KPI

Міністерство освіти та науки

Національний технічний університет України

«Київський політехнічний інститут»

Фізико-технічний інститут

**Лабороторна робота №1**

на тему:

**«Крокова регресія»**

з предмету

«**Аналіз даних та статистична обробка сигналів**»

**Виконала:**

студентка 6 курсу ФТІ

групи ФІ-31м

Донченко Валентина

**Перевірив:**

Архипов О.Є.

**Київ 2014**

1. **Теоретичні відомості**

Регресія – зміна функції в залежності від зміни одного чи декількох аргументів.

Вираз  відповідає *моделі класичної лінійної регресії*. Через те, що змінна *y* лінійно залежить від невідомих параметрів , , регресійна модель називається *лінійною* *за параметрами* (або просто лінійною), а співвідношення– *лінійним регресійним рівнянням*.

Незалежні змінні в рівнянні лінійної регресії називаються *регресорами*, параметри  – *коефіцієнтами регресії*. Регресор може співпадати з факторною змінною або може бути довільною функцією від факторів, що не містить невідомих параметрів.

Для регресійних моделей існує декілька покрокових процедур (англ. *stepwise* – поступово, поетапно) відбору змінних: процедура послідовного приєднання, процедура приєднання-видалення та процедура послідовного видалення.

Ми детально розглянемо один із можливих способів організації розрахунків в покроковій процедурі приєднання, який називається методом крокової регресії.

Метод крокової регресії дає можливість визначити структуру функціональної залежності між вихідною та незалежними змінними.

*Алгоритм методу крокової регресії* містить таку послідовність дій.

1. Сформувати набір регресорів, які можуть впливати на значення вихідної змінної, зокрема, фактори (тобто вхідні змінні), квадрати факторів, їх добутки тощо.

2. Розрахувавши вибіркові коефіцієнти парної кореляції кожного регресора з вихідною змінною *Z*, вибрати та ввести регресор , який має максимальний за модулем коефіцієнт кореляції, до складу початкового варіанта моделі – так званої моделі першого *l=*1наближення .

3. Обчислити вектор  оцінок параметрів моделі першого наближення, розрахувати модельні значення  та нев’язки цієї моделі: .

Н*ев’язка* (залишок, похибка) - різниця між значенням залежної змінної *zi* та її модельним значенням : , .

4. Розрахувати суму квадратів нев’язок моделі: .

5. Розрахувати вибіркові коефіцієнти кореляції не введених у модель регресорів з нев’язками , , де *l* – номер наближення для попереднього кроку підбору структури моделі, визначити найбільш корельований (максимальний за модулем коефіцієнта кореляції) з нев’язками регресор і ввести його до складу моделі чергового (*l*+1)-го наближення.

6. Обчислити вектор  оцінок параметрів ускладненої моделі на поточному *l*+1 кроці процедури підбору структури моделі , розрахувати нев’язки  та відповідну суму квадратів нев’язок .

7. Розрахувати -статистику Фішера для двох послідовно отриманих моделей для перевірки ефективності ускладнення моделі за формулою:

, (15)

де  – сума квадратів нев'язок для регресійної моделі попереднього -го кроку,

 – сума квадратів нев'язок для моделі поточного ()-го кроку,

*k* – ступінь вільності для моделі ()-го кроку.

*Ступінь вільності* *k* дорівнює різниці між кількістю спостережень (або обсягом вибірки) *n* та кількістю параметрів *m* в моделі, тобто

. (16)

В загальному випадку *F*-статистику розраховують за такою формулою:

, (17)

де *k*1 – кількість додатково введених регресорів у поточну (ускладнену) модель порівняно з попередньою (при ускладненні моделі на один елемент *k*1=1 і тоді з формули (23) отримуємо формулу (2)),

*k*2 – ступінь вільності для ускладненої моделі.

8. Порівняти розраховане значення *F*-статистики з критичним значенням , що визначається за статистичними таблицями (див. додаток), де  – довірча ймовірність.

Якщо , чергове () ускладнення моделі слід вважати доцільним: воно обумовило суттєве зменшення рівня нев’язок і покращило точність апроксимації моделлю вихідних даних. В цій ситуації виникає потреба в перевірці можливості подальшого ускладнення моделі. Тепер останню модель вже вважатимемо попередньою і задля її ускладнення перейдемо до виконання п.5.

Пункти 5–8 повторюються, поки покрокове ускладнення моделі є ефективним.

Якщо , останнє ускладнення структури моделі не ефективне, процедура крокової регресії переривається і в якості моделі для подальшої роботи використовується більш проста модель, а саме модель, отримана не на поточному, а на попередньому -ому кроці підбору елементів структури моделі.

9.Як тільки процес покрокового ускладнення моделі буде перервано, слід виконати перевірку значущості оцінок параметрів найкращої моделі за -критерієм Стьюдента:

, , (18)

Де  – оцінка *j*-го параметра моделі;

 – оцінка середнього квадратичного відхилення для -го параметра моделі.

10. Розраховані значення  порівняти з критичними  із статистичних таблиць *t*-розподілу статистики Стьюдента (див. додаток), де *р* – довірча ймовірність, зазвичай *р*=0,95, *k* – ступінь вільності для обраної моделі

Для значущих параметрів моделі обов’язкова умова . Це означає, що задіяна у моделі сукупність регресорів не містить надмірності, зокрема, відсутні регресори, лінійно залежні від інших.

Якщо будуть виявлені незначущі оцінки, для яких , треба по одному виключити кожен з регресорів з моделі та виконати для отриманих таким чином спрощених варіантів моделі перевірку значущості оцінок їх коефіцієнтів за критерієм Стьюдента. Серед спрощених варіантів, що успішно пройшли перевірку, визначити найкращий (шляхом проведення зіставлення та додаткового аналізу цих варіантів за значеннями їх коефіцієнтів детермінації , залишковими сумами квадратів нев’язок, *t*-статистиками), який надалі слід використовувати у якості прикладної математичної моделі.

Деякі визначення:

Сукупність значень, отриманих у ході спостереження ряду величин під час проведення досліджень (експериментів), називають *вибіркою*. Кількість спостережень (експериментів) *n* визначає обсяг вибірки.

Будь-яку функцію вибіркових даних, що не залежить від невідомих параметрів, називають *статистикою*. Статистика є випадковою величиною.

Статистику, обчислену за вибірковими даними, яку беруть як невідоме значення параметра, називають *оцінкою* цього параметра*.*

1. **Завдання:**

1. Методом крокової регресії обрати найкращу модель.

2. Для обосновання вибору кращої моделі побудувати графіки динаміки статистик R2 , S для всіх отриманих моделей.

3. Побудувати графік Yp і Y за допомогою Мастеру діаграм.

**Табл. 2. Критические значения статистик**

|  |  |
| --- | --- |
| **F(p,k1,k2)** | **t(p,k)** |
| F(0,95;1;40)=4,08 | t(0,95;40)=2,02 |
| F(0,95;1;30)=4,17 | t(0,95;30)=2,04 |
| F(0,95;1;29)=4,18 | t(0,95;29)=2,05 |
| F(0,95;1;28)=4,20 | t(0,95;28)=2,05 |
| F(0,95;1;27)=4,21 | t(0,95;27)=2,05 |
| F(0,95;1;26)=4,22 | t(0,95;26)=2,06 |

**3.Виконання роботи:**

Побудова коефіцієнтів парної кореляції кожного регресора:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **x1** | **x2** | **x3** | **x4** |
| Z | 0,156875 | 0,305786 | 0,250021 | -0,44035 |
| abs(z) | 0,156875 | 0,305786 | 0,250021 | 0,44035 |
| z-y1 | 0,053456 | 0,379516 | 0,302567 | -0,53619 |
| abs(z-y1) | 0,053456 | 0,379516 | 0,302567 | 0,536186 |
| z-y2 | 0,09603 | 0,108995 | 0,165003 | -0,74604 |
| abs(z-y2) | 0,09603 | 0,108995 | 0,165003 | 0,74604 |
| z-y3 | 0,122556 | 0,013992 | 0,353997 | -0,04964 |
| abs(z-y3) | 0,122556 | 0,013992 | 0,353997 | 0,049642 |
| z-y4 | 0,120635 | -0,21569 | 6,79E-17 | -0,02417 |
| abs(z-y4) | 0,120635 | 0,215693 | 6,79E-17 | 0,024168 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **x1^2** | **x2^2** | **x3^2** | **x4^2** | **x1x2** | **x1x3** | **x1x4** | **x2x3** | **x2x4** | **x3x4** |
| z | 0,142567 | 0,498623 | 0,181283 | 0,819197 | 0,271593 | 0,225891 | -0,33653 | 0,399728 | -0,50921 | -0,47895 |
| abs(z) | 0,142567 | 0,498623 | 0,181283 | 0,819197 | 0,271593 | 0,225891 | 0,336526 | 0,399728 | 0,509211 | 0,478954 |
| z-y1 | 0,029548 | 0,836595 | 0,595338 | -4,4E-16 | 0,467027 | 0,330264 | -0,6689 | 0,739103 | -0,57648 | -0,46984 |
| abs(z-y1) | 0,029548 | 0,836595 | 0,595338 | 4,41E-16 | 0,467027 | 0,330264 | 0,668905 | 0,739103 | 0,576479 | 0,469844 |
| z-y2 | 0,158966 | 4,59E-16 | 0,109099 | -1,2E-15 | 0,20139 | 0,181312 | -0,94444 | 0,069097 | 0,046321 | -0,00829 |
| abs(z-y2) | 0,158966 | 4,59E-16 | 0,109099 | 1,18E-15 | 0,20139 | 0,181312 | 0,944443 | 0,069097 | 0,046321 | 0,008293 |
| z-y3 | 0,205258 | 1,29E-15 | 0,12313 | -2,6E-15 | 0,200192 | 0,318473 | 5,61E-17 | 0,150746 | 0,135928 | 0,054656 |
| abs(z-y3) | 0,205258 | 1,29E-15 | 0,12313 | 2,56E-15 | 0,200192 | 0,318473 | 5,61E-17 | 0,150746 | 0,135928 | 0,054656 |
| z-y4 | 0,203415 | 1,38E-15 | 0,184736 | -3,4E-15 | -0,01743 | 0,033734 | -5,3E-16 | 0,170038 | 0,129022 | 0,078317 |
| abs(z-y4) | 0,203415 | 1,38E-15 | 0,184736 | 3,37E-15 | 0,017429 | 0,033734 | 5,26E-16 | 0,170038 | 0,129022 | 0,078317 |

Оцінка параметрів моделі:

|  |  |  |
| --- | --- | --- |
|  | **a1** | **ao** |
|  | 3,318851 | 6,529811 |
| Sa | 0,376921 | 1,305298 |
| R^2, Sz | 0,671084 | 6,275653 |
| F, k | 77,53097 | 38 |
| Sreg, Sost | 3053,466 | 1496,585 |
| **T** | 8,805167 | 5,002546 |

|  |  |  |  |
| --- | --- | --- | --- |
| A | 3,116635 | **3,274157** | **3,513919** |
| Sa | 0,335217 | 0,209183 | 0,79355 |
| R^2, Sz | 0,901411 | 3,481939 | #Н/Д |
| F, k | 169,148 | 37 | #Н/Д |
| Sreg, Sost | 4101,467 | 448,5843 | #Н/Д |
| F | 86,44093 |  |  |
| **T** | 9,297362 | 15,65209 | 4,428103 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | -2,05546 | 2,754077 | 3,351026 | 3,809173 |
| Sa | 0,097194 | 0,094328 | 0,057997 | 0,220024 |
| R^2, Sz | 0,992655 | 0,963477 | #Н/Д | #Н/Д |
| F, k | 1621,851 | 36 | #Н/Д | #Н/Д |
| Sreg, Sost | 4516,633 | 33,41836 | #Н/Д | #Н/Д |
| F | **447,2383** |  |  |  |
| **T** | 21,14801 | 29,19684 | 57,77978 | 17,31252 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| A | 0,521239 | -2,03954 | 2,702686 | 3,338758 | 3,887344 |
| Sa | 0,222137 | 0,091879 | 0,091584 | 0,054925 | 0,210084 |
| R^2, Sz | 0,993654 | 0,908308 | #Н/Д | #Н/Д | #Н/Д |
| F, k | 1370,014 | 35 | #Н/Д | #Н/Д | #Н/Д |
| Sreg, Sost | 4521,176 | 28,87583 | #Н/Д | #Н/Д | #Н/Д |
| F | **5,505946** |  |  |  |  |
| **T** | 2,346475 | 22,19804 | 29,51045 | 60,7876 | 18,50378 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | -0,37255 | 0,856463 | -2,05529 | 2,751749 | 3,345821 | 3,823203 |
| Sa | 0,19769 | 0,278631 | 0,089096 | 0,092171 | 0,053159 | 0,205658 |
| R^2, Sz | 0,994254 | 0,876909 | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| F, k | 1176,615 | 34 | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| Sreg, Sost | 4523,906 | 26,14496 | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| F | #**3,551334** | #Н/Д | #Н/Д | #Н/Д | #Н/Д | #Н/Д |
| **T** | 1,884498 | -3,07382 | 23,06827 | 29,85467 | 62,94037 | 18,59014 |

Графіки динаміки статистик R2 , S для всіх отриманих моделей:

Графік Yp і Y:

1. **Висновки**

За результатами виконаної роботи ми навчились користуватись методом крокової регресії та працювати з структурами регресійної моделі методом крокової регресії, а саме: перевіряти вихідні дані на наявність колінеарності між факторами, підбирати модель методом крокової регресії, проводити комплексний аналіз якості моделі і при необхідності уточнювати її структуру, оцінювати якість параметричної ідентифікації.