

# **IBM SPSS Statistics Base 24**

---

**Statistické procedury**



**ACREA CR, spol. s r.o.**  
Kraťovská 7, 110 00 Praha 1  
tel./fax: 234 721 444  
email: [info@acrea.cz](mailto:info@acrea.cz)  
[www.acrea.cz](http://www.acrea.cz)

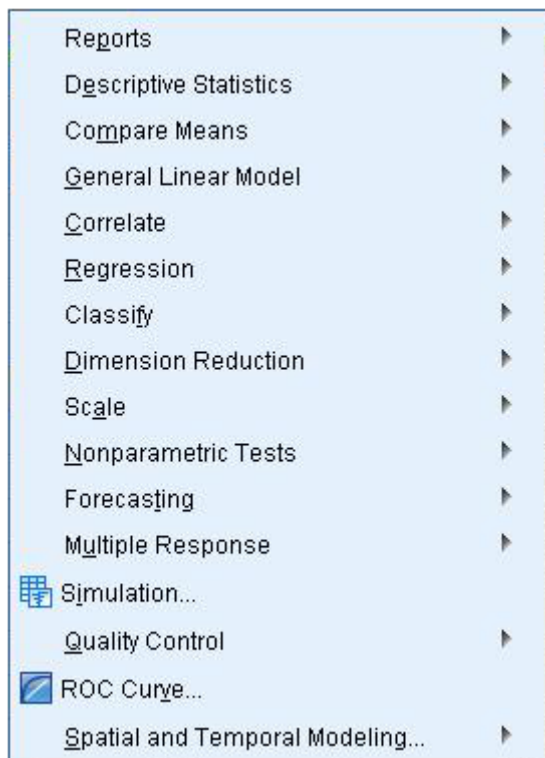


## Obsah

<b>1</b>	<b>Analýzy (Analyze)</b> .....	<b>5</b>
1.1	Zprávy (Reports) .....	8
1.1.1	Shrnutí informací o proměnných a datovém souboru (Codebook) .....	8
1.1.2	Kostky OLAP (OLAP Cubes) .....	9
1.1.3	Sumář případů (Case Summaries) .....	11
1.1.4	Sumář v řádcích (Summaries in Rows) .....	13
1.1.5	Sumář ve sloupcích (Summaries in Columns) .....	16
1.2	Popisné statistiky (Descriptive Statistics) .....	19
1.2.1	Četnosti (Frequencies) .....	19
1.2.2	Popisné statistiky (Descriptives) .....	21
1.2.3	Explorace (Explore) .....	23
1.2.4	Kontingenční tabulky (Crosstabs) .....	24
1.2.5	Poměrové statistiky (Ratio) .....	27
1.2.6	Graf kumulativních proporcí (P-P Plots) .....	29
1.2.7	Kvantilový graf (Q-Q Plots) .....	31
1.3	Porovnání průměrů (Compare Means) .....	33
1.3.1	Průměry (Means) .....	33
1.3.2	Jednovýběrový T-test (One-Sample T Test) .....	34
1.3.3	T-test pro dva nezávislé výběry (Independent-Samples T Test) .....	36
1.3.4	Párový T-test (Paired-Samples T Test) .....	38
1.3.5	Jednoduchá analýza rozptylu (One-Way ANOVA) .....	39
1.4	Obecný lineární model (General Linear Model) .....	41
1.4.1	Jednoduchý lineární model (Univariate) .....	41
1.5	Korelační analýza (Correlate) .....	44
1.5.1	Korelace (Bivariate Correlations) .....	44
1.5.2	Parciální korelace (Partial Correlations) .....	45
1.5.3	Vzdálenosti (Distances) .....	47
1.6	Regresní analýza (Regression) .....	49
1.6.1	Automatická tvorba lineárního modelu (Automatic Linear Modeling) .....	49
1.6.2	Lineární regrese (Linear Regression) .....	51
1.6.3	Prokládání křivky (Curve Estimation) .....	54
1.6.4	Ordinální regrese (Ordinal) .....	55
1.7	Klasifikační metody (Classify) .....	58
1.7.1	Dvoustupňové seskupování (TwoStep Cluster) .....	58
1.7.2	Seskupování metodou K-means (K-Means Cluster) .....	61
1.7.3	Hierarchické seskupování (Hierarchical Cluster) .....	63
1.7.4	Diskriminační analýza (Discriminant Analysis) .....	65
1.7.5	Analýza nejbližších sousedů (Nearest Neighbor Analysis) .....	67
1.8	Redukce dat (Dimension Reduction) .....	71
1.8.1	Faktorová analýza (Factor Analysis) .....	71
1.9	Škálování (Scale) .....	73
1.9.1	Analýza reliability (Reliability Analysis) .....	73
1.9.2	Mnohorozměrné škálování (Multidimensional Scaling) – ALSCAL .....	73

1.10	Neparametrické testy (Nonparametric Tests) .....	76
1.10.1	Jednovýběrové neparametrické testy (One Sample).....	76
1.10.2	Neparametrické testy pro nezávislé výběry (Independent Samples) .....	78
1.10.3	Neparametrické testy pro závislé výběry (Related Samples).....	80
1.10.4	Test Chí-kvadrát (Chi-square Test) .....	82
1.10.5	Binomický test (Binomial Test).....	84
1.10.6	Test pořadí (Runs Test).....	85
1.10.7	Jednovýběrový Kolmogorov-Smirnovův test (One-Sample Kolmogorov-Smirnov test) .....	87
1.10.8	Testy pro dva nezávislé výběry (Two-Independent-Samples Tests) .....	87
1.10.9	Testy pro několik nezávislých výběrů (Tests for Several Independent Samples) 89	
1.10.10	Testy pro dva závislé výběry (Two-Related-Samples Tests).....	90
1.10.11	Testy pro několik závislých výběrů (Tests for Several Related Samples) .....	92
1.11	Časové řady (Forecasting) .....	93
1.11.1	Graf časových řad (Sequence Charts) .....	93
1.11.2	Autokorelace (Autocorrelations).....	95
1.11.3	Vzájemné korelace (Cross-Correlations) .....	97
1.12	Mnohonásobné odpovědi (Multiple Response).....	100
1.12.1	Sady proměnných pro mnohonásobné odpovědi (Define Sets) .....	100
1.12.2	Výpočet četností pro mnohonásobné odpovědi (Frequencies).....	101
1.12.3	Kontingenční tabulky pro mnohonásobné odpovědi (Crosstabs).....	102
1.13	Simulace (Simulation).....	105
1.14	Kontrola kvality (Quality Control) .....	120
1.14.1	Řídící diagramy (Control Charts) .....	120
1.14.2	Paretův graf (Pareto Charts) .....	121
1.15	ROC křivka (ROC Curve).....	125
1.16	Časoprostorové modelování (Spatial and Temporal Modeling) .....	128
1.16.1	Geoprostorové modelování (Spatial Modeling).....	128
1.16.2	Zobecněná geoprostorová asociační pravidla (Geospatial Association Rules) 139	
1.16.3	Časoprostorové predikce (Spatio-Temporal Prediction) .....	150
<b>Příloha 1 – Podmíněné formátování tabulek .....</b>		<b>163</b>
<b>Rejstřík .....</b>		<b>166</b>
<b>Anglicko – český slovník .....</b>		<b>169</b>
<b>Česko – anglický slovník .....</b>		<b>174</b>

# 1 Analýzy (Analyze)



Nabídka **Analyze** obsahuje četné procedury pro statistickou analýzu dat. Vzhledem k tomu, že systém IBM SPSS Statistics je modulární a jednotlivé moduly jsou plně propojené, můžeme v tomto menu najít nejen volby, které patří do základního modulu Statistics Base, ale také nabídky a podnabídky dalších modulů (pokud jsou k dispozici). V následujících odstavcích se zaměříme pouze na ty, které jsou součástí modulu IBM SPSS Statistics Base. Při instalaci integračních pluginů pro jazyky Python nebo R se nabídka rovněž automaticky rozšíří o řadu ukázkových procedur. Jejich přehled je uveden ve skriptech *IBM SPSS Statistics Base 24: Datový soubor a práce výstupy* v přílohách.

Nabídka **Reports** umožňuje shrnout základní informace o proměnných a datovém souboru (*Codebook*), spočítat průměry, úhrny a další statistiky a zobrazit je v podrobněji tříděných tabulkách (*OLAP Cubes*) nebo vytvářet sumáře.

V nabídce **Descriptive Statistics** získáme základní informace o rozložení proměnných.

Procedury *Frequencies* a *Descriptives* zobrazí přehled vybraných popisných statistik doplněný zvolenými grafy. Tabulky lze rovněž třídit a provádět podrobnou analýzu číselných proměnných (*Explore*), nebo pracovat s kontingenčními tabulkami (*Crosstabs*). Procedura *Ratio* je určena k analýze poměru dvou veličin. Graf kumulativních proporcí (*P-P Plots*) a kvantilový graf (*Q-Q Plots*) slouží k porovnání výběrového rozložení proměnné s vybraným teoretickým rozdělením.

Volba **Compare Means** obsahuje procedury určené k porovnání středních hodnot a testování rozdílů mezi nimi. Kromě výpočtu průměrů a dalších statistik pro podrobnější třídění je možné zvolit jednovýběrový t-test (*One-Sample T Test*), testující hypotézu, že střední hodnota proměnné je rovna zadané konstantě; t-test pro dva nezávislé výběry (*Independent-Samples T Test*), který testuje hypotézu, že rozdíl středních hodnot dvou nezávislých výběrů je roven nule; párový t-test (*Paired-Samples T Test*) hypotézy, že rozdíl dvou závislých výběrů je roven nule a analýzu rozptylu jednoduchého třídění (*One-Way ANOVA*), ověřující hypotézu, že střední hodnoty ve skupinách podrobnějšího třídění podle zvoleného faktoru se rovnají.

Nabídka **General Linear Model** zahrnuje v základním modulu IBM SPSS Statistics Base jednoduchý lineární model (*Univariate*).

Ke sledování vzájemné závislosti dvou proměnných je určena nabídka **Correlate**, kde lze zvolit výpočet různých typů korelačních koeficientů (*Bivariate Correlations*), výpočet parciálních korelací (*Partial Correlations*) a výpočet vzdáleností (nebo jiných měr podobnosti či nepodobnosti) pro proměnné nebo případy (*Distances*).

Nabídka **Regression** umožňuje pomocí procedur *Linear* a *Automatic Linear Modeling* provést odhad koeficientů lineárního regresního modelu s jednou nebo více nezávislými proměnnými metodou nejmenších čtverců a zároveň dovoluje testovat tento model. Procedura *Automatic Linear Modeling* nabízí navíc automatickou přípravu dat a metody pro zvýšení přesnosti a stability modelu.

## Analýzy

### Analyze

Procedura *Curve Estimation* je určena k prokládání křivky s možností výběru z řady různých modelů (lineární, kvadratický, exponenciální atd.).

Pro klasifikaci dat (**Classify**) jsou určeny především tři základní metody seskupovací analýzy (*TwoStep*, *K-means* a *Hierarchical Cluster*), které se užívají k rozdělení souboru (případů nebo proměnných) do skupin tak, aby si prvky uvnitř skupin byly navzájem co nejvíce podobné. Diskriminační analýza (*Discriminant Analysis*) je vhodná k sestrojení klasifikačního modelu v situaci, kdy je třeba na základě několika změřených kvantitativních proměnných odhadnout, ke které skupině určitá jednotka patří. Analýza nejbližších sousedů (*Nearest Neighbor Analysis*) klasifikuje případy na základě jejich podobnosti s jinými případy.

Metody pro redukci dat (**Data Reduction**) jsou zde zastoupeny především faktorovou analýzou (*Factor*), která se zabývá rozbořením struktury vzájemných závislostí proměnných na základě předpokladu, že tyto závislosti jsou způsobeny působením určitého menšího počtu neměřitelných proměnných (faktorů).

Nabídka **Scale** lze využít pro analýzu reliability (*Reliability Analysis*), nebo pomocí mnohorozměrného škálování (*Multidimensional Scaling*) zobrazit analyzované mnohorozměrné objekty v prostoru zvolené dimenze tak, aby se pokud možno zachovaly vzdálenosti (nebo jiné míry podobnosti či nepodobnosti) mezi nimi.

Program nabízí rovněž množství neparametrických testů (**Nonparametric Tests**), z nichž mnohé jsou neparametrickou obdobou některého z dříve uvedených testů (*Two-Independent-Samples Tests*, *Tests for Several Independent Samples*), jiné testují například, zda rozdělení četností proměnné odpovídá určitému teoretickému rozdělení (*Binomial Test*, *Chi-Square Test*, *One-Sample Kolmogorov-Smirnov Test*, *Two-Related-Samples Tests*) nebo zda pořadí výskytu dvou hodnot proměnné je náhodné (*Runs Test*). Pro zadávání těchto testů je k dispozici rovněž novější rozhraní, které automaticky doporučí vhodné testy v závislosti na nastaveném typu proměnných (*One Sample*, *Independent Samples*, *Related Samples*).

Nabídka **Time Series** poskytuje základní grafické nástroje pro analýzu časových řad – graf časových řad (*Sequence Charts*), graf autokorelační funkce (*Autocorrelations*) a vzájemných korelací (*Cross-Correlations*). (Pro hlubší analýzu časových řad je určen speciální modul *IBM SPSS Forecasting*.)

Nástroje pro práci s mnohonásobnými odpověďmi (tj. otázkami, na něž lze uvést více než jednu odpověď) zahrnuje volba **Multiple Response**.

Nabídka **Simulations** je určena pro simulace dat metodou Monte Carlo. Simulace nacházejí uplatnění například při budování predikčních modelů s nejistými vstupy. V takových případech lze pomocí nich odhadnout pravděpodobnost různých situací. Data lze však simulovat i bez využití predikčního modelu tak, že rozložení simulovaných proměnných necháme odhadnout z pozorovaných dat, nebo ho zadáme přímo.

Nabídka **Quality Control** obsahuje grafy určené především pro kontrolu kvality – řídicí diagramy (*Control Charts*) a Paretův graf (*Pareto Charts*).

ROC křivku zobrazíme pomocí nabídky **ROC Curve**.

Volba **Spatial and Temporal Modeling** poskytuje nástroje pro geoprostorové modelování. To je určeno k hledání vztahů v datech obsahujících geoprostorové informace (tj. informace z map). Kromě toho lze do analýz zahrnout také časovou komponentu. Procedura nabízí dvě různé metody: zobecněná geoprostorová asociační pravidla (*Geospatial Association Rules*) a časoprostorové modelování (*Spatio-Temporal Prediction*).

**Slovník anglických pojmů**

<i>ANOVA</i>	<i>analýza rozptylu</i>
<i>chart</i>	<i>graf</i>
<i>compare</i>	<i>srovnání</i>
<i>crosstabs</i>	<i>kontingenční tabulka</i>
<i>curve</i>	<i>křivka</i>
<i>curve estimation</i>	<i>prokládání křivky</i>
<i>define</i>	<i>definovat, zadat</i>
<i>descriptive</i>	<i>popisný</i>
<i>distance</i>	<i>vzdálenost</i>
<i>estimation</i>	<i>odhad</i>
<i>explore</i>	<i>explorovat, prozkoumat</i>
<i>frequencies</i>	<i>četnosti</i>
<i>general</i>	<i>obecný</i>
<i>independent</i>	<i>nezávislý</i>
<i>list cases</i>	<i>vypsát případy</i>
<i>mean</i>	<i>průměr</i>
<i>multiple</i>	<i>mnohonásobný</i>
<i>paired</i>	<i>párový</i>
<i>report</i>	<i>zpráva</i>
<i>response</i>	<i>odpověď</i>
<i>sample</i>	<i>výběr</i>
<i>standard deviation</i>	<i>směrod. odchylka</i>
<i>table</i>	<i>tabulka</i>
<i>value</i>	<i>hodnota</i>

## 1.1 Zprávy (Reports)

### 1.1.1 Shrnutí informací o proměnných a datovém souboru (Codebook)

Procedura poskytuje přehled základních informací o vybraných proměnných, sadách proměnných i celém datovém souboru. Součástí výstupu může být například přehled definovaných vlastností proměnných, základní statistiky proměnných podle typu proměnných, název a umístění datového souboru, počet případů apod.

Na záložce **Variables** přeneseme do pole **Codebook Variables** proměnné, o nichž chceme získat podrobnější informace.

Záložka **Output** je určena ke specifikaci výstupů.

V části **Variable Information** označíme požadované informace o proměnných: pozice (pořadí) v datovém souboru (**Position**), popis proměnné (**Label**), typ (**Type**), formát (**Format**), způsob měření (**Measurement level**), role (**Role**), popisy hodnot (**Value labels**), chybějící hodnoty (**Missing values**), uživatelem definované atributy (**Custom attributes**) a systémové atributy (**Reserved attributes**).

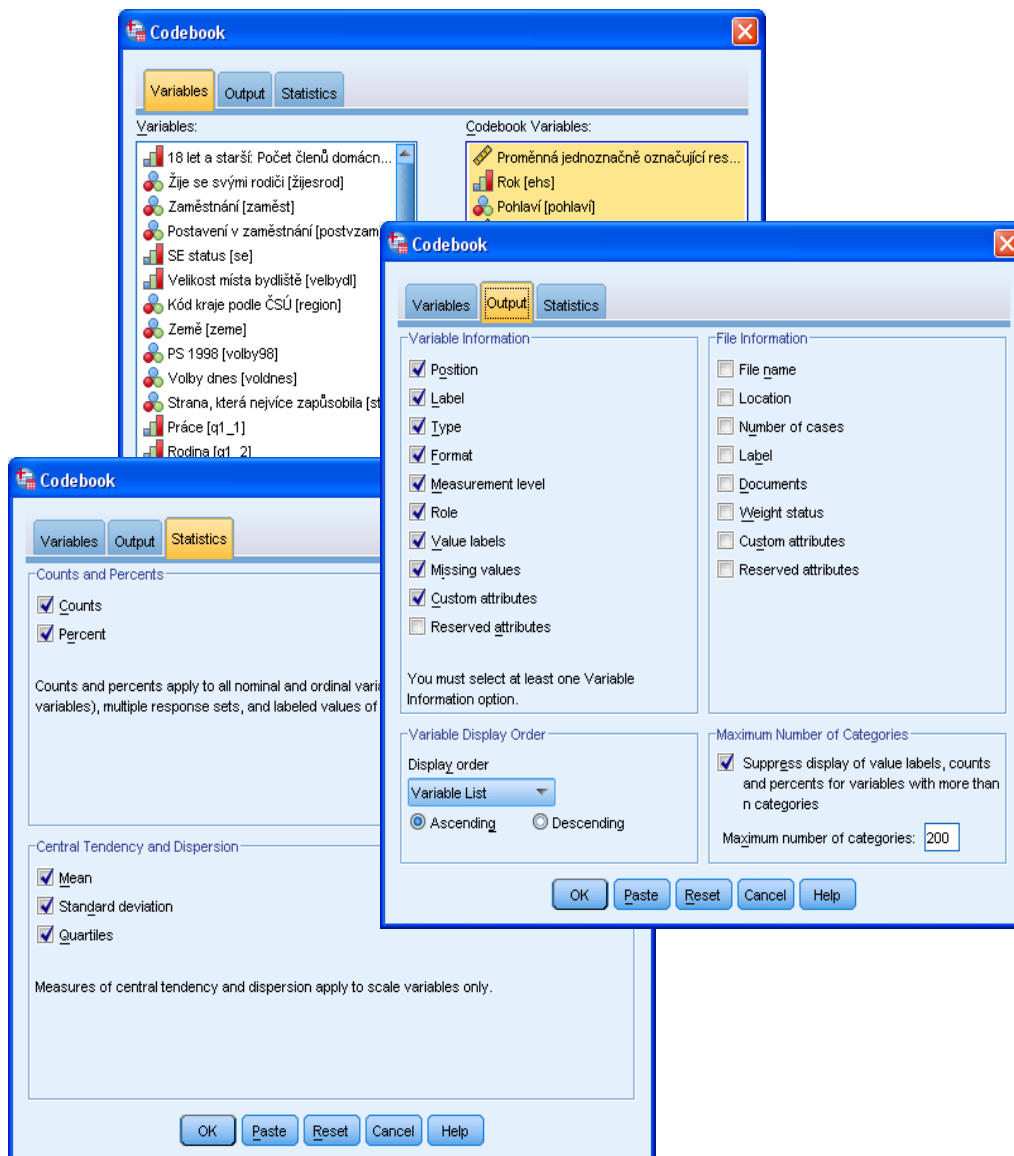
V části **File Information** zvolíme informace o datovém souboru: název (**File name**), umístění (**Location**), počet případů (**Number of cases**), popis souboru (**Label**), komentář k datovému souboru (**Documents**), informace, zda je soubor vážený (**Weight status**), uživatelem definované atributy (**Custom attributes**) a systémové atributy (**Reserved attributes**).

V části **Variable Display Order** určíme způsob řazení informací o proměnných: abecedně podle názvů proměnných (**Alphabetical**), podle pořadí proměnných v datovém souboru (**File**), podle pořadí proměnných na záložce Variables (**Variable List**), nebo podle způsobu měření proměnné (**Measurement level**). Řazení může být rostoucím (**Ascending**) nebo klesajícím (**Descending**) způsobem.

V části **Maximum Number of Categories** rozhodujeme, zda potlačit zobrazování informací o proměnných, které nabývají příliš velkého počtu hodnot (**Suppress display of value labels, counts and percents for variables with more than n categories**) a případně určíme maximální povolený počet kategorií (**Maximum number of categories**).

Na záložce **Statistics** zadáme statistiky, které se budou ve výstupu zobrazovat. Pro kategorizované proměnné (nominální i ordinální) jsou v části **Counts and Percents** k dispozici četnosti (**Counts**) a procenta (**Percent**). Pro číselné lze označit v části **Central Tendency and Dispersion** průměr (**Mean**), směrodatnou odchylku (**Standard deviation**) a kvartily (**Quartiles**).





***Shrnutí informací o proměnných a datovém souboru***

### ***Shrnutí informací o proměnných a datovém souboru***

- Analyze
  - Reports
    - Codebook

#### **1.1.2 Kostky OLAP (OLAP Cubes)**

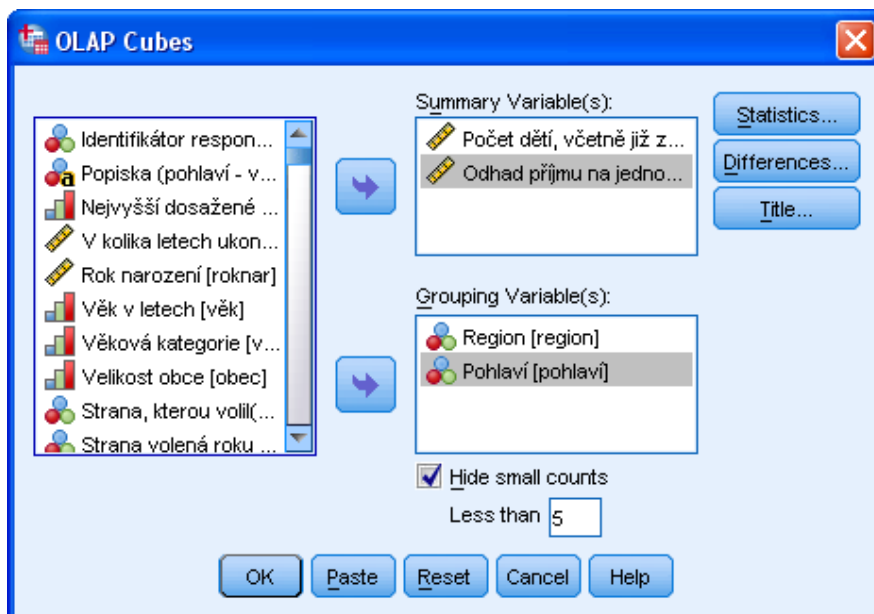
Procedura počítá celkové hodnoty, průměry a další statistiky pro číselné proměnné, které jsou tříděny podle kategorií jedné nebo více jiných proměnných. Výstup se zobrazí jako vícevrstevná

tabulka, kde každá vrstva představuje jednu kategorii podrobného třídění. Pohyb mezi vrstvami umožňují rozvírací seznamy v záhlaví tabulky.

Do pole **Summary Variable(s)** převedeme proměnnou (proměnné), jejichž sumární statistiky počítáme. Třídící proměnnou (proměnné) přeneseme šipkou do pole **Grouping Variable(s)**.

Zaškrtnutím volby **Hide small counts** a zadáním hraniční hodnoty četnosti  $n$  lze dosáhnout toho, že v buňkách s četností menší než zadaná hraniční hodnota  $n$  se zobrazí text „< $n$ “, kde místo  $n$  se použije zvolená hodnota.

Tlačítko **Statistics** slouží pro výběr statistik (průměr, medián, směrodatná odchylka, standardní chyba průměru, suma apod.). Tlačítko **Differences** dovoluje výpočet aritmetických nebo procentních rozdílů mezi proměnnými nebo skupinami. Pro zadání nadpisu a poznámky k tabulce je k dispozici tlačítko **Title**.



**Kostky OLAP**

### **Kostky OLAP**

- Analyze
  - Reports
    - OLAP Cubes

### Slovník anglických pojmů

<i>appear as</i>	<i>zobrazit se jako</i>
<i>arithmetic</i>	<i>početní</i>
<i>between</i>	<i>mezi</i>
<i>caption</i>	<i>legenda, vysvětlivky</i>
<i>case</i>	<i>případ</i>
<i>continue</i>	<i>pokračovat</i>
<i>display</i>	<i>zobrazit</i>
<i>difference</i>	<i>rozdíl</i>
<i>error</i>	<i>chyba</i>
<i>exclude</i>	<i>vynechat</i>
<i>first</i>	<i>první</i>
<i>groups</i>	<i>skupiny</i>
<i>grouping</i>	<i>třídící, seskupující</i>
<i>kurtosis</i>	<i>špičatost</i>
<i>last</i>	<i>poslední</i>
<i>layer</i>	<i>vrstva</i>
<i>listwise</i>	<i>pro celý případ</i>
<i>mean</i>	<i>průměr</i>
<i>median</i>	<i>medián</i>
<i>missing</i>	<i>vynechávaný</i>
<i>number</i>	<i>číslo</i>
<i>only</i>	<i>jen, pouze</i>
<i>percentage</i>	<i>procentní vyjádření</i>
<i>range</i>	<i>rozpětí</i>
<i>report</i>	<i>zpráva</i>
<i>show</i>	<i>ukázat</i>
<i>skewness</i>	<i>šikmost</i>
<i>standard deviation</i>	<i>směrodatná odchylka</i>
<i>statistics</i>	<i>statistiky</i>
<i>sum</i>	<i>suma</i>
<i>summarize</i>	<i>sumarizovat</i>
<i>title</i>	<i>nadpis, titul</i>
<i>valid</i>	<i>platný</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>variance</i>	<i>rozptyl</i>

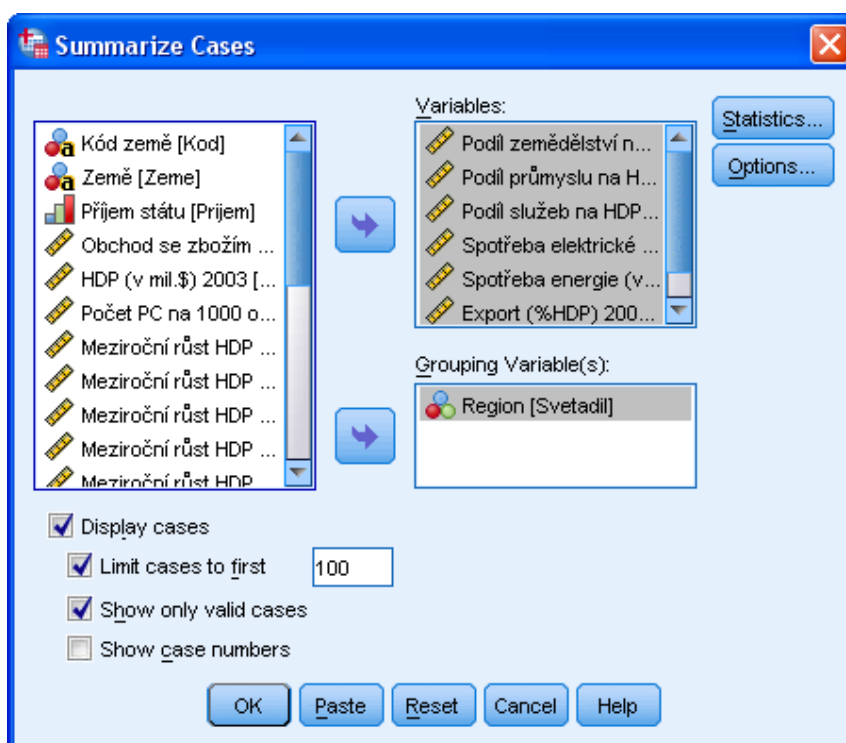
#### 1.1.3 Sumář případů (Case Summaries)

Procedura vytvoří tabulku zvolených statistik (průměr, medián, směrodatná odchylka, minimum, maximum, šikmost, špičatost atd.) pro jednotlivé proměnné nebo pro podrobnější třídění. Zadané statistiky se vztahují ke všem proměnným. V sumáři můžeme rovněž zobrazit jednotlivé případy, jejichž počet lze omezit.

Proměnnou (proměnné), pro kterou se sumář vypíše, převedeme do pole **Variables**. Do pole **Grouping Variable(s)** můžeme přenést třídící proměnnou (proměnné).

Pro zobrazení jednotlivých případů v sumáři zaškrtneme **Display cases**. Dále lze nastavit maximální počet případů zobrazených v sumáři (**Limit cases to first**), zobrazování pouze platných případů (**Show only valid cases**) a zobrazování čísel případů (**Show case numbers**).

Tlačítkem **Statistics** zvolíme statistiky, které přeneseme do pole **Cell Statistics**. Pro zadání nadpisu a vysvětlivky k tabulce je určeno tlačítko **Options**. Rovněž určíme, jakým způsobem zacházet s vynechanými hodnotami.



**Sumář případů**

## Sumář případů

- Analyze
  - Reports
    - Case Summaries

### Slovník anglických pojmů

<i>appear as</i>	zobrazit jako
<i>caption</i>	legenda, vysvětlivky
<i>case</i>	případ
<i>continue</i>	pokračovat
<i>display</i>	zobrazit
<i>error</i>	chyba
<i>exclude</i>	vynechat
<i>first</i>	první
<i>grouping</i>	třídící, seskupující
<i>last</i>	poslední
<i>layer</i>	vrstva
<i>listwise</i>	pro celý případ
<i>mean</i>	průměr
<i>missing</i>	vynechaný
<i>number</i>	číslo
<i>only</i>	jen, pouze
<i>range</i>	rozpětí
<i>report</i>	zpráva
<i>show</i>	ukázat
<i>standard deviation</i>	směrodatná odchylka
<i>statistic</i>	statistika
<i>sum</i>	suma
<i>summarize</i>	sumarizace
<i>title</i>	titul
<i>valid</i>	platný
<i>value</i>	hodnota
<i>variable</i>	proměnná
<i>variance</i>	rozptyl

#### 1.1.4 Sumář v řádcích (Summaries in Rows)

Procedura vytvoří přehledný sumář, který na rozdíl od většiny výstupů programu IBM SPSS Statistics nemá charakter pivotní tabulky, ale textového výstupu. V řádcích mohou být zobrazeny jednotlivé případy i zvolené statistiky, sumář může být podrobněji tříděn. Statistiky lze zadat odlišně pro jednotlivé stupně třídění. Ve sloupcích jsou zobrazeny vybrané proměnné.

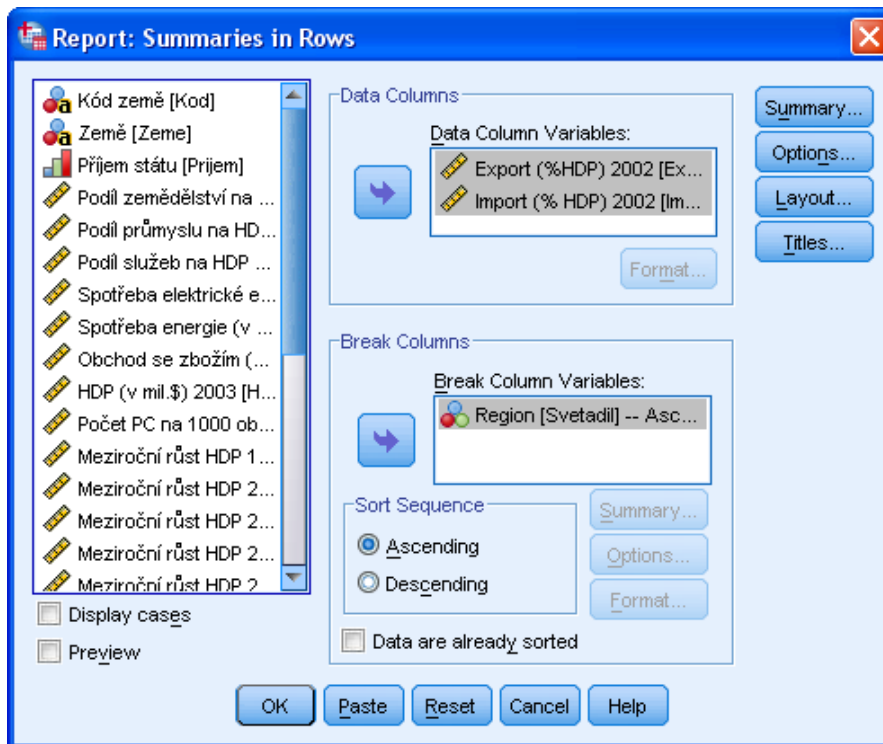
Do okna **Data Column Variables** převedeme proměnné, které budou tvořit sloupce sumáře. Tlačítko **Format** je určeno k úpravě názvů sloupců, zarovnání (doleva, doprostřed, doprava), umístění hodnot uvnitř sloupce, šířky sloupce a obsahu sloupce (hodnota nebo popis hodnoty).

Tlačítko **Summary** slouží k zadávání souhrnných statistik pro celou tabulku. Tlačítkem **Options** volíme metodu práce s vynechávanými hodnotami a číslo první stránky. Tlačítko **Layout** umožňuje vybrat nastavení stránky (první a poslední řádek a sloupec), umístění záhlaví a zápatí strany a umístění titulů sloupců. Tlačítkem **Titles** zadáváme záhlaví a zápatí strany.

Pro vytvoření tříděného sumáře převedeme třídící proměnnou (proměnné) do okna **Break Column Variables**.

Tlačítko **Summary** v oblasti **Break Columns** slouží k volbě statistik pro dané proměnné (stupeň třídění). Každé proměnné lze přiřadit jiné statistiky. Tlačítkem **Options** určíme způsob oddělování

jednotlivých kategorií třídící proměnné (prázdné řádky, každá kategorie na samostatné straně). Tlačítko **Format** je určeno k zadání formátu jednotlivých kategorií třídící proměnné (název sloupce, šířka sloupce, umístění hodnot ve sloupcích atd.).



**Řádkový sumář**

### **Sumář v řádcích**

- Analyze
  - Reports
    - Summaries in Rows

### Slovník anglických pojmů

<i>after</i>	<i>po</i>
<i>alignment</i>	<i>zarovnání</i>
<i>already</i>	<i>již</i>
<i>appear as</i>	<i>zobrazí se jako</i>
<i>ascending</i>	<i>vzestupně</i>
<i>before</i>	<i>před</i>
<i>begin</i>	<i>začít</i>
<i>border</i>	<i>orámování</i>
<i>bottom</i>	<i>dolní okraj</i>
<i>break</i>	<i>dělit, zlom, změna</i>
<i>case</i>	<i>případ</i>
<i>center</i>	<i>střed</i>
<i>column</i>	<i>sloupec</i>
<i>continue</i>	<i>pokračovat</i>
<i>descending</i>	<i>sestupný</i>
<i>end</i>	<i>skončit, konec</i>
<i>exclude</i>	<i>vyločit</i>
<i>final</i>	<i>finální, konečný</i>
<i>footer</i>	<i>zápatí</i>
<i>format</i>	<i>formát</i>
<i>grand total</i>	<i>celkový souhrn</i>
<i>high</i>	<i>vysoký, horní</i>
<i>layout</i>	<i>rozvržení</i>
<i>left</i>	<i>nalevo</i>
<i>length</i>	<i>délka</i>
<i>line</i>	<i>řádek</i>
<i>listwise</i>	<i>pro celý případ</i>
<i>low</i>	<i>nízký, spodní</i>
<i>mean</i>	<i>průměr</i>
<i>next</i>	<i>další</i>
<i>number</i>	<i>číslo</i>
<i>page</i>	<i>strana</i>
<i>point</i>	<i>bod</i>
<i>previous</i>	<i>předchozí</i>
<i>report</i>	<i>zpráva</i>
<i>right</i>	<i>napravo</i>
<i>row</i>	<i>řádek</i>
<i>same</i>	<i>tentýž</i>
<i>skewness</i>	<i>šikmost</i>
<i>sort</i>	<i>řadit</i>
<i>standard deviation</i>	<i>směrod.odchylka</i>
<i>summary</i>	<i>sumární statistiky</i>
<i>title</i>	<i>titul</i>
<i>top</i>	<i>horní okraj</i>
<i>underscore</i>	<i>podtrhnout</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>vertically</i>	<i>vertikálně, svisle</i>
<i>width</i>	<i>šířka</i>

### 1.1.5 Sumář ve sloupcích (Summaries in Columns)

Procedura vytvoří sumář, kde ve sloupcích budou zobrazeny zvolené statistiky vybraných proměnných, v řádcích kategorie třídění. Výstup je v textovém formátu a stejně jako u příkazu **Summaries in Rows** nemá charakter pivotní tabulky.

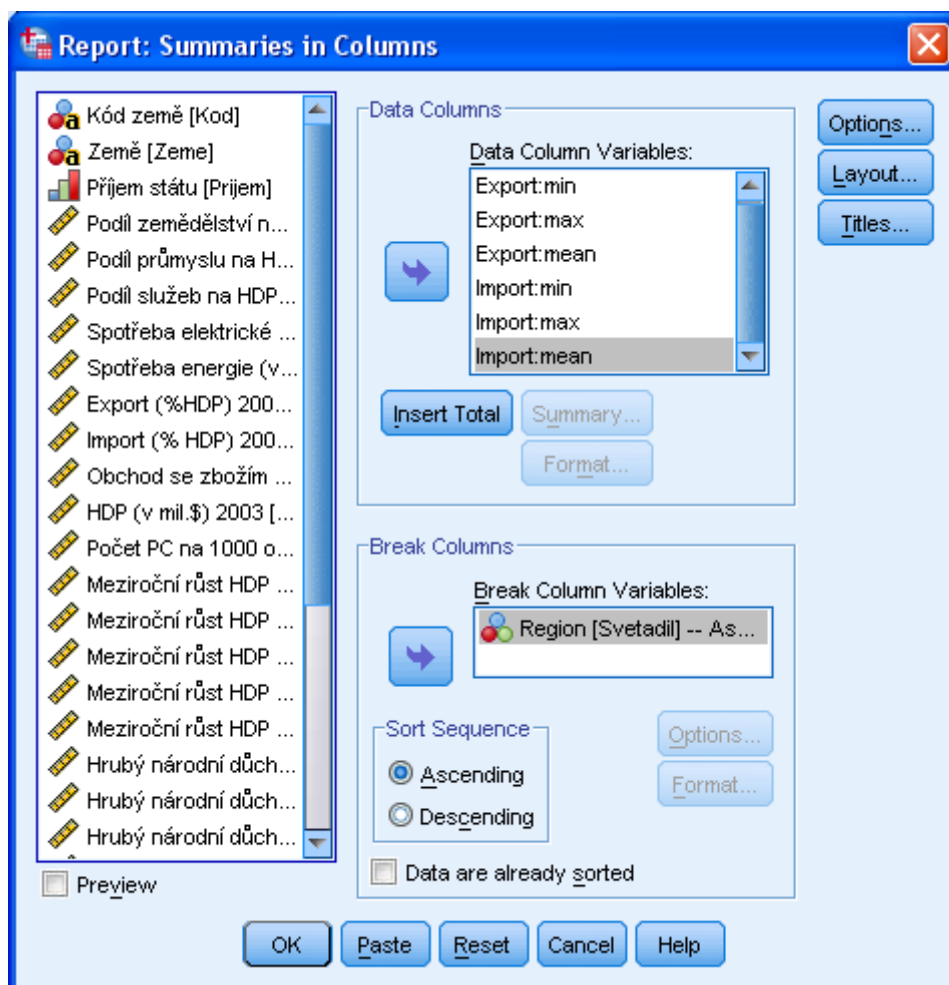
Proměnné, které budou tvořit sloupce sumáře, převedeme do okna **Data Column Variables**. Tlačítkem **Summary** přiřadíme označené proměnné statistiku (u každé proměnné může být jiná statistika, jedna proměnná se může vyskytovat s více různými statistikami).

Tlačítkem **Insert Total** lze do tabulky vložit sloupec vyjadřující úhrn alespoň dvou jiných sloupců. Vlastnosti sloupce blíže specifikujeme tlačítkem **Summary**. Tlačítko **Format** je určeno k zadávání formátu jednotlivých kategorií třídící proměnné (název sloupce, šířka sloupce, umístění hodnot ve sloupcích).

Tlačítko **Options** dovoluje zobrazit v tabulce celkové úhrny, zvolit metodu práce s vynechávanými hodnotami a nastavit číslo první stránky. Tlačítko **Layout** umožňuje nastavit některé parametry stránky, umístění záhlaví a zápatí strany a nadpisů sloupců. Tlačítkem **Titles** definujeme nadpis, záhlaví a zápatí stránky.

Pro vytvoření tříděného sumáře převedeme do okna **Break Column Variables** proměnnou (proměnné), která bude tvořit řádky sumáře. Tlačítko **Options** umožňuje zadat výpočet úhrnů pro jednotlivé stupně třídění (proměnné) zaškrtnutím políčka **Display subtotal**. Tlačítkem **Format** můžeme upravit vlastnosti sloupců popisujících kategorie třídění (název sloupce, šířka sloupce, umístění hodnot ve sloupcích).





Sloupcový sumář

### Sumář ve sloupcích

- Analyze
  - Reports
    - Summaries in Columns

### Slovník anglických pojmů

<i>alignment</i>	zarovnání
<i>already</i>	již
<i>appear as</i>	jsou zobrazeny jako
<i>ascending</i>	vzestupný
<i>before</i>	před
<i>begin</i>	začít
<i>border</i>	orámování
<i>bottom</i>	dolní okraj
<i>break</i>	dělit, zlom, změna
<i>case</i>	případ
<i>center</i>	střed
<i>column</i>	sloupec
<i>continue</i>	pokračovat
<i>descending</i>	sestupný
<i>end</i>	skončit, konec
<i>exclude</i>	vyločit
<i>final</i>	finální, konečný
<i>footer</i>	patka stránky
<i>format</i>	formát
<i>grand total</i>	celkový souhrn
<i>high</i>	vysoký, horní
<i>insert</i>	vložit
<i>layout</i>	rozvržení
<i>left</i>	nalevo
<i>length</i>	délka
<i>line</i>	řádek
<i>listwise</i>	celý seznam
<i>low</i>	nízký, spodní
<i>number</i>	číslo
<i>options</i>	volby
<i>page</i>	strana
<i>point</i>	bod
<i>previous</i>	předchozí
<i>report</i>	zpráva
<i>row</i>	řádek
<i>skewness</i>	zešikmení
<i>sort</i>	řadit
<i>std.deviation</i>	směrod.odchylka
<i>summary</i>	sumární statistiky
<i>title</i>	titul
<i>top</i>	horní okraj
<i>underscore</i>	podtrhnout
<i>value</i>	hodnota
<i>variable</i>	proměnná
<i>vertically</i>	vertikálně, svisle
<i>width</i>	šířka

## 1.2 Popisné statistiky (Descriptive Statistics)

### 1.2.1 Četnosti (Frequencies)

Procedura vytvoří tabulky četností zvolených proměnných, které doplní vybranými popisnými statistikami a grafy. Nabídka je určena především pro analýzu kategorizovaných dat.

Do okna **Variable(s)** zadáme sledované proměnné.

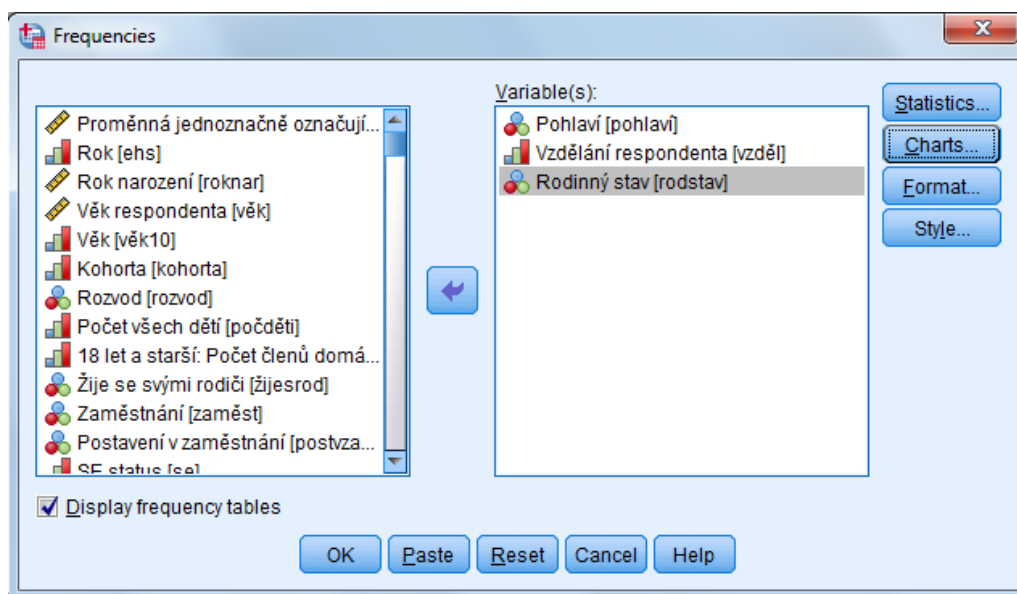
Políčko **Display frequency tables** řídí zobrazování tabulek četností.

Tlačítkem **Statistics** zadáme výpočet doplňujících statistik. Dialog nabízí výběr z těchto oblastí: percentily (**Percentile Values**), charakteristiky variability (**Dispersion**), odhady střední hodnoty (**Central Tendency**) a statistiky tvaru rozložení (**Distribution**).

Tlačítkem **Charts** nakreslíme graf. Typ grafu (**Chart Type**) vybereme z nabídky: sloupcový (**Bar charts**), koláčový (**Pie charts**) nebo histogram (**Histograms**). U histogramu dále určíme, zda proložit Gaussovu křivku (**With normal curve**), u koláčových a sloupcových grafů je třeba specifikovat, zda se mají zobrazovat četnosti či procenta (**Chart Values**).

Tlačítkem **Format** nastavíme formát výstupů. Sekce **Order by** určuje řazení četností: vzestupně, sestupně, podle hodnot nebo četností. Při analýze více proměnných najednou lze v sekci **Multiple Variables** uspořádat výstup po jednotlivých proměnných. Dialog nabízí také možnost potlačit tabulky s větším počtem kategorií (**Suppress tables with more than n categories**).

Tlačítko **Style** je určeno k tzv. podmíněnému formátování, které dovoluje nastavit formát pivotních tabulek již při jejich vytváření (podrobněji viz *Příloha 1 Podmíněné formátování tabulek*).



**Výpočet četností**

## Četnosti

- Analyze
  - Descriptive Statistics
    - Frequencies

### Slovník anglických pojmů

<i>add</i> .....	<i>přidat</i>
<i>ascending</i> .....	<i>vzestupně</i>
<i>bar chart</i> .....	<i>sloupcový graf</i>
<i>categories</i> .....	<i>kategorie</i>
<i>central</i> .....	<i>střední</i>
<i>compare</i> .....	<i>srovnávat</i>
<i>continue</i> .....	<i>pokračovat</i>
<i>counts</i> .....	<i>četnosti, počty</i>
<i>curve</i> .....	<i>křivka</i>
<i>cut points</i> .....	<i>dělicí body</i>
<i>descending</i> .....	<i>sestupně</i>
<i>dispersion</i> .....	<i>rozptýlení</i>
<i>display</i> .....	<i>zobrazit</i>
<i>distribution</i> .....	<i>distribuce, dělení</i>
<i>equal groups</i> .....	<i>shodné skupiny</i>
<i>format</i> .....	<i>formát tabulky</i>
<i>frequencies</i> .....	<i>četnosti</i>
<i>histogram</i> .....	<i>histogram</i>
<i>change</i> .....	<i>změnit</i>
<i>chart</i> .....	<i>graf</i>
<i>kurtosis</i> .....	<i>špičatost</i>
<i>label</i> .....	<i>popis, označení</i>
<i>midpoint</i> .....	<i>střední bod</i>
<i>mode</i> .....	<i>modus</i>
<i>multiple</i> .....	<i>mnohonásobný</i>
<i>normal curve</i> .....	<i>Gaussova křivka</i>
<i>order</i> .....	<i>řád, seřazení</i>
<i>organize</i> .....	<i>uspořádat</i>
<i>page</i> .....	<i>stránka</i>
<i>percentage</i> .....	<i>procentní vyjádření</i>
<i>percentile</i> .....	<i>percentily</i>
<i>quartile</i> .....	<i>kvartil</i>
<i>range</i> .....	<i>rozpětí</i>
<i>S.E. Mean</i> .....	<i>stand. chyba průměru</i>
<i>skewness</i> .....	<i>šikmost</i>
<i>stadandard deviation</i> .....	<i>směrodatná odchylka</i>
<i>summarize</i> .....	<i>sumarizovat</i>
<i>suppress</i> .....	<i>potlačit</i>
<i>tendency</i> .....	<i>tendence</i>
<i>type</i> .....	<i>typ</i>
<i>value</i> .....	<i>hodnota</i>
<i>variance</i> .....	<i>rozptyl</i>

## 1.2.2 Popisné statistiky (Descriptives)

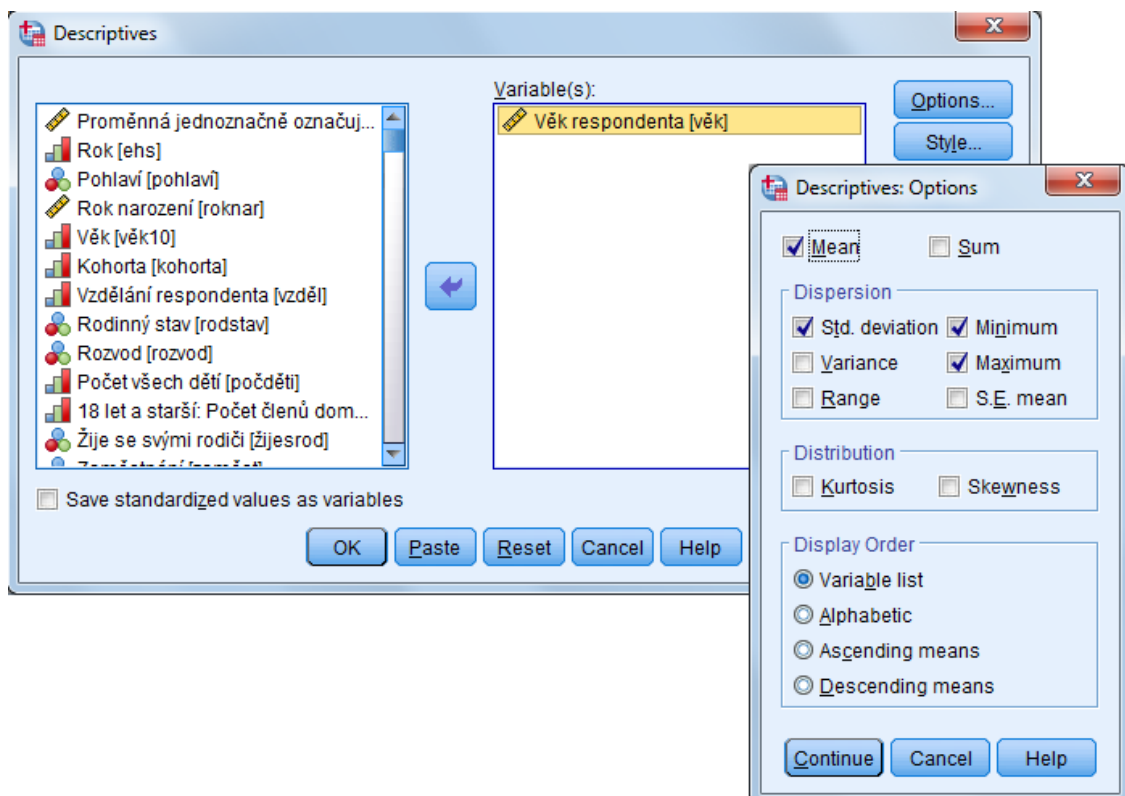
Procedura zobrazí přehlednou tabulku, kde v řádcích budou zvolené proměnné, ve sloupcích vybrané popisné statistiky. V dialogovém okně lze rovněž označit výpočet standardizovaných hodnot (z-skórů), které se zobrazí v datové matici jako nová proměnná. Procedura je určena především pro číselné proměnné.

Analyzované proměnné označíme a převedeme šipkou do pole **Variable(s)**.

Zaškrtnutím políčka **Save standardized values as variables** zadáme výpočet standardizovaných hodnot, které se vloží jako nové proměnné do datové matice. Tento postup je vhodný především před použitím některých metod, pokud jsou původní proměnné zjišťovány v odlišných (nesouměřitelných) jednotkách. Pomocí transformace na z-skóry je umístíme na společnou škálu, aby mohly být lépe srovnávány.

Tlačítko **Options** slouží pro výběr statistik: průměr (**Mean**), celkový součet (**Sum**), směrodatná odchylka (**Std. deviation**), rozptyl (**Variance**), rozpětí (**Range**), minimum (**Minimum**), maximum (**Maximum**), standardní chyba průměru (**S. E. mean**), špičatost (**Kurtosis**) a šikmost (**Skewness**). Sekce **Display Order** je určena k volbě pořadí, v jakém budou proměnné zobrazeny ve výstupní tabulce (podle seznamu, abecedně nebo podle průměrné hodnoty).

Tlačítko **Style** je určeno k tzv. podmíněnému formátování, které dovoluje nastavit formát pivotních tabulek již při jejich vytváření (podrobněji viz *Příloha 1 Podmíněné formátování tabulek*).



Popisné statistiky

## Popis číselných dat

- Analyze
  - Descriptive Statistics
    - Descriptives

### Slovník anglických pojmů

<i>alphabetic</i>	<i>abecedně</i>
<i>as</i>	<i>jako</i>
<i>ascending</i>	<i>vzestupně</i>
<i>both</i>	<i>obojí</i>
<i>confidence interval</i>	<i>interval spolehlivosti</i>
<i>continue</i>	<i>pokračovat</i>
<i>descending</i>	<i>sestupně</i>
<i>descriptive</i>	<i>popisný</i>
<i>dispersion</i>	<i>rozptýlení</i>
<i>display</i>	<i>zobrazit</i>
<i>estimation</i>	<i>odhad</i>
<i>exclude</i>	<i>vynechat</i>
<i>label</i>	<i>popis</i>
<i>level</i>	<i>úroveň</i>
<i>listwise</i>	<i>pro celý případ</i>
<i>mean</i>	<i>průměr</i>
<i>missing</i>	<i>chybějící, vynechaný</i>
<i>none</i>	<i>žádný</i>
<i>options</i>	<i>volby</i>
<i>order</i>	<i>řád, řazení</i>
<i>pairwise</i>	<i>po dvojicích</i>
<i>plot</i>	<i>graf</i>
<i>power</i>	<i>mocnina</i>
<i>range</i>	<i>rozpětí</i>
<i>report</i>	<i>vypsát</i>
<i>S.E. Mean</i>	<i>stand.chyba průměru</i>
<i>save</i>	<i>uložit</i>
<i>skewness</i>	<i>šikmost</i>
<i>spread</i>	<i>rozptýlení</i>
<i>standardized</i>	<i>standardizovaný</i>
<i>standard deviation</i>	<i>směrodatná odchylka</i>
<i>sum</i>	<i>suma</i>
<i>table</i>	<i>tabulka</i>
<i>together</i>	<i>dohromady, společně</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>variable list</i>	<i>seznam proměnných</i>
<i>variance</i>	<i>rozptyl</i>
<i>with</i>	<i>s</i>

### 1.2.3 Explorace (Explore)

Procedura je určena k velmi podrobné analýze číselných proměnných. Vypočítá sumární statistiky a nakreslí požadované typy grafů pro celý soubor nebo pro zvolené skupiny případů.

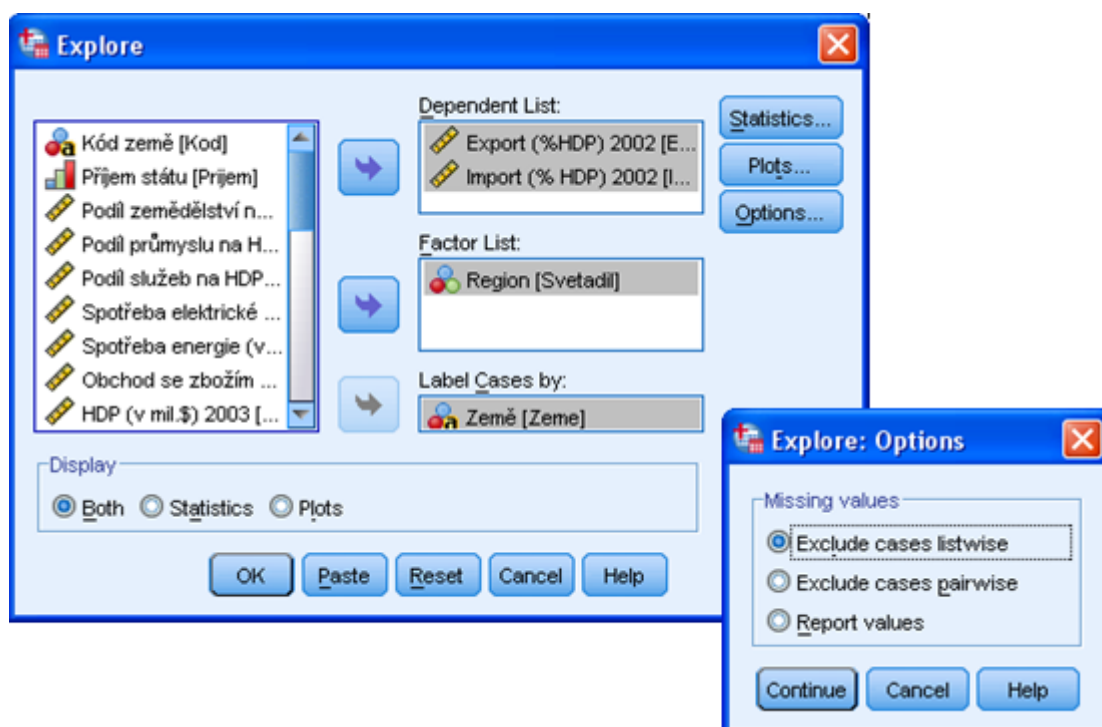
Analýzované proměnné převedeme do okna **Dependent List**. Skupiny definujeme pomocí kategorizované proměnné (proměnných), které převedeme do okna **Factor List**. Analýza se potom provede samostatně pro každou skupinu případů definovanou stejnou úrovní tohoto faktoru. Pokud mají být v grafech popsány jednotlivé případy (názvy zemí, měst, regionů atd.), převedeme příslušnou popisnou proměnnou do okna **Label Cases by**. Dále specifikujeme, zda zobrazit statistiky, grafy nebo oboje (**Statistics, Plots, Both**).

Pro zadání statistik slouží tlačítko **Statistics**. Označit můžeme: popisné statistiky (**Descriptives**) se zvoleným intervalem spolehlivosti pro průměr, odhady střední hodnoty (**M-estimators**), výpis pěti nejnižších a nejvyšších pozorování (**Outliers**) a percentily (**Percentiles**).

Tlačítko **Plots** je určeno k zadávání grafů. Lze vybrat: krabicové grafy (**Boxplots**), grafy rozložení (**Descriptive**), grafy a testy pro ověření shody s normálním rozložením (**Normality plots with tests**) a grafy závislosti polohy a rozptylu včetně Levenova testu (**Spread vs. Level with Levene Test**).

Tlačítkem **Options** zadáme způsob práce s chybějícími hodnotami:

- **Exclude cases listwise:** pokud obsahuje některá ze zadaných proměnných chybějící hodnotu, je případ vynechán také u ostatních proměnných,
- **Exclude cases pairwise:** chybějící hodnota se vynechá, platné hodnoty se u tohoto případu pro výpočet použijí,
- **Report values:** chybějící hodnoty proměnné v poli *Factor List* jsou považovány za samostatnou kategorii.



Explorace

## Explorace (Explore)

- Analyze
  - Descriptive Statistics
    - Explore

### Slovník anglických pojmů

<i>both</i>	<i>obojí</i>
<i>boxplot</i>	<i>boxplot</i>
<i>confid. interval</i>	<i>interval spolehlivosti</i>
<i>continue</i>	<i>pokračovat</i>
<i>dependent</i>	<i>závislá</i>
<i>descriptive</i>	<i>popisný</i>
<i>display</i>	<i>zobrazit</i>
<i>estimation</i>	<i>odhad</i>
<i>exclude</i>	<i>vynechat</i>
<i>explore</i>	<i>explorace</i>
<i>factor</i>	<i>faktor, nezávislá proměnná</i>
<i>labels</i>	<i>popisy</i>
<i>level</i>	<i>úroveň</i>
<i>listwise</i>	<i>pro celý případ</i>
<i>mean</i>	<i>průměr</i>
<i>missing</i>	<i>chybějící, vynechaný</i>
<i>none</i>	<i>žádný</i>
<i>options</i>	<i>volby</i>
<i>order</i>	<i>řád, řazení</i>
<i>outliers</i>	<i>vnější pozorování</i>
<i>pairwise</i>	<i>po dvojicích</i>
<i>percentiles</i>	<i>percentily</i>
<i>plot</i>	<i>graf</i>
<i>power</i>	<i>mocnina</i>
<i>range</i>	<i>rozpětí</i>
<i>report</i>	<i>vypsat</i>
<i>save</i>	<i>uložit</i>
<i>skewness</i>	<i>šikmost</i>
<i>spread</i>	<i>rozptýlení</i>
<i>stem-and-leaf</i>	<i>graf „lodyha s listy“</i>
<i>table</i>	<i>tabulka</i>
<i>together</i>	<i>dohromady, společně</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>variable list</i>	<i>seznam proměnných</i>

### 1.2.4 Kontingenční tabulky (Crosstabs)

Pomocí této procedury lze získat tabulky třídění druhého a vyšších stupňů a spočítat příslušné testy a míry, které se vztahují k dvourozměrným tabulkám.



Pokud se jedná například o tabulku třetího stupně, počítají se všechny míry a testy vždy postupně k tabulkám tak, jak jsou řazeny podle hodnoty třetí (kontrolní) proměnné. Pokud by například představovalo kontrolní proměnnou „pohlaví“, vznikla by jedna tabulka pro muže a druhá pro ženy a testy by se spočítaly pro každé pohlaví samostatně.

Nejprve v poli **Row(s)** zadáme proměnnou (proměnné), které budou tvořit řádky tabulky. Proměnnou (proměnné), které budou tvořit sloupce, umístíme do okna **Column(s)**. Pro každou kombinaci řádkové a sloupcové proměnné se potom vytvoří samostatná tabulka. Při zadávání třídění třetího a vyšších stupňů přeneseme další proměnnou (proměnné) do okna **Layer 1 of 1**. Tlačítkem **Next** lze pro každou proměnnou z tohoto okna přidat další dimenzi tabulky.

Pro vytvoření sloupcového grafu na základě tabulky zaškrtneme **Display clustered bar charts**. Chceme-li zobrazit pouze statistiky a grafy, ne však samotnou kontingenční tabulku, zaškrtneme políčko **Suppress tables**. Volbou **Display layer variables in table layers** se proměnné zadané v poli **Layer** zobrazí ve vrstvách výsledné tabulky. Není-li tato volba vybrána, pak jsou tyto proměnné umístěné v řádkové dimenzi výstupní tabulky.

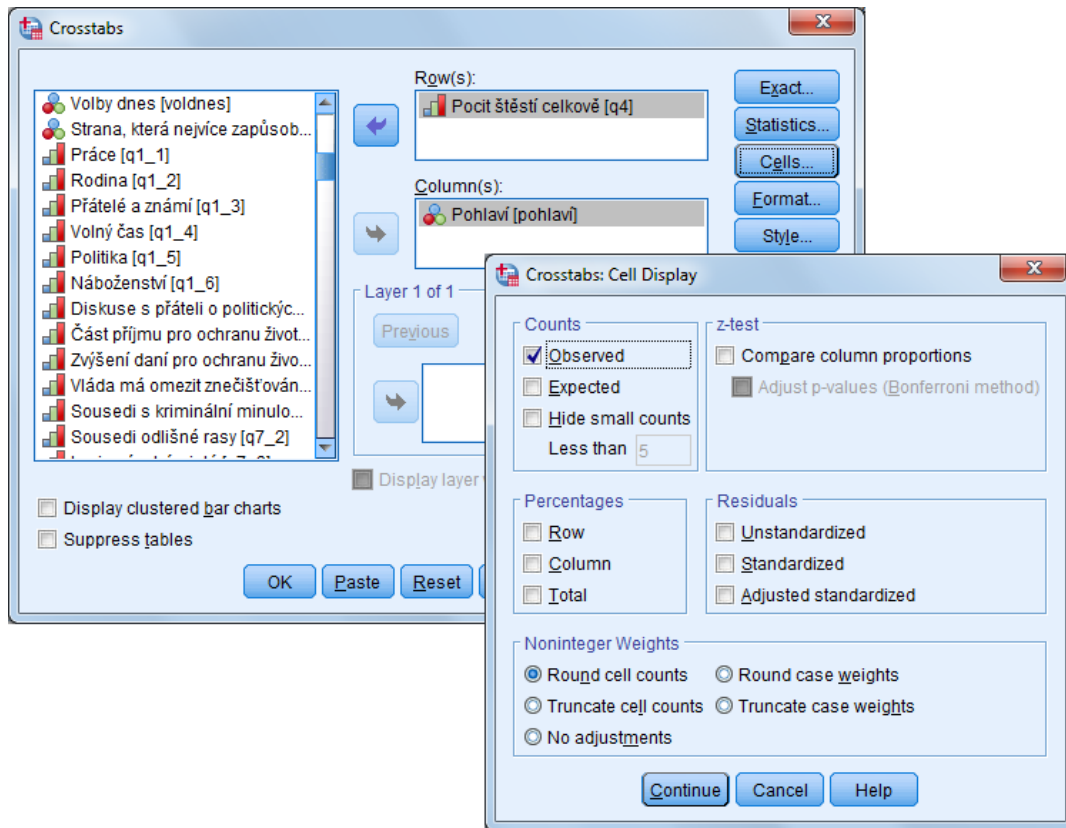
Zadání statistik pro tabulku se provádí tlačítkem **Statistics**. Lze volit:

- pro tabulku: chí-kvadrát, korelační koeficient,
- pro nominální data: kontingenční koeficient,  $F$  a Cramerovo  $V$ ,  $\Lambda$ , koeficient neurčitosti,
- pro pořadová (ordinální) data: Gama, Sommerovo  $D$ , Kendallovo  $\tau$ -B a  $\tau$ -C,
- další testy: Kapa, Risk, McNemarův,  $\epsilon$ , Cochranovy a Mantel-Haenszelovy statistiky.

Pro zadání statistik jednotlivých polí tabulky je určeno tlačítko **Cells**. Lze volit: pozorované nebo očekávané absolutní četnosti (**Counts**), řádková, sloupcová nebo celková procenta (**Percentages**), nestandardizovaná, standardizovaná nebo adjustovaná rezidua (**Residuals**). Volbou **Hide small counts** a zadáním hraniční hodnoty do pole **Less than** se v polích s četností menší než zadaná hraniční hodnota zobrazí text „< $n$ “, kde místo  $n$  je uvedena zvolená hodnota. Pomocí volby **Compare column proportions** lze provést z-test shody sloupcových procent pro každou dvojici kategorií sloupcové proměnné. Tento test je možné provést s Bonferroniho korekcí pro mnohonásobné testování zvolením **Adjust p-values (Bonferroni method)**. V části **Noninteger Weights** určíme způsob práce s neceločíselnými vahami.

Tlačítkem **Format** lze nastavit vzestupné nebo sestupné řazení řádků tabulky (**Row Order**).

Tlačítko **Style** je určeno k tzv. podmíněnému formátování, které dovoluje nastavit formát pivotních tabulek již při jejich vytváření (podrobněji viz *Příloha 1 Podmíněné formátování tabulek*).



### Zadávání kontingenční tabulky

## Kontingenční tabulky

- Analyze
  - Descriptive Statistics
    - Crosstabs

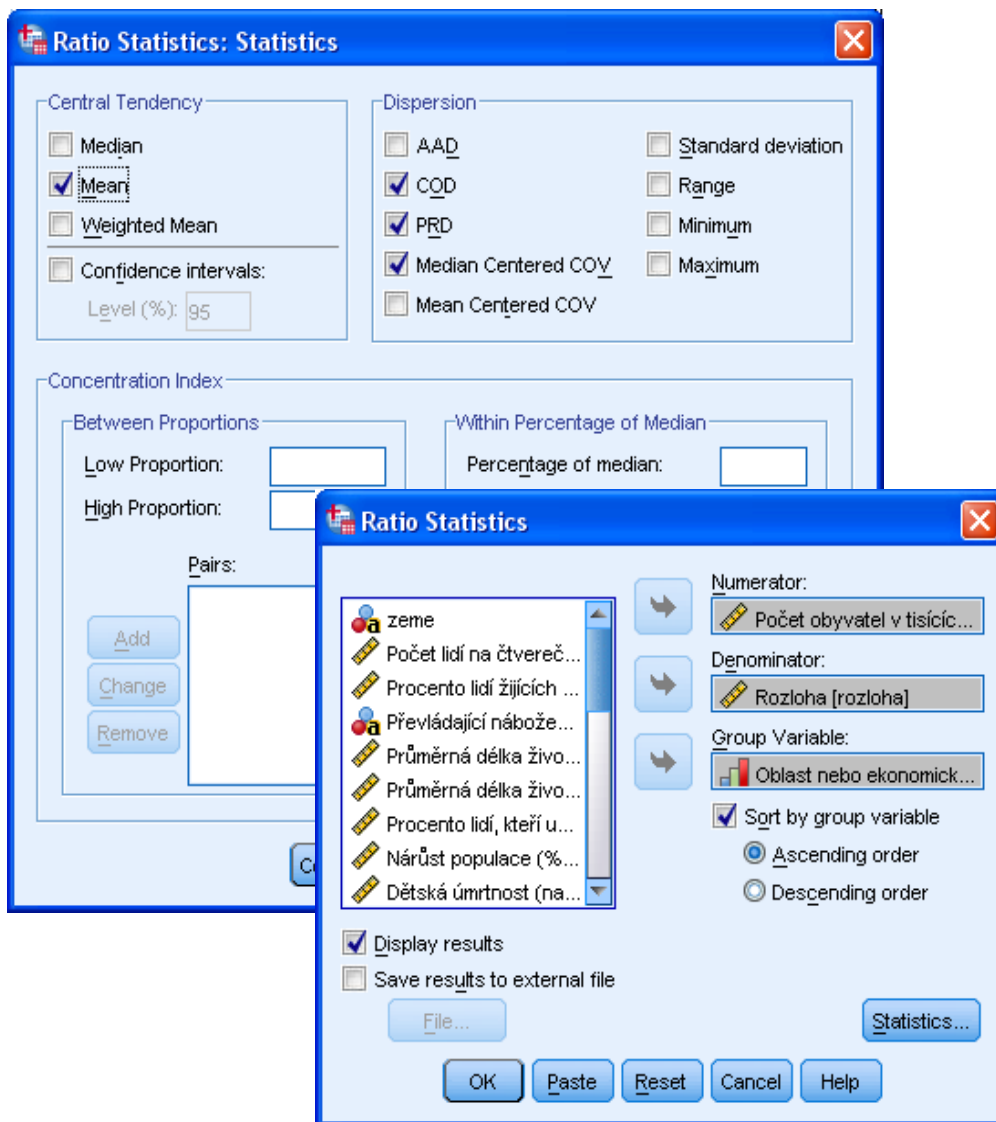
### Slovník anglických pojmů

<i>adj. (adjusted)</i> .....	<i>adjustovaný, upravený</i>
<i>ascending</i> .....	<i>vzestupný</i>
<i>column</i> .....	<i>sloupec</i>
<i>continue</i> .....	<i>pokračovat</i>
<i>correlation</i> .....	<i>korelace</i>
<i>count</i> .....	<i>počet, četnost</i>
<i>crosstab</i> .....	<i>kontingenční tabulka</i>
<i>descending</i> .....	<i>sestupný</i>
<i>expected</i> .....	<i>očekávaný</i>
<i>Chi-square</i> .....	<i>chí-kvadrát</i>
<i>interval</i> .....	<i>intervalový</i>
<i>label</i> .....	<i>popis</i>
<i>layer</i> .....	<i>vrstva</i>
<i>next</i> .....	<i>další</i>
<i>nominal</i> .....	<i>nominální</i>
<i>none</i> .....	<i>žádný</i>
<i>observed</i> .....	<i>pozorovaný</i>
<i>only</i> .....	<i>jen</i>
<i>order</i> .....	<i>řád, pořadí</i>
<i>ordinal</i> .....	<i>pořadový, ordinální</i>
<i>percentage</i> .....	<i>procentní vyjádření</i>
<i>residual</i> .....	<i>zbytek, reziduum</i>
<i>row</i> .....	<i>řádek</i>
<i>standardized</i> .....	<i>standardizovaný</i>
<i>suppress</i> .....	<i>potlačit</i>
<i>table</i> .....	<i>tabulka</i>
<i>total</i> .....	<i>celkem, celkový</i>
<i>unstandardized</i> .....	<i>nestandardizovaný</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>

#### 1.2.5 Poměrové statistiky (Ratio)

Pomocí této procedury lze získat souhrnné statistiky pro poměr dvou posuzovaných veličin včetně koeficientu rozptylu, variačního koeficientu, difference hodnotové příbuznosti a průměrné absolutní odchylky.

Do pole **Numerator** převedeme proměnnou reprezentující čitatel, do pole **Denominator** proměnnou představující jmenovatel. Pokud je třeba výstup třídit přes některou další proměnnou, zadáme ji do pole **Group Variable**. Pro definování statistik slouží tlačítko **Statistics**. Získané výsledky lze zobrazit do výstupového okna (**Display results**) nebo uložit do souboru (**Save results to external file**).



**Poměrové statistiky**

### **Poměrové statistiky**

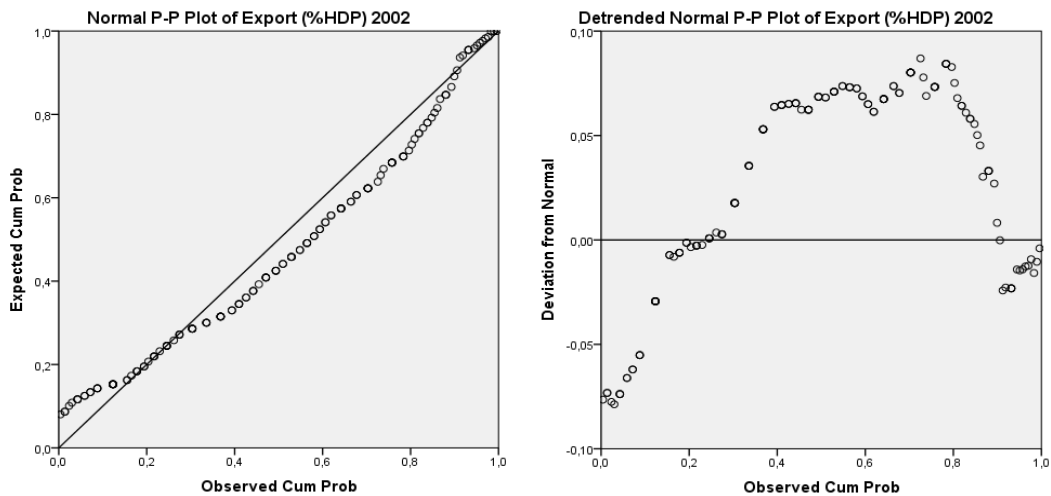
- Analyze
  - Descriptive Statistics
    - Ratio

### Slovník anglických pojmů

<i>appear as</i>	zobrazuje se jako
<i>arithmetic</i>	početní
<i>between</i>	mezi
<i>caption</i>	legenda, vysvětlivky
<i>case</i>	případ
<i>continue</i>	pokračovat
<i>denominator</i>	jmenovatel
<i>display</i>	zobrazit
<i>differences</i>	rozdíly
<i>error</i>	chyba
<i>exclude</i>	vynechat
<i>first</i>	první
<i>group</i>	skupina
<i>grouping</i>	třídící, seskupující
<i>kurtosis</i>	špičatost
<i>last</i>	poslední
<i>layer</i>	vrstva
<i>listwise</i>	pro celý případ
<i>mean</i>	průměr
<i>median</i>	medián
<i>missing</i>	vynechávaný
<i>number</i>	číslo
<i>numerator</i>	čitatel
<i>only</i>	jen, pouze
<i>percentage</i>	procentní
<i>range</i>	rozpětí
<i>report</i>	zpráva
<i>show</i>	ukázat
<i>skewness</i>	šikmost
<i>st. deviation</i>	směrodatná odchylka
<i>statistic</i>	statistika
<i>sum</i>	suma
<i>summarize</i>	sumarizace
<i>title</i>	titul
<i>valid</i>	platný
<i>value</i>	hodnota
<i>variable</i>	proměnná
<i>variance</i>	rozptyl

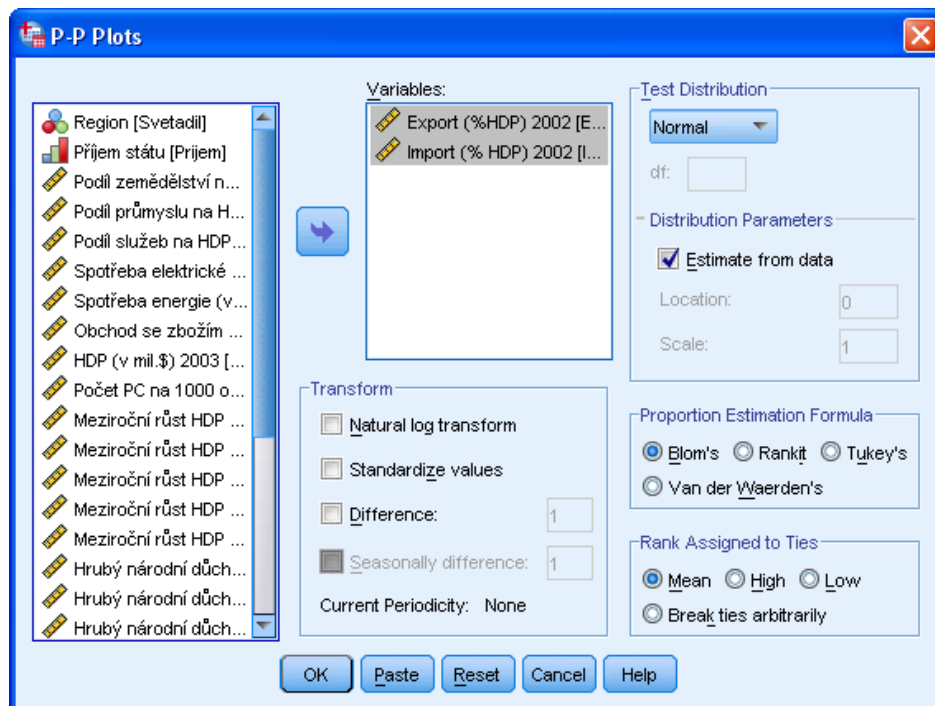
### 1.2.6 Graf kumulativních proporcí (P-P Plots)

Tento typ grafu se obvykle užívá jako doplněk dalších analýz při testování, zda má daná proměnná určitý typ rozdělení. V programu IBM SPSS Statistics je k dispozici srovnání s deseti různými typy rozdělení. Graf je založen na vynášení naměřených kumulativních proporcí proti hodnotám teoretického rozdělení. Při platnosti hypotézy by se zobrazené body měly nacházet v blízkosti diagonály. **Detrended Normal P-P Plot** vyjadřuje diference mezi body grafu kumulativních proporcí a diagonálou.



**P-P Plot**

Do pole **Variables** přeneseme analyzované proměnné. V poli **Test Distribution** zvolíme porovnávané teoretické rozdělení. Místo původních hodnot můžeme pracovat rovněž s transformovanými hodnotami (**Transform**). K dispozici jsou transformace přirozeným logaritmem, z-skóry, diference a sezónní diference pro časové řady s definovanou periodou. Podrobnosti výpočtu nastavíme v polích **Proportion Estimation Formula** a **Rank Assigned to Ties**.



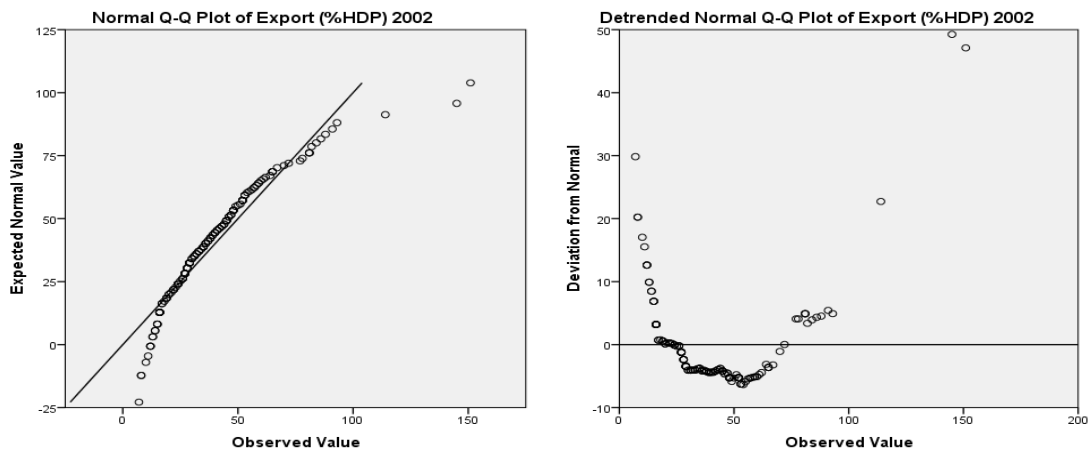
**Graf kumulativních proporcí**

### Graf kumulativních proporcí

- Analyze
  - Descriptive Statistics
    - P-P Plots

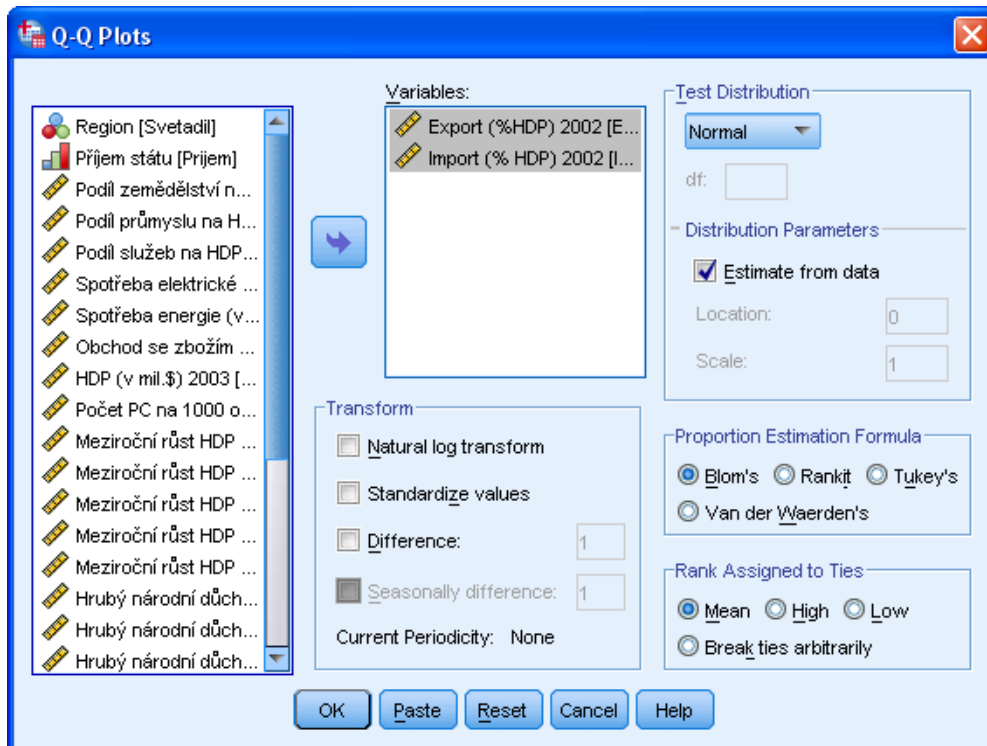
### 1.2.7 Kvantilový graf (Q-Q Plots)

Kvantilový graf je obdobou grafu kumulativních četností (viz výše), je však založen na kvantilech.



**Q-Q Plot**

Do pole **Variables** přeneseme analyzované proměnné. V poli **Test Distribution** zvolíme porovnávané teoretické rozdělení. Původní hodnoty můžeme nejprve transformovat (**Transform**). K dispozici jsou transformace přirozeným logaritmem, z-skóry, difference a sezónní difference pro časové řady s definovanou periodou. Podrobnosti výpočtu nastavíme v polích **Proportion Estimation Formula** a **Rank Assigned to Ties**.



**Kvantilový graf**

## Kvantilový graf

- Analyze
  - Descriptive Statistics
    - Q-Q Plots

### Slovník anglických pojmů

<i>continue</i>	<i>pokračovat</i>
<i>degrees of freedom</i>	<i>stupně volnosti</i>
<i>difference</i>	<i>rozdíl</i>
<i>distribution</i>	<i>rozložení, rozdělení</i>
<i>error</i>	<i>chyba</i>
<i>estimate</i>	<i>odhad</i>
<i>mean</i>	<i>průměr</i>
<i>periodicity</i>	<i>periodicita</i>
<i>proportion</i>	<i>proporce, podíl</i>
<i>quantile</i>	<i>kvantil</i>
<i>rank</i>	<i>pořadí</i>
<i>standardize</i>	<i>standardizovat</i>
<i>tie</i>	<i>shoda</i>
<i>transform</i>	<i>transformovat</i>
<i>variable</i>	<i>proměnná</i>



## 1.3 Porovnání průměrů (Compare Means)

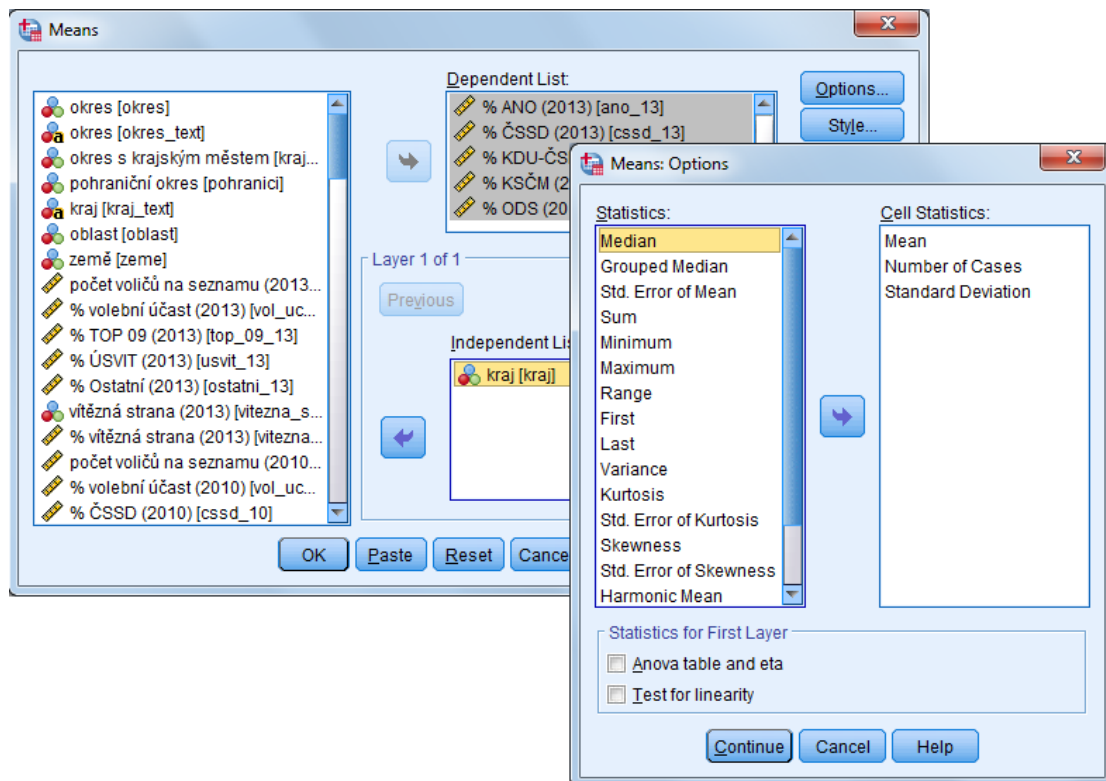
### 1.3.1 Průměry (Means)

Procedura počítá průměry a další popisné statistiky pro závislé proměnné. Zadáním jedné nebo více nezávislých proměnných lze získat také popisné statistiky ve skupinách určených kategoriemi těchto proměnných.

Do okna **Dependent List** převedeme analyzovanou proměnnou (proměnné). Třídící proměnnou zadáme do pole **Independent List**. Je-li třídících proměnných více, lze zobrazit pro každou z nich samostatnou tabulku nebo vytvořit podrobněji strukturovanou tabulku. V prvním případě pouze přeneseme všechny třídící proměnné do pole **Independent List**. Ve druhém přeneseme nejprve do pole **Independent List** proměnnou, která má vytvářet první stupeň třídění. Poté stiskneme tlačítko **Next** a stejným způsobem zadáme třídící proměnnou pro druhý a případně další stupeň. Tlačítka **Next** a **Previous** jsou určena k pohybu mezi vrstvami (v tomto případě se však nejedná o vrstvy ve stejném smyslu jako u pivotování, ale o podrobnější třídění tabulky).

Tlačítkem **Options** volíme počítané statistiky. Pro první úroveň třídění lze rovněž zadat výpočet analýzy rozptylu a testu linearitu (**Statistics for First Layer**).

Tlačítko **Style** je určeno k tzv. podmíněnému formátování, které dovoluje nastavit formát pivotních tabulek již při jejich vytváření (podrobněji viz *Příloha 1 Podmíněné formátování tabulek*).



**Výpočet statistik ve skupinách**

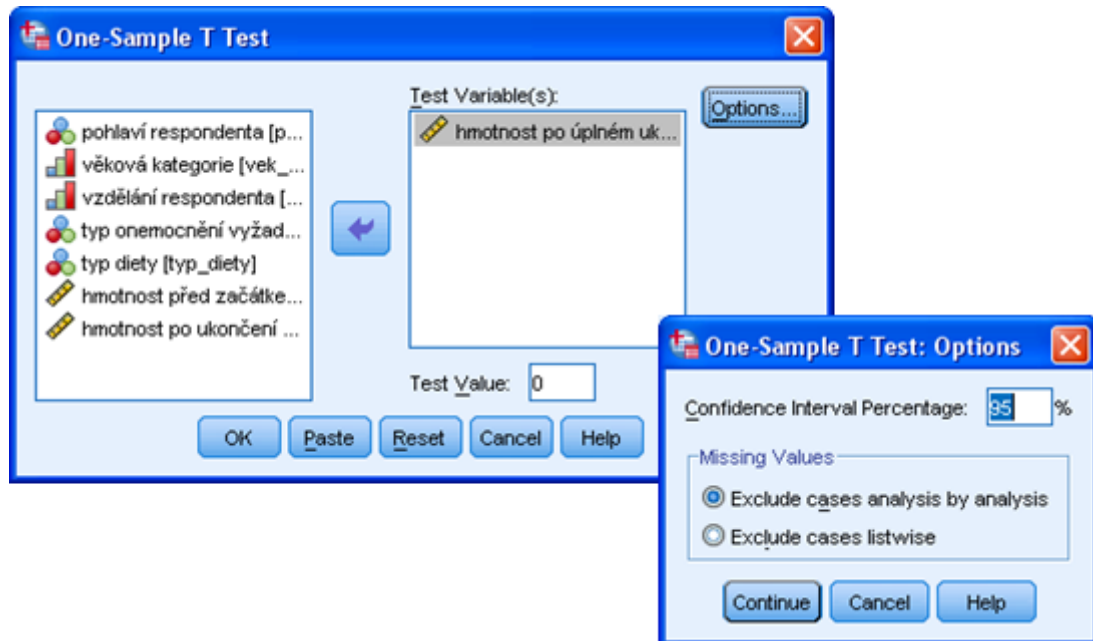
### Výpočet průměrů

- Analyze
  - Compare Means
    - Means

#### 1.3.2 Jednovýběrový T-test (One-Sample T Test)

Procedura testuje u jednotlivých proměnných hypotézu, že jejich střední hodnota je rovna zadané konstantě.

Do okna **Test Variable(s)** převedeme analyzovanou proměnnou (proměnné). V poli **Test Value** zadáme konstantu, se kterou budeme průměr porovnávat. Pomocí tlačítka **Options** lze volit interval spolehlivosti (**Confidence Interval Percentage**) a přístup ke zpracování chybějících hodnot (**Missing Values**).



*Jednovýběrový T-test*

### **Jednovýběrový T-test**

- Analyze
  - Compare Means
    - One-Sample T Test

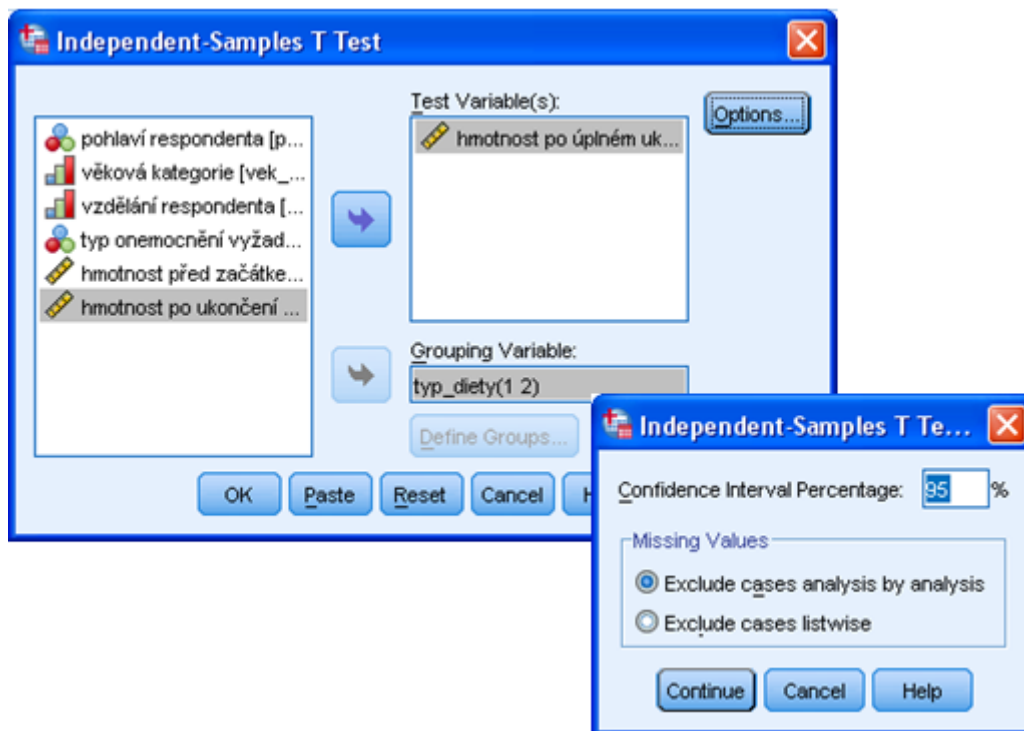
### **Slovník anglických pojmů**

<i>confidence interval</i>	<i>interval spolehlivosti</i>
<i>exclude</i>	<i>vynechat</i>
<i>listwise</i>	<i>celý seznam</i>
<i>missing</i>	<i>vynechávaná, chybějící</i>
<i>one-sample</i>	<i>jednovýběrový</i>
<i>options</i>	<i>volby</i>
<i>test value</i>	<i>testová hodnota,</i>
<i>test variable</i>	<i>testovaná proměnná</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

### 1.3.3 T-test pro dva nezávislé výběry (Independent-Samples T Test)

Procedura testuje rozdíl středních hodnot dvou skupin případů (nezávislých výběrů).

Analyzovanou proměnnou (proměnné) převedeme do okna **Test Variable(s)**, třídící proměnnou do pole **Grouping Variable**. Dále je nutné zadat tlačítkem **Define Groups** kategorie třídící proměnné, které definují rozdělení souboru na dvě skupiny, nebo hodnotu, která rozdělí interval na dvě části. Tlačítkem **Options** volíme interval spolehlivosti (**Confidence Interval Percentage**) a způsob práce s chybějícími hodnotami (**Missing Values**).



*T-test pro nezávislé skupiny*

#### **T-test pro nezávislé skupiny**

- Analyze
  - Compare Means
    - Independent-Samples T Test

### Slovník anglických pojmů

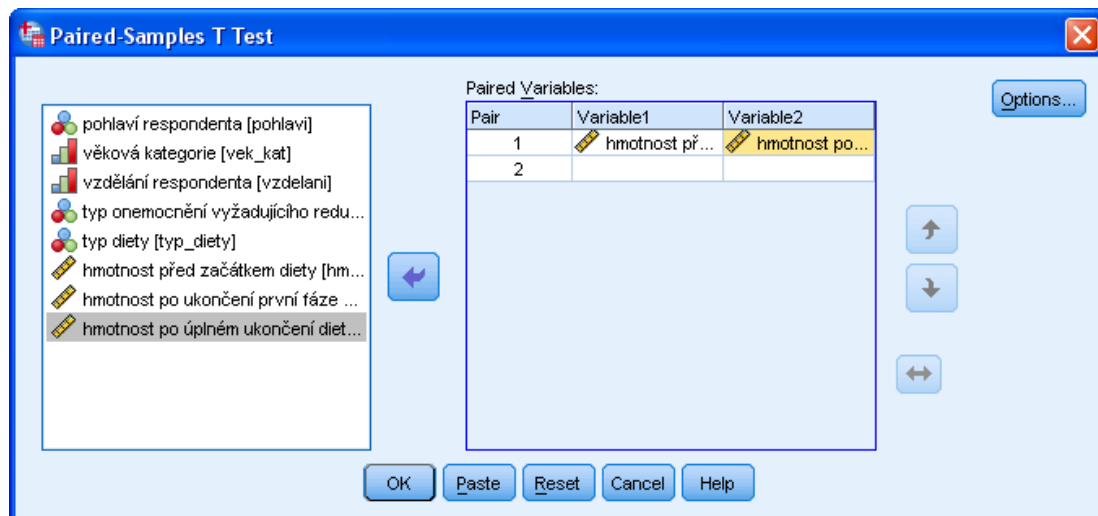
<i>ANOVA</i>	<i>analýza rozptylu</i>
<i>cell</i>	<i>buňka</i>
<i>compare</i>	<i>srovnat</i>
<i>confid.interval</i>	<i>interval spolehlivosti</i>
<i>count</i>	<i>počet (případů)</i>
<i>define</i>	<i>určit, definovat</i>
<i>dependent</i>	<i>závislý</i>
<i>display</i>	<i>zobrazit</i>
<i>exclude</i>	<i>vynechat</i>
<i>first</i>	<i>první</i>
<i>group</i>	<i>skupina</i>
<i>grouping</i>	<i>seskupovací</i>
<i>independent</i>	<i>nezávislý</i>
<i>label</i>	<i>popisek</i>
<i>layer</i>	<i>vrstva, úroveň</i>
<i>linearity</i>	<i>linearita, přímková závislost</i>
<i>list</i>	<i>seznam</i>
<i>listwise</i>	<i>celý seznam</i>
<i>mean</i>	<i>průměr</i>
<i>missing</i>	<i>vynechávaná, chybějící</i>
<i>next</i>	<i>další, následující</i>
<i>none</i>	<i>žádný</i>
<i>one-sample</i>	<i>jednovýběrový</i>
<i>only</i>	<i>jen</i>
<i>options</i>	<i>volby</i>
<i>previous</i>	<i>předchozí</i>
<i>standard deviation</i>	<i>směrod. odchylka</i>
<i>sum</i>	<i>suma</i>
<i>test value</i>	<i>testová hodnota,</i>
<i>test variable</i>	<i>testovaná proměnná</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>variance</i>	<i>rozptyl</i>

### 1.3.4 Párový T-test (Paired-Samples T Test)

Procedura porovnává průměry dvou proměnných. Pro každý případ spočítá rozdíl těchto proměnných a testuje, zda se střední hodnota diferencí liší od nuly.

Nejprve označíme první analyzovanou proměnnou a přeneseme ji šipkou do pole **Paired Variables**. Její název se automaticky zobrazí v části **Variable1**. Stejným způsobem přeneseme druhou proměnnou do pole **Variable2**.

Tlačítkem **Options** volíme interval spolehlivosti (**Confidence Interval Percentage**) a způsob práce s chybějícími hodnotami (**Missing Values**).



**Párový T-test**

### **Párový T-test**

- *Analyze*
  - *Compare Means*
    - *Paired-Samples T Test*

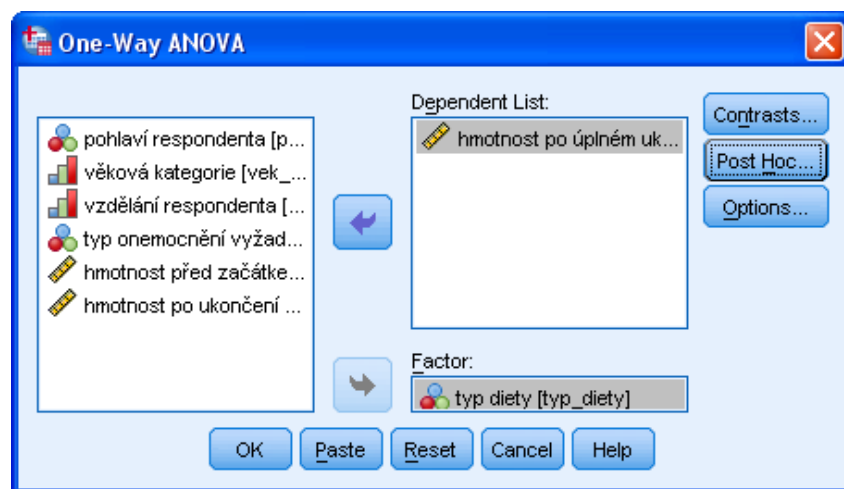
### **Slovník anglických pojmů**

<i>case</i>	<i>případ</i>
<i>confidence interval</i>	<i>interval spolehlivosti</i>
<i>exclude</i>	<i>vynechat</i>
<i>listwise</i>	<i>celý seznam</i>
<i>missing</i>	<i>vynechávaná, chybějící</i>
<i>options</i>	<i>volby</i>
<i>pair</i>	<i>pár</i>
<i>variable</i>	<i>proměnná</i>

### 1.3.5 Jednoduchá analýza rozptylu (One-Way ANOVA)

Procedura je určena k testování hypotézy, že střední hodnoty podrobnějšího třídění podle zvoleného faktoru se v jednotlivých skupinách rovnají. Jedná se o rozšíření dvouvýběrového T-testu. Procedura se aplikuje na závislou kvantitativní proměnnou za přítomnosti jedné nezávislé proměnné (faktor). Navíc obsahuje dva typy testů pro porovnání průměrů: kontrasty a post-hoc testy.

Analyzovanou proměnnou (proměnné) převedeme do okna **Dependent List**, zvolenou třídící proměnnou do okna **Factor**. Tlačítko **Contrasts** slouží k definování kontrastů, tlačítko **Post Hoc** k výběru párových testů. Po stisku tlačítka **Options** lze vybrat zobrazované statistiky (**Statistics**), nakreslit graf průměrů ve skupinách (**Means plot**) a určit přístup k zpracování vynechaných hodnot (**Missing Values**).



*Jednoduchá analýza rozptylu*

#### **Jednoduchá analýza rozptylu**

- *Analyze*
  - *Compare Means*
    - *One-Way ANOVA*

**Slovník anglických pojmů**

<i>average</i>	<i>průměr</i>
<i>confidence int.</i>	<i>interval spolehlivosti</i>
<i>continue</i>	<i>pokračovat</i>
<i>contrast</i>	<i>kontrast</i>
<i>current</i>	<i>současný</i>
<i>define</i>	<i>definovat, určit</i>
<i>dependent</i>	<i>závislý</i>
<i>display</i>	<i>zobrazit</i>
<i>exclude</i>	<i>vynechat</i>
<i>factor</i>	<i>faktor, nezávislá proměnná</i>
<i>group</i>	<i>skupina</i>
<i>harmonic aver.</i>	<i>harmonický průměr</i>
<i>independent</i>	<i>nezávislý</i>
<i>label</i>	<i>popis</i>
<i>list</i>	<i>seznam</i>
<i>listwise</i>	<i>celý</i>
<i>missing</i>	<i>chybějící</i>
<i>one-way</i>	<i>jedno třídění, jeden faktor</i>
<i>options</i>	<i>možnosti, volby</i>
<i>paired</i>	<i>párový, spárovaný</i>
<i>paired var.</i>	<i>spárované proměnné</i>
<i>range</i>	<i>rozpětí</i>
<i>selection</i>	<i>výběr</i>
<i>total</i>	<i>celkem</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>



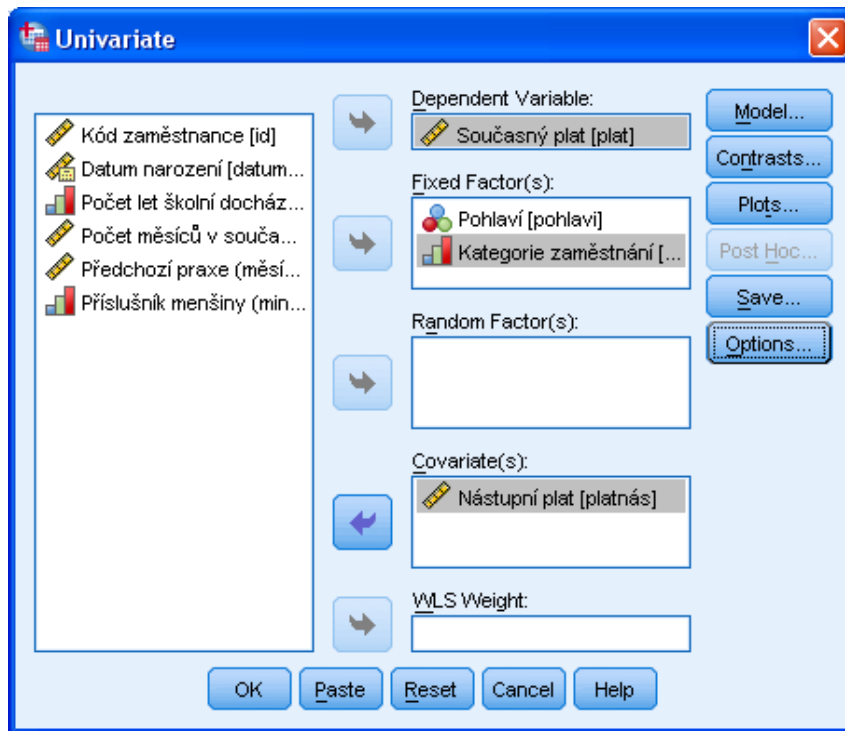
## 1.4 Obecný lineární model (General Linear Model)

### 1.4.1 Jednoduchý lineární model (Univariate)

Jednoduchý lineární model je obecnější metoda, která v sobě zahrnuje analýzu variance (ANOVA), analýzu kovariance (ANCOVA) a regresní analýzu. ANOVA analyzuje vztah mezi jednou závislou číselnou proměnnou a jednou nebo více nezávislými kategorizovanými proměnnými (faktory). Sledujeme-li navíc číselné proměnné (tzn. kovariáty), které mohou ovlivnit studovaný vztah, a chceme-li, aby případná závislost mezi číselnou proměnnou a faktory byla očištěna od tohoto vlivu, jedná se o metodu ANCOVA. Regresní analýza hodnotí vztah mezi jednou závislou číselnou proměnnou a jednou nebo několika nezávislými číselnými proměnnými.

Do pole **Dependent Variable** převedeme číselnou proměnnou, jejíž závislost na dalších proměnných analyzujeme. Do pole **Fixed Factor(s)** zadáme nezávislé kategorizované proměnné, u nichž předpokládáme, že v datovém souboru nabývají všech možných hodnot. Do pole **Random Factor(s)** můžeme zadat kategorizované proměnné, které v datovém souboru nabývají pouze výběr všech možných kategorií. Pole **Covariate(s)** specifikuje číselné nezávislé proměnné, tzv. kovariáty, u kterých se předpokládá lineární závislost s vysvětlovanou proměnnou. Pole **WLS Weight** umožňuje definovat váhy pro jednotlivé případy.

Tlačítko **Model** slouží ke specifikaci modelu (lze zadat například interakce mezi nezávislými proměnnými). Pomocí tlačítka **Contrasts** definujeme požadované kontrasty. Tlačítko **Plots** zobrazí vybrané grafy. Tlačítko **Post Hoc** nabízí řadu testů identifikujících skupiny, jejichž průměry se mezi sebou signifikantně liší. Tlačítko **Save** vytvoří nové proměnné v datovém souboru, které obsahují odhad hodnot závislé proměnné a další odvozené statistiky. Volbou **Options** nastavíme požadované výstupy.



*Jednoduchý lineární model*

### ***Jednoduchý lineární model***

- *Analyze*
  - *General Linear Model*
    - *Univariate*

### Slovník anglických pojmů

<i>add</i> .....	<i>přidat</i>
<i>assumed</i> .....	<i>předpokládaný</i>
<i>axis</i> .....	<i>osa</i>
<i>case</i> .....	<i>případ</i>
<i>change</i> .....	<i>změnit</i>
<i>close</i> .....	<i>zavřít</i>
<i>compare</i> .....	<i>porovnat</i>
<i>confidence interval</i> .....	<i>interval spolehlivosti</i>
<i>continue</i> .....	<i>pokračovat</i>
<i>contrast</i> .....	<i>kontrast, porovnat</i>
<i>custom</i> .....	<i>vlastní</i>
<i>dependent</i> .....	<i>závislý</i>
<i>descriptive</i> .....	<i>popisný</i>
<i>deviation</i> .....	<i>deviace, odchylka</i>
<i>display</i> .....	<i>zobrazit</i>
<i>distance</i> .....	<i>vzdálenost</i>
<i>equal</i> .....	<i>rovný, roven</i>
<i>error</i> .....	<i>chyba</i>
<i>estimate</i> .....	<i>odhad</i>
<i>file</i> .....	<i>soubor</i>
<i>first</i> .....	<i>první</i>
<i>full</i> .....	<i>plný, zcela</i>
<i>general</i> .....	<i>obecný</i>
<i>independent</i> .....	<i>nezávislý</i>
<i>last</i> .....	<i>poslední</i>
<i>line</i> .....	<i>čára</i>
<i>matrix</i> .....	<i>matice</i>
<i>mean</i> .....	<i>průměr</i>
<i>new</i> .....	<i>nový</i>
<i>none</i> .....	<i>žádný</i>
<i>options</i> .....	<i>volby</i>
<i>plot</i> .....	<i>graf</i>
<i>random</i> .....	<i>náhodný</i>
<i>remove</i> .....	<i>odstranit</i>
<i>save</i> .....	<i>uložit</i>
<i>separate</i> .....	<i>oddělený, oddělit</i>
<i>significance</i> .....	<i>význam</i>
<i>specify</i> .....	<i>upřesnit</i>
<i>standard error</i> .....	<i>standardní chyba</i>
<i>sum of squares</i> .....	<i>součet čtverců</i>
<i>univariate</i> .....	<i>s jednou proměnnou</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>
<i>variance</i> .....	<i>rozptyl</i>
<i>weight</i> .....	<i>váha</i>

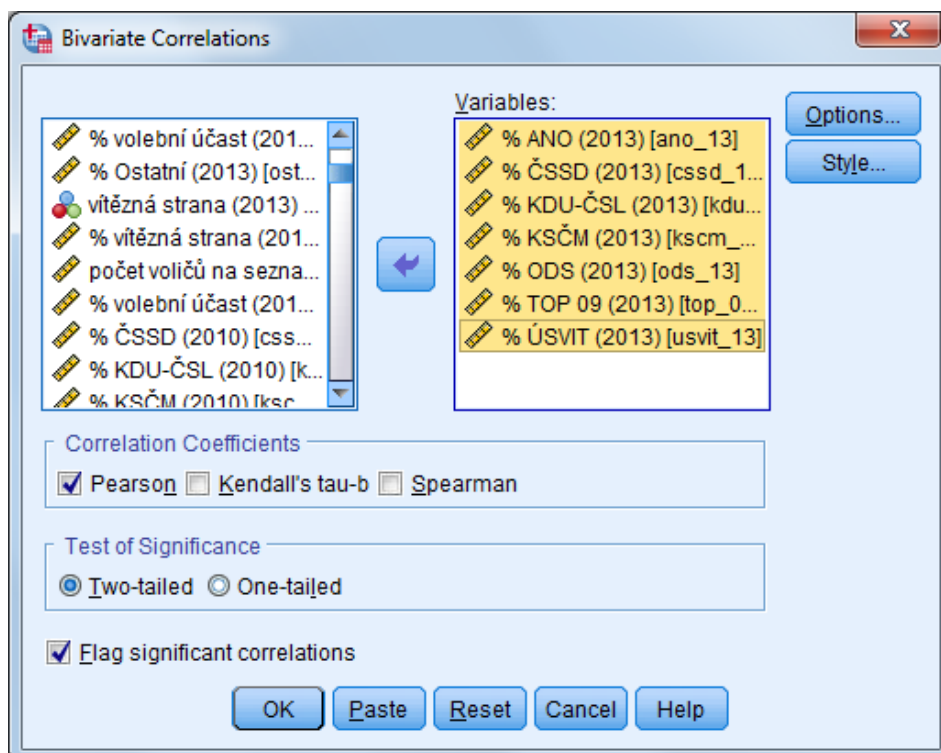
## 1.5 Korelační analýza (Correlate)

### 1.5.1 Korelace (Bivariate Correlations)

Procedura počítá Pearsonův lineární korelační koeficient, Spearmanův koeficient rho a Kendallovo tau-b pro dvojice proměnných.

Do pole **Variables** převedeme proměnné, jejichž korelaci zjišťujeme. Dále zvolíme typ korelačního koeficientu (**Correlation Coefficients**) a typ testu významnosti (**Test of Significance**). Chceme-li ve výstupu zvýraznit statisticky významné korelace, zaškrtneme volbu **Flag significant correlations**. Tlačítko **Options** slouží k volbě dalších statistik (**Statistics**) a způsobu práce s vynechanými hodnotami (**Missing Values**).

Tlačítko **Style** je určeno k tzv. podmíněnému formátování, které dovoluje nastavit formát pivotních tabulek již při jejich vytváření (podrobněji viz *Příloha 1 Podmíněné formátování tabulek*).



**Korelace**

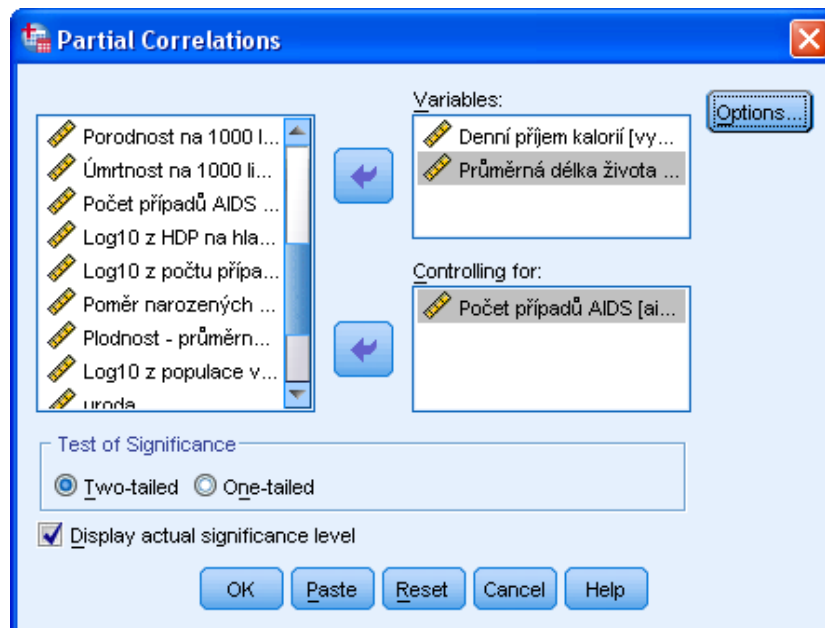
### **Korelace**

- Analyze
  - Correlate
    - Bivariate Correlations

### 1.5.2 Parciální korelace (Partial Correlations)

Procedura počítá parciální korelační koeficient, který vyjadřuje sílu lineárního vztahu mezi dvěma proměnnými po očištění od vlivu jedné nebo více jiných proměnných.

Do okna **Variables** převedeme proměnné, jejichž korelační koeficient chceme spočítat. Kontrolní proměnnou (proměnné) přeneseme do okna **Controlling for**. Dále zvolíme jednostrannou nebo oboustrannou variantu testu (**Test of Significance**). Tlačítkem **Options** zadáme požadované statistiky (**Statistics**) a rozhodneme o přístupu k vynechaným hodnotám (**Missing Values**).



*Parciální korelace*

### **Parciální korelace**

- Analyze
  - Correlate
    - Partial Correlations

**Slovník anglických pojmů**

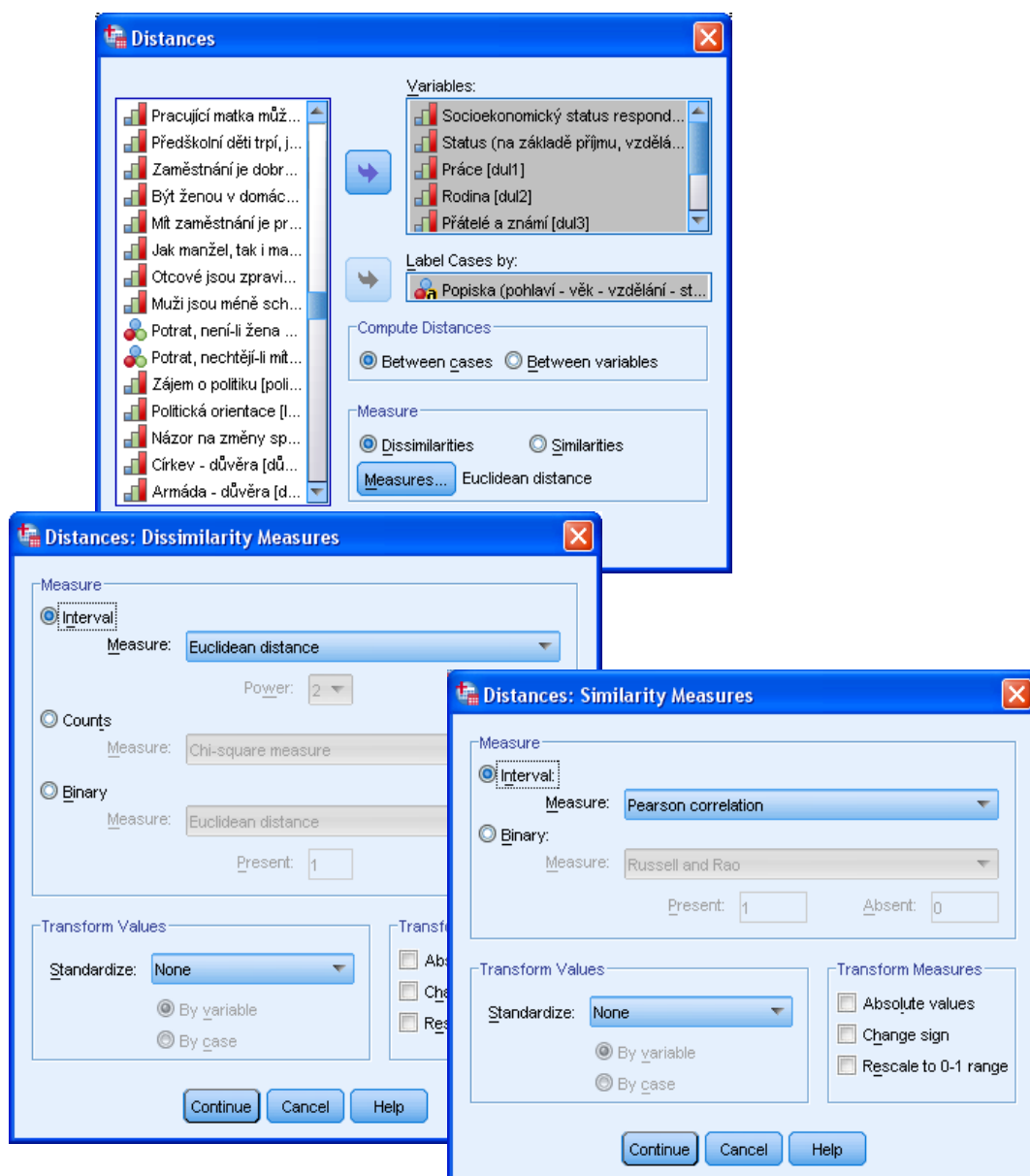
<i>absolute value</i>	<i>absolutní hodnota</i>
<i>actual</i>	<i>skutečný</i>
<i>between</i>	<i>mezi</i>
<i>case</i>	<i>případ</i>
<i>close</i>	<i>zavřít</i>
<i>compute</i>	<i>vypočítat</i>
<i>continue</i>	<i>pokračovat</i>
<i>controlling for</i>	<i>kontrolováno pro</i>
<i>correlate</i>	<i>korelovat</i>
<i>count</i>	<i>počet</i>
<i>covariance</i>	<i>kovariance</i>
<i>customized</i>	<i>vlastní nastavení</i>
<i>exclude</i>	<i>vyločit</i>
<i>change</i>	<i>změnit</i>
<i>interval</i>	<i>intervalový</i>
<i>label</i>	<i>označit</i>
<i>level</i>	<i>úroveň</i>
<i>listwise</i>	<i>celý seznam</i>
<i>mean</i>	<i>průměr</i>
<i>measure</i>	<i>měření</i>
<i>none</i>	<i>žádný</i>
<i>one-tailed</i>	<i>jednostranný (test)</i>
<i>options</i>	<i>možnosti, volby</i>
<i>pairwise</i>	<i>po dvojicích</i>
<i>partial</i>	<i>parciální</i>
<i>power</i>	<i>mocnina</i>
<i>present</i>	<i>je přítomno</i>
<i>rescale</i>	<i>změnit škálu</i>
<i>root</i>	<i>odmocnina</i>
<i>significance</i>	<i>významnost</i>
<i>significance level</i>	<i>hladina významnosti</i>
<i>stand.deviation</i>	<i>směrod.odchylka</i>
<i>transform</i>	<i>změnit</i>
<i>two-tailed</i>	<i>dvoustranný (test)</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>zero-order correlations</i>	<i>neparcializované korelace</i>

### 1.5.3 Vzdálenosti (Distances)

Procedura počítá různé míry podobnosti nebo nepodobnosti (vzdálenosti) pro proměnné nebo případy.

Do pole **Variables** zadáme proměnné, které určují profil případů (dimenze) při výpočtu vzdáleností mezi případy, nebo proměnné, jejichž vzdálenosti analyzujeme. V poli **Label Cases by** lze specifikovat proměnnou, která popisuje jednotlivé případy. V části **Compute Distances** určíme, zda se budou počítat vzdálenosti mezi případy (**Between cases**) nebo mezi proměnnými (**Between variables**).

V části **Measure** se rozhodneme pro výpočet měr podobnosti (**Similarities**) nebo nepodobnosti (**Dissimilarities**) a tlačítkem **Measures** detailně specifikujeme požadovanou míru. Podle potřeby lze rovněž transformovat data před výpočtem míry nebo přímo výsledné hodnoty.



Vzdálenosti

### Vzdálenosti

- Analyze
  - Correlate
    - Distances

#### Slovník anglických výrazů

<i>between</i> .....	<i>mezi</i>
<i>between cases</i> .....	<i>mezi případy</i>
<i>between var.</i> .....	<i>mezi proměnnými</i>
<i>binary</i> .....	<i>binární, dvouhodnotový</i>
<i>case</i> .....	<i>případ</i>
<i>change</i> .....	<i>změna</i>
<i>compute</i> .....	<i>vypočítat</i>
<i>continue</i> .....	<i>pokračovat</i>
<i>count</i> .....	<i>počet</i>
<i>disimilarities</i> .....	<i>nepodobnost</i>
<i>distance</i> .....	<i>vzdálenost</i>
<i>Euclidean distance.</i> .....	<i>eukleidovská vzdálenost</i>
<i>label</i> .....	<i>popis, označení</i>
<i>measure</i> .....	<i>měření</i>
<i>similarities</i> .....	<i>podobnosti</i>
<i>transform</i> .....	<i>změnit</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>



## 1.6 Regresní analýza (Regression)

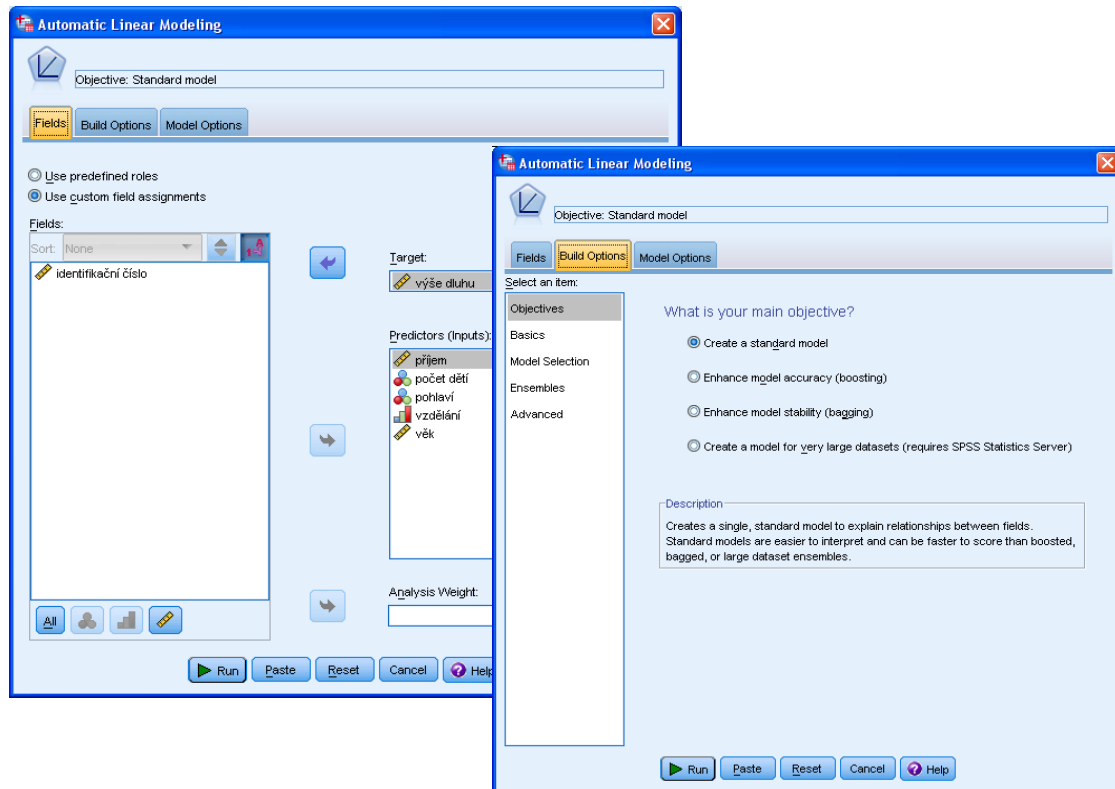
### 1.6.1 Automatická tvorba lineárního modelu (Automatic Linear Modeling)

Procedura je určena k modelování jedné číselné závislé proměnné na základě sady nezávislých prediktorů, které mohou být číselné i kategorizované.

Na záložce **Fields** zadáme do pole **Target** cílovou proměnnou, do pole **Predictors** nezávislé proměnné. Nastavení provedeme ručně, nebo označíme volbu **Use predefined roles**, čímž se proměnné umístí do příslušných polí automaticky dle své nastavené role (podrobněji o roli proměnných viz popis záložky Variable View, skripta *IBM SPSS Statistics Base 24*, část *Datový soubor a práce s výstupem*). Do pole **Analysis Weight** můžeme zadat váhy případů v modelu.

Na záložce **Build Options** určíme v nabídce **Objectives** cíl modelování: tvorba standardního lineárního regresního modelu (**Create a standard model**), dosažení co nejpřesnějšího modelu (**Enhance model accuracy (boosting)**), dosažení co nejstabilnějšího modelu (**Enhance model stability (bagging)**) nebo tvorba modelu pro rozsáhlá data s využitím IBM SPSS Statistics Server (**Create a model for very large datasets**). V nabídce **Basics** je k dispozici automatická příprava dat zahrnující zpracování proměnných ve formátu datum a čas, nastavení vhodných typů proměnných, ošetření odlehlých a vynechaných pozorování a sloučení kategorií. Dále lze v této části nastavit požadovaný interval spolehlivosti. V nabídce **Model Selection** specifikujeme metodu pro výběr prediktorů (**Model selection method**), případně provedeme další podrobnější nastavení dané metody. Pokud je cílem modelování vytvoření co nejpřesnějšího nebo nejstabilnějšího modelu, zvolíme v nabídce **Ensembles** podrobnosti metod boosting či bagging. V nabídce **Advanced** nastavíme parametry pro opětovnou replikaci výsledků.

Na záložce **Model Options** lze zadat uložení predikovaných hodnot do datové matice (**Save predicted values to the dataset**) a export výsledného modelu ve formátu XML souboru (**Export model**), např. pro skórování nových případů.



**Automatická tvorba lineárního modelu**

### **Automatická tvorba lineárního modelu**

- Analyze
  - Regression
    - Automatic Linear Modeling

### Slovník anglických pojmů

<i>all</i> .....	všechny
<i>confidence interval</i> .....	interval spolehlivosti
<i>forward</i> .....	postupné zahrnování
<i>include</i> .....	zahrnout
<i>mean</i> .....	průměr
<i>method</i> .....	metoda
<i>options</i> .....	volby
<i>predicted values</i> .....	predikované hodnoty
<i>regression</i> .....	regrese, regresní
<i>save</i> .....	uložit
<i>stepwise</i> .....	krokově
<i>value</i> .....	hodnota
<i>variable</i> .....	proměnná
<i>weight</i> .....	váha

#### 1.6.2 Lineární regrese (Linear Regression)

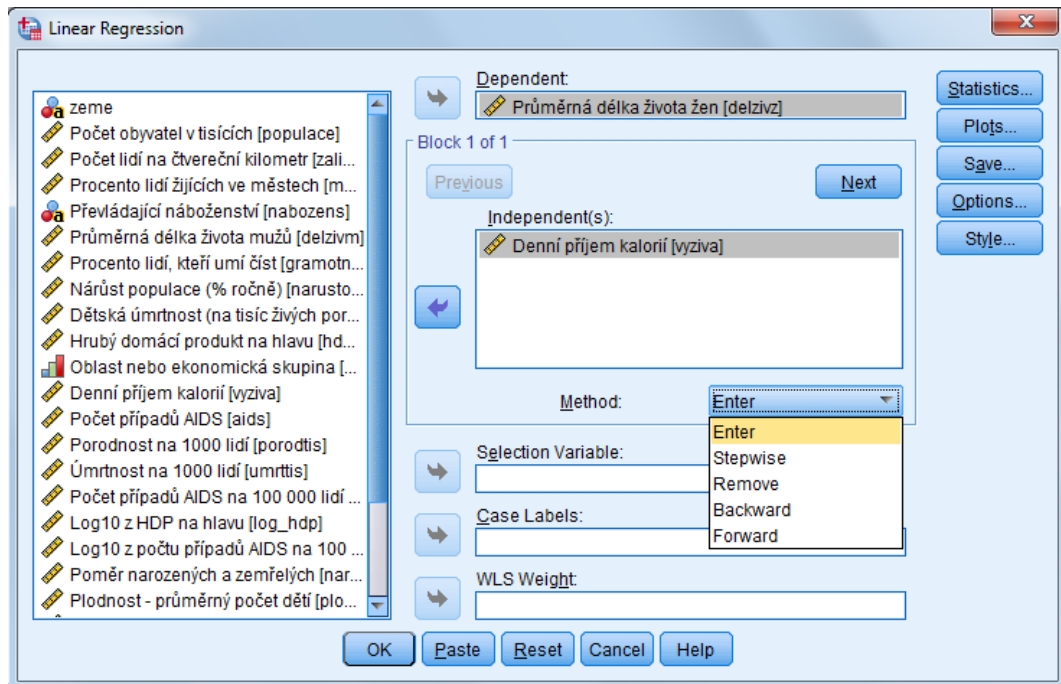
Procedura odhaduje metodou nejmenších čtverců koeficienty pro lineární model, který obsahuje jednu nebo více nezávislých proměnných.

Do okna **Dependent** převedeme tzv. závislou proměnnou, do okna **Independent(s)** proměnné, které vstupují do analýzy jako nezávislé. V poli **Method** lze zadat metodu výběru nezávislých proměnných. Chceme-li zvolit odlišné metody pro různé skupiny nezávislých proměnných, použijeme tlačítko **Next**.

Zadáním podmínky založené na rovnosti či nerovnosti hodnot proměnné v poli **Selection Variable** pomocí tlačítka **Rule** lze omezit případy vstupující do modelu. Dále můžeme přenést do pole **Case Labels** proměnnou, pomocí které popíšeme případy. Pole **WLS Weight** umožňuje definovat váhy pro jednotlivé případy.

Tlačítko **Statistics** slouží k volbě zobrazovaných statistik, pomocí tlačítka **Plots** lze vykreslit vybrané grafy. Tlačítko **Save** se užívá k uložení predikovaných hodnot (**Predicted Values**), vzdáleností (**Distances**), intervalů spolehlivosti (**Prediction Intervals**), reziduí (**Residuals**) a míry vlivu případů na průběh regrese (**Influence Statistics**) do datové matice. Kromě toho umožňuje exportovat model ve formátu XML (**Export model information to XML file**). Tlačítkem **Options** nastavíme kritéria dalších kroků analýzy: meze pravděpodobnosti F-testu, mezní hodnoty F-testu. Lze též zvolit metodu práce s vynechanými hodnotami: vynechání celého případu, vynechání případu pouze u proměnné, kde byla nalezena vynechaná hodnota, nahrazení vynechané hodnoty průměrem.

Tlačítko **Style** je určeno k tzv. podmíněnému formátování, které dovoluje nastavit formát pivotních tabulek již při jejich vytváření (podrobněji viz *Příloha 1 Podmíněné formátování tabulek*).



*Lineární regrese*

### *Lineární regresní analýza*

- Analyze
  - Regression
    - Linear Regression

### Slovník anglických pojmů

<i>adjusted</i>	<i>adjustovaný, upravený</i>
<i>all</i>	<i>všechny</i>
<i>backward</i>	<i>postupné vyjímání</i>
<i>block</i>	<i>blok (výpočtu)</i>
<i>case</i>	<i>případ</i>
<i>casewise</i>	<i>po případech</i>
<i>collinearity</i>	<i>kolinearita</i>
<i>confidence interval</i>	<i>interval spolehlivosti</i>
<i>constant</i>	<i>konstanta v rovnici</i>
<i>continue</i>	<i>pokračovat</i>
<i>covariance</i>	<i>kovariance</i>
<i>deleted</i>	<i>vyseknutý</i>
<i>dependent</i>	<i>závislý</i>
<i>descriptives</i>	<i>popisné statistiky</i>
<i>distance</i>	<i>vzdálenost</i>
<i>enter</i>	<i>určený vstup</i>
<i>equation</i>	<i>rovnice</i>
<i>estimation</i>	<i>odhad</i>
<i>F-value</i>	<i>hodnota statistiky F</i>
<i>forward</i>	<i>postupné zahrnování</i>
<i>include</i>	<i>zahrnout</i>
<i>independent</i>	<i>nezávislý</i>
<i>individual</i>	<i>individuální</i>
<i>influence st.</i>	<i>influenční statistiky, statistiky vlivu</i>
<i>label</i>	<i>popis</i>
<i>list</i>	<i>seznam</i>
<i>matrix</i>	<i>matice</i>
<i>mean</i>	<i>průměr</i>
<i>measure</i>	<i>měření</i>
<i>method</i>	<i>metoda</i>
<i>model fit</i>	<i>vhodný model</i>
<i>next</i>	<i>příští, další</i>
<i>options</i>	<i>volby</i>
<i>outliers</i>	<i>vnější pozorování</i>
<i>partial</i>	<i>dílčí, parciální</i>
<i>plot</i>	<i>graf</i>
<i>predicted value</i>	<i>predikované hodnoty</i>
<i>previous</i>	<i>předchozí</i>
<i>probability</i>	<i>pravděpodobnost</i>
<i>produce</i>	<i>produkovat</i>
<i>regression</i>	<i>regrese, regresní</i>
<i>remove</i>	<i>vyjmout</i>
<i>replace</i>	<i>zaměnit</i>
<i>residual</i>	<i>reziduální</i>
<i>S.E. of Mean</i>	<i>stand. chyba průměru</i>
<i>save</i>	<i>uložit</i>
<i>standardized</i>	<i>standardizovaný</i>
<i>standard deviation</i>	<i>směrod. odchylka</i>
<i>stepwise</i>	<i>krokově</i>
<i>unstandardized</i>	<i>nestandardizovaný</i>
<i>use</i>	<i>použít</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>weight</i>	<i>váha</i>
<i>WLS</i>	<i>metoda vážených nejmenších čtverců</i>

### 1.6.3 Prokládání křivky (Curve Estimation)

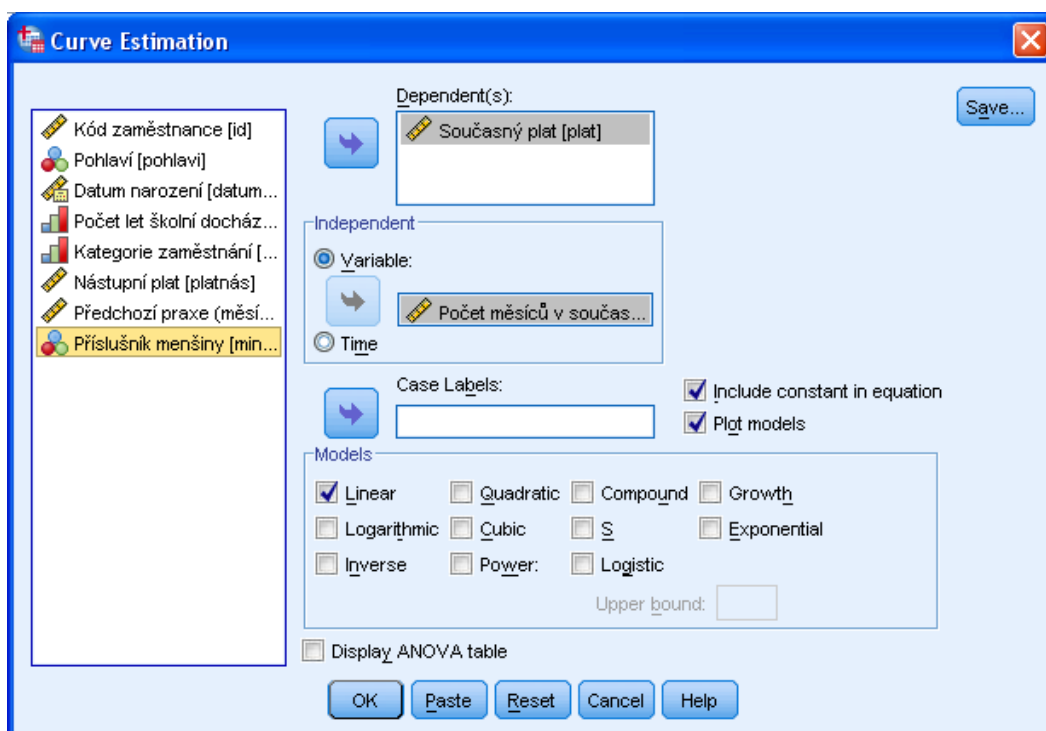
Procedura prokládá data na základě regresního modelu vybraný typ křivky. K dispozici je celkem 11 odlišných modelů (lineární, kvadratický, exponenciální atd.), které jsou transformacemi převoditelné na lineární regresi. Pro každou závislou proměnnou se odhaduje samostatný model (lze zadat i více modelů současně). Procedura zobrazí ve výstupu přehled základních charakteristik jednotlivých modelů a grafy. Do datové matice lze uložit jako nové proměnné predikované hodnoty, rezidua a konfidenční intervaly.

Do okna **Dependent(s)** převedeme závislou proměnnou (proměnné). Jestliže hodnoty závislé proměnné tvoří časovou řadu a její závislost na čase (tj. na čísle řádku v datové matici), zaškrtneme **Time** v sekci **Independent**. V ostatních případech převedeme do okna **Variable** proměnnou, která vstoupí do analýzy jako nezávislá. Proměnnou popisující jednotlivé případy můžeme zadat do okna **Case Labels**.

Jestliže má rovnice zahrnovat konstantu, zaškrtneme políčko **Include constant in equation**. Pro grafické znázornění výsledků označíme možnost **Plot models**. Sekce **Models** slouží k výběru modelů. Po zaškrtnutí **Display ANOVA table** se zobrazí tabulka analýzy rozptylu.

Tlačítko **Save** umožňuje uložit predikované hodnoty (**Predicted values**), jejich intervaly spolehlivosti (**Prediction intervals**) a rezidua (**Residuals**). Případy mohou být predikovány (**Predict Cases**) od prvního řádku vybraného pro analýzu do konce souboru nebo do zvoleného případu.

Pozn.: **Estimation period** je časový interval, na jehož hodnotách zakládáme odhad parametrů modelu. Ostatní hodnoty před použitím regrese ze souboru odfiltrujeme. Pro kontrolu bývá zvykem odfiltrovat také několik posledních pozorování, aby byla možnost na nich posoudit predikční sílu odvozeného modelu.



Prokládání křivky

## Prokládání křivky

- Analyze
  - Regression
    - Curve Estimation

### Slovník anglických výrazů

ANOVA .....	analýza rozptylu
case .....	případ
curve.....	křivka
dependent .....	závislý
display .....	zobrazit
equation .....	rovnice
estimation .....	odhad
include.....	zahrnout
independent.....	nezávislý
label.....	popis, označení
last case .....	poslední případ
method.....	metoda
observation .....	pozorování
plot.....	graf
predicted values.....	predikované hodnoty
regression.....	regrese
residual.....	residuální, reziduum
save .....	uložit
table .....	tabulka
time .....	čas
value.....	hodnota
variable .....	proměnná

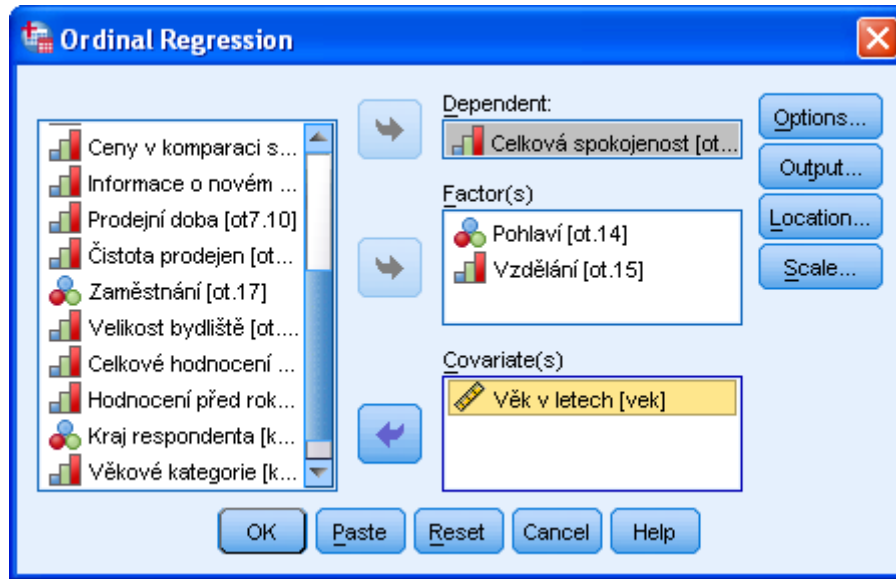
### 1.6.4 Ordinální regrese (Ordinal)

Ordinální regrese odhaduje hodnoty ordinální závislé proměnnou na základě jedné nebo více nezávislých proměnných, které mohou být číselné i kategorizované. Obyčejnou lineární regresi není vhodné v takovém případě použít, protože vzdálenosti mezi sousedními kategoriemi ordinální proměnné nemusejí být stejné.

Jestliže například vytváříme model pro odhad reakce pacienta na léčbu (žádná, slabá, střední, prudká), nemusí být rozdíl mezi žádnou a slabou reakcí stejný jako rozdíl mezi střední a prudkou reakcí, nebo lze tyto rozdíly jen obtížně kvantifikovat.

Do okna **Dependent** převedeme závislou proměnnou. Do pole **Factor(s)** zadáme kategorizované nezávislé proměnné, v poli **Covariate(s)** specifikujeme číselné nezávislé proměnné.

Tlačítko **Options** je určeno k volbě linkové funkce (**Link**), nastavení iteračních kritérií (**Iterations**) a dalších parametrů. Tlačítkem **Output** se řídí výstup procedury, tj. tabulky zobrazené ve výstupovém okně (**Display**), výsledky uložené do datové matice jako nové proměnné (**Saved Variables**) a zobrazení logaritmu věrohodnosti modelu (**Print Log-Likelihood**). Pomocí tlačítek **Location** a **Scale** podrobně specifikujeme model.



**Ordinalní regrese**

### **Ordinalní regrese**

- *Analyze*
  - *Regression*
    - *Ordinal*



### Slovník anglických výrazů

<i>confidence interval</i>	<i>interval spolehlivosti</i>
<i>convergence</i>	<i>konvergence</i>
<i>covariate</i>	<i>kovariáta</i>
<i>dependent</i>	<i>závislý</i>
<i>display</i>	<i>zobrazit</i>
<i>equation</i>	<i>rovnice</i>
<i>estimation</i>	<i>odhad</i>
<i>factor</i>	<i>faktor, nezávislá proměnná</i>
<i>include</i>	<i>zahrnout</i>
<i>independent</i>	<i>nezávislý</i>
<i>iteration</i>	<i>iterace</i>
<i>likelihood</i>	<i>věrohodnost</i>
<i>method</i>	<i>metoda</i>
<i>plot</i>	<i>graf</i>
<i>predicted values</i>	<i>predikované hodnoty</i>
<i>regression</i>	<i>regrese</i>
<i>residual</i>	<i>residuální, reziduum</i>
<i>save</i>	<i>uložit</i>
<i>table</i>	<i>tabulka</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

## 1.7 Klasifikační metody (Classify)

### 1.7.1 Dvoustupňové seskupování (TwoStep Cluster)

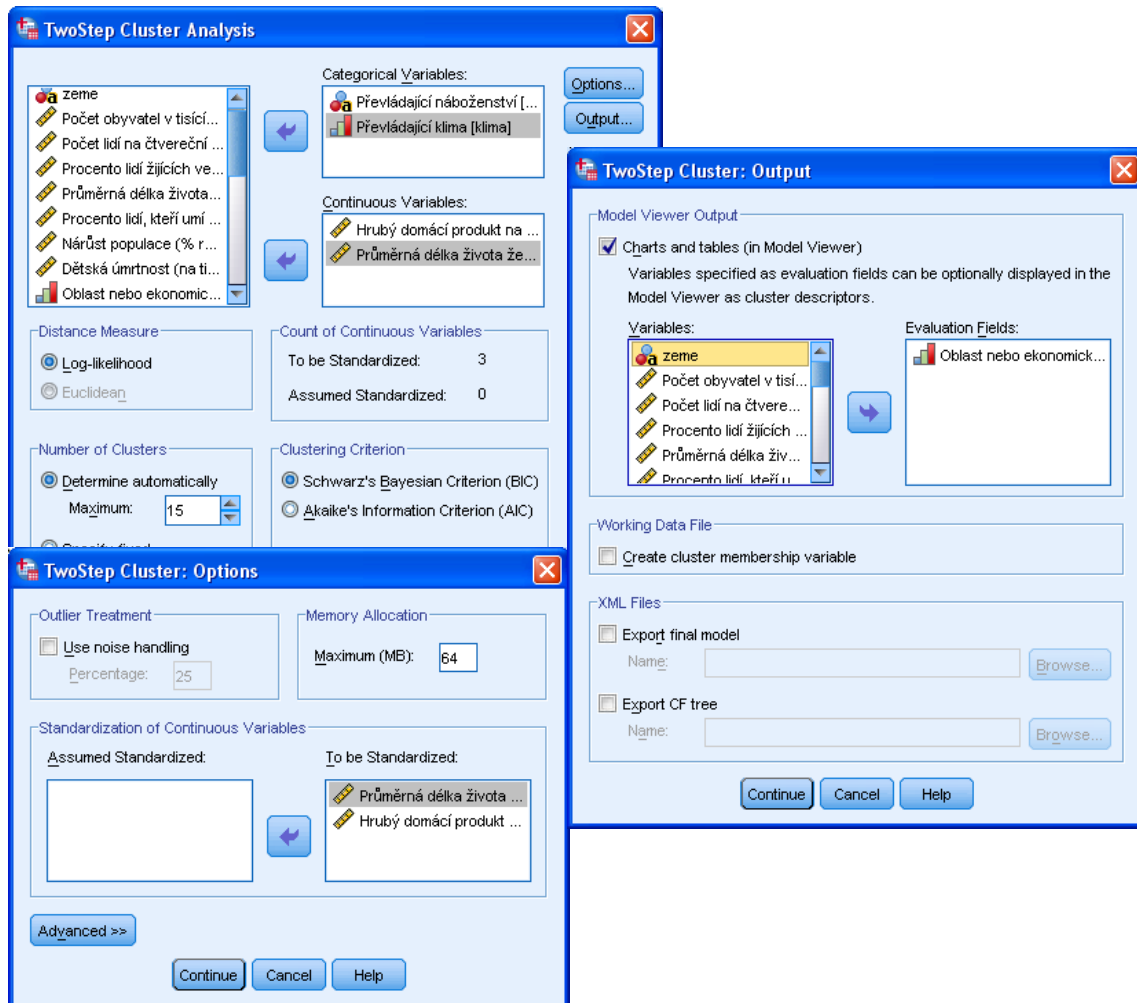
Procedura se užívá k rozdělení souboru (případů) do vnitřně co možná nejvíce homogenních skupin. Počet skupin, do kterých se jednotky rozdělí, můžeme stanovit sami nebo využít doporučení programu. Na rozdíl od ostatních klasifikačních modelů programu IBM SPSS Statistics je metoda TwoStep schopna pracovat s číselnými i kategorizovanými proměnnými.

Do pole **Categorical Variables** převedeme kategorizované vstupní proměnné, do pole **Continuous Variables** číselné proměnné (jeden z obou seznamů může být prázdný). V sekci **Distance Measure** zvolíme míru vzdálenosti: logaritmicko-věrohodnostní nebo euklidovskou (k dispozici pouze pro číselné proměnné).

V sekci **Number of Clusters** specifikujeme počet skupin, které se vytvoří. Počet může být zadán uživatelem nebo určen automaticky programem (zadáva se pouze maximální přípustná hodnota). Pro automatickou volbu určíme v části **Clustering Criterion** kritérium, na základě něhož je odvozen optimální počet klastrů: kritérium BIC nebo AIC.

Tlačítkem **Options** zadáme, zda předpokládáme „šum“ (**Use noise handling**) a jaké maximální procento případů v jedné skupině smí být šum. Pole **Memory Allocation** určuje, kolik MB paměti se alokuje pro výpočet. V sekci **Standardization of Continuous Variables** lze vybrat proměnné, které budou před výpočtem standardizovány. Po stisku tlačítka **Advanced** je možné ještě zadat parametry datové struktury CFT a případně její načtení z formátu XML.

Tlačítko **Output** dovoluje v části **Model Viewer Output** zadat další doplňující proměnné (mimo vstupních), které se budou podílet na popisu klastrů (**Evaluation Fields**). V části **Working Data File** můžeme uložit do datové matice jako novou proměnnou číslo klastru (**Create cluster membership variable**) nebo exportovat model či datovou strukturu CFT do formátu XML (**XML Files**).



Metoda TwoStep

## Dvoustupňové seskupování

- Analyze
  - Classify
    - TwoStep Cluster

### Slovník anglických pojmů

<i>advanced</i> .....	<i>pokročilý, složitý</i>
<i>allocation</i> .....	<i>alokace</i>
<i>assumed</i> .....	<i>předpokládán</i>
<i>automatically</i> .....	<i>automaticky</i>
<i>branch</i> .....	<i>větev</i>
<i>categorical</i> .....	<i>zde: kategorizovaná</i>
<i>change</i> .....	<i>změna</i>

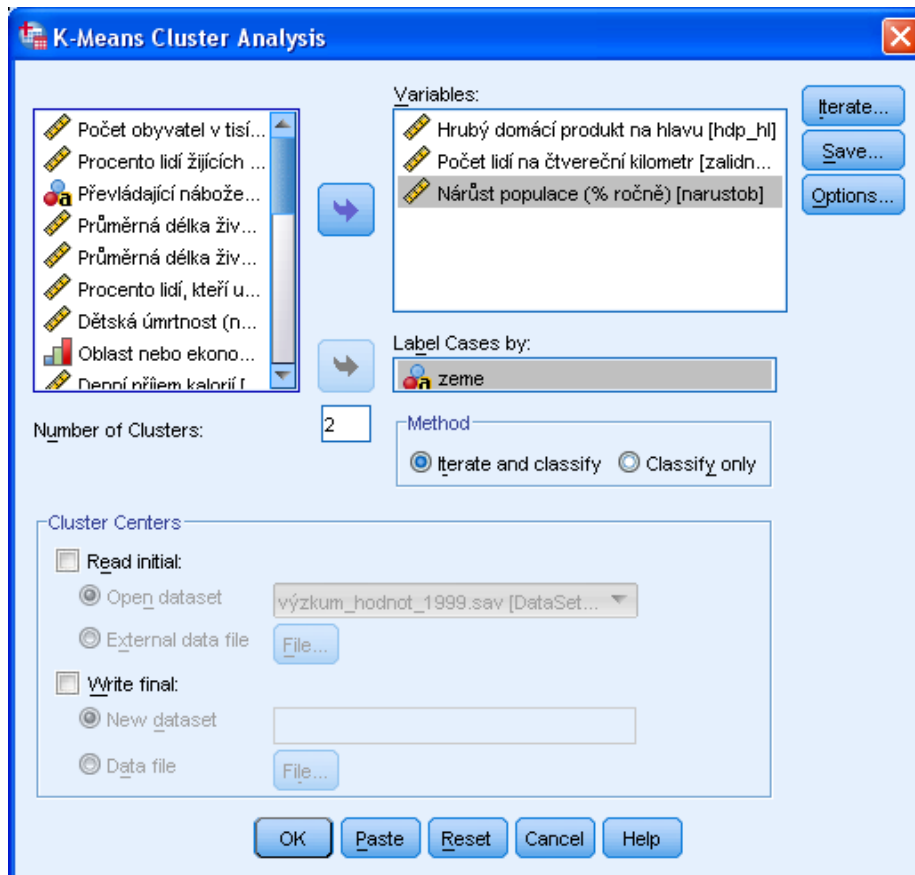
<i>Chi-square</i>	<i>chí kvadrát</i>
<i>classify</i>	<i>klasifikovat</i>
<i>cluster</i>	<i>skupina, seskupovací</i>
<i>confidence</i>	<i>spolehlivost</i>
<i>continuous</i>	<i>spojitý</i>
<i>count</i>	<i>počet</i>
<i>create</i>	<i>vytvořit</i>
<i>criterion</i>	<i>kritérium</i>
<i>depth</i>	<i>hloubka</i>
<i>descriptives</i>	<i>popisné statistiky</i>
<i>determine</i>	<i>stanovit</i>
<i>display</i>	<i>zobrazit</i>
<i>distance</i>	<i>vzdálenost</i>
<i>export</i>	<i>exportovat</i>
<i>Euclidean</i>	<i>euklidovský</i>
<i>file</i>	<i>soubor</i>
<i>final</i>	<i>konečný</i>
<i>fixed</i>	<i>pevný, fixní</i>
<i>frequencies</i>	<i>četnosti</i>
<i>handling</i>	<i>zacházení</i>
<i>importance</i>	<i>důležitost</i>
<i>insignificant</i>	<i>nevýznamný</i>
<i>initial</i>	<i>počáteční</i>
<i>leaf</i>	<i>list</i>
<i>level</i>	<i>úroveň</i>
<i>likelihood</i>	<i>věrohodnost</i>
<i>measure</i>	<i>měření, míra</i>
<i>membership</i>	<i>členství</i>
<i>memory</i>	<i>paměť</i>
<i>number</i>	<i>počet</i>
<i>node</i>	<i>uzel</i>
<i>noise</i>	<i>šum</i>
<i>none</i>	<i>žádný</i>
<i>omit</i>	<i>vynechat</i>
<i>options</i>	<i>volby</i>
<i>outlier</i>	<i>vzdálené pozorování</i>
<i>output</i>	<i>výstup</i>
<i>percentage</i>	<i>procento</i>
<i>pie</i>	<i>výšečový graf</i>
<i>plot</i>	<i>graf</i>
<i>possible</i>	<i>možný</i>
<i>rank</i>	<i>pořadí</i>
<i>save</i>	<i>uložit</i>
<i>significance</i>	<i>signifikance, význam</i>
<i>specify</i>	<i>určit, stanovit</i>
<i>standardized</i>	<i>standardizovaný</i>
<i>treatment</i>	<i>zacházení</i>
<i>tree</i>	<i>strom</i>
<i>threshold</i>	<i>práh</i>
<i>tuning</i>	<i>ladění</i>
<i>update</i>	<i>zde: načtení</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>within</i>	<i>uvnitř</i>

### 1.7.2 Seskupování metodou K-means (K-Means Cluster)

Metoda se užívá k rozdělení souboru (případů) do skupin na základě zvolených číselných charakteristik tak, aby si jednotky uvnitř skupin byly co nejvíce podobné. Jedná se o iterační metodu, která začíná určitým počátečním rozmístěním středů klastrů (souřadnice středů mohou být určeny algoritmem nebo načteny), které se postupně v jednotlivých krocích upravuje. V každém kroku se zároveň přehodnotí zařazení jednotek ke klastrům tak, aby každá jednotka patřila k nejbližšímu středu. Proces končí tehdy, pokud je splněno konvergenční kritérium nebo dosažen maximální počet iterací. Výhodou tohoto algoritmu je schopnost pracovat s relativně velkým počtem případů. Metoda vyžaduje předem zadat požadovaný počet skupin (klastrů). Pokud si však nejsme jistí, můžeme proceduru spustit vícekrát a porovnat různá řešení.

Do pole **Variables** převedeme analyzované proměnné. Přitom je nutné, aby se jednalo o číselné proměnné. Do pole **Label Cases by** můžeme zařadit proměnnou popisující případy. V poli **Number of Clusters** nastavíme požadovaný počet skupin. V sekci **Method** zadáváme, zda má algoritmus iterovat a poté klasifikovat (**Iterate and classify**), nebo pouze klasifikovat (**Classify only**). Druhá možnost se užívá především tehdy, pokud již máme klastry vytvořené a je třeba k nim pouze přiřadit nové případy. Počáteční středy (centroidy) skupin lze načíst ze souboru (**Read initial**) nebo naopak zapsat do souboru souřadnice výsledných centroidů (**Write final**).

Tlačítkem **Iterate** nastavíme maximální počet iterací a konvergenční kritéria (pouze pokud je zvolena iterační metoda). Tlačítkem **Save** uložíme číslo skupiny, do níž případ patří (**Cluster membership**) a vzdálenost jednotky od centroidu (**Distance from cluster center**). Tlačítko **Options** slouží k nastavení požadovaných výstupů (**Statistics**) (centroidy skupin, analýza rozptylu, informace o příslušnosti jednotlivých případů ke klastrům) a volbě způsobu práce s chybějícími hodnotami (**Missing Values**).



**Seskupování metodou K-means**

### **Seskupování metodou K-means**

- Analyze
  - Classify
    - K-Means Cluster Analysis

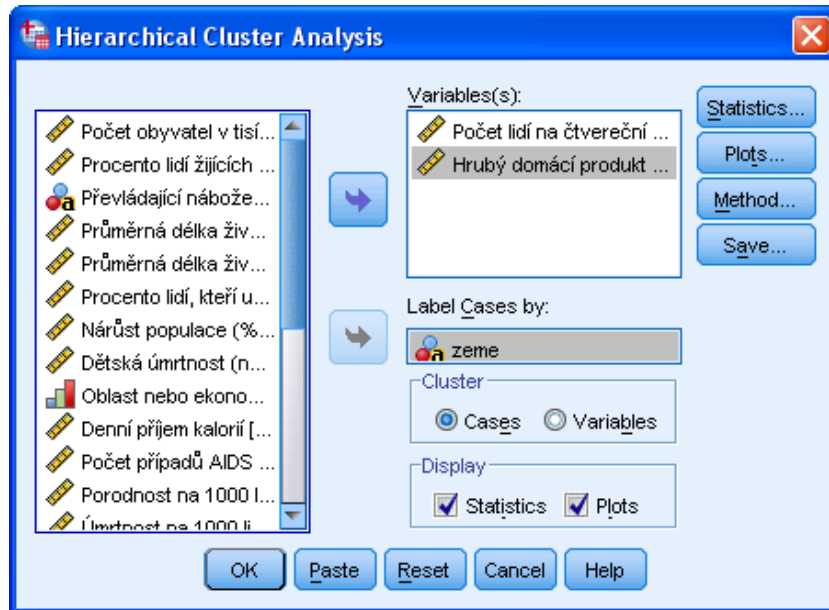
### Slovník anglických pojmů

<i>ANOVA</i>	<i>analýza rozptylu</i>
<i>case</i>	<i>případ</i>
<i>classify</i>	<i>klasifikovat</i>
<i>cluster</i>	<i>skupina, seskupovací</i>
<i>continue</i>	<i>pokračovat</i>
<i>file</i>	<i>soubor</i>
<i>final</i>	<i>konečný</i>
<i>label</i>	<i>popis, označení</i>
<i>listwise</i>	<i>celý případ</i>
<i>mean</i>	<i>průměr</i>
<i>membership</i>	<i>členství</i>
<i>missing</i>	<i>chybějící</i>
<i>number</i>	<i>počet</i>
<i>only</i>	<i>jen, pouze</i>
<i>options</i>	<i>volby</i>
<i>pairwise</i>	<i>po dvojicích</i>
<i>read</i>	<i>číst</i>
<i>save</i>	<i>uložit</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>
<i>write</i>	<i>psát</i>

### 1.7.3 Hierarchické seskupování (Hierarchical Cluster)

Metoda se užívá k rozdělení souboru (případů nebo proměnných) do skupin tak, aby si jednotky uvnitř skupin byly co nejvíce podobné. Je určena především pro menší datové soubory. Vstupní proměnné mohou být číselné, dichotomické nebo vyjadřovat četnosti (všechny proměnné by měly být stejného typu). Na začátku tvoří každá jednotka samostatnou skupinu. V každém kroku postupně dochází na základě jistých pravidel ke spojení dvou skupin, dokud všechny případy nepatří do jediné skupiny.

Do pole **Variable(s)** převedeme analyzované proměnné. Do pole **Label Cases by** můžeme zadat proměnnou popisující případy. Objekty seskupování určíme v sekci **Cluster**: případy (**Cases**), nebo proměnné (**Variables**). V části **Display** určíme, co se bude zobrazovat: statistiky (**Statistics**) nebo grafy (**Plots**). Tlačítko **Statistics** je určeno pro nastavení tabulkových výstupů, tlačítko **Plots** pro volbu grafů. Po stisku tlačítka **Method** lze zadat metodu seskupování (**Cluster Method**), míru vzdálenosti (**Measure**), transformaci hodnot (**Transform Values**) a transformaci míry vzdálenosti (**Transform Measure**). Tlačítko **Save** slouží k uložení čísla klastru do datové matice.



*Hierarchická seskupovací analýza*

### **Hierarchická seskupovací analýza**

- Analyze
  - Classify
    - Hierarchical Cluster Analysis



### Slovník anglických pojmů

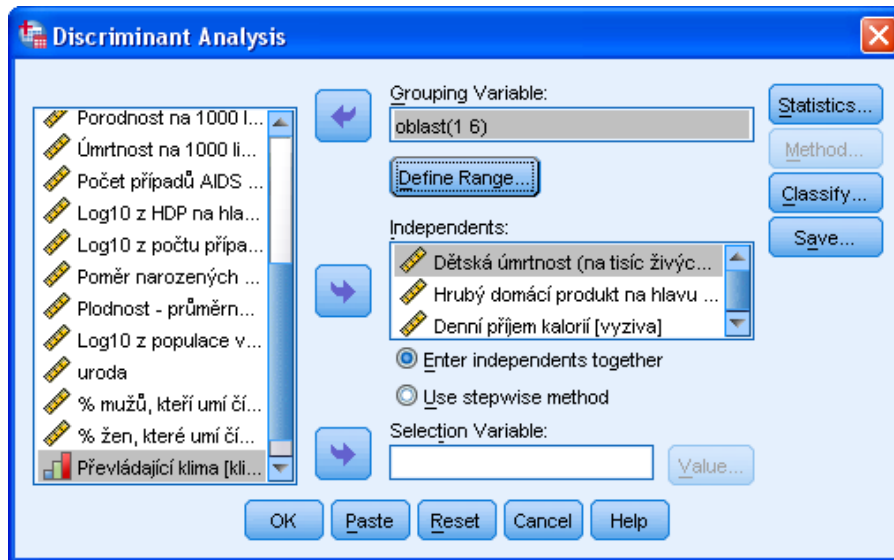
<i>case</i> .....	<i>případ</i>
<i>classify</i> .....	<i>klasifikovat</i>
<i>cluster</i> .....	<i>skupina, seskupovací</i>
<i>count</i> .....	<i>počet</i>
<i>display</i> .....	<i>zobrazit</i>
<i>distance</i> .....	<i>vzdálenost</i>
<i>hierarchical</i> .....	<i>hierarchický</i>
<i>label</i> .....	<i>popis, označení</i>
<i>matrix</i> .....	<i>matice</i>
<i>mean</i> .....	<i>průměr</i>
<i>measure</i> .....	<i>měření</i>
<i>membership</i> .....	<i>členství</i>
<i>none</i> .....	<i>žádný</i>
<i>plot</i> .....	<i>graf</i>
<i>range</i> .....	<i>rozpětí</i>
<i>save</i> .....	<i>uložit</i>
<i>start</i> .....	<i>začátek</i>
<i>stop</i> .....	<i>zastavení</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>

#### 1.7.4 Diskriminační analýza (Discriminant Analysis)

Diskriminační analýza je mnohorozměrná statistická metoda, která umožňuje diskriminovat (rozlišit) objekty ze dvou nebo více skupin na základě číselných proměnných a z nich odvozených kritérií. Užívá se k sestrojení prediktivních klasifikačních modelů, kdy na základě několika změřených proměnných rozhodujeme, ke které skupině určitá jednotka patří. Závislá proměnná je tedy nominální, nezávislé proměnné musí být číselné. Tuto techniku lze užít jak pro popis vztahů a komparační analýzu skupin, tak pro nalezení rozlišovacích pravidel a klasifikaci případů.

Do pole **Grouping Variable** přeneseme závislou proměnnou reprezentující skupiny. Hodnotami této proměnné musí být přirozená čísla, jejich rozsah zadáme tlačítkem **Define Range**. Případy mimo uvedený interval budou z dalších analýz vyloučeny. Nezávislé proměnné převedeme do pole **Independents**. Dále určíme, zda má být diskriminační funkce odhadnuta na základě všech zadaných nezávislých proměnných (**Enter independents together**), nebo zda chceme vybrat pouze ty, které hrají při klasifikaci významnou úlohu (**Use stepwise method**). Ve druhém případě je ještě nutné určit krokovou metodu (tlačítko **Method**), na základě které se výběr proměnných provede. Do pole **Selection Variable** můžeme zadat proměnnou, jejímiž hodnotami omezíme případy vstupující do analýzy (kritérium upřesníme tlačítkem **Value**).

Tlačítkem **Statistics** volíme zobrazované statistiky: popisné statistiky (**Descriptives**), koeficienty funkce (**Function Coefficients**), matice (**Matrices**). Tlačítkem **Classify** lze zadat apriorní pravděpodobnosti (**Prior Probabilities**), zobrazení výsledků (**Display**), nahrazení chybějících údajů průměrem (**Replace missing values with means**), použití kovarianční matice (**Use Covariance Matrix**) a grafy (**Plots**). Tlačítko **Save** umožňuje uložit predikce (predikované hodnoty, diskriminační skóry a pravděpodobnost predikované skupiny) do datového souboru jako nové proměnné a exportovat model do formátu XML.



*Diskriminační analýza*

### **Diskriminační analýza**

- Analyze
  - Classify
    - Discriminant Analysis

### Slovník anglických pojmů

<i>analysis</i>	.....	analýza
<i>classify</i>	.....	klasifikace
<i>compute</i>	.....	vypočítat
<i>define</i>	.....	definovat
<i>descriptives</i>	.....	popisné statistiky
<i>display</i>	.....	zobrazit
<i>distance</i>	.....	vzdálenost
<i>group</i>	.....	skupina
<i>independent</i>	.....	nezávislý
<i>matrix</i>	.....	matice
<i>mean</i>	.....	průměr
<i>membership</i>	.....	členství
<i>method</i>	.....	metoda
<i>missing</i>	.....	chybějící
<i>new</i>	.....	nový
<i>pairwise</i>	.....	párový
<i>probability</i>	.....	pravděpodobnost
<i>range</i>	.....	rozpětí
<i>replace</i>	.....	nahrazení
<i>save</i>	.....	uložit
<i>select</i>	.....	vybrat
<i>step</i>	.....	krok
<i>summary</i>	.....	sumární
<i>table</i>	.....	tabulka
<i>use</i>	.....	použít
<i>value</i>	.....	hodnota
<i>variable</i>	.....	proměnná
<i>variance</i>	.....	rozptyl

#### 1.7.5 Analýza nejbližších sousedů (Nearest Neighbor Analysis)

Analýza nejbližších sousedů je metoda klasifikace případů na základě jejich podobnosti s jinými případy.

Procedura předpokládá, že je správně nastaven způsob měření všech proměnných vstupujících do analýzy (datová matice, pohled na proměnné (*Variable View*), sloupec *Measure*).

Na záložce **Variables** definujeme proměnné vstupující do analýzy. V poli **Target** lze určit cílovou proměnnou, kterou má model předpovídat. Do pole **Features** zadáme proměnné, z nichž bude vycházet výpočet vzdálenosti případů. Pomocí zaškrtačovacího políčka **Normalize scale features** zvolíme, zda má být provedena standardizace vstupních číselných proměnných. V poli **Focal Case Identifier** můžeme určit proměnnou definující případy, které nás speciálně zajímají, tj. například nové případy, které mají být klasifikovány. Do pole **Case Label** lze zadat proměnnou popisující jednotlivé případy.

Na záložce **Neighbors** určíme, jakým způsobem budou nalezeny sousední případy. V části **Number of Nearest Neighbors (k)** definujeme počet sousedů, který může být určen pevně nebo nalezen automaticky na základě nejlepšího modelu pro různé hodnoty počtu sousedů. V části **Distance Computation** vybereme typ vzdálenosti pro určení podobnosti případů: euklidovská vzdálenost nebo manhattanská vzdálenost. Při zaškrtnutí políčka **Weight features by importance when computing**

**distances** budou vstupní proměnné při výpočtu vzdálenosti váženy na základě jejich normované důležitosti. V části **Prediction for Scale Target** volíme funkci pro výpočet predikce v případě, že cílová proměnná je číselná: průměr nebo medián nejbližších sousedů.

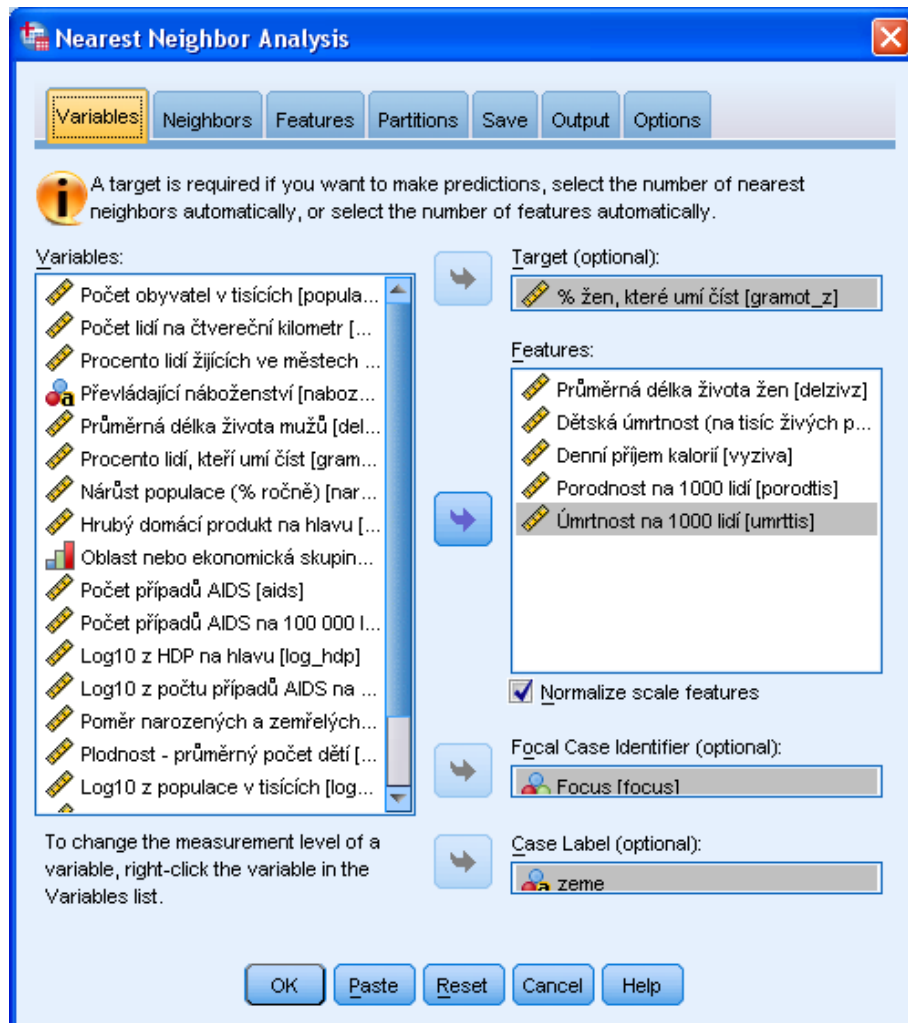
Na záložce **Features** lze zvolit automatický výběr vstupních proměnných a zadat požadavky na způsob výběru. Tyto možnosti jsou k dispozici pouze v případě, že je na záložce **Variables** definována cílová proměnná (**Target**) a alespoň jedna proměnná v poli **Features**. Zaškrtačací políčko **Perform feature selection** řídí, zda bude probíhat automatický výběr proměnných. V poli **Forward Selection** jsou proměnné, z nichž bude realizován výběr na základě metody **Forward**. Je-li však třeba zajistit, aby některé proměnné model určitě zahrnoval, přeneseme je do pole **Forced Entry**. V části **Stopping Criterion** definujeme kritérium pro vstup proměnných do modelu: zastavení přidávání proměnných, pokud již model obsahuje požadovaný počet vstupních proměnných nebo pokud se již model dostatečně nezlepšuje přidáním další proměnné.

Na záložce **Partitions** vybereme v části **Training and Holdout Partitions** způsob rozdělení datového souboru na trénovací a testovací množinu: náhodně (zadáme požadované procento případů v trénovací množině) nebo rozdělení na základě informace načtené ze zvolené proměnné. V oblasti **Cross Validation Folds** vybereme způsob rozdělení případů do skupin pro křížovou validaci: náhodně (zadáme požadovaný počet skupin) nebo na základě informace načtené ze zvolené proměnné. Volbou **Set seed for Mersenne Twister** a zadáním hodnoty **Seed** nastavíme počáteční hodnotu pro generátor náhodných čísel.

Na záložce **Save** definujeme, které proměnné mají být uloženy do datové matice a jakým způsobem. V části **Names of Saved Variables** určíme, zda mají být názvy nových proměnných vytvořeny automaticky nebo na základě zadání uživatele. V následující tabulce (**Variables to Save**) potom ve sloupci **Save** označíme požadované proměnné a případně upravíme jejich název ve sloupci **Variable of Root Name**. Uložit lze předpovídanou hodnotu nebo kategorii, předpovídanou pravděpodobnost pro všechny kategorie kategorizované cílové proměnné, informaci o zařazení do tréninkové nebo testovací množiny a informaci o zařazení do skupiny při křížové validaci.

Na záložce **Output** určíme v části **Viewer Output** požadovaný obsah výstupního objektu Model Viewer: tabulka s přehledem informací o analyzovaných případech a další tabulky a grafy. V části **Files** lze zvolit export celého modelu do formátu XML nebo uložit do datového souboru vzdálenosti případů ve středu zájmu (tj. případů, které jsou definované pomocí proměnné v poli **Focal Case Identifier** na záložce **Variables**) a jejich nejbližších sousedů.

Na záložce **Options** definujeme v části **User-Missing Values** způsob práce s uživatelem definovanými chybějícími hodnotami pro kategorizované proměnné.



**Analýza nejbližších sousedů**

### **Analýza nejbližších sousedů**

- Analyze
  - Classify
    - Nearest Neighbor

**Slovník anglických pojmů**

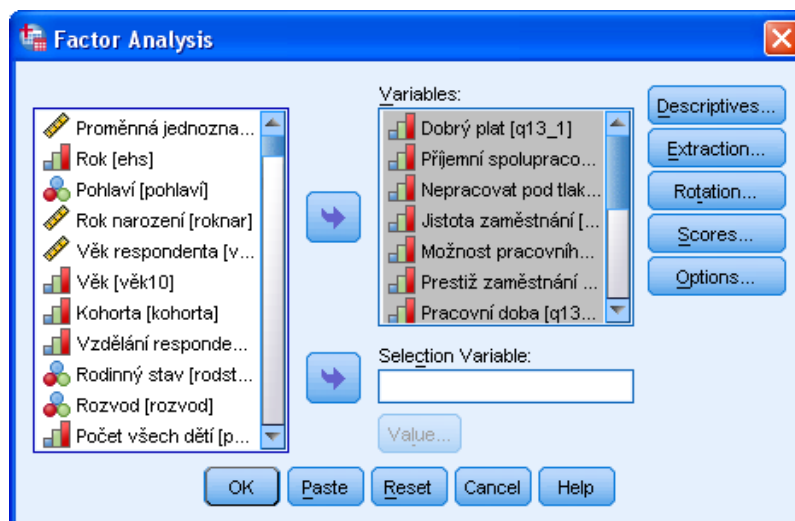
<i>analysis</i>	<i>analýza</i>
<i>display</i>	<i>zobrazit</i>
<i>distance</i>	<i>vzdálenost</i>
<i>mean</i>	<i>průměr</i>
<i>membership</i>	<i>členství</i>
<i>method</i>	<i>metoda</i>
<i>missing</i>	<i>chybějící</i>
<i>new</i>	<i>nový</i>
<i>probability</i>	<i>pravděpodobnost</i>
<i>range</i>	<i>rozpětí</i>
<i>save</i>	<i>uložit</i>
<i>select</i>	<i>vybrat</i>
<i>table</i>	<i>tabulka</i>
<i>use</i>	<i>použít</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

## 1.8 Redukce dat (Dimension Reduction)

### 1.8.1 Faktorová analýza (Factor Analysis)

Metoda se zabývá rozbořem struktury vzájemných závislostí proměnných na základě předpokladu, že tyto závislosti jsou důsledkem působení určitého menšího počtu neměřitelných, tzv. latentních, proměnných (faktorů). Cílem faktorové analýzy je nalézt a interpretovat společné faktory, které co nejjednodušeji objasní závislosti mezi proměnnými. Zároveň by měl být počet faktorů co nejmenší a faktory by měly co nejlépe vysvětlovat variabilitu sledovaných proměnných.

Analyzované proměnné převedeme do pole **Variables**. Pole **Selection Variable** využijeme, jestliže chceme pracovat pouze s některými případy v souboru. Tlačítkem **Descriptives** lze zobrazit ve výstupu další statistiky (**Statistics**) a matice (**Correlation Matrix**). Tlačítko **Extraction** umožňuje zvolit metodu extrakce (**Method**), analýzu korelační nebo kovariační matice (**Analyze**), zobrazení faktorů nebo grafu vlastních hodnot (**Display**), postup extrakce faktorů (**Extract**) a maximální počet iterací (**Maximum Iterations for Convergence**). Tlačítkem **Rotation** nastavíme metodu rotace (**Method**), zobrazení faktorů nebo grafu faktorových zátěží (**Display**) a maximální počet iterací (**Maximum Iterations for Convergence**). Tlačítkem **Scores** uložíme faktorové skóry (**Save as variables**), přitom volíme metodu jejich odhadu (**Method**), a zobrazíme jejich transformační matici (**Display factor score coefficient matrix**). Po stisku tlačítka **Options** lze vybrat způsob zacházení s chybějícími hodnotami (**Missing Values**) a formát zobrazení koeficientů (**Coefficient Display Format**).



Faktorová analýza

### Faktorová analýza

- Analyze
  - Dimension Reduction
    - Factor

**Slovník anglických pojmů**

<i>correlation</i>	<i>korelace</i>
<i>display</i>	<i>zobrazit</i>
<i>exclude</i>	<i>vyločit</i>
<i>level</i>	<i>úroveň</i>
<i>matrix</i>	<i>matice</i>
<i>mean</i>	<i>průměr</i>
<i>missing</i>	<i>chybějící</i>
<i>number</i>	<i>počet</i>
<i>plot</i>	<i>graf</i>
<i>replace</i>	<i>nahradit</i>
<i>rotation</i>	<i>rotace</i>
<i>save</i>	<i>uložit</i>
<i>significance</i>	<i>významnost</i>
<i>sorted by size</i>	<i>seřadit podle velikosti</i>
<i>suppress</i>	<i>potlačit</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

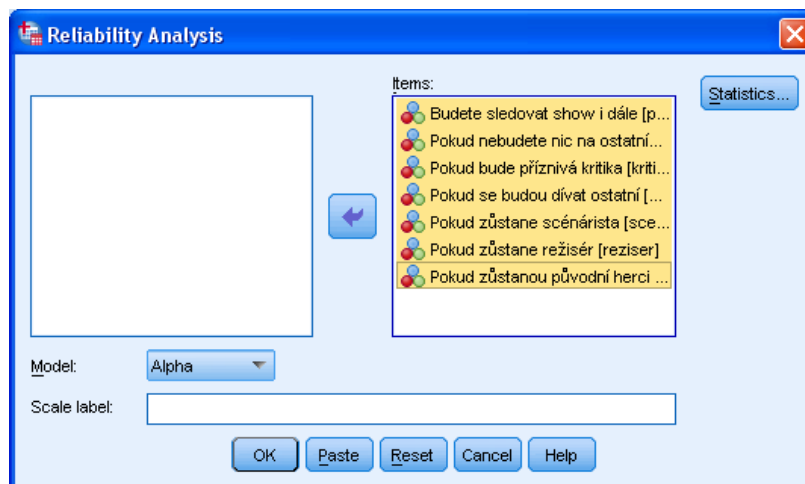


## 1.9 Škálování (Scale)

### 1.9.1 Analýza reliability (Reliability Analysis)

Analýza reliability studuje vlastnosti škály tvořené součtem (průměrem) několika položek. Procedura nabízí řadu klasických měr reliability a zároveň poskytuje informace o vzájemném vztahu individuálních položek i o jejich vztahu s celkovým skóre. Vnitrotřídní korelační koeficient měří homogenitu hodnocení různých hodnotitelů.

Vstupní proměnné (položky škály) převedeme do pole **Items**. V rozevíracím seznamu **Model** zvolíme jeden z nabízených modelů pro hodnocení míry vnitřní konzistence položek. Do pole **Scale label** lze zadat popis škály. Pomocí tlačítka **Statistics** nastavíme požadované výstupy: popisné statistiky položek, škály nebo škály po vyloučení jedné položky (**Descriptives for**), statistiky průměrů, rozptylů, kovariancí a korelací položek (**Summaries**), kovarianční či korelační matice položek (**Inter-Item**), tabulka analýzy rozptulu (**ANOVA Table**) a další.



*Analýza reliability*

### **Analýza reliability**

- Analyze
  - Scale
    - Reliability Analysis

### 1.9.2 Mnohorozměrné škálování (Multidimensional Scaling) – ALSCAL

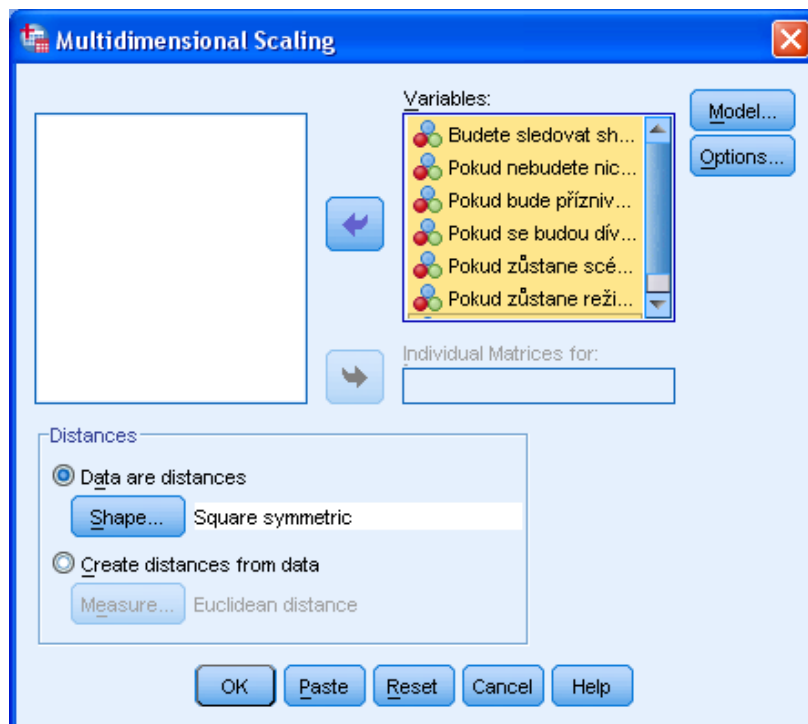
Pomocí mnohorozměrného škálování zobrazíme analyzované objekty (proměnné nebo případy) v rovině (nebo v prostoru vyšší dimenze) tak, aby se co možná nejlépe zachovaly vztahy mezi nimi (vzdálenosti či jiné míry nepodobnosti). Tímto způsobem získáme „mapu“, v níž jsou objekty

znázorněny. Z metod mnohorozměrného škálování obsahuje modul IBM SPSS Statistics Base pouze proceduru **ALSCAL**, procedury **PROXSCAL** a **PREFSCAL** patří do modulu *IBM SPSS Categories* a jsou k dispozici pouze tehdy, je-li nainstalován také tento modul.

Do pole **Variables** přeneseme vstupní proměnné. V případě, že se vzdálenosti počítají z dat, lze do pole **Individual Matrices for** zadat proměnnou, jejíž kategorie rozštěpí soubor na skupiny. Pro každou z nich se potom počítá matice vzdáleností a realizuje výpočet samostatně.

V části **Distances** zvolíme způsob odvození vzdáleností z dat. Pokud data představují přímo vzdálenosti, uijeme tlačítko **Shape** k určení typu matice vzdáleností: čtvercová symetrická (Square symmetric), čtvercová asymetrická (Square asymmetric) nebo obdélníková (Rectangular). Je-li třeba vzdálenosti spočítat z dat, nastavíme tlačítkem **Measure** vhodnou metriku pro daný typ proměnných, v případě potřeby zadáme požadovanou transformaci hodnot před výpočtem vzdáleností (Transform Values) a určíme, zda se mají spočítat vzdálenosti mezi řádky nebo sloupci datové matice (Create Distance Matrix).

Tlačítkem **Model** specifikujeme detaily modelu, pomocí tlačítka **Options** nastavíme požadované výstupy (Display) a kritéria konvergence (Criteria).



**Mnohorozměrné škálování**

## **Mnohorozměrné škálování**

- Analyze
  - Scale
    - Multidimensional Scaling

### **Slovník anglických pojmů**

<i>case</i> .....	<i>případ</i>
<i>change</i> .....	<i>změnit</i>
<i>close</i> .....	<i>zavřít</i>
<i>column</i> .....	<i>sloupec</i>
<i>consistency</i> .....	<i>konzistence</i>
<i>continue</i> .....	<i>pokračovat</i>
<i>correlation</i> .....	<i>korelace</i>
<i>define</i> .....	<i>zadat, definovat</i>
<i>descriptive</i> .....	<i>popisný</i>
<i>distance</i> .....	<i>vzdálenost</i>
<i>item</i> .....	<i>položka</i>
<i>label</i> .....	<i>popis, označení</i>
<i>lower</i> .....	<i>dolní</i>
<i>measure</i> .....	<i>míra</i>
<i>missing</i> .....	<i>chybějící</i>
<i>mixed</i> .....	<i>tříděný</i>
<i>multidimensional</i> .....	<i>mnohorozměrný</i>
<i>multiple</i> .....	<i>mnohonásobný</i>
<i>name</i> .....	<i>jméno</i>
<i>options</i> .....	<i>předvolby</i>
<i>probability</i> .....	<i>pravděpodobnost, signifikance</i>
<i>random</i> .....	<i>náhodný</i>
<i>remove</i> .....	<i>odstranit</i>
<i>scaling</i> .....	<i>škálování</i>
<i>shape</i> .....	<i>typ</i>
<i>strictly</i> .....	<i>přísně</i>
<i>two-way</i> .....	<i>dvojný</i>
<i>row</i> .....	<i>řádek</i>
<i>set</i> .....	<i>sada</i>
<i>table for</i> .....	<i>tabulka pro</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>

## 1.10 Neparametrické testy (Nonparametric Tests)

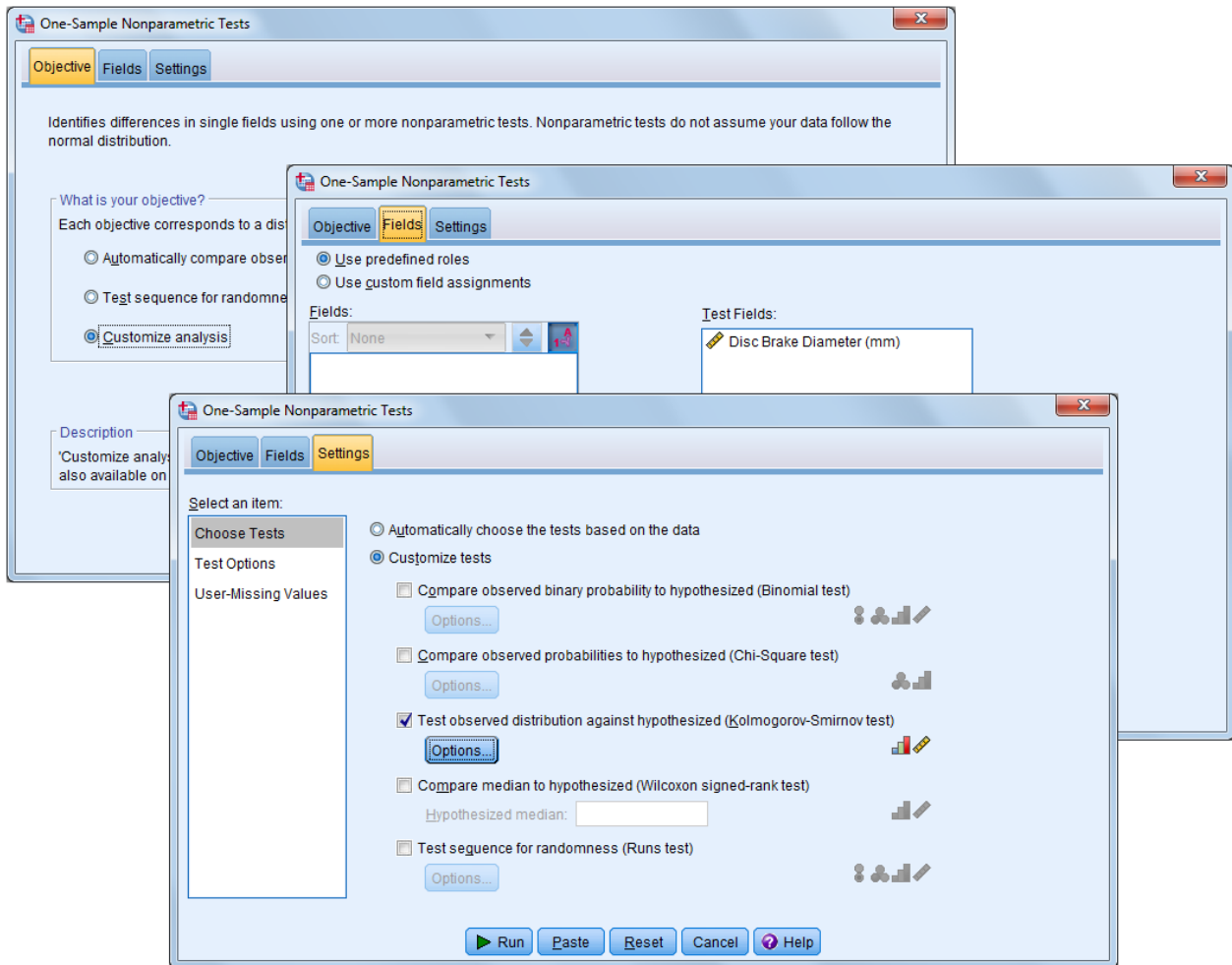
### 1.10.1 Jednovýběrové neparametrické testy (One Sample)

Procedura **One-Sample Nonparametric Tests** nabízí jednovýběrové neparametrické testy. K dispozici je binomický test, test Chí-kvadrát, jednovýběrový Kolmogorov-Smirnovův test, jednovýběrový Wilcoxonův znaménkový test a test pořadí. Výstup procedury je ve formě interaktivního objektu *Model Viewer*.

Na záložce **Objective** určíme cíl testování: automaticky porovnat rozložení dat s předpokládaným teoretickým rozložením (**Automatically compare observed data to hypothesized**), automaticky otestovat náhodnost výskytu hodnot (**Test sequence for randomness**), nebo zadat vlastní volbu testů (**Customize analysis**).

Na záložce **Fields** zadáme do pole **Test Fields** testované proměnné, což provedeme buď ručně, nebo označíme **Use predefined roles** a program automaticky vybere ty proměnné, u nichž atribut *Role* (datová matice, záložka *Variable View*) nabývá hodnot *Input*, *Target* nebo *Both*.

Na záložce **Settings** specifikujeme požadované testy. Při volbě **Automatically choose the tests based on the data** bude automaticky vybrán nejvhodnější test na základě cíle testování a typu proměnné. Při označení **Customize tests** můžeme zadat testy ručně. Tlačítkem **Options** provedeme detailní nastavení pro daný test. V části **Test Options** dále určíme požadovanou hladinu významnosti a interval spolehlivosti, v části **User-Missing Values** způsob zacházení s chybějícími hodnotami.



**Jednovýběrové neparametrické testy**

**Jednovýběrové neparametrické testy**

- Analyze
  - Nonparametric Tests
    - One Sample

### Slovník anglických výrazů

<i>all</i> .....	vše
<i>custom</i> .....	vlastní
<i>data</i> .....	data
<i>define</i> .....	definovat, zadat
<i>descriptive</i> .....	popisný
<i>equal</i> .....	rovny
<i>exclude</i> .....	vyločit
<i>expected</i> .....	očekávaný
<i>get from</i> .....	odečíst z
<i>list</i> .....	seznam
<i>mean</i> .....	průměr
<i>missing</i> .....	chybějící
<i>nonparametric</i> .....	neparametrický
<i>options</i> .....	volby
<i>point</i> .....	bod
<i>proportion</i> .....	proporce
<i>quartiles</i> .....	kvartily
<i>range</i> .....	rozpětí
<i>specified</i> .....	specifikovaný
<i>test</i> .....	testovaný
<i>use</i> .....	použit
<i>value</i> .....	hodnota
<i>variable</i> .....	proměnná

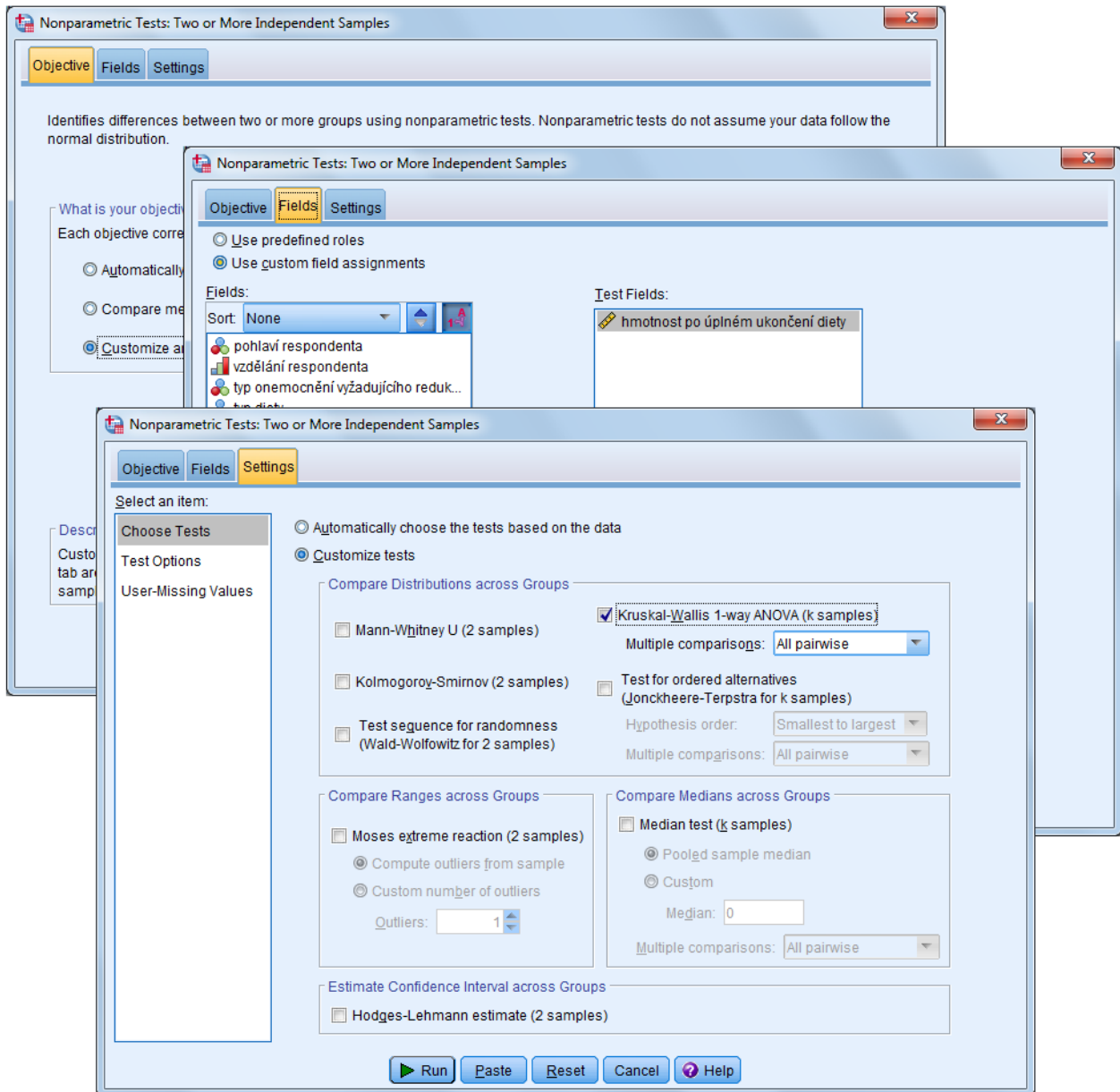
#### 1.10.2 Neparametrické testy pro nezávislé výběry (Independent Samples)

Procedura **Nonparametric Tests: Two or More Independent Samples** je určena k provedení neparametrických testů pro dva nebo více nezávislých výběrů (výběry jsou definovány kategoriemi jedné kategorizované proměnné). K dispozici jsou následující testy: Mann-Whitneyův U test, Kolmogorov-Smirnovův test, Wald-Wolfowitzův test pořadí, Kruskal-Wallisův test, Jonckheere-Terpstrův test, Mosesův test a mediánový test. Výstup procedury je ve formě objektu *Model Viewer*.

Na záložce **Objective** zvolíme cíl testování: automaticky porovnat rozložení dat ve skupinách (**Automatically compare distributions across groups**), porovnat mediány ve skupinách (**Compare medians across groups**) nebo zadat vlastní volbu testů (**Customize analysis**).

Na záložce **Fields** přeneseme do pole **Test Fields** testované proměnné, do pole **Groups** kategorizovanou proměnnou určující skupiny.

Na záložce **Settings** jsou k dispozici testy pro porovnání rozložení ve dvou nebo více skupinách (**Compare Distributions across Groups**), test pro porovnání rozpětí ve dvou skupinách (**Compare Ranges across Groups**) a test pro porovnání mediánů ve dvou nebo více skupinách (**Compare Medians across Groups**). Pro dva nezávislé výběry lze dále vypočítat interval spolehlivosti pro rozdíl mediánů (**Estimate Confidence Interval across Groups**). U testů pro více výběrů je možné zadat mnohonásobná srovnání výběrů.



**Neparametrické testy pro nezávislé výběry**

### Neparametrické testy pro nezávislé výběry

- Analyze
  - Nonparametric Tests
    - Independent Samples

### Slovník anglických pojmů

<i>define</i> .....	<i>zadat, definovat</i>
<i>distribution</i> .....	<i>rozdělení</i>
<i>group</i> .....	<i>skupina</i>
<i>independent</i> .....	<i>nezávislý</i>
<i>list</i> .....	<i>seznam</i>
<i>missing</i> .....	<i>chybějící</i>
<i>nonparametric</i> .....	<i>neparametrický</i>
<i>normal</i> .....	<i>normální</i>
<i>options</i> .....	<i>volby</i>
<i>sample</i> .....	<i>výběr</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>

### 1.10.3 Neparametrické testy pro závislé výběry (Related Samples)

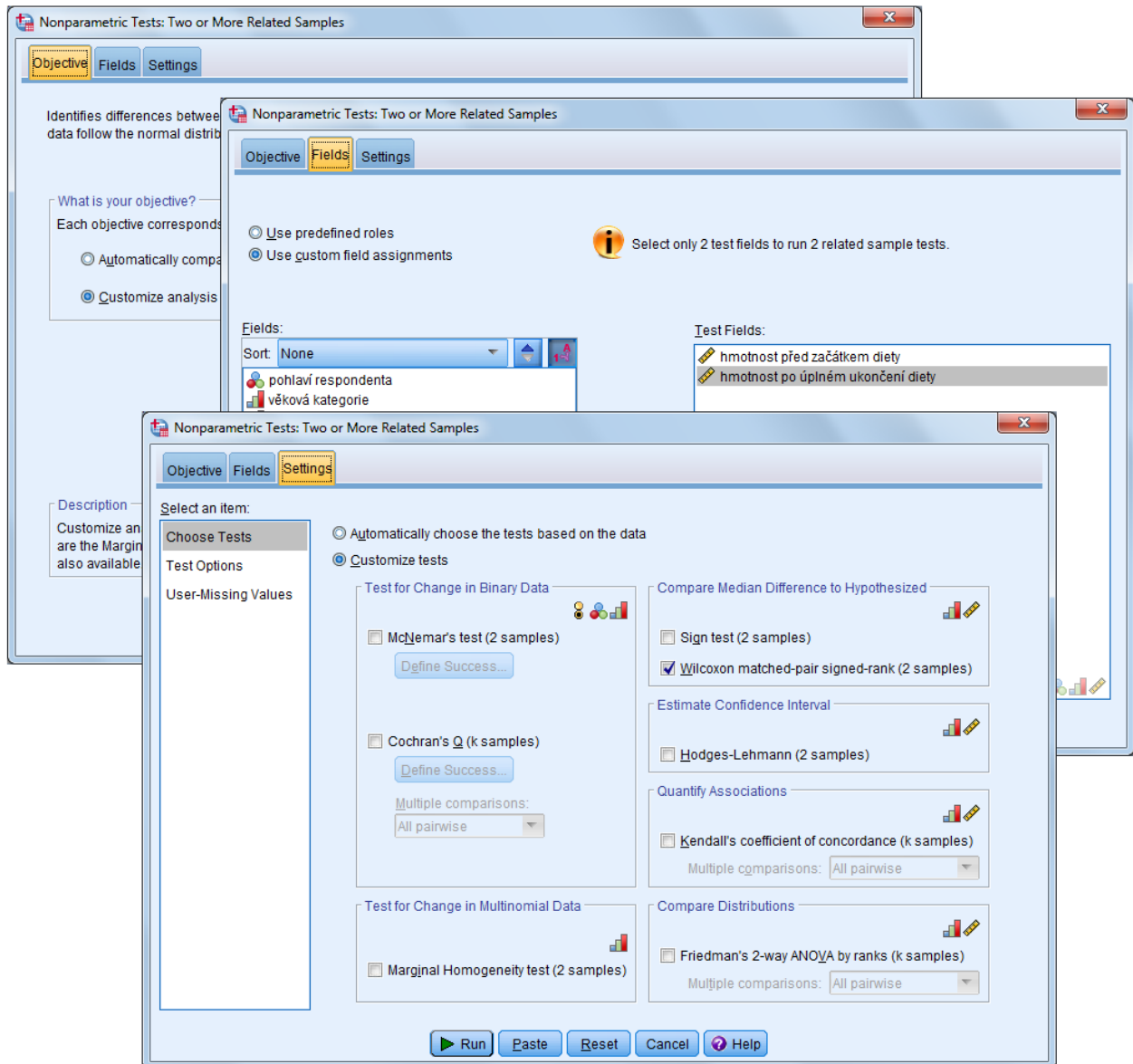
Procedura **Nonparametric Tests: Two or More Related Samples** nabízí neparametrické testy pro porovnání dvou nebo více závislých výběrů. K dispozici jsou následující testy: McNemarův test, Cochranův Q test, test marginální homogenity, znaménkový test, Wilcoxonův test, Kendallův test a Friedmanův test. Výstup procedury je ve formě objektu *Model Viewer*.

Na záložce **Objective** určíme, zda budou vhodné testy vybrány automaticky (**Automatically compare observed data to hypothesized**) nebo na základě vlastního nastavení (**Customize analysis**).

Na záložce **Fields** specifikujeme testované proměnné (dvě nebo více), které zadáme do pole **Test Fields**.

Na záložce **Settings** jsou k dispozici testy pro binární či multinomická data (**Test for Change in Binary/Multinomial Data**), testy pro porovnání mediánů (**Compare Median Difference to Hypothesized**), test síly asociace (**Quantity Associations**) a test pro srovnání rozložení ve výběrech (**Compare Distributions**). Pro dva závislé výběry lze dále vypočítat interval spolehlivosti pro medián rozdílu proměnných (**Estimate Confidence Interval**) a provést mnohonásobná srovnání pro více závislých výběrů.





**Neparametrické testy pro závislé výběry**

### Neparametrické testy pro závislé výběry

- Analyze
  - Nonparametric Tests
    - Related Samples

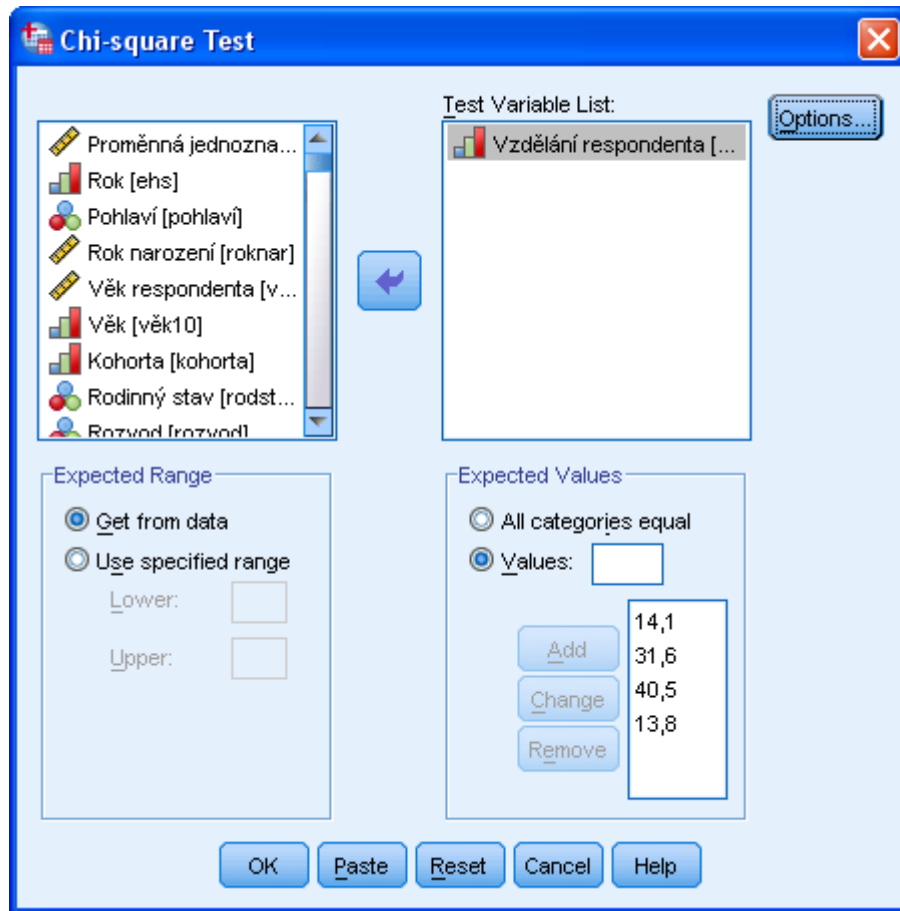
### Slovník anglických pojmů

<i>current</i> .....	<i>běžný, aktuální</i>
<i>define</i> .....	<i>zadat, definovat</i>
<i>list</i> .....	<i>seznam</i>
<i>options</i> .....	<i>volby</i>
<i>paired</i> .....	<i>párový</i>
<i>sample</i> .....	<i>výběr, soubor</i>
<i>selection</i> .....	<i>selekce, výběr</i>
<i>variable</i> .....	<i>proměnná</i>

#### 1.10.4 Test Chí-kvadrát (Chi-square Test)

Test Chí-kvadrát se užívá k posouzení, zda rozložení četností kategorizované proměnné odpovídá očekávanému rozložení kategorií (například při ověřování reprezentativnosti výběrového šetření).

Testovanou proměnnou (proměnné) převedeme do okna **Test Variable List**. V sekci **Expected Values** specifikujeme očekávané rozdělení pravděpodobností. Při označení **All categories equal** testujeme, zda jsou všechny kategorie zastoupeny rovnoměrně. Varianta **Values** ověřuje, zda rozložení četností odpovídá zadaným hodnotám očekávaných pravděpodobností. Do okna **Values** potom postupně zadáme očekávané pravděpodobnosti jednotlivých kategorií (nebo čísla jim úměrná tak, aby byly zachovány poměry) v pořadí kódování kategorií. V sekci **Expected Range** rozhodujeme, zda se test provede pro všechny kategorie dané proměnné (**Get from data**), nebo pouze pro kategorie vymezené dolní a horní hranicí intervalu (**Use specified range**). Tlačítko **Options** je určeno k nastavení požadovaných popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



*Test dobré shody*

### **Test Chí-kvadrát**

- *Analyze*
  - *Nonparametric Tests*
    - *Legacy Dialogs*
      - *Chi-Square*

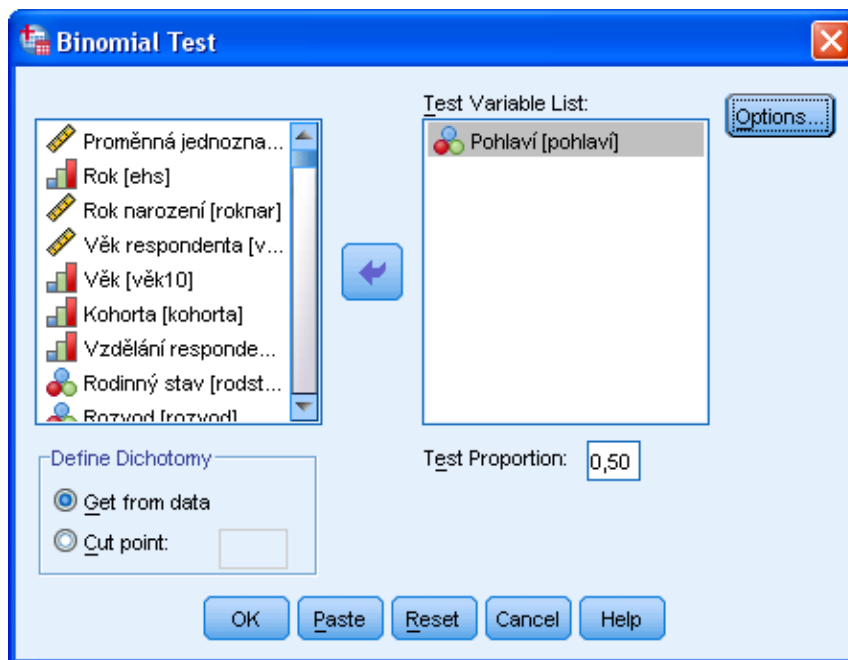
### Slovník anglických výrazů

<i>all</i> .....	vše
<i>data</i> .....	<i>data</i>
<i>define</i> .....	definovat, zadat
<i>descriptive</i> .....	popisný
<i>equal</i> .....	rovny
<i>exclude</i> .....	vyločit
<i>expected</i> .....	očekávaný
<i>get from</i> .....	odečíst z
<i>list</i> .....	seznam
<i>nonparametric</i> .....	neparametrický
<i>options</i> .....	volby
<i>point</i> .....	bod
<i>quartiles</i> .....	kvartily
<i>range</i> .....	rozpětí
<i>specified</i> .....	specifikovaný
<i>test</i> .....	testovaný
<i>use</i> .....	použit
<i>value</i> .....	hodnota
<i>variable</i> .....	proměnná

#### 1.10.5 Binomický test (Binomial Test)

Metoda testuje, zda pozorované četnosti dichotomické proměnné (proměnná nabývá pouze dvou hodnot) odpovídají očekávaným četnostem za předpokladu binomického rozdělení se zadaným parametrem pravděpodobnosti.

Testovanou proměnnou (proměnné) převedeme do okna **Test Variable List**. V okně **Test Proportion** zadáme pravděpodobnost výskytu první kategorie (pravděpodobnost pro druhou kategorii se spočítá automaticky). V sekci **Define Dichotomy** určíme typ testované proměnné. Pro dichotomickou proměnnou je určena volba **Get from data**, pro jiné než dichotomické proměnné volba **Cut point** (zadáme číslo oddělující hodnoty proměnné v první kategorii od hodnot odpovídajících druhé kategorii). Tlačítko **Options** slouží k výběru popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



*Binomický test*

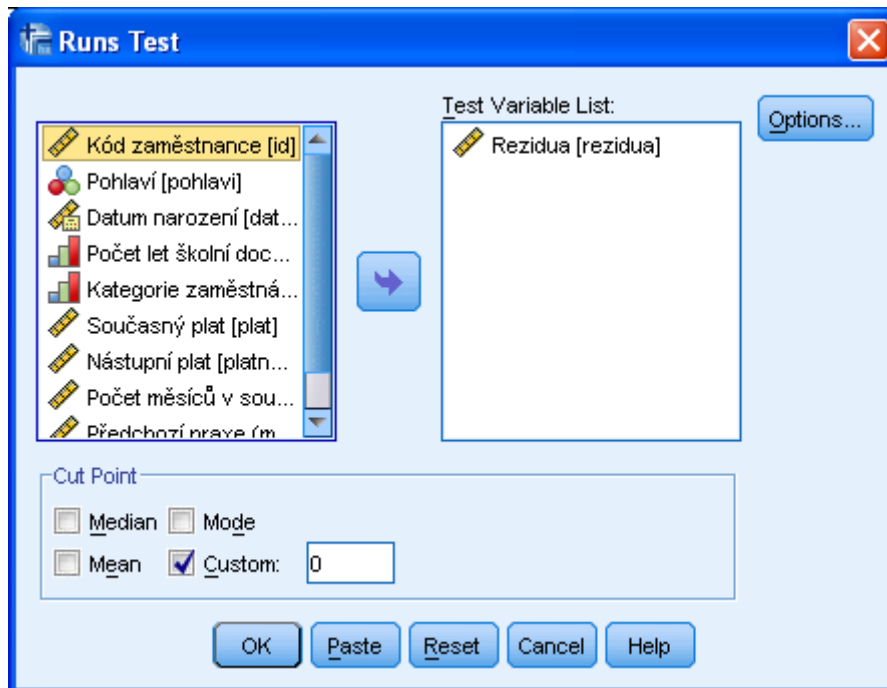
### **Binomický test**

- *Analyze*
  - *Nonparametric Tests*
    - *Legacy Dialogs*
      - *Binomial*

#### **1.10.6 Test pořadí (Runs Test)**

Metoda zjišťuje, zda je pořadí výskytu dvou hodnot proměnné náhodné.

Testovanou proměnnou (proměnné) převedeme do okna **Test Variable List**. V sekci **Cut Point** zvolíme hodnotu, která rozdělí data na dvě části a definuje kategorie, pro které bude testování probíhat. Lze zadat medián (**Median**), modus (**Mode**), průměr (**Mean**) nebo vlastní hodnotu (**Custom**). Tlačítko **Options** slouží k volbě popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



**Test pořadí**

### Test pořadí

- Analyze
  - Nonparametric Tests
    - Legacy Dialogs
      - Runs

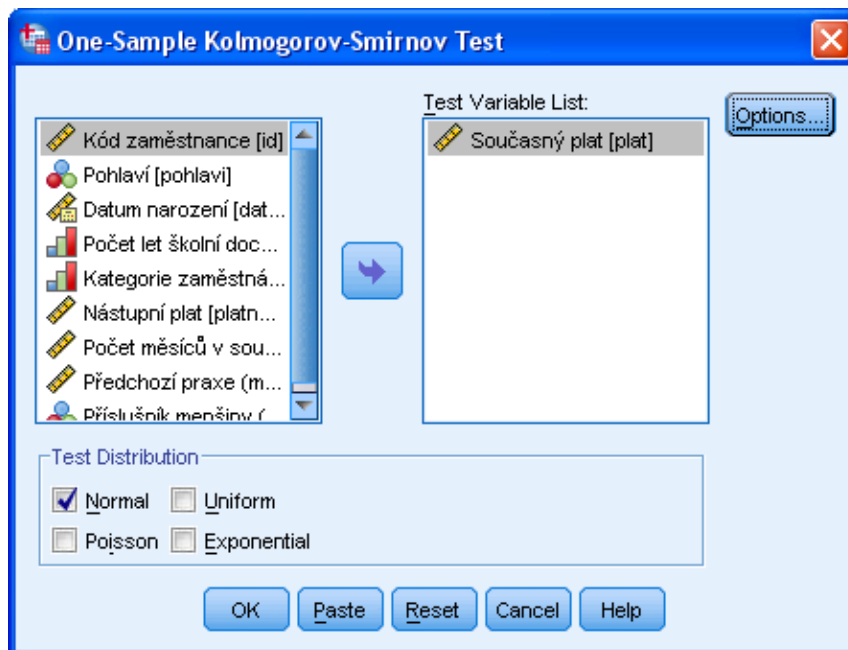
### Slovník anglických pojmů

<i>continue</i>	<i>pokračovat</i>
<i>custom</i>	<i>vlastní</i>
<i>define</i>	<i>definovat, zadat</i>
<i>equal</i>	<i>rovný, roven</i>
<i>get from</i>	<i>odečíst z</i>
<i>list</i>	<i>seznam</i>
<i>mean</i>	<i>průměr</i>
<i>missing</i>	<i>chybějící</i>
<i>nonparametric</i>	<i>neparametrický</i>
<i>options</i>	<i>volby</i>
<i>point</i>	<i>bod</i>
<i>proportion</i>	<i>proporce</i>
<i>test</i>	<i>test</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

### 1.10.7 Jednovýběrový Kolmogorov-Smirnovův test (One-Sample Kolmogorov-Smirnov test)

Metoda ověřuje, zda rozložení číselné proměnné odpovídá očekávanému teoretickému rozdělení (k dispozici je normální, rovnoměrné, Poissonovo nebo exponenciální).

Testovanou proměnnou (proměnné) převedeme do okna **Test Variable List**. V sekci **Test Distribution** volíme testované rozdělení: normální (**Normal**), rovnoměrné (**Uniform**), Poissonovo (**Poisson**) nebo exponenciální (**Exponential**). Tlačítko **Options** je určeno k výběru popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



*Kolmogorov-Smirnovův test*

### **Jednovýběrový Kolmogorov-Smirnovův test**

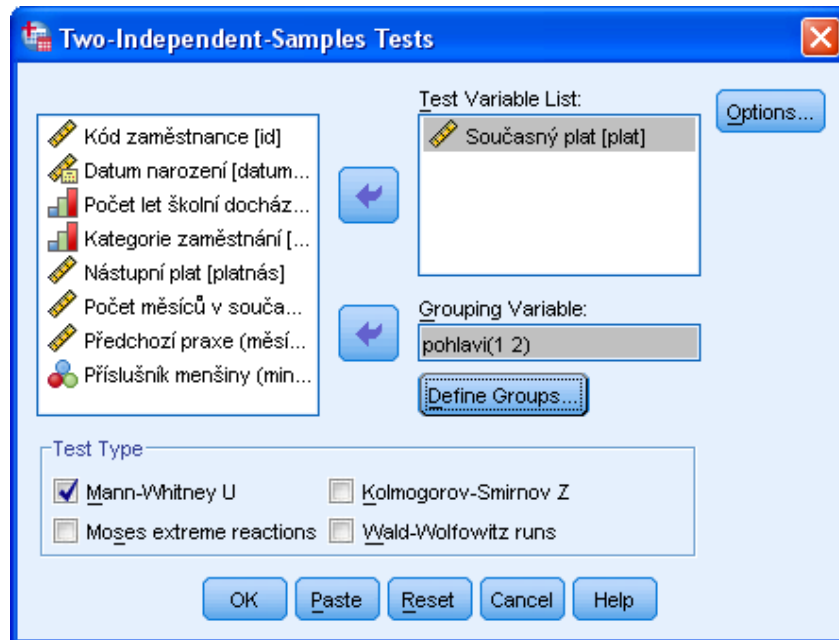
- Analyze
  - Nonparametric Tests
    - Legacy Dialogs
      - 1-Sample K-S

### 1.10.8 Testy pro dva nezávislé výběry (Two-Independent-Samples Tests)

Procedura zahrnuje čtyři testy (Mann-Whitneyův U test, Mosesův test, Kolmogorov-Smirnovův Z test a Waldův-Wolfowitzův test pořadí), které jsou neparametrickou obdobou dvouvýběrového t-testu, tj. porovnávají rozdělení hodnot ve dvou nezávislých skupinách.

Testovanou proměnnou (proměnné) převedeme do okna **Test Variable List**. V poli **Grouping Variable** zadáme proměnnou určující dvě nezávislé skupiny. Tlačítko **Define Groups** slouží k zadání hodnot, kterými definujeme první a druhou skupinu. V sekci **Test Type** zadáme jeden nebo více testů. Tlačítko

**Options** slouží k výběru popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



*Test dvou nezávislých výběrů*

### **Testy pro dva nezávislé výběry**

- *Analyze*
  - *Nonparametric Tests*
    - *Legacy Dialogs*
      - *2 Independent Samples*



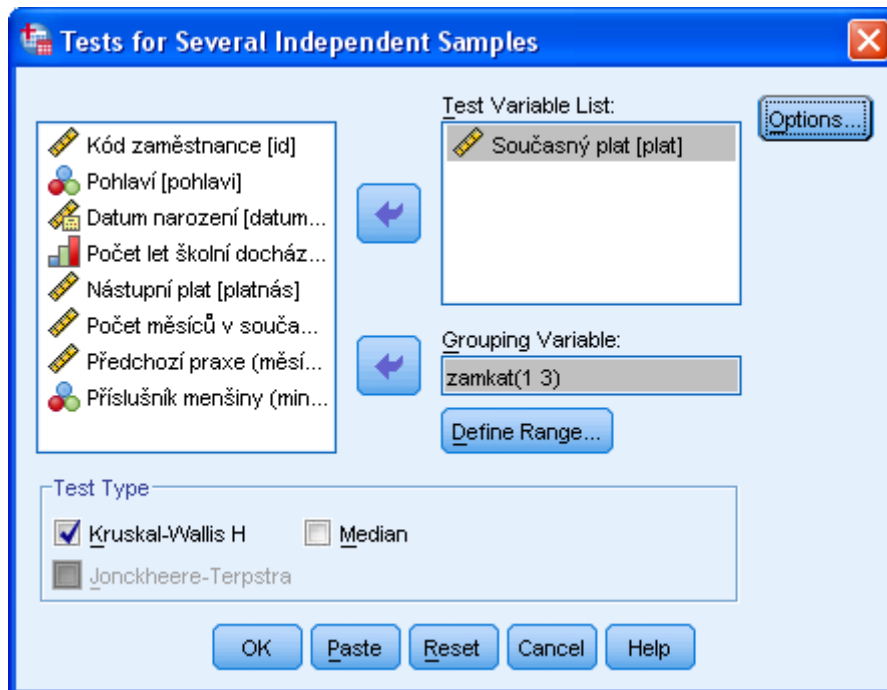
### Slovník anglických pojmů

<i>continue</i>	<i>pokračovat</i>
<i>define</i>	<i>zadat, definovat</i>
<i>distribution</i>	<i>rozdělení</i>
<i>group</i>	<i>skupina</i>
<i>independent</i>	<i>nezávislý</i>
<i>list</i>	<i>seznam</i>
<i>missing</i>	<i>chybějící</i>
<i>nonparametric</i>	<i>neparametrický</i>
<i>normal</i>	<i>normální</i>
<i>one-sample</i>	<i>jednovýběrový</i>
<i>options</i>	<i>volby</i>
<i>sample</i>	<i>výběr</i>
<i>uniform</i>	<i>rovnoměrný</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

#### 1.10.9 Testy pro několik nezávislých výběrů (Tests for Several Independent Samples)

Procedura zahrnuje tři testy (Kruskal-Wallisův H test, mediánový test a Jonckheere-Terpstra test), které jsou neparametrickou obdobou analýzy rozptylu (tj. porovnávají rozdělení hodnot v několika nezávislých skupinách). Test Jonckheere-Terpstra test je však dostupný pouze po nainstalování modulu IBM SPSS Exact Tests.

Testovanou proměnnou (proměnné) převedeme do okna **Test Variable List**. Proměnnou definující nezávislé skupiny převedeme do pole **Grouping Variable**. Tlačítkem **Define Range** zadáme minimální a maximální hodnoty rozpětí této proměnné. V sekci **Test Type** označíme požadované testy. Tlačítko **Options** slouží k výběru popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



Testy pro několik nezávislých výběrů

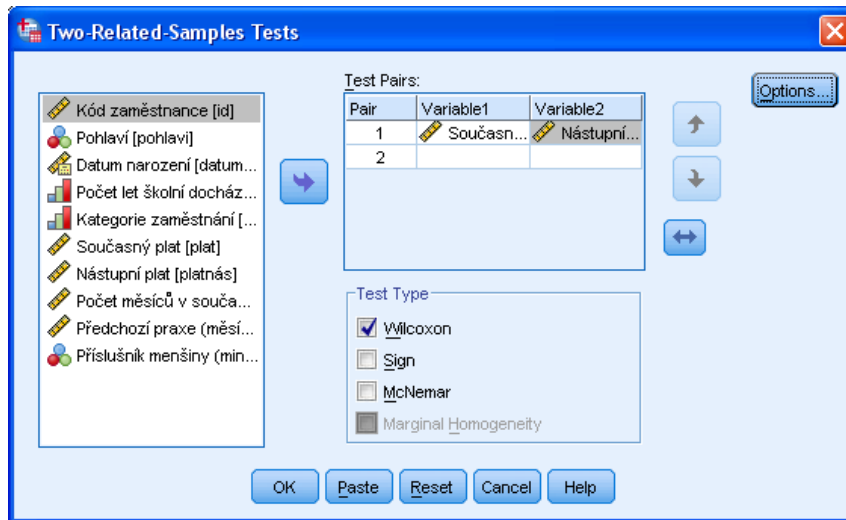
### Testy pro $k$ nezávislých výběrů

- Analyze
  - Nonparametric Tests
    - Legacy Dialogs
      - $K$  Independent Samples

#### 1.10.10 Testy pro dva závislé výběry (Two-Related-Samples Tests)

Procedura zahrnuje tři testy – Wilcoxonův test, znaménkový test a McNemarův test (test marginální homogenity je dostupný až po nainstalování modulu Exact Tests), které jsou neparametrickou obdobou párového t-testu (tj. porovnávají rozdělení hodnot dvou závislých proměnných).

Nejprve označíme první analyzovanou proměnnou a přeneseme ji šipkou do pole **Test Pairs**. Její název se automaticky zobrazí v části **Variable1**. Stejným způsobem přeneseme druhou proměnnou do pole **Variable2**. V sekci **Test Type** zadáme jeden nebo více testů. Tlačítko **Options** slouží k výběru popisných statistik (**Statistics**) a způsobu práce s chybějícími hodnotami (**Missing Values**).



Testy dvou závislých výběrů

### Testy pro dva závislé výběry

- Analyze
  - Nonparametric Tests
    - Legacy Dialogs
      - 2 Related Samples

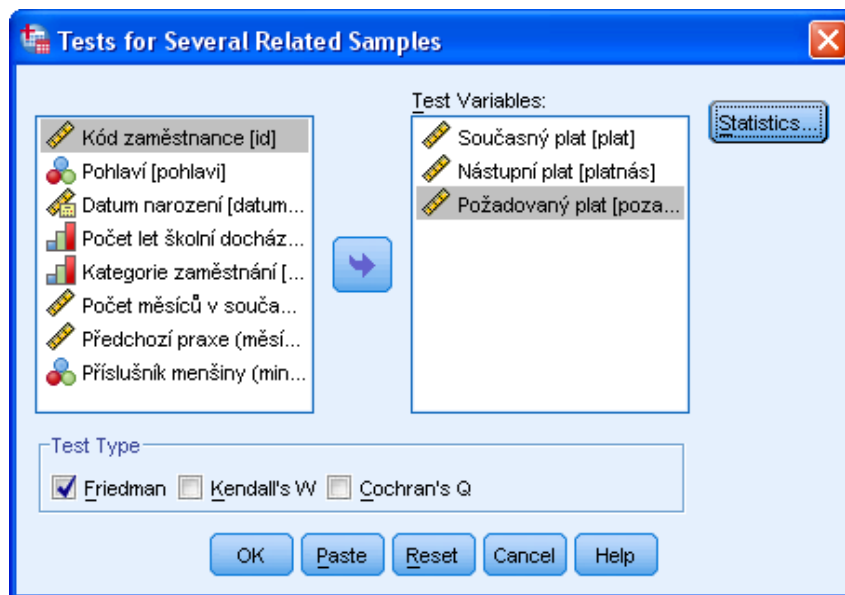
### Slovník anglických pojmů

<i>current</i> .....	<i>běžný, aktuální</i>
<i>define</i> .....	<i>zadat, definovat</i>
<i>group</i> .....	<i>skupina</i>
<i>independent</i> .....	<i>nezávislý</i>
<i>list</i> .....	<i>seznam</i>
<i>options</i> .....	<i>volby</i>
<i>paired</i> .....	<i>párový</i>
<i>sample</i> .....	<i>výběr, soubor</i>
<i>selection</i> .....	<i>selekce, výběr</i>
<i>variable</i> .....	<i>proměnná</i>

### 1.10.11 Testy pro několik závislých výběrů (Tests for Several Related Samples)

Procedura zahrnuje tři testy (Friedmanův test, Kendallův W test a Cochranův Q test), které porovnávají rozdělení hodnot ve více závislých výběrech.

Testovanou proměnnou (proměnné) převedeme do okna **Test Variables**. V sekci **Test Type** označíme testy, které se mají spočítat. Tlačítko **Statistics** slouží k výběru popisných statistik.



Testy pro několik závislých výběrů

### Testy pro několik závislých výběrů

- Analyze
  - Nonparametric Tests
    - Legacy Dialogs
      - K Related Samples

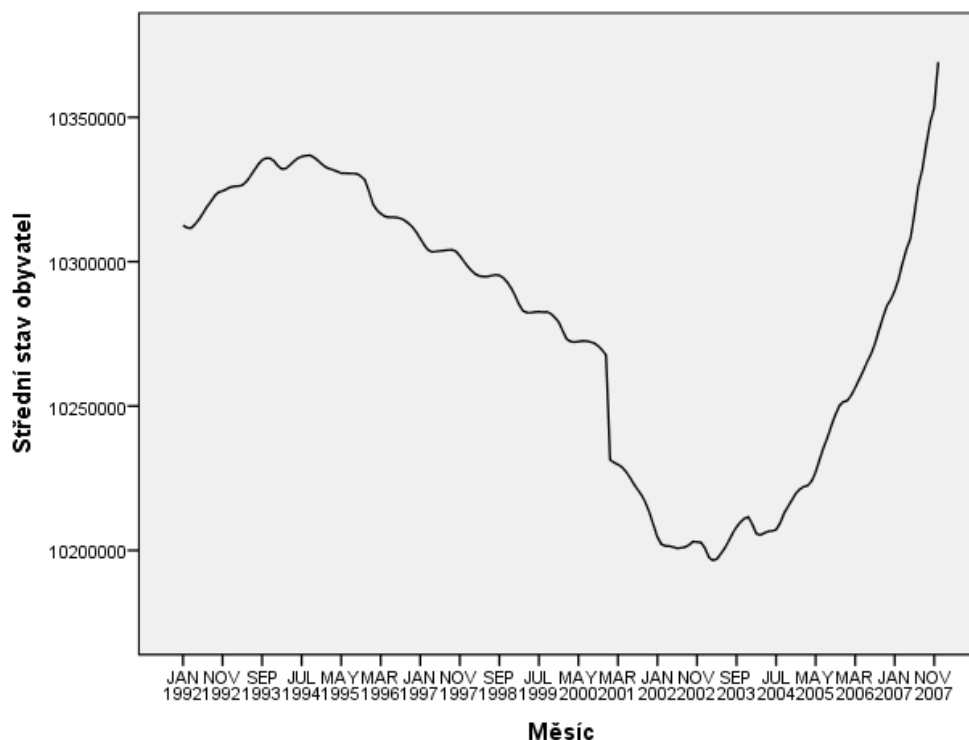
### Slovník anglických pojmů

<i>descriptive</i> .....	<i>popisný</i>
<i>list</i> .....	<i>seznam</i>
<i>nonparametric</i> .....	<i>neparametrický</i>
<i>related</i> .....	<i>závislý</i>
<i>value</i> .....	<i>hodnota</i>
<i>variable</i> .....	<i>proměnná</i>

## 1.11 Časové řady (Forecasting)

### 1.11.1 Graf časových řad (Sequence Charts)

Tento typ grafu zobrazí hodnoty zvolené proměnné (proměnných) pro jednotlivé případy tak, jak jsou za sebou řazeny v datové matici. Graf je proto vhodný především pro analýzu časových řad nebo pro situace, kdy jsou případy pro tento účel smysluplně setříděny.

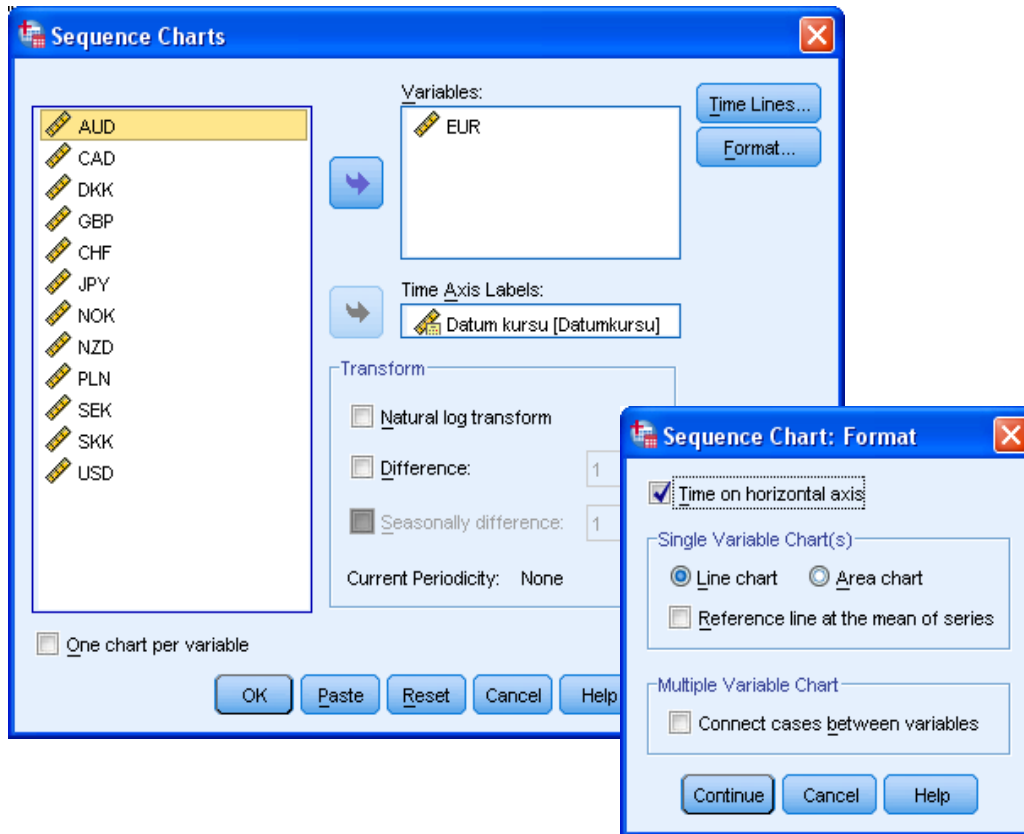


Graf časových řad

Do okna **Variables** zadáme analyzované proměnné, do pole **Time Axis Labels** můžeme přenést proměnnou popisující časové okamžiky. Dále lze zvolit některé transformace časové řady: **Natural log transform** pracuje s přirozeným logaritmem hodnot místo hodnot samotných, **Difference** počítá rozdíly dvou následujících pozorování a **Seasonally difference** vyjadřuje rozdíl vzhledem ke stejnému období předchozí periody. U posledních dvou transformací je nutné určit řád difference (zadáme-li například řád dva, znamená to, že se provede zvolená transformace a pro vzniklou řadu se stejný postup zopakuje ještě jednou). Volba **Seasonally difference** je k dispozici pouze tehdy, jestliže jsme pro časovou řadu definovali systémové datum obsahující periodu (*Data, Define date and time*).

Dále určíme, zda se každá proměnná zobrazí do jiného grafu (**One chart per variable**), nebo všechny do jednoho. Pomocí tlačítka **Time Lines** můžeme ve vybraných bodech grafu zobrazit referenční čáru. Tlačítko **Format** dovoluje upravit některé vlastnosti grafu: přesunout čas na vertikální osu, změnit typ

grafu na plošný, zobrazit referenční čáru v průměru hodnot nebo v grafu s více proměnnými spojit hodnoty odpovídající stejným x-ovým souřadnicím svislými čarami.



**Graf časových řad**

### **Graf časových řad**

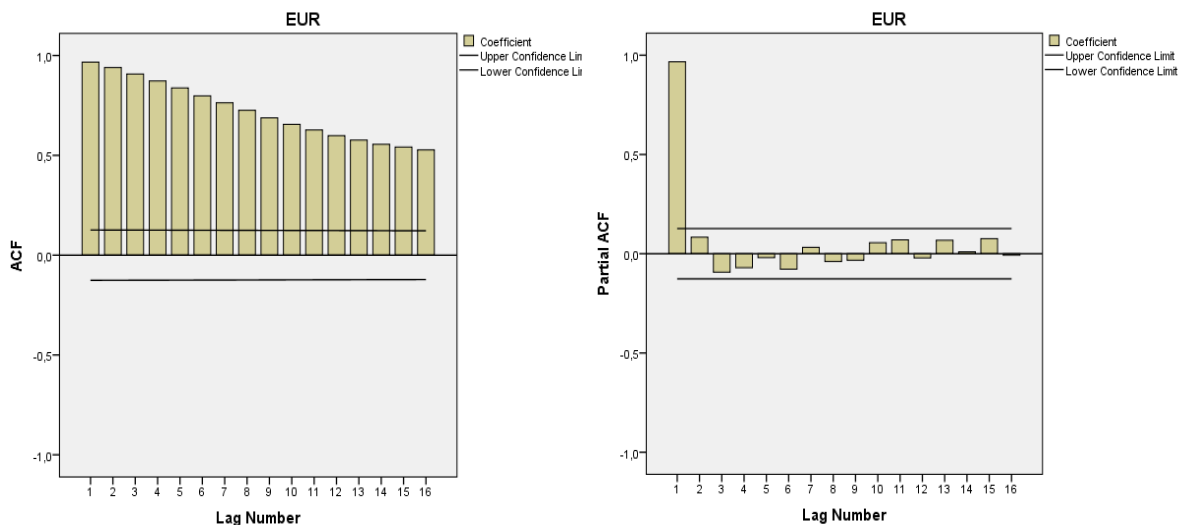
- Analyze
  - Forecasting
    - Sequence Charts

### Slovník anglických pojmů

<i>area</i>	<i>plošný graf</i>
<i>bar</i>	<i>sloupcový graf</i>
<i>boxplot</i>	<i>boxplot</i>
<i>chart</i>	<i>graf</i>
<i>cluster</i>	<i>skupinka</i>
<i>control</i>	<i>řízení procesů</i>
<i>difference</i>	<i>diference, rozdíl</i>
<i>drop</i>	<i>pokles (rozsah)</i>
<i>error bar</i>	<i>graf rozptýlení</i>
<i>error</i>	<i>chyba</i>
<i>high</i>	<i>vysoký</i>
<i>layer</i>	<i>vrstva</i>
<i>linegraph</i>	<i>spojnicový graf</i>
<i>line</i>	<i>čára</i>
<i>low</i>	<i>nízký</i>
<i>matrix</i>	<i>matice</i>
<i>multiple</i>	<i>mnohonásobný</i>
<i>overlay</i>	<i>ve vrstvách</i>
<i>pie</i>	<i>koláč</i>
<i>pie graph</i>	<i>výsečový graf</i>
<i>plot</i>	<i>graf</i>
<i>range</i>	<i>rozpětí</i>
<i>scatterplot</i>	<i>bodový, souřadnicový graf</i>
<i>scatter</i>	<i>rozložení</i>
<i>sequence</i>	<i>řada</i>
<i>simple</i>	<i>jednoduchý</i>
<i>stacked</i>	<i>skládáný</i>
<i>time series</i>	<i>časové řady</i>

#### 1.11.2 Autokorelace (Autocorrelations)

Procedura zobrazuje graf autokorelační funkce (ACF) a parciálních autokorelační funkce (PACF) pro jednu nebo více časových řad.

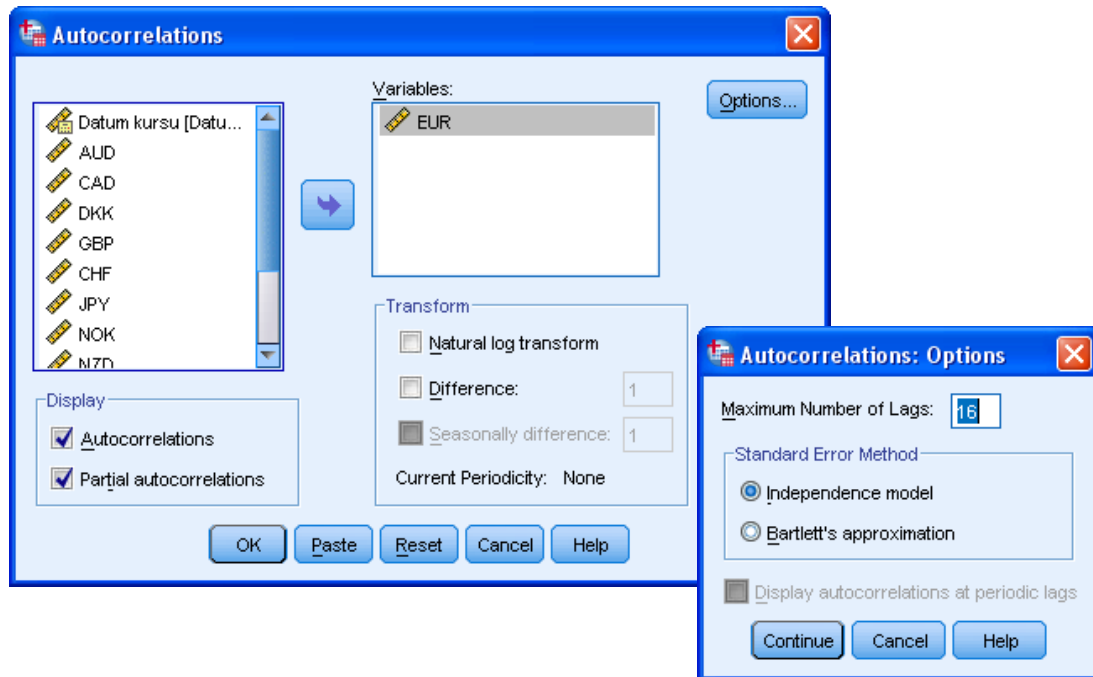


### Autokorelace

Do pole **Variables** zadáme analyzované proměnné. V části **Display** zaškrtneme požadované výstupy – graf autokorelační funkce (**Autocorrelations**), graf parciální autokorelační funkce (**Partial autocorrelations**). Dále lze určit transformace časové řady: **Natural log transform**, kdy se místo s původními hodnotami počítá s přirozeným logaritmem těchto hodnot, **Difference** pro výpočet rozdílů dvou následujících pozorování a **Seasonally difference** – rozdíly vzhledem ke stejnému období předchozí periody. U posledních dvou transformací je nutné určit řád difference (zadáme-li například řád dva, znamená to, že se provede zvolená transformace a pro vzniklou řadu se stejný postup zopakuje ještě jednou). Volba **Seasonally difference** je k dispozici pouze tehdy, jestliže jsme pro časovou řadu definovali systémové datum obsahující periodu (*Data, Define date and time*).

Dále zvolíme pomocí tlačítka **Options** maximální počet zpoždění zobrazený v grafech (**Maximum Number of Lags**) a vybereme v poli **Standard Error Method** metodu pro výpočet rozptylu odhadů. K dispozici jsou dvě metody: **Independence model** a **Bartlett's approximation**. Je-li určena perioda časové řady, lze zvolit možnost **Display autocorrelations at periodic lags** pro zobrazení autokorelací s periodickým zpožděním.





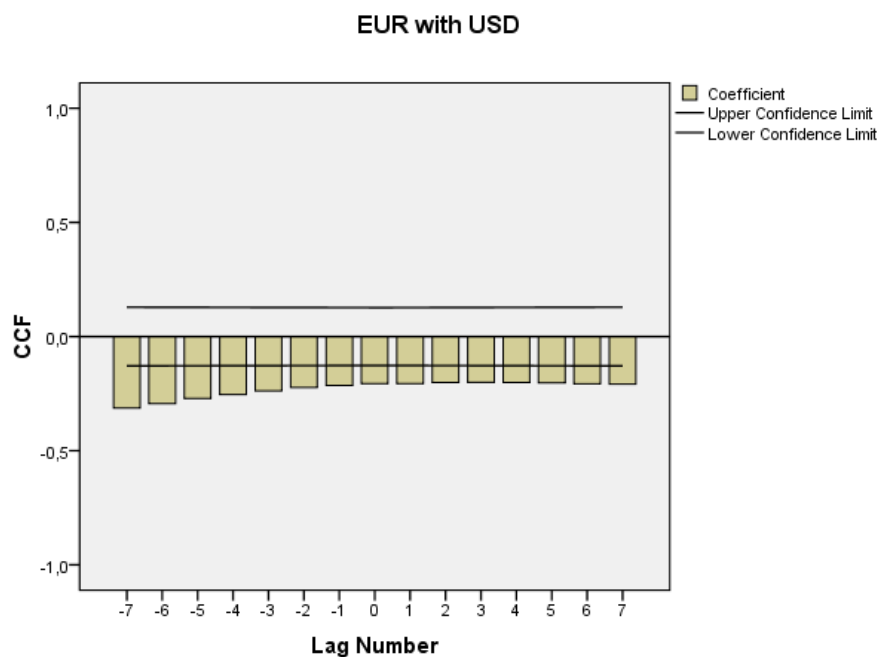
### **Autokorelace**

### **Autokorelace**

- Analyze
  - Forecasting
    - Autocorrelations

### **1.11.3 Vzájemné korelace (Cross-Correlations)**

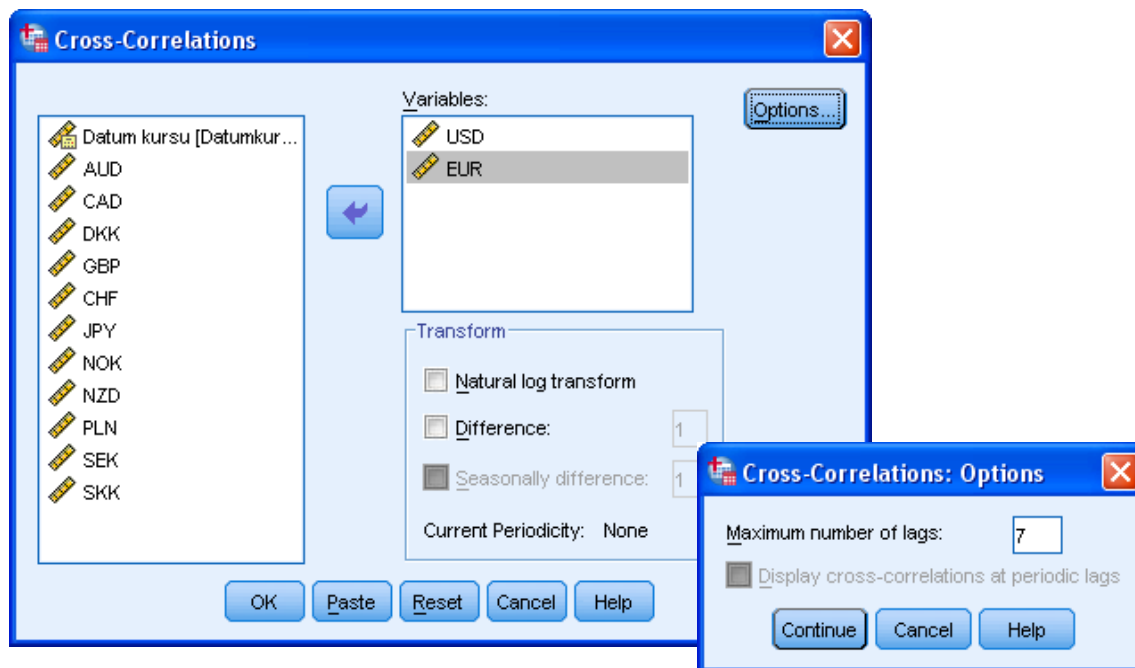
Procedura vytvoří graf vzájemné korelační funkce (CCF) pro dvě nebo více časových řad, a to pro posunutí dopředu i dozadu.



**Vzájemné korelace**

V poli **Variables** zadáme dvě nebo více analyzované proměnné (časové řady). Dále lze zvolit transformace časových řad: **Natural log transform** pro práci s přirozeným logaritmem řady, **Difference** pro rozdíly dvou následujících pozorování a **Seasonally difference** pro rozdíly vzhledem ke stejnému období předchozí periody. U posledních dvou transformací je nutné určit řád difference (zadáme-li například řád dva, znamená to, že se provede zvolená transformace a pro vzniklou řadu se stejný postup zopakuje ještě jednou). Volba **Seasonally difference** je k dispozici pouze tehdy, jestliže jsme pro časovou řadu definovali systémové datum obsahující periodu (*Data, Define date and time*).

Dále zvolíme tlačítkem **Options** maximální počet zpoždění zobrazený v grafu (**Maximum Number of Lags**). Je-li definována perioda časové řady, lze zvolit možnost **Display cross-correlations at periodic lags** pro zobrazení vzájemných korelací s periodickým zpožděním.



### Autokorelace

### Vzájemné korelace

- Analyze
  - Forecasting
    - Cross-Correlations

### Slovník anglických pojmů

<i>bar</i>	<i>sloupcový graf</i>
<i>chart</i>	<i>graf</i>
<i>cross-correlations</i>	<i>vzájemné korelace</i>
<i>difference</i>	<i>diference, rozdíl</i>
<i>error</i>	<i>chyba</i>
<i>lag</i>	<i>zpoždění</i>
<i>matrix</i>	<i>matice</i>
<i>method</i>	<i>metoda</i>
<i>periodicity</i>	<i>periodicita</i>
<i>plot</i>	<i>graf</i>
<i>range</i>	<i>rozpětí</i>
<i>sequence</i>	<i>řada</i>
<i>time series</i>	<i>časové řady</i>
<i>transform</i>	<i>transformovat</i>
<i>variable</i>	<i>proměnná</i>

## 1.12 Mnohonásobné odpovědi (Multiple Response)

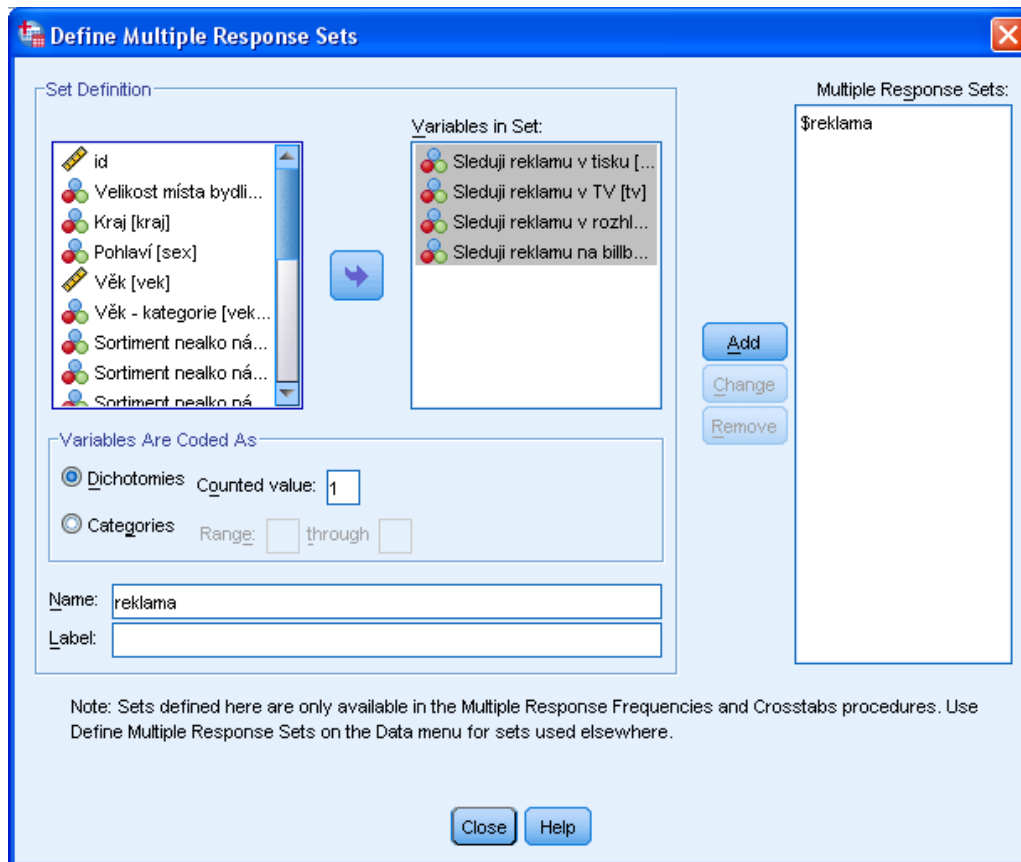
Nabídka je určena k analýze mnohonásobných odpovědí, tj. ke zpracování otázek, na které mohl respondent uvést více než jednu odpověď. Poté, co definujeme tzv. sadu proměnných (skupinu proměnných v níž je odpověď na danou otázku zakódována), můžeme pro ni vytvářet tabulky četností nebo kontingenční tabulky.

### 1.12.1 Sady proměnných pro mnohonásobné odpovědi (Define Sets)

Pomocí této procedury definujeme sadu proměnných pro analýzu mnohonásobných odpovědí.

Proměnné, které tvoří sadu, převedeme do pole **Variables in Set**. V poli **Variables Are Coded As** rozhodneme, jakým způsobem jsou odpovědi kódovány. Může se jednat o skupinu dichotomických proměnných (každé možné odpovědi odpovídá jedna proměnná, 1= uvedl, 0=neuvedl), nebo o skupinu multinomických proměnných (proměnné reprezentují postupně jednotlivé odpovědi respondenta a nabývají tolika hodnot, kolik je možných odpovědí). V prvním případě volíme možnost **Dichotomies** a do pole **Counted value** zadáme kód, který reprezentuje, že odpověď byla uvedena (obvykle 1). Ve výstupu se potom zobrazí přehled zastoupení této kategorie pro každou proměnnou sady samostatně i v úhrnu. V případě druhého typu kódování označíme volbu **Categories** a určíme kód první a poslední kategorie. Výstupní tabulka potom charakterizuje zastoupení jednotlivých kategorií v celé sadě.

Do pole **Name** napíšeme název sady proměnných, do pole **Label** zadáme její popis. Pomocí tlačítka **Add** přeneseme definovanou sadu do pole **Multiple Response Sets**. Pro změnu označené sady proměnných slouží tlačítko **Change**, odstranění se provádí tlačítkem **Remove**.



**Sady proměnných pro mnohonásobné odpovědi**

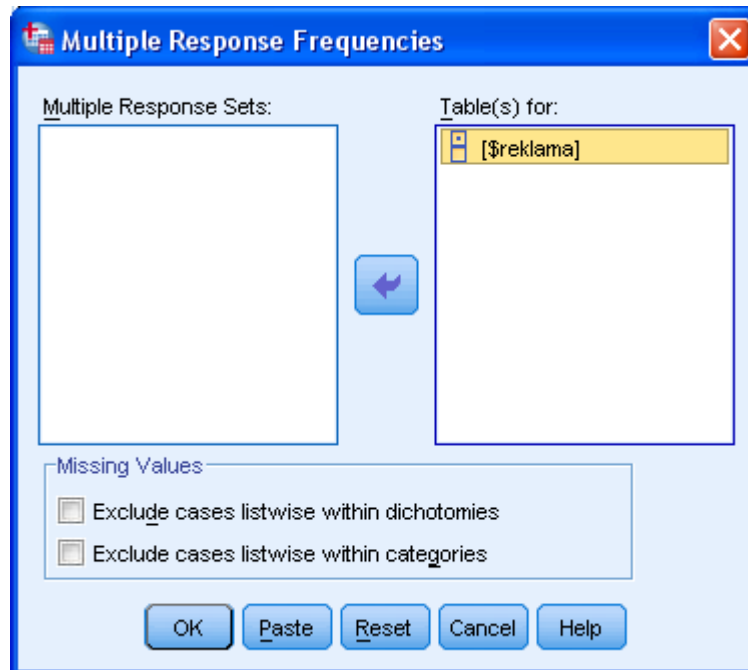
### **Sady proměnných pro mnohonásobné odpovědi**

- *Analyze*
  - *Multiple Response*
    - *Define Sets*

#### **1.12.2 Výpočet četností pro mnohonásobné odpovědi (Frequencies)**

Máme-li definovanou sadu proměnných, můžeme ji využít pro výpočet tabulky četností.

Analyzovanou proměnnou (proměnné) převedeme do pole **Table(s) for**. V sekci **Missing Values** rozhodneme, jak naložit s chybějícími hodnotami.



*Výpočet četností pro mnohonásobné odpovědi*

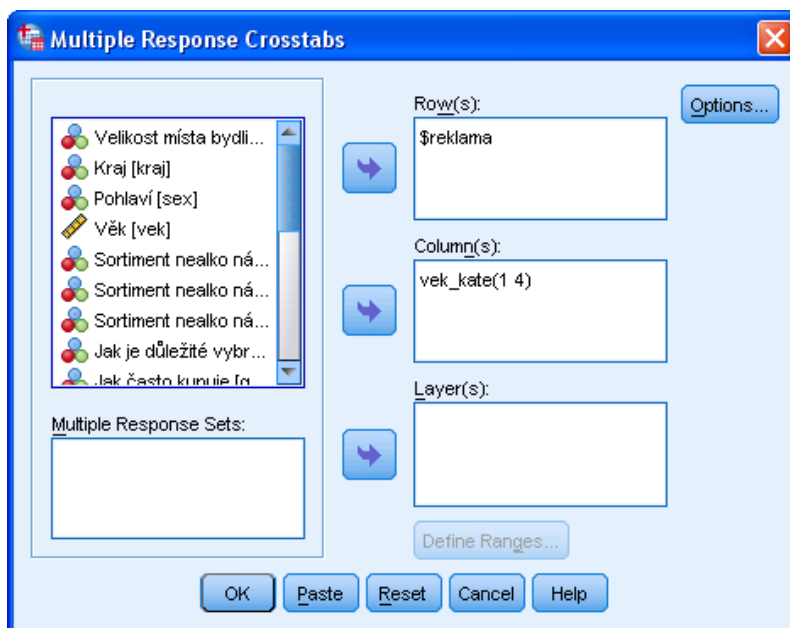
### **Výpočet četností pro mnohonásobné odpovědi**

- Analyze
  - Multiple Response
    - Frequencies

#### **1.12.3 Kontingenční tabulky pro mnohonásobné odpovědi (Crosstabs)**

Procedura dovoluje vytvářet kontingenční tabulky pro sady mnohonásobných odpovědí. Před použitím této nabídky je nejprve nutné definovat sadu proměnných (viz oddíl 1.12.1 *Sady proměnných pro mnohonásobné odpovědi (Define Sets)*, str. 100).

Proměnnou (sadu proměnných) tvořící řádky tabulky převedeme do pole **Row(s)**, proměnnou (sadu proměnných) tvořící sloupce do pole **Column(s)**. Jestliže chceme vytvořit tabulku třetího (nebo vyššího) stupně třídění, převedeme požadované proměnné do pole **Layer(s)**. U kategorizovaných proměnných určíme tlačítkem **Define Ranges** rozmezí zobrazených kategorií. Pomocí tlačítka **Options** lze zadat typ relativních četností – řádkové, sloupcové nebo celkové (**Cell percentages**), základ relativních četností – případy nebo odpovědi (**Percentages Based on**) a metodu práce s chybějícími hodnotami (**Missing Values**). Procedura rovněž dovoluje tabelovat skupiny dichotomií proti sobě (**Match variables across response sets**).



*Kontingenční tabulky pro mnohonásobné odpovědi*

### ***Kontingenční tabulky pro mnohonásobné odpovědi***

- *Analyze*
  - *Multiple Response*
    - *Crosstabs*

**Slovník anglických pojmů**

<i>add</i> .....	<i>přidat</i>
<i>case</i> .....	<i>případ</i>
<i>change</i> .....	<i>změnit</i>
<i>close</i> .....	<i>zavřít</i>
<i>coded as</i> .....	<i>kódovaný jako</i>
<i>column</i> .....	<i>sloupec</i>
<i>continue</i> .....	<i>pokračovat</i>
<i>define</i> .....	<i>zadat, definovat</i>
<i>label</i> .....	<i>popis, označení</i>
<i>layer</i> .....	<i>vrstva</i>
<i>multiple</i> .....	<i>mnohonásobný</i>
<i>name</i> .....	<i>jméno</i>
<i>options</i> .....	<i>předvolby</i>
<i>remove</i> .....	<i>odstranit</i>
<i>response</i> .....	<i>odpověď</i>
<i>row</i> .....	<i>řádek</i>
<i>set</i> .....	<i>sada</i>
<i>table for</i> .....	<i>tabulka pro</i>
<i>value</i> .....	<i>hodnota</i>



## 1.13 Simulace (Simulation)

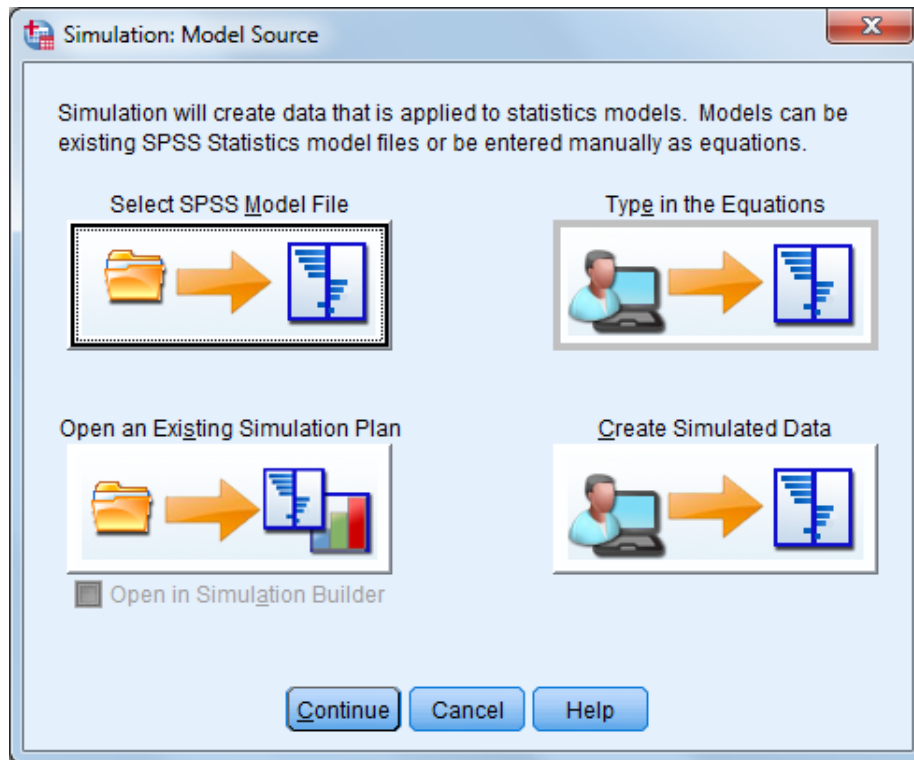
Procedura nabízí simulace dat metodou Monte Carlo. Simulace nacházejí uplatnění například při budování predikčních modelů s nejistými vstupy. V takových případech lze pomocí nich odhadnout pravděpodobnost různých situací. Data lze však simulovat i bez využití predikčního modelu.

Při užití predikčního modelu jsou vstupní proměnné modelovány na základě různých pravděpodobnostních rozdělení, nebo, pokud je jejich hodnota známá, jsou fixovány na tuto hodnotu. Nasimulované i pevně dané hodnoty mohou být následně použity pro výpočet výstupní proměnné na základě předem určeného predikčního modelu. Díky tomu, že se simuluje několik tisíc hodnot, je odhadnuto rozložení výstupní proměnné, z něhož lze odhadovat pravděpodobnost výskytu různých situací (např. poptávka převyší dostupné zdroje, prodeje poklesnou pod určitou mez apod.). Pokud simulujeme data bez užití predikčního modelu, necháme jejich rozložení odhadnout z pozorovaných dat, nebo ho zadáme přímo. Žádná proměnná potom není považována za výstupní.

Pro provedení simulace je třeba specifikovat predikční model (je-li vyžadován), pravděpodobnostní rozdělení nejistých vstupů, hodnoty fixovaných vstupů a korelace všech vstupních proměnných. Simulace může být potom provedena či uložena v podobě speciálního souboru obsahujícího plán simulace, který lze rovněž sdílet s dalšími uživateli a používat opakovaně.

Procedura je vybavena dvojím grafickým rozhraním. **Simulation Builder** je určen uživatelům, kteří navrhují a provádějí simulace. Toto rozhraní disponuje všemi dostupnými funkcemi pro návržení simulace, její uložení v podobě simulačního plánu, výběr požadovaných výstupů procedury a samotné provedení simulace. Pro uživatele, kteří mají soubor s plánem simulace a chtějí pomocí něho především simulaci provést, je k dispozici zjednodušené rozhraní **Run Simulation**, kde lze upravovat pouze vybrané parametry simulace a realizovat simulace při různých nastaveních. Obě rozhraní jsou interaktivní a poskytují kontextovou nápovědu, jejíž obsah se mění v závislosti na označené části dialogu či provedeném výběru.

Úvodní dialog procedury nabízí čtyři základní volby: simulace na základě predikčního modelu uloženého ve formě XML, PMML nebo ZIP souboru (**Select SPSS Model File**), simulace na základě modelu specifikovaného uživatelem pomocí zadaných rovnic (**Type in the Equations**), otevření souboru s plánem simulace a realizace simulací (**Open an Existing Simulation Plan**) nebo simulace dat bez užití predikčního modelu **Create Simulated Data**. Při výběru třetí možnosti lze otevřít rozhraní Simulation Builder (**Open in Simulation Builder**), nebo pouze zjednodušené rozhraní s omezenými možnostmi.



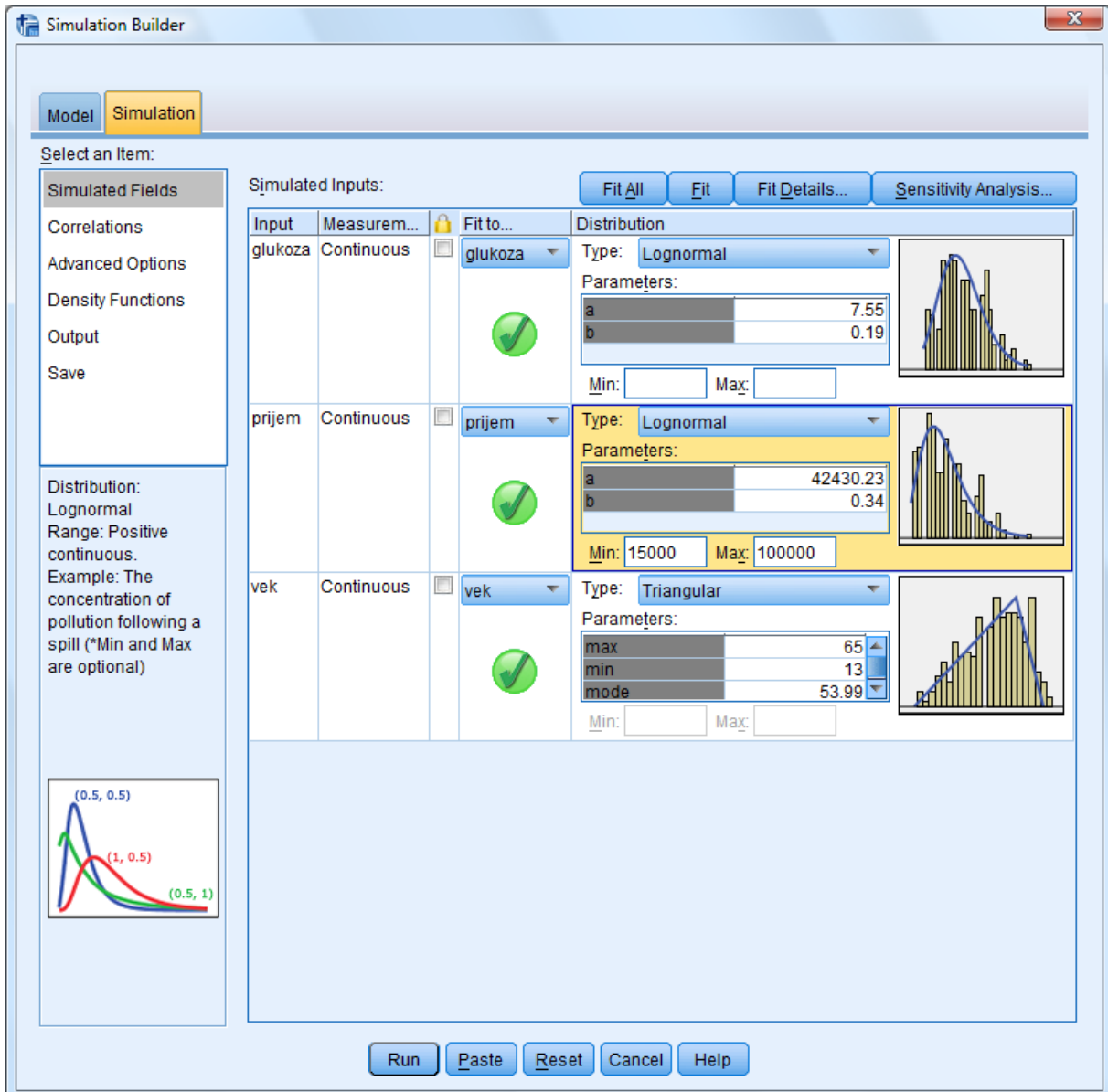
Úvodní dialog simulační procedury

Při volbě **Select SPSS Model File** je nutné nejprve načíst soubor, v němž je uložena specifikace predikčního modelu (formáty XML, PMML nebo ZIP). Tento soubor lze získat ve formátu XML jako výstup procedur pro tvorbu predikčních modelů (např. lineární či logistická regrese, rozhodovací stromy, seskupovací algoritmy apod.) nebo ve formátu ZIP z procedury Automatic Linear Modeling.

Následně se otevře rozhraní **Simulation Builder** na záložce **Simulation**, kde je třeba upřesnit potřebné parametry simulace.

Záložka **Simulation** je rozdělena do několika oblastí, mezi kterými se přechází výběrem položky ze seznamu **Select an Item** v levé části okna. V oblasti **Simulated Fields** se zobrazuje seznam vstupních proměnných načteného modelu (**Simulated Inputs**). Výstupní proměnná je určena modelem a bude pomocí něho dopočítána ze simulovaných hodnot vstupních proměnných, proto není v seznamu uvedena. Pro každou vstupní proměnnou je nutné určit pravděpodobnostní rozdělení a jeho parametry, nebo ji fixovat na pevnou hodnotu – provádí se výběrem rozdělení, resp. volby Fixed, v poli **Type** a vepsáním hodnot parametrů do tabulky **Parameters**, případně omezením simulovaných hodnot na rozmezí od minima (**Min**) do maxima (**Max**). Procedura umožňuje pracovat i s textovými vstupními proměnnými, pro něž je vyhrazen typ *Categorical*. Má-li být nastavení uzamčeno, a tím zamezeno dalším změnám, stačí zaškrtnout políčko ve sloupci s ikonou zámku. Pro proměnné napojené na proměnné datového souboru, lze pravděpodobnostní rozdělení určit automaticky tlačítkem **Fit All** (pro všechny proměnné) nebo **Fit** (pro vybranou proměnnou). V takovém případě software odhadne parametry různých teoretických rozdělení a vybere to, které nejlépe odpovídá datům. Podrobnosti o procesu hledání nejhodnějšího rozdělení si prohlédneme tlačítkem **Fit Details**, které zobrazí tabulku s teoretickými rozděleními, která byla vyzkoušena, odhadnutými hodnotami jejich parametrů a výsledky testu shody s daty z aktivního datového souboru. Vstupní proměnné modelu se automaticky spojí s proměnnými z aktivního datového souboru, které mají stejný název a

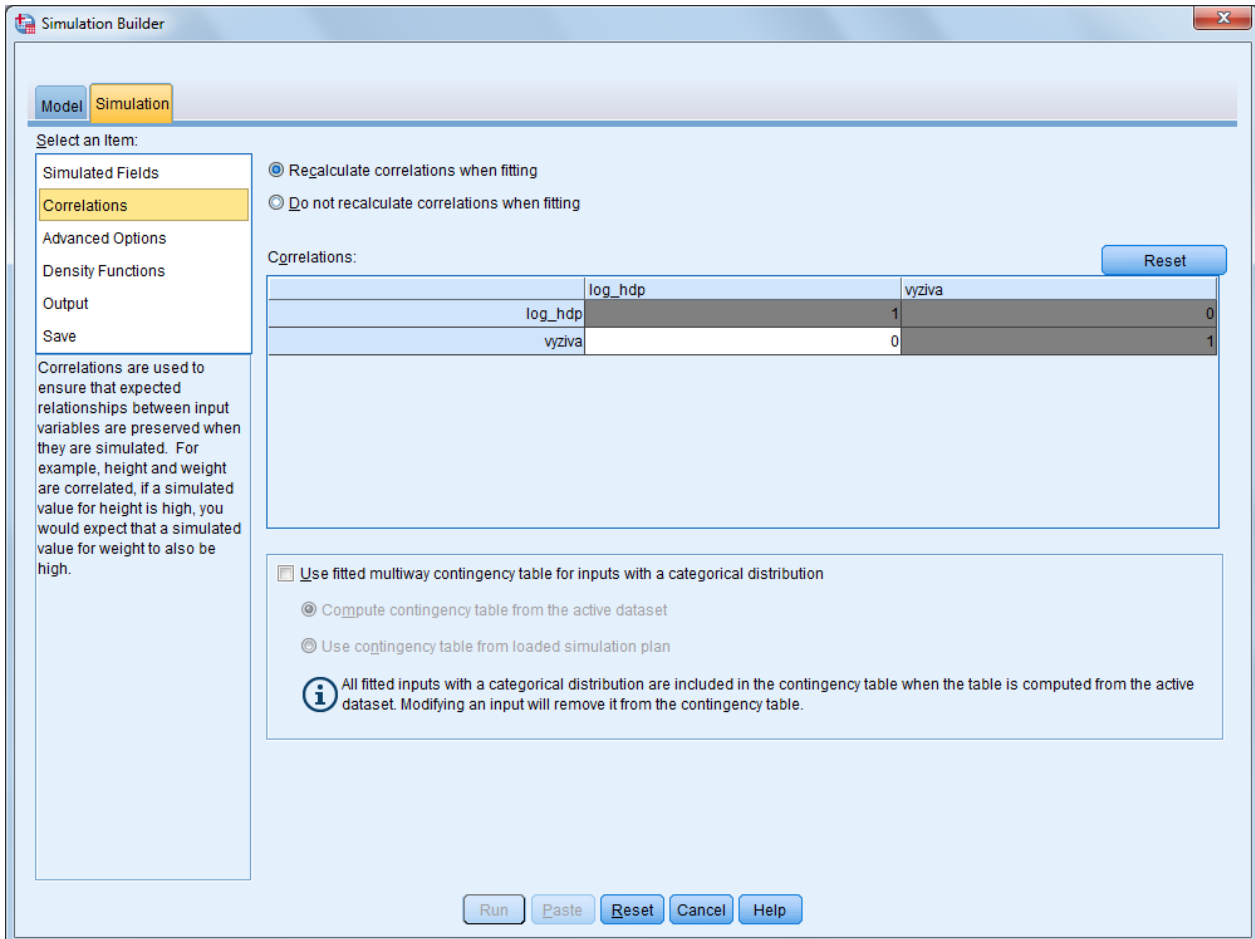
typ. Spojení je možné provést i ručně výběrem proměnné v poli **Fit to ...** nebo ponechat proměnnou nenapojenou. Pro jednotlivé proměnné lze zadat tlačítkem **Sensitivity Analysis** různé hodnoty některého z parametrů vybraného rozložení a provést tak srovnání výsledků při změně nastavení, tzv. analýzu senzitivity.



**Rozhraní Simulation Builder, záložka Simulation, položka Simulated Fields**

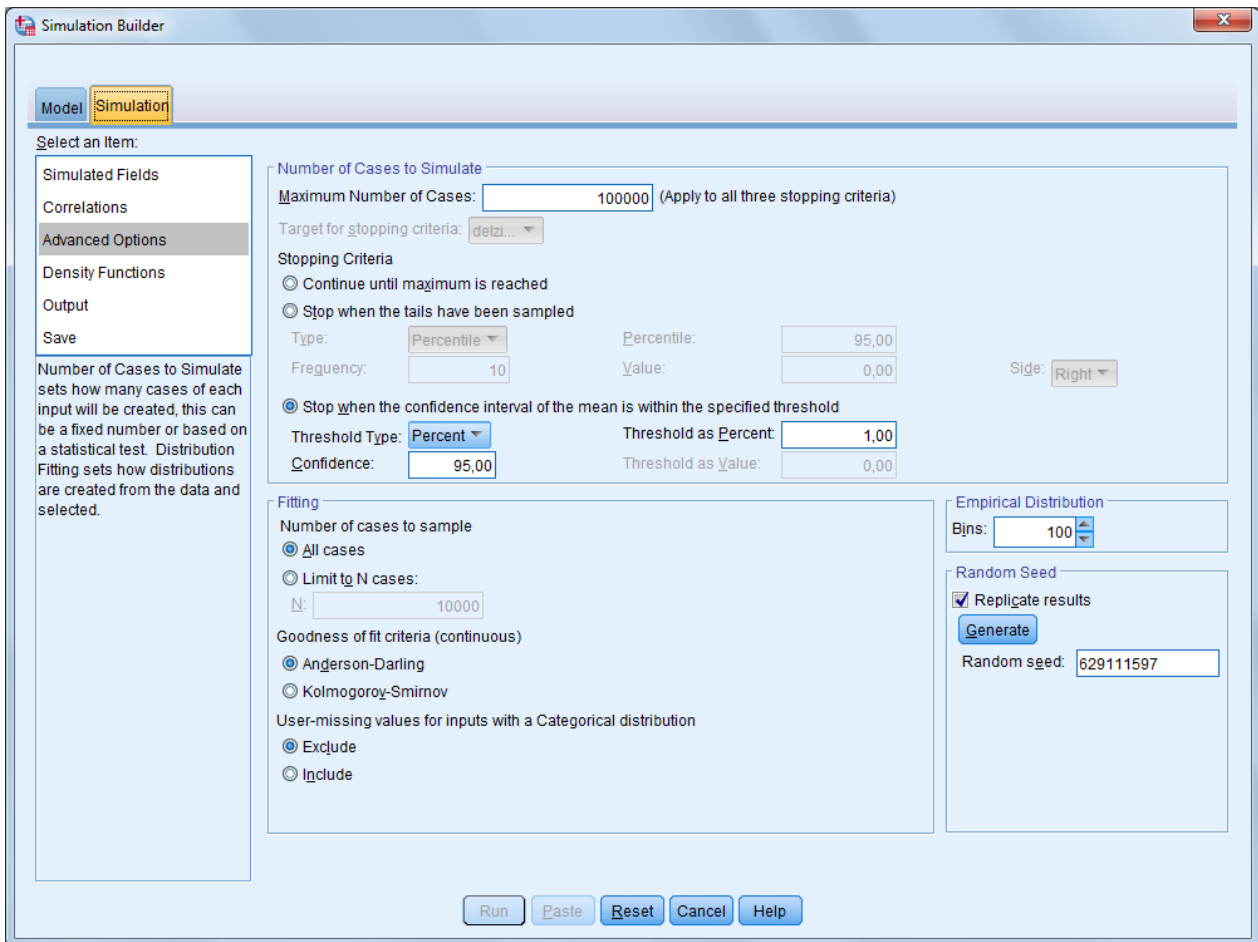
Položka **Correlations** ze seznamu **Select an Item**, slouží k zadání požadovaných korelací mezi simulovanými proměnnými. Bylo-li rozdělení proměnných odhadnuto tak, aby co nejlépe odpovídalo datům z aktivního datového souboru, jsou odhadnuty také korelace a zobrazeny v tabulce **Correlations**. Tato tabulka zároveň slouží i k ručnímu zadávání požadovaných hodnot. Volbou **Do not recalculate correlations when fitting** se zamezí změně korelací při opakovaném hledání nevhodnějšího rozdělení. Pro vynulování všech korelací je určeno tlačítko **Reset**.

Procedura rovněž umožňuje načíst asociace mezi nominálními proměnnými, jejichž rozložení odhadované z dat je nastaveno na typ *Categorical*, a použít je při simulaci. Asociace jsou dány celkovými procenty z mnohorozměrné kontingenční tabulky určené těmito proměnnými. Pro jejich načtení zaškrtneme volbu **Use fitted multiway contingency table for inputs with a categorical distribution**.



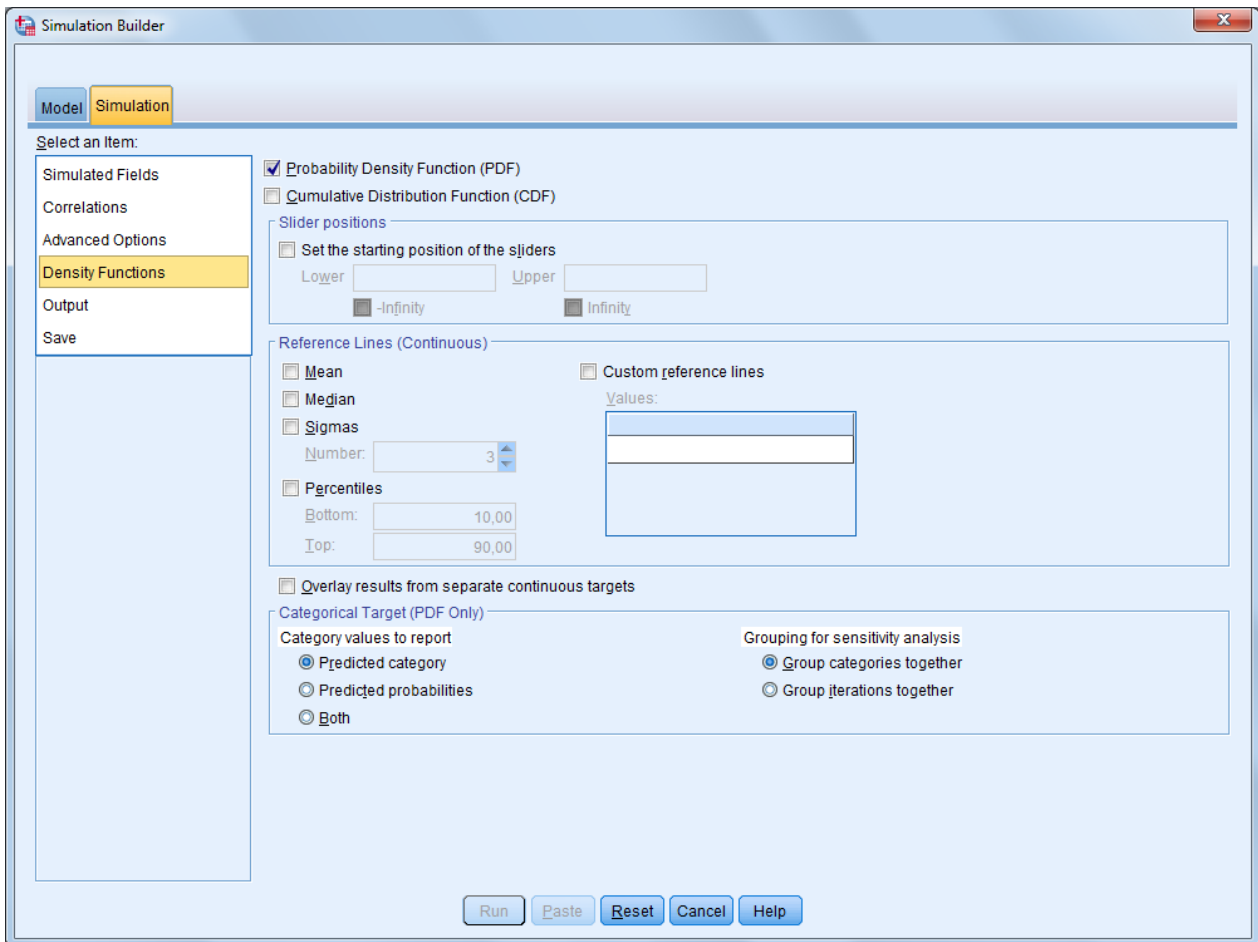
Rozhraní Simulation Builder, záložka Simulation, položka Correlations

Položka **Advanced Options** ze seznamu **Select an Item**, zobrazuje další nastavení simulace. Horní část specifikuje kritéria pro ukončení procesu simulace hodnot. Simulaci lze ukončit při dosažení požadovaného počtu případů (**Continue until maximum is reached**) uvedeného v poli **Maximum Number of Cases**, je-li nasimulován dostatečný počet extrémních případů (**Stop when the tails have been sampled**), nebo pokud interval spolehlivosti pro průměr dosáhl požadované přesnosti (**Stop when the confidence interval of the mean is within the specified threshold**). V části **Fitting** lze omezit počet případů používaných k odhadu nejvhodnějšího rozdělení (**Number of cases to sample**), vybrat test shody, který je kritériem při výběru nejvhodnějšího rozdělení (**Goodness of fit criteria**) a rozhodnout o zahrnutí nebo vyloučení uživatelem definovaných vynechaných hodnot pro proměnné s rozložením *Categorical* (**User-missing values for inputs with a Categorical distribution**). Pro výpočet empirických distribučních funkcí číselných vstupních proměnných lze v části **Empirical Distribution** zvolit počet intervalů (**Bins**). V části **Random Seed** lze zajistit replikovatelnost výsledku tak, že zaškrtneme políčko **Replicate results** a zadáme počáteční hodnotu pro generátor náhodných čísel (**Random seed**), nebo ji necháme vygenerovat tlačítkem (**Generate**).



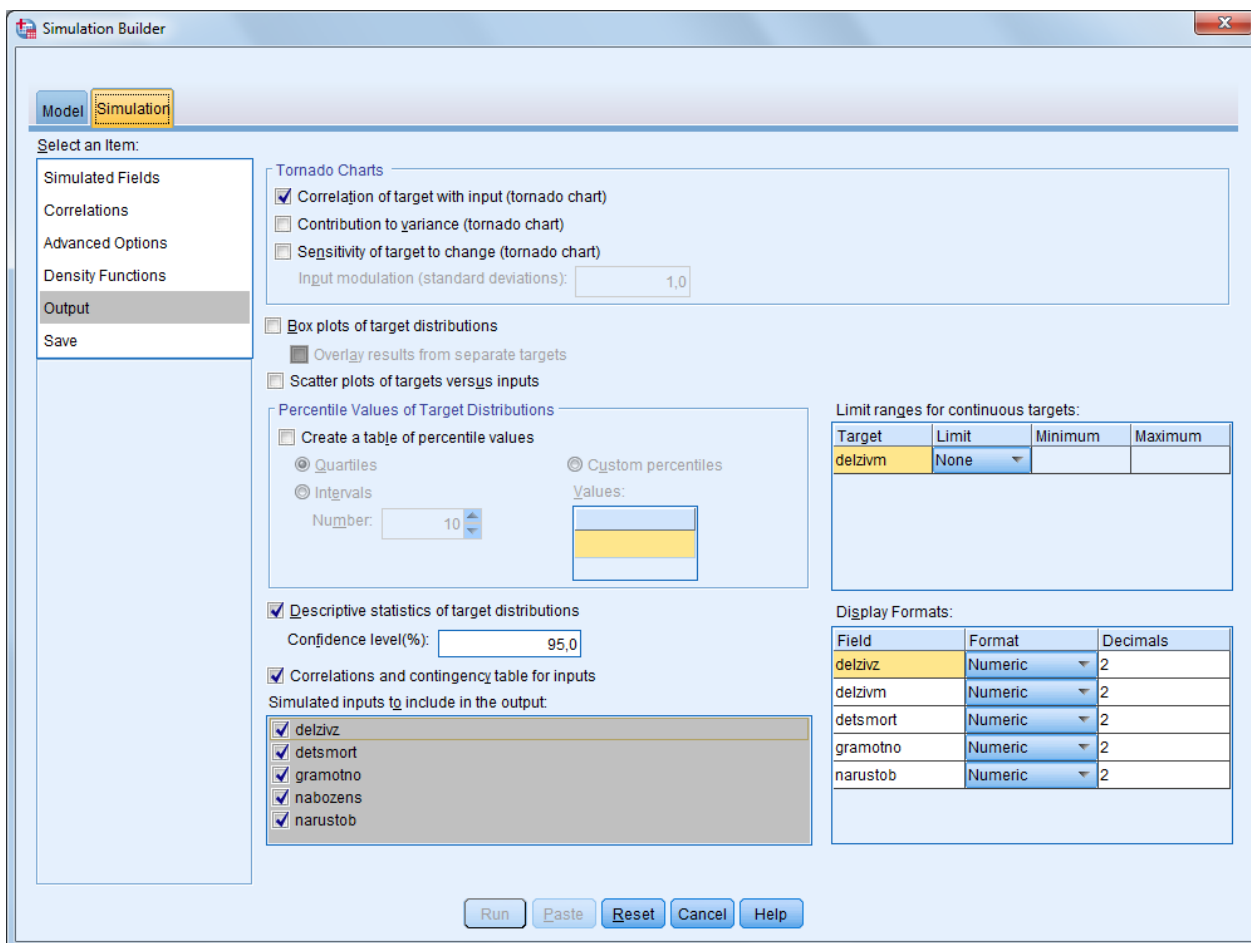
**Rozhraní Simulation Builder, záložka Simulation, položka Advanced Options**

V části **Density Functions** lze nastavit funkce, které se zobrazí ve výstupním grafu (**Probability Density Function, Cumulative Distribution Function**), nastavit polohu jezdců v grafu rozložení výstupní proměnné (**Slider positions**), přidat referenční čáry do grafu pro číselné výstupní proměnné (**Reference Lines (Continuous)**), nechat křivky pro více číselných závislých proměnných vykreslit do jednoho grafu (**Overlay results from separate continuous targets**) a v případě kategorizované závislé proměnné zvolit, zda se má reportovat predikovaná kategorie, její pravděpodobnost, nebo obojí (**Categorical Target (PDF Only)**). V případě analýzy senzitivity je třeba určit, zda se má výstupní skupinkový sloupcový graf seskupovat po kategoriích nebo po iteracích analýzy senzitivity (**Grouping for sensitivity analysis**).



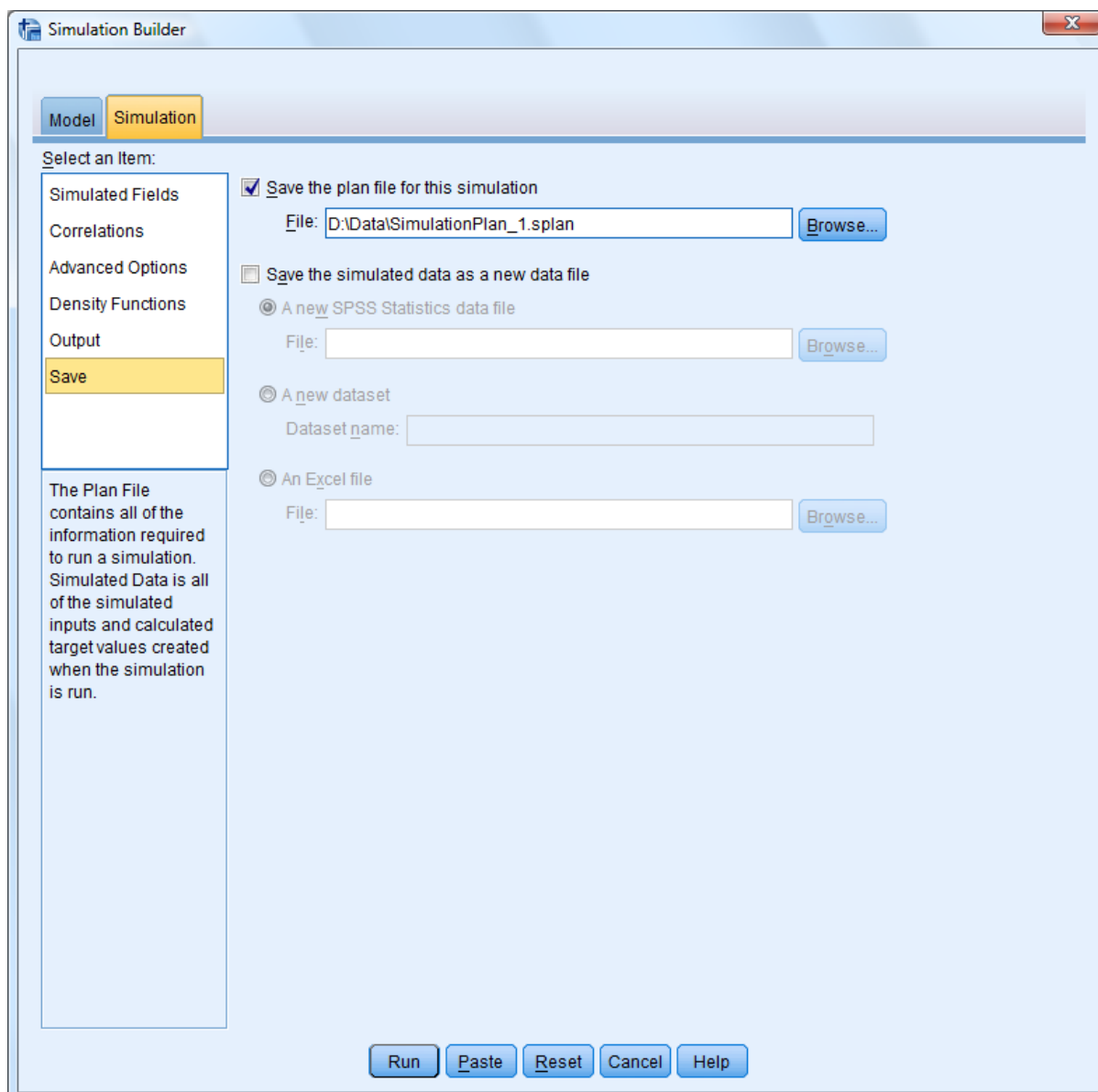
**Rozhraní Simulation Builder, záložka Simulation, položka Density Functions**

V oblasti **Outputs** lze přidat do výstupu grafy a tabulky. V horní části je nabídka tzv. tornádových grafů (**Tornado Charts**): graf korelací vstupních proměnných s výstupní proměnnou, graf příspěvků jednotlivých vstupních proměnných k variabilitě výstupní proměnné a graf citlivosti výstupní proměnné na změnu hodnot vstupních proměnných. Dále lze zobrazit boxploty závislých proměnných (**Box plots of target distributions**) a bodové grafy vztahů závislých proměnných s jednotlivými vstupními proměnnými, případně teplotní grafy (heat maps), pokud je alespoň jedna ze znázorňovaných proměnných typu *Categorical* (**Scatter plots of targets versus inputs**), přidat tabulku požadovaných percentilů závislých proměnných (**Percentile Values of Target Distributions**) a tabulku popisných statistik pro závislé proměnné a vybrané vstupní proměnné (**Descriptive statistics of target distributions a Simulated inputs to include in the output**) v požadovaných formátech (**Display Formats**). Je-li některá vstupní proměnná vynechána z tabulkových výstupů, není součástí ani výše zmíněných grafů. Zaškrtnutím položky **Correlations and contingency table for inputs** zobrazíme kontingenční tabulku pro nominální vstupní proměnné a korelační matici pro číselné a ordinální vstupní proměnné, které budou spočítány z již nasimulovaných dat. Generované hodnoty číselné výstupní proměnné lze omezit dolním a horním limitem (**Limit ranges for continuous targets**).



Rozhraní Simulation Builder, záložka Simulation, položka Output

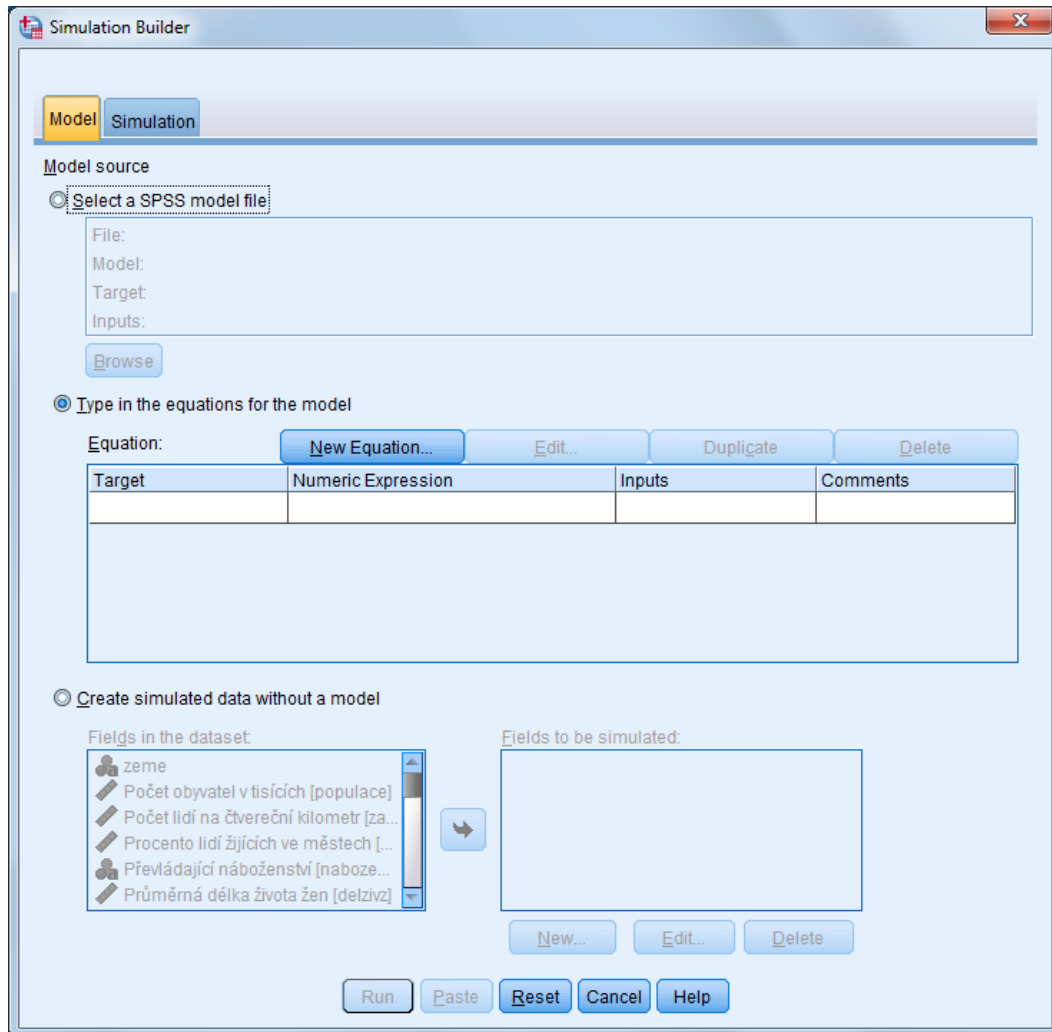
V části **Save** lze uložit soubor s plánem simulace (**Save the plan file for this simulation**) a nasimulované hodnoty do nového datového souboru, datového okna, nebo do Excelu (**Save the simulated data as a new data file**).



**Rozhraní Simulation Builder, záložka Simulation, položka Save**

Po výběru **Type in the Equations** v úvodním dialogu procedury, se otevře rozhraní **Simulation Builder** na záložce **Model**, kde je aktivní druhá volba **Type in the equations for the model**. Predikční model se v tomto případě určuje přímo, pomocí rovnic. Nová rovnice se definuje tlačítkem **New Equation**. Již zadaná rovnice se edituje tlačítkem **Edit**, duplikuje tlačítkem **Duplicate** a maže tlačítkem **Delete**.

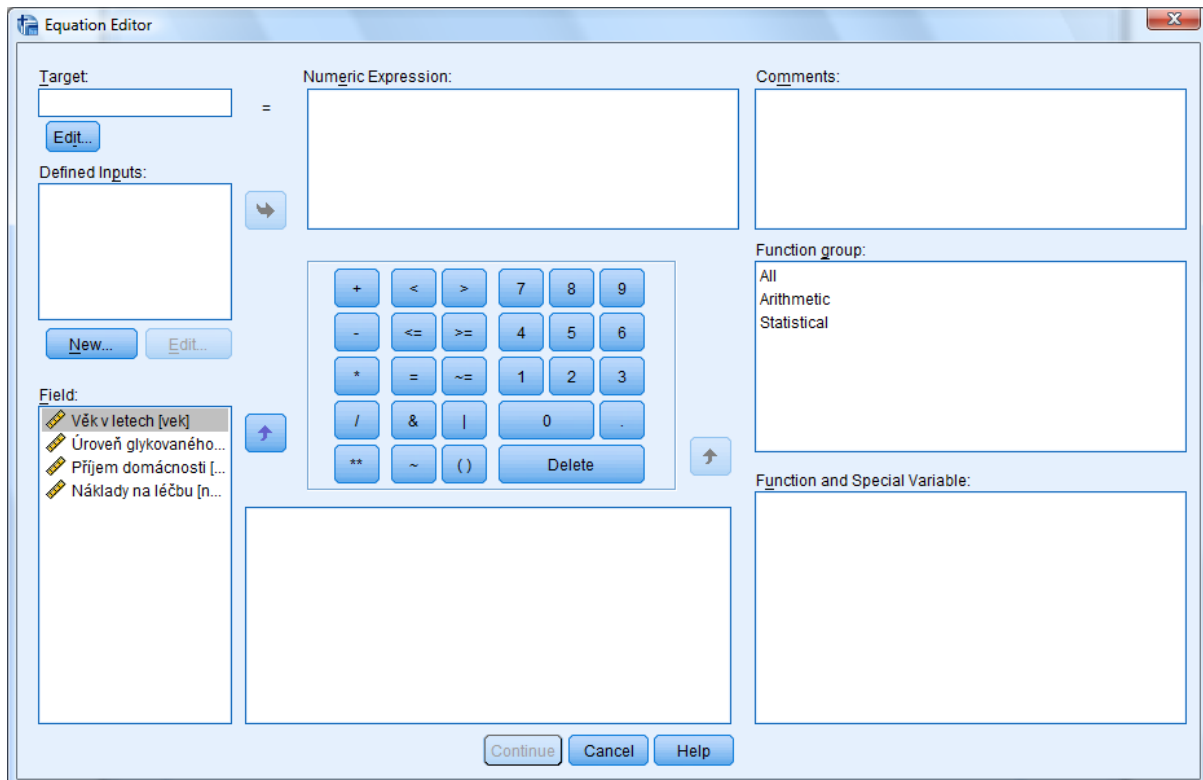




**Rozhraní Simulation Builder, záložka Model**

Tlačítkem **New Equation** se otevře rozhraní **Equation Editor**, ve kterém se specifikuje rovnice predikčního modelu. Do pole **Target** se vepíše název výstupní proměnné a tlačítkem **Edit ...** se případně změní její typ. Do pole **Numeric Expression** se vepíše rovnice modelu, v níž mohou vystupovat proměnné z aktivního datového souboru (**Field**), vstupní proměnné, které nejsou v aktivním datovém souboru (**Defined Inputs**), čísllice, operátory a funkce. Vstupní proměnné, které nejsou v aktivním datovém souboru, se definují tlačítkem **New ...** a případně se upravují tlačítkem **Edit ...**. K rovnici může být přidán také komentář (**Comments**).

Nastavení na záložce **Simulation** se poté provádějí stejným způsobem jako je popsáno výše pro případ zadávání predikčního modelu uloženého ve formě souboru (volba **Select SPSS Model File** z úvodního dialogu procedury).

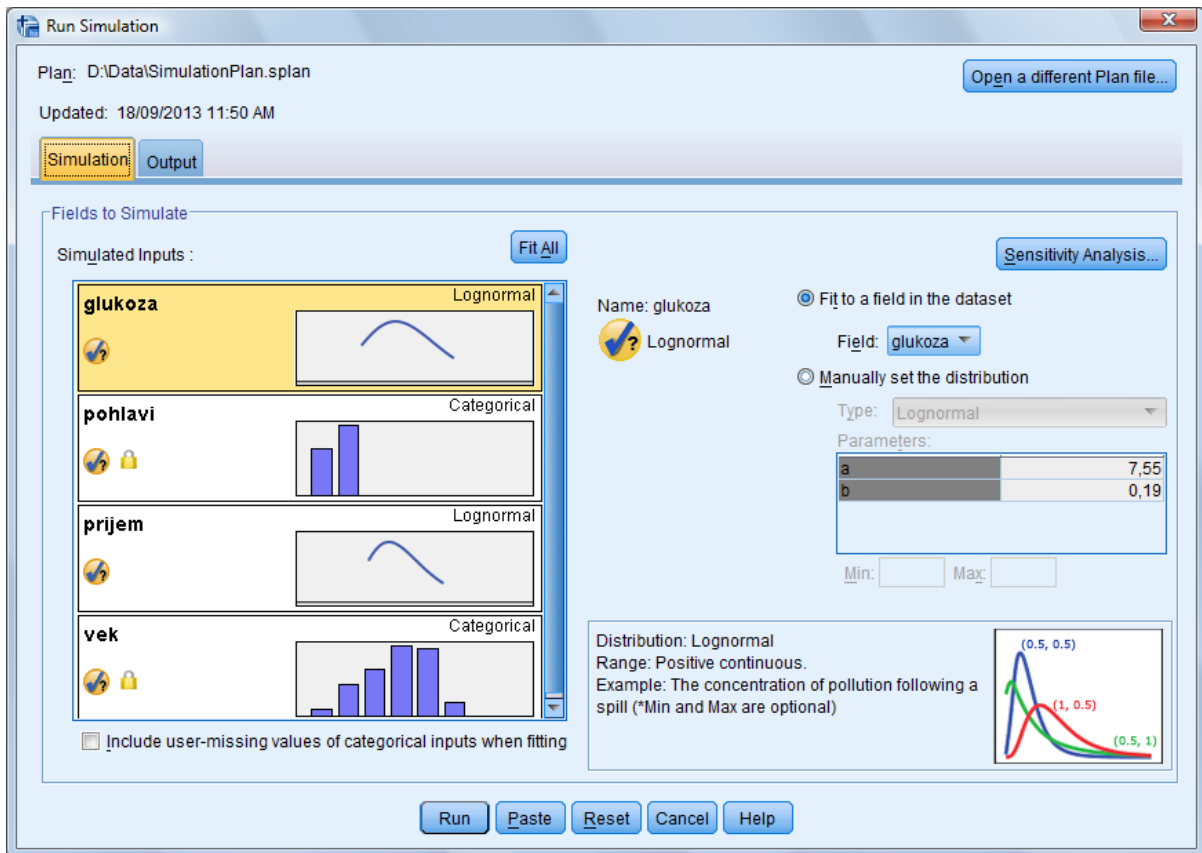


**Rozhraní Simulation Builder, Equation Editor**

Při volbě **Open an Existing Simulation Plan** v úvodním dialogu, procedury je třeba následně specifikovat příslušný soubor obsahující plán simulace, tj. soubor s příponou \*.splan. Je-li zároveň zaškrtnuta volba **Open in Simulation Builder**, otevře se rozhraní **Simulation Builder** popsané výše.

V případě, že plán zahrnuje načtení asociací mezi proměnnými typu Categorical jsme upozorněni, že nastavení pro tyto proměnné bude uzamčeno. V tom případě jsou tyto asociace načteny ze souboru s plánem simulace a v části **Correlation** je aktivní volba **Use contingency table from loaded simulation plan**. Chceme-li načíst asociace z aktivního datového okna, zvolíme **Compute contingency table from active dataset**. Následně můžeme pod položkou **Simulated Fields** v seznamu **Select an Item** aktualizovat nastavení proměnných tlačítkem **Fit All**, příp. **Fit**.

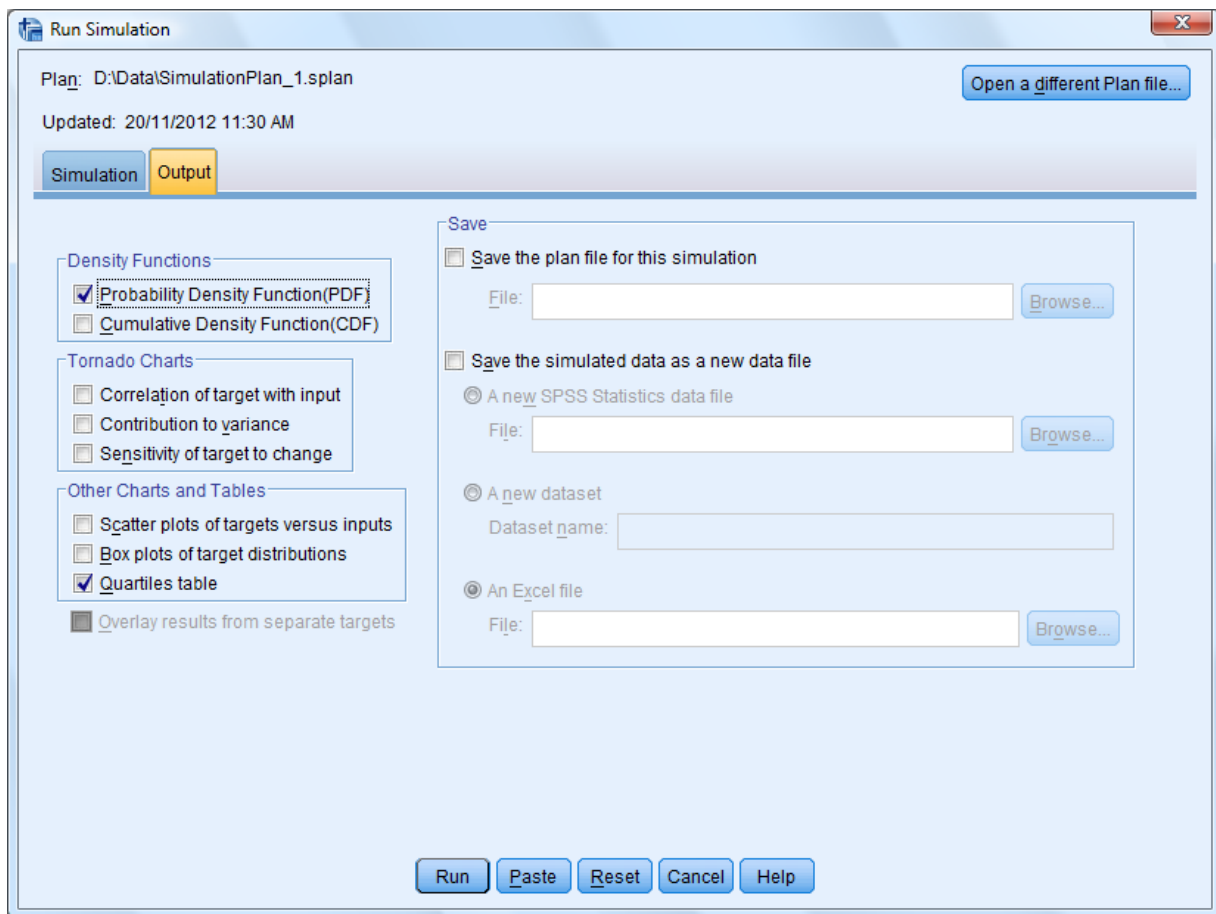
Jinak se otevře pouze zjednodušené rozhraní **Run Simulation**, které je opatřeno dvěma záložkami: **Simulation** a **Output**. Na záložce **Simulation** se zobrazují vstupní proměnné modelu (**Simulated inputs**), rozdělení, které pro ně bylo v plánu simulace nastaveno a jeho parametry. Jednotlivé proměnné, které nebyly uzamčeny pro další změny, je možné napojit na proměnné z aktivního datového souboru (**Fit to a field in the dataset**) a tlačítkem **Fit All** provést nalezení nejvhodnějšího rozdělení a jeho parametrů, nebo rozdělení či jeho parametry změnit ručně (**Manually set the distribution**). Pro jednotlivé proměnné lze také tlačítkem **Sensitivity Analysis** zadat různé hodnoty některého z parametrů a provést analýzu senzivity výsledku na změnu daného parametru. Zaškrtnutím volby **Include user-missing values of categorical inputs when fitting** zahrneme uživatelem definované vynechané hodnoty proměnných s rozdělením *Categorical* mezi platné při odhadu rozdělení na základě aktivního datového souboru.



**Rozhraní Run Simulation, záložka Simulation**

Na záložce **Output** jsou k dispozici některá z nastavení dostupných v rozhraní Simulation Builder popsaných výše: výběr typu zobrazené funkce (**Density Functions**), tornádové grafy (**Tornado Charts**), další grafy a tabulky (**Other Charts and Tables**), možnost zobrazení křivek pro více výstupních proměnných do jednoho grafu (**Overlay results from separate targets**), uložení souboru s plánem simulace a uložení nasimulovaných dat do souboru (**Save**).

Změnu souboru s plánem simulace lze provést tlačítkem **Open a different Plan file ...**

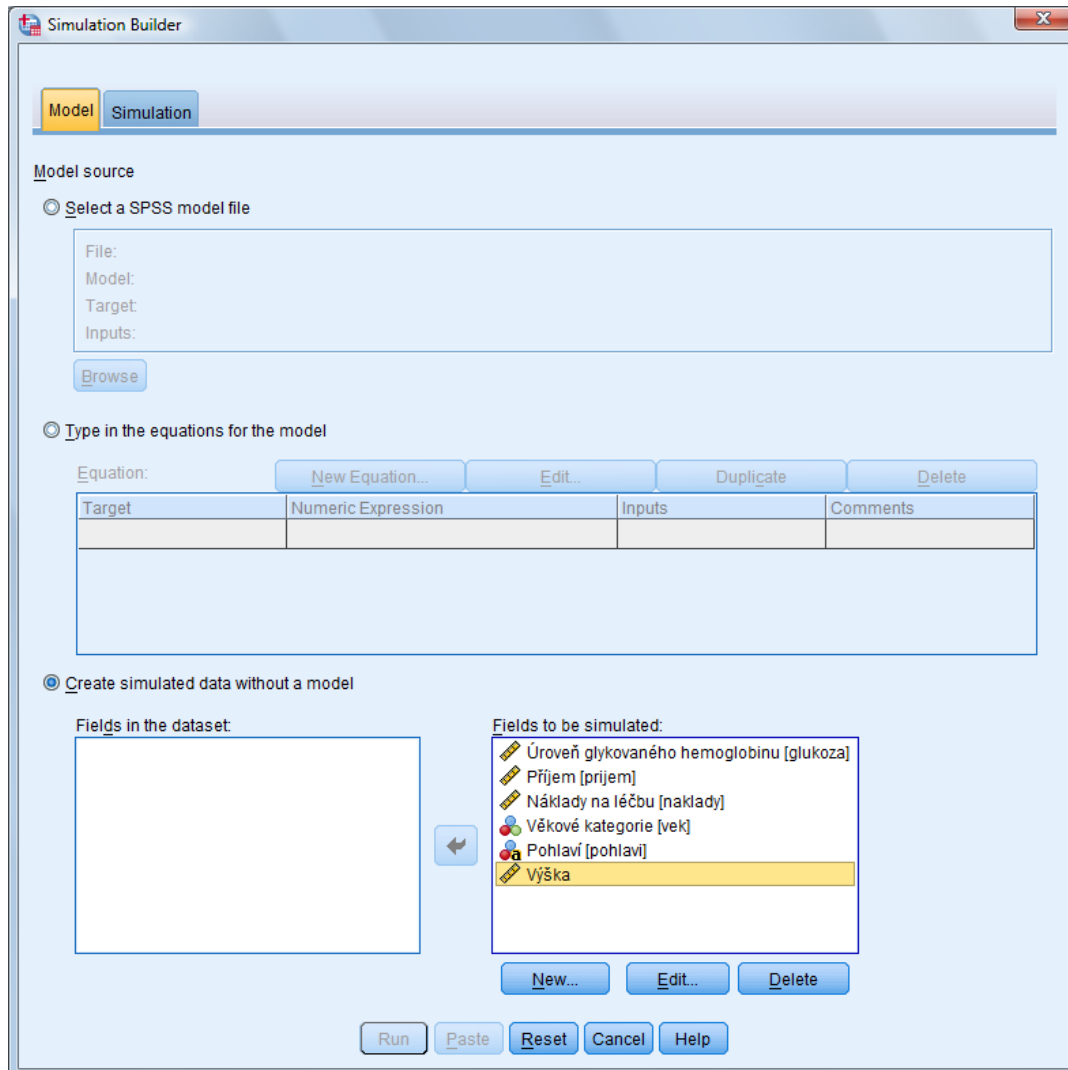


**Rozhraní Run Simulation, záložka Output**

Při volbě **Create Simulated Data** z úvodního dialogu nasimulujeme data bez užití predikčního modelu. Rozložení simulovaných dat necháme buď odhadnout z pozorovaných dat, nebo ho zadáme přímo. Žádná proměnná není považována za výstupní.

V rozhraní **Simulation Builder** na záložce **Model** je aktivní volba **Create simulated data without a model**. Zde vybereme proměnné z aktivního datového okna, které mají být simulovány na základě dat a přeneseme je do pole **Fields to be simulated**. Tlačítkem **New** dále zadáme další proměnné, jejichž rozdělení stanovíme přímo. Tlačítkem **Change** nastavení označené proměnné upravíme, tlačítkem **Delete** vybranou proměnnou odstraníme.

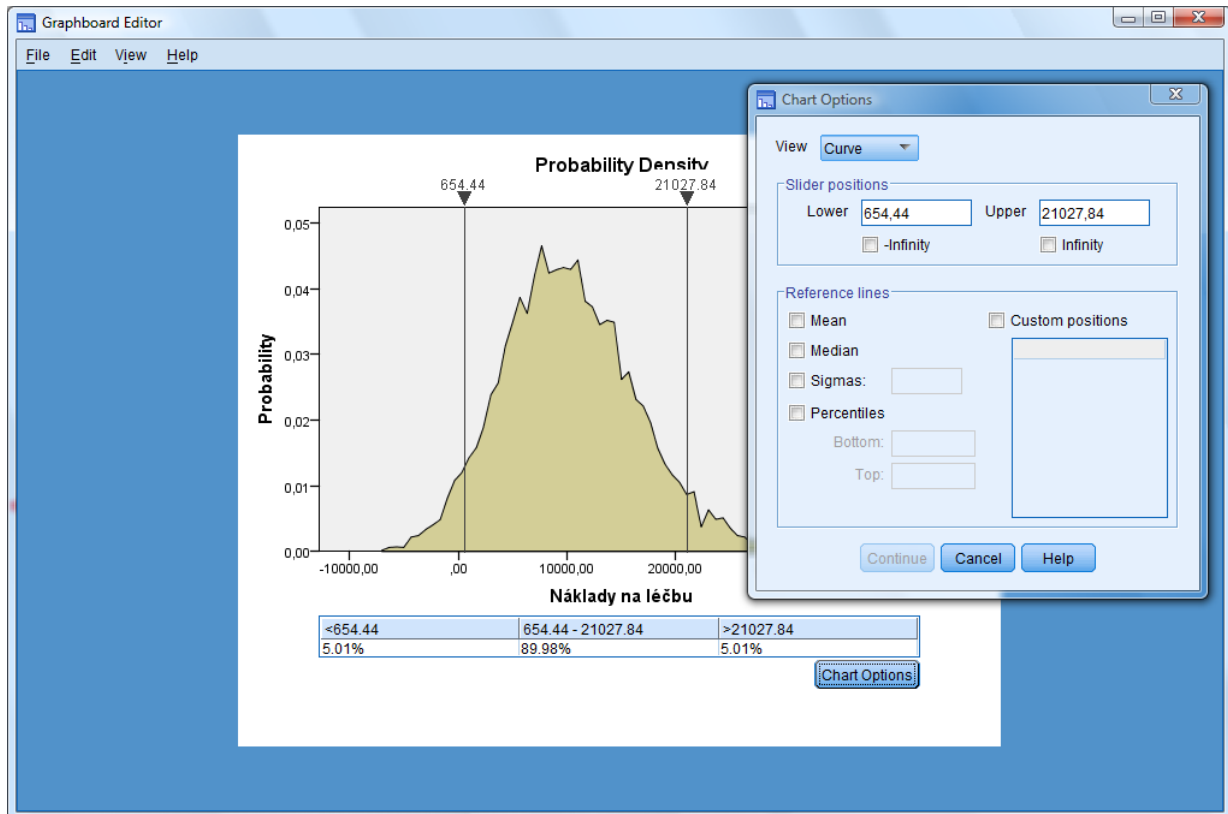
V zadávání pokračujeme na záložce **Simulation** stejným způsobem jako při práci s predikčními modely. Velká část voleb je však v tomto případě neaktivní a v několika případech je změněno defaultní nastavení ve srovnání s jinými způsoby zadávání simulací. Pod položkou **Advanced Options** v seznamu **Select an Item** je k dispozici pouze jedno kritérium pro zastavení simulování dat – zastavit simulace v okamžiku, kdy je dosaženo maximálního počtu případů (**Continue until maximum is reached**). V poli **Maximum Number of Cases** zadáme požadovaný počet případů. Pod položkou **Save** v seznamu **Select an Item** je defaultně nastaveno, že se nasimulovaná data zobrazí v novém datovém okně (**Save the simulated data as a new data file, A new dataset**), jehož název lze zadat v poli **Dataset name**.



**Simulace dat bez využití predikčního modelu**

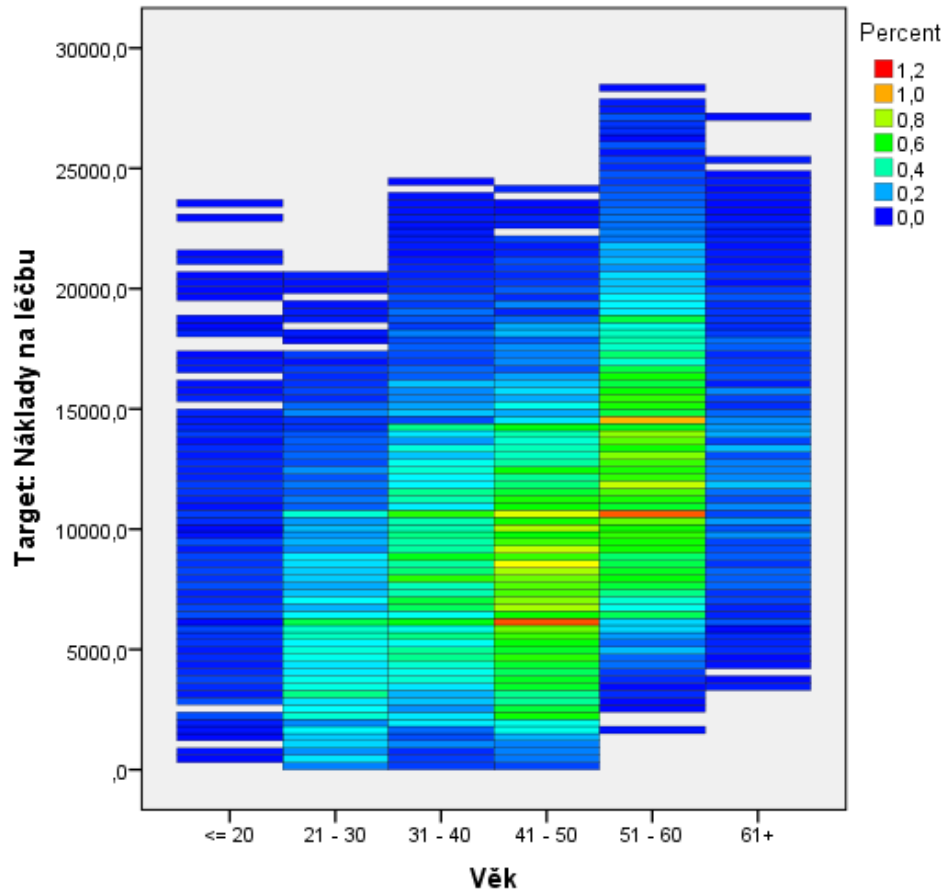
Výstup procedury je tvořen pivotními tabulkami a doprovázen grafy, které lze editovat v rozhraní **Graphboard Editor**.

Graf *Probability Density* je doplněn tabulkou obsahující polohu jezdců zobrazených v grafu a procento případů v jednotlivých intervalech určených těmito jezdcí. Jedná se o interaktivní výstup, v němž lze polohu jezdců měnit tlačítkem **Chart Options** (k dispozici v rozhraní *Graphboard Editor* po poklepnutím myši na graf). Pomocí tohoto tlačítka může uživatel vložit do grafu také referenční čáry.



Rozhraní Graphboard Editor, tlačítko Chart Options

Teplotní graf (*Heat map*) je sestaven z výstupní proměnné umístěné na osu Y a vstupní proměnné umístěné na osu X. Škála číselné proměnné je rozdělena do velkého počtu intervalů. Barva jednotlivých intervalů v kategoriích určených kategorizovanou proměnnou je dána procentem nasimulovaných případů spadajících do těchto intervalů.



Výstup procedury Simulation, teplotní graf (heat map)

## Simulace

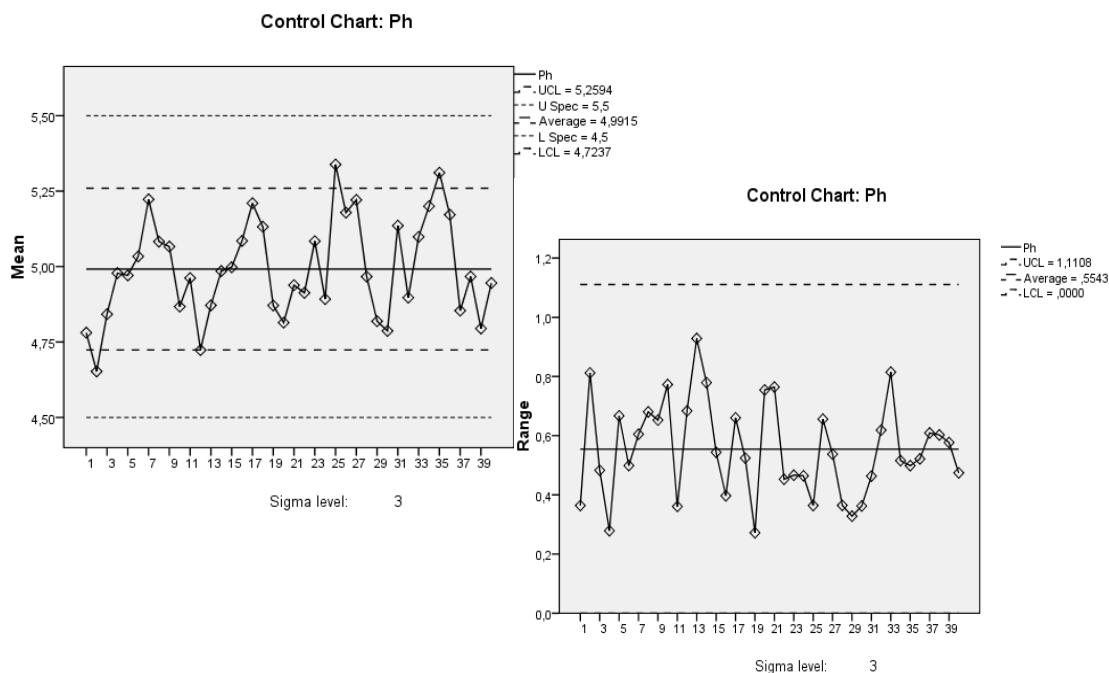
- Analyze
  - Simulations

## 1.14 Kontrola kvality (Quality Control)

### 1.14.1 Řídící diagramy (Control Charts)

Řídící diagramy zobrazují výstup procesu v průběhu času. Podle typu procesu vybereme vhodný typ řídicího diagramu:

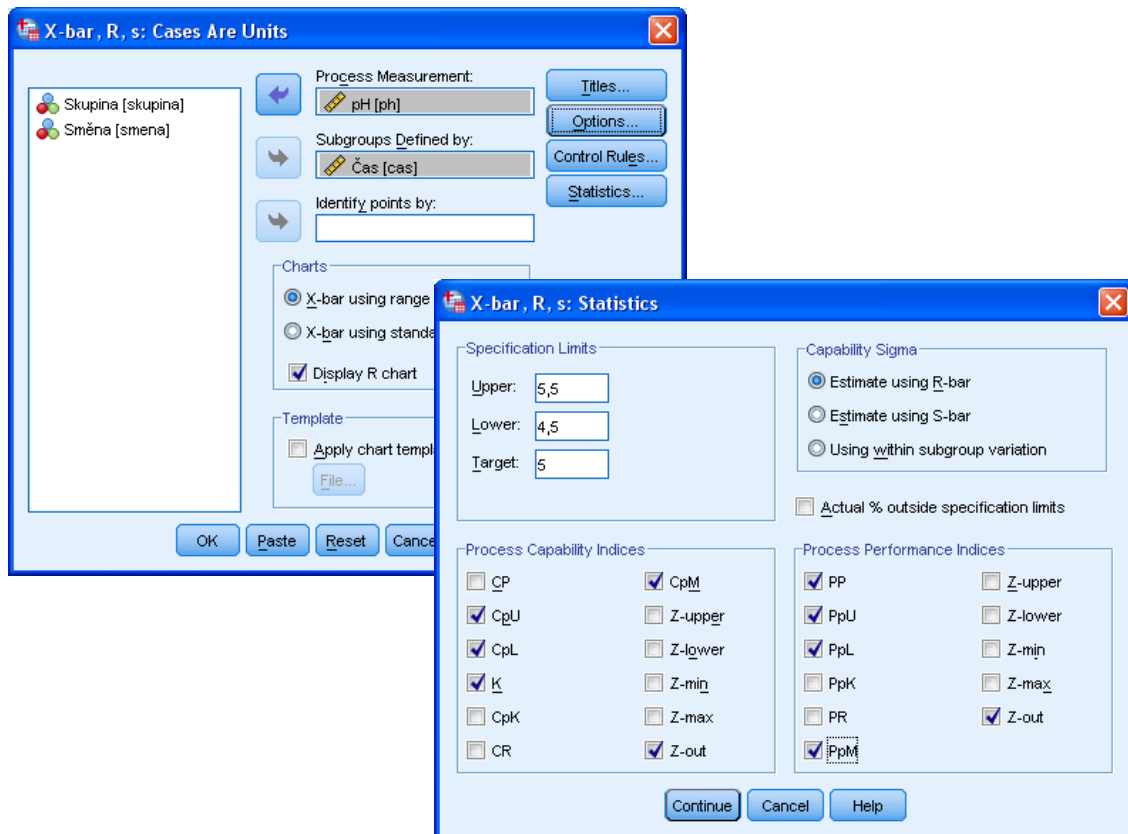
- **X-Bar, R, s**
- **Individuals, Moving Range**
- **p, np**
- **c, u.**



#### Řídící diagramy – kontrola průměru a rozpětí

Pro procesy s číselnými výstupy jsou k dispozici grafy **X-Bar, R chart** a **s chart** pro vícenásobná měření, nebo **Individuals** a **Moving Range** pro samostatná měření. Pokud je v procesu zaznamenáván počet defektů, pak lze použít grafy **p chart** a **np chart** pro počet defektních výrobků nebo **c chart** a **u chart** pro počet defektů na výrobku. Dialogová okna pro zadávání různých typů řídicích diagramů se vzájemně liší.





### Řídící diagramy – kontrola průměru a rozpětí

## Řídící diagramy

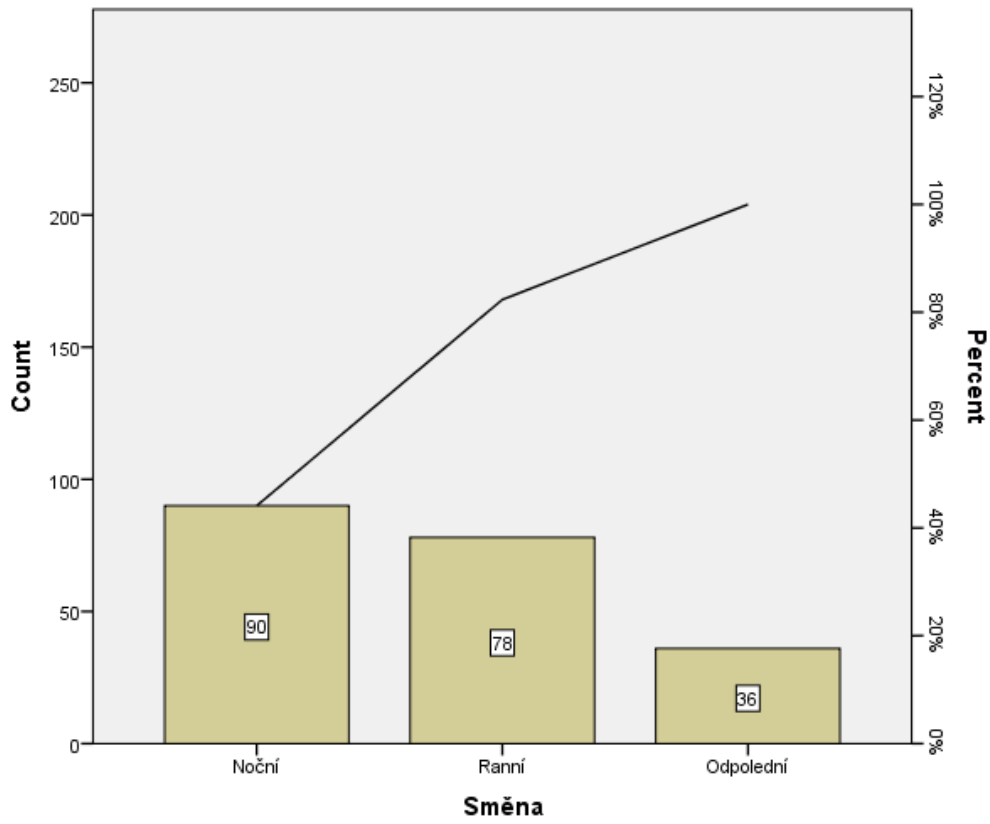
- Analyze
  - Quality Control
    - Control Charts

### 1.14.2 Paretův graf (Pareto Charts)

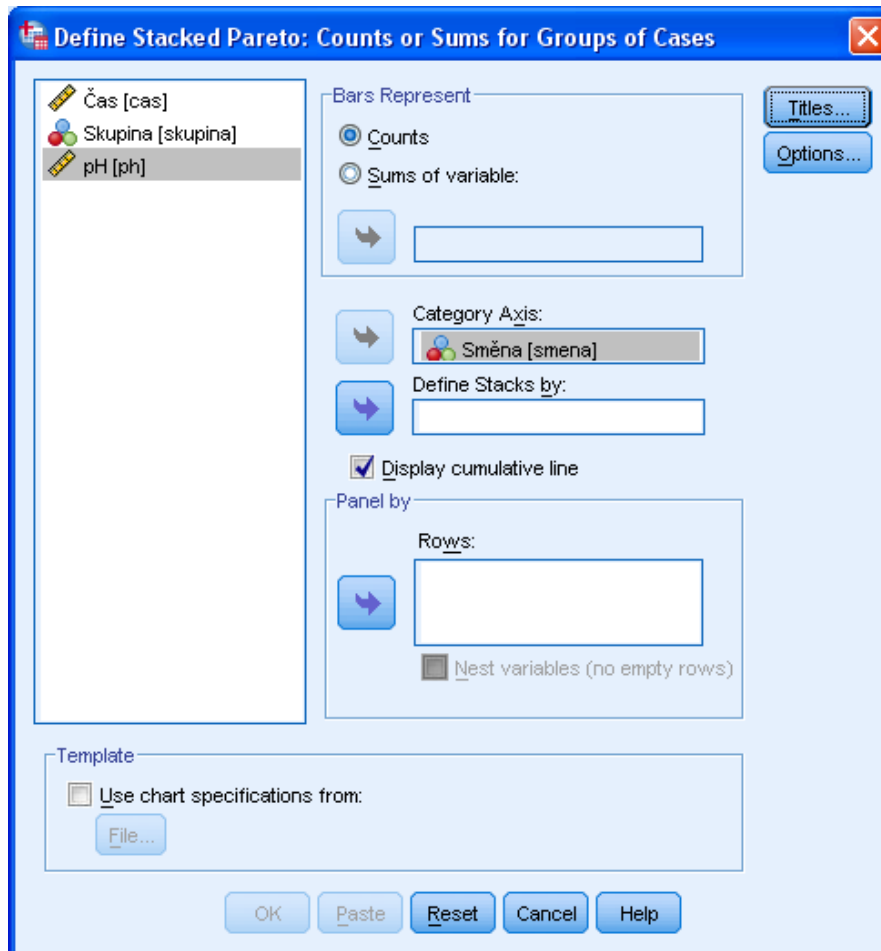
Paretův graf vzniká kombinací dvou grafů – sloupcového grafu uspořádaného podle četností kategorií a čárového grafu vyjadřujícího kumulativní procenta. Měřítka na svislé ose se pro každý graf liší. U skládaného grafu jsou sloupce navíc skládány na sebe – jednotlivé kategorie jsou odlišeny barvou a znázorněny v jednom sloupci nad sebou.

V úvodním dialogovém okně zadáme požadovaný typ grafu: jednoduchý (**Simple**) nebo skládaný (**Stacked**) a určíme, co mají sloupce zobrazovat: četnosti nebo úhrny pro skupiny případů (**Counts or sums for groups of cases**), úhrny pro samostatné proměnné (**Sums of separate variables**) nebo hodnoty jednotlivých případů (**Values of individual cases**).

Následující dialogové okno se mírně odlišuje podle typu zadávaného grafu, princip zadávání je však shodný jako u ostatních typů grafů (viz Skripta IBM SPSS Statistics Base 24, Grafika).



*Paretův graf*



**Zadávání Paretova grafu**

### **Paretův graf**

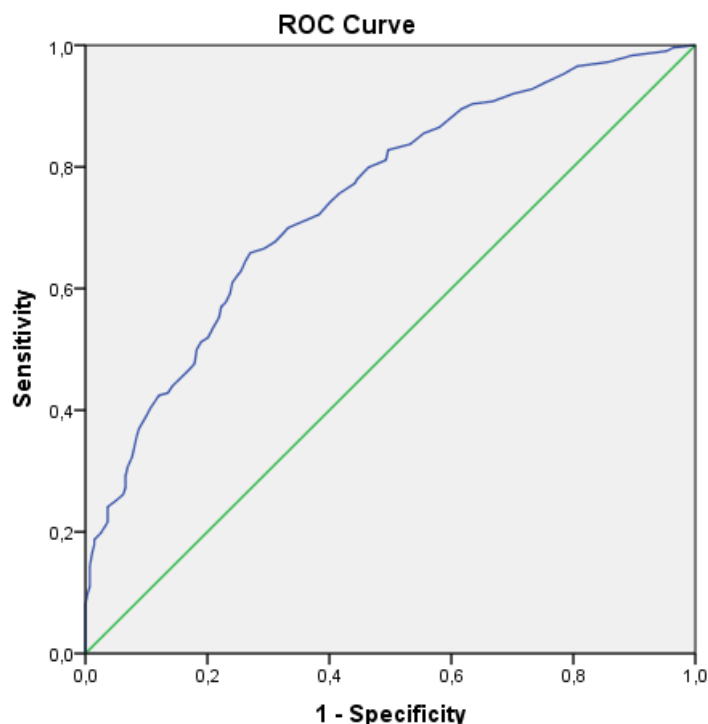
- Analyze
  - Quality Control
    - Pareto Charts

### **Slovník anglických pojmů**

<i>bar</i> .....	<i>sloupcový graf</i>
<i>case</i> .....	<i>případ</i>
<i>category</i> .....	<i>kategorizovaná</i>
<i>chart</i> .....	<i>graf</i>
<i>cluster</i> .....	<i>skupinka</i>
<i>control</i> .....	<i>řízení procesů</i>
<i>cumulative</i> .....	<i>kumulativní</i>
<i>error</i> .....	<i>chyba</i>
<i>footnote</i> .....	<i>poznámka</i>
<i>line</i> .....	<i>čára</i>
<i>measurement</i> .....	<i>míra, měření</i>
<i>plot</i> .....	<i>graf</i>
<i>process</i> .....	<i>proces</i>
<i>range</i> .....	<i>rozpětí</i>
<i>sequence</i> .....	<i>řada</i>
<i>simple</i> .....	<i>jednoduchý</i>
<i>stacked</i> .....	<i>skládáný</i>
<i>standard deviation</i> .....	<i>směrodatná odchylka</i>
<i>subgroup</i> .....	<i>podskupina</i>
<i>template</i> .....	<i>šablona</i>
<i>title</i> .....	<i>název</i>
<i>variable</i> .....	<i>proměnná</i>

## 1.15 ROC křivka (ROC Curve)

ROC křivka (**Receiver Operating Characteristics**) slouží ke zjišťování, jak dobře lze předpovídat hodnotu dichotomické proměnné pomocí číselné proměnné. Graf znázorňuje závislost dvou charakteristik – senzitivity a specifčnosti – pro různě zvolené hranice, kde senzitivita vyjadřuje procento správných pozitivních předpovědí ze všech skutečných pozitivních hodnot, specifčnost procento správných negativních předpovědí ze všech skutečných negativních hodnot. Pro lepší přehlednost se na vodorovnou osu vynáší hodnota **1 - specifčnost**, na svislou **senzitivita**.



Diagonal segments are produced by ties.

### ROC křivka

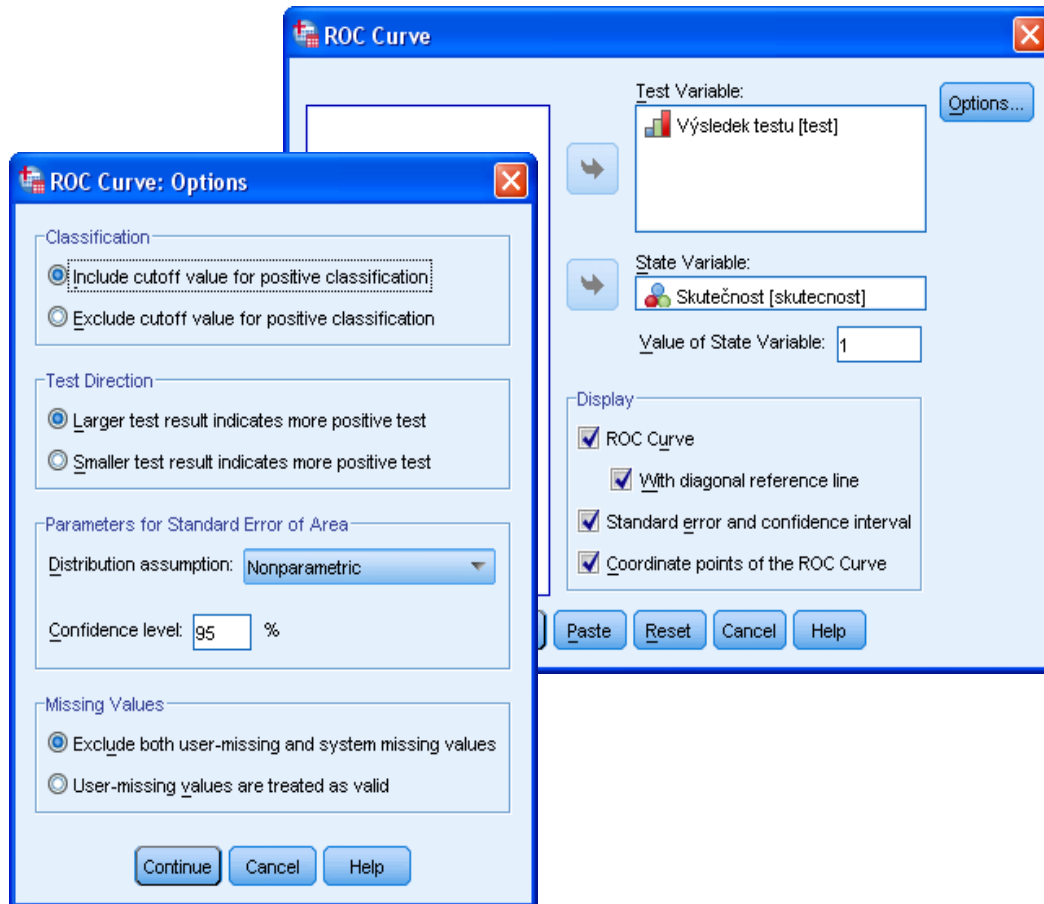
Pro dosažení co nejlepší předpovědi by měly oba ukazatele dosahovat co největší hodnoty. Model je tedy tím lepší, čím výše se nachází ROC křivka nad diagonálou. Křivka pod diagonálou signalizuje, že číselná proměnná předpovídá hodnoty dichotomické proměnné lépe v obráceném pořadí.

Při zadávání ROC křivky přeneseme do pole **Test Variable** testovanou číselnou proměnnou, do pole **State Variable** predikovanou dichotomickou proměnnou. V poli **Value of State Variable** zadáme hodnotu dichotomické proměnné (obvykle 1), kterou předpovídáme, nabyde-li testovaná číselná proměnná hodnotu nad příslušnou hranicí.

V oblasti **Display** dále označíme, zda chceme zobrazit ROC křivku (**ROC Curve**), úsečku znázorňující diagonálu (**With diagonal reference line**), tabulku obsahující plochu pod křivkou, standardní chybu a interval spolehlivosti (**Standard error and confidence interval**) a tabulku obsahující zlomové body s odpovídajícími hodnotami specifčnosti a senzitivity (**Coordinate point of the ROC Curve**).

Tlačítkem **Options** zadáváme některé doplňující údaje jako například způsob zacházení s chybějícími hodnotami (**Missing Values**), metodu výpočtu standardní chyby a požadovanou hladinu spolehlivosti

(Parameters for Standard Error of Area). Dále můžeme změnit směr testu (**Test Direction**) – hodnota uvedená v poli Value of State Variable bude predikována pro hodnoty číselné proměnné pod hranicí.



ROC křivka

### ROC křivka

- Analyze
  - ROC Curve

**Slovník anglických pojmů**

<i>assumption</i>	<i>předpoklad</i>
<i>chart</i>	<i>graf</i>
<i>classification</i>	<i>klasifikace</i>
<i>confidence interval</i>	<i>interval spolehlivosti</i>
<i>diagonal</i>	<i>diagonální</i>
<i>distribution</i>	<i>rozdělení</i>
<i>error</i>	<i>chyba</i>
<i>display</i>	<i>zobrazit</i>
<i>line</i>	<i>čára</i>
<i>missing</i>	<i>chybějící</i>
<i>parameter</i>	<i>parametr</i>
<i>plot</i>	<i>graf</i>
<i>standard error</i>	<i>standardní chyba</i>
<i>test direction</i>	<i>směr testu</i>
<i>value</i>	<i>hodnota</i>
<i>variable</i>	<i>proměnná</i>

## 1.16 Časoprostorové modelování (Spatial and Temporal Modeling)

### 1.16.1 Geoprostorové modelování (Spatial Modeling)

Geoprostorové modelování je určeno k hledání vztahů v datech obsahujících geoprostorové informace (tj. informace z map). Kromě toho lze do analýz zahrnout také časovou komponentu. Procedura nabízí dvě různé metody: zobecněná geoprostorová asociační pravidla (Geospatial Association Rules) a časoprostorové modelování (Spatio-Temporal Prediction).

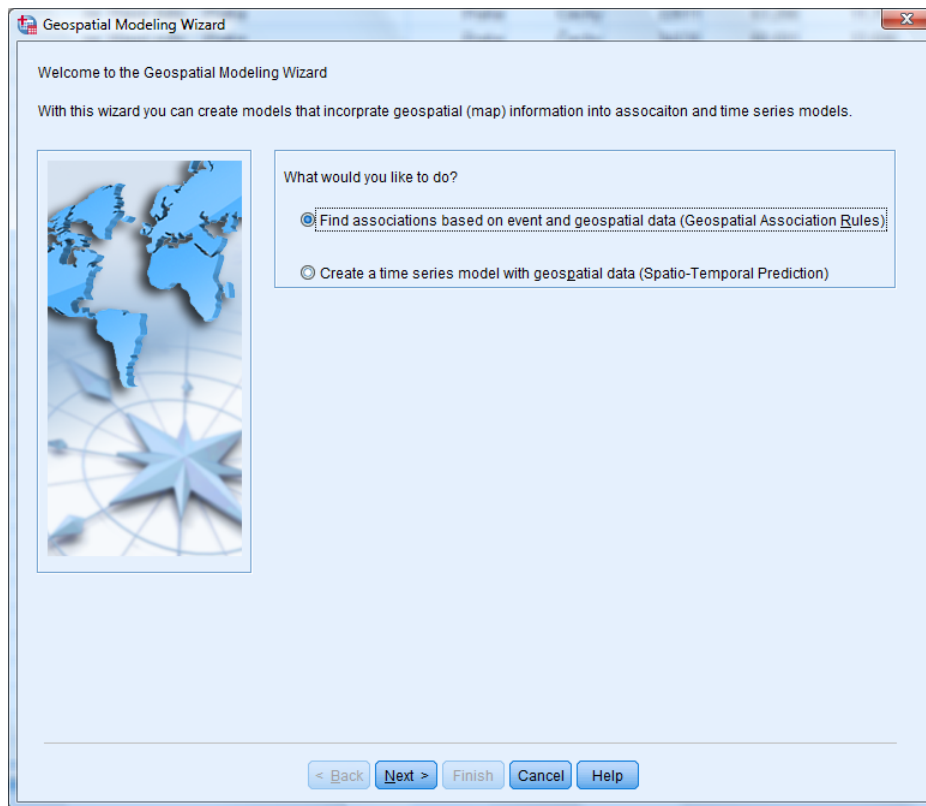
Geoprostorová asociační pravidla se užívají k hledání souvislostí v datech na základě prostorových i neprostorových vlastností. Například v datech o kriminalitě lze identifikovat vzory chování vycházející z polohy a demografických charakteristik místa činu. Na základě těchto vzorů se vytvoří pravidla, která předpovídají, kde se s vyšší pravděpodobností vyskytují určité typy kriminálních případů. Tato pravidla mají tvar implikace ( $X \Rightarrow Y$ ), která platí s určitou odhadnutou spolehlivostí.

Časoprostorové predikce využívají data o poloze, jednu nebo více časových informací a další vstupní proměnné pro predikci cílové proměnné. Závislou proměnnou může být hustota událostí v daných prostorových objektech (okresy, silnice, řeky apod.) nebo určitá charakteristika prostorových objektů v čase. Data obsahují pro každý prostorový objekt více řádků, které odpovídají hodnotám cílové proměnné i nezávislých proměnných pro každý časový interval. Vztahy mezi prostorovými objekty na základě modelu jsou vyjádřeny prostorovou kovariancí. Metoda rovněž umožňuje určit nejvíce podobné prostorové objekty a vytvářet jejich seskupení.

V úvodním dialogovém okně průvodce časovým a prostorovým modelováním specifikujeme v části **What would you like to do?** požadovaný typ analýzy:

- zobecněná geoprostorová asociační pravidla (**Find associations based on event and geospatial data (Geospatial Association Rules)**),
- časoprostorové predikce (**Create a time series model with geospatial data (Spatio-Temporal Prediction)**).





Úvodní dialog průvodce časovým a prostorovým modelováním (výběr metody)

Geoprostorové analýzy vyžadují mapové podklady, které definují geografické objekty a jejich souřadnice pro vykreslování do mapy a umožňují propojení s daty. Zadáváme je v následujícím dialogovém okně označeném **Add maps to the model**. Pro zobecněná geoprostorová asociační pravidla může být mapových podkladů více, pro časoprostorové predikce pouze jeden. Způsob zadávání je shodný u obou metod.

Mapový podklad je třeba mít k dispozici ve formátu od společnosti ESRI jako tři soubory SHP, DBF a PRJ. Pokud však soubor PRJ chybí, lze ho vynechat a v dalším dialogu specifikovat souřadnicový systém. Všechny tři soubory musí mít stejný název a musí být uloženy ve stejné složce nebo v jednom ZIP souboru rovněž se stejným názvem. Procedura také umožňuje vytvořit na základě těchto vstupů jeden soubor ve formátu MPLAN pro pozdější snadnější užívání (může obsahovat specifikace i více mapových podkladů).

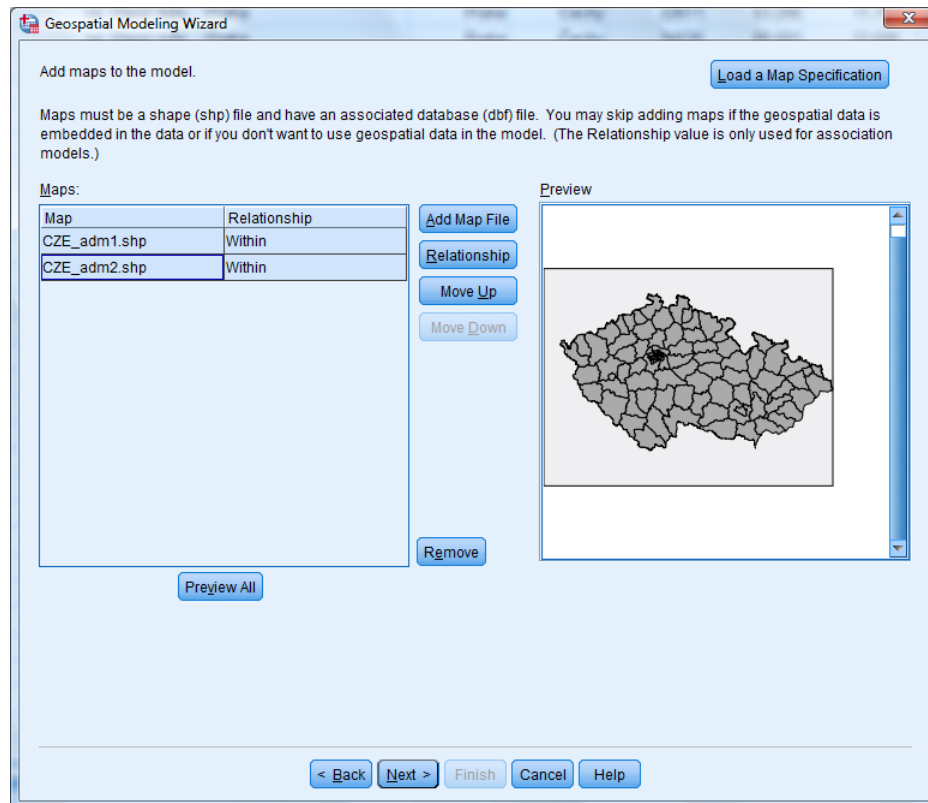
Pokud máme k dispozici přímo mapový podklad ve formátu MPLAN, zadáme ho tlačítkem **Load a Map Specification**.

V ostatních případech načteme mapové podklady tlačítkem **Add Map File** a určíme příslušný soubor ve formátu SHP nebo ZIP (odpovídající soubory DBF a PRJ již není třeba specifikovat).

V poli **Maps** se potom zobrazuje seznam již definovaných mapových podkladů, v poli **Preview** jejich náhled.

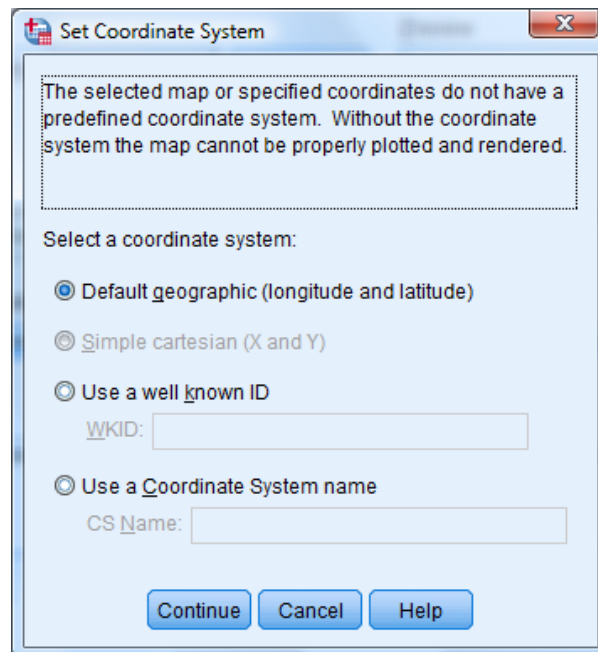
Pro zobecněná geoprostorová asociační pravidla definujeme tlačítkem **Relationship** vztah mezi zkoumanými událostmi a geoprostorovými objekty (podrobný popis dialogu tlačítka viz níže). Pro časoprostorové predikce není tlačítko k dispozici.

Pořadí mapových vrstev je určeno jejich pořadím v seznamu (první položka seznamu představuje nejnižší mapovou vrstvu). Pořadí lze upravovat tlačítky pro posouvání nahoru (**Move up**) a dolů (**Move down**). Tlačítkem **Remove** označenou položku odstaníme. Tlačítko **Preview All** zobrazí náhled všech vrstev.



**Definice mapových podkladů**

V případě, že při načítání mapového podkladu není dodán soubor PRJ, jsme v následujícím dialogu **Set Coordinate System** vyzváni ke specifikaci souřadnicového systému. Ten může být charakterizován čtyřmi různými způsoby: zeměpisnou šířkou a délkou (**Default geographic (longitude and latitude)**), kartézskými souřadnicemi na osách X a Y (**Simple Cartesian (X and Y)**), kódem WKID označujícím typ projekce (**Use a well known ID**) nebo názvem projekce (WKT) uvedeným v uvozovkách (**Use a Coordinate System name**).

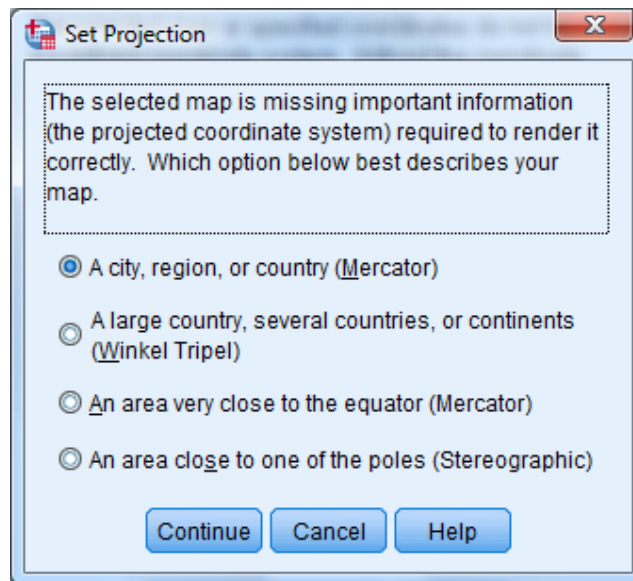


**Volba souřadnicového systému**

Jestliže z mapového podkladu nelze určit typ kartografické projekce (nejčastěji pokud chybí soubor PRJ), je třeba ji specifikovat v dialogovém okně **Set Projection**. Kartografická projekce je metoda zobrazování trojrozměrného zemského povrchu do roviny.

Podle toho, jakou oblast mapa zobrazuje, zvolíme jednu ze čtyř možností:

- **A city, region, or country (Mercator)** – pro mapy měst, regionů, jednotlivých zemí nebo jiných menších oblastí je vhodné Mercatorovo zobrazení. Je založeno na projekci ze středu glóbu na válcovou plochu jejíž osa je obvykle rovnoběžná se zemskou osou a dotýkající se glóbu na rovníku. Po rozvinutí do roviny vznikne pravoúhlá síť poledníků a rovnoběžek.
- **A large country, several countries, or continents (Winkel Tripel)** – pro velké země, skupiny zemí, kontinenty nebo jiná rozsáhlá území je vhodným typem projekce Winkel Tripel.
- **An area very close to the equator (Mercator)** – Mercatorovo zobrazení se užívá rovněž pro oblasti v blízkosti rovníku. Vzhledem k tomu, že válcová plocha se na rovníku dotýká glóbu, dochází zde tímto typem projekce k minimálnímu zkreslení.
- **An area close to one of the poles (Stereographic)** – pro zobrazování oblastí v blízkosti pólů se užívá stereografická projekce. Jedná se o projekci z protilehlého pólu do roviny, která se dotýká zobrazovaného pólu.



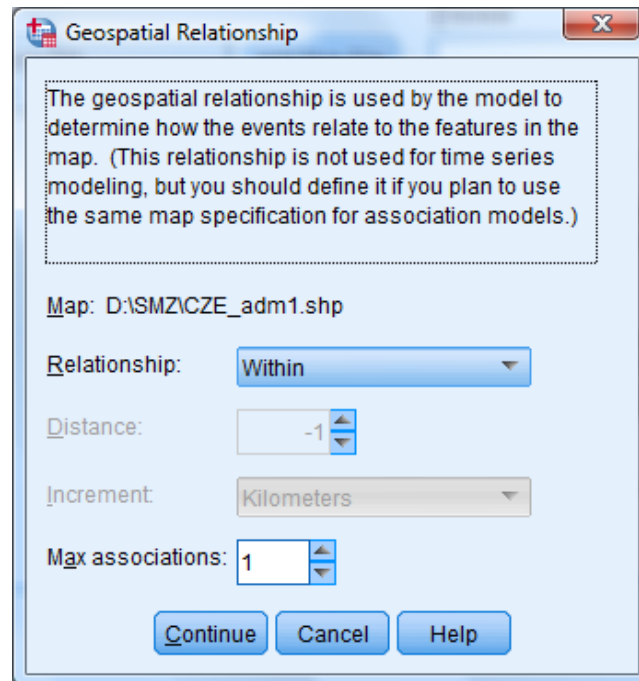
*Nastavení kartografické projekce*

Pro zobecněná geoprostorová asociační pravidla je nutné při zadávání mapových podkladů v dialogovém okně *Add maps to the model* definovat také tlačítkem *Relationship* vztah mezi zkoumanými událostmi a geoprostorovými objekty. Tato nastavení se vztahují pouze k datovým zdrojům asociovaným s mapami, které jsou v následujícím dialogovém okně *Data Sources* specifikovány jako kontextová data.

V dialogové okně **Geospatial Relationship** v seznamu **Relationship** určíme typ vztahu:

- **Within** – událost se odehrála uvnitř prostorového objektu (oblasti na mapě),
- **Close** – událost se odehrála v blízkosti určitého bodu nebo prostorového objektu (oblasti na mapě),
- **Contains** – oblast, ve které se událost odehrála, zahrnuje daný prostorový objekt,
- **Cross** – pro více map určuje místo, kde se protínají křivky (například silnice, řeky, železniční tratě apod.) nebo oblasti z různých map,
- **Intersects** – určuje místo, kde se protínají křivky (například silnice, řeky, železniční tratě apod.) nebo oblasti z různých map,
- **North of/South of/ East of/ West of** – událost se odehrála v oblasti severně/ jižně/ východně/ západně od specifikovaného bodu na mapě.

Pro některé typy vztahů je nutné dále specifikovat maximální vzdálenost od objektu (**Distance**) a jednotky (**Increment**), v nichž je uvedena (míle (*Miles*), stopy (*Feet*), kilometry (*Kilometers*) nebo metry (*Meters*)). Pole **Max associations** udává maximální počet asociací pro jednu událost.



*Specifikace geoprostorových vztahů*

Dialogové okno **Data Sources** je určeno ke specifikaci datových zdrojů. Datovým zdrojem mohou být soubory DBF, které jsou součástí mapových podkladů nebo otevřené datové soubory či externí soubory IBM SPSS Statistics (SAV). Dialogové okno se mírně odlišuje podle zvolené metody.

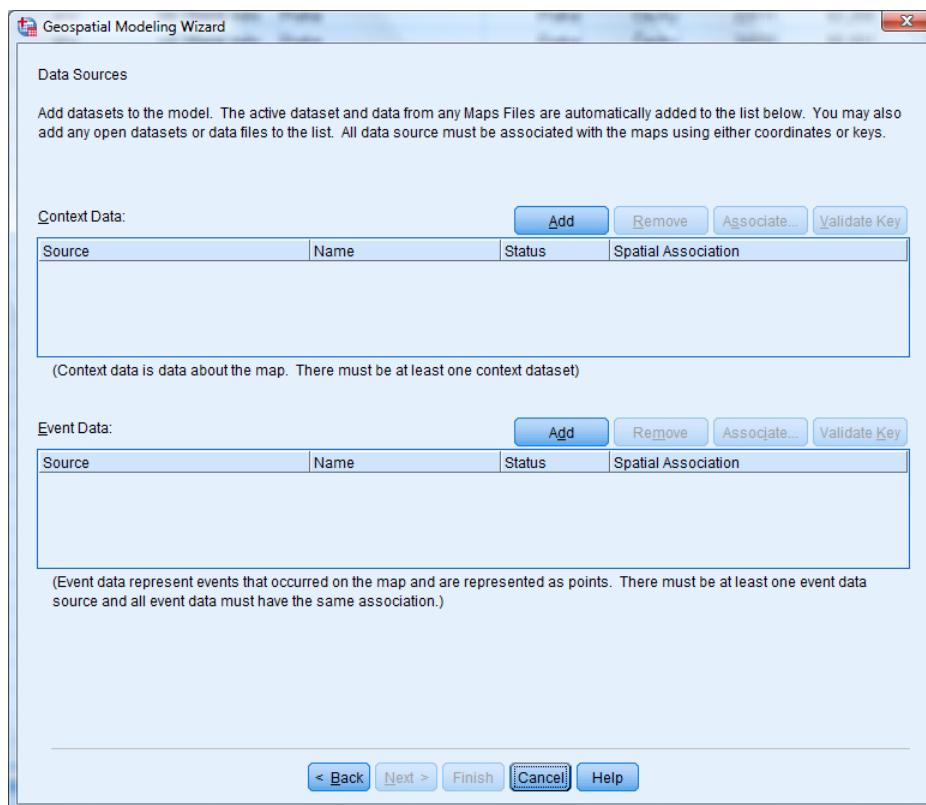
### Pro zobecněná geoprostorová asociační pravidla

V části **Context data** definujeme kontextová data, která popisují vlastnosti prostorových objektů na mapě nebo představují vstupy pro modelování.

V části **Event Data** specifikujeme data, která obsahují informace o zkoumaných událostech (kriminální případy, nehody apod.).

Datový zdroj přidáme do seznamu tlačítkem **Add**, odstraníme ho tlačítkem **Remove**. K propojení dat s mapovým podkladem je určeno tlačítko **Associate**, k ověření propojení na základě klíče slouží tlačítko **Validate Key**. Dialogy jednotlivých tlačítek jsou samostatně popsány níže.

Pro zobecněná geoprostorová asociační pravidla je třeba definovat alespoň jeden datový zdroj v části *Event Data*. Zároveň musí mít všechny datové soubory zadané v části *Event Data* stejný způsob asociace s identifikátory na mapě (souřadnice nebo klíčové hodnoty) a v případě klíčových hodnot musí být události asociovány se stejným typem mapových objektů (polygony, body, čáry apod.)

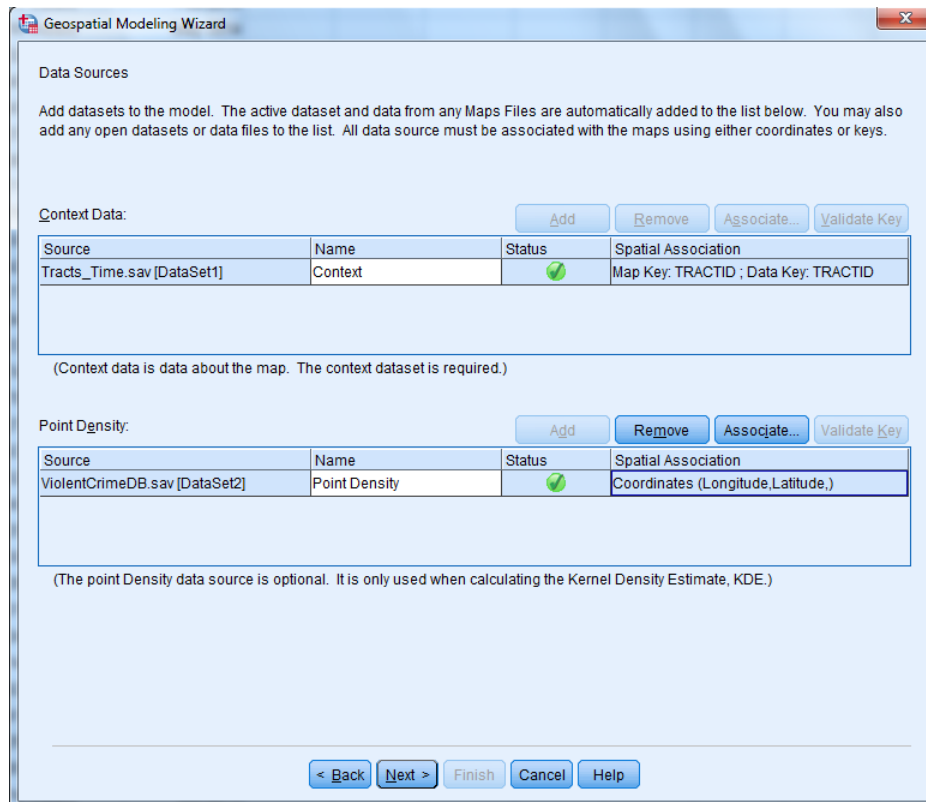


**Datové zdroje pro zobecněná geoprostorové asociační pravidla**

### Pro časoprostorové predikce

Dialogové okno pro časoprostorové modelování se liší tím, že obsahuje místo části *Event Data* část **Point Density**. Zde specifikujeme data, která obsahují časový interval a souřadnice událostí (zeměpisnou šířku a délku) pro odhad hustoty událostí.

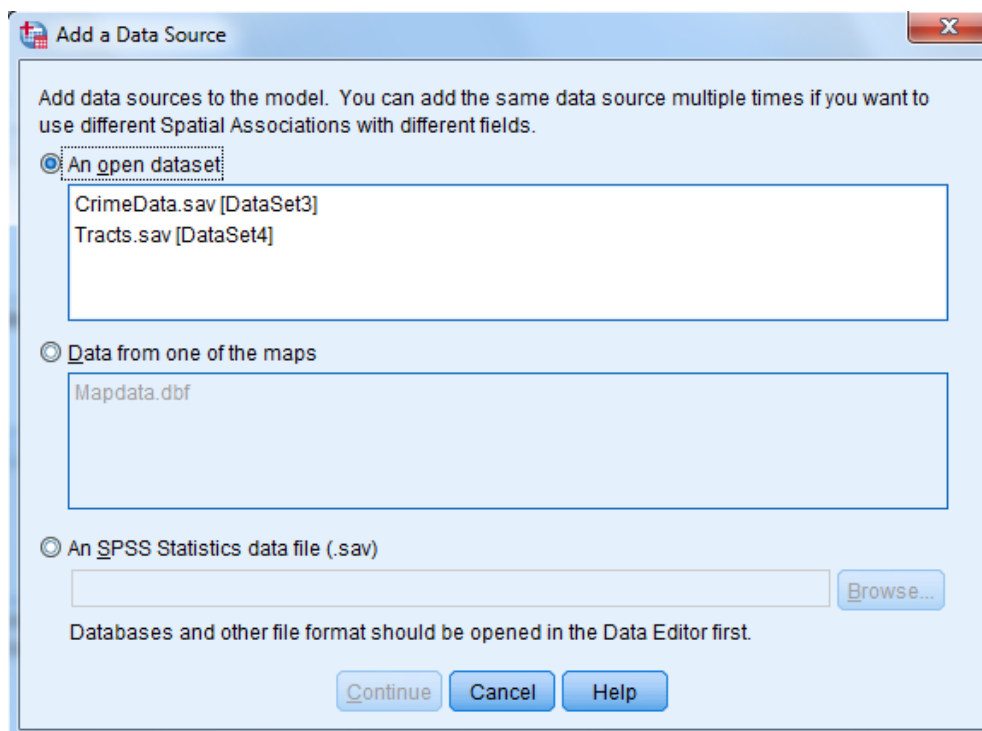
Pro časoprostorové modelování je nutné vždy definovat kontextová data. Pokud je pouze jeden datový zdroj (datový soubor bez asociované mapy), musí obsahovat souřadnicové hodnoty. V případě dvou datových souborů musí jeden obsahovat kontextová data a druhý data o hustově bodů (*Point Density*). Více než dva datové zdroje do analýzy nelze zahrnout.



**Datové zdroje pro časoprostorové predikce**

V dialogovém okně tlačítka **Add (Add a Data Source)** specifikujeme datový zdroj. Pro zadání vybereme jednu ze tří možností:

- **An open dataset** – otevřený datový soubor,
- **Data from one of the maps** – soubor ve formátu DBF, který je součástí mapových podkladů,
- **An SPSS Statistics data file (.sav)** – datový soubor IBM SPSS Statistics (SAV).

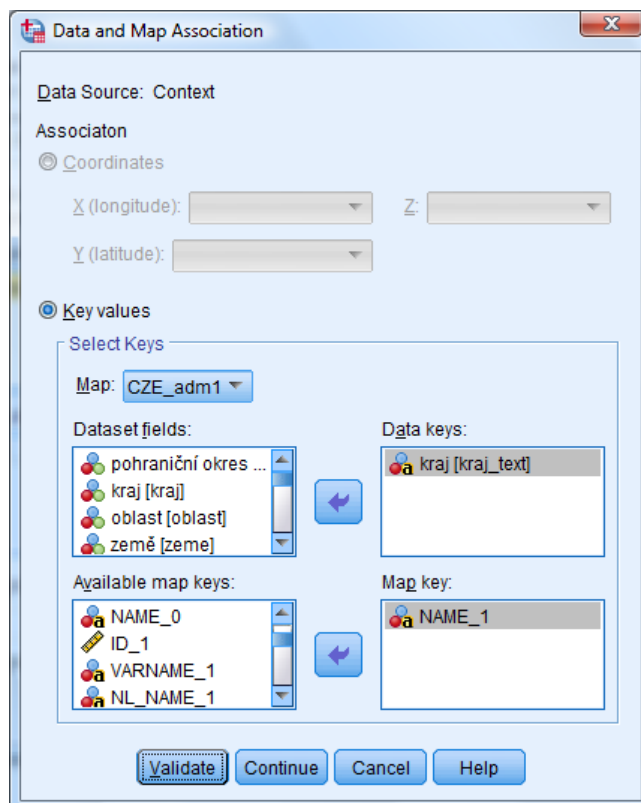


**Načítání dat**

Po specifikaci datového zdroje se zobrazí dialogové okno **Data and Map Association**, v němž definujeme propojení dat s mapou. Každý datový zdroj musí obsahovat jeden nebo více identifikátorů, které toto propojení umožní.

Propojení lze realizovat na základě souřadnic (**Coordinates**) nebo klíčových hodnot (**Key values**). V prvním případě je třeba definovat pole, která reprezentují kartézské souřadnice **X (longitude)** a **Y (latitude)** (obvykle zeměpisná šířka a délka). Pro zobecněná geoprostorová asociační pravidla lze zadat také souřadnici **Z**. Ve druhém případě definujeme klíčové proměnné v datech (**Data keys**) i v mapovém podkladu (**Map key**), jejichž hodnoty se musí shodovat, aby bylo propojení možné. Shodu ihned ověříme tlačítkem **Validate**.

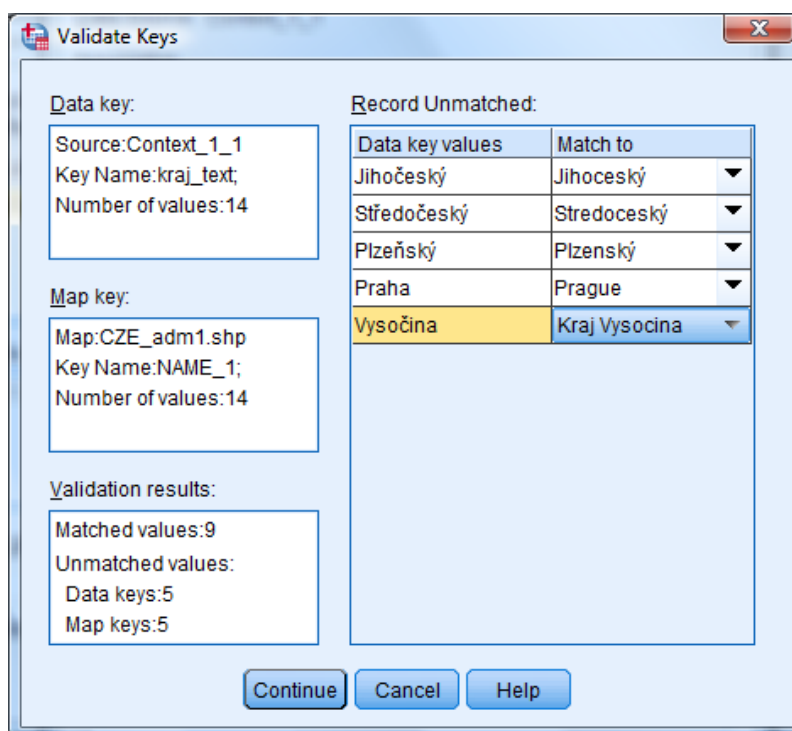




*Propojení dat s mapovým podkladem*

Dialogové okno **Validate Key** vyvoláme z předchozího dialogu tlačítkem *Validate*, nebo z dialogového okna *Data Sources* tlačítkem *Validate Key*. Ověřuje shodu hodnot klíčových proměnných v datech a v mapovém podkladu.

Pole **Data key**, **Map key** a **Validation results** shrnují postupně informace o klíčové proměnné v datech, klíčové proměnné v mapovém podkladu a výsledcích validace. Pole **Record Unmatched** obsahuje nespárované dvojice hodnot a návrh jejich spárování, které lze ručně upravit.



**Ověření shody identifikátorů**

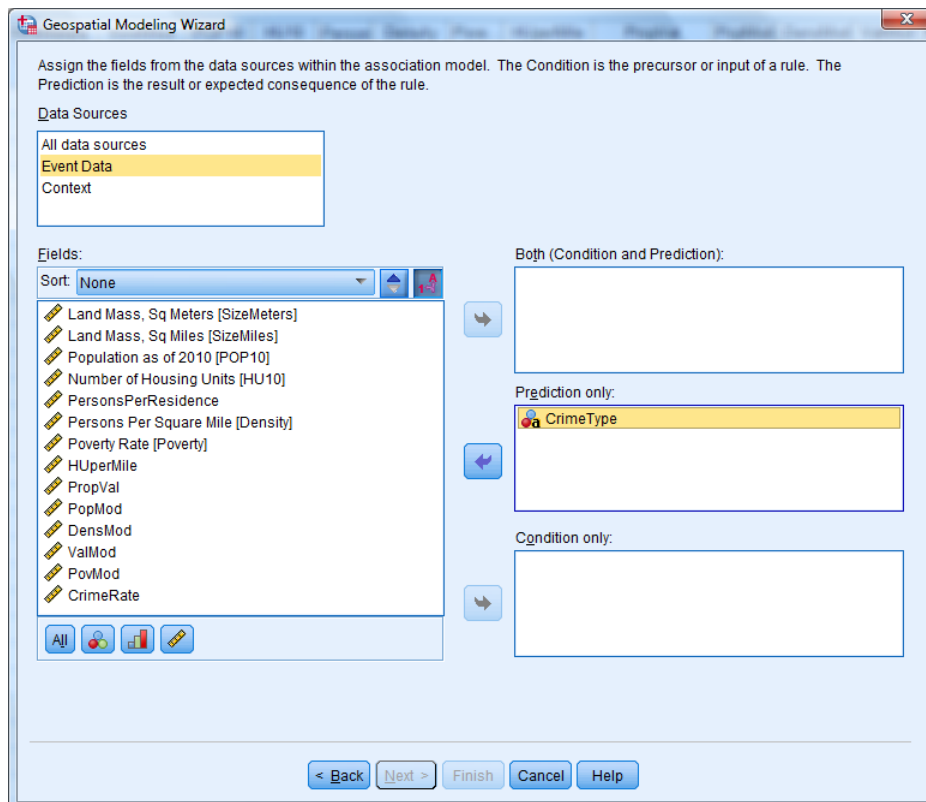
Další nastavení dialogů se pro zobecněná geoprostorová asociační pravidla a časoprostorové predikce liší, proto v popisu pokračujeme odděleně.

### 1.16.2 Zobecněná geoprostorová asociační pravidla (Geospatial Association Rules)

V následujícím dialogovém okně specifikujeme proměnné pro vytváření zobecněných geoprostorových asociačních pravidel.

Pole **Data Sources** obsahuje seznam všech načtených datových souborů. Na základě této volby se v poli **Fields** zobrazují všechny proměnné, nebo jen z vybraného datového souboru. Proměnné lze řadit (**Sort**) či filtrovat podle typu pomocí tlačítek pod seznamem.

Do pole **Both (Condition and Prediction)** zadáme proměnné, které mohou být součástí podmínek a zároveň mají být předpovídány asociačním pravidlem. Proměnné v poli **Prediction only** budou pouze předpovídány, proměnné v poli **Condition only** mohou být pouze součástí podmínek.



**Zadávání proměnných pro zobecněná geoprostorová asociační pravidla**

Následující dialogové okno **Choose settings for building the model and the output** je určeno ke specifikaci modelu a výstupů. Skládá se z více záložek, mezi kterými přepínáme v seznamu **Select an item**.

Na záložce **Output** nastavíme požadované výstupy ve formě pivotních tabulek nebo interaktivních výstupů (včetně map).

V části **Rules Tables** lze označit tabulky obsahující přehled pravidel spolu s hodnotami kritérií kvality. Pravidla jsou řazena sestupně podle vybraného kritéria, každá tabulka však zobrazuje také ostatní kritéria. Kvalitu asociačních pravidel vyhodnocujeme na základě těchto pěti kritérií:

- **Confidence** – procento správných predikcí na základě asociačního pravidla (podíl případů splňujících asociační pravidlo vzhledem k těm, které splňují jeho podmínku).
- **Rule Support** – procento pozorování splňujících asociační pravidlo vzhledem ke všem případům (pokud má například asociační pravidlo tvar  $(X=1) \Rightarrow (Y=2)$ , splňují ho případy s hodnotami  $X=1$  a  $Y=2$ ).
- **Lift** – podíl procenta správných predikcí na základě asociačního pravidla (*Confidence*) vzhledem k celkovému procentu případů splňujících předpověď pravidla. Vyjadřuje kolikrát se zlepší předpověď použitím asociačního pravidla vzhledem k náhodě.
- **Condition support** – podíl pozorování splňujících podmínku asociačního pravidla vzhledem ke všem případům.
- **Deployment** – procento pozorování, která nesplňují asociační pravidlo a splňují jeho podmínku vzhledem ke všem případům (tj.  $Deployment = (1 - Confidence) * Condition support$  nebo také  $Deployment = Condition support - Rule Support$ ).

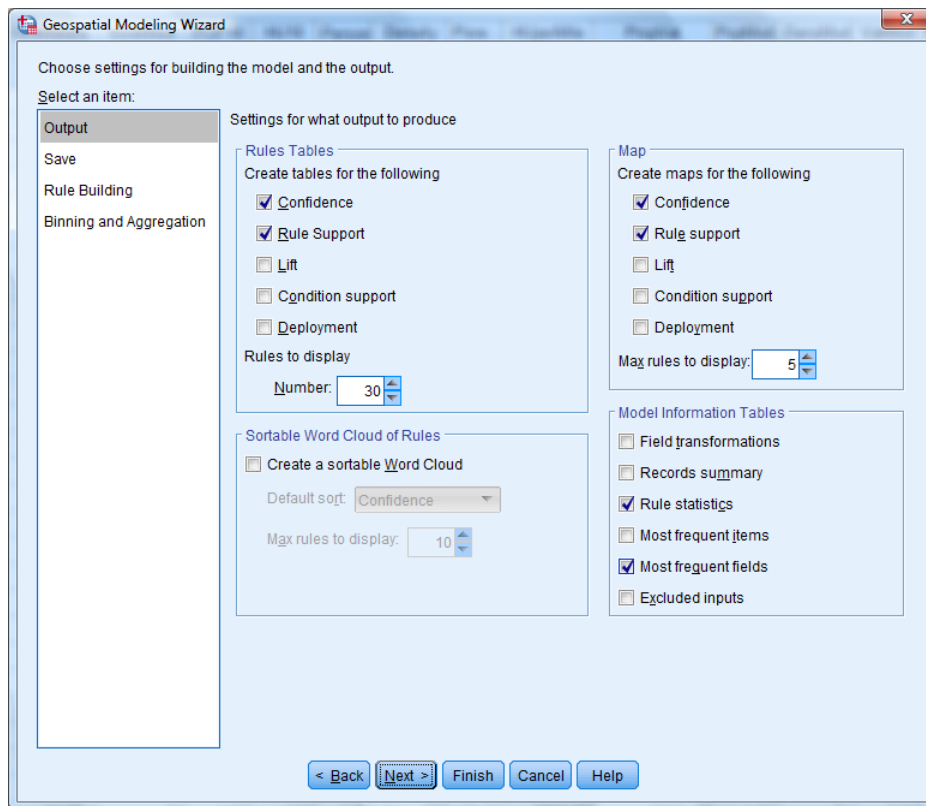
V poli **Number** lze nastavit maximální počet zobrazených nejlepších pravidel na základě zvoleného kritéria.

V části **Map** volíme interaktivní výstupy (přehled nejlepších pravidel ve formě sloupcového grafu a mapa) na základě vybraného kritéria kvality (**Confidence, Rule Support, Lift, Condition support a Deployment**). V poli **Max rules to display** nastavíme maximální počet zobrazených pravidel.

V části **Sortable Word Cloud of Rules** zadáváme zaškrtnutím políčka **Create a sortable Word Cloud** interaktivní výstup ve formě textového zápisu asociačních pravidel ve tvaru (*If ... Then ...*). Výstup zobrazuje nejlepší asociační pravidla na základě kritéria zvoleného v poli **Default sort** řazená podle tohoto kritéria. Volba kritéria zde však určuje pouze výchozí zobrazení, objekt zahrnuje seznamy nejlepších pravidel pro všech pět kritérií a lze mezi nimi interaktivně přepínat. V poli **Max rules to display** nastavíme maximální počet zobrazených pravidel.

V části **Model Information Tables** lze rozšířit výstup o další tabulky s informacemi o modelu:

- **Field transformation** – přehled provedených transformací proměnných,
- **Record Summary** – počet a procento zahrnutých a vynechaných případů,
- **Rule Statistics** – souhrnné statistiky kritérií kvality (průměr, minimum, maximum a směrodatná odchylka),
- **Most Frequent Items** – seznam položek nejčastěji se vyskytujících v pravidlech v rámci podmínky nebo předpovědi (například „*region = 4*“ nebo „*míra nezaměstnanosti > 5%*“),
- **Most Frequent Fields** – proměnné, které se nejčastěji vyskytují v pravidlech,
- **Excluded Inputs** – proměnné vyloučené z analýzy a důvod jejich vyřazení.



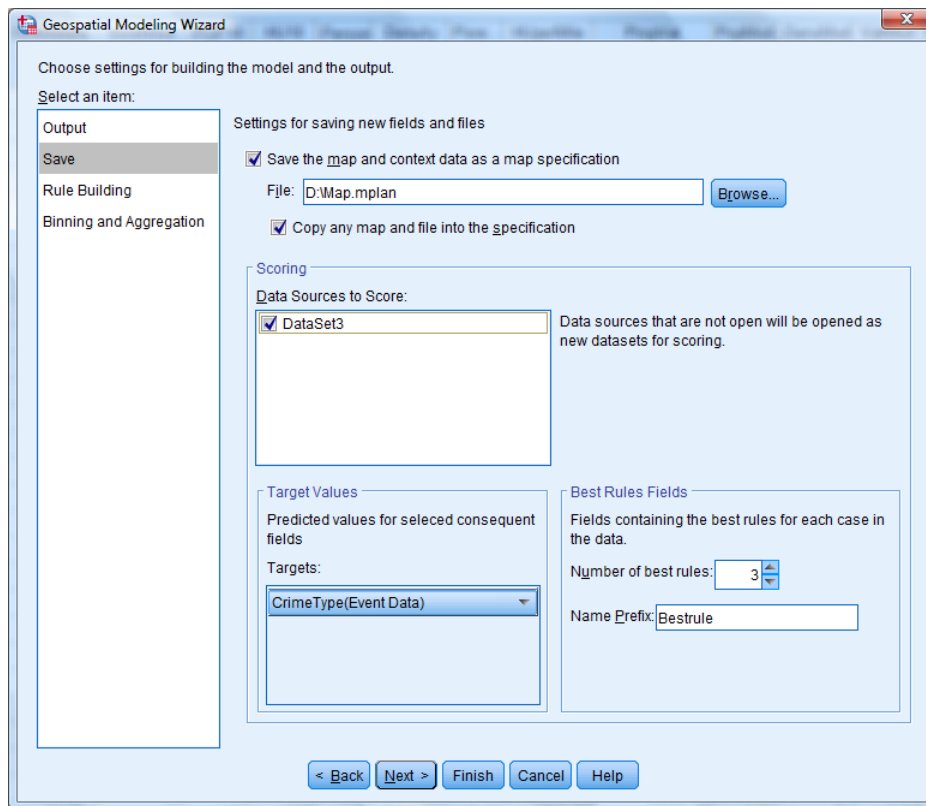
**Nastavení výstupů**

Na záložce **Save** nastavíme podrobnosti ohledně uložení mapových podkladů a nejlepších asociačních pravidel.

Zaškrťovací políčko **Save the map and context data as a map specification** uloží mapovou specifikaci do souboru *MPLAN*, který lze později užít při dalších analýzách. Tlačítkem **Browse** zadáme název a umístění souboru. Při zaškrtnutí políčka **Copy any map and data files into the specification** se do souboru uloží rovněž data ze všech mapových souborů, otevřených datových souborů i externích datových souborů užívaných v mapové specifikaci. Soubor *MPLAN* lze později užít pro definování mapových podkladů v dialogovém okně *Add maps to the model* tlačítkem *Load map specification* nebo v syntaxovém příkazu *SPATIAL ASSOCIATION RULES*.

V části **Scoring** nastavíme uložení výsledků nejlepších asociačních pravidel do datového souboru. V poli **Data Sources to Score** určíme datový soubor, do něhož se informace uloží. V části **Target values** zvolíme proměnné, jejichž předpovědi na základě nejlepších asociačních pravidel se uloží. Pro tyto proměnné se vytvoří vždy dvě nové proměnné: *Predicted value* a *Confidence value*. Sloupec *Predicted value* obsahuje předpovídanou hodnotu na základě nejlepšího asociačního pravidla, sloupec *Confidence value* hodnotu kritéria *Confidence* daného asociačního pravidla.

V části **Best Rules Fields** lze vytvořit nové proměnné obsahující informace o nejlepších asociačních pravidlech pro každý případ. Počet nejlepších pravidel zadáme v poli **Number of best rules**. Pro každé z nich se vytvoří v datech tři sloupce udávající předpovídanou hodnotu na základě daného pravidla, hodnotu kritéria *Confidence* daného pravidla a identifikační číslo pravidla. Nové proměnné budou mít název vytvořený pomocí předpony zadané poli **Name Prefix**. Pomocí identifikačního čísla lze potom ve výstupu dohledat detaily pravidla a hodnoty dalších kritérií kvality.



*Ukládání mapových podkladů a nejlepších asociačních pravidel*

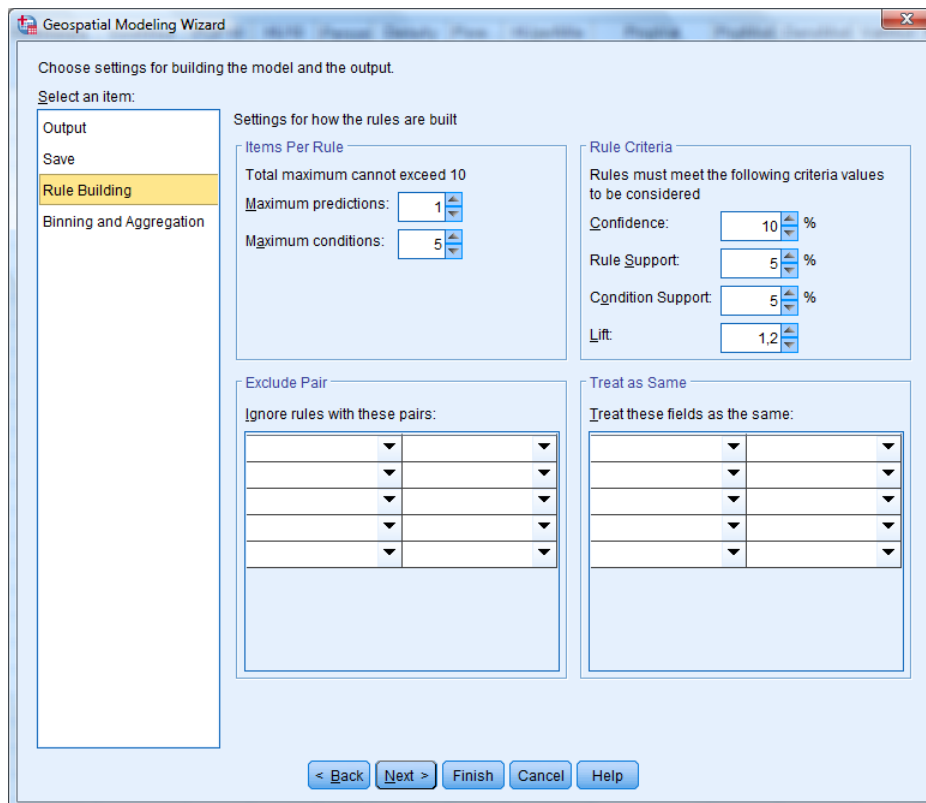
Na záložce **Rule Building** specifikujeme nastavení pro vytváření asociačních pravidel.

V části **Items Per Rule** definujeme maximální počet položek obsažených v předpovědi (**Maximum predictions**) a podmínce (**Maximum conditions**) asociačního pravidla. Jejich celkový počet nesmí přesáhnout deset. Například pravidlo  $(x=1 \text{ a } y=2) \Rightarrow (z=3)$  obsahuje celkem tři položky: dvě v podmínce ( $x=1$  a  $y=2$ ) a jednu v předpovědi ( $z=3$ ).

V části **Rule Criteria** určíme požadované minimální hodnoty kritérií kvality asociačních pravidel (**Confidence**, **Rule Support**, **Condition Support** a **Lift**).

Zobecněná geoprostorová asociační pravidla rovněž umožňují zamezit vytváření pravidel, která obsahují některé specifikované dvojice proměnných. Tyto dvojice zadáme v části **Exclude Pair**.

V některých situacích může být užitečné definovat dvojice proměnných, které budou v analýze považovány za stejné. Tyto dvojice zadáme v části **Treat as Same**.



**Nastavení pro vytváření asociačních pravidel**

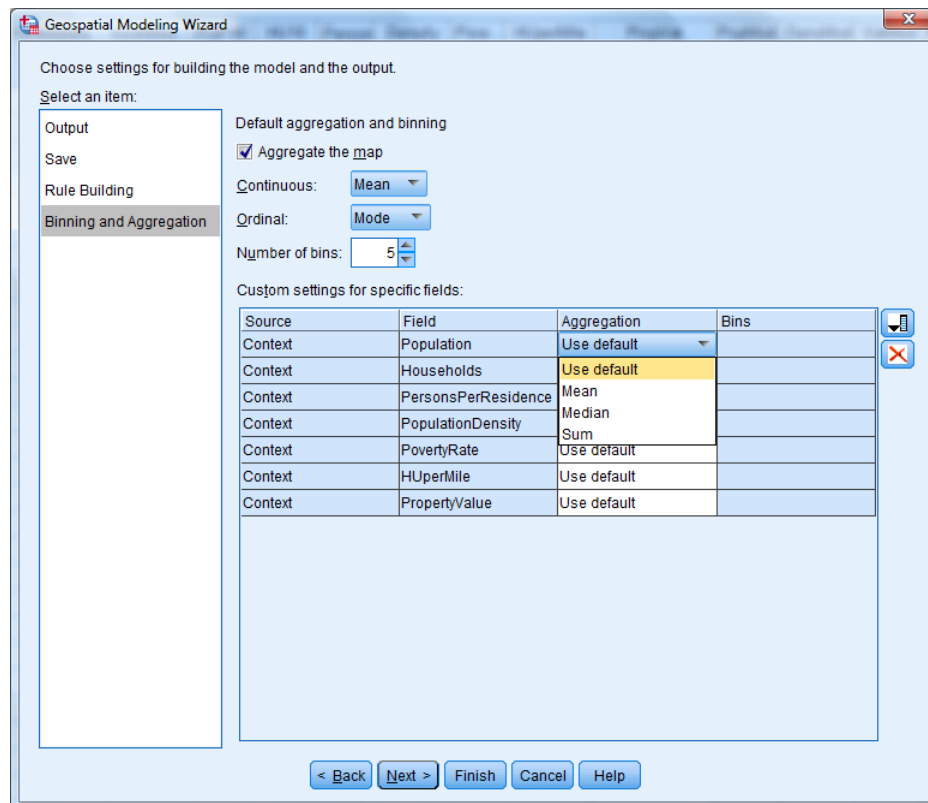
Na záložce **Binning and Aggregation** specifikujeme podrobnosti ohledně přípravy dat před vytvářením asociačních pravidel.

Agregace je nutná především v situaci, kdy záznamy v datech obsahují podrobnější informace, než jaké zobrazuje mapa (v datovém souboru jsou například demografické údaje za okresy, avšak v mapách jsou znázorněny pouze kraje). Při zaškrtnutí políčka **Aggregate the map** proběhne agregace na datových souborech i v mapách.

V rozevíracím seznamu **Continuous** nastavíme agregační funkci pro spojité proměnné, k dispozici jsou: průměr (*Mean*), medián (*Median*) a součet (*Sum*). Agregační funkci pro ordinální proměnné definujeme v poli **Ordinal** (medián (*Median*), modus (*Mode*), nejvyšší hodnota (*Highest*) nebo nejnižší hodnota (*Lowest*)). Nominální proměnné jsou vždy agregovány na základě modální hodnoty.

Před vyhledáním nejlepších asociačních pravidel se spojité proměnné kategorizují. Maximální počet kategorií (přesněji dělicích bodů při kategorizaci) nastavíme v poli **Number of bins**.

V poli **Custom settings for specific fields** lze nastavit agregační funkci a způsob kategorizace samostatně pro každou proměnnou. Kliknutím na ikonu vpravo se otevře dialog **Field Chooser**, kde vybereme proměnné, které se následně objeví v seznamu *Custom settings for specific fields*. Ve sloupci **Aggregation** nastavíme agregační funkci, ve sloupci **Bins** detaily kategorizace (viz níže).



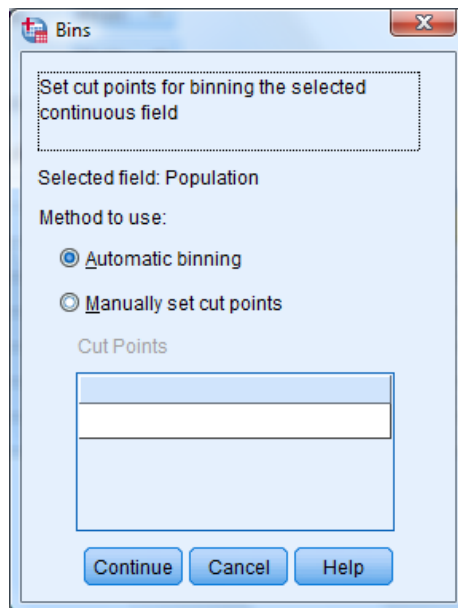
### Kategorizace a agregace

V dialogovém okně **Bins** zadáváme způsob kategorizace jednotlivých proměnných (otevře se z předchozího dialogu po poklepnání v části *Custom settings for specific fields* na některé z polí ve sloupci *Bins*).

K dispozici jsou dvě možnosti:

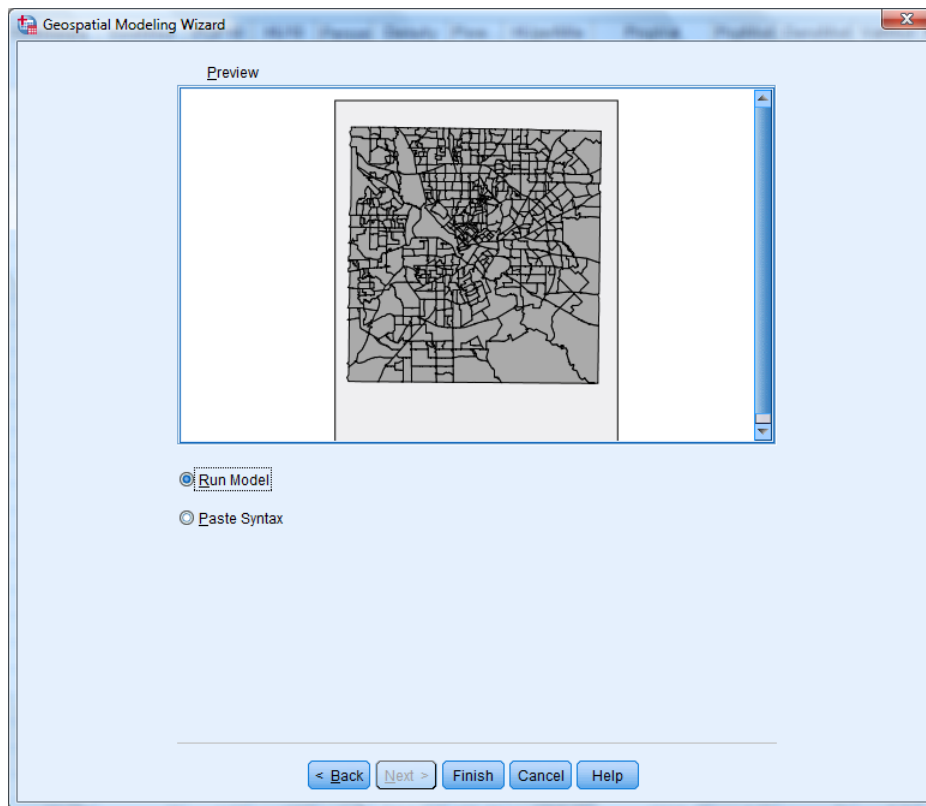
- **Automatic binning** – automatická kategorizace spojité proměnné na pevný počet kategorií daný na záložce *Binning and Aggregation*,
- **Manually set cut points** – kategorie definujeme ručně zadáním dělicích bodů v tabulce **Cut Points** (dělicí body jsou zahrnuty v horní hranici intervalů).





**Volba způsobu kategorizace**

Poslední dialogové okno zobrazuje náhled mapy (**Preview**). Rozhodneme zde, zda se má zadaná analýza provést (**Run Model**) nebo pouze zapsat syntaxový příkaz do syntaxového okna (**Paste Syntax**).



***Ukončení analýzy***

## Výstupy

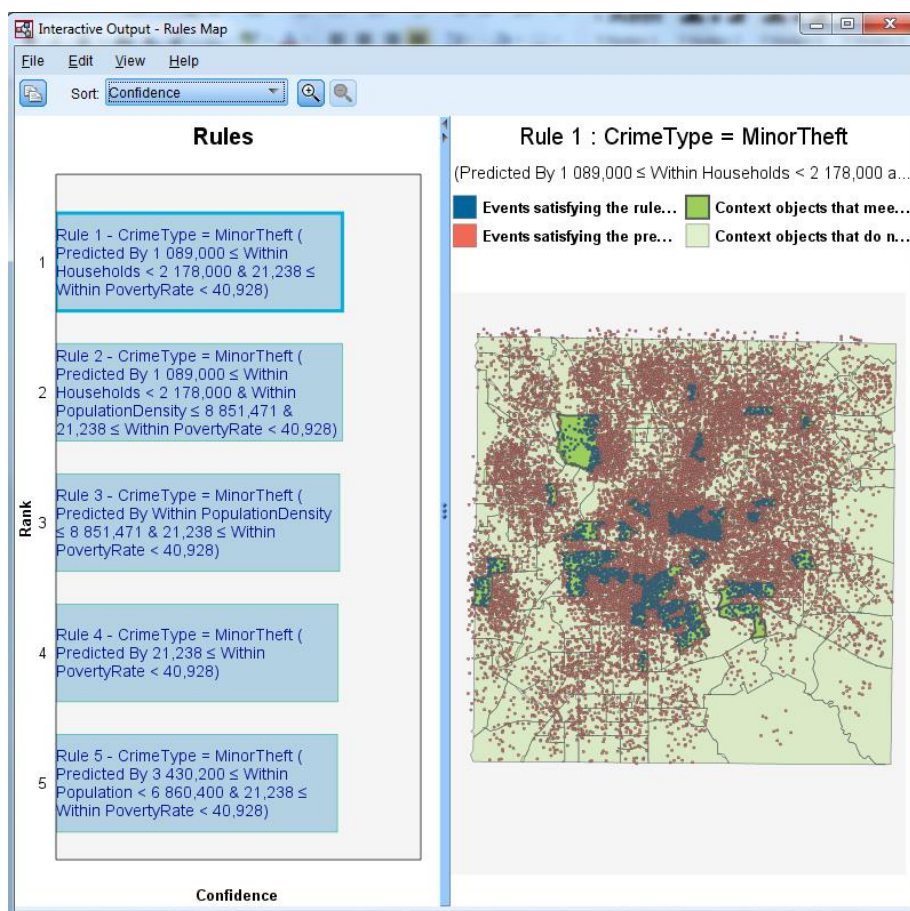
Zobecněná geoprostorová asociační pravidla nabízejí široký výběr různých typů výstupů, které mají formu pivotních tabulek nebo interaktivního výstupu (*Interactive Output*). Obecný způsob práce s interaktivními výstupy je popsán ve skriptech *IBM SPSS Statistics Base 24: Datový soubor a práce s výstupy*. Níže se zaměříme na specifika jednotlivých typů interaktivního výstupu.

### Přehled základních typů interaktivního výstupu

#### Mapa pravidel (Rules Map)

Interaktivní objekt *Rules Map* se skládá ze dvou panelů. Levý panel zobrazuje přehled nejlepších asociačních pravidel na základě daného kritéria ve formě sloupcového grafu, pravý panel znázorňuje mapu. Pravidla jsou řazena sestupně podle hodnoty kritéria, šířka sloupce je úměrná této hodnotě. Kritérium lze změnit v rozevíracím seznamu **Sort** nebo v nabídce **View, Sort**. Číslo asociačního pravidla (za slovem *Rule*) odpovídá pořadí při řazení dle kritéria *Confidence*.

Mapa se dle výchozího nastavení vztahuje k nejlepšímu asociačnímu pravidlu. Nadpis obsahuje předpověď pravidla, podmínku nalezneme v podnadpisu. Mapa zobrazuje všechny události splňující předpověď daného pravidla. Modrou barvou jsou znázorněny ty, které splňují podmínku; červenou barvou ty, které ji nespĺňují. Prostorové objekty vyhovující podmínce mají tmavě zelenou barvu; ostatní světle zelenou. Mezi pravidly přepínáme kliknutím na příslušný sloupec v levé části okna.

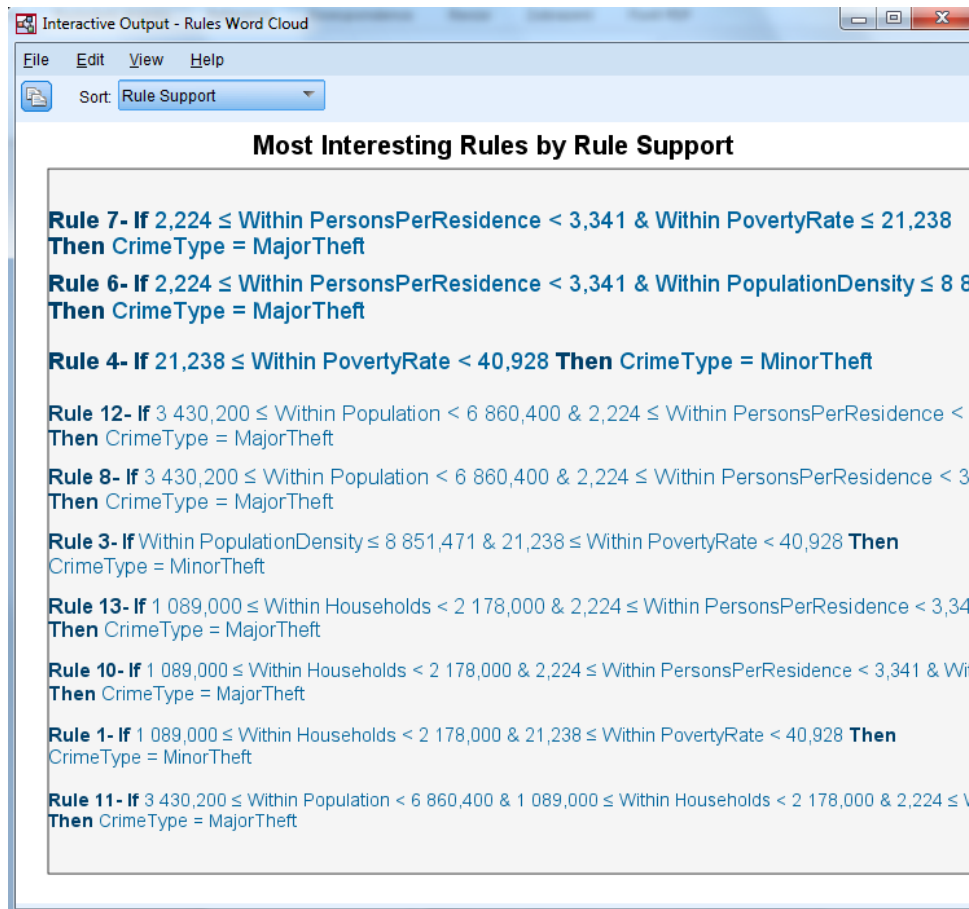


Interaktivní výstup – Map

### Oblak slov pravidel (Rules Word Cloud)

Interaktivní objekt *Rules Word Cloud* obsahuje výpis nejlepších asociačních pravidel na základě zvoleného kritéria řazených podle tohoto kritéria. Pravidla jsou zapsána ve tvaru: „Rule“ číslo pravidla „-If“ podmínka „Then“ předpověď. Velikost písma indikuje relativní důležitost pravidla na základě vybraného kritéria. Číslo asociačního pravidla (za slovem *Rule*) odpovídá pořadí při řazení podle spolehlivosti (*Confidence*).

Kritérium lze přepnout v poli **Sort** nebo v nabídce **View, Sort**, čímž se změní seznam zobrazených pravidel. Interaktivní výstup zahrnuje seznamy nejlepších pravidel pro všech pět kritérií, volba kritéria v dialogu při zadávání určuje pouze výchozí zobrazení.



Interaktivní výstup – Rules Word Cloud

### 1.16.3 Časoprostorové predikce (Spatio-Temporal Prediction)

Po navázání datových souborů na mapové podklady specifikujeme požadovaný model pro časoprostorové predikce.

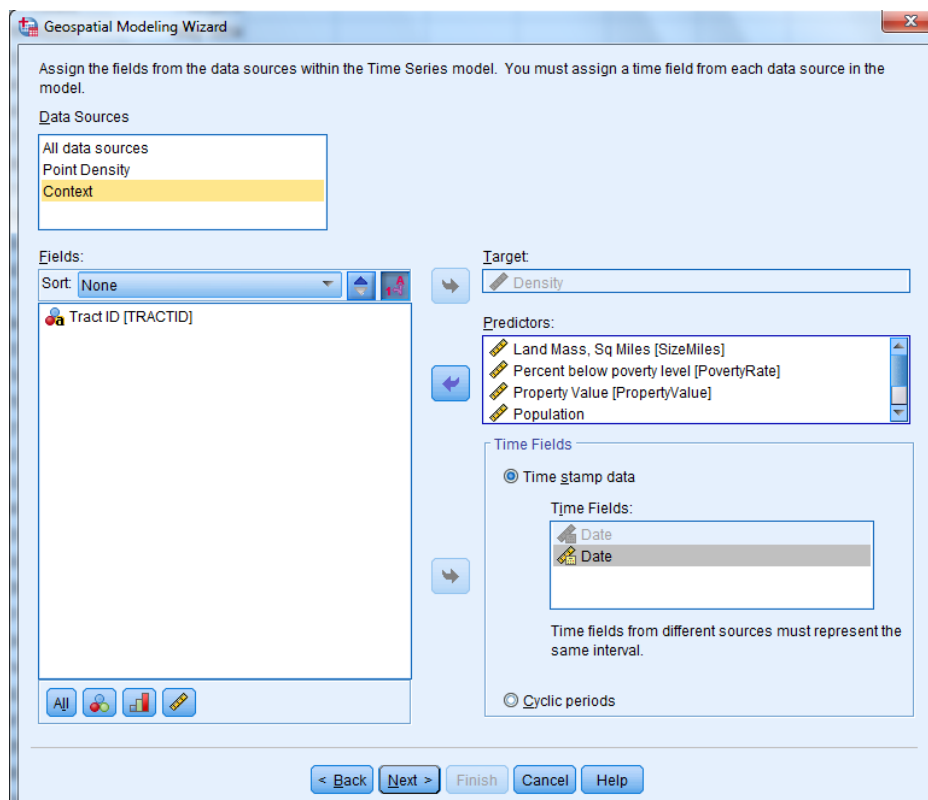
Pole **Data Sources** obsahuje seznam všech načtených datových souborů. Na základě této volby se v poli **Fields** zobrazují všechny proměnné, nebo jen z vybraného datového souboru. Proměnné lze řadit (**Sort**) či filtrovat podle typu pomocí tlačítek pod seznamem.

V poli **Target** definujeme cílovou proměnnou. Pokud jsme v předcházejícím kroku načteli data o hustově bodů (*Point Density*), do pole se automaticky vložila položka *Density* vyjadřující hustotu událostí na mapě (tuto volbu již nelze měnit). V ostatních případech zadáme číselnou proměnnou z kontextových dat.

V poli **Predictors** můžeme určit nezávislé proměnné (prediktory).

V části **Time Fields** zvolíme způsob definování časové proměnné:

- **Time stamp data** – zadání jedné nebo více časových proměnných z dat do **Time Fields**. V případě dvou datových zdrojů je třeba definovat časové proměnné pro oba soubory, časové intervaly se musí shodovat.
- **Cyclic Periods** – periodické cykly vyjadřující sezónnost (definují se v následujícím dialogovém okně na záložce *Time Intervals*).



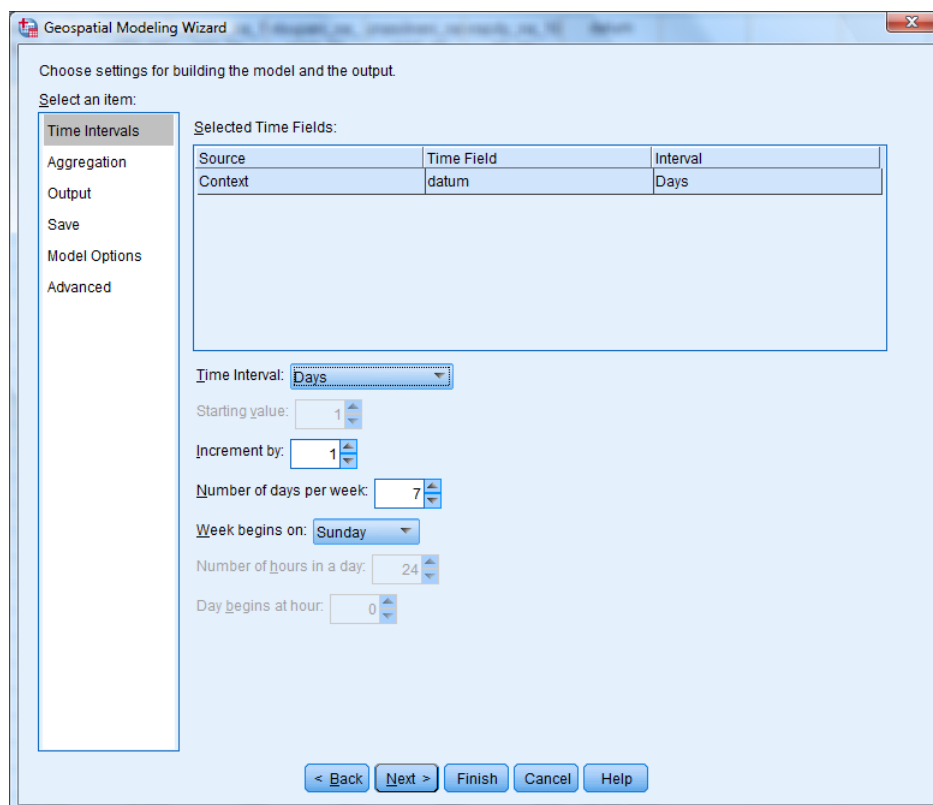
Zadávání proměnných pro časoprostorové predikce

Následující dialogové okno **Choose settings for building the model and the output** je určeno ke specifikaci modelu a výstupů. Skládá se z více záložek, mezi kterými přepínáme v seznamu **Select an item**.

Na záložce **Time Intervals** definujeme časový interval. Možnosti nastavení se liší podle volby v předcházejícím dialogovém okně v části *Time Fields*. Pokud jsme načetli časové proměnné z dat (*Time stamp data*), zobrazí se tyto informace v tabulce **Selected Time Fields**.

Časový interval pro analýzu nastavíme v rozevíracím seznamu **Time Interval**. V závislosti na této volbě jsou nutná některá další nastavení: počáteční hodnota (**Starting value**), časový přírůstek mezi pozorováními (**Increment by**), počet dní v týdnu (**Number of days per week**), počáteční hodnota pro týdny (**Weeks begins on**), počet hodin v jednom dni (**Number of hours in a day**) a první hodina dne (**Day begins at hour**). Pokud zadaná proměnná nemá formát data ani času, nastaví se automaticky v poli *Time Interval* hodnota *Periodic* a toto nastavení nelze změnit.

Procedura předpokládá, že záznamy reprezentují stejně široké časové intervaly. Na základě těchto specifikací je potom schopna detekovat chybějící pozorování, nebo více pozorování v jednom intervalu, která je nutné agregovat.



**Nastavení časových intervalů**

Pokud jsme v předchozím dialogovém okně v části *Time Fields* zvolili periodické cykly (*Cyclic periods*), musíme na záložce **Time Intervals** specifikovat proměnné obsahující cyklickou reprezentaci časových údajů. Tímto způsobem definujeme v datech sezónnost, tj. opakující se cyklické změny (například střídání měsíců v roce nebo dní v týdnu).

V dialogu lze zadat až tři úrovně periodicity, přičemž je řadíme sestupně podle délky časových intervalů (například v pořadí dny, hodiny, minuty). Tyto úrovně definujeme postupně v částech **First Cycle Field**, **Second Cycle Field**, **Third Cycle Field** v poli **Field**. V případě dvou datových zdrojů je třeba provést zadání pro každý z nich a délka odpovídajících proměnných se musí shodovat.

Hodnoty v polích **First cycle length** a **Second cycle length** udávají délku periody vzhledem k následující úrovni (pokud například první úroveň vyjadřuje roky a druhá měsíce, do pole *First cycle length* zadáme 12).

Do polí **Second starting value** a případně **Third starting value** zadáme počáteční hodnoty druhého a třetího cyklického období.

V poli **Increment** definujeme časový přírůstek mezi pozorováními na poslední úrovni.

**Nastavení periodických cyklů**

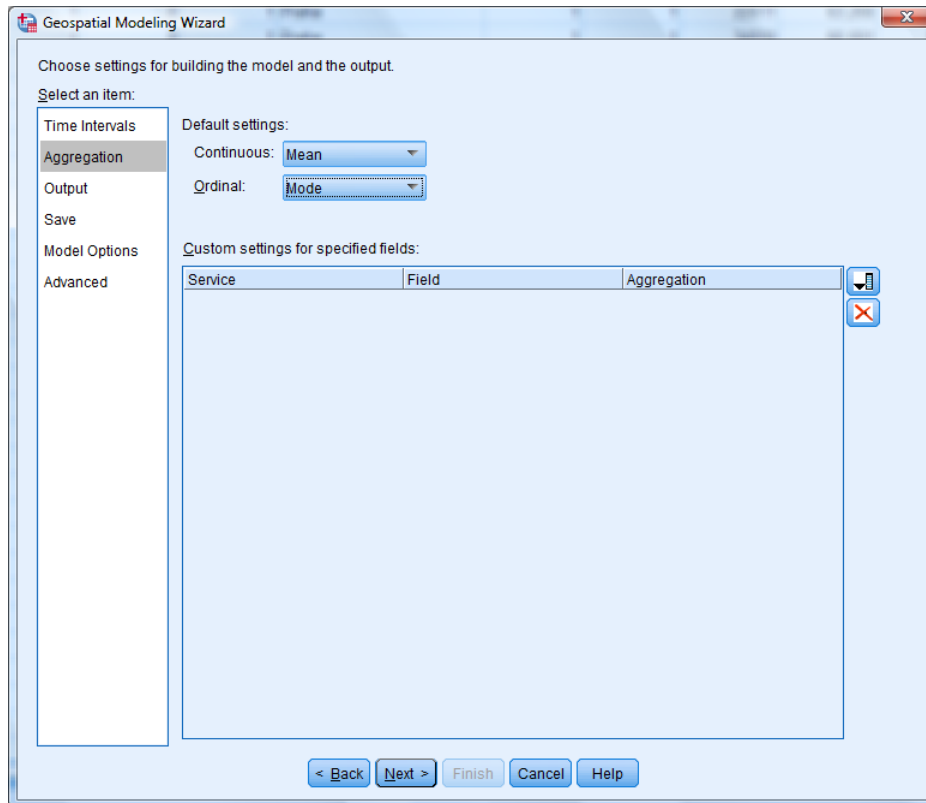
Záložka **Aggregation** je určena pro situace, kdy model zahrnuje prediktory definované v předchozím dialogovém okně v poli *Predictors*. Specifikujeme zde agregační funkce pro případ, že se v jednom časovém intervalu vyskytne více pozorování a data je nutné agregovat.

V rozevíracím seznamu **Continuous** nastavíme agregační funkci pro spojité proměnné, k dispozici jsou: průměr (*Mean*), medián (*Median*) a součet (*Sum*). Agregační funkci pro ordinální proměnné definujeme v poli **Ordinal** (medián (*Median*), modus (*Mode*), nejvyšší hodnota (*Highest*) nebo nejnížší hodnota (*Lowest*)). Nominální proměnné jsou vždy agregovány na základě modální hodnoty.

V poli **Custom settings for specific fields** lze nastavit agregační funkci a způsob kategorizace samostatně pro každou proměnnou. Kliknutím na ikonu vpravo se otevře dialog **Field Chooser**, kde vybereme proměnné, které se následně objeví v seznamu *Custom settings for specific fields*. Pole



**Service** charakterizuje datový zdroj, pole **Field** název proměnné. Ve sloupci **Aggregation** nastavíme agregační funkci.



### **Agregace**

Na záložce **Output** specifikujeme požadované výstupy.

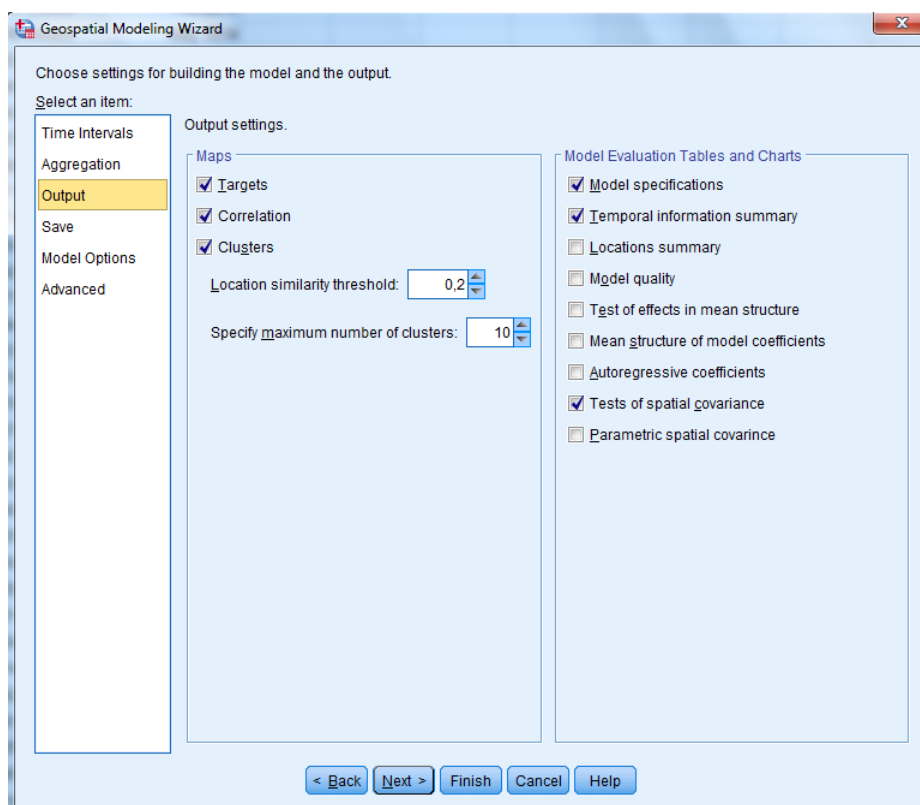
V části **Maps** jsou k dispozici tři typy interaktivních objektů obsahujících mapy:

- **Targets** – teplotní mapa korelací (*Correlation Heat Map*) zobrazuje pro vybraný prostorový objekt pět s ním nejsilněji korelovaných objektů.
- **Correlation** – mapa korelací (*Correlation Map*), znázorňuje korelace a kovariance mezi libovolnými pěti vybranými prostorovými objekty.
- **Clusters** – mapy shluků prostorových objektů (*Map by Closeness, Map by Size*). Jsou k dispozici pouze pro empirické modely (na záložce *Model Options, Spatial Covariance* je označeno *Nonparametric*). Pole **Location similarity threshold** udává požadovanou podobnost objektů pro vytvoření shluku (hodnota mezi 0 a 1). V poli **Specify maximum number of clusters** určíme maximální počet shluků.

V části **Model Evaluation Tables and Charts** zadáváme výstupy ve formě pivotních tabulek a grafů:

- **Model specifications** – souhrnné informace o nastavení modelu.
- **Temporal information summary** – informace o časových proměnných a časových intervalech užitých v modelu.
- **Location summary** – počet zahrnutých a vyloučených prostorových objektů.

- **Model quality** – informace o kvalitě modelu (*Mean Squared Error, Pseudo R Square, Pseudo Adjusted R Square*).
- **Test of effects in mean structure** – testy efektů pro složky modelu vyjadřující lineární závislost průměru na hodnotách prediktorů ( $X\beta$ ): testová statistika ( $F$ ), stupně volnosti ( $df1$ ,  $df2$ ) a signifikance ( $Sig.$ ).
- **Mean structure of model coefficients** – pro složky vyjadřující lineární závislost průměru na hodnotách prediktorů ( $X\beta$ ): odhady koeficientů (*Coefficients*), její standardní chyby (*Std. Error*), testová statistika ( $t$ ), signifikance ( $Sig.$ ) a intervaly spolehlivosti (*Confidence Interval*) pro každý člen modelu.
- **Autoregressive coefficients** – odhady autoregresních koeficientů pro všechna uvažovaná posunutí (*Coefficients*), její standardní chyby (*Std. Error*), testová statistika ( $t$ ), signifikance ( $Sig.$ ) a intervaly spolehlivosti (*Confidence Interval*).
- **Test of spatial covariance** – pro parametrické modely prostorové kovariance: test dobré shody prostorové kovarianční struktury (ověřuje, zda je vhodné modelovat prostorovou kovarianční strukturu parametricky nebo raději užít neparametrický model).
- **Parametric spatial covariance** – pro parametrické modely prostorové kovariance: graf znázorňující vztah mezi Euklidovskou vzdáleností prostorových objektů a odhadem prostorové kovariance.



**Nastavení výstupů**

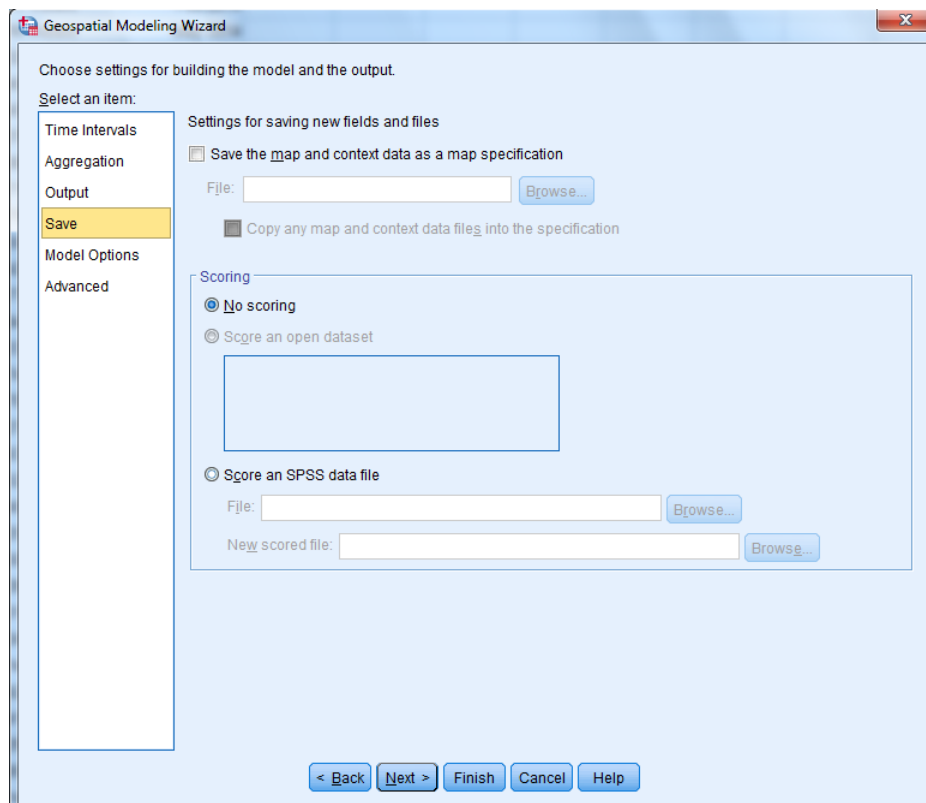
Na záložce **Save** nastavíme podrobnosti ohledně uložení mapových podkladů a predikcí modelu.

Zaškrtnuté políčko **Save the map and context data as a map specification** uloží mapovou specifikaci do souboru *MPLAN*, který lze později užít při dalších analýzách. Tlačítkem **Browse** zadáme název a umístění souboru. Při zaškrtnutí políčka **Copy any map and data files into the specification** se do souboru uloží rovněž data ze všech mapových souborů, otevřených datových souborů i externích datových souborů užívaných v mapové specifikaci. Soubor *MPLAN* lze později užít pro definování mapových podkladů v dialogovém okně *Add maps to the model* tlačítkem *Load map specification* nebo v syntaxovém příkazu *SPATIAL ASSOCIATION RULES*.

V části **Scoring** můžeme uložit do datového souboru: predikce cílové proměnné, odhad standardní chyby a dolní a horní mez predikčního intervalu. Data lze uložit do otevřeného datového okna nebo datového souboru IBM SPSS Statistics (SAV). Nesmí se však jednat o žádný ze zdrojových souborů užívaných modelem. Soubor musí obsahovat všechny časové proměnné a prediktory zahrnuté v modelu, hodnoty časových proměnných však musí být vyšší než ty, na nichž byl model vytvářen.

Pro specifikaci datového souboru zvolíme jednu z možností:

- **No scoring** – neukládat predikce do souboru,
- **Score an open dataset** – uložit do vybraného otevřeného datového okna,
- **Score an SPSS data file** – uložit do datového souboru IBM SPSS Statistics (v poli **File** specifikujeme cestu k souboru s požadovanými informacemi; data se uloží do nového souboru, který definujeme v poli **New scored file**).



***Uložení mapových podkladů a výsledků skórování***

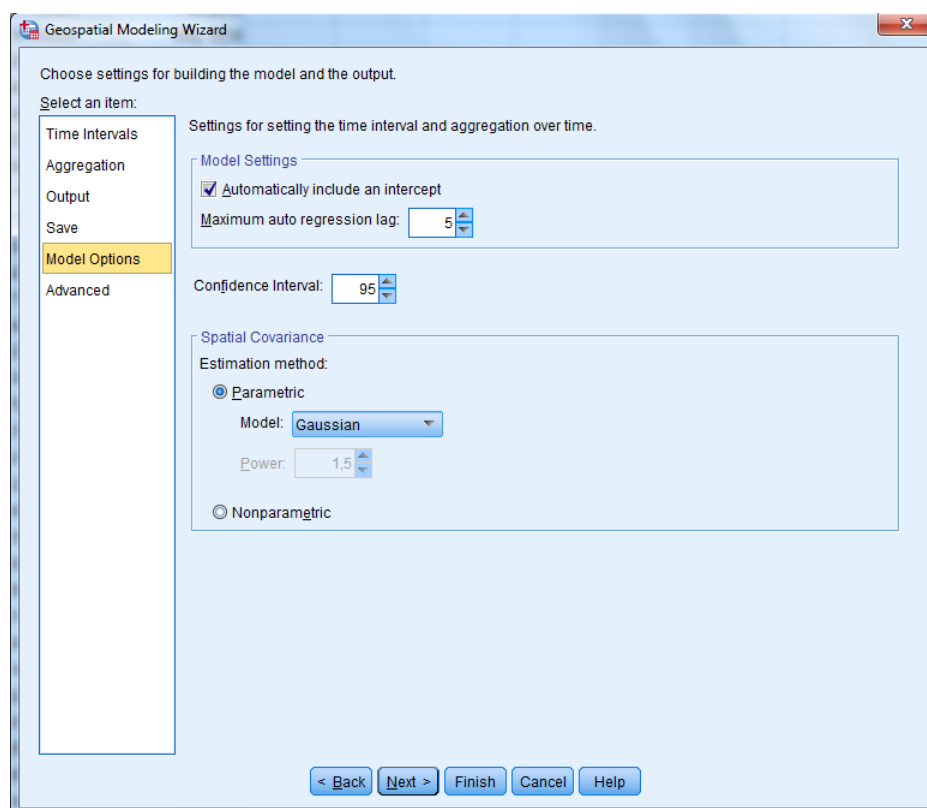
Na záložce **Model Options** provádíme základní nastavení modelu.

V části **Model Settings** určíme, zda má být do modelu zahrnuta regresní konstanta (**Automatically include an intercept**) a zadáme maximální počet uvažovaných autoregresních zpoždění (**Maximum auto regression lag**).

V poli **Confidence Interval** určíme požadovanou hladina spolehlivosti.

V části **Spatial Covariance** zadáme metodu odhadu prostorové kovarianční matice:

- **Parametric** – parametrický model, v poli **Model** zvolíme Gaussovský (**Gaussian**), exponenciální (**Exponential**) nebo mocninný exponenciální (**Power Exponential**). Pro mocninný exponenciální model dále zadáme mocninu v poli **Power** (musí být větší nebo rovna 1.1).
- **Nonparametric** – neparametrický model.



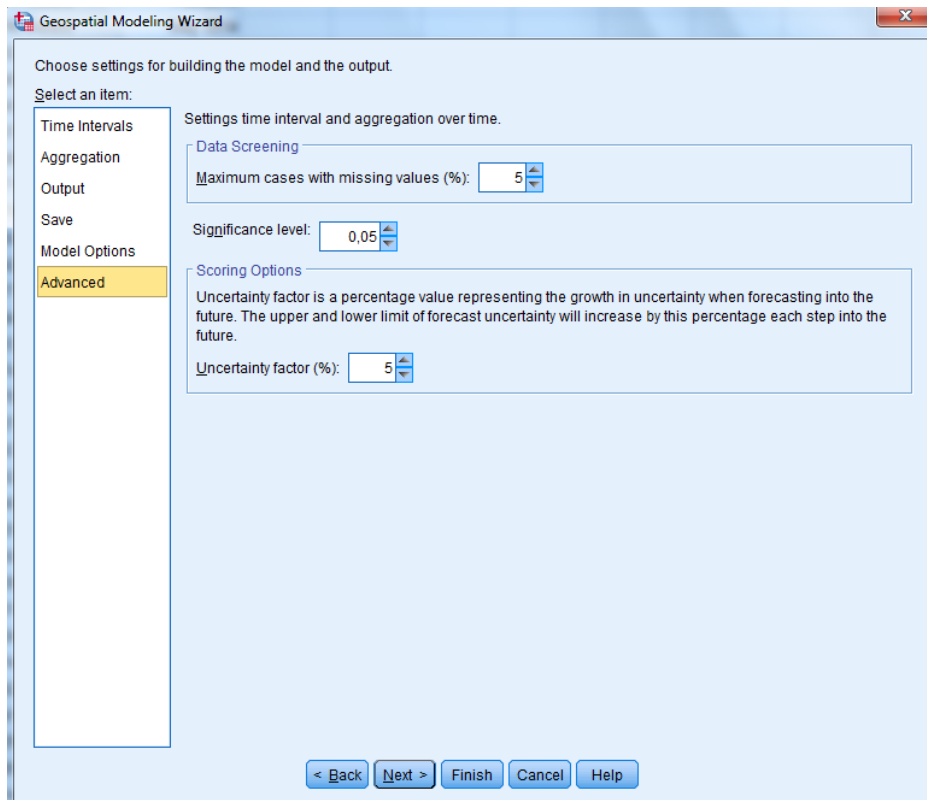
**Nastavení modelu**

Na záložce **Advanced** provádíme pokročilá nastavení modelu.

V části **Data Screening** v poli **Maximum cases with missing values (%)** určíme maximální povolené procento případů s vynechanými hodnotami.

Pole **Significance level** určuje hladinu významnosti pro testování vhodnosti parametrického modelu.

V části **Scoring Options** v poli **Uncertainty factor (%)** definujeme procentuální nárůst nejistoty pro předpovědi do budoucna. Při každé předpovědi do budoucna o jednu časovou jednotku, se rozšíří dolní i horní mez odhadu predikčního intervalu o dané procento.



**Pokročilá nastavení**

Poslední dialogové okno je shodné jako pro zobecněná geoprostorová asociační pravidla. Zobrazuje náhled mapy (**Preview**) a rozhodneme zde, zda se má zadaná analýza provést (**Run Model**) nebo pouze zapsat syntaxový příkaz do syntaxového okna (**Paste Syntax**).

## Výstupy

Časoprostorové predikce nabízejí široký výběr různých typů výstupů, které mají formu pivotních tabulek, grafu nebo interaktivního výstupu (*Interactive Output*). Obecný způsob práce s interaktivními výstupy je popsán ve skriptech *IBM SPSS Statistics Base 24: Datový soubor a práce s výstupy*. Níže se zaměříme na specifika jednotlivých typů interaktivního výstupu.

### Přehled základních typů interaktivního výstupu

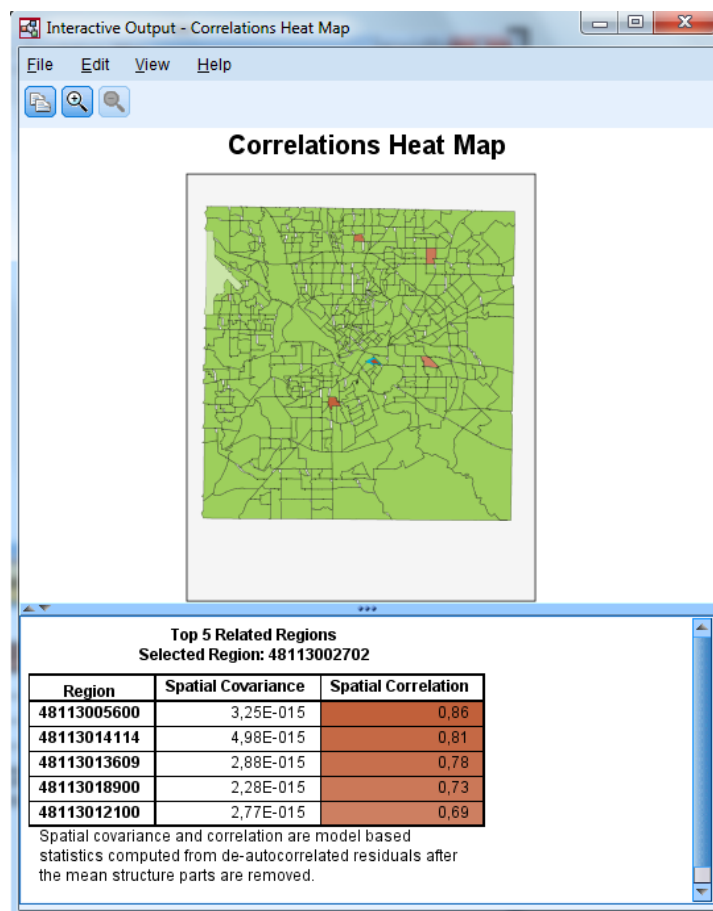
#### Teplotní mapa korelací (Correlation Heat Map)

Interaktivní výstup *Correlation Heat Map* znázorňuje pro vybraný prostorový objekt pět nejsilněji korelovaných objektů. K tomuto účelu kombinuje mapu (horní část okna) a tabulku (spodní část okna). Vybraný objekt je v mapě ohraničen modrou barvou, všech šest klíčových objektů je podbarveno hnědě. Intenzita hnědé barvy v mapě i v posledním sloupci tabulky odpovídá hodnotě prostorové korelace – čím silnější korelace, tím sytější barva. Pokud umístíme kurzor nad libovolný objekt v mapě, zobrazí se popisek s jeho identifikací.

Název tabulky obsahuje identifikaci vybraného objektu, řádky reprezentují pět s ním nejsilněji korelovaných objektů. Ve sloupcích nalezneme postupně: identifikaci objektu, prostorovou kovarianci (*Spatial Covariance*) a prostorovou korelaci (*Spatial Correlation*). Řádky jsou řazeny sestupně podle prostorové korelace. Kliknutím na libovolný řádek tabulky se v mapě zvýrazní modrým ohraničením také odpovídající objekt.

Pro výběr jiného objektu klikneme v mapě na příslušnou oblast. Mapa i tabulka se následně aktualizují.

Mapu přibližujeme, resp. oddalujeme lupou nebo v nabídce **View, Zoom**. Přiblíženou mapu lze posouvat směrem nahoru, dolů, doprava nebo doleva pomocí nabídky **View, Pan**.



Interaktivní výstup – Correlation Heat Map

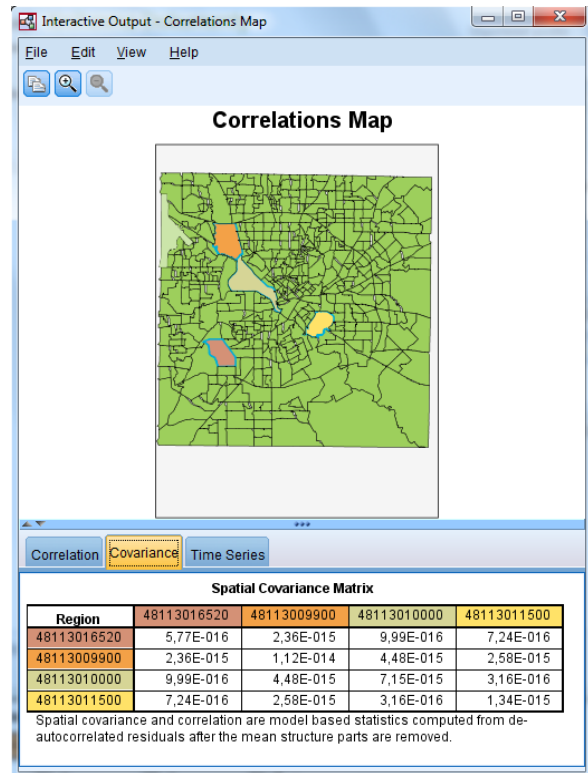
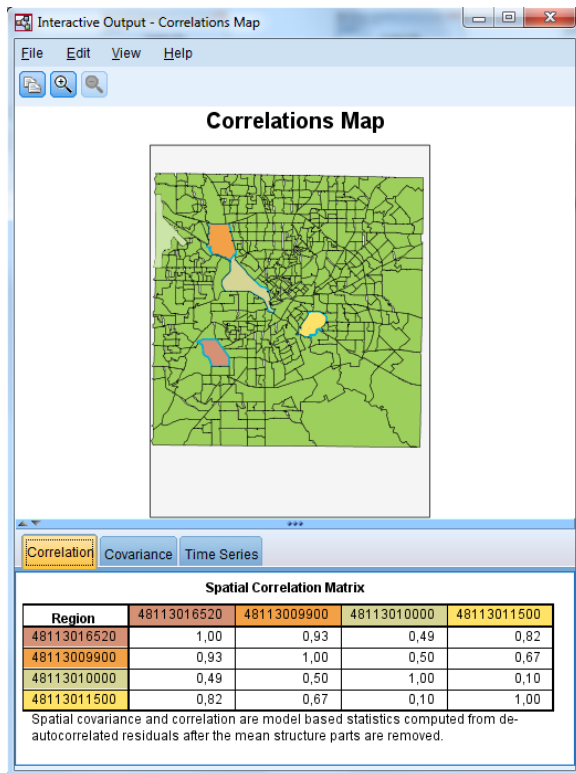
### Mapa korelací (Correlation Map)

Interaktivní výstup *Correlation map* zobrazuje prostorové korelace a kovariance až pěti vybraných prostorových objektů. V horní části okna se nachází mapa, dolní část se skládá ze tří záložek (**Correlation**, **Covariance** a **Time Series**), mezi kterými přecházíme kliknutím na záložku nebo přepínáním v nabídce **View**. Pokud umístíme kurzor nad libovolný objekt v mapě, zobrazí se popis s jeho identifikací.

Neoznačený objekt vybereme, resp. označený objekt odstraníme z výběru kliknutím na příslušnou oblast v mapě. Každý vybraný objekt je v mapě i v tabulkách podbarvený jinou barvou. Mapu přibližujeme, resp. oddalujeme lupou nebo v nabídce **View, Zoom**. Přiblíženou mapu lze posouvat směrem nahoru, dolů, doprava nebo doleva pomocí nabídky **View, Pan**.

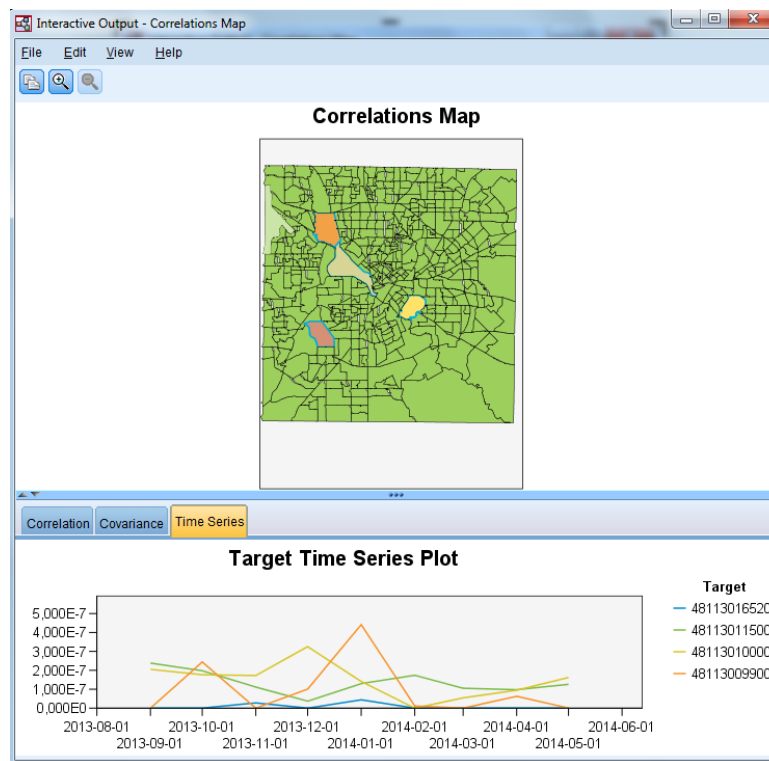
Záložky **Correlation** a **Covariance** obsahují matice prostorových korelací a kovariancí vybraných objektů.

Záložka **Time Series** znázorňuje odhad trendu cílové proměnné ve zvolených oblastech v celé délce časového intervalu. Pokud jsme pro vstup do modelu použili dva datové soubory (tj. také soubor *Point Density*), zobrazují křivky odhad trendu jádrové hustoty událostí (nejedná se přímo o míru výskytu událostí).



**Interaktivní výstup – Correlation Map, záložky Correlation a Covariance**





Interaktivní výstup – Correlation Map, záložka Time Series

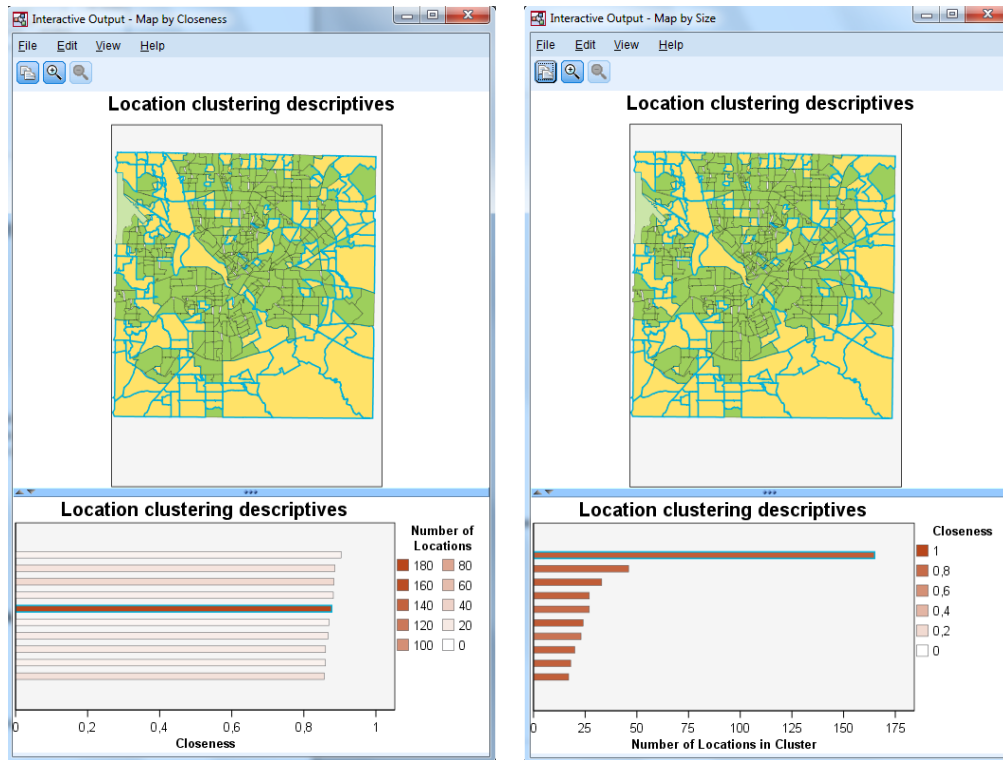
### Mapy shluků prostorových objektů (Map by Closeness, Map by Size)

Mapy shluků prostorových objektů zahrnují dva typy interaktivního výstupu: *Map by Closeness* a *Map by Size*. Oba se skládají z mapy (horní část okna) a sloupcového grafu (spodní část okna), liší se pouze informací zobrazenou ve sloupcovém grafu. Pro jejich vytvoření je třeba na záložce *Output* zaškrtnout volbu *Clusters* (rovněž zde zadáváme požadovanou podobnost objektů pro vytvoření shluku a maximální počet shluků).

Mapa zobrazuje shluky (klastry) nejvíce podobných prostorových objektů. Žlutou barvou jsou vždy znázorněny objekty jednoho shluku. Pokud umístíme kurzor nad libovolný objekt v mapě, zobrazí se popisek s jeho identifikací. Kliknutím na vybraný objekt přepneme na zobrazení daného shluku.

Každý sloupec grafu ve spodní části okna odpovídá jednomu shluku. Kliknutím na vybraný sloupec se v mapě zvýrazní žlutou barvou objekty tohoto shluku. V případě objektu *Map by Closeness* znázorňuje horizontální osa grafu charakteristiku *Closeness* vyjadřující kvalitu shluků. Sloupce jsou podbarvené podle počtu objektů ve shluku. Objekt *Map by Size* má tyto vlastnosti přehozené (horizontální osa grafu znázorňuje velikost shluku, podbarvení hodnotu charakteristiky *Closeness*).

Mapu přibližujeme, resp. oddalujeme lupou nebo v nabídce **View, Zoom**. Přiblíženou mapu lze posouvat směrem nahoru, dolů, doprava nebo doleva pomocí nabídky **View, Pan**.



Interaktivní výstupy – Map by Closeness, Map by Size

## Geoprostorové modelování

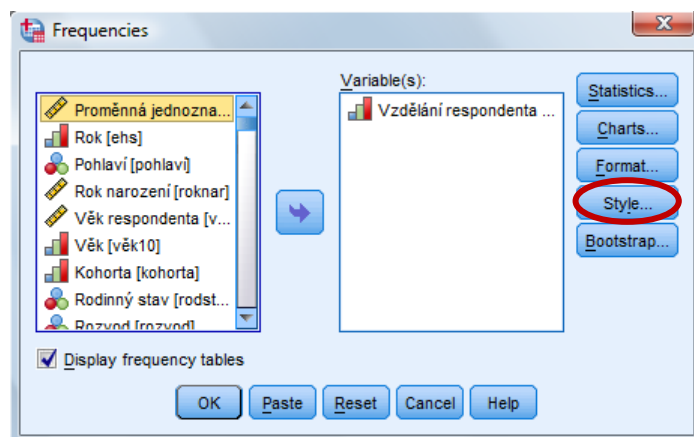
- Analyze
  - Spatial and Temporal Modeling
    - Spatial Modeling

## Příloha 1 – Podmíněné formátování tabulek

Některé analytické procedury IBM SPSS Statistics umožňují nastavit formát pivotních tabulek již při jejich vytváření pomocí tzv. podmíněného formátování. Podmíněné formátování tabulek je podporované následujícími procedurami:

- Četnosti (*Frequencies*),
- Popisné statistiky (*Descriptives*),
- Kontingenční a asociační tabulky (*Crosstabs*),
- Průměry (*Means*),
- Korelace (*Correlate*),
- Regresní analýza (*Regression*).

Dialogy těchto procedur obsahují tlačítko **Style**, které otevře dialog **Table Style**.



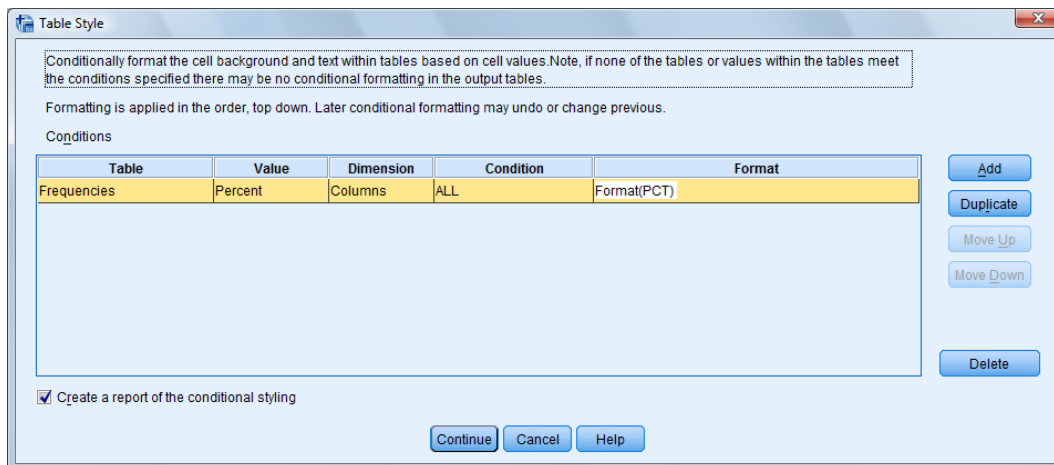
**Tlačítko Style pro zadání podmíněného formátování**

Dialog **Table Style** zůstává v principu stejný pro všechny podporované procedury, drobné odlišnosti se vztahují pouze k obsahu jednotlivých procedur.

Tabulky, na které se mají změny formátování aplikovat, určíme v poli **Table**: všechny tabulky, pro něž jsou změny relevantní (**All applicable tables**), nebo konkrétní typ tabulky vytvářený danou procedurou. V poli **Value** specifikujeme oblast tabulky – vybereme některou z nabízených statistik zobrazených v tabulce, vepíšeme vlastní slovní spojení určující řádek či sloupec tabulky, nebo vybereme celou datovou oblast. Pole **Dimension** určuje, zda se statistika resp. slovní spojení, specifikované v přechodím poli, má vyhledat v řádcích, ve sloupcích, nebo v obou dimenzích tabulky. V poli **Condition** zadáme podmínku definující, kdy se má formátování provádět (podrobněji viz popis dialogu **Table Style: Condition**). Pole **Format** definuje požadované změny formátu (podrobněji viz popis dialogu **Table Style: Format**).

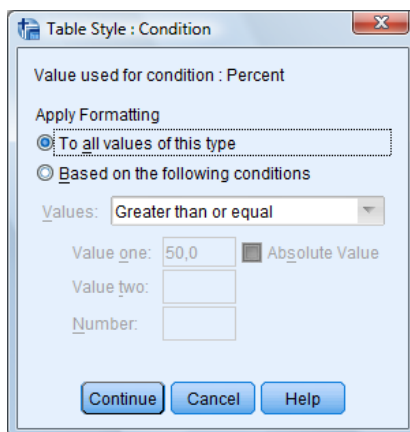
V pravé části dialogu **Table Style** se nachází pět tlačítek, která umožňují provést více podmíněných formátování najednou. Tlačítkem **Add** přidáme do seznamu nový prázdný řádek, ve kterém definujeme další podmíněné formátování. Tlačítkem **Duplicate** zkopírujeme označený řádek na poslední místo v seznamu podmínek a dále jej podle potřeby upravíme. Editace tabulky se provádějí v pořadí určeném v dialogu. Označený řádek posuneme v seznamu nahoru pomocí tlačítka **Move up** a dolů tlačítkem **Move down**, případně ho vymažeme tlačítkem **Delete**.

Při zaškrtnutí volby **Create a report of the conditional styling** se ve výstupu zobrazí tabulka **Output Modify Summary** s přehledem o počtu vybraných, změněných a vymazaných tabulek.



**Zadání podmíněného formátování**

V okně **Table Style: Condition** vybereme buď všechna datová pole, která vyhovují podmínkám definovaným v polích **Value** a **Dimension (To all values of this type)**, nebo pomocí jednoduchého logického výrazu výběr polí dále specifikujeme (**Based on the following conditions**).



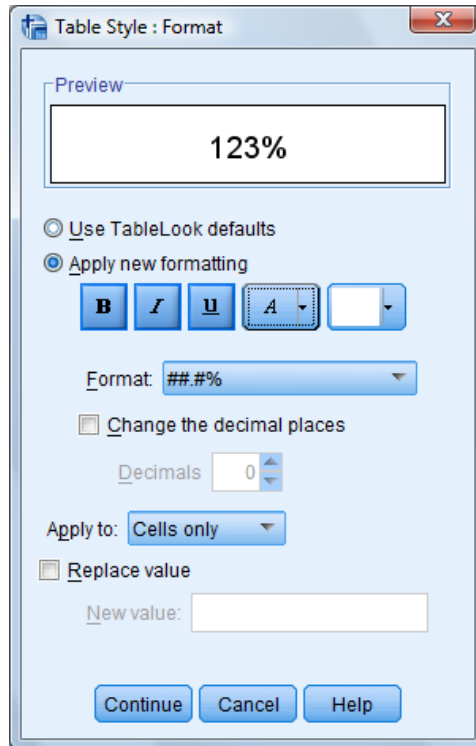
**Zadání podmínky**

V okně **Table Style: Format** určíme vzhled oblasti tabulky specifikované výše buď na základě šablony tabulky (**Use TableLook defaults**), nebo definujeme změny, které se mají provést (**Apply new formatting**). V případě druhé možnosti jsou k dispozici tato nastavení:

- řez písma,
- barva písma,
- barva pozadí,
- formát a případně počet zobrazených desetinných míst,

- nahrazení obsahu polí zadaným textem.

V poli **Apply to** vybereme, zda se mají změny aplikovat pouze na datová pole tabulky, nebo i na řádkový či sloupcový popis.



**Zadání změny formátování**

Podmíněné formátování tabulek lze zadávat také pomocí syntaxového příkazu **Output modify**. Bližší informace o tomto příkazu a jeho možnostech jsou uvedeny v manuálu *IBM SPSS Statistics 24 Command Syntax Reference*.

## Rejstřík

### A

Analýza nejbližších sousedů, 6, 67, 69  
 Analýza reliability, 73  
 Analyze, 5, 9, 10, 12, 14, 17, 20, 22, 24, 26, 28, 31, 32, 34, 35, 36, 38, 39, 42, 44, 45, 48, 50, 52, 55, 56, 59, 62, 64, 66, 69, 71, 73, 74, 77, 79, 81, 83, 85, 86, 87, 88, 90, 91, 92, 94, 97, 99, 101, 102, 103, 121, 123, 126  
 Analýzy, 5  
 Autocorrelations, 6, 95, 96, 97  
 Autokorelace, 95, 96, 97, 99  
 Automatic Linear Modeling, 5, 49, 50  
 Automatická tvorba lineárního modelu, 49, 50

### B

Binomial, 6, 84, 85  
 Binomial Test, 6, 84  
 Binomický test, 84, 85  
 Bivariate Correlations, 5, 44

### C

Case Summaries, 11, 12  
 Classify, 6, 58, 59, 61, 62, 64, 65, 66, 69  
 Codebook, 5, 8, 9  
 Compare Means, 5, 33, 34, 35, 36, 38, 39  
 Control Charts, 6, 120, 121  
 Correlate, 5, 44, 45, 48  
 Cross-Correlations, 6, 97, 99  
 Crosstabs, 24, 26, 102, 103  
 Curve Estimation, 6, 54, 55

### Č

Časoprostorové modelování, 128  
 Časoprostorové predikce, 150  
 Časové řady, 93  
 Četnosti, 19

### D

Data Reduction, 6, 71  
 Define Sets, 100, 101, 102  
 Descriptive Statistics, 5, 19, 20, 22, 24, 26, 28, 31, 32  
 Descriptives, 21, 22, 23, 65, 71, 73  
 Discriminant Analysis, 6, 65, 66  
 Diskriminační analýza, 6, 65, 66  
 Distances, 5, 47, 48, 51, 74  
 Dvoustupňové seskupování, 58, 59

### E

Explorace, 23, 24  
 Explore, 23, 24

### F

Factor, 23, 39, 41, 55, 71  
 Factor Analysis, 71  
 Faktorová analýza, 71  
 Frequencies, 19, 20, 101, 102

### G

General Linear Model, 5, 41, 42  
 Geoprostorové modelování, 128  
 Geospatial Association Rules, 139  
 Graf časových řad, 93, 94  
 Graf kumulativních proporcí, 5, 29, 30, 31

### H

Hierarchical Cluster, 6, 63, 64  
 Hierarchické seskupování, 63

### Ch

Chi-Square, 6, 83  
 Chi-square Test, 82

### I

Independent Samples, 6, 78, 79, 88, 90  
 Independent-Samples T Test, 5, 36

### J

Jednoduchá analýza rozptylu, 39  
 Jednoduchý lineární model, 41, 42  
 Jednovýběrové neparametrické testy, 76, 77  
 Jednovýběrový T-test, 34, 35

### K

Klasifikační metody, 58  
 K-Means Cluster, 61, 62  
 Kontingenční a asociační tabulky, 24, 26  
 Kontingenční tabulky pro mnohonásobné odpovědi, 102, 103  
 Kontrola kvality, 120  
 Korelace, 44  
 Korelační analýza, 44  
 Kostky OLAP, 9, 10  
 Kvantilový graf, 31, 32

## Rejstřík

### L

Linear Regression, 51, 52  
Lineární regrese, 51, 52

### M

Means, 33, 34, 39  
Mnohonásobné odpovědi, 100  
Mnohorozměrné škálování, 73, 74  
Multidimensional Scaling, 6, 73, 74  
Multiple Response, 6, 100, 101, 102, 103

### N

Nearest Neighbor, 6, 67, 69  
Nearest Neighbor Analysis, 6, 67  
Neparametrické testy, 76, 78, 79, 80, 81  
Neparametrické testy pro nezávislé výběry, 78, 79  
Neparametrické testy pro závislé výběry, 80, 81  
Nonparametric Tests, 6, 76, 77, 78, 79, 80, 81, 83, 85, 86, 87, 88, 90, 91, 92

### O

Obecný lineární model, 41  
OLAP Cubes, 5, 9, 10  
One Sample, 6, 76, 77  
One-Sample Kolmogorov-Smirnov test, 87  
One-Sample T Test, 5, 34, 35  
One-Way ANOVA, 5, 39  
Ordinal, 55, 56  
Ordinální regrese, 55, 56

### P

Paired-Samples T Test, 5, 38  
Parciální korelace, 45  
Pareto Charts, 6, 121, 123  
Paretův graf, 6, 121, 122, 123  
Párový T-test, 38  
Partial Correlations, 5, 45  
Podmíněné formátování tabulek, 163  
Poměrové statistiky, 27, 28  
Popisné statistiky, 19, 21  
Porovnání průměrů, 33  
P-P Plot, 5, 29, 30, 31  
Prokládání křivky, 54, 55  
Průměry, 33

### Q

Q-Q Plot, 5, 31, 32  
Quality Control, 6, 120, 121, 123

### R

Ratio, 5, 27, 28  
Redukce dat, 71  
Regresní analýza, 41, 49  
Regression, 5, 49, 50, 52, 55, 56  
Related Samples, 6, 80, 81, 91, 92  
Reliability Analysis, 6, 73  
Reports, 5, 8, 9, 10, 12, 14, 17  
ROC Curve, 6, 125, 126  
ROC křivka, 125, 126  
Runs, 6, 85, 86  
Runs Test, 6, 85

### Ř

Řídící diagramy, 120, 121

### S

Sady proměnných pro mnohonásobné odpovědi, 100, 101, 102  
Scale, 6, 55, 68, 73, 74  
Sequence Charts, 6, 93, 94  
Seskupování metodou K-means, 61, 62  
Shrnutí informací o proměnných a datovém souboru, 8, 9  
Simulace, 105  
Simulation, 105  
Spatial and Temporal Modeling, 128  
Spatial Modeling, 128  
Spatio-Temporal Prediction, 150  
Style, 19, 21, 25, 33, 44, 51, 163  
Sumář případů, 11, 12  
Sumář v řádcích, 13, 14  
Sumář ve sloupcích, 16, 17  
Summaries in Columns, 16, 17  
Summaries in Rows, 13, 14, 16

### Š

Škálování, 73

### T

Test Chí-kvadrát, 82, 83  
Test pořadí, 85, 86  
Tests for Several Independent Samples, 6, 89  
Tests for Several Related Samples, 92  
Testy pro dva nezávislé výběry, 87, 88  
Testy pro dva závislé výběry, 90, 91  
Testy pro několik nezávislých výběrů, 89, 90  
Testy pro několik závislých výběrů, 92  
Time Series, 6, 93, 94, 97, 99  
T-test pro dva nezávislé výběry, 36  
Two-Independent-Samples Tests, 6, 87

## Rejstřík

Two-Related-Samples Tests, 6, 90  
TwoStep Cluster, 58, 59

### U

Univariate, 5, 41, 42

### V

Výpočet četností pro mnohonásobné odpovědi,  
101, 102

Vzájemné korelace, 97, 98, 99  
Vzdálenosti, 47

### Z

Zobecněná geoprostorová asociační pravidla, 139  
Zprávy, 8



## Anglicko – český slovník

2-Digit ..... dvoumístný

## A

a, an ..... neurč. člen (= nějaký)  
 above value ..... nad hodnotou  
 absolute value ..... absolutní hodnota  
 actual ..... skutečný  
 add ..... přidat  
 adjusted ..... adjustovaný, upravený  
 advanced ..... pokročilý, složitý  
 after ..... po  
 aggregate ..... agregovat  
 alignment ..... zarovnání  
 all ..... vše, všechno, celé  
 allocation ..... alokace  
 alphabetic ..... abecední  
 already ..... již  
 analyze ..... analýza, analyzovat  
 ANOVA ..... analýza rozptylu  
 appear as ..... zobrazit se jako  
 append ..... přidat  
 apply ..... použít, aplikovat  
 approximately ..... přibližně  
 area ..... oblast, plošný graf  
 arithmetic ..... početní  
 arrow ..... šipka  
 as ..... jako  
 ascending ..... vzestupně, vzestupný  
 aspect ratio ..... poměr výšky a šířky  
 assign ..... označit, určit, provést  
 assumed ..... předpokládaný  
 automatic ..... automatický  
 available ..... dostupný  
 average ..... průměr  
 axis ..... osa

## B

back ..... zpět  
 backward ..... postupné vyjímání  
 bar ..... sloupec, sloupcový graf  
 bar chart ..... sloupcový graf  
 base ..... základ  
 based on ..... založeno na  
 become ..... stát se  
 before ..... před  
 begin ..... začít  
 belong ..... patřit, náležet  
 below value ..... pod hodnotou  
 between ..... mezi  
 binary ..... binární, dvouhodnotový  
 block ..... blok (výpočtu), část  
 bookmark ..... záložka  
 border ..... orámování, rámeček  
 both ..... obojí  
 bottom ..... dolní okraj  
 bound ..... mez  
 boxplot ..... boxplot  
 branch ..... větev

break ..... přerušení  
 break variable ..... dělicí proměnná  
 browse ..... prohlížet, probírat, procházet

## C

calculate ..... vypočítat  
 caption ..... legenda, poznámky, vysvětlivky  
 case ..... případ  
 casewise ..... po případech  
 category ..... kategorie  
 cell ..... pole (tabulky), buňka  
 center ..... střed  
 classify ..... klasifikovat, klasifikace  
 clear ..... čistit, vymazat  
 close ..... zavřít, uzavřít  
 cluster ..... skupina, seskupovací  
 cluster by ..... třídění podle  
 coach ..... průvodce  
 code ..... kód  
 coded as ..... kódovaný jako  
 collinearity ..... kolinearita  
 color ..... barva  
 column ..... sloupec  
 comma ..... čárka  
 command ..... příkaz  
 compare ..... srovnávat, srovnání, porovnat  
 compute ..... vypočítat, výpočet  
 condensed ..... zhuštěný  
 condition ..... podmínka, předpoklad  
 conf. interval ..... interval spolehlivosti  
 connector ..... spojnice  
 consistency ..... konzistence  
 constant ..... konstanta  
 continue ..... pokračovat  
 continuous ..... spojitý  
 contrast ..... kontrast, porovnat  
 control ..... řízení procesů  
 convert ..... převést  
 copy ..... kopírovat, kopie  
 correlation ..... korelace  
 count ..... počet, četnost  
 covariance ..... kovariance  
 create ..... vytvořit  
 criterion ..... kritérium  
 crosstabs ..... kontingenční tabulka  
 currency ..... měna, běžný, současný  
 current ..... běžné, stávající  
 current status ..... současný stav  
 curve ..... křivka  
 custom ..... vlastní  
 customized ..... vlastní nastavení  
 cut ..... vytrhnout, vystříhnout  
 cut points ..... dělicí body

## D

data ..... data  
 database ..... databáze  
 dataset ..... datový soubor  
 date ..... datum  
 decimal ..... desetinný  
 default ..... implicitní, výchozí

define ..... definovat  
 delete ..... vymazat, zrušit, smazat  
 delimiter ..... oddělovatel  
 denominator ..... jmenovatel  
 dependent ..... závislý  
 depth ..... hloubka  
 descending ..... sestupný, sestupně  
 describe ..... popsat  
 descriptive ..... popisný  
 design ..... vzhled, uspořádání  
 determine ..... stanovit  
 deviation ..... deviace, odchylka  
 dictionary ..... slovník  
 difference ..... diference, rozdíl  
 different ..... jiný, rozdílný, rozdílný  
 direction ..... orientace  
 directory ..... adresář, složka  
 discrete ..... samostatná, kategorizovaná  
 dissimilarities ..... nepodobnost  
 dispersion ..... rozptýlení  
 display ..... zobrazení  
 distance ..... vzdálenost  
 distribution ..... rozdělení, dělení  
 document ..... dokument  
 draft ..... pracovní  
 draft viewer ..... pracovní (textový) výstup  
 drop ..... pokles, rozsah, vynechat

E

each ..... každý  
 edit ..... úprava, upravovat, editace  
 element ..... prvek  
 end ..... konec  
 enter ..... určený vstup  
 equal ..... rovný, roven  
 equation ..... rovnice  
 error ..... chyba  
 error bar ..... graf rozptýlení  
 estimate ..... odhadovat  
 estimation ..... odhad  
 Euclidean ..... euklidovský  
 Euclidean distance ..... eukleidovská vzdálenost  
 every ..... každý  
 exactly ..... přesně  
 example ..... příklad  
 exclude ..... vynechat, vyloučit  
 execute ..... vykonat  
 exist ..... existovat  
 exit ..... východ, ukončení  
 expect ..... očekávat  
 explore ..... prozkoumat  
 export ..... exportovat  
 expression ..... vyjádření; výraz  
 external ..... vnější

F

factor ..... faktor, nezávislá proměnná  
 field ..... pole (databázové)  
 file ..... soubor  
 fill ..... výplň  
 filter ..... filtr, filtrovat, vynechat z analýz

final ..... finální, konečný  
 find ..... najít  
 first ..... první  
 fixed ..... fixované, pevně stanovené  
 font ..... typ písma  
 footer ..... zápatí  
 footnote ..... poznámka pod čarou  
 for example ..... například  
 format ..... formát  
 forward ..... postupné zahrnování  
 fraction ..... podíl  
 frame ..... rám, orámování  
 freefield ..... volně umístěné  
 frequencies ..... četnosti  
 from ..... z  
 full ..... plný, zcela  
 function ..... funkce  
 F-value ..... hodnota statistiky F

G

general ..... obecný  
 generate ..... generování, vytvoření  
 get from ..... odečíst z  
 global ..... celkový  
 go to ..... jdi na  
 grand total ..... celkový souhrn  
 graph ..... graf  
 grid line ..... mřížka  
 group ..... skupina  
 grouping ..... seskupující, třídící

H

handle ..... zacházet  
 harmonic aver. .... harmonický průměr  
 hide ..... skrýt  
 hierarchical ..... hierarchický  
 high ..... vysoký, výška, horní  
 histogram ..... histogram

C

change ..... vyměnit, změna  
 character ..... znak  
 chart ..... graf  
 Chi-square ..... chí-kvadrát  
 choose ..... zvolit

I

if ..... jestliže, když  
 image ..... obraz  
 immediately ..... okamžitě, ihned, přímo  
 importance ..... důležitost  
 include ..... zahrnout  
 independent ..... nezávislý  
 individual ..... individuální  
 infinite ..... nekonečný  
 influence statistics ..... statistiky vlivu  
 info ..... informace  
 inch ..... palec (délková míra)  
 initial ..... počáteční  
 inner ..... vnitřní  
 input ..... vstupní

insert ..... vložit  
 inside ..... uvnitř (intervalu)  
 insignificant ..... nevýznamný  
 interactive ..... interaktivní  
 interval ..... intervalový  
 item ..... položka

J

justification ..... zarovnání

K

keep ..... držet  
 key ..... klíč  
 key variable ..... klíčová proměnná  
 kind ..... druh  
 kurtosis ..... špičatost

L

label ..... označení, popis  
 language ..... jazyk  
 large ..... velký  
 last ..... poslední  
 layer ..... vrstva, úroveň  
 layout ..... rozvržení  
 leaf ..... list  
 left ..... nalevo  
 legend ..... legenda, popis  
 length ..... délka  
 level ..... úroveň  
 lighting ..... osvětlení  
 likelihood ..... věrohodnost  
 limit ..... hranice, limit  
 line ..... řádek, čára  
 linearity ..... linearita, přímková  
 linegraph ..... spojnicový graf  
 list ..... seznam  
 list cases ..... vypsát případy  
 listwise ..... pro celý seznam, celý případ  
 look ..... vzhled  
 look at ..... dívat se na  
 low ..... nízký, spodní, dolní

M

mail ..... pošta  
 marker ..... značka  
 match ..... spojení  
 matrix ..... matice  
 maximum ..... maximální hodnota  
 mean ..... průměr  
 measure ..... měření, míra  
 measurement ..... měření, měřící  
 median ..... medián  
 medium ..... střední  
 membership ..... členství  
 memory ..... paměť  
 menu ..... nabídka  
 merge ..... spojit, spojování  
 message ..... vzkaz  
 method ..... metoda  
 midpoint ..... střední bod  
 minimum ..... minimální hodnota

missing ..... chybějící, vynechaná  
 mixed ..... tříděný  
 mode ..... modus  
 model fit ..... vhodný model  
 modify ..... změnit  
 monospaced ..... neproporcionální  
 more than ..... více než  
 multidimensional ..... mnohorozměrný  
 multiple ..... mnohonásobný, vícečetný

N

name ..... jméno  
 negative ..... záporný  
 new ..... nový  
 next ..... další, příští, následující  
 no ..... ne, žádný  
 node ..... uzel  
 noise ..... šum  
 nominal ..... nominální  
 nonparametric ..... neparametrický  
 normal ..... normální  
 normal curve ..... Gaussova křivka  
 note ..... poznámka  
 number ..... číslo, počet  
 number of cases ..... počet případů  
 numerator ..... číselník  
 numeric ..... numerický, číselný

O

object ..... objekt  
 observation ..... pozorování  
 observe ..... pozorovat  
 old ..... starý  
 omit ..... vynechat  
 one-sample ..... jednovýběrový  
 one-tailed ..... jednostranný (test)  
 one-way ..... jedno třídění, jeden  
 only ..... jen, pouze  
 open ..... otevřít  
 option ..... možnost, volba, předvolba  
 order ..... pořadí  
 ordinal ..... ordinální, pořadová  
 organize ..... uspořádat  
 other ..... ostatní  
 outer ..... vnější  
 outlier ..... vzdálené pozorování  
 output ..... výstup, výstupní  
 outside ..... mimo (interval)  
 overlay ..... ve vrstvách  
 overwrite ..... přepsat

P

page ..... stránka, strana  
 paired ..... párový  
 paired variables ..... spárované proměnné  
 pairwise ..... po dvojicích  
 partial ..... dílčí, parciální  
 password ..... heslo  
 paste ..... přilepit, přenést, vložit  
 pattern ..... vzor  
 percentage ..... procento, procentní

percentile ..... percentil  
 period ..... tečka  
 pie ..... koláč (výšečový graf)  
 pivot ..... pivotní  
 place ..... místo  
 plot ..... graf  
 point ..... bod  
 positive ..... kladný  
 possible ..... možný  
 power ..... mocnina  
 predict ..... předpovědět  
 preference ..... preference, předvolba  
 prefix ..... předpona  
 present ..... prezentovat, představit, být přítomen  
 previous ..... předchozí  
 print ..... tisk  
 probability ..... pravděpodobnost  
 produce ..... produkovat  
 property ..... vlastnost  
 proportion ..... proporce

Q

quartile ..... kvartil  
 query ..... dotaz (databázový)

R

raise ..... vyzvednutí (do popředí)  
 random ..... náhodný  
 range ..... rozpětí, interval  
 rank ..... pořadí  
 read ..... číst  
 rearrange ..... přeskupit  
 recently ..... naposledy  
 recode ..... překódování  
 record ..... záznam, zaznamenat  
 regression ..... regrese, regresní  
 related ..... příbuzný, závislý  
 remove ..... zrušit, odstranit, vyjmout  
 repeat ..... opakovat  
 replace ..... zaměnit, nahradit, přepsat  
 report ..... zpráva, vypsát  
 represented ..... zastoupený  
 rescale ..... změnit škálu  
 reset ..... zrušit nastavení  
 residual ..... reziduum, reziduální, zbytek  
 response ..... odpověď  
 restructure ..... restrukturalizovat  
 retrieve ..... obnovené (převzaté)  
 right ..... napravo  
 root ..... odmocnina  
 rotate ..... otáčet, rotovat  
 rotation ..... rotace  
 round ..... zaokrouhlit  
 row ..... řádek  
 run ..... spustit, běžet

S

S.E. Mean ..... stand. chyba průměru  
 same ..... stejný, totožný  
 sample ..... výběr, výběrový soubor  
 satisfy ..... splnit, uspokojit

save ..... uložit  
 save as ..... uložit jako  
 scale ..... škála, číselný  
 scaling ..... škálování  
 scatter ..... rozložení  
 scatterplot ..... bodový, souřadnicový graf  
 script ..... skript (program)  
 scroll ..... rolovat  
 search ..... hledat  
 select ..... vybrat  
 selection ..... selekce, výběr  
 send ..... poslat  
 separate ..... samostatný, oddělený  
 separator ..... oddělovač  
 sequence ..... řada  
 sequential ..... následný, postupný  
 set ..... množina, soubor, sada  
 setting ..... nastavení  
 shape ..... typ, tvar  
 show ..... ukázat  
 significance ..... významnost  
 significant ..... významný  
 similar ..... podobný  
 similarity ..... podobnost  
 simple ..... jednoduchý  
 single ..... jednotlivý, jediný  
 size ..... velikost, rozsah  
 skewness ..... šikmost, zešikmení  
 slice ..... díl  
 small ..... malý  
 sort ..... seřadit, řadit  
 sort by size ..... seřadit podle velikosti  
 sound ..... zvuk  
 source ..... zdroj, zdrojový  
 space ..... mezera  
 spacing ..... rozteč, vzdálenost  
 specification ..... specifikace  
 specify ..... určit, stanovit, zpřesnit  
 spin ..... otáčení  
 split ..... rozdělení, rozštěpení  
 spread ..... rozptýlení  
 stacked ..... skládaný  
 standard ..... standardní  
 standard deviation ..... směrod. odchylka  
 standard error ..... standardní chyba  
 standardized ..... standardizovaný  
 start ..... začátek  
 statistic ..... statistika  
 stem-and-leaf ..... graf „lodyha s listy“  
 step ..... krok  
 stepwise ..... krokově  
 stop ..... zastavení  
 stored ..... uložený  
 strictly ..... přísně  
 string ..... znak, znakový  
 style ..... styl  
 suffix ..... přípona  
 sum ..... suma hodnot (součet)  
 sum of squares ..... součet čtverců  
 summarize ..... sumarizace  
 summary ..... sumární statistiky

suppress ..... potlačit  
 syntax ..... syntaxe  
 system-missing ..... systémově vynechávaný

T

table ..... tabulka  
 target ..... vybraný  
 template ..... šablona  
 temporary ..... dočasný  
 tendency ..... tendence  
 test ..... testovaný  
 test value ..... testová hodnota,  
 test variable ..... testovaná proměnná  
 threshold ..... práh  
 through ..... až  
 tick ..... značka  
 time ..... čas  
 time series ..... časové řady  
 title ..... titul  
 to be created ..... jež se mají vytvořit  
 together ..... dohromady, společně  
 tool ..... nástroj  
 toolbar ..... panel nástrojů  
 top ..... horní okraj  
 topic ..... téma  
 total perc. ..... celková procenta  
 transform ..... přeměnit  
 transformation ..... transformace  
 transpose ..... transponovat  
 treatment ..... nakládání  
 tree ..... strom  
 tuning ..... ladění  
 tutorial ..... konzultace  
 two-tailed ..... dvoustranný (test)  
 two-way ..... dvojný  
 type ..... typ

U

underscore ..... podtrhnout  
 undo ..... zpět, zrušení akce

uniform ..... rovnoměrný  
 unique ..... jedinečný  
 univariate ..... s jednou proměnnou  
 unpaired ..... nespárované  
 unselected ..... nevybraný  
 unstandardized ..... nestandardizovaný  
 unweighed ..... nevážený  
 update ..... zde: načtení  
 use ..... použít  
 user-missing ..... uživatelsky vynechávaný  
 utilities ..... pomocné funkce

V

valid ..... platný  
 value ..... hodnota  
 variable ..... proměnná  
 variable list ..... seznam proměnných  
 variance ..... rozptyl  
 vertically ..... vertikálně, svisle  
 view ..... pohled  
 viewer ..... prohlížeč, výstup  
 visible ..... viditelný

W

warning ..... upozornění, varování  
 weight ..... váha, vážit  
 weighted ..... vážený  
 width ..... šířka  
 window ..... okno  
 with ..... s  
 within ..... uvnitř  
 WLS ..... metoda vážených nejmenších čtverců  
 working ..... pracovní  
 write ..... psát, zapsat

Y

year ..... rok

Z

zero-order correlations. .... neparcializované korelace

## Česko – anglický slovník

### A

abecední ..... alphabetical  
 absolutní hodnota ..... absolute value  
 adresář ..... directory  
 agregovat ..... aggregate  
 alokace ..... allocation  
 analýza, analyzovat ..... analyze  
 analýza rozptylu ..... ANOVA  
 aplikovat ..... apply  
 automatický ..... automatic  
 až ..... through

### B

barva ..... color  
 běžet ..... run  
 běžný ..... current  
 binární ..... binary  
 blok (výpočtu) ..... block  
 bod ..... point  
 bodový ..... scatterplot  
 boxplot ..... boxplot  
 buňka ..... cell  
 být přítomen ..... present

### C

celé ..... all  
 celkem ..... total  
 celkový ..... total, global  
 celkový souhrn ..... grand total  
 celý případ ..... listwise

### Č

čára ..... line  
 čárka ..... comma  
 čas ..... time  
 časové řady ..... time series  
 četnosti ..... counts, frequencies  
 číselný ..... numeric  
 číslo ..... number  
 číst ..... read  
 čistit ..... clear  
 čítec ..... numerator  
 členství ..... membership  
 čtení ..... reading

### D

další ..... next  
 data ..... data  
 databáze ..... database  
 datový soubor ..... dataset  
 datum ..... date  
 definovat ..... define  
 dělení ..... distribution  
 dělicí bod ..... cut point  
 dělicí proměnná ..... break variable  
 délka ..... length  
 desetinný ..... decimal  
 deviace ..... deviation

diference ..... difference  
 díl ..... slice  
 dílčí ..... partial  
 kategorizovaná ..... discrete  
 dívat se na ..... look at  
 dočasný ..... temporary  
 dohromady ..... together  
 dokument ..... document  
 dolní ..... low  
 dolní okraj ..... bottom  
 dostupný ..... available  
 dotaz (databázový) ..... query  
 druh ..... kind  
 držet ..... keep  
 důležitost ..... importance  
 dvojný ..... two-way  
 dvouhodnotový ..... binary  
 dvoumístný ..... 2-Digit  
 dvoustranný (test) ..... two-tailed

### E

editace ..... edit  
 eukleidovská vzdálenost ..... Euclidean distance.  
 euklidovský ..... Euclidean  
 existovat ..... exist  
 exportovat ..... export

### F

faktor ..... factor  
 filtr, filtrovat ..... filter  
 finální ..... final  
 fixovaný ..... fixed  
 formát ..... format  
 funkce ..... function

### G

Gaussova křivka ..... normal curve  
 generovat ..... generate  
 graf ..... graph, chart, plot  
 graf „lodyha s listy“ ..... stem-and-leaf  
 graf rozptýlení ..... error bar

### H

harmonický průměr ..... harmonic average  
 heslo ..... password  
 hierarchický ..... hierarchical  
 histogram ..... histogram  
 hladina významnosti ..... significance level  
 hledat ..... search  
 hloubka ..... depth  
 hodnota ..... value  
 hodnota statistiky F ..... F-value  
 horní ..... high  
 horní okraj ..... top  
 hranice ..... limit

### C

chí-kvadrát ..... Chi-square  
 chyba ..... error  
 chybějící ..... missing

I

ihned ..... immediately  
 implicitní ..... default  
 individuální ..... individual  
 informace ..... info  
 interaktivní ..... interactive  
 interval ..... range, interval  
 interval spolehlivosti ..... confidence interval

J

jako ..... as  
 jazyk ..... language  
 jdi na ..... go to  
 jeden ..... one-way  
 jedinečný ..... unique  
 jediný ..... single  
 jedno třídění ..... one-way  
 jednoduchý ..... simple  
 jednostranný (test) ..... one-tailed  
 jednovýběrový ..... one-sample  
 jen ..... only  
 jestliže ..... if  
 jiný ..... different  
 již ..... already  
 jméno ..... name, název  
 jmenovatel ..... denominator

K

kategorie ..... category  
 každý ..... every, each  
 když ..... if  
 kladný ..... positive  
 klasifikace, klasifikovat ..... classify  
 klíč ..... key  
 kód ..... code  
 koláč (výšečový graf) ..... pie  
 kolinearita ..... collinearity  
 konec ..... end  
 konečný ..... final  
 konstanta ..... constant  
 kontingenční tabulka ..... crosstabs  
 kontrast ..... contrast  
 konzistence ..... consistency  
 konzultace ..... tutorial  
 kopie, kopírovat ..... copy  
 korelace ..... correlation  
 kovariance ..... covariance  
 kritérium ..... criterion  
 krok ..... step  
 krokově ..... stepwise  
 křivka ..... curve  
 kvartil ..... quartile

L

ladění ..... tuning  
 legenda ..... legend, caption  
 limit ..... limit  
 linearita ..... linearity  
 list ..... leaf

M

malý ..... small  
 matice ..... matrix  
 maximální hodnota ..... maximum  
 medián ..... median  
 měna ..... currency  
 měřit ..... measure  
 měření ..... measurement  
 metoda ..... method  
 metoda vážených nejmenších čtverců ..... WLS  
 mez ..... bound  
 mezera ..... space  
 mezi ..... between  
 mimo (interval) ..... outside  
 minimální hodnota ..... minimum  
 míra ..... measure  
 místo ..... place  
 mnohonásobný ..... multiple  
 mnohorozměrný ..... multidimensional  
 množina ..... set  
 mocnina ..... power  
 modus ..... mode  
 možnost ..... option  
 možný ..... possible  
 mřížka ..... grid line

N

nabídka ..... menu  
 nad hodnotou ..... above value  
 náhodný ..... random  
 nahradit ..... replace  
 najít ..... find  
 nakládání ..... treatment  
 nalevo ..... left  
 náležet ..... belong  
 naposledy ..... recently  
 napravo ..... right  
 například ..... for example  
 následný ..... sequential  
 následující ..... next  
 nastavení ..... setting  
 nástroj ..... tool  
 ne ..... no  
 nedokončený ..... undone  
 nekonečný ..... infinite  
 neparametrický ..... nonparametric  
 neparcializované korelace ..... zero-order correlations.  
 nepodobnost ..... dissimilarities  
 neproporcionální ..... monospaced  
 nespárované ..... unpaired  
 nestandardizovaný ..... unstandardized  
 nevážený ..... unweighed  
 nevybraný ..... unselected  
 nevýznamný ..... insignificant  
 nezávislá proměnná ..... factor  
 nezávislý ..... independent  
 nízký ..... low  
 nominální ..... nominal  
 normální ..... normal  
 nový ..... new  
 numerický ..... numeric

O

obecný ..... general  
objekt ..... object  
objevit se ..... appear  
oblast ..... area  
obnovené (převzaté) ..... retrieve  
obojí ..... both  
obraz ..... image  
očekávaný ..... expected  
oddělený ..... separate  
oddělovač ..... separator  
oddělovatel ..... delimiter  
odhad ..... estimation  
odchylka ..... deviation  
odmocnina ..... root  
odpověď ..... response  
odstranit ..... remove  
okamžitě ..... immediately  
okno ..... window  
opakovat ..... repeat  
orámování ..... frame, border  
ordinální ..... ordinal  
orientace ..... direction  
osa ..... axe  
ostatní ..... other  
osvětlení ..... lighting  
otáčení ..... spin  
otevřít ..... open  
označení ..... label

P

palec (délková míra) ..... inch  
paměť ..... memory  
panel nástrojů ..... toolbar  
parciální ..... partial  
párový ..... paired  
patřit ..... belong  
percentily ..... percentile  
pevně stanovený ..... fixed  
pivovní ..... pivot  
platný ..... valid  
plný ..... full  
plošný graf ..... area  
po ..... after  
po dvojicích ..... pairwise  
po případech ..... casewise  
počáteční ..... initial  
počet ..... count, number  
početní ..... arithmetic  
pod hodnotou ..... below value  
podíl ..... fraction  
podmínka ..... condition  
podobnost ..... similarities  
podobný ..... similar  
podtrhnout ..... underscore  
pokles ..... drop  
pokračovat ..... continue  
pokročilý ..... advanced  
pole (databázové) ..... field  
pole (tabulky) ..... cell  
položka ..... item

poměr výšky a šířky ..... aspect ratio  
pomocné funkce ..... utilities  
popis ..... legend, label  
popisný ..... descriptive  
popisy ..... labels  
popsat ..... describe  
porovnat ..... compare, contrast  
pořadí ..... rank, order  
pořadový ..... ordinal  
poslat ..... send  
poslední ..... last  
postupný ..... sequential  
pošta ..... mail  
potlačit ..... suppress  
pouze ..... only  
použít ..... use, apply  
použité ..... used  
poznámka ..... note, caption  
poznámka pod čarou ..... footnote  
pozorování ..... observation  
pozorovaný ..... observed  
pracovní ..... working  
pracovní výstup ..... draft viewer  
práh ..... threshold  
pravděpodobnost ..... probability  
predikované hodnoty ..... predicted val.  
prezentovat ..... present  
pro celý seznam ..... listwise  
pro nezávislé výběry ..... indep.-sample  
probírat ..... browse  
procento ..... percent  
produkovat ..... produce  
prohlížeč ..... viewer  
prohlížet ..... browse  
procházet ..... browse  
prokládání křivky ..... cur. estimation  
proměnná ..... variable  
proporce ..... proportion  
provést ..... assign  
prozkoumat ..... explore  
průměr ..... average, mean  
průvodce ..... coach  
prvek ..... element  
první ..... first  
před ..... before  
předchozí ..... previous  
předpokládaný ..... assumed  
předpona ..... prefix  
představit ..... present  
předvolba ..... option  
překódování ..... recode  
přeměnit ..... transform  
přenést ..... paste  
přepsat ..... overwrite, replace  
přerušit ..... break  
přeskupit ..... rearrange  
přesně ..... exactly  
převést ..... convert  
přibližně ..... approximately  
příbuzný ..... related  
přidat ..... add, append



příkaz ..... command  
 příklad ..... example  
 přilepit ..... paste  
 přímkový ..... linearity  
 přímo ..... immediately  
 případ ..... case  
 přípona ..... suffix  
 přísně ..... strictly  
 příští ..... next  
 psát ..... write

R

rám ..... frame  
 rámeček ..... border  
 regrese, regresní ..... regression  
 reziduum, reziduální ..... residual  
 restrukturalizovat ..... restructure  
 rok ..... year  
 rolovat ..... scroll  
 rotace ..... rotation  
 roven ..... equal  
 rovnice ..... equation  
 rovnoměrný ..... uniform  
 rovný ..... equal  
 rozdělení ..... split, distribution  
 rozdíl ..... difference  
 rozdílný ..... different  
 rozlišení tisku ..... print resolution  
 rozložení ..... scatter  
 rozpětí ..... range  
 rozptýl ..... variance  
 rozptýlení ..... dispersion, spread  
 rozsah ..... drop, size  
 rozštěpení ..... split  
 rozteč ..... spacing  
 rozvržení ..... layout

Ř

řada ..... sequence  
 řádek ..... row, line  
 řadit ..... sort  
 řádková procenta ..... row percentage  
 řízení procesů ..... control

S

s ..... with  
 s jednou proměnnou ..... univariate  
 sada ..... set  
 samostatný ..... separate  
 selekce ..... selection  
 seřadit ..... sort  
 seřadit podle velikosti ..... sort by size  
 seskupovací ..... cluster  
 seskupující ..... grouping  
 sestupný ..... descending  
 seznam ..... list  
 seznam proměnných ..... variable list  
 shodné skupiny ..... equal groups  
 skládaný ..... stacked  
 skript (program) ..... script  
 skrýt ..... hide

skrytý ..... hidden  
 skupina ..... group, cluster  
 skutečný ..... actual  
 sloupcová procenta ..... column perc.  
 sloupcový graf ..... bar chart  
 sloupec ..... column, bar  
 slovník ..... dictionary  
 složitý ..... advanced  
 složka ..... directory  
 smazat ..... delete  
 směrodatná odchylka ..... standard deviation  
 soubor ..... file, set  
 součet čtverců ..... sum of squares  
 souřadnicový graf ..... scatterplot  
 spárované proměnné ..... paired var.  
 specifikace ..... specification  
 splnit ..... satisfy  
 spodní ..... low  
 spojení ..... match  
 spojit ..... merge  
 číselný ..... continuous, scale  
 spojnice ..... connector  
 spojnicový graf ..... linegraph  
 spojovací proměnná ..... key variable  
 společně ..... together  
 spustit ..... run  
 srovnání, srovnávat ..... compare  
 stand. chyba průměru ..... S.E. Mean  
 standardizovaný ..... standardized  
 standardní ..... standard  
 standardní chyba ..... standard error  
 stanovit ..... determine, specify  
 starý ..... old  
 stát se ..... become  
 statistika ..... statistic  
 stávající ..... current  
 strana ..... page  
 strom ..... tree  
 střední ..... central, medium  
 střední bod ..... midpoint  
 styl ..... style  
 suma hodnot (součet) ..... sum  
 sumarizace ..... summarize  
 sumární statistiky ..... summary  
 svisle ..... vertically  
 syntaxe ..... syntax  
 systémově vynechávaný ..... system-missing

Š

šablona ..... template  
 šikmost ..... skewness  
 šipka ..... arrow  
 šířka ..... width  
 škála ..... scale  
 škálování ..... scaling  
 špičatost ..... kurtosis  
 šum ..... noise

T

tabulka ..... table  
 tečka ..... period

téma ..... topic  
 tendence ..... tendency  
 testová hodnota ..... test value  
 testovaná proměnná ..... test variable  
 testovaný ..... test  
 tisk ..... print  
 titul ..... title  
 totožný ..... same  
 transformace ..... transformation  
 transponování ..... transpose  
 třídění podle ..... cluster by  
 tvar ..... shape  
 typ ..... type

U

ukázat ..... show  
 ukončení ..... exit  
 uložený ..... stored  
 uložit ..... save  
 uložit jako ..... save as  
 upozornění ..... warning  
 úprava, upravit ..... edit  
 upravený ..... adjusted (adj.)  
 určený vstup ..... enter  
 určit ..... specify, assign  
 úroveň ..... level  
 uspokojit ..... satisfy  
 uspořádání ..... design  
 uspořádat ..... organize  
 uvnitř ..... within  
 uvnitř (intervalu) ..... inside  
 uzavřít ..... close  
 uzel ..... node  
 uživatelsky vynechávaný ..... user-missing

V

váha ..... weight  
 varování ..... warning  
 vážený ..... weighed  
 vážít ..... weight  
 ve vrstvách ..... overlay  
 velikost ..... size  
 velký ..... large  
 věrohodnost ..... likelihood  
 vertikálně ..... vertically  
 větev ..... branch  
 vhodný model ..... model fit  
 více než ..... more than  
 vícečetný ..... multiple  
 viditelný ..... visible  
 vlastní ..... custom  
 vlastní nastavení ..... customized  
 vlastnost ..... property  
 vložit ..... paste, insert  
 vnější ..... external, outer  
 vnitřní ..... inner  
 volba ..... option  
 volně umístěné ..... freefield  
 vrstva ..... layer  
 vstupní ..... input  
 vše ..... all

výběr ..... sample, selection  
 vybrat ..... select  
 výchozí ..... default  
 vyjádření; výraz ..... expression  
 vyjmout ..... remove  
 vykonat ..... execute  
 vyloučit ..... exclude  
 vymazat ..... clear, delete  
 vyměnit ..... change  
 vynechaná ..... missing  
 vynecháno z analýzy ..... filtered  
 vynechat ..... exclude, drop, omit  
 výplň ..... fill  
 vypočítat ..... compute, calculate  
 vypsát ..... report  
 vypsát případy ..... list cases  
 výsečový graf ..... pie graph  
 vyseknutý ..... deleted  
 vysoký ..... high  
 vystřihnout ..... cut  
 výstup ..... output, viewer  
 vysvětlivky ..... caption  
 výška ..... high  
 vytrhnout ..... cut  
 vytvoření ..... generate  
 vytvořit ..... create  
 významnost ..... significance  
 významný ..... significant  
 vyzvednutí (do popředí) ..... raise  
 vzdálené pozorování ..... outlier  
 vzdálenost ..... distance, spacing  
 vzestupný ..... ascending  
 vzhled ..... design, look  
 vzkaz ..... message  
 vzor ..... pattern

Z

Z ..... from  
 začátek ..... start  
 začít ..... begin  
 zahrnout ..... include  
 zacházení ..... handling  
 základ ..... base  
 založeno na ..... based on  
 záložka ..... bookmark  
 zaměnit ..... replace  
 zaokrouhlit ..... round  
 zápatí ..... footer  
 záporný ..... negative  
 zapsat ..... write  
 zarovnání ..... alignment, justification  
 zastavení ..... stop  
 zastoupený ..... represented  
 závislý ..... dependent, related  
 zavřít ..... close  
 záznam ..... record  
 zaznamenaný ..... recorded  
 zbytek ..... residual  
 zdroj ..... source  
 zdrojový ..... source  
 zešikmení ..... skewness

zhuštěný.....condensed  
změna.....change  
změnit.....modify  
změnit škálu.....rescale  
značka.....marker  
znak.....character, string  
znakový.....string  
zobrazit.....display  
zobrazit se jako.....appear as  
zpět.....back, undo

zpráva.....report  
zpřesnit.....specify  
zrušit akci.....undo  
zrušit nastavení.....reset  
zvolit.....choose  
zvuk.....sound

Ž  
žádný.....no, none