

Řešené statistické příklady v SPSS

Lektor: Ing. Martin Kořínek, Ph.D.

1 Krátce o SPSS

Statistických programů komerčně šířených existuje veliké množství. Jako nejpoblárnější příklady můžeme zmínit NCSS, SAS, S-Plus, Statistica, Stata, Minitab, Unistat nebo SPSS. To jsou tzv. statistické pakety, pokrývající celou škálu statistických metod. Jiné programy jsou specializované na analýzu některých dat (časové řady, kategoriální data, regresní analýza ap.). Všechny statistické programy však mají tyto základní funkce:

- import dat (vstup datové tabulky připravené v jiném programovém prostředí, třeba v Excelu nebo v Accessu),
- manipulace s daty (transformace, uspořádávání dat, výběry podmnožin datové matice, spojování datových matic),
- základní deskriptivní statistiky,
- grafické prostředky,
- ukládání dat k snadnému využití pro další zpracování,
- export dat (ve formátech vhodných pro jiné programové prostředí),
- prezentace výsledků ve formě souborů pro další zpracování textovými procesory.

Ovládání statistických programů je v současné době možné většinou přes menu a ikony podobně jako u ostatních programových produktů pracujících pod operačním systémem Windows (dříve převažovalo ovládání pomocí příkazového jazyka – command line –, které bylo poněkud náročnější nejen pro občasného uživatele nebo začátečníka).

Statistický paket SPSS je nejen na univerzitní půdě (a to nejen v České republice) poměrně značně rozšířen, což je tedy dostatečné objasnění, proč pro naše příklady byl vybrán právě tento statistický paket.

SPSS je jeden z nejstarších a nejrozšířenějších paketů. Na základě úspěšného paketu vznikla nadnárodní firma, která pohltila řadu dalších paketů (dosud existují např. i Systat, BMDP, Sigma Plot). Dnes podporuje firma SPSS mnoho produktů, nejznámější jsou Clementine (dataminig) a velmi názorný Amos (strukturní modely).

SPSS (zkratka původně znamenala Statistical Package for the Social Sciences) je americká softwarová firma sídlící v Chicagu, od roku 2009 vlastněná společností IBM, která ze statistického a analytického softwaru dodává především univerzální softwarový balík označovaný jako IBM SPSS. Předtím v letech 2009 až 2010 PASW (Predictive Analytics SoftWare), a před rokem 2009 SPSS. Software dodávaný od roku 1968 patří k předním hráčům na trhu analytických nástrojů.

Poznámka: Společnost SPSS CR, spol. s r.o., je výhradním distributorem softwaru SPSS a poskytovatelem analytických a statistických služeb v České a Slovenské republice. Společnost je součástí mezinárodní sítě SPSS, která působí ve více než 60 pobočkách po celém světě. Společnost SPSS CR byla založena v roce 1998 a nabízí široké portfolio programů zaměřených na statistické analýzy, business intelligence, data mining, na analýzy v marketingu, sběr dat a prezentaci výsledků. Dále poskytuje statistické služby, analýzy, modelování, predikce, řešení analytického CRM, nabízí výuku statistiky a poradenství v daném oboru

1.1 Základy ovládání

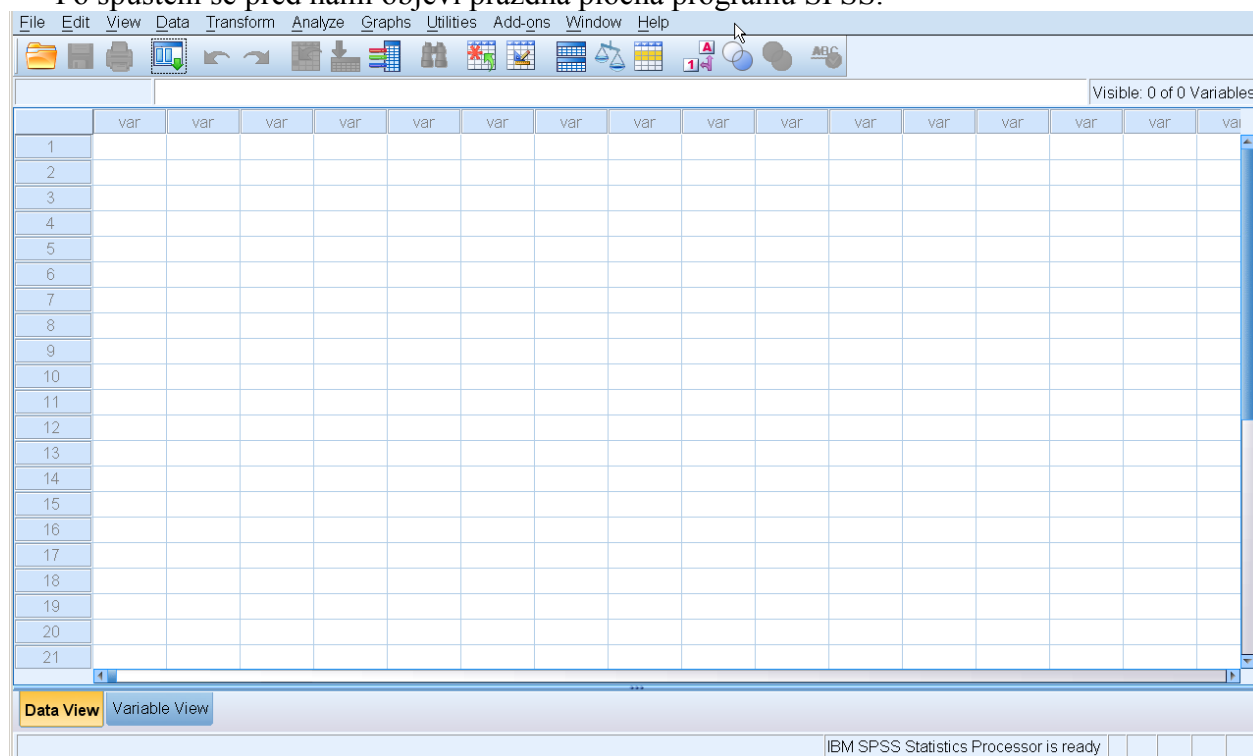
Poznámka: Následující text popisuje SPSS ve verzi 19 a pod operačním systémem MS Windows XP. Nicméně zde uváděné základy jsou platné pro ostatní verzi SPSS a operační systém MS Windows 7.

1.1.1 Spuštění

V prostředí MS Windows máme více možností, jak spustit SPSS. V podstatě se shodují s popisem spuštění MS Excel, který jsme si popsali výše a to včetně spuštění SPSS rovnou s některým již řešeným či vyřešeným příkladem či vstupními daty.

1.1.2 Popis pracovní plochy

Po spuštění se před námi objeví prázdná plocha programu SPSS.



Obrázek 1 Pracovní plocha SPSS

První řádek obsahuje menu programu (aktivuje se myší či klávesou Alt, případně klávesou Alt + písmenem z menu – tedy Alt-F pro menu File). Druhý řádek obsahuje ikony pro rychlejší spuštění nejpoužívanějších voleb.

Třetí řádek je vstupním řádkem – alternativa stejného řádku v MS Excel.

Předposlední řádek je informativním, kde například nalezneme informaci o aktuální pozici kurzoru. Na předposledním řádku nalezneme dvě záložky, jakoby listy. Ty slouží k přepínání mezi datovým listem a listem s popisem proměnných (s popisem sloupců). Ve „spreadsheetu“ můžeme přepínat mezi dvěma okny: Data View slouží pro zápis dat a Variable View pro práci s proměnnými.

No a poslední řádek nás informuje o názvu spuštěného programu a také o poloze kurzoru (včetně názvu vybraného menu atd.).

1.2 Okna programu SPSS

Program SPSS využívá čtyři základní typy oken:

- datové okno (Data Editor),
- výstupové okno (Viewer),
- syntaxové okno (Syntax) a
- skriptové okno (Script).

Jednotlivá okna mají odlišné funkce a tudíž i nabídky se v každém z nich mírně liší. **Datové okno** je určeno pro práci se vstupními daty a proměnnými. Skládá se ze dvou záložek – pohled na data (Data View) a pohled na proměnné (Variable View). Do **výstupového okna** se vkládají výstupy programu (tabulky, grafy, text, hlášení apod.). Program SPSS můžeme ovládat pomocí dialogových oken nebo pomocí jazyka syntax, který zapisujeme a spouštíme ze **syntaxového okna**. **Skriptové okno** slouží k psaní skriptů, tj. programů, které mohou automatizovat některé úlohy programu.

***Poznámka:** Náš popis se omezí na ovládání SPSS datového a výstupového okna. Ani syntaxí a už vůbec ne skriptováním se zabývat nebudeme – rozsah našeho materiálu to ani nepředpokládá.*

SPSS dovoluje pochopitelně otevřít současně více oken a přepínat mezi nimi. Tato vlastnost umožňuje například pracovat s několika výstupovými okny a řídit, do kterého z nich se výstupy zapíší. Z toho důvodu je třeba rozlišovat, které okno je aktivní (tzv. zaměřené), což poznáme podle zeleného křížku u ikony aktivního okna.

1.2.1 Datové okno (Data Editor)

Datové okno obsahuje informace o příkladech (datech) a proměnných. V záhlaví okna je uveden název souboru a v hranatých závorkách název datového okna (automaticky má tvar „DataSet+číslo“).

***Poznámka:** Pozor, v SPSS můžeme otevřít současně nejen několik datových souborů, ale dokonce několik verzí jednoho souboru. Proto je nezbytné rozlišovat, které okno je aktuální*

Datové okno je rozděleno do dvou listů: pohled na data (Data View) a pohled na proměnné (Variable View). Mezi listy přecházíme standardním způsobem známým z tabulkových procesorů

***Poznámka:** S daty v data pohledu (Data View) pracujeme velice podobně jako s obvyklou tabulkou ve spreadsheetu. Proto se detailním popisem nebudeme zde zabývat.*

Na záložce Variable View nalezneme informace o proměnných. Každý řádek zde představuje jednu proměnnou, ve sloupcích jsou charakteristiky proměnných (jméno, typ, počet míst, počet desetinných míst, popis, popis hodnot, uživatelem definované chybějící hodnoty, zarovnání sloupce, způsob měření).

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	VAR00001	Numeric	8	2		None	None	8	Right	Scale	Input
2	VAR00002	Comma	8	2		None	None	8	Right	Ordinal	Input

Obrázek 2 Okno Variable View v SPSS

V některých polích lze text přímo editovat, jinde je třeba pro zobrazení dialogu poklepat myší na tři tečky, které se zde po označení objeví. Většinu zadaných informací je rovněž možné kopírovat z jednoho pole na další.

Zobrazení sloupců ve Variable View lze přizpůsobit požadavkům uživatele a podle potřeb definovat další vlastnosti proměnných nebo naopak některé sloupce skrýt (viz View, Customize Variable View a Data, New Custom Attribute).

Nyní se pozastavíme u některých atributů u proměnných.

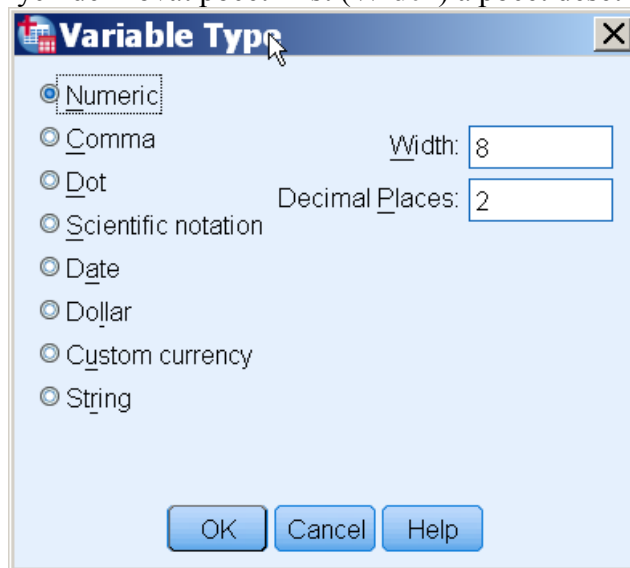
Název proměnné (**Name**) může obsahovat až 64 znaků, v praxi je však užitečné užívat spíše krátké názvy a případnou podrobnější informaci o proměnné doplnit do pole Label.

Poznámka: Název proměnné nesmí obsahovat mezery nebo různé speciální znaky (tečka, dvojtečka, čárka, střední apod.) ani nesmí začínat číslicí.

SPSS pracuje s těmito následujícími typy proměnné (**Type**):

- číselná (Numeric, Comma, Dot, Scientific Notation)
- datum (Date)
- textová (String)
- číselná obsahující měnu nebo jednotku (Dollar, Custom currency)

Dále lze u proměnných definovat počet míst (**Width**) a počet desetinných míst (**Decimals**)

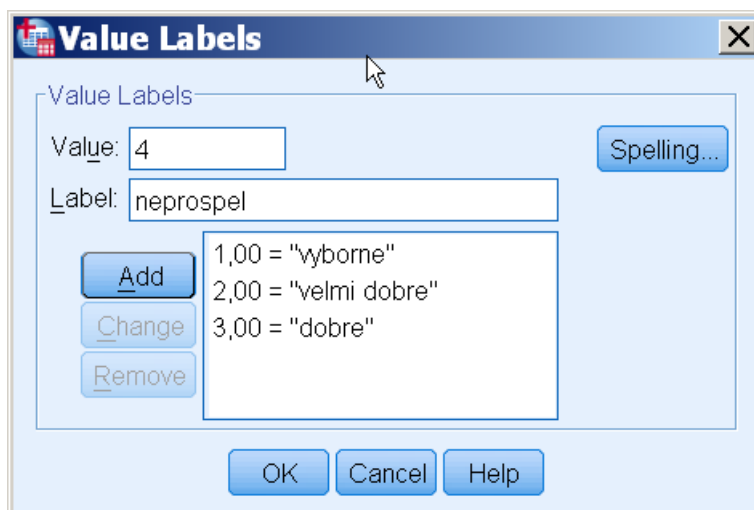


Obrázek 3 Nabídka typu proměnných v SPSS

Název proměnné bývá obvykle krátký. Ke každé proměnné však lze připojit také slovní komentář – popis proměnné (**Label**) – tak, aby byl zcela jasný její význam. Tento popis se rovněž může objevovat ve všech tabulkách a grafech (místo názvu, nebo současně s ním).

A kdy využít popis hodnot proměnné (**Values**)? Při zpracování dat je vhodnější pracovat s číselnými kódy než s textem. Pro porozumění proměnné ale potřebujeme znát také význam kódů – tzv. popisy hodnot. Popisy hodnot se potom mohou objevovat místo kódu, nebo společně s ním, ve výsledných tabulkách a grafech.

V dialogu vyplníme hodnotu (Value) i její popis (Label) a potvrdíme tlačítkem Add.



Obrázek 4 Použití volby Values Label v SPSS (Variable View)

Způsob měření (**Measure**) vyjadřuje typ proměnné. Statistika standardně rozlišuje proměnnou číselnou (scale), nominální (nominal) nebo ordinální (ordinal).

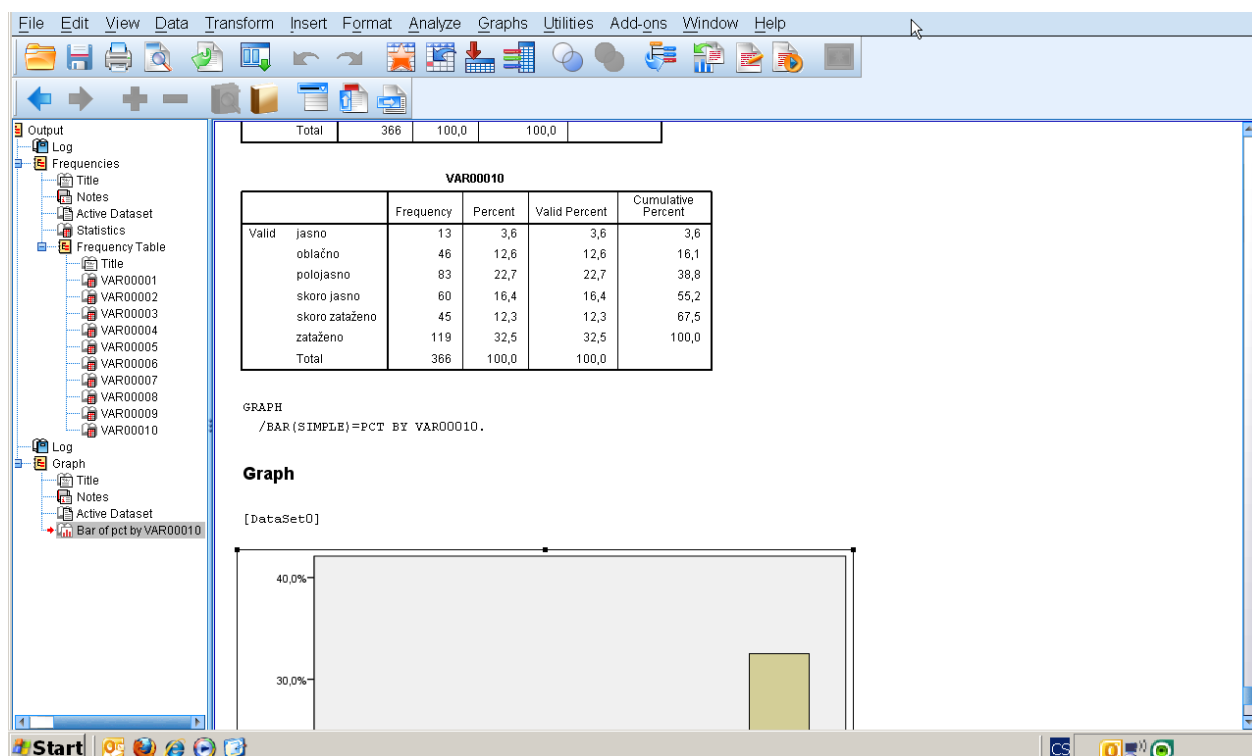
***Poznámka:** Jen malé opakování – číselné proměnné nabývají obvykle většího množství různých hodnot a uvedená hodnota u nich byla přímo změřena nebo jinak zjištěna. Ordinální i nominální proměnné jsou kategorizované proměnné. Kategorie ordinálních proměnných jsou uspořádané. Kategorie nominálních proměnných naopak uspořádané nejsou.*

1.2.2 Výstupové okno (Viewer)

Do výstupového okna se zaznamenává převážná většina výstupů SPSS (tabulky, grafy, text, hlášení, rozhodovací stromy apod.) v podobě, kterou lze do značné míry dále editovat nebo jinak upravovat. Okno je rozděleno na dvě části – na levé straně je obsah uspořádaný do stromové struktury, v pravé části se nacházejí jednotlivé objekty výstupu.

Názvy položek obsahu snadno upravíme poté, co na zvolenou položku dvakrát poklepeme myší. Pomocí myši nebo s využitím nabídek a ikon lze také měnit uspořádání položek nebo jejich hierarchii. Po označení určitého objektu ve výstupovém okně se u něj objeví červená šipka a můžeme ho smazat, zkopírovat a vložit na jiné místo, přenést myší, zobrazit nebo skrýt atd.

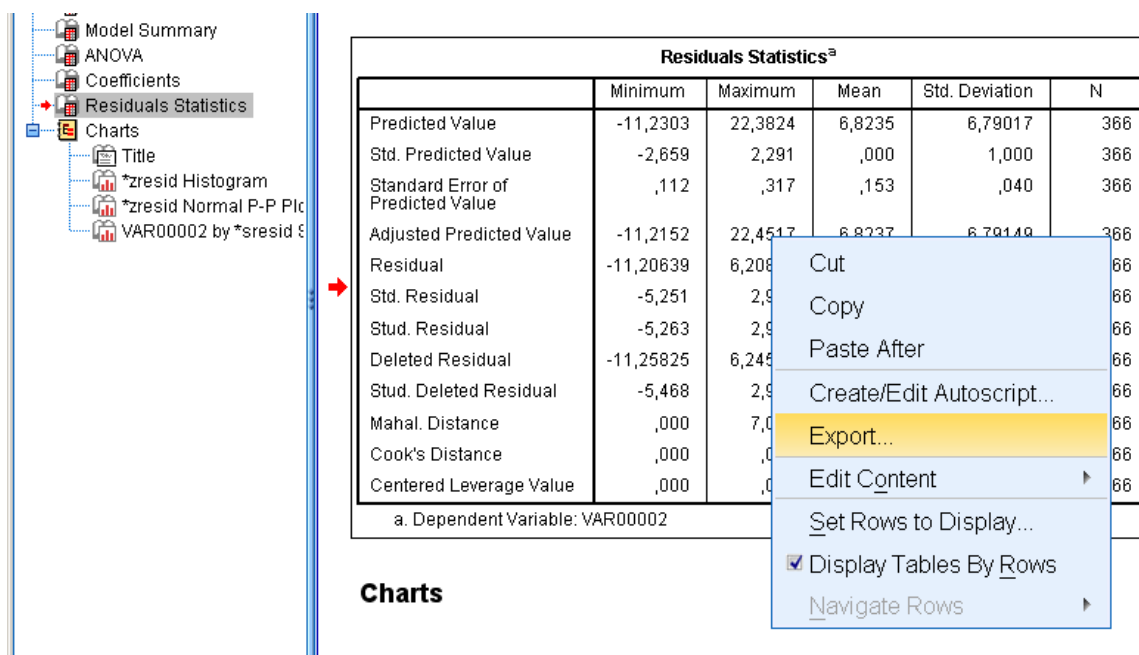
Jednotlivé objekty výstupového okna lze rovněž dále upravovat. Pro přístup k editacím je nutné na objekt dvakrát poklepat myší a uvést ho do editačního modu. Uživatelské nabídky i ikony se přizpůsobí danému objektu, v některých případech se objekt dokonce otevře v novém okně. Úpravy ukončíme kliknutím myši mimo objekt ve výstupovém okně, nebo uzavřením editačního okna.



Obrázek 5 SPSS - okno Output

Poznamenejme si, že jednotlivé sekce okna Viewer lze pomocí Windows schránky zkopírovat do téměř libovolné Windows aplikace – nejčastěji asi použijeme kopírování do MS Word či MS Excel.

Náš další obrázek ukazuje situaci, kdy jsme vybrali jednu tabulku z výstupového okna a pomocí pravého tlačítka myši jsme si zvolili možnost Exportu obsahu těchto výsledků.



Obrázek 6 SPSS - okno Viewer a volba pro Export vybrané části

1.3 Vybrané důležitější operace v SPSS

1.3.1 Nový soubor

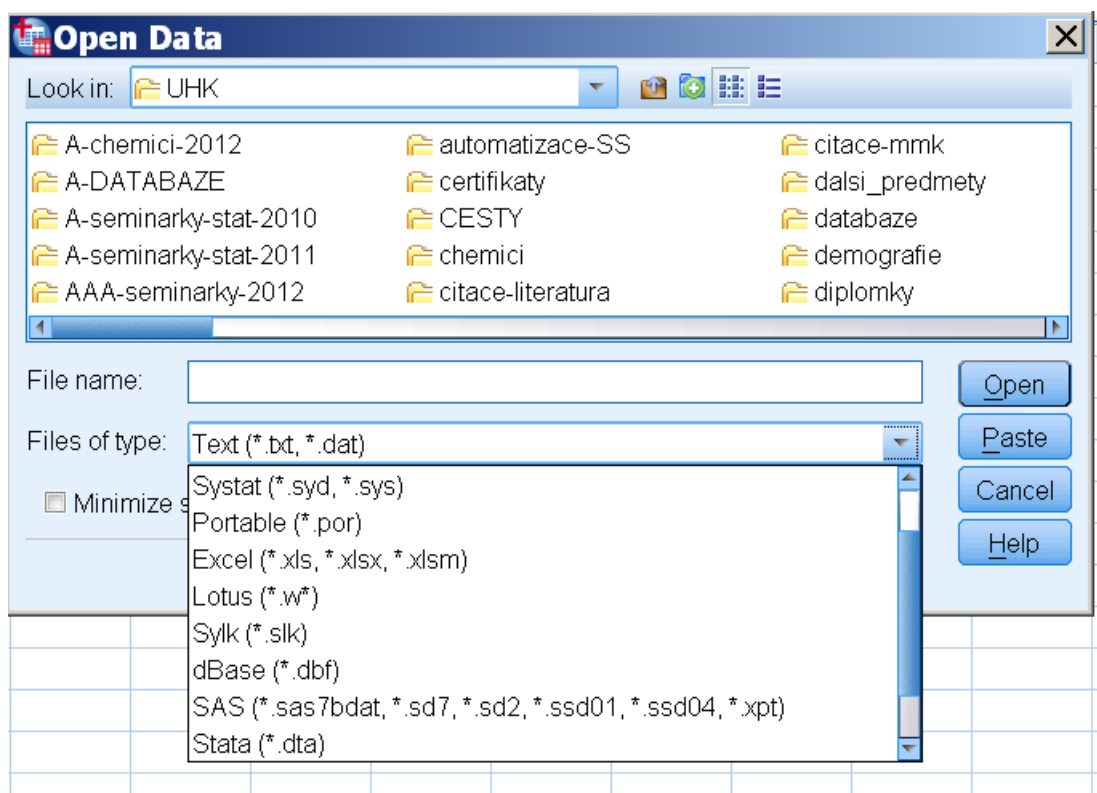
Pro vytvoření nového souboru zvolíme menu File/New/Data. Otevře se prázdné datové okno, do kterého můžeme zadávat nový příklad. Nový datový soubor má standardně přiřazeno jméno „Untitled+číslo“ (například Untitled2).

Poznámka: Nezapomeňme, že v SPSS můžeme mít otevřeno více oken, více nových souborů (příkladů).

1.3.2 Import dat

Vedle vkládání nových dat (či případně otevření souboru s již jednou uloženými daty pomocí menu File/Open/Data) můžeme data do SPSS importovat data (soubory, příklady) a to z mnoha formátů.

Po otevření menu File/Read Text Data si můžeme vybrat například z formátů typu Excel, Lotus, dBase, Sylk a z mnoha dalších (včetně obecného formátu TXT. V neposlední řadě lze do NCSS importovat data z jiných statistických paketů (SAS, Stata, Systat).



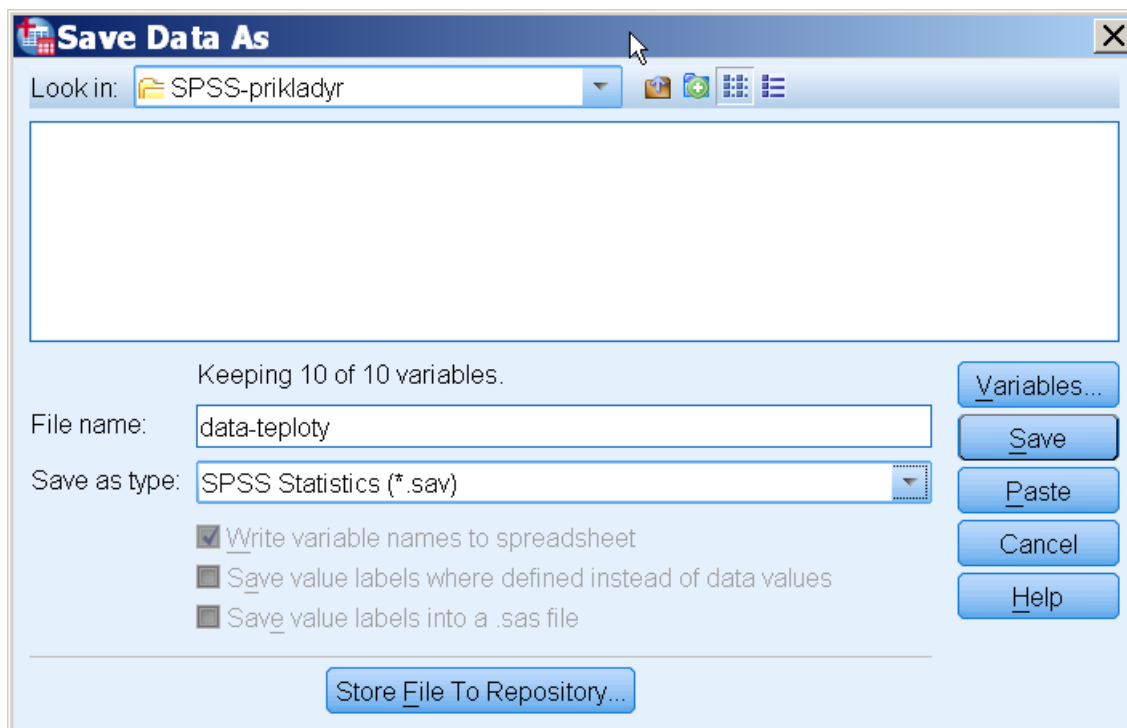
Obrázek 7 SPSS - import dat

Po vybrání daného datového zdroje (souboru) se spustí průvodce, který nám s importem dat pomůže (například u Excel souboru nám poradí, který list chceme importovat a jakou oblast dat – zda-li celou tabulku či bez nadpisu atd.).

1.3.3 Uložení datového souboru

Rozpracovaný datový soubor je vhodné pravidelně ukládat, abychom o data nepřišli. Pro ukládání slouží menu File/Save.

Pokud ukládáme poprvé či pokud chceme umístění a název datového souboru změnit, zvolíme menu File/Save As. Ve standardním okně můžeme vybrat složku, do které se soubor uloží a rovněž název datového souboru.



Obrázek 8 Uložení datového souboru v SPSS

Poznamenejme, že vedle uložení do standardního formátu SPSS můžeme náš datový soubor exportovat (nabídka Save as type) do mnoha formátů – výčet je velmi podobný s nabídkou pro import dat.

1.4 Statistické procedury

V této části si již nastíníme, jakým způsobem budeme spouštět vybranou statistickou analýzu našeho datového souboru.

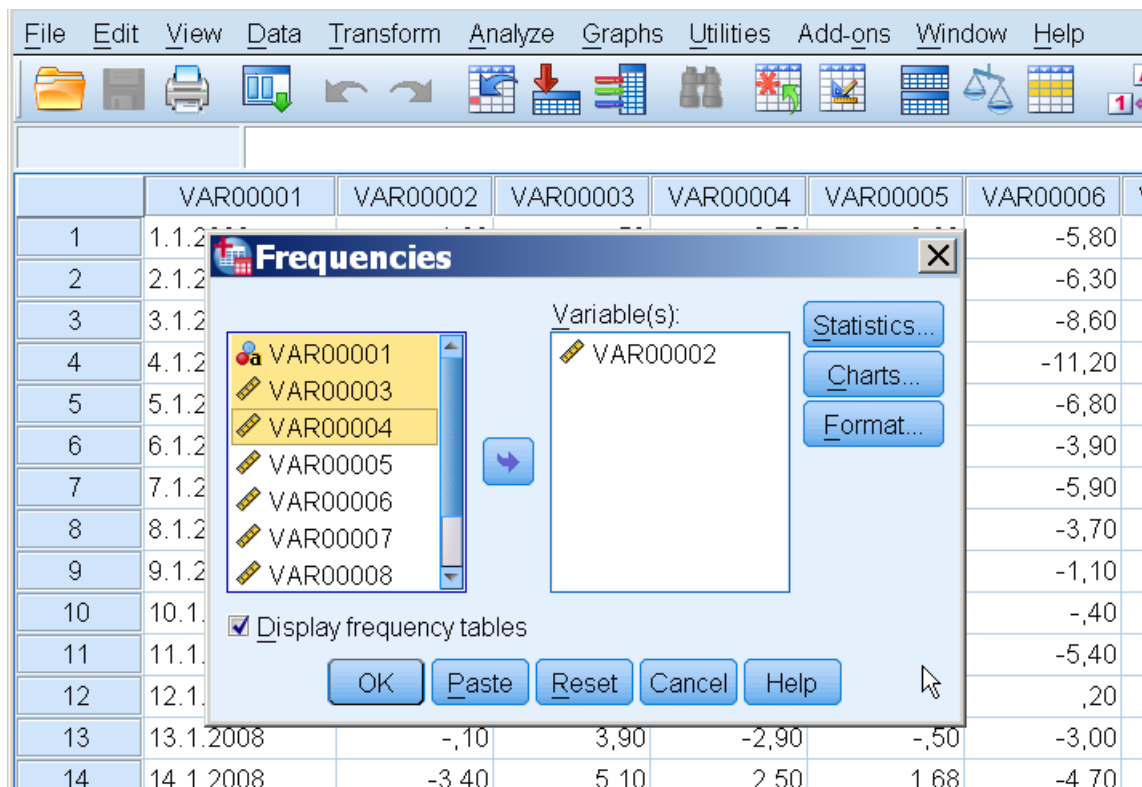
1.4.1 Dvě hlavní nabídky v SPSS

Program SPSS ovládáme pomocí těchto dvou důležitých oken.

- **Analyze** – zobrazuje nabídku celé řady statistických procedur, operací. Musíme ovšem upozornit na to, že program sice nabídne všechny statistické procedury, ale statistik musím vybrat právě tu, která je pro konkrétní data (příklad) jedinečně správná, odpovídající. Proto je nutné, abych program SPSS ovládali pouze statisticky poučení uživatelé.
- **Graphs** je nabídka pro grafické znázornění dat, které máme v aktivním datovém okně.

1.4.2 Postup před spuštěním procedury

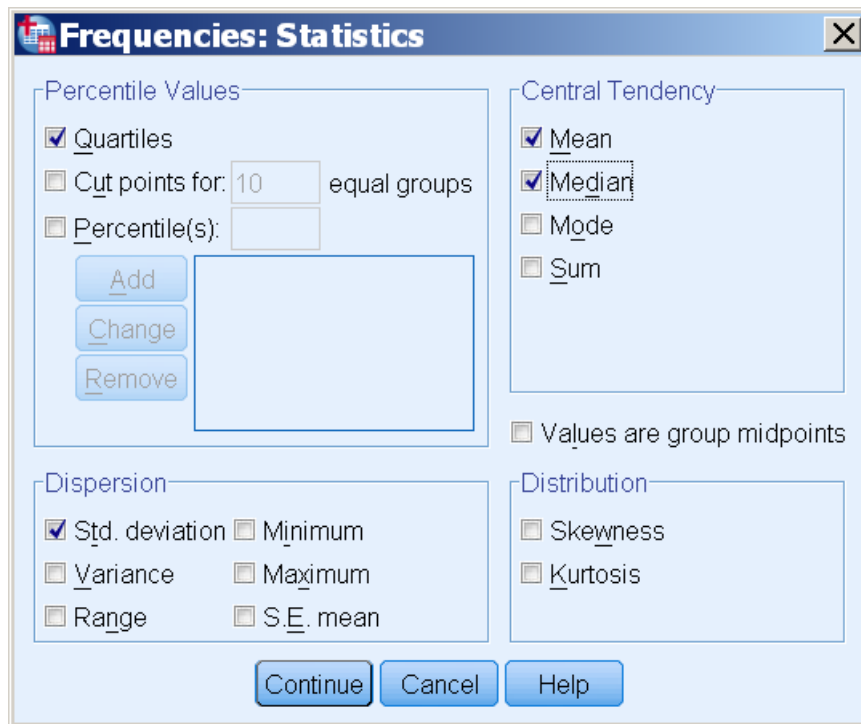
Je nutno poznamenat, že po vybrání libovolné statistické procedury musíme nejprve v okně zadat, které proměnné požadujeme zpracovávat – nemusíme analyzovat všechny „sloupce“, ale pouze některé.



Obrázek 9 Výběr proměnných pro zpracování (SPSS - Analyze)

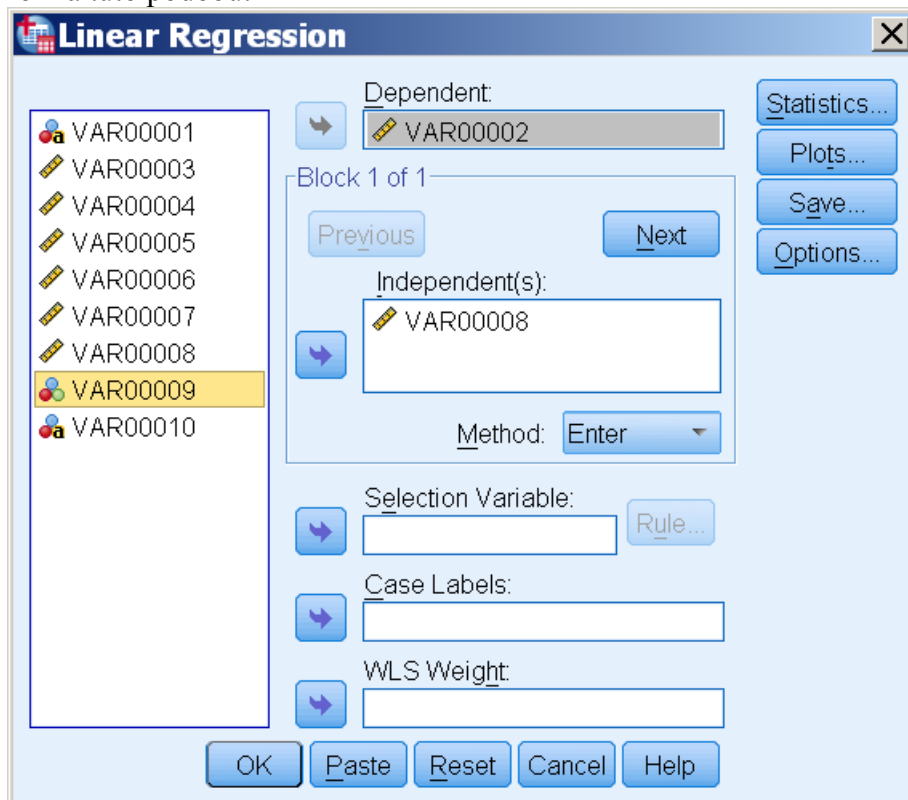
V dialogovém okně si z levého sloupce vybereme proměnné, které přesuneme do okna pravého (lze, jak je vidět na našem obrázku, využít i označení i více proměnných najednou).

V dalším kroku musíme detailně specifikovat, které statistiky požadujeme ve vybrané proceduře zpracovat. Výběr je snadný, volíme v dialogovém okně, které se objeví pod nabídkou Statistics (případě Options). Například při statistické proceduře Descriptive Statistics/Frequencies nabídka vypadá takto:



Obrázek 10 Nabídka statistik při proceduře Frequencies

Je zcela pochopitelné, že podoba oken a nabídky jsou pro jednotlivé procedury individuální. Například pro proceduru Lineární regrese (Analyze/Regression/Linear) vstupní dialogové okno má tuto podobu:



Obrázek 11 SPSS nabídka dialogového okna pro lineární regresi

Po zadání parametrů, tedy po konkretizaci statistik, které požadujeme zjistit, tlačítkem OK spustíme výpočet. Výsledek se objeví v novém, výstupovém okně (Viewer).

Sylabus kurzu Řešené statistické příklady v SPSS

Lektor: Ing. Martin Kořínek, Ph.D.

Statistický paket SPSS

Krátce o SPSS

Základy ovládání

Spuštění

Popis pracovní plochy

Okna programu SPSS

Datové okno (Data Editor)

Výstupové okno (Viewer)

Vybrané důležitější operace v SPSS

Nový soubor

Import dat

Uložení datového souboru

Statistické procedury

Dvě hlavní nabídky v SPSS

Postup před spuštěním procedury

Řešené příklady

Tabulky a grafy (Příklad 1 – 4)

Míry polohy (průměr, medián, modus, kvantily), Příklad 5 - 7

Míry variability (variační rozpětí, rozptyl, směrodatná odchylka, variační koeficient)

Test chí-kvadrát (kontingenční tabulka)