

Національний технічний університет України

“Київський політехнічний інститут”

АНАЛІЗ ДАНИХ І СТАТИСТИЧНА ОБРОБКА СИГНАЛІВ

# Лабораторна робота №1:

Крокова регресія

Виконав:

Студент гр. ФІ-31м

Ушаков Владислав

Варіант 10

Перевірив:

Архіпов О.Є.

**Назва роботи:** Крокова регресія.

**Мета роботи:** вибір структури регресійної моделі методом крокової регресії.

**Ключові поняття:** мультиколінеарність, кореляція, метод крокової регресії, регресори, ступінь вільності, оцінка якості моделі.

**Порядок виконання:**

1. Обстежити вихідні дані на присутність колінеарних незалежних факторів. При виявленні мультиколінеарності необхідно позбутися однієї змінної в парах, що мають високі коефіцієнти парної кореляції.
2. Розрахувати на базі незалежних факторів множину регресорів:
3. За допомогою статистичної функції КОРРЕЛ(массив\_1, массив\_2) знайти коефіцієнти кореляції залежної змінної з усіма регресорами.

Таблиця кореляцій

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  | max |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

1. Вибрати за допомогою статистичної функції МАКС регресор, що має максимальне за модулем значення коефіцієнта парної кореляції із залежною змінною . Включити цей регресор у модель

|  |  |
| --- | --- |
|  | (-1-) |

1. Методом найменших квадратів за допомогою статистичної функції ЛИНЕЙН знайти оцінки коефіцієнтів , .
2. Розрахувати модельні значення за формулою (-1-), підставивши отримані за допомогою функції ЛИНЕЙН оцінки коефіцієнтів.
3. Знайти нев'язку і визначити коефіцієнти кореляції нев’язки з усіма регресорами. Занести дані в таблицю кореляцій.
4. Вибрати регресор (див. таблицю), що має максимальне за модулем значення коефіцієнта парної кореляції із змінною . Включити цей регресор в ускладнену модель:

|  |  |
| --- | --- |
|  | (-2-) |

1. Для нової ускладненої моделі (-2-) за допомогою МНК розрахувати оцінки коефіцієнтів , , .
2. Розрахувати для двох моделей і -статистику:

|  |  |
| --- | --- |
|  | (-3-) |

де

– сума квадратів нев’язок для простої моделі -го кроку,

– сума квадратів нев’язок для простої моделі -го кроку,

– сума квадратів нев’язок для простої моделі -го кроку,

Значення , та можна знайти у додатковій статистиці функції ЛИНЕЙН.

1. Перевірити умову , виконання якої свідчить про доцільність ускладнення моделі, що суттєво збільшило точність апроксимації моделлю вихідних даних. Якщо ця нерівність не виконується, процедура крокової регресії переривається, і в якості моделі для подальшої роботи використовується більш проста модель, тобто модель, отримана не на поточному, а на попередньому кроці підбору елементів структури моделі.
2. Як тільки процес крокового ускладнення моделі виявиться неефективним, слід виконати за -критерієм Стьюдента

|  |  |
| --- | --- |
|  | (-4-) |

перевірку значущості оцінок параметрів моделі, знайдених на попередніх кроках. Для значущих параметрів моделі обов'язкова умова . Якщо будуть виявлені незначущі оцінки, для яких , то треба виключити їх з моделі, перевіривши точність спрощеної моделі за -статистикою.

Табличні значення та , де – довірча ймовірність, можна знайти у додатку.

1. П.п..3–12 повторюються, поки крокове ускладнення моделі буде ефективним.
2. Якщо виникають сумніви щодо ускладнення або спрощення моделі, провадиться комплексний аналіз якості моделі за -, -статистиками, за показниками коефіцієнта детермінації , за значеннями стандартної похибки оцінювання залежної змінної . Значну частину з цієї інформації можна отримати з додаткової статистики, яку повертає функція ЛИНЕЙН.
3. Побудувати графіки залежної змінної та кращої моделі.

**Результат роботи:**

Множина вихідних даних зображена на рис. 1.

Далі наведена таблиця розрахунків кореляцій до 4-го кроку (табл.2). На цьому кроці коефіцієнт буде дорівнювати , з чого можна зробити висновок про достовірність знайденої моделі.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Z** | 0.032 | -0.106 | 0.349 | 0.001 | 0.027 | 0.062 | -0.098 | 0.969 | -0.028 | 0.133 | 0.047 | 0.413 | -0.167 | -0.094 | 0.202 | -0.244 | -0.200 | 0.093 | 0.133 | 0.092 |
| **abs(Z)** | 0.032 | 0.106 | 0.349 | 0.001 | 0.027 | 0.062 | 0.098 | **0.969** | 0.028 | 0.133 | 0.047 | 0.413 | 0.167 | 0.094 | 0.202 | 0.244 | 0.200 | 0.093 | 0.133 | 0.092 |
| **Z-Y\_1** | 0.273 | -0.195 | 0.626 | -0.046 | -0.104 | 0.332 | -0.196 | 0.000 | -0.020 | 0.128 | 0.223 | 0.430 | -0.029 | 0.109 | 0.693 | -0.366 | -0.177 | 0.489 | 0.103 | -0.044 |
| **abs(Z-Y\_1)** | 0.273 | 0.195 | 0.626 | 0.046 | 0.104 | 0.332 | 0.196 | 0.000 | 0.020 | 0.128 | 0.223 | 0.430 | 0.029 | 0.109 | **0.693** | 0.366 | 0.177 | 0.489 | 0.103 | 0.044 |
| **Z-Y\_2** | 0.257 | -0.101 | 0.955 | -0.133 | -0.103 | 0.189 | -0.109 | 0.000 | -0.112 | 0.088 | 0.125 | -0.059 | 0.133 | 0.146 | 0.000 | -0.063 | -0.088 | 0.859 | 0.446 | -0.165 |
| **abs(Z-Y\_2)** | 0.257 | 0.101 | **0.955** | 0.133 | 0.103 | 0.189 | 0.109 | 0.000 | 0.112 | 0.088 | 0.125 | 0.059 | 0.133 | 0.146 | 0.000 | 0.063 | 0.088 | 0.859 | 0.446 | 0.165 |
| **Z-Y\_3** | -0.098 | -0.192 | 0.000 | 0.151 | 0.950 | 0.066 | 0.047 | 0.000 | 0.134 | 0.302 | 0.053 | -0.022 | -0.179 | 0.059 | 0.000 | -0.209 | -0.118 | 0.013 | 0.276 | 0.732 |
| **abs(Z-Y\_3)** | 0.098 | 0.192 | 0.000 | 0.151 | **0.950** | 0.066 | 0.047 | 0.000 | 0.134 | 0.302 | 0.053 | 0.022 | 0.179 | 0.059 | 0.000 | 0.209 | 0.118 | 0.013 | 0.276 | 0.732 |
| **Z-Y\_4** | 0.097 | -0.125 | 0.000 | -0.197 | 0.000 | 0.237 | 0.203 | 0.000 | -0.228 | -0.257 | 0.260 | -0.023 | 0.078 | 0.380 | 0.000 | -0.080 | -0.088 | -0.108 | 0.331 | -0.013 |
| **abs(Z-Y\_4)** | 0.097 | 0.125 | 0.000 | 0.197 | 0.000 | 0.237 | 0.203 | 0.000 | 0.228 | 0.257 | 0.260 | 0.023 | 0.078 | **0.380** | 0.000 | 0.080 | 0.088 | 0.108 | 0.331 | 0.013 |

Табл. 2. Таблиця кореляція

Рис. 1. Вихідна множина даних

Отримані моделі:

,

|  |  |
| --- | --- |
| 8.506032 | 4.688938 |
| 0.353093 | 1.309443 |
| 0.938544 | 6.583343 |
| 580.3294 | 38 |
| 25151.71 | 1646.935 |

*,*

|  |  |  |
| --- | --- | --- |
| 8.459211 | 5.141947 | 5.062733 |
| 0.258025 | 0.878879 | 0.958553 |
| 0.968077 | 4.808495 |
| 561.0146 | 37 |
| 25943.14 | 855.5003 |

,

|  |  |  |  |
| --- | --- | --- | --- |
| 8.140553 | 5.619882 | 3.092144 | 5.804724 |
| 0.054942 | 0.184003 | 0.108301 | 0.201534 |
| 0.99865 | 1.002535 |
| 8875.768 | 36 |
| 26762.46 | 36.18273 |

,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 8.100065 | 5.701198 | 3.300547 | 0.227544 | 5.467748 |
| 0.00039 | 0.0013 | 0.000801 | 0.000267 | 0.001474 |
| 1 | 0.007063 |
| 1.34E+08 | 35 |
| 26798.64 | 0.001746 |

**Висновки:**

В результаті виконання лабораторної роботи: