# Modelová rozložení

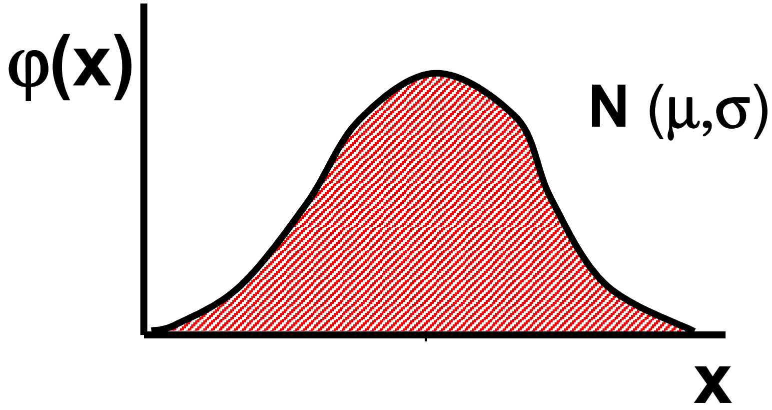
## Normální rozložení jako statistický model

## Aplikace modelových rozložení

## Přehled modelových rozložení

Vytvořil Institut biostatistiky a analýz, Masarykova univerzita
*J. Jarkovský, L. Dušek*

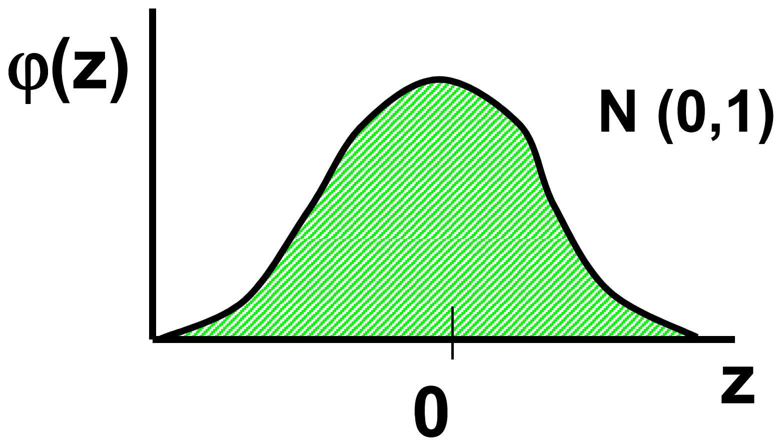
# Rozložení hodnot jako model: Normální rozložení




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

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

**Standardizovaná forma**

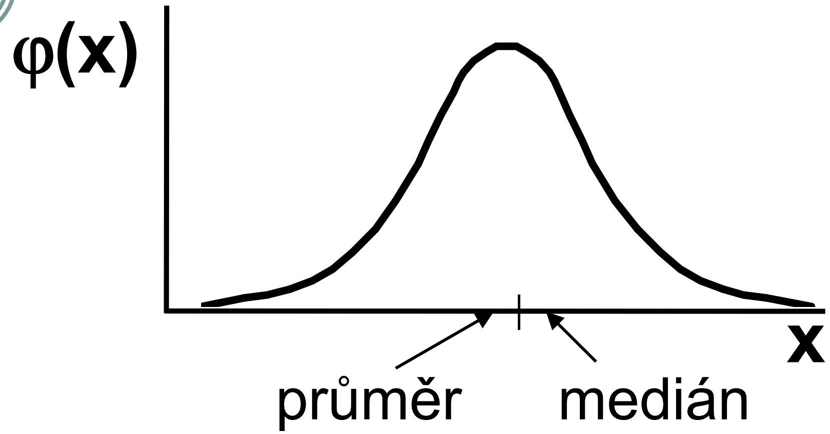



$$\varphi(z) = \frac{1}{\sqrt{2.\pi}}.e^{-\frac{z^2}{2}}$$

**Tabelovaná podoba**

# Parametry charakterizující normální rozložení a jejich význam

$$E(x) \sim \bar{x} \sim \mu$$
$$D(x) \sim s^2 \sim \sigma^2$$




a)

$$\mu \sim \bar{x}$$

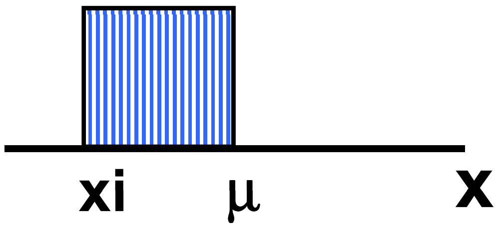
**průměr - ukazatel středu**

b)

$$\sigma^2 \sim s^2$$

**rozptyl**

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$




c)

$$\sigma \sim s$$

**směrodatná odchylka**

$$s = \sqrt{s^2}$$

**Pravidlo ± 3s**

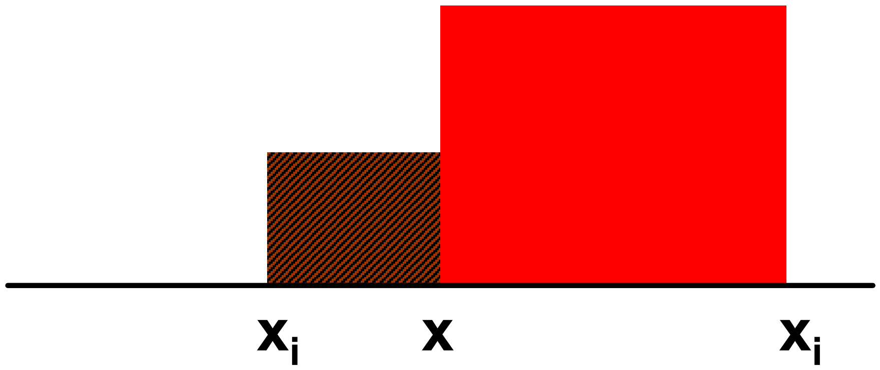
d)

**koeficient variance**

$$c = s / \bar{x}$$

# Rozptyl není univerzálním ukazatelem variability

$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n - 1}$$





$\Rightarrow$ neúměrně zvýší $s^2$

# Normální rozložení jako model

## *I. Použitelnost modelu*

**A) X: spojitý znak - hmotnost jedince (myši)**

**1,2; 1,4; 1,6; 1,8; 2,0; 2,4; 3.8**

n = 7 opakování

**medián** = 1,8

**průměr** = $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{7}\sum_{i=1}^{7} x_i = \frac{1}{7}(1{,}2+1{,}4+1{,}6+1{,}8+2{,}0+2{,}4+3{,}8) = \frac{1}{7}14{,}2 = 2{,}03$

**rozptyl ($s^2$) =** $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{7}(x_i-2{,}03)^2}{6} = 0{,}766$

**sm. odchylka (s) =** $\sqrt{s^2} = \sqrt{0{,}766} = 0{,}875$

**?** **Je předpoklad normálního rozložení oprávněný ?**
**Jaký předpokládáte možný rozsah hodnot tohoto znaku ?** **?**

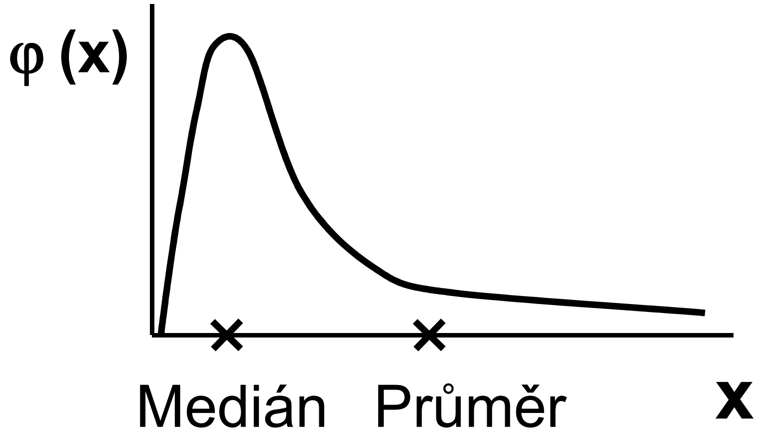
# Stručný přehled modelových rozložení I.

| Rozložení | Parametry | Stručný popis |
| --- | --- | --- |
| Normální | **Průměr ($\mu$)**<br>**Rozptyl ($\sigma^2$)** | Symetrická funkce popisující intervalovou hustotu četnosti; nejpravděpodobnější jsou průměrné hodnoty znaku v populaci. |
| Log-normální | **Medián**<br>**Geometrický průměr**<br>**Rozptyl ($\sigma^2$)** | Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení. |
| Weibullovo | **$\alpha$ - parametr tvaru**<br>**$\beta$ - 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**<br>**Geometrický průměr**<br>**Rozptyl ($\sigma^2$)** | Funkce intervalové hustoty četnosti, která po logaritmické transformaci nabude tvaru normálního rozložení. |
| Triangulární | **$f(x) = [b - ABS (x - a)] / b^2$**<br>**$a - b < x < a + b$** | Pravděpodobnostní funkce pro typ rozložení, kdy jsou střední hodnoty výrazně pravděpodobnější než hodnoty okrajové. |
| Gamma | **Parametry distribuční funkce:**<br>**$\alpha$ - parametr tvaru**<br>**$\beta$ - parametr rozsahu hodnot** | Umožňuje flexibilně modelování distribučních funkcí nejrůznějších tvarů. Např. $\chi^2$ rozložení je rozložení typu Gamma. Gamma rozložení s a = 1 je známo jako exponenciální rozložení. |

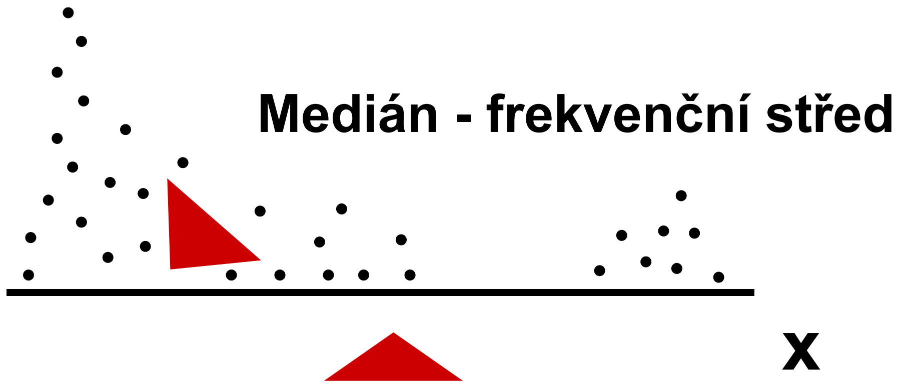
# Stručný přehled modelových rozložení II.

| Rozložení | Parametry | Stručný popis |
|---|---|---|
| **Beta** | **Parametry distribuční funkce:**<br>**$\alpha$ - parametr tvaru**<br>**$\beta$ - 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**<br>**Průměr**<br>**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.<br>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. |

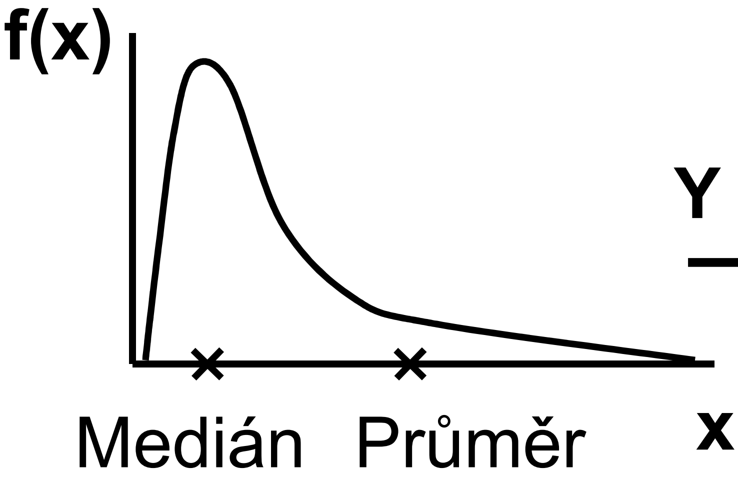
# Log-normální rozložení jako častý model reálných znaků



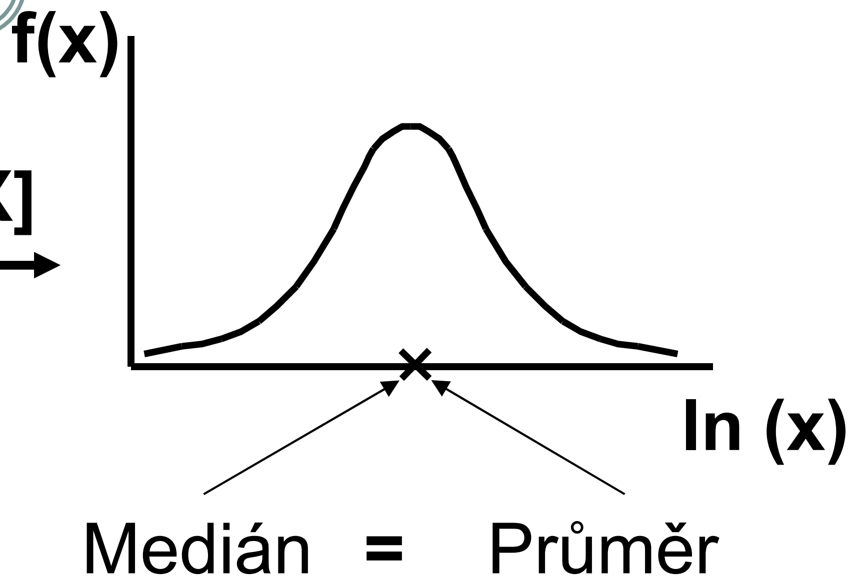

**U asymetrických rozložení je medián velmi vhodným alternativním ukazatelem středu**



# Log-normální rozložení lze jednoduše transformovat




Y = Ln [X]




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

**$\bar{Y}$ ± Standardní chyba**

**EXP (Y) = Geometrický průměr X**

# Parametry rozložení

- Soubor dat (řada čísel) můžeme charakterizovat parametry jeho rozložení
- Hlavní skupiny těchto parametrů můžeme charakterizovat jako ukazatele:
  - Středu (medián, průměr, geometrický průměr)
  - Šířky rozložení (rozsah hodnot, rozptyl, směrodatná odchylka)
  - Tvaru rozložení (skewness, kurtosis)
  - Kvantily rozložení – kolik % řady dat leží nad a pod kvantilem