

Předmět: 1.3 Statistická analýza jednorozměrných dat

Přednášející: Prof. RNDr. Milan Meloun, DrSc.

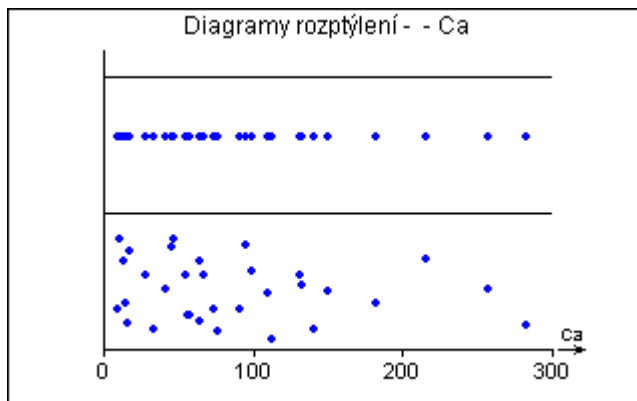
Úloha 1. Statistická analýza velkých výběrů

Zadání: Při analýze vzorků půd odebraných v porostech LS Nové Město na Moravě byl stanoven obsah přístupného vápníku a hořčíku ve výluhu chloridem amonným. Vybrány byly lokality na kterých jsou v současné době dospělé porosty se sníženým zakmeněním, ve kterých jsou připravovány bukové podsadby.

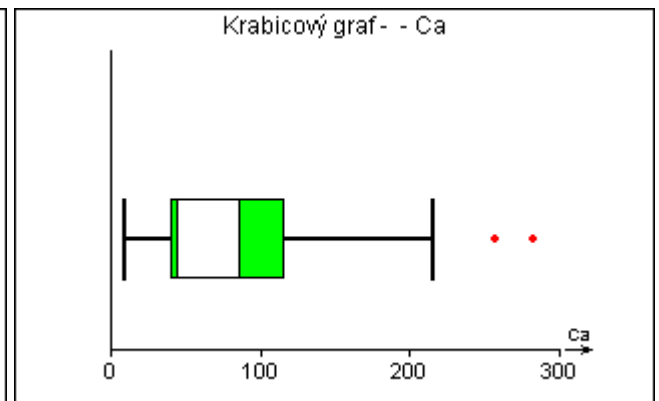
Data

revír	1	1	1	1	1	1	2	2	2	2	2	2
Ca (mg/kg)	257	76	45	42	132	58	74.0	216.0	95.0	110.0	99	34
Mg (mg/kg)	39.5	32.0	18.8	20.7	47.4	14.3	22.6	12.8	19.3	21.9	10.5	5.2
revír	2	2	2	2	2	2	2	2	2	2	2	3
Ca (mg/kg)	131	14	150	16	18	64	64	10.0	67.0	11	55.0	112.0
Mg (mg/kg)	18.4	7.4	22.9	10.2	12.8	21.5	10.7	4.6	9.6	7.1	5.7	9.5
revír	3	3	3	4	4	5	5	5				
Ca (mg/kg)	56.0	141	28	182	282	47	15	91				
Mg (mg/kg)	11.8	35.6	19.0	38.5	12.3	8.7	17.5					

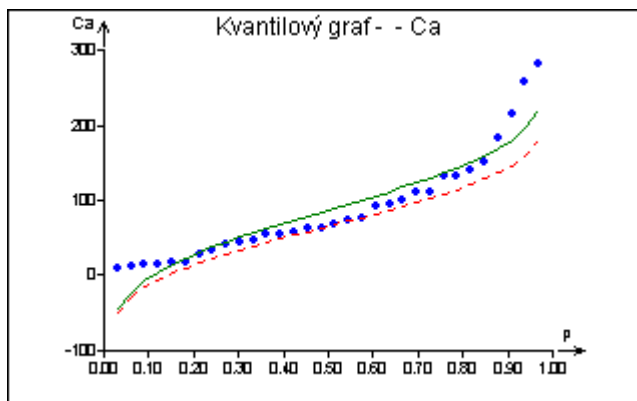
Užitý program: QCExpert, ADSTAT1.25



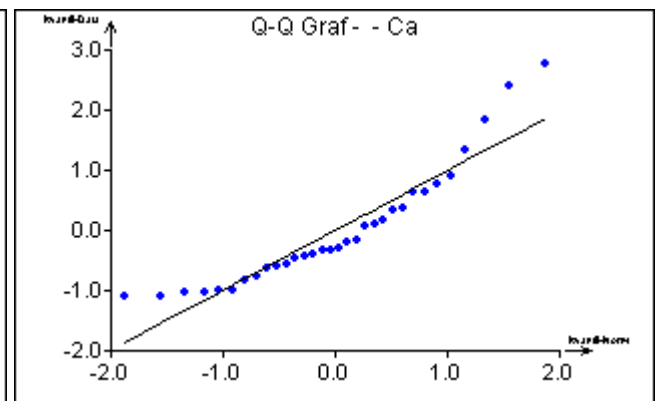
Obr. 1 – Diagramy rozptýlení



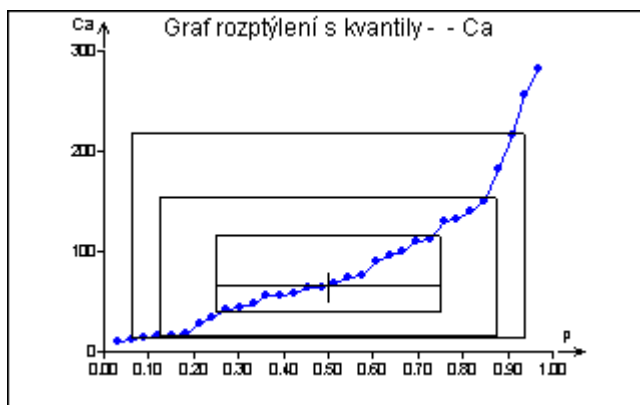
Obr. 2 – Krabicový graf



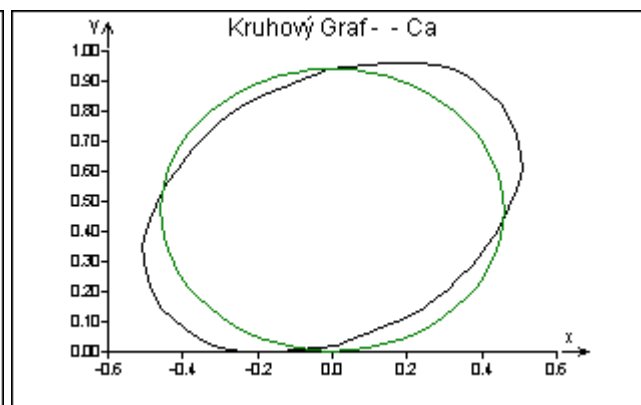
Obr. 3 – Kvantilový graf



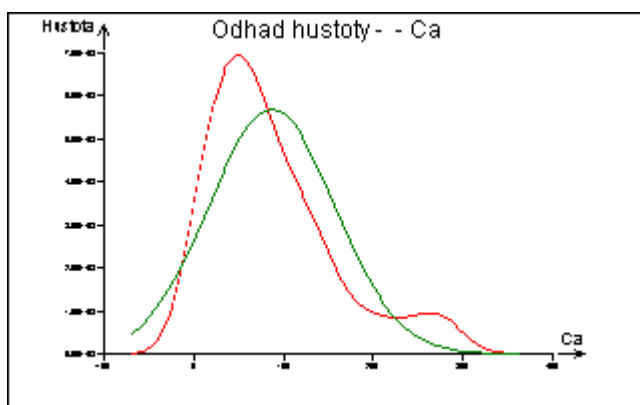
Obr. 4 – Q-Q graf



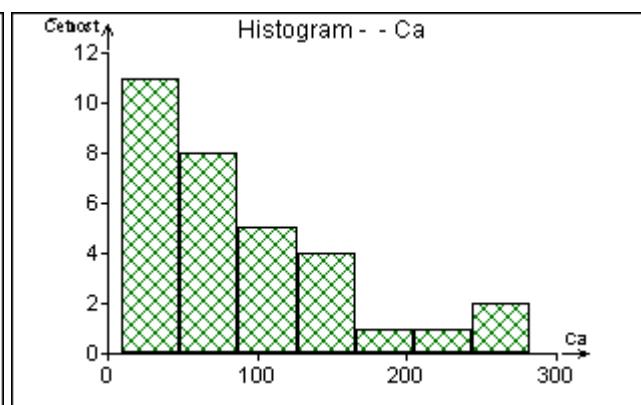
Obr. 5 – Graf rozptýlení s kvantily



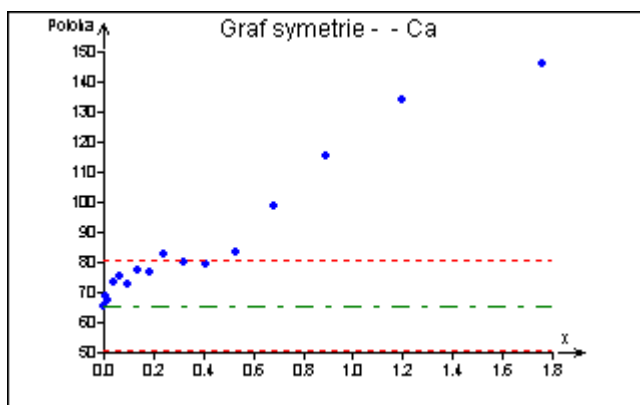
Obr. 6 – Kruhový graf



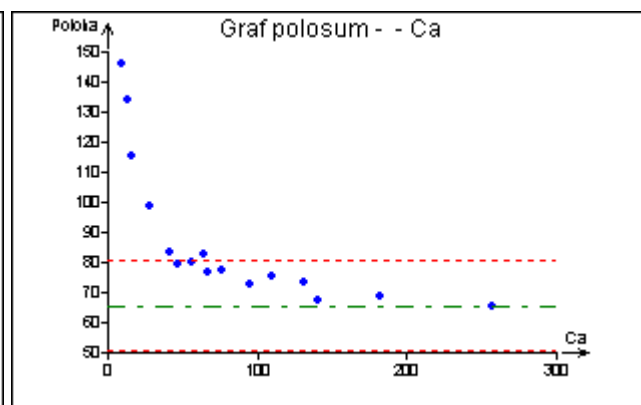
Obr. 7 – Odhad hustoty



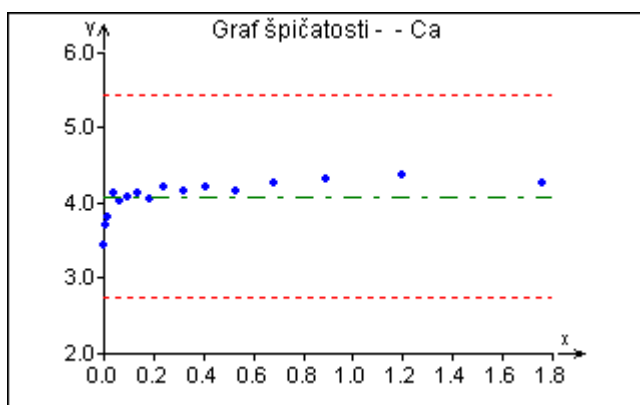
Obr. 8 - Histogram



Obr. 9 – Graf symetrie



Obr. 10 – Graf polosum



Obr. 11 – Graf špičatosti

Bodový graf (obr. 1) naznačuje roztažení do pravé strany, kde jsou dva až tři podezřelé body, které by mohly být odlehlými hodnotami. Medián bude lepším odhadem střední hodnoty než aritmetický průměr, protože rozdělení je zřejmě asymetrické.

V krabicovém grafu (obr. 2) jsou vpravo dva odlehlé body, rozdělení dat je levostranně asymetrické.

Kvantilový (obr. 3) a Q-Q graf (obr. 4) ukazují na odchylky od symetrie. Data „utíkají“ z přímky Q-Q grafu, resp. křivky kvantilového grafu na obou koncích.

Grafy polosum (obr. 10) a symetrie (obr. 9) ukazují na možnou asymetrii – část bodů v grafu polosum neleží na rovnoběžce s osou x a na grafu symetrie je část bodů mimo interval spolehlivosti mediánu.

Graf rozptýlení s kvantily (obr. 5) indikuje dva až tři možné odlehlé body v horní části a jeden až dva možné odlehlé body v dolní části. Obdélníky jsou posunuty dolů ve směru osy y, což svědčí o asymetrii rozdělení dat výběru.

Kruhový graf (obr. 6) je mírně vychýlený v diagonálním směru, naznačuje možnou asymetrii.

Odhad hustoty (obr. 7) a histogram (obr. 8) ukazují odchylku od symetrického rozdělení v pravé části, data jsou posunuta k levé straně. Zřejmě se jedná o asymetrické rozdělení, nebude možné použít aritmetický průměr.

Základní předpoklady (ADSTAT)

KLASICKÉ ODHADY PARAMETRŮ:

Průměr: 8.7250E+01 Rozptyl: 4.9383E+03
Směrodatná odchylka: 7.0273E+01 Šikmost: 1.2070E+00
Špičatost: 3.9114E+00

TEST NORMALITY:

Tabulkový kvantil $\chi^2(1-\alpha,2)$: 5.9915E+00
 χ^2 -statistika: 1.1871E+01
Závěr: **Předpoklad normality zamítnut**
Vypočtená hladina významnosti: 2.6435E-03

TEST NEZÁVISLOSTI:

Tabulkový kvantil $t(1-\alpha/2, n+1)$: 2.0345E+00
Test autokorelace: 5.1072E-01
Závěr: **Předpoklad nezávislosti přijat**
Vypočtená hladina významnosti: 3.0647E-01

PŘEDPOKLAD HOMOGENITY VÝBĚRU:

Aritmetický průměr: 8.7250E+01
Rozptyl: 4.9383E+03
Směrodatná odchylka: 7.0273E+01
Vnitřní meze: Spodní mez: -1.4048E+02 Horní mez: 2.9998E+02

MINIMÁLNÍ VELIKOST VÝBĚRU:

pro 25% relativní chybu směrodatné odchylky: $n = 13$
pro 10% relativní chybu směrodatné odchylky: $n = 74$
pro 5% relativní chybu směrodatné odchylky: $n = 292$

DETEKCE ODLEHLÝCH BODŮ:

Ve výběru nejsou odlehlé body

Z přehledu základních předpokladů vyplývá, že odchylka šikmosti je od hodnoty pro normální rozdělení velká – $g_1 = 1.2$, odchylka špičatosti tak velká není, $g_2 = 3.91$, hodnoty pro normální rozdělení jsou $g_1 = 0$, $g_2 = 3$.

Porovnání rozdělení (ADSTAT)

Hladina významnosti alfa: 0.050

Název výstupního souboru: Ca2.TXT

LINEARITA V GRAFU KVANTIL-KVANTIL (Q-Q):

Čís.	Rozdělení	Směrnice	Úsek	Korelační koeficient
0	Laplaceovo	5.0192E+01	8.7250E+01	9.3237E-01
1	Normální	6.7788E+01	8.7250E+01	9.3880E-01
2	Exponenciální	7.5021E+01	1.3967E+01	9.9285E-01
3	Rovnoměrné	2.2272E+02	-2.4112E+01	9.2192E-01
4	Lognormální	4.1342E+01	2.2873E+01	9.6550E-01
5	Gumbelovo	4.9169E+01	1.1474E+02	8.5308E-01

KVANTILOVÉ MÍRY

Kvantil	P	Spodní	Horní	Polorozptyl
Medián	0.5	6.5500E+01	-	-
Kvartil	0.25	3.6000E+01	1.3075E+02	9.4750E+01
Oktil	0.125	1.5125E+01	1.7775E+02	1.6263E+02
Sedecil	0.0625	1.1188E+01	2.5544E+02	2.4425E+02

Závěr EDA: Rozdělení dat výběru není symetrické, data pocházejí z exponenciálního rozdělení, projevuje se posun doleva, který je patrný z některých grafů průzkumové analýzy dat. Pro odhad střední hodnoty nelze použít aritmetický průměr, který se od mediánu významně liší. Průměr $\bar{x} = 87.25$, medián $\tilde{x}_{0.5} = 65.50$.

ANALÝZA DAT VÝBĚRU (QCE)

Klasické parametry :

Průměr:	87.25
Spodní mez:	61.91397525
Horní mez:	112.5860247
Rozptyl:	4938.258065
Směr. odchylka:	70.27274055
Šikmost:	1.207040653
Odchylka od 0:	Významná
Špičatost:	3.911408352
Odchylka od 3:	Nevýznamná
Polosuma:	146
Modus:	24.63636364

Robustní parametry :

Medián:	65.5
IS spodní:	31.68091605
IS horní:	99.31908395
Median. směr. odch.:	16.58193723
Medianový rozptyl:	274.9606423

TRANSFORMACE DAT (ADSTAT)

Prostá mocninná transformace:

Optimální hodnoty mocniny pro vybraná kritéria:

Optimální mocnina: 2.6667E-01 pro šikmost:	3.3657E-02
Optimální mocnina: -4.0000E-01 pro špičatost:	3.0673E+00
Optimální mocnina: 1.3333E-01 pro asymetrii:	3.0677E-03
Optimální mocnina: -1.3333E-01 pro asymetrii, rob.:	1.4132E-03
Optimální mocnina: -4.0000E+00 pro Hinkley-asymetrii:	2.4039E-05

Zvolená mocnina: 0.27

Průměr: 3.0840E+00
Rozptyl: 5.2025E-01
Směrodatná odchylka: 7.2128E-01
Šikmost: 3.3657E-02
Špičatost: 2.2978E+00
Opravený průměr: 6.8266E+01

Box-Coxova transformace:

Optimální hodnoty mocniny pro vybraná kritéria:

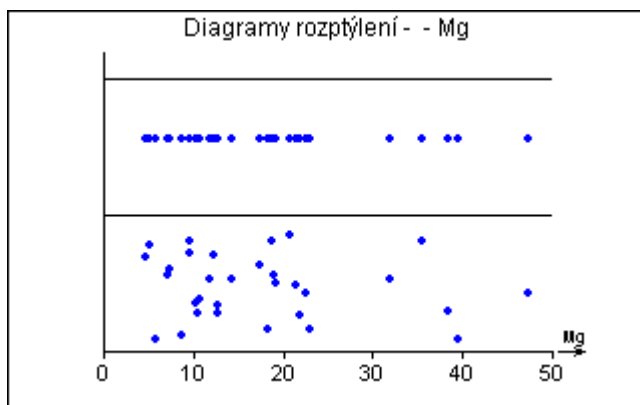
Optimální mocnina: 2.6667E-01 pro šikmost:	3.3657E-02
Optimální mocnina: -4.0000E-01 pro špičatost:	3.0673E+00
Optimální mocnina: 1.3333E-01 pro asymetrii:	3.0677E-03
Optimální mocnina: -1.3333E-01 pro asymetrii, rob.:	1.4132E-03
Optimální mocnina: -4.0000E+00 pro Hinkley-asymetrii:	6.0097E-06
Optimální mocnina: 2.6667E-01 pro věrohodnost:	-1.2848E+02

Zvolená mocnina: 0.27

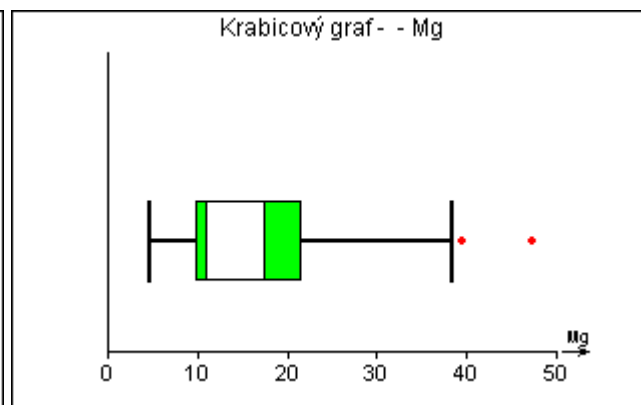
Průměr: 7.8152E+00
Rozptyl: 7.3160E+00
Směrodatná odchylka: 2.7048E+00
Šikmost: 3.3657E-02
Špičatost: 2.2978E+00
Opravený průměr: 6.8266E+01

Závěr: Transformace byla zcela oprávněná, retransformovaný průměr se přiblížil hodnotě mediánu. Obě transformace, mocninná i Box-Coxova, poskytují stejný výsledek pro opravený průměr.

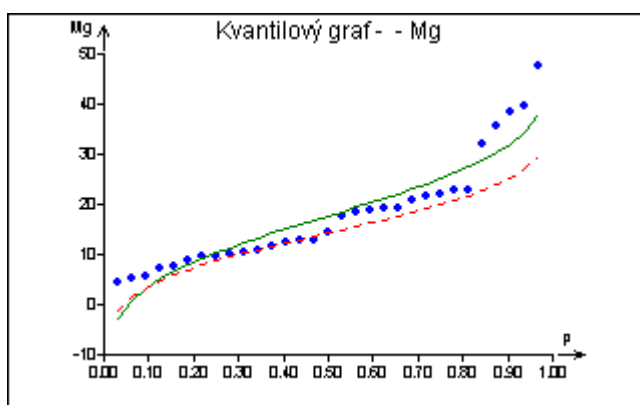
Výběr pochází z exponenciálního rozdělení, nejlepší odhad stření hodnoty získáme provedenou transformací dat. $\bar{x}_R = 68.27$, 95% interval spolehlivosti $L_D = 42.93$, $L_H = 93.61$.
--



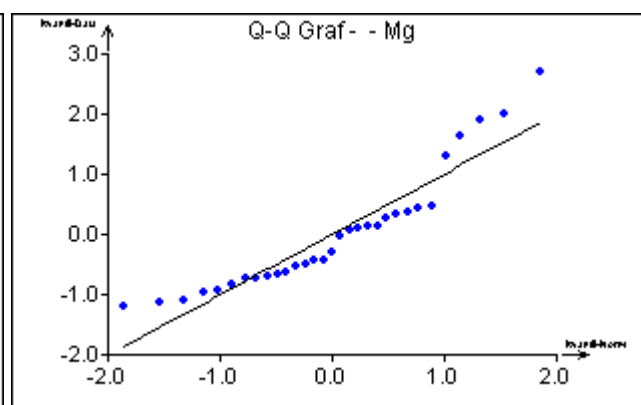
Obr. 12 – Diagramy rozptýlení



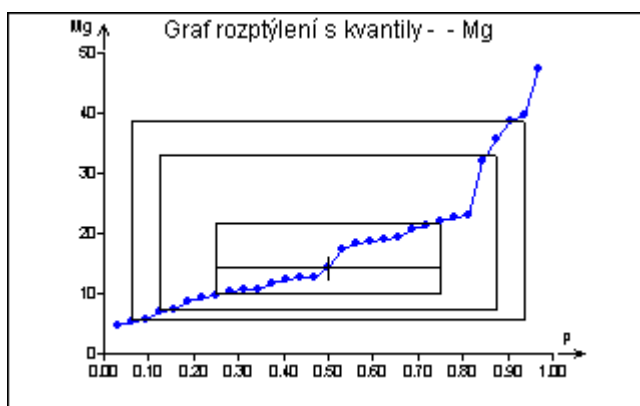
Obr. 13 – Krabicový graf



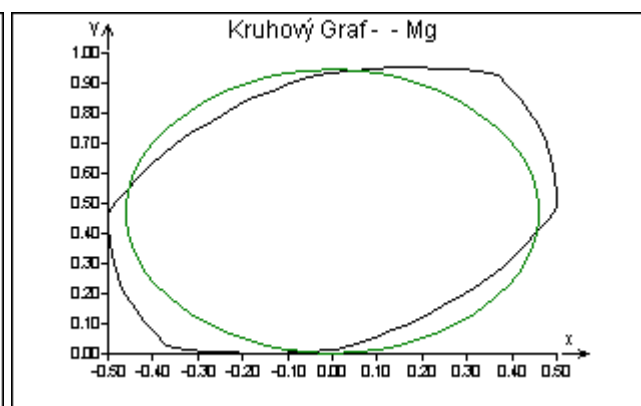
Obr. 14 – Kvantilový graf



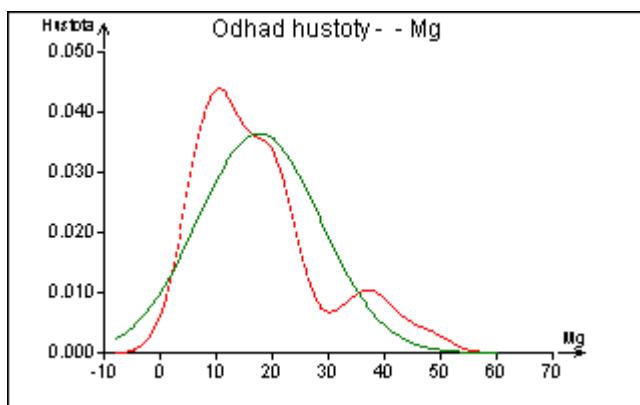
Obr. 15 – Q-Q graf



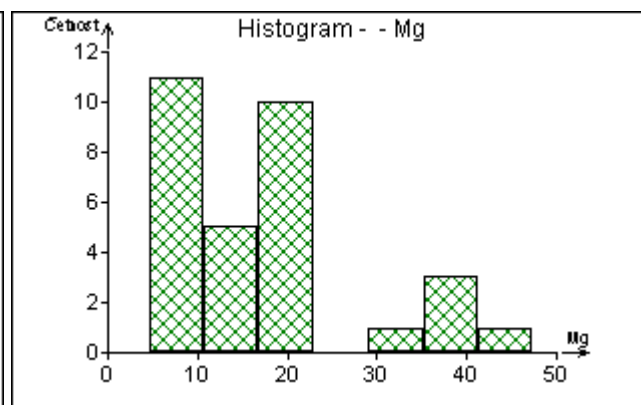
Obr. 16 – Graf rozptýlení s kvantily



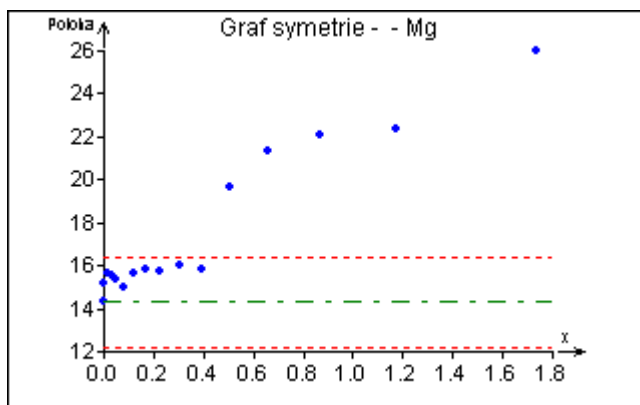
Obr. 17 – Kruhový graf



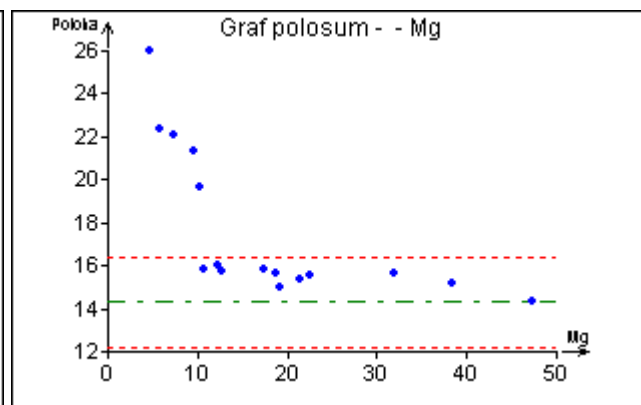
Obr. 18 – Odhad hustoty



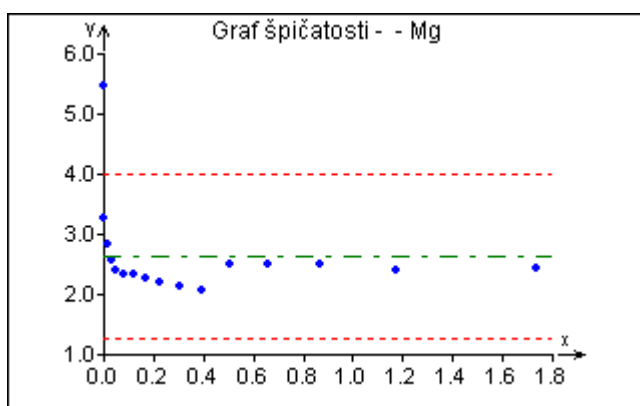
Obr. 19 - Histogram



Obr. 20 – Graf symetrie



Obr. 21 – Graf polosum



Obr. 22 – Graf špičatosti

Bodový graf (obr. 12) ukazuje větší koncentraci dat v levé části, v pravé části je jeden podezřelý bod, který by mohl být odlehlým bodem, v krabicovém grafu (obr. 13) jsou v pravé části dva odlehlé body, pravý „fous“ je výrazně delší než levý, hlavní část dat je posunuta doleva. Rozdělení dat je zřejmě asymetrické.

Kvantilový (obr. 14) a Q-Q graf (obr. 15) ukazují odchylky od symetrie na obou koncích. Vpravo na hoře je patrný jeden podezřelý bod.

Grafy polosum (obr. 21) a symetrie (obr. 20) naznačují odchylku od asymetrie – část bodů v grafu polosum neleží na rovnoběžce s osou x a na grafu symetrie je část bodů mimo interval spolehlivosti mediánu.

Graf rozptýlení s kvantily (obr. 16) indikuje dva až tři možné odlehlé body v horní části a jeden až dva možné odlehlé body v dolní části. Je patrný mírný posun obdélníků dolů ve směru osy y.

Kruhový graf (obr. 17) je vychýlený v diagonálním směru, což naznačuje možnou asymetrii v rozdělení dat výběru.

Odhad hustoty (obr. 18) a histogram (obr. 19) – většina dat je koncentrována v levé části histogramu, to způsobuje odchylky od symetrie patrné v grafu odhadu hustoty.

Základní předpoklady (ADSTAT)

KLASICKÉ ODHADY PARAMETRŮ:

Průměr: 1.7703E+01 Rozptyl: 1.1993E+02

Směrodatná odchylka: 1.0951E+01 Šikmost: 1.1044E+00

Špičatost: 3.5024E+00

TEST NORMALITY:

Tabulkový kvantil $\chi^2(1-\alpha,2)$: 5.9915E+00

χ^2 -statistika: 8.6135E+00

Závěr: **Předpoklad normality zamítnut**

Vypočtená hladina významnosti: 1.3477E-02

TEST NEZÁVISLOSTI:

Tabulkový kvantil $t(1-\alpha/2, n+1)$: 2.0369E+00

Test autokorelace: 1.4144E+00

Závěr: **Předpoklad nezávislosti přijat**

Vypočtená hladina významnosti: 8.3455E-02

PŘEDPOKLAD HOMOGENITY VÝBĚRU:

Aritmetický průměr: 1.7703E+01

Rozptyl: 1.1993E+02

Směrodatná odchylka: 1.0951E+01

Vnitřní meze: Spodní mez: -1.5280E+01 Horní mez: 4.6880E+01

MINIMÁLNÍ VELIKOST VÝBĚRU:

pro 25% relativní chybu směrodatné odchylky: $n = 11$

pro 10% relativní chybu směrodatné odchylky: $n = 64$

pro 5% relativní chybu směrodatné odchylky: $n = 251$

DETEKCE ODLEHLÝCH BODŮ:

Bod číslo 5 (horní): 4.7400E+01

Počet odlehlých bodů: 1

Parametry s vynechanými odlehlými hodnotami:

Průměr: 1.6713E+01

Rozptyl: 9.2638E+01

Směrodatná odchylka: 9.6248E+00

Šikmost: 9.8450E-01

Špičatost: 3.2917E+00

Z přehledu základních předpokladů vyplývá, že odchylka šikmosti je od hodnoty pro normální rozdělení velká – $g_1 = 1.1$, odchylka špičatosti tak velká není, $g_2 = 3.5$, hodnoty pro normální rozdělení jsou $g_1 = 0$, $g_2 = 3$.

Z výsledků analýz půd nelze hodnoty vynechávat, došlo by ke zkreslení.

Porovnání rozdělení (ADSTAT)

Hladina významnosti α : 0.050

Název výstupního souboru: PHKCL1.TXT

LINEARITA V GRAFU KVANTIL-KVANTIL (Q-Q):

Čís.	Rozdělení	Směrnice	Úsek	Korelační koeficient
0	Laplaceovo	7.8453E+00	1.7703E+01	9.3351E-01
1	Normální	1.0604E+01	1.7703E+01	9.4172E-01
2	Exponenciální	1.1618E+01	6.3614E+00	9.8512E-01
3	Rovnoměrné	3.4866E+01	2.7001E-01	9.2633E-01
4	Lognormální	6.3924E+00	7.7616E+00	9.5407E-01
5	Gumbelovo	7.7442E+00	2.2030E+01	8.6123E-01

KVANTILOVÉ MÍRY

Kvantil	P	Spodní	Horní	Polorozptyl
Medián	0.5	1.4300E+01	-	-
Kvartil	0.25	9.6000E+00	2.1900E+01	1.2300E+01
Oktil	0.125	7.1000E+00	3.5600E+01	2.8500E+01
Sedecil	0.0625	5.2000E+00	3.9500E+01	3.4300E+01

Závěr EDA: Rozdělení dat výběru není symetrické, data pocházejí z exponenciálního rozdělení, projevuje se posun doleva, který je patrný z některých grafů průzkumové analýzy dat. Pro odhad střední hodnoty nelze použít aritmetický průměr, který se od mediánu liší. Průměr $\bar{x} = 17.7$, medián $\tilde{x}_{0.5} = 14.3$.

ANALÝZA DAT VÝBĚRU (QCE)

Klasické parametry:

Průměr:	17.70322581
Spodní mez:	13.6863344
Horní mez:	21.72011722
Rozptyl:	119.9263226
Směr. odchylka:	10.95108774
Šikmost:	1.104356701
Odchylka od 0:	Významná
Špičatost:	3.502388062
Odchylka od 3:	Nevýznamná
Polosuma:	26
Modus:	7.918951613

Robustní parametry:

Medián:	14.3
IS spodní:	8.9858261
IS horní:	19.6141739
Median. směr. odch.:	2.602088612
Medianový rozptyl:	6.770865143

TRANSFORMACE DAT (ADSTAT)

Prostá mocninná transformace:

Optimální hodnoty mocniny pro vybraná kritéria:

Optimální mocnina:	-1.3333E-01	pro šikmost:	1.5403E-01
Optimální mocnina:	-6.6667E-01	pro špičatost:	3.0160E+00
Optimální mocnina:	-2.6667E-01	pro asymetrii:	4.8467E-03
Optimální mocnina:	-1.3333E-01	pro asymetrii, rob.:	6.1545E-03
Optimální mocnina:	-4.0000E+00	pro Hinkley-asymetrii:	7.8923E-04

Zvolená mocnina: -0.13

Průměr:	7.0020E-01
Rozptyl:	3.2509E-03
Směrodatná odchylka:	5.7017E-02
Šikmost:	1.5403E-01
Špičatost:	2.3388E+0
Opravený průměr:	1.4482E+01

Box-Coxova transformace:

Optimální hodnoty mocniny pro vybraná kritéria:

Optimální mocnina:	-1.3333E-01	pro šikmost:	1.5403E-01
Optimální mocnina:	-6.6667E-01	pro špičatost:	3.0160E+00
Optimální mocnina:	-2.6667E-01	pro asymetrii:	4.8467E-03
Optimální mocnina:	-1.3333E-01	pro asymetrii, rob.:	6.1545E-03
Optimální mocnina:	-4.0000E+00	pro Hinkley-asymetrii:	1.9731E-04
Optimální mocnina:	-1.3333E-01	pro věrohodnost:	-6.8417E+01

Zvolená mocnina: -0.13

Průměr: 2.2485E+00

Rozptyl: 1.8286E-01

Směrodatná odchylka: 4.2762E-01

Šikmost: -1.5403E-01

Špičatost: 2.3388E+00

Opravený průměr: 1.4482E+01

Závěr: Transformace byla zcela oprávněná, retransformovaný průměr ($\bar{x}_R = 14.48$) se přiblížil hodnotě mediánu ($\tilde{x}_{0.5} = 14.3$). Obě transformace, mocninná i Box-Coxova, poskytují stejný výsledek pro opravený průměr.

Výběr pochází z exponenciálního rozdělení, nejlepší odhad stření hodnoty získáme provedenou transformací dat. $\bar{x}_R = 14.48$, 95% interval spolehlivosti $L_D = 10.46$, $L_H = 18.5$.
