

Лабораторна робота №2

Статистичний аналіз даних

Мета роботи: Ознайомитись з основними функціями бібліотеки NumPy та SciPy для описової статистики, перевірки статистичних гіпотез, кореляційного аналізу та лінійної регресії.

Короткі теоретичні відомості

Статистичними називають гіпотези про вигляд розподілу генеральної сукупності або про параметри відомих розподілів.

Основною (нульовою) називають висунуту гіпотезу і позначають H_0 .

Альтернативною (конкурентною) називають гіпотезу, що суперечить основній, її позначають H_1 . Альтернативні гіпотези бувають двосторонніми та односторонніми.

Алгоритм перевірки гіпотез:

- 1) Визначити параметр, стосовно якого потрібно перевірити гіпотезу
- 2) Визначити основну гіпотезу H_0
- 3) Визначити гіпотезу H_1 , двосторонню чи односторонню, альтернативну до гіпотези H_0
- 4) Обрати статистичний критерій для перевірки
- 5) Визначити критерій відхилення основної гіпотези, наприклад, р-значення нижче за певний рівень значущості α (найчастіше 0,05)
- 6) Використати відповідну функцію і отримати значення статистичного критерію та р-значення
- 7) Зробити висновки: чи потрібно відхилити основну гіпотезу.

Коваріація та коефіцієнт кореляції Пірсона показують міру лінійного зв'язку між випадковими величинами. Коефіцієнт кореляції рангу Спірмена є непараметричним показником монотонності зв'язку між двома наборами даних. На відміну від кореляції Пірсона, кореляція Спірмена не передбачає нормального розподілу обох наборів даних. Коефіцієнт кореляції рангу Кендала або тау-коефіцієнт є непараметричним показником подібності впорядкування даних.

Проста лінійна регресійна модель:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Завдання до лабораторної роботи

Створити програму, яка за даними файлу виконує завдання відповідно до варіанту.

Оформити звіт. Звіт повинен містити:

- титульний лист;
- код програми;
- результати виконання коду;
- висновки щодо результатів (відповіді на завдання, прийняття/відхилення гіпотез, висновки про наявність зв'язку і т.д.)

Ім'я	Змінна
номер	
pres.abs	0=Жаби спостерігались 1=ні
northing	відстань від точки відліку на північ
easting	відстань від точки відліку на схід
altitude	висота (м)
distance	відстань до найближчого поселення (м)
NoOfPools	кількість можливих груп для розмноження
NoOfSites	кількість можливих місць для розмноження в радіусі 2 км
avrain	середня кількість опадів весною (мм)
meanmin	середня мінімальна температура
meanmax	середня максимальна температура

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?