**Тести Чоу**

**Зміст**

Вступ

Застосування тесту

Реалізація тесту Чоу

Тести на стійкість

Тест Чоу на невдачу прогнозу

Приклад

F-тест на стабільність коефіцієнтів

Приклад

Висновок

Література

# Вступ

Нерідко на практиці ми стикаємося всього лише з двома групами даних. Наприклад, важливо визначити значущість структурних зрушень між двома періодами часу, або ж значущість відмінностей у виробничих функціях для двох галузей, або ж значущість відмінностей між споживчими функціями у двох країнах [НАК, c. 196].

Доти, доки кількість спостережень у кожній групі перевищує кількість оцінюваних параметрів, запропонований підхід можна застосувати безпосередньо. Якщо ж в одній із груп кількість спостережень менше кількості параметрів, то цей випадок потребує особливого дослідження, яке виконав Г. Чоу (незалежно від нього – Ф. Фішер).

Нехай маємо сукупність спостережень, яка містить дві групи даних, що різняться між собою певними якісними ознаками. Постають такі запитання:

1. Чи можемо ми об'єднати ці групи для побудови економетричної моделі за всією сукупністю спостережень?

2. Як відрізняються вільні члени моделей, побудованих окремо за двома групами даних?

3. Чи однакові оцінки параметрів моделі, що характеризують вплив пояснювальної змінної на залежну?

Якщо йдеться лише про відмінності вільних членів, то групи спостережень можна об'єднати в одну сукупність і скоригувати вільний член введенням до моделі фіктивних змінних.

Часто трапляється так, що окремі чинники, які ви хотіли б ввести в регресійну модель, є якісними за своєю природою і, отже, не вимірюються в числовій шкалі. Наведемо декілька прикладів [ДОУ, с. 262].

1. Досліджується залежність між тривалістю здобутої освіти і доходом, і у вибірці представлені особи як чоловічої, так і жіночої статі. Потрібно з'ясувати, чи обумовлює стать відмінність в результатах.

2. Досліджується залежність між доходом і споживанням, і вибірка включає як україномовні сім'ї, так і сім'ї, що розмовляють російською. Потрібно з'ясувати, чи має істотне значення ця етнічна відмінність.

3. Досліджуються чинники, що визначають інфляцію, і в деякі роки періоду спостережень уряд проводив політику регулювання доходів. Потрібно перевірити, чи надало це який-небудь вплив на досліджувану залежність.

У кожному з цих прикладів одним з можливих рішень було б оцінювання окремих регресій для двох вказаних категорій з подальшим з'ясуванням, чи розрізняються одержані коефіцієнти. Інший можливий підхід до рішення полягає в оцінюванні єдиної регресії з використанням всієї сукупності спостережень і вимірюванням ступеня впливу якісного чинника за допомогою введення так званої фіктивної змінної. Другий підхід володіє двома важливими перевагами: по-перше, є простий спосіб перевірки, чи є дія якісного чинника значущим; по-друге, за умови виконання певних припущень регресійні оцінки виявляються ефективнішими.

# Застосування тесту

Тест Чоу дозволяє оцінити значущість поліпшення регресійної моделі після розділення початкової вибірки на частини. Це модифікація однопараметричної експоненціальної моделі з корекцією коефіцієнта лінійного тренда. У методі Чоу відбувається адаптація параметра до змін в діанаміке ряду.

Прикладом використання даного критерію може служити завдання, яке вирішував Чоу. Він тестував свій метод на рядах місячних даних про операції на різні види продукції: рукавички, змащувальні матеріали, сальники, підшипники і т.д. Дані були різноманітними зразками поведінки економічних тимчасових рядів, включаючи циклічний рух. У 59 випадків з 60 пропонований метод показав переваги перед стандартною процедурою і в одному випадку результати були майже однакові.

Проілюструємо метод використання фіктивних змінних на прикладі регресійного аналізу основних чинників, що впливають на вагу новонароджених немовлят. Якщо ви думаєте, що ця тема не представляє достатнього інтересу для економіста, то помиляєтеся. Типовий економіст, що працює в прикладній сфері, не витрачає увесь свій час на створення макроекономічних моделей. Значно вірогідніше, що він бере участь в роботі, що має більш безпосереднє відношення до практики, наприклад, займається аналізом використання ресурсів в якій-небудь конкретній сфері. Оскільки в більшості країн на медичне обслуговування прямує достатньо велика частина особистих і суспільних ресурсів, цілком виправданий інтерес до нього з боку економістів; на цю тему є обширна економічна література, причому об'єм її постійно зростає. Не зважаючи на те, що найбільшою увагою засобів масової інформації звичайно користується невідкладна медична допомога, найважливішими з погляду витрат є акушерськая допомога, догляд за літніми і турбота про осіб, що мають психічні захворювання.

Оскільки вартісний вираз результатів медичної допомоги звичайно є вельми спірним предметом, неможливо провести задовільний порівняльний аналіз витрат і результатів по більшості видів витрат на медицину. Натомість широко використовується наступний підхід: береться який-небудь показник успіху (або невдачі), визначаються основні чинники, його що обумовлюють, і потім робиться спроба знайти найбільш ефективний шлях досягнення заданої мети, вираженої цим показником.

У принципі завдання виявлення основних чинників, що формують значення цільового показника, повинна розв'язуватися фахівцями у області медичної статистики, а завдання якнайкращого розподілу ресурсів – економістами; але в охороні здоров'я, як і в інших галузях прикладної економіки, високопрофесійним економістам часто доводиться виходити за безпосередні рамки своєї дисципліни і проводити такий статистичний аналіз, перш ніж взятися до своєї основної роботи. У області акушерської допомоги двома головними показниками є дитяча смертність і вага новонароджених. Оскільки коефіцієнт смертності новонароджених для більшості країн дуже низький, для дослідження визначальних його чинників потрібні вибірки великого об'єму, і тому вага новонароджених у багатьох випадках є практичнішим альтернативним показником.

# Реалізація тесту Чоу

Іноді вибірка спостережень складається з двох або більше підвибірок, і важко встановити, чи слід оцінювати одну об’єднану регресію або окремі регресії для кожної підвибірки [ДОУ, с. 282]. На практиці проблема вибору стоїть звичайно не так жорстко, оскільки можуть бути деякі можливості об'єднання підвибірок при використанні відповідних фіктивних змінних і фіктивних змінних для коефіцієнта нахилу, щоб зробити менш строгим припущення про те, що всі коефіцієнти повинні бути однаковими для кожної підвибірки. До цього питання ми ще повернемося.

Припустимо, що є вибірка, що складається з двох підвибірок, і що виникає питання, чи слід об'єднати їх для оцінювання загальної регресії P або оцінити окремі регресії А і В. Позначимо суми квадратів залишків для регресій підвибірок UA і UB. Хай UPA і UPB – суми квадратів залишків в об'єднаній регресії для спостережень, що відносяться до двох даних підвибірок. Оскільки окремі регресії для підвибірок повинні відповідати спостереженням щонайменше так само добре, якщо не краще, ніж об'єднана регресія, то UA < UPA і UB < UPB. Отже, (UA + UB) < UP, де загальна сума квадратів залишків в об'єднаній регресії UP рівна сумі UPA і UPB .

Це пояснюється на малюнку.

Припустимо, що є дані тимчасового ряду по двох змінним і що в період вибірки відбулася структурна зміна, що розділяє спостереження на підвибірки А і В. На мал. Б регресії для підвибірок забезпечують цілком адекватну відповідність даним, обумовлюючи низькі значення UA і UB. Якби потрібно було оцінити об'єднану регресію, як на мал. А, то залишки в обох підвибірках в цілому були б значно більше.

Рівність між UP і (UA + UB) матиме місце тільки при збігу коефіцієнтів регресії для об'єднаної регресії і регресій підвибірок. У загальному випадку при розділенні вибірки спостерігатиметься поліпшення якості рівняння, що можна представити як (UP – UA – UB). Це має свою ціну: використовуються (k + 1) додаткових ступенів свободи, оскільки замість (k + 1) параметрів для однієї об'єднаної регресії ми тепер повинні оцінити в сумі (2k + 2) параметрів (k – число пояснюючих змінних, одиниця відповідає постійному члену). Проте, після розділення вибірки залишається непояснена сума квадратів залишків (UA + UB) і, крім того (n – 2k – 2) ступенів свободи.

Тепер ми можемо визначити, чи є значущим поліпшення якості рівняння після розділення вибірки. Для цього використовується F-статистика:

|  |  |  |
| --- | --- | --- |
| Покращення якості рівняння / Використані ступені свободи | = |  |
| Непояснена дисперсія / Число решти ступенів свободи |

яка розподілена з (k + 1) і (n – 2k – 2) ступенями свободи.

Тепер, наприклад, давайте повернемося до випадку парної регресійної залежності ваги новонароджених від інтенсивності куріння їх матерів, і хай ми ще не вирішили, чи слід об'єднувати підвибірки, включаючих 584 матері, яка раніше не народжувала, і 380 матерів, які раніше народжували. Оцінювання об'єднаної регресії і регресій для підвибірок дає результати, показані в таблиці.

|  |  |  |  |
| --- | --- | --- | --- |
| Вибірка | Оцінене рівняння | R2 | Сума квадратів залишків |
| Об’єднана вибірка | y = 3418 – 7,2x (1) | 0,012 | 158,6 ∙ 106 |
| Первістки | y = 3363 – 4,0x (2) | 0,004 | 91,2 ∙ 106 |
| Непервістки | y = 3506 – 12,1x (3) | 0,039 | 63,5 ∙ 106 |

Відповідна F-статистика, отже, рівна:

Критичне значення F з 2 і 960 мірами свободи складає 6,91 (при рівні значущості в 0,1%), тому ми робимо висновок, що не слід оцінювати об'єднану регресію.

Регресії для підвибірок ідентичні регресіям, представленим співвідношеннями (2) і (3), і це не простий збіг. У основній регресії (1) складова, не пов'язана з фіктивною змінною, включає постійний член і показник залежності від інтенсивності куріння. До цього додаються фіктивна змінна, що дозволяє розрізняти значення постійного члена для первістків і дітей, що народилися не першими, і фіктивна змінна для коефіцієнта нахилу, що також дозволяє розрізняти коефіцієнти при показнику інтенсивності куріння для двох даних підвибірок. Отже, в (1) ми не задаємо наперед який-небудь коефіцієнт однаковим для обох підвибірок і, таким чином, одержуємо такі ж оцінки коефіцієнтів, як і в окремих регресіях для підвибірок.

Розглядаючи лише співвідношення (1), ми можемо перевірити, чи виправдана ця гнучкість, з'ясувавши, чи вносять вказані фіктивні змінні, як група, значущий внесок в пояснюючу здатність рівняння. Сума квадратів залишків, якщо фіктивні змінні не включені в рівняння, складає 158,6 ∙ 106, а коли вони включені в рівняння, ця сума рівна 154,7 ∙ 106. Отже, F-статистика для перевірки пояснюючої здатності фіктивних змінних як групи має вигляд:

тобто вона в точності така ж, як в тесті Чоу.

Можна показати, що це загальний результат. Вибір між використанням розглянутої процедури тесту Чоу або оцінюванням складної регресії з фіктивними змінними на основі співвідношення (1) залежатиме від цілей, які ставить перед собою дослідник. Тест Чоу виконується швидше, і він достатній, якщо потрібно тільки встановити, що залежності в підвибірках в деякій мірі розрізняються. Оцінювання регресії з фіктивними змінними більш інформативно в тому відношенні, що воно дозволяє виконувати F-тесты з розглядом внеску кожної фіктивної змінної, а також всієї групи в цілому і може привести до компромісу, в якому дослідник припускає, що деякі коефіцієнти однакові в обох підвибірках, і використовує фіктивні змінні для диференціації значень решти коефіцієнтів.

# Тести на стійкість

Тести на стійкість для регресійної моделі призначені для оцінки того, наскільки поведінка моделі в послявибірковому періоді порівнянна з її поведінкою в період вибірки, на якій вона була одержана. У основі організації тестів на стійкість можуть лежати два принципи. Один підхід – зосередитися на передбачаючій здатності моделі; інший підхід – оцінити, чи відбувається зрушення параметрів в період прогнозу.

## Тест Чоу на невдачу прогнозу

Як ми бачили в попередньому розділі, помилку прогнозу можна розрахувати, додавши набір фіктивних змінних для спостережень періоду прогнозу. Тепер цілком природно визначити, чи істотно помилка прогнозу відрізняється від нуля, і ми можемо зробити це за допомогою F-тесту на сумісну пояснюючу здатність фіктивних змінних. Сумістивши період вибірки і період прогнозу, ми оцінимо рівняння регресії спочатку без набору фіктивних змінних, а потім – разом з цим набором. Позначимо одержані суми квадратів відхилень як RSST+m і RSSDT+m, де нижній індекс показує число спостережень в регресії, а верхній індекс «D» означає включення в рівняння фіктивних змінних. За допомогою F-тесту, описаного в розділі, ми можемо визначити, чи було істотним поліпшення якості рівняння після додавання набору фіктивних змінних. Його можна представити у вигляді (RSST+m – RSSDT+m); число фіктивних змінних рівне m; сума квадратов відхилень після включення фіктивних змінних складає RSSDT+m; число мір свободи, що залишається, рівне числу спостережень в суміщеній вибірці (T + m) за вирахуванням числа оцінених параметрів (k + m + 1). У результаті значення F-статистики складе:

Насправді для реалізації тесту навіть не потрібно оцінювати рівняння регресії з фіктивними змінними, оскільки значення RSSDT+m рівно значенню RSST – сумі квадратів відхилень для рівняння регресії, оціненого на періоді вибірки. Якість цієї регресії в точності таке ж, як і у регресії для перших T спостережень в рівнянні з фіктивними змінними, і відхилення тут ті ж самі. Для останніх m спостережень в рівнянні з фіктивними змінними немає відхилень, оскільки включення спеціальної фіктивної змінної для кожного спостереження гарантує точність рівняння для цих спостережень. У результаті значення RSSDT+m у точності таке ж, як і значення RSST, і F-статистика може бути переписана як

Цей тест відомий як тест Чоу і був названий так на ім'я свого творця Г. Чоу (Chow, 1960), інтерпретація тесту, що проте приводиться тут, була запропонована дещо пізніше X. Песараном, Р. Смітом і С. Ео.

## Приклад

Функція попиту на продукти харчування спочатку була оцінена на даних за період 1959-1979 рр., і RSST = 0,0052, а потім – на даних за період 1959-1983 рр., RSST+m = 0,0070. Як наслідок значення F-статистики рівне:

Критичне значення F-статистики з 4 і 18 мірами свободи при 5-процентному рівні значущості рівне 2,93, тому ми не відкидаємо нульову гіпотезу про стабільність коефіцієнтів рівняння регресії.

## F-тест на стабільність коефіцієнтів

Якщо є прийнятні спостереження за період прогнозу, то можна провести F-тест на наявність структурного перелому, описаний в розділі, і оцінити, чи значущо розрізняються коефіцієнти періоду вибірки і періоду прогнозу. Для реалізації цього тесту спочатку необхідно оцінити роздільно рівняння регресії для періоду вибірки і періоду прогнозу, а потім – спільно для цих двох періодів. Після цього потрібно перевірити, чи значущо поліпшується якість рівняння при розділенні загального періоду оцінки регресії на період вибірки і період прогнозу. Підтвердження цієї гіпотези може служити свідчення того, що коефіцієнти регресії нестабільні.

##

## Приклад

При оцінюванні функції попиту на продукти харчування з використанням спостережень за 1959-1979 рр. як період вибірки, а за 1980-1983 рр. – як період прогнозу, суми квадратів відхилень для періоду вибірки, періоду прогнозу і суміщеного періоду дорівнювали 0,0052; 0,0002 і 0,0070 відповідно. Оцінка окремих рівнянь регресії для двох періодів призводить до втрати трьох мір свободи, і число мір свободи, що залишається після оцінювання шести параметрів (двох постійних членів, двох коефіцієнтів при logx, двох коефіцієнтів при logp), рівне 19. У результаті ми одержуємо наступну F-статистику, розподілену з 3 і 19 мірами свободи:

Критичне значення t-статистики з таким числом мір свободи при 5-процентному рівні значущості рівне 3,13, що дозволяє нам зробити висновок про відсутність явної нестабільності коефіцієнтів.

# Висновок

Побудова регресійних моделей на сьогодні, поза сумнівом, є найбільш широко вживаним методом багатовимірного статистичного аналізу соціологічних даних. За останні декілька років більше половини статей, що аналізують емпіричні дані, засновані на використанні регресійних моделей.

Достатньо поширені регресійні методи і серед російських соціологів, фахівців, що використовують дослідні методики. Разом з тим багато особливостей і обмеження регресійних моделей звичайно залишаються поза сферою уваги дослідників, що, часом, призводить до неточних, або просто помилкових результатів.

Традиційна модель множинного лінійного регресійного аналізу має на увазі пошук показників (що позначаються X), що визначають значення окремої кількісної змінної, що позначається Y. Структура зв'язку в даній моделі передбачається лінійною. Іншими словами, шукається наступна форма залежності:

Y = B0 + B1X1 + B2X2 + ... +BnXn + U,

де U – так званий залишковий член, що фіксує ту частину інформації Y, яка не пояснюється іксами.

Регресійний аналіз показує, по-перше, якість моделі, тобто ступінь того, наскільки дана сукупність іксів пояснює Y. Показник якості називається коефіцієнтом детерміації R2 і показує, який відсоток інформації Y можна пояснити поведінкою іксів. По-друге, регресійний аналіз обчислює значення коефіцієнтів В, тобто визначає, з якою силою кожний з Х впливає на Y.

Методологічним недоліком такого підходу є те, що дана залежність шукається єдиною для всієї сукупності опитаних респондентів. Іншими словами, ми припускаємо, що для всіх людей характер залежності Y від іксів єдиний. У тому випадку, коли вибіркова сукупність достатньо однорідна, такого роду допущення має під собою певні підстави. Проте, якщо аналізуються, скажімо, детермінанти електоральних переваг на основі даних всеросійської вибірки, допущення про однорідність цих детермінантів для чукотського оленяря і для московського професора виглядає не дуже переконливим.

Єдина форма рівняння в цій ситуації сильно огрублює реальну залежність, якість моделі неминуче виявляється вельми низькою, а сенс регресійних коефіцієнтів, що фіксують ступінь впливу іксів на Y, можна прирівняти до горезвісного показника "середньої температури по лікарні".

Цілком очевидно, що набагато розумніше будувати окремі моделі для груп респондентів, що істотно розрізняються між собою. Проте доведення такого підходу до логічного завершення чревате небезпекою повного релятивізму. Дійсно, завжди можна знайти більш-менш переконливі аргументи на користь того, що з аналізованої проблеми механізми формування оцінок різні у жінок і чоловіків, у городян і сільських жителів, у інженерів і робочих і т.д. і т.п. Отже, для кожної групи необхідно будувати свою модель, що не дуже конструктивно, оскільки кількість таких моделей обмежується лише фантазією соціолога по розбиттю всієї сукупності на окремі групи.

Виявляється, проте, що є певні формальні критерії, що дозволяють визначати межі груп, для яких діють однакові, або різні механізми.

Отже, ми розглянули статистичний тест, що дозволяє оцінити значущість поліпшення регресійної моделі після розділення початкової вибірки на частини. Одним з обмежень лінійної регресії є те, що для різних інтервалів значень незалежної змінної характер її зв'язку з вихідною змінною може мінятися. Наприклад, із збільшенням віком клієнта його кредитний рейтинг може збільшуватися. Але дана закономірність не справедлива для всіх віків. Після певного віку (50–55 років), люди частіше хворіють, їм складніше знайти роботу і т.д., тому після, скажімо, 50 років спостерігається зворотна залежність.

Очевидно, що будь-яка модель, яка апроксимує таку закономірність єдиною лінійною залежністю, навряд чи буде точною. Виходом з ситуації є розділення діапазону значень вхідної змінної на два, в межах кожного з яких залежність між нею і вихідній змінній монотонна і побудова рівняння регресії для кожного одержаного піддіапазону. Виникає питання: як розбити початкову множину так, щоб одержане розбиття забезпечило кращу апроксимацію? Для цього звичайно будують безліч розбиття, для кожного визначають значущість поліпшення моделі і вибирають те, яке забезпечило велику значущість. Для оцінки такої значущості і використовується тест Чоу.

# Література

1. Елисеева И. И. Эконометрика. М.: ФиС. – 2004, 344 ст. [ЕЛИ]
2. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2005. – 311 с. [КРЕ]
3. Лещинський О. Л. Економетрія. – К.:МАУП 2003. – 208 с. [ЛЕЩ]
4. Лук’яненко І. Г., Краснікова Л. П. – Економетрика. – К.:Знання 1998. – 494 с. [ЛУК]
5. Наконечний C. І., Терещенко Т. О. Економетрія. – К.:КНЕУ, 2006. – 528 с. [НАК]