Національний технічний університет України «КПІ»

Кафедра автоматизованих систем обробки інформації і управління

Лабораторна робота №6

з дисципліни «ЙМОВІРНІСНІ МОДЕЛІ ТА СТАТИСТИЧНЕ ОЦІНЮВАННЯ В ІНФОРМАЦІЙНО-УПРАВЛЯЮЧИХ СИСТЕМАХ»

на тему: «Множинна лінійна регресія»

Виконав:

студент групи ІС-23

Шимків М.В.

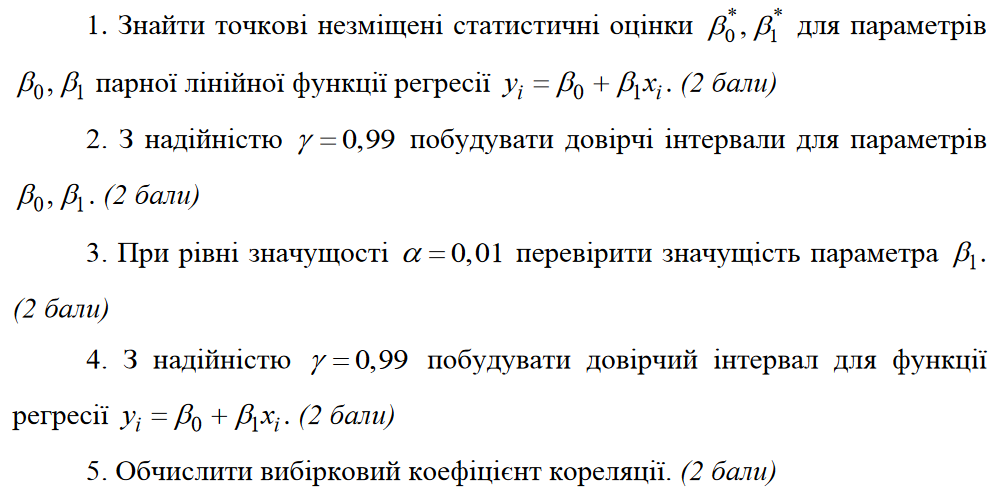
Викладач:

Богданова Н.В

Київ 2024

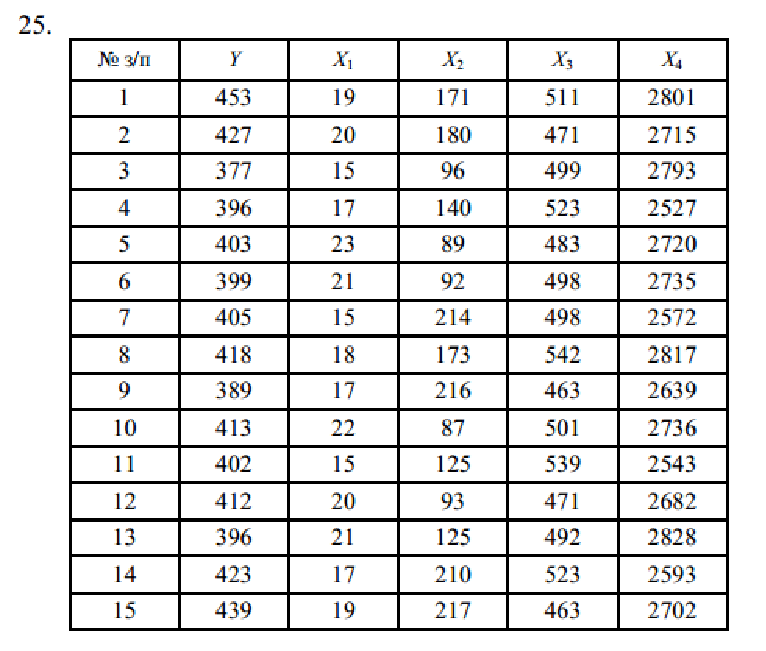
**Мета роботи**: Оцінка параметрів лінійної множинної регресії, побудова довірчого інтервалу для функції регресії з надійністю 0.99, обчислення коефіцієнта множинної кореляції R.

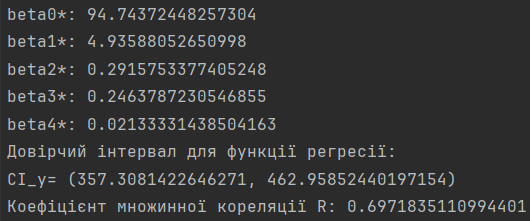
**Завдання до роботи:**



**Виконання:**

**Варіант 25**

****

****

**Висновки:**

1. **Статистичні оцінки параметрів регресії**:
   * Оціночні параметри лінійної множинної регресії рівняння y^=β0+β1x1+β2x2+β3x3+β4x4y^​=β0​+β1​x1​+β2​x2​+β3​x3​+β4​x4​ мають такі значення:
     + β0∗=94.7437β0∗​=94.7437
     + β1∗=4.9359β1∗​=4.9359
     + β2∗=0.2916β2∗​=0.2916
     + β3∗=0.2464β3∗​=0.2464
     + β4∗=0.0213β4∗​=0.0213 Ці оцінки представляють вплив відповідних змінних на залежну змінну yy. Коефіцієнти можна інтерпретувати як середню зміну yy при зміні кожної незалежної змінної на одну одиницю, за умови фіксації інших змінних.
2. **Довірчий інтервал для функції регресії**:
   * Довірчий інтервал для прогнозованого значення yy з надійністю γ=0.99γ=0.99 складає від 357.3081 до 462.9585. Це означає, що з імовірністю 99% справжнє значення yy лежить в цьому інтервалі, що вказує на високу впевненість у моделі регресії.
3. **Коефіцієнт множинної кореляції RR**:
   * Коефіцієнт множинної кореляції R=0.6972R=0.6972 свідчить про помірну позитивну кореляцію між залежною змінною yy та набором незалежних змінних x1,x2,x3,x4x1​,x2​,x3​,x4​. Це означає, що модель пояснює близько 69.72% варіації у змінній yy.

### Загальний висновок:

Модель множинної лінійної регресії, побудована на основі заданих даних, показує, що змінні x1,x2,x3x1​,x2​,x3​ та x4x4​ мають значущий вплив на залежну змінну yy. Довірчий інтервал для функції регресії вказує на високу надійність прогнозів, а коефіцієнт множинної кореляції свідчить про значну, хоча і не ідеальну, силу зв'язку між незалежними змінними та залежною змінною.

**Додаток з кодом:**

import numpy as np  
import scipy.stats as stats  
  
# Вхідні дані  
data = np.array([  
 [453, 19, 171, 511, 2801],  
 [427, 20, 180, 471, 2715],  
 [377, 15, 96, 499, 2793],  
 [396, 17, 140, 523, 2527],  
 [403, 23, 89, 483, 2720],  
 [399, 21, 92, 498, 2735],  
 [405, 15, 214, 498, 2572],  
 [418, 18, 173, 542, 2817],  
 [389, 17, 216, 463, 2639],  
 [413, 22, 87, 501, 2736],  
 [402, 15, 125, 539, 2543],  
 [412, 20, 93, 471, 2682],  
 [396, 21, 125, 492, 2828],  
 [423, 17, 210, 523, 2593],  
 [439, 19, 217, 463, 2702]  
])  
  
# Розділення даних на Y (залежна змінна) та X (незалежні змінні)  
Y = data[:, 0]  
X = data[:, 1:]  
  
# Крок 1: Знайти точкові незміщені статистичні оцінки b0\*, b1\*, b2\*, b3\*, b4\*  
X\_with\_intercept = np.hstack((np.ones((X.shape[0], 1)), X))  
beta = np.linalg.inv(X\_with\_intercept.T @ X\_with\_intercept) @ X\_with\_intercept.T @ Y  
  
beta0 = beta[0]  
beta1 = beta[1]  
beta2 = beta[2]  
beta3 = beta[3]  
beta4 = beta[4]  
  
print("beta0\*:", beta0)  
print("beta1\*:", beta1)  
print("beta2\*:", beta2)  
print("beta3\*:", beta3)  
print("beta4\*:", beta4)  
  
# Крок 2: Побудувати довірчий інтервал для функції регресії з надійністю у = 0.99  
n = len(Y)  
p = X.shape[1]  
  
gamma = 0.99  
t\_value = stats.t.ppf(1 - (1 - gamma) / 2, n - p - 1) # двобічний довірчий інтервал  
  
Y\_pred = X\_with\_intercept @ beta  
residuals = Y - Y\_pred  
residual\_var = np.sum(residuals \*\* 2) / (n - p- 1)  
  
# обчислення вектора залишкових помилок  
e = Y - Y\_pred  
  
# Обчислення стандартного відхилення  
SE = np.sqrt(np.sum(e \*\* 2) / (n - p - 1))  
  
# обчислення довірчого інтервалу  
CI\_lower = np.mean(Y\_pred) - t\_value \* SE  
CI\_upper = np.mean(Y\_pred) + t\_value \* SE  
  
print("Довірчий інтервал для функції регресії:")  
print("CI\_y=", (CI\_lower, CI\_upper))  
  
  
# Крок 3: Обчислити коефіцієнт множинної кореляції в  
Y\_mean = np.mean(Y)  
SS\_total = np.sum((Y - Y\_mean) \*\* 2)  
SS\_reg = np.sum((Y\_pred - Y\_mean) \*\* 2)  
R = np.sqrt(SS\_reg / SS\_total)  
  
print("Коефіцієнт множинної кореляції R:", R)