МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ МІСЬКОГО ГОСПОДАРСТВА ІМЕНІ О. М. БЕКЕТОВА

МЕТОДИЧНІ ВКАЗІВКИ
ДО ПРАКТИЧНИХ ЗАНЯТЬ І САМОСТІЙНОЇ РОБОТИ З КУРСУ

ЕКОНОМЕТРІЯ

(для студентів галузі знань 0306 «Менеджмент і адміністрування» напряму 6.030601 «Менеджмент» заочної форми навчання ЦПО та ЗН)

ХАРКІВ
ХНУМГ
2013

Методичні вказівки до практичних занять і самостійної роботи з курсу «Економетрія» (для студентів галузі знань 0306 «Менеджмент і адміністрування» напряму 6.030601 «Менеджмент» заочної форми навчання ЦПО та ЗН) / Харк. нац. ун-т міськ. госп-ва ім. О. М. Бекетова; уклад.: О. О. Воронков. - Х.: ХНУМГ, 2013. – 40 с.

Укладач: О. О. Воронков

Рекомендовано кафедрою економіки підприємств міського господарства, протокол № 12 від 07.06.12 р.

ЗАГАЛЬНІ ПОЛОЖЕННЯ

Методичні вказівки спрямовані на допомогу студентам оволодіти практичними навичками з побудови економетричних моделей реальних економічних явищ з метою прогнозування за різними економічними показниками.

Дисципліна «Економетрія» є нормативною дисципліною циклу природничо-наукової та загальноекономічної підготовки бакалаврів за напрямом 6.030601 «Менеджмент і адміністрування». Відповідно до навчального плану її вивчають у 6 семестрі 3 курсу. Обсяг практичних занять становить 4 аудиторних години (2 практичних заняття). Обсяг самостійної роботи становить 64 години. Вивчення дисципліни «Економетрія» спрямовано на підготовку висококваліфікованих фахівців, що володіють теоретичними знаннями та практичними навичками з питань визначення кількісних та якісних взаємозв'язків економічних об'єктів і процесів за допомогою математичних і статистичних методів та моделей.

Відповідно до робочої програми курсу «Економетрія» у методичних вказівках до практичних занять розглянуто найважливіші теми змістового модуля 1 «Економетрічне моделювання як метод наукового пізнання. Методи побудови загальної лінійної моделі», зокрема побудова лінійної моделі парної регресії та множинної регресії за методом найменших квадратів. Знання й навички, що отримані під час вивчення цих тем, є основою для вивчення наступних складніших тем курсу, та найчастіше застосовуються у практичній діяльності.

У методичних вказівках до самостійної роботи для кожної теми зазначено обсяг витрат часу на вивчення, що відповідає програмі курсу. Наприкінці методичних вказівок наведено список основних і додаткових підручників, які рекомендується використовувати. Кожна тема супроводжується посиланнями на відповідні їй сторінки підручників. Після вивчення теоретичного матеріалу треба дати відповіді на запитання для самоперевірки з теми, а також вирішити задачі, пропоновані для самостійного розв'язання. Для полегшення роботи перед задачами для самостійного розв'язання наведене розв'язання аналогічних прикладів.

МЕТОДИЧНІ ВКАЗІВКИ ДО ПРАКТИЧНИХ ЗАНЯТЬ

Практичне заняття 1
ПОБУДОВА ЗАГАЛЬНОЇ ЛІНІЙНОЇ МОДЕЛІ

Мета – оволодіння основними етапами економетричного моделювання та визначення параметрів моделі за методом найменших квадратів.

Вказівки до виконання завдання

Щоб знайти пояснену частину, тобто величину $M_x(Y)$, якщо визначено характер експериментальних даних та виділено певний набір пояснюючих змінних, потрібно знати умовні розподіли випадкової величини Y. На практиці це майже ніколи не має місця, тому точне визначення поясненої частини є неможливим. У таких випадках застосовують стандартну процедуру згладжування експериментальних даних. Ця процедура полягає у двох етапах:

- визначають параметричне сімейство, до якого належить шукана функція $M_x(Y)$ (її розглядають як залежність від значень пояснюючих змінних X). Це сімейство може бути множиною лінійних функцій, степеневих функцій та ін.;
- визначають оцінки параметрів цієї функції за допомогою одного з методів математичної статистики.

У переважній більшості випадків економетричні моделі обирають лінійними. Окрім відносної простоти лінійної моделі, для такого вибору є дві істотні причини. Перша причина: якщо випадкова величина (X, Y) має спільний нормальний розподіл, тоді, як відомо, рівняння регресії є лінійними. Припущення про нормальний розподіл можна обґрунтувати за допомогою граничних теорем теорії ймовірностей. Друга причина, за якою лінійна регресійна модель є переважнішою за інші, - це менший ризик істотної помилки прогнозу.

Лінійна регресія збігається до знаходження рівняння виду

$$\hat{y}_x = a + bx \text{ або } y = a + bx + \varepsilon .$$

Рівняння виду $\hat{y}_x = a + bx$ дозволяє за заданим значенням фактору x знаходити теоретичні значення результативної ознаки, підставляючи до нього фактичні значення фактору x.

Побудова лінійної регресії збігається до оцінки її параметрів – a та b. Класичний підхід до оцінювання цих параметрів лінійної регресії ґрунтується на методі найменших квадратів (МНК). МНК відповідно до теореми Гауса-Маркова дає найкращі оцінки цих параметрів.

МНК дозволяє дістати такі оцінки параметрів a та b, за яких сума квадратів відхилень фактичних значень результативної ознаки y від теоретичних $\hat{y}_x$ є мінімальною:

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_{xi}\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2 \to \min .$$

Тобто з усієї множини ліній лінія регресії на графіку обирається так, щоб сума квадратів відстаней між статистичними точками та цією лінією була б мінімальною.

Задача 1.1

За даними проведеного опитування восьми груп сімей відомі дані зв'язку витрат населення на продукти харчування з рівнем доходів сім'ї. Дані наведені в табл. 1.1.

Таблиця 1.1 – Вихідні дані

Витрати на продукти харчування, y, тис. грн.	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Доходи сім'ї, x, тис. грн.	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

Припустимо, що зв'язок між доходами сім'ї та витратами на продукти харчування є лінійним. Для підтвердження цього припущення побудуємо поле кореляції, наведене на рис. 1.1.

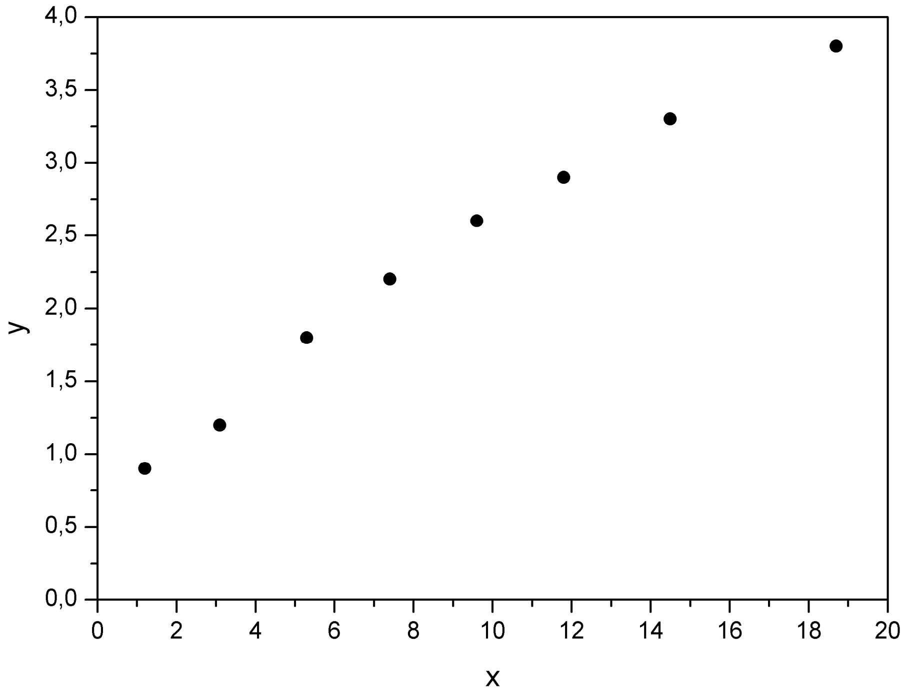

Рис. 1.1 – Поле кореляції

З графіка видно, що точки вибудовуються у певну пряму лінію.

Для зручності подальших обчислень складемо таблицю 1.2.

Розрахуємо параметри лінійного рівняння парної регресії $\hat{y}_x = a + b \cdot x$. Для цього скористаємось формулами:

$$b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{26{,}09 - 8{,}95 \cdot 2{,}34}{30{,}56} = 0{,}168;$$

$$a = \overline{y} - b \cdot \overline{x} = 2{,}34 - 0{,}168 \cdot 8{,}95 = 0{,}836.$$

Таблиця 1.2 – Проміжкові розрахунки

	x	y	$x \cdot y$	x^2	y^2	$\hat{y}_x$	$y - \hat{y}_x$	$\left(y - \hat{y}_x\right)^2$	A_i, %
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1,2	0,9	1,08	1,44	0,81	1,038	–0,138	0,0190	15,33
2	3,1	1,2	3,72	9,61	1,44	1,357	–0,157	0,0246	13,08
3	5,3	1,8	9,54	28,09	3,24	1,726	0,074	0,0055	4,11
4	7,4	2,2	16,28	54,76	4,84	2,079	0,121	0,0146	5,50
5	9,6	2,6	24,96	92,16	6,76	2,449	0,151	0,0228	5,81
6	11,8	2,9	34,22	139,24	8,41	2,818	0,082	0,0067	2,83
7	14,5	3,3	47,85	210,25	10,89	3,272	0,028	0,0008	0,85
8	18,7	3,8	71,06	349,69	14,44	3,978	–0,178	0,0317	4,68
Разом	71,6	18,7	208,71	885,24	50,83	18,717	–0,017	0,1257	52,19
Середнє значення	8,95	2,34	26,09	110,66	6,35	2,34	–	0,0157	6,52
σ	5,53	0,935	–	–	–	–	–	–	–
σ^2	30,56	0,874	–	–	–	–	–	–	–

Дістали рівняння: $\hat{y}_x = 0,836 + 0,168 \cdot x$. Тобто із збільшенням доходу сім'ї на 1000 грн. витрати на харчування збільшуються на 168 грн.

Як було зазначено вище, рівняння лінійної регресії завжди доповнюється показником тісноти зв'язку – лінійним коефіцієнтом кореляції r_{xy}:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,168 \cdot \frac{5,53}{0,935} = 0,994 \cdot$$

Наближеність коефіцієнта кореляції до 1 вказує на тісний лінійний зв'язок між ознаками.

Коефіцієнт детермінації $r_{xy}^2 = 0,987$ показує, що рівняння регресії пояснює 98,7% дисперсії результативної ознаки, а на долю інших факторів припадає лише 1,3%.

Оцінимо якість рівняння регресії в цілому за допомогою F- критерію Фішера. Розрахуємо фактичне значення F - критерію:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot \left(n - 2\right) = \frac{0,987}{1 - 0,987} \cdot 6 = 455,54 \cdot$$

Табличне значення ($k_1 = 1,\ \ k_2 = n - 2 = 6,\ \ \alpha = 0,05$): $F_{\text{табл}} = 5,99$. Оскільки $F_{\text{факт}} > F_{\text{табл}}$, визнаємо статистичну значущість рівняння в цілому.

Для оцінки статистичної значущості коефіцієнтів регресії та кореляції розрахуємо t-критерій Стьюдента та довірчі інтервали кожного з показників. Розрахуємо випадкові помилки параметрів лінійної регресії та коефіцієнту кореляції $\left(S_{\text{ост}}^2 = \frac{\sum\left(y - \hat{y}_x\right)^2}{n-2} = \frac{0,1257}{8-2} = 0,021 \right)$:

$$m_b = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}} = \frac{\sqrt{0,021}}{5,53 \cdot \sqrt{8}} = 0,0093,$$

$$m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n} = \frac{\sqrt{0,021 \cdot 885,24}}{5,53 \cdot 8} = 0,0975, \qquad m_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0,987}{6}} = 0,0465.$$

Фактичні значення t- статистик: $t_b = \frac{0,168}{0,0093} = 18,065$, $t_a = \frac{0,836}{0,0975} = 8,574$, $t_r = \frac{0,994}{0,0465} = 21,376$. Табличне значення t- критерію Стьюдента при $\alpha = 0,05$ та кількості ступенів свободи $\nu = n - 2 = 6$ дорівнює $t_{\text{табл}} = 2,447$. Оскільки $t_b > t_{\text{табл}}$, $t_a > t_{\text{табл}}$ та $t_r > t_{\text{табл}}$, визнаємо статистичну значущість параметрів регресії та показника тісноти зв'язку. Розрахуємо довірчі інтервали для параметрів регресії a та b: $a \pm t \cdot m_a$; $b \pm t \cdot m_b$. Отримаємо, що $a \in [0,597;\ \ 1,075]$; $b \in [0,145;\ \ 0,191]$.

Середня помилка апроксимації (знаходимо за допомогою стовпця 10 таблиці 1.2 за формулою $A_i = \left| \frac{y_i - \hat{y}_{x_i}}{y_i} \right| \cdot 100\%$) $\bar{A} = 6,52\%$ говорить про гарну якість рівняння регресії, тобто свідчить про гарний підбор моделі до вихідних даних.

Визначимо прогнозне значення результативного фактору $\hat{y}_p$ при значенні ознаки-фактору, що складає 110% від середнього рівня $x_p = 1,1 \cdot \bar{x} = 1,1 \cdot 8,95 = 9,845$, тобто визначимо витрати на харчування, якщо доходи сім'ї складуть 9,85 тис. грн.

$$\hat{y}_p = 0,836 + 0,168 \cdot 9,845 = 2,490 \text{ (тис. грн.)}$$

Отже, якщо доходи сім'ї складуть 9,845 тис. грн., витрати на харчування складатимуть 2,490 тис. грн.

Визначимо довірчий інтервал прогнозу. Помилка прогнозу

$$m_{\hat{y}_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{n \cdot \sigma_x^2}} = \sqrt{0,021 \cdot \left(1 + \frac{1}{8} + \frac{(9,845 - 8,95)^2}{8 \cdot 30,56}\right)} = 0,154,$$

а довірчий інтервал ($\hat{y}_p - \Delta_{\hat{y}_p} \le \hat{y}_p \le \hat{y}_p + \Delta_{\hat{y}_p}$):

$$2,113 < \hat{y}_p < 2,867.$$

Отже, прогноз є статистично надійним.

На одному графіку побудуємо вихідні дані та лінію регресії (рис. 1.2):

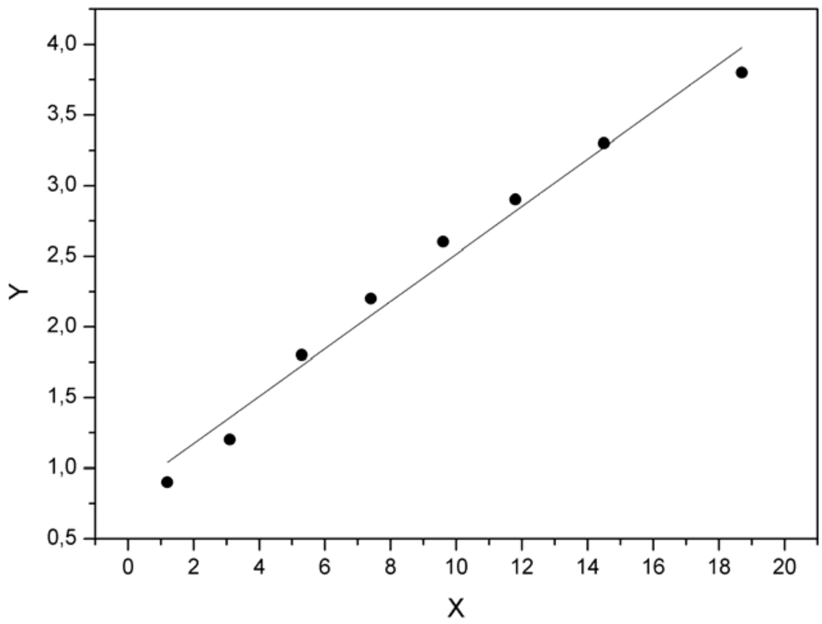

Рис. 1.2 – Лінія регресії

Задача 1.2

За територіями регіону наведені дані у в таблиці 1.3.

Таблиця 1.3 – Вихідні дані

Номер регіону	**Середньодушовій прожитковий мінімум на день одного працездатного, грн. (х)**	**Середньоденна заробітна плата, грн. (у)**
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Потрібно побудувати лінійне рівняння парної регресії у від х. Розрахувати лінійний коефіцієнт парної кореляції та середню помилку апроксимації. Оцінити статистичну значущість параметрів регресії та кореляції. Виконати прогноз заробітної платні при прогнозному значенні среднедушового прожиткового мінімуму х, що становить 107% від середнього рівня. Оцінити точність прогнозу, розрахувавши помилку прогнозу та його довірчий інтервал.

Розв'язання

1. Для розрахунку параметрів рівняння лінійної регресії будуємо розрахункову таблицю (табл. 1.4).

Таблиця 1.4 – Проміжкові розрахунки

	x	y	xy	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	A^i
1	78	133	10374	6084	17689	149	-16	12,0
2	82	148	12136	6724	21904	152	-4	2,7
3	87	134	11658	7569	17956	157	-23	17,2
4	79	154	12166	6241	23716	150	4	2,6
5	89	162	14418	7921	26244	159	3	1,9
6	106	195	20670	11236	38025	174	21	10,8
7	67	139	9313	4489	19321	139	0	0,0
8	88	158	13904	7744	24964	158	0	0,0
9	73	152	110196	5329	23104	144	8	5,3
10	87	162	14094	7569	26244	157	5	3,1
11	76	159	12084	5776	25281	147	12	7,5
12	115	173	19895	13225	29929	183	-10	5,8
Сума	1027	1869	161808	89907	294377	1869	0	68,8
Середнє	85,6	155,8	13484,0	7492,3	24531,4			5,7
σ	12,95	16,53						
σ^2	167,7	273,4						

$$b=\frac{\overline{yx}-\overline{y}\,\overline{x}}{\sum x^2-(\overline{x})^2}=\frac{13484-85{,}6\cdot155{,}8}{7492{,}3-85{,}6^2}=\frac{151{,}8}{164{,}94}=0{,}92 \qquad a=\overline{y}-b\overline{x}=155{,}8-0{,}92\cdot85{,}6=77{,}0$$

Одержано рівняння регресії: $y=77{,}0+0{,}92x$.

Із збільшенням середньодушового прожиткового мінімуму на 1 грн. середньоденна заробітна платня зростає в середньому на 0,92 грн.

2. Тісноту лінійного зв'язку оцінимо за допомогою коефіцієнту кореляції:

$$r_{xy}=b\frac{\sigma_x}{\sigma_y}=0{,}92\cdot\frac{12{,}95}{16{,}53}=0{,}721; \qquad r_{xy}^2=0{,}52.$$

Це означає, що 52% варіації заробітної платні (у) пояснюється варіацією фактору х - середньодушового прожиткового мінімуму. Якість моделі визначає середня помилка апроксимації:

$$\overline{A}=\frac{1}{n}\sum A_i=\frac{68{,}9}{12}=5{,}7\%.$$

Якість побудованої моделі оцінюється як добра, оскільки А не перевершує 8 - 10%.

3. Оцінку статистичної значущості параметрів регресії проведемо за допомогою t-статистики Стьюдента і шляхом розрахунку довірчого інтервалу кожного з показників.

Висуваємо гіпотезу H_0 про статистично незначущу відмінність показників від нуля: $a=b=r_{xy}=0$.

$t_{табл}$ для кількості ступенів свободи df= n-2 = 12-2 = 10 і α= 0,05 складе 2,23.

Визначимо випадкові помилки m_a, m_b, m_r:

$$m_a=12{,}6\frac{\sqrt{89907}}{12\cdot12{,}95}=24{,}3; \qquad m_b=\frac{126}{12{,}95*\sqrt{12}}=0{,}281; \qquad m_r=\sqrt{\frac{1-0{,}52}{12-2}}=0{,}219.$$

Тоді

$$t_a = \frac{77}{24,3} = 3,2; \quad t_b = \frac{0,92}{0,281} = 3,3; \quad t_r = \frac{0,721}{0,219} = 3,3.$$

Фактичні значення t-статистики перевершують табличні значення:

$$t_a = 3,2 > t_{табл} = 2,3; \; t_b = 3,3 > t_{табл} = 2,3; \; t_r = 3,3 > t_{табл} = 2,3,$$

тому гіпотеза H_0 відхиляється, тобто a, b і r_{xy} не випадково відрізняються від нуля, а є статистично значущими.

Розрахуємо довірчий інтервал для a і b. Для цього визначимо граничну помилку для кожного показника:

$$\Delta a = 2,23 * 24,3 = 54; \quad \Delta b = 2,23 * 0,281 = 0,62.$$

Довірчі інтервали:

$$\gamma_a = a \pm \Delta a = 77 \pm 54; \; \gamma_{amin} = 77 - 54 = 23; \; \gamma_{amax} = 77 + 54 = 131;$$

$$\gamma_b = b \pm \Delta b = 0,92 \pm 0,62; \; \gamma_{bmin} = 0,92 - 0,62 = 0,30; \; \gamma_{bmax} = 0,92 + 0,62 = 1,54.$$

Аналіз верхньої і нижньої меж довірчих інтервалів призводить до висновку про те, що з імовірністю $p = 1 - \alpha = 0,95$ параметри a і b, знаходячись у вказаних межах, не приймають нульових значень, тобто не є статистично незначущими та істотно відмінні від нуля.

4. Одержані оцінки рівняння регресії дозволяють використовувати його для прогнозу. Якщо прогнозне значення прожиткового мінімуму складе:

$$x_p = \overline{x} \cdot 1,07 = 85,6 \cdot 1,07 = 91,6 \text{ тис. грн.},$$

тоді прогнозне значення прожиткового мінімуму складе:

$$\hat{y}_p = 77 + 0,92 \cdot 91,6 = 161 \text{ тис. грн.}$$

5. Помилка прогнозу складе:

$$m_{\hat{y}_p} = 12,6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91,6 - 85,6)^2}{12 \cdot 12,95^2}} = 13,2 \text{ тис. грн.}$$

Гранична помилка прогнозу, яка у 95% випадків не буде перевищеною, складе:

$$\Delta_{\hat{y}_p} = t_{табл} \cdot m_{\hat{y}_p} = 2,23 \cdot 13,2 = 29,4.$$

Довірчий інтервал прогнозу:

$$\gamma_{\hat{y}_p} = 91,6 \pm 29,4; \; \gamma_{\hat{y}_p \min} = 91,6 - 29,4 = 62,2 \text{ грн.}; \; \gamma_{\hat{y}_p \max} = 91,6 + 29,4 = 121 \text{ грн.}$$

Виконаний прогноз середньомісячної заробітної платні виявився надійним ($p = 1 - \alpha = 1 - 0,05 = 0,95$), але неточним, оскільки діапазон верхньої та нижньої меж довірчого інтервалу складає 1,95 разу.

Практичне заняття 2
МНОЖИННА РЕГРЕСІЯ. УЗАГАЛЬНЕНИЙ МЕТОД НАЙМЕНШИХ КВАДРАТІВ (УМНК)

Мета – оволодіння основними етапами економетричного моделювання та визначення параметрів моделі при порушенні властивості гомоскедастичності.

Вказівки до виконання завдання

Множинна регресія - рівняння зв'язку з кількома незалежними змінними:

$$\hat{y} = f(x_1, x_2, ..., x_m),$$

де y - залежна змінна (результативна ознака); x_1, x_2, x_m - незалежні змінні (фактори).

Для оцінки параметрів рівняння множинної регресії застосовують метод найменших квадратів (МНК). Для лінійних рівнянь будують наступну систему нормальних рівнянь, розв'язання якої дозволяє одержати оцінки параметрів регресії:

$$\begin{cases} \sum y = na + b_1 * \sum x_1 + b_2 \sum x_2 + ... + b_m \sum x_m \\ \sum yx_1 = a\sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + ... + b_m \sum x_1 x_m \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum yx_m = a\sum x_m + b_1 \sum x_1 x_m + b_2 \sum x_2 x_m + ... + b_m \sum x_m^2 \end{cases}.$$

Для її розв'язання можна застосувати метод визначників:

$$a = \frac{\Delta a}{\Delta};\ b_1 = \frac{\Delta b_1}{\Delta};\ b_m = \frac{\Delta b_m}{\Delta},$$

де $\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & \cdots & \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \cdots & \sum x_m x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \cdots & \sum x_m x_2 \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_m & \sum x_1 x_m & \sum x_2 x_m & \cdots & \sum x_m^2 \end{vmatrix}$ - визначник системи;

Δa, Δb_1, …, Δb_m - часткові визначники, які виходять шляхом заміни відповідного стовпця матриці визначника системи даними лівої частини системи.

Інший вид рівняння множинної регресії - рівняння регресії в стандартизованому масштабі:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \cdots + \beta_m t_{x_m},$$

де $t_y = \frac{y - \bar{y}}{\sigma_y}$; $t_{x_1} = \frac{x_1 - \overline{x_1}}{\sigma_{x_1}}$ - стандартизовані змінні; β_i - стандартизовані коефіцієнти регресії.

До рівняння множинної регресії в стандартизованому масштабі застосуємо МНК. Стандартизовані коефіцієнти регресії (β–коефіцієнти) визначають з наступної системи рівнянь:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_2 x_1} + \beta_3 r_{x_3 x_1} + \cdots + \beta_m r_{x_m x_1} \\ r_{yx_2} = \beta_1 r_{x_2 x_1} + \beta_2 + \beta_3 r_{x_3 x_2} + \cdots \beta_m r_{x_m x_2} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ r_{yx_m} = \beta_1 r_{x_m x_1} + \beta_2 r_{x_m x_2} + \beta_3 r_{x_3 x_m} + \cdots + \beta_m \end{cases}$$

Зв'язок коефіцієнтів множинної регресії b_i із стандартизованими коефіцієнтами b описується співвідношенням

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}.$$

Параметр a визначається як $a = \overline{y} - b_1\overline{x_1} - b_2\overline{x_2} - \ldots - b_m\overline{x_m}$.

Середні коефіцієнти еластичності для лінійної регресії розраховують за формулою

$$\overline{E}_{yx_j} = b_j \frac{\overline{x}_j}{\overline{y}}.$$

Для розрахунку часткових коефіцієнтів еластичності застосовують наступну формулу:

$$E_{yx_1} = b_1 \frac{x_1}{\hat{y}_{x_1 x_1, x_2, , x_{l-1}, x_{l+1}, , x_m}}.$$

Тісноту сумісного впливу чинників на результат оцінює індекс множинної кореляції:

$$R_{yx_1x_2,,x_m} = \sqrt{1 - \frac{\sigma^2_{y\,\text{зал}}}{\sigma^2_y}}.$$

Значення індексу множинної кореляції лежить у межах від 0 до 1 і повинне перевищувати або дорівнювати максимальному парному індексу кореляції:

$$R_{yx_1x_2,,x_m} \geq r_{yx_i}, \quad (i = \overline{1,m}).$$

Індекс множинної кореляції для рівняння в стандартизованому масштабі можна записати у вигляді

$$R_{yx_1x_2,,x_m} = \sqrt{\sum \beta_i r_{yx_i}}.$$

Якщо залежність лінійна, коефіцієнт множинної кореляції можна визначити через матрицю парних коефіцієнтів кореляції:

$$R_{yx_1x_2,,x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

де

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \cdots r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \cdots r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \cdots r_{x_2x_m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{yx_m} & r_{x_mx_1} & r_{x_mx_2} & \cdots 1 \end{vmatrix}$$ - визначник матриці парних коефіцієнтів кореляції;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \cdots r_{x_1x_m} \\ r_{x_2x_1} & 1 & \cdots r_{x_2x_m} \\ \cdots & \cdots & \cdots \\ r_{x_mx_1} & r_{x_mx_2} & \cdots 1 \end{vmatrix}$$ - визначник матриці кореляції, міжфакторної кореляції.

Часткові коефіцієнти (або індекси) кореляції, що вимірюють вплив на у фактора x, при незмінному рівні інших факторів, можна визначити за формулою

$$r_{yx_l x_1 x_2,\dots,x_{l-1}x_{l+1},\dots,x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \dots x_l \dots x_m}}{1 - R^2_{yx_1 x_2 \dots x_{l-1} x_{l+1} \dots x_m}}}\ .$$

Часткові коефіцієнти кореляції змінюються в межах від -1 до +1.

Якість побудованої моделі в цілому оцінює коефіцієнт (індекс) детермінації. Коефіцієнт множинної детермінації розраховується як квадрат індексу множинної кореляції:

$$R^2_{yx_1 x_2 \dots x_m}\ .$$

Скорегований індекс множинної детермінації містить поправку на кількість ступенів свободи і розраховується за формулою

$$\hat{R}^2 = 1 - (1 - R^2)\frac{(n-1)}{n-m-1},$$

де n - кількість спостережень, m - кількість факторів.

Значущість рівняння множинної регресії в цілому оцінюється за допомогою F-критерію Фішера:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}.$$

Частковий F-критерій оцінює статистичну значущість присутності кожного з факторів у рівнянні. У загальному вигляді для чинника x_l частковий F-критерій визначиться як

$$F_{част_{x_l}} = \frac{R^2_{yx_1 \dots x_l \dots x_m} - R^2_{yx_1 \dots x_{l-1} x_{l+1} \dots x_m}}{1 - R^2_{yx_1 \dots x_l \dots x_m}} \cdot \frac{n-m-1}{1}.$$

Оцінка значущості коефіцієнтів чистої регресії за допомогою t-критерію Стьюдента збігається до обчислення значення

$$t_{b_l} = \frac{b_l}{m_{b_l}} = \sqrt{F_{x_l}}\ ,$$

де m_{bl} - середня квадратична помилка коефіцієнта регресії b_l, її можна визначити за наступною формулою

$$m_{b_l} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \dots x_m}}}{\sigma_{x_l}\sqrt{1 - R^2_{x_l x_1 \dots x_m}}} \cdot \frac{1}{\sqrt{n-m-1}}\ .$$

Під час побудови рівняння множинної регресії може виникнути проблема мультиколінеарності факторів, їх тісного лінійного зв'язку.

Вважають, що дві змінні явно колінеарні, тобто знаходяться одна з одною в лінійній залежності, якщо $r_{xixj} \geq 0{,}7$.

Величина парних коефіцієнтів кореляції виявляє лише явну колінеарність факторів. Найбільші труднощі у використанні апарату множинної регресії виникають за наявності мультиколінеарності факторів. Чим сильніша мультиколінеарність факторів, тим менше надійність оцінки розподілу суми поясненої варіації за окремими факторами за допомогою методу найменших квадратів.

Для оцінки мультиколінеарності факторів може використовуватися визначник матриці парних коефіцієнтів кореляції між факторами.

Якби фактори не корелювали між собою, то матриця парних коефіцієнтів кореляції між ними була б одиничною матрицею, оскільки всі недіагональні елементи r_{xixj} ($x_i \neq x_j$) дорівнювали би нулю. Так, для рівняння, що включає три пояснюючих змінних,

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$$

матриця коефіцієнтів кореляції між факторами має визначник, що дорівнює 1:

$$\mathrm{Det}|R| = \begin{vmatrix} r_{x_1x_1} & r_{x_2x_1} & r_{x_3x_1} \\ r_{x_1x_2} & r_{x_2x_2} & r_{x_3x_2} \\ r_{x_1x_3} & r_{x_2x_3} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 .$$

Якщо ж, навпаки, між чинниками існує повна лінійна залежність і всі коефіцієнти кореляції дорівнюють 1, то визначник такої матриці дорівнює 0:

$$\mathrm{Det}|R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0 .$$

Чим ближчий до 0 визначник матриці міжфакторної кореляції, тим сильніше мультиколінеарність факторів та ненадійніше результати множинної регресії. І навпаки, чим ближче до 1 визначник матриці міжфакторної кореляції, тим менше мультиколінеарність факторів.

Перевірку мультиколінеарності факторів можна провести методом випробування гіпотези про незалежність змінних H_0:Det|R|=1. Доведено, що величина $\left[n-1-\frac{1}{6}\cdot(2m+5)\lg \mathrm{Det}R\right]$ має наближений розподіл χ^2 з $\left[\frac{1}{2}n(n-1)\right]$ ступенями свободи. Якщо фактичне значення χ^2 перевершує табличне (критичне) $\chi^2_{факт} > \chi^2_{табл}$, то гіпотеза H_0 відхиляється. Це означає, що $\mathrm{Det}R \neq 1$, недіагональні ненульові коефіцієнти кореляції вказують на колінеарність факторів. Мультиколінеарність визнається доведеною.

Для застосування МНК потрібно, щоб дисперсія залишків була гомоскедастичною. Це означає, що для кожного значення фактору x_i залишки ε_i мають однакову дисперсію. Якщо ця умова не виконується, то має місце гетероскедастичність.

При порушенні гомоскедастичності ми маємо нерівності

$$\sigma^2_{\varepsilon i} \neq \sigma^2_{\varepsilon j} \neq \sigma^2, \quad j \neq i .$$

Якщо обсяг вибірки малий, для оцінки гетероскедастичності може використовуватися метод Гольдфельда-Квандта. Основна ідея тесту Гольдфельда-Квандта полягає в наступному:

1) впорядкування n спостережень у міру зростання змінної x;

2) виключення з розгляду C центральних спостережень; при цьому (n - C) : 2 > p, де p - кількість оцінюваних параметрів;

3) розділення сукупності з (n - C) спостережень на дві групи (відповідно з малими та з великими значеннями фактору x) і визначення за кожною групою рівнянь регресії;

4) визначення залишкової суми квадратів для першої (S_1) та другої (S_2) груп

і знаходження їх відношення: R = $S_1 : S_2$.

Якщо нульова гіпотеза про гомоскедастичність виконується, відношення R задовольнятиме F-критерію із ступенями свободи ((n - C - 2p) : 2) для кожної залишкової суми квадратів. Чим більше величина R перевищує табличне значення F-критерію, тим більше порушено передумову про рівність дисперсій залишкових величин.

Задача 2.1

Нехай є наступні дані (умовні) про змінний видобуток вугілля на одного робітника y (т), потужності шару x_1 (м) та рівні механізації робіт x_2 (%), що характеризують процес видобутку вугілля у 10 шахтах.

Таблиця 2.1 – Вихідні дані

№	1	2	3	4	5	6	7	8	9	10
x_1	8	11	12	9	8	8	9	9	8	12
x_2	5	8	8	5	7	8	6	4	5	7
y	5	10	10	7	5	6	6	5	6	8

Припустимо, що між змінними y, x_1, x_2 існує лінійна кореляційна залежність та знайдемо рівняння регресії y на x_1 та x_2.

Для зручності подальших обчислень складаємо таблицю ($\varepsilon = y - \hat{y}_x$) (табл. 2.2).

Таблиця 2.2 - Проміжкові розрахунки

№	x_1	x_2	y	x_1^2	x_2^2	y^2	$x_1 \cdot x_2$	$x_1 \cdot y$	$x_2 \cdot y$	$\hat{y}_x$	ε^2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	8	5	5	64	25	25	40	40	25	5,13	0,016
2	11	8	10	121	64	100	88	110	80	8,79	1,464
3	12	8	10	144	64	100	96	120	80	9,64	0,127
4	9	5	7	81	25	49	45	63	35	5,98	1,038
5	8	7	5	64	49	25	56	40	35	5,86	0,741
6	8	8	6	64	64	36	64	48	48	6,23	0,052
7	9	6	6	81	36	36	54	54	36	6,35	0,121
8	9	4	5	81	16	25	36	45	20	5,61	0,377
9	8	5	6	64	25	36	40	48	30	5,13	0,762
10	12	7	8	144	49	64	84	96	56	9,28	1,631
Сума	94	63	68	908	417	496	603	664	445	68	6,329
Середнє значення	9,4	6,3	6,8	90,8	41,7	49,6	60,3	66,4	44,5	–	–
σ^2	2,44	2,01	3,36	–	–	–	–	–	–	–	–
σ	1,56	1,42	1,83	–	–	–	–	–	–	–	–

Для визначення параметрів рівняння регресії в цьому випадку необхідно розв'язати наступну систему нормальних рівнянь:

$$\begin{cases} 10a + 94b_1 + 63b_2 = 68, \\ 94a + 908b_1 + 603b_2 = 664, \\ 63a + 603b_1 + 417b_2 = 445. \end{cases}$$

Звідки отримаємо, що $a = -3{,}54$, $b_1 = 0{,}854$, $b_2 = 0{,}367$. Тобто рівняння множинної регресії має вигляд:

$$\hat{y}_x = -3{,}54 + 0{,}854 \cdot x_1 + 0{,}367 \cdot x_2.$$

Воно показує, що при збільшенні тільки потужності шару x_1 (при незмінному x_2) на 1 м здобування вугілля на одного робітника y зросте у середньому на 0,854 т, а при збільшенні тільки рівня механізації робіт x_2 (при незмінному x_1) на 1% – у середньому на 0,367 т.

Знайдемо рівняння множинної регресії у стандартизованому масштабі:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon,$$

при цьому визначимо стандартизовані коефіцієнти регресії:

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,854 \cdot \frac{1,56}{1,83} = 0,728, \qquad \beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,367 \cdot \frac{1,42}{1,83} = 0,285 \cdot$$

Отже отримали рівняння

$$\hat{t}_y = 0,728 \cdot t_{x_1} + 0,285 \cdot t_{x_2}.$$

Стандартизовані коефіцієнти регресії можна порівнювати один з одним. Отже можна побачити, що потужність шару впливає на змінний видобуток вугілля більше за рівень механізації робіт.

Порівнювати вплив факторів на результат можна так само за допомогою середніх коефіцієнтів еластичності:

$$\bar{Э}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}},$$

або

$$\bar{Э}_1 = 0,854 \cdot \frac{9,4}{6,8} = 1,18, \qquad \bar{Э}_2 = 0,367 \cdot \frac{6,3}{6,8} = 0,34 \cdot$$

Отже, збільшення тільки потужності шару (від свого середнього значення) або тільки рівня механізації робіт на 1% збільшує в середньому змінний видобуток вугілля на 1,18% або 0,34% відповідно. Це підтверджує, що вплив на результат y фактору x_1 є більшим за фактор x_2.

Задача 2.2

Оцінимо якість рівняння, що отримане у попередній задачі. Спочатку визначимо парні коефіцієнти кореляції:

$$r_{yx_1} = \frac{\overline{y \cdot x_1} - \bar{y} \cdot \bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = \frac{66,4 - 6,8 \cdot 9,4}{1,83 \cdot 1,56} = 0,869; \qquad r_{yx_2} = \frac{\overline{y \cdot x_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = \frac{44,5 - 6,8 \cdot 6,3}{1,83 \cdot 1,42} = 0,639;$$

$$r_{x_1x_2} = \frac{\overline{x_1 \cdot x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{60,3 - 9,4 \cdot 6,3}{1,56 \cdot 1,42} = 0,488 \cdot$$

Їх значення вказують на досить тісний зв'язок змінного видобутку вугілля на одного робітника y з потужністю шару x_1 та на помірний зв'язок з рівнем механізації робіт x_2. У той самий час міжфакторний зв'язок $r_{x_1x_2}$ не дуже сильний ($r_{x_1x_2} = 0,49 < 0,7$). Це показує, що обидва фактори є інформативними, тобто необхідно включити до моделі x_1 та x_2.

Визначимо сукупний коефіцієнт кореляції $R_{yx_1x_2}$. Для цього спочатку знайдемо визначник матриці парних коефіцієнтів кореляції

$$\Delta r = \begin{vmatrix} 1 & 0,87 & 0,64 \\ 0,87 & 1 & 0,49 \\ 0,64 & 0,49 & 1 \end{vmatrix} = 0,139064$$

та визначник матриці міжфакторної кореляції:

$$\Delta r_{11} = \begin{vmatrix} 1 & 0,49 \\ 0,49 & 1 \end{vmatrix} = 0,7599.$$

Тоді коефіцієнт множинної кореляції визначимо за формулою:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} = \sqrt{1 - \frac{0,139064}{0,7599}} = 0,904.$$

Можна сказати, що 81,7% (коефіцієнт детермінації $R^2_{yx_1x_2} = 0,817$) варіації результативної ознаки пояснюється варіацією представлених у рівнянні ознак-факторів, що вказує на досить тісний зв'язок ознак з результатом.

Приблизно той самий результат (розходження пов'язані з помилками округлень) дістанемо для коефіцієнта множинної регресії, якщо скористаємося формулами:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{0,6329}{3,36}} = 0,901; \quad R_{yx_1x_2} = \sqrt{\sum \beta_i \cdot r_{yx_i}} = \sqrt{0,728 \cdot 0,87 + 0,285 \cdot 0,64} = 0,903.$$

Скорегований коефіцієнт множинної детермінації

$$\hat{R} = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-m-1} = 1 - \left(1 - 0,817\right) \cdot \frac{10-1}{10-2-1} = 0,765$$

вказує на помірний зв'язок між результатом та ознаками. Це зумовлено малою кількістю спостережень.

Визначимо часткові коефіцієнти кореляції за формулами:

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}} = \sqrt{1 - \frac{1 - 0,817}{1 - 0,408}} = 0,831; \quad r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}} = \sqrt{1 - \frac{1 - 0,817}{1 - 0,755}} = 0,503.$$

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} = \frac{0,869 - 0,639 \cdot 0,488}{\sqrt{\left(1 - 0,489^2\right)\left(1 - 0,639^2\right)}} = 0,830;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} = \frac{0,639 - 0,869 \cdot 0,488}{\sqrt{\left(1 - 0,488^2\right)\left(1 - 0,869^2\right)}} = 0,498.$$

Звідси можна зробити висновок, що фактор x_1 надає сильніший вплив на результат, ніж фактор x_2.

Оцінимо надійність рівняння регресії в цілому та показника зв'язку за допомогою F - критерію Фішера. Фактичне значення F - критерію:

$$F_{\text{факт}} = \frac{R^2}{1 - R^2} \cdot \frac{n-m-1}{m} = \frac{0,817}{1 - 0,817} \cdot \frac{10-2-1}{2} = 15,63.$$

Табличне значення F - критерію за п'ятивідсотковий рівень значущості ($\alpha = 0,05$, $k_1 = 2$, $k_2 = 10 - 2 - 1 = 7$): $F_{\text{табл}} = 4,74$. Оскільки $F_{\text{факт}} = 15,63 > F_{\text{табл}} = 4,10$, то рівняння визнається статистично значущим.

Оцінимо доцільність включення фактору x_1 після фактору x_2 та x_2 після x_1 за допомогою частки F - критерію Фішера:

$$F_{x_1}=\frac{R^2_{yx_1x_2}-r^2_{yx_2}}{1-R^2_{yx_1x_2}}\cdot(n-3)=\frac{0,817-0,408}{1-0,817}\cdot 7=15,65;\quad F_{x_2}=\frac{R^2_{yx_1x_2}-r^2_{yx_1}}{1-R^2_{yx_1x_2}}\cdot(n-3)=\frac{0,817-0,755}{1-0,817}\cdot 7=2,37.$$

Табличне значення частки F - критерію за п'ятивідсотковий рівень значущості ($\alpha=0,05$, $k_1=1$, $k_2=10-2-1=7$): $F_{\text{табл}}=5,59$. Оскільки $F_{x_1}=15,65>F_{\text{табл}}=5,59$, а $F_{x_2}=2,37<F_{\text{табл}}=5,59$, то включення фактору x_1 до моделі статистично виправдане, коефіцієнт чистої регресії b_1 є статистично значущим, а додаткове включення фактору x_2, після того, як уже введений фактор x_1, є недоцільним.

Рівняння регресії, що включає тільки один значущий аргумент x_2:

$$\hat{y}=-2,754+1,016x_1.$$

Задача 2.3

За 30 територіями є дані, представлені в таблиці 2.3.

Таблиця 2.3 – Вихідні дані

Ознака	Середнє значення	Середнє квадратичне відхилення	Лінійний коефіцієнт парної кореляції
Середньоденний душовий дохід, грн. (y)	86,8	11,44	-
Середньоденна заробітна плата одного працюючого, грн. (x_1)	54,9	5,86	r_{yx1} =0 ,8405
Середній вік безробітного, років (x_2)	33,5	0,58	r_{yx2} = -0,2101 r_{x1x2} = -0,1160

1. Побудувати рівняння множинної регресії в стандартизованій і природній формі; розрахувати часткові коефіцієнти еластичності, порівняти їх з β_1 і β2, пояснити відмінності між ними.
2. Розрахувати лінійні коефіцієнти часткової кореляції та коефіцієнт множинної кореляції, порівняти їх з лінійними коефіцієнтами парної кореляції, пояснити відмінності між ними.
3. Розрахувати загальний та часткові F-критерії Фішера.

Розв'язання

1. Лінійне рівняння множинної регресії y від x_1 та x_2 має вигляд:

$$y = a + b_1x_1 + b_2x_2.$$

Для розрахунку його параметрів застосуємо метод стандартизації змінних і побудуємо шукане рівняння в стандартизованому масштабі:

$$t_y = \beta_1 t_{x1} + \beta_2 t_{x2}.$$

Розрахунок β–коефіцієнтів здійснимо за формулами

$$\beta_1=\frac{r_{yx1}-r_{yx2}r_{x1x2}}{1-r^2_{x1x2}}=\frac{0,8405-0,2101\cdot 0,116}{1-0,116^2}=\frac{0,8161}{0,9865}=0,8273;$$

$$\beta_2=\frac{r_{yx2}-r_{yx1}r_{x1x2}}{1-r^2_{x1x2}}=\frac{-0,2101-0,8405\cdot 0,116}{1-0,116^2}=\frac{-0,1126}{0,9865}=-0,1141.$$

Отримаємо рівняння

$$t_y = 0,8273t_{x1} - 0,1141t_{x2}.$$

Для побудови рівняння в природній формі розрахуємо b_1 та b_2, використовуючи формули для переходу від β_i до b_i

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}; \qquad b_1 = \frac{0,8273 \cdot 11,44}{5,86} = 1,6151; \qquad b_2 = \frac{-0,1141 \cdot 11,44}{0,58} = -2,2505.$$

Значення а визначимо із співвідношення

$$a = \overline{y} - b_1\overline{x}_1 - b_2\overline{x}_2 = 86,8 - 1,6151 \cdot 54,9 + 2,2505 \cdot 33,5 = -73,52$$

$$y = -73,52 + 1,62x_1 - 2,25x_2.$$

Для характеристики відносної сили впливу x_1 та x_2 на у розрахуємо середні коефіцієнти еластичності:

$$\overline{E}_{yxj} = b_j \frac{\overline{x}_j}{\overline{y}}; \qquad \overline{E}_{yx1} = \frac{1,62 \cdot 54,9}{86,8} = 1,025\%; \qquad \overline{E}_{yx2} = \frac{-2,25 \cdot 33,59}{86,8} = -0,868\%.$$

Із збільшенням середньої заробітної платні x_1 на 1% від її середнього рівня середній душовий дохід у зростає на 1,02% від свого середньою рівня; при підвищенні середнього віку безробітного x_2 на 1% средньодушовій дохід у знижується на 0,87% від свого середнього рівня. Очевидно, що сила впливу середньої заробітної платні x_1 на середній душовий дохід у виявилася більшою, ніж сила впливу середнього віку безробітного x_2. До аналогічних висновків про силу зв'язку приходимо при порівнянні модулів значень β_1 та β_2:

$$|\beta_1| = |0,8273| > |\beta_2| = |-0,1141|.$$

Відмінності у силі впливу фактору на результат, одержані порівнянням $\overline{E}_{yxj}$ та β_j, пояснюються тим, що коефіцієнт еластичності виходить із співвідношення середніх, а β–коефіцієнт - із співвідношення середніх квадратичних відхилень.

2. Лінійні коефіцієнти часткової кореляції розрахуємо за рекурентною формулою:

$$r_{yx1x2} = \frac{r_{yx1} - r_{yx_2} \cdot r_{x1x2}}{\sqrt{\left(1 - r^2_{yx2}\right)\left(1 - r^2_{x1x2}\right)}} = \frac{0,8405 - 0,210` \cdot 0,116}{\sqrt{(1 - 0,2101^2)(1 - 0,116^2)}} = 0,8404;$$

$$r_{yx2x1} = \frac{r_{yx2} - r_{yx_1} \cdot r_{x1x2}}{\sqrt{\left(1 - r^2_{yx1}\right)\left(1 - r^2_{x1x2}\right)}} = \frac{-0,2101 + 0,8405` \cdot 0,116}{\sqrt{(1 - 0,8405^2)(1 - 0,116^2)}} = -0,209;$$

$$r_{x1x2y} = \frac{r_{x1x2} - r_{yx_1} \cdot r_{yx2}}{\sqrt{\left(1 - r^2_{yx1}\right)\left(1 - r^2_{yx2}\right)}} = \frac{-0,116 + 0,08405` \cdot 0,2101}{\sqrt{(1 - 0,8405^2)(1 - 0,2101^2)}} = 0,1144.$$

Якщо порівняти значення коефіцієнтів парної та часткової кореляції, то приходимо до висновку, що із-за слабкого міжфакторного зв'язку ($r_{x1x2} = 0,116$) коефіцієнти парної та часткової кореляції відрізняються неістотно: висновки про тісноту та напрям зв'язку на підставі коефіцієнтів парної та часткової кореляції співпадають:

$R_{yx1} = 0,8405$; $r_{yx2} = -0,2101$; $r_{x1x2} = 0,116$; $r_{yx1x2} = 08404$; $r_{yx2x1} = -0,2092$; $r_{x1x2y} = 0,1144$.

Розрахунок лінійного коефіцієнта множинної кореляції здійснимо з використанням коефіцієнтів r_{yxj} та β_j.

$$R_{yx1x2} = \sqrt{r_{yx1}\beta_1 + r_{yx2}\beta_2} = \sqrt{0{,}8405 \cdot 0{,}8273 + 0{,}2101 \cdot 0{,}1141} = \sqrt{0{,}7193} = 0{,}8481.$$

Залежність у від x_1 та x_2 характеризується як тісна, в якій 72% варіації середнього душового доходу визначаються варіацією врахованих в моделі факторів: середньої заробітної платні та середнього віку безробітного. Інші чинники, що не включені у модель, складають відповідно 28% від загальної варіації у.

3. Загальний F-критерій перевіряє гіпотезу H_0 про статистичну значущість рівняння регресії та показника тісноти зв'язку ($R^2 = 0$):

$$F_{факт} = \frac{R^2_{yx1x2}}{1 - R^2_{yx1x2}} \cdot \frac{n-m-1}{m} = \frac{0{,}7193}{0{,}2807} \cdot \frac{27}{2} = 34{,}6; \qquad F_{табл} = 3{,}4; \ \alpha=0{,}05.$$

Порівнюючи $F_{табл}$ та $F_{факт}$, приходимо до висновку про необхідність відхилити гіпотезу H_0, оскільки $F_{табл} = 3{,}4 < F_{факт} = 34.6$. З імовірністю $1 - \alpha = 0{,}95$ робимо висновок про статистичну значущість рівняння в цілому та показника тісноти зв'язку R_{yx1x2}, які сформувалися під невипадковою дією факторів x_1 та x_2.

Часткові F-критерії F_{x1} та F_{x2} оцінюють статистичну значущість присутності чинників x_1 та x_2 у рівнянні множинної регресії, оцінюють доцільність включення до рівняння одного фактору після іншого фактору, тобто F_{x1} оцінює доцільність включення у рівняння фактору x_1, після того, як в нього був включений фактор x_2. Відповідно F_{x2} вказує на доцільність включення у модель фактору x_2 після фактору x_1.

$$F_{x1факт} = \frac{R^2_{yx1x2} - r^2_{yx2}}{1 - R^2_{yx1x2}} \cdot \frac{n-m-1}{1} = \frac{0{,}8481^2 - 0{,}2101^2}{1 - 0{,}8481^2} \cdot \frac{30-2-1}{1} = 64{,}9;$$

$$F_{табл} = 4{,}21; \ \alpha=0{,}05.$$

Порівнюючи $F_{табл}$ і $F_{факт}$, приходимо до висновку про доцільність включення у модель фактору x_1 після фактору x_2, оскільки $F_{x1факт}=64{,}9 > F_{табл}$. Гіпотезу H_0 про неістотність приросту R^2_y за рахунок включення додаткового фактора x_1 відхиляємо та приходимо до висновку про статистично підтверджену доцільність включення фактору x_1 після фактора x_2.

Доцільність включення у модель фактору x_2 після фактору x_1 перевіряє F_{x2}:

$$F_{x2факт} = \frac{R^2_{yx1x2} - r^2_{yx1}}{1 - R^2_{yx1x2}} \cdot \frac{n-m-1}{1} = \frac{0{,}8481^2 - 0{,}8405^2}{1 - 0{,}8481^2} \cdot \frac{30-2-1}{1} = 1{,}234.$$

Мале значення $F_{x2факт}$ (дещо більше 1) свідчить про статистичну незначущість приросту r^2_{yx1}, за рахунок включення у модель фактора x_2 після фактора x_1. Отже, підтверджується нульова гіпотеза H_0 про недоцільність включення у модель фактора x_2 (середній вік безробітного). Це означає, що парна регресійна модель залежності середнього доходу від середньої заробітної платні є достатньо статистично значущою, надійною і що немає необхідності покращувати її, включаючи додатковий чинник x_2 (середній вік безробітного).

МЕТОДИЧНІ ВКАЗІВКИ ДО САМОСТІЙНОЇ РОБОТИ

ЗМ 1.1. Економетричне моделювання. Побудова загальної лінійної моделі

Тема 1. ПРЕДМЕТ І ЗАДАЧІ ДИСЦИПЛІНИ (2 години)

1. Предмет, методи й завдання дисципліни
2. Етапи економетричного моделювання
3. Економетрична модель, класифікація

Джерела: [1] с. 6-23; [2] с. 15-41; [4] с. 4-16.

Запитання для самоконтролю

1. Наведіть визначення економетрії.
2. За якими ознаками економетрію визначають як відокремлену науку?
3. З якими науками пов'язана економетрія?
4. З якою метою проводять економетричні дослідження?
5. Охарактеризуйте етапи економетричного дослідження. Які питання вирішує економетрія?
6. Яку роль відіграє математична статистика у формуванні економетричних методів?
7. Які типи даних використовують у екометричних дослідженнях. Які проблеми даних виникають?
8. Поясніть складові загальної економетричної моделі.
9. Поясніть відмінності між ендогенними і екзогенними змінними.
10. У чому полягає ідентифікація та верифікація економетричної моделі?
11. Поясніть терміни «регресія», «умовне математичне сподівання» та «збурення».
12. Які класи моделей використовують для аналізу або прогнозу в економетрії?
13. Які властивості повинна мати регресійна модель з одним рівнянням, побудована на основі просторової вибірки?
14. Поясніть терміни «гомоскедастичність» та «гетероскедастичність».
15. Яку вибірку спостережень називають часовим рядом?
16. Які економетричні моделі належать до систем одночасних рівнянь?

Тема 2. ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ (14 годин)

1. Побудова загальної лінійної моделі
2. Лінійна модель парної регресії
3. Оцінка значущості рівняння лінійної регресії

Джерела: [1] с. 50-75; [2] с. 43-77; [4] с. 16-31.

Запитання для самоконтролю

1. Поясніть терміни «рівняння регресії» і «функція регресії».
2. Які помилки належать до помилок специфікації? До помилок виміру?
3. На яких підставах вибирають вид регресійної залежності?
4. Як розраховують залишкову дисперсію? Загальну дисперсію ознаки y?
5. У чому полягає сутність методу найменших квадратів?
6. Якими властивостями має володіти лінійна модель, щоб оцінки її параметрів мали найменшу дисперсію в класі всіх лінійних незміщених оцінок?
7. Поясніть, що таке коефіцієнт регресії? Коефіцієнт кореляції? Коефіцієнт детермінації?
8. Як визначають статистичну значущість рівняння регресії в цілому?

9. З якою метою визначають стандартну помилку коефіцієнта регресії? Який критерій для цього використовують?

10. Як визначають довірчий інтервал для коефіцієнта регресії?

11. З якою метою і як визначають значущість лінійного коефіцієнта кореляції?

12. У якому випадку використовують лінійну модель множинної регресії?

13. Що таке коефіцієнти «чистої» регресії? Який в них економічний зміст?

14. У якому випадку використовують стандартизовані коефіцієнти регресії?

15. Що характеризують показник множинної кореляції та показник детермінації?

16. Що таке лінійний коефіцієнт множинної кореляції і скорегований індекс множинної кореляції?

17. Для чого розраховують часткові коефіцієнти кореляції?

18. Як оцінюють значущість рівняння множинної регресії в цілому та значущість коефіцієнтів чистої регресії?

19. Поясніть, у чому полягає процедура відбору факторів під час побудови рівняння регресії за методом виключення?

20. Охарактеризуйте помилки специфікації моделі. У чому вони полягають?

21. Поясніть зміст коефіцієнта регресії, назвіть способи його оцінювання, покажіть, як його використовують для розрахунку мультиплікатора у функції споживання.

22. Поясніть, що розуміють під кількістю ступенів свободи. Як її визначають для факторної та залишкової сум квадратів?

23. Поясніть призначення та концепцію критерію Фішера.

24. Як оцінюють значущість параметрів рівняння регресії?

25. На яких посиланнях базується використання методу найменших квадратів?

26. Які властивості оцінюваних параметрів забезпечує використання методу найменших квадратів?

27. Поясніть сутність середньої помилки апроксимації. Як її визначають?

28. Що оцінюють за допомогою коефіцієнта еластичності?

Приклад 2.1. Дані про залежність змінного видобутку вугілля одного робітника Y (т) та потужністю шару X (м) на 10 шахтах наведені у таблиці

i	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
x_i	8	11	12	9	8	8	9	9	8	12
y_i	5	10	10	7	5	6	6	5	6	8

Знайти лінійне рівняння регресії Y за X.

Розв'язання

Лінійне рівняння парної регресії має вигляд

$$\hat{y}_x = a + b \cdot x .$$

та скористаємось формулами:

$$b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{66{,}4 - 9{,}4 \cdot 6{,}8}{2{,}71} = 1{,}016 \,;$$

$$a = \overline{y} - b\overline{x} = 6{,}8 - 1{,}016 \cdot 9{,}4 = -2{,}75 \,.$$

Для визначення його параметрів зробимо обчислення у таблиці:

	x	y	$x \cdot y$	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	$\left(y-\hat{y}_x\right)^2$	A_i, %
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	8	5	40	64	25	5,38	-0,38	0,1422	15,33
2	11	10	110	121	100	8,43	1,57	2,4768	13,08
3	12	10	120	144	100	9,44	0,56	0,3107	4,11
4	9	7	63	81	49	6,39	0,61	0,3679	5,5
5	8	5	40	64	25	5,38	-0,38	0,1422	5,81
6	8	6	48	64	36	5,38	0,62	0,3881	2,83
7	9	6	54	81	36	6,39	-0,39	0,1548	0,85
8	9	5	45	81	25	6,39	-1,39	1,9417	4,68
9	8	6	48	64	36	5,38	0,62	0,3881	
10	12	8	96	144	64	9,44	-1,44	2,0812	
Разом	96	71	668	913	502,00	75,00	8,00	17,39	52,19
Середнє значення	9,4	6,8	66,4	90,8	49,6	6,80	0,00	0,84	6,52
σ	1,65	1,93	–	–	–	1,67	0,97	0,93	–
σ^2	2,71	3,73	–	–	–	2,80	0,93	0,87	–

Отримали рівняння

$$\hat{y} = 2{,}75 + 1{,}016x\,,$$

з якого випливає, що при збільшенні потужності шару X на 1 м видобуток на одного робітника Y зростає у середньому на 1,016 т. Вільний член у цьому рівнянні не має реального змісту.

Приклад 2.2. Для умов попереднього прикладу обчислити коефіцієнт кореляції між змінними X та Y.

Розв'язання

Для визначення коефіцієнта кореляції скористаємось формулою

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = 1{,}016 \cdot \frac{1{,}65}{1{,}93} = 0{,}866\,.$$

Таке значення коефіцієнту кореляції вказує на тісний лінійний зв'язок між ознаками. Коефіцієнт детермінації $r_{xy}^2 = 0{,}75$ показує, що рівняння регресії пояснює 75% дисперсії результативної ознаки, а на долю інших факторів припадає 25%.

Приклад 2.3. Для умов прикладу 2.1 оцінити якість рівняння регресії в цілому за допомогою F - критерію Фішера та статистичну значущість коефіцієнтів регресії та кореляції.

Розв'язання

Для оцінки якості рівняння регресії в цілому за допомогою F - критерію Фішера розрахуємо його фактичне значення:

$$F = \frac{r_{xy}^2}{1-r_{xy}^2} \cdot (n-2) = \frac{0{,}75}{1-0{,}75}(10-2) = 24\,.$$

Табличне значення ($k_1 = 1$, k_2=n-2=8, $\alpha = 0{,}05$): $F_{табл} = 5{,}32$. Оскільки

$F_{\text{факт}} > F_{\text{табл}}$, визнаємо статистичну значущість рівняння в цілому.

Для оцінки статистичної значущості коефіцієнтів регресії та кореляції розрахуємо t - критерій Стьюдента та довірчі інтервали кожного з показників.

Розрахуємо випадкові помилки параметрів лінійної регресії та коефіцієнту кореляції

$$S^2_{\text{зал}} = \frac{\sum(y-\hat{y}_x)^2}{n-2} = \frac{17{,}39}{8} = 2{,}17; \quad m_b = \frac{s_{\text{зал}}}{\sigma_x \cdot \sqrt{n}} = \frac{\sqrt{2{,}17}}{1{,}65 \cdot \sqrt{10}} = 0{,}28$$

$$m_a = s_{\text{зал}} \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n} = \frac{\sqrt{2{,}17 \cdot 913}}{1{,}65 \cdot 10} = 2{,}7; \quad m_r = \sqrt{\frac{1-r^2}{n-2}} = \sqrt{\frac{1-0{,}75}{8}} = 1{,}77.$$

Тоді фактичні значення t - статистик:

$$t_b = \frac{b}{m_b} = \frac{1{,}016}{0{,}28} = 3{,}63; \quad t_a = \frac{a}{m_a} = \frac{2{,}75}{2{,}7} = 1{,}02; \quad t_r = \frac{r}{m_r} = \frac{0{,}866}{1{,}77} = 0{,}49.$$

Табличне значення t - критерію Стьюдента при $\alpha = 0{,}05$ та кількості ступенів свободи $\nu = n-2 = 8$ дорівнює $t_{\text{таб}} = 2{,}31$. Отже $t_b > t_{\text{табл}}$, параметр b є статистично значущим, $t_a < t_{\text{табл}}$ і параметр a є незначущим, $t_r < t_{\text{табл}}$, показник тісноти зв'язку так само є незначущим.

Розрахуємо довірчі інтервали для параметрів регресії a та b: $a \pm t \cdot m_a$; $b \pm t \cdot m_b$. Отримаємо, що $a \in$[-5,45; -0,05]; $b \in$[0,74; 1,3].

Задачі для самостійного розв'язання

Задача 2.1. Є наступні дані про рівень механізації робіт X(%) та продуктивності праці Y (т/год) для 14 однотипних підприємств:

x_i	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Необхідно:

а) оцінити тісноту і напрям зв'язку між змінними за допомогою коефіцієнта кореляції;

б) знайти рівняння регресії Y за X.

Задача 2.2. За результатом досліджень кореляційної залежності між ціною на нафту X та індексом нафтових компаній Y одержані наступні дані:

x =16,2 (грош. од), y = 4000 (умовн. од) s_x^2 =4, s_y^2 =500, Cov (X,Y)=40.

Необхідно:

а) скласти рівняння регресії Y за X;

б) використовуючи рівняння регресії, знайти середнє значення індексу при ціні на нафту 16,5 грош. од.

Задача 2.3. За даними задачі 2.1:

а) знайти рівняння регресії Y за X;

б) знайти коефіцієнт детермінації R^2 та пояснити його смисл;

в) перевірити значущість рівняння регресії на 5%-му рівні за F-критерієм;

г) оцінити середню продуктивність праці на підприємствах з рівнем механізації робіт 60% та побудувати для неї 95%-ий довірчий інтервал; аналогічний довірчий інтервал знайти для індивідуальних значень продуктивності праці на тих самих підприємствах.

Задача 2.4.. За даними 30 нафтових компаній одержано наступне рівняння регресії між оцінкою Y (грош. од.) та фактичною вартістю X (грош. од.) цих компаній: $y_x = 0{,}8750x + 295$. Знайти: 95% - ві довірчі інтервали для середнього та індивідуального значень оцінки підприємств, фактична вартість яких склала 1300 грош. од., якщо коефіцієнт кореляції між змінними дорівнює 0,76, а середнє квадратичне відхилення змінної X дорівнює 270 грош. од.

Тема 3. МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ (12 годин)

1. Лінійна модель множинної регресії
2. Оцінка значущості множинної регресії і показники якості моделі
3. Мультиколінеарність та її вплив на оцінки параметрів моделі
4. Лінійні регресійні моделі з гетероскедастичними залишками
5. Узагальнений метод найменших квадратів (УМНК)
6. Регресійні моделі із змінною структурою. Фіктивні змінні

Джерела: [1] с. 82-115, 150-167; [2] с. 109-166, 182-214, ; [4] с. 31-68.

Запитання для самоконтролю

1. У чому полягає специфікація моделі множинної регресії?
2. Сформулюйте вимоги до факторів щодо включення їх до моделі множинної регресії.
3. До яких ускладнень призводить мультиколінеарність факторів? Як її подолати?
4. Які коефіцієнти використовують для оцінювання порівняльного впливу факторів на результативну ознаку?
5. Яке призначення часткової кореляції під час побудови моделі множинної кореляції?
6. Як пов'язані один з одним t-критерій Стьюдента для оцінювання значущості b_i та часткові F-критерії?
7. Охарактеризуйте основні посилання щодо застосування методу найменших квадратів для побудови регресійної моделі.
8. Охарактеризуйте сутність аналізу залишків регресійної моделі.
9. Як перевірити наявність гомо- або гетероскедастичності залишків?
10. Сформулюйте умови застосування узагальненого методу найменших квадратів.
11. Якими властивості повинні мати оцінки параметрів регресії?
12. Як впливає на параметри множинної лінійної моделі порушення умови, що математичне сподівання збурювання ε дорівнює нулю?
13. Як впливає на параметри множинної лінійної моделі порушення умови, що дисперсія збурювання ε постійною?
14. Який спосіб використовують для вивчення гомо- і гетероскедастичності?
15. На підставі якого дослідження роблять висновок про автокорельованість залишків?
16. Як поводяться при недотриманні основних передумов МНК?
17. Чим відрізняється узагальнений МНК від звичайного? У яких випадках його використовують?
18. Які властивості притаманні оцінкам параметрів моделі, отриманим на основі узагальненого МНК?

Приклад 3.1. Для дослідження залежності між продуктивністю праці (X_1), віком (X_2) і виробничим стажем (X_3) була проведена вибірка з 100 робітників тієї самої спеціальності. Обчислені парні коефіцієнти кореляції виявилися значущими і склали: r_{12}=0,20; r_{13}=0,41; r_{23}=0,82. Обчислити часткові коефіцієнти кореляції та оцінити їх значущість на рівні α=0,05.

Розв'язання

Часткові коефіцієнти кореляції обчислюють за формулою

$$r_{ijk} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{\left(1-r_{ik}^2\right)\left(1-r_{jk}^2\right)}}, \qquad r_{123} = \frac{0{,}2^2 - 0{,}41 \cdot 0{,}82}{\sqrt{(1-0{,}41)^2(1-0{,}82)^2}} = -0{,}26 .$$

аналогічно отримаємо r_{132}=0,44; r_{231}=0,83.

Оцінимо значущість r_{123}. Значення статистики t-критерію за формулою при n' = n-p+ 2= 100-3+2=99 (за абсолютною величиною)

$$|r_1| = \frac{|-0{,}26|\sqrt{99-2}}{\sqrt{1-(-0{,}26)^2}} = 2{,}65$$

більше за табличне $t_{0,95;97}$=1,99), отже, частковий коефіцієнт кореляції r_{123} є значущим. Аналогічно встановлюється значущість інших часткових коефіцієнтів кореляції.

Порівнюючи часткові коефіцієнти кореляції r_{ijk} з відповідними парними коефіцієнтами, бачимо, що за рахунок «очищення зв'язку» найбільшій зміні піддався коефіцієнт кореляції між продуктивністю праці (X_1) та віком (X_2) робітників (змінилося не тільки його значення, але й знак r_{12}=0,20; r_{123}=-0,26, причому обидва ці коефіцієнти є значущими).

Отже, між продуктивністю праці (X_1) та віком (X_2) робітників існує прямий кореляційний зв'язок (r_{12}=0,20). Якщо ж усунути (елімінувати) вплив змінної «виробничий стаж» (X_3), то в чистому вигляді продуктивність праці (X_1) знаходиться у зворотному за напрямом (і слабкому за тіснотою) зв'язку з віком робітників (X_2) (r_{123}=-0,26). Це цілком зрозуміло, якщо розглядати вік тільки як показник працездатності організму на певному етапі його життєдіяльності. Так само можуть бути інтерпретовані й інші часткові коефіцієнти кореляції.

Задачі для самостійного розв'язання

Задача 3.1. Є наступні дані про споживання певного продукту Y (умовн. од.) залежно від рівня урбанізації (частки міського населення) X_1, відносного освітнього рівня X_2 та відносного заробітку X_3 для дев'яти географічних районів:

і (номер району)	x_{i1}	x_{i2}	x_{i3}	y_i
1	42,2	11,2	31,9	167,1
2	48,6	10,6	13,2	174,4
3	42,6	10,6	28,7	160,8
4	39,0	10,4	26,1	162,0
5	34,7	9,3	30,1	140,8
6	44,5	10,8	8,5	174,6
7	39,1	10,7	24,3	163,7
8	40,1	10,0	18,6	174,5
9	45,9	12,0	20,4	185,7
Середні	41,85	10,62	24,42	167,07

Стандартні відхилення σ_{x1} = 4,176; σ_{x2} = 0,7463; σ_{x3} = 7,928; σ_y=12,645. Кореляційна матриця має вигляд:

	X_1	X_2	X_3	Y
X_1	1	0,684	-0,616	0,802
X_2	0,684 -	1	-0,173	0,770
X_3	0,616	-0,173	1	-0,629
Y	0,802	0,770	-0,629	1

Використовуючи покрокову процедуру відбору найбільш інформативних пояснюючих змінних, визначити відповідну регресійну модель, виключивши при цьому мультиколінеарність. Оцінити значущість коефіцієнтів регресії одержаної моделі за t-критерієм.

Задача 3.2. Є наступні дані про вагу Y (у фунтах) та вік X (у тижнях) 13 індичок, вирощених в областях А, В, С.

i	x_i	y_i	Область походження
1	28	12,3	A
2	20	8,9	A
3	32	15,1	A
4	22	10,4	A
5	29	13,1	B
6	27	12,4	B
7	28	13,2	B
8	26	11,8	B
9	21	11,5	C
10	27	14,2	C
11	29	15,4	C
12	23	13,1	C
13	25	13,8	C

Є підстава вважати, що на вагу індичок робить вплив не тільки їх вік, але й область походження. Необхідно:

а) знайти рівняння парної регресії Y за X та оцінити його значущість;

б) ввести відповідні фіктивні змінні та знайти загальне рівняння множинної регресії за всіма пояснюючими змінними (включаючи фіктивні);

в) оцінити значущість загального рівняння множинної регресії за F-критерієм та значущість його коефіцієнтів за t-критерієм на рівні α=0,05;

г) простежити за зміною скорегованого коефіцієнта детермінації при переході від парної до множинної регресії;

д) оцінити на рівні α=0,05 значущість відмінності між вільними членами рівнянь, що одержуються із загального рівняння множинної регресії Y для кожної області.

Задача 3.3. Під час побудови лінійної залежності витрат на одяг від доходу за вибіркою для 10 жінок одержані наступні суми квадратів та добутків спостережень:

$$\sum_{1}^{10} x_i = 110, \quad \sum_{1}^{10} x_i^2 = 1540, \quad \sum_{1}^{10} y_i = 60, \quad \sum_{1}^{10} x_i y_i = 828, \quad \sum_{1}^{10} y_i^2 = 448.$$

Аналогічні обчислення сум за вибіркою з 5 чоловіків дали:

$$\sum_{1}^{5} x_i = 35, \quad \sum_{1}^{5} x_i^2 = 325, \quad \sum_{1}^{5} y_i = 15, \quad \sum_{1}^{5} x_i y_i = 140, \quad \sum_{1}^{5} y_i^2 = 61.$$

За загальною (об'єднаною) вибіркою оцінено регресію з використанням фіктивної змінної Z (Z= 1 для чоловіка і Z= 0 для жінки), яка має вигляд:

$$\hat{y} = -0{,}06 + 0{,}438x + 0{,}46z + 0{,}072(zx).$$

На рівні значущості α=0,05 перевірити гіпотезу, що функція споживання одна і та сама для чоловіків та жінок, якщо виконані всі передумови класичної нормальної лінійної регресії.

Задача 3.4. З метою дослідження впливу чинників X_1 - середньомісячної кількості профілактичних налаштувань автоматичної лінії та X_2 - середньомісячної кількості обривів нитки на чинник Y - середньомісячну характеристику якості тканини (у балах) за даними 37 підприємств легкої промисловості були обчислені парні коефіцієнти кореляції: r_{y1} = 0,105, r_{y2} = 0,024 та r_{12}=0,996. Визначити часткові коефіцієнти кореляції r_{y12} та r_{y21} і оцінити їх значущість на 5%-му рівні.

ЗМ 1.2. Економетричні моделі динаміки. Система структурних рівнянь

Тема 4. ЧАСОВІ РЯДИ І ПРОГНОЗУВАННЯ (10 годин)

1. Загальні відомості про часові ряди і завдання їх аналізу
2. Автокореляція рівнів часового ряду
3. Моделювання тенденції часового ряду
4. Моделювання сезонних коливань
5. Автокореляція залишків часового ряду. Критерій Дарбіна-Уотсона

Джерела: [1] с. 133-149, 167-185; [2] с. 296-324, 436-446, ; [4] с. 69-80.

Запитання для самоконтролю

1. Наведіть визначення часового ряду.
2. Чим відрізняються часові ряди від звичайних просторових вибірок?
3. Які фактори впливають на рівні часового ряду?
4. Які компоненти містить реальний часовий ряд?
5. Поясніть адитивну і мультиплікативну моделі часового ряду.
6. Поясніть, що таке тренд часового ряду і які види тренду зустрічаються?
7. Як визначають значущість тренду?
8. Перелічіть основні етапи аналізу часових рядів.
9. Перелічіть найпоширеніші методи аналізу часових рядів.
10. Поясніть, що таке автокореляція рівнів часового ряду. Як її можна виміряти?
11. Перелічіть властивості коефіцієнта автокореляції.
12. Що таке автокореляційна функція часового ряду і корелограма?
13. З якою метою виконують аналітичне вирівнювання часового ряду?
14. Які функції найчастіше застосовують для побудови трендів?
15. Який метод дозволяє визначити параметри тренду?
16. Які методи використовують для згладжування часового ряду?
17. В чому полягає метод ковзних середніх?
18. Поясніть, як провадиться усунення сезонних коливань за методом ковзної середньої?

19. З якою метою провадиться спектральний аналіз часового ряду?
20. Поясніть, у чому полягає зв'язний аналіз часових рядів?
21. Перелічіть основні елементи часового ряду.
22. Що являє собою автокореляція рівнів часового ряду та як її оцінюють кількісно?
23. Приведіть визначення автокореляційної функції часового ряду.
24. Перелічіть основні види трендів.
25. Яку інтерпретацію мають параметри лінійного та експоненційного трендів?
26. Запишіть загальний вигляд мультиплікативної та адитивної моделей часового ряду.
27. Охарактеризуйте основні етапи побудови мультиплікативної та адитивної моделей часового ряду.
28. З якою метою виявляють наявність сезонного ефекту?

Приклад 4.1. Часовий ряд попиту y_t наведено у таблиці.

Рік, t	1	2	3	4	5	6	7	8
Попит, y_t	213	171	291	309	317	362	351	361

За даними таблиці визначить середнє значення, середнє квадратичне відхилення, коефіцієнти автокореляції (для лагів т=1;2) і частковий коефіцієнт автокореляції 1-го порядку.

Розв'язання

Знаходимо середнє значення часового ряду за формулою:

$$\bar{y}_t = \frac{\sum_{t=1}^{n} y_t}{n} = \frac{213+171+291+309+317+362+351+361}{8} = 296{,}88 \text{ (од.)}.$$

Дисперсію і середнє квадратичне відхилення обчислимо за співвідношенням

$$s_t^2 = \overline{y_t^2} - \bar{y}_t^2 = 92478{,}38 - 296{,}88^2 = 4343{,}61 \qquad s_t = \sqrt{4343{,}61} = 65{,}31 \text{ (од.)}.$$

де $$\overline{y_t^2} = \frac{\sum_{t=1}^{n} y_t^2}{n} = \frac{213^2+171^2+291^2+309^2+317^2+362^2+351^2+361^2}{8} = 92478{,}38.$$

Знайдемо коефіцієнт автокореляції $r(\tau)$ часового ряду (для лага $\tau = 1$), тобто коефіцієнт кореляції між послідовностями семи пар спостережень y_t та y_{t+1} (t= 1,2...,7):

y_t	213	171	291	309	317	362	351
y_{t+1}	171	291	309	317	362	351	361

Обчислюємо необхідні суми:

$$\sum_{t=1}^{7} y_t = 213 + 171 + 291 + 309 + 317 + 362 + 351 = 2014;$$

$$\sum_{t=1}^{7} y_t^2 = 213^2 + 171^2 + 291^2 + 309^2 + 317^2 + 362^2 + 351^2 = 609506;$$

$$\sum_{t=1}^{7} y_{t+\tau} = 171 + 291 + 309 + 317 + 362 + 351 + 361 = 2162;$$

$$\sum_{t=1}^{7} y_{t+\tau}^2 = 171^2 + 291^2 + 309^2 + 317^2 + 362^2 + 351^2 + 361^2 = 694458;$$

$$\sum_{t=1}^{7} y_t y_{t+\tau} = 213\cdot171 + 171\cdot291 + 291\cdot309 + 309\cdot317 + 317\cdot362 + 362\cdot351 + 351\cdot361 = 642583.$$

Тепер обчислимо коефіцієнт автокореляції

$$r(1)=\frac{7\cdot 642583-2014\cdot 2162}{\sqrt{7\cdot 609506-2014^2}\sqrt{7\cdot 694458-2162^2}}=0{,}725\,.$$

Коефіцієнт автокореляції r(2) для лага τ = 2 між членами ряду y_t і y_{t+2} (t= 1,2...,6) за шістьма парами спостережень обчислюємо аналогічно: r(2) = 0,842.

Для визначення часткового коефіцієнта кореляції 1-го порядку $r_{част}(2)$= =r_{021} між членами ряду y_t і y_{t+2} при виключенні впливу y_{t+1} спочатку знайдемо (по аналогії з попереднім) коефіцієнт автокореляції r(1,2) між членами ряду y_{t+1} і y_{t+2}: r(1,2)=0,825, а потім обчислимо $r_{част}(2)$:

$$r_{чст}(2)=r_{021}=\frac{0{,}842-0{,}725\cdot 0{,}825}{\sqrt{1-0{,}725^2}\sqrt{1-0{,}825^2}}=0{,}627\,.$$

Знання автокореляційних функцій $r(\tau)$ і $r_{част}(\tau)$ може надати істотну допомогу під час підбору та ідентифікації моделі аналізованого часового ряду і статистичній оцінці його параметрів.

Приклад 4.2. За даними прикладу 4.1 знайти рівняння невипадкової складової (тренду) для часового ряду y_t, вважаючи тренд лінійним.

Розв'язання

Відповідно до методу найменших квадратів параметри лінійної залежності визначають з системи нормальних рівнянь. З урахуванням, що значення t утворюють натуральний ряд чисел від 1 до n, суми $\sum_{t=1}^{n} t$ та $\sum_{t=1}^{n} t^2$ можна виразити як кількість членів ряду n за відомими формулами:

$$\sum_{t=1}^{n} t=\frac{n(n+1)}{2}=\frac{8\cdot 9}{2}=36\,,\qquad \sum_{t=1}^{n} t^2=\frac{n(n+1)(2n+1)}{6}=\frac{8\cdot 9\cdot 17}{6}=204\,.$$

Далі

$$\sum_{t=1}^{8} y_t = 213+171+291+309+317+362+351+361=2375\ ;$$

$$\sum_{t=1}^{8} y_t^2 = 213^2+171^2+291^2+309^2+317^2+362^2+351^2+361^2=739827\,;$$

$$\sum_{t=1}^{8} y_t t = 213\cdot 1+171\cdot 2+291\cdot 3+309\cdot 4+317\cdot 5+362\cdot 6+351\cdot 7+361\cdot 8=11766\ .$$

Система нормальних рівнянь має вигляд:

$$\begin{cases} 8b_0+36b_1=2375 \\ 36b_0+204b_1=11766\ . \end{cases}$$

Отримаємо b_0=181,32 та b_1=25,679. Отже рівняння тренду має вигляд:

$$\hat{y}_t=181{,}32+25{,}679t\,,$$

тобто попит щорічно збільшується у середньому на 25,7 одиниць.

Перевіримо значущість отриманого рівняння тренду за F-критерієм на 5%-му рівні значущості. Обчислимо дисперсії.

Дисперсія, зумовлена регресією

$$Q_R = \sum_{t=1}^{n}\left(\hat{y}_t - \bar{y}_t\right)^2 = \sum b_1^2\left(t-\bar{t}\right)^2 = b_1^2\left(\sum_{t=1}^{n} t^2 - \frac{\left(\sum_{t=1}^{n} t\right)^2}{n}\right) = 25{,}679^2\left(204 - \frac{36^2}{8}\right) = 27695{,}3.$$

Дисперсія загальна

$$Q = \sum_{t=1}^{n}\left(y_t - \bar{y}_t\right)^2 = \sum y_t^2 - \frac{\left(\sum_{t=1}^{n} y_t\right)^2}{n} = 739827 - \frac{2375^2}{8} = 34748{,}9$$

Дисперсія залишкова

$$Q_{зал} = Q\text{-}Q_R = 34748{,}9\text{-}27695{,}3 = 7053{,}6.$$

Обчислимо значення статистики

$$F = \frac{Q_R(n-2)}{Q_{зал}} = \frac{27695{,}3 \cdot 6}{7053{,}6} = 23{,}56\,.$$

Оскільки $F > F_{0,05;1;6}$, рівняння тренду є значущим.

Задачі для самостійного розв'язання

Задача 4.1. Дані про врожайність озимої пшениці y (ц/га) за 10 років наведено у таблиці 4.1.

Таблиця 4.1 - Врожайність озимої пшениці y (ц/га)

t	1	2	3	4	5	6	7	8	9	10
y_t	16,3	20,2	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

Знайти середнє значення, середнє квадратичне відхилення та коефіцієнти автокореляції для лагов $\tau = 1;2$ часового ряду.

Задача 4.2. За даними табл. 4.1 знайти рівняння тренду часового ряду y_t, вважаючи що він лінійний. Перевірте значущість тренду часового ряду на рівні 0,05.

Задача 4.3. У таблиці 4.2 представлені дані, що відображають динаміку зростання доходів на душу населення y_t (грош. од.) за восьмирічний період:

Таблиця 4.2 - Доходи на душу населення y_t (грош. од.)

t	1	2	3	4	5	6	7	8
y_t	1133	1222	1354	1389	1342	1377	1491	1684

Вважаючи, що тренд лінійний і умови класичної моделі виконані, знайти рівняння тренду, оцінити його значущість на рівні 0,05 та дати точковий і з надійністю 0,95 інтервальний прогнози середнього та індивідуального значень доходів на дев'ятий рік.

Тема 5. СИСТЕМА СТРУКТУРНИХ РІВНЯНЬ (8 годин)

1. Поняття системи структурних рівнянь
2. Структурна та приведена форми моделі
3. Проблема ідентифікації
4. Методи оцінки параметрів структурної форми моделі

Джерела: [1] с. 224-242; [2] с. 246-284; [4] с. 81-94.

Запитання для самоконтролю

1. Які основні причини використання систем одночасних рівнянь?
2. Перелічіть можливі способи побудови систем рівнянь. Чим вони відрізняються один від одного?
3. Як пов'язані одна з одною структурна та приведена форми моделі? У чому їх основна розбіжність?
4. Чому для оцінювання систем одночасних рівнянь майже не використовують звичайний метод найменших квадратів?
5. У чому полягає сутність непрямого методу найменших квадратів?
6. У чому полягають проблеми ідентифікації моделі? Які умови ідентифікації існують?
7. Охарактеризуйте причини неідентифікованості та надідентифікованості систем структурних рівнянь?

Приклад 5.1. Розглянемо модель виду:

$$\begin{cases} y = a_1 + b_1(C + D) + \varepsilon_1 \\ C = a_2 + b_2 y + b_3 y_{-1} + \varepsilon_2 \end{cases},$$

де y – валовий національний дохід; y_{-1} - валовий національний дохід попереднього року; C - особисте споживання; D - кінцевий попит (окрім особистого споживання); ε_1 та ε_2 - випадкові складові.

Інформація про приріст чинників за дев'ять років наведена у таблиці

Рік	D	y_{-1}	y	C
1	-6,8	46,7	3,1	7,4
2	22,4	3,1	22,8	30,4
3	-17,3	22,8	7,8	1,3
4	12	7,8	21,4	8,7
5	5,9	21,4	17,8	25,8
6	44,7	17,8	37,2	8,6
7	23,1	37,2	35,7	30
8	51,2	35,7	46,6	31,4
9	32,3	46,6	56	39,1
Сума	167,5	239,1	248,4	182,7

Для цієї системи отримана система приведених рівнянь:

$$\begin{cases} y = 8{,}219 + 0{,}6688D + 0{,}261y_{-1} \\ C = 8{,}636 + 0{,}3384D + 0{,}202y_{-1} \end{cases}.$$

Провести ідентифікацію моделі та розрахувати параметри першого рівняння структурної моделі.

Розв'язання

Модель містить дві ендогенні змінні (у та С) та дві екзогенні змінні (D та y_{-1}). Друге рівняння є точно ідентифікованим, оскільки містить дві ендогенні змінні та не містить одну екзогенну змінну з системи.

Перше рівняння є надідентифікованим, оскільки у ньому на параметри при С та D накладено обмеження – вони повинні дорівнювати один одному. У цьому рівнянні міститься одна ендогенна змінна у. Змінна С у цьому рівнянні не розглядається як ендогенна, оскільки вона міститься у рівнянні разом з змінною D. У цьому рівнянні відсутня одна екзогенна змінна, існуюча у системі. Отже, за рахувальним правилом отримуємо 1+1=2:D+1>H. Це перевищує кількість ендогенних змінних у цьому рівнянні, відповідно система є надідентифікованою.

Для визначення параметрів надідентифікованої моделі скористаємося двокроковим методом найменших квадратів.

На першому кроці за точно ідентифікованим другим рівнянням визначимо теоретичні значення ендогенної змінної С. Для цього у приведене рівняння

$$C = 8{,}636 + 0{,}3384D + 0{,}202y_{-1}$$

Підставимо значення D та y_{-1} з умови. Отримаємо

$$\hat{C}_1 = 15{,}8;\ \hat{C}_2 = 16{,}8;\ \hat{C}_3 = 7{,}4;\ \hat{C}_4 = 14{,}3;\ \hat{C}_5 = 15;\ \hat{C}_6 = 27{,}4;$$

$$\hat{C}_7 = 24;\ \hat{C}_8 = 33{,}2;\ \hat{C}_9 = 29.$$

На другому кроці за надідентифікованим рівнянням структурної форми моделі змінюєм фактичні значення С на теоретичні $\hat{C}$ та розраховуємо нову змінну $\hat{C}$+D у наступній таблиці

Рік	**D**	$\hat{C}$	$\hat{C}$+D
1	-6,8	15,8	9
2	22,4	16,8	39,2
3	-17,3	7,4	-9,9
4	12	14,3	26,3
5	5,9	15	20,9
6	44,7	27,4	72,1
7	23,1	24	47,1
8	51,2	33,2	84,4
9	32,3	29	61,3
Сума	167,5	182,9	350,4

Далі до надідентифікованого рівняння застосуємо метод найменших квадратів. Позначимо нову змінну $\hat{C}$+D = Z та розв’яжемо рівняння

$$Y = a_1 + b_1Z.$$

Система нормальних рівнянь має вигляд

$$\begin{cases}\sum y = na_1 + b_1\sum Z\\ \sum y\cdot Z = a_1\cdot\sum Z + b_1\sum Z^2\end{cases};\qquad \begin{cases}248{,}4 = 9a_1 + 350{,}4b_1\\ 13508{,}71 = 350{,}4a_1 + 21142{,}02b_1\end{cases};$$

$$a_1=7{,}678;\quad b_1=0{,}512.$$

Отже, перше рівняння структурної моделі матиме вигляд

$$Y=7{,}678+0{,}512(C+D).$$

Задачі для самостійного розв'язання

Задача 5.1. Розглядається система рівнянь:

$$Y_1 = \beta X + \gamma Y_2 + \varepsilon_1$$
$$Y_2 = \delta Y_1 + \varepsilon_2$$

Перевірити, чи є ця система ідентифікуємою. Чи зміниться відповідь, якщо у склад регресорів другого рівняння включити: а) константу; б) змінну X.

Задача 5.2. До системи двох рівнянь виду:

$$Y_1 = \beta_1 X_1 + \gamma_1 Y_2 + \varepsilon_1$$
$$Y_2 = \beta_2 X_2 + \gamma_2 Y_1 + \varepsilon_2$$

Застосовано непрямий метод найменших квадратів. Для коефіцієнтів приведеної форми

$$Y_1 = \beta_1 X_1 + \gamma_1 Y_2 + \varepsilon_1$$
$$Y_2 = \beta_2 X_2 + \gamma_2 Y_1 + \varepsilon_2$$

Отримані наступні оцінки c_1=2,2, c_2=0,4, c_3=0,08, c_4=-0,5.

Знайти оцінки двокрокового методу найменших квадратів щодо системи.

Задача 5.3. Під час оцінювання системи

$$Y_1 = \beta_1 X_1 + \gamma_1 Y_2 + \varepsilon_1$$
$$Y_2 = \beta_2 X_2 + \gamma_2 Y_1 + \varepsilon_2$$

за двокроковим та трикроковим методами найменших квадратів отримано однакові оцінки. Чи будуть оцінки, визначені за звичайним методом найменших квадратів, спроможними?

ЗМ 1.3. Методи аналізу на підставі статистичних рівнянь. Модель з автокорельованими залишками. Моделі розподіленого лагу

Тема 6. НЕЛІНІЙНІ ОДНОФАКТОРНІ МОДЕЛІ (10 годин)

1. Види нелінійних моделей парної регресії
2. Моделі нелінійні щодо пояснюючих змінних
3. Моделі нелінійні за оцінюваними параметрами
4. Виробнича функція Кобба-Дугласа

Джерела: [1] с. 124-130; [2] с. 77-108; [4] с. 95-104.

Запитання для самоконтролю

1. Якою нелінійною функцією можна замінити параболу другої степені, якщо не спостережується зміна спрямованості зв’язку ознак?

2. Перелічіть види моделей нелінійних щодо уведених змінних та оцінюваних параметрів.

3. Охарактеризуйте відмінності у застосуванні методу найменших квадратів

до моделей нелінійних щодо змінних та до моделей нелінійних щодо оцінюваних параметрів.

4. Як визначають коефіцієнти еластичності за різними видами регресійних моделей?

5. Які показники кореляції, що використовують за нелінійні співвідношення досліджуваних ознак?

6. Як обирають перетворення лінеаризації для моделей із внутрішнею нелінійністю?

7. За якими ознаками класифікують методи оцінювання параметрів нелінійних моделей?

Приклад 6.1. Нехай відомі значення двох ознак:

i	y	x
1	68,8	45,1
2	61,2	59
3	59,9	57,2
4	56,7	61,8
5	55	58,8
6	54,3	47,2
7	49,3	55,2

Визначити параметри степеневої функції та рівнобічної гіперболи.

Розв'язання

Для побудови степеневої моделі $y = ax^b$ треба лінеаризувати змінні. Виконаємо це шляхом логарифмування обох частин рівняння

$$\text{Lg}y = \lg a + b\lg x;$$
$$Y = C + bX,$$

де $Y = \lg y$, $C = \lg a$, $X = \lg x$.

Скористаємось методом найменших квадратів та одержимо лінійне рівняння

$$\hat{Y} = 2{,}278 - 0{,}298X.$$

Виконаємо потенціювання та одержимо рівняння

$$\hat{y} = 10^{2,278} \cdot x^{-0,298} = 189{,}7 \cdot x^{-0,298}.$$

Для побудови рівнобічної гіперболи $y = a + b \cdot \frac{1}{x}$ лінеаризуємо її заміною $z = \frac{1}{x}$. Тоді отримаємо $y = a + bz$. Параметри a та b за методом найменших квадратів мають наступні значення: a=38,5, b=1051,4. Рівняння рівнобічної гіперболи має вигляд

$$y = 38{,}5 + 1051{,}4 \cdot \frac{1}{x}.$$

Задачі для самостійного розв'язання

Задача 6.1. Визначити параметри експонентної функції для умови прикладу 6.1.

Задача 6.2. Нелінійне однофакторне рівняння регресії має вигляд

$$y = f(x) + \varepsilon = \alpha e^{\beta x} + \varepsilon.$$

Треба розкласти функцію f(x) в ряд Тейлора другого порядку у точці x_0=0 та визначити, чому дорівнює границя розкладання в ряд n-го порядку при $n \to \infty$.

Задача 6.3. Є нелінійне однофакторне рівняння регресії

$$y_t = \alpha_0 + \alpha_1 x_t^{\alpha_2}.$$

Записати систему нормальних рівнянь для визначення оцінок параметрів α_0, α_1 та α_2.

Задача 6.4. Чи можна наступні рівняння перетворити на рівняння, лінійні за параметрами?

а) $y = \alpha \cdot e^{\beta x} \cdot \varepsilon$;

б) $y = \alpha \cdot e^{-\beta x} + \varepsilon$;

в) $y = e^{\alpha + \beta x + \varepsilon}$;

г) $y = \dfrac{\alpha}{\beta - x} + \varepsilon$.

Тема 7. РЕГРЕСІЙНІ ДИНАМІЧНІ МОДЕЛІ (8 годин)

1. Причини виникнення лагових ефектів в економетричних моделях
2. Методи оцінки параметрів з урахуванням лагових ефектів

Джерела: [1] с. 191-215; [2] с. 454-494; [4] 105-116.

Запитання для самоконтролю

1. Наведіть приклади економічних завдань, для яких економетричне моделювання потребує використання моделей з розподіленим лагом та моделей авторегресії.
2. Що в економіці називають лагом? Що є причиною лага?
3. Які регресійні моделі називають авторегресійними?
4. Які проблеми виникають під час оцінювання параметрів авторегресійних моделей?
5. Які гіпотези висувають щодо залишків лагових моделей?
6. Охарактеризуйте інтерпретацію параметрів моделі з розподіленим лагом. Перелічіть абсолютні та відносні чинники сили зв'язку моделі з розподіленим лагом.
7. Охарактеризуйте інтерпретацію параметрів моделі авторегресії.

Приклад 7.1. На підставі часових рядів, що характеризують чисту продукцію та капітальні вкладення, треба побудувати взаємну кореляційну функцію. Вихідні дані наведені у таблиці:

Рік	**Капітальні вкладення, тис. грн.**	**Чиста продукція, тис. грн.**	**Рік**	**Капітальні вкладення, тис. грн.**	**Чиста продукція, тис. грн.**
1	3857	24334	11	17006	72165
2	4686	28678	12	17352	78743
3	5515	33021	13	17838	80381
4	5209	32432	14	18878	82204
5	7522	40325	15	19090	77833
6	10390	49334	16	20016	81413
7	13678	54717	17	17736	77484
8	15976	53818	18	11951	75443
9	13880	55968	19	11469	85038
10	13949	61517	20	9068	75809

Розв'язання

Виконаємо розрахунки коефіцієнтів кореляції для різних значень τ за формулою:

$$r_\tau = \frac{(n-\tau)\sum_{t=1}^{n-\tau} y_t x_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} x_{t+\tau}}{\sqrt{\left[(n-\tau)\sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t\right)^2\right]\left[(n-\tau)\sum_{t=1}^{n-\tau} x_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} x_{t+\tau}\right)^2\right]}}$$

та зведемо їх у таблицю

τ	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
r_τ	0,89	0,86	0,89	0,92	0,92	0,85	0,75	0,64	0,4	0,55

Отже, найбільше значення коефіцієнту кореляції дорівнює 0,92, він відповідає двом значенням τ (4 та 5), тобто найбільшого впливу капітальних вкладень на обсяг чистої продукції треба очікувати впродовж четвертого та п'ятого років.

Динамічна модель розподіленого лага має вигляд:

$$y_t = a_0 x_t + a_1 x_{t-3} + a_2 x_{t-4} + a_3 x_{t-5} + u_t,$$

де y_t – чиста продукція в період t; a_j, j=0,1,2,3 - вагові коефіцієнти лагових змінних; $x_{t-\tau}$, τ=3,4,5 - капітальні вкладення в період t-τ.

Задачі для самостійного розв'язання

Задача 7.1. Оцінюється модель Y=βX+ε за звичайним методом найменших квадратів. Одержано рівняння виду

$$\hat{y} = 1{,}2x.$$

Проте відомо, що вимірювання регресора X виконуються з помилками, дисперсія яких дорівнює 10. Чи можна вважати, що істинне значення коефіцієнта β є додатним, якщо: а) дисперсія X дорівнює 100; б) дисперсія X дорівнює 10?

Задача 7.2. Для оцінки лагового рівняння $y_t=\gamma y_{t-1}+\varepsilon_t$ застосований звичайний метод найменших квадратів та одержано рівняння

$$\hat{y} = 0{,}9y_{t-1}, \quad d = 1{,}95.$$

Помилки регресії є стаціонарним авторегресійним процесом першого порядку. Чи можна вважати, що коефіцієнт λ: а) перевищує 0,5; б) перевищує 0,7, якщо обсяг вибірки великий.

СПИСОК ДЖЕРЕЛ

1. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов/ Под ред. проф. Н.Ш.Кремера.- М.: ЮНИТИ-ДАНА, 2002.- 311 с.

2. Эконометрика: учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева. – М.: Финансы и статистика, 2007. – 576 с.

3. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордиенко и др.; Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002 – 192 с.

4. Воронков О. О., Конспект лекцій з курсу «Економетрія» (для студ. галузі

знань 0306 - «Менеджмент і адміністрування» напряму 6.030601 - «Менеджмент» заочної форми навчання) / Харк. нац. акад. міськ. госп-ва; авт.: І. А. Ачкасов, О. О. Воронков, Т. Б. Воронкова – Х.: ХНАМГ, 2011.- 120 с.

5. Наконечный С.И., Терещенко Т.П. Эконометрия, - К.:КНЕУ, 2001.

6. Лещинський О. Л. Економетрія. – К.: МАУП, 2003. – 208 с.

7. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс. – М.: Дело, 2001. -400 с.

8. Тихомиров Н. П., Дорохина Е. Ю. Эконометрика. – М.: Издательство «Экзамен», 2003. – 512 с.

Додаток

Таблиця F-розподілу Фішера- Снедекора
для рівня значущості α=0,05 (5%)

n-m-1	m									
	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,20	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,68	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	1,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98

Таблиця t-розподілу Стьюдента
(критичні значення t (α, n-m-1))

Тест	**Рівень значущості α**							
Двобічний	**50 %**	**20 %**	**10 %**	**5 %**	**2 %**	**1 %**	**0,2 %**	**0,1 %**
Однобічний	**25 %**	**10 %**	**5 %**	**2,5 %**	**1 %**	**0,5 %**	**0,1 %**	**0,05 %**
n-m-1								
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31	636,62
2	0,861	1,886	2,920	4,303	6,965	9,925	22,327	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587

ЗМІСТ

Навчальне видання

Методичні вказівки
до практичних занять і самостійної роботи
з курсу

ЕКОНОМЕТРІЯ

(для студентів галузі знань 0306 «Менеджмент і адміністрування» напряму 6.030601 «Менеджмент» заочної форми навчання ЦПО та ЗН»)

Укладач: **ВОРОНКОВ** Олексій Олександрович

Відповідальний за випуск *А. Є. Ачкасов*

За редакцією укладача

Комп'ютерне верстання *К. А. Алексанян*

План 2012, поз. 592 М

Підп. до друку 02.10.2012 Формат 60x84/16
Друк на різографі. Ум. друк. арк. 2,3
Зам. № Тираж 50 пр.

Видавець і виготовлювач:
Харківський національний університет міського господарства
імені О. М. Бекетова,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@ksame.kharkov.ua
Свідоцтво суб'єкта видавничої справи:
ДК № 4064 від 12.05.2011р.