**Міністерство освіти і науки України**

**Національний технічний університет України**

**«Київський політехнічний інститут імені Ігоря Сікорського»**

**Факультет інформатики та обчислювальної техніки**

**Кафедра обчислювальної техніки**

**Лабораторна робота №2**

з дисципліни

«Аналіз даних з використанням мови Python»

Виконала: Викладач:

студентка групи ІП-03 Тимофєєва Ю.

Величко Ірина Євгенівна

Київ 2022

**Завдання**

Варіант №4

Файл frogs.csv.

1. На якій середній відстані від поселення спостерігаються жаби, а на

якій – ні, порівняти ці середні значення.

2. Перевірити чи нормально розподілена середня кількість опадів.

3. Чи є зв’язок між кількістю місць для розмноження та відстанню до

поселення?

4. Перевірити за допомогою статистичних гіпотез чи однакові середні

висоти, на яких спостерігаються жаби і на яких – ні?

**Код програми**

import pandas as pd

import scipy.stats

data = pd.read\_csv('frogs.csv')

avg = data.groupby('pres.abs').mean()

print("Average distance where frogs are present: ", avg.at[0, 'distance'])

print("Average distance where frogs are NOT present: ", avg.at[1, 'distance'])

print()

print("Testing 'avrain' values for normal distribution: ", scipy.stats.normaltest(data['avrain']))

print()

print(scipy.stats.normaltest(data['NoOfSites']))

print(scipy.stats.normaltest(data['distance']))

print(scipy.stats.spearmanr(data['NoOfSites'], data['distance']))

print()

IsPres = data.loc[data['pres.abs'] == 0, 'altitude']

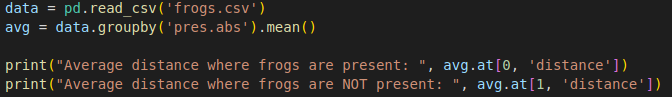
NotPres = data.loc[data['pres.abs'] == 1, 'altitude']

print("Null: ", scipy.stats.ttest\_ind(IsPres, NotPres))

print("Alternative: ", scipy.stats.ttest\_ind(IsPres, NotPres, alternative='less'))

**Результати та висновки**

1. На якій середній відстані від поселення спостерігаються жаби, а на якій – ні, порівняти ці середні значення.





Середня відстань, де спостерігаються жаби набагато більша за середню відстань, де жаби не спостерігаються.

1. Перевірити чи нормально розподілена середня кількість опадів.



Так як pvalue > 0.05, можна вважати що середня кількість опадів розподілена нормально.

1. Чи є зв’язок між кількістю місць для розмноження та відстанню поселення?





Перевіримо, що дві вибірки розподілені нормально. Так як в обох випадках значення pvalue << 0.05, набори даних не мають нормального розподілу. Тому використовуємо кореляцію Спірмена.

Коефіцієнт кореляцїї дорівнює -0.497, що вказує на зв’язок між наборами даних: коли відстань до поселення зменшується, кількість місць зростає.

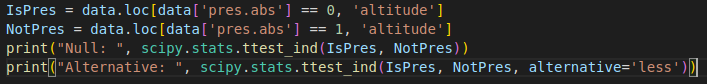
1. Перевірити за допомогою статистичних гіпотез чи однакові середні

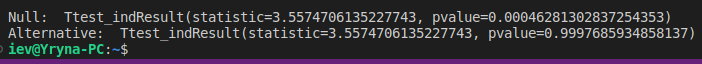
висоти, на яких спостерігаються жаби і на яких – ні?

Нульова гіпотеза: середні висоти однакові (Н0).

Альтернативна гіпотеза (одностороння): середня висота, де спостерігаються жаби, менша за середню висоту, де вони не спостерігаються (Н1).

Перевіримо гіпотезу за допомогою scipy.stats.ttest\_ind(); критерій відхилення основної гіпотези р < 0.05.





При використання відповідної функції (перевіряємо основну гіпотезу) вийшло pvalue << 0.05 - відкидаємо основну гіпотезу.

При перевірці альтернативної виходить, що pvalue = 0.999, отже середня висота, де спостерігаються жаби, майже завжди менша за середню висоту, де вони не спостерігаються (у заданій вибірці).