

Тема 2. Методи побудови загальної лінійної моделі

Лекція 3. Імовірнісний зміст простої регресії

Зміст

1. Узагальнена регресійна модель.
2. Критерій Стюдента для перевірки на значущість параметрів $\hat{a}_0$ і $\hat{a}_1$ знайдених за методом найменших квадратів.
3. Знаходження інтервалів довіри для параметрів a_0 і a_1 за t -розподілом Стюдента.
4. Прогнозування за моделями простої лінійної регресії.

1. Узагальнена регресійна модель.

Економетрична модель – це функція чи система функцій, що описує кореляційно-регресійний зв'язок між економічними показниками, один чи кілька з них є залежною змінною, інші – незалежні.

Незалежні змінні моделі називаються *пояснюючими*, наперед заданими змінними. Залежні змінні називаються *пояснюваними* змінними.

Побудова будь-якої економетричної моделі здійснюється як послідовність певних кроків.

Крок 1. Знайомство з економічною теорією, висунення гіпотези взаємозв'язку. Чітка постановка задачі.

Крок 2. Специфікація моделі.

Специфікація моделі – це аналітична форма економетричної моделі. На основі досліджуваних чинників вона складається з певного виду функції чи функцій, що використовуються для побудови моделей, має імовірнісні характеристики, які притаманні стохастичним залишкам моделі.

Крок 3. Формування масивів вхідної інформації згідно з метою та завданнями дослідження.

Крок 4. Оцінка параметрів економетричної моделі методом найменших квадратів. Аналіз залишків дає змогу перевірити чи не суперечить специфікація моделі передумовам “класичної” лінійної регресії.

Крок 5. Якщо деякі передумови моделі не виконуються, то для продовження аналізу треба замінити специфікацію або застосувати інші методи оцінювання параметрів.

Крок 6. Проведення аналізу достовірності моделі та прогнозу за побудованою моделлю.

Ми розглянули вибіркові лінійні економетричні моделі ($\hat{Y} = \hat{a}_0 + \hat{a}_1 X$), які правильні лише для однієї вибірки. Якщо розглянути всю генеральну сукупність, з якої формується вибірка, то правильною для неї буде модель:

$$Y = a_0 + a_1 X + u$$

Ця модель називається *узагальненою економетричною моделлю*, де a_0 і a_1 – це дійсні параметри всієї генеральної сукупності, а u – *випадкова*, не спостережувана величина (залишок). Виникає питання, як за значеннями параметрів вибіркової моделі ($\hat{a}_0$ і $\hat{a}_1$) зробити висновок щодо параметрів генеральної сукупності (a_0 і a_1).

Доведено, що якщо параметри вибіркової лінійної моделі розраховані за 1МНК, то при певних класичних припущеннях математичні сподівання $\hat{a}_0$ і $\hat{a}_1$ дорівнюють параметрам a_0 і a_1 узагальненої економетричної моделі:

$$M\hat{a}_0 = a_0; M\hat{a}_1 = a_1.$$

Розглянемо припущення, які складають основу класичного економетричного аналізу. **Передумови використання методу найменших квадратів:**

1. Математичне сподівання залишків u має дорівнювати «0»: $M(u) = 0$.
2. Незалежні змінні не повинні бути мультиколінеарними, тобто залежними між собою $|X'X| \neq 0$. Це означає, що матриця X має повний ранг.
3. Незалежні змінні не повинні бути пов'язаними із залишками: $M(x'u) = 0$
4. Повинна спостерігатись гомоскедастичність або однакова дисперсія випадкової величини u . Це означає, що всі випадкові величини u_i , не залежно від номера спостереження мають однакову дисперсію:

$$\sigma_u^2 = \text{const}; M(u'u) = \sigma^2 E, \text{ де } E - \text{одинична матриця.}$$

5. Економетричну модель визначено вірно за умов, якщо:

- 1) в модель включені потрібні змінні;
- 2) правильно вибрана форма моделі.

[До змісту](#)

2. Критерій Стюдента для перевірки на значущість параметрів $\hat{a}_0$ і $\hat{a}_1$ знайдених за методом найменших квадратів.

Перевірку гіпотези про значущість параметрів економетричної моделі можна виконати згідно з t- критерієм:

$$t_j = \frac{|\hat{a}_j|}{\sigma_{aj}}.$$

Обчислене значення t- критерію порівнюється з табличним для вибраного рівня довіри і n-m ступенів вільності. Якщо $t_{\text{факт}} > t_{\text{табл}}$, то відповідний параметр економетричної моделі є достовірним.

Дисперсії параметрів економетричної моделі можна визначити за формулами:

$$\sigma_{A0}^2 = \frac{\sigma_u^2 \sum X^2}{n \sum (X - \bar{X})^2} = \frac{3,325 \times 104784}{10 \times 3376} = 117,62, \sigma_{A0} = 10,85;$$

$$\sigma_{A1}^2 = \frac{\sigma_u^2}{\sum (X - \bar{X})^2} = \frac{3,325}{3376} = 0,00098, \sigma_{A1} = 0,031.$$

Отже, перевіримо гіпотези про значущість оцінок параметрів моделі побудованої на основі вихідних даних, наведених у табл. 1

$$t_1 = \frac{|\hat{a}_1|}{\sigma_{a1}} = \frac{0,5}{0,031} = 16,13;$$

$$t_0 = \frac{|\hat{a}_0|}{\sigma_{a0}} = \frac{3,84}{10,85} = 0,35.$$

Якщо ступінь вільності n-m = 10-2=8 і рівень значущості $\alpha=0,05$ $t_{\text{табл}}=2,306$. Оскільки $t_{1\text{факт}} > t_{\text{табл}}$, то параметр $\hat{a}_1$ є значущим, $t_{0\text{факт}} < t_{\text{табл}}$, то параметр $\hat{a}_0$ є незначущим.

3. Знаходження інтервалів довіри для параметрів a_0 і a_1 за t -розподілом Стюдента.

Для того, щоб визначити, як параметри $\hat{a}_0$ і $\hat{a}_1$, розраховані за методом НМК, тісно пов'язані з параметрами a_0 і a_1 генеральної сукупності, потрібно побудувати інтервали довіри для параметрів узагальненої економетричної моделі. Тобто такі інтервали, в які з заданою ймовірністю попадають їхні значення. Для цього спочатку розрахуємо t - статистику для кожного з параметрів:

$$t = \frac{\hat{a}_i - a_i}{\hat{\sigma}_{a_i}}.$$

Далі вибираємо рівень значущості α , або рівень довіри, який дорівнює $p = 1 - \alpha$. Знаходимо табличне значення t -критерію t_α для ступенів вільності $(n-m)$. Тоді можна записати так: ймовірність того, що t -статистики попадають в інтервал t_α :

$$P\{-t_\alpha < t < t_\alpha\} = 1 - \alpha, \quad t = \frac{\hat{a}_i - a_i}{\hat{\sigma}_{a_i}}, \quad \text{тоді} \quad P\left\{-t_\alpha < \frac{\hat{a}_i - a_i}{\hat{\sigma}_{a_i}} < t_\alpha\right\} = 1 - \alpha$$

$-t_\alpha \hat{\sigma}_{a_i} < \hat{a}_i - a_i < t_\alpha \hat{\sigma}_{a_i}; \quad t_\alpha \hat{\sigma}_{a_i} < -\hat{a}_i + a_i < -t_\alpha \hat{\sigma}_{a_i} + \hat{a}_i < a_i - t_\alpha \hat{\sigma}_{a_i} < \hat{a}_i -$ інтервал довіри для параметру економетричної моделі.

Зробимо відповідні розрахунки за даними нашого прикладу: $t_\alpha = t_{0,05}(8) = 2,306$, $\hat{Y} = 3,8 + 0,5X$; $Y = 3,8 + 0,5X + u$; $\sigma_{a_0} = 3,212$; $\sigma_{a_1} = 0,0314$.

Тоді інтервали довіри будуть мати вигляд:

$$0,5 - 2,306 * 0,0314 < a_1 < 0,5 + 2,306 * 0,0314$$

$$0,5 - 0,0724 < a_1 < 0,5 + 0,0724$$

$$0,4276 < a_1 < 0,5724$$

$$3,8 - 2,306 * 3,212 < a_0 < 3,8 + 2,306 * 3,212$$

$$3,8-7,4080 < a_0 < 3,8+7,4080$$

$$-3,608 < a_0 < 11,208.$$

[До змісту](#)

4. Прогнозування за моделями простої лінійної регресії.

Для подальшого аналізу побудованої економетричної моделі визначимо коефіцієнт детермінації. Він показує на скільки відсотків варіація залежної змінної визначається варіацією незалежної змінної.

Перевірити на значущість коефіцієнт детермінації можна за допомогою F-критерію:

$$F_{k-1, n-k} = \frac{R^2}{k-1} / \frac{1-R^2}{n-k}.$$

Фактичне значення F-критерію порівнюється з табличним при ступенях вільності $k-1$ і $n-k$ і при вибраному рівні значущості. Якщо $F_{\text{факт}} > F_{\text{табл}}$, то гіпотеза про значущість коефіцієнта детермінації підтверджується, у противному разі – відкидається. У нашому випадку :

$$F_{2-1, 10-2} = \frac{0,98^2}{2-1} / \frac{1-0,98^2}{10-2} = 194,02.$$

Табличне значення F-критерію при ступенях вільності 1 і 8 і рівні значущості 0,95 ($F_{1,8}(0,95)$) дорівнює 5,32. Оскільки $F_{\text{факт}} > F_{\text{табл}}$, то можна зробити висновок про значущість коефіцієнта детермінації.

Одним із важливих завдань економетричного моделювання є оцінка прогнозного значення залежної змінної за умови, що пояснювальна змінна задана на перспективу. На основі побудованої економетричної моделі можна отримати точковий прогноз залежної змінної на перспективу. Нехай $x_{\text{прогн}}=130$ ум. од., тоді $y_{\text{прогн}}=3,8 + 0,5 x_{\text{прогн}} = 3,8 + 0,5 \times 130=68,8$ ум. од..

На основі точкового прогнозу можна побудувати інтервальний прогноз скориставшись формулою:

$$y_{\text{прогн}} \pm t_{\text{табл}} \sigma_u \sqrt{1 + \frac{1}{n} + \frac{\Sigma(x_{\text{прогн}} - \bar{x})^2}{\Sigma(x - \bar{x})^2}}.$$

Для нашої моделі за даними табл. 1 інтервальний прогноз для залежної змінної буде такий:

$$68,8 \pm 2,306 \times \sqrt{3,325} \sqrt{1 + \frac{1}{10} + \frac{\Sigma(130 - 100,8)^2}{3376}} = 68,8 \pm 4,204 \times 1,163 = 68,8 \pm 4,889 .$$

Отже з імовірністю $1-\alpha = 0,95$ можна стверджувати, що при рівні фондомісткості в розмірі 130 ум. од. витрати ресурсів на одиницю продукції будуть коливатись в межах від 63, 911 ($68,8-4,889 = 63, 911$) до 73,689 ($68,8 + 4,889 = 73,689$) ум. од.

[До змісту](#)