

Popisná statistika

pro to jednoduché metody, že hlavně není potřeba žádná teorie.

Patří sem např. tabulka četností, histogram, výběrový průměr, výběrový průměr, rozptyl, výběrová odchylka, výběrová směrodatná odchylka, křivkový diagram, rozptylový diagram atd.

Tabulka četností: máme naměřené hodnoty $x_1, x_2, \dots, x_n$

upřesňujeme je: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

necht $a_1 < a_2 < \dots < a_m$ jsou všechny navzájem různé hodnoty souboru

$m_1, m_2, \dots, m_m$ jsou jejich četnosti; $n = \sum_{j=1}^m m_j$

vytvoříme tabulku četností

a_1	a_2	...	a_m
m_1	m_2	...	m_m

nebo relativních četností

a_1	a_2	...	a_m
m_1/n	m_2/n	...	m_m/n

(relativní četnost $\frac{m_i}{n}$ je odhadem pravděpodobnosti, že náhodná veličina X máji hodnotu a_i)

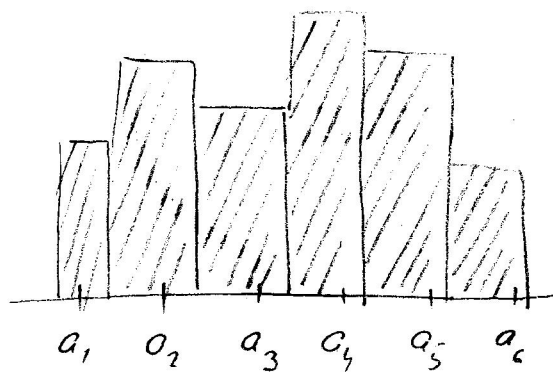
nebo kumulativních četností

a_1	a_2	...	a_m
N_1	N_2	...	N_m

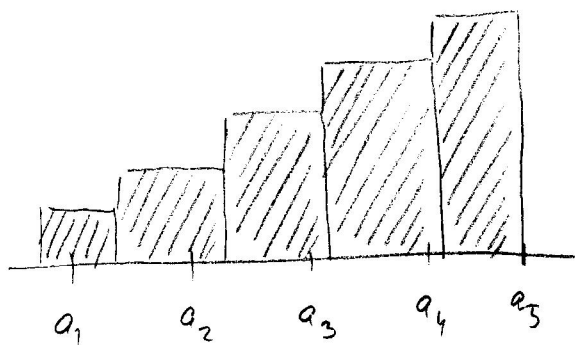
tedy $N_j = \sum_{i=1}^j m_i$

(relativní kumulativní četnost $\frac{N_j}{n}$ je odhadem pravděpodobnosti $P(X \leq j) = P(X < j+1) = F(j+1)$... distribuční funkce)

Histogram je grafická podoba tabulek četosti



histogram relativních četostí
je odhadem hustoty



histogram kumulativních
relativních četostí je
odhadem distribuční
funkce

Výběrový průměr $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ je odhadem střední hodnoty
 $EX = \sum_i x_i P(X=x_i)$

Výběrový medián je "prostřední" hodnota v uspořádaném

souboru, tj.
$$\tilde{x} = \begin{cases} x_{(\frac{n+1}{2})} & \text{po } n \text{ liché} \\ \frac{1}{2} (x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{po } n \text{ sudé} \end{cases}$$

je odhadem medianu $\tilde{\mu}$, který je definován vztahem

$$F(\tilde{\mu}) \leq \frac{1}{2}, \quad F(\tilde{\mu}+0) \geq \frac{1}{2}$$

limita zprava

Medan je najčelnější hodnota

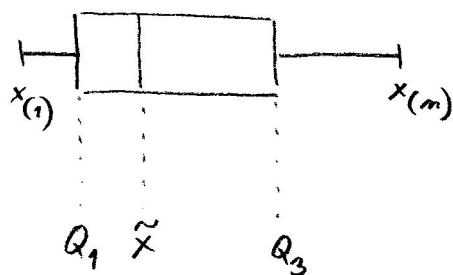
Výběrová rozpětí $R = X_{(n)} - X_{(1)}$

Průměrná odchylka $d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$

Výběrový rozptyl $M^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ je odhadem rozptylu $E(X - EX)^2$

odmocnina z něj je výběrová směrodatná odchylka.

Krácový diagram vypadá nějak takhle (není to jediná možnost):



kde $\tilde{x}$ je median

Q_1 je dolní kvartil $x_{0,25}$

Q_3 je horní kvartil $x_{0,75}$

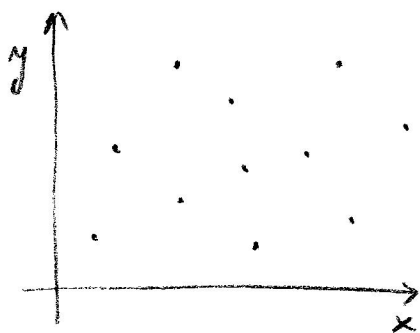
(obecně p -tý kvantil $x_p = \begin{cases} x_{(\lceil np \rceil + 1)} & \text{pro } np \neq \lceil np \rceil \\ \frac{1}{2} (x_{(np)} + x_{(np+1)}) & \text{pro } np = \lceil np \rceil \end{cases}$ celí)

$R_Q = Q_3 - Q_1$ je kvartilové rozpětí

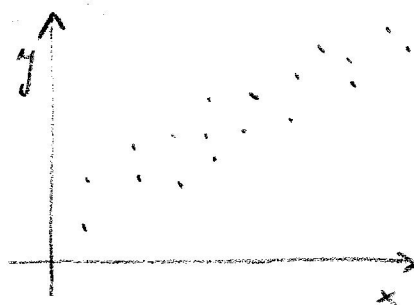
Rozptylový diagram se používá ke měření závislosti mezi veličinami X a Y

máme poradení $x_1, x_2, \dots, x_n$ a $y_1, y_2, \dots, y_n$

vypočítáme dvojice $(x_1, y_1), \dots, (x_n, y_n)$ a naneseme je do grafu



takže to vypadá,
že jsou X a Y závislé



tedy je závislost
(přímá úměrnost)

Jste-li tabulkám čísel: Pokud je soubor příliš mnoha různých hodnot, rozdělíme jejich obor na intervaly (třídy):

$$(-\infty, t_0), (t_0, t_1), (t_1, t_2), \dots, (t_m, \infty)$$

a každé třídě zvolíme střed $a_j = \frac{t_{j-1} + t_j}{2}$

a_0 se volí tak, aby to byl střed intervalu (a_0, a_1)

$$a_{m+1} \quad \text{---} \quad t_m \quad \text{---} \quad a_{m+1}$$

učiníme třídní četnosti n_j jako počet hodnot x_i , které patří do j -tého intervalu

sestavíme tabulku třídních četností

Příklad : Quicksort - algoritmus M1 (přesně je malý lež' bym')

(5)

P-čas je čas potřebný na výběr pivotu

Q-čas je čas spotřebovaný rekursivním quicksortem

I-čas je čas na dořídování malých polí Insertionsortem

Čas je celkový čas třídění

1 000 000 prvků (MN, M1, M3)

čas	I-čas	P-čas	Čas	M1	Výměny	Porovnání	Q-čas	I-čas	P-čas	Čas	M3	Výměny	Porovnání
76	1.19	0.00	7.95	1.	4100362	23391548	6.84	1.17	0.00	8.01	1.	4094311	19850507
88	1.18	0.00	8.06	2.	4109733	23693092	6.94	1.18	0.00	8.12	2.	4126967	19089632
11	1.18	0.00	8.29	3.	4114931	22602791	6.77	1.19	0.00	7.96	3.	4098052	19640826
12	1.19	0.00	8.31	4.	4051794	24107207	7.03	1.18	0.00	8.21	4.	4093828	20067844
37	1.19	0.00	8.06	5.	4083053	23268242	6.88	1.19	0.00	8.07	5.	4064715	20258908
37	1.18	0.00	8.05	6.	4054783	24534597	7.10	1.18	0.00	8.28	6.	4108642	19373468
80	1.18	0.00	7.98	7.	4106845	22960107	6.84	1.17	0.00	8.01	7.	4090494	19629649
88	1.18	0.00	8.06	8.	4103625	22791147	6.78	1.17	0.00	7.95	8.	4082225	19802586
94	1.19	0.00	8.13	9.	4069077	23652237	6.87	1.18	0.00	8.05	9.	4100272	19489754
91	1.17	0.00	8.08	10.	4087163	23237522	6.77	1.18	0.00	7.95	10.	4115292	19326990
25	1.19	0.00	8.44	11.	4113420	23055301	6.82	1.18	0.00	8.00	11.	4105098	19565539
88	1.18	0.00	8.06	12.	4068719	24453598	7.07	1.19	0.00	8.26	12.	4111603	19424510
93	1.18	0.00	8.11	13.	4137795	22489281	6.78	1.18	0.00	7.96	13.	4108325	19427871
91	1.17	0.00	8.08	14.	4118836	22799494	6.79	1.18	0.00	7.97	14.	4107414	19406223
02	1.18	0.00	8.20	15.	4120516	22832905	6.79	1.18	0.00	7.97	15.	4082516	19754591
37	1.18	0.00	8.05	16.	4115326	22636947	6.75	1.19	0.00	7.94	16.	4113465	19339860
09	1.18	0.00	8.27	17.	4124258	22927205	6.78	1.18	0.00	7.96	17.	4110898	19387273
96	1.18	0.00	8.14	18.	4106038	22846174	6.80	1.18	0.00	7.98	18.	4115370	19328065
93	1.18	0.00	8.11	19.	4060726	24266218	7.00	1.18	0.00	8.18	19.	4103859	19486129
88	1.17	0.00	8.05	20.	4088484	23539988	6.93	1.17	0.00	8.10	20.	4082030	20203015
84	1.18	0.00	8.02	21.	4125048	22669153	6.83	1.19	0.00	8.02	21.	4099764	19585975
02	1.19	0.00	8.11	22.	4091661	23873178	6.93	1.18	0.00	8.11	22.	4111102	19445783
37	1.18	0.00	8.05	23.	4109483	22916798	6.84	1.19	0.00	8.03	23.	4096939	19678438
12	1.18	0.00	8.30	24.	4129074	22315426	6.63	1.17	0.00	7.80	24.	4114048	19278849
99	1.18	0.00	8.17	25.	4071611	23853352	7.03	1.18	0.00	8.21	25.	4103438	19616061
94	1.18	0.00	8.13	Ø	4098494	23268540	6.86	1.18	0.00	8.04	Ø	4101627	19578334

Podane analízat posledni skope - čas

vybraný soubor: $n = 25$

7,80; 7,94; 7,95; 7,95; 7,96; 7,96; 7,96; 7,97; 7,97; 7,98;
 8,00; 8,01; 8,01; 8,02; 8,03; 8,05; 8,07; 8,10; 8,11; 8,12; 8,18;
 8,21; 8,21; 8,26; 8,28

min = 7,80, max = 8,28, rozsah = 0,48

průměr $\bar{x} = 8,04$

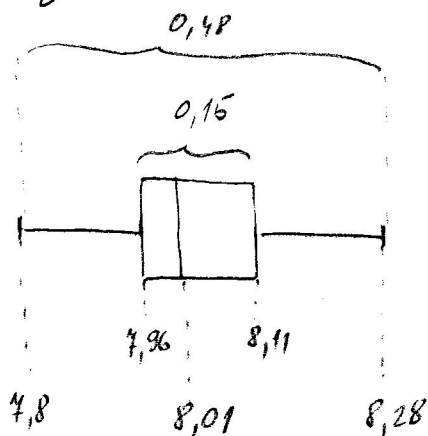
median $\tilde{x} = 8,01$

dolní kvartil $x_{0,25} = x_{(\lfloor 25 \cdot 0,25 \rfloor + 1)} = x_{(4)} = 7,96$

horní kvartil $x_{0,75} = x_{(\lfloor 25 \cdot 0,75 \rfloor + 1)} = x_{(19)} = 8,11$

kvartilové rozpětí $R_Q = 0,15$

boxplot diagram



modus : 7,96

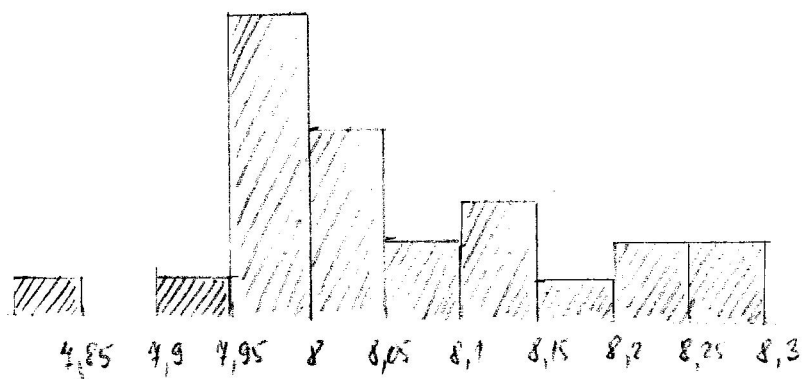
průměrná odchylka $d = 0,0912$

směrodatná odchylka $m = 0,068$

V souboru je poměrně mnoho různých hodnot, uděláme tedy
tabulku četností

interval :	četnost :	relativní četnost
$(-\infty; 4,85)$	1	0,04
$[4,85; 4,9)$	0	0
$[4,9; 4,95)$	1	0,04
$[4,95; 8)$	8	0,32
$[8; 8,05)$	5	0,2
$[8,05; 8,1)$	2	0,08
$[8,1; 8,15)$	3	0,12
$[8,15; 8,2)$	1	0,04
$[8,2; 8,25)$	2	0,08
$[8,25; 8,3)$	2	0,08
$[8,3; \infty)$	0	0
	25	1

histogram



data file: quicksort.sta [25 cases with 10 variables]

	MN	M1	M3	M3-5	M3-7	M3-11	M3-17	M3-5-11	M3-5-17
	1 VAR1	2 VAR2	3 VAR3	4 VAR4	5 VAR5	6 VAR6	7 VAR7	8 VAR8	9 VAR9
1	7.950	8.010	7.710	7.530	7.580	7.520	7.500	7.510	7.530
2	8.060	8.120	7.580	7.520	7.420	7.460	7.480	7.450	7.470
3	8.290	7.960	7.630	7.520	7.530	7.520	7.470	7.520	7.520
4	8.310	8.210	7.680	7.670	7.510	7.470	7.460	7.480	7.430
5	8.060	8.070	7.740	7.540	7.500	7.540	7.470	7.500	7.330
6	8.050	8.280	7.670	7.640	7.480	7.520	7.480	7.520	7.490
7	7.980	8.010	7.630	7.560	7.520	7.530	7.460	7.480	7.480
8	8.060	7.950	7.670	7.530	7.580	7.490	7.470	7.520	7.430
9	8.130	8.050	7.640	7.530	7.540	7.500	7.440	7.460	7.450
10	8.080	7.950	7.600	7.440	7.500	7.530	7.480	7.440	7.420
11	8.440	8.000	7.610	7.460	7.470	7.600	7.420	7.490	7.560
12	8.060	8.260	7.610	7.610	7.520	7.490	7.470	7.480	7.440
13	8.110	7.960	7.630	7.680	7.650	7.470	7.470	7.520	7.420
14	8.080	7.970	7.610	7.660	7.550	7.480	7.440	7.510	7.470
15	8.200	7.970	7.680	7.560	7.420	7.470	7.430	7.460	7.480
16	8.050	7.940	7.630	7.520	7.500	7.500	7.460	7.530	7.540
17	8.270	7.960	7.670	7.520	7.440	7.460	7.410	7.440	7.470
18	8.140	7.980	7.610	7.500	7.500	7.500	7.470	7.470	7.470
19	8.110	8.180	7.670	7.540	7.550	7.550	7.490	7.450	7.480
20	8.050	8.100	7.710	7.520	7.500	7.510	7.480	7.490	7.490
21	8.020	8.020	7.550	7.600	7.550	7.460	7.480	7.530	7.500
22	8.110	8.110	7.600	7.460	7.490	7.510	7.510	7.500	7.440
23	8.050	8.030	7.620	7.630	7.560	7.490	7.500	7.520	7.420
24	8.300	7.800	7.580	7.530	7.510	7.380	7.470	7.490	7.460
25	8.170	8.210	7.660	7.510	7.520	7.410	7.480	7.520	7.400

	10 VAR10
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	

Popis: naměřené časy ($n \cdot s$) různých verzí algoritmu Quicksort
(na postupnostech délky 1 000 000 prvků)

MN - máhodná volba prvku.

M1 - prvok je vždy levý krajní

M3 - -1- median ze 3

M3-5 - -1- 5, při kladných postupnostech median ze 3

M3-7 - -1- 7 -1- 3

M3-11 - -1- 11 -1- 3

M3-17 - -1- 17 -1- 3

M3-5-11 - -1- 11 -1- 5, pak ze 3

M3-5-17 - -1- 17 -1- 5 -1- 3

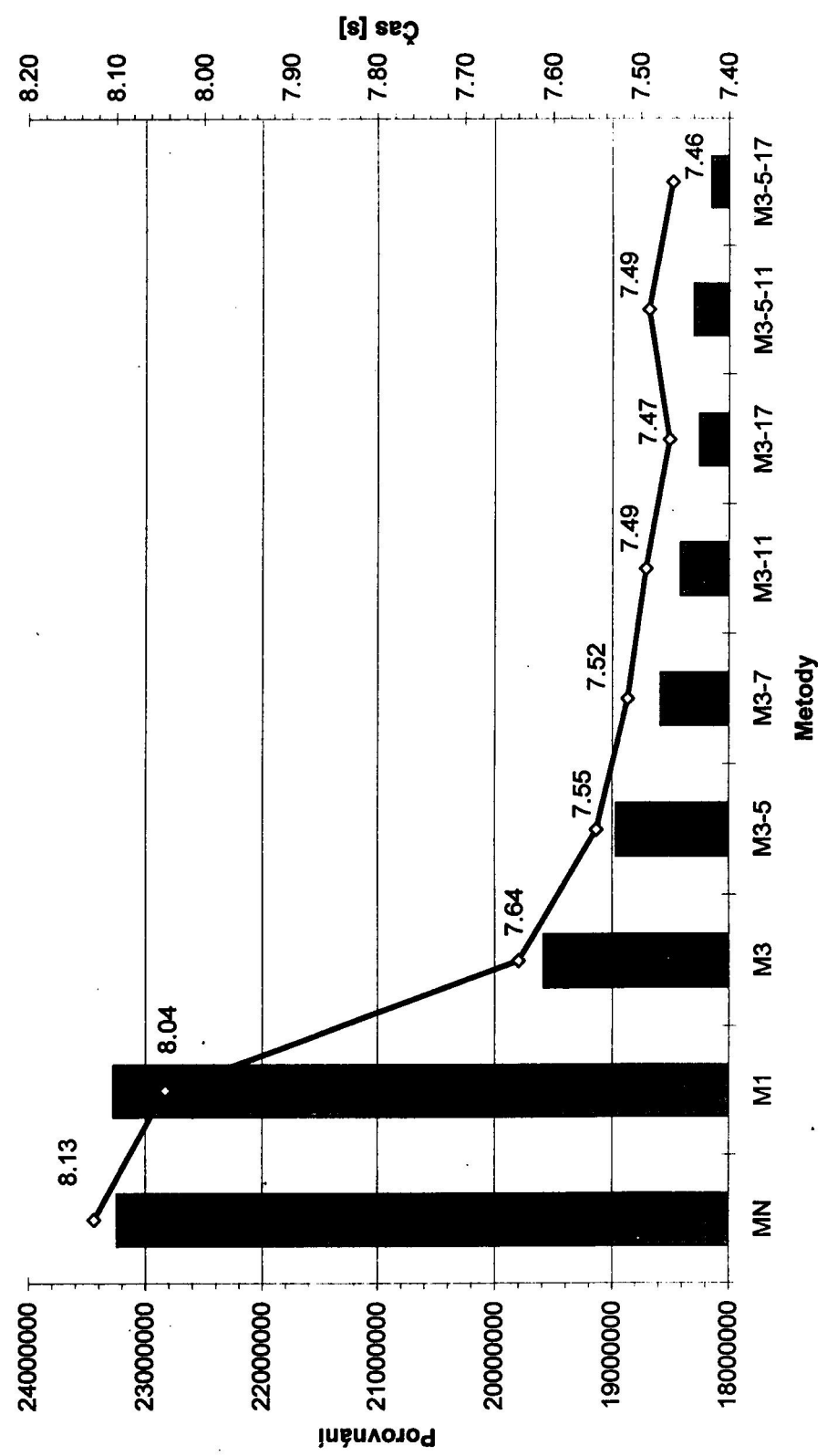
Mex je kromě prvku vždy vždy experimentální.

Kladné postupnosti se obtížněji inspekcionem

8a

Globalní měření manometry (příměry)

1 000 000 prvků



STAT. BASIC STATS	Descriptive Statistics (quicksort.sta)						
Variable	Valid N	Mean	Confid. -95.000%	Confid. +95.000%	Median	Minimum	Maximum
VAR1	25	8.125200	8.077016	8.173384	8.080000	7.950000	8.440000
VAR2	25	8.044000	7.996396	8.091604	8.010000	7.800000	8.280000
VAR3	25	7.639600	7.620779	7.658421	7.630000	7.550000	7.740000
VAR4	25	7.551200	7.524176	7.578224	7.530000	7.440000	7.680000
VAR5	25	7.515600	7.494617	7.536583	7.510000	7.420000	7.650000
VAR6	25	7.494400	7.476091	7.512709	7.500000	7.380000	7.600000
VAR7	25	7.467600	7.457611	7.477589	7.470000	7.410000	7.510000
VAR8	25	7.491200	7.479206	7.503194	7.490000	7.440000	7.530000
VAR9	25	7.463600	7.443382	7.483818	7.470000	7.330000	7.560000
VAR10	0	--	--	--	--	--	--

STAT. BASIC STATS	Descriptive Statistics (quicksort.sta)	
Variable	Range	Std.Dev.
VAR1	.490000	.116730
VAR2	.480000	.115326
VAR3	.190000	.045596
VAR4	.240000	.065468
VAR5	.230000	.050833
VAR6	.220000	.044355
VAR7	.100000	.024201
VAR8	.090000	.029057
VAR9	.230000	.048980
VAR10	--	--

STAT. BASIC STATS	VAR2 (quicksort.sta)					
Category	Count	Cumul. Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumul. % of All
7.7000 < x <= 7.8000	1	1	4.00000	4.0000	4.00000	4.0000
7.8000 < x <= 7.9000	0	1	0.00000	4.0000	0.00000	4.0000
7.9000 < x <= 8.0000	10	11	40.00000	44.0000	40.00000	44.0000
8.0000 < x <= 8.1000	7	18	28.00000	72.0000	28.00000	72.0000
8.1000 < x <= 8.2000	3	21	12.00000	84.0000	12.00000	84.0000
8.2000 < x <= 8.3000	4	25	16.00000	100.0000	16.00000	100.0000
Missing	0	25	0.00000		0.00000	100.0000

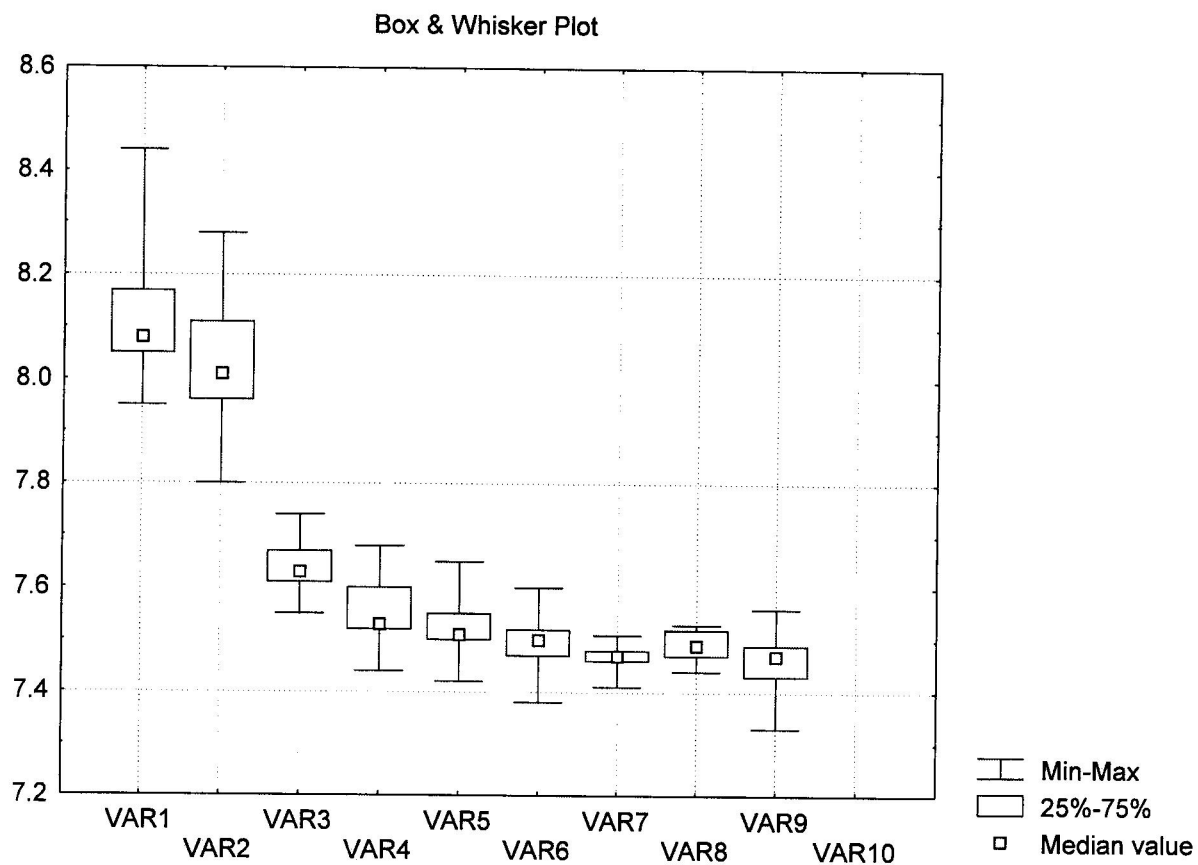
Popis: frekvenci tabulka po vari M1

(Histogram se mi nepostarila vytisknout)

STAT. BASIC STATS	Descriptive Statistics (quicksort.sta)						
Variable	Valid N	Mean	Median	Minimum	Maximum	Lower Quartile	Upper Quartile
VAR1	25	8.125200	8.080000	7.950000	8.440000	8.050000	8.170000
VAR2	25	8.044000	8.010000	7.800000	8.280000	7.960000	8.110000
VAR3	25	7.639600	7.630000	7.550000	7.740000	7.610000	7.670000
VAR4	25	7.551200	7.530000	7.440000	7.680000	7.520000	7.600000
VAR5	25	7.515600	7.510000	7.420000	7.650000	7.500000	7.550000
VAR6	25	7.494400	7.500000	7.380000	7.600000	7.470000	7.520000
VAR7	25	7.467600	7.470000	7.410000	7.510000	7.460000	7.480000
VAR8	25	7.491200	7.490000	7.440000	7.530000	7.470000	7.520000
VAR9	25	7.463600	7.470000	7.330000	7.560000	7.430000	7.490000
VAR10	0	--	--	--	--	--	--

STAT. BASIC STATS	Descriptive Statistics (quicksort.sta)	
Variable	Range	Quartile Range
VAR1	.490000	.120000
VAR2	.480000	.150000
VAR3	.190000	.060000
VAR4	.240000	.080000
VAR5	.230000	.050000
VAR6	.220000	.050000
VAR7	.100000	.020000
VAR8	.090000	.050000
VAR9	.230000	.060000
VAR10	--	--

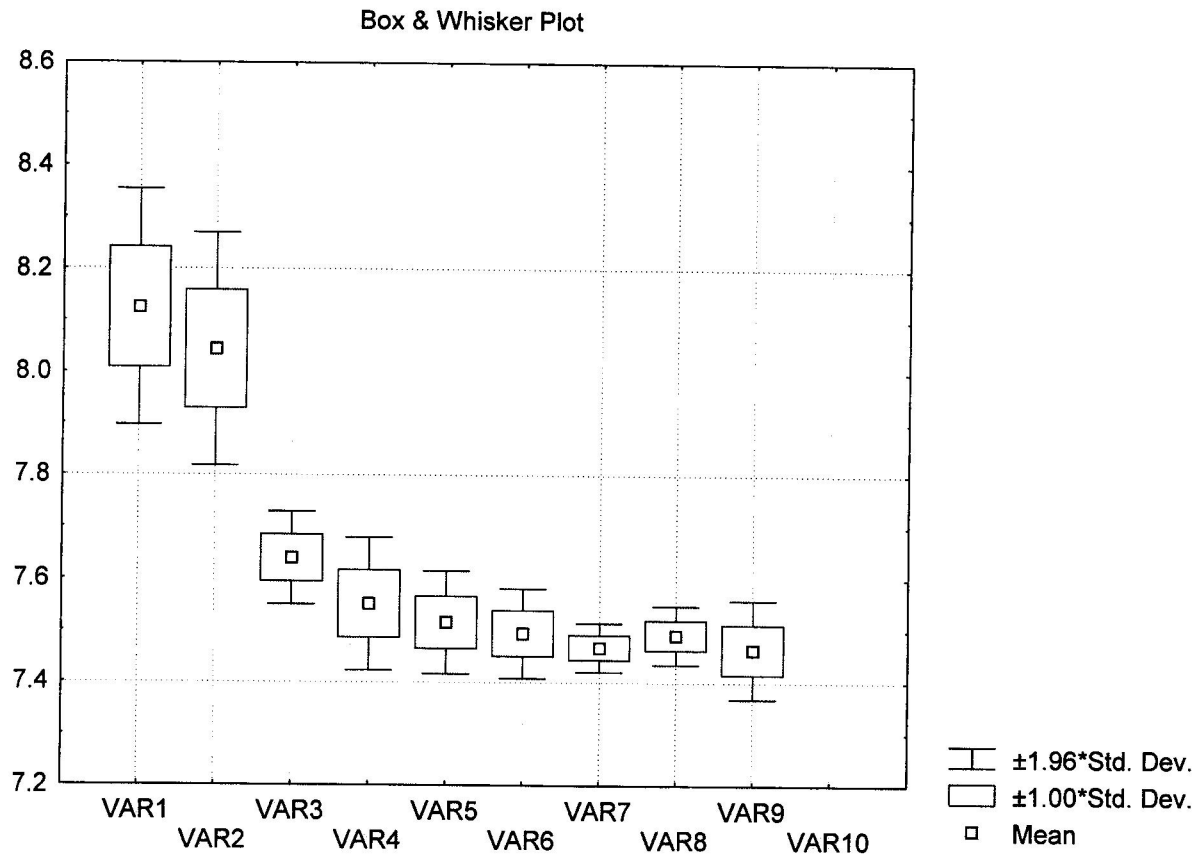
Popis: Nekteré popisné statistiky pro všechny nové.



Popis: Keatkové diagramy založené na predchádzajúcich štatistikách.
Umožňujú porovnať údaje medzi sebou.

STAT. BASIC STATS	Descriptive Statistics (quicksort.sta)				
Variable	Valid N	Mean	Confid. -95.000%	Confid. +95.000%	Std.Dev.
VAR1	25	8.125200	8.077016	8.173384	.116730
VAR2	25	8.044000	7.996396	8.091604	.115326
VAR3	25	7.639600	7.620779	7.658421	.045596
VAR4	25	7.551200	7.524176	7.578224	.065468
VAR5	25	7.515600	7.494617	7.536583	.050833
VAR6	25	7.494400	7.476091	7.512709	.044355
VAR7	25	7.467600	7.457611	7.477589	.024201
VAR8	25	7.491200	7.479206	7.503194	.029057
VAR9	25	7.463600	7.443382	7.483818	.048980
VAR10	0	--	--	--	--

Popis : Datě popisné statistiky pro každý řádek.



Popis: Katicové diagramy po veľkosti nerie.

Kejmanie je porovnanie s katicovými diagramy nábožními na mediane.