

**Univerzita Pardubice
Fakulta chemicko-technologická
Katedra analytické chemie**

STATISTICKÉ ZPRACOVÁNÍ EXPERIMENTÁLNÍCH DAT

STATISTICKÁ ANALÝZA JEDNOROZMĚRNÝCH DAT

Seminární práce 1

Brno, 2002

**Ing. Pavel Němec
ÚKZÚZ Brno**

Obsah

Statistická analýza velkých výběrů.....	3
Statistická analýza malých výběrů dle Horna.....	5
Statistické testování.....	9

Příklad 1. Statistická analýza velkých výběrů

Obsah nitrátového dusíku v ornici zemědělských půd

V rámci Bazálního monitoringu zemědělských půd byly na území jihomoravského regionu odebrány vzorky ornice, a v nich stanoven na nitrátově selektivní elektrodě podle Šenkýře a Petra obsah nitrátového dusíku (NO_3^- v mg.kg^{-1} půdy). Naleznete ve výběru odlehle hodnoty, určete typ rozdělení a míru variability.

Data: Obsah nitrátového dusíku v ornici (mg.kg^{-1})

7.1 12.0 10.6 18.3 9.4 7.9 7.9 6.5 8.2 3.9 12.8 3.5 43 12.0 8.2 9.8 20.4 8.5
7.6 9.3 18.9 13.9 4.3 8.2 7.4 8.8 12.0 11.2 16.6 11.8 9.8 3.3 3.6 6.8 22.6 7.1
10.6 3.6 25.8 22.6 13.0

Program: ADSTAT – jednorozměrná data – exploratorní analýza

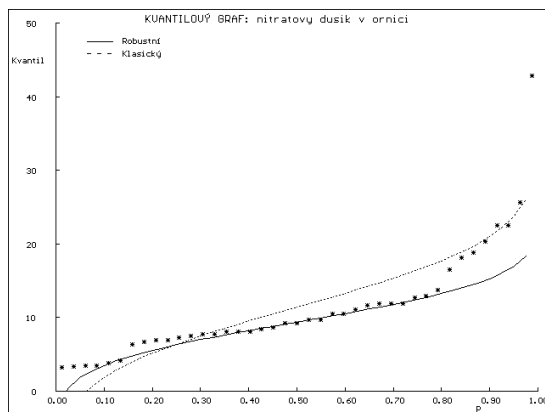
Řešení:

I. Exploratorní analýza spojitá

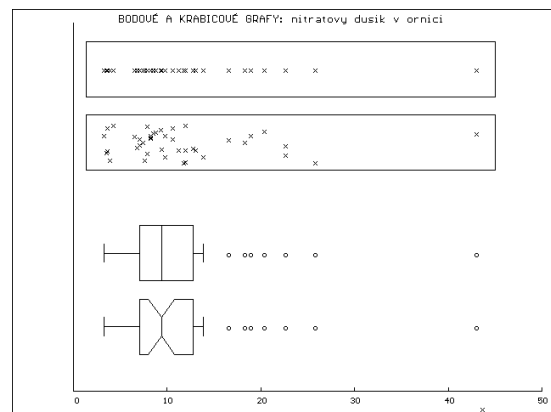
Klasické odhady parametrů:

Medián:	9.4	Průměr:	11.434
Rozptyl:	55.871	Šikmost:	2.1373
Špičatost:	9.0609	Směrodatná odchylka:	7.4747

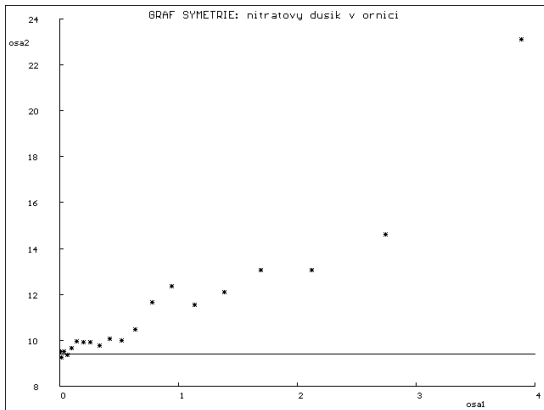
Grafická analýza



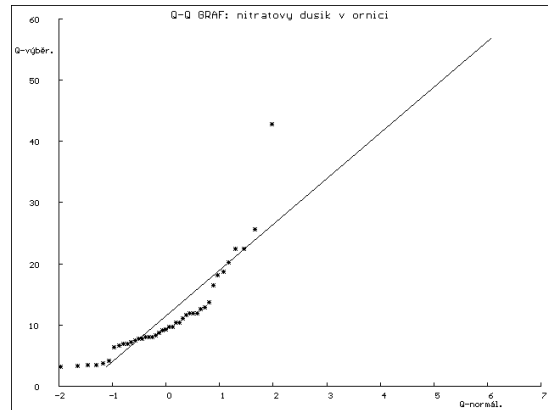
Obr. 1 Kvantilový graf



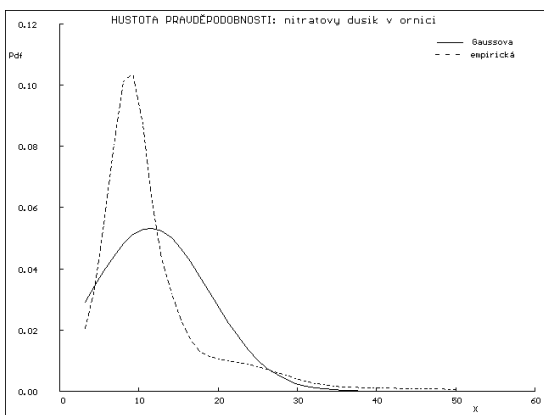
Obr. 2 Diagram rozptýlení a krabicový graf



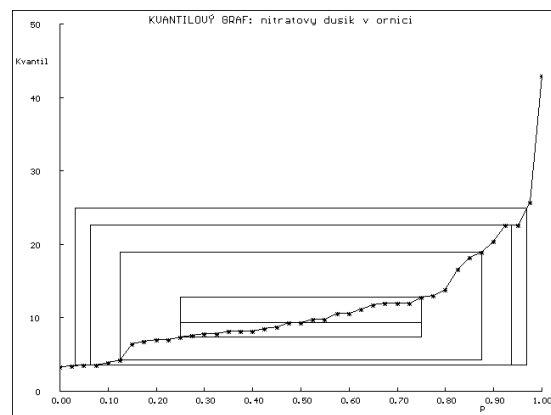
Obr. 3 Graf symetrie



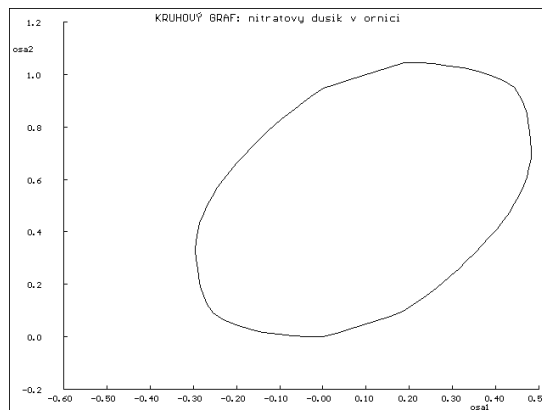
Obr. 4 Q-Q graf



Obr 5 Hustota pravděpodobnosti



Obr. 6 Kvantilový graf



Obr. 7 Kruhový graf

Závěr: Asymetrie, nahore dva odlehlé body, sešikmení k nižším hodnotám, nelze použít aritmetický průměr.

II. Základní předpoklady

1. Předpoklad normality zamítnut
2. Předpoklad nezávislosti přijat
3. Existence dvou odlehlých bodů (13 – 43.0, 39 – 25.8), vzhledem k jedinečnosti dat a relativně malém výběru nelze body vylučovat.
4. typ rozdělení – lognormální (korelační koeficient = 0.98713)

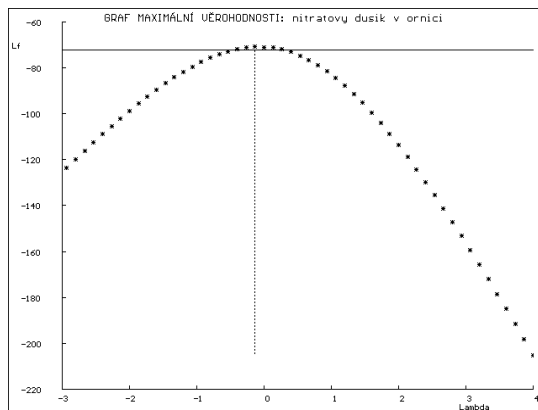
III. Mocinná transformace

1. Mocinná transformace prostá

Zvolená mocnina:	-0.13
Směrodatná odchylka	0.56705
Šikmost	0.99543
Špičatost	2.9525
Opravený průměr	9.4777

2. Box- - Coxova transformace

Zvolená mocnina:	-0.13
Směrodatná odchylka	0.42528
Šikmost	-0.99543
Špičatost	2.9525
Opravený průměr	9.4777



Obr. 8 Graf maximální věrohodnosti

Na grafu maximální věrohodnosti je patrné, že bod [1.0] se nenachází pod segmentem - transformace je nezbytná.

2. Robustní odhady parametrů:

Medián:	9.4000
Dolní mez	7.4824
Horní mez	11.318

Závěr: Analyzovaná data vykazují výrazné lognormální rozdělení, sešikmení k nižším hodnotám. Aritmetický průměr je chybným odhadem střední hodnoty. Vhodným postupem je použití mocninné nebo Box-Coxovy transformace či mediánu. Dle hodnoty intervalu spolehlivosti mediánu obsah nitrátového dusíku kolísá kolem odhadu střední hodnoty mezi 7.48 až 11.31 mg.kg⁻¹ půdy ($\alpha = 0.05$).

Příklad 2. Statistická analýza malých výběrů dle Horna

Obsah přístupného draslíku v ornici

V rámci ověřování půdní analytické metody Mehlich III bylo v rámci jednoho produkčního honu odebráno 13 vzorků půdy z ornice. Kromě ostatních živin byl ve výluhu Mehlich III též stanoven obsah přístupného draslíku. Určete střední hodnotu a interval spolehlivosti podle Horna a porovnejte je s klasickými a robustními statistikami polohy a rozptýlení z ADSTATu.

Data: Obsah přístupného draslíku v ornici (Mehlich III, mg.kg⁻¹)

199 206 306 366 213 159 276 169 191 313 307 223 202

Řešení: 1. Hornův postup pivotů

Uspořádání dat vzestupně

159 169 191 199 202 206 213 223 276 306 307 313 366

Hloubka pivotu

$$H = \text{int}[(n + 1) / 2] / 2$$

$$H = \text{int}[(13 + 1) / 2] / 2 - 3.5 \Leftrightarrow 3$$

Pivoty

$$H_D = x_3 = 191$$

$$H_H = x_{(n+1-H)} = x_{11} = 307$$

Pivotová polosuma

$$P_L = (H_D + H_H) / 2 = (191 + 307) = 249$$

Pivotové rozpětí

$$R_L = H_D - H_H = 307 - 191 = 116$$

95 % interval spolehlivosti střední hodnoty μ

$$P_L - R_L * t_{L,0,0.975}(n) \leq \mu \leq P_L + R_L * t_{L,0,0.975}(n)$$

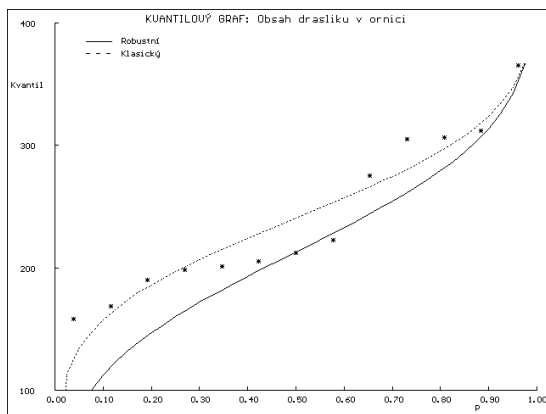
$$249 - 116 * 0.608 \leq \mu \leq 249 + 116 * 0.608$$

$$178.47 \leq \mu \leq 319.528$$

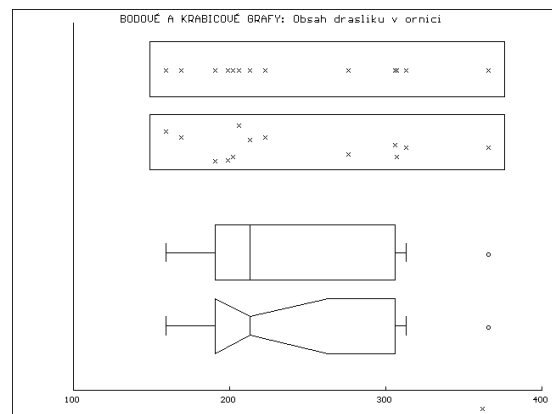
praktické hodnoty

$$178 \leq \mu \leq 320$$

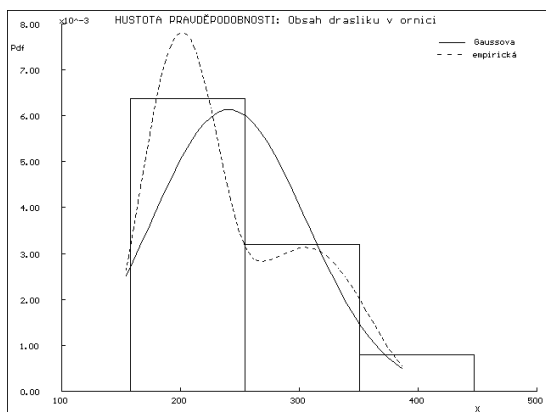
2. Exploratorní analýza



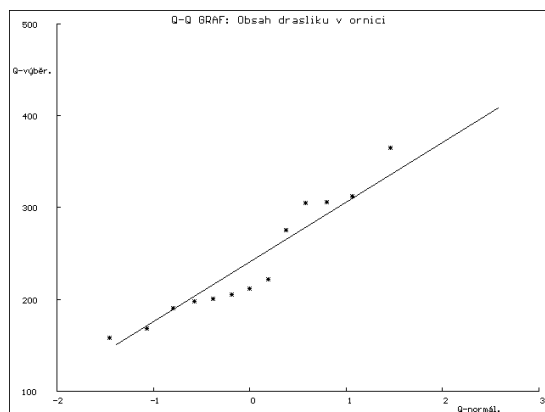
Obr. 1 Kvantilový graf



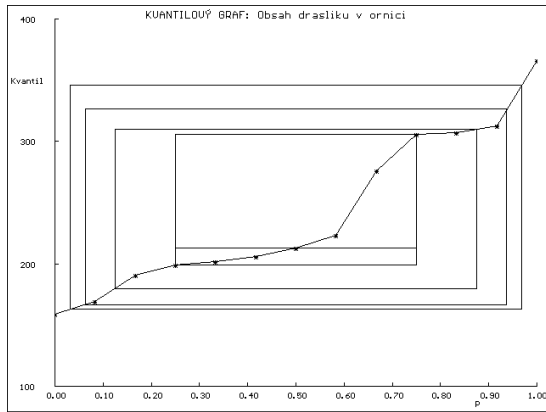
Obr. 2 Bodový a krabicový graf



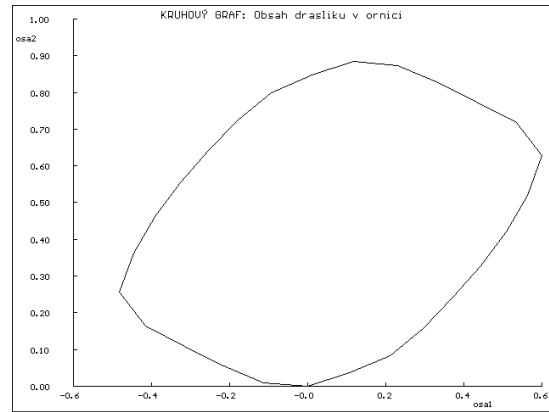
Obr. 3 Hustota pravděpodobnosti



Obr. 4 Q – Q graf



Obr. 5 Kvantilový graf



Obr. 6 Kruhový graf

Závěr: Asymetrie, dole i nahoře jeden odlehlý bod (graficky)

Klasické odhady parametrů

Medián:	213.0	Průměr:	240.8
Rozptyl:	4220	Šikmost:	0.541
Špičatost:	2.012	Směrodatná odchylka:	64.962
Dolní mez (průměr)	201.51	Horní mez(průměr)	280.03
Dolní mez (medián)	144.31	Horní mez(medián)	281.69

Základní předpoklady

Předpoklad normality přijat.

Předpoklad nezávislosti přijat.

Ve výběru nejsou odlehlé body.

Typ rozdělení – rovnoměrné (korelační koeficient = 0.95920)

Závěr: Asymetrie, výběr s dvouvrcholovým rozdělením hustoty pravděpodobnosti. Graficky dva odlehlé body. Hornův průměr je blízký klasickým odhadům – Horn = 249 mg.kg⁻¹ K, aritmetický průměr = 241 mg.kg⁻¹ K.

Nejlépe o střední hodnotě vypovídá medián, což lze prokázat i přehlédnutím dat.

Příklad 3. Statistické testování

Porovnání obsahů thalia v zemědělských půdách (test shodnosti)

Při náhodném screeningu půdních vzorků byl v geologicky specifické oblasti Českomoravské vrchoviny zjištěn v půdě vysoký obsah thalia. Geologicky specifická oblast (melanokratiní granitické horniny svrchního paleozoika) byla podrobně vzorkována a ve vzorcích stanoveno thalium po rozkladu lučavkou královskou metodou ICP – MS. Pro porovnání byla vybrána a shodným způsobem zkoumána polabská aluvia. Určete zda je mezi oběma oblastmi statisticky průkazný obsah thalia na hladině stat.významnosti $\alpha = 0.05$.

Data: Obsah thalia v půdě (Tl v mg.kg⁻¹)

1. výběr - Českomoravská vrchovina

1.612 1.704 1.066 1.102 1.900 1.360 1.842 1.314 1.802 1.496 1.620 0.688 1.932
3.462 2.484 2.134 2.998 2.380 2.130 1.278 1.764 1.096 2.162 2.264 2.446 1.262
1.770 1.038 2.306 2.166 1.120 1.554 1.668 2.112 2.072 1.632 2.242 0.658 2.624
1.366 2.408 1.844 1.370 0.736 0.874 0.516 0.734 1.760 1.190 1.746 1.832 1.956
1.768 1.920 2.180 1.400 1.476 0.690 1.574 0.786 0.446 0.620 2.712 2.376 0.548
1.506 1.664 2.096 1.350 2.144 1.722 1.290 1.066 0.448 0.462

2. výběr – aluvia v Polabí

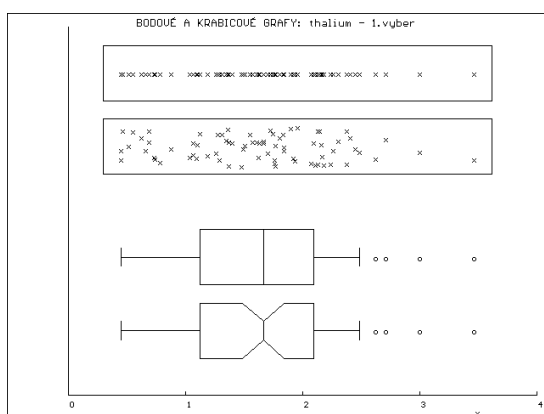
0.166 0.294 0.266 0.098 0.194 0.148 0.080 0.144 0.268 0.196 0.164 0.164 0.138
0.244 0.224 0.394 0.378 0.414 0.406 0.360 0.204 0.218 0.242 0.210 0.244 0.180
0.134 0.260 0.284 0.194

Program: ADSTAT - Porovnání dvou výběrů

I. Ověření normality

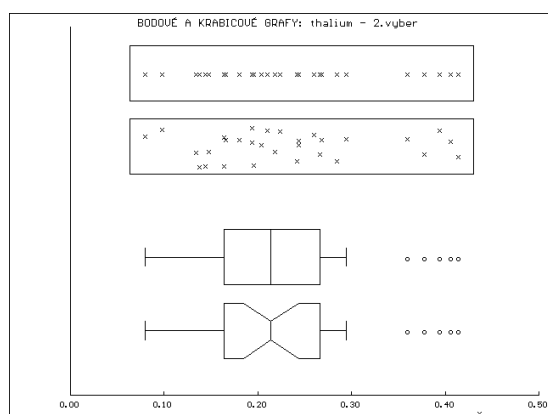
Grafy exploratorní analýzy

1. výběr



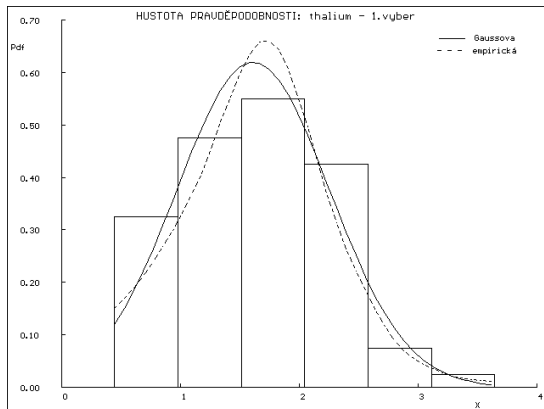
Obr 1. Bodový a krabicový graf

2. výběr

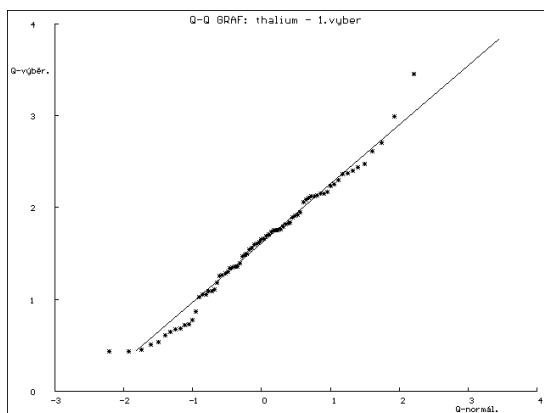


Obr 2. Bodový a krabicový graf

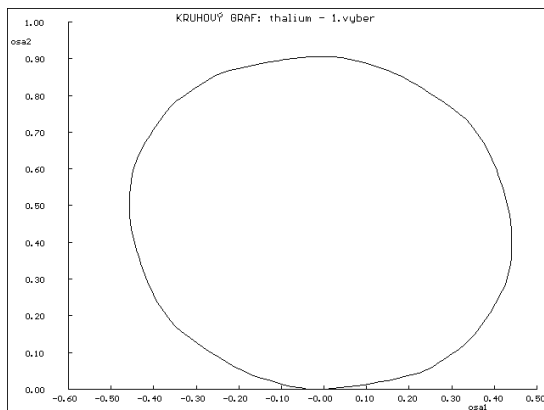
1. výběr



Obr. 3 Hustota pravděpodobnosti

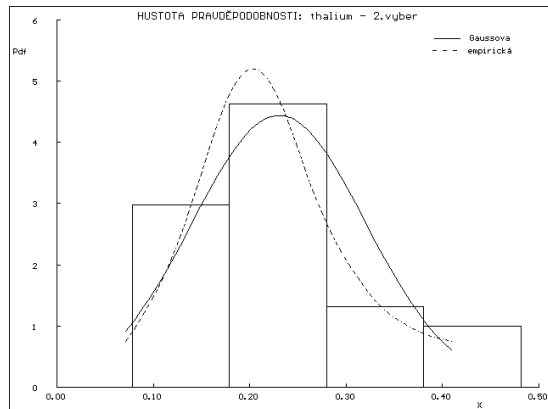


Obr. 5 Q – graf

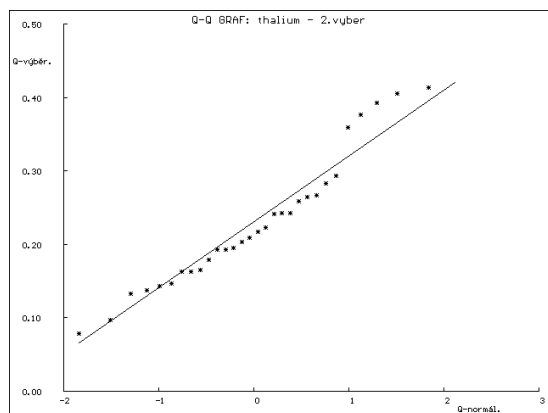


Obr. 7 Kruhový graf

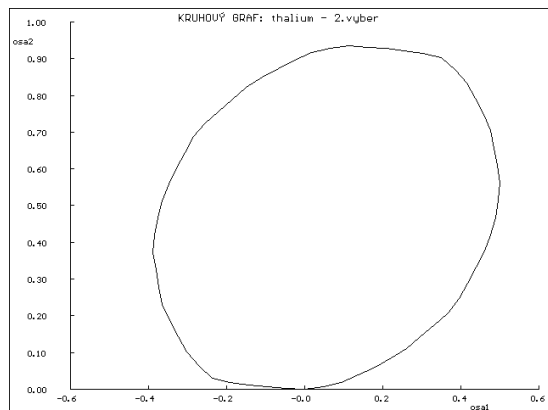
2. výběr



Obr. 4 Hustota pravděpodobnosti



Obr. 6 Q – graf



Obr. 8 Kruhový graf

Závěr grafické analýzy: 1.výběr gaussovské symetrické rozdělení, nahoře několik odlehlých bodů. 2. výběr má velice mírnou asymetrii k nižším hodnotám, gaussovské rozdělení a několik odlehlých bodů dole i nahoře.

Klasické odhady parametru

	1. výběr	2.výběr
Průměr	1.611	0.230
Medián	1.664	0.221
Šikmost	0.1166	0.5914
Špičatost	2.8095	2.6029

Základní předpoklady

1.výběr

Předpoklad normality přijat

|Předpoklad nezávislosti přijat

Ve výběru nejsou odlehlé body

Normální rozdělení, $r = 0.99157$

2.výběr

Předpoklad normality přijat

Předpoklad nezávislosti přijat

Ve výběru nejsou odlehlé body

Normální rozdělení, $r = 0.97417$

II. Porovnání dvou výběrů

Test homogenity rozptylů

Jackknife F test

Tabulkový kvantil $F(1 - \alpha/2, Df1, Df2) = 3.8242$

F – statistika = 16.805

Hladina významnosti = 0.0000

Závěr: Rozptyly se považují za rozdílné, nulová hypotéza zamítnuta.

Test shody průměrů

Tabulkový kvantil $t(1 - \alpha / 2, Df1) = 1.9893$

T2 statistika = 18.130

Hladina významnosti = 0.0000

Závěr: Průměry se považují za rozdílné, nulová hypotéza zamítnuta.

Závěr: Geologicky podmíněné anomální podloží průkazně zvyšuje obsah rizikového thalia v ornici oproti geologicky nezatížené oblasti.

Shodnost obsahu chloru v upravené vodě s normou (test správnosti)

V průběhu jednoho dne byl v úpravně sledován po hodinách obsah chloru. Porovnejte dosažené výsledky s normou – 0.3 mg.l^{-1} .

Data: Obsah chloru (mg.l^{-1})

0.10 0.15 0.25 0.15 0.30 0.25 0.25 0.30 0.35 0.55 0.70 0.70 0.80 0.65 0.55 0.50 0.30
0.35 0.30 0.25 0.25 0.20 0.15

Program: ADSTAT – analýza jednoho výběru

Řešení:

Exploratorní analýza prokázala mírnou asymetrii k nižším hodnotám. Předpoklady normality a nezávislosti potvrzeny. Ve výběru nejsou odlehlé body. Lze použít aritmetický průměr.

Analýza jednorozměrného výběru obsahu chloru

Klasické odhady parametrů

Průměr	= 0.3630
Směrodatná odchylka	= 0.20238
Dolní mez 95 % intervalu spolehlivosti	= 0.27553
Horní mez 95 % intervalu spolehlivosti	= 0.45056

Robustní odhady parametrů

Medián	= 0.3000
Směrodatná odchylka	= 0.1852
Dolní mez 95 % intervalu spolehlivosti	= 0.20672
Horní mez 95 % intervalu spolehlivosti	= 0.39328

Adaptivní odhady parametrů (Hoggovy odhady)

Průměr	= 0.36304
Směrodatná odchylka	= 0.20238
Dolní mez 95 % intervalu spolehlivosti	= 0.27553
Horní mez 95 % intervalu spolehlivosti	= 0.45056

Závěr: S 95 % statistickou jistotou byly nalezeny intervalové odhady obsahu chloru pro aritmetický průměr v rozmezí $0.28 - 0.45 \text{ mg.l}^{-1}$, pro medián $0.21 - 0.39 \text{ mg.l}^{-1}$. Naměřené výsledky vyhovují normě v případě klasických i robustních postupů.

Porovnání dvou extrakčních postupů pro stanovení výměnné půdní reakce (párový test)

V rámci přechodu na ISO postupy v analýze půd byl národní postup stanovení výměnné půdní reakce (pH/ 0.2M KCl) porovnán v s postupem dle ISO (pH/0.01 M CaCl₂). Určete zda je mezi výsledky obou metod průkazný rozdíl na hladině $\alpha = 0.05$.

Data: výměnná půdní reakce

1. výběr pH/KCl

7.5 6.1 5.9 5.8 4.8 6.2 6.7 6.7 7.1 7.6 5.5 7.4 5.4 6.2 6.8 6.4 6.1 7.0 6.3 6.7 7.5
6.9 5.3 6.2 6.1 6.2 5.8 5.2 7.4

2. výběr pH/0.01 CaCl₂

7.4 6.3 6.2 6.1 5.2 6.4 6.9 6.7 7.0 7.4 5.8 7.2 5.6 6.4 7.0 6.6 6.3 7.0 6.6 6.9 7.4
7.0 5.6 6.4 6.2 6.4 5.9 5.5 7.2

Program: ADSTAT – porovnání dvou výběrů – párový test

Řešení:

Počet stupňů volnosti	= 28
Tabulkový kvantil $t(1 - \alpha/2, Df1)$	= 2.048
t-statistika	= -9.2800
vypočtená hladina významnosti	= 0.0000

Závěr: Průměry se považují za rozdílné, nulová hypotéza zamítnuta.

Závěr: ISO postup stanovení pH půd poskytuje statisticky významné odlišné výsledky ($\alpha = 0.05$) oproti původnímu národnímu postupu.