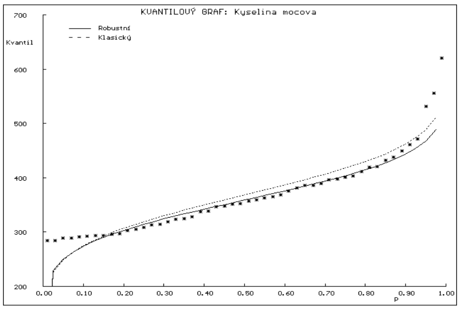
Vzorová úloha 3.1 *Analýza velkého výběru*

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?

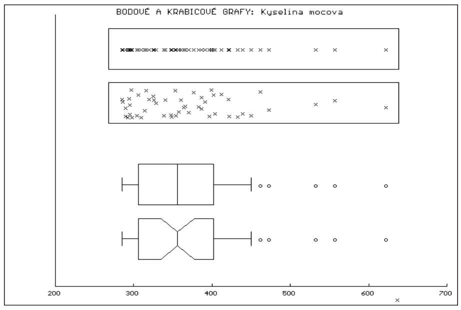
Řešení:

1. Průzkumová (exploratorní) analýza dat

Z grafických diagnostik průzkumové analýzy dat jsou uvedeny pouze čtyři nejdůležitější.

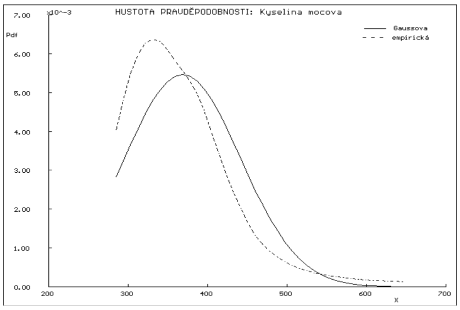


Obr. 3.2 Kvantilový graf

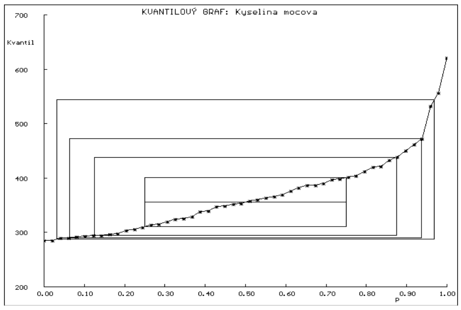


Obr. 3.3 Diagramy rozptýlení a krabicové grafy

Grafy ukazují, že výběrové rozdělení je asymetrické a silně sešikmené. 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.



Obr. 3.4 Graf hustoty pravděpodobnosti



Obr. 3.5 Graf rozptýlení s kvantily

2. Ověření základních předpokladů o výběru:

Předpoklad o normalitě rozdělení je zamítnut, protože hodnota testovacího kritéria χ^2_{exp} je vyšší než tabulkový kvantil $\chi^2_{1-\alpha}(2)$. Předpoklad nezávislosti je přijat, protože hodnota testovacího kritéria je nižší než tabulkový kvantil. Data nejsou homogenní, mimo modifikované vnitřní hradby B^*_D = 109.62 μmol/l, B^*_H = 602.38 μmol/l leží hodnota č. 17, a to 622.0 μmol/l. Po odstranění této odlehlé hodnoty by byl aritmetický průměr $\overline{x}$ = 363.29 μmol/l a směrodatná odchylka s = 63.875 μmol/l. Protože se však jedná o biochemická data, nelze však odlehlou hodnotu vyloučit z dat. Znamenalo by to zde totiž ztrátu důležité biochemické informace.

Základní předpoklady výběru **úlohy B2.25** (ADSTAT)

(a) Test normality: tabulkový kvantil $\chi^2_{1-\alpha}(2)$:	5.992
Odhad χ^2_{exp} statistiky:	29.199
Závěr: Předpoklad normality zamítnut na spočtené hladině významnosti α = 4.566E-07.	
(b) Test nezávislosti: tabulkový kvantil $t_{1-\alpha/2}(n+1)$:	2.008
Odhad von Neumannovy statistiky t_n:	1.0266
Závěr: Předpoklad nezávislosti přijat na spočtené hladině významnosti α = 0.155.	
(c) Detekce odlehlých bodů: metodou modifikované vnitřní hradby	
Dolní vnitřní hradba B^*_D:	109.62
Horní vnitřní hradba B^*_H:	602.38

Závěr: Ve výběru je 1 odlehlý bod, a to bod č. 17 (horní) o hodnotě 622.0.

(d) Opravené parametry výběru s vynechanými odlehlými hodnotami:

Odhad aritmetického průměru $\bar{x}$:	363.29
Odhad směrodatné odchylky s:	63.88
Odhad šikmosti $\hat{g}_1$:	0.97
Odhad špičatosti $\hat{g}_2$:	3.81

3. Mocninná transformace:
Při pokusu o transformaci dat poskytuje prostá mocninná transformace hodnotu opraveného průměru $\bar{x}_R$ = 350.91 μmol/l, zatímco Box - Coxova transformace $\bar{x}_R$ = 362.17 μmol/l, s odhadem šikmosti g_1 = 0.81 a špičatosti g_2 = 3.38, což je bližší parametrům normálního rozdělení. U mocninné transformace výpočet šikmosti a špičatosti selhává. Věrohodnější se zde proto jeví odhad, získaný metodou Boxovy-Coxovy transformace.

Box-Coxova transformace u dat výběru **úlohy B2.25** (ADSTAT)

Box-Coxova transformace:

Odhad optimálního exponentu λ	-2.13
Opravený odhad průměru původních dat $\bar{x}_R$	350.91

4. Odhady polohy, rozptylu a tvaru rozdělení:
Rozdělení souboru vykazuje mírné sešikmení. Soubor obsahuje jeden výrazně odlehlý bod. Mocninná transformace selhává, Box-Coxova transformace přináší zlepšení parametrů šikmosti a špičatosti souboru a je robustní vůči odlehlé hodnotě. Dobrým odhadem střední hodnoty se jeví také uřezané aritmetické průměry. Výpočet aritmetického průměru při vynechání odlehlého bodu podstatně zlepší variabilitu souboru a potvrzuje především robustnost retransformovaného průměru vůči odlehlé hodnotě.

Analýza jednorozměrného výběru dat **úlohy B2.25** (ADSTAT)

(1) Odhady klasických parametrů:

Odhad aritmetického průměru $\bar{x}$	368.46
Odhad směrodatné odchylky s	73.04
Odhad šikmosti $\hat{g}_1$	1.33
Odhad špičatosti $\hat{g}_2$	5.00
Dolní mez 95.0% intervalu spolehlivosti L_D	347.70
Horní mez 95.0% intervalu spolehlivosti L_H	389.22

(2) Odhady ostatních parametrů:

Odhad modu $\hat{x}_M$	293.00
Odhad polosumy $\hat{x}_P$	453.50

(3) Robustní odhady parametrů:

Medián $\tilde{x}_{0.5}$	356.00
Odhad směrodatné odchylky mediánu $s(\tilde{x}_{0.5})$	85.48
Dolní mez 95.0% intervalu spolehlivosti L_D	331.52
Horní mez 95.0% intervalu spolehlivosti L_H	380.48
Odhad 40%ního uřezaného průměru $\bar{x}(40\%)$	356.20
Odhad směrodatné odchylky $s(40\%)$	73.69
Odhad winsorizovaného průměru $\bar{x}_w(40\%)$	355.24
Odhad směrodatné odchylky $s_w(40\%)$	31.26

Dolní mez 95.0% intervalu spolehlivosti L_D	332.98
Horní mez 95.0% intervalu spolehlivosti L_H	379.42
Odhad *M*-odhadu střední hodnoty $\hat{\mu}_M$	358.17
Odhad směrodatné odchylky s_M	63.09
Dolní mez 95.0% intervalu spolehlivosti L_D	339.29
Horní mez 95.0% intervalu spolehlivosti L_H	377.05
(4) Hoggovy adaptivní odhady parametrů:	
Hoggův průměr $\hat{\mu}_M$	356.46
Odhad směrodatné odchylky s_M	72.72
Dolní mez 95.0% intervalu spolehlivosti L_D	335.79
Horní mez 95.0% intervalu spolehlivosti L_H	377.13

Vzorová úloha 3.2 *Analýza malého výběru dle Horna*

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

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

1. Pořádkové statistiky: seřadíme prvky od nejmenší do největší hodnoty

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: vyčíslíme hloubku pivotu

n = **8**, sudé

$$H = int \frac{\frac{n+1}{2}+1}{2}$$

int(2.75) ≈ 2

3. Pivoty: Dolní pivot $x_D = x_{(H)}$ $x_{(2)} = 0.49$

Horní pivot $x_H = x_{(n+1-H)}$ $x_{(7)} = 3.32$

4. Pivotová polosuma $P_L = \frac{x_D + x_H}{2}$ = 1.905

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

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

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

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

$$0.31 \le \mu \le 3.50$$

7. Závěr: Bodový odhad míry polohy je 1.91, míry rozptýlení 2.83 a intervalový odhad míry polohy je $0.31 \le \mu \le 3.50$.

(b) Užijeme také počítačové analýzy jednorozměrných dat: z průzkumové analýzy dat a ověření předpokladů o výběru plyne závěr, že rozdělení výběru pochází z Gaussova rozdělení, prvky výběru jsou nezávislé a ve výběru nejsou odlehlé body.

Analýza jednorozměrného výběru **úlohy B3.01** (ADSTAT)

(1) Test normality: tabulkový kvantil $\chi^2_{1-\alpha}(2)$:	5.992
Odhad χ^2_{exp} statistiky:	0.809
Závěr: Předpoklad normality přijat na spočtené hladině významnosti α = 0.6674.	
(2) Test nezávislosti: tabulkový kvantil $t_{1-\alpha/2}(n+1)$:	2.262
Odhad von Neumannovy statistiky t_n:	1.077
Závěr: Předpoklad nezávislosti přijat na spočtené hladině významnosti α = 0.155.	
(3) Detekce odlehlých bodů: metodou modifikované vnitřní hradby	
Závěr: Ve výběru nejsou odlehlé body.	
(4) Prostá mocninná transformace:	
Odhad optimálního exponentu λ	0.53
Odhad průměru transformovaných dat $\bar{y}$	1.246
Opravený odhad průměru původních dat $\bar{x}_R$	1.510
(5) Box-Coxova transformace:	
Odhad optimálního exponentu λ	0.53

Odhad průměru transformovaných dat $\bar{y}$	0.461
Opravený odhad průměru původních dat $\bar{x}_R$	1.510
(6) Odhady klasických parametrů:	
Odhad aritmetického průměru $\bar{x}$	1.736
Odhad směrodatné odchylky s	1.283
Odhad šikmosti $\hat{g}_1$	0.46
Odhad špičatosti $\hat{g}_2$	1.99
Dolní mez 95.0% intervalu spolehlivosti L_D	0.664
Horní mez 95.0% intervalu spolehlivosti L_H	2.809
(7) Robustní odhady parametrů:	
Medián $\tilde{x}_{0.5}$	1.545
Odhad směrodatné odchylky mediánu $s(\tilde{x}_{0.5})$	1.347
Odhad 40%ního uřezaného průměru $\bar{x}(40\%)$	1.545
Odhad směrodatné odchylky $s(40\%)$	1.347
Dolní mez 95.0% intervalu spolehlivosti L_D	-0.821
Horní mez 95.0% intervalu spolehlivosti L_H	3.912
Odhad *M*-odhadu střední hodnoty $\hat{\mu}_M$	1.679
Odhad směrodatné odchylky s_M	1.234
Dolní mez 95.0% intervalu spolehlivosti L_D	0.602
Horní mez 95.0% intervalu spolehlivosti L_H	2.756
(8) Hoggovy adaptivní odhady parametrů:	
Hoggův průměr $\hat{\mu}_M$	1.736
Odhad směrodatné odchylky s_M	1.283
Dolní mez 95.0% intervalu spolehlivosti L_D	0.664
Horní mez 95.0% intervalu spolehlivosti L_H	2.809

Vzorová úloha 3.3 *Test střední hodnoty (Test správnosti)*

Na úloze *Test správnosti koncentrace vápníku* ukážeme užití testu správnosti. Výrobce kontrolního komerčního materiálu uvádí koncentraci vápníku 2.20 mmol/1. Jsou naměřené výsledky správné ?

Data: Koncentrace vápníku [mmol/l] v komerčním materiálu:

2.26	2.16	2.18	2.15	2.23	2.25	2.19	2.18	2.16	2.20
2.19	2.22	2.19	2.21	2.25	2.29	2.26	2.15	2.18	

Řešení: Z exploratorní analýzy dat byla zjištěna mírná asymetrie, posun k nižším hodnotám, v horní části řady pořádkových statistik 3 podezřelé body. Ze základních předpokladů vyplývá, že data jsou homogenní, soubor neobsahuje odlehlé hodnoty. Data mají normální rozložení a jsou nezávislá.

Analýza jednorozměrných dat koncentrace vápníku (ADSTAT)

(1) Odhady klasických parametrů:	
Odhad aritmetického průměru $\bar{x}$	2.205
Odhad směrodatné odchylky s	0.041
Odhad šikmosti $\hat{g}_1$	0.44
Odhad špičatosti $\hat{g}_2$	2.12
Dolní mez 95.0% intervalu spolehlivosti L_D	2.185
Horní mez 95.0% intervalu spolehlivosti L_H	2.225
(2) Robustní odhady parametrů:	
Medián $\tilde{x}_{0.5}$	2.190
Odhad směrodatné odchylky mediánu $s(\tilde{x}_{0.5})$	0.056
Dolní mez 95.0% intervalu spolehlivosti L_D	2.161
Horní mez 95.0% intervalu spolehlivosti L_H	2.219
Odhad 40%ního uřezaného průměru $\bar{x}(40\%)$	2.194
Odhad směrodatné odchylky $s(40\%)$	0.040
Dolní mez 95.0% intervalu spolehlivosti L_D	2.171
Horní mez 95.0% intervalu spolehlivosti L_H	2.219
Odhad *M*-odhadu střední hodnoty $\hat{\mu}_M$	2.203
Odhad směrodatné odchylky s_M	0.041
Dolní mez 95.0% intervalu spolehlivosti L_D	2.183
Horní mez 95.0% intervalu spolehlivosti L_H	2.224
(3) Hoggovy adaptivní odhady parametrů:	
Hoggův průměr $\hat{\mu}_M$	2.205
Odhad směrodatné odchylky s_M	0.041
Dolní mez 95.0% intervalu spolehlivosti L_D	2.185
Horní mez 95.0% intervalu spolehlivosti L_H	2.225
(4) Prostá mocninná transformace:	
Odhad optimálního exponentu λ	-4.00
Opravený odhad průměru původních dat $\bar{x}_R$	2.204

Závěr: Pro 95%ní statistickou jistotu byly nalezeny následující intervalové odhady: pro aritmetický průměr $\bar{x}$ je interval $2.19 < \mu < 2.23$, pro medián $\tilde{x}_{0.5}$ pak $2.16 < \mu < 2.22$. Z uvedených intervalových odhadů vyplývá, že obsah vápníku 2.20 mmol/l leží v rozmezí zadané normy a naměřené výsledky jsou správné.

Vzorová úloha 3.4 *Test shodnosti středních hodnot*

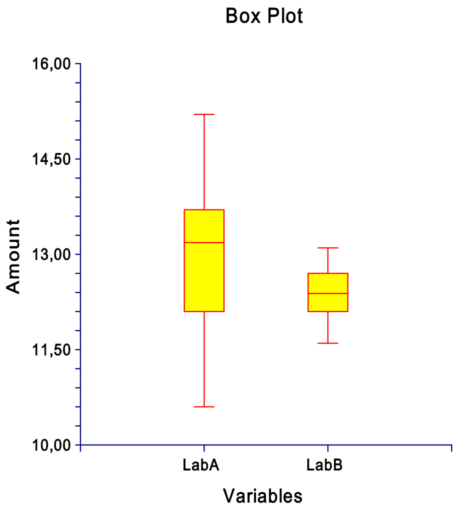
Na vzorové úloze *Porovnání práce dvou laborantek* ukážeme aplikaci testu shodnosti. Dvě laborantky prováděly analýzu koncentrace mukoproteinů v kontrolním vzorku. Určete, zda obě dospěly ke stejným výsledkům.Test shodnosti proveďte na hladině významnosti $\alpha = 0.05$.

Data: Stanovení koncentrace mukoproteinů v g/l laborantkou A a laborantkou B:

A: 11.6 12.1 13.2 12.1 10.6 13.3 13.7 14.4 15.2 13.6 13.7 12.4 12.5

B: 12.4 12.8 12.3 12.7 12.4 12.5 11.9 13.1 12.7 12.5 11.8 11.6 12.3

Řešení: Z ověření základních předpokladů pro jednotlivé výběry vyplývá, že data v obou výběrech jsou nezávislá, homogenní bez odlehlých bodů, test normality u obou výběrů prokázal Gaussovo rozdělení.



Obr. 3.6 Krabicový graf porovnání dvou výběrů, ***NCSS2000***

Porovnání dvou výběrů (ADSTAT)

(1) Klasické odhady parametrů:

Parametr	**Výběr 1**	**Výběr 2**	**Celkově**
Četnost	13	13	26
Průměr	12.954	12.385	12.669
Rozptyl	1.5160	0.1764	0.81243
Šikmost	-0.07	-0.34	-0.10
Špičatost	2.58	2.50	4.21

(2) Test homogenity rozptylu H$_0$: $s_1^2 = s_2^2$**.** Fisher-Snedecor F-test:

Počet stupňů volnosti $n_1 - 1$	12
Počet stupňů volnosti $n_2 - 1$	12
Tabulkový kvantil $F_{1-\alpha}(n_1 - 1, n_2 - 1)$	3.277
Experimentální F_{exp} statistika	8.594

Závěr: Rozptyly se považují za rozdílné, H$_0$ zamítnuta

(3) Test shody průměrů H$_0$: $\mu_1 = \mu_2$: Studentův t-test (pro různé rozptyly)

Tabulkový kvantil $t_{1-\alpha/2}(v)$	2.120
T_2 statistika	1.578

Závěr: Průměry se považují za shodné, H$_0$ přijata

Závěr: Na hladině významnosti $\alpha = 0.05$ potvrzuje oboustranný klasický test shodu středních hodnot obou výběrů i při významné odlišnosti obou rozptylů. Obě laborantky dosáhly stejných výsledků, i když každá s jinou variabilitou.

Vzorová úloha 3.5 *Párový test*

Na úloze **H3.22** *Zkoušení obsahu niklu v drátu a svarovém kovu* u dat ukážeme párový test, n = 45.

Řešení: Párový test **úlohy H3.22** (ADSTAT): Průměrný rozdíl $\overline{D}$ je 0.04356.

Rozptyl s_D^2	0.03004
Počet stupňů volnosti *n-1*	44
Tabulkový kvantil $t_{1-\alpha/2}(n-1)$	2.015
t_{exp} statistika	9.728

Závěr: Průměry se považují za rozdílné, H_0 zamítnuta na hladině významnosti 0.000.

Závěr: Párový test zamítl hypotézu o shodě obsahu niklu v drátu a svárovém kovu.