

# Laserová dopplerovská anemometrie, statistická práce z předmětu Pravděpodobnost a statistika 1

Jindřich Dvořáček

1. 9. 2025

## Obsah

<b>1 Úvod</b>	<b>1</b>
<b>2 Teorie</b>	<b>2</b>
<b>3 Statistická práce</b>	<b>5</b>
3.1 Explorační analýza dat . . . . .	5
3.2 Test normality . . . . .	5
3.3 Intervalové odhady, test hypotézy . . . . .	5
3.4 Dodatečné informace . . . . .	6
3.5 Fitovaný histogram . . . . .	7
<b>4 Závěr</b>	<b>7</b>
<b>5 Kód</b>	<b>8</b>

## 1 Úvod

Zdrojem dat pro naši statistickou práci je úloha č. 18 z fyzikálního předmětu „Praktikum III - Optika“ nesoucí název „Laserová dopplerovská anemometrie“, ve které jde o měření rychlostí částic ve vodě. Tyto rychlosti by měly mít gaussovské rozdělení s neznámými parametry. Naším úkolem v této statistické práci bude rozhodnout, zda-li jsou rychlosti rozděleny normálně a s jakými parametry. Provedeme také jednoduchý test hypotézy.

Úlohu jsme neměřili, data byla získána od studenta. Budeme vycházet z tabulky 1, kde částice označuje kolikátou částici měřil,  $\Delta t$  je časová odlehlost maxim,  $n$  je počet maxim,  $f$  je přepočtená frekvence,  $v_x$  je průmět rychlosti do  $x$  směru a  $\sigma$  značí chybu měření. Sami jsme si data uložili do formátu .csv, který jsme pak nahráli do Pythonu. Pro vysvětlení podstaty

měření a značení viz sekce 2. Dále však budeme pracovat pouze se sloupci  $v_x$ , tedy samotnými rychlostmi. Pro jednoduchost nebudeme brát v potaz chyby měření těchto rychlostí, protože nevíme, jak s nimi naložit.

částica	$\Delta t$ (ms)	$n$	$f$ (ms <sup>-1</sup> )	$\sigma_f$ (ms <sup>-1</sup> )	$v_x$ (μm/ms)	$\sigma_{v_x}$ (μm/ms)	částica	$\Delta t$ (ms)	$n$	$f$ (ms <sup>-1</sup> )	$\sigma_f$ (ms <sup>-1</sup> )	$v_x$ (μm/ms)	$\sigma_{v_x}$ (μm/ms)
1	7,48	5	0,67	0,04	20,17	1,40	41	7,10	4	0,56	0,04	17,00	1,24
2	11,82	4	0,34	0,01	10,21	0,47	42	6,87	4	0,58	0,04	17,57	1,32
3	7,32	5	0,68	0,05	20,61	1,46	43	8,59	5	0,58	0,03	17,57	1,07
4	9,50	7	0,74	0,04	22,24	1,24	44	10,87	5	0,46	0,02	13,88	0,69
5	7,57	4	0,53	0,03	15,95	1,09	45	8,24	4	0,49	0,03	14,65	0,93
6	7,49	5	0,67	0,04	20,15	1,40	46	9,37	4	0,43	0,02	12,88	0,73
7	16,51	5	0,30	0,01	9,14	0,33	47	13,25	6	0,45	0,02	13,67	0,58
8	9,57	4	0,42	0,02	12,61	0,70	48	10,27	4	0,39	0,02	11,75	0,61
9	9,83	5	0,51	0,03	15,35	0,83	49	8,71	5	0,57	0,03	17,32	1,05
10	9,98	5	0,50	0,03	15,12	0,81	50	6,51	4	0,61	0,05	18,54	1,47
11	10,45	5	0,48	0,02	14,44	0,74	51	10,56	5	0,47	0,02	14,29	0,73
12	7,09	4	0,56	0,04	17,03	1,24	52	5,97	4	0,67	0,06	20,22	1,74
13	6,10	4	0,66	0,05	19,79	1,66	53	9,02	5	0,55	0,03	16,73	0,98
14	8,28	5	0,60	0,04	18,22	1,15	54	7,64	4	0,52	0,03	15,80	1,08
15	8,38	5	0,60	0,04	18,01	1,13	55	6,23	4	0,64	0,05	19,38	1,60
16	7,66	5	0,65	0,04	19,70	1,34	56	9,19	4	0,44	0,02	13,14	0,76
17	13,20	5	0,38	0,01	11,43	0,48	57	11,34	8	0,71	0,03	21,29	1,02
18	10,09	3	0,30	0,01	8,97	0,48	58	8,77	5	0,57	0,03	17,21	1,03
19	14,85	5	0,34	0,01	10,16	0,39	59	9,01	5	0,55	0,03	16,75	0,98
20	15,06	5	0,33	0,01	10,02	0,38	60	8,66	5	0,58	0,03	17,42	1,06
21	8,54	4	0,47	0,03	14,14	0,87	61	6,69	4	0,60	0,04	18,04	1,39
22	8,74	4	0,46	0,03	13,81	0,83	62	6,87	4	0,58	0,04	17,57	1,32
23	10,71	5	0,47	0,02	14,09	0,71	63	7,11	4	0,56	0,04	16,98	1,24
24	7,66	4	0,52	0,03	15,76	1,07	64	8,79	5	0,57	0,03	17,17	1,03
25	9,25	5	0,54	0,03	16,31	0,93	65	8,89	4	0,45	0,03	13,58	0,80
26	9,73	5	0,51	0,03	15,51	0,85	66	7,10	4	0,56	0,04	17,00	1,24
27	10,27	5	0,49	0,02	14,69	0,77	67	6,69	4	0,60	0,04	18,04	1,39
28	9,49	5	0,53	0,03	15,90	0,89	68	7,28	4	0,55	0,04	16,58	1,18
29	7,52	4	0,53	0,04	16,05	1,11	69	7,52	4	0,53	0,04	16,05	1,11
30	7,34	4	0,54	0,04	16,45	1,16	70	8,06	4	0,50	0,03	14,98	0,97
31	8,18	4	0,49	0,03	14,76	0,94	71	7,12	4	0,56	0,04	16,95	1,23
32	9,86	5	0,51	0,03	15,30	0,83	72	7,94	5	0,63	0,04	19,00	1,25
33	9,85	5	0,51	0,03	15,32	0,83	73	6,81	5	0,73	0,05	22,16	1,68
34	9,37	5	0,53	0,03	16,10	0,91	74	5,85	5	0,85	0,07	25,79	2,26
35	7,59	4	0,53	0,03	15,90	1,09	75	9,43	5	0,53	0,03	16,00	0,90
36	7,58	4	0,53	0,03	15,93	1,09	76	8,06	5	0,62	0,04	18,72	1,21
37	9,44	5	0,53	0,03	15,98	0,90	77	6,81	4	0,59	0,04	17,73	1,34
38	9,67	5	0,52	0,03	15,60	0,86	78	7,76	5	0,64	0,04	19,44	1,30
39	11,64	6	0,52	0,02	15,56	0,73	79	8,85	4	0,45	0,03	13,64	0,81
40	8,53	5	0,59	0,03	17,69	1,09	80	8,12	5	0,62	0,04	18,58	1,20

Obrázek 1: Původní data

## 2 Teorie

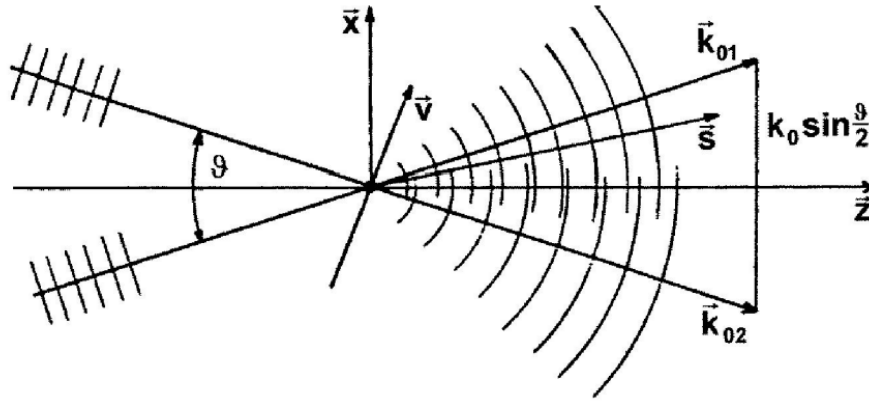
Následující část není nutná k samotné statistické práci. Poskytuje intuitivní vysvětlení, jak funguje měření.

Metoda laserové dopplerovské anemometrie je jedna z nejpřesnějších metod měření rychlostí. Podstata tkví v Dopplerově principu, který říká, že frekvence elektromagnetického záření (laserové světlo) se změní po odrazu nebo rozptýlení částic v závislosti na rychlosti částice.

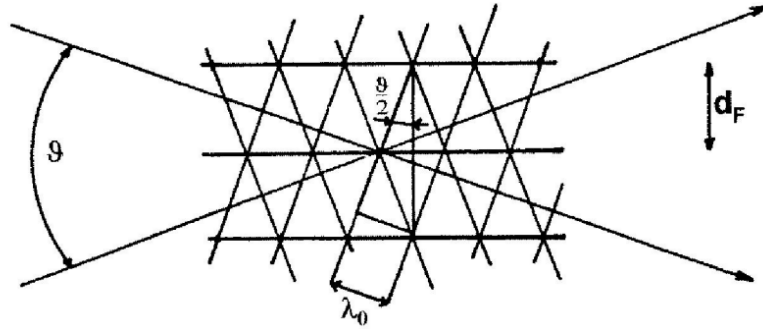
Při experimentu se používá tzv. diferenciálního uspořádání kvůli odstranění směrové závislosti vektoru rychlosti částice, kde se laserový svazek rozdělí na dva koherentní svazky, které následně svírají úhel  $\vartheta$  a díky tomu vznikne *interference ve sbíhavých svazcích*, viz obrázek 2 a 3. V průsečíku se objeví 3D interferenční pole, které ve 2D znázorníme jako soustavu ekvidistantních proužků, které, jak plyne z teorie a geometrie, mají vzdálenost

$$d_F = \frac{\lambda_0}{2 \sin(\vartheta/2)}, \quad (1)$$

kde  $\lambda_0$  je vlnová délka laseru a  $\vartheta$  úhel mezi paprsky, viz obrázek 3.



Obrázek 2: Diferenciální dopplerovský signál, převzato z [1]



Obrázek 3: Interferenční pole v průsečíku laserových paprsků, převzato z [1]

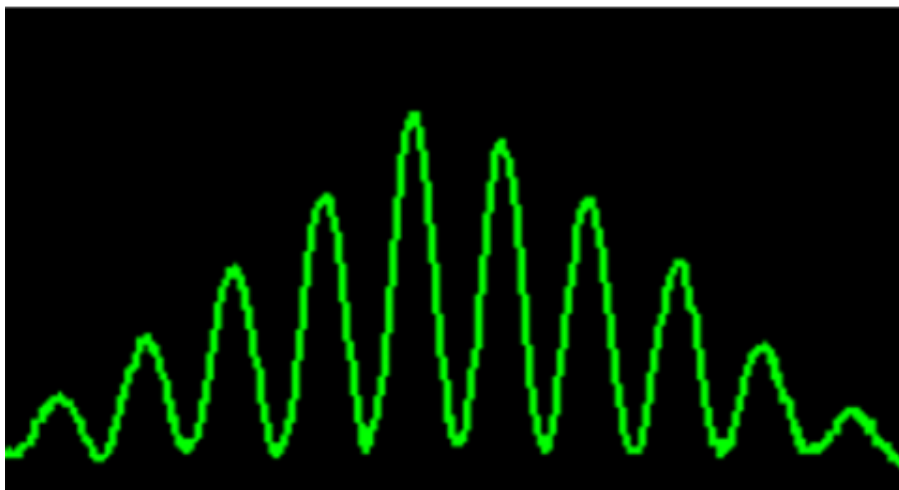
Při průletu částice **světlym** proužkem dojde k rozptylu světla, které můžeme zaznamenat. Částice během pohybu prolétne několika proužky a tedy frekvenci tohoto rozptylování spočteme snadno dle zákonů kinematiky. Čas je dráha lomená rychlostí a frekvence je reciproký čas, neboli

$$t = \frac{s}{v} \iff f = \frac{v}{s}. \quad (2)$$

Rychlost, nebo přesněji její *průmět* do osy  $x$ , jako na obrázku 2, chceme změřit a dráhu známe, to je  $d_F$ . Když dáme vztahy dohromady, dostaneme

$$f_D = \frac{v_x}{d_F} = v_x \frac{2 \sin(\vartheta/2)}{\lambda_0}. \quad (3)$$

Ukázka výstupního signálu je na obrázku 4. Vidíme zde soustavu píků odpovídající jednotlivým rozptylům na světlych proužcích. Soustava vykazuje gaussové rozdělení, jelikož příčná intenzita laseru je gaussovská.



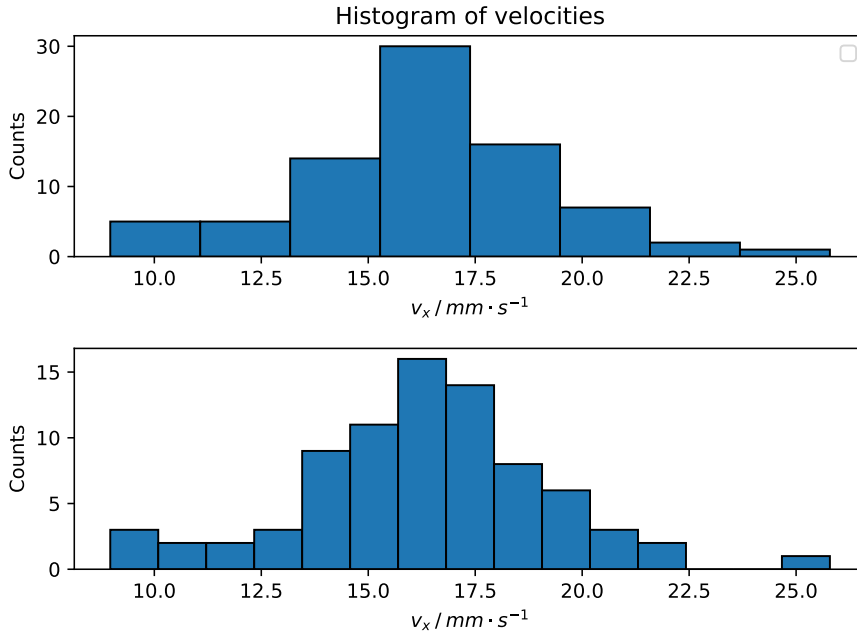
Obrázek 4: Ukázka vhodného dopplerovského signálu na osciloskopu, převzato z [1]

## 3 Statistická práce

### 3.1 Explorační analýza dat

Nejprve jsme si zobrazili histogram rychlostí s 8, resp. 15 biny, viz obrázek 7. Dle [2] je totiž optimální počet binů roven <sup>1</sup>

$$m_{opt} = \frac{\log N}{\log 2} + 1 \approx 8. \quad (4)$$



Obrázek 5: Histogram rychlostí

### 3.2 Test normality

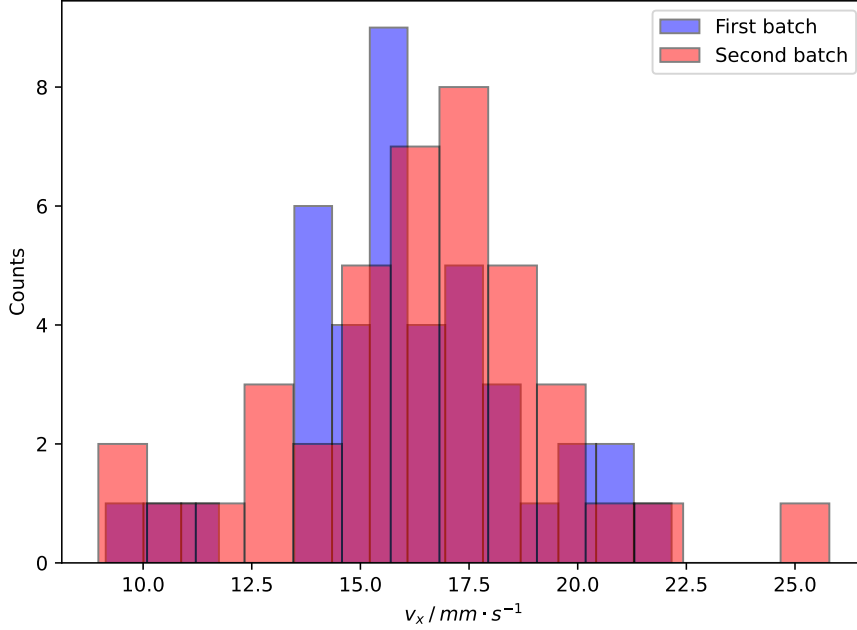
Před statistickým zpracováním jsme potřebovali ověřit, zda-li jsou naměřená data z normálního rozdělení. Pro to jsme použili Shapiro-Wilkův test normality. Nulová hypotéza  $H_0$  zní: „Naměřená data pocházejí z normálního rozdělení.“ Zjistili, že  $p = 0.20$ . To znamená, že s pravděpodobností 0.20 naměříme data, která mají stejně nebo hůře normální rozdělení, než máme my. Data tedy z normálního rozdělení nejspíš pocházejí.

### 3.3 Intervalové odhady, test hypotézy

Poté, co jsme potvrdili, že data mohou být normální, jsme mohli provést intervalový odhad  $\mu$  a provést jednoduchý test hypotézy. Na to jsme si data rozdělili do dvou polovin. Do první části jsme dali liché částice, do druhé sudé částice. Neudělali jsme rozdělení v půlce datasetu pro minimalizaci vnesení chyby výběru – experimentátor mohl prvních či posledních pár měření pokazit, což by se projevilo jen v jednom datasetu. Z prvního datasetu chceme zjistit informace

<sup>1</sup>Základ logaritmu je v této rovnici nepodstatný.

o problematice a prohlásit data za „vzorové“. Druhou polovinu pak použijeme na test hypotézy s parametry zjištěnými z první poloviny dat. Pro oba histogramy viz obrázek 6.



Obrázek 6: Histogram rychlostí, části dat

Předtím jsme provedli stejný test normality na oba datasets. Vychází pro první dataset  $p = 0.48$  a pro druhý  $p = 0.42$ . Opět tedy můžeme předpokládat normálnost rozdělení. Z první poloviny jsme odhadli střední hodnotu  $\mu_1$ , odchylku  $\sigma_1$ , standardní chybu  $\text{se}_1$  a spočetli 95% konfidenční interval podle známých vzorců z [3]. Vychází

$$\mu_1 = 16.14 \text{ mm s}^{-1}, \sigma_1 = 2.70 \text{ mm s}^{-1}, \text{se}_1 = 0.43 \text{ mm s}^{-1} \text{ 95-CI} = (15.28; 17.00). \quad (5)$$

Následně jsme vzali druhou polovinu dat, kterou můžeme považovat za nezávislou od té první, taktéž jsme odhadli střední hodnotu  $\mu_2$  a zkoumali jsme, zda-li  $\mu_2$  padne do konfidenčního intervalu. Vychází  $\mu_2 = 16.39 \text{ mm s}^{-1}$ , tedy střední hodnota druhé poloviny dat leží v **95-CI**. Můžeme tedy usoudit, že druhý dataset má stejnou střední hodnotu jako ten první.

### 3.4 Dodatečné informace

Další informaci o datech může poskytnout šikmost, skewness, která se spočte jako třetí centrální moment podělený druhým centrálním momentem umocněným na  $3/2$ , [2]

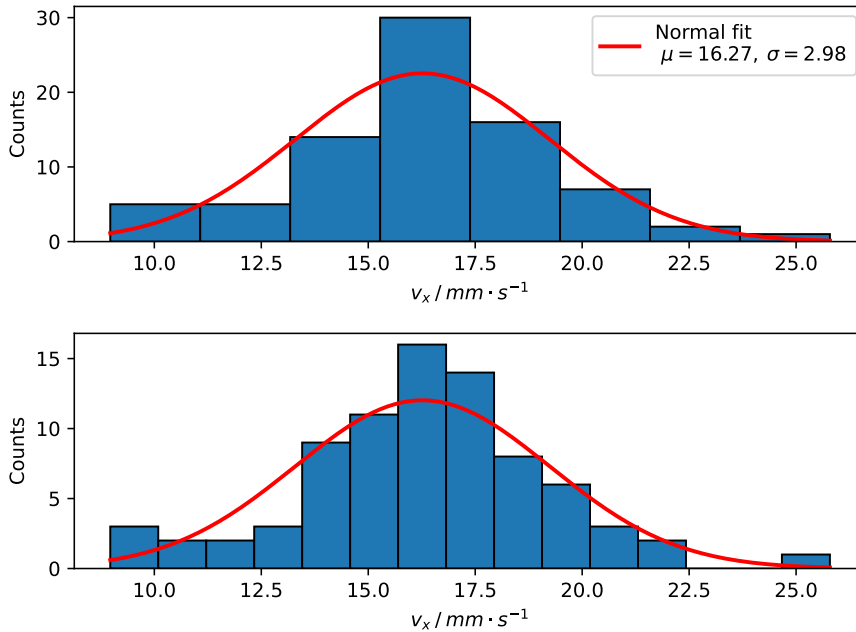
$$\gamma_3 = \frac{\mathbb{E}[(x - \mu)^3]}{(\mathbb{E}[(x - \mu)^2])^{3/2}}. \quad (6)$$

Platí, že šikmost normálního rozdělení je nulová [2]. Pro naše data vychází šikmost  $\gamma_3 = 0.008$ , z čehož plyne, že data jsou dobře centrovaná. Toto číslo nás utvrzuje v přesvědčení, že data pocházejí z normálního rozdělení.

### 3.5 Fitovaný histogram

Na závěr naší práce jsme nafitovali původní histogram pomocí hustoty pravděpodobnosti normálního rozdělení, kterou jsme správně nanormovali na náš histogram. Nafitovaný histogram je na obrázku 7. Vychází

$$\mu = 16.27 \text{ mm s}^{-1}, \sigma = 2.98 \text{ mm s}^{-1}. \quad (7)$$



Obrázek 7: Nafitovaný histogram rychlostí

## 4 Závěr

Na získaných datech rychlostí proudících částic ve vodě jsme ověřili, že data pravděpodobně pocházejí z normálního rozdělení,  $p$  hodnota Shapiro-Wilkova testu vyšla  $p = 0.20$ . Pro vyšší jistotu bychom potřebovali větší počet částic. Po rozdělení dat na dvě poloviny jsme provedli bodový odhad střední hodnoty  $\mu_1 = 16.14 \text{ mm s}^{-1}$  a odchylky  $\sigma_1 = 2.70 \text{ mm s}^{-1}$  a dále jsme zkonstruovali 95% konfidenční interval z první poloviny pro test hypotézy na střední hodnotu druhé poloviny. Tento test hypotézy je poněkud umělý, cílem bylo si především vyzkoušet provést nějaký test. Zjistili jsme, že střední hodnota  $\mu_2$  je v 95%-CI, tedy že oba datasety mají pravděpodobně stejnou střední hodnotu. Pro doplnění informací o datech jsme spočetli šikmost dat  $\gamma_3 = 0.008$  a nakonec jsme histogram nafitovali normální distribucí, čímž jsme zároveň provedli bodový odhad  $\mu = 16.27 \text{ mm s}^{-1}$  a  $\sigma_1 = 2.98 \text{ mm s}^{-1}$  pro všechna data.

## 5 Kód

Níže se nachází kód použitý pro zpracování. Adresa pro ukládání obrázků byla odstraněna a nahrazena textem „my personal filepath“.

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from scipy.stats import shapiro
4 from scipy.stats import t
5 from scipy.stats import skew
6 from scipy.stats import norm
7
8 #closes all previous plots
9 plt.close("all")
10
11 def t_confidence_interval(data, alpha=0.05):
12     #computing basic statistics and the (1-alpha)-CI
13     average = np.mean(data)
14     stdeva = np.std(data) #standard deviation
15     se = stdeva/np.sqrt(len(data)) #standard error
16     df = len(data)-1 #degrees of freedom to be passed to t-test
17     t_alpha_half = t.ppf(1-(alpha/2), df)
18     #returns mean, stdeva, se and lower and upper bound of 1-alpha CI
19     return (average, stdeva, se,
20             average + t_alpha_half*se, average - t_alpha_half*se)
21
22
23
24 # Loading data
25 data_raw = np.genfromtxt("data.csv", delimiter=";", skip_header=1)
26 data = data_raw[:, 5]
27 #splitting data into two batches for statistic testing
28 batch1 = data[:, 2]
29 batch2 = data[1::2]
30
31 ###Explorative analysis###
32 #Optimal no. of bins acc. to H. A. Sturges, J.
33 #American Statistical Association, 65--66 (1926).
34 m_opt = (np.log(len(data))/np.log(2))+1
35
36 #Plotting histograms
37 fig, ax = plt.subplots(nrows = 2, ncols=1)
38 counts1, bins1, _ = ax[0].hist(data, bins=int(np.ceil(m_opt)),
    edgecolor="black")
```



```

39 ax[0].set_xlabel(r'$v_x \text{ : / \text{ : mm \cdot s}^{-1}$')
40 ax[0].set_ylabel("Counts")
41 # ax[0].set_title("Histogram of velocities")
42 #My preferred no. of bins
43 counts2, bins2, _ = ax[1].hist(data, bins=15, edgecolor="black")
44 ax[1].set_xlabel(r'$v_x \text{ : / \text{ : mm \cdot s}^{-1}$')
45 ax[1].set_ylabel("Counts")
46
47 # Fit normal distribution
48 mu_fit, std_fit = norm.fit(data)
49
50 # Generate x values
51 x = np.linspace(min(data), max(data), 1000)
52 pdf = norm.pdf(x, mu_fit, std_fit)
53
54 # Scale PDF for first histogram
55 bin_width1 = bins1[1] - bins1[0]
56 pdf_scaled1 = pdf * len(data) * bin_width1
57 ax[0].plot(x, pdf_scaled1, color='red', linewidth=2,
58           label=f"Normal fit\n $\mu = \{mu\_fit:.2f\}, \text{ : } \sigma = \{std\_fit:.2f\}$")
59 ax[0].legend()
60
61 # Scale PDF for second histogram
62 bin_width2 = bins2[1] - bins2[0]
63 pdf_scaled2 = pdf * len(data) * bin_width2
64 ax[1].plot(x, pdf_scaled2, color='red', linewidth=2)
65
66 fig.tight_layout()
67 fig.savefig("my personal filepath", format="pdf")
68
69 #Plotting both batches
70 fig2, ax2 = plt.subplots(1,1)
71 ax2.hist(batch1, bins=15, color='blue', edgecolor="black", alpha =
       0.5, label='First batch')
72 ax2.hist(batch2, bins=15, color='red', edgecolor="black", alpha =
       0.5, label='Second batch')
73 ax2.set_xlabel(r'$v_x \text{ : / \text{ : mm \cdot s}^{-1}$')
74 ax2.set_ylabel("Counts")
75 # ax2.set_title('Both batches')
76 ax2.legend()
77 fig2.tight_layout()
78 fig2.savefig("my personal filepath", format="pdf")

```

```

79
80
81 ###Test of normality - Shapiro--Wilk test###
82 stat, p = shapiro(data)
83 print(f"Shapiro-Wilk test of all data: W={stat:.4f}, p={p:.4f}")
84
85 stat_batch1, p_batch1 = shapiro(batch1)
86 print(f"Shapiro-Wilk test of first batch: W={stat_batch1:.4f},
      p={p_batch1:.4f}")
87
88 stat_batch2, p_batch2 = shapiro(batch2)
89 print(f"Shapiro-Wilk test of second batch: W={stat_batch2:.4f},
      p={p_batch2:.4f}")
90
91 ###Interval estimate of mu and sigma###
92
93 #computing mu, sigma and rejection region from the first batch
94 mu, sigma, se, upper, lower = t_confidence_interval(batch1)
95 print("First batch's statistics below:")
96 print(f'Mu = {mu:.2f}, sigma = {sigma:.2f}, se = {se:.2f}, the 95%
      confidence interval for mu is: [{lower:.2f}, {upper:.2f}]')
97
98 #performing estimate of mu from the second batch
99 average2 = np.mean(batch2)
100 print(f"Mean for the second batch is: {average2:.2f}")
101 if average2 <=upper and average2>=lower:
102     print(f"Mu estimate mu = {average2:.2f} lays in the 95%
          confidence interval.")
103 else:
104     print(f"Mu estimate {average2:.2f} does not lay in the 95%
          confidence interval.")
105
106 ###computing skewness###
107 skew = skew(data)
108 print(f"Skewness is: {skew:.4f}.")
109
110 plt.show()
111 -----
112 VYSTUP Z KONZOLE
113
114 Shapiro-Wilk test of all data: W=0.9788, p=0.2045
115 Shapiro-Wilk test of first batch: W=0.9739, p=0.4752
116 Shapiro-Wilk test of second batch: W=0.9721, p=0.4174

```

```
117 First batch's statistics below:
118 Mu = 16.14, sigma = 2.70, se = 0.43, the 95% confidence interval for
    mu is: [15.28, 17.00]
119 Mean for the second batch is: 16.39
120 Mu estimate mu = 16.39 lays in the 95% confidence interval.
121 Skewness is: 0.0082.
```

## Reference

- [1] Kolektiv ZFP KVOF MFF UK. *Studijní text a pokyny k měření k úloze „Laserová dopplerovská anemometrie“*. Online. [cit. 2025-08-27]. Dostupné z: [https://physics.mff.cuni.cz/vyuka/zfp/\\_media/zadani/texty/txt\\_318.pdf](https://physics.mff.cuni.cz/vyuka/zfp/_media/zadani/texty/txt_318.pdf).
- [2] HRUŠKA, Petr. Histogram, momenty rozdělení, Presentace z předmětu „Úvod do praktické fyziky“, prezentováno v ak. roce 2023/2024. Online. S. 12. [cit. 2025-08-27]. Dostupné z: <https://physics.mff.cuni.cz/kfnt/vyuka/upf/hruska/UPF-prednaska4-PH.pdf>
- [3] ŠÁMAL, Robert. Skripta k předmětu NMAI059 Pravděpodobnost a statistika 1. Online. 2025. Dostupné z: <https://iuuk.mff.cuni.cz/~samal/vyuka/PSt1/skripta.pdf>. [cit. 2025-08-27].