Inovace bakalářského studijního oboru Aplikovaná chemie

http://aplchem.upol.cz

CZ.1.07/2.2.00/15.0247

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
OKRESNÍ HOSPODÁŘSKÁ KOMORA OLOMOUC

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Základy zpracování dat

chemometrie, statistika

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
OKRESNÍ HOSPODÁŘSKÁ KOMORA OLOMOUC
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Inovace bakalářského studijního oboru Aplikovaná chemie

Doporučená literatura

- Pytela O., Chemometrie pro organické chemiky (prodejna skript)
- Hendl J., Přehled statistických metod zpracování dat, Portál, Praha 2004
- Otyepka, Banáš (web: fch.upol.cz - skripta)

Pravděpodobnost elementární základy

Náhodné jevy

- náhodný jev je výsledek náhodného pokusu
- náhodný pokus
  - různé možné výsledky
  - nelze určit, který v daném pokusu získáme
  - pokusy lze opakovat a jednotlivá opakování se neovlivňují

Náhodné jevy

- všechny možné výsledky = prostor náhodných výsledků
  - př. mince „orel“, „panna“
- množina výsledků = náhodný jev
- náhodný jev – jeden výsledek = elementární jev
  - př. kostka „sudá“ – náhodný jev; padnutí dvou ok – elementární jev
- jev jistý = všechny možné výsledky náhodného pokusu
- jev nemožný
- sjednocení jevů (padne 2 nebo 3) $A \cup B$
- současný výskyt (padne sudá a 2) $A \cap B$
- vylučující-se jevy, disjunktní jevy $A \cap B = \text{prázdná}$

Pravděpodobnost

- tzv. statistická definice
- pravděpodobnost náh. jevu *A* je číslo *P*(*A*) ~ relativní četnost jevu *A*

$$\lim_{n\to\infty}\frac{m_A}{n}=P(A)$$

$$P(A)\in\langle 0,1\rangle$$

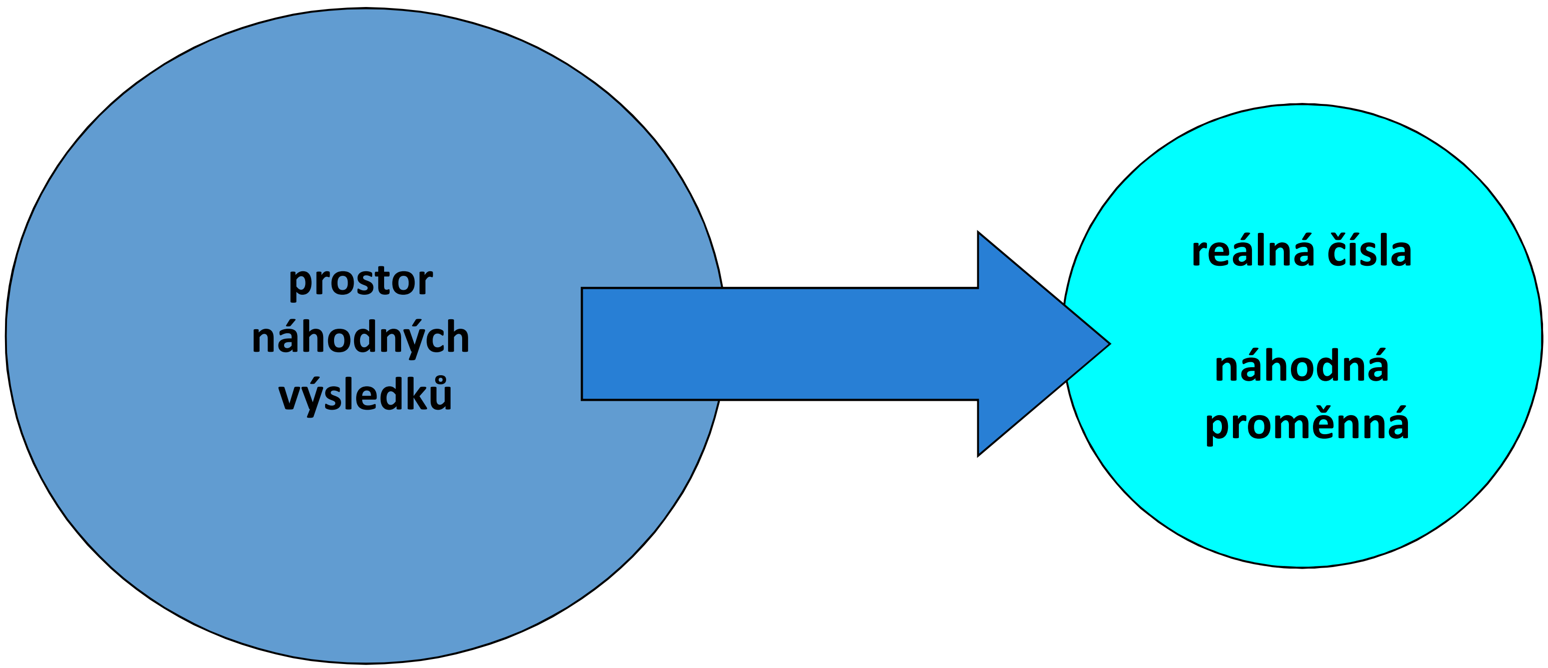
$$P(\Omega)=1$$

$$P(\varnothing)=0$$

$$P(-A)=1-P(A)$$

Náhodná proměnná

- výsledku náh. pokusu přiřadíme číslo → náh. proměnná



Rozdělení náhodné proměnné

- pravděpodobnost, se kterou náhodná proměnná nabývá danou hodnotu (nebo interval) – pravděpodobnostní rozdělení (rozdělení) náhodné proměnné
- náhodná proměnná
  - diskrétní (házení kostkou, ruleta ...)
  - spojitá (měření délek, apod.)

př. rozdělení

- kostka – diskrétní rozdělení

x_i	1	2	3	4	5	6
p_i	1/6	1/6	1/6	1/6	1/6	1/6

- náh. proměnná se řídí daným zákonem rozdělení

Distribuční funkce

- *F*(*x*), nejúplnější popis pravd. chování náh. proměnné *X*

$$F(x) = P(X \leq x),\ x \in (-\infty, \infty)$$

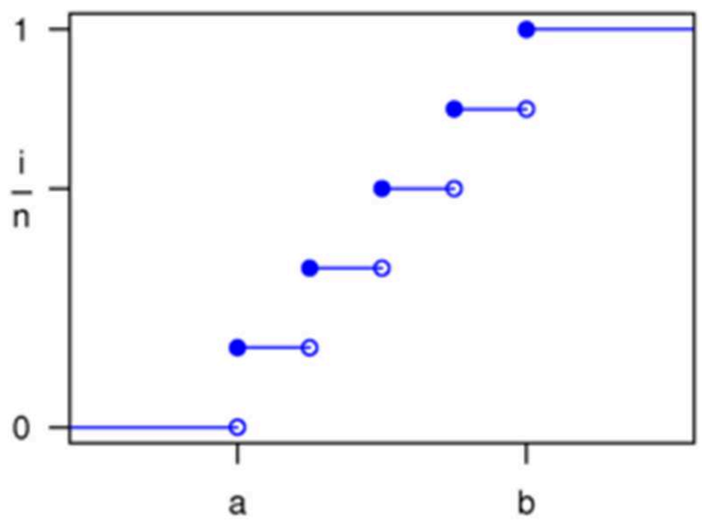
$$F(x) \in \langle 0, 1 \rangle$$

$$\lim_{x \to -\infty} F(x) = 0$$

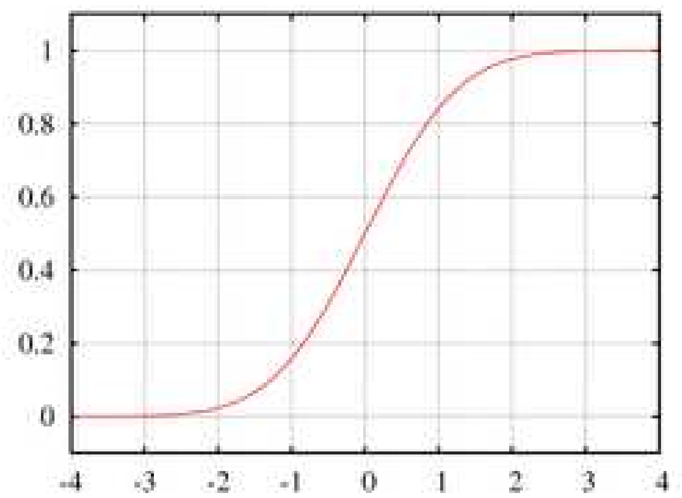
$$\lim_{x \to \infty} F(x) = 1$$

neklesající, nemusí být spojitá

Distribuční funkce



diskrétní



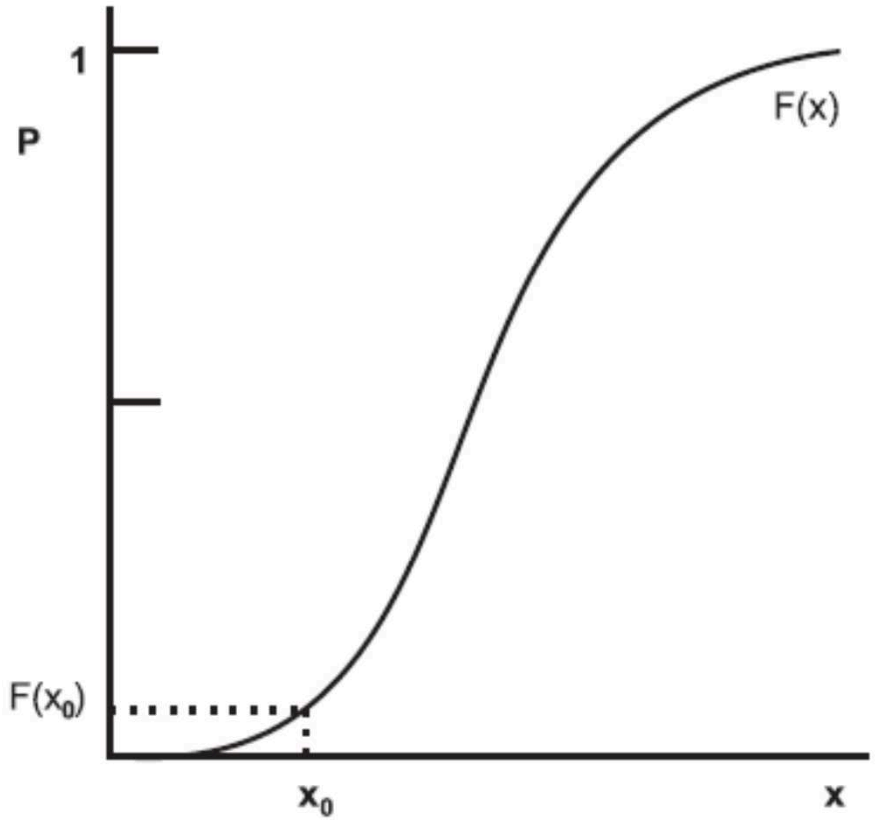
spojité

Hustota pravděpodobnosti

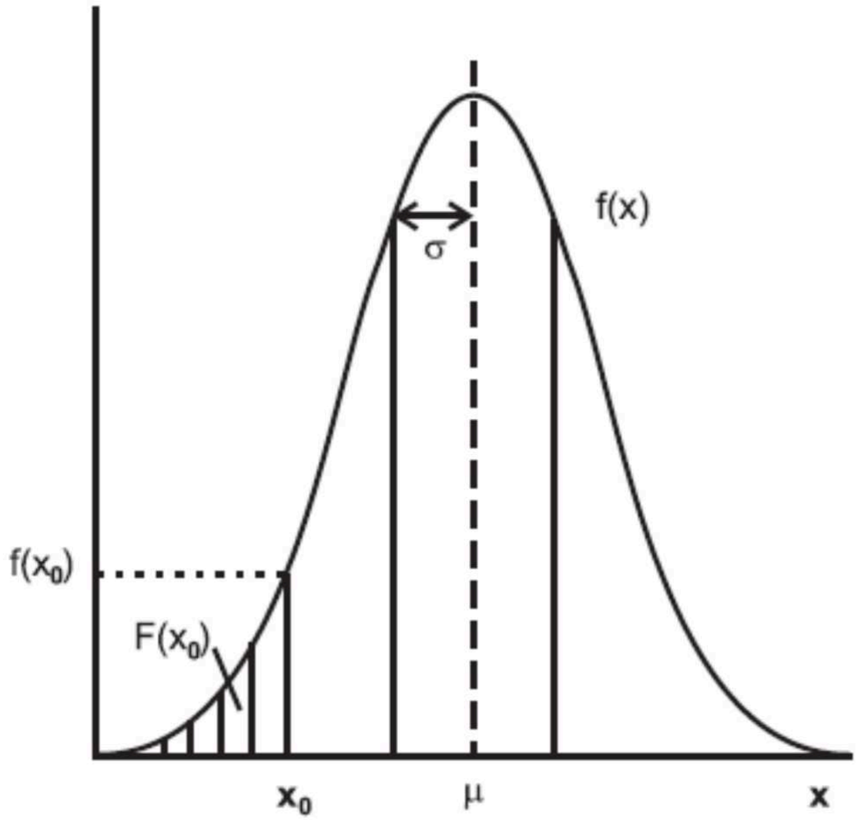
- má-li $F(x)$, pro všechna x derivaci, pak derivace $\mathrm{d}F(x)/\mathrm{d}x$ je hustota pravděpodobnosti $f(x)$
  - někdy se nazývá frekvenční funkce
  - rozdělení s hustotou jsou vždy spojitá!

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

Hustota pravděpodobnosti

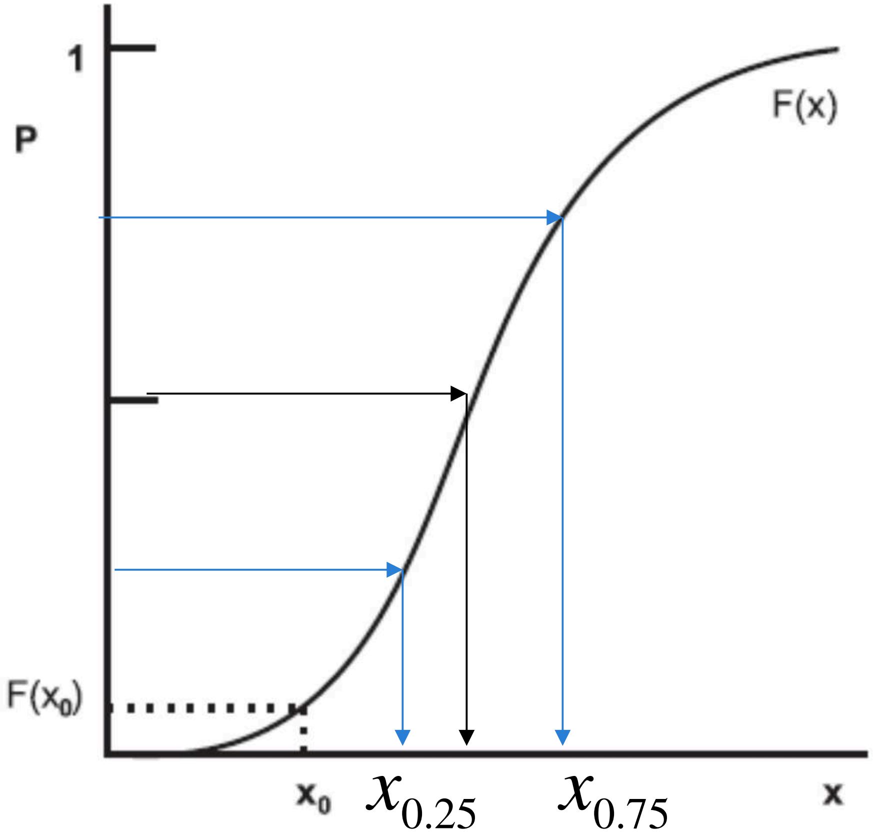


Distribuční funkce



Hustota pravděpodobnosti
frekvenční funkce

Kvantil



$x_{0.25}$ $x_{0.75}$

$x_{0.5}$
medián

horní kvartil

dolní kvartil

$$Q = F^{-1}$$

kvantilová funkce

Normální rozdělení

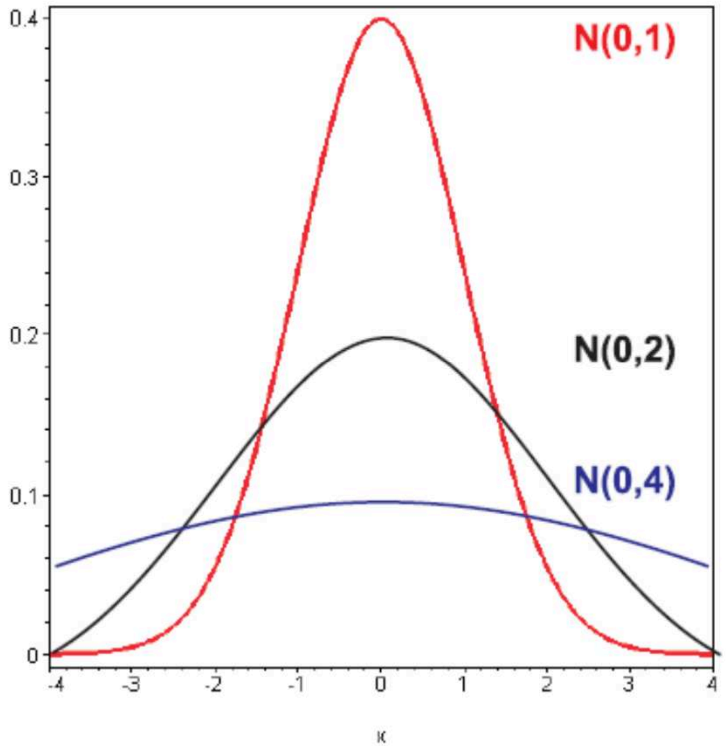
- též Gaussovo rozdělení

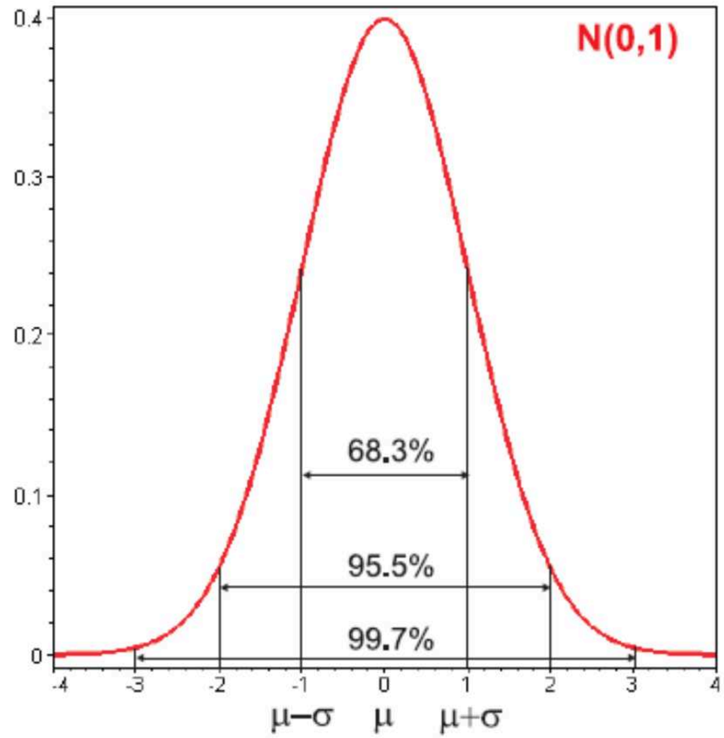
$$N\left(\mu,\sigma^2\right)$$

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

$$X \sim N\left(\mu,\sigma^2\right)$$

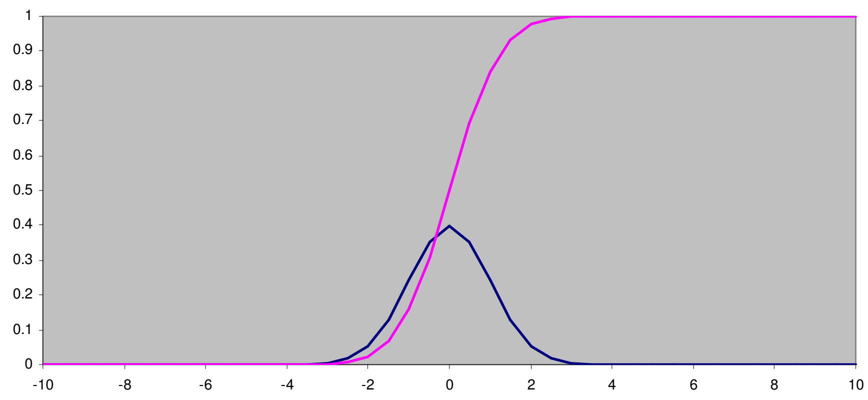
Normální rozdělení





Normální rozdělení a EXCEL

- cvičení – EXCEL
  - *pomocí EXCELu vytvořte graf frekvenční a distribuční funkce normálního rozdělení, sledujte chování na zvolených parametrech μ a σ^2*
  - *nápověda: funkce statistické, NORMDIST*
  - *jaký je rozdíl mezi NORMDIST a NORMSDIST?*
  - *nalezněte v EXCELU příslušné kvantilové funkce.*



Obecné a centrální momenty

- střední hodnota = 1. ob. moment

$$E(X) = \int_{-\infty}^{\infty} x \mathrm{d}F_x(x) = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

$$\mu_k' = E(X)^k$$

$$\operatorname{var}(X) = E(X - E(X))^2$$

$$\mu_k = E(X - E(X))^k$$

$$N(\mu, \sigma^2)$$

$$E(x) = \mu$$

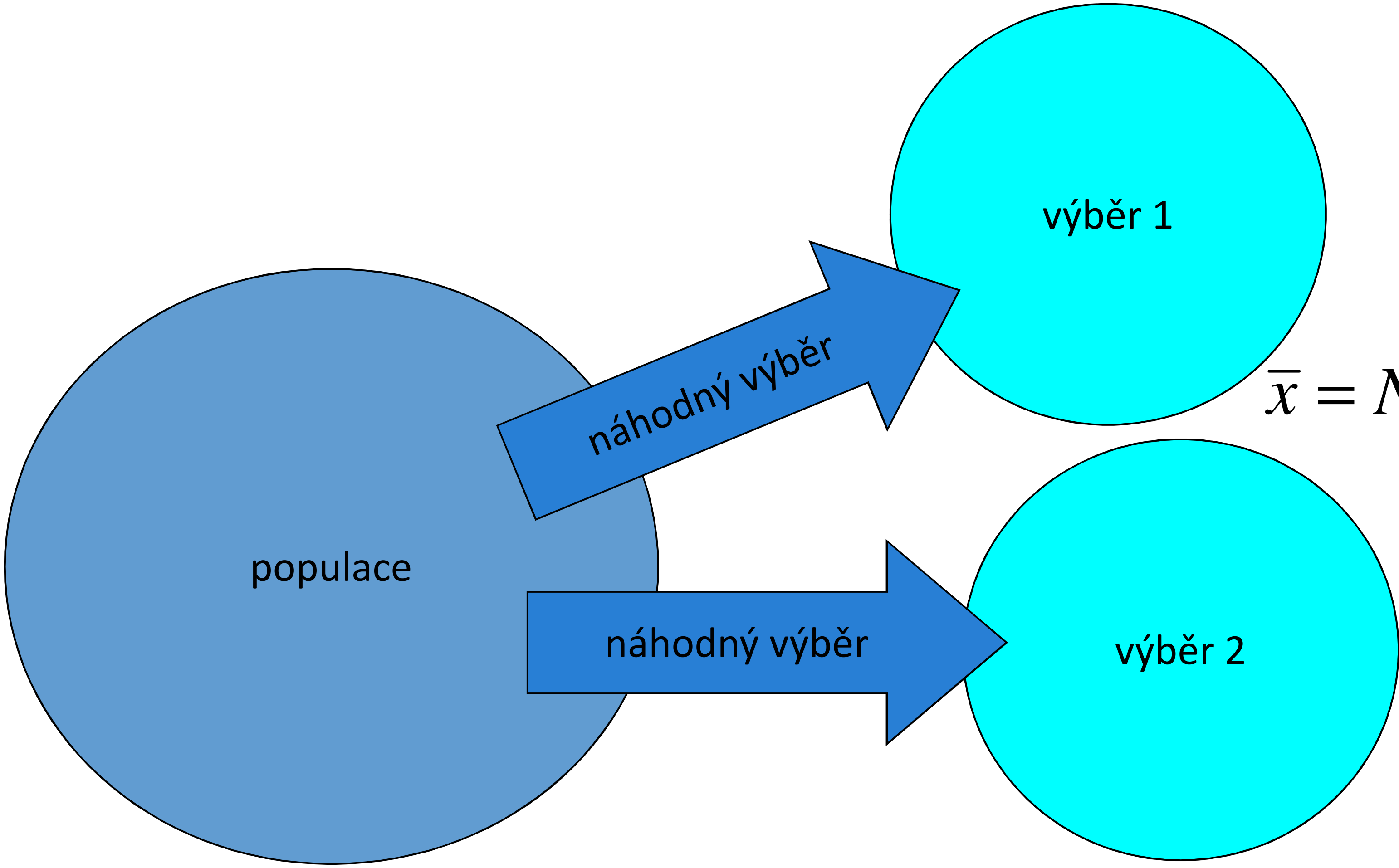
$$\operatorname{var}(x) = \sigma^2$$

Důležité vlastnosti

$$\begin{aligned} E(\alpha) &= \alpha \\ E(\alpha X + \beta) &= \alpha E(X) + \beta \\ E(X + Y) &= E(X) + E(Y) \end{aligned}$$

$$\begin{aligned} var(X) &= E(X^2) - (E(X))^2 \\ var(\alpha X + \beta) &= \alpha^2 var(X) \end{aligned}$$

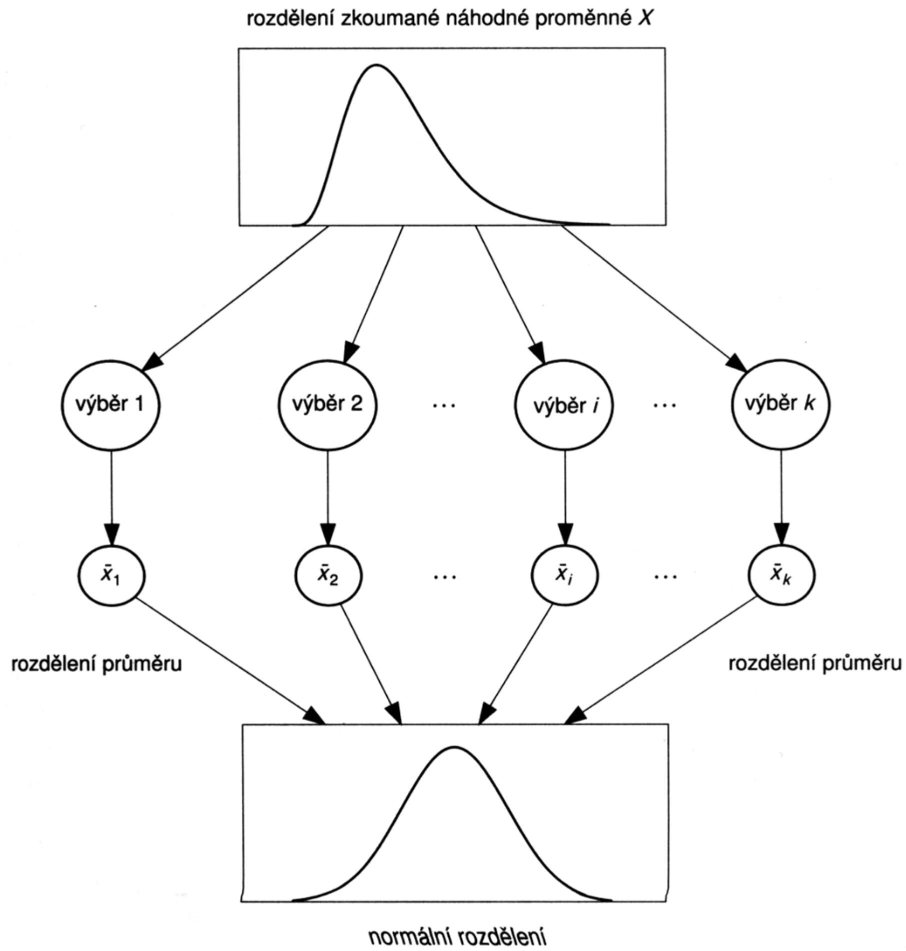
Výběrové rozdělení stř. hod.



$$\overline{x} = N\left(\mu, \frac{\sigma^2}{n}\right)$$

Centrální limitní teorém

- významné postavení norm. rozdělení



Normování

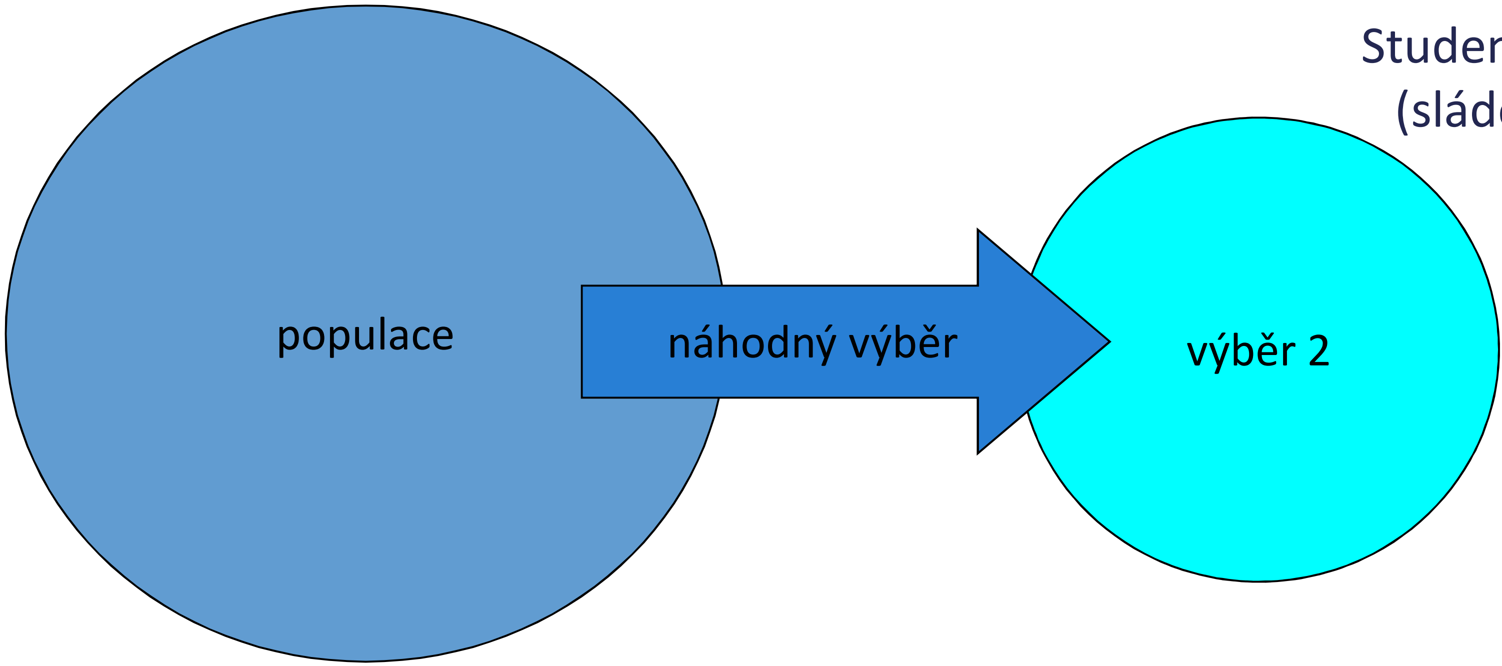
$$N(\mu, \sigma^2) \rightarrow N(0,1)$$

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

Výběrové rozdělení stř. hod.
neznáme σ

$$\frac{E(X)-\mu}{s_{E(X)}} = \frac{E(X)-\mu}{s/\sqrt{n}} \sim t(\nu = n-1)$$

Studentovo rozdělení
Student = W. S. Gosset
(sládek u Guinesse)



Výběrové rozdělení rozptylu

- chí-kvadrát χ^2
  - EXCEL = CHIDIST, CHIINV
- Poměr rozptylů
- F-rozdělení (Fisherovo)
  - EXCEL = FDIST, FINV

Kovariance, korelace

- kovariance

$$cov(X,Y) = E(X - E(X))(Y - E(Y))$$

$$cov(X,X) = var(X)$$

$$var(X + Y) = var(X) + var(Y) + 2cov(X,Y)$$

- korelace

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

$$|\rho_{X,Y}| \leq 1, \rho_{X,X} = 1$$