

PSY117 2016

Statistická analýza dat v psychologii

Přednáška 3

Transformace skóru a kvantily normálního rozložení

Transformace skóru (dat)

Pro usnadnění porozumění a možnost dalších analýz často přepočítáváme hodnoty proměnných, aby měla lepší vlastnosti

- Usnadnění interpretace – *lineární transformace*
 - např. vynásobení 10 nebo 100 pro odstranění desetinných míst
 - tvar rozložení zůstává zachován
 - možnost sjednocení různých proměnných na stejnou škálu, měřítko ... Standardizace

 - Změna tvaru rozložení – *nelineární transformace*
 - log/exp fce, (od)mocniny, Tukey: „ladder of powers“ Hendl kap. o EDA.
 - Též „normalizace“ rozložení – normalizované skóry

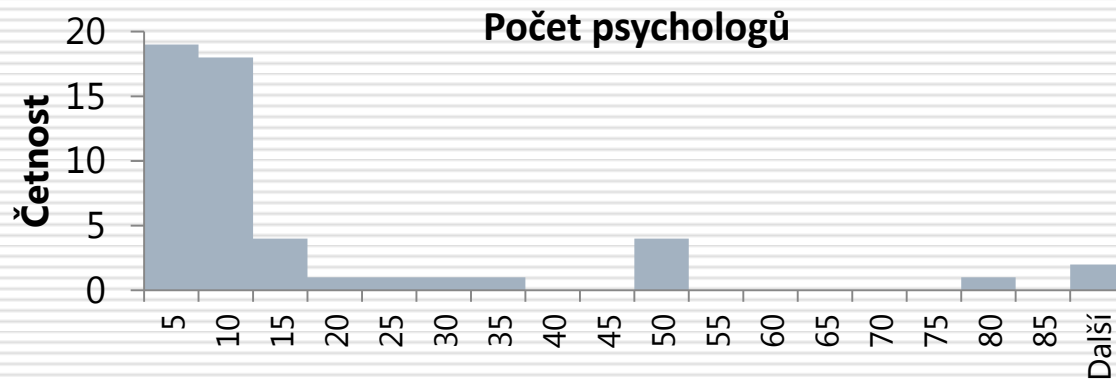
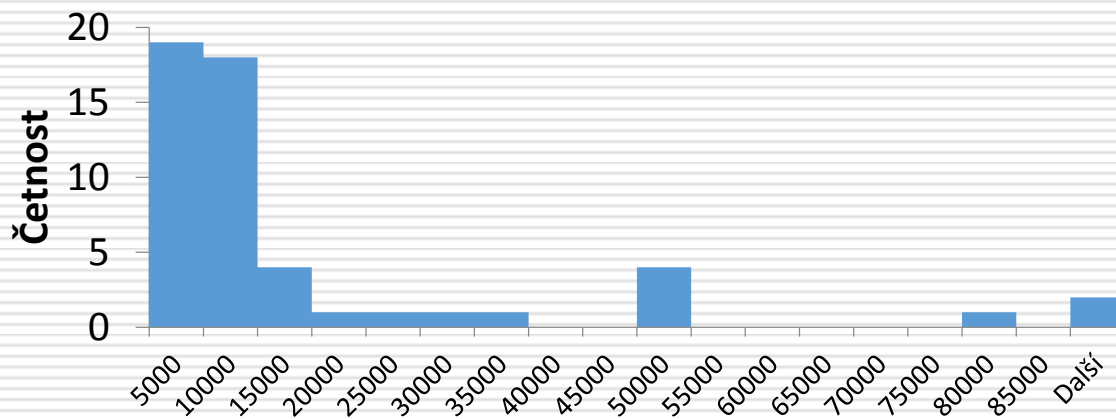
 - Změna úrovně měření – *pořadová transformace (ranking)*
-

Lineární transformace 1

□ Např. počtu psychologů na tisíce

- Tvar rozložení zachován
- Popisné statistiky se předpověditelně změní
- M , SD , Md , IQR , min , max jsou tisíckrát menší
- s^2 (VAR)?

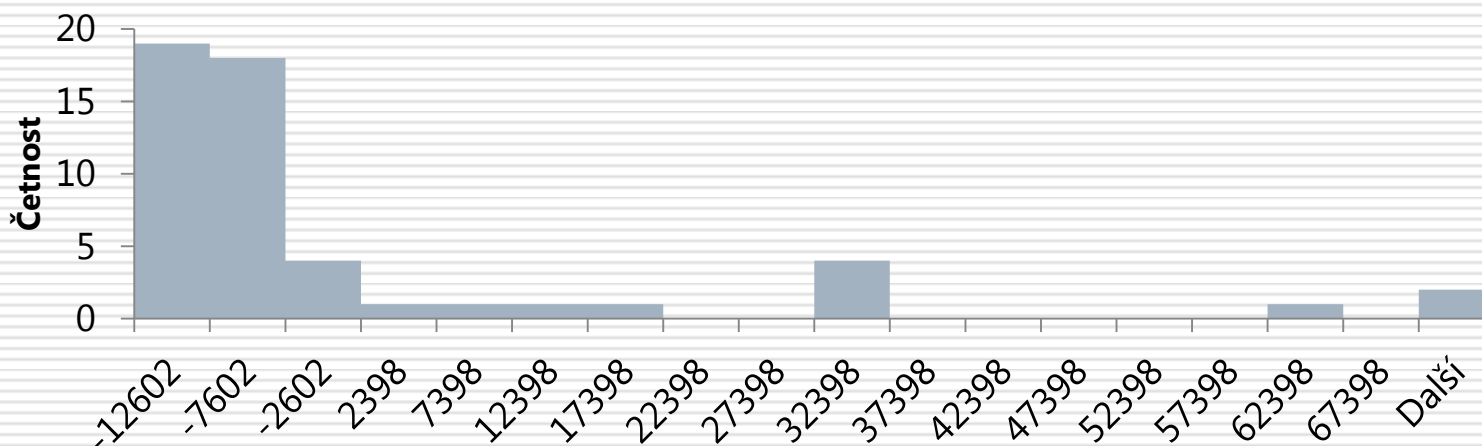
poc_psy	v tisících
5000	5
1200	1,2
1000	1
10000	10
12000	12
4000	4
1500	1,5
10000	10
100000	100
10000	10
12000	12
10000	10
150000	150
35000	35
80000	80
50000	50
17000	17
1385	1,385
2000	2
10000	10
5000	5
10000	10
9999	9,999
50000	50
10000	10
10000	10
12500	12,5
3000	3
15000	15
3 000	3
6000	6
2000	2
5000	5
10000	10
10000	10
10000	10
5000	5
5000	5
5743	5,743
8000	8
3500	3,5
1500	1,5
25000	25
10000	10
10000	10
50000	50
30000	30
50000	50
5000	5
7000	7
5000	5
1000	1



Počet psychologů v tisících

Lineární transformace 2

- Deviační skóry x_i - odečtení průměru
 - Tvar rozložení zůstává zachován
 - Popisné statistiky – CT jsou o průměr menší, variabilita beze změn
 - Snadnější interpretace jednotlivých skóru



Rozdíl mezi odhadem a průměrem odhadu počtu psychologů

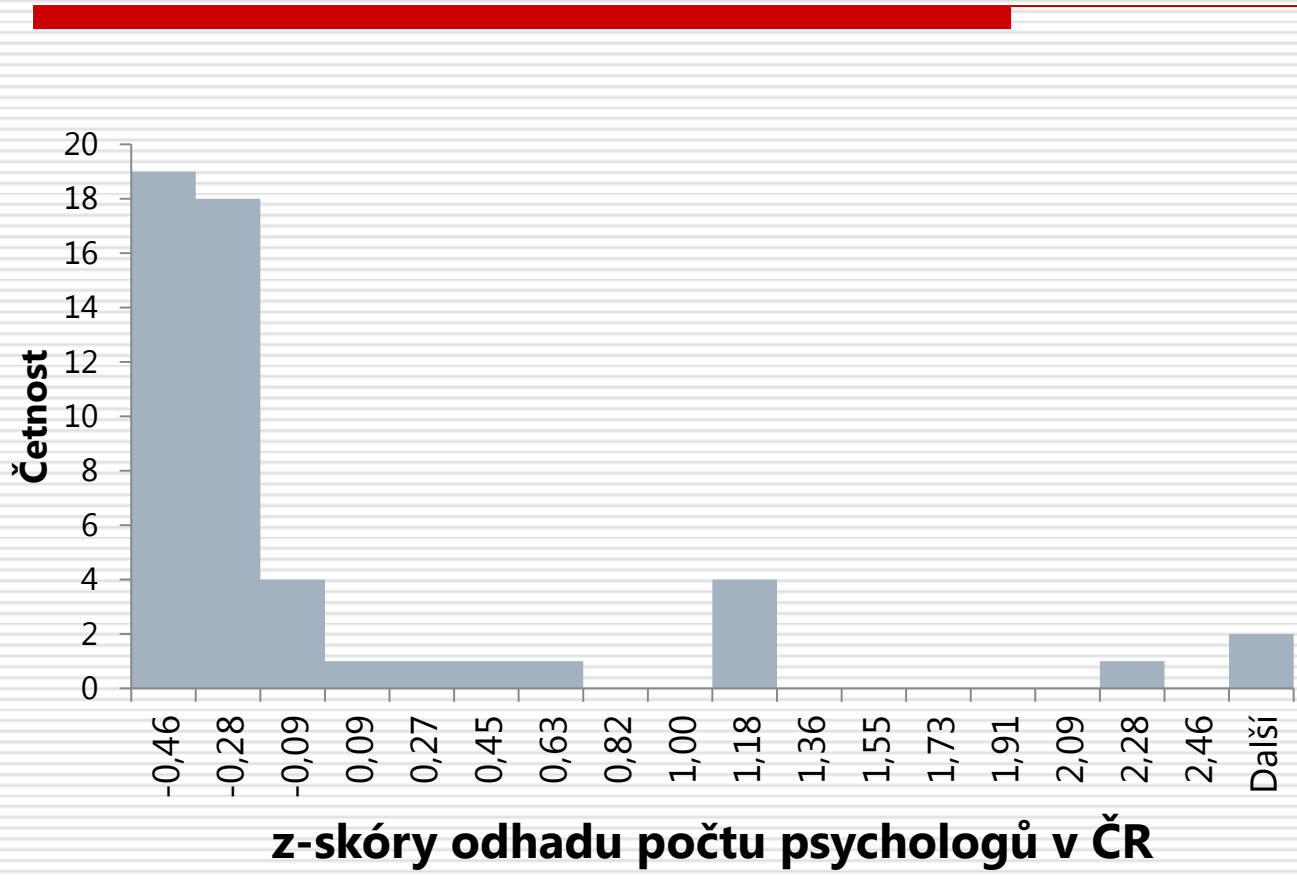
	poc. psy	v tisících	dev
	5000	5	-12602
	1200	1,2	-16402
	1000	1	-16602
	10000	10	-7602
	12000	12	-5602
	4000	4	-13602
	1500	1,5	-16102
	10000	10	-7602
	100000	100	82398
	10000	10	-7602
	12000	12	-5602
	10000	10	-7602
	150000	150	132398
	35000	35	17398
	80000	80	62398
	50000	50	32398
	17000	17	-602
	1385	1,385	-16217
	2000	2	-15602
	10000	10	-7602
	5000	5	-12602
	10000	10	-7602
	9999	9,999	-7603
	50000	50	32398
	10000	10	-7602
	10000	10	-7602
	15000	15	-2602
	3000	3	-14602
	6000	6	-11602
	2000	2	-15602
	5000	5	-12602
	15000	15	-5102
	3000	3	-14602
	15000	15	-2602
	3000	3	-14602
	6000	6	-11602
	2000	2	-15602
	5000	5	-12602
	10000	10	-7602
	10000	10	-7602
	5000	5	-12602
	5000	5	-12602
	5743	5,743	-11859
	8000	8	-9602
	3500	3,5	-14102
	1500	1,5	-16102
	25000	25	7398
	10000	10	-7602
	10000	10	-7602
	50000	50	32398
	30000	30	12398
	50000	50	32398
	5000	5	-12602
	7000	7	-10602
	5000	5	-12602
	1000	1	-16602
M	17602	17,60	0,00
SD	27410	27,41	27410
VAR	751320188	751,3202	751320188
Me	10000	10,000	-7602
QQR	7125	7,125	7125
min	1000	1	-16602,4423
max	150000	150	132397,558

Lineární transformace - standardizace z-skóry, standardizované skóry

- Nejobvyklejší lineární transformace - **standardizace**
 - transformace sady skóru, aby $m = 0$, $s = 1$
 - **jednotkou měření se stává s** , možnost srovnávání různých škál (ale pozor rozdíly v rozložení zůstávají!)

$$z_i = (X_i - m) / s$$

- s . je zajímavá zvláště u normálně rozložených dat, protože známe řadu jeho percentilů zpaměti
 - u přibližně normálně rozložených dat o lidech je většina (přes 90%) lidí mezi -3 a 3



	poc_psy	v tisících	dev	z
	5000	5	-12602	-0,46
	1200	1,2	-16402	-0,60
	1000	1	-16602	-0,61
	10000	10	-7602	-0,28
	12000	12	-5602	-0,20
	4000	4	-13602	-0,50
	1500	1,5	-16102	-0,59
	10000	10	-7602	-0,28
	100000	100	82398	3,01
	10000	10	-7602	-0,28
	12000	12	-5602	-0,20
	10000	10	-7602	-0,28
	150000	150	132398	4,83
	35000	35	17398	0,63
	80000	80	62398	2,28
	50000	50	32398	1,18
	17000	17	-602	-0,02
	1385	1,385	-16217	-0,59
	2000	2	-15602	-0,57
	10000	10	-7602	-0,28
	5000	5	-12602	-0,46
	10000	10	-7602	-0,28
	9999	9,999	-7603	-0,28
	50000	50	32398	1,18
	10000	10	-7602	-0,28
	10000	10	-7602	-0,28
	12500	12,5	-5102	-0,19
	3000	3	-14602	-0,53
	15000	15	-2602	-0,09
	3 000	3	-14602	-0,53
	6000	6	-11602	-0,42
	2000	2	-15602	-0,57
	5000	5	-12602	-0,46
	10000	10	-7602	-0,28
	10000	10	-7602	-0,28
	10000	10	-7602	-0,28
	5000	5	-12602	-0,46
	5000	5	-12602	-0,46
	5743	5,743	-11859	-0,43
	8000	8	-9602	-0,35
	3500	3,5	-14102	-0,51
	1500	1,5	-16102	-0,59
	25000	25	7398	0,27
	10000	10	-7602	-0,28
	10000	10	-7602	-0,28
	10000	10	-7602	-0,28
	50000	50	32398	1,18
	30000	30	12398	0,45
	50000	50	32398	1,18
	5000	5	-12602	-0,46
	7000	7	-10602	-0,39
	5000	5	-12602	-0,46
	1000	1	-16602	-0,61
M	17602	17,60	0,00	0,00
SD	27410	27,41	27410	1
VAR	751320188	751,3202	751320188	1
Md	10000	10,000	-7602	0
IQR	7125	7,125	7125	0,25994
min	1000	1	-16602,4423	-0,6057
max	150000	150	132397,558	4,830226

Skóry odvozené ze z-skórů

Používané primárně v psychodiagnostických metodách

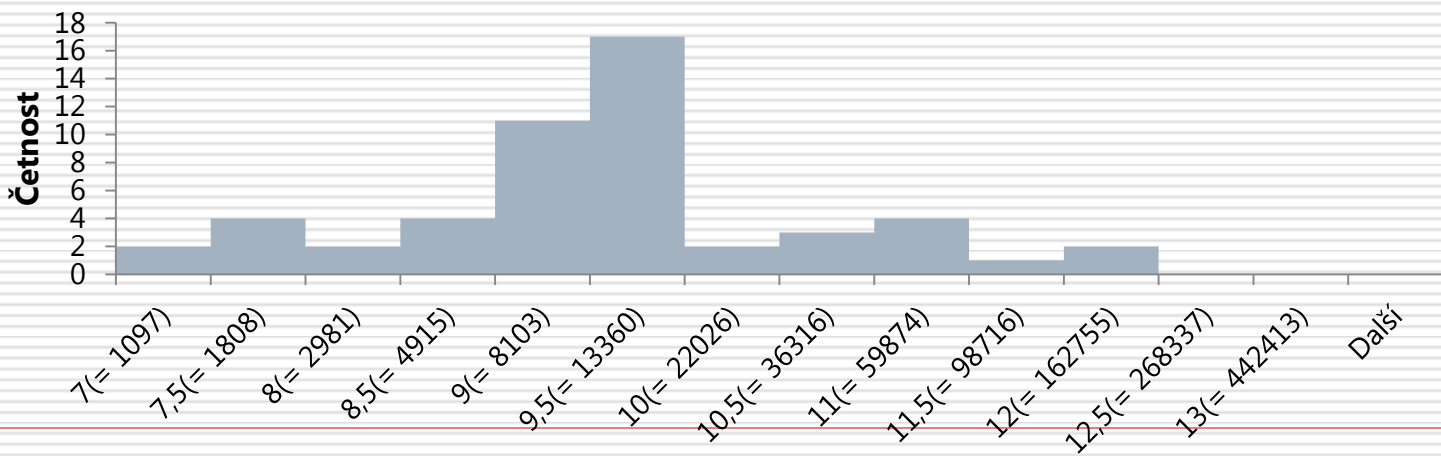
- ***IQ skóry*** ($m=100, s=15$)
- ***T skóry*** ($m=50, s= 10$)

- **staniny**, staninové skóry (standard *nine*) ($m=5, s= 2$) + kategorizace zaokrouhlením na celá čísla
- **steny**, stenové skóry (standard *ten*) ($m=5,5, s= 2$) + kategorizace zaokrouhlením na celá čísla

Pořád je podmínkou správné interpretace normální rozložení měřené proměnné v dané populaci!

Nelineární transformace 1

- Změna rozložení (obv. směrem k normálnímu)
 - Pro smysluplnější využití momentových statistik
 - Pro možnost využití analytických technik, které nějakou podobu rozložení vyžadují
- Popisné statistiky se mění složitěji
- Př. logaritmus počtu psychologů

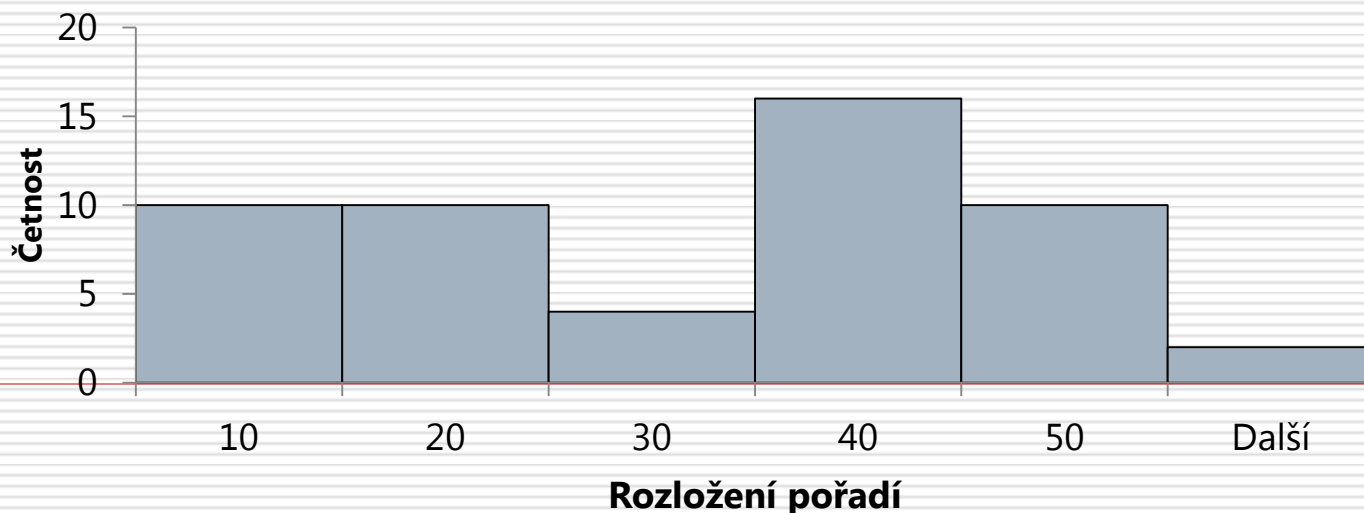


Přirozený logaritmus odhadu počtu psychologů

Nelineární transformace 2

□ Transformace na pořadí – ranking

- Eliminace odlehlých hodnot, odhlédnutí od velikosti rozdílů mezi lidmi
- Obvykle vzestupně (nejnižší hodnota má pořadí 1)
- Stejně hodnoty dostávají průměrné pořadí (=RANK.AVG)
- Výsledné rozložení je (přibližně) uniformní



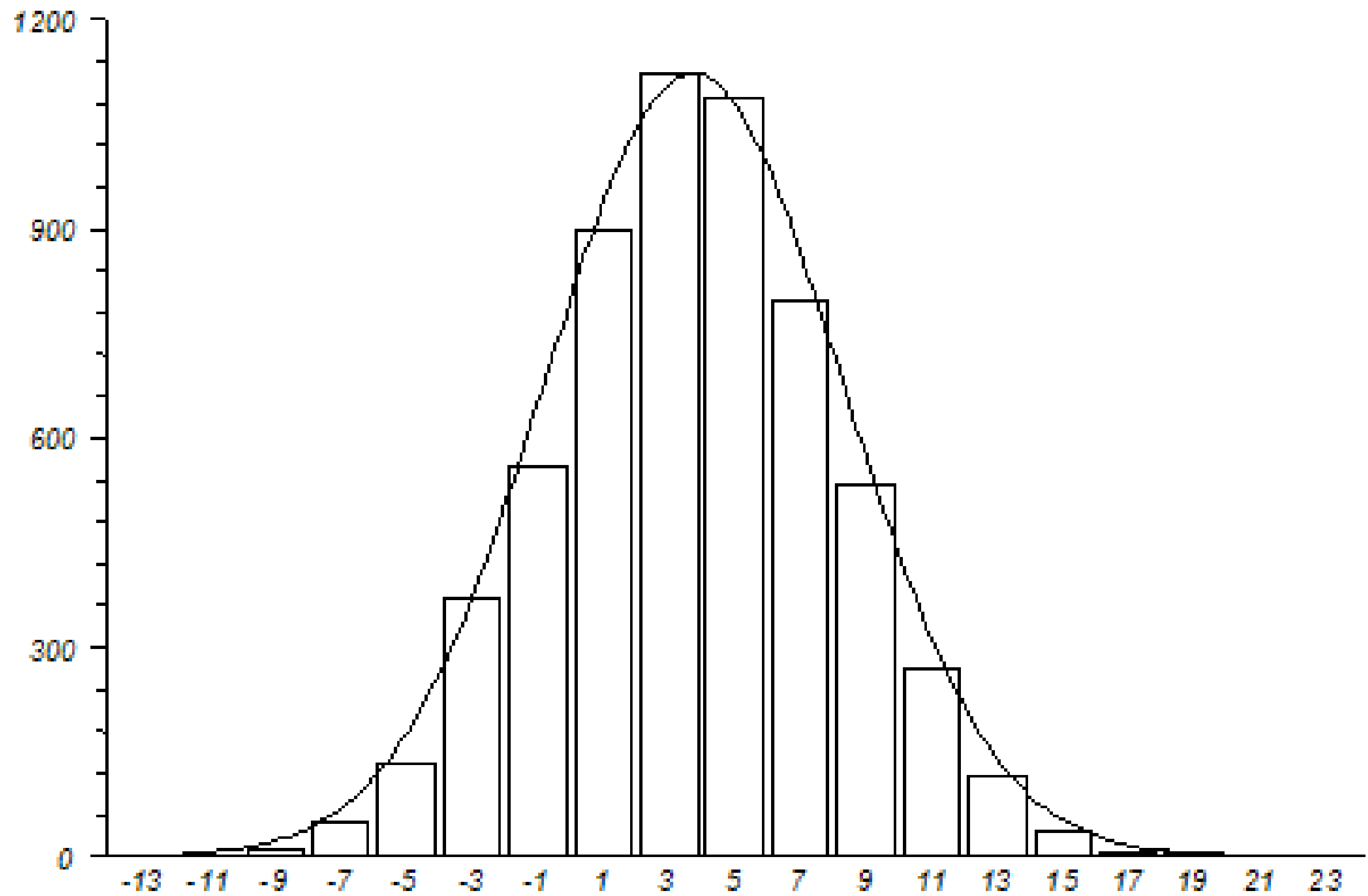
Transformace na percentily

- Zvláštní (standardizovaná) podoba transformace na pořadí
 - Používá se při tvorbě norem psychodiagnostických metod
-

Psychodiagnostická kalkulačka

- ❑ Převody různých skóru online.
 - ❑ Nástroj vyvíjí Hynek Cígler a Martin Šmíra
 - ❑ <http://kalkulacka.testforum.cz/transformace-skoru>
-

Histogram for Normal Distribution (mean = 3.8, sd = 4.3)



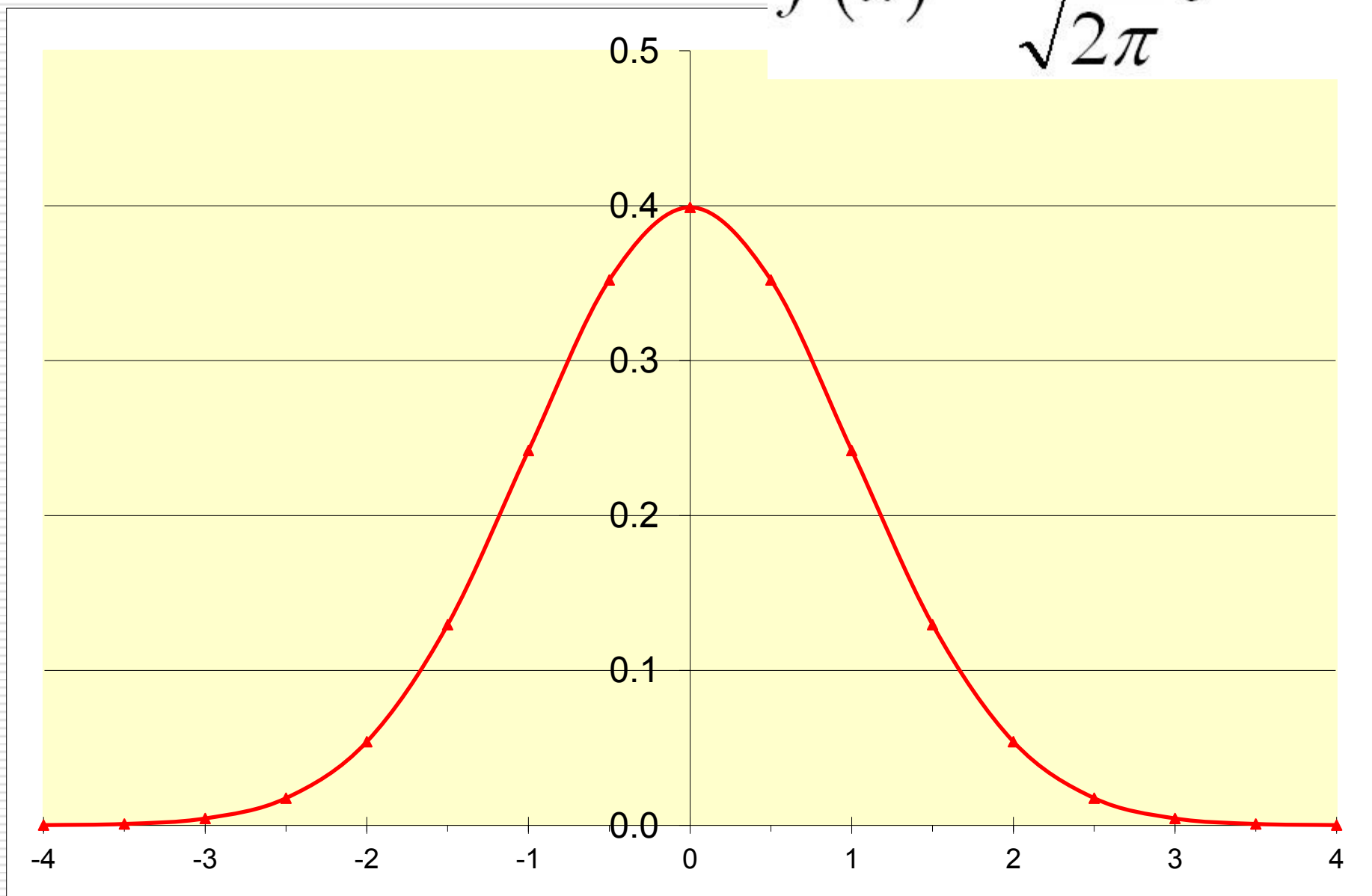
Mid Points for Normal Distribution (mean = 3.8, sd = 4.3)

Normální rozložení

Gaussovo, bell-curve

- Rozložení...
 - ...náhodných chyb
 - ...jevů v přírodě ovlivněných mnoha nezávislými faktory
 - Dlouhá historie – od 17. stol.
 - DeMoivre – sázení
 - Laplace a Gauss – chyby v astronomických pozorováních
 - Quetelet – lidi, *l'homme moyen*
-

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu-x)^2}{2\sigma^2}}$$



K čemu/proč normální rozložení?

- Mnoho proměnných je takto rozloženo
 - Můžeme pak hádat, kolik jakých hodnot v populaci je
 - Mnoho statistických postupů s normálním rozložením pracuje
 - v různých modifikacích a rolích
-

Vlastnosti normálního rozložení

https://en.wikipedia.org/wiki/Normal_distribution

- Symetrické, unimodální
- Průměr=medián=modus
- V hodnotách $M \pm SD$ se mění prohnutí
- Je-li $SD = 1$, pak plocha pod křivkou je 1
- Zešikmení (skewness) je 0
- Strmost (kurtosis) je 3
 - často se prezentuje hodnota $K-3$
- *Forma, od níž odrážíme popis pozorovaných rozložení*

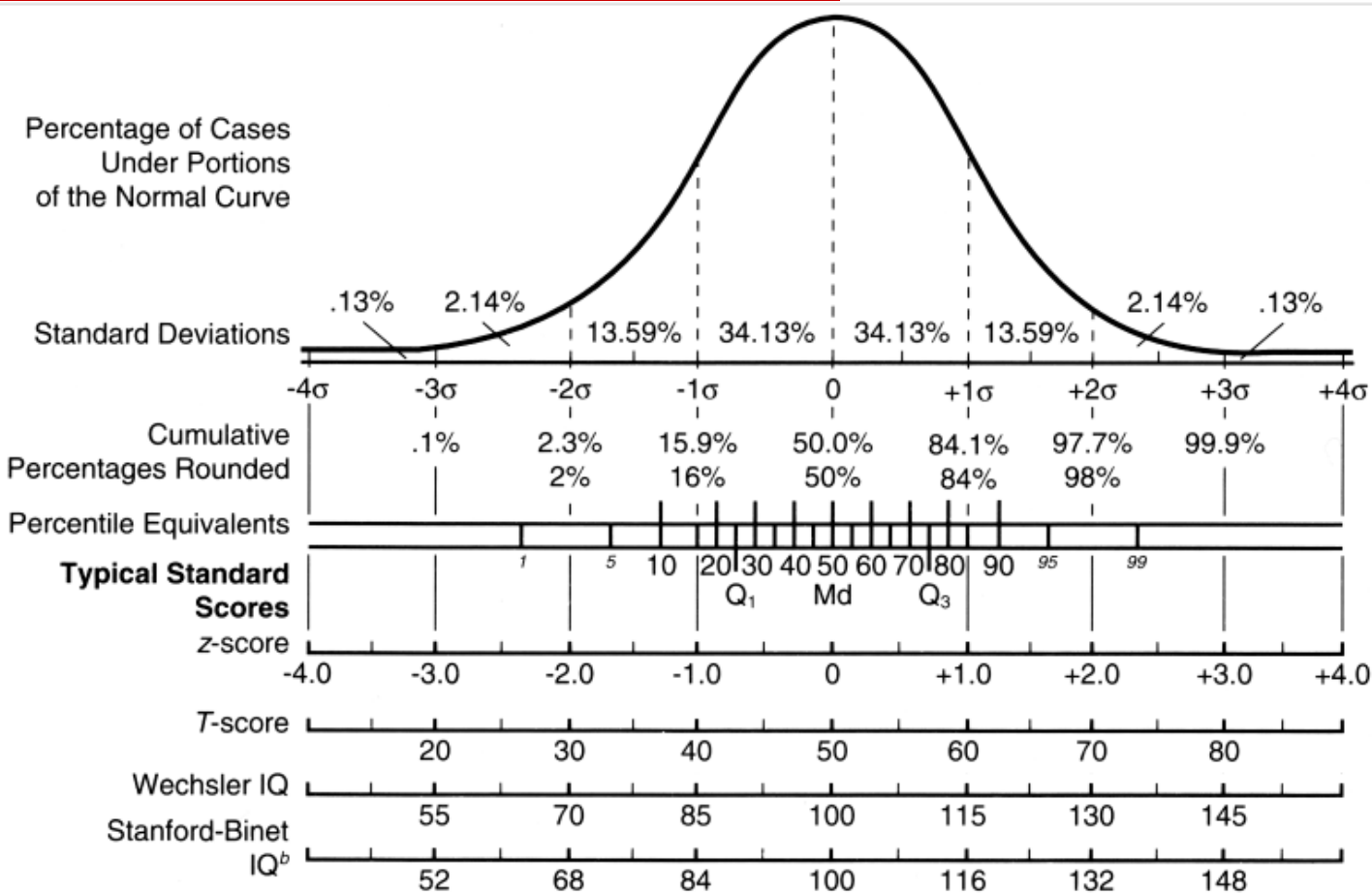
$$Skewness = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{SD(x)} \right)^3$$

$$Kurtosis = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{SD(x)} \right)^4$$

Mnohost normálních rozložení

- Jeden tvar, ale různé umístění na různých škálách (M) a různé roztažení (SD)
 - <http://www.intmath.com/counting-probability/normal-distribution-graph-interactive.php>
 - Standardizované rozložení $N(0,1)$
 - tj. převedení normálně rozložené proměnné na z-skóry
-

Kvantily standardního normálního rozložení $N(0;1)$ alias oblasti pod křivkou normálního rozložení



Kvantily přesněji v MS Excel

- **NORM.S.DIST**(z;1) – udává percentil odpovídající zadanému z-skóru, tj. kolik lidí má z-skór roven z nebo menší
 - Procento lidí v rozmezí z-skóruů =
 $\text{NORM.S.DIST}(\text{vyšší } z;1) - \text{NORM.S.DIST}(\text{nižší } z;1)$
- **NORM.S.INV**(p) – udává z-skór odpovídající zadanému percentilu

Bez toho **S.** poskytují funkce tutéž informaci pro normální rozložení s jiným M a SD

Kvantily přesněji postaru

<i>z</i>	<i>0.00</i>	<i>0.01</i>	<i>0.02</i>	<i>0.03</i>	<i>0.04</i>	<i>0.05</i>	<i>0.06</i>	<i>0.07</i>	<i>0.08</i>	<i>0.09</i>
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2089	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1336	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.9985

Statistické zkratky a značky

- různé systémy, je třeba dobře popisovat
- N, n = velikost vzorku, podvzorku(skupiny)
- X_i = skór i-té osoby u proměnné X
- x_i = deviační skór, odchylka od průměru
- M, m, \bar{x} = průměr
- SD, s = směrodatná odchylka
- VAR, s^2 = rozptyl



**Yo Momma
is so Mean,
she has no
Standard Deviation**