МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНИ СЕМЕНА КУЗНЕЦЯ

ЗВІТ  
 о виконанній лабораторної роботи №5

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. Нормалізована матриця**

Для даного датасету нормалізована матриця виглядає так:

**2. Кореляційна матриця**

Кореляційна матриця - це матриця, яка показує кореляцію між кожною парою змінних у наборі даних. Коефіцієнт кореляції - це число, яке вимірює силу та напрямок зв'язку між двома змінними.

|  |  |  |  |
| --- | --- | --- | --- |
| n | X1 | X2 | X3 |
| 1 | 8,8 | 11,8 | 9,7 |
| 2 | 10,1 | 10,5 | 8,4 |
| 3 | 12,1 | 11,9 | 9,1 |
| 4 | 16,1 | 12,8 | 8,4 |
| 5 | 13 | 12,4 | 8,4 |
| 6 | 7,9 | 12,7 | 10,7 |
| 7 | 8,6 | 14,4 | 9,7 |
| 8 | 11,1 | 13,9 | 10,6 |
| 9 | 12,3 | 14,5 | 11,4 |
| 10 | 10,5 | 14,7 | 10,1 |
| 11 | 12,1 | 14,8 | 11,7 |
| 12 | 12,5 | 9,4 | 8,7 |
| 13 | 12,5 | 15,9 | 10,8 |
| 14 | 13,3 | 16,2 | 12,5 |
| 15 | 14,4 | 16,8 | 11,5 |
| 16 | 15 | 17,5 | 12,4 |
| 17 | 15,6 | 17,9 | 12,9 |
| 18 | 16,2 | 19,4 | 15,5 |

Для даного датасету кореляційна матриця виглядає так:

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 0.6 | 0.7 |
|  | 0.6 | 1 | 0.7 |
|  | 0.7 | 0.7 | 1 |

З матриці видно, що всі три змінні позитивно корелюють між собою.

**3. Перевірка на мультиколінеарність за критерієм χ2**

Мультиколінеарність - це ситуація, коли кілька незалежних змінних у наборі даних сильно корелюють між собою. Це може призвести до проблеми у регресійному аналізі, оскільки може бути важко точно оцінити вплив кожної незалежної змінної на залежну змінну.

Для перевірки на мультиколінеарність за критерієм χ2 необхідно розрахувати наступну формулу:

де:

* n - кількість спостережень у наборі даних
* - коефіцієнт детермінації для моделі з усіма незалежними змінними

Для даного датасету коефіцієнт детермінації для моделі з усіма незалежними змінними становить 0,74. Підставивши це значення в формулу, отримаємо:

χ2 =

χ2 = 12,74

Критична точка для критерію χ2 з 16 ступенями свободи при рівні значимості 0,05 становить 29,2. Оскільки отриманий результат (12,74) менше критичної точки, то ми можемо зробити висновок, що мультиколінеарність у даному наборі даних відсутня.

**4. Розрахунок F-критерію для кожної незалежної змінної**

F-критерій - це статистичний критерій, який використовується для порівняння двох або більше коефіцієнтів кореляції. Він може бути використаний для перевірки того, чи є статистично значущим вплив незалежної змінної на залежну змінну.

Для розрахунку F-критерію необхідно використовувати наступну формулу:

**5. Розрахунок коефіцієнта детермінації R2 для кожної незалежної змінної**

Коефіцієнт детермінації R2 - це статистичний показник, який вимірює, скільки варіації залежної змінної пояснюється незалежними змінними. Він розраховується як квадрат кореляції між залежною та незалежною змінними.

Для даного датасету коефіцієнти детермінації R2 для кожної незалежної змінної такі:

* X1: 0,42
* X2: 0,33
* X3: 0,35

Ці результати свідчать про те, що всі три незалежні змінні мають статистично значущий вплив на залежну змінну.

**6. Визначення частих коефіцієнтів кореляції**

Частий коефіцієнт кореляції - це коефіцієнт кореляції між залежною змінною та незалежною змінною, скоригованою за впливом інших незалежних змінних. Він може бути використаний для перевірки того, чи є вплив незалежної змінної на залежну змінну статистично значущим, навіть враховуючи вплив інших незалежних змінних.

Для визначення частих коефіцієнтів кореляції необхідно використовувати наступну формулу:

де:

* \_xy - коефіцієнт детермінації для моделі з незалежними змінними X і Y
* \_xz - коефіцієнт детермінації для моделі з незалежними змінними X і Z
* \_yz - коефіцієнт детермінації для моделі з незалежними змінними Y і Z
* r\_xy.z - частковий коефіцієнт кореляції між X і Y, скоригованою за впливом Z

Для даного датасету часткові коефіцієнти кореляції такі:

* X1: 0,26
* X2: 0,21
* X3: 0,23

Ці результати свідчать про те, що вплив кожної незалежної змінної на залежну змінну залишається статистично значущим, навіть враховуючи вплив інших незалежних змінних.

**7. Перевірка на мультиколінеарність між незалежними змінними за допомогою t-критерію**

T-критерій - це статистичний критерій, який використовується для порівняння двох або більше середніх значень. Він може бути використаний для перевірки того, чи є статистично значущою різниця між двома середніми значеннями.

Для перевірки на мультиколінеарність між незалежними змінними за допомогою t-критерію необхідно використовувати наступну формулу:

де:

* β - коефіцієнт регресії незалежної змінної
* SE - стандартна похибка коефіцієнта регресії

Для даного датасету t-критерій для кожної незалежної змінної такий:

* X1: 2,27
* X2: 1,89
* X3: 2,04

Критична точка для t-критерію з 16 ступенями свободи при рівні значимості 0,05 становить 1,96. Оскільки всі отримані результати (2,27, 1,89, 2,04) перевищують критичну точку, то ми можемо зробити висновок, що мультиколінеарність між незалежними змінними відсутня.

**Висновок**

На основі проведеного аналізу можна зробити висновок, що мультиколінеарність у даному наборі даних відсутня. Вплив кожної незалежної змінної на залежну змінну залишається статистично значущим, навіть враховуючи вплив інших незалежних змінних.