МІЖРЕГІОНАЛЬНА
АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ЩОДО ЗАБЕЗПЕЧЕННЯ ПРАКТИЧНИХ І ЛАБОРАТОРНИХ ЗАНЯТЬ

з дисципліни

“ЕКОНОМЕТРІЯ”

(для бакалаврів)

Київ
ДП «Видавничий дім «Персонал»
2009

Підготовлено кандидатом фіз.-мат. наук, професором кафедри математики *І. І. Юртиним* і кандидатом фіз.-мат. наук, доцентом кафедри математики *О. О. Юньковою*

Затверджено на засіданні кафедри математики (протокол № 2 від 26.10.07)

Схвалено Вченою радою Міжрегіональної Академії управління персоналом

Юртин І. І., Юнькова О. О. Методичні рекомендації щодо забезпечення практичних і лабораторних занять з дисципліни “Економетрія” (для бакалаврів). — К.: ДП «Вид. дім «Персонал», 2009. — 31 с.

Методичні рекомендації містять пояснювальну записку, тематичний план дисципліни, зміст практичних завдань, вказівки до виконання лабораторних робіт та зразки розв’язування деяких завдань, а також список літератури.

ПОЯСНЮВАЛЬНА ЗАПИСКА

Мета викладання дисципліни — дати студентам знання і сформувати такі навички: 1) створення математичних моделей, які описують кількісні залежності між економічними показниками; 2) оцінювання параметрів таких моделей на основі статистичних даних щодо значень відповідних показників; 3) перевірки адекватності розроблених моделей реальним економічним явищам та процесам; 4) застосування цих моделей для аналізу і прогнозування розвитку досліджуваних явищ.

Основні завдання викладання дисципліни — дати студентам систематизовані знання:

- суті й етапів економетричного дослідження;
- основних принципів та прийомів математичного моделювання залежностей між економічними показниками;
- методів оцінювання параметрів регресійних рівнянь та програмного забезпечення обчислень.

А також сформувати уміння:

- постановки та формалізації задач економетричного моделювання;
- класифікації моделей;
- оцінювання параметрів парної лінійної регресії та її аналізу;
- оцінювання параметрів парної нелінійної регресії та її аналізу;
- побудови моделі лінійної множинної регресії та оцінювання параметрів методом найменших квадратів;
- перетворення нелінійних залежностей між показниками до лінійного вигляду й оцінювання їх параметрів методом найменших квадратів;
- побудови динамічних моделей на основі часових рядів;
- оцінювання параметрів систем одночасних рівнянь;
- використання програмного забезпечення Excel на ПЕОМ при проведенні розрахунків та аналізу результатів;
- проведення аналізу побудованих моделей та розробки практичних рекомендацій з їх застосування.

ТЕМАТИЧНИЙ ПЛАН
дисципліни
“ЕКОНОМЕТРІЯ”

№ пор.	Назва змістового модуля і теми	Лекції	Практ.	Самост. роб. студ.	Разом (годин)
	Змістовий модуль I. Парна і багатофакторна лінійна регресія				
1	Вступ. Математичне моделювання як метод наукового пізнання економічних явищ і процесів	2	0	5	7
2	Побудова та дослідження парної лінійної регресійної моделі	4	4	4	12
3	Загальна лінійна економетрична модель. Багатофакторна регресія	4	2	4	10
	Змістовий модуль II. Особливості застосування МНК для багатофакторних моделей				
4	Мультиколінеарність	2	2	6	10
5	Автокореляція в економетричних моделях динаміки	2	2	6	10
6	Гомо- та гетероскедастичність	2	2	6	10
	Змістовий модуль III. Динамічні моделі				
7	Моделювання часових рядів	2	0	2	4
8	Лагові моделі в економіці	2	0	8	8
	Змістовий модуль IV. Системи одночасних рівнянь				
9	Одночасні рівняння	2	0	8	10
Разом годин: 81		22	12	49	81

ЗМІСТ ПРАКТИЧНИХ ЗАНЯТЬ

На практичних заняттях, загальний обсяг яких складає 12 год., необхідно виконати лабораторні роботи з таких тем:

1. Побудова та дослідження парної регресії.
2. Побудова та дослідження багатофакторної регресії.
3. Перевірка даних на наявність мультиколінеарності.

Перша лабораторна робота виконується на трьох практичних заняттях, друга — на двох практичних занять, третя — на одному практичному занятті.

Рекомендації щодо виконання лабораторних робіт наведені у конкретних задачах.

Лабораторна робота 1. Парна регресія

***Тема*: Парна лінійна регресія**

1. Побудова системи нормальних рівнянь.
2. Оцінювання параметрів регресії.
3. Побудова графіків залежностей.
4. Коефіцієнти кореляції та детермінації. Значущість коефіцієнта кореляції.
5. Перевірка моделі на адекватність за F-критерієм Фішера.
6. Коефіцієнт еластичності.
7. Прогнозування за моделлю.
8. Побудова та дослідження нелінійної моделі.

Література [1–4; 8; 11; 21]

Задача 1. Нехай задано обсяги споживання X (у. о.) домогосподарства протягом року на підставі вибірки $n = 12$ спостережень (щомісячно впродовж року), які наведені в табл. 1.

Таблиця 1

t_i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150

На основі цих статистичних даних:

1) побудувати графік парної лінійної регресії $x(t) = \hat{a}_0 + \hat{a}_1 t$;
2) оцінити всі її параметри;
3) визначити довірчі інтервали для параметрів регресії a_0, a_1 при рівні значущості $\alpha = 0{,}05$;

4) знайти коефіцієнти детермінації R^2 та кореляції R;
5) обчислити прогнозні значення показника X для наступних трьох місяців ($x(13), x(14), x(15)$);
6) побудувати графік парної лінійної регресії $y(x) = a_0 + a_1 x$;
7) виконати попередні пункти 2–5 для моделі $y(t) = b_0 + b_1 t$.

Розв'язання

Виконувати завдання будемо засобами Excel. Час t у цій задачі є незалежною змінною, а витрати на споживання x — залежною змінною.

Початкову інформацію запишемо у таблиці Excel: заголовки робочої таблиці будемо писати у першому рядку, числові дані — починаючи з другого рядка. У перших двох стовпцях запишемо початкові дані: значення незалежної змінної — у стовпці A (діапазон A2 : A13), значення залежної — у стовпці B (діапазон B2 : B13).

Для обчислення оцінок параметрів лінійної моделі $x(t) = \hat{a}_0 + \hat{a}_1 t$ використаємо формули:

$$\hat{a}_1 = \frac{\overline{tx} - \bar{t} \cdot \bar{x}}{\overline{t^2} - (\bar{t})^2}, \quad \hat{a}_0 = \bar{x} - \hat{a}_1 \bar{t}.$$

Для цього виконаємо додаткові розрахунки.

Обчислимо для всіх спостережень квадрати незалежної змінної $t_i^2,\ i = 1, 2, ..., n,$ і попарні добутки значень залежної і незалежної змінних $t_i\, x_i, i = 1, 2, ..., n$. Для цього у комірках C3 і D3 набрати відповідно формули: “=A2*A2” і “=A2*B2”. Після отримання результату продовжити формули у діапазонах C4 : C13; D4 : D13 (курсор у правому нижньому кутку у вигляді знака “+”).

Середні значення $\bar{t}, \bar{x}, \overline{t^2}, \overline{tx}$ визначаємо за допомогою функції СРЗНАЧ програми Excel і, підставляючи їх у наведені формули, отримаємо: $\hat{a}_0 = 3{,}78$, $\hat{a}_1 = 96{,}08$.

Отже, рівняння парної лінійної регресії має вигляд:

$$x(t) = 96{,}08 + 3{,}78t.$$

У стовпці E запишемо розрахункові (модельні) значення $\hat{x}_i$, які обчислюємо за формулою

$$\hat{x}_i = 96{,}08 + 3{,}78 t_i.$$

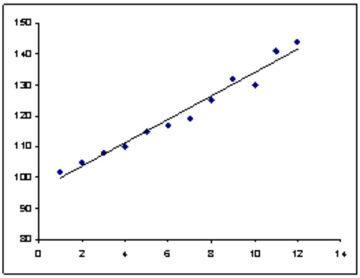


Рис. 1

Застосовуючи майстер діаграм Excel, графічно зобразимо початкові статистичні дані (t_i, x_i) — кореляційне поле та модельні значення $(t_i, \hat{x}_i)$ — графік тренду.

Залишки моделі $u_i = x_i - \hat{x}_i$, $i = 1,..., n$, та їх квадрати обчислимо у стовпцях F і G, набираючи формули в F2 (“=B2-E2”) та в G2 (“=F2*F2”) і продовжуючи їх для всіх спостережень. Суму квадратів усіх залишків $\sum_{i=1}^{n} u_i^2$ обчислимо за допомогою функції СУММ, а середнє цих же значень дає вибіркову дисперсію залишків $S_u^2 = \frac{1}{n}\sum_{i=1}^{n} u_i^2$ (функція СРЗНАЧ).

Далі обчислюємо коефіцієнти детермінації R^2 та кореляції R:

$$R^2 = 1 - \frac{S_u^2}{S_x^2},$$

де $S_x^2 = \overline{x^2} - (\overline{x})^2$; $R = \sqrt{R^2}$.

Маємо: $R^2 = 0{,}97$, $R = 0{,}99$.

Обчислене значення коефіцієнта кореляції дає змогу зробити висновок про сильну (пряму) лінійну залежність між змінними t та x. Це також підтверджується розташуванням точок (t_i, x_i) і графіка тренду на кореляційному полі.

Прогнозоване споживання для наступних трьох місяців також визначається за формулою регресії, тобто $X(t) = 96{,}08+3{,}78*t$, де $t = 13,14,15$. Значення легко отримати, задаючи прогнозні значення незалежної змінної у першому стовпці і продовжуючи формулу у стовпці модельних значень.

Побудоване рівняння регресії у будь-якому разі потребує певної інтерпретації та аналізу.

У нашому прикладі коефіцієнт $\hat{a}_1$ може розглядатися як гранична схильність до споживання. Фактично він показує, на яку величину зміниться обсяг споживання у наступному місяці, якщо тенденції минулого періоду залишаться незмінними.

Вільний член $\hat{a}_0$ рівняння регресії визначає прогнозне значення x при змінній t, що дорівнює нулю (тобто автономне споживання). У нашому випадку значення $\hat{a}_0 = 96{,}08$ у. о. Цей параметр може визначати накопичені або позичені кошти.

Необхідно пам'ятати, що емпіричні коефіцієнти регресії $\hat{a}_0$ та $\hat{a}_1$ є лише оцінками теоретичних коефіцієнтів a_0 та a_1 а власне рівняння відображає лише загальну тенденцію у поведінці розглянутих змінних.

При змінюванні статистичної бази (початкових даних) результати оцінювання, очевидно, відрізнятимуться від попередніх, але з високою ймовірністю можуть опинитися в певних межах — у межах довірчого інтервалу параметрів регресії. Цей інтервал визначається для кожного параметра за формулами:

$$\text{для } \hat{a}_0 \ (\hat{a}_0 - \Delta_0,\ \hat{a}_0 + \Delta_0),$$
$$\text{для } \hat{a}_1 \ (\hat{a}_1 - \Delta_1,\ \hat{a}_1 + \Delta_1),$$

де $\Delta_0 = t_{\text{табл}}\sqrt{\sigma_u^2 c_{00}}$; $\Delta_1 = t_{\text{табл}}\sqrt{\sigma_u^2 c_{11}}$; для парної регресії $\sigma_u^2 = \dfrac{1}{n-2}\sum_{i=1}^{n} u_i^2$ — незміщена дисперсія залишків;

$c_{00} = \dfrac{\overline{t^2}}{n\sigma_t^2}$, $c_{11} = \dfrac{1}{n\sigma_t^2}$, $\sigma_t^2 = \overline{t^2} - (\overline{t})^2$; $t_{\text{табл}} = t(\dfrac{\alpha}{2}, n-2)$ — табличне значення розподілу Стьюдента. Його визначають із стандартної таблиці розподілу Стьюдента [3, дод. 2] або за допомогою статистичної функції СТЬЮДРАСПОБР, параметри якої вводять з клавіатури: у запиті "Вероятность" — 0,05/2; у запиті "Степени свободы" — 12-2·

Для аналізу щільності лінійної залежності обчислимо коефіцієнт кореляції:

$$r_{tx} = \frac{\overline{tx} - \overline{x}\cdot\overline{t}}{\sqrt{\overline{t^2} - (\overline{t})^2}\cdot\sqrt{\overline{x^2} - (\overline{x})^2}}.$$

Таблиця розрахунків набуває вигляду:

A	B	C	D	E	F	G	H
T	x	t^2	tx	$\hat{x}$	u	u^2	x^2
1	102	1	102	99,85	2,14	4,58	10404
2	105	4	210	103,64	1,35	1,84	11025
3	108	9	324	107,42	0,57	0,33	11664
4	110	16	440	111,20	-1,20	1,460	12100
5	115	25	575	114,99	0,008	0	13225
6	117	36	702	118,77	-1,778	3,150	13689
7	119	49	833	122,55	-3,55	12,66	14161
8	125	64	1000	126,34	-1,34	1,80	15625
9	132	81	1188	130,12	1,87	3,51	17424
10	130	100	1300	133,90	-3,91	15,27	16900
11	141	121	1551	137,69	3,31	10,95	19881
12	144	144	1728	141,47	2,52	6,38	20736
$\overline{t}=$	$\overline{x}=$	$\overline{t^2}=$	$\overline{tx}=$			$\sum u^2=$	$\overline{x^2}=$
6,5	120,66	54,166	829,41			61,94	14736,16
$a_1=$	3,78	Sa1=	0,21		$Su^2=$	5,16	
$a_0=$	96,07	Sa0=	1,53		Sigma $u^2=$	6,19	
$a_0 \in$	(93,30	98,85)			Ttabl=	1,81	
$a_1 \in$	(3,41	4,16)			$C_{00}=$	0,38	
			$R_{tx}=$	0,99	$C_{11}=$	0,007	
			$R^2=$	0,97			

Задача 2. Парна нелінійна регресія

Є такі статистичні дані, які відображають залежність між двома показниками X та Y:

X	1	1,5	2	3	3,5	4,5	6	7	8	9	11	12
Y	48	45	27	10	11	5	1	9	8	13	30	25

За наведеними статистичними даними:

- оцінити параметри парної лінійної моделі $y = a_0 + a_1 x$ (виконати пункти 1–5 завдання 1);
- підібрати іншу форму регресійного рівняння, обчислити параметри моделі (виконати пункти 1–5 завдання 1);
- визначити довірчі інтервали для статистичних даних y_i.

Розв'язання

Здійснивши побудову парної лінійної моделі $y = a_0 + a_1 x$ за алгоритмом пунктів 1–5 завдання 1, дістанемо такі результати:

$y = 26{,}2 - 1{,}2x,\ R^2 = 0{,}083,\ R = 0{,}288.$

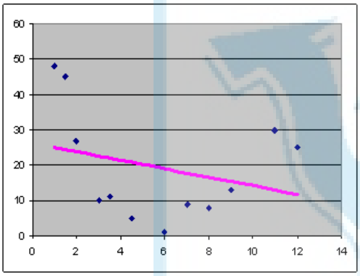


Рис. 2

Очевидно, що побудована модель не передає залежності між показниками. Розташування точок на (рис. 2) наводить на думку про існування іншої форми залежності. Такою залежністю може бути параболічна: $y = a_0 + a_1 (x - x_0)^2$, де $a_0,\ a_1$ — невідомі параметри моделі, а x_0 — відомий параметр, який визначається експериментально.

Згідно із графіком такою величиною може бути вершина параболи, тобто величина $x_0 = 6$. Надалі її можна уточнювати таким чином, щоб відобразити сильнішу залежність між x та y. Для цього застосовується значення коефіцієнта кореляції R.

Для знаходження параметрів $a_0,\ a_1$ здійснюється лінеаризація моделі шляхом заміни $t = (x - x_0)^2$, яка приводить до лінійної моделі:

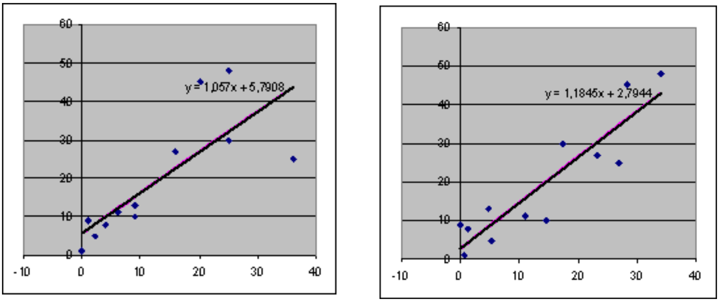
$$y = b_0 + b_1 t\ .$$

Статистичні дані при цьому набувають вигляду:

T	33,87	28,30	23,23	14,59	11,02	5,38	0,67	0,03	1,39	4,75	17,47	26,83
Y	48	45	27	10	11	5	1	9	8	13	30	25

Здійснивши побудову парної лінійної моделі $y = b_0 + b_1 t$ за алгоритмом пунктів 1–5 завдання 1, дістанемо такі результати:

$y = 5{,}79 + 1{,}06t,\ R^2 = 0{,}62,\ R = 0{,}78.$



$x_0 = 6,\ R = 0{,}78$ $\qquad$ $x_0 = 6{,}82,\ R = 0{,}908$

Рис. 3

Величину R можна збільшити, змінюючи параметр x_0. Наприклад, для $x_0 = 6{,}82$ маємо $R = 0{,}908$, що є кращим, ніж у попередньому випадку. Кореляційне поле і графіки трендів для різних значень x_0 зображено на рис. 3.

Довірчі, або надійні інтервали для y_i визначаються за формулою

$$\left(\hat{y}_i - \Delta y_i,\ \hat{y}_i + \Delta y_i\right),$$

де $\Delta y_i = t_{\text{табл}} \dfrac{S}{\sqrt{n}} \sqrt{1 + \dfrac{(t_i - \bar{t}\,)^2}{\sigma_t^2}}$; $\quad t_{\text{табл}} = t(\alpha / 2, n-2)$ — табличне значення розподілу Стьюдента; S — стандартне відхилення залишків $u_i = y_i - \hat{y}_i$; $\sigma_x^2 = \overline{t^2} - (\bar{t})^2$ — дисперсія статистичної вибірки t_i.

Обчислимо довірчі інтервали, графічно їх зобразимо на рис. 4.

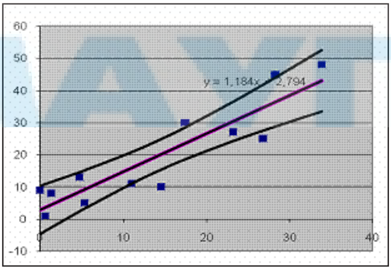


Рис. 4

Лабораторна робота 2. Багатофакторна регресія

***Тема*: Багатофакторна регресія**

1. Оцінювання параметрів рівнянь.
2. Застосування вбудованих функцій Excel.
3. Коефіцієнти кореляції та детермінації. Значущість коефіцієнта кореляції.
4. Перевірка моделі на адекватність за *F*-критерієм Фішера.
5. Коефіцієнт еластичності.
6. Прогнозування за моделлю.

Література [1–4; 8; 11; 21]

Задача 3

На основі $n = 15$ статистичних даних певного регіону:

1) визначити параметри лінійної моделі залежності прибутку підприємства (Y) від рівня інвестицій (I), витрат на рекламу (Cr) та заробітної плати (L);
2) оцінити коефіцієнт детермінації R^2.

i	Y	$X_1\,(I)$	$X_2\,(Cr)$	$X_3\,(L)$
1	15,70	17,37	5,28	1,42
2	17,34	18,24	6,47	1,58
3	21,57	22,47	6,98	1,98
4	33,50	18,47	7,05	2,04
5	32,30	16,82	7,94	2,38
6	37,90	17,60	8,12	3,48
7	40,78	17,12	8,69	3,07
8	48,02	19,81	9,31	3,84
9	43,30	18,67	10,45	4,28
10	49,57	20,83	10,47	4,67
11	52,14	22,84	13,48	5,98
12	55,17	28,85	15,78	6,51
13	59,18	29,61	17,65	7,82
14	62,22	35,67	18,47	8,58
15	77,58	47,87	19,64	9,47

Розв'язання

1-й крок. Оцінювання параметрів

Загальна лінійна модель має вигляд:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + u, \qquad (1)$$

де y — результативна (залежна) змінна; Y — прибуток підприємства; x_1, x_2, x_3 — незалежні, факторні змінні (I, Cr і L відповідно); $a_0, a_1, ..., a_m$ — параметри моделі; u — випадкова складова регресійного рівняння.

1.1. Оцінювання параметрів моделі $a_0, a_1, ..., a_m$ виконаємо методом найменших квадратів, матричний запис якого має вигляд:

$$A = (X^T X)^{-1}(X^T Y),$$

де $A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix}$ – вектор невідомих параметрів.

Складемо вектор-стовпець і матрицю спостережень у вигляді:

$$Y = \begin{pmatrix} 15{,}7 \\ 17{,}34 \\ 21{,}57 \\ 33{,}5 \\ 32{,}3 \\ 37{,}9 \\ 40{,}78 \\ 48{,}02 \\ 43{,}3 \\ 49{,}57 \\ 52{,}14 \\ 55{,}17 \\ 59{,}18 \\ 62{,}22 \\ 77{,}58 \end{pmatrix}; \qquad X = \begin{pmatrix} 1 & 17{,}37 & 5{,}28 & 1{,}42 \\ 1 & 18{,}24 & 6{,}47 & 1{,}58 \\ 1 & 22{,}47 & 6{,}98 & 1{,}98 \\ 1 & 18{,}47 & 7{,}05 & 2{,}04 \\ 1 & 16{,}82 & 7{,}94 & 2{,}38 \\ 1 & 17{,}6 & 8{,}12 & 3{,}48 \\ 1 & 17{,}12 & 8{,}69 & 3{,}07 \\ 1 & 19{,}81 & 9{,}31 & 3{,}84 \\ 1 & 18{,}67 & 10{,}45 & 4{,}28 \\ 1 & 20{,}83 & 10{,}47 & 4{,}67 \\ 1 & 22{,}84 & 13{,}48 & 5{,}98 \\ 1 & 28{,}85 & 15{,}78 & 6{,}51 \\ 1 & 29{,}61 & 17{,}65 & 7{,}82 \\ 1 & 35{,}67 & 18{,}47 & 8{,}58 \\ 1 & 47{,}87 & 19{,}64 & 9{,}47 \end{pmatrix}.$$

Стовпчик одиниць у матриці X відповідає коефіцієнту 1 при параметрі a_0.

Виконувати розрахунки будемо за допомогою вбудованих функцій Excel:

1) X' – функція ТРАНСП(*массив*) (ТРАНСПонована матриця) із категорії “ссылки и массивы”;
2) $X'X, X'Y, A$ – функція МУМНОЖ(*массив*1, *массив*2) (Матричное УМНОЖение) із категорії “математические”;
3) $(X'X)^{-1}$ – функція МОБР(*массив*) (Матрица ОБРатная) також із категорії “математические”.

Для роботи із вказаними функціями потрібно:

1) виділити місце під результат (діапазон комірок);
2) викликати функцію (натиснути кнопку f_x на панелі інструментів, вказати категорію, вибрати функцію);
3) вказати аргументи функції (у тому порядку, як вони записані у формулі);
4) після виходу з діалогового вікна функції у рядку формул натиснути ліву клавішу мишки (аргументи виділяться рамками), а потім одночасно натиснути на клавіатурі три клавіші (Ctrl+Shift+Enter).

Врешті отримаємо такі результати:

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 17,37 & 18,24 & 22,47 & 18,47 & 16,82 & 17,6 & 17,12 & 19,81 & 18,67 & 20,83 & 22,84 & 28,85 & 29,61 & 35,67 & 47,87 \\ 5,28 & 6,47 & 6,98 & 7,05 & 7,94 & 8,12 & 8,69 & 9,31 & 10,45 & 10,47 & 13,48 & 15,78 & 17,65 & 18,47 & 19,64 \\ 1,42 & 1,58 & 1,98 & 2,04 & 2,38 & 3,48 & 3,07 & 3,84 & 4,28 & 4,67 & 5,98 & 6,51 & 7,82 & 8,58 & 9,47 \end{pmatrix};$$

$$X'X = \begin{pmatrix} 15 & 352,24 & 165,78 & 67,1 \\ 352,24 & 9335,74 & 4404,383 & 1858,071 \\ 165,78 & 4404,38 & 2147,268 & 914,9516 \\ 67,1 & 1858,07 & 914,9516 & 397,2576 \end{pmatrix};$$

$$(X'X)^{-1}=\begin{pmatrix} 2,14866 & -0,0276 & -0,51745 & 0,958316 \\ -0,0276 & 0,00428 & -0,0056 & -0,00245 \\ -0,5174 & -0,0056 & 0,18797 & -0,31932 \\ 0,95831 & -0,0024 & -0,31932 & 0,58755 \end{pmatrix};$$

$$X'Y=\begin{pmatrix} 646,27 \\ 16861,1 \\ 8209,78 \\ 3498,18 \end{pmatrix}; \qquad A=\begin{pmatrix} 26,10789 \\ -0,2518 \\ -2,72767 \\ 11,85602 \end{pmatrix}.$$

1.2. Запишемо функцію регресії з урахуванням знайдених оцінок коефіцієнтів моделі:

$$\hat{y}=26,10789-0,2518x_1-2,72767x_2+11,85602x_3. \qquad (2)$$

Модельні значення $\hat{y}_i$ зручно розташувати у тих же рядках, де записано початкові дані. Обчислити їх можна двома способами:

- за формулою $\hat{y}=\hat{a}_0+\hat{a}_1x_1+\hat{a}_2x_2+\hat{a}_3x_3$, яка набирається у першому рядку і продовжується в усіх інших. Значення параметрів $\hat{a}_0, \hat{a}_1, ..., \hat{a}_m$ у кожному рядку є незмінними, тому їх адреса фіксується (клавіша F4);
- за формулою $\hat{Y}=XA$, яка реалізується вбудованою функцією МУМНОЖ(X,A) (місце під її результат вказується перед викликом функції).

Отже, ми побудували лінійну модель (2) залежності прибутку від інвестицій, витрат на рекламу та заробітної плати. Наступним кроком наших досліджень є проведення дисперсійно-кореляційного аналізу та аналізу залишків.

2-й крок. Обчислення якісних характеристик моделі

2.1. Обчислимо залишки моделі $u_i=y_i-\hat{y}_i$, $i=1,2,...,15$, та їх квадрати аналогічно тому, як це виконувалось у першому завданні.

2.2. Обчислимо виправлену (незміщену) середньоквадратичну похибку дисперсії залишків:

$$\sigma_u = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m-1}} = \sqrt{\frac{\sum_{i=1}^{n} u_i^2}{n-m-1}}.$$

Маємо $\sigma_u = 5{,}7357$.

2.3. Обчислимо коефіцієнт детермінації за формулою

$$R^2 = 1 - \frac{\sum_{i=1}^{n} u_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

Отримаємо $R^2 = 0{,}91436$.

Висновок. *Оскільки коефіцієнт детермінації наближається до одиниці, то варіація залежної змінної Y значною мірою визначається варіацією незалежних змінних.*

Для знаходження параметрів регресії $\hat{a}_0, \hat{a}_1, ..., \hat{a}_m$ можна використати функцію ЛИНЕЙН, для якої вказати: 1) відомі значення Y; 2) відомі значення X (без стовпця одиниць); 3) константу "1"; 4) статистику "1". Результати цієї функції відображаються у таблиці:

$\hat{a}_m$	$\hat{a}_{m-1}$	...	$\hat{a}_2$	$\hat{a}_1$	$\hat{a}_0$
$S_{\hat{a}_m}$	$S_{\hat{a}_{m-1}}$	...	$S_{\hat{a}_2}$	$S_{\hat{a}_1}$	$S_{\hat{a}_0}$
R^2	σ_u				
$F_{\text{експ}}$	k				
$S_{\text{регр}}$	$\sum u_i^2$				

S_{a_i} ($i = \overline{0;m}$) — стандартні значення похибок для параметрів моделі $a_0, a_1, ..., a_m$;

$F_{\text{експ}}$ — спостережуване (експериментальне) значення F-статистики;

k — кількість ступенів вільності ($k = n–m–1$, де n — кількість спостережень; m— кількість факторних змінних моделі);

$S_{\text{регр}}$ — регресійна сума $\left(S_{\text{регр}} = \sum\left(\hat{y}_i - \overline{y}\right)^2\right)$.

Для заданих статистичних даних функція ЛИНЕЙН дає такі результати.

11,85602	-2,72767	-0,2518	26,10789
4,396525	2,486744	0,375368	8,407584
0,91436	5,735708	#Н/Д	#Н/Д
39,14827	11	#Н/Д	#Н/Д
3863,741	361,8818	#Н/Д	#Н/Д

Як бачимо, результати функції ЛИНЕЙН, повністю збігаються з результатами, одержаними на 2-му кроці.

3-й крок. Перевірка статистичних гіпотез

3.1. Перевіримо значущість вибіркового коефіцієнта кореляції. Обчислимо $R = \sqrt{R^2}$ — коефіцієнт кореляції (характеризує щільність лінійного зв'язку всіх незалежних змінних x_j (j = 1, 2, 3) із залежною змінною y):

$$R = 0{,}956222.$$

Коефіцієнт кореляції R, близький до одиниці, свідчить про те, що існує тісний лінійний зв'язок усіх незалежних змінних x_1, x_2, x_3 із залежною змінною y.

Однак потрібна ще перевірка його значущості, яка здійснюється за критерієм Стьюдента.

3.2. *Гіпотеза 1.* (H_0: $R = 0$).

Обчислимо t-статистику за формулою

$$t = \frac{R\sqrt{n-m-1}}{\sqrt{1-R^2}}; \qquad t = 37{,}03215.$$

Знайдемо $t_{табл}$ = $t(\alpha/2, n-m-1)$ — табличне значення t-розподілу з рівнем значущості $\alpha = 0{,}05$ і $(n-m-1) = 11$ ступенями вільності. Його можна визначити за таблицею розподілу Стьюдента або за допомогою статистичної функції СТЬЮДРАСПОБР(вероятность, степени свободы), де "вероятность" — 0,025 ($\alpha/2$), "степени свободы" — 11. Отримаємо:

$$t_{табл} = t\,(0{,}025;\,11) = \text{СТЬЮДРАСПОБР}\,(0{,}025;\,11) = 2{,}201.$$

Оскільки $|t| > t_{табл}$, то можна зробити висновок про достовірність коефіцієнта кореляції, який характеризує щільність зв'язку між залежною і незалежними змінними моделі.

3.3. Для вибраного рівня значущості α = 0,05 і ступеня вільності $k = n-m-1 = 11$ запишемо межі надійності для множинного коефіцієнта кореляції R:

$$(R–\Delta R;\ R+\Delta R),$$

де $\Delta R = t_{\text{табл}} \dfrac{1-R}{\sqrt{15}} =$ 2,201·(1–0,956222)/$\sqrt{15}$ = 0,029311; довірчий інтервал для множинного коефіцієнта кореляції R:

$$(R–\Delta R;\ R+\Delta R) = (0{,}926911;\ 0{,}985533).$$

3.4. Перевіримо значущість (адекватність) моделей загалом.

Гіпотеза 2. (H_0: $R^2 = 0$, що рівносильно $a_1 = a_2 = a_3 = 0$).

Обчислимо F-статистику за формулою

$$F_{\text{експ}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}; \qquad F = 39{,}14827.$$

Знайдемо табличне значення F-статистики:

$F_{\text{табл}} = F(\alpha, m, n-m-1)$, $F_{\text{табл}}$ = FРАСПОБР(0,05; 3; 11) = 3,59 (статистична функція FРАСПОБР(вероятность, степени свободы 1, степени свободы 2).

Порівняємо $F_{\text{табл}}$ з $F_{\text{експ}}$.

Оскільки $F_{\text{експ}} > F_{\text{табл}}$, то *нульова гіпотеза відхиляється, тобто коефіцієнти детермінації є значущими*. Це означає, що гіпотеза $a_1 = a_2 = a_3 = 0$ відхиляється, тобто показник y істотно залежить хоча б від одного із факторів x_1, x_2, x_3.

3.5. Перевіримо значущість кожного параметра моделі за критерієм Стьюдента.

Гіпотеза 3. (H_0: $a_0 = 0, a_1 = 0, a_2 = 0, a_3 = 0$).

Обчислимо експериментальне значення t-статистики для кожного параметра a_j, j = 0, 1, 2, 3:

$$t_j = \frac{a_j}{S_{a_j}},$$

де S_{a_0}, $S_{a_1}, S_{a_2}, S_{a_3}$ – стандартні похибки параметрів a_0, a_1, a_2, a_3, які обчислені на другому кроці. Їх можна взяти з другого рядка таблиці, яка є результатом функції ЛИНЕЙН.

Отже, маємо:

$$t_0 = 3{,}11,\ t_1 = -0{,}67,\ t_2 = -1{,}10,\ t_3 = 2{,}70.$$

Порівняємо абсолютні величини цих t-статистик з табличним значенням $t_{\text{табл}} = t(0{,}025;\ 11) = 2{,}201$.

Матимемо:

$|t_0| > t_{\text{табл}}$. Гіпотеза $a_0 = 0$ відхиляється. Цей параметр істотно відмінний від нуля, він вагомий.

$|t_1| < t_{табл}$. Гіпотеза $a_1 = 0$ приймається. Цей параметр випадково відмінний від нуля, він невагомий.

$|t_2| < t_{табл}$. Гіпотеза $a_2 = 0$ приймається. Цей параметр випадково відмінний від нуля, він невагомий.

$|t_3| > t_{табл}$. Гіпотеза $a_3 = 0$ відхиляється. Цей параметр істотно відмінний від нуля, він вагомий.

Внаслідок перевірки цієї гіпотези робимо такі висновки: фактори x_1, x_2 слабо впливають на показник y; фактор x_3 сильно впливає на показник y. Також є істотним постійний фактор, який визначається параметром моделі $a_0 = 0$.

4-й крок. Обчислення довірчих інтервалів регресії

Надійні зони регресії (довірчі інтервали для значень y_i) обчислимо за формулою

$$(y_i - \Delta y_i;\ y_i + \Delta y_i),$$

де $\Delta y_i = t_{табл}\, \sigma_u \sqrt{X_j (X^T X)^{-1} X_j^T}$; $t_{табл} = t(\alpha/2, n-m-1) = t(0{,}025{,}11)$ — табличне значення t-розподілу з $(n-m-1)$ =11 ступенями вільності і рівнем значущості $\alpha = 0{,}05$ ($t_{табл.}(0{,}025;\ 11) = 2{,}201$ — обчислене у попередньому пункті); σ_u — незміщена дисперсія залишків рівняння ($\sigma_u = 5{,}7357$ з п. 2.2); X_i — i-й рядок матриці спостережень X; X_i^T — i-й стовпець транспонованої матриці X^T.

Розрахунки потрібно виконати поетапно.

Обчислення під знаком кореня виконуються за допомогою математичної функції МУМНОЖ та функції СУММПРОИЗВ:

1) помножити матрицю X на обернену $\left(X^T X\right)^{-1}$ — отримати матрицю розмірності (15×4);

2) помножити кожний її рядок на відповідний рядок матриці X і отримати стовпець з 15 елементів.

Обчислити корені квадратні від кожного елемента останнього стовпця (математична функція КОРЕНЬ) і за обчисленими раніше $t_{табл}$ і σ_u визначити межі довірчих інтервалів для кожного модельного значення — два нових вектори-стовпці по 15 елементів.

5-й крок. Обчислення довірчих інтервалів параметрів

Довірчі інтервали для окремого параметра моделі a_j, $j = 0, 1, 2, ..., m$ обчислюються за формулою

$$(a_j - t_{\text{табл}}\sqrt{\sigma_u^2 c_{jj}};\ a_j + t_{\text{табл}}\sqrt{\sigma_u^2 c_{jj}}),$$

$$\text{або } (a_j - t_{\text{табл}}\,\sigma_u\sqrt{c_{jj}};\ a_j + t_{\text{табл}}\,\sigma_u\sqrt{c_{jj}}),$$

де $t_{\text{табл}}$, σ_u — визначені в попередніх пунктах; c_{jj} — діагональні елементи матриці $C = (X^T X)^{-1}$(для зручності можна переписати їх у вигляді окремого стовпця і розрахунки виконати за спільною формулою, зафіксувавши адреси $t_{\text{табл}}$ і σ_u). Як і в попередньому пункті буде утворено ще два нових стовпця, тепер по чотири елементи.

6-й крок. Прогнозування

Прогнозування за моделлю складається з двох варіантів: точкового та інтервального прогнозів.

6.1. Точковий прогноз.

За моделлю (2) обчислюють значення залежної змінної Y_p для заданих прогнозних значень $X1_p, X2_p, X3_p$.

6.2. Інтервальний прогноз.

Межі надійних інтервалів індивідуальних прогнозованих значень обчислюють за формулою

$$(y_{pi} - \Delta y_{pi};\ y_{pi} + \Delta y_{pi}),$$

де $\Delta y_{pi} = t_{\text{табл}}\,\sigma_u\sqrt{1 + X_p (X^T X)^{-1} X_p^T}$; $t_{\alpha/2,k}$, σ — визначені в попередніх пунктах; X_p — вектор-рядок незалежних змінних, що лежить за межами базового періоду; X_p^T — транспонований у стовпець рядок X_p, $\hat{Y}_p = X_p A$ — точкова оцінка математичного сподівання прогнозного значення Y_p, яка може розглядатися також як індивідуальне значення залежної змінної для відповідного вектора незалежних змінних.

6.3. Межі надійних інтервалів для математичного сподівання значення y_{pi} знаходять за формулою

$$(\hat{Y}_p - t_{\text{табл}}\,\sigma_u\sqrt{X_p (X^T X)^{-1} X_p^T} \le M(Y_p(X_p)) \le$$

$$\le \hat{Y}_p + t_{\alpha/2,k}\sigma\sqrt{X_p (X^T X)^{-1} X_p^T}).$$

Розрахунки під знаком кореня виконуються, як описано раніше, із застосуванням функцій МУМНОЖ і СУММПРОИЗВ, значення $t_{\alpha/2,k}$ і σ ті ж самі, що і раніше.

Лабораторна робота 3. Перевірка на наявність мультиколінеарності

1. Алгоритм Фаррара-Глобера тестування наявності мультиколінеарності.
2. Висновки стосовно методів оцінювання параметрів моделі.
3. Метод головних компонент.

Література [1–4]

Наявність мультиколінеарності в моделі свідчить про те, що фактори x_1, x_2, x_3 є залежними між собою. Це приводить до того, що не можна вказати, який же вплив кожного фактора на показник y. Перевірка здійснюється за алгоритмом Фаррара-Глобера. Наведемо покроковий алгоритм.

Алгоритм Фаррара-Глобера.

1-й крок

Нормалізувати змінні $x_1, x_2, ..., x_m$ моделі, для чого обчислити

$$x_{ij}^* = \frac{\left(x_{ij} - \overline{x}_j\right)}{\sqrt{n\sigma_{x_j}^2}}, \text{ або } x_{ij}^* = \frac{(x_{ij} - \overline{x}_j)}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}},$$

де n — кількість спостережень, ($i = 1, 2, ..., n$); m — кількість незалежних змінних ($j = 1,2, ... m$); $\overline{x}_j$ — середня арифметична j-ї незалежної змінної; σ_{xj}^2 — дисперсія j-ї незалежної змінної.

2-й крок

Побудувати нову матрицю X^*, елементами якої є нормалізовані незалежні змінні x_{ij}^*, і обчислити кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь):

$$R = X^{*tr}X^* = \begin{pmatrix} 1 & r_{12} & ... & r_{1m} \\ r_{21} & 1 & ... & r_{2m} \\ ... & ... & ... & ... \\ r_{m1} & r_{m2} & ... & 1 \end{pmatrix},$$

де X^{*tr} — транспонована матриця до матриці X^* (елементи матриці R характеризують тісноту зв'язку однієї незалежної змінної з іншою ($r_{ij} = r_{x_i x_j}$ — парні коефіцієнти кореляції).

3-й крок

3.1. Обчислити $|R|$ — визначник кореляційної матриці R (для цього застосувати функцію Excel МОПРЕД);

3.2. Визначити критерій χ^2, як

$$\chi^2 = -\left(n-1-\frac{1}{6}(2m+5)\right)\cdot \ln|R|.$$

3.3. Порівняти значення χ^2 з табличним при $\frac{1}{2}m(m-1)$ ступенях вільності і рівні значущості α (якщо $\chi^2 > \chi^2_{\text{табл}}$, де $\chi^2_{\text{табл}}$ = ХИ2ОБР (α, $m\ (m-1)/2$), то в масиві незалежних змінних існує мультиколінеарність).

4-й крок

Визначити матрицю C:

$$C = R^{-1} = (X^{*tr}X^{*})^{-1} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix}.$$

5-й крок

5.1. Розрахувати F-критерії:

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

де c_{kk} — діагональні елементи матриці C (C_{11}, C_{22}, ... C_{mm}).

5.2. Значення критеріїв F_k порівняти з табличним при $(n-m)$ і $(m-1)$ ступенях вільності і рівні значущості α (якщо $F_k > F_{\text{табл}}$, то відповідна k-та незалежна змінна мультиколінеарна з іншими).

5.3. Розрахувати коефіцієнти детермінації для кожної змінної:

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

6-й крок

Знайти часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що всі інші змінні x_{l1}, x_{l2}, ... , x_{lm} не впливають на цей зв'язок (існування парної мультиколінеарності):

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

де c_{kj} — елементи матриці C, що знаходиться в k-му рядку та j-му стовпці, $k = 1, 2, ..., m; j = 1, 2,..., m$, c_{kk} і c_{jj} — діагональні елементи матриці C.

Однак, якщо порівняти конкретні числові значення часткових і парних коефіцієнтів, то можна побачити, що часткові значно менші за парні. Тому на основі знання парних коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно ще виконати сьомий крок.

7-й крок

7.1. Розрахувати t-критерії:

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

7.2. Значення критеріїв t_{kj} порівняти з табличними при $(n-m)$ ступенях вільності і рівні значущості α; якщо $t_{kj} > t_{табл}$, де $t_{табл} = t(\alpha/2, n-m)$, то між незалежними змінними x_k і x_j існує мультиколінеарність.

Висновок

1. Між незалежними змінними може існувати лінійна залежність, але вона може не бути явищем мультиколінеарності змінних і тому не буде негативно впливати на кількісні параметри моделі, які розраховані за допомогою звичайного МНК.

2. Якщо $F_k > F_{табл.}$, то x_k залежить від усіх інших незалежних змінних і треба вирішити питання про її виключення з переліку змінних.

3. Якщо $t_{kj} > t_{табл.}$, то x_k і x_j тісно пов'язані між собою.

4. Аналізуючи F- і t-критерії, робимо висновок, яку із змінних можна виключити з багатофакторної моделі (зрозуміло, що при цьому потрібно виходити з економіко-логіко-теоретичних міркувань).

5. Якщо після вилучення певної змінної ми ще не позбавились мультиколінеарності, то оцінювання параметрів моделі слід здійснити за допомогою іншого методу, наприклад методу головних компонентів (або одного з його модифікацій).

Приклад. Дослідження наявності мультиколінеарності на основі алгоритму Фаррара-Глобера.

Розглянемо задачу дослідження впливу на економічний показник y (реальне споживання країни, млрд грн.) трьох факторів: x_1 (купівля та оплата товарів та послуг, млрд грн.), x_2 (всього заощаджень від загального грошового доходу, % від загальної суми доходу), x_3 (рівень ставки ПДВ, %). Необхідно перевірити фактори на мультиколінеарність.

№	$y(i)$	$x_1(i)$	$x_2(i)$	$x_3(i)$
1	25,74	4,69	11,97	29,23
2	25,34	5,64	13,43	29,35
3	31,26	6,26	12,92	33,40
4	33,50	6,99	14,74	30,97
5	32,30	6,36	14,64	32,92
6	38,90	7,60	17,10	37,27
7	41,58	7,12	15,63	30,97
8	48,02	6,81	15,35	33,58
9	43,30	8,67	15,85	35,62
10	51,78	7,83	18,05	34,99
11	52,14	7,84	17,24	39,34
12	54,94	8,85	20,52	41,50
13	59,18	9,61	19,18	45,58
14	62,22	10,67	19,03	41,08
15	63,62	11,04	21,45	40,54
16	65,01	11,85	22,25	42,75
17	67,78	12,94	24,75	43,89
18	71,45	14,24	25,03	41,95
19	75,24	15,67	27,87	44,06
20	77,38	16,33	30,48	46,77

Розв'язання

1-й крок

Нормалізуємо змінні x_1, x_2, x_3 економетричної моделі, для чого обчислимо:

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{n\sigma_{x_j}^2}}, \text{ або } x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}}$$

де $n = 20$ — кількість спостережень ($i = 1, 2, ..., 20$); $m = 3$ — кількість незалежних змінних ($j = 1, 2, 3$); $\bar{x}_j$ — середня арифметична j-ї незалежної змінної; σ_{xj}^2 — дисперсія j-ї незалежної змінної:

$$\sigma_{xj}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2.$$

Розрахунки за цією формулою проводяться поетапно:

- під кожним стовпцем незалежних змінних обчислити середнє значення:

$$\bar{x}_1 = 9{,}3505; \quad \bar{x}_2 = 18{,}874; \quad \bar{x}_3 = 37{,}788;$$

- від кожного елемента у стовпцях незалежних змінних відняти відповідне середнє значення, тобто обчислити $(x_{ij} - \bar{x}_j)$ (у результаті утвориться робоча матриця Р1 розмірності (15×3));
- отримані різниці піднести до квадрату, тобто обчислити $(x_{ij} - \bar{x}_j)^2$ (ще одна матриця Р2 такої ж розмірності, як і Р1);
- додати елементи у кожному стовпці матриці Р2, тобто отримати $n\sigma_{xj}^2 = \sum_{i=1}^{n}\left(x_{ij} - x_j\right)^2$ або застосувати до стовпців Р1 функцію СУММКВ;
- обчислити їх корені квадратні, тобто $\sqrt{n\sigma_{xj}^2}$ для $j = 1, 2, 3$;
- розділити елементи відповідних стовпців матриці Р1 на значення $\sqrt{n\sigma_{xj}^2}$ ($j = 1, 2, 3$).

У результаті отримаємо нову матрицю X^*, елементами якої є нормалізовані незалежні змінні x_{ij}^*:

$$X^* = \begin{pmatrix}
-0{,}309287 & -0{,}302374 & -0{,}339018 \\
-0{,}246242 & -0{,}238430 & -0{,}334264 \\
-0{,}205096 & -0{,}260766 & -0{,}173826 \\
-0{,}156651 & -0{,}181056 & -0{,}270089 \\
-0{,}198460 & -0{,}185436 & -0{,}192841 \\
-0{,}116169 & -0{,}077695 & -0{,}020520 \\
-0{,}148024 & -0{,}142077 & -0{,}270089 \\
-0{,}168596 & -0{,}154340 & -0{,}166696 \\
-0{,}045160 & -0{,}132441 & -0{,}085883 \\
-0{,}100905 & -0{,}036088 & -0{,}110840 \\
-0{,}100242 & -0{,}071564 & 0{,}061481 \\
-0{,}033215 & 0{,}072089 & 0{,}147047 \\
0{,}017221 & 0{,}013401 & 0{,}308673 \\
0{,}087566 & 0{,}006832 & 0{,}130409 \\
0{,}112121 & 0{,}112820 & 0{,}109018 \\
0{,}165875 & 0{,}147858 & 0{,}196565 \\
0{,}238212 & 0{,}257350 & 0{,}241725 \\
0{,}324485 & 0{,}269613 & 0{,}164874 \\
0{,}419385 & 0{,}393997 & 0{,}248460 \\
0{,}463185 & 0{,}508307 & 0{,}355814
\end{pmatrix}$$

2-й крок

Обчислимо кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь)

$$R = X^{*tr}X^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix},$$

де X^{*tr} — транспонована матриця X^*. Елементи матриці R характеризують щільність зв'язку однієї незалежної змінної з іншою, тобто $r_{ij} = r_{x_i x_j}$ — парні коефіцієнти кореляції. Застосувавши функцію МУМНОЖ, отримаємо

$$R = \begin{pmatrix} 1 & 0{,}92883 & 0{,}819534 \\ 0{,}92883 & 1 & 0{,}831277 \\ 0{,}81953 & 0{,}83127 & 1 \end{pmatrix}.$$

3-й крок

3.1. Обчислимо $|R|$ — визначник кореляційної матриці R за допомогою математичної функції МОПРЕД:

$$|R| = 0{,}00881.$$

3.2. Визначимо значення критерію χ^2, як

$$\chi^2 = -(n-1-\frac{1}{6}(2m+5))\ln|R|;\quad \chi^2 = 81{,}23005.$$

3.3. Порівняємо значення χ^2 з табличним $\chi^2_{\text{табл}} = \chi^2(\alpha; m(m-1)/2)$ при $\frac{1}{2}m(m-1) = 3$ ступенях вільності і рівні значущості $\alpha = 0{,}05$ (таблиця розподілу χ^2 або статистична функція ХИ2ОБР):

$$\chi^2_{\text{табл}} = \text{ХИ2ОБР}(0{,}05; 3) = 7{,}814724.$$

Оскільки $\chi^2 > \chi^2_{\text{табл}}$, то в масиві незалежних змінних існує мультиколінеарність в сукупності.

4-й крок

Визначимо матрицю C:

$$C = R^{-1} = (X^{*tr}X^*)^{-1} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix}.$$

Скориставшись математичною функцією МОБР, отримаємо:

$$C = \begin{pmatrix} 24{,}00377 & -22{,}82912 & -0{,}7311322 \\ -22{,}82912 & 26{,}20416 & -3{,}235457 \\ -0{,}731132 & -3{,}235457 & 4{,}5144749 \end{pmatrix}.$$

5-й крок

5.1. Розрахуємо *F*-критерії:

$$F_k = \frac{(c_{kk}-1)(n-m)}{(m-1)}, \quad k=1,\,2,\,3.$$

Винесемо в окремий стовпець c_{kk} — діагональні елементи матриці C, помножимо кожен з них на (15-3)/(3-1) = 6, отримаємо:

$$F_1 = 195{,}5320; \quad F_2 = 214{,}2354; \quad F_3 = 29{,}87303.$$

5.2. Значення критеріїв F_k порівняємо з табличним при $(n-m) = 17$ і $(m-1) = 2$ ступенях вільності і рівні значущості α=0,05 (таблиця *F*-розподілу або статистична функція FРАСПОБР):

$$F_{\text{табл}} = \text{FРАСПОБР}(0{,}05;\ 17;\ 2) = 19{,}43703.$$

Оскільки $F_1 > F_{\text{табл}}$, $F_2 > F_{\text{табл}}$, $F_3 > F_{\text{табл}}$, то робимо висновок, що перша, друга і третя незалежні змінні мультиколінеарні з іншими.

5.3. Розрахуємо коефіцієнти детермінації для кожної змінної:

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

(При розрахунках можна використати виписані раніше значення c_{kk}.)

$$R_1^2 = 0{,}958339; \quad R_2^2 = 0{,}961838; \quad R_3^2 = 0{,}778490.$$

6-й крок

Знайдемо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що всі інші змінні $x_{l1}, x_{l2}, \ldots, x_{lm}$ не впливають на цей зв'язок (існування парної мультиколінеарності).

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

де c_{kj} — елемент матриці C, що знаходиться в k-му рядку та j-му стовпці, коли $k \neq j$, тобто елементи c_{12}, c_{13}, c_{23}; c_{kk} і c_{jj} — діагональні елементи матриці C (c_{11}, c_{22}, c_{33} у відповідних комбінаціях).

$$r_{12} = 0{,}910257; \qquad r_{13} = 0{,}070234; \qquad r_{23} = 0{,}297472.$$

Однак, якщо порівняти абсолютні значення часткових і парних коефіцієнтів, то можна побачити, що перші значно менші за останні. Тому на основі знання парних коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно ще виконати сьомий крок.

7-й крок.

7.1. Розрахуємо t-критерії:

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

$t_{12} = 9{,}064506$; $t_{13} = 0{,}290302$; $t_{23} = 1{,}284666$.

7.2. Значення критеріїв t_{kj} порівняємо з табличними при $(n-m) = 17$ ступенях вільності і рівні значущості $\alpha = 0{,}05$ (таблиця розподілу Стьюдента або статистична функція СТЬЮДРАСПОБР): $t_{табл} = 2{,}109818$.

Оскільки $t_{12} > t_{табл}$, $t_{13} < t_{табл}$, $t_{23} < t_{табл}$, то між першою і другою незалежними змінними існує мультиколінеарність.

Якщо F-критерій більше табличного значення, а це значить, що k-та змінна залежить від всіх інших в масиві, то необхідно вирішувати питання про її виключення з переліку незалежних змінних моделі.

Якщо t_{kj}-критерій більше табличного, то ця пара змінних (x_k і x_j) тісно взаємопов'язана. Звідси, аналізуючи рівень обох критеріїв F і t, можна зробити обґрунтований висновок про те, яку із змінних необхідно виключити із дослідження чи замінити іншою. Але заміна масиву незалежних змінних завжди повинна узгоджуватись з економічною доцільністю, що випливає з мети дослідження.

СПИСОК ЛІТЕРАТУРИ

Основна

1. *Грубер Й*. Економетрія: Вступ до множинної регресії та економетрії: У 2 т. — К.: Нічлава, 1998. — Т. 1. — 384 с.; 1999. — Т. 2. — 308 с.
2. *Доугерти К*. Введение в эконометрику. — М.: ИНФРА-М, 1997. — 402 с.
3. *Економетрія*: Навч. посіб. для студ. вищ. навч. закл. / О. Л. Лещинський, В. В. Рязанцева, О. О. Юнькова. — К.: МАУП, 2003. — 208 с.
4. *Корольов О. А*. Економетрія: Навч. посіб. — К.: КНТЕУ, 2000. — 660 с.
5. *Лук'яненко І. Г., Краснікова Л. І.* Економетрика: Підручник. — К.: Знання; КОО, 1998. — 494 с.
6. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс. — М.: Дело, 1998. — 248 с.
7. *Наконечний С. І., Терещенко Т. О., Романюк Т. П.* Економетрія: Підручник. — 2-ге вид., допов. та перероб. — К.: КНЕУ, 2000. — 296 с.

Додаткова

1. *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики: Учеб. для вузов. — М.: ЮНИТИ, 1998. — 1022 с.
2. *Джонстон Дж.* Эконометрические методы. — М.: Статистика, 1980. — 444 с.
3. *Дрейпер С.* Прикладной регрессионный анализ. — М.: Мир, 1988. — Т. 1–2.
4. *Катышев П. К., Пересецкий А. А.* Сборник задач к начальному курсу эконометрики. — М.: Дело, 1999. — 72 с.
5. *Маленво Э.* Статистические методы в эконометрии. — М.: Статистика, 1975. — Вып. 1. — 423 с.; 1976. — Вып. 2. — 325 с.
6. *Винн Р., Холден К.* Введение в прикладной эконометрический анализ. — М.: Финансы и статистика, 1981. — 294 с.
7. *Клас А., Гергели К., Колек Ю., Шуян И.* Введение в эконометрическое моделирование. — М.: Статистика, 1978. — 152 с.
8. *Тинтнер Г.* Введение в эконометрию. — М.: Статистика, 1965. — 361 с.

9. *Клейнер Б. Г.* Производственные функции. — М.: Финансы и статистика, 1995.
10. *Алмон Д.* Система функций потребления и ее оценка для Бельгии // Экономика и матем. методы. — 1978. — XIV, вып. 3. — С. 480-502.
11. *Рабочая* книга по прогнозированию / Отв. ред. И. В. Бестужев-Лада. — М.: Мысль, 1982. — 430 с.
12. *Хазанова Л. Э.* Математическое моделирование в экономике: Учеб. пособие. — М.: БЕК, 1998. — 141 с.
13. *Дубров А. М., Лагоша Б. А., Хрусталев Е. Ю.* Моделирование рисковых ситуаций в экономике и бизнесе: Учеб. пособие. — М.: Финансы и статистика, 1999. — 176 с.
14. *Жданов С. А.* Экономические модели и методы в управлении. — М.: Дело и сервис, 1998. — 176 с.
15. *Єлейко В. М.* Основи економетрії. — Львів.: Марка Лтд, 1995. — 191 с.
16. *Емельянов А. С.* Эконометрия и прогнозирование. — М.: Экономика, 1985. — 117 с.

ЗМІСТ

Відповідальний за випуск *А. Д. Вегеренко*
Редактор *С. Г. Рогузько*
Комп'ютерне верстання *О. Л. Тищенко*

Зам. № ВКЦ-3516
Підп. до друку 01.06.09. Формат $60\times84/_{16}$. Папір офсетний.
Друк ротаційний трафаретний.
Ум. друк. арк. 1,74. Обл.-вид. арк. 1,6. Наклад 50 пр.

Міжрегіональна Академія управління персоналом (МАУП)
03039 Київ-39, вул. Фрометівська, 2, МАУП

ДП "Видавничий дім "Персонал"
03039 Київ-39, просп. Червонозоряний, 119, літ. XX

Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК № 3262 від 26.08.2008