Klasifikace objektů

hledání struktury a vzájemných vazeb v objektech

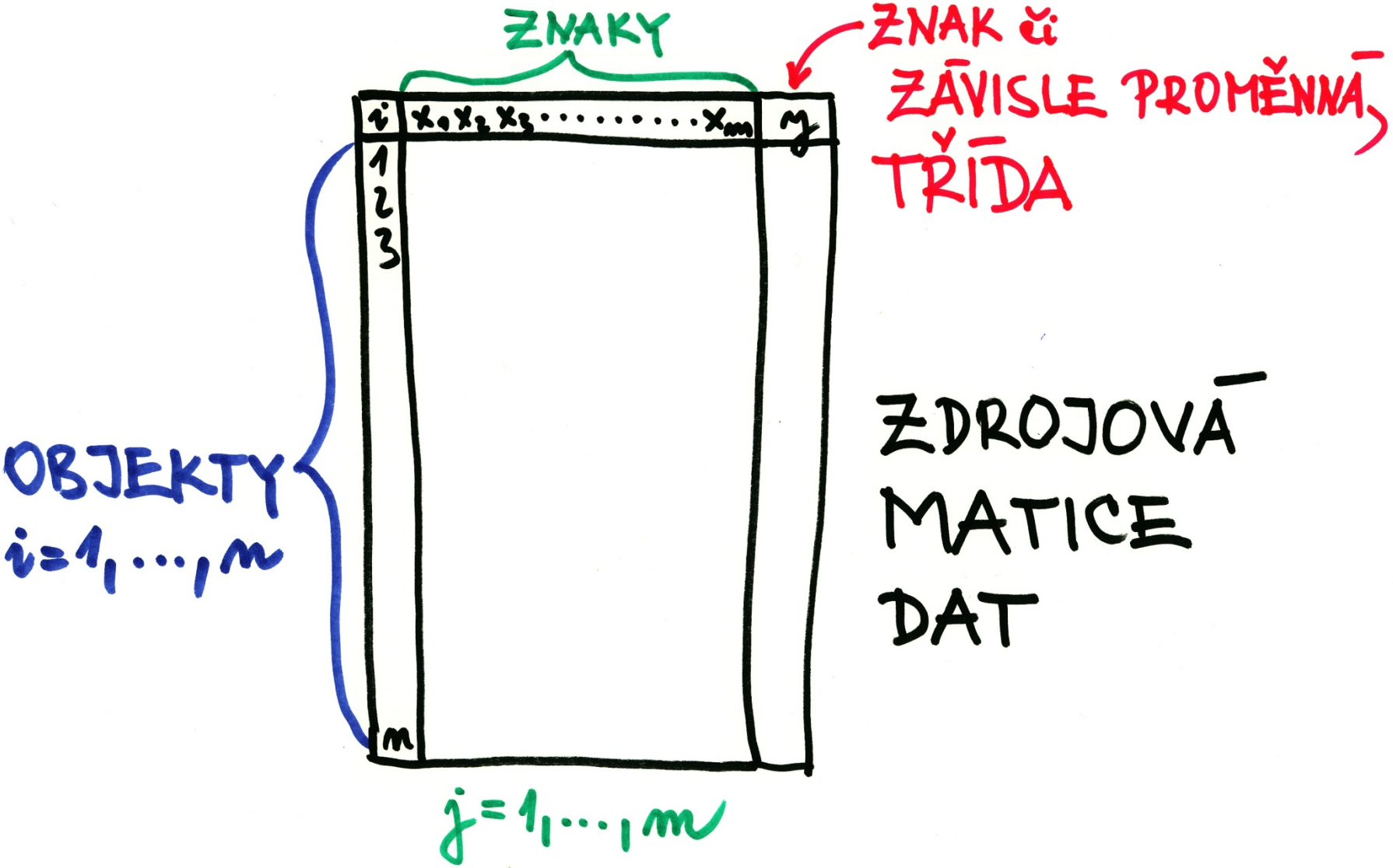
1) *Diskriminační analýza DA:* nový objekt se zařadí do již existující třídy.

2) *Analýza shluků CLU*: neuspořádanou skupinu objektů lze uspořádat do několika vnitřně sourodých tříd či shluků

3) *Vícerozměrné škálování MDS*: hledá strukturu a vazby mezi objekty na základě jejich podobnosti.

Analyzovaný výběr jsou trénovací data, která obsahují pro každý objekt jak výstup y, tak i hodnoty všech znaků $\boldsymbol{x}$, $\{y; x_1, x_2, \ldots x_m\}$

Klasifikovaný výběr: na základě analyzovaného výběru sestavit predikční model, který umožní zařazení nových objektů do tříd.



ZDROJOVÁ MATICE DAT

KLASIFIKACE OBJEKTŮ

y $\quad$ $x_1\ x_2\ x_3\ x_4 \ldots\ldots\ x_m$

→ PCA

→ FA

→ SHLUKY

→ MDS

DISKRIMINACE DA

(NORMALITA)

y $\quad$ $x_1\ x_2\ x_3\ x_4 \ldots\ldots\ x_m$

TŘÍDA $\quad$ PROMĚNNÉ

LOGISTICKÁ REGRESE, Logit

(NENORMALITA)

DISKRIMINAČNÍ ANALÝZA

TRÉNOVÁNÍ S UČITELEM (SUPERVISED LEARNING)

$$\boxed{y} \quad \boxed{x_1 \; x_2 \; x_3 \; x_4 \; \cdots\cdots \; x_m} \longrightarrow \text{MODEL (FISHEROVA FUNKCE)}$$

1. TRÉNOVACÍ DATA
(ANALYZOVANÝ SOUBOR)

$$\boxed{} \quad \boxed{x_1 \; x_2 \; x_3 \; x_4 \; \cdots\cdots \; x_m} \longleftarrow \text{VYUŽITÍ MODELU}$$

2. KLASIFIKACE OBJEKTŮ
(KLASIFIKOVANÝ SOUBOR)

TRÉNOVÁNÍ BEZ UČITELE (UNSUPERVISED LEARNING)

$\square$ $\boxed{x_1\ x_2\ x_3 \cdots\cdots\ x_m}$

1. obvykle PCA, FA, SHLUKY
2. DA – KLASIFIKAČNÍ SOUBOR

1) **Statistické učení s učitelem (supervised learning):** u dat analyzovaného výběru se sestaví predikční model $y = f(\boldsymbol{x})$, který umožní predikovat výstup pro nový objekt z dat klasifikovaného výběru x_0.

2) **Statistické učení bez učitele (unsupervised learning):** k dispozici jsou určité znaky pro objekty a nikoliv výstupy. Úlohou je pak pouze stanovit shluky.

Výstupem klasifikace jsou kvalitativní proměnné označující pouze třídy (jako je zdravý, nemocný, nebo nevyhovující, dobrý, vyhovující atd.).

Často jsou výstupem pouze dvě kategorie (binární výstup) kódované jako 0, 1, respektive NULA, JEDNA, nebo -1, 1.

Když $\boldsymbol{G}$ obsahuje pouze dvě třídy (binární proměnné 0 a 1), využívá se náhrady $\boldsymbol{G}$ kvantitativním výstupem y s tím, že predikce leží v intervalu [0, 1] a zařazuje do třídy NULA pro ≤ 0.5, resp. do třídy JEDNA pro ≥ 0.5.

Diskriminační analýza DA

hodnotí rozdíly mezi dvěma nebo více skupinami objektů charakterizovaných pomocí několika znaků.

Dělí se na

1) Techniky, které *vysvětlují rozdíly* mezi předem danými shluky objektů.
2) Techniky, u kterých je cílem *klasifikace objektů do skupin*. Jsou porovnávány znaky jednoho objektu (např. charakteristiky sloučenin, vlastnosti pacientů, ...) se znaky ostatních.

Příklad: Analyzovaný výběr obsahuje *třídu přežívajících* pacientů a *třídu nepřežívajících*. Dle klinických vyšetření chceme rozhodnout, kam nového pacienta z klasifikovaného výběru zařadit. Potřebujeme znaky k rozlišení dosud nezařazených pacientů do tříd.

Klasická DA (Ronald Fisher, 1936): zkoumání vztahu mezi skupinou p nezávislých znaků (*diskriminátory*) a jednou kvalitativní závisle proměnnou y, nabývající hodnotu 0 (objekt je v 1. třídě), nebo hodnotu 1 (objekt je ve 2. třídě).

Třídy jsou zřetelně *odlišené*. Každý objekt patří do jedné ze dvou tříd.

Cílem je nalézt znaky, které úspěšně přispívají do procesu *třídění*.

DISKRIMINAČNÍ ANALÝZA

DISKRIMINÁTORY, $j = 1, \dots, m$

OBJEKTY

$i = 1, \dots, n$

i	x_1	x_2	x_3	...	x_m	y
1						
2						
3						
4						
⋮						
n						

TŘÍDA

ANALYZOVANÝ SOUBOR

V datech **analyzovaného výběru** jsou pomocí znaků objekty umístěny do tříd.
V **klasifikovaném výběru** je třeba nalézt *predikční model,* který umístí **nové objekty** do stávajících tříd.

Řešení: v analyzovaném výběru

π_1 značí apriorní pravděpodobnost příslušnosti objektu do první třídy,
$\pi_2 = 1 - \pi_1$ značí pravděpodobnost příslušnosti objektu do druhé třídy.

Pomocí Bayesovy věty pak určíme aposteriorní pravděpodobnost příslušnosti k j-té skupině ($j = 1, 2$)

$$P(G = j \mid \boldsymbol{x}) = \frac{f_j(\boldsymbol{x})\ \pi_j}{\sum_{i=1}^{2} f_i(\boldsymbol{x})\ \pi_i}, \quad j = 1,2$$

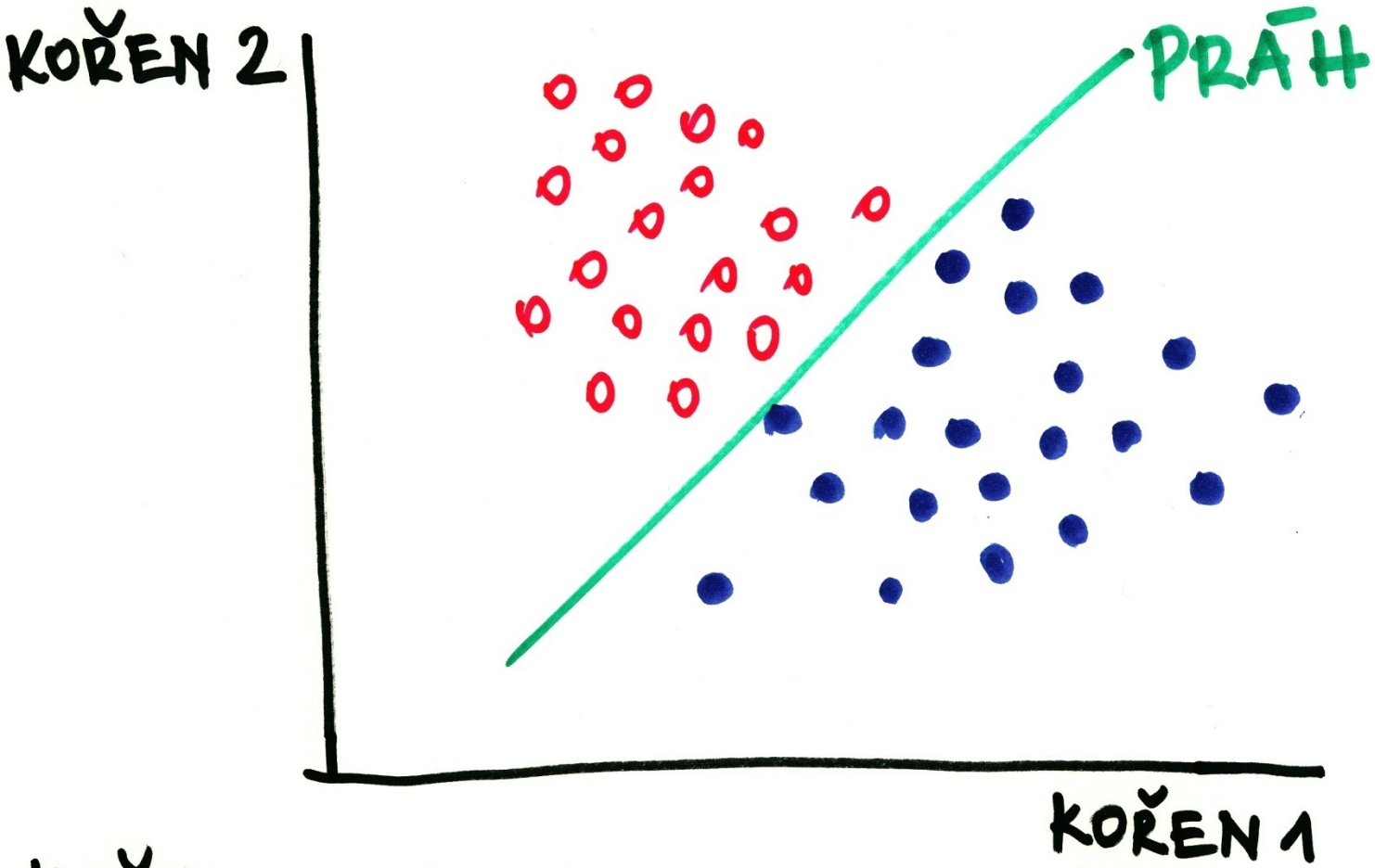
Zápis se dá snadno rozšířit i pro více kategorií.

Třídění:

Objekt zařadíme do první skupiny, pokud $\pi_1 f_1(x) > \pi_2 f_2(x)$.
Pro zařazení objektu do první skupiny musí platit, že

$$f_1(x)/f_2(x) > \pi_2/\pi_1$$

FISHEROVA LINEÁRNÍ DISKRIMINAČNÍ FUNKCE LDA



KOŘEN 1: $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_m x_m$)

KOŘEN 2: $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_m x_m$)

x_i: DISKRIMINÁTORY, ZNAKY

Zařazovací pravidla DA

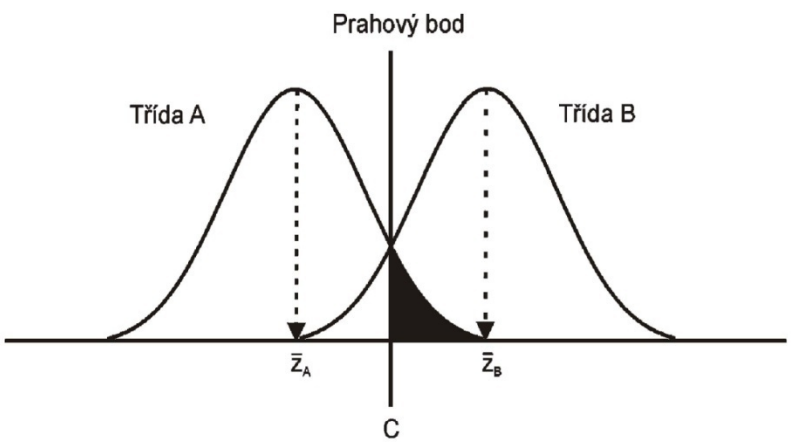
účelem je klasifikace do jedné ze dvou tříd, A nebo B

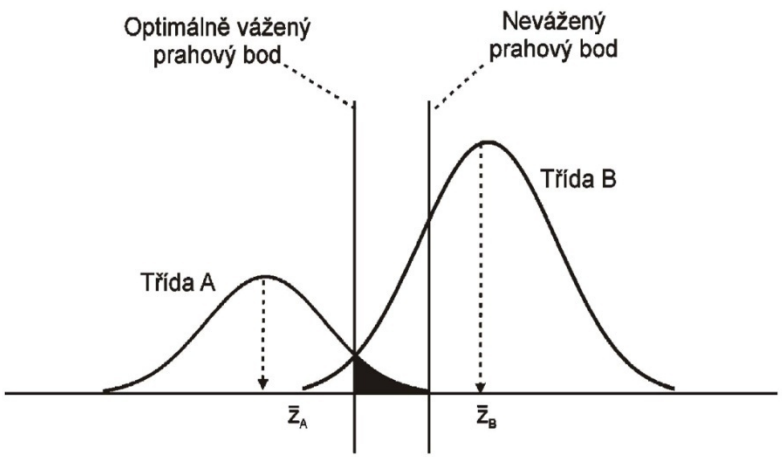
a) Klasifikace na základě jediného znaku *x* (s normálním rozdělením):

1) ve třídě A jde o rozdělení $N(\mu_A, \sigma_A^2)$,
2) ve třídě B jde o rozdělení $N(\mu_B, \sigma_B^2)$,
3) nový objekt má hodnotu jediného znaku x,
4) vybrat třídu, pro kterou je x blíže ke střední hodnotě dané třídy μ_A či μ_B.

b) Určíme *prahový bod* $C = (\mu_A + \mu_B)/2$:

- je-li $x < C$, pak se objekt zařadí do první třídy, A,
- je-li $x \geq C$ pak se objekt zařadí do druhé třídy, B.





1) $\sigma_A^2 = \sigma_B^2$ pravděpodobnost nesprávné klasifikace je šrafovaně. Tato souvisí se vzdáleností obou rozdělení a s jejich rozptyly.
2) $\sigma_A^2 < \sigma_B^2$ je pravděpodobnost nesprávné klasifikace pro třídu A větší než pro třídu B. Bude proto třeba penalizovat C s ohledem na nestejný rozptyl.

1) Pro shodné kovarianční matice $C_A = C_B$ a stejné apriorní pravděpodobnosti zařazení do kategorií $\pi_1 = \pi_2$ bude **dělicí čarou** pro obě třídy přímka $a_1 x_1 + a_2 x_2 = C$ jako *lineární diskriminace LDA*.

2) Pro různé kovarianční matice $C_A \neq C_B$ bude **dělící čára** definovaná polynomem druhého stupně jako *kvadratická diskriminace QDA*.

Zařazování do tříd se takto převede na výpočet hodnoty funkce $Z_i = a_1 x_{1i} + a_2 x_{2i}$ a porovnání Z_i s prahem C.

Lineární LDA a kvadratická QDA diskriminační funkce

Dle typu hustot pravděpodobnosti pro znaky $f_1(\boldsymbol{x})$ a $f_2(\boldsymbol{x})$:

1) Pro normální rozdělení, lišící se jen středními hodnotami tříd je *lineární diskriminační analýzu LDA*.

2) Pro normální rozdělení, lišící se středními hodnotami a kovariančními maticemi tříd je *kvadratickou diskriminační analýzu QDA*.

3) Pro směsi normálních rozdělení je *nelineární diskriminační funkce*.

4) Pro neparametrické hustoty rozdělení znaků ve třídách je *flexibilní diskriminační funkce*.

Pro vícerozměrné Gaussovo rozdělení v i-té třídě má odpovídající hustota pravděpodobnosti tvar

$$f_i(\boldsymbol{x}) = \frac{1}{(2\pi)^{m/2}\ \det(\boldsymbol{C}_i)^{1/2}} \exp\left[-\frac{1}{2}\ (\boldsymbol{x} - \mu_i)^T\ \boldsymbol{C}_i^{-1}\ (\boldsymbol{x} - \mu_i)\right]$$

kde $m = 2$ je počet znaků.

Lineární diskriminační funkce LDA

Pro dvě třídy 1 a 2 vychází LDA z předpokladu, že $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \boldsymbol{C}$.

Logaritmus poměru aposteriorních pravděpodobností je

$$\ln\frac{P(G = 1 \mid x)}{P(G = 2 \mid x)} = \ln\left(\frac{f_1(\boldsymbol{x})}{f_1(\boldsymbol{x})} + \ln\frac{\pi_1}{\pi_2}\right) =$$

$$= \ln\frac{\pi_1}{\pi_2} - 0.5(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T\ \boldsymbol{C}^{-1}\ (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \boldsymbol{x}^T\boldsymbol{C}^{-1}\ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

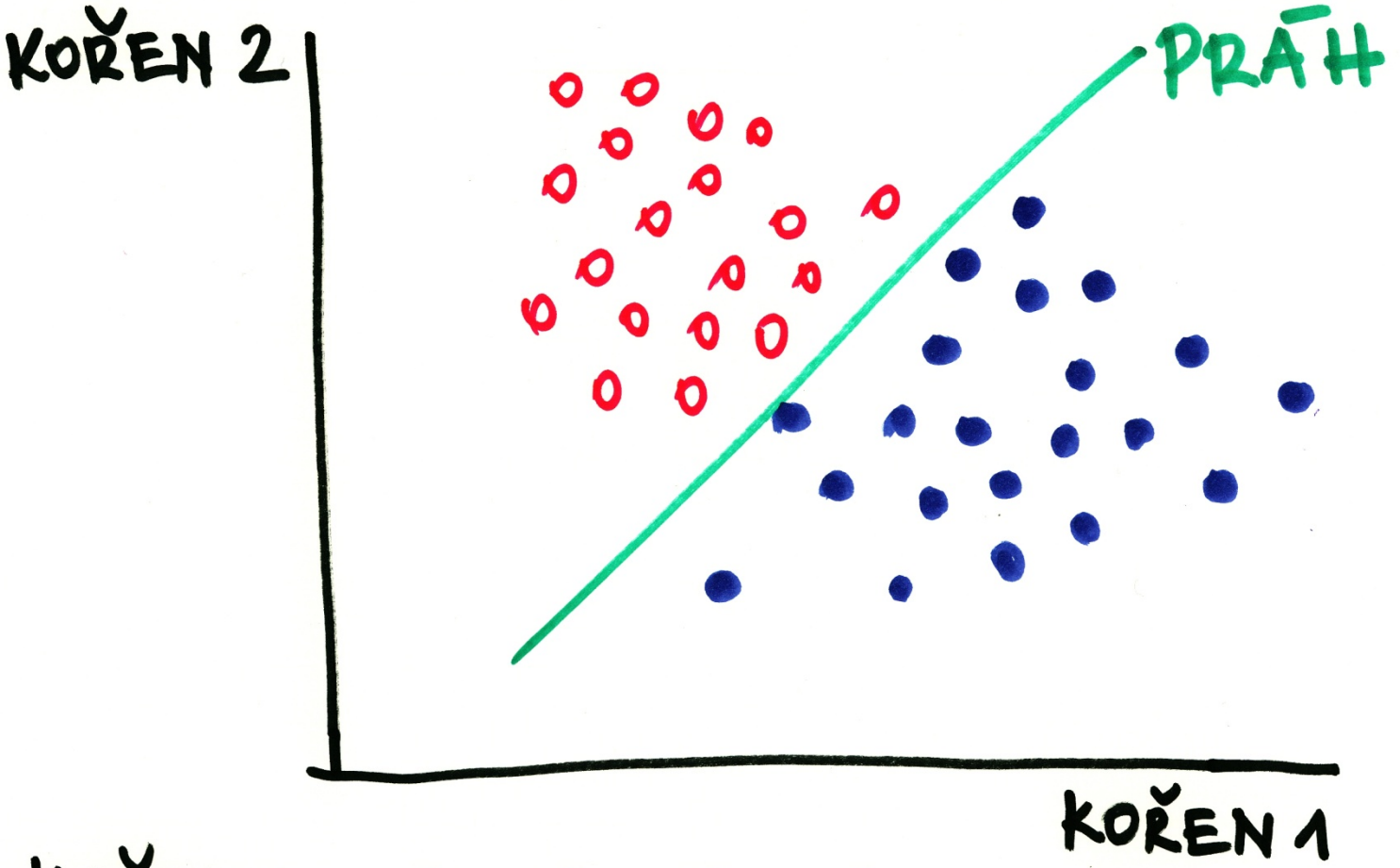
Vzhledem k $\boldsymbol{x} = (x_1, x_2)^T$ jde o lineární funkci a dělicí funkcí bude přímka.

Dělicí funkce je množinou bodů $\boldsymbol{x}$, pro které platí

$$P(G = 1/\boldsymbol{x}) = P(G = 0/\boldsymbol{x}).$$

Pro více znaků jsou dělicí čáry lineárními úseky procházejícími průsečíky jednotlivých elips konstantní hustoty pro jednotlivé třídy.

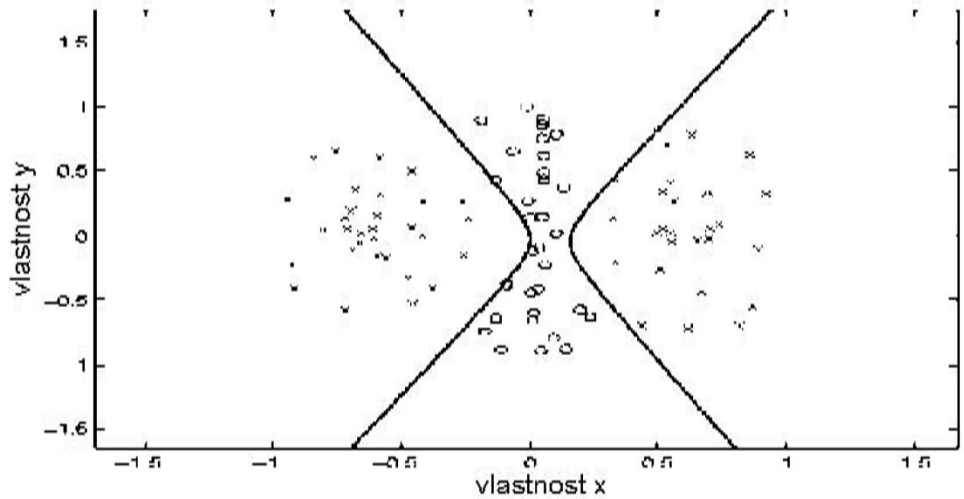
FISHEROVA LINEÁRNÍ DISKRIMINAČNÍ FUNKCE LDA



KOŘEN 1: $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_m x_m$)

KOŘEN 2: $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_m x_m$)

x_i DISKRIMINÁTORY, ZNAKY



LDA pro dva znaky a tři kategorie: kdy rozdíly jsou pouze
ve středních hodnotách $\mu_1 \neq \mu_2$ a shodné rozptyly
$\sigma_1^2 = \sigma_2^2$.

Pravidlo pro zařazení do první třídy je ve tvaru $\boldsymbol{a}^T \boldsymbol{x} + \boldsymbol{b} > 0$, kde $\boldsymbol{a}^T$ je vektor koeficientů u lineárního členu $\boldsymbol{a}^T = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{C}^{-1}$ a absolutní člen je $\boldsymbol{b} = -0.5\, \boldsymbol{a}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) - \ln(\pi_2/\pi_1)$

Lineární diskriminační funkce je $L(\boldsymbol{x}) = \boldsymbol{a}^T \boldsymbol{x}$.

Rozdělení funkce je normální a závisí na Mahalanobisově vzdálenosti mezi středními hodnotami v obou třídách

$$D_M^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{C}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

Pro obě třídy pak platí,

$$E(L(\boldsymbol{x})) = D_M^2/2$$
$$D(L(\boldsymbol{x})) = D_M^2.$$

Dělicí rovina (přímka pro $m = 2$) je určena rovnicí $L(\boldsymbol{x}) + \boldsymbol{b} = \boldsymbol{a}^T \boldsymbol{x} + \boldsymbol{b} = 0$.

Chyba nesprávné klasifikace ω se vyčíslí

$$\omega = \pi_1 \, \varphi\left(\frac{\ln(\pi_1/\pi_2) - D_M^2/2}{D_M}\right) + \pi_2 \, \varphi\left(\frac{-\ln(\pi_2/\pi_1) - D_M^2/2}{D_M}\right)$$

kde $\varphi(x)$ je distribuční funkce normovaného normálního rozdělení.

Pro neinformativní apriorní pravděpodobnosti $\pi_1 = \pi_2 = 0.5$ vyjde jednoduchý vztah pro **chybu nesprávné klasifikace** $\omega = \varphi(-D_M/2)$.
Chyba nesprávné klasifikace do první třídy je stejná jako chyba klasifikace do druhé třídy.

Místo lineární diskriminační funkce je možné použít **lineární diskriminační kritérium**, které nevyžaduje stanovení koeficientů ***a, b.***

Kritérium $Lk_j(\boldsymbol{x})$ se počítá pro každou j-tou třídu zvlášť.

Při klasifikaci objektů se pak objekt zařazuje do l-té třídy, pro kterou vyjde $Lk_l(\boldsymbol{x}) = \max \{Lk_j(\boldsymbol{x})\}$.

Pro LDA má ***lineární diskriminační kritérium*** tvar

$$LK_j(\boldsymbol{x}) = \boldsymbol{x}^{\boldsymbol{T}} C^{-1} \boldsymbol{\mu_j} - 0.5\ \mu_j^T \boldsymbol{C}^{-1} \mu_j + \ln \pi_j$$

Příklad: použití LDA pro dvě třídy $m = 2$. Vychází se ze známých matic:

$\boldsymbol{X}_1$ rozměru $n_1 \times m$ pro třídu 1, $\boldsymbol{X}_2$ rozměru $n_2 \times m$ pro třídu 2.

Jednotlivé objekty v matici $\boldsymbol{X}$ všech dat se zařadí do tříd podle výstupu $\boldsymbol{G}$.

Postup:

1) Vyčíslí se výběrové průměry $\bar{x}_1$ a $\bar{x}_2$ a společná kovarianční matice

$$\boldsymbol{S} = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n_1 + n_2 - 2}$$

2a) Nejjednodušší je předpoklad $\pi_1 = \pi_2 = 0.5$.

2b) Pokud je výběr informativní a byl pořízen jako celek a pak rozdělen do skupin, je možné použít relativních četností

$$\hat{\pi}_1 = n_1/(n_1 + n_2) \quad \text{a} \quad \hat{\pi}_2 = n_2/(n_1 + n_2)$$

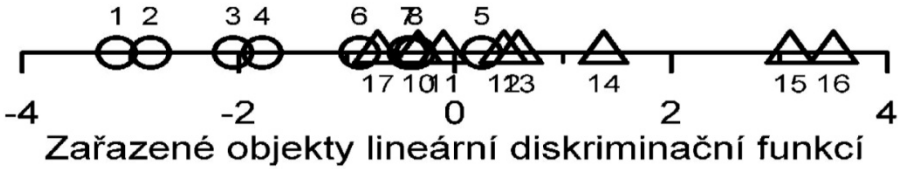
3) Za normality se určí koeficienty Fisherovy lineární diskriminační funkce z odhadů

$$\hat{\boldsymbol{a}} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)\,\boldsymbol{S} \quad \text{a} \quad \hat{\boldsymbol{b}} = -0.5\,\boldsymbol{a}^T(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) - \ln(\hat{\pi}_2/\hat{\pi}_1)$$

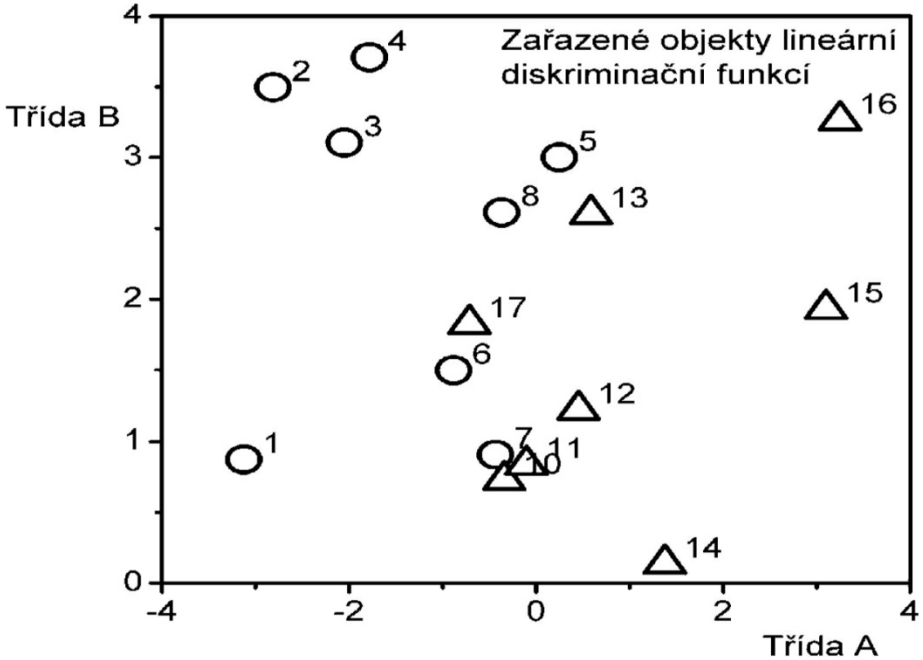
4) Při zařazování nových objektů s hodnotami znaků $\boldsymbol{x}_0$ se použije pravidlo, že do první skupiny je objekt klasifikován pokud platí $\hat{\boldsymbol{a}}^T \boldsymbol{x}_0 + \hat{\boldsymbol{b}} \geq 0$.

5) V opačném případě se klasifikuje do druhé skupiny.

Ukázka 1: zařazování Fisherovou lineární diskriminační funkcí:



Ukázka 2: zařazování Fisherovou lineární diskriminační funkcí:



Příklad 4.22 ***Třídění lebek Tibeťanů lineární diskriminační funkcí***

Databáze lebek na pohřebištích v Tibetu svědčí o dvou skupinách lidí:

prvních 13 bylo nalezeno v hrobech v Sikkimu a okolí,

druhých 15 lebek na bojištích okolo Lhasy.

Předpokládejme, že máme data o 2 třídách tibetských lebek.

Data: *i* index lebky,

x1 největší délka lebky [mm],

x2 největší horizontální šířka lebky [mm],

x3 výška lebky [mm],

x4 výška horní části obličeje [mm],

x5 šířka obličeje mezi body lícních kostí [mm].

x1	***x2***	***x3***	***x4***	***x5***
190.5	152.5	145	73.5	136.5
...	...	...	...	...
...	...	...	...	...
172.5	132	125.5	63	121
167	130	125.5	69.5	119.5
1825	131	135	68.5	136

Řešení:

1) **Analyzovaný výběr**: na trénovacích datech vytvoříme třídy

První třída (13 lebek z hrobů v Sikkimu) vede ke středním hodnotám

[174.82, 139.35, 132.00, 69.82, 130.35]

a kovarianční matici $S_1 = \begin{vmatrix} 45.53 & & & & \\ 25.22 & 57.81 & & & \\ 12.39 & 11.88 & 36.09 & & \\ 22.15 & 7.52 & -0.31 & 20.94 & \\ 27.97 & 48.06 & 1.41 & 16.77 & 66.21 \end{vmatrix}$.

Druhá třída (15 lebek z bojišť v Lhase) vede ke středním hodnotám

[185.73, 138.73, 134.77, 76.47, 137.50]

a kovarianční matici $S_2 = \begin{vmatrix} 74.42 & & & & \\ -9.52 & 37.35 & & & \\ 22.74 & -11.26 & 36.32 & & \\ 17.79 & 0.70 & -10.72 & 15.30 & \\ 11.13 & 9.46 & 7.20 & 8.66 & 17.96 \end{vmatrix}$.

Koeficienty diskriminační funkce $a_1, \ldots, a_5$, jsou vyčísleny podle vztahu

$$\boldsymbol{a} = \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) = [-0.09, 0.16, 0.01, -0.18, -0.18]$$

a vedou k *průměrům obou tříd*: $\bar{Z}_1 = -28.71$ a $\bar{Z}_2 = -32.21$.

Optimální prahový bod C,

dle kterého se budou nezařazené objekty třídit do první nebo druhé třídy, se vyčíslí jako polosuma obou průměrů dle vztahu

$$C = (\bar{Z}_1 + \bar{Z}_2)/2 = (-28.71 + (-32.21))/2 = -30.46.$$

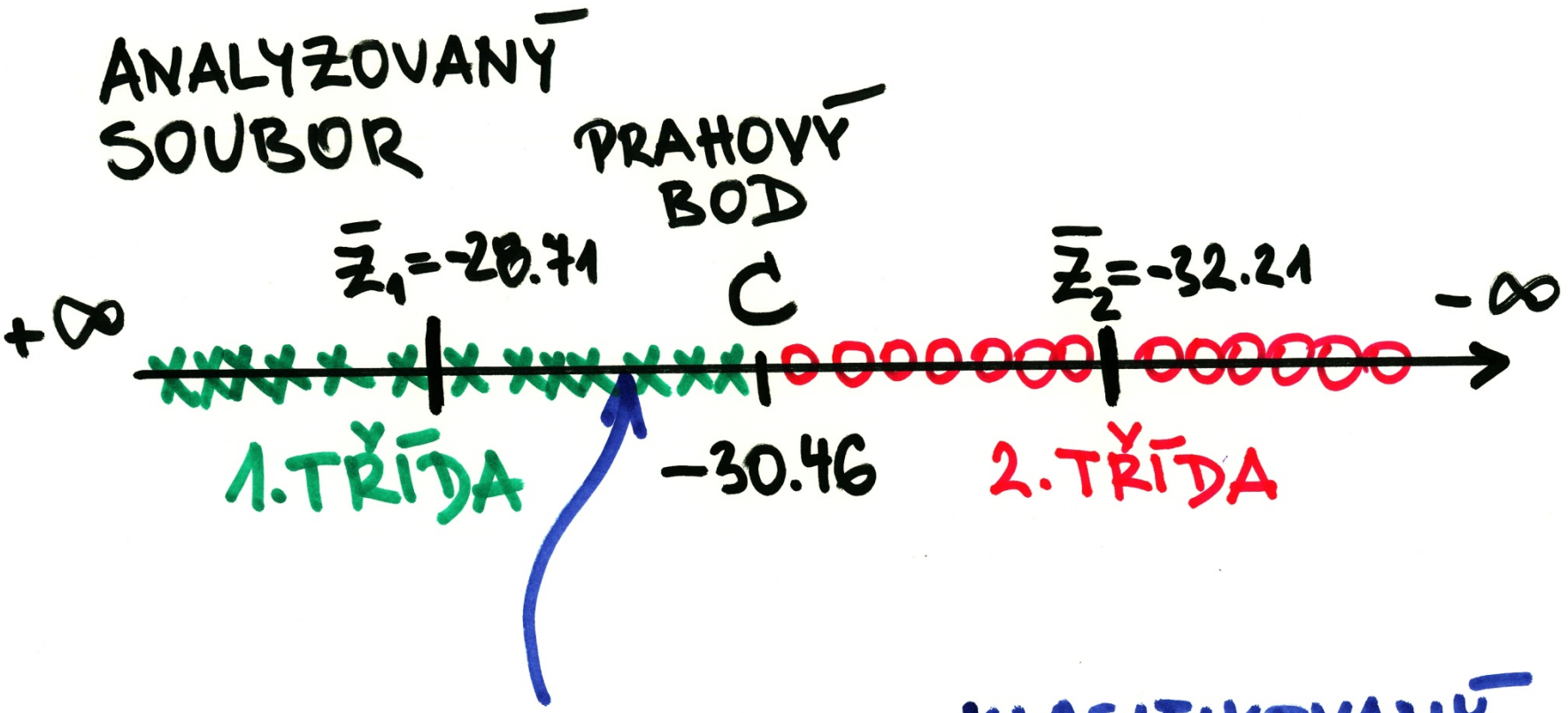
Diskriminace:

2) **Klasifikovaný výběr:** na nových datech provedeme zařazení dosud nezařazených objektů

Vezmeme *data pro lebku prvního Tibeťana* a pokusíme se ji zařadit do 1. nebo do 2. třídy. Vyčísleme proto pro ni hodnotu *lineární diskriminační funkce*

$$Z_1 = -0.09 \times 190.5 + 0.16 \times 152.5 + 0.01 \times 145.0 - 0.18 \times 73.5 - 0.18 \times 136.5 = -29.74,$$

Závěr: Protože lineární diskriminační funkce $Z_1 = -29.74$ je menší než optimální prahový bod $C = -30.46$, *patří lebka prvního Tibeťana do první třídy*.



-29.74

PRVNÍ TIBEŤAN

(KLASIFIKOVANÝ PRVEK, OBJEKT)

KLASIFIKOVANÝ SOUBOR

Kvadratická diskriminační funkce QDA

Pokud nejsou kovarianční matice stejné, vede pravidlo pro zařazení do první skupiny $f_1(x)\,\pi_1 > f_1(x)\,\pi_2$ ke kvadratické nerovnosti

$$\boldsymbol{x}^T \boldsymbol{G} \boldsymbol{x} + \boldsymbol{h}^T \boldsymbol{x} + \boldsymbol{C} > 0$$

kde matice $$\boldsymbol{G} = 0.5\,(\boldsymbol{C}_2^{-1} - \boldsymbol{C}_1^{-1})$$

vektor $$\boldsymbol{h}^T = \boldsymbol{\mu}_1^T \boldsymbol{C}_1^{-1} - \boldsymbol{\mu}_2^T \boldsymbol{C}_2^{-1}$$

a konstanta $$C = 0.5 \ln \frac{\det(\boldsymbol{C}_2)}{\det(\boldsymbol{C}_1)} - 0.5 \left(\boldsymbol{\mu}_1^T \boldsymbol{C}_1^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T \boldsymbol{C}_2^{-1} \boldsymbol{\mu}_2 \right) - \ln(\pi_2/\pi_1)$$

Platí-li pro nové $\boldsymbol{x}_0$ tato kvadratická nerovnost, zařazuje se objekt do skupiny 1 a v opačném případě do skupiny 2.

Lze také definovat **kvadratické diskriminační kritérium**

$$QK_j(\boldsymbol{x}) = -0.5 \ln \det(\boldsymbol{C}_j) - 0.5 (\boldsymbol{x} - \boldsymbol{\mu}_j)^T \boldsymbol{C}_1^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_j) + \ln \pi_j$$

Objekt $\boldsymbol{x}_0$ se pak zařazuje do třídy, které odpovídá maximální hodnota $Qk_j(\boldsymbol{x}_0)$.

Při kvadratické diskriminační analýze se objekty zařazují do tříd podle minima Mahalanobisových vzdáleností od středů tříd $\boldsymbol{\mu}_j$.

Úprava prahového bodu

Volba prahového bodu *C* poskytuje požadovaný poměr apriorních pravděpodobností π_1 a π_2.

Optimální volba prahového bodu C je daná vzorcem $C = \frac{\bar{Z}_1 + \bar{Z}_2}{2} + \ln\frac{\pi_1}{\pi_2}$

a když bude $\pi_1 = \pi_2 = 0.5$, bude prahový bod *C* roven $C = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$

Standardizované koeficienty: hodnoty koeficientů $a_1, a_2, \ldots, a_p$ **nejsou** přímo porovnatelné.

Relativní vliv na každou proměnnou v diskriminační funkci získáme ze *standardizovaných diskriminačních koeficientů*.

Standardizované koeficienty se vypočtou vynásobením koeficientů a_i odpovídající směrodatnou odchylkou s_i.

Volba znaků, diskriminátorů

Znaky musí zajistit přesné zařazení objektů do tříd čili diskriminaci.

Principem selekce znaků je zajištění dostatečné separability tříd a maximalizace některé zvolené míry:

-Začneme se všemi znaky. Pak se a vypouštějí se takové znaky, které provedou nedostatečnou separaci.

-V mnoha situacích může být DA užita jako exploratorní pomůcka.

-Nejprve jsou do dat zahrnuty všechny využitelné znaky. Na začátku není známo, které znaky jsou k zařazení objektů do tříd účinné.

- V diskriminační analýze místo testování změny hodnoty čtverce korelačního koeficientu R^2 přidáním nebo odebráním proměnné testujeme změnu Mahalanobisovy vzdálenosti D_M^2.

-Užívají se stejná testační kritéria jako při výběru nezávisle proměnných v lineární regresní analýze.

-*Krokový výběr znaků* kombinuje jak jejich přidávání, tak i jejich odstraňování.

-První znak, zahrnutý do modelu ve výběrovém kritériu, má největší přijatelnou hodnotu.

-*Vybírání znaků* se ukončí, když žádné další znaky nesplňují zaváděcí nebo odstraňovací kritérium.

Kritéria k vybírání znaků

Wilkovo kritérium λ: když znak poskytuje v diskriminační funkci nejmenší hodnotu Wilkova kritéria λ, je zahrnut do modelu:
- k zavedení nebo odstranění znaku je dovolen jeden krok.
- maximální počet kroků je roven dvojnásobku počtu znaků.

- ***tolerance*** je mírou stupně lineární asociace mezi znaky. Pro *i*-tý znak platí vztah $1 - R_i^2$, kde R_i^2 je čtverec vícenásobného korelačního koeficientu, když je uvažován *i*-tý znak za závisle proměnnou a když je uvažována regresní rovnice mezi tímto *i*-tým znakem a ostatními znaky.

- ***využití tolerance***: malé hodnoty tolerance indikují, že *i*-tý znak je lineární kombinací ostatních znaků. Znaky s tolerancí menší než 0.001 nejsou do modelu zařazeny.

Testování: významnost změny Wilkova kritéria λ po zavedení znaku do modelu nebo odstranění z modelu je založena na testačním kritériu *F*.

Na začátku procesu vybírání znaků: tolerance a minimum tolerance jsou položeny rovny 1, protože v modelu zatím nejsou znaky.

N=143	Výsledky diskriminační funkční analýzy (B23) Počet prom. v modelu: 5; grupovací: B23x6-GROUP (3 skup) Wilk. lambda: ,10863 přibliž F (10,272)=55,328 p<0,0000					
	Wilk. Lambda	Parc. Lambda	F na vyj (2,136)	p-hodn.	Toler.	1-toler. R^2
B23x1-RELWT	0,117426	0,925075	5,50754	0,005012	0,745200	0,254800
B23x2-GLUFAST	0,149708	0,725594	25,71637	0,000000	0,149227	0,850773
B23x3-GLUTEST	0,199867	0,543500	57,11504	0,000000	0,143100	0,856900
B23x4-INSTEST	0,122473	0,886949	8,66731	0,000286	0,813954	0,186046
B23x5-SSPG	0,112112	0,968922	2,18109	0,116853	0,562221	0,437779

Test statistické významnosti dotyčného diskriminátoru

N=150	Výsledky diskriminační funkční analýzy (73Iris) Počet prom. v modelu: 4; grupovací: Trida (3 skup) Wilk. lambda: ,02344 přibliž F (8,288)=199,15 p<0,0000					
	Wilk. Lambda	Parc. Lambda	F na vyj (2,144)	p-hodn.	Toler.	1-toler. R^2
SEPALLENGTH	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007
SEPALWIDTH	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALLENGTH	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
PETALWIDTH	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686

F-test významnosti každého znaku: hodnota F pro změnu Wilkova kritéria λ při přidání znaku do modelu se vyčíslí dle

$$F_{\text{změny}} = \frac{n - g - p}{g - 1} \left(\frac{\dfrac{1 - \lambda_{p+1}}{\lambda_p}}{\dfrac{\lambda_{p+1}}{\lambda_p}} \right)$$

kde n je celkový počet objektů,
g udává počet tříd,
p je počet znaků,
λ_p značí Wilkovo lambda před přidáním a
λ_{p+1} je Wilkovo lambda po přidání znaku do modelu.

Testování: Do modelu je zařazen ten znak, který způsobuje nejmenší hodnotu Wilkova kritéria λ

Raovo *V* kritérium, známé jako ***Lawley-Hotellingova stopa***, je

$$V = (n - g) \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij} \sum_{k=1}^{g} (\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j)$$

kde p udává počet znaků v modelu,

g značí počet tříd,

n_k je velikost k-té třídy,

$\bar{x}_{ik}$ je střední hodnota i-tého znaku v k-té třídě,

$\bar{x}$ je průměr i-tého znaku pro všechny třídy kombinované dohromady a

w^*_{ij} představuje prvek inverzní kovarianční matice mezi třídami $\boldsymbol{B}^{-1}$.

Testování: čím větší jsou rozdíly mezi středními hodnotami (průměry) tříd, tím větší je hodnota Raova V.

Může se však stát, že znak po zařazení do modelu sníží hodnotu Raova V.

Mahalanobisova vzdálenost $D_{1,2}^2$ je zobecněná míra vzdálenosti mezi dvěma třídami 1 a 2 definovaná vztahem

$$D_{1,2}^2 = (n - g) \sum_{i=1}^{m} \sum_{j=1}^{m} w_{ij} (\bar{x}_{i1} - \bar{x}_{i2})(\bar{x}_{j1} - \bar{x}_{j2})$$

kde m udává počet znaků v modelu,

$\bar{x}_{i1}$ je průměr i-tého znaku ve třídě 1,

w_{ij} je prvek inverzní kovarianční matice $\boldsymbol{B}^{-1}$.

Testování: kritérium všech párů tříd vyčísleno je jako první. Znak, který měl pro dvě od začátku nejtěsnější třídy, nejmenší hodnotu $D_{1,2}^2$, je zařazen do modelu.

***F*-testační kritérium mezi třídami** testuje nulovou hypotézu H_0: *dva vektory středních hodnot tříd objektů jsou stejné.*

Kritérium je postaveno na Mahalanobisově vzdálenosti $D_{1,2}^2$.

Testační kritérium *F* je definováno vztahem

$$F = \frac{(n - 1 - p)\, n_1\, n_2}{p(n - 2)\,(n_1 + n_2)}\, D_{1,2}^2$$

a může být použito k výběru znaků.

Testování: V každém kroku je zařazen do modelu ten znak, který vykazuje největší hodnotu kritéria F.

Po zavedení dalších znaků do modelu jsou sledovány změny hodnoty testačního kritéria $F_{\text{změny}}$.

Proces přidávání znaků do modelu buď pokračuje, nebo je zastaven terminačním kritériem.

F-test vyšetření přidaného 1 znaku:

Testační kritérium je

$$F = \frac{(n_1 \; n_2)(n_1 + n_2 - p - 2)(D^2_{p+1} - D^2_p)}{(n_1 + n_2)(n_1 + n_2 - 2) + (n_1 \; n_2) \; D^2_p}$$

které má Fisherovo-Snedecorovo rezdělení s 1 a $(n_1 + n_2 + - p - 2)$ stupni volnosti.

Testování: Vyšetřujeme, zda přidaný znak x_{p+1} zlepší zařazení objektů o znacích $x_1, x_2, ..., x_p$ významně zvětší hodnotu Mahalanobisovy vzdálenosti $D_{p+1}{}^2$, čili H_0: $D_{p+1}{}^2 = D_p{}^2$.

F-test o přidání *p* znaků současně:

Testační kritérium je

$$F = \frac{(n_1 \; n_2)(n_1 + n_2 - p - q - 1)(D^2_{p+q} - D^2_p)}{q \; (n_1 + n_2)(n_1 + n_2 - 2) + (n_1 \; n_2) \; D^2_p}$$

má *F*-rozdělení s q a $(n_1 + n_2 + - p - q - 2)$ stupni volnosti.

Testování: vyšetřujeme, zda další znaky $x_{p+1}, x_{p+2}, ..., x_{p+q}$ zlepší predikci, H_0: $D^2{}_{p+q} = D^2{}_p$.

Procento správně zařazených objektů

je jedním z indikátorů efektivnosti diskriminační funkce.

Aktuální diskriminační skóre ve třídách: "Dobrá" diskriminační funkce je taková, která má hodně mezitřídní proměnlivosti vzhledem k proměnlivostem přes všechny ostatní třídy čili koeficienty diskriminačních funkcí jsou voleny tak, že poměr sumy čtverců mezi třídami a součtem čtverců přes třídy je co možná největší.

Test H_0: *V souborech není rozdílu mezi třídními průměry* je založen na Wilkově kritériu λ.

Pro dvě třídy představuje **Wilkovo kritérium λ** poměr sumy čtverců mezi třídami a totální sumy čtverců čili podíl celkového rozptylu v diskriminač-ním skóre, který je nevysvětlen rozdíly mezi třídami.

Testování:

a) *Malé hodnoty* λ jsou spojeny s diskriminačními funkcemi, které mají hodně proměnlivosti mezi třídami a málo proměnlivosti přes třídy.

b) *Hodnota* $\lambda = 1$ se objevuje, když průměr diskriminačního skóre je stejný ve všech třídách a když není žádná mezitřídní proměnlivost.

Veličina λ je transformována do proměnné, která má χ^2-rozdělení. Poskytuje pouze test, že průměry tříd jsou stejné. Malé rozdíly mohou být statisticky významné, ale nedovolují dobrou diskriminaci mezi třídami.

Jsou-li průměry a kovariance matic shodné, diskriminace není možná.

Nesprávně zařazené objekty jsou zde v učících datech označeny *

Případ	Klasifikace případů (73Iris) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	1 p=,33333	2 p=,33333	3 p=,33333
1	SETOSA	SETOSA	VERSICOL	VIRGINIC
2	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
3	VERSICOL	VERSICOL	VIRGINIC	SETOSA
4	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
* 5	VIRGINIC	VERSICOL	VIRGINIC	SETOSA
6	SETOSA	SETOSA	VERSICOL	VIRGINIC
7	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
8	VERSICOL	VERSICOL	VIRGINIC	SETOSA
* 9	VERSICOL	VIRGINIC	VERSICOL	SETOSA
10	SETOSA	SETOSA	VERSICOL	VIRGINIC
11	VERSICOL	VERSICOL	VIRGINIC	SETOSA
* 12	VERSICOL	VIRGINIC	VERSICOL	SETOSA
13	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
14	VERSICOL	VERSICOL	VIRGINIC	SETOSA
15	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
16	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
17	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
18	SETOSA	SETOSA	VERSICOL	VIRGINIC
19	VERSICOL	VERSICOL	VIRGINIC	SETOSA
20	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
21	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
22	VERSICOL	VERSICOL	VIRGINIC	SETOSA
23	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
24	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
25	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
26	SETOSA	SETOSA	VERSICOL	VIRGINIC
27	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
28	VERSICOL	VERSICOL	VIRGINIC	SETOSA
29	VERSICOL	VERSICOL	VIRGINIC	SETOSA
30	VERSICOL	VERSICOL	VIRGINIC	SETOSA
31	SETOSA	SETOSA	VERSICOL	VIRGINIC
32	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
33	VERSICOL	VERSICOL	VIRGINIC	SETOSA
34	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
35	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
36	SETOSA	SETOSA	VERSICOL	VIRGINIC
37	SETOSA	SETOSA	VERSICOL	VIRGINIC
38	VERSICOL	VERSICOL	VIRGINIC	SETOSA
39	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
40	SETOSA	SETOSA	VERSICOL	VIRGINIC
41	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
42	SETOSA	SETOSA	VERSICOL	VIRGINIC
43	VERSICOL	VERSICOL	VIRGINIC	SETOSA
44	SETOSA	SETOSA	VERSICOL	VIRGINIC
45	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
46	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
47	SETOSA	SETOSA	VERSICOL	VIRGINIC
48	VERSICOL	VERSICOL	VIRGINIC	SETOSA
49	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
50	SETOSA	SETOSA	VERSICOL	VIRGINIC

Případ	Klasifikace případů (73Iris) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	1 p=,33333	2 p=,33333	3 p=,33333
51	SETOSA	SETOSA	VERSICOL	VIRGINIC
52	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
53	SETOSA	SETOSA	VERSICOL	VIRGINIC
54	SETOSA	SETOSA	VERSICOL	VIRGINIC
55	SETOSA	SETOSA	VERSICOL	VIRGINIC
56	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
57	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
58	SETOSA	SETOSA	VERSICOL	VIRGINIC
59	SETOSA	SETOSA	VERSICOL	VIRGINIC
60	SETOSA	SETOSA	VERSICOL	VIRGINIC
61	VERSICOL	VERSICOL	VIRGINIC	SETOSA
62	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
63	SETOSA	SETOSA	VERSICOL	VIRGINIC
64	SETOSA	SETOSA	VERSICOL	VIRGINIC
65	VERSICOL	VERSICOL	VIRGINIC	SETOSA
66	VERSICOL	VERSICOL	VIRGINIC	SETOSA
67	SETOSA	SETOSA	VERSICOL	VIRGINIC
68	SETOSA	SETOSA	VERSICOL	VIRGINIC
69	VERSICOL	VERSICOL	VIRGINIC	SETOSA
70	VERSICOL	VERSICOL	VIRGINIC	SETOSA
71	SETOSA	SETOSA	VERSICOL	VIRGINIC
72	SETOSA	SETOSA	VERSICOL	VIRGINIC
73	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
74	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
75	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
76	VERSICOL	VERSICOL	VIRGINIC	SETOSA
77	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
78	SETOSA	SETOSA	VERSICOL	VIRGINIC
79	SETOSA	SETOSA	VERSICOL	VIRGINIC
80	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
81	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
82	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
83	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
84	VERSICOL	VERSICOL	VIRGINIC	SETOSA
85	VERSICOL	VERSICOL	VIRGINIC	SETOSA
86	VERSICOL	VERSICOL	VIRGINIC	SETOSA
87	SETOSA	SETOSA	VERSICOL	VIRGINIC
88	SETOSA	SETOSA	VERSICOL	VIRGINIC
89	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
90	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
91	SETOSA	SETOSA	VERSICOL	VIRGINIC
92	VERSICOL	VERSICOL	VIRGINIC	SETOSA
93	VERSICOL	VERSICOL	VIRGINIC	SETOSA
94	VERSICOL	VERSICOL	VIRGINIC	SETOSA
95	SETOSA	SETOSA	VERSICOL	VIRGINIC
96	SETOSA	SETOSA	VERSICOL	VIRGINIC
97	VERSICOL	VERSICOL	VIRGINIC	SETOSA
98	VERSICOL	VERSICOL	VIRGINIC	SETOSA
99	VERSICOL	VERSICOL	VIRGINIC	SETOSA
100	SETOSA	SETOSA	VERSICOL	VIRGINIC

Příklad 4.23 *Logistická diskriminace k určení rakoviny prostaty*

Rakovina lymfatických uzlin. Rozhodující metodou vyšetření je laparotomie, vyjádřená proměnnou *B412x6,* kde 0 značí nepřítomnost rakoviny a 1 přítomnost rakoviny. Brownovo vyšetření pěti diskriminantů u 53 pacientů mělo nahradit toto obtížnější vyšetření:

Data:

B412x1 značí věk pacienta,
B412x2 je hladina sérové kyselé fosfatázy v Kingových-Armstrongových jedn.,
B412x3 rentgenového vyšetření: 0 značí nepřítomnost a 1 přítomnost tumoru,
B412x4 velikost tumoru rektálně: 0 = malý, 1 = velký,
B412x5 závěr patologického bodování z biopsie: 0 méně vážný, 1 velmi vážný.

B412x1	***B412x2***	***B412x3***	***B412x4***	***B412x5***	***B412x6***
66	0.48	0	0	0	0
..	..	..	..	..	..
..	..	..	..	..	..
64	0.89	1	1	0	1
68	1.26	1	1	1	1

Řešení:

1) **Analyzovaný výběr**: odhady parametrů (směrodat. odchylky v závorce) jsou b_0 1.52 (3.56), b_1 0.10 (0.06), b_2 2.64 (1.33), b_3 1.68 (0.80), b_4 2.04 (0.83), b_5 0.35 (0.80).

Dvě třídy: 1. třída (rakovinu má), 2. třída (rakovinu nemá).

Pacient rakovinu lymfatických uzlin má (je zařazen do 1. třídy), je-li splněna nerovnost

$$1.52 - 0.10\, x_1 + 2.64\, x_2 + 1.68\, x_3 + 2.04\, x_4 + 0.35\, x_5 > 0.$$

vykazuje-li pacient **kladnou hodnotu** v této nerovnosti, má rakovinu uzlin a je zařazen do 1. třídy, zatímco vykazuje-li **zápornou hodnotu** v této nerovnosti, nemá rakovinu a je zařazen do 2. třídy.

2) **Klasifikovaný výběr:** dosadíme hodnoty prvního pacienta do této nerovnosti a vyčíslíme

$$1.52 - 0.10 \times 66 + 2.64 \times 0.48 + 1.68 \times 0 + 2.04 \times 0 + 0.35 \times 0 = \mathbf{-3.81.}$$

Závěr:

Protože **výsledek v nerovnosti -3.81 je záporný**, je pacient zařazen do 2. třídy a **nemá rakovinu** lymfatických uzlin.

Průběh diagnostikování DA

Úlohy diskriminační analýzy se řeší v šesti krocích:

1. krok: Cíle diskriminační analýzy

Zajímáme se o správné zařazování objektů do tříd:

a) Vyšetřujeme, zda existují statisticky významné rozdíly mezi profily průměrného skóre znaků pro více předem definovaných tříd?

b) Hledáme, který ze znaků se projevuje nejvíce v rozdílových profilech průměrného skóre dvou či více tříd?

c) Na základě skóre znaků stanovíme postup k zařazování objektů do tříd.

d) Určíme počet a složení nalezených tříd znaků.

DA pracuje nejlépe, když jde o jedinou závisle proměnnou a několik metrických nezávislých proměnných či znaků.

DA poskytuje ve znacích objektivní vyčíslení rozdílů mezi třídami.

DA je dosti podobná vícerozměrné analýze rozptylu, umožňuje pohled do jednotlivých znaků a definuje rozměry diskriminace mezi třídami.

DA poskytuje podklad k zařazování objektů do tříd.
DA určuje diskriminační funkce k zařazování nových objektů do předem nadefinovaných tříd.

2. krok: Formulace úlohy, volba znaků

1) Volby závisle proměnné a několika nezávislých proměnných, znaků:

Závisle proměnná: je většinou nemetrická, kategorická proměnná. Každý objekt může být umístěn pouze do jediné třídy. Obvykle se závisle proměnná týká dvou tříd - dichotomie, např. *dobrý* versus *špatný,* nebo multichotomie, např. *lékař, řidič, učitel.*

Znaky: jsou znaky, metrické proměnné.

2) Určení velikosti výběru k odhadu diskriminační funkce:

DA je značně citlivá *na poměr velikosti výběru ku počtu znaků.*

Empirická úmluva: poměr 20 objektů na jeden znak.

Minimální velikost výběru: 5 objektů na jeden znak.

Rozdílná velikost tříd: způsobí problémy při odhadování diskriminační funkce a zařazování objektů, větší třídy mají větší šanci při zařazování.

3) Dělení výběru ke klasifikačním účelům.

Výběr je často rozdělen (v poměru 50/50 nebo 60/40 nebo 75/25) na:

Analyzovaný výběr k výstavbě diskriminační funkce.

Klasifikovaný výběr k testování diskriminační funkce.

Kategorizované třídy jsou nestejné, pak i velikosti vybraných tříd pro klasifikovaný výběr měly být úměrné jejich podílu v celkovém výběru:

např. výběr 50 mužů a 50 žen, pak bude klasifikovaný výběr 25 mužů a 25 žen,
výběr 70 žen a 30 mužů, pak bude klasifikovaný výběr 35 žen a 15 mužů.

4) Chybí-li u některého objektu nějaký znak:

DA se omezí na malý počet úplně definovaných objektů, což může způsobit v DA určité problémy.

3. krok: Předpoklady diskriminační analýzy

- Předpokladem DA je *vícerozměrná normalita znaků,* jinak se užije alternativně *logistická regrese.*
- *Nestejné* kovarianční matice mohou negativně ovlivnit klasifikační proces LDA.
- Příliš malý výběr a nestejné kovarianční matice nepříznivě ovlivní statistickou významnost odhadů.
- *Multikolinearita* značí velkou korelaci dvou znaků. To je nevhodné při užití krokové metody.
- Klasická DA vychází z předpokladu, že *všechny vztahy jsou lineárního charakteru.*
- *Odlehlé hodnoty* mají rovněž špatný dopad na klasifikační schopnost DA.

4.krok: Nalezené řešení a dosažená těsnost proložení

Dělení souboru: objekty souboru náhodně rozdělíme do dvou výběrů:

1. ***Analyzovaný výběr*** se použije k vyčíslení diskriminační funkce na základě známých a dostatečně popsaných objektů.
2. ***Klasifikovaný výběr*** se použije ke klasifikační matici a zařazování nových objektů.

Predikční schopnost diskriminační funkce vyjadřuje počet správně klasifikovaných objektů.

Diagnostiky každého objektu mohou ukázat na správnost zařazení.

Odvození diskriminační funkce: *přímá metoda* a *kroková metoda.*

Přímá metoda: diskriminační funkce je vyčíslena na základě všech znaků navzdory jejich rozličné diskriminační síle.

Kritéria posouzení statistické významnosti diskriminace

Wilkovo kritérium λ je poměr sumy čtverců mezi třídami a celkového součtu čtverců:

Velké $\lambda = 1$ ukazuje, že průměry tříd jsou shodné.

Malé $\lambda = 0$ ukazuje, že průměry tříd jsou velmi rozdílné a většinu variability lze přičíst rozdílům mezi průměry tříd.

Raovo největší vlastní číslo vyhodnocuje pouze první diskriminační funkci. Mahalanobisova D^2 a Raovo V jsou nejvhodnějšími mírami k posouzení obecné vzdálenosti.

Mahalanobisova vzdálenost D^2 je však nespolehlivá, když počet znaků roste, protože neprovádí žádnou redukci dimenze.

Kritéria těsnosti proložení

a) **Diskriminační Z-skóre pro každý objekt:** vyčíslí se *Z*-skóre nebo-li Fisherova lineární diskriminační funkce pro *k*-tý objekt $L(x_k)$.

Vlastnosti Z-skóre:

1) *Z*-skóre je metrická proměnná a nabízí přímé průměry k porovnání objektů.

2) Objekty s podobnými *Z*-skóre jsou podobné ve znacích, které tvoří tuto funkci, na rozdíl od objektů s nepodobnými skóre.

3) Standardizovaná verze diskriminační funkce je užitečnější k interpretaci.

4) Vysvětlení koeficientů je podobné jako ve vícenásobné lineární regresi.

5) Znaky s velkými koeficienty přispívají více do celkové diskriminační funkce.

Případ	Nestandardizovaná kanonická skóre (73Iris)		
	Skup.	Kořen1	Kořen2
1	SETOSA	7,67197	-0,13489
2	VIRGINIC	-6,80015	0,58090
3	VERSICOL	-2,54868	-0,47220
4	VIRGINIC	-6,65309	1,80532
5	VIRGINIC	-3,81516	-0,94299
6	SETOSA	7,21262	0,35584
7	VIRGINIC	-5,10556	1,99218
8	VERSICOL	-3,49805	-1,68496
9	VERSICOL	-3,71590	1,04451
10	SETOSA	8,68104	0,87759
11	VERSICOL	-2,29249	-0,33286
12	VERSICOL	-4,49847	-0,88275
13	VIRGINIC	-4,96774	0,82114
14	VERSICOL	-1,09043	-1,62658
15	VIRGINIC	-5,06601	-0,02627
16	VIRGINIC	-5,50748	-0,03581
17	VIRGINIC	-6,79602	1,46069
18	SETOSA	6,25140	0,43970
19	VERSICOL	-2,42997	-0,96613
20	VIRGINIC	-5,88615	2,34509
21	VIRGINIC	-6,60882	1,75164
22	VERSICOL	-2,44848	0,79596
23	VIRGINIC	-6,84736	2,42895
24	VIRGINIC	-7,41817	-0,17312
25	VIRGINIC	-4,67800	-0,49910
26	SETOSA	8,61367	0,40325
27	VIRGINIC	-5,64500	1,67772
28	VERSICOL	-1,45928	0,02854
29	VERSICOL	-1,79771	0,48439
30	VERSICOL	-0,99761	-0,49053
31	SETOSA	6,75895	-0,75900
32	VIRGINIC	-4,68315	0,33203
33	VERSICOL	-1,10669	-1,75225
34	VIRGINIC	-5,17956	-0,36348
35	VIRGINIC	-5,80704	2,01020
36	SETOSA	7,99129	0,08638
37	SETOSA	8,33042	0,22813
38	VERSICOL	-2,93398	0,02738
39	VIRGINIC	-5,22049	1,46820
40	SETOSA	7,24101	-0,27262
41	VIRGINIC	-5,72318	1,29328
42	SETOSA	6,41444	1,24730
43	VERSICOL	-1,27227	-1,21458
44	SETOSA	6,93412	-0,70552
45	VIRGINIC	-7,58120	-0,98072
46	VIRGINIC	-4,37150	-0,12130
47	SETOSA	6,82946	-0,54496
48	VERSICOL	-2,40157	-1,59458
49	VIRGINIC	-5,27916	-0,04246
50	SETOSA	6,76470	-0,50515
51	SETOSA	8,08190	0,76339
52	VIRGINIC	-4,07704	0,52324

Případ	Nestandardizovaná kanonická skóre (73Iris)		
	Skup.	Kořen1	Kořen2
53	SETOSA	6,55893	-0,38922
54	SETOSA	6,77138	-0,97063
55	SETOSA	7,95876	-0,16496
56	VIRGINIC	-5,10749	-2,13059
57	VIRGINIC	-6,51910	0,29698
58	SETOSA	7,57247	-0,80546
59	SETOSA	9,84984	1,58594
60	SETOSA	6,85944	1,05165
61	VERSICOL	-1,20117	0,08444
62	VIRGINIC	-4,08087	0,18594
63	SETOSA	7,12869	-0,78666
64	SETOSA	8,06180	0,30042
65	VERSICOL	-2,76810	0,03220
66	VERSICOL	-0,77685	-1,65916
67	SETOSA	6,82308	0,46301
68	SETOSA	7,18677	-0,36099
69	VERSICOL	-2,58980	-0,17461
70	VERSICOL	0,30749	-1,31887
71	SETOSA	9,15824	2,73760
72	SETOSA	8,13231	0,51446
73	VIRGINIC	-6,79672	0,86309
74	VIRGINIC	-6,52450	2,44504
75	VIRGINIC	-5,50748	-0,03581
76	VERSICOL	-1,61616	-0,47010
77	VIRGINIC	-4,58372	-0,85682
78	SETOSA	7,37495	0,56584
79	SETOSA	9,12635	1,22443
80	VIRGINIC	-6,29201	0,46718
81	VIRGINIC	-4,99550	0,18777
82	VIRGINIC	-3,93985	0,61402
83	VIRGINIC	-5,60546	-0,34074
84	VERSICOL	-0,22267	-1,58467
85	VERSICOL	-2,00599	-0,90542
86	VERSICOL	-1,18166	-0,53757
87	SETOSA	9,46768	1,82523
88	SETOSA	7,06201	-0,66340
89	VIRGINIC	-9,17147	-0,74826
90	VIRGINIC	-4,76454	-2,15574
91	SETOSA	7,70195	1,46172
92	VERSICOL	-1,75020	-0,82118
93	VERSICOL	-1,95842	-0,35156
94	VERSICOL	-2,10361	1,19157
95	SETOSA	7,60529	-0,01163
96	SETOSA	6,56055	-1,01516
97	VERSICOL	-1,19376	-2,63446
98	VERSICOL	-0,60552	-1,94298
99	VERSICOL	-0,89870	-0,90494
100	SETOSA	7,48983	-0,26538
101	SETOSA	6,81320	-0,67063
102	VIRGINIC	-6,27284	1,64948
103	VERSICOL	-1,42159	-0,55124
104	VIRGINIC	-6,22824	-0,71272

Případ	Nestandardizovaná kanonická skóre (73Iris)		
	Skup.	Kořen1	Kořen2
105	VERSICOL	-1,85893	0,31901
106	SETOSA	7,68823	-0,00922
107	SETOSA	7,91794	0,67512
108	VIRGINIC	-5,36071	0,64612
109	VERSICOL	-1,15809	-2,64341
110	VIRGINIC	-6,31693	-0,96898
111	VIRGINIC	-6,32774	-1,38329
112	SETOSA	7,34306	-0,94732
113	VERSICOL	-2,14258	0,08878
114	VERSICOL	-2,47946	-1,94074
115	SETOSA	8,39739	0,64736
116	VERSICOL	-1,32553	-0,16287
117	VERSICOL	-3,83526	-1,40596
118	VERSICOL	-2,25741	-1,42679
119	VERSICOL	-1,25571	-0,54642
120	VERSICOL	0,47597	-0,79991
121	VERSICOL	-1,54948	-0,59336
122	VIRGINIC	-6,85056	0,82983
123	VIRGINIC	-5,20383	1,14477
124	SETOSA	8,58243	1,83449
125	SETOSA	7,78075	0,58434
126	VIRGINIC	-6,85281	2,71759
127	VIRGINIC	-4,44073	1,34724
128	VERSICOL	-2,66606	-0,64250
129	VERSICOL	-0,37837	0,08664
130	VERSICOL	-2,41695	-0,09278
131	VIRGINIC	-5,45010	-0,20774
132	VIRGINIC	-5,66034	0,83271
133	VERSICOL	-1,95558	-1,15435
134	SETOSA	7,21930	-0,10965
135	SETOSA	7,32680	-1,07299
136	SETOSA	5,66188	-1,93436
137	VIRGINIC	-5,95824	-0,09402
138	SETOSA	8,07836	0,96858
139	SETOSA	8,02097	1,14050
140	VERSICOL	-2,26247	-1,58725
141	VERSICOL	-1,43756	-0,13442
142	VERSICOL	-2,45906	-0,93528
143	SETOSA	7,49680	-0,18838
144	SETOSA	7,58648	1,20797
145	SETOSA	7,92462	0,20964
146	VIRGINIC	-6,75926	1,60023
147	VERSICOL	-3,51848	0,16059
148	VIRGINIC	-7,83947	2,13973
149	SETOSA	8,31445	0,64495
150	VERSICOL	-0,29318	-1,79872

b) **Vyhodnocení třídních rozdílů v diskriminačních Z-skóre:** určení rozdílu mezi objekty každé třídy v termínech diskriminačního *Z*-skóre.

Souhrnou mírou třídních rozdílů: je porovnání těžiště třídy čili průměru diskriminačního *Z*-skóre pro všechny objekty ve třídě.

Mírou úspěchu diskriminační analýzy: je schopnost definovat diskriminační funkce, které vedou k významně rozličným třídním těžištím v jednotkách Mahalanobisovy vzdálenosti D^2.

Rozdíly mezi třídami na každé diskriminační funkci však vizuální inspekce neodhalí.

c) Přiřazení predikční schopnosti tříd:

Závisle proměnná je nemetrická a proto není možné k vyhodnocení predikční schopnosti použít míru R^2.

U diskriminační analýzy je procento správně klasifikovaných objektů označeno pojmem *hit poměr*, který je analogií vůči regresnímu R^2.

	Průměry kan. proměnných (73Iris)	
Skup.	Kořen1	Kořen2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

d) Určení prahového bodu:

Při konstrukci klasifikačních matic je třeba určit optimální prahový bod *C*.

1) **Pro třídy stejné velikosti**: optimální prahový bod *C* diskriminace bude $C = \frac{\bar{Z}_1 + \bar{Z}_2}{2}$,

kde $\bar{Z}_1$ je těžiště třídy 1 a Z_2 je těžiště třídy 2.

Skup.	Průměry kan. proměnných (73Iris)	
	Kořen1	Kořen2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

2) **Pro nestejné třídní velikosti:** je *vážený průměr třídních těžišť* optimálním prahovým bodem

$C = \frac{n_1 \bar{Z}_1 + n_2 \bar{Z}_2}{n_1 + n_2}$, kde n_1 je počet prvků ve třídě 1, n_2 je počet prvků ve třídě 2.

Diskriminační skóre Z_n pro klasifikované objekty jsou porovnávána s prahovým bodem *C* dle schématu

prvek bude zařazen do třídy 1, když je $Z_n < C$, nebo **bude zařazen do třídy 2, když je $Z_n > C$**.

Test: *Studentův t-test klasifikační správnosti* podle testačního kritéria $t = \frac{p - 0.5}{\sqrt{\frac{0.5\,(1 - 0.5)}{n}}}$,

kde *p* je podíl správně zařazených prvků a *n* je velikost výběru.

e) Měření predikční schopnosti diskriminační funkce:

Hit poměr odhalí, jak dobře diskriminační funkce zařazuje objekty. Nejprve se určuje podíl správně klasifikovaných pravděpodobností:

a) *Stejné třídní velikosti:* klasifikační apriorní pravděpodobnost π_i je rovna reciproké hodnotě počtu tříd k, $C_{Pr} = 1/k$.
Například: pro dvojtřídní funkci $k = 2$ je pravděpodobnost $C_{Pr} = 0.50$, pro trojtřídní $k = 3$ pak $C_{Pr} = 0.33$, atd.

b) *Nestejné třídní velikosti:* klasifikační apriorní pravděpodobnost π_i, je založena na velikosti největší třídy a nazývá se *kritérium maximální pravděpodobnosti*. Určuje se vyčíslením procenta celkového výběru, tvořeného největším ze dvou (či více) tříd.
Například: velikosti dvou tříd jsou 65 a 35, a kritérium maximální pravděpodobnost bude $C_{Pr} = 65\%$ značící počet správných klasifikací.

c) *Kritérium poměrné pravděpodobnosti* se užije pro nestejně veliké třídy a klasifikování prvků do dvou či více tříd dle vzorce $\pi_1 = p^2 + (1 - p)^2$, kde p je podíl prvků v 1. třídě a $(1 - p)$ je podíl prvků ve 2. třídě.
Například: když máme třídy 75% a 25%, bude $\pi_1 = 0.75^2 + (1 - 0.75)^2 = 0.625$.

Statistické míry klasifikace spolehlivosti diskriminace:

1) **Pressovo *q*-kritérium** vyjadřuje míru porovnání počtu správných klasifikací vůči celkové velikosti výběru a počtu tříd. Vyčíslí se vztahem

$$q = \frac{(n - n_s k)^2}{n(k - 1)}$$

kde n je velikost výběru,
n_s udává počet objektů správně klasifikovaných a
k značí počet tříd,

Testování: vyšetřuje diskriminační sílu klasifikační matice v porovnání s modelem pravděpodobnosti: vypočtené q se porovnává s $\chi^2_{1-\alpha}(1)$ při dané α, a to když q překročí $\chi^2_{1-\alpha}(1)$, klasifikační matice se jeví statisticky lepší než klasifikační pravděpodobnost.

Vlastnosti testu:

Test je citlivý na velikost výběru, protože velké výběry jsou náchylné snadněji vykazovat statistickou významnost než výběry malé.

I když se totiž velikosti výběru zvyšují, menší klasifikační poměr bude stále významný.

Například, pro $n = 50$, $n_s = 42$ a $k = 2$ bude $q = (50 - 42 \times 2)^2/(50(2 - 1)) = 23.12$ a kritická hodnota pro $\alpha = 0.01$ je $\chi^2_{1-\alpha}(1) = 6.63$.

Závěr: Predikce jsou statisticky významnější než apriorní pravděpodobnost π_i, která uvádí správnou klasifikaci pro 50 %.

Například zvětšíme velikost výběru na 100 a klasifikační poměr zůstane 84 %, Pressovo q se zvýší na 46.24.

2) **Diagnostiky jednotlivých objektů:** stanovení těsnosti proložení modelem spočívá na vyšetření predikčních výsledků celé řady objektů.

Hledáme, který objekt byl chybně klasifikován a který nereprezentuje zbývající objekty třídy.

I když znázorníme, které objekty jsou správně a které nesprávně klasifikovány, stále potřebujeme míru podobnosti objektů vůči zbytku třídy.

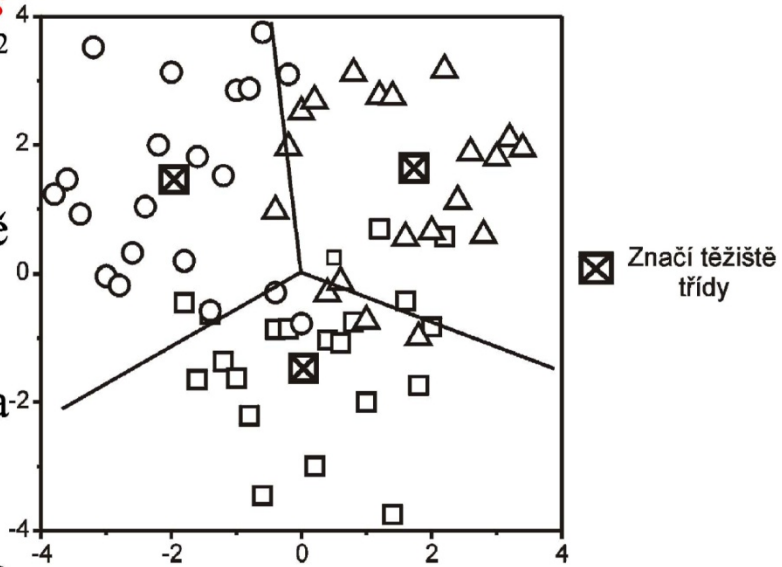
Podobnost objektu vůči ostatním objektům třídy:

Na základě vyšetření Mahalanobisovy vzdálenosti D^2 mezi posuzovaným objektem a těžištěm třídy.

Objekty blíže k těžišti jsou v této třídě reprezentativnější než objekty vzdálenější od těžiště.

Grafická analýza: definujeme vybočující objekty a sledujeme jejich vliv v teritoriální mapě.

Nachází-li se objekty poblíž těžiště ukazuje to jednak na charakteristiku třídy, ale také na rozptýlení prvků ve třídě.



Obrázek teritoriální mapy.

Případ	Mahalanobisovy vzdálenosti (73Iris) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
1	SETOSA	0,2419	90,6602	181,5587
2	VIRGINIC	208,5713	27,3188	1,8944
3	VERSICOL	105,2663	2,2329	13,0720
4	VIRGINIC	207,9180	31,7492	4,4506
* 5	VIRGINIC	133,0668	5,2529	7,2359
6	SETOSA	1,3337	84,0118	170,0569
7	VIRGINIC	173,1838	26,5620	11,0484
8	VERSICOL	131,6617	8,4307	14,7647
* 9	VERSICOL	130,8624	8,6697	6,5068
10	SETOSA	2,2864	113,6509	210,0239
11	VERSICOL	99,2338	1,2963	13,8174
* 12	VERSICOL	149,0303	8,4393	4,8645
13	VIRGINIC	158,9817	12,7512	1,2342
14	VERSICOL	79,1079	1,4076	26,6531
15	VIRGINIC	161,8529	12,1703	1,9781
16	VIRGINIC	174,0819	16,0529	2,3902
17	VIRGINIC	209,0295	29,5143	1,9395
18	SETOSA	2,7690	67,4717	145,7007
19	VERSICOL	104,5191	2,7936	15,7979
20	VIRGINIC	188,8750	28,1929	5,6252
21	VIRGINIC	213,4056	37,9708	11,1557
22	VERSICOL	102,0479	3,2965	11,7819
23	VIRGINIC	215,2034	36,5459	6,1622
24	VIRGINIC	230,2236	35,8899	7,4448
25	VIRGINIC	160,5946	17,3403	11,3925
26	SETOSA	3,9513	113,1501	210,1669
27	VIRGINIC	183,5631	26,1715	7,1685
28	VERSICOL	86,5713	5,0343	23,2534
29	VERSICOL	88,8713	1,8095	16,2189
30	VERSICOL	75,6723	1,8657	25,0269
31	SETOSA	1,8029	73,8199	159,0406
32	VIRGINIC	153,7618	11,9778	3,9269
33	VERSICOL	80,1609	1,9167	27,3454
34	VIRGINIC	166,9178	14,4570	4,2028
35	VIRGINIC	184,8178	24,9964	3,8858
36	SETOSA	0,7032	97,5630	190,4399
37	SETOSA	1,7964	105,3213	200,5307
38	VERSICOL	117,0332	7,6733	14,2231
39	VIRGINIC	176,5572	26,7790	11,6559
40	SETOSA	2,0048	84,0333	171,8626
41	VIRGINIC	179,4320	19,8406	1,1728
42	SETOSA	3,6979	72,9994	150,5149
43	VERSICOL	81,0528	0,6990	23,4829
44	SETOSA	2,5780	78,0003	164,4747
45	VIRGINIC	238,3291	39,3966	11,6650
46	VIRGINIC	144,7976	8,0379	3,5786
47	SETOSA	1,9783	75,7291	160,9768
48	VERSICOL	106,2031	3,8280	18,6165
49	VIRGINIC	171,0985	17,3642	5,5252
50	SETOSA	2,5908	75,1950	159,8312

Případ	Mahalanobisovy vzdálenosti (73Iris) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
51	SETOSA	2,4781	102,3241	194,2383
52	VIRGINIC	137,0315	7,0426	3,3147
53	SETOSA	5,2822	74,2231	156,9431
54	SETOSA	2,4034	74,2557	160,0999
55	SETOSA	1,7097	97,4818	190,7249
56	VIRGINIC	175,5659	21,1319	15,8330
57	VIRGINIC	199,8679	23,3820	0,8865
58	SETOSA	2,3656	89,6421	181,4171
59	SETOSA	10,6473	145,3975	249,2639
60	SETOSA	3,4054	80,7331	162,2563
61	VERSICOL	80,3117	3,7493	23,8726
62	VIRGINIC	137,5205	6,8231	3,9018
63	SETOSA	2,0313	80,9713	169,1870
64	SETOSA	0,2911	98,8847	191,7886
65	VERSICOL	110,8837	4,6621	12,5128
66	VERSICOL	74,9365	3,0905	30,8988
67	SETOSA	0,8497	76,3812	159,0772
68	SETOSA	1,1793	82,0178	169,6370
69	VERSICOL	104,2282	0,9804	10,7555
70	VERSICOL	56,3716	5,6238	41,1703
71	SETOSA	9,9923	133,8673	229,4021
72	SETOSA	0,5956	100,9232	193,8540
73	VIRGINIC	215,4602	34,8047	8,7073
74	VIRGINIC	206,8479	34,3116	6,4435
75	VIRGINIC	174,0819	16,0529	2,3902
76	VERSICOL	86,4677	1,0306	19,2454
77	VIRGINIC	155,5493	13,3989	9,0850
78	SETOSA	2,8189	88,9556	175,7645
79	SETOSA	7,8957	128,3151	227,3521
80	VIRGINIC	194,3124	22,4317	1,3114
81	VIRGINIC	161,2735	13,3248	3,1597
82	VIRGINIC	134,1908	6,9612	4,0938
83	VIRGINIC	177,3643	16,9118	3,2303
84	VERSICOL	66,8487	5,5980	37,6079
85	VERSICOL	94,3909	0,7785	16,9879
86	VERSICOL	80,2017	2,8341	24,6553
87	SETOSA	7,8843	135,8763	236,1242
88	SETOSA	1,1800	79,0946	166,4767
89	VIRGINIC	286,5761	58,0811	17,1857
90	VIRGINIC	158,7299	10,7184	8,1964
91	SETOSA	1,9448	95,9400	183,1140
92	VERSICOL	90,7313	2,1032	20,1282
93	VERSICOL	94,7422	3,0717	18,2833
94	VERSICOL	96,7766	5,2776	15,5110
95	SETOSA	0,1174	89,5104	179,5753
96	SETOSA	3,8501	71,6410	155,9269
97	VERSICOL	87,7219	6,1713	33,0998
98	VERSICOL	72,2566	3,1074	32,9760
99	VERSICOL	73,7968	1,0745	26,0469
100	SETOSA	0,5533	87,2894	177,0701

Případ	Mahalanobisovy vzdálenosti (73Iris) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
101	SETOSA	2,0867	75,2937	160,7244
102	VIRGINIC	195,4075	26,1183	2,2160
103	VERSICOL	82,4650	0,5454	20,5015
104	VIRGINIC	196,3446	23,4415	5,7537
105	VERSICOL	89,8691	1,3404	15,6756
106	SETOSA	0,1432	91,1053	181,8208
107	SETOSA	0,5788	97,1651	188,0006
108	VIRGINIC	172,4599	18,4859	4,2928
109	VERSICOL	86,9920	6,0973	33,3304
110	VIRGINIC	200,8481	25,7886	8,0346
111	VIRGINIC	197,3391	21,2942	4,4828
112	SETOSA	1,4419	84,1230	174,4342
113	VERSICOL	96,7106	2,3964	15,0578
114	VERSICOL	109,7475	5,2503	20,2811
115	SETOSA	1,5991	107,1781	201,8772
116	VERSICOL	81,5154	2,1405	21,8933
117	VERSICOL	133,9424	4,8762	7,8489
118	VERSICOL	102,7239	3,3849	18,8980
119	VERSICOL	80,4464	1,6652	22,9223
120	VERSICOL	56,1418	9,5513	45,1437
121	VERSICOL	84,7852	0,3734	19,4217
122	VIRGINIC	209,7892	28,0552	1,6141
123	VIRGINIC	166,6071	16,5332	2,3445
124	SETOSA	4,4186	115,7274	208,9456
125	SETOSA	0,3669	94,1941	184,1690
126	VIRGINIC	215,4522	37,2362	6,0931
127	VIRGINIC	146,6684	11,3725	2,7214
128	VERSICOL	107,2758	1,7068	12,0394
129	VERSICOL	64,8054	3,7695	30,4000
130	VERSICOL	103,3818	3,5491	14,4894
131	VIRGINIC	170,9715	13,7008	0,9189
132	VIRGINIC	177,3138	18,0392	1,0116
133	VERSICOL	94,1567	1,0258	18,2519
134	SETOSA	1,4143	83,3404	170,5934
135	SETOSA	1,8033	83,9406	174,4348
136	SETOSA	15,8000	64,9036	144,3573
137	VIRGINIC	187,7461	21,1036	4,0176
138	SETOSA	2,5326	102,6988	194,0759
139	SETOSA	1,4696	100,8776	191,3738
140	VERSICOL	102,1168	2,3796	18,2508
141	VERSICOL	84,1656	2,7309	21,5264
142	VERSICOL	105,9403	3,7242	16,4216
143	SETOSA	1,0549	88,0678	177,7126
144	SETOSA	1,0069	92,3452	179,2350
145	SETOSA	0,3420	96,1764	188,2198
146	VIRGINIC	216,2140	37,6554	10,0253
147	VERSICOL	124,8026	4,6670	6,2599
148	VIRGINIC	245,7223	47,8027	10,2841
149	SETOSA	1,2454	105,2551	199,3038
150	VERSICOL	68,4668	5,4822	37,4651

5. krok: Interpretace výsledků

Tři metody určují relativní důležitost znaku:

1) **Standardizované diskriminační koeficienty:** interpretace diskriminačních funkcí vyšetřuje znaménko a velikost *standardizovaných diskriminačních koeficientů* $\boldsymbol{a}^T = [a_0, a_1, \ldots, a_p]$, které představují relativní příspěvek svého znaku do Fisherovy lineární diskriminační funkce:

Diskuze:

1) Znaky *s relativně velkými koeficienty* přispívají více do diskriminační síly diskriminační funkce než znaky s menšími koeficienty.

2) Znaménko ukazuje, že znak dává buď kladný, nebo záporný příspěvek.

3) Malý koeficient indikuje buď, že odpovídající znak je nevýznamný k určování vztahu nebo je neúplným vztahem, protože je zde vysoký stupeň multikolinearity.

4) Problémem je také značná nestabilita diskriminačních koeficientů.

Proměnná	Standardiz. koeficienty (73Iris) pro kanonické proměnné	
	Kořen1	Kořen2
SEPALLENGTH	0,42695	0,012408
SEPALWIDTH	0,52124	0,735261
PETALLENGTH	-0,94726	-0,401038
PETALWIDTH	-0,57516	0,581040
Vlastní	32,19193	0,285391
KumPodíl	0,99121	1,000000

2) Diskriminační zátěže (označováné také *strukturní korelace*):

a) Jsou mírou korelace mezi každým znakem x_i a diskriminační funkcí Z_k.

b) Vystihují rozptyl, který sdílejí znaky x_i s diskriminační funkcí Z_k,

c) Jsou to faktorové zátěže při posouzení relativního příspěvku každého znaku do diskriminační funkce.

Vyčíslí se:

1) Pro každý objekt hodnota diskriminační funkce Z_k a korelační koeficient mezi Z_k a znakem x_i.

2) Samostatnou *korelační matici* lze vyčíslit pro každou třídu.

3) Výsledky lze dále kombinovat k vyčíslení *společné korelační matice* mezi všemi třídami.

4) Můžeme také uvažovat všechny objekty dohromady a vyčíslit *celkovou korelační matici.*

	Faktorová strukturní matice (73Iris) Korelační proměnné - Kanonické kořeny (vnitřní korelace)	
Proměnná	Kořen1	Kořen2
SEPALLENGTH	-0,222596	0,310812
SEPALWIDTH	0,119012	0,863681
PETALLENGTH	-0,706065	0,167701
PETALWIDTH	-0,633178	0,737242

6. krok: Ověření výsledků

Konečné stadium DA se týká potvrzení diskriminačních výsledků.

DA má tendenci přeceňovat hit poměr.

Metoda dělení do skupin (cross-validation) je velmi užitečná.

Rozdělení výběru:

1) Soubor je náhodně rozdělen *na analyzovaný výběr* a *na klasifikovaný výběr*.

2) Místo dělení na *analyzovaný* a *klasifikovaný* můžeme zcela náhodně rozdělit soubor několikrát.

Testujeme potvrzení diskriminační funkce pomocí *klasifikační matice* a *hit poměru*.

Když *najdeme znaky*, které mají největší vliv na diskriminaci mezi třídami, je dalším krokem profilování charakteristiky tříd, založené na třídních průměrech.

Příklad 4.24 *Botanické třídění kosatců diskriminační analýzou*

Ve Fisherově úloze o rozměrech okvětních lístků u 150 kosatců se analyzují květy tří základních tříd: (1) *Iris setosa,* (2) *Iris versicolor,* (3) *Iris virginica.* Květy kosatců jsou popsány čtyřmi znaky: délkou kališních lístků v mm anglicky *lsepal* a jejich šířkou *wsepal,* dále délkou korunních plátků v mm *lpetal* a jejich šířkou *wpetal*.

Každý objekt je popsán $p = 4$ znaky, a to *SepalLength, SepalWidth, PetalLength, PetalWidth*.

Data: Data jsou v pořadí proměnných *E418a, E418b, E418c, E418d:*

5.1	3.5	1.4	0.2	5.4	3.9	1.3	0.4	5.2	4.1	1.5	0.1	5.3	3.7	1.5	0.2
4.9	3.0	1.4	0.2	5.1	3.5	1.4	0.3	5.5	4.2	1.4	0.2	5.0	3.3	1.4	0.2
4.7	3.2	1.3	0.2	5.7	3.8	1.7	0.3	4.9	3.1	1.5	0.2	7.0	3.2	4.7	1.4
4.6	3.1	1.5	0.2	5.1	3.8	1.5	0.3	5.0	3.2	1.2	0.2	6.4	3.2	4.5	1.5
5.0	3.6	1.4	0.2	5.4	3.4	1.7	0.2	5.5	3.5	1.3	0.2	6.9	3.1	4.9	1.5
5.4	3.9	1.7	0.4	5.1	3.7	1.5	0.4	4.9	3.6	1.4	0.1	5.5	2.3	4.0	1.3
4.6	3.4	1.4	0.3	4.6	3.6	1.0	0.2	4.4	3.0	1.3	0.2	6.5	2.8	4.6	1.5
5.0	3.4	1.5	0.2	5.1	3.3	1.7	0.5	5.1	3.4	1.5	0.2	5.7	2.8	4.5	1.3
4.4	2.9	1.4	0.2	4.8	3.4	1.9	0.2	5.0	3.5	1.3	0.3	6.3	3.3	4.7	1.6
4.9	3.1	1.5	0.1	5.0	3.0	1.6	0.2	4.5	2.3	1.3	0.3	4.9	2.4	3.3	1.0
5.4	3.7	1.5	0.2	5.0	3.4	1.6	0.4	4.4	3.2	1.3	0.2	6.6	2.9	4.6	1.3
4.8	3.4	1.6	0.2	5.2	3.5	1.5	0.2	5.0	3.5	1.6	0.6	5.2	2.7	3.9	1.4
4.8	3.0	1.4	0.1	5.2	3.4	1.4	0.2	5.1	3.8	1.9	0.4	5.0	2.0	3.5	1.0
4.3	3.0	1.1	0.1	4.7	3.2	1.6	0.2	4.8	3.0	1.4	0.3	5.9	3.0	4.2	1.5
5.8	4.0	1.2	0.2	4.8	3.1	1.6	0.2	5.1	3.8	1.6	0.2	6.0	2.2	4.0	1.0
5.7	4.4	1.5	0.4	5.4	3.4	1.5	0.4	4.6	3.2	1.4	0.2	6.1	2.9	4.7	1.4

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica

Řešení: zdrojová matice 150 řádků a 4 sloupce neobsahuje žádné chybějící prvky.

1. Znaky

1.1 Aritmetický průměr [mm] objektů ve třídě Setosa G_1, Versicolor G_2, Virginica G_3.

Znak	**Setosa G_1**	**Versicolor G_2**	**Virginica G_3**	**Celkově $G_1+G_2+G_3$**
SepalLength	50.06	59.36	65.88	58.43333
SepalWidth	34.28	27.7	29.74	30.57333
PetalLength	14.62	42.6	55.52	37.58
PetalWidth	2.46	13.26	20.26	11.99333
Počet	50	50	50	150

Tabulka obsahuje průměry každého znaku pro každou třídu kosatců G_1, G_2 a G_3.

1.2 Směrodatné odchylky [mm] objektů ve třídě Setosa G_1, Versicolor G_2, Virginica G_3.

Znak	Setosa G_1	Versicolor G_2	Virginica G_3	Celkově $G_1+G_2+G_3$
SepalLength	3.524897	5.161712	6.358796	8.280662
SepalWidth	3.790644	3.137983	3.224966	4.358663
PetalLength	1.73664	4.69911	5.518947	17.65298
PetalWidth	1.053856	1.977527	2.7465	7.622377
Počet	50	50	50	150

Lineární diskriminační analýza předpokládá, že kovarianční matice jsou stejné pro každou třídu. Tabulka posoudí předpoklad, zda jsou směrodatné odchylky v třídách zhruba stejné.

1.3 Celkové korelace a kovariance znaků.

Znak	SepalLength	SepalWidth	PetalLength	PetalWidth
SepalLength	68.56935	-4.243401	127.4315	51.62707
SepalWidth	-0.117570	18.99794	-32.96564	-12.16394
PetalLength	0.871754	-0.428440	311.6278	129.5609
PetalWidth	0.817941	-0.366126	0.962865	58.10063

Korelace jsou v dolní levé části, kovariance jsou v pravé horní části matice. Rozptyly znaků jsou na diagonále matice.

1.4 Mezitřídní korelace a kovariance znaků.

Znak	SepalLength	SepalWidth	PetalLength	PetalWidth
SepalLength	3160.607	-997.6334	8262.42	3563.967
SepalWidth	-0.745075	567.2466	-2861.98	-1146.633
PetalLength	0.994135	-0.812838	21855.14	9338.7
PetalWidth	0.999768	-0.759258	0.996232	4020.667

Tabulka obsahuje korelace a kovariance, vytvořené za použití průměrů tříd objektů místo jednotlivých objektů. Korelace jsou v dolní levé části, mezitřídní kovariance jsou na diagonále matice a v horní pravé části matice.

1.5 Vnitrotřídní korelace a kovariance znaků.

Znak	SepalLength	SepalWidth	PetalLengthPetalWidth
SepalLength	26.50082	9.272109	16.751433.840136
SepalWidth	0.530236	11.53878	5.5243543.27102
PetalLength	0.756164	0.377916	18.518784.266531
PetalWidth	0.364506	0.470535	0.4844594.188163

V datech byly třídní průměry objektů odečteny. Korelace jsou v dolní levé části, vnitrotřídní kovariance jsou na diagonále a v pravé horní části matice.

1.6 Vyšetření vlivu jednotlivých znaků.

	Při odstranění dotyčného znaku		Pro dotyčný samotný znak		R^2		
Diskrimin.	Wilk. λ	*F*-test	Spočt. α	Wilk. λ	*F*-test	Spočt. α	Ostatní *X*
SepalLength	0.938463	4.720	0.010329	0.381294	119.26	0.000000	0.858612
SepalWidth	0.766480	21.94	0.000000	0.599217	49.16	0.000000	0.524007
PetalLength	0.669206	35.59	0.000000	0.058628	1180.2	0.000000	0.968012
PetalWidth	0.743001	24.90	0.000000	0.071117	960.01	0.000000	0.937850

Vliv jednotlivých znaků na diskriminační analýzu: *spočtená hladina významnosti při odstranění dotyčného znaku* je vypočtená u *F*-testu při odstranění dotyčného znaku. Znak je důležitý, je-li tato hodnota menší než uživatelem zadaná hladina významnosti $\alpha = 0.05$. Zde jsou všechny čtyři znaky statisticky významné. *Wilk. λ pro dotyčný samotný znak* značí hodnotu za použití jediného znaku, který je důležitý, je-li tato hodnota menší než uživatelem zadaná hladina významnosti $\alpha = 0.05$.

1.7 Automatická volba účinných znaků

Úkolem v diskriminační analýze je výběr znaků. Z velké skupiny možných znaků je třeba vybrat menší výběr těch nejlepších, maximálně 8 proměnných.

Krok	Činnost v kroku	Diskrimin.	% změny lambda	v *F*-test	Spočtená hladina α	Wilkovo lambda
0	Žádné					1.000000
1	Zavedeno	PetalLength	94.14	1180.16	0.000000	0.058628
2	Zavedeno	SepalWidth	37.09	43.04	0.000000	0.036884
3	Zavedeno	PetalWidth	32.29	34.57	0.000000	0.024976
4	Zavedeno	SepalLength	6.15	4.72	0.010329	0.023439
.. .		.....			...	.....

Detail ve 4. kroku automatického volby znaku:

Status	Diskrimin.	% změny lambda	v *F*-test	Spočtená hladina α	R^2 na ostatních *X*
Uvnitř	SepalLength	6.15	4.72	0.010329	0.858612
Uvnitř	SepalWidth	23.35	21.94	0.000000	0.524007
Uvnitř	PetalLength	33.08	35.59	0.000000	0.968012
Uvnitř	PetalWidth	25.70	24.90	0.000000	0.937850

Celkové Wilkovo $\lambda = 0.023439$

Automatický výběr znaků se plní krokově - nejprve se nalezne nejlepší znak a potom druhý nejlepší. Postupný krokový proces přidávání nejlepšího zbývajícího znaku s ověřením, zda by jeden aktivní znak mohl být odebrán, pokračuje, dokud je ještě nějaký znak k dispozici. *% změny v lambda* značí procento snížení v hodnotě lambda, které je výsledkem tohoto kroku. Wilkovo kritérium λ je analogické (1 - R^2) ve vícenásobné regresi. Vhodnější model sníží Wilkovo kritérium λ .

2. Diskriminační funkce.

2.1 Odhady diskriminačních koeficientů $a_0, a_1, \ldots, a_p$ diskriminační funkce Z_k pro třídy Setosa G_1, Versicolor G_2, Virginica G_3.

Znak	Setosa G_1	Versicolor G_2	Virginica G_3
Absolutní člen	-85.20985	-71.754	-103.2697
SepalLength	2.354417	1.569821	1.244585
SepalWidth	2.358787	0.707251	0.3685279
PetalLength	-1.643064	0.5211451	1.276654
PetalWidth	-1.739841	0.6434229	2.107911

Tabulka obsahuje celkem tři diskriminační funkce, jednu pro každou třídu ve sloupci. Tato diskriminační funkce je určena k zařazování nových, dosud nezařazených objektů do tříd čili ke klasifikaci, a proto se v této formě nazývá *klasifikační funkce.*

Klasifikační funkce *pro první třídu* G_1: -85.20985+2.35442*SepalLength +2.35879*SepalWidth-1.64306*PetalLength-1.73984*PetalWidth.

Tyto tři funkce mohou být použity k zařazení nových kosatců do navržených tříd.

2.2 Odhady regresních parametrů a_0 , a_1 , ..., a_p lineárního regresní modelu pro každou třídu Setosa G_1, Versicolor G_2, Virginica G_3

Znak	Setosa G_1	Versicolor G_2	Virginica G_3
Absolutní člen	0.1182229	1.577059	-0.6952819
SepalLength	6.602977E-03	-2.015369E-03	-4.587608E-03
SepalWidth	2.428479E-02	-4.456162E-02	2.027684E-02
PetalLength	-2.246571E-02	2.206692E-02	3.987911E-04
PetalWidth	-5.747273E-03	-4.943066E-02	5.517793E-02

Tabulka obsahuje regresní parametry a_0 , a_1 , ..., a_p lineárního regresního modelu. Zařazení objektů do třídy se provede tak, že se u každého objektu vybere vždy třída s nejvyšší hodnotou diskriminačního skóre.

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica

3. Klasifikace objektů.

3.1 Zařazení (klasifikace) objektů klasifikační funkcí do tříd tvoří klasifikační matici kosatců ve třídách Setosa G_1, Versicolor G_2, Virginica G_3:

	Nalezená predikovaná třída			
Daná třída	Setosa G_1	Versicolor G_2	Virginica G_3	Celkově $G_1+G_2+G_3$
Setosa	50	0	0	50
Versicolor	0	34	16	50
Virginica	0	7	43	50
Celkově	50	41	59	150

Redukce v klasifikační správnosti v důsledku proměnných $X = 77.0$ %.

Klasifikační matice ukazuje, jak diskriminační funkce zařazují objekty do tří tříd. Bylo-li dosaženo perfektního zařazení, obdržíme v matici mimo diagonálu nuly. *Redukce v klasifikační správnosti* obsahuje procento snížení v zařazovací správnosti, dosažené diskriminačními funkcemi vůči očekávanému zařazení dle původní závisle proměnné.

3.2 Přehled chybně zařazených objektů v řádcích do tříd Setosa G_1, Versicolor G_2, Virginica G_3 dle klasifikačních funkcí proti původní závisle proměnné.

			Procento zařazení do jednotlivé třídy		
Objekt	**Daná třída**	**Nalezená třída**	**Setosa G_1**	**Versicolor G_2**	**Virginica G_3**
5	Virginica	Versicolo	-1.8	58.6	43.1
9	Versicolo	Virginica	10.3	20.2	69.5
22	Versicolo	Virginica	18.8	22.6	58.6
28	Versicolo	Virginica	22.1	35.5	42.4
29	Versicolo	Virginica	22.1	27.4	50.6
38	Versicolo	Virginica	10.6	38.3	51.1
45	Virginica	Versicolo	-31.4	66.4	65.0
57	Virginica	Versicolo	-18.6	83.9	34.7
62	Versicolo	Virginica	24.4	34.0	41.6
66	Versicolo	Virginica	11.9	37.9	50.2
70	Versicolo	Virginica	12.1	41.5	46.3
78	Virginica	Versicolo	-7.3	58.4	48.9
91	Virginica	Versicolo	-16.1	83.8	32.3
95	Versicolo	Virginica	23.7	14.3	62.0
106	Versicolo	Virginica	20.7	30.7	48.7
111	Virginica	Versicolo	-21.4	63.8	57.6
112	Virginica	Versicolo	-23.9	71.8	52.1
114	Versicolo	Virginica	17.1	35.6	47.2
117	Versicolo	Virginica	22.1	38.9	39.0
130	Versicolo	Virginica	30.9	32.4	36.8
131	Versicolo	Virginica	14.0	39.6	46.4
142	Versicolo	Virginica	21.4	38.6	40.0
148	Versicolo	Virginica	6.8	36.8	56.4

V řádku je hodnota pravděpodobnosti (v procentech), že objekt se nachází v dané třídě kosatců. Hodnota blízko 100 % značí, že objekt patří do dotyčné třídy.

3.3 Zařazení všech objektů diskriminačními funkcemi do tříd Setosa G_1, Versicolor G_2, Virginica G_3:

			Procento zařazení do jednotlivé třídy		
Objekt	Daná třída	Nalezená třída	Setosa G_1	Versicolor G_2	Virginica G_3
1	Setosa	Setosa	92.4	21.6	-14.0
2	Virginica	Virginica	-16.4	34.9	81.5
3	Versicolo	Versicolo	10.8	47.2	42.0
..	.....	.....	.....	.....	.....
..	.....	.....	.....	.....	.....
150	Setosa	Setosa	101.8	5.4	-7.2

Tabulka obsahuje pro každý objekt skutečnou třídu kosatců, dále predikovanou třídu kosatců a procento pravděpodobnosti zařazení do jednotlivé třídy kosatců.

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica

4. Klasifikace objektů kanonickou korelační analýzou.

4.1 Analýza kanonických proměnných:

Fn	$C^{-1}B$ vlast.číslo	Ind. %	Total %	Kanon. korel.	Kanon. korel2	Čitatel *F*-test	Jmenov.SV	SpočtenáSV	Wilk. λ α kritérium
1	32.191929	99.1	99.1	0.9848	0.9699	199.1	8.0	288.00	.00000.023439
2	0.285391	0.9	100.0	0.4712	0.2220	13.8	3.0	145.00	.00000.777973

F-test testuje, zda tato funkce a další jsou statisticky významné.

Tabulka obsahuje zařazování objektů kanonickou korelační analýzou.

U kanonické korelační analýzy jsou dva typy znaků: první skupina obsahuje znaky a druhá závisle proměnné.

Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (73Iris)					
	Vlastní číslo	Kan. R	Wilk. Lambda	Chi-kv.	sv	p-hodn.
0	32,19193	0,984821	0,023439	546,1153	8	0,000000
1	0,28539	0,471197	0,777973	36,5297	3	0,000000

4.2 Odhady koeficientů u kanonických proměnných

Znak	1. kanonická proměnná	2. kanonická proměnná
Absolutní člen	-2.105106	6.661473
SepalLength	-0.082938	-0.002410
SepalWidth	-0.153447	-0.216452
PetalLength	0.220121	0.093192
PetalWidth	0.281046	-0.283919

Obsahuje kanonické koeficienty k výpočtu kanonického skóre.

4.3 Kanonické proměnné u třídních průměrů

Třída kosatců	Kanonická funkce 1	Kanonická funkce 2
Setosa	-7.6076	-0.215133
Versicolor	1.82505	0.7278996
Virginica	5.78255	-0.5127666

Skup.	Průměry kan. proměnných (73Iris)	
	Kořen1	Kořen2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

Tabulka obsahuje výsledky kanonických koeficientů k výpočtu kanonické funkce pro průměry, těžiště u každé třídy.

4.4 Standardizované kanonické koeficienty u kanonických proměnných

Znak	1. kanonická proměnná	2. kanonická proměnná
SepalLength	-0.426955	-0.012408
SepalWidth	-0.521242	-0.735261
PetalLength	0.947257	0.401038
PetalWidth	0.575161	-0.581040

Tabulka obsahuje standardizované kanonické koeficienty.

Proměnná	Standardiz. koeficienty (73Iris) pro kanonické proměnné	
	Kořen1	Kořen2
SEPALLENGTH	0,42695	0,012408
SEPALWIDTH	0,52124	0,735261
PETALLENGTH	-0,94726	-0,401038
PETALWIDTH	-0,57516	0,581040
Vlastní	32,19193	0,285391
KumPodíl	0,99121	1,000000

4.5 Korelace původních a kanonických proměnných

Znak	1. kanonická proměnná	2. kanonická proměnná
SepalLength	0.222596	-0.310812
SepalWidth	-0.119012	-0.863681
PetalLength	0.706065	-0.167701
PetalWidth	0.633178	-0.737242

Proměnná	Faktorová strukturní matice (73Iris) Korelační proměnné - Kanonické kořeny (vnitřní korelace)	
	Kořen1	Kořen2
SEPALLENGTH	-0,222596	0,310812
SEPALWIDTH	0,119012	0,863681
PETALLENGTH	-0,706065	0,167701
PETALWIDTH	-0,633178	0,737242

Tabulka obsahuje zátěže původních znaků a kanonických proměnných. Každá hodnota představuje korelaci mezi kanonickou proměnnou a znakem.

5. Skóre k zařazování objektů do tříd.

5.1 Lineární diskriminační skóre všech objektů k zařazení do tříd G_1, G_2 a G_3.

Objekt	Daná třída	Nalezená predikovaná třída		
		Skóre třídy G_1	Skóre třídy G_2	Skóre třídy G_3
		Setosa	Versicolor	Virginica
1	Setosa	83.86837	38.65921	-6.790054
2	Virginica	1.230765	91.857	104.5692
..	.....	.....	.....	.....
150	Setosa	98.72371	46.71882	-0.3055334

Tabulka obsahuje hodnoty vyčíslených lineárních diskriminačních skóre pro všech 150 kosatců.

5.2 Regresní skóre všech objektů k zařazení do tříd G_1, G_2 a G_3:

		Nalezená predikovaná třída			
Objekt	Daná třída	Skóre třídy G_1	Skóre třídy G_2		Skóre třídy G_3
		Setosa	Versicolor		Virginica
1	Setosa	0.923755	0.215832		-0.139588
2	Virginica	-0.163732	0.348623		0.815109
3	Versicolo	0.107759	0.471953		0.420288
..	.....	.....	.....	.....	
..	.....	.....	.....	.....	
150	Setosa	1.018238	0.053607		-0.071844

Tabulka obsahuje vyčíslená predikovaná skóre, založená na regresních koeficientech.

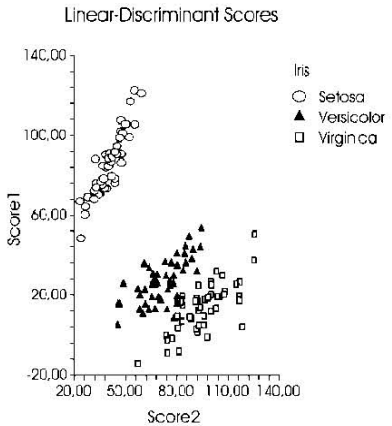
5.2 Kanonická skóre všech objektů k zařazení do dvou tříd G_1 a G_2.

		Nalezená predikovaná třída		
Objekt	Daná třída	Skóre třídy G_1		Skóre třídy G_2
1	Setosa	-7.671967		0.134894
2	Virginica	6.800150		-0.580895
3	Versicolo	2.548678		0.472205
..	.....	.....	.....	
..	.....	.....	.....	
150	Setosa	-8.314449		-0.644953

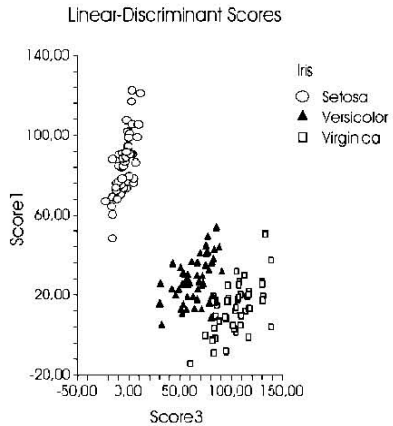
Tabulka obsahuje skóre kanonických proměnných pro každý řádek u všech 150 kosatců.

6. Klasifikační grafy.

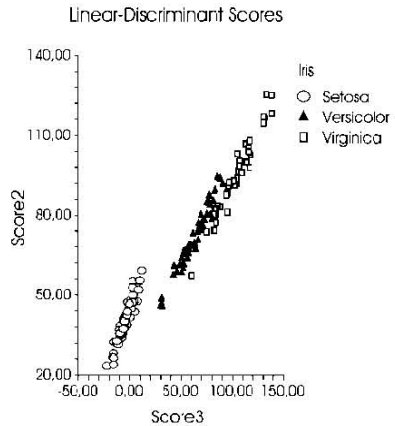
Užijeme grafy: (a) lineárních diskriminačních skóre, (b) regresních skóre nebo (c) kanonických skóre.



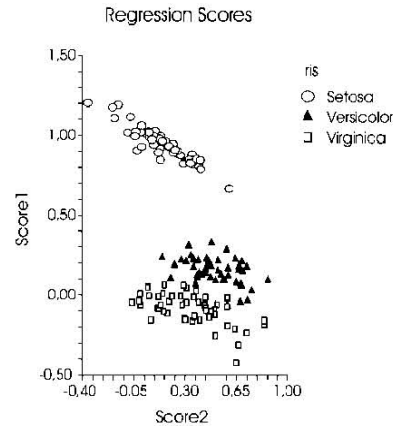
Obr. 7.9a Graf lineárního diskriminačního skóre 1 a 2 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).



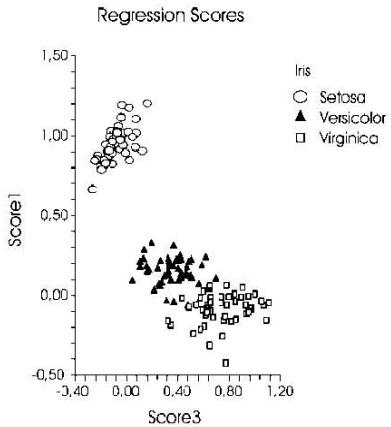
Obr. 7.9b Graf lineárního diskriminačního skóre 1 a 3 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).



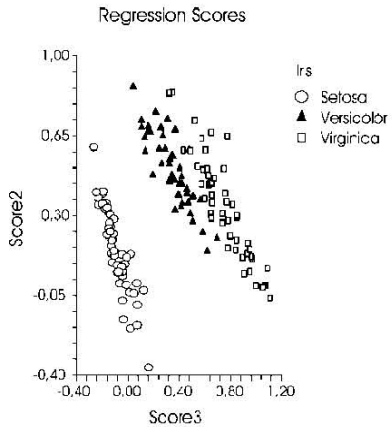
Obr. 7.9c Graf lineárního diskriminačního skóre 2 a 3 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).



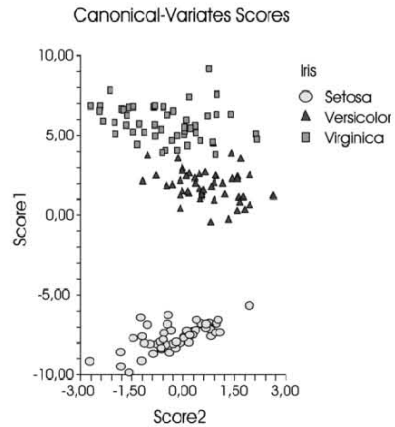
Obr. 7.9d Graf regresního skóre 1 a 2 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).



Obr. 7.9e Graf regresního skóre 1 a 3 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).



Obr. 7.9f Graf regresního skóre 2 a 3 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).



Obr. 7.9g Graf kanonických proměnných 1 a 2 u 4 znaků 150 kosatců zdrojové matice dat *Iris* (STATISTICA).

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica

Cvičení v programu STATISTICA

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica

STATISTICA Cz - [Data: 73Iris.sta (5s krát 150ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Fisher (1936) iris data: length & width of sepals and petals, 3 types of Iris

	1 SEPALLENGTH	2 SEPALWIDTH	3 PETALLENGTH	4 PETALWIDTH	5 Trida
1	5	3,3	1,4	0,2	SETOSA
2	6,4	2,8	5,6	2,2	VIRGINIC
3	6,5	2,8	4,6	1,5	VERSICOL
4	6,7	3,1	5,6	2,4	VIRGINIC
5	6,3	2,8	5,1	1,5	VIRGINIC
6	4,6	3,4	1,4	0,3	SETOSA
7	6,9	3,1	5,1	2,3	VIRGINIC
8	6,2	2,2	4,5	1,5	VERSICOL
9	5,9	3,2	4,8	1,8	VERSICOL
10	4,6	3,6	1	0,2	SETOSA
11	6,1	3	4,6	1,4	VERSICOL
12	6	2,7	5,1	1,6	VERSICOL
13	6,5	3	5,2	2	VIRGINIC
14	5,6	2,5	3,9	1,1	VERSICOL
15	6,5	3	5,5	1,8	VIRGINIC
16	5,8	2,7	5,1	1,9	VIRGINIC
17	6,8	3,2	5,9	2,3	VIRGINIC
18	5,1	3,3	1,7	0,5	SETOSA
19	5,7	2,8	4,5	1,3	VERSICOL
20	6,2	3,4	5,4	2,3	VIRGINIC
21	7,7	3,8	6,7	2,2	VIRGINIC
22	6,3	3,3	4,7	1,6	VERSICOL
23	6,7	3,3	5,7	2,5	VIRGINIC
24	7,6	3	6,6	2,1	VIRGINIC
25	4,9	2,5	4,5	1,7	VIRGINIC
26	5,5	3,5	1,3	0,2	SETOSA
27	6,7	3	5,2	2,3	VIRGINIC
28	7	3,2	4,7	1,4	VERSICOL
29	6,4	3,2	4,5	1,5	VERSICOL
30	6,1	2,8	4	1,3	VERSICOL
31	4,8	3,1	1,6	0,2	SETOSA
32	5,9	3	5,1	1,8	VIRGINIC
33	5,5	2,4	3,8	1,1	VERSICOL
34	6,3	2,5	5	1,9	VIRGINIC

1. Fisherova úloha kosatce (iris, 1936) třídí kosatce do tří tříd podle svých 4 znaků, a to dély a šířky kališních lístků a délky a šířky korunních plátků. Poslední sloupec dat je kódovaný znak.

Čtenář u své úlohy zamění okvětní lístky za znaky své úlohy. Poslední sloupec bude vždy kódovací znak.

STATISTICA Cz - [Data: 73Iris.sta (5s krát 150ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Obnovit... Ctrl+R
Analýza skupin
Základní statistiky/tabulky
Vícerozměrná regrese
ANOVA
Neparametrická statistika
Prokládání rozdělení
Pokročilé lineární/nelineární modely
Vícerozměrné průzkumné techniky
Průmyslová statistika & Six Sigma
Analýza síly testu
Neuronové sítě
Vytěžování dat
QC Data mining & Analýza hlavních příčin
Text & Document Mining, Web Crawling
Statistiky bloku dat
STATISTICA Visual Basic
Pravděpodobnostní kalkulátor

Shluková analýza
Faktorová analýza
Hlavní komponenty & klasifikační analýza
Kanonická analýza
Analýza spolehlivosti/prvků
Klasifikační stromy
Korespondenční analýza
Vícerozměrné škálování
Diskriminační analýza
Modely obecné diskriminační analýzy

Fisher (1936) iris data: length & widt

	1 SEPALLENGTH	2 SEPALWIDTH			5 Trida
1	5	3,3			TOSA
2	6,4	2,8			GINIC
3	6,5	2,8			RSICOL
4	6,7	3,1			GINIC
5	6,3	2,8			
6	4,6	3,4			
7	6,9	3,1			
8	6,2	2,2			
9	5,9	3,2			
10	4,6	3,6			
11	6,1	3			
12	6	2,7			
13	6,5	3			
14	5,6	2,5			
15	6,5	3			
16	5,8	2,7			GINIC
17	6,8	3,2			GINIC
18	5,1	3,3	1,7	0,5	SETOSA
19	5,7	2,8	4,5	1,3	VERSICOL
20	6,2	3,4	5,4	2,3	VIRGINIC
21	7,7	3,8	6,7	2,2	VIRGINIC
22	6,3	3,3	4,7	1,6	VERSICOL
23	6,7	3,3	5,7	2,5	VIRGINIC
24	7,6	3	6,6	2,1	VIRGINIC
25	4,9	2,5	4,5	1,7	VIRGINIC
26	5,5	3,5	1,3	0,2	SETOSA
27	6,7	3	5,2	2,3	VIRGINIC
28	7	3,2	4,7	1,4	VERSICOL
29	6,4	3,2	4,5	1,5	VERSICOL
30	6,1	2,8	4	1,3	VERSICOL
31	4,8	3,1	1,6	0,2	SETOSA
32	5,9	3	5,1	1,8	VIRGINIC
33	5,5	2,4	3,8	1,1	VERSICOL
34	6,3	2,5	5	1,9	VIRGINIC
35	6,4	3,2	5,3	2,3	VIRGINIC
36	5,2	3,4	1,4	0,2	SETOSA
37	4,9	3,6	1,4	0,1	SETOSA
38	5,4	3	4,5	1,5	VERSICOL
39	7,9	3,8	6,4	2	VIRGINIC
40	4,4	3,2	1,3	0,2	SETOSA
41	6,7	3,3	5,7	2,1	VIRGINIC
42	5	3,5	1,6	0,6	SETOSA
43	5,8	2,6	4	1,2	VERSICOL
44	4,4	3	1,3	0,2	SETOSA

2. Zvolí se *Diskriminační analýza* v bloku *Vícerozměrné průzkumové metody*.

STATISTICA Cz - [Data: 73Iris.sta (5s krát 150ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Fisher (1936) iris data: length & width of sepals and petals, 3 types of Iris

	1 SEPALLENGTH	2 SEPALWIDTH	3 PETALLENGTH	4 PETALWIDTH	5 Trida
1	5	3,3	1,4	0,2	SETOSA
2	6,4	2,8	5,6	2,2	VIRGINIC
3					
4					
5					
6					
7					
8					
9	5,9	3,2	4,8	1,8	VERSICOL
10	4,6	3,6	1	0,2	SETOSA
11	6,1	3	4,6	1,4	VERSICOL
12	6	2,7	5,1	1,6	VERSICOL
13	6,5	3	5,2		VIRGINIC
14	5,6	2,5	3,9		
15	6,5	3	5,5		
16	5,8	2,7	5,1		
17	6,8	3,2	5,9		
18	5,1	3,3	1,7		
19	5,7	2,8	4,5		
20	6,2	3,4	5,4		
21	7,7	3,8	6,7		
22	6,3	3,3	4,7		
23	6,7	3,3	5,7		
24	7,6	3	6,6		
25	4,9	2,5	4,5		
26	5,5	3,5	1,3		
27	6,7	3	5,2		
28	7	3,2	4,7		
29	6,4	3,2	4,5		
30	6,1	2,8			
31	4,8	3,1			
32	5,9	3			
33	5,5	2,4			
34	6,3	2,5			
35	6,4	3,2			
36	5,2	3,4			
37	4,9	3,6			
38	5,4	3			
39	7,9	3,8			
40	4,4	3,2			
41	6,7	3,3			
42	5	3,5			
43	5,8	2,6			
44	4,4	3			
45	7,7	2,8			
46	6,3	2,7			

3. Zadají se všechny nezávisle *Proměnné* a jedna *Grupovací proměnná.*

Diskriminační funkční analýza: 73Iris.sta

Základní nastavení

Proměnné

grupovací: nic

nezávislé: nic

Kódy pro grupovací prom.: nic

Další možnosti (kroková analýza)

Pro pokročilé funkce diskriminační analýzy nebo rozsáhlejší analytické problémy použijte Modely obecné diskriminační analýzy.

OK

Storno

Možnosti

Otevřít data

SELECT CASES f v

ChD vynechána

celé případy

substituce průměrem

Výběr jednu grupovací proměnnou a seznam nezávislých proměnných

1-SEPALLENGTH
2-SEPALWIDTH
3-PETALLENGTH
4-PETALWIDTH
5-Trida

Vybrat vše | Dl. názvy | Detaily

Groupovací proměnná: 5

1-SEPALLENGTH
2-SEPALWIDTH
3-PETALLENGTH
4-PETALWIDTH
5-Trida

Vybrat vše | Dl. názvy | Detaily

Seznam nezáv. proměnných: 1-4

OK

Storno

Volba "Ukázat pouze odpovídající proměnné" slouží k předzpracování seznamů proměnných a zobrazení kategorických a spojitých proměnných. Pro více informací stiskněte F1.

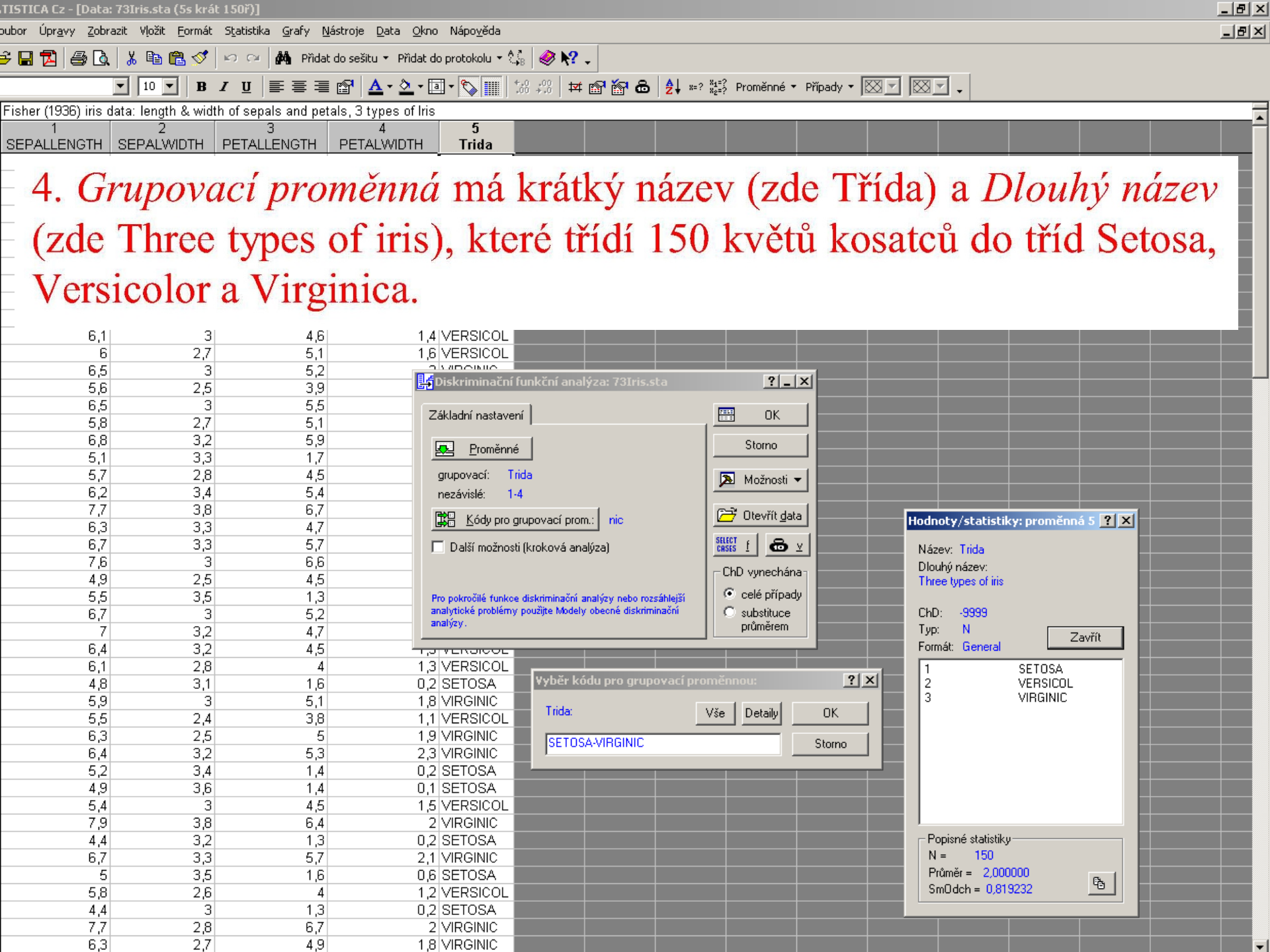
Ukázat pouze odpovídající proměnné

Diskriminační funkčn...

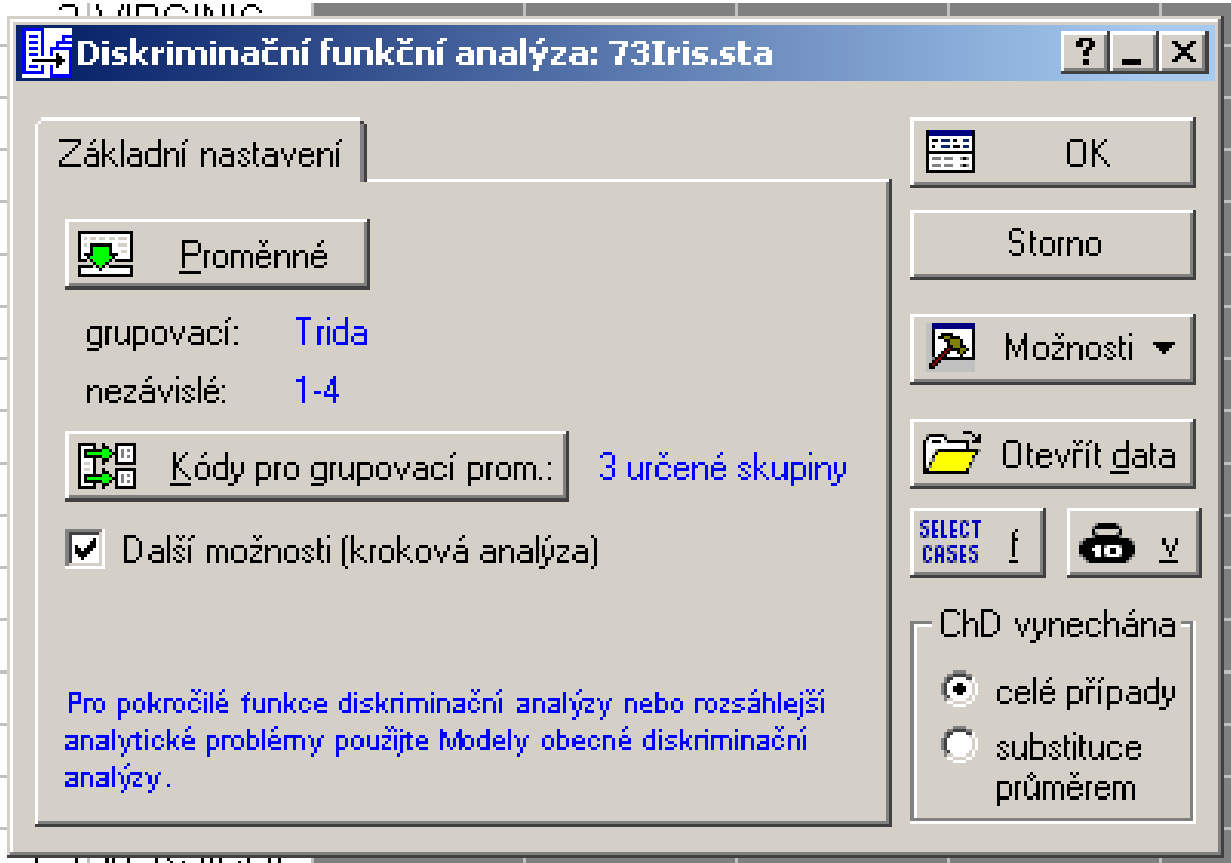
Připravena | Ř1;S5 | SETOSA | Filtr - | Váhy: VYPN | ABC | 123 | ZÁZN

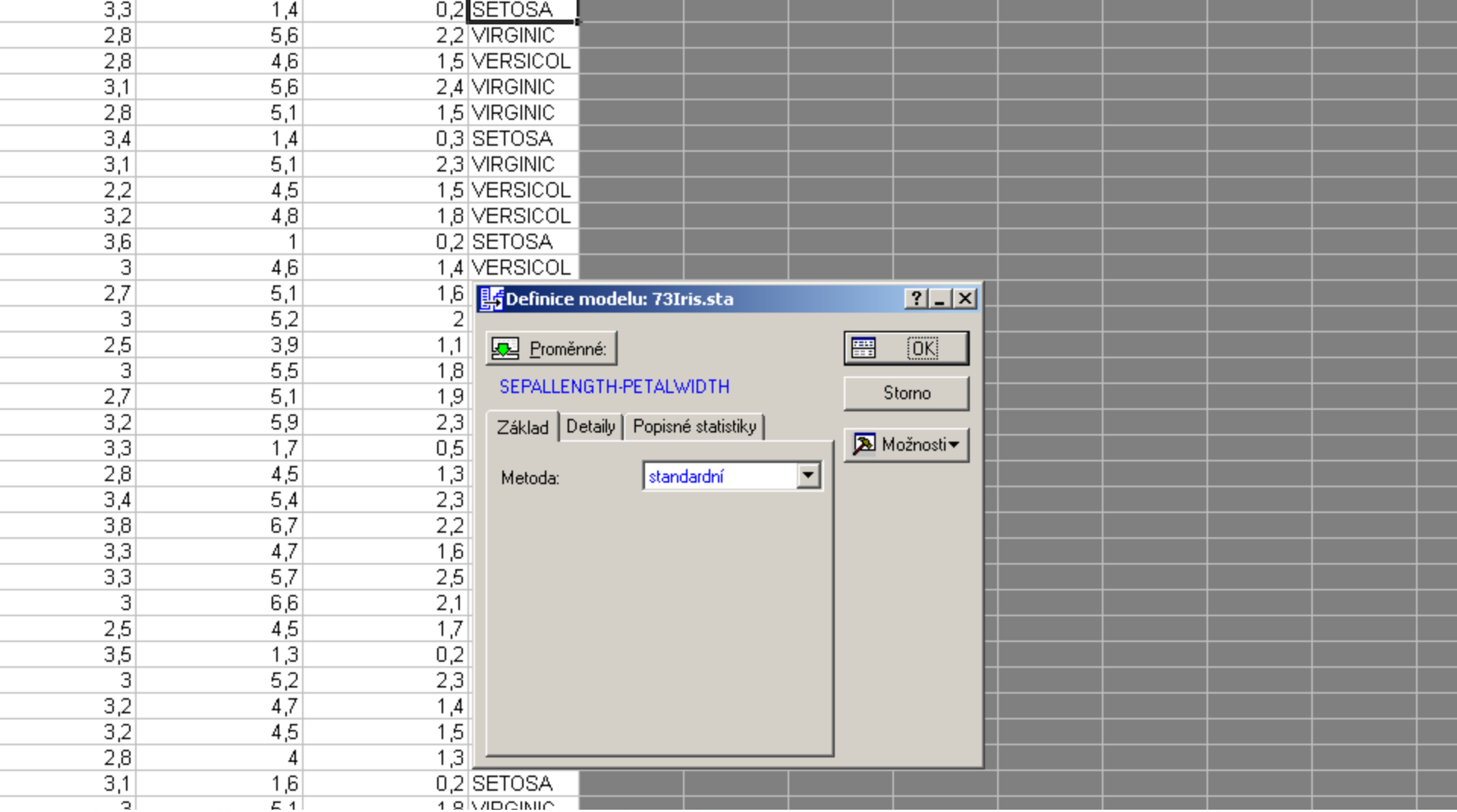
Start | Doručená pošta... | Total Command... | Zoner Photo St... | Anglicko-český -... | Microsoft Power... | STATISTICA C... | 8:53

4. *Grupovací proměnná* má krátký název (zde Třída) a *Dlouhý název* (zde Three types of iris), které třídí 150 květů kosatců do tříd Setosa, Versicolor a Virginica.

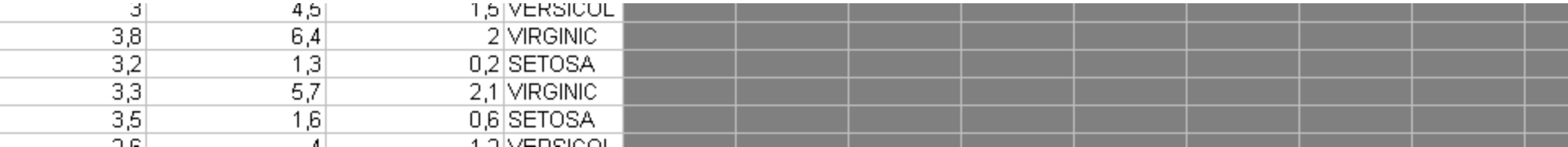


5. V okénku *Další možnosti* se zaškrtne *Kroková analýza*. Díry ve zdrojové matici vedou buď k vyškrtnutí řádku, objektu (značeno *Celé případy*) nebo k dodatečnému vyplnění díry průměrem (značeno *Substituce průměrem*).

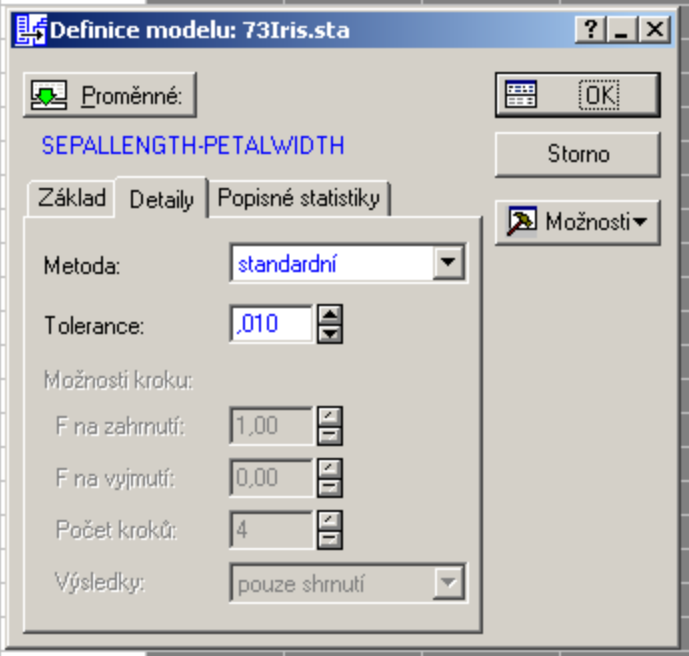




6. Definice modelu má tři záložkya to *Základ, Detaily, Popisné statistiky*, které si nyní postupně prohlédneme.



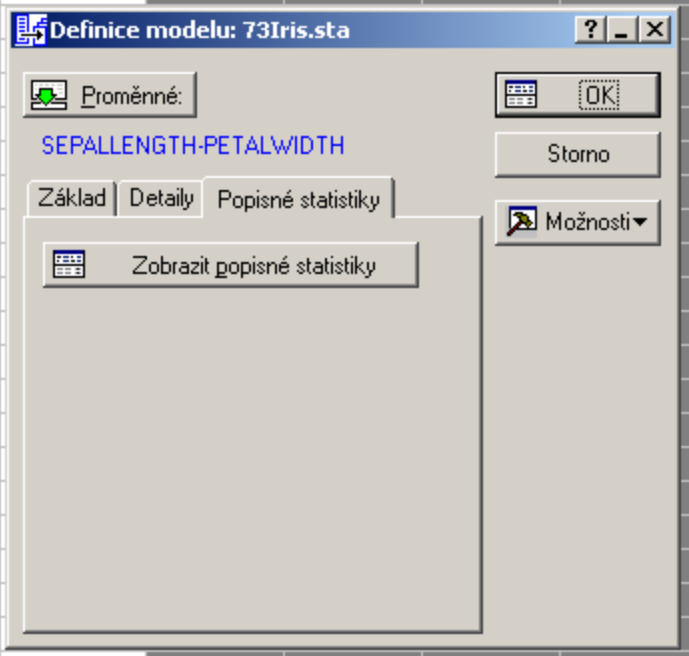
3,3	1,4	0,2	SETOSA
2,8	5,6	2,2	VIRGINIC
2,8	4,6	1,5	VERSICOL
3,1	5,6	2,4	VIRGINIC
2,8	5,1	1,5	VIRGINIC
3,4	1,4	0,3	SETOSA
3,1	5,1	2,3	VIRGINIC
2,2	4,5	1,5	VERSICOL
3,2	4,8	1,8	VERSICOL
3,6	1	0,2	SETOSA
3	4,6	1,4	VERSICOL
2,7	5,1	1,6	
3	5,2	2	
2,5	3,9	1,1	
3	5,5	1,8	
2,7	5,1	1,9	
3,2	5,9	2,3	
3,3	1,7	0,5	
2,8	4,5	1,3	
3,4	5,4	2,3	
3,8	6,7	2,2	
3,3	4,7	1,6	
3,3	5,7	2,5	
3	6,6	2,1	
2,5	4,5	1,7	
3,5	1,3	0,2	
3	5,2	2,3	
3,2	4,7	1,4	
3,2	4,5	1,5	
2,8	4	1,3	
3,1	1,6	0,2	SETOSA



7. Záložka *Detaily*: zvolí se zde *Metoda* diskriminace a zadají se hodnoty tolerance, pokud je jiná než defaultně zadaná.

3,6	1,4	0,1	SETOSA
3	4,5	1,5	VERSICOL
3,8	6,4	2	VIRGINIC
3,2	1,3	0,2	SETOSA
3,3	5,7	2,1	VIRGINIC
3,5	1,6	0,6	SETOSA
2,6	4	1,2	VERSICOL

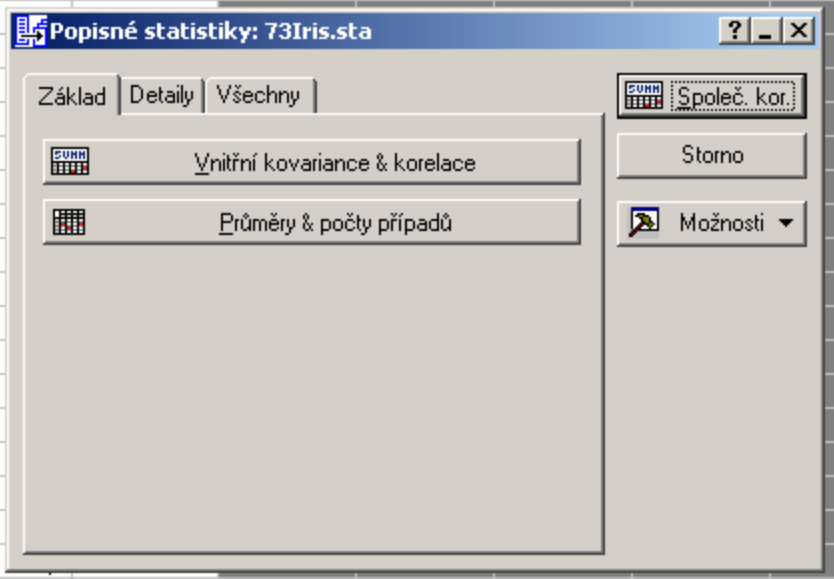
3,3	1,4	0,2	SETOSA
2,8	5,6	2,2	VIRGINIC
2,8	4,6	1,5	VERSICOL
3,1	5,6	2,4	VIRGINIC
2,8	5,1	1,5	VIRGINIC
3,4	1,4	0,3	SETOSA
3,1	5,1	2,3	VIRGINIC
2,2	4,5	1,5	VERSICOL
3,2	4,8	1,8	VERSICOL
3,6	1	0,2	SETOSA
3	4,6	1,4	VERSICOL
2,7	5,1	1,6	
3	5,2	2	
2,5	3,9	1,1	
3	5,5	1,8	
2,7	5,1	1,9	
3,2	5,9	2,3	
3,3	1,7	0,5	
2,8	4,5	1,3	
3,4	5,4	2,3	
3,8	6,7	2,2	
3,3	4,7	1,6	
3,3	5,7	2,5	
3	6,6	2,1	
2,5	4,5	1,7	
3,5	1,3	0,2	
3	5,2	2,3	
3,2	4,7	1,4	
3,2	4,5	1,5	
2,8	4	1,3	
3,1	1,6	0,2	SETOSA
3	5,1	1,8	VIRGINIC
2,4	3,8	1,1	VERSICOL
2,5	5	1,9	VIRGINIC
3,2	5,3	2,3	VIRGINIC



8. V záložce *Popisné statistiky* klikneme na *Zobrazit popisné statistiky*.

3,2	1,3	0,2	SETOSA
3,3	5,7	2,1	VIRGINIC
3,5	1,6	0,6	SETOSA

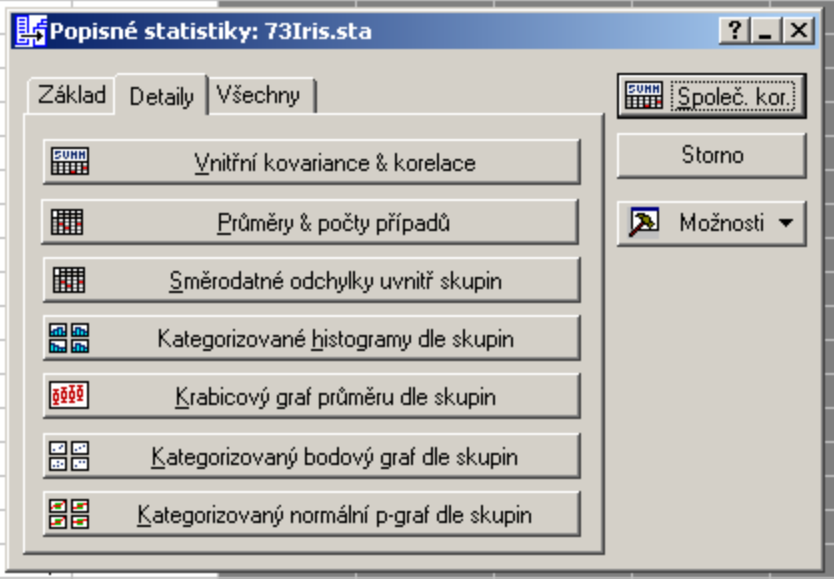
3,3	1,4	0,2	SETOSA
2,8	5,6	2,2	VIRGINIC
2,8	4,6	1,5	VERSICOL
3,1	5,6	2,4	VIRGINIC
2,8	5,1	1,5	VIRGINIC
3,4	1,4	0,3	SETOSA
3,1	5,1	2,3	VIRGINIC
2,2	4,5	1,5	VERSICOL
3,2	4,8	1,8	VERSICOL
3,6	1	0,2	SETOSA
3	4,6	1,4	VERSICOL
2,7	5,1	1,6	VERSICOL
3	5,2		
2,5	3,9		
3	5,5		
2,7	5,1		
3,2	5,9		
3,3	1,7		
2,8	4,5		
3,4	5,4		
3,8	6,7		
3,3	4,7		
3,3	5,7		
3	6,6		
2,5	4,5		
3,5	1,3		
3	5,2		
3,2	4,7		
3,2	4,5		
2,8	4	1,3	VERSICOL
3,1	1,6	0,2	SETOSA



9. Záložka *Základ* přináší dvě možnosti volby. Zvolíme však nejprve záložku *Detaily*.

3	4,5	1,5	VERSICOL
3,8	6,4	2	VIRGINIC
3,2	1,3	0,2	SETOSA
3,3	5,7	2,1	VIRGINIC
3,5	1,6	0,6	SETOSA

3,3	1,4	0,2	SETOSA
2,8	5,6	2,2	VIRGINIC
2,8	4,6	1,5	VERSICOL
3,1	5,6	2,4	VIRGINIC
2,8	5,1	1,5	VIRGINIC
3,4	1,4	0,3	SETOSA
3,1	5,1	2,3	VIRGINIC
2,2	4,5	1,5	VERSICOL
3,2	4,8	1,8	VERSICOL
3,6	1	0,2	SETOSA
3	4,6	1,4	VERSICOL
2,7	5,1	1,6	VERSICOL
3	5,2		
2,5	3,9		
3	5,5		
2,7	5,1		
3,2	5,9		
3,3	1,7		
2,8	4,5		
3,4	5,4		
3,8	6,7		
3,3	4,7		
3,3	5,7		
3	6,6		
2,5	4,5		
3,5	1,3		
3	5,2		
3,2	4,7		
3,2	4,5		
2,8	4	1,3	VERSICOL
3,1	1,6	0,2	SETOSA
3	5,1	1,8	VIRGINIC
2,4	3,8	1,1	VERSICOL
3,3	5,7	2,1	VIRGINIC
3,5	1,6	0,6	SETOSA



10. Záložka *Detaily* přináší sedm voleb: zvolíme nejprve *Průměry a počty případů.*

STATISTICA Cz - [PS 1* - Průměry (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

PS 1*
- Diskriminační analýza (73Iris.
  - Dialog popisných statistik
    - Kovariance (73Iris.st
    - Vnitřní korelace (73Ir
    - Průměry (73Iris.sta)

Trida	Průměry (73Iris.sta)				
	SEPALLENGTH	SEPALWIDTH	PETALLENGTH	PETALWIDTH	N platn.
SETOSA	5,006000	3,428000	1,462000	0,246000	50
VERSICOL	5,936000	2,770000	4,260000	1,326000	50
VIRGINIC	6,588000	2,974000	5,552000	2,026000	50
Všechny	5,843333	3,057333	3,758000	1,199333	150

11. Tabulka *Průměry* ukazuje průměry čtyř měr okvětních lístků ve třech třídách.

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

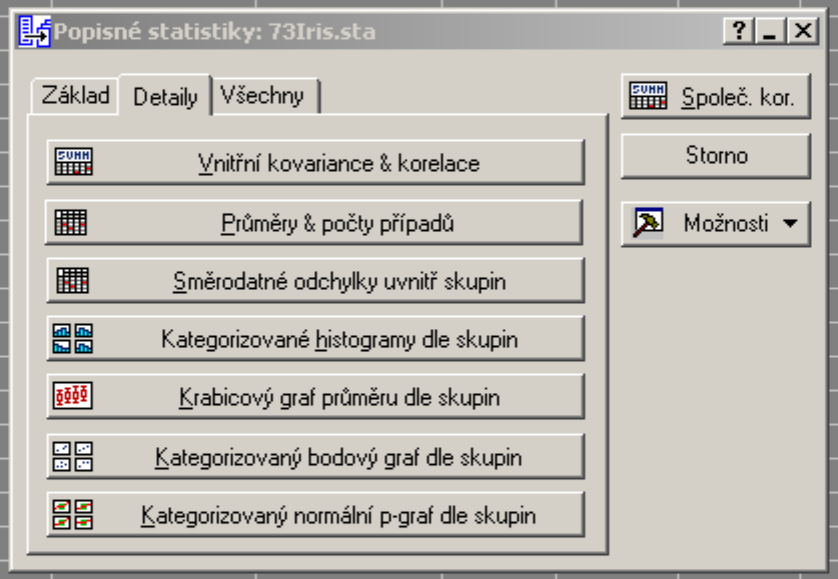
Iris Versicolor

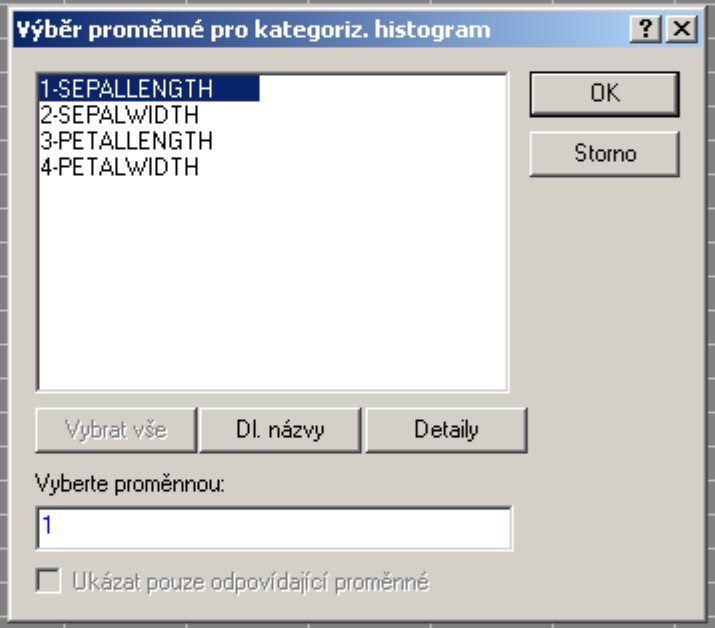
Iris Virginica

Iris Virginica

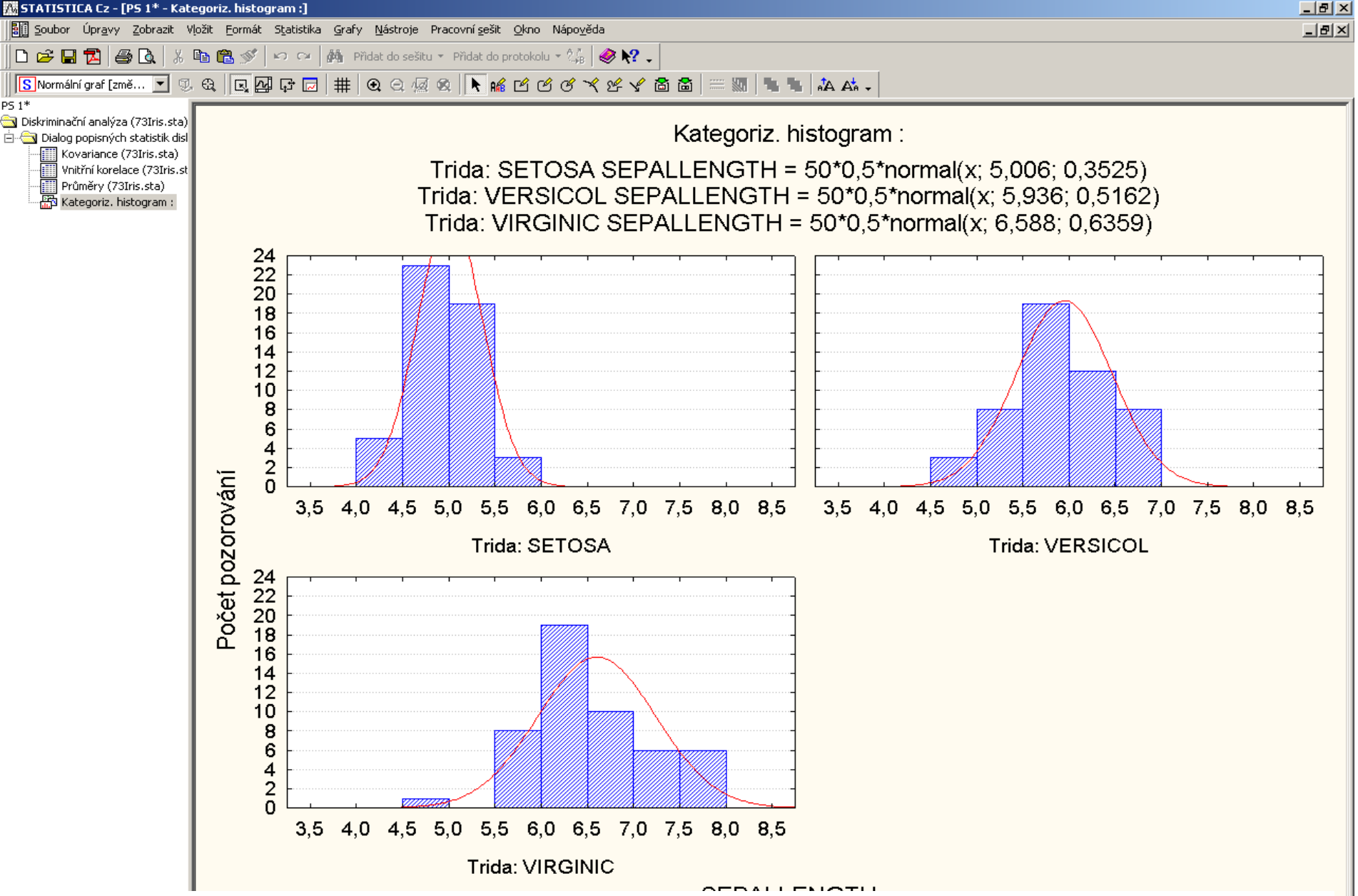
Iris Virginica

SEPALLENGTH	SEPALWIDTH	PETALLENGTH	PETALWIDTH	N platn.
5,006000	3,428000	1,462000	0,246000	50
5,936000	2,770000	4,260000	1,326000	50
6,588000	2,974000	5,552000	2,026000	50
5,843333	3,057333	3,758000	1,199333	150





12. Zvolíme první znak, *SEPALLENGTH*, pro kategorizovaný histogram čili celkem 3 histogramy, pro každou třídu kosatců jeden histogram.



13. Obrázek odkrývá, zda tento znak vykazuje normální rozdělení, a to porovnáním těsnosti proložení červené křivky histogramem.

STATISTICA Cz - [PS 1* - Kategoriz. histogram :]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Normální graf [změ...

PS 1*

- Diskriminační analýza (73Iris.sta)
  - Dialog popisných statistik disl
    - Kovariance (73Iris.sta)
    - Vnitřní korelace (73Iris.st
    - Průměry (73Iris.sta)
    - Kategoriz. histogram :

Kategoriz. histogram :

Trida: SETOSA SEPALLENGTH = 50*0,5*normal(x; 5,006; 0,3525)
Trida: VERSICOL SEPALLENGTH = 50*0,5*normal(x; 5,936; 0,5162)
Trida: VIRGINIC SEPALLENGTH = 50*0,5*normal(x; 6,588; 0,6359)

Počet pozorování

24 22 20 18 16 14 12 10 8 6 4 2 0

3,5 4,0 4,5 5,0 5,5 6,0 6,5 7,0 7,5 8,0 8,5

Trida: VIRGINIC

SEPALLENGTH

Popisné statistiky: 73Iris.sta

Základ | Detaily | Všechny

- Vnitřní kovariance & korelace
- Průměry & počty případů
- Směrodatné odchylky uvnitř skupin
- Kategorizované histogramy dle skupin
- Krabicový graf průměru dle skupin
- Kategorizovaný bodový graf dle skupin
- Kategorizovaný normální p-graf dle skupin

Společ. kor. | Storno | Možnosti

Výběr proměnné pro kategoriz. histogram

1-SEPALLENGTH
2-SEPALWIDTH
3-PETALLENGTH
4-PETALWIDTH

OK | Storno

Vybrat vše | Dl. názvy | Detaily

Vyberte proměnnou: 2

Ukázat pouze odpovídající proměnné

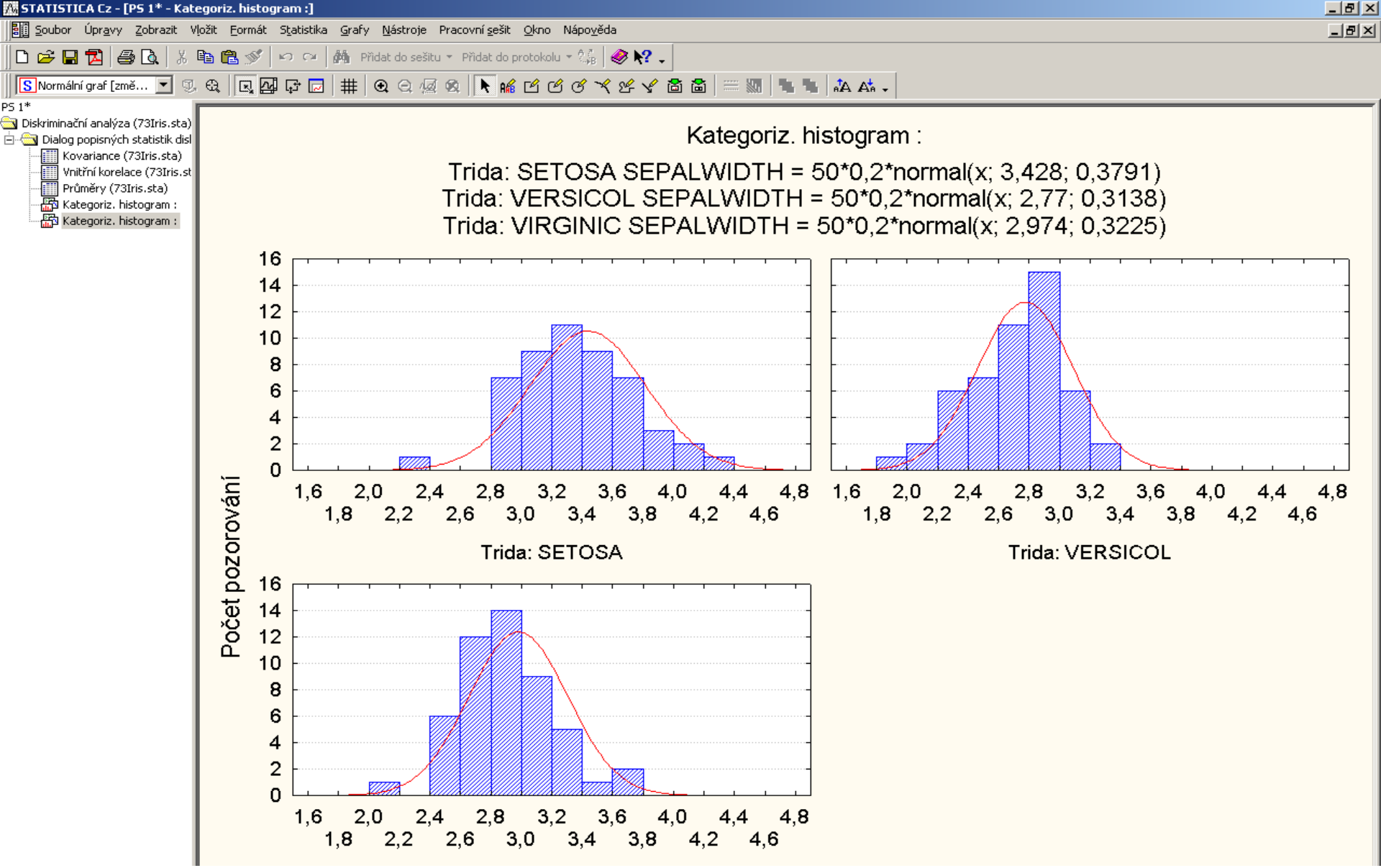
14. Zadá se druhý znak, *SEPALWIDTH.*

Kovariance (73Iris.sta) | Vnitřní korelace (73Iris.sta) | Průměry (73Iris.sta) | Kategoriz. histogram :

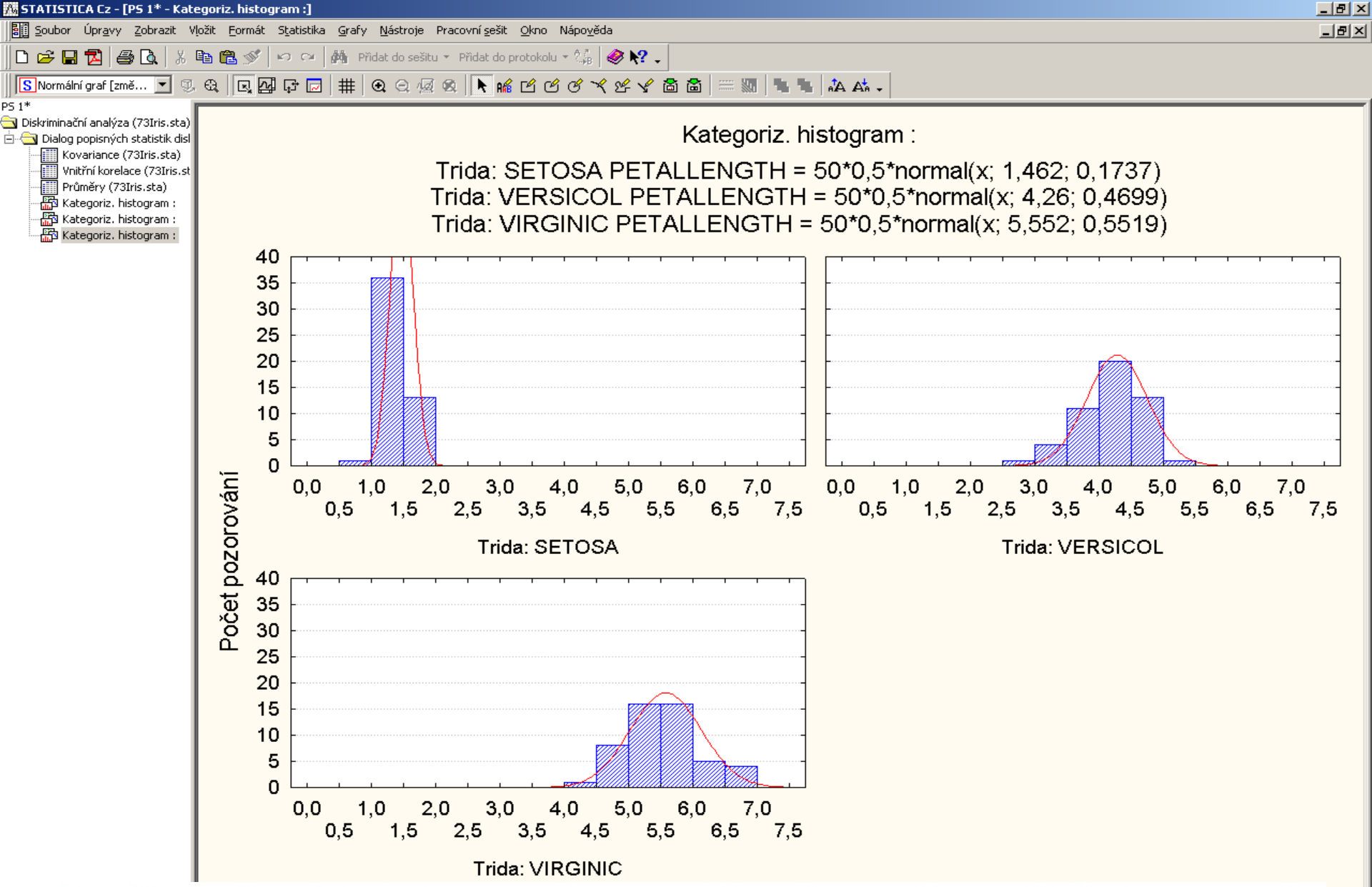
Popisné statistiky: 73...

Připravena

Start | Doručená pošt... | Total Command... | Zoner Photo St... | Anglicko-český ... | Microsoft Powe... | STATISTICA ... | 9:05



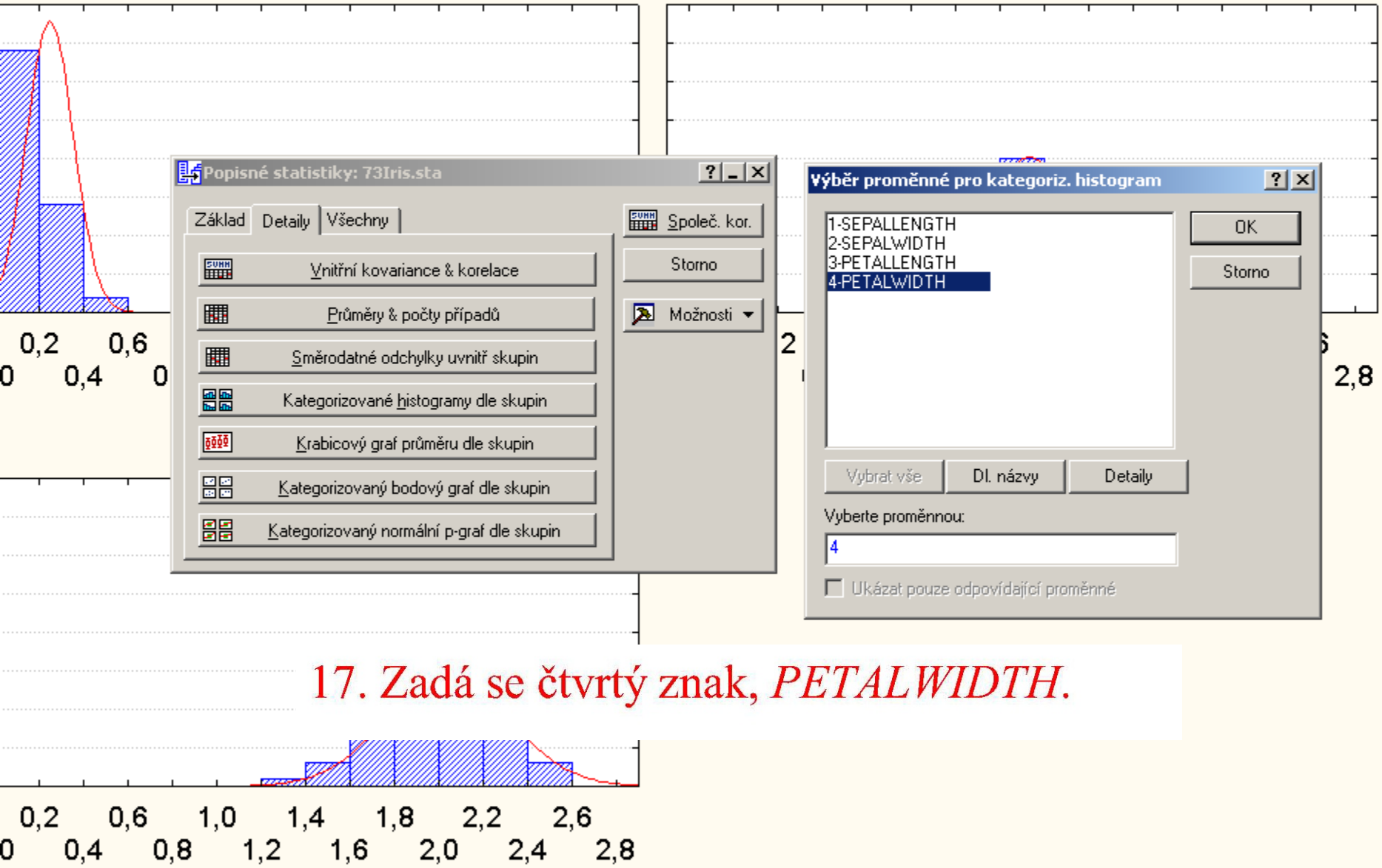
15. Obrázek odkrývá normální rozdělení ve všech třídách kosatců: *SEPALWIDTH.*



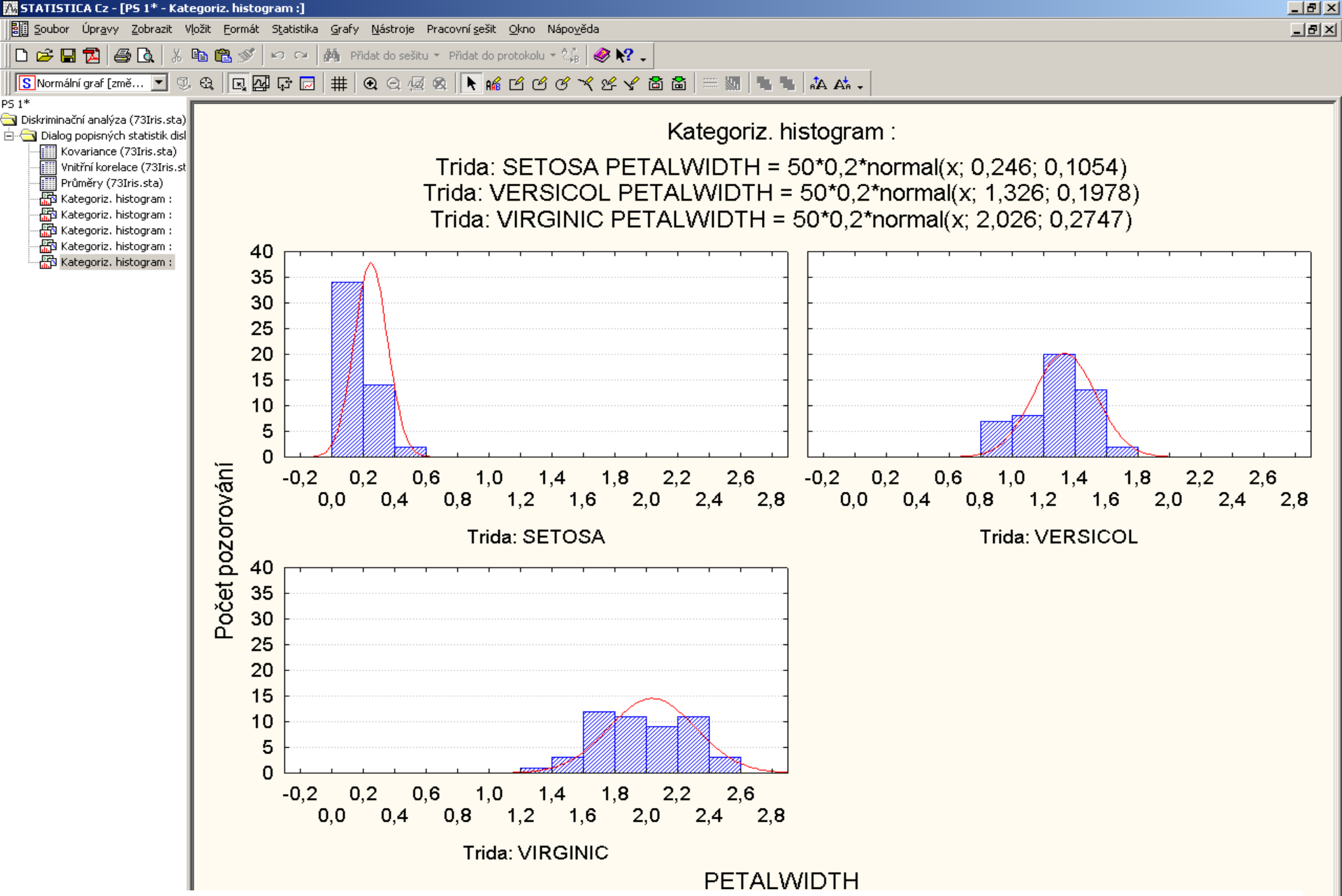
16. Obrázek odkrývá normální rozdělení ve všech třídách kosatců: *PETEALLENGTH.*

Kategoriz. histogram :

Trida: SETOSA PETALWIDTH = 50*0,2*normal(x; 0,246; 0,1054)
Trida: VERSICOL PETALWIDTH = 50*0,2*normal(x; 1,326; 0,1978)
Trida: VIRGINIC PETALWIDTH = 50*0,2*normal(x; 2,026; 0,2747)

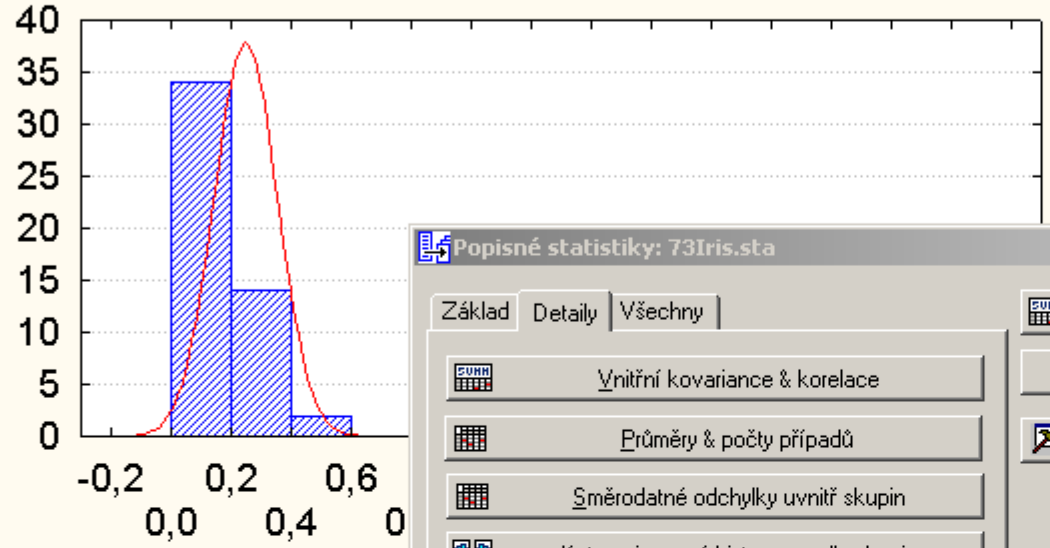


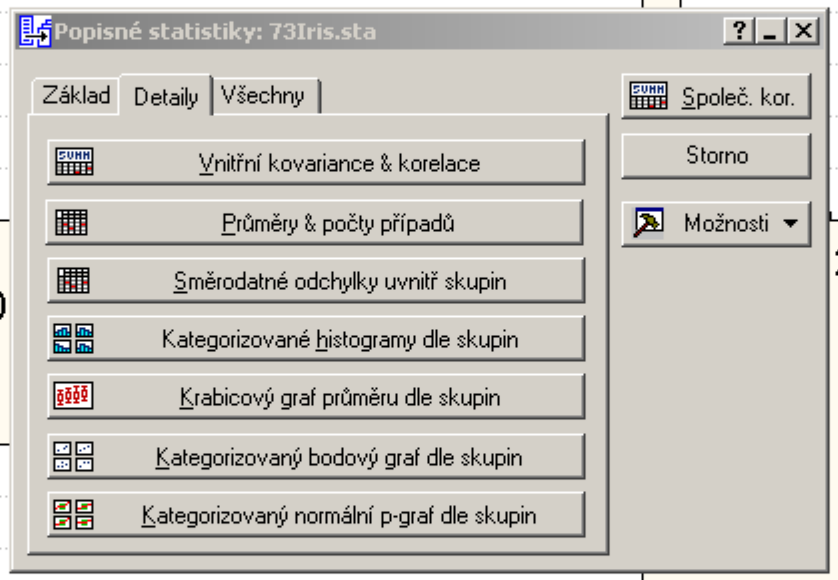
17. Zadá se čtvrtý znak, *PETALWIDTH*.

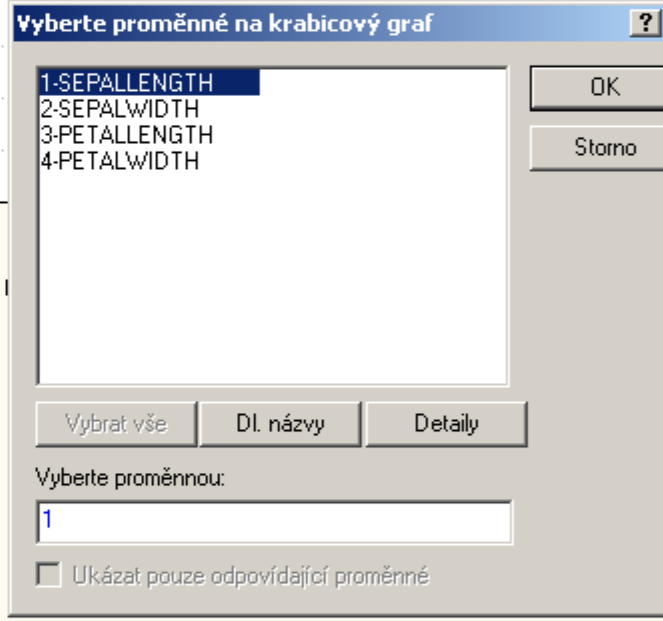


18. Obrázek odkrývá normální rozdělení ve všech třídách kosatců: *PETALWIDTH.*

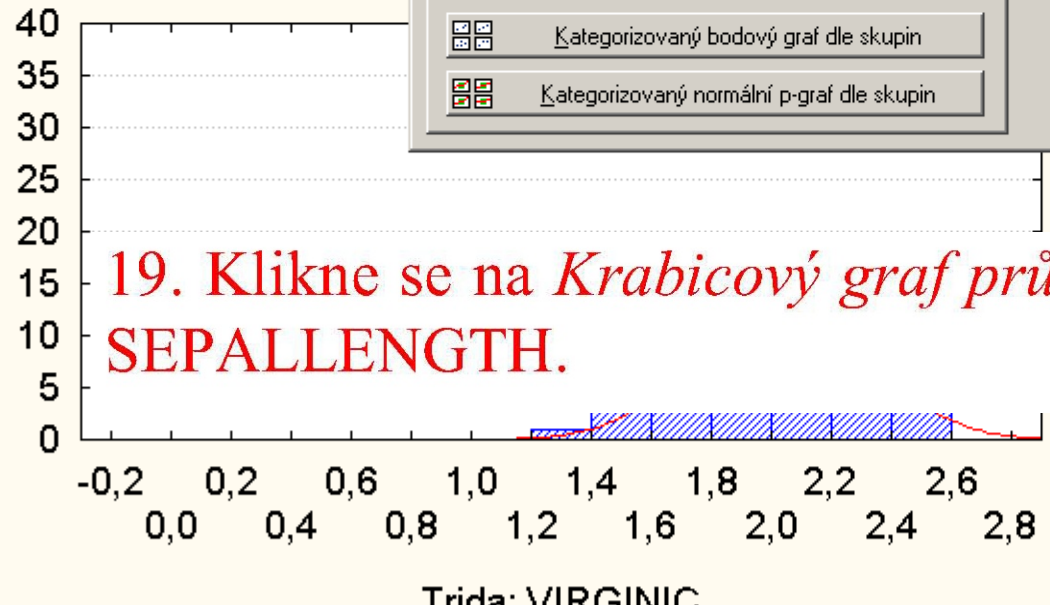
Trida: SETOSA PETALWIDTH = 50*0,2*normal(x; 0,246; 0,1054)
Trida: VERSICOL PETALWIDTH = 50*0,2*normal(x; 1,326; 0,1978)
Trida: VIRGINIC PETALWIDTH = 50*0,2*normal(x; 2,026; 0,2747)





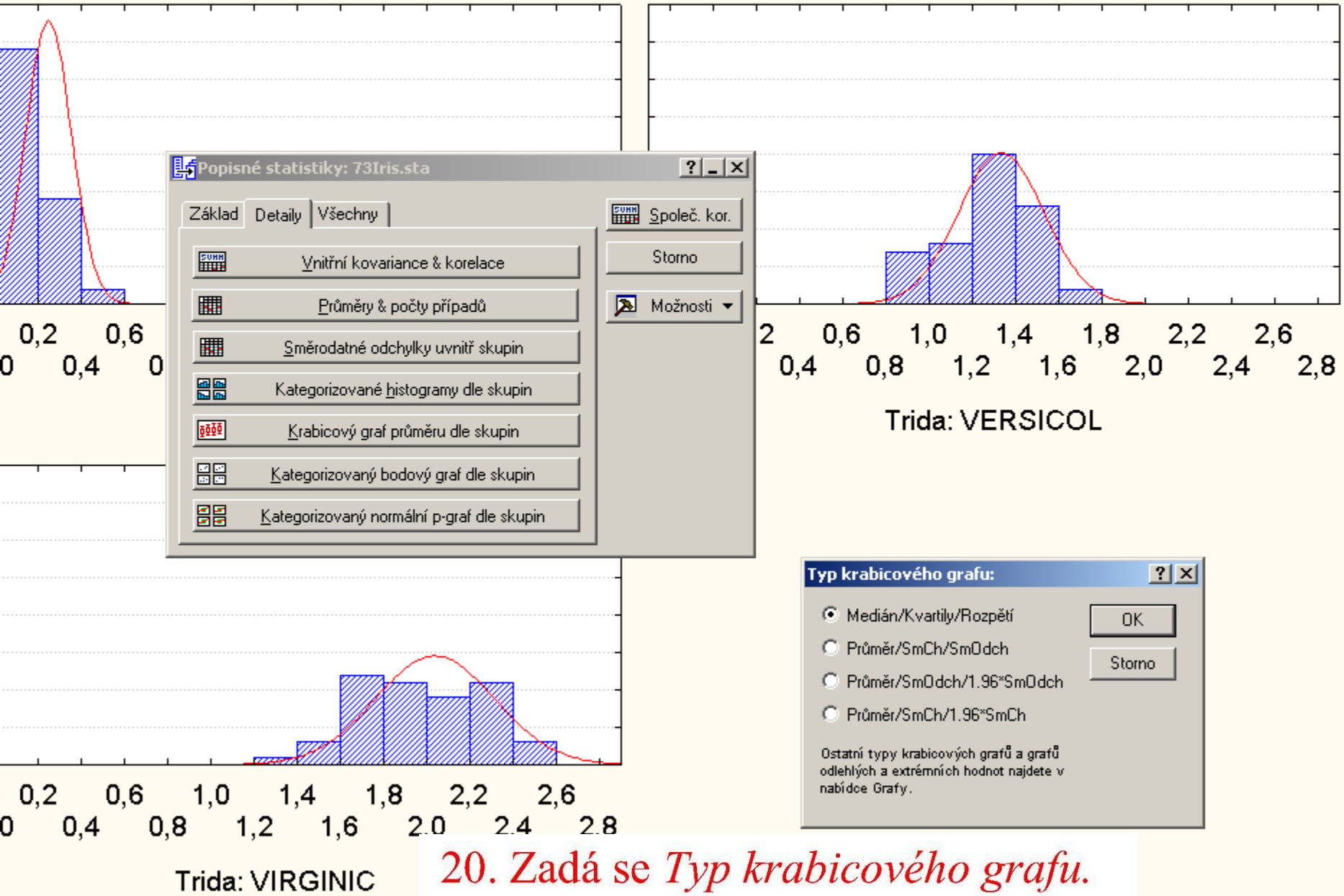


19. Klikne se na *Krabicový graf průměru dle skupin* pro první znak: SEPALLENGTH.

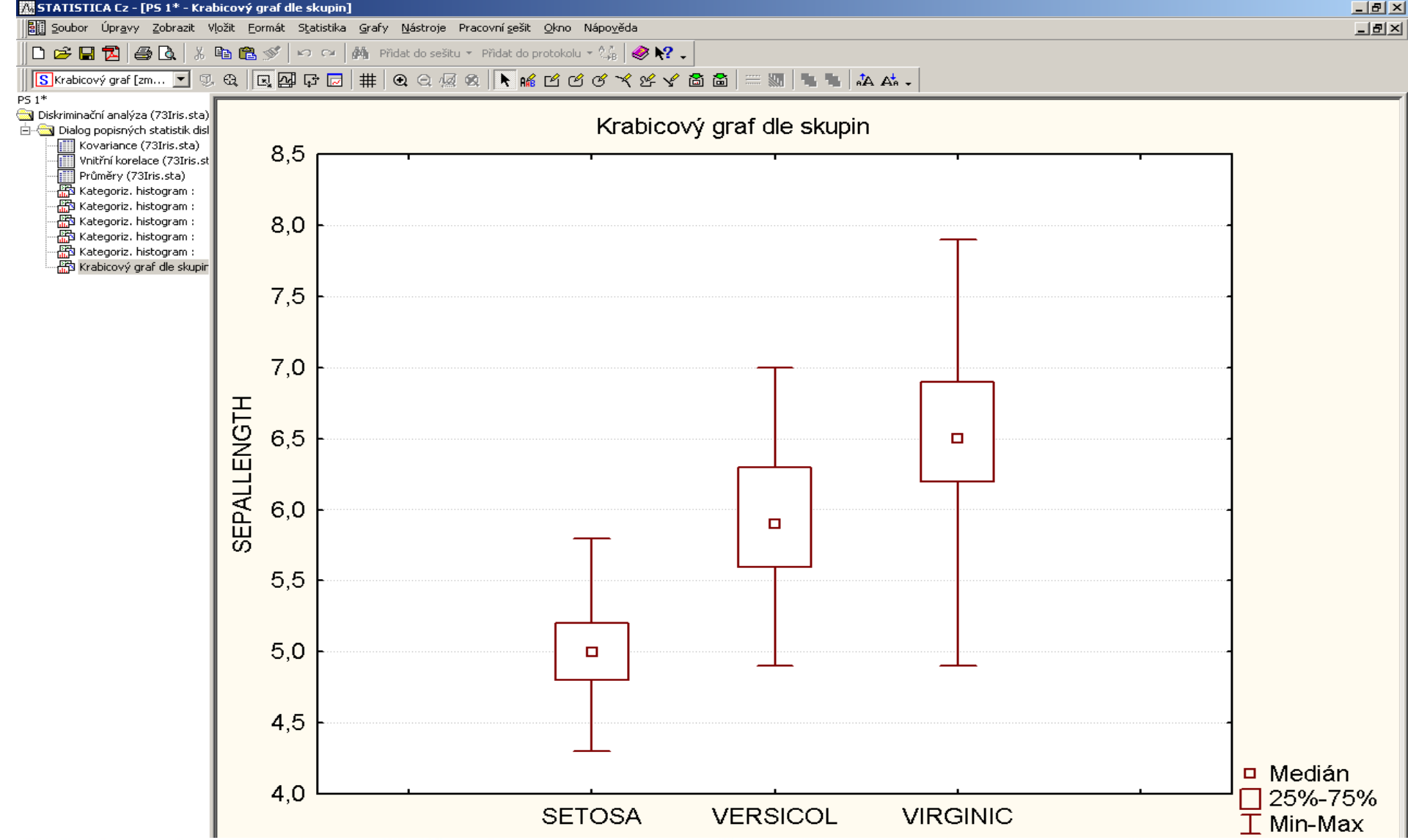


Trida: VIRGINIC

Trida: SETOSA PETALWIDTH = 50*0,2*normal(x; 0,246; 0,1054)
Trida: VERSICOL PETALWIDTH = 50*0,2*normal(x; 1,326; 0,1978)
Trida: VIRGINIC PETALWIDTH = 50*0,2*normal(x; 2,026; 0,2747)



20. Zadá se *Typ krabicového grafu.*



21. *Krabicový graf dle skupin* ukazuje míry polohy, rozptýlení, tvaru a vybočující hodnoty včetně symetrie rozdělení u tří tříd kosatců SETOSA, VERSICOLOR, VIRGINICA.

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

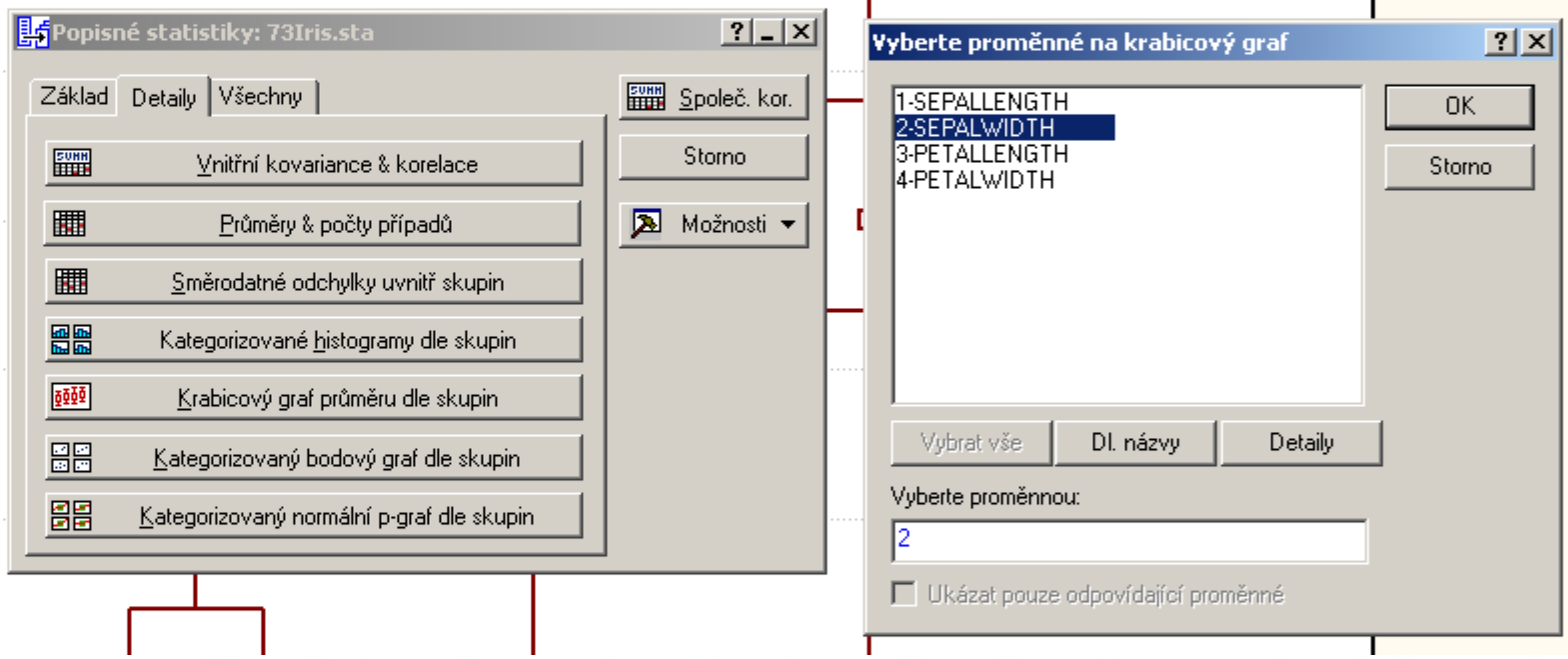
Iris Versicolor

Iris Virginica

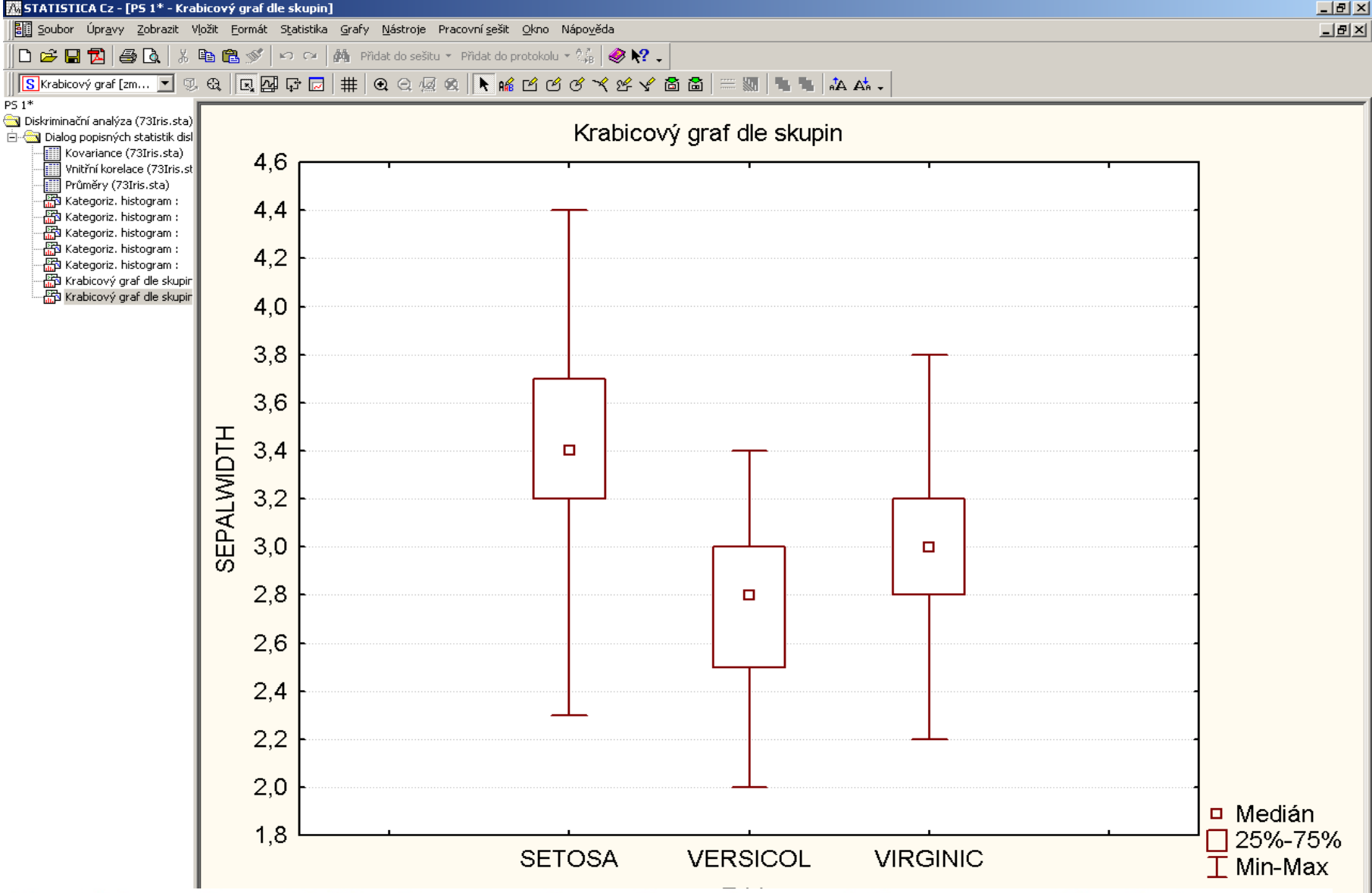
Iris Virginica

Iris Virginica

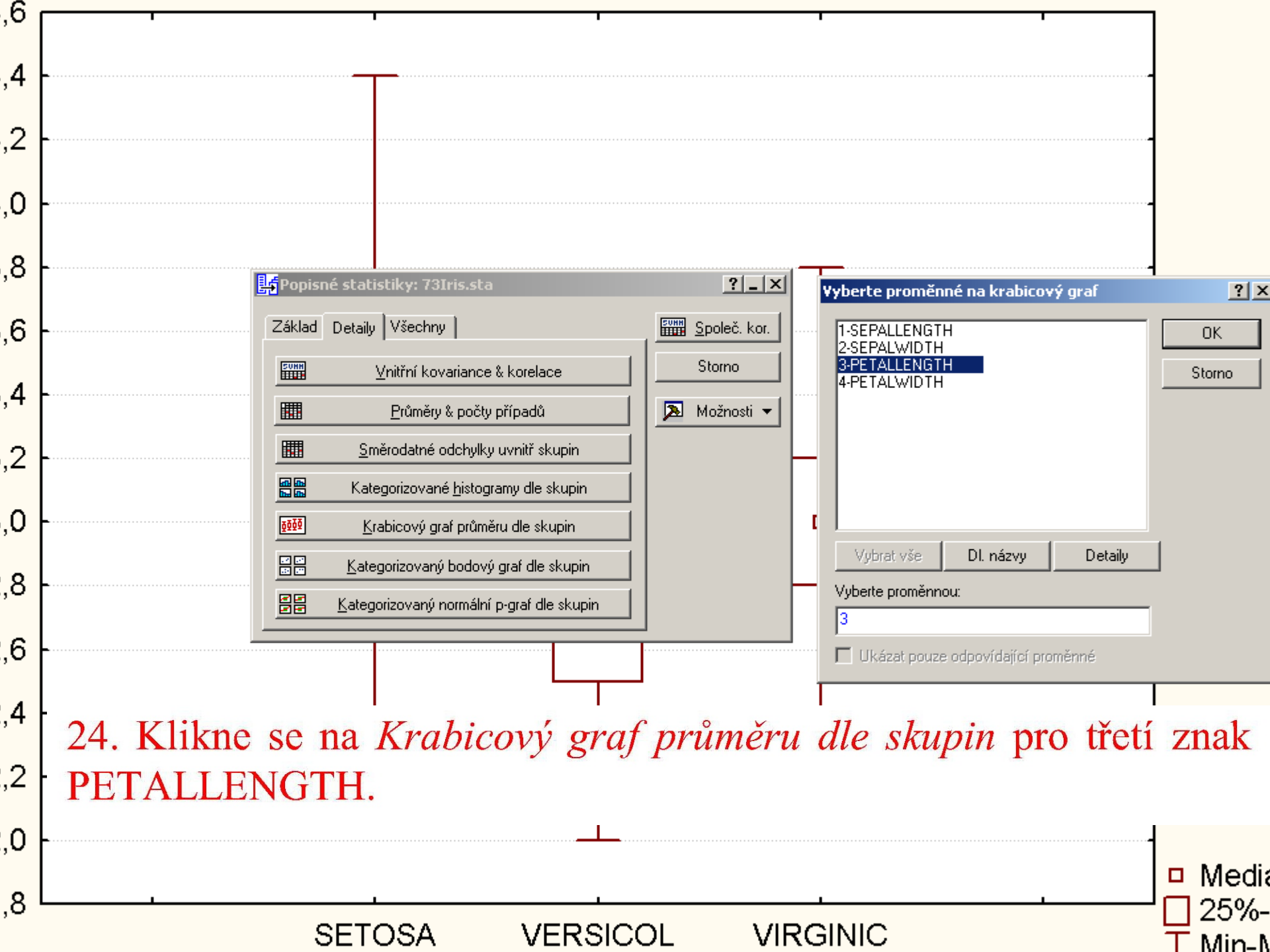
Krabicový graf dle skupin



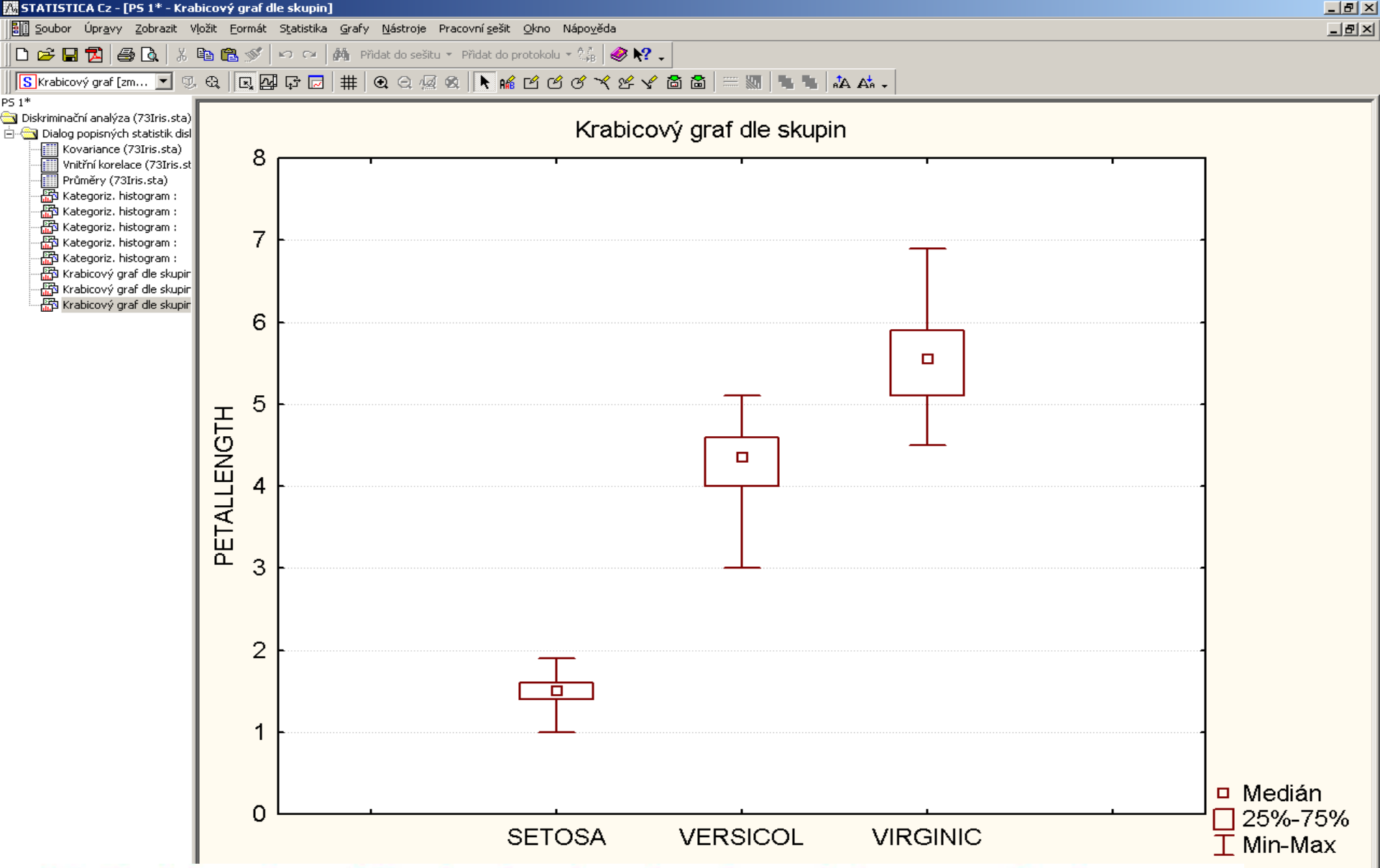
22. Klikne se na *Krabicový graf průměru dle skupin* pro druhý znak: SEPALWIDTH.



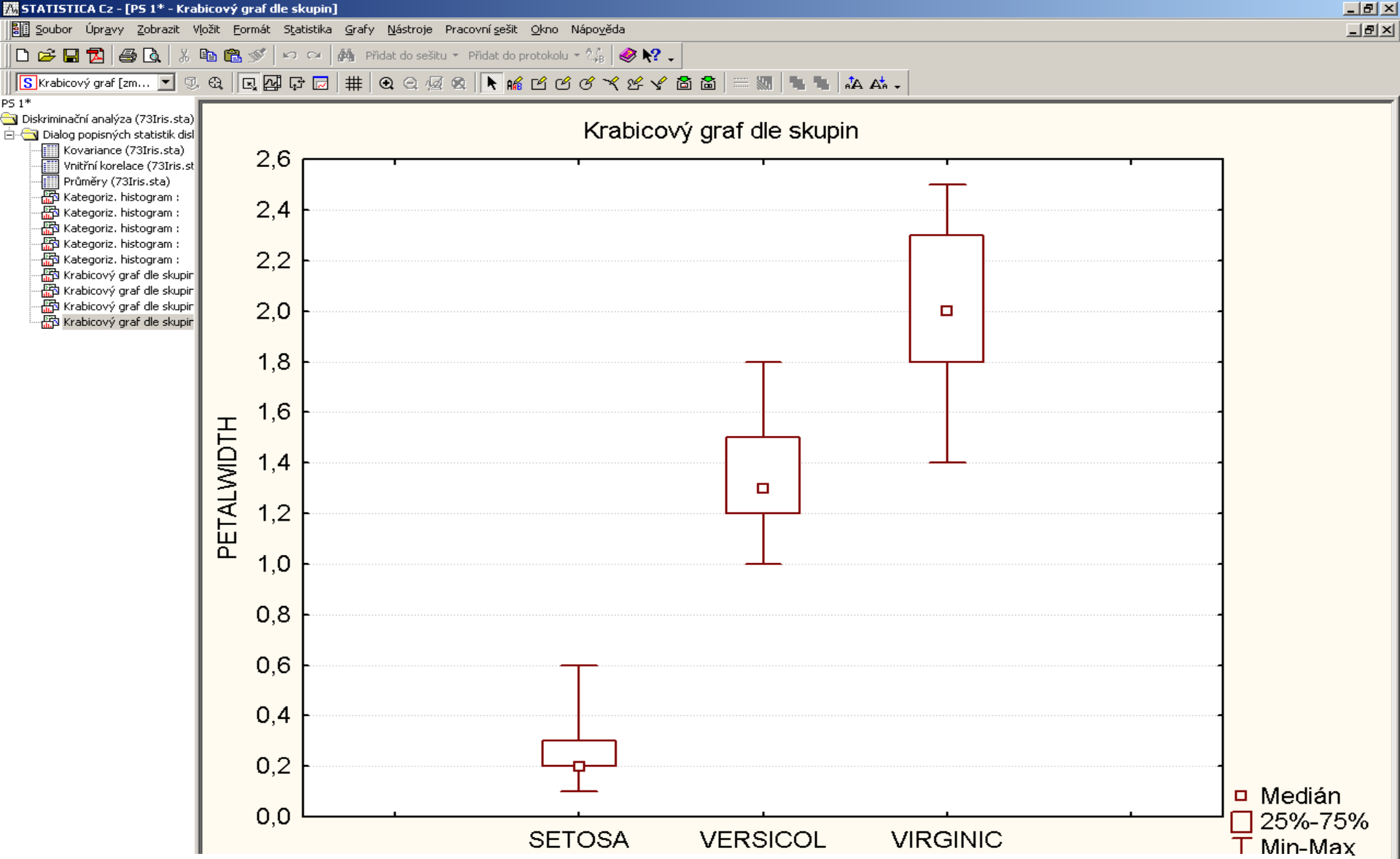
23. *Krabicový graf dle skupin* ukazuje míry polohy, rozptýlení, tvaru a vybočující hodnoty včetně symetrie rozdělení u tří tříd kosatců SETOSA, VERSICOLOR, VIRGINICA.



24. Klikne se na *Krabicový graf průměru dle skupin* pro třetí znak PETALLENGTH.



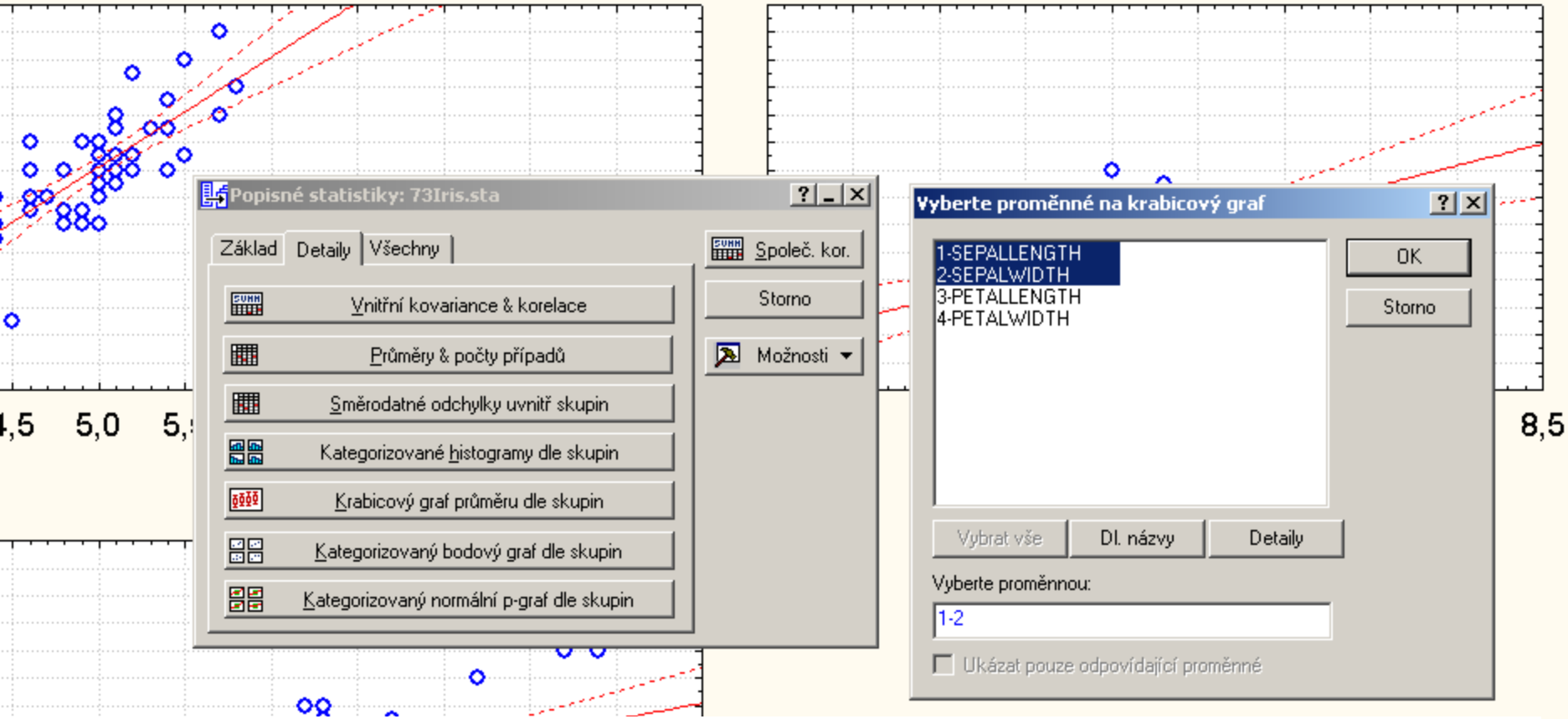
25. Krabicový graf dle skupin ukazuje míry polohy, rozptýlení, tvaru a vybočující hodnoty včetně symetrie rozdělení u tří tříd kosatců SETOSA, VERSICOLOR, VIRGINICA.



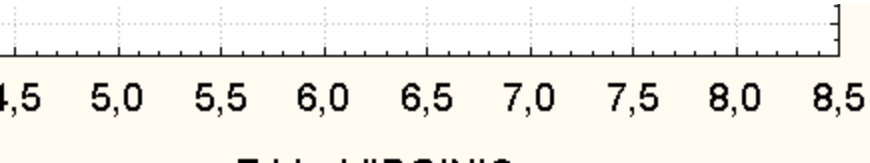
26. Krabicový graf dle skupin ukazuje míry polohy, rozptýlení, tvaru a vybočující hodnoty včetně symetrie rozdělení u tří tříd kosatců SETOSA, VERSICOLOR, VIRGINICA.

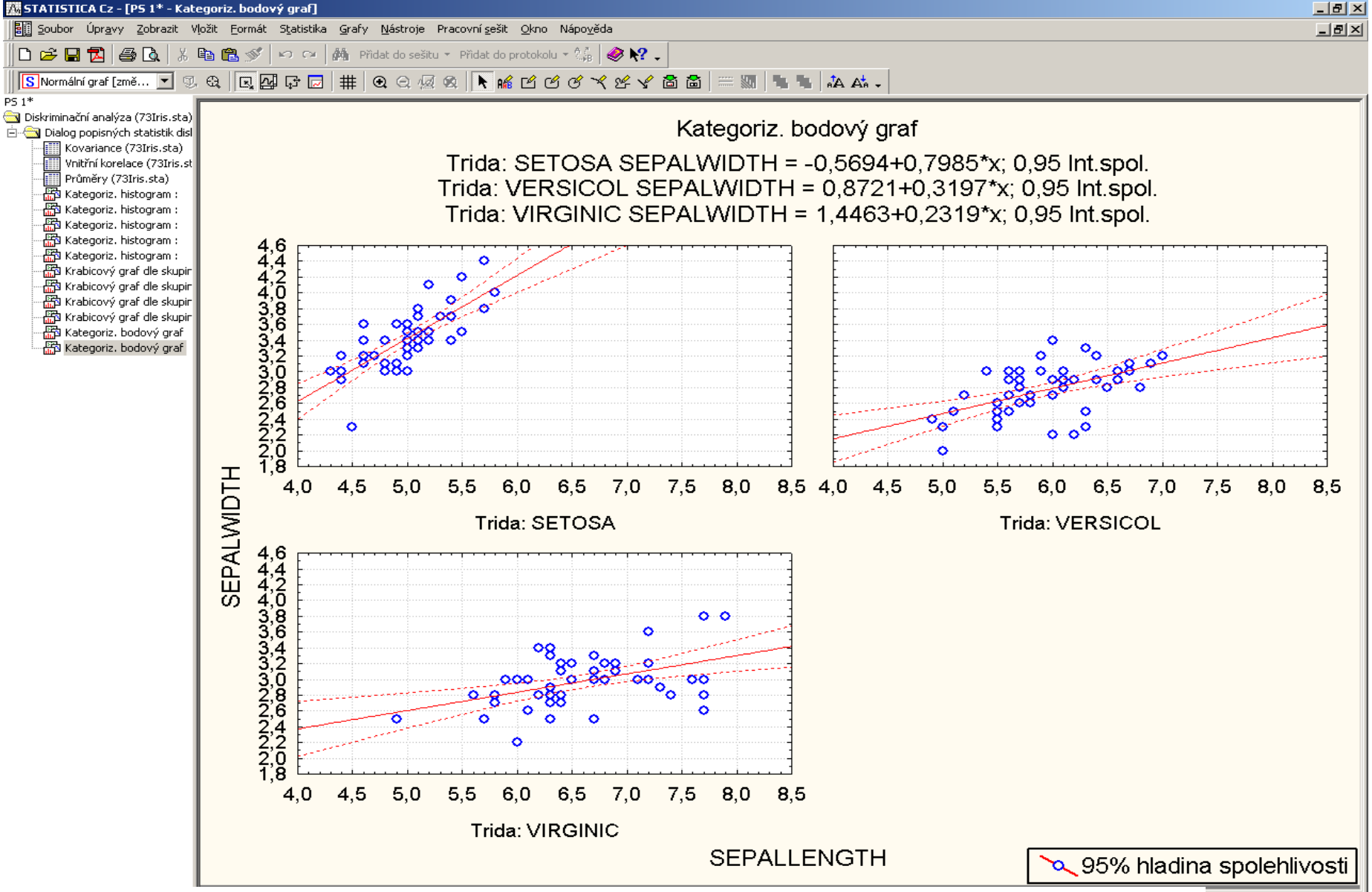
Kategoriz. bodový graf

Trida: SETOSA SEPALWIDTH = -0,5694+0,7985*x; 0,95 Int.spol.
Trida: VERSICOL SEPALWIDTH = 0,8721+0,3197*x; 0,95 Int.spol.
Trida: VIRGINIC SEPALWIDTH = 1,4463+0,2319*x; 0,95 Int.spol.



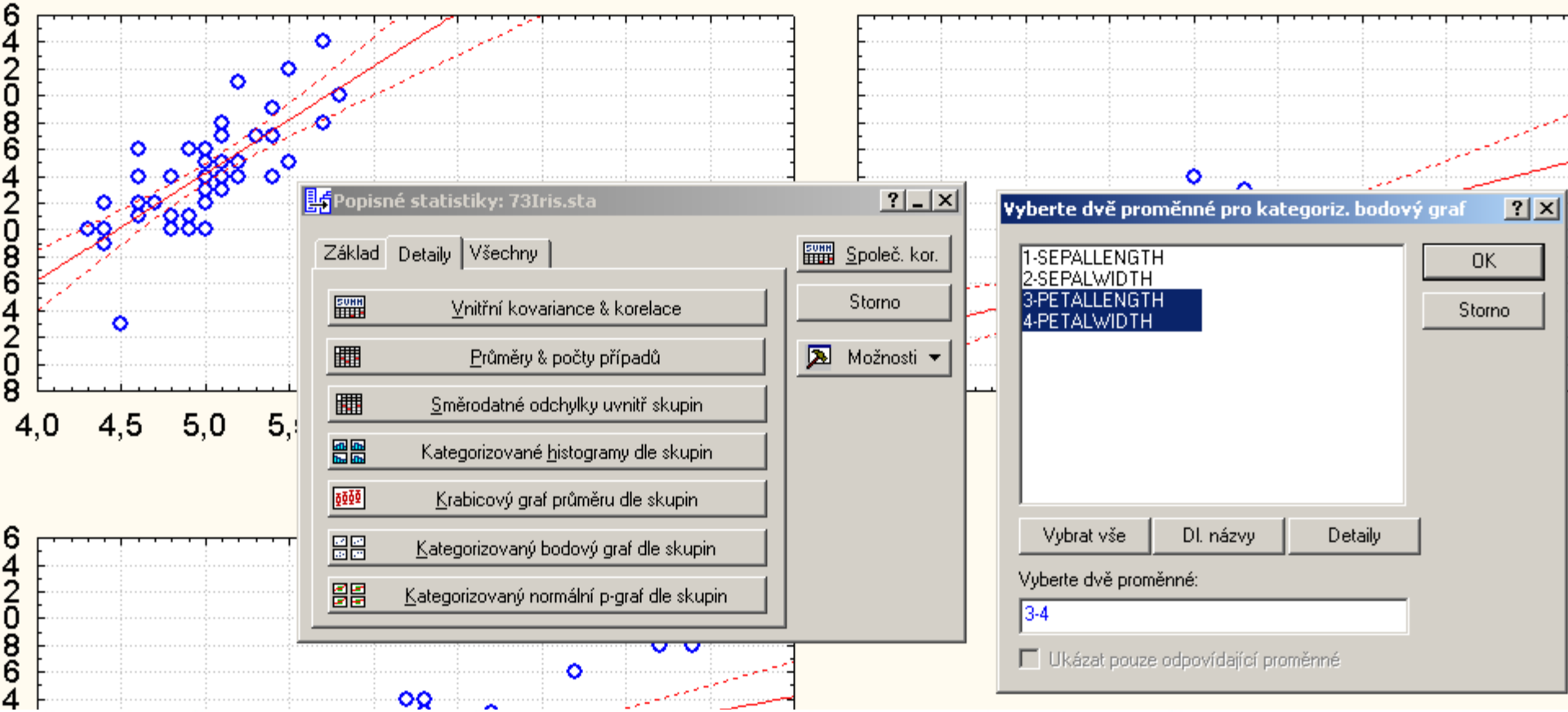
27. Zadá se kategorizovaný bodový rozptylový graf pro první dva znaky: 1. SEPALLENGTH a 2.SEPALWIDTH.



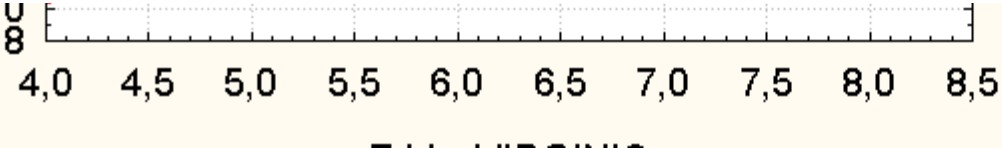


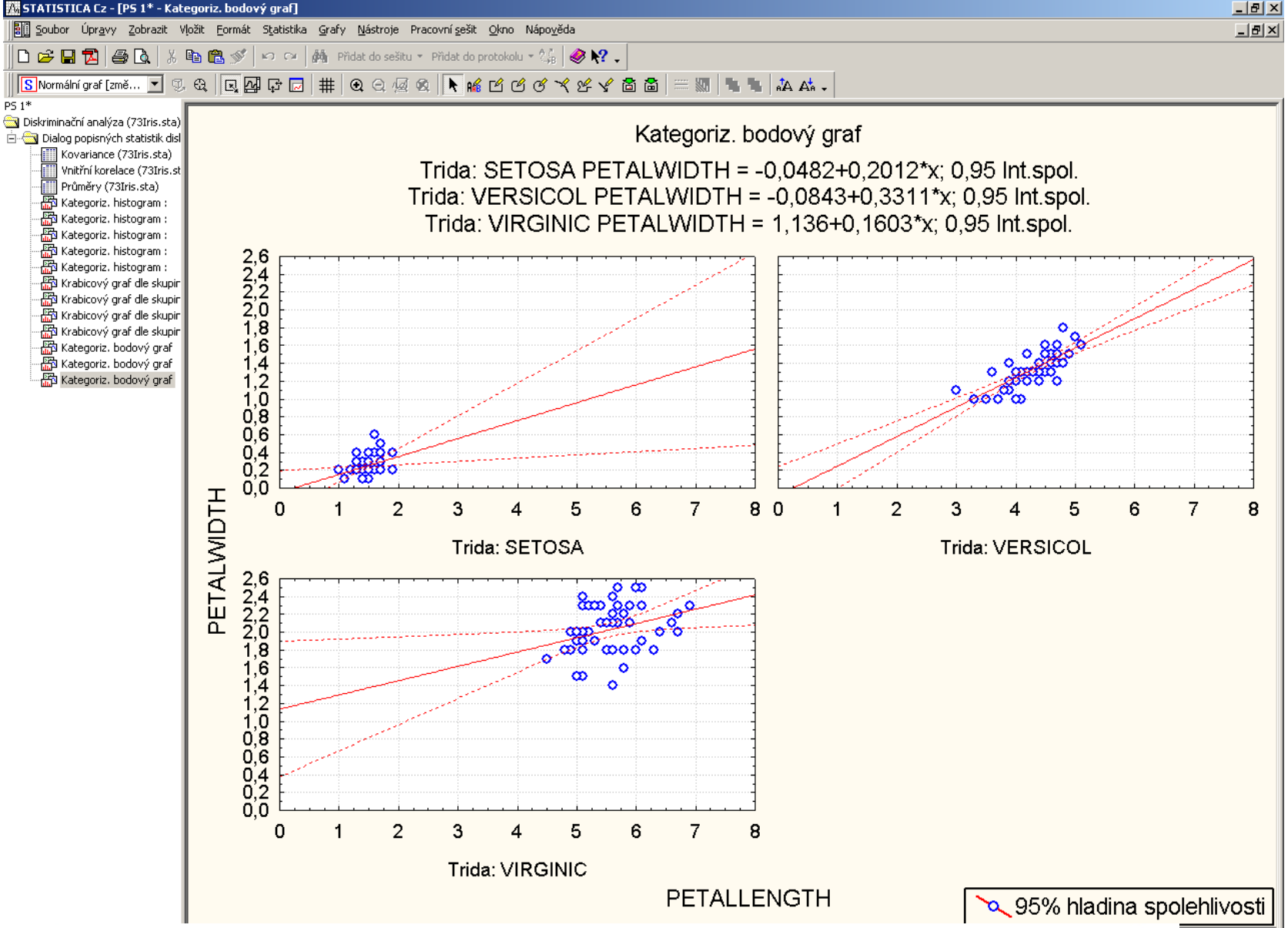
28. Když by bodový rozptylový graf ukazoval pouze **jediný** mrak bodů, pak diskriminační analýza nemá šanci rozdělit kosatce do tříd. Už v tomto grafu jsou vidět zřetelné skupiny bodů.

Trida: SETOSA SEPALWIDTH = -0,5694+0,7985*x; 0,95 Int.spol.
Trida: VERSICOL SEPALWIDTH = 0,8721+0,3197*x; 0,95 Int.spol.
Trida: VIRGINIC SEPALWIDTH = 1,4463+0,2319*x; 0,95 Int.spol.

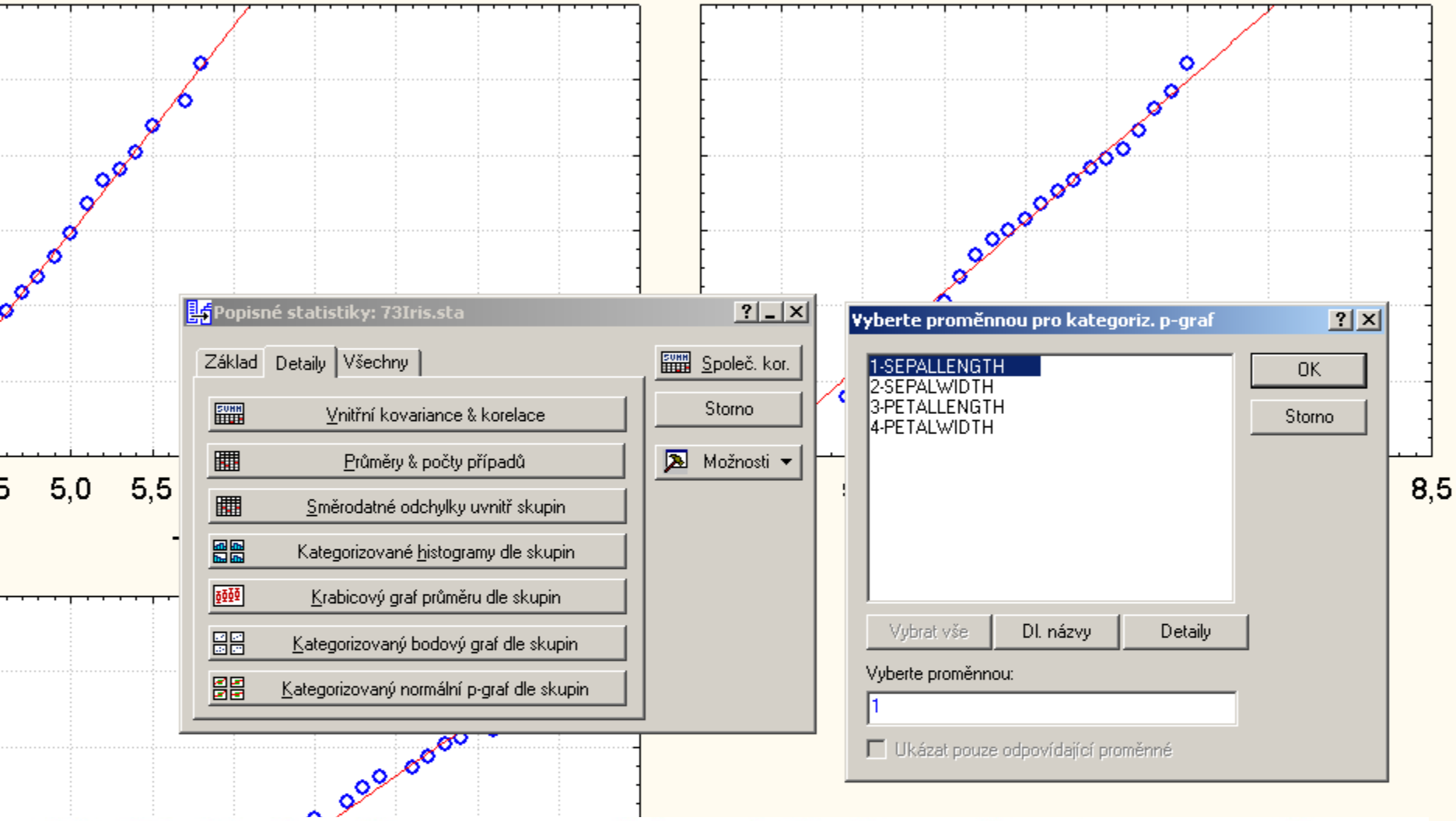


29. Zadá se kategorizovaný bodový rozptylový graf pro druhé dva znaky: 3. PETALLENGTH a 4.PETALWIDTH.



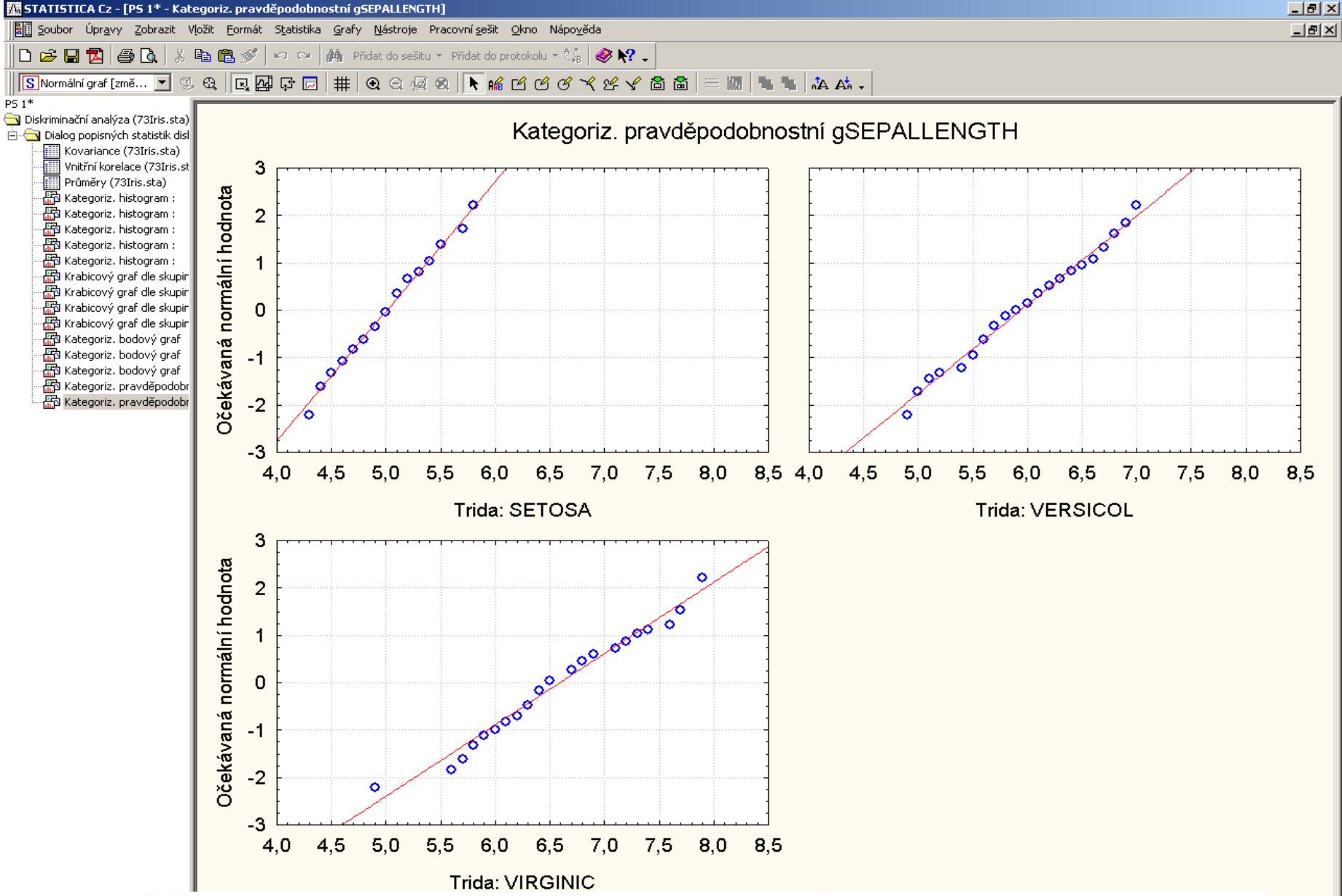


30. Když by bodový rozptylový graf ukazoval pouze **jediný** mrak bodů, pak diskriminační analýza nemá šanci rozdělit kosatce do tříd. Už v tomto grafu jsou vidět zřetelné skupiny bodů.

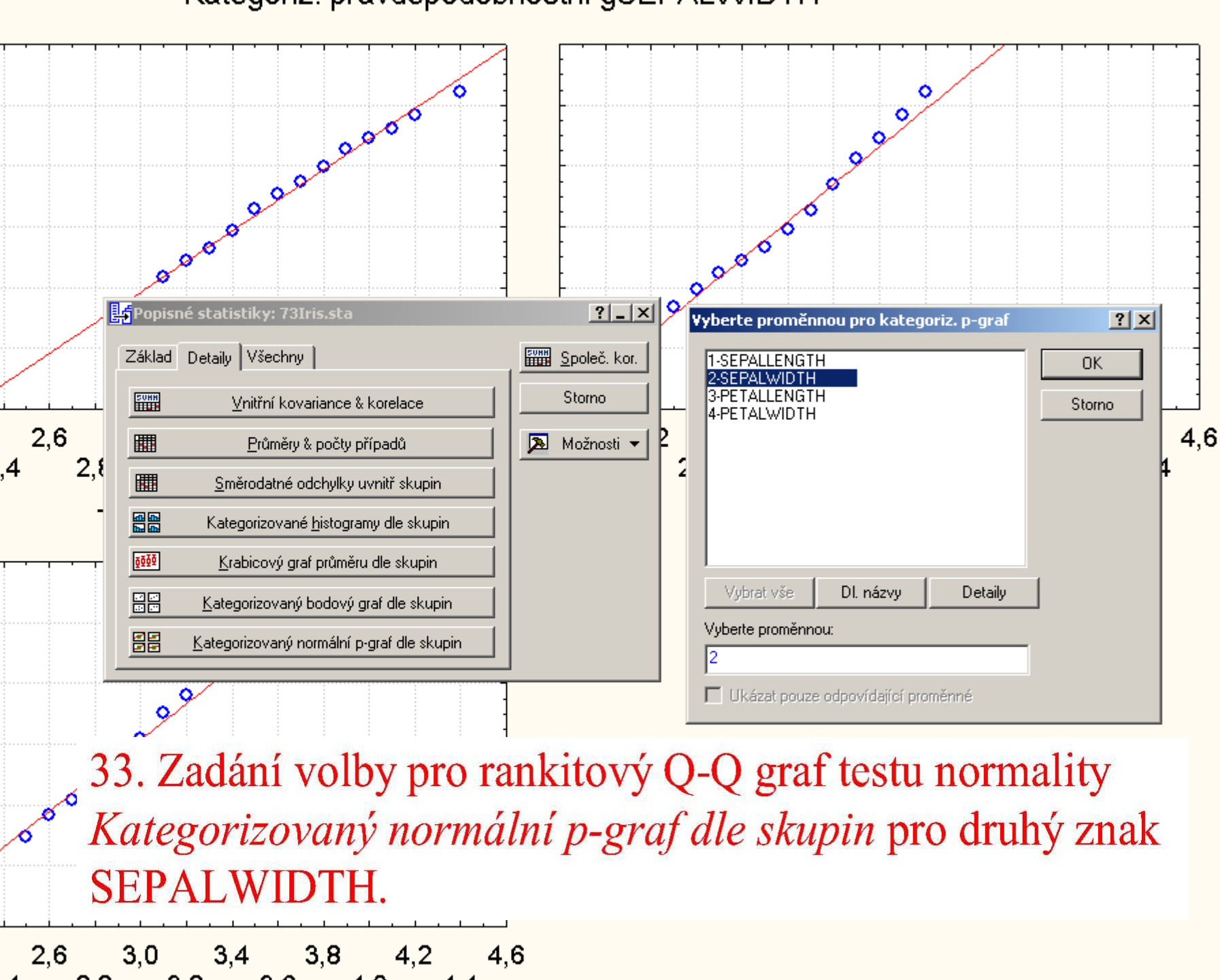


31. Zadání volby pro rankitový Q-Q graf testu normality *Kategorizovaný normální p-graf dle skupin* pro první znak SEPALLENGTH.

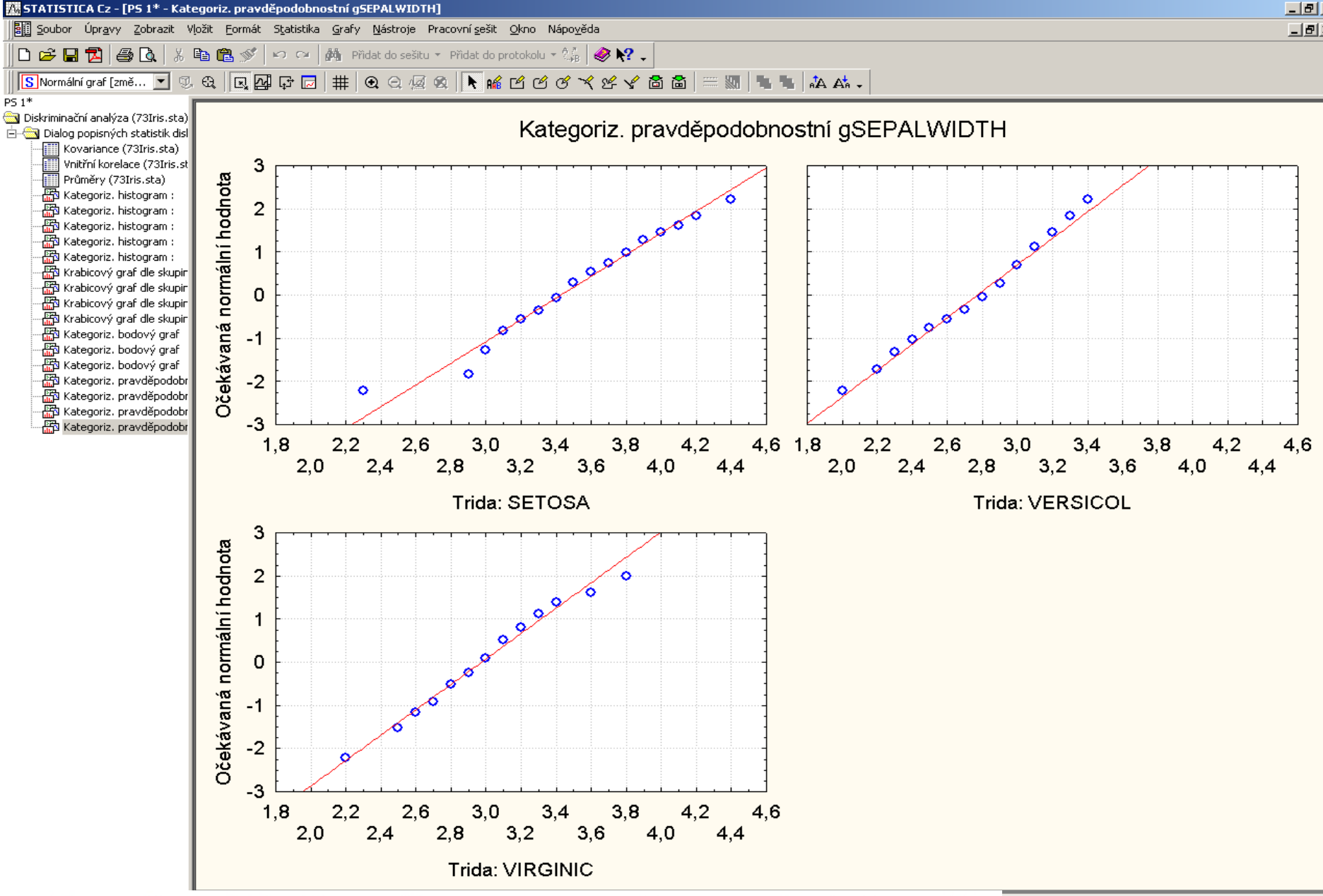
5 5,0 5,5 6,0 6,5 7,0 7,5 8,0 8,5



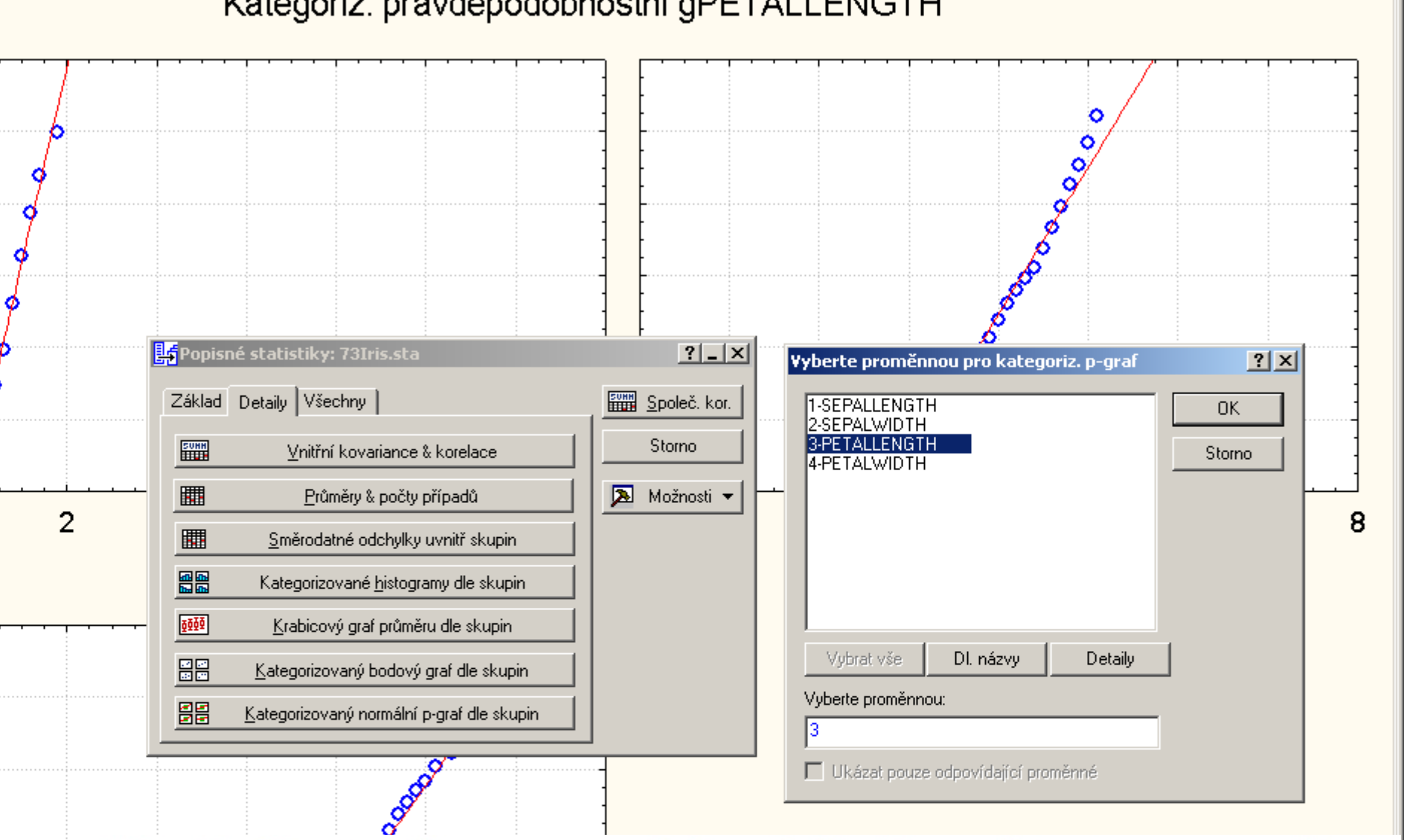
32. Rankitový Q-Q graf testu normality pro první znak SEPALLENGTH.



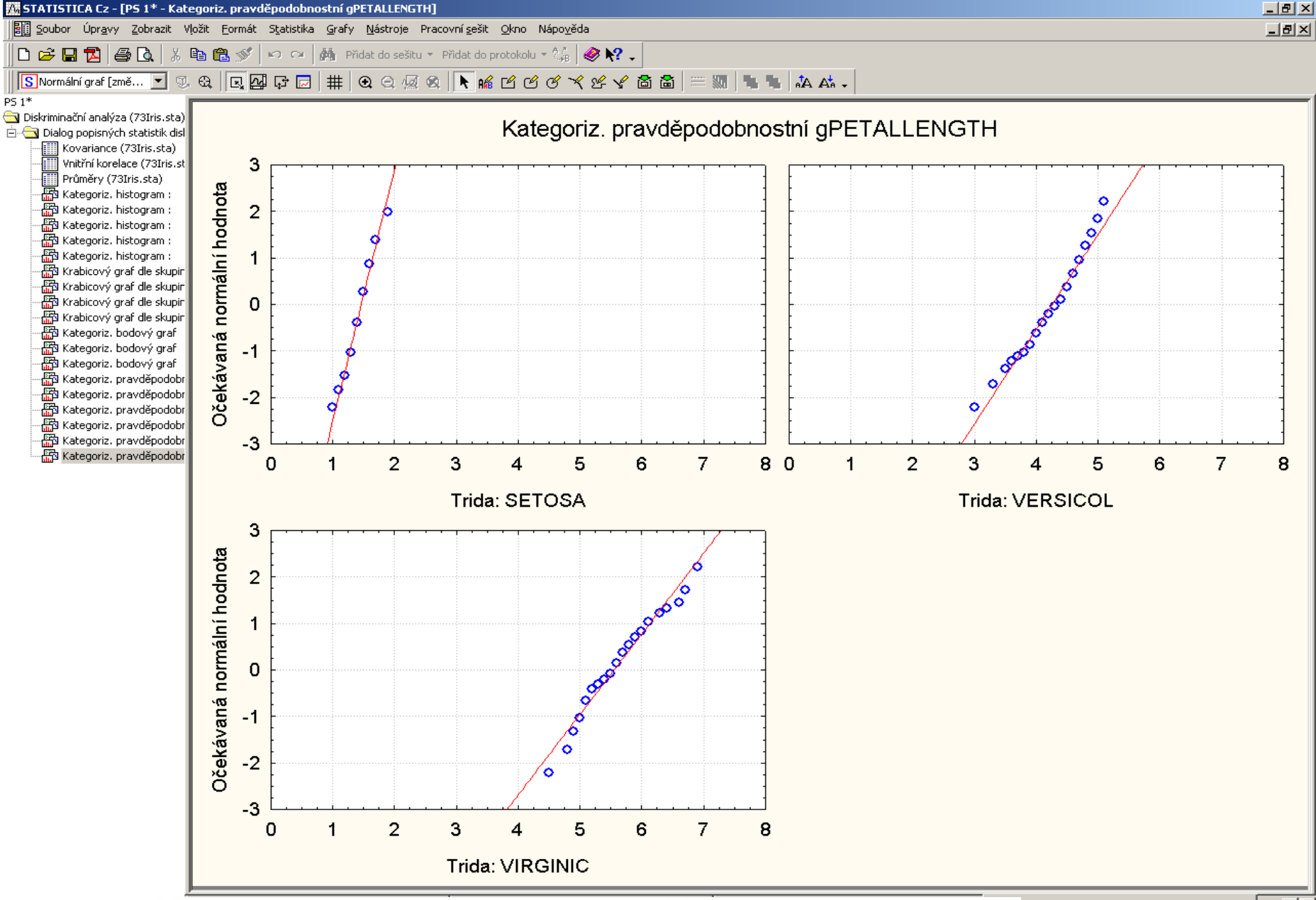
33. Zadání volby pro rankitový Q-Q graf testu normality *Kategorizovaný normální p-graf dle skupin* pro druhý znak SEPALWIDTH.



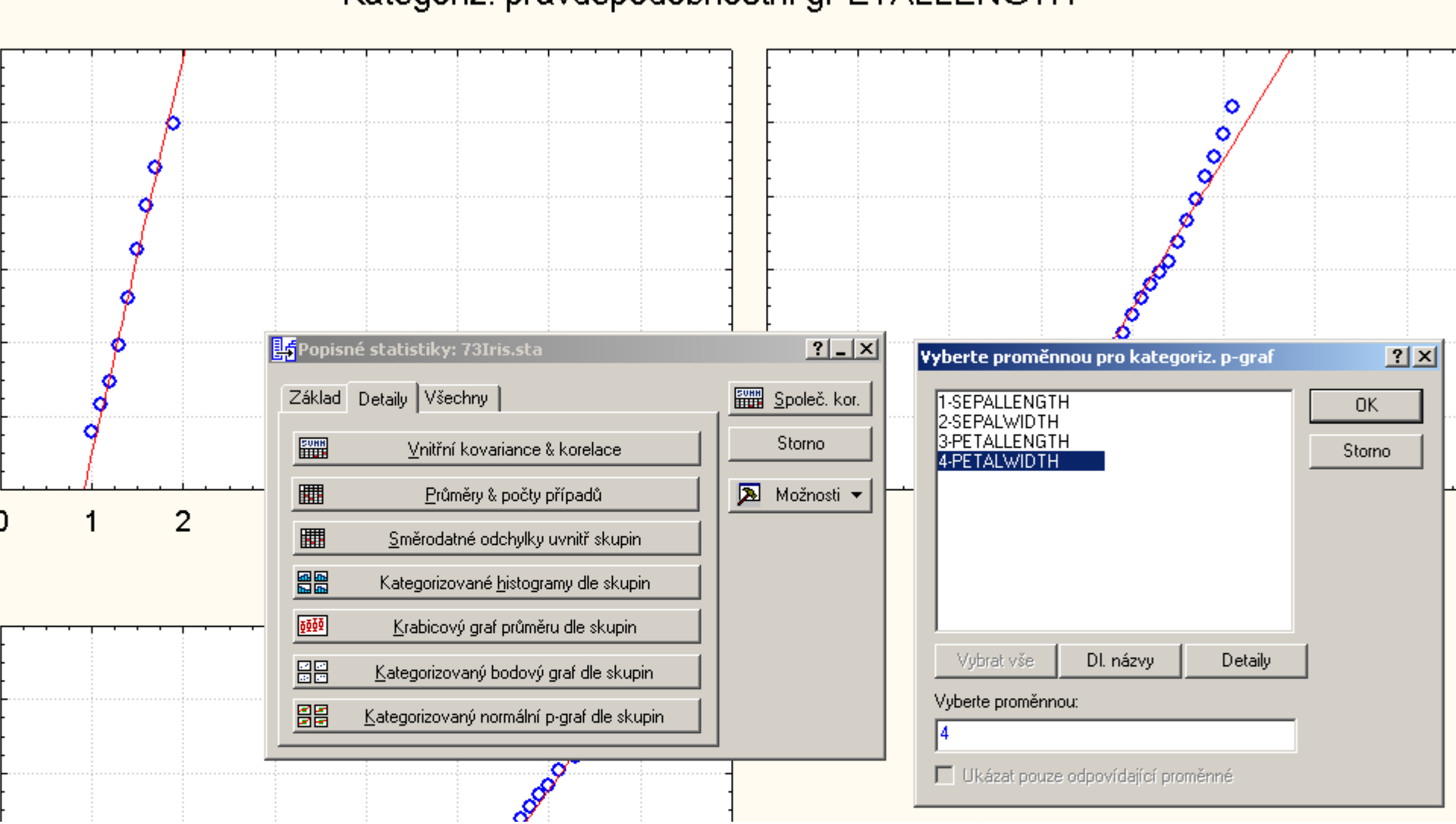
34. Rankitový Q-Q graf testu normality pro druhý znak SEPALWIDTH.



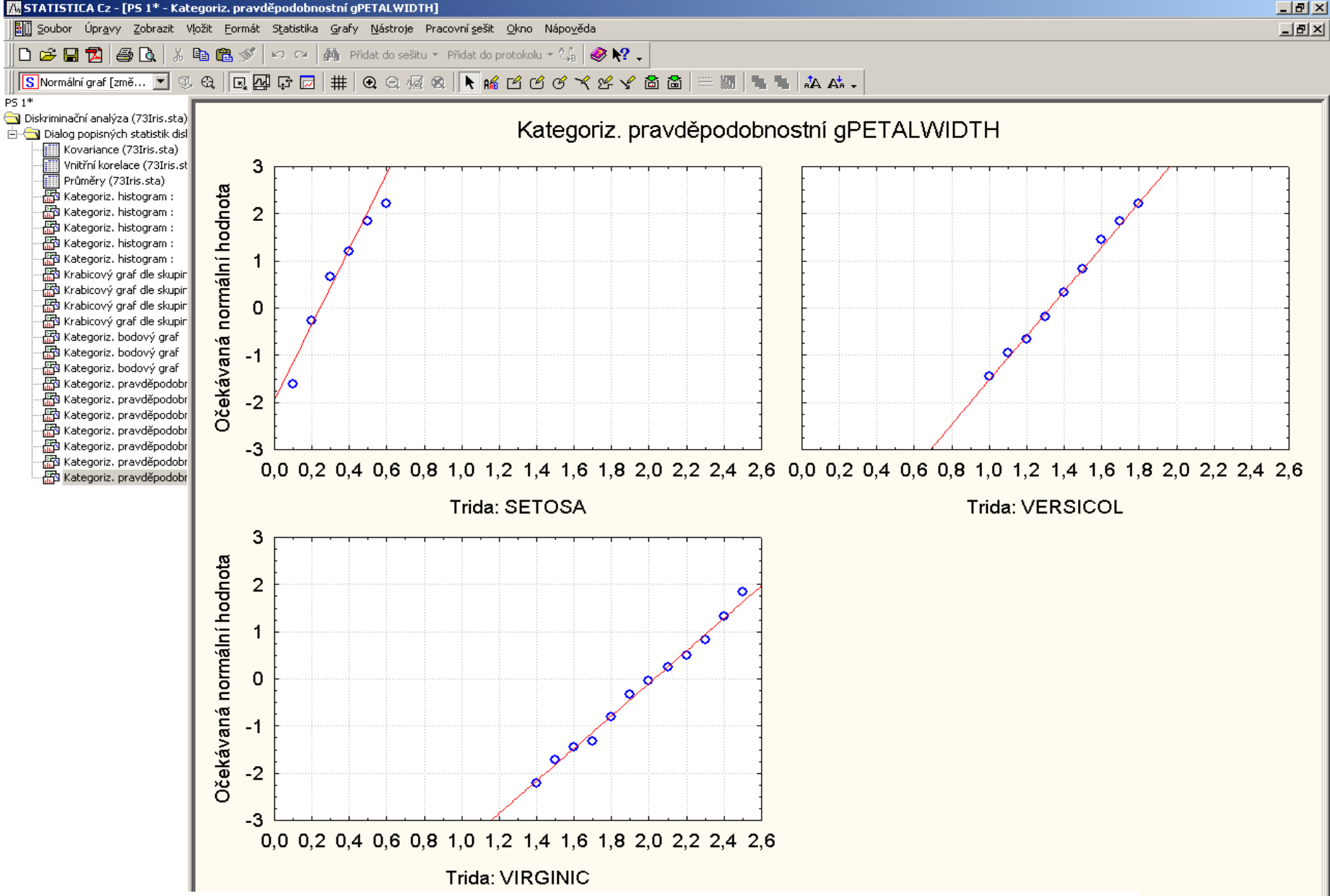
35. Zadání volby pro rankitový Q-Q graf testu normality *Kategorizovaný normální p-graf dle skupin* pro třetí znak PETALLENGTH.



36. Rankitový Q-Q graf testu normality pro třetí znak PETALLENGTH.

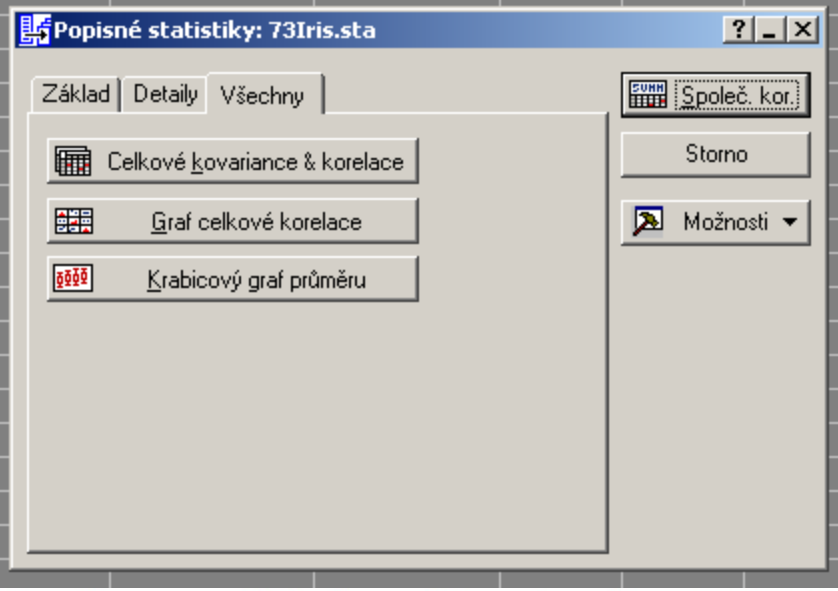


37. Zadání volby pro rankitový Q-Q graf testu normality *Kategorizovaný normální p-graf dle skupin* pro čtvrtý znak PETALWIDTH.



38. Rankitový Q-Q graf testu normality pro třetí znak PETALWIDTH.

SEPALWIDTH	0,53	1,00	0,38	0,47
PETALLENGTH	0,76	0,38	1,00	0,48
PETALWIDTH	0,36	0,47	0,48	1,00



39. V záložce *Všechny* u bloku *Popisní statistiky* zadáme *Celkové kovariance a korelace.*

STATISTICA Cz - [PS 1* - Korelace (celkové) (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

PS 1*

- Diskriminační analýza (73Iris.sta)
  - Dialog popisných statistik disl
    - Kovariance (73Iris.sta)
    - Vnitřní korelace (73Iris.st
    - Průměry (73Iris.sta)
    - Kategoriz. histogram :
    - Kategoriz. histogram :

Proměnná	Korelace (celkové) (73Iris.sta)			
	SEPALLENGTH	SEPALWIDTH	PETALLENGTH	PETALWIDTH
SEPALLENGTH	1,00	-0,12	0,87	0,82
SEPALWIDTH	-0,12	1,00	-0,43	-0,37
PETALLENGTH	0,87	-0,43	1,00	0,96
PETALWIDTH	0,82	-0,37	0,96	1,00

40. Korelační matice formou Pearsonových korelačních koeficientů ukáže na **celkové korelace** mezi 4 znaky, SEPALLENGTH, SEPALWIDTH, PETALLENGTH a PETALWIDTH.

- Kategoriz. bodový graf
- Kategoriz. bodový graf
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Korelace (73Iris.sta 5s*1
- Kovariance (73Iris.sta)
- Vnitřní korelace (73Iris.st
- Kovariance (celkové) (73
- Korelace (celkové) (73Iri

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

Iris Versicolor

Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica

STATISTICA Cz - [PS 1* - Vnitřní korelace (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

PS 1*

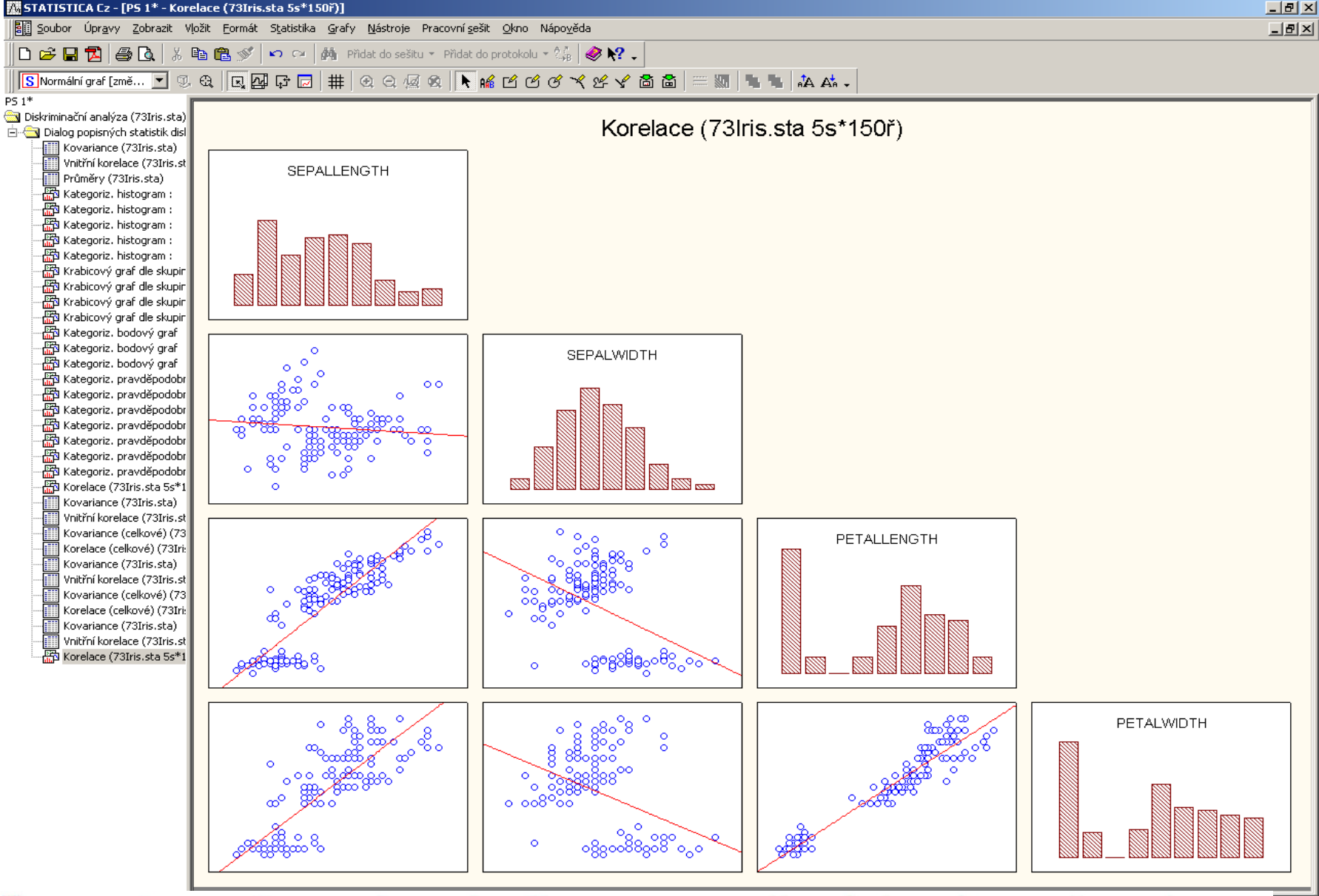
- Diskriminační analýza (73Iris.sta)
  - Dialog popisných statistik disl
    - Kovariance (73Iris.sta)
    - Vnitřní korelace (73Iris.st
    - Průměry (73Iris.sta)
    - Kategoriz. histogram :
    - Kategoriz. histogram :

Vnitřní korelace (73Iris.sta)

Proměnná	SEPALLENGTH	SEPALWIDTH	PETALLENGTH	PETALWIDTH
SEPALLENGTH	1,00	0,53	0,76	0,36
SEPALWIDTH	0,53	1,00	0,38	0,47
PETALLENGTH	0,76	0,38	1,00	0,48
PETALWIDTH	0,36	0,47	0,48	1,00

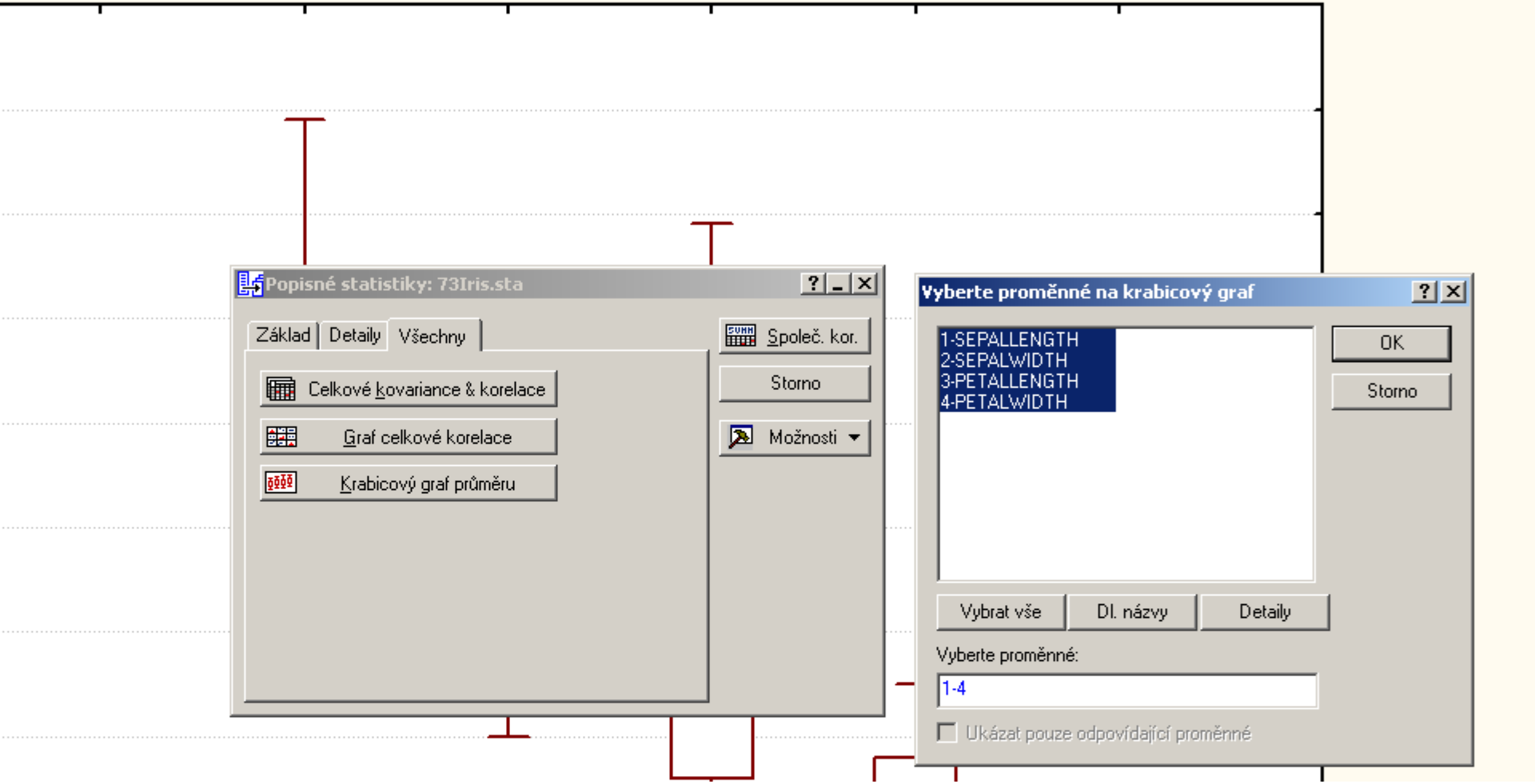
41. Korelační matice formou Pearsonových korelačních koeficientů ukáže na **vnitřní korelace** mezi 4 znaky, SEPALLENGTH, SEPALWIDTH, PETALLENGTH a PETALWIDTH..

- Kategoriz. bodový graf
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Kategoriz. pravděpodobr
- Korelace (73Iris.sta 5s*1
- Kovariance (73Iris.sta)
- Vnitřní korelace (73Iris.st
- Kovariance (celkové) (73

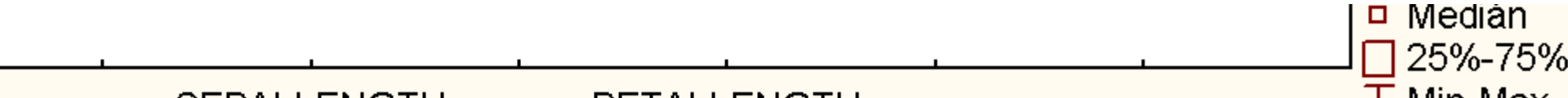


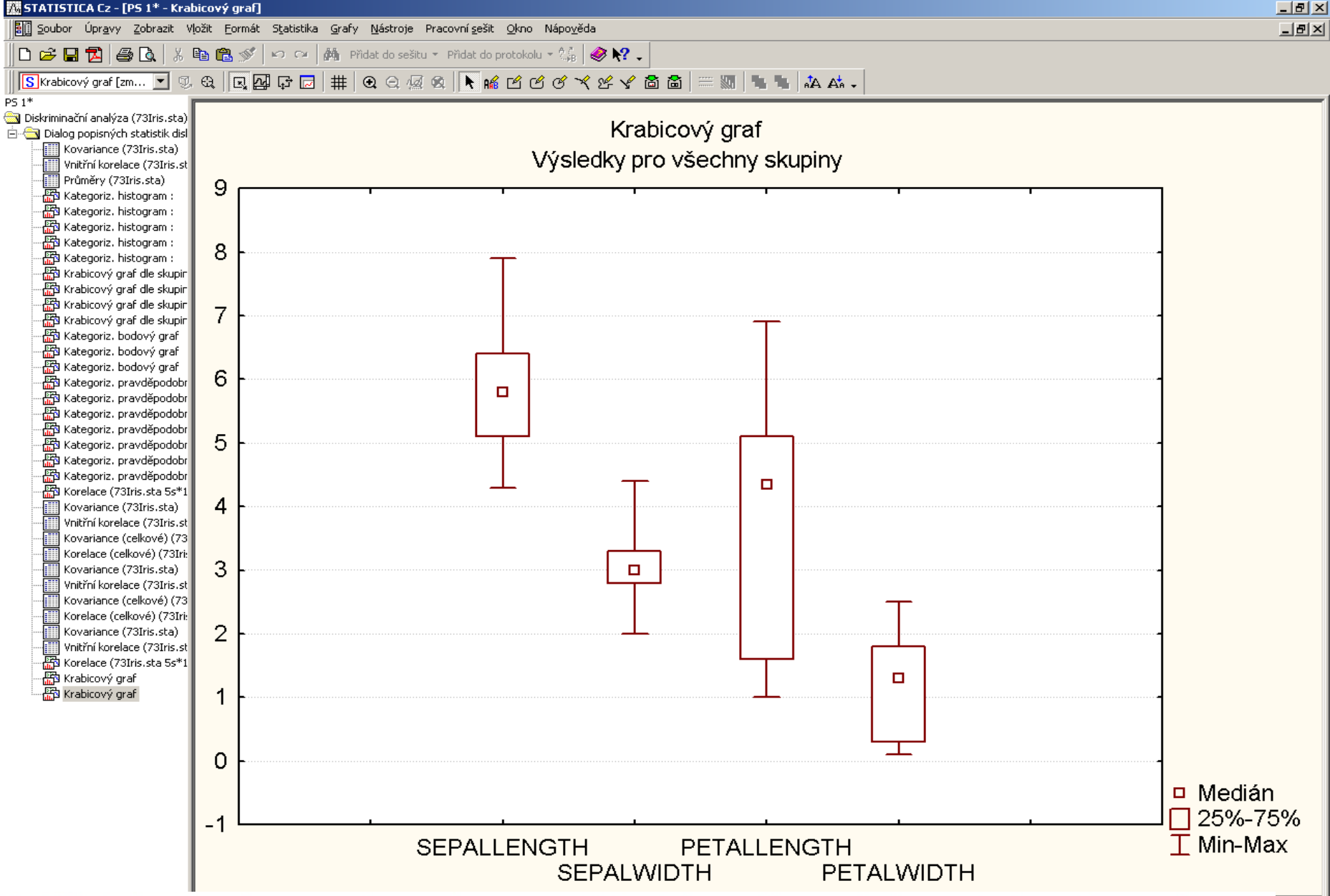
44. Červené přímky v oknech diagramu *Celkové korelace* ukazují na lineární závislost mezi dvěma znaky. Osamocené shluky bodů v každém grafu ukazují, že lze diskriminovat kosatce do několika tříd.

Výsledky pro všechny skupiny

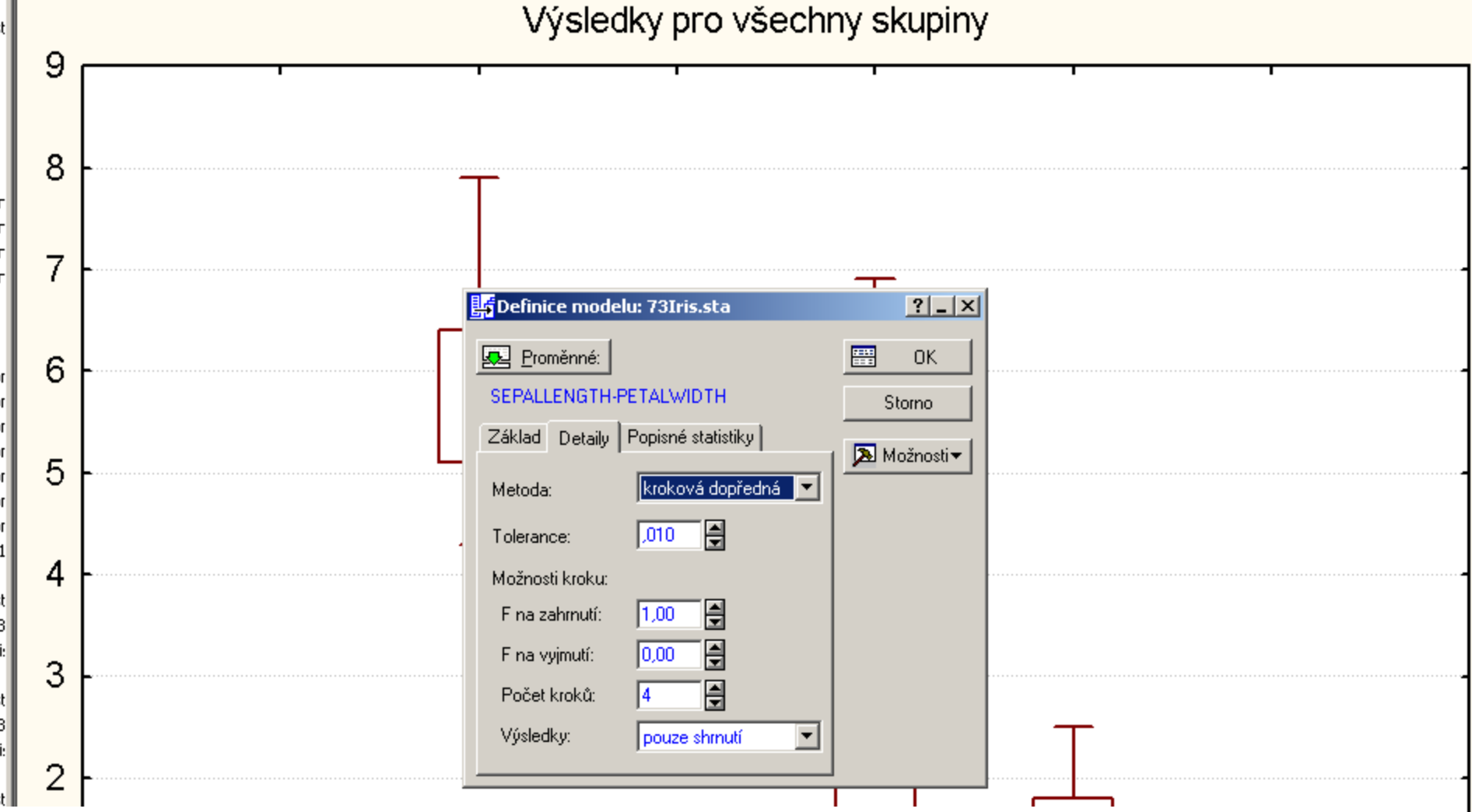


45. Zadání *Krabicového grafu průměrů pro všechny skupiny* pro vyšetřované 4 znaky.

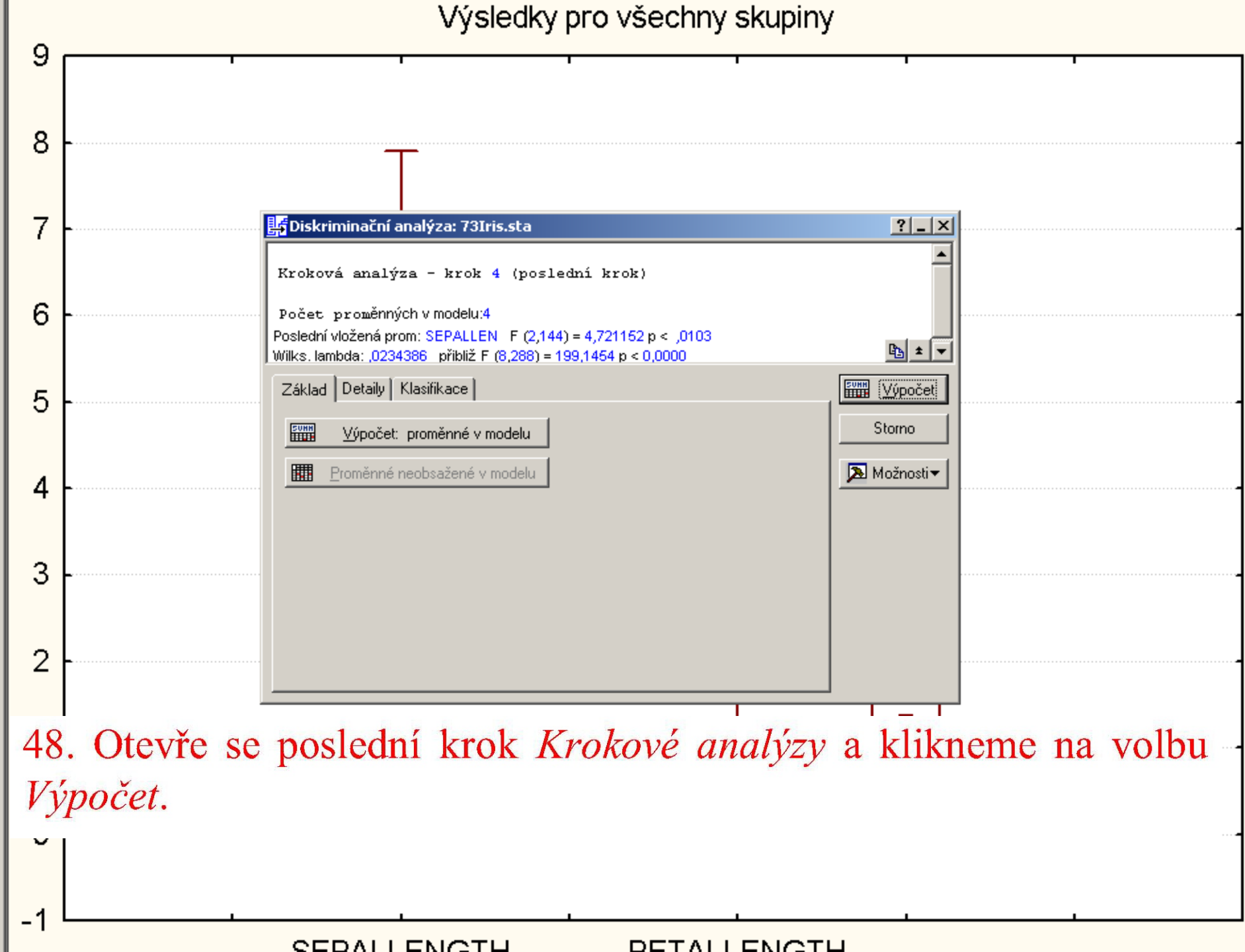




46. Tři druhy kosatců vedou na *Krabicový graf* 4 znaků. Ukazují na míry polohy, rozptýlení a tvaru všech kosatců.



47. V záložce *Detaily* u otevřené *Definice modelu* se zadá *Metoda diskriminace kroková dopředná* a dále se ponechá defaultní velikost *Tolerance* rovna 0.01. V okně *Výsledky* otevřeme roletu a zadáme *Pouze shrnutí* čímž si zvolíme poslední krok postupné krokové metody.



48. Otevře se poslední krok *Krokové analýzy* a klikneme na volbu *Výpočet*.

Výsledky diskriminační funkční analýzy (73Iris.sta)
krok 4, poč. prom. v modelu: 4; grupovací: Trida (3 skup)
Wilks. lambda: ,02344 přibliž F (8,288)=199,15 p<0,0000

N=150	Wilks. Lambda	Parc. Lambda	F na vyj (2,144)	Úroveň p	Toler.	1-toler. R^2
PETALLENGTH	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
SEPALWIDTH	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALWIDTH	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686
SEPALLENGTH	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007

49. Program seřadí znaky a vybere přitom nejprve takový znak, který nejlépe a nejvíce přispívá k diskriminaci objektů do tříd. Jde o znak s největší hodnotou *F*, větší hodnotou *F* než bylo *F* zadáno na začátku, při hledání nejlepších znaků se za počáteční hodnotu totiž volí obvykle nula. Když bychom ale naopak chtěli všechny znaky z modelu odebrat, zadáme na počátku pro *F* vysokou hodnotu, například 9999.

Tolerance $1\text{-}R^2$ je mírou redundance (= nadbytečnosti) znaku. Vykazuje-li totiž znak toleranci třeba 0.01, pak je z 99% nadbytečný vůči dříve zadaným ostatním znakům. Když jeden nebo více znaků jsou příliš redundantní, diskriminaci nelze provést. Položíme proto mezní hodnotu tolerance rovnu 0.01.

Wilkovo λ je standardní statistika testu významnosti diskriminační síly sledovaného znaku v modelu. Nabývá hodnot od 1, značící pak žádnou diskriminační sílu znaku, do 0.0, značící perfektní diskriminační silu znaku.

Výsledky diskriminační funkční analýzy (73Iris.sta)

Výsledky diskriminační funkční analýzy (73Iris.sta)
krok 4, poč. prom. v modelu: 4; grupovací: Trida (3 skup)
Wilks. lambda: ,02344 přibliž F (8,288)=199,15 p<0,0000

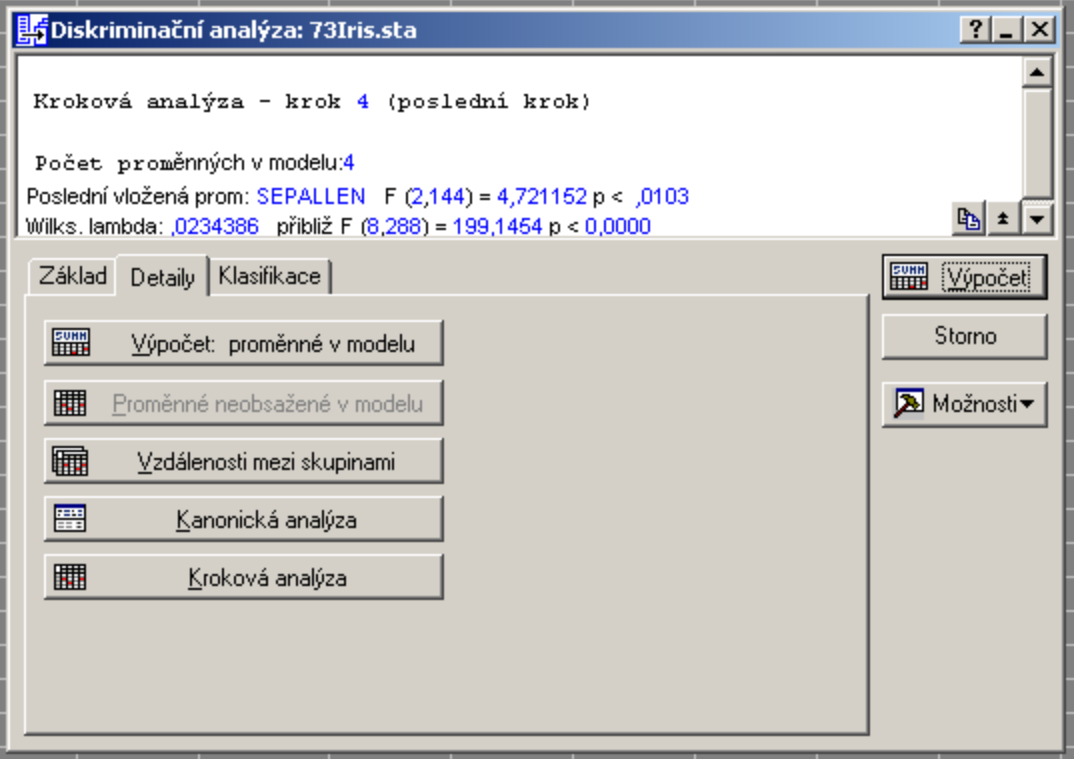
N=150	Wilks. Lambda	Parc. Lambda	F na vyj (2,144)	Úroveň p	Toler.	1-toler. R^2
PETALLENGTH	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
SEPALWIDTH	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALWIDTH	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686
SEPALLENGTH	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007

Parciální Wilkovo λ je jednotkový příspěvek dotyčného znaku k diskriminaci objektů do tříd. Čím je jeho hodnota bližší k 0, tím větší diskriminační síla tohoto znaku. Výpočet začíná s počáteční hodnotou rovnou 0.0. Wilkovo λ konvertuje do *F* kritéria a odpovídající vypočtené hladiny významnosti *p*.

Ukazuje se zde, že *PETALLENGTH* přispívá nejvíce, *PETALWIDTH* jako druhý, pak *SEPALWIDTH* a nejméně přispívá k diskriminaci kosatců do tříd *SEPALLENGTH*. Čím menší je totiž λ, tím více přispívá znak k diskriminaci. Z toho plyne, že PETAL je hlavním znakem, který vlastně vůbec umožňuje roztřídit kosatce do tříd.

Výsledky diskriminační funkční analýzy (73Iris.sta)

N=150	Wilks. Lambda	Parc. Lambda	F na vyj (2,144)	Úroveň p	Toler.	1-toler. R^2
PETALLENGTH	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
SEPALWIDTH	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALWIDTH	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686
SEPALLENGTH	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007



50. Protože více k diskriminaci kosatců pomůže *Kanonická analýza*, klikneme na ni.

N=150	Wilks. Lambda	Parc. Lambda	F na vyj (2,144)	Úroveň p	Toler.	1-toler. R^2
PETALLENGTH	0,035025	0,669206	35,59018	0,000000	0,365126	0,634874
SEPALWIDTH	0,030580	0,766480	21,93593	0,000000	0,608859	0,391141
PETALWIDTH	0,031546	0,743001	24,90433	0,000000	0,649314	0,350686
SEPALLENGTH	0,024976	0,938464	4,72115	0,010329	0,347993	0,652007

51. Program nyní vyčíslí rozličné nezávislé, orthogonální diskriminační funkce. Maximální počet těchto funkcí je buď omezen počtem znaků nebo počtem tříd minus jedna. Záleží, který z těchto dvou počtů představuje menší číslo. V naší úloze jde o 2 diskriminační funkce, totiž 3 třídy minus 1. Klikneme na *Test Chí-kvadrát postupných kořenů.*

N=150	Výsledky diskriminační funkční analýzy (73Iris) Počet prom. v modelu: 4; grupovací: Trida (3 skup) Wilk. lambda: ,02344 přibliž F (8,288)=199,15 p<0,00					
	Wilk. Lambda	Parc. Lambda	F na vyj (2,144)	p-hodn.	Toler.	
SEPALLENGTH	0,024976	0,938464	4,72115	0,010329	0,347993	0
SEPALWIDTH	0,030580	0,766480	21,93593	0,000000	0,608859	0
PETALLENGTH	0,035025	0,669206	35,59018	0,000000	0,365126	0
PETALWIDTH	0,031546	0,743001	24,90433	0,000000	0,649314	0

STATISTICA Cz - [PS 1* - Test chí-kvadrát po odstranění post. kořenů (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

PS 1*

- Diskriminační analýza (73Iris.sta)
  - Dialog popisných statistik disl
    - Kovariance (73Iris.sta)
    - Vnitřní korelace (73Iris.st
    - Průměry (73Iris.sta)
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Krabicový graf dle skupir

Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (73Iris.sta)					
	Vlastní číslo	Kan. R	Wilks. Lambda	Chi-kv.	sv	Úroveň p
0	32,19193	0,984821	0,023439	546,1153	8	0,000000
1	0,28539	0,471197	0,777973	36,5297	3	0,000000

Kořeny odstraněny	Test chí-kvadrát po odstranění post. ko			
	Vlastní číslo	Kan. R	Wilk. Lambda	Chi-kv.
0	32,19193	0,984821	0,023439	546,115
1	0,28539	0,471197	0,777973	36,529

52. Nejprve určíme, zda obě diskriminační funkce (či kanonické kořeny) jsou statisticky významné. První řádek tabulky *Testu Chí-kvadrát* testuje statistickou významnost všech (zde 2) kanonických kořenů při jejich postupném odstraňování. Druhý řádek se týká všech kořenů zbývajících ale bez prvního, atd. Tabulka proto odpoví, kolik a které kořeny jsou statisticky ještě významné. Zde jsou obě diskriminační funkce čili kanonické kořeny statisticky významné. Když proto budeme proměřovat u nových kosatců dvě míry okvětních lístků, PETALS a SEPALS, zaručíme, že je pak možné roztřídit kosatce do tříd.

STATISTICA Cz - [PS 1* - Standardiz. koeficienty (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Dat

Přidat do sešitu

Arial 10

PS 1*

- Diskriminační analýza (73Iris.sta)
  - Dialog popisných statistik disl
    - Kovariance (73Iris.sta)
    - Vnitřní korelace (73Iris.st
    - Průměry (73Iris.sta)
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Krabicový graf dle skupir
    - Krabicový graf dle skupir

Proměnná	Standardiz. koeficienty pro kanonické proměnn	
	Kořen1	Kořen2
PETALLENGTH	-0,94726	-0,401038
SEPALWIDTH	0,52124	0,735261
PETALWIDTH	-0,57516	0,581040
SEPALLENGTH	0,42695	0,012408
Vlastní	32,19193	0,285391
KumPodíl	0,99121	1,000000

Proměnná	Standardiz. koeficie pro kanonické prom	
	Kořen1	K
SEPALLENGTH	0,42695	
SEPALWIDTH	0,52124	
PETALLENGTH	-0,94726	
PETALWIDTH	-0,57516	
Vlastní	32,19193	
KumPodíl	0,99121	

53. *Tabulka standardizovaných koeficientů:* Původní koeficienty ve spojení s původními daty poskytují skóre diskriminační funkce. Standardizované koeficienty patří ke standardizovaným znakům a jsou v normovaných vzájemně porovnatelných škálách. První diskriminační funkce je silně vážená délkami a šířkami PETALS. Zbývající dva znaky ale také trochu přispívají. Druhá diskriminační funkce je poznamenána více znakem SEPALWIDTH a daleko však méně PETALWIDTH a PETALLENGTH.

STATISTICA Cz - [PS 1* - Faktorová strukturní matice (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

PS 1*

- Diskriminační analýza (73Iris.sta)
  - Dialog popisných statistik disl
    - Kovariance (73Iris.sta)
    - Vnitřní korelace (73Iris.st
    - Průměry (73Iris.sta)
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Kategoriz. histogram :
    - Krabicový graf dle skupir
    - Krabicový graf dle skupir
    - Krabicový graf dle skupir
    - Krabicový graf dle skupir
    - Kategoriz. bodový graf
    - Kategoriz. bodový graf
    - Kategoriz. bodový graf
    - Korelace (celkové) (73Iri

Proměnná	Faktorová strukturní m Korelační proměnné - I (vnitřní korelace)	
	Kořen1	Kořen2
PETALLENGTH	-0,706065	0,167701
SEPALWIDTH	0,119012	0,863681
PETALWIDTH	-0,633178	0,737242
SEPALLENGTH	-0,222596	0,310812

Proměnná	Faktorová strukturní matice Korelační proměnné - Kano (vnitřní korelace)	
	Kořen1	K
SEPALLENGTH	-0,222596	
SEPALWIDTH	0,119012	
PETALLENGTH	-0,706065	
PETALWIDTH	-0,633178	

54. *Faktorová strukturní matice:* v tabulce uvedené koeficienty představují korelace mezi znaky a diskriminačními funkcemi a užívají se proto k objasnění "významu" diskriminačních funkcí. Bývá často výhodné pojmenovat či názvem přidělit význam těmto diskriminačním funkcím podobně jako se to dělá v FA.

STATISTICA Cz - [PS 1* - Průměry kan. proměnných (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu

Arial 10

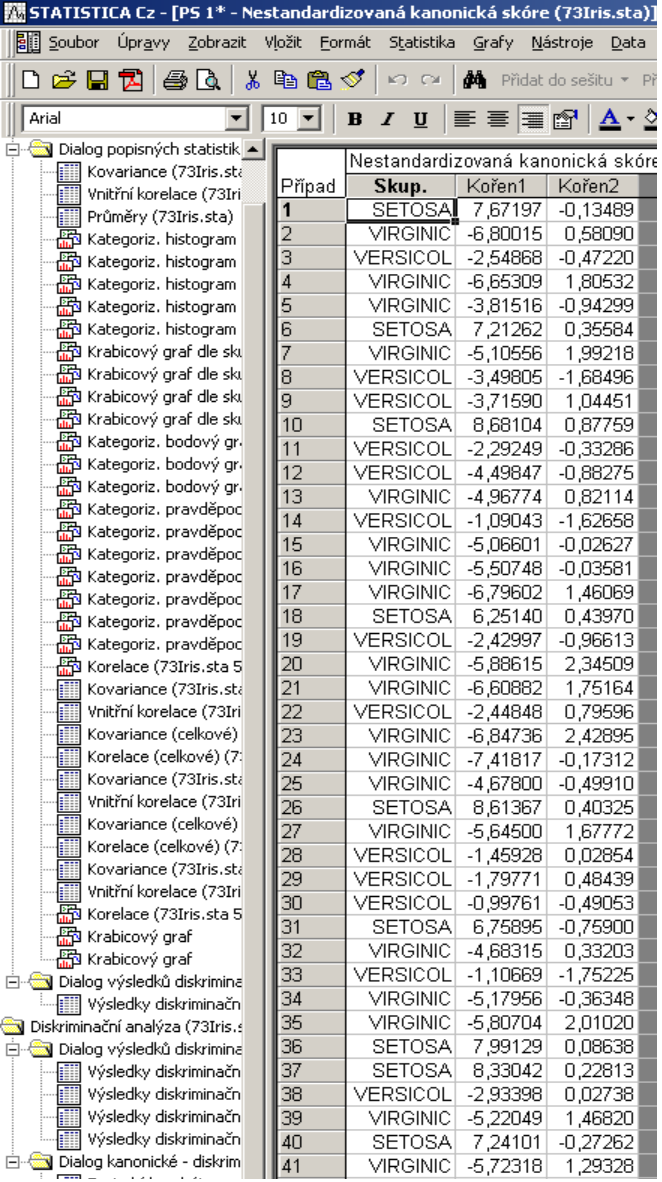
Diskriminační analýza (73Iris.s
Dialog popisných statistik
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Průměry (73Iris.sta)
Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram

Skup.	Průměry kan. proměnných (7	
	Kořen1	Kořen2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

Skup.	Průměry kan. proměnnýc	
	Kořen1	Ko
SETOSA	7,60760	
VERSICOL	-1,82505	-
VIRGINIC	-5,78255	

55. *Průměry kanonických proměnných:* je třeba určit povahu diskriminace kosatců v každém kanonickém kořeni. Zdá se, že první diskriminační funkce diskriminuje většinou mezi třídou Setosa a ostatními kosatci. Kanonický průměr Setosy se zcela liší od ostatních dvou tříd. Druhá diskriminační funkce rozlišuje pak mezi Versicolorem a ostatními kosatci, avšak tato diskriminace je mnohem menší než u první funkce.

Kategoriz. pravděpoc
Kategoriz. pravděpoc
Korelace (73Iris.sta 5
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Kovariance (celkové)
Korelace (celkové) (7:
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Kovariance (celkové)



Případ	Nestandardizovaná kanonická skóre (73Iris)		
	Skup.	Kořen1	Kořen2
1	SETOSA	7,67197	-0,13489
2	VIRGINIC	-6,80015	0,58090
3	VERSICOL	-2,54868	-0,47220
4	VIRGINIC	-6,65309	1,80532
5	VIRGINIC	-3,81516	-0,94299
6	SETOSA	7,21262	0,35584
7	VIRGINIC	-5,10556	1,99218
8	VERSICOL	-3,49805	-1,68496
9	VERSICOL	-3,71590	1,04451
10	SETOSA	8,68104	0,87759
11	VERSICOL	-2,29249	-0,33286
12	VERSICOL	-4,49847	-0,88275
13	VIRGINIC	-4,96774	0,82114
14	VERSICOL	-1,09043	-1,62658
15	VIRGINIC	-5,06601	-0,02627
16	VIRGINIC	-5,50748	-0,03581
17	VIRGINIC	-6,79602	1,46069
18	SETOSA	6,25140	0,43970
19	VERSICOL	-2,42997	-0,96613
20	VIRGINIC	-5,88615	2,34509
21	VIRGINIC	-6,60882	1,75164
22	VERSICOL	-2,44848	0,79596
23	VIRGINIC	-6,84736	2,42895
24	VIRGINIC	-7,41817	-0,17312
25	VIRGINIC	-4,67800	-0,49910
26	SETOSA	8,61367	0,40325
27	VIRGINIC	-5,64500	1,67772
28	VERSICOL	-1,45928	0,02854
29	VERSICOL	-1,79771	0,48439
30	VERSICOL	-0,99761	-0,49053
31	SETOSA	6,75895	-0,75900
32	VIRGINIC	-4,68315	0,33203
33	VERSICOL	-1,10669	-1,75225
34	VIRGINIC	-5,17956	-0,36348
35	VIRGINIC	-5,80704	2,01020
36	SETOSA	7,99129	0,08638
37	SETOSA	8,33042	0,22813
38	VERSICOL	-2,93398	0,02738
39	VIRGINIC	-5,22049	1,46820
40	SETOSA	7,24101	-0,27262
41	VIRGINIC	-5,72318	1,29328
42	SETOSA	6,41444	1,24730
43	VERSICOL	-1,27227	-1,21458
44	SETOSA	6,93412	-0,70552
45	VIRGINIC	-7,58120	-0,98072
46	VIRGINIC	-4,37150	-0,12130
47	SETOSA	6,82946	-0,54496
48	VERSICOL	-2,40157	-1,59458
49	VIRGINIC	-5,27916	-0,04246
50	SETOSA	6,76470	-0,50515
51	SETOSA	8,08190	0,76339
52	VIRGINIC	-4,07704	0,52324

Případ	Nestandardizovaná kanonická skóre (73Iris)		
	Skup.	Kořen1	Kořen2
53	SETOSA	6,55893	-0,38922
54	SETOSA	6,77138	-0,97063
55	SETOSA	7,95876	-0,16496
56	VIRGINIC	-5,10749	-2,13059
57	VIRGINIC	-6,51910	0,29698
58	SETOSA	7,57247	-0,80546
59	SETOSA	9,84984	1,58594
60	SETOSA	6,85944	1,05165
61	VERSICOL	-1,20117	0,08444
62	VIRGINIC	-4,08087	0,18594
63	SETOSA	7,12869	-0,78666
64	SETOSA	8,06180	0,30042
65	VERSICOL	-2,76810	0,03220
66	VERSICOL	-0,77685	-1,65916
67	SETOSA	6,82308	0,46301
68	SETOSA	7,18677	-0,36099
69	VERSICOL	-2,58980	-0,17461
70	VERSICOL	0,30749	-1,31887
71	SETOSA	9,15824	2,73760
72	SETOSA	8,13231	0,51446
73	VIRGINIC	-6,79672	0,86309
74	VIRGINIC	-6,52450	2,44504
75	VIRGINIC	-5,50748	-0,03581
76	VERSICOL	-1,61616	-0,47010
77	VIRGINIC	-4,58372	-0,85682
78	SETOSA	7,37495	0,56584
79	SETOSA	9,12635	1,22443
80	VIRGINIC	-6,29201	0,46718
81	VIRGINIC	-4,99550	0,18777
82	VIRGINIC	-3,93985	0,61402
83	VIRGINIC	-5,60546	-0,34074
84	VERSICOL	-0,22267	-1,58467
85	VERSICOL	-2,00599	-0,90542
86	VERSICOL	-1,18166	-0,53757
87	SETOSA	9,46768	1,82523
88	SETOSA	7,06201	-0,66340
89	VIRGINIC	-9,17147	-0,74826
90	VIRGINIC	-4,76454	-2,15574
91	SETOSA	7,70195	1,46172
92	VERSICOL	-1,75020	-0,82118
93	VERSICOL	-1,95842	-0,35156
94	VERSICOL	-2,10361	1,19157
95	SETOSA	7,60529	-0,01163
96	SETOSA	6,56055	-1,01516
97	VERSICOL	-1,19376	-2,63446
98	VERSICOL	-0,60552	-1,94298
99	VERSICOL	-0,89870	-0,90494
100	SETOSA	7,48983	-0,26538
101	SETOSA	6,81320	-0,67063
102	VIRGINIC	-6,27284	1,64948
103	VERSICOL	-1,42159	-0,55124
104	VIRGINIC	-6,22824	-0,71272

Případ	Nestandardizovaná kanonická skóre (73Iris)		
	Skup.	Kořen1	Kořen2
105	VERSICOL	-1,85893	0,31901
106	SETOSA	7,68823	-0,00922
107	SETOSA	7,91794	0,67512
108	VIRGINIC	-5,36071	0,64612
109	VERSICOL	-1,15809	-2,64341
110	VIRGINIC	-6,31693	-0,96898
111	VIRGINIC	-6,32774	-1,38329
112	SETOSA	7,34306	-0,94732
113	VERSICOL	-2,14258	0,08878
114	VERSICOL	-2,47946	-1,94074
115	SETOSA	8,39739	0,64736
116	VERSICOL	-1,32553	-0,16287
117	VERSICOL	-3,83526	-1,40596
118	VERSICOL	-2,25741	-1,42679
119	VERSICOL	-1,25571	-0,54642
120	VERSICOL	0,47597	-0,79991
121	VERSICOL	-1,54948	-0,59336
122	VIRGINIC	-6,85056	0,82983
123	VIRGINIC	-5,20383	1,14477
124	SETOSA	8,58243	1,83449
125	SETOSA	7,78075	0,58434
126	VIRGINIC	-6,85281	2,71759
127	VIRGINIC	-4,44073	1,34724
128	VERSICOL	-2,66606	-0,64250
129	VERSICOL	-0,37837	0,08664
130	VERSICOL	-2,41695	-0,09278
131	VIRGINIC	-5,45010	-0,20774
132	VIRGINIC	-5,66034	0,83271
133	VERSICOL	-1,95558	-1,15435
134	SETOSA	7,21930	-0,10965
135	SETOSA	7,32680	-1,07299
136	SETOSA	5,66188	-1,93436
137	VIRGINIC	-5,95824	-0,09402
138	SETOSA	8,07836	0,96858
139	SETOSA	8,02097	1,14050
140	VERSICOL	-2,26247	-1,58725
141	VERSICOL	-1,43756	-0,13442
142	VERSICOL	-2,45906	-0,93528
143	SETOSA	7,49680	-0,18838
144	SETOSA	7,58648	1,20797
145	SETOSA	7,92462	0,20964
146	VIRGINIC	-6,75926	1,60023
147	VERSICOL	-3,51848	0,16059
148	VIRGINIC	-7,83947	2,13973
149	SETOSA	8,31445	0,64495
150	VERSICOL	-0,29318	-1,79872

56. *Nestandardizované kanonické skóre:* Tabulka nestandardizovaných kanonických kořenů ukazuje na diskriminaci všech 150 kosatců. Samozřejmě by bylo zde názornější provést grafické zobrazení.

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

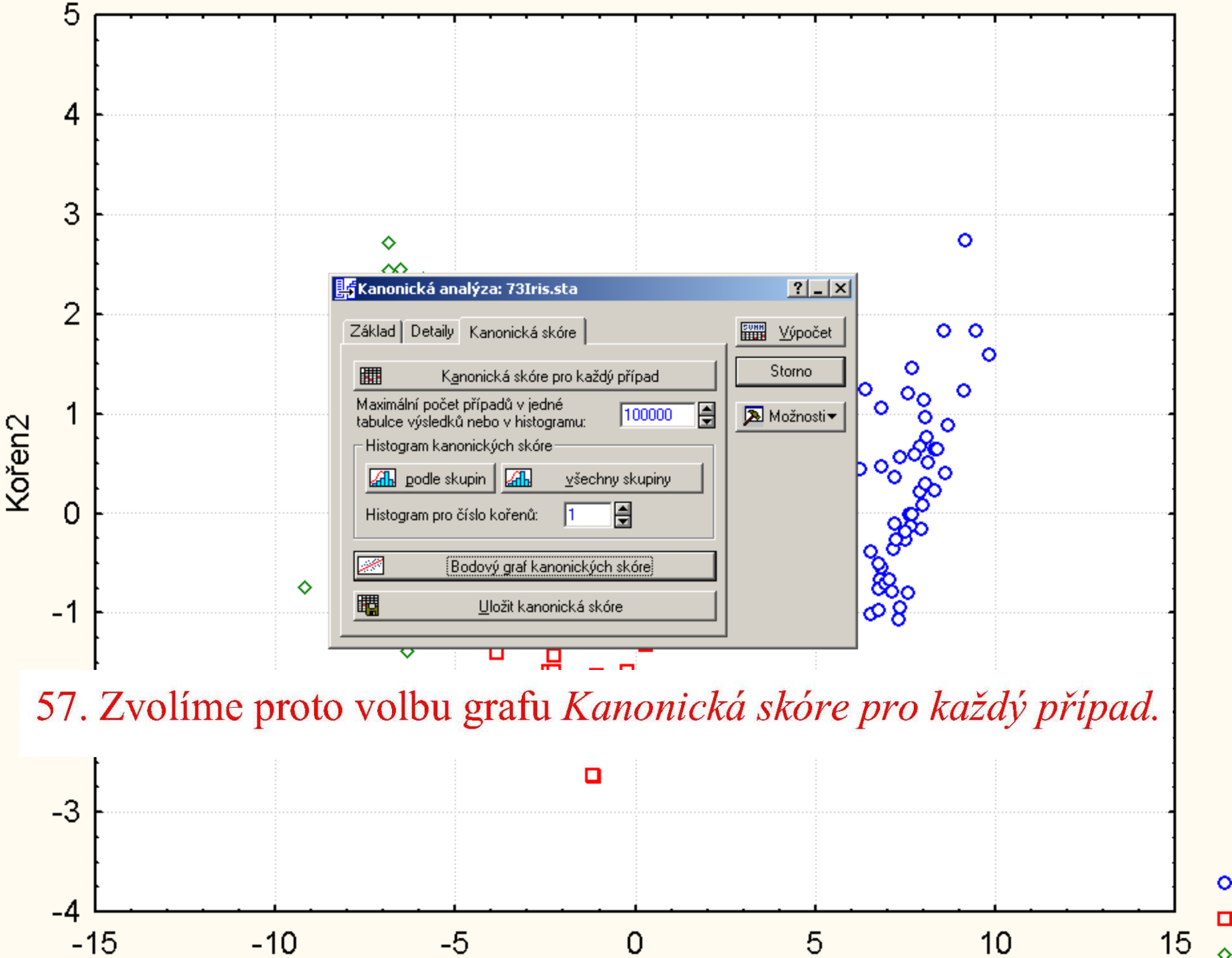
Iris Versicolor

Iris Versicolor

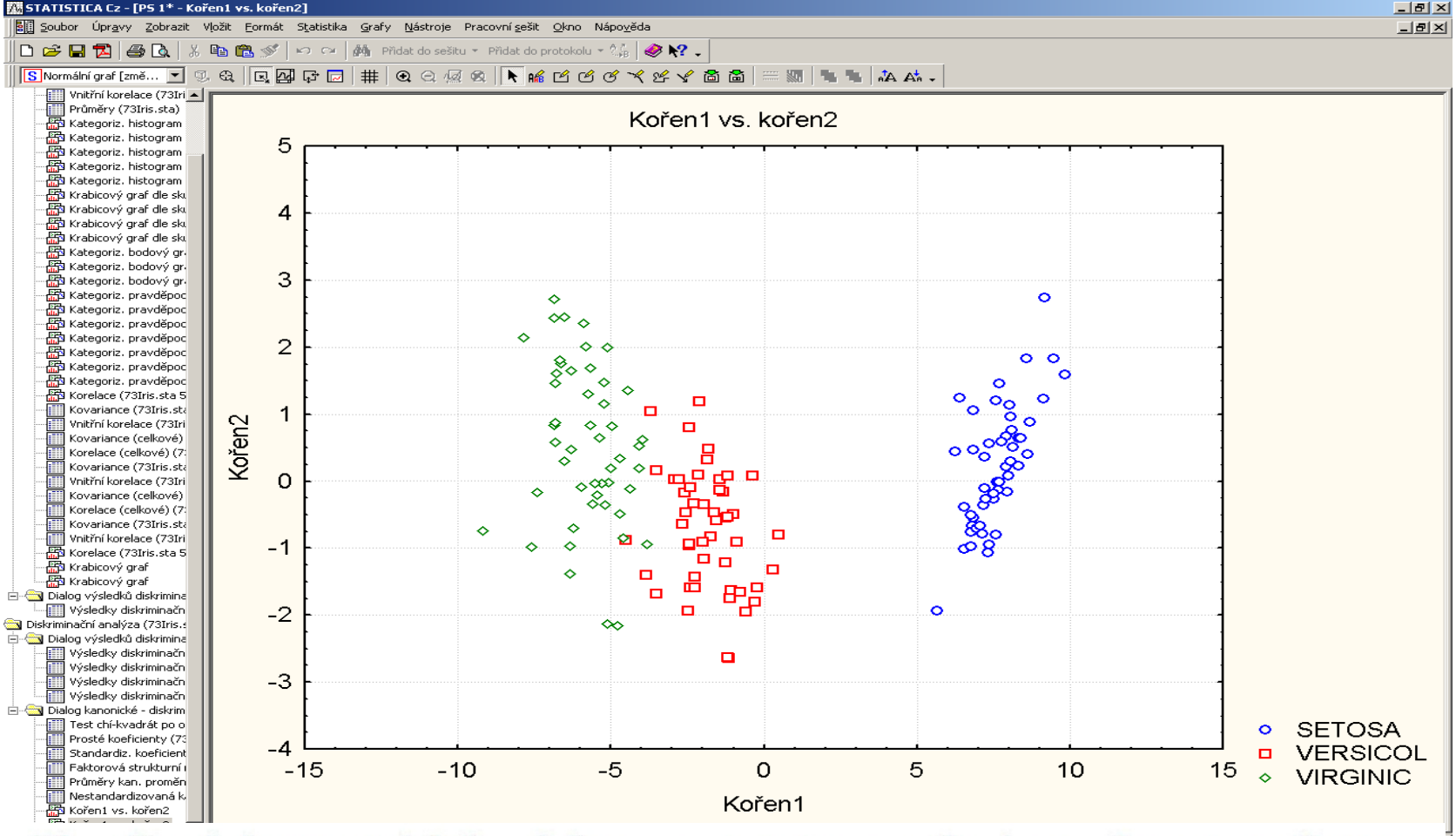
Iris Virginica

Iris Virginica

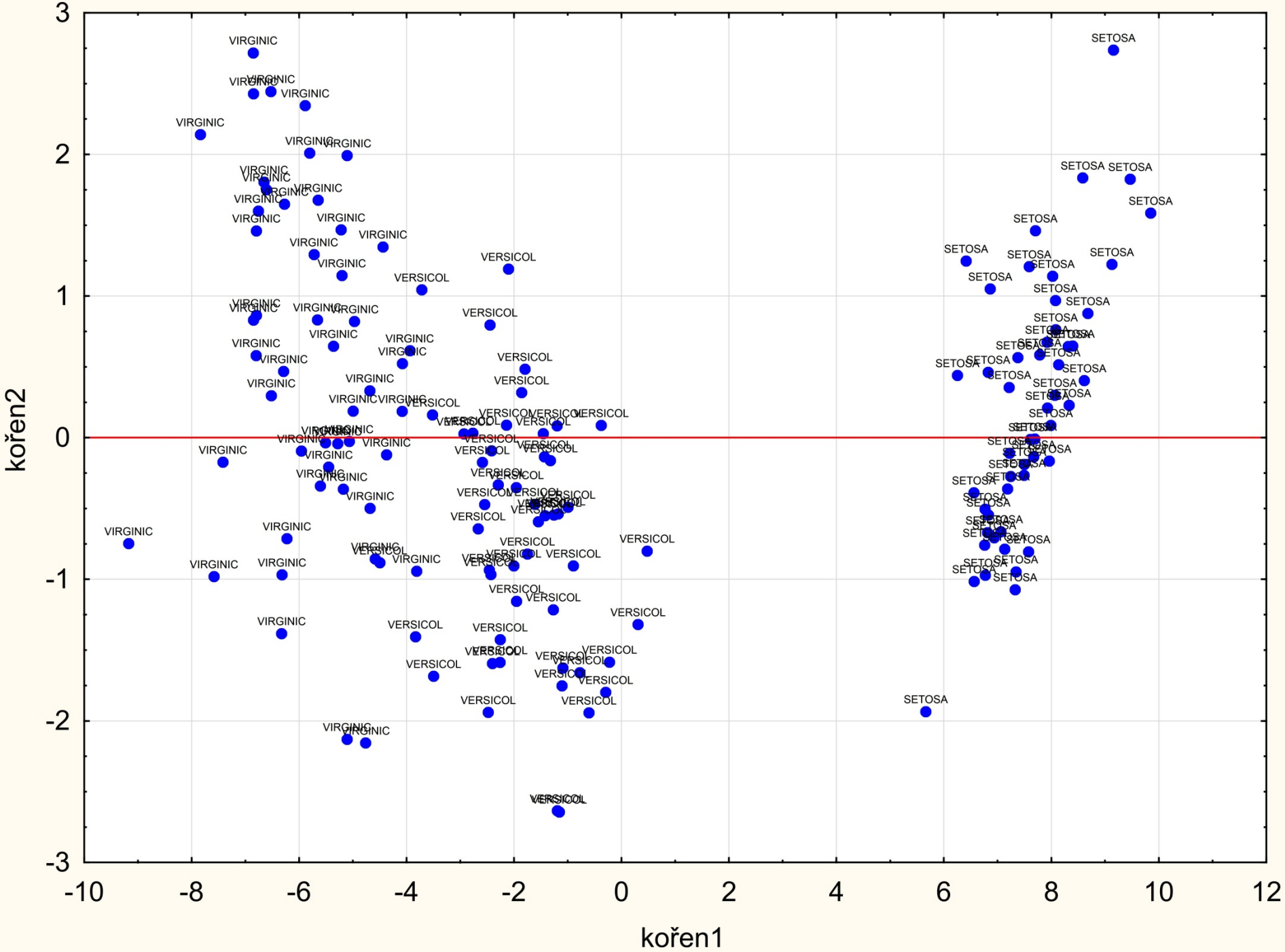
Iris Virginica

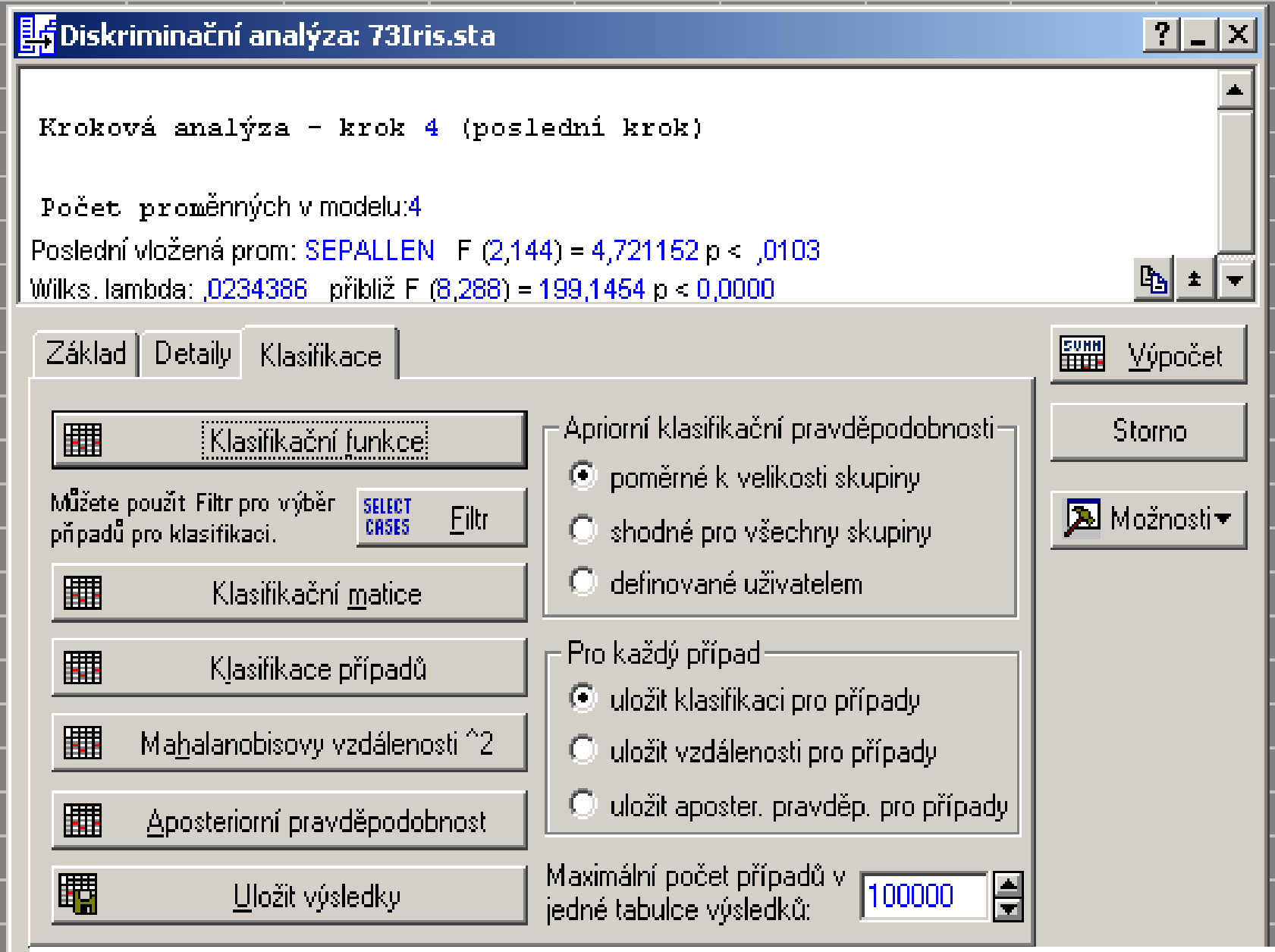


57. Zvolíme proto volbu grafu *Kanonická skóre pro každý případ.*



58. *Graf kanonických skóre:* v tomto grafu jsou kosatce Setosa zobrazeny modře daleko vpravo od ostatních. Kořen 1 (čili první diskriminační funkce) silně diskriminuje Setosu od ostatních kosatců. Kořen 2 (druhá diskriminační funkce) poskytuje diskriminaci mezi Versicolor se zápornými hodnotami Kořene 2 a ostatními kosatci, převážně kladného Kořene 2. Diskriminace však není tak zřetelná jako je tomu u prvního Kořene 1.


kořen2
kořen1
3
2
1
0
-1
-2
-3
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
VIRGINIC
VERSICOL
SETOSA



59. Následuje zadání záložky *Klasifikace*, ve které zvolíme *Klasifikační funkce*.

STATISTICA Cz - [PS 1* - Klasifikační funkce; grupovací : Trida (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram
Krabicový graf dle sku
Krabicový graf dle sku
Krabicový graf dle sku
Krabicový graf dle sku
Kategoriz. bodový gr.

Proměnná	Klasifikační funkce; grupovací : Trida (73Iris.sta)		
	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
PETALLENGTH	-16,4306	5,2115	12,767
SEPALWIDTH	23,5879	7,0725	3,685
PETALWIDTH	-17,3984	6,4342	21,079
SEPALLENGTH	23,5442	15,6982	12,446
konst.	-86,3085	-72,8526	-104,368

60. *Tabulka klasifikační funkce:* klasifikační funkce **se nesmí zaměnit** za diskriminační funkci. Klasifikační funkce se počítají pro každou třídu a mohou být použity přímo ke klasifikaci tříd. Každý květ kosatce (čili každý objekt) by měl být klasifikován do té třídy, ve které dosahuje své neivvšší klasifikační skóre.

Proměnná	Klasifikační funkce; grupovací : Trida (73Iris)		
	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SEPALLENGTH	23,5442	15,6982	12,446
SEPALWIDTH	23,5879	7,0725	3,685
PETALLENGTH	-16,4306	5,2115	12,767
PETALWIDTH	-17,3984	6,4342	21,079
Konstant	-86,3085	-72,8526	-104,368

STATISTICA Cz - [PS 1* - Klasifikační matice (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10

Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram
Krabicový graf dle sku
Krabicový graf dle sku
Krabicový graf dle sku
Krabicový graf dle sku
Kategoriz. bodový gr
Kategoriz. bodový gr

Klasifikační matice (73Iris.sta)
Řádky: pozorované klasifikace
Sloupce: předpovězené klasifikace

Skup.	% správných	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SETOSA	100,0000	50	0	0
VERSICOL	96,0000	0	48	2
VIRGINIC	98,0000	0	1	49
Celkem	98,0000	50	49	51

61. *Klasifikační matice:* v nadpisech sloupců je uvedena dotyčná *a priori* klasifikační pravděpodobnost. Z tabulky lze snadno odečíst chybnou klasifikaci do jednotlivých tříd.

Kategoriz. pravděpoc
Kategoriz. pravděpoc
Korelace (73Iris.sta 5
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Kovariance (celkové)
Korelace (celkové) (7
Kovariance (73Iris.sta
Vnitřní korelace (73Iri

Klasifikační matice (73Iris)
Řádky: pozorované klasifikace
Sloupce: předpovězené klasifikace

Skup.	% správnýc	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SETOSA	100,0000	50	0	0
VERSICOL	96,0000	0	48	2
VIRGINIC	98,0000	0	1	49
Celkem	98,0000	50	49	51

STATISTICA Cz - [PS 1* - Klasifikace případů (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu Proměnné Případy

Klasifikace případů (73Iris.sta)
Nesprávná klasifikace je označena *

Případ	Pozorov. Klasif.	1 p=,33333	2 p=,33333	3 p=,33333
1	SETOSA	SETOSA	VERSICOL	VIRGINIC
2	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
3	VERSICOL	VERSICOL	VIRGINIC	SETOSA
4	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
* 5	VIRGINIC	VERSICOL	VIRGINIC	SETOSA
6	SETOSA	SETOSA	VERSICOL	VIRGINIC
7	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
8	VERSICOL	VERSICOL	VIRGINIC	SETOSA
* 9	VERSICOL	VIRGINIC	VERSICOL	SETOSA
10	SETOSA	SETOSA	VERSICOL	VIRGINIC
11	VERSICOL	VERSICOL	VIRGINIC	SETOSA
* 12	VERSICOL	VIRGINIC	VERSICOL	SETOSA
13	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
14	VERSICOL	VERSICOL	VIRGINIC	SETOSA
15	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
16	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
17	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
18	SETOSA	SETOSA	VERSICOL	VIRGINIC
19	VERSICOL	VERSICOL	VIRGINIC	SETOSA
20	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
21	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
22	VERSICOL	VERSICOL	VIRGINIC	SETOSA
23	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
24	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
25	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
26	SETOSA	SETOSA	VERSICOL	VIRGINIC
27	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
28	VERSICOL	VERSICOL	VIRGINIC	SETOSA
29	VERSICOL	VERSICOL	VIRGINIC	SETOSA
30	VERSICOL	VERSICOL	VIRGINIC	SETOSA
31	SETOSA	SETOSA	VERSICOL	VIRGINIC
32	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
33	VERSICOL	VERSICOL	VIRGINIC	SETOSA
34	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
35	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
36	SETOSA	SETOSA	VERSICOL	VIRGINIC
37	SETOSA	SETOSA	VERSICOL	VIRGINIC
38	VERSICOL	VERSICOL	VIRGINIC	SETOSA
39	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
40	SETOSA	SETOSA	VERSICOL	VIRGINIC
41	VIRGINIC	VIRGINIC	VERSICOL	SETOSA
42	SETOSA	SETOSA	VERSICOL	VIRGINIC

62. *Klasifikace objektů* (zde kosatců): první sloupec přináší skutečnou třídu. Druhý sloupec pak první možnost klasifikace, další sloupce pak další možnosti klasifikace. Hvězdičkou * je zde označena chybná klasifikace.

STATISTICA Cz - [PS 1* - Mahalanobisovy vzdálenosti (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Mahalanobisovy vzdálenosti (73Iris.sta)
Nesprávná klasifikace je označena *

Případ	Pozorov. Klasif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
1	SETOSA	0,2419	90,6602	181,5587
2	VIRGINIC	208,5713	27,3188	1,8944
3	VERSICOL	105,2663	2,2329	13,0720
4	VIRGINIC	207,9180	31,7492	4,4506
* 5	VIRGINIC	133,0668	5,2529	7,2359
6	SETOSA	1,3337	84,0118	170,0569
7	VIRGINIC	173,1838	26,5620	11,0484
8	VERSICOL	131,6617	8,4307	14,7647
* 9	VERSICOL	130,8624	8,6697	6,5068
10	SETOSA	2,2864	113,6509	210,0239
11	VERSICOL	99,2338	1,2963	13,8174
* 12	VERSICOL	149,0303	8,4393	4,8645
13	VIRGINIC	158,9817	12,7512	1,2342
14	VERSICOL	79,1079	1,4076	26,6531
15	VIRGINIC	161,8529	12,1703	1,9781
16	VIRGINIC	174,0819	16,0529	2,3902
17	VIRGINIC	209,0295	29,5143	1,9395
18	SETOSA	2,7690	67,4717	145,7007
19	VERSICOL	104,5191	2,7936	15,7979
20	VIRGINIC	188,8750	28,1929	5,6252
21	VIRGINIC	213,4056	37,9708	11,1557
22	VERSICOL	102,0479	3,2965	11,7819
23	VIRGINIC	215,2034	36,5459	6,1622
24	VIRGINIC	230,2236	35,8899	7,4448
25	VIRGINIC	160,5946	17,3403	11,3925
26	SETOSA	3,9513	113,1501	210,1669
27	VIRGINIC	183,5631	26,1715	7,1685
28	VERSICOL	86,5713	5,0343	23,2534
29	VERSICOL	88,8713	1,8095	16,2189
30	VERSICOL	75,6723	1,8657	25,0269
31	SETOSA	1,8029	73,8199	159,0406
32	VIRGINIC	153,7618	11,9778	3,9269
33	VERSICOL	80,1609	1,9167	27,3454
34	VIRGINIC	166,9178	14,4570	4,2028
35	VIRGINIC	184,8178	24,9964	3,8858
36	SETOSA	0,7032	97,5630	190,4399
37	SETOSA	1,7964	105,3213	200,5307
38	VERSICOL	117,0332	7,6733	14,2231
39	VIRGINIC	176,5572	26,7790	11,6559
40	SETOSA	2,0048	84,0333	171,8626

63. *Mahalanobisovy vzdálenosti:* jde o míru vzdálenosti ve vícerozměrném prostoru. Nesprávné zařazení při klasifikaci je označeno hvězdičkou.

STATISTICA Cz - [PS 1* - Aposteriorní pravděpodobnosti (73Iris.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram
Kategoriz. histogram
Krabicový graf dle sku
Krabicový graf dle sku
Krabicový graf dle sku
Krabicový graf dle sku
Kategoriz. bodový gr.
Kategoriz. bodový gr.
Kategoriz. bodový gr.
Kategoriz. pravděpoc
Kategoriz. pravděpoc
Kategoriz. pravděpoc
Kategoriz. pravděpoc
Kategoriz. pravděpoc
Kategoriz. pravděpoc
Kategoriz. pravděpoc
Korelace (73Iris.sta 5
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Kovariance (celkové)
Korelace (celkové) (7:
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Kovariance (celkové)
Korelace (celkové) (7:
Kovariance (73Iris.sta
Vnitřní korelace (73Iri
Korelace (73Iris.sta 5
Krabicový graf
Krabicový graf
Dialog výsledků diskrimina
Výsledky diskriminačn
Diskriminační analýza (73Iris.s
Dialog výsledků diskrimina
Výsledky diskriminačn

Aposteriorní pravděpodobnosti (73Iris.sta)
Nesprávná klasifikace je označena *

Případ	Pozorov. Klasif.	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
1	SETOSA	1,000000	0,000000	0,000000
2	VIRGINIC	0,000000	0,000003	0,999997
3	VERSICOL	0,000000	0,995590	0,004410
4	VIRGINIC	0,000000	0,000001	0,999999
* 5	VIRGINIC	0,000000	0,729388	0,270612
6	SETOSA	1,000000	0,000000	0,000000
7	VIRGINIC	0,000000	0,000428	0,999572
8	VERSICOL	0,000000	0,959573	0,040427
* 9	VERSICOL	0,000000	0,253228	0,746772
10	SETOSA	1,000000	0,000000	0,000000
11	VERSICOL	0,000000	0,998093	0,001907
* 12	VERSICOL	0,000000	0,143392	0,856608
13	VIRGINIC	0,000000	0,003146	0,996854
14	VERSICOL	0,000000	0,999997	0,000003
15	VIRGINIC	0,000000	0,006084	0,993916
16	VIRGINIC	0,000000	0,001078	0,998922
17	VIRGINIC	0,000000	0,000001	0,999999
18	SETOSA	1,000000	0,000000	0,000000
19	VERSICOL	0,000000	0,998502	0,001498
20	VIRGINIC	0,000000	0,000013	0,999987
21	VIRGINIC	0,000000	0,000002	0,999998
22	VERSICOL	0,000000	0,985835	0,014165
23	VIRGINIC	0,000000	0,000000	1,000000
24	VIRGINIC	0,000000	0,000001	0,999999
25	VIRGINIC	0,000000	0,048620	0,951380
26	SETOSA	1,000000	0,000000	0,000000
27	VIRGINIC	0,000000	0,000075	0,999925
28	VERSICOL	0,000000	0,999889	0,000111
29	VERSICOL	0,000000	0,999257	0,000743
30	VERSICOL	0,000000	0,999991	0,000009
31	SETOSA	1,000000	0,000000	0,000000

64. *A posteriorní pravděpodobnosti:* Každá třída je dokumentována svou vypočtenou čili *a posteriorní* pravděpodobností. Největší hodnota pravděpodobnosti ve sloupečku ukazuje, že kosatec náleží do třídy dotyčného sloupce.

Kosatce (Iris)

Iris Setosa

Wild Iris - *Iris setosa*

Iris Versicolor

Iris Virginica

Kosatce (Iris)

Iris Setosa

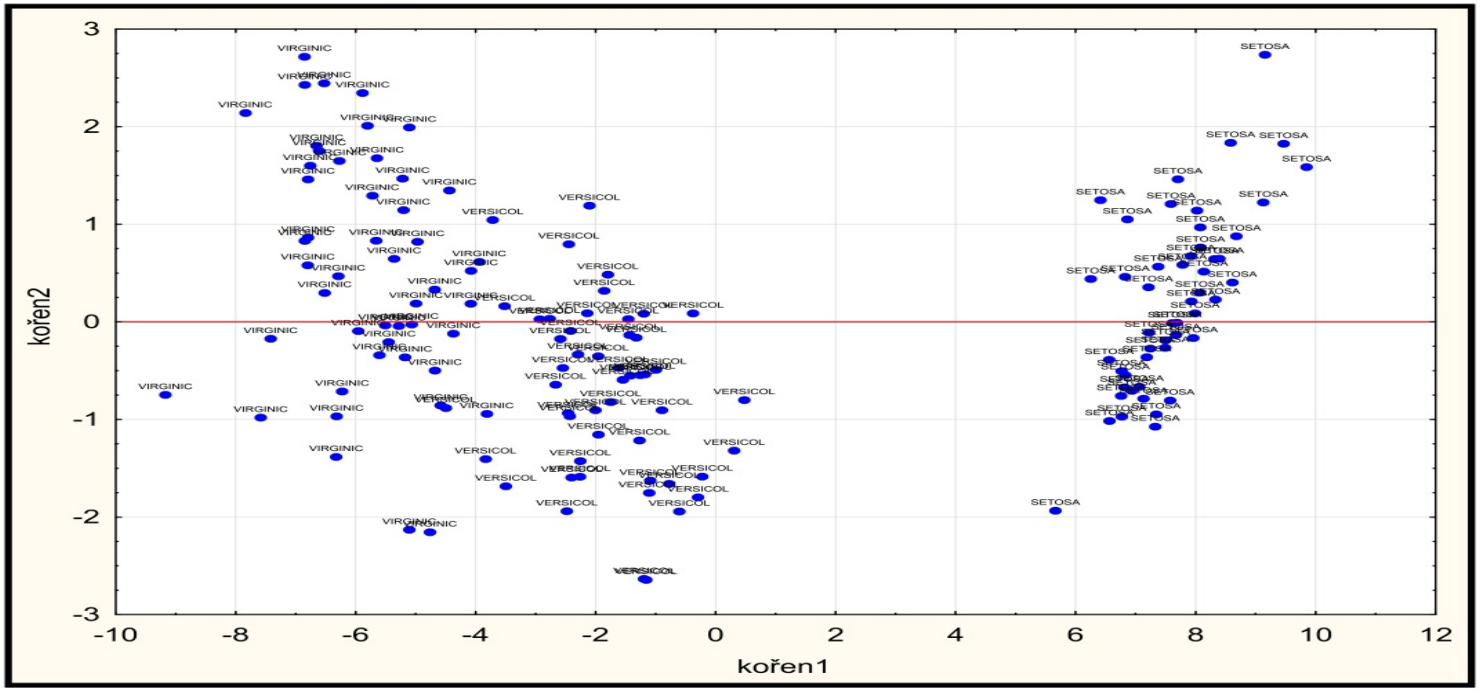
Iris Versicolor

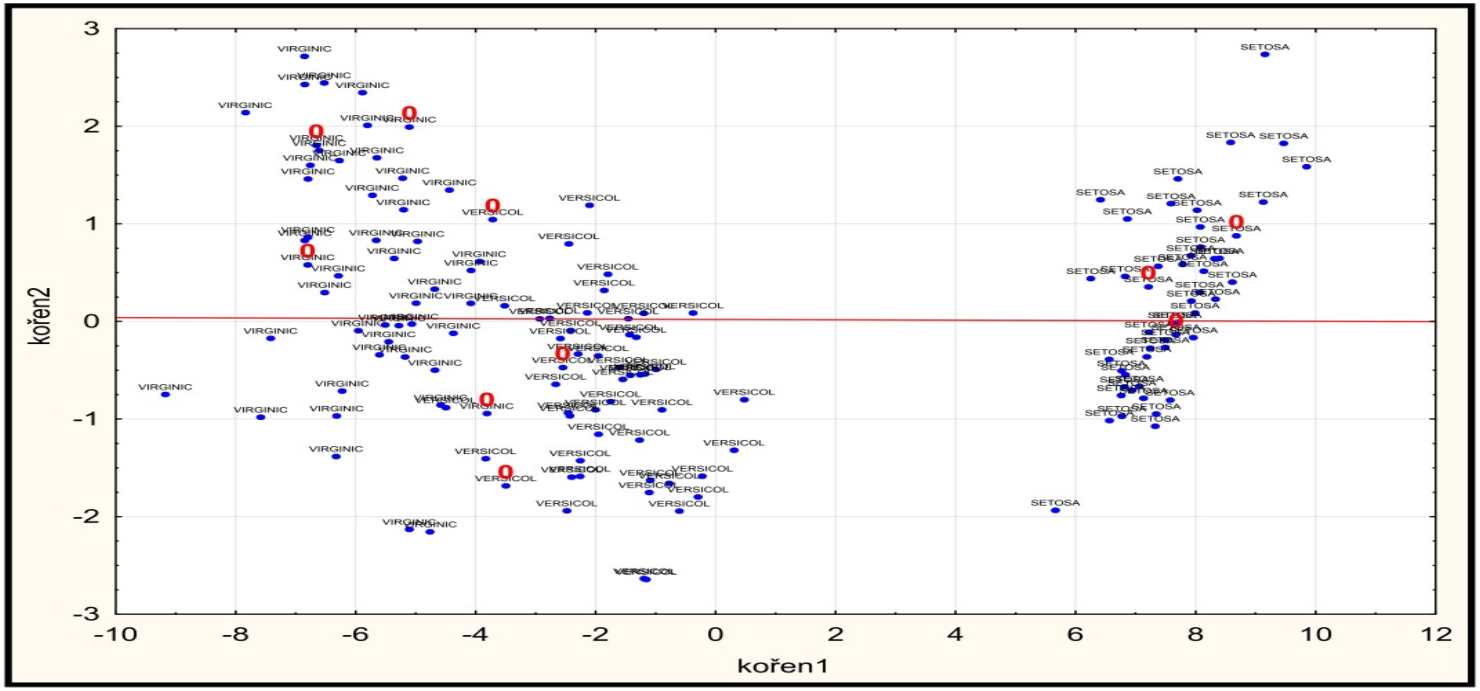
Iris Versicolor

Iris Virginica

Iris Virginica

Iris Virginica


kořen2
kořen1
3
2
1
0
-1
-2
-3
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
VIRGINIC
VERSICOL
SETOSA


kořen2
kořen1
3
2
1
0
-1
-2
-3
-10
-8
-6
-4
-2
0
2
4
6
8
10
12
VIRGINIC
VERSICOL
SETOSA

Skup.	Průměry kan. proměnných (73Iris)	
	Kořen1	Kořen2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

Skup.	Klasifikační matice (73Iris) Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace			
	% správnýc	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SETOSA	100,0000	50	0	0
VERSICOL	96,0000	0	48	2
VIRGINIC	98,0000	0	1	49
Celkem	98,0000	50	49	51

Skup.	Průměry kan. proměnných (73Iris-klas2)	
	Kořen1	Kořen2
SETOSA	7,60760	0,215133
VERSICOL	-1,82505	-0,727900
VIRGINIC	-5,78255	0,512767

Skup.	Klasifikační matice (73Iris-klas2) Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace			
	% správnýc	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SETOSA	100,0000	50	0	0
VERSICOL	96,0000	0	48	2
VIRGINIC	98,0000	0	1	49
Celkem	98,0000	50	49	51

Proměnná	Klasifikační funkce; grupovací : Trida (73Iris)		
	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SEPALLENGTH	23,5442	15,6982	12,446
SEPALWIDTH	23,5879	7,0725	3,685
PETALLENGTH	-16,4306	5,2115	12,767
PETALWIDTH	-17,3984	6,4342	21,079
Konstant	-86,3085	-72,8526	-104,368

Proměnná	Faktorová strukturní matice (73Iris) Korelační proměnné - Kanonické kořeny (vnitřní korelace)	
	Kořen1	Kořen2
SEPALLENGTH	-0,222596	0,310812
SEPALWIDTH	0,119012	0,863681
PETALLENGTH	-0,706065	0,167701
PETALWIDTH	-0,633178	0,737242

Proměnná	Klasifikační funkce; grupovací : Trida (73Iris-klas2)		
	SETOSA p=,33333	VERSICOL p=,33333	VIRGINIC p=,33333
SEPALLENGTH	23,5442	15,6982	12,446
SEPALWIDTH	23,5879	7,0725	3,685
PETALLENGTH	-16,4306	5,2115	12,767
PETALWIDTH	-17,3984	6,4342	21,079
Konstant	-86,3085	-72,8526	-104,368

Proměnná	Faktorová strukturní matice (73Iris-klas2) Korelační proměnné - Kanonické kořeny (vnitřní korelace)	
	Kořen1	Kořen2
SEPALLENGTH	-0,222596	0,310812
SEPALWIDTH	0,119012	0,863681
PETALLENGTH	-0,706065	0,167701
PETALWIDTH	-0,633178	0,737242

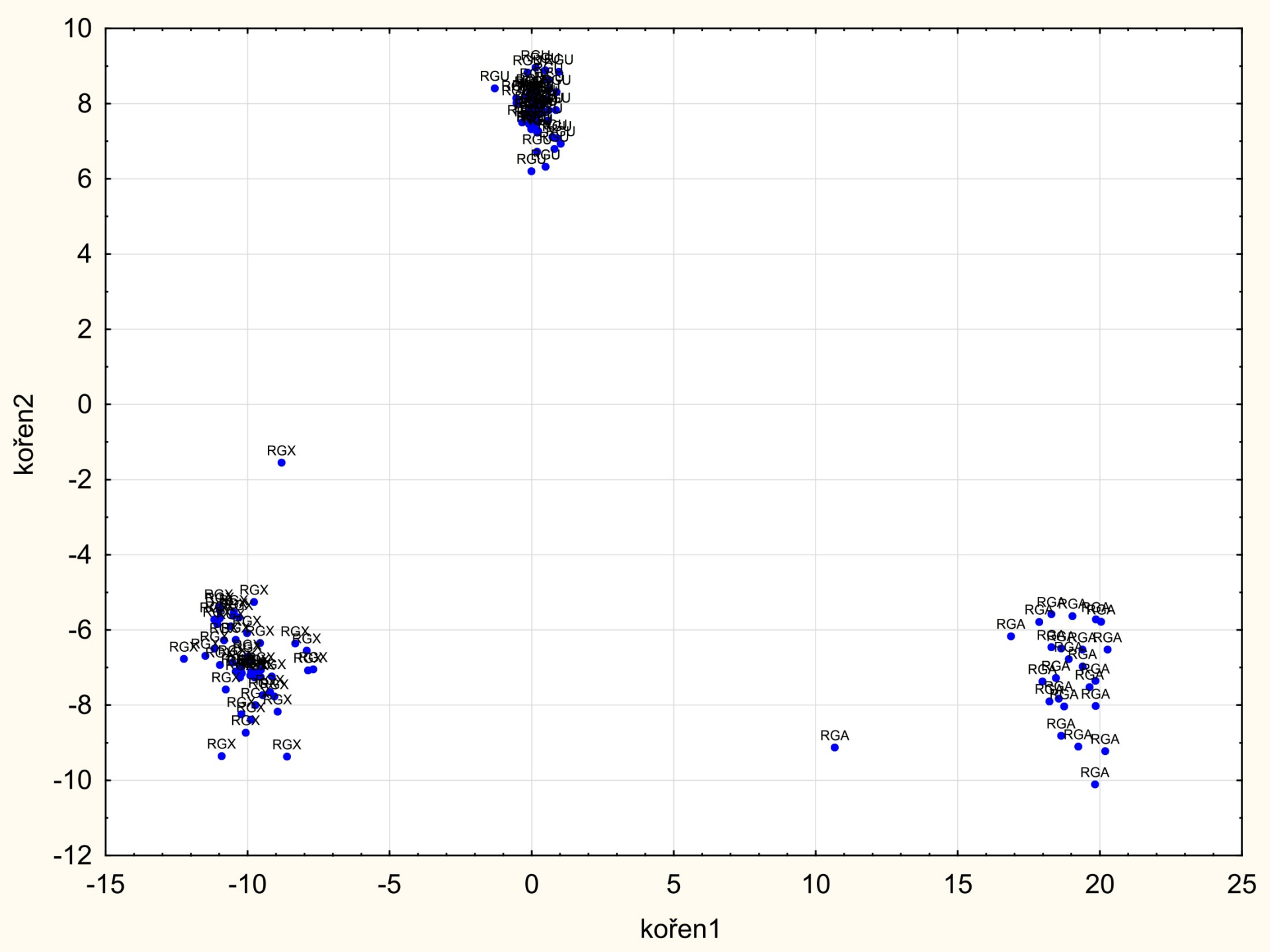
Úloha C32 ***Klasifikace a struktura znaků oxidu titaničitého dle chemických a fyzikálních vlastností*** **(DA)**

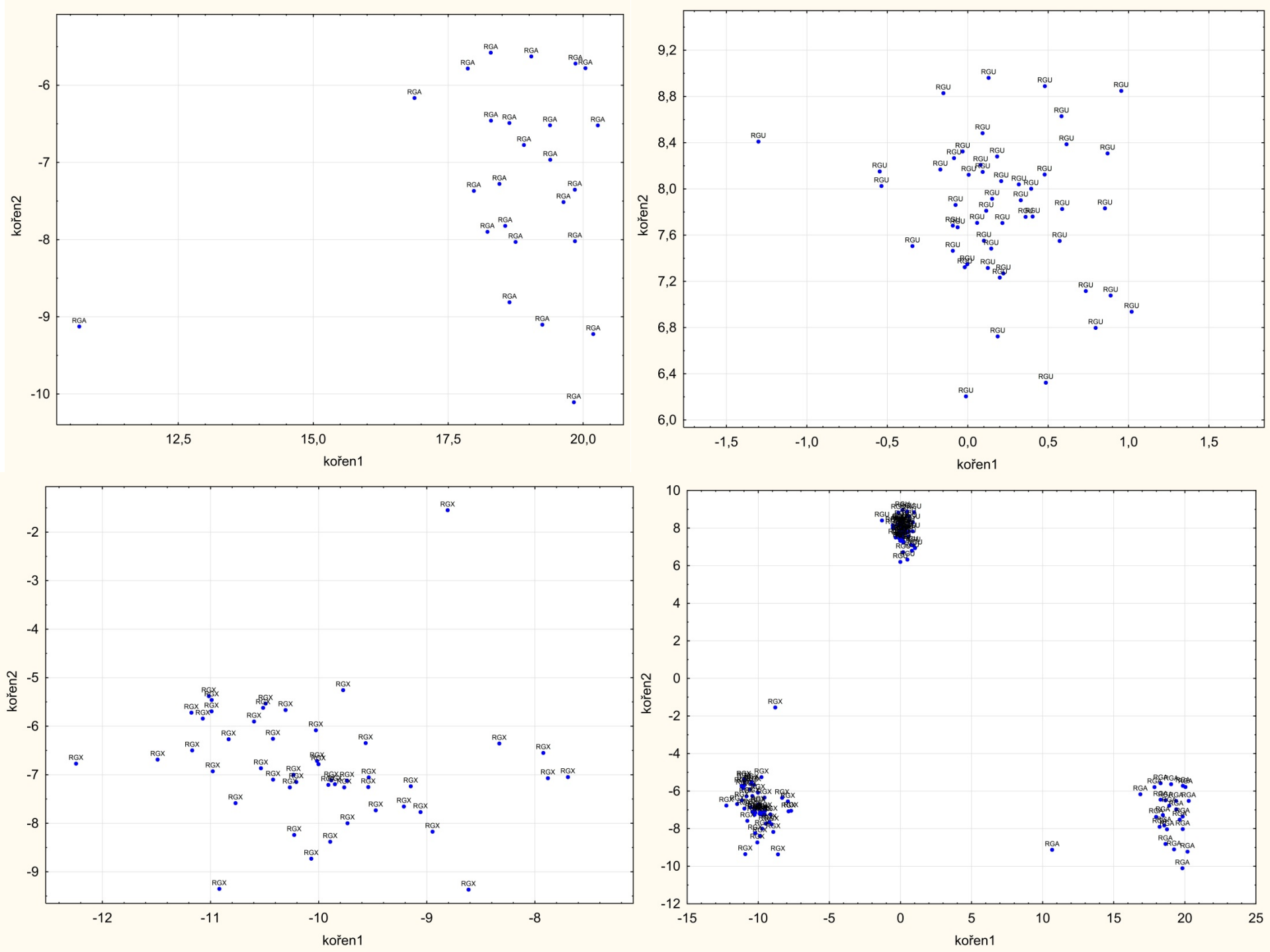
Výsledky analýz tří skupin bílého pigmentu oxidu titaničitého. Proveďte klasifikaci diskriminační analýzou (DA).

◯ ***Data:*** Popis znaků ve sloupcích: Data **ID**, Zařazení do tříd **Klas**, tj. typy RGA, RG*X*, RGU tvoří tréninková data. Diskriminátory:

Ti obsah Ti (%),
TL obsah těkavých látek (%),
SpO je spotřeba oleje (g/100 g pigmentu),
Barv značí barvivost,
Podt značí podtón,
Si značí obsah Si (%),
Al značí obsah Al (%).

ID	Ti	TL	SpO	Barv	Podt	Si	Al	Klas
1	93.78	0.5	24.2	1870	10	2.174	3.511	RGA
.....	.....	.....	.....	.....	.....	.....	.....	.....
144	98.21	0.2	19	1860	13	0.354	1.014	RGX


kořen2
kořen1
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-15
-10
-5
0
5
10
15
20
25
RGU
RGX
RGA


kořen2
kořen1
-6
-7
-8
-9
-10
12,5
15,0
17,5
20,0
RGA
9,2
8,8
8,4
8,0
7,6
7,2
6,8
6,4
6,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
RGU
-2
-3
-4
-5
-6
-7
-8
-9
-12
-11
-10
-9
-8
RGX
10
8
6
4
2
0
-2
-4
-6
-8
-10
-12
-15
-10
-5
0
5
10
15
20
25

N=144	Výsledky diskriminační funkční analýzy (C32) Počet prom. v modelu: 7; grupovací: klas (3 skup) Wilk. lambda: ,00018 přibliž F (14,270)=1436,8 p<0,0000					
	Wilk. Lambda	Parc. Lambda	F na vyj (2,135)	p-hodn.	Toler.	1-toler. R^2
Ti	*0,000268*	*0,654340*	*35,65741*	*0,000000*	*0,019606*	*0,980394*
TL	0,000177	0,988528	0,78338	0,458925	0,892277	0,107723
SpO	*0,000210*	*0,836343*	*13,20851*	*0,000006*	*0,538857*	*0,461143*
Barv	0,000177	0,989290	0,73079	0,483425	0,735078	0,264922
Podt	*0,000223*	*0,785127*	*18,47337*	*0,000000*	*0,903931*	*0,096069*
Si	*0,000205*	*0,857679*	*11,20079*	*0,000032*	*0,134136*	*0,865864*
Al	*0,000350*	*0,501522*	*67,09041*	*0,000000*	*0,029380*	*0,970620*

Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (C32)					
	Vlastní číslo	Kan. R	Wilk. Lambda	Chi-kv.	sv	p-hodn.
0	99,74799	0,995025	0,000175	1193,457	14	0,00
1	55,57775	0,991123	0,017675	556,915	6	0,00

Proměnná	Klasifikační funkce; grupovací : klas (C32)		
	RGA p=,18056	RGU p=,47222	RGX p=,34722
Ti	170186	170628	170317
TL	-17743	-17802	-17766
SpO	-393	-392	-378
Barv	24	24	24
Podt	-530	-540	-532
Si	154897	154935	154649
Al	188781	189293	188809
Konstant	-8486890	-8529983	-8499422

Skup.	Klasifikační matice (C32) Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace			
	% správnýc	RGA p=,18056	RGU p=,47222	RGX p=,34722
RGA	100,0000	26	0	0
RGU	100,0000	0	68	0
RGX	100,0000	0	0	50
Celkem	100,0000	26	68	50

Skup.	Průměry kan. proměnných (C32)	
	Kořen1	Kořen2
RGA	18,6583	-7,28080
RGU	0,2259	7,79704
RGX	-10,0096	-6,81795

Proměnná	Faktorová strukturní matice (C32) Korelační proměnné - Kanonické kořeny (vnitřní korelace)	
	Kořen1	Kořen2
Ti	-0,864629	-0,163046
TL	0,232129	0,017429
SpO	0,237965	-0,160084
Barv	0,019496	0,007877
Podt	-0,153956	-0,300398
Si	0,874747	-0,256509
Al	0,697907	0,407525

Proměnná	Standardiz. koeficienty (C32) pro kanonické proměnné	
	Kořen1	Kořen2
Ti	-0,72373	4,17371
TL	0,02890	-0,11067
SpO	-0,38308	-0,40158
Barv	0,02454	-0,11927
Podt	0,03374	-0,49075
Si	0,57408	0,86484
Al	-0,05554	4,15558
Vlastní	99,74799	55,57775
KumPodíl	0,64219	1,00000

Případ	Klasifikace případů (C32) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	1 p=,18056	2 p=,47222	3 p=,34722
2	RGA	RGA	RGU	RGX
3	RGA	RGA	RGU	RGX
4	RGA	RGA	RGU	RGX
5	RGA	RGA	RGU	RGX
6	RGA	RGA	RGU	RGX
7	RGA	RGA	RGU	RGX
8	RGA	RGA	RGU	RGX
9	RGA	RGA	RGU	RGX
10	RGA	RGA	RGU	RGX
11	RGA	RGA	RGU	RGX
12	RGA	RGA	RGU	RGX
13	RGA	RGA	RGU	RGX
14	RGA	RGA	RGU	RGX
15	RGA	RGA	RGU	RGX
16	RGA	RGA	RGU	RGX
17	RGA	RGA	RGU	RGX
18	RGA	RGA	RGU	RGX
19	RGA	RGA	RGU	RGX
20	RGA	RGA	RGU	RGX
21	RGA	RGA	RGU	RGX
22	RGA	RGA	RGU	RGX
23	RGA	RGA	RGU	RGX
24	RGA	RGA	RGU	RGX
25	RGA	RGA	RGU	RGX
26	RGA	RGA	RGU	RGX
27	RGA	RGA	RGU	RGX
28	RGU	RGU	RGA	RGX
29	RGU	RGU	RGA	RGX
30	RGU	RGU	RGA	RGX
31	RGU	RGU	RGA	RGX
32	RGU	RGU	RGA	RGX
33	RGU	RGU	RGA	RGX
34	RGU	RGU	RGA	RGX
35	RGU	RGU	RGA	RGX
36	RGU	RGU	RGA	RGX
37	RGU	RGU	RGA	RGX
38	RGU	RGU	RGA	RGX
39	RGU	RGU	RGA	RGX
40	RGU	RGU	RGA	RGX
41	RGU	RGU	RGA	RGX
42	RGU	RGU	RGA	RGX
43	RGU	RGU	RGA	RGX
44	RGU	RGU	RGA	RGX
45	RGU	RGU	RGA	RGX
46	RGU	RGU	RGA	RGX
47	RGU	RGU	RGA	RGX
48	RGU	RGU	RGA	RGX
49	RGU	RGU	RGA	RGX
50	RGU	RGU	RGA	RGX
51	RGU	RGU	RGA	RGX

Případ	Mahalanobisovy vzdálenosti (C32) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	RGA p=,18056	RGU p=,47222	RGX p=,34722
2	RGA	10,1351	623,1396	900,3110
3	RGA	10,0932	541,3123	852,3933
4	RGA	8,0935	500,6362	783,0421
5	RGA	5,0655	574,8432	802,9057
6	RGA	4,8171	614,9479	883,0971
7	RGA	8,5472	550,8557	828,4168
8	RGA	7,8208	536,6371	808,1469
9	RGA	23,7401	525,8623	823,0444
10	RGA	3,3003	562,5682	813,3089
11	RGA	3,3003	562,5682	813,3089
12	RGA	8,3897	641,5818	899,2757
13	RGA	6,4779	599,4259	834,2990
14	RGA	10,8236	590,3975	827,4419
15	RGA	5,9304	590,4635	869,6200
16	RGA	79,4508	407,4983	444,9338
17	RGA	8,4340	612,0465	922,2669
18	RGA	8,1832	572,3218	897,4604
19	RGA	4,8689	575,7915	867,9421
20	RGA	9,1453	581,9042	909,0862
21	RGA	7,2044	619,5528	829,3455
22	RGA	4,3358	565,1535	839,9256
23	RGA	11,4189	479,1127	730,1782
24	RGA	4,1241	548,6656	787,1074
25	RGA	7,0487	650,5714	864,3236
26	RGA	10,4597	692,3977	921,9067
27	RGA	13,0826	708,4438	904,7653
28	RGU	521,2330	9,3222	318,7957
29	RGU	536,7763	8,9696	275,9360
30	RGU	536,7763	8,9696	275,9360
31	RGU	543,4659	7,2903	293,4542
32	RGU	519,7687	6,6857	287,3059
33	RGU	519,7687	6,6857	287,3059
34	RGU	593,7277	10,4871	342,0955
35	RGU	575,4103	4,3821	331,6899
36	RGU	648,0285	6,1246	311,1339
37	RGU	648,0285	6,1246	311,1339
38	RGU	563,7344	1,9133	301,4111
39	RGU	530,8040	2,9600	311,8223
40	RGU	530,8040	2,9600	311,8223
41	RGU	530,8040	2,9600	311,8223
42	RGU	615,5059	3,3827	344,1681
43	RGU	558,3248	3,6562	330,2850
44	RGU	553,2376	8,2193	340,4370
45	RGU	550,5502	3,6501	321,8599
46	RGU	581,4843	2,4549	352,4435
47	RGU	581,4843	2,4549	352,4435
48	RGU	567,2060	0,1514	327,9353
49	RGU	582,9557	0,3958	326,2716
50	RGU	594,5043	1,5918	322,6992

Případ	Aposteriorní pravděpodobnosti (C32) Nesprávná klasifikace je označena *			
	Pozorova Klasif.	RGA p=,18056	RGU p=,47222	RGX p=,34722
2	RGA	1,000000	0,000000	0,000000
3	RGA	1,000000	0,000000	0,000000
4	RGA	1,000000	0,000000	0,000000
5	RGA	1,000000	0,000000	0,000000
6	RGA	1,000000	0,000000	0,000000
7	RGA	1,000000	0,000000	0,000000
8	RGA	1,000000	0,000000	0,000000
9	RGA	1,000000	0,000000	0,000000
10	RGA	1,000000	0,000000	0,000000
11	RGA	1,000000	0,000000	0,000000
12	RGA	1,000000	0,000000	0,000000
13	RGA	1,000000	0,000000	0,000000
14	RGA	1,000000	0,000000	0,000000
15	RGA	1,000000	0,000000	0,000000
16	RGA	1,000000	0,000000	0,000000
17	RGA	1,000000	0,000000	0,000000
18	RGA	1,000000	0,000000	0,000000
19	RGA	1,000000	0,000000	0,000000
20	RGA	1,000000	0,000000	0,000000
21	RGA	1,000000	0,000000	0,000000
22	RGA	1,000000	0,000000	0,000000
23	RGA	1,000000	0,000000	0,000000
24	RGA	1,000000	0,000000	0,000000
25	RGA	1,000000	0,000000	0,000000
26	RGA	1,000000	0,000000	0,000000
27	RGA	1,000000	0,000000	0,000000
28	RGU	0,000000	1,000000	0,000000
29	RGU	0,000000	1,000000	0,000000
30	RGU	0,000000	1,000000	0,000000
31	RGU	0,000000	1,000000	0,000000
32	RGU	0,000000	1,000000	0,000000
33	RGU	0,000000	1,000000	0,000000
34	RGU	0,000000	1,000000	0,000000
35	RGU	0,000000	1,000000	0,000000
36	RGU	0,000000	1,000000	0,000000
37	RGU	0,000000	1,000000	0,000000
38	RGU	0,000000	1,000000	0,000000
39	RGU	0,000000	1,000000	0,000000
40	RGU	0,000000	1,000000	0,000000
41	RGU	0,000000	1,000000	0,000000
42	RGU	0,000000	1,000000	0,000000
43	RGU	0,000000	1,000000	0,000000
44	RGU	0,000000	1,000000	0,000000
45	RGU	0,000000	1,000000	0,000000
46	RGU	0,000000	1,000000	0,000000

	C32		
	1 klas	2 kořen1	3 kořen2
1	RGA	19,85	-7,35
2	RGA	19,04	-5,63
3	RGA	17,86	-5,78
4	RGA	18,23	-7,90
5	RGA	19,64	-7,51
6	RGA	18,63	-6,49
7	RGA	18,29	-6,46
8	RGA	18,29	-5,58
9	RGA	18,45	-7,28
10	RGA	18,45	-7,28
11	RGA	19,85	-8,02
12	RGA	18,75	-8,03
13	RGA	18,55	-7,82
14	RGA	19,39	-6,97
15	RGA	10,67	-9,13
16	RGA	20,27	-6,52
17	RGA	19,86	-5,72
18	RGA	19,39	-6,52
19	RGA	20,04	-5,78
20	RGA	18,64	-8,81
21	RGA	18,90	-6,77
22	RGA	16,87	-6,17
23	RGA	17,98	-7,37
24	RGA	19,24	-9,10
25	RGA	20,19	-9,22
26	RGA	19,83	-10,11
27	RGU	1,02	6,94
28	RGU	-0,01	6,20
29	RGU	-0,01	6,20
30	RGU	0,19	6,72
31	RGU	0,49	6,32
32	RGU	0,49	6,32
33	RGU	0,18	8,28
34	RGU	0,32	8,04
35	RGU	-1,30	8,41
36	RGU	-1,30	8,41
37	RGU	-0,02	7,32
38	RGU	0,73	7,12
39	RGU	0,73	7,12
40	RGU	0,73	7,12
41	RGU	-0,15	8,83
42	RGU	0,59	7,83
43	RGU	0,85	7,83
44	RGU	0,57	7,55
45	RGU	0,58	8,63
46	RGU	0,58	8,63
47	RGU	0,39	8,00
48	RGU	0,09	8,15
49	RGU	-0,17	8,17
50	RGU	0,22	7,27
51	RGU	0,13	8,96

	C32		
	1 klas	2 kořen1	3 kořen2
52	RGU	0,09	8,48
53	RGU	0,15	7,48
54	RGU	0,15	7,91
55	RGU	0,95	8,85
56	RGU	0,95	8,85
57	RGU	0,48	8,89
58	RGU	0,06	7,71
59	RGU	0,06	7,71
60	RGU	-0,34	7,51
61	RGU	-0,06	7,67
62	RGU	-0,09	7,68
63	RGU	-0,09	7,68
64	RGU	0,11	7,81
65	RGU	-0,00	7,35
66	RGU	0,10	7,55
67	RGU	0,10	7,55
68	RGU	0,61	8,39
69	RGU	0,61	8,39
70	RGU	0,87	8,31
71	RGU	0,87	8,31
72	RGU	0,36	7,76
73	RGU	0,36	7,76
74	RGU	-0,03	8,32
75	RGU	-0,03	8,32
76	RGU	0,08	8,21
77	RGU	-0,54	8,02
78	RGU	-0,55	8,15
79	RGU	0,22	7,71
80	RGU	0,21	8,07
81	RGU	0,21	8,07
82	RGU	0,33	7,90
83	RGU	-0,09	8,27
84	RGU	-0,09	7,46
85	RGU	-0,08	7,86
86	RGU	0,20	7,23
87	RGU	0,89	7,08
88	RGU	0,80	6,80
89	RGU	0,01	8,12
90	RGU	0,12	7,32
91	RGU	0,12	7,32
92	RGU	0,48	8,12
93	RGU	0,48	8,12
94	RGU	0,40	7,76
95	RGX	-10,98	-6,93
96	RGX	-10,77	-7,58
97	RGX	-10,27	-7,26
98	RGX	-7,88	-7,07
99	RGX	-9,06	-7,77
100	RGX	-9,91	-7,21
101	RGX	-9,47	-7,73
102	RGX	-9,15	-7,24

	C32		
	1 klas	2 kořen1	3 kořen2
103	RGX	-9,76	-7,26
104	RGX	-10,92	-9,35
105	RGX	-8,81	-1,55
106	RGX	-9,77	-5,26
107	RGX	-10,07	-8,73
108	RGX	-8,95	-8,17
109	RGX	-8,33	-6,36
110	RGX	-7,70	-7,05
111	RGX	-8,61	-9,37
112	RGX	-9,89	-8,38
113	RGX	-7,92	-6,55
114	RGX	-11,49	-6,69
115	RGX	-12,24	-6,77
116	RGX	-11,18	-5,72
117	RGX	-10,83	-6,27
118	RGX	-10,60	-5,90
119	RGX	-11,17	-6,50
120	RGX	-9,74	-7,12
121	RGX	-9,54	-7,25
122	RGX	-10,99	-5,46
123	RGX	-10,31	-5,67
124	RGX	-10,53	-6,86
125	RGX	-9,85	-7,19
126	RGX	-9,57	-6,35
127	RGX	-10,21	-7,15
128	RGX	-10,21	-7,15
129	RGX	-10,42	-7,10
130	RGX	-10,23	-8,24
131	RGX	-10,02	-6,72
132	RGX	-10,00	-6,78
133	RGX	-9,88	-7,12
134	RGX	-9,73	-8,00
135	RGX	-9,54	-7,05
136	RGX	-9,21	-7,65
137	RGX	-10,23	-7,00
138	RGX	-11,07	-5,84
139	RGX	-10,03	-6,08
140	RGX	-10,49	-5,53
141	RGX	-10,42	-6,26
142	RGX	-11,02	-5,38
143	RGX	-10,51	-5,62
144	RGX	-10,99	-5,69

Úloha B4.21 *Ergometrické vyšetření ischemie myokardu (DA)*

U pacientů s ICHS byly ergometricky v intervalech měřeny biochemické parametry a výkon na kole (W/kg). Ke stanovení ischemií modifikovaného albuminu (IMA) byla odebírána krev v časových úsecích "před zátěží“ a "po zátěži". Výskyt ischemie byl indikován pomocí thaliové scintigrafie.

◯ *Data: Data ISCHEM o **85 pacientech** obsahují*
i (ID) značící identifikační číslo pacienta,
x1 (IMA1) koncentraci IMA měřenou před zátěží v klidu,
x2 (IMA2) značí koncentraci IMA měřenou při vrcholu zátěže ve svém maximu,
x3 (IMA3) značí koncentraci IMA měřenou 30 min po zátěži na ergometru,
x4 (IMA4) značí koncentraci IMA měřenou 60 min po zátěži,
x5 (HB1) značí koncentraci hemoglobinu naměřenou před zátěží,
x6 (HTK1) hodnota hematokritu vyjadřující poměr plasmy a krevních buněk před zátěží,
x7 (HB2) značí koncentraci hemoglobinu po zátěži,
x8 (HTK2) značí hodnotu hematokritu naměřenou po zátěži,
x9 (CRP) značí C-reaktivní bílkovinu, která se vyskytuje při zánětlivých procesech a je markrem možného vzniku ischemie srdeční,
x10 (LAKTAT1) značí kyselinu mléčnou, která vzniká ve svalech při tělesné námaze a která byla naměřena před zátěží,
x11 (LAKTAT2) značí značí kyselinu mléčnou, která vzniká ve svalech při tělesné námaze a která byla naměřena po zátěži,
x12 (ERGOPOZ) je označení jednotlivců, kteří měli pozitivní nález na EKG,
x13 (VYŠLAP) značí jednotlivce, kteří dokončili (1) a kteří pro vyčerpání nedokončili ergometrii (2),
x14 (W/kg) značí dosažený maximální výkon na ergometru, přepočtený na 1 kg hmotnosti,
x15 (SKUP) značí rozdělení pacientů do skupin, podle závažnosti onemocnění,
x16 (THAL) značí výsledek scintilační zobrazovací metody, kdy pacient vypije roztok s thaliem značkovaného obsahu a následnou obrazovou analýzou cév a srdce je potvrzeno nebo zamítnuto ischemické onemocnění srdce.

i	***x1***	***x2***	***x3***	***x4***	***x5***	***x6***	***x7***	***x8***	***x9***
	x10	***x11***	***x12***	***x13***	***x14***	***x15***	***x16***		
1	0.301	0.303	0.397	0.643	158	0.48	158	0.49	16.3
	2.7	10.21	1	1	2.82	4	2		
...	...	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...		
85	0.54	0.47	0.48	0.55	137	0.4	145	0.43	2.42
	1.45	5.22	2	1	1.75	3			

Data B4.21

STATISTICA Cz - [Data: B401.sta (237s krát 287ř)]

184 B420x6 DLOKET	185 B421i ID	186 B421x1 IMA1	187 B421x2 IMA2	188 B421x3 IMA3	189 B421x4 IMA4	190 B421x5 HB1	191 B421x6 HTK1	192 B421x7 HB2	193 B421x8 HTK2	194 B421x9 CRP	195 B421x10 LAKTAT1	196 B421x11 LAKTAT2	197 B421x12 ERGOPOZ	198 B421x13 VYSLAP	199 B421x14 W/KG	200 B421x15 SKUP	201 B421x16 THAL	202 B422i ID	203 B422x1 GROUP	204 B422x2 IDPAC
1	1	0.301	0.303	0.397	0.643	158	0.48	158	0.49	16.3	2.7	10.21	1	1	2.82	4	2	1	1	9
2	2	0.639	0.252	0.397	0.275	132	0.42	144	0.45	5.4	2.01	4.53	1	1	1.1	1	*	2	1	7
3	3	0.487	0.399	0.691	0.912	137	0.43	150	0.46	2	1.39	4.4	2	1	2.09	3	2	3	1	10
4	4	34	0.345	0.28	0.1	[illegible]	0.44	155	0.48	11.6	1.99	8.82	2	1	1.49	1	1	4	1	1
5	5	0.4[illegible]	0.192	0.9[illegible]	[illegible]	136	0.42	148	0.45	2	1.41	7.2	1	1	1.37	1	2	5	1	3
6	6	0.796	0.396	0.521	0.295	146	0.44	161	0.5	17.5	1.81	8.82	1	2	1.87	2	*	6	1	5
7	7	1.244	0.425	0.278	0.328	130	0.39	142	0.44	10.8	1.65	6.57	2	1	1.19	2	2	7	1	1
8	8	0.375	0.35	0.346	0.447	150	0.45	172	0.52											
9	9	0.798	0.871	0.734	0.442	140	0.43	141	0.44											
10	10	0.401	0.384	0.425	0.423	143	0.44	157	0.49											
11	11	0.428	0.3	0.333	0.374	158	0.49	166	0.52											
12	12	0.545	0.804	0.497	0.534	154	0.48	162	0.52											
13	13	0.383	0.422	0.486	0.686	167	0.53	172	0.53											
14	14	0.391	0.362	0.438	0.596	144	0.42	161	0.49											
15	15	0.647	0.408	0.519	0.397	122	0.37	138	0.4											
16	16	0.596	0.367	0.726	0.347	165	0.51	168	0.54											
17	17	0.412	0.368	0.424	0.447	121	0.38	137	0.43											
18	18	0.396	0.371	0.435	0.405	150	0.44	160	0.48											
19	19	0.6	0.576	0.43	0.253	130	0.4	135	0.42											
20	20	0.329	0.585	0.379	0.571	139	0.41	148	0.47											
21	21	0.525	0.512	0.547	0.39	123	0.39	128	0.4											
22	22	0.468	0.715	0.431	0.429	142	0.43	153	0.47											
23	23	0.438	0.532	0.403	0.51	135	0.41	146	0.44											
24	24	0.379	0.334	0.394	0.355	146	0.43	156	0.48											
25	25	0.464	0.438	0.395	0.454	142	0.43	152	0.46											
26	26	0.372	0.384	0.312	0.351	141	0.44	149	0.49											
27	27	0.364	0.762	0.477	0.469	138	0.41	147	0.44											
28	28	0.304	0.325	0.359	0.321	111	0.34	118	0.37											
29	29	0.533	0.628	0.447	0.538	132	0.4	143	0.43											
30	30	0.449	0.463	0.408	0.458	159	0.5	158	0.49											
31	31	0.273	0.378	0.414	0.37	132	0.41	152	0.46											
32	32	0.439	0.423	0.281	0.396	154	0.46	162	0.49											
33	33	0.477	0.357	0.443	0.328	138	0.41	148	0.44											
34	34	0.487	0.349	0.327	0.366	135	0.41	142	0.43											
35	35	0.355	0.371	0.375	0.43	153	0.47	163	0.5											
36	36	0.393	0.511	0.418	0.576	153	0.48	152	0.49											
37	37	0.431	0.362	0.449	0.423	149	0.45	151	0.47											
38	38	0.474	0.408	0.43	0.487	131	0.4	132	0.41											
39	39	0.394	0.401	0.335	0.599	150	0.46	155	0.48											
40	40	0.39[illegible]	0.362	0.415	0.78[illegible]	149	0.46	159	0.51											

Zadání výpočtu DA

STATISTICA Cz - [Data: B401.sta (237s krát 287ř)]

Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Data Okno Nápověda

- Obnovit... Ctrl+R
- Základní statistiky/tabulky
- Vícenásobná regrese
- ANOVA
- Neparametrická statistika
- Prokládání rozdělení
- Rozdělení & simulace
- Pokročilé lineární/nelineární modely
- Vícerozměrné průzkumné techniky
- Průmyslová statistika & Six Sigma
- Analýza síly testu
- Automatizované neuronové sítě
- PLS, PCA, vícerozměrné/dávkové SPC
- Variance Estimation and Precision
- Statistiky bloku dat
- STATISTICA Visual Basic
- Dávková (po skupinách) analýza
- Pravděpodobnostní kalkulátor

- Shluková analýza
- Faktorová analýza
- Hlavní komponenty & klasifikační analýza
- Kanonická analýza
- Analýza spolehlivosti/položek
- Klasifikační stromy
- Korespondenční analýza
- Vícerozměrné škálování
- Diskriminační analýza
- Modely obecné diskriminační analýzy

Zadání proměnných

Výběr jednu grupovací proměnnou a seznam nezávislých proměnných

201 - B421x16	211 - B423i
202 - B422i	212 - B423x1
203 - B422x1	213 - B423x2
204 - B422x2	214 - B423x3
205 - B422x3	215 - B423x4
206 - B422x4	216 - B423x5
207 - B422x5	217 - B423x6
208 - B422x6	218 - B424i
209 - B422x7	219 - B424x1
210 - B422x8	220 - B424x2

191 - B421x6	201 - B421x16
192 - B421x7	202 - B422i
193 - B421x8	203 - B422x1
194 - B421x9	204 - B422x2
195 - B421x10	205 - B422x3
196 - B421x11	206 - B422x4
197 - B421x12	207 - B422x5
198 - B421x13	208 - B422x6
199 - B421x14	209 - B422x7
200 - B421x15	210 - B422x8

OK

Storno

[Svazky]...

Pro zobrazení odpovídajících proměnných zvolte "Ukázat pouze...". Pro více informací stiskněte F1.

Vybrat vše | Podrobn. | Přiblížit

Vybrat vše | Podrobn. | Přiblížit

Groupovací proměnná: 201

Seznam nezáv. proměnných: 186-200

Průběh výpočtu DA

Diskriminační analýza: B401.sta

```
Počet proměnných v modelu:15

Wilks. lambda: .3557155   přibliž F (30,102) = 2.300692 p <  .0011
```

Základní nastavení | Detaily | Klasifikace

Výpočet: proměnné v modelu

Proměnné neobsažené v modelu

Výpočet

Storno

Možnosti

Anal. sk.

Výsledky diskriminační funkční analýzy (B401.sta)
Počet prom. v modelu: 15; grupovací: B421x16 (3 skup)
Wilks. lambda: .35572 přibliž F (30,102)=2.3007 p< .0011

N=68	Wilks. Lambda	Parc. Lambda	F na vyj (2,51)	p-hodn.	Toler.	1-toler. R^2
B421x1	0.359050	0.990714	0.239022	0.788274	0.792714	0.207286
B421x2	0.4246	0.8378	4.9353	0.0110	0.7969	0.2031
B421x3	0.359484	0.989516	0.270181	0.764327	0.750422	0.249578
B421x4	0.395285	0.899896	2.836606	0.067906	0.701408	0.298592
B421x5	0.375870	0.946379	1.444801	0.245280	0.025654	0.974346
B421x6	0.377387	0.942575	1.553557	0.221336	0.027352	0.972648
B421x7	0.4175	0.8520	4.4280	0.0169	0.0425	0.9575
B421x8	0.4009	0.8872	3.2407	0.0473	0.0520	0.9480
B421x9	0.384100	0.926100	2.034821	0.141181	0.767322	0.232678
B421x10	0.365271	0.973839	0.685012	0.508661	0.790165	0.209835
B421x11	0.358681	0.991731	0.212609	0.809185	0.779348	0.220652
B421x12	0.367781	0.967194	0.864931	0.427165	0.705663	0.294337
B421x13	0.374493	0.949858	1.346119	0.269337	0.755379	0.244621
B421x14	0.378078	0.940852	1.603097	0.211248	0.494063	0.505937
B421x15	0.4793	0.7421	8.8623	0.0005	0.5726	0.4274

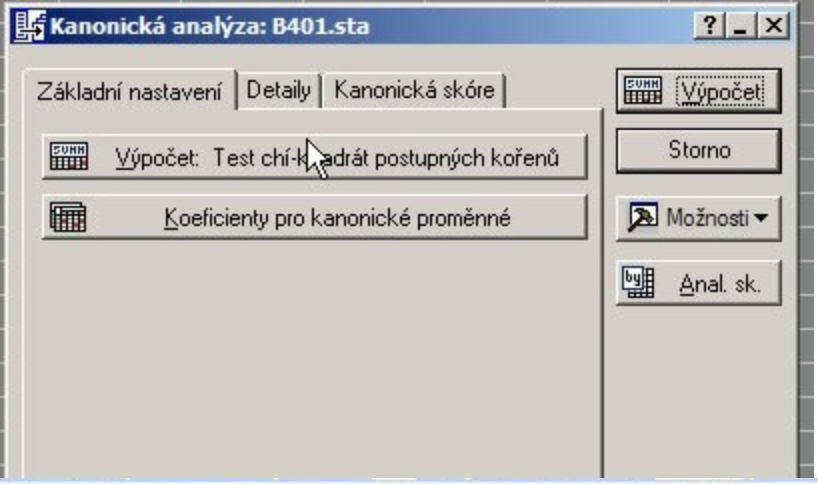
Výstavba Fisherovy diskriminační funkce.

Červeně jsou statisticky významné parametry.

Testační kritéria jsou
Wilkovo lambda
Parc. Wilk. lambda
F
p spočtené
Tolerance
1-Tolerance

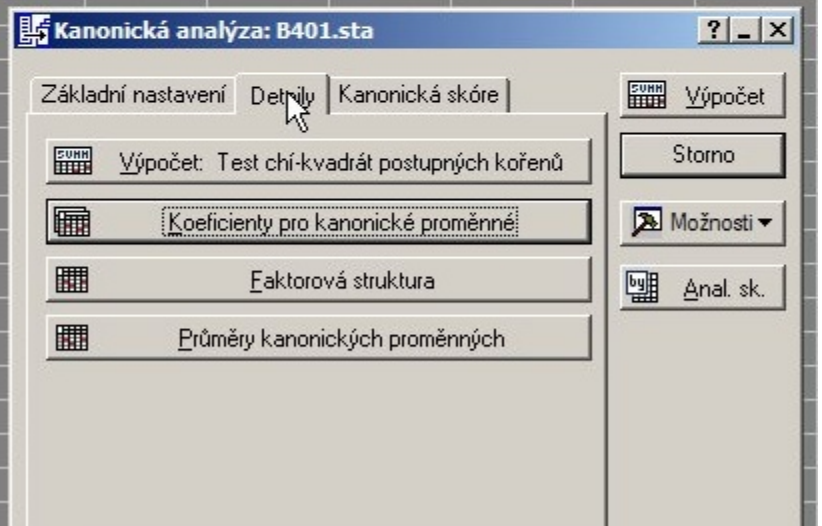
p-hodnot (B401.sta)

B421x16	G_1:1	G_2:2	*
G_1:1		0.063462	0.061393
G_2:2	0.063462		0.001397
*	0.061393	0.001397	

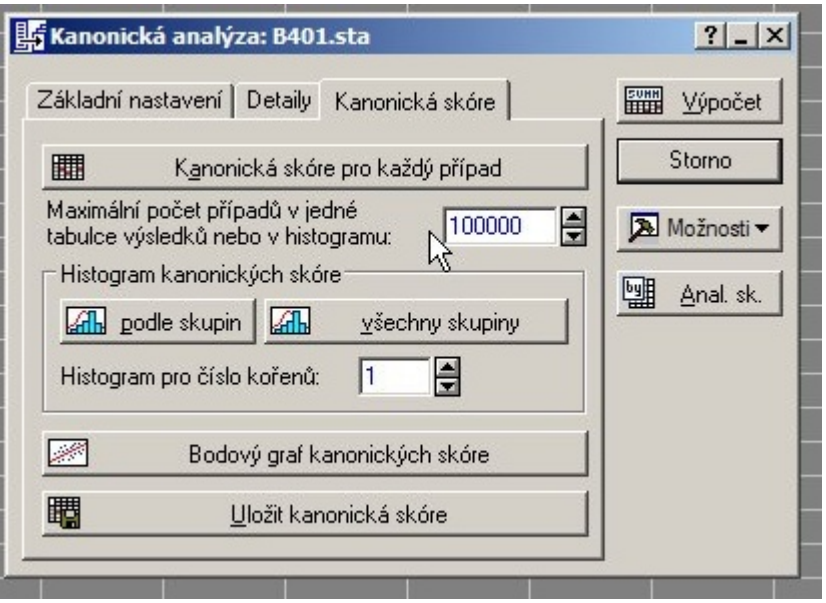


Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (B401.sta)					
	Vlastní číslo	Kan. R	Wilks. Lambda	Chi-kv.	sv	p-hodn.
0	0.904615	0.689173	0.355715	59.95020	30	0.000934
1	0.476013	0.567890	0.677501	22.58198	14	0.067421

Proměnná	Standardiz. koeficienty (B401.sta) pro kanonické proměnné	
	Kořen1	Kořen2
B421x1	-0.12338	0.11792
B421x2	-0.20143	-0.75581
B421x3	-0.14465	0.11182
B421x4	-0.48298	0.31462
B421x5	-2.00135	0.76302
B421x6	1.72689	-1.45539
B421x7	2.65348	-0.65991
B421x8	-2.08299	0.57436
B421x9	0.21368	-0.48103
B421x10	-0.02431	0.31904
B421x11	0.02288	0.17924
B421x12	-0.08888	0.36403
B421x13	0.37369	-0.01284
B421x14	-0.48796	-0.14336
B421x15	0.96477	0.16067
Vlastní	0.90461	0.47601
KumPodíl	0.65522	1.00000

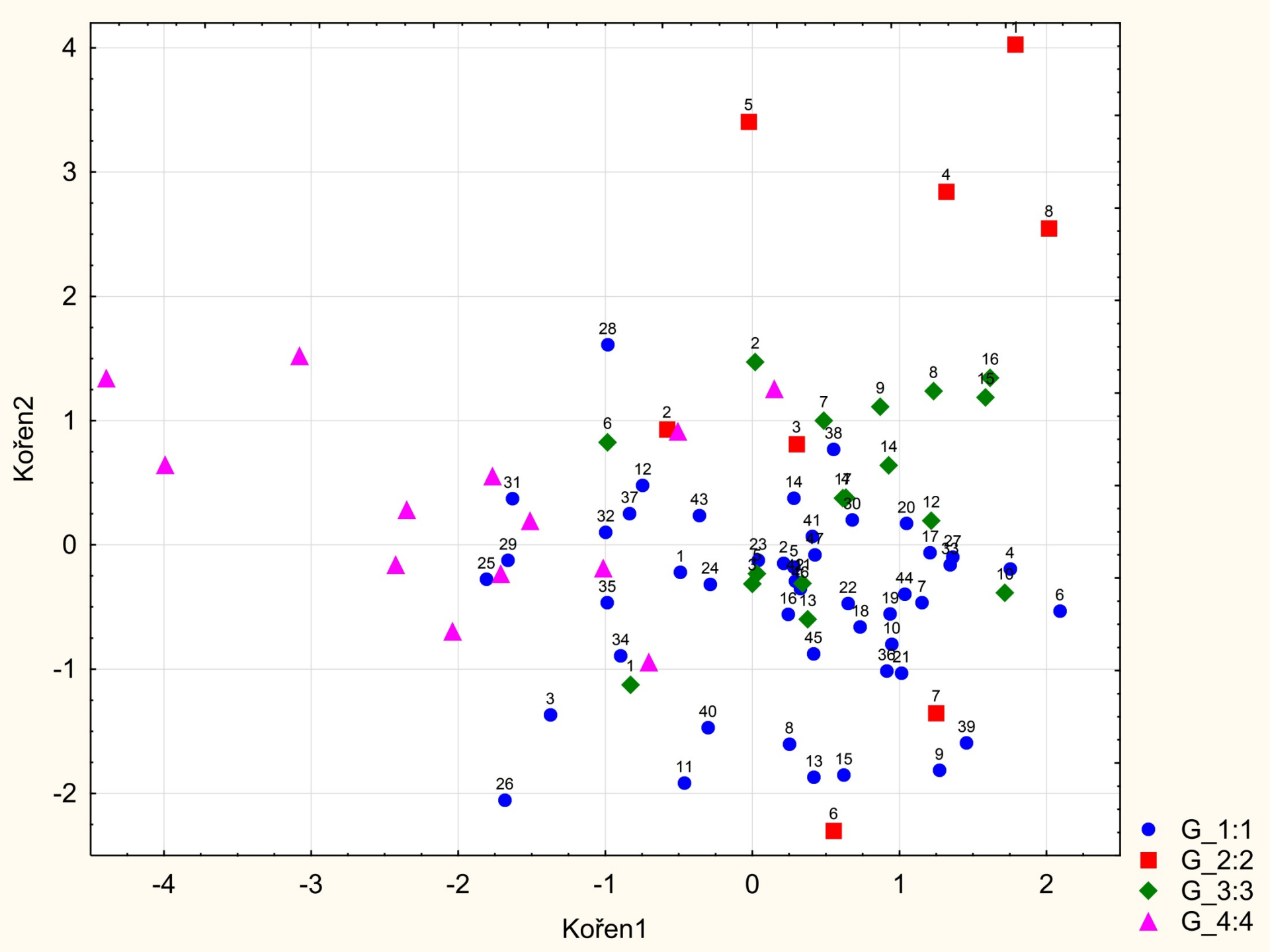


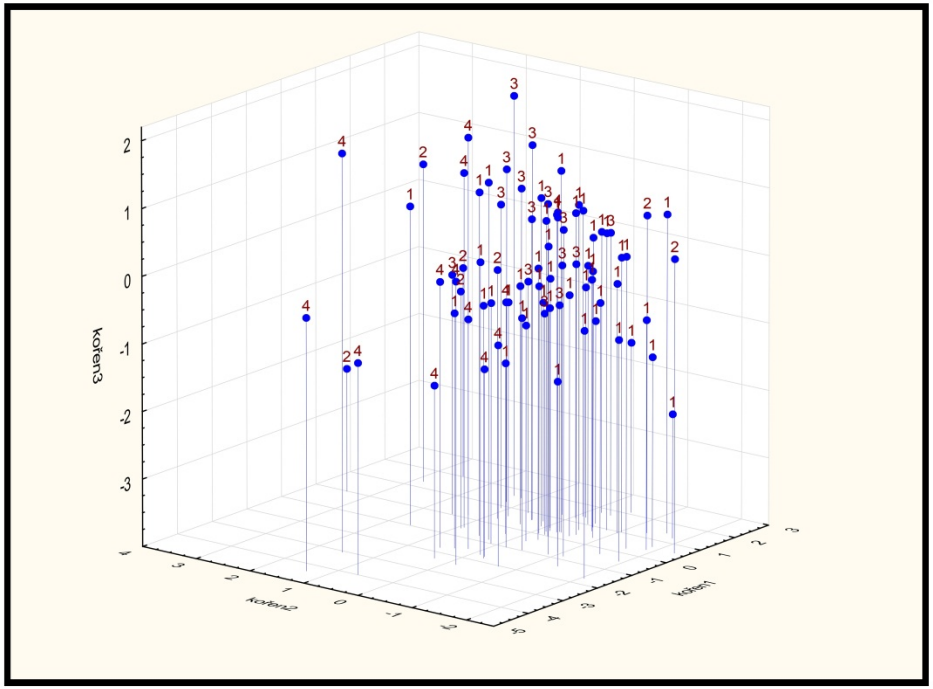
Skup.	Průměry kan. proměnných (B401.sta)	
	Kořen1	Kořen2
G_1:1	0.076601	-1.72620
G_2:2	-0.897345	0.24264
*	1.112423	0.28957

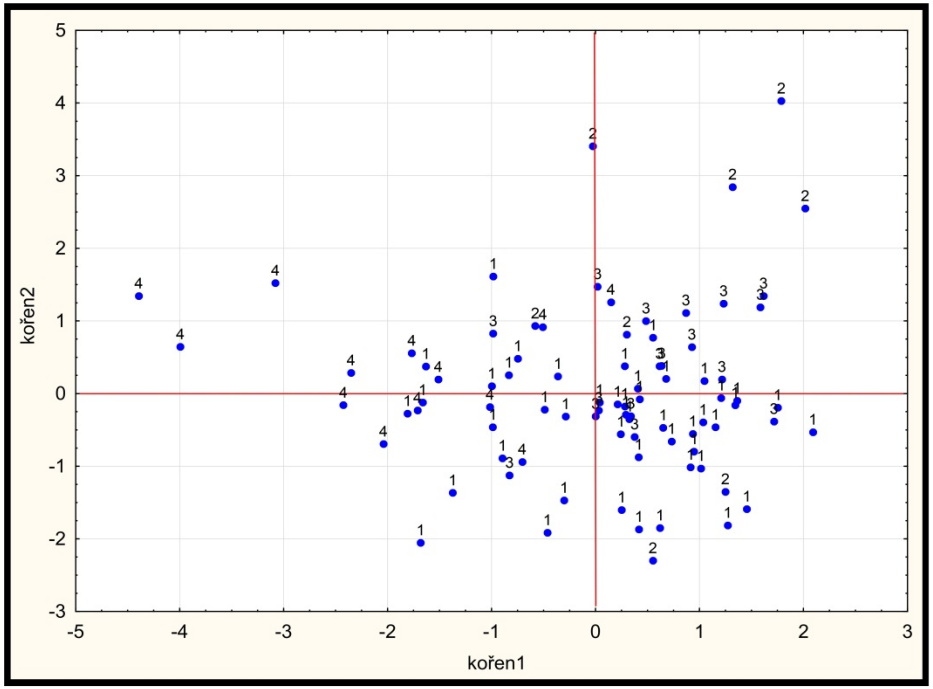


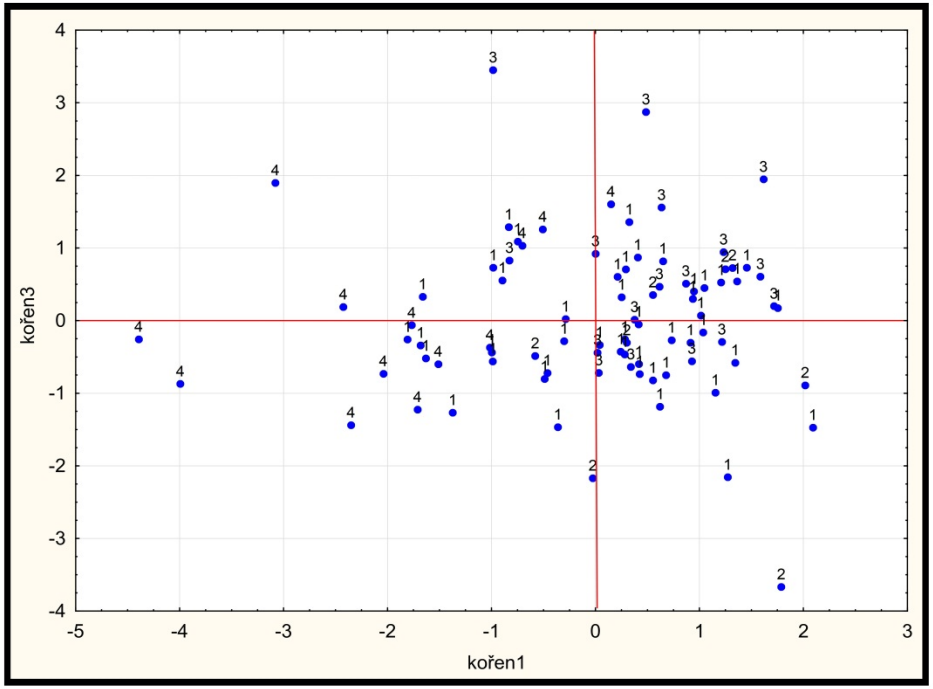
Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (B401.sta)					
	Vlastní číslo	Kan. R	Wilks. Lambda	Chi-kv.	sv	p-hodn.
0	0.904615	0.689173	0.355715	59.95020	30	0.000934
1	0.476013	0.567890	0.677501	22.58198	14	0.067421

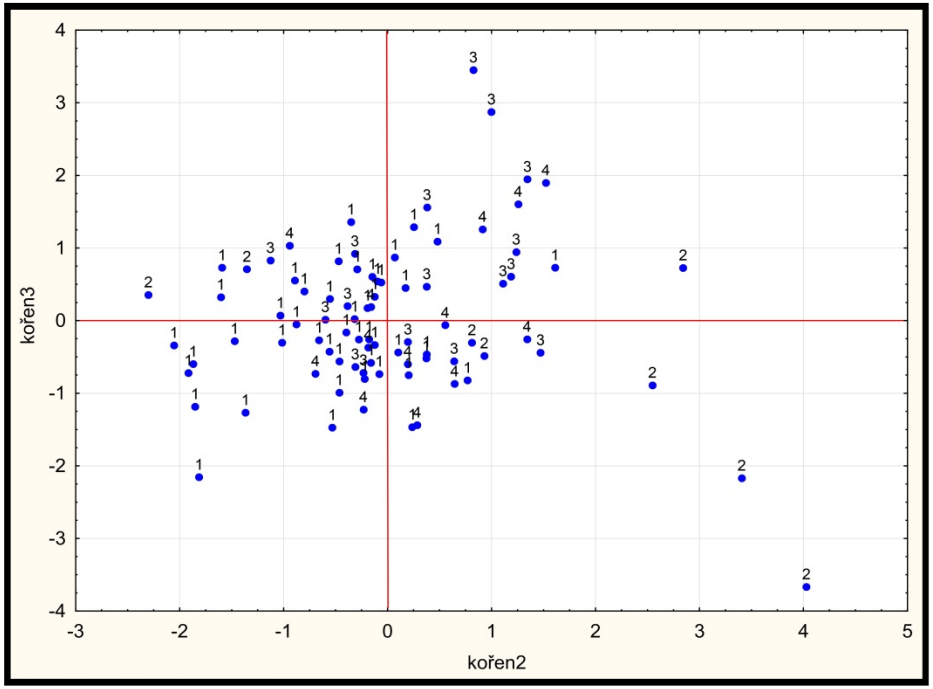
Případ	Nestandardizovaná kanonická sk		
	Skup.	Kořen1	Kořen2
1	G_2:2	-0.17774	-0.37886
2	*	0.47621	0.47074
3	G_2:2	-1.79542	1.84447
4	G_1:1	-0.36953	-3.87873
5	G_2:2	-1.28770	-0.31768
6	*	1.00999	-1.21573
7	G_2:2	-0.52083	1.10283
8	*	1.53731	0.36076
9	G_1:1	-0.08195	-3.54414
10	G_2:2	-1.14078	-0.41446
11	G_1:1	3.04259	-1.26613
12	G_1:1	-0.68331	-3.21713
13	G_1:1	0.07257	-1.38681
14	G_2:2	-0.98484	0.79145
15	G_2:2	0.02077	0.32253
16	G_2:2	-1.75972	-0.10411
17	G_2:2	-1.15030	0.81671
18	G_2:2	-1.15432	0.55227
19	*	1.95069	0.28499
20	G_2:2	-1.04715	0.32267
21	G_1:1	1.01711	-0.67960
22	G_2:2	-1.67959	-1.78078
23	*	1.89896	-0.11144
24	G_1:1	-0.84934	-0.47018
25	G_2:2	-0.82992	-0.04805
26	G_2:2	-0.06394	-0.49767
27	G_2:2	-0.85264	-2.40076
28	G_2:2	-1.11914	2.20831
29	G_2:2	-1.63526	-0.38159
30	*	3.60252	-1.71191
31	G_2:2	-0.15577	0.02375
32	G_1:1	-0.40876	-0.99193
33	*	1.65006	1.07805
34	*	0.43371	-0.68851
35	*	-0.51045	-0.35048
36	G_2:2	-2.52761	-0.69459
37	*	0.48002	1.35970
38	G_2:2	-2.77443	0.91539
39	G_2:2	-1.71526	-0.40234
40	G_2:2	-0.94963	1.33509


Kořen2
Kořen1
G_1:1
G_2:2
G_3:3
G_4:4


kořen3
kořen2
kořen1


kořen2
kořen1


kořen3
kořen1


kořen3
kořen2

Úloha B23 *Porovnání analýzy krve diabetiků a zdravých jedinců*

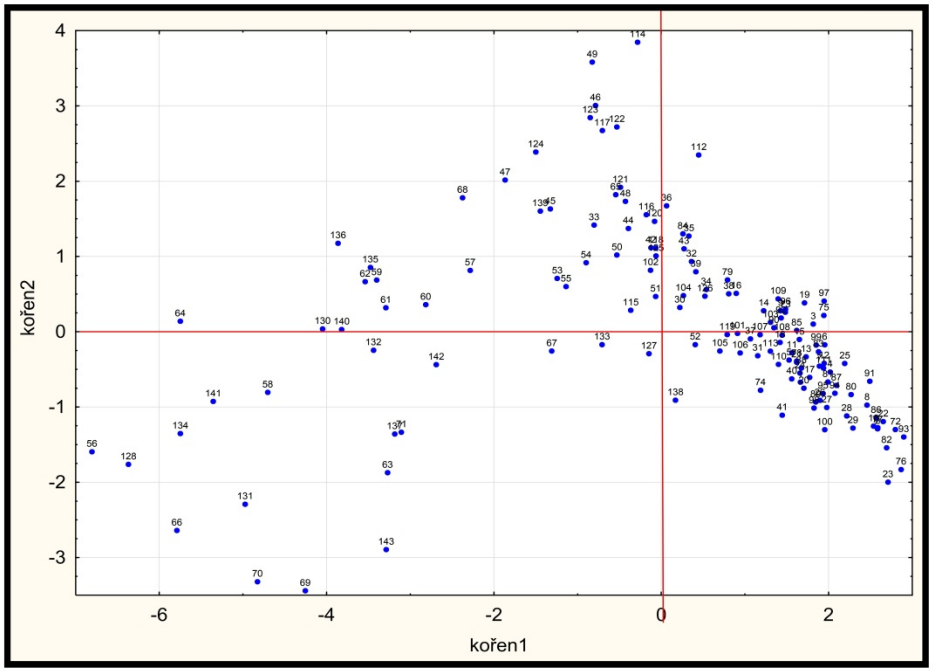
Jsou porovnány výsledky rozboru krve zdravých jedinců a dvou skupin diabetiků. Postavte soubor diskriminačních funkcí, které mohou předpovědět třídu a který rozlišuje mezi třídami podle hodnot ostatních kvantitativních proměnných.

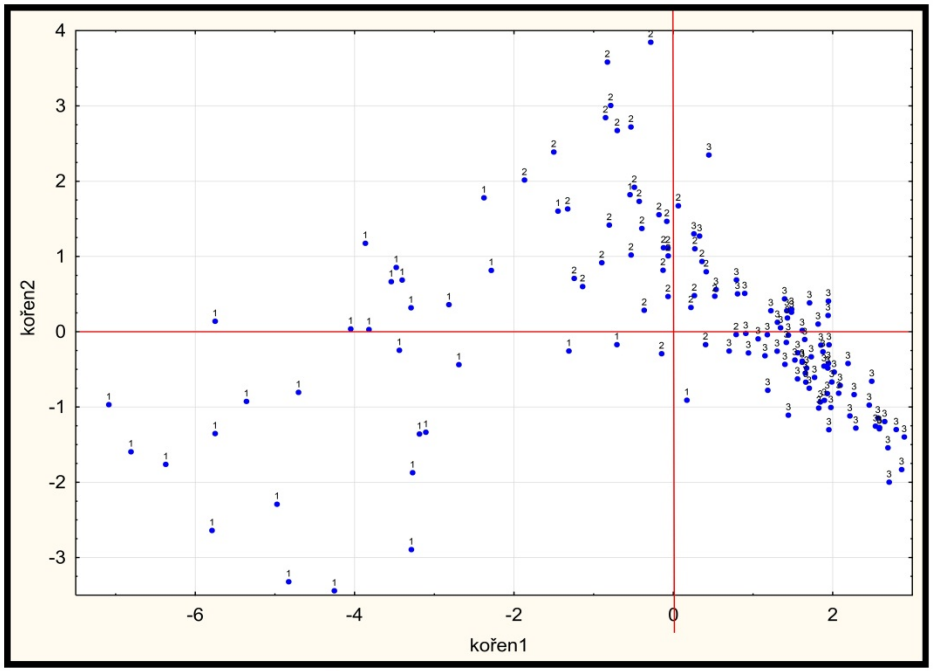
◯ ***Data:*** Výběr DIABETES se týká vyšetření 144 jedinců pro 6 znaků:

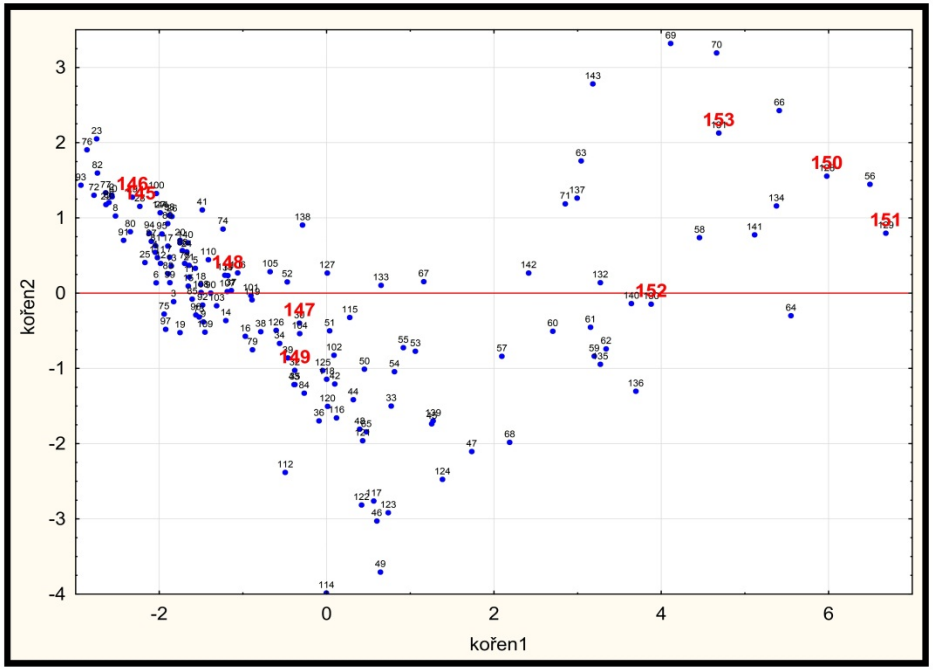
i (ID) identifikační číslo pacienta,
*x*1 (RELWT) značí relativní hmotnost pacienta,
*x*2 (GLUFAST) značí půst plasmové glukózy,
*x*3 (GLUTEST) značí prověrku plasmové glukózy,
*x*4 (INSTEST) značí plasmový inzulin v průběhu prověrky,
*x*5 (SSPG) značí ustálený rovnovážný stav plazmové glukózy,
*x*6 (GROUP) značí klinickou skupinu diabetes:

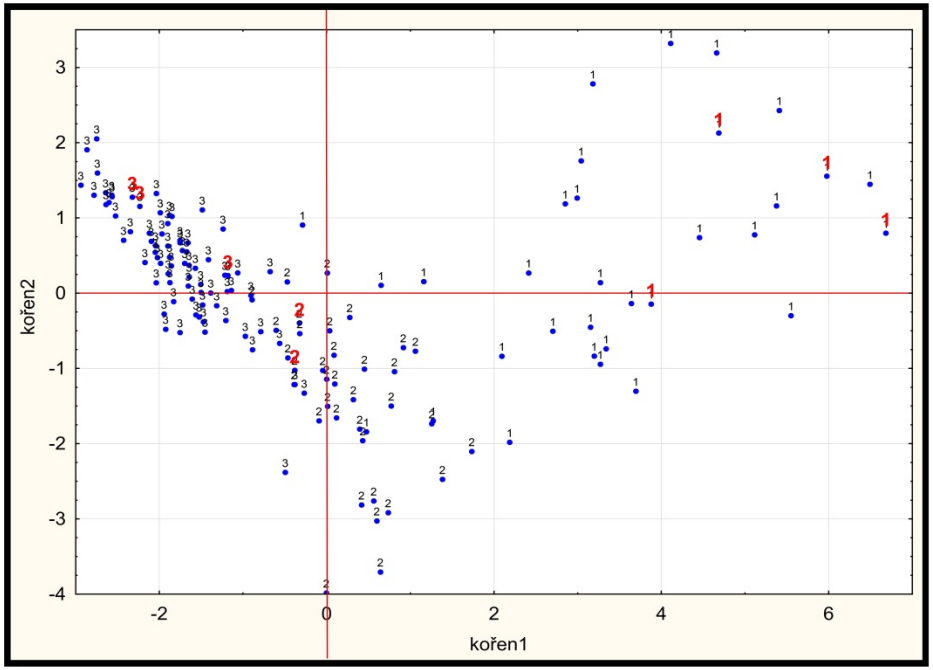
1 je zřetelný, vyložený diabetik,
2 je chemický,
3 je normální zdravý člověk.

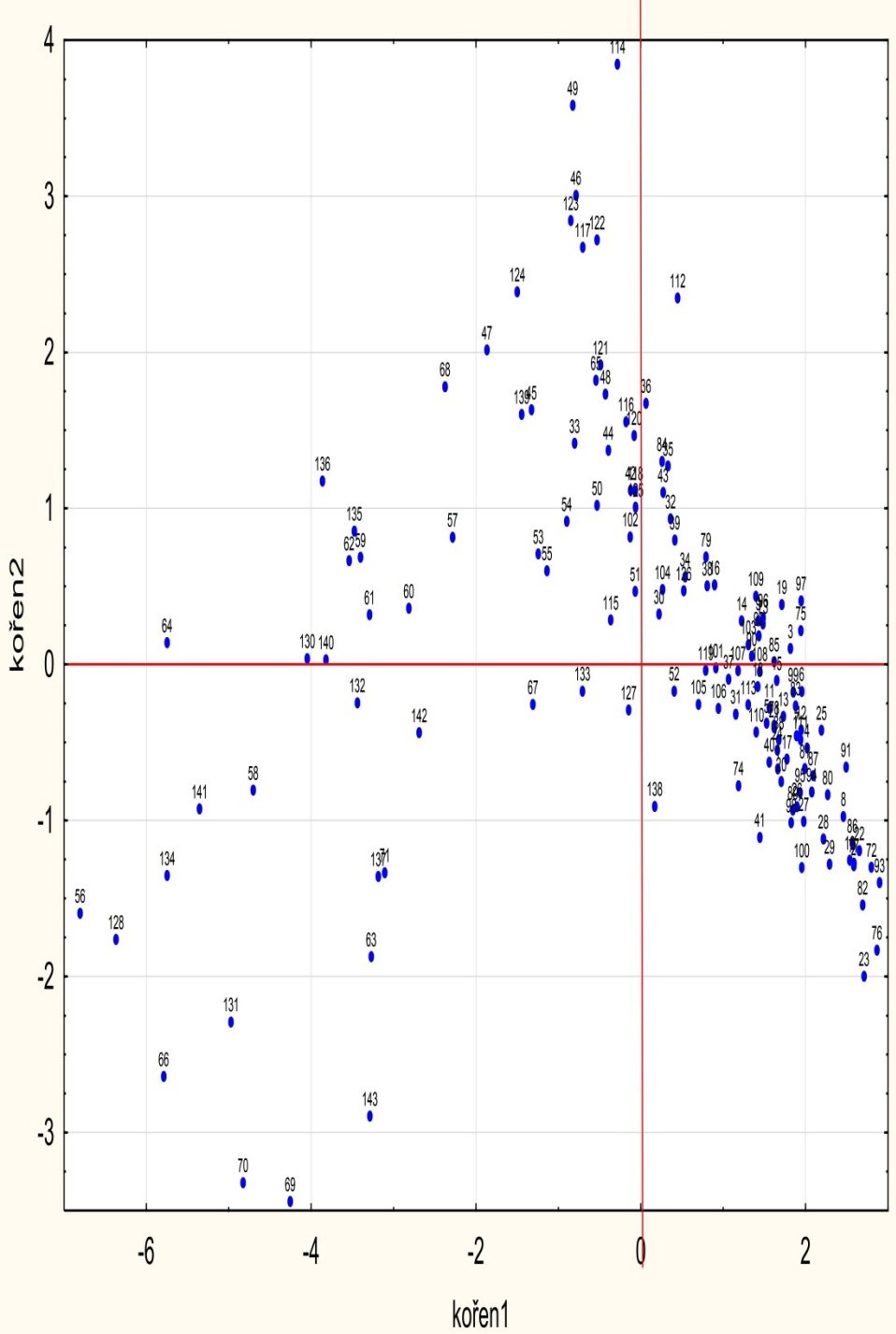
i	*x1*	*x2*	*x3*	*x4*	*x5*	*x6*
1	**0.81**	**80**	**356**	**124**	**55**	**3**
...	**...**	**...**	**...**	**...**	**...**	**...**
144	**1.11**	**328**	**1246**	**124**	**442**	**1**

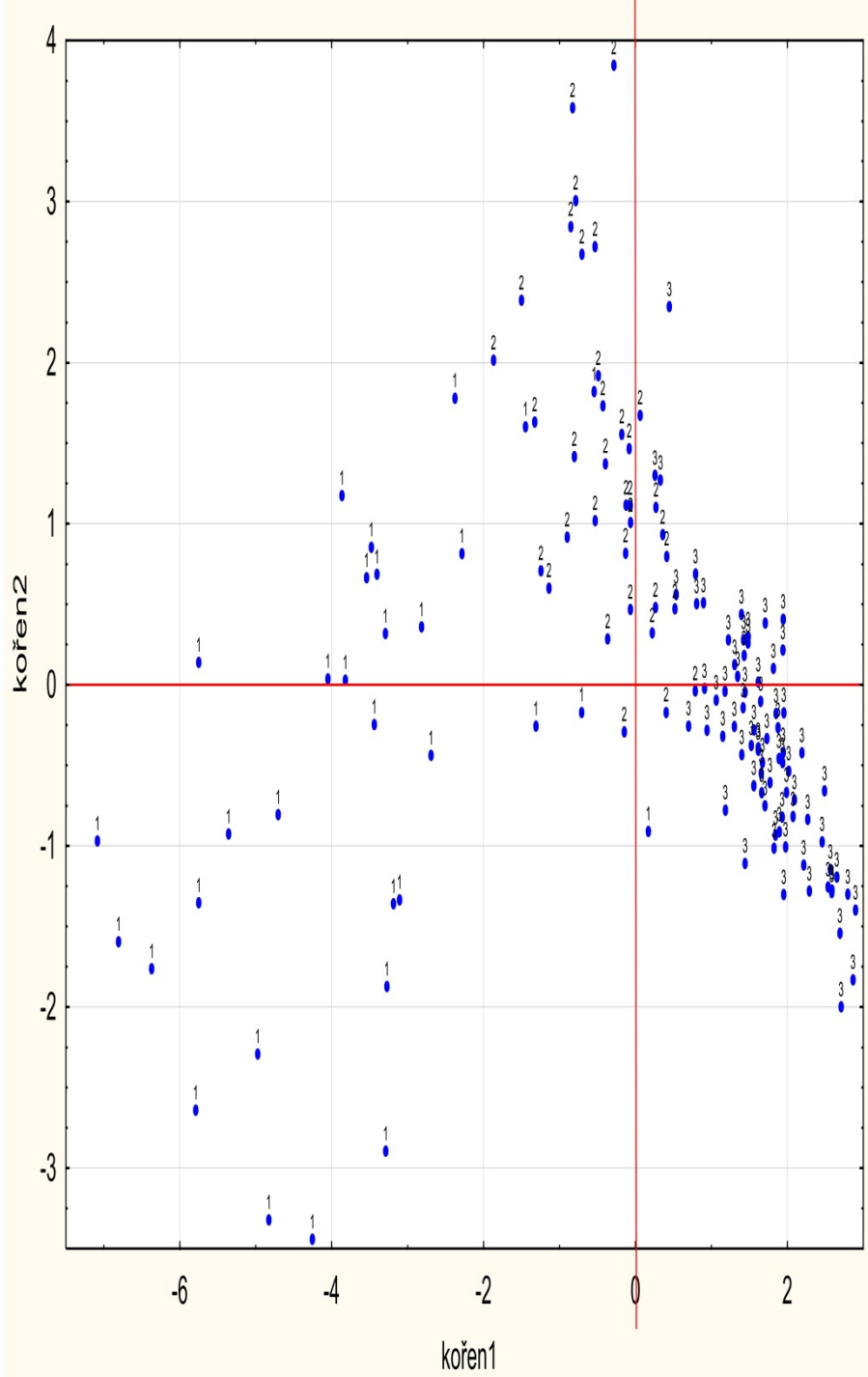

kořen2
kořen1

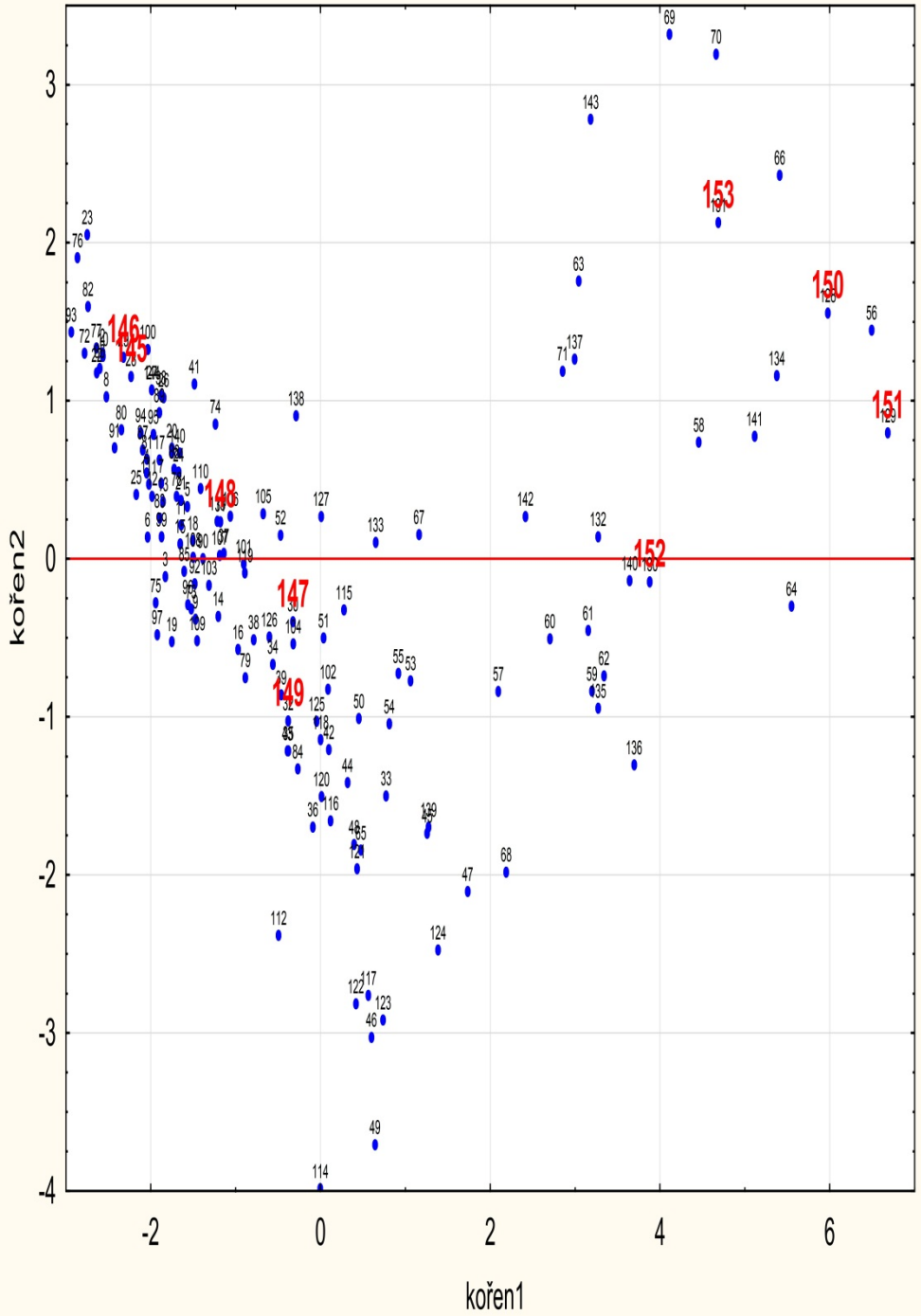

kořen2
kořen1

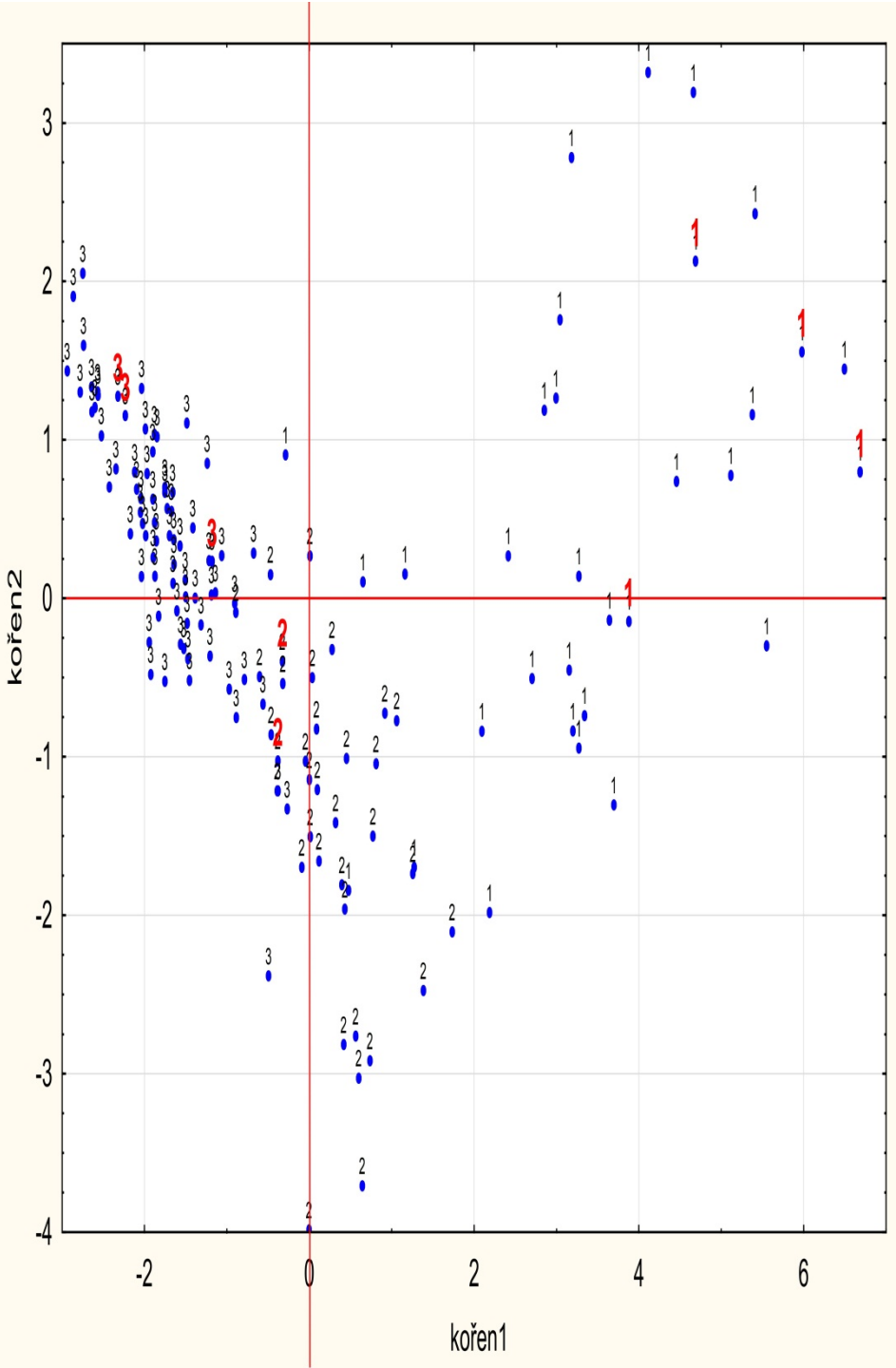

kořen2
kořen1


kořen2
kořen1


kořen2
kořen1


kořen2
kořen1


kořen2
kořen1
3
2
1
0
-1
-2
-3
-4
-2
0
2
4
6
146
145
148
147
149
152
153
150
151


kořen2
kořen1
3
2
1
0
-1
-2
-3
-4
-2
0
2
4
6

Úloha B40 ***Vliv vlákninové diety ve snídaňových lupínkách na nadváhu žen***

Ženy konzumují lupínky různých typů, např. otrubové lupínky, gumové, obojí a konečně i s kontrolním druhem lupínek bez vlákniny k snídani tolik, na kolik mají hlad. Kontigenční tabulka ukazuje vztah 4 různých druhů lupínek a 4 hladin nevolnosti z přesycení. χ^2-test může otestovat, zda je přesycení nezávislé na druhu lupínek.

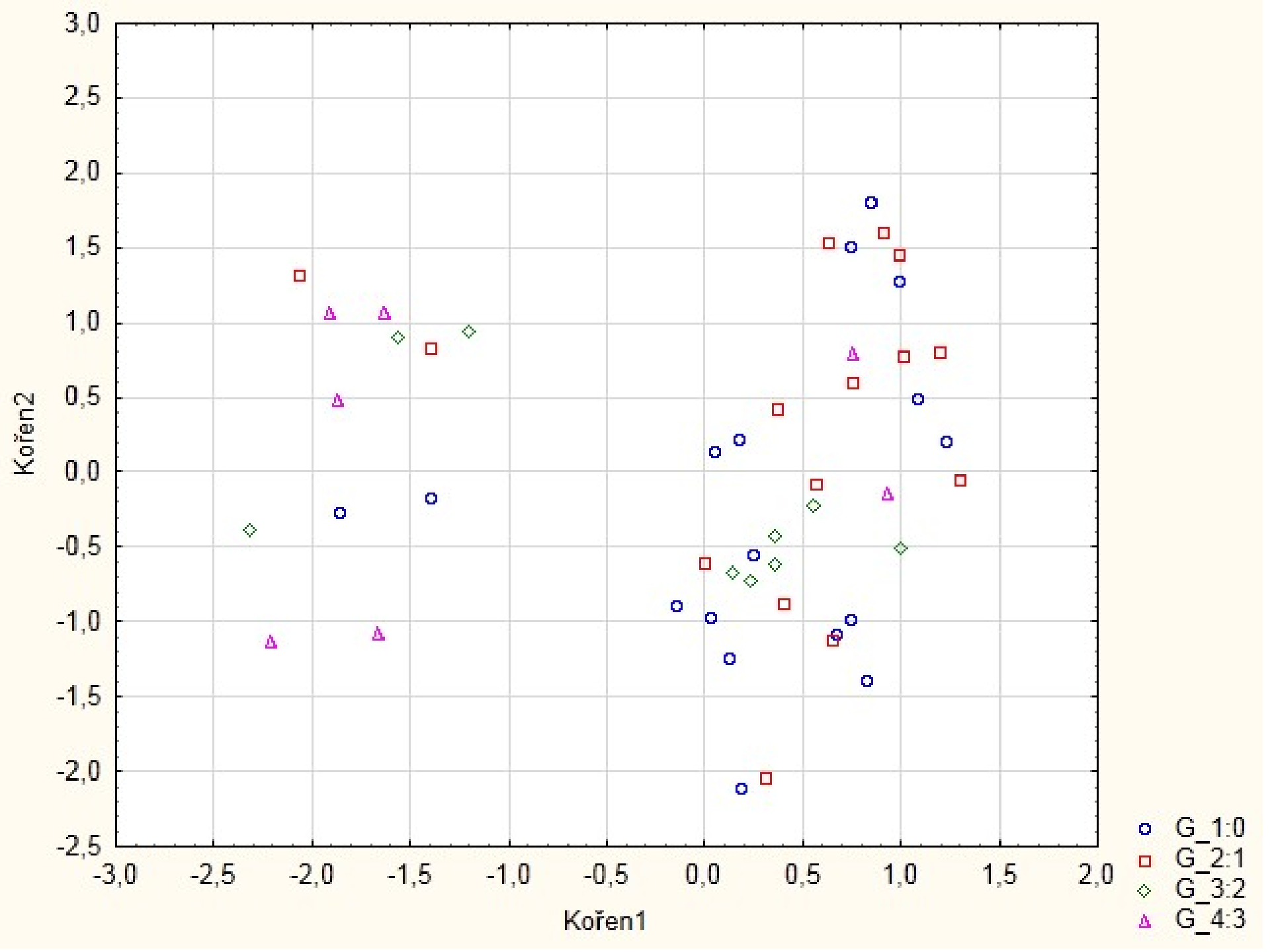
U diskriminační analýzy vezměte za závisle proměnnou údaje o stupni přesycení *x*5 (BLOAT) a ostatní znaky pak za nezávisle diskriminátory.

◯ ***Data:*** Celkem 12 žen v řádcích bylo testováno na 3 diety ve sloupcích a na kontrolní dietu. Před každým jídlem otylky jedly lupínky a) s otrubami, b) s gumovými vlákny, c) s kombinací obou, a konečně d) bez vláken. Celkem bylo provedeno 48 pokusů a užito 5 sledovaných znaků.

Datový výběr CRACKER obsahuje:

i (ID) pořadové číslo ženy,
*x*1 (CRACKER) značící typ vlákna v lupínkách,
*x*2 (DIET) značí jednu ze 4 diet čili typu lupínek,
*x*3 (IDFEM) index sledované ženy,
*x*4 (DIGESTED) značí strávené kalorie.
*x*5 (BLOAT) značí stupeň přesycení a nadýmání po jídle: 0 žádný, 1 malý, 2 střední a 3 vysoký.

i	*x1*	*x2*	*x3*	*x4*	*x5*
1	control	1	3	1772.84	0
	...	...	...	...	...
12	bran	4	12	2287.52	0


3,0
2,5
2,0
1,5
1,0
0,5
0,0
-0,5
-1,0
-1,5
-2,0
-2,5
Kořen2
-3,0
-2,5
-2,0
-1,5
-1,0
-0,5
0,0
0,5
1,0
1,5
2,0
Kořen1
G_1:0
G_2:1
G_3:2
G_4:3

Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (B40)					
	Vlastní číslo	Kan. R	Wilk. Lambda	Chi-kv.	sv	p-hodn.
0	0,279382	0,467303	0,724296	13,86989	12	0,309094
1	0,063846	0,244979	0,926651	3,27568	6	0,773535
2	0,014390	0,119106	0,985814	0,61437	2	0,735514

Proměnná	Klasifikační funkce; grupovací : BLOAT (B40)			
	G_1:0 p=,35417	G_2:1 p=,31250	G_3:2 p=,18750	G_4:3 p=,14583
CRACKER	112,23	112,32	111,81	110,86
DIET	51,62	51,67	51,23	50,78
IDFEM	-1,46	-1,49	-1,35	-1,36
DIGESTED	-0,00	-0,00	-0,00	-0,00
Konstant	-5824,98	-5829,56	-5781,35	-5682,67

Skup.	Průměry kan. proměnných (B40)		
	Kořen1	Kořen2	Kořen3
G_1:0	0,26422	-0,231565	-0,091501
G_2:1	0,37329	0,306967	0,024591
G_3:2	-0,27171	-0,190684	0,212517
G_4:3	-1,09226	0,149752	-0,103715

Úloha E23 *Klasifikace trubců a královen vos dle tělesných měr*

Jeanne, Graf a Yandell (1995) vyšetřovali vosy a proměřili 13 znaků u 50 trubců a 50 královen. Měření byla kódována dotyčnou částí těla, H značí hlavu, T značí trup či krunýř, W značí křídlo a G rodidla. U těchto částí těla byly proměřeny šířka w, výška h a délka l. Gonadium má uvedeny dvě délky G1l a G2l a tři šířky G1wa, G1wb a G2wa. Je třeba provést klasifikaci vos a rozhodnout zda se trubci liší svými rozměry těla od královen. Lze diskriminační analýzou rozlišit, zda se velikosti těl královen a trubců výrazně liší? Je klasifikace vos do dvou tříd provedená na základě diskriminačních parametrů shodná s původní klasifikací na královny a trubce?

◯ *Data:* Výběr WASP obsahuje 100 vos v řádcích dle 13 znaků ve sloupcích:
i (ID) pořadové číslo vosy,
$x1$ (CASTE) značící druh vosy: 0 značí královnu a 1 trubce, Q královny, W trubce,
$x2$ (l) značí délku těla,
$x3$ (Wl) značí délku křídla,
$x4$ (Hh) značí výšku hlavy,
$x5$ (Hw) značí šířku hlavy,
$x6$ (Th) značí výšku trupu,
$x7$ (Tw) značí šířku trupu,
$x8$ (G1l) značí délku gonadia 1,
$x9$ (G2wa) značí šířku gonadia 2a,
$x10$ (G2l) značí délku gonadia 2,
$x11$ (Hl) značí délku hlavy,
$x12$ (G1wb) značí šířku gonadia 1b,
$x13$ (G1wa) značí šířku gonadia 1a,
$x14$ (G1h) značí výšku gonadia 1.

i	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
1	0	5.420	9.09	2.39	2.64	2.36	2.25	3.4	3.21	3.380	0.387	0.13	0.39	0.235
		...	...	...	...	...	...	...	...	...	...	...	...	...
100	1	5.72	10.01	2.69	2.96	2.63	2.36	3.63	3.03	3.550	0.431	0.129	0.378	0.21

N=100	Výsledky diskriminační funkční analýzy (E23) Počet prom. v modelu: 13; grupovací: CASTE (2 skup) Wilk. lambda: ,07240 přibliž F (13,86)=84,757 p<0,0000					
	Wilk. Lambda	Parc. Lambda	F na vyj (1,86)	p-hodn.	Toler.	1-toler. R^2
L	0,072408	0,999899	0,00869	0,925930	0,526939	0,473061
WL	*0,078137*	*0,926581*	*6,81433*	*0,010667*	*0,401716*	*0,598284*
HH	0,074253	0,975051	2,20056	0,141616	0,522212	0,477788
HW	*0,080447*	*0,899977*	*9,55797*	*0,002683*	*0,292731*	*0,707269*
TH	0,072438	0,999484	0,04443	0,833550	0,517266	0,482734
TW	*0,093191*	*0,776898*	*24,69665*	*0,000003*	*0,312300*	*0,687700*
G1L	0,073680	0,982635	1,51977	0,221014	0,428199	0,571801
G2WA	0,072973	0,992152	0,68023	0,411789	0,696144	0,303856
G2L	0,072926	0,992788	0,62475	0,431463	0,419627	0,580372
HL	*0,078068*	*0,927394*	*6,73296*	*0,011126*	*0,637000*	*0,363000*
G1WB	0,072400	1,000000	0,00000	1,000000	0,629069	0,370931
G1WA	*0,092923*	*0,779142*	*24,37779*	*0,000004*	*0,649371*	*0,350629*
G1H	0,073605	0,983631	1,43116	0,234863	0,755097	0,244903

Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (E23)					
	Vlastní číslo	Kan. R	Wilk. Lambda	Chi-kv.	sv	p-hodn.
0	12,81212	0,963120	0,072400	240,2375	13	0,00

Proměnná	Klasifikační funkce; grupovací : CASTE (E23)	
	Q p=,50000	W p=,50000
L	112,18	111,45
WL	-43,51	-31,64
HH	911,36	945,68
HW	-251,60	-193,01
TH	347,55	350,84
TW	-162,64	-250,50
G1L	195,39	211,40
G2WA	21,25	17,30
G2L	-83,90	-78,32
HL	2972,11	3227,95
G1WB	965,45	965,78
G1WA	227,14	-88,93
G1H	791,86	683,60
Konstant	-2150,22	-2334,19

Skup.	Klasifikační matice (E23) Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace		
	% správnýc	Q p=,50000	W p=,50000
Q	100,0000	50	0
W	100,0000	0	50
Celkem	100,0000	50	50

Případ	Klasifikace případů (E23) Nesprávná klasifikace je označena *		
	Pozorova Klasif.	1 p=,50000	2 p=,50000
1	Q	Q	W
2	Q	Q	W
3	Q	Q	W
4	Q	Q	W
5	Q	Q	W
6	Q	Q	W
7	Q	Q	W
8	Q	Q	W
9	Q	Q	W
10	Q	Q	W
11	Q	Q	W
12	Q	Q	W
13	Q	Q	W
14	Q	Q	W
15	Q	Q	W
16	Q	Q	W
17	Q	Q	W
18	Q	Q	W
19	Q	Q	W
20	Q	Q	W
21	Q	Q	W
22	Q	Q	W
23	Q	Q	W
24	Q	Q	W
25	Q	Q	W
26	Q	Q	W
27	Q	Q	W
28	Q	Q	W
29	Q	Q	W
30	Q	Q	W
31	Q	Q	W
32	Q	Q	W
33	Q	Q	W
34	Q	Q	W
35	Q	Q	W
36	Q	Q	W
37	Q	Q	W
38	Q	Q	W
39	Q	Q	W
40	Q	Q	W
41	Q	Q	W
42	Q	Q	W
43	Q	Q	W
44	Q	Q	W
45	Q	Q	W
46	Q	Q	W
47	Q	Q	W
48	Q	Q	W
49	Q	Q	W
50	Q	Q	W

Případ	Klasifikace případů (E23) Nesprávná klasifikace je označena *		
	Pozorova Klasif.	1 p=,50000	2 p=,50000
51	W	W	Q
52	W	W	Q
53	W	W	Q
54	W	W	Q
55	W	W	Q
56	W	W	Q
57	W	W	Q
58	W	W	Q
59	W	W	Q
60	W	W	Q
61	W	W	Q
62	W	W	Q
63	W	W	Q
64	W	W	Q
65	W	W	Q
66	W	W	Q
67	W	W	Q
68	W	W	Q
69	W	W	Q
70	W	W	Q
71	W	W	Q
72	W	W	Q
73	W	W	Q
74	W	W	Q
75	W	W	Q
76	W	W	Q
77	W	W	Q
78	W	W	Q
79	W	W	Q
80	W	W	Q
81	W	W	Q
82	W	W	Q
83	W	W	Q
84	W	W	Q
85	W	W	Q
86	W	W	Q
87	W	W	Q
88	W	W	Q
89	W	W	Q
90	W	W	Q
91	W	W	Q
92	W	W	Q
93	W	W	Q
94	W	W	Q
95	W	W	Q
96	W	W	Q
97	W	W	Q
98	W	W	Q
99	W	W	Q
100	W	W	Q

Případ	Mahalanobisovy vzdálenosti (E23) Nesprávná klasifikace je označena *		
	Pozorova Klasif.	Q p=,50000	W p=,50000
1	Q	20	64
2	Q	10	71
3	Q	21	79
4	Q	11	36
5	Q	14	54
6	Q	8	66
7	Q	17	49
8	Q	6	42
9	Q	7	70
10	Q	9	70
11	Q	21	42
12	Q	5	68
13	Q	5	54
14	Q	10	57
15	Q	7	54
16	Q	18	49
17	Q	17	85
18	Q	12	71
19	Q	12	65
20	Q	13	61
21	Q	11	42
22	Q	8	62
23	Q	17	93
24	Q	9	55
25	Q	15	69
26	Q	8	48
27	Q	9	57
28	Q	6	67
29	Q	8	55
30	Q	7	60
31	Q	18	64
32	Q	7	63
33	Q	9	81
34	Q	9	44
35	Q	20	90
36	Q	15	80
37	Q	18	67
38	Q	16	93
39	Q	34	59
40	Q	6	66
41	Q	18	68
42	Q	16	65
43	Q	25	94
44	Q	6	45
45	Q	12	68
46	Q	17	56
47	Q	6	44
48	Q	14	53
49	Q	9	71
50	Q	10	52

Případ	Mahalanobisovy vzdálenosti (E23) Nesprávná klasifikace je označena *		
	Pozorova Klasif.	Q p=,50000	W p=,50000
51	W	91	17
52	W	36	22
53	W	53	12
54	W	64	13
55	W	52	8
56	W	48	17
57	W	53	8
58	W	43	12
59	W	52	7
60	W	39	17
61	W	73	12
62	W	54	12
63	W	72	15
64	W	53	12
65	W	82	17
66	W	58	13
67	W	72	14
68	W	62	13
69	W	85	15
70	W	63	12
71	W	69	10
72	W	55	7
73	W	54	13
74	W	34	17
75	W	62	10
76	W	69	13
77	W	63	11
78	W	52	10
79	W	53	8
80	W	76	7
81	W	99	16
82	W	71	11
83	W	66	8
84	W	72	11
85	W	108	29
86	W	54	19
87	W	60	8
88	W	46	14
89	W	79	13
90	W	67	16
91	W	74	8
92	W	54	11
93	W	59	12
94	W	86	13
95	W	53	10
96	W	56	13
97	W	56	12
98	W	72	17
99	W	55	18
100	W	75	11

Případ	Nestandardizovaná kanonická skóre (E23)	
	Skup.	Kořen1
1	Q	3,130
2	Q	4,287
3	Q	4,107
4	Q	1,783
5	Q	2,814
6	Q	4,066
7	Q	2,285
8	Q	2,530
9	Q	4,425
10	Q	4,268
11	Q	1,482
12	Q	4,419
13	Q	3,490
14	Q	3,331
15	Q	3,302
16	Q	2,134
17	Q	4,816
18	Q	4,136
19	Q	3,715
20	Q	3,434
21	Q	2,181
22	Q	3,843
23	Q	5,385
24	Q	3,213
25	Q	3,820
26	Q	2,807
27	Q	3,364
28	Q	4,327
29	Q	3,315
30	Q	3,758
31	Q	3,274
32	Q	3,962
33	Q	5,114
34	Q	2,485
35	Q	4,941
36	Q	4,616
37	Q	3,400
38	Q	5,395
39	Q	1,754
40	Q	4,196
41	Q	3,575
42	Q	3,432
43	Q	4,823
44	Q	2,753
45	Q	4,007
46	Q	2,789
47	Q	2,656
48	Q	2,743
49	Q	4,339
50	Q	2,950
51	W	-5,257
52	W	-0,987

Případ	Nestandardizovaná kanonická skóre (E23)	
	Skup.	Kořen1
53	W	-2,918
54	W	-3,567
55	W	-3,113
56	W	-2,159
57	W	-3,173
58	W	-2,222
59	W	-3,172
60	W	-1,574
61	W	-4,318
62	W	-3,001
63	W	-4,029
64	W	-2,883
65	W	-4,622
66	W	-3,153
67	W	-4,098
68	W	-3,470
69	W	-4,974
70	W	-3,627
71	W	-4,151
72	W	-3,416
73	W	-2,888
74	W	-1,161
75	W	-3,683
76	W	-3,929
77	W	-3,728
78	W	-2,930
79	W	-3,165
80	W	-4,873
81	W	-5,872
82	W	-4,189
83	W	-4,143
84	W	-4,266
85	W	-5,560
86	W	-2,458
87	W	-3,665
88	W	-2,225
89	W	-4,638
90	W	-3,644
91	W	-4,602
92	W	-3,010
93	W	-3,326
94	W	-5,191
95	W	-3,039
96	W	-3,033
97	W	-3,080
98	W	-3,844
99	W	-2,649
100	W	-4,501
101	----	1893,274

Skup.	Průměry kan. proměnných (E23)
	Kořen1
Q	3,54343
W	-3,54343

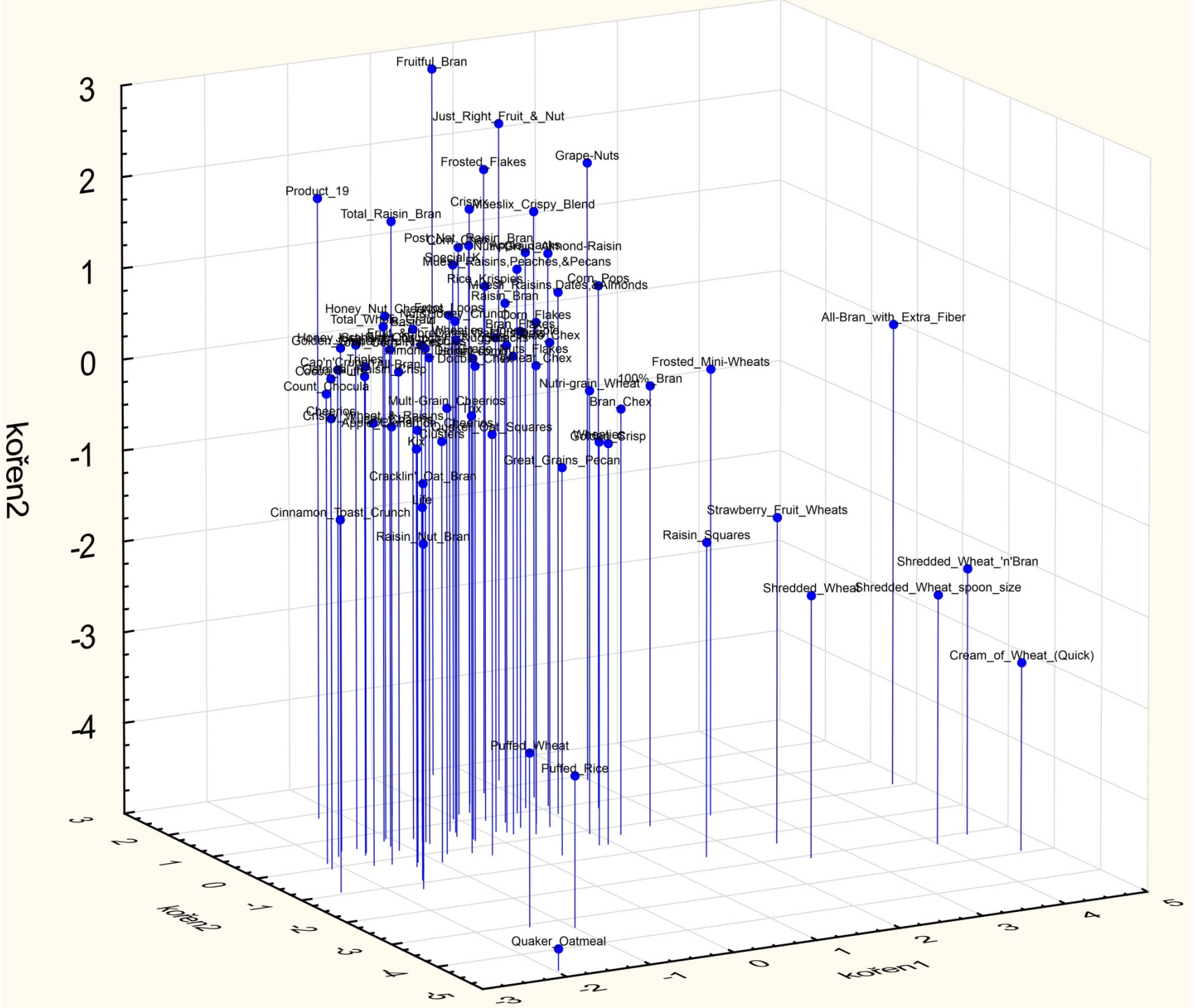
Úloha E22 *Zdravotní klasifikace 77 druhů obilninových lupínek k snídani*

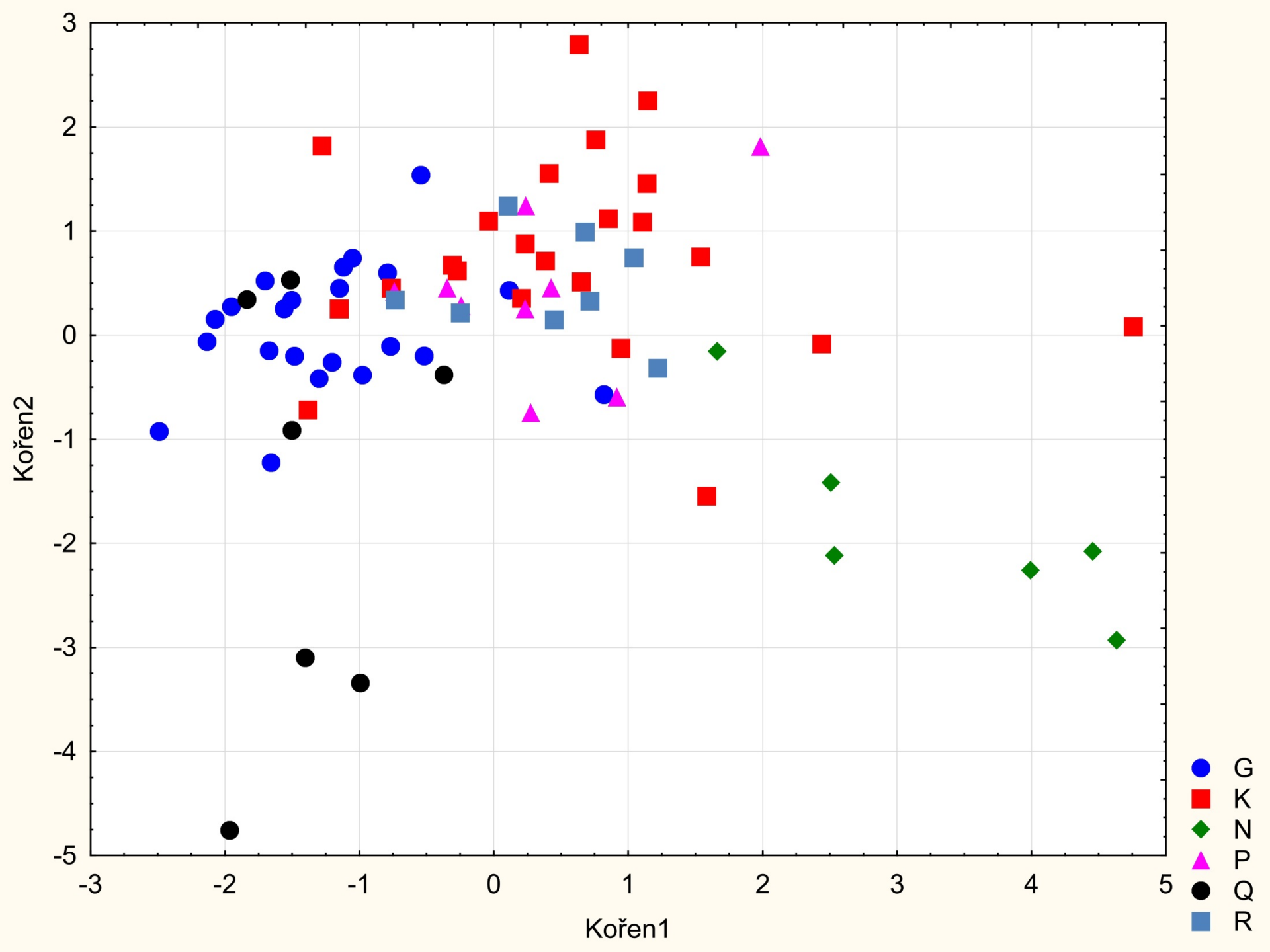
Zdravotní norma u lupínek požaduje, že dospělí by měli denně zkonzumovat ne více než 30% kalorií ve formě tuku, muži potřebují okolo 63 g bílkovin a ženy 50 g bílkovin a zbytek poskytnou uhlovodíky (1 cal = 4.184 joulu). 1 g tuku obsahuje asi 9 kalorií, bílkoviny a uhlovodíky 4 kalorie na 1 g. Vhodná dieta by měla obsahovat 20 - 35 g dietní vlákniny. Níže analyzovaná data obsahují počet kalorií na 1 porci, gramy bílkovin a tuků, mg Na^+ a K^+iontů, gramy vlákniny, uhlovodíků a cukru, typické procento FDA RDA vitaminů, hmotnost jedné porce a počet šálků na 1 porci podávanou. Data obsahují umístění na regále v prodejně: nahoře, ve středu či dole u země.

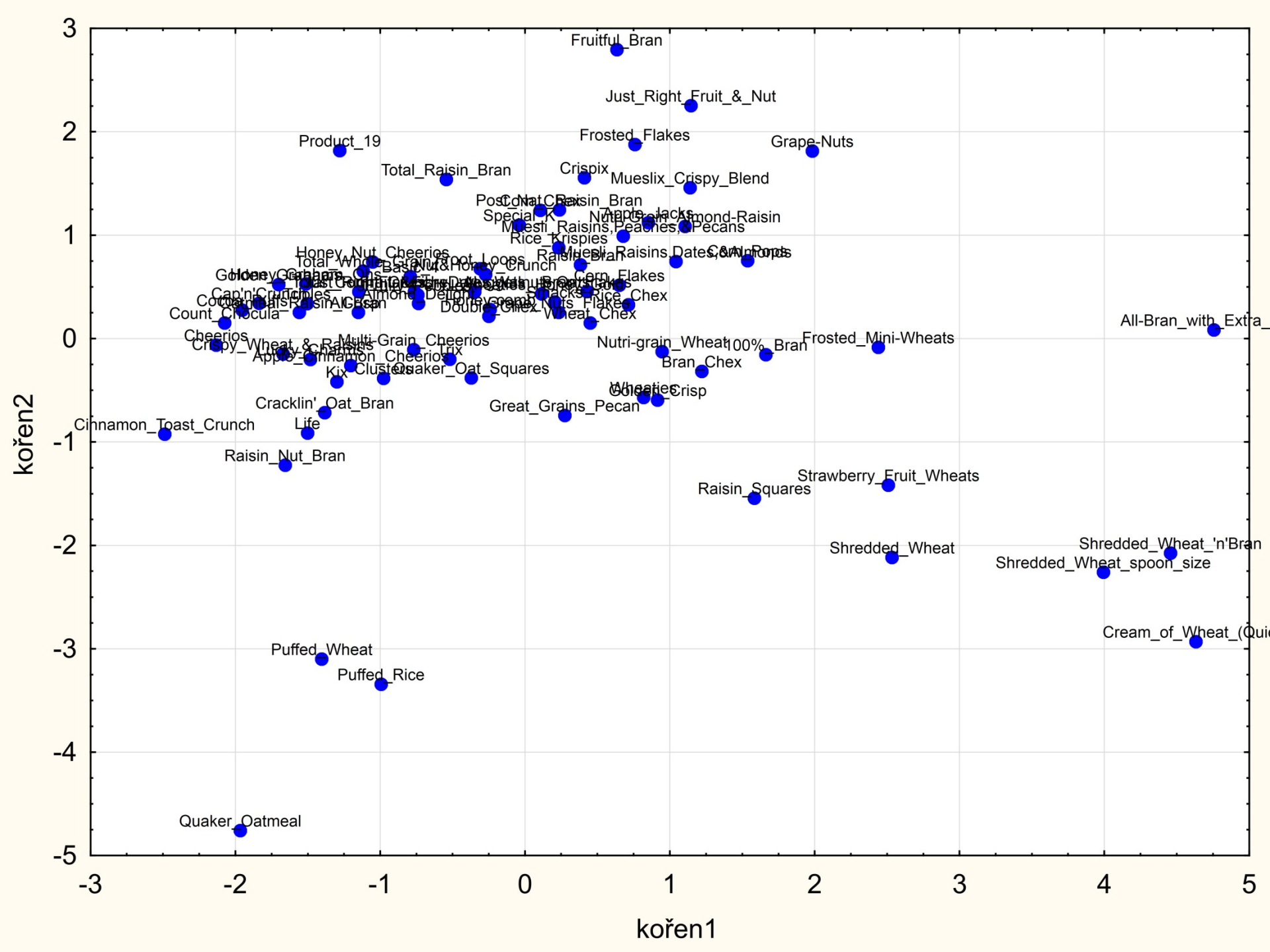
◯ ***Data:*** Výběr CEREALS obsahuje 77 druhů obilninových lupínek v řádcích a 15 znaků:

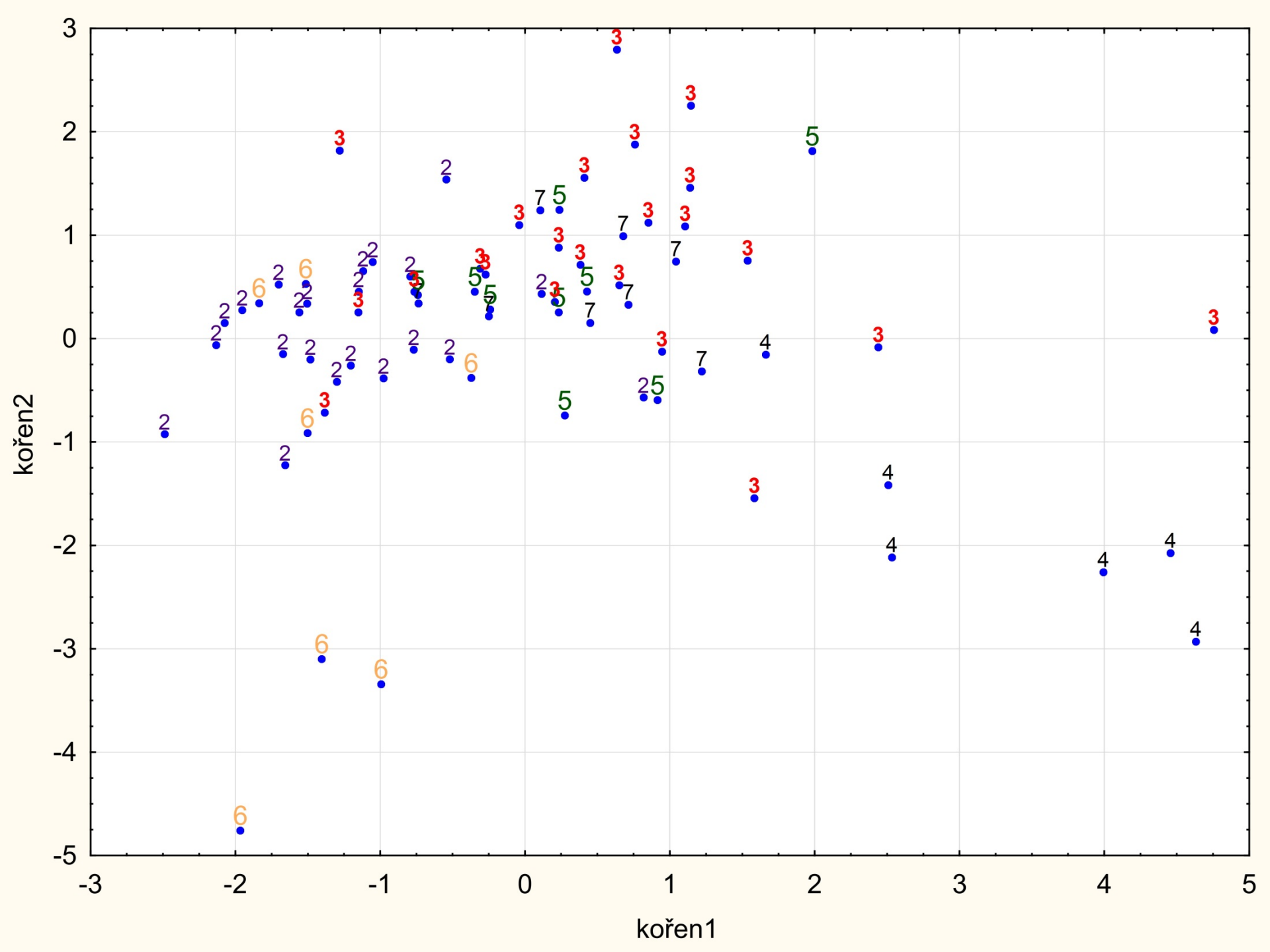
i (ID) je index druhu lupínek, *j* (J) je název druhu obilninových lupínek,
*x*1 (MFR) značí výrobce lupínek: A je American Home Food Products; G je General Mills; K je Kelloggs; N je Nabisco; P je Post; Q je Quaker Oats; R je Ralston Purina,
*x*2 (TYPE) značí formu v jaké je lupínka konzumována: C je studená, H je horká,
*x*3 (CALORIES) značí počet kalorií jedné porce,
*x*4 (PROTEIN) značí obsah bílkovin v gramech v jedné porci,
*x*5 (FAT) značí obsah tuku v gramech v jedné porci,
*x*6 (SODIUM) značí obsah sodných iontů v miligramech v jedné porci,
*x*7 (FIBER) značí obsah dietní vlákniny v jedné porci,
*x*8 (CH) značí obsah komplexních uhlovodíků v gramech,
*x*9 (SUGARS) značí obsah cukrů v gramech v jedné porci,
*x*10 (POTASS) značí obsah draselných iontů v miligramech v jedné porci,
*x*11 (VITAMINS) značí obsah vitaminů a minerálů v jedné porci (0, 25, nebo 100) ukazuje typické % doporučených FDA,
*x*12 (SHELF) značí umístění na regále, jehož výška je číslovaná směrem od země 1 (nejníže), 2, nebo 3 (nejvýše),
*x*13 (WEIGHT) značí hmotnost jedné porce v uncích,
*x*14 (CUPS) značí počet šálků v jedné porci,
*x*15 (RATING) značí hodnocení, ocenění druhu lupínek.
Číslo -1 značí chybějící numerickou informaci.

i	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
100%_Bran	N	C	70	4	1	130	10	5	6	280	25	3	1	0.33	68.40
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Wheaties_Honey_Gold	G	C	110	2	1	200	1	16	8	60	25	1	1	0.75	36.19


kořen2
kořen2
kořen1
Fruitful_Bran
Just_Right_Fruit_&_Nut
Frosted_Flakes
Grape-Nuts
Product_19
Total_Raisin_Bran
Mueslix_Crispy_Blend
Raisins,Peaches,&Pecans
Corn_Pops
Raisin_Bran
Honey_Nut_Cheerios
Corn_Flakes
All-Bran_with_Extra_Fiber
Frosted_Mini-Wheats
100%_Bran
Nutri-grain_Wheat
Count_Chocula
Mult-Grain_Cheerios
Bran_Chex
Cheerios
Oat_Squares
Kix
Clusters
Great_Grains_Pecan
Cracklin'_Oat_Bran
Life
Cinnamon_Toast_Crunch
Raisin_Nut_Bran
Strawberry_Fruit_Wheats
Raisin_Squares
Shredded_Wheat_'n'Bran
Shredded_Wheat_spoon_size
Cream_of_Wheat_(Quick)
Puffed_Wheat
Puffed_Rice
Quaker_Oatmeal


Kořen2
Kořen1
3
2
1
0
-1
-2
-3
-4
-5
-3
-2
-1
0
1
2
3
4
5
G
K
N
P
Q
R


kořen2
kořen1
3
2
1
0
-1
-2
-3
-4
-5
-3
-2
-1
0
1
2
3
4
5
Fruitful_Bran
Just_Right_Fruit_&_Nut
Product_19
Frosted_Flakes
Grape-Nuts
Total_Raisin_Bran
Crispix
Mueslix_Crispy_Blend
Raisin_Bran
Special_K
Apple_Jacks
Almond-Raisin
Muesli_Raisins,Peaches,&Pecans
Rice_Krispies
Honey_Nut_Cheerios
Muesli_Raisins,Dates,&Almonds
Corn_Flakes
Chex
Wheat_Chex
Count_Chocula
Cheerios
Multi-Grain_Cheerios
Trix
Kix
Clusters
Quaker_Oat_Squares
Nutri-grain_Wheat
100%_Bran
Frosted_Mini-Wheats
All-Bran_with_Extra_
Bran_Chex
Wheaties
Golden_Crisp
Great_Grains_Pecan
Cracklin'_Oat_Bran
Cinnamon_Toast_Crunch
Life
Raisin_Nut_Bran
Strawberry_Fruit_Wheats
Raisin_Squares
Shredded_Wheat
Shredded_Wheat_'n'Bran
Shredded_Wheat_spoon_size
Cream_of_Wheat_(Qui
Puffed_Wheat
Puffed_Rice
Quaker_Oatmeal


kořen2
kořen1
3
2
1
0
-1
-2
-3
-4
-5
-3
-2
-1
0
1
2
3
4
5

Kořeny odstraněny	Test chí-kvadrát po odstranění post. kořenů (E22)					
	Vlastní číslo	Kan. R	Wilk. Lambda	Chi-kv.	sv	p-hodn.
0	1,824023	0,803676	0,100675	149,2308	70	0,000000
1	0,971859	0,702043	0,284308	81,7502	52	0,005270
2	0,428362	0,547629	0,560616	37,6167	36	0,395068
3	0,199115	0,407495	0,800763	14,4424	22	0,885038
4	0,041442	0,199482	0,960207	2,6394	10	0,988685

Proměnná	Klasifikační funkce; grupovací : MFR (E22)				
	G_1:2 p=,28947	G_2:3 p=,30263	G_3:4 p=,07895	G_4:5 p=,11842	G_5:6 p=,10526
TYPE	431	429	443	429	431
CALORIES	1444741	1444724	1444645	1444730	1444725
PROTEIN	-21234981	-21234726	-21233572	-21234813	-21234741
FAT	10976930	10976795	10976200	10976840	10976805
SODIUM	353534	353530	353511	353531	353530
FIBER	-22336642	-22336371	-22335156	-22336463	-22336389
CH	-7087424	-7087338	-7086952	-7087368	-7087344
SUGARS	4703311	4703254	4702999	4703273	4703257
POTASS	220361	220359	220347	220360	220359
VITAMINS	332383	332379	332361	332380	332379
SHELF	1966	1966	1963	1966	1967
WEIGHT	41729	41722	41718	41721	41730
CUPS	-5056	-5058	-5059	-5059	-5054
RATING	6487518	6487440	6487088	6487467	6487445
Konstant	-178184354	-178180076	-178160726	-178181531	-178180305

Skup.	Klasifikační matice (E22) Řádky: pozorované klasifikace Sloupce: předpovězené klasifikace						
	% správnýc	G_1:2 p=,28947	G_2:3 p=,30263	G_3:4 p=,07895	G_4:5 p=,11842	G_5:6 p=,10526	G_6:7 p=,10526
G_1:2	90,90909	20	0	0	0	0	2
G_2:3	65,21739	5	15	2	0	0	1
G_3:4	83,33334	0	1	5	0	0	0
G_4:5	11,11111	1	7	0	1	0	0
G_5:6	50,00000	3	1	0	0	4	0
G_6:7	75,00000	0	2	0	0	0	6
Celkem	67,10526	29	26	7	1	4	9

Skup.	Průměry kan. proměnných (E22)				
	Kořen1	Kořen2	Kořen3	Kořen4	Kořen5
G_1:2	-1,21444	0,06672	0,347723	0,471965	-0,043966
G_2:3	0,59012	0,77854	-0,481083	-0,000789	0,178348
G_3:4	3,29766	-1,82526	0,310731	0,465967	-0,040646
G_4:5	0,30440	0,39909	-0,600635	-0,318510	-0,468234
G_5:6	-1,57511	-1,96403	-0,538354	-0,532021	0,129005
G_6:7	0,40252	0,46219	1,407895	-0,754767	0,036400

	E22			
	1 MFR	2 TŘÍDA_1	3 TŘÍDA_2	4 TŘÍDA_3
100%_Bran	4	3,00	5,00	4,00
100%_Natural_Bran	6	6,00	2,00	5,00
All-Bran	3	2,00	5,00	3,00
All-Bran_with_Extra_Fiber	3	4,00	3,00	5,00
Almond_Delight	7	3,00	2,00	5,00
Apple_Cinnamon_Cheerios	2	2,00	3,00	5,00
Apple_Jacks	3	3,00	5,00	2,00
Basic_4	2	2,00	3,00	5,00
Bran_Chex	7	7,00	3,00	2,00
Bran_Flakes	5	3,00	5,00	2,00
Cap'n'Crunch	6	2,00	5,00	3,00
Cheerios	2	2,00	7,00	6,00
Cinnamon_Toast_Crunch	2	2,00	6,00	7,00
Clusters	2	2,00	5,00	3,00
Cocoa_Puffs	2	2,00	5,00	3,00
Corn_Chex	7	7,00	3,00	2,00
Corn_Flakes	3	3,00	7,00	2,00
Corn_Pops	3	3,00	5,00	7,00
Count_Chocula	2	2,00	5,00	6,00
Cracklin'_Oat_Bran	3	2,00	6,00	5,00
Cream_of_Wheat_(Quick)	4	4,00	7,00	3,00
Crispix	3	3,00	5,00	7,00
Crispy_Wheat_&_Raisins	2	2,00	5,00	6,00
Double_Chex	7	3,00	2,00	5,00
Froot_Loops	3	3,00	5,00	2,00
Frosted_Flakes	3	3,00	5,00	7,00
Frosted_Mini-Wheats	3	3,00	5,00	4,00
Fruit_&FibreDates,Walnuts,Oats	5	3,00	2,00	5,00
Fruitful_Bran	3	3,00	5,00	2,00
Fruity_Pebbles	5	2,00	3,00	5,00
Golden_Crisp	5	3,00	5,00	2,00
Golden_Grahams	2	2,00	3,00	5,00
Grape_Nuts_Flakes	5	3,00	5,00	2,00
Grape-Nuts	5	3,00	5,00	7,00
Great_Grains_Pecan	5	5,00	3,00	7,00
Honey_Graham_Ohs	6	2,00	3,00	5,00
Honey_Nut_Cheerios	2	2,00	3,00	5,00
Honey-comb	5	3,00	2,00	5,00
Just_Right_Crunchy__Nuggets	3	2,00	3,00	5,00
Just_Right_Fruit_&_Nut	3	3,00	5,00	2,00
Kix	2	2,00	7,00	3,00
Life	6	2,00	6,00	3,00
Lucky_Charms	2	2,00	3,00	5,00
Muesli_Raisins,Dates,&Almonds	7	7,00	3,00	5,00
Muesli_Raisins,Peaches,&Pecans	7	7,00	3,00	5,00
Mueslix_Crispy_Blend	3	3,00	5,00	2,00
Multi-Grain_Cheerios	2	2,00	7,00	3,00
Nut&Honey_Crunch	3	3,00	5,00	2,00
Nutri-Grain_Almond-Raisin	3	3,00	7,00	5,00
Nutri-grain_Wheat	3	3,00	5,00	2,00
Oatmeal_Raisin_Crisp	2	2,00	5,00	3,00

Případ	Mahalanobisovy vzdálenosti (E22) Nesprávná klasifikace je označena *						
	Pozorova Klasif.	G_1:2 p=,28947	G_2:3 p=,30263	G_3:4 p=,07895	G_4:5 p=,11842	G_5:6 p=,10526	G_6:7 p=,10526
* 1	G_3:4	23,67547	14,49502	20,66662	14,43710	26,03329	21,96582
2	G_5:6	35,58430	52,52211	60,02314	47,30113	22,74284	48,62678
* 3	G_2:3	20,09037	24,10233	44,16545	21,87566	26,03981	24,13942
* 4	G_2:3	64,71687	43,41764	34,68192	47,00818	69,83513	52,78951
* 5	G_6:7	16,37051	15,06332	36,89524	15,92531	19,00193	20,61881
6	G_1:2	5,40308	11,03883	28,03071	9,65832	10,45411	10,06924
7	G_2:3	18,30098	9,03338	27,40379	9,86129	23,79709	18,42899
8	G_1:2	13,09016	14,57751	35,17149	15,03295	21,39900	20,63593
9	G_6:7	14,66622	14,75315	15,31579	15,55116	25,55887	11,23554
* 10	G_4:5	14,35637	9,81977	24,87019	9,15198	19,08937	13,82327
* 11	G_5:6	5,77882	11,82327	36,36992	9,61232	10,91140	10,60018
12	G_1:2	24,79615	33,45683	56,60040	36,13863	30,39525	30,12784
13	G_1:2	9,71267	21,90228	41,46841	18,96708	11,93702	17,22026
14	G_1:2	4,74111	7,59010	24,79315	5,45201	6,22078	8,01250
15	G_1:2	4,80237	11,57031	36,17770	9,24290	10,40923	12,33160
16	G_6:7	10,25421	9,84974	26,81234	11,33692	22,92979	6,35456
17	G_2:3	10,28487	8,80283	19,14694	11,27683	21,22288	8,25804
18	G_2:3	18,32846	7,81883	19,88651	8,21970	23,43952	15,61031
19	G_1:2	5,17311	12,97497	37,19010	10,30457	10,68096	13,02045
* 20	G_2:3	6,98859	12,35021	29,48117	9,40485	7,00034	9,69923
21	G_3:4	90,23396	80,53103	50,19286	83,51861	91,23771	75,24791
22	G_2:3	17,94987	10,90472	32,77249	12,20890	25,31488	13,60959
23	G_1:2	5,36220	10,15614	32,53907	7,16568	7,45248	13,26080
* 24	G_6:7	7,42208	6,92455	23,17559	5,95053	12,56218	9,81476
25	G_2:3	10,77894	7,98326	28,74222	8,14334	15,63217	13,82776
26	G_2:3	17,05134	10,51666	30,00130	10,68734	28,90620	12,53016
27	G_2:3	23,80770	12,31225	14,08917	12,59490	27,51070	18,39841
* 28	G_4:5	7,20303	6,97584	24,72184	5,59774	13,67794	10,56400
29	G_2:3	22,40768	13,06559	39,84023	13,57404	36,42472	22,61537
* 30	G_4:5	7,32444	8,85568	28,28620	7,12187	13,49087	10,89511
* 31	G_4:5	20,09370	17,97444	22,26156	17,65175	26,12609	24,65039
32	G_1:2	7,28401	13,36476	37,36532	11,76675	14,64743	12,63442
* 33	G_4:5	10,49788	6,02179	21,97656	6,82477	13,45312	10,94296
* 34	G_4:5	34,15321	19,29721	35,69373	19,24667	40,61960	23,86510
35	G_4:5	19,50293	16,94977	26,87641	14,78474	17,74596	16,17608
* 36	G_5:6	8,11504	11,67789	36,48456	10,40755	12,75189	10,85011
37	G_1:2	7,73618	10,52831	32,76487	10,08227	15,74143	10,77352
* 38	G_4:5	11,30825	11,13246	27,23473	11,12059	17,71410	14,37552
* 39	G_2:3	12,74213	14,74201	34,06980	16,75882	20,81820	17,94605
40	G_2:3	21,54010	13,81091	32,41000	17,42981	38,00913	20,24961
41	G_1:2	12,74540	19,30053	35,68530	20,60431	17,99603	16,96744
* 42	G_5:6	8,55634	14,95569	31,35201	14,33574	10,02466	17,33527
43	G_1:2	5,88188	10,75389	31,17042	9,53385	9,43523	14,11615
44	G_6:7	40,02333	35,88518	46,30773	34,70576	46,62595	22,47350
45	G_6:7	38,17924	35,05224	48,67453	34,15369	45,61559	21,47292
46	G_2:3	21,55923	13,77907	29,44186	13,48039	32,36572	20,63890
47	G_1:2	4,82811	9,96248	24,15476	9,83688	12,67367	7,65936
48	G_2:3	8,59048	7,36834	26,08059	6,23183	14,31444	8,15105
49	G_2:3	14,28566	9,52377	21,17522	10,67775	26,24445	10,08102
50	G_2:3	13,52704	7,44929	17,12940	9,42917	16,85745	14,39585

Případ	Mahalanobisovy vzdálenosti (E22) Nesprávná klasifikace je označena *						
	Pozorova Klasif.	G_1:2 p=,28947	G_2:3 p=,30263	G_3:4 p=,07895	G_4:5 p=,11842	G_5:6 p=,10526	G_6:7 p=,10526
51	G_1:2	7,69126	11,30153	35,35495	8,88399	11,81928	17,17745
* 52	G_4:5	13,33706	10,48618	28,58726	9,18266	24,47307	15,82537
* 53	G_2:3	16,30254	17,54034	47,41578	22,04887	29,26045	23,02103
54	G_5:6	26,24629	32,35653	35,22444	28,35830	14,29940	31,38314
55	G_5:6	27,25809	32,35383	40,78497	28,15808	13,14488	33,82783
* 56	G_5:6	10,87573	10,90455	25,43018	10,05180	12,93504	15,39709
57	G_5:6	76,12154	83,30858	88,32105	77,87903	53,10953	91,04832
* 58	G_2:3	11,84151	12,40198	23,87890	11,67770	25,31079	13,90317
59	G_1:2	7,42175	14,85682	30,43906	11,88633	7,01816	16,54562
* 60	G_2:3	20,72703	16,23555	13,23060	14,08997	21,01864	22,77087
61	G_6:7	13,34493	12,90696	20,98839	13,81160	23,91018	8,45653
* 62	G_2:3	8,85954	8,96365	22,91102	10,40044	20,91880	5,31691
63	G_3:4	27,92197	21,77083	9,78657	20,80016	26,45654	20,54344
64	G_3:4	42,06213	30,82609	6,77227	31,14136	45,27572	30,58792
65	G_3:4	39,02975	30,40701	7,23471	30,54216	42,31019	29,93259
66	G_2:3	12,64321	8,42111	24,72960	7,55464	16,54875	15,35314
67	G_2:3	31,28401	25,43623	48,37295	28,80089	36,64025	32,96488
68	G_3:4	22,07296	13,34621	6,70308	14,13385	23,09563	20,96351
69	G_1:2	12,97295	17,90410	37,81858	20,21426	22,71355	18,37724
70	G_1:2	24,07286	25,29600	47,48998	26,79062	40,25020	34,96743
71	G_1:2	13,44690	15,84134	38,70824	18,42268	21,38998	19,52563
72	G_1:2	8,84013	14,86837	36,47711	13,53853	15,58159	11,14789
73	G_1:2	4,74256	6,11306	21,34700	5,22357	8,84380	9,61106
74	G_6:7	8,95875	9,87137	18,31320	10,47662	18,91396	5,54275
* 75	G_1:2	8,74328	8,75884	12,00980	10,29119	15,94339	6,40153
* 76	G_1:2	6,60407	7,88046	20,01625	7,72927	16,78503	4,44612

Případ	Klasifikace případů (E22) Nesprávná klasifikace je označena *						
	Pozorova Klasif.	1 p=,28947	2 p=,30263	3 p=,07895	4 p=,11842	5 p=,10526	6 p=,10526
* 1	G_3:4	G_2:3	G_4:5	G_3:4	G_1:2	G_6:7	G_5:6
2	G_5:6	G_5:6	G_1:2	G_4:5	G_6:7	G_2:3	G_3:4
* 3	G_2:3	G_1:2	G_4:5	G_2:3	G_6:7	G_5:6	G_3:4
* 4	G_2:3	G_3:4	G_2:3	G_4:5	G_6:7	G_1:2	G_5:6
* 5	G_6:7	G_2:3	G_1:2	G_4:5	G_5:6	G_6:7	G_3:4
6	G_1:2	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
7	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
8	G_1:2	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
9	G_6:7	G_6:7	G_2:3	G_1:2	G_4:5	G_3:4	G_5:6
* 10	G_4:5	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
* 11	G_5:6	G_1:2	G_4:5	G_2:3	G_6:7	G_5:6	G_3:4
12	G_1:2	G_1:2	G_6:7	G_5:6	G_2:3	G_4:5	G_3:4
13	G_1:2	G_1:2	G_5:6	G_6:7	G_4:5	G_2:3	G_3:4
14	G_1:2	G_1:2	G_4:5	G_2:3	G_5:6	G_6:7	G_3:4
15	G_1:2	G_1:2	G_4:5	G_2:3	G_5:6	G_6:7	G_3:4
16	G_6:7	G_6:7	G_2:3	G_1:2	G_4:5	G_5:6	G_3:4
17	G_2:3	G_2:3	G_6:7	G_1:2	G_4:5	G_3:4	G_5:6
18	G_2:3	G_2:3	G_4:5	G_6:7	G_1:2	G_3:4	G_5:6
19	G_1:2	G_1:2	G_4:5	G_5:6	G_2:3	G_6:7	G_3:4
* 20	G_2:3	G_1:2	G_5:6	G_4:5	G_6:7	G_2:3	G_3:4
21	G_3:4	G_3:4	G_6:7	G_2:3	G_4:5	G_1:2	G_5:6
22	G_2:3	G_2:3	G_4:5	G_6:7	G_1:2	G_5:6	G_3:4
23	G_1:2	G_1:2	G_4:5	G_5:6	G_2:3	G_6:7	G_3:4
* 24	G_6:7	G_2:3	G_1:2	G_4:5	G_6:7	G_5:6	G_3:4
25	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
26	G_2:3	G_2:3	G_4:5	G_6:7	G_1:2	G_5:6	G_3:4
27	G_2:3	G_2:3	G_4:5	G_3:4	G_6:7	G_1:2	G_5:6
* 28	G_4:5	G_2:3	G_1:2	G_4:5	G_6:7	G_5:6	G_3:4
29	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
* 30	G_4:5	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
* 31	G_4:5	G_2:3	G_4:5	G_1:2	G_3:4	G_6:7	G_5:6
32	G_1:2	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
* 33	G_4:5	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
* 34	G_4:5	G_2:3	G_4:5	G_6:7	G_1:2	G_3:4	G_5:6
35	G_4:5	G_4:5	G_2:3	G_6:7	G_1:2	G_5:6	G_3:4
* 36	G_5:6	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
37	G_1:2	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
* 38	G_4:5	G_2:3	G_1:2	G_4:5	G_6:7	G_5:6	G_3:4
* 39	G_2:3	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
40	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_3:4	G_5:6
41	G_1:2	G_1:2	G_6:7	G_2:3	G_5:6	G_4:5	G_3:4
* 42	G_5:6	G_1:2	G_5:6	G_2:3	G_4:5	G_6:7	G_3:4
43	G_1:2	G_1:2	G_2:3	G_4:5	G_5:6	G_6:7	G_3:4
44	G_6:7	G_6:7	G_2:3	G_4:5	G_1:2	G_5:6	G_3:4
45	G_6:7	G_6:7	G_2:3	G_4:5	G_1:2	G_5:6	G_3:4
46	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_3:4	G_5:6
47	G_1:2	G_1:2	G_6:7	G_2:3	G_4:5	G_5:6	G_3:4
48	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
49	G_2:3	G_2:3	G_6:7	G_4:5	G_1:2	G_3:4	G_5:6
50	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4

Případ	Klasifikace případů (E22) Nesprávná klasifikace je označena *						
	Pozorova Klasif.	1 p=,28947	2 p=,30263	3 p=,07895	4 p=,11842	5 p=,10526	6 p=,10526
51	G_1:2	G_1:2	G_4:5	G_2:3	G_5:6	G_6:7	G_3:4
* 52	G_4:5	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
* 53	G_2:3	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
54	G_5:6	G_5:6	G_1:2	G_4:5	G_2:3	G_6:7	G_3:4
55	G_5:6	G_5:6	G_1:2	G_4:5	G_2:3	G_6:7	G_3:4
* 56	G_5:6	G_2:3	G_1:2	G_4:5	G_5:6	G_6:7	G_3:4
57	G_5:6	G_5:6	G_1:2	G_4:5	G_2:3	G_3:4	G_6:7
* 58	G_2:3	G_1:2	G_2:3	G_4:5	G_6:7	G_3:4	G_5:6
59	G_1:2	G_1:2	G_5:6	G_4:5	G_2:3	G_6:7	G_3:4
* 60	G_2:3	G_3:4	G_4:5	G_2:3	G_1:2	G_5:6	G_6:7
61	G_6:7	G_6:7	G_2:3	G_1:2	G_4:5	G_3:4	G_5:6
* 62	G_2:3	G_6:7	G_1:2	G_2:3	G_4:5	G_5:6	G_3:4
63	G_3:4	G_3:4	G_2:3	G_6:7	G_4:5	G_1:2	G_5:6
64	G_3:4	G_3:4	G_2:3	G_6:7	G_4:5	G_1:2	G_5:6
65	G_3:4	G_3:4	G_2:3	G_6:7	G_4:5	G_1:2	G_5:6
66	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
67	G_2:3	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6	G_3:4
68	G_3:4	G_3:4	G_2:3	G_4:5	G_1:2	G_6:7	G_5:6
69	G_1:2	G_1:2	G_2:3	G_6:7	G_4:5	G_5:6	G_3:4
70	G_1:2	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
71	G_1:2	G_1:2	G_2:3	G_4:5	G_6:7	G_5:6	G_3:4
72	G_1:2	G_1:2	G_6:7	G_2:3	G_4:5	G_5:6	G_3:4
73	G_1:2	G_1:2	G_2:3	G_4:5	G_5:6	G_6:7	G_3:4
74	G_6:7	G_6:7	G_1:2	G_2:3	G_4:5	G_3:4	G_5:6
* 75	G_1:2	G_6:7	G_2:3	G_1:2	G_4:5	G_3:4	G_5:6
* 76	G_1:2	G_6:7	G_1:2	G_2:3	G_4:5	G_5:6	G_3:4