



Міністерство освіти і науки України  
Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
Факультет інформатики та обчислювальної техніки  
Кафедра інформаційних систем та технологій

Лабораторна робота №2  
з дисципліни Аналіз даних з використанням мови Python  
Знайомство з бібліотекою SciPy. Статистичний аналіз даних

Виконав:  
студент групи ІА-11:  
Никифоров М. С.

Перевірила:  
ст. викладач  
Тимофєєва Ю.С.

## Завдання 1

Визначити на якій середній відстані від поселення спостерігаються жаби, а на якій – ні. Порівняти та проаналізувати отримані значення.

```
frog_Observed = frog_data[frog_data["pres.abs"] == 1]
frog_NOT_Observed = frog_data[frog_data["pres.abs"] == 0]

frog_Observed_distance = np.mean(frog_Observed["distance"]).round(0)
frog_NOT_Observed_distance = np.mean(frog_NOT_Observed["distance"]).round(0)

print("Середня відстань до населеного пункту у місцях, де жаби спостерігалися --- ",
print("Середня відстань до населеного пункту у місцях, де жаби НЕ спостерігалися --- ")

Середня відстань до населеного пункту у місцях, де жаби спостерігалися --- 854.0
Середня відстань до населеного пункту у місцях, де жаби НЕ спостерігалися --- 2573.0
```

Відповідно, можна зробити висновок: можливість спостерігати жаб вища у місцях, які знаходяться ближче до населених пунктів.

## Завдання 2

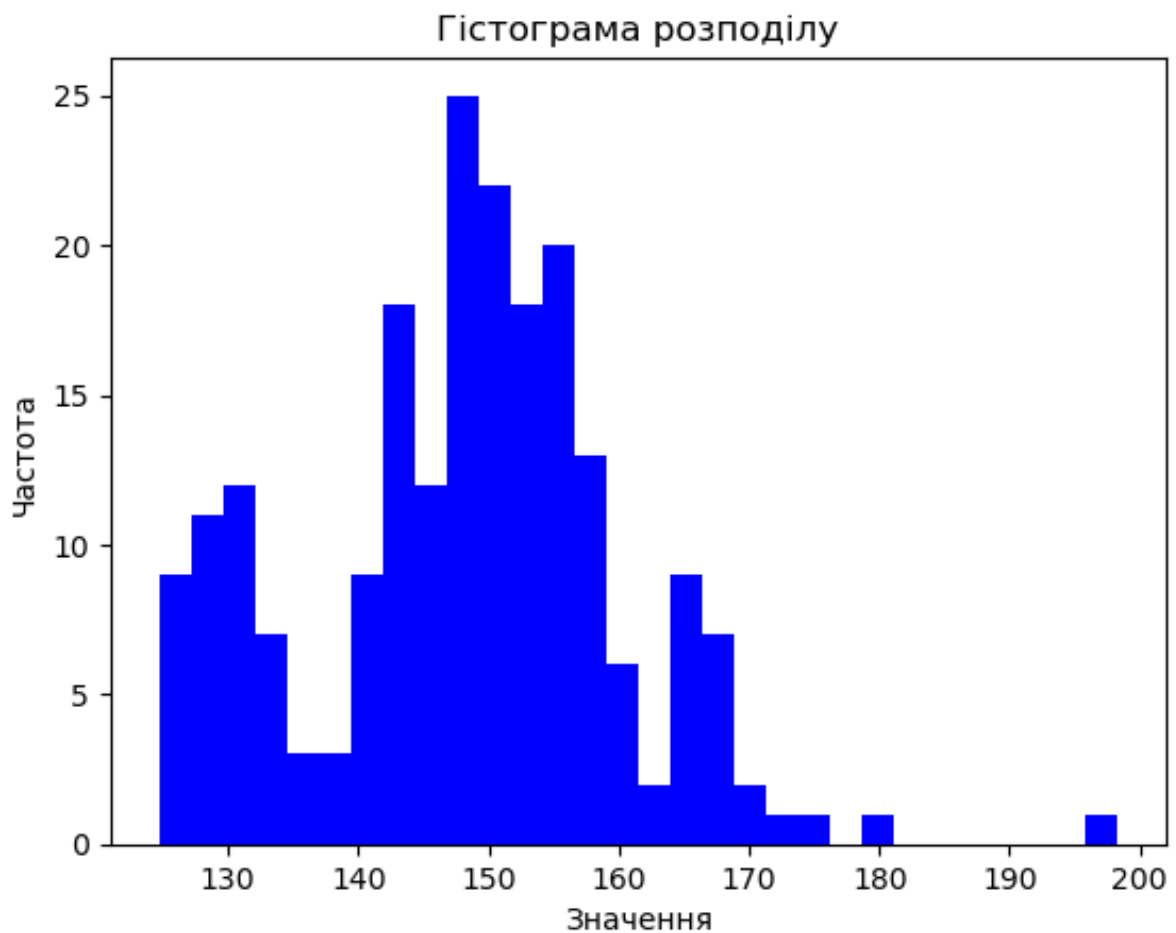
Перевірити чи нормально розподілена середня кількість опадів.

*Основна гіпотеза* - вибірка, яка описує середню кількість опадів нормально розподілена

*Альтернативна гіпотеза* - вибірка розподілена не нормально

*Рівень значущості* - 0.05

```
: avrain = np.array(frog_data["avrain"].round(2))
plt.hist(avrain, bins=30, color='b')
plt.xlabel('Значення')
plt.ylabel('Частота')
plt.title('Гістограма розподілу')
plt.show()
print(stats.normaltest(avrain).pvalue)
```



0.08013357573404803

Оскільки р-значення вище заданого рівня значущості, то ми **приймаємо** основну гіпотезу. Робимо висновок, що середня кількість опадів розподілена нормально

### Завдання 3

Визначити, чи є зв'язок між кількістю місць для розмноження та відстанню до поселення.

*Основна гіпотеза* - коефіцієнт кореляції дорівнює 0 [дані незалежні]

*Альтернативна гіпотеза* - коефіцієнт кореляції НЕ дорівнює 0

*Критерій перевірки* - Критерій Спірмена

*Рівень значущості* - 0.05

```
NoOfSites = np.array(frog_data["NoOfSites"].round(2))
distance = np.array(frog_data["distance"].round(2))
print(stats.spearmanr(NoOfSites, distance).pvalue)
```

```
1.1854086024923914e-14
```

Оскільки р-значення майже нульове, тобто менше нашого рівня значущості, то ми відхиляємо основну гіпотезу. Відповідно, приймаємо альтернативну гіпотезу, що стверджує про залежність наших даних.

## Завдання 4

Перевірити за допомогою статистичних гіпотез чи однакові середні висоти, на яких спостерігаються жаби і на яких – ні.

*Основна гіпотеза* - різниця математичних сподівань дорівнює нулю

*Альтернативна гіпотеза* - різниця математичних сподівань НЕ дорівнює нулю

Критерій перевірки -

$$T_0 = \frac{(\bar{x}_1 - \bar{x}_2) - c_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

*Рівень значущості* - 0.05

```
frog_Observed_Altitude = frog_Observed["altitude"]
frog_NOT_Observed_Altitude = frog_NOT_Observed["altitude"]

print(stats.ttest_ind(frog_Observed_Altitude, frog_NOT_Observed_Altitude).pvalue)

0.00046281302837254353
```

Оскільки р-значення менше нашого рівня значущості, то ми відхиляємо основну гіпотезу. Відповідно, приймаємо альтернативну гіпотезу. Отже, середні висоти у місцях де спостерігалися та не спостерігалися жаби - різні.

## **Висновки:**

У ході цієї лабораторної роботи ми ознайомились з основними функціями бібліотек NumPy та SciPy, які дозволяють проводити аналіз даних в наукових і дослідницьких проектах. Ми вивчили можливості цих бібліотек для виконання описової статистики, перевірки статистичних гіпотез та кореляційного аналізу.

Під час виконання роботи ми виконали практичні завдання з використанням функцій цих бібліотек, що дозволило нам зрозуміти їхні можливості і переваги в аналізі даних. Ми засвоїли важливі навички та інструменти для обробки та аналізу даних у наукових дослідженнях.