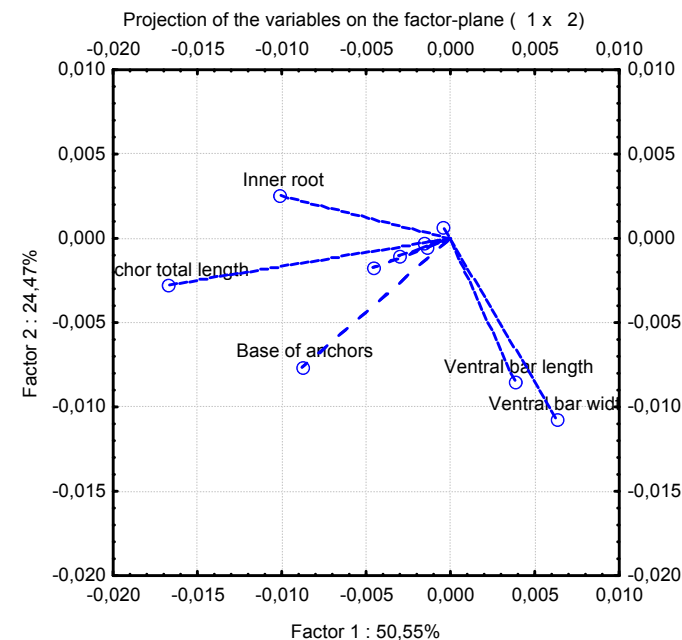


## Vícerozměrné metody

# STATISTIKA

## Vícerozměrné metody

1. Vstupní data pro vícerozměrné analýzy
2. Metriky podobností a vzdáleností
3. Cluster Analysis
4. Principal component analysis
5. Correspondence analysis
6. Canonical analysis
7. Discriminant analysis
8. Factor analysis
9. Multidimensional scaling

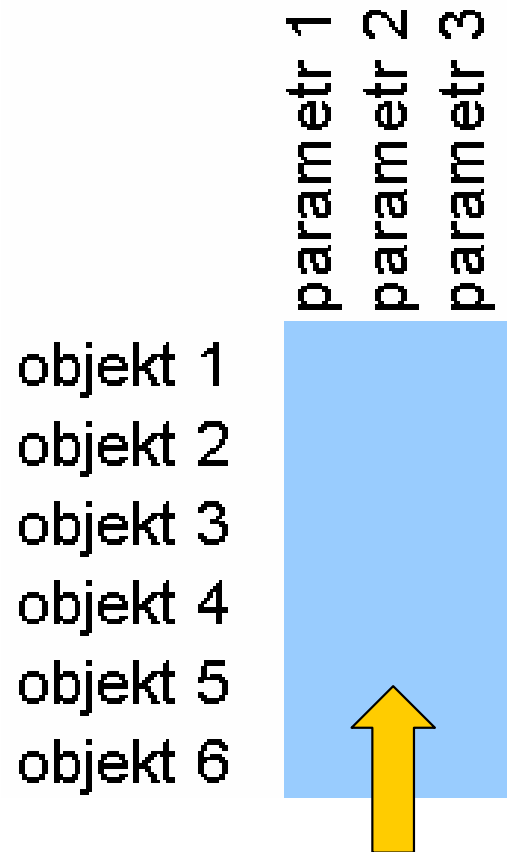


# Úvod do vícerozměrných metod I.

- ◆ **Vícerozměrné metody:** Název vícerozměrné vychází z typu vstupních dat, tato data jsou tvořena jednotlivými objekty (i.e. klienti) a každý z nich je charakterizován svými parametry (věk, příjem atd.) a každý z těchto parametrů můžeme považovat za jeden rozměr objektu.
- ◆ **Maticová algebra:** Základem práce s daty a výpočtů vícerozměrných metod je maticová algebra, matice tvoří jak vstupní, tak výstupní data a probíhají na nich výpočty.
- ◆  **$N \times P$  matice:**  $N$  objektů s  $p$  parametry pak vytváří tzv.  $N \times P$  matici, která je prvním typem vstupu dat do vícerozměrných analýz.
- ◆ **Asociační matice:** Na základě těchto matic jsou počítány matice asociační na nichž pak probíhají další výpočty, jde o čtvercové matice obsahující informace o podobnosti nebo rozdílnosti (tzv. **metriky**) buď objektů (Q mode analýza) nebo parametrů (R mode analýza). Měřítko podobnosti se liší podle použité metody a typu dat, některé metody umožňují použití uživatelských metrik.

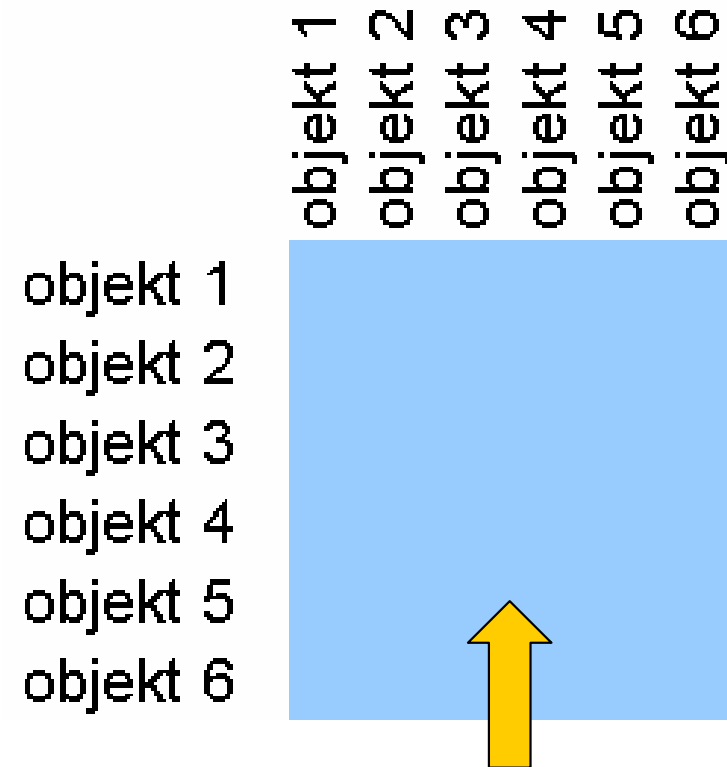
# Vstupní matice vícerozměrných analýz

## NxP MATICE



Hodnoty parametrů pro jednotlivé objekty

## ASOCIAČNÍ MATICE



Korelace, kovariance, vzdálenost, podobnost

## Úvod do vícerozměrných metod II.

### SHLUKOVÁ ANALÝZA

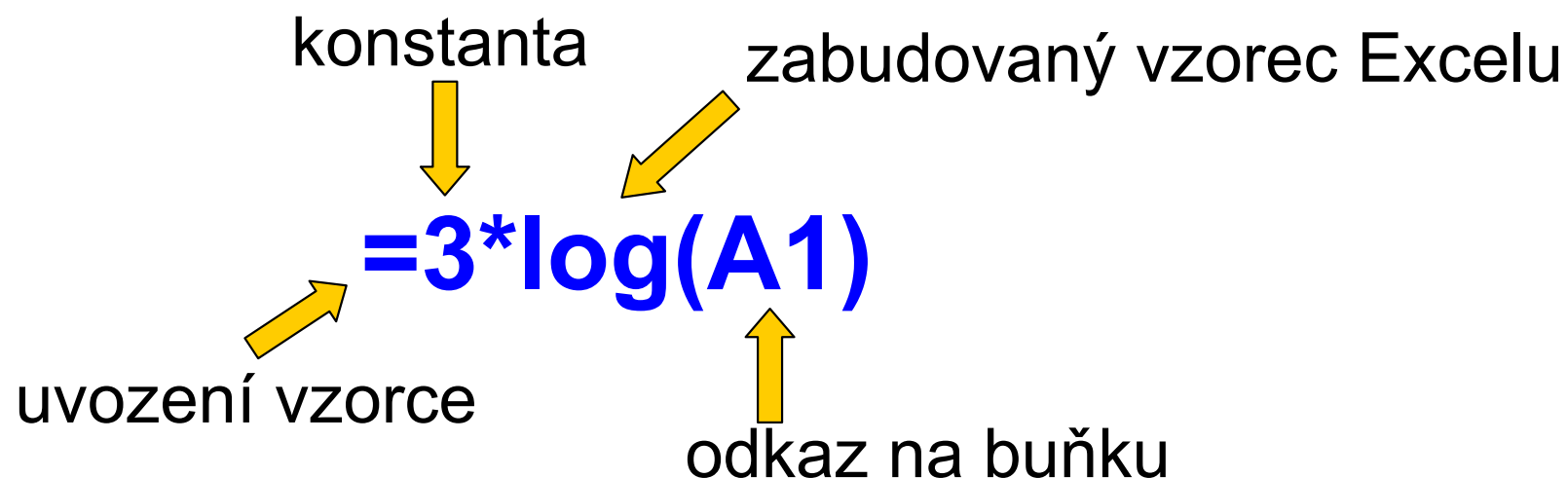
- ◆ vytváření shluků objektů na základě jejich podobnosti
- ◆ identifikace typů objektů

### ORDINAČNÍ METODY

- ◆ zjednodušení vícerozměrného problému do menšího počtu rozměrů
- ◆ principem je tvorba nových rozměrů, které lépe vyčerpávají variabilitu dat

# Vzorce v Excelu

- vpisují se do buněk sešitu
- vzorce jsou vždy uvozeny = (lze též + -)
- aritmetické operátory + zabudované funkce Excelu
- pro „sčítání“ nečíselných položek se používá &
- výpočet je založen buď na číselných konstantách nebo odkazech na buňky



# Vzorce v Excelu – odkazy na buňku – styl A1

## Relativní odkazy

- **A1** = buňka 1. řádku sloupci A
- **A1:B6** = blok buněk – levý horní roh je v 1. řádku, sloupec A, pravý dolní na řádku 6, sloupec B
- relativní odkaz se při automatickém vyplnění buněk vzorcem posune

**Absolutní odkaz** – odkaz na buňku je pevně dán, při kopírování nebo automatickém vyplnění se nemění, lze uzamknout jak řádky, tak sloupce samostatně

uzamčení sloupce → **\$A\$1** ← uzamčení řádku

# Maticové vzorce v Excelu

- výpočty z matic dat
- zadávání je ukončeno stiskem CTRL+SHIFT+ENTER

Vzorec je založen na těchto dvou maticích dat

16			
17	10	2	
18	12	3	
19	5	4	
20	8	5	
21	4	8	
22	7	9	
23	9	11	
24	suma součinů řádků	310	
25			

**{=SUMA(A17:A23\*B17:B23)}**

Násobení řádků matic

Celkové sečtení

**Nezbytné pro operace s maticemi.**



# Měření vzdálenosti objektů

Euklidovská vzdálenost

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Vážená euklidovská vzdálenost

$$d_{ij} = \sqrt{\sum_{k=1}^p w_k^2 (x_{ik} - x_{jk})^2}$$

$i, j$  – označení objektů

$d_{ij}$  – vzdálenost objektů  $i$  a  $j$

$p$  – počet parametrů

$k$  –  $k$ -tý parametr

$w_k$  – váha parametru  $k$

Minkowski (power distance)

$$d_{ij} = \sqrt[\lambda]{\sum_{k=1}^p |x_{ik} - x_{jk}|^\lambda}$$

- - celé číslo
- = 1 Manhattan (city block)
- = 2 Euklidovská vzdálenost

Chebychev

$$d_{ij} = \max |x_{ik} - x_{jk}|$$

# Měření podobnosti objektů

## Binární koeficienty podobnosti

		Objekt 1	
		1	0
Objekt 2	1	a	b
	0	c	d

a, b, c, d = počet případů, kdy souhlasí binární charakteristika objektu 1 a 2  
 $a+b+c+d=p$

Symetrické binární koeficienty - není rozdíl mezi případem 1-1 a 0-0

Simple matching coefficient

$$S(x_1, x_2) = \frac{a + d}{p}$$

Hamman, Yule coefficient, Pearson's  $\chi^2$  (phi) a další koeficienty

## Asymetrické binární koeficienty – odstranění double zero

Jaccard`s coefficient

$$S(x_1x_2) = \frac{a}{a+b+c}$$

Sorensen`s coefficient

$$S(x_1x_2) = \frac{2a}{2a+b+c}$$

Řada dalších koeficientů dávajících různou váhu jednotlivým kombinacím parametrů

## Kvantitativní koeficienty

Obdoby binárních koeficientů pro více parametrů než 0/1

Simple matching coefficient pro více parametrů

$$S(x_1x_2) = \frac{\textit{souhlas}}{p} \quad p = \text{počet parametrů}$$

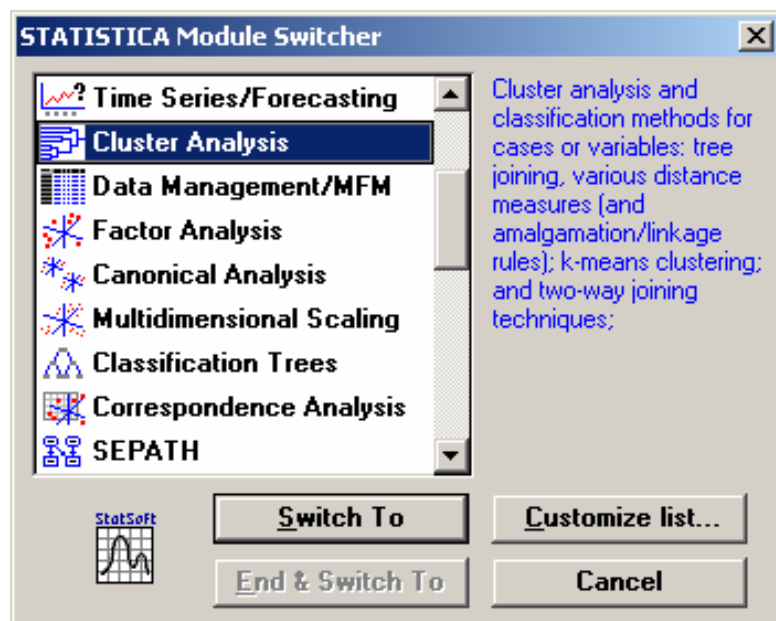
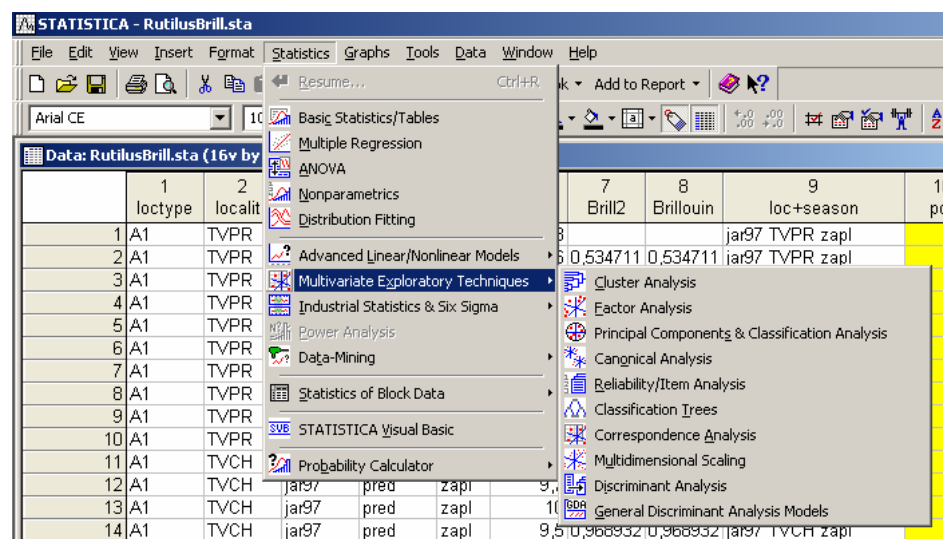
## Gowerův koeficient

Zahrnutí podobnosti podle různých typů parametrů – binární, kvalitativní a semikvantitativní i kvantitativní (odlišný výpočet pro jednotlivé typy). Celkový součet podobností je podělen počtem parametrů. Může zahrnovat podmínku nepočítat s chybějícími parametry – Kronecker`s delta.

Více informací a další měření vzdáleností a podobností najdete v knize **LEGENDRE, P. & LEGENDRE, L. (1998). *Numerical ecology*. Elseviere Science BV, Amsterodam.**

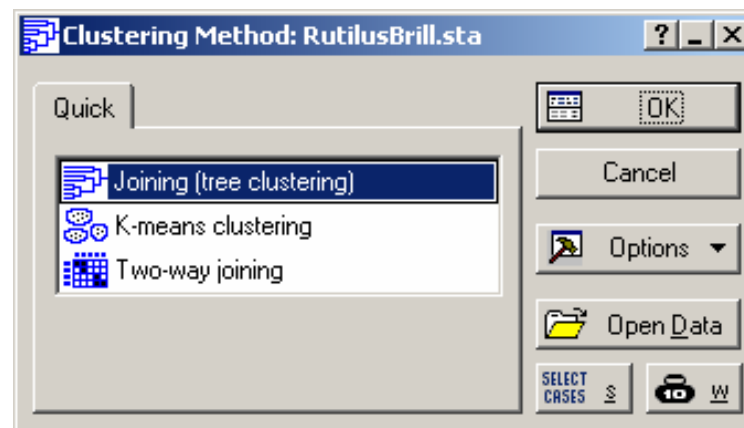
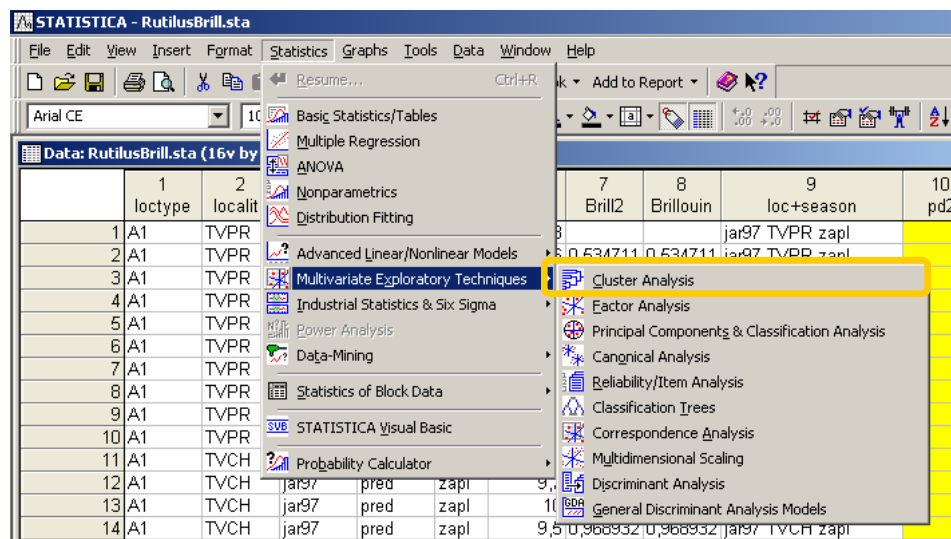
# Vícerozměrné metody v *STATISTICA*

**Statistica 6** – nabídková větev *Multivariate Exploratory Techniques* v menu *Statistics*



**Statistica 5.5** – několik samostatných modulů volitelných ze základní nabídky (*Cluster Analysis*, *Factor Analysis*, *Canonical Analysis*, *Multidimensional Scaling*, *Correspondence Analysis* a jiné)

## Statistics >> Multivariate Exploratory Techniques >> Cluster Analysis



- **Joining (tree clustering)** – hierarchické shlukování, podle vzdálenosti mezi objekty jsou tyto skládány do skupin pomocí různých algoritmů.
- **K – means clustering** (hypotéza existence  $x$  clusterů a její ověření analogické k ANOVA – sestavení clusterů tak aby se minimalizovala jejich vnitřní variabilita a maximalizovala variabilita mezi clustery), nehierarchické shlukování
- **Two-way joining** (shlukování je prováděno zároveň na základě jak objektů, tak parametrů)

# “Klasická“ shluková analýza hierarchicky spojující objekty do skupin podle vzdálenosti v asociační matici

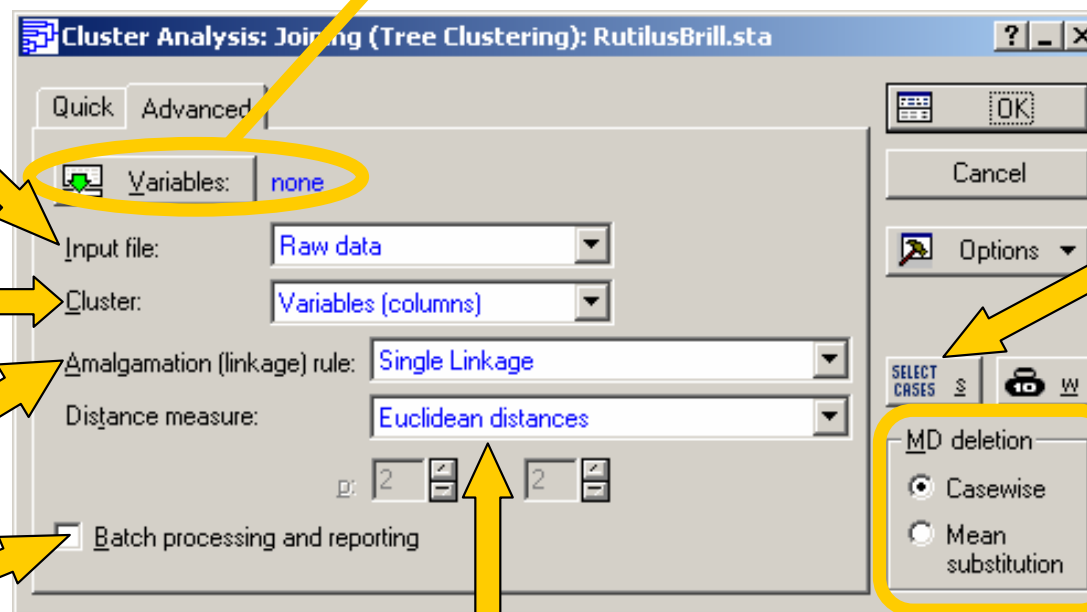
Vstupní soubor je matice objekty  $x$  parametry nebo matice vzdáleností

Mají být shlukovány sloupce nebo řádky vstupní matice objekty  $x$  parametry?

Shlukovací algoritmus

Automatizovaný výstup

Vybrání proměnných pro výpočet

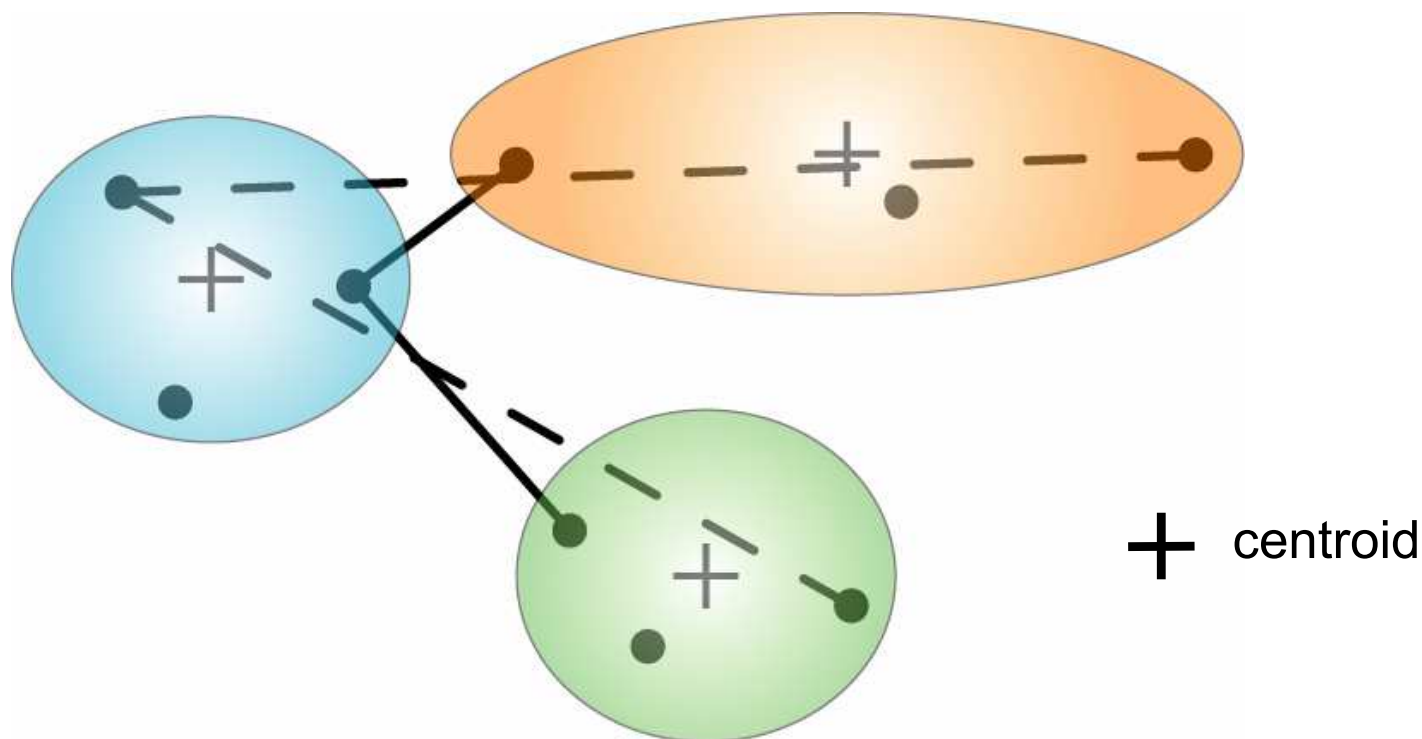


Výběr z dat

Použitá vzdálenost mezi objekty (jen matice objekty  $x$  parametry)

Smazání chybějících dat nebo jejich nahrazení průměrem

## Joining (Tree Clustering) – shlukovací algoritmy



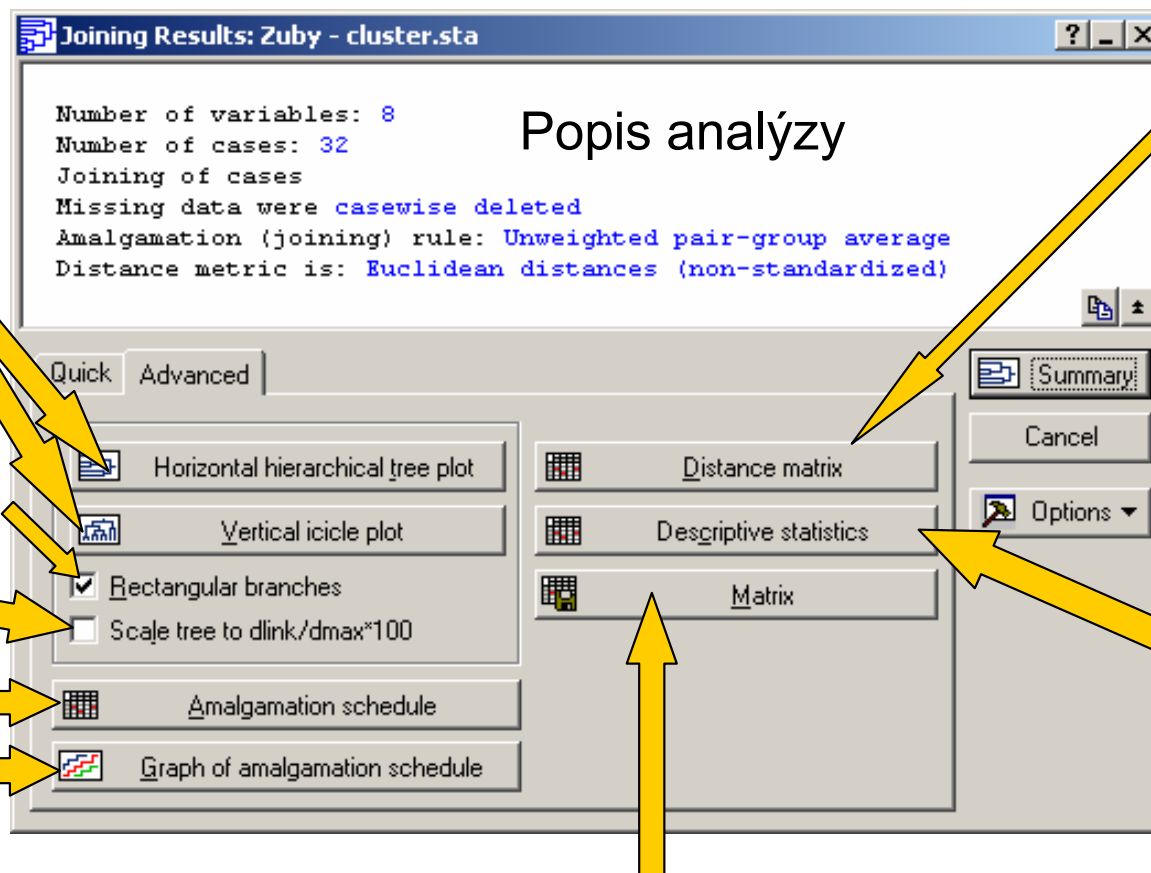
- Na tuto vzdálenost se ptá **single linkage**
- - Na tuto vzdálenost se ptá **complete linkage**

Další metody počítají s **průměrnou vzdáleností** všech objektů shluků nebo vzdáleností **centroidů** (vzdálenost může být **vážena** velikostí shluků).

**Wardova metoda** se snaží minimalizovat variabilitu uvnitř shluků.



Výsledky programu Statistica se typicky dělí na záložky **Quick** (nejdůležitější výstupy) a **Advanced** (podrobnější analýza, nastavení vlastností výstupů)



Horizontální a vertikální dendrogram

Pravoúhlé větve stromu

Vzdálenost v %

Postup skládání stromu v podobě tabulky a grafu

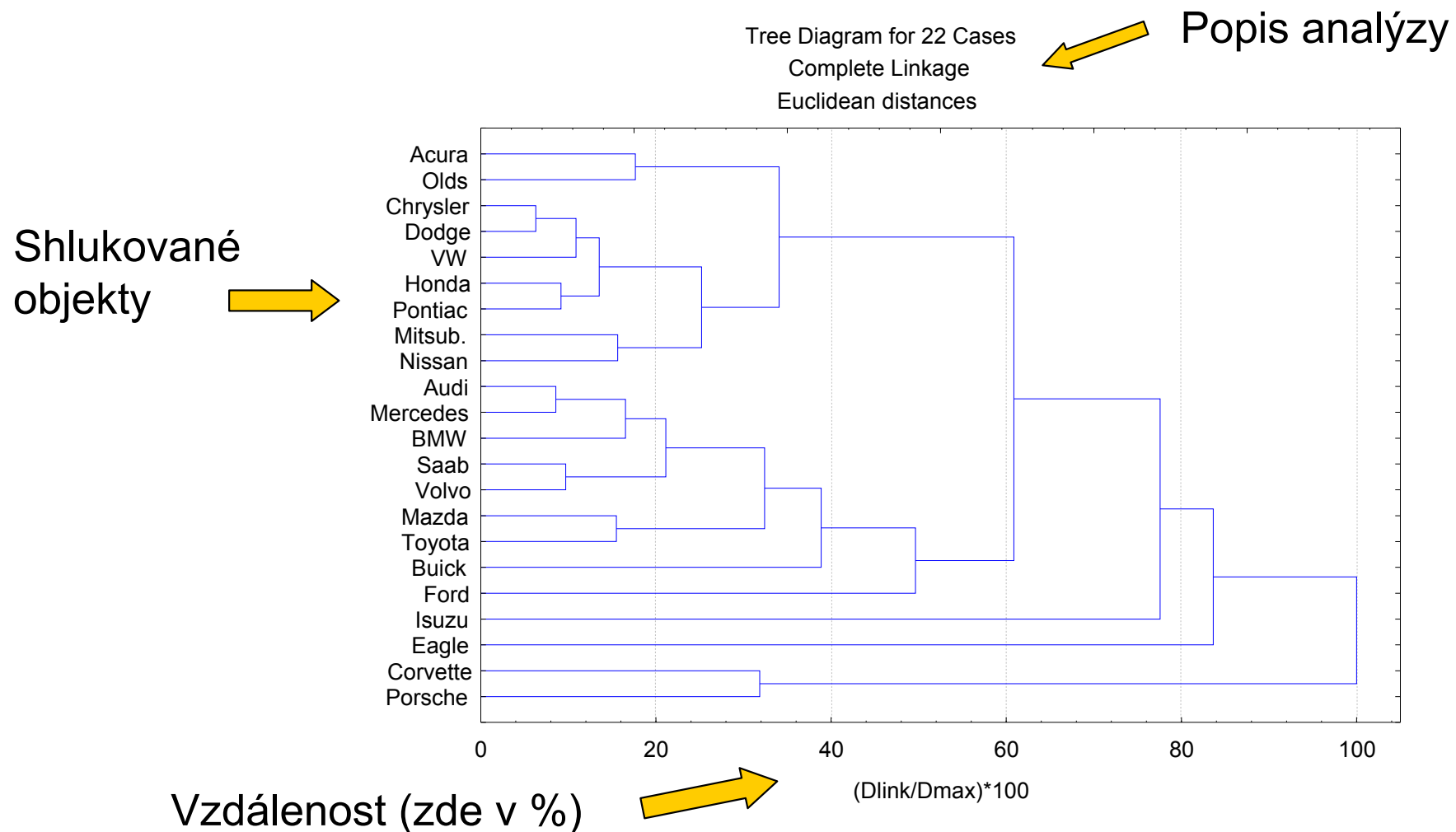
Popis analýzy

Matrice vzdáleností

Popis objektů (průměr a SD)

Export matice vzdáleností (podle zvolené metriky) do speciálního souboru Statistica pro matice vzdáleností

**Dendrogram** představuje grafický výstup shlukové analýzy, kde jsou objekty propojeny tak, jak postupovalo jejich shlukování



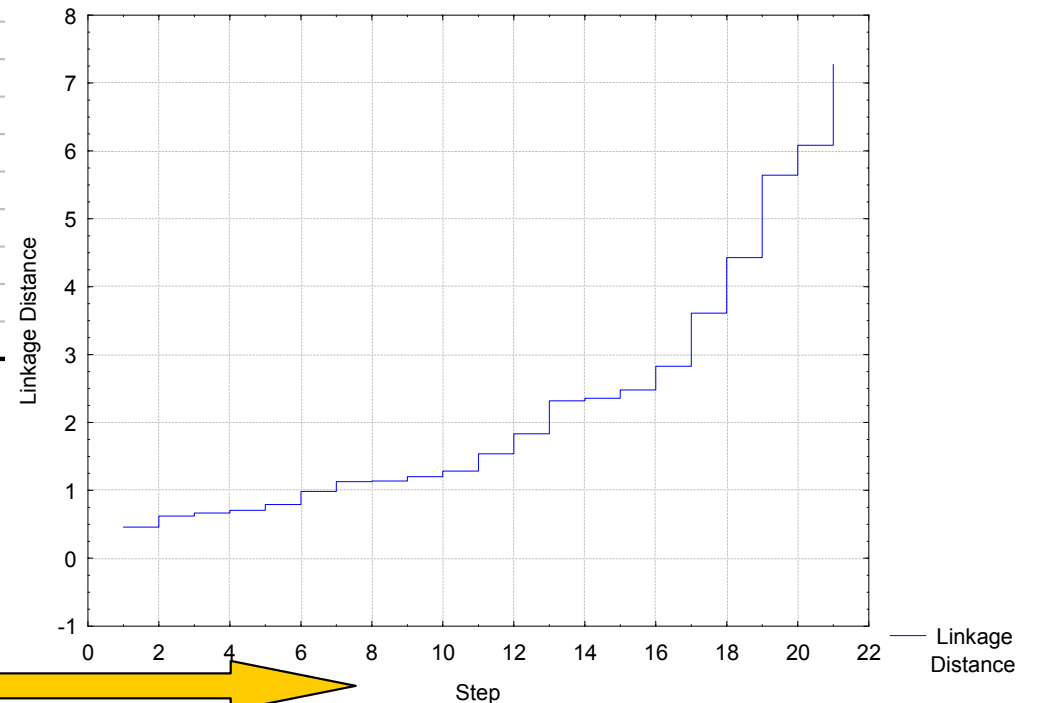
**Amalgamation schedule a graf** poskytují uživateli přehled nad celým procesem shlukování, tj. při jaké vzdálenosti a jaké objekty nebo jejich skupiny se shlukly

Vzdálenost na níž došlo k shlukování

Amalgamation Schedule (Cars.sta)					
Complete Linkage					
Euclidean distances					
Linkage distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5
,4580483	Chrysler	Dodge			
,6231085	Audi	Mercedes			
,6670490	Honda	Pontiac			
,7060042	Saab	Volvo			
,81339	Chrysler	Dodge	VW		
,98471	Chrysler	Dodge	VW	Honda	
1,127473		Toyota			
1,137488	Mitsub.	Nissan			
1,202407	Audi	Mercedes	BMW		
1,284603	Acura	Olds			
1,537968	Audi	Mercedes	BMW	Saab	
1,834401	Chrysler	Dodge	VW	Honda	

Shlukované objekty

Plot of Linkage Distances across Steps  
Euclidean distances



Kroky shlukování

## Joining (Tree Clustering) – asociační matice

**Asociační matice** představují speciální typ souborů programu Statistica (přípona .smx), jde o čtvercové matice nesoucí informaci o vztazích mezi řádky a sloupci, tvoří alternativní vstup pro vícerozměrné analýzy, některé analýzy lze provádět pouze na datech v tomto formátu. Na rozdíl od běžných souborů obsahují 4 speciální řádky, pro správnou funkci je nezbytné dodržet jejich přesnou syntaxi.

### Vlastní matice vzdáleností

Průměr a SD proměnných  
(není nutné pro matici  
podobností a nepodobností)

Počet případů = počet z něž  
byla matice vytvořena, ne počet  
jejích řádků

	Var 1	Var 2	Var 3
Var 1	1.00	.20	.30
Var 2	.20	1.00	.10
Var 3	.30	.10	1.00
Means	12	11	10
Std. Dev.	3	5	2
No. Cases	50		
Matrix	1		

Typ matice 1 = korelace, 2 = podobnosti, 3 = nepodobnosti, 4 = kovariance

# Shluková analýza K-means clustering

**K-means clustering** se snaží rozdělit objekty do zadaného počtu shluků tak, aby byla minimalizována variabilita uvnitř shluků a maximalizována mezi shluky

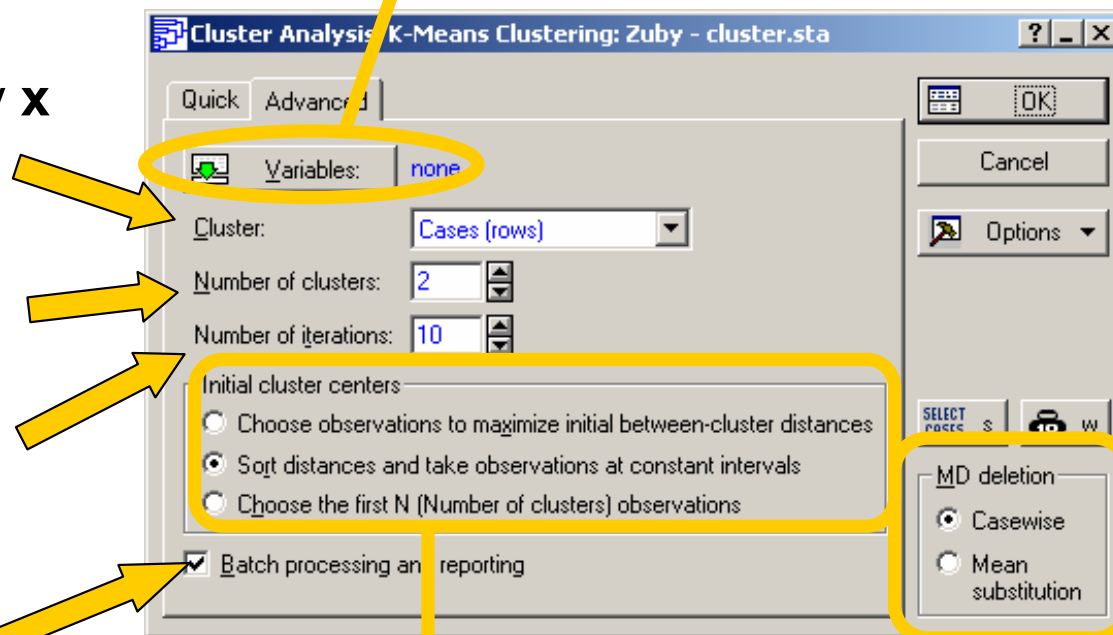
Vybrání proměnných pro výpočet

Mají být shlukovány sloupce nebo řádky vstupní matice objekty  $x$  parametry?

Počet očekávaných shluků

Počet iterací – kroků výpočtu

Automatizovaný výstup



Nastavení počátečních shluků, od nichž se výpočet odvíjí

Smazání chybějících dat nebo jejich nahrazení průměrem

## K-means clustering - výsledky

**K-means clustering** pracuje s objekty pouze na základě Euklidovské vzdálenosti, na tuto skutečnost je nezbytné pamatovat pokud tato metrika není pro data vhodná.

Euklidovská vzdálenost středu shluků

ANOVA pro jednotlivé proměnné

Graf průměrů jednotlivých proměnných v shlucích

Průměr, rozptyl, SD parametrů v shlucích

Objekty v shlucích a jejich vzdálenost od centroidu

Uloží příslušnost k shluku doplněnou o vzdálenost k centroidu pro všechny objekty (+ vybrané parametry).

# K-means clustering – tabulky výsledků

ANOVA jednotlivých parametrů rozdělených podle shluků

Variable	Cluster Means (O2_Statistica_Cluster.sta)			
	Cluster No. 1	Cluster No. 2	Cluster No. 3	Cluster No. 4
Var1	2,000000	1,285714	2,933333	0,000000
Var2	2,833333			
Var3	1,000000			
Var4	0,833333			
Var5	2,000000			
Var6	2,500000			
Var7	3,000000			
Var8	3,000000			

Euclidean Distances between Clusters (O2\_Statistica\_Cluster.sta)  
Distances below diagonal  
Squared distances above diagonal

Cluster Number	No. 1	No. 2	No. 3	No. 4
No. 1	0,000000	0,971825	1,033266	0,971825
No. 2	0,971825	0,000000	0,971825	0,971825
No. 3	1,033266	0,971825	0,000000	0,971825
No. 4	0,971825	0,971825	0,971825	0,000000

Analysis of Variance (O2_Statistica_Cluster.sta)						
Variable	Between SS	df	Within SS	df	F	signif. p
Var1	32,60685					
Var2	25,53542					
Var3	5,46875					
Var4	6,66667					
Var5	24,15446					
Var6	24,91786					
Var7	22,13542					
Var8	11,47500					

Descriptive Statistics for Cluster 3  
Cluster contains 15 cases

Variable	Mean	Standard Deviation	Variance
Var1	2,933333	0,258199	0,066667
Var2	2,600000	0,632455	0,400000
Var3	1,000000	0,000000	0,000000
Var4	1,000000	0,000000	0,000000
Var5	3,600000	0,507092	0,257143
Var6	3,400000	0,736789	0,542857
Var7	1,333333	0,000000	0,000000
Var8	1,800000	0,000000	0,000000

Members of Cluster Number 4 (O2\_Statistica\_Cluster.sta) and Distances from Respective Cluster Center  
Cluster contains 4 cases

	reindeer	elk	deer	moose
Distance	0,176777	0,176777	0,176777	0,176777

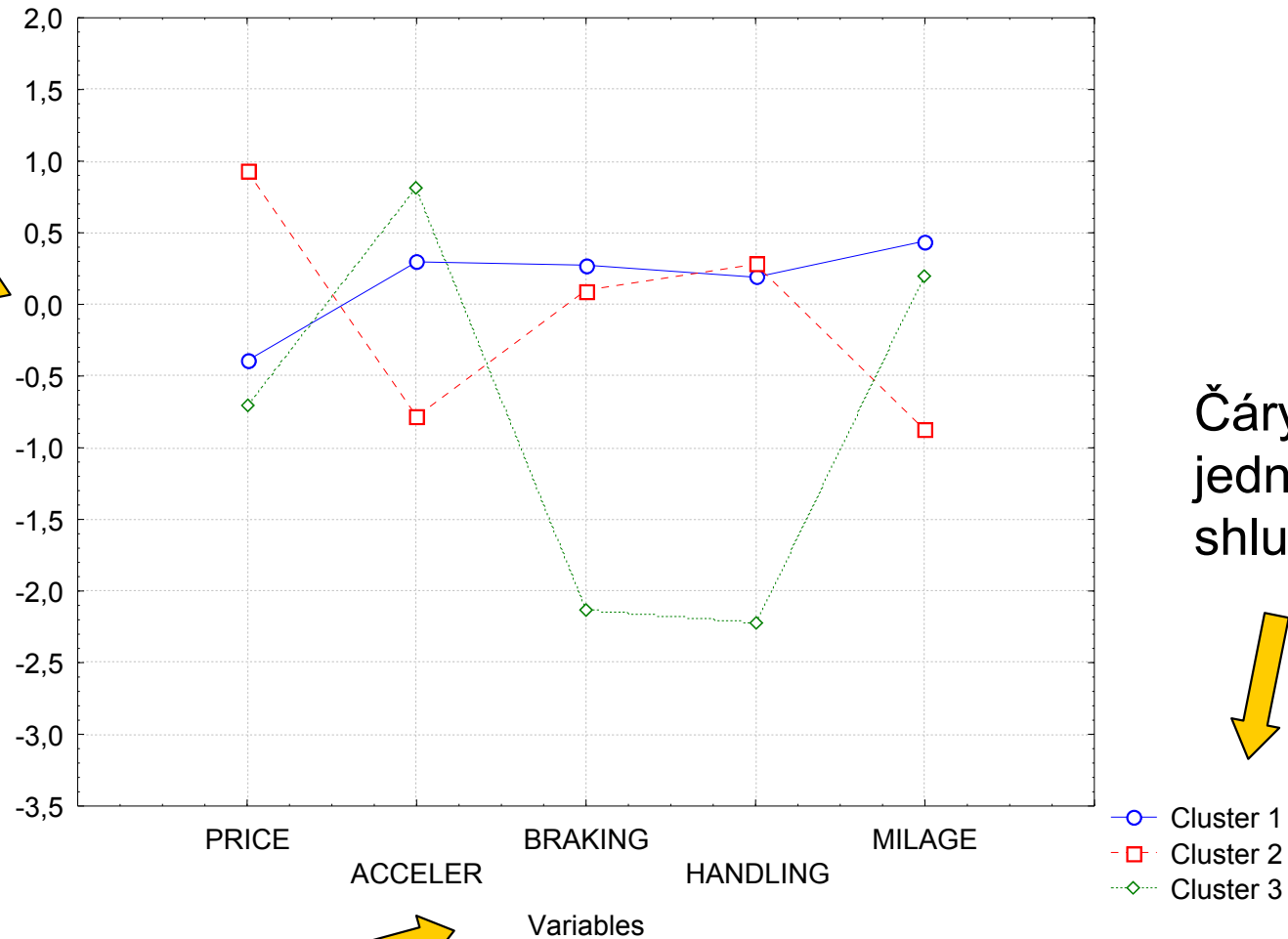
Střed a vzdálenosti středů shluků

Popisná statistika shluků

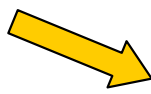
Členové shluku a jejich vzdálenost od středu shluku

# K-means clustering – průměry parametrů

Plot of Means for Each Cluster



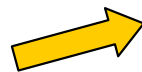
Průměry v  
shlucích



Čáry pro  
jednotlivé  
shluky



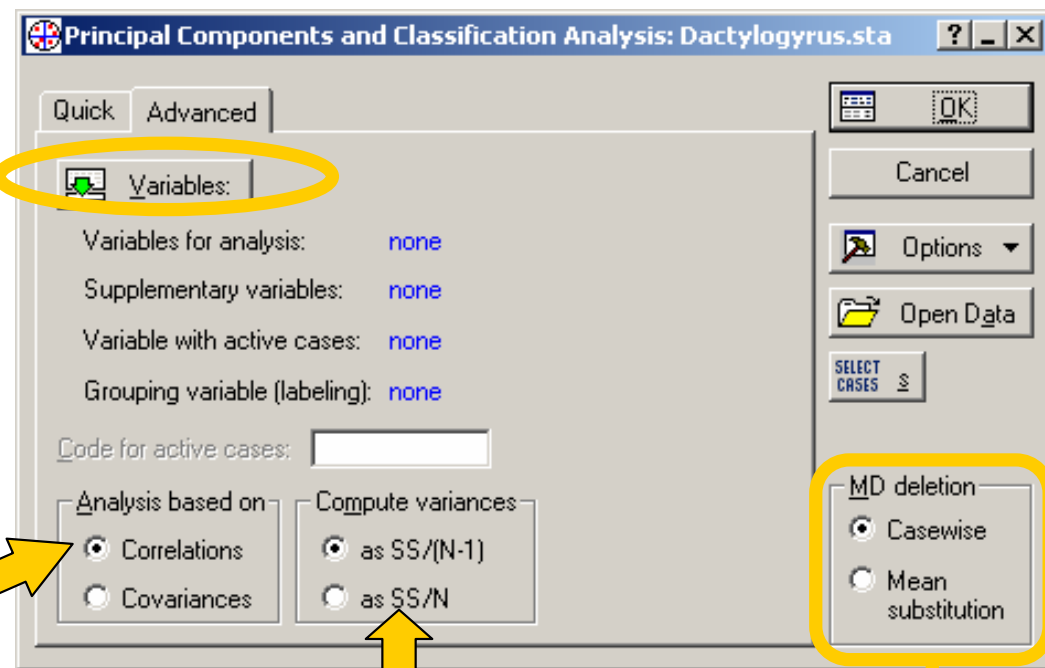
Jednotlivé parametry





## Statistics >> Multivariate Exploratory Techniques >> Principal components ...

1. **proměnné pro výpočet**
2. **supplementary variables**  
nejsou použity pro výpočet, ale objeví se na výsledku
3. **active cases** – vybrání cases (řádků), které se použijí pro výpočet, ostatní se mohou pouze zobrazit
4. **grouping variables** – pro označení skupin objektů



Analýza je založena na matici korelací (standardizace proměnných) nebo kovariancí (vliv rozdílných rozptylů)

Pro výpočet rozptylu se používá  $n$  nebo  $n-1$ .

Smazání chybějících dat nebo jejich nahrazení průměrem

# Principal component analysis – výsledky quick

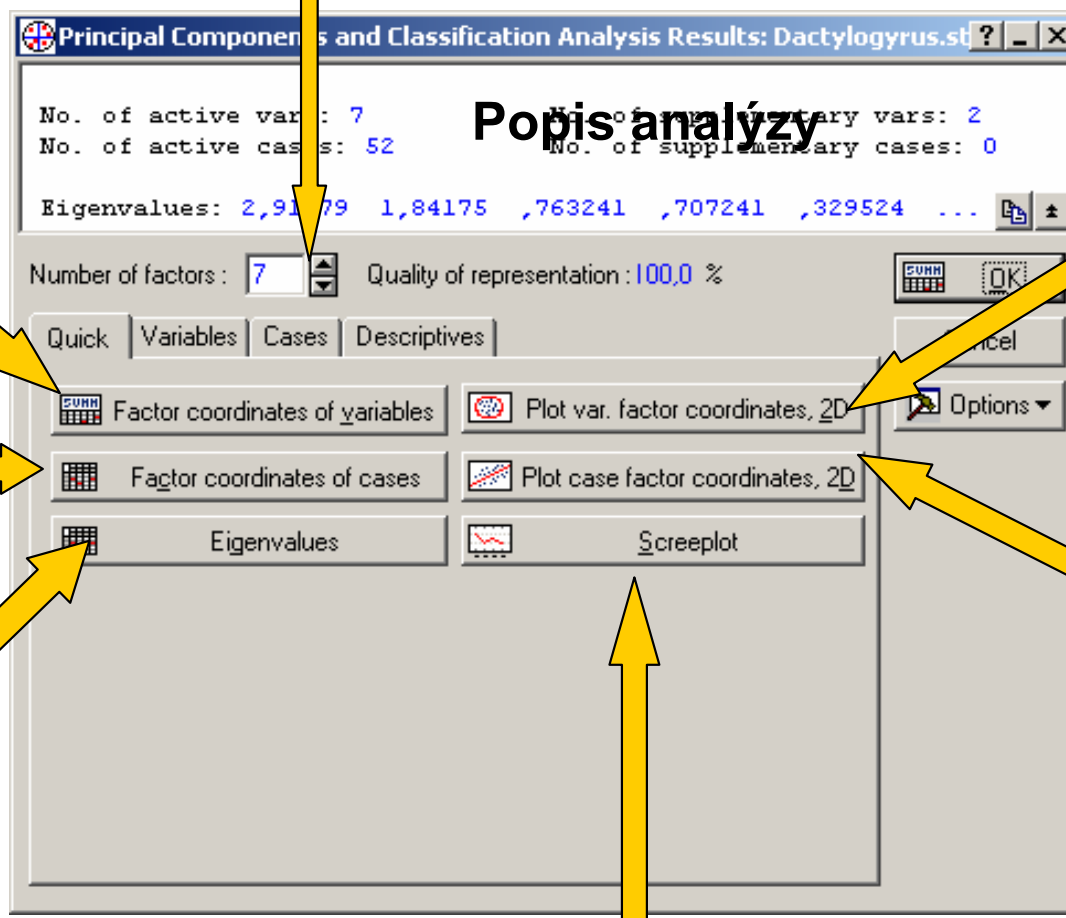
Počet faktorů

Koordináty parametrů na faktorových osách

Koordináty objektů na faktorových osách

**Eigenvalues**  
~ variabilita vyčerpaná faktorovými osami, jejich součet = počet parametrů

Grafické znázornění eigenvalues



2D graf parametrů vzhledem k faktorovým osám

2D graf objektů vzhledem k faktorovým osám

## Factor coordinates of variables = korelace

parametry

Variable	Factor coordinates of the variables, based on correlations (04_PCA.sta)						
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
druh1	0,870988	0,088622	0,311857	-0,045525	-0,074689	0,210660	-0,290247
druh2	0,189727	0,471241	-0,317073	-0,766312	0,179947	0,129830	0,070210
druh3	0,868428	-0,055174	0,381471	0,109716	-0,006402	0,073756	0,311478
druh4	0,308144	-0,120995	-0,808490	0,272021	-0,311633	0,253811	0,022504
druh5	0,570394	0,603301	-0,277173	0,207165	-0,006431	-0,435705	-0,032504
druh6	-0,383706	0,576749	0,297602	-0,069102	-0,649814	0,052953	0,041551
druh7	-0,297247	0,688877	0,058910	0,463059	0,380940	0,271819	0,014072

Faktorové osy

Pozice parametrů  
na faktorových  
osách

## Factor coordinates of cases

objekty

Case	Factor coordinates of cases, based on correlations (04_PCA.sta)							Sezona
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	
HEL	-0,61623	-1,33188	-0,47377	-0,04405	0,03909	0,14791	-0,14760	autumn 97
HVE	-0,50161	-1,09014	-0,01107	0,02535	0,27617	-0,25525	-0,06605	autumn 97
MEL	-0,74946	-0,82916	0,43497	-0,66209	0,17579	-0,65651	-0,47288	autumn 97
ROH	-1,53568	3,32389	0,72130	-0,52632	-1,82878	0,21582	0,10822	autumn 97
STR	-0,96700	2,57010	0,32726	1,51434	0,37240	-0,54825	-0,30062	autumn 97
HEL	-0,68288	-1,24455	0,43192	-0,27091	0,50764	-0,35601	-0,16454	spring 97
HVE	1,64883	0,78839	-2,36792	1,77638	0,02305	-1,36117	0,28317	spring 97
MEL	-0,46771	-0,94204	0,00428	0,43506	0,81362	0,21496	0,06124	spring 97
ROH	4,59087	0,67263	1,33177	-0,00030	-0,21815	0,16225	-0,44771	spring 97
STR	-0,60010	-0,84114	-0,57108	0,15838	-0,24892	0,49180	-0,09555	spring 97
HEL	-0,33993	-1,14800	0,65528	-0,57325	0,27923	-0,21073	0,07271	spring 98
ROH	-1,01963	-0,42177	1,06891	-0,45581	-0,71097	-0,17656	-0,57878	spring 98
MNV	0,14927	0,39321	-1,07576	0,83232	0,75186	-0,85026	-0,32807	autumn 97
PLA	0,25085	0,64376	0,37760	0,07708	0,31431	0,38923	0,01011	autumn 97

Faktorové osy

Příslušnost objektů do  
skupin

Pozice parametrů  
na faktorových  
osách

**Eigenvalues** vyjadřují variabilitu vyčerpanou faktorovými osami, jejich hodnoty slouží při rozhodnutí kolik faktorových os je pro nás zajímavých

Eigenvalue

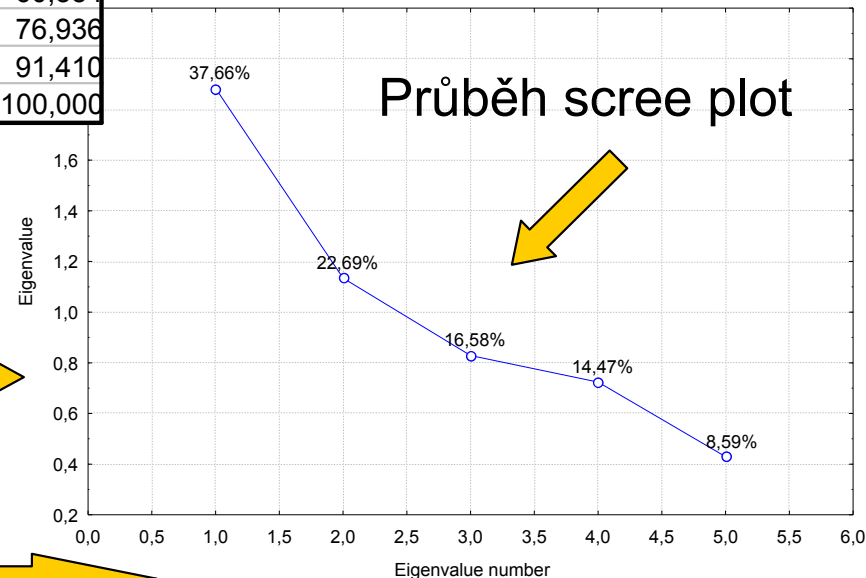
variabilita vyčerpaná příslušnou osou

Kumulativní eigenvalue/vyčerpaná variabilita

Eigenvalues of correlation matrix, and associated statistics for Active variables only

Value number	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	1,883153	37,66307	1,883153	37,663
2	1,134548	22,69096	3,017701	60,354
3	0,829119	16,58238	3,846820	76,936
4	0,723700	14,47401	4,570521	91,410
5	0,429479	8,58959	5,000000	100,000

Eigenvalues of correlation matrix  
Active variables only



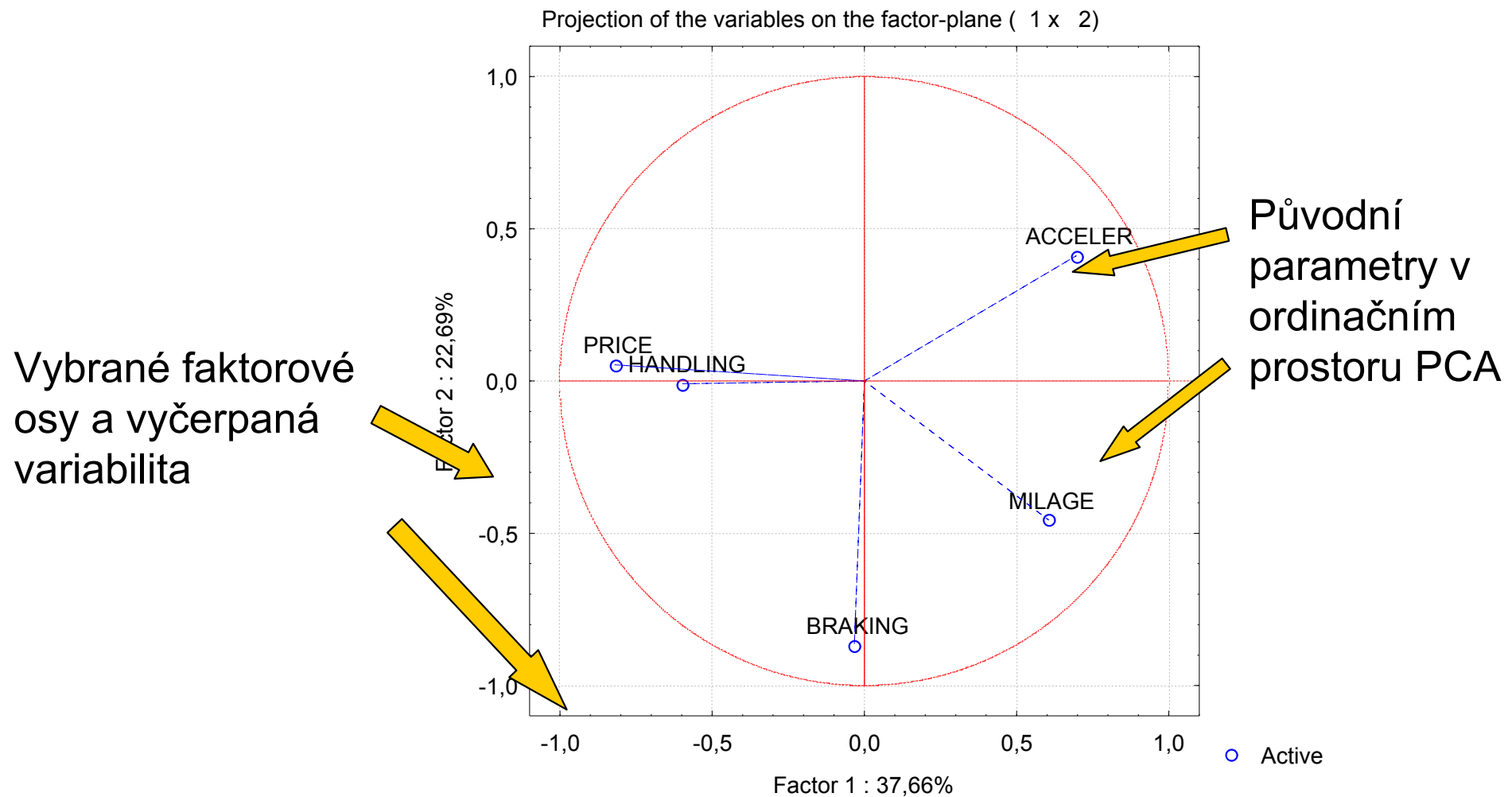
Eigenvalue



Principal component vytvořená PCA

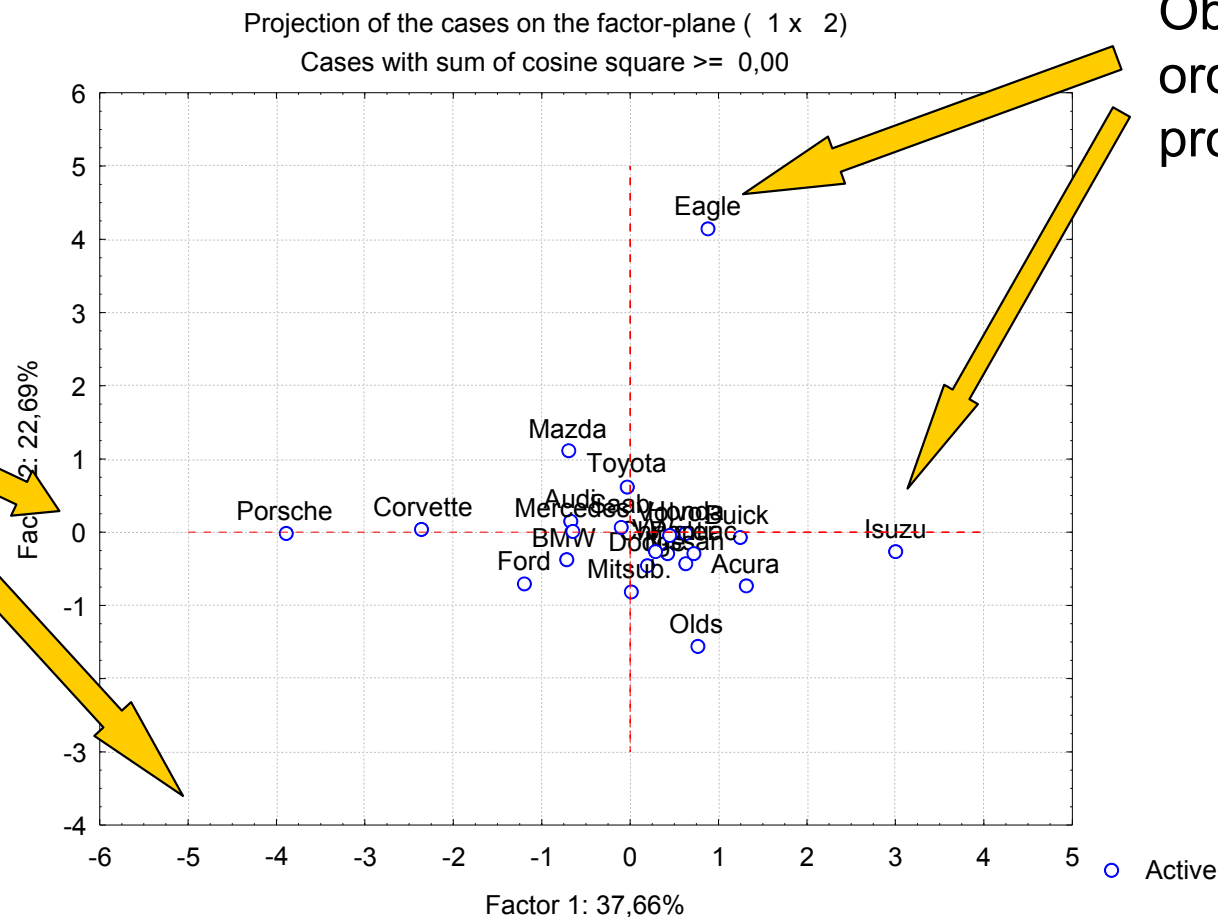


**Plot variables factor coordinates** – vynáší do prostoru faktorových os původní parametry, zobrazuje jejich korelaci s faktorovými osami



**Plot variables cases coordinates** Výpočet je založen na původní NxP matici a matici eigenvektorů, zobrazuje vzájemné vzdálenosti objektů

Vybrané faktorové osy a vyčerpaná variabilita



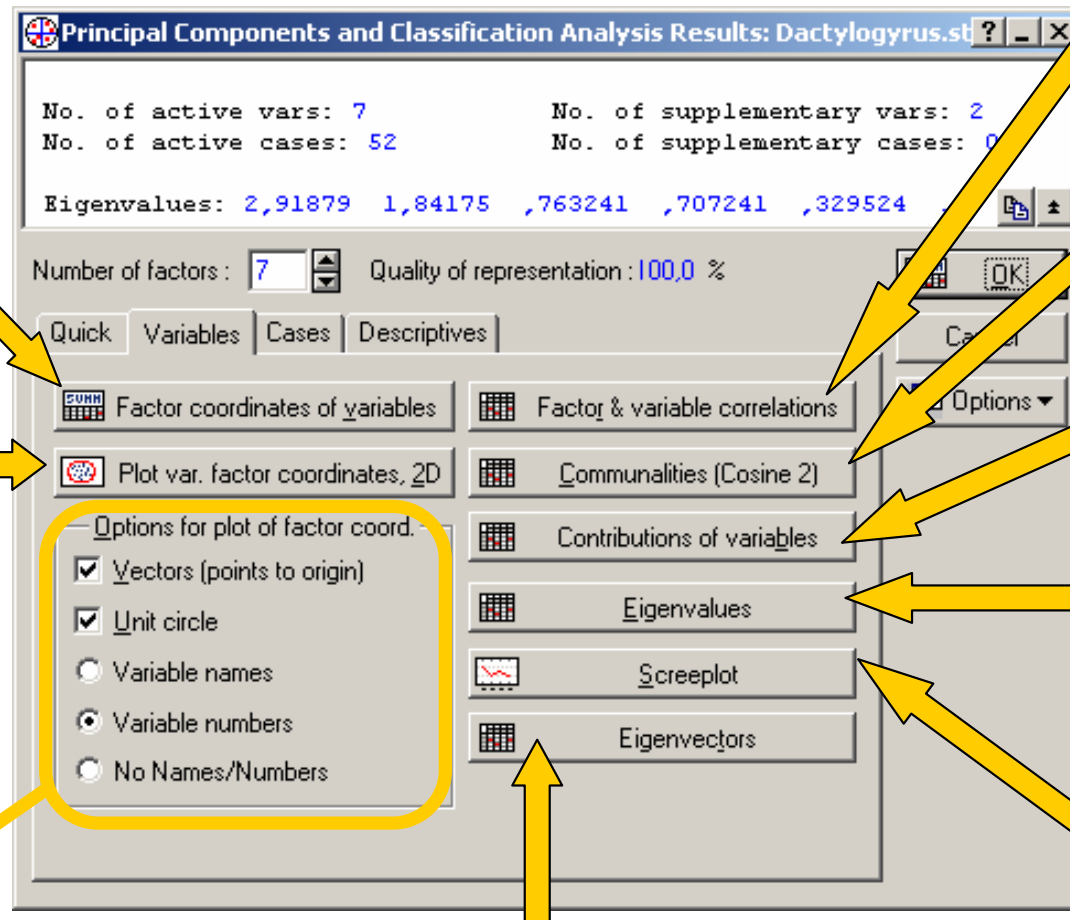
Objekty v ordinačním prostoru PCA

# Principal component analysis – výsledky II parametry

Koordináty parametrů na faktorových osách

2D graf parametrů vzhledem k faktorovým osám

Nastavení grafu



Eigenvectors – vektory faktorů v původním prostoru

Korelace proměnných a faktorů

Podíl variability proměnných vyčerpaný daným počtem faktorů

Příspěvek proměnných k jednotlivým faktorům

Eigenvalues

Grafické znázornění eigenvalues

## Eigenvectors

parametry

eigenvektor

Eigenvectors of correlation matrix (04\_PCA.sta)  
Active variables only

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
druh1	0,588908	0,074402	0,288827	-0,047020	-0,089109	0,334078	-0,665185
druh2	0,128282	0,395628	-0,293658	-0,791463	0,214690	0,205892	0,160906
druh3	0,580416	-0,046321	0,353300	0,113317	-0,007638	0,116967	0,713842
druh4	0,208348	-0,101580	-0,748786	0,280949	-0,371802	0,402509	0,080069
druh5	0,385665	0,506499	-0,256705	0,213964	-0,007673	-0,690967	-0,074492
druh6	-0,259438	0,484207	0,275625	-0,071370	-0,775276	0,083976	0,095226
druh7	-0,200980	0,578343	0,054560	0,478256	0,454490	0,431067	0,032251

faktory

## Contribution of variables

Variable contributions, based on correlations (04\_PCA.sta)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
druh1	0,346813	0,005536	0,083421	0,002211	0,007940	0,111608	0,442471
druh2	0,016456	0,156522	0,086235	0,626413	0,046092	0,042391	0,025891
druh3	0,336883	0,002146	0,124821	0,012841	0,000058	0,013681	0,509570
druh4	0,043409	0,010319	0,560680	0,078932	0,138236	0,162013	0,006411
druh5	0,148738	0,275541	0,065897	0,045781	0,000059	0,477436	0,005549
druh6	0,067308	0,274456	0,075969	0,005094	0,601053	0,007052	0,009068
druh7	0,040393	0,317481	0,002977	0,228729	0,206561	0,185819	0,001040

## Communalities

Communalities, based on correlations (04\_PCA.sta)

Variable	From 1 factor	From 2 factors	From 3 factors	From 4 factors	From 5 factors	From 6 factors	From 7 factors
druh1	0,758619	0,766473	0,863728	0,865800	0,871379	0,915756	1,000000
druh2	0,035996	0,258064	0,358600	0,945834	0,978215	0,995071	1,000000
druh3	0,736899	0,739943	0,885463	0,897500	0,897541	0,902981	1,000000
druh4	0,094953	0,109592	0,763249	0,837244	0,934360	0,998779	1,000000
druh5	0,325349	0,689322	0,766146	0,809064	0,809105	0,998944	1,000000
druh6	0,147230	0,479869	0,568436	0,573211	0,995470	0,998274	1,000000
druh7	0,088356	0,562907	0,566377	0,780801	0,925916	0,999802	1,000000

Příspěvek parametru k variabilitě faktoru



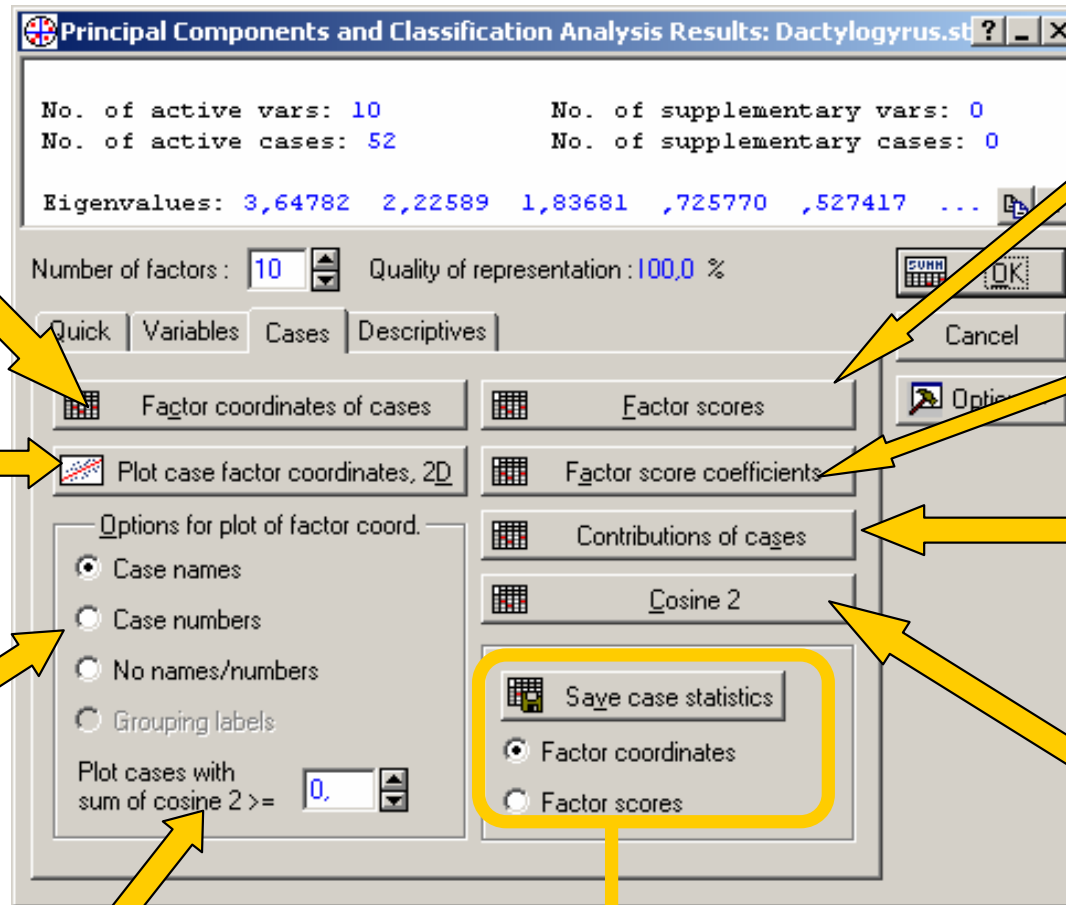
# Principal component analysis – výsledky III objekty

Koordináty objektů  
na faktorových  
osách

2D graf objektů  
vzhledem k  
faktorovým osám

Nastavení grafu

Výběr objektů podle sumy  $\cos^2$   
objektu pro dané faktory



Factor scores

Factor scores  
coefficients

Příspěvek  
proměnných k  
jednotlivým  
faktorům

$\cos^2$  úhlu mezi  
faktorem a  
vektorem objektu  
(communalities)

Uložit koordináty nebo scores objektů

## Factor scores

Factor scores, based on correlations (U4\_PCA.sta)  
Labelling variable: Sezona

Case	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Sezona
HEL	-0,41665	-1,11817	-0,43878	-0,04549	0,04664	0,23456	-0,33826	autumn 97
HVE	-0,33916	-0,91522	-0,01025	0,02618	0,32949	-0,40478	-0,15138	autumn 97
MEL	-0,50674	-0,69611	0,40285	-0,68382	0,20973	-1,04113	-1,08374	autumn 97
ROH	-1,03833	2,79055	0,66803	-0,54360	-2,18187	0,34226	0,24803	autumn 97
STR	-0,65383	2,15771	0,30309	1,56404	0,44430	-0,86945	-0,68897	autumn 97
HEL	-0,46172	-1,04486	0,40003	-0,27980	0,60565	-0,56458	-0,37709	spring 97
HVE	1,11484	0,66189	-2,19306	1,83468	0,02750	-2,15863	0,64897	spring 97
MEL	-0,31624	-0,79088	0,00396	0,44933	0,97071	0,34089	0,14034	spring 97
ROH	3,10406	0,56471	1,23343	-0,00031	-0,26027	0,25731	-1,08555	spring 97
STR	-0,40575	-0,70617	-0,52891	0,16358	-0,29698	0,77993	-0,26446	spring 97

objekty

Faktorové osy

Příslušnost objektů do skupin

Factor coordinates dělené odmocninou eigenvalue

## Factor scores coefficients

Factor score coefficients, based on correlations (U4\_PCA.sta)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
druh1	0,398184	0,062464	0,267498	-0,048563	-0,106314	0,52980	-1,52446
druh2	0,086736	0,332148	-0,271972	-0,817438	0,256141	0,32652	0,36876
druh3	0,392442	-0,038888	0,327210	0,117036	-0,009112	0,18549	1,63587
druh4	0,140872	-0,085281	-0,693490	0,290170	-0,443587	0,63832	0,00000
druh5	0,260763	0,425229	-0,237748	0,220986	-0,009155	-1,09557	0,17072
druh6	-0,175416	0,406514	0,255271	-0,073712	-0,924961	0,13317	0,21824
druh7	-0,135890	0,485545	0,050531	0,493953	0,542240	0,68361	0,07391

parametry

Faktorové osy

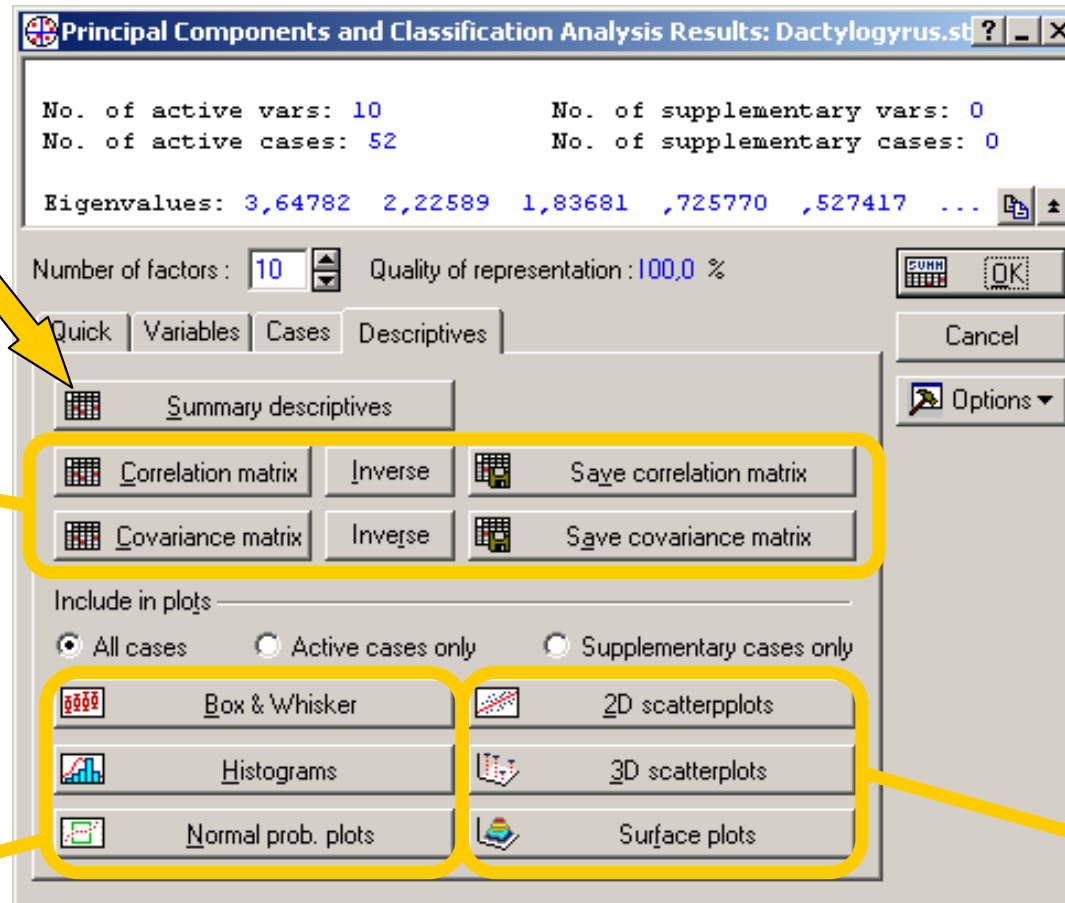
Eigenvektory podělené odmocninou eigenvalue

# Principal component analysis – popisná statistika

Průměr a SD  
proměnných

Korelační a  
kovarianční  
matice  
proměnných,  
inverze, uložení

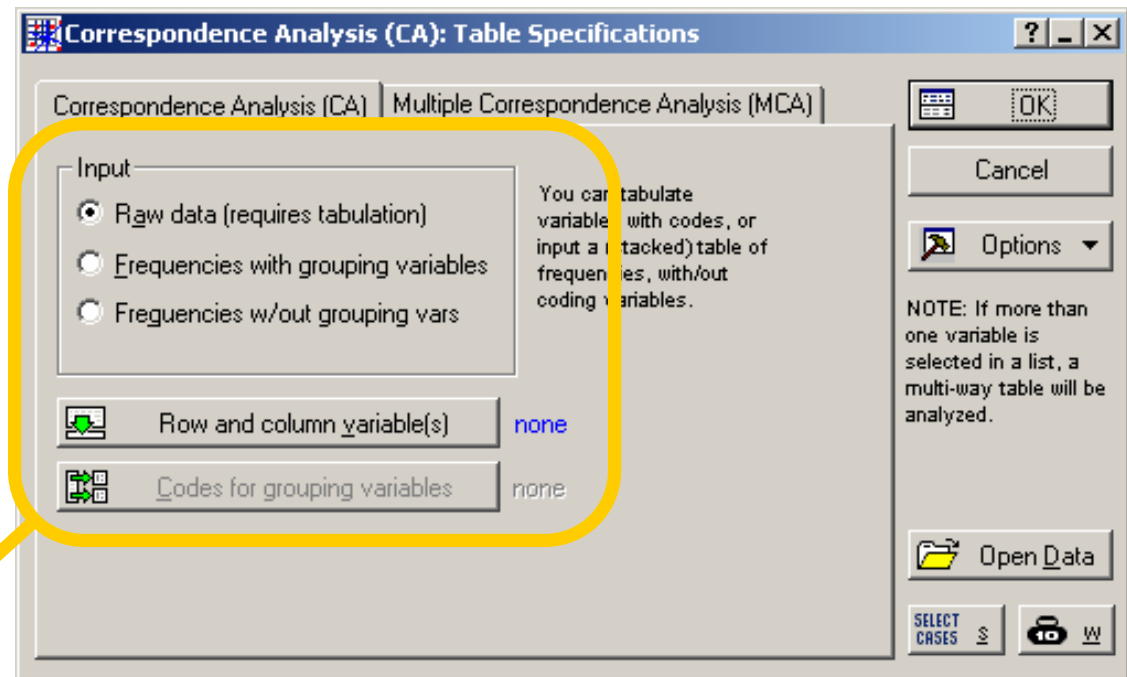
Popisné grafy  
jednotlivých  
proměnných



Zobrazení objektů  
podle různých  
proměnných

## Statistics >> Multivariate Exploratory Techniques >> Correspondence Analysis

Podstatou **korespondenční analýzy** je analýza kontingenčních tabulek, tj. tabulek kde průsečíky řádků a sloupců obsahují frekvenci dané kombinace



### Nastavení vstupních dat

**raw data** – proměnné s názvy řádků a sloupců tabulky – frekvence se dopočítají

**frekvence s třídící proměnnou** – sloupec názvů řádků, názvů sloupců, sloupec s frekvencemi

**frekvence bez řídící proměnné** – klasická tabulka – řádky X sloupce, na průsečíku frekvence

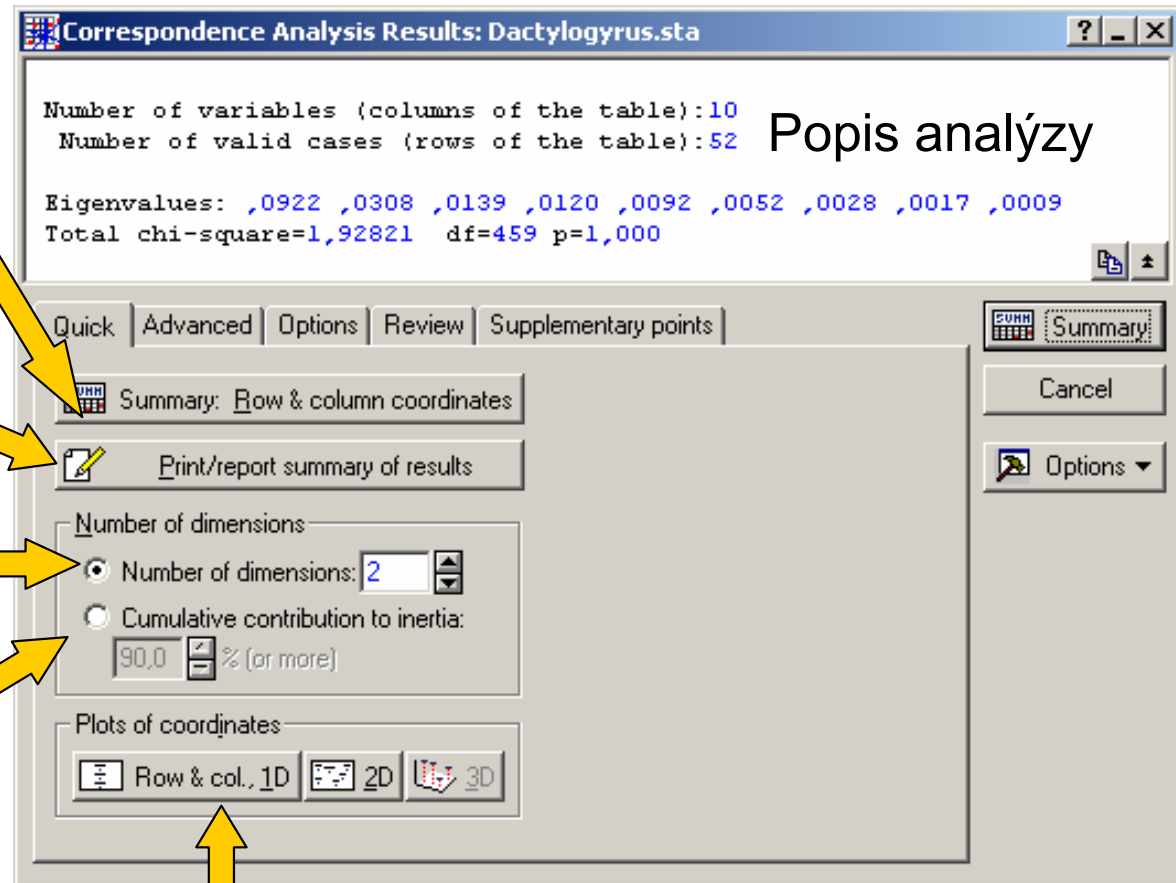
# Correspondence analysis – výsledky Quick

Koordináty řádků a sloupců v souřadném systému

Výstup všech základních výsledků

Počet rozměrů pro grafy a tabulky

Vybere počet os, vyčerpávajících určitou hodnotu inertia



1D, 2D, 3D grafy řádků a sloupců v souřadném systému

## Row and column coordinates

Koordináty v ordinačním prostoru CA

Obdoba kvality ( $\cos^2$  a relative inertia pro jednotlivé dimenze

Sloupce nebo řádky  
(samostatné  
tabulky)

Column Coordinates and Contributions to Inertia (05\_Correspondence.sta  
Input Table (Rows x Columns): 33 x 26  
Standardization: Row and column profiles

Column Name	Column Number	Coordn. Dim.1	Coordn. Dim.2	Mass	Quality	Relative Inertia	Inertia Dim.1	Cosine2 Dim.1	Inertia Dim.2	Cosine2 Dim.2
druh1	1	0,36515	-0,34050	0,123461	0,091865	0,109809	0,021925	0,049138	0,028166	0,042727
druh2	2	0,36340	-0,26874	0,1453295	0,503895	0,059935	0,079724	0,327038	0,064419	0,178857
druh3	3	0,10790	0,2126	0,006517	0,00012	0,012127	0,000101	0,002051	0,000580	0,007961
druh4	4	0,09817	0,071	0,019551	0,000754	0,012191	0,000251	0,005066	0,000197	0,002689
druh5	5	0,31392	-0,1479	0,010138	0,020860	0,015811	0,0001331	0,020710	0,000403	0,004249
druh6	6	0,14225	0,10737	0,001448	0,000220	0,020962	0,000039	0,000458	0,000096	0,000761
druh7	7	0,07044	0,00442	0,028602	0,080429	0,013987	0,000107	0,003326	0,004879	0,058103
druh8	8	0,32368	-0,2926	0,046343	0,020416	0,069855	0,006467	0,022783	0,001524	0,003633
druh9	9	0,26360	0,02069	0,008689	0,020002	0,007658	0,000804	0,025843	0,000007	0,000159
druh10	10	0,18414	0,09563	0,105720	0,050888	0,029033	0,004774	0,000472	0,001903	0,010916

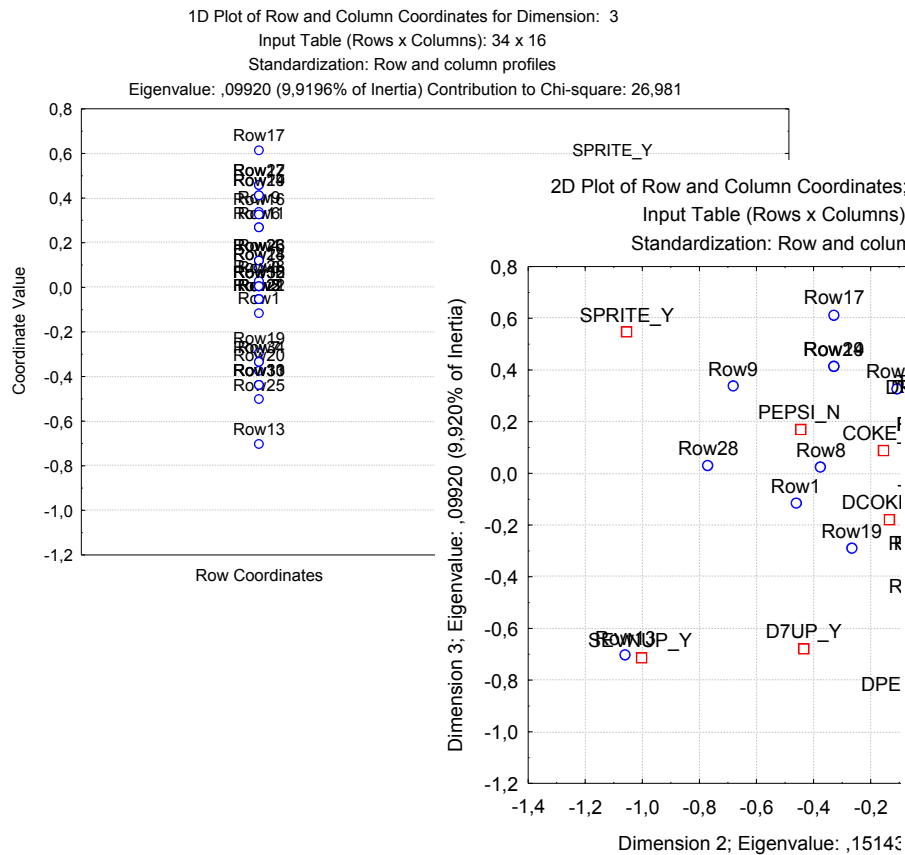
Celkový podíl řádku v tabulce  
relativních frekvencí (výpočet závisí  
na nastavení záložky options)

Kvalita zobrazení daného  
bodu daným počtem  
dimenzí (proporce bodu k  
celkové inertii dané  
počtem dimenzí)

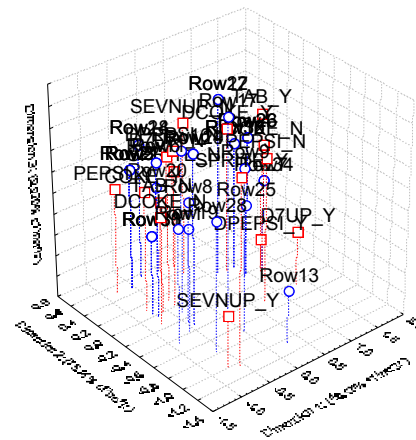
Podíl bodu na celkové  
inertii (neovlivněno  
počtem dimenzí)

# Grafy CA 1D,2D,3D

Grafy mohou být generovány pro všechny kombinace dimenzí.



3D Plot of Row and Column Coordinates; Dimension: 1 x 2 x 3  
 Input Table (Rows x Columns): 34 x 16  
 Standardization: Row and column profiles



Grafy obsahují koordináty jak řádků, tak sloupců původní tabulky.

○ Row.Coords  
 □ Col.Coords

# Correspondence analysis - nastavení

Počet rozměrů pro grafy a tabulky

Vybere počet os, vyčerpávajících určitou hodnotu inertia

Způsob standardizace koordinátů

1. Interpretace vzdáleností v rámci řádků i sloupců
2. Kanonická standardizace
3. Interpretace jen v rámci řádků
4. Interpretace jen rámci sloupců

Correspondence Analysis Results: Frekvence.sta

Number of variables (columns of the table): 26  
Number of valid cases (rows of the table): 33

Eigenvalues: ,7508 ,5082 ,4073 ,3549 ,2427 ,1961 ,1302 ,0843 ,0703 ...  
Total chi-square=16853,1 df=800 p=0,000

Quick | Advanced | Options | Review | Supplementary points

Number of dimensions

Number of dimensions: 2

Cumulative contribution to inertia: 90,0 % (or more)

Standardization of coordinates

Row & column profiles

Canonical standardization

Row profiles (interpret row dist.)

Column profiles (interpret col. dist.)

Summary

Cancel

Options



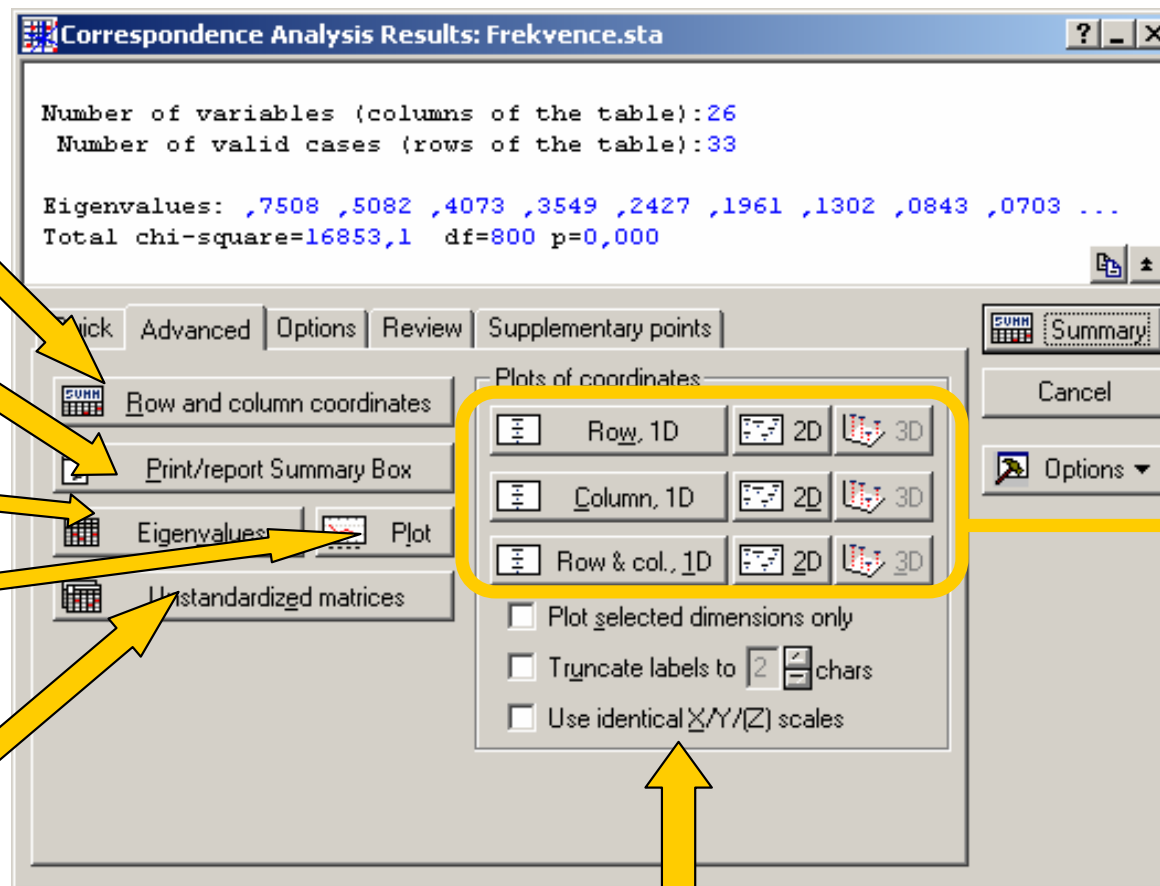
# Correspondence analysis – výsledky Advanced

Koordináty řádků a sloupců v souřadném systému

Výstup všech základních výsledků

Eigenvalues ~ inertia  
“vysvětlená“ sloupce + graf

Tabulka frekvencí výskytu + nestandardizované koordináty



1D, 2D, 3D grafy řádků a sloupců v souřadném systému

Nastavení grafů – výběr os, zkrácení popisek, identické měřítko

Jednotlivé dimenze

Eigenvalues and Inertia for all Dimensions (Beyrage.st  
Input Table (Rows x Columns): 34 x 16  
Total Inertia=1,0000 Chi2=272,00 df=495 p=1,0000

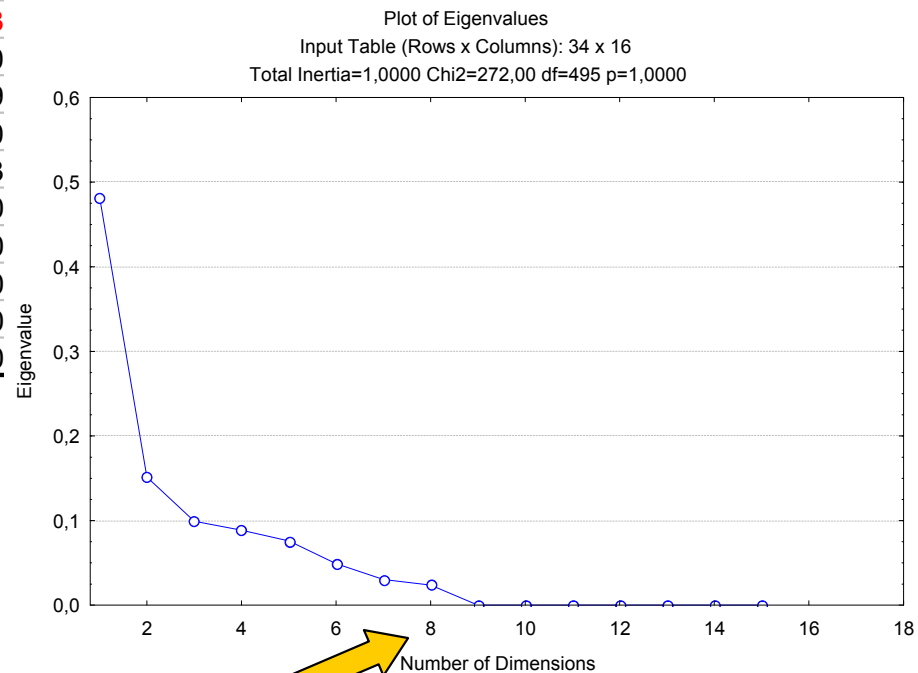
Number of Dims.	Singular Values	Eigen-Values	Perct. of Inertia	Cumulatv Percent	Chi Squares
1	0,694078	0,481745	48,17449	48,1745	131,0346
2	0,389141	0,151430	15,14304	63,3175	41,1891
3	0,314954	0,099196	9,91959	73,23	
4	0,297800	0,088685	8,86851	82,10	
5	0,275500	0,075900	7,59004	89,69	
6	0,221580	0,049098	4,90976	94,60	
7	0,174100	0,030311	3,03109	97,63	
8	0,153736	0,023635	2,36348	100,00	
9	0,000000	0,000000	0,00000	100,00	
10	0,000000	0,000000	0,00000	100,00	
11	0,000000	0,000000	0,00000	100,00	
12	0,000000	0,000000	0,00000	100,00	

eigenvalue

% inertia a kumulativní inertia vybraná dimenzí

Vysvětlený  $\chi^2$

eigenvalue



Počet dimenzí



**Korespondenční analýza** analyzuje kontingenční tabulky, k původní tabulce frekvencí je vytvořena tabulka očekávaných frekvencí a tyto dvě tabulky jsou pomocí  $\chi^2$  srovnány, analýza hledá takové nové dimenze, které vyčerpávají maximální část celkové  $\chi^2$  hodnoty (tzv. inertia)

Matrix of Relative Frequencies (05\_Correspondence.sta)  
Input Table (Rows x Columns): 33 x 26

	druh1	druh22	druh23	druh24	druh25	druh26	Total
HEL	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,016655
TYM	0,001448	0,000000	0,000000	0,000000	0,000000	0,000000	0,005612
KYJ	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,007983
MOR	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,001448
KYJ	0,001086	0,000000	0,000000	0,001267	0,000000	0,000000	0,179536
MOR	0,000362	0,000000	0,000000	0,000000	0,000181	0,000000	0,034938
KYJ	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,003077
MOR	0,033128	0,000000	0,000000	0,000000	0,000000	0,000000	0,051774
Total	0,123461	0,000905	0,000362	0,002896	0,002715	0,000000	1,000000

řádky

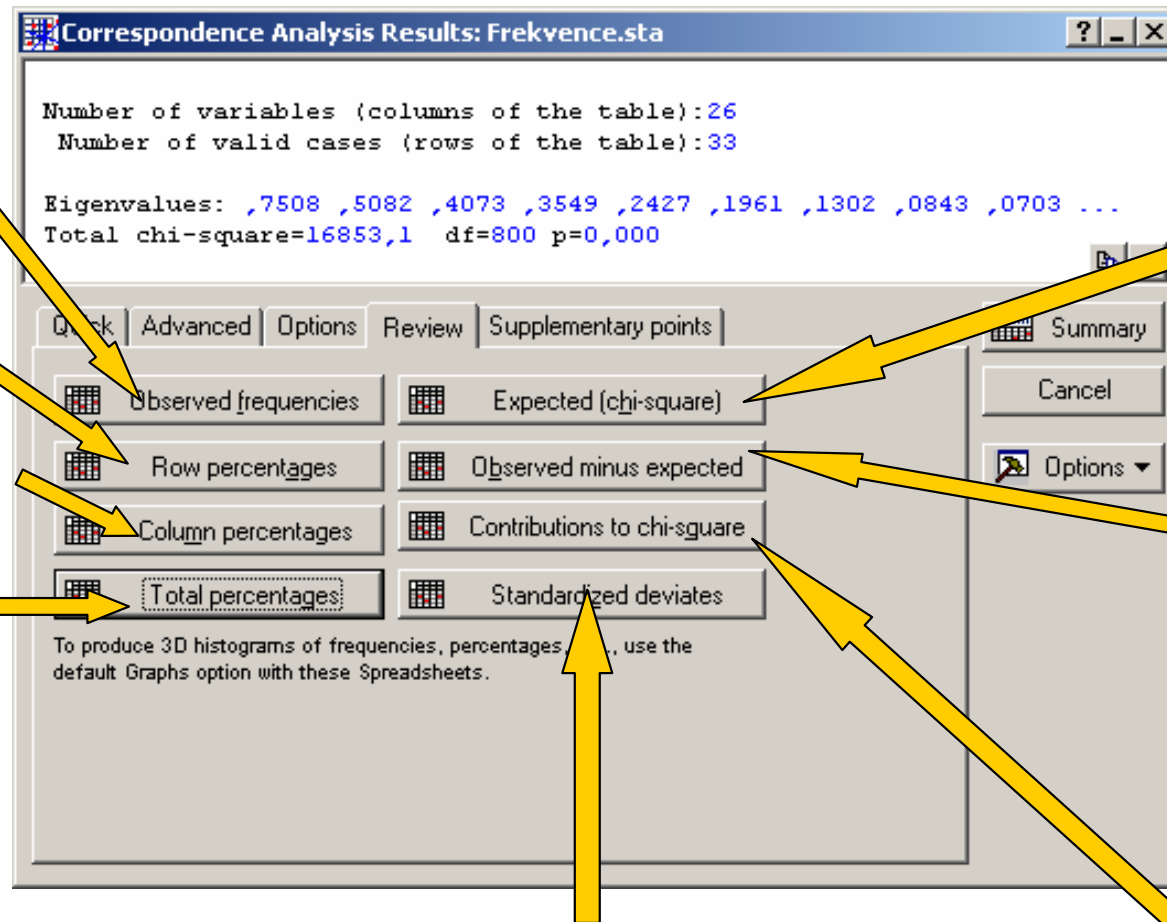
relativní frekvence bodů v původní matici

sloupce

Podíly řádků a sloupců

suma celé matice

# Correspondence analysis – přehledy



Pozorované četnosti

Podíly v řádcích

Podíly v sloupcích

Podíly v celé tabulce

Očekávané četnosti

Rozdíl pozorovaných a očekávaných frekvencí

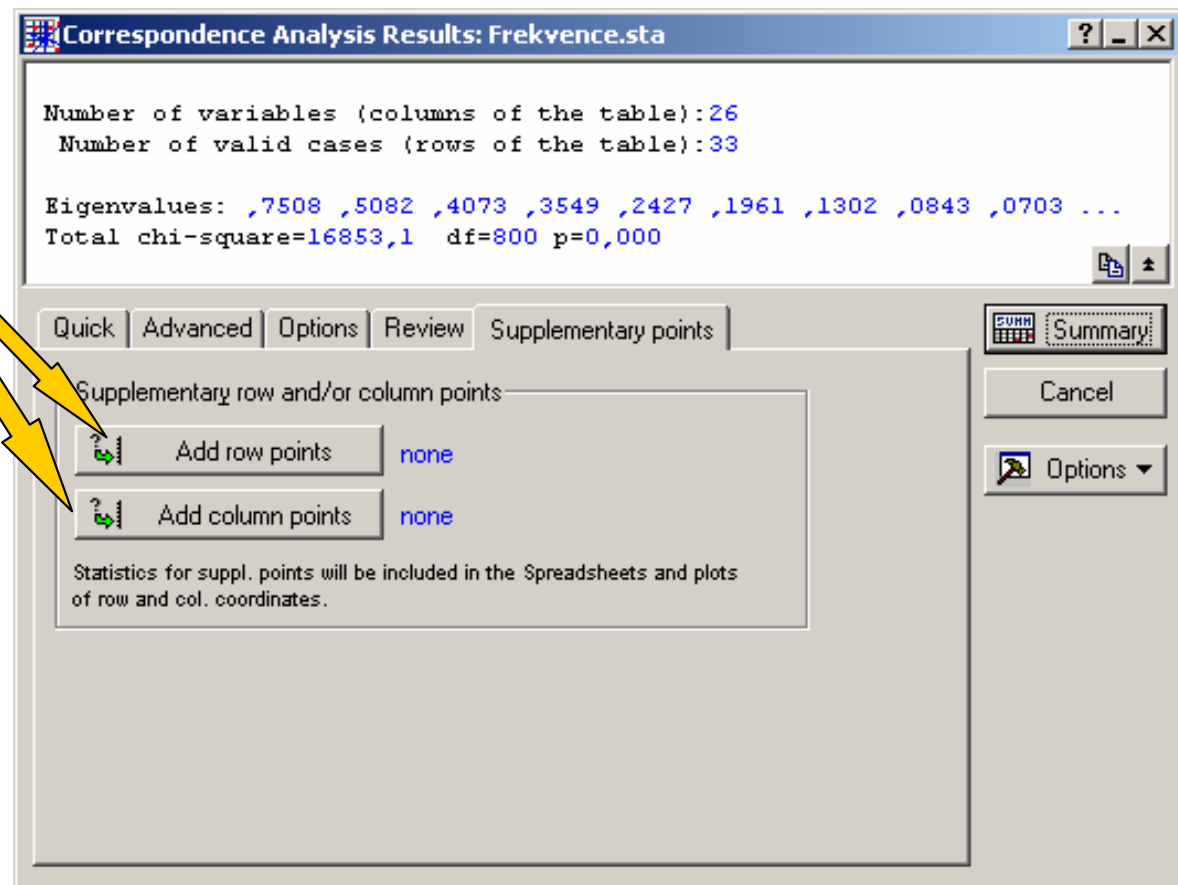
Standardized deviates – odmocnina vlivu na Chi – square + doplnění znaménka

Vliv jednotlivých položek tabulky na celkový Chi-square

K výsledkům analýzy je možné přidat další řádky nebo sloupce, jejichž pozice v souřadném prostoru se spočítají na základě CA, ale její výpočet neovlivní (obdobu supplementary variables a ne-active cases u PCA)

Přidání dalších řádků nebo sloupců

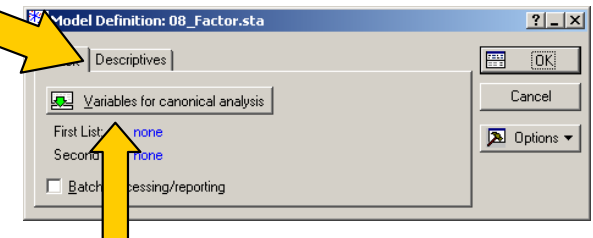
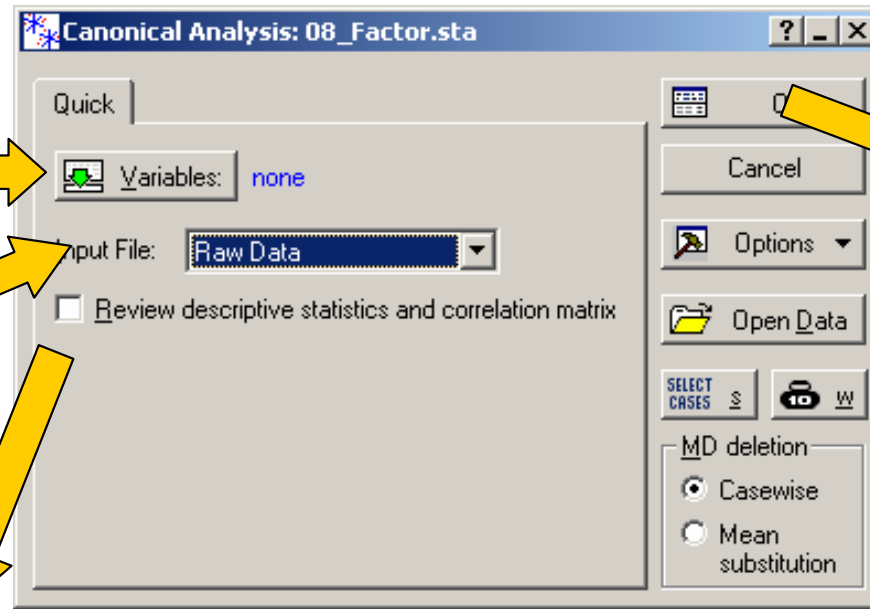
Přidají se do analýzy na základě již spočítaných parametrů



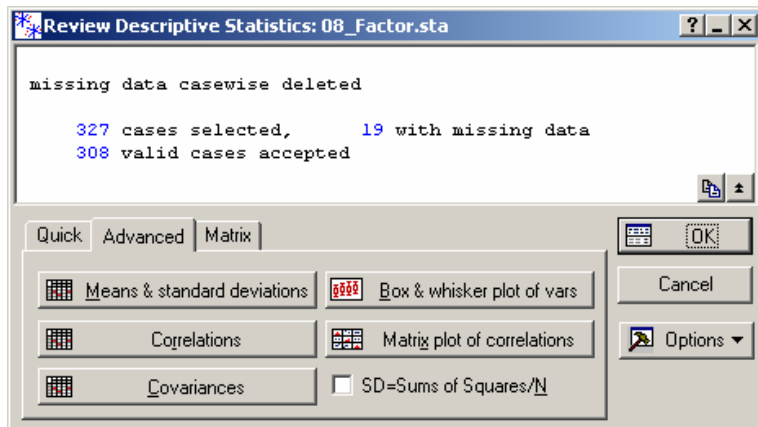
# Statistics >> Multivariate Exploratory Techniques >> Canonical Analysis

Výběr proměnných

Typ vstupních dat – n x p matice nebo korelační matice



Výběr proměnných pro jednotlivé tabulky



Deskriptivní statistika a korelační matice  
Grafy a popisná statistika

**Canonical analysis** umožňuje nadefinovat některé z parametrů jako nezávislé, jiné jako závislé, smyslem je studium vztahů těchto skupin parametrů

Canonical R: ,6134441  
Chi-Square: 148,2822 df = ( 12) p = 0,000000  
Number of valid cases: 308

	No. of vars.	Variance extracted	Total redundancy given the other set
Left set:	4	73,486764010%	14,930043926%
Right set:	3	100,000000000%	19,312344202%

Quick Canonical factors Factor structures Canonical scores Summary

Summary: Canonical results  
Eigenvalues  
Plot of eigenvalues  
Chi square tests

souhrn analýzy

eigenvalues

graf eigenvalues

$\chi^2$  testy – slouží pro rozhodnutí kolik canonical roots vybrat jako reprezentační

Canonical scores a weights (užití pro výpočet scores a interpretaci canonical roots)

Canonical R: ,6134441  
Chi-Square: 148,2822 df = ( 12) p = 0,000000  
Number of valid cases: 308

	No. of vars.	Variance extracted	Total redundancy given the other set
Left set:	4	73,486764010%	14,930043926%
Right set:	3	100,000000000%	19,312344202%

Quick Canonical factors Factor structures Canonical scores Summary

Left & right set canonical weights  
Save canonical scores

## Souhrn výsledků

Canonical Analysis Sun		
Canonical R: ,61344		
Chi2(12)=148,28 p=0,00		
	Left Set	Right Set
N=308		
<b>No. of variables</b>	4	3
Variance extracted	73,4868%	100,000%
Total redundancy	14,9300%	19,3123%
Variables:		
1	ROZ1	ROZ5
2	ROZ2	ROZ6
3	ROZ3	ROZ8
4	ROZ4	

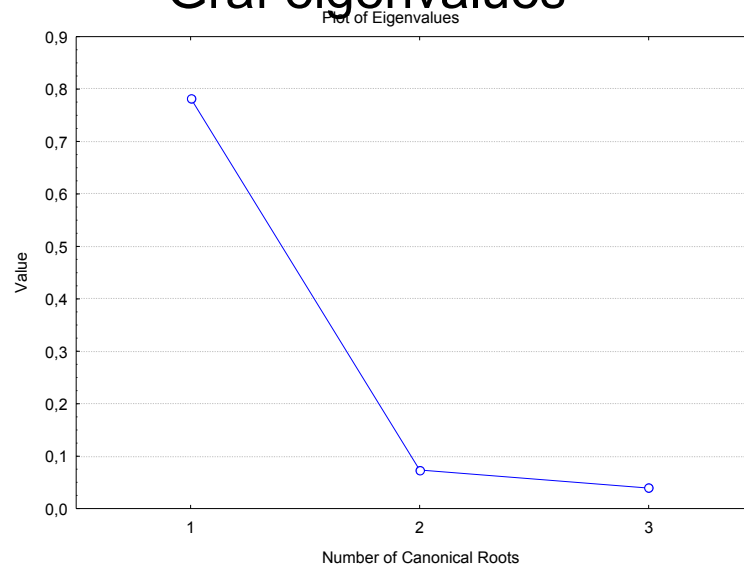
## $\chi^2$ testy

Root Removed	Chi-Square Tests with Successive Roots Removed (08_Fa					
	Canonical R	Canonical R-sqr.	Chi-sqr.	df	p	Lambda Prime
0	0,613444	0,376314	148,2822	12	0,000000	0,613006
1	0,130514	0,017034	5,2336	6	0,514234	0,982876
2	0,009591	0,000092	0,0279	2	0,986159	0,999908

## Eigenvalues

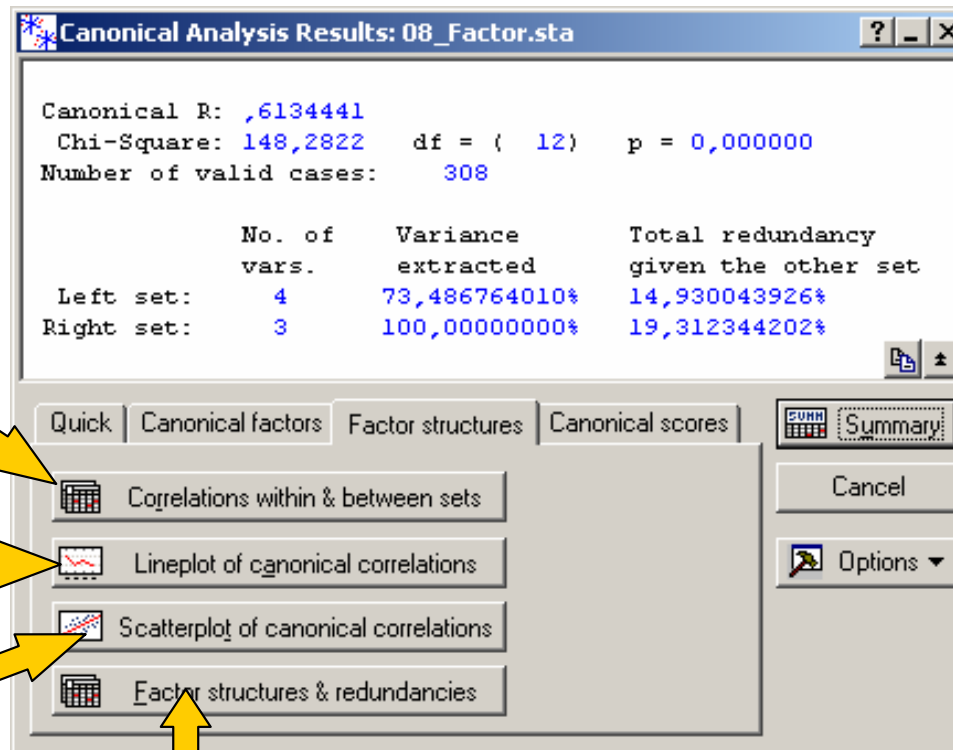
Root	Eigenvalues (08_Factor.sta)		
	Canonical Root 1	Canonical Root 2	Root 3
<b>Value</b>	0,376314	0,017034	0,000092

## Graf eigenvalues





## Canonical analysis – výsledky II



Korelace mezi proměnnými v rámci tabulek a mezi nimi

Graf kanonických korelací

XY graf faktorů proti sobě

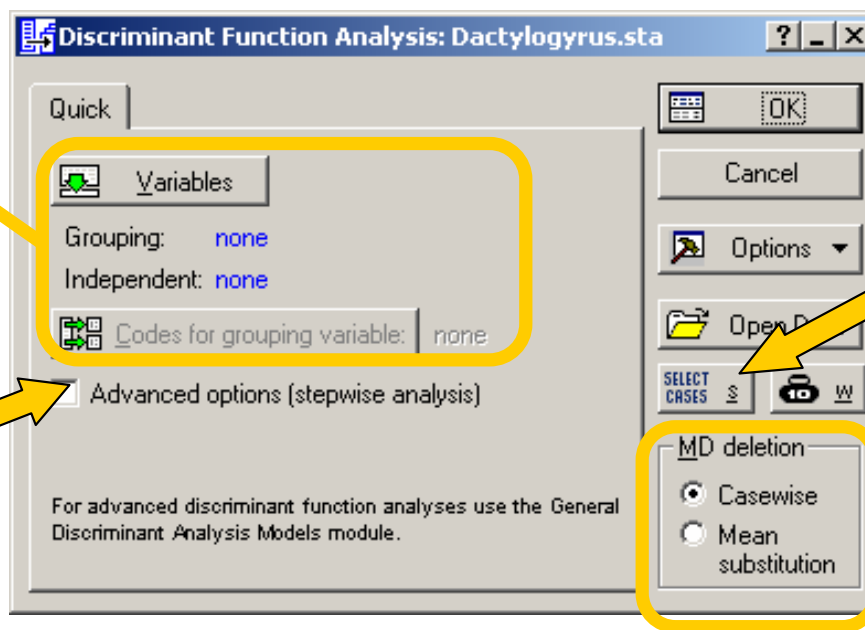
Struktura faktorů a redundance

## Statistics >> Multivariate Exploratory Techniques >> Discriminant Analysis

**Diskriminační analýza** na základě námi daného rozdělení objektů do skupin vytváří model pro jejich rozdělení podle parametrů

Nastavení proměnných s hodnotami a se skupinami + definice rozlišovaných skupin

Rozšířené možnosti specifikování modelu



Výběr z dat

Smazání chybějících dat nebo jejich nahrazení průměrem

# Discriminant analysis – definice modelu

Rychlé nastavení metody

Typ metody:

- Standartní
- Forward stepwise
- Backward stepwise

Nastavení stepwise metod

Model Definition: PCAall.STA

Variables: [List of variables]

Method: Standard

Tolerance: .010

Stepwise options:

- E to enter: 1,00
- F to remove: 0,00
- Number of steps: 4
- Display results: Summary only

Descriptives:

- Pooled within-groups covariances & correlations
- Means & number of cases
- Within-groups standard deviations
- Categorized histogram by group
- Box plot of means by group
- Categorized scatterplot by group
- Categorized normal probability plot by group

Review Descriptive Statistics: 06\_Discriminant.sta

Quick Within All cases

Pooled within-groups covariances & correlations

Means & number of cases

Within-groups standard deviations

Categorized histogram by group

Box plot of means by group

Categorized scatterplot by group

Categorized normal probability plot by group

## Discriminant analysis – výsledky

Popis výsledků – příspěvek jednotlivých proměnných k diskriminaci objektů

Vzdálenosti diskriminovaných skupin

Kanonická analýza

Discriminant Function Analysis Results: Dactylogyrus.sta

Popis analýzy

Number of variables in the model: 10

Wilks' Lambda: ,7183904 approx. F (10,41) = 1,607203 p < ,1389

Quick Advanced Classification

Summary: Variables in the model

Variables not in the model

Distances between groups

Perform canonical analysis

Stepwise analysis summary

Summary

Cancel

Options

# Discriminant analysis – výsledky tabulky

F spojené s danou WL

Wilk`s Lambda po vyjmutí parametru  
(0=perfektní diskriminace, 1=žádná diskriminace)

p spojené s daným  
F to remove

Discriminant Function Analysis Summary (06\_Discriminant.sta)  
No. of vars in model: 7; Grouping: D, UH (2 grps)  
Wilk's Lambda: ,01612 approx. F (7, 478)=4167,7 p<0,0000

N=486	Wilk's Lambda	Partial Lambda	F-remove (1, 478)	p-level	Toler.	1-Toler. (R-Sqr.)
ROZ1	0,016500	0,976988	11,2589	0,000856	0,432261	0,567739
ROZ2	0,026691	0,958971	313,4290	0,000000	0,581312	0,415688
ROZ3	0,017607	0,916584	44,0711	0,000000	0,751944	0,276056
ROZ4	0,017084	0,918588	28,5772	0,000000	0,481628	0,535372
ROZ5	0,016169	0,917022	1,4279	0,232698	0,681677	0,332323
ROZ6	0,016212	0,914356	2,7133	0,100175	0,901271	0,097729
ROZ8	0,016610	0,910503	14,5281	0,000156	0,781792	0,237208

R<sup>2</sup> (spjato s tolerance)

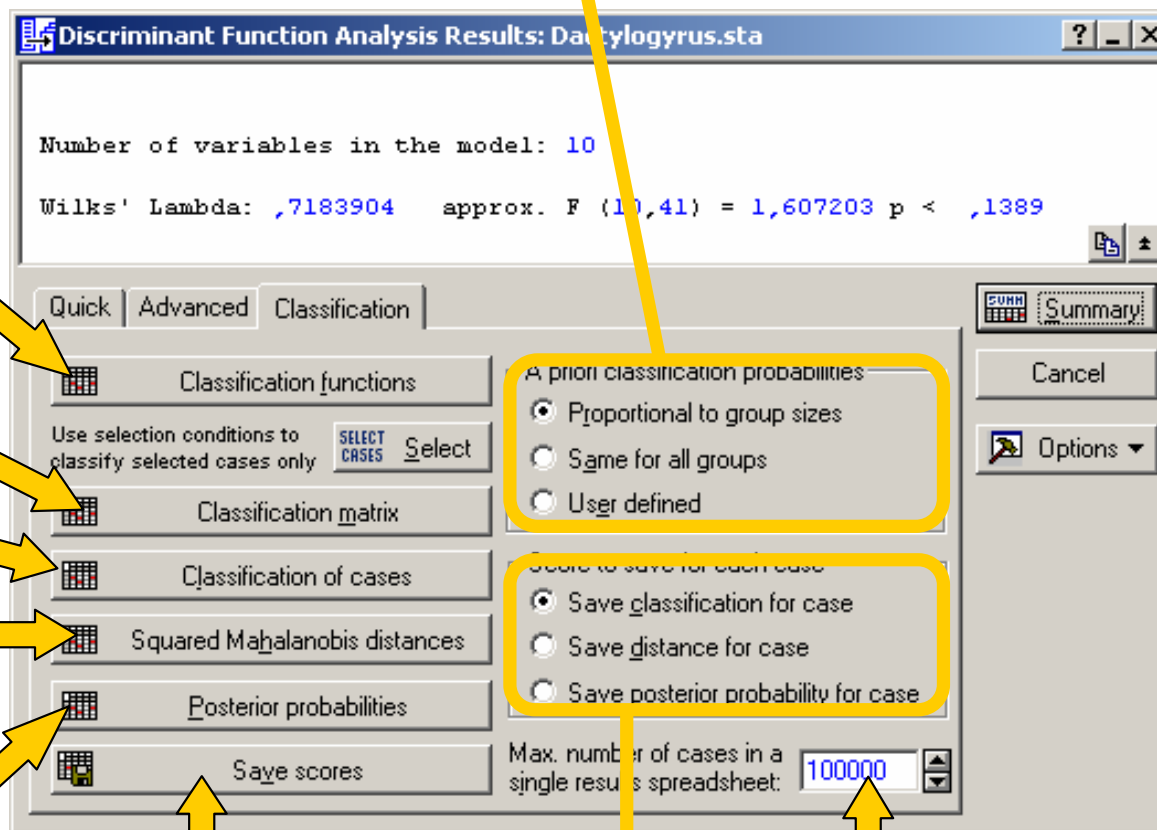
parametry

Wilk`s Lambda spojená s unikátním příspěvkem parametru k diskriminační síle modelu

Tolerance = měřítko redundance

# Discriminant analysis – výsledky klasifikace

Předem nastavená pravděpodobnost zařazení do skupiny



Klasifikační funkce

Pozorované a vypočítané příslušnosti do skupin

Klasifikace objektů

Mahalanobisova vzdálenost<sup>2</sup> objektů od centroidů skupin

Pravděpodobnost zařazení

Uložení klasifikace (jaký parametr a kolik objektů uložit)

# Discriminant analysis – výsledky klasifikace

## Koeficienty klasifikační funkce

Variable	Classification Function	
	PBIN p=,63374	PAN p=,36626
ROZ1	956,913	1923,03
ROZ2	6960,975	11766,81
ROZ3	5447,041	7612,83
ROZ4	1054,730	2527,01
ROZ5	28,245	509,99
ROZ6	2333,167	1509,32
ROZ8	2047,701	1062,15
Constant	-359,064	-861,43

Objekt patří do skupiny pro kterou mu vyšla vyšší hodnota funkce

## Vzdálenost do centroidů

### Vzdálenost od centroidů

Case	Observed Classif.	Squared Mahalanobis Distances from Centroids Incorrect classifications are marked	
		PBIN p=,63374	PAN p=,36626
408	PBIN	4,4234	328,2919
101	PAN	140,4836	25,3236
374	PBIN	7,4163	295,4637
375	PBIN	3,3083	262,1007
376	PBIN	4,5284	298,2785
289	PAN	264,9879	8,8166
290	PAN	240,0623	6,6247
291	PAN	265,5203	2,6734
301	PAN	248,9952	8,3603
605	PBIN	5,9409	289,9818
606	PBIN	5,5818	292,0057

Objekt

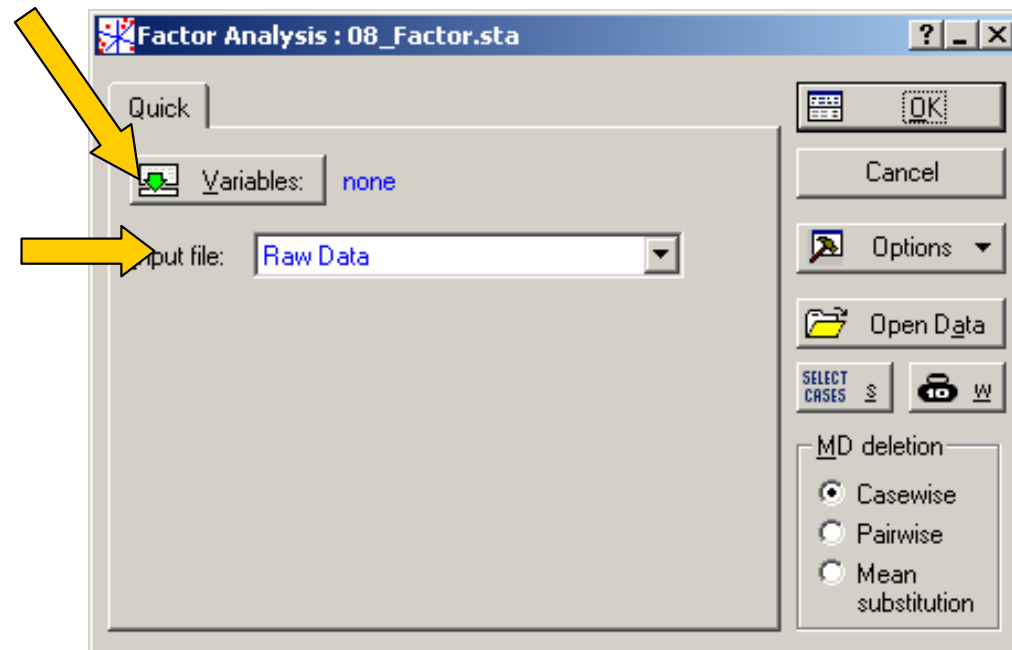
Jeho klasifikace

## Statistics >> Multivariate Exploratory Techniques >> Factor Analysis

**Faktorová analýza** - Účelem je zjištění struktury vztahů proměnných na základě korelace a redukce počtu proměnných.

Výběr proměnných

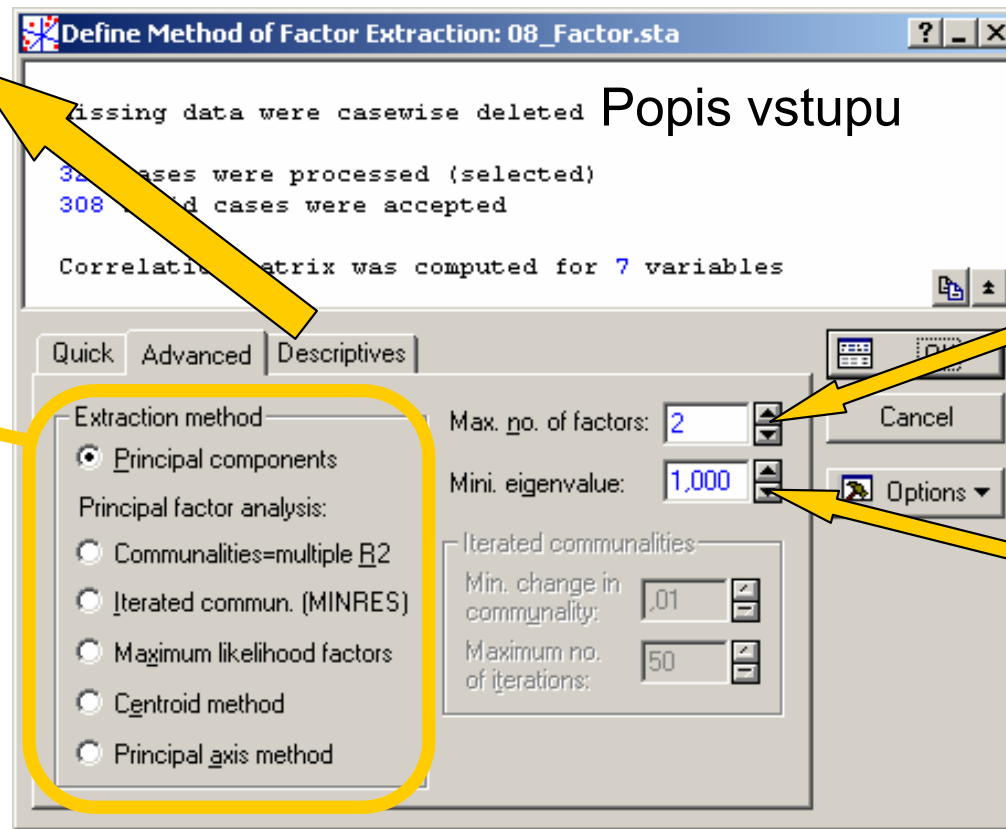
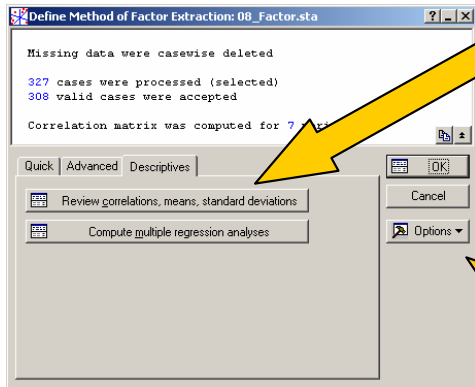
Typ vstupního souboru (matice  $n \times p$  nebo asociační matice korelací)





# Faktorová analýza - extrakce faktorů

Deskripce – popis parametrů, korelace, multiple regression atd.



Typ extrakce jednotlivých faktorů

Popis vstupu

Počet faktorů, které mají být extrahovány

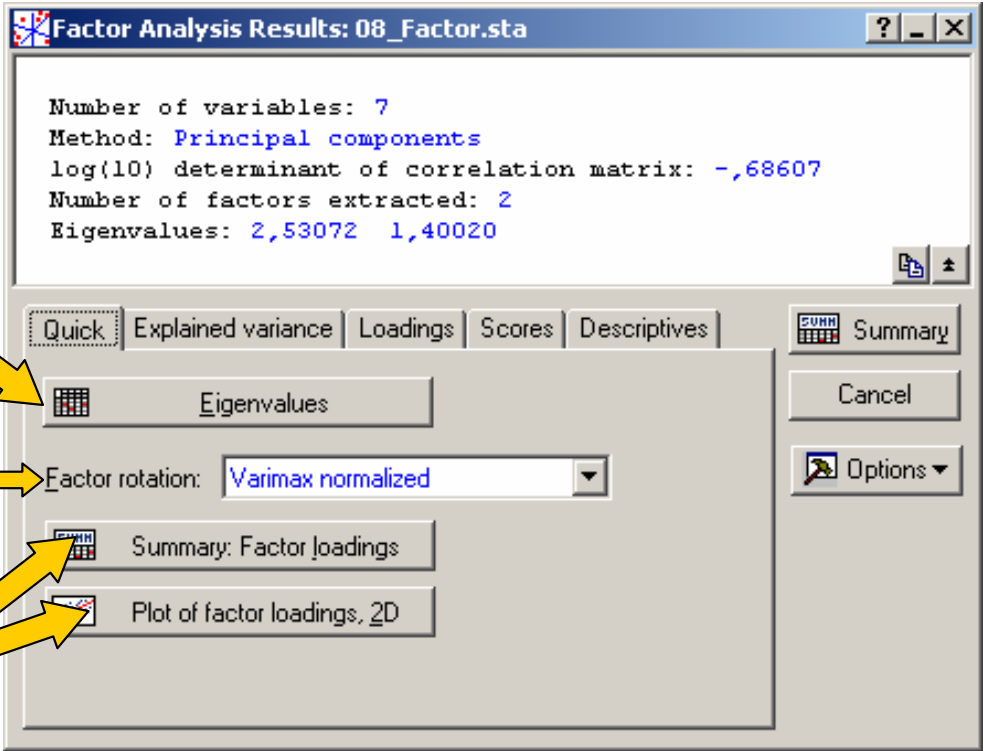
Jeich minimální eigenvalue

# Faktorová analýza - výsledky

Eigenvalues

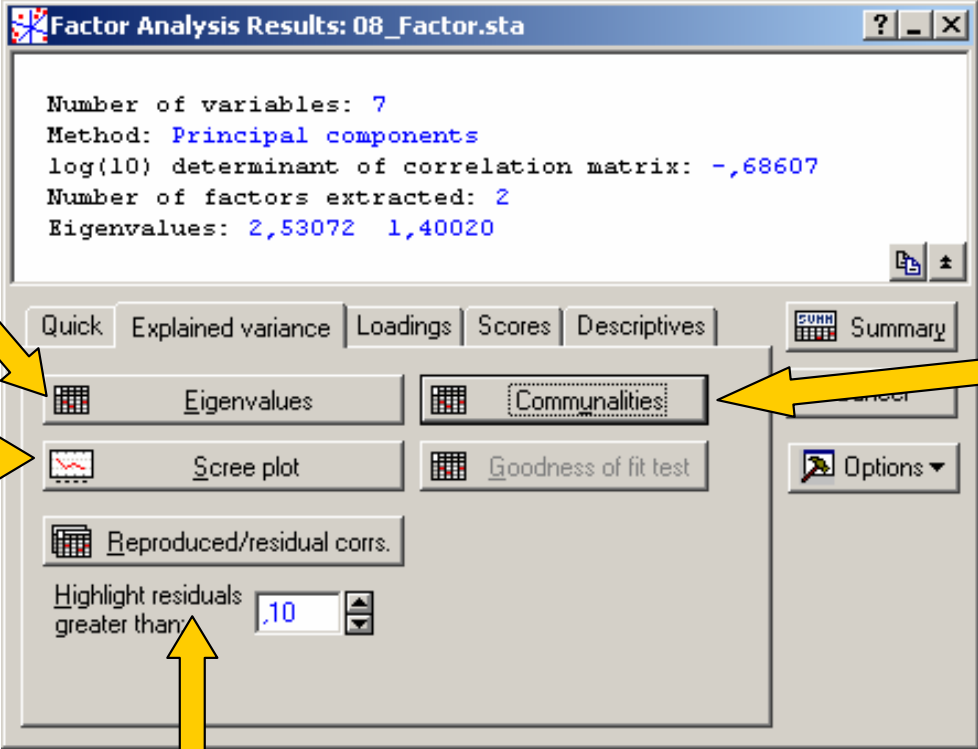
Způsob rotace

Factor loadings a jeho graf



The screenshot shows the 'Factor Analysis Results: 08\_Factor.sta' dialog box. The text area displays the following information: Number of variables: 7; Method: Principal components; log(10) determinant of correlation matrix: -,68607; Number of factors extracted: 2; Eigenvalues: 2,53072 1,40020. The 'Quick' tab is selected, showing options for 'Eigenvalues', 'Factor rotation: Varimax normalized', 'Summary: Factor loadings', and 'Plot of factor loadings, 2D'. A 'Cancel' button and an 'Options' dropdown are also visible.

## Faktorová analýza – výsledky II



The screenshot shows the 'Factor Analysis Results: 08\_Factor.sta' dialog box. The text area contains the following information:

```
Number of variables: 7  
Method: Principal components  
log(10) determinant of correlation matrix: -,68607  
Number of factors extracted: 2  
Eigenvalues: 2,53072 1,40020
```

The dialog box has several tabs: 'Quick', 'Explained variance', 'Loadings', 'Scores', 'Descriptives', and 'Summary'. The 'Quick' tab is active, showing options for 'Eigenvalues', 'Communalities', 'Scree plot', 'Goodness of fit test', and 'Reproduced/residual corr.'. A 'Highlight residuals greater than' field is set to '.10'. Yellow arrows point from external text labels to these specific options.

Eigenvalues

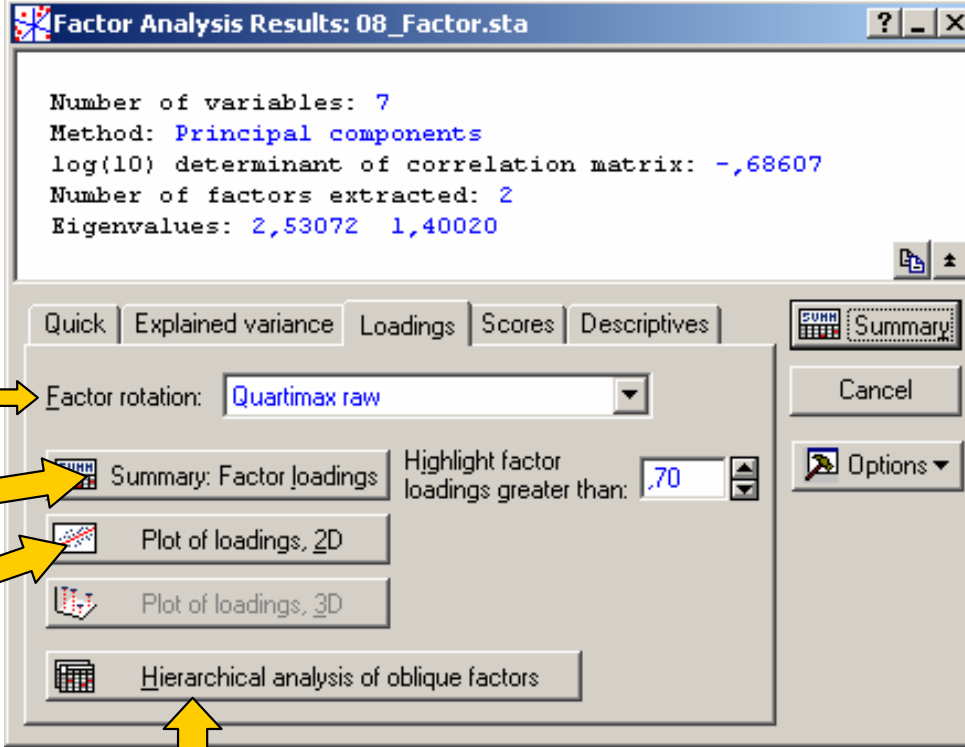
Scree plot (všechny eigenvalues)

Communalities

Highlight residuals greater than .10

Reprodukovaná a residuální korelační matice

## Faktorová analýza – výsledky III



The screenshot shows the 'Factor Analysis Results: 08\_Factor.sta' dialog box. The text area contains the following information:

```
Number of variables: 7  
Method: Principal components  
log(10) determinant of correlation matrix: -,68607  
Number of factors extracted: 2  
Eigenvalues: 2,53072 1,40020
```

The dialog box has several tabs: 'Quick', 'Explained variance', 'Loadings', 'Scores', and 'Descriptives'. The 'Quick' tab is active. The 'Factor rotation' dropdown menu is set to 'Quartimax raw'. The 'Highlight factor loadings greater than:' field is set to '.70'. The 'Options' dropdown menu is open, showing 'Summary: Factor loadings', 'Plot of loadings, 2D', 'Plot of loadings, 3D', and 'Hierarchical analysis of oblique factors'. The 'Hierarchical analysis of oblique factors' option is highlighted with a yellow arrow.

Annotations with yellow arrows point to the following elements:

- 'Způsob rotace' (Rotation method) points to the 'Factor rotation' dropdown menu.
- 'Factor loadings a jeho graf' (Factor loadings and its graph) points to the 'Summary: Factor loadings' and 'Plot of loadings, 2D' options.
- 'Hierarchical analysis of oblique factors' points to the 'Hierarchical analysis of oblique factors' option.

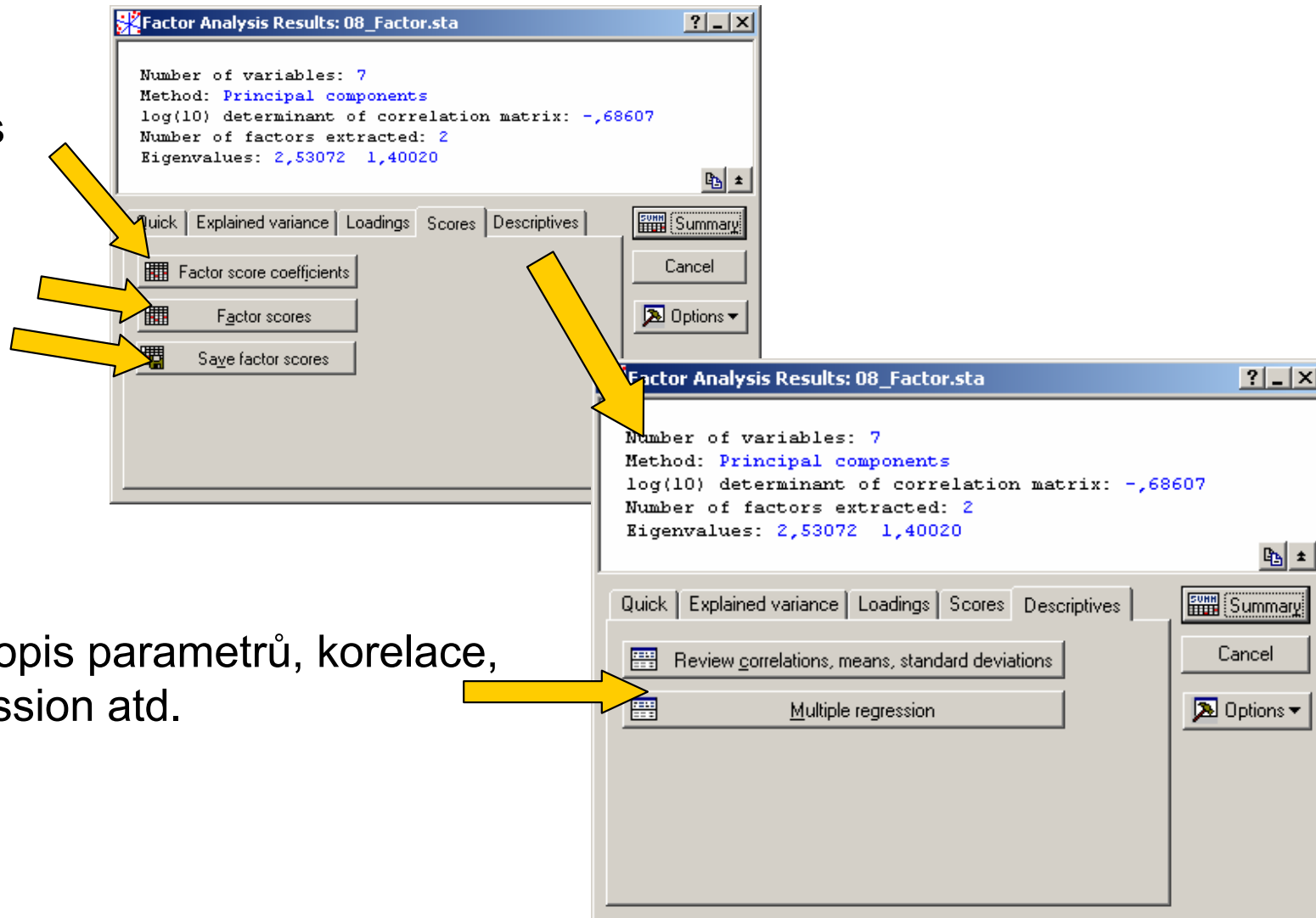
Hierarchical analysis of oblique factors - dvoustupňová analýza (nejprve výběr shluků proměnných podle jejich „unikátnosti“, pak tvorba sekundárních (se sdílenou variabilitou) a primárních faktorů (shluky podobných proměnných))

# Faktorová analýza – výsledky IV

Factors scores coefficients

Factor scores a jejich uložení

Deskripce – popis parametrů, korelace, multiple regression atd.

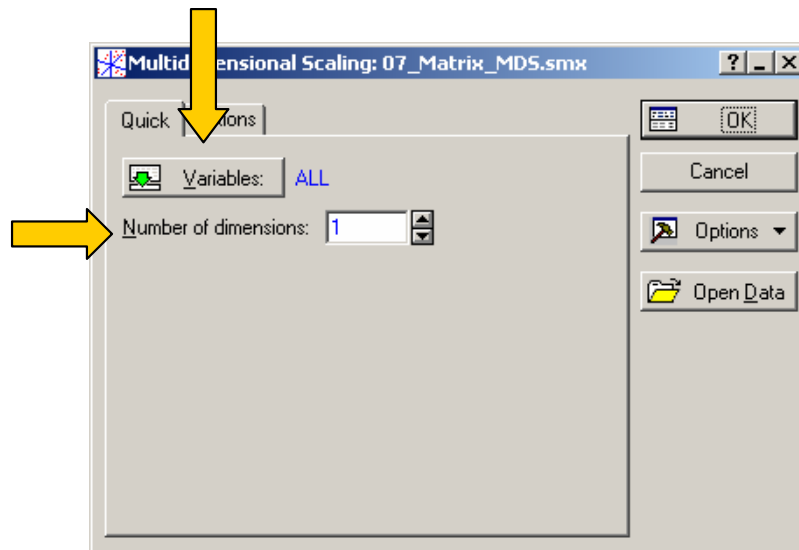


## Statistics >> Multivariate Exploratory Techniques >> Multidimensional Scaling

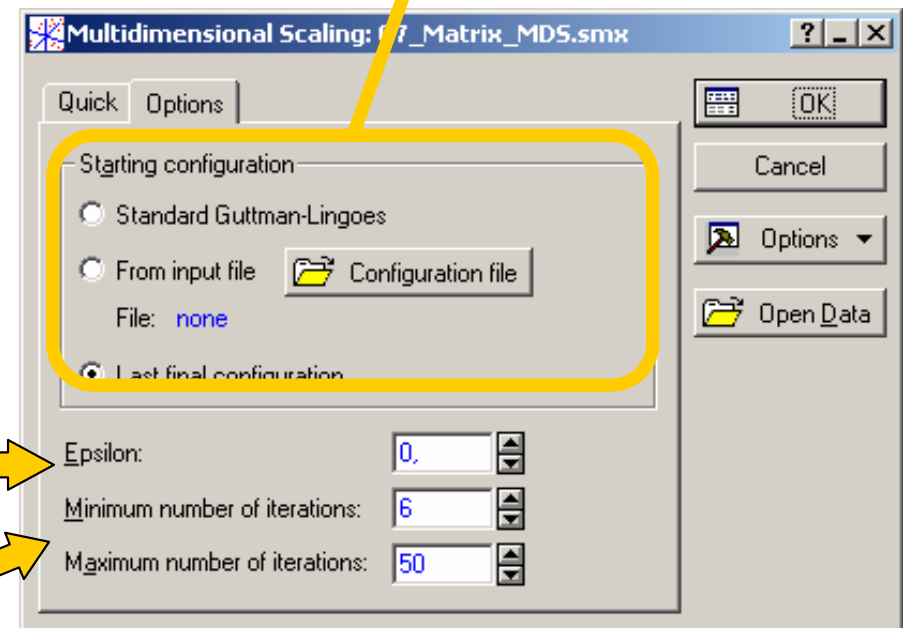
**Multidimensional scaling** dokáže na základě asociační matice s libovolnou metrikou vytvořit její Euklidovskou reprezentaci (příklad: na základě tabulky vzdáleností měst vytvoří mapu).

Výběr parametrů (vstupní soubor musí mít formát asociační matice)

Počet dimenzí k extrakci



Počáteční konfigurace



Vzdálenosti menší než jsou považovány za 0

Počty iterací

## Multidimensional Scaling - výpočet

**Multidimensional scaling** může sloužit pro přípravu podkladů pro k-means clustering pokud nemůžeme na naše data použít Euklidovskou vzdálenost. Metoda je výpočetně velmi náročná.

Parameter Estimation: 07\_Matrix\_MDS.smx

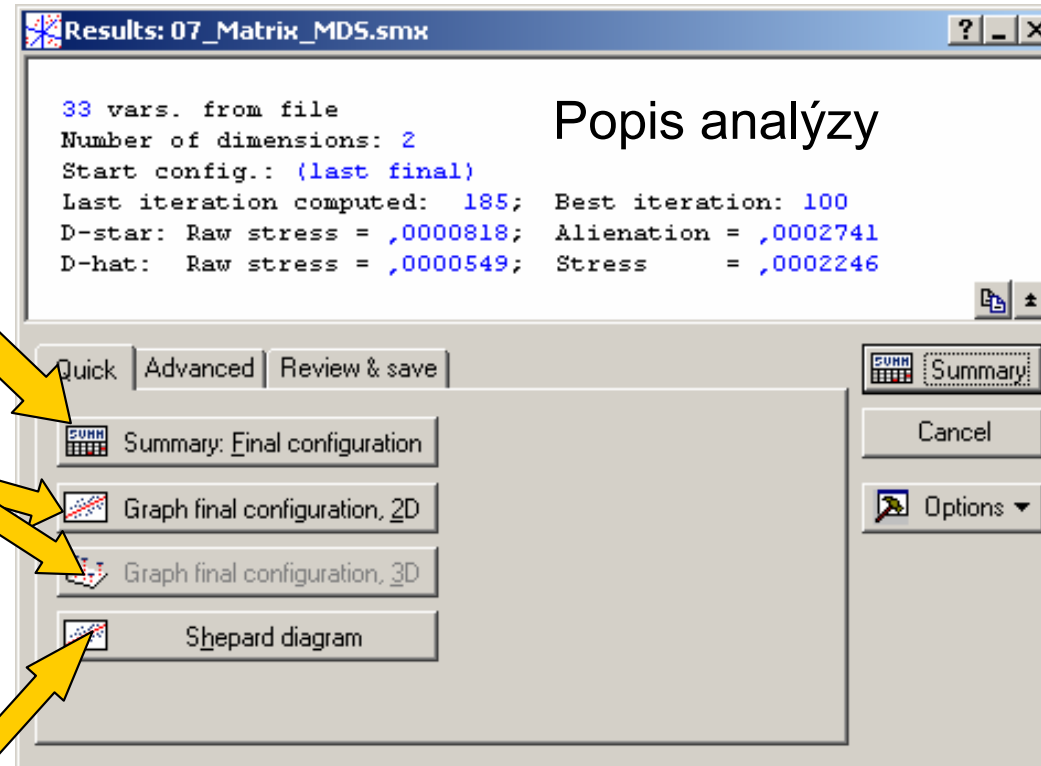
iter.	[dim=1]	D-star	D-star	D-hat	d-hat	
s: t:	cosin	step	raw stress	alienation	raw stress	stress
59	1	,758	,081		,0000005	,0000214
60	1	,518	,051		,0000004	,0000197
61	1	,672	,055		,0000004	,0000183
62	1	,891	,099		,0000003	,0000159
63	1	,826	,098		,0000002	,0000141
64	1	,424	,050		,0000002	,0000129
65	1	,515	,043		,0000002	,0000122
66	1	,901	,094		,0000001	,0000107
67	1	,942	,141		,0000001	,0000088
68	1	,604	,069		,0000001	,0000080
69	1	,262	,041		,0000001	,0000075
70	1	,770	,063		,0000001	,0000068
71	1	,939	,122		,0000000	,0000058
72	1	,802	,098		,0000000	,0000051
73	1	,360	,048		,0000000	,0000047
72 *			,0000000	,0000057	,0000000	,0000047
56	1	,624	,054		,0000010	,0000300
57	1	,795	,074		,0000008	,0000271
58	1	,850	,096		,0000006	,0000238

Estimation procedure converged

Cancel OK

Parametry měnící se při přepočtech

# Multidimensional Scaling – výsledky Quick



Výstup nových dimenzí + charakteristiky

Výstupní 2D a D graf

Shepard diagram ~ věrnost reprezentace



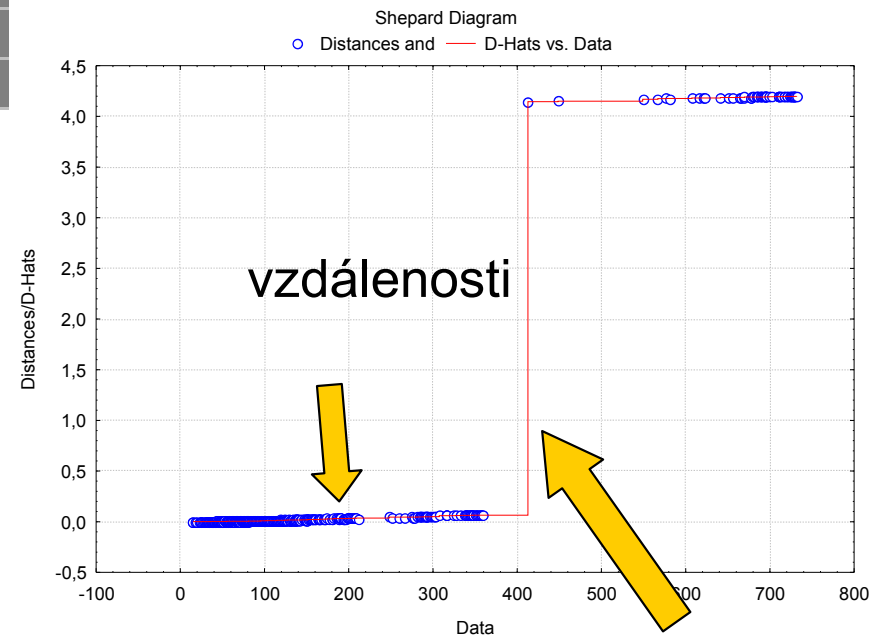
# Multidimensional Scaling – výsledky tabulky

Final Configuration (07\_Matrix\_MDS.smx)  
 D-star: Raw stress = ,0000818; Alienation = ,0002741  
 D-hat: Raw stress = ,0000549; Stress = ,0002246

objekty

	DIM. 1	DIM. 2	
HEL	-0,254837	0,000931	Nové dimenze
HVE	-0,254777	0,000982	
MEL	-0,254272	0,000494	
ROH	-0,255098	-0,002506	

## Shepard diagram



**Stress** – měřítko reprezentace, čím nižší, tím lepší reprezentace  
**Alienation** – cizost, čím nižší, tím lepší reprezentace

D-hat ~ průběh vzdáleností při dobré reprezentaci

# Multidimensional Scaling – výsledky Advanced

The screenshot shows a window titled "Results: 07\_Matrix\_MDS.smx" with the following text:

```
33 vars. from file
Number of dimensions: 4
Start config.: (last final)
Last iteration computed: 270; Best iteration: 100
D-star: Raw stress = ,0868132; Alienation = ,0089284
D-hat: Raw stress = ,0559948; Stress = ,0071707
```

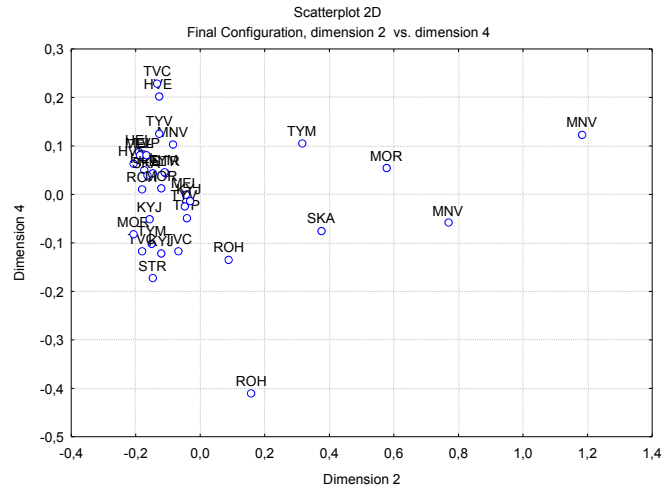
Below the text is a grid of buttons for various outputs:

Summary: Final configuration	Graph final configuration, 2D
D-hat values	Graph final configuration, 3D
D-star values	Graph D-hat vs. distances
Distance matrix	Graph D-star vs. distances
Summary statistics	Shepard diagram

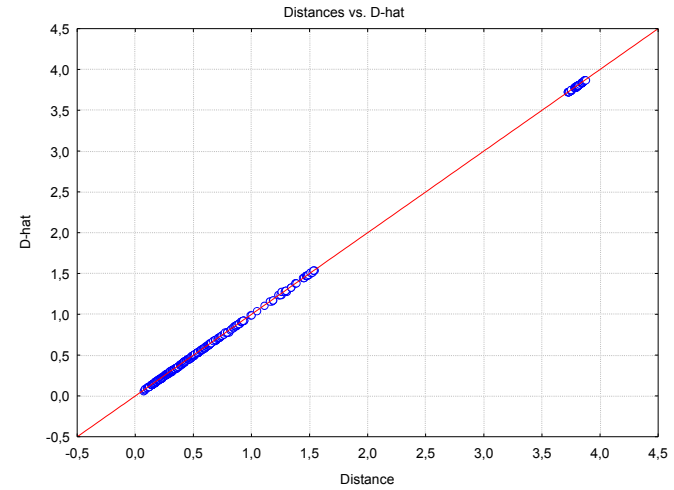
Yellow arrows point from the following text labels to the corresponding buttons in the screenshot:

- Výstup nových dimenzí + charakteristiky** points to the "Advanced" tab.
- Výstupní 2D a 3D graf** points to "Graph final configuration, 2D" and "Graph final configuration, 3D".
- D-hat, D-star** points to "D-hat values" and "D-star values".
- Matrice vzdáleností (reprodukovaná)** points to "Distance matrix".
- Sumární hodnoty (reprodukovaná vzdálenost, D-hat, D-star)** points to "Summary statistics".
- Shepard diagram** points to "Shepard diagram".
- D-hat, D-star versus reprodukovaná vzdálenost ~ věrnost reprodukce** points to "Graph D-hat vs. distances" and "Graph D-star vs. distances".

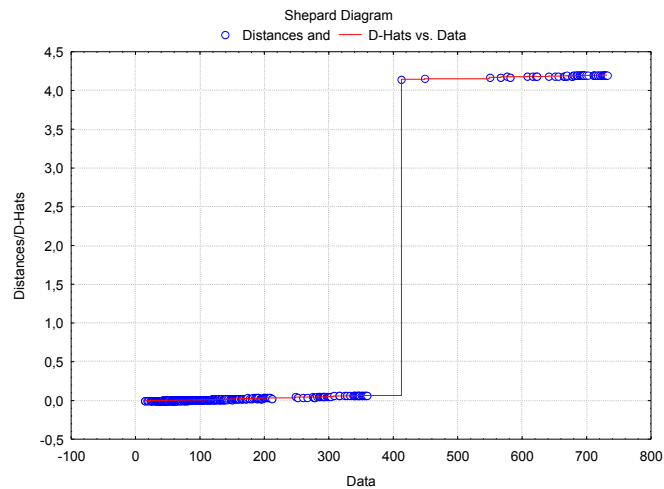
## 2D graf



## Vzdálenosti x D-har (D-star)



## Shepard diagram



## 3D graf

