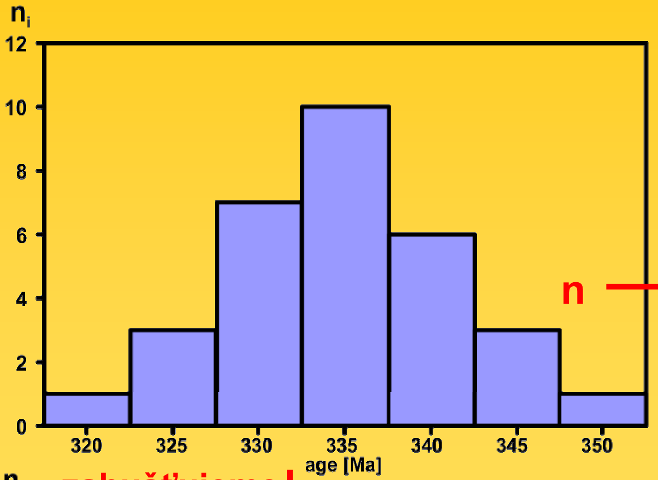
Základy zpracování geologických dat

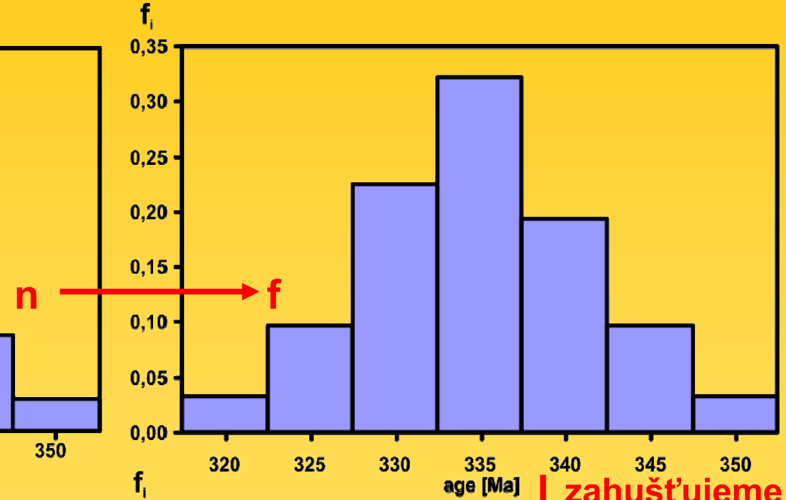
Rozdělení pravděpodobnosti

R. Čopjaková

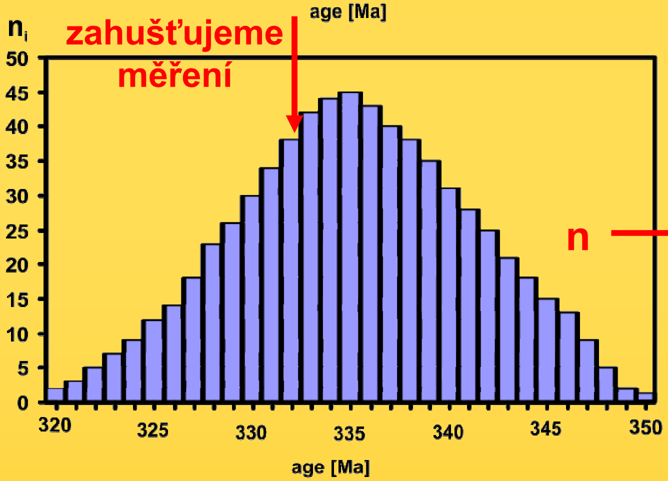
Od četnosti k pravděpodobnosti

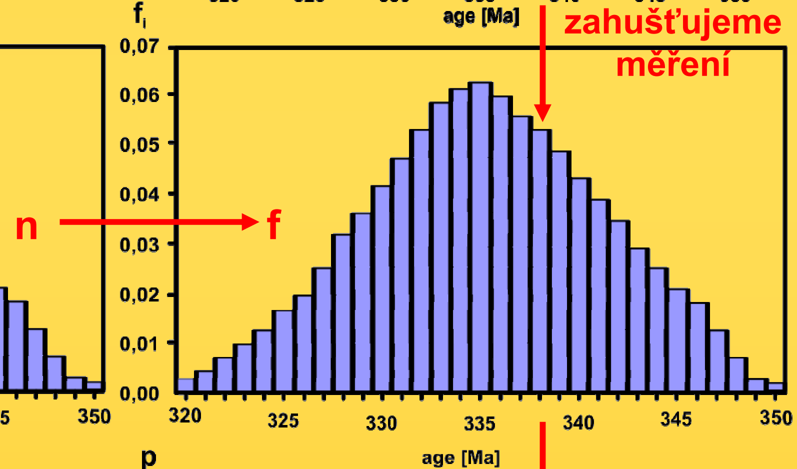
střed int	n_i	f_i
320	1	0,03
325	3	0,10
330	7	0,23
335	10	0,32
340	6	0,19
345	3	0,10
350	1	0,03
suma	31	1

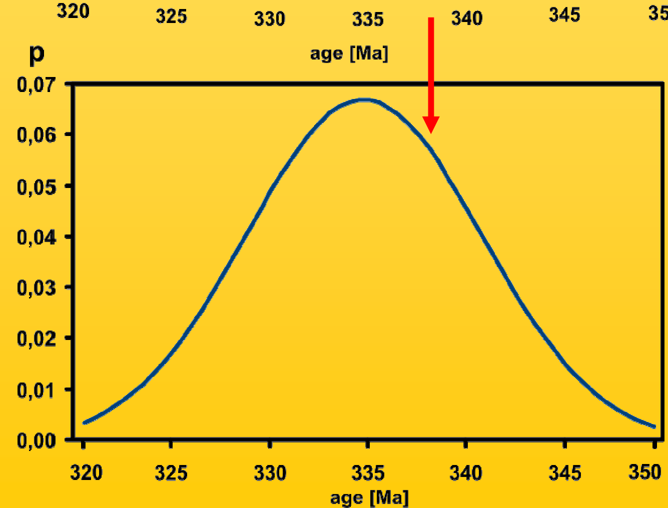




střed int	n_i	f_i
320	2	0,00
321	3	0,00
322	5	0,01
323	7	0,01
324	9	0,01
325	12	0,02
326	14	0,02
327	18	0,03
328	23	0,03
329	26	0,04
330	30	0,04
331	34	0,05
332	38	0,06
333	42	0,06
334	44	0,07
335	45	0,07
336	43	0,06
337	40	0,06
338	38	0,06
339	35	0,05
340	31	0,05
341	28	0,04
342	25	0,04
343	21	0,03
344	18	0,03
345	15	0,02
346	13	0,02
347	9	0,01
348	5	0,01
349	2	0,00
350	1	0,00
suma	676	1







Hustota rozdělení pravděpodobností
frekvenční funkce
pravděpodobnostní funkce

Rozdělení pravděpodobnosti

- Spojité náhodné veličiny
- Diskrétní náhodné veličiny

Normální rozdělení (spojité)

Hustota rozdělení pravděpodobnosti (neboli frekvenční funkce)

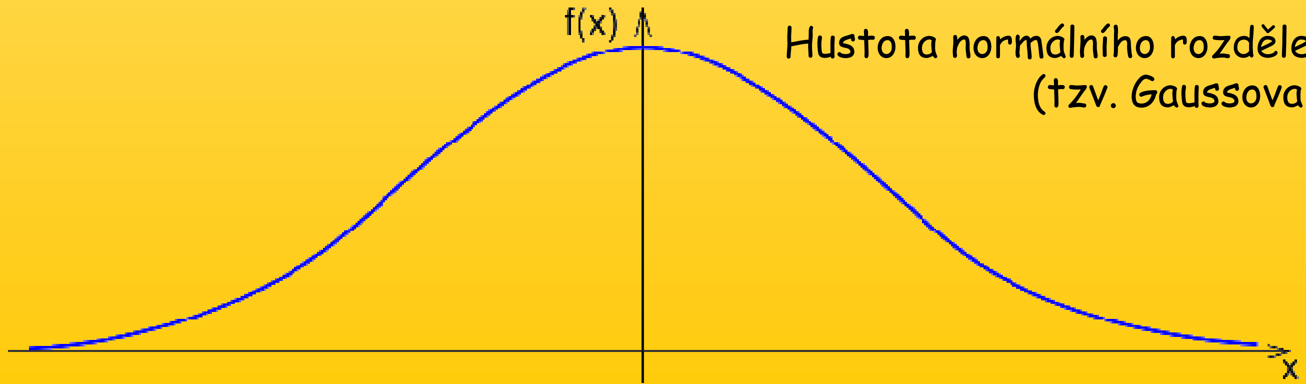
- Normální rozdělení pravděpodobnosti s parametry μ a σ^2 je definováno hustotou pravděpodobnosti ve tvaru

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

  - kde μ = (Ex) je střední hodnota normálního rozdělení
  - a $\boldsymbol{\sigma^2}$ je rozptyl, (**σ** je směrodatná odchylka)

základní soubor		výběrový soubor	
μ	$\longrightarrow$	$\bar{x}$	$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$
$\boldsymbol{\sigma^2}$	$\longrightarrow$	$\boldsymbol{S_x^2}$	$S_X^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})^2$

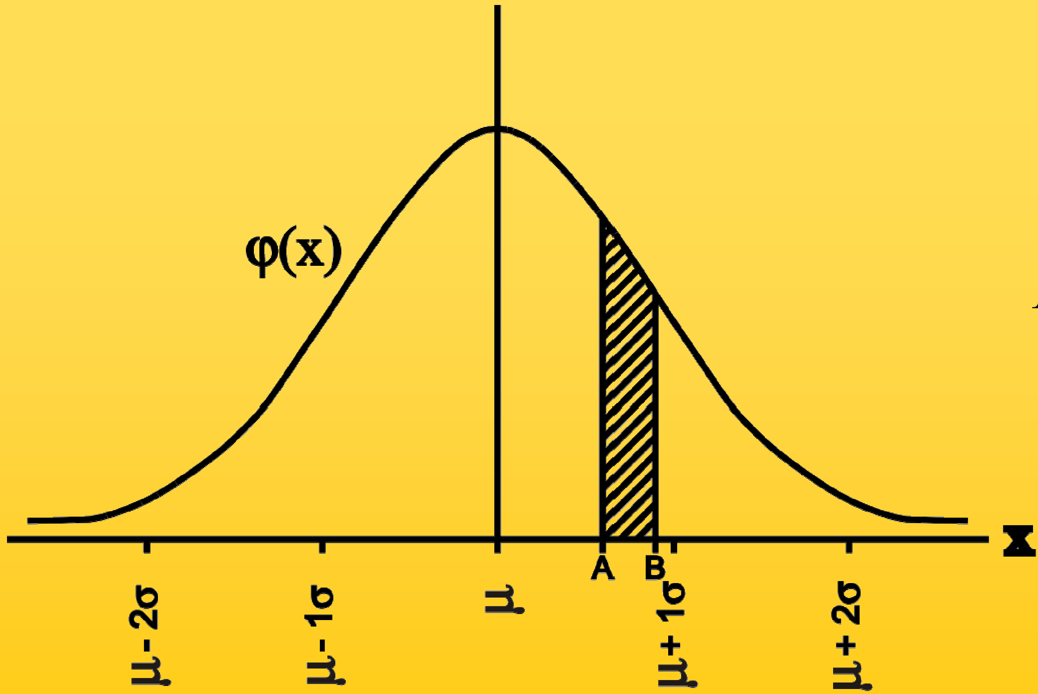
- Střední hodnota (Ex) normálního rozložení by se měla blížit aritmetickému průměru naměřených hodnot
- Normální rozdělení se značí $\mathrm{N}(\mu, \sigma^2)$



Hustota normálního rozdělení pravděpodobnosti (tzv. Gaussova křivka)

Normální rozdělení

- Plocha pod křivkou hustoty má velikost 1.
- Protože hustota je symetrická kolem μ, znamená to, že μ dělí plochu pod křivkou na dvě stejné části - každá z nich má tedy velikost 1/2.
- Ve vzdálenosti jedné směrodatné odchylky se nacházejí **inflexní body** fce
- Pravděpodobnost P, že náhodná veličina bude mít hodnotu padnoucí do intervalu <A;B> je rovna ploše vyšrafované v grafu.



$$P = \int_A^B \varphi(x)\,dx$$

Frekvenční křivka normálního rozdělení

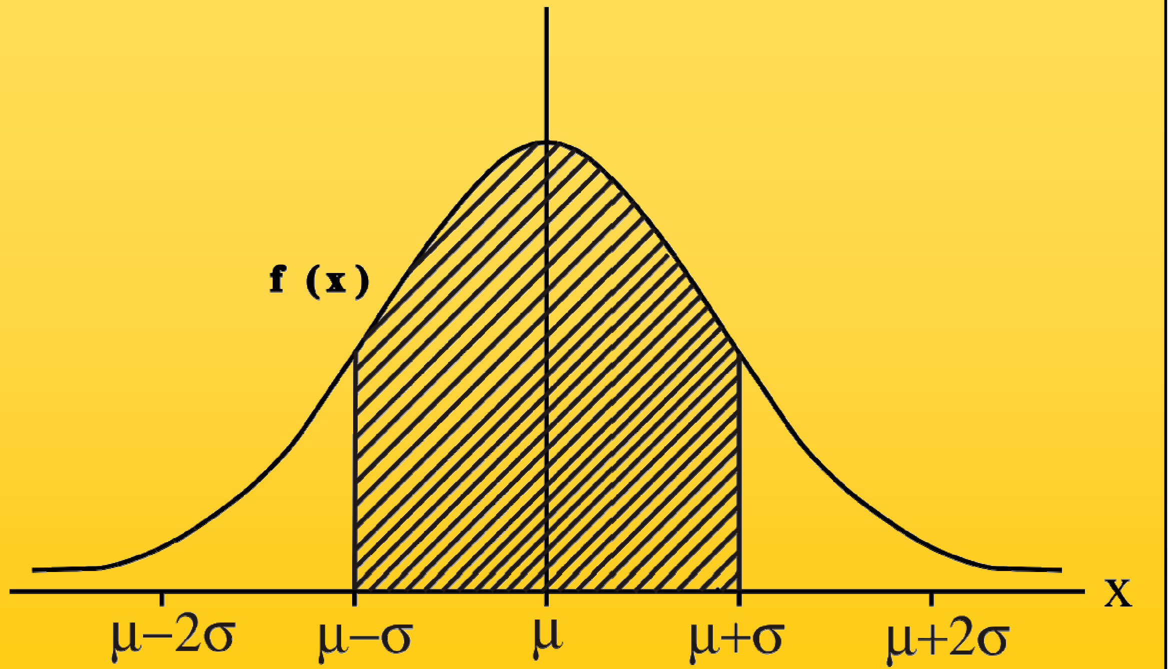
rozptyl $$Sx^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

směrodatná odchylka $$Sx = \sqrt{Sx^2}$$

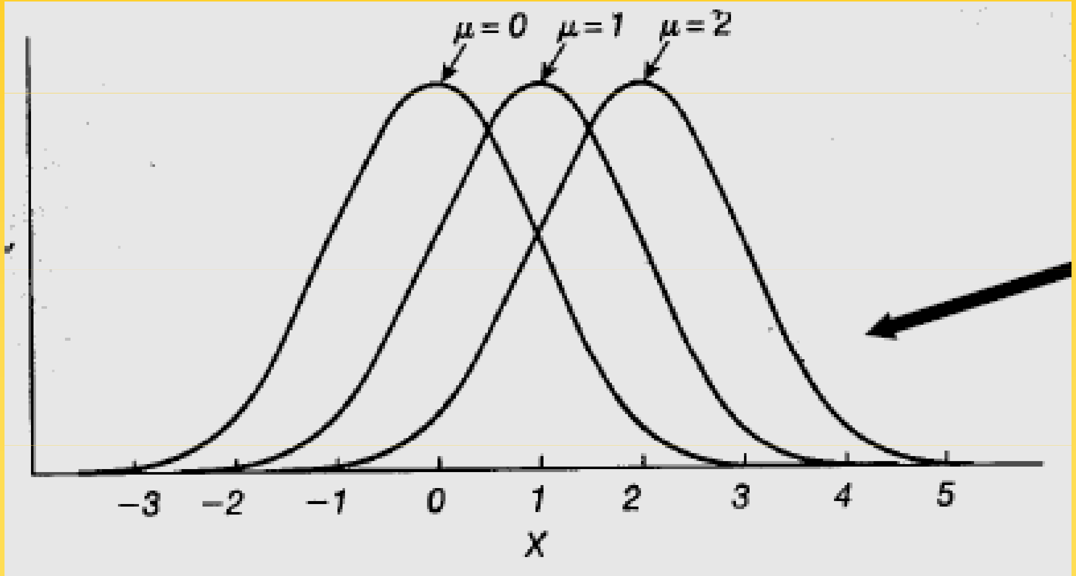
Rozptyl Sx^2 určuje průměrnou čtvercovou odchylku od aritmetického průměru
Směrodatná odchylka Sx nám umožňuje odhadnout interval, ve kterém očekáváme naměřenou hodnotu

Pravděpodobnost P, že měření bude zatíženo chybou o velikosti padnoucí do intervalu <A;B> je rovna ploše vyšrafované v grafu.

A	B	P
$\mu - \sigma$	$\mu + \sigma$	0,682
$\mu - 2\sigma$	$\mu + 2\sigma$	0,954
$\mu - 3\sigma$	$\mu + 3\sigma$	0,997
$\mu - \infty$	$\mu + \infty$	1,000



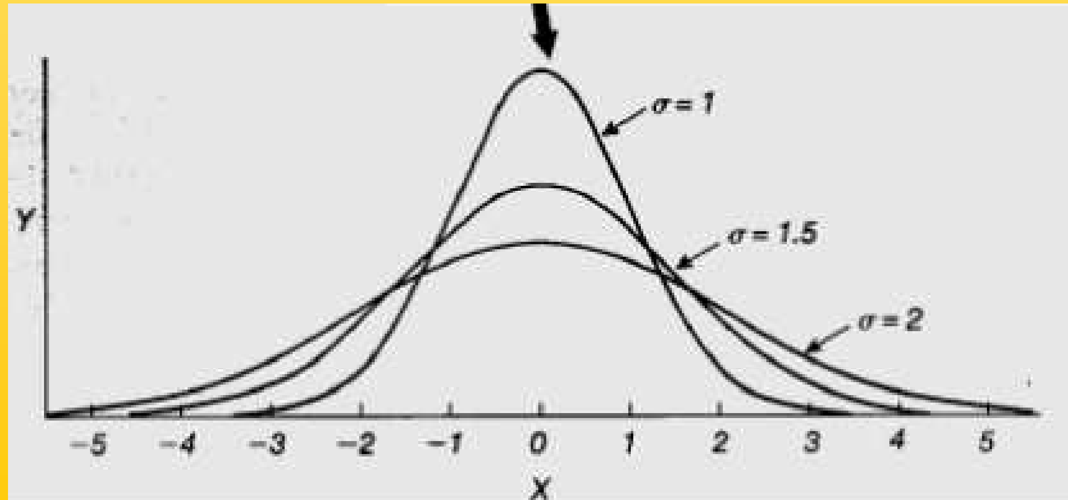
Normální rozdělení



Hustota rozdělení pravděpodobností

a) Pro stejné σ a různé μ

b) Pro různé σ a stejné μ



- Normální rozdělení má zásadní význam v teorii pravděpodobnosti a matematické statistice a řídí se jím (alespoň "přibližně") mnoho náhodných veličin. Nejběžnějším typem takových veličin jsou náhodné chyby (chyby měření). Proto se normálnímu rozdělení někdy říká rozdělení chyb.

- Rovněž mnohé náhodné veličiny v geologii se řídí tímto rozdělením nebo jejich rozdělení jím může být velmi dobře aproximováno.

- Někdy se rovněž můžeme setkat s označením Gaussova křivka pro označení hustoty normálního rozdělení, podle jednoho z *praotců* tohoto rozdělení.

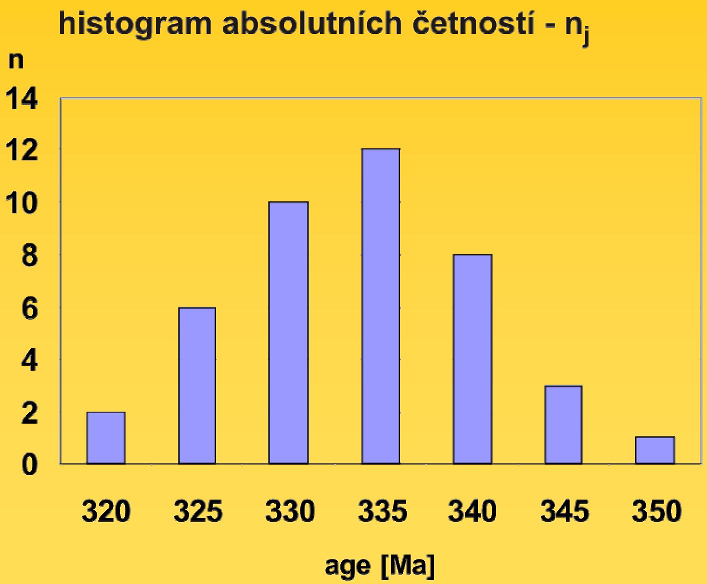
- Normální rozdělení je jednovrcholové rozdělení symetrické okolo střední hodnoty, kterou budeme značit μ. Hustota pravděpodobnosti má *zvonovitý* tvar - maxima dosahuje ve střední hodnotě.
- "Konce" tohoto rozdělení vypadají, jako by se již dotýkaly osy x, nikdy se jí však nedotknou, i když jsou jí tím blíže, čím více se vzdalujeme od střední hodnoty μ - ať již doleva či doprava.

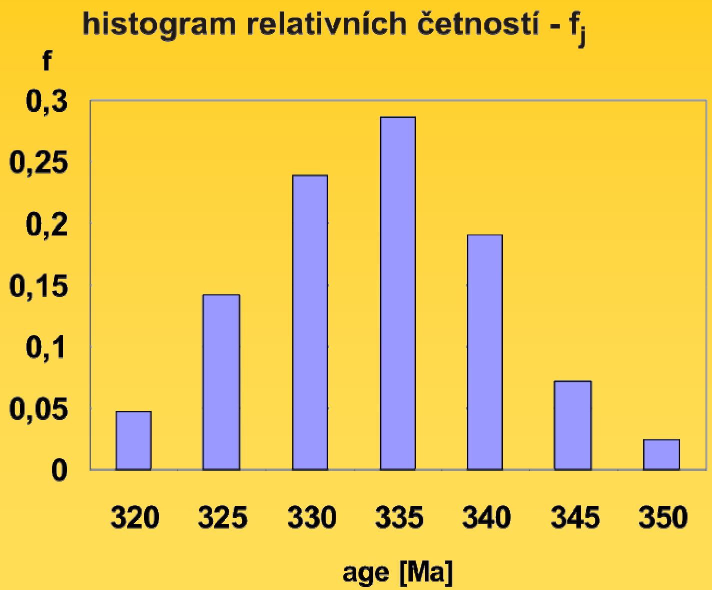
- Normální rozdělení je jednoznačně určeno střední hodnotou a rozptylem, jež jsou jeho parametry. Pokud tedy tyto dvě charakteristiky známe, můžeme určit lehce již vše ostatní - to je tvar celého rozdělení.

- Dobře aproximuje řadu jiných (i diskrétních) pravděpodobnostních rozdělení.

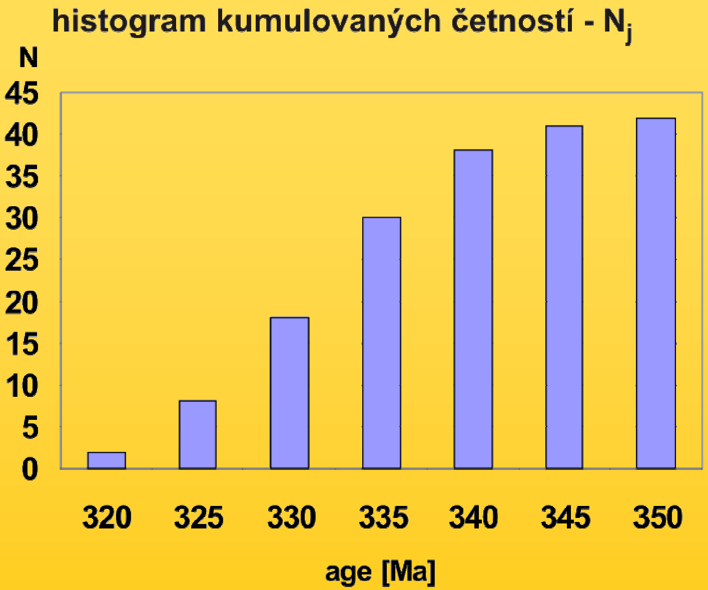
- Při řešení pravděpodobnostních úloh se často předpokládá, že sledovaná náhodná veličina má normální rozdělení, ačkoliv její skutečné rozdělení má jen podobný tvar, tzn. je jednovrcholové a přibližně symetrické. Tento postup je samozřejmě teoreticky podložen, jak dále uvidíme, a je velmi výhodný, neboť usnadňuje teoretické řešení mnoha problémů i praktické výpočty.

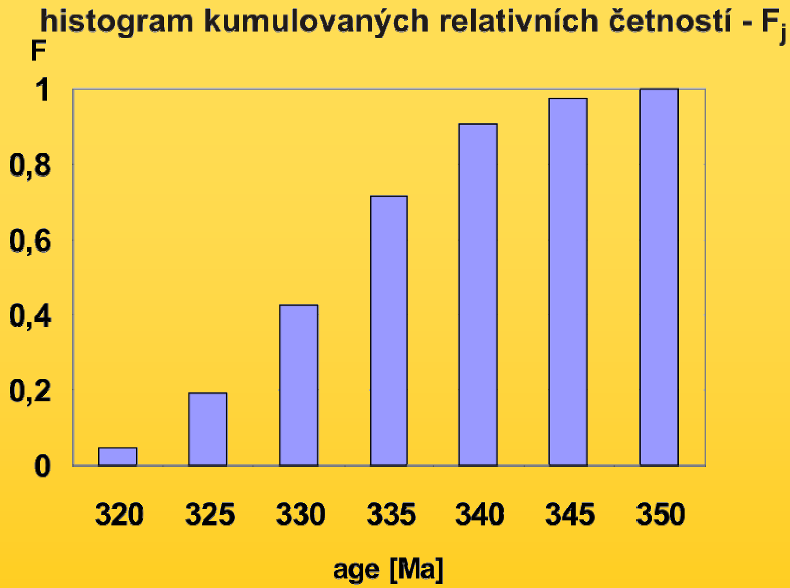
Vztah mezi frekvenční a distribuční funkcí





f(x)





F(x)

pro distribuční fci diskrétní náhodné veličiny platí: $F(x) = P(X \le x)$ a tedy $F(x) = \sum_{x_i \le x} p(x_i)$

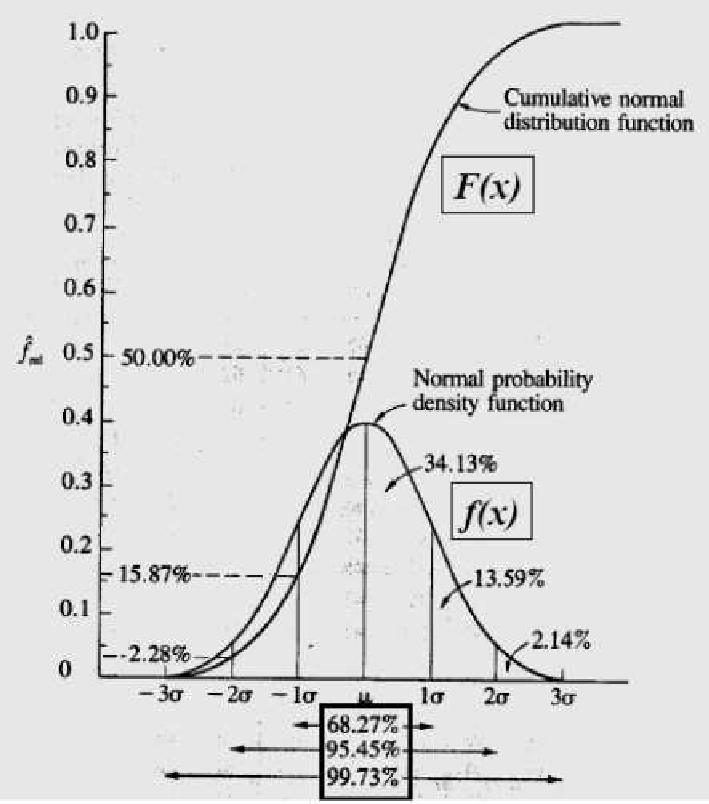
Normální rozdělení

- Distribuční funkce normálního rozdělení – spojitá funkce

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\, \mathrm{d}t$$

- Plocha pod křivkou hustoty má velikost 1. Jinými slovy, integrál z hustoty pravděpodobnosti normálního rozdělení přes celý definiční obor náhodné veličiny je roven jedné.

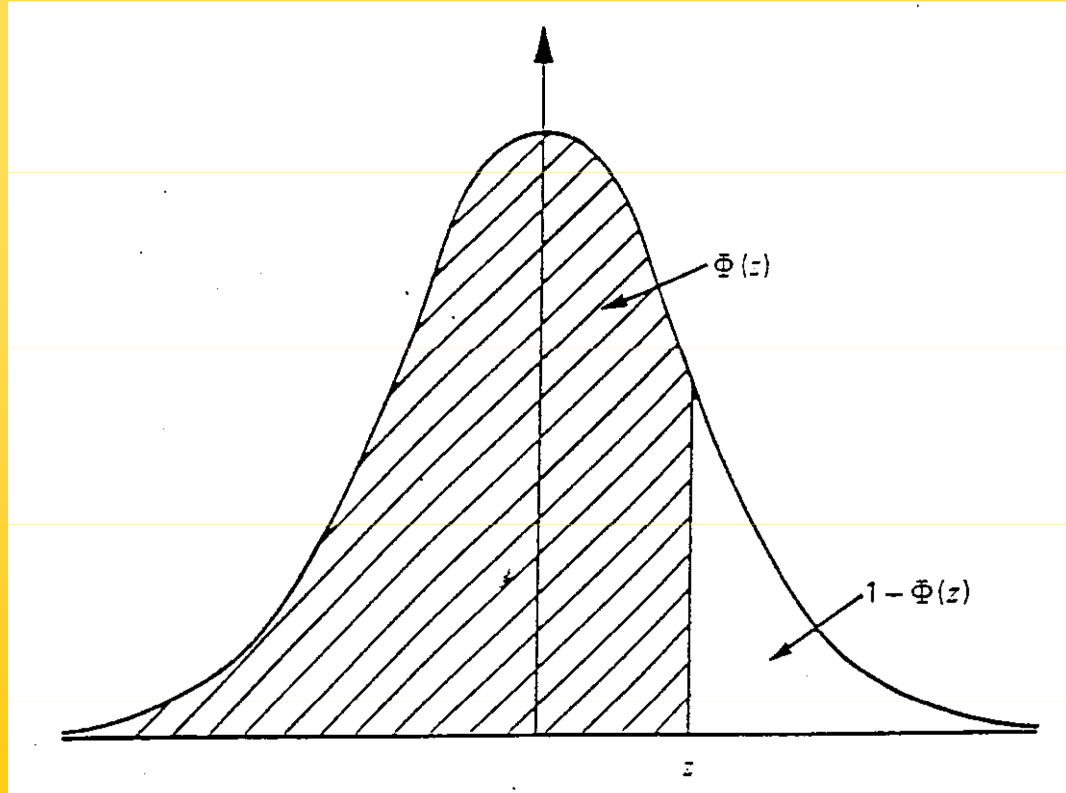
Vztah mezi frekvenční a distribuční funkcí spojité náhodné veličiny



Pro případ normálního rozdělení

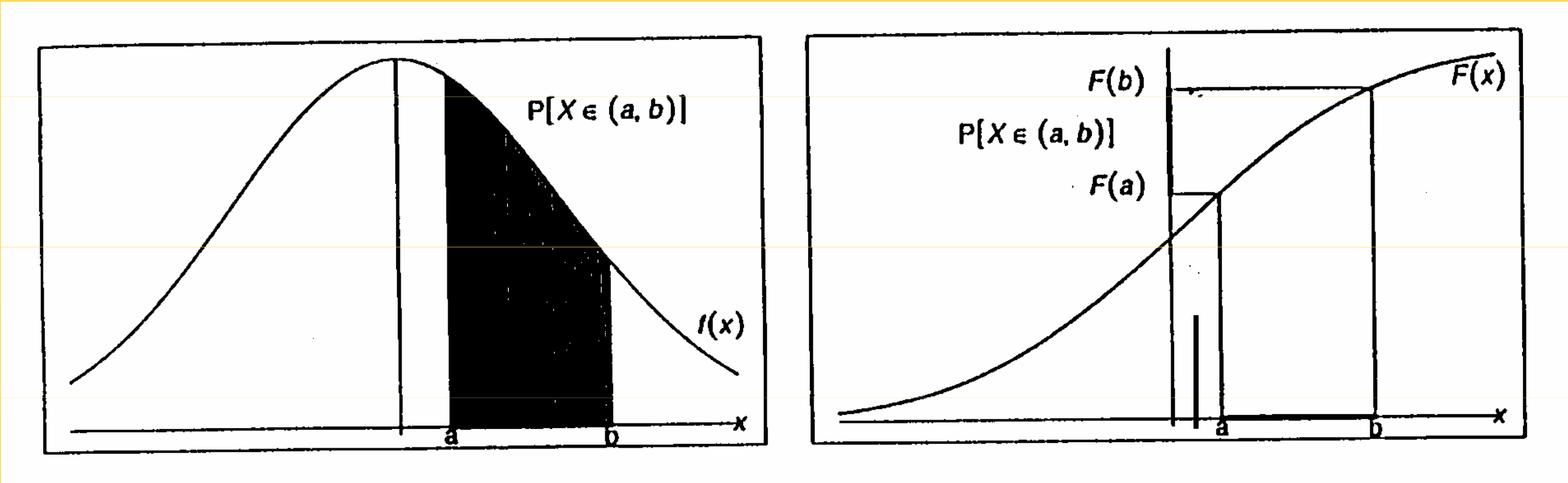
distribuční funkce $F(x)$ se dá vyjádřit

pomocí integrálu: $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$



Velikost vybarvené plochy pod frekvenční funkcí odpovídá hodnotě distribuční funkce *F* v bodě z

Pravděpodobnost P, že náhodná veličina bude mít hodnotu padnoucí do intervalu <A;B> je rovna ploše vyšrafované v grafu frekvenční funkce nebo úseku na ose y pro x z intervalu <A;B> v grafu distribuční funkce.



Co je to distribuční funkce

(cumulative probability)?

$$\boxed{F(x) = P\,(X \le x)}$$

Distribuční funkce v bodě x je rovna pravděpodobnosti jevu, že náhodná veličina X nepřevýší hodnotu x.

Diskrétní n.v.

$$F(x) = \sum_{x_i \le x} P(x_i)$$

Spojitá n.v.

$$F(x) = \int_{-\infty}^{x} f(x)d(x)$$

Transformace normálního rozdělení na standardizované normální rozdělení

- Frekvenční funkce normálního rozdělení

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

- Náhodnou veličinu X s normálním rozdělením N(μ, σ^2) transformuji pomocí tzv. Z-transformace na veličinu se standardním normálním rozdělením, tedy s μ **= 0** a σ^2 **= 1**

Z-transformace $z = (\frac{x-\mu}{\sigma})$

- Normované (nebo standardizované) normální rozdělení má tedy hustotu pravděpodobnosti

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

- Standardizované normální rozdělení se značí N(0, 1)
- Využívá se často při práci s vícerozměrnými soubory dat

V geologii mají normální rozdělení např. tyto náhodné veličiny

- topografický reliéf
- hustota hornin
- obsah hlavních oxidů v horninách
- stanovení stáří hornin