

# Analýza dat pro Neurovědy



RNDr. Eva Koritáková, Ph.D.  
doc. RNDr. Ladislav Dušek, Dr.

# Blok 2

Jak medicínská data správně  
testovat.

# Osnova

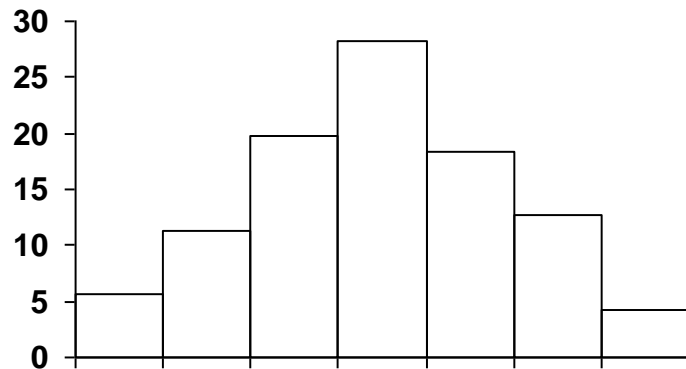
---

1. Vybraná modelová rozdělení
2. Transformace dat
3. Intervaly spolehlivosti
4. Formulování hypotéz nad medicínskými daty
5. Hladina významnosti a síla testu
6.  $p$ -hodnota

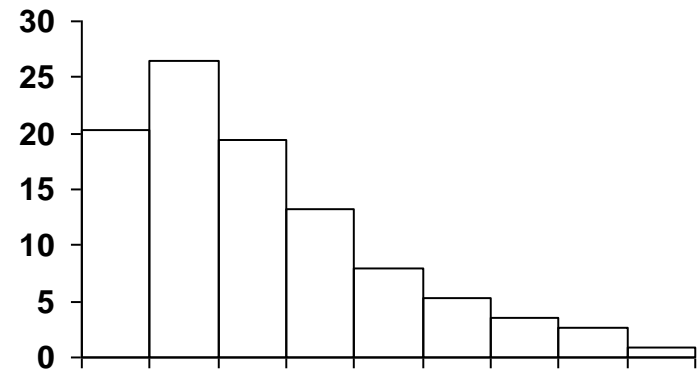
# 1. Vybraná modelová rozdělení

# Motivace

## Symetrická data



## Asymetrická data



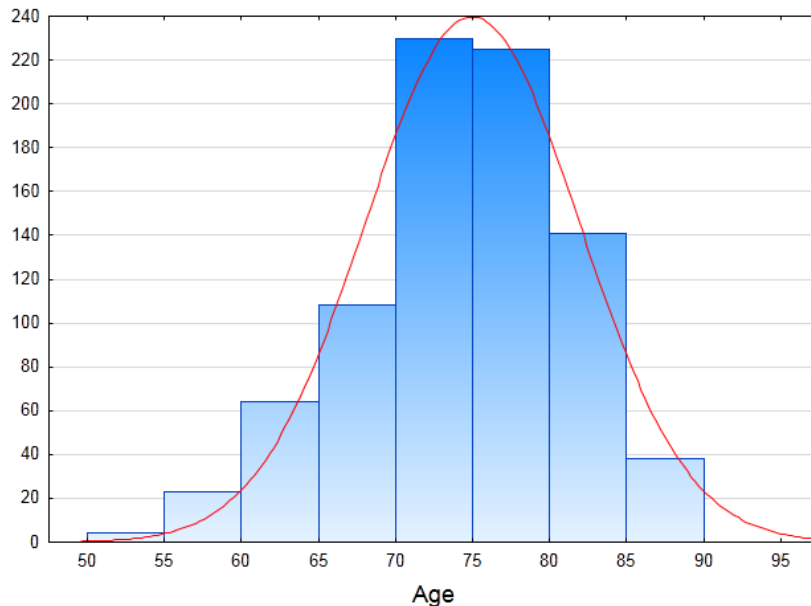
# K čemu je nám znalost o modelových rozděleních?

- **Popis vlastností cílové populace** – na základě pozorovaných dat (histogram, box plot, popisné statistiky) jsme schopni usuzovat na charakter rozdělení pravděpodobnosti sledované veličiny. Dokonce jsme schopni otestovat míru shody s teoretickým rozdělením.
- **Srovnání vlastností cílové populace/populací** – na základě pozorovaných dat a našich předpokladů o teoretickém modelu (hypotéz) jsme schopni pomocí statistických testů srovnávat vlastnosti jedné nebo více cílových populací.
- **Predikce vlastností cílové populace** – nevyvrátíme-li na základě pozorovaných dat platnost teoretického modelu, jsme schopni se ptát, jak a s jakou pravděpodobností se bude cílová populace v budoucnu chovat.

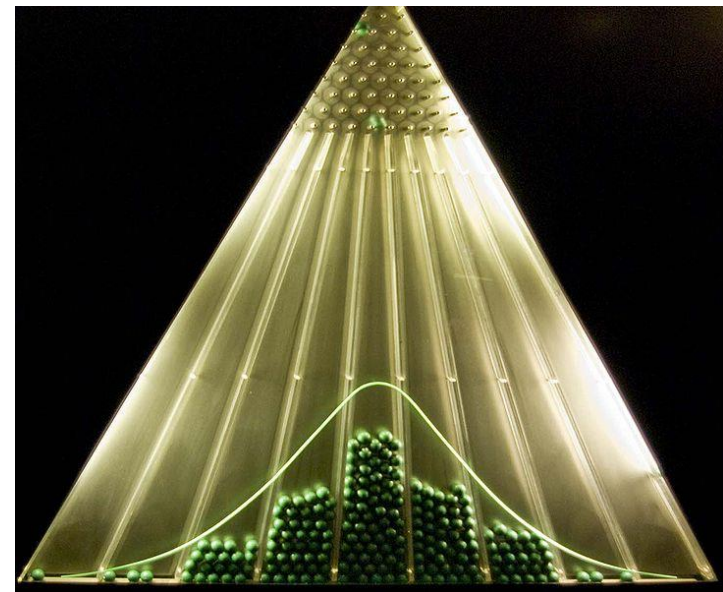
# Normální rozdělení

- jiný název – Gaussovo rozdělení
- základní rozdělení – u mnoha klinických a biologických veličin: tělesná výška, délka končetin a kostí, krevní tlak,...
- hodnoty veličiny se symetricky shlukují kolem středu, variabilita je dána aditivním vlivem mnoha „slabě působících faktorů“

**Příklad - věk**



**Příklad vzniku normálního rozdělení – Galtonova deska**

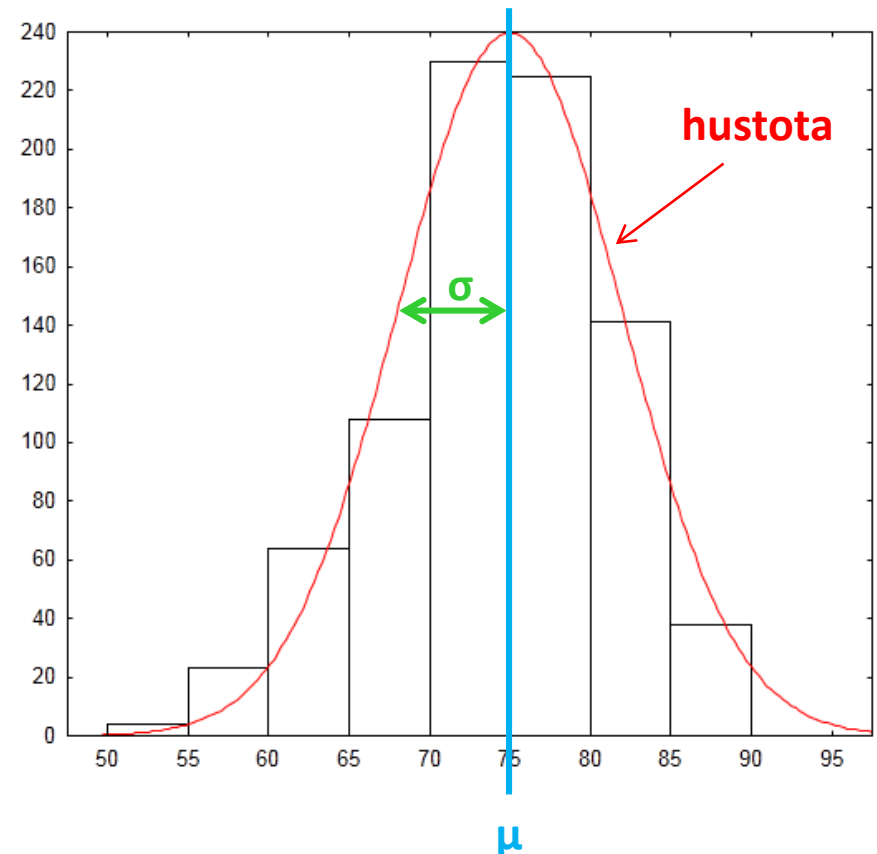


# Normální rozdělení

- **střední hodnota** – sumární statistika středu dat (tzn. číslo, které zastoupí střední, typickou, průměrnou hodnotu)
  - u normálního rozd. označení:  $\mu$
- **rozptyl** – sumarizace variability (tzn. odlišnosti jedinců zahrnutých ve výběrovém souboru);
  - u normálního rozd. označení:  $\sigma^2$
- tvar rozdělení nám popisuje **hustota** (hustota normálního rozdělení – tzv. Gaussova křivka):

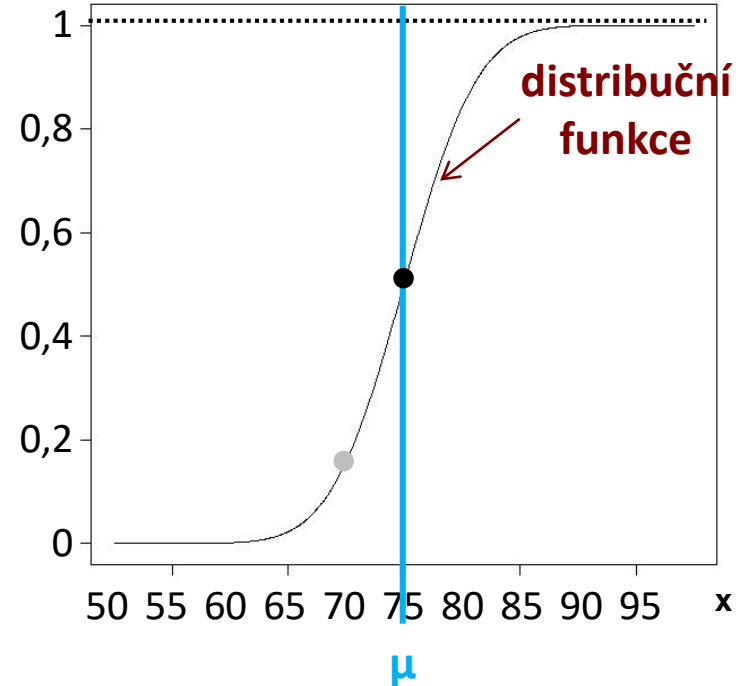
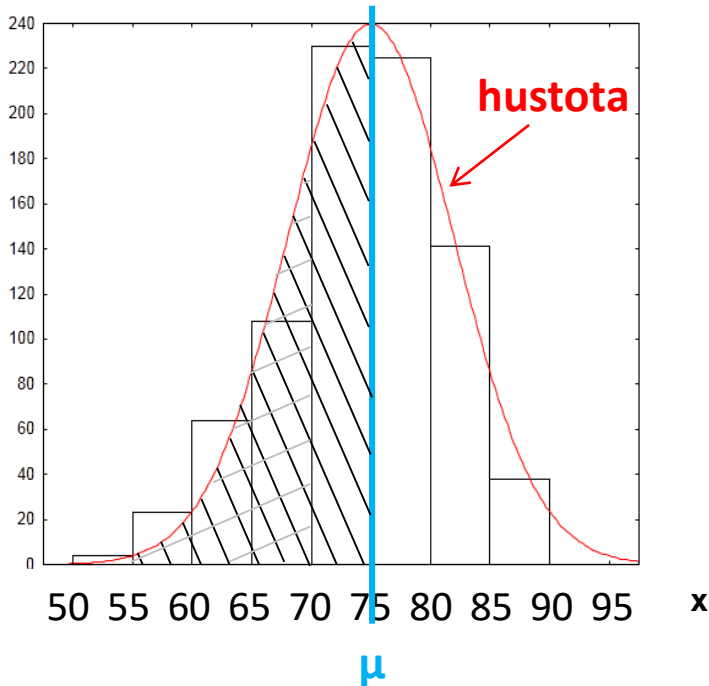
$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

- značení:  $N(\mu, \sigma^2)$





# Normální rozdělení – distribuční funkce



interval	$d(I)$	$n(I)$	$n(I)/n$	$N(x'')$	$F(x'')$
<50,55)	5	4	0,005	4	0,005
<55,60)	5	23	0,028	27	0,033
<60,65)	5	64	0,077	91	0,110
...					

$d(I)$  – šířka intervalu

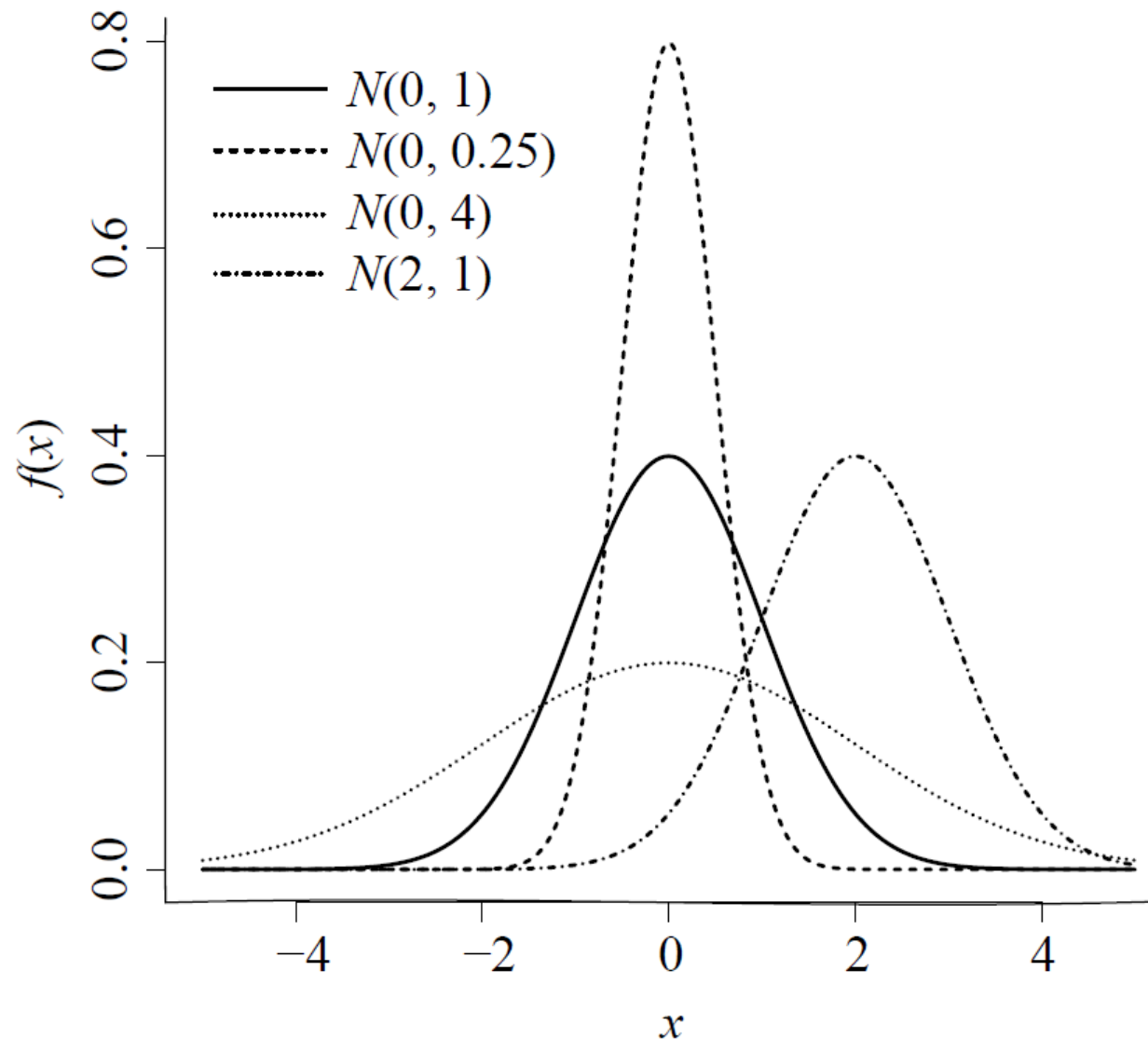
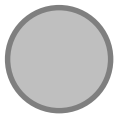
$n(I)$  – absolutní četnost

$n(I) / n$  – intervalová relativní četnost

$N(x'')$  – intervalová kumulativní četnost do horní hranice  $x''$

**$F(x'')$  – intervalová relativní kumulativní četnost do horní hranice  $X''$**

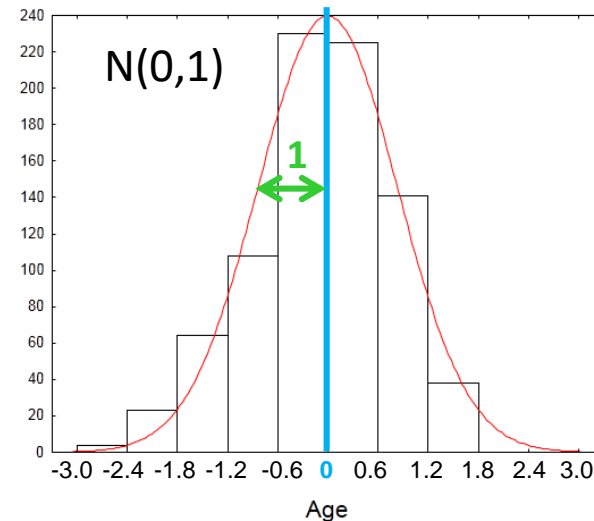
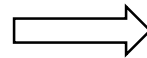
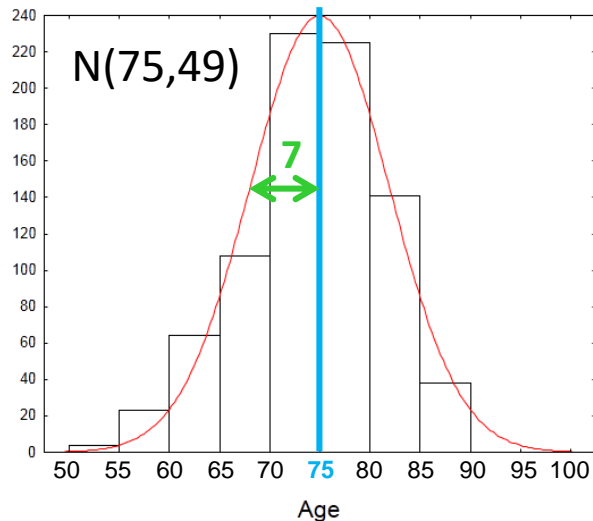
# Normální rozdělení – různé $\mu$ a $\sigma^2$



# Standardizované normální rozdělení

- Jakékoliv normální rozdělení může být převedeno na tzv. standardizované normální rozdělení:

$$X \sim N(\mu, \sigma^2) \rightarrow Z = \frac{X - \mu}{\sqrt{\sigma^2}} \sim N(0,1)$$

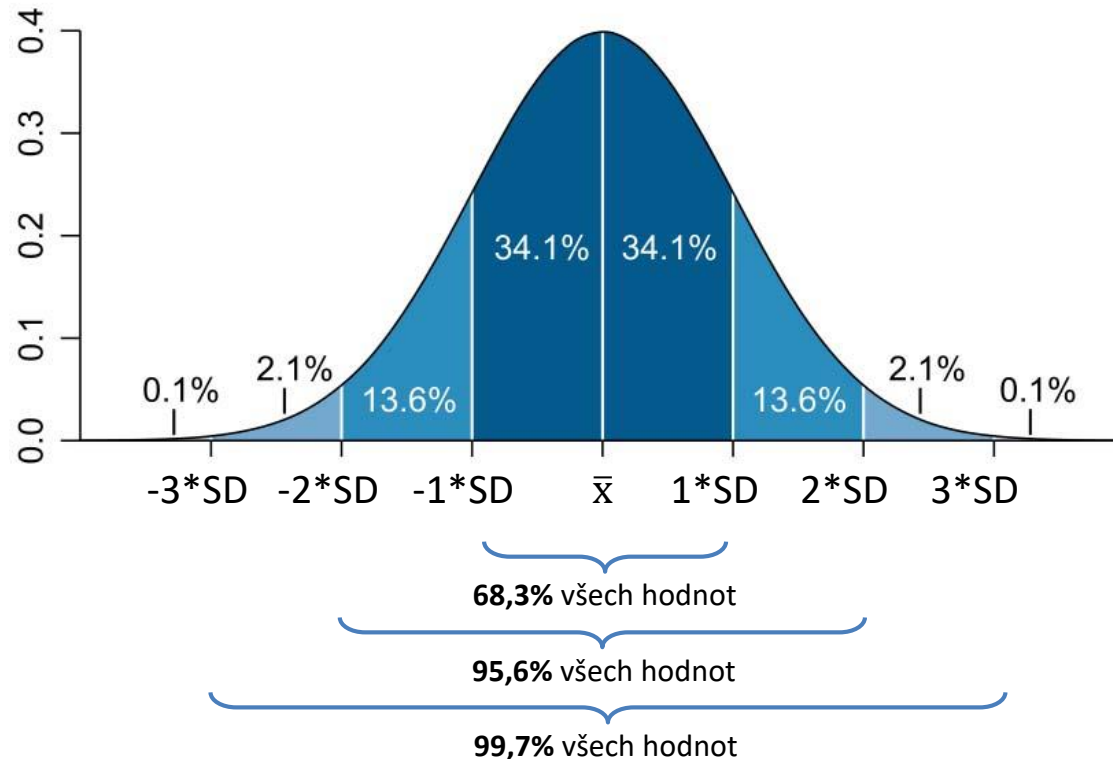


→ střední hodnota rovna 0, rozptyl roven 1

- Hustota pravděpodobnosti:  $f(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
- Klíčové rozdělení řady testů.**
- Výhoda je, že všechny hodnoty distribuční i kvantilové funkce jsou tabelovány a obsaženy ve všech dostupných softwarech.

# Normální rozdělení – pravidlo $\pm 3$ SD (sigma)

- U normálního rozdělení lze vyčíslit procento hodnot, které by se měly vyskytovat v rozmezí  $\pm k$  násobku směrodatné odchylky ( $SD=\sigma$ ) od průměru.
- Lze říci, že v rozmezí průměr  $\pm 3*SD$  by se mělo vyskytovat přes 99,5% všech hodnot.

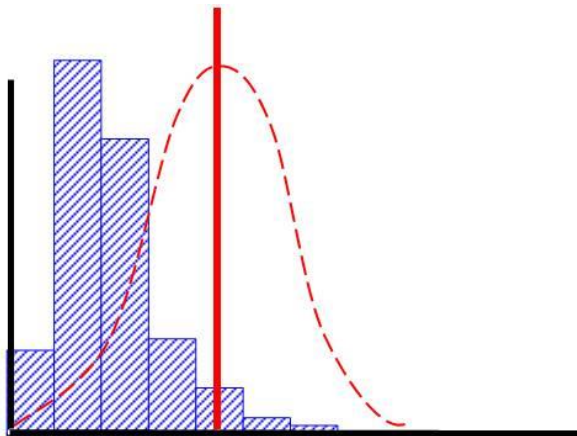


- Použití: orientační ověření normality dat, identifikace odlehlých hodnot

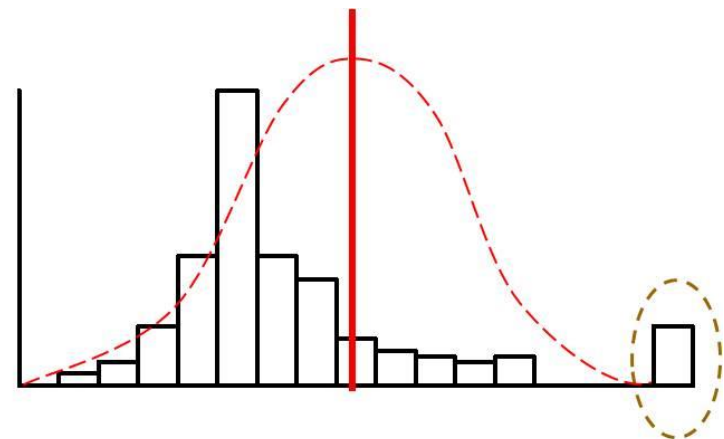
# Normalita dat

- Normalita je klíčovým předpokladem řady statistických metod – zejména testů a modelů.
- Není-li splněna podmínka normality hodnot, je špatně celý model, se kterým daná metoda pracuje, což vede k neinterpretovatelným závěrům.
- Její ověření je tak stejně důležité jako výběr správného testu.
- Pro ověření normality existuje řada testů a grafických metod.

Rozdělení není normální

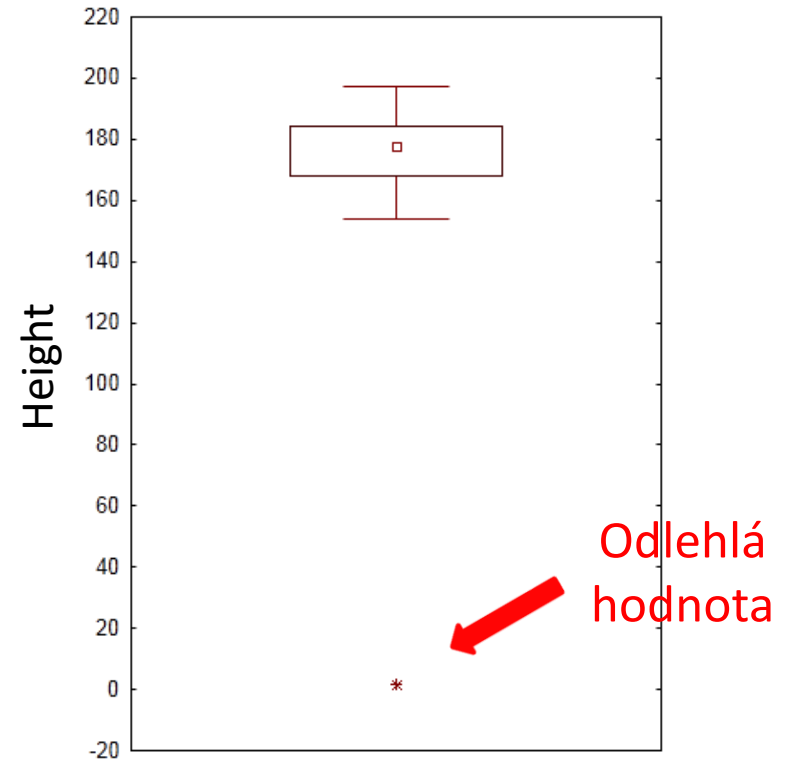
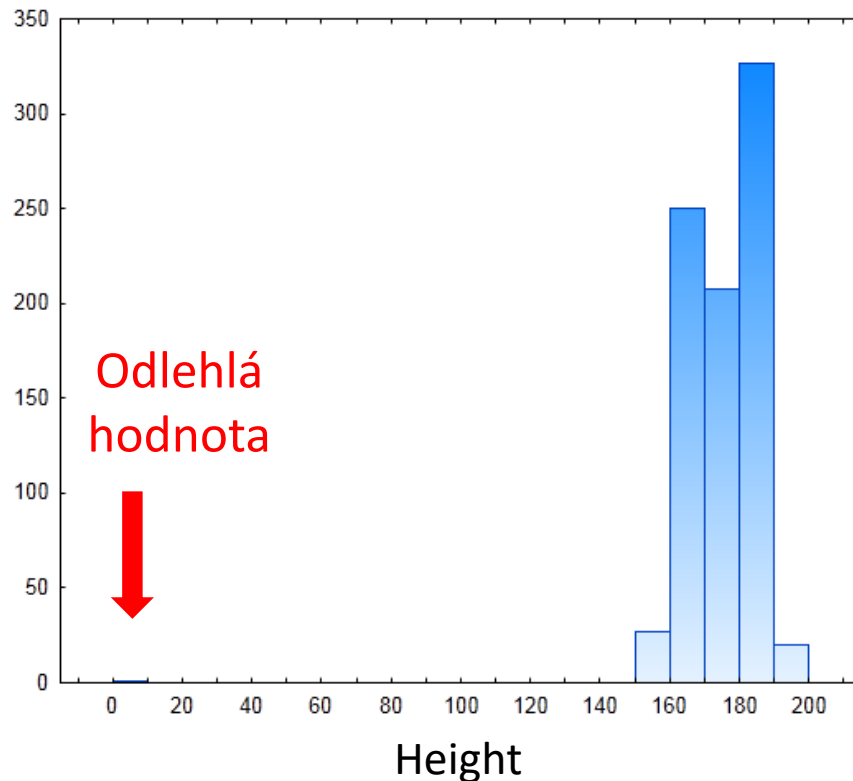


Odlehlá hodnota



# Odlehlá hodnota

- Netypické pozorování
- Závisí však na naší znalosti dané problematiky, jestli je daná hodnota možná či nikoliv!
- Grafická identifikace: pomocí histogramu a krabicového grafu



# Odlehlá hodnota

- Identifikace pomocí popisných statistik: srovnání mediánu a průměru a pomocí směrodatné odchylky

	Count	Mean	Median	Minimum	Maximum	Standard Deviation
Height	833	175,97	178,00	1,58	197,00	10,97
Height_cor	833	176	178	154	197	9

	Count	Mean	Median	Minimum	Maximum	Standard Deviation
Height	20	166,30	174,00	1,58	193,00	39,60
Height_cor	20	174,2	174,0	158,0	193,0	8,9

- U velkého datového souboru bude průměr méně ovlivněn odlehlou hodnotou, z popisných statistik nemusíme poznat, že by tam mohla být odlehlá hodnota -> vždy provádět vizualizaci dat!

# Úkol 1

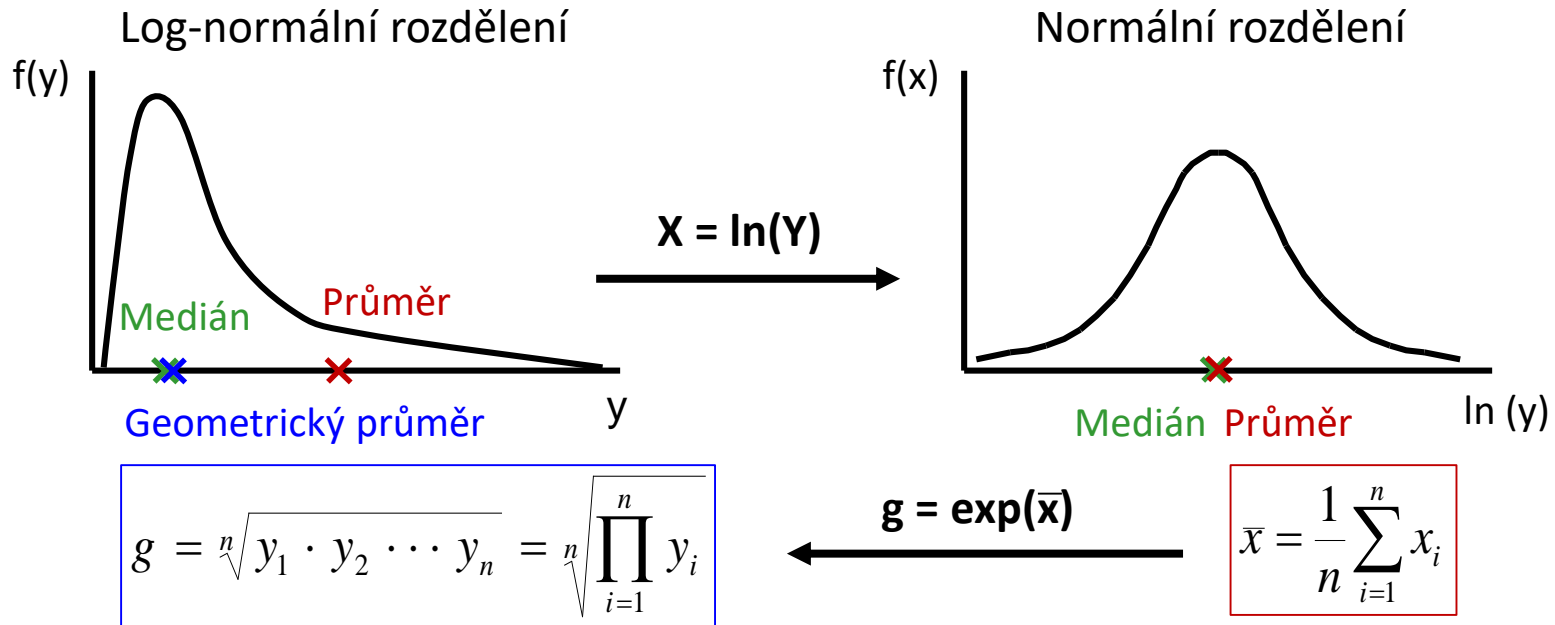
---

- Zjistěte, zda má MMSE skóre normální rozdělení – použijte histogram, krabicový graf a popisnou statistiku.



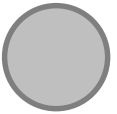
# Logaritmicko-normální rozdělení

- u zešikmeného rozdělení nám často (ale ne vždy!) může pomoci proměnnou transformovat pomocí **logaritmické transformace**:  $X = \ln(Y)$  nebo  $X = \ln(Y+1)$ , pokud data obsahují 0



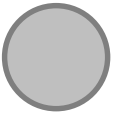
- můžeme použít přirozený logaritmus ( $\ln$ ), dvojkový logaritmus ( $\log_2$ ) nebo dekadický logaritmus ( $\log_{10}$ )
- Příklady veličin s log-normálním rozdělením: tělesná hmotnost, délka inkubační doby infekčního onemocnění, řada krevních parametrů (např. počet krevních buněk v daném objemu krve, sérový bilirubin u pacientů s cirhózou), počet bakteriálních buněk v daném objemu,...

# Stručný přehled rozdělení I.



Rozdělení	Parametry	Popis	Graf
<b>Normální</b> $N(\mu, \sigma^2)$	Průměr Rozptyl	Praktická významnost, spojitě. $EX = \mu, DX = \sigma^2$ Př. délkové rozměry těla	
<b>Log-normální</b> $lnN(\mu, \sigma^2)$	Geometrický průměr Rozptyl	Praktická významnost, spojitě. $EX = e^{\mu + \sigma^2 / 2}, DX = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$ Př. objemové rozměry, hmotnost	
<b>Studentovo t</b> $t(k)$	Stupně volnosti (uvažuje velikost vzorku) Průměr, Rozptyl	Teoretická významnost, spojitě. Aproximace normálního rozd. pro malé soubory, pro větší soubory ( $n > 100$ ) se limitně blíží normálnímu rozd. Teoretický základ $t$ testu.	
<b>Chí-kvadrát</b> $\chi^2(k)$	Stupně volnosti (uvažuje velikost vzorku)	Teoretická významnost, spojitě. Porovnávání četností jevů ve 2 a více kategoriích, výpočet intervalu spolehlivosti pro rozptyl.	

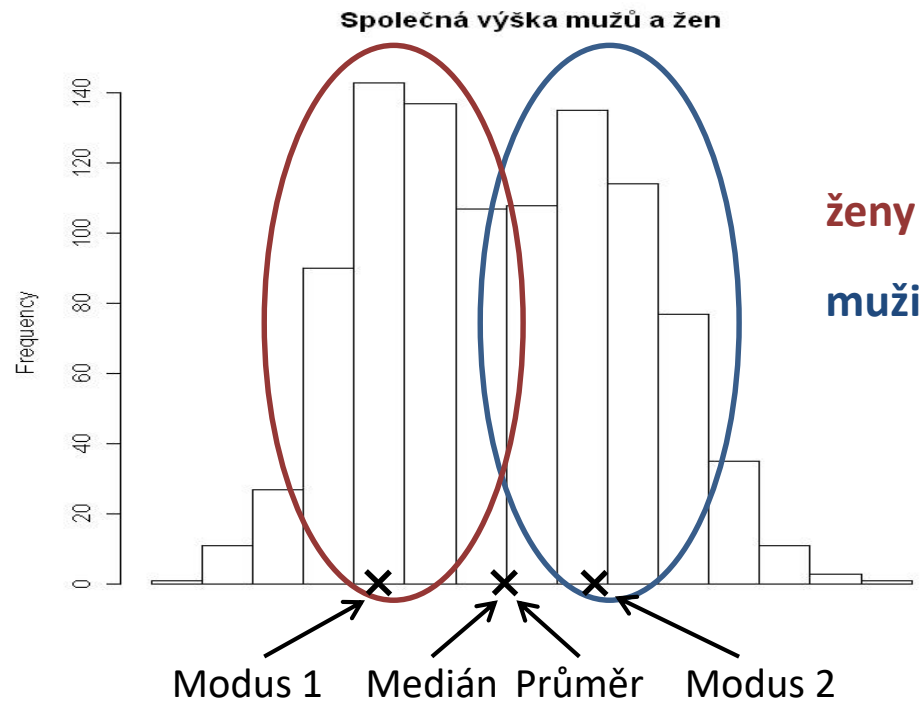
# Stručný přehled rozdělení II.



Rozdělení	Parametry	Popis	Graf
<b>Fisherovo F</b> $F(k_1, k_2)$	Dvojitě stupně volnosti (uvažuje velikost dvou vzorků)	Teoretická významnost, spojitě. Základ ANOVA testu a F-testu, výpočet intervalu spolehlivosti pro podíl rozptylů.	
<b>Exponenciální</b> $Exp(\lambda)$	Průměr Rozptyl	Praktická významnost, spojitě. $EX=1/\lambda$ , $DX=1/\lambda^2$ Popisuje dobu mezi událostmi, význam v analýze přežití, zobecněním je Weibullovo a Gamma rozdělení. Př. doba od diagnózy do úmrtí	
<b>Binomické</b> $Bi(n, \pi)$	Průměr Rozptyl	Praktická významnost, diskrétní. $EX=n\pi$ , $DX=n\pi(1-\pi)$ Popisuje počet výskytů sledované události v $n$ nezávislých pokusech. Př. výskyt nežádoucích účinků léků.	
<b>Poissonovo</b> $Po(\lambda)$	Průměr Rozptyl	Praktická významnost, diskrétní. $EX=\lambda$ , $DX=\lambda$ Popisuje počet výskytů sledované události na danou jednotku času, plochy... Př. počet krvinek v poli mikroskopu.	

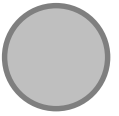
# Bimodální rozdělení

- Představuje většinou problém, neboť se zřejmě jedná o směs dvou souborů s unimodálním rozdělením.
- Bimodální rozdělení má např. tento tvar:

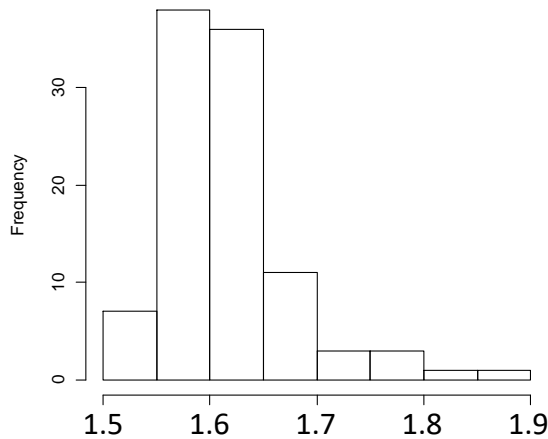


- Nutná další analýza: Co způsobuje bimodalitu? Umožňuje proměnná rozlišit kategorie lidí (např. pacienti od kontrol)? Je vzorek reprezentativní?

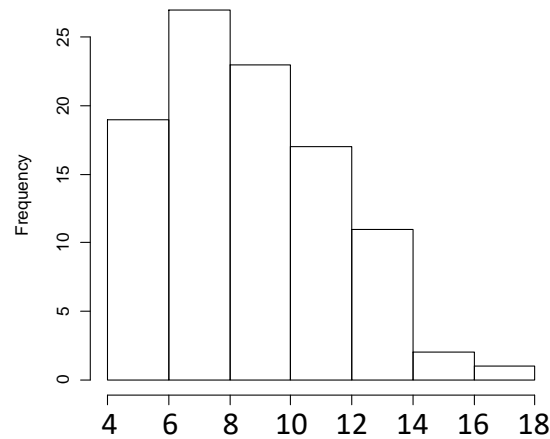
# Úkol 2 - Přiřadte k daným veličinám jejich název a typ rozdělení.



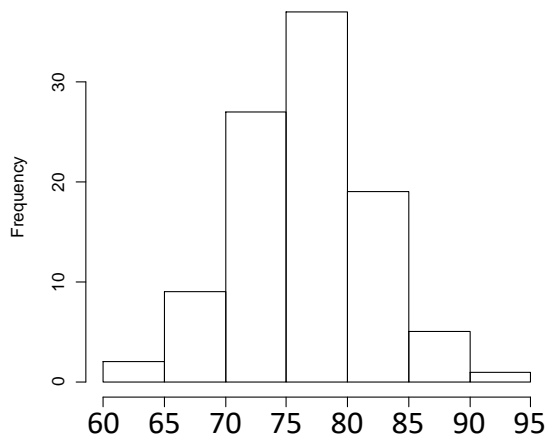
**X1:** 1.58 1.55 1.67 1.69 1.57



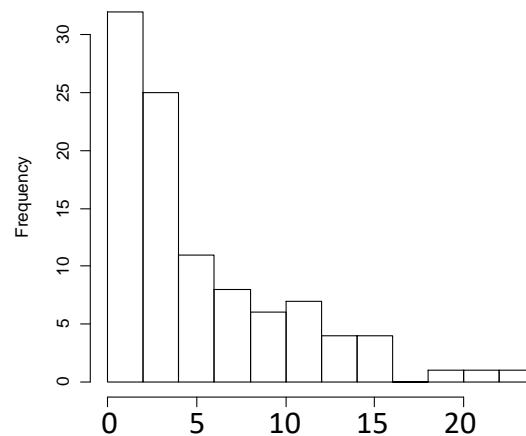
**X2:** 10 12 8 7 10



**X3:** 79.5 89.2 75.3 77.8 90.0



**X4:** 0.49 0.78 6.01 0.47 4.70



## Vybraná rozdělení:

- I. Normální rozdělení
- II. Logaritmicko-normální rozdělení
- III. Poissonovo rozdělení
- IV. Exponenciální rozdělení

## Veličiny:

- a) Doba od zahájení léčby do kompletní remise u pacienta s chronickou myeloidní leukémií (v letech)
- b) Plocha kůže člověka (v m<sup>2</sup>)
- c) Diastolický tlak (v mm Hg)
- d) Počet příjezdů sanitky do okresní nemocnice za hodinu

## 2. Transformace dat

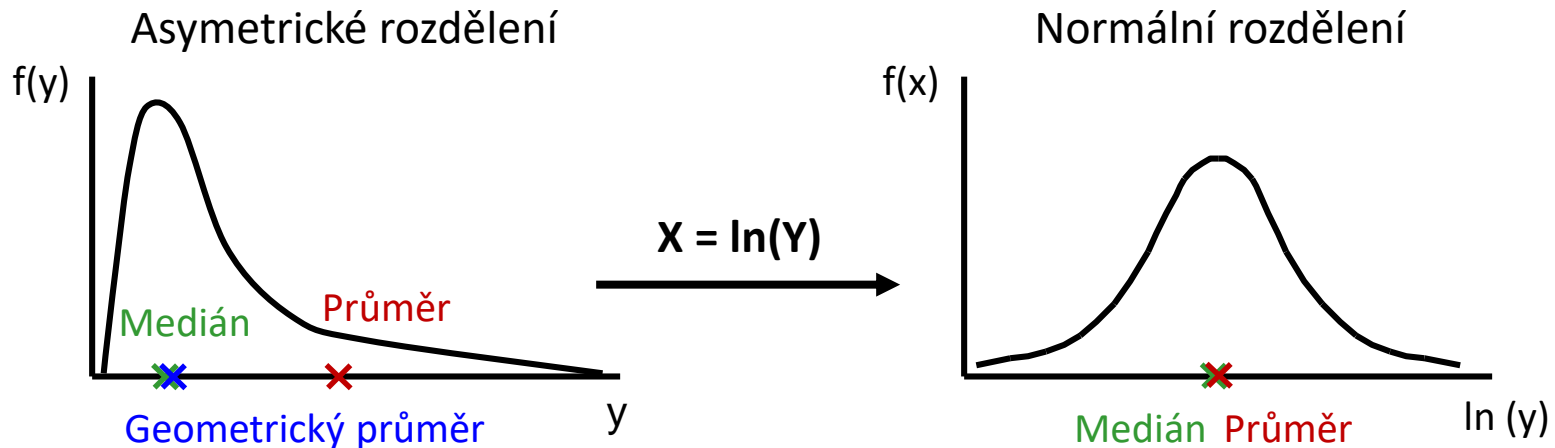
# Význam transformací

---

- Transformace umožní změnit rozsah hodnot proměnné, změnit typ rozložení apod.
- Hlavní cíle transformací:
  1. Normalizace dat – převod na normální rozdělení
  2. Standardizace dat – převod na standardizované normální rozdělení
  3. Centrování dat
  4. Lepší interpretace dat

# Normalizace dat

- Převod na normální rozdělení (normalita je předpokladem řady statistických testů).
- Např. **logaritmická transformace**:  $X = \ln(Y)$  nebo  $X = \ln(Y+1)$ , pokud data obsahují hodnotu 0

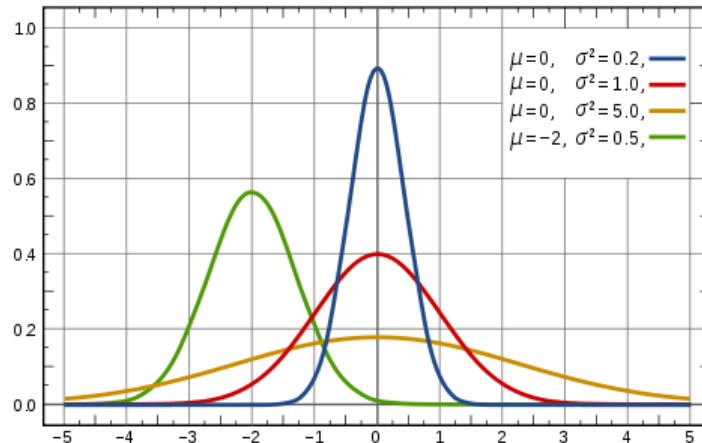


- Další příklady:
  - **odmocninová transf.** (pro proměnné s Poissonovým rozložením nebo obecně data typu počet jedinců, buněk apod.:  $X = \sqrt{Y}$  nebo  $X = \sqrt{Y + 1}$ )
  - **arcsin transformace** (pro proměnné s binomickým rozložením)
  - **Box-Coxova transformace**



# Standardizace dat

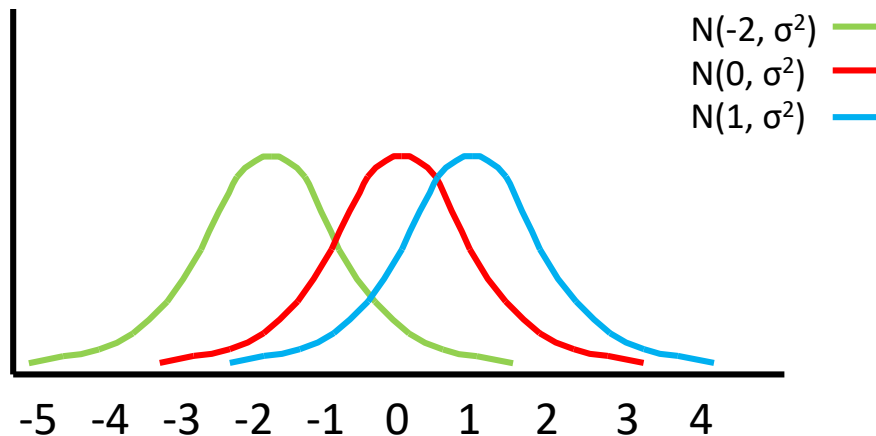
- Převod proměnné s normálním rozdělením na standardizované normální rozdělení:  $N(\mu, \sigma^2) \rightarrow N(0,1)$
- Důvod: řada statistických metod byla odvozena pro standardizované normální rozdělení,  $N(0,1)$ . Děláme to tedy opět kvůli lepší možnosti hodnocení dat.
- Standardizace:  $u_i = \frac{x_i - \bar{x}}{s}$
- Obrázek – standardizace je převod „modré“, „zelené“ a „okrové“ na „červenou“.



- z-skóre vlastně vyjadřuje, o kolik směrodatných odchylek se i-tá hodnota odchýlila od průměru.

# Centrování dat

- Odečtení průměru od dat – získáme novou proměnnou, která bude mít střední hodnotu rovnu nule:  $N(\mu, \sigma^2) \rightarrow N(0, \sigma^2)$
- Důvod: Centrování je důležitou podmínkou některých pokročilých statistických metod (např. klasifikačních).
- Centrování:  $u_i = x_i - \bar{x}$
- Obrázek – centrování je převod „modré“ a „zelené“ na „červenou“.



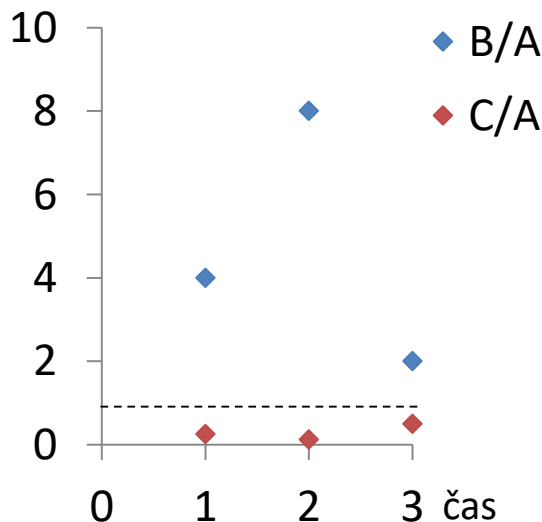
# Transformace kvůli lepší interpretaci dat

- Příklad: Microarray experiment se dvěma vzorky, měříme intenzitu exprese genu XY v jedné tkáni (hodnota intenzity  $A_{XY}$ ) a v druhé tkáni (hodnota intenzity  $B_{XY}$ ).
- Následně hodnoty převádíme na logaritmus se základem 2 jejich podílu:

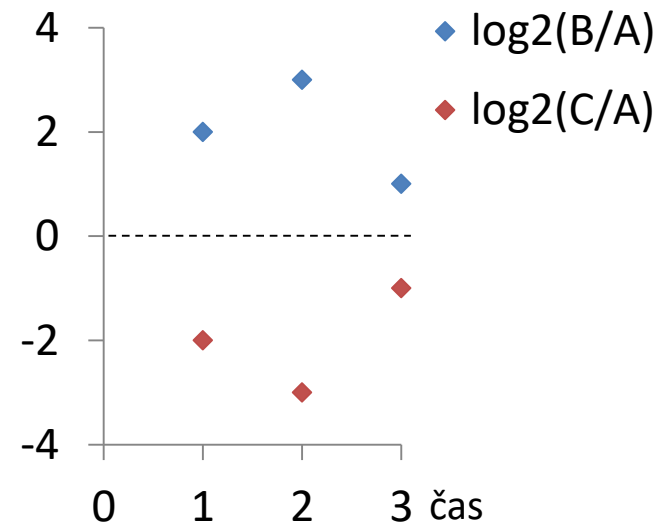
$$Z_{XY} = \log_2\left(\frac{B_{XY}}{A_{XY}}\right)$$

- Umožní nám to posoudit kolikrát byla exprese jednoho genu větší/menší než druhého genu (2x, 4x, 8x, 16x,...).

čas	B/A	C/A
1	4	1/4
2	8	1/8
3	2	1/2



$\log_2$



# Další příklady transformací – odvozené typy dat

- **Procento** (percentage) – sledujeme-li např. zlepšení v určitém parametru, je výhodné sledovat procentuální zlepšení. Příklad: ejekční frakce levé srdeční komory.
- **Podíl** (ratio) – mnoho indexů je odvozeno jako podíl dvou měřených veličin. Příklad: BMI
- **Pořadí** (rank) – místo absolutních hodnot známe někdy jen jejich pořadí. Jedná se sice o ztrátu určitého množství informace, nicméně i pořadí lze v analýze využít.
- **Skóre** (score) – jedná se o uměle vytvořené hodnoty charakterizující určitý stav, který nelze jednoduše měřit jako číselné hodnoty. Příklad: indexy kvality života.

# Kategorizace

- Vytvoření kvalitativní proměnné z kvantitativní proměnné.

Primární  
data

Age  
84  
76  
79  
89  
71  
70  
88  
86  
.  
.  
.  
.  
.  
.  
.  
.  
.  
n=833

Kategorizace



Frekvenční tabulka

	n(x)	N(x)	p(x)	F(x)
<60	23	23	2,8	2,8
60-69	126	149	15,1	17,9
70-79	467	616	56,1	73,9
>80	217	833	26,1	100,0

**x:** Kategorizovaný věk

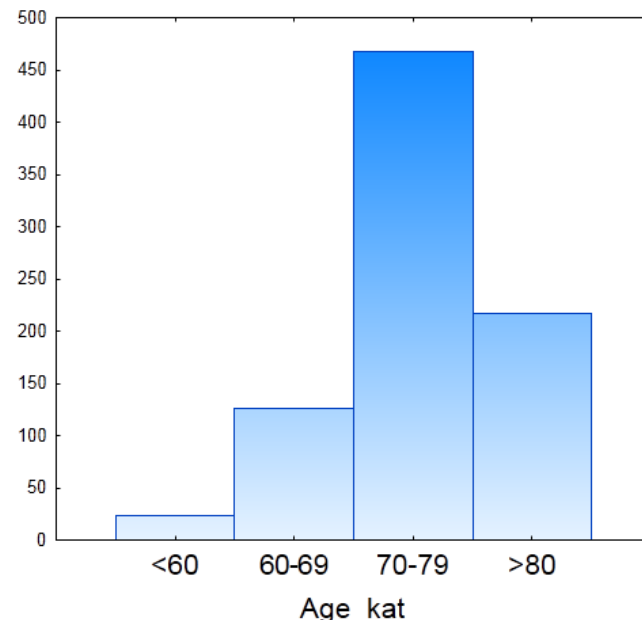
**n(x)** – absolutní četnost x

**N(x)** – kumulativní četnost hodnot nepřevyšujících x;  $N(x) = \sum_{t \leq x} n(t)$

**p(x)** – relativní četnost;  $p(x) = n(x) / n$

**F(x)** – kumulativní relativní četnost hodnot nepřevyšujících x;  $F(x) = N(x) / n$

Vizualizace



# Úkol 3

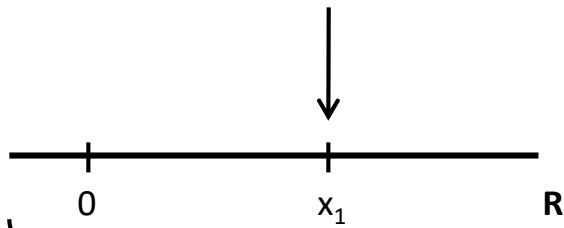
---

- Vytvořte novou proměnnou, která bude obsahovat logaritmovaný objem amygdaly.
- Vytvořte novou proměnnou, která bude obsahovat kategorizovanou váhu.

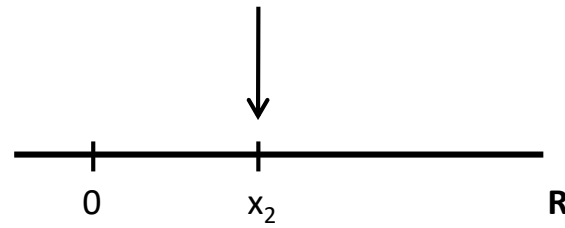
# 3. Intervaly spolehlivosti

# Intervaly spolehlivosti – motivace

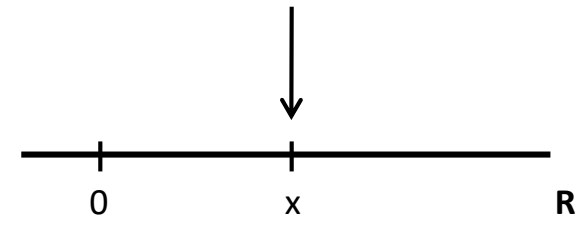
Výběr číslo 1



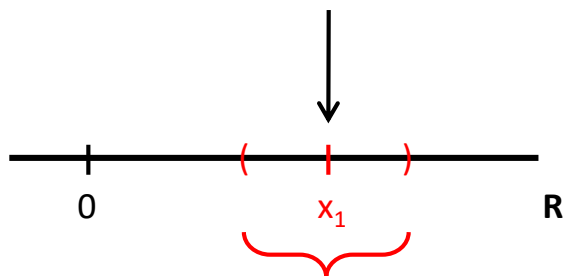
Výběr číslo 2



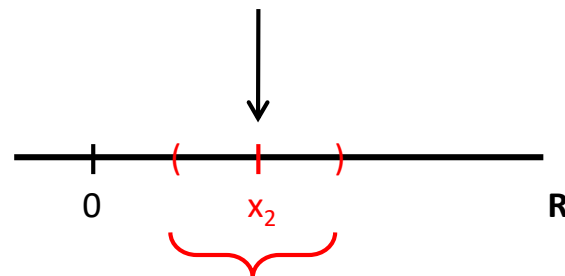
Celá cílová populace



Pracujeme-li s výběrem z cílové populace, je třeba na základě variability pozorovaných dat spočítat tzv. interval spolehlivosti pro bodový odhad.



Interval spolehlivosti na základě výběru číslo 1.



Interval spolehlivosti na základě výběru číslo 2.

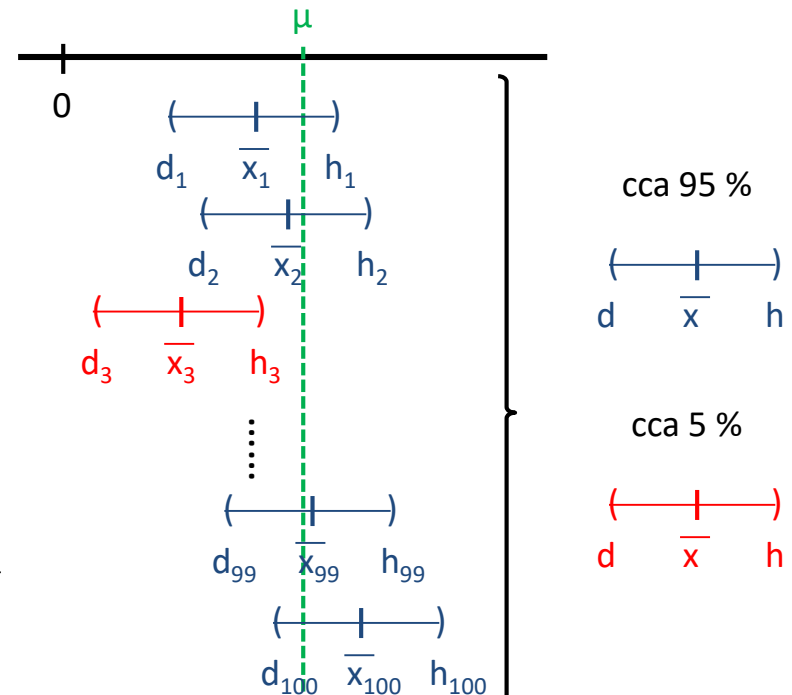
Umíme-li „změřit“ celou cílovou populaci, nepotřebujeme interval spolehlivosti, protože jsme schopni odhadnout sledovaný parametr přesně – v praxi je tato situace nereálná.



# Interval spolehlivosti (IS) – interpretace

- Interval spolehlivosti ukazuje, jak přesný je výpočet průměru.
- 95% interval spolehlivosti vymezuje prostor kam s 95% pravděpodobností padne populační průměr vypočtený při dalším vzorkování populace (za stejných podmínek a o stejné velikosti vzorku). Tedy 95% interval spolehlivosti obsahuje populační průměr s rizikem  $\alpha=0,05$  (5%).
- Čím je interval spolehlivosti užší, tím přesnější je náš odhad průměru (tím víc se náš odhad průměru pomocí našeho vzorku blíží populačnímu průměru).

- 95% interval spolehlivosti - ilustrace: Pokud bychom opakovaně vybírali skupiny subjektů o stejné velikosti a počítali průměr a interval spolehlivosti, tak 95% intervalů spolehlivosti by pokrývalo populační průměr  $\mu$  a 5% intervalů spolehlivosti by populační průměr nepokrývalo.



# Střední chyba průměru

- Nebo též standardní chyba průměru („standard error“) – značka  $SE$ .
- **Neplést se  $SD$  (směrodatnou odchylkou)!!!**
- Pozor na rozdíl mezi  $SD$  a  $SE$  v tabulkách a grafech v člancích a knihách!
- $SE = \frac{SD}{\sqrt{n}}$
- $SE$  je založena na směrodatné odchylce dat a počtu hodnot.
- Říká, jak přesný je výpočet průměru:
  - velký počet subjektů ( $n$ ), z nichž počítáme průměr → tím menší je  $SE$  (tzn. tím přesnější je průměr)
  - malý počet subjektů ( $n$ ), z nichž počítáme průměr → tím větší je  $SE$  (tzn. tím méně přesný je průměr)
- $SE$  se používá při výpočtu intervalů spolehlivosti.

# Interval spolehlivosti - poznámka

- Interval spolehlivosti (Confidence Interval – CI)
- Interval spolehlivosti pro průměr se vypočítá jako:

$$\bar{x} - SE \cdot 1,96 \leq \mu \leq \bar{x} + SE \cdot 1,96$$

- Interval spolehlivosti má smysl počítat pouze v případě, že mají data normální rozdělení!
- Interval spolehlivosti počítá pouze s variabilitou danou náhodným výběrem, nepočítá se zdroji systematického zkreslení – např.
  - Měření krevního tlaku může být systematicky zkresleno starým měřidlem („technical bias“).
  - Měření krevního tlaku může být systematicky zkresleno tím, že se do studie přihlásí pouze určitá skupina osob („selection bias“).

# Převody mezi $SD$ , $SE$ a $CI$

$\bar{x} \pm SD$

$\bar{x}$ : 170 cm  
SD: 10 cm



N=100

N=900

$\bar{x} \pm SE$

$$SE = SD / \sqrt{N}$$

$$SE = 10 / \sqrt{100}$$

$$SE = 10 / 10 = 1$$



$$SE = SD / \sqrt{N}$$

$$SE = 10 / \sqrt{900}$$

$$SE = 10 / 30 = 0,33$$



$\bar{x} \pm CI$

$$CI \cong 2 * SE$$

$$CI \cong 2 * 1 \cong 2$$



$$CI \cong 2 * SE$$

$$CI \cong 2 * 0,33 \cong 0,66$$



$\bar{x}$  - průměr,  $SD$  - směrodatná odchylka,  $SE$  - střední chyba průměru,  
 $CI$  - intervaly spolehlivosti

# Interval spolehlivosti pro $\mu$

$$P(D \leq \text{odhad} \leq H) > 1 - \alpha$$

## Obecný tvar intervalu spolehlivosti (IS):

$$\text{Odhadovaný parametr} \pm \text{Chyba odhadu} * \text{Kvantil modelového rozložení pro } (1-\alpha/2)$$

## Interval spolehlivosti pro $\mu$ :

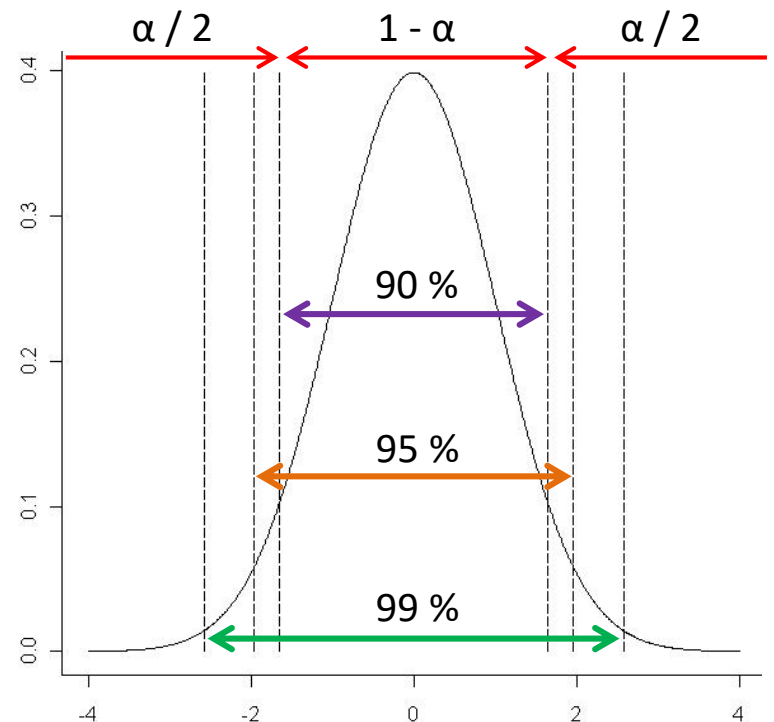
$$\underbrace{\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}}_{\text{dolní mez IS (D)}} \leq \mu \leq \underbrace{\bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}}_{\text{horní mez IS (H)}}$$

dolní mez IS (D)

horní mez IS (H)

- $\bar{x}$  ... výběrový průměr
- $\sigma$  ... směrodatná odchylka
- $n$  ... velikost výběrového souboru
- $z_{1-\alpha/2}$  ... kvantil standardizovaného normálního rozdělení
- $\alpha$  ... riziko
- $\frac{\sigma}{\sqrt{n}}$  ... střední chyba odhadu průměru

## Kvantily standardizovaného normálního rozdělení



$$z_{0,005} = -2,58$$

$$2,58 = z_{0,995}$$

$$z_{0,025} = -1,96$$

$$1,96 = z_{0,975}$$

$$z_{0,050} = -1,64$$

$$1,64 = z_{0,950}$$

# Ovlivnění šířky intervalu spolehlivosti

## Interval spolehlivosti:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$$

Co ovlivňuje šířku intervalu spolehlivosti?

- **Velikost vzorku** – s rostoucí velikostí vzorku je IS užší (máme více informace, a tak je odhad přesnější)
- **Variabilita náhodné veličiny** – čím náhodná veličina vykazuje větší variabilitu, tím je IS pro odhad střední hodnoty širší, tedy odhad je méně přesný.
- **Spolehlivost, kterou požadujeme** – s rostoucí spolehlivostí (tzn. menším  $\alpha$ ), je IS širší, neboť požadujeme větší jistotu, že náš interval skutečně pokrývá hodnotu neznámého parametru). Standardně se používá 95% IS (odpovídající riziku  $\alpha=5\%$ ), ale v literatuře se lze setkat i s 90% anebo 99% IS.

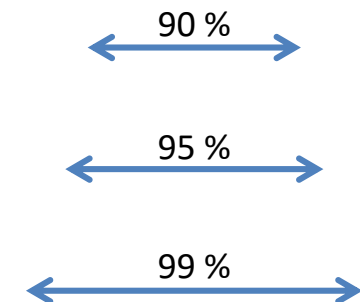
## Interval spolehlivosti je tedy UŽŠÍ s:

- **větším N** (větší velikostí vzorku)

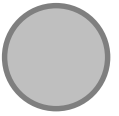
$$n=20: \quad \left( \begin{array}{c} | \\ d \quad \frac{1}{x} \quad h \end{array} \right)$$

$$n=100: \quad \left( \begin{array}{c} | \\ d \quad \frac{1}{x} \quad h \end{array} \right)$$

- menší variabilitou dat
- **menší spolehlivostí**



# Interval spolehlivosti pro $\mu$ při neznámém $\sigma$

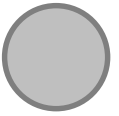


- IS pro  $\mu$  při známém  $\sigma$ : 
$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$$
- **IS pro  $\mu$  při neznámém  $\sigma$ :** 
$$\bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$
- Přesnou hodnotu populační  $\sigma$  v praxi většinou neznáme  $\rightarrow$  snažíme se ji odhadnout pomocí výběrové směrodatné odchylky  $s$ : 
$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$
- $t_{1-\alpha/2}(n-1)$  je kvantil Studentova  $t$  rozdělení
- **Příklad:** V našem souboru má 833 lidí průměrný věk roven 74,8 let a směrodatná odchylka věku je 6,9 let. Vypočtete 95% IS pro odhad střední hodnoty věku.
- Řešení:  
 $n = 833$   
 $\bar{x} = 74,8$  let  
 $s = 6,9$  let

$$\bar{x} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1) \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$
$$74,8 - \frac{6,9}{\sqrt{833}} t_{1-0,05/2}(833-1) \leq \mu \leq 74,8 + \frac{6,9}{\sqrt{833}} t_{1-0,05/2}(833-1)$$

$$74,3 \leq \mu \leq 75,3$$

# Další druhy intervalů spolehlivosti



- Interval spolehlivosti **pro rozdíl průměrů dvou výběrů** (jde nám např. o srovnání objemu hippocampu u pacientů a kontrol):

$$\bar{X} - \bar{Y} - t_{1-\alpha/2}(n_1 + n_2 - 2) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq \bar{X} - \bar{Y} + t_{1-\alpha/2}(n_1 + n_2 - 2) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- Interval spolehlivosti **pro odhad rozptylu**:

$$\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{\alpha/2}^2(n-1)}$$

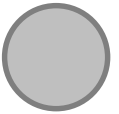
- Interval spolehlivosti **pro podíl rozptylů dvou výběrů** (lze ho použít pro hodnocení homogenity rozptylů dvou výběrů, která je jedním z předpokladů v testování hypotéz):

$$\frac{s_2^2}{s_1^2} F_{\alpha/2}(n_1 - 1, n_2 - 1) \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \frac{s_2^2}{s_1^2} F_{1-\alpha/2}(n_1 - 1, n_2 - 1)$$

- Druhů intervalů spolehlivosti je ještě mnohem více – např. IS pro medián, pro podíl,...



# Neparametrické metody pro konstrukci IS



- **Bootstrap** – je založen na principu opakovaného vzorkování naměřených dat s vracením, kdy pro vytvoření nového vzorku dat může být každý prvek použit více než jednou, právě jednou anebo není použit vůbec (ovšem se zachováním celkové velikosti souboru  $n$  i velikosti jednotlivých skupin). Pro každý vzorek je vypočítán výběrový průměr, tyto výběrové průměry seřadíme podle velikosti a vypočítáme 2,5% a 97,5% kvantil (stejně jako jsme počítali 80% kvantil na slidu 32), které nám dají dolní a horní mez pro 95% IS.
- **Jackknife** – opakovaný výpočet sledované charakteristiky je prováděn vždy s vynecháním právě jednoho pozorování. Tento postup nám stejně jako v případě metody bootstrap poskytuje představu o rozsahu hodnot, ve kterých se námi sledovaná charakteristika může pohybovat, budeme-li považovat naměřená data za reprezentativní vzorek z cílové populace.

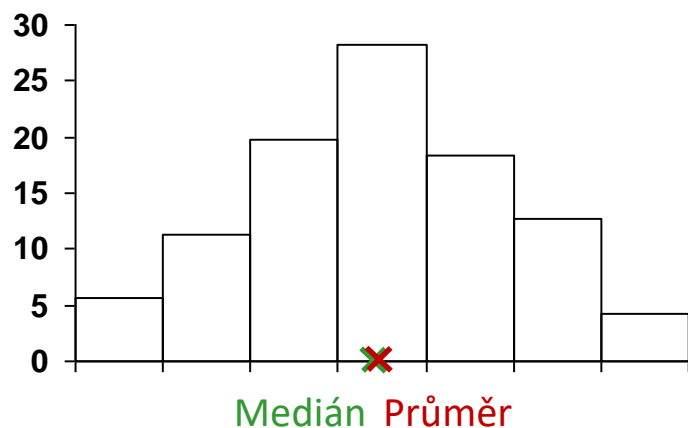
# Úkol 4

---

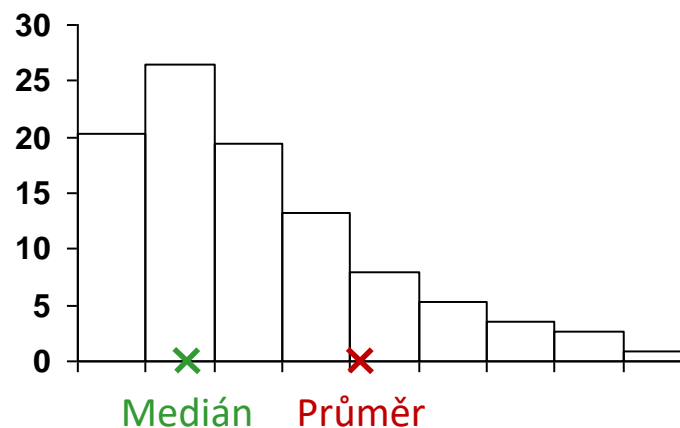
- Vypočtete průměr, střední chybu průměru (SE) a intervaly spolehlivosti pro všech šest mozkových struktur a MMSE skóre.
- Zamyslete se nad tím, zda mělo vůbec smysl počítat intervaly spolehlivosti pro všechny výše uvedené proměnné.

# Popis kvantitativních dat – shrnutí

## Symetrická data



## Asymetrická data



	Age
N	833
Průměr (Mean)	74,8
Směrodatná odchylka (SD)	6,9
95% interval spolehlivosti (CI)	74,3-75,3
Medián (Median)	75,0
Minimum	54,0
Maximum	90,0

	MMSE
N	833
Medián (Median)	27
Minimum	18
Maximum	30

# Ukázka popisné sumarizace

**Table 1 Patient demographics and baseline characteristics (intent-to-treat population)**

Characteristic	Placebo (n = 125)	Lipegfilgrastim 6 mg SC (n = 250)
Age, years		
Mean $\pm$ SD	58.7 $\pm$ 8.5	58.2 $\pm$ 8.5
$\leq 64$ , n (%)	94 (75.2)	193 (77.2)
65–74, n (%)	29 (23.2)	54 (21.6)
$\geq 75$ , n (%)	2 (1.6)	3 (1.2)
Weight, kg		
Mean $\pm$ SD	70.4 $\pm$ 13.4	69.0 $\pm$ 12.9
$\leq 60$ , n (%)	34 (27.2)	70 (28.0)
>60 to $\leq 75$ , n (%)	53 (42.4)	106 (42.4)
>75, n (%)	38 (30.4)	74 (29.6)
Sex, n (%)		
Female	20 (16.0)	30 (12.0)
Male	105 (84.0)	220 (88.0)
Time since diagnosis, months		
Mean $\pm$ SD	3.4 $\pm$ 9.1	2.4 $\pm$ 6.2
Median (range)	1.0 (0–58.0)	1.0 (0–52.0)

Zdroj: Volovat et al., *Phase III, randomized, double-blind, placebo-controlled, multicenter study of lipegfilgrastim in patients with non-small cell lung cancer receiving myelosuppressive therapy*, Springerplus. 2015; 4:316.

# 4. Formulování hypotéz nad medicínskými daty

# Statistické testování - cíle

---

## 1. Srovnávat.

- Jednu proměnnou s předpokládanou hodnotou
- Dvě nebo více proměnných mezi sebou

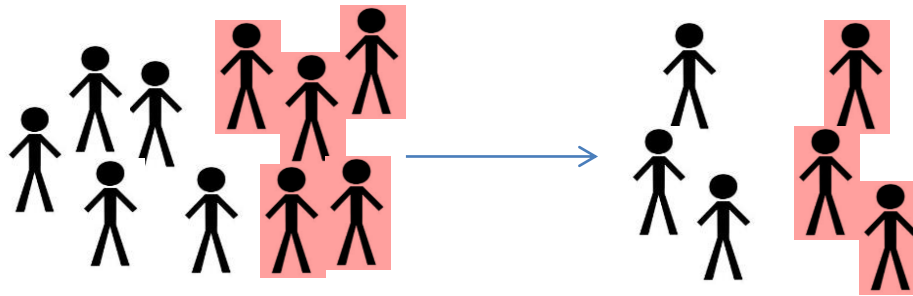
## 2. Hodnotit změnu proměnné vzhledem k vnějšímu zásahu.

## 3. Zjistit závislost dvou proměnných.

## 4. Zjistit typ rozdělení proměnné.

# Statistické testování - postup

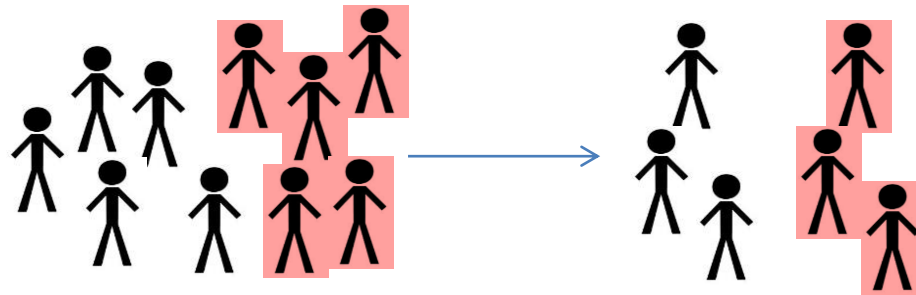
1. **Sestavíme hypotézu** k ověření (např. chceme ověřit, jestli se pacienti a zdravá populace liší v nějakém parametru)
2. **Vybereme** vhodný statistický test
3. **Stanovíme velikost vzorku** a **provedeme výběr z populace** (např. vybereme pacienty a zdravé lidi a naměříme zkoumaný parametr)



4. **Aplikujeme** vhodný **statistický test** a **rozhodneme**, jestli hypotézu zamítáme, nebo ne

# Statistické testování – **nesprávný** postup

1. Provedeme **výběr z populace** (např. vybereme pacienty a zdravé lidi a naměříme zkoumaný parametr)

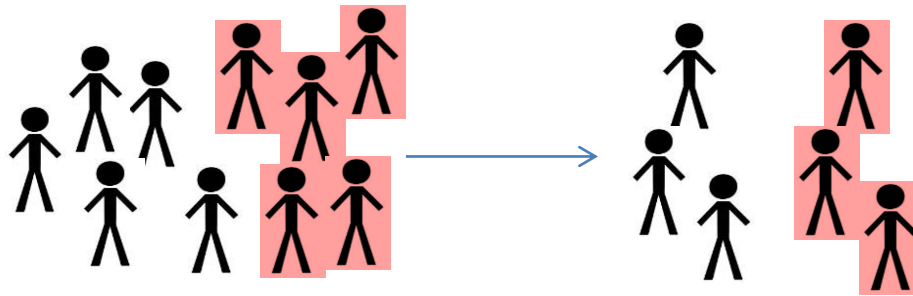


2. **Sestavíme hypotézu** k ověření (např. chceme ověřit, jestli se pacienti a zdravá populace liší v nějakém parametru)
3. **Vybereme** vhodný statistický test
4. **Aplikujeme** vhodný **statistický test** a **rozhodneme**, jestli hypotézu zamítáme, nebo ne



# Statistické testování - postup

1. **Sestavíme hypotézu** k ověření (např. chceme ověřit, jestli se pacienti a zdravá populace liší v nějakém parametru)
2. **Vybereme vhodný statistický test**
3. **Stanovíme velikost vzorku a provedeme výběr z populace** (např. vybereme pacienty a zdravé lidi a naměříme zkoumaný parametr)



4. **Aplikujeme vhodný statistický test a rozhodneme, jestli hypotézu zamítáme, nebo ne**

# 1. Sestavení hypotézy

- Statistické testy testují **nulovou hypotézu** ( $H_0$  – „null hypothesis“) – **tvrzení**, že se něco **nestalo** nebo **neprojevilo** (tzn. že efekt je nulový)
  - Není rozdíl v systolickém tlaku mezi skupinami A a B
  - Nepřítomnost efektu zlepšení stavu při nové léčbě v porovnání se standardníJe to **opak** toho, co chceme experimentem prokázat.  $H_0 : \theta = \theta_0$

- **Alternativní hypotéza** ( $H_1$  – „alternative hypothesis“) je tvrzení, které vymezuje, jaká situace nastává, když nulová hypotéza neplatí (tzn. efekt není nulový)

- Skupina A **nemá stejný** systolický tlak jako skupina B (**oboustranná alternativa**)

$$H_1 : \theta \neq \theta_0$$

- Skupina A **má menší** systolický tlak jako skupina B (**jednostranná alternativa**)

$$H_1 : \theta < \theta_0$$

- Skupina A **má větší** systolický tlak jako skupina B (**jednostranná alternativa**)

$$H_1 : \theta > \theta_0$$

# Hypotézy – příklady

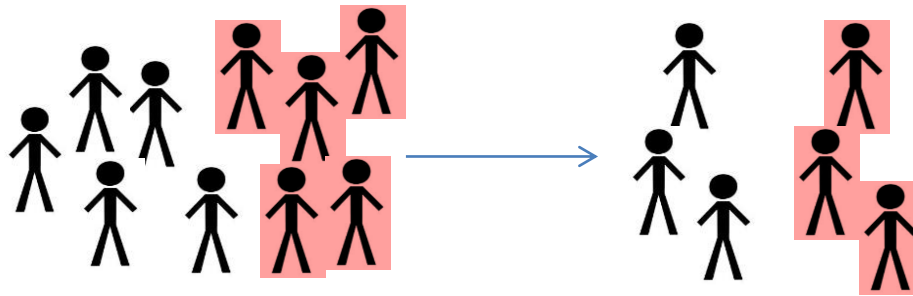
- 1. Je objem mozkových komor u pacientů s Alzheimerovou chorobou větší než u zdravých lidí?** Označme střední hodnotu objemu komor u pacientů symbolem  $\theta_1$  a střední hodnotu objemu komor u zdravých lidí  $\theta_2$ .  
**Nulová hypotéza:**  $H_0 : \theta_1 = \theta_2$  Není rozdíl v objemu komor u pacientů a kontrol.  
**Alternativní hypotéza:**  $H_1 : \theta_1 > \theta_2$  Objem u komor větší u pacientů než u kontrol.
- 2. Je průměrná hodnota MMSE skóre u pacientů s Alzheimer. chorobou menší než průměrná hodnota celé populace?** Označme střední hodnotu MMSE u pacientů symbolem  $\theta_1$  a u celé populace symbolem  $\theta_0$ .  
**Nulová hypotéza:**  $H_0 : \theta_1 = \theta_0$   
**Alternativní hypotéza:**  $H_1 : \theta_1 < \theta_0$
- 3. Liší se objem hipokampu u pacientů s Alzheimer. chorobou (AD), pacientů s mírnou kognitivní poruchou (MCI) a zdravých lidí (CN)?** Označme střední hodnotu objemu hipokampu u jednotlivých skupin symboly  $\theta_{AD}$ ,  $\theta_{MCI}$ ,  $\theta_{CN}$ .  
**Nulová hypotéza:**  $H_0 : \theta_{AD} = \theta_{MCI} = \theta_{CN}$   
**Alternativní hypotéza:**  $H_1 : \text{Nejméně jedno } \theta \text{ je odlišné od ostatních.}$

# Proč nulová hypotéza vyjadřuje nepřítomnost efektu?

- Nulová hypotéza je formulována jako opak toho, co chceme experimentem prokázat, proto, že je vždy jednodušší zamítnout hypotézu (na to stačí jeden případ, že hypotéza neplatí) než potvrdit hypotézu.
- Pokud se nám nepodaří nulovou hypotézu vyvrátit (tedy zamítnout), **mluvíme o nezamítnutí nulové hypotézy, ne o přijetí nulové hypotézy!!!** (možná jen nemáme dostatek důkazů, dostatečně velký soubor...)
- Platnost nulové hypotézy ověřujeme pomocí **statického testu** – rozhodovací pravidlo, které pozorovaným datům přiřadí právě jedno ze dvou možných rozhodnutí: nulovou hypotézu  $H_0$  na základě dat nezamítáme nebo nulovou hypotézu  $H_0$  zamítáme.

# Statistické testování - postup

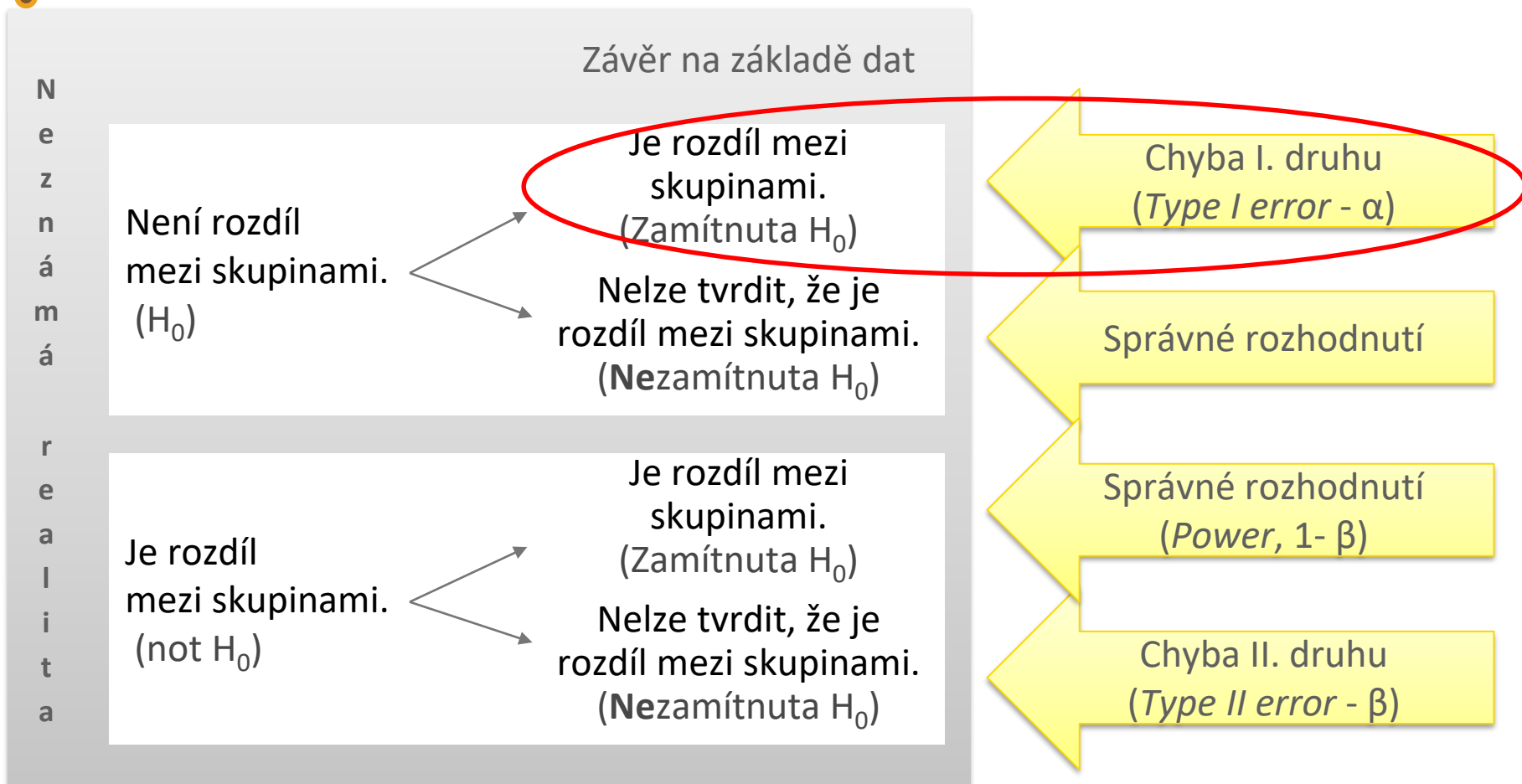
1. **Sestavíme hypotézu** k ověření (např. chceme ověřit, jestli se pacienti a zdravá populace liší v nějakém parametru)
2. **Vybereme vhodný statistický test**
3. **Stanovíme velikost vzorku a provedeme výběr z populace** (např. vybereme pacienty a zdravé lidi a naměříme zkoumaný parametr)



4. **Aplikujeme vhodný statistický test** a rozhodneme, jestli hypotézu zamítáme, nebo ne

# 5. Hladina významnosti a síla testu

# Co se při testování hypotéz může stát



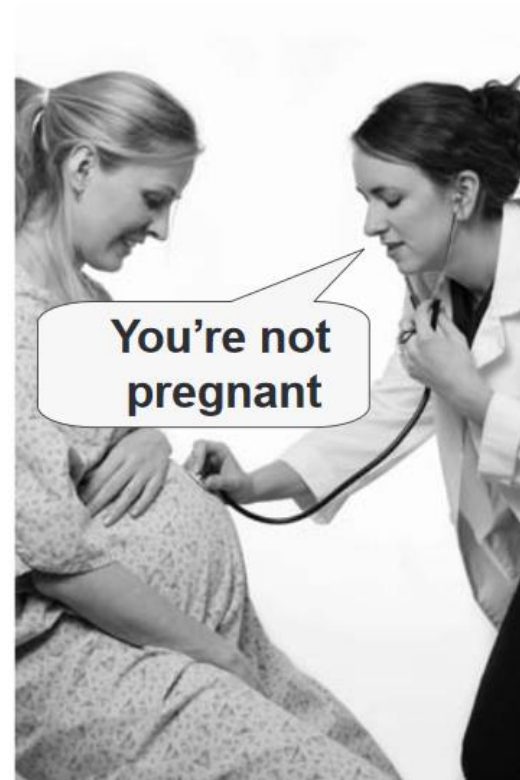
- **chyba I. druhu** – falešně pozitivní závěr testu – tzn. nesprávné zamítnutí nulové hypotézy (ve skutečnosti není rozdíl mezi skupinami, ale náš závěr z dat je opačný)
- **chyba II. druhu** – falešně negativní závěr testu – tzn. nerozpoznání neplatné nulové hypotézy (rozdíl mezi skupinami skutečně existuje, my ho ale nejsme schopni na základě dat statisticky prokázat)

# Co se při testování hypotéz může stát

**Type I error**  
(false positive)



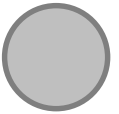
**Type II error**  
(false negative)



Zdroj: Paul D. Ellis, The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results, 2010



# Analogie se soudním procesem



- Ctíme presumpci nevinny = předpokládáme, že nulová hypotéza platí.
- **Požadujeme důkaz pro prokázání viny = na základě dat chceme ukázat, že nulová hypotéza neplatí.**

Rozhodnutí	Skutečnost	
	$H_0$ platí	$H_0$ neplatí
$H_0$ nezamítneme	správné přijetí platné nulové hypotézy	chyba II. druhu
$H_0$ zamítneme	chyba I. druhu	správné zamítnutí neplatné nulové hypotézy

- Když nám bude stačit málo důkazů, zvýší se procento odsouzených nevinných = **chyba I. druhu**, ale zároveň se zvýší i procento odsouzených, kteří jsou skutečně vinni = **správné zamítnutí neplatné nulové hypotézy**.
- Když budeme požadovat hodně důkazů, zvýší se procento nevinných, kteří budou osvobozeni = **správné přijetí platné nulové hypotézy**, ale zároveň se zvýší i procento vinných, kteří budou osvobozeni = **chyba II. druhu**.
- Proč hlídat spíše  $\alpha$  než  $\beta$ ?

Benjamin Franklin: „*It is better that 100 guilty persons should escape than that one innocent person should suffer.*“

# Chyba I. druhu ( $\alpha$ )

- Neboli **hladina významnosti testu (*level of significance*)**.
- Pravděpodobnost, že **ve skutečnosti není rozdíl mezi skupinami, ale náš závěr z dat je opačný**.
- Potřeba více hlídat – nastavíme fixně a poté se snažíme minimalizovat chybu II. druhu  $\beta$ , tzn. maximalizovat sílu testu  $1-\beta$ .
- Nejčastější hodnoty hladiny významnosti testu:
  - $\alpha = 0,05$  (tzn. 5%)
  - $\alpha = 0,01$  (tzn. 1%)
  - $\alpha = 0,001$  (tzn. 0,1%)

# Síla testu (*power*, $1-\beta$ )

- Pravděpodobnost, že zamítneme  $H_0$  ve chvíli, kdy  $H_0$  opravdu neplatí – tzn. **prokážeme rozdíl tam, kde skutečně existuje.**
- **Sílu testu zpravidla požadujeme 0,8 (80%) nebo vyšší.**
- **Vyšší počet vzorků nám umožní zvýšit sílu testu**
- Snažíme se sílu testu **optimalizovat** (tedy snížit  $\beta$ ) při **zachování** zvolené hladiny významnosti testu  $\alpha$
- **Prakticky:** *Snažíme se zjistit, kolik maximálně je třeba experimentálních subjektů (pozorování) k tomu, aby měl výsledný test dostatečnou sílu k zamítnutí nulové hypotézy, bude-li tato hypotéza skutečně neplatná).*
- K výpočtu vzorku je potřebné vědět, **jak velký efekt chceme prokázat** (čím nižší efekt, tím vyšší počet subjektů)

# Síla testu ( $1-\beta$ )

---

- Proč je důležité optimalizovat velikost vzorku před provedením studie?
  - 1.
  - 2.
  - 3.
- Rizika neplánovaného počtu subjektů ve studii:
  - **malý vzorek**
  - **velký vzorek**

# Síla testu ( $1-\beta$ )

- Proč je důležité optimalizovat velikost vzorku před provedením studie?
  1. **etické aspekty** – nelze zbytečně léčit lidi
  2. **ekonomické aspekty** – zbytečné plýtvání prostředky
  3. **statistické vlastnosti** – při velkém N lze prokázat cokoliv
- Rizika neplánovaného počtu subjektů ve studii:
  - **malý vzorek** – ztráta času, nemožnost prokázat rozdíl mezi srovnávanými skupinami pacientů
  - **velký vzorek** – ztráta času a prostředků, průkaz klinicky nevýznamného rozdílu mezi srovnávanými skupinami pacientů

# Faktory ovlivňující sílu testu

- Vychází se z výpočtu intervalu spolehlivosti:

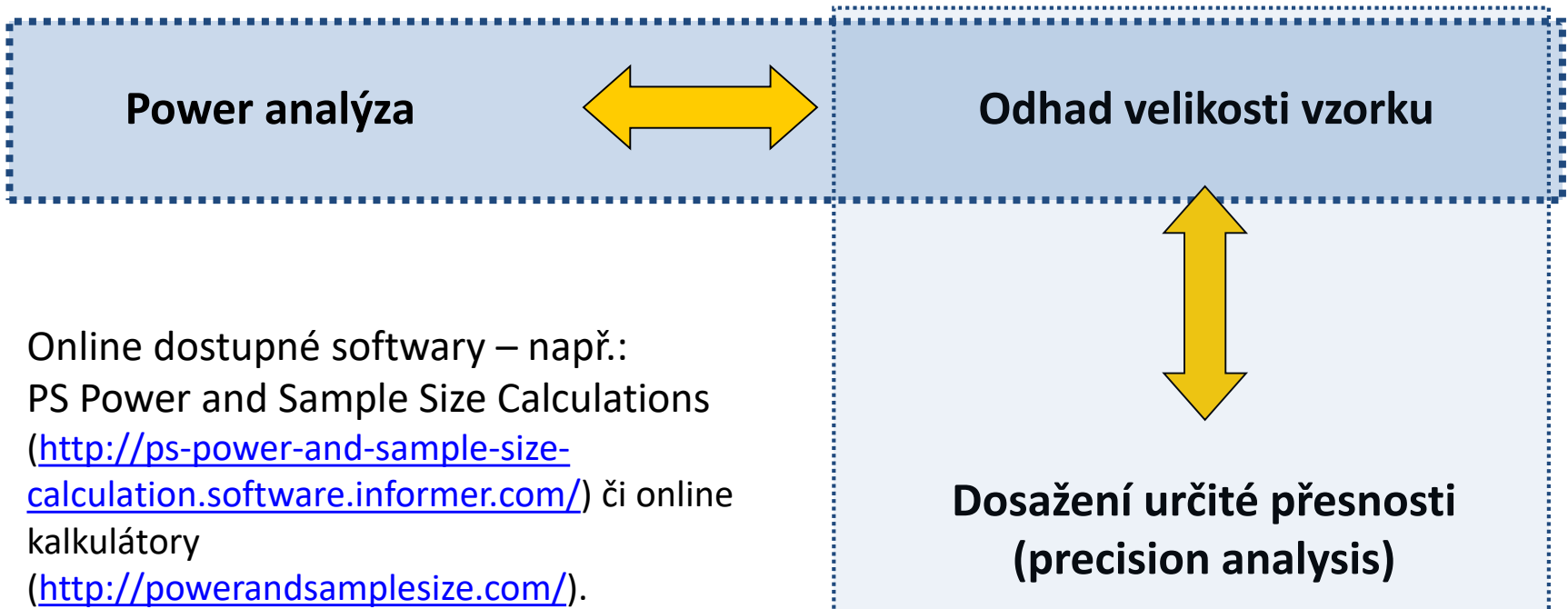
$$\text{Odhadovaný parametr} \pm \text{Kvantil modelového rozložení pro } (1-\alpha/2) * \frac{\sigma}{\sqrt{n}}$$

Chyba odhadu

- **Velikost vzorku:** čím více pozorování (informace o platnosti nulové hypotézy), tím větší má test sílu. Síla testu roste s odmocninou z  $n$ .
- **Velikost efektu (účinku):** velikost rozdílu v neznámých parametrech také ovlivňuje sílu testu. Vždy je jednodušší identifikovat jako významný velký efekt, např. velký rozdíl ve středních hodnotách objemu prostaty dvou populací. Naopak je těžší prokázat jako významný menší efekt (menší rozdíl).
- **Variabilita dat:** variabilita dat zvyšuje variabilitu odhadů a ztěžuje tak rozhodnutí o  $H_0$ . Čím více jsou pozorované hodnoty variabilní, tím více dat bude potřeba pro přesný odhad velikosti účinku (rozdílu).
- **Hladina významnosti:** snížíme-li hladinu významnosti testu (např. zvolíme 0,01 místo 0,05), bude obtížnější  $H_0$  zamítnout → sníží se síla testu.

# Power analýza a optimalizace velikosti vzorku

- Power analýza (analýza síly testu) a optimalizace velikosti vzorku (sample size estimation) jsou dvě strany téže mince.
- Obě vycházejí z testování hypotéz, jednou však máme jako předpoklad požadovanou sílu testu a chceme optimalizovat  $n$ , podruhé jsme limitováni  $n$  a ptáme se, jaké jsme v našich podmínkách schopni dosáhnout síly testu.



# Odhad velikosti vzorku - ukázka

Power and Sample Size Home / Calculators / Compare 2 Means: 2-Sample, 2-Sided Equality

SPONSORED SEARCHES

sample size calculator

g power effect size

population and sample

statistical power analysis

Calculate:  Sample Size

Vypočtená velikost vzorku B:  Sample Size,  $n_B$

Síla testu:  Power,  $1 - \beta$

Hladina významnosti:  Type I error rate,  $\alpha$

Group 'A' mean,  $\mu_A$  (průměr skupiny A)

Group 'B' mean,  $\mu_B$  (průměr skupiny B)

Standard Deviation,  $\sigma$  (směrodatná odchylka)

Sampling Ratio,  $\kappa = n_A/n_B$

Calculate

Při zvýšení síly testu na 0,9 bude potřeba min. 21 subjektů v každé skupině.

<http://powerandsamplesize.com>



# Power analýza - ukázka

Power and Sample Size Home / Calculators / Compare 2 Means: 2-Sample, 2-Sided Equality

SPONSORED SEARCHES

sample size calculator    g power effect size  
population and sample    statistical power analysis

Calculate:  Power

Velikost skupiny B:  Sample Size,  $n_B$

Vypočtená síla testu:  Power,  $1 - \beta$

Hladina významnosti:  Type I error rate,  $\alpha$

Group 'A' mean,  $\mu_A$  (průměr skupiny A)  
 Group 'B' mean,  $\mu_B$  (průměr skupiny B)  
 Standard Deviation,  $\sigma$  (směrodatná odchylka)  
 Sampling Ratio,  $\kappa = n_A/n_B$

Calculate

Pokud bychom měli 20 subjektů v každé skupině, síla testu by byla 0,886 (88,6%).

<http://powerandsamplesize.com>

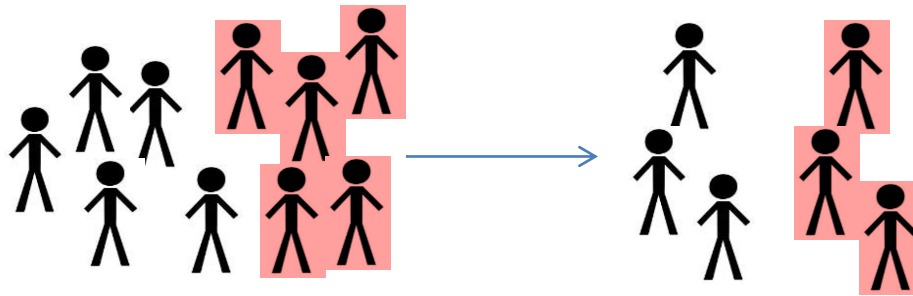
# Plánování klinického hodnocení fáze I - IV

COMPARISON OF CLINICAL TRIAL PHASES

	PHASE I	PHASE II	PHASE III	PHASE IV
<b>OBJECTIVES:</b>	Determine the metabolic and pharmacological actions and the maximally tolerated dose	Evaluate effectiveness, determine the short-term side effects and identify common risks for a specific population and disease	Obtain additional information about the effectiveness on clinical outcomes and evaluate the overall risk-benefit ratio in a demographically diverse sample	Monitor ongoing safety in large populations and identify additional uses of the agent that might be approved by the FDA
<b>FACTORS TO BE IDENTIFIED:</b>	<ul style="list-style-type: none"> <li>-Bioavailability</li> <li>-Bioequivalence</li> <li>-Dose proportionality</li> <li>-Metabolism</li> <li>-Pharmacodynamics</li> <li>-Pharmacokinetics</li> </ul>	<ul style="list-style-type: none"> <li>-Bioavailability</li> <li>-Drug-disease interactions</li> <li>-Drug-drug interactions</li> <li>-Efficacy at various doses</li> <li>-Pharmacodynamics</li> <li>-Pharmacokinetics</li> <li>-Patient safety</li> </ul>	<ul style="list-style-type: none"> <li>-Drug-disease interactions</li> <li>-Drug-drug interactions</li> <li>-Dosage intervals</li> <li>-Risk-benefit information</li> <li>-Efficacy and safety for subgroups</li> </ul>	<ul style="list-style-type: none"> <li>-Epidemiological data</li> <li>-Efficacy and safety within large, diverse populations</li> <li>-Pharmacoeconomics</li> </ul>
<b>DATA FOCUS:</b>	<ul style="list-style-type: none"> <li>-Vital signs</li> <li>-Plasma and serum levels</li> <li>-Adverse events</li> </ul>	<ul style="list-style-type: none"> <li>-Dose response and tolerance</li> <li>-Adverse events</li> <li>-Efficacy</li> </ul>	<ul style="list-style-type: none"> <li>-Laboratory data</li> <li>-Efficacy</li> <li>-Adverse events</li> </ul>	<ul style="list-style-type: none"> <li>-Efficacy</li> <li>-Pharmacoeconomics</li> <li>-Epidemiology</li> <li>-Adverse events</li> </ul>
<b>DESIGN FEATURES:</b>	<ul style="list-style-type: none"> <li>-Single, ascending dose tiers</li> <li>-Unblinded</li> <li>-Uncontrolled</li> </ul>	<ul style="list-style-type: none"> <li>-Placebo controlled comparisons</li> <li>-Active controlled comparisons</li> <li>-Well-defined entry criteria</li> </ul>	<ul style="list-style-type: none"> <li>-Randomized</li> <li>-Controlled</li> <li>-2-3 treatment arms</li> <li>-Broader eligibility criteria</li> </ul>	<ul style="list-style-type: none"> <li>-Uncontrolled</li> <li>-Observational</li> </ul>
<b>DURATION:</b>	Up to 1 month	Several months	Several years	Ongoing (following FDA approval)
<b>POPULATION:</b>	Healthy volunteers or individuals with the target disease (such as cancer or HIV)	Individuals with target disease	Individuals with target disease	Individuals with target disease, as well as new age groups, genders, etc.
<b>SAMPLE SIZE:</b>	20 to 80	200 to 300	Hundreds to thousands	Thousands
<b>EXAMPLE:</b>	Study of a single dose of Drug X in normal subjects	Double-blind study evaluating safety and efficacy of Drug X vs. placebo in patients with hypertension	Study of Drug X vs. standard treatment in hypertension study	Study of economic benefit of newly-approved Drug X vs. standard treatment for hypertension

# Statistické testování - postup

1. **Sestavíme hypotézu** k ověření (např. chceme ověřit, jestli se pacienti a zdravá populace liší v nějakém parametru)
2. **Vybereme vhodný statistický test**
3. **Stanovíme velikost vzorku a provedeme výběr z populace** (např. vybereme pacienty a zdravé lidi a naměříme zkoumaný parametr)



4. **Aplikujeme vhodný statistický test a rozhodneme, jestli hypotézu zamítáme, nebo ne**

# 6. p-hodnota

# p-hodnota („p-value“, „p-level“)

- Neboli **dosažená hladina významnosti testu**.
- Značka:  $p$
- Je to pravděpodobnost, s jakou bychom mohli obdržet pozorovaná data nebo data stejně, či ještě více odporující nulové hypotéze, za předpokladu, že je nulová hypotéza pravdivá.
- Čím menší je  $p$ , tím neudržitelnější čili méně důvěryhodná je nulová hypotéza.
- Hodnocení, kdy je výsledek testu statisticky významný:
  - Máme zvolenu hladinu významnosti testu (např.  $\alpha=0,05$ ).
  - Dvě možné situace:
    1.  $p < \alpha$  – **zamítáme  $H_0$**  – statisticky významný výsledek testu
    2.  $p \geq \alpha$  – **nezamítáme  $H_0$**

# p-hodnota – ukázka I

**Table 3 Incidence of severe neutropenia in cycles 1–4 (per-protocol population)\***

	Cycle 1		Cycle 2		Pegfilgrastim 6 mg
	Pegfilgrastim 6 mg	Lipegfilgrastim 6 mg	Pegfilgrastim 6 mg	Lipegfilgrastim 6 mg	
n/N	48/94	41/94	20/93	8/94	11/91
(%)	(51.1)	(43.6)	(21.5)	(8.5)	(12.1)
OR (95% CI)	0.745 (0.405%, 1.369%)		0.291 (0.110%, 0.769%)		0.676
<i>p</i> value <sup>†</sup>	0.3409		0.0130		

\*Multiple mentions per patient are possible.

<sup>†</sup>Pegfilgrastim 6 mg vs. lipegfilgrastim 6 mg (*p* values are based on a null hypothesis of OR = 1).  
CI confidence interval, OR odds ratio.

statisticky  
nevýznamný  
výsledek

statisticky  
významný  
výsledek

Zdroj: Bondarenko et al., *Efficacy and safety of lipegfilgrastim versus pegfilgrastim: a randomized, multicenter, active-control phase 3 trial in patients with breast cancer receiving doxorubicin/docetaxel chemotherapy*, BMC Cancer. 2013; 13:386.

# p-hodnota – ukázka II

**Table 4 Incidence of severe and very severe neutropenia in cycles 1, 2, 3, and 4 (ITT population)**

Cycle	Placebo			Lipegfilgrastim 6 mg SC			Lipegfilgrastim 6 mg SC vs. placebo		
	<i>N</i>	<i>n</i>	%	<i>N</i>	<i>n</i>	%	Odds ratio	95% CI	<i>P</i> value*
Severe neutropenia (grade 4, ANC <0.5 × 10 <sup>9</sup> /L)									
1	125	74	59.2	249	80	32.1	0.325	0.206–0.512	<0.0001
2	105	55	52.4	215	36	16.7	0.156	0.086–0.282	<0.0001
3	92	47	51.1	188	26	13.8	0.115	0.057–0.229	<0.0001
4	81	45	55.6	169	25	14.8	0.121	0.062–0.238	<0.0001
All	125	100	80.0	249	103	41.4	0.176	0.105–0.294	<0.0001
Very severe neutropenia (ANC <0.1 × 10 <sup>9</sup> /L)									
1	125	18	14.4	249	27	10.8	0.700	0.365–1.342	NS
2	105	10	9.5	215	8	3.7	0.298	0.099–0.895	0.031
3	92	9	9.8	188	9	4.8	0.421	0.156–1.138	NS
4	81	11	13.6	169	8	4.7	0.260	0.098–0.687	0.007
All	125	33	26.4	249	40	16.1	0.516	0.300–0.888	0.017

Based on data actually collected.

ANC absolute neutrophil count, CI confidence interval, ITT intent to treat, NS not significant, SC subcutaneously.

\**P* values based on a null hypothesis of odds ratio = 1.

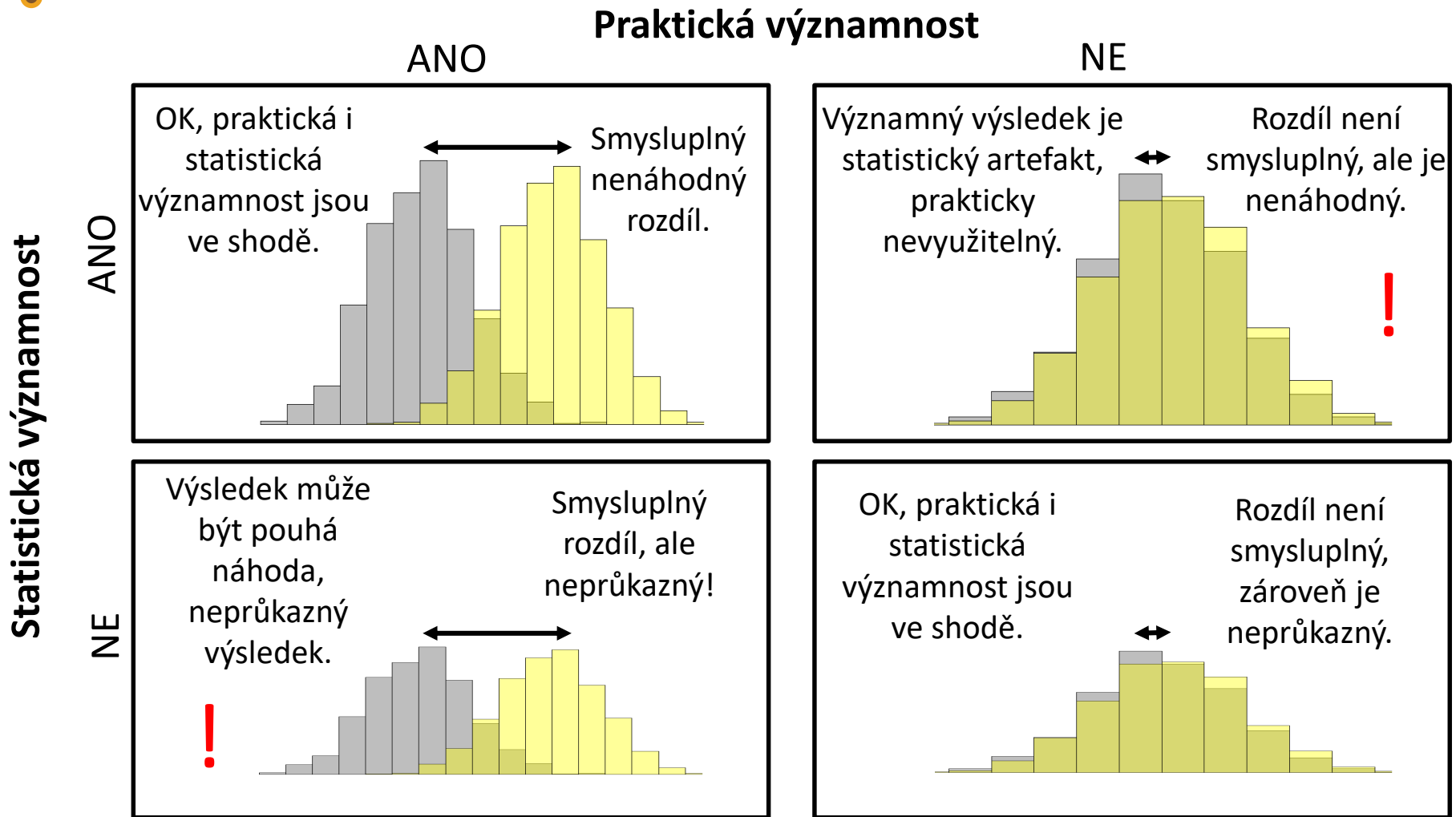
Zdroj: Volovat et al., *Phase III, randomized, double-blind, placebo-controlled, multicenter study of lipegfilgrastim in patients with non-small cell lung cancer receiving myelosuppressive therapy*, Springerplus. 2015; 4:316.

# Důležité poznámky k testování hypotéz

- **Nezamítnutí nulové hypotézy neznamena automaticky její přijetí!** Může se jednat o situaci, kdy pro zamítnutí nulové hypotézy nemáme dostatečné množství informace.
- **Dosažená hladina významnosti testu** (ať už 0,05, 0,01 nebo 0,10) **nesmí být slepě brána jako hranice pro existenci/neexistenci testovaného efektu.** Neexistuje jasná hranice pro významnost či nevýznamnost – často je velmi malý rozdíl mezi p-hodnotou 0,049 a p-hodnotou 0,051.
- **Malá p-hodnota nemusí znamenat velký efekt.** Hodnota testové statistiky a odpovídající p-hodnota může být ovlivněna velkou velikostí vzorku a malou variabilitou pozorovaných dat.
- **Výsledky testování musí být nahlíženy kriticky** – jedná se o závěr založený „pouze“ na jednom výběrovém souboru.

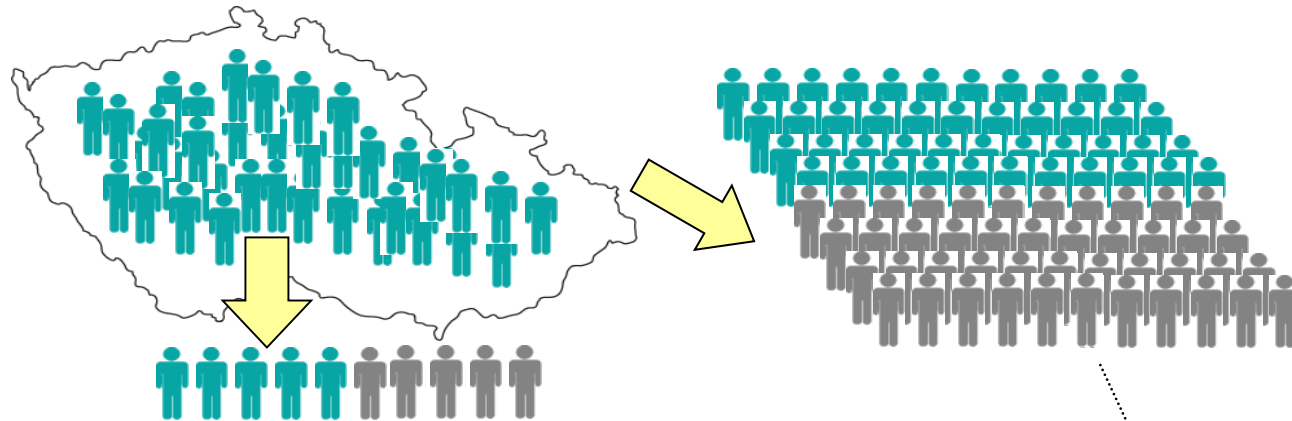


# Statistická vs. praktická (klinická) významnost



Statisticky nevýznamný výsledek neznamená, že pozorovaný rozdíl ve skutečnosti neexistuje! Může to být způsobeno nedostatečnou informací v pozorovaných datech!

# Vliv velikosti vzorku na výsledky testování



Dvě skupiny pacientů s nepatrným rozdílem v dané charakteristice, který ale není klinicky významný.

$n_1 = 10, n_2 = 10$   
 $p = 0,797$

$n_1 = 100, n_2 = 100$   
 $p = 0,140$

$n_1 = 1000, n_2 = 1000$   
 $p < 0,001$

**Statistická významnost způsobená velkým N**

# Hodnocení velikosti účinku

- absolutní velikost účinku při srovnání dvou skupin – rozdíl odhadů průměrů:  $\bar{x}_1 - \bar{x}_2$
- koeficienty velikosti účinku:
  - dosažený efekt standardizují a jsou tak využitelné pro srovnávání různých experimentů (uplatnění v metaanalýzách)
  - např. Cohenův koeficient  $d$ :  $d = \frac{\bar{x}_1 - \bar{x}_2}{s}$ , kde  $s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ 
    - velký efekt:  $d > 0,8$
    - střední efekt:  $0,5 < d \leq 0,8$
    - malý efekt:  $0,2 < d \leq 0,5$
    - zanedbatelný efekt:  $d \leq 0,2$
- korelační koeficienty (hodnocení míry vztahu dvou proměnných)

# Shrnutí klíčových pojmů analýzy dat

- **Významnost** – viz. předcházející slidy.
- **Zkreslení výsledků** („biased results“) – zkreslení způsobené starým nebo nenakalibrovaným měřidlem („technical bias“), zkreslení nevhodným výběrem subjektů („selection bias“), sledování zavádějícího faktoru namísto faktoru, který je pravou příčinou sledovaného výsledku.
- **Reprezentativnost** – experimentální vzorek musí svými charakteristikami odpovídat cílové populaci.
- **Srovnatelnost** – pokud chceme srovnávat skupiny mezi sebou, musí být skupiny srovnatelné. Pokud nemůžeme provést randomizaci (tzn. náhodné rozdělení subjektů do skupin), musíme hlídat, aby skupiny byly srovnatelné. Pokud nejsou, můžeme vytvořit podskupiny a ty srovnávat mezi sebou, nebo se snažíme odstranit vliv „nechtěných“ faktorů.
- **Spolehlivost** – sumarizace sledované proměnné jedním číslem (např. průměrem) není dostatečná, protože nepostihujeme variabilitu dat – průměr vypočítaný z dat 10 lidí bude určitě méně přesný (spolehlivý) než průměr vypočítaný z dat 1000 lidí → průměr doplníme o interval spolehlivosti.

# Poděkování...

Příprava výukových materiálů předmětu „DSAN01 Analýza dat pro Neurovědy“ byla finančně podporována prostředky projektu FRVŠ č. 942/2013 „Inovace materiálů pro interaktivní výuku a samostudium předmětu Analýza dat pro Neurovědy“

