

# Варіант 4

## Перевірка гіпотез про вид розподілу

1. За допомогою графічного представлення вибірки висунути гіпотезу про вид розподілу.

Вхідні дані подані на таблиці

122	102	124	93	96	103	90	82	99	94
117	101	98	103	117	102	84	93	113	99
89	118	90	106	113	103	125	99	97	106
109	112	104	109	110	86	107	100	90	92
103	97	95	94	90	96	100	101	120	101
99	104	103	118	97	107	114	89	87	93
96	104	82	77	104	101	103	97	101	104
91	92	94	102	110	105	102	96	108	114
94	95	104	84	113	97	111	98	108	112
117	113	105	90	110					

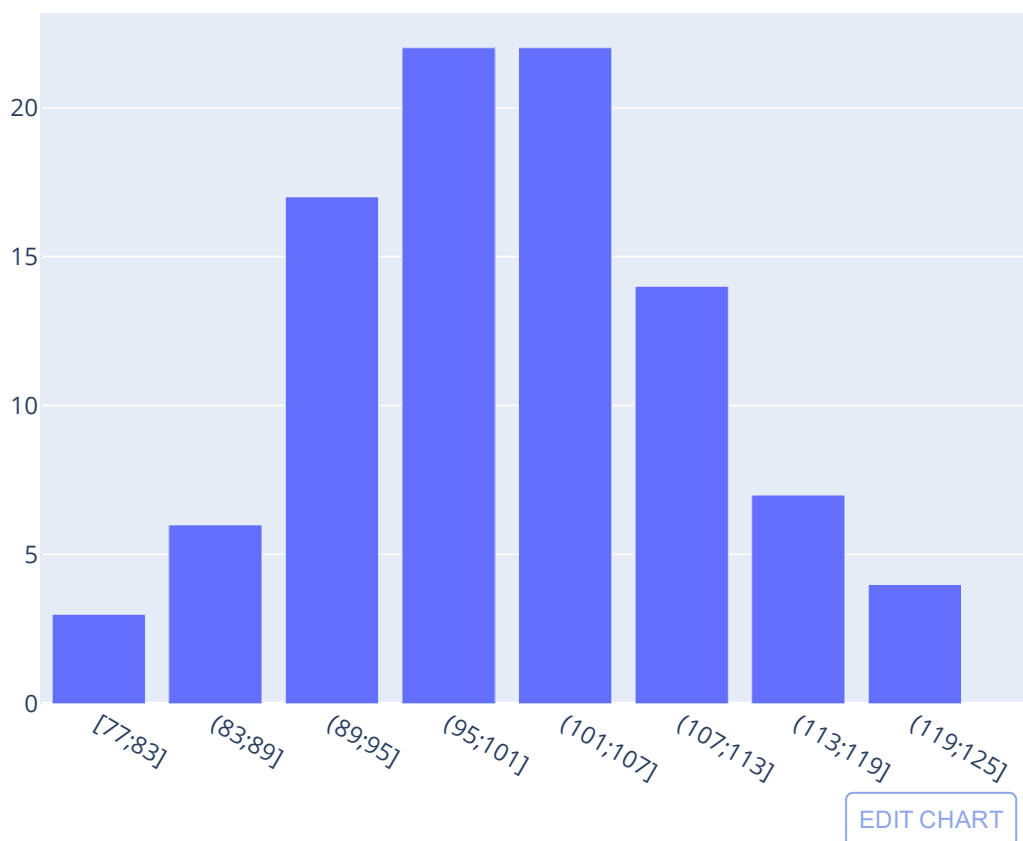
Вираховуємо частоту кожного з проміжків і зберігаємо його в словник `historgam`  
Будуємо гістограму по обчисленим даним

In [1]:

```
from sys import path
path.append('.')
import mathstat
import chart_studio.plotly as py
import plotly.graph_objects as go

data = [122, 102, 124, 93, 96, 103, 90, 82, 99, 94, 117, 101, 98, 103, 117, 102, 84,
93, 113, 99, 89, 118, 90, 106, 113, 103, 125, 99, 97, 106, 109, 112, 104, 109, 110,
86, 107, 100, 90, 92, 103, 97, 95, 94, 90, 96, 100, 101, 120, 101, 99, 104, 103, 118
, 97, 107, 114, 89, 87, 93, 96, 104, 82, 77, 104, 101, 103, 97, 101, 104, 91, 92, 94
, 102, 110, 105, 102, 96, 108, 114, 94, 95, 104, 84, 113, 97, 111, 98, 108, 112, 117
, 113, 105, 90, 110]
historgam = mathstat.histogramData(data)
h = mathstat.h_func(data)
plot_data = [go.Bar(x=list(historgam.keys()),
y=list(historgam.values()))]
py.iplot(plot_data, filename='HW5-basic_bar')
```

Out[1]:



Робимо припущення, що це нормальний розподіл.

## 2.Перевірити гіпотезу за допомогою критерію $\chi^2$ на рівні значущості $\alpha = 0.01$

Обчислюємо вибірку середню  $\bar{x}$  і вибіркве середньоквадратичне відхилення  $\sigma_x^*$ .

In [2]:

```
sample_mean = mathstat.calc_sample_mean(data)
sample_deviation = mathstat.math.sqrt(mathstat.calc_variance(data))
print(f"{sample_mean=}\n{sample_deviation=}")
```

```
sample_mean=101.46315789473684
sample_deviation=9.905671449720447
```

В нас є кілька інтервалів з відносною частотою  $< 5$  тому об'єднаємо їх **вручну**

In [3]:

```
polygon = mathstat.histogramData(data, polygon = True)
x_i, n_i = list(polygon.keys()), list(polygon.values())
x_i.append(x_i[-1]+h)
x_i = list(map(lambda i:i-h/2, x_i))
x_i = mathstat.np.delete(mathstat.np.array(x_i), [1, len(x_i)-2])
n_i[1], n_i[-2] = n_i[0]+n_i[1], n_i[-1]+n_i[-2]
n_i = mathstat.np.array(n_i[1:-1])
print(f"{x_i=}\n{n_i=}")
```

```
x_i=array([ 77.,  89.,  95., 101., 107., 113., 125.])
n_i=array([ 9, 17, 22, 22, 14, 11])
```

Пронормуємо X, тобто перейдемо до випадкової величини

$$Z = \frac{X - \bar{x}}{\sigma_x^*}$$

In [4]:

```
Z_i = (x_i-sample_mean)/sample_deviation
Z_i[0]=mathstat.np.NINF
Z_i[-1]=mathstat.np.inf
print(f"{Z_i=}")
```

```
Z_i=array([          -inf, -1.25818406, -0.65247045, -0.04675684,  0.558956
77,
          1.16467038,           inf])
```

Обчислюємо теоретичні частоти  $n'_i = n \cdot P_i$  де  $n$  – об'єм вибірки (сума всіх частот)

$$P_i = \Phi(z_{i+1}) - \Phi(z_i)$$

– функція Лапласа, значення якої знаходяться за таблицею, в нашому випадку за модулем `scipy.special.erf` але на практичних ми використовуємо нормоване значення, тому аргумент ділимо на  $\sqrt{2}$  а значення на 2, тому в модулі `mathstat` вже зроблена виправлена функція

In [5]:

```
P_i = mathstat.norm_erf(Z_i[1:]) - mathstat.norm_erf(Z_i[:-1])
print(f"{P_i=}")
n_prime_i = len(data)*P_i
print(f"{n_prime_i=}")
```

```
P_i=array([0.1041626 , 0.15288627, 0.22430465, 0.23055087, 0.16601939,
          0.12207622])
n_prime_i=array([ 9.89544697, 14.52419523, 21.30894163, 21.90233299, 15.
77184208,
          11.5972411 ])
```

Знаходимо спостережуване значення критерію

$$\chi^2_{\text{спост.}} = \sum \frac{(n_i - n'_i)^2}{n'_i}$$

In [6]:

```
chi_2 = sum((n_i - n_prime_i)**2 / n_prime_i)
alpha = 0.01
k = len(n_i) - 3
print(f"{chi_2=}\n{alpha=}\n{k=}")
```

```
chi_2=0.7557135579690212
alpha=0.01
k=3
```

За таблицею критичних точок розподілу  $\chi^2$  (в нашому випадку за функцією з модуля  $s$ ), за заданим рівнем значущості  $\alpha$  і кількістю ступенів вільності  $k$  знаходимо критичну точку

In [7]:

```
from scipy.stats import chi2
chi_critical = chi2.ppf(1-alpha, k)
print(f"Критична точка хі-квадрат = {chi_critical}")
```

Критична точка хі-квадрат = 11.344866730144373

Оскільки  $\chi_{\text{спост.}} < \chi_{\text{крит.}}$   $0.76 < 11.34$ , немає підстав відхилити гіпотезу про нормальний розподіл  $X$