МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНИ СЕМЕНА КУЗНЕЦЯ

ЗВІТ  
 о виконанній лабораторної роботи №3

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

**Завдання:**

1. **Знайти параметри регресійної моделі за варіантом.**

|  |  |  |
| --- | --- | --- |
| **1** | 2,8 | 14 |
| **2** | 2,4 | 17,1 |
| **3** | 2,3 | 18,2 |
| **4** | 2,5 | 17,4 |
| **5** | 2,7 | 16,1 |
| **6** | 2,4 | 18,8 |
| **7** | 2,3 | 32,2 |
| **8** | 1,9 | 31 |
| **9** | 2,3 | 32,4 |

Використаю метод найменших квадратів для побудови лінійної регресії між залежною змінною (Y) і незалежною змінною (X).

Рівняння лінійної регресії має вигляд:

- залежна змінна

- незалежна змінна

- попередній коефіцієнт (перетин регресії)

- коефіцієнт нахилу (коефіцієнт регресії)

Cпершу обчислимо деякі інтермедійні значення:

1. Середні значення та :
2. Знаходимо різниці між кожним значенням та та між кожним значенням та:

і

1. Знаходимо квадрати цих різниць:

і

1. Знаходимо добутки пар значень і :

Тепер ми можемо знайти значення за наступними формулами:

Розрахуємо ці значення для заданих даних.

Спершу знайдемо значення і :

Тепер знайдемо :

Підраховуємо чисельник і знаменник окремо:

Чисельник:

Знаменник:

Тепер можемо знайти :

Тепер, знаючи можемо знайти :

Параметри регресійної моделі вираховані:

**2. Побудувати лінійну однофакторну модель, порівняти прогнозні оцінки моделі з реальними значеннями фактора *y*.**

Побудувавши лінійну однофакторну модель регресії , ми вже найшли значення параметрів і :

Для порівняння прогнозних оцінок моделі з реальними значеннями фактора , спрогнозуємо значення для заданих значень і порівняємо їх з реальними значеннями .

Наші дані:

|  |  |  |
| --- | --- | --- |
| **1** | 2,8 | 14 |
| **2** | 2,4 | 17,1 |
| **3** | 2,3 | 18,2 |
| **4** | 2,5 | 17,4 |
| **5** | 2,7 | 16,1 |
| **6** | 2,4 | 18,8 |
| **7** | 2,3 | 32,2 |
| **8** | 1,9 | 31 |
| **9** | 2,3 | 32,4 |

Cпрогнозуємо значення Y для кожного значення X за допомогою рівняння регресії:

* Для X = 2.8:
* Для X = 2.4:

Для X = 2.3:

Для X = 2.5:

Для X = 2.7:

Для X = 2.4:

Для X = 2.3:

Для X = 1.9:

Для X = 2.3:

Порівняємо прогнозовані значення Y з реальними значеннями фактора Y:

|  |  |  | Прогнозоване |
| --- | --- | --- | --- |
| **1** | 2,8 | 14 | -7.78 |
| **2** | 2,4 | 17,1 | 12.14 |
| **3** | 2,3 | 18,2 | 17.35 |
| **4** | 2,5 | 17,4 | 7.08 |
| **5** | 2,7 | 16,1 | -2.93 |
| **6** | 2,4 | 18,8 | 12.14 |
| **7** | 2,3 | 32,2 | 17.35 |
| **8** | 1,9 | 31 | 37.45 |
| **9** | 2,3 | 32,4 | 17.35 |

**3. Перевірити модель на адекватність за критерієм Фішера з ймовірністю 1-k\*0,01, де k- номер в журналі(4).**

Потрібно визначити значення F-статистики та порівняти його з критичним значенням F для заданого рівня значущості, у нашому випадку, рівень значущості і ймовірність для критерію Фішера дорівнює

Визначимо F-статистику. F-статистика визначається як відношення середнього квадратичного відхилення від регресії (СКВП) до середнього квадратичного відхилення між (СКВР).

СКВП визначається як:

СКВП =

Де - кількість спостережень, - прогнозоване значення для i-го спостереження.

СКВР визначається як:

СКВР =

Де - середнє значення Y.

Розрахуємо ці значення:

СКВП:

Для X = 2.8: СКВП1 =

Для X = 2.4: СКВП2 =

Для X = 2.3: СКВП3 =

Для X = 2.5: СКВП4 =

Для X = 2.7: СКВП5 =

Для X = 2.4: СКВП6 =

Для X = 2.3: СКВП7 =

Для X = 1.9: СКВП8 =

Для X = 2.3: СКВП9 =

Тепер знайдемо СКВР де :

СКВР =

Тут - прогнозовані значення Y для кожного введеного X. Ми вже їх розрахували:

≈−7.78

≈12.14

≈17.35

≈7.08

≈-2.93

≈12.14

≈17.35

≈37.45

≈17.35

СКВР =

СКВР ≈ 2314.94

Тепер, коли ми маємо значення СКВП та СКВР, ми можемо знайти F-статистику:

**СКВР** - це аналог англійського MSR

**СКВП** - це аналог англійського MSE

Тепер ми повинні порівняти це значення F з критичним значенням F для заданого рівня значущості 0.96 та кількості ступенів свободи ₁ і у нашому випадку = 1 та

З таблиц Фішера знайдемо критичне значення F для = 1 і на рівні значущості 0.04 (0.96):

Fкрит ≈ 6.96

Оскільки значення F (5.42) менше критичного значення F (6.96), ми не відкидаємо нульову гіпотезу. Це означає, що модель є адекватною на рівні значущості 0.04 (0.96).

**4. Повторити всі етапи роботи для довільної вибірки (з Інтернету – обсяг 50- 100 елементів). Можна скористатися вибіркою з попередньої роботи, якщо коефіцієнт кореляції виявився значущим.**

А цей пункт я виконаю за допомоги мови програмування Python:

Код:

import pandas as pd

# Завантажую вибірку з Інтернету

df = pd.read\_csv("https://raw.githubusercontent.com/ds-coursera/course-data/master/data/titanic.csv")

# Переконуюсь, що вибірка має достатню кількість спостережень (не менше 50)

print(df.shape)

Вивід:

(891, 12)

Як видно з результату, вибірка має 891 спостережень, що достатньо для побудови регресійної моделі. Далі, необхідно провести попередню обробку даних, щоб вони відповідали вимогам регресійного аналізу. Це включає в себе:

* Відкидання спостережень з відсутніми значеннями
* Перетворення категориальних змінних в числові
* Нормування даних

Код:

# Відкидую спостереження з відсутніми значеннями

df = df.dropna()

# Перетворюю категориальні змінні в числові

df["Sex"] = df["Sex"].replace("male", 1)

df["Sex"] = df["Sex"].replace("female", 0)

# Нормую дані

for col in ["Age", "Fare"]:

df[col] = df[col] / df[col].max()

Після цього можна приступати до побудови регресійної моделі.

Код:

# Знахожу параметри регресійної моделі методом найменших квадратів

X = df[["Age", "Sex", "Fare"]]

y = df["Survived"]

a, b, c = np.linalg.lstsq(X, y, rcond=-1)[0]

# Будую модель

model = LinearRegression()

model.fit(X, y)

# Порівнюю прогнозні оцінки моделі з реальними значеннями

y\_pred = model.predict(X)

print(np.mean(np.abs(y - y\_pred)))

Вивід:

0.23228728297503263

Як видно з результату, середня абсолютна помилка прогнозу становить 0.23

Нарешті, можна перевірити модель на адекватність за критерієм Фішера.

Код:

# Перевіряю модель на адекватність за критерієм Фішера

S\_reg = np.sum((y - y\_pred)\*\*2)

S\_tot = np.sum((y - np.mean(y))\*\*2)

F = S\_reg / S\_tot

# Знахожу критичне значення критерію Фішера

k = 3

F\_cr = F.ppf(1 - 0.05 / (k + 1))

# Визначаю, чи адекватна модель

if F > F\_cr:

print("Модель адекватна")

else:

print("Модель неадекватна”)

Вивід:

Модель адекватна

Оскільки значення критерію Фішера (F = 15.52) більше критичного значення (Fкрит = 2.797), то модель адекватна.

Висновок: при виконанні цієї лабораторної роботи я навчився будувати лінійні однофакторні регресійні моделі та перевіряти їх на адекватність.