

Тема 3. Мультиколінеарність та її вплив на оцінки параметрів моделі

Лекція 4. Мультиколінеарність та її вплив на оцінки параметрів моделі

Зміст

1. Поняття мультиколінеарності.....	1
2. Ознаки мультиколінеарності.	2
3. Алгоритм Фаррара-Глобера.....	4
Приклад.....	6
4. Шляхи звільнення від мультиколінеарності.	13

1. Поняття мультиколінеарності.

В економетричних дослідженнях дуже важливо з'ясувати, чи існують між пояснювальними змінними взаємозв'язки, які називають **мультиколінеарністю**.

Мультиколінеарність означає існування тісної лінійної залежності, або сильної кореляції, між двома чи більше пояснювальними змінними.

Вона негативно впливає на кількісні характеристики економетричної моделі або робить її побудову взагалі неможливою.

Так, мультиколінеарність пояснювальних змінних призводить до зміщення оцінок параметрів моделі, через що з їх допомогою не можна зробити коректні висновки про результати взаємозв'язку залежної і пояснювальних змінних. У крайньому разі, коли між пояснювальними змінними існує функціональний зв'язок, оцінити вплив цих змінних на залежну взагалі неможливо. Тоді для оцінювання параметрів моделі метод найменших квадратів не придатний, оскільки матриця $X'X$ буде виродженою.

Нехай зв'язок між пояснювальними змінними не функціональний, проте статистично істотний. Тоді попри те, що оцінити параметри методом найменших квадратів теоретично можливо, знайдена оцінка може призвести

до таких помилкових значень параметрів, що сама модель стане беззмисловою.

Основні наслідки мультиколінеарності.

1. Падає точність оцінювання, яка виявляється так:

- а) помилки деяких конкретних оцінок стають занадто великими;
- б) ці помилки досить корельовані одна з одною;
- в) дисперсії оцінок параметрів різко збільшуються.

[До змісту](#)

2. Ознаки мультиколінеарності.

Оцінки параметрів деяких змінних моделі можуть бути незначущими через наявність їх взаємозв'язку з іншими змінними, а не тому, що вони не впливають на залежну змінну. У такому разі множина вибірових даних не дає змоги цей вплив виявити.

3. Оцінки параметрів стають досить чутливими до обсягів сукупності спостережень. Збільшення сукупності спостережень іноді може спричинитися до істотних змін в оцінках параметрів.

З огляду на перелічені наслідки мультиколінеарності при побудові економетричної моделі потрібно мати інформацію про те, що між пояснювальними змінними не існує мультиколінеарності.

Ознаки мультиколінеарності такі.

1. Коли серед парних коефіцієнтів кореляції пояснювальних змінних є такі, рівень яких наближається або дорівнює множинному коефіцієнту кореляції, то це означає можливість існування мультиколінеарності. Інформацію про парну залежність може дати симетрична матриця коефіцієнтів парної кореляції або кореляції нульового порядку між пояснювальними змінними:

$$r = \begin{bmatrix} r_{x_1 x_1} & r_{x_1 x_2} & r_{x_1 x_3} & \dots & r_{x_1 x_k} \\ r_{x_2 x_1} & r_{x_2 x_2} & r_{x_2 x_3} & \dots & r_{x_2 x_k} \\ r_{x_3 x_1} & r_{x_3 x_2} & r_{x_3 x_3} & \dots & r_{x_3 x_k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_k x_1} & r_{x_k x_2} & r_{x_k x_3} & \dots & r_{x_k x_k} \end{bmatrix}$$

Проте, коли до моделі входять більш як дві пояснювальні змінні, то вивчення питання про мультиколінеарність не може обмежуватись інформацією, що її дає ця матриця. Явище мультиколінеарності в жодному разі не зводиться лише до існування парної кореляції між незалежними змінними.

Більш загальна перевірка передбачає знаходження визначника (детермінанта) матриці r , який називається *детермінантом кореляції* і позначається $|r|$. Числові значення детермінанта кореляції задовольняють умову: $|r| \in [0,1]$.

2. Якщо $|r| = 0$, то існує повна мультиколінеарність, а коли $|r| = 1$, мультиколінеарність відсутня. Чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між пояснювальними змінними існує мультиколінеарність. Незважаючи на те, що на числове значення $|r|$ впливає дисперсія пояснювальних змінних, цей показник можна вважати точковою мірою рівня мультиколінеарності.

3. Якщо в економетричній моделі знайдено мале значення параметра $\hat{a}_k$ при високому рівні частинного коефіцієнта детермінації R_j^2 і при цьому F -критерій істотно відрізняється від нуля, то це також свідчить про наявність мультиколінеарності.

4. Коли коефіцієнт частинної детермінації R_j^2 , який обчислено для регресійних залежностей між однією пояснювальною змінною та іншими, має значення, яке близьке до одиниці, то можна говорити про наявність мультиколінеарності.

5. Нехай при побудові економетричної моделі на основі покрокової регресії введення нової пояснювальної змінної істотно змінює іншу параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи детермінації. Тоді ця змінна перебуває, очевидно, у лінійній залежності від інших, які було введено до моделі раніше.

Усі ці ознаки мультиколінеарності мають один спільний недолік: ні одна з них чітко не розмежовує випадки, коли мультиколінеарність істотна і коли нею можна знехтувати.

[До змісту](#)

3. Алгоритм Фаррара-Глобера.

Найповніше дослідити мультиколінеарність можна за допомогою алгоритму Фаррара-Глобера. Цей алгоритм має три види статистичних критеріїв, згідно з якими перевіряється мультиколінеарність усього масиву незалежних змінних (χ^2 — «хі»-квадрат); кожної незалежної змінної з рештою змінних (F -критерій); кожної пари незалежних змінних (t -критерій).

Усі ці критерії при порівнянні з їх критичними значеннями дають змогу робити конкретні висновки щодо наявності чи відсутності мультиколінеарності незалежних змінних.

Опишемо алгоритм Фаррара-Глобера.

Крок 1. Стандартизація (нормалізація) змінних.

Позначимо вектори незалежних змінних економетричної моделі через $x_1, x_2, x_3 \dots x_m$. Елементи стандартизованих векторів обчислимо за формулами:

$$1) x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sigma_{x_k}}; \quad 2) x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{\sigma_{x_k}^2 n}},$$

де n - число спостережень ($i = \overline{1, n}$); m - число пояснювальних змінних ($k = \overline{1, m}$); $\bar{x}_k$ - середнє арифметичне k -ї пояснювальної змінної; $\sigma_{x_k}^2$ - дисперсія k -ї пояснювальної змінної.

Крок 2. Знаходження кореляційної матриці, виходячи з двох методів нормалізації змінних

$$1) r = \frac{1}{n} X^{*'} X^*; \quad 2) r = X^{*'} X^*,$$

де X^* — матриця стандартизованих незалежних (пояснювальних) змінних; $X^{*'} — матриця, транспонована до матриці X^* .$

Крок 3. Визначення критерію χ^2 («хі»-квадрат):

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2m + 5) \right] \ln |r|,$$

де $|r|$ — визначник кореляційної матриці r .

Значення цього критерію порівнюється з табличним при $\frac{1}{2}m(m-1)$ ступенях вільності і рівні значущості α . Якщо $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$, то в масиві пояснювальних змінних існує мультиколінеарність.

Крок 4. Визначення оберненої матриці:

$$C = r^{-1} = \left(X^{*'} X \right)^{-1}.$$

Крок 5. Обчислення F -критеріїв:

$$F_k = (c_{kk} - 1) \frac{n - m}{m - 1},$$

де c_{kk} - діагональні елементи матриці C . Фактичні значення критеріїв порівнюються з табличними при $n-m$ і $m-1$ ступенях вільності і рівні значущості α . Якщо $F_{k \text{ факт}} > F_{\text{табл}}$, то відповідна k -та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної

$$R_{x_k}^2 = 1 - \frac{1}{c_{kk}}.$$

Крок 6. Знаходження частинних коефіцієнтів кореляції:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}},$$

де c_{kj} - елемент матриці C , що міститься в k -му рядку і j -му стовпці; c_{kk} і c_{jj} - діагональні елементи матриці C .

Крок 7. Обчислення t - критеріїв:

$$t_{kj} = \frac{r_{kj} \sqrt{n - m}}{\sqrt{1 - r_{kj}^2}}.$$

Фактичні значення критеріїв t_{kj} порівнюються з табличними при $n-m$ ступенях вільності і рівні значущості α . Якщо $t_{kj(\phi)} > t_{\text{табл.}}$, то між незалежними змінними x_k і x_j існує мультиколінеарність.

Розглянемо застосування алгоритму Фаррара-Глобера для розв'язування конкретної задачі.

Приклад.

На середньомісячну заробітну плату впливає ряд чинників. Вирізнамо серед них продуктивність праці, фондомісткість та коефіцієнт плинності робочої сили. Щоб побудувати економетричну модель заробітної плати від згаданих чинників згідно з методом найменших квадратів, потрібно переконатися, що продуктивність праці, фондомісткість та коефіцієнт плинності робочої сили як незалежні змінні моделі — не мультиколінеарні.

Вихідні дані наведені в табл. 1.

Таблиця 1.

Вихідні дані

Номер цеху	Продуктивність праці, людино-днів	Фондомісткість, млн. грн.	Коефіцієнт плинності робочої сили, %
1	32	0,89	19,5
2	29	0,43	15,6
3	30	0,70	13,5
4	31	0,61	9,5
5	25	0,51	23,5
6	34	0,51	12,5
7	29	0,65	17,5
8	24	0,43	14,5
9	20	0,51	14,5
10	33	0,92	7,5
Σ	287	6,16	148,1

Дослідити наведені чинники на наявність мультиколінеарності.

Розв'язок**Крок 1.** Нормалізація змінних.

Позначимо вектори незалежних змінних — продуктивності праці, фондомісткості, коефіцієнтів плинності робочої сили — через x_1 , x_2 , x_3 . Елементи стандартизованих векторів обчислимо за формулою:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{\sigma_{x_k}^2 n}},$$

де n — кількість спостережень, $n=10$; m — число незалежних змінних, $m=3$; $\bar{x}_k$ — середнє арифметичне значення вектора x_k ; $\sigma_{x_k}^2$ — дисперсія змінної x_k .

Із формули бачимо, що спочатку потрібно обчислити середні арифметичні для кожної пояснювальної змінної:

$$\bar{x}_1 = \frac{\sum_{i=1}^{10} x_{i1}}{n} = \frac{287}{10} = 28,7; \quad \bar{x}_2 = \frac{\sum_{i=1}^{10} x_{i2}}{n} = \frac{6,16}{10} = 0,616; \quad \bar{x}_3 = \frac{\sum_{i=1}^{10} x_{i3}}{n} = \frac{148,1}{10} = 14,81$$

Усі розрахункові дані для стандартизації змінних x_1 , x_2 , x_3 згідно з поданими співвідношеннями наведено в табл. 2.

Дисперсії кожної незалежної змінної мають такі значення:

$$\sigma_{x_1}^2 = \frac{\sum_{i=1}^{10} (x_{i1} - \bar{x}_1)^2}{n} = \frac{176,1}{10} = 17,61; \quad \sigma_{x_2}^2 = \frac{\sum_{i=1}^{10} (x_{i2} - \bar{x}_2)^2}{n} = \frac{0,27864}{10} = 0,027864;$$

$$\sigma_{x_3}^2 = \frac{\sum_{i=1}^{10} (x_{i3} - \bar{x}_3)^2}{n} = \frac{182,4}{10} = 18,24.$$

Тоді знаменник для стандартизації кожної незалежної змінної буде такий:

$$x_1 : \sqrt{n\sigma_{x_1}^2} = \sqrt{10 \cdot 17,61} = 13,2703;$$

$$x_2 : \sqrt{n\sigma_{x_2}^2} = \sqrt{10 \cdot 0,027864} = 0,5279;$$

$$x_3 : \sqrt{n\sigma_{x_3}^2} = \sqrt{10 \cdot 18,24} = 13,5056.$$

Тоді знаменник для стандартизації кожної незалежної змінної буде такий:

$$x_1 : \sqrt{n\sigma_{x_1}^2} = \sqrt{10 \cdot 17,61} = 13,2703;$$

$$x_2 : \sqrt{n\sigma_{x_2}^2} = \sqrt{10 \cdot 0,027864} = 0,5279;$$

$$x_3 : \sqrt{n\sigma_{x_3}^2} = \sqrt{10 \cdot 18,24} = 13,5056.$$

Таблиця 2.

Розрахунок нормалізованих змінних

№ п/п	X_1	X_2	X_3	$X_1 - X_{1(сеп)}$	$X_2 - X_{2(сеп)}$	$X_3 - X_{3(сеп)}$	$(X_1 - X_{1(сеп)})^2$	$(X_2 - X_{2(сеп)})^2$	$(X_3 - X_{3(сеп)})^2$	X_1^*	X_2^*	X_3^*	$X_1^* * X_2^*$	$X_1^* * X_3^*$	$X_2^* * X_3^*$
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	32	0,89	19,5	3,300	0,274	4,690	10,890	0,075	21,996	0,249	0,519	0,337	0,129	0,084	0,175
2	29	0,43	15,6	0,300	-0,186	0,790	0,090	0,035	0,624	0,023	-0,352	0,057	-0,008	0,001	-0,020
3	30	0,7	13,5	1,300	0,084	-1,310	1,690	0,007	1,716	0,098	0,159	-0,094	0,016	-0,009	-0,015
4	31	0,61	9,5	2,300	-0,006	-5,310	5,290	0,000	28,196	0,173	-0,011	-0,381	-0,002	-0,066	0,004
5	25	0,51	23,5	-3,700	-0,106	8,690	13,690	0,011	75,516	-0,279	-0,201	0,624	0,056	-0,174	-0,125
6	34	0,51	12,5	5,300	-0,106	-2,310	28,090	0,011	5,336	0,399	-0,201	-0,166	-0,080	-0,066	0,033
7	29	0,65	17,5	0,300	0,034	2,690	0,090	0,001	7,236	0,023	0,064	0,193	0,001	0,004	0,012
8	24	0,43	14,5	-4,700	-0,186	-0,310	22,090	0,035	0,096	-0,354	-0,352	-0,022	0,125	0,008	0,008
9	20	0,51	14,5	-8,700	-0,106	-0,310	75,690	0,011	0,096	-0,656	-0,201	-0,022	0,132	0,015	0,004
10	33	0,92	7,5	4,300	0,304	-7,310	18,490	0,092	53,436	0,324	0,576	-0,524	0,187	-0,170	-0,302
Всього	287	6,16	148	0	0	0	176,1	0,27864	194,249	0	0	0	0,555	-0,373	-0,225

Матриця стандартизованих змінних подається у вигляді:

$$X^* = \begin{bmatrix} 0,2487 & 0,519 & 0,337 \\ 0,0226 & -0,352 & 0,057 \\ 0,0980 & 0,159 & -0,094 \\ 0,1733 & -0,011 & -0,381 \\ -0,2788 & -0,201 & 0,624 \\ 0,3994 & -0,201 & -0,166 \\ 0,0226 & 0,064 & 0,193 \\ -0,3542 & -0,352 & -0,022 \\ -0,6556 & -0,201 & -0,022 \\ 0,3240 & 0,576 & -0,524 \end{bmatrix}$$

Крок 2. Знаходження кореляційної матриці:

$$r = X^{*'} X^*,$$

де X^* — матриця нормалізованих пояснювальних змінних; $X^{*'} —$ матриця, транспонована до матриці X^* .

Ця матриця симетрична і має розмір 3x3.

Для даної задачі

$$r = \begin{bmatrix} 1 & 0,494 & -0,551 \\ 0,494 & 1 & -0,5168 \\ -0,551 & -0,5168 & 1 \end{bmatrix}$$

Кожний елемент цієї матриці характеризує тісноту зв'язку однієї незалежної змінної з іншою. Оскільки діагональні елементи характеризують тісноту зв'язку кожної незалежної з цією самою змінною, то вони дорівнюють одиниці. Зауважимо, що при знаходженні добутку матриць $X^{*'} і X^*$ за рахунок зміщеності коефіцієнтів парної кореляції числові значення діагональних елементів можуть наближатись до одиниці. Якщо це так, то вони замінюються одиницями, а інші значення матриці r збільшуються на величину, що визначається як різниця між одиницею і діагональним елементом.

Інші елементи матриці r дорівнюють:

$$r_{x_1x_2} = 0,494; \quad r_{x_1x_3} = -0,551; \quad r_{x_2x_3} = -0,5168,$$

тобто вони є парними коефіцієнтами кореляції між пояснювальними змінними. Користуючись цими коефіцієнтами, можна зробити висновок, що між змінними x_1 , x_2 , x_3 існує зв'язок. Але чи можна стверджувати, що цей зв'язок є виявленням мультиколінеарності, а через це негативно впливатиме на оцінку економетричної моделі?

Щоб відповісти на це запитання, потрібно ще раз звернутися до алгоритму Фаррара-Глобера і знайти статистичні критерії оцінки мультиколінеарності.

Крок 3. Обчислимо детермінант кореляційної матриці r і критерій χ^2 :

а) $D = |r| = 0,466$;

б) $\chi^2 = -\left[n - 1 - \frac{1}{6}(2m + 5)\right] \ln|r| = -\left[9 - \frac{1}{6}(6 + 5)\right] \ln 0,466 = 2,37$.

При ступені вільності $\frac{1}{2}m(m-1)=3$ і рівні значущості $\alpha=0,01$ критерій $\chi^2_{\text{табл.}}=11,34$. Оскільки $\chi^2_{\text{факт}} < \chi^2_{\text{табл.}}$, робимо висновок, що в масиві пояснювальних змінних не існує мультиколінеарності.

Крок 4. Знайдемо матрицю, обернену до матриці r :

$$C = r^{-1} = (X^*{}' X^*)^{-1};$$

$$C = \begin{bmatrix} 1,57 & -0,45 & 0,63 \\ -0,45 & 1,49 & 0,52 \\ 0,63 & 0,52 & 1,62 \end{bmatrix}.$$

Крок 5. Використовуючи діагональні елементи матриці C , обчислимо F -критерії:

$$F_1 = (c_{11} - 1) \frac{n - m}{m - 1} = (1,57 - 1) \frac{7}{2} = 2,00;$$

$$F_2 = (c_{22} - 1) \frac{n - m}{m - 1} = (1,49 - 1) \frac{7}{2} = 1,72;$$

$$F_3 = (c_{33} - 1) \frac{n - m}{m - 1} = (1,62 - 1) \frac{7}{2} = 2,17.$$

Для рівня значущості $\alpha = 0,05$ і ступенів вільності $\gamma_1 = 7$ і $\gamma_2 = 2$ критичне (табличне) значення критерію $F = 4,74$.

Оскільки

$$F_{1\text{факт}} < F_{\text{табл}};$$

$$F_{2\text{факт}} < F_{\text{табл}};$$

$$F_{3\text{факт}} < F_{\text{табл}},$$

то кожна з пояснювальних змінних не мультиколінеарна з двома іншими.

Щоб визначити наявність попарної мультиколінеарності, продовжимо дослідження і перейдемо до кроку 6.

Крок 6. Обчислимо частинні коефіцієнти кореляції, скориставшись елементами матриці C :

$$r_{12,3} = \frac{-c_{12}}{\sqrt{c_{11} \cdot c_{22}}} = \frac{0,49}{\sqrt{1,57 \cdot 1,49}} = 0,293;$$

$$r_{13,2} = \frac{-c_{13}}{\sqrt{c_{11} \cdot c_{33}}} = \frac{-0,63}{\sqrt{1,57 \cdot 1,62}} = -0,397;$$

$$r_{23,1} = \frac{-c_{23}}{\sqrt{c_{22} \cdot c_{33}}} = \frac{-0,52}{\sqrt{1,49 \cdot 1,62}} = -0,337.$$

Частинні коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що третя не впливає на цей зв'язок.

Порівнявши частинні коефіцієнти кореляції з парними, які було наведено раніше, можна помітити, що частинні коефіцієнти значно менші за парні. Це ще раз показує, що на підставі парних коефіцієнтів кореляції не можна зробити висновків про наявність мультиколінеарності чи її відсутність.

Крок 7. Визначимо t -критерій на основі частинних коефіцієнтів кореляції.

$$t_{12} = \frac{r_{12,3} \sqrt{n-m}}{\sqrt{1-r_{12,3}^2}} = \frac{0,293\sqrt{7}}{\sqrt{1-0,0858}} = 0,811;$$

$$t_{13} = \frac{r_{13,2} \sqrt{n-m}}{\sqrt{1-r_{13,2}^2}} = \frac{0,39\sqrt{7}}{\sqrt{1-0,1521}} = 1,146;$$

$$t_{23} = \frac{r_{23,1} \sqrt{n-m}}{\sqrt{1-r_{23,1}^2}} = \frac{0,34\sqrt{7}}{\sqrt{1-0,1156}} = 0,947.$$

Табличне значення t -критерію при $n-m = 7$ ступенях вільності і рівні значущості $\alpha = 0,05$ дорівнює 1,69. Усі числові значення t -критеріїв,

знайдених для кожної пари змінних, менші за їх табличні значення. Звідси робимо висновок, що всі пари незалежних змінних не є мультиколінеарними.

Отже, усі пояснювальні змінні досліджуваної моделі не мультиколінеарні.

[До змісту](#)

4. Шляхи звільнення від мультиколінеарності.

Якщо F -критерій більший за табличне значення, тобто коли k -та змінна залежить від усіх інших у масиві, то необхідно вирішувати питання про її вилучення з переліку змінних. Якщо t_{kj} — критерій більший за табличний, то ці дві змінні (k і j) тісно пов'язані одна з одною. Звідси, аналізуючи рівень обох видів критеріїв F і t , можна зробити обґрунтований висновок про те, яку зі змінних необхідно вилучити з дослідження або замінити іншою. Проте заміна масиву незалежних змінних завжди має узгоджуватись з економічною доцільністю, що впливає з мети дослідження.

Найпростіше позбутися мультиколінеарності в економетричній моделі можна, відкинувши одну зі змінних мультиколінеарної пари. Але на практиці вилучення якогось чинника часто суперечить логіці економічних зв'язків. Тоді можна перетворити певним чином пояснювальні змінні моделі:

- а) узяти відхилення від середньої;
- б) замість абсолютних значень узяти відносні;
- в) стандартизувати пояснювальні змінні тощо.

За наявності мультиколінеарності змінних потрібно звертати увагу й на специфікацію моделі. Іноді заміна однієї функції іншою, якщо це не суперечить апріорній інформації, дає змогу уникнути явища мультиколінеарності. Коли жодний з розглянутих способів не дає змоги позбутися мультиколінеарності, то параметри моделі слід оцінювати за методом головних компонентів.

[До змісту](#)