

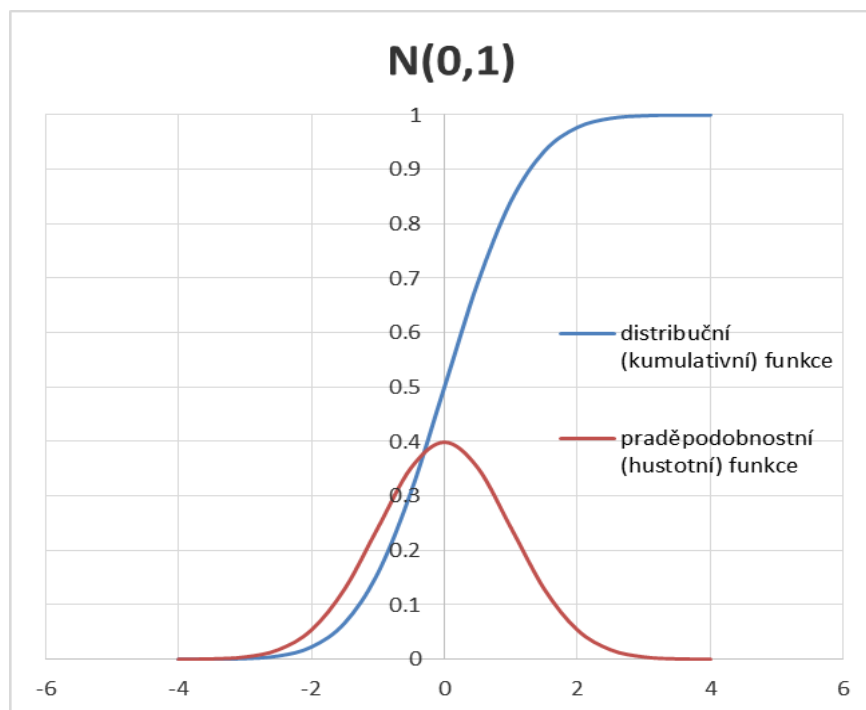
sestrojení grafu Gaussovy funkce

použijte funkci =normsdist
použijte funkci =normdist

=NORM.S.DIST(z;kur
=NORM.DIST(z;x0;s;C

x	$x_0=0$ $s=1$ P(x)	CDF(x)
-4.0		
-3.5		
-3.0		
-2.5		
-2.0		
-1.5		
-1.0		
-0.5		
0.0		
0.5		
1.0		
1.5		
2.0		
2.5		
3.0		
3.5		
4.0		

nul)
3/1)

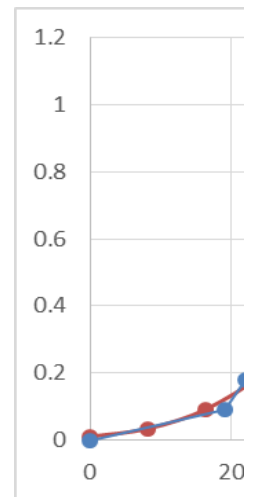


empirická distribuční funkce (EDF, kumulativní)

z každého souboru dat (náhodná data) se dá vytvořit kumulativní distribuční funkce



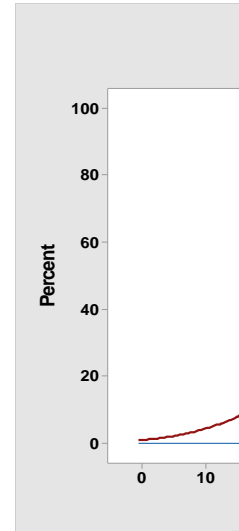
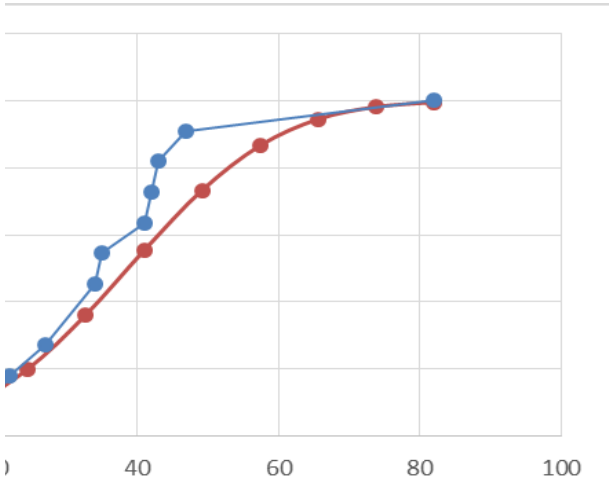
	bez dupl.	=countif() abs_četnost	rel_četnost	kumul. F(x)
data	0	0		0
22				
82				
27				
43				
19				
47				
41				
34				
34				
42				
35				
průměr				
sm.odch.				
počet =				



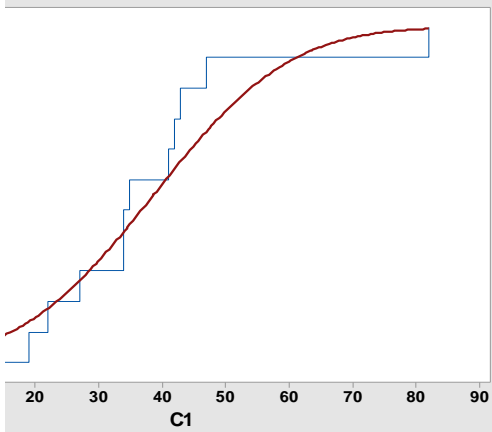


normální distribuce $N(\text{prum}, s^2)$
krok= 0

x	norm.dist
---	-----------



Empirical CDF of C1
Normal



Mean	38.73
StDev	16.85
N	11

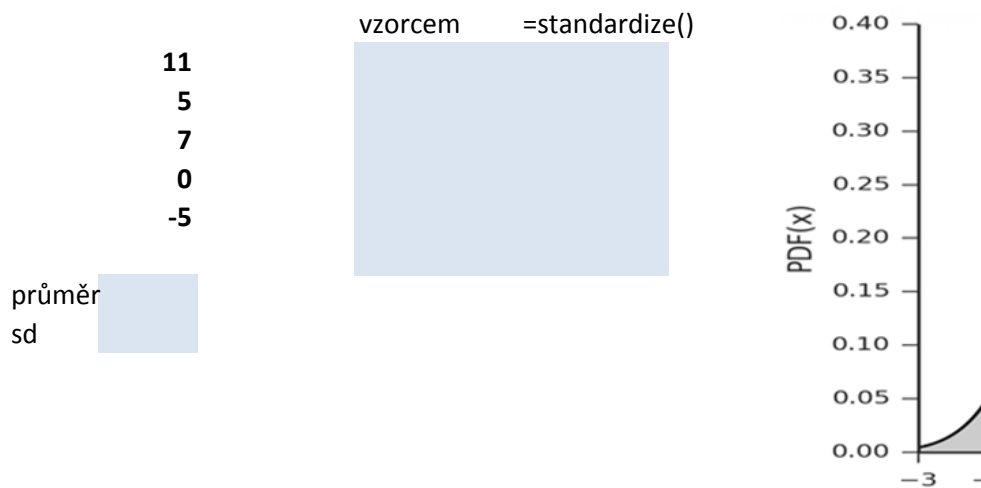
Normální rozdělení je nejdůležitější mezi rozděleními, musí se testovat, chceme-li použít parametrické
 častým problémem je malý počet dat (ale to je problém i u jiných testů) -
 vzhledem k náhodnému charakteru dat, není možné dokázat normalitu u malého počtu hodnot
 můžeme použít některý software, doporučený = Gnumeric (Lilliefors)

vysvětlení z-skóre

z-skóre

$$z = \frac{x - \mu}{\sigma}$$

the z-score is the number of population standard deviations that the data
 For sample data, the z Score is the number of sample standard deviations



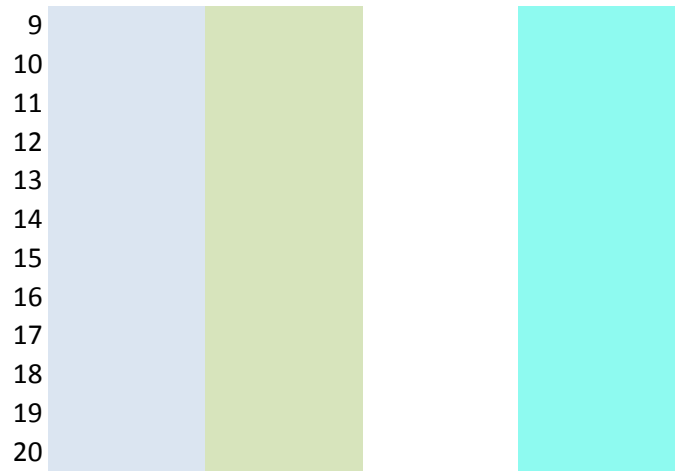
Normal Probability Plot
 Q-Q graf

Data
4.0
4.7
4.9
5.0
5.0
5.4
5.5
6.1

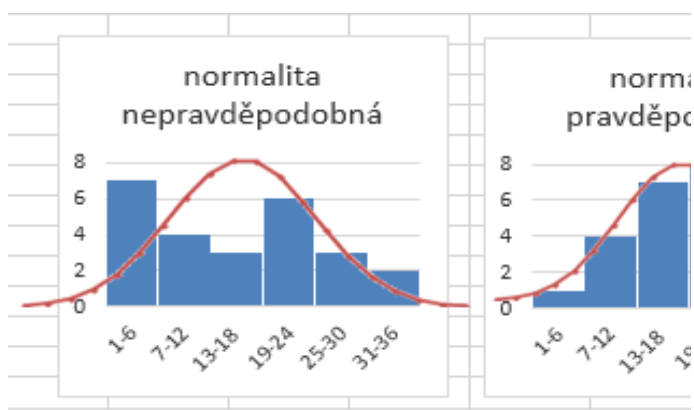
tabulka	
počet	
průměr	
SD	

i	Data	z-skóre	(i-0.5)/n	norm.s.invG
1				
2				
3				
4				
5				
6				
7				
8				

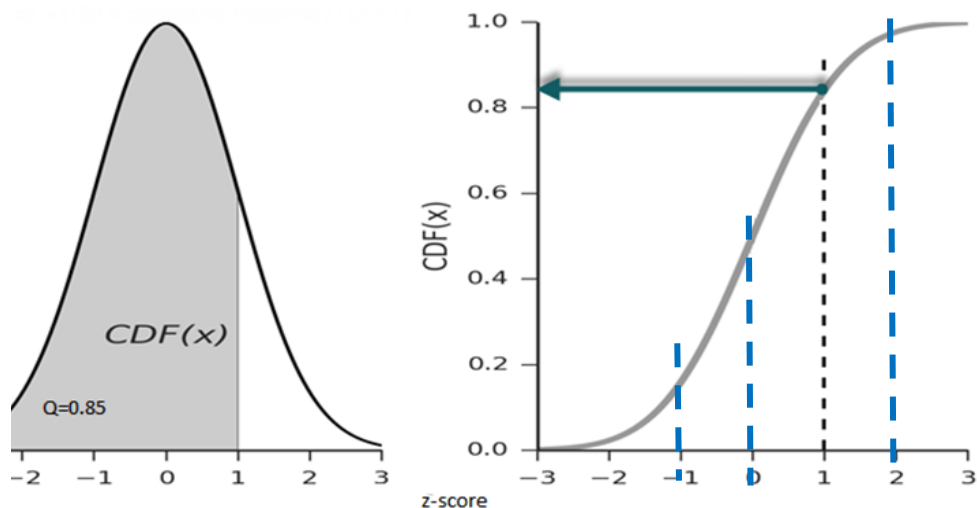
6.3
6.8
7.1
8.0
8.8
9.7
9.8
10.1
10.6
10.7
10.8
11.7



ě testy, které mají větší sílu.



value is from the population mean.
that a data value is from the sample mean.



výstup z Gnumeric

Lillieforsův (Kolmogorov)

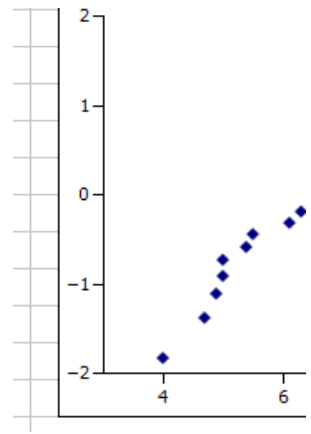
Alfa

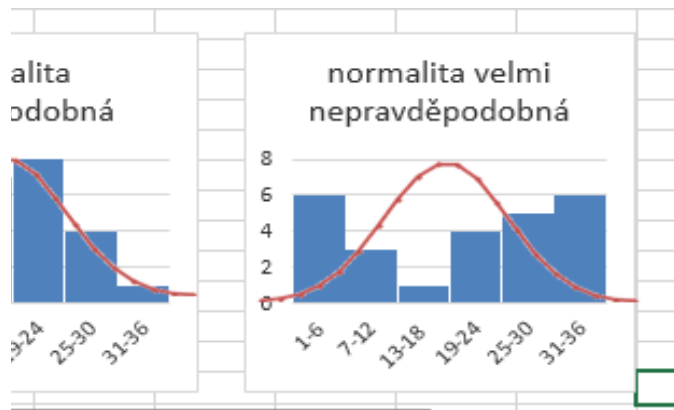
p-hodnota

Statistika

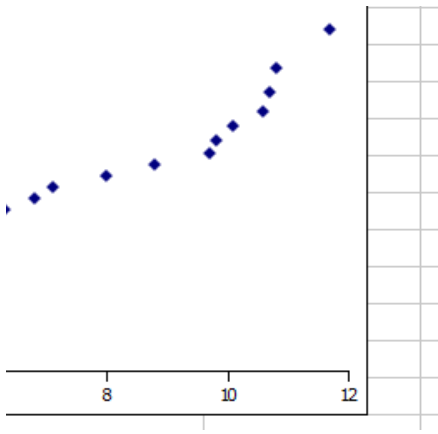
N

Závěr



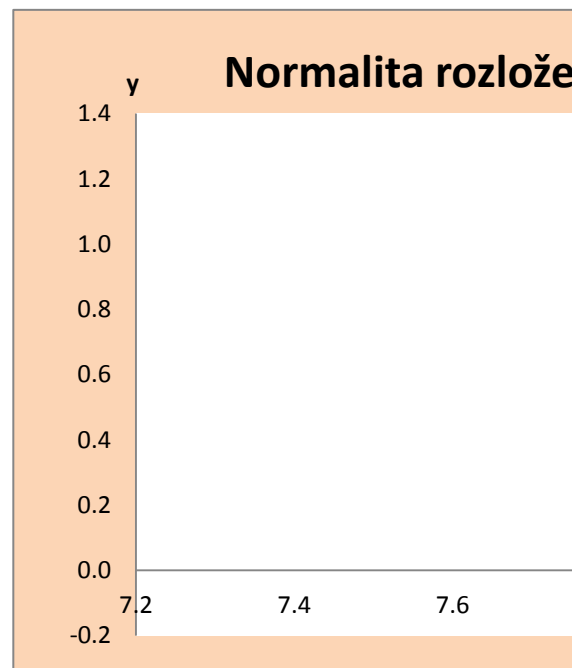
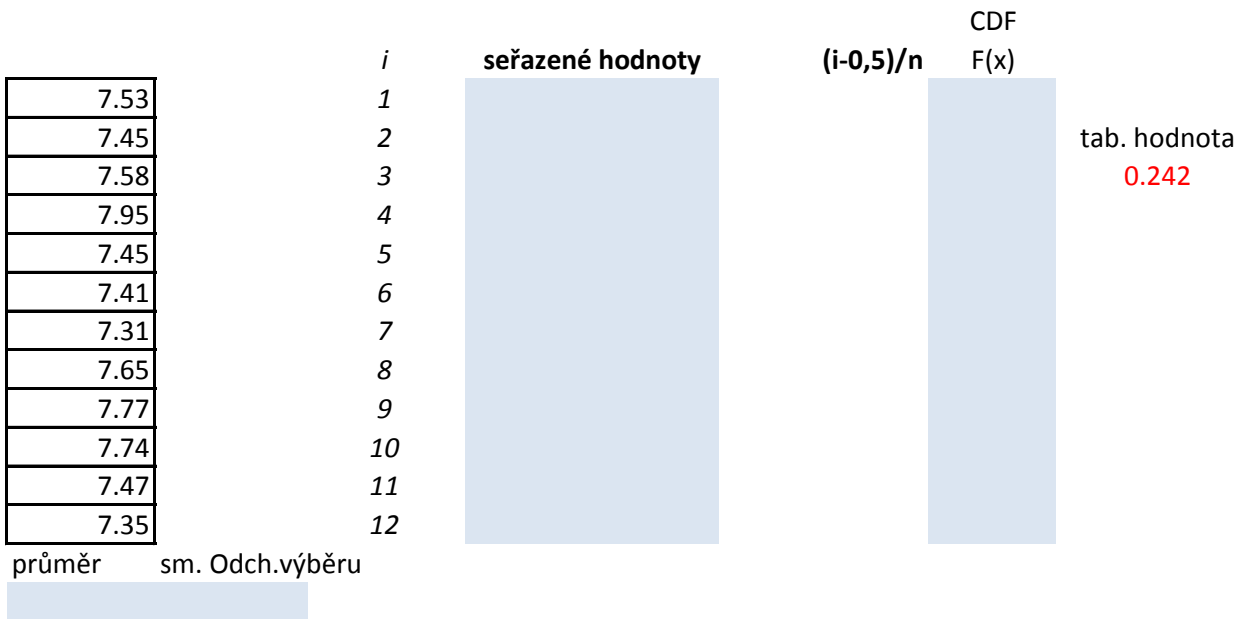


rovň-Smírnovi	Data
	0,05
	0,23724996112
	0,15488614216
	20
	Možná normální



Lillieforsův (KS) test normality srovnává, zda rozdíl N.rozdělení od 2 EDF z mírně posunutých P nepřekročí K.S. předpokládá znalost skutečné stř. hodnoty a sm.odchylky, L. vychází z jejich odhadu = jiná tabulka kr

I



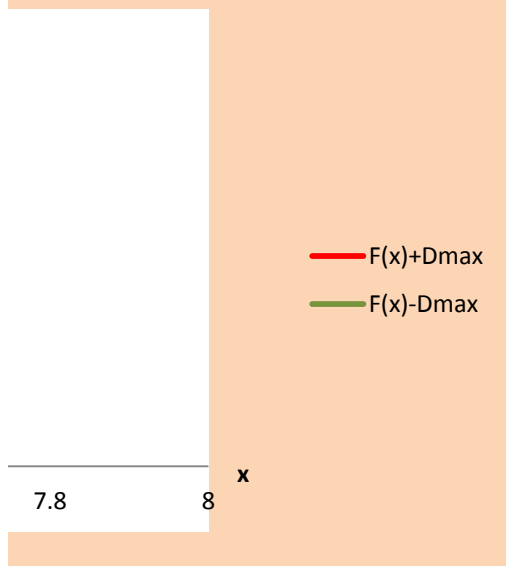
í kritickou hodnotu.
itických hodnot

$$\max_x |F(x) - S_n(x)| \leq D_{n,\alpha}$$

$$F(x)+0,242 \quad F(x)-0,242$$

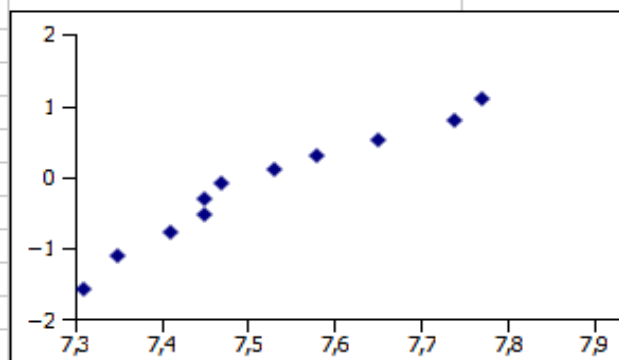
$n \setminus \alpha$	0.01	0.05	0.10	0.15
4	0.417	0.381	0.352	0.319
5	0.405	0.337	0.315	0.299
6	0.364	0.319	0.294	0.277
7	0.348	0.300	0.276	0.258
8	0.331	0.285	0.261	0.244
9	0.311	0.271	0.249	0.233
10	0.294	0.258	0.239	0.224
11	0.284	0.249	0.230	0.217
12	0.275	0.242	0.223	0.212
13	0.268	0.234	0.214	0.202
14	0.261	0.227	0.207	0.194
15	0.257	0.220	0.201	0.187
16	0.250	0.213	0.195	0.182
17	0.245	0.206	0.189	0.177
18	0.239	0.200	0.184	0.173
19	0.235	0.195	0.179	0.169
20	0.231	0.190	0.174	0.166
25	0.203	0.180	0.165	0.153
30	0.187	0.161	0.144	0.136
OVER 30	1.031	0.886	0.805	0.768
	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}

ení dat dle Lillieforse



takto vypadá výstup z Gnumer

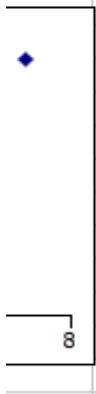
Lilliefors (Kolmogorov-Smirnov) Test		Sloupec 1
Alpha		
p-Value		0,4185071
Statistic		0,172290246
N		
Conclusion		Possibly norr



	0.20
)	0.300
)	0.285
)	0.265
)	0.247
)	0.233
)	0.223
)	0.215
)	0.206
)	0.199
)	0.190
)	0.183
)	0.177
)	0.173
)	0.169
)	0.166
)	0.163
)	0.160
)	0.149
)	0.131
)	0.736
	\sqrt{n}

ic

0,05
 44726
 515192
 12
 mal



Náhodnost prvků výběrového souboru je pilířem statistiky.

bez ní nelze vztahovat závěry z analýzy výběru na populaci

Jak ověřit, že data jsou náhodná?

nebo že zvolený klíč rozdělení pacientů na skupiny lék/placebo je náhodný?

<https://publi.cz/books/201/01.html#3>

RUNS test (Wald-Wolfowitz runs test)

"run" je motiv/série/běh stejných **binárních** (dichotomických) kódů, srovnáváme jejich **POČET** (shodu) s normálním rozdělením

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1$$

přibl. pro $n_1, n_2 > 20$ $H_0: \pi=0,5$

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

pokud n_1 nebo $n_2 < 20$ pak se použije TABULKA <http://www.real-statistics.com/statistics-tables>

n1	n2	r - lower	r - upper	n1	n2	r - lower	r - upper	n1	n2	r - lower	r - upper
20	20	15	27	19	19	14	26	18	18	13	25
20	19	14	26	19	18	14	25	18	17	13	24
20	18	14	26	19	17	13	25	18	16	12	24
20	17	14	25	19	16	13	24	18	15	12	23
20	16	13	24	19	15	12	23	18	14	11	22
20	15	13	24	19	14	12	22	18	13	11	21
20	14	12	23	19	13	11	21	18	12	10	21
20	13	11	22	19	12	11	21	18	11	10	19
20	12	11	21	19	11	10	20	18	10	9	18
20	11	10	20	19	10	9	19	18	9	9	19
20	10	10	19	19	9	9	19	18	8	8	16
20	9	9	19	19	8	8	16	18	7	7	
20	8	8	16	19	7	7		18	6	6	
20	7	7		19	6	7		18	5	6	
20	6	7		19	5	6		18	4	5	
20	5	6		19	4	5		18	3	4	
20	4	5		19	3	4		18	2	3	
20	3	4		19	2	3					
20	2	3									

Příklady: proměnná může být sama **dichotomická**, např. pohlaví

trik, jak v Excelu spočítat výskyt z
=DÉLKA(řetězec)-DÉLKA(DOSADIT

Je následující pořadí výběru muž/žena 26 osob z populace náhodné?
 MMMMŽŽMMŽMMMMMMMMŽŽŽMMŽMMMMŽŽ 26

n1=
 n2=
 n=
 runs=
 min.poč=
 max.počet=

n1	n2	r - lower	r - upper	n1
17	17	12	24	16
17	16	12	23	16
17	15	12	22	16
17	14	11	22	16
17	13	11	21	16
17	12	10	20	16
17	11	10	19	16
17	10	9	18	16
17	9	8	19	16
17	8	8	16	16
17	7	7		16
17	6	6		16
17	5	5		16
17	4	5		16
17	3	4		16
17	2	3		

jak by to vyšlo nenáhodně?

*pokud proměnná není binární, musíme ji takovouto vyrobit:
 např. poloha od mediánu nebo sudé/liché*

Zjistěte, zda data výběru byla vygenerována náhodně

data
56
67
38
-8
66
79
34
2
0
12
78
23
90
4
78

median=
 n1=
 n2=
 runs=
 podle tabulky (7,7)...<4,12>
 H0....

n1	n2	r - lower	r - upper	n1	n2	r - lower
14	14	10	20	12	12	8
14	13	10	19	12	11	8
14	12	9	19	12	10	8
14	11	9	18	12	9	7
14	10	8	17	12	8	7
14	9	8	16	12	7	6
14	8	7	15	12	6	5
14	7	6	14	12	5	5
14	6	6		12	4	4
14	5	5		12	3	3
14	4	4		12	2	3
14	3	3				
14	2	3				
13	13	9	19	11	11	8
13	12	8	19	11	10	7
13	11	8	18	11	9	7
13	10	8	17	11	8	6
13	9	7	16	11	7	6
13	8	7	15	11	6	5
13	7	6	14	11	5	5
13	6	6		11	4	4
13	5	5		11	3	3
13	4	4		11	2	
13	3	3				
13	2	3				

jak by to vyšlo nenáhodně:

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15

median=
 n1=
 n2=
 runs=
 podle tabulky (7,7)...<4,12>
 H0...



Zjistěte, zda data výběru byla vygenerována náhodně

data	6	2	7	0	0	1	7	3	0	5	0	8	4	6
	S	S	L	S	S	L	L	L	S	L	S	S	S	S
				SS	L	SS	LLL	S	L	SSSSS	L	SS	LL	

<http://www.real-statistics.com/non-parametric-tests/one-sample-runs-test/>

[/runs-test-table/](#)

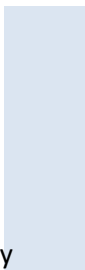
řetězcí
T(řetězec;"znak";"")

n2	r - lower	r - upper	n1	n2	r - lower	r - upper
16	12	22	15	15	11	21
15	11	22	15	14	10	21
14	11	21	15	13	10	20
13	10	21	15	12	9	19
12	10	21	15	11	9	18
11	9	19	15	10	8	17
10	9	18	15	9	8	19
9	8	19	15	8	7	15
8	7	16	15	7	7	14
7	7		15	6	6	
6	6		15	5	5	
5	5		15	4	4	
4	5		15	3	4	
3	4		15	2	3	
2	3					

r - upper	n1	n2	r - lower	r - upper	n1	n2	r - lower	r - upper
18	10	10	7	15	8	8	5	13
17	10	9	6	15	8	7	5	12
16	10	8	6	14	8	6	4	11
15	10	7	6	13	8	5	4	10
15	10	6	5	12	8	4	4	
13	10	5	4		8	3	3	
12	10	4	4		8	2		
	10	3	3					
	10	2			7	7	4	12
					7	6	4	11
					7	5	4	10
					7	4	3	
					7	3	3	
					7	2		
16	9	9	6	14				
16	9	8	6	13	6	6	4	10
15	9	7	5	13	6	5	4	9
14	9	6	5	12	6	4	3	8
13	9	5	4		6	3	3	
12	9	4	4		6	2		
	9	3	3					
	9	2			5	5	3	9
					5	4	3	8
					5	3		
					5	2		



8	7	0	6	5	5
S	L	S	S	L	L



=n
=n1
=n2

(12,8) podle tabulky

H0....

=runs