МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНИ СЕМЕНА КУЗНЕЦЯ

ЗВІТ  
 о виконанній лабораторної роботи №6

з дисципліни:

«Основи математичного моделювання»

Варіант №4

Виконав:  
 Студент групи  
 факультету Інформаційні технології

спеціальності Кібербезпека

Ф.І.П. Бойко В.В.

Перевірила: Шаповалова O.O.

Харків-2023

**1. Перевірка вибірки на гетероскедастичність за допомогою критерію однорідності вибірки Бартлета**

Перевірка гетероскедастичності за допомогою тесту Бартлетта:

У цьому коді ми використовуємо тест Бартлетта для перевірки гетероскедастичності між змінними X1, X2 та Y у наших даних.

1. Спочатку імпортуємо бібліотеки: **pandas** для роботи з даними та **scipy.stats** для використання тесту Бартлетта.
2. Ми створюємо DataFrame **data**, який містить наші дані.
3. Далі ми використовуємо функцію **bartlett()** для обчислення статистики тесту та p-значення для кожної змінної окремо (X1, X2, Y).
4. На виході ми отримуємо значення статистики та p-значення, які можуть бути використані для визначення гетероскедастичності.

Код:

import numpy as np

from scipy.stats import levene

n = 100 # Розмір вибірки

# Генеруємо випадкові дані для трьох змінних (X1, X2, Y)

data = {

'X1': np.random.normal(50, 10, n),

'X2': np.random.normal(10, 5, n),

'Y': np.random.normal(130, 15, n)

}

# Виконуємо тест Лівеня для всіх змінних

stat, p\_value = levene(data['X1'], data['X2'], data['Y'])

print(f'Statistic = {stat}, p-value = {p\_value}’)

Output:

Statistic = 27.42887060976287, p-value = 1.1717010859924181e-11

**2. Перевірка вибірки на гетероскедастичність за допомогою тесту Голдфельда–Квандта**

Перевірка гетероскедастичності за допомогою тесту Голдфельда-Квандта:

В цьому коді ми використовуємо тест Голдфельда-Квандта для перевірки гетероскедастичності між двома групами даних (перші 10 спостережень і наступні 10).

1. Ми імпортуємо необхідні бібліотеки та створюємо DataFrame **data**, як у попередньому завданні.
2. Розділяємо дані на дві групи: **group1** та **group2**.
3. Далі використовуємо функцію **fligner()** для обчислення статистики тесту та p-значення для кожної змінної окремо (X1, X2, Y) між двома групами.
4. Результати тесту вказують на те, чи є статистично значущі відмінності у дисперсії між двома групами.

Щоб код запрацював, треба додати до першого коду:

from scipy.stats import fligner

# Розділімо дані на дві групи для тесту Голдфельда-Квандта

group1 = data.iloc[:10]

group2 = data.iloc[10:]

# Виконуємо тест Голдфельда-Квандта для кожної змінної

stat, p\_value = fligner(group1['X1'], group2['X1'], group1['X2'], group2['X2'], group1['Y'], group2['Y'])

print(f'Statistic = {stat}, p-value = {p\_value}')

І тепер код працює:

Код:

import pandas as pd

from scipy.stats import bartlett

# Створення DataFrame з даними

data = pd.DataFrame({

'X1': [65, 47, 61, 44, 47, 61, 45, 47, 58, 57, 69, 46, 55, 59, 52, 58, 47, 65, 59, 58],

'X2': [10, 5, 10, 13, 9, 14, 12, 15, 13, 27, 8, 8, 24, 8, 8, 15, 12, 16, 23, 16],

'Y': [113, 124, 124, 122, 128, 140, 117, 113, 122, 139, 126, 120, 125, 118, 122, 133, 136, 146, 148, 136]

})

# Виконуємо тест Бартлетта для всіх змінних

stat, p\_value = bartlett(data['X1'], data['X2'], data['Y'])

print(f'Statistic = {stat}, p-value = {p\_value}')

from scipy.stats import fligner

# Розділімо дані на дві групи для тесту Голдфельда-Квандта

group1 = data.iloc[:10]

group2 = data.iloc[10:]

# Виконуємо тест Голдфельда-Квандта для кожної змінної

stat, p\_value = fligner(group1['X1'], group2['X1'], group1['X2'], group2['X2'], group1['Y'], group2['Y'])

print(f'Statistic = {stat}, p-value = {p\_value}’)

Output:

Statistic = 6.050694013097125, p-value = 0.0485409743112169

Statistic = 7.641797775076178, p-value = 0.17711315272208344

**3. Перевірка на гомоскедастичність для вибірки обсягом 50-100 дослідів**

Генерую замість знаходження у інтернеті, я ж айтівець.

Перевірка гомоскедастичності для згенерованої вибірки:

У цьому коді ми генеруємо випадкові дані для трьох змінних (X1, X2, Y) та використовуємо тест Лівеня для перевірки гомоскедастичності.

1. Ми імпортуємо необхідні бібліотеки та генеруємо випадкові дані для трьох змінних, зазначаючи середні значення та стандартні відхилення для кожної змінної.
2. Потім ми використовуємо функцію **levene()** для обчислення статистики тесту та p-значення для кожної змінної окремо (X1, X2, Y).
3. Результати тесту допомагають визначити, чи є гомоскедастичність (рівність дисперсій) серед різних змінних.

Код:

import numpy as np

from scipy.stats import levene

n = 100 # Розмір вибірки

# Генеруємо випадкові дані для трьох змінних (X1, X2, Y)

data = {

'X1': np.random.normal(50, 10, n),

'X2': np.random.normal(10, 5, n),

'Y': np.random.normal(130, 15, n)

}

# Виконуємо тест Лівеня для всіх змінних

stat, p\_value = levene(data['X1'], data['X2'], data['Y'])

print(f'Statistic = {stat}, p-value = {p\_value}')

Output:

Statistic = 32.69600381642888, p-value = 1.4665909283481577e-13

Аналізуємо результати тестів на гетероскедастичність та гомоскедастичність на основі статистики та p-значень:

Перше завдання:

* + Статистика тесту Бартлетта = 27.43
  + p-значення = 1.17e-11

Результати першого завдання свідчать про те, що гетероскедастичність між змінними X1, X2 та Y є статистично значущою, оскільки p-значення дуже мале (менше за рівень значущості 0.05).

Друге завдання:

* + Статистика тесту Голдфельда-Квандта для X1 = 6.05
  + p-значення для X1 = 0.0485
  + Статистика тесту Голдфельда-Квандта для X2 = 7.64
  + p-значення для X2 = 0.1771

Результати другого завдання показують, що немає статистично значущих відмінностей у дисперсії між двома групами даних (перші 10 спостережень і наступні 10) для змінних X1 та X2, оскільки p-значення не менше за рівень значущості 0.05.

Третє завдання:

* + Статистика тесту Лівеня = 32.70
  + p-значення = 1.47e-13

Результати третього завдання свідчать про те, що гомоскедастичність (рівність дисперсій) серед різних змінних (X1, X2, Y) в згенерованих даних є статистично значущою, оскільки p-значення дуже мале (менше за рівень значущості 0.05).

Загальний висновок:

* У першому завданні спостерігається гетероскедастичність між змінними X1, X2 та Y.
* У другому завданні немає статистично значущих відмінностей у дисперсії між двома групами для змінних X1 та X2.
* У третьому завданні спостерігається гомоскедастичність серед різних змінних X1, X2 та Y в згенерованих даних.