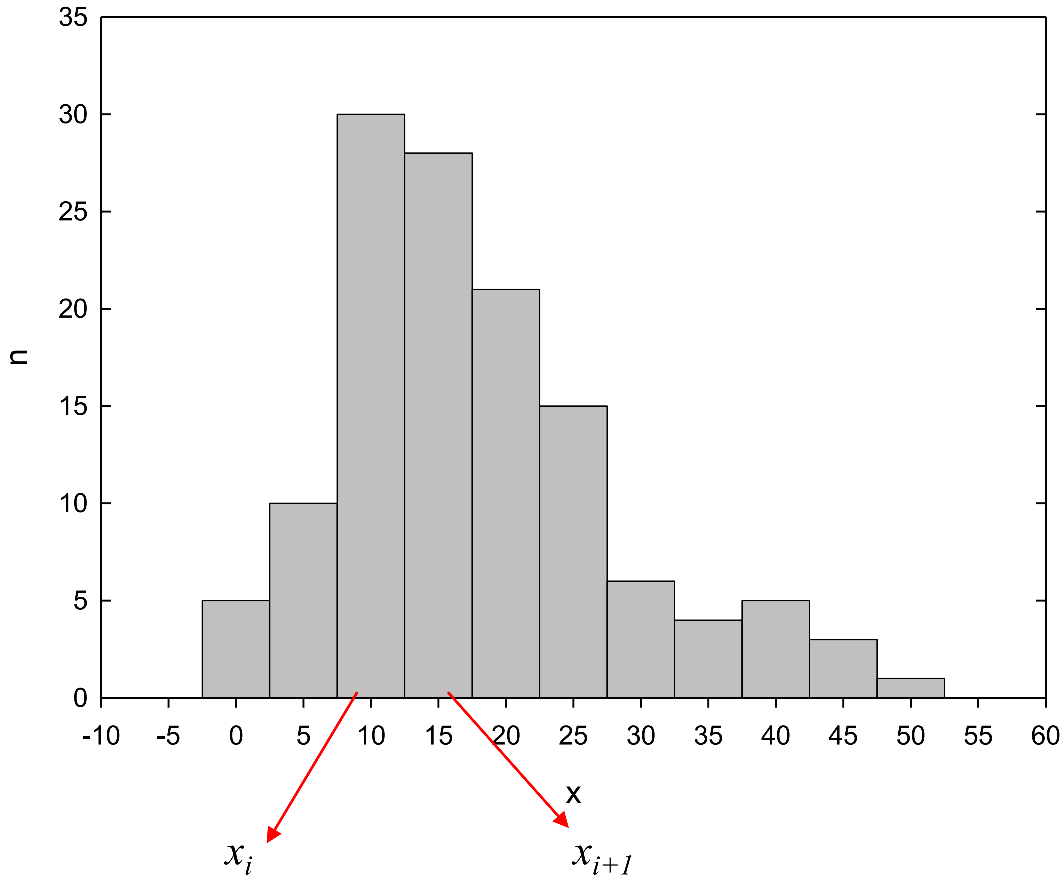
Histogram

Histogram – způsob jak experimentálně zjistit hustotu pravděpodobnosti z experimentálních dat



šířka binu: $\Delta_i = x_{i+1} - x_i$

plocha histogramu: $\sum_{i=1}^{m} n_i \Delta_i$

$\Downarrow$

normalizovaný histogram:

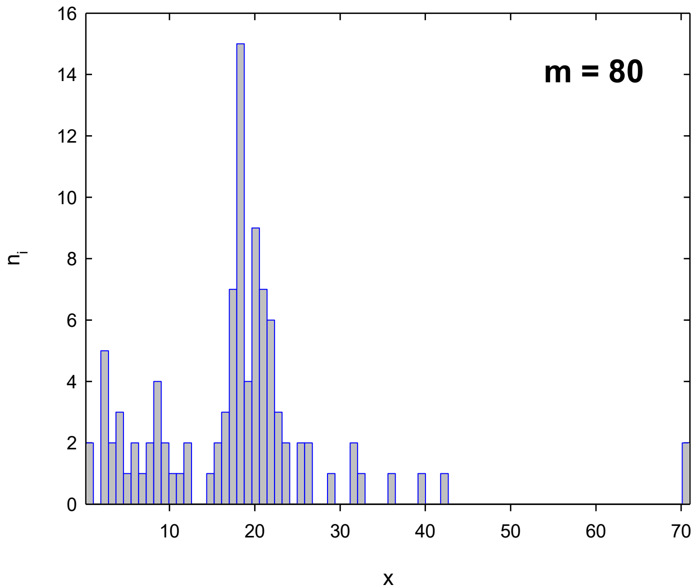
$$\xi_i = \frac{n_i}{\Delta_i N}, \quad \text{kde} \quad N = \sum_{i=1}^{m} n_i$$

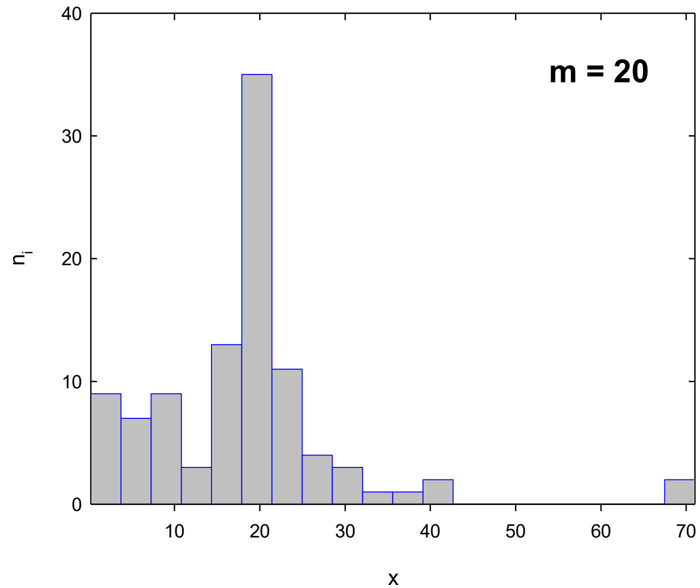
plocha normovaného histogramu: $\sum_{i=1}^{m} \xi_i \Delta_i = 1$

hustota pravděpodobnosti:

$$f(x_i) = \lim_{\substack{\Delta_i \to 0 \\ N \to \infty}} \xi_i = \lim_{\substack{\Delta_i \to 0 \\ N \to \infty}} \frac{n_i}{\Delta_i N}$$

Histogram – šířka binu





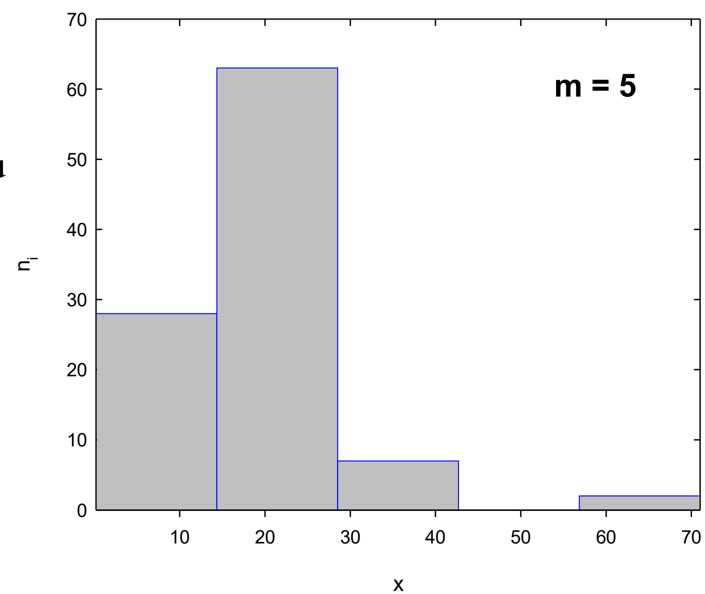
- **šířka binu** $w = \dfrac{x_{\max} - x_{\min}}{m}$

počet binů

Excel $m = \left[\sqrt{N}\right]$

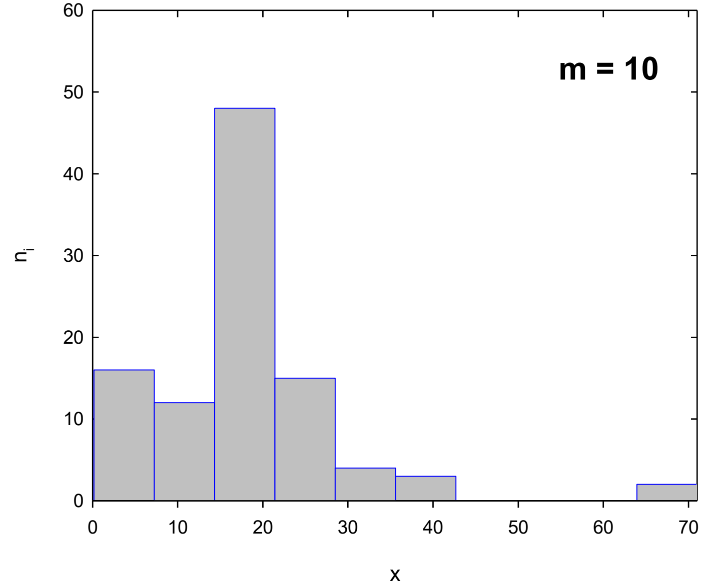
$$m = \left[\frac{\log N}{\log 2} + 1\right]$$

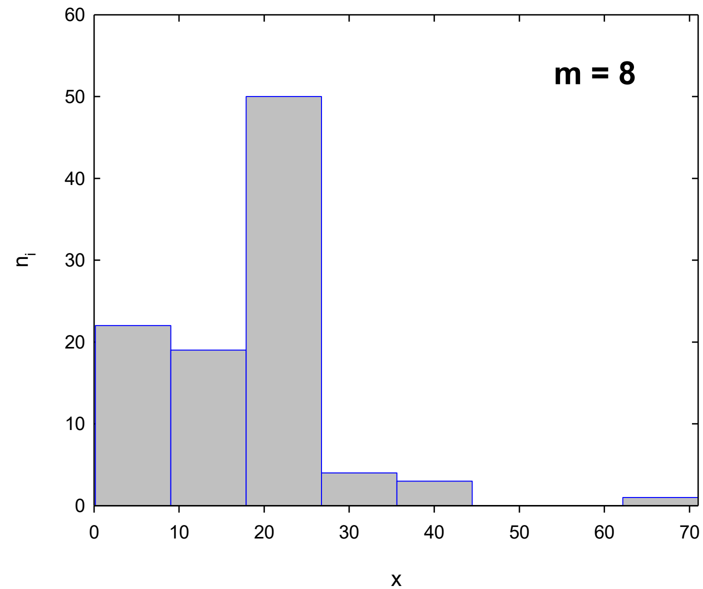
H. A. Sturges, J. American Statistical Association, 65–66 (1926).



16.17759	18.75516	8.60201
22.36831	41.95208	3.18462
3.29369	19.23135	15.66299
17.96900	8.88075	21.10663
18.52658	32.60371	2.28124
17.63568	4.27135	16.14332
17.79473	18.43469	35.88762
39.80907	23.99716	28.72841
18.25682	18.94920	0.17358
20.63264	8.22661	31.91945
25.89910	17.88642	70.80681
17.57289	17.96704	9.47664
18.74632	20.07927	23.20253
8.46536	7.04639	6.16414
21.63599	12.39286	15.65710
31.43157	18.06331	7.47195
2.71104	17.36080	20.18533
9.89574	17.95492	1.98676
18.16503	7.71726	17.71942
20.18927	20.49528	21.70207
11.27086	21.00411	21.28737
2.49163	25.37069	16.99344
11.77613	21.77872	18.19663
0.25810	24.99534	19.87326
4.53349	21.43774	5.84716
21.22557	10.56477	71.01371
20.04356	4.50194	18.09185
18.79175	23.01736	21.75327
20.86614	20.48741	17.09857
17.80408	20.42592	15.19833
18.29748	20.22713	19.04226
18.08830	21.35032	22.89348
4.65786	26.23743	
21.52645	2.10586	

Histogram – šířka binu





- **šířka binu** $$w = \frac{x_{\max} - x_{\min}}{m}$$

Excel $$m = \left[\sqrt{N}\right] = 10$$

$$n = \left[\frac{\log N}{\log 2} + 1\right] = 8$$

H. A. Sturges, J. American Statistical Association, 65–66 (1926).

16.17759	18.75516	8.60201
22.36831	41.95208	3.18462
3.29369	19.23135	15.66299
17.96900	8.88075	21.10663
18.52658	32.60371	2.28124
17.63568	4.27135	16.14332
17.79473	18.43469	35.88762
39.80907	23.99716	28.72841
18.25682	18.94920	0.17358
20.63264	8.22661	31.91945
25.89910	17.88642	70.80681
17.57289	17.96704	9.47664
18.74632	20.07927	23.20253
8.46536	7.04639	6.16414
21.63599	12.39286	15.65710
31.43157	18.06331	7.47195
2.71104	17.36080	20.18533
9.89574	17.95492	1.98676
18.16503	7.71726	17.71942
20.18927	20.49528	21.70207
11.27086	21.00411	21.28737
2.49163	25.37069	16.99344
11.77613	21.77872	18.19663
0.25810	24.99534	19.87326
4.53349	21.43774	5.84716
21.22557	10.56477	71.01371
20.04356	4.50194	18.09185
18.79175	23.01736	21.75327
20.86614	20.48741	17.09857
17.80408	20.42592	15.19833
18.29748	20.22713	19.04226
18.08830	21.35032	22.89348
4.65786	26.23743	
21.52645	2.10586	

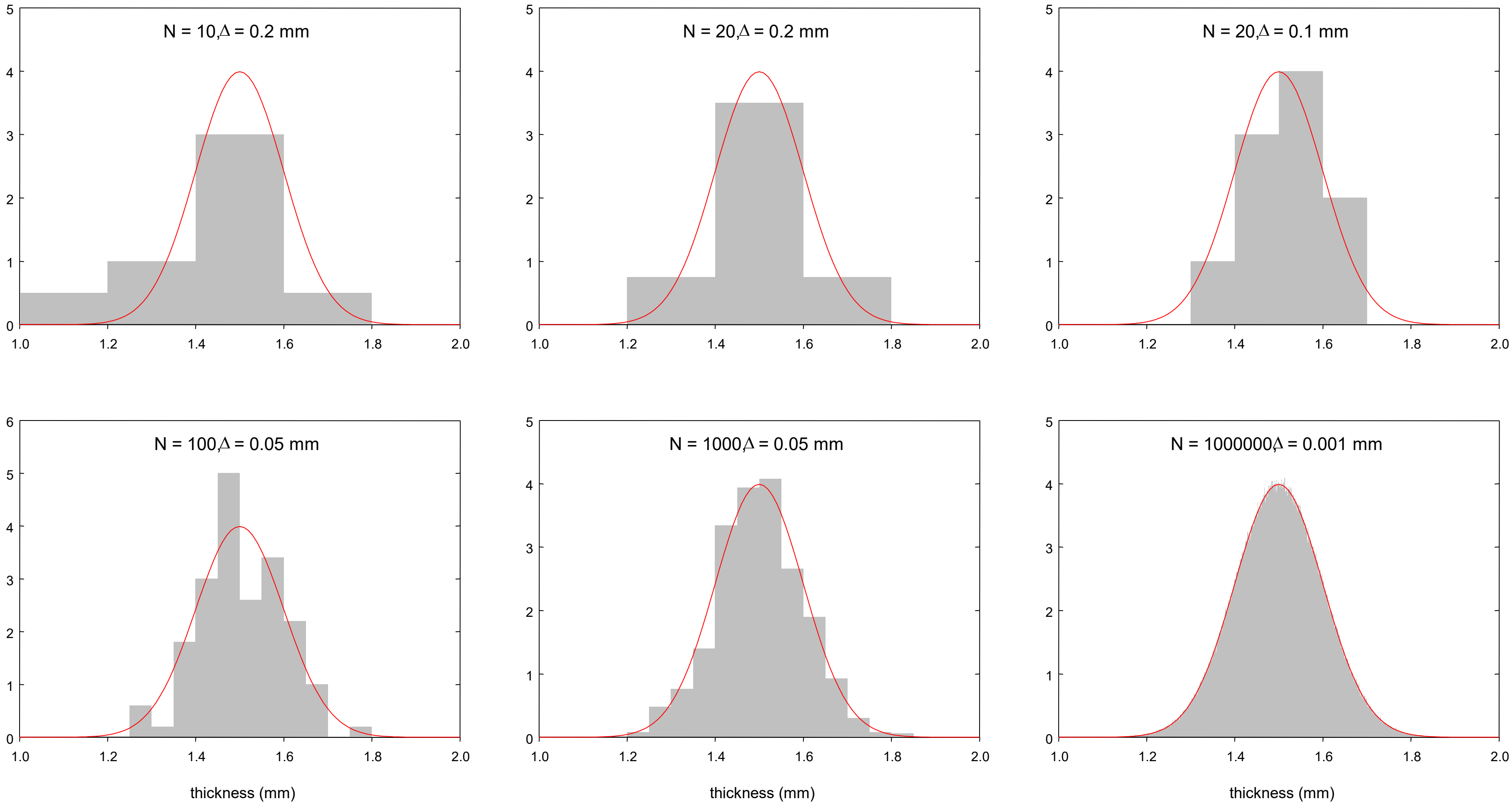
Hustota pravděpodobnosti – Měření tloušťky vzorku

μ = 1.5 mm, σ = 0.1 mm

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

- počet naměřených hodnot: N
- šířka binu: Δ

$$f(x_i)\approx\frac{n_i}{\Delta N} \qquad N\to\infty, \Delta\to 0$$



Momenty

• operátor **střední (očekávané) hodnoty** $\mu = E[x]$

• **diskrétní** náhodná proměnná: $\mu \equiv E[x] = \sum_i x_i P_i$

• **spojitá** náhodná proměnná: $\mu \equiv E[x] = \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x$

• **rozptyl (variance)**: $\sigma^2 = V[x] = \mu_2' = E\left[(x-\mu)^2\right] = E\left[x^2\right] - \left(E[x]\right)^2$

• **standardní odchylka**: $\sigma = \sqrt{V[x]}$

***n*-tý moment:** $\mu_n = E\left[x^n\right]$

***n*-tý centrální moment:** $\mu_n' = E\left[(x-\mu)^n\right]$

• $\mu_1' = 0$ • $\mu_2' = \sigma^2$

Momenty vyšších řádů

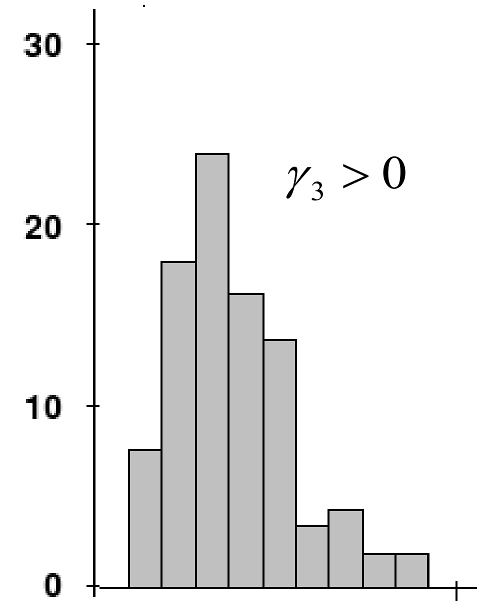
- operátor střední (očekávané) hodnoty $\mu = E[x]$
- diskrétní náhodná proměnná: $$\mu \equiv E[x] = \sum_i x_i P_i$$
- spojitá náhodná proměnná: $$\mu \equiv E[x] = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$$

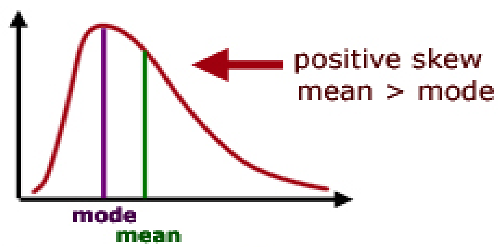
šikmost (skewness): $$\gamma_3 = \frac{\mu_3'}{\sigma^3} = \frac{E\left[(x-\mu)^3\right]}{\left(E\left[(x-\mu)^2\right]\right)^{\frac{3}{2}}}$$

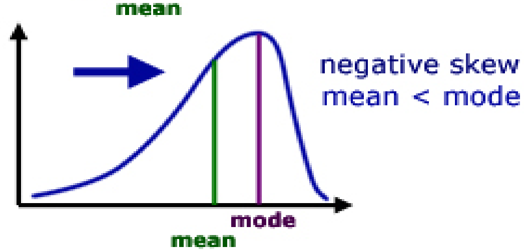
špičatost (kurtosis): $$\gamma_4 = \frac{\mu_4'}{\sigma^4} = \frac{E\left[(x-\mu)^4\right]}{\left(E\left[(x-\mu)^2\right]\right)^{2}}$$

normální rozdělení: $\gamma_4 = 3$

dodatečná (excesivní) špičatost: $\gamma_4 - 3$



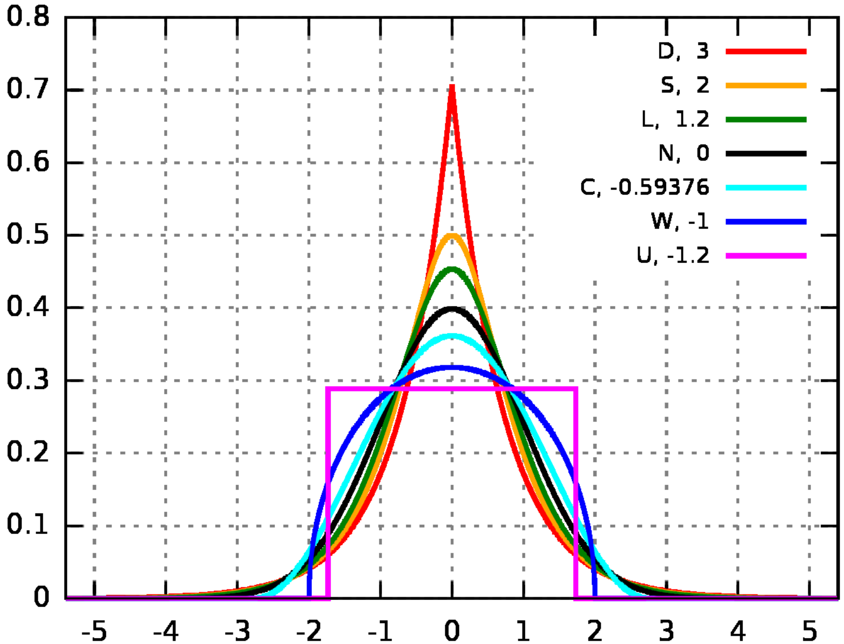


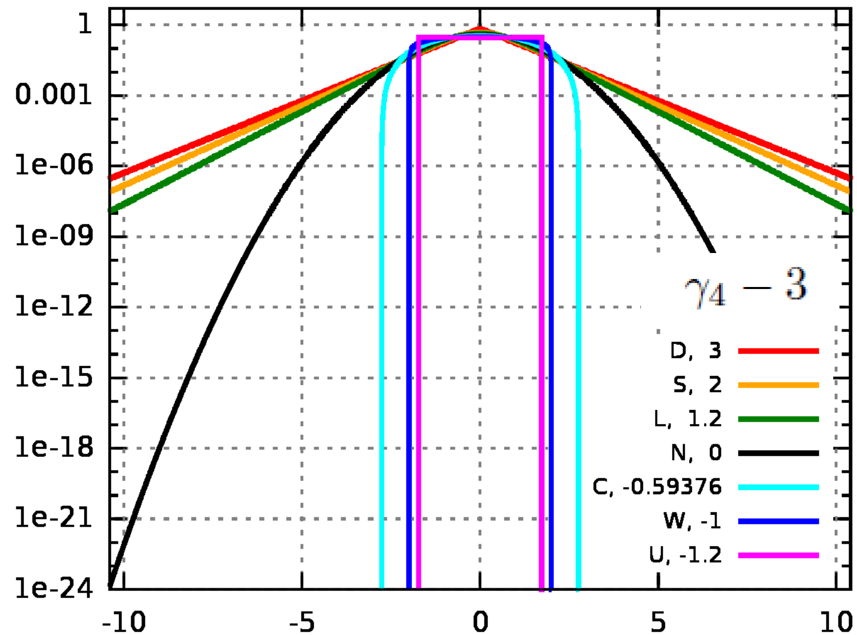


Špičatost

špičatost (kurtosis): $\gamma_4 = \dfrac{\mu_4'}{\sigma^4} = \dfrac{E\left[(x-\mu)^4\right]}{\left(E\left[(x-\mu)^2\right]\right)^2}$

$\gamma_4 - 3$

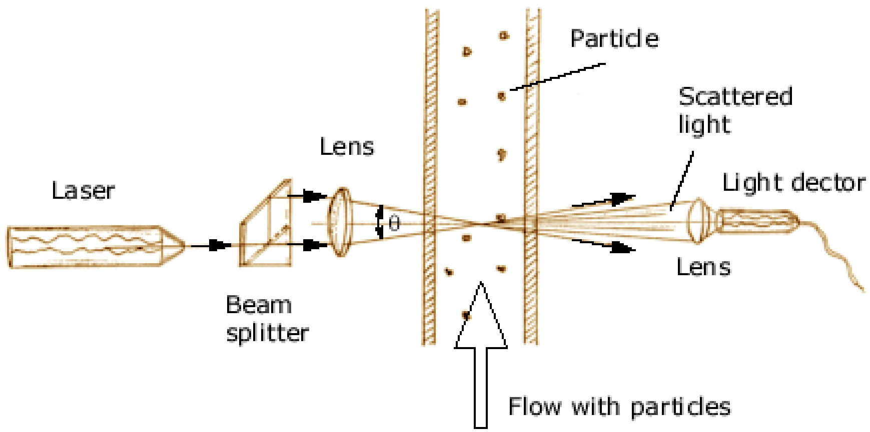


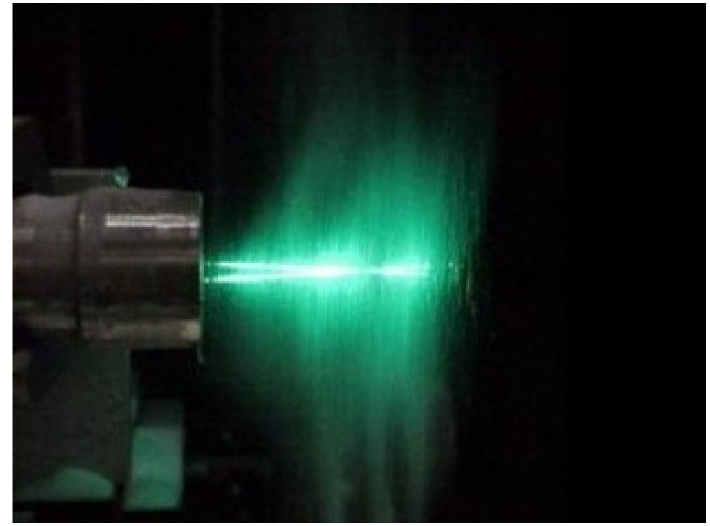


https://en.wikipedia.org/wiki/Kurtosis

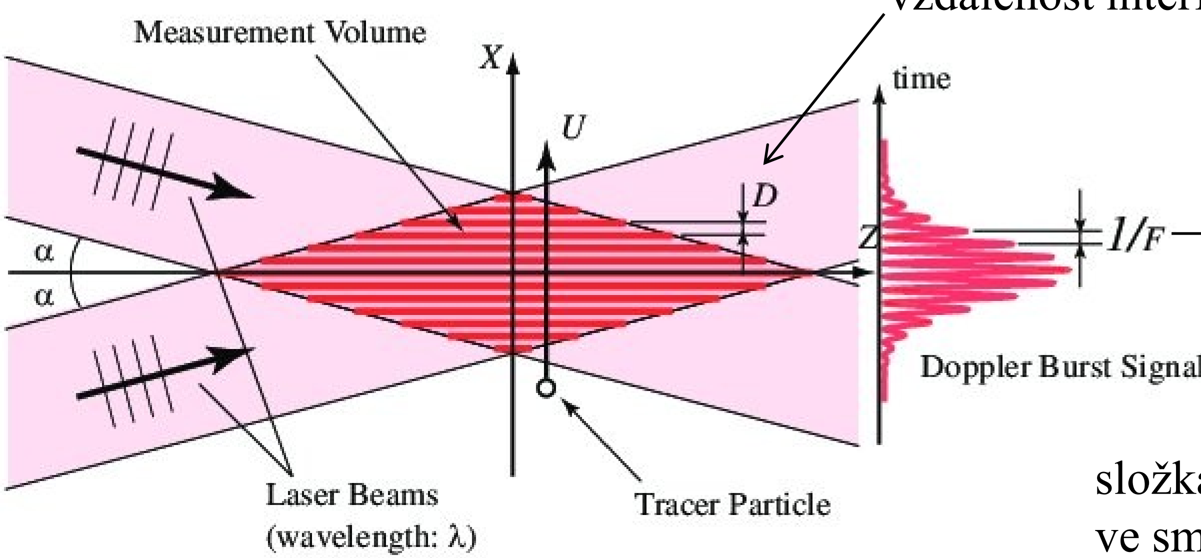
Příklad – laserová Dopplerovská anemometrie

• měření rychlosti částic





vzdálenost interferenčních proužků: $D = \dfrac{\lambda_0}{2\sin\alpha}$

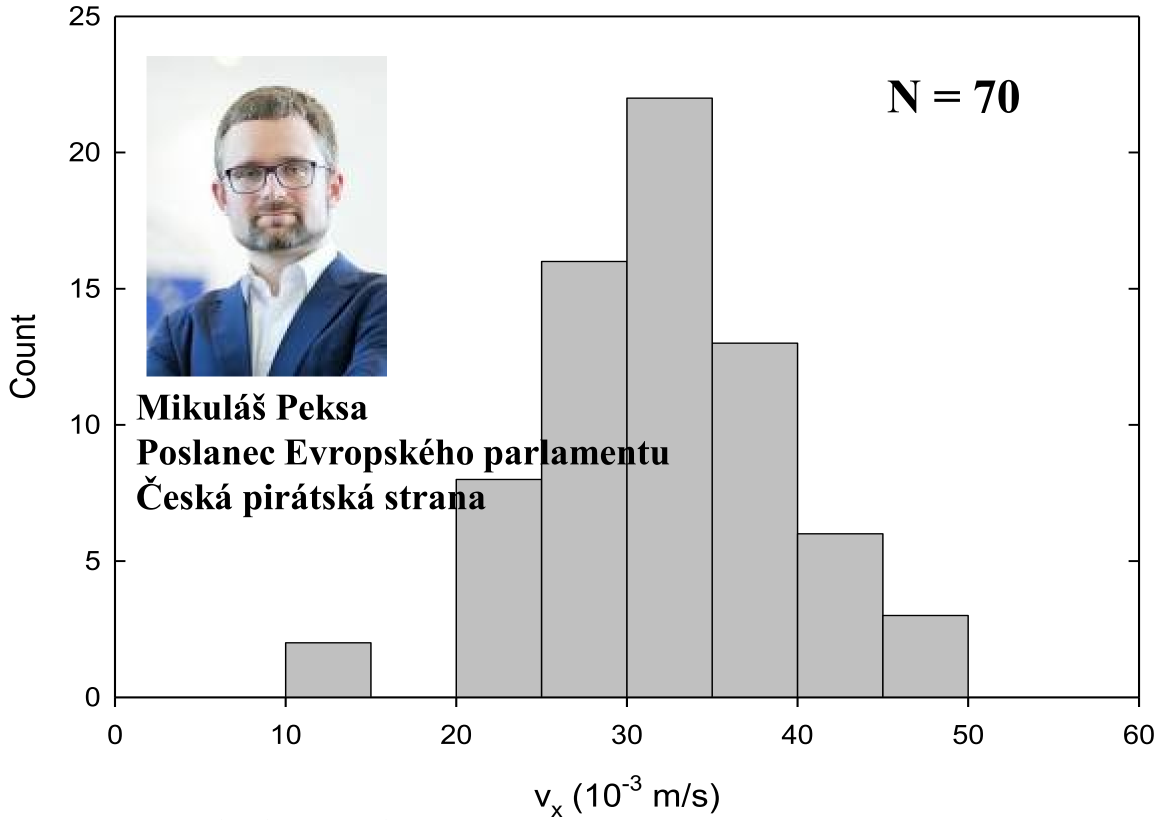


změříme frekvenci F
detekovaných signálů

složka rychlosti částice
ve směru kolmo na proužky $v_x = \dfrac{\lambda_0}{2\sin\alpha}F$

Příklad – laserová Dopplerovská anemometrie

Histogram průmětu rychlosti částic



14.3. 2007 Mikuláš Peksa

Příklad – laserová Dopplerovská anemometrie

částice „ideálního plynu“

hmotnost částice

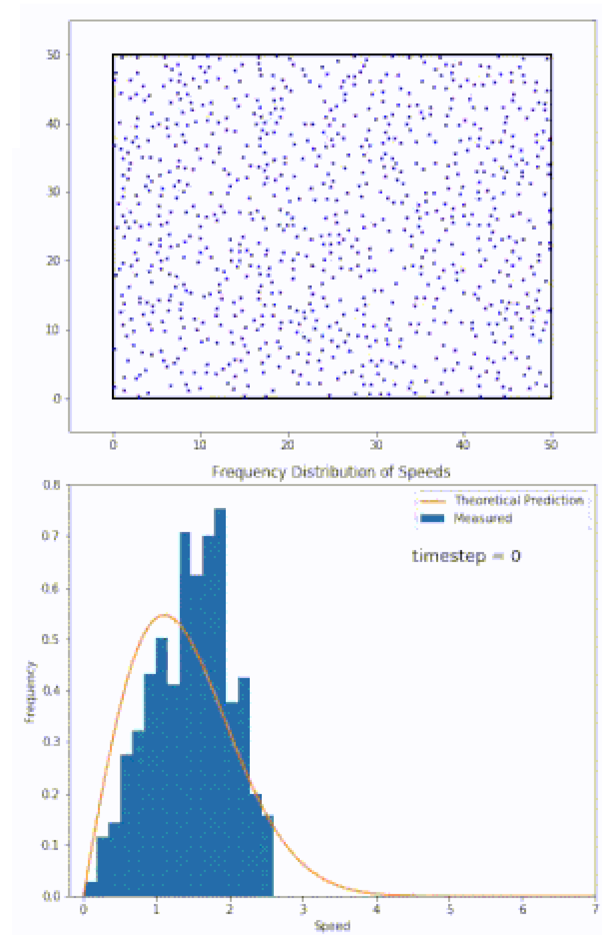
Maxwell-Boltzmanovo rozdělení $$f(v) = \sqrt{\frac{2}{\pi}} \left(\frac{m}{kT}\right)^{3/2} v^2 \exp\left[-\frac{mv^2}{2kT}\right]$$

Boltzmannova konstanta *teplota*

To je ale rozdělení velikostí rychlostí $$v = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

Složky rychlostí v_i mají normální rozdělení

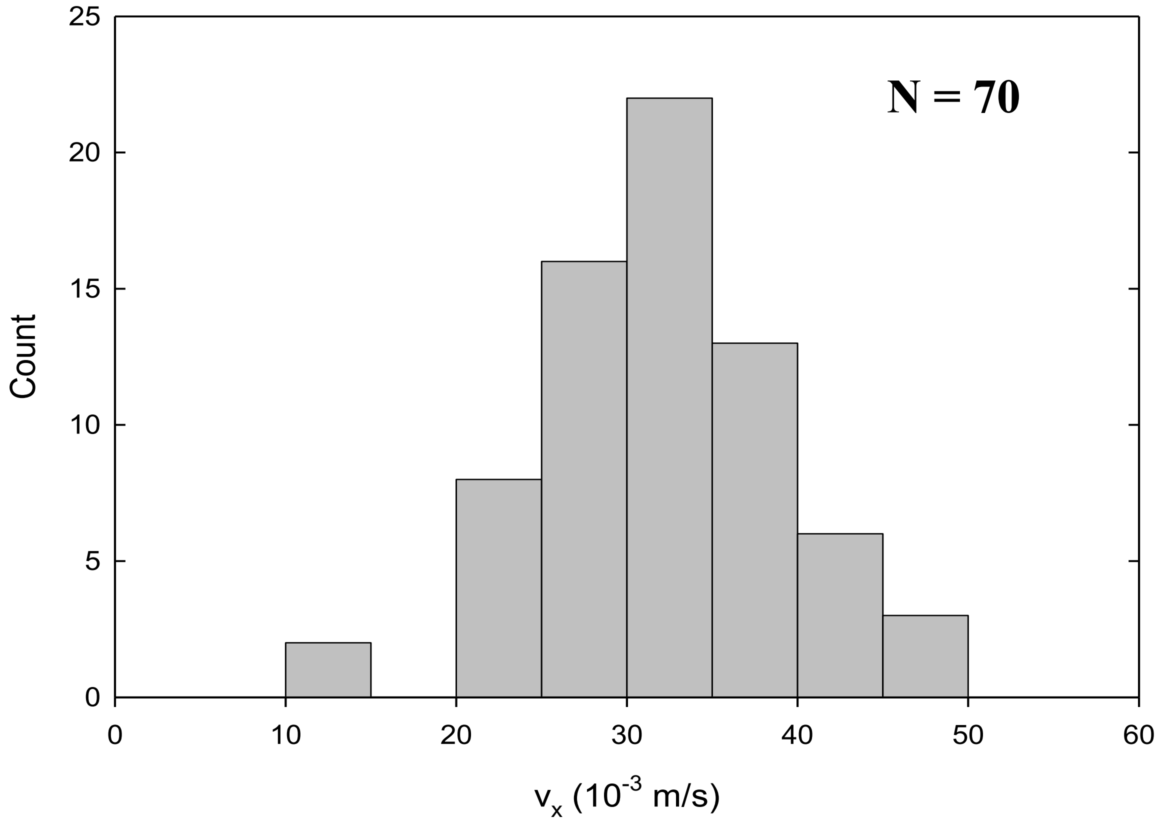
$$f(v_i) = \sqrt{\frac{m}{2\pi kT}} \exp\left[\frac{-mv_i^2}{2kT}\right]$$



https://en.wikipedia.org/wiki/Maxwell-Boltzmann_distribution

Příklad – laserová Dopplerovská anemometrie

Histogram průmětu rychlosti částic



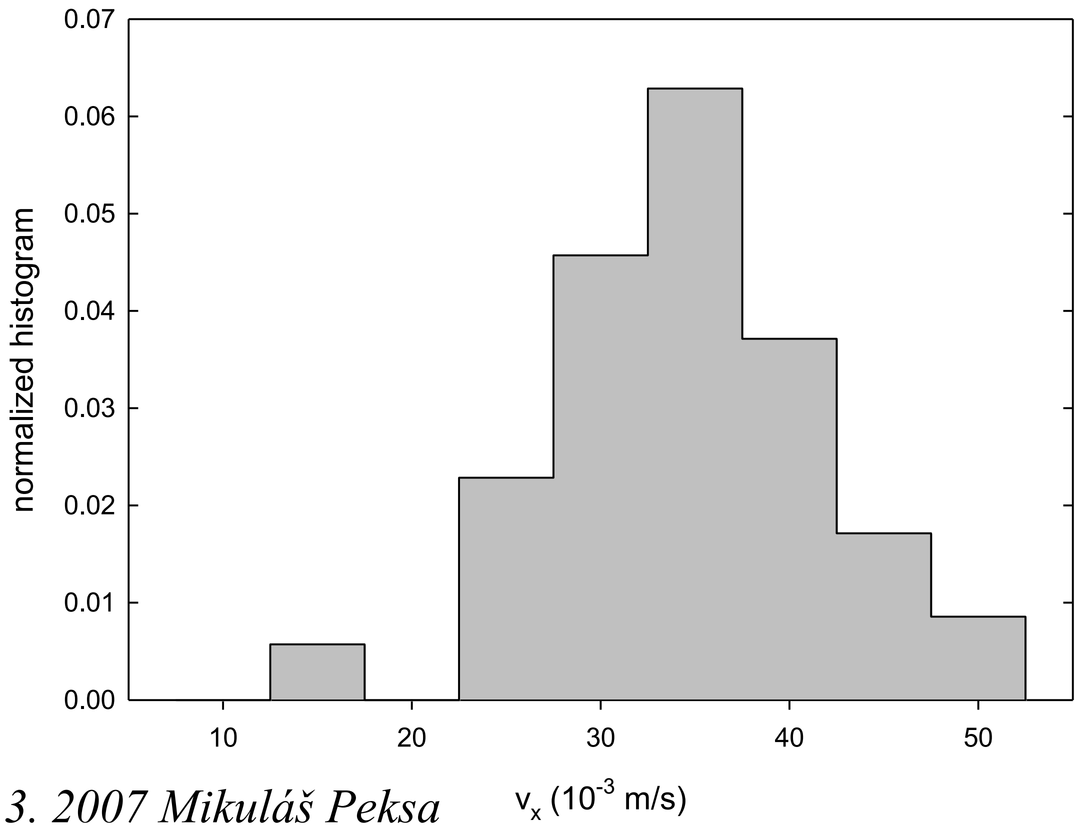
Je to normální rozdělení?

Je šikmost nulová?

Je dodatečná špičatost nulová?

Příklad - Dopplerovská anemometrie

- odhad očekávané hodnoty: $\hat{\mu} = \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = (32.5 \times 10^{-3})\ \mathrm{m\ s^{-1}}$
- odhad standardní odchylky: $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} = (6.7 \times 10^{-3})\ \mathrm{m\ s^{-1}}$



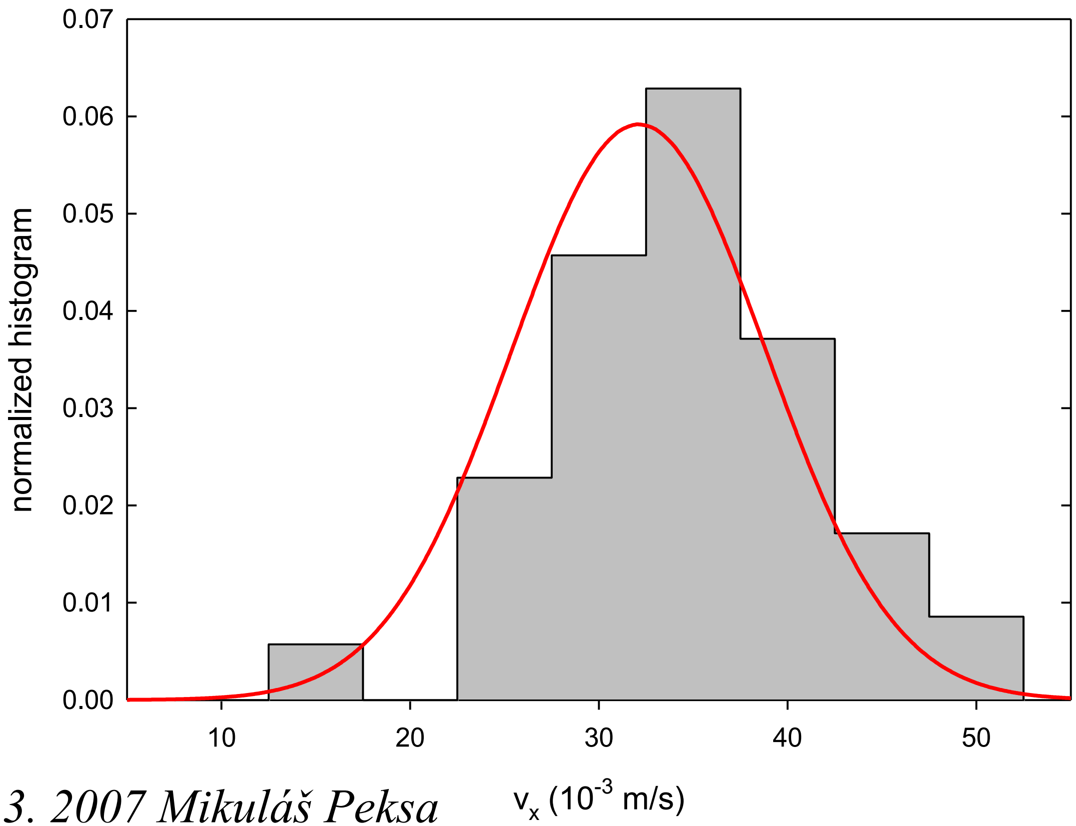
Má normální rozdělení?

Je šikmost nulová?

Je dodatečná špičatost nulová?

Příklad - Dopplerovská anemometrie

- odhad očekávané hodnoty: $\hat{\mu} = \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = (32.5 \times 10^{-3})\ \mathrm{m\ s^{-1}}$
- odhad standardní odchylky: $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} = (6.7 \times 10^{-3})\ \mathrm{m\ s^{-1}}$



Má normální rozdělení?

Je šikmost nulová?

Je dodatečná špičatost nulová?

Gaussián

$$f(x) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left[-\frac{(x-\overline{x})^2}{2\hat{\sigma}^2}\right]$$

14.3. 2007 Mikuláš Peksa

Příklad - Dopplerovská anemometrie

- odhad očekávané hodnoty: $\hat{\mu} = \overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = (32.5 \times 10^{-3})\ \mathrm{m\ s^{-1}}$
- odhad standardní odchylky: $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} = (6.7 \times 10^{-3})\ \mathrm{m\ s^{-1}}$
- odhad šikmosti: $\hat{\gamma}_3 = \frac{\hat{\mu}'_3}{\hat{\sigma}^3} = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^3}{\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^{3/2}} = -0.16$ (Fisher-Pearson)
- odhad šikmosti: $\hat{\gamma}_3 = \frac{N^2}{(N-1)(N-2)}\frac{\hat{\mu}'_3}{\hat{\sigma}^3} = \frac{\sqrt{N(N-1)}}{N-2}\frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^3}{\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^{3/2}} = -0.17$

(korigovaný Fisher-Pearson)

- odhad chyby odhadu šikmosti: $\sigma_{skew} = \sqrt{\frac{6N(N-1)}{(N-2)(N+1)(N+3)}} \approx \sqrt{\frac{6}{N}} = 0.3$

výsledný odhad šikmosti: $\hat{\gamma}_3 = -0.2 \pm 0.3$

Příklad - Dopplerovská anemometrie

- odhad očekávané hodnoty: $\hat{\mu} = \overline{x} = \dfrac{1}{N}\sum_{i=1}^{N} x_i = (32.5 \times 10^{-3})\ \mathrm{m\ s^{-1}}$
- odhad standardní odchylky: $\hat{\sigma} = \sqrt{\dfrac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N-1}} = (6.7 \times 10^{-3})\ \mathrm{m\ s^{-1}}$
- odhad dodatečné špičatosti: $\hat{\gamma}_4 - 3 = \dfrac{\hat{\mu}_4'}{\hat{\sigma}^4} - 3 = \dfrac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^4}{\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^2} = 0.07$ (předpojatý)

$$\hat{\gamma}_4 - 3 = \frac{(N+1)N(N-1)}{(N-2)(N-3)} \frac{\sum_{i=1}^{N}(x_i - \overline{x})^4}{\left[\sum_{i=1}^{N}(x_i - \overline{x})^2\right]^2} - 3\frac{(N-1)^2}{(N-2)(N-3)} = 0.26$$

(nepředpojatý pro normální rozdělení)

- odhad chyby odhadu dodatečné špičatosti: $\hat{\sigma}_{kurt} = \sqrt{\dfrac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)}} = 0.56$

  pro velká N: $\hat{\sigma}_{kurt} \approx 2\sqrt{\dfrac{6}{N}} = 0.59$

výsledný odhad dodatečné špičatosti: $\hat{\gamma}_4 - 3 = 0.3 \pm 0.6$