

KAPITOLA 1

PRÁCE S HROMADNÝMI DATY PŘED JEJICH ANALÝZOU

1. 0 STRUČNÉ SEZNÁMENÍ SE S SPSS

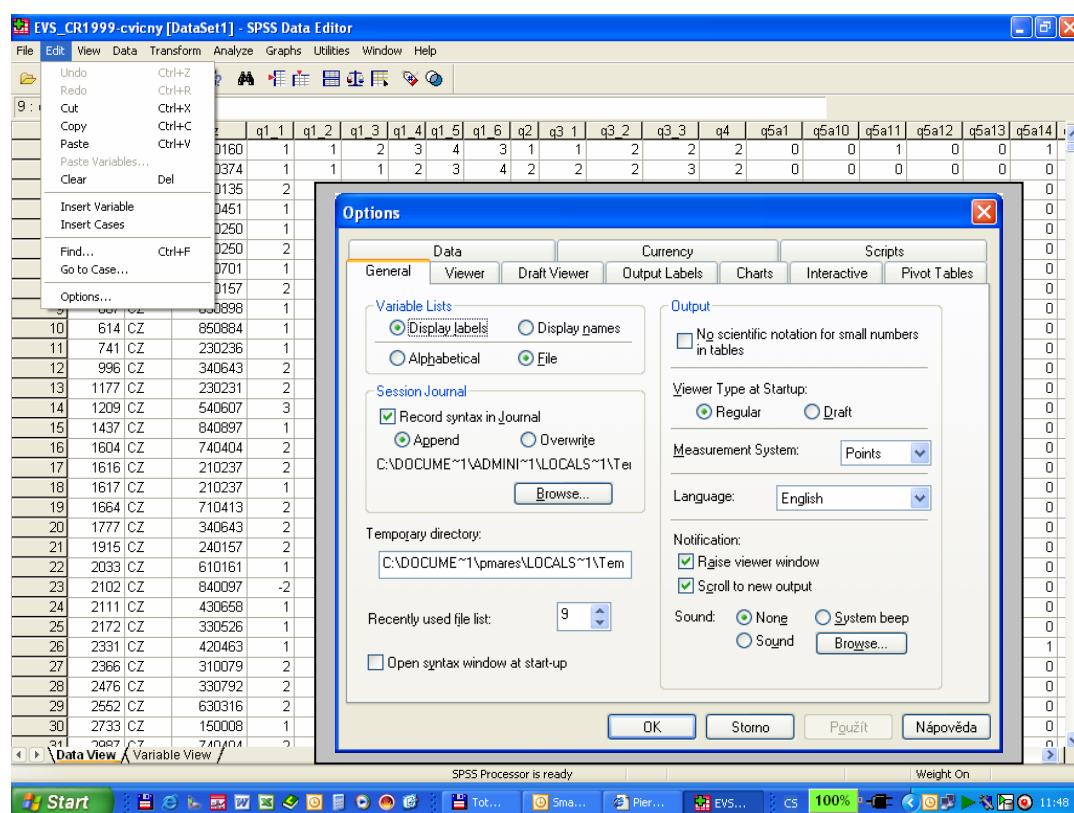
1. 0. 1 ZÁKLADNÍ INFORMACE

SPSS je jeden z řady programů pro statistické zpracování hromadných dat. S charakteristikami programu se nejlépe seznámíte na příslušných websites – najdete si je¹, kde si lze stáhnout i orientační popisy jednotlivých modulů a některé základní popisy procedur obsažených v programu.

1.0. 2 NASTAVENÍ VNITŘNÍHO PROSTŘEDÍ

Před zahájením analýz v SPSS je vhodné si nejdříve nastavit prostředí programu. Pod tlačítkem *EDIT* se skrývá volba *Options* (viz obr. 1), kde lze navolit zejména podobu výstupů (výsledků výpočtů). Například grafickou podobu tabulek, popisky proměnných, lze také naučit program správné češtině apod.

Obr. 1.1: Okno pro nastavení vnitřního prostředí programu



¹ Odkaz najdete i v menu HELP – SPSS Home Page. To je ovšem centrální anglická website. Existuje i česká verze respektive stránky českého zastoupení firmy.

1. 0. 3 OKNA, S NIMIŽ PRACUJEME (Data View, Variable View, Output)

Každá analýza dat začíná nahráváním dat. Abychom data mohli nahrávat, musíme (1) nejdříve definovat jednotlivé proměnné, jimž pak (2) přiřazujeme výzkumem zjištěné konkrétní hodnoty. K operacím (1) a (2) slouží okna *Data View* a *Variable View*. Obě okna jsou interaktivní, takže do nich můžeme psát.

Data View (viz obr. 1.2) obsahuje matici dat, ve které řádky znamenají případy (cases) výzkumné jednotky – většinou jde o osoby respondenty/respondentky, výzkumnými jednotkami však mohou být i skupiny, územní celky, předměty jako texty apod. Sloupce matice jsou proměnné, neboli charakteristiky těchto zkoumaných jednotek, jejich vlastnosti. Každá jednotka tedy představuje vektor a číslce v něm představují kódy hodnot proměnných (u nominálních a ordinálních proměnných) nebo čísla (u spojitých, to je kardinálních proměnných) popisujících vlastnosti/charakteristiky jednotky.²

Obr. 1.2: Datová matice (*Data View*)

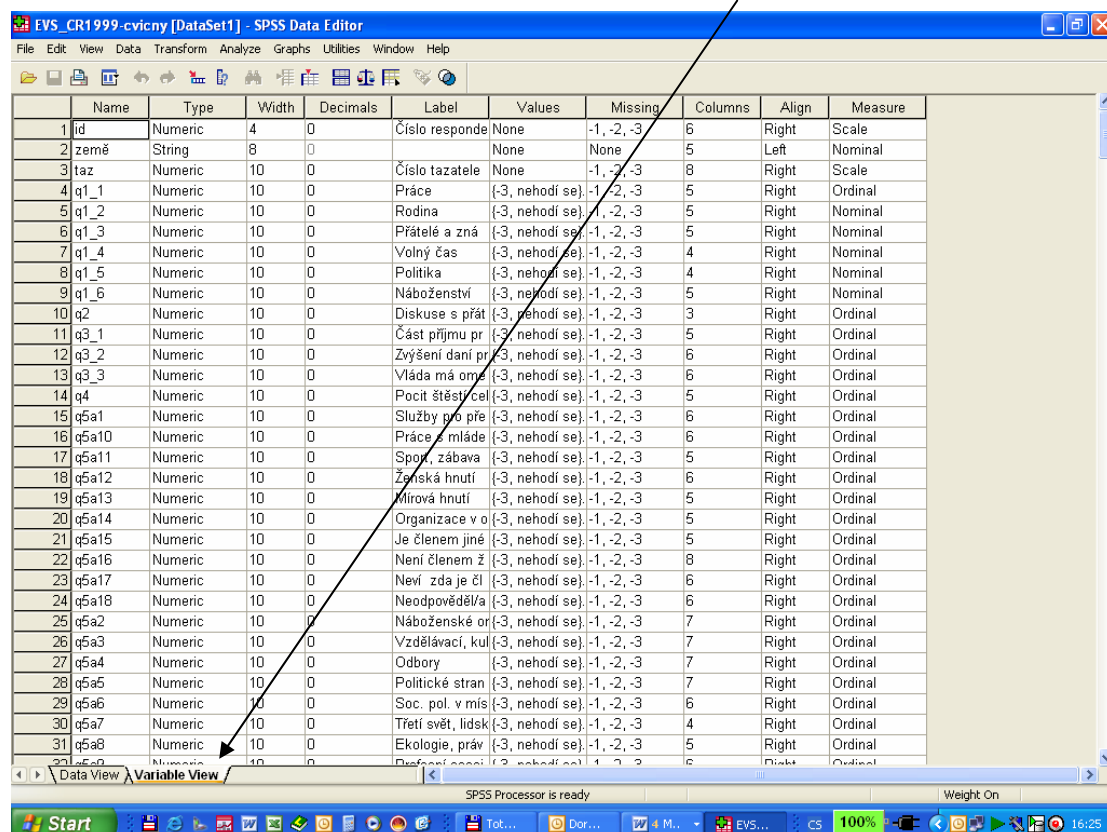
	id	země	taz	q1_1	q1_2	q1_3	q1_4	q1_5	q1_6	q2	q3_1	q3_2	q3_3	q4	q5a1	q5a10	q5a11	q5a12	q5a13	q5a14
1	109	CZ	510160	1	1	2	3	4	3	1	1	2	2	2	0	0	1	0	0	1
2	126	CZ	850374	1	1	1	2	3	4	2	2	2	3	2	0	0	0	0	0	0
3	148	CZ	610135	2	2	2	2	3	4	2	2	2	2	2	0	0	0	0	0	0
4	193	CZ	240451	1	1	1	2	3	4	2	1	1	3	2	0	0	1	0	0	0
5	231	CZ	220250	1	1	2	2	3	4	2	-1	2	3	2	0	0	0	0	0	0
6	238	CZ	220250	2	1	3	3	1	2	2	3	2	1	2	0	0	0	0	0	0
7	538	CZ	850701	1	1	2	1	2	4	2	3	3	1	2	0	0	0	0	0	0
8	589	CZ	240157	1	2	2	1	2	3	2	1	2	3	2	1	0	0	1	0	0
9	607	CZ	850898	1	1	2	2	2	2	2	-2	-2	-2	2	0	0	0	0	0	0
10	614	CZ	850884	1	1	3	1	3	4	2	4	4	2	2	0	0	0	0	0	0
11	741	CZ	230236	1	1	1	1	3	4	3	3	2	3	2	0	0	1	0	0	0
12	996	CZ	340643	2	1	2	2	3	4	2	-1	2	3	2	0	0	0	0	0	0
13	1177	CZ	230237	2	1	2	2	3	3	1	1	2	3	1	0	0	0	0	0	0
14	1209	CZ	540677	3	1	3	2	3	3	1	1	3	1	2	0	0	0	0	0	0
15	1437	CZ	840697	1	1	1	3	3	3	2	2	2	3	2	0	0	1	0	0	0
16	1604	CZ	740404	2	1	2	2	2	3	2	2	2	3	2	0	0	1	0	0	0
17	1616	CZ	210237	2	1	1	3	3	-1	2	3	2	3	2	0	0	0	0	0	0
18	1617	CZ	210237	1	2	2	2	3	4	2	-1	-1	4	2	0	0	0	0	0	0
19	1664	CZ	710413	2	2	2	1	2	4	2	2	2	3	2	0	0	0	0	0	0
20	1777	CZ	340643	2	1	3	-1	3	-1	2	-1	2	-1	2	0	0	0	0	0	0
21	1915	CZ	240157	2	2	2	2	3	3	2	3	2	3	2	0	0	0	0	0	0
22	2033	CZ	610161	1	1	2	3	3	4	2	3	3	2	2	0	0	0	0	0	0
23	2102	CZ	840097	-2	1	1	2	3	4	2	2	2	3	2	0	0	0	0	0	0
24	2111	CZ	430658	1	1	1	1	1	4	2	1	1	2	2	0	0	0	0	0	0
25	2172	CZ	330526	1	1	2	2	3	4	2	2	2	3	2	0	0	0	0	0	0
26	2331	CZ	420463	1	1	2	2	3	4	1	1	2	3	1	1	0	0	0	0	1
27	2365	CZ	310079	2	1	2	2	3	2	3	2	2	2	2	1	0	1	0	0	0
28	2476	CZ	330792	2	1	2	3	3	3	2	1	2	3	2	0	0	0	0	0	0
29	2552	CZ	630316	2	1	2	2	2	3	2	2	3	1	2	0	0	1	0	0	0
30	2733	CZ	150008	1	1	2	2	2	3	1	2	1	4	2	0	0	0	0	0	0

Okno *Variable View* (viz obr 1.3) představuje popis proměnných. Je to v SPSS zabudovaný speciální tabulkový procesor, který tento popis umožňuje. Bez popisu proměnných bychom konkrétní hodnoty proměnných nemohli nahrávat, proto popis proměnných musí vždy předcházet nahrávání dat. Při popisu proměnných vlastně převádíme náš dotazník, jeho jednotlivé otázky či položky, do formalizované podoby, kterou vyžaduje SPSS.

Jednotlivé proměnné, což jsou zkratkovitě vyjádřené jednotlivé otázky z dotazníku, jsou zde – na rozdíl od datové matice – umístěny v řádcích. Sloupce tohoto procesoru pak udávají jejich základní charakteristiky: jméno proměnné, její popis (název), popisky jednotlivých variant proměnné, chybějící hodnoty atd.

² Pozici řádků a sloupců lze měnit pomocí menu *DATA – TRANSPOSE*. Děláme to někdy proto, když je např. výstupní tabulka příliš široká a nevešla by se na šířku tisku.

Obr. 1.3: Tabulkový procesor pro popis proměnných (Variable View)



Třetím základním oknem je okno výstupu, *Output* (viz obr. 1.4a a 1.4b), které se automaticky otevře v okamžiku, kdy zadáme nějaký výpočet. Objevují se v něm výsledky požadovaných výpočtů (tabulky, grafy atd.). Ty zde můžeme editovat.³ Klikneme-li dvakrát na výstup, který chceme editovat, objeví se poněkud jiná nabídka a my můžeme měnit jeho grafickou podobu, měnit texty popisků apod. Postup je naznačen níže v následujících obrázcích. Editovat můžeme především prostřednictvím menu *EDIT*, *FORMAT* nebo také *PIVOT* (již méně často), kde se nabízí zejména již zmíněná a užitečná operace záměny sloupců a řádků.

Etická vsuvka: hovoříme-li o editování, máme na mysli pochopitelně pouze editování grafické podoby výstupů. V žádném případě není možné v tabulkách editovat, to je měnit, jejich číselné hodnoty! Obsah výstupů neboli výpočtů z analýz je ve vědě nedotknutelný! Přepis hodnot ve vypočtených tabulkách nebo údajích je ve vědě tím nejhorším zločinem, ještě horším, než plagiarismus. Je to hanebný čin, který má pro jeho aktéra závažné důsledky.

Obr. 1.4a: Ukázka výstup výpočtu distribuce četnosti (příkaz *Frequencies*)

³ Pozor, prosím. Pokud máte potřebu editovat své výsledky, veškerou editaci výstupů provádějte zde. Po přenesení výsledků do textové procesoru Word (viz dále) je to již prakticky vyloučeno, a to i v případě, kdy – navzdory varování, která najdete v textu dále, – použijete pro přenesení výstupu do Wordu příkazu export. Takto přenesené tabulky/grafy se sice editovat dají, ale při návratu z editačního režimu se tabulka obvykle rozpadne a je nepřehlédná až nečitelná.

Output1 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Output
Frequencies
Title
Notes
Active Dataset
Statistics
Rodinný stav

Frequencies

[DataSet1] C:\Documents and Settings\pmareš\Documents\Moje soubory\04PREDMETY_A_PREDNASKY\02UNDERG

Statistics

Rodinný stav

N	Valid	1893
	Missing	15

Rodinný stav

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ženatý/vdaná	1201	63,0	63,4	63,4
	vdovec/ vdova	190	10,0	10,0	73,5
	rozvedený/á	141	7,4	7,4	80,9
	odloučení	7	,4	,4	81,3
	nikdy neoženěn/ neprovdána	354	18,6	18,7	100,0
	Total	1893	99,2	100,0	
Missing	nehodí se	5	,3		
	neodpověděl/a	10	,5		
	Total	15	,8		
Total		1908	100,0		

Select Last Output

SPSS Processor is ready

Obr. 1.4b: Editace výstupu

Output1 - SPSS Viewer

File Edit View Insert Pivot Format Analyze Graphs Utilities Window Help

Output
Frequencies
Title
Notes
Active Dataset
Statistics
Rodinný stav
Pie Chart

Cell Properties...
Table Properties...
TableLooks...
Font...
Footnote Marker...
Set Data Cell Widths...
Autofit
Renumber Footnotes
Rotate Inner Column Labels
Rotate Outer Row Labels
Keep Together
Break Here
Remove Keep Together
Remove Break Here

Documents and Settings\pmareš\Documents\Moje soubory\04PREDMETY_A_PREDNASKY\02UNDE

1893
15

Rodinný stav

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ženatý/vdaná	1201	63,0	63,4	63,4
	vdovec/ vdova	190	10,0	10,0	73,5
	rozvedený/á	141	7,4	7,4	80,9
	odloučení	7	,4	,4	81,3
	nikdy neoženěn/ neprovdána	354	18,6	18,7	100,0
	Total	1893	99,2	100,0	
Missing	nehodí se	5	,3		
	neodpověděl/a	10	,5		
	Total	15	,8		
Total		1908	100,0		

Tabulka po dvojím kliknutí v editačním režimu
Všimněte si změněného menu.

SPSS Processor is ready

H: 208 , W: 386 pt

Výsledky, které se objeví v okně *Output* lze uchovat uložením příkazem *SAVE AS*. Uloží se v novém souboru s příponou *.spo*. Jednotlivé výstupy i celek lze také exportovat do Wordu, to však nedoporučujeme, neboť tabulky se často rozpadnou.. Lepší je v menu *EDIT* tabulku zablokovat a pomocí *COPY OBJECT* (nebo také příkazem Ctrl. K) ji vkopírovat (vložit, Ctrl. V) do vašeho textu, který píšete v textovém editoru.

Syntax a Script

SPSS byl původně vyvinut v operačním systému DOS, takže místo klikáním na tlačítka v menu bylo nutno zadávat příkazy ve formě vět, jejichž syntax byl předepsán. I ve verzi pracující pod Windows lze příkaz zadat nejen pomocí menu, ale i pomocí psaného příkazu. K tomu slouží okno Syntax (menu FILES). Například výpočet rozložení hodnot proměnné q89 a grafu tohoto rozložení (viz dále) lze zadat příkazem:

```
FREQUENCIES VARIABLES=q89 /PIECHART FREQ /ORDER = ANALYSIS.
```

(Pro samotný výpočet by ovšem stačilo: FREQUENCIES VARIABLES=q89.

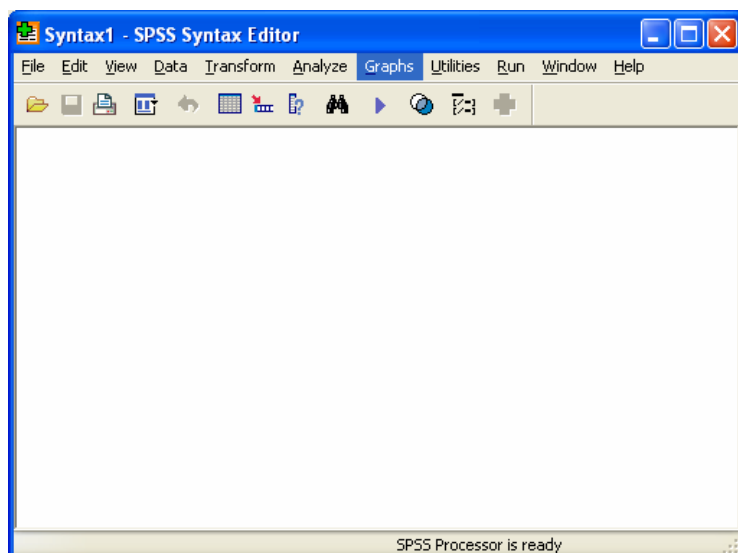
Tečka na konci příkazu je bytostně důležitá. Pokud ji zapomenete, program neví, kde jeden příkaz končí a druhý začíná, takže výpočet odmítne.)

Pokud si v EDIT – OPTION – DRAFT VIEWER (tedy v příkazech, jimiž nastavujeme vnitřní prostředí SPSS) zatržením kolonky u DISPLAY COMMAND IN LOG tuto funkci nastavíte, máte možnost si po každém výpočtu zadaném v menu na prvních řádcích výsledků ve výstupu Output přečíst i text příkazu výpočet zadávající.

Nová okna *Syntax* a *Script* lze otevřít v menu FILE – NEW a do otevřených oken lze psát konvenčním jazykem SPSS příkazy. Obsah okna lze uložit jako soubor syntaxí s příponou *.sps* a SCRIPT s příponou *.sbs*. Soubory se syntaxem obsahují příkazy, které umožňují zadávat a spouštět statistické procedury (které jsou jinak v menu ANALYZE) a příkazy k transformaci dat (které jsou jinak v menu TRANSFORM). Skriptové soubory dovolují manipulovat s výstupy (oba typy souboru lze pro práci s daty kombinovat). Neděste se, tato nyní naprosto nepochopitelná hatmatilka se po vám po několika sezeních nad SPSS a práci s ním natolik dostane do krve, že se stane běžnou součástí vašeho datově analytického žargonu.

Ukázka konverzace odposlechnutá během cvičení z analýzy dat:
 Ivo: „Lucko, kde najdu ten zatracený *CROSSTABS*, abych ho mohl spustit?“
 Lucka: „V *DESCRIPTIVES*.“
 Ivo: Aha.... Ale jak to, že se mně netisknou tabulky???
 Lucka: „Ukaž, no jo, tys od minulé výpočtu zapomněl zrušit *Supress tables*.“

Obr. 1.5: Okno pro psaní příkazů ve formě syntaxe



S psanými příkazy většinou nepracujeme, existují však užitečné výjimky. Zmíníme tři z nich:

- Syntaxe je výhodné použít při transformaci existujících proměnných v novou proměnnou za pomoci logických podmínek – viz příslušnou kapitolu o transformaci proměnných a proceduře IF.
- Je výhodné zapsat si syntakticky zadání rutinně opakovaného výpočtu s různými daty. Například tehdy, když úplně stejným způsobem zpracováváte začátkem každého měsíce data, která vám přicházejí ze statistického výkazu úřadu práce o počtech a struktuře nezaměstnaných.). Jednou napsaný příkaz (skript) vám slouží tak dlouho, jak zůstává výpočet neměnný. Pak stačí, abyste si otevřeli matici s novými daty a na něj pustili skript uložený na disku vašeho počítače prostřednictvím příkazu RUN.
- U složitějších výpočtů vícerozměrných analýz je potřeba, abyste si všechny příkazy k analýzám uchovávali ve svém výpočetním archívu. Nikdy totiž nevíte, kdy si budete muset ověřit, zdali jste postupovali správně – a bez archivace syntaxe výpočtu toho nebudete schopni.

1. 0. 4 HELP A TUTORIAL

SPSS má poměrně rozsáhlý a dobře zpracovaný HELP, který obsahuje i základní uvedení do programu (TUTORIAL). Rozhodně stojí za prohlédnutí.

1. 1 DATA

1. 1. 1 CO JE TO MATICE DAT

Při statistické analýze dat pracujeme s číslicemi, které mají určitý význam (kategorizovaná data) nebo s čísly. Abychom mohli tuto analýzu provádět, musíme tato data dostat do počítače a vytvořit v něm matici dat. Protože jde o zpracování hromadných dat, pracujeme s *hodnotami proměnných* neboli s kvantifikovanými charakteristikami/vlastnostmi *případů* – to je respondentů či jiných objektů, popřípadě jevů.. Matici tvoří tedy *případy* (obvykle řádky matice) versus *proměnné* (obvykle sloupce matice) a obsah matice tvoří *hodnoty příslušných proměnných* charakterizujících jednotlivé případy.

Případy jsou popsány svými vlastnostmi (atributy) – variantami neboli hodnotami proměnných, které jsou jejich logickými uskupeními. Například proměnná vzdělání může být uskupením

možných nejvyšších dosažených stupňů vzdělání: základní, středoškolské, vysokoškolské (které lze popřípadě dále členit: vysokoškolské nižšího typu (Bc.), vyššího typu (Mgr.) popř. s vědeckou hodností (PhDr, PhD. Apod.). Každému případu přidělujeme jeho identifikační číslo – id – a ideální je, máme-li stejným číslem označený i originál, z něhož data o případu (nejčastěji respondentovi/respondence – tedy dotazníku) čerpáme. Jen tak můžeme v případě nejasností porovnat zdroj dat s jejich záznamem v matici (proto dotazníky nikdy neničte, ale archivujte je) a provést dodatečnou kontrolu. A že se chyby při nahrávání dat vyskytují, je mnohokrát potvrzenou zkušeností.⁴

Co jsou *proměnné* již víme, stejně jako víme, že existují proměnné kategorizované (nominální a ordinální), a proměnné spojité (kardinální). U nominálních proměnných je spojení číslice (numerického kódu) a vlastnosti zcela arbitrární, takže bychom proměnnou „rodinný stav“ mohli kódovat např. 1 = svobodný/á, 5 = ženatý/vdaná, 6 = rozvedený/á a 9 = ovdovělý/á.), u ordinálních proměnných číslice označují pozici varianty na škále, aniž by cokoli říkali o vzdálenosti mezi těmito pozicemi (vzdělání: základní = 1, střední = 2, vysokoškolské = 3 apod.). Je dobré si toto zvolené přiřazení číslic k charakteristikám pamatovat (je sice již v dotazníku, ale do něho nemůžeme stále nahlížet, a proto při definici matice vedle definice proměnných musíme také určit, jak ji pojmenujeme, kolik bude mít desetinných míst v případě spojité proměnné a jejich popisu, jaký verbální význam mají jméno proměnné a číslice hodnot proměnné v případě kategorizovaných proměnných). U spojité proměnné jde o konkrétní číslo, které přímo vyjadřuje množství příslušné vlastnosti.

Každý případ tak představuje vektor obsahující *hodnoty příslušných proměnných* (každá varianta každé proměnné má přiřazenu číslici). Vektory plníme do matice: co řádek, to případ (např. respondent) a co sloupec, to proměnná. Vše ilustruje obr. 1.6

Obr. 1.6: Ukázka matice dat s pěti proměnnými a dvěma případy

	id	pohlaví	vek	vzdel	prav_lev	var	var	var	var	var
1	1080	2	34	2	1					
2	1081	1	45	4	5					
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

1 = muž
2 = žena

1 = základní, nevyučen/a
2 = základní, vyučen/a
3 = středoškolské
4 = vysokoškolské

1 = krajní levice
2 = levice
3 = střed
4 = pravice
5 = krajní pravice

PŘÍPAD (CASE/RESPONDENT) 1:
 pořadové číslo (ID) = 1080
 žena (pohlaví=2)
 věk: 34 let
 vzdělání: vyučená
 pozice na škále politická levice či pravice = 1 signalizuje levicovou orientaci

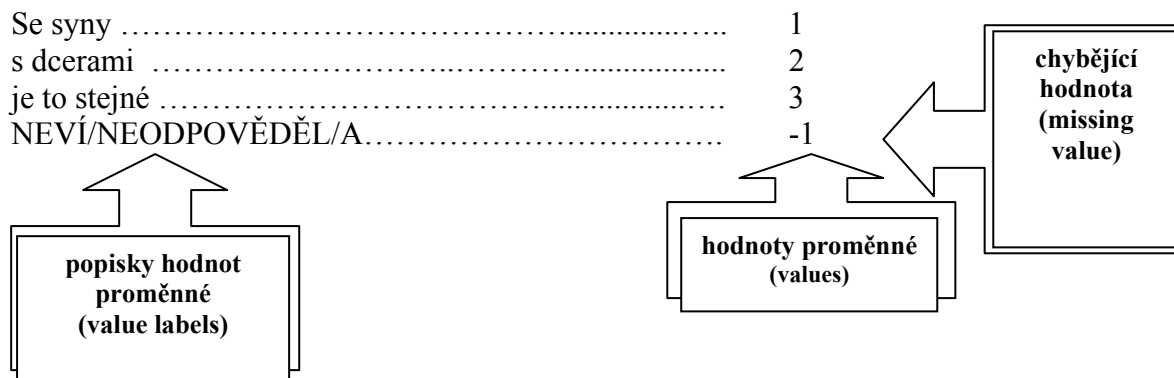
PŘÍPAD (CASE/RESPONDENT) 2:
 pořadové číslo (ID) = 1081
 muž (pohlaví=1)
 věk: 45 let
 vzdělání: vysokoškolské
 pozice na škále politická levice či pravice = 5 signalizuje pravicovou orientaci

hrávání dat do počítače se nevyhnou ani přírodní vědy.

Matice dat je tedy souborem kódů, za nimiž se skrývají konkrétní kvalitativní nebo kvantitativní vlastnosti jednotek našeho výzkumu. Tato data mohou být dále upravována (například pomocí transformací proměnných nebo výběrem případů) a – především – analyzována.

Demonstrace otázky v dotazníku jako (kategorizované) proměnné:

36. „S kým jsou podle Vás větší výdaje, se syny nebo dcerami anebo je to stejné?“



Každá proměnná musí mít své jméno. Jména můžeme dávat různá, ale SPSS má svá přísná pravidla, jak tato jména mohou vypadat. Každé jméno musí začínat písmenem, znaky se mohou používat alfanumerické, ale bez diakritiky: např. P1, A26, X26. Při vymýšlení jmen si můžeme mnemotechnicky usnadňovat práci s maticí při analýze. Například rezervujeme všechna jména začínající písmenem A pro důvody, které lidé uvádějí, proč chtějí mít dítě, písmeno B pro důvody, proč dítě mít nechtějí Písmeno D pak demografické charakteristiky respondentů (, , pohlaví, věk, rodinný stav apod. Nebo můžeme používat i nápovědných jmen proměnných: sex (pro pohlaví, ne pro frekvenci koitu), příjem, vzdělání, job (povolání) apod. Nejen jména (tam je počet znaků limitován, ale i popisy proměnných (*variable labels*, které používáme pro lepší orientaci v tom, co znamenají zkratkovitá jména proměnných) by měly být krátké. Na rozdíl od jmen proměnných u jejich popisků lze používat diakritických znamének (možnost psát česky si lze nastavit v menu EDIT – OPTIONS).

Demonstrace baterie otázek v dotazníku jako sady (kategorizovaných) proměnných:

1. Řekněte prosím o každé z následujících skutečností, jak je ve Vašem životě důležitá:

	Velmi důležitá	Dost důležitá	Ne příliš důležitá	Vůbec ne důležitá	Neví	Neodpověď(a)	
A Práce	1	2	3	4	-1	-2	Q1a
B Rodina	1	2	3	4	-1	-2	Q1b
C Přátelé a známí	1	2	3	4	-1	-2	Q1c
D Volný čas	1	2	3	4	-1	-2	Q1d
E Politika	1	2	3	4	-1	-2	Q1a
F Náboženství	1	2	3	4	-1	-2	Q1f

Zde je každý řádek proměnnou s variantami/oborem hodnot <1;4>; záporné hodnoty představují missing values (v tomto příkladu rozlišujeme jejich význam, protože pro analýzy může být důležité, zdali ti, kdo na tyto otázky neodpovídali, tak dělali z důvodů, že se nemohli rozhodnout (-1) nebo že odmítli odpovědět (-2). Můžeme zvolit jména

proměnných například: Q1a až Q1f nebo Q1_1 až Q1_6 apod. tak, aby napovídala, že všech 6 proměnných má něco společného.

1. 1. 2 DEFINICE JEDNOTLIVÝCH PROMĚNNÝCH

Abychom mohli matici naplnit daty, musíme ji, jak jsme naznačili výše, nejprve definovat. Děje se tak v okně VARIABLE VIEW. Jde o tyto úkony:

- Připsání jména proměnné, určení jejího místa v matici (sloupce/sloupců).
- Definice charakteru proměnné jako numerické či stringové (alfaznakové, kterou počítač chápe jako označení a neprovádí s ní početní operace).
- Připsání širšího/podrobnějšího označení proměnné (*variable labels*).
- Připsání verbálního označení jednotlivým hodnotám (kategorizované) proměnné (*value labels*).

Labels zpřehledňují tištěné výstupy, neboť přiřazují k jménům proměnných (jež mohou mít dle konvence pouze 8 znaků) i vysvětlující popis. Např. q1_2 může být jméno proměnné neboli *name variable*) a „důležitost rodiny v životě“ může být vysvětlující popis jména této proměnné neboli *label variable*). Její varianty „velmi důležitá“, „dosti důležitá“, „nepříliš důležitá“, „vůbec nedůležitá“ a „nevím“ jsou pak *value labels* dané proměnné.

Pozor: V případě spojitých proměnných nedávají *value labels* smysl a nepoužíváme jich!

- Určení počtu desetinných míst (v případě spojitých proměnných).
- Definování tzv. missing value.

Obvykle jsou missing value výsledkem toho, že pro daný případ nemáme v konkrétní proměnné žádný údaj – respondent/ka na otázku neodpověděl/a nebo na ni odmítl/a odpovědět, tazatel/ka jeho/její odpověď nezaznamenal/a, při kontrole dotazníků se nám jeho/její odpověď zdála vysoce nevěrohodná či rozporná s jinými odpověďmi: například uvedl/a, že nemá děti, ale uvádí věk svého prvního dítěte apod.

Někdy do missing value přerazujeme některé hodnoty proměnných při jejich transformaci. Týká se to například varianty *nevím*, která sice někdy může být součástí ordinální proměnné jakožto její středová hodnota (1 = s vládou jsem spokojen, 2 = nevím, 3 = s vládou jsem nespokojen), častější jsou ale případy, kdy ji používáme jen proto, abychom nenutili respondenta/respondentku zaujímat postoj, který nemá. V další analýze se pak často soustředíme jen na ty, kdo postoj zaujaly a v modulu TRANSFORM – RECODE přiřadíme odpovědím „nevím“ číslici označující chybějící hodnotu (*missing value*). Často je to záporná hodnota.⁵

⁵ K tomu, jak se v některých konkrétních statistických technikách zachází s missing value, se ještě vrátíme. Většinou se s případy obsahující missing value nepracuje.

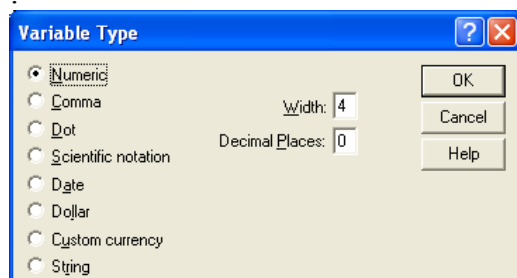
Definované proměnné (okno VARIABLE VIEW)

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	
1	id	Numeric	4	0	Číslo respondenta	None	-1, -2, -3	6	Rig
2	země	String	8	0		None	None	5	Left
3	taz	Numeric	10	0	Číslo tazatele	None	-1, -2, -3	8	Rig
4	q1_1	Numeric	10	0	Práce	{1, velmi důležité}	-1, -2, -3	5	Rig
5	q1_2	Numeric	10	0	Rodina	{1, velmi důležité}	-1, -2, -3	5	Rig
6	q1_3	Numeric	10	0	Prátelé a zná	{1, velmi důležité}	-1, -2, -3	5	Rig
7	q1_4	Numeric	10	0	Volný čas	{1, velmi důležité}	-1, -2, -3	5	Rig
8	q1_5	Numeric	10	0	Politika	{1, velmi důležité}	-1, -2, -3	5	Rig
9	q1_6	Numeric	10	0	Náboženství	{1, velmi důležité}	-1, -2, -3	5	Rig
10	q2	Numeric	10	0	Diskuse s přát	{1, často}...	-1, -2, -3	5	Rig
11	q3_1	Numeric	10	0	Část příjmu pr	{1, rozhodně s	-1, -2, -3	5	Rig
12	q3_2	Numeric	10	0	Zvýšení daní pr	{1, rozhodně s	-1, -2, -3	5	Rig
13	q3_3	Numeric	10	0	Vláda má ome	{1, rozhodně s	-1, -2, -3	5	Rig
14	q4	Numeric	10	0	Pocit štěstí cel	{1, velmi šťast	-1, -2, -3	5	Rig
15	q5a1	Numeric	10	0	Služby pro pře	{0, ne}...	-1, -2, -3	5	Rig
16	q5a10	Numeric	10	0	Práce s mláde	{0, ne}...	-1, -2, -3	5	Rig
17	q5a11	Numeric	10	0	Sport, zábava	{0, ne}...	-1, -2, -3	5	Rig
18	q5a12	Numeric	10	0	Ženská hnutí	{0, ne}...	-1, -2, -3	5	Rig
19	q5a13	Numeric	10	0	Mírová hnutí	{0, ne}...	-1, -2, -3	5	Rig
20	q5a14	Numeric	10	0	Organizace v o	{0, ne}...	-1, -2, -3	5	Rig
21	q5a15	Numeric	10	0	Je členem jiné	{0, ne}...	-1, -2, -3	5	Rig
22	q5a16	Numeric	10	0	Není členem ž	{0, ne}...	-1, -2, -3	5	Rig
23	q5a17	Numeric	10	0	Neví, zda je čl	{0, ne}...	-1, -2, -3	5	Rig

Obr. 1.7: Ukázka popisu proměnných

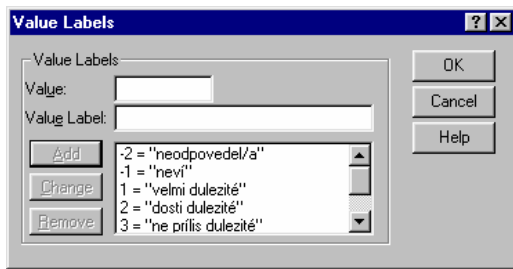
Některé sloupce v okně VARIABLE VIEW můžeme vyplnit přímo, jiné nám nabídnou po kliknutí předdefinované volby. Příkladem je definice typu proměnných (současně určujeme počet desetinných míst u spojitých znaků a šíři neboli počet míst, které proměnná vyžaduje – například v případě věku potřebujeme nejméně width 2, ale 3 je již zbytečné, pokud nechceme zpovídat stoleté a starší respondenty/respondentky.

:



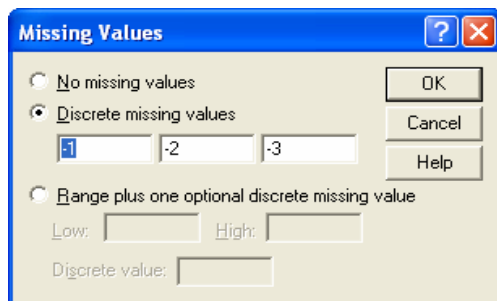
Klikneme-li na VALUES, můžeme opatřit proměnnou popisky jejich hodnot neboli *value labels* (variable label napíšeme přímo do sloupce LABELS, stejně jako jméno proměnné do sloupce NAME). Do kolonky value (viz obr. 1.8) vepíšeme číslíci hodnoty, tabulátorem či pomocí myši přejdeme do kolonky value label a popisek vepíšeme. Spodní tlačítka nám umožní takto definovaný label přidat a v seznamu labels pak provádět změny. Nakonec vše odsouhlasíme (OK).

Obr. 1.8.: Okno pro popis variant znaků (value labels)



Podobně můžeme definovat i chybějící proměnné (*missing values*), což jsou hodnoty, které nevházejí (pokud si to výslovně nepřejeme a nezadáme příkazem) do analýzy. SPSS nám k tomu nabízí speciální okno, jehož ukázkou uvádí obr. 1.9. Zde se badatel rozhodl, že pro chybějící hodnoty bude rezervovat výrazy -1, -2 a -3. Jistě podle ukázky přijdete sami na to, jaké další možnosti, jak definovat missing values, se nabízejí.

Obr. 1.9.: Okno pro definování chybějících hodnot



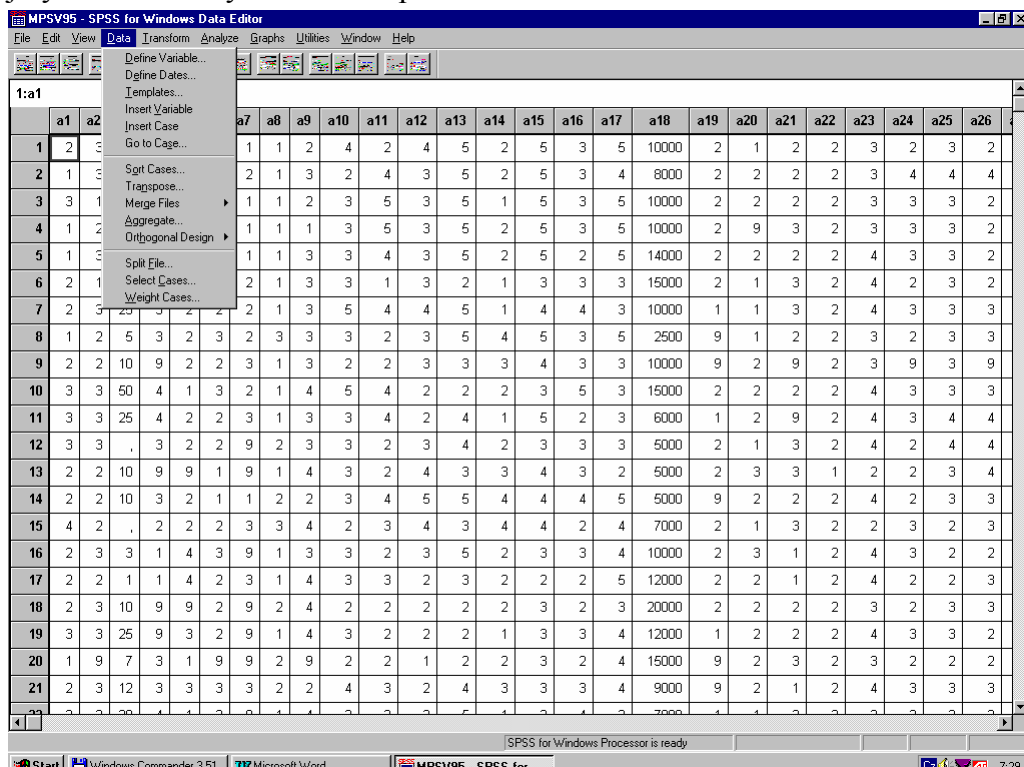
1. 1. 3 PLNĚNÍ MATICE DAT

Data můžete dostat do matice různými způsoby. Důležité jsou pro nás zejména:

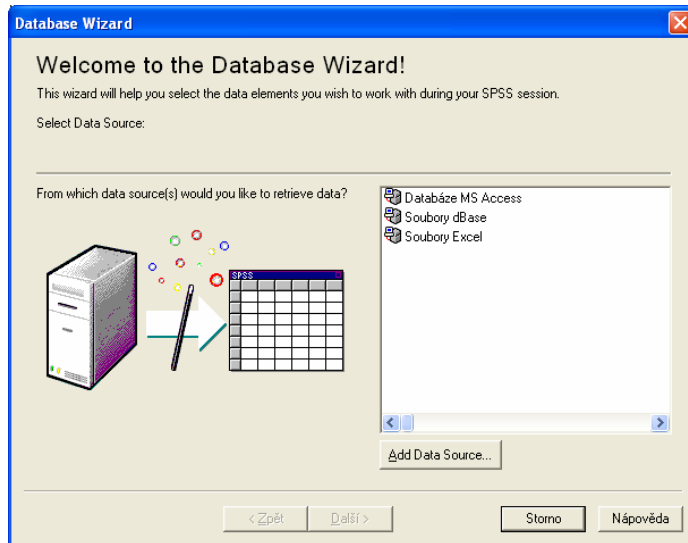
- Plnění námi definované matice daty.
- Import dat ze souboru jiného typu (z textového editoru, databáze či tabulkového procesoru (spreadsheetu), např. programu jako je Excel).

Můžeme ovšem také použít i dříve nebo někým jiným vytvořenou matici dat (tzv. systémový soubor).

Nejčastějším případem je plnění námi definované matice dat přepisem údajů z dotazníků nebo jiných záznamových archů do předem definované matice.



1. 1. 4 Import dat ze souboru jiného typu

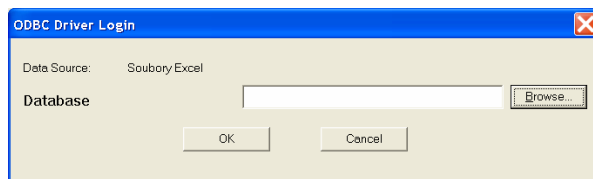


FILE → OPEN DATABASE → NEW QUERY

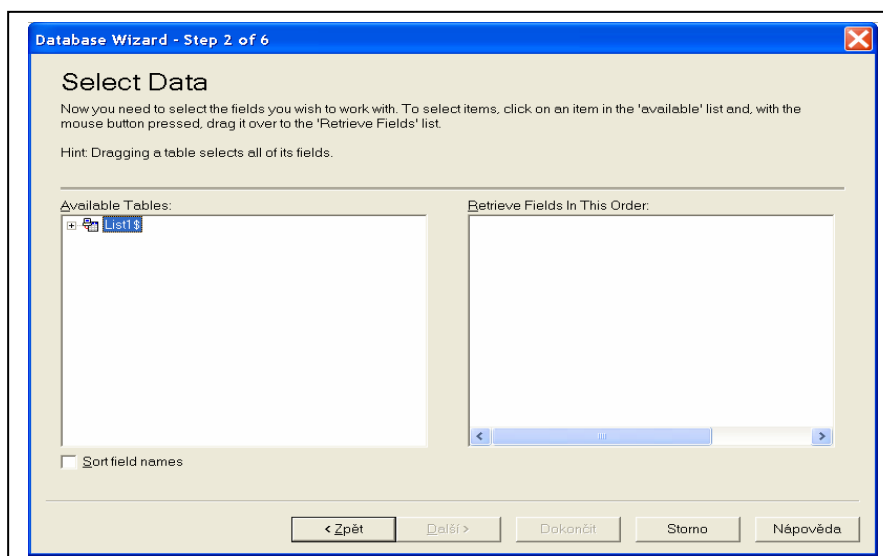
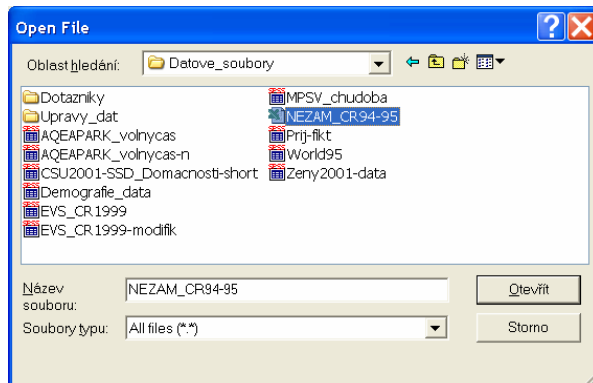
Zvolíme typ souboru (např. EXCEL files).

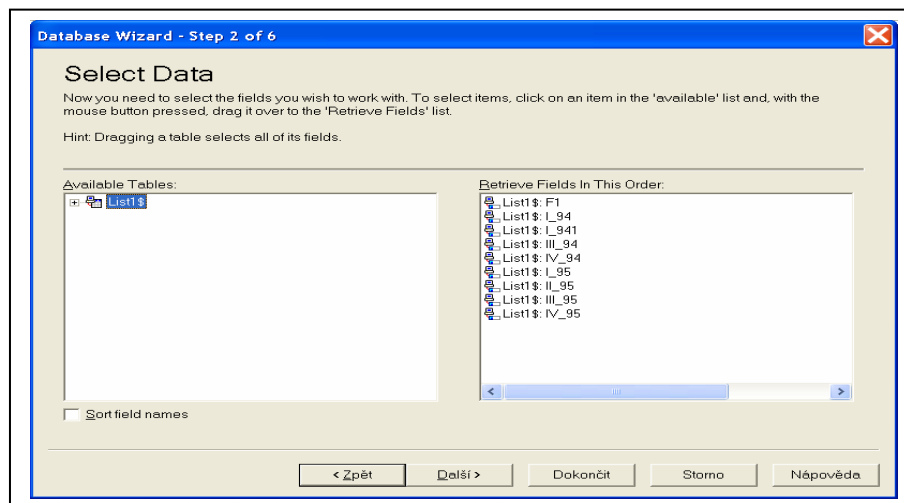
Najdeme příslušný soubor pomocí Browse: SOUBORY EXCEL

Dále najdeme na disku příslušný soubor.



Otevřeme ho a odsouhlasíme (OK v ODBC Driver Login). V tomto případě jde o soubor v němž jsou pro jednotlivá čtvrtletí let 1994 a 1995 zachyceny míry nezaměstnanosti v jednotlivých okresech České republiky. Pak přetáhneme pomocí myši List z levého do pravého okna.





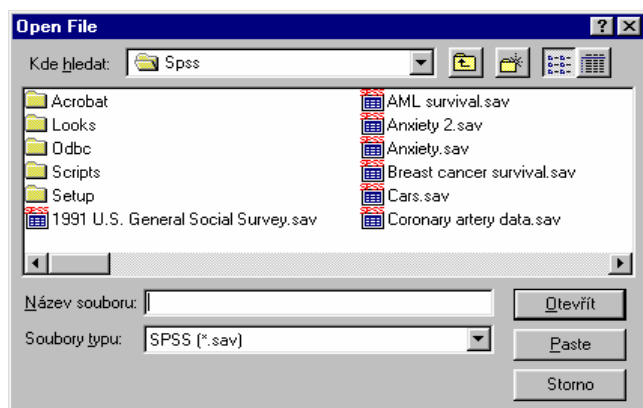
Pomocí *Další* mohu omezit přetahované případy, nebo mohu *Dokončit*. Obsah excelového souboru je přetážen do systémového souboru SPSS. Je to matice dat i se sloupcem představujícím jména bývalých krajů (proměnnou F1 mohu v okně *VARIABLES VIEW* přejmenovat) a jmény proměnných (jednotlivá čtvrtletí let 1994 a 1995). Data v matici představují příslušné míry nezaměstnanosti v daných krajích (kraje jsou případy) v těchto čtvrtletích (čtvrtletí jsou proměnnými a data v dané kolonce vždy hodnotou dané proměnné – svou povahou jsou to kardinální/spojité proměnné).

	F1	I_94	I_941	III_94	IV_94	I_95	II_95	III_95	IV_95	var	var	var	var
1	PRAHA	,30	,30	,30	,30	,30	,20	,30	,30				
2	STR_C	3,37	2,80	2,80	2,90	2,80	2,50	2,70	2,60				
3	JIH_C	2,70	2,00	2,10	2,30	2,20	1,80	2,00	2,00				
4	ZAP_C	2,68	2,20	2,20	2,20	2,20	2,00	2,00	2,20				
5	SEV_C	4,48	4,00	4,20	4,40	4,50	4,30	4,70	4,80				
6	VYCH_C	2,80	2,40	2,60	2,50	2,30	2,10	2,30	2,30				
7	JIH_M	3,63	3,20	3,20	3,30	3,10	2,80	3,00	2,90				
8	SEV_M	6,27	5,60	5,60	5,60	5,40	4,80	5,00	4,80				
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													
33													
34													

1. 2 PRÁCE SE SYSTÉMOVÝMI SOUBORY

1. 2. 1 OTEVŘENÍ EXISTUJÍCÍHO SOUBORU

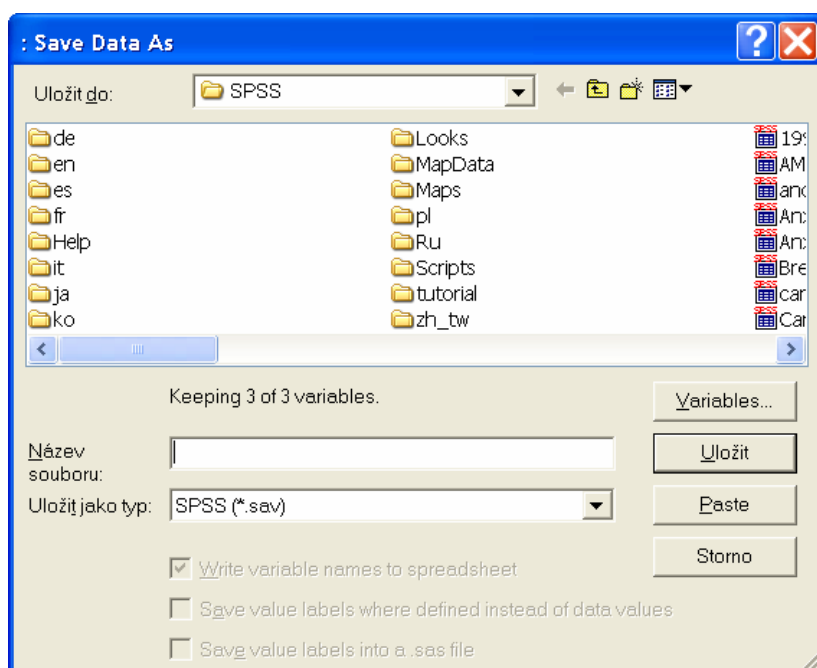
Již existující soubory otvíráme stejně jako jakkoliv jiné soubory v jiných programech. Tedy, po spuštění programu SPSS, klikáme postupně na tlačítka FILE → OPEN → DATA, v otevřeném okně potom najdeme to správné místo na disku, kde



máme soubor uložen.

1. 2. 2 UKLÁDÁNÍ VYTVOŘENÉHO SYSTÉMOVÉHO SOUBORU

Máme-li matici naplněnou našimi daty, snažíme se tuto matici zachovat pro další zpracování tím, že ji uložíme jako systémový soubor. SPSS takovým souborům při jejich uložení přidává příponu .sav – podle ní tyto soubory můžete identifikovat.



Pozor: Ukládejte soubor v každém kroku popisu a plnění matice stejně jako po každé změně, kterou v ní provedete (např. po přidání případu nebo vytvoření nových proměnných – viz lekci věnovanou transformaci proměnných). Ponechávejte přitom (samozřejmě pod různými názvy):

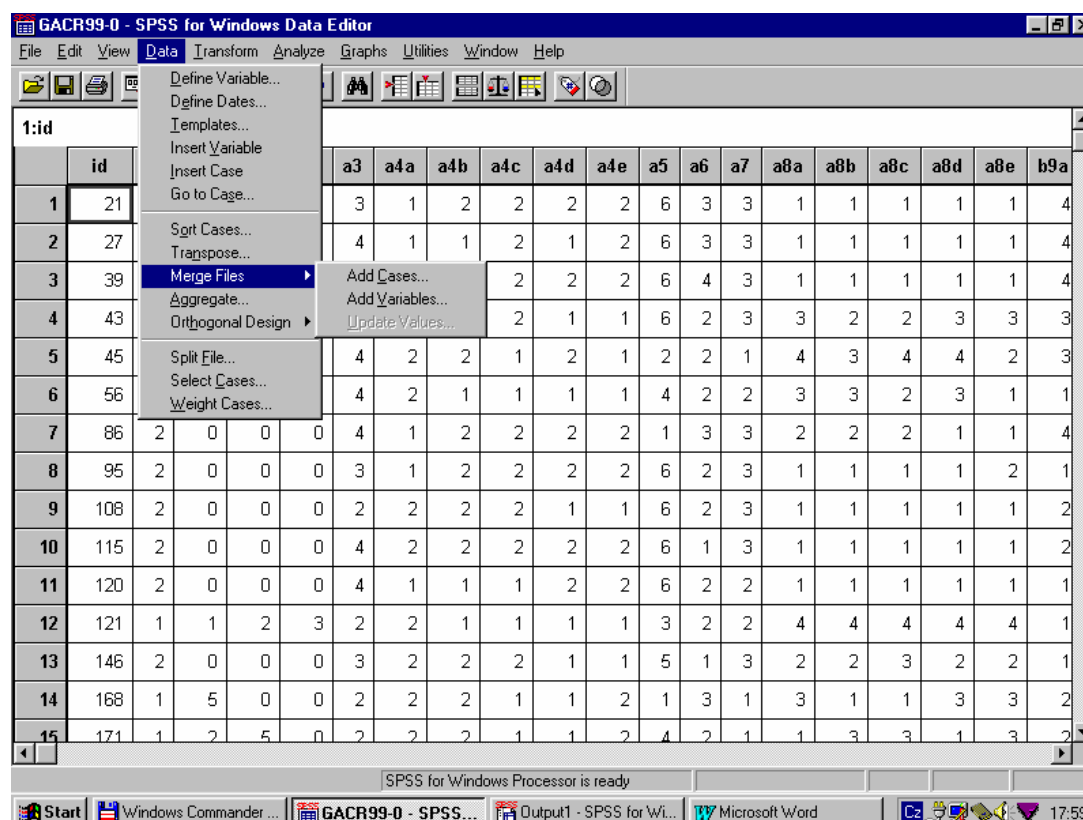
- Pramenný soubor (což je naplněná a zkontrolovaná původní matice, v níž nebyly provedeny žádné další změny).
- Předposlední podobu souboru (po předposledních provedených změnách).
- Poslední podobu souboru (po posledních provedených změnách).

Vyplatí se Vám to! Někdy se mohou totiž naplnit i katastrofické scénáře a při práci s poslední verzí souboru o něj můžete v důsledku technických potíží programu nebo počítače přijít (uchovávejte proto raději starší verze souboru i mimo harddisk svého počítače). Nebo se může stát, že omylem provedete v matici při transformaci proměnných nevratné změny, jak ukážeme v lekci o transformaci proměnných).

1. 3 PRÁCE SE SOUBORY

1. 3. 1 SLUČOVÁNÍ SOUBORŮ

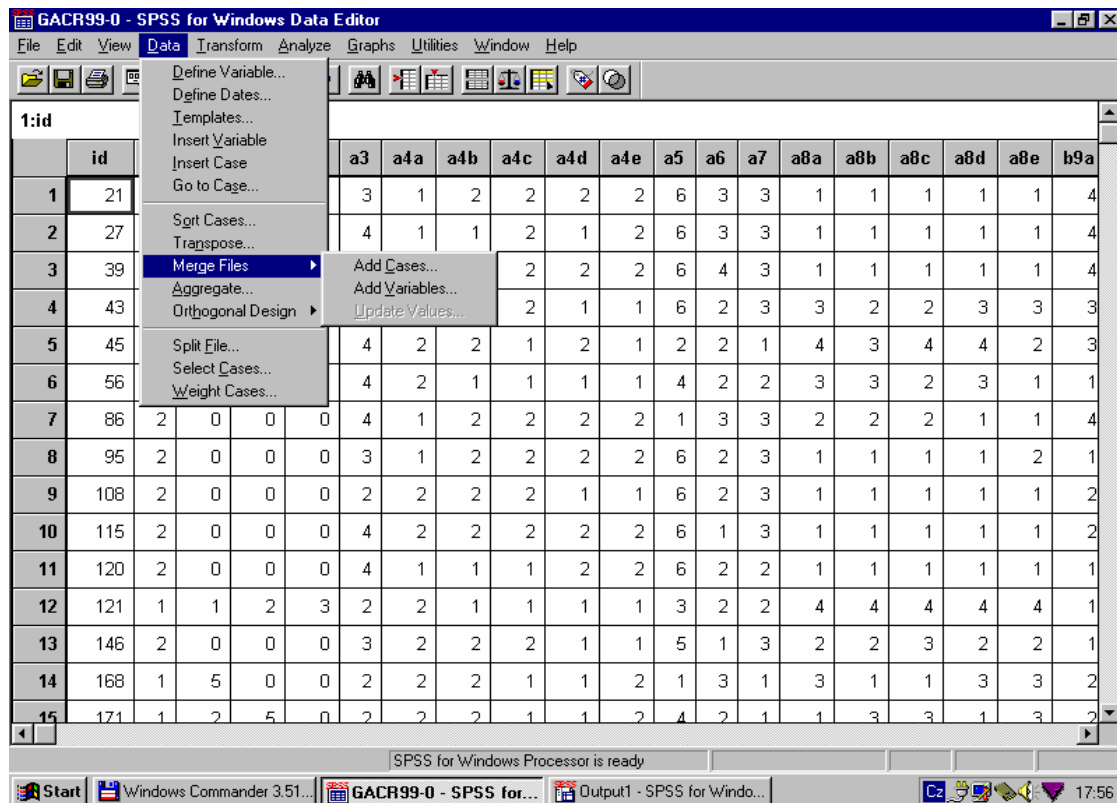
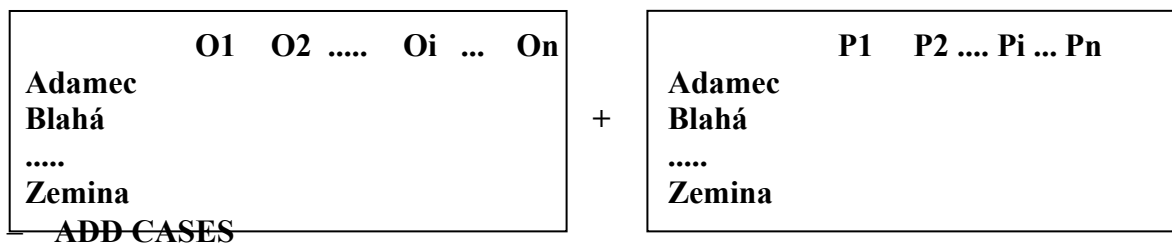
ADD VARIABLES



Máme v jedné databázi (matici) údaje o osobních charakteristikách studentů a v druhé databázi (matici) údaje o jejich prospěchu. Chceme je dostat do jedné matice všech údajů o studentech. Pořadí studentů musí být ve slučovaných maticích shodné, nebo musíme mít znak, který každého

studenta jednoznačně definuje. K proměnných jednoho souboru se přidají proměnné dalšího souboru, jak naznačuje schéma 1.

Schéma 1: Struktura dvou matic, které se budou slučovat



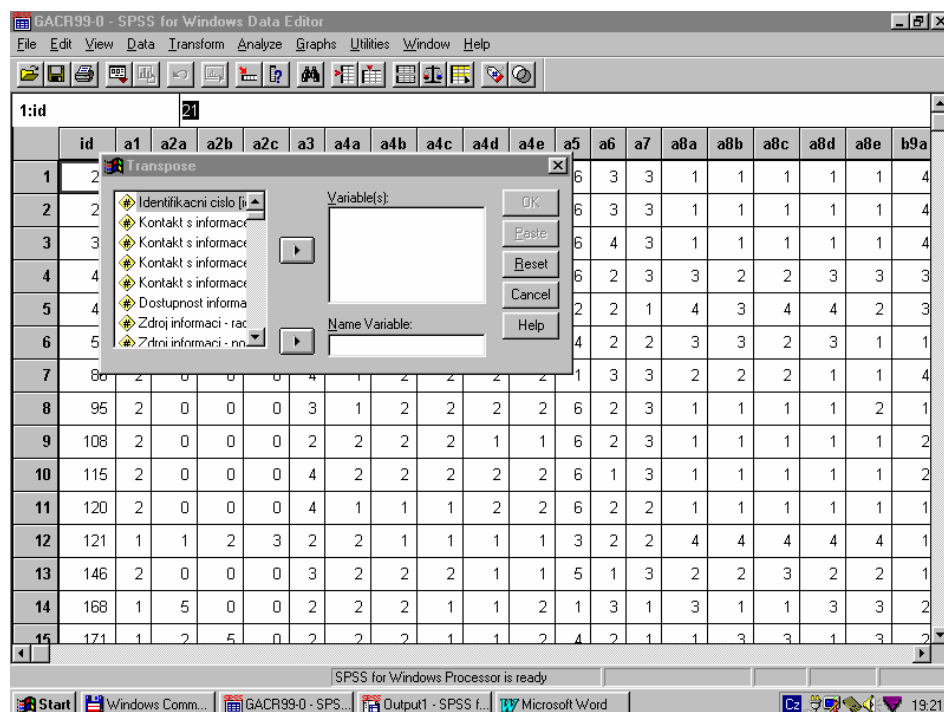
Máme personální databáze jednotlivých imatrikulačních ročníků studentů (každý ročník je samostatná matice dat) a chceme vytvořit jednotnou databázi studentů všech ročníků (jednu matici). Struktura matice je stejná: sledují se stejné proměnné (charakteristiky studentů) a v maticích jsou uvedeny ve stejném pořadí. K případům jednoho souboru se přidají případy druhého souboru.

	A1	A2	A3	A4	A5	A6	Ai	An
Adamec										
Blahá										
.....										
Zemina										

+

	A1	A2	A3	A4	A5	A6	Ai	An
Deml										

Stará
.....
Vechtr



1. 3. 2 TRANSPOSE

Příkaz TRANSPOSE vytváří nový datový soubor ve kterém jsou:

- původní řádky (případy) sloupce (proměnnými)
- původní sloupce (proměnné) řádkami (případy)

Automaticky se vytvářejí nová jména proměnných

1. 4 VÝBĚR PŘÍPADŮ Z VÝBĚROVÉHO SOUBORU

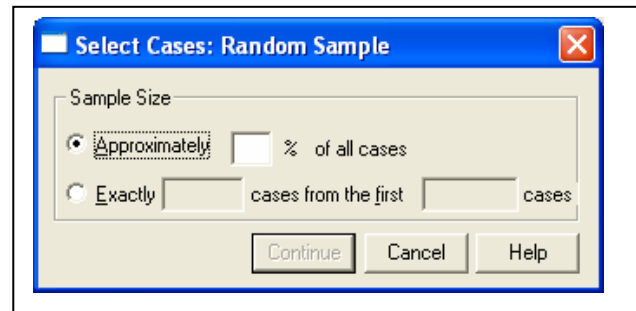
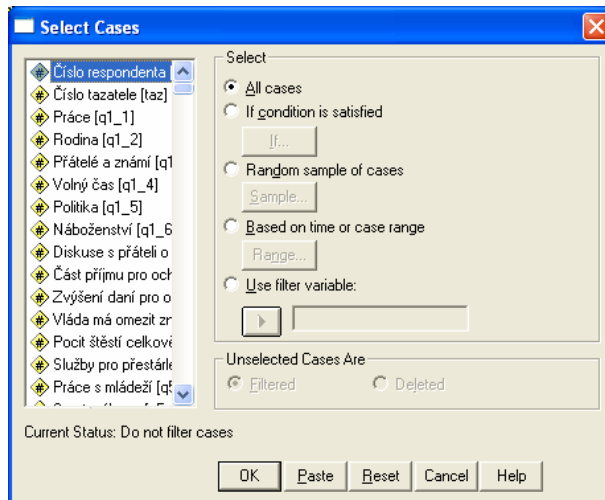
Výběr případů představuje manipulaci s datovým souborem, která nám umožní pracovat pouze s určitým podsouborem případů. Pomocí procedury SELECT CASES můžeme požadovaný podsoubor definovat:

- a) Podsoubor náhodně vybraných případů, máme-li například příliš velký základní soubor, jako tomu může být v případě dat z mikrocensu apod. Důvod redukce velikosti našeho výběrového souboru může být v tomto případě technický – operace probíhají rychleji. Navíc, kupodivu, mohou být naše výsledky přesnější.

Podsoubor vybraný na základě výzkumné otázky. Chceme provádět výpočty například jen s lidmi, kdo preferují určitou politickou stranu nebo s lidmi určité věkové skupiny (například s osobami staršími 60 let) apod. Práce s podsoubory předpokládá, že výběrový soubor je natolik velký, že má statistický smysl z něj vybírat soubor menší, podsoubor.

- b) 1. 4. 1 NÁHODNÝ VÝBĚR (RANDOM SAMPLE OF CASES)

Tato operace dovoluje vytvořit ze souboru náhodný výběr (omezit počet jeho jednotek při zachování reprezentativity souboru – samozřejmě, pokud byl reprezentativní původní soubor).



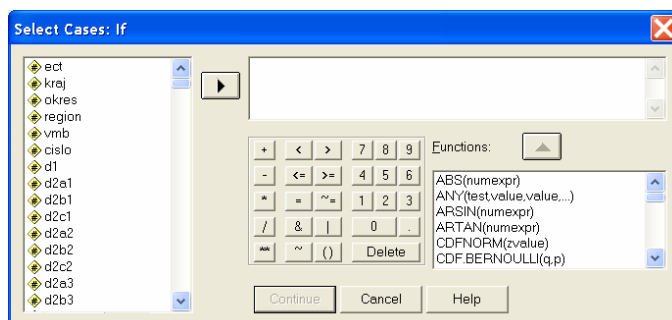
Můžeme zadat procento případů z celku, které chceme do výběru zahrnout, nebo přesný počet případů

Můžete nechat vybrat přibližný podíl z původního souboru, který stanovíme, nebo určitý počet případů (do rámečku *from the first cases* vypíšeme celkový počet jednotek původního souboru nebo někdy – spíše výjimečně – výběr omezíme jen na určitý počet případů). Co se týče rozhodnutí, co s nevybranými případy, používejte raději variantu *unselect cases are filtered*. Filtr lze odstranit a dále pracovat s celým souborem, pokud použijete variantu *unselect cases are deleted*, jsou všechny nevybrané případy (v matici se projeví přeškrtnutím jejich *id*) odstraněny. Pak musíte být velmi opatrní a přemýšlet, zdali takto redukovaný soubor chcete uložit, nebo. Pokud ho uložíte pod stejným jménem, původní soubor se přepíše a vám zůstane jen soubor s vybranými jednotkami (a právem také jen oči pro pláč, pokud byste neměli poslední podobu souboru zálohovanou).

1. 4. 2 IF CONDITION IS SATISFIED:

Někdy se může stát, že nás analyticky zajímají jen menší podsoubory (například jen ženy, nebo jen osoby se středoškolským vzděláním a vyšším, popřípadě jen osoby bydlící v Praze), a proto si je vybíráme, abychom další analytické výpočty prováděli jen s nimi.. Je pochopitelné, že je můžeme vybírat jen podle známých – zjištěných – charakteristik: pokud jsme například v dotazníku nezjišťovali místo bydliště respondenta, nemůžeme obyvatele Prahy vybrat.

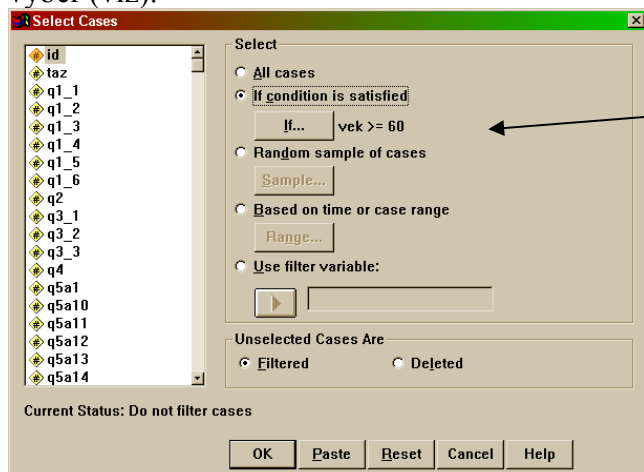
Podsoubory, s nimiž chceme pracovat, určujeme pomocí podmínky: do okénka vyklikáme nebo vypíšeme podmínku, např. $SEX=1$ (chceme-li pracovat jen s muži a víme, že v proměnné SEX 1=muž), $OBEC=15$ (chceme-li pracovat jen s obyvateli Prahy a víme, že v proměnné $OBEC$ Praha=15), $VZDEL > 2$ (chceme-li pracovat s osobami, jež mají středoškolské a vysokoškolské vzdělání, a víme, že v proměnné $VZDEL$ osoba se středoškolským vzděláním=3 a osoba s vysokoškolským vzděláním = 4).



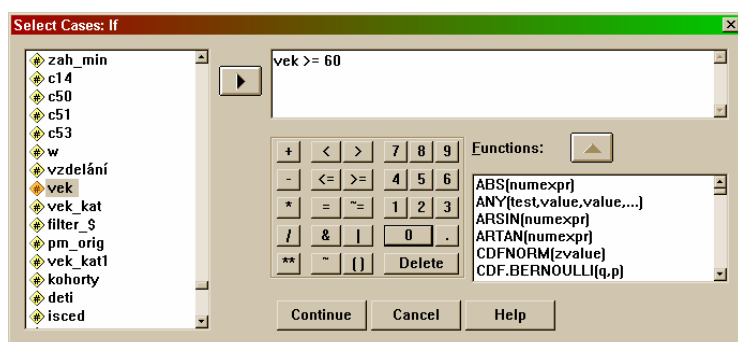
Např. nás může zajímat analýza lidí ve věku 60 let a starších. K výběru takového podsouboru použijeme proceduru

Data – Select Cases – If condition is satisfied

Po kliknutí na tlačítko **If...** se objeví dialogové okno, do nějž vepíšeme příslušnou podmínku pro výběr (viz).



Toho dosáhneme klikneme-li na *If* a zadáme-li podmínku výběru (viz následující obrázek)



Když si pro kontrolu necháme udělat rozložení takto redukováného souboru, získáme výsledek, který je uveden dole v tabulce (viz). Podmínky lze samozřejmě různě kombinovat, např. bylo by možné získat podsoubor mužů ve věku 60+ let, kteří ještě pracují apod.

Před přechodem k výpočtům opět s celým souborem nezapomeňte filtraci odstranit. Chceme-li v průběhu práce s daty ukončit práci s vybraným podsouborem a vrátit se k celému souboru, klikneme v SELECT CASES na RESET (filtr je odstraněn) nebo na ALL CASES (filtr je pouze vypnut, lze ho opětovně použít). Někdy se stane, zejména v časové tísní, že na to člověk zapomene, používá stále vybraný podsoubor a výsledky vydává za produkt výpočtů s celým souborem.⁶

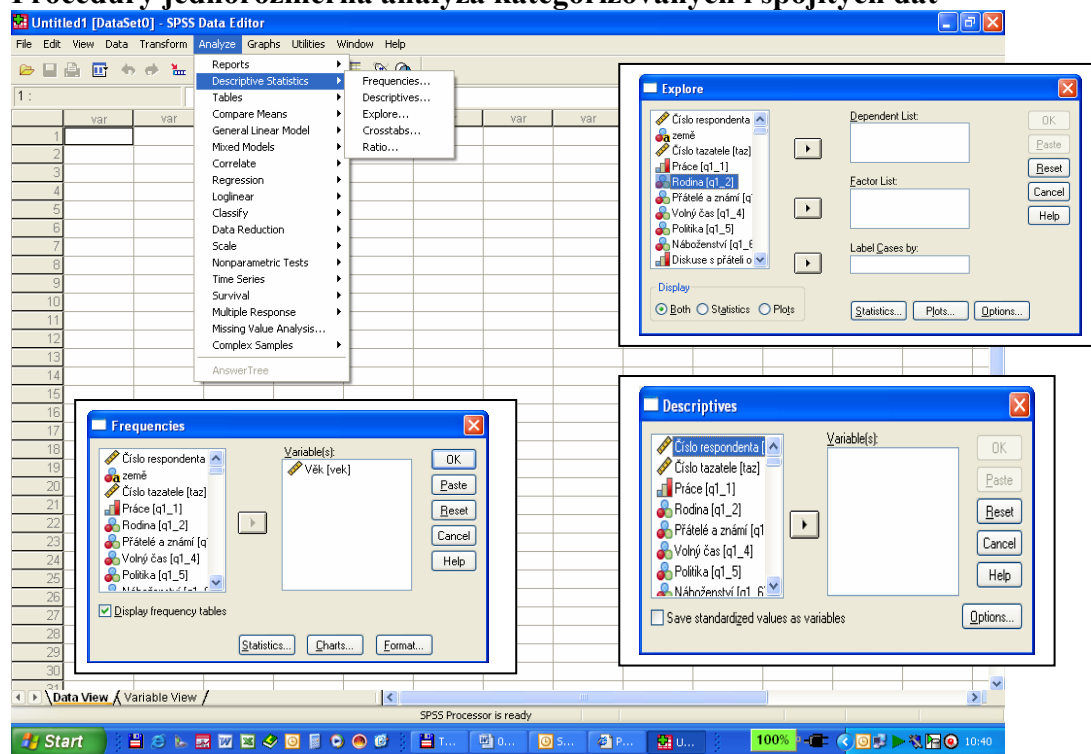
1. 5 ZÁKLADNÍ NABÍDKA SPSS

⁶ Pokud chcete začít pracovat s celým souborem a použili jste volbu *Deleted*, musíte si znovu otevřít původní soubor.

Pokusme se stručně ukázat, co využijeme ze široké nabídky SPSS pro naši práci.

1. 5. 1 POPISNÁ ANALÝZA SLOŽENÍ SOUBORŮ

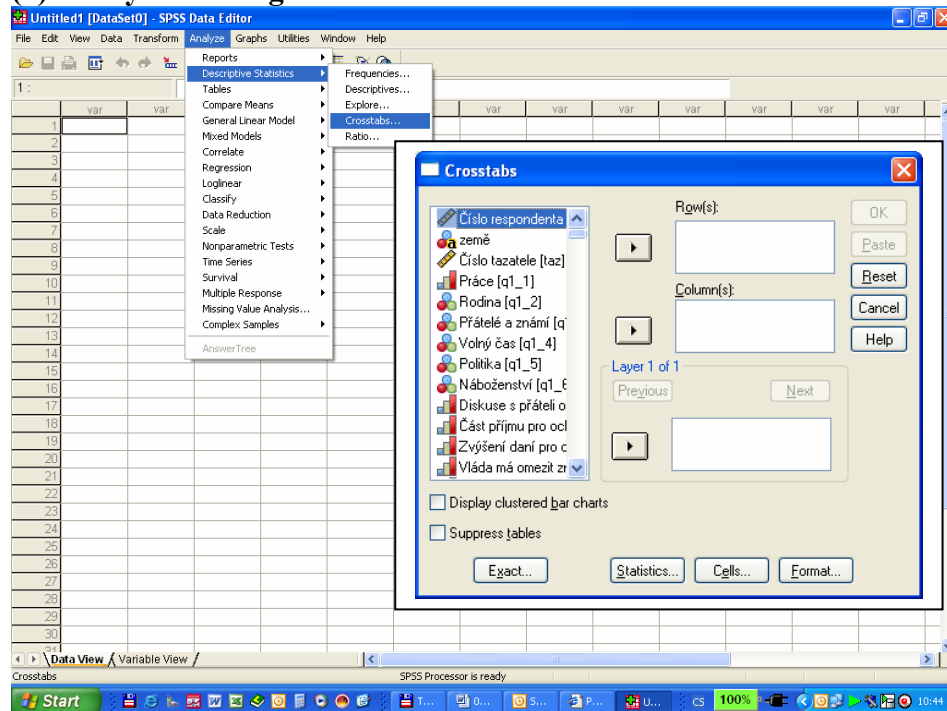
Procedury jednorozměrná analýza kategorizovaných i spojitých dat



Zahrnuje procedury FREQUENCIES, DESCRIPTIVES a EXPLORE.

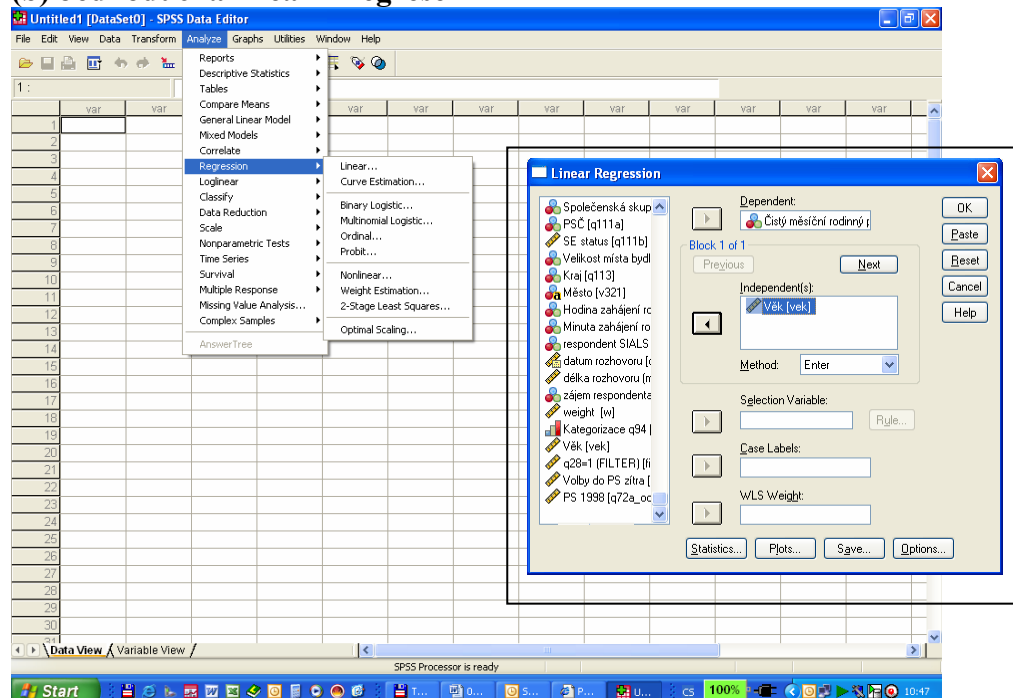
1. 5. 2 ANALÝZA VZTAHŮ MEZI DVĚMA CHARAKTERISTIKAMI VÝZKUM-NÝCH JEDNOTEK

(a) Analýza kontingenčních tabulek



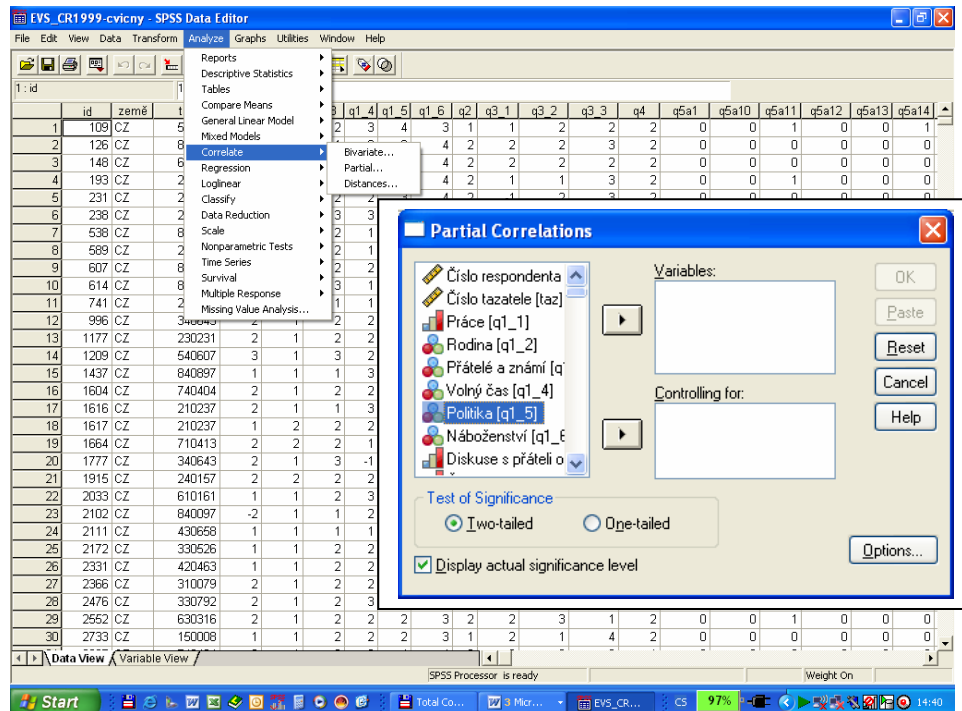
Procedura CROSSTABS.

(b) Jednoduchá lineární regrese



Procedura REGRESION - LINEAR.

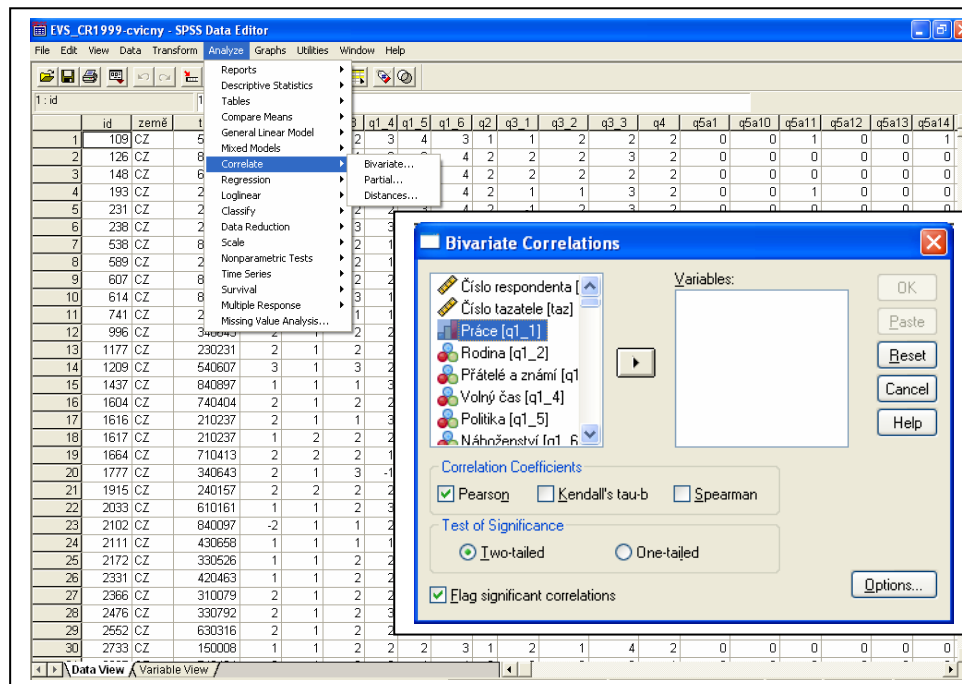
1. 5. 4 ELABORACE (ZJIŠŤOVÁNÍ EXISTENCE VLIVU TŘETÍ PROMĚNNÉ NA VZTAH MEZI 2 PROMĚNNÝMI.



Procedura CORRELATE – PARTIAL (CORRELATION)

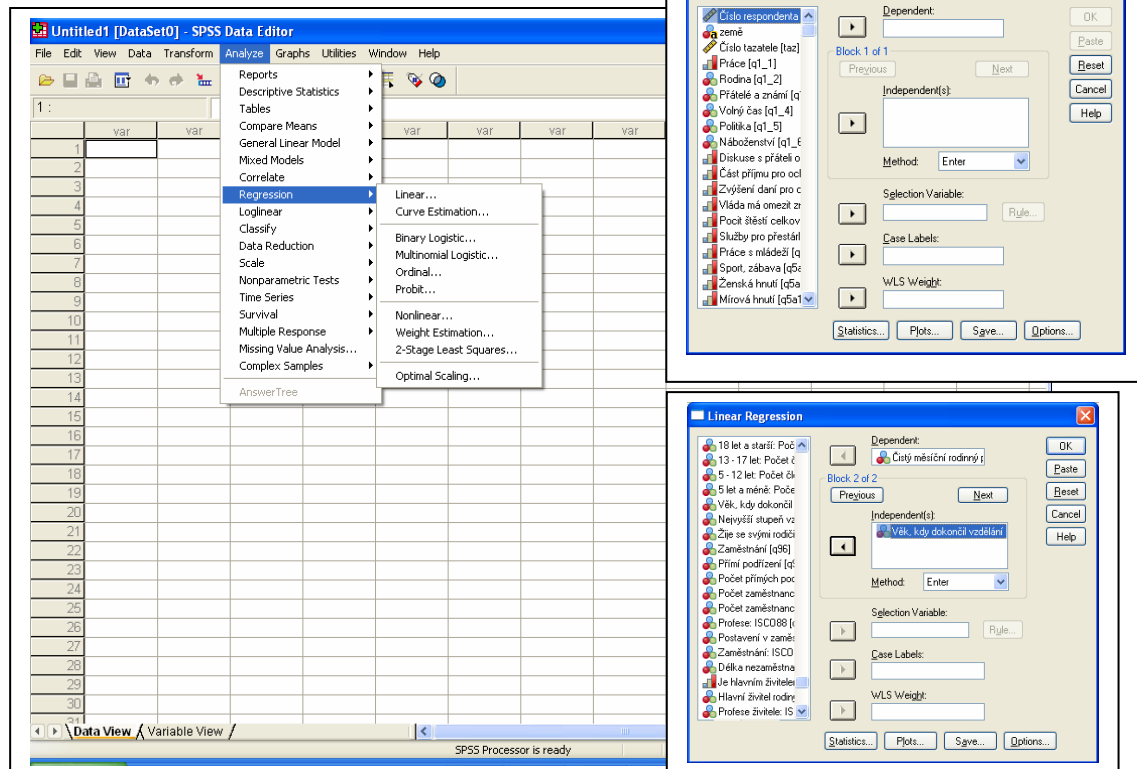
1. 5. 5 VZTAHY MEZI VÍCE PROMĚNNÝMI

(a) Matice koeficientů asociace a jejich analýza



Procedura CORRELATE – BIVARIATE

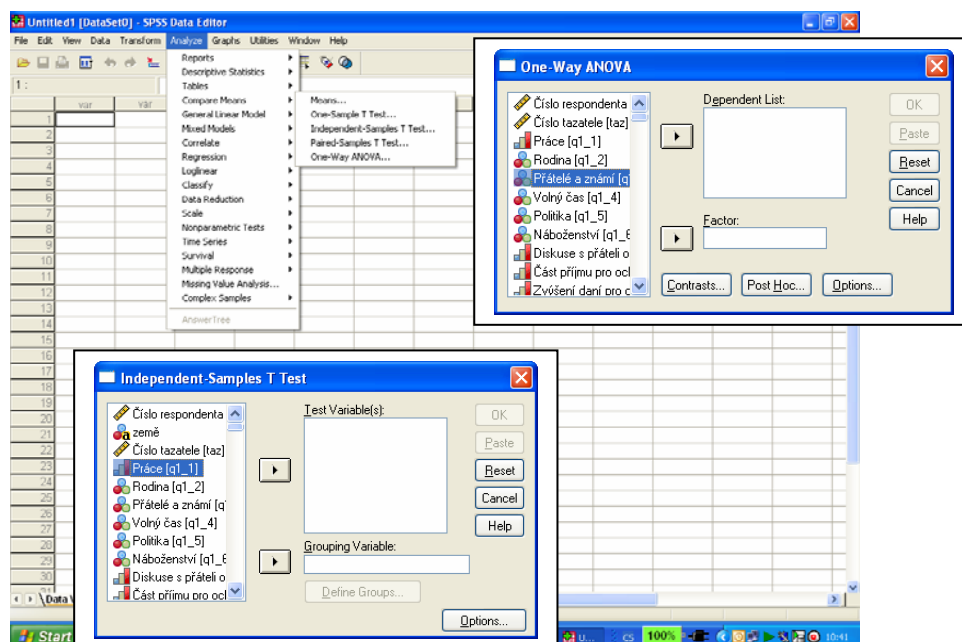
(b) Mnohonásobná lineární regrese



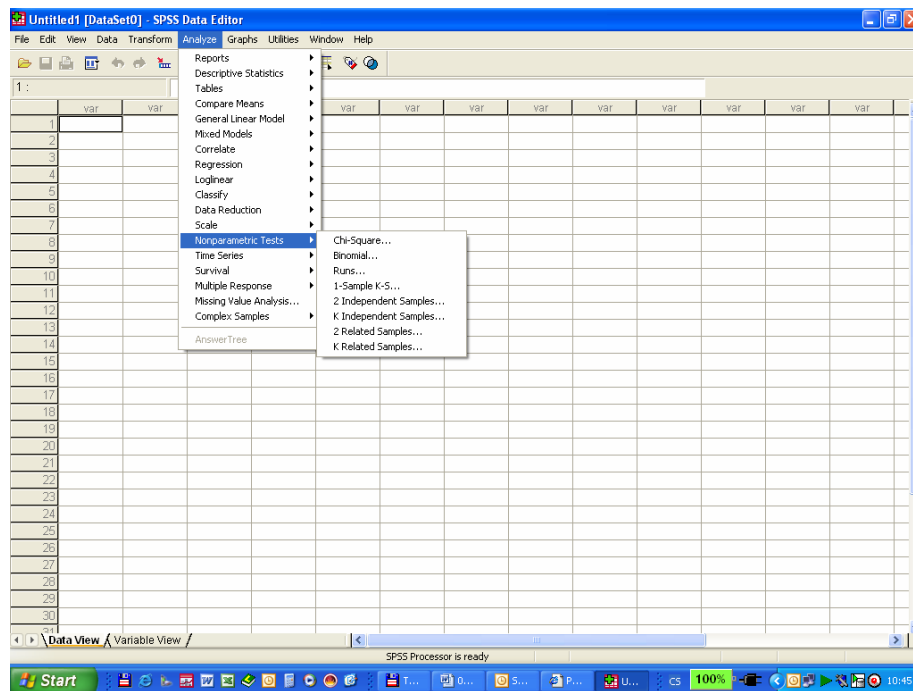
1. 5. 4 TESTOVÁNÍ SHODY A ROZDÍLU

Srovnávání skupin na základě středních hodnot jejich kardinálních charakteristik a testování hypotéz o shodě těchto středních hodnot (t-testy pro srovnání 2 průměrů a analýza variancí pro srovnání více průměrů)

Parametrické testy

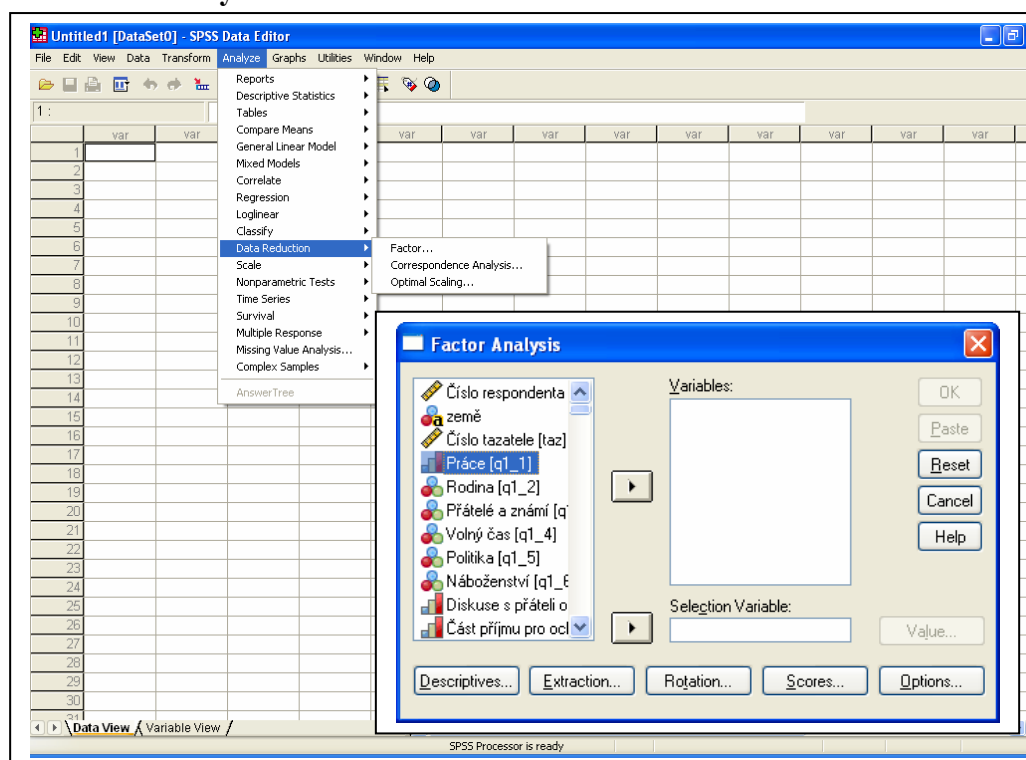


Procedury COMPARE MEANS – INDEPENDENT SAMPLES T TEST a další
Neparametrické testy



1. 5. 5 JEDNODUCHÉ HLEDÁNÍ STRUKTUR V ŘADĚ PROMĚNNÝCH

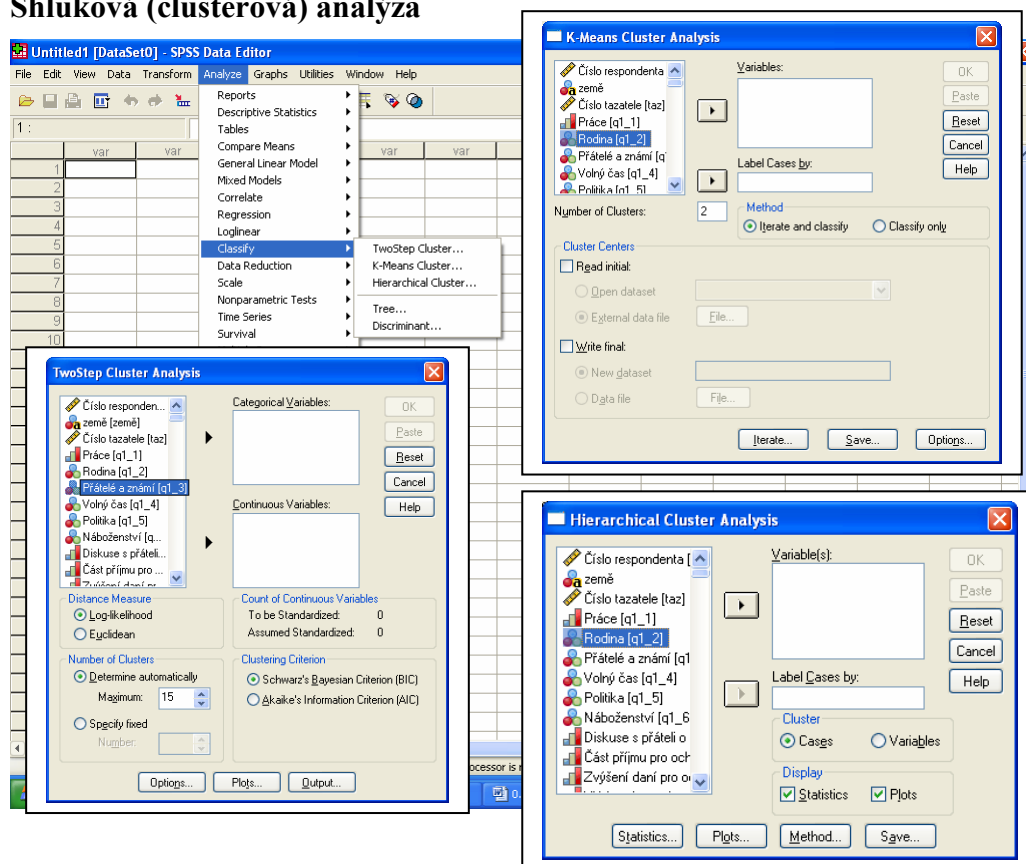
Faktorová analýza



Procedura DATA REDUCTION – FACTOR ANALYSIS

1. 5. 6 JEDNODUCHÉ HLEDÁNÍ TYPŮ VÝZKUMNÝCH JEDNOTEK

Shluková (clusterová) analýza



Procedura CLASIFY – TWO-STEP CLUSTER ANALYSIS

K MEANS CLUSTER ANALYSIS

HIERARCHICAL CLUSTER ANALYSIS

1. 5. 7 UPOZORNĚNÍ

SPSS nabízí velké množství statistických procedur, které sociolog málokdy detailně zvládne a ne všechny ve své praxi využije (ať již je to limitováno povahou jeho dat nebo rozsahem jeho znalostí statistiky). S některými z nich – s těmi nejfrekventovanějšími a jednoduššími se postupně seznámíme. Vždy je však třeba mít na mysli, že:

- Bychom neměli naše otázky (a tedy i charakter získávaných dat) přizpůsobovat statistickým procedurám, ale naopak bychom měli vyhledávat procedury, umožňující maximální využití našich dat. To vyžaduje, abychom:
 - Měli alespoň povědomí o tom, co každá z procedur nabízí.
 - Věděli, jaká má každá procedura požadavky na povahu škál, s nimiž pracuje. To neruší požadavek důrazu na věcnou stránku výzkumu, upozorňuje nás to však na to, že základní plán analýzy dat musíme mít již při koncipování výzkumu. Abychom mohli použít například faktorovou analýzu (její aplikace ovšem musí vycházet z toho, že nám pomůže smysluplně odpovědět naše výzkumné otázky a nikoliv z toho, že jsme se ji právě naučili používat nebo že je módní), musíme již ve svém dotazníku pro ni připravit vhodné

otázky (položky): musíme mít baterii alespoň 6-10 otázek se stupnicemi o stejném rozsahu s alespoň 5 hodnotami.

- Musíme mít stále na paměti, že program spočítá vše, co mu zadáme a nepřemýšlí (ani nemůže, není to myslící bytost, byť máme někdy tendenci jej personifikovat) o smysluplnosti zadání, ani o dodržení určitých požadavků na analýzu. My musíme např. vědět (ne ON), že nemá příliš smysl sledovat souvislost dvou proměnných pouze prostřednictvím charakteristiky chí kvadrát (vysvětlíme v příslušné kapitole), že není možné počítat z kódů nominální proměnné průměr nebo že není možné, aby při aplikaci lineární regrese byla závisle proměnná dichotomické povahy.. Pokud si některá základní pravidla statistiky nebudeme pamatovat, může se stát, že budeme často počítat (a přesností několik desetinných míst) a pak i sofistikovaně interpretovat naprosté nesmysly.