

Lab 0 – rozkład normalny

Arkadiusz Kurnik, Jan Cichoń

Zadanie 1:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import chi2

samples = np.array([12.237, -9.712, -9.218, -7.235, -6.455, -4.869, -4.842, -4.407, -3.460, -2.527, -1.764, -1.711, -0.613,
                    0.252, 0.363, 1.193, 1.720, 2.185, 3.379, 5.496, 6.511, 8.722, 10.292, 19.126])

## Zadanie 1 ##
print("ZADANIE 1\n")

mean_x = samples.mean()
std_x = samples.std()
std_x_bessel = np.sqrt(np.var(samples, ddof=1))

print("a):")
print(f"Średnia arytmetyczna  $\mu = \{mean\_x\}$ ")
print(f"Odchylenie standardowe  $\hat{\sigma} = \{std\_x\}$ \n")

print("b):")
print(f"Średnia obliczona ręcznie  $\mu = \{samples.sum() / len(samples)\}$ ")
print(f"Odchylenie standardowe  $\hat{\sigma} = \{std\_x\_bessel\}$ \n")

#Wykres
plt.figure(figsize=(10, 6))
plt.hist(samples, bins=10, alpha=0.7, color='blue', edgecolor='black')

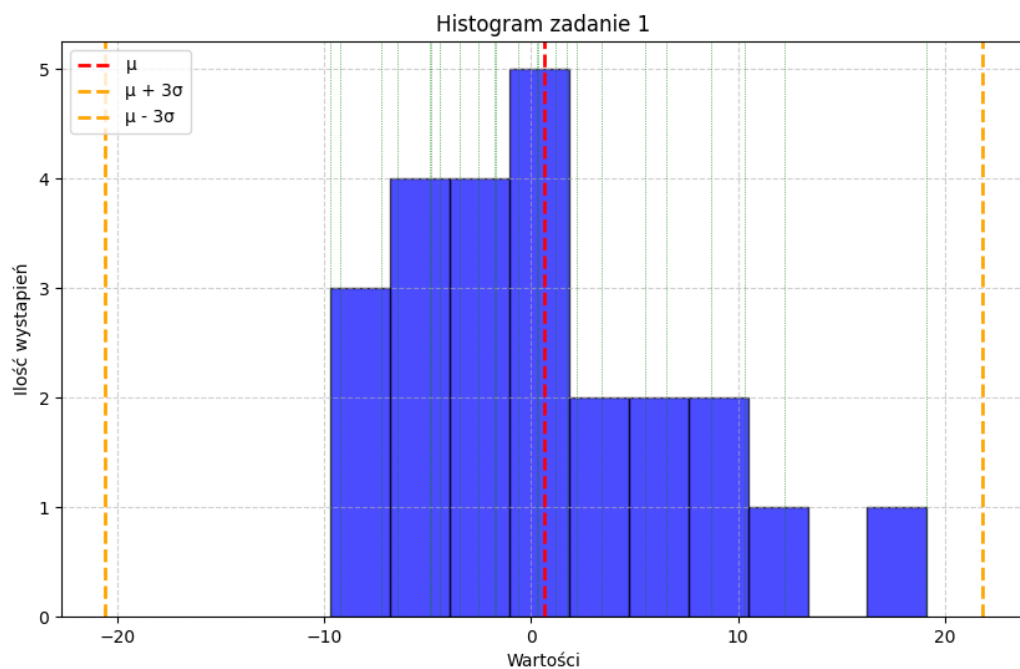
plt.axvline(mean_x, color='red', linestyle='dashed', linewidth=2, label='μ')
plt.axvline(mean_x + 3 * std_x_bessel, color='orange', linestyle='dashed', linewidth=2, label='μ + 3σ')
plt.axvline(mean_x - 3 * std_x_bessel, color='orange', linestyle='dashed', linewidth=2, label='μ - 3σ')

for value in samples:
    plt.axvline(value, color='green', linestyle='dotted', linewidth=0.5)

plt.title("Histogram zadanie 1")
plt.xlabel("Wartości")
plt.ylabel("Ilość wystąpień")
plt.grid(True, linestyle="---", alpha=0.6)
plt.legend()

plt.show()

print("Funkcja std() z bibliotki numpy realizuje wzór na odchylenie standardowe bez korekcji Bessela,\n"
      "więc wynik można uznać za prawidłowy dla bardzo dużej ilości próbek.\n"
      "W obliczeniach za pomocą wzorów została uwzględniona korekcja Bessela,\n"
      "pozwala ona skorygować odchylenie dla małych prób, zwiększając jego wartość, co daje lepszą estymację niepewności.\n")
```



ZADANIE 1

a):

Średnia arytmetyczna $\mu = 0.6109583333333336$

Odchylenie standardowe $\sigma = 6.911972833901854$

b):

Średnia obliczona ręcznie $\mu = 0.6109583333333336$

Odchylenie standardowe $\sigma = 7.060634418343209$

Funkcja `std()` z biblioteki `numpy` realizuje wzór na odchylenie standardowe bez korekcji Bessela, więc wynik można uznać za prawidłowy dla bardzo dużej ilości próbek. W obliczeniach za pomocą wzorów została uwzględniona korekcja Bessela, pozwala ona skorygować odchylenie dla małych próbek, zwiększając jego wartość, co daje lepszą estymację niepewności.

Zadanie 2:

```
## Zadanie 2 ##
print("ZADANIE 2\n")

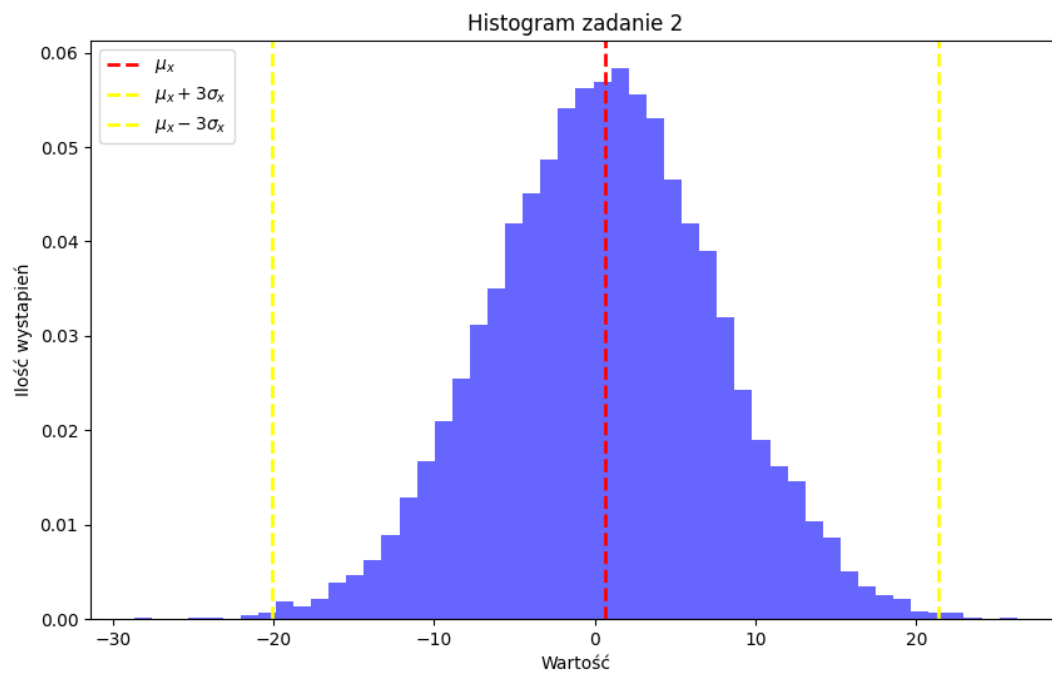
random_samples = np.random.normal(mean_x, std_x, 10000)

#Wykres
plt.figure(figsize=(10, 6))
plt.hist(random_samples, bins=50, density=True, alpha=0.6, color='blue')

plt.axvline(mean_x, color='red', linestyle='dashed', linewidth=2, label='${\mu}_x$')
plt.axvline(mean_x + 3 * std_x, color='yellow', linestyle='dashed', linewidth=2, label='${\mu}_x + 3{\sigma}_x$')
plt.axvline(mean_x - 3 * std_x, color='yellow', linestyle='dashed', linewidth=2, label='${\mu}_x - 3{\sigma}_x$')

plt.title("Histogram zadanie 2")
plt.xlabel("Wartość")
plt.ylabel("Ilość wystapień")
plt.legend()

plt.show()
```



Zadanie 3:

```
## Zadanie 3 ##
print("ZADANIE 3\n")

subset = np.random.choice(random_samples, 24, replace=False)
estimated_mean = np.mean(subset)
estimated_std = np.std(subset)

#Wykres
plt.hist(subset, bins=10, density=True, alpha=0.6, color='blue')

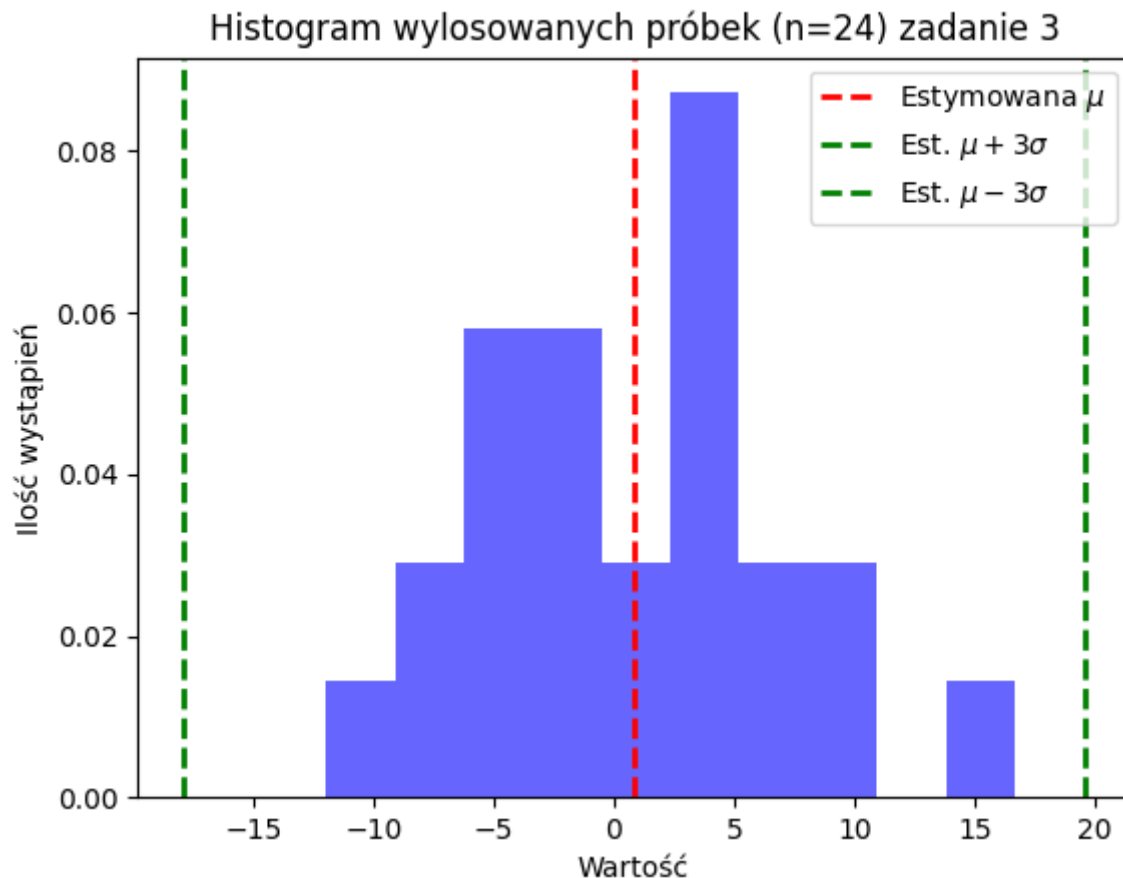
plt.axvline(estimated_mean, color='red', linestyle='dashed', linewidth=2, label='Estymowana  $\mu$ ')
plt.axvline(estimated_mean + 3 * estimated_std, color='green', linestyle='dashed', linewidth=2, label='Est.  $\mu + 3\sigma$ ')
plt.axvline(estimated_mean - 3 * estimated_std, color='green', linestyle='dashed', linewidth=2, label='Est.  $\mu - 3\sigma$ ')

plt.legend()
plt.title("Histogram wylosowanych próbek (n=24) zadanie 3")
plt.xlabel("Wartość")
plt.ylabel("Ilość wystąpień")

plt.show()

print(f"Estymowana średnia  $\mu$ : {estimated_mean}")
print(f"Estymowane odchylenie standardowe  $\sigma$ : {estimated_std}\n")

print("Dobranie zbyt małej próby powoduje zaburzenie rzeczywistych wyników obliczeń dla całej populacji.\n"
      "Odpowiednia ilość losowo wybranych próbek ze zbioru, również podlega rozkładowi normalnemu.\n")
```



ZADANIE 3

Estymowana średnia μ : 1.3354732476867357

Estymowane odchylenie standardowe σ : 4.836072493144196

Dobranie zbyt małej próby powoduje zaburzenie rzeczywistych wyników obliczeń dla całej populacji.
Odpowiednia ilość losowo wybranych próbek ze zbioru, również podlega rozkładowi normalnemu.

Zadanie 4:

```
## Zadanie 4 ##
print("ZADANIE 4\n")

# Przedział ufności dla średniej (95%)
t = 2.064
mean_lower = mean_x - std_x_bessel * t
mean_upper = mean_x + std_x_bessel * t

# Przedział ufności dla odchylenia standardowego (95%)
alpha = 0.05
df = len(samples) - 1 # Stopnie swobody

chi2_lower = chi2.ppf(alpha / 2, df)
chi2_upper = chi2.ppf(1 - alpha / 2, df)

std_lower = np.sqrt((df * std_x_bessel**2) / chi2_upper)
std_upper = np.sqrt((df * std_x_bessel**2) / chi2_lower)

print(f"Estymowana średnia  $\mu$ : {mean_x}")
print(f"95% przedział ufności dla średniej  $\mu$ : ({mean_lower}, {mean_upper})")
print(f"Estymowane odchylenie standardowe  $\sigma$ : {std_x_bessel}")
print(f"95% przedział ufności dla odchylenia standardowego  $\sigma$ : ({std_lower}, {std_upper})\n")
```

ZADANIE 4

```
Estymowana średnia  $\mu$ : 0.6109583333333336
95% przedział ufności dla średniej  $\mu$ : (-13.96219110612705, 15.184107772793718)
Estymowane odchylenie standardowe  $\sigma$ : 7.060634418343209
95% przedział ufności dla odchylenia standardowego  $\sigma$ : (5.487621009635744, 9.904379856713213)
```

Zadanie 5:

```
## Zadanie 5 ##
print("ZADANIE 5\n")

sorted_samples = np.sort(samples)
median = np.median(sorted_samples)

print(f"mediana m = {median}")

# Według tabeli z artykułu (95%)
k = 7
n = len(sorted_samples) - k + 1

print(f"Numery próbek: od {k} do {n}")
print(f"95% Przedział ufności dla mediany: ({sorted_samples[k]}, {sorted_samples[n]})\n")

print("Jeśli dane mają rozkład normalny średnia jest dobrym estymatorem, natomiast jeśli mają wartości odstające, lepsza będzie mediana")
```

ZADANIE 5

```
mediana m = -0.1805
Numery próbek: od 7 do 18
95% Przedział ufności dla mediany: (-3.46, 5.496)
```