Náhodná proměnná

Náhodná proměnná může mít rozdělení

- diskrétní ($x_1, x_2, \ldots, x_n$)
- spojité ($<x_1; x_2>$)

Poznámky:

1. Fyzikální veličiny jsou zpravidla spojité, ale změřené hodnoty jsou diskrétní.
2. Pokud je rozptyl měření podstatně větší než nejmenší rozlišitelná hodnota, lze měřenou veličinu považovat za spojitou (Meloun & Militký, 2013).

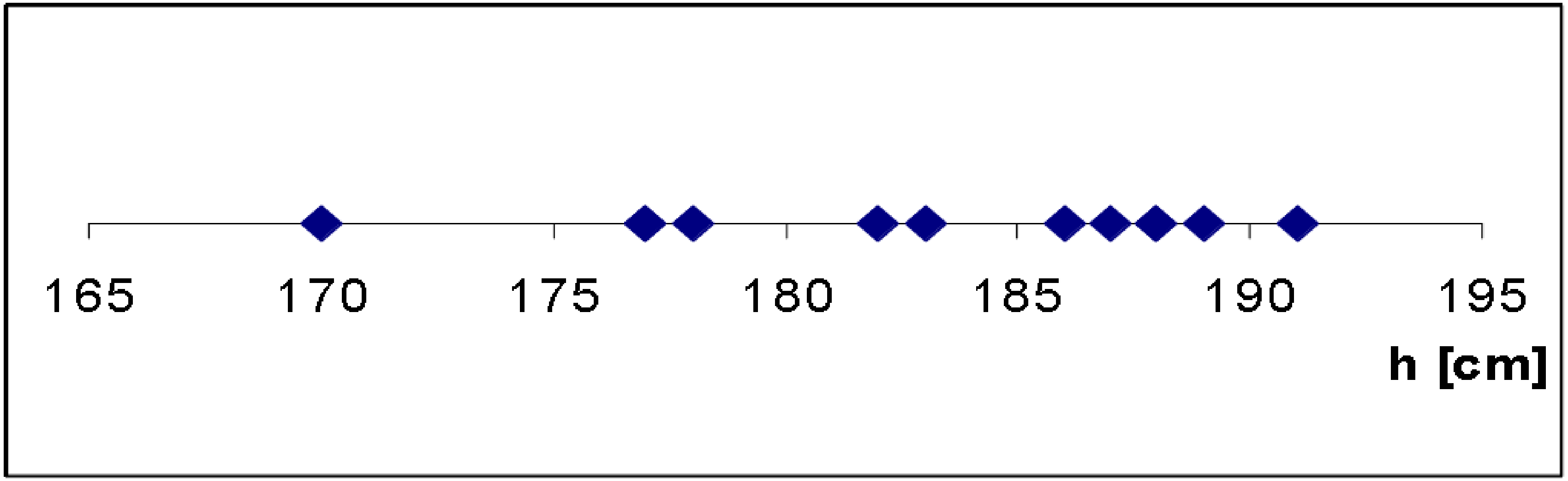
Meloun, M., Militký, J. (2013). *Kompendium statistického zpracování dat*. Praha: Karolinum.

Diagram rozptýlení

Neroztříděná data - výšky 12 studentů FT [cm]:

177 170 182 183 186 188 191 177 189 178 187 188

Diagram rozptýlení (Neubauer et al., 2012)



Používá se pro malý počet hodnot (do 30).

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

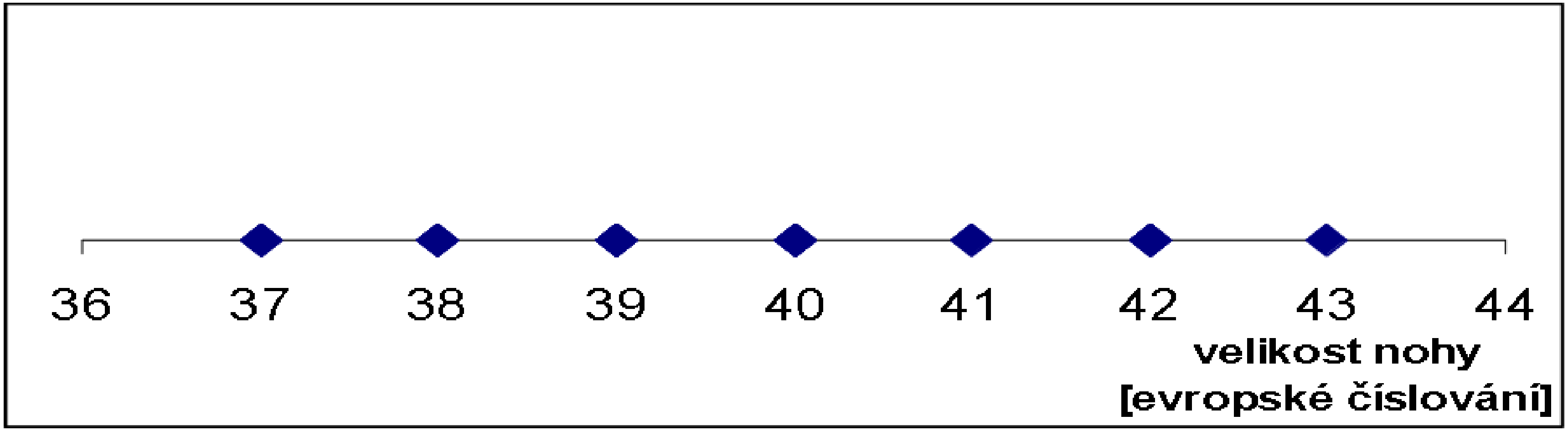
Diagram rozptýlení – nová data

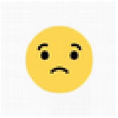
Velikost nohy 32 studentek FT [evropské číslování]

40 38 40 42 40 37 40 39 37 42 39 38 39 38 41 38

42 39 42 43 40 41 41 37 39 38 40 39 37 41 40 41

Diagram rozptýlení (Neubauer et al., 2012):



Tudy cesta nevede 

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Četnost – bodové rozdělení

Velikost nohy 32 studentek FT [evropské číslování]

40 38 40 42 40 37 40 39 37 42 39 38 39 38 41 38

42 39 42 43 40 41 41 37 39 38 40 39 37 41 40 41

Hodnota znaku x_i	*Absolutní četnost n_i*	*Relativní četnost p_i*	*Absolutní kumulativní četnost N_i*	*Relativní kumulativní četnost P_i*
37	4	0,125	4	0,125
38	5	0,156	9	0,281
39	6	0,188	15	0,469
40	7	0,219	22	0,688
41	5	0,156	27	0,844
42	4	0,125	31	0,969
43	1	0,031	32	1,000
Σ	**32**	**1,000**		

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

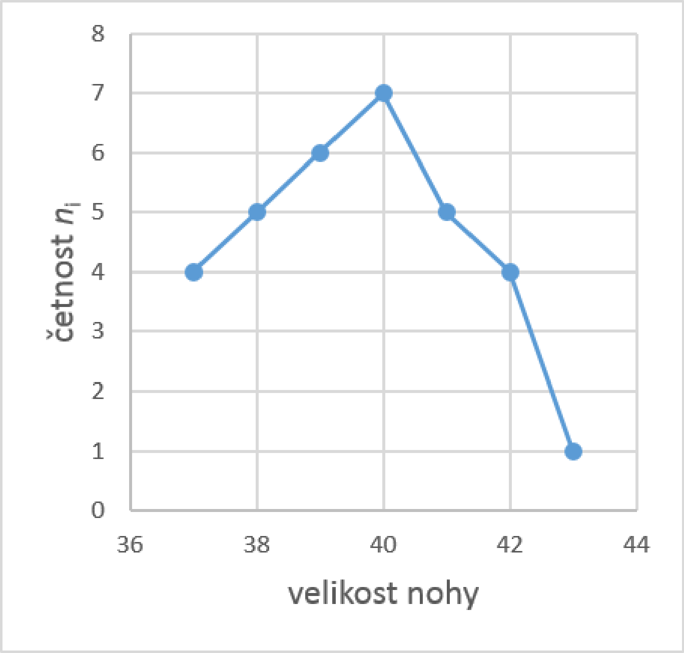
Četnost – bodové rozdělení

Velikost nohy 32 studentek FT [evropské číslování]

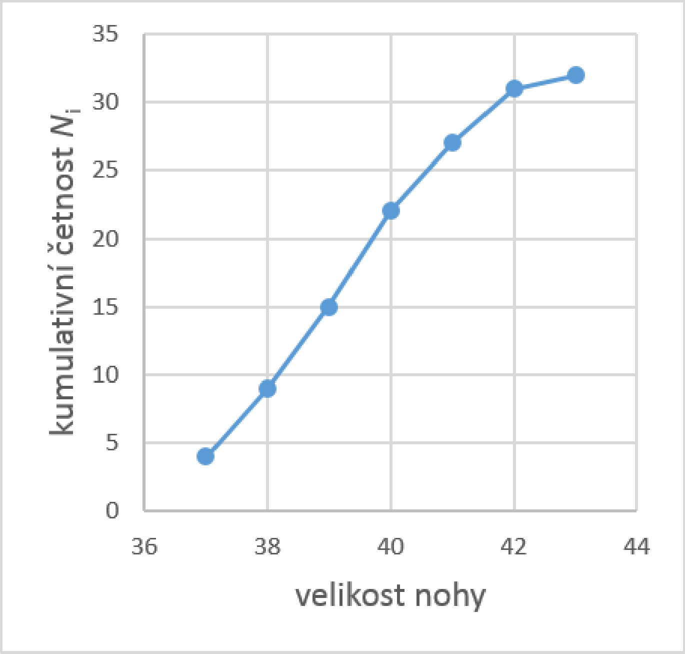
40 38 40 42 40 37 40 39 37 42 39 38 39 38 41 38

42 39 42 43 40 41 41 37 39 38 40 39 37 41 40 41

Polygon četností
(Neubauer et al., 2012)



Součtová křivka
(Neubauer et al., 2012)



Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Četnost – bodové rozdělení

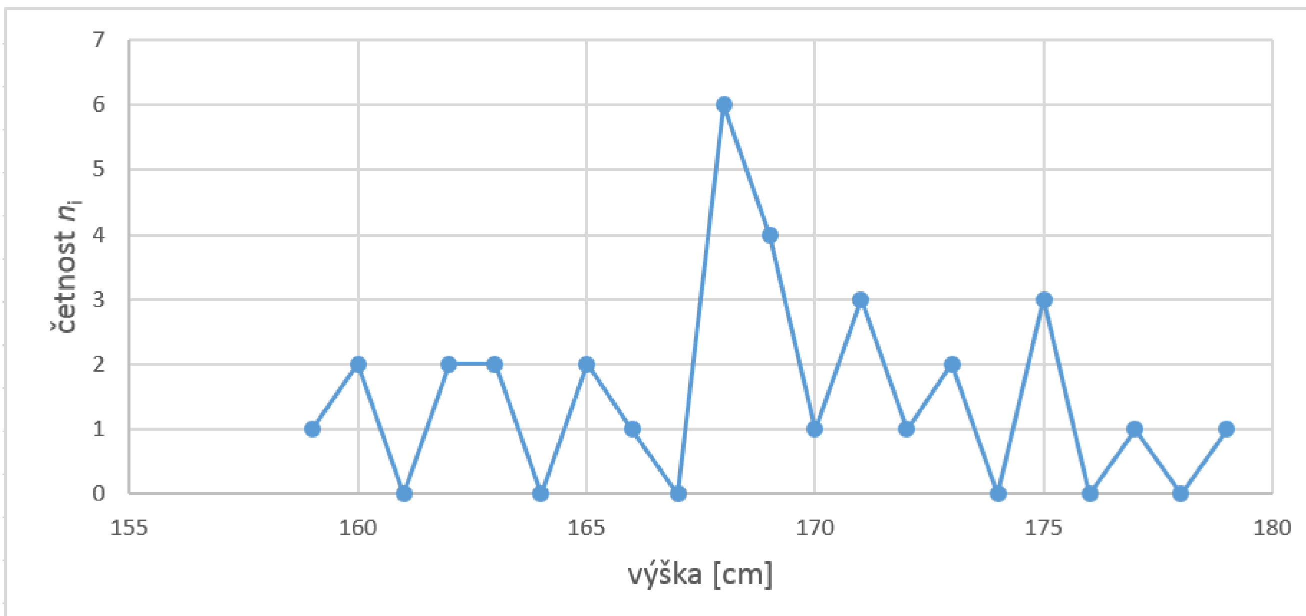
Výška 32 studentek FT [cm]:

172 169 163 171 165 169 169 165 168 175 163

168 175 169 168 171 168 163 168 173 159 168

162 163 162 179 160 177 173 160 175 170 171

polygon četností
(Neubauer et al., 2012):



Tudy cesta nevede

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Četnost - intervalové rozdělení

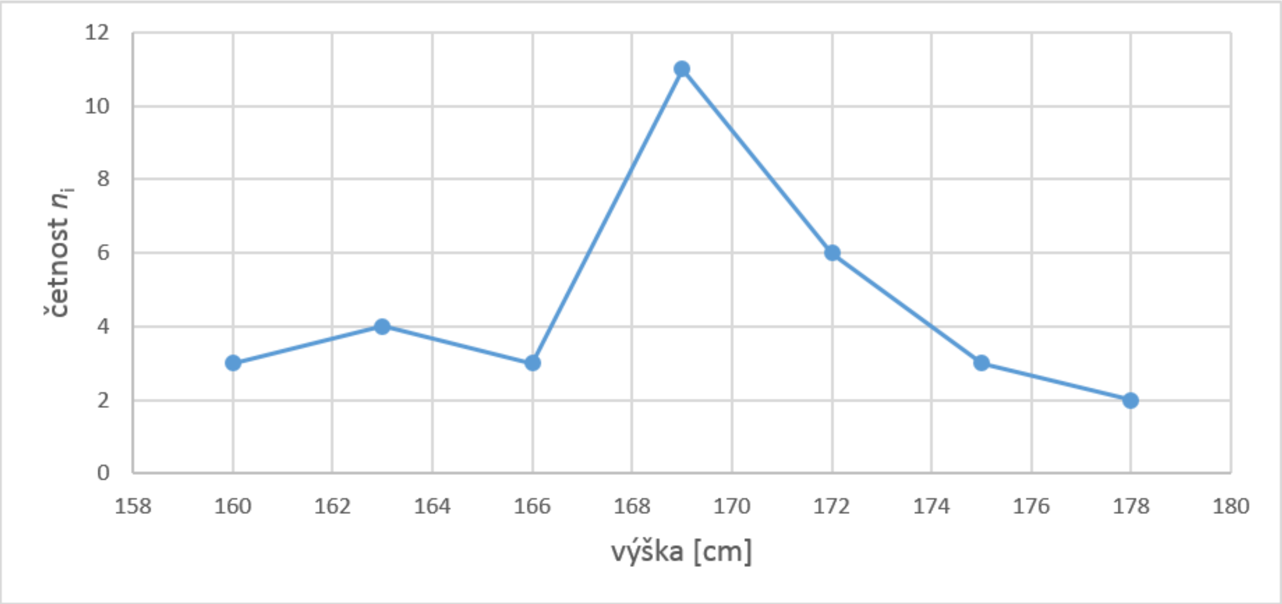
Výška 32 studentek FT [cm]:

172 169 161 171 165 169 169 165 168 175 163

168 175 169 168 171 168 163 168 175 159 168

162 163 162 179 160 177 169 160 175 170 171

polygon četností
(Neubauer et al., 2012):



doporučený počet tříd $k \approx 2{,}46\left(N-1\right)^{0{,}4}$ nebo $k \approx \sqrt{N}$

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Četnost absolutní a relativní

U spojité veličiny je pravděpodobnost výskytu dané hodnoty x nekonečně malá. $P(x) \rightarrow 0$

Je proto lepší zjišťovat pravděpodobnost výskytu hodnoty x v nějakém intervalu od x_1 do x_2. **ozn. $P(x_1, x_2)$**

Počet výsledků spadajících do daného intervalu nazýváme (absolutní) **četnost Q**.

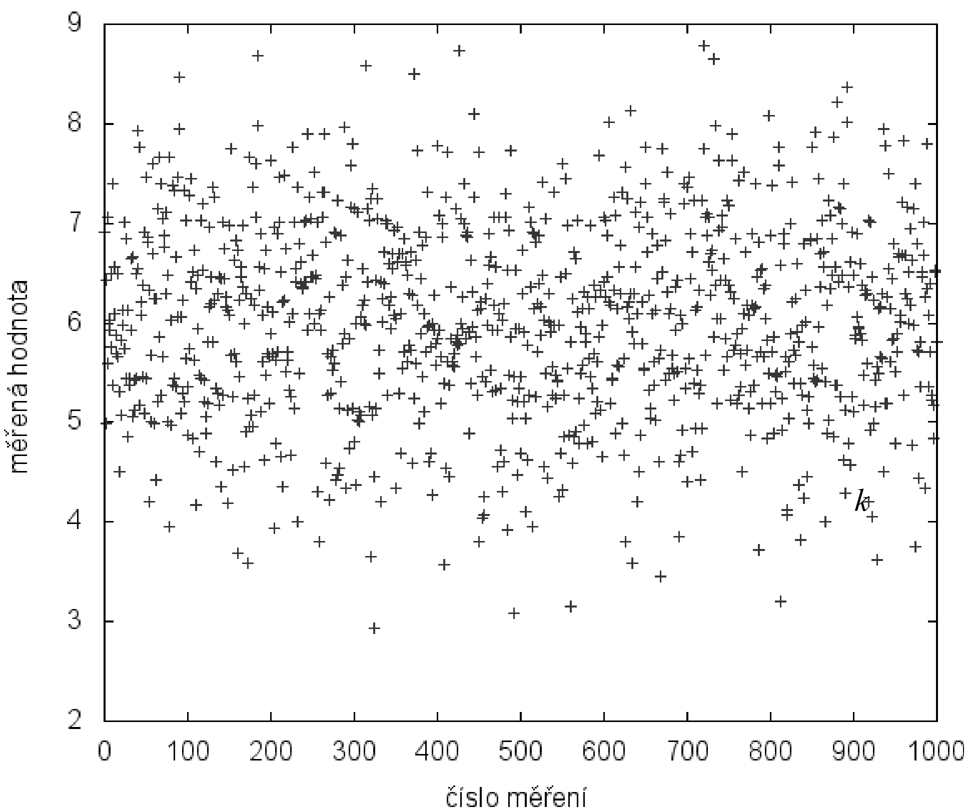
Výhodnější je používat **relativní četnost q = Q/N**, kde N je počet měření

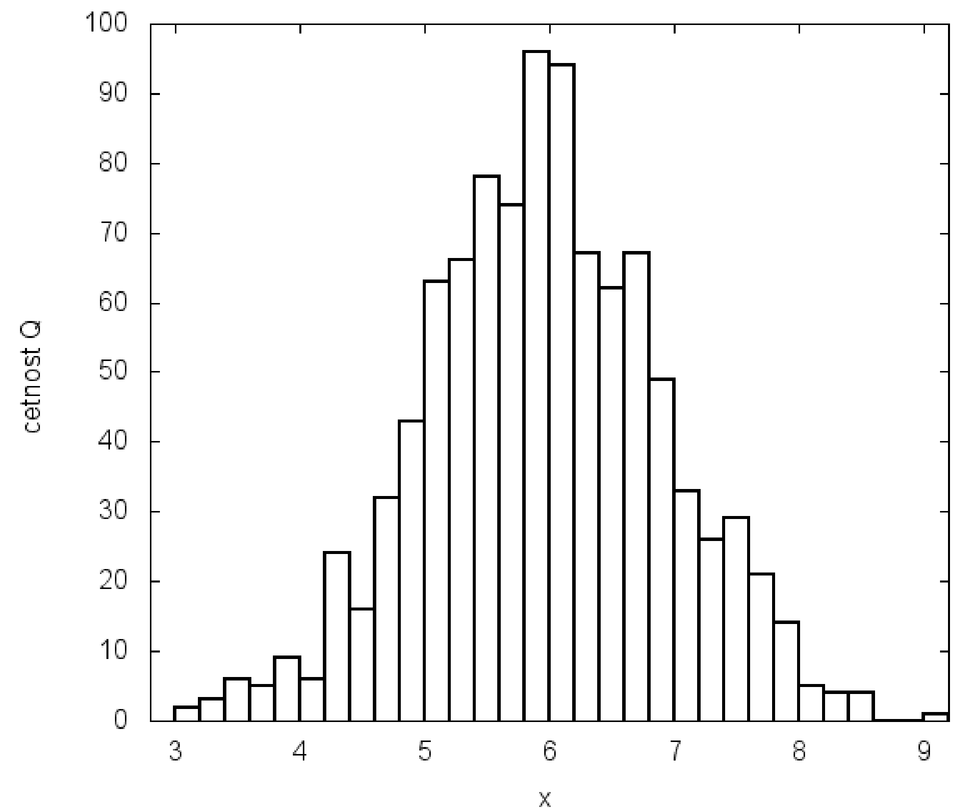
(Anděl, 2011).

Anděl, J. (2011). *Základy matematické statistiky*. Praha: MATFYZPRESS.

Histogram

Četnosti měření graficky znázorníme **histogramem** (Budíková et al.,2010).



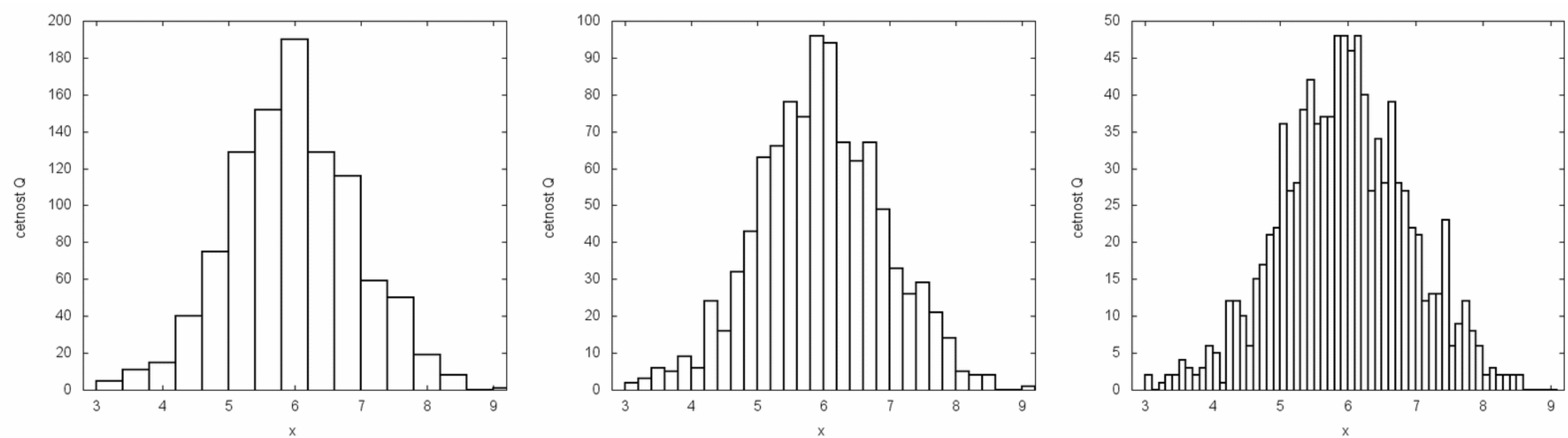


doporučený počet tříd histogramu

$$k \approx 2{,}46\left(N-1\right)^{0{,}4} \text{ nebo } k \approx \sqrt{N}$$

Budíková, M., Králová, M., Maroš, B. (2010). *Průvodce základními statistickými metodami.* Praha: GRADA.

Histogram - počet tříd

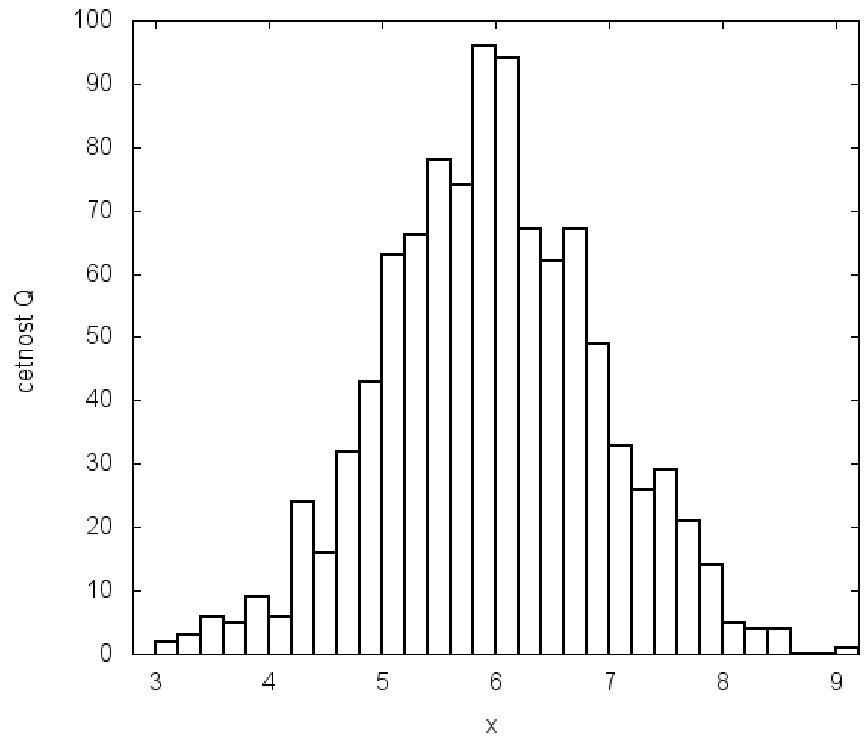


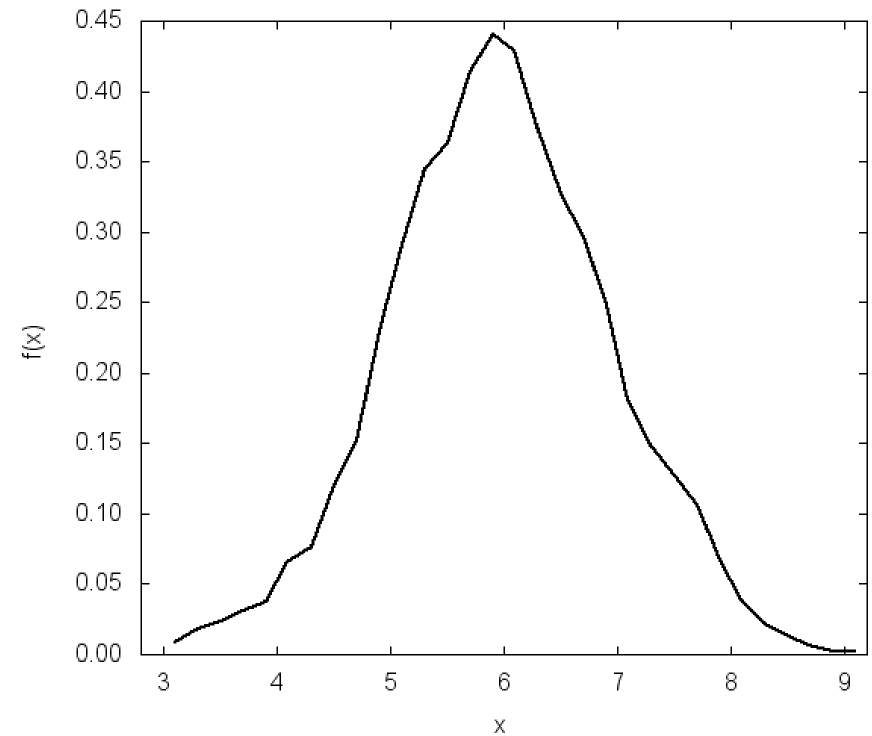
doporučený počet tříd histogramu (Budíková et al.,2010):

$$k \approx 2{,}46\left(N-1\right)^{0{,}4} \ \text{nebo} \ k \approx \sqrt{N}$$

Budíková, M., Králová, M., Maroš, B. (2010). *Průvodce základními statistickými metodami.* Praha: GRADA.

Hustota pravděpodobnosti





$$f(x) = \lim_{\Delta x \to 0} \frac{P(x, x + \Delta x)}{\Delta x}$$

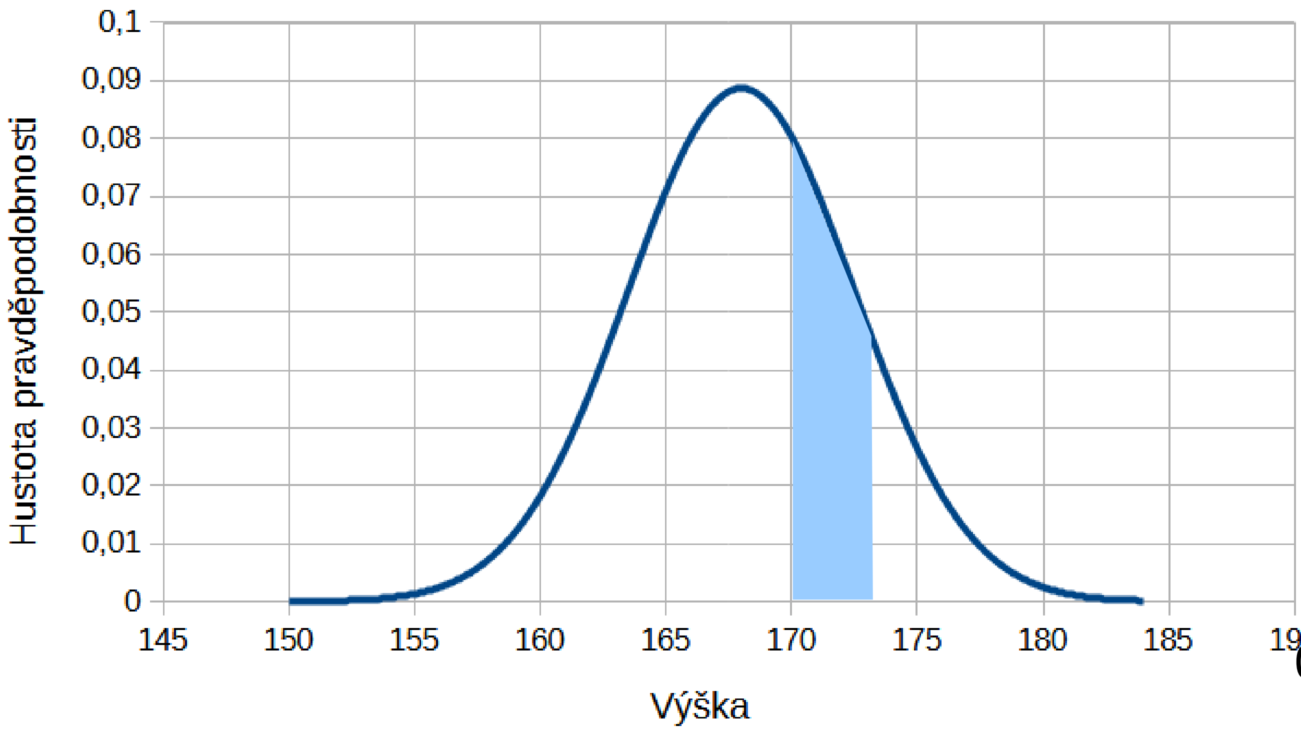
$$P(a, b) = \int_a^b f(x)\,\mathrm{d}x$$

(Neubauer et al., 2012)

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

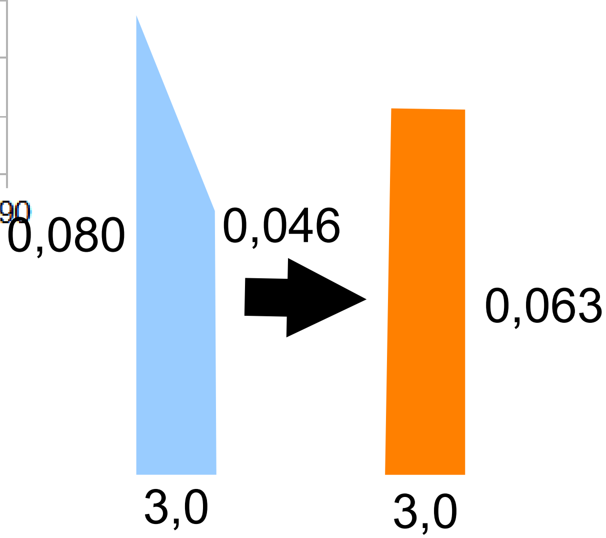
Hustota pravděpodobnosti

Výšky 20 studentek, seřazené [cm]:
158; 163; 163; 165; 166; 167; 167; 167; 168; 169;
169; 170; 170; 170; 170; 172; 173; 174; 174; 177



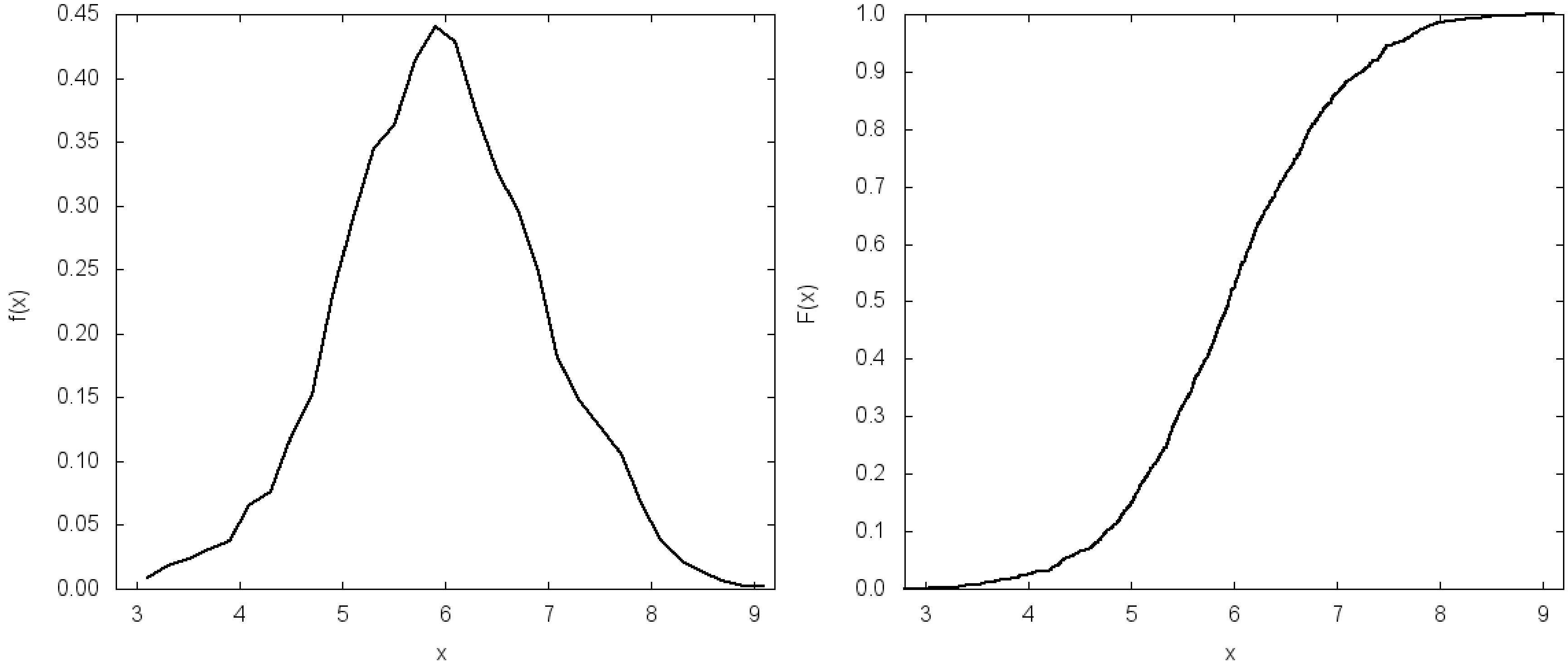
Pravděpodobnost odpovídá ploše pod křivkou.

Kolik výšek bude mezi 170 a 173 cm?



3 * 0,063 = 0,19 = 19 %
19 % z 20 je 3,8 -> asi 4 výšky

Distribuční funkce



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

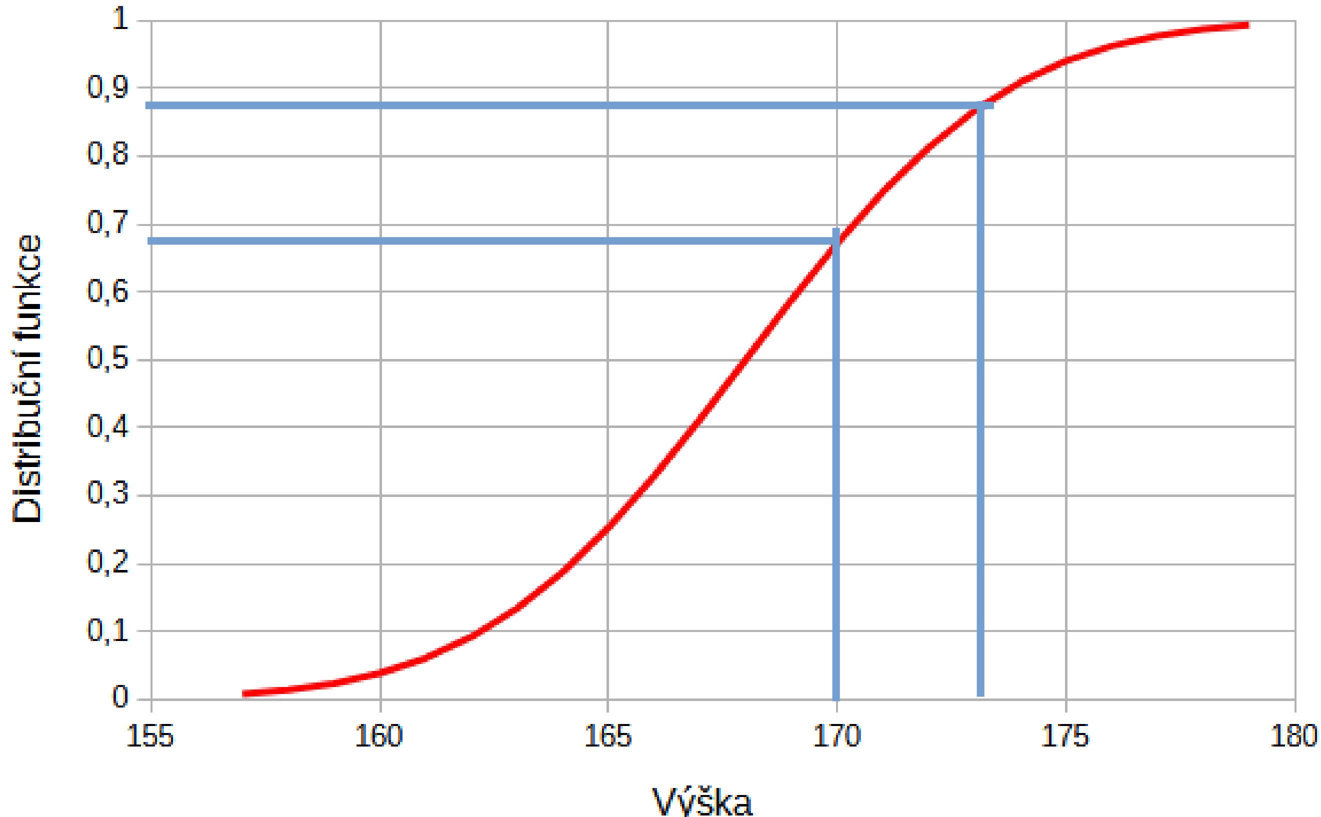
Funkce hustoty pravděpodobnosti nebo distribuční funkce v sobě nesou kompletní informaci o náhodném rozdělení!

(Neubauer et al., 2012).

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Distribuční funkce

Výšky 20 studentek, seřazené [cm]:
158; 163; 163; 165; 166; 167; 167; 167; 168; 169;
169; 170; 170; 170; 170; 172; 173; 174; 174; 177

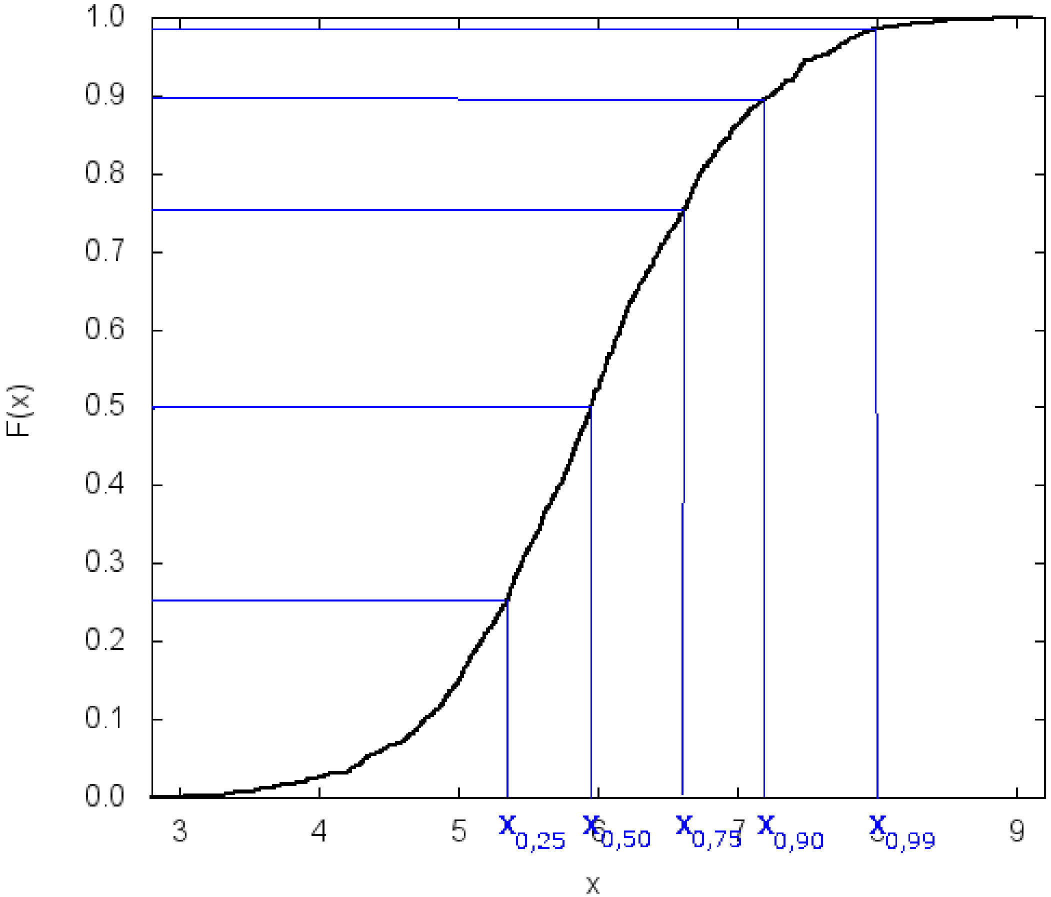


Kolik výšek bude mezi 170 a 173 cm?

Menších než 170 je 69 % výšek.
Menších než 173 je 88 % výšek.
Mezi 170 a 173 je 88-69=19 % výšek.

Kvantily

Kvantil x_p je hodnota znaku, pro kterou platí, že 100p % jednotek uspořádaného souboru má hodnotu menší nebo rovnu x_p (Neubauer et al., 2012).



25% kvantil $x_{0,25}$ - dolní kvartil

75% kvantil $x_{0,75}$ - horní kvartil

50% kvantil $x_{0,50}$ - **medián**

90% kvantil $x_{0,90}$ - 9. decil

99% kvantil $x_{0,99}$ - 99. percentil

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Aritmetický průměr

Aritmetický průměr **diskrétní** náhodné veličiny *x*

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i$$

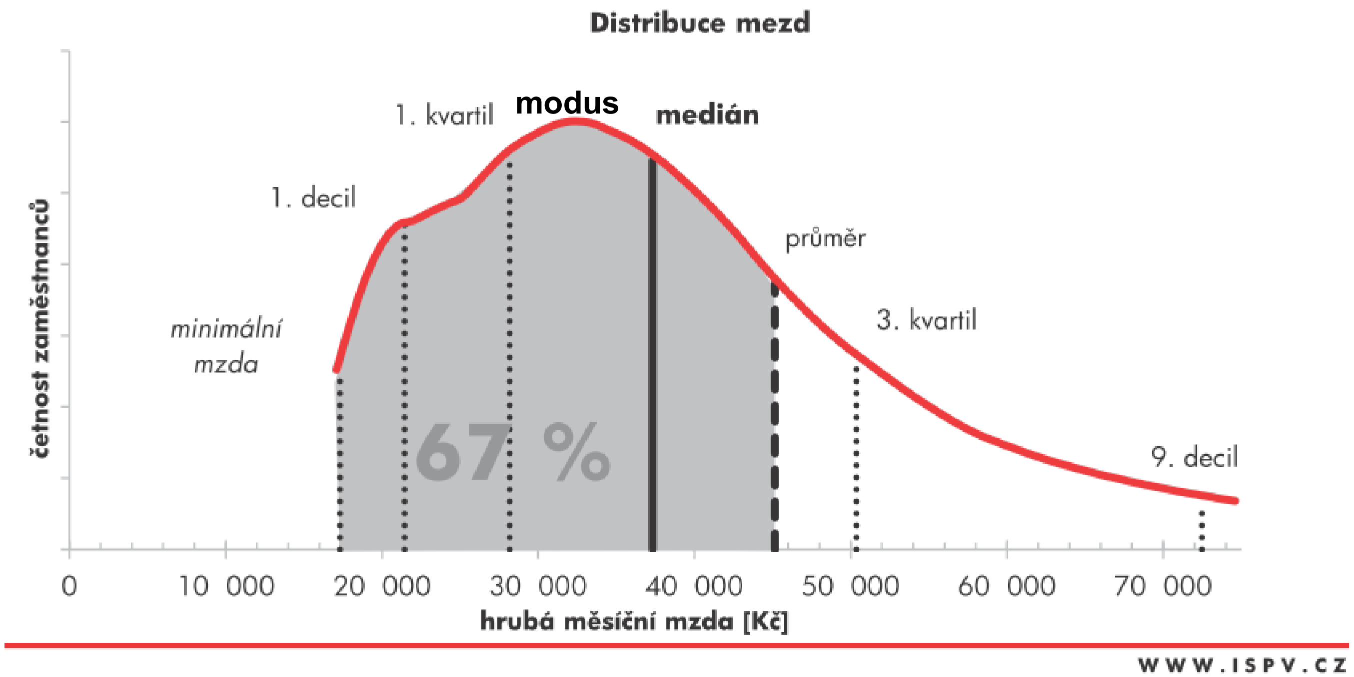
Ukážeme dále, že nejlepším odhadem střední hodnoty μ je aritmetický průměr $E(x)$
(Meloun & Militký, 2013)

$$\mu = E(x)$$

Meloun, M., Militký, J. (2013). *Kompendium statistického zpracování dat*. Praha: Karolinum.

Srovnání různých charakteristik polohy

Distribuce mezd 1. pololetí 2023

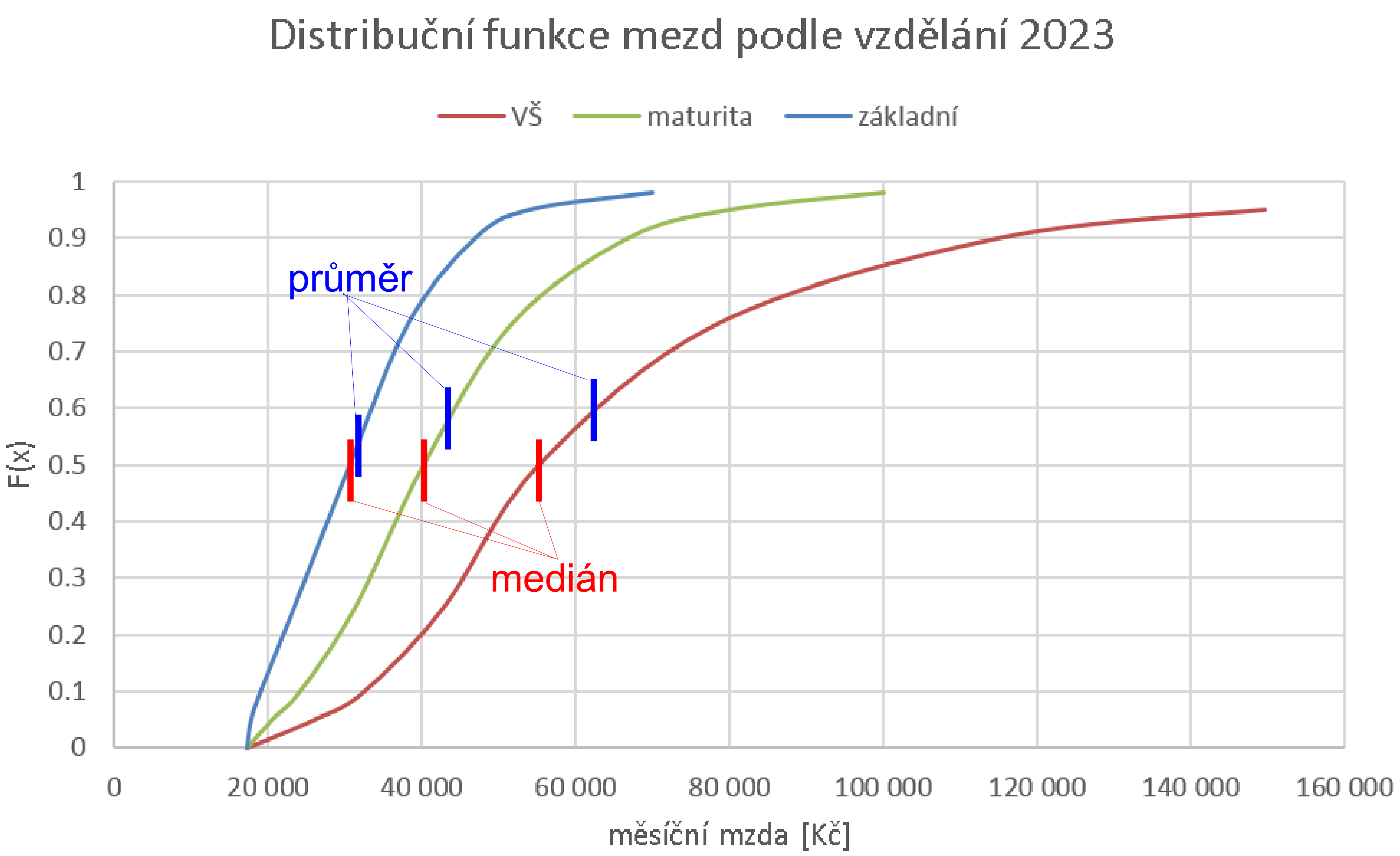


Aritmetický průměr
- součet hodnot vydělený jejich počtem.

Medián
- kvantil $x_{0,50}$
- stejně hodnot pod mediánem jako nad ním.

Modus
- nejpravděpodobnější hodnota.

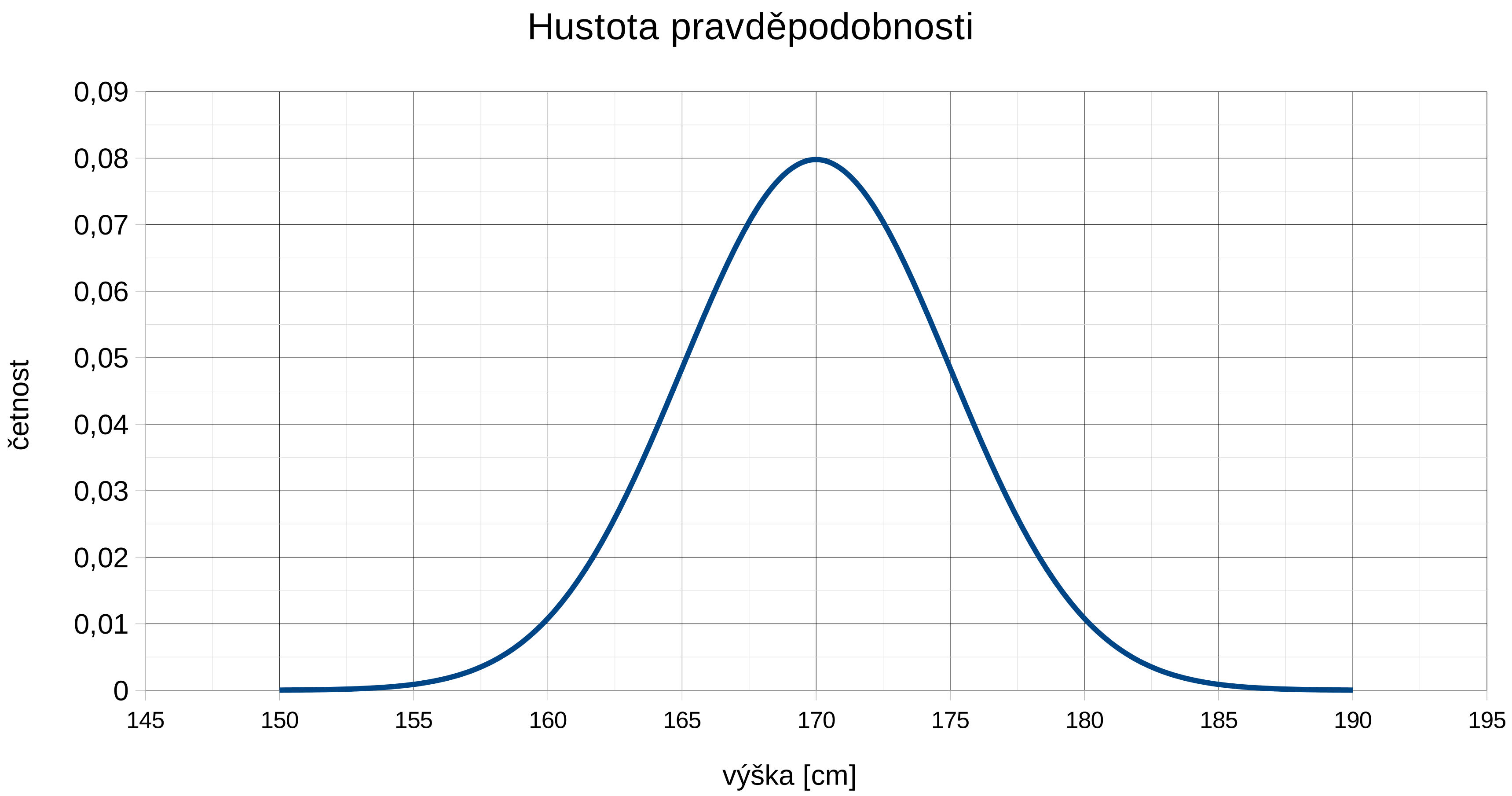
Distribuční funkce - mzdy



http://csu.gov.cz

Procvičování

Hustota pravděpodobnosti výšky studentek



Odhadněte:

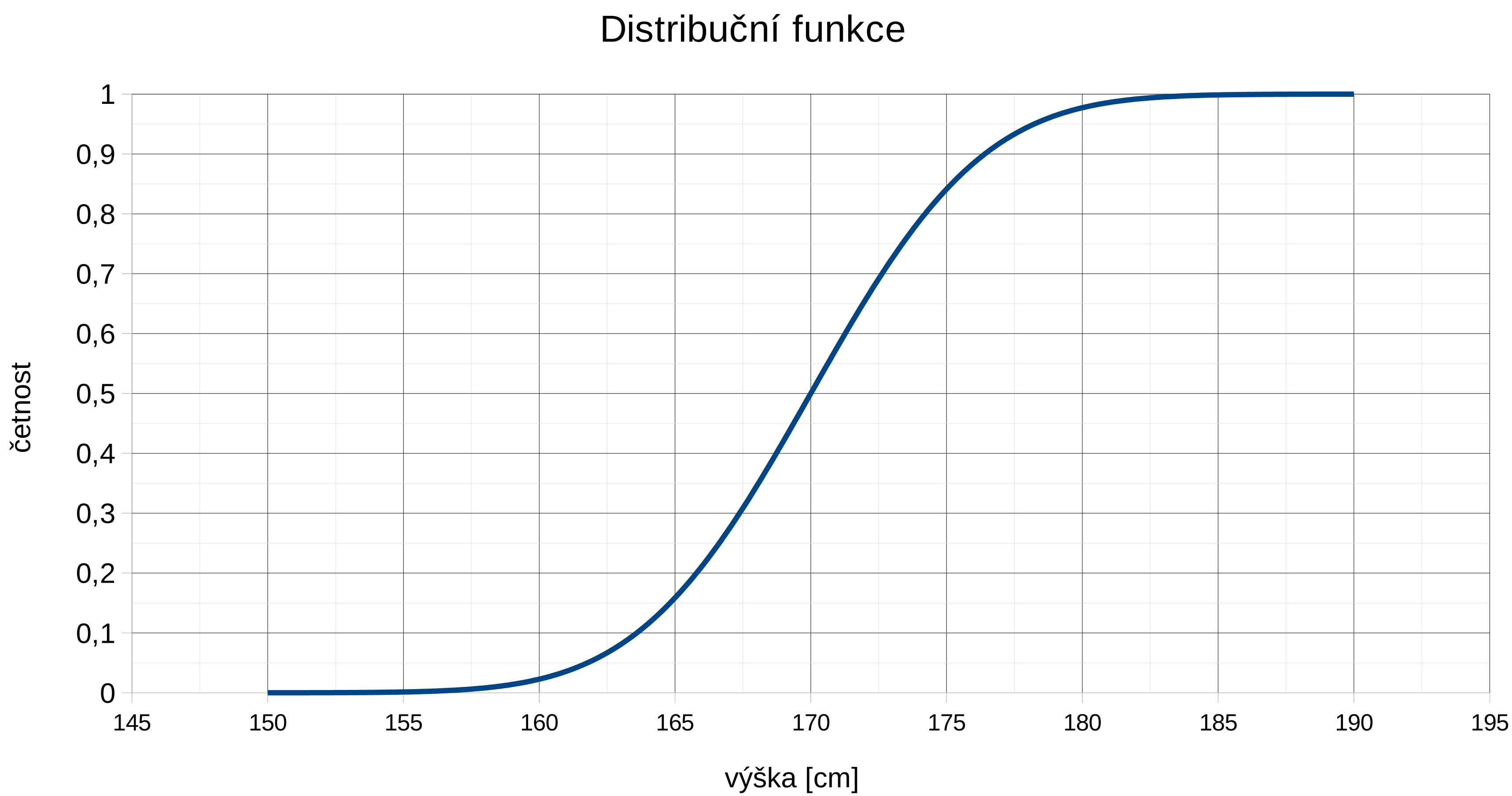
a) Medián výšky studentek.

b) Kolik studentek má výšku mezi 160 a 165 cm.

c) Kolik studentek má výšku mezi 170 a 172 cm.

Procvičování

Distribuční funkce výšky studentek



Odhadněte:

a) Medián výšky studentek.

b) Kolik studentek má výšku mezi 160 a 165 cm.

c) Dolní kvartil výšky studentek.

d) 6. decil výšky studentek.

e) 35. percentil výšky studentek.

Různé charakteristiky variability

Variabilita popisuje rozsah souboru.

Různé charakteristiky variability:

- rozdíl maximální a minimální hodnoty
- kvartilové rozpětí (rozdíl horního a dolního kvartilu)
- decilové rozpětí (rozdíl devátého a prvního decilu)
- percentilové rozpětí (rozdíl 99. a 1. percentilu)
- ...

(Meloun & Militký, 2013)

Neubauer, J., Sedlačík, M., Kříž, O. (2012). *Základy statistiky Aplikace v technických a ekonomických oborech*. Praha: GRADA.

Rozptyl a směrodatná odchylka

Nejčastěji používanou charakteristikou variability (tj. míry odchylky jednotlivých hodnot od střední hodnoty) je **rozptyl** náhodné veličiny *x*.

Pro **diskrétní** náhodnou veličinu

$$D(x) = \frac{1}{N}\sum_{i=1}^{N}\{x_i - E(x)\}^2$$

Odmocnina z rozptylu se nazývá **směrodatná odchylka** – ozn. σ

$$\sigma = \sqrt{D(x)}$$

(Meloun & Militký, 2013).

Meloun, M., Militký, J. (2013). *Kompendium statistického zpracování dat*. Praha: Karolinum.

Různé charakteristiky koncentrace

Kromě střední hodnoty charakterizující polohu rozdělení a směrodatné odchylky charakterizující variabilitu (šířku) rozdělení existují i **charakteristiky koncentrace** informující o tvaru rozdělení.

Budeme používat

1. koeficient šikmosti (**šikmost**) a
2. koeficient špičatosti (**špičatost**).

Šikmost informuje o **souměrnosti** rozdělení.

Špičatost informuje o **koncentraci prostředních hodnot**.

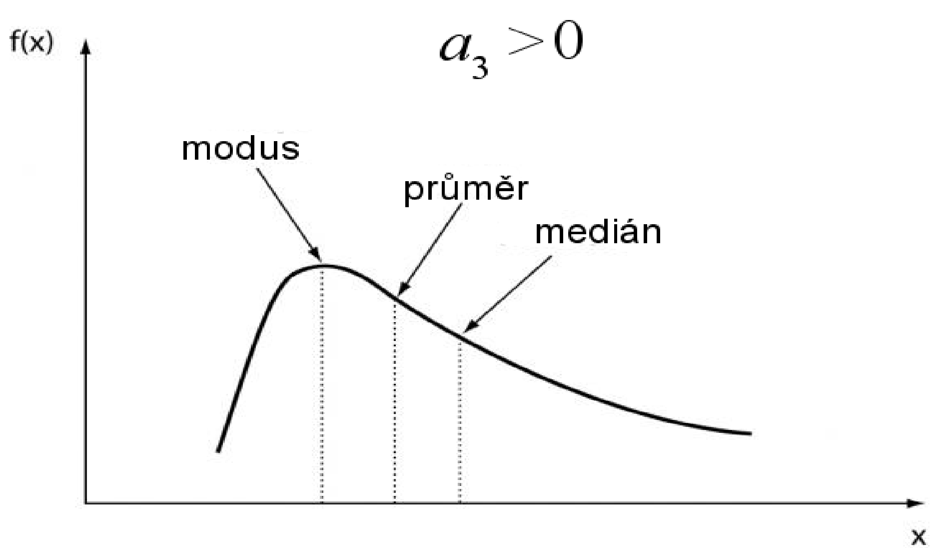
Meloun, M., Militký, J. (2013). *Kompendium statistického zpracování dat*. Praha: Karolinum.

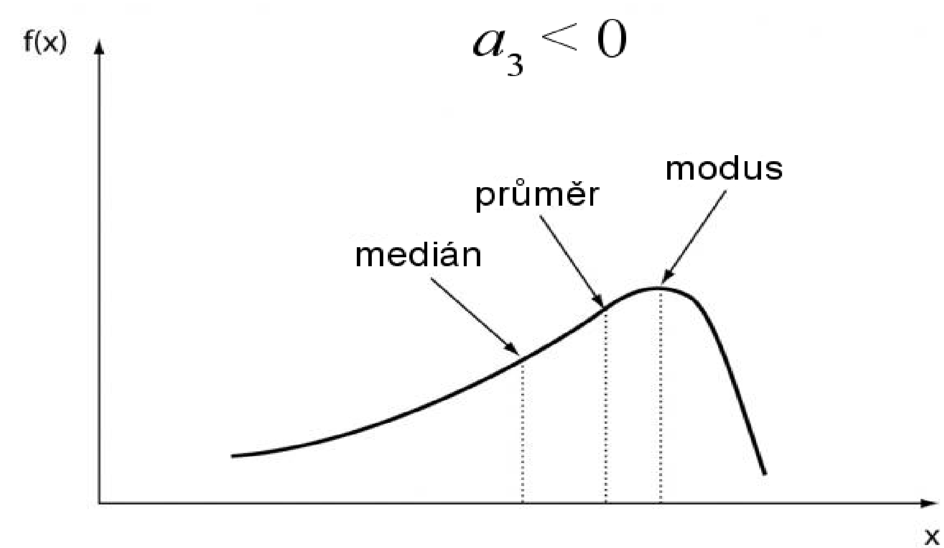
Šikmost

Šikmost je charakteristika rozdělení náhodné veličiny, která popisuje jeho symetrii (Meloun & Militký, 2013).

Šikmost a_3 je definována vztahem:

$$a_3 = \frac{\sum_{i=1}^{n} (x_i - \mu)^3}{N \sigma^3}$$





pro $a_3 = 0$ je rozdělení symetrické

Meloun, M., Militký, J. (2013). *Kompendium statistického zpracování dat*. Praha: Karolinum.

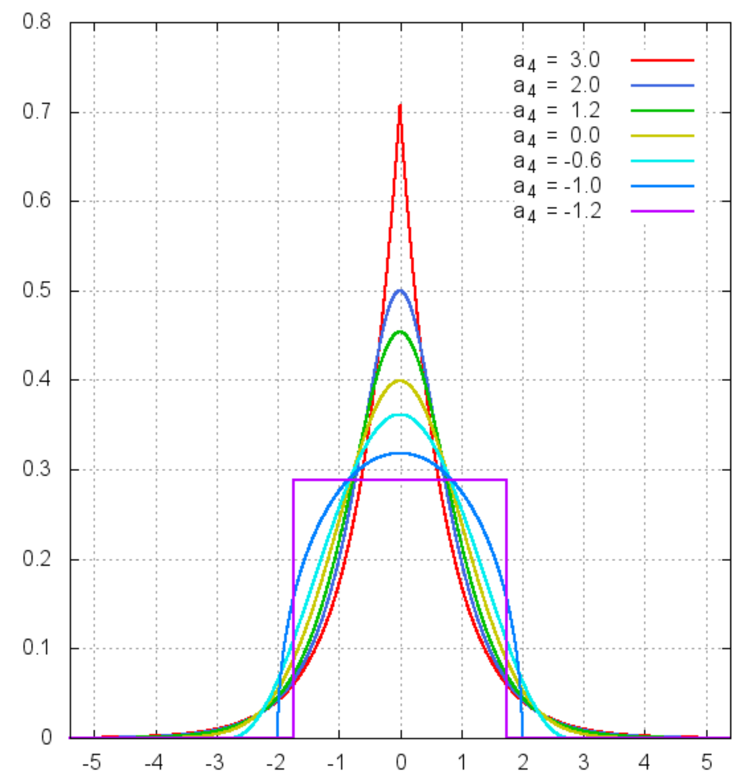
Špičatost

Špičatost (koeficient špičatosti) popisuje koncentraci rozdělení kolem středu (Meloun & Militký, 2013).

- normální rozdělení má nulovou špičatost
- při kladné špičatosti je křivka hustoty pravděpodobnosti špičatější než u normálního rozdělení
- při záporné špičatosti je křivka hustoty pravděpodobnosti plošší než u normálního rozdělení

Šikmost a_4 je definována vztahem:

$$a_4 = \frac{\sum_{i=1}^{n} (x_i - \mu)^4}{n \sigma^4} - 3$$



Meloun, M., Militký, J. (2013). *Kompendium statistického zpracování dat*. Praha: Karolinum.

Výpočet šikmosti a špičatosti v Excelu

V Excelu existuje pro výpočet šikmosti funkce a_3* =SKEW()
a pro výpočet špičatosti funkce a_4*=KURT().
Bohužel jsou tyto funkce definovány pomocí jiných vzorců:

$$a_3^* = \frac{N}{(N-1)(N-2)} \frac{\sum_{i=1}^{n} (x_i - \mu)^3}{\sigma^3}$$

$$a_4^* = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \frac{\sum_{i=1}^{n} (x_i - \mu)^4}{\sigma^4} - \frac{3(N-1)^2}{(N-2)(N-3)}$$

Hodnoty lze přepočítat podle vztahů:

$$a_3 = \frac{(N-2)}{\sqrt{(N(N-1))}} a_3^* \qquad a_4 = \frac{(N-2)(N-3)}{N^2-1} a_4^* - \frac{6}{N+1}$$