

Bodové míry polohy, rozptýlení a tvaru

1. Mírou polohy je výběrový průměr \bar{x} , který je také *prvním centrálním momentem*,

$$E(\hat{x}) = \mu \quad D(\hat{x}) = \frac{\sigma^2}{n}$$

2. Mírou variability je výběrový rozptyl s^2 , který je také *druhým centrálním momentem*.

$$E(s^2) = \sigma^2 \quad D(s^2) = \frac{\sigma^4}{n} \left[g_2 - \frac{n-3}{n-1} \right]$$

3. Módus, \hat{x}_M je lokální maximum na grafu hustoty pravděpodobnosti čili nejčastější prvek výběru.

5. Kvantilové odhady jsou robustnější, méně citlivé na vybočující hodnoty než je tomu u momentových odhadů.

6. Pro rovnoměrné rozdělení dat je vydatným odhadem polohy výběrová polosuma \hat{x}_P definovaná vztahem

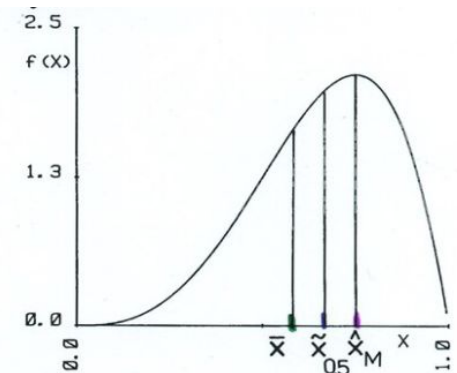
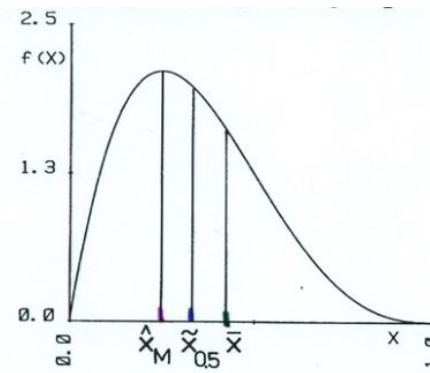
$$\hat{x}_P = \frac{x_{(1)} + x_{(n)}}{2}$$

kde $x_{(1)}$ je minimální a $x_{(n)}$ maximální prvek ve výběru.

Rozptyl odhadu polosumy je $D_R(\hat{x}_P) = \frac{6 \sigma^2}{(n-1)(n-2)}$

7. U řady měřicích přístrojů se měří za podmínek konstantní relativní chyby měření. Pro rozptyl měření platí, že

$$\sigma_x^2 = x_i^2 \delta^2$$



Vztah mezi módem \hat{x}_M , průměrem \bar{x} a mediánem $\hat{x}_{0.5}$ pro sešikmená beta rozdělení (a) k nižším, (b) k vyšším hodnotám

4. Výběrový medián, $\hat{x}_{0.5}$ dělí výběr na dvě části, každá obsahuje 50% prvků. Pro prvky seřazené vzestupně $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ (pořádkové statistiky), je **pro n liché** $\hat{x}_{0.5} = x_{(k)}$, kde $k = (n+1)/2$,

a pro n sudé $\hat{x}_{0.5} = [x_{(k)} + x_{(k+1)}]/2$, kde $k = n/2$.

8. Má-li každé měření x_i normální rozdělení s hustotou pravděpodobnosti a rozptylem $\sigma^2(x_i)$, lze pro odhad střední hodnoty odvodit vztah tzv. vážený aritmetický průměr s vahami $1/\sigma^2(x_i)$,

$$\hat{x}_w = \frac{\sum_{i=1}^n \frac{x_i}{\sigma^2(x_i)}}{\sum_{i=1}^n \frac{1}{\sigma^2(x_i)}}$$

Rozptyl tohoto odhadu má tvar

$$D(\bar{x}_w) = \frac{1}{\sum_{i=1}^n \frac{1}{\sigma^2(x_i)}}$$

9. Když se provádí měření za podmínek konstantní relativní chyby δ , má **vážený odhad střední hodnoty** tvar

$$\bar{x}_w = \frac{\sum_{i=1}^n \frac{1}{x_i}}{\sum_{i=1}^n \frac{1}{x_i^2}}$$

a pro rozptyl tohoto odhadu platí $D(\bar{x}_w) = \frac{\delta^2}{\sum_{i=1}^n \frac{1}{x_i}}$

10. Pro odhad variability je možné užít kvantilové odhady, např. **interkvartilové rozpětí**

$$R = 0.7413 (\bar{x}_{0.75} - \bar{x}_{0.25})$$

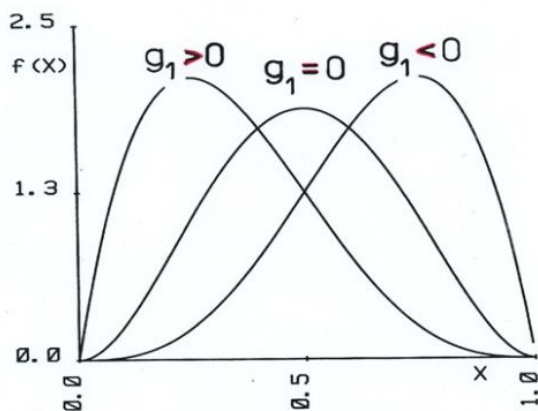
kde $\bar{x}_{0.75}$ je odhad horního a $\bar{x}_{0.25}$ odhad dolního kvantilu.

13. **Momentový odhad šikmosti** g_1 je prvním parametrem tvaru rozdělení

$$\hat{g}_1 = \frac{\sqrt{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[\sum_{i=1}^n (x_i - \bar{x})^4 \right]^{3/2}}$$

a střední hodnota pro výběry z normálního rozdělení je rovna nule, $E(\hat{g}_1) = 0$, a pro asymptotický rozptyl odhadu platí

$$D(\hat{g}_1) = \frac{(n-2)}{(n+1)(n+3)}$$



11. **Průměrná absolutní odchylka** d

$$d = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^n |x_i - \mu|$$

faktor $\sqrt{\pi/2}$ zajišťuje pro normální rozdělení asymptoticky přiblížení k směrodatné odchylce σ .

12. Pro relativní rozptýlení dat se užívá relativní směrodatná odchylka nebo-li **variační koeficient** $\delta = \sigma/\mu$ a odhad je

$$\hat{\delta} = \frac{S}{\bar{X}}$$

Pro rozptyl tohoto odhadu platí $D(\hat{\delta}) \approx \delta^2 \frac{n + \delta^2 (2n + 1)}{2n(n-1)}$.

14. **Momentový odhad špičatosti** g_2 je druhým parametrem tvaru rozdělení

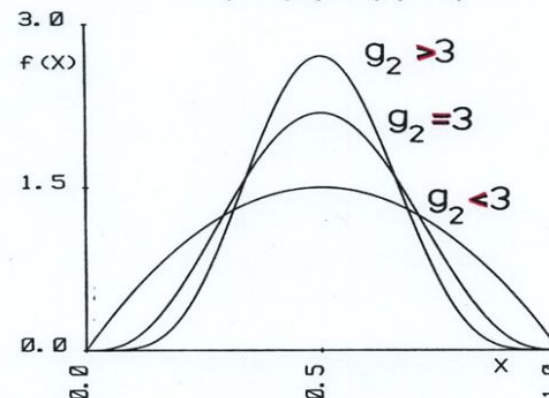
$$\hat{g}_2 = \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2}$$

a střední hodnota pro výběry z normálního rozdělení je

$$E(\hat{g}_2) = 3 - \frac{6}{n+1}$$

a pro asymptotický rozptyl tohoto odhadu platí

$$D(\hat{g}_2) \approx \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$



Kdy použít polosumy?

Polosuma je efektivnější než \bar{x} pro $g_2 > 2.2$.

- Pro rozdělení s plochými vrcholy se doporučuje použití **kvartilové polosumy**

$$P_F = (\tilde{x}_{0.75} - \tilde{x}_{0.25}) / 2$$

kde $\tilde{x}_{0.75}$ resp. $\tilde{x}_{0.25}$ je horní, resp. dolní kvartil

- V případě ohraničených rozdělení (arkussínové a lichoběžníkové třídy) je efektivní tzv. **polosuma**

$$\hat{x}_P = (x_{\max} - x_{\min}) / 2$$

kde x_{\max} je maximální a x_{\min} je minimální prvek výběru

Kombinovaný odhad centrální hodnoty

- Pro symetrická rozdělení s vybočujícími hodnotami je doporučen za odhad středu symetrie čili centrální hodnoty použít **medián** dle vzorce

$$\tilde{x}_C = \text{med}\{\bar{x}, \tilde{x}_{0.5}, \hat{x}_P, PF, \bar{x}(0.25)\}$$

kde $\text{med}\{\cdot\}$ označuje medián z prvků v závorce.

Pro odhad rozptylu odhadu \tilde{x}_C je možno použít interkvartilové délky

$$k_{0.9} = (\tilde{x}_{0.95} - \tilde{x}_{0.05}) / 2$$

$$D(\tilde{x}_C) = k_{0.9}^2 / (2.72N)$$

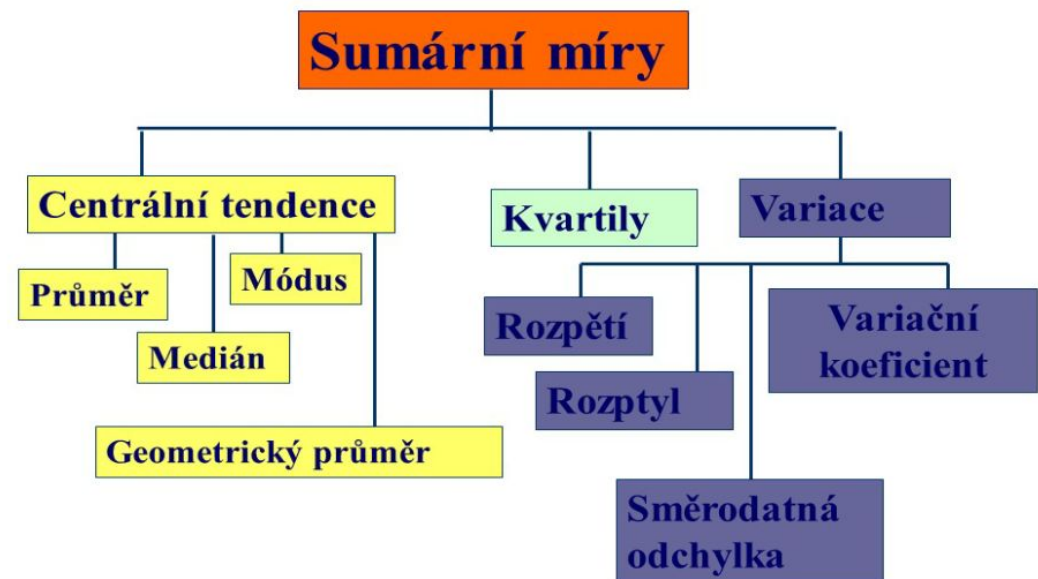
Uřezaný průměr

- α -uřezaný průměr $\bar{x}(\alpha)$ je definován vztahem

$$\bar{x}(\alpha) = \frac{1}{n - 2M} \sum_{i=M+1}^{N-M} x_{(i)}$$

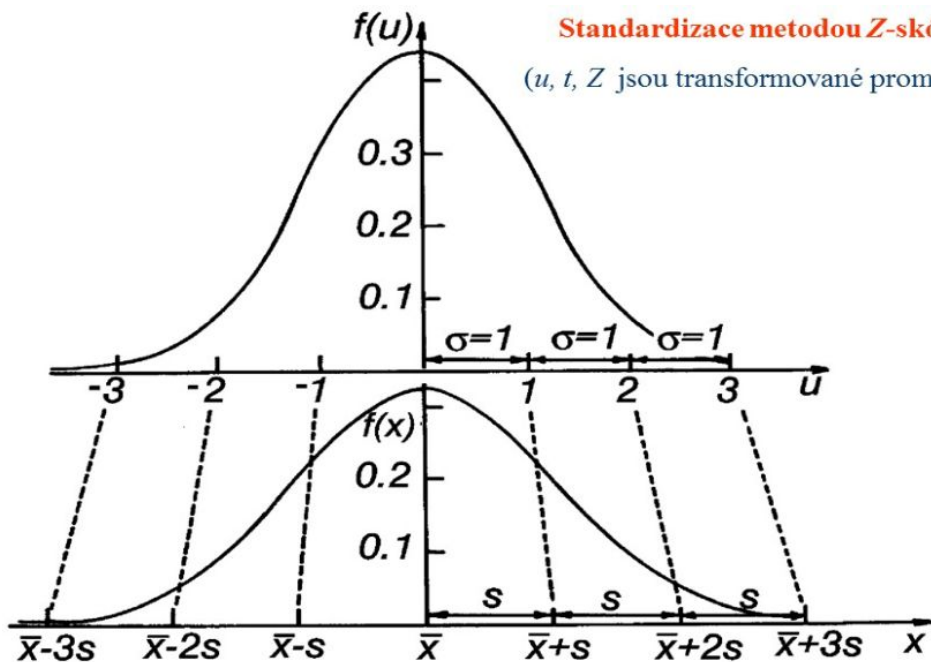
kde $M = \text{int}(\alpha N / 100)$ je celá část výrazu $\alpha N / 100$ a $x_{(i)}$ jsou pořádkové statistiky (vzestupně seříděné prvky výběru).

SUMARIZACE DAT



Standardizace metodou Z-skóre

(u, t, Z jsou transformované proměnné)



Úlohy k procvičení odhadů střední hodnoty

M. Meloun, J. Militký:
Kompedium statistického zpracování dat,
Karolinum Praha 2012

Analýza velkého výběru

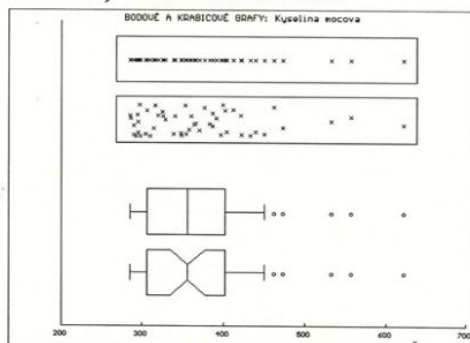
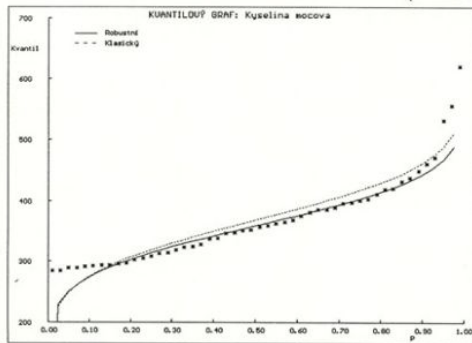
Na úloze B2.25 Koncentrace kyseliny močové v krvi dárců ukážeme postup analýzy velkého výběru s odlehlými prvky pro určení typu rozdělení koncentrace kyseliny močové u 50 dárců krve. Jaká je míra polohy a rozptýlení uvedeného výběru?

Data: Koncentrace kyseliny močové [$\mu\text{mol/l}$]:

295	325	320	316	314	310	306	304	298	297
462	294	292	290	286	285	622	557	532	473
397	450	439	433	422	421	412	404	402	399
358	391	387	387	383	377	370	366	364	360
295	354	353	349	348	340	339	329	326	290

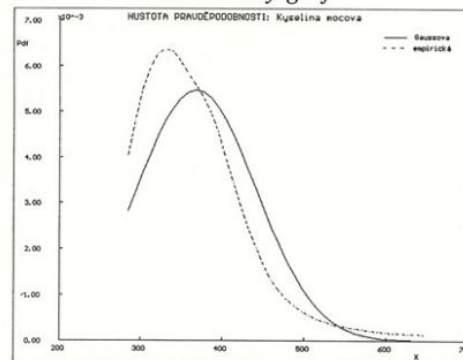
Řešení:

1. PRŮZKUMOVÁ (EXPLORATORNÍ) ANALÝZA DAT



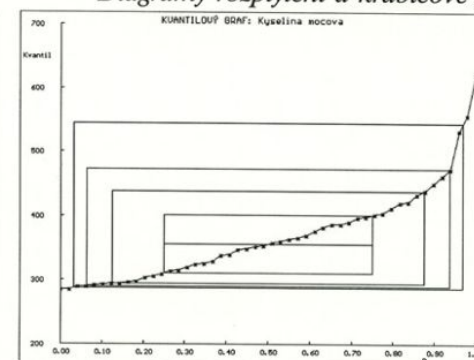
14

Kvantilový graf



Graf hustoty pravděpodobnosti

Diagramy rozptýlení a krábicevé grafy



Graf rozptýlení s kvantily

Poznatky z grafů:

1. Výběrové rozdělení je asymetrické a silně sešikmené.
2. V horní části pořádkových statistik lze indikovat 3 až 5 podezřelých bodů, z nich se 3 jeví jako vysloveně odlehlé. Nelze proto použít klasických odhadů polohy a rozptýlení.

2. OVĚŘENÍ ZÁKLADNÍCH PŘEDPOKLADŮ O VÝBĚRU

Ověření předpokladů o výběru (výstup programu ADSTAT)

(a) TEST NORMALITY:

Tabulkový kvantil $\chi^2(1-\alpha, 2)$: 5.9915
χ^2 -statistika	: 29.199
Závěr: Předpoklad normality zamítnut	
Vypočtená hladina významnosti	: 4.5661E-07

(b) TEST NEZÁVISLOSTI:

Tabulkový kvantil $t(1-\alpha/2, n+1)$: 2.0076
Test autokorelace	: 1.0266
Závěr: Předpoklad nezávislosti přijat	
Vypočtená hladina významnosti	: 0.1547

(c) DETEKCE ODLEHLÝCH BODŮ: modifikované vnitřní hradby:

Dolní vnitřní hradba	: 109.62
Horní vnitřní hradba	: 602.38
Bod číslo 17 (horní)	: 622.00

Parametry s vynechanými odlehlými hodnotami:

Průměr	: 363.29
Směrodatná odchylka	: 63.875
Šikmost	: 0.9700
Špičatost	: 3.8090

3. MOCNINNÁ TRANSFORMACE

Transformace dat (výstup programu ADSTAT)

(a) PROSTÁ MOCNINNÁ TRANSFORMACE:	Opravený průměr: 350.91
(b) BOX-COXOVA TRANSFORMACE:	Opravený průměr: 362.17

Poznátky z výstupu:

1. Rozdělení souboru vykazuje mírné sešikmení.
2. Soubor obsahuje jeden výrazně odlehlý bod.
3. Mocninná transformace zde selhává.
4. Box-Coxova transformace přináší zlepšení parametrů šikmosti a špičatosti souboru a je robustní vůči odlehlé hodnotě. Věrohodnější se proto jeví odhad, získaný metodou Boxovy-Coxovy transformace, a to **362.17 $\mu\text{mol/l}$** .

Poznátky z výstupu:

1. Předpoklad o normalitě rozdělení je zamítnut, protože hodnota testovacího kritéria χ^2 je vyšší než tabulkový kvantil.
2. Předpoklad nezávislosti je přijat, protože hodnota testovacího kritéria je nižší než tabulkový kvantil.
3. Data nejsou homogenní, mimo modifikované vnitřní hradby $B_D = 109.62 \mu\text{mol/l}$, $B_H = 602.38 \mu\text{mol/l}$ leží hodnota č. 17, a to $622.0 \mu\text{mol/l}$.
4. Po odstranění odlehlé hodnoty je aritmetický průměr $\bar{x} = 363.29 \mu\text{mol/l}$ a směrodatná odchylka $s = 63.875 \mu\text{mol/l}$.
5. Protože se však jedná o biochemická data, nelze odlehlou hodnotu vyloučit. Znamenalo by to zde totiž ztrátu informace.

4. ODHADY POLOHY, ROZPTYLU A TVARU

Odhady polohy, rozptýlení a tvaru (výstup programu ADSTAT)

(a) PARAMETRY TVARU:			
Šikmost	: 1.3299		
Špičatost	: 5.0034		
(b) KLASICKÉ ODHADY PARAMETRŮ :			
Průměr	: 368.46		
Směr. odchylka	: 73.044		
95.0% spolehlivost: Mez spodní:	347.70	horní:	389.22
(c) OSTATNÍ ODHADY POLOHY:			
Odhad modu	: 293.00		
Odhad polosumy	: 453.50		
(d) ROBUSTNÍ ODHADY PARAMETRŮ :			
Medián	: 356.00		
Směr. odchylka mediánu	: 85.476		
95.0% spolehlivost: Mez spodní:	331.52	horní:	380.48
Uřezání 5% (pro P=0.05):			
Průměr	: 361.38		
Směr. odchylka	: 68.636		
95.0% spolehlivost: Mez spodní:	340.94	horní:	381.82
Uřezání 10% (pro P=0.10):			
Průměr	: 358.35		
Směr. odchylka	: 67.344		
95.0% spolehlivost: Mez spodní:	339.10	horní:	377.60
Biweight:			
Průměr	: 358.17		
Směr. odchylka	: 63.086		
95.0% spolehlivost: Mez spodní:	339.29	horní:	377.05
(e) ADAPTIVNÍ ODHADY PARAMETRŮ: Hoggovy odhady:			
Průměr	: 356.46		
Směr. odchylka	: 72.722		
95.0% spolehlivost: Mez spodní:	335.79	horní:	377.13

Analýza malého výběru

Poznátky z výstupu:

- Dobrým odhadem střední hodnoty se jeví uřezané aritmetické průměry ale především předešle retransformovaný průměr metodou dle Boxe a Coxe, a to **362.17 μmol/l**.
- Výpočet aritmetického průměru při vynechání odlehlého bodu podstatně zlepšil variabilitu souboru a potvrzuje především robustnost retransformovaného průměru vůči odlehlé hodnotě.

5. Pivotové rozpětí $R_L = x_H - x_D$ 3.32 - 0.49 = 2.83

6. 95%ní interval spolehlivosti střední hodnoty μ : $t_{L, 1-\alpha/2} = 0.564$

$$P_L - R_L t_{L, 1-\alpha/2}(n) \leq \mu \leq P_L + R_L t_{L, 1-\alpha/2}(n)$$

$$1.905 - 2.83 \times 0.564 \leq \mu \leq 1.905 + 2.83 \times 0.564$$

$$0.31 \leq \mu \leq 3.50$$

7. Závěr: Bodový odhad míry polohy je **1.91 g/l**, míry rozptýlení **2.83** a intervalový odhad míry polohy je **0.31 g/l $\leq \mu \leq 3.50$ g/l**.

Na vzorové úloze **B3.01 Střední hodnota haptoglobinu v lidském krevním séru** ukážeme Hornův postup analýzy malých výběrů.

Data: Koncentrace haptoglobinu [g l⁻¹]: 1.82 3.32 1.07 1.27 0.49 3.79 0.15 1.98

Řešení: Hornův postup pivotů pro malé výběry ($4 < n < 20$):

1. Pořádkové statistiky:

i	1	2	3	4	5	6	7	8
$x_{(i)}$	0.15	0.49	1.07	1.27	1.82	1.98	3.32	3.79

2. Hloubka pivotu:

$$H = \text{int} \frac{\frac{m+1}{2} + 1}{2}$$

$n = 8$, sudé

$\text{int}(2.75) \approx 2$

3. Pivoty: Dolní pivot $x_D = x_{(H)}$

Horní pivot $x_H = x_{(n+1-H)}$

$x_{(2)} = 0.49$

$x_{(7)} = 3.32$

4. Pivotová polosuma $P_L = \frac{x_D + x_H}{2}$

= 1.905

Úloha B3.03 Vliv glucagonu na koncentraci krevního cukru (Horn)

Je třeba vyšetřit vliv farmaka glucagonu na snížení hladiny krevního cukru. Po 15 minutách od dávkování glucagonem byla u 8 pokusných krys hladina 300 mg/100 ml významně snížena. Byly naměřeny následující hodnoty krevního cukru v mg/100 ml krve. Jde o symetrické rozdělení? Určete odhad střední hodnoty krevního cukru. Aplikujte i Hornův postup. ([19] s. 43). Obsahuje intervalový odhad hodnotu 300 mg/100 ml?

Data: Hladina krevního cukru [mg/100 ml]: 270 275 265 250 280 245 265 260.

Analýza malých výběrů

N:	8
Střední hodnota:	262,5
Spodní mez (5%):	250,775
Horní mez (95%):	274,225
Spodní mez (2.5%):	248,4
Horní mez (97.5%):	276,6
Pivotové rozpětí:	25

Exponenciální transformace dat:

Zvolený parametr:	-0,2319526672
Opravený průměr:	264,8761089
Spodní mez:	254,2174579
Horní mez:	273,7301717

Box-Coxova transformace dat:

Pravděpodobnost:	33,1473236485966 %
Zvolený parametr:	5,271528625
Věrohodnost:	12,96717158
Opravený průměr:	264,6230685

Klasické parametry:

Sloupce:	B303
Průměr:	263,75
Spodní mez:	253,8202874
Horní mez:	273,6797126
Směr. odchylka:	11,87734939
Šikmost:	-0,299073999
Odchylka od 0:	Nevýznamná
Špičatost:	2,024835763
Odchylka od 3:	Nevýznamná

Test normality

Sloupce:	B303
Normalita:	Přijata
Testové kritérium:	0,3424940481
Kritický kvantil $\chi^2(22)$:	5,991464547
p-hodnota:	0,842613402

Momentový

Robustní parametry:

Medián:	265
IS spodní:	243,8868957
IS horní:	286,1131043
Medianová směr. odchylka:	8,928735432
Medianový rozptyl:	79,72231641

Úloha B3.05 Stanovení čistoty kalciferolu (Horn)

V kontrolní laboratoři se stanovuje čistota kalciferolu metodou vnějšího standardu na kapalinovém chromatografu. Vyšetřete požadavky kladené na výběr a určete parametry polohy a rozptýlení. Které diagnostiky ukazují, že ve výběru jsou odlehla měření? Je rozdělení symetrické? Na jaký tvar rozdělení ukazuje koeficient šikmosti a špičatosti? Stanovte odhad střední hodnoty a její 95% interval spolehlivosti. Aplikujte také Hornův postup. U kterého odhadu polohy vychází interval spolehlivosti širší?

Data: Procento čistoty kalciferolu [%]: 97.6 99.0 99.5 98.8 101.0 99.2 98.7.

Analýza malých výběrů

N :	7
Střední hodnota :	99,1
Spodní mez (5%) :	98,66
Horní mez (95%) :	99,54
Spodní mez (2.5%) :	98,524
Horní mez (97.5%) :	99,676
Pivotové rozpětí :	0,8

Exponenciální transformace dat :

Zvolený parametr :	0,2067108154
Opravený průměr :	99,03081787
Spodní mez :	98,19008805
Horní mez :	100,0388569

Box-Coxova transformace dat :

Pravděpodobnost :	43,9756030776374 %
Zvolený parametr :	-3,900748825
Věrohodnost :	11,43571082
Opravený průměr :	99,03107677

Klasické parametry :

Sloupec :	B305
Průměr :	99,11428571
Spodní mez :	98,1674608
Horní mez :	100,0611106
Směr. odchylka :	1,023765226
Dolní mez :	0,6597075629
Horní mez :	2,254398712
Šikmost :	0,5347700116
Odchylka od 0 :	Nevýznamná
Špičatost :	3,179825844
Odchylka od 3 :	Nevýznamná

Test normality

Normalita :	Přijata
Testové kritérium :	0,9181604091
Kritický kvantil $\chi^2(22)$:	5,991464547
p-hodnota :	0,6318645645

Momentový

Přijata
0,9181604091
5,991464547
0,6318645645

Robustní parametry :

Medián :	99
IS spodní :	96,87763951
IS horní :	101,1223605
Medianová sm. od :	0,8673628705
Medianový rozptyl :	0,7523183492

25 ADSTAT: Opravený průměr: 99,101

Úloha B3.14 Test správnosti nalezeného obsahu penicilinu v krvi vůči deklarovanému (Horn)

Vzorek s deklarovaným obsahem penicilinu 2.4 mg.l⁻¹ byl proměřen HPLC analýzou. Zjistěte, zda stanovená koncentrace odpovídá požadavku. Užití intervalový odhad na hladině významnosti $\alpha = 0.05$.

Data: Obsah penicilinu v krvi [mg. l⁻¹]: 2.2 2.3 2.5 2.1 2.3 2.4 2.5.

Analýza malých výběrů

N :	7
Střední hodnota :	2,35
Spodní mez (5%) :	2,185
Horní mez (95%) :	2,515
Spodní mez (2.5%) :	2,134
Horní mez (97.5%) :	2,566
Pivotové rozpětí :	0,3

Exponenciální transformace dat :

Zvolený parametr :	-0,1936454773
Opravený průměr :	2,34040184
Spodní mez :	2,191312348
Horní mez :	2,465808129

Box-Coxova transformace dat :

Zvolený parametr :	4,076744461
Opravený průměr :	2,336400621

Klasické parametry :

Sloupec :	B314
Průměr :	2,328571429
Spodní mez :	2,190212055
Horní mez :	2,466930802
Směr. odchylka :	0,1496026483
Dolní mez :	0,09640296038
Horní mez :	0,3294349222
Šikmost :	-0,1975061961
Odchylka od 0 :	Nevýznamná
Špičatost :	1,846763241
Odchylka od 3 :	Nevýznamná

Test normality

Sloupec :	B314
Normalita :	Přijata
Testové kritérium :	0,1491712813
Kritický kvantil $\chi^2(22)$:	5,991464547
p-hodnota :	0,9281279852

Momentový

Přijata
0,1491712813
5,991464547
0,9281279852

Robustní parametry :

Sloupec :	B314
Medián :	2,3
IS spodní :	2,050310531
IS horní :	2,549689469
Medianová směr. od :	0,1020426907

27 ADSTAT: Opravený průměr: 2,340

Úloha B3.09 Test správnosti koncentrace cyclosporinu metodou HPLC (Horn)

Pro studii biologické dostupnosti cyclosporinu A byl zakoupen roztok této látky v metanolu. Deklarovaná koncentrace cyclosporinu A byla 20 ng/ml. Při HPLC analýzách byly naměřeny koncentrace uvedené v Datech. Test správnosti je třeba provést na hladině významnosti $\alpha = 0.05$. Obsahuje intervalový odhad střední hodnoty číslo 20 ng/ml?

Data: Koncentrace cyclosporinu A [ng/ml]: 19.96 20.05 20.00 19.99 20.01 19.98 20.00 20.02 20.01 19.93.

Analýza malých výběrů

N :	10
Střední hodnota :	19,995
Spodní mez (5%) :	19,97931
Horní mez (95%) :	20,01069
Spodní mez (2.5%) :	19,97496
Horní mez (97.5%) :	20,01504
Pivotové rozpětí :	0,03

Exponenciální transformace dat :

Zvolený parametr :	-0,1681041718
Opravený průměr :	19,99734999
Spodní mez :	19,97277985
Horní mez :	20,01922664

Box-Coxova transformace dat :

Zvolený parametr :	4,934959793
Opravený průměr :	19,99729821

Klasické parametry :

Sloupec :	B309
Průměr :	19,995
Spodní mez :	19,97133428
Horní mez :	20,01866572
Směr. odchylka :	0,03308238874
Dolní mez :	0,0227552317
Horní mez :	0,06039554574
Šikmost :	-0,4269918497
Odchylka od 0 :	Nevýznamná
Špičatost :	2,993764333
Odchylka od 3 :	Nevýznamná

Test normality

Normalita :	Přijata
Testové kritérium :	0,7466989989
Kritický kvantil $\chi^2(22)$:	5,991464547
p-hodnota :	0,6884245868

Momentový

Přijata
0,7466989989
5,991464547
0,6884245868

Robustní parametry :

Medián :	20
IS spodní :	19,93074902
IS horní :	20,06925098
Medianová sm. od :	0,0306128072
Medianový rozptyl :	0,0009371439644

26 ADSTAT: Opravený průměr: 19,995

Úloha E2.10 Vyšetření symetrie výběrového rozdělení obsahu dusičnanů v pitné vodě

Fotometrickou metodou se salicylanem sodným byl stanoven obsah dusičnanů [mg.l⁻¹] v reálném vzorku pitné vody. Aplikujte postup exploratorní analýzy EDA, rozeberte její diagnostiky a učinite závěry o typu rozdělení. Jsou ve výběru dat odlehle hodnoty? Prokažte, zda je v tomto případě výhodnější mocinná transformace dat nebo užití robustních odhadů?

Data: Obsah dusičnanů v pitné vodě [mg.l⁻¹]: 57.56 57.80 58.59 56.72 59.33 58.27 56.65 57.03 56.58 55.71 58.00 57.08 58.41 53.64 57.13 58.04 58.45 57.92 56.21

Klasické parametry :

Sloupec :	E210
Průměr :	57,32210526
Spodní mez :	56,70454908
Horní mez :	57,93966145
Rozptyl :	1,641673099
Směr. odchylka :	1,281277917
Dolní mez :	0,968149780
Horní mez :	1,89478522
Robustní směr.odch.	1,245384
Šikmost :	-1,13258890
Odchylka od 0 :	Významná
Špičatost :	4,701557261
Odchylka od 3 :	Nevýznamná
Polosuma :	56,485
Modus :	57,98821053
Geometrický průměr :	57,30830226
Harmonický průměr :	57,29425437

Robustní parametry :

Sloupec :	E210
Medián :	57,56
IS spodní :	54,51039133
IS horní :	60,60960867
Medianová směr. odchylka :	1,451557275
Medianový rozptyl :	2,107018521
10% Průměr :	57,42058824
10% IS spodní :	56,90524942
10% IS horní :	57,93592705
10% Směr. odchylka :	0,8099892274
10% Rozptyl :	0,6560825485
40% Průměr :	57,47272727
40% IS spodní :	56,85295035
40% IS horní :	58,09250419
40% Směr. odchylka :	0,4256846819
40% Rozptyl :	0,1812074484

Analýza malých výběrů

N :	19
Střední hodnota :	57,46
Spodní mez (5%) :	56,87518
Horní mez (95%) :	58,04482
Spodní mez (2.5%) :	56,77474
Horní mez (97.5%) :	58,14526
Pivotové rozpětí :	1,62

Test normality

Sloupec :	E210
Normalita :	Přijata
Testové kritérium :	4,738286485
Kritický kvantil $\chi^2(22)$:	5,991464547
p-hodnota :	0,09262990487

28 ADSTAT: Opravený průměr: 57.308

Úloha E2.37 Výčetní tloušťka na kmeni 1.3 m nad zemí stromů ve stejnověkém kontrolním smrkovém porostu

Měřením Fluryho průměrkou byla v roce 1996 získána data pro výčetní tloušťku, vyjádřenou v cm a měřenou na kmeni 1.3 m nad zemí stromů ve stejnověkém kontrolním smrkovém porostu. Aplikujte postup postupného vyšetření v pořadí: 1. Průzkumová analýza spojitých dat (EDA), 2. Ověření předpokladů o výběru, 3. Transformace dat, 4. Statistická analýza jednorozměrných dat (CDA). Rozeberte a vysvětlete diagnostiky a učíte své závěry.

Data: Výčetní tloušťka stromů smru [cm]: 17.1 16.3 17.2 9.2 12.4 12.9 13.8 14.7
15.6 16.2 11.4 13.2 14.5 10.8 12.7 15.4 12.9 12.3

Klasické parametry :

Sloupec : E237
Průměr : 13,50481928
Spodní mez : 12,95828075
Horní mez : 14,05135781
Rozptyl : 6,26485454
Směr. odchylka : 2,502969145
Dolní mez : 2,176958534
Horní mez : 2,945719414
Robustní směr.odch. : 2,52042
Šikmost : -0,1954438735
Odchylka od 0 : Nevýznamná
Špičatost : 2,8021386
Odchylka od 3 : Nevýznamná
Polosuma : 14,15
Modus : 13,78582903
Geometrický průměr : 13,26024415
Harmonický průměr : 12,99843785

Robustní parametry :

Sloupec : E237
Medián : 13,6
IS spodní : 11,87453895
IS horní : 15,32546105
Medianová směr. odchylka : 0,8673628705
Medianový rozptyl : 0,7523383492
10% Průměr : 13,55205479
10% IS spodní : 12,98039101
10% IS horní : 14,12371858
10% Směr. odchylka : 1,87670666
10% Rozptyl : 3,522027888
20% Průměr : 13,62769231
20% IS spodní : 13,03715878
20% IS horní : 14,21822584
20% Směr. odchylka : 1,473333811
20% Rozptyl : 2,170712518
40% Průměr : 13,66326531
40% IS spodní : 13,08909607
40% IS horní : 14,23743454
40% Směr. odchylka : 0,9176120464
40% Rozptyl : 0,8420118676

Exponenciální transformace dat:

Zvolený parametr : -0,07294082642
Opravený průměr : 13,59294802
Spodní mez : 13,04473818
Horní mez : 14,13255062

Test normality

Sloupec : E237
Normalita : Přijata
Testové kritérium : 0,7745526268
Kritický kvantil $\chi^2(22)$: 5,991464547
p-hodnota : 0,6789034789

Momentový

Sloupec : E237
Normalita : Přijata
Testové kritérium : 0,7745526268
Kritický kvantil $\chi^2(22)$: 5,991464547
p-hodnota : 0,6789034789

29 ADSTAT: Opravený průměr: 13.582

Úloha S2.04 Vyšetření rozdělení výsledků Amthaireova testu u 98 studentů

V rámci přijímacího řízení absolvují uchazeči o studium na vysoké škole Amthauerův test struktury inteligence. Výsledky testu se vyjadřují prostřednictvím tzv. celkového hrubého skóre. Ze studentů přijatých ke studiu během 4 let byl proveden náhodný výběr 98 studentů. Proveďte průzkumovou analýzu dat EDA a ověření předpokladů pro tento výběr. Zkonstruujte bariérově-číslicové schéma formou sedmipísmenového zápisu výběru. Je rozdělení symetrické? U kolika diagnostik je shoda v indikaci odlehklých bodů?

Data: Výsledek Amthaireova testu u 98 studentů [skóre]: 77 105 110 88 128 104
104 129 96 82 120 102 80 103 101 147 112 120
104 99 146 99 104 109 116 124 132 125
109 134 113 118 122 127 131 110 117

Klasické parametry :

Sloupec : S204
Průměr : 109,877551
Spodní mez : 106,5465237
Horní mez : 113,2085783
Rozptyl : 276,0467073
Směr. odchylka : 16,61465339
Dolní mez : 14,60309195
Horní mez : 19,27943153
Robustní směr.odch. : 13,3434
Šikmost : -0,02934327
Odchylka od 0 : Nevýznamná
Špičatost : 2,931690997
Odchylka od 3 : Nevýznamná
Polosuma : 110
Modus : 107,2803546
Geometrický průměr : 108,5938474
Harmonický průměr : 107,2606122

Robustní parametry :

Sloupec : S204
Medián : 109
IS spodní : 96,34209412
IS horní : 121,6579059
Medianová směr. odchylka : 6,3776681
Medianový rozptyl : 40,67465123
10% Průměr : 109,8977273
10% IS spodní : 106,4921516
10% IS horní : 113,3033029
10% Směr. odchylka : 12,39299338
10% Rozptyl : 153,5862849
20% Průměr : 110,0897436
20% IS spodní : 106,6204725
20% IS horní : 113,5590147
20% Směr. odchylka : 9,138962508
20% Rozptyl : 83,52063573
40% Průměr : 109,7586207
40% IS spodní : 106,6846925
40% IS horní : 112,8325489
40% Směr. odchylka : 5,03382548
40% Rozptyl : 25,3393989

Test normality Momentový

Sloupec : S204
Normalita : Přijata
Testové kritérium : 0,05592345827
Kritický kvantil $\chi^2(22)$: 5,991464547
p-hodnota : 0,9724255817

Exponenciální transformace

Zvolený parametr : -0,01022338867
Opravený průměr : 109,9613753
Spodní : 106,6272654
Horní : 113,2886613

31 ADSTAT: Opravený průměr: 104.55

Úloha H2.05 Klasifikace kvality kaolinu dle obsahu oxidu hlinitého

Pro keramickou výrobu je požadován kaolin o obsahu alespoň 37 % oxidu hlinitého ([18], s. 247). Rozhodněte, zda výběr splňuje základní požadavky, především normální rozdělení. Na jaký tvar rozdělení ukazují koeficienty šikmosti a špičatosti? Je nutné užít mocninnou transformaci? Jde o symetrické rozdělení? Splňuje tento výběr obsahu oxidu hlinitého Al_2O_3 v kaolinu požadavek normy 37 %? Kolik procent hodnot leží pod hodnotou 37 % Al_2O_3 a kolik nad ní?

Data: Obsah oxidu hlinitého v kaolinu [%]: 37.76 38.15 37.55 37.95 38.32 37.59 37.81 37.84
37.45 37.81 37.36 37.28

Klasické parametry :

Sloupec : H205
Průměr : 37,888125
Spodní mez : 37,6749211
Horní mez : 38,1013289
Rozptyl : 0,3496931452
Směr. odchylka : 0,5913485818
Dolní mez : 0,4758992219
Horní mez : 0,7820211254
Robustní směr.odch. : 0,607866
Šikmost : -0,5751321964
Odchylka od 0 : Nevýznamná
Špičatost : 3,722868194
Odchylka od 3 : Nevýznamná
Polosuma : 37,485
Modus : 37,70640152
Geometrický průměr : 37,88362632
Harmonický průměr : 37,87909886

Robustní parametry :

Sloupec : H205
Medián : 37,825
IS spodní : 36,42541022
IS horní : 39,22458978
Medianová směr. odchylka : 0,6862370946
Medianový rozptyl : 0,47092135
10% Průměr : 37,91857143
10% IS spodní : 37,71801804
10% IS horní : 38,11912481
10% Směr. odchylka : 0,4035437154
10% Rozptyl : 0,1628475302
20% Průměr : 37,91
20% IS spodní : 37,69779439
20% IS horní : 38,12220561
20% Směr. odchylka : 0,3135025853
20% Rozptyl : 0,09828387097
40% Průměr : 37,89555556
40% IS spodní : 37,67850266
40% IS horní : 38,11260845
40% Směr. odchylka : 0,1931575391
40% Rozptyl : 0,03730983493

Test normality Momentový

Sloupec : H205
Normalita : Přijata
Testové kritérium : 2,431067039
Kritický kvantil $\chi^2(22)$: 5,991464547
p-hodnota : 0,2965517559

Exponenciální transformace

Zvolený parametr : -0,2095489502
Opravený průměr : 37,94325355
Spodní mez : 37,72950424
Horní mez : 38,14201767

30 ADSTAT: Opravený průměr: 37.871

Úloha S2.19 Symetrie rozdělení výnosu žita a brambor u zemědělských farem

Porovnejte vyrovnanost 20 zemědělských farem z hlediska hektarových výnosů žita a brambor. Určete typ rozdělení obou výběrů. Která plodina má větší absolutní variabilitu? Kterou relativní míru variability využijete k požadovanému srovnání? U kolika diagnostik je shoda v indikaci odlehklých bodů? Zkonstruujte bariérově-číslicové schéma formou sedmipísmenového zápisu výběru. Vede graf maximální věrohodnosti k nutnosti použití transformace dat?

Data: S219a Výnos žita [q/ha] u zemědělských farem: 18.0 18.2 18.5 18.6 19.0 19.1 19.3 19.4
19.5 19.7 19.8 20.0 20.2 20.4 20.5 20.8 21.0 21.1
21.2 21.5

Klasické parametry :

Sloupec : S219a
Průměr : 19,79
Spodní mez : 19,30270164
Horní mez : 20,27729836
Rozptyl : 1,084105263
Směr. odchylka : 1,041203757
Dolní mez : 0,7918254556
Horní mez : 1,520752674
Robustní směr.odch. : 1,11195
Šikmost : -0,05233721
Odchylka od 0 : Nevýznamná
Špičatost : 1,941704158
Odchylka od 3 : Nevýznamná
Polosuma : 19,75
Modus : 19,67761905
Geometrický průměr : 19,76388281
Harmonický průměr : 19,73768785

Robustní parametry :

Sloupec : S219a
Medián : 19,75
IS spodní : 17,8811942
IS horní : 21,6188058
Medianová směr. odchylka : 0,8928735432
Medianový rozptyl : 0,7972231641
10% Průměr : 19,79444444
10% IS spodní : 19,26891773
10% IS horní : 20,31997116
10% Směr. odchylka : 0,8727165694
10% Rozptyl : 0,7616342105
20% Průměr : 19,80625
20% IS spodní : 19,24384242
20% IS horní : 20,36865758
20% Směr. odchylka : 0,7236384675
20% Rozptyl : 0,5236526316
40% Průměr : 19,80833333
40% IS spodní : 19,19621923
40% IS horní : 20,42044744
40% Směr. odchylka : 0,441605921
40% Rozptyl : 0,1950157895

Analýza malých výběrů

N : 20
Střední hodnota : 19,9
Spodní mez (5%) : 19,2934
Horní mez (95%) : 20,5066
Spodní mez (2.5%) : 19,1854
Horní mez (97.5%) : 20,6146
Pivotové rozpětí : 1,8

Test normality Momentový

Sloupec : S219a
Normalita : Přijata
Testové kritérium : 0,07200256672
Kritický kvantil $\chi^2(22)$: 5,991464547
p-hodnota : 0,9646390555

Exponenciální transformace

Zvolený parametr : -0,03867721558
Opravený průměr : 19,808957
Spodní : 19,31763453
Horní : 20,29149613

32 ADSTAT: Opravený průměr: Nelze, příliš velký výběr

