

1.9 Klasifikační metody (Classify)

1.9.1 Dvoustupňové seskupování (TwoStep Cluster)

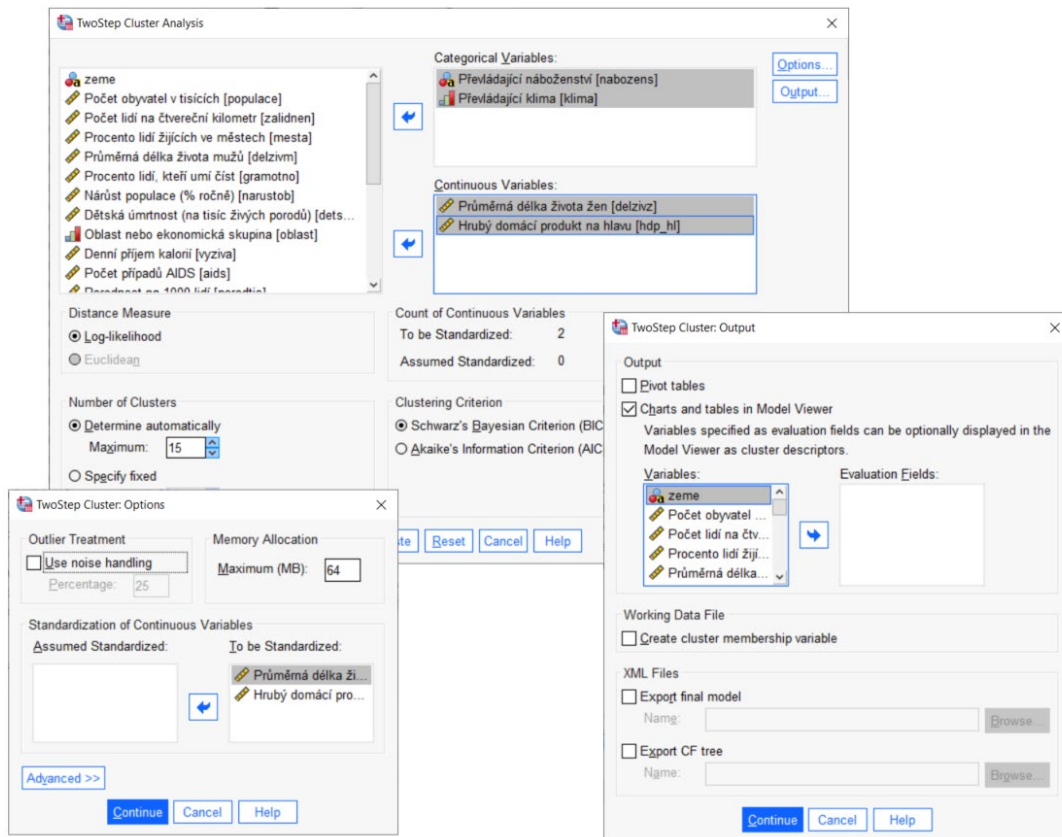
Procedura se užívá k rozdělení souboru (případů) do vnitřně co možná nejvíce homogenních skupin. Počet skupin, do kterých se jednotky rozdělí, můžeme stanovit sami nebo využít doporučení programu. Na rozdíl od ostatních klasifikačních modelů programu IBM SPSS Statistics je metoda TwoStep schopna pracovat 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**). Rovněž můžeme zvolit, jestli má být výstup vypsán ve formě pivotačních tabulek (**Pivot tables**). 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>čí 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>

1.9.2 Seskupování metodou K-means (K-Means Cluster)

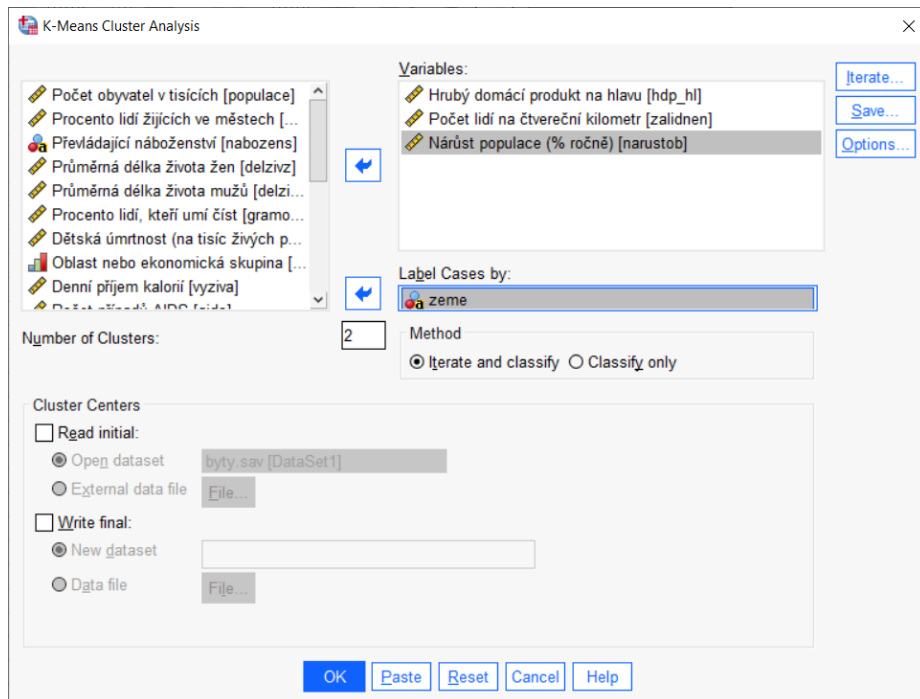
Metoda se užívá k rozdělení souboru (případů) do skupin na základě zvolených číselných charakteristik tak, aby si jednotky uvnitř skupin byly co nejvíce podobné. Jedná se o iterační metodu, která začíná určitým počátečním rozmístěním středů klastrů (souřadnice středů mohou být určeny algoritmem nebo 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). Pokud chceme, aby se středy clusterů aktualizovali průběžně po každém zařazení jednoho případu, zvolíme možnost **Use running means**.

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

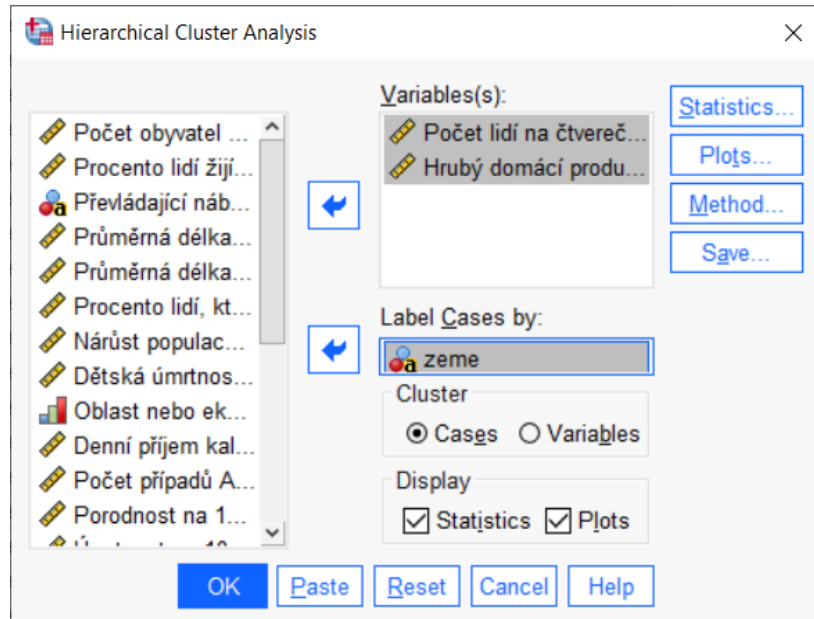
<i>ANOVA</i>	<i>analýza rozptylu</i>
<i>case</i>	<i>případ</i>
<i>classify</i>	<i>klasifikovat</i>
<i>cluster</i>	<i>skupina, seskupovací</i>
<i>continue</i>	<i>pokračovat</i>
<i>file</i>	<i>soubor</i>
<i>final</i>	<i>konečný</i>
<i>label</i>	<i>popis, označení</i>
<i>listwise</i>	<i>celý případ</i>
<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>

1.9.3 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 (dendrogram nebo rampouchový graf). Po stisku tlačítka **Method** lze zadat jednu ze sedmi nabízených metod 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>

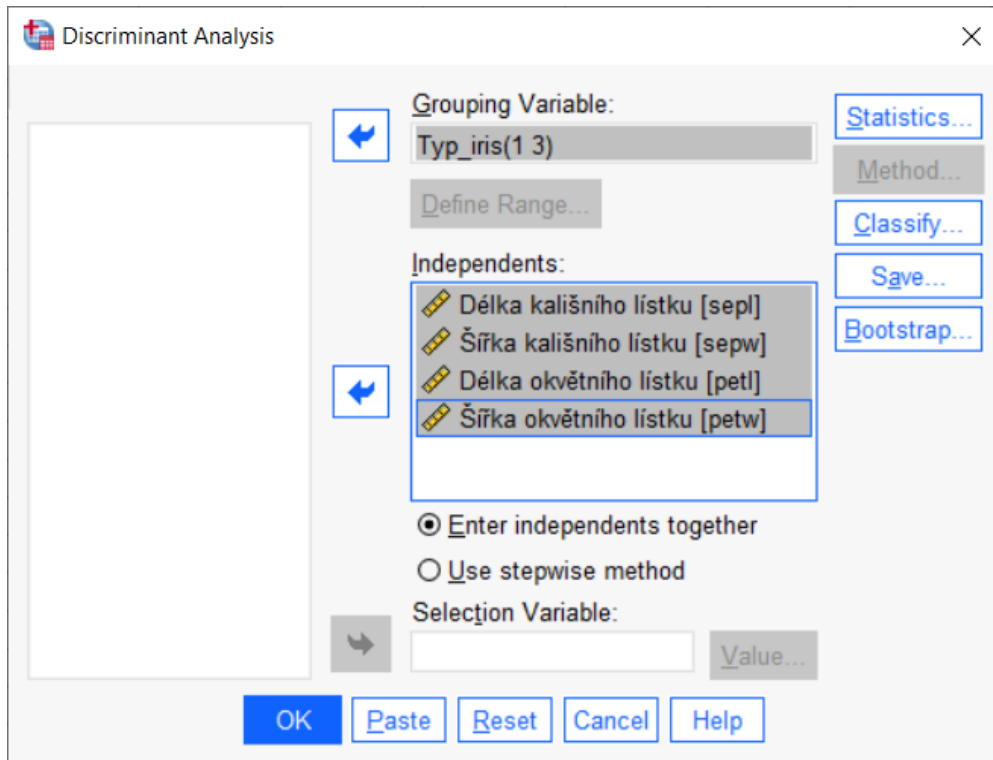
1.9.4 Diskriminační analýza (Discriminant Analysis)

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 sestavení 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 skóre a pravděpodobnost predikované skupiny) do datového souboru jako nové proměnné a exportovat model do formátu XML.

Tlačítkem **Bootstrap** zadáme výpočet odhadů standardních chyb a intervalů spolehlivosti na základě metody *bootstrap* pro koeficienty diskriminačních a klasifikačních funkcí.



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

1.9.5 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.

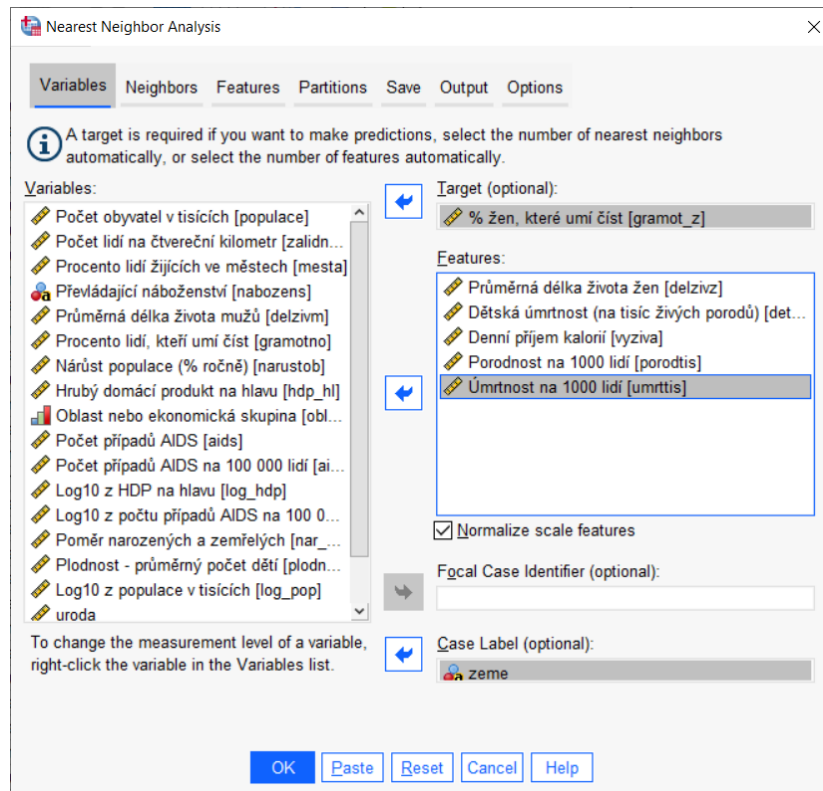
Na záložce **Features** lze zvolit automatický výběr vstupních proměnných a zadat požadavky na způsob výběru. Tyto možnosti jsou k dispozici pouze v případě, že je na záložce **Variables** definována cílová proměnná (**Target**) a alespoň jedna proměnná v poli **Features**. Zaškrtačací políčko **Perform 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

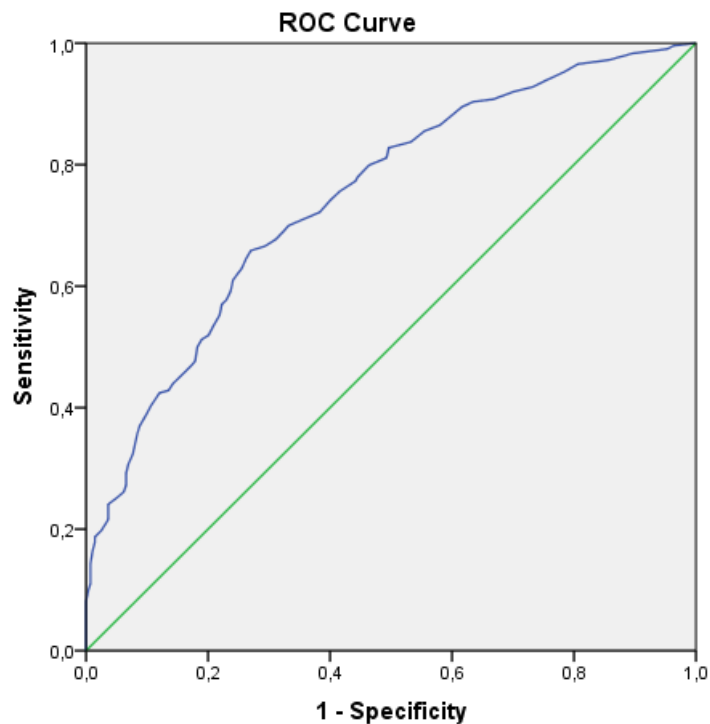
- Analyze
 - Classify
 - Nearest Neighbor

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>

1.9.6 ROC křivka (ROC Curve)

ROC křivka (**Receieve 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.

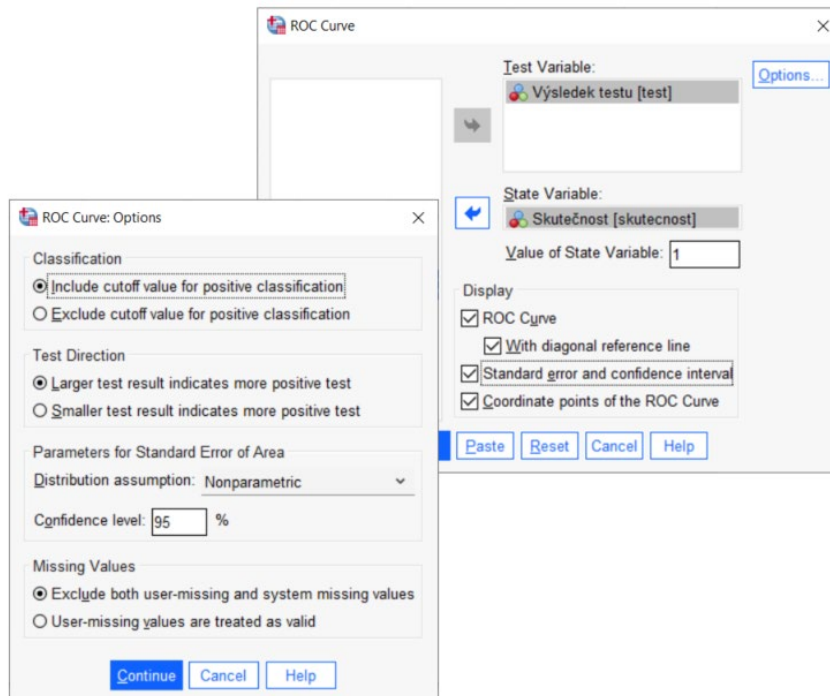
ROC křivka

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 specifičnosti a sensitivity (**Coordinate point of the ROC Curve**).

Tlačítkem **Options** zvolíme, zda bude hraniční hodnota testované proměnné odpovídat pozitivní či negativní klasifikaci (**Classification**). Dále v tomto dialogovém okně 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

1.9.7 ROC analýza (ROC Analysis)

Procedura ROC Analysis nabízí jiný typ analýzy založené na ROC křivce, včetně možnosti porovnání dvou skupin či volby párového designu.

Do pole **Test Variable** přeneseme testovanou číselnou proměnnou, do pole **State Variable** predikovanou dichotomickou proměnnou. V poli **Value of State Variable** zadáme hodnotu dichotomické proměnné (obvykle 1), kterou předpovídáme, pokud bude testovaná číselná proměnná hodnotu nad příslušnou hranicí.

Zaškrtnutí políčko **Paired-sample design** umožňuje zadat párový design.

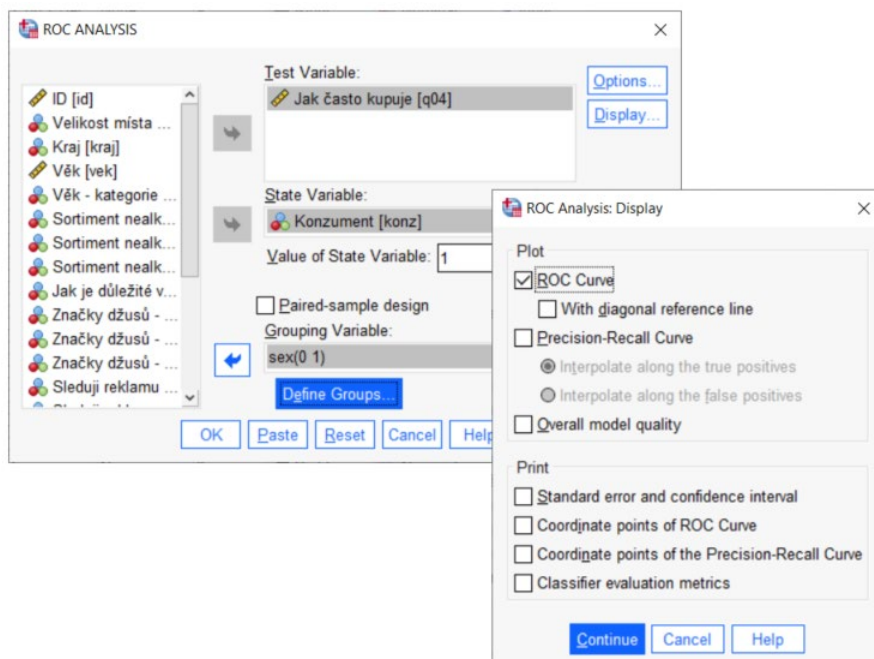
Pro porovnání dvou skupin je určeno pole **Grouping Variable**, do něhož zadáme proměnnou určující skupiny. Pomocí tlačítka **Define Groups** specifikujeme kódy skupin.

Tlačítkem **Options** zadáváme některé doplňující údaje jako například způsob zacházení s vynechanými 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, a určit, zda bude hraniční hodnota testované proměnné odpovídat pozitivní či negativní klasifikaci (**Classification**).

V oblasti **Display** dále určíme požadované grafy (**Plot**) a tabulky (**Print**):

- ROC křivka (**ROC Curve**) a úsečku znázorňující diagonálu (**With diagonal reference line**),
- křivka Precision-Recall a volba způsobu její interpolace (**Precision-Recall Curve**) – křivka proti sobě vynáší dvě charakteristiky – *prediction* a *recall*, kde *prediction* vyjadřuje procento správných pozitivních předpovědí ze všech pozitivních předpovědí a *recall* (odpovídá senzitivitě) vyjadřuje procento správných pozitivních předpovědí ze všech skutečně pozitivních hodnot,
- celková míra kvality modelu (**Overall model quality**) – odpovídá dolní hranici intervalu spolehlivosti pro ROC křivku, která by měla být větší než 0.5,
- standardní chyba a interval spolehlivosti ROC křivky (**Standard error and confidence interval**),
- tabulka obsahující souřadnice zlomových bodů ROC křivky (**Coordinate point of ROC Curve**),
- tabulka obsahující souřadnice zlomových bodů křivky Precision-Recall (**Coordinate point of the Precision-Recall Curve**)
- tabulka obsahující hodnocení úspěšnosti klasifikátoru, např. Giniho index (**Classifier evaluation metrics**).



ROC analýza

ROC analýza

- Analyze

- ROC Analysis

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>