

ZPRACOVÁNÍ DAT VE STOPOVÉ ANALÝZE

Jiří Militký Katedra textilních materiálů, Textilní fakulta, Technická universita v Liberci, Liberec, e- mail jiri.militky@vslib.cz

Milan Meloun, Katedra analytické chemie, Universita Pardubice, Pardubice

1. Úvod

Stopová analýza má zvláštní postavení mezi analytickými postupy s ohledem na koncentrační rozmezí analyzovaných látek. V řadě případů se rozmezí analyzovaných látek pohybuje v několika řádech, což omezuje použití standardních statistických metod založených na předpokladu konstantního rozptylu resp. **aditivního modelu měření**. [1] .

Pro tuto situaci se může použít transformace stabilizující rozptyl nebo **multiplikativní model měření**. To vede k logaritmické transformaci dat [1]. Nevýhodou této transformace je fakt, že při nízkých koncentracích je absolutní chyba měření velmi malá (blízká 0), což odporuje realitě. V práci [2] je navržen postup kombinující oba modely měření a odstraňující jejich nevýhody. Tento postup však vyžaduje poměrně složité výpočty a není pro praktické měření příliš použitelný.

V tomto příspěvku jsou nejdříve popsány základní modely měření a jejich základní výhody resp. nevýhody.

Druhá část je věnována přehledu metod standardní statistické analýzy dat pro aditivní (a pro logaritmické transformace multiplikativní) model měření.

2. Modely měření

Omezte se na nejjednodušší úkoly stanovení koncentrace analytu. Účelem je odhad parametru polohy a stanovení jeho neurčitosti. Standardní model měření je **aditivní**, t.j.

$$x = \mu + \varepsilon \quad (1)$$

kde μ je skutečná hodnota měřené veličiny (koncentrace analytu) a ε je náhodná chyba měření. Standardní statistická analýza vychází z těchto předpokladů:

- střední hodnota chyb měření je nulová, t.j. $E(\varepsilon) = 0$,
- rozptyl chyb měření je konstantní, t.j. $D(\varepsilon) = \sigma^2$
- chyby jsou vzájemně nezávislé .t.j. $E(\varepsilon_i * \varepsilon_j) = 0$
- chyby mají normální rozdělení t.j. $\varepsilon \approx N(0, \sigma^2)$

Bez újmy na obecnosti lze přijmout první předpoklad . V případě , kdy $E(\varepsilon) = k$, se tato konstanta vlastně přičte k μ a resultují systematicky vychýlené odhady střední hodnoty, protože vyjde, že $E(x) = \mu + k$.

Druhý předpoklad o konstantnosti rozptylu je více omezující. Je ekvivalentní požadavku konstantní aditivní chyby měřicího přístroje $\Delta \approx \sigma$. Pro aditivní model je tedy relativní chyba hyperbolicky klesající funkcí úrovně měřené veličiny

$$\delta = \Delta / x \quad (2)$$

Pro zachování vhodné relativní přesnosti měření je pak nezbytné volit minimální hranici x_p od které je relativní chyba maximálně rovna hraniční hodnotě δ_h . Platí, že

$$x_p = \Delta / x_h \quad (3)$$

Třetí předpoklad o nezávislosti měření je v řadě případů splněn. Potíže mohou nastat např. u kinetických experimentů, kde může dojít vlivem vzorkování na jedné soustavě ke vzniku autokorelace. Nejjednodušší je model autokorelace I.řádu, kdy

$$\varepsilon_i = \rho * \varepsilon_{i-1} + u_i \quad (4)$$

Zde u_i je náhodná veličina s konstantním rozptylem Platí, že $u_0 = 0$. Autokorelační koeficient ρ je korelační koeficient mezi dvojicemi x_j a x_{j+1} , $i = 1, \dots, M - 1$. Lze určit, že

$$D(u) = D(x) = \frac{\sigma^2}{1 - \rho^2} \quad (5)$$

Je patrné, že v případě výrazné autokorelace dojde ke zvýšení rozptylu..Autokorelační koeficient ρ se obvykle odhaduje pomocí vztahu

$$\hat{\rho} = \frac{\sum_{j=1}^{N-1} (x_j - \bar{x}) * (x_{j+1} - \bar{x})}{[s^2(N-1)]} \quad (6)$$

kde s^2 je výběrový rozptyl a \bar{x} je aritmetický průměr.. Orientačně platí, že pokud leží $\hat{\rho}$ v intervalu

$$-2 / \sqrt{M} \leq \hat{\rho} \leq 2 / \sqrt{M}$$

lze považovat ρ za nevýznamný. Přesnější testy významnosti jsou uvedeny v [1].

Pro posouzení autokorelace lze použít také modifikovaný von Neumanův poměr

$$V^2 = \frac{\sum_{j=1}^{M-1} (x_j - x_{j+1})^2}{\sum_{j=1}^M (x_j - \bar{x})^2} \quad (7)$$

Lze ukázat, že pro velká M má veličina V^2 přibližně normální rozdělení s parametry

$$E(V^2) = 1 \quad D(V^2) = \frac{M - 2}{M^2 - 1} \approx \frac{1}{M + 2}$$

Orientačně platí, že pokud leží V^2 mimo interval $1 \pm 2 / \sqrt{(M + 2)}$, nejsou výběry nezávislé.

Pokud vyjde V^2 menší než spodní mez tohoto intervalu, projevuje se v datech trend,

nebo pomalé cyklické změny. Pokud vyjde V^2 větší než horní mez tohoto intervalu, projevují se v datech rychlé cyklické změny.

Čtvrtým, nejvíce restriktivní, je předpoklad, že chyby mají normální rozdělení. Tento předpoklad je potřebný pro konstrukci intervalů spolehlivosti (neurčitosti výsledků měření) resp. testování hypotéz. Pokud je k dispozici dostatek dat, lze odhadnout rozdělení chyb \mathcal{E} z rozdělení x měření, protože pro model (1) je tvar hustoty pravděpodobnosti totožný.

Normální rozdělení lze chápat jako jednoho z členů třídy eliptických symetrických rozdělení, pro které platí že se liší pouze délkou konců. Ve stopové analýze je však častějším jevem asymetrické rozdělení dat zešikmené k vyšším hodnotám. Důvodem je zejména frekventovaný výskyt hodnot těsně nad limitou citlivosti měřících přístrojů. Pro odstranění této asymetrie se často používá vhodná transformace $h(x)$. Ta však v případě platnosti modelu (1) vede ke vzniku nekonstantního rozptylu

$$D(h(x)) = \left[\frac{dh(x)}{dx} \right]^2 * \sigma^2 \quad (8)$$

Např. pro běžně doporučovanou logaritmickou transformaci $h(x) = \ln(x)$ vyjde

$$D(h(x)) = \left(\frac{\sigma}{x} \right)^2 = \delta^2 \quad (9)$$

To znamená, že místo konstantní absolutní chyby je v této transformaci konstantní relativní chyba (variační koeficient), což odporuje přijatému modelu měření. Korektní analýza zde vyžaduje přímé použití zešikmeného rozdělení a konstrukci nesymetrických intervalů spolehlivosti.

Multiplikativní model měření je založen na předpokladech konstantní relativní chyby a nezápornosti měření (jde o fyzikální veličiny související s hmotou). Výsledek měření je modelován vztahem

$$x = \mu * \exp(\varepsilon) \quad (10)$$

Zde ε má stejné vlastnosti jako u modelu aditivního (rov.(1)). Po korektní logaritmické transformaci přechází tento model na aditivní model v logaritmech, tedy

$$\ln(x) = \ln(\mu) + \varepsilon \quad (11)$$

Nevýhodou multiplikativního modelu je především to, že pro velmi nízké koncentrace resp. malé μ vychází absolutní chyba měření příliš nízká [4]. V případě, že se použije logaritmická transformace na model (1) vyjde

$$\ln(x) = \ln(\mu + \varepsilon) = \ln \mu + \ln(1 + \varepsilon / \mu) \quad (12)$$

S využitím Taylorova rozvoje lze psát

$$\ln(x) \approx \ln(\mu + \varepsilon) = \ln \mu + \varepsilon / \mu - 0.5 * (\varepsilon / \mu)^2 + 0.33 * (\varepsilon / \mu)^3 - \dots (13)$$

Pro malé relativní chyby měření $\delta = \sigma / \mu$ lze pak s využitím tohoto vztahu nalézt výrazy pro střední hodnotu a rozptyl $\ln(x)$ ve tvaru

$$E(\ln x) = \ln \mu - 0.5 * \delta^2 - 0.75 * \delta^4 \quad (14)$$

a

$$D(\ln x) = \delta^2 + 2.5 * \delta^4 + 4.66 * \delta^6 \quad (15)$$

Je tedy patrné, že použití nesprávného předpokladu ovlivní jak střední hodnotu tak i rozptyl.

Aditivní i multiplikativní model lze vyjádřit jako speciální případy mocninné třídy modelů měření, která je charakterizována tím, že transformací obou stran pomocí funkce $h(\cdot)$ vyjde aditivní model

$$h(x) = h(\mu) + \varepsilon \quad (16)$$

S výhodou se jako funkce $h(\cdot)$ používá Box Coxova třída polynomických transformací ve tvaru

$$h(x) = \frac{(x-1)^\lambda}{\lambda} \quad \lambda \neq 0$$

$$h(x) = \ln(x) \quad \lambda = 0$$

kde λ je parametr transformace. Pro $\lambda = 1$ resultuje aditivní model měření a pro $\lambda = 0$ model multiplikativní. Lze ukázat, že vhodným odhadem parametru μ (neznámá koncentrace) je výběrový medián, který je invariantní vůči monotónní transformaci. S využitím Taylorova rozvoje lze vyjádřit rov. (16) ve tvaru

$$x \approx \mu + \varepsilon / \mu^{1-\lambda} \quad (17)$$

Pro případ, že rozptyl $D(\varepsilon) = \sigma^2$ je malý jde o aditivní model s nekonstantními chybami, pro který lze použít jako odhad μ vážený aritmetický průměr s vahami úměrnými $\mu^{-(1-\lambda)/2}$. Pro odhad parametru λ je možno použít metodu maximální věrohodnosti. Pokud je ε rozdělené v souladu s předpoklady aditivního modelu měření (normalita a nezávislost) má logaritmus věrohodnostní funkce tvar

$$\ln L(\lambda) = \sum (\lambda - 1) * \ln(x_i) - \frac{1}{2\sigma^2} \sum [h(x_i) - h(\mu)]^2 \quad (18)$$

Pro pevné λ lze určit maximálně věrohodný odhad rozptylu ve tvaru

$$\hat{\sigma}_c^2 = \frac{1}{N} \sum [h(x_i) - h(\mu)]^2 \quad (19)$$

kde se za $h(\mu)$ dosazuje aritmetický průměr transformovaných dat

$$h(\mu) \approx \frac{1}{N} \sum h(x_i) \quad (20)$$

Po dosazení do věrohodnostní funkce resultuje vztah

$$\ln L^*(\lambda) = \sum (\lambda - 1) * \ln(x_i) - \frac{N * \ln \hat{\sigma}_c^2}{2} \quad (21)$$

Maximalizací $\ln L^*(\lambda)$ podle λ (viz.[1]) lze pak snadno určit maximálně věrohodný odhad $\hat{\lambda}$ parametru transformace λ . Na základě asymptotického $(1 - \alpha)$ % ního intervalu spolehlivosti tohoto parametru lze sestavit nerovnost

$$\ln L(\lambda) \geq \ln L(\hat{\lambda}) - 0.5 * \chi_{1-\alpha}^2(I) \quad (22)$$

Všechna λ splňující tuto nerovnost leží v intervalu spolehlivosti a jsou tedy přijatelná. Toho lze snadno využít pro rozlišení mezi aditivním a multiplikativním modelem měření. V rovnici (22) označuje $\chi_{1-\alpha}^2(I)$ kvantil chí kvadrát rozdělení s 1 stupněm volnosti.

Platí, že:

- pokud obsahuje 95% ní interval spolehlivosti také jedničku, volí se aditivní model.
- pokud obsahuje 95% ní interval spolehlivosti nulu a nikoliv jedničku, volí se multiplikativní model.
- v ostatních případech je možné zvolit obecný model ve tvaru $\ln(x) = \ln(\mu) + \varepsilon_1 + \ln\left(1 + \frac{\varepsilon_2}{\mu * \exp(\varepsilon_1)}\right)$ a použít pro další analýzu postup navržený v [1].

Jak je patrné, lze s výhodou použít postupu mocninné transformace včetně příslušného software pro rozlišení mezi dvěma základními modely měření a navíc je k dispozici celý postup statistické analýzy včetně zpětné transformace.

Pro odstranění nevýhod aditivního modelu měření (konstantní rozptyl v celém rozmezí dat a multiplikativního modelu (neúnosně malá absolutní chyba v okolí nuly) je v práci [2] navržen **kombinovaný model měření** ve tvaru

$$x = \mu * \exp(\varepsilon_1) + \varepsilon_2 \quad (23)$$

kde se předpokládá, že obě chyby mají normální rozdělení $\varepsilon_1 \approx N(0, \sigma_1^2)$ a $\varepsilon_2 \approx N(0, \sigma_2^2)$. Pokud se týkají chyby měření především měřicího přístroje je možné použít také zjednodušující předpoklad, že jsou oba rozptyly stejné t.j. $\sigma_1^2 \approx \sigma_2^2$. Tento model tedy pro větší úroveň měřené veličiny (koncentrace) používá předpoklad konstantní relativní chyby a pro nízké úroveň měřené veličiny předpoklad konstantní absolutní chyby. V oblasti malých koncentrací vychází tedy relativní chyba větší. To bylo experimentálně potvrzeno pro případ sledování úroveň pesticidů v rozmezí 1 ppb až 100 jednotek (Horowitzova trumpetka) [2]. Pro odhad obou rozptylů je v práci [2] navržena metoda maximální věrohodnosti. S využitím rov (12) lze vyjádřit model (23) v logaritmické transformaci

$$\ln x = \ln \mu + \varepsilon_1 + \ln \left[1 + \frac{\varepsilon_2}{\mu * \exp(\varepsilon_1)} \right] \quad (24)$$

Třetí člen rov. (24) se dá rozvinout do Taylorovy řady (viz rov.(14)). Při omezení na první sčítanec pak vyjde

$$\ln x \approx \ln \mu + \varepsilon_1 + \left[\frac{\varepsilon_2}{\mu * \exp(\varepsilon_1)} \right] \quad (25)$$

Tento vztah ukazuje, že:

- pro hodně malé ε_1 vyjde aditivní model v logaritmech $\ln x \approx \ln \mu + \left[\frac{\varepsilon_2}{\mu} \right]$ s nekonstantním rozptylem rovným $\delta = \sigma_2 / \mu$.
- pro hodně malé ε_2 , resp. pro srovnatelné ε_1 a ε_2 a velké μ vyjde aditivní model v logaritmech $\ln x \approx \ln \mu + \varepsilon_1$ s konstantním rozptylem rovným σ_1^2

Rozptyl měření je možné určit ve tvaru

$$D(x) = \sqrt{\sigma_1^2 + \mu^2 * \exp(\sigma_2^2) * (\exp(\sigma_2^2) - 1)} \quad (26)$$

Pro nízké úrovně x, kde dominuje druhý člen rov. (23) je možno použít předpoklad normality a přibližný 95 % ní interval spolehlivosti je $x \pm 1.96 * \sqrt{D(x)}$.

Pro nízké úrovně x, kde dominuje první člen rov. (23) je možno použít předpoklad logaritmicko normálního rozdělení a přibližný 95 % ní interval spolehlivosti je

$$\exp(\ln(x)) + 1.96 * \sigma_2, \quad \exp(\ln(x)) - 1.96 * \sigma_2.$$

Podrobnosti o určení odhadů obou rozptylů lze nalézt v práci [2].

3. Zpracování dat

Dále uvedené techniky jsou vhodné především pro **aditivní model měření** a po logaritmické transformaci též pro **multiplikativní model**. Používají se často obecně i pro základní charakterizaci dat při rutinní analýze. Jsou uvedeny pouze vybrané jednodušší techniky zaměřené na odhady parametrů polohy a variability. Řada komplexnějších postupů je popsána v práci [1].

3.1 Nekategorizovaná data

Nechť $\{x_i\}$ $i = 1, \dots, N$ je náhodný výběr

- složený z nezávislých prvků
- homogenní
- normálně rozdělený (*není u všech postupů nezbytné*)

Ověření těchto předpokladů (viz [1])

Normalita: Q - Q grafy, vynášejí se $x_{(i)}$ proti u_{Pi} , $Pi = i/(N+1)$

Vybočující hodnoty: metoda barier (viz [1]). Jednodušší je momentová metoda. Platí zde, že vybočující jsou takové x_j , pro které je

$$|\bar{x} * -x_j| \geq K_c . s *$$

kde

$$K_c = 1.55 + [0.8 \cdot \sqrt{g_2^* - 1} \cdot \log(N/10)]$$

\bar{x}^* , s^* odhadnutý střední hodnoty a směrodatné odchylky bez podezřelých bodů.
 g_2^* odhad špičatosti bez vybočujících bodů

$$g_2^* = \frac{N \cdot \sum (x_i - \bar{x})^4}{[\sum (x_i - \bar{x})^2]^2}$$

Postup je vhodný pro $N \geq 20$. Pro $N = 20$ vyjde pro různá rozdělení:

- Normální rozdělení $g_2 = 3$ $K_c = 1.89$
- Rovnoměrné rozdělení $g_2 = 1.8$ $K_c = 1.77$
- Laplaceovo rozdělení $g_2 = 6$ $K_c = 2.09$

ODHADY PARAMETRŮ

Polohy:

průměr \bar{x}

medián $\tilde{x}_{0.5}$

Rozptýlení:

rozptyl s^2

směrodatná odchylka s

variační koeficient $v = s / \bar{x} \cdot 100$

Přesnosti odhadů se charakterizují pomocí rozptylů:

Pro případ normálního rozdělení dat $x_i \sim N(\mu, \sigma^2)$ je $D(\bar{x}) = \frac{\sigma^2}{N}$

$$D(s^2) = \frac{2 \cdot \sigma^4}{N-1} \quad D(\tilde{x}_{0.5}) = \frac{\pi \cdot \sigma^2}{2 \cdot N} \quad D(s) = \frac{\sigma^2}{2 \cdot (N-1)}$$

$$D(v) = \delta^2 \cdot \left[\frac{N + \delta^2 \cdot (2 \cdot N + 1)}{2 \cdot N \cdot (N + 1)} \right], \text{ kde } \delta = \frac{\mu}{\sigma} \text{ je populační variační koeficient}$$

Klasická statistická analýza:

a) \bar{x} , s^2 a 100. $(1 - \alpha)$ % ní interval spolehlivosti střední hodnoty

$$\bar{x} - t_{1-\alpha/2}(N-1) \cdot \frac{s}{\sqrt{N}} \leq \mu \leq \bar{x} + t_{1-\alpha/2}(N-1) \cdot \frac{s}{\sqrt{N}}$$

b) Variační koeficient v [%] a interval spolehlivosti pro δ asymptoticky:

$$\frac{v}{100} - Z_{1-\alpha/2} \cdot \sqrt{D(v)} \leq \delta \leq \frac{v}{100} + Z_{1-\alpha/2} \cdot \sqrt{D(v)}$$

Robustní statistická analýza:

Medián $\tilde{x}_{0.5}$, robustní odhad rozptylu s_R^2 a interval spolehlivosti pro populační medián *Med*. Pro odhad směrodatné odchylky mediánu platí

$$s_R = \frac{x_{(N-k+1)} - x_{(k)}}{2 \cdot Z_{1-\alpha/2}} \quad \text{kde } k = \frac{N+1}{2} - Z_{1-\alpha/2} \cdot \sqrt{N/4}$$

Obyčejně se volí $\alpha = 0.05$ ($u_{1-\alpha/2} = 1.96$). Pro interval spolehlivosti je pak

$$\tilde{x}_{0.5} - t_{1-\alpha/2}(N-1) \cdot s_R \leq \text{Med} \leq \tilde{x}_{0.5} + t_{1-\alpha/2}(N-1) \cdot s_R$$

Extrémně malé výběry: (N ≤ 20)

! Vždy vysoká nejistota → velký vliv vybočujících měření. Vhodné metody lze nalézt v [1].

3.2 Kategorizovaná data

Vznikají tříděním číselných údajů do intervalů, které jsou třídami nového znaku.

Jednotlivým třídám přiřazujeme číselné hodnoty x_j (střed intervalu,...).

Diskrétní, kardinální, četností kategorizace, pseudokategorizace.

Příklad: Délka částic:

Třída	x_j	n_j	f_j	F_j
21 - 23	22	50	0.2	0.2
23 - 25	24	80	0.32	0.52
25 - 27	26	72	0.288	0.808
27 - 29	28	48	0.192	1

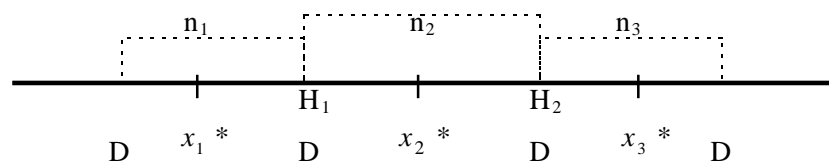
N = 250

Třídy:

- přirozené číselné vyjádření (počet vad, mikroorganismů atd.)
- sloučením údajů

$$A = \{a_1, a_2, \dots, a_{K'}\} \Rightarrow x = \{x_1, x_2, \dots, x_K\} \quad \text{kde } K < K'$$

Sloučení údajů:



$$x_i^* \quad \text{průměr} \quad x_i^* = D_i + (H_i - D_i) / 2 = (D_i + H_i) / 2$$

$$[D_i, H_i] \quad \text{třídní interval}$$

$$\Delta x_i = H_i - D_i \quad \text{délka třídy}$$

$$N = \sum_{i=1}^K n_i$$

Parametry: D_1 , K , Δx_i . Platí tato heuristická pravidla

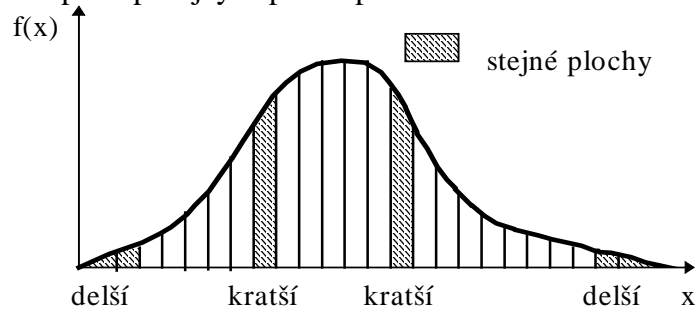
$$0 \leq k \cdot \Delta x - [x_{(N)} - x_{(1)}] \leq \Delta x \cdot D_1 + K \cdot \Delta x > x_{(N)}$$

$$0 < x_{(l)} - D_l \leq \Delta x$$

Volba K:

$N > 100$	$K = \text{int}[10 \cdot \log(N)]$
$40 \leq N \leq 100$	$K = \text{int}[2 \cdot \sqrt{N}]$
$N < 100$	$K = \text{int}[1 + 1.44 \cdot \ln(N)]$
Universálně	$K = \text{int}[2.64 \cdot (N - 1)^{0.4}]$

Nekonstatní délka tříd - princip stejných pravděpodobností.



Charakteristiky polohy:

Me . . . mediánová kategorie

a) Konstantní Δx

Medián $\tilde{x}_{0.5} = x_{Me} + \Delta x / 2 - \Delta x \cdot \frac{F_{Me} - 0.5}{l}$

b) obecně: $\tilde{x}_{0.5} = x_{Me} - \Delta x_{Me} \cdot \frac{F_{Me} - 0.5}{f_{0.5}}$

Aritmetický průměr:

$$\bar{x} = \sum_{i=1}^K f_i \cdot x_i^* = \frac{1}{N} \sum_{i=1}^K n_i \cdot x_i^*$$

1. Vlastnosti: $\bar{x} = x_l^* \rightarrow f_l = 1$

2. $\bar{x} = x_K^* \rightarrow f_K = 1$ Platí, že $\sum (x_i - A)^2 \rightarrow \min$ pro $A = \bar{x}$

Geometrický průměr: (kladná data \rightarrow velký rozsah)

$$G = \prod_{j=1}^K (x_j^*)^{f_j} = \exp \left[\sum_{i=1}^C f_i \cdot \ln(x_i^*) \right]$$

Charakteristiky rozptýlení:

Dorvar $Dor var = 2 \cdot \Delta x \cdot \sum_{i=1}^K F_i \cdot (1 - F_i)$

Rozptyl
$$s^2 = \sum_{i=1}^K f_i \cdot (x_i^* - \bar{x})^2 = \sum_{i=1}^K (f_i \cdot x_i^{*2} - \bar{x}^2)$$

Vlastnosti:

1. $s^2 = 0$ všechna data v jedné třídě $f_i = 1$
2. Maximálně $s_{\max}^2 = (x_K^* - x_1^*)^2 / 4 \Rightarrow f_1 = f_K = 0.5$
3. Čím větší s^2 tím více se data vzdalují od \bar{x} .

Směrodatná odchylka:

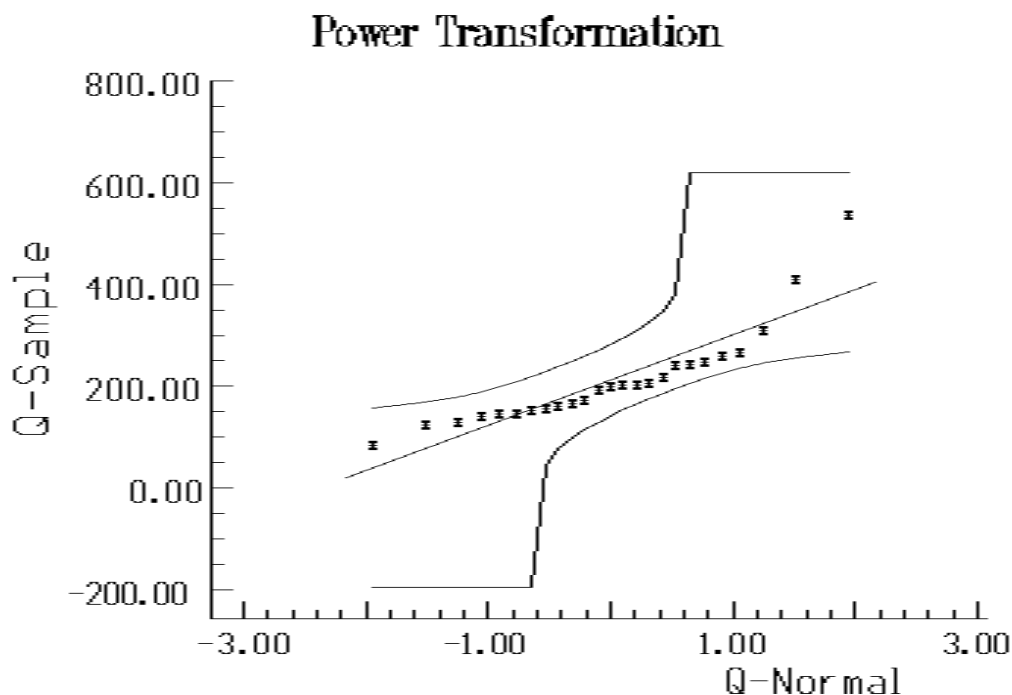
$$s = \sqrt{s^2} = \sqrt{\sum_{i=1}^K (f_i \cdot x_i^{*2} - \bar{x}^2)}$$

Konstrukce intervalů spolehlivosti a testy jsou stejné jako u nekategorizované proměnné.

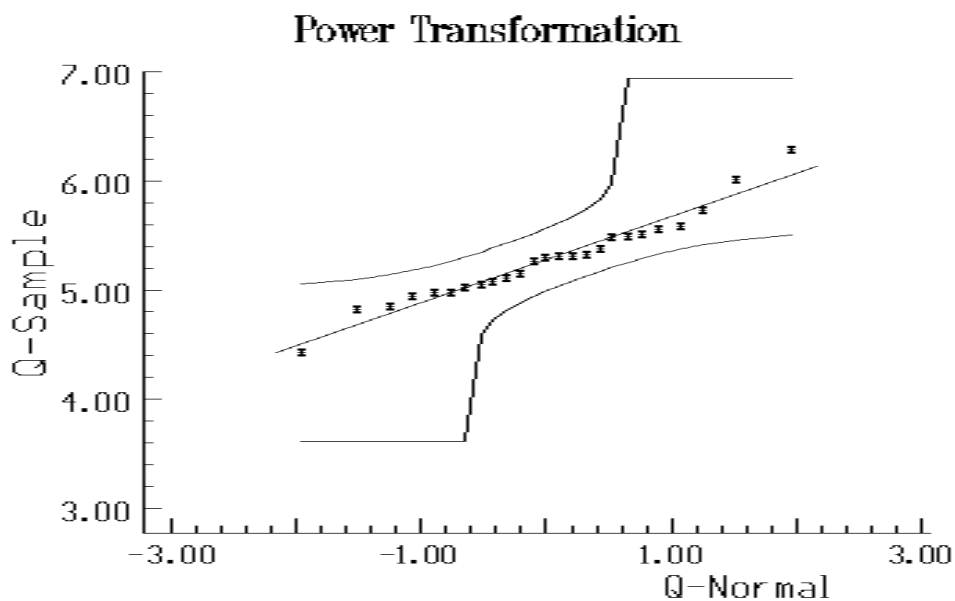
4. Příklad

Stopový obsah antimonu měděné rudě)

Byl stanoven obsah antimonu u N=25 vzorků měděné rudy a stanoven obsah antimonu v ppm. Na obr. 1 je rankitový graf pro původní data a na obr. 2 rankitový graf pro logaritmickou transformaci. Je patrné, že původní data mají nesymetrické rozdělení, které se blíží v logaritmické transformaci rozdělení normálnímu.

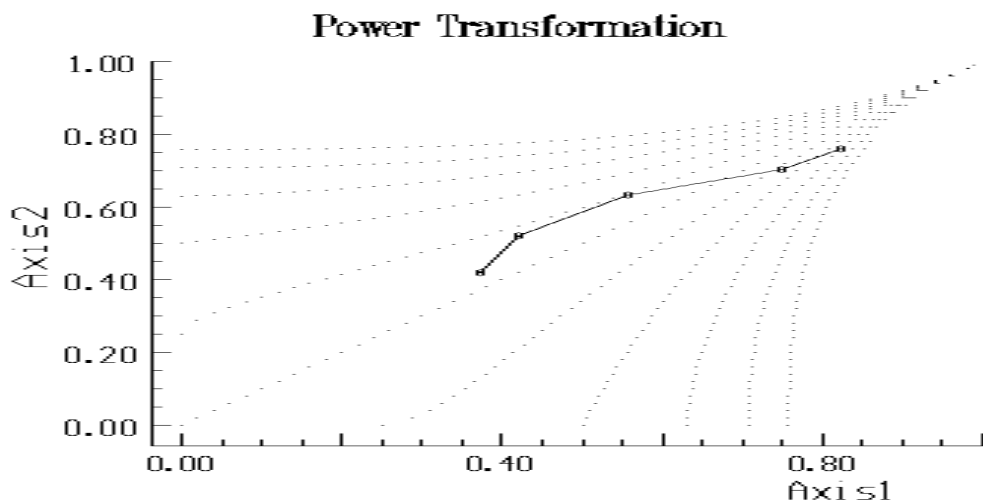


Obr.1. Rankitový graf pro původní data



Obr2. Rankitový graf pro data v logaritmické transformaci

Na obr. 3 je graf Hines Hinesové, který indikuje, že optimální je Transformace v okolí nuly.



Obr. 3 Graf Hines Hinesové

Optimální mocnina pro Box Coxovu transformaci vyšla -0.4. Odpovídající interval spolehlivosti obsahuje nulu ale nikoliv jedničku. Dále jsou shrnuty výsledky pro oba modely a pro robustní analýzu založenou na mediánu. Je patrné, že korektnější multiplikativní model vede k výsledkům blízkým robustní analýze založené na mediánu

Multiplikatívni model

Průměr 196.04
Směrodatná odchylka 76.925
95.0 % interval spolehlivosti:
Dolní mez: 166.73 Horní mez: 230.51

Aditivní model

Průměr 212.14
Směrodatná odchylka 95.90
95.0 % interval spolehlivosti:
Dolní mez: 172.55 Horní mez: 251.73

ROBUSTNÍ (Median):

Medián 199.6
Směrodatná odchylka 88.44
95.0 % interval spolehlivosti:
Dolní mez: 161.82 Horní mez: 237.38

5. Závěr

Je patrné, že statistické zpracování dat ve stopové analýze má celou řadu specifických zvláštností, které je třeba brát v úvahu. Je vždy výhodné začít porovnáním modelů měření a až poté zvolit další cestu. Pro dostatečné počty dat je nejjednodušší sledovat změny variačního koeficientu s úrovní měření x . Podle typu závislosti lze vybrat aditivní (klesající hyperbola), multiplikatívni (konstanta) resp kombinovaný (kombinace předchozího) model měření.

Poděkování: Tato práce vznikla s podporou grantu MŠMT č. VS 97084, grantu GAČR . 106/99/1184 a výzkumného záměru MŠMT č.J11/98:244101113

6. Literatura

- [1] Meloun M., Militký J.: *Zpracování experimentálních dat*, East Publishing Praha 1998
- [2] Rocke D., Lorenzato S.: *Technometrics* 37,176 (1995)
- [3] Horowitz W.: *Analytical Chemistry* 54,67A (1982)
- [4] Massart D.L. a kol. : *Chemometrics a textbook*, Elsevier Amsterdam 1988

Název souboru: stopov1
Adresář: E:\Pom
Šablona: D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název: Zpracování dat ve stopové analýze
Předmět:
Autor: Militky
Klíčová slova:
Komentáře:
Datum vytvoření: 14.09.00 13:40
Číslo revize: 2
Poslední uložení: 14.09.00 13:40
Uložil: Milan Meloun
Celková doba úprav: 0 min.
Poslední tisk: 14.09.00 13:42
Jako poslední úplný tisk
Počet stránek: 12
Počet slov: 2 667 (přibližně)
Počet znaků: 15 207 (přibližně)