Přednáška 4

Modelová rozložení

Normální rozložení jako statistický model

Aplikace modelových rozložení

Přehled modelových rozložení

Anotace

- Klasickým postupem statistické analýzy je na základě vzorku cílové populace identifikovat typ a charakteristiky modelového rozložení dat, využít jeho matematického modelu k popisu reality a získané výsledky zobecnit na hodnocenou cílovou populaci.
- Využití tohoto přístupu je možné pouze v případě shody reálných dat s modelovým rozložením, v opačném případě hrozí získání zavádějících výsledků.
- Nejklasičtějším modelovým rozložením, od něhož je odvozena celá řada statistických analýz je tzv. normální rozložení, známé též jako Gaussova křivka.

All models are wrong but some are useful.

George Box, 1978

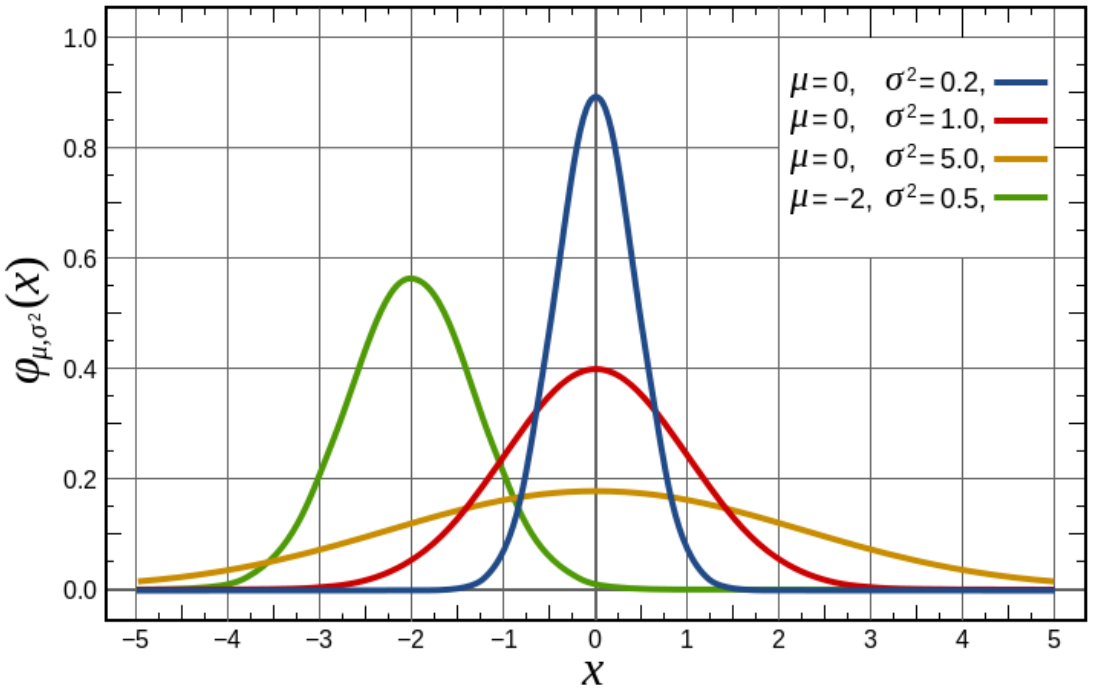
Normální rozdělení

- Nejklasičtějším modelovým rozložením, od něhož je odvozena celá řada statistických analýz je tzv. normální rozložení, známé též jako Gaussova křivka.

- Popisuje rozdělení pravděpodobnosti spojité náhodné veličiny: např. výška v populaci, chyba měření…

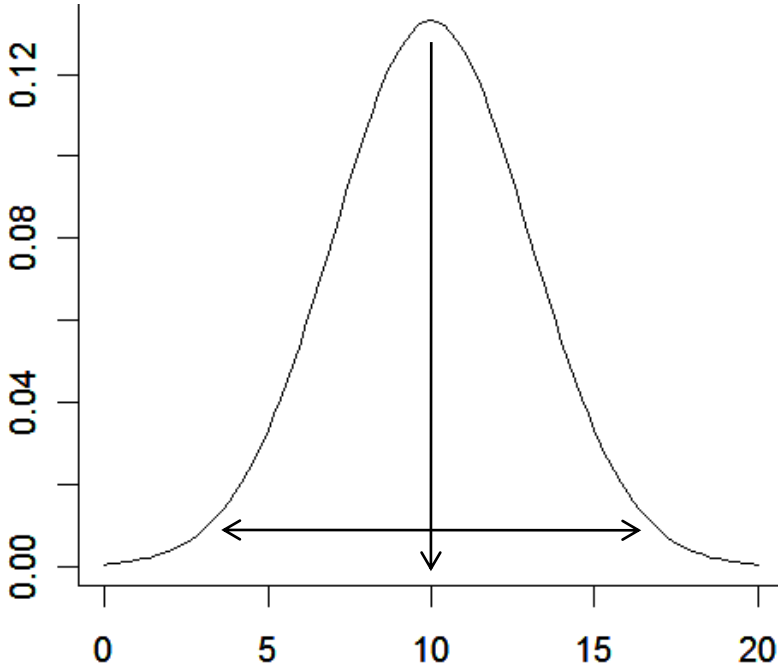
- Je kompletně popsáno dvěma parametry:
  - μ – střední hodnota
  - σ^2 – rozptyl
  - Označení: N(μ, σ2)

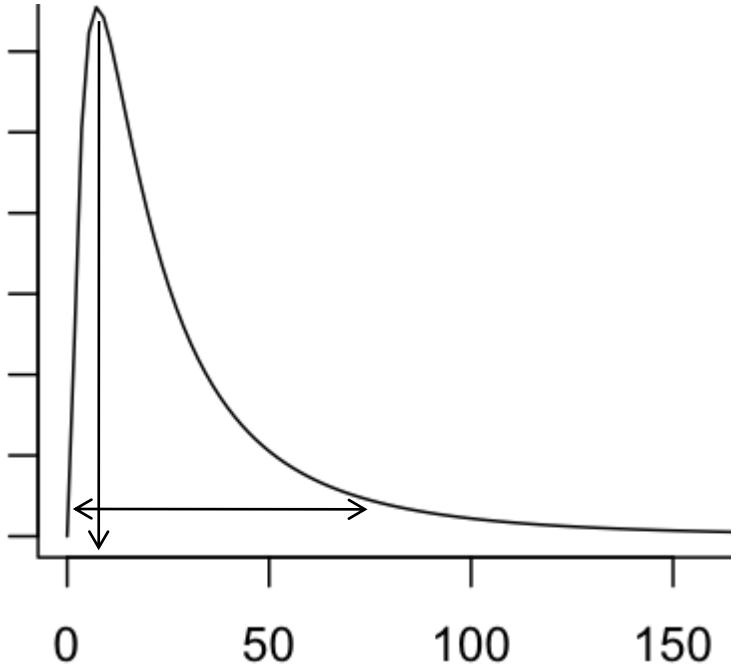
- Normalita je klíčovým předpokladem řady statistických metod
- Pro ověření normality existuje řada testů a grafických metod



Popis rozdělení kvantitativních dat: co chceme u dat popsat?

- Kvantitativní data – těžiště a rozsah pozorovaných hodnot.





Výpočet charakteristik normálního rozdělení: průměr

- μ – průměr rozdělení (cílová populace)
- $\bar{x}$ – průměr rozložení vzorkovaných dat (odhad průměru cílové populace)
- Průměr lze spočítat z libovolných kvantitativních dat, ale pouze za některých situací jej lze považovat za ukazatel středu dat (symetrické, normální rozdělení dat)
- Odlehlé hodnoty a asymetrie dat výrazně ovlivňují výsledek výpočtu průměru

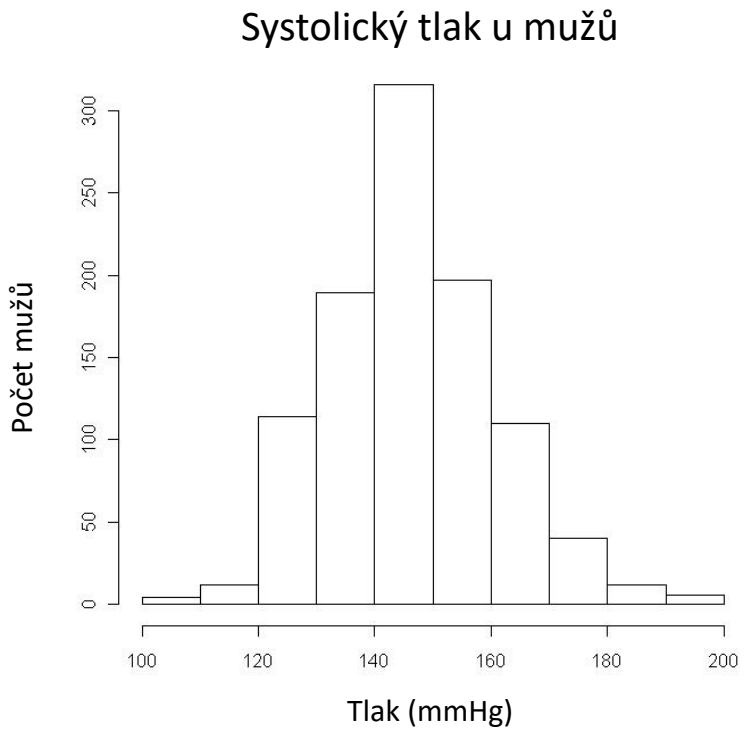
N=5

Objekt	Hodnota
x_1	5
x_2	3
x_3	4
x_4	7
x_5	2

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{21}{5} = 4{,}2$$

Průměr vs. medián

- Máme-li symetrická data, je výsledek výpočtu průměru i mediánu podobný.
- Vše je OK.



Průměr = 149,9 mmHg

Medián = 150,0 mmHg

Průměr vs. medián

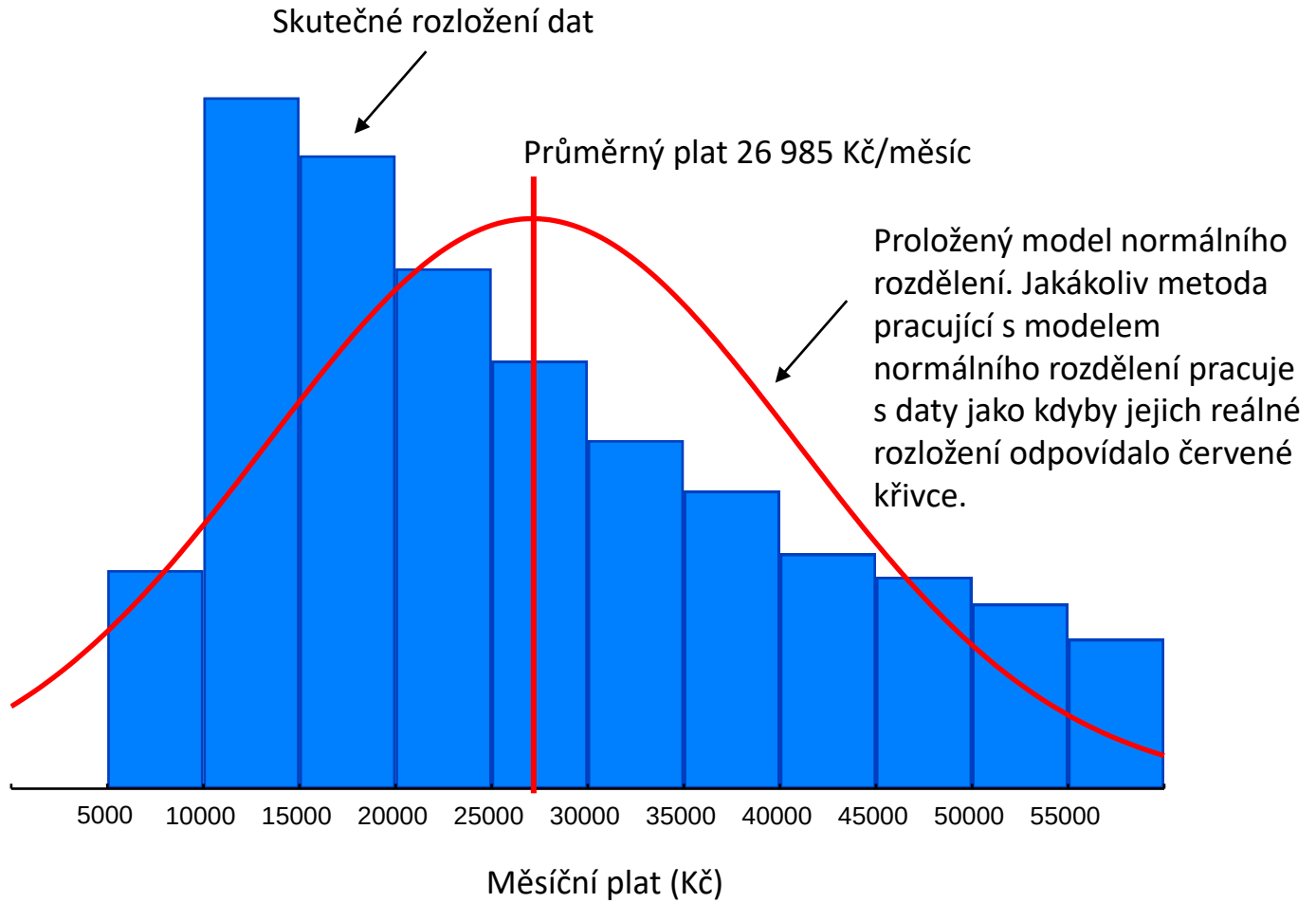
- Nemáme-li symetrická data, je výsledek výpočtu průměru i mediánu rozdílný.
- Není to OK. Výpočet průměru je v tuto chvíli nevhodný!

- Příklad 1: známkování ve škole
  - Student A: 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 5

    Průměr = 1,35 Medián = 1,00
  - Student B: 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 2

    Průměr = 1,13 Medián = 1,00

- Příklad 2: plat v ČR



Popis „těžiště“ – míry polohy

- Mějme pozorované hodnoty: $x_1, x_2, \ldots, x_n$
- Seřaďme je podle velikosti: $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$

- **Minimum a maximum** – nejmenší a největší pozorovaná hodnota nám dávají obraz o tom, kde se na ose x pohybujeme.

$$x_{\min} = x_{(1)}$$
$$x_{\max} = x_{(n)}$$

- **Průměr** – charakterizuje hodnotu, kolem které kolísají ostatní pozorované hodnoty. Je to fyzikální obraz těžiště stejně hmotných bodů ose x.

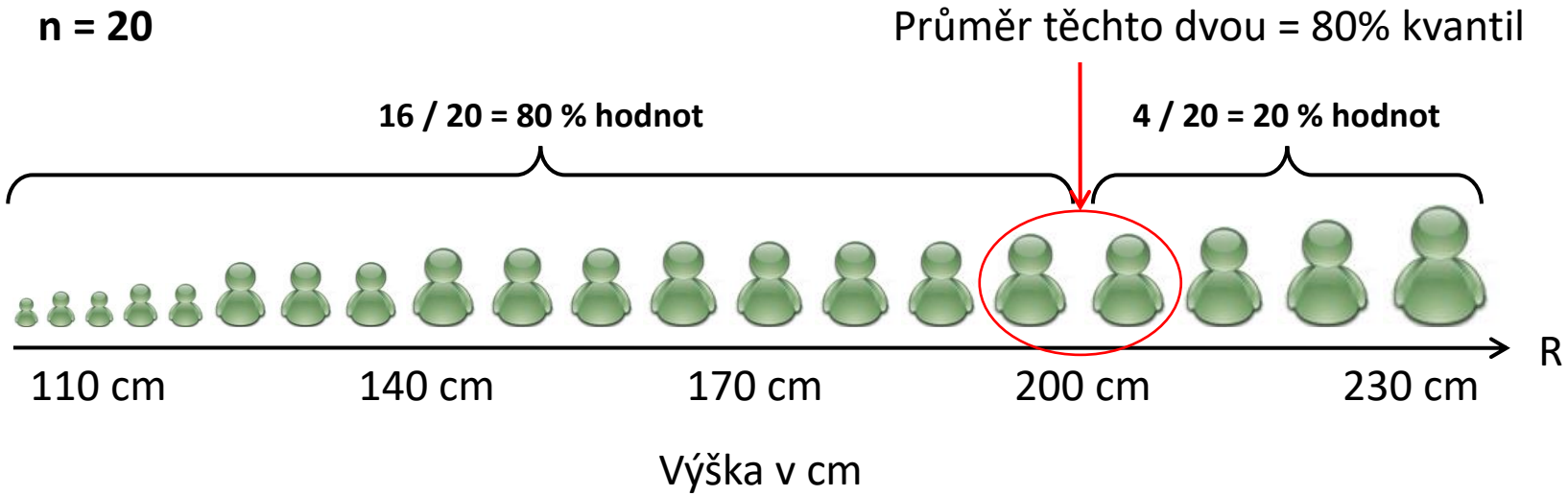
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- **Medián** – je to prostřední pozorovaná hodnota. Dělí pozorované hodnoty na dvě půlky, půlka hodnot je menší a půlka hodnot je větší než medián.

$$\tilde{x} = x_{((n+1)/2)} \quad \text{pro } n \text{ liché}$$
$$\tilde{x} = \tfrac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right) \quad \text{pro } n \text{ sudé}$$

Pojem kvantil

- Laicky lze kvantil definovat jako číslo na reálné ose, které rozděluje pozorovaná data na dvě části: p% kvantil rozděluje data na p % hodnot a (100-p) % hodnot.
- Máme soubor 20 osob, u nichž měříme výšku. Chceme zjistit 80% kvantil souboru pozorovaných dat.

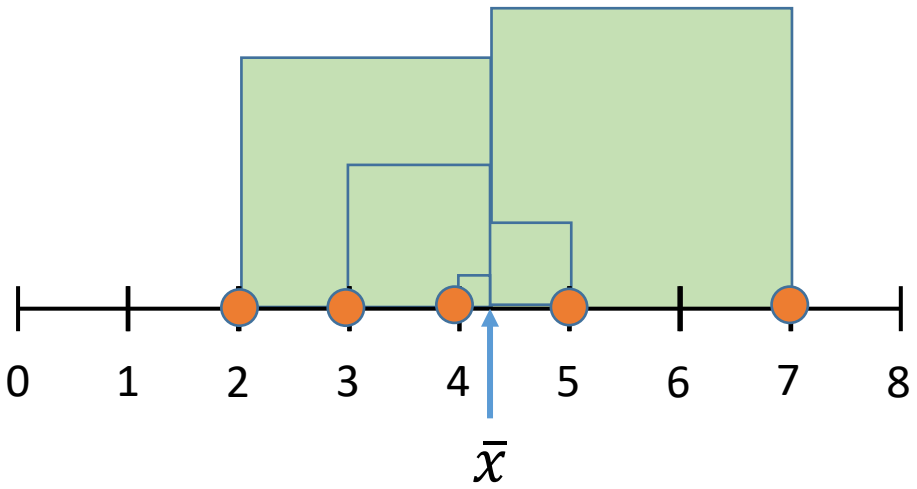


Výpočet charakteristik normálního rozdělení: rozptyl a směrodatná odchylka

- σ^2 – rozptyl rozdělení (cílová populace)
- s^2 – rozptyl rozložení vzorkovaných dat (odhad rozptylu cílové populace)

N=5

Objekt	Hodnota
x_1	5
x_2	3
x_3	4
x_4	7
x_5	2



$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1} = \frac{14{,}8}{4} = 3{,}7$$

$$s = \sqrt{s^2} = \sqrt{3{,}7} = 1{,}92$$

- Směrodatná odchylka (s, SD=standard deviation) = druhá odmocnina z rozptylu (snazší interpretovatelnost)
- N-1 nebo N ? Dělení N-1 je výpočet rozptylu vzorku, dělení N je pro celou populaci (výjimečně)

Popis „rozsahu“ – míry variability

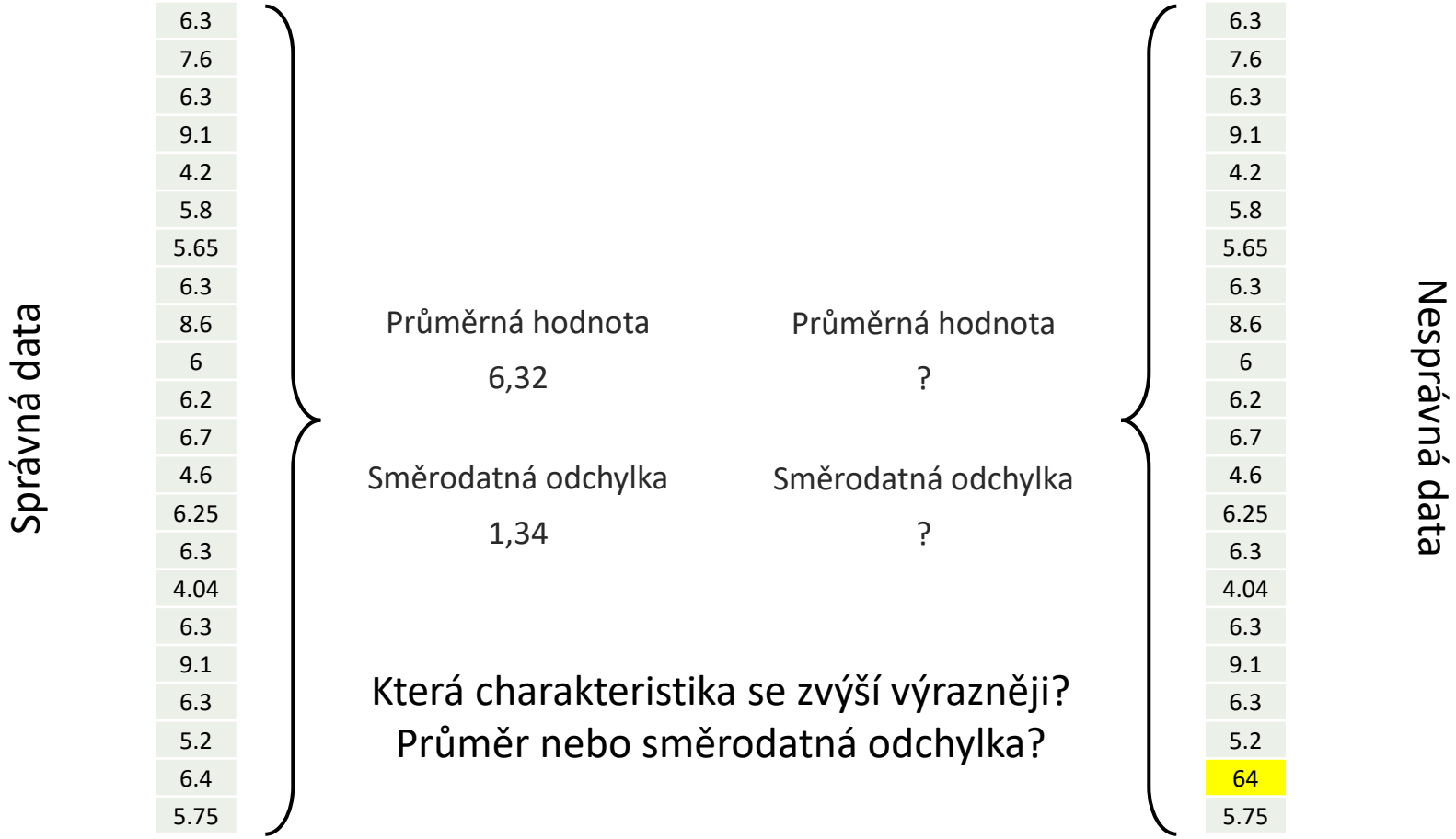
- Nejjednodušší charakteristikou variability pozorovaných dat je rozsah hodnot (rozpětí) = maximum – minimum. Je snadno ovlivnitelný netypickými (odlehlými) hodnotami.
- **Kvantilové rozpětí** je definováno p% kvantilem a (100-p)% kvantilem a je méně ovlivněno odlehlými hodnotami. Speciálním případem je kvartilové rozpětí, které pokrývá 50 % pozorovaných hodnot.
- **Rozptyl** – průměrný čtverec odchylky od průměru. Velmi ovlivnitelný odlehlými hodnotami.

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - x)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - nx^2\right)$$

- **Směrodatná odchylka** – odmocnina z rozptylu. Výhodou směrodatné odchylky je, že má stejné jednotky jako pozorovaná data.
- **Koeficient variance** - podíl směrodatné odchylky ku průměru (u normálního rozložení by se 95% hodnot mělo vejít do průměr ±3 SD), pokud je SD větší než 1/3 průměru jsou teoreticky pravděpodobné záporné hodnoty v rozložení – ukazatel problémů s normalitou dat

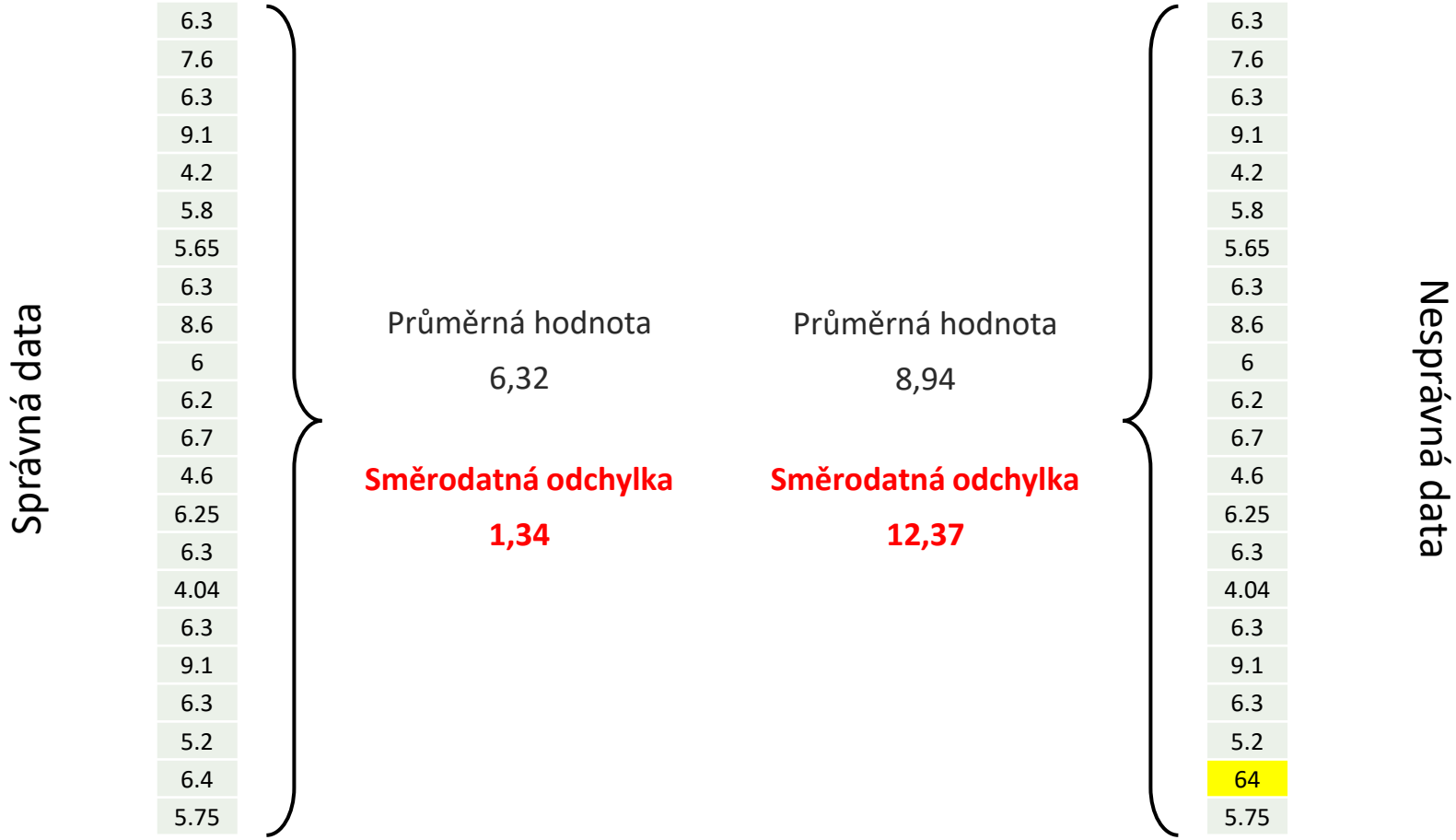
Normální rozdělení: vliv odlehlé hodnoty na popisné statistiky

- Cílem je určit průměrnou hladinu cholesterolu vybrané populace mužů (hodnoty v mmol/l)



Normální rozdělení: vliv odlehlé hodnoty na popisné statistiky

- Cílem je určit průměrnou hladinu cholesterolu vybrané populace mužů (hodnoty v mmol/l)



Identifikace odlehlých hodnot

- Na menších souborech stačí vizualizace.
- Na větších datových souborech nelze bez vizualizace a popisných statistik.

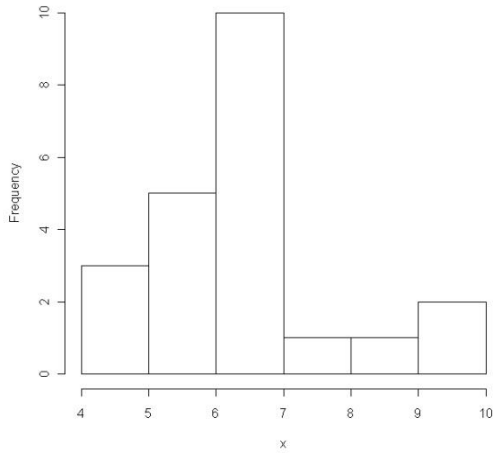
- Grafická identifikace: pomocí histogramu a box plotu.
- Identifikace pomocí popisných statistik: srovnání mediánu a průměru.

Identifikace odlehlých hodnot – příklad

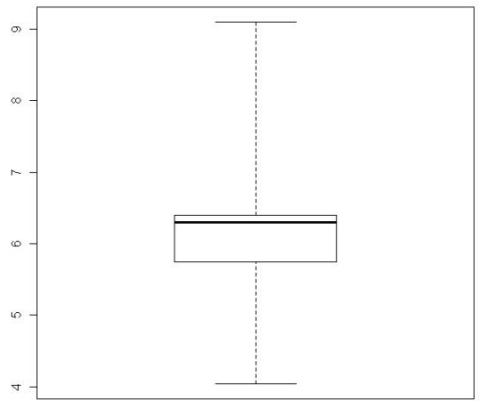
Správná data

6.3
7.6
6.3
9.1
4.2
5.8
5.65
6.3
8.6
6
6.2
6.7
4.6
6.25
6.3
4.04
6.3
9.1
6.3
5.2
6.4
5.75

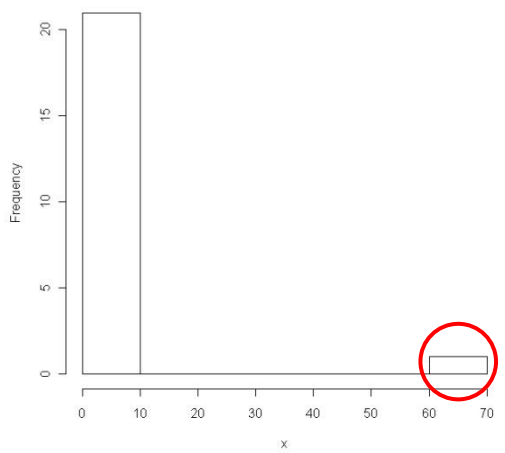
Histogram



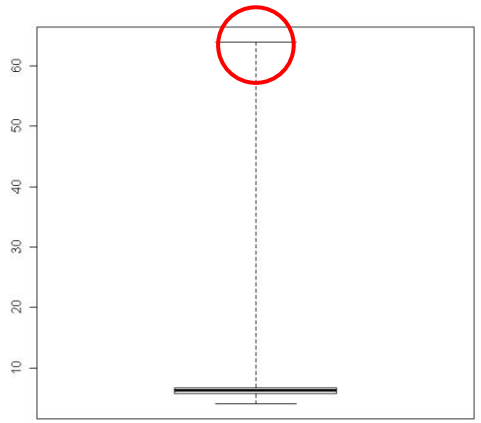
Box plot



Histogram



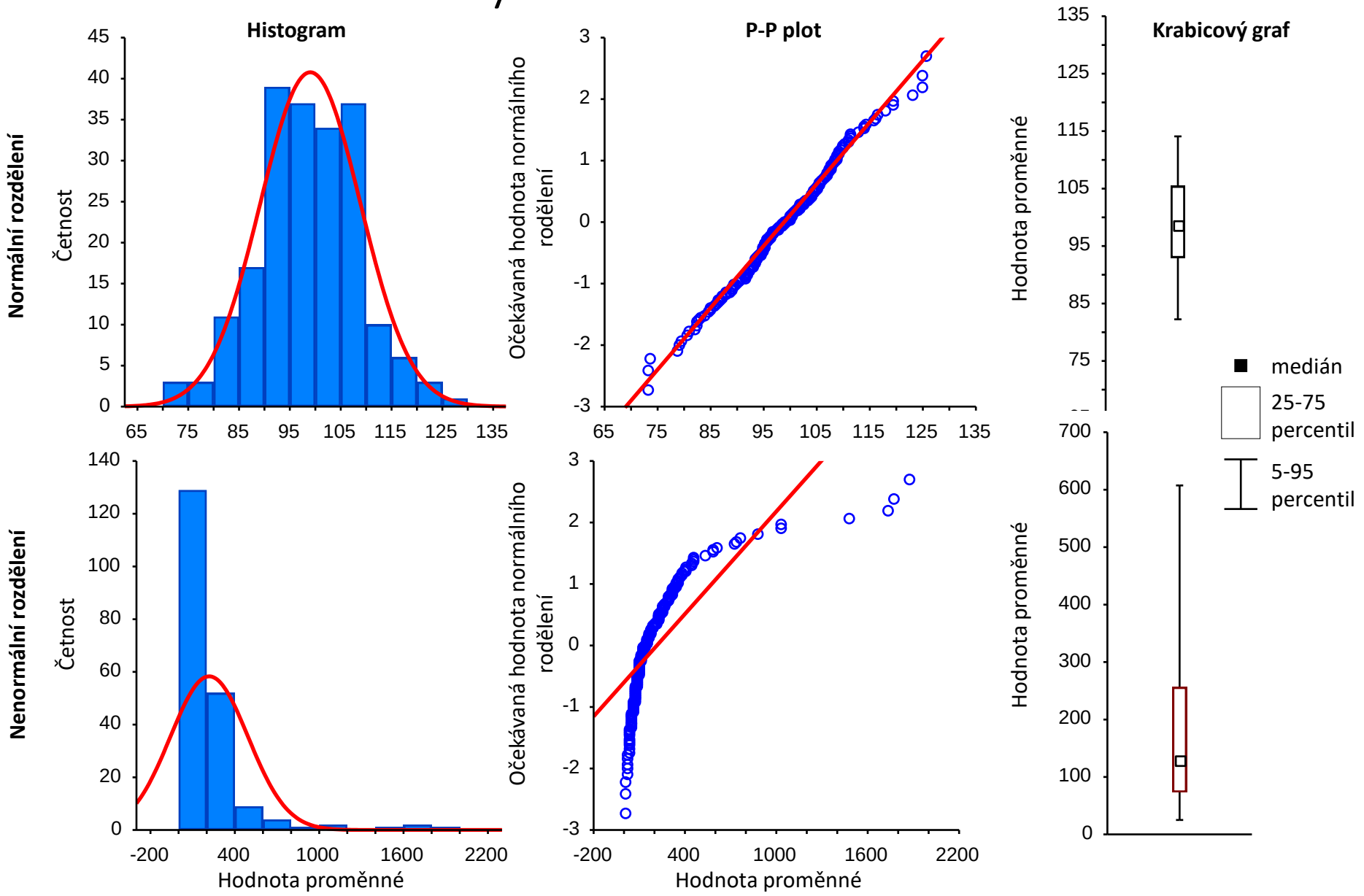
Box plot



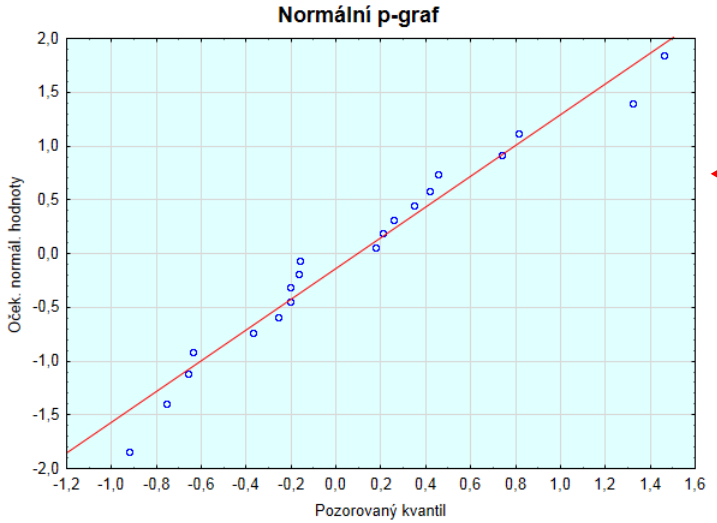
Nesprávná data

6.3
7.6
6.3
9.1
4.2
5.8
5.65
6.3
8.6
6
6.2
6.7
4.6
6.25
6.3
4.04
6.3
9.1
6.3
5.2
64
5.75

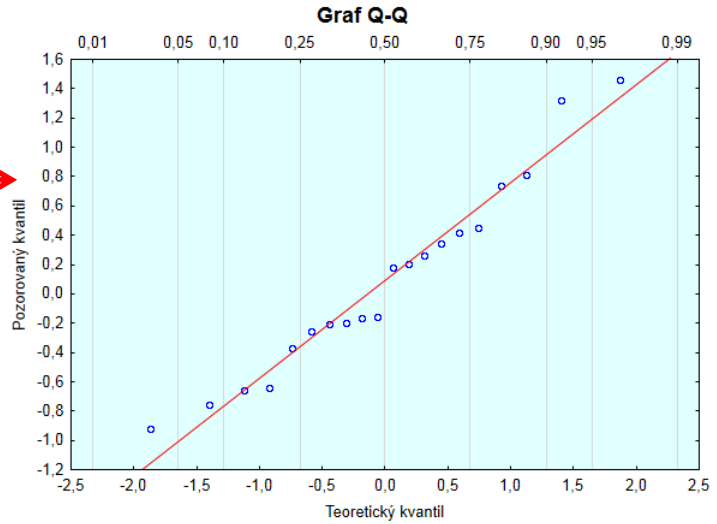
Vizuální hodnocení normality



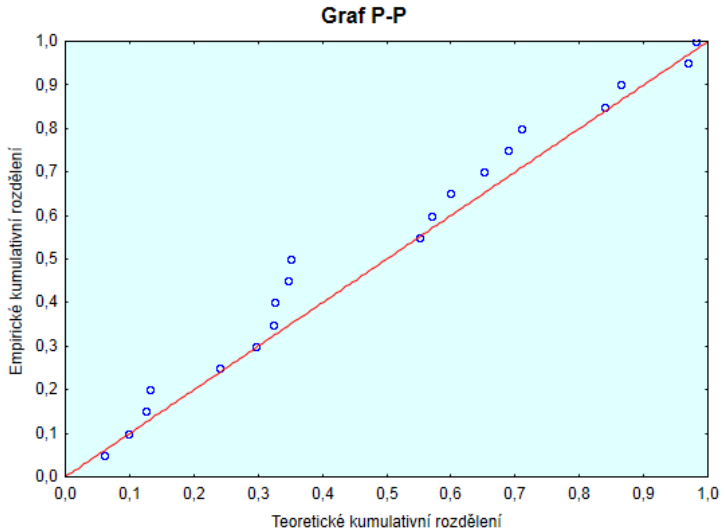
Rozdíl mezi N-P, Q-Q, P-P grafem



???



- ***Pouze výměna os***
- ***Znázorněn pozorovaný a teoretický kvantil***



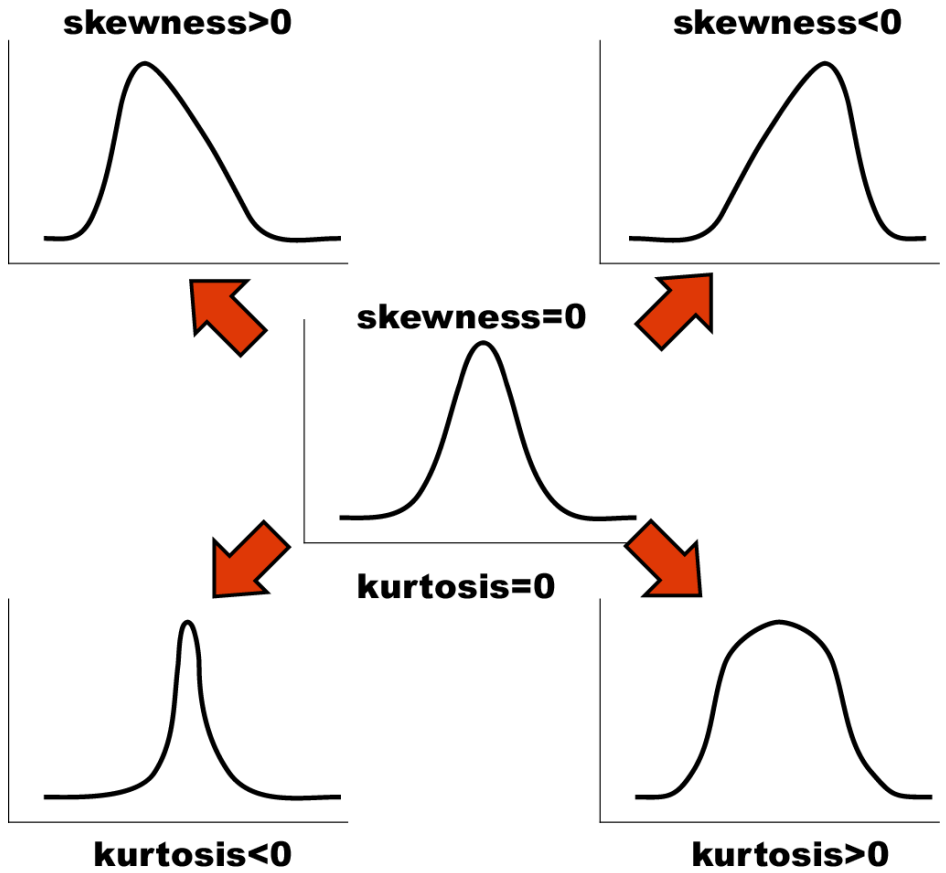
- ***Vykresleno kumulativní rozdělení***

PAMATUJ:

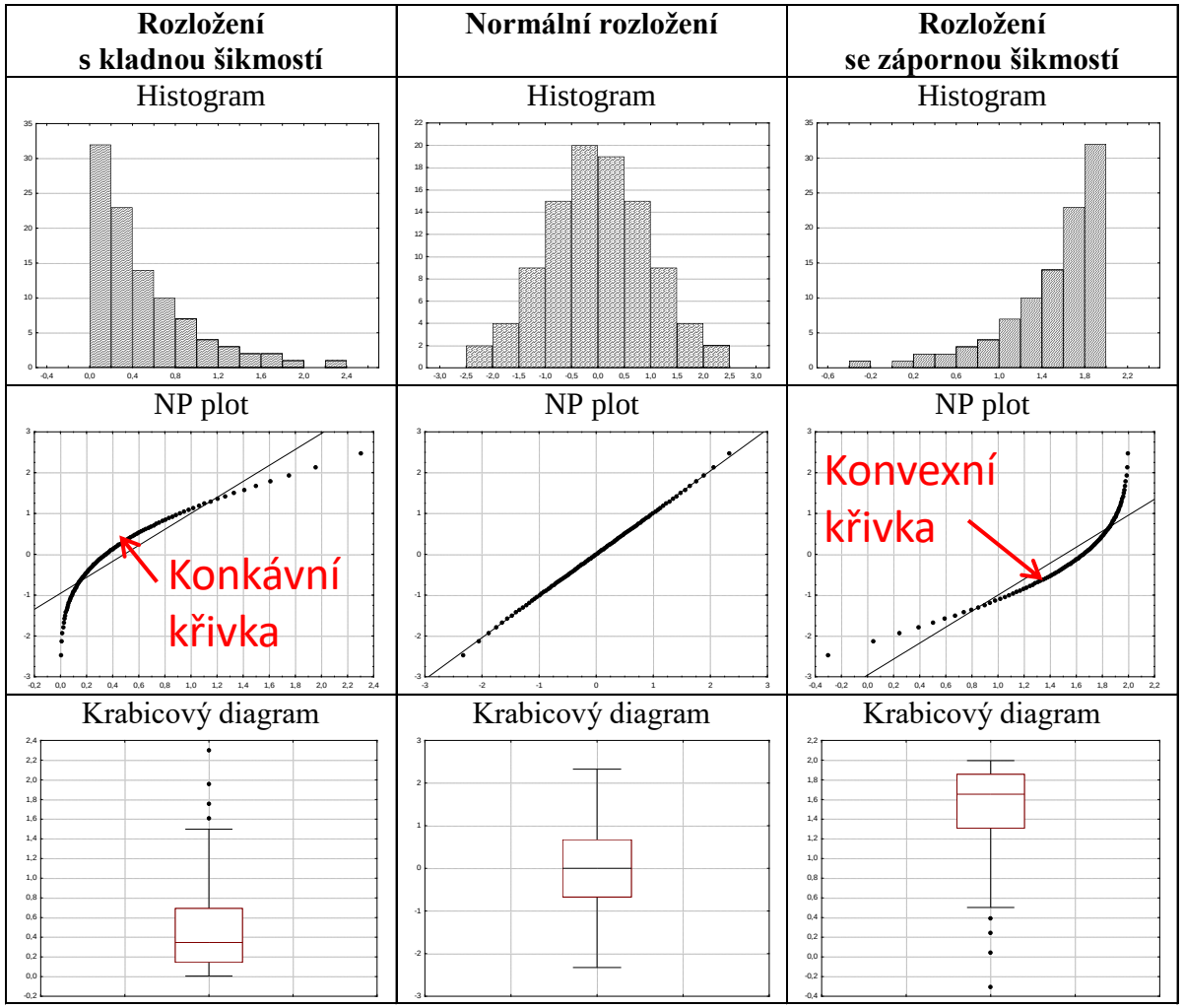
Pocházejí-li data z normálního rozložení, pak body budou ležet okolo přímky

Ukazatele tvaru rozložení

- **Skewness** – ukazatel „šikmosti“ rozložení, asymetrie rozložení
- **Kurtosis** – ukazatel „špičatosti/plochosti“ rozložení



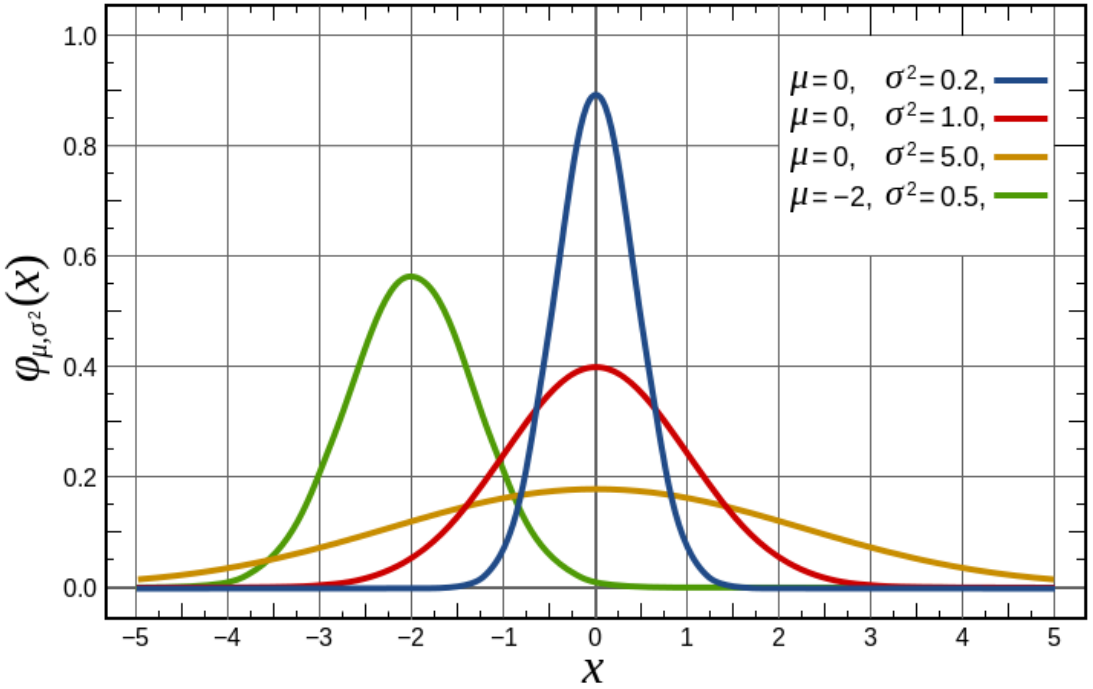
Jak se projeví asymetrie dat v diagnostických grafech?

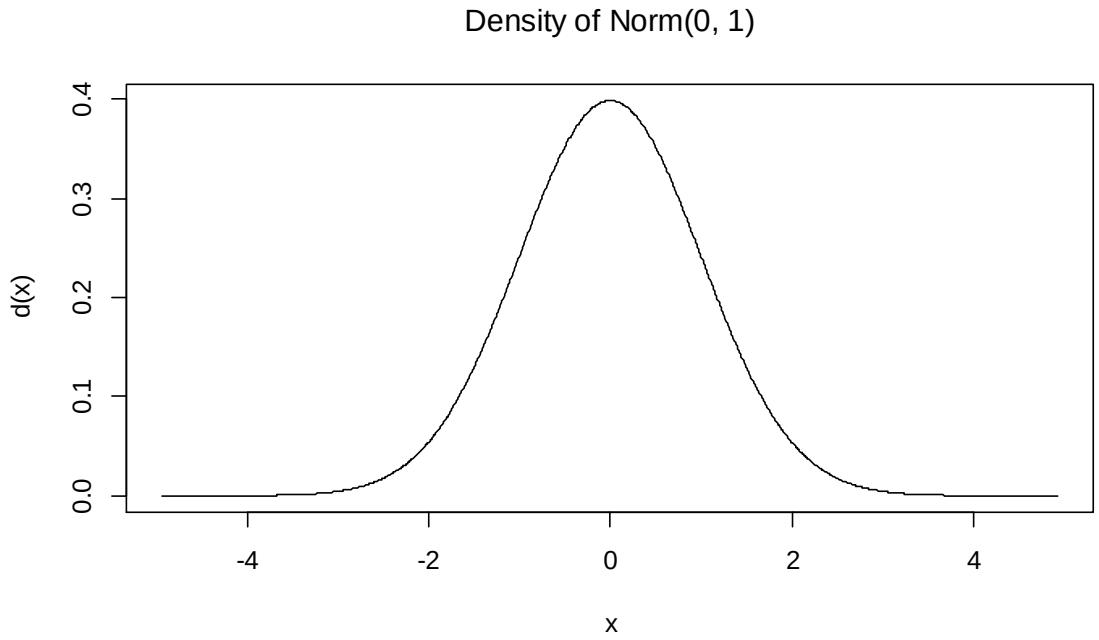


Výukové materiály: Výpočetní statistika, RNDr. Marie Budíková, Dr., 2011

Standardní normální rozdělení

- Speciální případ normálního rozdělení s N(μ=0, σ^2=1) - standardizovaná forma využívaná:
  - ve statistických výpočtech
  - pro srovnání extrémnosti / průměrnosti hodnot u proměnných s různými rozsahy nebo jednotkami
  - Jednoduchá interpretace – základní hodnoty vhodné zapamatovat





Přepočet na standardní normální rozdělení

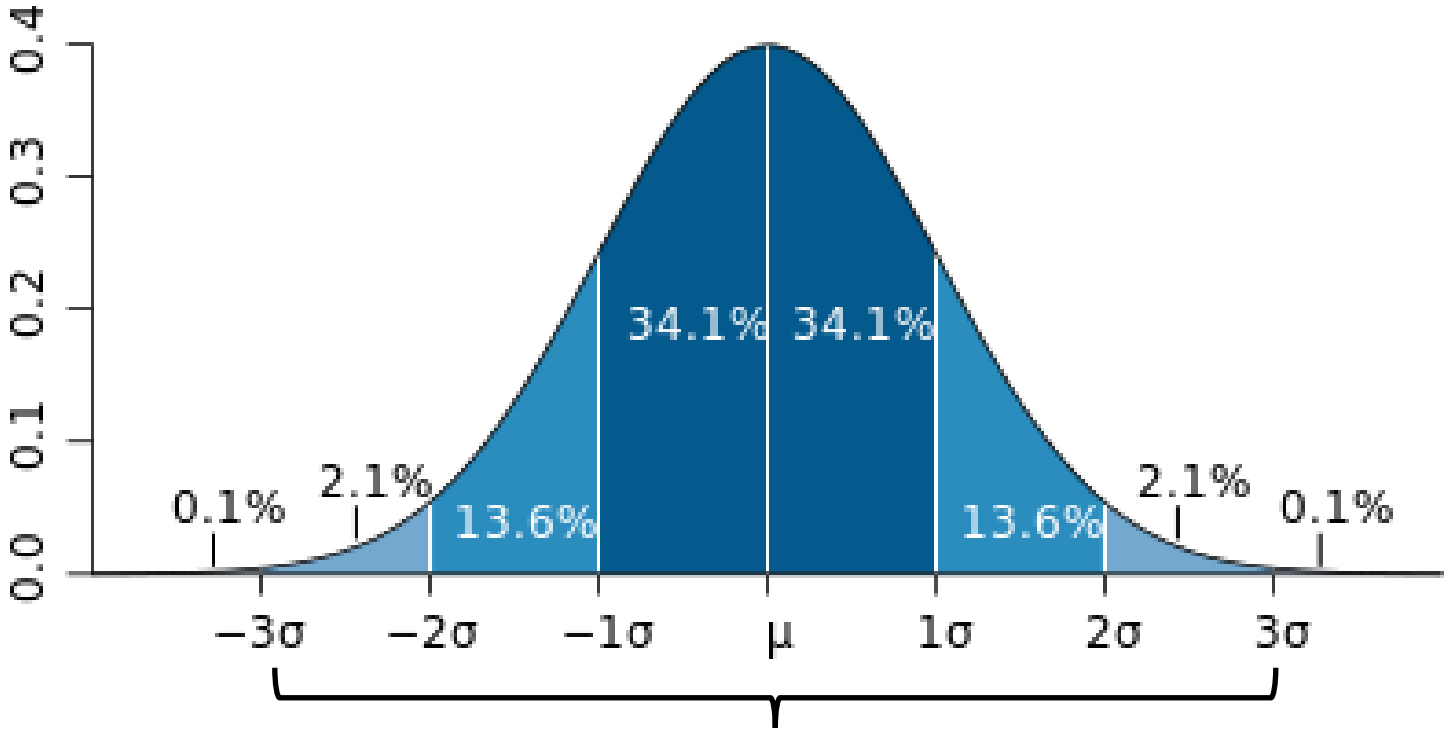
- Tzv. Z skóre – kromě statistických výpočtů využíváno např. v diagnostických skóre (osteoporóza) nebo pro srovnávání extrémnosti / průměrnosti proměnných s různými rozsahy nebo jednotkami (např. měření polutantů)
- Využití při výpočtu standardizovaných charakteristik (např. kovariance -> korelační koeficient)
- Ve vícerozměrné analýze používáno pro dosažení stejné váhy různých proměnných ve výpočtu
- Tabelovaná forma -> využití ve výpočtech

Objekt	Hodnota	Standardizovaná hodnota (z)
x_1	5	0.42
x_2	3	-0.62
x_3	4	-0.10
x_4	7	1.46
x_5	2	-1.14
průměr	4,2	0
s	1,92	1

$$z_i = \frac{x_{i-\mu}}{\sigma}$$

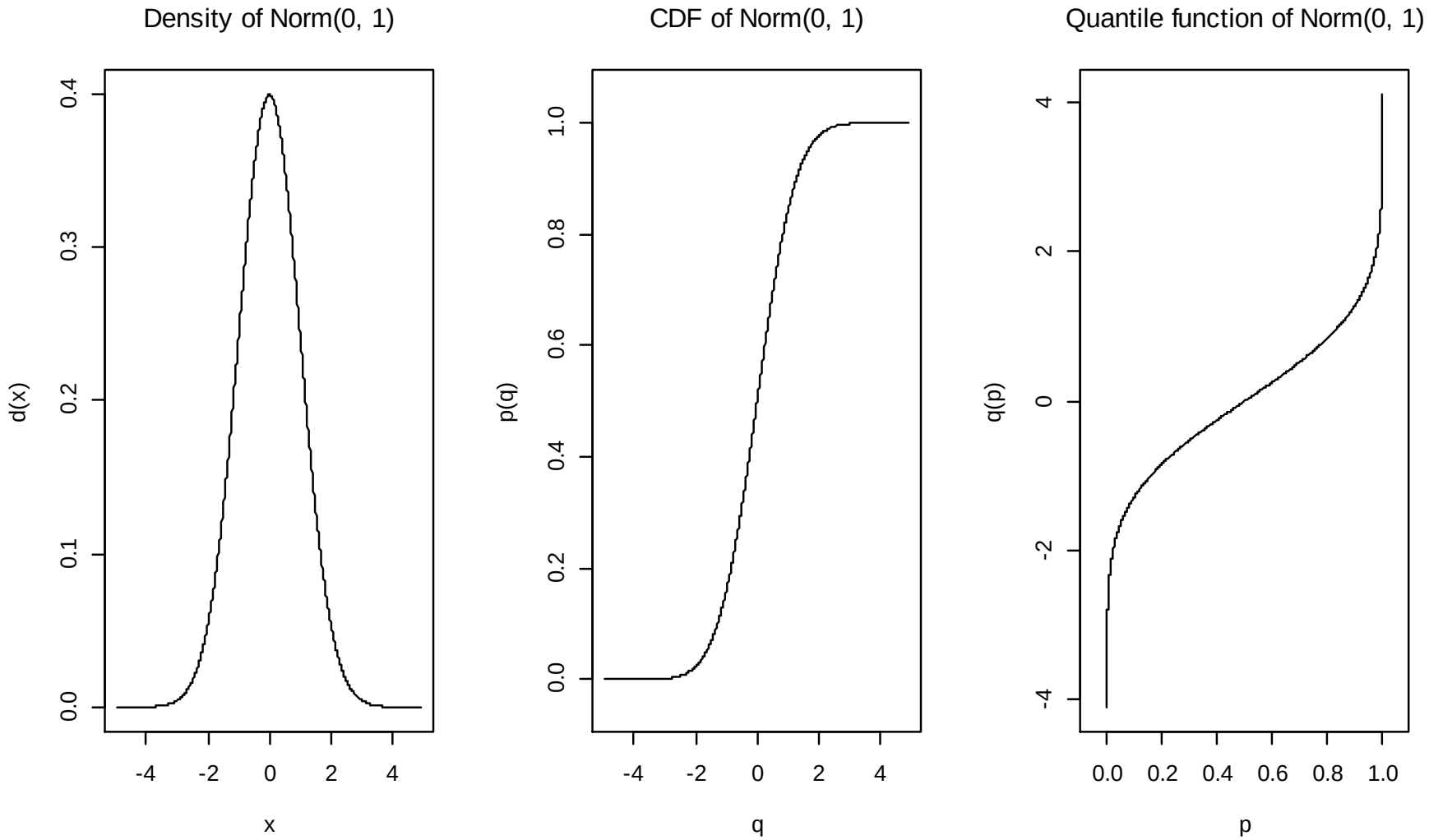
Pravidlo 3 sigma

- V rozmezí μ ± 3σ by se mělo vyskytovat 99,7 % všech hodnot
- Vhodné znát pro orientační posouzení rozsahu dat
- U proměnných, které nemohou být záporné využití pro orientační posouzení normality



99,7 % všech hodnot

Standardizované normální rozdělení a jeho charakteristiky

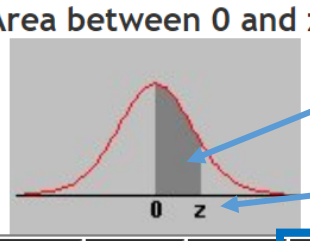


Statistické tabulky

- Přehledné vyjádření distribuční funkce pro modelová rozdělení
- V předpočítačovém období základní pomůcka, nyní hlavně výukový význam
- http://www.statsoft.com/Textbook/Distribution-Tables (potřebné i pro zkoušku)

Druhé desetinné místo hledaného **z**

Celá část a první desetinné místo hledaného **z**



Plocha pod křivkou standardního normálního rozdělení (= pravděpodobnost) mezi průměrem a hledaným **z**

Hledané z (hodnota standardního normálního rozdělení)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852

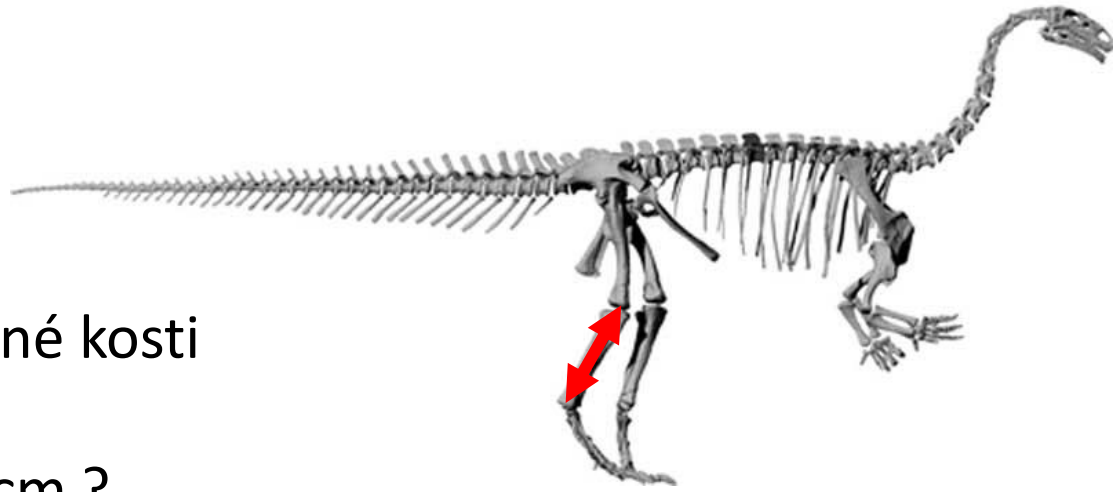
Plocha pod křivkou standardního normálního rozdělení mezi průměrem a hledaným **z**
Zde pro **z=0.46** to je **0.1772** (mezi průměrem a **z=0.46** leží **17.7%** rozdělení)

Využití statistických modelů

1. Máme nějaký znak v populaci, který chceme pro účely analýz nahradit statistickým modelem (de facto to děláme při každém výpočtu průměru, který považujeme za ukazatel středu)
2. Ověříme předpoklad, že je znak rozložen podle daného modelu = **Platí vybraný model?** Např. vizuální posouzení normality nebo její testování.
3. Spočítáme charakteristiky modelu (průměr a směrodatná odchylka v případě normálního rozdělení)
4. Převedeme na standardní formu modelu (standardní normální rozdělení v případě normálního rozdělení)
5. Využijeme známé vlastnosti rozdělení pro odpověď na položené otázky (distribuční funkce, její hodnoty ve statistických tabulkách)

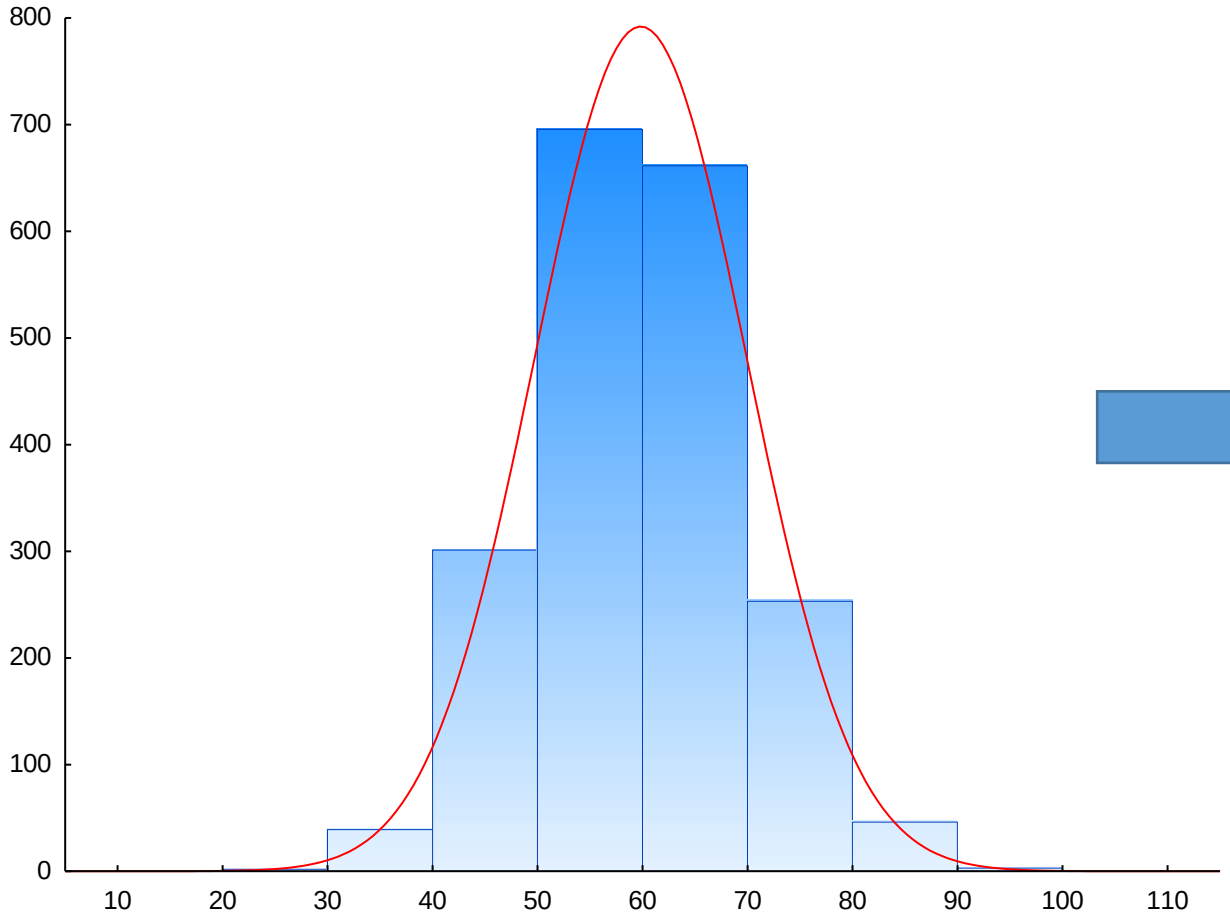
Příklad aplikace modelu normálního rozdělení

- Máme data z průzkumu kostí prehistorického zvířete
  - N=2 000
  - Průměrná délka = 60 cm
  - Směrodatná odchylka = 10 cm

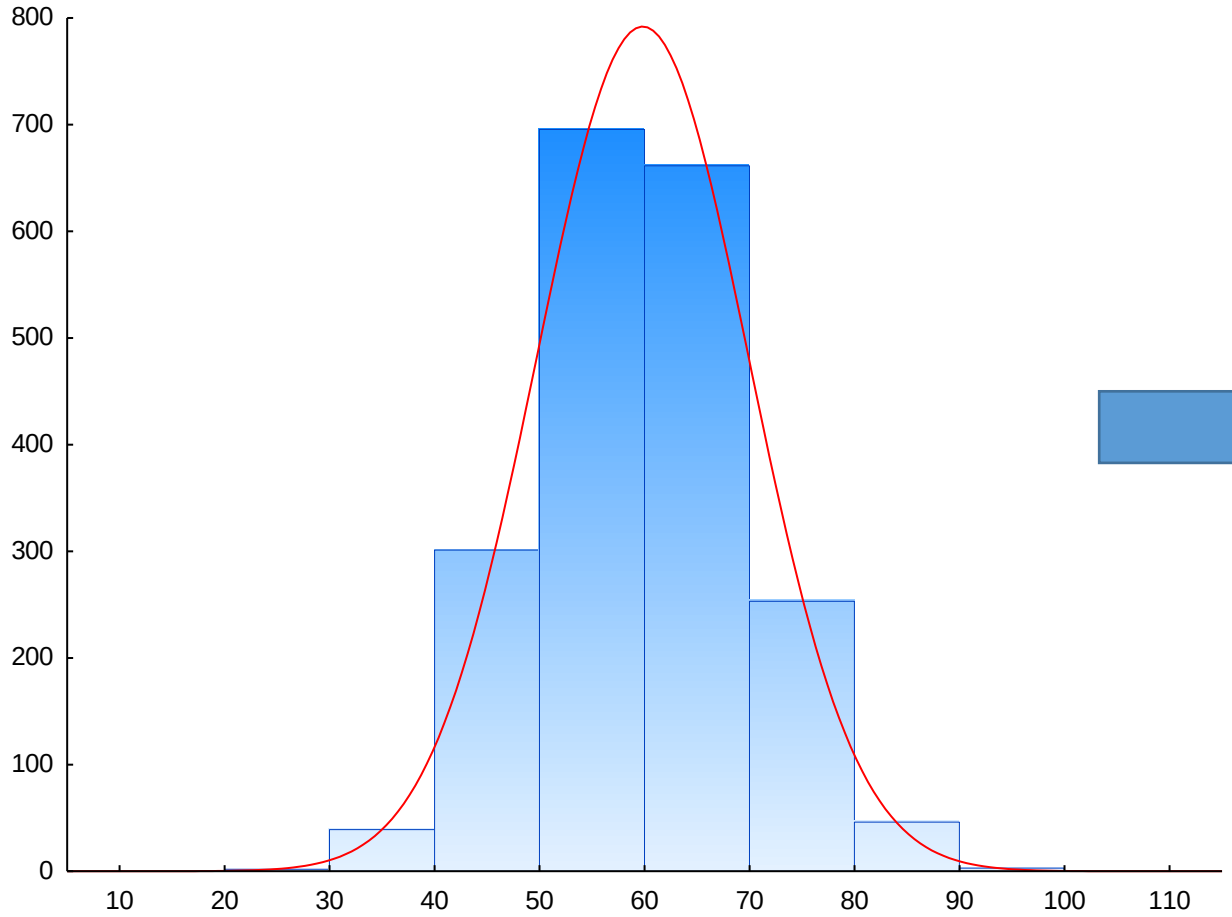


- Výzkumné otázky:
  - Jaká je pravděpodobnost, že by velikost dané kosti překročila velikost 66 cm?
  - Kolik kostí mělo zřejmě délku větší než 66 cm ?
  - Jaký podíl kostí ležel svou délkou v rozsahu od 60 cm do 66 cm ?

Ověření rozložení dat a výběr statistického modelu



Ověření rozložení dat a výběr statistického modelu

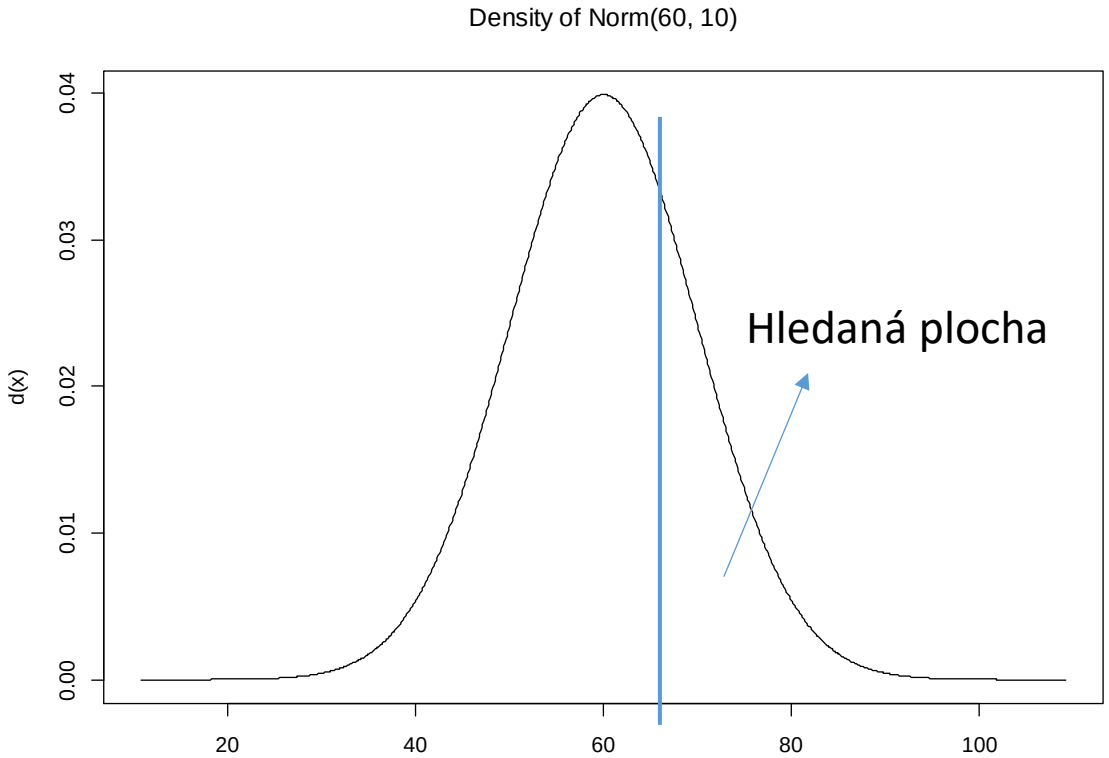


Předpoklad normálního rozdělení dat se zdá oprávněný.

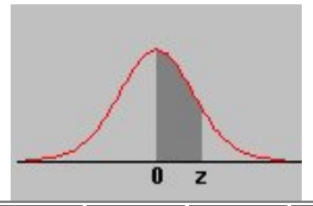
Jaká je pravděpodobnost, že by velikost dané kosti překročila velikost 66 cm?

- Přepočet hledané hodnoty na standardizovanou formu normálního rozdělení

$$z = \frac{x - \mu}{\sigma} = \frac{66 - 60}{10} = 0{,}6$$



Area between 0 and z



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852

$$P(x > 66) = 1 - P(x \le 66) = 1 - P\left(\frac{x - m}{s} \le \frac{66 - 60}{10}\right) = 1 - F(0{,}6) = 0{,}27425$$

Aplikace modelu normálního rozdělení

- Kolik kostí mělo zřejmě délku větší než 66 cm ?

$$P(x > 66) * n = 0{,}27425 * 2000 = 548$$

- Jaký podíl kostí ležel svou délkou v rozsahu x od 60 cm do 66 cm ?

$$P(60 < x < 66) = P\left(\frac{60-60}{10} < Z < \frac{66-60}{10}\right) = F(0{,}6) - F(0) = 0{,}22575$$

  - 22,6% kostí leží v rozsahu 60-66cm

Stručný přehled modelových rozložení I

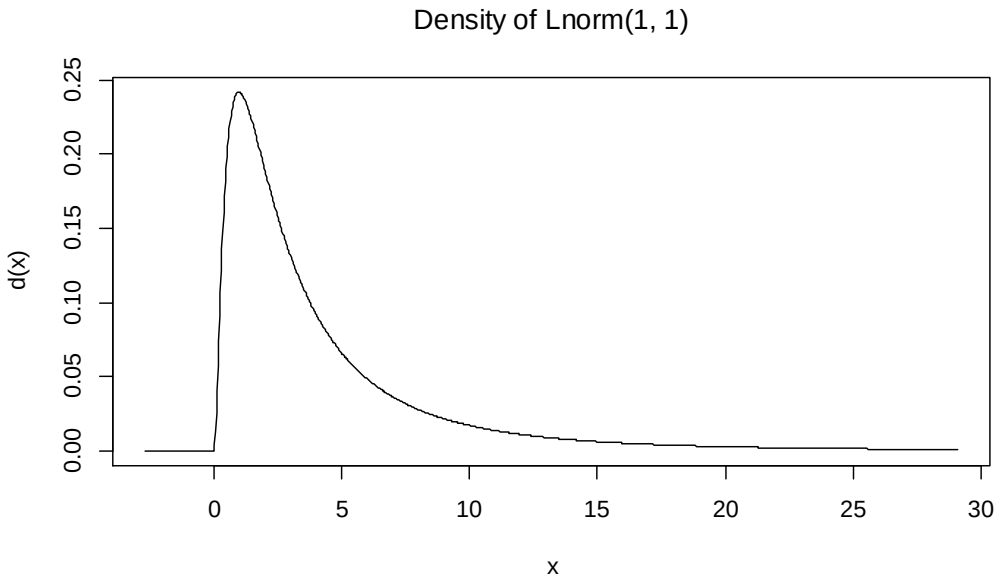
Rozložení	Parametry	Stručný popis
Normální	Průměr (μ) Rozptyl (σ^2)	Symetrická funkce popisující intervalovou hustotu četnosti; nejpravděpodobnější jsou průměrné hodnoty znaku v populaci.
Log-normální	Medián Geometrický průměr Rozptyl (σ^2)	Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení.
Weibullovo	α - parametr tvaru β - parametr rozsahu hodnot	Změnou parametru a lze modelovat distribuci doby přežití, např. stresovaného organismu. Rozložení využívané i jako model k odhahu LC_{50} nebo EC_{50} u testů toxicity.
Rovnoměrné	Medián Geometrický průměr Rozptyl (σ^2)	Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení.
Triangulární	$f(x) = [b - ABS(x - a)] / b^2$ $a - b < x < a + b$	Pravděpodobnostní funkce pro typ rozložení, kdy jsou střední hodnoty výrazně pravděpodobnější než hodnoty okrajové.
Gamma	Parametry distribuční funkce: α - parametr tvaru β - parametr rozsahu hodnot	Umožňuje flexibilně modelování distribučních funkcí nejrůznějších tvarů. Např. χ^2 rozložení je rozložení typu Gamma. Gamma rozložení s a = 1 je známo jako exponenciální rozložení.

Stručný přehled modelových rozložení

Rozložení	Parametry	Stručný popis
Beta	Parametry distribuční funkce: α - parametr tvaru β - parametr rozsahu hodnot	Pravděpodobnostní funkce pro proměnnou omezenou rozsahem do intervalu [0; 1]. Je matematicky komplikovanější, ale velmi flexibilní při popisu změn hodnot proměnné v ohraničeném intervalu.
Studentovo	Stupně volnosti - uvažuje velikost vzorku Průměr Rozptyl	Simuluje normální rozložení pro menší vzorky čísel. Pro větší soubory (n > 100) se limitně blíží k normálnímu rozložení.
Pearsonovo	Stupně volnosti - uvažuje velikost vzorku	Slouží především k porovnání četností jevů ve dvou a více kategoriích. Používá se k modelování rozložení odhadu rozptylu normálně rozložených dat.
Fisher-Snedecorovo	Dvojí stupně volnosti - uvažuje velikost dvou vzorků	Používá se k testování hodnot průměrů - F test pro porovnání dvou výběrových rozptylů; F test, ANOVA atd.

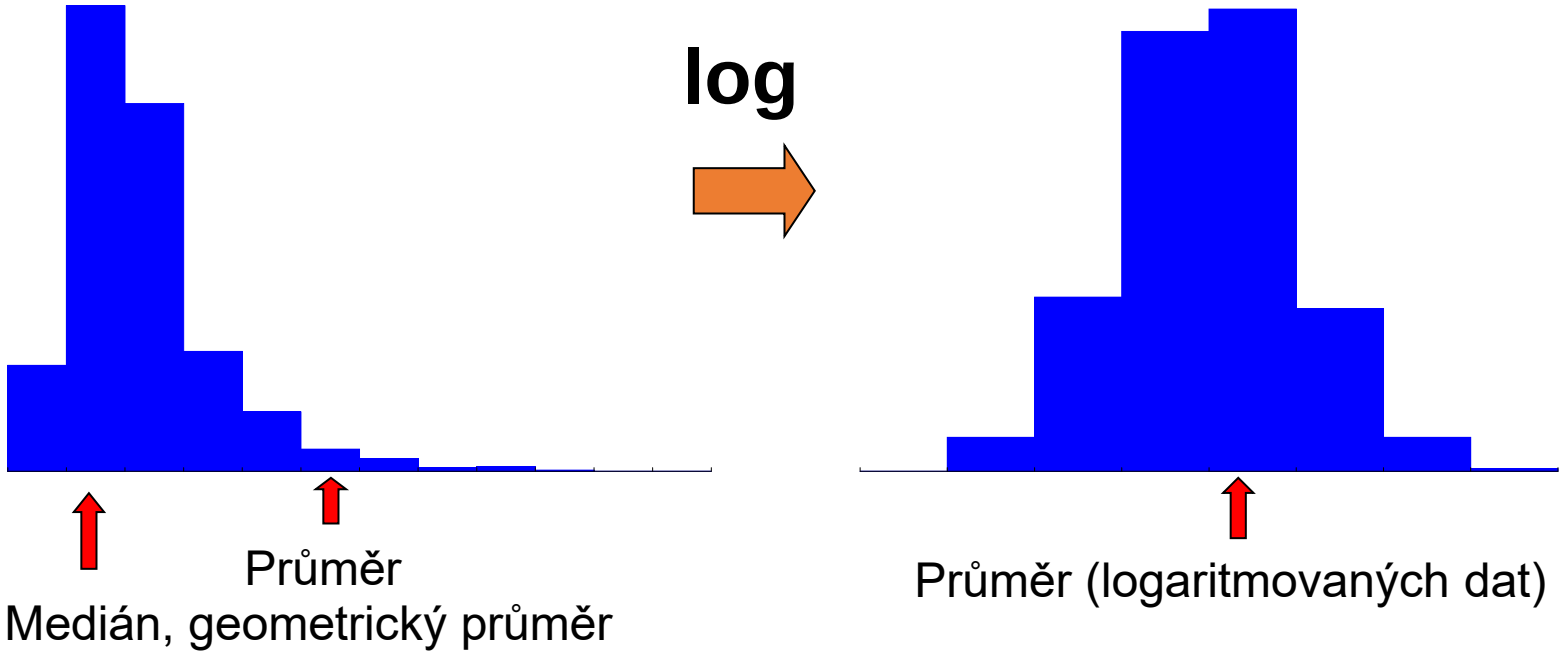
Lognormální rozdělení

- Asymetricky rozložená data – velmi častá v biologii (ale i jinde, např. platy)
- Spolu s normálním rozdělením nejčastější model
- S rozdělením je spjat geometrický průměr jako ukazatel středu

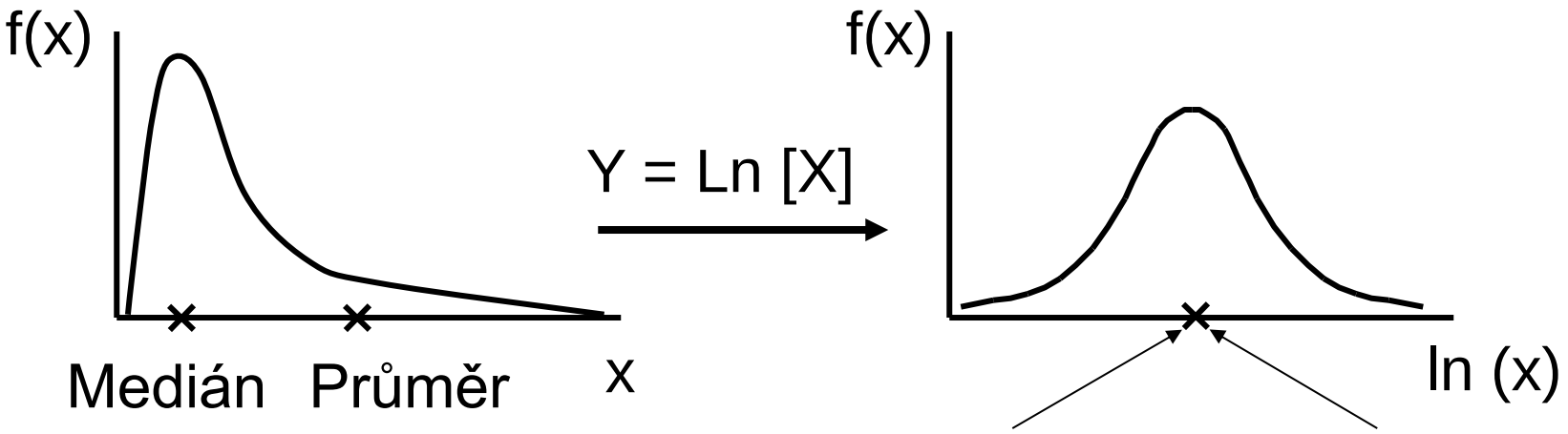


Logaritmická transformace

- Geometrický průměr – antilogaritmus průměru logaritmovaných dat, je vhodný pro doleva asymetrická data (lognormální rozložení), která jsou v biologii velmi častá, jeho hodnota v podstatě odpovídá mediánu
- Takto asymetrická data je možné převést logaritmickou transformací na normální rozložení



Geometrický průměr



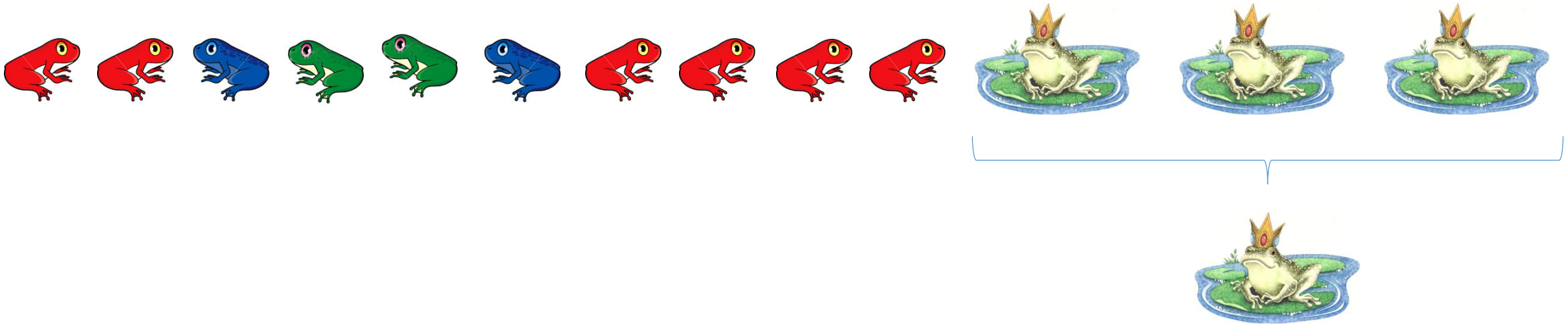
EXP (Y) = Geometrický průměr X

$$\overline{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}$$

$\overline{Y}$ ± Standardní chyba

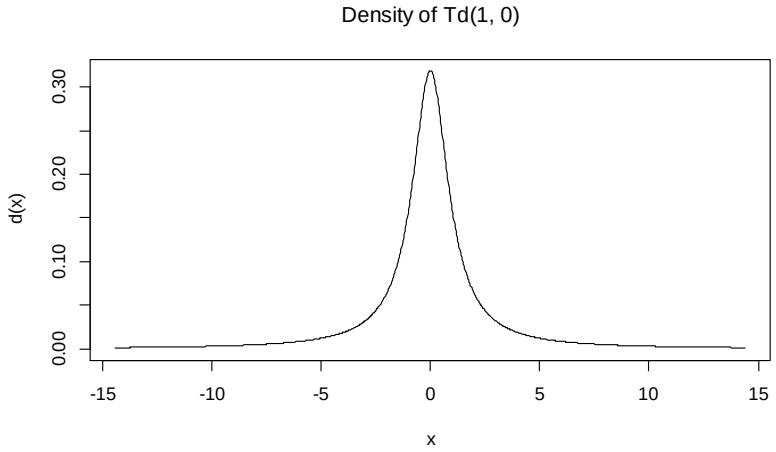
Stupně volnosti

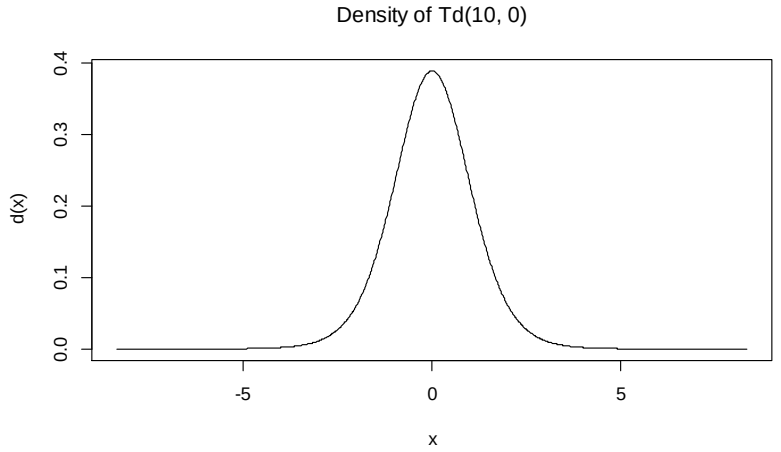
- Nezávislé jednotky informace
- Spjaty s počtem objektů, popřípadě skupin v datech
- Klesají s výpočtem každé souhrnné statistiky (=odečítáme od celkového počtu vzniklé závislé statistiky)

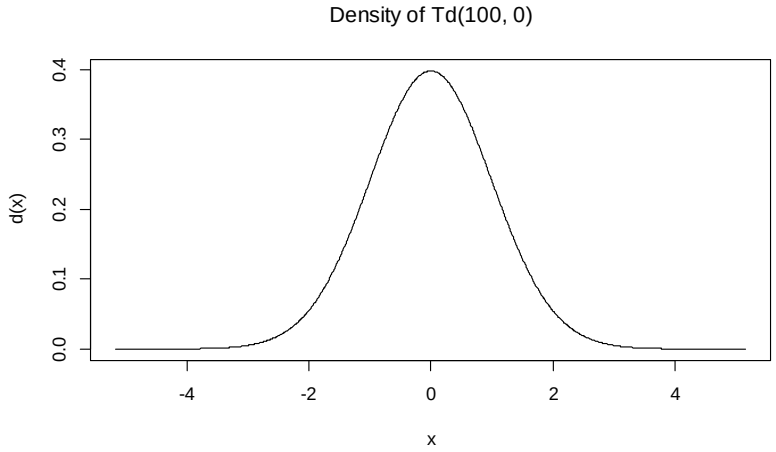


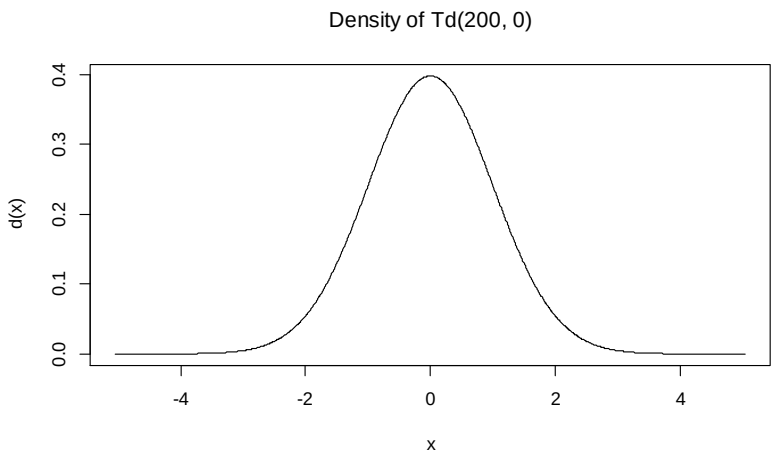
Studentovo rozdělení

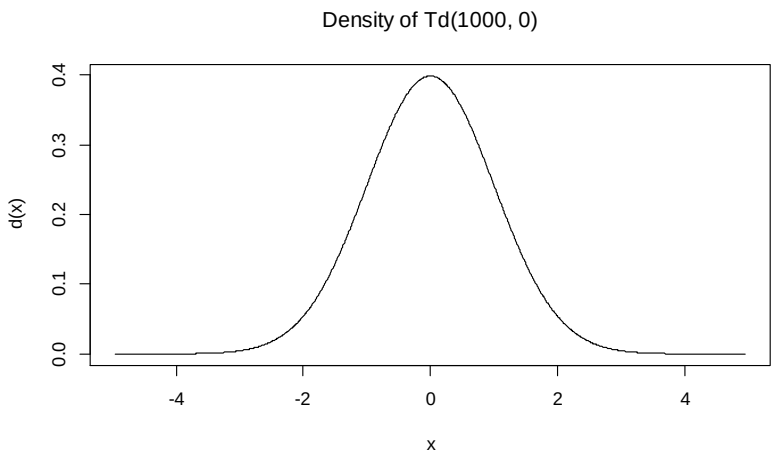
- Pro reálnější popis reality než umožňuje normální rozdělení
- Stupně volnosti – ve vazbě na velikost vzorku







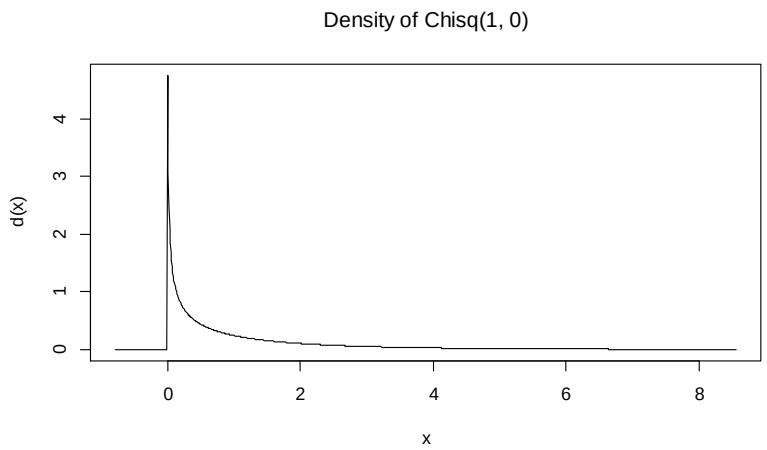


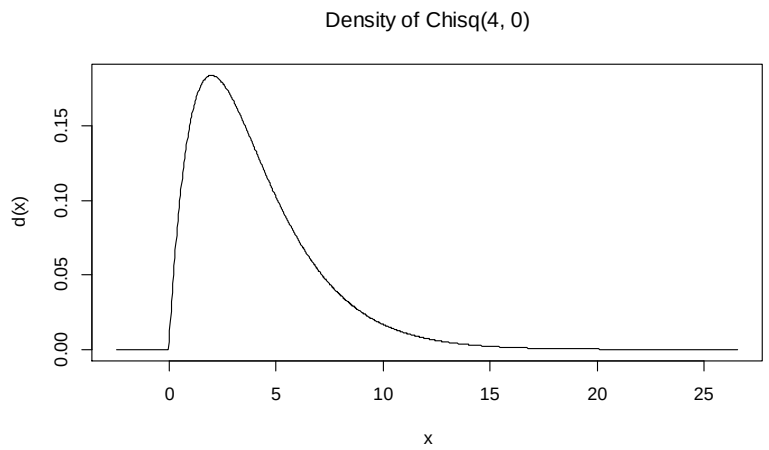


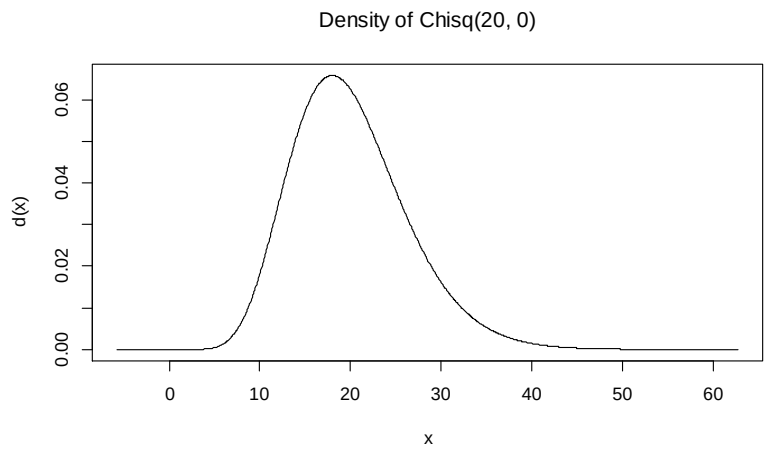
William Sealy Gosset
Publikace pod pseudonymem Student
t rozdělení na základě experimentů s kvasinkami

Pearsonovo (Chi-kvadrát) rozdělení

- Pro data, která nemohou být principiálně nikdy záporná
- Tvar ovlivněn stupni volnosti
- Očekávané a pozorované počty, rozptyly
- Často v genetice

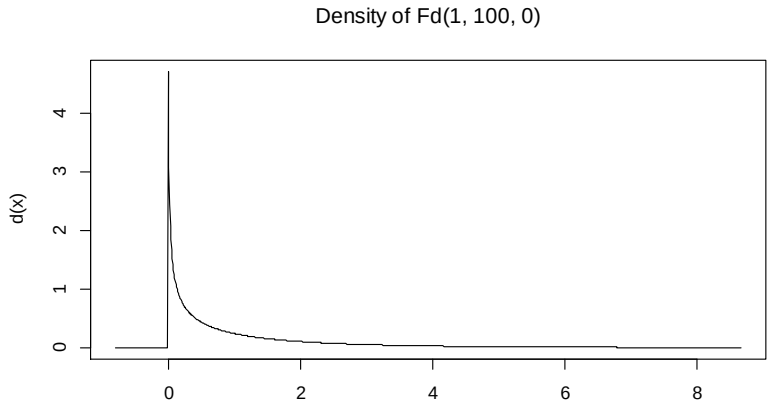


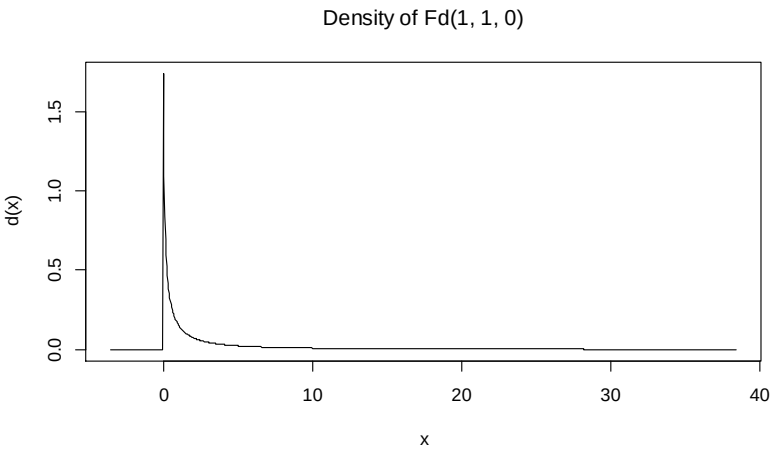


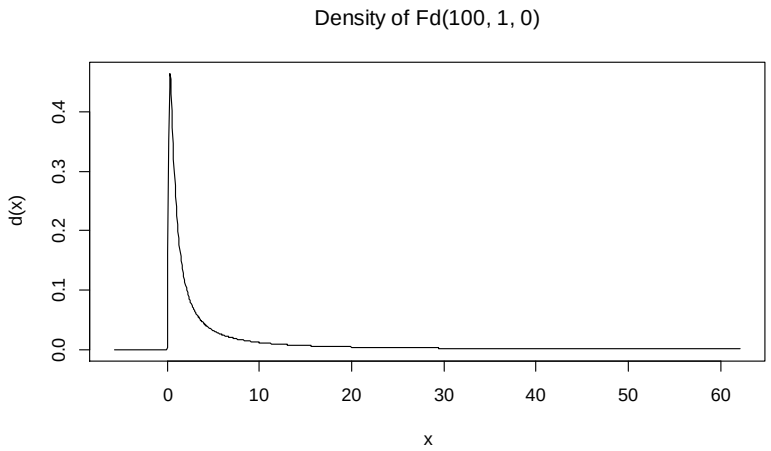


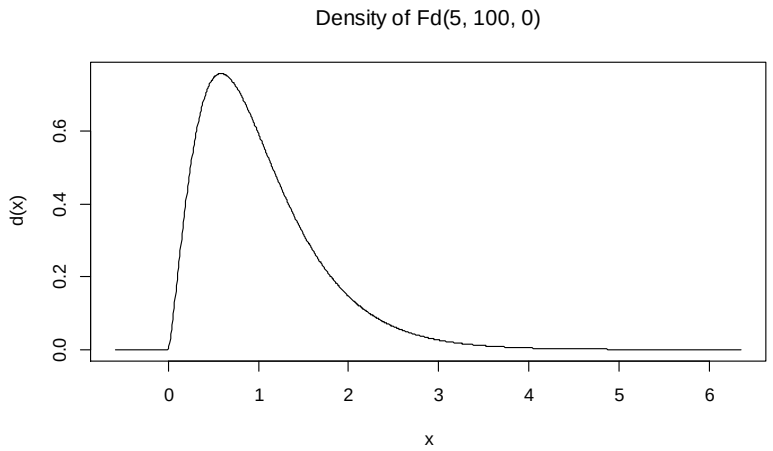
Fisher-Snedecorovo rozdělení

- Pro data, která nemohou být principiálně nikdy záporná
- Typicky poměr dvou rozptylů – využití v řadě, zejména pokročilejších statistických testů
- Dva různé stupně volnosti









Transformace dat - legitimní úprava rozložení

- **Základní typy transformací vedou k normalitě rozložení nebo k homogenitě rozptylu**
- **<u>Logaritmická transformace</u>**
- Logaritmická transformace je velmi vhodná pro data s odlehlými hodnotami na horní hranici rozsahu. Při porovnání průměrů u více souborů dat je pro tuto transformaci indikující situace, kdy se s rostoucím průměrem mění proporcionálně i směrodatná odchylka, a tedy jednotlivé proměnné mají stejný koeficient variance, ačkoli mají různý průměr.
- Za takovéto situace přináší logaritmická transformace nejen zeslabení asymetrie původního rozložení, ale také vyšší homogenitu rozptylu proměnných. Pro transformaci se nejčastěji používá přirozený logaritmus a pokud jsou v původním souboru dat nulové hodnoty, je vhodné použít operaci $Y = \ln (X+1)$.
- Je-li průměr logaritmovaných dat (tedy průměrný logaritmus) zpětně transformován do původních hodnot, výsledkem není aritmetický, ale geometrický průměr původních dat.

Transformace dat - legitimní úprava rozložení

- **Základní typy transformací vedou k normalitě rozložení nebo k homogenitě rozptylu**
- **<u>Odmocninová transformace</u>**

- Transformace je vhodná pro proměnné mající Poissonovo rozložení, tedy proměnné vyjadřující celkový počet nastání určitého jevu (spíše vzácného) v n nezávisle opakovaných pokusech. Obecněji lze tento typ transformace doporučit v případě normalizace dat typu počtu jedinců (buněk, apod.). Jde o transformaci:

- $Y = \sqrt{x}$ nebo $Y = \sqrt{x+1}$ nebo $Y = \sqrt{x} + \sqrt{x+1}$

- Transformace s přičtenou hodnotou 1 jsou efektivní, pokud X nabývá velmi malých nebo nulových hodnot. Situace indikující vhodnost odmocninové transformace je také proporcionalita výběrového rozptylu a průměru, tedy obecně jestliže s2x = k (výběrový průměr).

Transformace dat - legitimní úprava rozložení

- **Základní typy transformací vedou k normalitě rozložení nebo k homogenitě rozptylu**
- **Arcsin transformace**
- Tzv. úhlová transformace - velmi vhodná pro data typu podílů výskytu určitého jevu (znaku) mezi n hodnocenými jedinci - tedy pro data mající binomické rozložení. Pokud se určitý znak vyskytuje r-krát mezi n možnostmi (jedinci, opakováními), pak lze vyjádřit relativní četnost jeho výskytu jako p = r/n s variabilitou p.(1-p)/n. Arcsin transformace odstraní ze souborů dat podíly blízké 0 nebo 1, a tak efektivně sníží variabilitu odhadů středu. Transformace však není schopná odstranit variabilitu vyvolanou rozdílným počtem opakování v jednotlivých variantách - v takovém případě lze doporučit provedení vážených transformací dat. Velmi častou formou této transformace je:

$$Y = \arcsin \sqrt{p}$$

- - tedy transformace podílů do hodnot, jejichž sinus je roven druhé odmocnině původních hodnot. Pokud celkový počet jedinců (opakování), mezi kterými je výskyt znaku monitorován, je n < 50, pak lze doporučit velmi efektivní empirická opatření pro transformaci podílů blízkých 0 nebo 1. Pro tento případ lze nahrazovat nulové podíly hodnotou 1/4n a 100 % podíly hodnotou (n-1/4)/n. Pokud se mezi hodnotami vyskytuje větší množství krajních hodnot (menší než 0,2 a větší než 0,8), lze doporučit transformaci:

$$Y = \frac{1}{2}\left[\arcsin \sqrt{\frac{x}{n+1}} + \arcsin \sqrt{\frac{x+1}{n+1}}\right]$$