Grafické diagnostiky EDA při vícerozměrné klasifikaci těžkých kovů v půdách

Prof. RNDr. Milan Meloun, DrSc.,
Katedra analytické chemie, Univerzita Pardubice, 532 10 Pardubice,
milan.meloun@upce.cz,

Ing. Miluše Galusková,
Státní zdravotní ústav v Ostravě, Partyzánske nám. 7, 702 00 Ostrava,
miluse.galuszkova@zuova.cz

Prof. Ing. Jiří Militký, CSc.,
Katedra textilních materiálů, Technická univerzita Liberec, 461 17 Liberec,
jiri.militky@tul.cz

Souhrn: *Účelem průzkumové analýzy dat (EDA), jako prvního kroku v analýze jednorozměrných dat, je odhalení jejich statistických zvláštností pomocí 12 grafických diagnostik, ověření základních předpokladů a případná mocninná transformace. Při průzkumové analýze složitějších, nákladných nebo unikátních měření je totiž nutné posoudit zvláštnosti chování dat ještě před vlastní, rutinní statistickou analýzou. Jedině tak lze zabránit provádění numerických výpočtů bez hlubších statistických souvislostí. EDA je pak obzvláště důležitá u aplikace regulačních diagramů při kontrole kvality.* **.** *Vícerozměrná statistická analýza je založena na latentních proměnných, které jsou lineární kombinací původních. Zdrojová matice dat obsahuje kvalitativní znaky v 11 sloupcích a sledované půdy jako objekty v 80 řádcích zdrojové matice dat. Cílem analýzy dat je nalézt shluk podobných si půd a podobných znaků, jež půdy popisují. Podobnost půd je posuzována na základě jistých podobností či vzdálenosti půd v 11-rozměrném prostoru všech znaků dle kritéria, že čím je Mahalonobisova vzdálenost shluků půd větší, tím menší je jejich vzájemná podobnost. Strukturu a vazby mezi sledovanými znaky vystihují metody snížení dimensionality. Rozptylový diagram skóre zobrazuje objekty, rozptýlené v rovině prvních dvou hlavních komponent (PCA) či faktorů (FA). Graf komponentních vah porovnává vzdálenosti (podobnosti) mezi znaky, kde krátká vzdálenost značí silnou korelaci dvou znaků. Znaky ale také povodně lze seskupovat do shluků hierarchicky, a to dle předem zvoleného způsobu metriky a výsledkem je dendrogram. Původních 11 sledovaných znaků lze zredukovat ve tři latentní proměnné, faktory nebo hlavní komponenty.*

1. Úvod

Experimentální data se v laboratoři často vyznačují nekonstantním rozptylem, malým počtem, asymetrickým rozdělením a porušením základních předpokladů, kladených na výběr. Uveďme nejprve tři etapy obecné osnovy analýzy výběru dat.

A. V průzkumové analýze dat se vyšetřují *statistické zvláštnosti dat,* jako je lokální koncentrace dat, tvarové zvláštnosti rozdělení dat a přítomnost podezřelých hodnot. Odhalí se také anomálie a odchylky rozdělení výběru od typického rozdělení, obyčejně normálního (Gaussova). Interaktivní statistická analýza na počítači tento postup ulehčuje, většina statistického softwaru nabízí řadu diagnostických grafů a diagramů. Pokud je rozdělení dat nevhodné pro standardní statistickou analýzu (tj. většinou je asymetrické), provádí se nejprve vhodná úprava dat. Pokud bylo indikováno zešikmené rozdělení nebo

rozdělení s dlouhými konci, pomocníkem je mocninná a Boxova-Coxova transformace. Transformace je vhodná především při asymetrii rozdělení původních dat, ale také při nekonstantnosti rozptylu.

B. Pro případ rutinních měření se ověří *základní předpoklady,* kladené na výběr, jako jsou nezávislost prvků, homogenita výběru, dostatečný rozsah výběru a rozdělení výběru. Jsou-li závěry tohoto kroku optimistické, následuje vyčíslení klasických odhadů polohy a rozptýlení, tj. obyčejně aritmetického průměru a rozptylu. Dále se vyčíslí intervaly spolehlivosti, následované testováním statistických hypotéz. V pesimistickém případě následuje další pokus o úpravu dat.

C. V konfirmatorní analýze je nabízena paleta rozličných odhadů polohy, rozptýlení a tvaru, jež lze rozdělit do dvou skupin: na *klasické odhady* a na *robustní odhady* (necitlivé na odlehlé prvky výběru, resp. další předpoklady o datech). Z nabídky odhadů parametrů vybírá uživatel uvážlivě ty, jež mají statistický smysl a odpovídají závěrům průzkumové analýzy dat a ověření předpokladů o výběru.

Postup statistické analýzy jednorozměrných dat[1, 3, 5], prováděné v interaktivním režimu na počítači, lze shrnout do bloků, i když lze jednotlivé bloky provádět samostatně: blok A, blok B, blok A+B, blok B+C a konečně všechny bloky A+B+C.

2. Přehled operací analýzy jednorozměrných dat

A. Průzkumová (exploratorní) analýza dat (EDA):

Odhalení stupně symetrie a špičatosti výběrového rozdělení;
Indikace lokální koncentrace výběru dat;
Nalezení vybočujících a podezřelých prvků ve výběru;
Porovnání výběrového rozdělení dat s typickými rozděleními;
Mocninná transformace výběru;
Boxova-Coxova transformace výběru.

B. Ověření předpokladů o datech:

Ověření nezávislosti prvků výběru;
Ověření homogenity rozdělení výběru;
Určení minimálního rozsahu výběru;
Ověření normality rozdělení výběru.

C. Konfirmatorní analýza dat (CDA) - parametry polohy, rozptýlení a tvaru:

Klasické odhady (bodové a intervalové) z výběru;
Robustní odhady (bodové a intervalové) z výběru.

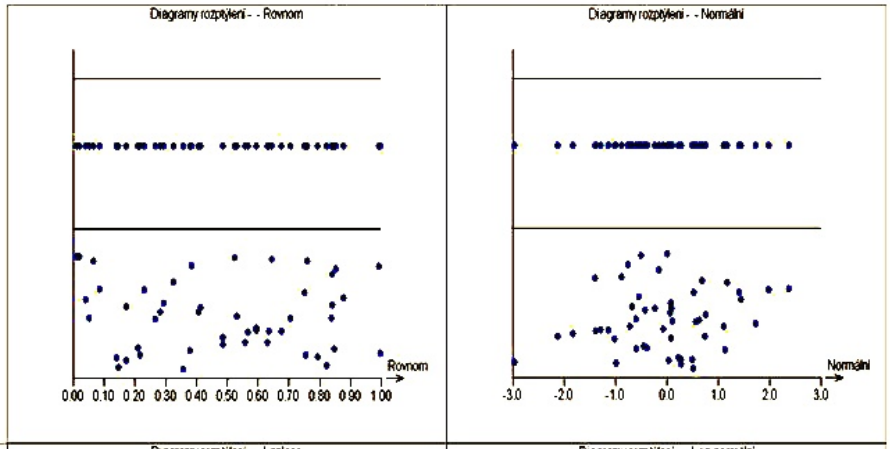
D. Vícerozměrná analýza zdojové matice dat

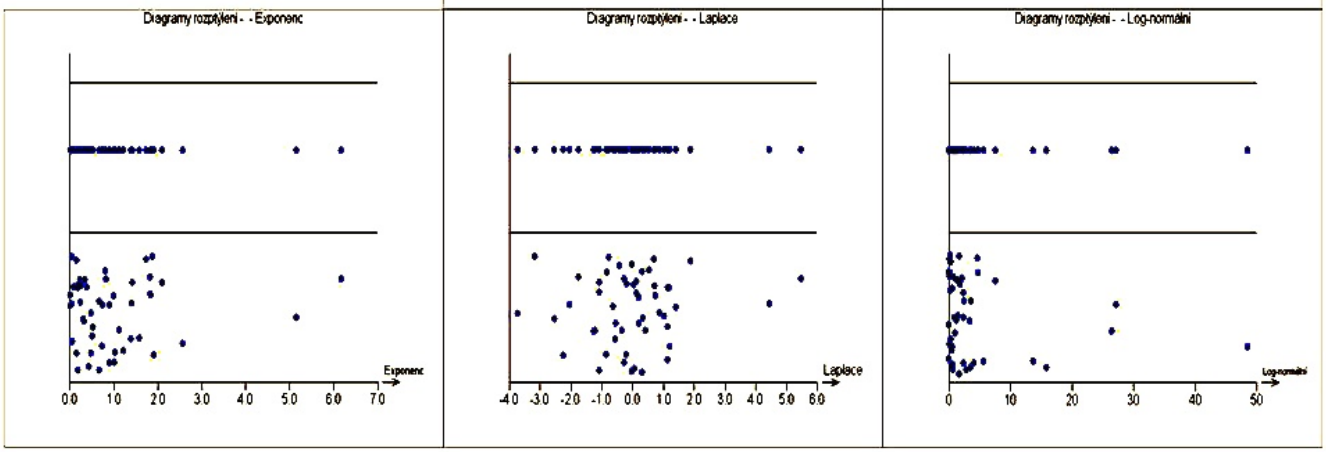
Metoda hlavních komponent a faktorová analýza.
Analýza shluků.
Klasifikace objektů diskriminační analýzou.

3. Grafické diagnostiky exploratorní analýzy

Diagram rozptýlení a rozmítnutý diagram rozptýlení

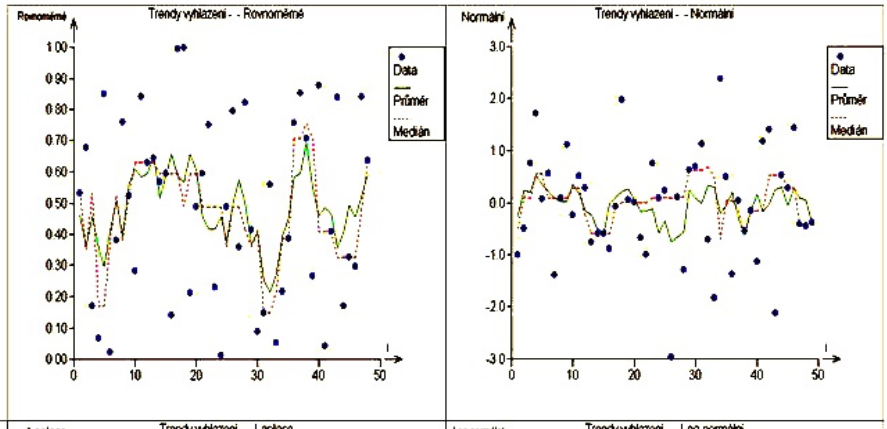
Představuje jednorozměrnou projekci kvantilového grafu do osy *x*. Ukazuje na lokální koncentrace dat a indikuje podezřelá a odlehlá měření. Rozmítnutý diagram rozptýlení představuje také projekci kvantilového grafu, body jsou však pro lepší přehlednost vhodně rozmítnuté. Popis osy y zde nemá význam. Aby nedošlo ke splývání shodných nebo blízkých dat, jsou ve spodní polovině grafu zobrazena táž data, ale náhodně rozmítnutá ("rozházená") na ose y.

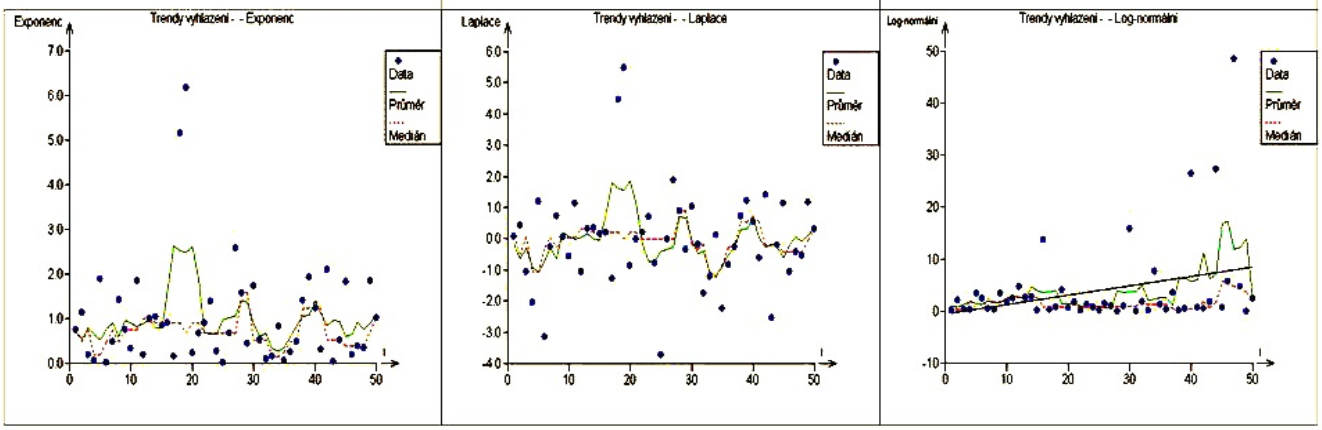




Trendy vyhlazení

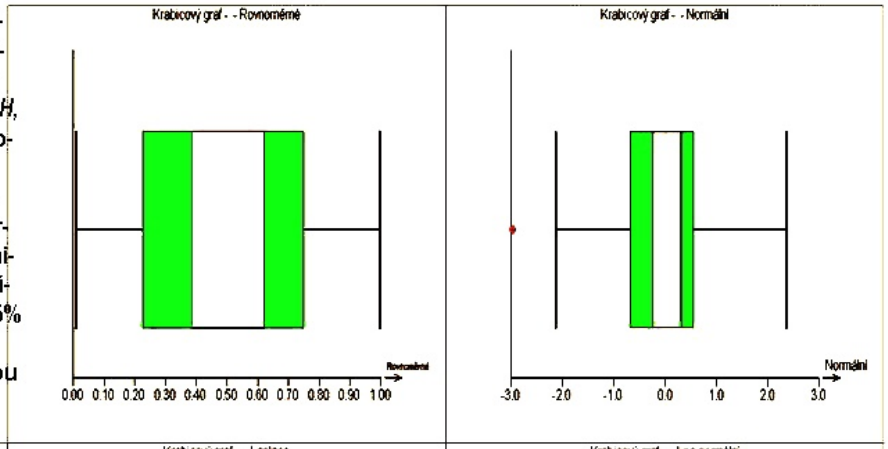
Grafické znázornění trendu v datech pomocí klouzavého průměru (plná křivka) a klouzavého mediánu (přerušovaná křivka). Čím je řád trendu větší, tím jsou křivky hladší, méně citlivé na lokální poruchy a zachycují spíše globální dlouhodobý trend. Menší řád zachycuje spíše lokální chování dat. Klouzavý medián je méně citlivý (robustní) na lokální poruchy v datech a jednotlivá odlehlá měření a tedy vhodnější pro tyto případy.

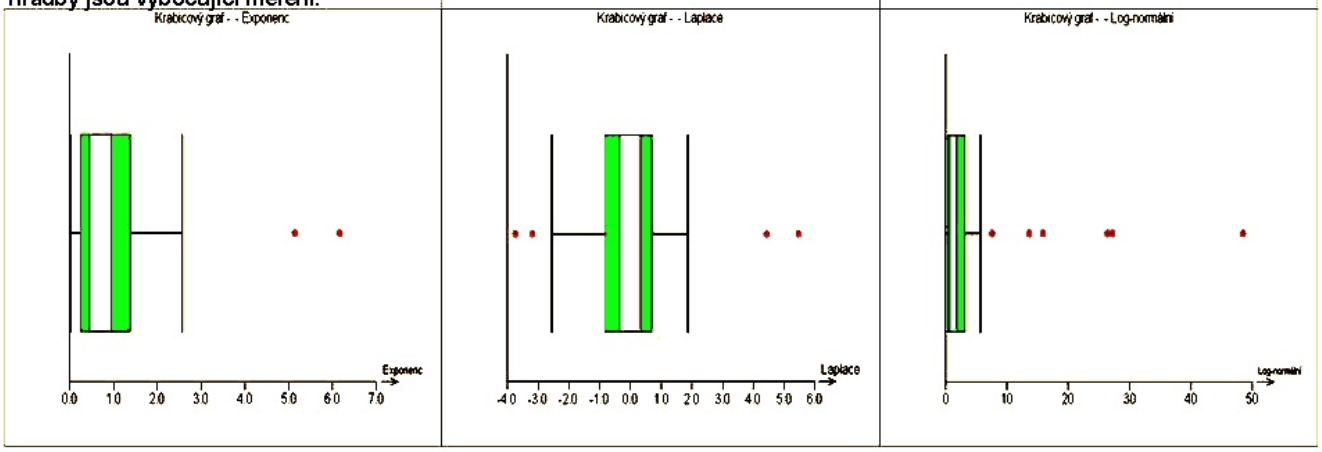




Krabicový graf a vrubový krabicový graf

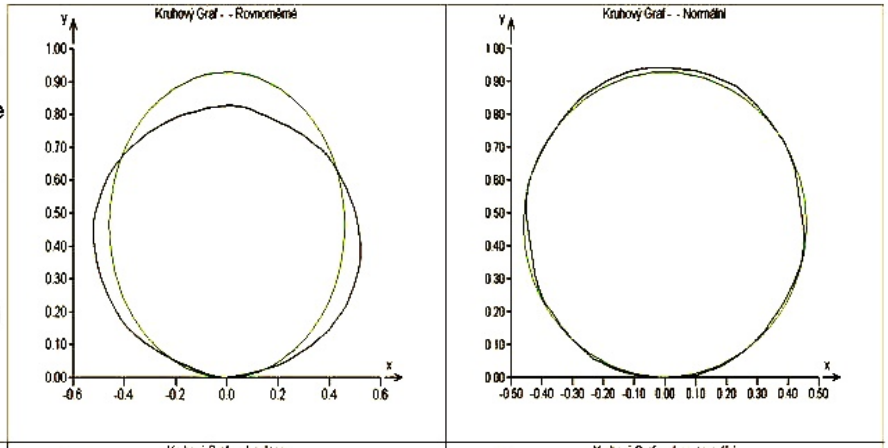
Umožňuje znázornění robustního odhadu polohy, mediánu *M*, symetrie v okolí kvartilů, u konců rozdělení a identifikaci odlehlých dat. Úsečky jsou ukončeny *vnitřními hradbami BH* , *BD*. Prvky mimo interval [*BH*, *BD*,] jsou podezřelé, obvykle vybočující body a jsou znázorněny kroužky. Vrubový krabicový graf umožňuje posouzení variability mediánu, $ID \leq M \leq IH$. IS bývá znázorněn bílým proužkem. Větší obdélník ohraničuje 50% dat, horní okraj zeleného obdélníku odpovídá 75% kvantilu, spodní okraj 25% kvantilu, střed bílého pruhu v zeleném obdélníku mediánu, dva černé proužky jsou tzv. vnitřní hradby. Data mimo vnitřní hradby jsou vybočující měření.

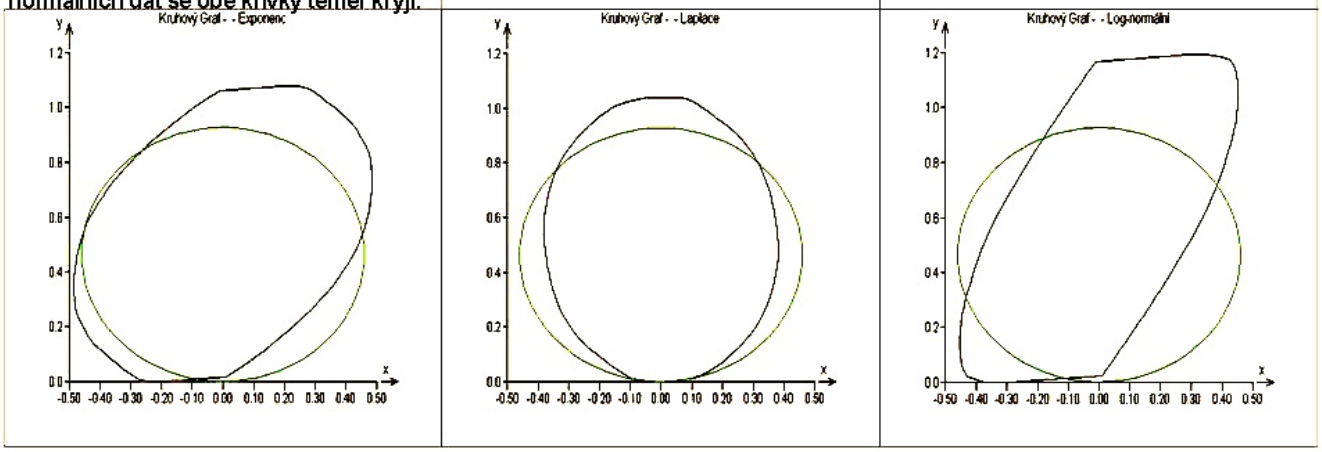




Kruhový graf

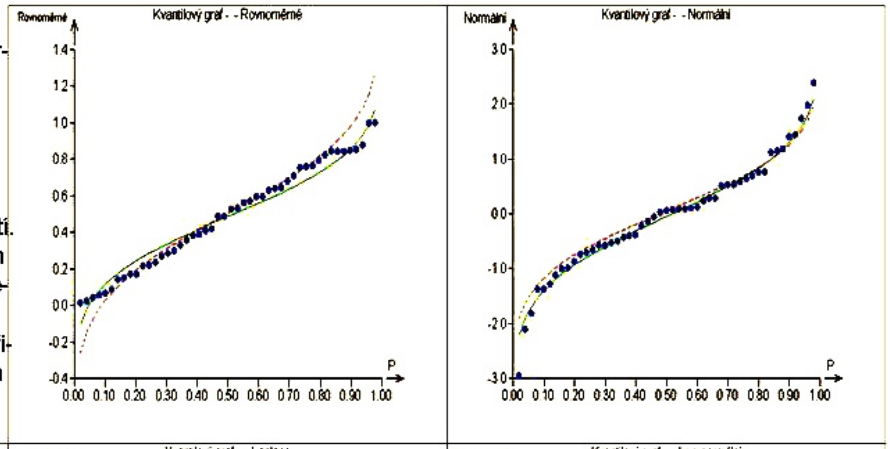
Kruhový graf slouží k vizuálnímu ověření hypotézy, že výběr pochází ze symetrického (nejčastěji Gaussova) rozdělení. V takovém případě je grafem regulární, konvexní polygon, blízký kružnici. Odchylky od kružnice ukazují na jiné než symetrické rozdělení výběru: (1) protáhlý elipsovitý tvar, s hlavní osou umístěnou úhlopříčně, ukazuje na asymetrické rozdělení, (2) elipsovitý tvar podél osy *x* ukazuje na rovnoměrné rozdělení. Graf slouží ke komplexnímu vizuálnímu posou-zení normality na základě kombinace šik-mosti a špičatosti. Zelený kruh (elipsa) je optimální tvar pro normální rozdělení, černý "kruh" představuje data. V případě normálních dat se obě křivky téměř kryjí.

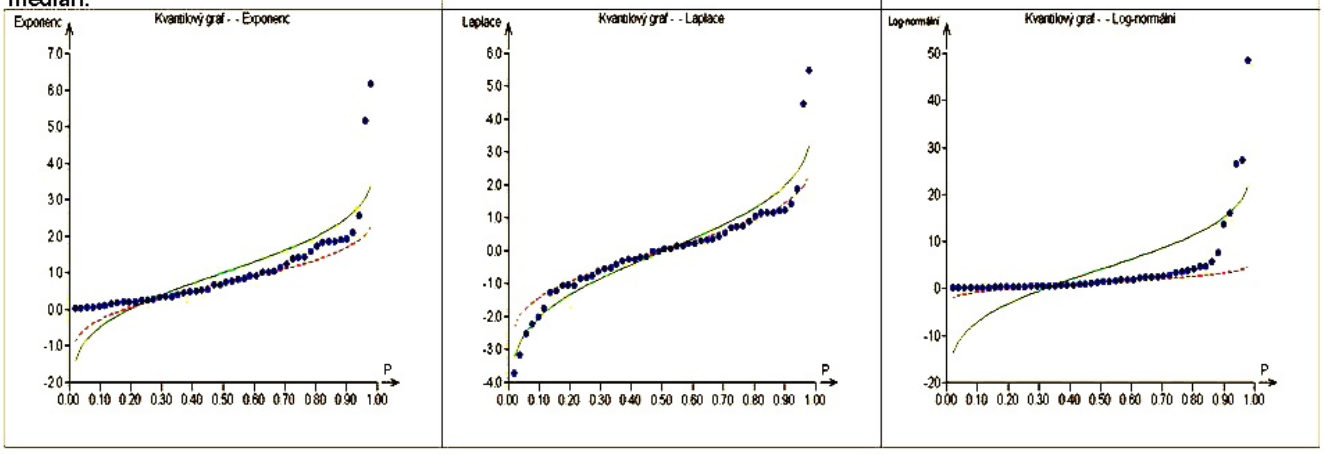




Kvantilový graf

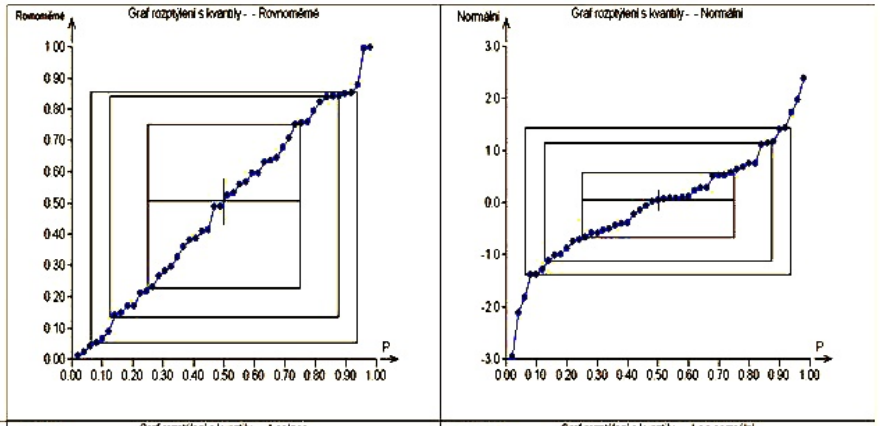
Znázorní data a rozliší tvar rozdělení, které může být symetrické, zešikmené k vyšším nebo nižším hodnotám. K porovnání s normálním rozdělením se do grafu zakreslují i kvantilové funkce normálního rozdělení: (1) *klasických odhadů* parametrů polohy a rozptýlení, tj. aritmetického průměru a směrodatné odchylky a (2) *robustních odhadů*, tj. medián a interkvartilové rozpětí. Zelená křivka odpovídá funkci s klasickým průměrem a rozptylem (nerobustní), červená křivka odpovídá mediánu a mediánové odchylce (robustní). Podle toho, která z křivek lépe prokládá data, je vhodné zvolit za odhad střední hodnoty průměr nebo medián.

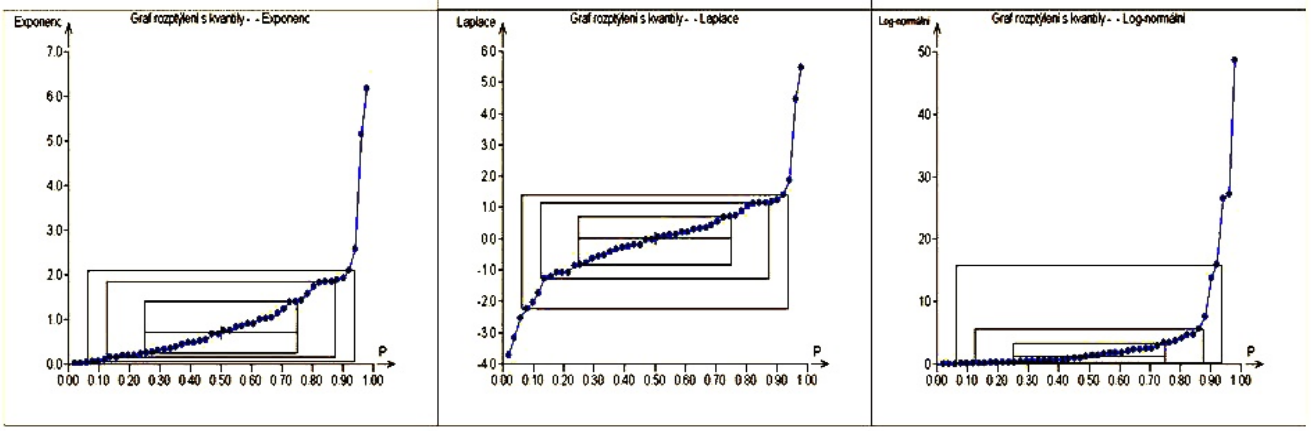




Graf rozptýlení s kvantily

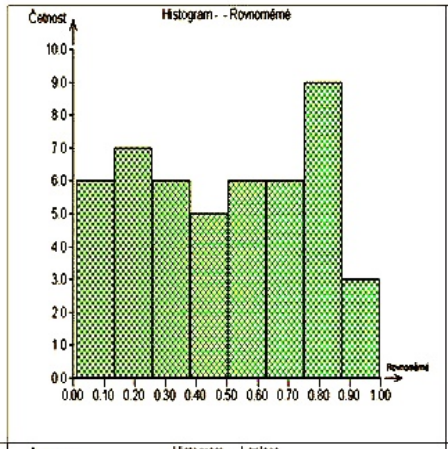
Základem je odhad kvantilové funkce výběru, který se získá spojením bodů $\{x_{(i)}, P_i\}$ lineárními úseky. Pro symetrická rozdělení má kvantilová funkce sigmoidální tvar. 1. *Symetrické unimodální rozdělení výběru* obsahuje obdélníky symetricky uvnitř sebe. 2. *Nesymetrická rozdělení* mají vzdálenosti mezi dolními hranami obdélníků *F*, *E* a *D* pro rozdělení zešikmené k vyšším hodnotám výrazně kratší než mezi horními. 3. *Odlehlá pozorování* jsou indikována. 4. *Vícemodální rozdělení:* na kvantilové funkci uvnitř obdélníku *F* je několik úseků s téměř nulovými směrnicemi.

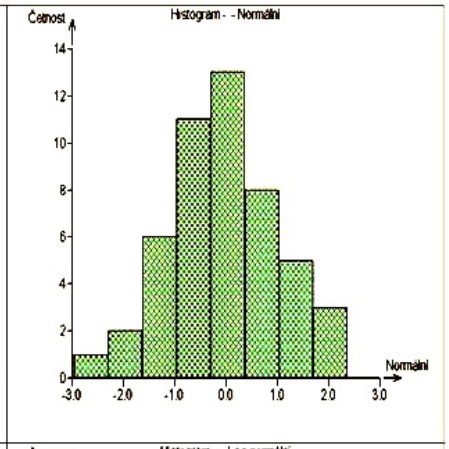


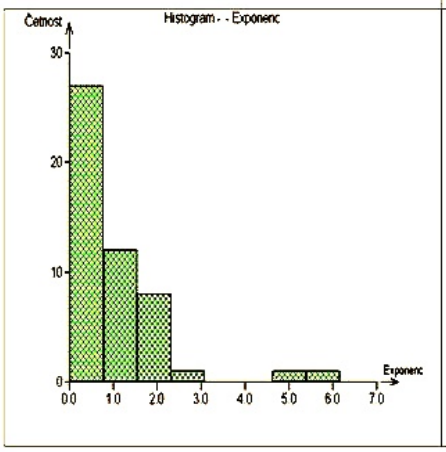


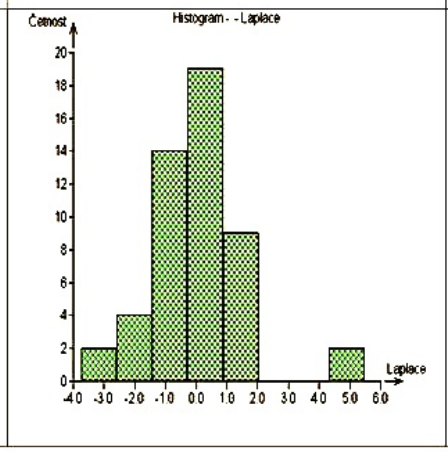
Histogram

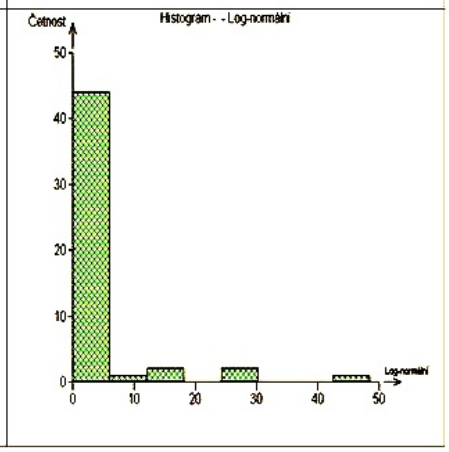
Jde o obrys sloupcového grafu, kde jsou na ose *x* jednotlivé třídy. Výšky sloupců odpovídají empirickým hustotám pravděpodobnosti. Kvalitu histogramu ovlivňuje volba počtu tříd *L*. Pro přibližně symetrická rozdělení výběru lze vyčíslit počet tříd *L* podle *L* = int(2*sqrt n), kde funkce int(*x*) je celočíselná část čísla *x*. optimální počet tříd je stanovován automaticky dle počtu dat.





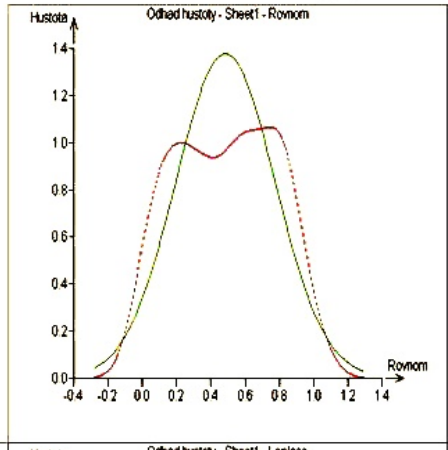


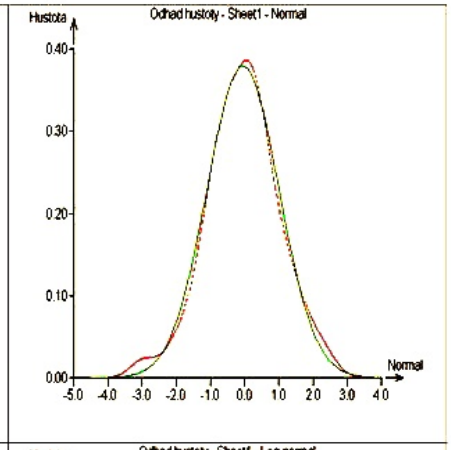


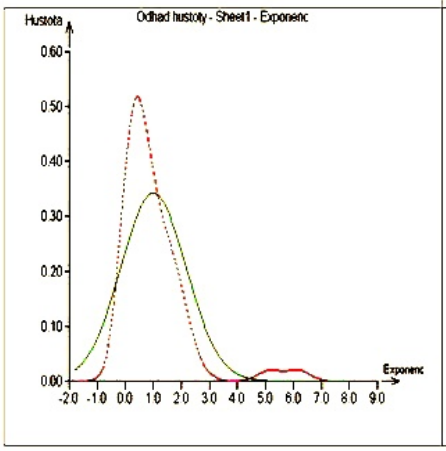


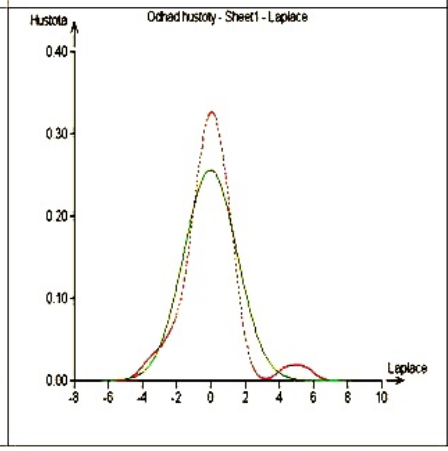
Graf jádrového odhadu hustoty pravděpodobnosti

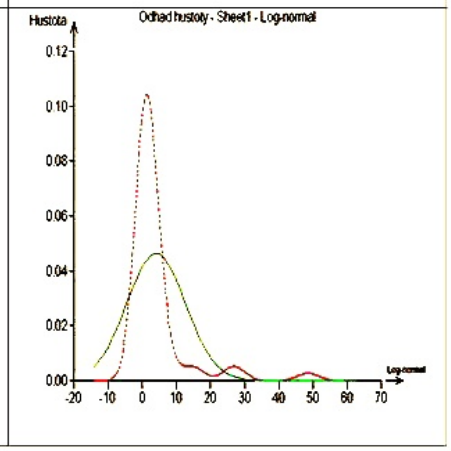
Porovnání průběhu hustoty pravděpodobnosti normálního rozdělení (plná čára) s jádrovým odhadem hustoty vypočítaným na základě dat (přerušovaná čára). Jádrový odhad používá Gaussovské jádro. Hladkost křivky jádrového odhadu je dána parametrem "Vyhlazení hustoty". Jsou-li data nehomogenní a tvoří shluky, může se objevit více maxim na jádrovém odhadu. V případě normality a většího množství dat jsou si obě křivky blízké. Je však třeba si uvědomit, že při příliš malém parametru vyhlazení hustoty se objeví maxima pro každá data.





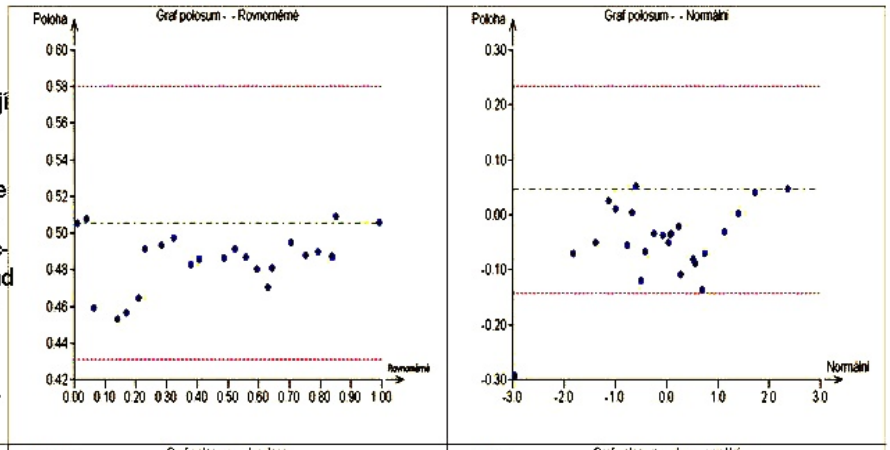


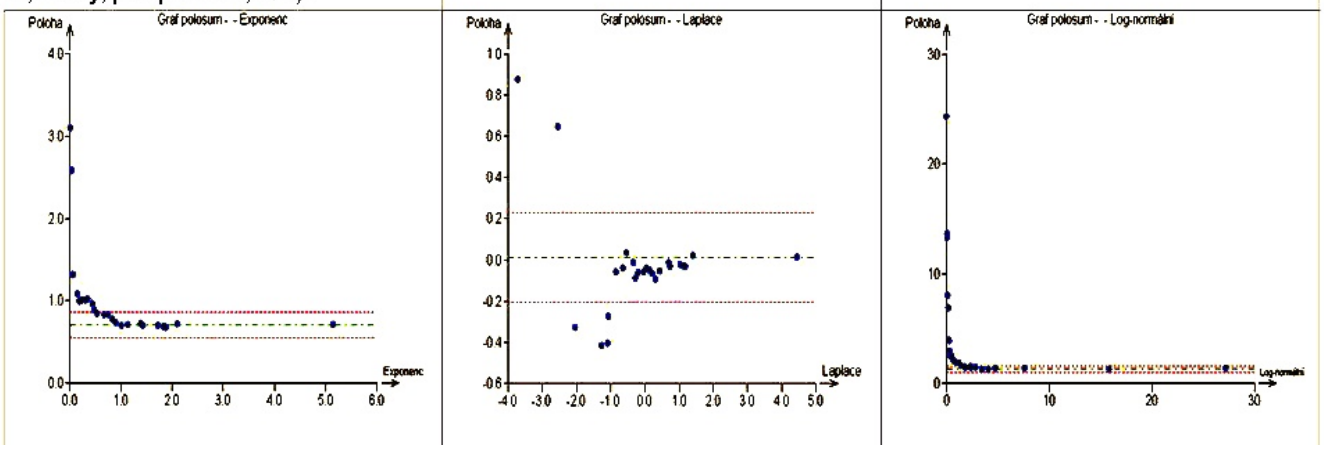




Graf polosum

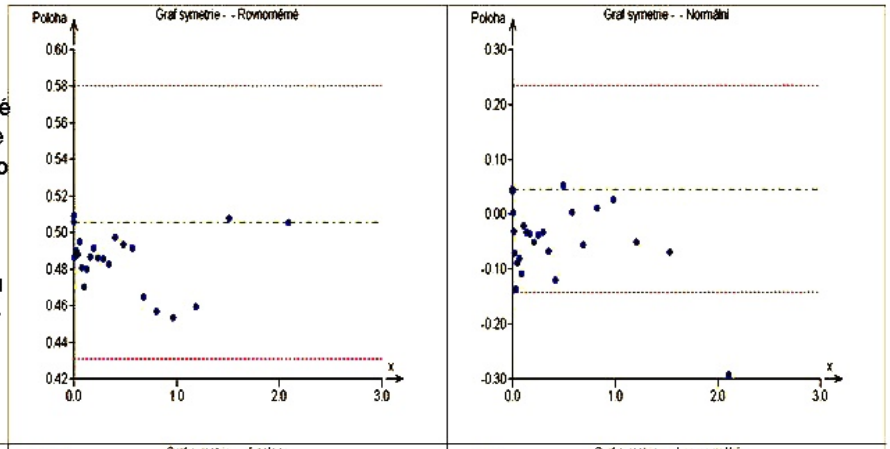
Pro symetrické rozdělení je grafem horizontální přímka, určená rovnicí *y* = *M*. Body oscilují okolo horizontální přímky a vykazují tak náhodný shluk a měřítko osy *y* je silně detailní. *Pro asymetrické rozdělení* vykazuje nenáhodný trend, body neoscilují okolo horizontální přímky a měřítko osy *y* není detailní. Citlivý indikátor asymetrie rozdělení. Horizontální přímka představuje medián a červené přerušované meze jeho interval spolehlivosti. V případě asymetrického rozdělení vykazují body výrazný trend (rostoucí pro zápornou šikmost, nebo klesající, pro kladnou šikmost) výrazně překračující přerušované meze. Body jsou konstruovány ze dvojic dat (první, poslední; druhý, předposlední; atd.).

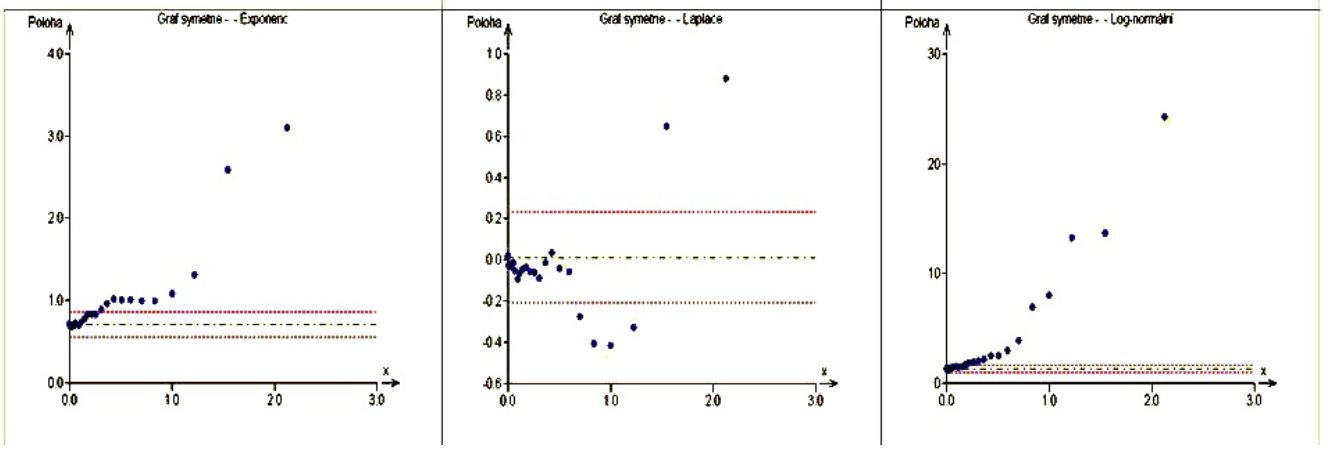




Graf symetrie

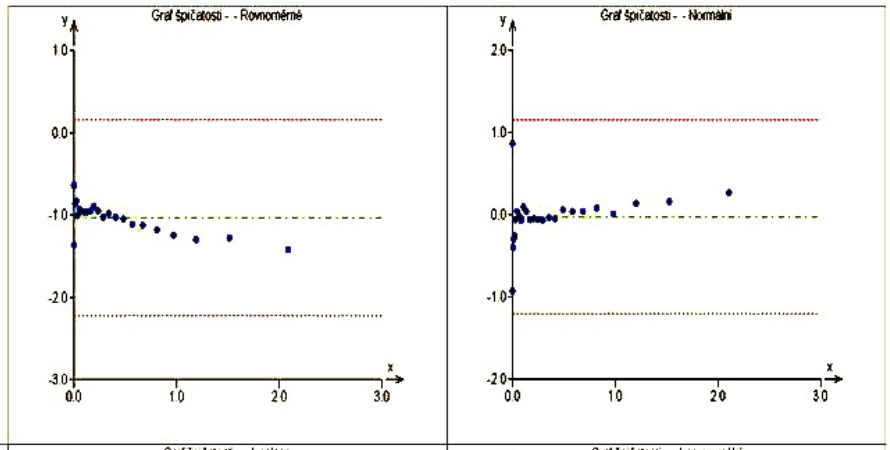
Symetrická rozdělení jsou charakterizována přímkou *y* = *M*. *Pro asymetrické rozdělení* nemá přímka nulovou směrnici. Směrnice je odhadem šikmosti. Asymetrické rozdělení vykazuje body uspořádané v trendu nějaké křivky. Směrnice trendu je úměrná šikmosti. V případě asymetrického rozdělení vykazují body výrazný trend (rostoucí pro zápornou šikmost, nebo klesající, pro kladnou šikmost) výrazně překračující přerušované meze. Body jsou konstruovány ze dvojic dat (první, poslední; druhý, předposlední; atd.), proto označením bodu jsou označena dvě příslušná data v tabulce.

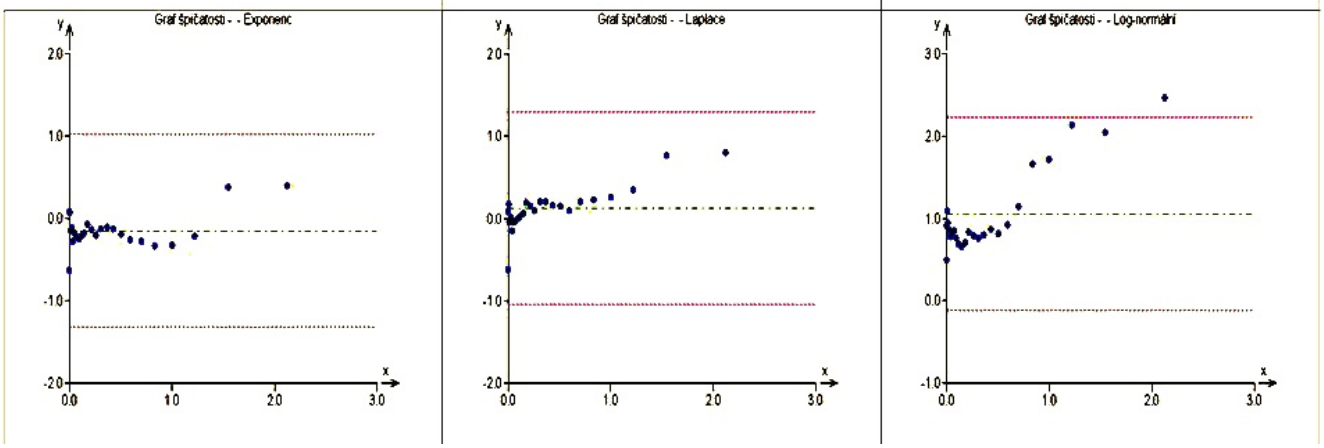




Graf špičatosti

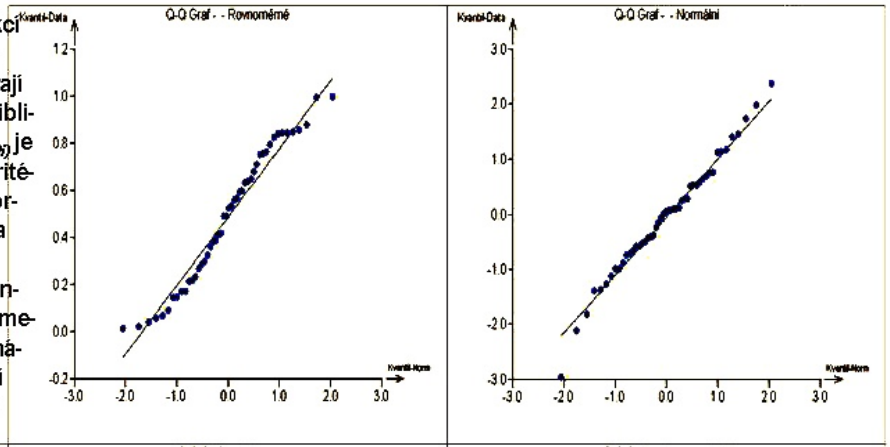
Pro normální rozdělení je grafem horizontální přímka a body leží převážně na této přímce. Pokud body tvoří nenáhodný trend, odpovídá hodnota směrnice této aktuální přímky parametru špičatosti. Má podobný význam jako dva předchozí grafy. Směrnice případného trendu je úměrná odchylky špičatosti od 3. V případě výrazně nenormální špičatosti rozdělení vykazují body výrazný trend. Body jsou konstruovány ze dvojic dat (první, poslední; druhý, předposlední; atd.), proto označením bodu jsou označena dvě příslušná data v tabulce.

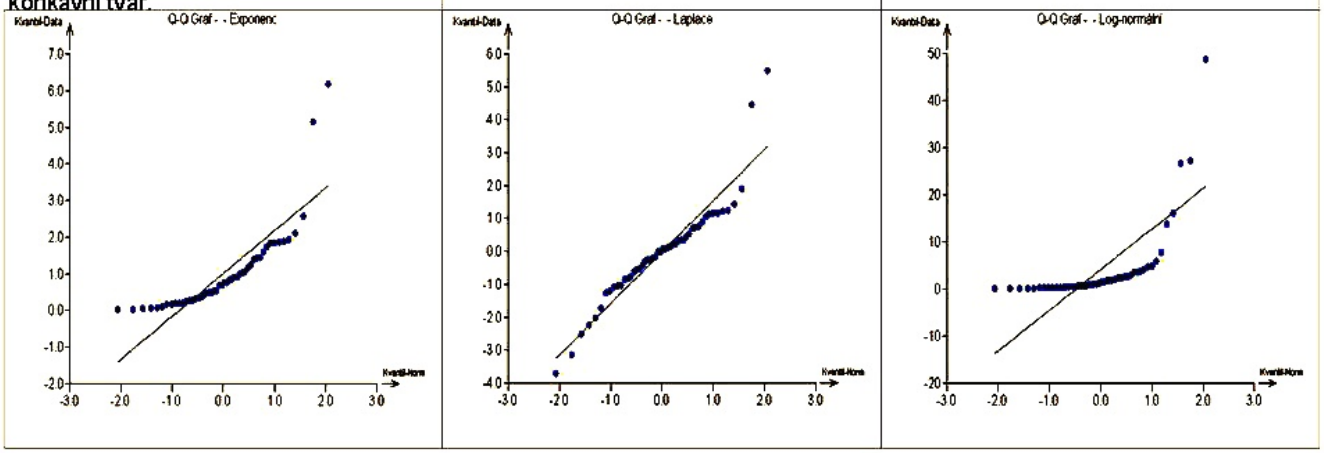




Kvantil-kvantilový Q-Q graf

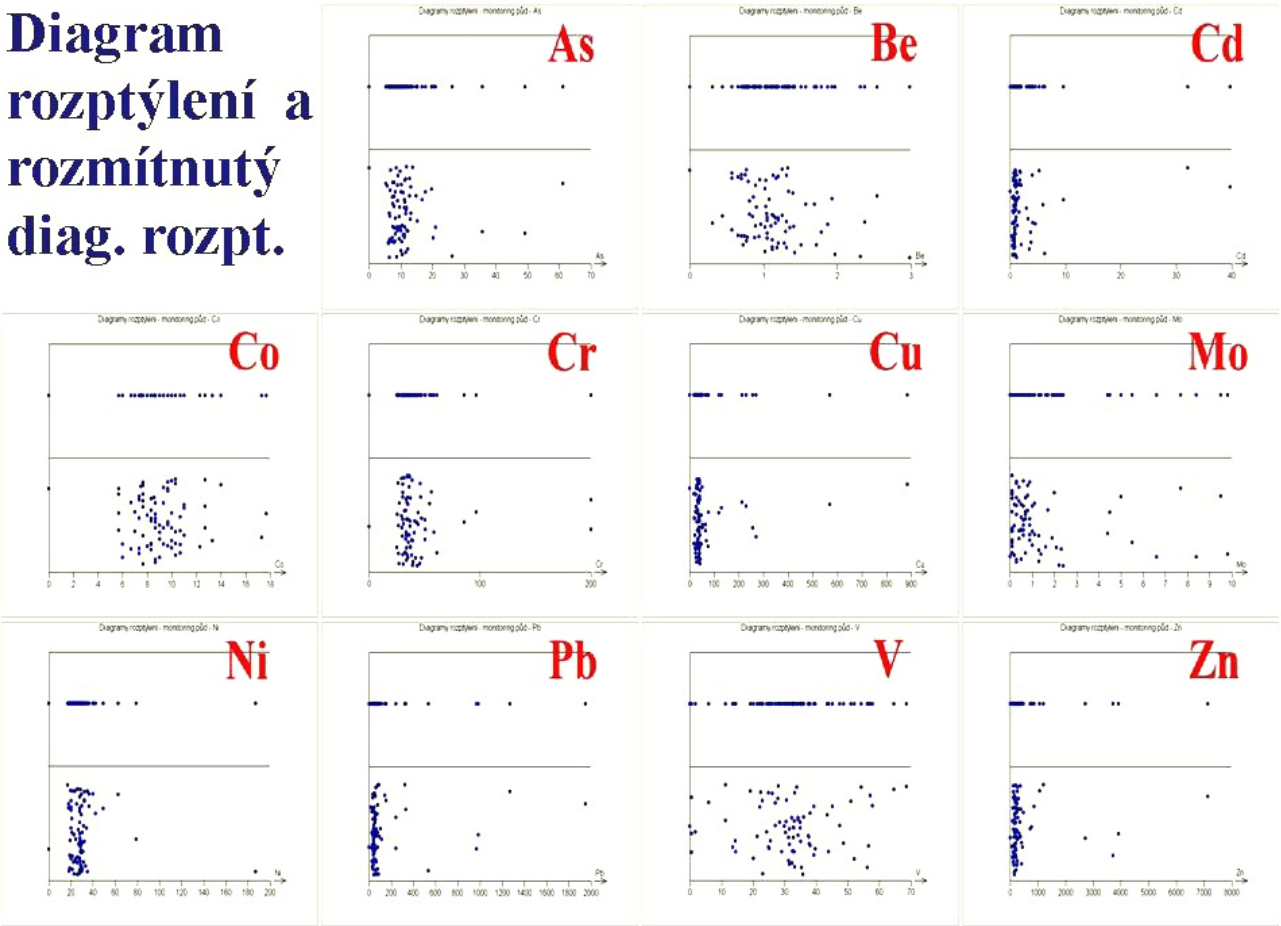
Osa x: $Q_{T(P)}$, osa y: $x_{(i)}$. Posoudí shodu výběrového rozdělení, jež je charakterizováno kvantilovou funkcí $Q_{E(P)}$ s kvantilovou funkcí zvoleného teoretického rozdělení $Q_{T(P)}$. Za odhad kvantilové funkce výběru se využívají pořádkové statistiky $x_{(i)}$. Při shodě platí přibližná rovnost kvantilů a závislost $x_{(i)}$ na $Q_{T(P_i)}$ je přibližně přímka. Korelační koeficient je kritériem těsnosti proložení přímky. Testuje normalitu a odlehlá měření; pro normální data bez odlehlých měření má tvar přímky, a s odlehlými měřeními má tvar přímky s koncovými body mimo tuto přímku; pro sešikmená data s kladnou šikmostí (např. lognormální, exponenciální) má nelineární konvexní tvar a se zápornou šikmostí má nelineární konkávní tvar.



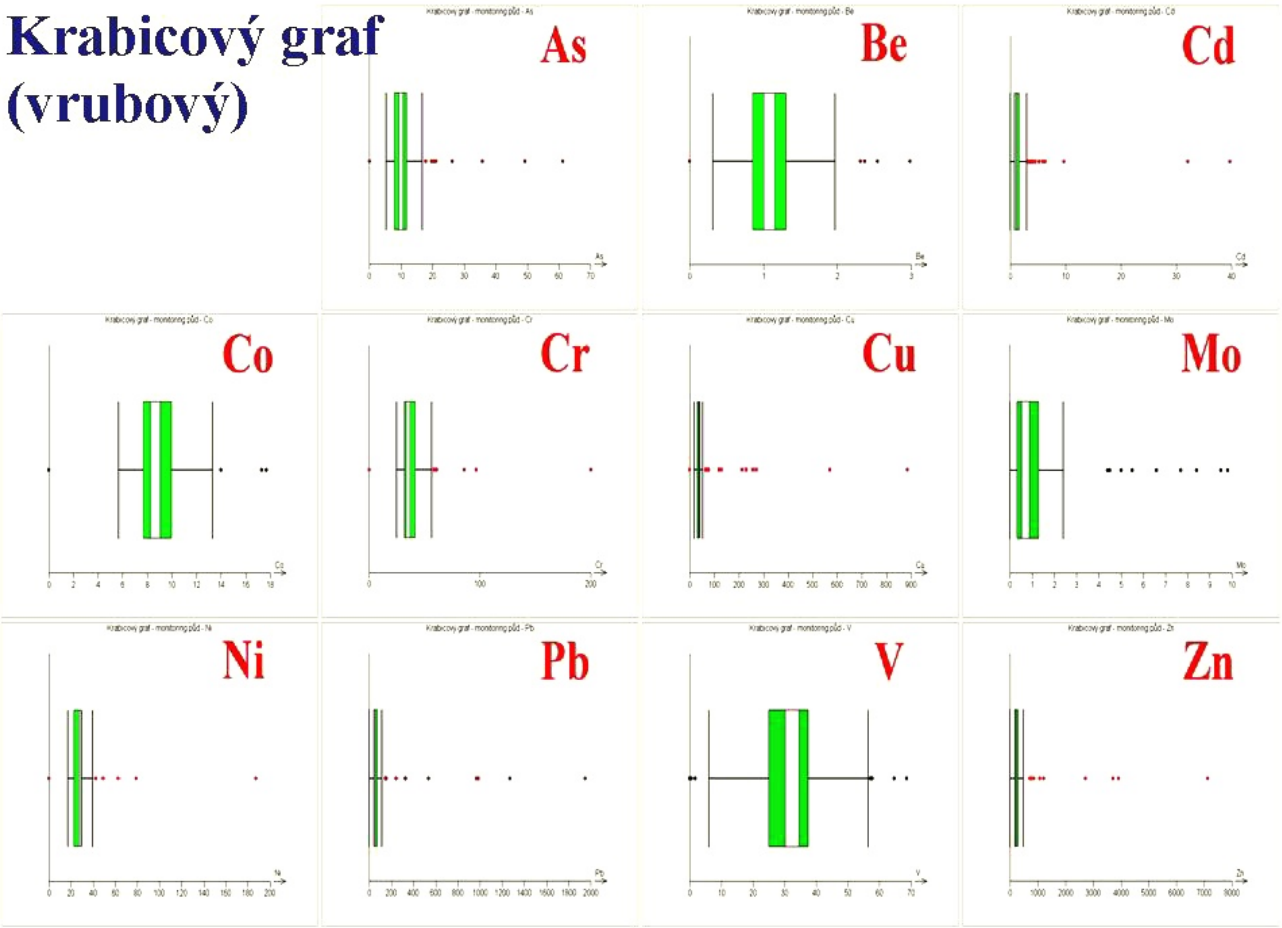


4. Monitoring půd
4.1 Exploratorní analýza jednorozměrných výběrů dat

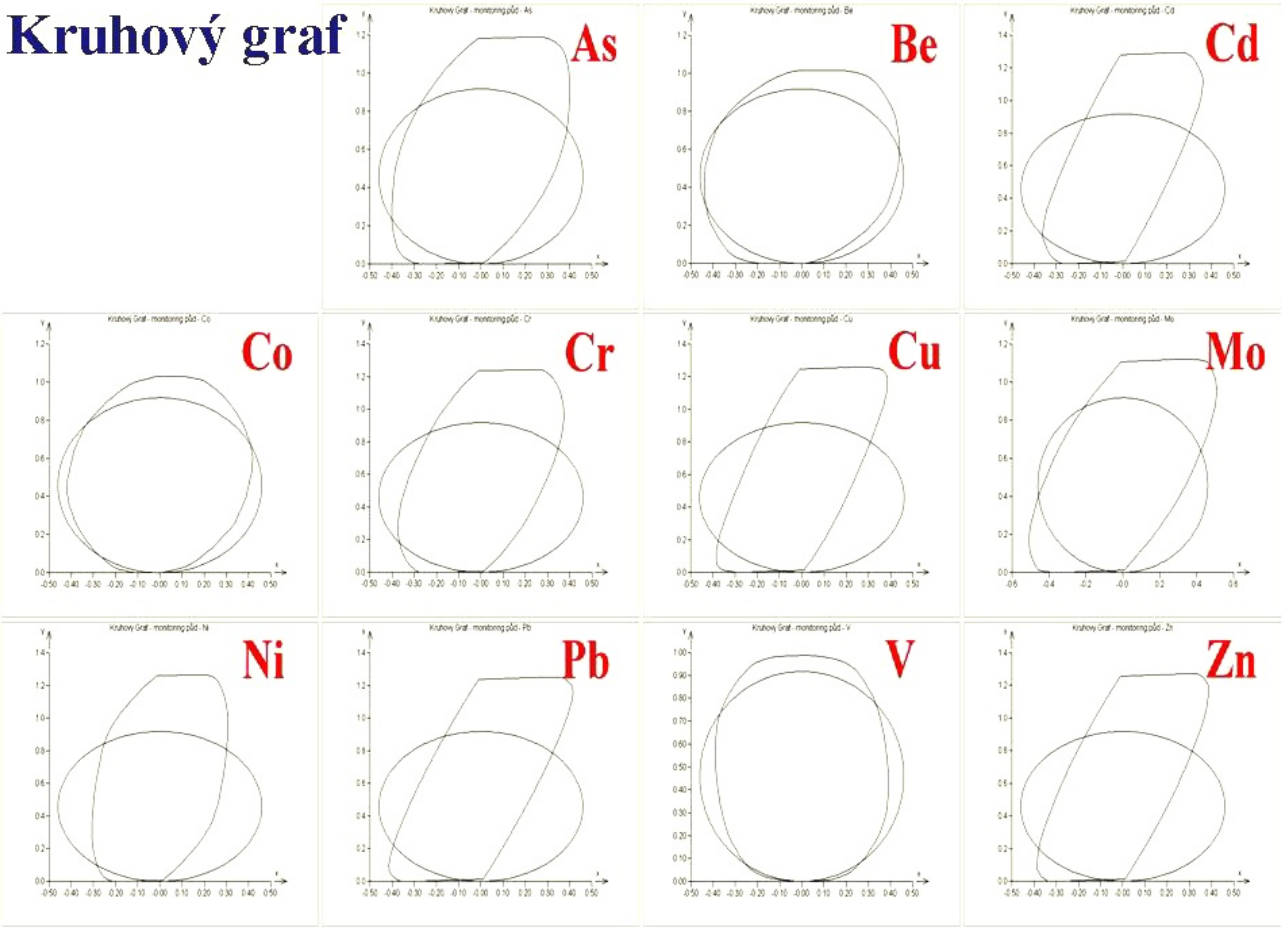
V kontaminované průmyslové oblasti Moravy byl prováděn monitoring půd. Půdy byly mineralizovány lučavkou královskou, v mineralizátech byly stanoveny kovy metodou AAS. Byla získána zdrojová matice dat 80 objektů (čili vzorků půd), 11 znaků čili obsahů prvků, převážně těžkých kovů. Exploratorní analýzou zdrojové matice dat EDA s využitím především grafických diagnostik byla vyšetřována symetrie rozdělení prvků výběru, přítomnost odlehlých hodnot a předpoklady o reprezentativním náhodném výběru jako jsou homogenita výběru, nezávislost a normální rozdělení prvků výběru. Následující grafy exploratorní analýzy přináší výsledky tohoto vyšetření.



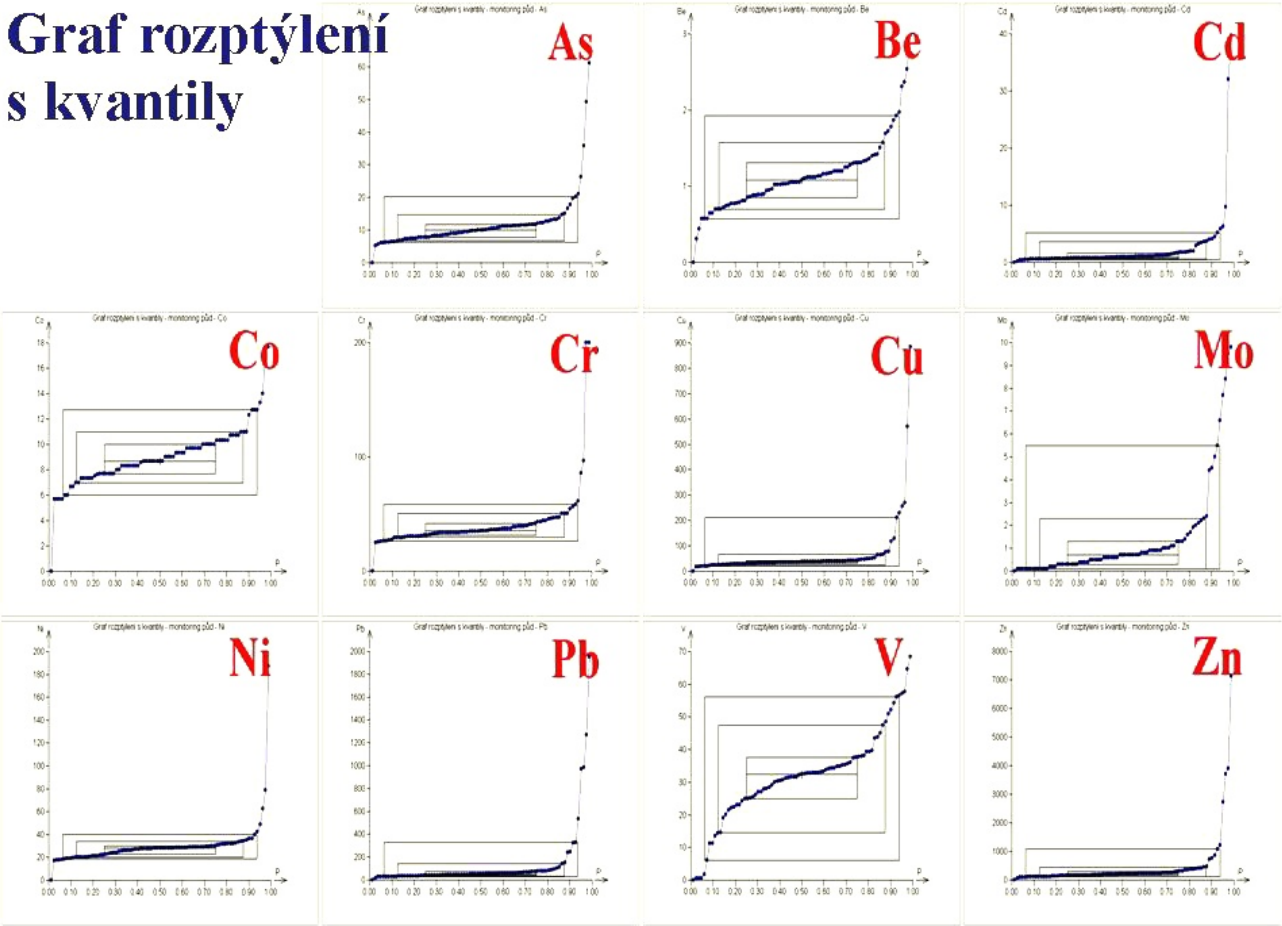
Krabicový graf (vrubový)



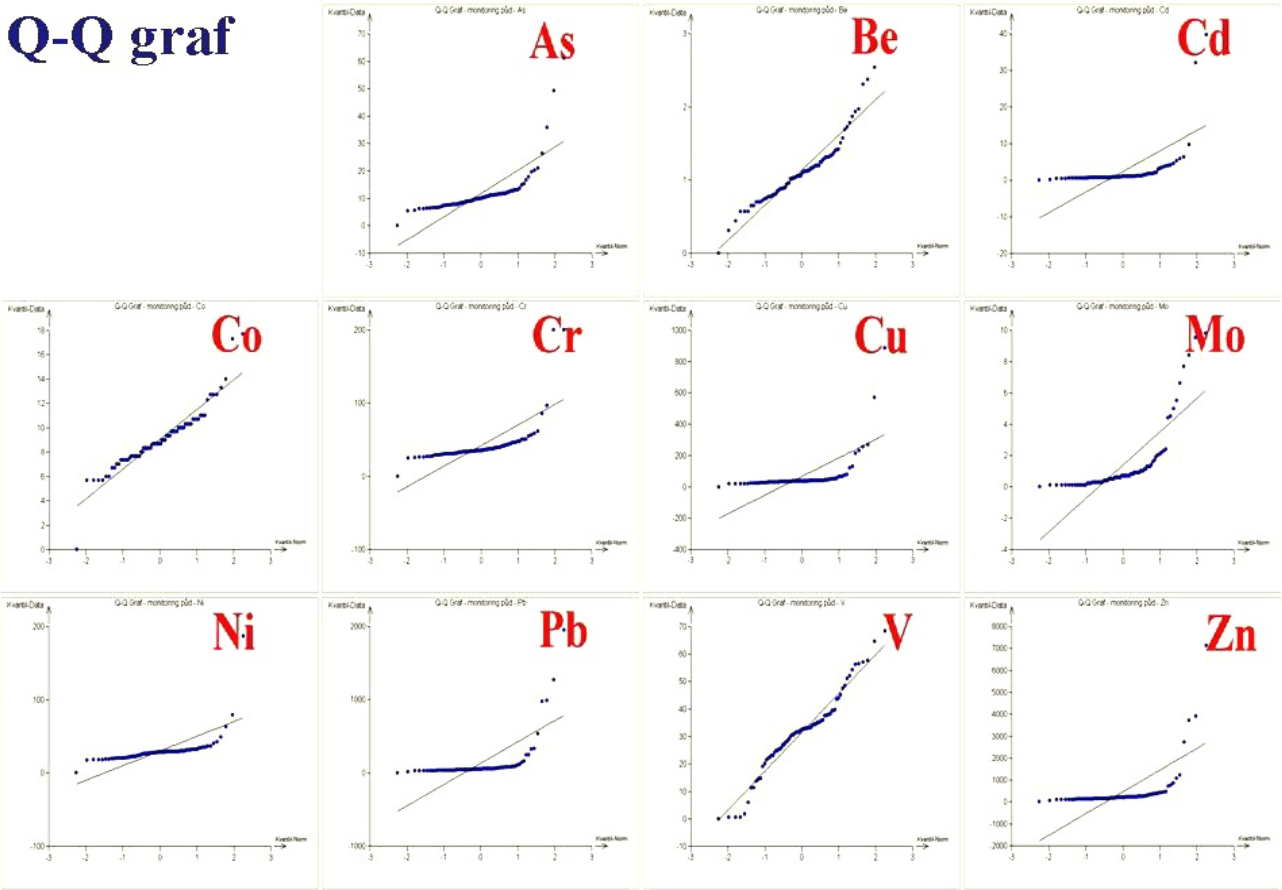
Kruhový graf

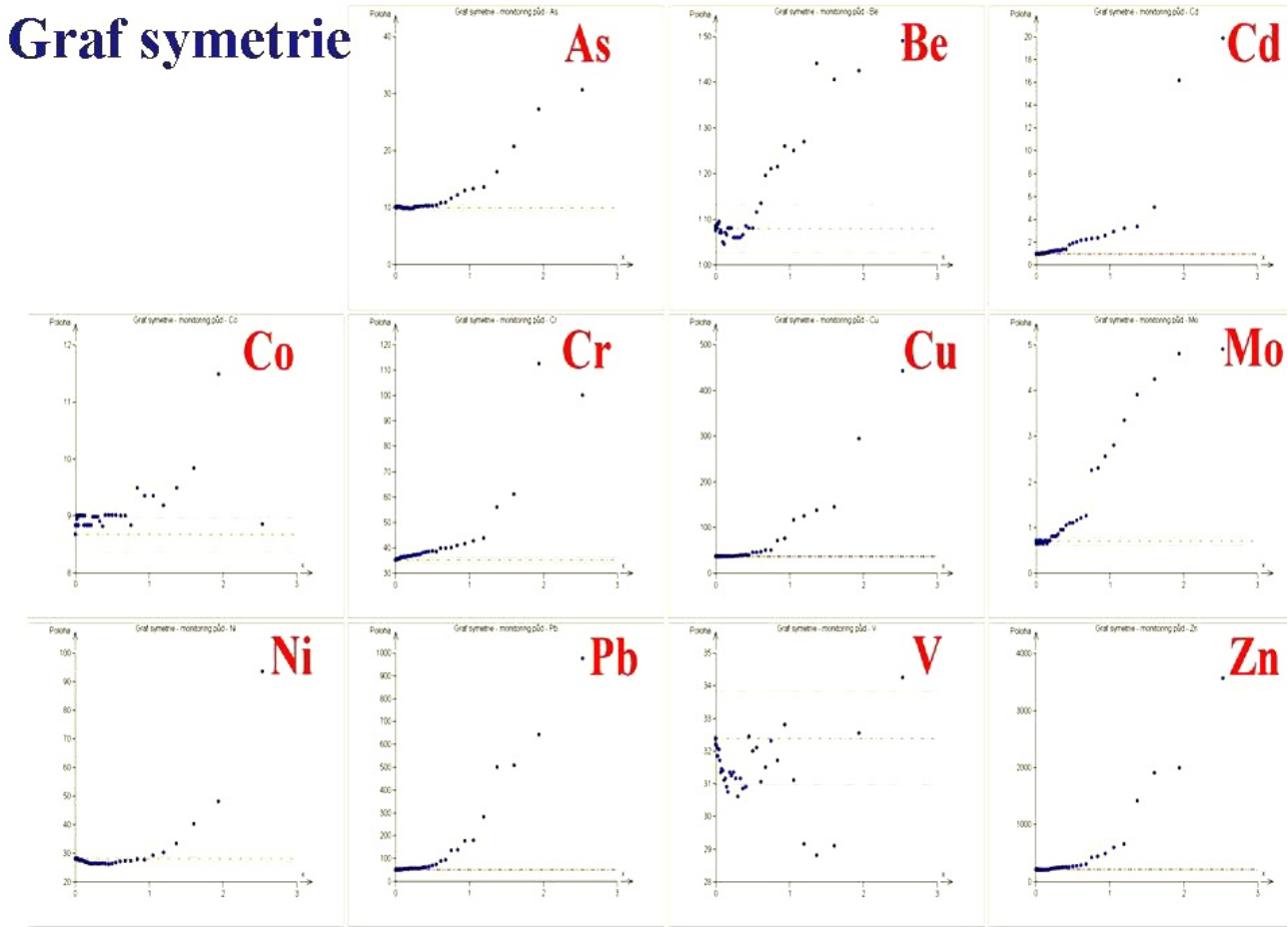


Graf rozptýlení s kvantily



Q-Q graf

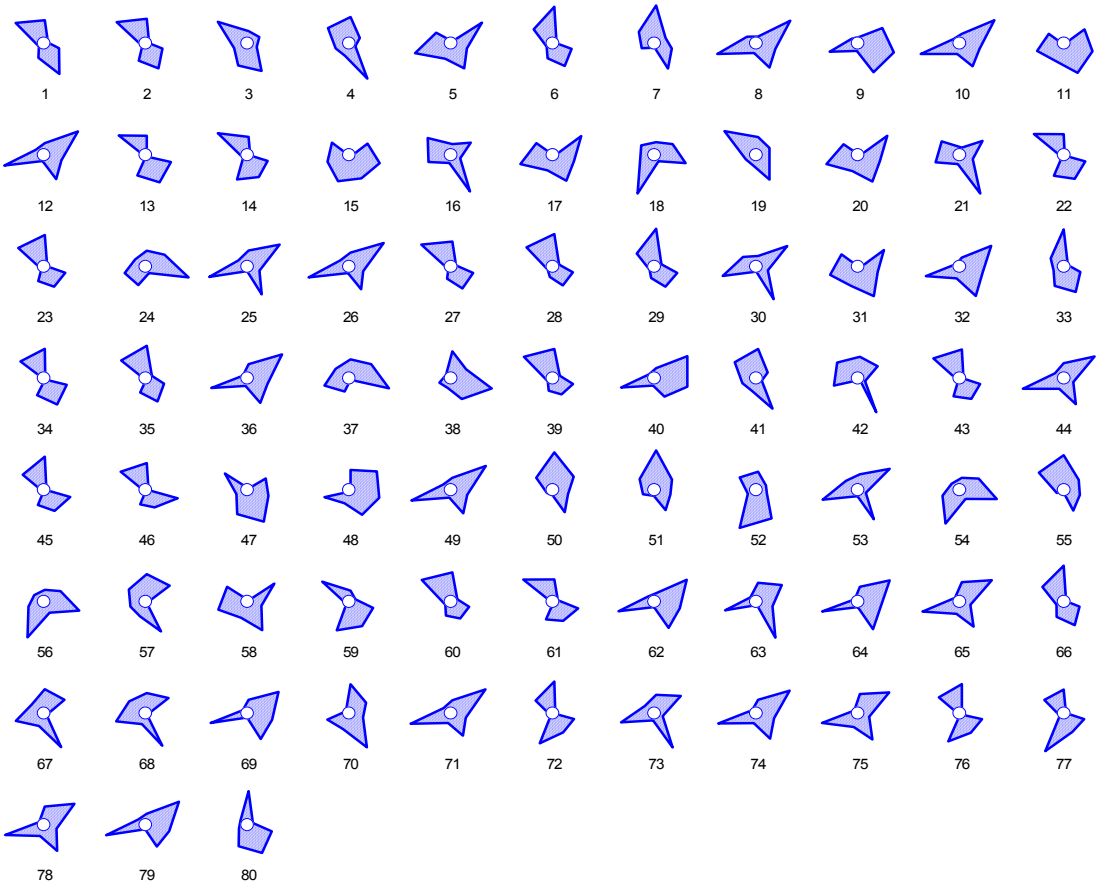




4.2 Vícerozměrná analýza dat

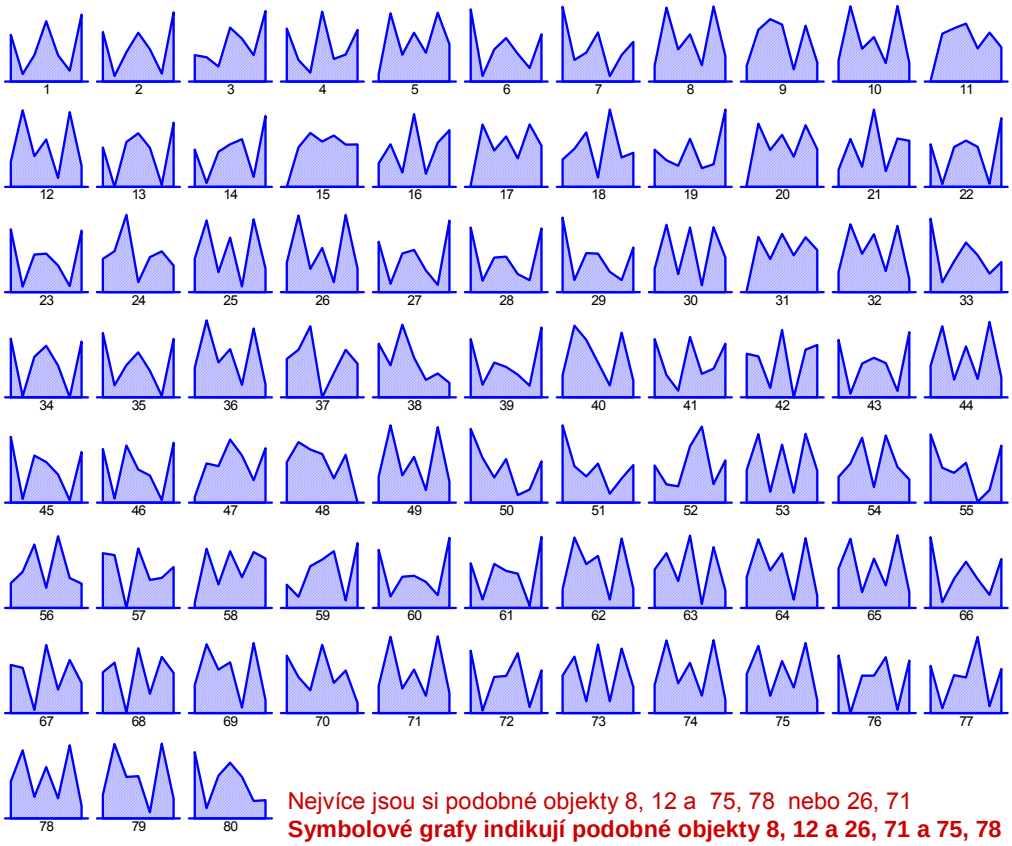
S využitím metody hlavních komponent PCA a faktorové analýzy FA, doplněné o shlukovou analýzu a korelační matice je třeba diagnostikovat odlehlé body (vzorky půd), které nejsou společné pro většinu znaků, např. 26, 79. Z důvodu nebezpečí ztráty informace, kterou obsahují nebudou odlehlé body vyřazeny ze zdrojové matice Monitoring půd. Popisné statistiky parametrů polohy a rozptýlení slouží také k posouzení intenzity vztahu mezi proměnnými a určení vhodného počtu proměnných. Krabicové grafy zobrazují proměnlivost znaků Cd, Pb, Ni, Cu, Co, Cr, Zn, As, Be, V, Mo. Na obrázcích jsou krabicové grafy pro data standardizovaná, které znázorňují proměnlivost znaků matice zdrojových dat Monitoring půd.

Ikonový graf (půdy monitoring 11v*80c)



Zleva doprava:
V
Cd
Pb
Zn
Cu
Be
Cr

Ikonový graf (půdy monitoring 11v*80c)



Zleva doprava:
V
Cd
Pb
Zn
Cu
Be
Cr

Nejvíce jsou si podobné objekty 8, 12 a 75, 78 nebo 26, 71

Symbolové grafy indikují podobné objekty 8, 12 a 26, 71 a 75, 78

Pro účely průzkumové analýzy vícerozměrných dat se používá různých technik umožňujících jejich grafické zobrazení ve dvourozměrném souřadnicovém systému a u novějších výpočetních prostředků i v třírozměrném souřadnicovém systému. Toto zobrazení umožňuje:a) identifikovat vektory $\boldsymbol{x_i}$, $i = 1, ..., n$, nebo jejich složky, které se jeví jako vybočující, b) indikovat různé struktury v datech, jako jsou shluky, jež ukazují na heterogenitu použitého výběru nebo přítomnost různých dílčích výběrů s odlišným chováním.

Na základě těchto informací a výsledků testů normality (popř. grafických ekvivalentů těchto testů) pak může být před vlastní statistickou analýzou provedena řada různých korekcí vedoucích k odstranění nehomogenity výběru a přiblížení se k vícerozměrné normalitě.

Většina používaných technik pro zobrazení vícerozměrných dat se dá zařadit do jedné ze dvou základních skupin, kterými jsou *zobecněné rozptylové diagramy* a *symbolové grafy*.

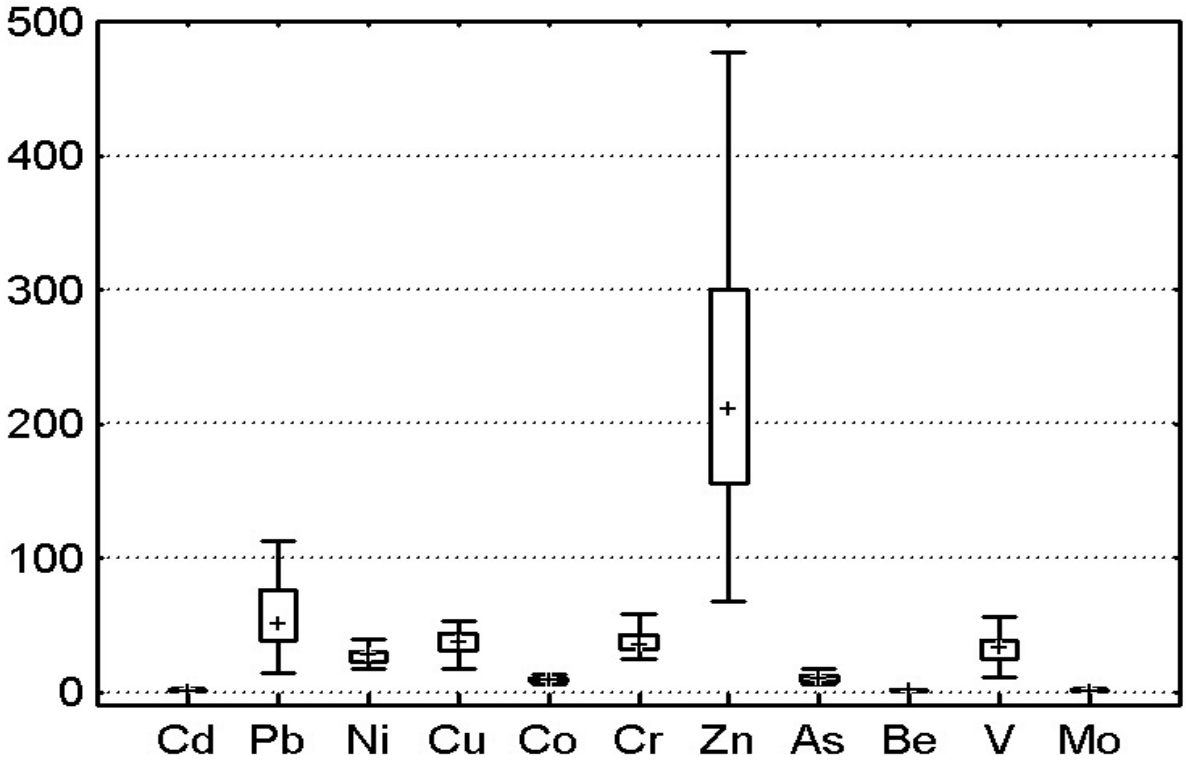
Pro základní případ dvojice náhodných znaků ($m = 2$) lze konstruovat rozptylové grafy, které umožňují sledovat statistické zvláštnosti dat. Ke konstrukci výběrového rozdělení nebo jeho porovnání s rozděleními teoretickými je možné použít i histogramy, neparametrické odhady hustoty a jiné grafy. Problémy však nastávají u vícerozměrných dat pro $m > 2$, kdy je třeba buď volit několik různých grafů, či vhodným způsobem provést transformace na dvoudimenzionální data.

Nelze obecně říci, která metoda zobrazení vícerozměrných dat je nejlepší. Závisí to na počtu znaků m vektorů $\boldsymbol{x_i}$, $i = 1, ..., n$, počtu měření a specifických zvláštnostech dat.

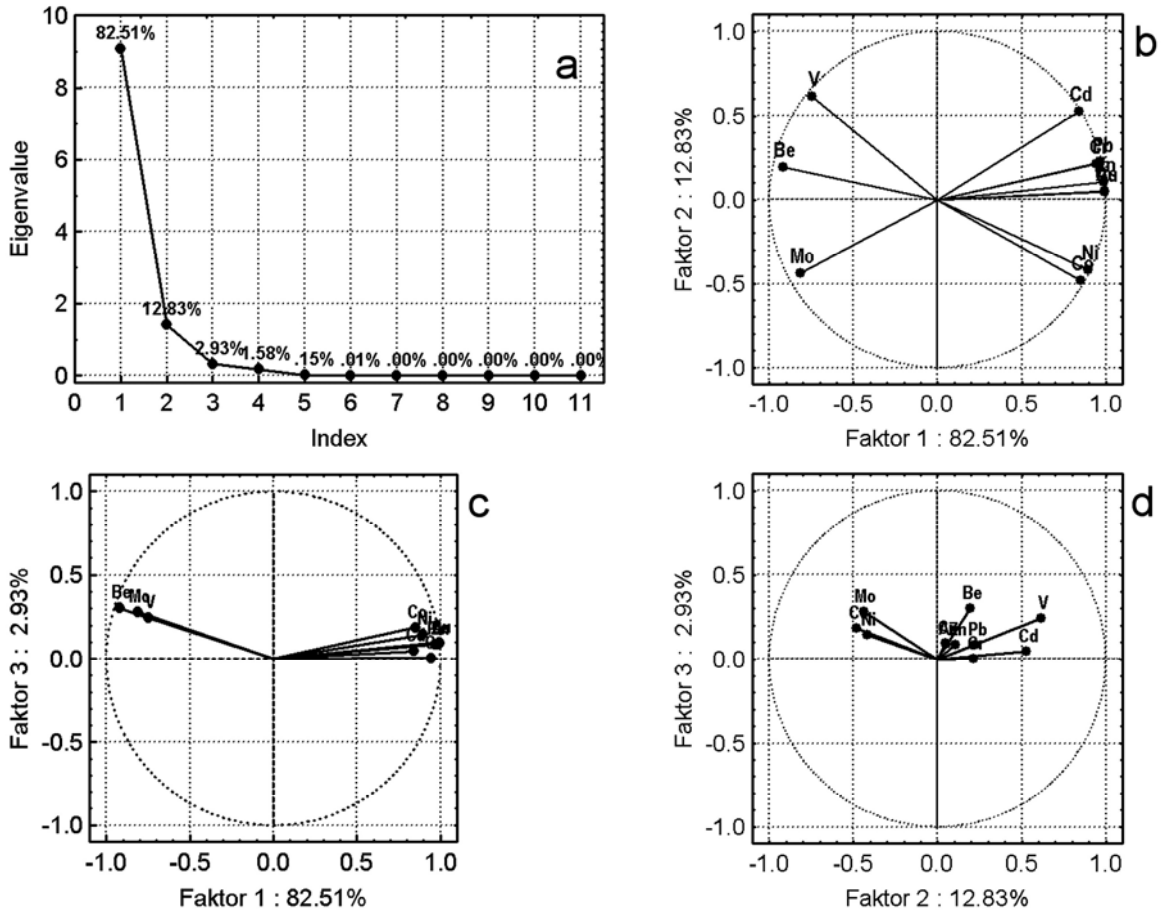
Cílem vícerozměrné klasifikace čili třídění je nalezení vnitřních vztahů mezi znaky, představovanými zde koncentracemi 11 prvků. Zdrojovou matici 80 vzorků půd je třeba roztřídit do určitého počtu shluků při respektování 11 prvků, které jsou v půdách obsaženy a stanoveny. Ke klasifikaci se využije metoda hlavních komponent, faktorová analýza a metoda shluků. Dle závěrů v dendrogramu nalezených shluků je potom aplikována i diskriminační analýza.

Tabulka 1. Přehled popisných statistik a ověření předpokladů o výběru 11 prvků Cd, Pb, Ni, Cu, Co, Cr, Zn, As, Be, V, Mo včetně klasických a robustních odhadů polohy, rozptýlení a tvaru ze 80 vzorků půd. Je aplikována mocninná a Box-Coxova transformace dat. [a]Maximalní povolená koncentrace normou je uvedena. Hodnoty koncentrace jsou v mg . kg^{-1}

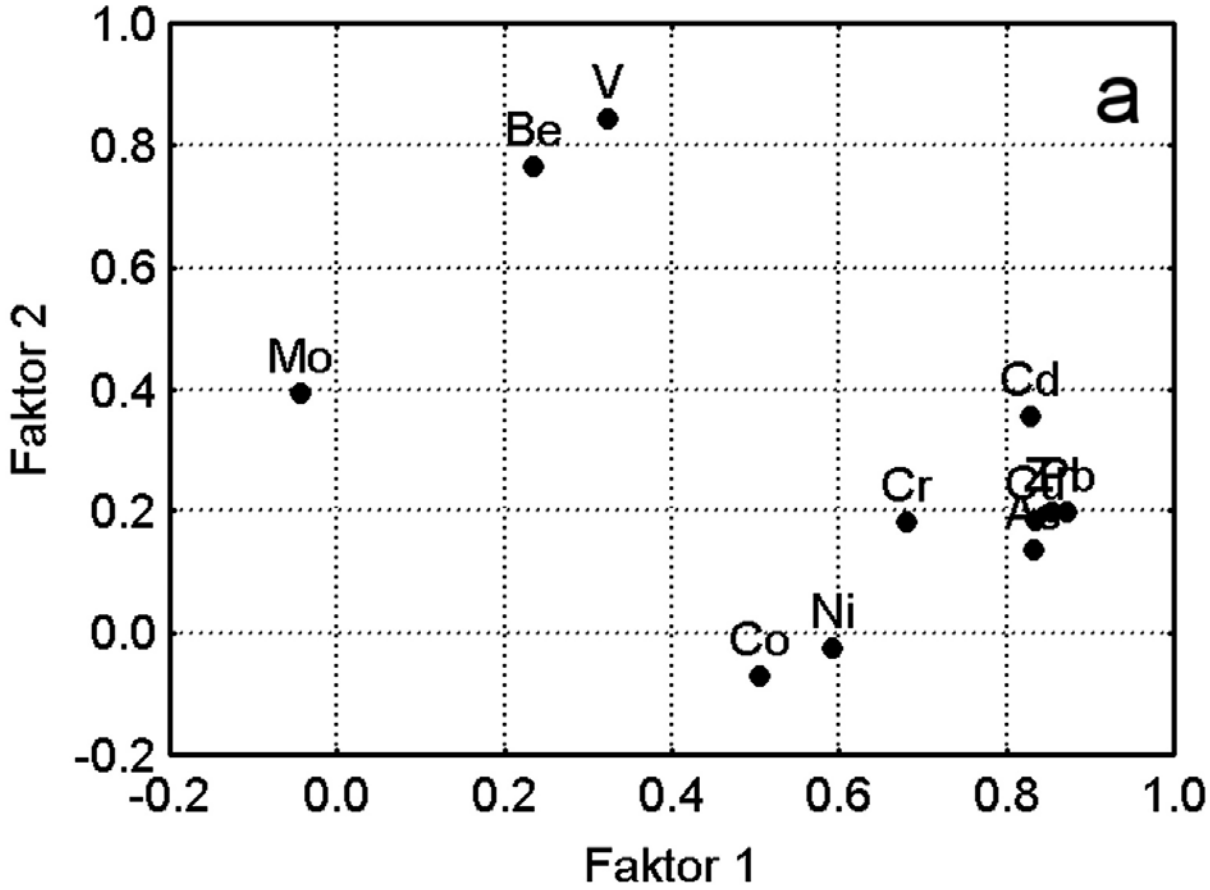
Odhady	**Zn**	**Pb**	**Cu**	**Cr**	**V**	**Ni**	**As**	**Co**	**Cd**	**Be**	**Mo**
					Povolená koncentrace a mez detekce LOD						
Povolená koncentrace[a]	200.0	140.0	100.0	200.0	220.0	80.0	30.0	50.0	1.0	7.0	5.0
Mez detekce LOD	2	0.9	0.7	0.9	0.3	0.9	0.07	0.09	0.07	0.02	0.1
					Klasické odhady polohy, rozptýlení a tvaru						
Minimum x_1	68.0	14.0	17.7	24.9	0.50	17.0	5.33	5.67	0.20	0.31	0.10
Maximum x_n	7130.0	1950.0	885.0	200.0	68.50	187.0	61.20	17.70	39.7	2.98	9.80
Průměr $\bar{x}$	464.68	133.26	66.81	42.32	31.54	30.34	11.92	9.17	2.40	1.11	1.41
Směrodatná odchylka s	995.14	292.78	119.48	27.89	13.74	19.90	8.42	2.23	5.65	0.47	2.13
Šikmost g_1	4.94	4.42	5.19	4.72	0.05	6.43	3.99	1.37	5.53	1.41	2.61
Kurtosis g_2	29.44	23.75	32.49	26.50	3.60	49.78	20.98	6.27	34.13	5.78	8.86
					Robustní odhady polohy						
Median $\tilde{x}_{0.5}$	210.5	51.0	37.4	35.1	32.45	28.0	10.05	8.67	0.97	1.05	0.70
Uřezaný průměr $\bar{x}$ (10%), n = 64	233.78	60.0	39.2	36.87	31.88	27.3	10.20	8.95	1.25	1.09	0.86
Uřezaný průměr $\bar{x}$ (25%), n = 40	210.10	52.1	37.1	35.94	31.86	27.7	10.05	8.91	1.01	1.08	0.69
Uřezaný průměr $\bar{x}$ (45%), n = 8	210.3	50.8	37.3	35.30	32.35	28.0	10.09	8.79	0.97	1.09	0.68
					Rozdělení výběru						
Rozdělení, Korelace r v rankitovém grafu	Lognorm. 0.9057	Lognorm. 0.9182	Lognorm. 0.9059	Lognorm. 0.9149	Laplac. 0.9872	Lognorm 0.8783	Lognorm. 0.9661	Exp. 0.9728	Lognorm. 0.8696	Exp., 0.9825	Lognorm. 0.9587
					Test homogenity						
Určených outlierů	9	9	10	4	1	4	5	2	11	2	9
					Jarque-Berra test normality, kritická hodnota pro α = 0.05 činí $\chi^2_{0.95}(2) = 5.99$						
Testační kritérium C_1	3171.9	2020.7	3895.8	2550.1	1.77	9405.2	1536.3	71.8	4347.7	61.0	255.0
Normalita je	Zamítn.	Zamítn.	Zamítn.	Zamítn.	Přijato	Zamítn.	Zamítn.	Zamítn.	Zamítn.	Zamítn.	Zamítn.
					Transformace dat						
Mocn. nebo Box-Cox, exponent λ	0.00	-0.67	-0.93	-2.67	0.93	-1.20	-1.07	-0.40	-0.53	0.13	-0.13
Retransformovaný průměr $\bar{x}_R$	250.6	53.53	37.20	34.24	31.3	26.57	9.75	8.84	1.05	1.04	0.61

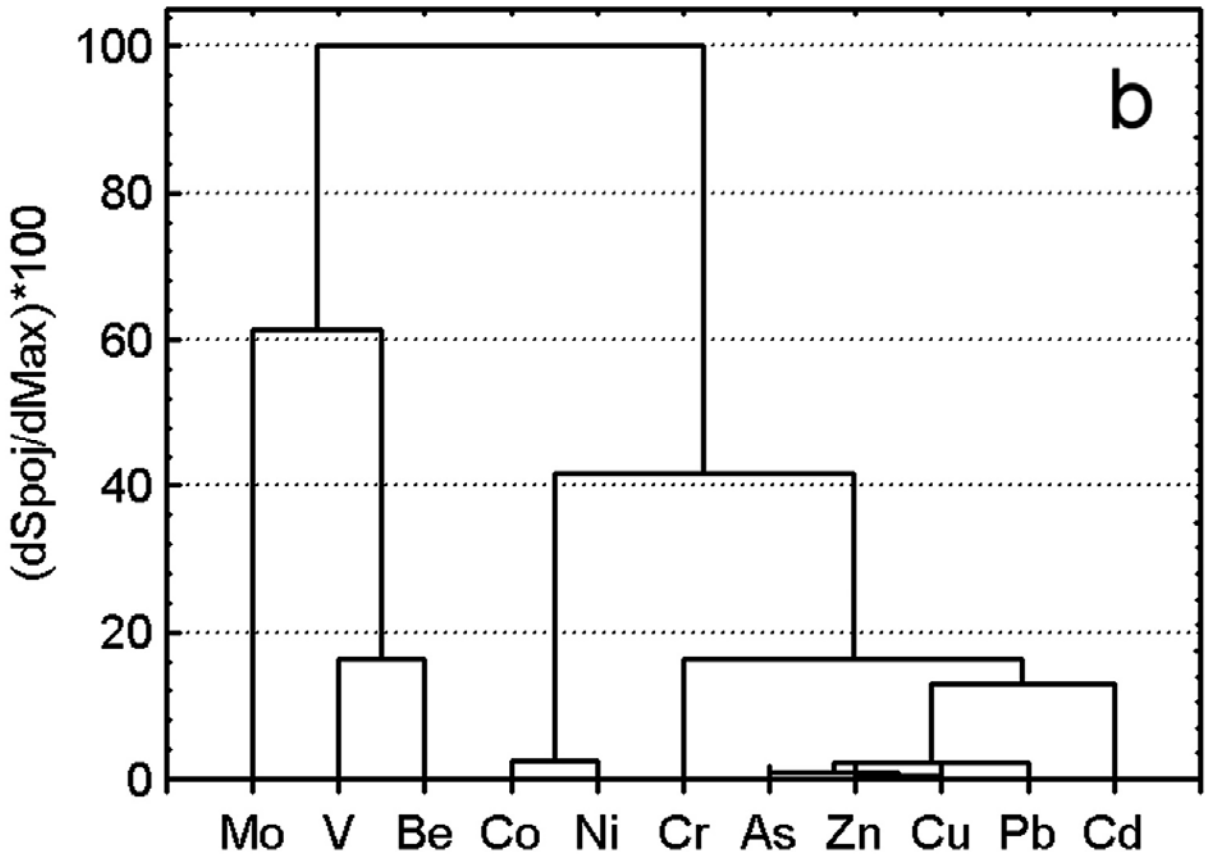


Krabicový graf všech 11 proměnných znaků čili elementů v půdě ukazuje na dostatečnou proměnlivost znaků v analyzované zdrojové matici.

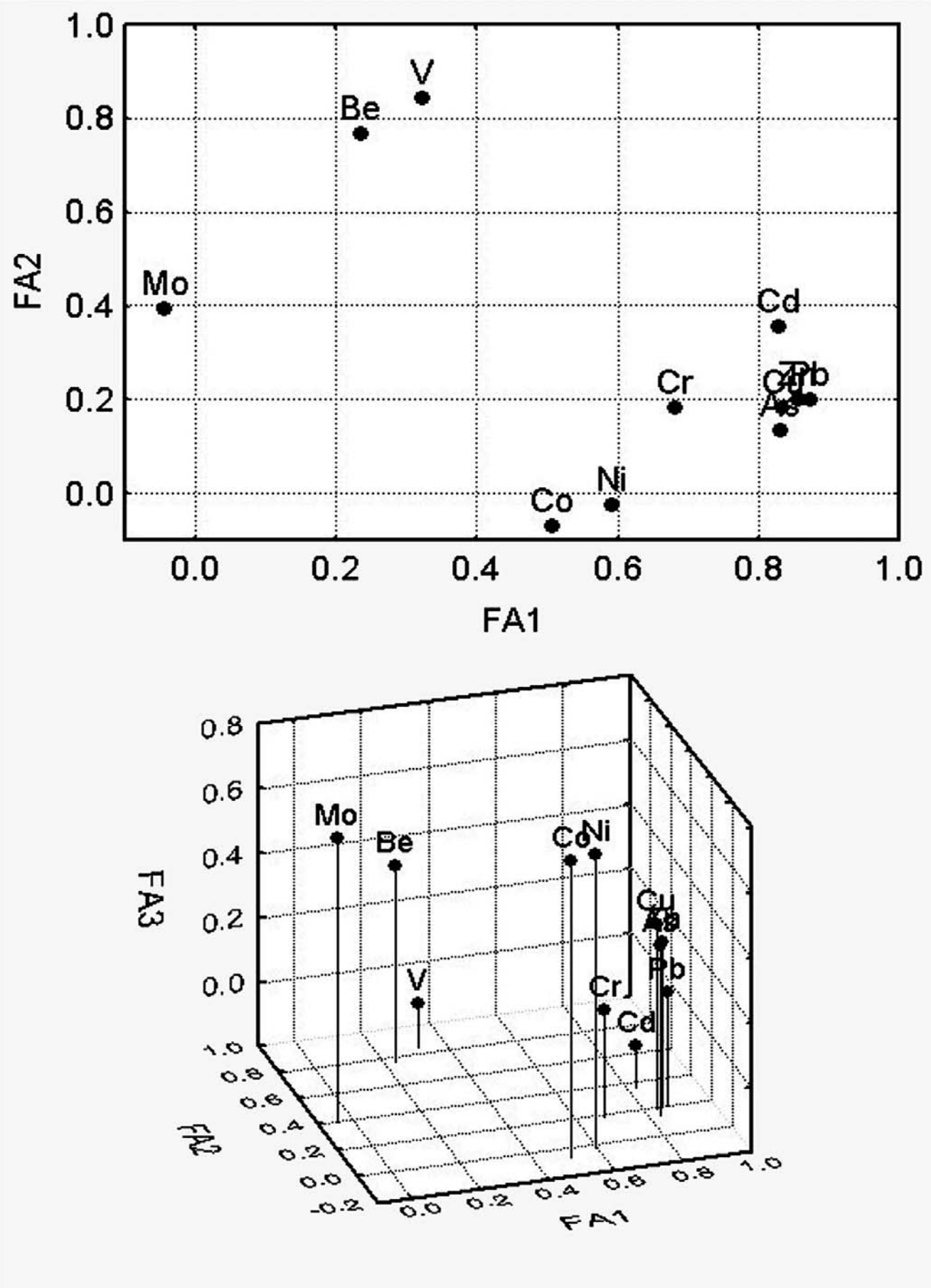


(a) Cattelův indexový graf vlastních čísel indikuje 2 až 3 hlavní komponenty k následnému využití. Graf komponentních vah dat po transformaci ukazuje na korelaci znaků v případě (b) PC1-PC2, (c) PC1-PC3, (d) PC2-PC3.

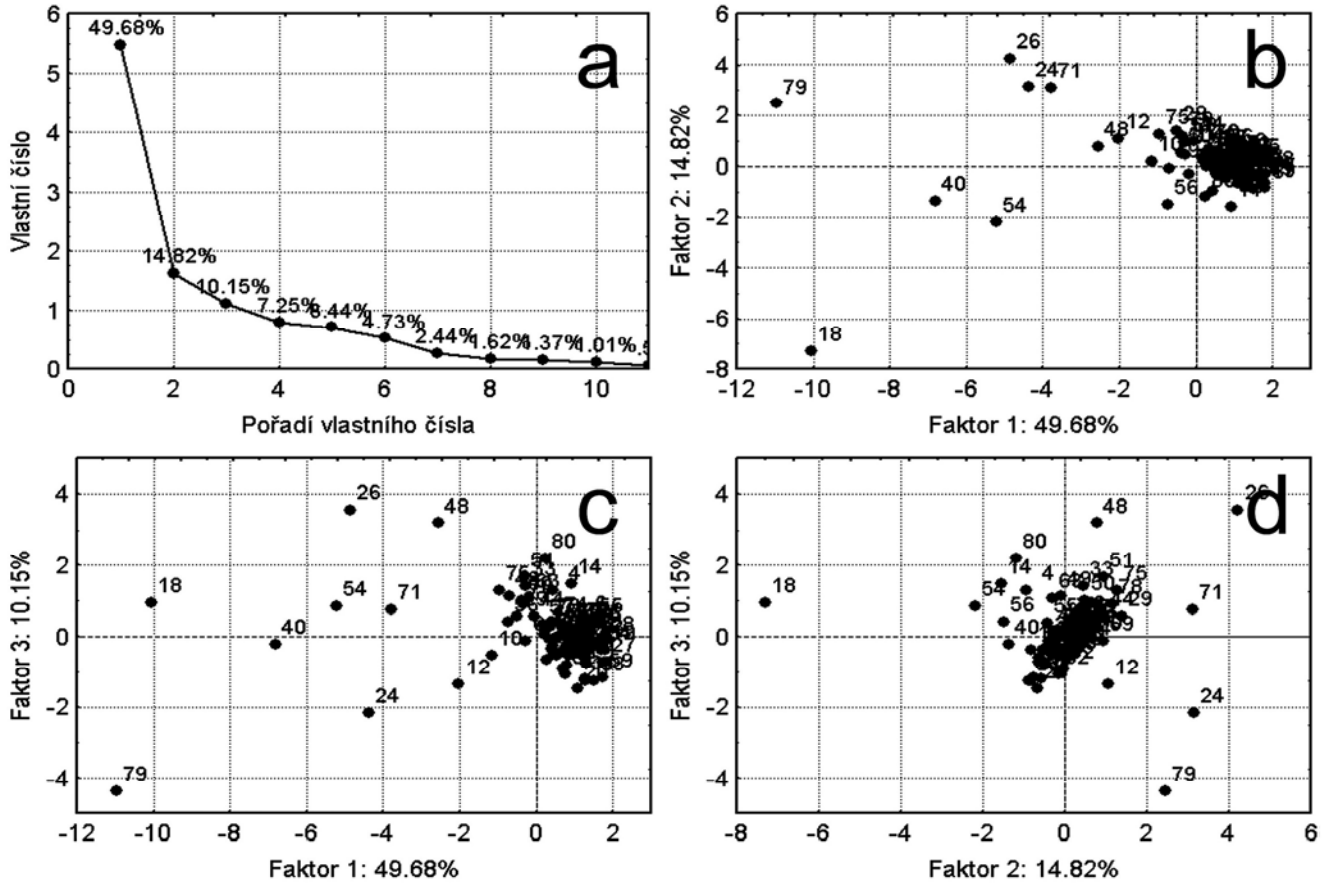




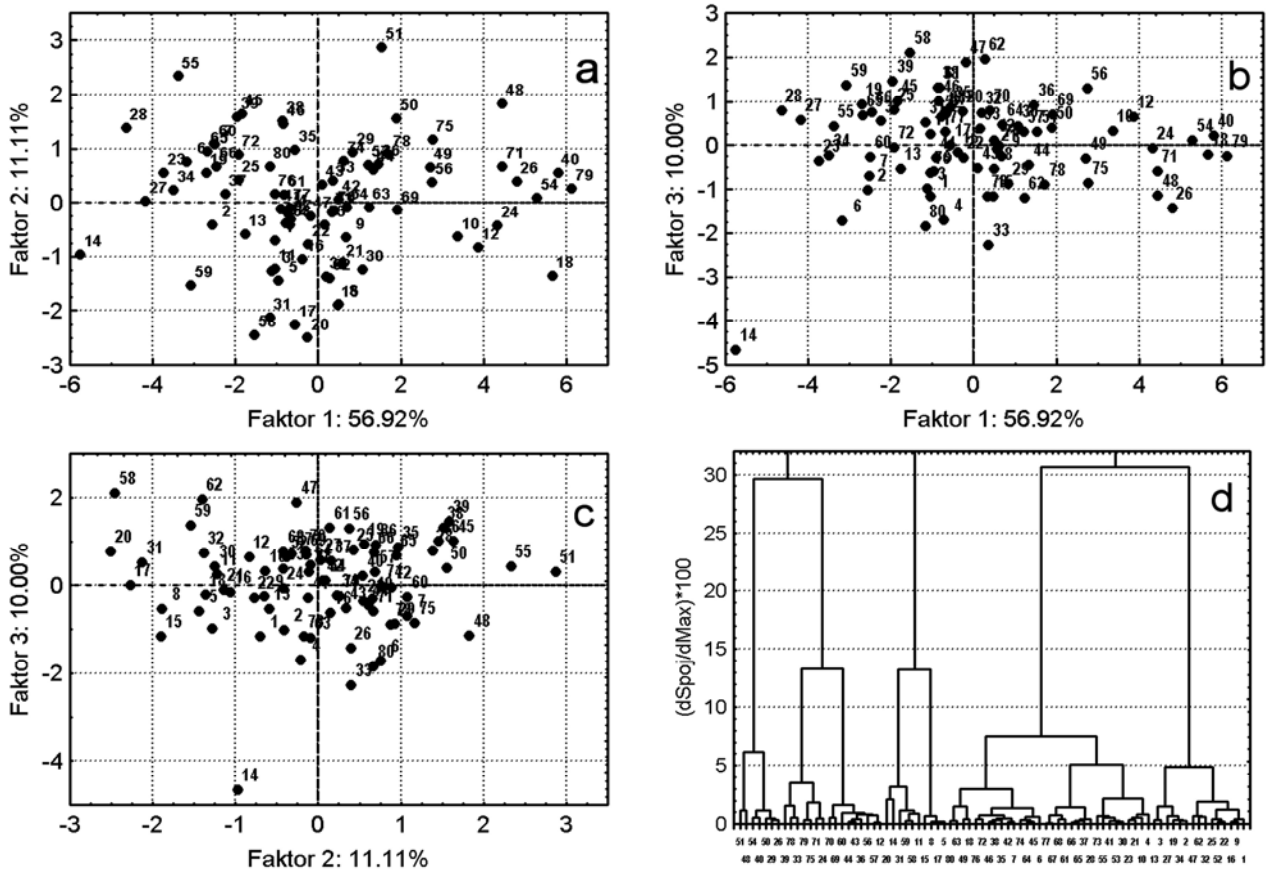
Porovnání grafu faktorových vah u dat po Box-Coxově transformaci s dendrogramem aglomerativní hierarchickou metodou váženého průměru u týchž dat.



2D- a 3D-projekce grafu faktorových vah po rotaci Varimax u předem transformovaných dat. Z grafů je možné indikovat shluky podobných stanovovaných prvků v 80 zeminách zdrojové matice dat.

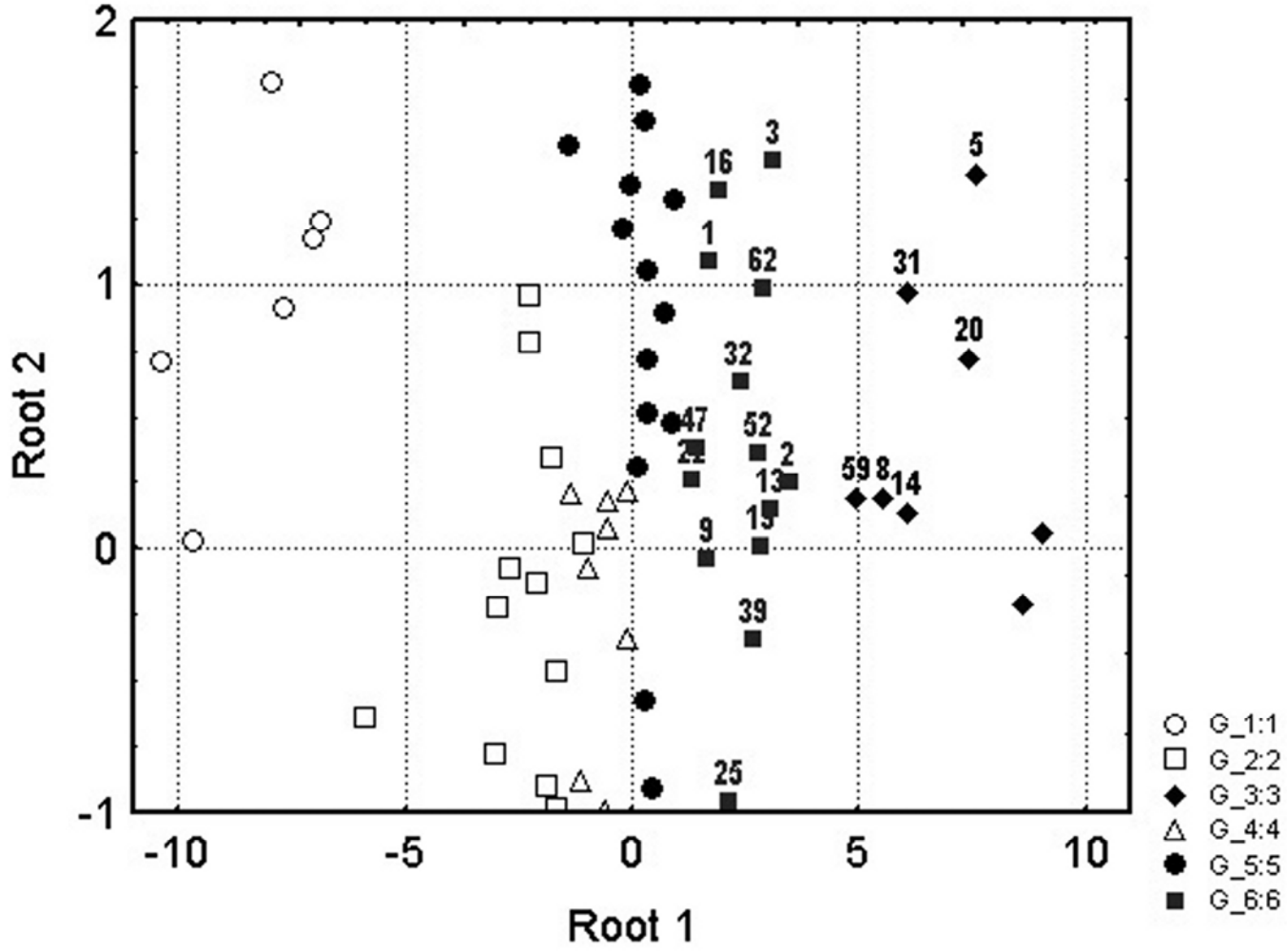


(a) Cattelův indexový graf vlastních čísel určuje 3 využitelné faktory, (b) až (d) jsou grafy faktorového skóre FA bez rotace pro transformovaná data.



(a) až (c) jsou grafy faktorového skóre vzorků půd, (d) dendrogram podobných objektů aglomerativní hierarchickou metodou váženého průměru u týchž dat.

.



Graf klasifikačního skóre objektů po CDF diskriminaci vykazuje 6 tříd rozdklasifikovaných objektů.

Závěr: Grafy klasifikačního skóre roztřídily objekty do 5 až 6 tříd při respektování 11 prvků. Dosažené závěry této klasifikace jsou ve shodě i s ostatními metodami vícerozměrné analýzy PCA, FA a CA.

Poděkování: Autoři vyslovují svůj dík za finační podporu vědeckého záměru č. MSM0021627502.

Doporučená literatura

[1] Siotani M., Hayakawa T., Fujikoshi Y.: *Modern Multivariate Statistical Analysis*, A Graduate Course and Handbook. American Science Press, Columbia 1985.

[2] Kendall M. G., Stuart A.: *The Advanced Theory of Statistics*, Vol. III. New York 1966.

[3] James W., Stein C.: *Estimation with Quadratic Loss*, Proceed. 4th Berkeley Symp. on Math. Statist., p. 361, 1961.

[4] Guanadeskian R., Kettenring J. R.: *Biometrics* **28**, 80 (1972).

[5] Campbell N. A.: *Appl. Statist.*, 29, 231 (1980).

[6] Hu J., Skrabal P., Zollinger H.: *Dyes and Pigments*, **8**, 189 (1987).

[7] Chambers J. M., Cleveland W. S., Kleiner B., Tukey P. A.: *Graphical Methods for Data Analysis*. Duxburg Press, Belmont, California 1983.

[8] Barnett V., (Edit.): *Interpreting Multivariate Data*. Wiley, Chichester 1981, kap. 6.

[9] Jolliffe I. T.: *Principal Component Analysis*. Springer Verlag, New York 1986.

[10] Barnett V., (Edit.): *Interpreting Multivariate Data*. Wiley, Chichester 1981, kap. 12.

[11] Everitt B. S.: *Graphical Techniques for Multivariate Data*. London 1978.

[12] Andrews D. F.: *Biometrics*, **28**, 125 (1972).

[13] Kulkarni S. R., Paranjape S. R.: *Commun. Statist.*, **13**, 2511 (1984).

[14] Guanadeskian R.: *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, New York 1977.

[15] Kleiner B., Hartigan J. A., *J. Amer. Statist. Assoc.*, **76**, 260 (1981).

[16] Kres H.: *Statistical Tables for Multivariate Analysis*. Springer, New York 1983.

[17] Seber G. A. F.: *Multivariate Observations*. Wiley, New York 1984.

[18] Stryjewska E., Rubel S., Henrion A., Henrion G.: *Z. Anal. Chem.*, **327**, 679 (1987).

[19] Mudholkar G. S., Trivedi M. S., Lin T. C.: *Technometrics*, **24**, 139 (1982).

[20] Johnson R.A., Wichern D.W.: *Applied Multivariate Statistical Analysis,* Prentice Hall, 1982

[21] Ajvjazin S., Bežajeva Z., Staroverov O.: *Metody vícerozměrné analýzy,* SNTL Praha 1981

[22] Meloun M., Militký J. , Forina M.: *Chemometrics for Analytical Chemistry, Volume 1. PC-Aided Statistical Data Analysis,* Ellis Horwood, Chichester 1992.

[23] Brereton R. G. *Multivariate Pattern Recognition in Chemometrics, Illustrated by Case Studies*, Elsevier 1992,

[24] Krzanowski W. J.: *Principles of Multivariate Analysis, A User's Perspective*, Oxford Science Publications 1988,

[25] Jeffers J. N. R., *Applied Statistician*, **16**, 225 (1967).

[26] Meloun M. , Militký J., *Statistické zpracování experimentálních dat*, Plus Praha 1994, Academia Praha 2004.

[27] Martens H., Naes T., *Multivariate calibration*, Wiley (1989) Chichester.

[28] Thomas E. V., *Anal. Chem.*, **66** (1994) 795A-804A.

[29] Malinowski F., Howery D., *Factor Analysis in Chemistry*, Wiley (1980) New York.

[30] Meloun M. , Militký J., *Kompendium statistického zpracování experimentálních dat*, Academia Praha 2002, Academia Praha 2006.