НАЦІОНАЛЬНА АКАДЕМІЯ УПРАВЛІННЯ

ЕКОНОМЕТРИКА

НАВЧАЛЬНО-МЕТОДИЧНІ МАТЕРІАЛИ

(ЗАОЧНЕ ВІДДІЛЕННЯ)

Київ - 2018

Укладачі:
Лопатін Олексій Костянтинович
професор, доктор фізико-математичних наук,
Черненко Ольга Борисівна,
старший викладач

ЗМІСТ

1. Задачі эконометрики

Можна виділити три основних задачі економетрики.

A. Підбір підхожої моделі, що адекватна наявним даним. При цьому може бути відсутня економічна теорія для опису досліджуваного явища.

B. Методи оцінювання параметрів економетричної моделі. Типовим формулюванням може бути таке:
з ймовірністю 95% можна стверджувати, що невідомий параметр β знаходиться в інтервалі:

$$\beta_1 \leqslant \beta \leqslant \beta_2,$$

де β_1, β_2 – відомі числа.

C. Економічне прогнозування на базі побудованих моделей.

Нижче наведені приклади розв'язання сформульованих задач для часових рядів (ковзкі середні, экспоненційне сгладжування) і не часових рядов – двухфакторна регресія.

2. Сгладжування даних і прогнозування на основі ковзких середнїх

Найбільше поширений і простий засіб визначення тенденції развитку– це сгладжування часового ряду. Основна ідея сгладжування– заміна фактичних рівнів часового ряду розрахунковими даними.

Приклад 1. Маються дані на попит пилососів за 12 місяців

Таблиця 1

Місяць (t)	y_i (Попит)	y_i^* (Попит за ковзким середнім
1	46	–
2	56	–
3	54	–
4	43	52
5	57	51
6	56	51
7	67	52
8	62	60
9	50	62
10	56	60
11	47	56
12	56	51

Потрібно побудувати криву згладжування і дати прогноз попиту на пилососи на 13 і 14 місяць.

Р о з в'я з а н н я. Попит y_{t+1}^* за методом ковзкого середнього визначається як середнє арифметичне від попиту за попередні k місяців. Число k називається *базою*.

Приймемо $k = 3$. Тоді

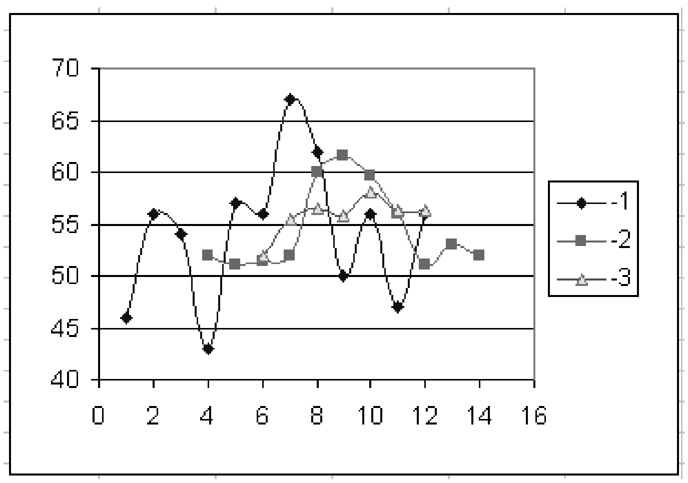
$$y_4^* = \frac{y_3 + y_3 + y_1}{3} = \frac{46 + 56 + 54}{3} = 52,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_{12}^* = \frac{y_{11} + y_{10} + y_9}{3} = \frac{56 + 47 + 56}{3} = 51.$$

Значення y_i^* записано в третій стовпчик Таблиці 1.

Прогноз на 13 і 14 місяці

$$y^*_{13} = \frac{y_{12} + y_{11} + y_{10}}{3} = \frac{56 + 47 + 56}{3} = 53,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$y^*_{14} = \frac{y_{13} + y_{12} + y_{11}}{3} = \frac{53 + 56 + 47}{3} = 52.$$

На мал. 1 показана крива попиту для вибірки 1 і згладжена крива 2 для $k = 3$.



Мал. 1

Зауваження 1. *Прогнозування даних на заданий часовий інтервал (у розглядованому випадку [1, 12]) називається* ***інтерполяцією****. Прогнозування даних поза інтервалом (у розглядованому випадку [1, 12]), називається* ***екстраполяцією****. Отже, прогноз на* 13 *і* 14 *місяці— екстраполяція.*

Зауваження 2. *Крива, що згладжує, приблизно описує теоретичну модель для всієї генеральної сукупності*

$$y_t = b + \varepsilon_t,$$

де $b = \text{const}$, ε_t *– випадкова помилка.*

Зауваження 3. *Немає чіткого правила для вибору* k *– бази методу.Чим більше база, тим більше сгладжування. (Порівняєте криві сгладжування 2 і 3 на мал. 1, виконані для* $k = 3$ *та* $k = 6$ *відповідно.)*

3. Експоненційне сгладжування та прогнозування

Експоненційне сгладжування є подальшим узагальненням сгладжування за ковзким середнім. Попит y^*_{t+1} за методом експоненційного сгладжування визначається за рекуррентною формулою

$$y^*_{t+1} = \alpha y_t + (1-\alpha)y^*_t.$$

Тут α – константа сгладжування, $0 \leqslant \alpha \leqslant 1$. При $t = 1$, y^*_1 не визначаються. Для $t = 2$ покладаємо $y^*_2 = y_2$.

Приклад 1 (продовження) Побудувати криву сгладжування для $\alpha = 0.1$ і дати прогноз попиту на пилососи на 13 місяць.

Р о з в'я з а н н я.

$$y^*_2 = \alpha y_2 + (1-\alpha)y^*_2 = 0.1 \cdot 56 + 0.9 \cdot 46 = 47,$$
$$y^*_4 = \alpha y_3 + (1-\alpha)y^*_3 = 0.1 \cdot 54 + 0.9 \cdot 47 = 47,7$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

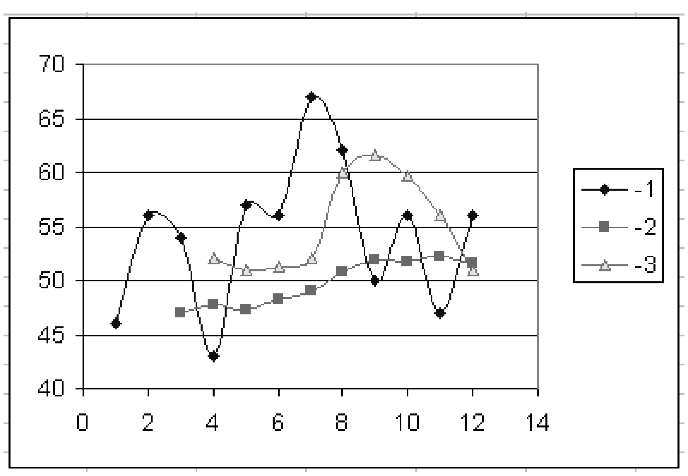
Подальші розрахунки наведені в Таблиці 2.

Таблиця 2

Місяц (t)	y_i (Попит)	y_i^* (експоненціальні значення)
1	**46**	–
2	**56**	**46**
3	**54**	$0,1\times 56+0,9\times 46=47$
4	**43**	$0,1\times 54+0,9\times 47=47,7$
5	**57**	$0,1\times 43+0,9\times 47,7=47,2$
6	**56**	$0,1\times 57+0,9\times 47,2=48,2$
7	**67**	$0,1\times 56+0,9\times 48,2=49,0$
8	**62**	$0,1\times 67+0,9\times 48,9=50,8$
9	**50**	$0,1\times 62+0,9\times 50,8=51,9$
10	**56**	$0,1\times 50+0,9\times 51,9=51,7$
11	**47**	$0,1\times 56+0,9\times 51,7=52,1$
12	**56**	$0,1\times 47+0,9\times 52,1=51,6$

Прогноз на 13 місяць

$$y_{13}=\alpha y_{12}+(1-\alpha)y_{12}^*=0,1\times 56+0,9\times 51,6=52,1.$$



Мал. 2

Зауваження 4. *Чим менше α, тим краще сгладжування. Практично α береться від* 0.01 *до* 0.3.

На мал. 3 подані (1) — крива вихідних даних, (2) — крива експоненційного сгладжування, (3) — крива сгладжування за ковзким середнім

4. Побудова лінії регресії та її економічний зміст

Регресійний аналіз визначає зв'язок між залежною змінною, наприклад, попит продукції і незалежними (пояснюючими) змінними, наприклад, витрати на рекламу продукції, її ціна і.д.

Сама проста регресійна модель передбачє, що залежна змінна є лінійною функцією від незалежної змінної x:

$$y = \alpha + \beta x + \varepsilon, \tag{1}$$

де α, β — коефіцієнти, ε — випадкова помилка.

Співвідношення (1) — це гадане теоретичне співвідношення між змінними x і y у всій генеральній сукупності, причому α, β і ε — невідомі. У дійсності, ми маємо вибірку з n пар чисел $(x_1, y_1), \ldots, (x_n, y_n)$. По цій вибірці будується вибіркова лінія регресії

$$y' = a + bx, \tag{2}$$

яка є лише деяким наближенням до співвідношень (1).

Зауваження 5. *Передбачається, що випадкові помилки ε мають нульове математичне сподівання і однакову дисперсію для всіх значень y, що спостерігаються*

Побудова вибіркової лінії регресії

Коефіцієнти a, b у (2) знаходяться за методом найменших квадратів (МНК). Останнє означає, що сума квадратів відхилень

$$S = \sum e_i^2 = e_1^2 + \cdots + e_n^2,$$

де

$$e_i = y_i - y_i^{'} = y_i - a - bx_i,$$

найменша.

Коефіцієнт b визначається за формулою

$$b = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sum (x - \overline{x})^2} \tag{3}$$

або за еквівалентною формулою

$$b = \frac{\sum xy - n\overline{xy}}{\sum x^2 - n\overline{x}^2}.$$

Коефіцієнт a знаходиться за формулою

$$a = \overline{y} - b\overline{x}. \tag{4}$$

Тут

$$\overline{x} = \frac{\sum x}{n}, \quad \overline{y} = \frac{\sum y}{n}. \tag{5}$$

Знак Σ означає підсумовування від $i = 1$ до n, тобто, наприклад, $\sum x = \sum\limits_{i=1}^{n} x_i$ і т.д.

Зауваження 6. *Вибіркова лінія регресії (2), побудована за МНК, проходить через точку $\overline{x}, \overline{y}$ (5).*

Приклад 2. Власник ательє бажає з'ясувати співвідношення між віком швейної машини x (роки) і витратами на її ремонт y (долари). Є вибіркові дані

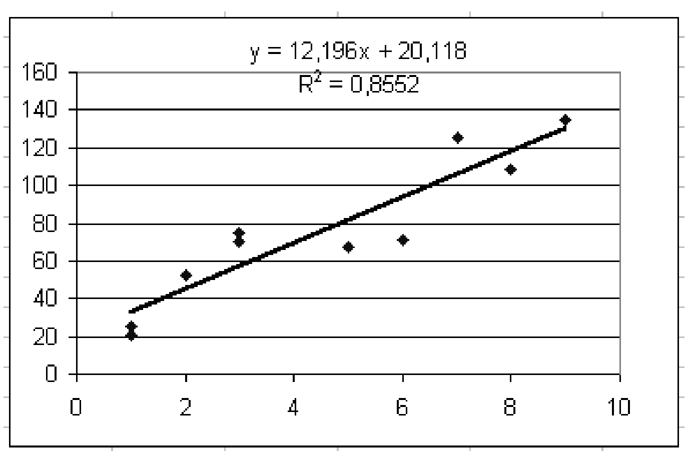
Таблица 3

N	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	$y-\overline{y}$	$(x-\overline{x})(y-\overline{y})$
1	8	109	3.5	12.25	34	19
2	3	75	-1.5	2.25	0	0
3	1	21	-3.5	12.25	-54	189
4	9	135	4.5	20.25	60	270
5	5	67	0.5	0.25	-8	-4
6	7	125	2.5	6.25	50	125
7	6	71	1.5	2.25	-4	-6
8	2	52	-25	6.25	-23	57.5
9	1	25	-3.5	12.25	-50	175
10	3	70	-1.5	2.25	-5	7.5
Сума	45	750	0	76.5	0	933
Середнє	4.5	75				

Тут Среднє $= \frac{\text{Сумма}}{10}$. По формулах (3) і (4) знаходимо $b = \frac{933}{76.5} \approx 12.2$, $a = 75 - 12.2 \cdot 4.5 \approx 20.1$. Отже,

$$y' = 20.1 + 12.2x. \tag{6}$$

Діаграма розсіювання та вибіркова лінія регресії зображені на мал. 3.



Мал. 3

Економічна інтепретація вибіркової лінії регресії (6)

- коефіцієнт нахилу $12,2$ показує, що якщо вік швейної машини збільшити на 1 рік, витрати на її ремонт збільшуються на $\$$ 12.2.
- якщо машина нова $(x = 0)$, то витрати на її ремонт складуть $\$$ 20.1. Цей висновок суперечить здоровому глузду.

Означення 1. *Регресійне співвідношення між змінними x і y називається **еластичним**, якщо зміна змінної x на 1% призводить до збільшення y більш, ніж на 1%, і **не є еластичним** у противному випадку.*

Для розрахунку еластичності застосовується формула

$$\frac{\Delta y}{y} \cdot 100 = \frac{1}{1 + \dfrac{a}{bx}}. \tag{7}$$

Підставляючи у формулу **(7)** коефіцієнти **(6)**, знходимо

$$\frac{\Delta y}{y} \cdot 100 = \frac{1}{1 + \dfrac{20.1}{12.2x}} < 1,$$

тобто наше регресійне співвідношення **(7)** не є еластичним.

5. Використання вибіркової лінії регресії для прогнозування й оцінювання

Задача А. $100(1 - \alpha)\%$ довірчий інтервал для невідомих (істинних) значень y у точці x_0 по заданій виборці, знаходиться за формулою

$$y' \pm t_{1-\frac{\alpha}{2}}\, S_{y/x}\sqrt{1+\frac{1}{n}+\frac{(x_0-\overline{x})^2}{\Sigma(x-\overline{x})^2}},$$

де

$$S_{y/x}\sqrt{\frac{\Sigma(y-y')^2}{n-2}}=\sqrt{\frac{\Sigma e^2}{n-2}}, \qquad (9)$$

α — рівень значущості. Звичайно $\alpha = 0.05$ або 0.01.

$t_{1-\frac{\alpha}{2}}$ — число, що знаходиться за таблицею розподілу Стьюдента.

Потрібно також знати число ступенів свободи df. Для нашого прикладу, тому що $n = 10$, $df = n - 2 = 8$.

Якщо $\alpha = 0.05$, то $t_{0,95} = 2.306$;

Якщо $\alpha = 0.01$, то $t_{0,95} = 3.3554$.

Задача В. $100(1-\alpha)\%$ довірчий інтервал для оцінки середнього значення $\overline{y}$ у точці x_0 задається формулою

$$y' \pm t_{1-\frac{\alpha}{2}}\, S_{y/x}\sqrt{1+\frac{(x_0-\overline{x})^2}{\Sigma(x-\overline{x})^2}}. \qquad (10)$$

Задача А називається задачею *прогнозування* y, а задача В — задачею *оцінювання* середнього значення $\overline{y}$.

Якщо точка x_0, що вибирається для прогнозування або оцінювання, належить розглянутому у вибірці інтервалу значень x, то маємо задачу *інтерполяції*, у противнім випадку — задачу *екстраполювання*.

Приклад 2 (продовження). Знайти оцінку витрат на ремонт швейної машини у віці $x_0 = 4$ роки і віці $x_0 = 10$ років. Побудуйте відповідні довірчі інтервали.

Розв'язання. Всі дані для застосування формул **(9)**, **(10)** обчислюємо в таблиці 4

$$y'|_{x_0=a} = 12.2\cdot 4 + 20.1 = 68.9,\ \ S_{y/x} = \sqrt{\frac{1927.1}{8}} = 15.52.$$

95% довірчий інтервал для $x_0 = 4$

$$68.5 \pm 2.306 \cdot 15,32\sqrt{1+\frac{1}{8}+\frac{(4-4.5)^2}{76.5}} \Rightarrow 68.9 \pm 38; \quad [30.9; 106.9].$$

Таблиця 4

No	x	y	$x-\overline{x}$	$(x-\overline{x})^2$	y'	$e=y-y'$	e^2
1	8	109	3.5	12.25	117.7	-8.7	75.7
2	3	75	-1.5	2.25	56.7	18.3	334.9
3	1	21	-3.5	12.25	32.3	-11.3	127.7
4	3	135	4.5	20.25	129.9	5.1	26.0
5	5	67	0.5	0.25	81.1	-14.1	198.8
6	7	125	2.5	6.25	105.5	19.5	330.3
7	6	71	1.5	2.25	93.3	-22.3	497.3
8	2	52	-2.5	6.25	44.5	7.5	56.3
9	1	25	-3.5	12.25	32.3	-7.3	53.3
10	3	70	-1.5	2.25	56.7	13.3	176.3
Сума	45	750	0	76.5		0	1927.1
Середнє	4.5	75					

Інтепретація результату. У 95 випадках із 100 витрати на ремонт швейної машини чотирирічного віку будуть знаходитися в межах від \$ 30.9 до \$ 106.9.
Аналогічно будуємо довірчий інтервал для точки $x_0 = 10$.

$$y_{x_0=10} = 12.2 \cdot 10 + 20.1 = 142.1$$

95% довірчий інтервал для $x_0 = 10$

$$142,1 \pm 44.1; \quad [98; 186.2].$$

Іншими словами, у 95 випадках із 100 витрати на ремонт швейної машини віку 10 років будуть знаходитися в межах від \$ 98 до \$ 186.2.

Приклад 2 (продовження). Знайти оцінку середніх витрат на ремонт швейної машини у віці $x_0 = 4$ і $x_0 = 10$ років.

Відповідно до формули (10) для $x_0 = 4$ маємо

$$68.9 \pm 2.306 \cdot 15.52\sqrt{1 + \frac{(4-4.5)^2}{76.5}},$$

$$68.9 \pm 35.8; \quad [33.1; 104.7],$$

тобто, у 95 випадках із 100 середні витрати на ремонт швейної машини віку $x_0 = 4$ роки будуть знаходитися в межах \$ 33.1 до \$104.7.

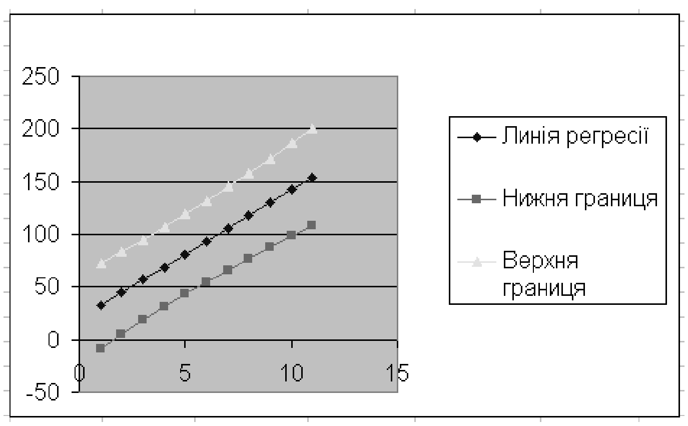
Для точки $x_0 = 10$ маємо

$$142.1 \pm 42.3; \quad [99.8; 184.4].$$

На мал. 4 зображені вибіркова лінія регресії

$$y' = 12.2x + 20.1$$

і довірчий коридор для середніх витрат на ремонт швейної машини.



Мал. 4

6. Оцінка якості вибіркової лінії регресії

6.1. Знаходження коефіцієнта детермінації

Коефіцієнт детермінації визначається за формулою

$$R^2 = \frac{\Sigma(y' - \overline{y})^2}{\Sigma(y - \overline{y})^2}. \quad (11)$$

Значення R^2, близькі до 1, говорять про те, що точки діаграми розсіювання достатньо "щільно" групуються поблизу вибіркової лінії регресії, тобто вибіркова лінія регресії адекватно відображає реальну залежність.

Приклад 2 (продовження). Необхідні дані для обчислення R^2 за формулою (11) наведені в таблиці

Таблиця 5

No	x	y	y'	$y' - \overline{y}$	$(y - \overline{y})^2$	$(y' - \overline{y})^2$	$(y - \overline{y})^2$
1	8	109	117.7	42.7	34	1823	1156
2	3	75	56.7	-18.3	0	335	0
3	1	21	32.3	-42.7	-54	1823	2916
4	3	135	129.9	54.3	60	3014	3600
5	5	67	81.1	6.1	-8	37	64
6	7	125	105.5	30.5	50	930	2510
7	6	71	93.3	18.3	-4	335	16
8	2	52	44.5	-30.5	-23	930	524
9	1	25	32.3	-42.7	-50	1823	2500
10	3	70	56.7	-18.3	-5	335	25
Сума							
Середнє						11 386	13 306

$$R^2 = \frac{11386}{13306} = 0,8557.$$

Висновок 1. *Точки диаграми розсіювання достатньо щільно групуються біля лінії регресії* $y = 12.2x + 20.1$.

6.2. Перевірка вибіркової лінії регресії на адекватність із використанням критерію Фішера

Якщо коефіцієнт детермінації R^2 близький до одиниці, то вибіркова лінія регресії адекватно відображає реальний процес.

Якщо коефіцієнт детермінації R^2 близький до нуля, то вибіркова лінія дисперсії неадекватна істинній залежності.

Якщо ж R^2 приймає знаки посередині інтервалу, наприклад, $0.4\ -0.5$, то може знадобитися додаткова перевірка.

Першим кроком є упорядкування дисперсійної таблиці вигляду (таблиця ANOVA)

Таблиця 6

Джерело варіації	Сума квадратів (*SS*)	Кількисть ступенів свободи (*df*)	Середнє квадратів (*MS*)	Статистика (*F*)
Регресія	SSR	**1**	$MSR = SSR/1$	
Помилки	SSE	$n-2$	$MSE = SSE/(n-2)$	$F = \frac{MSR}{MSE}$
Загальна сума	SST	$n-1$		

Тут

$$SST = \Sigma(y - \overline{y})^2;$$
$$SSR = \Sigma(y' - \overline{y})^2;$$
$$SSE = \Sigma(y - y')^2.$$

Приклад 2 (продовження). Для упорядкування таблиці скористаємося числовими даними з таблиці 5.

Таблиця 7

Джерело	SS	*df*	MS	F
Регресія	11 386	1	11 386	42.27
Помилки	1927	8	240.9	
Загальна сума	13 313	9		

Алгоритм перевірки адекватності вибіркової регресійної моделі розбиваємо на окремі кроки

1. Гіпотези:

H_0 : не існує лінійного зв'язку між y і x $\quad (\beta = 0)$.

H_1 : є лінійний зв'язок між y і x $\quad (\beta \neq 0)$.

2. Статистика для теста. Використовуємо статистику

$$F = \frac{MSR}{MSE}.$$

Якщо H_0 вірна і допущення приймається, то F має розподіл Фішера з 1 і 8 ступенями свободи.

3. Рівень значимості приймемо $\alpha = 0.01$.

4. Правило прийняття рішення. За таблицею знаходимо

$$F(1, 8, \alpha) = 11.26.$$

Якщо обчислене значення $F > 11,26$, то H_0 відхиляється.

5. Обчислення. Див. таблицю 7.

6. Статистичне рішення: Оскільки

$$47.27 \geqslant 11.26,$$

ми відхиляємо гіпотезу H_0.

Висновок 2. *Ми переконуємося, що дані вибірки з високим ступенем ймовірності підтверджує наявність лінійного зв'язку між x і y.*

7. Задачі для самостійного розв'язання

Сгладжування даних і прогнозування на основі ковзких середніх

Побудуйте прогноз на 13 і 14 місяці с базою k.

1. Грошові надходження фірми (у тис. гривень) за останні 12 місяців склали.

місяць	1	2	3	4	5	6
надходження	24	27	16	23	21	44
місяць	7	8	9	10	11	12
надходження	48	32	12	37	38	31

База $k = 3$.

2. Роздрібна ціна на виріб за 12 місяців складала

місяць	1	2	3	4	5	6
ціна	1,45	1,19	2,29	1,65	2,49	4,79
місяць	7	8	9	10	11	12
ціна	3,75	1,59	1,99	3.79	4,25	2,89

База $k = 5$.

3. Середня зарплата пана N за 12 місяців складала

місяць	1	2	3	4	5	6
зарплата	818	942	963	893	819	941
місяць	7	8	9	10	11	12
зарплата	935	865	840	937	954	946

База $k = 7$.

Експоненційне сгладжування та прогнозування

Побудуйте прогноз рівня продажів на 7 місяць для заданого α.

4. За перші півроку фірма здійснила наступну кількість продажів

місяць	1	2	3	4	5	6
продажі	288	247	117	227	239	101

Прийняти $\alpha = 0,2$.

5. Ризик компанії, що здійснює інвестиції, оцінюється по 10 бальній шкалі. Для останніх 6 проектів він склав

місяць	1	2	3	4	5	6
ризик	3	4	9	6	1	7

Прийняти $\alpha = 0,1$.

6. Чистий прибуток компанії N за 6 місяців задає таблиця

місяць	1	2	3	4	5	6
прибуток	83	41	29	32	1	70

Прийняти $\alpha = 0,2$.

Регресійний аналіз

В наступних задачах:

a) Знайдіть коефіцієнти лінії регресії і проведіть її економічний аналіз.

b) Для значень x_{01} і x_{02} дайте прогноз залежної змінної і побудуйте 95% довірчий інтервал для неї. Дайте практичну інтепретацію результатів.

c) Знайдіть коефіцієнт детермінації. Складіть дисперсійну таблицю. Перевірте побудовану модель на адекватність для рівня значимості $\alpha = 0,05$, використовуючи критерій Фішера. Дайте практичну інтепретацію результатів.

7. Є вибірка даних (x, y), де x - вік працюючого (роки), y - його річна зарплата (у тисячах гривень)

вік (x)	29	48	41	54	44	57
зарплата (y)	24	27	16	23	21	44

Точки для прогнозу: $x_{01} = 50$ і $x_{02} = 60$.

8. Рух готівки на фірмі та зміна активів фірми задає таблиця

активи (x)	8,3	3,2	12	2,7	4,3	2,1
готівка (y)	1219	100	65	2073	338	173

Точки для прогнозу: $x_{01} = 5$ і $x_{02} = 13.$

9. Кількість робітників (сотні) на фірмі та чистий прибуток (мільони) задає таблиця

кільк. робітників (x)	2,4	13,2	2,4	26,2	26,2	5,3
прибуток (y)	152	116	390	102	135	47

Точки для прогнозу: $x_{01} = 10$ і $x_{02} = 2.$

Література

1. Доугерти К., Введение в эконометрику, Москва, Инфра-М, 1997, 404 стор.

2. Лук'яненко І.Г., Краснікова Л.І., Економетрика: Підручник, Київ, Товариство Знання, КОО, 1988, 361 стор.

3.Лук'яненко І.Г., Краснікова Л.І.,Економетрика, Практикум з використанням компютера, Київ, Товариство Знання, КОО, 1988, 218 стор.

4. Daniel W.W., Essentials of business statistics, Boston and others, Houghton Mifflin Company, 1988,500 p.

5. Економетрична сторінка Олександра Циплакова http://www.nsu.ru/ef/tsy/ecmr/index.htm

6. Носко В.П., Эконометрика для начинающих, Москва, Институт єкономики переходного периода, 2000, 267 стр.(Доступна на сайте http://www.iet.ru/archiv/zip/nosko.zip).

7. Наконечний С.І., Терещенко Т.О., Романюк Т.П., Економетрія: Підручник, Київ, КНЕУ, 2000.