**Однофакторна регресійна модель**

1 Застосування методу найменших квадратів для визначення параметрів однофакторної лінійної моделі

2 Перевірка моделі на адекватність

3 Нелінійна однофакторна модель

**|фактор|**

Нехай|нехай| є вибірка з|із| генеральної сукупності  і відповідна до неї вибірка з|із| генеральної сукупності . Тоді можна записати наступне|слідуючий| співвідношення:

, (1.25)

де *ƒ* – деяка детермінована функція;

*ε* – випадкова величина, що має математичне очікування|чекання| рівне нулю. Функцію *ƒ* називають функцією регресії або просто регресією *у|в,біля|(х),* вона вказує|вказувати| на наявність статистичного зв'язку між ознаками *Y*|в,біля| і *X*. Характер|вдача| цього зв'язку визначає вигляд|вид| функції регресії (лінійна, логарифмічна, поліноміальна тощо). У статистиці функцію*ƒ*, деяку детерміновану складову моделі (1.25) також називають трендом.

### 1 Застосування методу найменших квадратів для визначення параметрів однофакторної лінійної моделі

Перше завдання|задача|, яке виникає під час побудови|шикування| моделі перед дослідником – це визначити вид залежності між факторами, тобто характер|вдача| *ƒ*. Якщо *r≠*0, , то між *X* і *Y* |в,біліснує лінійний статистичний зв'язок, тобто|цебто| в моделі (1.25) функція лінійна:

.

Отже, модель (1.25) має вигляд|вид|:

. (1.26)

Модель (1.26) називають лінійною однофакторною моделлю, причому роль фактора відіграє *X*. Далі необхідно визначити коефіцієнти моделі *α1, α0.* Для цього до масиву дослідних даних застосовують метод найменших квадратів.

Якби існувала чисто функціональна залежність *Y*|в,біля| від *X*, то в моделі (1.26)   
*ε* = 0 і модель мала б вигляд|вид|:

*.*

На практиці маємо

. (1.27)

Оскільки|тому що| *ε* в моделі (1.27) має математичне очікування|чекання| рівне нулю, тоді *ε* повинно мати мале значення:

.

Розглянемо|розгледимо|  і оберемо *α*1, *α0* такими, щоб сума була найменшою. Введемо|запровадимо| функцію двох змінних:

**,(1.28)

і шукатимемо α1, α0 такі, щоб функція *F* в формулі (1.28) при цих значеннях досягла найменшого значення, тобто|цебто| шукатимемо 

Необхідною умовою екстремуму функції є|з'являтися,являтися| рівність нулю її частинних похідних. Знайдемо частинні похідні  та  функції *F* і прирівняємо їх до нуля.

 ⇔ ⇔

.

Доданки, що відповідають невідомим *α1, α0,* залишимо в лівій частині|частка| рівняння, а інші – перенесемо до правої|вправо|.

. (1.29)

Таким чином, маємо систему двох лінійних алгебраїчних рівнянь (1.29) для знаходження параметрів моделі α1, α0*.* Цю систему рівнянь називають нормальною системою рівнянь. Її розв'язанням є|з'являтися,являтися| шукані параметри регресійної моделі (1.26).

Розв'яжемо (1.29):

, (1.30)

де .

Підставимо α0 в друге рівняння системи (1.29):

,

,

. (1.31)

Використовуємо вибіркові оцінки. Оцінки дисперсій факторів:

, (1.32)

. (1.33)

Оцінка коваріації між *Х* і *Y* |в,біля|:

. (1.34)

Оцінне значення коефіцієнта кореляції:

. (1.35)

Оцінки середньоквадратичних відхилень відповідно|відповідно до| *Х* і *Y*|в,біля|:

 . (1.36)

Тоді співвідношення (1.31) з|із| урахуванням (1.32)–(1.36) можна привести до вигляду|вид|:

,

, (1.37)

.

### 2 Перевірка моделі на адекватність

Під перевіркою моделі на адекватність розуміють перевірку того, що побудована|споруджена| модель адекватно (реально) оцінює досліджувану залежність між ознаками:

. (1.38)

За формулою (1.38) оцінимо значення розрахункове значення , яке знаходять|находити| підставляючи в модель відповідне значення *хi*. Якщо модель адекватна, різниця між реальним значенням ознаки і його оцінкою невелика: .

Адекватність моделі оцінюють за *F-*статистикою або *F-*співвідношенням:

, (1.39)

де *k* – число незалежних змінних в моделі. Для однофакторної моделі *k*=1, для двофакторної – *k*=2, для трифакторної – *k*=3 тощо.

Перевірку здійснюють за допомогою таблиць Фішера, оскільки *F-* співвідношення|ставлення| (1.39) розподілене за законом розподілу Фішера з|із| числом ступенів свободи 1(для однофакторної моделі), *n-*2. Для заданого рівня значущості *q* (наприклад, *q*=0,05) на перетині числа ступенів свободи 1 (у першому стовпці) і числа ступенів свободи (*n*-2) (у першому рядку) знаходять|находити| *F1,n-2,q*. Якщо значення F, знайдене за формулою (1.39) більше табличного значення *F1,n-2,q*, то з|із| вірогідністю|ймовірність| помилки *q* модель адекватно відображає|відбивати| процес, інакше – модель неадекватна з|із| довірчою вірогідністю|ймовірність| *α*.

**Приклад 1.13** Задана вибірка з генеральної сукупності *Х =* (1, -1, 1, 1, -1, 0) і відповідна до неї вибірка з генеральної сукупності *Y* = (0, -2, 1, 0, -2, -1). Побудувати лінійну модель і перевірити її на адекватність із довірчою вірогідністю 95%.

**Розв’язання.** Заповнимо кореляційну таблицю (табл. 1.4).

Таблиця 1.4 – Кореляційна таблиця

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| n |  |  |  | 2 | 2 |  | - | (-)2 | - | (-)2 |
| 1 | 1 | 0 | 0 | 1 | 0 | 0,310345 | 0,977011 | 0,954551 | 0,31034 | 0,0963 |
| 2 | -1 | -2 | 2 | 1 | 4 | -2,03448 | -1,36782 | 1,870921 | -0,03448 | 0,0012 |
| 3 | 1 | 1 | 1 | 1 | 1 | 0,310345 | 0,977011 | 0,954551 | -0,68966 | 0,4756 |
| 4 | 1 | 0 | 0 | 1 | 0 | 0,310345 | 0,977011 | 0,954551 | 0,31034 | 0,0963 |
| 5 | -1 | -2 | 2 | 1 | 4 | -2,03448 | -1,36782 | 1,870921 | -0,03448 | 0,0012 |
| 6 | 0 | -1 | 0 | 0 | 1 | -0,86207 | -0,1954 | 0,038182 | 0,13793 | 0,0190 |
|  | 1 | -4 | 5 | 5 | 10 | -4 | 0 | 6,643678 | 0 | 0,6896 |

За формулам (1.30), (1.31) розрахуємо  і :



У результаті модель прийме вигляд: =1,17*хi* – 0,86.

Розрахуємо модельні значення  = 1,17-0,86, *i* = 1,…,6;

=1,17\*1-0,86=0,31;

=1,17\*(-1)-0,86=-2,03;

=0,31;

=0,31;

=-2,03;

=-0,86.

Занесемо значення в табл. 1.4 Знайдемо математичне очікування  і заповнимо чотири правих стовпця таблиці. Розрахунки зручно виконувати в середовищі Excel.

Розрахуємо значення *F*-статистики, .

За таблицею (додаток Г) або за допомогою вбудованої функції FРАСПРОБР знайдемо =7,7; порівняємо розрахункове і табличне значення. Оскільки 39,12>7,7, можна вважати, що модель =1,17*хi* – 0,86 адекватна з рівнем значущості 5%.

**3 Нелінійна однофакторна економетрична модель**

**|фактор|**

Нехай|нехай| *Х* – незалежна, а *Y* |в,біля| – залежна змінна, і між цими змінними існує деякий зв'язок:

, (1.40)

де  – деяка нелінійна функція, причому вигляд її невідомий;

ε – випадкова величина з|із| математичним очікуванням|чекання| рівним нулю.

Нелінійна функція  називається кривою зростання|зріст|. Як криві зростання|зріст| можуть виступати |вирушати|поліноми різного ступеня:|міра|

.

Представлена|розгледіти| вище лінійна модель – це поліном першого ступеня. Можливість|спроможність| вибору полінома пояснюється тим, що гладку функцію можна з|із| деякою мірою точності апроксимувати поліномами. На практиці будують поліноми 2-го ступеня і перевіряють моделі на адекватність. Коефіцієнти знаходять|находити| за допомогою МНК.

Наприклад, у разі|в разі| полінома 2-го ступеня  знаходять|находити| різниці , зводять|підносити| їх до квадрата і мінімізують:

()2 min.

Для цього розглядають|розглядувати| частинні похідні і прирівнюють їх до нуля: , внаслідок чого отримують|одержувати| систему рівнянь з|із| трьома невідомими (). Така система рівнянь називається нормальною, а її розв’язання|розв'язання,вирішення,розв'язування| дозволяє отримати|одержати| оцінки параметрів .

Аналогічним чином діють і у випадку|в разі| поліномів *n*-ступеня|міра|, записуючи|занотовувавши| систему *n+1* рівнянь з|із| *n+1* невідомими.

**Завдання для самоперевірки**

1 Побудуйте лінійну модель для наступних пар вибіркових значень (*х, у*) та перевірте модель на адекватність з довірчою вірогідністю 90%.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *х* | 2 | 2,1 | 2,3 | 2,6 | 2,8 | 2,8 | 2,2 | 2,3 | 2,6 |
| *y* | 48,1 | 45,1 | 39,7 | 37,2 | 31,7 | 30,1 | 39,9 | 38,8 | 38,4 |

2 Побудуйте лінійну модель для наступних пар вибіркових значень (*х, у*) та перевірте модель на адекватність з рівнем значущості 5%.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 2,4 | 2,2 | 2,6 | 2,6 | 2,3 | 2,2 | 2,8 | 2,8 | 2,6 |
| *y* | 40,2 | 39,4 | 35,1 | 38,4 | 38,8 | 39,9 | 30 | 31,7 | 37,2 |

3 Побудуйте лінійну модель для наступних пар вибіркових значень (*х, у*) та перевірте модель на адекватність з довірчою вірогідністю 0,95.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *x* | 1,9 | 2 | 2,2 | 1,9 | 2,4 | 1,9 | 2,1 | 2,4 | 2,2 |
| *y* | 39,1 | 37,8 | 35,8 | 40,2 | 29,7 | 38,2 | 34,3 | 30,1 | 34,5 |

**Контрольні питання**

1 У чому сутність методу найменших квадратів?

2 Який характер має зв’язок між коефіцієнтом кореляції і параметрами моделі?

3 За допомогою якого критерію перевіряють адекватність однофакторної лінійної моделі?

4 Що таке тренд?

5 Який характер в загальному випадку може мати регресійна модель?

6 Яким чином використовують табульовані значення розподілу Фішера під час перевірки моделі на адекватність?