mriežka

Z dôvodu prehľadnosti je vhodné mať mriežku zapnutú, aby sa nám jednotlivé prípady nemiešali.

Príkaz Opisy hodnôt (Value Labels)

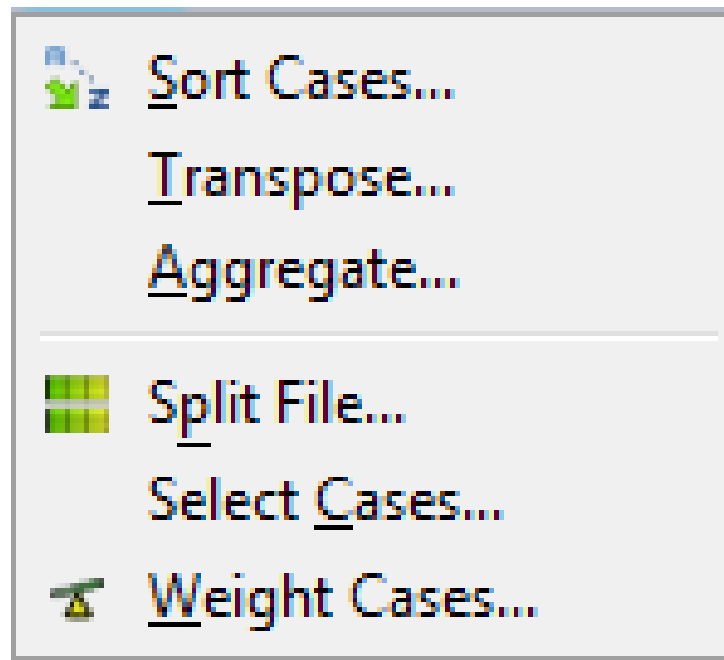
Zobrazí, resp. skryje Opisy hodnôt. Pre užívateľa to prakticky znamená že namiesto kódov, napr. 0 alebo 1, uvidí opisy hodnôt napr. muž a žena. Klikneme na Zobrazit' (View) a následne na Označenie hodnôt (Value Labels).

Príkaz Prepínanie medzi záložkami dátového okna (Data)

Zobrazí, resp. skryje Dátové Zobrazenie (Data View), resp. Zobrazenie Premenných (Variable View). Kliknem na Zobrazenie (View) a následne na Dáta (Data) a premenné (Variables).

4.3 Ponuka Dáta (Data)

V ponuke dáta je možné dátové prípady *Triediť (Sort Cases)* a taktiež *Transponovať dátovú maticu (Transpose)*. V ponuke sa nachádza aj možnosť dáta *Agregovať (Aggregate)*, ako aj dáta rozdeliť podľa potreby analýzy príkazom *Rozdeliť súbor (Split File)*, *Vybrať dáta (Select Cases)* a *Vážiť dáta (Weight Cases)*. Táto ponuka je zameraná na prácu s dátami a ich prípravu, úpravu na analýzu.

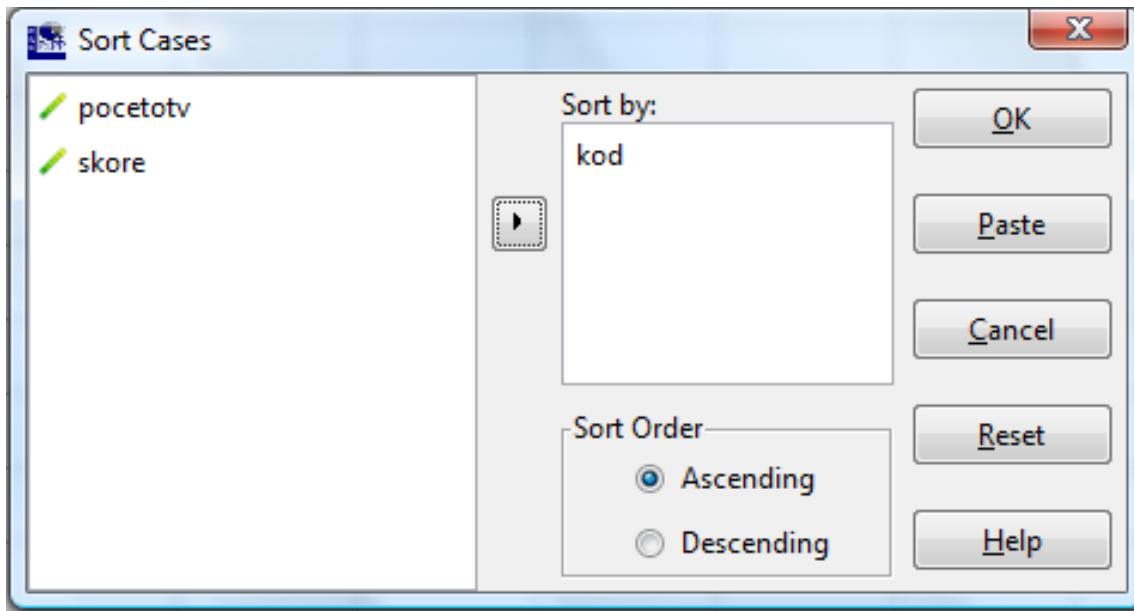


Obr. 4.3.1 Ponuka Dáta (Data)

Príkaz Zoradiť Prípady (Sort Cases)

Zoradí dáta v súbore podľa vopred zadaných kritérií.

V dialógovom okne vyberieme premennú, ktorú chceme zoradiť a označíme ju. Šípkou ju presunieme do okna *Zoradiť podľa (Sort by)*. V zaškrťavacích políčkach môžeme vybrať *Spôsob zoradenia (Sort Order)*, a to *Vzostupný (Ascending)* alebo *Zostupný (Descending)*.

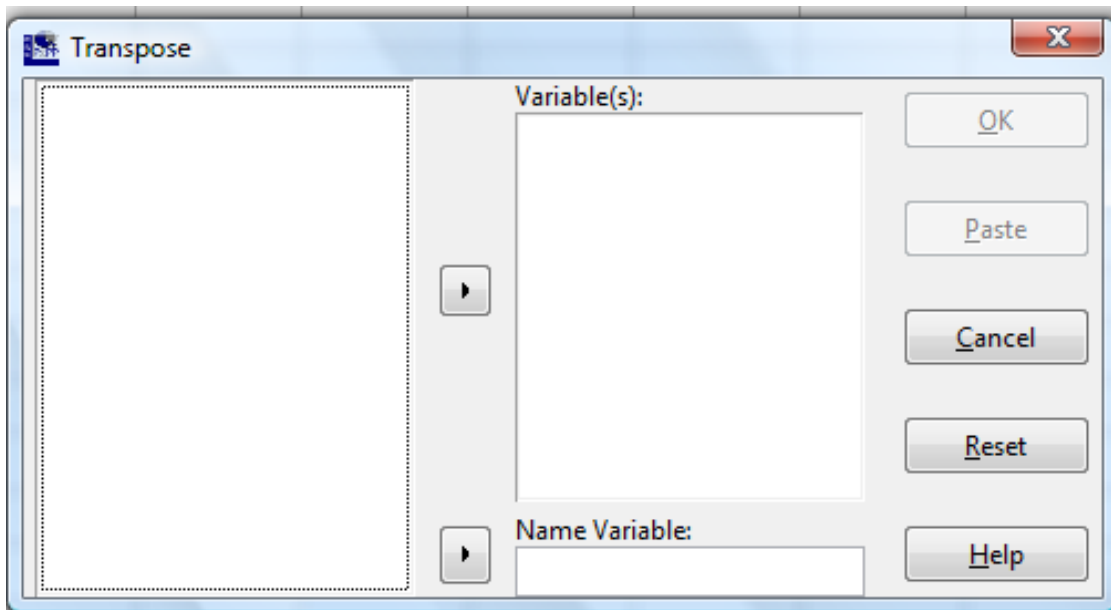


Obr. 4.3.2 Dialógové okno Zoradiť Prípady (*Sort Cases*)

Klikneme na *Dáta (Data)* a následne na *Zoradiť prípady (Sort Cases)*.

Príkaz Transpozícia dátovej matice (Transpose)

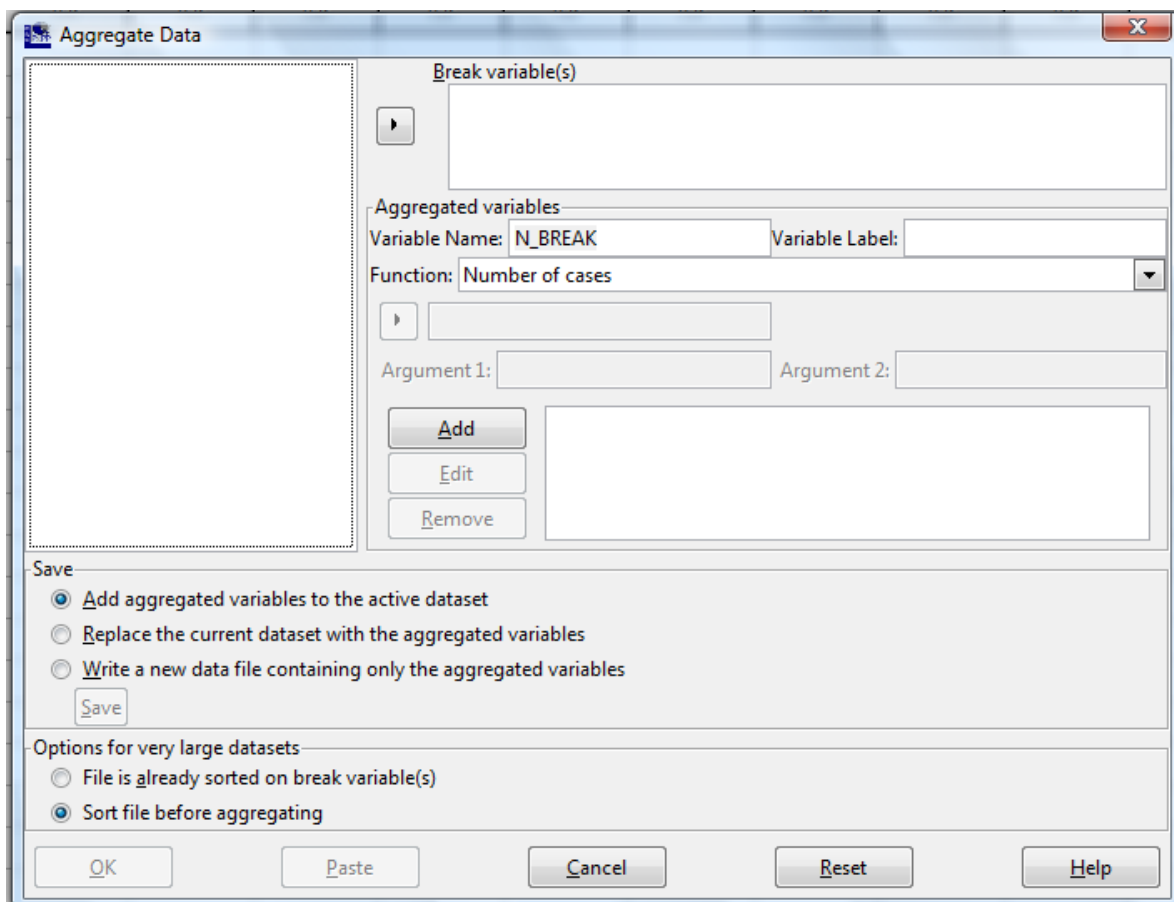
- Transpozícia dátovej matice zmení riadky v stĺpce a naopak. Je možné transponovať všetky alebo len niektoré vybrané premenné. Všetky premenné sa nám zobrazujú v ľavej časti okna. Vybrané premenné, ktoré chceme transponovať presunie do okna *Premenné (Variable)*. *Meno premennej (Name Variable)* je tá premenná, ktorá bude novými názvami pre transponované premenné. Klikneme na *Dáta (Data)* a následne na *Transponovať (Transpose)*.



Obr. 4.3.3 Transpozícia dátovej matice

Príkaz Agregovanie dát (Aggregate)

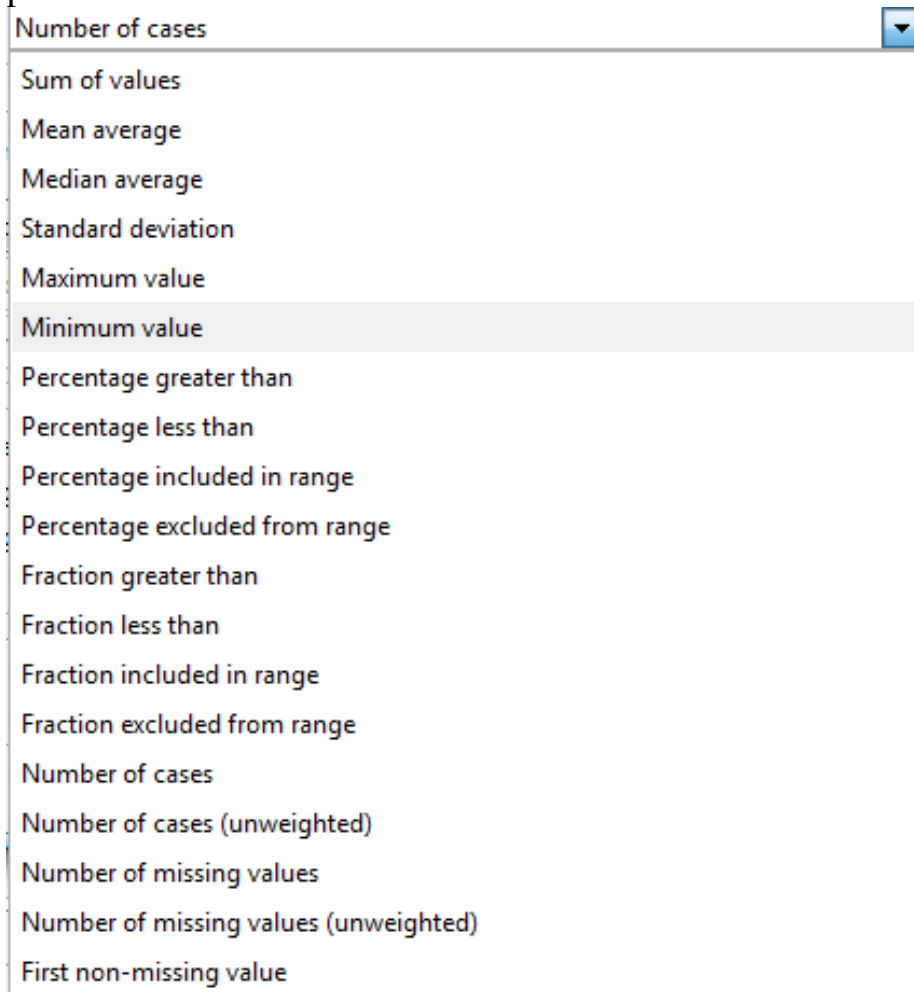
Vytvorí nový súbor agregovaním dát z pôvodného súboru.



Obr. 4.3.4 Agregovanie dát

Umožňuje agregovať (zoradovať, triediť) dáta pomocou 19 funkcií. V ľavej časti sú uvedené všetky premenné v dátovom súbore. Po výbere konkrétnej premennej, podľa ktorej sa bude agregovať súbor, túto presunieme do *okna Rozdeľovacia premenná/é (Break variable(s))*. U agregovanej premennej je možné zmeniť pôvodné meno a nové *Meno premennej (Variable Name)* a taktiež aj *Opis (Variable Label)*

Funkcia (Function) je funkcia podľa ktorej sa agregujú dáta. Užívateľovi ponúkne pole zo zoznamom s 19 funkciami.



Obr. 4.3.5 Funkcia, podľa ktorej sú agregované dáta

Po agregácii dát sa nás program PSPP opýta na spôsob uloženia. *Uložiť (Save)*, obsahuje tri zaškrťavacie políčka, ktoré užívateľovi ponúkajú tri rôzne možnosti ako uložiť výsledok agregovania dát.

Uložiť agregované premenné do aktívneho dátového súboru (Add aggregated variables to the active dataset).

Nahradiť súčasný dátový súbor súborom s agregovanými premennými (Replace the current dataset with the aggregated variables).

Vytvor nový dátový súbor obsahujúci iba agregované premenné (Write a new data file containing only aggregated variables).

Možnosti pre veľké dátové súbory (Options for very large datasets) obsahujú dve zaškrtačacie políčka:

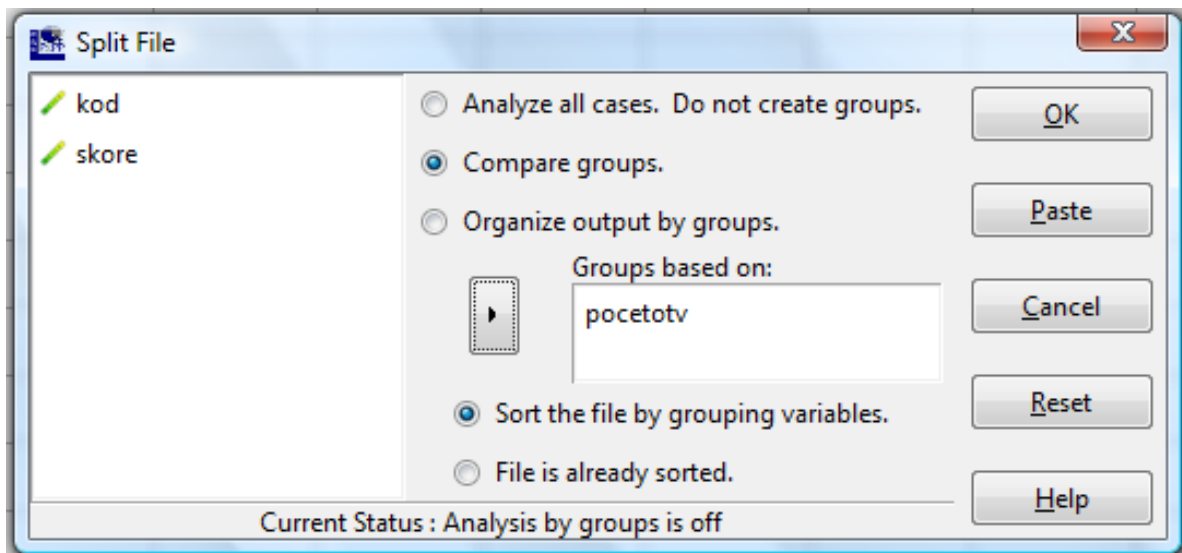
- Súbor je už rozdelený podľa agregovanej premennej (File already sorted on break variable(s)). Ďalej máme k dispozícii príkaz *Zoraď súbor pred agregovaním (Sort file before aggregating)*. Agregovanie dát spustíme kliknutím na *Dáta (Data)* a potom *Agregovať (Aggregate)*

Príkaz Rozdelenie dátového súboru (Split file)

Príkaz rozdelí dátový súbor a používame ho v prípade ak chceme rozdeliť na základe konkrétnych kritérií súbor na skupiny, ktoré môžeme potom analyzovať. V ľavej časti sa nachádzajú všetky premenné v súbore. V pravej časti máme tri zaškrtačacie políčka, ktoré sa týkajú toho, ako budú organizované výstupy. V prípade, ak nechceme porovnávať skupiny, ale analyzovať všetky prípady spolu, tak zaškrtneme políčko : *Analyzovať všetky prípady. Nevytvárať skupiny (Analyze all cases. Do not create groups.)*. Ak chceme výstup, kde budú všetky skupiny uvedené spolu s ostatnými a aj s grafmi, potom zaškrtneme: *Porovnať skupiny (Compare groups)*. Ak chceme, aby bola každá skupina uvedená samostatne aj s celým výstupom, tak zaškrtneme políčko *Organizovať output podľa skupín (Organize output by groups)*. Premennú/é, ktorú chceme triediť presunieme do okna *Skupiny založené na (Groups based on)*. PSPP zoradí súbor podľa zoskupujúcej premennej, keď zaklikneme zaškrtačacie políčko: *Zoraď súbor podľa zoskupujúcich premenných (Sort the file by grouping variables)*. Ak je súbor už zoradený, potom môžeme pre urýchlenie zaklinúť : *Súbor je už zoradený (File is already sorted)*. Z ponuky dáta ide o jeden z najčastejšie používaných príkazov.

Rozdelenie dátového súboru

Na rozdelenie dátového súboru budeme postupovať nasledovne. Klikneme na *Dáta (Data)* a potom na *Rozdelenie súboru (Split file)*, potom sa nám zobrazí nasledovné okno.

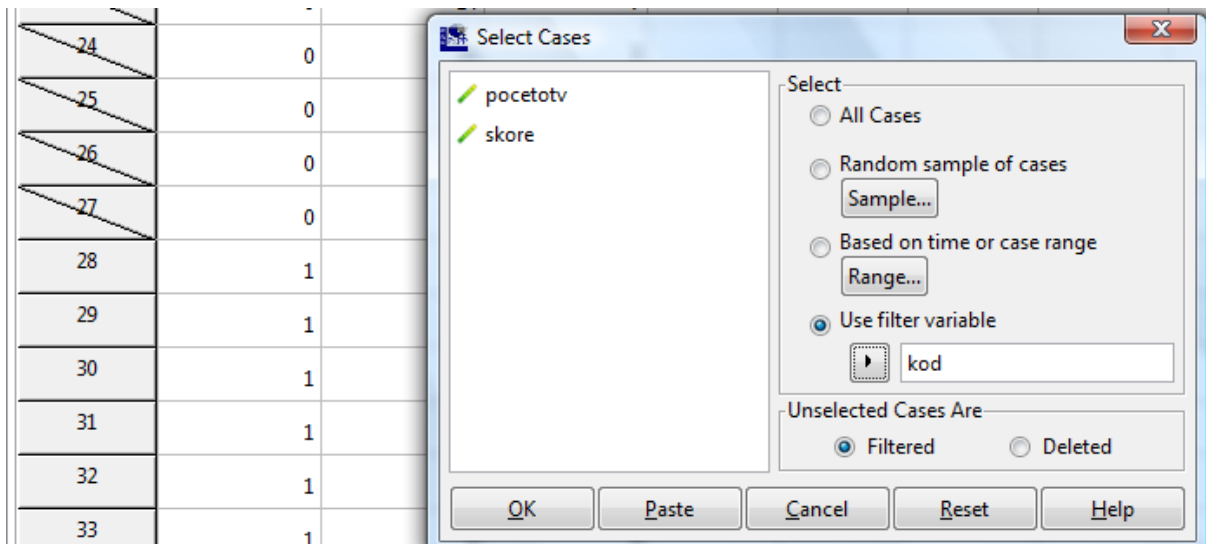


Obr. 4.3.6 Rozdelenie dátového súboru

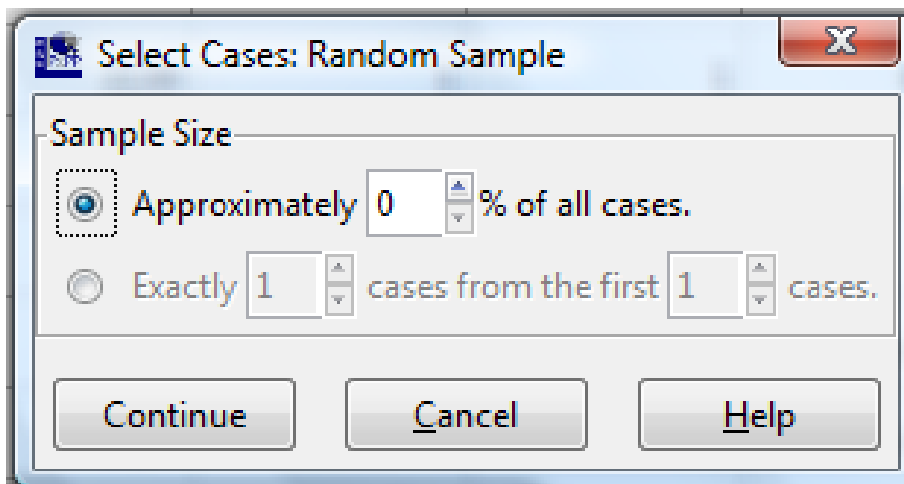
Aktívny užívateľ programu PSPP bude ponuku Split File využívať pomerne často. Zvlášť v prípadoch kedy bude analyzovať viacero skupín a bude chcieť vytvoriť výstupy pre každú zo skupín aby mohol predbežne porovnať výsledky. V takomto prípade zaškrtnie Organizovať výstup podľa skupín (*Organize output by groups*).

Príkaz Výber prípadov (Select Cases)

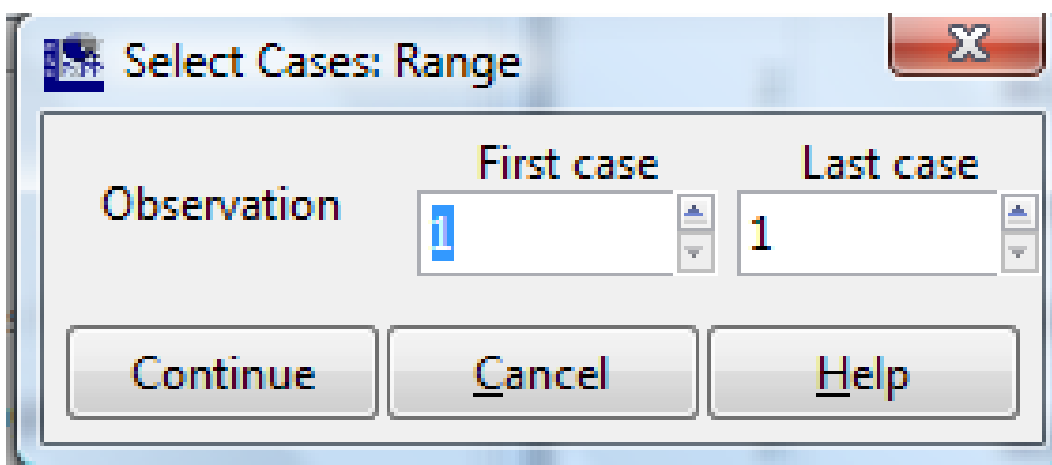
Vyselektuje konkrétne prípady podľa stanovených kritérií. Výber prípadov z celého súboru pre ďalšiu analýzu pomocou príkazu *Vyber prípady (Select Cases)*. Vybrať môžeme *Všetky Prípady (All Cases)*, *Náhodným výberom prípadov (Random sample of cases)*, Výberom v časovom alebo prípadovom intervale. Náhodný výber prípadov ponúka príkazové tlačidlo *Sample*, ktoré pri otvorení ponúkne náhodný výber prostredníctvom dvoch možností (Obr. 4.3.7 Výber prípadov: náhodný výber) a to ako: *Približné percento všetkých prípadov (Approximately % of all cases)*, alebo *Presný počet prípadov z prvých x prípadov (Exactly cases from the first x cases)*. Výber v časovom alebo prípadovom intervale umožňuje vybrať prípady v intervale od *1 Prvého prípadu (First Case)* po *Posledný prípad (Last Case)*. Posledné možnosti, ktoré toto okno ponúka sú spôsoby ako nakladať s Dátami, ktoré neboli prostredníctvom hore uvedených filtrov vybrané. Pre tieto nevybrané dáta sú dve možnosti: buď budú nevybrané prípady *Filtrované*, resp. *Vynechané z analýzy (Filtered)*, alebo *Zmazané (Deleted)*.



Obr. 4.3.8 Výber Prípadov



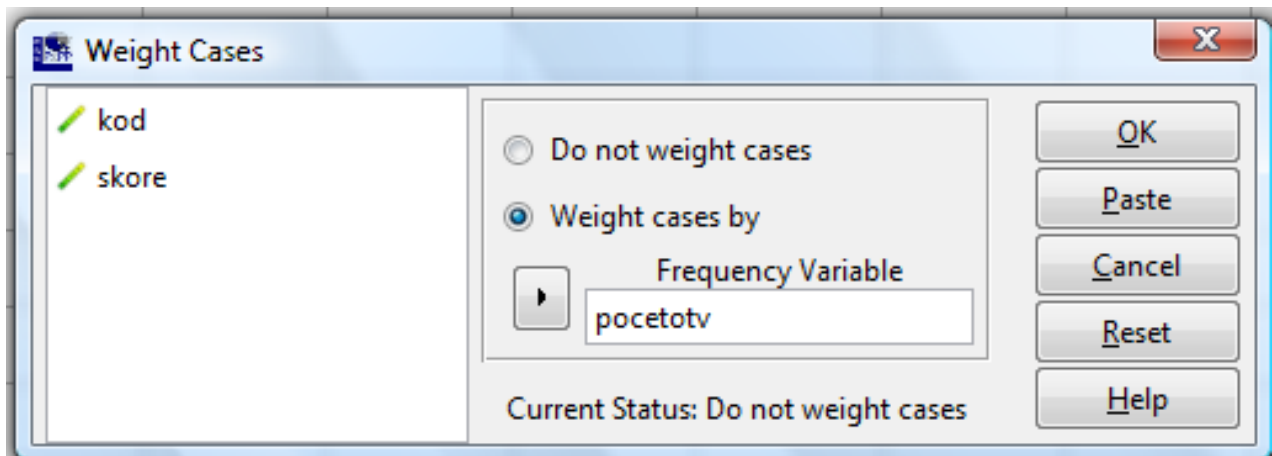
Obr. 4.3.9 Výber prípadov: Náhodný výber



Obr. 4.3.10 Výber prípadov z intervalu

Príkaz Váženie prípadov (Weight Cases)

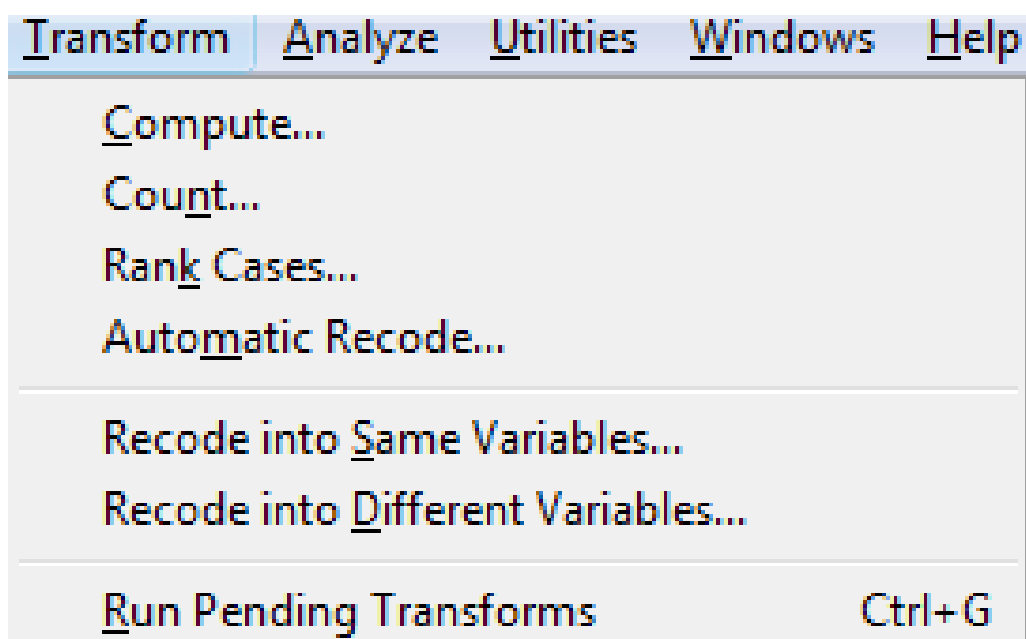
Príkaz priradí prípadom vybranú váhu a používame ho vtedy ak chceme prípadom priradiť vyššiu alebo nižšiu váhu. Klikneme na *Dáta (Data)* a *Vážiť prípady (Weight cases)*. V ľavej časti sú všetky premenné a ak chceme vybrať premennú, podľa ktorej budeme prípady vážiť, tak zaškrtneme tlačidlo *Vážiť prípady podľa (Weight cases by)* a presunieme premennú do Frequency Variable. Ak chceme zrušiť váženie prípadov, tak zaškrtneme *Nevážiť prípady (Do not weight cases)*.



Obr. 4.3.11 Váženie prípadov

4.4 Ponuka Transformácia (Transform)

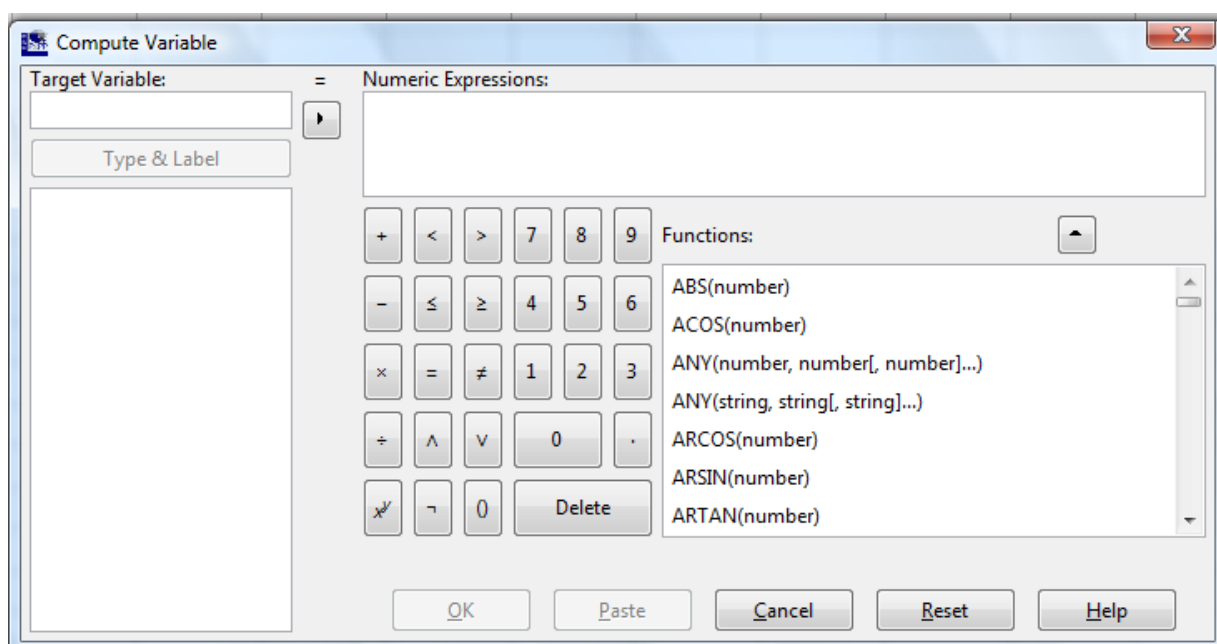
Ponuka Transform slúži k odvodzovaniu nových premenných z už existujúcich a k ich transformáciám. Obsahuje tieto možnosti: *Výpočet novej premennej (Compute)*, *Počet (Count)*, *Výpočet poradovej štatistiky (Rank Cases)*, *Automatické Prekódovanie (Automatic Recode)*, *Rekódovanie hodnôt do tej istej premennej (Recode into Same Variables)*, *Rekódovanie hodnôt do inej premennej (Recode into Different Variables)*. Posledný príkaz, *Spustenie odložených transformácií (Run Pending Transforms)*, spustí odložené transformácie.



Obr. 4.4.1 Transformácia dát

Príkaz Výpočet novej premennej (Compute)

Príkaz vypočíta novú premennú z existujúcich premenných. Okno obsahuje v ľavej časti *Názov premennej (Target Variable)*, ktorú budeme vytvárať. Pod názvom sa nachádza *Typ premennej* a jej *Opis (Type&Label)*. Vpravo dole je zoznam všetkých premenných, ktoré sa v súbore nachádzajú. V pravej časti je *Numerické vyjadrenie (Numeric Expressions)* a *Zoznam funkcií (Functions)*. Tento zoznam funkcií je bohatý a obsahuje desiatky rôznych funkcií. V strede sa nachádza kalkulačka.

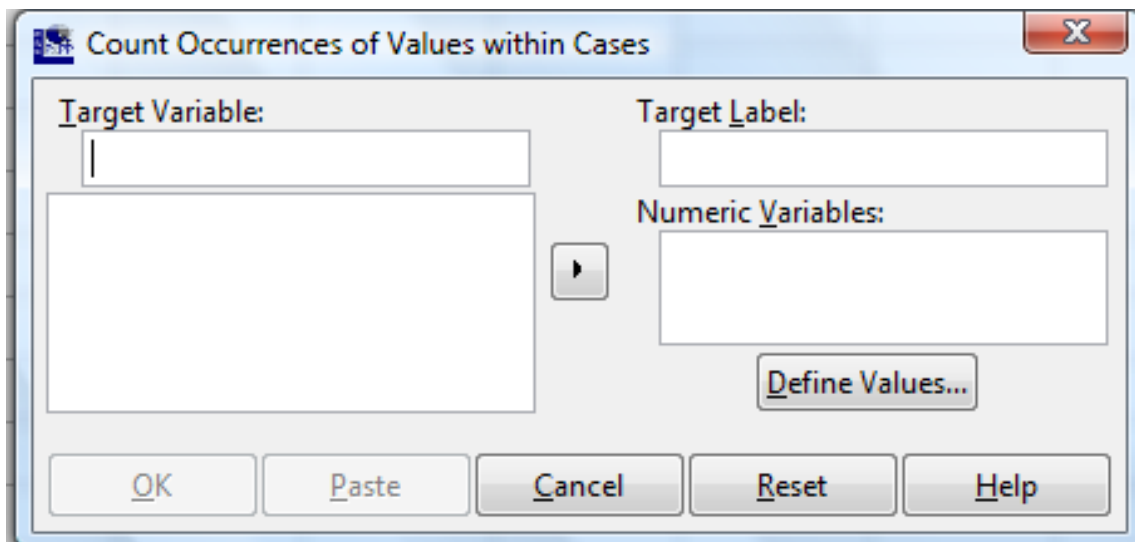


Obr. 4.4.2 Výpočet novej premennej

Pri výpočte novej premennej klikneme na Transformovať (Transform) potom na Vypočítať (Compute). Následne v dialógovom okne vyberieme konkrétnu funkciu, príkaz a premennú, na ktorú sa má daný príkaz aplikovať.

Príkaz Počet (Count)

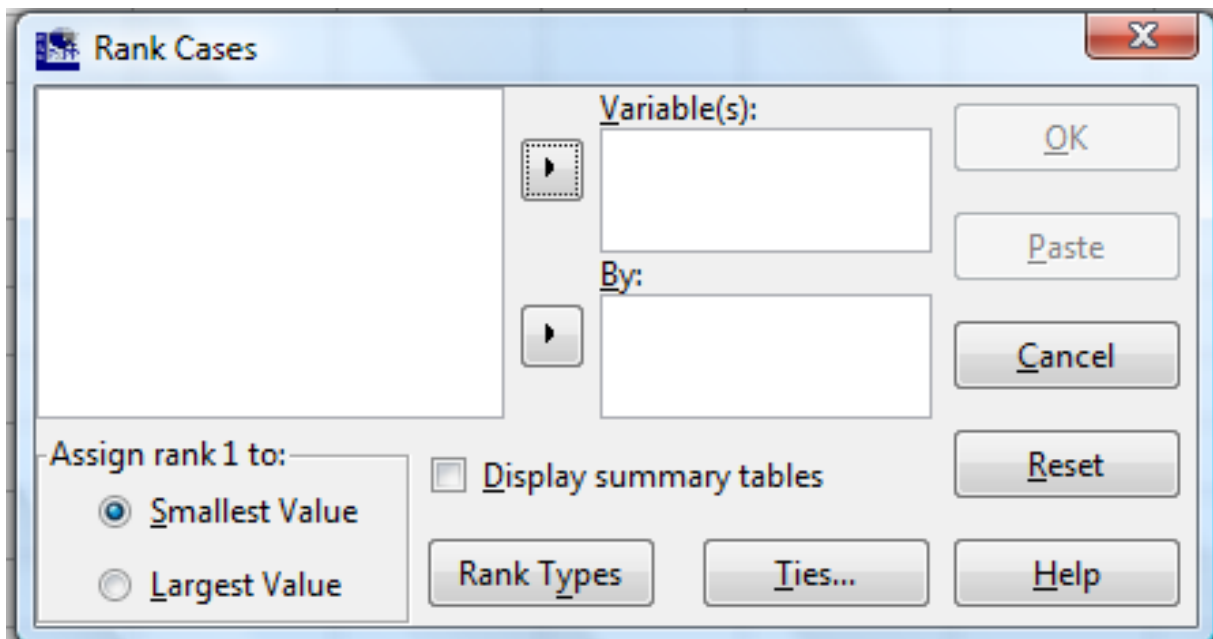
Tento príkaz spočíta počet prípadov výskytu konkrétnej hodnoty. Klikneme na Transformovať (Transform) a potom na Počet (Count). Nová premenná (Target Variable) bude vyjadrovať počet hodnôt výskytu konkrétnej hodnoty, ktorej môžeme priradiť jej Opis (Target Label). Do poľa Číselné premenné (Numeric Variables) prenesieme tie premenné, u ktorých chceme výskyt hodnôt sledovať.



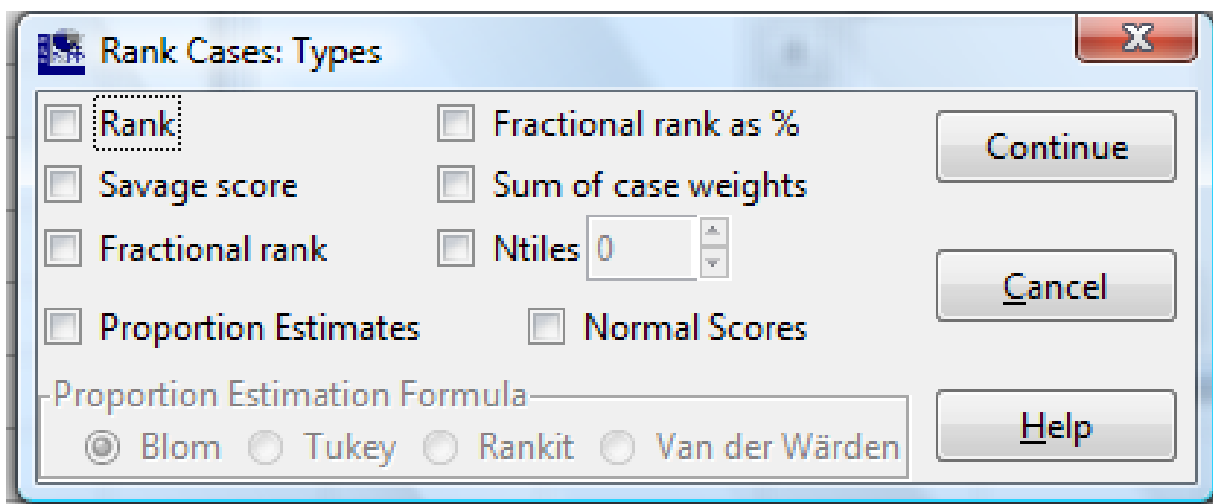
Obr. 4.4.3 Príkaz Počet (Count)

Príkaz Zoradiť prípady (Rank Cases)

Tento príkaz zoradí prípady podľa stanovených kritérií. Klikneme na Transformovať (Transform) a potom na Zoradiť prípady (Rank Cases). Príkaz Zoradiť prípady obsahuje v ľavej časti okna všetky premenné súboru. Po výbere tej, ktorú chceme zoradiť ju presunieme po poľa Premenná/é (Variable(s)). Ak chceme radieť prípady podľa dvoch premenných do podskupín, tak druhú premennú prevedieme do poľa Podľa (By). Priradiť poradové číslo 1 (Assign rank 1 to) môžeme dvoma spôsobmi, a to: Najmenšej hodnote (Smallest Value), alebo Najväčšej hodnote (Largest Value). Príkazové tlačidlo Typy Radenia (Rank Types) otvorí okno s ponukou spôsobov radenia dát (Obr. 4.4.4). Poradie (Rank), Savageho skóre (Savage score), Poradie v tvare zlomku (Fractional rank), Odhad proporcií (Proportion Estimates), Poradie v tvare zlomku ako percentá (Fractional rank as %), Suma vážených prípadov (Sum of case weights), kategorizácia podľa kvantilov (Ntiles) alebo Normálne skóre (Normal Score). Tlačítka (Ties) použijeme v prípade ak chceme priradiť poradie prípadom, ktoré majú rovnakú hodnotu.



Obr. 4.4.5 Zoradiť prípady (Rank Cases)

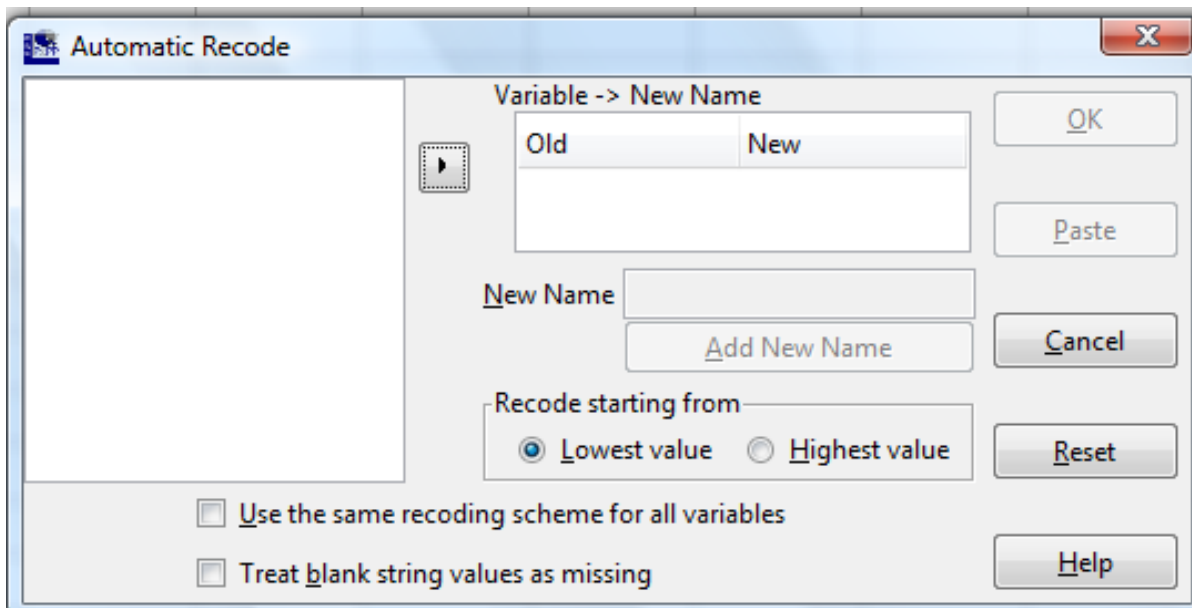


Obr. 4.4.6 Zoradiť prípady, Typ Rank Cases, Typ

Príkaz Automatické Rekódovanie (Automatic Recode)

Automaticky prekóduje hodnoty premennej na celé čísla a zoradí ich od 1 vyššie. Klikneme na *Transformovať (Transform)* a potom na *Automatické rekódovanie (Automatic Recode)*. V ľavej časti sa nachádzajú názvy všetkých premenných, z ktorých vyberieme tú, ktorú chceme prekódovať a presunieme ju do poľa *Premenná-> Nový názov (Variable -> New Name)*. V poli *Nový názov (New Name)* zadáme nový názov premennej a klikneme na *Pridať Nový Názov (Add New Name)*. Rekódovať môžeme začať od *Najnižšej hodnoty (Lowest value)* alebo od *Najvyššej hodnoty (Highest value)*. Dole máme uvedené dve zaškrtnuté políčka:

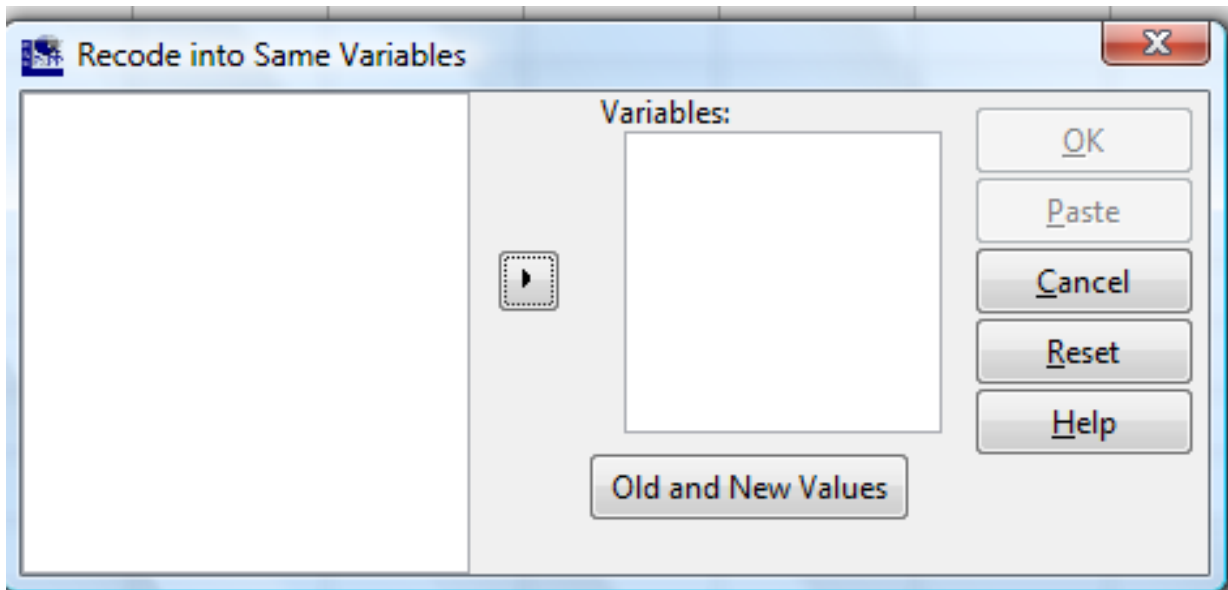
Použiť rovnakú rekódovaciu schému pre všetky premenné (Use the same recoding scheme for all variables) a Zaradiť prázdne hodnoty ako chýbajúce (Treat blank string values as missing).



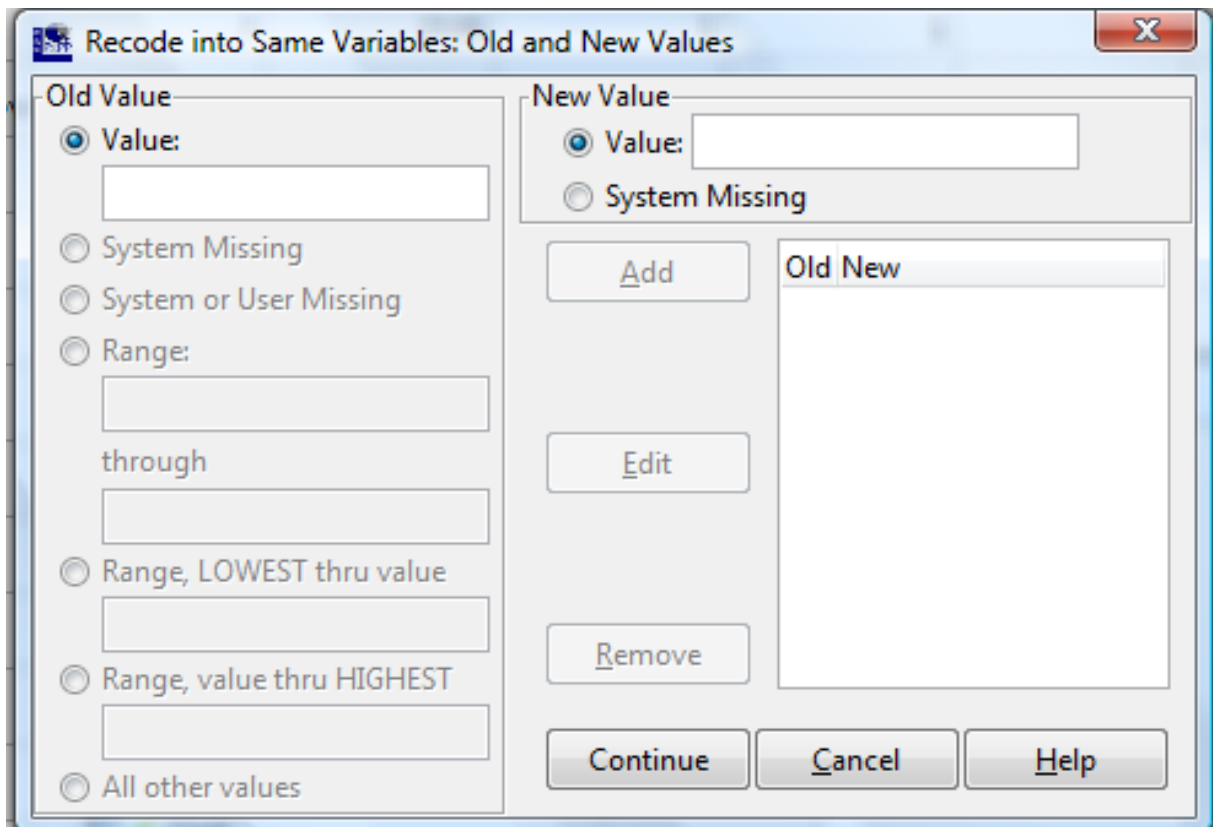
Obr. 4.4.7 Automatické Rekódovanie (Automatic Recode)

Príkaz Rekódovanie do tej istej premennej (Recode into Same Variables)

Príkaz zmení kódovanie premennej a prepíše ju. V prípade, že by sme chceli vytvoriť z rekódovanej premennej novú premennú tak použijeme nasledujúci príkaz. Po výbere premennej ju presunie do poľa *Premenné (Variables)*. Príkazové tlačidlo *Nové a Staré Hodnoty (Old and New Values)* slúži na spresnenie systému rekódovania. Užívateľ môže zadať zmeny postupne *Starú hodnotu (Old Value)* za *Novú (New Value)*, alebo zvoliť *Systémovo vynechané hodnoty (System Missing)*, zvoliť všetky hodnoty ako vynechané *Systémovo alebo Užívateľsky Vynechané (System or User Missing)*. Ďalej je možnosť vybrať hodnoty z *Intervalu (Range)*. V tomto prípade je potrebné zadať *Najnižšiu hodnotu v Intervale (Range, LOWEST thru value)* ako aj *Najvyššiu v Intervale (Range, thru value HIGHEST)*. Poslednou možnosťou je zmeniť všetky ostatné hodnoty, ktoré neboli zadané.



Obr. 4.4.8 Rekódovanie do tej istej premennej



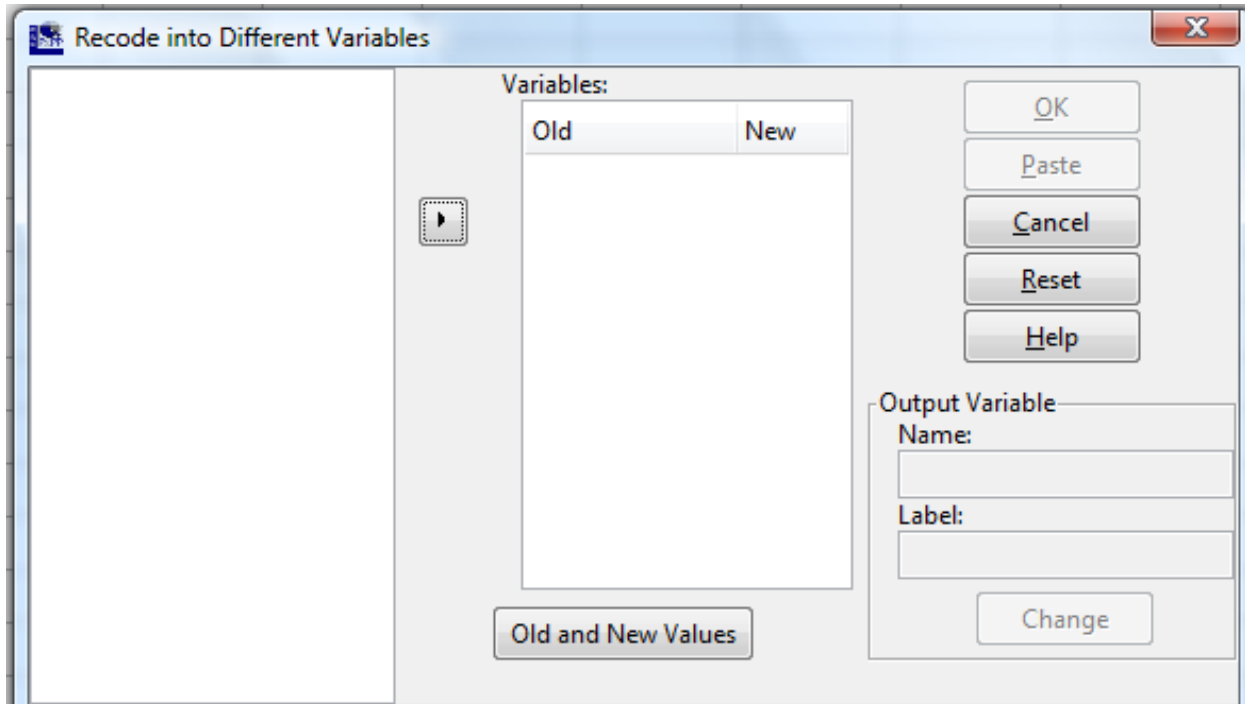
Obr. 4.4.9 Príkazové tlačidlo Old and New Values

Príkaz Rekódovanie do inej premennej (Recode into Different Variables)

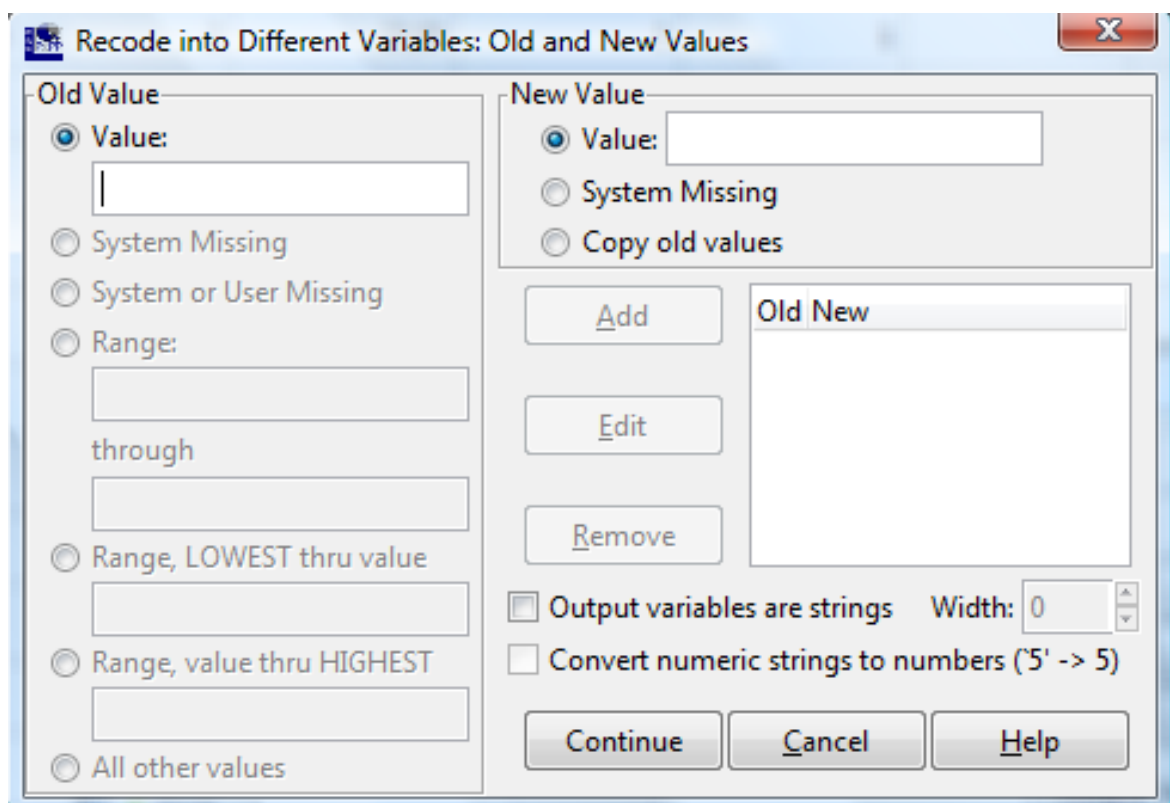
Tento príkaz rekóduje do novej premennej. Premennú, ktorú chceme rekódovať presunieme z ľavej časti okna do *Premenných (Variables)*. Na presný postup rekódovania klikneme na tlačidlo *Staré a Nové Hodnoty (Old and New Values)*. Zobrazí sa nám okno Obr. 4.4.10 veľmi podobné ako v predchádzajúcom príkaze

Obr. 4.4.11 s malými rozdielmi, ktoré opíšeme. Okno obsahuje zaškrtačacie tlačidlá. Ak chceme aby Výstupné premenné boli znaky, tak zaškrtneme (*Output variables are strings*) a ak chceme, aby boli čísla, tak zaškrtneme *Konvertuj numerické znaky na čísla* (*Convert numeric strings to numbers*).

Príkazové tlačidlo



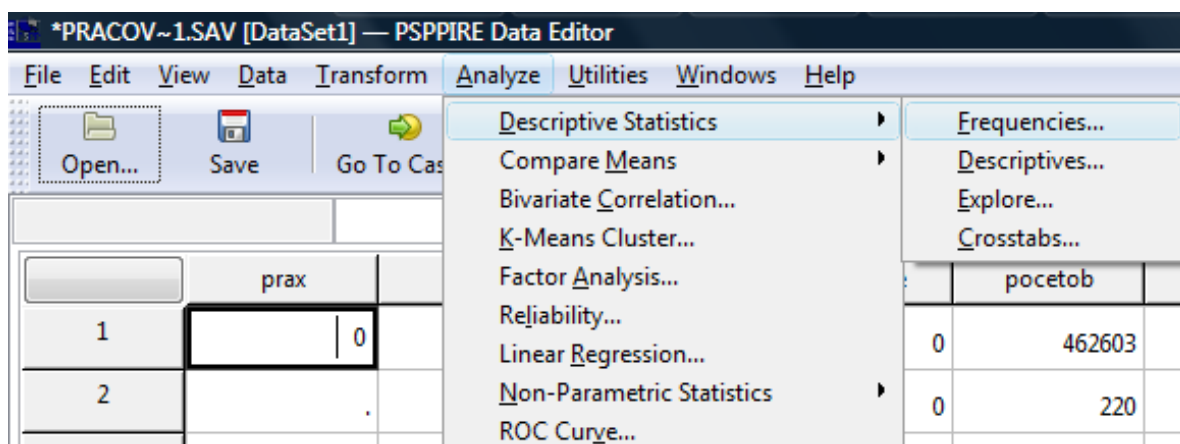
Obr. 4.4.12 Rekódovanie do inej premennej



Obr. 4.4.13 Rekódovanie do inej premennej, Staré a nové hodnoty (**Old and New Values**)

4.5 Ponuka Analýza (Analyze)

Ponuka Analyze je zameraná na štatistickú analýzu dát. Táto ponuka je najdôležitejšia z celého programu a tvorí jeho jadro. Každý z príkazov sa ďalej delí a my podrobne opíšeme každý z nich, no v tejto podkapitole ich predstavíme len sumárne. Podrobne si opíšeme jednotlivé štatistické metódy v ďalších kapitolách knihy, ktoré sú nazvané podľa konkrétnych štatistických metód. Prvou operáciou uvedenou v ponuke je *Deskriptívna štatistika (Descriptive Statistics)*, ktorá slúži na základný štatistický popis skúmaných dát.



Obr. 4.5.1 Ponuka Analyzovať (Analyze)

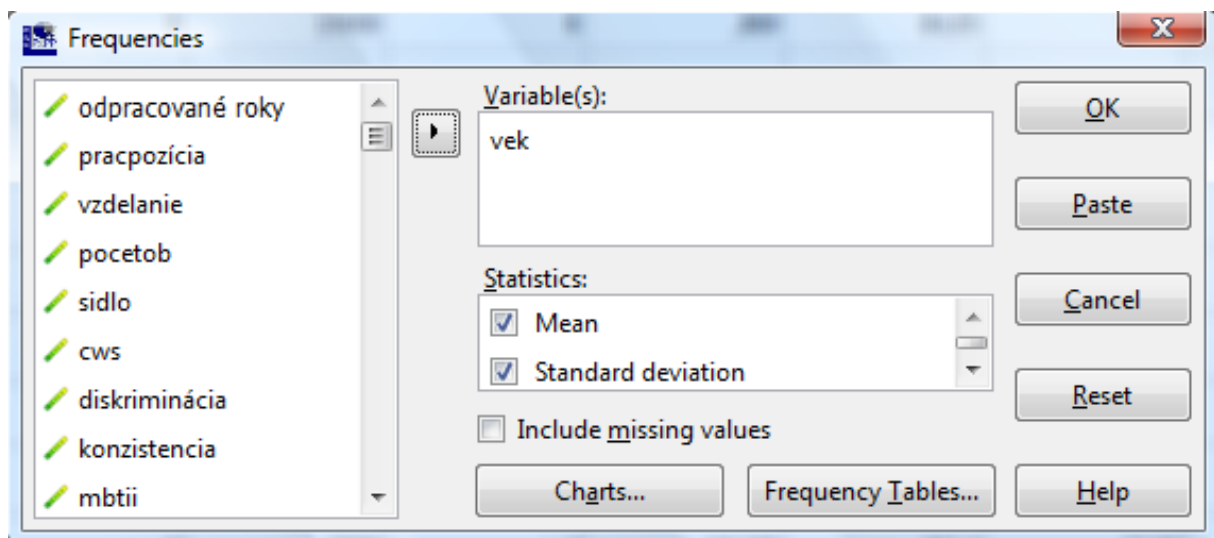
Opisná štatistika Descriptive Statistics

V rámci deskriptívnej štatistiky nájdeme *Frekvencie (Frequencies)*, *Deskriptívnu štatistiku (Descriptives)*, *Exploračnú analýzu (Explore)* a *Krížové tabuľky (Crosstabs)*. Aj pri zložitejších analýzach je vhodné výsledky interpretovať a uvádzať spolu s výsledkami deskriptívnej štatistiky. Opisnú štatistiku, konkrétne príklady, interpretáciu i teoretické vysvetlenia pomerne podrobne opisujeme v kapitole s názvom Opisné Štatistiky.

Početnosť, Frekvencie (Frequencies)

Príkaz *Početnosť (Frequencies)* vytvorí tabuľky početnosti vybranej premennej alebo premenných, ktoré môže kombinovať s deskriptívnou štatistikou a zvolenými grafmi. Frekvenčnú analýzu taktiež opisujeme v samostatnej kapitole s názvom Frekvenčná Analýza.

Okno príkazu obsahuje *Premennú/é (Variables (s))*, pre ktoré chceme početnosť vyjadriť, *Deskriptívnu štatistiku (Statistics)*. V spodnej časti dole sa nachádza možnosť zobrazenia dát *Grafmi (Charts)* a aj *Frekvenčnými Tabuľkami (Frequency Tables)*.

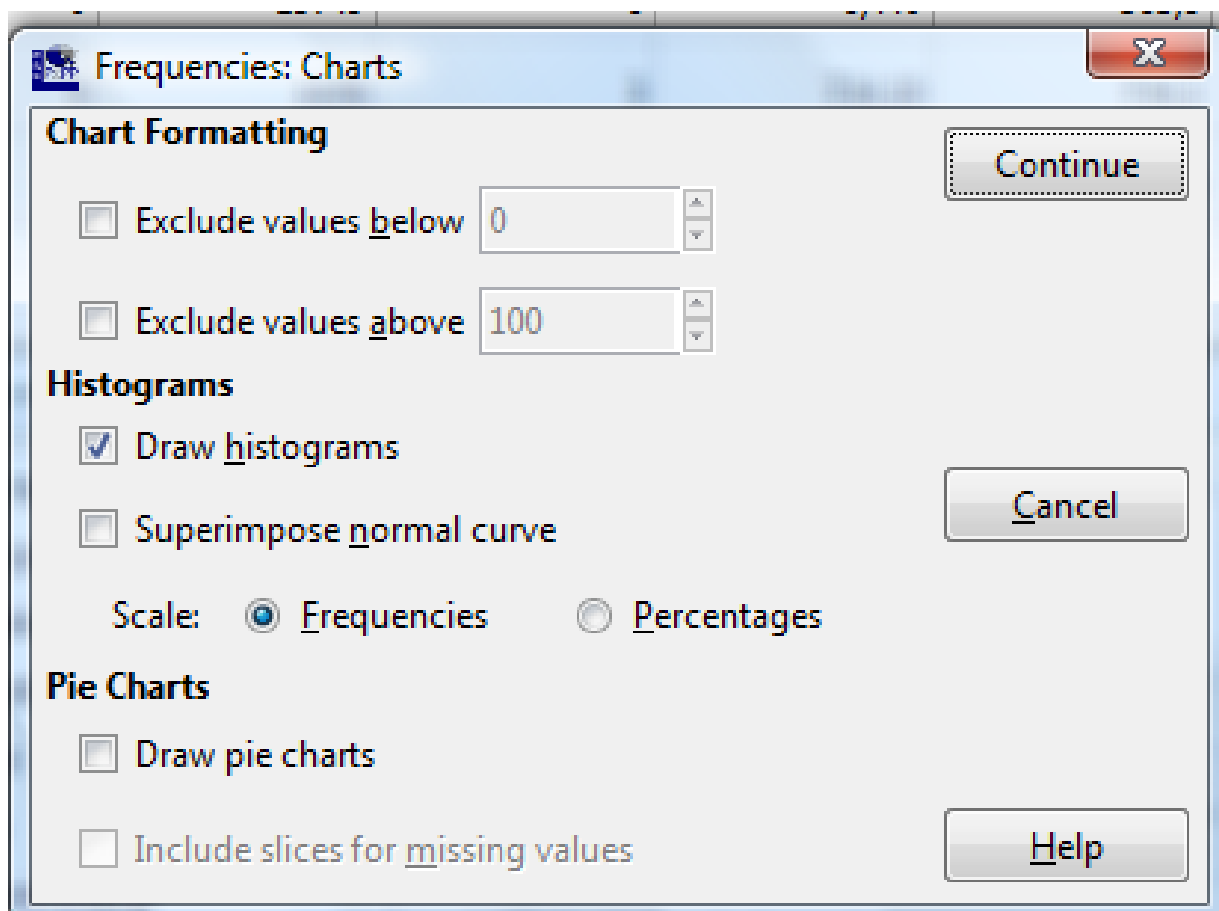


Obr. 4.5.2 Príkaz *Početnosť (Frequencies)*

Príkazové tlačidlo *Grafy (Charts)* otvorí nový dialógové okno, v ktorom sa nachádzajú tri oddelenia so zaškrávacími tlačidlami. Prvým je *Formátovanie grafu (Chart Formatting)*, ktoré obsahuje dve možnosti : *Vylúčiť hodnoty pod (Exclude values below)*, ktorá vylúči z grafu hodnoty menšie ako nami stanovené číslo. Druhou možnosťou je *Vylúčiť hodnoty nad (Exclude values above)*. Ďalším oddelením sú *Histogramy (Histograms)*, ktoré obsahujú dve zaškrávacie políčka *Vytvoriť histogramy (Draw histograms)* a *Prekryť krivkou normálneho rozdelenia (Superimpose normal curve)*. Škála na základe ktorej sa bude histogram tvoriť môže byť *Frekvenčná (Frequencies)* alebo *Percentuálna (Percentages)*. Posledným oddelením sú *Koláčové grafy (Pie charts)*, ktoré obsahujú

zaškrtnuté políčko *Vytvoríť koláčový graf (Draw pie charts)* a možnosť označiť chýbajúce hodnoty ako samostatný výsek v grafe (*Include slices for missing values*).

Deskriptívna štatistika (Statistics) obsahuje nasledovné funkcie: *Mean* – priemer, *Median* – medián, *Mode* – modus, *Sum* – súčet, *S.E. mean* - štandardná chyba priemeru (štandardná odchýlka výberového priemeru). *Std. deviation* - štandardná odchýlka, *Variance* – rozptyl, *Range* - variačné rozpätie, *Minimum* - minimálna hodnota, *Maximum* - maximálna hodnota, *Skewness* - koeficient šikmosti (charakterizuje symetriu) a jeho štandardná chyba, *Kurtosis* - koeficient špicatosti (charakteristika strmosti) a jeho štandardná chyba.



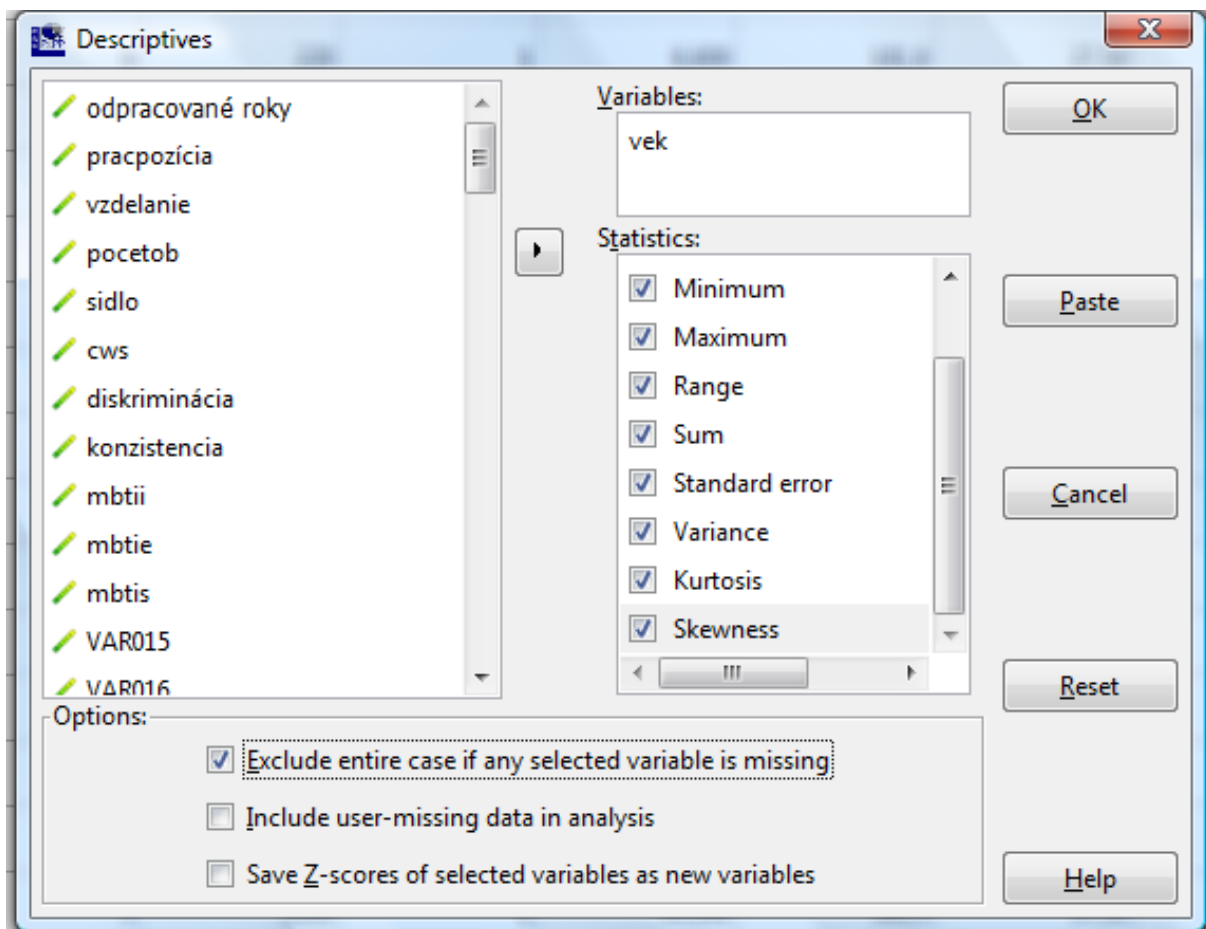
Obr. 4.5.3 Početnosť, tvorba Grafov

Deskriptívna štatistika (Descriptives)

Vypočíta deskriptívne štatistiky pre zvolenú premennú. Postup je nasledovný. V ľavej časti dialógového okna sa nachádzajú ako zvyčajne všetky premenné. Po výbere konkrétnej premennej, u ktorej chceme vypočítať deskriptívne štatistiky, túto presunieme do poľa *Premenné (Variables)* V pravej časti sa nachádza zoznam štatistických premenných, z ktorých sa bude deskriptívna štatistika počítať.

V spodnej časti dialógového okna sú uvedené *Možnosti (Options)*. Tie umožňujú nasledovné operácie s dátami. *Vylúčiť celý prípad, ak akákoľvek premenná chýba (Exclude case if any selected variable is missing)*. *Zahrň chýbajúce dáta do analýzy (Include user-missing data in analysis)* *Ulož Z skóre vybraných premenných ako nové premenné (Save Z-scores of selected variables as new variables)*.

Čo sa týka štatistických funkcií uvedených v okne *Štatistiky (Statistics)*, tie sú rovnaké ako v predchádzajúcom príkaze, a preto ich nebudeme dvojmo uvádzať.



Obr. 4.5.4 Sprievodca deskriptívnou štatistikou (Descriptives)

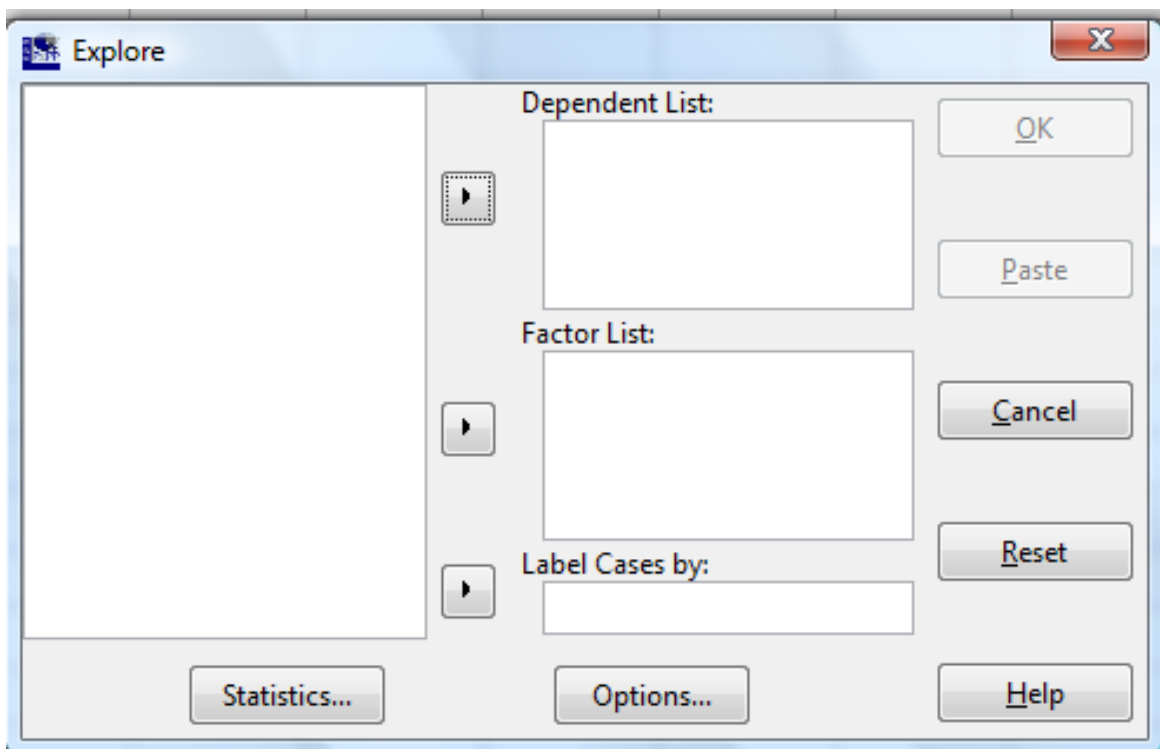
Pri otvorení sprievodcu pre deskriptívnu štatistiku postupujeme nasledovne. Klikneme na *Analyzovať (Analyze)*, potom na *Deskriptívnu štatistiku (Descriptive statistics)* a nakoniec na *Deskriptívu (Descriptives)*.

Reportovanie výsledkov deskriptívnej štatistiky. Vo vedeckých článkoch sa deskriptívna štatistika reportuje zvyčajne nasledovne. Uvádza sa priemer, po anglicky Mean, (nie Average ako by laický predkladateľ očakával) vo forme skratky, napr. $M = 2,17$. Za priemerom sa vždy udáva štandardná smerodajná odchýlka (standard deviaton), v skratke napr. $SD = 0,57$. Súčasne je potrebné uviesť počet prípadov, resp. respondentov vo forme $n = 78$.

Explorácia (Explore)

Príkaz vypočíta zvolené sumárne štatistiky a nakreslí grafy pre celý súbor alebo jeho časť. Túto časť analýzy znova podrobne opisujeme v kapitolách s názvom: Analýza vzájomných vzťahov u nominálnych premenných a Analýza vzájomných vzťahov u ordinálnych premenných.

Príkaz Explorácia ponúka rozsiahlu možnosť analýzy *Závislých premenných (Dependent List)* a ich grafického zobrazenia podľa vybraných faktorov, *Nezávislých premenných (Factor List)*. Tieto presúvame po ich výbere do ľavej časti okna.



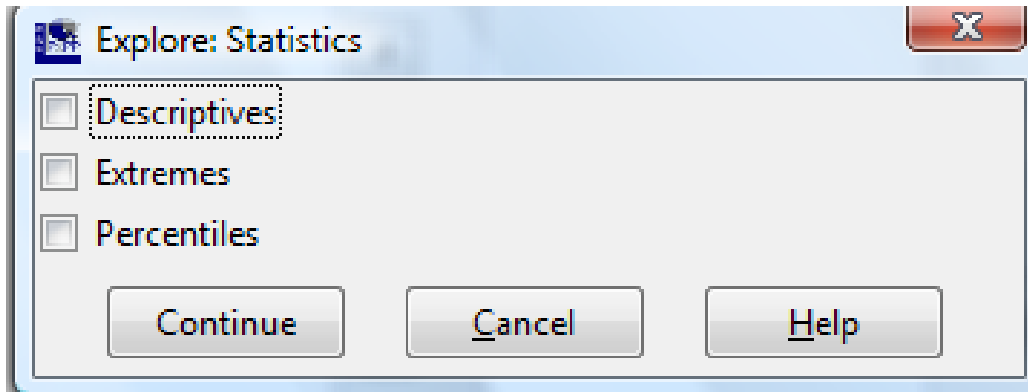
Obr. 4.5.5 Príkaz Explorácia (Explore)

V spodnej časti okna sa nachádzajú dve príkazové tlačidlá : *Štatistika (Statistics)* a *príkaz Možnosti (Options)*.

Príkaz *Opisné štatistiky (Descriptives)* obsahuje základné opisné štatistiky, teda: priemer s 95 % intervalom spoľahlivosti a ich spodnú a hornú hranicu intervalu, medián, 5% useknutý priemer (priemer po vylúčení 5% najmenších a 5% najväčších hodnôt), rozptyl, štandardná odchýlka, hodnota, maximálna hodnota, variačné a kvartilové rozpätie, koeficient šikmosti a koeficient strmosti.

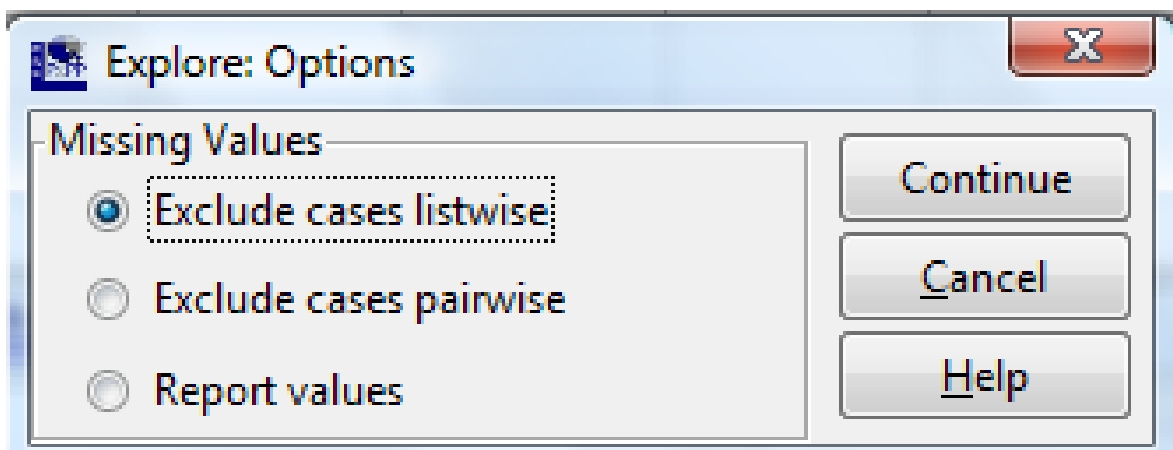
Zaškrtávacie tlačidlo *Extrémy (Extremes)* zobrazí päť najväčších hodnôt a päť najmenších hodnôt ako extrémny. Výstup je pre celý súbor ako aj samostatne pre jednotlivé faktory.

Zaškrtávacie tlačidlo *Percentily (Percentiles)* vypočíta metódou HAVERAGE (diskrétna interpolácia) vypočíta nasledujúce percentily: 5, 10, 25, 50, 75, 90 a 95.



Obr. 4.5.6 Príkaz Explorácia (Explore) príkaz Štatistika (Statistics)

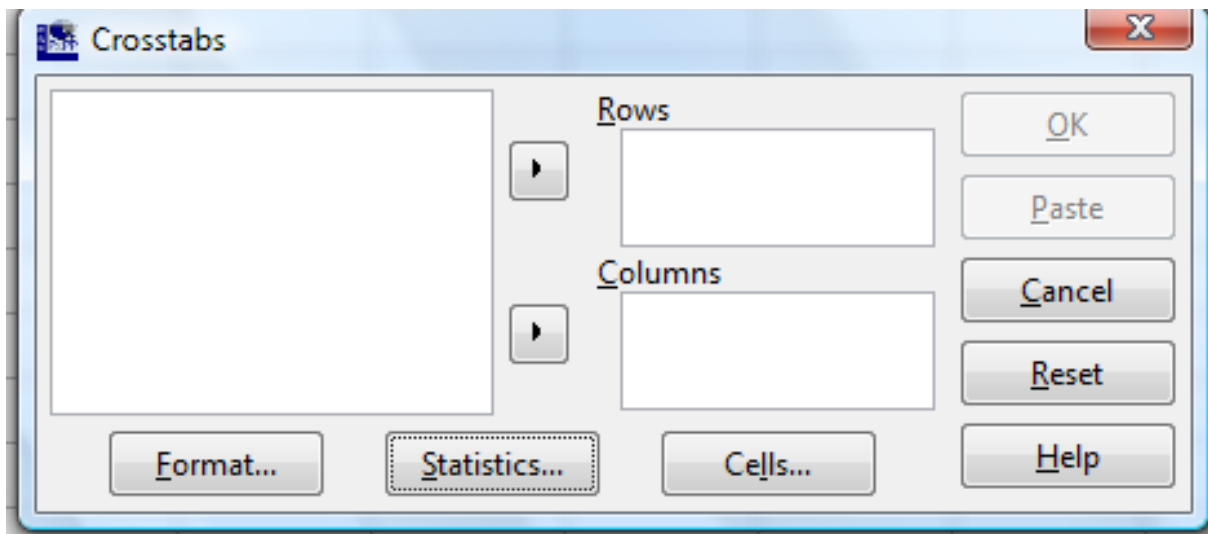
Príkaz *Možnosti (Options)* obsahuje tri príkazové tlačidlá: *Vylúčiť hodnoty pre celý prípad (Exclude cases listwise)* - ak aspoň jedna z analyzovaných premenných (závislých premenných alebo faktorov) má v určitom prípade (v riadku v editore dát) systémovú alebo užívateľskú chýbajúcu hodnotu, prípad je vylúčený z analýzy, či tvorby grafu a to aj u ostatných premenných. *Vylúčiť hodnoty po dvojiciach (Exclude cases pairwise)* pri analýze vynechá chýbajúcu hodnotu, no ostatné hodnoty použije. *Vypísať vylúčené hodnoty (Report values)* prípady s chýbajúcimi hodnotami závislých premenných sú vylúčené zo spracovania a vo výstupe je o nich podaná informácia; chýbajúce hodnoty faktorov (ako také) sú zahrnuté do spracovania.



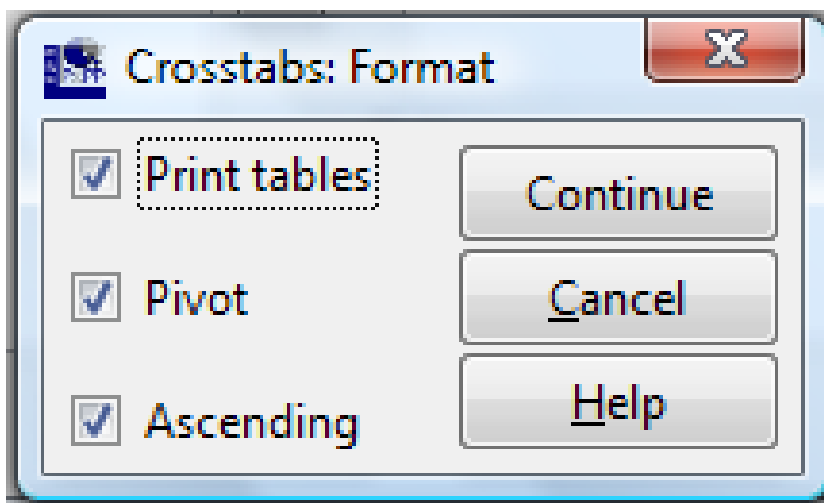
Obr. 4.5.7 Príkaz Explorácia (Explore) príkaz Možnosti (Options)

Kontingčné tabuľky (Crosstabs)

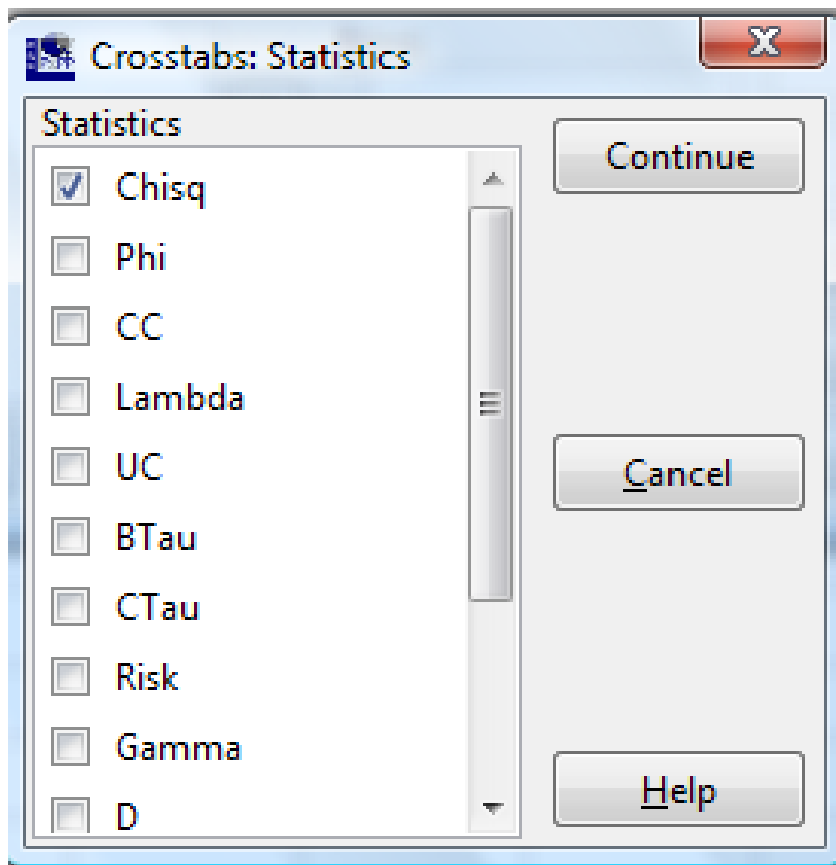
Tvorí tabuľky druho a viacstuňového triedenia s možnosťami bivariačných štatistík.



Obr. 4.5.8 Kontingčné tabuľky



Obr. 4.5.9 Kontingčné tabuľky, príkazové tlačidlo Formát (Format).



Obr. 4.5.10 Kontingenčné tabuľky, príkazové tlačidlo Štatistiky (Statistics).

Okno *Štatistiky (Statistics)* obsahuje dole uvedené štatistické metódy. Tieto podrobne opisujeme aj s príkladmi v kapitole s názvom *Skúmanie vzájomných vzťahov medzi premennými*. Preto ich v tejto časti len v krátkosti vymenujeme. *Chí kvadrát (Chisq)* testuje hypotézu o nezávislosti riadkovej a stĺpcovej premennej.

Phi (Phi) vypočíta koeficient *Phi*

Contingency coefficient (CC), *kontingenčný koeficient*.

Koeficient Lambda (Lambda)

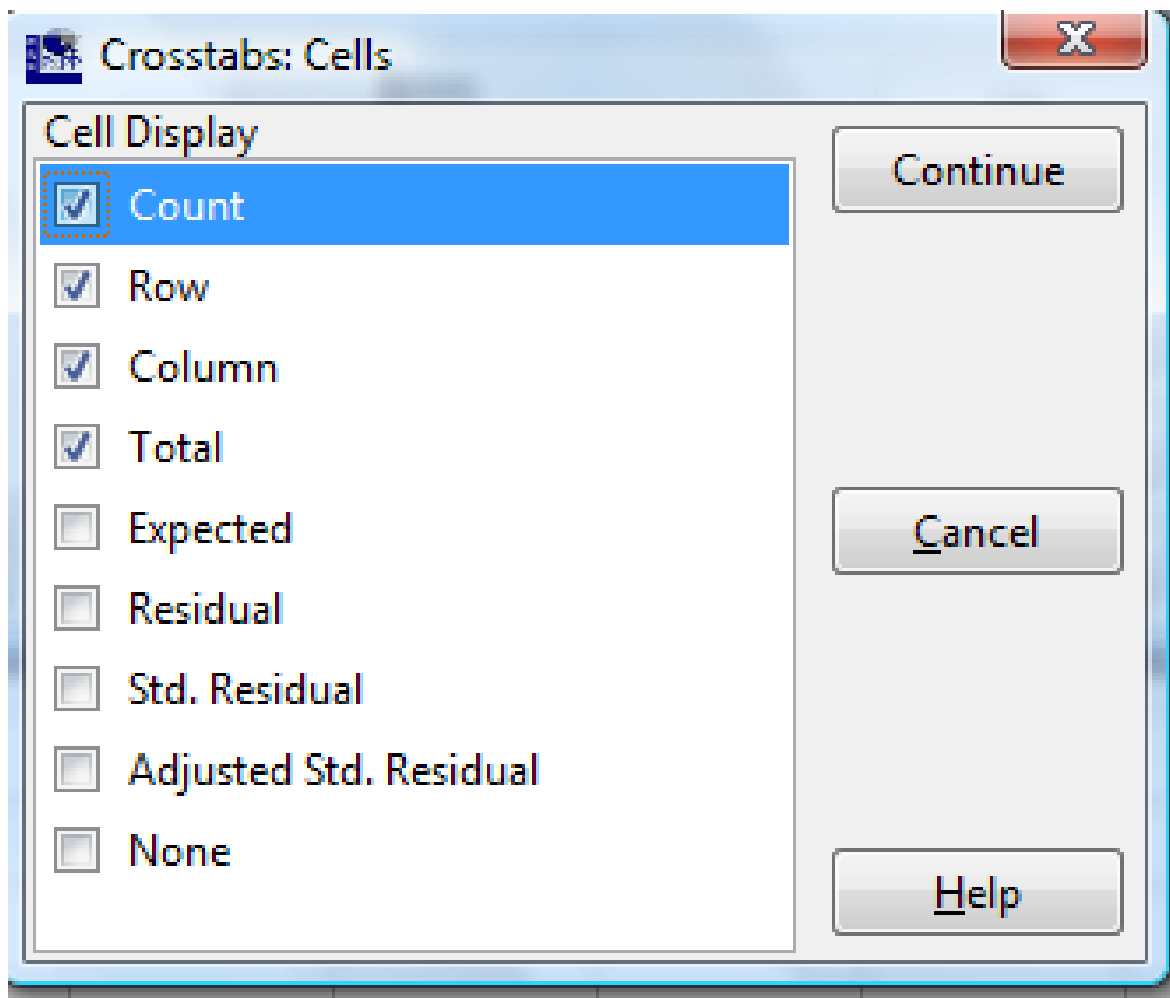
Koeficient neurčitosti (UC – Uncertainly coefficient)

Kendallove tau b (BTau) a kendallove tau c (CTau)

Risk, risk koeficient

Gamma, koeficient Gamma

Sommerovo d (D)



Obr. 4.5.11 Kontingenčné tabuľky, príkazové tlačidlo Bunky(Cells).

Pri krížovej tabuľke si môžeme zaškrtnúť formát zobrazenia pomocou stĺpcov, riadkov alebo ich kombinácií a môžeme zaškrtnúť aj dole uvedené možnosti zobrazenia dát.

Počet (Count)

Riadok (Row) zobrazí riadkové percentá

Stĺpec (Column) zobrazí stĺpcové percentá

Spolu (Total) vypočíta percentá spolu

Očakávané početnosti (Expected)

Reziduálne početnosti (Residual) Zobrazí reziduálne početnosti, teda rozdiel medzi pozorovanými a očakávanými početnosťami.

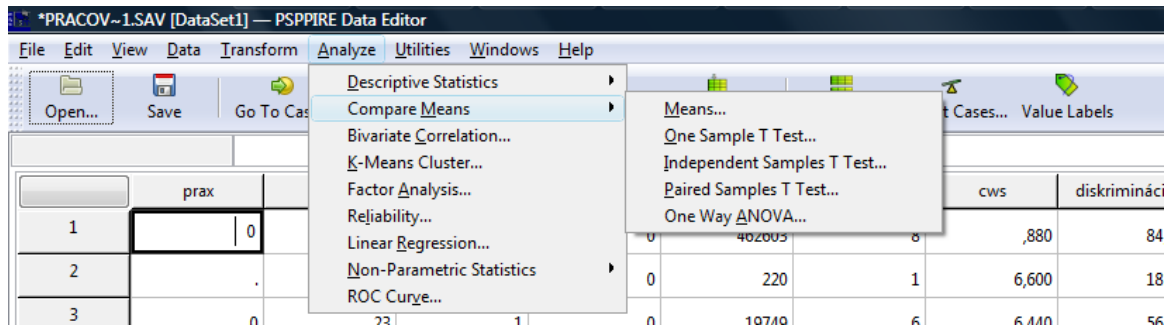
Štandardizované Reziduálne početnosti (Std. Residual)

Adjustované reziduálne početnosti (Adjusted Std. Residual)

Compare Means

Táto časť analýzy je zameraná na porovnávanie stredových hodnôt (priemerov) a obsahuje parametrické testy. Pri používaní paramterických testov musí byť

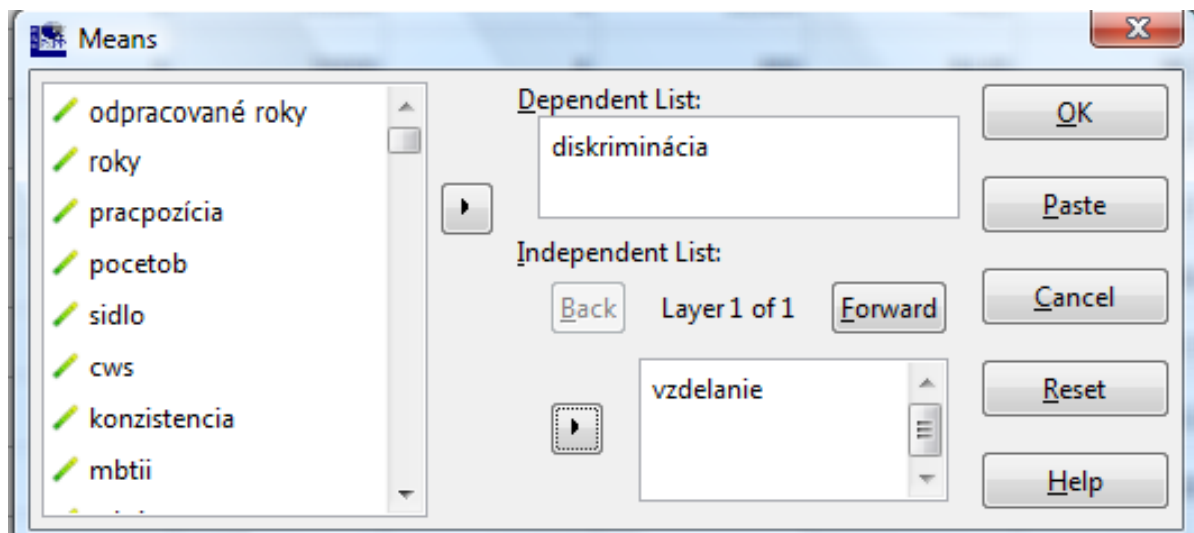
splnených viacero predpokladov, ktoré podrobne opisujeme v kapitole s názvom Parametrické testy a podmienky ich použitia. V tejto ponuke užívateľ nájde: T testy (pre porovnanie 2 skupín alebo skupiny s konštantou) a Analýzu rozptylu ANOVU (pre porovnanie viac ako 2 skupín). Každý z týchto testov podrobne opisujeme v rovnomenných kapitolách ďalej v tejto knihe. Preto ich nebudeme v tejto podkapitole opisovať a čitateľa odkážeme na podrobný opis i s príkladmi, ktorý nájde v nasledujúcich kapitolách pomenovaných podľa názvov jednotlivých štatistických metód.



Obr. 4.5.12 Analyzovať (Analyze), Porovnať priemery (Compare Means).

Priemery (Means)

Štatistická metóda na porovnanie priemerov u dvoch skupín. Spolu s priermi vypočíta štandardnú smerodajnú odchýlku a početnosti.



Obr. 4.5.13 Sprievodca analýzou Priemerov (Means)

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
diskriminácia: vzdelanie	176	61.9718%	108	38.0282%	284	100%
diskriminácia:	176	61.9718%	108	38.0282%	284	100%

Report

	vzdelanie	Mean	N	Std. Deviation
diskriminácia	0	434,33	157,00	286,29
	1	382,28	19,00	185,49

Report

	Mean	N	Std. Deviation
diskriminácia	428,71	176,00	277,25

Obr. 4.5.14 Výstup z analýzy Priemery (Means)

Univariačná analýza (Univariate analysis)

Slúži na skúmanie vplyvu jednotlivých faktorov na závislú premennú. V našej publikácii sa jej nebudeme venovať.

Korelačná analýza (Bivariate correlation)

Korelačná analýza skúma vzájomné vzťahy medzi premennými, ktoré vyjadruje číslom od 0 do 1, resp. od -1 do 0, ak je vzťah nepriamy. Softvér PSPP na rozdiel napríklad od MS Excelu zobrazí vo výstupnej časti aj hladinu významnosti, toho konkrétneho korelačného vzťahu. Korelačnú analýzu podrobne opisujeme v kapitole s názvom *Skúmanie vzájomných vzťahov u intervalových premenných*, podkapitola Korelácia.

Klastrová analýza (K – Means Cluster)

Klastrová analýza sa používa na zoskupovanie premenných, ktoré sú v rámci skupiny čo najpodobnejšie a súčasne sa líšia od ostatných skupín (Řehák, Brom, 2015). V tejto publikácii sa klastrovej analýze nebudeme venovať.

Faktorová analýza (Factor analysis)

Faktorovú analýzu podrobne opisujeme v kapitole s názvom *Faktorová analýza* a podkapitole *Exploračná faktorová analýza*.

Analýza Reliability (Reliability)

Analýza reliability slúži na otestovanie vnútornej konzistentnosti rôznych dotazníkov. Skúma odpovede všetkých respondentov na jednotlivé otázky. Respondenti by mali odpovedať konzistentne, resp. nenáhodne a systematicky. Softvér PSPP umožňuje merať reliability dvoma bežne používanými metódami : Split half a Cronbach alpha (označuje sa α). Výsledok merania obidvoch je číslo v rozpätí od 0 po 1 a platí veľmi jednoduché pravidlo, čím je α vyššia, tým je lepšie. Najkvalitnejšie dotazníky dosahujú skóre od 0,9 po 1. Vo vedeckej obci sa bežne tolerujú hodnoty od 0,7 vyššie, no pod túto hranicu by sme klesnúť už nemali. Metóda Split – half je založená na predpoklade, že tak ako respondent odpovedá na jednu polovicu otázok, potom v rovnakom duchu by mal odpovedať aj na druhú polovicu. Analýzu reliability podrobne rozpracúvame v kapitole s názvom *Analýza Reliability*.

Regresné analýzy (Regression)

Regresné analýzy aj s príkladmi sme spracovali v kapitole s názvami *Lineárna regresná analýza* a *Binárna logistická regresia*.

Neparametrické štatistiky (Non – Parametric Statistics)

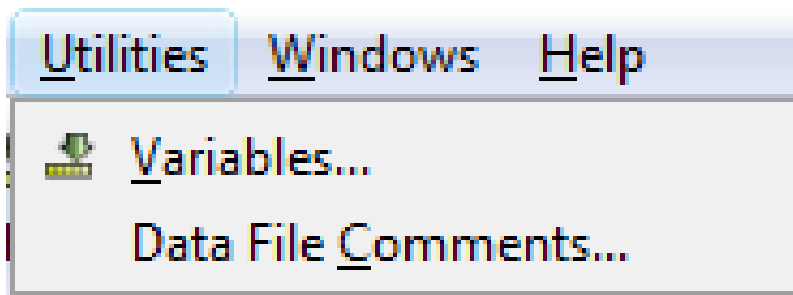
K neparametrickým štatistickým metódam patrí aj Wilcoxonov test, ktorý je ekvivalentom párového t testu. Wilcoxonov test sme podrobne spracovali v kapitole *Neparametrické testy*, v podkapitole *Wilcoxonov test*.

ROC Curve (Receiver Operating Characteristic - ROC)

Krivka, ktorá sa používa hodnotenie a optimalizáciu binárneho systému triedenia. Ukazuje vzťah medzi špecifickosťou a senzitivitou určitého testu. V tejto publikácii sa ROC krivke nebudeme venovať.

4.6 Ponuka Pomocné funkcie (Utilities)

Ponuka pomocné funkcie obsahuje príkazy *Premenné (Variables)*, ktorá zobrazí podrobné informácie o vybranej premennej. A subponuka *Komentár (Data File Comments)* umožňuje pridať vlastný komentár ku konkrétnemu dátovému súboru.



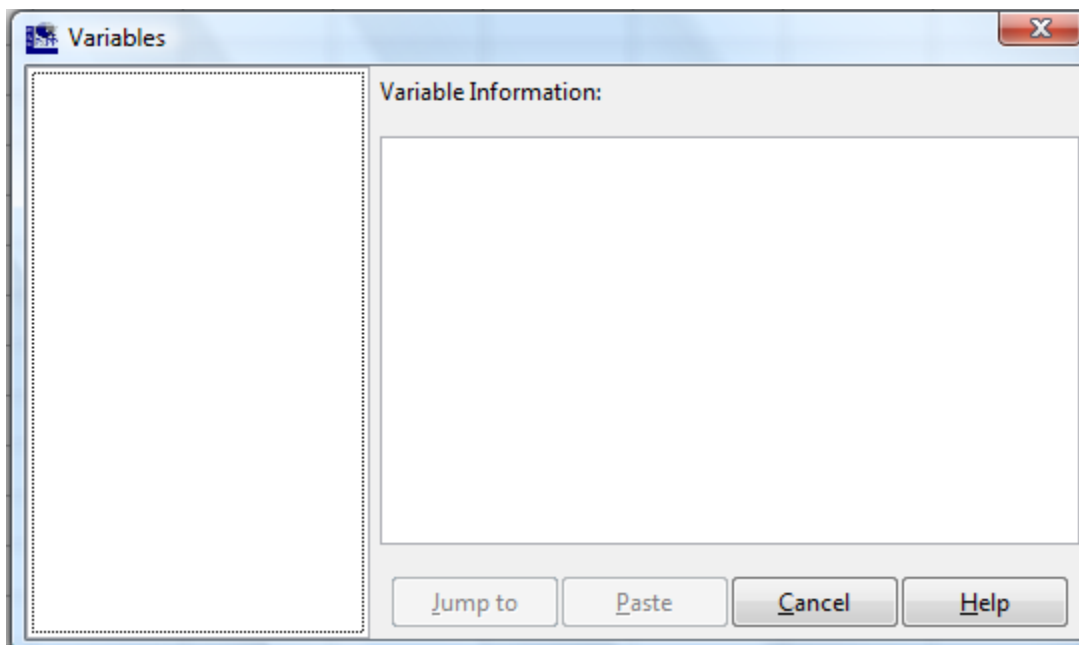
Obr. 4.6.1 Pomocné funkcie (Utilities)

Príkaz Premenné (Variables)

Zobrazí v ľavej časti okna zoznam všetkých premenných, ktoré sú aktuálnom dátovom dokumente. Klikneme na *Pomocné funkcie (Utilities)* a potom na *Premenné (Variables)*. V pravej časti okna sa nachádzajú *Informácie o premennej (Variable Information)*, ktoré sa aktivujú kliknutím na vybranú premennú a sú v tomto poradí: *Názov premennej, popis (Label), Typ (Type), Chýbajúce hodnoty (Missing Values), Typ premennej (Measurement Level) a Definované popisy hodnôt (Value Labels)*.

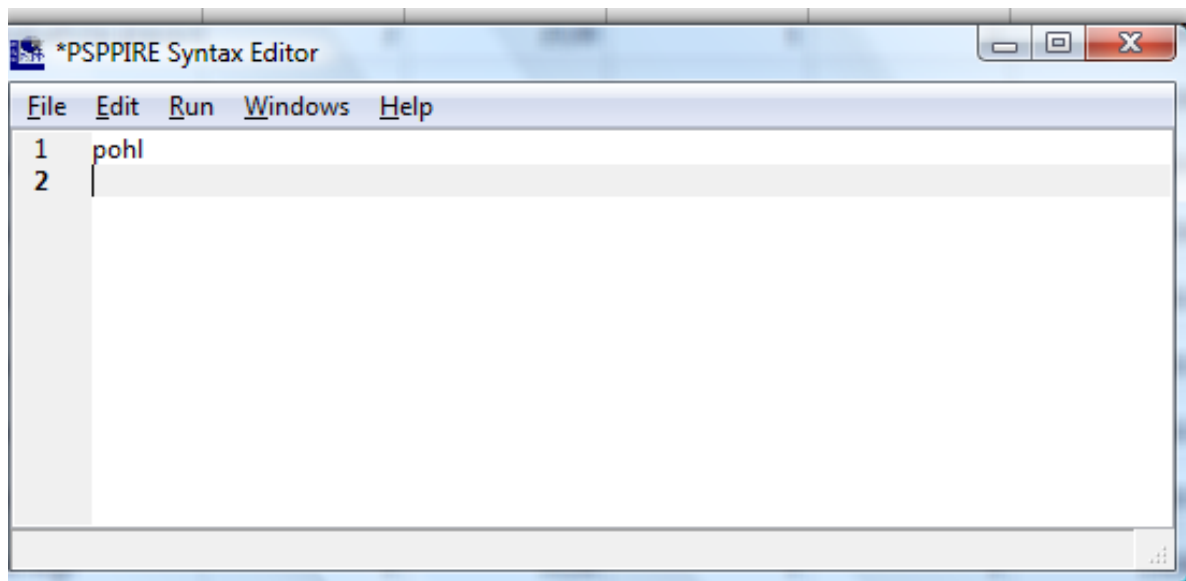
V spodnej časti okna je príkazové tlačidlo *Chod' na (Jump to)*, ktorý presunie kurzor v dátovom hárku na tú premennú, ktorá je práve označená v ľavej časti okna v *Premenných (Variables)* a súčasne sa okno s premennými zavrie.

Tlačítko *Prilepiť (Paste)* preniesie vybranú premennú do syntaxového okna.



Obr. 4.6.2 Príkaz Premenné (Variables)

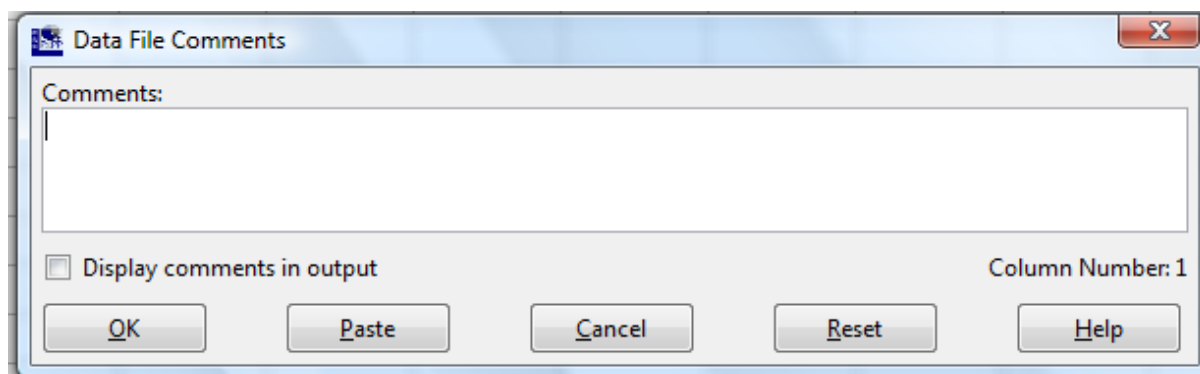
Tlačítko *Prilepiť (Paste)* preniesie vybranú premennú do syntaxového okna.



Obr. 4.6.3 Prenesenie vybranej premennej do syntaxového okna

Príkaz Komentáre (Data files comments)

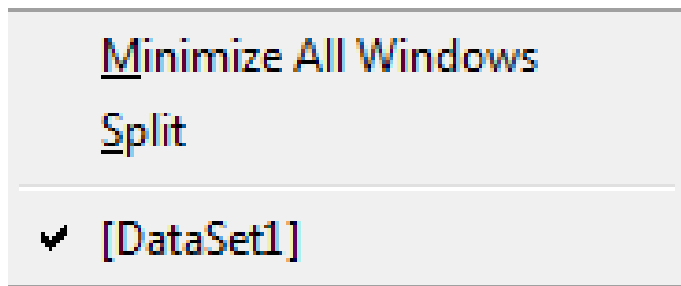
Tento príkaz pridáva komentáre k dátovému súboru. Zaštrtávacie políčko *Zobrazit' komentáre v outpute (Display comments in output)* po zaškrtnutí zobrazí komentár v outpute. Klikneme na *Pomocné funkcie (Utilities)* a potom *Komentáre k dátovému súboru (Data file comments)*.



Obr. 4.6.4 Príkaz Komentár (Data file comments)

4.7 Ponuka Okná (Windows)

Ponuka Okná obsahuje dva príkazy. Prvým je *Minimalizovanie všetkých okien (Minimize All Windows)*, ktorý minimalizuje všetky otvorené okná v programe PSPP. Druhý príkaz v tejto ponuke *Rozdeliť (Split)*, rozdelí súbor na štyri okná.



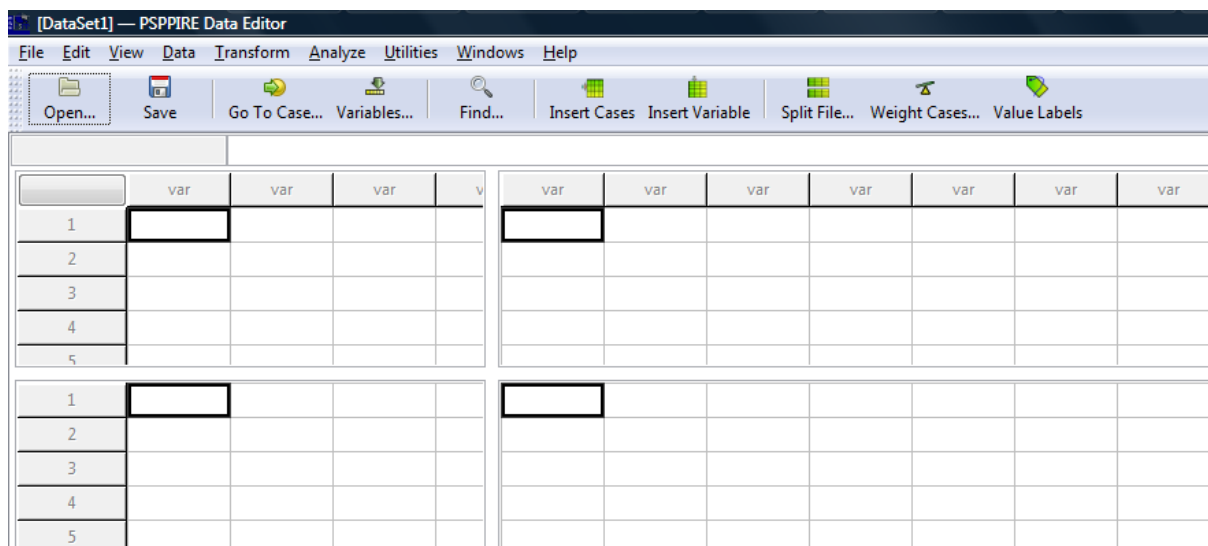
Obr. 4.7.1 Ponuka Okná (Windows)

Príkaz minimalizovať všetky okná (Minimize all windows)

Minimalizuje všetky otvorené okná. Klikneme na *Okná (Windows)* a potom *Minimalizovať všetky okná (Minimize all windows)*.

Príkaz Rozdeliť (Split)

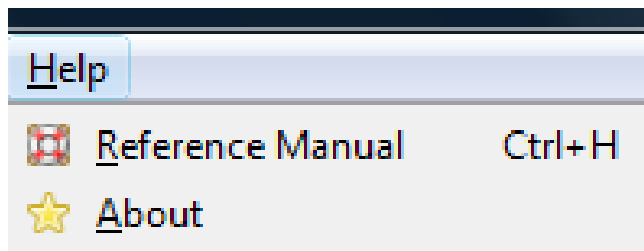
Príkaz rozdelí súbor na štyri okná a umožňuje listovať v nich samostatne. Rozdelenie súboru prebieha od označenej bunky vľavo hore. Klikneme na *Okná (Windows)* a potom na *Rozdelenie (Split)*.



Obr. 4.7.2 Rozdelenie súboru na štyri okná.

4.8 Ponuka Pomoc (Help)

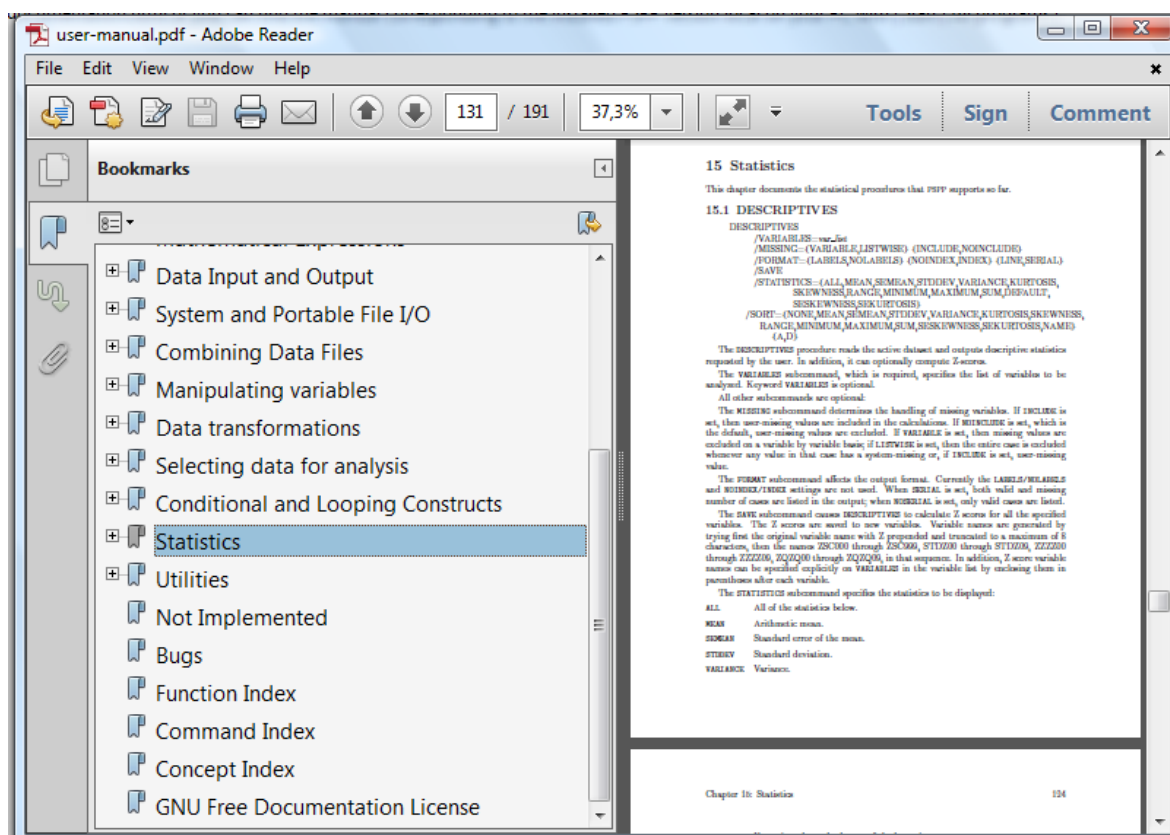
Ponuka *Pomoc (Help)* obsahuje príkazy *Manuál (Reference Manual)*, ktorý zobrazí manuál k programu PSPP a príkaz *O programe (About)*. Tento príkaz informuje o programe PSPP pomocou príkazových tlačidiel *Zásluhy (Credits)* a *Licencia (License)*.



Obr. 4.8.1 Ponuka Pomoc (Help)

Príkaz Manuál (Reference Manual)

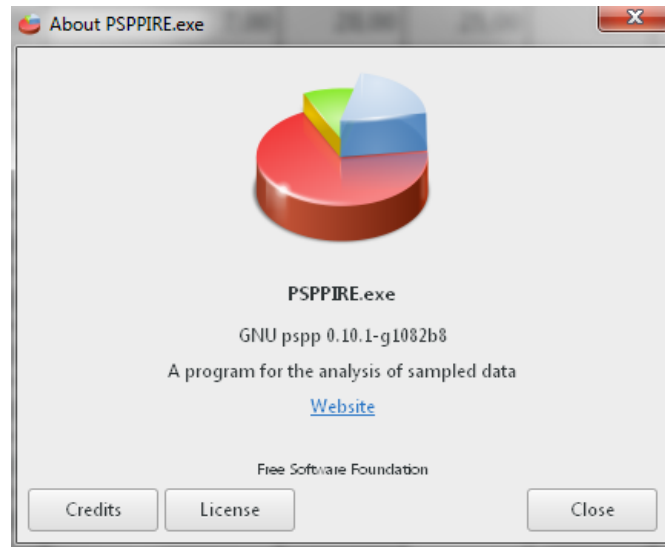
Otvorí manuál programu PSPP. Klikneme na *Pomoc (Help)* a *Manuál (Reference Manual)*.



Obr. 4.8.2 Príkaz Manuál (Reference Manual)

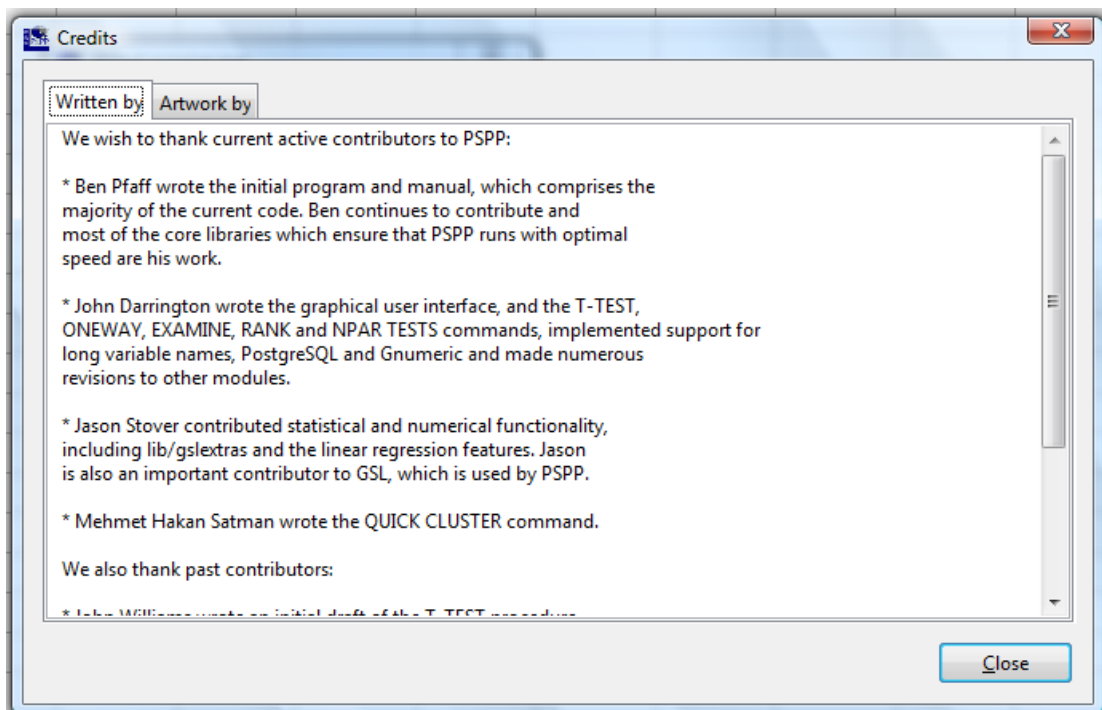
Príkaz O programe (About)

Otvorí informácie o programe PSPP



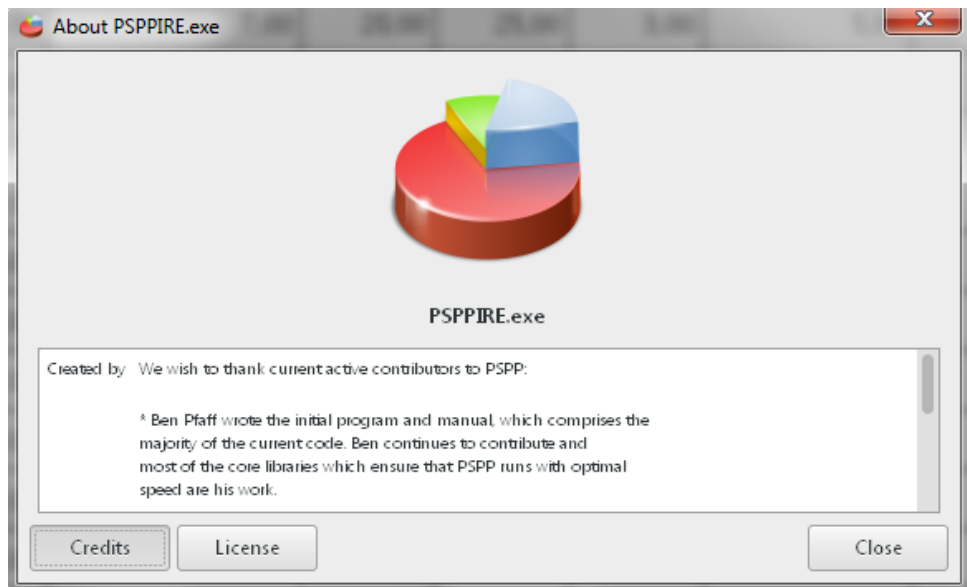
Obr. 4.8.3 Príkaz O programe (About)

Príkazové tlačidlo *Zásluhy (Credits)* otvorí dialógové okno so zásluhami programátorov a tvorcov programu PSPP.



Obr. 4.8.4 Obsah dialógového okna Zásluhy (Credits)

Príkazové tlačidlo *Licencia (License)* otvorí dialógové okno s licenčnými podmienkami programu PSPP.



Obr. 4.8.5 Obsah dialógového okna Licencia (License)

Príkazové tlačidlo *Zavrieť* (*Close*) zatvorí okno *O programe* (*About*).

5 VEDECKÝ VÝSKUM

V nasledujúcich podkapitolách v krátkosti vysvetlíme základné pojmy a postupy práce pri vedeckom výskume i pri analýze dát.

5.1 Populácia

Základná populácia je množina všetkých možných skúmaných objektov. Môžeme ju nazvať aj **Základný súbor**, ktorý Hindls a kolektív (2007, str. 15) definujú ako: „štatistický súbor všetkých jednotiek, ktorý je predmetom sledovania...“. Táto populácia je tvorená jednotlivými **štatistickými jednotkami (prvkami)**, ktorými môžu byť jednotliví ľudia, podniky, zvieratá, jednotky produkcie, skupiny, alebo čokoľvek čo skúmame. Napríklad populáciu v triede tvoria všetci žiaci, ktorí triedu navštevujú. Jej štatistickou jednotkou je jeden žiak. Populáciu školy tvoria všetci študenti školy a populáciu študentov v danom meste tvoria všetci študenti v danom meste. Skúmanie žiakov konkrétnej triedy je pomerne jednoduché, pretože ich nie je veľa, a počas vyučovania sú všetci dostupní na jednom mieste. No pri praktickom skúmaní je neraz populácia veľmi veľká (všetci voliči v SR, všetci žiaci, všetci muži), ťažko dostupná (personálni manažéri), nie ochotná poskytnúť dáta (konšpirátori) a podobne. V týchto prípadoch nie je možné skúmať všetkých jedincov v populácii. Preto sa pri skúmaní rôznych javov spoliehame iba na časť populácie tzv. výber. **Výber, respektíve výberový súbor** je: „podmnožinou základnej populácie“ (Hendl a kol., 2014, str. 18). A z výsledkov skúmania vo výberovej vzorke robíme úsudky o vlastnostiach celej populácie. Ak chceme poznať názory dospelých Slovákov ohľadom triedenia odpadu, nemusíme sa pýtať na názory každého jedného dospelého občana v SR. Stačí ak zhromaždíme dostatočne veľkú a reprezentatívnu vzorku a pri správne aplikovaných metódach zberu získame údaje. Hoci nebudeme mať odpovede všetkých dospelých Slovákov, ale len dostatočne veľký a reprezentatívny výberový súbor, naše údaje budú reprezentovať populáciu primerane dobre na to, aby sme mohli s určitou istotou tvrdiť, že sú veľmi podobné názorom celej populácie a mohli s z nich robiť závery. Vo výskume sa snažíme aby bol výber čo najväčší a obsahoval čo najvyšší počet jednotiek (prvkov). Počet jednotiek vo výbere sa nazýva **Rozsah výberu** a označuje sa n .

Náhodný výber znamená, že každá jednotka v populácii má rovnakú pravdepodobnosť byť vybratá do výberového súboru (Hendl a kol., 2014 str. 19). Pri skúmaní to znamená, že máme možnosť mať informácie o každej z jednotiek v populácii a my náhodným generátorom čísel vyberieme počet, ktorý potrebujeme. Hoci si študenti neraz myslia, že dotazníky poslali náhodne svojim

známym sú náhodným výberom, no v skutočnosti nebola dodržaná podmienka, že každý dospelý má rovnakú pravdepodobnosť byť členom prieskumu.

Zámerný výber v tomto prípade zámerne (cielené) vyberáme z populácie určité skupiny, ktoré majú špecifické vlastnosti a tie chceme skúmať. Napr. manažéri s viac než tridsať ročnou skúsenosťou, s ročným príjmom nad 50 000 Eur a s vysokoškolským vzdelaním. **Hromadný výber** je vtedy ak náhodne nevyberieme len jednotlivca, ale celé súbory- skupiny jednotlivcov, napríklad celé školské triedy, obyvateľov v celých bytovkách a podobne (Juszczyk, 2003, str. 81)

Pri skúmaní len podmnožiny základnej populácie, t. j. výberu si musíme byť vedomí toho, že naše výsledky môžu (a v praxi aj jednoznačne sú) viac, alebo menej, odlišné od základnej populácie. Pri skúmaní sa teda dopúšťame určitých chýb.

Prvou chybou je **Výberová náhodná chyba (random error)**, ktorá označuje rozdiel medzi výberovou štatistikou (naším výberom) a parametrom populácie. Táto chyba bola spôsobená viacerými náhodami pri zhromažďovaní výberovej vzorky. U náhodnej chyby predpokladáme, že odchýlky od parametra populácie sú aj v pozitívnom aj v negatívnom smere. Nie sú systematické a hoci znemožňujú presnejšie určiť parameter v populácii, vzájomne sa neutralizujú.

Druhou chybou, ktorá sa vyskytuje je **Chyba systematická (systematic error)**. Táto chyba systematicky skresľuje hodnotu výberovej štatistiky voči parametru populácie určitým smerom (pozitívne alebo negatívne k hodnote parametra v populácii). Tu pozorujeme systematicky skreslené hodnoty, čiže dochádza k preferovaniu určitých hodnôt.

Pri skúmaní akéhokoľvek spoločenského javu, stavu alebo problému sa chybám samozrejme nevyhneme. Na porozumenie a následne minimalizovanie efektu chýb máme k dispozícii viacero metód a nástrojov, ktoré si v krátkosti opíšeme v nasledujúcej kapitole.

5.2 Správnosť merania

Hendl (2009) uvádza, že k hlavným konceptom kvality merania patrí reliabilita (spoľahlivosť), validita a objektivita. K správnosti merania môžeme ešte zaradiť senzitivnosť a špecifickosť merania. Všetky tieto pojmy si v krátkosti vysvetlíme.

Objektivita merania

Objektivita merania sa vzťahuje tomu, či sú výsledky merania nezávislé na výskumníkovi, ktorý realizuje meranie. Teda ako veľmi ovplyvňuje skúmateľ (vedome, či nevedome) skúmaného participanta, resp. objekt skúmania. Zvlášť v psychológii alebo sociológii môže vplyv výskumníka významne meniť výsledky. Preto sa snažíme vplyv subjektivity minimalizovať, čo môžeme dosiahnuť nasledovnými krokmi:

- Znížime počet výskumníkov ideálne na 1 osobu, ktorá prichádza do styku s participantmi. Tým sa minimalizuje vplyv rôznych prístupov, postupov a komunikácie, ktoré sú nevyhnutne spojené s viacerými výskumníkmi, ktorí robia rovnaký výskum.
- Participantom prezentujeme inštrukciu písomne. Vyhneme sa tak tomu, aby sme participantov ovplyvňovala svojím zovňajškom, náladou, spôsobom prezentácie, alebo zabudnutím časti inštrukcie.
- Nedovolíme participantom komunikovať spolu počas riešenia. Odstránime tak možnosť opisovania, napomáhania a vzájomného ovplyvňovania.

Na určovanie objektivity merania používame viacero metód (Hendl, 2009). Najbežnejšou je korelovanie výsledkov hodnotenia dvoch hodnotiteľov. Ďalej môžeme použiť Kappa koeficient vzájomnej zhody.

Validita merania

Halama (2011, str. 108) definuje validitu merania ako: „neprítomnosť konštantných chýb v nameraných hodnotách“. Mareš a kol. (2015, str. 38) definovali validitu merania ako: „Meranie je validné, ak meria to, čo myslíme (očakávame), že meria“. Existuje viacero druhov validity, s ktorými musíme pri výskume uvažovať. Pelikán (2011, str. 58) uvádza štyri typy validity: prediktívnu, súbežnú, obsahovú a konštruktovú. Podrobnejšie si ich opíšeme.

Prediktívna validita hovorí o tom, že keď nameriame určitú úroveň vlastnosti, tak sa táto bude prejavovať aj v reálnom správaní jednotlivca. Napríklad, ak v dotazníku zistíme, že daná osoba je extrovert, potom očakávame, že jednotlivec sa bude aj extrovertne správať. Teda bude mať široké spektrum priateľov, známych, bude ľahko nadväzovať kontakty s cudzími ľuďmi a podobne. Ak by sme v dotazníku zistili, že osoba je extrovertná, no v „skutočnom“ živote sa aktívne vyhýba ostatným ľuďom, potom náš dotazník má nízku prediktívnu validitu. Pelikán (2011, str. 58) opisuje prediktívnu validitu, ako: „kritérium mimo používanej výskumnej metódy, techniky“. Teda určité konkrétne správanie, prípadne výsledky správania.

Súbežná validita hovorí o tom, že náš dotazník merajúci napríklad extroverziu by mal súvisieť s inými dotazníkmi merajúcimi extroverziu. Teda ak respondent v už overenom a etablovanom dotazníku, prípadne viacerých dotazníkoch skóroval ako silne extrovertný, potom by mal aj v našom dotazníku skórovať podobne.

Konštruktová validita znamená, že ak : „určitá premenná, alebo ešte lepšie meraný konštrukt vykazuje empiricky také vzťahy s ďalšími premennými aké sú predpokladané teóriou“ Pelikán (2011, str. 62). Heretik a kol. (2005, str. 93) opisuje uplatnenie konštruktivej validity: „pri hodnotení prepojenia použitej metódy a teórie či modelu ako určitý jav funguje.“ Náš merací instrument (dotazník, úloha a pod.) bude mať dobrú konštruktovú validitu ak výsledky získané z jeho aplikácie sú v zhode s teóriou na ktorej bol postavený.

Konštruktová validita má dve zložky: **konvergentnú validitu** a **diskriminačnú validitu**. Pri konvergentnej validite predpokladáme, že dva nástroje ktoré majú merať rovnaký konštrukt budú vykazovať vysokú vzájomnú koreláciu. A naopak pri diskriminačnej validite očakávame, že dva nástroje merajúce rôzne konštrukty (napr. extroverziu a svedomitosť) budú mať nulovú, prípadne nízku vzájomnú koreláciu.

Reliabilita merania

Halama (2011, str. 90) definuje reliabilitu ako: „vlastnosť testu, ktorá hovorí o miere náhodných, premenlivých chýb“. Reliabilitu definuje Mareš a kol. (2015, str. 39) ako: „rozsah, v ktorom spôsob merania dáva konzistentné výsledky“. Čo sa týka reliability merania, pre podrobnejší popis a spôsob výpočtu si treba pozrieť kapitolu Analýza Reliability. V krátkosti si len uvedieme, že existuje viacero postupov metód na overovanie reliability. Prvým je opakovanie merania (test retest), čo je metóda na overenie si ako je test stabilný v čase, preto sa niekedy nazýva aj koeficient stability. Na meranie sa používa Pearsonov korelačný koeficient (Hendl, 2009). Druhým postupom je použitie iného testu, ktorý meria rovnaký konštrukt. Štatistické metódy na overovanie reliability merania u dotazníkov sú: rozdeľovanie testu (split-half) a Cronbachova alfa

Senzitivita merania znamená, že získame pozitívnu odpoveď ak je jav u osoby skutočne zastúpený. A naopak **špecifickosť merania** je schopnosť testu dať negatívnu odpoveď, ak jav u osoby nie je zastúpený (Mareš a kol., 2015)

5.3 Premenné a ich typy

Premenné sú charakteristiky, znaky štatistických jednotiek (prvkov) základného súboru, ktoré môžu nadobúdať viacero hodnôt (Hendl a kol., 2014). **Premenná** môže byť opísaná aj ako vlastnosť, ktorá sa mení, premieňa, na rozdiel od konštanty, ktorá zostáva nemenná (Halama, 2011). Hindls a kolektív (2007, str. 14) definujú premenné ako : „vlastnosti štatistických jednotiek“. Premenné rozdeľujeme do dvoch skupín a to: **Kvalitatívna premenná** (nominálna a ordinálna), **Kvantitatívna premenná** (intervalová a pomerová) (Terek, 2013, str.16). Kvalitatívne premenné môžeme vyjadriť slovami napríklad: muž –žena, bratislavský kraj, trnavský kraj atď. Kvantitatívne vyjadrujeme číslami napríklad 238 Eur, 37 stupňov Celzia, 45 sekúnd atď.

Premenné môžeme merať na 4 typoch škál, stupníc:

- 1. Nominálna premenná (Nominal):** pozostáva z kategórií z ktorých vyberáme. Je to najnižšia úroveň triedenia (Ritomský, 1999). Je opísaná kódmi, ktoré boli umelo (arbitrárne) vytvorené a potom boli pridelené jednotlivým kategóriám. Napríklad Bratislavský kraj = 0, Trnavský = 1, Nitriansky = 2, atď. Keďže kód pre premenné určujeme my, potom si premenné môžeme

nakódovať ľubovoľne a tak Bratislavský kraj môže byť aj 25, Trnavský môže byť aj 10, alebo akékoľvek iné číslo okrem už použitých. Ak táto premenná má len dve možnosti napríklad : muž – žena, áno - nie, potom hovoríme o dichotomickej nominálnej premennej, ak má viac kategórií (bratislavský kraj, nitriansky atď..) potom hovoríme o polytomickej nominálnej premennej. Nominálna premenná je premenná, ktorá prináša najmenej informácií, a možnosti budúcej analýzy, ak máme len nominálne premenné sú veľmi obmedzené. Dotazník v ktorom sa napríklad pýtame na názory ohľadom súčasnej vlády a formulovanú otázku: „Súhlasíte s politikou vlády?“ a dáva možnosti odpovedí len áno a nie, núti respondentov vybrať si len jednu z dvoch možností. Pritom len minimum z nich by svoju spokojnosť opísalo len slovom áno, respektíve nie. Softvér PSPP označuje túto premennú *Nominal*.

2. **Ordinálna premenná (Ordinal):** zoraďuje (order) premennú do kategórií. Príkladom môže byť škála : („Veľmi sa mi páči, páči sa mi, páči sa mi trochu, trochu sa mi nepáči, nepáči sa mi, veľmi sa mi nepáči“ atď.). Pri ordinálnej premennej však vieme určiť, ktoré tvrdenie je nad druhým, je silnejšie, je viac. Konkrétne čo sa týka obľuby tak: „veľmi sa mi páči“, je viac než: „páči sa mi“, a to je viac než: „páči sa mi trochu.“ Avšak neplatí, že medzi kategóriami sú rovnaké vzťahy. Keďže ide o poradovú premennú, tak v tomto prípade nemôžeme exaktne povedať že tvrdenie: „Veľmi sa mi páči“, je dvakrát silnejšie než: „páči sa mi“. Pre niektorých respondentov to môže byť dvakrát silnejšie tvrdenie, no pre iných aj desaťkrát silnejšie tvrdenie. Jediné čo môžeme konštatovať, že „Veľmi sa mi páči“ je viac (je silnejšie) než „Páči sa mi“. Príkladom môže byť aj poradie časov riešenia, kde sme si nezaznamenali presný čas, ale len prvý, druhý, tretí atď. Softvér PSPP označuje túto premennú *Ordinal*.

Kardinálne premenné sú premenné, kde ich číselné hodnoty vyjadrujú skutočný stav skúmanej vlastnosti a nie len umelo(arbitrárne) vytvorený kód. Napríklad počet absolvovaných rokov dochádzky (skutočný počet rokov strávený v školách) oproti bratislavský kraj (kód 0). Kardinálne premenné podľa škál delíme na intervalové a pomerové. Softvér PSPP označuje tieto premenné pojmom *Scale*.

3. **Intervalová premenná (Interval):** hodnoty premennej sa nachádzajú v stanovenom intervale, ale v intervale sa nenachádza absolútna nula 0. Vzťahy medzi nimi sú presné, teda vieme presne definovať veľkosť rozdielu: napr. 4 je dvakrát toľko ako 2 a polovica z 8. Teda medzi postupne rastúcimi jednotkami je presne rovnaká vzdialenosť, medzi 1 a 2 je rozdiel rovnaký ako medzi 2 a 3. Ako už bolo vysvetlené hore toto neplatí o ordinálnej škále. Obmedzením intervalovej premennej, že nemáme nulovú hodnotu, sme len niekde v intervale (napr. od 40 do 80) a tak nevieme určiť o koľko je niektorá kategória silnejšia než iná. Konkrétne keď nevieme či začíname pri čísle 10,

alebo -50, alebo 35, preto nemôžeme tvrdiť, že 80 je dvakrát silnejšie než 40, pretože toto platí len ak začíname na 0, ale nie ak začíname na 10, -50, alebo 35. Pri intervalovej premennej už máme k dispozícii rozsiahly štatistický aparát na rozdiel od nominálnej, či ordinálnej premennej.

4. **Pomerová premenná (Ratio)** je rovnaká ako intervalová, ale navyše obsahuje aj 0. Je to najpresnejší spôsob merania. Môžeme sem zaradiť premenné ako vek, počet rokov návštevy škôl, počet bodov v teste, výška investície, počet spomenutých slov.

Intervalové alebo pomerové premenné môžeme ešte vyjadrovať dvomi spôsobmi. Buď máme určitú presne definovanú stupnicu hodnôt, ktoré premenná nadobúda ako napríklad odpracovaný čas v zamestnaní, kde máme hodnoty 0,1,2 roky atď. Táto sa nazýva **Diskrétna premenná (discrete)**. Diskrétna premenná nadobúdajú určitý konečný (spočítateľný) počet hodnôt. Ale čas odpracovaný čas v zamestnaní by sme mohli merať oveľa podrobnejšie a jemnejšie. Konkrétne aj na mesiace, týždne, dni, či dokonca na hodiny. Táto sa nazýva **Spojité premenná (continuos)**. Takže odpracovaný čas v zamestnaní môže byť 12 rokov, čo je diskrétna premenná, ktorá nadobúda hodnoty od 0 cez 1,2,3 ... až napríklad do 50. Ale odpracovaný čas môže byť aj 144,521 mesiacov (12 rokov * 12 mesiacov + 0,521 mesiaca) a to je už spojitá premenná. Z praktického hľadiska však zvyčajne nepracujeme so spojitými premennými a v rámci zjednodušenia si z nich robíme diskrétna premenné, teda napr. čas zaokrúhlime na 3 desatinné miesta, odpracovaný čas na roky, mesiace, počet bodov zo skúšky na celé čísla a podobne. Pri navrhovaní výskumu, respektíve zberu dát je veľmi dôležité vopred si dôkladne premyslieť ako by sme chceli a ako objektívne môžeme merať jednotlivé premenné. Je to dôležité v prvom rade preto, lebo rôzne typy škál umožňujú len konkrétne štatistické metódy, testy a postupy. Na niektoré premenné nemôžeme použiť niektoré štatistické metódy a tým naše možnosti analýzy a aj výsledkov sú obmedzené. Najmenej a „najhrubšie“ štatistické metódy vieme použiť na nominálne premenné. Ordinálne premenné už umožňujú viac i podrobnejších štatistických metód a najlepšie z hľadiska analýzy sú kardinálne premenné.

Pred akoukoľvek štatistickou analýzou je potrebné poznať s akými dátami pracujeme, či na ne môžeme aplikovať určité testy alebo nie. Rovnako dôležité je poznať premenné dokonca ešte pred samotnou analýzou, aby sme boli schopní dobre si nastaviť výskum a pripraviť zber dát. Ak nemáme dobre zadané a nastavené premenné ešte pred zberom a zozbierame dáta, potom sa bežne stáva, že ani najlepšia štatistická analýza nám nepomôže získať zmysluplné dáta. V rámci vlastného výskumu sa snažte o to, aby ste mali čo najväčší počet intervalových, resp. pomerových premenných, pretože na tieto premenné sa viažu najcitlivejšie štatistické metódy a sú analyzovateľné všetkými dostupnými štatistickými metódami, nástrojmi a testami.

Závislé a nezávislé premenné

Ak medzi premennými predpokladáme vzájomný príčinný vzťah, potom môžeme hovoriť o závislej (závisí od nezávislej premennej) a nezávislej premennej (priamo ovplyvňuje závislú premennú). Napríklad u inteligentného žiaka očakávame že počet bodov zo skúšky (závislá premenná) je ovplyvnený, závisí od počtu hodín prípravy (nezávislá premenná). Pri experimentálnych výskumoch výskumník manipuluje nezávislou premenou aby zistil ako tá potom ovplyvňuje závislú premennú.

Nezávislá premenná (Independent variable - IV) sa v literatúre označuje aj ako prediktívna premenná, prediktor alebo vysvetľujúca premenná. **Závislá premenná** (Dependent variable - DV) sa niekedy označuje aj odpoveďová či cieľová.

Vo výskume si musíme byť vedomí toho, že väčšinou závislá premenná je ovplyvnená viacerými nezávislými premennými, nielen jednou. Súčasne je bežné že závislá premenná je ovplyvnená rušivou (cofound variable) premennou, ktorá ako tretia vstupuje do vzťahu nezávislá verzus závislá premenná. Rušivá premenná je vo vzťahu so závislou premennou a jej pôsobenie ovplyvňuje vzájomný vzťah medzi nezávislou a závislou premennou.

Latentné a manifestné premenné

Predovšetkým v psychológii sme nútení merať rôzne psychologické konštrukty, ktoré sa neprejavujú jednoznačne a viditeľne jedným znakom. Napríklad motivácia, osobnostné charakteristiky ako introverzia, preferencie k intuícii a stovky iných konštruktov. Nemáme k dispozícii jednu otázku pomocou ktorej by sme vedeli zistiť stav napríklad motivácie. Takéto premenné sa nazývajú latentné a nakoľko ich nevieme merať priamo snažíme sa ich zachytiť prostredníctvom iných premenných, ktorých prejavy sú spojené s nami hľadanou latentnou, skrytou premennou. Preto: **Latentná premenná** je definovaná ako skutočnú vlastnosť, ktorú chceme merať, no neexistuje priamy spôsob jej merania. Taký prejav latentnej premennej, ktorý dokážeme merať sa nazýva **Manifestná premenná**. Hendl (2009, str. 505) definuje latentnú premennú ako: „priamo nemerateľnú, hypotetickú premennú stojacu v pozadí jednej alebo niekoľkých merateľných premenných.“

Praktický príklad premennej

Môžeme si ho ilustrovať na premennej s názvom: Počet bodov z vedomostného testu. Pri tejto premennej vieme, že môže dosiahnuť hodnoty od 0 (študent nevedel úplne nič) až po napr. 100 bodov. Pritom vieme, že 10 bodov je presne polovica z 20 bodov a zároveň škála obsahuje absolútnu nulu. Ak učiteľ presne a podrobne boduje jednotlivé úlohy, študent môže dosiahnuť aj napríklad 10,25 bodu.

Na základe týchto skutočností môžeme konštatovať, že ide o kvantitatívnu a kardinálnu premennú, ktorá je pomerová a je spojitá. V PSPP by sme ju označili ako škálová (scale). Ak by sme skúmali napríklad aj ako dlho sa študent pripravoval na test, potom by sme mohli v tomto konkrétnom prípade prípravy na test povedať nasledovné. Počet hodín prípravy je nezávislá premenná a počet bodov je závislá. Logicky predpokladáme, že počet bodov závisí od počtu hodín prípravy. To či je premenná závislá alebo nezávislá, je určené tým, čo chceme skúmať a ich pozície sa môžu meniť. Ak by sme skúmali subjektívnu hodnotenie úrovne študentov u učiteľa založené na známkach, potom by sme si mohli stanoviť, že počet bodov z vedomostného testu ovplyvňuje (nezávislá premenná) pohľad učiteľa na študenta (závislá). Inými slovami, vnímanie študenta učiteľom závisí od počtu bodov z vedomostného testu.

5.4 Výskum a jeho typy

Táto kniha sa zaoberá empirickým výskumom, ktorý vychádza z filozofického smeru empirizmu. Tento smer považuje skúsenosť za základ, alebo zdroj znalostí (Černík, Viceník, 2011). Existuje samozrejme viacero rôznych iných typov výskumu a napríklad Punch (2015, str. 11) uvádza: „teoretický výskum, analytický výskum, konceptuálne – filozofický výskum a historický výskum“. Zjednodušene by sme mohli v sociálnych vedách kategorizovať výskum na dva základné typy, a to kvantitatívny a kvalitatívny. Kvalitatívny výskum podľa Hendla (2016, str. 46, tab. 2.4) používa metódy ako pozorovanie, texty a dokumenty, interview a audio a video záznamy“. Z jednotlivých metód vidíme, že v prípade kvalitatívneho výskumu je práca s dátami a konkrétnymi štatistickými testami pomerne obmedzená. Naopak hlavné metódy kvantitatívneho výskumu podľa Hendla (2016, str. 44, tab. 2.2) sú: „štatistické skúmanie, experiment, oficiálne štatistiky, štruktúrované pozorovanie, obsahová analýza“. V prípade týchto štatistických metód je práca s dátami, číslami a rôznymi štatistickými testami oveľa viac rozšírená. Naša publikácia sa zaoberá v prvom rade kvantitatívnym výskumom a najzákladnejšími i najbežnejšími testami používanými v tomto type výskumu. Malá časť z týchto metód je korelačný a experimentálny výskum, ktoré si nasledujúcich odstavcoch podrobnejšie vysvetlíme.

Korelačný výskum

Korelačný výskum je v podstate skúmanie ako spolu dve alebo viac premenných súvisia a vieme vyjadriť dokonca aj ako silno. Pomocou korelačného výskumu však nevieme vysvetliť príčinnú súvislosť. Ak nám pri korelačnom výskume vyjde, že premennú spolu súvisia, nikdy nemôžeme z toho usudzovať, že jedna príčinne ovplyvňuje druhú. Pri korelačnom výskume skôr pozorujeme a meriame svet okolo nás, avšak neovplyvňujeme ho na rozdiel od experimentálneho

výskumu, kde priamo zasahujeme a ovplyvňujeme jednu alebo viac premenných. (Walker, 2013). Korelačný výskum je rozsiahlo používaný, pretože sa dá ľahko spracovať a aj interpretovať. Problémom je, že tento typ výskumu sa dostáva len na povrch problému. Teda dokáže nám povedať len, že existuje súvislosť medzi niektorými premennými, ale nič viac. Ak by sme tieto vzťahy chceli lepšie pochopiť, tak musíme spraviť ďalší výskum, respektíve inú analýzu. Niektorí skúsení výskumníci tvrdia, že korelačný výskum vytvára viac otázok ako vlastne rieši. Z toho dôvodu sa pri hlbšom skúmaní určitého problému korelačný výskum používa na prvých stupňoch na získanie určitého základného prehľadu vzťahov medzi premennými. Ďalej pri korelačnom výskume je dôležité uvedomiť si určité praktické obmedzenia. Napríklad ak máme stovky až tisícky jednotiek v súbore (respondentov) a robíme rozsiahlu korelačnú maticu, tak sa nám objaví určite viacero, neraz veľmi vysoko, signifikantných korelácií. Nie je spôsobené len tým, že by sme objavili niečo výnimočné, ale zvyčajne má veľký vplyv počet respondentov. Pri určitom zjednodušení môžeme konštatovať, že čím viac respondentov máme (stovky až tisíce), tým budú vzťahy medzi nimi signifikantnejšie. Korelačný výskum je aj inými slovami meranie vzájomných závislostí medzi premennými, ktoré môžeme realizovať rôznymi štatistickými metódami a postupmi. Tieto sú podrobne vysvetlené v nasledujúcich kapitolách o meraní vzájomných závislostí.

Experimentálny výskum

Pri experimentálnom výskume sa snažíme zistiť ako presne 1 premenná (nezávislá) ovplyvňuje druhú premennú (závislú). Teda ako závislá premenná závisí od nezávislej. Výskumník pri tomto type výskumu sa snaží zachovať všetky ostatné podmienky nezmenené, okrem jednej, ktorú premyslene sám mení, a potom sleduje jej dopad na celkový výsledok (Chajdiak, 2010). Zmenil sa, v akom smere a v akej intenzite, alebo sa vôbec nezmenil? Oproti korelačnému výskumu však dokáže presne vymedziť vzťahy medzi premennými. Je však prácnejší na prípravu, kde je dôležité dobre premyslieť celý experiment a pripraviť sa aj na zber dát. Musíme vytvoriť minimálne dve skupiny, experimentálnu a kontrolnú. Experimentálna je vystavená skúmanému podnetu (nezávislej premennej) a kontrolná nie.

5.5 Testovanie hypotéz

Testovanie hypotéz by sme mohli nazvať aj spôsobom rozhodovania sa za neurčitosti (Chajdiak, 2007, str. 169). Je to spôsob overovania platnosti tvrdení. Pacáková a kolektív (2009, str. 130) definuje testovanie hypotéz ako: „overovanie platnosti štatistických hypotéz pomocou údajov zistených vo výberovom súbore“. Hendel (2009, str.25) definuje hypotézu nasledovne: „Hypotéza je tvrdenie o podstate určitej situácie v skúmanej oblasti. Ide o návrh, predstavu o vzťahu

medzi uvažovanými výskumnými premennými.“ Podrobnejšie definujú a vymenovávajú výskumné hypotézy autori Mareš a kolektív (2015, str. 40) ako: „Predbežný predpoklad, domnienka o 1) existencii a 2) príčinách javov, o 3) vzťahoch medzi javmi, o 4) priebehu nejakého procesu, o 5) zmene apod.“

Stanovovanie hypotéz by malo byť založené už na buď na priamych existujúcich predchádzajúcich zisteniach z minulosti, alebo na základe úvah, extrapolácií alebo predpokladov iných zistení. Preto je vhodné pred stanovením samotnej hypotézy v krátkosti uviesť literárne zdroje a vysvetliť prečo a ako sme si danú hypotézu takto konkrétne stanovili. Hypotézy potvrdzujeme, alebo vyvraciamy na základe empirických (získaných z vonkajšieho sveta) zistení, faktov a dokladov. Ondrejko (2007, str. 139) uvádza, že hypotézu možno verifikovať alebo falzifikovať iba vtedy ak: „hypotéza používa presne vymedzené pojmy, a ktorá je jednoznačná i terminologicky“.

Postup tvorby a testovania hypotéz na základe použitia p- hodnoty definuje Hendel (2009, str. 182) ako aj autori Neubauer, Sedlačík, Kríž (2016, str. 197) nasledovne:

1. Formulácia výskumnej otázky vo forme nulovej a alternatívnej štatistickej hypotézy.

Pri testovaní hypotéz si vždy stanovujeme dve hypotézy, ktoré sa majú zvyčajne skratky H_0 a H_1 . H_0 je nazývaná nulovou hypotézou a H_1 sa nazýva hypotéza jedna, prípadne alternatívna hypotéza a pri experimentoch tiež experimentálna hypotéza. Obidve sa týkajú toho istého tvrdenia, pričom nulová hypotéza (H_0) tvrdí, že neexistuje žiaden rozdiel, prípadne efekt u skúmaných premenných a alternatívna hypotéza (H_1) tvrdí, že rozdiel, efekt existuje (Field, 2013, str. 870). Zjednodušene nulová a alternatívna hypotéza sú opačné tvrdenia. Vždy platí len jedna z nich. To, ktorá platí overujeme pomocou rôznych štatistických testov.

Alternatívna hypotéza môže byť **jednostranná**, alebo **obojustranná**. **Jednostranná hypotéza** je vtedy ak očakávame že bude testovaný parameter určite väčší, alebo naopak určite menší. Napríklad počet chybných výrobkov po školení zamestnancov klesne, teda jednoznačne očakávame zníženie chybovosti. Netvrdíme len že sa chybovosť zmení, my priamo očakávame pokles chybovosti. Alebo spokojnosť zamestnancov vo firme po zvýšení plátov vzrastie, teda očakávame určite rast spokojnosti. Obojustranná alternatívna hypotéza tvrdí, že zmena nastane, avšak nevieme či smerom nahor, alebo nadol.

Príklady alternatívnej, experimentálnej (H_1) hypotézy:

- Skupina po školení dosiahne vyššiu produktivitu ako skupina bez školenia (jednostranná alternatívna hypotéza).
- Podniky, ktoré majú napísanú informačnú stratégiu investujú viac do informačných technológií, než podniky, ktoré ju nemajú napísanú (jednostranná).
- Skúsení manažéri s viac než 10 ročnou praxou pri posúdení podnikateľského plánu prisudzujú parametru ziskovosti vyššiu

subjektívnu váhu, v porovnaní s manažérmi s menej než 10 ročnou praxou (jednostranná).

- Podnikatelia, ktorí založili viac ako tri podniky, vnímajú riziko pri založení konkrétneho podniku inak, ako tí, ktorí založili len 1 podnik (obojsstranná).

A naopak príklady nulovej hypotézy (H_0)

- Neexistuje štatisticky významný rozdiel v produktivite medzi skupinou bez školenia a skupinou so školením.
- Neexistuje štatisticky významný rozdiel vo výške investícií do informačných technológií medzi podnikmi, ktoré majú napísanú Informačnú stratégiu a tými, ktoré ju nemajú napísanú.
- Neexistuje štatisticky významný rozdiel v prisudzovaní subjektívnej váhy parametru ziskovosť pri posudzovaní podnikateľského plánu medzi manažérmi s 10 ročnou praxou a manažérmi s menej než 10 ročnou praxou.
- Nie je štatisticky významný rozdiel medzi podnikateľmi ktorí založili viac ako tri podniky, vo vnímaní rizika pri založení konkrétneho podniku, a takými, ktorí založili len 1 podnik.

Pri testovaní hypotéz začíname s testovaním nulovej hypotézy. Postup testovania hypotéz spočíva v tom, že najprv formálne očakávame, že platí nulová hypotéza, teda, že rozdiel, efekt neexistuje a až na základe empirických faktov môžeme tvrdiť že platí alternatívna hypotéza, teda opak nulovej hypotézy. Teda, že sme zistili, že nulová hypotéza neplatí, preto potom s určitou pravdepodobnosťou platí alternatívna hypotéza (H_1). Inými slovami, my dokážeme, že sme nezistili, že neexistuje rozdiel. Ako výskumníkov a vedcov nás väčšinou zaujíma práve to, že existuje rozdiel, teda, že sme napr. zistili, že: školenie zlepšuje produktivitu, informačná stratégia znamená väčšie investície, a, že skúsení manažéri priradujú vyššiu váhu ziskovosti, než nováčikovia atď. Preto výskumník očakáva, že nulová hypotéza sa zamietne a potvrdia sa jeho očakávania spojené s platnosťou alternatívnej hypotézy Hendel (2009). Tým sa potvrdia jeho predpoklady o tom, aké sú vzťahy medzi skúmanými premennými.

V reálnom svete vieme, že len v málo veciach si môžeme byť úplne istí na celých 100 percent. Vo väčšine prípadov sme si veľmi istí napr. na 80 percent, prípadne extrémne istí napr. na 98 %, ale reálne vždy pripúšťame, že existuje malá, niekedy aj veľmi malá pravdepodobnosť (napr. 0,005 %), že by sme sa mohli mýliť. Taktiež pri testovaní hypotéz musíme určitú pravdepodobnosť chybovania brať do úvahy. Tá pravdepodobnosť chybovania je spôsobená tým, že nemáme celú populáciu vo svojom výskume a máme len výber z nej, ktorý môže v niektorých prípadoch reprezentovať len veľmi malú časť populácie. Ten výber môže byť zhodou rôznych náhod práve taký, že nám poskytuje celkom mylné predstavy o celej populácii. Tejto problematike sa venujeme v ďalšom kroku pri postupe testovania hypotézy.

2. Zvolenie prijateľnej úrovne chyby rozhodovania.

V tomto kroku si sami stanovíme hladinu významnosti, na ktorej budeme hypotézy testovať. Hladina významnosti je pravdepodobnosť toho, že sa zamietne nulová hypotéza, napriek tomu, že v skutočnosti platí. Inak sa nazýva aj falošné zamietnutie v skutočnosti pravdivej nulovej hypotézy. Nazýva sa **Chyba prvého druhu** a označuje sa gréckym písmenom alfa: α . Inými slovami, že zistíme, že v našom výbere existuje rozdiel (efekt), no v skutočnosti rozdiel, efekt v populácii neexistuje. Prakticky by sme chceli aby sme sa tejto chybe vyhli a tak *hladinu významnosti (inak aj pravdepodobnosť chyby prvého druhu)* stanovujeme prísne. Teda, že pravdepodobnosť, že spravíme túto chybu bola čo najnižšia. V praxi sa bežne stanovuje pravdepodobnosť chyby 0,05, teda 5% alebo 0,01, teda 1% (Neubauer, Sedlačík, Kříž, 2016, str. 197).

Zároveň platí, že ak si stanovujeme, že neexistuje rozdiel, teda H_0 platí, potom sa môžeme dopustiť presne opačnej chyby. To znamená, že prijmem H_0 aj keď v skutočnosti neplatí. Inými slovami, napriek tomu, že v skutočnosti v populácii existuje rozdiel, my sme v našom výbere chybné zistili, že neexistuje. Naším testovaním sme nedokázali odhaliť existujúci rozdiel. Táto druhá chyba sa nazýva **Chyba II druhu** a označuje sa β . Pravdepodobnosť vyčíslenú ako $1 - \beta$ sa nazýva **Sila testu**. Chyba I a chyba II druhého druhu sú vo vzájomnom vzťahu, ktorý zobrazuje nasledovná tabuľka.

Tabuľka č. 5.5 Schéma testovania hypotéz (Hendl, 2009, str. 191)

	Náš záver o teste	
Objektívna skutočnosť	H_0 platí	H_1 neplatí
H_0 platí	Záver bol správny	Chyba I. druhu
H_1 neplatí	Chyba II. druhu	Záver bol správny

3. Výpočet konkrétnej testovacej štatistiky

Nakoľko si môžeme stanovovať veľmi rôzne hypotézy, na ich overenie môžeme použiť tiež pomerne veľa rôznych štatistických testov a metód. K variabilite potencionálne použiteľných metód prispieva tiež niekedy rozdielny charakter dát a typov premenných. Pri týchto testoch sledujeme hladinu významnosti, ktorú sme si stanovili na úrovni $\alpha = 0,05$. Predpokladajme, že by sme chceli otestovať vyššie uvedené príklady hypotéz. Na ich overenie použijeme nezávislý t – test pre dva výbery (viac o teste čitateľ nájde v samostatnej podkapitole v knihe). Nulová hypotéza predpokladá, že medzi skupinami nie je rozdiel. Pomocou softvéru PSPP aplikujeme t – test a sledujeme výsledky pre hodnotu p – v softvéri označenú ako Sigma. Ak sme predpokladali, že skupiny sa medzi sebou nebudú líšiť (nulová hypotéza), tak hodnota sigma bude vyššia než 0,05, potom nulovú hypotézu zamietame a prijímame alternatívnu.

4. Záverečné odporúčanie

Na základe výsledkov štatistických testov buď a) nepreukážeme rozdiel (efekt), a potom prijímame nulovú hypotézu H_0 , b) preukážeme rozdiel alebo efekt a prijímame alternatívnu hypotézu H_1 . Alternatívnu hypotézu prijímame na určitej hladine významnosti.

Z praktického hľadiska je schopnosť premýšľať o probléme a z týchto systematických úvah si stanovovať dobre premyslené hypotézy patrí k základným predpokladom kvalitného vedeckého myslenia a práce. Preto si aj my budeme pri rôznych testoch opakovane stanovovať hypotézy a budeme ich overovať.

6 ZÁKLADNÁ ANALÝZA DÁT

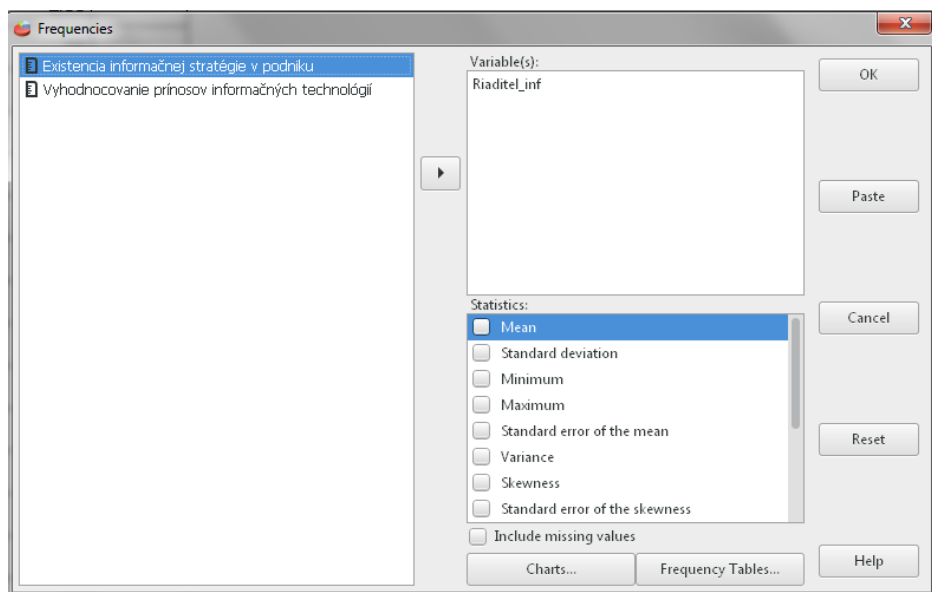
Postup dátovej analýzy by mal vždy začať dôkladným prevereníím prvotných dát, pretým než sa pustíme do podrobnejších analýz. Na to nám slúžia dva štatistické nástroje: frekvenčná analýza (frequency analysis) a opisná štatistika (descriptive statistics). V tomto prípade skúmame hodnoty jednej premennej a nevzťahujeme ich k hodnotám iných premenných. Preto tento typ analýz sa nazýva jednorozmerná štatistická analýza.

6.1 Frekvenčná analýza

Je založená na sčítaní prípadov v jednotlivých kategóriách. Teda koľko jednotiek máme v tej ktorej kategórii. Napr. koľko mužov je v súbore, koľko žien alebo aj koľko respondentov si neuviedlo svoje pohlavie a program ich zobrazuje ako nezaradených. Frekvenčná analýza je prvotným a najjednoduchším opisom dát. Čo sa týka vedeckého publikovania, tak početnosť jednotlivých prípadov sa vždy uvádza pri výsledkoch všetkých štatistických testov. Program PSPP štandardne uvádza početnosti jednotiek v rámci výstupov väčšiny štatistických testov, a preto nie je potrebné robiť frekvenčné analýzy druhý krát.

1 úloha: V dátovom súbore s názvom Príklad č. 3 Riaditeľ IT vyčíslite koľko podnikov má riaditeľa informatiky a koľko nie z celkového počtu podnikov.

Riešenie: Klikneme na *Analyzovať (Analyze)* vyberieme *Opisná štatistika (Descriptive statistics)* a potom *Frekvencie (Frequencies)*. V dialógovom okne nám ponúka *Výber konkrétnej premennej (Variable(s))*, z ktorej chceme počítať frekvenčné štatistiky a typ štatistiky (Statistics). Dole sú uvedené grafy (Charts) a frekvenčné tabuľky (Frequency Tables). Vyberieme premennú *Riaditeľ_inf*, a presunieme ju do políčka *premenné (Variable(s))* a spustíme OK.



Obr. 6.1.1 Frekvenčná analýza postup

Existencia pozície riaditeľa informačných technológií

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
nie je riaditeľ	,00	114	50,67	50,67	50,67
je riaditeľ	1,00	111	49,33	49,33	100,00
<i>Total</i>		225	100,0	100,0	

Obr. 6.1.2 Výsledky frekvenčnej analýzy

Výsledok: V tabuľke v stĺpci *Frekvencie (Frequency)* vidíme, že je 144 podnikov (50,67 %), ktoré nemajú riaditeľa informatiky a 111, ktoré majú riaditeľa (49,33%).

Uvádzanie výsledkov: Naš príklad : Z 225 skúmaných podnikov 114 (50,67%) nemalo riaditeľa informačných technológií a 111 (49,33%) malo riaditeľa.

6.2 Opisné charakteristiky

K základným štatistickým nástrojom, ktoré charakterizujú akýkoľvek súbor nameraných premenných môžeme zaradiť nasledovné, ktoré rozdelíme do týchto skupín:

Ukazovatele centrálnej tendencie (Measures of central tendency). Tri najviac používané sú: Aritmetický priemer (Mean), Medián (Median) - hodnota, ktorá rozdeľuje súbor na dve rovnako veľké časti a Modus (Mode). – najčastejšie sa vyskytujúca hodnota v skúmanom súbore.

Ukazovatele variability : Rozptyl (Variability) a Smerodajná odchýlka (STDEV) a môžeme sem zaradiť aj variačné rozpätie. Variačné rozpätie vypočítame ako

rozdiel najvyššej hodnoty, maximálnej (funkcia MAX) a minimálnej – najnižšej hodnoty v súbore (funkcia MIN).

Miery špicatosti a šikomosti : Šikomosť (SKEW), ktorá meria zošikmenosť, resp nesymetriu dát. Špicatosť (KURT) meria odchýlku skúmaného rozdelenia od normálneho rozdelenia.

V nasledujúcich podkapitolách každú z nich podrobne opíšeme a zdefinujeme.

6.3 Ukazovatele centrálnej tendencie

Tieto ukazovatele by sme mohli zjednodušene charakterizovať ako najtypickejšie hodnoty dát, ktoré reprezentujú súbor. Cieľom týchto štatistických ukazovateľov je čo najlepšie opísať súbor dát (dataset) aby používateľ mal predstavu s akými hodnotami konkrétnych premenných pracuje.

Aritmetický priemer (Mean, M)

Aritmetický priemer je súčet všetkých hodnôt premennej delený ich počtom. Ide o najjednoduchšiu a najčastejšie používanú charakteristiku polohy, umiestnenia v súbore. Je súčasťou výpočtu mnohých ďalších štatistík a metód analýzy, a preto je jeho používanie tak široké. Priemer môžeme počítať ak naša premenná je intervalová, resp. pomerová a nie ak sú dáta ordinálne. Priemer nie je vhodný ak máme extrémne hodnoty v dátach a dáta nie sú rozdelené symetricky. Jeho najväčšou nevýhodou je, že priemer je „citlivý“ na zmeny hodnôt a pri zmene hodnoty väčšinou jedinej jednotky sa mení tiež. Ak priemer vnímame ako určitý opis dát, musíme si byť vedomí skreslení, ktoré nám prináša. Ak poznáme len priemer nevieme nič o variabilite v súbore, teda o tom aké hodnoty nadobúda konkrétna premenná. Napríklad ak vieme že priemerný hrubý zárobok vo firme je 1500 Eur. Tento priemer však mohol vzniknúť rôznymi spôsobmi:

$$A: 1500 + 1600 + 1400 + 1450 + 1550 = 7500 / 5 = 1500$$

$$B: 4500 + 780 + 720 + 710 + 790 = 7500/5 = 1500$$

Ako vidíme v prípade A sú príjmy veľmi podobné a každý jeden prípad sa moc neodlišuje od priemeru samotného. V prípade B máme rovnaký priemer ako v prvom prípade, ale každá hodnota sa od priemeru značne odlišuje a niektoré sa odlišujú veľmi aj medzi sebou navzájom. Z tohto dôvodu je potrebné uvádzať nielen priemer, ale spolu s priemerom aj mieru variability. Tou môže byť rozptyl alebo častejšie uvádzaná smerodajná odchýlka, ktorá je druhou odmocninou rozptylu. Ďalším dôvodom prečo sa uvádza priemer, počet štatisticky analyzovaných jednotiek a smerodajná odchýlka sú meta – analýzy. Tieto sa počítajú z jednotlivých samostatných analýz a sumarizovaním všetkých malých výskumov dokážeme pomerne presne zistiť aký je efekt javu v celej populácii. Na to však potrebujeme konkrétne výsledky opisných štatistík.

Modus (Mode, Mo)

Modus je najčastejšie sa vyskytujúca hodnota v súbore, teda tá, ktorá má najväčšiu početnosť. Pokým aritmetický priemer sa reportuje vždy, keď máme dáta, z ktorých má význam ho počítať, modus sa vo výskumných a vedeckých článkoch reportuje pomerne málo. Príčinou je, že dáta v súbore reprezentuje v porovnaní s priemerom a mediánom najmenej presne.

6.4 Kvantily

Na lepší opis celého súboru môžeme použiť kvantily. Kvantily sú charakteristiky polohy, teda umiestnenia. Existujú medzi nimi presne vyjadriteľné lineárne vzťahy. Ku kvantilom patria: medián (MEDIAN), percentily, kvartily, decily. Čo sa týka vzájomných vzťahov, uvádzame niektoré z nich, konkrétne medián je aj 50-ty percentil a tiež druhý kvartil. Ďalej Maximum je stý percentil a štvrtý kvartil. Konkrétne hodnoty si čitateľ môže porovnať na doterajších výsledkoch.

Medián (Median, Me)

Medián je hodnota, ktorá zoradený súbor podľa veľkosti hodnôt premennej (od najmenšieho po najväčšie) rozdelí na dve rovnako početné polovice. Teda prvá polovica súboru má menšie hodnoty než medián a druhá polovica väčšie. Čo sa týka ukazovateľov polohy, tak medián je najstabilnejší, teda pri pridávaní alebo odoberaní jednotiek sa medián zvyčajne nezmení, pokým priemer sa zmení vždy (pokiaľ pridávaná hodnota nie je presne totožná s priemerom). Medián je teda stredom rozdelenia dát a v súvzťažnosti s inými kvantilmi je súčasne aj 50tym percentilom.

Medián sa bežne využíva v sociálnom výskume na delenie súboru na dve skupiny, ktoré sú potom porovnávané voči iným premenným. Napríklad máme dáta o tom, koľko hodín sa študenti pripravovali na skúšku a máme aj body zo skúšky. Ak by sme chceli porovnať dosiahnuté body zo skúšky usilovne študujúcich študentov s menej usilovnými, potom vypočítame medián z počtu hodín prípravy. To bude hodnota, ktorá rozdelí študentov na 2 skupiny, nad mediánom budú usilovnejšie študujúci a pod mediánom budú menej usilovne študujúci. Potom pomocou nezávislého dvojvýberového t – testu porovnáme body zo skúšky týchto dvoch skupín.

Maximálna hodnota (Maximum, Max) a Minimálna hodnota (Minimum, Min)

Maximálna hodnota (Maximum, Max) je najvyššia hodnota v súbore a minimálna (Minimum, Min) je najnižšia hodnota v súbore. Pri reportovaní výskumu niekedy býva zvykom uviesť minimálne a maximálne hodnoty spolu ostatnými opisnými

štatistikami. Rozdiel medzi najvyššou a najnižšou hodnotou v súbore sa nazýva aj Variačné rozpätie.

Šikmosť (Skewness)

Šikmosť rozdelenia znamená, či sa viac v súbore vyskytujú vyššie alebo nižšie hodnoty skúmanej premennej. Ak vyjde koeficient šikmosti nula, potom ide o symetrické rozdelenie. Ak je menší ako nula ide o vpravo zošikmené rozdelenie (viac väčších hodnôt), ak je koeficient vyšší ako 0 ide o vľavo zošikmené rozdelenie (viac menších hodnôt). Pri praktickom výskume sa nám len výnimočne podarí vytvoriť taký výber z populácie (výberový súbor), ktorý by mal symetrické rozdelenie. Takmer vždy bude zošikmený vpravo alebo vľavo, niekedy viac alebo menej.

Špicatosť (Kurtosis)

Ak premenné zobrazíme graficky a väčšina hodnôt premennej sa nachádza blízko priemeru a máme len pomerne málo vysokých a nízkych hodnôt potom graf nadobudne typický ostrý vrchol. Takéto rozdelenie nazývame špicaté. Ak je výsledok väčší ako 0 potom je to špicatejšie rozdelenie a ak je výsledok menší ako 0 potom je rozdelenie plochejšie. Špicatejšie rozdelenie znamená, že väčšina hodnôt je blízko priemeru. Plochejšie rozdelenie znamená viac extrémnejších hodnôt.

6.5 Ukazovatele variability

Ukazovatele variability nás informujú o miere premenlivosti, variability dát. K najpoužívanejším patria sem rozptyl, smerodajná odchýlka, variačné rozpätie interkvantilové rozpätie. Ak výberový súbor obsahuje jednotky (napr. respondentov), ktoré nadobúdajú extrémne hodnoty, teda extrémne vysoké alebo nízke v porovnaní s ostatnými hodnotami, tak tieto výrazne ovplyvňujú rozptyl a teda aj smerodajnú odchýlku.

Rozptyl (Variance, Var)

Rozptyl je priemerná kvadratická odchýlka od aritmetického priemeru.

Smerodajná odchýlka (Standard deviation, SD)

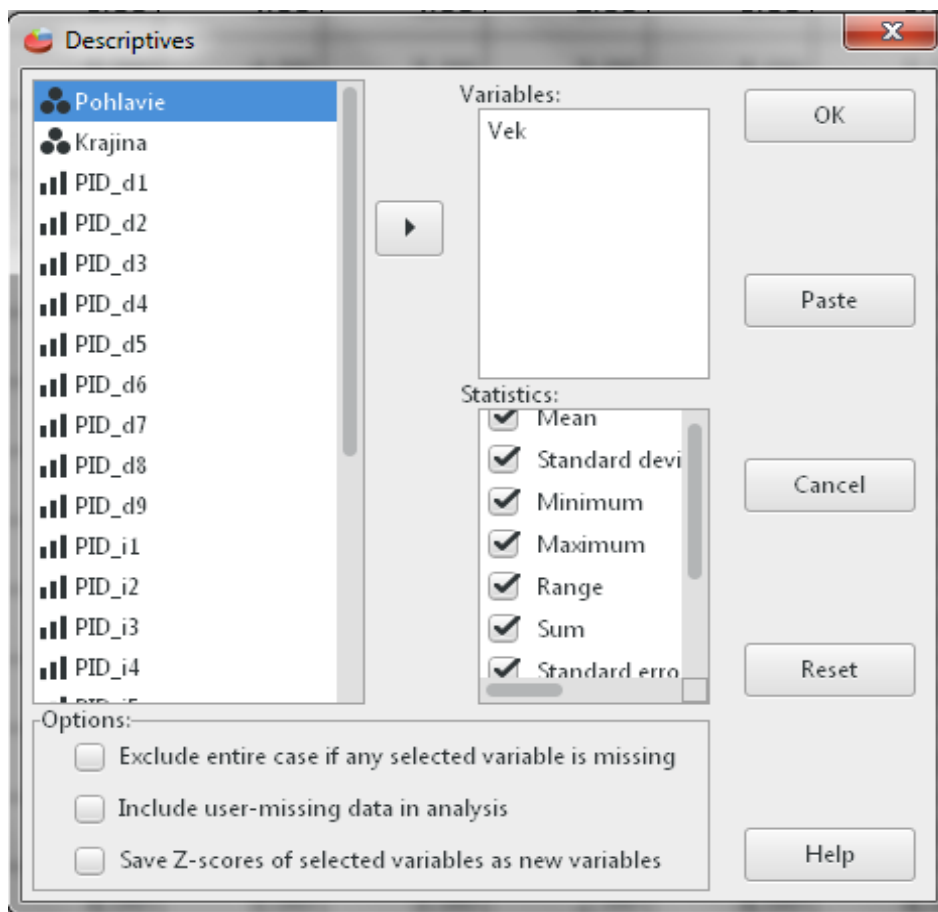
Smerodajná odchýlka (STDEV) je štatistickou funkciou, ktorá meria variabilitu v súbore. Vyjadruje mieru rozptýlenia hodnôt od aritmetického priemeru. Smerodajná odchýlka je druhou odmocninou rozptylu. Výhodou smerodajnej odchýlky oproti rozptylu je, že jej výsledok je v rovnakej mierke akou boli merané prvotné dáta.

Táto funkcia nevyhnutnou charakteristikou pri akomkoľvek serióznom štatistickom opise súboru a takmer vždy sa uvádza spolu s priemerom.

K smerodajnej odchýlke sa viaže zákon 3 sigma, ktorý znamená nasledovné. Keď pripočítame a aj odpočítame smerodajnú odchýlku od priemeru, tak nám vznikne interval, ktorý v sebe zahŕňa 68,27 % hodnôt v súbore (v prípade normálneho rozdelenia hodnôt v súbore). Ak k priemeru pripočítame, resp. odpočítame 2 smerodajné odchýlky, potom v tomto intervale bude 95,45 % napozorovaných hodnôt. Ak to isté urobíme s tromi smerodajnými odchýlkami, potom bude v súbore 99,73% napozorovaných hodnôt. Pri štyroch smerodajných odchýlkach od priemeru sa v tomto rozpätí nachádza 99,994 % napozorovaných hodnôt.

1. úloha: Pomocou opisných štatistík charakterizujte, opíšte vek respondentov v príklade č. 1 – Faktorová analýza

Postup: V tomto príklade vypočítame výsledky pre všetky opisné štatistiky a potom sa im budeme podrobnejšie venovať pri vysvetľovaní konkrétnych opisných štatistík. Klikneme na *Analyzovať (Analyze)* vyberieme *Opisná štatistika (Descriptive statistics)* a potom znova *Opisná štatistika (Descriptives)*. V dialógovom okne nám ponúka výber konkrétnych *Premenných (Variables)*, z ktorej chceme počítať frekvenčné štatistiky a typ štatistiky (*Statistics*). V konkrétnych štatistikách máme uvedené prakticky všetky bežne používané opisné štatistické metódy. V našom príklade ich zaškrtneme všetky a potom ich budeme interpretovať v jednotlivých častiach opisných štatistík.



Obr. 6.5.1 Sprievodca metódou Opisných štatistík pre premennú Vek

```
DESCRIPTIVES
DESCRIPTIVES
/VARIABLES= Vek
/STATISTICS=ALL.
```

Valid cases = 872; cases with missing value(s) = 379.

Variable	N	Mean	S.E. Mean	Std Dev	Variance	Kurtosis	S.E. Kurt	Skewness	S.E. Skew	Range	Minimum	Maximum	Sum
Vek	493	27,93	,37	8,30	68,89	2,16	,22	1,42	,11	50,00	18,00	68,00	13770,00

Obr. 6.5.2 Výsledok opisnej štatistiky pre premennú vek

Výsledok: Na základe údajov v tabuľke výsledkov môžeme konštatovať, že z celkového počtu 872 prípadov (riadky v tabuľke dát) 379 respondentov neuviedlo svoj vek, resp. nemáme k nim túto informáciu. Zostalo nám teda $872 - 379 = 493$ respondentov. Ich priemerný vek (*Mean*) bol 27,93 roka. *Smerodajná chyba priemeru (Standard Error of Mean, S.E. Mean)* = 0,37 nám opisuje odchýlku od nami získaného priemeru a teoretického rozdelenia premennej. Čím je naša vzorka väčšia, tým je smerodajná chyba priemeru nižšia. Princíp smerodajnej chyby je rovnaký aj u *Smerodajnej chyby šikmosti (S.E. Skew)* aj u *Smerodajnej chyby špicatosti (S.E. Kurtosis)*.

V našom príklade sme zistili, že šikmosť je 1,42. Táto hodnota je vyššia ako nula, čo znamená, že je tam viac študentov, ktorí majú nižší vek.

V našom prípade je Špicatosť 2,16, to znamená špicatejšie rozdelenie. Teda väčšina respondentov má vek blízko priemeru.

V našom príklade bol Rozptyl veku 68,89. Odmocnina z rozptylu je Štandardná smerodajná odchýlka, teda $\sqrt{68,89} = 8,3$ roka.

Minimálna hodnota (Minimum) bola v našom prípade 18 rokov a *Maximálna hodnota (Maximum)* bola 68 rokov. Rozdiel medzi týmito hodnotami je Variáčne rozpätie (Range) = 50.

Delenie súboru pomocou opisných štatistík

Opisné štatistiky sa používajú aj na delenie súboru na viac častí. Robíme tak vtedy, keď pomocou 1 premennej, napr. veku, rozdelíme súbor na skupiny. Napríklad vytvoríme skupinu najmladších desať percent súboru a najstarších desať percent nášho súboru. Potom tieto skupiny porovnávame medzi sebou na základe hodnôt druhej premennej, napríklad skóre v dotazníku intuície a zisťujeme, či sú mladší respondenti štatisticky významne odlišní od starších respondentov v intuícii. V nasledovných odstavcoch uvediem tri metódy najčastejšie používané na delenie súboru na konkrétne časti.

Metóda Medián split.

V praktickom výskume sa často používa metóda delenia dát na dva súbory Median split. Metóda spočíva v tom, že súbor rozdelíme na dve časti podľa mediánu, kde jedna polovica má nižšie hodnoty a druhá vyššie. Potom tieto dva súbory zvyčajne pomocou T – testov porovnávame voči iným premenným. Táto metóda zahŕňa všetky jednotky v súbore, teda aj tie, ktoré sú blízko mediánu a aj tie ktoré sú viac vzdialené. Ak by sme sa chceli zamerať na tie, ktoré dosahujú tie najvyššie, resp. najnižšie hodnoty, inými slovami aj extrémnejšie hodnoty tak môžeme použiť nasledovné metódy delenia súboru.

Metóda Smerodajnej odchýlky

Rozdelí súbor tak, že pripočíta k priemeru smerodajnú odchýlku. Získame tak hodnotu priemer + smerodajná odchýlka. Všetky jednotky v súbore, ktoré majú túto a vyššiu hodnotu sú jedna extrémna skupina. Druhú skupinu získame tak, že smerodajnú odchýlku odpočítame od priemeru a všetky menšie od získanej hodnoty budú v tejto skupine. Na základe zákona 3 sigma vieme, že 68, 27 % jednotiek tam nebude patriť a zostane nám 15,86% s najvyššími hodnotami a tiež 15,86% jednotiek s najnižšími hodnotami. Metóda smerodajnej odchýlky sa môže použiť násobne, teda k priemeru pripočítame dvojnásobok smerodajnej odchýlky a dvojnásobok odpočítame. Takto vylúčime 95,45 % jednotiek a zostane nám

4,55 % ktoré rozdelíme na najvyšších a najnižších 2,28 %. Pri tomto postupe je však potrebné mať veľmi veľké vzorky (minimálne stovky, lepšie tisíce jednotiek), aby sme po vylúčení tak veľkého percenta jednotiek mali v súbore, ktorý nám zostane ešte zmysluplné počty na ďalšiu analýzu.

Metóda Kvantilov, decilov a percentilov

Táto metóda pracuje na rovnakom princípe ako predchádzajúce dve, no za deliace kritérium používa konkrétne ukazovatele polohy, ktoré si sami stanovíme (napr. kvantily, decily alebo percentily). Napríklad si povieme, že pre potreby výskumu nás bude zaujímať 10 percent najlepších, ktorých porovnáme s 10 percentami najhorších.

7 ANALÝZA RELIABILITY

Na analýzu reliability softvér PSPP ponúka dve metódy : Rozdeľovanie testu (split-half) a Cronbachovu alfu. V nasledujúcich podkapitolách si opíšeme na akom princípe sú tieto metódy založené.

Rozdeľovanie testu (split-half)

Pri hodnotení kvality dotazníka, ktorý má merať určitú latentnú premennú, predpokladáme, že všetky otázky z dotazníka by sa mali vzťahovať k danej latentnej premennej. Konkrétne z logického hľadiska by sa nemalo stať, že respondenti na 8 z nich odpovedajú podobne a na jednu úplne opačne než na ostatné. Inými slovami odpovede na každú otázku by mali byť konzistentné s ostatnými odpoveďami a tým aj s celkovým dotazníkom. Najjednoduchším spôsobom ako overiť či respondenti odpovedajú konzistentne je metóda rozdeľovania testu (split-half). Táto metóda rozdelí odpovede náhodným spôsobom u každého participanta (resp. prípadu skúmania) na dve polovice, potom vypočíta skóre pre každú polovicu a koreluje ich vzájomne. Teda skúma či jedna polovica dotazníka súvisí z druhou polovicou. Táto korelácia by v ideálnom prípade mala byť veľmi vysoká. Ak je nízka, potom časť testu meria niečo iné (inú latentnú premennú) než druhá časť. Praktickým problémom tejto jednoduchej metódy je, že súbor otázok môžeme rozdeliť na dve polovice rôznym spôsobom. Hoci ho delíme náhodne, rôzne delenia súboru produkujú rôzne polovice a potom aj rôzne výsledky, ktoré sa zvyčajne odlišujú len málo, ale predsa sa líšia. Preto bola vyvinutá presnejšia metóda, ako sa tomuto problému vyhnúť. Volá sa Cronbachova alfa.

Cronbachova alfa

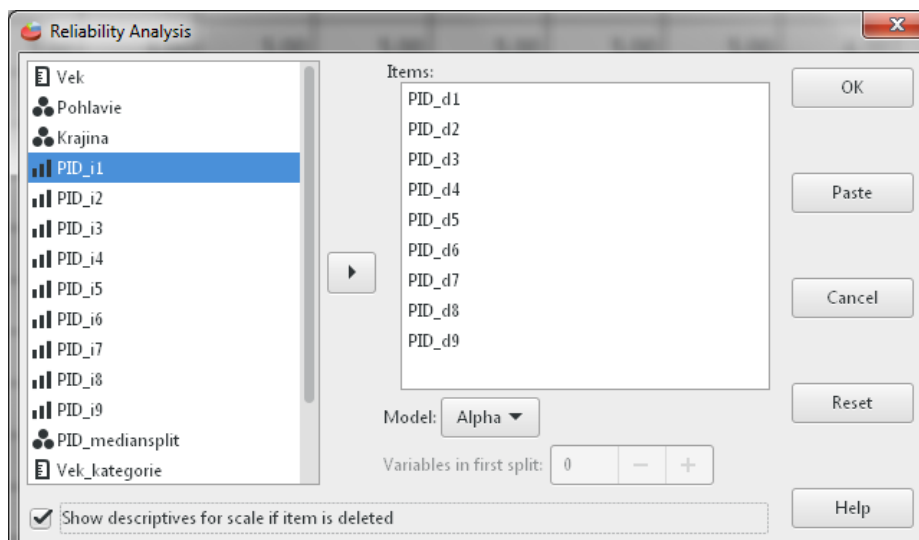
Princíp je rovnaký ako u metódy rozdeľovania súboru (split half). Rovnako rozdelí súbor na dve polovice ale súčasne vygeneruje všetky možné kombinácie akými je možné súbor rozdeliť na dve polovice. Týmto získame všetky možné spôsoby ako sa dal súbor rozdeliť a potom z nich počítame vzájomné korelácie a priemer z týchto korelácií je Cronbachova alfa. Týmto zahrnieme všetky možné existujúce kombinácie a tým sa vyhneme akejkol'vek chybe spôsobenej len čiastočným výberom. Je to najpoužívanejšia štatistická metóda na meranie reliability. Rovnako ako pri metóde rozdeľovania testu platí, že čím sa výsledok Cronbachovej alfy viac blíži k hodnote 1 tým reliabnejšie dotazník meria.

V praxi sa samozrejme stretávame s rozličnými hodnotami, a preto je dôležité určiť si určité intervaly. Field a Hole (2010, str. 48) uvádza, že hladina Cronbachovej alfy nad úrovňou 0,8 je akceptovateľná úrovne nižšie naznačujú už pomerne nereliabnú škálu. Anýžová (2015, str. 66) uvádza, že hodnota 0,7 je už spodná hranica spoľahlivosti testu, a preto pod touto hranicou môžeme hovoriť o skúmanej škále, dotazníku ako o nedostatočne reliabnej.

Pri vedeckom skúmaní akéhokoľvek dotazníka sa vždy uvádza aj údaj o nameranej reliabilite. Umožňuje to iným vedcom porovnať si výsledky a je to aj informácia pre nás samotných aké kvalitné dáta máme my. Nízka reliabilita našich výsledkov v porovnaní s inými autormi môže naznačovať rozličné chyby, ktorých sme sa sami dopustili. Napríklad zle nakódované dáta, chyby v dátach pri nahrávaní do programu, nezaujímaví respondenti, vyplnené odpovede bez čítania otázky a mnohé iné rôzne príčiny.

1. úloha: Vypočítajte reliabilitu škály deliberácia z dotazníka Preferencia pre deliberáciu a intuíciu (PID) pomocou metódy Cronbachova alfa.

Riešenie: Klikneme na menu *Analyzovať (Analyze)* a potom vyberieme *Reliabilitu (Reliability)*. Dotazník PID má dve škály, deliberáciu (racionalitu) a Intuíciu. Každá z týchto škál má 9 otázok, ktoré sú označené skratkou d pre deliberáciu (konkrétne PID_d1, PID_d2 atď.) a skratkou i pre intuíciu (konkrétne PID_i1, PID_i2 atď.). Keďže chceme analyzovať škálu deliberácia, musíme vybrať všetkých deväť otázok, ktoré sa vzťahujú k deliberácii. Dole je uvedený Model analýzy a máme možnosť vybrať si z dvoch modelov: *Cronbachova alfa* a *Split half*. Vyberieme Cronbachovu alfu. V ľavej časti dole dialógového okna je možnosť: *Zobraziť údaje pre škálu ak je otázka vymazaná (Show descriptives for scale if item is deleted)*. Táto možnosť nám umožňuje preskúmať vplyv každej otázky na celkovú reliabilitu. Vo výsledkoch uvidíme akú celkovú reliabilitu by mal dotazník ak by sme vymazali konkrétnu otázku. Tým zistíme presne, ktoré otázky nám reliabilitu dotazníka zvyšujú a ktoré ju znižujú. Na záver už len klikneme *OK*. Zaškrtnuté políčka, vybraný model a jednotlivé položky analýzy vidíme v nasledujúcom obrázku (obr. 1.1)



Obr. 6.5.1 Sprievodca analýzou reliability škály deliberácia z dotazníka Preferencia pre deliberáciu a intuíciu (PID)

Interpretácia: Výsledok anlyzy vidíme v nasledujúcom obrázku (obr. 1.2) V hornej tabuľke (*Case processing summary*) máme počet analyzovaných jednotiek = respondentov (*Cases – Valid*), ktorí vyplnili daný dotazník a bolo ich 872.

Reliabilita škály delibarácia dotazníka PID meraná pomocou Cronbachovej alfy je $\alpha = 0,82$ a je uvedená v strednej tabuľke s názvom *Reliability Statistics*. Hodnotu 0,82 môžeme považovať za veľmi dobrú.

V spodnej tabuľke máme výsledky pre každú z otázok uvedené takom vo formáte, že vidíme aké hodnoty priemeru, rozptylu, korelácie otázky s celkovým dotazníkom a aj Cronbachovej alfy by sme dosiahli, ak by sme danú otázku vymazali. Toto je pre nás pomerne užitočná informácia, pretože vieme ako každá jedna otázka zo subškály dotazníka vplýva na celkové výsledky dotazníka.

```
RELIABILITY
RELIABILITY
/VARIABLES= PID_d1 PID_d2 PID_d3 PID_d4 PID_d5 PID_d6 PID_d7 PID_d8 PID_d9
/MODEL=ALPHA
/SUMMARY = TOTAL.
```

Scale: ANY

Case Processing Summary

		N	%
Cases	Valid	872	100,00
	Excluded	0	,00
	Total	872	100,00

Reliability Statistics

Cronbach's Alpha	N of Items
,82	9

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
PID_d1	30,04	26,83	,62	,79
PID_d2	30,07	27,10	,62	,79
PID_d3	30,03	27,94	,47	,81
PID_d4	30,57	26,29	,55	,80
PID_d5	30,82	28,59	,36	,82
PID_d6	30,09	27,35	,55	,80
PID_d7	30,35	26,39	,62	,79
PID_d8	30,60	27,43	,47	,81
PID_d9	30,80	27,95	,45	,81

Obr. 6.5.2 Výsledky analýzy reliability metódou Cronbachova alfa

Reportovanie výsledkov: Náš príklad : Reliabilita škály delibarácia z dotazníka PID (Preference for deliberation and intuition, Betsch, 2004) je $\alpha = 0,82$. Reliabilita jednotlivých otázok sa pohybovala od $\alpha = 0,79$ (PID_d1) po $\alpha = 0,81$ (PID_d3).

Iné príklady (domáce):

„V slovenskom prostredí skúmali Hanák, Čavojová, Ballová Mikušková (2012) reliabilitu dotazníka na 60-tich administratívnych a manažérskych pracovníkoch a zistili úroveň Cronbachovej alfy pre Uvažovanie 0,857 a pre Intuíciu 0,634. Čavojová, Ballová Mikušková, Hanák (2013) zistili Cronbachovu alfu ($n=169$) pre Intuíciu 0,73 a pre Uvažovanie 0,82. Čo sa týka merania reliability pomocou Cronbachovej alfy, tak sme na základe výsledkov iných autorov nútení

konštatovať, že sa jedná o hraničné hodnoty (menšie než 0,8).“ (Hanák, 2013, str. 83)

Zahraničné príklady (anglické):

“In this study, the internal consistency by Cronbach’s alpha for PID-D was measured as .795 and for PID-I as .744, which is an acceptable level and in accordance with Slovak population (n = 750) where was 0.827 for PID.D and 0.738 for PID-I (Ballová Mikušková et al., 2014). Levels of Cronbach’s alpha in this study are comparable to other cross- cultural backgrounds, where Betsch (2004) found .77 for PID intuition and .79 for PID deliberation for 2132 Dutch participants. Richetin, Perugini, Adjali, Hurling (2007) found in a sample of 299 British students lower levels .62 for PID intuition and .77 PID deliberation. The best results for internal consistency was found by Witteman et al. (2009) who found .87 for PID Intuition and .85 for PID deliberation in a group of 405 Dutch students.” (Hanák, 2014)

8 SKÚMANIE VZÁJOMNÝCH VZŤAHOV MEDZI PREMENNÝMI

V predchádzajúcich kapitolách sme sa zaoberali jednou premennou (jednorozmerná štatistická analýza), no vo vedeckých či praktických výskumoch nás zvyčajne zaujíma viacero premenných. Tieto premenné neexistujú izolovane a nezávisle od ostatných, v skutočnosti nájdeme medzi nimi rôzne vzťahy, neraz veľmi komplikované (Pacáková et. al., 2009). Preto aj meranie vzťahov medzi dvoma (dvojrozmerná štatistická analýza) alebo viacerými premennými (viacrozmerná štatistická analýza) je jedným z najčastejšie aplikovaných štatistických postupov. Pri tejto analýze skúmame ako sa mení jedna premenná vo vzťahu k druhej premennej.

Na skúmanie vzájomných vzťahov medzi premennými máme k dispozícii viacero štatistických metód. Najčastejšie sa v literatúre uvádza korelácia a regresia. Tie však môžeme aplikovať len na intervalové premenné. Avšak to, akú metódu zvolíme, závisí v prvom rade od charakteru premenných, s ktorými pracujeme. Iné štatistické metódy používame ak sú premenné nominálne, iné ak sú ordinálne a zas iné ak sú intervalové. Čiže charakter premenných určuje možnosti a typ štatistickej analýzy. Z hľadiska prehľadnosti a jednoduchosti pre čitateľa som sa rozhodol členiť nasledovné kapitoly podľa charakteru premenných a nie podľa príbuznosti štatistických metód, softvérovej ponuky alebo iných kritérií.

8.1 Skúmanie vzájomných vzťahov u nominálnych premenných

Nominálne premenné sú pomerne „hrubé“ a možnosti ich skúmania sú ďaleko viac obmedzené než pri ostatných typoch škál merania premenných (Field, 2013, Mareš, et. al. 2015, str. 268). Niekedy sa vzťahy medzi nominálnymi premennými zobrazujú pomocou kontingenčných tabuliek (Crosstabs). Program PSPP nám pri umožňuje vypočítať vzťahy medzi nominálnymi premennými viacerými štatistickými metódami.

Kontingenčná tabuľka (Crosstabs)

Je tabuľkové zobrazenie vzťahu dvoch premenných založené na početnosti výskytu. Vieme ju použiť nielen na nominálne ale aj na ordinálne premenné. Čím viac má premenná kategórií, tým je tabuľka väčšia, menej prehľadná a zároveň v každej kategórii je menej jednotiek. Toto tabuľkové zobrazenie sa používa na vizuálne identifikovanie vzťahov medzi premennými. Kontingenčná tabuľka je v mnohých dole uvedených štatistických metódach súčasťou výsledkovej tabuľky. Treba však poznamenať, že je to skôr grafická metóda, podobne ako graf závislostí, ktorá na rozdiel od dole uvedených štatistických metód neprodukuje určitý koeficient, hodnotu, či výsledok na základe ktorého vieme z dát robiť konkrétne štatistické úsudky.

Koeficient Phi (phi coefficient - ϕ)

Je štatistická metóda, ktorou vieme merať vzťah medzi dvoma dichotomickými premennými. Napríklad pohlavie (muž - žena) a fajčenie (áno - nie). Výsledok nadobúda hodnoty v intervale 0 až 1. Ak by jedna z premenných mala viac než dve kategórie, potom by mohol byť výsledok mimo intervalu 0 až 1, a preto sa tomto prípade používa Cramerovo V (Field, 2014, str. 725). Ritomský a Hanks (1994, str. 138) uvádzajú, že Phi koeficient je špeciálny prípad Cramerovho V pre kontingenčnú tabuľku v kategórii 2*2. Softvér PSPP používa skratku pre tento koeficient PHI.

Koeficient kontingencie (Coefficient of contingency)

Meria vzájomnú asociáciu medzi premennými. Mareš a kolektív (2015) uvádzajú, že hodnota tohto koeficientu príliš závisí od počtu riadkov stĺpcov a ani pri perfektnej závislosti nedosahuje adekvátnu hodnotu 1. Preto neodporúčajú ho používať. Každopádne softvér PSPP ho má v ponuke a označuje ho skratkou CC.

Cramérovo V (Cramér's V)

Meria vzájomný vzťah medzi dvoma premennými z ktorých je jedna dichotomická a druhá nadobúda vyšší počet kategórií než dve. Môžeme ho použiť aj v prípade ak obidve premenné majú vyšší počet kategórií než dve (Mareš a kol., 2015). Výsledok nadobúda hodnoty v intervale 0 až 1, kde 0 žiadny vzťah a 1 úplný vzťah (Hendl, 2009) Ak použijeme Cramérovo V prípade dvoch dichotomických premenných, potom výsledok bude totožný s koeficientom Phi. Softvér PSPP neobsahuje samostatne uvedenú skratku pre Cramérovo V, avšak pri výpočte koeficientu Phi, softvér vypočíta aj Cramérovo V a uvedie ho vo výsledkovom okne. Budíková, Králová a Maroš (2010, str. 214) uvádzajú nasledovné interpretačné intervaly pre koeficient Cramérovo V.

Hodnoty medzi

0 až 0,1 znamenajú zanedbateľnú závislosť

0,1 až 0,3 slabá závislosť

0,3 až 0,7 stredná závislosť

0,7 až 1 silná závislosť.

Goodmanova – Kruskalova lambda (Goodman and Kruskal's lambda - λ)

Používa sa pri dvoch premenných u ktorých vieme určiť predpokladanú nezávislú a závislú premennú. Koeficient lambda nám vyjadruje proporcionálnu redukciu chyby pri meraní závislej premennej. Ritomský a Hanks (1994, str. 138) ju definuje ako: „relatívne zníženie pravdepodobnosti chybnnej predikcie jednej premennej pri znalosti hodnoty druhej premennej.“ Nadobúda hodnoty v intervale 0 až 1. Nula znamená že nezávislá premenná nie je v žiadnom vzťahu k závislej. 1 znamená že ju predpovedná jednoznačne. Napríklad hodnota $\lambda = 0,27$

znamená, že sme zredukovali chybu o 27% pri predpovedaní závislej premennej (Mareš a kol., 2015). Softvér PSPP používa skratku pre tento koeficient LAMBDA.

Eta koeficient (Eta coefficient)

Je koeficient, ktorý meria vzťahy medzi nominálnou premennou a intervalovou. Preto by sme ho mohli zaradiť aj do kapitoly o meraní vzťahov pre intervalové premenné. Pri porovnávaní skupín pomocou t-testov, ANOVY alebo stredných hodnôt, by sme tento koeficient mali uvádzať vo výsledkoch tiež. Nadobúda hodnoty od 0 do 1 a softvér PSPP používa skratku pre tento koeficient ETA. Jeho výhodou je, že dokáže zachytiť aj nelineárny vzťah medzi premennými. Druhá mocnina koeficientu Eta sa označuje η^2 udáva aký podiel variability intervalovej premennej vysvetľuje nominálna premenná Rimančík (2007). V tomto je zhodná s koeficientom determinácie R^2 , ktorý si vysvetlíme pri reportovaní výsledkov regresnej analýzy.

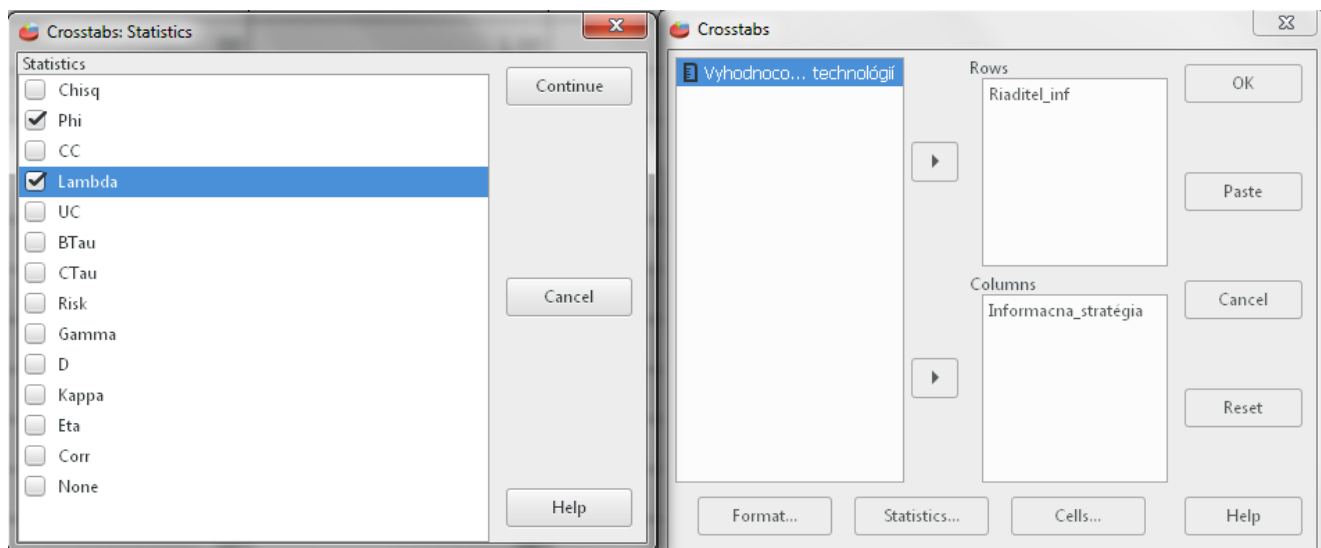
1 úloha: Zistite či existuje vzájomná závislosť medzi tým existenciou pracovnej pozície riaditeľ informatiky a existenciou vypracovanej informačnej stratégie podniku.

Riešenie: Dáta sú uvedené v príklade č. 3, v súbore s názvom Riaditeľ IT.sav. V tomto súbore máme 225 skutočných podnikov v SR, ktorí v roku 2014 uviedli či majú pozíciu riaditeľa informatiky (áno = 1, nie = 0) a či majú vypracovanú informačnú stratégiu podniku (áno = 1, nie = 0). Pretože obe premenné sú dichotomické, tak musíme na vypočítanie vzájomnej súvislosti medzi nimi použiť koeficient Phi. Zároveň predpokladáme, že medzi premenými existuje príčinná súvislosť, konkrétne, že ak existuje pracovná pozícia riaditeľa informatiky, tak ten potom vypracuje informačnú stratégiu podniku. Preto použijeme aj Goodmanovu – Kruskalovu lambdu ako druhú štatistickú metódu na overenie našich hypotéz. Tie si stanovíme nasledovne:

H_1 : Existencia pracovnej pozície riaditeľa informatiky priamo pozitívne vplýva na vypracovanie informačnej stratégie podniku.

H_0 : Existencia pracovnej pozície riaditeľa informatiky nevplýva na vypracovanie informačnej stratégie podniku.

Klikneme (Analyze), potom Opisné štatistiky (Descriptive Statistics) a Krížové tabuľky (Crosstabs). Vyberieme nasledovné dve premenné: do okna Rows presunieme premennú Riaditel_inf a do okna Columns premennú Informacna_strategia. V dialógovom okne si dole otvoríme Štatistiky (Statistics) a vyberieme si Koeficient Phi - Phi a Goodmanovu – Kruskalovu lambdu - Lambda. Potom klikneme Ok.



Obr. 8.1.1 Sprievodca pre koeficient chí a Goodmanovu – Kruskalovu lambdou

Výsledky: Program PSPP vytvorí 4 tabuľky, prvá je sumárny počet analyzovaných prípadov, druhá sú konkrétne početnosti pre jednotlivé skupiny. Druhá tabuľka je kontingenčná tabuľka (Crosstabs). V tejto tabuľke máme uvedené konkrétne početnosti, ich percentuálne vyjadrenie v stĺpcoch (columns), riadkoch (rows) i sumárne (total).

Summary.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Existencia pozície riaditeľa informačných technológií * Existencia informačnej stratégie v podniku	225	100,0%	0	0,0%	225	100,0%

Existencia pozície riaditeľa informačných technológií * Existencia informačnej stratégie v podniku [count, row %, column %, total %].

Existencia pozície riaditeľa informačných technológií	Existencia informačnej stratégie v podniku		Total
	nie je info stratégia	je info stratégia	
nie je riaditeľ	86,00 75,44%	28,00 24,56%	114,00 100,00%
je riaditeľ	81,13% 38,22%	23,53% 12,44%	50,67% 50,67%
Total	20,00 18,02% 18,87% 8,89%	91,00 81,98% 76,47% 40,44%	111,00 100,00% 49,33% 49,33%
	106,00 47,11% 100,00% 47,11%	119,00 52,89% 100,00% 52,89%	225,00 100,00% 100,00% 100,00%

Obr. 8.1.2 Kontingenčná tabuľka ako súčasť výstupu úlohy č. 1.

Ďalej si vysvetlíme čo obsahuje tretia a štvrtá tabuľka. V tretej tabuľke je uvedený koeficient $\Phi = 0,58$ a Cramerovo V , ktoré je tiež rovné $0,58$. Obidva koeficienty sú zhodné pretože obe skúmané premenné boli dichotomické. Vo štvrtej tabuľke s názvom Directional measures je uvedený výsledok pre Goodmanovu – Kruskalovu lambdu. Ak informačná stratégia závisí (je závislá premenná – dependent variable) od pozície riaditeľa, potom je lambda rovná $0,55$ a je

štatisticky vysoko významná $p < 0,001$. Na základe týchto výsledkov zamietame nulovú hypotézu H_0 a prijímame alternatívnu hypotézu H_1 .

Symmetric measures.

Category	Statistic	Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Nominal by Nominal	Phi	,58			
	Cramer's V	,58			
N of Valid Cases		225			

Directional measures.

Category	Statistic	Type	Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Nominal by Nominal	Lambda	Symmetric	,56	,06	8,09	,000
		Existencia pozície riaditeľa informačných technológií Dependent	,57	,06	6,26	,000
		Existencia informačnej stratégie v podniku Dependent	,55	,07	5,83	,000
	Goodman and Kruskal tau	Existencia pozície riaditeľa informačných technológií Dependent	,33			
		Existencia informačnej stratégie v podniku Dependent	,33			

Obr. 8.1.3 Vzájomná závislosť medzi pozíciou riaditeľa informatiky a existenciou informačnej stratégie v podniku počítanej pomocou koeficientu Phi a Goodmanovou – Kruskalovou lambdou.

Uvádzanie výsledkov: Vytvorenie pozície riaditeľa informatiky úzko súvisí s tým či podnik má vypracovanú informačnú stratégiu, alebo nie, koeficient Phi = 0,58, Cramerovo V = 0,58. Na základe výsledkov konštatujeme, že ak vymenujeme riaditeľa informatiky (nezávislá premenná), ten potom vytvorí informačnú stratégiu podniku (závislá premenná) Goodmanovou – Kruskalovou lambdu, kde $\lambda = 0,55$, $p < 0,001$.

8.2 Skúmanie vzájomných vzťahov u ordinálnych premenných

Sú poradové premenné a na meranie vzťahov medzi premennými môžeme použiť dole uvedené štatistické metódy, ktoré sú neparametrické štatistické metódy. Ich princíp je založený na vzájomnom porovnávaní hodnôt premenných medzi jednotlivými jednotkami v súbore a odlišujú sa v tom ktoré páry jednotiek berú do úvahy pri výpočte. Porovná sa respondent č. 1 s respondentom č. 2, potom respondent č. 3 s respondentom č. 4, atď. Vzájomné porovnanie hodnôt ich odpovedí môže patriť do piatich možných kategórií.

Konkordantné, súhlasné (Concordant), keď hodnoty oboch premenných sú vyššie alebo nižšie u druhého respondenta (druhého prípadu).

Diskordantné, nesúhlasné (Discordant) keď hodnota jednej premennej je vyššia alebo nižšia u druhého respondenta (druhého prípadu) a u druhej premennej je tomu presne naopak.

Spriahnuté poradia keď hodnoty premennej sú u oboch respondentov rovnaké a tak môžu nastať ďalšie tri prípady

Spriahnuté poradia (Tied ranks) v prvej premennej, spriahnuté poradia v druhej premennej, spriahnuté v oboch premenných.

Na lepšie vysvetlenie si uvedieme konkrétny príklad. V dátovom súbore č.1 sú uvedené odpovede respondentov na všetky otázky dotazníka PID ako aj sumáre skóre testu kognitívnej reflexie – CRT. Vybrali sme 4 respondentov od

poradového čísla 29 až 31 a uvádzame aj konkrétne hodnoty pre dve skúmané premenné.

Tabuľka č. 1. Výber dvoch premenných u respondentov 29 až 32 z príkladu č.1.

Číslo respondenta	PID škála uvažovanie otázka z dotazníka č. 1	Test kognitívnej reflexie – CRT, súčet všetkých 3 otázok
29	5	0
30	5	2
31	3	0
32	4	0

V príklade, ktorý uvedieme budeme porovnávať 29-teho respondenta s 30-tym respondentom, atď. Pozrime sa hodnoty otázky č.1 a skóre CRT u 29 respondenta, teda 5 a 0, u 30-teho je to 5 a 2. V prvej premennej (otázka č. 1) sú hodnoty premennej zhodné (obidve 5) u druhej je nárast z 0 na 2. Sú teda spriahnuté v prvej premennej. U respondenta 30 a 31 sú konkordantné. U 31 prvého a 32 druhého respondenta sú spriahnuté v druhej premennej.

V nasledovných štatistických metódach uvedieme ktoré z piatich možných kombinácií sú používané pri výpočte.

Goodmanova – Kruskalova gamma (Goodman and Kruskal's gamma - G)

Je založená na porovnávaní súhlasných a nesúhlasných párov a vypočítame ju nasledovne:

$$(KP - DP) / (KP + DP)$$

Kde : KP – Počet konkordantných párov, DP - Počet diskordantných párov

Mareš a kolektív (2015) uvádzajú, že všetky páry so spriahnutým poradím sú z menovateľa vylúčené. Nadobúda hodnoty od - 1 až 1 Softvér PSPP používa skratku pre tento koeficient GAMMA

Kenadallovo tau b (Kendall's tau b - τ)

Počíta sa na rovnakom princípe ako hore uvedená gamma, no z menovateľa nevyklučuje páry so spriahnutým poradím. PSPP používa skratku pre tento koeficient BTAU.

Kenadallovo tau c (Kendall's tau c - τ)

Výpočet je rovnaký ako v prípade Kenadallovo tau b a v menovateli je aj počet jednotiek vo výbere. Výsledok u obidvoch tau nadobúda hodnoty v intervale -1 až 0, resp. 0 až 1. PSPP používa skratku pre tento koeficient CTAU. Kenadallovo tau umožňuje jednoduchú pravdepodobnostnú interpretáciu na rozdiel od Pearsonovho alebo Spearmanovho koeficientu (Hendl, 2009, str. 271)

Somersove d (Somers' D)

Somersove d nám umožňuje pracovať s nezávislou a závislou premennou. Teda nielen skúmame vzťah medzi premennými, vieme aj overiť jeho príčinnú

závislosť. Inými slovami či jedna premenná priamo ovplyvňuje druhú premennú. Základný vzorec na výpočet je rovnaký ako pri *Gamme*, avšak v menovateli sa vynechávajú všetky prípady, ktoré majú spriahnuté poradie u nezávislej premennej. PSPP používa skratku pre tento koeficient D.

Spearmanovo rho (Spearman correlation)

Je korelačný koeficient, ktorý pracuje s poradím premenných. Rovnako ako predchádzajúce štatistické metódy, je to koeficient neparametrický, tak ho môžeme aplikovať aj na nie normálne distribuované dáta, čo prakticky znamená, že je odolný voči extrémnym hodnotám (outliers) a taktiež dokáže zachytiť aj nie lineárny vzťah medzi premennými. Nadobúda hodnoty -1 až 0 alebo 0 až 1. Rovnako ako pri iných mierach asociácie platí, že čím bližšie je koeficient k 1, tým je silnejšia priama poradová závislosť medzi premennými X a Y (Budíková a kol., 2010, str. 220). PSPP používa skratku pre tento koeficient CORR.

Cohenova Kappa (Cohen's Kappa)

Je štatistická metóda, ktorá sa používa na meranie vzájomnej zhody. Tento koeficient sa v minulosti a aj v súčasnosti pomerne často používa pri posudzovaní zhody expertov pri riešení rôznych úloh (Hanák, 2014). Pri výpočte berie do úvahy aj možnosť náhodnej zhody medzi hodnotiteľmi. Nadobúda hodnoty od -1 až +1. PSPP používa skratku pre tento koeficient KAPPA.

Uviedli sme viaceré metódy na výpočet koeficientov merajúcich vzťahy medzi ordinálnymi premennými. Ak by sme použili na rovnaký príklad všetky, potom dosiahneme štyri rôzne koeficienty a vyvstáva nám otázka ktorý z nich uviesť a prečo. Keď poznáme veľmi dôkladne charakter našich dát a spôsob ako jednotlivé metódy počítajú výsledok, potom dokážeme vyargumentovať prečo sme konkrétnu metódu použili. Ak si chceme zjednodušiť prácu, a stačí nám jednoduchšie a povrchnejšie riešenie problému, potom použijeme pravidlo, ktoré hovorí, že uvádzame ten výsledok, ktorý má najnižšie hodnoty zo všetkých.

Úloha 2: V súbore s názvom Príklad č.1 respondenti riešili logickú úlohu s názvom kognitívna reflexná úloha CRT a vyplňali dotazník o intuícii a uvažovaní (deliberácii) PID (Betsch, 2004). V úlohe CRT sú tri logické „chytáky“, ktoré vyžadujú dôkladne si prečítať a premyslieť odpoveď, inak sa respondent nechá zvyčajne nachytať a odpovie nesprávne. Respondent môže mať úspešnosť 0 – nevyriešil správne nič až 3 – vyriešil všetko. V seba-výpoved'ovom dotazníku PID odpovedá respondent na otázky ohľadom jeho intuície a racionality. Otázka č. 1 znie takto: „Predtým než spravím rozhodnutie, najprv si ho premyslím.“ Odpoveď je na 5 stupňovej Likertovej škále, 0 veľmi nesúhlasím a 5 veľmi súhlasím. Zaujímalo nás či odpoveď na túto konkrétnu otázku z dotazníku PID súvisí s úspešnosťou riešenia CRT. Predpokladáme, že tí, ktorí veľmi premýšľajú svoje rozhodnutia, dobre premyslia aj logické „chytáky“

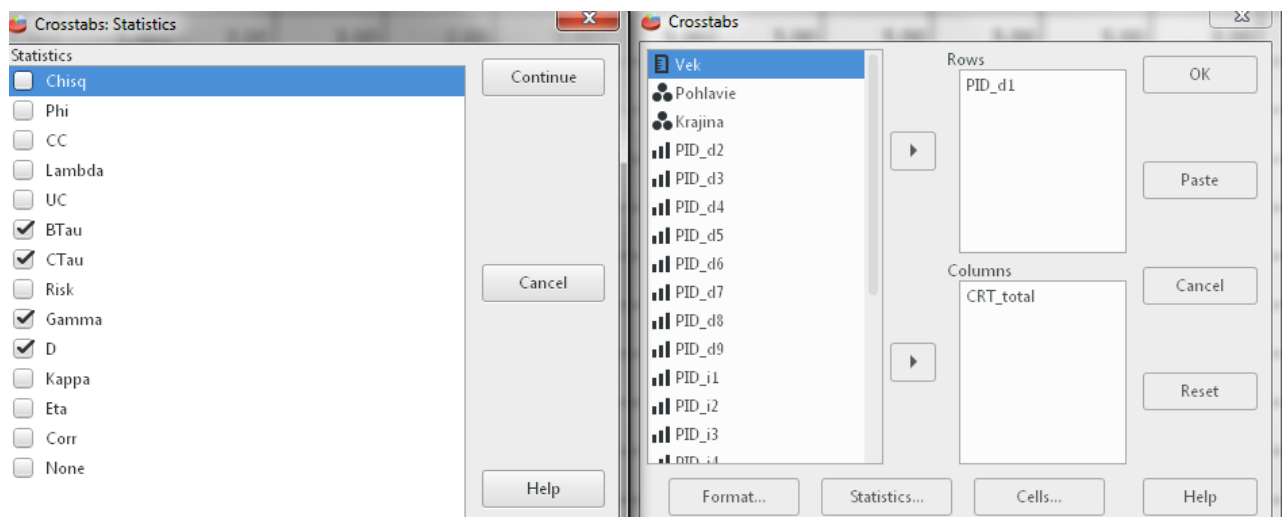
a nedajú sa nachytať. Konkrétne, že vysoké skóre v premýšľaní o rozhodnutí, bude znamenať viac správnych odpovedí na logické „chytáky“. Na základe týchto predpokladov sme si stanovili hypotézy nasledovne:

H₁: Odpovede na otázku č. 1 zo subškály PID uvažovanie: „Predtým než spravím rozhodnutie, najprv si ho premyslím“ budú v pozitívnom vzťahu k úspešnosti riešenia logickej kognitívno – reflexnej úlohe, CRT.

H₀: Odpovede na otázku č. 1 zo subškály PID uvažovanie: „Predtým než spravím rozhodnutie, najprv si ho premyslím“ nie sú vo vzťahu k úspešnosti riešenia logickej kognitívno – reflexnej úlohe, CRT.

Riešenie: Otázka z dotazníka PID je ordinálna premenná, ktorá má 5 stupňov. Riešenie logických chytákov v tomto prípade môžeme definovať tiež ako ordinálnu premennú, ktorá má 4 stupne (0 až 3). Na analýzu vzájomných závislostí použijeme Goodmanovu – Kruskalovu gammu, Kenadallovo tau b, Kenadallovo tau c aj Somersove d.

Klikneme Analyzovať (Analyze), potom Opisné štatistiky (Descriptive Statistics) a Krížové tabuľky (Crosstabs). Vyberieme nasledovné dve premenné: do okna Rows presunieme premennú PID_d1 a do okna Columns premennú CRT_total. V dialógovom okne si dole otvoríme Štatistiky (Statistics) a vyberieme si: Goodmanova – Kruskalovu gammu , Kenadallovo tau b, Kenadallovo tau c aj Somersove d . Potom klikneme Ok.



Obr. 8.2.1 Sprievodca pre tieto štatistické metódy: Goodmanova – Kruskalovu gammu , Kenadallovo tau b, Kenadallovo tau c aj Somersove d

Výsledok: V našom prípade sme dosiahli hodnoty -0,03, ktoré sú veľmi blízke nule. To znamená, že to, ako sami seba označili pri premýšľaní pri svojom rozhodovaní, nesúvisí vôbec s tým, ako úspešne naozaj riešili skutočné logické

úlohy. Koeficient je dokonca záporný, čo by pri vyššom zápornom skóre mohlo znamenať že vysoké skóre v premýšľaní o svojom rozhodnutí, znamená nízku úspešnosť riešených úloh. Avšak koeficient je tak blízko hodnote nula, že najskôr znamená, že skúmanými premennými neexistuje vzájomná asociácia. Sommersove d umožňuje overiť si vzťahy medzi premennými vo forme závislej a nezávislej premennej. Skóre v v CRT úlohách (CRT_total) je závislá premenná (Dependent) a odpoveď na otázku č. 1 je nezávislá. Očakávali sme, že premýšľanie pred prijatím rozhodnutia bude pozitívne ovplyvňovať úspešnosť pri riešení „chytákov“ CRT. No Sommersove $d = -0,03$, $p = 0,457$ znamená skutočnú neexistenciu vzťahu. Preto prijímame nulovú hypotézu H_0 .

Symmetric measures.

Category	Statistic	Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-b	-,03	,04	-,74	
	Kendall's tau-c	-,03	,03	-,74	
	Gamma	-,05	,06	-,74	
N of Valid Cases		446			

Directional measures.

Category	Statistic	Type	Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal by Ordinal	Somers' d	Symmetric	-,03		-,74	,457
		PID_d1 Dependent	-,03	,04	-,74	,457
		CRT_total Dependent	-,03	,04	-,74	,457

Obr. 8.2.2 Výsledok analýzy závislostí pre ordinálne premenné.

Uvádzanie výsledku: Odpoveď na otázku č. 1 z dotazníka PID, subškála deliberácia : „Predtým než spravím rozhodnutie, najprv si ho premyslím“ nesúvisí s úspešnosťou riešenia kognitívne reflexných úloh CRT, Kendallovou tau – b = -0,03, v prípade CRT uvedenej ako závislá premenná je vzťah neexistujúci a štatisticky nevýznamný, Somerovo d = -0,03, p = 0,46.

8.3 Skúmanie vzájomných vzťahov u intervalových premenných

Na skúmanie vzájomných závislostí u intervalových premenných používame najznámejšiu parametrickú metódu nazvanú korelácia.

8.3.1 Korelácia

Korelácia je štatistická metóda, pomocou ktorej môžeme vypočítať mieru vzájomnej lineárnej závislosti medzi dvomi premennými (Field, 2013). Inými slovami môžeme aj povedať, že: „ hodnoty jednej premennej majú tendenciu sa vyskytovať spoločne s hodnotami inej premennej“ (Hendl, 2009, str. 250).

Výsledkom výpočtu štatistickej metódy korelácie je korelačný koeficient označovaný malým písmenom r, ktorý môže nadobúdať len hodnoty od -1 do 0

alebo od 0 do +1. Mínus 1 znamená absolútnu nepriamu závislosť (premenné sú perfektne negatívne korelované), 0 znamená žiadnu, neexistujúcu lineárnu závislosť a 1 znamená absolútnu priamu závislosť (premenné sú perfektne pozitívne korelované) medzi dvoma premennými. Ak korelácia nadobúda hodnotu 1 alebo -1, tak z hodnôt jednej premennej potom vieme presne určiť hodnotu druhej premennej (Martin, Bateson, 2009, str. 140). Teda stačí ak poznáme jednu hodnotu.

Príkladom korelácie, kde koeficient r je kladný, môže byť napríklad školenie zamestnancov ohľadom zvyšovania kvality výroby. Teda čím viac sa zamestnanci školili a vzdelávali, tým viac rástol počet kvalitne spracovaných produktov. Nepriamy korelačný vzťah by mohol byť medzi hodinami strávenými školením o zvyšovaní kvality a poklesom počtu nepodarkov. Teda čím viac hodín sa vzdelávali v zvyšovaní kvality, tým viac klesal počet nepodarkov.

V sociálnom a ekonomickom výskume korelačné koeficienty takmer nikdy nenadobúdajú hodnoty blízke alebo rovné 1, ale sa k nim len viac, či menej, blížia, a korelačný koeficient dosahuje akékoľvek hodnoty v rámci intervalu. Z výšky korelačného koeficientu koštatujeme, ako úzko spolu dve premenné súvisia. S rastúcou hodnotou korelačného koeficientu smerom k 1, resp. -1, rastie aj testnosť väzby medzi premennými (Brož, Bezvoda, 2008, str. 319). Stanovenie intenzity vzájomnej závislosti na základe korelačného koeficientu, teda, či je vysoký, mierny, alebo nízky, závisí od konkrétnej vednej oblasti a predmetu skúmania. V prírodných vedách (napr. fyzike) sa korelačné koeficienty môžu blížiti alebo dokonca dosahujú hodnoty ako 1, alebo -1. Vo finančnom manažmente napríklad medzi niektorými finančnými ukazovateľmi dosiahneme vyššie hodnoty korelačného koeficientu než napr. v sociálnych manažérskych výskumoch týkajúcich sa motivácie alebo spokojnosti. Preto pokým v jednej vedeckej alebo odbornej oblasti (finančný manažment, materiálové technológie, technika) môže byť rovnaký korelačný koeficient považovaný za priemerný, v inej (psychológia, sociálna práca, pedagogika, manažment, marketingové výskumy a pod.) je vnímaný ako vysoký.

Chajdiak (2009, str. 79-80) uvádza hodnoty koeficientu korelácie 0,8 až 1 (-0,8 až -1) ako silné, teda medzi premennými existuje silná vzájomná závislosť. Hodnoty 0,4 až 0,8 (-0,4 až -0,8) sú stredne silné a od 0 do 0,4 (-0,4 až 0) sú považované za slabé.

Tabuľka č. 8.3 Cohen (1988, in: Rimančík (2007), str. 73) uvádza hodnoty korelačných koeficientov a ich interpretáciu pre psychológiu.

Číselná veľkosť koeficientu	Sila vzájomnej závislosti
0 až 0,1	triviálna
0,1 až 0,29	Malá
0,3 až 0,49	Stredná
0,5 až 0,69	Veľká
0,7 až 0,89	Veľmi veľká
0,9 až 1	Takmer dokonalá

V štatistickom softvéri sa korelácia pomerne jednoducho počíta a tiež ľahko interpretuje, avšak táto metóda nemôže byť používaná v niektorých konkrétnych prípadoch, pretože jej uplatnenie má svoje obmedzenia. Konkrétne pri korelácii si musíme byť vedomí nasledovných obmedzení:

1. Korelácia neznamená príčinnú súvislosť. Teda ak dve premenné spolu korelujú (súvisia) ešte to neznamená, že jedna priamo ovplyvňuje druhú. A to platí bez ohľadu toho aký silný je korelačný koeficient. Napríklad ak v medzinárodných provnaniach pomocou korelácie zistíme, že výška finančného príjmu koreluje s obezitou, určite to neznamená, že čím viac sa nám bude zvyšovať náš osobný príjem, tým priamo lineárne budeme aj samozrejme viac priberať.
2. Naopak ak premenné sú vo vzájomnom príčinnom vzťahu, korelácie existuje, no z korelačného vzťahu nevieme určiť, ktorá je závislá a nezávislá premenná. Teda ktorá premenná ovplyvňuje, a ktorá je ovplyvňovaná. V tomto prípade je dôležité logicky uvažovať a argumentovať, ktorá premenná mohla ovplyvniť druhú. Môžeme si vytvoriť obidve varianty a potom logicky skúmať, ktorá z nich je pravdivá a ktorá nie.
3. Korelácia meria lineárny vzťah a iné typy vzťahov nezachytí a tak sa nám premenné budú javiť ako nesúvisiace aj keď existuje medzi nimi vzájomný vzťah. Napríklad ak je medzi premenými exponenciálny, logaritmický alebo iný typ vzťahu, korelácia to nedokáže zachytiť.
4. Korelácia je veľmi ovplyvniteľná odľahlými hodnotami. To znamená, že niekoľko jednotiek v súbore, ktoré nadobúdajú extrémne hodnoty významne ovplyvní celý korelačný koeficient (Field, 2013).

Koreláciu dvoch premenných môžeme počítať pomocou Spearmanovho alebo Pearsonovho korelačného testu v závislosti od toho akú formu distribúcie dát má naša vzorka. Parametrické testy štatistickej významnosti korelačného koeficientu používame vtedy ak majú premenné normálne rozdelenie. V tomto prípade používame **Pearsonov korelačný koeficient**. Ten je uvedený v programe PSPP v menu Analyzovať ako Bivariate Correlation.

V prípade, že sme rozdelenie našich dát nie je normálne, alebo ho nevieme overiť, musíme používať neparametrické testy korelačného koeficientu (Pacáková, 2015, str. 220), ktoré sme si uviedli v predchádzajúcej kapitole. V psychologickom výskume sa dlhodobo vedie diskusia ohľadom toho aký typ korelačnej analýzy použiť na ordinálne dáta (veľmi súhlasím, čiastočne súhlasím, súhlasím atď.). Podľa správnosti by sa mala používať Spearmanova korelácia, no dokonca aj v najrenomovanejších psychologických časopisoch na svete sa bežne používa na tento typ dát Pearsonov korelačný koeficient.

My sa budeme snažiť podľa správnosti najprv test na overenie ako sú rozložené dáta a potom podľa výsledkov tohto testu použijeme v prípade normálne rozložených dát Pearsonov korelačný koeficient, alebo ak nie sú rozložené normálne potom Spearmanov korelačný koeficient.

1 úloha: V databázovom súbore č. 2 *Študenti body a pamäť* máme skutočné výsledky z testovania 22 ročných vysokoškolských študentov manažmentu. Títo v priebehu semestra riešili viacero rôznych úloh (posudzovanie podnikateľského plánu v časovom strese a bez časového stresu, pamäťové úlohy) a taktiež máme ich body z teoretickej skúšky a z rôznych semestrálnych projektov. Zistite pomocou korelačnej analýzy, či existuje medzi bodmi dosiahnutými zo softvéru o jednoduchom účtovníctve a bodmi zo softvéru o podvojných účtovníctve vzájomná závislosť.

Riešenie: Študenti riešili 2 rôzne projekty a prvým projektom bol zvládnutie účtovníckeho softvéru pre jednoduché účtovníctvo (študenti mali body za správne vytvorenie komplexného príkladu o 10 rôznych operáciách). Body sú uvedené pod názvom `Body_jednod_ucto`. Druhou premennou sú body za prípadovú štúdiu zo softvéru spracúvajúceho podvojných účtovníctvo a sú uvedené pod názvom `Body_podvojn_ucto`. Všetky uvedené projekty sú kódované ako intervalové premenné. Teda študent z nich mohol získať 0 až maximálny počet bodov, ktorý bol 5 pre jednoduché účtovníctvo a 7 pre podvojných účtovníctvo. Pri riešení príkladu budeme postupovať nasledovne.

1. Krok: Overíme si normalitu rozloženia dát.

Ako už bolo hore spomenuté nemôžeme spraviť ihneď korelačný test, musíme najprv zistiť ako sú rozložené dáta pre jednotlivé premenné. Preto ako prvý vykonáme test normality rozloženia dát. Z viacerých dostupných testov použijeme Kolmogorov-Smirnov test.

2. Krok Podľa typu rozloženia dát použijeme konkrétny korelačný test. Ak budú dáta rozložené normálne, potom použijeme Pearsonov korelačný test. Ak dáta nebudú rozložené normálne, použijeme Spearmanov korelačný test.

Budeme porovnávať počet dosiahnutých bodov z prípadovej štúdie v softvéri o jednoduchom účtovníctve s počtom bodov v softvéri o podvojných účtovníctve.

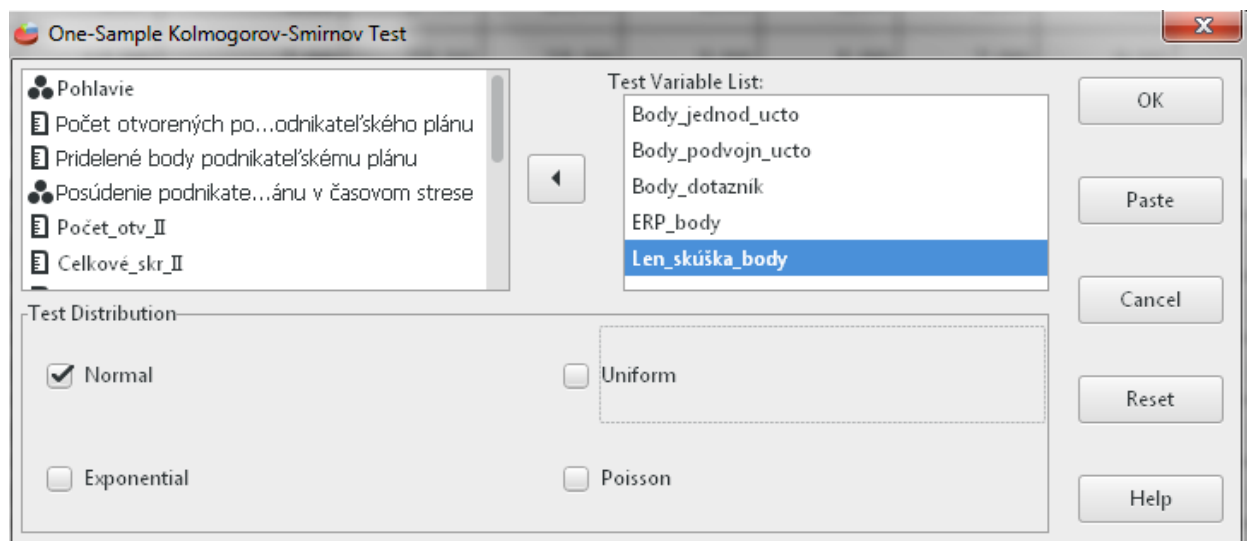
Predpokladáme, že tieto premenné spolu veľmi úzko súvisia, teda ak dosiahne respondent vysoký počet bodov z prípadovej štúdie jednoduchého účtovníctva, taktiež dosiahne aj vysoký počet bodov z podvojného účtovníctva. Náš predpoklad je:: Počet bodov za prípadovú štúdiu z jednoduchého účtovníctva súvisí s počtom bodov za prípadovú štúdiu z podvojného účtovníctva. Pred realizovaním korelačných analýz je vhodné vopred ujasniť si, aké máme predpoklady, aké výsledky očakávame a tie exaktne definovať prostredníctvom hypotéz. Preto si stanovíme tieto hypotézy.

H_1 Medzi premennými Body_jednod_ucto a Body_podvojn_ucto existuje vzájomná závislosť meraná korelačným koeficientom.

H_0 : Medzi premennými Body_jednod_ucto a Body_podvojn_ucto neexistuje vzájomná závislosť meraná korelačným koeficientom.

1. Krok: Overenie normality rozloženia dát.

Na overenie normality rozloženia dát použijeme Kolmogorov Smirnovov test. Klikneme na Analyzovať (Analyze), potom na Neparametrické štatistiky (NonParametric Statistics) a vyberieme 1 výberový K – S (1Sample K - S). Toto je skratka pre Kolmogorov – Smirnovov test. Následne vyberieme premenné a dole zvolíme Test Distribúcie (Test Distribution). V našom prípade vyberieme Normálne rozloženie (Normal). Na záver klikneme OK.



Obr. 8.3.1 Overenie normality rozloženia dát pomoc Kolmogorovho-Smirnovho testu

Výsledok je zobrazený ako tabuľka s viacerými dátami. Nás však zaujíma spodný riadok, v ktorom je uvedená hladina významnosti (*Asymp. Sig. 2 tailed*). Platí, že ak táto hladina je nižšia než 0,05, potom sú výsledky nie normálneho rozloženia štatisticky významné. Čiže ak sú nižšie než 0,05, potom dáta nie sú rozložené normálne. To je náš prípad pre obidve premenné. Takéto nie normálne rozloženie dát je to spôsobené tým, že študenti sa veľmi dobre pripravili na každú časť

hodnotenia a takmer všetci dosahovali plné počty. Inými slovami takmer všetci mali z úlohy jednoduché účtovníctvo maximum, teda 5 bodov, len málo študentov malo 4 body a len minimum z nich 3 body. Nikto nemal 2, 1, alebo 0 bodov. Preto dáta nie sú rozložené normálne a musíme použiť Spearmanov test korelácie.

NPAR TESTS

NPAR TEST

/KOLMOGOROV-SMIRNOV (NORMAL) = Body_jednod_ucto Body_podvojn_ucto Body_dotaznik ERP_body Len_skuška_body

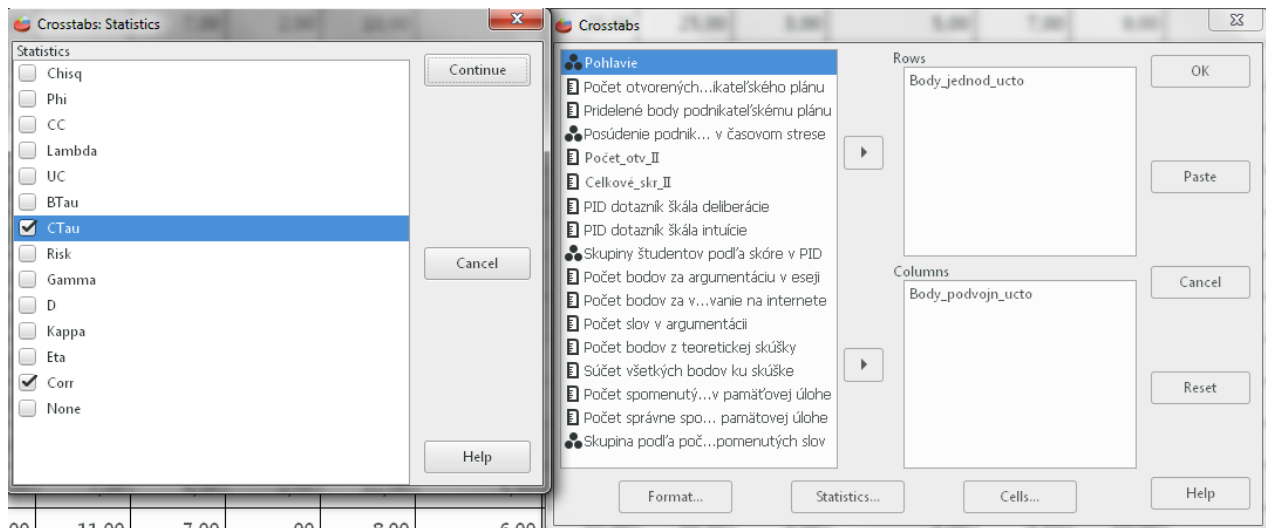
One-Sample Kolmogorov-Smirnov Test

		<i>Body zo softvéru jednoduché účtovníctvo</i>	<i>Body zo softvéru podvojné účtovníctvo</i>
<i>N</i>		75	75
<i>Normal Parameters</i>	<i>Mean</i>	4,93	6,62
	<i>Std. Deviation</i>	,22	,57
<i>Most Extreme Differences</i>	<i>Absolute</i>	,52	,29
	<i>Positive</i>	,38	,25
	<i>Negative</i>	-,52	-,29
<i>Kolmogorov-Smirnov Z</i>		4,54	2,53
<i>Asymp. Sig. (2-tailed)</i>		,000	,000

Obr. 8.3.2 Výsledky testu Kolmogorov-Smirnov Test normality rozloženia dát

2. Krok Výber korelačného testu.

Na základe výsledkov Kolmogorovho-Smirnovho testu vieme, že dáta nie sú rozložené normálne a preto musíme použiť Spearmanovu koreláciu. Táto sa nachádza v analyzovať (Analyze), potom Opisné štatistiky (Descriptive Statistics) a Krížové tabuľky (Crosstabs). Vyberieme nasledovné dve premenné: Body_jednod_ucto a Body_podvojn_ucto. V dialógovom okne si dole otvoríme Štatistiky (Statistics) a vyberieme si, konkrétnu štatistickú metódu, ktorú chceme spustiť. V našom prípade je to Spearmannova korelácia, ktorá je označená ako Corr a aj Kendallovo Tau, označené ako CTau. Klineme OK.



Obr. 8.3.3 Sprievodca Spearmanovou koreláciou a Kendallovým tau.

Výsledok: Softvér PSPP vytvorí tri tabuľky a výsledky Spearmanovho korelačného koeficientu sú uvedené v poslednej, tretej z nich. V našom prípade vidíme, že Spearmanov korelačný koeficient je $r = 0,19$ a hodnota Kendallovho tau je ešte nižšia $0,09$. A je uvedený aj výsledok Pearsonovej korelácie $r = 0,22$.

Symmetric measures.

Category	Statistic	Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal by Ordinal	Kendall's tau-c	,09	,06	1,43	
	Spearman Correlation	,19	,12	1,64	
Interval by Interval	Pearson's R	,22	,13	1,94	
N of Valid Cases		75			

Obr. 8.3.4 Výsledok Spearmanovej Korelácie a Kendallovho tau.

Ako už bolo hore uvedené korelačný koeficient môže nadobúdať hodnoty 0 až 1, resp. -1. V našom prípade je najlepším ukazovateľom Kendallovho tau – c, ktoré dosahuje hodnotu $0,09$, čo je veľmi nízka hodnota blízka nule. Počet bodov z úlohy jednoduché účtovníctvo nesúvisí s počtom bodov z úlohy podvojnú účtovníctvo. Na základe týchto výsledkov zamietame hypotézu H_1 a prijímame nulovú hypotézu H_0 .

Uvádzanie výsledkov: Počet bodov, ktorý študenti získali z prípadovú štúdiu v softvéri o jednoduchom účtovníctve závisí minimálne od počtu bodov, ktoré študenti získali za prípadovú štúdiu v softvéri o podvojnú účtovníctve, Kendallovho tau $c = 0,09$, Spearman $r = 0,19$.

8.3.2 Výber konkrétnej štatistickej metódy

Najlepší prehľadom ohľadom aplikovateľných metód skúmania vzájomných vzťahov medzi premennými s akým som sa v literatúre kedy stretol je prehľad uvedený v tabuľke č. 9.1 od autorov Petra Mareša, Ladislava Rabušica a Petra Soukupa v knihe Analýza sociálnovedných dat (nejen) v SPSS, strana 289. Kompletne prebratý prehľad od nich uvádzam v nasledujúcej tabuľke č. X. Pre záujemcov o štatistiku môžem ich knihu vrele odporučiť.

Tabuľka č 8.3.2 Prehľad mier asociácie a ich charakteristiky (prevzaté z : Mareš et. al. 2015, str. 289, tabuľka č. 9.1)

Úroveň merania	Počet kategórií	Štatistická metóda v PSPP	Vhodný koeficient
1. Nominálna/Nominálna	2 x 2	Crosstabs	Phi, Lambda
2. Nominálna/Nominálna	3+ x 2+	Crosstabs	Cramerovo V, Lambda
3. Nominálna/Ordinalná	3+ x 3+	Crosstabs	Cramerovo V, Lambda
4. Nominálna/Intervalová	Nominálna nezávislá	a) Crosstabs – pokiaľ má intervalová premenná málo kategórií b) Means, ANOVA	a) Eta b) Eta
5. Ordinalná /Ordinalná	Obidve premenné s malým počtom kategórií	Crosstabs	Gama, Kendallove tau b – pre štvorcovú tabuľku, Sommersovo d, Kendallove tau c – pre obdĺžnikovú tabuľku
6. Ordinalná /Ordinalná	Jedna premenná s mnohými kategóriami	Crosstabs	Kendallove tau c
7. Ordinalná /Ordinalná	Obidve premenné s mnohými kategóriami	Crosstabs Poradová korelácia	Kendallove tau c Kendallove tau b Spearmanovo rho

			Poradová korelácia	
8.	Ordinálna /Intervalová	Obidve premenné s niekoľkými kategóriami	a) Crosstabs b) porovnávanie priemerov, pokiaľ je závislá premenná intervalová (ANOVA)	a)Eta a rovnaké koeficienty ako v bode 5 b)Eta
9.	Ordinálna /Intervalová	Ordinálna s niekoľkými kategóriami, intervalová s mnohými	a) Means b) poradová korelácia	Eta Kendallove tau
10.	Ordinálna /Intervalová	Obidve premenné s mnohými kategóriami	poradová korelácia	Kendallove tau Spearmanovo rho
11.	Intervalová /Intervalová		Correalate, regresia alebo bodový graf	Pearsonovo r, Regresia

V nasledujúcich kapitolách si vysvetlíme skúmanie závislostí v tých prípadoch, kde očakávame, že jedna alebo viac premenných priamo ovplyvňujú hodnoty druhej premennej. Skúmaním príčinnej lineárnej závislosti sa zaoberá regresná analýza.

8.4 Lineárna regresná analýza

Korelačná analýza nám povie, či existuje medzi premennými určitá závislosť, avšak nevieme vyčísliť aký konkrétny vzťah medzi premennými existuje. Naznačuje určitú možnosť predpovedania, no len všeobecnú a nie jednoznačnú. Taktiež nevieme pri korelačnej analýze z hodnôt jednej premennej predpovedať hodnoty druhej. **Lineárna regresná analýza** tieto nedostatky v určovaní vzťahov medzi premennými odstraňuje, konkrétne je to štatistická metóda pomocou ktorej môžeme preskúmať príčinný vzťah medzi dvomi alebo viacerými premennými a dokážeme ho exaktne vyčísliť. Dokážeme určiť veľkosť vplyvu a dokonca určiť predpokladané hodnoty aké bude nadobúdať závislá premenná ak poznáme hodnotu nezávislej premennej. Pri regresii predpokladáme, že jedna premenná závisí od druhej premennej. Tá, ktorá závisí, sa volá **Závislá premenná** (dependent variable – DV, cieľová premenná a označujeme ju Y). Tá, ktorá

ovplyvňuje sa nazýva **Nezávislá premenná** (independent variable – IV, označujeme ju X) alebo aj prediktor či regresor. Nezávislých premenných, ktoré ovplyvňujú závislú môže byť, a v praxi zvyčajne aj vždy je, viacero. Čím viac nezávislých premenných, tým presnejšie vieme opísať vývoj závislej premennej. Vzájomný vzťah premenných pri regresnej analýze vieme potom exaktne vyčísliť v nasledovnej rovnici. Pre 1 nezávislú premennú.

$$Y = b_0 + b_1 * X + e$$

Pre viac nezávislých premenných.

$$Y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots b_n * x_n + e$$

Kde:

Y – je závislá premenná, ktorej hodnoty predpovedáme pomocou nezávislej premennej

X - je nezávislá premenná, alebo x_1 až x_n sú nezávislé premenné.

b_0 - je bod, kde regresná priamka pretína os y, zároveň udáva odhadovanú hodnotu závislej premennej Y, ak sa hodnota nezávislej premennej X rovná nule. Nazýva sa aj Lokujúca konštanta (Šoltés, 2015, str. 150).

b_1 je regresný koeficient a určuje smer priamky. Taktiež vyjadruje o koľko sa zvýši odhadovaná hodnota závislej premennej Y ak hodnotu nezávislej premennej X zvýšime o jednu jednotku.

e - je chybová hodnota.

Pred aplikovaním regresnej analýzy je dôležité najprv si dôkladne premyslieť a logicky zdôvodniť ktorá premenná je závislá a ktorá je nezávislá.

Regresnú analýzu však nemôžeme aplikovať na akékoľvek dáta. Existujú určité podmienky a obmedzenia regresnej analýzy.

Lineárna regresná analýza dokáže identifikovať vzťah medzi premennými iba ak je lineárny. Na identifikovanie nelineárneho vzťahu môžeme použiť špecifické softvérové analýzy, program PSPP ich žiaľ neponúka.

Pri lineárnej regresnej analýze musí byť závislá premenná meraná na intervalovej úrovni (Mareš a kol. 2015)

Pri malých súboroch musíme skontrolovať či sú dáta normálne rozložené.

Rozptýlenosť bodov okolo regresnej priamky je rovnaká pre všetky uvažované hodnoty X. Túto podmienku nazývame homoskedascita.

1 úloha: V príklade č. 2 máme študijné výsledky študentov manažmentu z predmetu Manažérska informatika. Premenná s názvom Len_skuška_body obsahuje počet dosiahnutých bodov z teoretickej skúšky. Okrem študijných výsledkov študenti riešili viacero úloh a jedna z nich bola pamäťová úloha, kde sme počas semestra požiadali študentov aby si spomenuli na konkrétny obsah úlohy, ktorú riešili 3 týždne dozadu. Za každú spomenutú konkrétnu charakteristiku, ktorú si dokázali spomenúť, dostali bod. V tomto príklade

vychádzame z teoretického predpokladu, že dobrý výkon v pamäťovej úlohe, bude mať pozitívny efekt na výsledky v teoretickej skúške, založenej na memorovaní informácií. Inými slovami, že tí, ktorí majú dobrú schopnosť pamätať si, budú mať lepšie výsledky v teoretickej skúške.

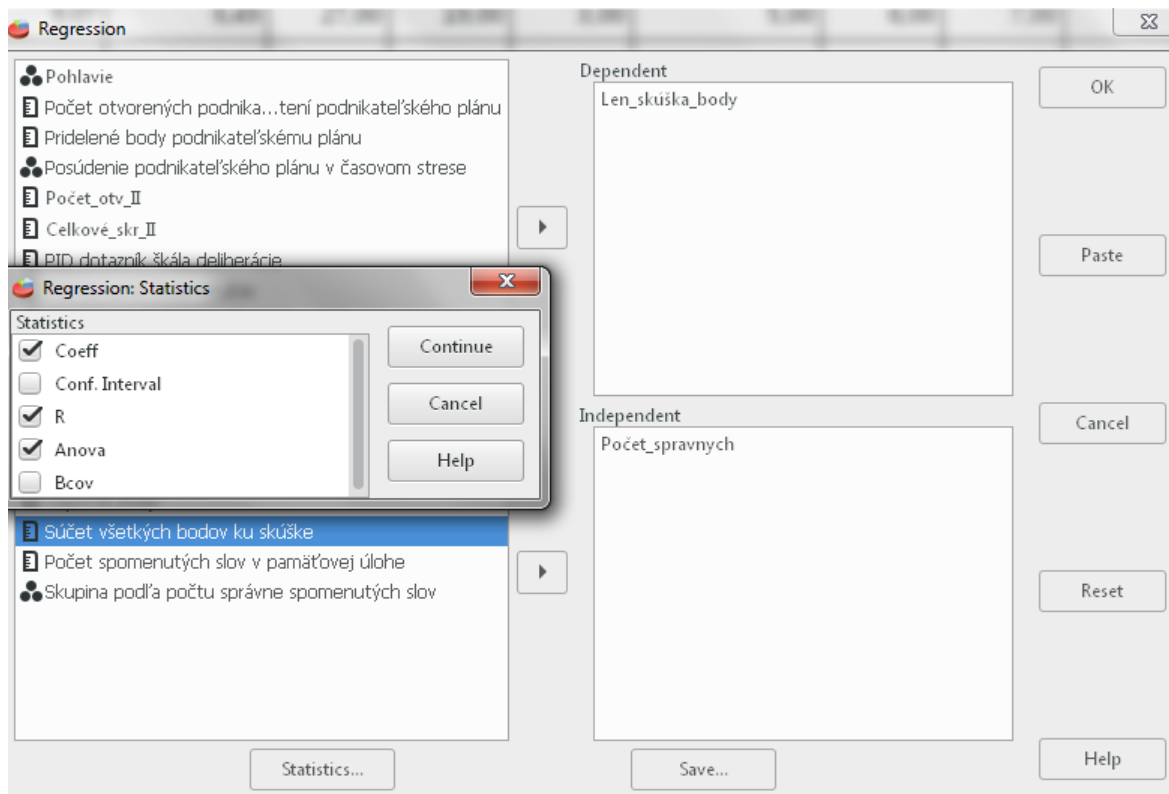
Riešenie: Najprv si musíme jednoznačne definovať závislú a nezávislú premennú. V našom prípade predpokladáme, že počet bodov z teoretickej skúšky je závislá premenná. Závisí od počtu bodov z pamäťovej úlohy, ktorá je nezávislá premenná. Predpokladáme, že čím lepší výsledok z pamäťovej úlohy študent dosiahne, tým viac bodov z teoretickej skúšky získa, pretože tá závisí od schopnosti memorovať sa fakty. Je dôležité naše predpoklady exaktne sformulovať vo forme hypotéz.

Hypotézy:

H_0 : Počet bodov z pamäťovej úlohy nemá štatisticky významný vplyv meraný lineárnou regresiou na dosiahnutý počet bodov z teoretickej skúšky.

H_1 : Počet bodov z pamäťovej úlohy má štatisticky významný vplyv meraný lineárnou regresiou na dosiahnutý počet bodov z teoretickej skúšky.

Konkrétny postup riešenia je nasledovný. Klikneme na Analyzovať (Analyze) vyberieme si Regresiu (Regression) a z dvoch dostupných možností si vyberieme Lineárnu (Linear). V sprievodcovi lineárnou regresiou musíme definovať závislú premennú (dependent variable). V našom prípade je to premenná Len_skuška_body. Presunieme ju do pravého prázdneho okna pre závislú premennú. Nezávislá premenná (independent variable) – tá ktorá ovplyvňuje závislú premennú je Počet_spravnych. Ďalej klikneme na Štatistiku (Statistics). Tu môžeme zaškrtnúť všetky políčka. Ďalej si podrobnejšie vysvetlíme, ktoré čo znamenajú.



Obr. 8.4.1 Sprievodca lineárnou regresiou

Výsledok:

Po spustení analýzy program PSPP vytvorí štyri tabuľky, ktoré si ďalej podrobne opíšme.

Model Summary (Počet bodov z teoretickej skúšky)

R	R Square	Adjusted R Square	Std. Error of the Estimate
,05	,00	-,01	10,29

ANOVA (Počet bodov z teoretickej skúšky)

	Sum of Squares	df	Mean Square	F	Sig.
Regression	22,93	1	22,93	,22	,643
Residual	7734,72	73	105,96		
Total	7757,65	74			

Coefficients (Počet bodov z teoretickej skúšky)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	47,02	2,52	,00	18,69	,000	42,00	52,03
Počet správne spomenutých slov v pamäťovej úlohe	,29	,63	,05	,47	,643	-,96	1,54

Coefficient Correlations (Počet bodov z teoretickej skúšky)

Model	Počet správne spomenutých slov v pamäťovej úlohe
	Covariances

Obr. 8.4.2 Výsledok lineárnej regresie

Prvá tabuľka s názvom Sumarizácia Modelu (Model Summary) opisuje regresný model. V tabuľke sú uvedené tieto výsledky štatistickej analýzy.

R je Pearsonov korelačný koeficient medzi skúmanými premennými. Vo výstupe programu PSPP o regresnej analýze však nenadobúda nikdy záporné hodnoty.

Tak ako pri korelačnej analýze aj tu platí, čím sú hodnoty bližšie číslu 1, tým je vzťah medzi premennými silnejší. V našom prípade je 0,05, čo je veľmi nízka hodnota, blízka nule. To znamená že medzi našimi premennými je veľmi slabý, prakticky neexistujúci vzťah.

R Square sa nazýva **Koeficient determinácie**. Vysvetľuje koľko percent variability závislej premennej Y – Počet bodov z teoretickej skúšky ovplyvňuje nezávislá premenná X – počet správne spomenutých slov v pamäťovej úlohe. Je druhou mocninou korelačného *R*. Teoreticky tiež koeficient determinácie môže nadobúdať hodnoty od 0 do 1, čo vieme transformovať pomocou vynásobenia číslom 100 na percentá, teda od 0 % po 100%.

Koeficient determinácie = variabilita vysvetlená naším modelom / celková variabilita (Hendl, 2009, str. 281).

Čím sa viac blíži k 1, tým viac variability je vysvetľuje daný model. *R Square* je v našom prípade takmer rovné nule, čo znamená, že počet bodov za pamäťovú úlohu nevysvetľuje nič ohľadom počtu bodov z teoretickej skúšky. Môžeme teda konštatovať, že vysoký, alebo nízky počet správne spomenutých slov v pamäťovej úlohe nemá žiadny vplyv na počet bodov dosiahnutý na teoretickej skúške, pretože vysvetľuje 0 % variability. Čiže počet bodov z teoretickej skúšky závisí od najrôznejších iných faktorov ako napríklad motivácia študovať, disciplína, intelekt a mnohých ďalších.

Nastavená hodnota spoľahlivosti R (*Adjusted R Square*) je upravená hodnota, ktorá berie do úvahy počet jednotiek v súbore.

V druhej tabuľke s názvom ANOVA program PSPP udáva nakoľko regresný model dobre opísal naše dáta. Opíšeme si posledné dva stĺpce, ktoré obsahujú hodnotu štatistickej významnosti (*Sig*) pre *F*. $F = 0,22$ a je štatisticky vysoko nevýznamné $p = 0,64$. PSPP vytvoril najlepší možný model z našich dát, ale ten nedokáže dobre opísať vzájomný vzťah medzi premennými.

V tabuľke s názvom koeficienty (Coefficients) sú uvedené výsledky, ktoré presne definujú vzťah medzi našimi skúmanými premennými. Stĺpec B Neštandardizované koeficienty (Unstandardized Coefficients) sú koeficienty, ktoré uvádzame v regresnej rovnici. Konštanta (Constant) je hodnota $b_0 = 47,02$. Hodnota koeficientu $b_1 = 0,29$. Ďalej je dôležité poznamenať, že hodnota *Sig* = 0,64, je oveľa vyššia než hranica štatistickej významnosti $p < 0,05$, potrebná na prijatie alternatívnej hypotézy H_1 . Taktiež interval spoľahlivosti (Confidence interval - CI) obsahuje hodnotu nula CI[-0,96, 1,54]. Na základe týchto výsledkov vieme zostaviť regresnú rovnicu z našich dát, ktorá vyzerá nasledovne.

$Y = 47,02 + 0,29 * X$, teda:

*Počet bodov v teoretickej skúške = 47,02 + 0,29 * Počet správne spomenutých slov v pamäťovej úlohe.*

Ako už bolo hore uvedené vzťah medzi premennými prakticky neexistuje a ako vidíme aj v regresnej rovnici vplyv bodov je minimálny. Najúspešnejší študenti

8.5 Binárna logistická regresia

Logistická regresia je druh regresie, kde závislá premenná je meraná v ordinálnej, alebo v nominálnej forme. Ak je závislá premenná meraná v binárnej forme, potom používame pojem **Binárna logistická regresia**, ktorú softvér PSPP označuje ako Binary Logistic. Keďže binárna logistická regresia pracuje s nominálnymi premennými, preto princíp výpočtu tohto špecifického typu regresnej analýzy je založený na pravdepodobnosti. Field (2013, str. 762) uvádza, že: „ V logistickej regresii namiesto predpovedania hodnoty premennej Y (závislej) z hodnôt premennej X (nezávislej), predpovedáme *pravdepodobnosť* výskytu hodnoty Y, pri známych hodnotách premennej X.“

Matematicky vyjadrený vzťah medzi premennými je nasledovný:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 * x_1)}}$$

Kde:

e – je základ prirodzeného algoritmu.

Ostatné premenné v rovnici sú rovnaké ako pri lineárnej regresii, opísanej v predchádzajúcej kapitole. Pri binárnej logistickej regresii je závislá premenná len v dvoch formách 0 – jav nenastal, 1 jav nastal (Terek, Horníková, Labudová, 2010, str. 171). My pomocou nezávislých premenných skúmame, ako tieto nezávislé premenné ovplyvnili finálnu závislú premennú, inými slovami, či viedli k tomu, že jav nastal 1, alebo viedli k tomu, že jav nenastal.

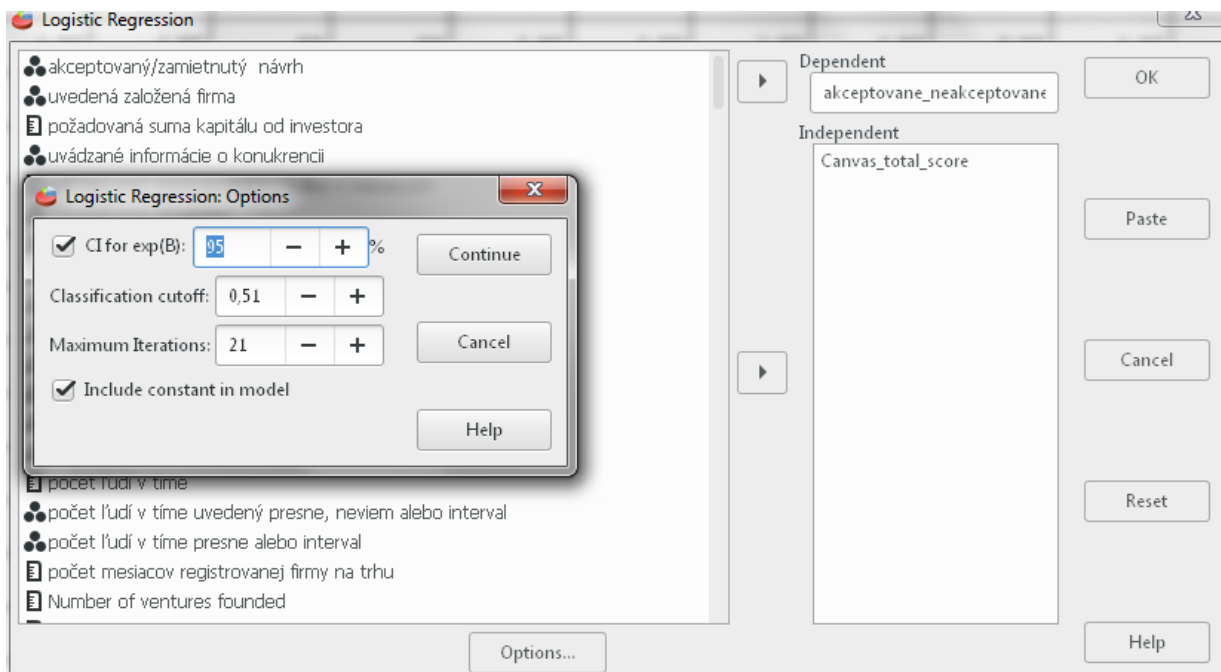
1 úloha: V databázovom súbore č. 7 s názvom Canvas metódy a startup.sav je uvedených 200 skutočných začínajúcich firiem, ktoré sa uchádzali o finančné prostriedky od externých investorov. Na to aby postúpili do vyššieho kola hodnotenia boli na úvod posudzované špecialistami, ktorí hodnotili kvalitu spracovania podnikateľského plánu. 84 firiem posunuli ďalej a 116 odmietli pre nedostatočnú kvalitu. Akceptované boli kódované ako 1 a neakceptované boli kódované ako 0. Akceptovanie alebo odmietnutie bola závislá premenná. Canvas metodika je založená na definovaní, opísaní dôležitých podnikových charakteristík ako kto sú kľúčoví zákazníci, aká je hodnota ponúkaného produktu, odkiaľ a ako sa získajú príjmy a podobne. Tieto charakteristiky boli obodované na škále od 0 po 10 a boli sčítané. Súčet všetkých bodov sme nazvali Celkové skóre canvas (Total canvas score) a v našom príklade bola nezávislá premenná. V našom výskume chceme zistiť, či mali jednotlivé charakteristiky vplyv na akceptovanie alebo odmietnutie. Konkrétne zadanie je nasledovné. Zistite či celkové skóre v Canvas metodike ohodnocovania podnikateľských príležitostí predpovedá úspešnosť akceptovania podnikateľského plánu externými odborníkmi. Hypotézy si stanovíme nasledovne:

H₁: Celkové skóre v Canvas metodike ohodnocovania podnikateľských príležitostí predpovedá úspešnosť akceptovania podnikateľského plánu externými odborníkmi.

H₀: Celkové skóre v Canvas metodike ohodnocovania podnikateľských príležitostí nemá vplyv na úspešnosť akceptovania podnikateľského plánu externými odborníkmi.

Riešenie: V tomto príklade je závislá premenná (dependent variable -DV) dichotomická, teda nadobúda dve hodnoty 0 – odmietnuté, 1 akceptované. Nezávislá premenná celkové skóre v Canvas metodike je intervalová, a preto použijeme logistickú regresiu.

Klikneme na Analyzovať (Analyze) vyberieme si Regresiu (Regression) a z dvoch dostupných možností si vyberieme Binárnu Logistickú (Binary Logistics). V sprievodcovi binárnou logistickou regresiou musíme definovať závislú premennú. V našom prípade je to premenná akceptovane_neakceptovane. Nezávislá premenná – tá ktorá ovplyvňuje závislú premennú je celkové skóre v canvas nazvané: Canvas_total_score. Ďalej klikneme na Možnosti (Options) a v sprievodcovi zaškrtneme nastavenia intervalov spoľahlivosti (confidence intervals), počet iterácií výpočtu zvýšime na 21 (Maximum iterations), a Clasification cutoff zvýšime na 0,51 a zaškrtneme políčko Konštanta v modeli (Include constant in model). Zvýšenie o jednu jednotku v oboch hore uvedených príkladoch sme zaklikli preto, aby nás softvér pustil ďalej a vypočítal výsledok. Klikneme na Pokračovať (Continue) následne OK.



Obr. 8.5.1 Binárna logistická regresia

Výsledok: V prvej tabuľke výsledkov (Dependent Variable Encoding) je uvedené kódovanie premenných, v druhej Počet analyzovaných prípadov (Case Processing Summary). Zaujímajú nás zvyšné tri. V tretej tabuľke Opis modelu (Model Summary) je uvedené koľko variancie model vysvetľuje. Sú tu dve metódy **Cox & Snell R Square** = 0,13 a **Nagelkerke R Square** = 0,18. Pri vynásobení hodnôt číslom 100, dostaneme 13% a 18 % variancie v akceptovaní alebo odmietnutí závisí od dosiahnutého skóre Canvas. Vidíme, že tieto metódy vypočítali odlišný výsledok a pri reportovaní hodnôt odporúčam uvádzať koeficient Nagekerkeho, ktorý môže dosiahnuť hodnotu 1, na rozdiel od Cox & Snellhovo, ktorý ju nedosiahne ani pri ideálnych podmienkach.

V štvrtej tabuľke sú uvedené jednotlivé početnosti. Najdôležitejšia je posledná tabuľka. V tejto sú uvedené konkrétne regresné koeficienty a ostatné štatistiky opisujúce vzájomný vzťah medzi týmito dvoma premennými. Vysvetlíme si ich podrobnejšie.

Wald nám hovorí či b koeficient pre prediktor je štatisticky významne odlišný od nuly. V našom prípade je Wald(ovo) z rovné 23,73 a $p < 0,001$.

Exp(B) je **Exponenciálne B**, ktoré vyjadruje zvýšenie **pomeru šancí (odds ratio)** udalosti pri zvýšení hodnoty prediktora o jednotku. **Pomer šancí** Field (2013, str. 880) definuje ako: „pomer šancí výskytu udalosti v jednej skupine v porovnaní s výskytom udalosti v druhej skupine.“ Field uvádza príklad, kde šanca výskytu určitej udalosti v jednej skupine je 4 a v druhej skupine je len 0,25, tak vzájomný pomer je $4/0,25 = 16$. Ak je rovné 1, pomer šancí v jednej aj v druhej je rovnaký a potom zmeny prediktora nemajú žiaden vplyv. Ak je väčšie ako 1, potom zvyšovaním prediktora rastú šance na výskyt udalosti. Terek a kolektív (2010, str. 175) definujú pomer šancí ako: „podiel pravdepodobnosti, že sledovaná udalosť nastane ($Y = 1$) a pravdepodobnosti, že sledovaná udalosť nenastane ($Y = 0$)“.

V našom prípade pomer šance byť v skupine akceptovaných oproti skupine neakcepovaných rastie ak zvýšime celkové skóre v canvas (*Total canvas score*) o jednu jednotku. Konkrétne $Exp(B) = 1,06$.

Na základe hore uvedených výsledkov prijímame alternatívnu hypotézu H_1 a zamietame nulovú hypotézu H_0 .

9 FAKTOROVÁ ANALÝZA

Faktorová analýza pomocou skúmania vzájomných vzťahov medzi premennými sa snaží identifikovať faktory, ktoré sú redukciou počtu skúmaných premenných. Inými slovami faktorová analýza identifikuje tie premenné, ktoré spolu súvisia do menšieho počtu faktorov. Halama (2011, str. 138) definuje faktorovú analýzu ako: „skupinu štatistických metód, ktoré sú používané na identifikáciu faktorov, t.j. akýchsi latentných premenných, ktorými možno vysvetliť vzťahy medzi pozorovateľnými premennými.“

Field (2013, str. 666.) vymenúva tri oblasti využitia exploračnej faktorovej analýzy:

Poznať štruktúru skupiny premenných. Prostredníctvom premenných následne môžeme poznať faktor, ktorý sýtia. To nám dáva predstavu o čom je približne meraný faktor.

Ak chceme vytvoriť dotazník na meranie latentnej premennej. Faktorová analýza sa rozsiahlo využíva pri skúmaní konštruktivej validity testov (pozri podkapitulu validita). Pomáha nám odhaliť koľko faktorov vlastne test otázkami (premenými) vlastne meria a v akom sú vzájomnom vzťahu (súvisí s bodom 1). Ak napríklad tvorcovia testu vytvoria test, o ktorom tvrdia, že meria dve latentné premenné a faktorová analýza ukáže napr. 4 rovnocenné faktory, potom je nevyhnutné test prepracovať a zmeniť otázky (premenné).

Zmenšiť dátový súbor a pritom zachovať tak veľa pôvodnej informácie ako je len možné. Tým, že v skupine premenných identifikuje niekoľko faktorov, prispieva k znižovaniu ich počtu, a preto faktorové analýzy sa niekedy nazývajú aj metódami redukcie premenných. Tak napríklad z rozsiahleho dotazníka o 40 otázkach vieme vytvoriť napríklad iba 5 faktorov, pričom takmer každá alebo každá z otázok bude patriť k niektorému z faktorov. Faktorová analýza sa pomerne často používa pri skúmaní dotazníkov alebo výpovedí. V tomto prípade si však treba byť vedomí skutočnosti, že prácou len s faktormi získanými redukciou časť informácie strácame.

Poznáme exploračnú a konfirmačnú faktorovú analýzu a softvér PSPP nám umožňuje počítať len exploračnú faktorovú analýzu. U premenných a ich vzájomných vzťahov, ktoré doteraz neboli skúmané a nemáme k nim už overený teoretický rámec aplikujeme exploračnú faktorovú analýzu. Čiže hľadáme aké faktory stoja v pozadí. Naopak u existujúcich teoretických predpokladov o vzájomných vzťahoch medzi premennými, kde je definovaná určitá štruktúra s predpokladanými a prípadne už inými vedcami overenými faktormi, používame konfirmačnú faktorovú analýzu. Jej úlohou je faktory potvrdiť (konfirmovať). Prakticky všetky psychologické testy musia byť pred používaním overené faktorovou analýzou. Pri faktorovej analýze treba mať na pamäti, že z výsledkov konkrétnej faktorovej analýzy nevieme robiť štatistické úsudky na celú populáciu,

teda na základný súbor, pretože sa vzťahujú len na naše dáta (výberový súbor), ktoré sme zozbierali (Mareš a kolektív, 2015).

9.1 Exploračná faktorová analýza

Pri hľadaní faktorov postupujeme prostredníctvom skúmania premenných o ktorých sa domnievame, že súvisia s hľadanými faktormi. Ak tieto premenné spolu korelujeme získame korelačnú maticu, v ktorej sú vzájomné korelácie všetkých premenných. Tieto vzájomné korelácie sú základom štatistickej metódy nazývanej **Exploračná faktorová analýza**. Pri skúmaní týchto korelácií zvyčajne zistíme, že niektoré premenné spolu silne korelujú a vytvárajú určitú skupinu. Táto skupina silne vzájomne korelujúcich premenných zvyčajne však len slabo koreluje s inými premennými, resp. s inými skupinami premenných. Skupina vzájomne korelovaných premenných, a súčasne nesúvisiacich inou korelovanou skupinou premenných, indikuje existenciu faktora. **Faktor** Halama (2011, str. 139) definuje ako : „ určitú vlastnosť, ktorá je sama o sebe nemeateľná, ale odráža sa v určitých prejavoch.“ Field ich definuje aj ako latentné premenné.

Faktorové zaťaženie, sýtenie alebo faktorový náboj (factor loading) znamená aká časť variability konkrétnej premennej X_i je vysvetlená konkrétnym faktorom F . Niekedy sa definuje aj ako korelácia premennej s faktorom (Halama, 2011). Nadobúda hodnoty od -1 (nepriamy vzťah) po 0 (s faktorom nesúvisí vôbec) až po 1. Čím sú hodnoty bližšie k 1 tým daná premenná lepšie sýti daný faktor. Pri tvorbe dotazníkov alebo analýze dát, kde máme možnosť vylučovať jednotlivé premenné (či už otázky z dotazníka alebo položky analýzy), najprv vylučujeme tie, ktoré majú najnižšie sýtenie. Čím nižšie faktorové zaťaženie, tým viac premenná meria niečo iné, a nie zamýšľaný faktor.

Jedným z výstupov exploračnej faktorovej analýzy je aj **Faktorová matica**. Tá nám zobrazuje faktorové zaťaženie jednotlivých premenných voči konkrétnym faktorom. V tejto matici prehľadne vidíme, aké majú jednotlivé premenné vzťahy voči konkrétnym faktorom, teda či faktor sýtia pozitívne, negatívne alebo je ich vplyv malý (pohybuje sa okolo nuly). Ideálny stav je, ak majú silné faktorové zaťaženie k jednému faktorovi a k ostatným dokonca záporné prípadne nízke.

Ako už bolo spomenuté v odseku faktorové zaťaženie faktory vysvetľujú časť variability jednotlivých premenných. Čím viac vysvetľujú tým premenná lepšie sýti faktor. Inými slovami tým máme lepšiu reálnu premennú, ktorá dobre opisuje latentnú premennú. Pri analýzach vieme zistiť aj sumárne skóre, teda koľko percent variability premenných vysvetľujú faktory a tým vieme porovnávať jednotlivé kombinácie premenných.

Ďalším výstupom exploračnej faktorovej analýzy je takzvaná **Vlastná hodnota** (Eigenvalue). Podľa Hendla (2009, str. 508) je to: „číslo, ktoré vo faktorovej analýze udáva pre daný faktor, koľko vysvetľuje variability zo systému premenných, ktoré sledujeme.“ Čím má faktor menšiu hodnotu, tým menej prispieva k vysvetleniu variability premenných.

Na to aby sme mohli realizovať exploračnú faktorovú analýzu musíme mať dostatočný počet skúmaných jednotiek vo výbere (napr. osôb). Bežne sa udáva minimálny počet respondentov na 1 otázku v dotazníku ako 10, niektorí autori uvádzajú aj 15. Teda ak máme dotazník, ktorý má 20 otázok, musíme mať minimálne $20 \times 15 = 300$ respondentov.

Pri zadávaní parametrov v sprievodcovi si musíme v programe PSPP definovať niektoré podmienky. Preto si jednotlivé kroky opíšeme podrobnejšie.

Extrakcia faktorov

Je proces identifikácie jednotlivých latentných faktorov zo skupiny premenných. Existuje viacero metód používaných na extrakciu faktorov a program PSPP používa analýzu hlavných komponentov (principal component analysis) a metódu hlavných osí (principal axis factoring).

Rotácia faktorov

Rotácia faktorov nám slúži na určenie ako sa prvotné faktory, ktoré sme extrahovali líšia jeden od druhého a aké premenné súvisia s ktorými faktormi. Rotáciu osí faktorov si môžeme predstaviť ako rotovanie dvoch úsečiek (samotné úsečky by sme zjednodušene mohli nazvať aj faktormi) okolo centrálného bodu do polohy, tak, aby sa premenné čo najlepšie priradili k jednotlivým faktorom.

Pri exploračnej faktorovej analýze sa používa viacero typov rotácie. Základné rozdelenie je kolmé, inak nazývané aj **Ortogonálne rozdelenie**, a **Šikmé rozdelenie**. Kolmé rozdelenie faktorov používame ak predpokladáme, že jednotlivé faktory sú nezávislé, teda spolu nekorelujú a sú jedinečné. Osi faktorov sa nastavujú ľubovoľne v priestore, ale vždy budú na seba kolmé. Pri šikmom sa osi nastavujú tak, aby sa premenné mohli k nim čo najlepšie priradiť, bez ohľadu na to v akej polohe osi faktorov k sebe zostanú.

Program PSPP nám ponúka štyri typy rotácie:

Žiadnu (None)

Varimax rotácia je kolmý (ortogonálny) typ rotácie. Táto metóda je navrhnutá tak, aby každá premenná sýtala len jeden faktor, teda minimalizuje počet premenných, ktoré majú vysoký náboj. Je v programe prednastavená, pretože sa najľahšie interpretuje a má aj najširšie možnosti aplikácie.

Quartimax rotácia je taktiež kolmý typ rotácie. Jeho výhodou je, že minimalizuje počet faktorov.

Equimax rotácia kombinuje obidva predchádzajúce. Snaží sa vybrať minimálny počet faktorov a súčasne každá premenná sýti len jeden faktor. Je to tiež ortogonálny typ rotácie.

Určovanie počtu faktorov v modeli znamená stanoviť si hranicu, ktoré faktory ešte budem brať do modelu.

Sutinový graf (scree plot) je graf zložený z vlastných hodnôt, ktoré sú na Y-ovej osi a faktorov na X-ovej osi zoradených od najvyšších po najnižšie. Preto

má graf vždy klesajúci tvar, avšak miera klesania je rozdielna, a preto tvar grafu a jeho interpretácia je dôležitá. Prvé vlastné hodnoty sú zvyčajne niekoľkonásobne väčšie než nasledujúce. Ako vidíme v grafe č. 1 prvá vlastná hodnota rovná sa 4,51, druhá je 2,62 a ostatné sú výrazne menšie. Graf začal najprv prudko klesať a potom klesal už len pozvoľne. Táto grafická metóda sa používa na určenie, ktoré faktory v analýze ponecháme, a ktoré vylúčime. Okrem sutinového grafu môžeme použiť aj viaceré iné metódy, založené na iných kritériách, pomocou ktorých budeme vylučovať, respektíve ponechávať.

Metódy, pravidiel pre separáciu faktorov:

- Tie faktory, kde bol v sutinovom grafe pokles najvyšší, ponecháme v modeli, ostatné vylúčime.
- Vylúčime všetky tie, ktorých vlastná hodnota je menšia než 1.
- Vylúčime všetky tie, ktorých vlastná hodnota je pod priemerom tzv. Kaiserovo kritérium.
- Stanovíme si percento variability, ktoré chceme vysvetliť (napr. 80 %) a zahrnieme tam všetky faktory od najviac sýtených dole, pokiaľ nenaplníme stanovené percento.
- Kombinácia hore uvedených pravidiel

Všimnime si prosím, že tieto metódy sú pomerne rôzne a uplatnením týchto metód na rovnaký príklad by viedlo k rôznemu počtu faktorov. Preto treba dôkladne zvážiť akú metódu výberu faktorov zvolíme.

Exploračná faktorová analýza je pomerne rozsiahla analýza, ktorej výstupom je mnoho tabuliek a tak vyvstáva otázka, čo všetko uviesť vo výsledkoch. Pri uvádzaní výsledkov by sme mali uviesť nasledovné minimum.

Počet faktorov, ktorých vlastná hodnota je vyššia ako 1

Percento variability u jednotlivých faktorov a prípadne kumulatívne percento u najvýznamnejších faktorov.

Faktorovú maticu s faktorovým zaťažením jednotlivých premenných

Pri reportovaní výsledkov závisí ich forma prezentácie od miesta, ktoré máme k dispozícii. Ak píšeme záverečnú prácu (diplomovú alebo dizertačnú) alebo rozsiahlu správu o testovaní, či monografiu, potom je vhodné reportovať čo najviac výsledkov. Robíme tak preto, aby iní výskumníci a vedci dokázali v budúcnosti porovnať naše výsledky s ich vlastnými výsledkami a prípadne aby mohli byť naše výsledky mohli byť porovnané v rozsiahlych medzinárodných databázach a meta - štúdiách. To znamená tabuľky s faktormi, kde sú uvedené vlastné hodnoty ako aj percentuálne vyjadrenie vysvetlenia variability a tiež maticu s faktorovým zaťažením všetkých premenných. V prípade, že reportujeme výsledky vo vedeckom článku a sme limitovaný obmedzeným priestorom, potom môžeme uviesť len tie najzásadnejšie výsledky.

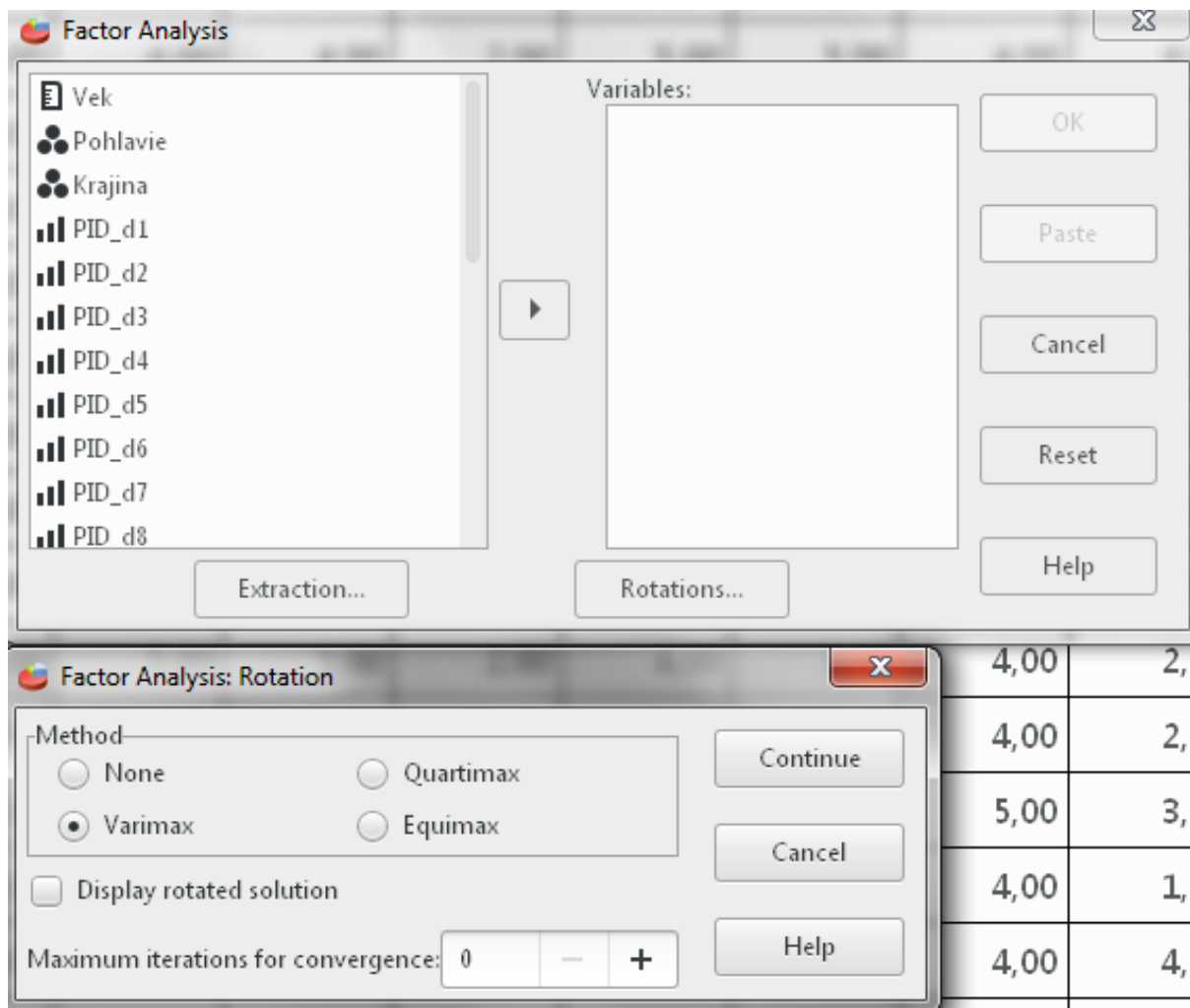
1. úloha: V nasledujúcom príklade budeme testovať už etablovaný psychologický dotazník vytvorený za účelom merať dve latentné premenné: preferenciu k intuícii a k deliberácii. Dotazník sa nazýva Preference for Intuition and Deliberation a používa skratku PID, a vytvorila ho vedkyňa Cornélia Betschová (Betsch, 2004) a testovala ho na nemeckých študentoch. Má 18 otázok a polovica z nich by mala sýtiť latentnú premennú intuíciu a druhá polovica otázok (v našom prípade premenných) deliberáciu, inými slovami aj sklon k uvažovaniu a racionalite.

V našom príklade vidíme, že každá z otázok je označená na základe nasledovného princípu: PID_1d, kde 1d znamená, že ide o prvú otázku zo subškály deliberácia. Rovnako PID_1i, kde 1i znamená, že ide o prvú otázku zo subškály intuícia.

Riešenie: To či použijeme exploračnú alebo konfirmačnú faktorovú analýzu je v tomto prípade relatívne hraničné, nakoľko by sme mali testovať už teoretický konštrukt, a preto by sme mali preferovať konfirmačnú analýzu. Avšak rozhodli sme sa použiť exploračnú faktorovú analýzu. Naša úvaha bola založená na nasledovných argumentoch. Nakoľko nemecká a slovenská kultúra sú pomerne odlišné, rovnako iné je u nás oproti Nemecku aj vnímanie intuície a sklonu k uvažovaniu – deliberácii, a preto sme sa rozhodli, že aplikujeme Exploračnú faktorovú analýzu. Okrem kultúrnej odlišnosti zdôrazňujeme aj prvé testovanie populácie na Slovensku, preto sú naše výsledky úplne nové. Ak by sme však testovali znova nemeckú populáciu, už by sme rozhodne mali zvoliť Konfirmačnú faktorovú analýzu.

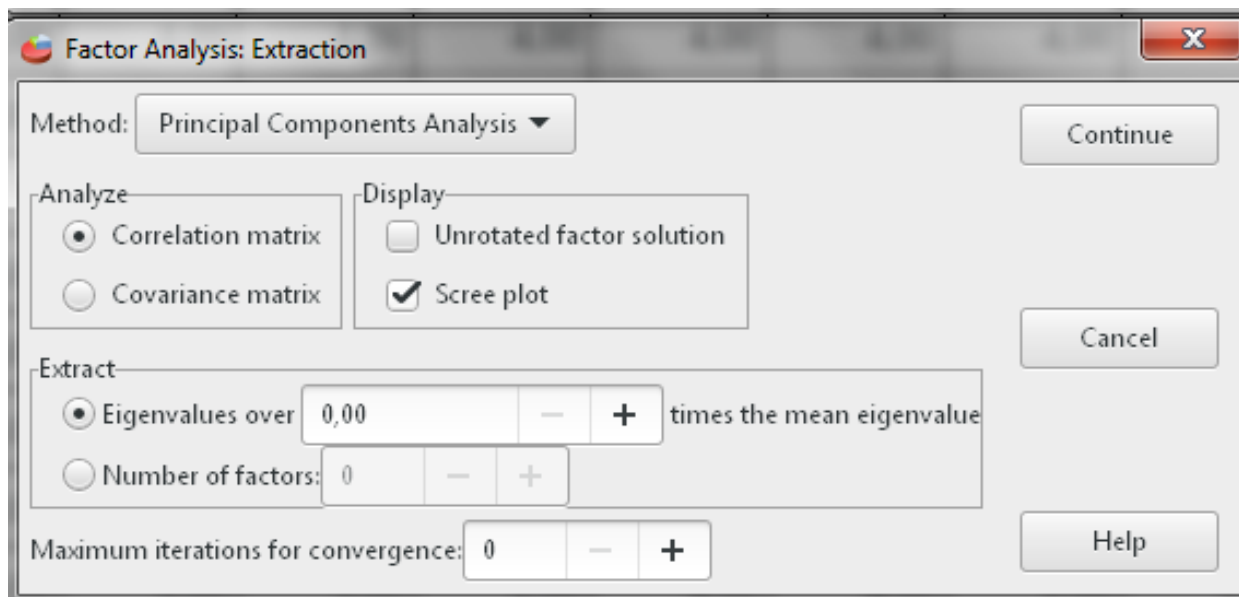
Keď sme sa rozhodli pre exploračnú faktorovú analýzu na začiatok zväžíme viacero špecifik, ktoré sa viažu k nášmu prípadu. Autorka dotazníka tvrdí, že dotazník meria intuíciu a deliberáciu, teda dva faktory. Ďalej tvrdí, že intuícia nie je opakom deliberácie, a preto sú obidva konštrukty nezávislé, teda konštatuje, že spolu by nemali korelovať. Preto zvolíme ortogonálny typ rotácie, teda taký, kde sú osi na seba vzájomne kolmé. Pri výbere ortogonálneho typu rotácie máme viacero možností ako Varimax, Quartimax a iné, z ktorých môžeme vybrať. Ak chceme čo najlepšie spoznať vzťah jednotlivých otázok k dvom latentným premenným je vhodné v tomto prípade vybrať rotáciu Varimax. Táto metóda rotácie nebude znižovať počet faktorov a ukáže nám aj iné faktory ktoré sú „skryté“ v odpovediach na otázky ohľadom intuície a deliberácie.

Klikneme Analyzovať (Analyze), potom Faktorová analýza (Factor Analysis). Následne presunieme všetky otázky z dotazníka zľava do pravej časti s názvom Premenné (Variables). Potom klikneme na nastavenie Rotácie (Rotation). Zo štyroch ponúkaných metód vyberieme Varimax. Potom klikneme na *Pokračovať* (Continue) a ideme nastavovať parametre v *Extrakcii* (Extraction).



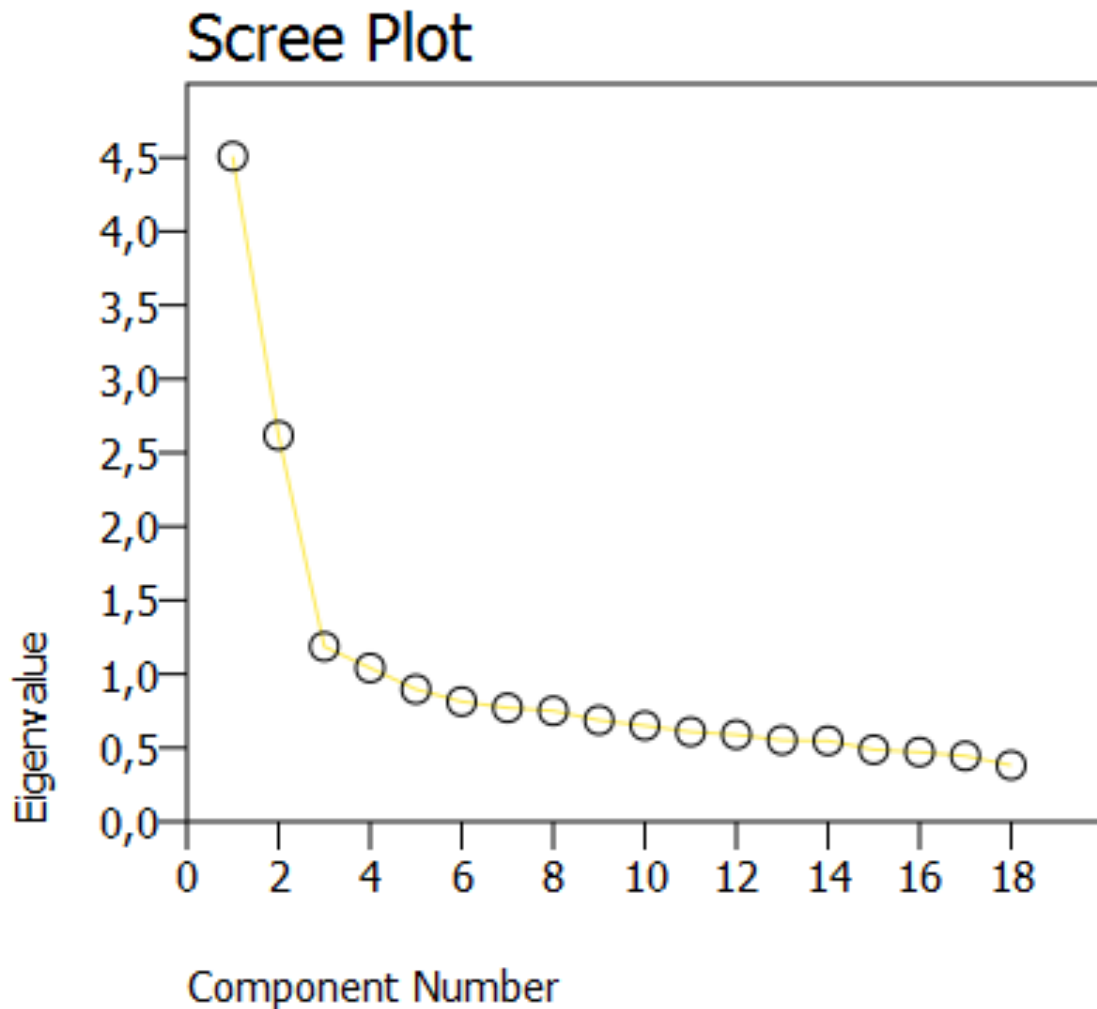
Obr. 9.1.1 Sprievodca faktorovou analýzou

Ďalej klikneme na *Extrakciu (Extraction)*. Z ponúkaných metód vyberieme *Analýzu hlavných komponentov (Principal component analysis)*. Ďalej v ponuke *Analyzovať (Analyze)* zaškrtneme korelačnú maticu (*Correlation matrix*). ponuke *Zobraziť (Display)* zaškrtneme *Sutinový graf (Scree plot)*. V ponuke *Extrakcia (Extraction)* nám softvér PSPP ponúka možnosť vybrať *Vlastné hodnoty (Eigenvalues)* väčšie než hranica, ktorú si stanovíme sami. Pre dôkladné pochopenie dát, neodporúčam, aby sa stanovoval akýkoľvek limit pre vlastné hodnoty. Napríklad zobrazíť hodnoty väčšie než 0,5. Ak nestanovíme žiaden limit tak softvér zobrazí všetky existujúce faktory na základe čoho si dokážeme spraviť predstavu nielen o počte faktorov, ale aj o ich relatívnom význame voči ostatným. Taktiež stanovovanie konkrétneho počtu faktorov (*Number of factors*) necháme bezo zmeny.



Obr. 9.1.2 Sprievodca extrakciou faktorov

Výsledky: Výstupom je viacero tabuliek a graf, ktoré si opíšeme a budeme ich interpretovať.



Obr. 9.1.3 Sutinový graf (Scree plot)

V sutinovom grafe vidíme všetky nájdené faktory s ich *Vlastnými hodnotami (eigenvalues)*. Pri prvých dvoch faktoroch je zrejmy najvýraznejší pokles. U tretieho, štvrtého aj piateho vidíme ešte určitý pokles, ale nasledovné faktory už klesajú minimálne. Preto prvé dva faktory sú najvýznamnejšie v porovnaní s ostatnými. Sutinový graf je dobrý na prvotnú a grafickú identifikáciu, avšak ďalej budeme pracovať už len s číselnými výsledkami, ktoré sú uvedené v nasledujúcich tabuľkách.

Total Variance Explained

<i>Component</i>	<i>Initial Eigenvalues</i>		
	<i>Total</i>	<i>% of Variance</i>	<i>Cumulative %</i>
1	4,51	25,05	25,05
2	2,62	14,54	39,59
3	1,19	6,59	46,17
4	1,04	5,78	51,96
5	,90	4,98	56,94
6	,81	4,52	61,45
7	,77	4,28	65,74
8	,75	4,17	69,91
9	,69	3,82	73,72
10	,65	3,61	77,33
11	,61	3,38	80,71
12	,59	3,27	83,98
13	,55	3,06	87,05
14	,55	3,04	90,09
15	,49	2,71	92,80
16	,47	2,61	95,41
17	,44	2,47	97,88
18	,38	2,12	100,00

Obr. 9.1.4 Identifikované faktory, vlastné hodnoty (eigenvalues) a percentá vysvetlenej variancie

V obrázku č. x sú uvedené všetky identifikované faktory v stĺpci *Komponenty (Component)*. Je ich 18, avšak majú rozdielnu výšku *Vlastnej hodnoty (eigenvalue)*, ktorá je v stĺpci *Total*. Prvý faktor má *Vlastnú hodnotu* 4,51 a vysvetľuje 25,05 % variability. Druhý faktor má už len približne polovičnú hodnotu z prvého teda 2,62 a vysvetľuje 14,54 %. Od dotazníka sme očakávali, že v podstate všetky otázky budú súvisieť s intuíciou alebo z uvažovaním – deliberáciou. Inými slovami, že bude pozostávať z dvoch jednoznačných faktorov, ktoré budú vysvetľovať veľkú väčšinu variability. To sa nám potvrdilo čiastočne. Síce dva faktory sú dominantné, avšak vysvetľujú spolu len 39,59 % variability. To je pomerne málo na dotazník, ktorý má merať len dva faktory.

Vlastná hodnota (eigenvalue) je vyššia ako číslo 1 aj pre tretí (1,19) i pre štvrtý faktor (1,04). Takže máme tu ešte dva faktory, ktoré sme neočakávali.

V *Matici komponentov (Component matrix)* je uvedený faktorový náboj, sýtenie pre každú otázku z dotazníka. V prvom riadku je otázka PID_d1, čiže zo škály uvažovanie. Vidíme, že táto otázka má pre prvý najsilnejší faktor sýtenie 0,63 a pre druhý faktor záporné sýtenie rovné -0,44. Rovnako ostatné otázky zo škály uvažovanie majú pre prvý faktor pozitívne sýtenie. Môžeme tento faktor nazvať uvažovanie – deliberácia. Naopak pre druhý faktor sú faktorové záťažé negatívne pre otázky zo škály uvažovanie a pozitívne pre otázky zo škály intuícia. Tento druhý faktor môžeme teda nazvať intuícia. Čiže máme dva faktory, kde prvý sýtia všetky otázky zo škály deliberácia a druhý sýtia všetky otázky zo škály intuícia. Poďme sa teraz pozrieť na to v akom vzťahu sú otázky z deliberácie voči faktoru intuícia a v akom vzťahu sú otázky intuície voči faktoru deliberácia. Očakávame, že budú v negatívnom vzťahu. Teda že otázka z intuície nebude pozitívne sýtiť faktor uvažovania a naopak, že otázky z deliberácie nebudú pozitívne sýtiť faktor intuícia. Pre faktor deliberácia (1 faktor) vidíme že otázky z intuície PID_i1 až PID_i9 sú tiež pozitívne sýtené. Mali by byť sýtené negatívne. To nie je pozitívna správa. Naproti tomu pre faktor intuícia (2 faktor), je každá otázka z deliberácie (PID_d1 až PID_d2) sýtená negatívne. To je podľa očakávaní a je to pozitívna správa. Keď uvažujeme nad interpretáciou môžeme povedať, že prvý faktor je zmiešaný, nakoľko ho pozitívne sýtia všetky otázky z deliberácie (čo je podľa našich očakávaní) ale aj všetky z intuície, čo je proti našim očakávaniam. Preto nemôžeme tvrdiť, že prvý faktor je len deliberatívny.

	Component Matrix																
	Component																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
PID_d1	,63	-,44	-,15	-,15	,08	-,42	-,14	-,05	-,09	-,11	,14	-,08	-,12	,15	,11	,06	-,20
PID_d2	,66	-,34	-,09	-,06	,06	,32	,10	,02	-,11	-,01	-,17	-,24	,00	,36	-,11	,00	-,17
PID_d3	,62	-,06	,08	-,35	-,18	-,08	,31	-,12	,06	,00	-,28	-,11	-,09	-,25	-,14	,04	-,16
PID_d4	,54	-,39	,18	,32	-,10	,05	-,05	-,05	-,05	-,02	,06	-,29	-,23	-,10	,20	,23	,39
PID_d5	,42	-,15	,32	,60	-,18	,00	-,26	,07	,23	,04	-,11	,11	-,11	,01	,02	-,18	-,32
PID_d6	,64	-,21	,10	-,06	-,15	,00	-,10	-,03	,29	-,23	,21	,02	,36	,12	-,37	-,07	,19
PID_d7	,63	-,41	-,15	-,06	,09	-,05	-,14	,11	,06	,48	-,09	,21	,09	-,04	-,13	,15	,16
PID_d8	,50	-,36	-,18	-,10	,43	,19	-,25	,03	-,14	-,12	,10	,05	,08	-,41	,13	-,19	-,10
PID_d9	,54	-,12	,18	,30	,08	,00	,59	-,08	-,27	-,04	,18	,30	-,01	-,03	-,04	,03	-,03
PID_i1	,56	,37	-,21	-,08	-,06	,03	-,15	-,06	-,24	-,06	-,11	,24	-,11	,39	,18	-,12	,16
PID_i2	,29	,54	-,13	,28	-,16	-,05	-,22	-,14	-,36	-,07	-,15	-,05	,38	-,16	-,07	,29	-,08
PID_i3	,18	,43	-,43	,25	,43	,02	,08	-,28	,42	-,04	,01	,00	-,13	,06	-,05	,24	-,06
PID_i4	,59	,23	-,01	-,25	-,30	,08	,04	-,12	,31	-,10	-,08	,17	,03	-,15	,36	,01	,06
PID_i5	,47	,60	-,09	-,08	-,25	,09	,01	-,03	-,02	,29	,41	-,18	,00	-,01	,07	-,07	-,18
PID_i6	,39	,40	-,17	,02	,01	-,02	,09	,77	,06	-,13	,00	,00	-,09	-,06	-,03	,14	,00
PID_i7	,26	,50	,40	-,21	,11	,05	-,28	-,10	-,12	-,04	,04	,09	-,43	-,08	-,34	,07	,04
PID_i8	,48	,52	,02	,17	,24	-,17	,12	-,01	,00	,08	-,14	-,21	,07	-,06	-,03	-,46	,22
PID_i9	,19	,27	,69	-,18	,39	-,02	,00	,07	,08	,05	-,02	-,04	,28	,16	,24	,19	-,06

Obr. 9.1.5 Faktorové záťažé pre každú z otázok dotazníka PID

Uvádzanie výsledkov: Na analýzu vnútornej štruktúry dotazníka PID slovenskom prostredí bola zvolená exploračná faktorová analýza, konkrétne metóda hlavných komponentov (principal axis factoring), Varimax rotácia. Analýza identifikovala

4 faktory, ktorých vlastná hodnota (Eigenvalue) bola vyššia ako 1, a hoci prvé dva boli silnejšie ako ostatné, okrem nich sme našli aj ďalších 2 faktory. Prvý faktor vysvetľoval 25,05 % variability a druhý 14,54 %, spolu vysvetľovali spolu 39,59 %, tretí faktor 6,59% a štvrtý 5,78%.

Všetky otázky, ktoré mali sýtiť latentný faktor Deliberáciu boli v pozitívnom vzťahu s týmto faktorom a faktorové zaťaženie sa pohybovalo od 0,42 pre *PID_d5* po 0,66 pre *PID_d2*. Pri latentnej premennej Intuícia bolo faktorové zaťaženie všetkých otázok merajúcich intuíciu bolo v pozitívnom vzťahu a pohybovalo sa od 0,19 pre *PID_i9* po 0,54 pre *PID_i9*. Avšak ako problematické pri tomto dotazníku vnímame, že prvý faktor bol zmiešaný a pozitívne ho sýtili nielen otázky zo škály deliberácia, ale aj všetky otázky zo škály intuícia. Čiže nemôžeme hovoriť o prvom faktore ako čisto deliberatívnom, ale ako zmiešanom, čo je úplnom rozpore s pôvodnými predpokladmi o jednoznačnosti faktorov merajúcich len deliberáciu a len intuíciu.

LOGISTIC REGRESSION

LOGISTIC REGRESSION akceptovane_neakceptovane WITH Canvas_total_score
/CRITERIA = CUT(0.5) ITERATE(21)
/NOORIGIN.

Dependent Variable Encoding

Original Value	Internal Value
Neakceptované	0
Akceptované	1

Case Processing Summary

Unweighted Cases	N	Percent
Included in Analysis	200	100,00
Missing Cases	0	,00
Total	200	100,00

note: Estimation terminated at iteration number 4 because parameter estimates changed by less than 0,001

Model Summary

Step 1	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	244,20	,13	,18

Classification Table

Observed		Predicted			
		akceptovaný/zamietnutý návrh Neakceptované	Akceptované	Percentage Correct	
Step 1	akceptovaný/zamietnutý návrh	Neakceptované	89	27	76,72
		Akceptované	43	41	48,81
Overall Percentage					65,00

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Total canvas score	,05	,01	23,73	1	,000	1,06
	Constant	-2,26	,43	27,22	1	,000	,10

Obr. 9.1.6 Výsledok binárnej logistickej regresie

Ak by sme hodnoty vložili do hore uvedenej rovnice, potom vieme vypočítať konkrétne pravdepodobnosti byť akceptovaný pre každé canvas skóre.

10 POROVNÁVANIE STREDOVÝCH HODNÔT

Porovnanie stredových hodnôt (priemerov) je jedna najčastejšie používaných metód dvojrozsmernej štatistickej analýzy. Porovnáваме hodnoty dvoch premenných. Pri porovnávaní priemerov nás zaujíma, či sa tieto medzi sebou štatisticky významne odlišujú, teda či odlišnosť medzi nimi je systematická, a nie len spôsobená náhodou. Z výsledkov našich výberových súborov potom usudzujeme na celú populáciu.

Porovnanie hodnôt premennej v jednotlivých skupinách patrí k najčastejšie používaným štatistickým metódam. V praktickej aplikácii bežne riešime úlohy ako napríklad či sa chlapci odlišujú od dievčat v dochádzke, či zamestnanci dosahujú lepšie pracovné výsledky po školení, než dosahovali pred školením, alebo či naša experimentálna intervencia (napr. navodenie časového stresu) mala vplyv na množstvo požadovaných informácií pred prijatím rozhodnutia. Pri výpočte priemerných hodnôt v jednotlivých skupinách pravdepodobne vždy bude existovať rozdiel medzi jednou skupinou a druhou. Rozdiel v priemeroch môže byť veľmi malý a aj veľký, no len na základe priemeru nemôžeme povedať, či je rozdiel spôsobený náhodnou chybou alebo je systematický. Preto používame viacero testov, ktoré nám umožňujú porovnať stredové hodnoty – priemery, a určiť či sa medzi sebou skupiny štatisticky významne líšia alebo nie. Testy môžeme rozdeliť na dve veľké skupiny a to parametrické a neparametrické, kde pracujeme s poradiami a mediánmi. To aký typ testu použijeme závisí od splnenia určitých podmienok pre dáta v našom skúmanom súbore. Najprv si vysvetlíme parametrické testy.

10.1 Parametrické testy a podmienky ich použitia

V softvéri PSPP sú uvedené všetky parametrické testy v menu *Analyzovať (Analyze)*, kde je uvedená možnosť porovnať priemery, inými slovami aj stredové hodnoty (*Compare Means*). V okne zo zoznamom nám ponúka päť rôznych testov, ktoré ďalej podrobnejšie opíšeme. Všetky testy v tejto časti sú parametrické testy. Parametrické testy vychádzajú z určitých predpokladov, ktoré musia byť najprv splnené, a až potom môžeme parametrické testy použiť. Zoznam podmienok pre použitie parametrických testov je podľa Mareša a kolektívu (2015, str. 140) nasledovný:

1. Normalita rozloženia

Predpokladáme, že naše dáta pochádzajú z populácie, ktorá má normálne rozloženie (Meloun, Militký, 2012, str. 480). Predpoklad normálneho rozloženia

dát overíme pomocou Kolgomorovho – Smirnovho testu. Tento test je založený na porovnávaní distribúcie našej vzorky s normálne rozloženou vzorkou, ktorá má rovnaký priemer a štandardnú smerodajnú odchýlku. Preto pri reálne práci s dátami najprv vypočítame pre skúmané premenné Kolgomorov – Smirnov test a na základe výsledkov z tohto testu potom používame parametrické, alebo neparametrické testy.

Pri tomto predpoklade spomenieme aj minimálnu veľkosť vzorky, ktorú potrebujeme, aby sme mohli zodpovedne aplikovať parametrické testy. Samozrejme čím väčšia vzorka, tým lepšie, avšak nie vždy máme možnosť mať dostatočne veľkú vzorku. Preto sa uvádza ako hraničný počet cez 30 jednotiek v súbore ako minimálna dolná hranica.

2. Zhoda rozptylov (homoskedasticita)

Správnosť výsledkov z parametrických testov závisí aj od rozptylov v jednotlivých skúmaných skupinách, a preto musí byť splnený predpoklad, že rozptyl v jednotlivých skupinách je približne rovnaký. Hoci priemery môžu byť takmer úplne rovnaké, alebo rovnaké, ale keď v jednej skupine sú dáta pomerne blízko priemeru a v druhej sú zas extrémne vysoké i nízke, avšak ich priemer je medzi nimi, potom je rozptyl u prvej a druhej skupiny je veľmi rozdielny aj keď priemery sú rovnaké. Predpoklad približne podobných rozptylov sa nazýva **Predpoklad homogenity rozptylu (Assumption of homogeneity of variance)**. Ďalším predpokladom je **Predpoklad sféricity (Sphericity assumption)**, ktorý požaduje aby rozptyly všetkých možných rozdielov merania v rámci jednotlivých jednotiek v súbore boli zhodné.

Tieto dva predpoklady vieme overiť jednoducho pomocou Levenovho testu, ktorý program PSPP štandardne uvádza vždy spolu s výsledkami parametrických testov. Levenov test aj Kolgomorov – Smirnov test uvádzajú vo výsledkoch aj hodnotu štatistickej významnosti (sigma - Sig) a pri oboch testoch ak je táto menšia než 0,05, potom sú výsledky štatisticky významné a teda sú porušené *Predpoklady sféricity a Homogenity rozptylov* u Levenovho testu a *Predpoklad normálneho rozloženia* dát v prípade Kolgomorovho testu. V týchto prípadoch potom nemôžeme použiť parametrické testy, ale musíme skupiny porovnať pomocou neparametrických testov.

3. Intervalové dáta

Premenná, pomocou ktorej porovnáваме skupiny by mala byť intervalová alebo pomerová, teda kardinálna premenná. Podľa správnosti by sme nemali porovnávať premenné merané na ordinálnej škále (veľmi súhlasím, súhlasím, čiastočne súhlasím atď.), aj keď tento predpoklad sa v praxi často porušuje. Čitateľ sa stretne s článkami aj v renomovaných časopisoch kedy sa na ordinálne dáta použili parametrické testy. Argumentuje sa tým, že ak máme škálu s dostatočne veľa stupňami, potom porušenie tohto predpokladu nemá až tak významné následky.

4. Nezávislosť merania

Pri tejto podmienke predpokladáme, že dáta merané na jednom subjekte (jednotke v súbore) nie sú ovplyvnené druhým subjektom.

Čiže ak naše premenné spĺňajú hore uvedené predpoklady potom môžeme použiť na analýzu parametrické testy. Program PSPP umožňuje počítať nasledovné parametrické testy:

Jednovýberový t – test (One sample T - Test). Je test, ktorý používame na porovnanie priemeru v našom výberovom súbore (jedna skupina – jeden výber) s určitou konkrétnou hodnotou priemeru v celej populácii. Napríklad vieme, že celonárodný priemer (základný súbor) na maturitnom teste z matematiky bol 75 bodov a my pomocou tohto testu porovnáme výsledky v našej skupine (výberový súbor) s touto hodnotou.

Dvojvýberový t - Test (Independent sample T- Test) Je to najčastejšie používaný test na porovnanie priemerov intervalovej premennej u dvoch rozdielnych skupín (dva výbery). Používa sa pri testovaní hypotéz, kde u nulovej hypotézy tvrdíme, že dve rôzne skupiny dosahujú v nejakej premennej rovnaký priemer, respektíve priemer premennej v jednej skupine sa štatisticky významne nelíši od priemeru v inej skupine. Test nám odpovie na otázku, či sa skupiny štatisticky významne medzi sebou líšia. Tento typ t testu sa používa veľmi často pri experimentálnom výskume, kde porovnáme dve skupiny, pričom jedna z nich (experimentálna skupina) bola vystavená určitej experimentálnej manipulácii a druhá nie (kontrolná skupina).

Párový T- test na strednú hodnotu. (Paired Sample T- Test) Používame ak testujeme 1 skupinu dva krát a zaujíma nás, či sa medzi sebou líšia v priemeroch po nejakej udalosti, napríklad pred školením a po školení. Inými slovami porovnáme priemery, ktoré dosiahol 1 respondent prvý a druhý krát, teda párové hodnoty.

Analýza rozptylu (variability) (ANOVA: Analysis of variance). Je ekvivalentom dvojvýberového t- testu pre viac než 2 skupiny. Teda ak porovnáme len dve skupiny, potom používame dvojvýberový t – test a ak porovnáme viac ako 2 skupiny, potom používame metódu ANOVA. Pri tomto štatistickom teste zistujeme či sa v jednej premennej líši viacero porovnávaných skupín.

10.1.1 Veľkosť účinku (effect size)

Podľa úrovne hladiny významnosti hovoríme o štatistickej významnosti našich výsledkov, avšak existuje ešte vecná významnosť. Vecná významnosť vyjadrená pojmom Veľkosť účinku je matematický vzorec, pomocou ktorého vieme zistiť nielen či sa skupiny medzi sebou štatisticky významne líšia, ale aj ako veľmi silno sa medzi sebou líšia. Čím väčší rozdiel medzi porovnávanými skupinami,

tým väčší je efekt účinku. Na výpočet veľkosti účinku máme viacero metód, avšak pri t – testoch a Anove sú najčastejšie používané takzvané Cohenovo d a Pearsonov korelačný koeficient. Pearsonov korelačný koeficient nadobúda hodnoty 0 až 1 a Cohenovo d môže nadobudnúť aj hodnoty vyššie než 1 (Soukup, 2013, str. 130). V dole uvedenej tabuľke č. X máme hodnoty a verbálne vyjadrenie pre obidva koeficienty. Je to štandardizovaný parameter, a tým nám umožňuje porovnávať medzi sebou rôzne štúdie a to aj vtedy ak tieto použili rôzne škály (Ptáček, , Raboch, 2010). Veľkosť efektu je závislá od počtu respondentov. Teda ak chceme dosiahnuť silné efekty, potom musíme veľa respondentov. Pre štatistické metódy sa počíta rôzne a pre každú metódu si ho budeme uvádzať osobitne.

Pearsonovo r

Rovnako ako pri korelácii výsledok nadobúda hodnoty od 0 po 1 a pre testy môžeme vypočítať veľkosť účinku z informácií, ktoré sú výstupom samotného t – testu. Vzorec na výpočet pre t testy má nasledovný konkrétny tvar (Field, 2002 str. 166):

Pre t testy:

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

Kde: r – Pearsonovo r, t – testovacie kritérium, df – počet stupňov voľnosti,

Pre Anovu :

Je niekedy dole uvedený koeficient nazývaný aj koeficient Eta η^2 (Field, Hole, 2003, str. 180)

$$r^2 = \frac{SS_M}{SS_T}$$

Kde: SS_M je rozptyl medzi skupinami (Between Groups) SS_T je a celkový rozptyl (Total).

Ak máme rovnako veľké skupiny pri ANOVE môžeme použiť aj nasledovný vzorec tzv. Omega na výpočet veľkosti účinku, ktorý je presnejší než hore uvedený vzorec (Field, Hole, 2003, str. 181).

$$\omega^2 = \frac{MS_M - MS_R}{MS_R + ((n - 1) * MS_R)}$$

Výhodou Pearsonovho r oproti Cohenovmu d je, že vieme z neho vypočítať aj **Koeficient determinácie**. Hodnotu r umocníme na druhú a výsledok vynásobíme 100 a získame percento variability, ktoré vysvetľuje koľko variability generuje pôsobenie nezávislej premennej na závislú.

Cohenovo d

Pre t testy:

$$d = \frac{x_1 - x_2}{\sqrt{\frac{s_1 + s_2}{2}}}$$

Kde: x_1 a x_2 sú priemery porovnávaných skupín a s_1 a s_2 sú ich smerodajné odchýlky.

Veľkosť účinku	Hodnota Cohenovho d	Interval pre Pearsonovo r
Nízka	0,2 až 0,5	0,1 až 0,3
Stredná	0,5 až 0,8	0,3 až 0,5
Vysoká	0,81 až 1,19	0,5 a viac
Veľmi vysoká	1,2 až 1,99	
Extrémne vysoká	2 a viac	Blízka 1

Tabuľka č. 10. Hodnoty veľkosti účinku (Cohen, 1988, Sawilowsky, 2009, Mareš et al. 2015)

Nakoľko softvér PSPP tento test nepočíta, avšak v súčasnosti ho už takmer všetky vedecké časopisy vyžadujú, musíme si ho dopočítať ručne.

10.1.2 Dvojvýberový t – test (Independent sample t -test)

Tento štatistický test používame ak chceme porovnať jednu alebo viac premenných pre dve nezávislé skupiny. Pomocou tohto testu vieme určiť, či sa jedna skupina štatisticky významne líši od druhej skupiny v hodnotách skúmanej premennej. Napríklad či sa líši počet bodov zo skúšky (skúmaná premenná) u mužov (prvá skupina) a u žien (druhá skupina). Každý jednotlivec sa môže nachádzať len v jednej zo skupín, teda buď je zaradený ako muž, alebo je zaradený ako žena. Preto sa tento test nazýva aj dvojvýberový (independent sample), lebo máme dve nezávislé skupiny.

1 úloha: V dátovom súbore č. 2 s názvom: Študenti body a pamäť porovnajte počet bodov zo skúšky u mužov a žien.

V tomto jednoduchom príklade vychádzame z predpokladu, že ženy budú usilovnejšie študentky a na skúšku sa pripravujú lepšie, a preto budú mať vyššie priemerný počet bodov než muži. Na základe týchto predpokladov si hypotézy stanovíme nasledovne:

H_1 : Ženy dosiahli štatisticky významne vyšší počet bodov zo skúšky ako muži.

H_0 : Medzi ženami a mužmi nie je štatisticky významný rozdiel v počte dosiahnutých bodov zo skúšky.

Riešenie:

1. Krok : Overenie predpokladov (homogenita rozptylu, sféricita a normalita rozloženia) na to, aby sme mohli použiť parametrický test.

2. Krok : Ak budú splnené predpoklady, potom použiť parametrický *Dvojvýberový t – test (Independent Samples T test)*.

1. Krok: Prvým predpokladom je *Normalita rozloženia dát* a ten si overíme pomocou Kolmogorovho – Smirnovho testu. Ďalším predpokladom je overenie si Homogenity rozptylu a sféricity a tie otestujeme pomocou Levenovho testu, ktorý je súčasťou výsledkov Dvojvýberového t testu. Tretím predpokladom sú intervalová, resp. pomerová premenná. Študenti mohli celkovo dosiahnuť 0 až 100 bodov a teda predpoklad pomerovej premennej spĺňame. Posledným predpokladom je nezávislosť merania, teda či výsledky testu u muža mohli ovplyvňovať výsledky u ženy. Predpokladáme, že nie. Zo štyroch predpokladov dva spĺňame, normalitu overíme Kolmogorovým – Smirnovovým testom a Levenov test bude uvedený vo výsledkoch t – testu. Začneme Kolmogorovým testom. Postup pri tomto teste sme si podrobnejšie opísali v kapitole o skúmaní vzájomných vzťahov medzi ordinálnymi premennými a tak v tejto časti už uvedieme len výsledok.

One-Sample Kolmogorov-Smirnov Test

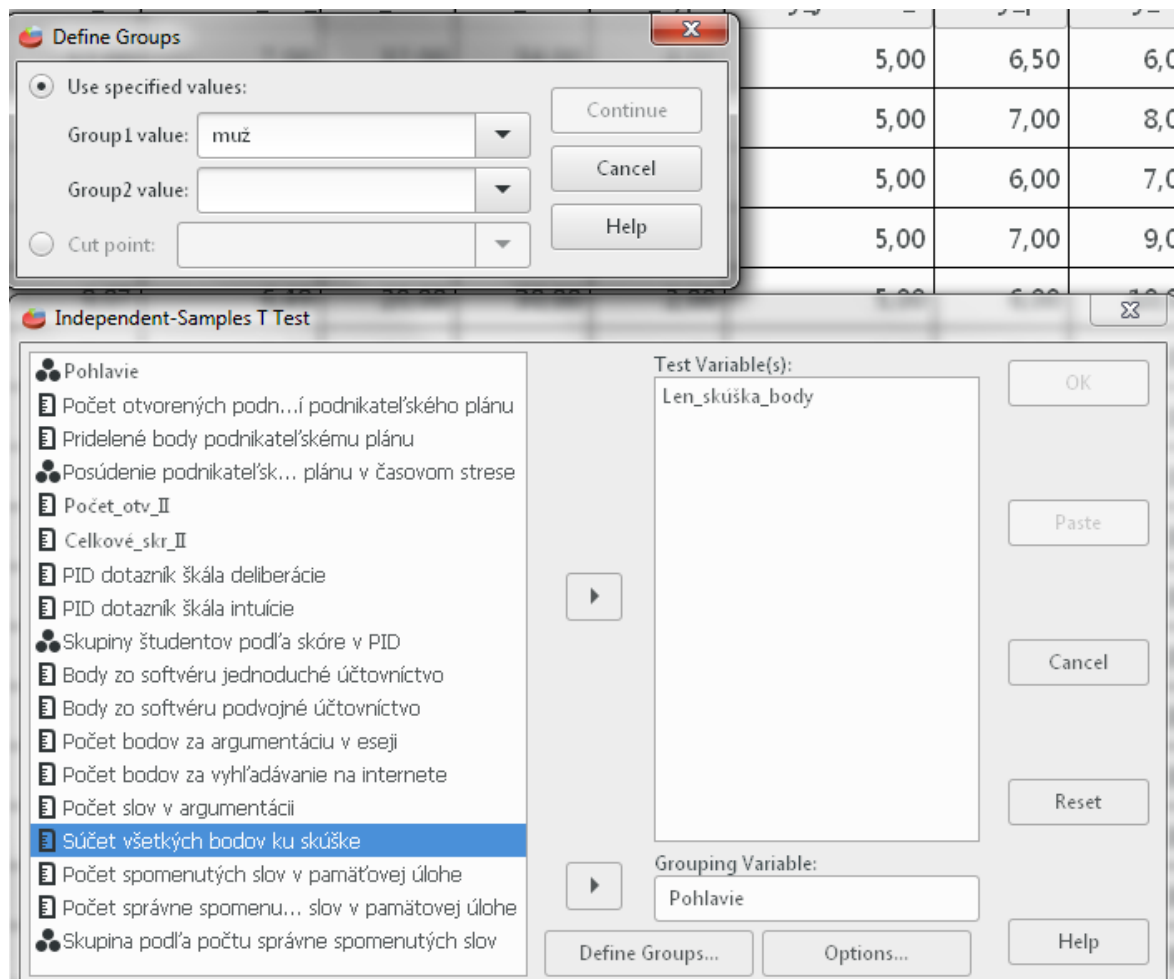
		Počet bodov z teoretickej skúšky
<i>N</i>		75
<i>Normal Parameters</i>	<i>Mean</i>	48,05
	<i>Std. Deviation</i>	10,24
<i>Most Extreme Differences</i>	<i>Absolute</i>	,08
	<i>Positive</i>	,07
	<i>Negative</i>	-,08
<i>Kolmogorov-Smirnov Z</i>		,68
<i>Asymp. Sig. (2-tailed)</i>		,745

Obr. 10.1.1 Výsledok Kolmogorovho – Smirnovho testu normality rozloženia dát

Keďže hodnota je Sigma je 0,745 čo je viac než 0,05, konštatujeme že test nie normálneho rozloženia nie je štatisticky významný. Môžeme preto pokračovať ďalej a

Klikneme na *Analyzovať (Analyze)* a vyberieme si *Porovnanie priemerov (Compare Means)* a potom si vyberieme *Dvojvýberový t – test (Independent Samples T)*. V dialógovom okne si vyberieme premennú: *Len_skuska_body* a presunieme ju vpravo do: *Testovanej premennej (Test Variable(s))*. Ďalej musíme vybrať skupiny, ktoré budeme porovnávať, v spodnej časti dialógového okna sa nachádzajú *Skupiny premennej (Grouping Variable)*, sem presunieme premennú *Pohlavie*. Ďalej klikneme na *Definovať skupiny (Define Groups)* a softvér nás v novom dialógovom okne požiada, aby sme vybrali dve skupiny.

Premennú pohlavie sme si definovali ako binárnu a teda nadobúda hodnoty 0 (muž) a 1(žena). Klikneme pokračovať a spustíme analýzu OK.



Obr. 10.1.2 Sprievodca pre Dvojvýberový t test s rovnosťou rozptylov (*Independent sample T- Test*)

Výsledok: Program PSPP uvádza dve tabuľky výsledkov. V prvej tabuľke sú uvedené počty mužov ($N = 22$) a opisné štatistiky ako priemer $Mean = 47,32$, smerodajná odchýlka $SD = 9,29$ a Smerodajná chyba priemeru $S.E. Mean = 1,98$. Rovnaké štatistické metódy a ich výsledky sú uvedené aj pre druhú skupinu – ženy, kde $Mean = 48,35$, $SD = 10,68$. Vidíme, že ženy dosiahli viac bodov zo skúšky, avšak to či je tento rozdiel štatisticky významný alebo nie, zistíme z druhej tabuľky.

V druhej tabuľke sú uvedené už konkrétne výsledky dvoch testov. Prvým je Levenov test, ktorý meria homogenitu rozptylov ako aj sféricitu dát v súbore. Najprv musí byť splnená podmienka, že naše dáta v skúmanom súbore neporušujú predpoklad homogenity rozptylu a sféricity, a až potom môžeme interpretovať výsledky samotného T – testu.

Výsledky Levenovho testu interpretujeme tak, že ak hodnota p je menšia než 0,05, potom nám test potvrdil, že sú porušené dané predpoklady. Potom musíme použiť neparametrický štatistický test na porovnanie dvoch skupín, konkrétne Mann-Whintey test. Ak je hodnota vyššia ako 0,05, až potom môžeme interpretovať druhú časť tabuľky, ktorá uvádza výsledky dvojvýberového t – testu. V našom prípade je v druhom stĺpci je uvedená hodnota Levenovho testu p (Sig.) = 0,355, čo je viac než podmienka 0,05. Môžeme teda interpretovať výsledky t – testu. Pri prvom pohľade skontrolujeme hodnotu p (Sig.), ktorá je v našom prípade 0,694, ak predpokladáme, že naše skupiny sa nebudú odlišovať (Equal variance assumed). To je oveľa viac než všeobecne akceptovaná hranica pre $p < 0,05$. To znamená, že ženy majú síce vyšší priemerný počet bodov ako muži, avšak tento rozdiel nie je štatisticky významný, a preto na základe týchto výsledkov zamietame hypotézu H_1 a prijímame nulovú hypotézu.

Group Statistics

	Pohlavie	N	Mean	Std. Deviation	S.E. Mean
Počet bodov z teoretickej skúšky	muž	22	47,32	9,29	1,98
	žena	53	48,35	10,68	1,47

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Počet bodov z teoretickej skúšky	Equal variances assumed	,87	,355	-,40	73,00	,694	-,103	2,61	-,624	4,17
	Equal variances not assumed			-,42	44,92	,677	-,103	2,46	-,600	3,93

Obr. 10.1.3 Výsledky pre Dvojvýberový t test s rovnosťou rozptylov (*Independent sample T- Test*)

Vo výsledkoch však nie je uvedená veľkosť účinku a ako bolo vysvetlené v kapitole veľkosť účinku, a preto si túto budeme musieť dopočítať samostatne. Veľkosť účinku (effect size) vypočítame nasledovne.

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

$$r = \sqrt{\frac{(-0,4)^2}{(-0,4)^2 + 73}}$$

$$r = \sqrt{\frac{0,16}{73,16}}$$

$$r = 0,047$$

Ako vidíme hodnota $r = 0,047$ je veľmi malá (menšia než 0,1) a je blízka 0, čo znamená minimálny, respektíve neexistujúci efekt.

Uvádzanie výsledkov: Muži ($M = 47,32$; $SD = 9,29$) sa štatisticky významne neodlišujú ($t(73) = -0,4$; $p = 0,69$; $r = 0,047$) od žien ($M = 48,35$; $SD = 10,68$) v počte bodov z teoretickej skúšky.

10.1.3 Analýza rozptylu -variability (ANOVA: Analysis of variance)

Je parametrická štatistická metóda vytvorená na vzájomné porovnávanie skupín, ktorých počet je viac než dve. Zjednodušene by sme mohli povedať, že je to t test pre viac ako dve skupiny. Existuje viacero typov analýzy rozptylu (viacero ANOV), no my sa v tejto časti zameriame na základnú a najjednoduchšiu z nich - ANOVU pre 1 faktor (*One way ANOVA*). Inak sa nazýva aj Jednofaktorová analýza rozptylu. Pri tejto ANOVE porovnáваме viacero skupín (viac než 2) a hľadáme či sa štatisticky významne odlišujú len v 1 faktore (Meloun, Militký, 2012b, str. 530).

1 úloha: V príklade č. 7. Canvas metódy a startup je opísaný podnikateľský zámer pomocou metódy Canvas. Podnikatelia, žiadajúci o financie od nezávislých investorov, opísali svoj zámer a ten bol následne obodovaný pomocou metódy Canvas. Podľa toho ako dobre, podrobne a odborne opísali svoj zámer, dostávali body za jednotlivé kategórie ako napríklad segmenty zákazníkov, ponúkaná hodnota a ostatné, ktoré sa sčítali a vzniklo Celkové skóre Canvas, ktoré je v našom príklade č. 7 posledný stĺpec s názvom: *Canvas_total_score*. Čím je toto skóre vyššie, tým podrobnejší, odbornejší a kvalitnejší podnikateľský plán predložili. Rôzni podnikatelia mali značne odlišné skúsenosti, a pre väčšinu z nich bol tento ich zámer vôbec prvým pokusom začať podnikat'. Iní už vlastnili existujúcu firmu a niektorí mali aj dve, či dokonca tri firmy. V stĺpci s názvom: N je uvedený počet firiem, ktoré podnikatelia založili v minulosti. Na základe počtu založených firiem rozdelíme podnikateľov do 4 skupín (0 - nemá firmu, 1 - založil jednu firmu, 2 - dve firmy, 3 - tri firmy). Zaujímá nás, či tí, ktorí majú skúsenosti v podnikaní (založili 1 alebo viac firiem), dokážu lepšie vypracovať podnikateľské zámery a plány a dosiahli iný počet bodov z Canvas. Predpokladáme, že premenná: *Celkové Canvas skóre* sa bude medzi skupinami líšiť. Na základe týchto predpokladov sme si stanovili nasledovné hypotézy:

H_1 : Podnikatelia rozdelení do štyroch skupín podľa počtu založených firiem sa štatisticky významne líšia v dosiahnutom *Celkovom Canvas skóre*.

H_0 : Neexistuje štatisticky významný rozdiel v *Celkovom Canvas skóre* medzi podnikateľmi, ktorí založili rôzny počet firiem.

Riešenie: Postup riešenia je nasledovný:

1. Krok: Overiť predpoklady na použitie parametrického testu ANOVA (normalita rozloženia dát, homogenita rozptylov, kardiálna premenná, nezávislosť meraní u skúmanej premennej *Celkové Canvas skóre*).

2. Krok: Ak sú splnené predpoklady parametrických testov, potom použijeme parametrickú metódu na porovnanie skupín, v našom prípade Analýzu rozptylu – ANOVU, ak nie sú, potom použijeme neparametrickú metódu pre viac skupín než dve, a tou je Kruskal – Walisov Test.

Prvým krokom je teda overiť normalitu rozloženia dát. Použijeme na to Kolmogorov – Smirnov test. Postup pri tomto teste opisujeme napríklad v záložke na stránke s názvom: *Meranie vzájomných vzťahov u ordinálnych premenných*. A preto tu uvádzame už iba výsledok. Z výsledku vidíme, že hodnota Sigma je 0,476, čo je oveľa viac než požadovaná hladina významnosti 0,05, čo znamená, že test nie je štatisticky významný a teda dáta sú rozložené normálne. Naša premenná je pomerová (počet bodov môže byť 0 až 100) a zároveň každé meranie je nezávislé, takže ďalšie predpoklady spĺňame. Čo sa týka posledného predpokladu, a teda homogenitu rozptylov i sféricitu zistíme na základe výsledkov Levenovho testu, ktorý je súčasťou výsledkov ANOVY, a preto môžeme už aplikovať parametrický test ANOVA.

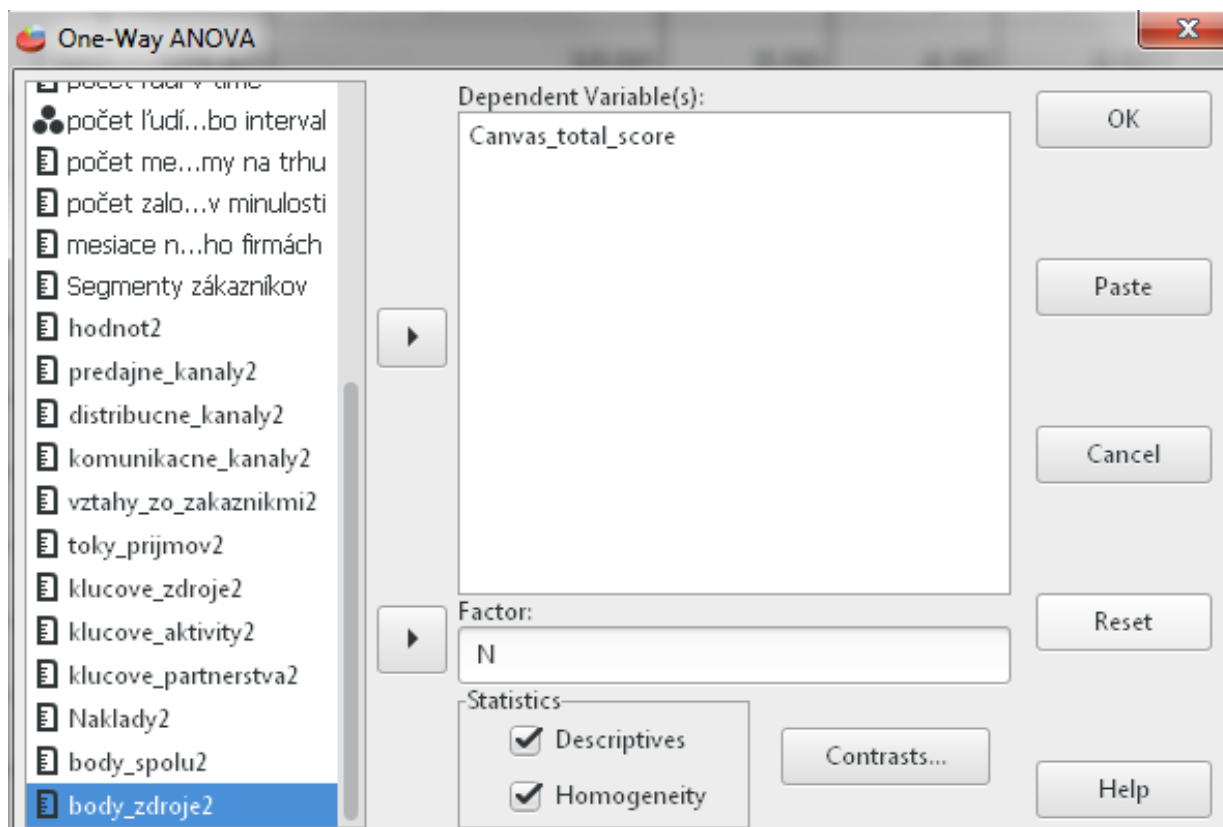
One-Sample Kolmogorov-Smirnov Test

		<i>Celkové skóre canvas</i>
<i>N</i>		200
<i>Normal Parameters</i>	<i>Mean</i>	35,08
	<i>Std. Deviation</i>	15,30
<i>Most Extreme Differences</i>	<i>Absolute</i>	,06
	<i>Positive</i>	,06
	<i>Negative</i>	-,05
<i>Kolmogorov-Smirnov Z</i>		,84
<i>Asymp. Sig. (2-tailed)</i>		,476

Obr. 10.1.4 Výsledok Kolmogorovho – Smirnovho testu.

Krok 2. Postup pri parametrickom teste ANOVA je nasledovný.

Klikneme na *Analyzovať (Analyze)* a potom na *Porovnať priemery (Compare Means)* a vyberieme *Jednofaktorovú Analýzu rozptylu (One – Way ANOVA)*. V sprievodcovi presunieme premennú *Canvas_total_score* do okna *Závislé premenné (Dependent variable(s))*. Ďalej vyberieme premennú *Počet založených firiem v minulosti – N* a presunieme ju do políčka *Faktor (Factor)*. V zaškrtávacích políčkach dole zaškrtneme *Opisné štatistiky (Descriptives)* a *test Homogenity dát (Homogeneity)*. Na záver klikneme *Ok*.



Obr. 10.1.5 Sprievodca Analýzou rozptylu (ANOVA)

Výsledky: Program PSPP ako výsledok vytvorí tri tabuľky. V prvej tabuľke sa nachádzajú opisné štatistiky pre každú skupinu. Máme 4 skupiny žiadateľov čo sa týka ich predchádzajúcich skúseností v podnikaní. Prvá skupina je najpočetnejšia ($n = 136$), označili sme ju 0, a táto nezaložila v minulosti žiaden podnik. Ich priemerné skóre v bodovaní Canvas je: $M = 32,13$ bodu, smerodajná odchýlka $SD = 14,63$ a v riadku sú uvedené aj ďalšie opisné štatistiky pre skupinu 0. V druhom riadku sú uvedené opisné štatistiky pre Celkové Canvas skóre u skupiny podnikateľov, ktorí založili 1 firmu a tret'om i štvrtom riadku je princíp rovnaký ako v prvých dvoch. Keď porovnáme priemery (stĺpec Mean v prvej tabuľke) vidíme, že skupiny sa medzi sebou pomerne výrazne líšia v priemerných hodnotách z *Celkového Canvas skóre*, otázkou však je, či tieto rozdiely sú aj štatisticky a následne aj vecne významné. Na to nám odpovie tretia tabuľka a výpočet vecnej významnosti.

V druhej tabuľke sú výsledky testu pre normalitu rozloženia dát, v tomto prípade výsledky Levenovho testu = 0,3 pri hladine významnosti $p = 0,992$. Táto úroveň hladiny významnosti je vysoko nad podmienkou $p < 0,05$, a teda test je štatisticky nevýznamný, čo znamená, že dáta pre premennú Celkové skóre canvas neporušujú predpoklad homogenity rozptylov a sféricity. Tieto výsledky sú spolu s výsledkami Kolmogorovho – Smirnovho testu potvrdením, že boli splnené predpoklady na aplikovanie parametrického testu.

Tretia tabuľka obsahuje výsledky samotnej ANOVY a je pre nás najdôležitejšia. Hovorí nám o tom, či sa naše 4 skupiny medzi sebou štatisticky významne líšia. V poslednom stĺpci Sig. je uvedená hladina významnosti, ktorá je menej než 0,001, čo znamená, že skupiny sa medzi sebou štatisticky významne líšia čo sa týka premenej: *Celkové skóre Canvas*. Na základe týchto výsledkov prijímame prvú hypotézu H_1 a zamietame nulovú hypotézu. Hoci skupiny sa medzi sebou štatisticky významne líšia, musíme dopočítať aj vecnú významnosť (veľkosť účinku – efekt size), aby sme zistili aký veľký je rozdiel.

Descriptives									
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
Celkové skóre canvas	,00	136	32,13	14,63	1,25	29,65	34,61	1,00	65,00
	1,00	38	40,63	14,81	2,40	35,76	45,50	5,00	73,00
	2,00	19	38,47	13,86	3,18	31,79	45,15	18,00	60,00
	3,00	7	52,86	14,99	5,67	38,99	66,72	26,00	70,00
	Total	200	35,08	15,30	1,08	32,94	37,21	1,00	73,00

Test of Homogeneity of Variances				
	Levene Statistic	df1	df2	Sig.
Celkové skóre canvas	,03	3	196	,992

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
Celkové skóre canvas	Between Groups	4783,82	3	1594,61	7,47	,000
	Within Groups	41824,05	196	213,39		
	Total	46607,87	199			

Obr. 10.1.6 Výsledok analýzy rozptylu (ANOVA)

Veľkosť účinku (effect size)

Vypočítame ju z nasledovného vzorca, kde dáta použijeme z poslednej tabuľky výsledkov, kde rozptyl medzi skupinami (Between Groups) sa označuje aj ako $SS_M = 4783,82$ a celkový rozptyl (Total) $SS_T = 46\ 607,87$. Dosadením do vzorca a vypočítaním dostaneme hodnotu $r = 0,32$, čo je stredná hodnota účinku (je väčšia ako 0,3).

$$r^2 = \frac{SS_M}{SS_T}$$

$$r^2 = \frac{4783,82}{46\ 607,87}$$

$$r = \sqrt{0,1026}$$

$$r = 0,32$$

Uvádzanie výsledkov: Jednotlivé skupiny podnikateľov, vytvorené podľa počtu založených firiem v minulosti, sa vzájomne štatisticky významne líšia v tom, aké Celkové skóre Canvas dosiahol nimi predložený podnikateľský plán, $F(3,196) = 7,47$, $p < 0.001$, so stredným efektom účinku $r = 0,32$.

10.1.4 Párový t test

Párový t test používame na porovnanie dvoch hodnôt premennej od toho istého respondenta.

1 úloha: Študentom manažmentu bolo ukázaných 25 podnikateľských charakteristík, na základe ktorých je možné pomerne dôkladne opísať podnikateľský plán ako napr. cena, ziskovosť, súčasná konkurencia a podobne. Potom boli o 3 týždne neskôr požiadaní, aby si spomenuli na čo najviac podnikateľských charakteristík a zapísali si ich. Tieto zápisy boli následne porovnávané, či sa zhodujú so skutočnosťou a bol zaznamenaný počet tých ktoré boli správne. Zaujímá nás ako veľa toho študenti zabudli a či existuje štatisticky významný rozdiel medzi počtom všetkých spomenutých a počtom tých správne spomenutých.

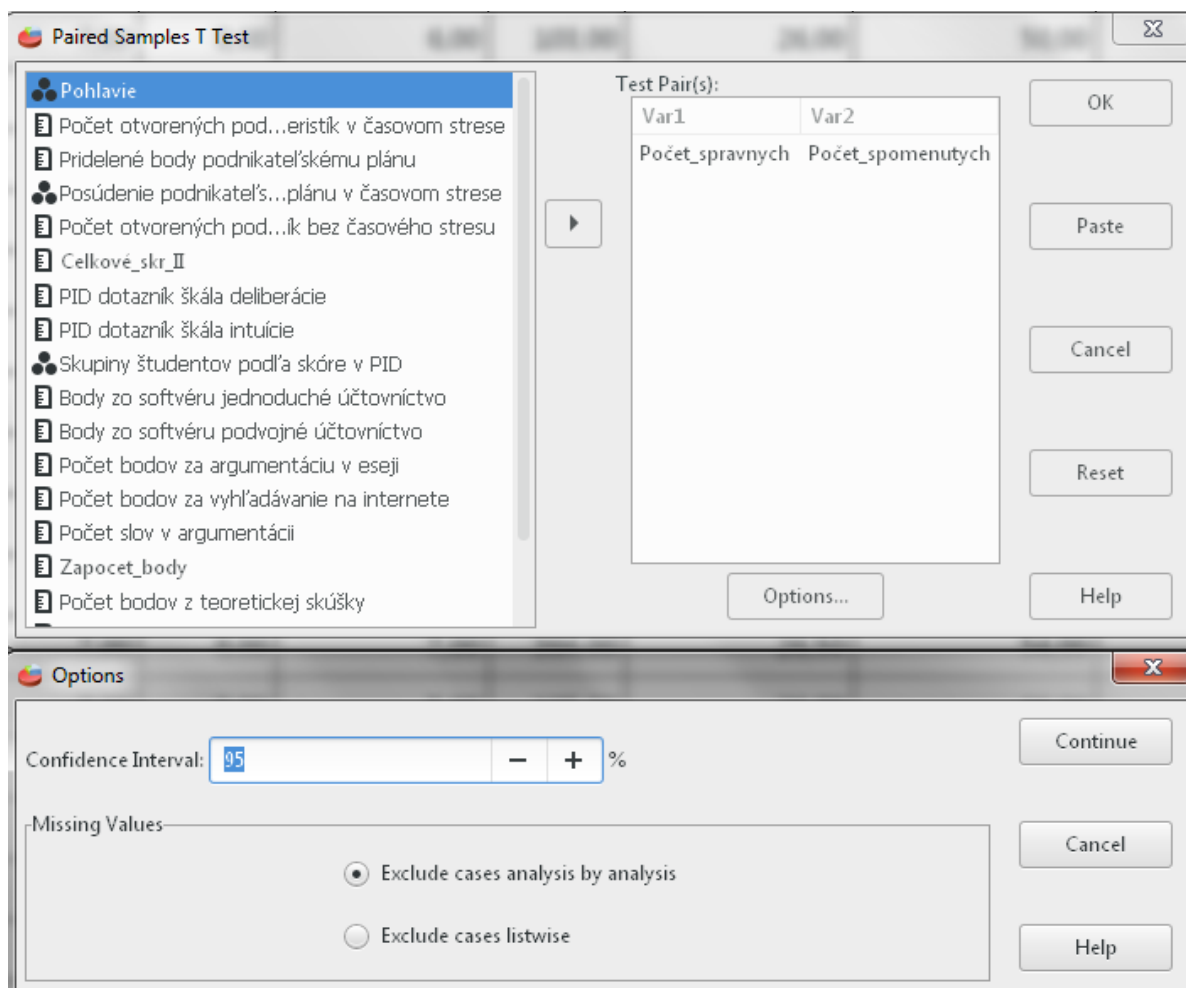
Budeme teda porovnávať dve premenné u každého respondenta a to počet všetkých spomenutých a počet správne spomenutých.

Riešenie: Podľa toho ako budú dáta rozložené použijeme: A) ak budú rozložené normálne použijeme Párový t- test, ak nebudú potom použijeme Wilcoxonov test. V prvom kroku otestujeme podmienky na použitie parametrického testu. Postup je rovnaký ako u hore uvedenej ANOVY. Začneme s overením rozloženia dát pomocou Kolmogorovho-Smirnovho testu. Tento sme už v učebnici opísali a tak teraz len uvedieme jeho výsledok. Ako vidíme u obdivoch skúmaných premenných je hodnota Sigmy väčšia než 0,05 a tak môžeme použiť parametrický Párový t test. Taktiež naše dáta sú merané vo forme pomerovej premennej. Podmienky pre aplikáciu parametrického testu sú splnené.

	Počet spomenutých slov v pamäťovej úlohe	Počet správne spomenutých slov v pamäťovej úlohe
N		75
Normal Parameters	Mean	6,64
	Std. Deviation	3,03
Most Extreme Differences	Absolute	,15
	Positive	,15
	Negative	-,13
Kolmogorov-Smirnov Z		1,27
Asymp. Sig. (2-tailed)		,100

Obr. 10.1.7 Výsledok Kolmogorovho-Smirnovho testu

Ďalej budeme postupovať nasledovne. Klikneme na *Analyzovať (Analyze)* a potom na *Porovnať priemery (Compare Means)* a vyberieme *Párový t test (Paired Sample t - test)*. Vyberieme premenné: *Počet spomenutých slov v pamäťovej úlohe (Počet_spomenutych)* a *Počet správne spomenutých (Pocet_spravnych)*. Ďalej klikneme na *Možnosti (Options)* a tu ponecháme nastavený *Interval spoľahlivosti (Confidence Interval)* na úrovni 95 %. Potom klikneme *Pokračovať (Continue)* a *Ok*.



Obr. 10.1.8 Sprievodca párovým testom

Výsledky:

Výsledkom testu sú tri tabuľky. Prvá uvádza deskriptívne štatistiky pre skúmanú skupinu. Vidíme že študent si priemerne celkovo spomenul na $M = 6,64$ podnikateľských charakteristík, $SD = 3,03$. Avšak správnych bolo priemerne len $3,54$, $SD = 1,91$, čo približne len polovica.

V druhej tabuľke je uvedená vzájomná korelácia medzi počtom správne spomenutých a všetkých spomenutých. Korelačný koeficient $r = 0,65$; $p < 0,001$. Výsledky samotného párového testu sú uvedené v tabuľke č. 3. V prvom stĺpci (Mean) je uvedený rozdiel medzi celkový počet spomenutých a správne spomenutých čo je $3,11$. Rozdiel medzi skúmanými premennými, $t(74) = -11,73$ je štatisticky významný kde $p < 0,001$.

Paired Sample Statistics

	Mean	N	Std. Deviation	S.E. Mean
Pair 1 Počet správne spomenutých slov v pamätovej ulohe	3,54	75	1,91	,22
Počet spomenutých slov v pamätovej ulohe	6,64	75	3,03	,35

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 Počet správne spomenutých slov v pamätovej ulohe & Počet spomenutých slov v pamätovej ulohe	75	,65	,000

Paired Samples Test

	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Paired Differences				
				Lower	Upper			
Pair 1 Počet správne spomenutých slov v pamätovej ulohe - Počet spomenutých slov v pamätovej ulohe	-3,11	2,29	,26	-3,63	-2,58	-11,73	74	,000

Obr. 10.1.9 Výsledok párového testu

Dopočítanie vecnej významnosti (veľkosti účinku – effect size) Nakoľko program PSPP neuvádza veľkosť účinku pre párový test, avšak pri uvádzaní výsledkov sa vyžaduje, a preto ho musíme dopočítať samostatne. Vzorec na výpočet je rovnaký ako pri dvojvýberovom t test s rovnosťou rozptylov (independent sample t - test). V našom prípade je postup nasledovný:

Veľkosť účinku výpočet (effect size):

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

$$r = \sqrt{\frac{(-11,73)^2}{(-11,73)^2 + 74}}$$

$$r = \sqrt{\frac{137,59}{211,59}}$$

$$r = 0,806$$

Uvádzanie výsledkov: Študenti (n =75) si priemerne spomenuli na 6,64 podnikateľských charakteristik, SD = 3,03, avšak v priemere bolo len 3,54 správne spomenutých SD = 1,91. Tento rozdiel je štatisticky významný, t (74) = -11,73 p < 0,001 s veľmi silným efektom účinku r =0,81.

10.1.5 Jednovýberový t – test (One sample T - Test)

Ako už bolo v úvode kapitoly spomenuté, tento test používame na porovnanie výsledkov s konštantou, respektíve s hodnotou priemeru, ktorá nám je známa. Pomocou testu zistíme, či sa priemer pre našu skupinu štatisticky významne líši od už nám známeho priemeru.

1 úloha: V súbore s názvom Príklad č. 2 Študenti body a pamäť máme výsledky študentov zo skúšky a ich celkový počet bodov uvedených ako premenná s názvom *Skúška_celkom*. Z minulého roka máme priemerný počet bodov zo skúšky, ktorý bol 70. Zaujímá nás či sa priemerný počet bodov z minulého roka (70 bodov) štatisticky významne odlišuje od priemerného počtu bodov, ktorý získali študenti v tomto roku. Na výpočet použijeme *Jednovýberový t test*.

Hypotézy si stanovíme nasledovne:

H_1 : Priemerný počet bodov zo skúšky v tomto roku sa štatisticky významne odlišuje od priemerného počtu bodov v minulom roku, ktorý bol 70.

H_0 : Nie je štatisticky významný rozdiel v priemernom počte dosiahnutých bodov v tomto roku oproti priemeru z minulého roka.

Riešenie:

1. Krok : Overenie predpokladov (normalita rozloženia, homogenita rozptylov, sféricita, kardinálna premenná a nezávislosť merania) na to, aby sme mohli použiť parametrický *Jednovýberový t test (One Sample t - test)*.

2. Krok : Ak budú splnené predpoklady, až potom môžeme použiť parametrický *Jednovýberový t test*.

Najprv overíme normalitu pomocou Kolmogorovho Smirnovho testu. Hladina významnosti je $p = 0,36$, čo je viac než 0,05 a teda test nie je štatisticky významný. Preto môžeme konštatovať, že naša dáta sú rozložené normálne. Naše dáta sú merané na pomerovej škále, teda študent mohol mať súčet všetkých bodov aj 0 aj 100. Študenti sa vzájomne neovplyvňovali a tak môžeme konštatovať nezávislosť merania. Na základe hore uvedeného tvrdíme, že predpoklady pre parametrické testy boli splnené a môžeme použiť parametrický *Jednovýberový t test*. Postup je nasledovný.

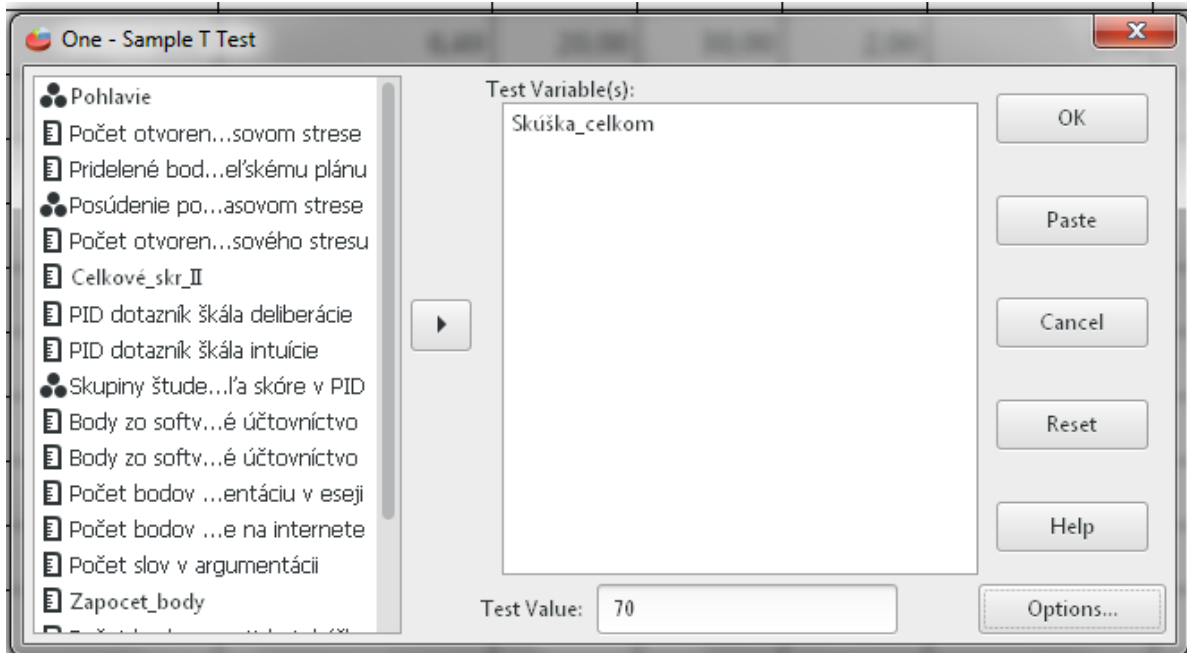
One-Sample Kolmogorov-Smirnov Test

		<i>Súčet všetkých bodov ku skúške</i>
<i>N</i>		75
<i>Normal Parameters</i>	<i>Mean</i>	73,75
	<i>Std. Deviation</i>	13,40
<i>Most Extreme Differences</i>	<i>Absolute</i>	,11
	<i>Positive</i>	,05
	<i>Negative</i>	-,11
<i>Kolmogorov-Smirnov Z</i>		,92
<i>Asymp. Sig. (2-tailed)</i>		,360

Obr. 10.1.10 Výsledok Kolmogorovho – Smirnovho testu

Klikneme na *Analyzovať (Analyze)* a potom na *Porovnať priemery (Compare Means)* a vyberieme *Jednovýberový t test (One Sample t - test)*. Vyberieme

premennú: *Skúška_celkom* a presunieme ju vpravo do okna *Testované premenné (Test Variables)*. Potom dole nastavíme Testovanú hodnotu (Test Value) na 70. Následne klikneme Ok.



Obr. 10.1.11 Sprievodca Jednovýberovým t testom

Výsledky: Výsledkom sú dve tabuľky.

One-Sample Statistics				
	N	Mean	Std. Deviation	S.E. Mean
Súčet všetkých bodov ku skúške	75	73,75	13,40	1,55

One-Sample Test						
	Test Value = 70,000000					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Súčet všetkých bodov ku skúške	2,42	74	,018	3,75	,66	6,83

Obr. 10.1.12 Výsledky Jednovýberového t testu

V prvej sú uvedené opisné štatistiky pre našu vzorku študentov ($n = 75$), ako priemer bodov zo skúšky $M = 73,75$, štandardná smerodajná odchýlka $SD = 13,4$. V druhej tabuľke sú uvedené výsledky t testu. V stĺpci s názvom Rozdiel Priemerov (Mean Difference) vidíme, že rozdiel medzi celkovými bodmi v minulom roku a našou vzorkou je 3,75. V stĺpci Sigma vidíme $p = 0,018$, čo znamená, že priemer našej skupiny sa štatisticky významne líši od testovanej hodnoty 70 bodov. Teda priemerné hodnoty v tomto roku sú štatisticky významne odlišné od priemerných hodnôt v minulom roku. Na základe týchto výsledkov zamietame nulovú hypotézu H_0 a prijímame H_1 , ktorá tvrdí, že priemery sú odlišné.

Na výpočet vecnej významnosti- veľkosti účinku použijeme Cohenovo d

$$\text{Cohenovo } d = \frac{\text{Rozdiel priemerov (Mean Difference)}}{\text{Štandardná smerodajná odchýlka SD (Std. Deviation)}}$$

$$\text{Cohenovo } d = \frac{3,75}{13,4}$$

$$\text{Cohenovo } d = 0,2798$$

Uvádzanie výsledkov: Priemerná hodnota bodov zo skúšky $M = 73,75$; $SD = 13,4$. sa od testovanej hodnoty 70 líši štatisticky významne $t(74) = 2,42$; $p = 0,018$ so strednou veľkosťou účinku $r = 0,28$.

10.2 Neparametrické testy

V prípade ak naše dáta nie sú rozložené normálne, potom musíme na porovnanie použiť neparametrické testy. Zjednodušene by sme mohli povedať že neparametrické testy sú ekvivalentom pre parametrické pre nie normálne rozložené dáta.

10.2.1 Wilcoxonov test

Jeden z dostupných neparametrických testov, ktoré program PSPP ponúka je **Wilcoxonov test**. Tento test je ekvivalentom párového t testu.

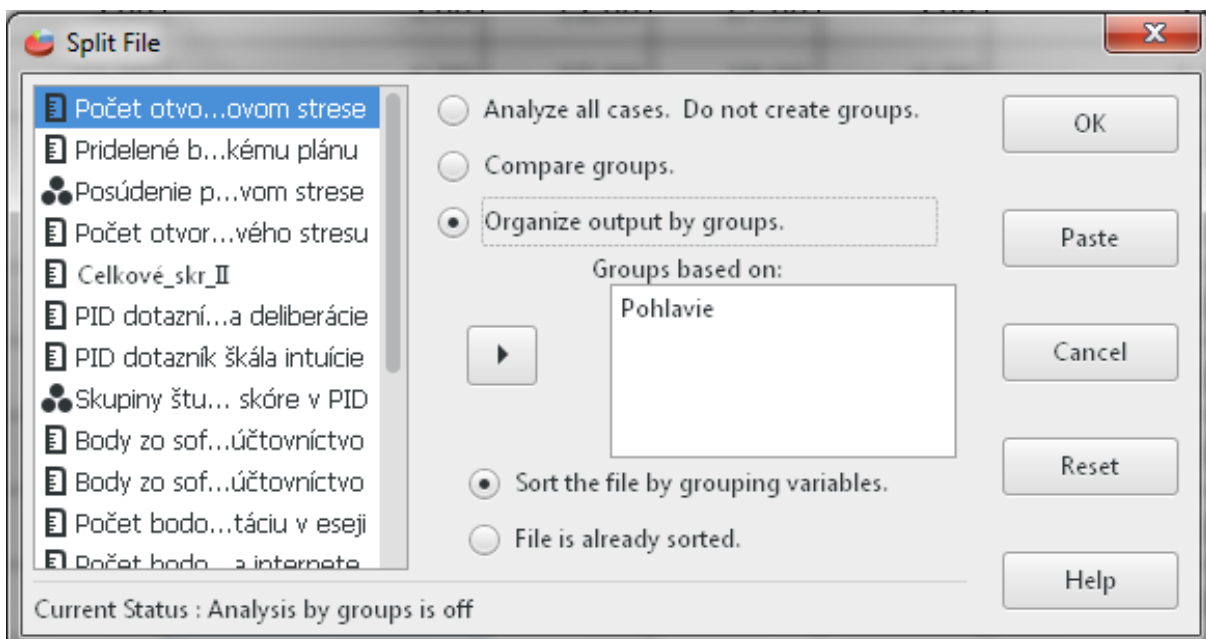
1 úloha: Porovnajete počet pridelených bodov podnikateľskému plánu (celkové hodnotenie) ženami v podmienke časového stresu a v podmienke bez časového stresu. Zaujímá nás, či ženy v prvom a v druhom hodnotení celkové skóre pre podnikateľský plán zmenili. Konkrétnejšie predpokladáme, že v podmienke bez časového stresu budú mať dostatok času na prečítanie všetkých podnikateľských charakteristík, nájdu také, ktoré majú nízku úroveň a potom budú hodnotiť podnikateľský plán negatívnejšie a pridelia mu menej bodov ako v časovom strese. V tomto prípade máme dve hodnotenia od jednej osoby, preto musíme použiť párový t test ak budú dáta rozložené normálne, alebo Wilcoxonov test ak nebudú rozložené normálne. Na základe hore uvedeného si hypotézy stanovíme nasledovne.

H_1 : Ženy v podmienke časového stresu ohodnotili podnikateľský plán pozitívnejšie ako v podmienke bez časového stresu.

H_0 : Celkové hodnotenie podnikateľského plánu ženami sa medzi podmienkami štatisticky významne nelíši.

Riešenie: Na to aby sme mohli z príkladu č. 2 vypočítať hodnoty pre ženy, musíme súbor rozdeliť na dve skupiny: muži a ženy. Je viacero spôsobov akým môžeme postupovať, ale za jeden z najjednoduchších považujem nástroj programu PSPP na rozdelenie súboru a následne aj výsledkov podľa jednotlivých skupín. Softvér takto bude počítat z hodnôt pre konkrétne skupiny, ktoré mu zadáme a výsledky bude uvádzať podľa jednotlivých skupín (muži a ženy). Postup na rozdelenie súboru na skupiny je nasledovný.

Klikneme na *Dáta (Data)* a vyberieme *Rozdeliť súbor (Split file)*. Zobrazí sa nám sprievodca, kde zaškrtneme *Organizovať výstup podľa skupín (Organize output by groups)* a presunieme premennú *Pohlavie* vpravo do kategórie *Skupiny založené na (Groups based on)*. Na záver klikneme *Ok*. Odteraz budeme mať všetky výsledky testov zobrazené samostatne za každú skupinu. Ak budeme chcieť vrátiť pôvodný stav, kedy bude uvádzať jeden výsledok pre celý súbor, potom zaklikneme príkaz: *Analyzuj všetky prípady. Nevytváraj skupiny (Analyze all cases. Do not create groups)*.



Obr. 10.2.1 Sprievodca rozdelením súboru podľa pohlavia

Keď sme skupiny skúmaný súbor rozdelili na dve skupiny podľa pohlavia, potom môžeme overiť, ako sú dáta rozložené v jednotlivých skupinách. V prípade, ak sú rozložené normálne, potom použijeme Párový t test a ak nie sú tak použijeme Wilcoxonov test. Na overenie normality rozloženia dát použijeme Kolmogorovov – Smirnovov test.

One-Sample Kolmogorov-Smirnov Test

		<i>Pridelené body podnikateľskému plánu</i>	<i>Celkové_skr_II</i>
<i>N</i>		22	22
<i>Normal Parameters</i>	<i>Mean</i>	6,86	6,75
	<i>Std. Deviation</i>	,99	1,64
<i>Most Extreme Differences</i>	<i>Absolute</i>	,24	,15
	<i>Positive</i>	,22	,12
	<i>Negative</i>	-,24	-,15
<i>Kolmogorov-Smirnov Z</i>		1,11	,71
<i>Asymp. Sig. (2-tailed)</i>		,153	,692

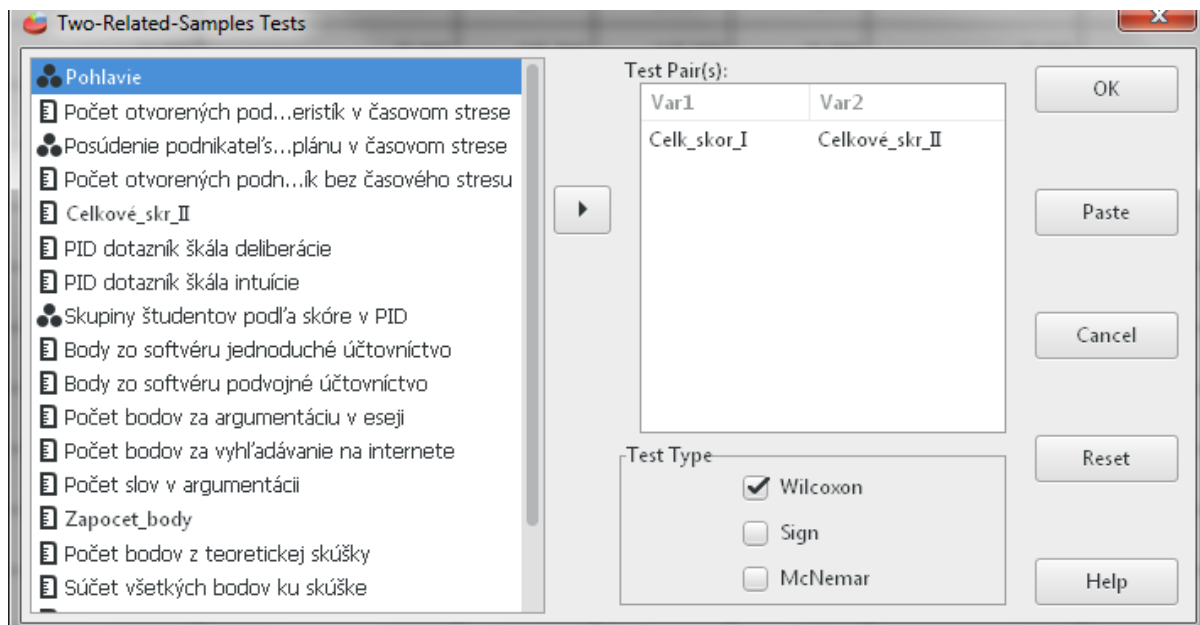
One-Sample Kolmogorov-Smirnov Test

		<i>Pridelené body podnikateľskému plánu</i>	<i>Celkové_skr_II</i>
<i>N</i>		53	51
<i>Normal Parameters</i>	<i>Mean</i>	6,87	6,35
	<i>Std. Deviation</i>	1,18	1,25
<i>Most Extreme Differences</i>	<i>Absolute</i>	,26	,19
	<i>Positive</i>	,17	,16
	<i>Negative</i>	-,26	-,19
<i>Kolmogorov-Smirnov Z</i>		1,90	1,38
<i>Asymp. Sig. (2-tailed)</i>		,001	,033

Obr. 10.2.2 Výsledky Kolmogorovho- Smirnovho testu pre rozdelený súbor na skupinu muži (n = 22) a ženy (n = 53)

Výsledkom sú dve tabuľky, prvá je pre mužov, ktorých počet je 22 a druhá je pre ženy, ktorých počet je 53. Ako vidíme, v prvej tabuľke (muži) Kolmogorov Smirnov test nie je štatisticky významný, konštatujeme, že dáta sú rozložené normálne. Naopak v druhej skupine (ženy) je hladina významnosti 0,01 pre Pridelené body v prvom hodnotení a 0,033 pre Celkové_str_II. Test je štatisticky významný a teda dáta nie sú rozložené pre skupinu žien normálne. Zároveň v tejto tabuľke vidíme priemerné hodnotenie podnikateľského plánu (pridelené body) v dvoch podmienkach.

Keďže naše dáta nie sú rozložené normálne na porovnanie prvého a druhého hodnotenia u žien použijeme Wilcoxonov test. Budeme postupovať nasledovne. Klikneme na *Analyzovať (Analyze)* a vyberieme si *Neparametrické Štatistiky (Non-parametric statistics)* a ďalej *Testy pre dve súvisiace skupiny (Two – Related – Samples Tests)*. Následne presunieme premenné *Celk_skor_I* a *Celkove_skr_II* do okna *Testované Páry (Test Pair(s))* a v spodnej časti okna zaškrtneme *Wilcoxonov test (Wilcoxon)*. Potom klikneme Ok. Výsledky tohto testu budú znova rozdelené na tabuľky pre ženy a tabuľky pre mužov.



Obr. 10.2.3 Sprievdca Wilcoxonovým testom

Výsledky: Keďže sme na začiatku rozdelili súbor (Príkaz Split file) na dve skupiny, tak softvér PSPP vytvorí dve tabuľky výsledkov pre každú skupinu (muži a ženy), teda spolu 4. Prvá tabuľka sú výsledky pre mužov, avšak nakoľko výsledky Kolmogorovho Smirnovho testu ukázali, že dáta sú rozložené normálne, mali by sme túto skupinu testovať parametrickým párovým t testom a nie Wilcoxonovým testom. Posledné dve tabuľky sú pre ženy.

Štvrtá tabuľka sú výsledky samotného Wilcoxonovho testu kde $Z = -2,28$, a sú štatisticky významné $p = 0,023$. Na základe týchto výsledkov prijímame hypotézu H_1 a zamietame nulovú hypotézu.

Ranks

		N	Mean Rank	Sum of Ranks
Pridelené body podnikateľskému plánu - Celkové_skr_II	Negative Ranks	8	6,94	55,50
	Positive Ranks	7	9,21	64,50
	Ties	7		
	Total	22		

Test Statistics

	Pridelené body podnikateľskému plánu - Celkové_skr_II
Z	-,26
Asymp. Sig. (2-tailed)	,796

Ranks

		N	Mean Rank	Sum of Ranks
Pridelené body podnikateľskému plánu - Celkové_skr_II	Negative Ranks	12	18,00	216,00
	Positive Ranks	26	20,19	525,00
	Ties	13		
	Total	51		

Test Statistics

	Pridelené body podnikateľskému plánu - Celkové_skr_II
Z	-2,28
Asymp. Sig. (2-tailed)	,023

Obr. 10.2.4 Výsledok Wilcoxonovho testu pre rozdelený súbor muži a ženy

Vecnú významnosť (veľkosť účinku – effect size) vypočítame pre Wilcoxonov test nasledovne:

$$r = \frac{Z}{\sqrt{N}}$$

$$r = \frac{-2,28}{\sqrt{51}}$$

$$r = -0,32$$

Uvádzanie výsledkov: Ženy ohodnotili rovnaký podnikateľský plán v podmienke časového stresu pozitívnejšie ako bez časového obmedzenia a výsledky boli štatisticky významné $Z = -2,28$, $p = 0,023$ so stredne silným efektom účinku $r = -0,32$.

11 ZÁVER

Kvalitná analýza dát je základným predpokladom pre napísanie hodnotného vedeckého článku, vypracovanie záverečnej práce na úrovni alebo klúčom, ktorý otvára cestu k zahraničným publikáciám. Požiadavky na univerzitných študentov z hľadiska vedeckej hĺbky a úrovne rastú na všetkých stupňoch štúdia. Tieto trendy sa budú v budúcnosti len posilňovať.

Študenti i ostatní, ktorí mali záujem o analýzu však dlhodobo čelili problému nedostatku, respektíve úplnej absencie kvalitného softvéru, ktorý by bol zdarma. Softvér PSPP je opensource softvér umožňujúci základné až mierne pokročilé štatistické analýzy. Je stabilný, rýchly, hardvérovo nenáročný a neustále sa zlepšujúci.

Predkladaná publikácia spolu s webstránkou <http://statistikapspp.sk/> zaplňajú medzeru na trhu a umožňujú užívateľom softvér PSPP zvládnuť a pochopiť základné princípy štatistickej analýzy.

Veríme, že čitatelia a študenti sociálnych vied túto knihu ocenia a stane sa pre nich prvou štatistickou knihou, ktorá otvára cesty a záujem o podrobnejšie vysvetlenie a hlbšie pochopenie štatistických metód a postupov.

12 POUŽITÁ LITERATÚRA

1. Anýžová, P. (2015). Srovnatelnost postojových škál v komparativním výzkumu. Univerzita Palackého v Olomouci. ISBN 978-80-244-4447-5
2. Betsch, C. (2004). „Präferenz für Intuition und Deliberation (PID): Inventar zur Erfassung von affekt- und kognitionsbasiertem Entscheiden“. Zeitschrift für differentielle und diagnostische psychologie, Volume 25, No. 4, 179–197.
3. Brož, M., Bezvoda, V. 2008. Microsoft Excel Vzorce, funkce, výpočty. Computer Press, Brno. ISBN 80-251-1088-5
4. Budíková, M., Králová, M., Maroš, B. Průvodce základními statistickými metodami. Grada publishing, a.s., Praha, ISBN 978-80-247-3243-5
5. Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.
6. Field, A., & Hole, G. J. (2002). *How to design and report experiments*. Sage.
7. Halama, P. (2005). *Princípy psychologické diagnostiky*. Typy Universitatis Tyrnaviensis.
8. Hanák (2014) Are deliberative people more consistent in decision making? In *Advances in cross-cultural decision making / recenzenti: Abbe, A., Appleget, J.* - [s.l.] : AHFE Conference, 2014. - ISBN 978-1-4951-2095-4. - pp. 43-52, [1,1 AH].
9. Hanák, R. (2014) Rozhodovanie expertov v personálnom manažmente. Bratislava: Vydavateľstvo EKONÓM, ISBN 978-80-225-3920-3
10. Hanák, R., Sirota, M., & Juanchich, M. (2013). Experts use compensatory strategies more often than novices in hiring decisions. *Studia Psychologica*, 55(4), 251.
11. Hendl, J. (2009) Přehled statistických metod. Praha, Portál. ISBN 978-80-7367-482-3
12. Hendl, J. (2014) *Statistika v aplikacích*. Portál, Praha. ISBN 978-80-262-0700-9
13. Hendl, J. (2016) Kvalitativní výzkum: základní metody a aplikace. Portál. Praha, ISBN 978-80-262-0982-9
14. Heretik, A. jr, Ritomský, A., Novotný V., Heretik, A., Pečeňák, J. (2005). *Validita a reliabilita merania úzkostlivosti v epidemiologickom výskume*. In. *Metódy Emprickej psychológie I*. Univerzita Konštantína Filozofa
15. Hindls, R., Hronová, S., Seger, J., & Fischer, J. (2007). *Statistika pro ekonomy*. Professional publishing.
16. Chajdiak, J. (2009). Štatistika v exceli 2007. *Bratislava, Statis*, 151-156.
17. Chajdiak, J. (2010). *Štatistika jednoducho*. Statis.
18. Chajdiak, J. (2013). *Analýza dotazníkových údajov*, Statis.
19. Juszczak, S. (2003). Metodológia empirických výskumov v spoločenských vedách. Iris.

20. Mareš, P., Rabušic, L., & Soukup, P. (2015). *Analýza sociálněvědních dat (nejen) v SPSS*. Masarykova univerzita.
21. Martin, P., & Bateson, P. P. G. (2009). *Úvod do teorie a metodologie měření chování*. Portál.
22. Meloun, M., & Militký, J. (2012). *Interaktivní statistická analýza dat*. Karolinum. ISBN 978-80-246-2173-9
23. Meloun, M., & Militký, J. (2012). *Kompendium statistického zpracování dat*. Karolinum. ISBN 978-80-246-2196-8
24. Mikušková, E. B., Hanák, R., & Čavojevová, V. (2015). Appropriateness of Two Inventories Measuring Intuition (the Pid and the Rei) for Slovak Population. *Studia Psychologica*, 57(1), 63.
25. Neubauer, J., Sedlačík, M., & Kříž, O. (2012). *Základy statistiky: Aplikace v technických a ekonomických oborech*. Grada.
26. Ondrejkovič (2007) *Úvod do metodológie spoločenskovedného výskumu*. Veda
27. Pacáková, V. a kol. (2009): *Štatistické metódy pre ekonómov*. Bratislava: Iura Edition. ISBN 987-80-8078-284-9
28. Pacáková, V. a kol. (2015): *Štatistická indukcia pre ekonómov a manažérov*. Wolters Kluwer. ISBN: 9788081680816
29. Pelikán, J. *Základy empirického výskumu pedagogických jevů*. Praha: Karolinum, 1998. 270 s. ISBN 80-7184-569-8.
30. Ptáček, R., Raboch J. (2010) Určení rozsahu souboru a power analýza v psychiatrickém výzkumu. *Česká a Slovenská Psychiatrie* 106 (1) 33- 41
31. Punch, K. (2008). *Úspěšný návrh výzkumu*. Praha: Portál. ISBN 978-80-7367-468-7.
32. Rimančík, M. (2007) *Štatistika pre prax*. ISBN 978-80-969813-1-1
33. Ritomský, A. (1994) *SPSS pre Windows*. C.T.S Bratislava, ISBN 80-967050-1-6
34. Ritomský, A. (1999). *Deskripcia dát pomocou SPSS: sondy do súčasnej rodiny a domácnosti*. Medzinárodné stredisko pre štúdium rodiny. ISBN 80-967419-4-2
35. Řehák, J., Brom, O. (2015) *SPSS – Praktická analýza dat*. Computer Press, Brno, ISBN 978-80-251-4609-5
36. Sawilowsky, S (2009). "New effect size rules of thumb.". *Journal of Modern Applied Statistical Methods*. 8 (2): 467–474
37. Soukup, P. (2013). Věcná významnost výsledků a její možnosti měření. *Data a výzkum - SDA Info*, Vol. 7, No. 2: 125-148.
38. Šoltés, Erik, (2015). *Štatistické metódy pre ekonómov*. Bratislava: Wolters Kluwer. ISBN: 9788081682346
39. Terek, M. (2013). *Interpretácia štatistiky a dát*. Equilibria. ISBN 978-80-8143-100-5
40. Terek, M., Horníková, A., Labudová, V. (2010): *Hĺbková analýza údajov*, Wolters Kluwer (Iura Edition), ISBN: 9788080783365

41. Walker, I. A. N. *Výzkumné metody a statistika. Praha: Grada, 2013* (pp. 37-40). ISBN 978-80-247-3920-5.

13 REGISTER

Analýza rozptylu	125
Analýza rozptylu -variability	132
Aritmetický priemer	76
Binárna logistická regresia	108
Cohenova Kappa	92
Cohenovo d	126, 141
Cox & Snell R Square	110
Cramérovo V	87
Cronbachova alfa	82
Dátové Zobrazenie	11
diskrétna premenná	67
Dvojvýberový t - Test	125
Equimax rotácia	114
Eta koeficient	88
Exploračná faktorová analýza	113
Exponenciálne B	111
Extrakcia faktorov	114
Faktor	113
Faktorová analýza	112
Faktorová matica	113
Faktorové zaťaženie	113
Goodmanova – Kruskalova gamma	92
Goodmanova – Kruskalova lambda	88
Chyba II druhu	73
Chyba prvého druhu	72
chyba systematická	63
Intervalová premenná	66
Jednostranná hypotéza	71
Jednovýberový t – test	125, 138
Kenadallovo tau b	92
Koeficient determinácie	106, 127
Koeficient Eta, (Eta coefficient)	88
Koeficient kontingencie	87
Koeficient Phi	87
Konštruktívna validita	64
Kontingenčná tabuľka	87
Kvalitatívna premenná	65
Kvantitatívna premenná	65
Latentná premenná	68
Lineárna regresná analýza	103

Manifestná premenná	68
Maximálna hodnota	78
Medián	77
Medián split	81
Minimálna hodnota	78
Modus	77
Nagelkerke R Square	110
Náhodný výber	62
Nastavená hodnota spoľahlivosti R	106
Nezávislá premenná	68, 103
Nezávislosť merania	125
Nominálna premenná	65
Normalita rozloženia	124
Objektivita merania	63
Ordinálna premenná	66
Ortogonalne rozdelenie	114
Párový t test	136
Párový T- test na strednú hodnotu	125
Pearsonov korelačný koeficient	97
Pomer šancí	111
Pomerová premenná	67
Porovnávanie stredových hodnôt	123
Prediktívna validita	64
Predpoklad homogenity rozptylu	124
Predpoklad sféricity	124
Premenná	65
Quartimax rotácia	114
Reliabilita merania	65
Rotácia faktorov	114
Rozdeľovanie testu	82
Rozptyl	78
rozsah výberu	62
Senzitivita merania	65
Sila testu	73
Smerodajná odchýlka	79
Somersove d	92
Spearmanovo rho	92
Spojité premenná	67
Stiahnutie programu	7
Šutinový graf	115
Šikmé rozdelenie	114
Šikmosť	78
špecifickosť merania	65

Špicatost'	78
Testovanie hypotéz	70
Validita	64
Validita merania	64
Varimax rotácia	114
Veľkosť účinku	5, 126, 127, 131, 135, 138
Vlastná hodnota	113
Výber	62
výberová náhodná chyba	63
výberový súbor	62
Wald	110
Wilcoxonov test	142
Základná populácia	62
Základný súbor	62
Zámerný výber	63
Závislá premenná	68, 103
Zhoda rozptylov (homoskedasticita)	124
Zobrazenie Premenných	13