

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

**Методичні рекомендації
до виконання лабораторних робіт
з навчальної дисципліни
"ПРОГНОЗУВАННЯ
СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ"
для студентів напряму підготовки
6.030502 "Економічна кібернетика"
денної форми навчання**

Харків. ХНЕУ ім. С. Кузнеця, 2015

Затверджено на засіданні кафедри економічної кібернетики.
Протокол № 8 від 15.01.2015 р.

Самостійне електронне текстове мережне видання

Укладачі: Л. С. Гур'янова
О. А. Сергієнко

М 54 Методичні рекомендації до виконання лабораторних робіт з навчальної дисципліни "Прогнозування соціально-економічних процесів" для студентів напряму підготовки 6.030502 "Економічна кібернетика" денної форми навчання : [Електронне видання] / уклад. Л. С. Гур'янова, О. А. Сергієнко. – Х. : ХНЕУ ім. С. Кузнеця, 2015. – 130 с. (Укр. мов.)

Подано плани лабораторних занять із навчальної дисципліни, визначено мету й засоби виконання завдань до лабораторних робіт.

Рекомендовано для студентів напряму підготовки 6.030502 "Економічна кібернетика".

Вступ

В умовах ринкової економіки істотно змінюються інформаційні запити керуючих структур за обсягом, складом, вірогідністю й оперативністю інформації. У зв'язку з цим для керівників різних рівнів зростає роль прогнозів у прийнятті обґрунтованих управлінських рішень. Методи соціально-економічного прогнозування стають важливим інструментом у діяльності планових, аналітичних, маркетингових відділів виробничих підприємств і об'єднань, торговельних, страхових компаній, банківських, урядових закладів.

Процедура побудови моделі прогнозування досить складна та вимагає значних розрахунків, проведення численних обчислювальних експериментів. Тому розробка соціально-економічних прогнозів неможлива без використання сучасних пакетів програм, що реалізують відповідні технології. Метою лабораторних робіт із дисципліни "Прогнозування соціально-економічних процесів" є закріплення теоретичного та практичного матеріалу, набуття навичок роботи з пакетами прикладних програм, що підтримують реалізацію різних підходів до побудови прогнозів у соціально-економічних системах. Для успішного виконання лабораторних робіт до кожного завдання даються методичні вказівки.

Змістовий модуль 1

Методи та моделі короткострокового прогнозування

Лабораторна робота 1

Критерії відповідності для перевірки гіпотез про нормальний закон розподілу

Мета – формування у студентів навичок застосування критеріїв відповідності для перевірки гіпотез про нормальний закон розподілу, придбання навичок роботи в модулі *Basic Statistics/Tables*.

Умови завдання

Необхідно провести аналіз статистичних характеристик рядів у модулі *Basic Statistics/Tables* ППП *Statistica*:

1. Розрахувати статистичні характеристики ряду (середнє, дисперсію, середнє квадратичне відхилення, моду, медіану, коефіцієнти асиметрії та ексцесу).

2. Побудувати гістограму та полігон розподілу випадкової величини, зробити висновки щодо характеру закону розподілу.

3. За допомогою критеріїв Пірсона та Колмогорова – Смірнова перевірити гіпотезу про нормальний закон розподілу.

Методичні рекомендації до виконання завдання

Для рішення й аналізу задач даного типу в ППП *Statistica* передбачений модуль *Basic Statistics/Tables* (Основні статистики й таблиці). Слід розглянути порядок роботи в даному модулі.

1. Запуск *Statistica* і підготовка даних.

У меню програм необхідно вибрати програму *Statistica*, після її запуску вибрати у меню пункт *File/New* для підготовки власних даних. З'явиться діалогове вікно, у якому необхідно вказати кількість змінних (*Number of variables*) і кількість випадків (*Number of Cases*). Після введення натиснути кнопку вікна *OK* (рис. 1.1).

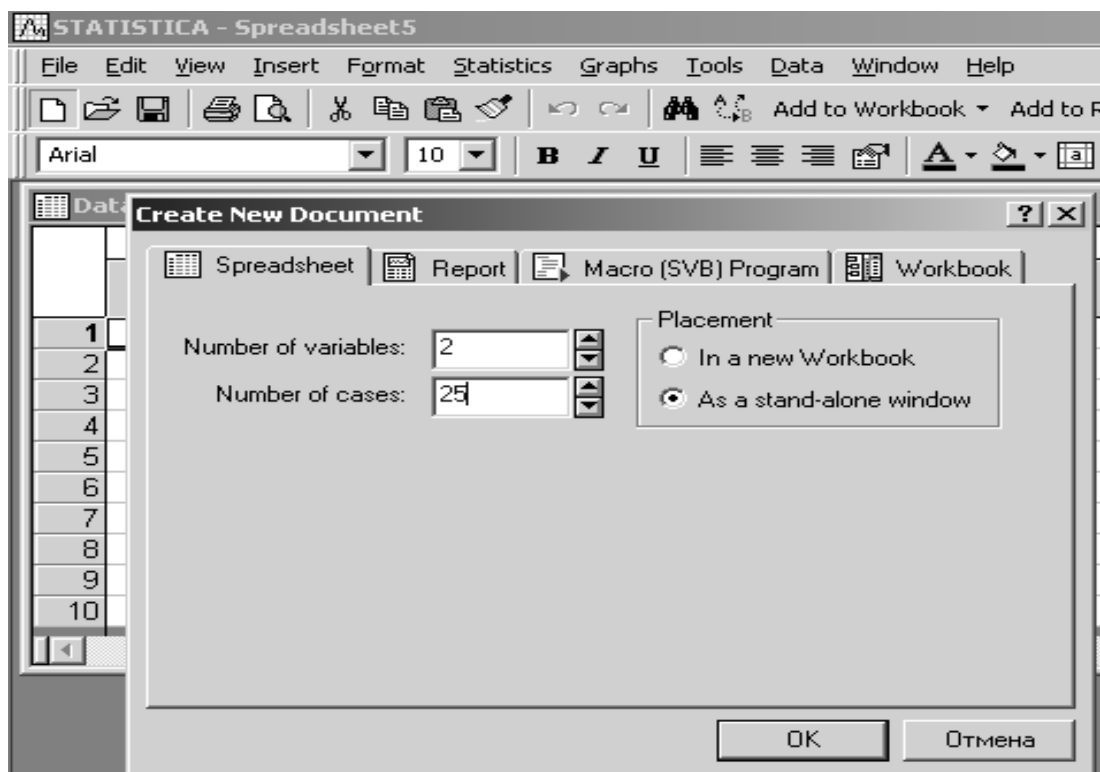


Рис. 1.1. Визначення кількості змінних і спостережень

З'явиться порожнє поле, що становить таблицю розміром 25*2: 25 спостережень, 2 змінні (рис. 1.2).

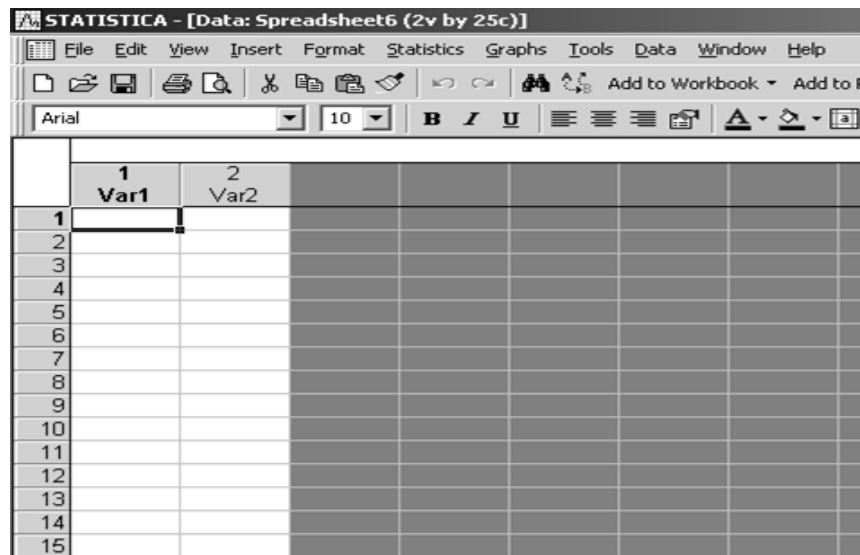


Рис. 1.2. Поле даних

Кожен елемент даних, тобто значення показника, займає одну комірку поля даних. Після заповнення всіх комірок поля даних буде одержано таблицю, наведену на рис. 1.3.

2		
% dohod		
1187477		
793821		
876148		
389719		
459234		
451074		
328131		
209010		
273945		
167741		
232158		
175292		
111185		
70674		
145468		
132243		
120243		
100261		
104326		
114054		
42602		
35241		
71296		
91436		
86384		

Рис. 1.3. Вихідні дані

2. Розрахунки.

Слід розрахувати основні статистичні характеристики ряду (середнє, дисперсію, середнє квадратичне відхилення, моду, медіану, розмах, коефіцієнти асиметрії й ексцесу).

Щоб почати обчислювальні процедури, необхідно увійти в позицію меню *Statistics/Basic Statistics/Tables* (рис. 1.4). Після підтвердження вибору модуля з'явиться діалогове вікно, що дозволяє задати напрямок аналізу *Descriptive statistics* (Описові статистики), подане на рис. 1.5.

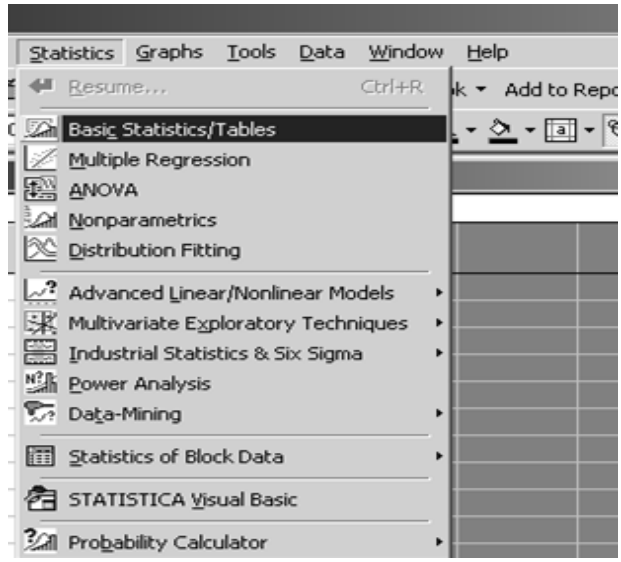


Рис. 1.4. Вибір модуля

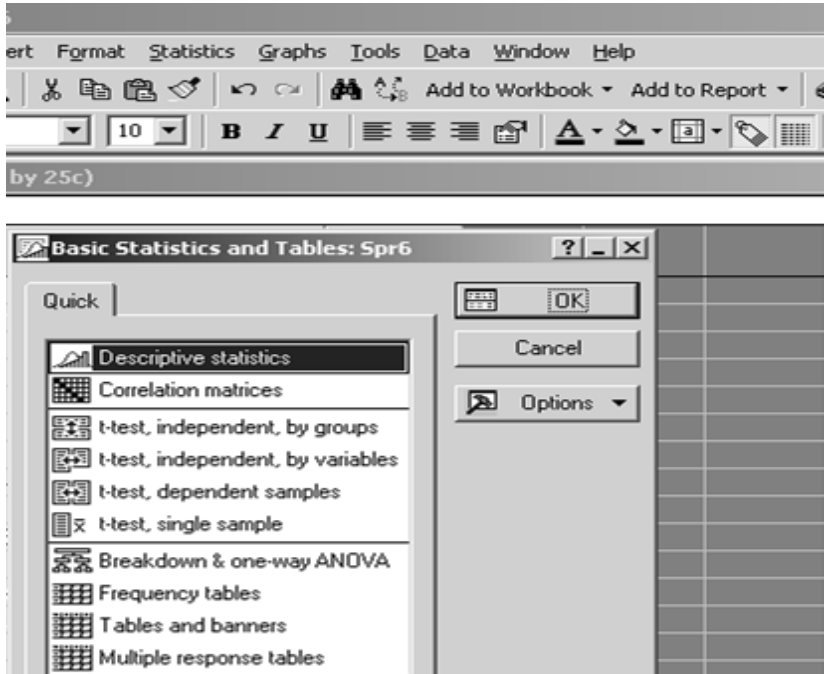


Рис. 1.5. Вибір напрямку аналізу

Після вибору напрямку аналізу з'явиться стартова панель модуля, де необхідно задати вихідні параметри: *Variable* (Змінні) і відповідний набір процедур для подальшого аналізу (рис. 1.6).

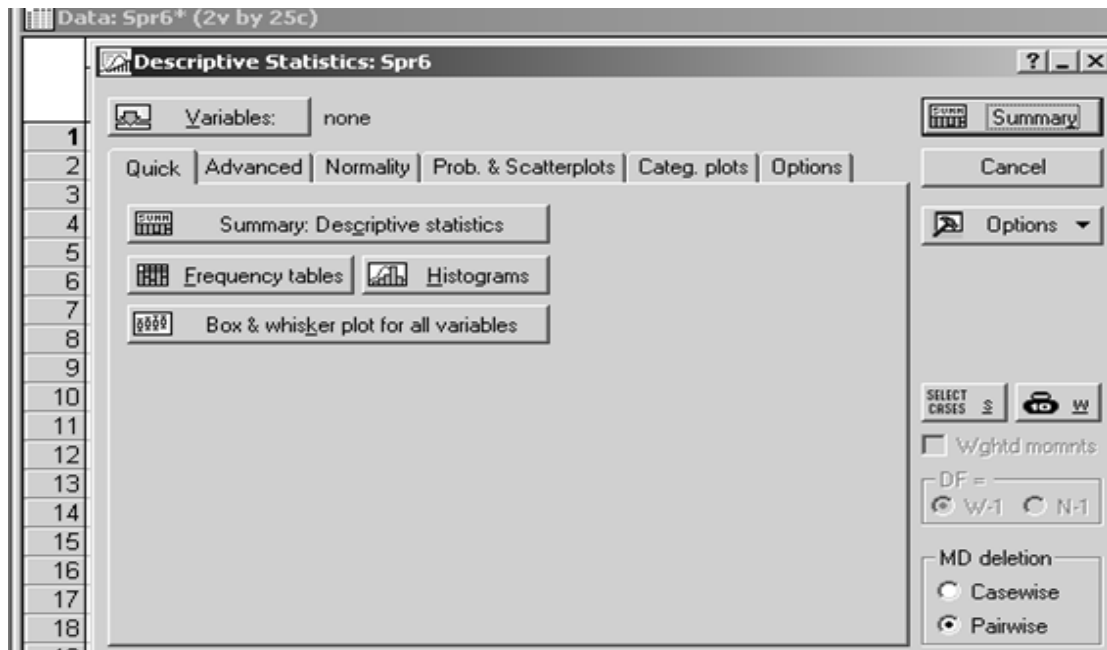


Рис. 1.6. Стартова панель модуля

Слід ініціювати кнопку *Variable* (Змінні) і у вікні, що з'явилося, указати показники, за якими здійснюється аналіз. Після зазначення змінних необхідно підтвердити свій вибір натисканням кнопки *OK* (рис. 1.7).

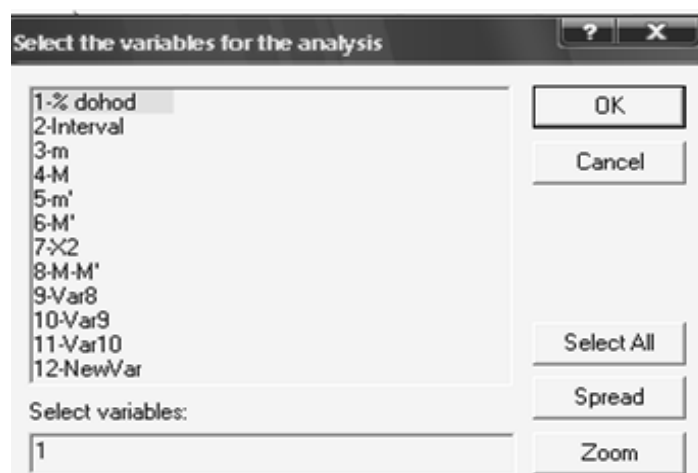


Рис. 1.7. Вибір змінних для аналізу

Далі, ініціювавши вкладиш *Advanced*, необхідно виділити основні статистики для розрахунку (рис. 1.8).

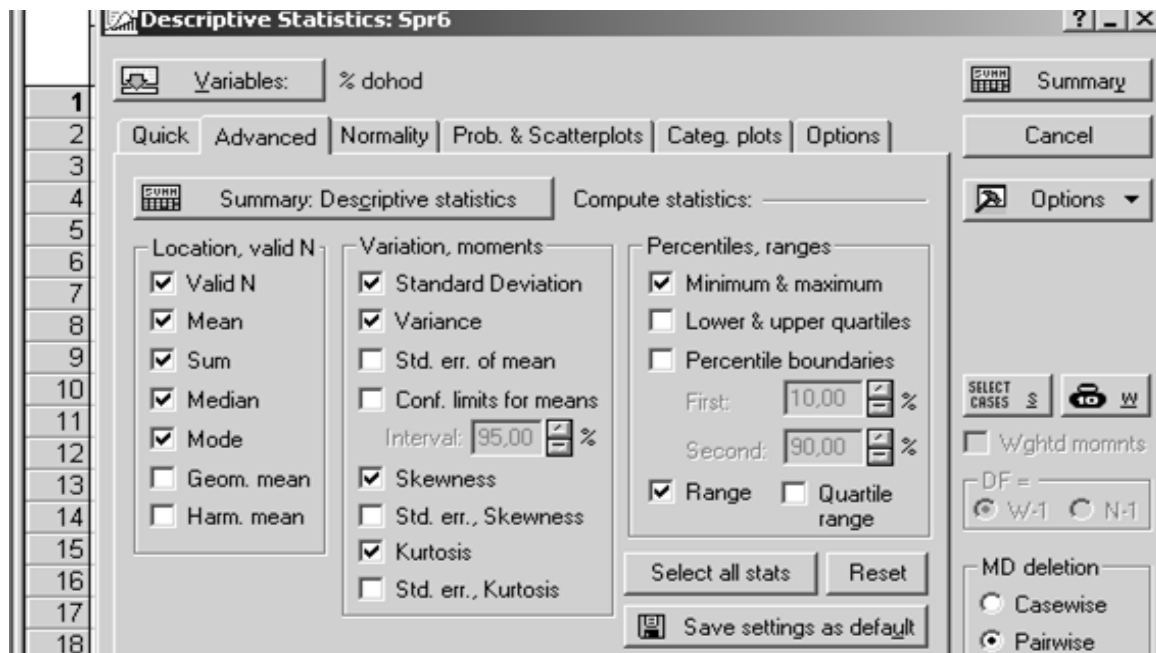


Рис. 1.8. Вибір описових статистик

Такими є: *Valid N* (число спостережень), *Mean* (середнє), *Sum* (сума значень), *Median* (медіана), *Mode* (мода), *Standard Deviation* (середнє квадратичне відхилення), *Variance* (дисперсія), *Skewness* (коефіцієнт асиметрії), *Kurtosis* (коефіцієнт ексцесу), *Min & Max* (мінімум і максимум). Результати розрахунку описових статистик для даної вибірки будуть одержані натисканням клавіші *Summary*. Вони наведені на рис. 1.9.

Descriptive Statistics (Spr6)						
Variable	Valid N	Mean	Median	Sum	Minimum	Maximum
% dohod	25	270766,5	145468,0	6769163	35241,00	1187477

Variance	Std.Dev.	Skewness	Kurtosis
8,370020E+10	289309,9	1,999460	3,715184

Рис. 1.9. Описові статистики

3. Графічне подання вибірових даних та їх групування.

Слід побудувати гістограму й полігон розподілу випадкової величини та провести угруповання вибірки.

Для наочності подання досліджуваної сукупності слід побудувати полігон розподілу. Для цього необхідно зайти в меню *Graphs/2D Graphs/Scaterplots* (рис. 1.10), вибрати змінні, задати параметри графіка (рис. 1.11) і побудувати полігон розподілу випадкової величини (рис. 1.12).

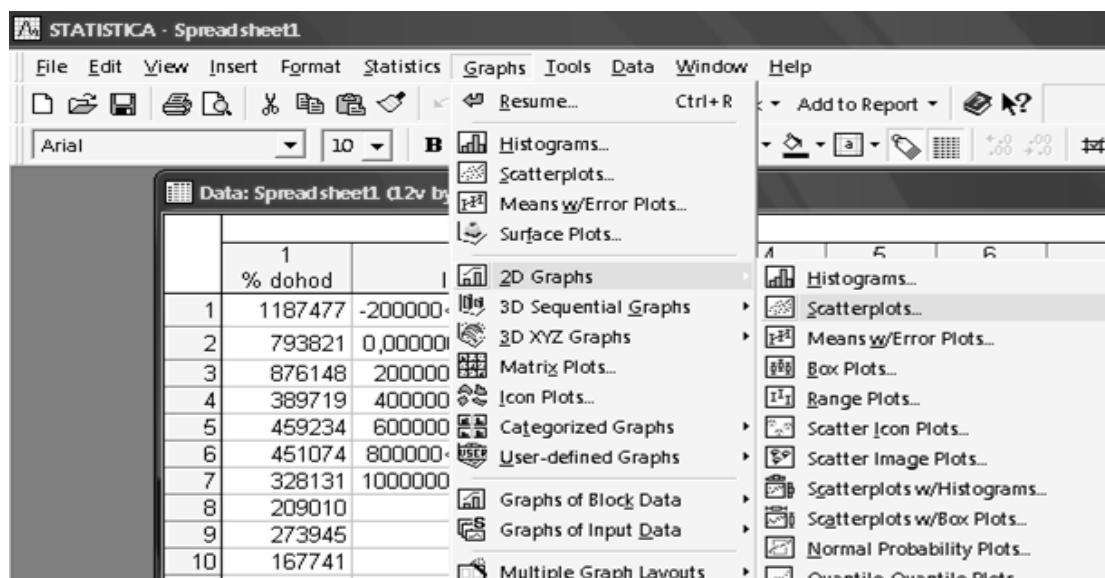


Рис. 1.10. Вибір типу графіка

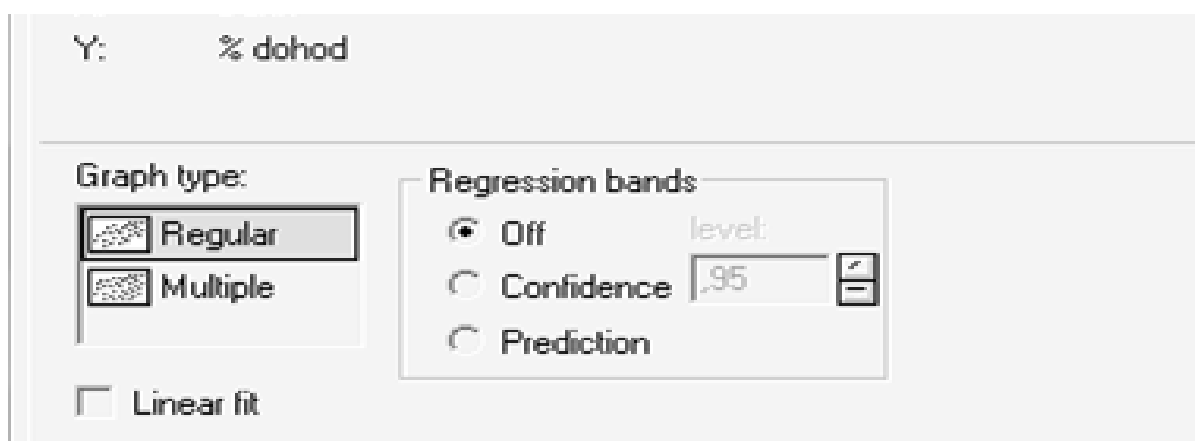


Рис. 1.11. Вибір параметрів графіка

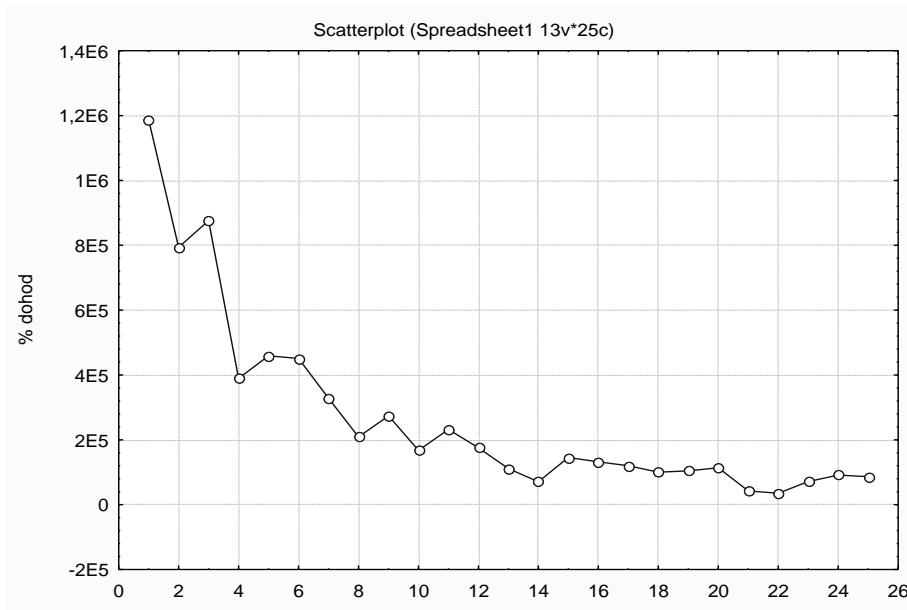


Рис. 1.12. Полігон розподілу випадкової величини

Подальший аналіз здійснюється в рамках перевірки вибірки на нормальний закон розподілу. Для проведення угруповання вибірки на стартовій панелі модуля вибрати вкладку *Normality*, де можна задавати бажану кількість інтервалів і критерій Колмогорова – Смірнова для тестування вибірки (рис. 1.13).

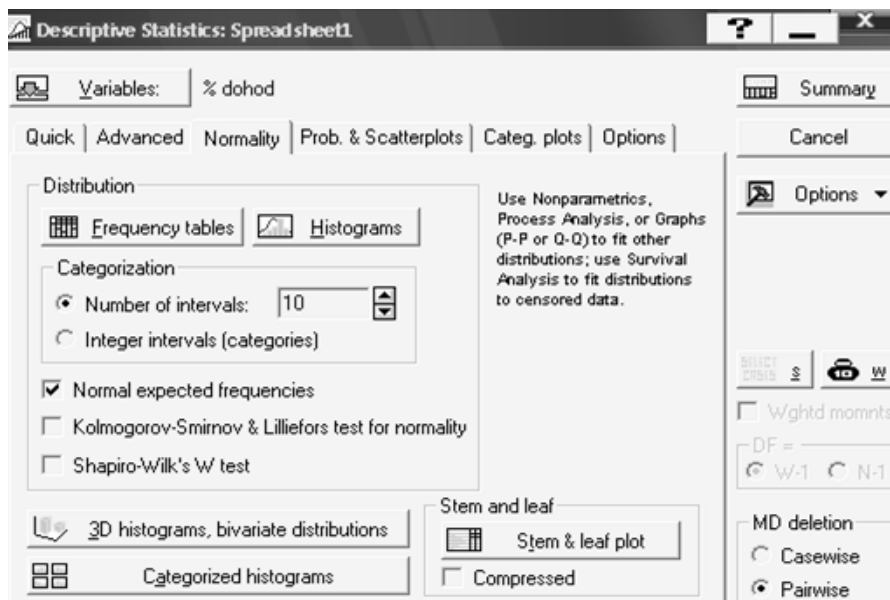


Рис. 1.13. Вибір параметрів угруповання випадкової величини

Ініціювавши клавішу *Frequency tables (Таблиці частот)*, буде одержано таку таблицю (рис. 1.14).

Frequency table: % dohod (Spr6) K-S d=,23308, p<,15 ; Lilliefors p<,01						
Category	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
-200000,<x<=0,000000	0	0	0,00000	0,0000	0,00000	0,0000
0,000000<x<=200000,0	15	15	60,00000	60,0000	60,00000	60,0000
200000,0<x<=400000,0	5	20	20,00000	80,0000	20,00000	80,0000
400000,0<x<=600000,0	2	22	8,00000	88,0000	8,00000	88,0000
600000,0<x<=800000,0	1	23	4,00000	92,0000	4,00000	92,0000
800000,0<x<=1000000,	1	24	4,00000	96,0000	4,00000	96,0000
1000000,<x<=1200000,	1	25	4,00000	100,0000	4,00000	100,0000
Missing	0	25	0,00000		0,00000	100,0000

%	Expected Count	Cumulative Expected	Percent Expected	Cumulative % Expected
00	4,366527	4,36653	17,46611	17,46611
00	5,718006	10,08453	22,87202	40,33813
00	6,726783	16,81132	26,90713	67,24526
00	4,999658	21,81097	19,99863	87,24389
00	2,347086	24,15806	9,38835	96,63224
00	0,695495	24,85355	2,78198	99,41422
00	0,129962	24,98352	0,51985	99,93407
00				

Рис. 1.14. Результат угруповання вибірки

Як видно, вихідна сукупність розбита на 7 інтервалів, у кожному інтервалі розраховані такі характеристики: *Count* (частота), *Cumulative Count* (накопичена частота), *Percent of Valid* (% від загальної частоти), *Cumul % of Valid* (накопичений % від загальної частоти), *% of all Cases* (% від загального числа спостережень), *Cumulative % of all Cases* (накопичений % від загального числа спостережень), *Expected Count* (теоретична частота), *Cumulative Expected* (накопичена теоретична частота), *% Expected* (% від загальної теоретичної частоти), *Cumulative % Expected* (накопичений % від загальної теоретичної частоти).

Ініціювавши клавішу *Histograms* (вкладиш *Normality*), буде одержано таку гістограму розподілу з накладеною кривою нормального закону розподілу (рис. 1.15).

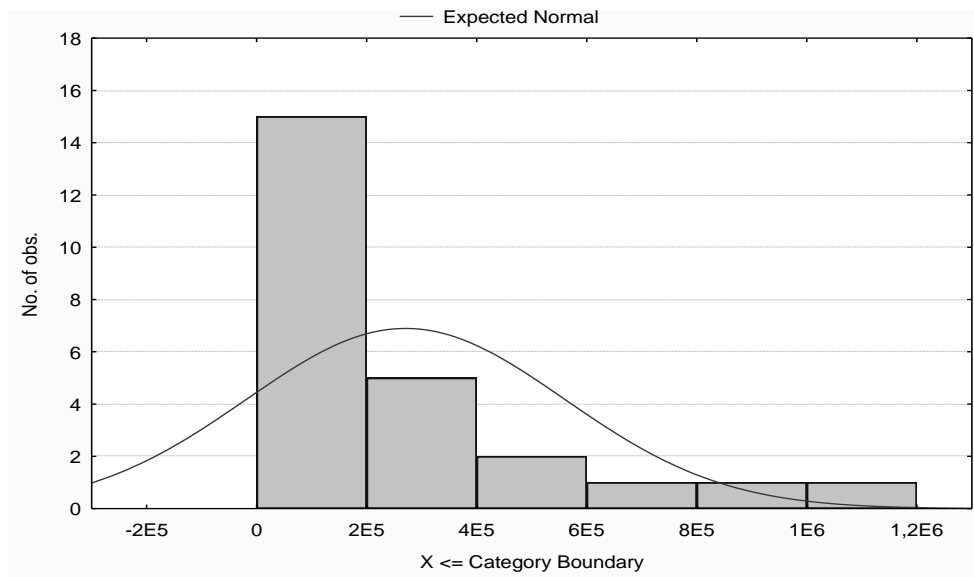


Рис. 1.15. Гістограма розподілу

4. Аналіз характеру закону розподілу випадкової величини.

Подальший аналіз вибірки передбачає розрахунок критерію Пірсона та Колмогорова – Смірнова для формування відповідних висновків про характер закону розподілу. Для визначення характеру закону розподілу та його відповідності нормальному закону необхідно дослідити за допомогою графіків порівняння емпіричних та теоретичних частот і накопичених частот. Вихідні дані для побудови графіків та розраховані значення критерію Пірсона та Колмогорова – Смірнова наведені на рис. 1.16. Для побудови графіків інтервальних значень частоти розподілу досліджуваної сукупності необхідно зайти в меню *Graphs/2D Graphs/Scatterplots*, вибрати змінні та задати параметри графіка (рис. 1.17). На рис. 1.18 та рис. 1.19 наведені графіки порівняння емпіричних та теоретичних частот та накопичених емпіричних та теоретичних частот, які дозволяють зробити висновки про відповідність нормальному закону розподілу й визначити розбіжність частот у кожному із досліджуваних інтервалів.

	1	2	3	4	5	6	7	8
	% dohod	Interval	m	M	m'	M	X ²	M-M
1	1187477	-200000<x<=0,000000	0	0	4,366527	4,36653	4,36653	-4,36653
2	793821	0,000000<x<=200000	15	15	5,718006	10,08453	15,06739	4,91547
3	876148	200000<x<=400000	5	20	6,726783	16,81132	0,44327	3,18868
4	389719	400000<x<=600000	2	22	4,999658	21,81097	1,79971	0,18903
5	459234	600000<x<=800000	1	23	2,347086	24,15806	0,77315	-1,15806
6	451074	800000<x<=1000000	1	24	0,695495	24,85355	0,13332	-0,85355
7	328131	1000000<x<=1200000	1	25	0,129962	24,98352	5,82450	0,01648
8	209010					X ²	28,4078616	
9	273945						Kolmogorov	0,98309

Рис. 1.16. Аналіз закону розподілу випадкової величини

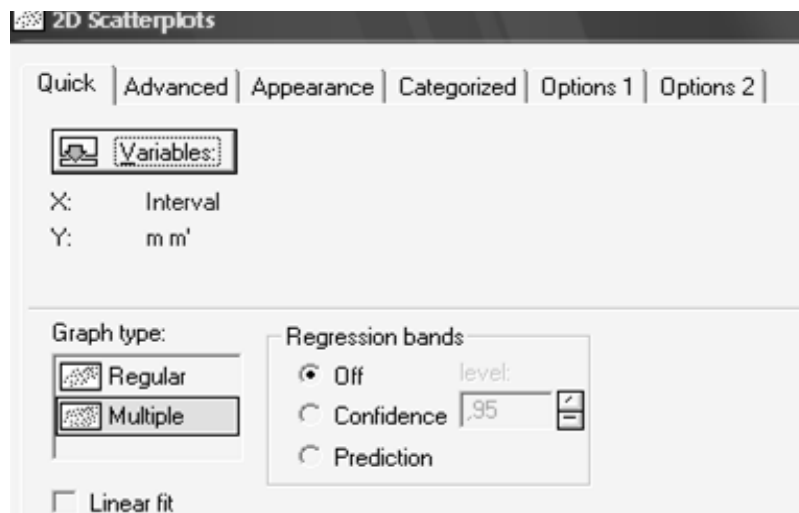


Рис. 1.17. Вибір змінних для побудови графіків

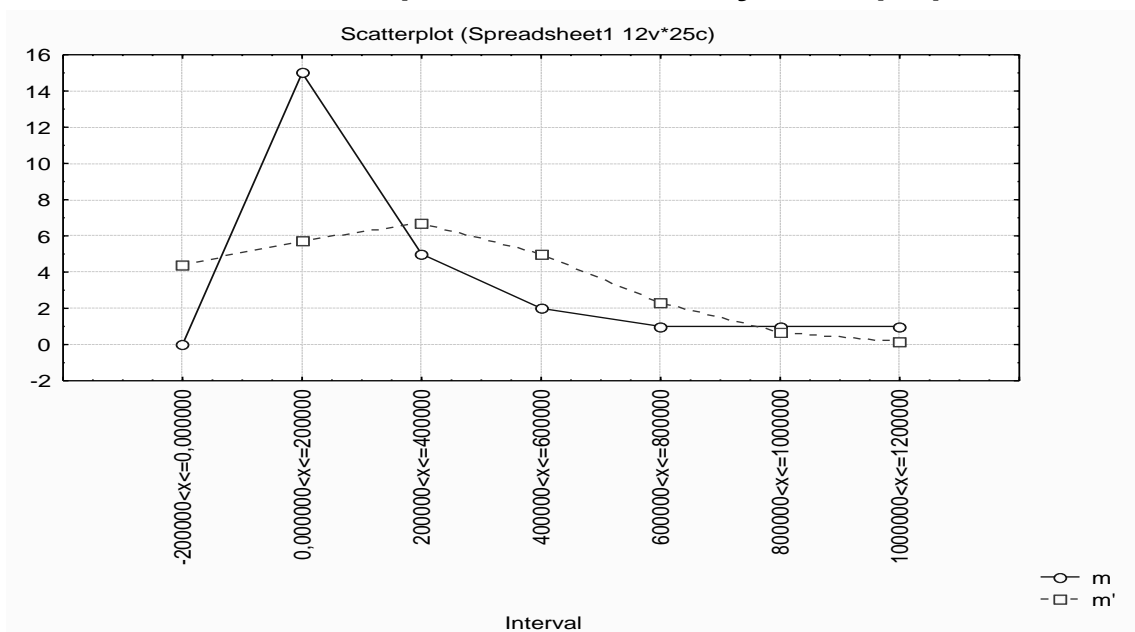


Рис. 1.18. Графік порівняння емпіричних та теоретичних частот

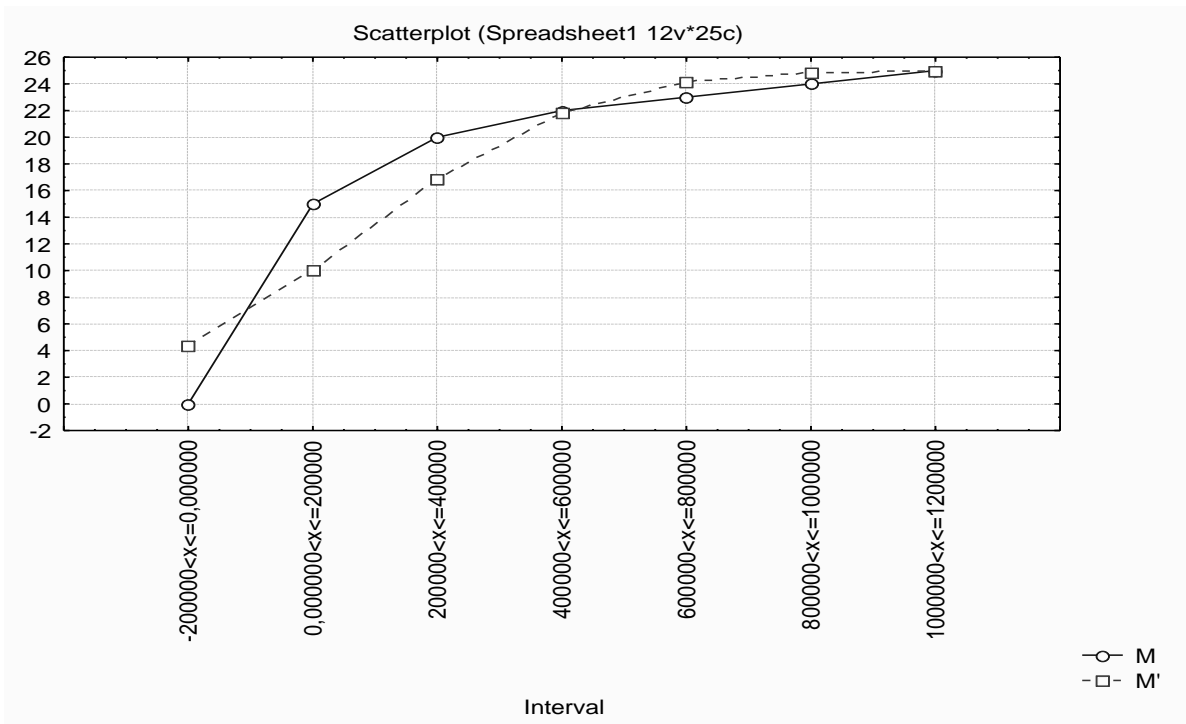


Рис. 1.19. **Графік порівняння накопичених емпіричних та теоретичних частот**

Слід зробити висновки про угруповання даних за величиною показника % доходу. Порівняти отримані значення з табличними за відповідними критеріями та зробити висновки про характер закону розподілу.

Лабораторна робота 2 Оцінювання якості прогнозів

Мета – формування у студентів навичок застосування критеріїв оцінювання якості прогнозів та проведення порівняльного аналізу альтернативних варіантів моделі прогнозування за допомогою програми *Microsoft Excel*.

Умови завдання

У табл. 2.1 наведені фактичні дані індексу зниження собівартості продукції (y_j); розрахункові дані, отримані на основі моделей 1 (y_{1j}) і 2 (y_{2j}), що містять різний набір факторів, які впливають на індекс собівартості продукції.

Вихідні дані

№ періоду	1	2	3	4	5	6	7	8	9	10	11	12
y_j	13.6	18.1	32.3	30.1	46.3	46.6	53.1	53.2	52.6	56.5	62.5	62
Значення моделі 1, y_{1j}	12.87	38.10	37.06	30.10	39.93	41.80	40.98	52.73	53.39	65.09	57.80	57.06
Значення моделі 2, y_{2j}	12.77	38.44	37.38	29.44	38.53	41.69	42.67	54.81	54.94	63.94	55.74	56.54

Необхідно розрахувати міри точності прогнозу. Провести порівняльний аналіз, здійснити вибір моделі для прогнозування індексу зниження собівартості продукції.

Методичні рекомендації до виконання завдання

Слід розглянути основні характеристики, що використовуються для оцінювання якості прогнозу. Першу групу утворюють абсолютні показники похибки прогнозу. Вони дозволяють кількісно визначити величину розбіжності прогнозу та факту спостереження в одиницях вимірюваної змінної. Необхідно розрізняти такі характеристики:

– похибка прогнозу:

$$e_j = y_j - y_j,$$

де y_j – фактичне значення змінної для j -го виміру;

y_j – прогнозне значення досліджуваної змінної;

– абсолютна похибка прогнозу:

$$\Delta_j = y_j - y_j ;$$

– середня абсолютна похибка прогнозу (*mean absolute error*):

$$MAE = \frac{\sum_{j=1}^N |y_j - y_j|}{N};$$

– середнє квадратичне відхилення (*root mean squared error*):

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (y_j - y_j)^2}{N}};$$

– середній квадрат похибки (*mean square error*):

$$MSE = \frac{1}{N} \sum_{j=1}^N e_j^2 ;$$

– сума квадратів похибки (*sum square error*):

$$SSE = \sum_{j=1}^N e_j^2.$$

Найчастіше останні два показники використовуються у процесі вибору оптимального методу прогнозування. У більшості статистичних пакетів дані показники використовуються як критерії вибору моделі.

Необхідно виконати розрахунок показників якості для наведеного прикладу. Результати розрахунків подані на рис. 2.1.

	A	B	C	D	E	F	G	H	I	J	K	L
	№ періоду	y_j	Значення моделі 1, \hat{y}_{1j}	Значення моделі 2, \hat{y}_{2j}	$\Delta_1 = y_j - \hat{y}_{1j} $	$\Delta_2 = y_j - \hat{y}_{2j} $	$(y_j - \hat{y}_{1j})^2$	$(y_j - \hat{y}_{2j})^2$			Модель 1	Модель 2
1	1	13.6	12.87	12.77	0.73	0.83	0.53	0.6889			M.A.E.	
2	2	18.1	38.1	38.44	20.00	20.34	400.00	413.716			5.69	6.14
3	3	32.3	37.06	37.38	4.76	5.08	22.66	25.8064			RMSE	
4	4	30.1	30.1	29.44	0.00	0.66	0.00	0.4356			7.93	8.03
5	5	46.3	39.93	38.53	6.37	7.77	40.58	60.3729			MSE	
6	6	46.6	41.8	41.69	4.80	4.91	23.04	24.1081			62.9	64.4
7	7	53.1	40.98	42.67	12.12	10.43	146.89	108.785			SSE	
8	8	53.2	52.73	54.81	0.47	1.61	0.22	2.5921			754.83	772.84
9	9	52.6	53.39	54.94	0.79	2.34	0.62	5.4756				
10	10	56.5	65.09	63.94	8.59	7.44	73.79	55.3536				
11	11	62.5	57.8	55.74	4.70	6.76	22.09	45.6976				
12	12	62	57.06	56.54	4.94	5.46	24.40	29.8116				

Рис. 2.1. Результати розрахунків показників якості

Розрахунки проводяться в пакеті *Excel* за допомогою вбудованих функцій:

ABS () – функція, що повертає модуль (абсолютне значення) числа;

СРЗНАЧ () – функція, що розраховує середнє арифметичне значення;

СУММ () – функція, що розраховує суму.

Виходячи з наведених на рис. 2.1 розрахунків можна зробити висновки, що перша модель є більш прийнятною для опису досліджуваної сукупності.

Значення всіх перерахованих показників залежать від масштабу вимірів, що у ряді випадків, зокрема під час міжоб'єктних зіставлень, зменшує об'єктивність оцінок. Для того, щоб уникнути цього, використовують відносні показники виміру похибки прогнозу, виражені або в частках одиниці, або у відсотках. Дані показники утворюють другу групу, відносні похибки прогнозу:

– відносна похибка прогнозу:

$$\varepsilon_j = \frac{y_j - y_j}{y_j} \cdot 100 \%;$$

– середня абсолютна процентна похибка (*mean absolute percentage error*):

$$MAPE = \frac{1}{N} \sum_{j=1}^N \frac{y_j - y_j}{y_j} \cdot 100 \%. \quad (2.1)$$

Показник, як правило, використовується для порівняння точності прогнозів різнорідних об'єктів прогнозування, оскільки в ньому застосовуються відносні величини. Уважається, що прогноз має високу точність, якщо $MAPE < 10 \%$. Прогноз має задовільну точність, якщо значення даного показника знаходиться між 10 і 20 %.

– середня процентна похибка (*mean percentage error*):

$$MAPE = \frac{1}{N} \sum_{j=1}^N \frac{y_j - y_j}{y_j} \cdot 100 \%.$$

MPE характеризує відносний ступінь зміщеності прогнозу. За умови, що втрати у ході прогнозування, пов'язані із завищенням фактичного майбутнього значення, урівноважуються заниженням, ідеальний прогноз повинен бути незміщеним, і обидві міри повинні прагнути до нуля. Середня процентна помилка не визначена за нульових даних і не повинна перевищувати 5 %.

Третю групу утворюють коефіцієнти невідповідності (розбіжності), запропоновані Тейлом. З їх допомогою рівняються дані ретроспективної інформації y_j , використовуваної під час розробки оператора прогнозування, з результатами прогнозування y_j .

$$K_T = \frac{\sum_{j=1}^N (y_j - y_j)^2}{\sum_{j=1}^N y_j^2};$$
$$K_T = \frac{\sum_{j=1}^N (y_j - y_j)^2}{\sum_{j=1}^N y_j^2 + \sum_{j=1}^N y_j^2}.$$

Друга форма коефіцієнта Тейла найпоширеніша на практиці. Якщо не робиться спеціальних застережень, то за замовчуванням використовується саме вона. Іноді для визначення точності прогнозу застосовують порівняльні методи. Так у формулах коефіцієнтів невідповідності Тейла замість значення прогнозного показника – y_j , може стояти його бажане значення, якийсь "еталон".

Прикладом іншого підходу можна назвати розрахунок коефіцієнта кореляції між прогнозованими значеннями показника та фактичними його значеннями:

$$R_{yy} = \frac{cov(y\hat{y})}{\sigma_y \sigma_{\hat{y}}}$$

де $cov(y\hat{y})$ – коефіцієнт коваріації прогнозних і фактичних значень змінних;
 $\sigma_{\hat{y}}$; σ_y – стандартні помилки прогнозного та фактичного рядів.

Однак під час аналізу й інтерпретації останнього показника не слід забувати про специфіку тлумачення значень коефіцієнта парної кореляції.

Результати розрахунків цих критеріїв для наведеного прикладу подані на рис. 2.2.

№ періоду	y_j	Значення моделі 1, \hat{y}_1	Значення моделі 2, \hat{y}_2	$y_j - \hat{y}_1$	$y_j - \hat{y}_2$	$\frac{y_j - \hat{y}_1}{y_j} \cdot 100\%$	$\frac{y_j - \hat{y}_2}{y_j} \cdot 100\%$	$\varepsilon_1 = \frac{ y_j - \hat{y}_1 }{y_j} \cdot 100\%$	$\varepsilon_2 = \frac{ y_j - \hat{y}_2 }{y_j} \cdot 100\%$	Модель 1	Модель 2
1	13.6	12.97	12.77	0.73	0.83	5.37	6.10	5.37	6.10	М.А.Р.Е.	
2	18.1	38.1	38.44	-20	-20.34	-110.50	-112.38	110.50	112.38	17.55	18.64
3	32.3	37.06	37.38	-4.76	-5.08	-14.74	-15.73	14.74	15.73	М.Р.Е.	
4	30.1	30.1	29.44	0	0.66	0.00	2.19	0.00	2.19	-6.11	-6.16
5	46.3	39.93	38.53	6.37	7.77	13.76	16.78	13.76	16.78	K_T	
6	46.6	41.8	41.69	4.8	4.91	10.30	10.54	10.30	10.54	0.121	0.123
7	53.1	40.98	42.67	12.12	10.43	22.82	19.64	22.82	19.64	R_{yy}	
8	53.2	52.73	54.81	0.47	-1.61	0.88	-3.03	0.88	3.03		
9	52.6	53.39	54.94	-0.79	-2.34	-1.50	-4.45	1.50	4.45	0.793	0.790
10	56.5	65.09	63.94	-8.59	-7.44	-15.20	-13.17	15.20	13.17		
11	62.5	57.8	55.74	4.7	6.76	7.52	10.82	7.52	10.82		
12	62	57.06	56.54	4.94	5.46	7.97	8.81	7.97	8.81		

Рис. 2.2. Результати розрахунків показників якості

Розрахунки проводяться в пакеті *Excel* за допомогою вбудованих функцій:

СУММКВ () – функція, що розраховує суму квадратів;

КОВАР (,) – функція, що розраховує коваріацію, середнє попарних добутків відхилень;

СТАНДОТКЛОН () – функція, що розраховує стандартне відхилення;

КОРРЕЛ (,) – функція, що розраховує кореляцію між двома множинами даних.

Виходячи з наведених на рис. 2.1 та рис. 2.2 розрахунків, можна зробити висновок, що перша з розглянутих моделей має більшу точність і є більш кращою у порівнянні з другою моделлю.

Лабораторна робота 3

Перевірка стаціонарності часових рядів

Мета – формування у студентів навичок перевірки часових рядів на наявність тренда на підставі критеріїв Стюдента, Фішера, Фостера-Стюарта, критерію серій за допомогою програми *Microsoft Excel*.

Умови завдання

У табл. 3.1 наведені дані, що відображають динаміку продажів підприємства (y_j).

Таблиця 3.1

Динаміка продажів підприємства, тис. грн

t	Sales (y_j)	t	Sales (y_j)
1	127,96	15	131,7
2	127,9	16	132,49
3	128,37	17	132,45
4	130,1	18	131,72
5	129,66	19	131,38
6	128,79	20	132,37
7	129,83	21	133,51
8	130,23	22	132,53
9	130,22	23	133,13
10	129,66	24	131,92
11	129,52	25	131,39
12	130,43	26	131,15
13	130,58	27	130,95
14	131,54	28	129,38

Необхідно побудувати графік вихідних даних. Провести перевірку часового ряду на наявність тренда на підставі критеріїв Стьюдента, Фішера, Фостера-Стюарта, критерію серій. Порівняти результати оцінювання за різними методами. Зробити висновки щодо стаціонарності часового ряду.

Методичні рекомендації до виконання завдання

Слід розглянути методи перевірки стаціонарності часових рядів.

Часовий ряд називається стаціонарним, якщо в ньому відсутня тенденція розвитку. Це значить, що рівні ряду варіюють навколо середнього рівня, відхилення від якого становлять випадкові коливання. Модель рівня часового ряду в цьому випадку має вигляд:

$$y_t = \mu + \varepsilon_t,$$

де y_t – рівні часового ряду;

μ – середній за період рівень часового ряду;

ε_t – випадкова складова.

Графічно стаціонарний ряд можна подати в такий спосіб (рис. 3.1).

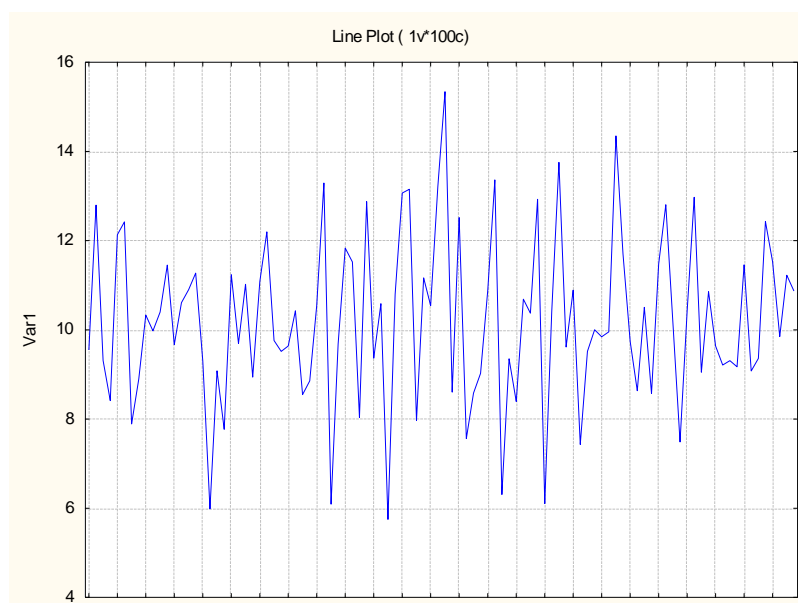


Рис. 3.1. Стаціонарний часовий ряд

Визначення стаціонарності часового ряду проводиться таким чином:
оцінювання корелограми й перевірка гіпотези про стаціонарність ряду за допомогою аналізу кореляційної функції;

використання критерію Стьюдента для перевірки значимості відмінностей оцінок математичних очікувань на різних відрізках часового ряду;

перевірка гіпотези про сталість дисперсії;

перевірка гіпотези про незалежність помилок і ін.

Найпоширенішим у практиці тестування рядів на стаціонарність є використання статистичних гіпотез про відсутність тенденцій за рядом. Якщо формулювати більш строго, то слід перевірити ряд на випадковість розподілу. Часто використовуваними в цих цілях є: t-критерій, F-критерій, критерій Фостера-Стюарта, критерій серій, заснований на медіані вибірки. Далі наведено розрахунок цих критеріїв.

У меню програм необхідно обрати програму *Microsoft Excel*, після її запуску ввести вихідні дані так, як це показано на рис. 3.2.

	A	B	C
1		Date	Sales
2	1	19.11.2014	127,96
3	2	20.11.2014	127,9
4	3	21.11.2014	128,37
5	4	22.11.2014	130,1
6	5	23.11.2014	129,66
7	6	26.11.2014	128,79
8	7	27.11.2014	129,83
9	8	28.11.2014	130,23
10	9	29.11.2014	130,22
11	10	30.11.2014	129,66
12	11	03.12.2014	129,52
13	12	04.12.2014	130,43
14	13	05.12.2014	130,58
15	14	06.12.2014	131,54
16	15	07.12.2014	131,7
17	16	10.12.2014	132,49
18	17	11.12.2014	132,45
19	18	12.12.2014	131,72
20	19	13.12.2014	131,38
21	20	14.12.2014	132,37
22	21	17.12.2014	133,51
23	22	18.12.2014	132,53
24	23	19.12.2014	133,13
25	24	20.12.2014	131,92
26	25	21.12.2014	131,39
27	26	24.12.2014	131,15
28	27	25.12.2014	130,95
29	28	26.12.2014	129,38

Рис. 3.2. Вихідні дані

Для побудови графіка виділити комірки C1:C29 із вихідними даними разом із заголовком Sales та обрати в пункті меню "Вставка" підпункт "Графік" (Графік з маркерами). Результат наведено на рис. 3.3.

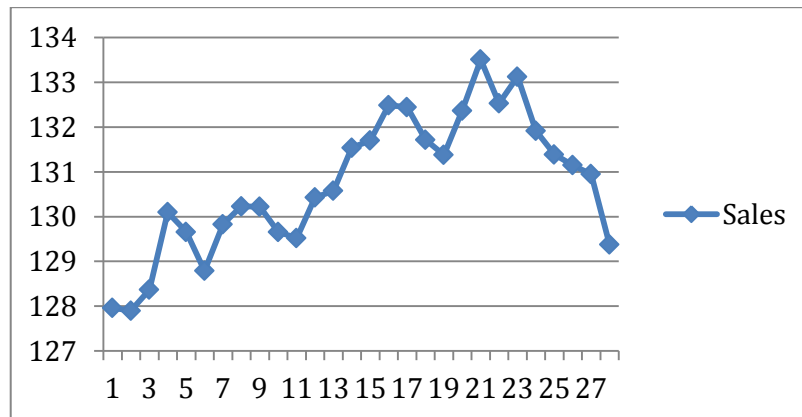


Рис. 3.3. Графік вихідних даних

У ході визначення тренда в дисперсії використовується критерій Фішера. Вихідний ряд y_1, y_2, \dots, y_n розподіляється на дві сукупності обсягом n_1 та n_2 $n_1 \approx \frac{n}{2}, n_2 = n - n_1$:

$y_1, y_2, \dots, y_k,$

$y_{k+1}, y_{k+2}, \dots, y_n.$

У даному випадку: $n_1 = \frac{28}{2} = 14, n_2 = 28 - 14 = 14.$

Результат наведено на рис. 3.4.

D	E	F
	Group 1	Group 2
1	127,96	131,70
2	127,90	132,49
3	128,37	132,45
4	130,10	131,72
5	129,66	131,38
6	128,79	132,37
7	129,83	133,51
8	130,23	132,53
9	130,22	133,13
10	129,66	131,92
11	129,52	131,39
12	130,43	131,15
13	130,58	130,95
14	131,54	129,38

Рис. 3.4. Результат розподілу вихідних даних на 2 групи

Для кожної із сукупностей розрахувати середнє значення та дисперсію, як показано на рис. 3.5.

	D	E	F
16			
17	Average	=СРЗНАЧ(E2:E16)	=СРЗНАЧ(F2:F16)
18	Variance	=ДИСП(E2:E16)	=ДИСП(F2:F16)
19	n_i	=СЧЕТ(E2:E16)	=СЧЕТ(F2:F16)
20	$n_i - 1$	=E19-1	=F19-1
21			

Рис. 3.5. Формули для розрахунку

На рис. 3.6. наведені результати розрахунку середніх значень та дисперсій для кожної із групи даних.

16			
17	Average	129,6	131,9
18	Variance	1,09603	1,05939
19	n_i	14	14
20	$n_i - 1$	13	13

Рис. 3.6. Результати розрахунків

Розрахункове значення критерію Фішера визначається за формулою:

$$F_{\text{теор}} = \begin{cases} \frac{S_2^2}{S_1^2}, \text{ якщо } S_2^2 > S_1^2 \\ \frac{S_1^2}{S_2^2}, \text{ якщо } S_1^2 > S_2^2 \end{cases}$$

Розрахункове значення порівнюється з табличним для рівня значущості $\alpha = 0,05$ та числа ступенів волі $k_1 = n_1 - 1$ і $k_2 = n_2 - 1$ (де k_1 – відповідає більшій дисперсії). Табличне значення можна визначити за формулою: $F_{\text{ТАСПОБР}}(0,05;E20;F20)$. Результати розрахунків наведені на рис. 3.7.

	B	C	D	E	F	G	H
1	Date	Sales		Group 1	Group 2		
2	19.11.2014	127,96	1	127,96	131,70		
3	20.11.2014	127,9	2	127,90	132,49		
4	21.11.2014	128,37	3	128,37	132,45		
5	22.11.2014	130,1	4	130,10	131,72		
6	23.11.2014	129,66	5	129,66	131,38	F =	1,03459
7	26.11.2014	128,79	6	128,79	132,37		
8	27.11.2014	129,83	7	129,83	133,51	F_{tab} =	2,5769
9	28.11.2014	130,23	8	130,23	132,53		
10	29.11.2014	130,22	9	130,22	133,13		
11	30.11.2014	129,66	10	129,66	131,92		
12	03.12.2014	129,52	11	129,52	131,39		
13	04.12.2014	130,43	12	130,43	131,15		
14	05.12.2014	130,58	13	130,58	130,95		
15	06.12.2014	131,54	14	131,54	129,38		
16	07.12.2014	131,7					
17	10.12.2014	132,49	Average	129,6	131,9		
18	11.12.2014	132,45	Variance	1,09603	1,05939		
19	12.12.2014	131,72	n_i	14	14		
20	13.12.2014	131,38	$n_i - 1$	13	13		

Рис. 3.7. Результати розрахунків

Якщо $F_{\text{теор}} \geq F_{\text{табл}}(\alpha, k_1, k_2)$, то підтверджується гіпотеза про наявність тренда у дисперсії. У даному випадку $F_{\text{теор}} < F_{\text{табл}}(\alpha, k_1, k_2)$, можна зробити висновок, що тренд у часовому ряді відсутній.

Після аналізу наявності тренда в дисперсії необхідно перейти до аналізу наявності тренда в середньому за допомогою *методу порівняння середніх*. Цей метод базується на t-критерії, що розраховується за формулою:

$$t_{\text{розр}} = \frac{y_1 - y_2}{\sqrt{\frac{n_1 - 1 \cdot S_1^2 + n_2 - 1 \cdot S_2^2}{n_1 + n_2}}} \cdot \frac{n_1 \cdot n_2 \cdot (n_1 + n_2 - 2)}{n_1 + n_2}$$

Формули для розрахунків наведені на рис. 3.8.

F	G	H
Group 2		
131,7		
132,49		
132,45		
131,72		
131,38	F = =E18/F18	
132,37		
133,51	F_{tab} = =FRАСПОБР(0,05;E20;F20)	
132,53		
133,13	t = =ABS(E17-F17)/КОРЕНЬ(E20*E18+F20*F18)*КОРЕНЬ((E19*F19*(E19+F19-2))/(E19+F19))	
131,92		
131,39	t_{tab} = =СТЬЮДРАСПОБР(0,05;E19+F19-2)	
131,15		
130,95		
129,38		
=CPЗНАЧ(F2:F16)		
=ДИСП(F2:F16)		
=СЧЕТ(F2:F16)		
=F19-1		

Рис. 3.8. Формули для розрахунків

Розрахункові значення критерію Стьюдента порівнюються з табличним ($t_{\text{табл}}$) для рівня значущості $\alpha = 0,05$ та числа ступенів волі $k = n - 2$. У даному випадку $t_{\text{розр}} > t_{\text{табл}}$, тобто підтверджується гіпотеза про наявність тренда в середньому. Оскільки $y_2 > y_1$, то тренд є зростаючим.

Результати розрахунків наведені на рис. 3.9.

	D	E	F	G	H
1		Group 1	Group 2		
2	1	127,96	131,70		
3	2	127,90	132,49		
4	3	128,37	132,45		
5	4	130,10	131,72		
6	5	129,66	131,38	F =	1,03459
7	6	128,79	132,37		
8	7	129,83	133,51	F_{tab} =	2,5769
9	8	130,23	132,53		
10	9	130,22	133,13	t =	5,69425
11	10	129,66	131,92		
12	11	129,52	131,39	t_{tab} =	2,05553
13	12	130,43	131,15		
14	13	130,58	130,95		
15	14	131,54	129,38		
16					
17	Average	129,6	131,9		
18	Variance	1,09603	1,05939		
19	n_i	14	14		
20	$n_i - 1$	13	13		

Рис. 3.9. Результати розрахунків

Критерій Фостера-Стюарта використовується для перевірки тренда як середніх, так і дисперсій. Метод базується на аналізі рекордних значень ряду. Значення рівня u_i є рекордним, якщо воно більше або менше усіх попередніх. Визначити значення u_i і v_i за формулами:

$$u_i = \begin{cases} 1, \text{ якщо } u_i \text{ більше усіх попередніх,} \\ 0, \text{ у протилежному випадку} \end{cases};$$

$$v_i = \begin{cases} 1, \text{ якщо } u_i \text{ менше усіх попередніх,} \\ 0, \text{ у протилежному випадку} \end{cases}.$$

Для цього створити заголовок таблиці в I1, J1, K1, L1. У строчці з номером 2 поставити прочерки. У третій строчці в комірку I ввести формулу =ЕСЛИ(C3=МАКС(\$C\$2:C3);1;0). У цій формулі зафіксувати тільки початок діапазону. Потім формулу продублювати до 28-го елемента включно. У стовбці J у комірку J3 ввести подібну формулу: =ЕСЛИ(C3=МИН(\$C\$2:C3);1;0) і продублювати її. Формули для розрахунків наведені на рис. 3.10.

	I	J	K	L
1	u_i	v_i	s_i	d_i
2	---	---	---	---
3	=ЕСЛИ(C3=МАКС(\$C\$2:C3);1;0)	=ЕСЛИ(C3=МИН(\$C\$2:C3);1;0)	=I3+J3	=I3-J3
4	=ЕСЛИ(C4=МАКС(\$C\$2:C4);1;0)	=ЕСЛИ(C4=МИН(\$C\$2:C4);1;0)	=I4+J4	=I4-J4
5	=ЕСЛИ(C5=МАКС(\$C\$2:C5);1;0)	=ЕСЛИ(C5=МИН(\$C\$2:C5);1;0)	=I5+J5	=I5-J5
6	=ЕСЛИ(C6=МАКС(\$C\$2:C6);1;0)	=ЕСЛИ(C6=МИН(\$C\$2:C6);1;0)	=I6+J6	=I6-J6
7	=ЕСЛИ(C7=МАКС(\$C\$2:C7);1;0)	=ЕСЛИ(C7=МИН(\$C\$2:C7);1;0)	=I7+J7	=I7-J7
8	=ЕСЛИ(C8=МАКС(\$C\$2:C8);1;0)	=ЕСЛИ(C8=МИН(\$C\$2:C8);1;0)	=I8+J8	=I8-J8
9	=ЕСЛИ(C9=МАКС(\$C\$2:C9);1;0)	=ЕСЛИ(C9=МИН(\$C\$2:C9);1;0)	=I9+J9	=I9-J9
10	=ЕСЛИ(C10=МАКС(\$C\$2:C10);1;0)	=ЕСЛИ(C10=МИН(\$C\$2:C10);1;0)	=I10+J10	=I10-J10
11	=ЕСЛИ(C11=МАКС(\$C\$2:C11);1;0)	=ЕСЛИ(C11=МИН(\$C\$2:C11);1;0)	=I11+J11	=I11-J11
12	=ЕСЛИ(C12=МАКС(\$C\$2:C12);1;0)	=ЕСЛИ(C12=МИН(\$C\$2:C12);1;0)	=I12+J12	=I12-J12
13	=ЕСЛИ(C13=МАКС(\$C\$2:C13);1;0)	=ЕСЛИ(C13=МИН(\$C\$2:C13);1;0)	=I13+J13	=I13-J13
14	=ЕСЛИ(C14=МАКС(\$C\$2:C14);1;0)	=ЕСЛИ(C14=МИН(\$C\$2:C14);1;0)	=I14+J14	=I14-J14
15	=ЕСЛИ(C15=МАКС(\$C\$2:C15);1;0)	=ЕСЛИ(C15=МИН(\$C\$2:C15);1;0)	=I15+J15	=I15-J15
16	=ЕСЛИ(C16=МАКС(\$C\$2:C16);1;0)	=ЕСЛИ(C16=МИН(\$C\$2:C16);1;0)	=I16+J16	=I16-J16
17	=ЕСЛИ(C17=МАКС(\$C\$2:C17);1;0)	=ЕСЛИ(C17=МИН(\$C\$2:C17);1;0)	=I17+J17	=I17-J17
18	=ЕСЛИ(C18=МАКС(\$C\$2:C18);1;0)	=ЕСЛИ(C18=МИН(\$C\$2:C18);1;0)	=I18+J18	=I18-J18
19	=ЕСЛИ(C19=МАКС(\$C\$2:C19);1;0)	=ЕСЛИ(C19=МИН(\$C\$2:C19);1;0)	=I19+J19	=I19-J19
20	=ЕСЛИ(C20=МАКС(\$C\$2:C20);1;0)	=ЕСЛИ(C20=МИН(\$C\$2:C20);1;0)	=I20+J20	=I20-J20
21	=ЕСЛИ(C21=МАКС(\$C\$2:C21);1;0)	=ЕСЛИ(C21=МИН(\$C\$2:C21);1;0)	=I21+J21	=I21-J21
22	=ЕСЛИ(C22=МАКС(\$C\$2:C22);1;0)	=ЕСЛИ(C22=МИН(\$C\$2:C22);1;0)	=I22+J22	=I22-J22
23	=ЕСЛИ(C23=МАКС(\$C\$2:C23);1;0)	=ЕСЛИ(C23=МИН(\$C\$2:C23);1;0)	=I23+J23	=I23-J23
24	=ЕСЛИ(C24=МАКС(\$C\$2:C24);1;0)	=ЕСЛИ(C24=МИН(\$C\$2:C24);1;0)	=I24+J24	=I24-J24
25	=ЕСЛИ(C25=МАКС(\$C\$2:C25);1;0)	=ЕСЛИ(C25=МИН(\$C\$2:C25);1;0)	=I25+J25	=I25-J25
26	=ЕСЛИ(C26=МАКС(\$C\$2:C26);1;0)	=ЕСЛИ(C26=МИН(\$C\$2:C26);1;0)	=I26+J26	=I26-J26
27	=ЕСЛИ(C27=МАКС(\$C\$2:C27);1;0)	=ЕСЛИ(C27=МИН(\$C\$2:C27);1;0)	=I27+J27	=I27-J27
28	=ЕСЛИ(C28=МАКС(\$C\$2:C28);1;0)	=ЕСЛИ(C28=МИН(\$C\$2:C28);1;0)	=I28+J28	=I28-J28
29	=ЕСЛИ(C29=МАКС(\$C\$2:C29);1;0)	=ЕСЛИ(C29=МИН(\$C\$2:C29);1;0)	=I29+J29	=I29-J29
30	Sum		=СУММ(K3:K29)	=СУММ(L3:L29)

Рис. 3.10. Формули для розрахунків

Далі визначити величини $s_i = u_i + v_i$ та $d_i = u_i - v_i$, а також:

$$S = \sum_{i=2}^n s_i, \quad D = \sum_{i=2}^n d_i.$$

Статистика S використовується для перевірки тренда в дисперсії, статистика d – для виявлення тренда в середньому.

Розрахувати для кожної з цих статистик значення критерію Стьюдента за формулами:

$$t_D = \frac{D}{\sigma_D n}, \quad t_S = \frac{S - \mu n}{\sigma_S n},$$

де μn , $\sigma_S n$, $\sigma_D n$ – табличні значення.

Розрахункові значення критерію Стьюдента порівнюються з табличним $t_{\text{табл}}^S$ та $t_{\text{табл}}^d$ залежно від заданого значення рівня значущості α та числа ступенів волі $k = n - 3$. Зробити висновки:

а) якщо $t_D > t_p$, $t_S > t_p$ – приймається гіпотеза про наявність тренда, при цьому, якщо $D > 0$, то тренд є висхідним, а якщо $D < 0$, то тренд спадний;

б) якщо $t_D < t_p$, $t_S > t_p$ – приймається гіпотеза про наявність тренда в дисперсії;

в) $t_D > t_p$, $t_S < t_p$ – не можна ні підтвердити, ні спростувати гіпотезу про наявність тренда;

г) $t_D < t_p$, $t_S < t_p$ – приймається гіпотеза про відсутність тренда.

Розрахувати статистики для даного прикладу:

$$S = \sum_{i=2}^{28} s_i = 10, \quad D = \sum_{i=2}^{28} d_i = 8,$$

$$t_D = \frac{D}{\sigma_D n} = 3,27, \quad t_S = \frac{S - \mu n}{\sigma_S n} = 2,13.$$

Порівняння розрахункових значень критерію Стьюдента з табличним $t_{\text{табл}}$ для рівня значущості $\alpha = 0,05$ та числа ступенів волі $k = 25$, що дорівнює 2,06, дозволяє зробити висновок про наявність тренда в часовому ряді.

Критерій серій. Нехай є вибірка y_1, y_2, \dots, y_N . Розташувати елементи вибірки в порядку зростання, тобто, в так званий варіаційний ряд (так що, наприклад, $y_{(1)}$ – це найменше із усіх вибірових значень y_1, y_2, \dots, y_N ; y_N – найбільше із усіх вибірових даних).

У якості вибіркового значення медіани Me береться середній (за розташуванням) елемент варіаційного ряду, тобто:

$$Me = \begin{cases} \frac{1}{2} y_{\frac{N}{2}} + y_{\frac{N}{2}+1}, & \text{якщо } N \text{ парне;} \\ y_{\frac{N+1}{2}}, & \text{якщо } N \text{ непарне.} \end{cases}$$

Результати розрахунків наведені на рис. 3.11.

	C	D	E	F	G	H
1	Sales					
2	127,96	Me	130,77			
3	127,9					
4	128,37					
5	130,1					
6	129,66					
7	128,79					
8	129,83					
9	130,23					
10	130,22					
11	129,66					
12	129,52					
13	130,43					
14	130,58					
15	131,54					
16	131,7					
17	132,49					
18	132,45					

Рис. 3.11. Результати розрахунків

Слід вернуться до вихідної вибірки y_1, y_2, \dots, y_N і замість кожного y_j поставити плюс, якщо $y_j > Me$, і мінус, якщо $y_j < Me$ (члени вибірки, рівні медіані в отриманій в такий спосіб послідовності плюсів і мінусів, опускаються). Результати розрахунків наведені на рис. 3.12. Отримана послідовність плюсів і мінусів характеризується загальним числом серій $v(n)$ і довжиною самої довгої серії $\tau(n)$, де під "серією" розуміється послідовність плюсів або мінусів, що ідуть підряд. Очевидно, що якщо спостереження незалежні, то чергування плюсів і мінусів у послідовності повинне бути більш-менш "випадковим", тобто ця послідовність не повинна містити занадто довгих серій. У даному випадку загальне число серій v дорівнює 3, а довжина самої довгої серії τ – 14.

	C	D	E	F	G	H
1	Sales			Ряд знаків		
2	127,96	Me	130,77	-		
3	127,9			-		
4	128,37			-		
5	130,1			-		
6	129,66			-		
7	128,79			-		
8	129,83			-		
9	130,23			-		
10	130,22			-		
11	129,66			-		
12	129,52			-		
13	130,43			-		
14	130,58			-		
15	131,54			+		
16	131,7			+		
17	132,49			+		
18	132,45			+		
19	131,72			+		
20	131,38			+		

Рис. 3.12. Результати розрахунків

У даному критерії розглядається одночасно пара критичних статистик $(v_N; \tau_N)$ Для величини рівня значимості $\alpha = 0,05$ справедливі такі нерівності:

$$\tau_{max} N < (3,3 \cdot \log_{10} N + 1);$$

$$v_N > (0,5 \cdot (N + 1 - 1,96 \cdot \sqrt{N - 1})).$$

У випадку, якщо хоча б одна з нерівностей виявиться порушеною, то гіпотеза про те, що вихідні результати спостереження є незалежними, відкидається.

Критичні значення для аналізованого часового ряду дорівнюють відповідно $\tau_p = 5,77$, $v_p = 8,41$. Отже, за досліджуваним критерієм гіпотеза про стаціонарність часового ряду відхиляється.

Лабораторна робота 4

Рішення та аналіз моделей простої екстраполяції й інтерполяції одномірних процесів

Мета – формування у студентів навичок побудови базових моделей прогнозування одномірних процесів, що містить побудову графіків, аналіз моделей, розрахунок прогнозів та інтерпретацію отриманих результатів за допомогою програми *Microsoft Excel*.

Умови завдання

Побудувати прогнозну модель валового доходу підприємства. Часовий ряд, що характеризує динаміку зміни валового доходу підприємства за 2007 – 2014 рр., наведено у табл. 4.1.

Таблиця 4.1

Динаміка валового доходу підприємства, тис. грн

Роки	2007	2008	2009	2010	2011	2012	2013	2014
y_j	6 103,6	5 532,4	4 173,5	6 340,6	7 045,9	8 004,3	12 062,5	15 036,0

Необхідно на основі даних 2007 – 2014 рр. розрахувати прогнозні значення валового доходу підприємства на 2015 – 2017 рр., використовуючи: лінійні тренди з оцінками параметрів за методом двох крайніх і середніх групових точок; показники середнього темпу зростання та середнього абсолютного приросту. Оцінити якість моделей прогнозування за критерієм середньої абсолютної відсоткової помилки.

Методичні рекомендації до виконання завдання

Сутність методу прогнозування на основі прямої, проведеної через дві крайні точки, полягає у такому. Із упорядкованої вибірки беруть дві крайні точки x_1, y_1 і x_n, y_n і через них проводять пряму вигляду $y = a_0 + a_1x$.

Оцінки параметрів обчислюються за формулами:

$$a_1 = \frac{y_n - y_1}{x_n - x_1}, a_0 = y_1 - a_1x_1$$

Отримані таким чином оцінки є незміщеними, але вони не будуть обґрунтованими і тим більш ефективними.

Результати розрахунків параметрів лінійної моделі тренда за допомогою методу двох крайніх крапок для даного випадку наведені на рис. 4.1.

	A	B	C	D	E
1	Роки	t	Yt		
2	2007	1	6103,6		
3	2008	2	5532,4	a1	1276,06
4	2009	3	4173,5	a0	4827,54
5	2010	4	6340,6		
6	2011	5	7045,9		
7	2012	6	8004,3		
8	2013	7	12062,5		
9	2014	8	15036		

Рис. 4.1. Результати розрахунків

Таким чином, отримана лінійна модель тренда має вигляд:

$$y_t = 4827,54 + 1276,06t,$$

де y_t – теоретичні значення досліджуваного показника;
 t – період часу.

Прогнозні значення валового доходу підприємства на 2015 – 2017 рр. за цією моделлю дорівнюють (рис. 4.2):

	A	B	C	D	E	F
1	Роки	t	Yt	\hat{y}_t		
2	2007	1	6103,6	6103,6		
3	2008	2	5532,4	7379,7		
4	2009	3	4173,5	8655,7		
5	2010	4	6340,6	9931,8		
6	2011	5	7045,9	11207,8		
7	2012	6	8004,3	12483,9		
8	2013	7	12062,5	13760,0		
9	2014	8	15036	15036,0		
10	2015	9		16312,1		
11	2016	10		17588,1		
12	2017	11		18864,2		

Рис. 4.2. Прогнозні значення валового доходу підприємства

Слід розглянути сутність методу прогнозування по прямій, проведеної по координатах двох середніх групових точок. Метод полягає у тому, що вихідну сукупність розподіляють на дві або три приблизно рівні частини, а потім знаходять координати середніх точок для крайніх груп.

Нехай x_1, y_1, x_2, y_2 , – координати середніх точок для крайніх груп. Тоді оцінки лінійної моделі a_0, a_1 обчислюються за формулами:

$$a_1 = \frac{y_2 - y_1}{x_2 - x_1}, a_0 = y_1 - a_1 x_1.$$

Оцінки, отримані цим методом, є незміщеними і обґрунтованими, але неефективними.

За методом двох середніх групових точок знайти оцінки параметрів моделі a_0, a_1 , розділивши попередньо вибірку на дві частини: 1 група – дані з 2007 по 2010 рр., 2 група – дані з 2011 по 2014 рр. Розрахувати середні значення у кожній групі (рис. 4.3).

E3		fx		=CPЗНАЧ(B2:B5)		
	A	B	C	D	E	F
1	Роки	t	Yt			
2	2007	1	6103,6			
3	2008	2	5532,4	\bar{x}_I	2,5	
4	2009	3	4173,5	\bar{x}_{II}	6,5	
5	2010	4	6340,6	\bar{y}_I	5537,5	
6	2011	5	7045,9	\bar{y}_{II}	10537,2	
7	2012	6	8004,3			
8	2013	7	12062,5			
9	2014	8	15036			

Рис. 4.3. Результати розрахунків середніх за групами

Знайти оцінки параметрів моделі за методом двох крайніх точок (рис. 4.4).

E7					fx	=(E6-E5)/(E4-E3)
	A	B	C	D	E	
1	Роки	t	Yt			
2	2007	1	6103,6			
3	2008	2	5532,4	\bar{x}_y		2,5
4	2009	3	4173,5	\bar{x}_{yz}		6,5
5	2010	4	6340,6	\bar{y}_z		5537,5
6	2011	5	7045,9	\bar{yz}		10537,2
7	2012	6	8004,3	a1		1249,9
8	2013	7	12062,5	ao		2412,7
9	2014	8	15036			

Рис. 4.4. Результати розрахунків

Таким чином, отримана лінійна модель має вигляд:

$$y_t = 2412,7 + 1249,9t.$$

Прогнозні значення валового доходу підприємства за цією моделлю на 2015 – 2017 рр. дорівнюють (рис. 4.5):

D10						fx	=2412,7+1249,9*B10
	A	B	C	D	E	F	
1	Роки	t	Yt	\hat{y}_t			
2	2007	1	6103,6	3662,6			
3	2008	2	5532,4	4912,5			
4	2009	3	4173,5	6162,4			
5	2010	4	6340,6	7412,3			
6	2011	5	7045,9	8662,2			
7	2012	6	8004,3	9912,1			
8	2013	7	12062,5	11162,0			
9	2014	8	15036	12411,9			
10	2015	9		13661,8			
11	2016	10		14911,7			
12	2017	11		16161,6			

Рис. 4.5. Прогнозні значення валового доходу підприємства

Подальший аналіз та прогнозування динаміки доходу підприємства слід провести, використовуючи показники середнього темпу зростання та середнього абсолютного приросту.

Середній темп зростання може бути виражено формулою:

$$\tau = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Метод екстраполяції на основі середнього темпу зростання полягає у такому: якщо в основу прогностичного розрахунку покладено середній темп зростання, то значення рівня, що екстраполюється, одержують за формулою:

$$y_{i+L} = y_i^* \cdot \tau^L,$$

де τ – середній темп зростання;

y_i^* – рівень, прийнятий за базу для екстраполяції;

L – період попередження.

Результати розрахунків наведені на рис. 4.6.

E10		fx		=D9*D13^C10		
	A	B	C	D	E	F
1	Роки	t	L	Yt	\hat{y}_t	
2	2007	1		6103,6		
3	2008	2		5532,4		
4	2009	3		4173,5	5014,7	
5	2010	4		6340,6	3451,1	
6	2011	5		7045,9	6421,6	
7	2012	6		8004,3	7303,4	
8	2013	7		12062,5	8450,3	
9	2014	8		15036	13512,8	
10	2015	9	1		17102,8	
11	2016	10	2		19453,7	
12	2017	11	3		22127,8	
13	$\bar{\tau}$			1,14		
14						

Рис. 4.6. Результати розрахунків

Середній абсолютний приріст (Δ) є узагальнюючим показником швидкості зміни явища у часі. Даний показник дає можливість встановити, наскільки в середньому за одиницю часу повинен збільшуватися рівень ряду (в абсолютному вираженні), щоб, починаючи від начального рівня, за дану кількість періодів часу, досягти кінцевого рівня. Для його визначення використовується формула:

$$\Delta = \frac{y_n - y_1}{n - 1}$$

Прогнозування на основі середнього абсолютного приросту проводиться таким чином:

$$y_{i+L} = y_i^* + \Delta \cdot L,$$

де Δ – середній абсолютний приріст;

y_i^* – рівень, прийнятий за базу для екстраполяції;

L – період попередження.

Результати розрахунків наведені на рис. 4.7.

E10		fx		= \$D\$9+\$D\$13*C10	
	A	B	C	D	E
1	Роки	t	L	Yt	\hat{y}_t
2	2007	1		6103,6	
3	2008	2		5532,4	
4	2009	3		4173,5	4961,2
5	2010	4		6340,6	3208,5
6	2011	5		7045,9	6419,6
7	2012	6		8004,3	7281,5
8	2013	7		12062,5	8321,1
9	2014	8		15036	12913,8
10	2015	9	1		16312,1
11	2016	10	2		17588,1
12	2017	11	3		18864,2
13	$\bar{\Delta y}$			1276,06	

Рис. 4.7. Результати розрахунків

Значення середньої абсолютної процентної похибки (*mean absolute percentage error*), розраховані за формулою (2.1) (див. лабораторну роботу 2), наведені в табл. 4.2.

Оцінювання якості моделей

Методи	За методом крайніх точок	За методом середніх групових точок	За середнім темпом зростання	За середнім абсолютним приростом
Умовне позначення моделі	Модель 1	Модель 2	Модель 3	Модель 4
MAPE, %	40,82	23,43	15,43	16,42

Як видно з табл. 4.2, найбільш високу точність прогнозу забезпечує модель на підставі середнього темпу зростання.

Графік динаміки зміни валового доходу підприємства із прогнозними значеннями наведено на рис. 4.8.

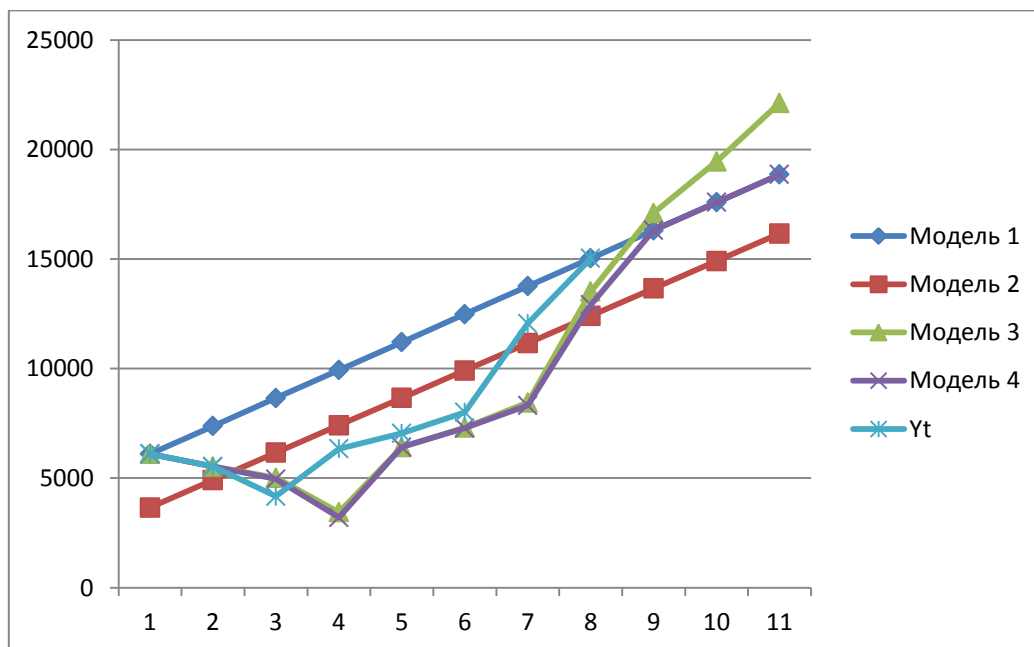


Рис. 4.8. Прогнозування валового доходу

Як видно з рис. 4.8, ретроспективний прогноз за методом двох крайніх точок (модель 1) є досить завищеним. Задовільну точність прогнозу за критерієм MAPE забезпечують моделі на підставі середнього абсолютного приросту, середнього темпу зростання. Перспективний прогноз за лінійним трендом та середнім абсолютним приростом співпадає: прогнозування за середнім абсолютним приростом еквівалентне представленню ряду в вигляді рівняння прямої. Перспективний прогноз за середнім темпом зростання на 2015 – 2017 рр. складає: 17 102,8 тис. грн; 19 453,7 тис. грн; 22 127,8 тис. грн відповідно. Цей прогноз можна прийняти як основний.

Лабораторна робота 5

Методи розв'язання адаптивних моделей прогнозування

Мета – закріплення теоретичного й практичного матеріалу, придбання навичок побудови адаптивних моделей прогнозування у модулі *Time Series/Forecasting – Часові ряди/Прогнозування* ППП *Statistica*.

Умови завдання

У табл. 5.1 наведені дані, що відображають динаміку курсу акцій підприємства (y_t).

Таблиця 5.1

Динаміка курсу акцій підприємства у травні 2009 р., грн

T	1	2	3	4	5	6	7	8	9	10
y_t	6,01	5,98	5,98	5,98	5,60	4,55	4,55	4,20	3,90	4,20
T	11	12	13	14	15	16	17	18	19	20
y_t	4,00	4,00	4,00	4,00	4,00	4,00	4,00	3,60	2,60	2,61

Необхідно

1. Побудувати графік вихідних даних.
2. Провести експоненціальне згладжування при $\alpha = 0,9$. У якості початкового значення експонентної середньої взяти середнє арифметичне значення рівнів ряду.
3. Побудувати модель Хольта при $\alpha_1 = 0,8$, $\alpha_2 = 0,3$. У якості початкових значень коефіцієнтів моделі Хольта взяти МНК-оцінки лінійного тренду.
4. Оцінити точність отриманих варіантів прогнозовної моделі та обрати найкращий із них за критерієм середньої абсолютної процентної помилки.
5. Зробити прогноз курсу акцій підприємства на 5 днів.

Методичні рекомендації до виконання завдання

Слід розглянути реалізацію моделей експоненціального згладжування у пакеті прикладних програм *Statistica*. Аналіз часових рядів здійснюється в модулі *Advanced Linear/Nonlinear Models/Time Series/Forecasting – Часові ряди/Прогнозування*. Стартова панель модуля наведена на рис. 5.1.

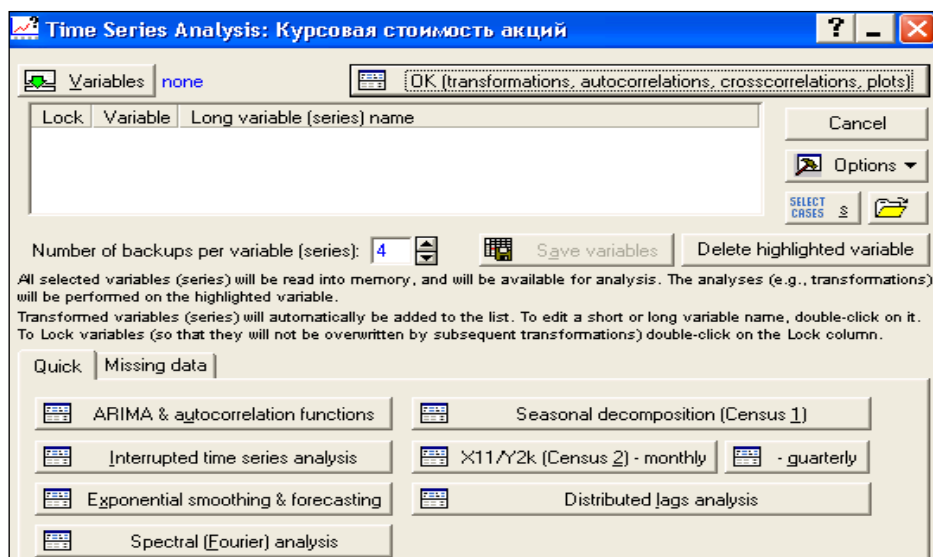


Рис. 5.1. Стартова панель модуля *Time Series/Forecasting*

Із метою побудови графіка вихідних даних обрати в пункті меню *Graphs* підпункт *Scatterplots (Точкові графіку)*. У процесі установки характеристик графіка обрати *Graph type – Regular (Простий)*, поставити відмітку *Linear fit (Лінійний тренд)*, а в якості змінних, що відбиваються на графіку, обрати по осі *X – Період часу*, а по осі *Y – Курс акцій*. Результат наведено на рис. 5.2.

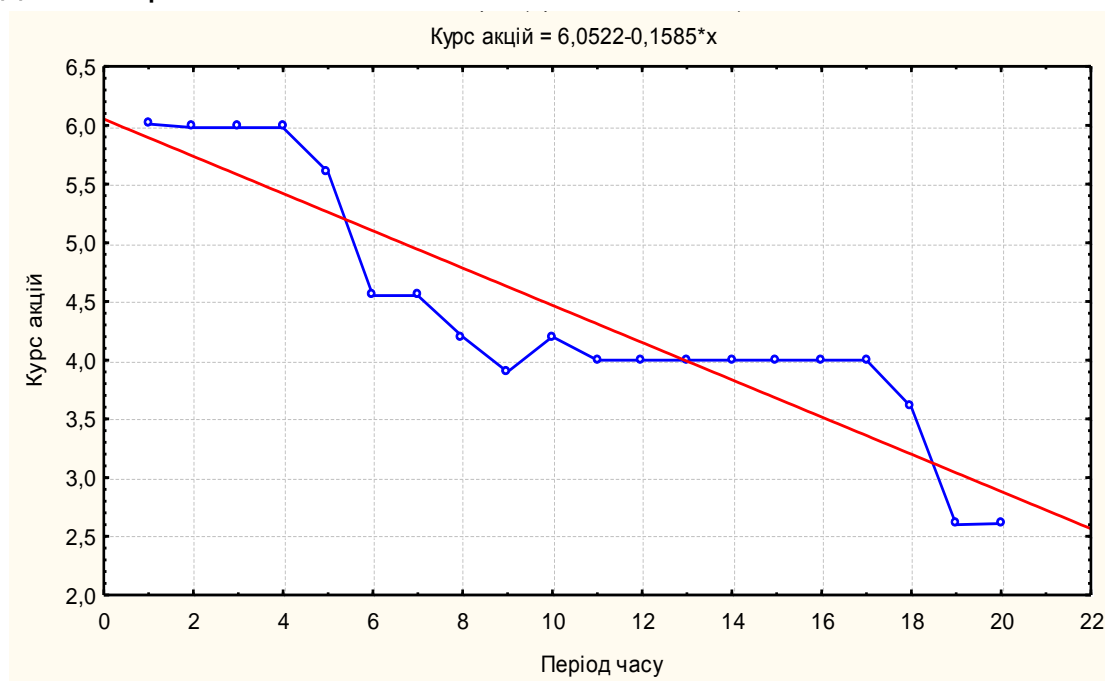


Рис. 5.2. Графік вихідних даних

Спочатку необхідно провести експоненціальне згладжування вихідних даних. Для цього вибрати вкладку *Exponential smoothing & forecasting – Експоненціальне згладжування і прогнозування*. У верхній частині стартової панелі модуля необхідно вказати ім'я змінної, тобто відмітити *Курс акцій* (рис. 5.3).

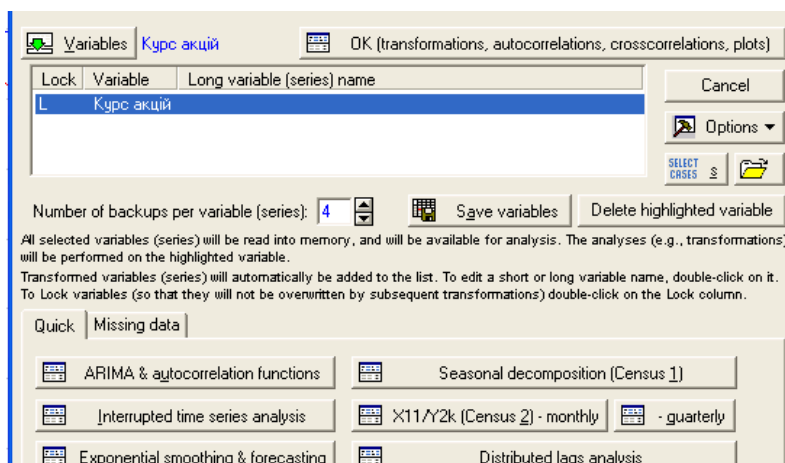


Рис. 5. 3. Вибір змінної для прогнозування

У даному вікні необхідно задати параметри експоненціального згладжування (рис. 5.4). Так, змінна для аналізу – *Курс акцій*, лаг сезонної компоненти (*Seasonal component*) – відсутній, тому що дані подані за 20 днів протягом одного місяця.

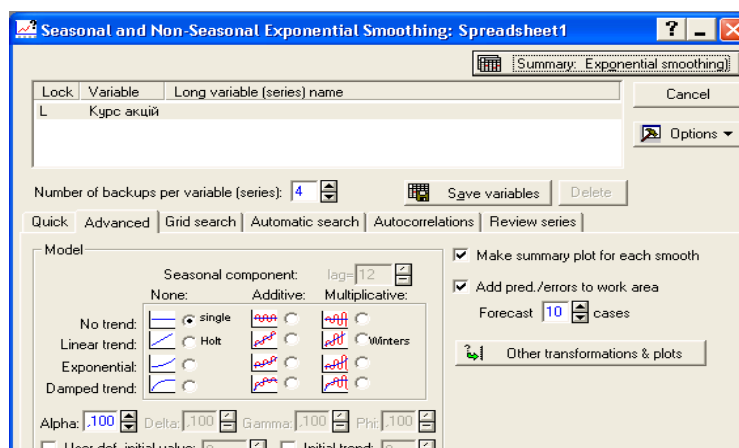


Рис. 5.4. Визначення характеристик моделі експоненціального згладжування

У середній частині вікна наведені можливості вибору різних типів моделі з урахуванням або без урахування сезонних факторів, у правій частині вікна – період прогнозування, що складає 10 кроків, і його можна змінювати залежно від мети прогнозування та характеру вихідних даних.

У нижній частині вказані значення та способи визначення параметрів згладжування. Параметри згладжування можуть задаватися безпосередньо дослідником, або за допомогою оптимізаційних процедур під час натискання вкладки *Automatic search* – *Автоматичний пошук* із урахуванням таких критеріїв: *mean squared error* – середня квадратична помилка, *mean absolute error* – середня абсолютна помилка, *mean absolute error* – середня абсолютна процентна помилка (рис. 5.5).

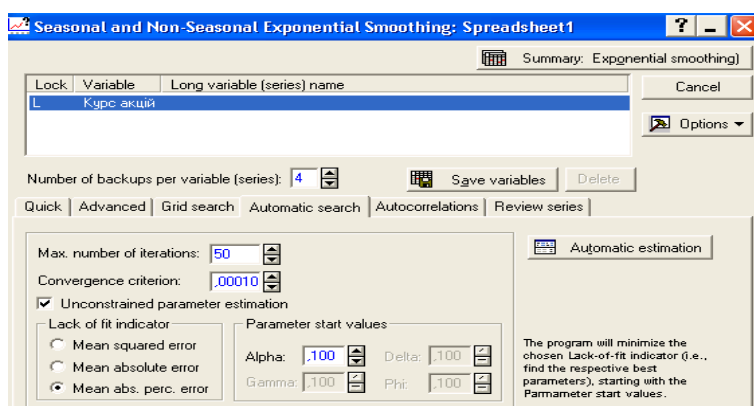


Рис. 5.5. Вікно вибору критерію пошуку оптимальних параметрів згладжування

Слід продемонструвати на основі вихідних даних основні типи моделей:

1. Побудувати модель експоненціального згладжування без урахування тренда та сезонної компоненти.

Для визначення найкращих параметрів моделі використати вбудовану оптимізаційну процедуру (рис. 5.5) із урахуванням середньої абсолютної процентної помилки. Після чого натиснути кнопку *Automatic estimation* – *Автоматичне оцінювання*.

Результатом оцінювання параметрів моделі є графік вихідних даних, згладжених і прогнозних (рис. 5.6); таблиця оцінок якості моделі (рис. 5.7); таблиця з вихідними даними, згладженими даними (*Smoothed Series*) і помилками (*Resids*) (рис. 5.8).

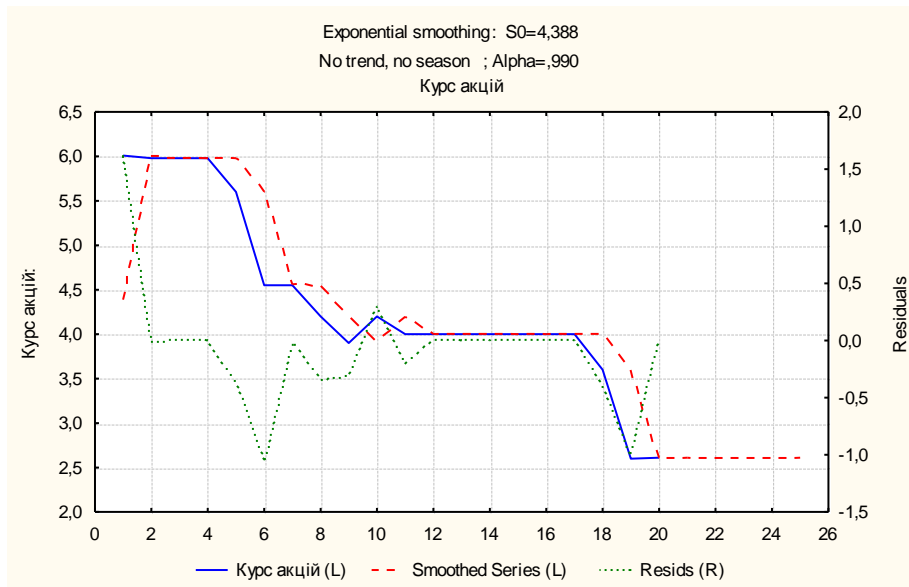


Рис. 5.6. Графік вихідних, згладжених і прогнозних значень курсу акцій

Exponential smoothing No trend, no season Курс акцій	
Summary of error	Error
Mean error	-0,089797959596
Mean absolute error	0,281694458542
Sums of squares	5,395795122273
Mean square	0,269789756113
Mean percentage error	-3,358651790040
Mean abs. perc. error	6,764546563301

Рис. 5.7. Оцінки якості моделі згладжування

Як видно з рис. 5.7, величина середньої абсолютної процентної помилки досить невелика (менше 10 %), що свідчить про високу прогностичну якість отриманої моделі експоненціального згладжування.

Exponential smoothing: S0=4,388 (Spr No trend, no season ; Alpha=,990 Курс акцій			
Case	Курс акцій	Smoothed Series	Resids
1	6,010000	4,388000	1,62200
2	5,980000	5,993780	-0,01378
3	5,980000	5,980138	-0,00014
4	5,980000	5,980001	-0,00000
5	5,600000	5,980000	-0,38000
6	4,550000	5,603800	-1,05380
7	4,550000	4,560538	-0,01054
8	4,200000	4,550105	-0,35011
9	3,900000	4,203501	-0,30350
10	4,200000	3,903035	0,29696
11	4,000000	4,197030	-0,19703
12	4,000000	4,001970	-0,00197
13	4,000000	4,000020	-0,00002
14	4,000000	4,000000	-0,00000
15	4,000000	4,000000	-0,00000
16	4,000000	4,000000	-0,00000
17	4,000000	4,000000	-0,00000
18	3,600000	4,000000	-0,40000
19	2,600000	3,604000	-1,00400
20	2,610000	2,610040	-0,00004
21		2,610000	
22		2,610000	
23		2,610000	
24		2,610000	
25		2,610000	

Рис. 5.8. Вихідні дані, згладжені дані, помилки

2. Далі необхідно визначити параметри моделі експоненціального згладжування із заданим параметром згладжування $\alpha = 0,9$. У ролі критерію якості моделі є середня абсолютна процентна помилка. У результаті розрахунків одержано модель з такими характеристиками (рис. 5.9 – 5.11).

Як видно з рис. 5.10, величина середньої абсолютної процентної помилки досить невелика (менше за 10 %), але отримана на підставі автоматичного пошуку параметра згладжування попередня модель має кращі прогностичні властивості.

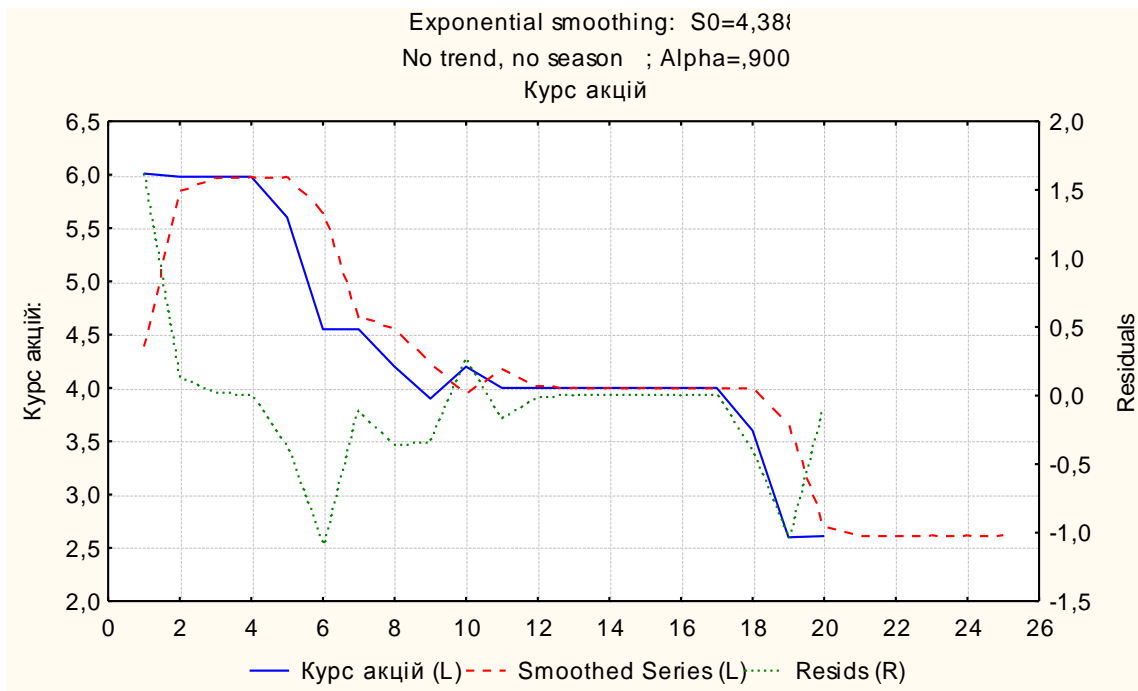


Рис. 5.9. Графік вихідних, згладжених і прогнозних значень курсу акцій

	Exponential smoo No trend, no seas Курс акцій
	Error
Summary of error	
Mean error	#####
Mean absolute error	0,301768875686
Sums of squares	5,583336561212
Mean square	0,279166828060
Mean percentage error	#####
Mean abs. perc. error	7,280472645208

Рис. 5.10. Оцінки якості моделі згладжування

3. Побудувати модель із урахуванням лінійного тренда, але без урахування сезонних чинників, тобто використовуючи процедуру згладжування за Холтом (рис. 5.12). Із метою пошуку оптимальних параметрів моделі слід використати вбудовану оптимізаційну процедуру. Буде отримано графік такої моделі (рис. 5.13) і відповідні характеристики (рис. 5.14 і 5.15).

Exponential smoothing: S0=4,388 No trend, no season ; Alpha=,900 Курс акцій			
Case	Курс акцій	Smoothed Series	Resids
1	6,010000	4,388000	1,622000
2	5,980000	5,847800	0,132200
3	5,980000	5,966780	0,013220
4	5,980000	5,978678	0,001322
5	5,600000	5,979868	-0,379868
6	4,550000	5,637987	-1,087987
7	4,550000	4,658799	-0,108800
8	4,200000	4,560880	-0,360880
9	3,900000	4,236080	-0,336080
10	4,200000	3,933609	0,266391
11	4,000000	4,173360	-0,173360
12	4,000000	4,017336	-0,017336
13	4,000000	4,001734	-0,001734
14	4,000000	4,000173	-0,000173
15	4,000000	4,000017	-0,000017
16	4,000000	4,000002	-0,000002
17	4,000000	4,000000	-0,000000
18	3,600000	4,000000	-0,400000
19	2,600000	3,640000	-1,040000
20	2,610000	2,704000	-0,094000
21		2,619400	
22		2,619400	
23		2,619400	
24		2,619400	
25		2,619400	

Рис. 5.11. Вихідні дані, згладжені дані, помилки моделі експоненціального згладжування з $\alpha = 0,9$

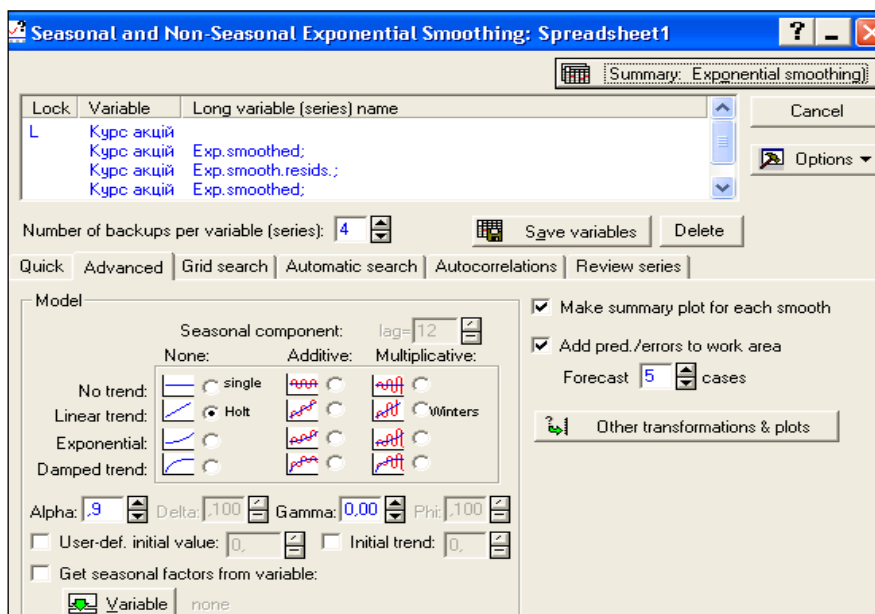


Рис. 5.12. Вибір моделі експоненціального згладжування за Холтом

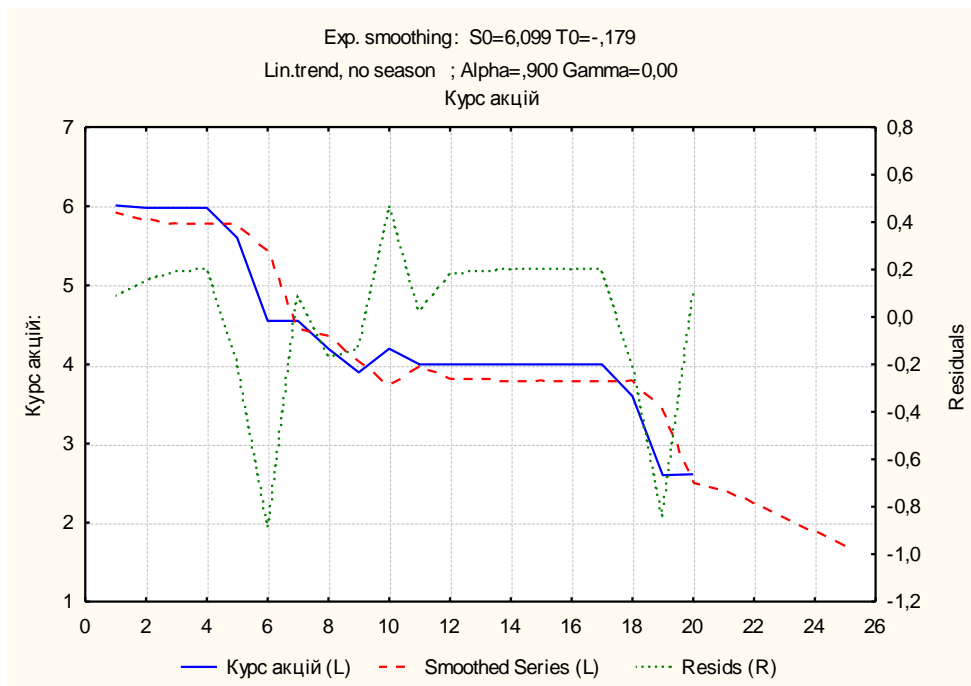


Рис. 5.13. Графік вихідних, згладжених і прогнозних значень курсу акцій

Exp. smoothing: S0=6,099 T0=-,179 Lin.trend, no season Курс акцій	
Error	
Summary of error	
Mean error	0,0043883690717
Mean absolute error	0,2455914327916
Sums of squares	2,1928515216265
Mean square	0,1096425760813
Mean percentage error	-0,5170679716612
Mean abs. perc. error	6,2925972676334

Рис. 5.14. Критерії якості моделі згладжування

Абсолютна процентна помилка невелика і відносно невеликі інші помилки (рис. 5.14), що говорить про досить високу прогностичну якість моделі.

4. Визначити параметри моделі експоненціального згладжування із заданими параметрами згладжування $\alpha_1 = 0,8$, $\alpha_2 = 0,3$. У ролі критерію якості моделі є середня абсолютна процентна помилка. У результаті розрахунків буде одержано модель із такими характеристиками (рис. 5.16 – 5.18).

Exp. smoothing: S0=6,099 T0=-,17 Lin.trend, no season ; Alpha=,900			
Курс акцій			
Case	Курс акцій	Smoothed Series	Resids
1	6,010000	5,920520	0,089476
2	5,980000	5,822100	0,157896
3	5,980000	5,785260	0,194733
4	5,980000	5,781570	0,198427
5	5,600000	5,781210	-0,181210
6	4,550000	5,439170	-0,889170
7	4,550000	4,459970	0,090030
8	4,200000	4,362050	-0,162050
9	3,900000	4,037250	-0,137250
10	4,200000	3,734770	0,465227
11	4,000000	3,974530	0,025470
12	4,000000	3,818500	0,181496
13	4,000000	3,802900	0,197096
14	4,000000	3,801340	0,198656
15	4,000000	3,801180	0,198816
16	4,000000	3,801170	0,198826
17	4,000000	3,801170	0,198830
18	3,600000	3,801170	-0,201170
19	2,600000	3,441170	-0,841170
20	2,610000	2,505170	0,104830
21		2,420570	
22		2,241620	
23		2,062670	
24		1,883720	
25		1,704780	

Рис. 5.15. Вихідні дані, згладжені дані, помилки за моделлю Холта

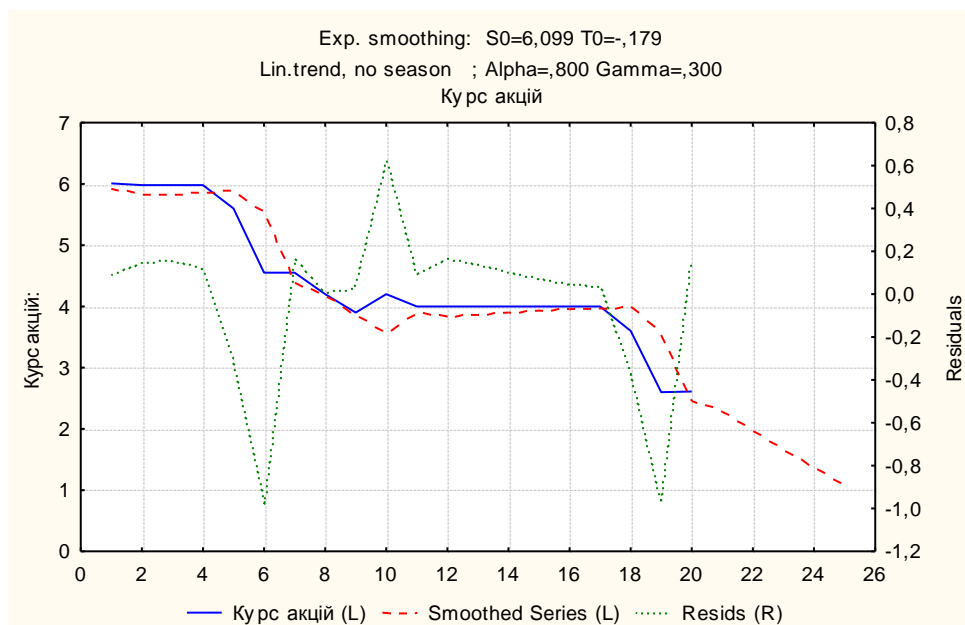


Рис. 5.16. Графік вихідних, згладжених і прогнозних значень курсу акцій

Exp. smoothing: S0=6,099 T0=-,177 Lin.trend, no season Курс акцій	
Error	
Summary of error	
Mean error	-0,0254618325619
Mean absolute error	0,2374700434202
Sums of squares	2,7136897692562
Mean square	0,1356844884628
Mean percentage error	-1,2565515153154
Mean abs. perc. error	6,2200475272717

Рис. 5.17. Критерії якості моделі згладжування

Exp. smoothing: S0=6,099 T0=-,177 Lin.trend, no season ; Alpha=,8000 Курс акцій			
Case	Курс акцій	Smoothed Series	Resids
1	6,010000	5,920520	0,089479
2	5,980000	5,834630	0,145369
3	5,980000	5,828340	0,151659
4	5,980000	5,863480	0,116519
5	5,600000	5,898470	-0,298470
6	4,550000	5,529830	-0,979830
7	4,550000	4,380950	0,169050
8	4,200000	4,191740	0,008259
9	3,900000	3,875880	0,024119
10	4,200000	3,578500	0,621500
11	4,000000	3,908180	0,091819
12	4,000000	3,836150	0,163849
13	4,000000	3,861070	0,138929
14	4,000000	3,899390	0,100609
15	4,000000	3,931200	0,068799
16	4,000000	3,954080	0,045920
17	4,000000	3,969670	0,030329
18	3,600000	3,980070	-0,380070
19	2,600000	3,570930	-0,970930
20	2,610000	2,456080	0,153919
21		2,278050	
22		1,976880	
23		1,675720	
24		1,374560	
25		1,073390	

Рис. 5.18. Вихідні дані, згладжені дані, помилки за моделлю Холта

Величина абсолютної процентної помилки невелика (рис. 5.17) – мінімальна у порівнянні з побудованими раніше моделями. Таким чином, найбільш високу точність прогнозу забезпечує модель Холта з параметрами згладжування $\alpha_1 = 0,8$, $\alpha_2 = 0,3$.

Змістовий модуль 2

Методи та моделі середньострокового, довгострокового прогнозування

Лабораторна робота 6

Реалізація моделей виділення та вирівнювання трендів

Мета – закріплення теоретичного та практичного матеріалу, формування навичок побудови нелінійних моделей тренда у модулі *Nonlinear Estimation/Advanced Linear/Nonlinear Models* ППП *Statistica*.

Умови завдання

У таблиці 6.1 наведені дані, що відображають динаміку чистого прибутку промислового підприємства за 10 років (y_t).

Таблиця 6.1

Динаміка чистого прибутку підприємства, млн грн

T	1	2	3	4	5	6	7	8	9	10
y_t	2	2,5	2,2	2,5	3,4	3,7	4	4,2	6	9,7

Необхідно

1. Побудувати графік залежності Y_t від t .
2. Виходячи з аналізу графіка зробити припущення щодо виду функціональної залежності між показниками Y_t і t .
3. У модулі *Nonlinear Estimation/Advanced Linear/Nonlinear Models* пакету *Statistica* побудувати 2 – 3 найбільш вірогідні залежності Y_t від часу.
4. Провести аналіз якості побудованих моделей та вибрати найкращу (оцінити статистичну значущість параметрів моделі тренда, проаналізувати значення коефіцієнтів кореляції та ін.).
5. Зробити прогноз на наступні 2 роки.

Методичні рекомендації до виконання завдання

На прикладі вихідних даних, що наведені в табл. 6.1, розглянути процес підбору та побудови нелінійної моделі тренда в пакеті *Statistica*.

Сформувати таблицю вихідних даних у ППП *Statistica* (рис. 6.1).

Для побудови графіка функціональної залежності обрати пункт головного меню *Graphs*, підпункт *Scatterplots* (точковий графік) (рис. 6.1). У вікні, що відкрилось, обрати тип графіка *Regular* (простий), та натискаючи на кнопку *Variables*, по осі ординат обрати змінну Y_t , по осі абсцис – змінну t .

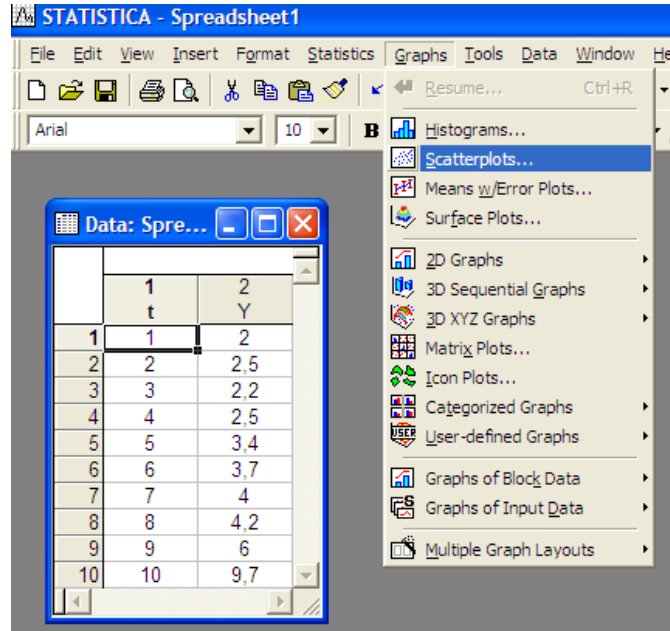


Рис. 6.1. Вихідні дані

У результаті отримано графік, що відображає залежність Y_t від часу (рис. 6.2).

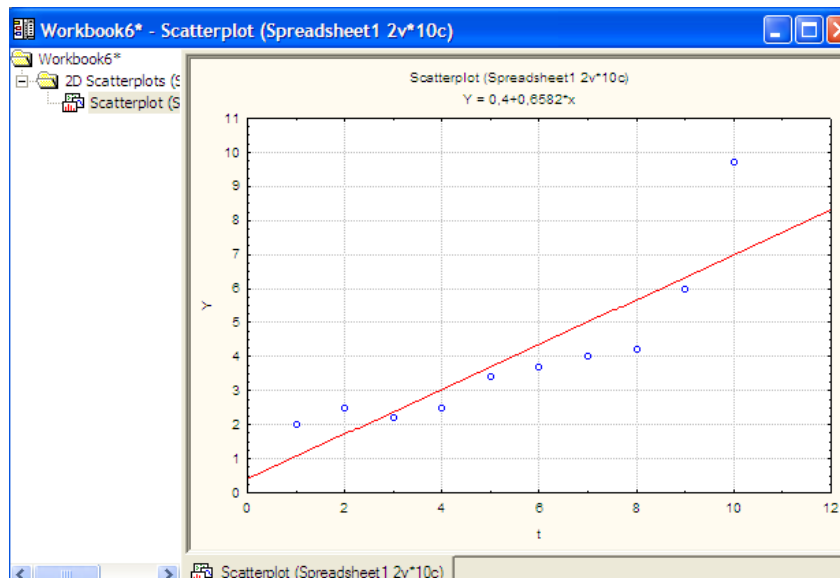


Рис. 6.2. Графік вихідних даних

Аналіз графіка вихідних даних дозволяє зробити висновок, що в якості конкуруючих варіантів моделі тренда можна виокремити такі криві росту: експонентна функція ($y = a_0 a_1^t$), степенева функція ($y = a_0 t^{a_1}$), поліном другого ступеня ($y = a_0 + a_1 t + a_2 t^2$).

Знайти параметри конкуруючих варіантів моделі тренда та оцінити їх прогнозні властивості.

Слід розглянути процес побудови експонентної функції ($y = a_0 a_1^t$). Для оцінювання параметрів моделі обрати пункт головного меню *Statistics*, підпункт *Nonlinear Estimation (нелінійне оцінювання)/Advanced Linear/Nonlinear Models (лінійні та нелінійні моделі)* (рис. 6.3).

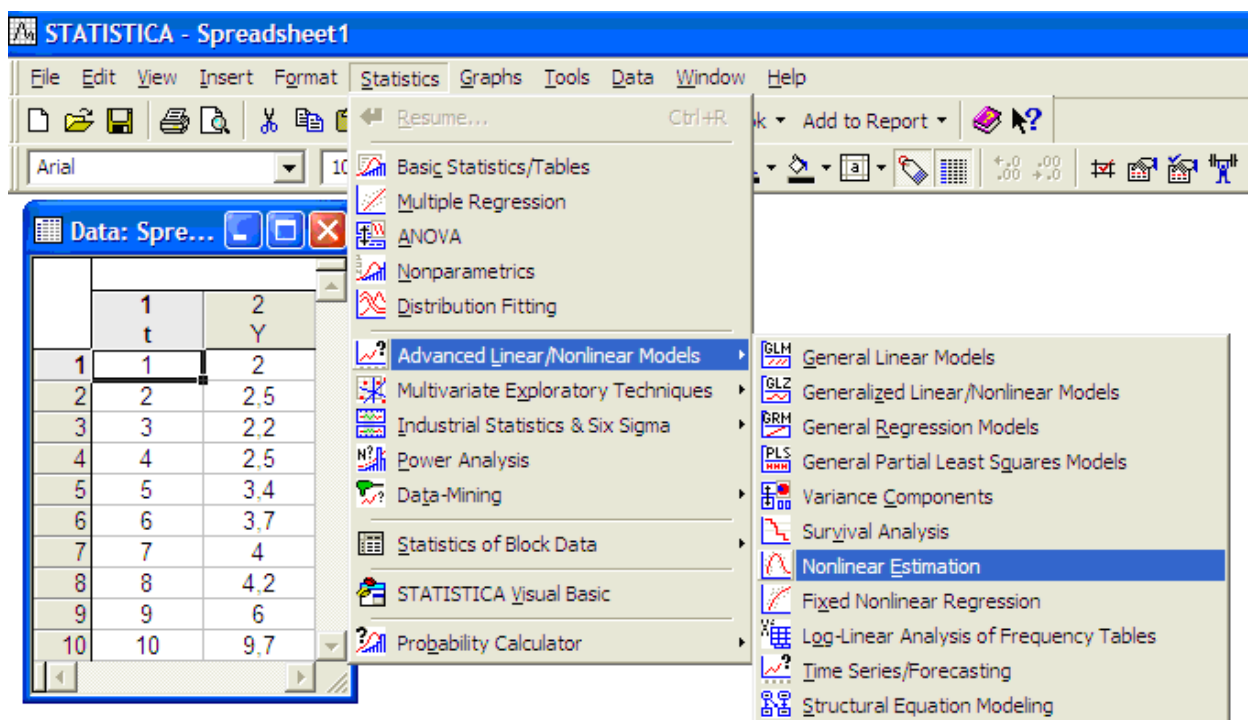


Рис. 6.3. Модуль нелінійного оцінювання

Вид стартової панелі модуля наведено на рис. 6.4. У вікні модуля подано такі види нелінійного оцінювання:

User – specified regression, least squares (визначена користувачем регресія з похибками за методом найменших квадратів);

User – specified regression, custom loss function (визначена користувачем регресія з заданою функцією похибок);

Quick Logit regression (логіт регресія);

Quick Probit regression (пробіт регресія);

Exponential growth regression (регресія експоненційного зростання);

Piecewise linear regression (кусково-лінійна регресія).

Для вирішення поставленої задачі необхідно вибрати опцію *User – specified regression, least squares* (рис. 6.4).

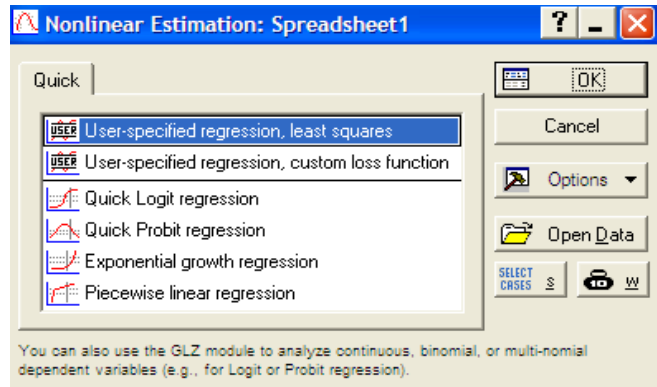


Рис. 6.4. Стартова панель модуля *Nonlinear Estimation*

Після чого у вікні *Function to be estimated* задати вид оцінюваної моделі тренда: $y = a_0 a_1^t$ (рис. 6.5). Слід використовувати спеціальні математичні позначки аналогічні до набору формул в *Excel*. Наприклад, експоненційна функція повинна бути задана як $Y = a0*a1^{**t}$ або $Y = a0*a1^t$.

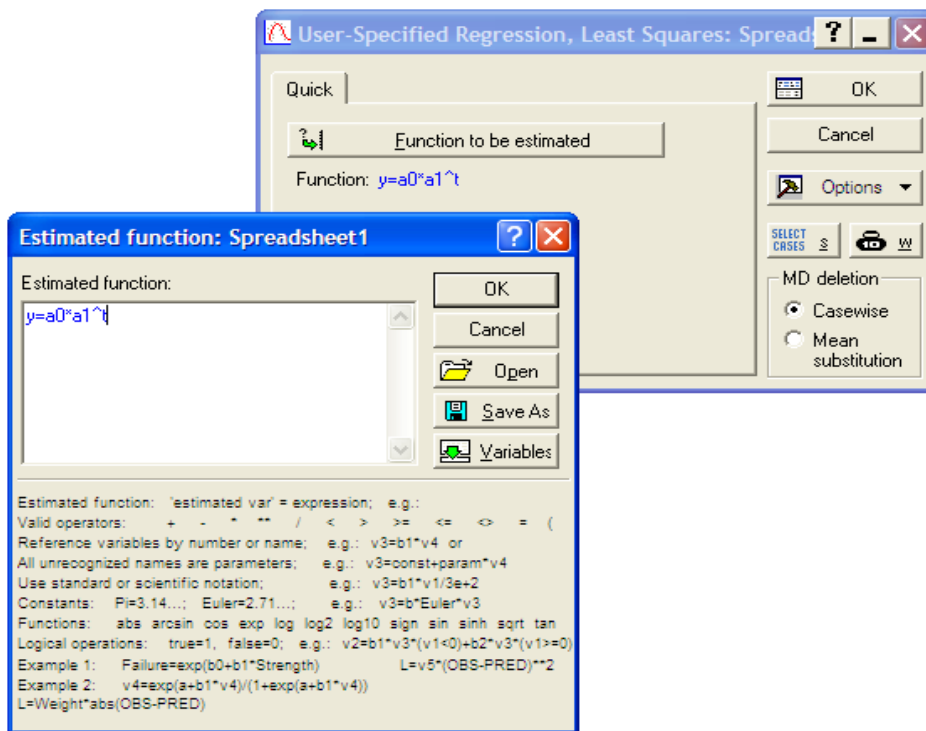


Рис. 6.5. Визначення виду функції оцінювання

У наступному вікні необхідно задати метод оцінювання параметрів нелінійної моделі тренда *Levenberg-Marquardt* (Левенберга – Марквардта) або *Gauss-Newton* (Гауса – Ньютона), за необхідності на вкладці *Advanced* можна визначити параметри ітераційної процедури та початкові значення оцінюваних параметрів і запустити процедуру оцінювання (рис. 6.6).

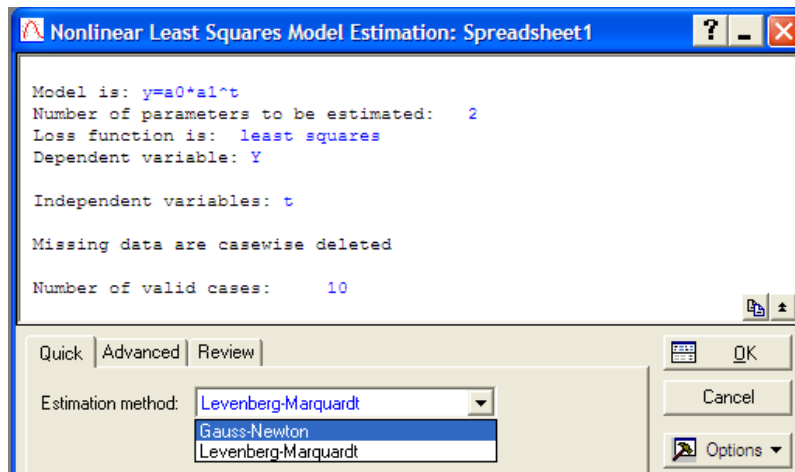


Рис. 6.6. Вибір методу оцінювання параметрів моделі

Результати побудови експоненційної моделі тренда можна проаналізувати за допомогою вікна, наведеного на рис. 6.7.

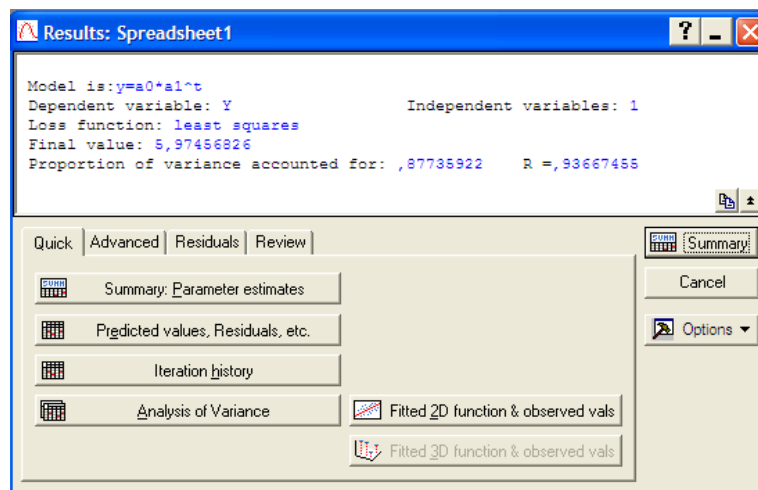


Рис. 6.7. Вікно результатів побудови експоненційної моделі тренда

У даному вікні подані результати моделі (вид досліджуваної моделі, залежна та незалежні змінні, вид функції оцінювання), у нижній інформаційній частині наведені опції для всебічного аналізу моделі.

Якість отриманої моделі оцінюється за допомогою значень: *Final value of Loss function* (Кінцеве значення функції втрат – Сума квадратів похибок моделі), *Proportion of variance accounted for* (Відсоток поясненої дисперсії), *R* (Коефіцієнт кореляції).

Оцінки параметрів моделі можна одержати, ініціювавши опцію *Quick/Summary: Parameter estimates* (Результат: параметри моделі) (рис. 6.7). Результат оцінювання параметрів наведено на рис. 6.8.

	Estimate	Standard error	t-value df = 8	p-level	Lo. Conf Limit	Up. Conf Limit
a0	1.101189	0.273196	4.03076	0.003784	0.471197	1.731181
a1	1.224090	0.036037	33.96767	0.000000	1.140989	1.307192

Рис. 6.8. Результат оцінювання параметрів моделі

У даному вікні наведені оцінки параметрів моделі (*Estimates*), значення середнє квадратичних відхилень параметрів (*Standard error*), статистики Стьюдента (*t-value*), рівень значимості статистики Стьюдента (*p-level*), нижня та верхня межі інтервальних оцінок параметрів (*Lo. Conf Limit; Up. Conf Limit*).

Для подальшого аналізу необхідно отримати теоретичні значення залежної змінної та похибки моделі, ініціювавши клавішу (*Predicted values, Residuals*) на вкладці *Quick*. Результати наведені на рис. 6.9.

	Observed	Predicted	Residuals
1	2.000000	1.347955	0.65205
2	2.500000	1.650018	0.84998
3	2.200000	2.019772	0.18023
4	2.500000	2.472383	0.02762
5	3.400000	3.026420	0.37358
6	3.700000	3.704612	-0.00461
7	4.000000	4.534780	-0.53478
8	4.200000	5.550981	-1.35098
9	6.000000	6.794902	-0.79490
10	9.700000	8.317575	1.38243

Рис. 6.9. Вихідні дані, прогнозовані дані та похибки моделі

Для порівняння емпіричних (досліджуваних) та теоретичних (розрахованих за моделлю) значень залежної змінної слід побудувати графік. Вибір опцій графічного аналізу отриманих результатів та типу графіка подані на рис. 6.10. На рис. 6.11 зображено графік порівняння теоретичних та емпіричних значень залежної змінної.

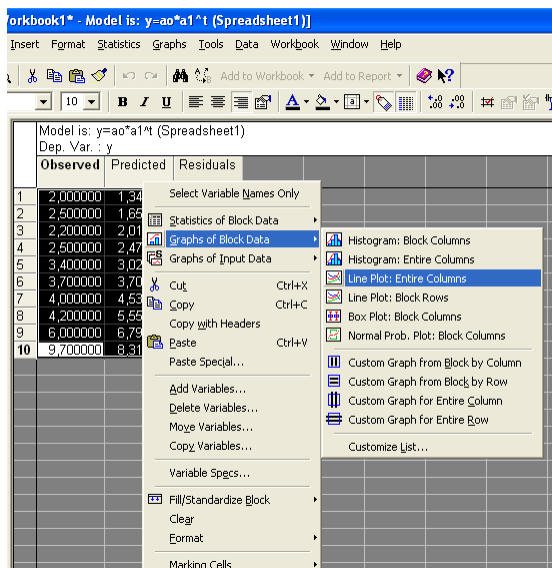


Рис. 6.10. Вибір опцій графічного аналізу змінних

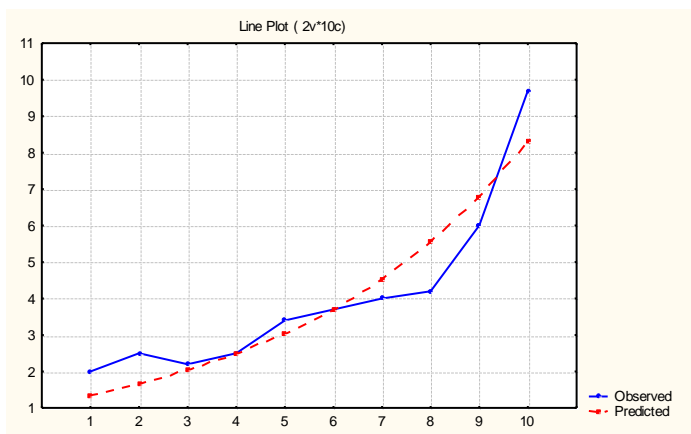


Рис. 6.11. Графік теоретичних та емпіричних значень залежної змінної

Всебічний аналіз помилок моделі можна отримати в опціях *Residuals* (Похибки) (рис. 6.12).

Гістограму помилок та графік розподілу помилок на нормальному ймовірнісному папері наведено на рис. 6.13, 6.14.

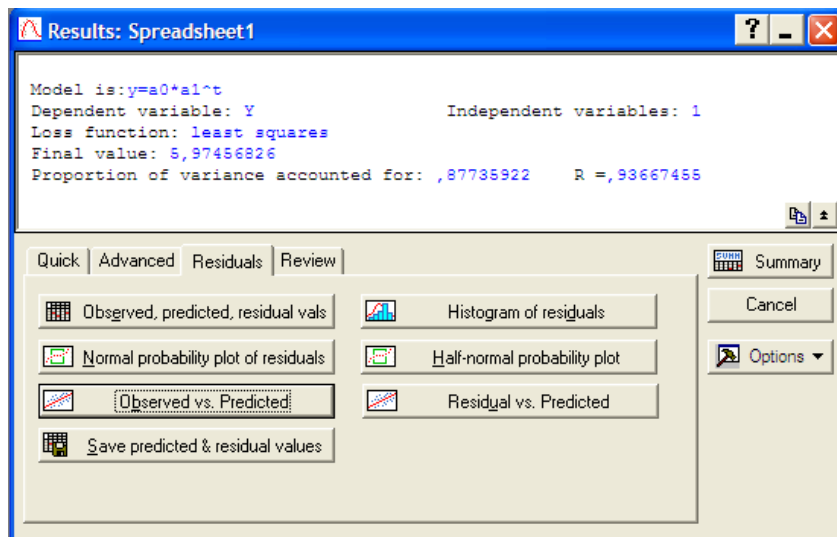


Рис. 6. 12. Меню аналізу помилок моделі

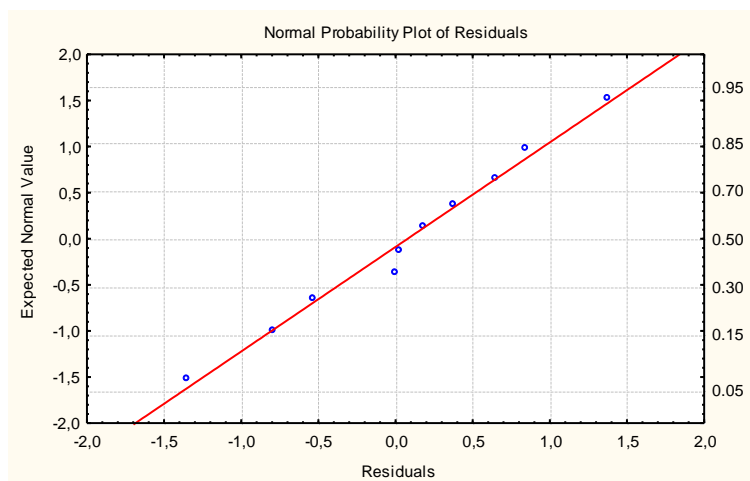


Рис. 6.13. Графік розподілу помилок на нормальному ймовірнісному папері

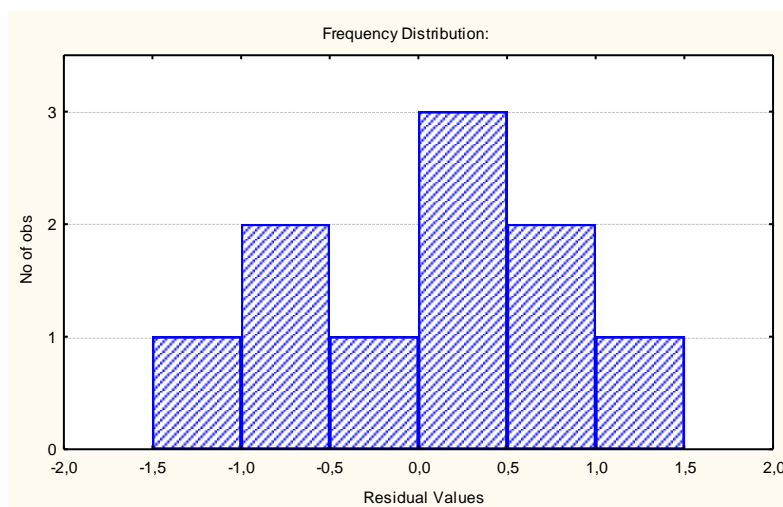


Рис. 6.14. Гістограма розподілу помилок

Таким чином, експоненційна модель тренда досліджуваного часового ряду має вид:

$$Y = 1,101 \cdot 1,224^t.$$

Виходячи з того, що індекс кореляції для моделі дорівнює $R=0,93667$, відсоток поясненої дисперсії – 87,73 %, параметри моделі статистично значущі за критерієм Стьюдента, можна зробити висновок, що побудована модель може бути використана для прогнозування.

Аналогічно визначаються характеристики степеневної та поліноміальної моделей тренда. Порівняльний аналіз якості конкуруючих варіантів моделі тренда наведено в табл. 6.2.

Таблиця 6.2

Порівняльний аналіз якості нелінійних моделей тренда

Вид моделі	Значення критерію Стьюдента	Коефіцієнт кореляції	Відсоток поясненої дисперсії
$Y = 1,101 \cdot 1,224^t$	$t_{a_0} = 4,03,$ $t_{a_1} = 33,97$	0,93667	87,73 %
$Y = 0,5119 \cdot t^{1,1649}$	$t_{a_0} = 1,5916,$ $t_{a_1} = 3,8339$	0,85614	73,29 %
$Y = 3,1417 - 0,7127t + 0,1246t^2$	$t_{a_0} = 3,23358,$ $t_{a_1} = 1,75629,$ $t_{a_2} = 3,46654$	0,94970	90,19 %

Аналіз даних, наведених у табл. 6.2, дозволяє зробити висновок, що найбільш високу прогностичну якість має експоненційна модель тренда. Слід знайти прогнозні значення чистого прибутку підприємства за цією моделлю. Для цього додати в таблиці з вихідними даними кількість рядків, рівну періоду упередження 2 та додаткову колонку Y для відображення прогнозу за моделлю. Розрахувати прогнозні значення показника чистого прибутку на 2 роки уперед можливо шляхом завдання в області специфікації змінної *Long Name* побудованої моделі тренда. Результати розрахунку прогнозних значень показника чистого прибутку підприємства наведено на рис. 6.15.

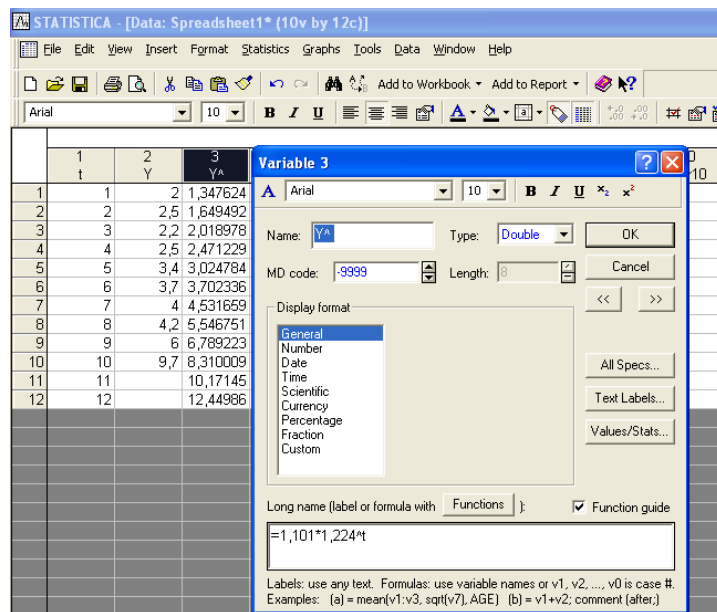


Рис. 6.15. Побудова прогнозу чистого прибутку підприємства на 2 роки вперед

Лабораторна робота 7

Моделі прогнозування багатомірних процесів

Мета – опанування студентами навичок побудови моделей багатомірних процесів, що містить змістовний опис моделі, оцінку та аналіз параметрів моделі, оцінку адекватності моделі та аналіз результатів моделювання, формування навичок побудови й аналізу багатофакторних моделей у модулі *Multiple Regression* ППП *Statistica*.

Умови завдання

Необхідно перевірити наявність лінійного множинного зв'язку між відповідними показниками в модулі *Multiple Regression* ППП *Statistica*:

1. Побудувати лінійну багатофакторну економетричну модель (включити усі відповідні фактори) і визначити всі її характеристики (параметри моделі, середні квадратичні відхилення параметрів моделі, дисперсію та середнє квадратичне відхилення похибок моделі, коефіцієнти множинної кореляції і детермінації).

2. Перевірити статистичну значимість параметрів моделі, коефіцієнта множинної кореляції. Перевірити адекватність моделі за допомогою критерію Фішера.

3. Навести таблиці з теоретичними значеннями залежного показника та значень похибок моделі. Побудувати графік лінійної функції з довірчими інтервалами. Знайти прогнозне значення залежної змінної та довірчі інтервали, якщо відомі дані про майбутні значення незалежних показників.

4. Перевірити модель на наявність мультиколінеарності. Навести матрицю парних кореляцій для факторних ознак. За методом Феррара – Глобера оцінити суттєвість мультиколінеарності.

5. Навести результати дослідження моделі за критерієм Дарбіна – Уотсона і нециклічного коефіцієнта автокореляції. Зробити висновки щодо наявності автокореляції. Побудувати гістограму і графік розподілу похибок. Навести групування даних за значеннями похибок, дати економічну інтерпретацію.

6. Зробити висновки щодо адекватності побудованої багатofакторної моделі, дати економічну інтерпретацію моделі в цілому.

7. Виключити з моделі фактори, які найменше впливають на залежну змінну або взаємозалежні між собою (використовувати результати критерію Стюдента та коефіцієнти парних кореляцій). Визначити всі вищевказані характеристики побудованих моделей, зробити висновки щодо їх адекватності.

8. Побудувати і дати інтерпретацію моделей, побудованих на основі методів покрокового включення і покрокового виключення змінних.

9. Якщо в моделі присутня мультиколінеарність, то для оцінювання параметрів використати метод рідж-регресії. Визначити всі характеристики моделі. Навести графіки зміни значень оцінок параметрів рідж-моделі залежно від значення параметра. Оцінити ступінь зсуву оцінок параметрів.

10. Зробити порівняльний аналіз побудованих моделей. Визначити найбільш адекватну й економічно інтерпретовану модель.

Методичні рекомендації до виконання завдання

Для побудови та всебічного аналізу множинних лінійних моделей у ППП *Statistica* передбачений модуль *Multiple Regression* (*Множинна регресія*). Слід розглянути порядок роботи в даному модулі у ході побудови багатofакторної моделі.

1. Запуск *Statistica* і підготовка даних.

У меню програм вибрати програму *Statistica*, після її запуску вибрати у меню пункт *File/New* для підготовки власних даних. Після уведення натиснути кнопку вікна *OK*. Після заповнення всіх комірок поля даних буде одержано таблицю, аналогічну наведеній на рис. 7.1.

	1 X1	2 X2	3 X3	4 Y
1	0,45	0,83	0,69	0,13
2	0,81	0,8	0,83	0,77
3	0,46	0,66	0,16	0,25
4	1,08	1,11	0,65	0,5
5	0,39	0,67	0,34	0,12
6	0,18	0,19	0,82	0,22
7	0,54	0,35	5,36	0,82
8	0,46	0,23	0,54	0,91
9	0,87	0,81	1,94	0,96
10	0,71	0,9	0,22	0,16
11	0,07	0,13	0,83	0,11
12	0,14	0,24	0,35	0,02
13	0,03	0,05	0,66	0,01
14	0,136	0,176	0,18	0,043
15	0,073	0,064	0,16	0,969
16	0,105	0,066	0,65	0,172
17	0,15	0,182	0,34	0,108
18	0,189	0,359	0,82	0,064
19	9,155	1,234	0,79	4,654
20	4,744	0,339	0,43	0,076
21	63,899	16,332	0,15	25,152
22	7,314	1,004	0,03	1,087
23	0,85	0,453	0,126	0,024
24	0,184	2,635	0,445	0,914
25	0,01	0,031	0,087	0,001

Рис. 7.1. Вихідні дані

де x_1 , x_2 , x_3 – факторні незалежні змінні;
 y – результуюча залежна змінна.

2. Побудова лінійної багатофакторної моделі.

Щоб почати обчислювальні процедури, необхідно увійти в позицію меню *Statistics/Multiple Regression*. Після підтвердження вибору модуля з'явиться стартова панель даного модуля, де необхідно задати змінні для аналізу.

Ініціювати кнопку *Variables (Змінні)* і у вікні, що з'явилося, вказати *Dependent* (залежну) і *Independent* (незалежну) змінні для побудови багатофакторної регресійної моделі. Вибір змінних подано на рис. 7.2. Після вказівки змінних підтвердити свій вибір натисканням кнопки *OK*.

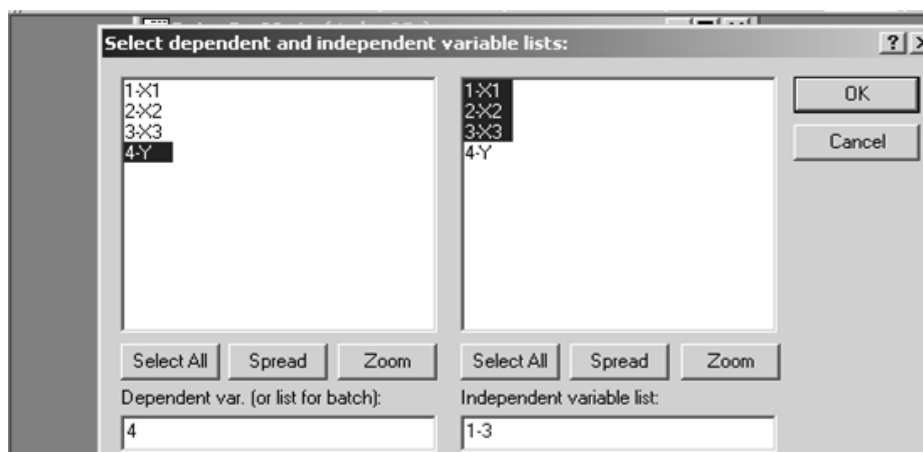


Рис. 7.2. Вибір змінних для аналізу

Побудувати лінійну багатофакторну модель і визначити всі її характеристики. Перевірити статистичну значимість параметрів моделі та адекватність моделі за критерієм Фішера. Результати побудови моделі наведені на рис. 7.3.

Regression Summary for Dependent Variable: Y (Spr38)						
R= ,99329417 R ² = ,98663330 Adjusted R ² = ,98472378						
F(3,21)=516,69 p<0,0000 Std.Error of estimate: ,61913						
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			-0,159728	0,165543	-0,964874	0,345589
X1	0,747946	0,119810	0,293843	0,047069	6,242776	0,000003
X2	0,253338	0,119732	0,396141	0,187223	2,115880	0,046472
X3	0,036745	0,025417	0,175450	0,121359	1,445707	0,163019

Рис. 7.3. Результати багатофакторного регресійного аналізу

Для перевірки гіпотези про значимість регресійної моделі використовується дисперсійний аналіз. Для цього необхідно ініціювати кнопку *Advanced/ANOVA*. Результати дисперсійного аналізу для досліджуваної моделі наведені на рис. 7.4. У даній таблиці наведено суму квадратів відхилень за регресією (*Sums of Squares Regress*), суму квадратів відхилень похибок (*Sums of Squares Residual*), дисперсію похибок (*Mean Squares Residual*) та критерій Фішера (*F*).

Analysis of Variance; DV: Y (Spreadsheet17)					
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	591,6353	1	591,6353	1283,945	0,000000
Residual	10,5983	23	0,4608		
Total	602,2336				

Рис. 7.4. Таблиця дисперсійного аналізу

3. Перевірка моделі на наявність мультиколінеарності.

Наступним кроком дослідження багатофакторної регресійної моделі є перевірка наявності мультиколінеарності в моделі.

Одним із способів перевірки моделі на мультиколінеарність є розрахунок матриці парних кореляцій. У меню аналізу моделі (рис. 7.5) ініціювати кнопку *Descriptive statistics/Correlations* (Описові статистики/Кореляція). Матриця коефіцієнтів парних кореляцій наведена на рис. 7.6.

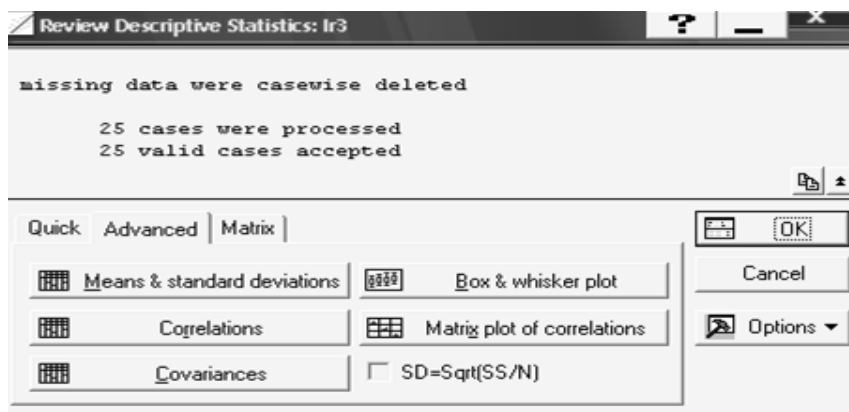


Рис. 7.5. Вікно вибору оцінки кореляційної залежності

Variable	Correlations (Spr38)			
	X1	X2	X3	Y
X1	1,000000	0,977544	-0,120649	0,991162
X2	0,977544	1,000000	-0,115205	0,980255
X3	-0,120649	-0,115205	1,000000	-0,082679
Y	0,991162	0,980255	-0,082679	1,000000

Рис. 7.6. Матриця коефіцієнтів парних кореляцій

Після ініціювання опції *Matrix plot of correlations* (Кореляційні графіки), будуть отримані гістограми і діаграми розсіву досліджуваних змінних у моделі (рис. 7.7).

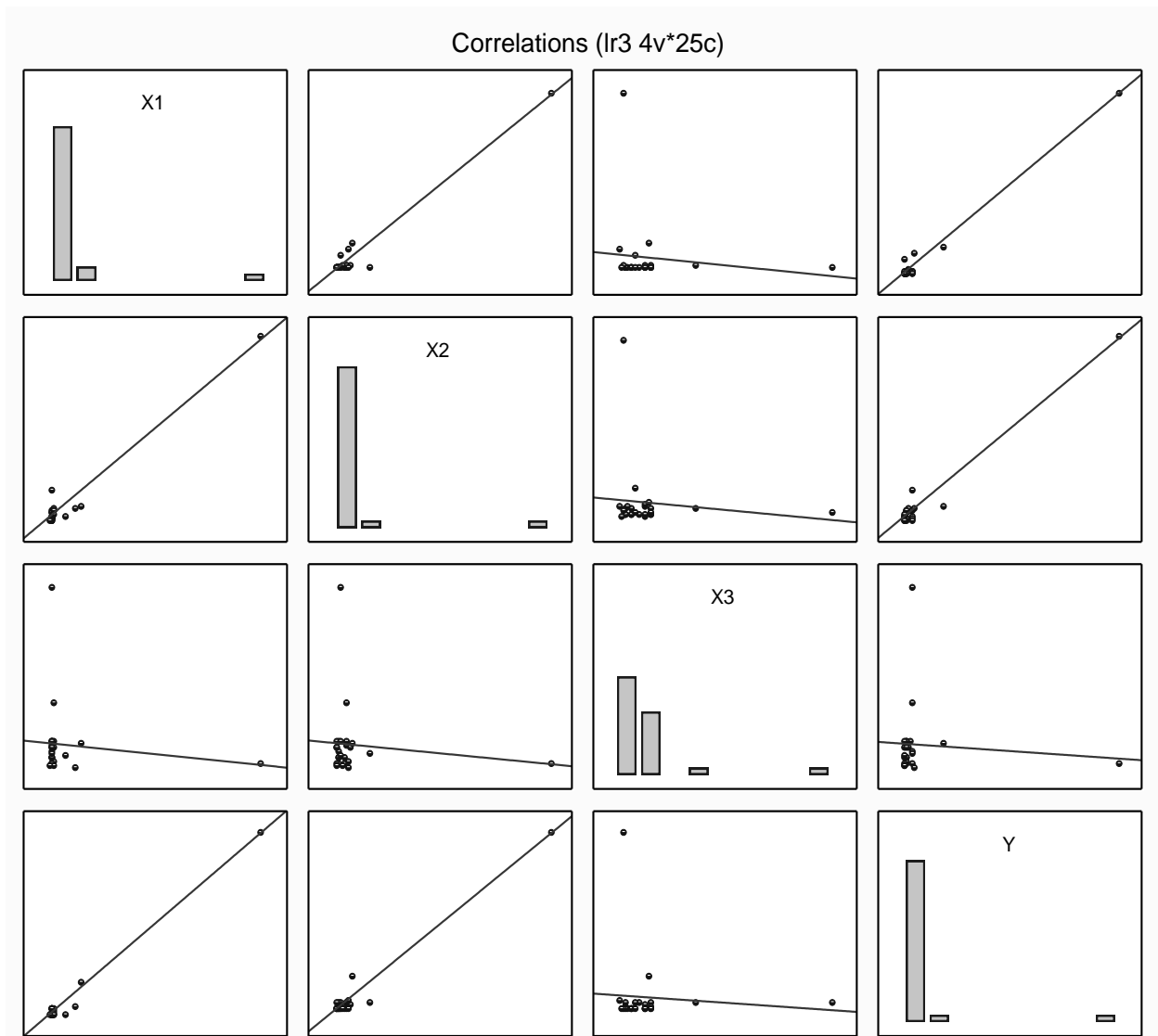


Рис. 7.7. Гістограми і діаграми розсіву досліджуваних змінних

Коефіцієнти парних кореляцій та діаграми розсіву свідчать про сильний ступінь лінійного взаємозв'язку між досліджуваними парами змінних.

Більш детальний аналіз передбачає розрахунок частинних кореляцій та розрахунок критерію максимальної спряженості для кожної з незалежних змінних. Ініціювавши опцію *Partial correlations* (Частинні кореляції) (рис. 7.8), буде отримана така таблиця (рис. 7.9). Дана таблиця містить значення стандартизованих коефіцієнтів регресії (*Beta in*); частинних кореляцій (*Partial correlations*), що відображають ступінь впливу кожної незалежної змінної на результуючу, за умови, що інші змінні не впливають на даний зв'язок; півчастинних кореляцій (*Semipartial correlations*);

коефіцієнт детермінації (*R square*) між даною змінною і другими незалежними змінними, які входять у рівняння регресії, що відображає міру максимальної спряженості; толерантність моделі (*Tolerance*), яка розраховується як $(1 - R\ square)$; значення критерію Стюдента (*t*) для перевірки гіпотези про значимість частинних коефіцієнтів кореляції з числом ступенів волі; рівень значимості (*p - level*); імовірність відхилення гіпотези про значимість частинних коефіцієнтів кореляції.

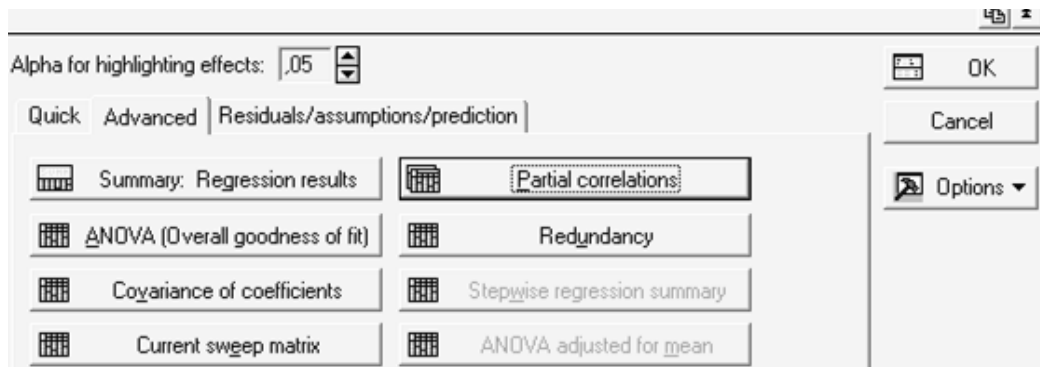


Рис. 7.8. Вікно вибору оцінки значимості незалежних змінних

Variable	Variables currently in the Equation; DV: Y (Spreadsheet17)						
	Beta in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(21)	p-level
X1	0,747946	0,806125	0,157500	0,044342	0,955658	6,242776	0,000003
X2	0,253338	0,419196	0,053382	0,044400	0,955600	2,115880	0,046472
X3	0,036745	0,300862	0,036474	0,985276	0,014724	1,445707	0,163019

Рис. 7.9. Значимість змінних у рівнянні регресії

Оцінювання ступеня впливу незалежних змінних на результуючий показник можливо виконати натисканням клавіші *Redundancy* (*Збитковість*), результати якого наведено на рис. 7.10.

Variable	Redundancy of Independent Variables; R-square column contains R-square of variable with all other independent variables			
	Toleran.	R-square	Partial Cor.	Semipart Cor.
X1	0,044342	0,955658	0,806125	0,157500
X2	0,044400	0,955600	0,419196	0,053382
X3	0,985276	0,014724	0,300862	0,036474

Рис. 7.10. Оцінювання надміру незалежних змінних

Для оцінювання ступеня мультиколінеарності за алгоритмом Феррара – Глобера використовуються частинні коефіцієнти кореляції між факторними змінними та їх статистична значимість, для їх розрахунку необхідно дослідити модель без залежної змінної, зробивши одну з факторних змінних залежною, та визначити дані характеристики (рис. 7.11).

Variables currently in the Equation; DV: X1 (I3)							
Variable	Beta in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(22)	p-level
X2	0,976606	0,977242	0,970104	0,986728	0,013272	21,60825	0,000000
X3	-0,008139	-0,038366	-0,008085	0,986728	0,013272	-0,18008	0,858735

Variables currently in the Equation; DV: X2 (I3)							
Variable	Beta in	Partial Cor.	Semipart Cor.	Tolerance	R-square	t(22)	p-level
X1	0,977879	0,977242	0,970736	0,985444	0,014556	21,60825	0,000000
X3	0,002775	0,013073	0,002755	0,985444	0,014556	0,06132	0,951656

Рис. 7.11. Оцінювання рівня зв'язку між незалежними змінними

Таким чином, значення коефіцієнтів частинних кореляцій дорівнює: $r_{12} = 0,9766$, $r_{13} = -0,0081$, $r_{23} = 0,002775$. Значимість часткових коефіцієнтів кореляції за допомогою t – критерію Стьюдента: $t_{12} = 21,60825$, $t_{13} = -0,18008$, $t_{23} = 0,06132$. Значення критеріїв t_{kj} порівняти з табличним при $(n - m)$ ступенях вільності і рівні значимості α . Якщо $t_{kj} > t_{\text{табл}}$, то між незалежними змінними x_k та x_j існує мультиколінеарність.

$t_{\text{табл}}(0.05,22) = 2,07$, отже $t_{12} > t_{\text{табл}}$, то можемо зробити висновок, що між змінними x_1 та x_2 існує тісний лінійний зв'язок (мультиколінеарність).

4. Розрахунок та дослідження помилок моделі.

Для проведення даного аналізу необхідно розрахувати теоретичні значення залежної змінної та помилки моделі, навести результати дослідження моделі за критерієм Дарбіна – Уотсона й нециклічного коефіцієнта автокореляції та зробити висновки про наявність автокореляції.

Щоб розрахувати й проаналізувати залишки, у нижній частині вікна результатів регресійного аналізу є опція *Perform residual analysis (Всебічний аналіз залишків)*. Ініціювавши дану опцію, буде одержано меню для аналізу помилок.

Ініціювавши клавішу аналізу помилок *Summary: Residuals & Predicted*, буде отримано таблицю зі спостережуваними значеннями залежної змінної (*Observed value*), теоретичними значеннями залежної змінної (*Predicted value*) і помилками моделі (*Residual*) та графіком розсіву помилок моделі в діапазоні $\pm 3\sigma$, який аналізує властивість сталості дисперсії помилок (рис. 7.12).

Raw Residuals							Raw Residual (lr3)			
							Dependent variable: Y			
Case	-3s	.	.	0	.	.	+3s	Observed Value	Predicted Value	Residual
1	.	.	.	*	.	.	.	0,13000	0,42236	-0,29236
2	*	.	.	0,77000	0,54082	0,22918
3	.	.	.	*	.	.	.	0,25000	0,26496	-0,01496
4	.	.	.	*	.	.	.	0,50000	0,71138	-0,21138
5	.	.	.	*	.	.	.	0,12000	0,27994	-0,15994
6	*	.	.	0,22000	0,11230	0,10770
7	.	.	.	*	.	.	.	0,82000	1,07801	-0,25801
8	*	.	0,91000	0,16130	0,74870
9	*	.	.	0,96000	0,75716	0,20284
10	.	.	.	*	.	.	.	0,16000	0,44403	-0,28403
11	.	.	.	*	.	.	.	0,11000	0,05796	0,05204
12	.	.	.	*	.	.	.	0,02000	0,03789	-0,01789
13	.	.	.	*	.	.	.	0,01000	-0,01531	0,02531
14	*	.	.	0,04300	-0,01846	0,06146
15	*	.	0,96900	-0,08485	1,05385
16	*	.	.	0,17200	0,01131	0,16069
17	*	.	.	0,10800	0,01610	0,09190
18	.	.	.	*	.	.	.	0,06400	0,18189	-0,11789
19	*	4,65400	3,15785	1,49615
20	.	*	0,07600	1,44400	-1,36800
21	.	.	.	*	.	.	.	25,15200	25,11234	0,03966
22	.	*	1,08700	2,39243	-1,30543
23	.	.	.	*	.	.	.	0,02400	0,29160	-0,26760
24	.	.	.	*	.	.	.	0,91400	1,01625	-0,10224
25	*	.	.	0,00100	-0,12924	0,13024
Minimum	.	*	0,00100	-0,12924	-1,36800
Maximum	*	25,15200	25,11234	1,49615
Mean	.	.	.	*	.	.	.	1,52976	1,52976	0,00000
Median	.	.	.	*	.	.	.	0,17200	0,27994	0,02531

Рис. 7.12. Аналіз помилок моделі та графік їх розсіву

У меню аналізу помилок, ініціювавши кнопку *Advanced/Durbin-Watson statistic* (Статистика Дарбіна – Уотсона), буде одержано значення автокореляції помилок моделі за критерієм Дарбіна – Уотсона та значення нециклічного коефіцієнта автокореляції (рис. 7.13).

Durbin-Watson d (Spi and serial correlation)		
	Durbin-Watson d	Serial Corr.
Estimate	2,46423	-0,23898

Рис. 7.13. Автокореляція помилок моделі

Значення отриманих коефіцієнтів порівнюються з табличними значеннями і робляться висновки про наявність у моделі автокореляції помилок. Для подальшого всебічного аналізу помилок необхідно побудувати гістограму й графік розподілу помилок на нормальному ймовірнісному папері (рис. 7.14, 7.15).

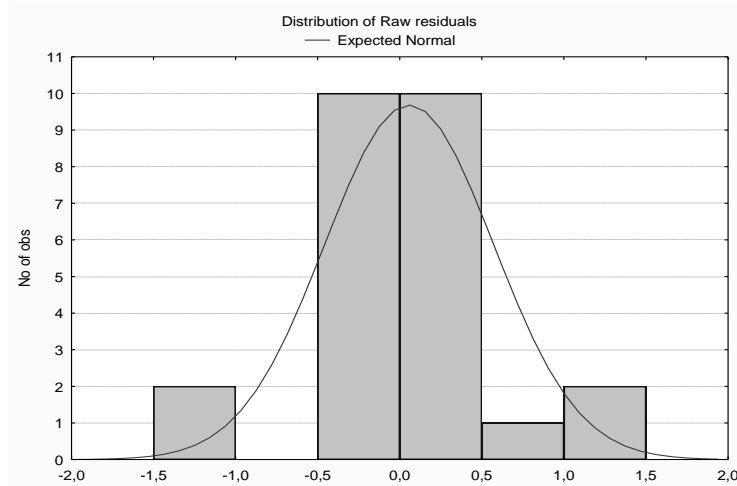


Рис. 7.14. Гістограма розподілу помилок

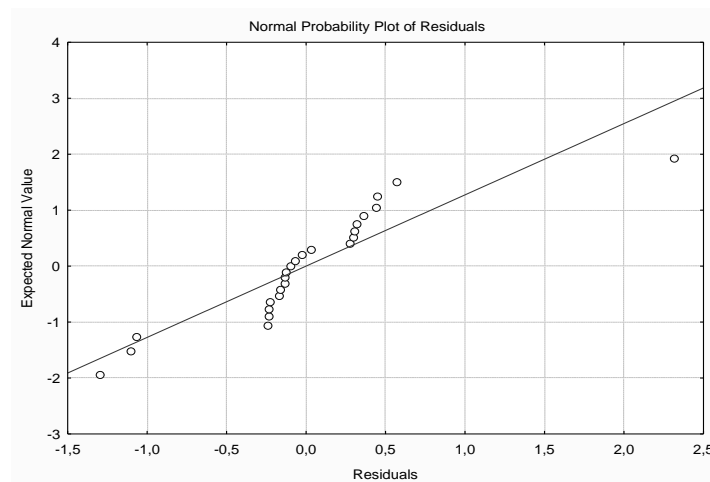


Рис. 7.15. Графік розподілу помилок на нормальному ймовірнісному папері

За отриманими результатами необхідно згрупувати об'єкти за додатними та від'ємними значеннями помилок, дати економічну інтерпретацію даному угрупованню.

5. Побудова моделей за методами покрокового включення та покрокового виключення незалежних змінних.

В умовах мультиколінеарності незалежних змінних ефективним методом оцінювання параметрів економетричних моделей є реалізація покрокової регресії, яка передбачає оцінювання параметрів моделі через коефіцієнти кореляцій.

У модулі *Multiple Regression* реалізовано метод покрокового включення змінних (*Forward stepwise*) і метод покрокового виключення (*Backward stepwise*). Вибір методів здійснюється на стартовій панелі у меню *Advanced* ініціюванням опції *Advanced options (stepwise or ridge regression)* (рис. 7.16).

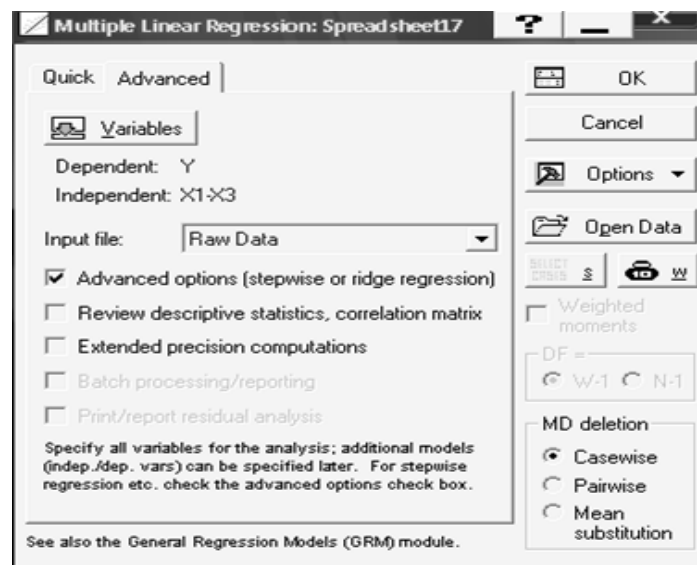


Рис. 7.16. Вибір методу покрокової регресії

Вибір методу оцінювання, порогові значення F-критерію включення або виключення, послідовність подання результатів вибирається у вкладці *Stepwise* (рис. 7.17).

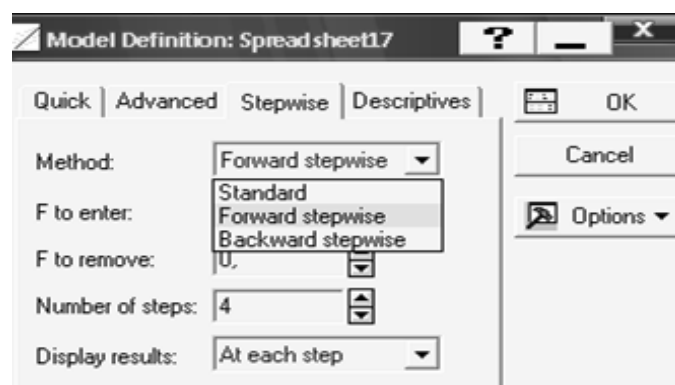


Рис. 7.17. Вибір методів оцінювання параметрів

Послідовність етапів реалізації алгоритму покрокового включення (*Forward stepwise*) подано на рис. 7.18.

```

Multiple Regression Results (Step 0)
Dependent: Y           Multiple R = 0,00000000    F = 0,000000
                        R?= 0,00000000          df = 0,24
No. of cases: 25      adjusted R?= 0,00000000      p = -0,00000
                        Standard error of estimate: 5,009297816
Step 0: No variables in the regression equation

Multiple Regression Results (Step 1)
Dependent: Y           Multiple R = ,99116180    F = 1283,945
                        R?= ,98240171          df = 1,23
No. of cases: 25      adjusted R?= ,98163656      p = 0,000000
                        Standard error of estimate: ,678818769
Intercept: ,081243426 Std.Error: ,1416544 t( 23) = ,57353 p = ,5719
                        X1 beta=,991

Multiple Regression Results (Step 2)
Dependent: Y           Multiple R = ,99262428    F = 737,4498
                        R?= ,98530296          df = 2,22
No. of cases: 25      adjusted R?= ,98396686      p = 0,000000
                        Standard error of estimate: ,634287475
Intercept: -,030737796 Std.Error: ,1428533 t( 22) = -,2152 p = ,8316
                        X1 beta=,741          X2 beta=,256

Multiple Regression Results (step 3, final solution)
no other F to enter exceeds specified limit
Dependent: Y           Multiple R = ,99329417    F = 516,6896
                        R?= ,98663330          df = 3,21
No. of cases: 25      adjusted R?= ,98472378      p = 0,000000
                        Standard error of estimate: ,619134289
Intercept: -,159727778 Std.Error: ,1655426 t( 21) = -,9649 p = ,3456
                        X1 beta=,748          X2 beta=,253          X3 beta=,037

Multiple Regression Results (step 3, final solution)
no other F to enter exceeds specified limit
Dependent: Y           Multiple R = ,99329417    F = 516,6896
                        R?= ,98663330          df = 3,21
No. of cases: 25      adjusted R?= ,98472378      p = 0,000000
                        Standard error of estimate: ,619134289
Intercept: -,159727778 Std.Error: ,1655426 t( 21) = -,9649 p = ,3456
                        X1 beta=,748          X2 beta=,253          X3 beta=,037

```

Рис. 7.18. Реалізація моделі покрокового включення змінних

Після закінчення процедури покрокової регресії в нижній інформаційній частині вікна стане активною опція *Stepwise regression summary* (*Результати покрокової регресії*) (рис. 7.19).

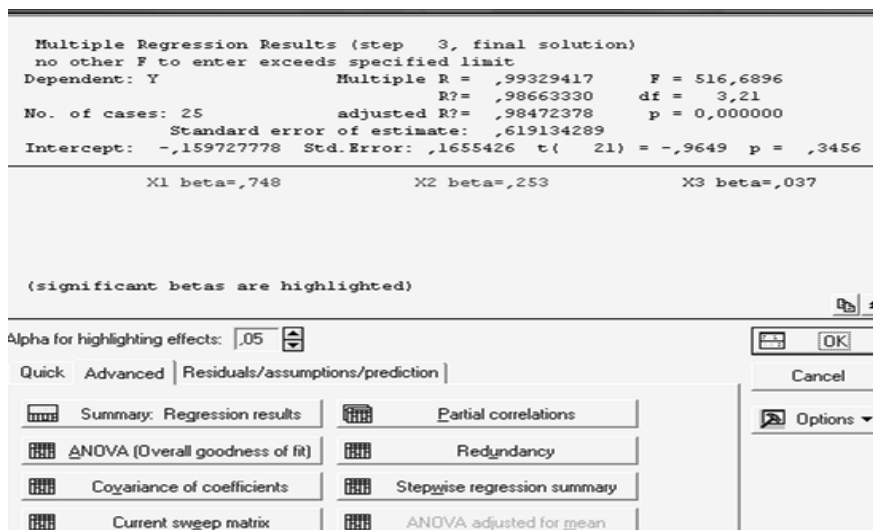


Рис. 7.19. Вікно вибору результатів покрокової регресії

Натиснувши дану опцію, буде отримано таблицю результатів по-крокової регресії, наведена на рис. 7.20, де відображено адекватність моделі на кожному з етапів та зміну характеристик моделі для кожного кроку.

Summary of Stepwise Regression; DV: Y (lr3)							
Variable	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
X1	1	0,991162	0,982402	0,982402	1283,945	0,000000	1
X2	2	0,992624	0,985303	0,002901	4,343	0,048989	2
X3	3	0,993294	0,986633	0,001330	2,090	0,163019	3

Рис. 7.20. Результати покрокової регресії включення

Аналіз результатів регресійної моделі методом покрокового включення змінних моделі (адекватність, статистична значимість) буде отримано після ініціювання опції *Summary: Regression results*, яку наведено на рис. 7.21.

Regression Summary for Dependent Variable: Y (Spreadsheet17)						
R= ,99329417 R ² = ,98663330 Adjusted R ² = ,98472378 F(3,21)=516,69 p<0,0000 Std.Error of estimate: ,61913						
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			-0,159728	0,165543	-0,964874	0,345589
X1	0,747946	0,119810	0,293843	0,047069	6,242776	0,000003
X2	0,253338	0,119732	0,396141	0,187223	2,115880	0,046472
X3	0,036745	0,025417	0,175450	0,121359	1,445707	0,163019

Рис. 7.21. Регресійна модель методом покрокового включення

Послідовність етапів реалізації алгоритму покрокового виключення (*Backward stepwise*) наведено на рис. 7.22.

```

Multiple Regression Results (Step 0)
Dependent: Y                Multiple R = ,99329417    F = 516,6896
                             R2= ,98663330    df = 3,21
No. of cases: 25            adjusted R2= ,98472378    p = 0,000000
                             Standard error of estimate: ,619134289
Intercept: -,159727778 Std.Error: ,1655426 t( 21) = -,9649 p = ,3456
X1 beta=,748                X2 beta=,253                X3 beta=,037

Multiple Regression Results (Step 1)
Dependent: Y                Multiple R = ,99262428    F = 737,4498
                             R2= ,98530296    df = 2,22
No. of cases: 25            adjusted R2= ,98396686    p = 0,000000
                             Standard error of estimate: ,634287475
Intercept: -,030737796 Std.Error: ,1428533 t( 22) = -,2152 p = ,8316
X1 beta=,741                X2 beta=,256

Multiple Regression Results (step 2, final solution)
no other F to remove is less than specified limit
Dependent: Y                Multiple R = ,99116180    F = 1283,945
                             R2= ,98240171    df = 1,23
No. of cases: 25            adjusted R2= ,98163656    p = 0,000000
                             Standard error of estimate: ,678818769
Intercept: ,081243426 Std.Error: ,1416544 t( 23) = ,57353 p = ,5719
X1 beta=,991

```

Рис. 7.22. Реалізація моделі покрокового виключення змінних

Натиснувши опцію *Stepwise regression summary*, буде отримана таблиця результатів покрокової регресії виключення для кожного етапу, подана на рис. 7.23.

Summary of Stepwise Regression; DV: Y (lr3)							
Variable	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
X3	-1	0,992624	0,985303	-0,001330	2,090069	0,163019	2
X2	-2	0,991162	0,982402	-0,002901	4,342880	0,048989	1

Рис. 7.23. Результати покрокової регресії виключення

Регресійна модель методом покрокового виключення змінних подана на рис. 7.24.

Regression Summary for Dependent Variable: Y (Spreadsheet1)						
R= ,99116180 R ² = ,98240171 Adjusted R ² = ,98163656 F(1,23)=1283,9 p<0,0000 Std.Error of estimate: ,67882						
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(23)	p-level
Intercept			0,081243	0,141654	0,57353	0,571850
X1	0,991162	0,027661	0,389395	0,010867	35,83218	0,000000

Рис. 7.24. Регресійна модель методом покрокового виключення

Аналіз побудованих моделей дозволить вибрати найбільш адекватну модель із точки зору опису реальних економічних процесів, їх взаємозв'язків та цілей вирішення задач.

6. Побудова моделі на основі методу рідж-регресії, звільнення від мультиколінеарності.

Одним із методів, що дозволяє скорегувати матрицю незалежних змінних за наявності в моделі мультиколінеарності, є метод рідж-регресії. У модулі *Multiple Regression* реалізацію даного методу можна здійснити на стартовій панелі у меню *Advanced* ініціюванням опції *Advanced options (stepwise or ridge regression)*.

Для вибору методу рідж-регресії і параметра зсуву λ у меню *Advanced* необхідно ініціювати опцію *Ridge regression; lambda* (рис. 7.25).

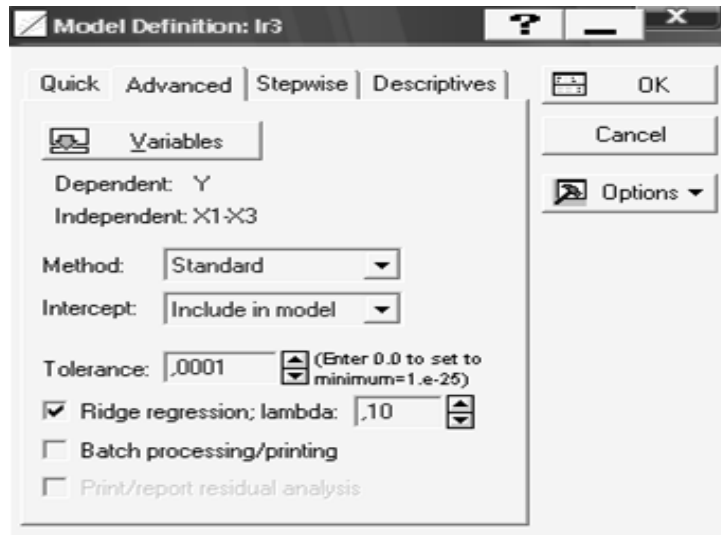


Рис. 7.25. Вибір методу рідж-регресії і параметра зсуву λ

Результати регресійних моделей методом рідж-регресії (адекватність, статистична значимість) з різними значеннями параметра зсуву λ можна отримати ініціювавши опцію *Summary: Regression results*, які наведено на рис. 7.26 та 7.27.

Ridge Regression Summary for Dependent Variable: Y (lr3)						
l=,10000 R= ,96780176 R?= ,93664024 Adjusted R?= ,92758884 F(3,21)=103,48 p<,00000 Std.Error of estimate: 1,3480						
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			-0,127295	0,342878	-0,371254	0,714167
X1	0,521135	0,114292	0,204737	0,044902	4,559671	0,000171
X2	0,430861	0,114231	0,673730	0,178621	3,771846	0,001120
X3	0,027120	0,052697	0,129493	0,251613	0,514651	0,612171

Рис. 7.26. Рідж-регресія при $\lambda=0,1$

Ridge Regression Summary for Dependent Variable: Y (lr3)						
l=,90000 R= ,82181444 R?= ,67537898 Adjusted R?= ,62900454 F(3,21)=14,564 p<,00002 Std.Error of estimate: 3,0511						
N=25	Beta	Std.Err. of Beta	B	Std.Err. of B	t(21)	p-level
Intercept			0,395520	0,709551	0,557422	0,583133
X1	0,348421	0,105264	0,136883	0,041355	3,309961	0,003331
X2	0,336603	0,105246	0,526340	0,164571	3,198262	0,004321
X3	-0,000981	0,090430	-0,004685	0,431777	-0,010850	0,991446

Рис. 7.27. Рідж-регресія при $\lambda=0,9$

Дослідження ступеня зміни параметрів моделі з різними значеннями параметра зсуву λ наведено в табл. 7.1.

Таблиця 7.1

Параметри моделі рідж-регресії

λ	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,99
a0	-0,160	-0,127	-0,039	0,043	0,117	0,183	0,244	0,299	0,349	0,396	0,434
a1	0,294	0,205	0,188	0,177	0,168	0,161	0,154	0,148	0,142	0,137	0,133
a2	0,396	0,674	0,671	0,651	0,629	0,606	0,585	0,564	0,545	0,526	0,511
a3	0,175	0,129	0,097	0,072	0,052	0,036	0,023	0,012	0,003	-0,005	-0,010

Графік гребневого сліду оцінок параметрів, який відображує ступінь зміщення оцінок регресійної моделі, наведено на рис. 7.28.

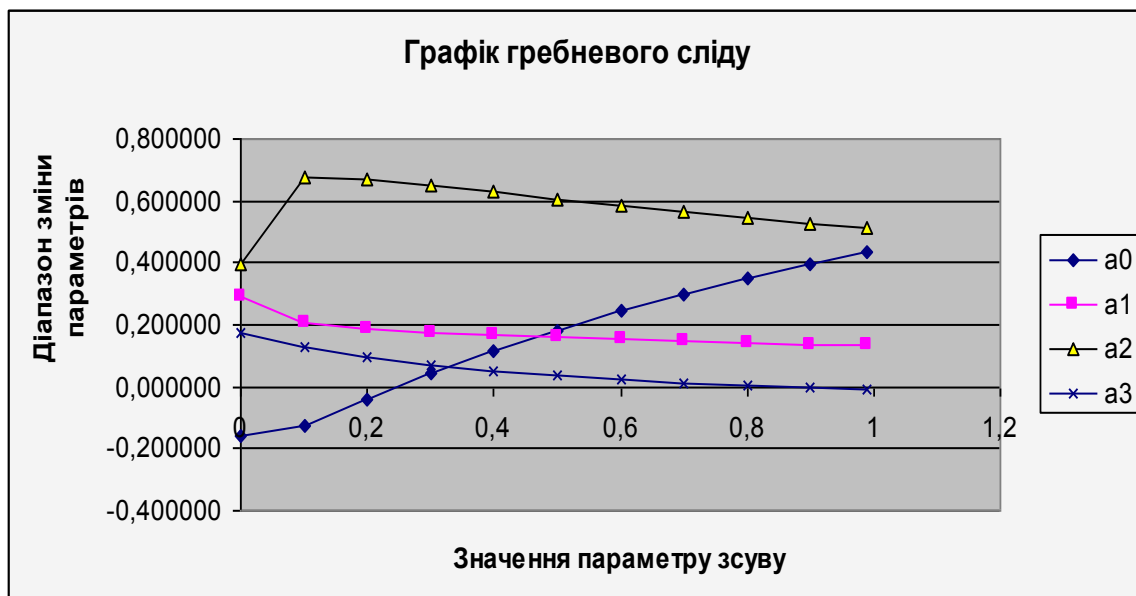


Рис. 7.28. Графік гребневого сліду оцінок параметрів

7. Прогнозування.

Оскільки побудовані моделі є адекватними, їх параметри статистично значимі, то за моделлю можна скласти прогнози. Щоб розрахувати прогнозні значення залежної змінної, у нижній частині вікна результатів регресійного аналізу є опція *Predict dependent variable* (Прогнозування залежної змінної), ініціювавши дану опцію, необхідно вказати значення незалежних змінних, для яких необхідно спрогнозувати залежну величину.

Результати прогнозування для трифакторної та однофакторної економетричної моделі наведені на рис. 7.29 та 7.30.

Прогнозне значення залежної змінної (*Predicted*) для трифакторної моделі = 1,5315; і довірчі інтервали для прогнозного значення:

$$1,2740 \leq 1,5315 \leq 1,7891.$$

Variable	Predicting Values for (I _{r3}) variable: Y		
	B-Weight	Value	B-Weight * Value
X1	0,29384	3,72000	1,09309
X2	0,39614	1,20000	0,47536
X3	0,17545	0,70000	0,12281
Intercept			-0,15972
Predicted			1,53155
-95,0%CL			1,27402
+95,0%CL			1,78907

Рис. 7.29. Результати прогнозу для трифакторної моделі

Variable	Predicting Values for (I _{r3}) variable: Y		
	B-Weight	Value	B-Weight * Value
X1	0,389395	3,720000	1,448548
Intercept			0,081243
Predicted			1,529791
-95,0%CL			1,248942
+95,0%CL			1,810640

Рис. 7.30. Результати прогнозу для однофакторної моделі

Прогнозне значення залежної змінної (*Predicted*) для однофакторної моделі = 1,5298; і довірчі інтервали для прогнозного значення:

$$1,2489 \leq 1,5298 \leq 1,8106.$$

Таким чином, можна зробити висновки, що прогнози за моделями майже не відрізняються, однак довірчі інтервали зміни залежної змінної для однофакторної моделі є більш широкими.

Лабораторна робота 8

Моделювання періодичних складових часового ряду

Мета – закріплення теоретичного та практичного матеріалу, формування навичок побудови моделей часового ряду з урахуванням періодичних складових за допомогою ППП *Statistica*.

Умови завдання

В таблиці 8.1 наведені дані, що відображають динаміку обсягу продажів підприємства (y_t).

Таблиця 8.1

Вихідні дані

№ року	№ кварталу	t	Обсяг продажів (y_t), тис. шт.
1	I	1	100.80
	II	2	98.10
	III	3	99.36
	IV	4	106.74
2	I	5	102.96
	II	6	100.08
	III	7	101.16
	IV	8	108.72
3	I	9	105.12
	II	10	102.42
	III	11	102.78
	IV	12	110.70
4	I	13	106.74
	II	14	104.76
	III	15	104.94
	IV	16	112.68
5	I	17	108.91
	II	18	106.41
	III	19	107.11
	IV	20	114.74
6	I	21	110.96

Необхідно

1. Навести графік динаміки показника та проаналізувати характер зміни значень показника у часі.
2. Побудувати мультиплікативну та адитивну тренд-сезонну модель часового ряду. Оцінити точність отриманих варіантів прогнозу моделі.
3. Провести моделювання періодичної складової часового ряду за допомогою *dumty*-змінних. Побудувати лінійну багатофакторну модель і визначити всі її характеристики (параметри моделі, середні квадратичні відхилення параметрів моделі, дисперсію та середнє квадратичне відхи-

лення похибок моделі, коефіцієнти множинної кореляції та детермінації). Перевірити статистичну значимість параметрів моделі, коефіцієнта множинної кореляції. Перевірити адекватність моделі за допомогою критерію Фішера.

4. Сформувати детрендований ряд. За допомогою спектрального аналізу визначити наявність значущих періодичних складових. Провести аналіз характеристик моделі періодичної складової часового ряду. Знайти розрахункові значення рівнів ряду з урахуванням трендової та періодичної складової. Оцінити точність отриманої комбінованої прогнозованої моделі.

5. Провести порівняльний аналіз якості отриманих варіантів прогнозованої моделі, визначити найкращий з них за критерієм середньої абсолютної процентної похибки. Знайти прогноз обсягу продажів підприємства в наступному році.

Методичні рекомендації до виконання завдання

Для побудови та всебічного аналізу моделей динаміки у ППП *Statistica* передбачений модуль *Time Series/Forecasting (Часові ряди/Прогнозування)*.

Вихідні дані динаміки об'єму продажів для побудови моделей наведені на рис. 8.1.

	1 Year	2 Quarter	3 t	4 Y
1	1	I	1	100.8
2		II	2	98.1
3		III	3	99.36
4		IV	4	106.74
5	2	I	5	102.96
6		II	6	100.08
7		III	7	101.16
8		IV	8	108.72
9	3	I	9	105.12
10		II	10	102.42
11		III	11	102.78
12		IV	12	110.7
13	4	I	13	106.74
14		II	14	104.76
15		III	15	104.94
16		IV	16	112.68
17	5	I	17	108.91
18		II	18	106.41
19		III	19	107.11
20		IV	20	114.74
21	6	I	21	110.96

Рис. 8.1. Вихідні дані динаміки об'єму продажів

1. Дослідження характеру поведінки показника.

Для побудови графіка вихідних даних вибрати в пункті меню *Graphs* підпункт *Scatterplots (Точкові графіку)*. У ході задавання характеристик графіка вибрати *Graph type – Regular (Простий)*, поставити оцінку *Linear fit (Лінійний тренд)*, у якості змінних, відбиваних на графіку, вибрати по осі *X* *Період часу*, а по осі *Y* – *Об'єм продажу*. Результат наведено на рис. 8.2.

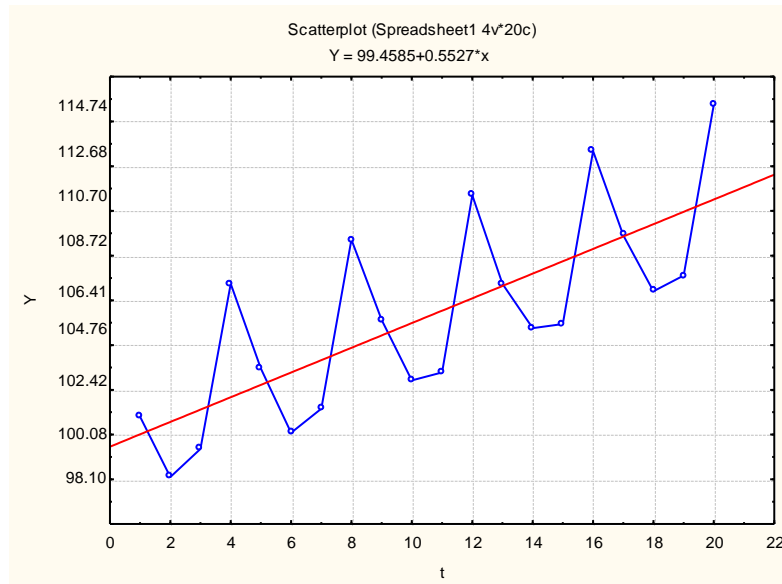


Рис. 8.2. Графік вихідних даних динаміки об'єму продажів

Графічний аналіз часового ряду (див. рис. 8.2) свідчить про наявність трендової компоненти: є стійка тенденція зростання обсягів продажів протягом останніх п'яти років. Рисунок показує, що характер тенденції близький до лінійного розвитку, також видно сезонні коливання (період коливань дорівнює одному року).

Аналіз часових рядів здійснюється в модулі *Advanced Linear/Nonlinear Models/Time Series/Forecasting (Часові ряди/Прогнозування)* (рис. 8.3). Стартова панель модуля наведена на рис. 8.4.

У даному модулі реалізовані такі методи аналізу часових рядів:

ARIMA & autocorrelation function – моделі авторегресії і проінтегрованої ковзної середньої;

Interrupted time series analysis – моделі аналізу перерваних часових рядів (моделі інтервенції);

Exponential smoothing & forecasting – моделі експоненційного згладжування та прогнозування;

Spectral (Fourier) analysis – моделі спектрального аналізу Фур'є;
Seasonal Decomposition (Census 1) – моделі сезонної декомпозиції;
X11/Y2k (Census 2) monthly/quarterly – спеціальні моделі сезонної декомпозиції;
Distributed Lags Analysis – моделі аналізу розподілених лагів.

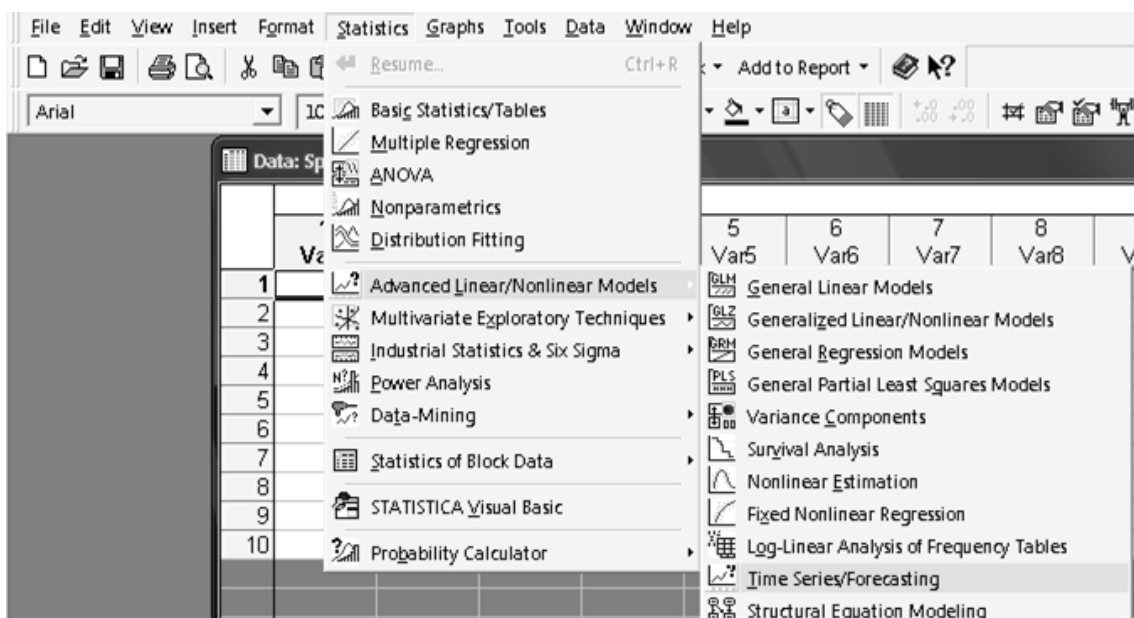


Рис. 8.3. Вибір модуля *Time Series/Forecasting*

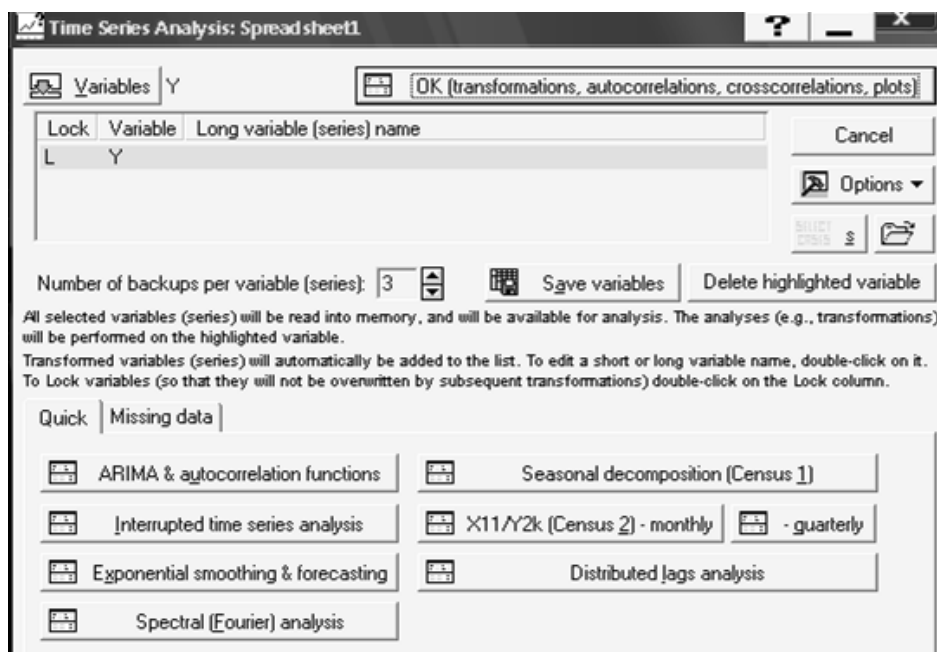


Рис. 8.4. Стартова панель модуля *Time Series/Forecasting*

2. Побудова мультиплікативної та адитивної тренд-сезонної моделі часового ряду.

Проведемо декомпозицію часового ряду на такі складові: трендову, сезонну та випадкову, використовуючи мультиплікативну модель декомпозиції часового ряду. Для цього слід вибрати опцію *Seasonal Decomposition* (Сезонна декомпозиція) у стартовій панелі модуля (див. рис. 8.4.). Ініціювавши дану опцію, буде отримано вікно для вибору типу аналізу (рис. 8.5).

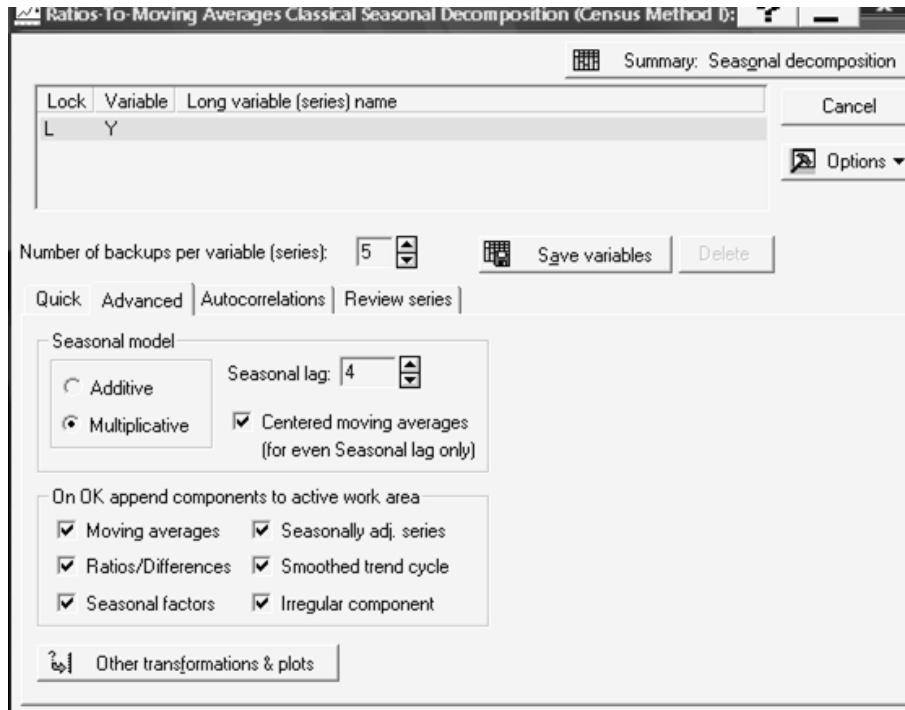


Рис. 8.5. Параметри моделі сезонної декомпозиції

У даному вікні у вкладці *Advanced* необхідно задати параметри сезонної декомпозиції:

змінна для аналізу – Y (об'єм продажів);

вид моделі (*Seasonal model*) – *Multiplicative* (Мультиплікативна);

лаг сезонного компонента (задається число сезонних індексів) (*Seasonal lag*) – 4 (квартальні дані);

центрування ковзної середньої (*Centered moving averages*).

У цьому ж вікні вибрати додаткові компоненти, які необхідно розрахувати (див. рис. 8.5), що повинні бути відбиті в робочій області:

Moving averages – ковзні середні;

Ratios/Differences – відношення/різниці (якщо модель мультиплікативна, береться відношення, адитивна – різниця вихідного та згладжуваного ряду);

Seasonal factors – сезонні індекси;

Seasonally adj. series – ряд, скорегований на сезонність (без сезонної компоненти);

Smoothed trend cycle – згладжена трендово-циклічна складова;

Irregular component – нерегулярна (випадкова) складова.

На рис. 8.6 наведені результати моделі сезонної декомпозиції об'єму продажів, які можна отримати, ініціювавши клавішу *Summary: Seasonal Decomposition* у верхній частині вікна на рис. 8.5.

Seasonal Decomposition: Multipl. season (4) (Spreadsheet1)							
Y							
Case	Y	Moving Averages	Ratios	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irreg. Compon.
1	100.8000			100.3765	100.4220	100.5936	0.998293
2	98.1000			97.5371	100.5771	100.9061	0.996740
3	99.3600	101.2500	98.1333	97.6807	101.7192	101.5310	1.001854
4	106.7400	101.7900	104.8630	104.4058	102.2357	102.0531	1.001790
5	102.9600	102.2850	100.6599	100.3765	102.5739	102.5209	1.000516
6	100.0800	102.7350	97.4157	97.5371	102.6071	102.9401	0.996765
7	101.1600	103.2300	97.9948	97.6807	103.5619	103.4960	1.000637
8	108.7200	103.7700	104.7702	104.4058	104.1322	104.0650	1.000645
9	105.1200	104.3550	100.7331	100.3765	104.7258	104.5818	1.001376
10	102.4200	104.7600	97.7663	97.5371	105.0062	105.0080	0.999983
11	102.7800	105.2550	97.6486	97.6807	105.2204	105.4218	0.998090
12	110.7000	105.6600	104.7700	104.4058	106.0286	105.9575	1.000671
13	106.7400	106.2450	100.4659	100.3765	106.3397	106.5043	0.998454
14	104.7600	106.7850	98.1037	97.5371	107.4053	107.0791	1.003046
15	104.9400	107.2800	97.8188	97.6807	107.4317	107.5330	0.999058
16	112.6800	107.8225	104.5051	104.4058	107.9251	108.0160	0.999158
17	108.9100	108.2350	100.6236	100.3765	108.5015	108.5148	0.999877
18	106.4100	108.7775	97.8235	97.5371	109.0969	109.0471	1.000457
19	107.1100	109.2925	98.0031	97.6807	109.6532	109.5550	1.000896
20	114.7400	109.8050	104.4943	104.4058	109.8981	110.0317	0.998786
21	110.9600			100.3765	110.5439	110.2701	1.002483

Рис. 8.6. Результат мультиплікативної сезонної декомпозиції

Таким чином, можна отримати такі компоненти:

Seasonal factors – сезонні індекси;

Seasonally adj. series – ряд, скорегований на сезонність (без сезонної компоненти);

Irregular component – випадкова складова.

Отримані результати необхідно скопіювати в таблицю з вихідними даними, користуючись контекстним меню (*Copy with Headers/Paste*) для подальшого їх аналізу та побудови графіків. Для побудови графіка ряду, скорегованого на сезонність, необхідно перебуваючи в таблиці вихідних даних, вибирати в пункті меню *Graphs* підпункт *Scatterplots* (*Точкові графіки*). У процесі задавання характеристик графіка слід вибрати *Graph type* – *Regular* (*Простий*), поставити оцінку *Linear fit* (*Лінійний тренд*), у якості змінних, відбиваних на графіку, вибрати по осі *X* *Період часу (t)*, а по осі *Y* – *Seasonally adj. series*. Результат наведено на рис. 8.7.

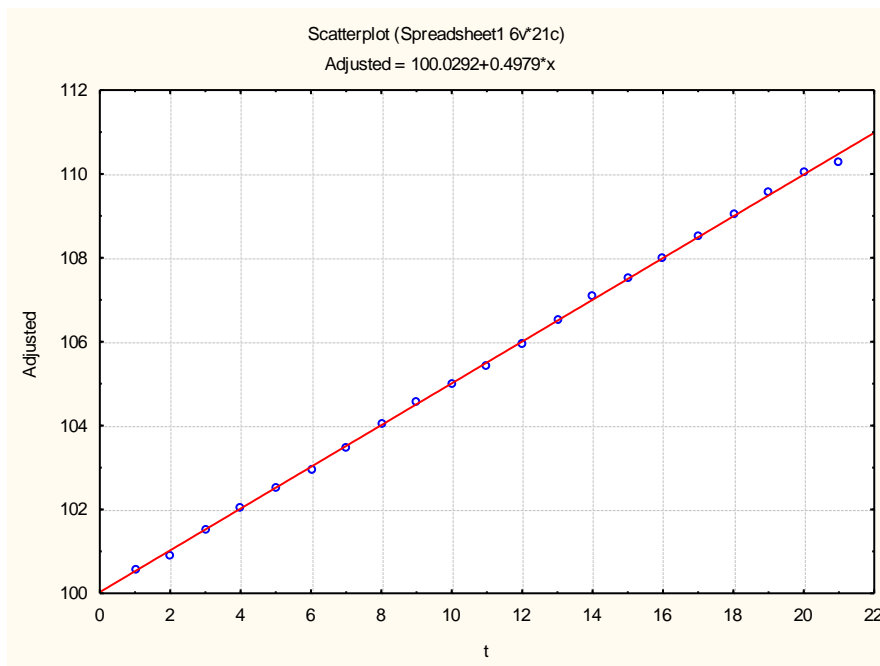


Рис. 8.7. Графік ряду, скорегованого на сезонність

Із аналізу графіка можна зробити висновок про існування лінійного тренда. Оцінювання параметрів тренда за допомогою МНК здійснюється у модулі *Multiple Regression* (*Множинна регресія*). У якості залежної змінної необхідно задати *Seasonally adj. series*, а незалежної – *Період часу (t)*. Результат побудови моделі тренда наведено на рис. 8.8.

Regression Summary for Dependent Variable: Adjusted (Spreadsh						
R= .99971965 R ² = .99943937 Adjusted R ² = .99940986						
F(1, 19)=33872. p<0.0000 Std.Error of estimate: .07507						
N=21	Beta	Std. Err. of Beta	B	Std. Err. of B	t(19)	p-level
Intercept			100.0292	0.033969	2944.714	0.000000
t	0.999720	0.005432	0.4979	0.002705	184.042	0.000000

Рис. 8.8. Оцінка параметрів лінійного тренда

Таким чином, модель тренда буде мати такий вигляд:

$$T = 100,0292 + 0,4979t .$$

Значення трендової складової (T) знаходяться як теоретичні значення за побудованою моделлю тренда (ініціювавши клавішу аналізу помилок *Summary: Residuals & Predicted*, теоретичні значення залежної змінної (*Predicted value*)).

Графік сезонної складової (*Seasonal factors*) наведений на рис. 8.9.

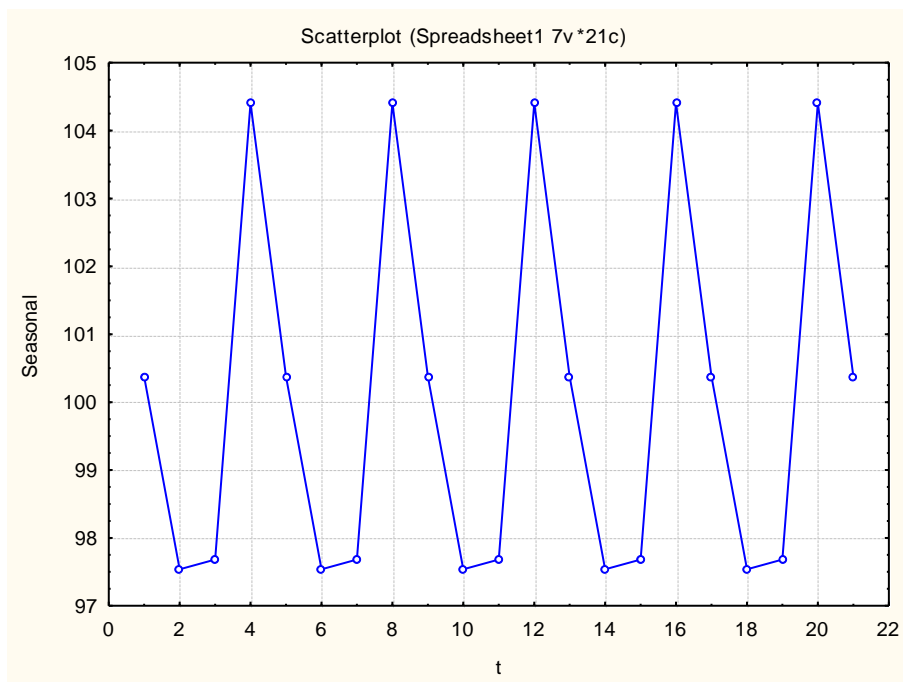


Рис. 8.9. Графік сезонної складової

Спостерігається стійке збільшення обсягів продажів, що повторюється в 4-му і в 1-му кварталах порівняно з 3-м і 2-м кварталами, причому найістотніші "сплески" в динаміці показника є чітко видимими у 4-му кварталі.

Розрахунок прогнозних значень часового ряду можливий шляхом задавання формули розрахунку $Prognoz = (100.0292 + 0.4979 * t) * Seasonal\ factors / 100\ %$ у області специфікації змінної *Long name* (рис. 8.10). Графік розрахункових значень обсягів продажу наведений на рис. 8.11.

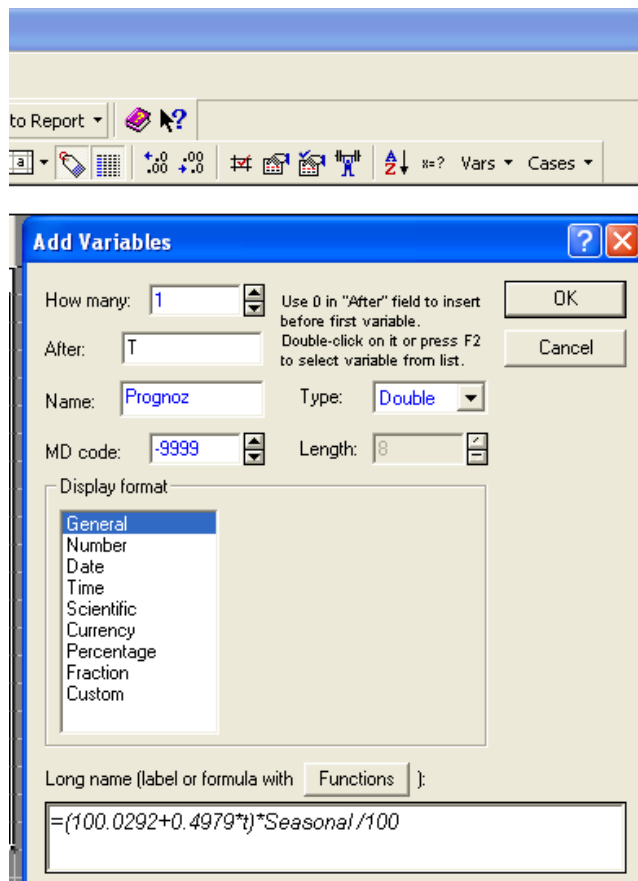


Рис. 8.10. Розрахунок прогнозних значень обсягів продажу

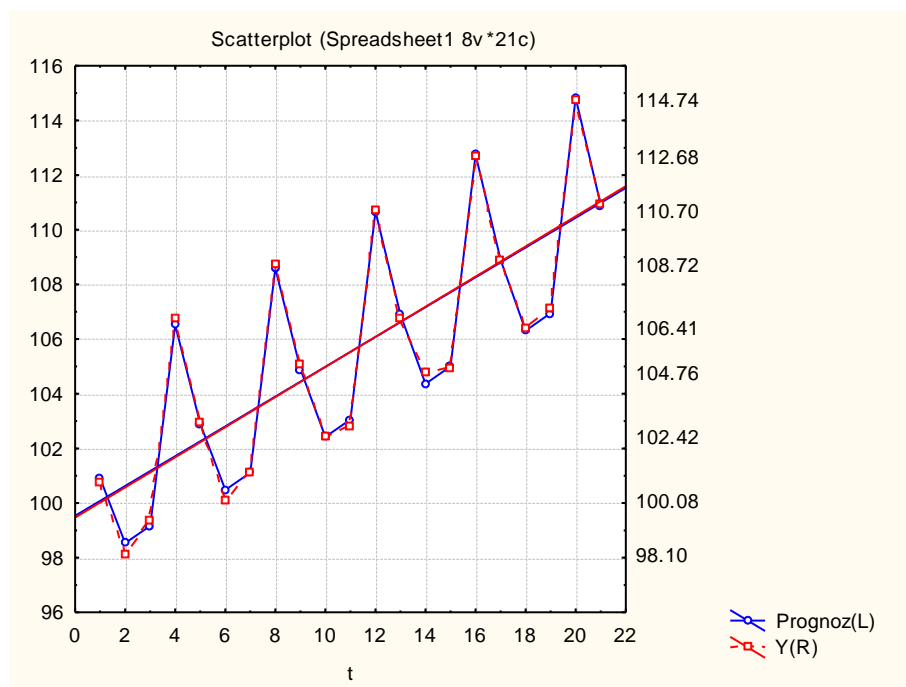


Рис. 8.11. Графік розрахункових та фактичних значень обсягів продажу

На рис. 8.12 наведені всі компоненти мультиплікативної моделі декомпозиції, теоретичні значення об'ємів продажів та похибки за даною моделлю.

	1 Year	2 Quater	3 t	4 Prognoz	5 R	6 Y	7 Seasonal	8 Adjusted	9 T
1	1	I	1	100.9055	-0.106	100.8	100.3765	100.5936	100.5271
2		II	2	98.53686	-0.437	98.1	97.5371	100.9061	101.0250
3		III	3	99.16825	0.192	99.36	97.6807	101.5310	101.5229
4		IV	4	106.5156	0.224	106.74	104.4058	102.0531	102.0208
5	2	I	5	102.9046	0.055	102.96	100.3765	102.5209	102.5187
6		II	6	100.4794	-0.399	100.08	97.5371	102.9401	103.0166
7		III	7	101.1137	0.046	101.16	97.6807	103.4960	103.5145
8		IV	8	108.595	0.125	108.72	104.4058	104.0650	104.0123
9	3	I	9	104.9037	0.216	105.12	100.3765	104.5818	104.5102
10		II	10	102.422	-0.002	102.42	97.5371	105.0080	105.0081
11		III	11	103.0591	-0.279	102.78	97.6807	105.4218	105.5060
12		IV	12	110.6743	0.026	110.7	104.4058	105.9575	106.0039
13	4	I	13	106.9028	-0.163	106.74	100.3765	106.5043	106.5018
14		II	14	104.3645	0.395	104.76	97.5371	107.0791	106.9997
15		III	15	105.0045	-0.064	104.94	97.6807	107.5330	107.4976
16		IV	16	112.7536	-0.074	112.68	104.4058	108.0160	107.9954
17	5	I	17	108.9019	0.008	108.91	100.3765	108.5148	108.4933
18		II	18	106.3071	0.103	106.41	97.5371	109.0471	108.9912
19		III	19	106.9499	0.160	107.11	97.6807	109.5550	109.4891
20		IV	20	114.833	-0.093	114.74	104.4058	110.0317	109.9870
21	6	I	21	110.901	0.059	110.96	100.3765	110.2701	110.4849

Рис. 8.12. Значення виділених компонент ряду динаміки

Для перевірки випадкової компоненти на нормальний закон розподілу необхідно виділити стовпець *Irregular component (R)*, натискаючи праву кнопку, викликати контекстне меню, вибрати *Graphs of Input Data/Histogram (R)/Normal Fit ma Graphs of Input Data/Probability Plot (R)/Normal Probability*, як це показано на рис. 8.13. Гістограма випадкової компоненти та графік випадкової компоненти на нормальному ймовірнісному папері подано на рис. 8.14 і 8.15.

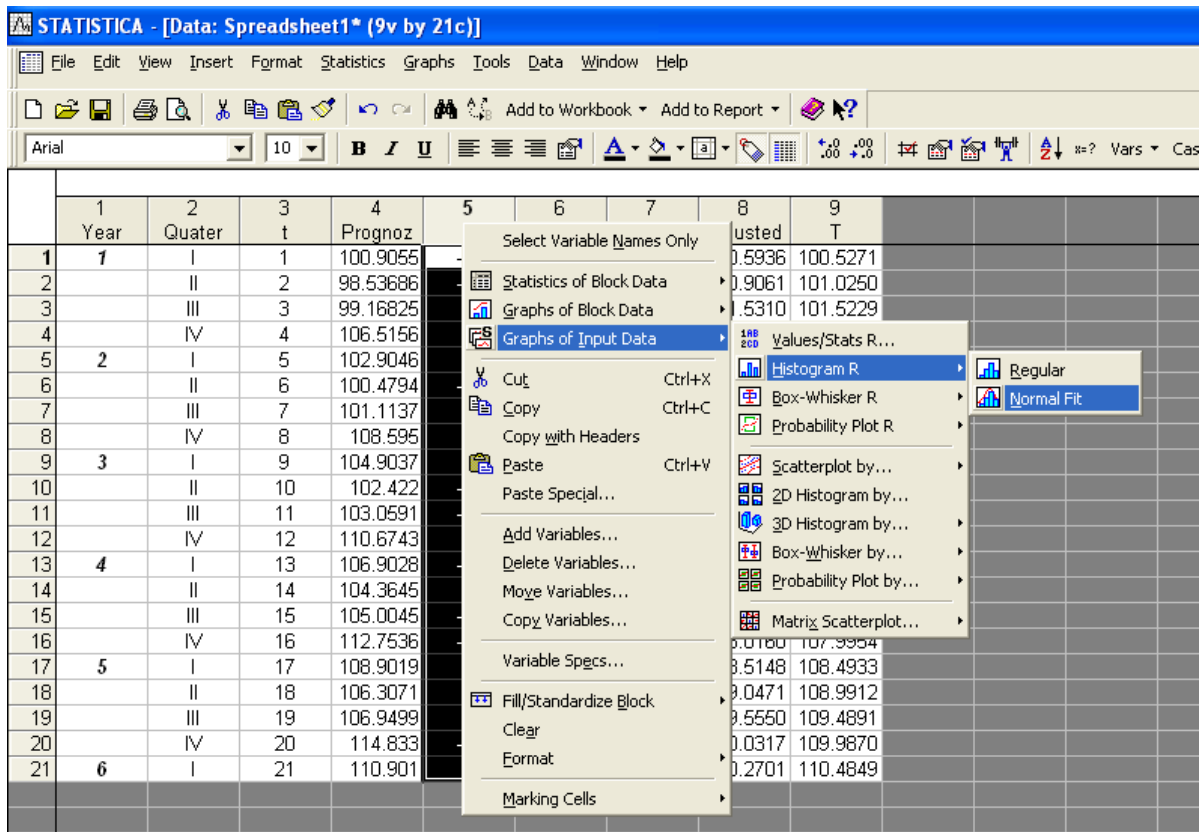


Рис. 8.13. Побудова графіків аналізу випадкової компоненти

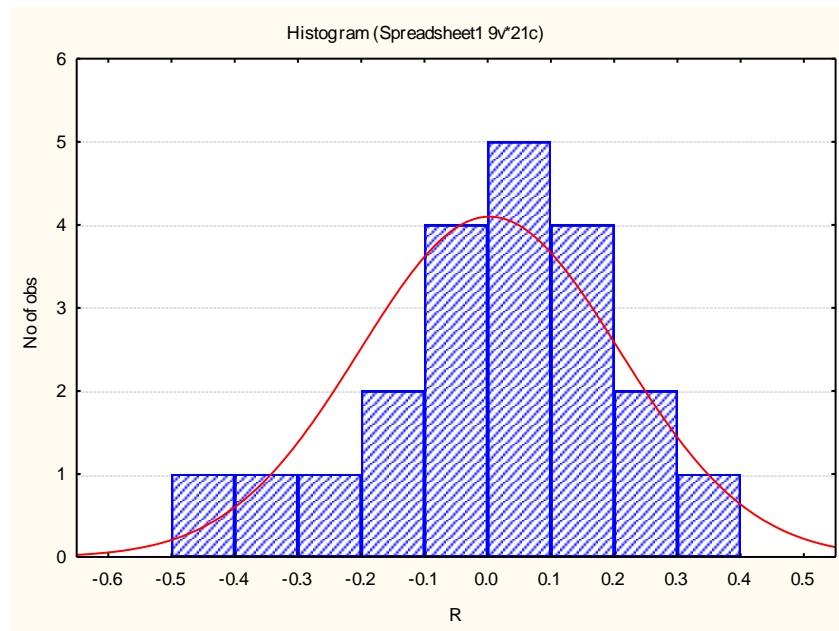


Рис. 8.14. Гістограма випадкової компоненти

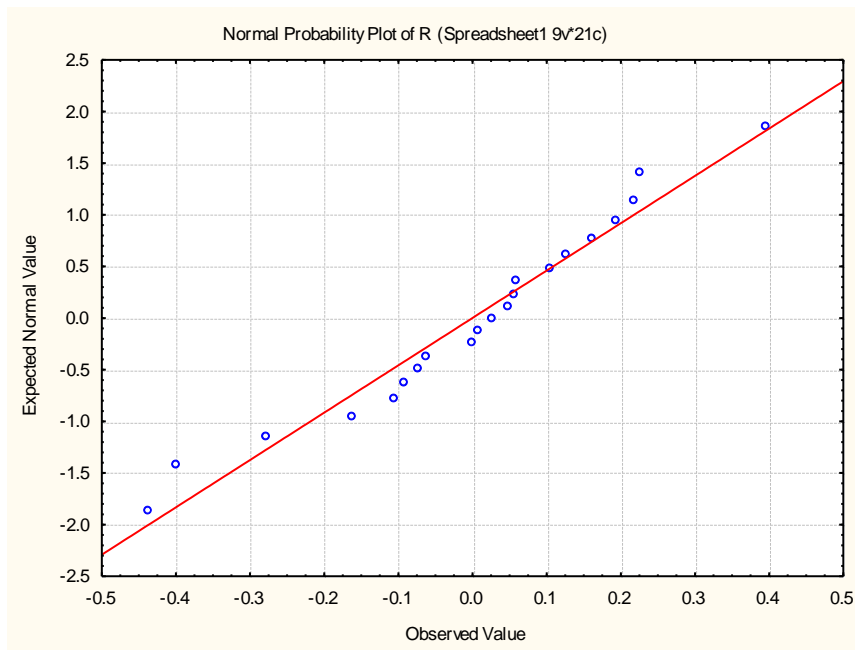


Рис. 8.15. Графік випадкової компоненти на нормальному ймовірнісному папері

Слід провести декомпозицію часового ряду на такі складові: трендову, сезонну й випадкову, використовуючи адитивну модель часового ряду. Ініціювавши дану опцію, буде отримано вікно для вибору типу аналізу (рис. 8.16).

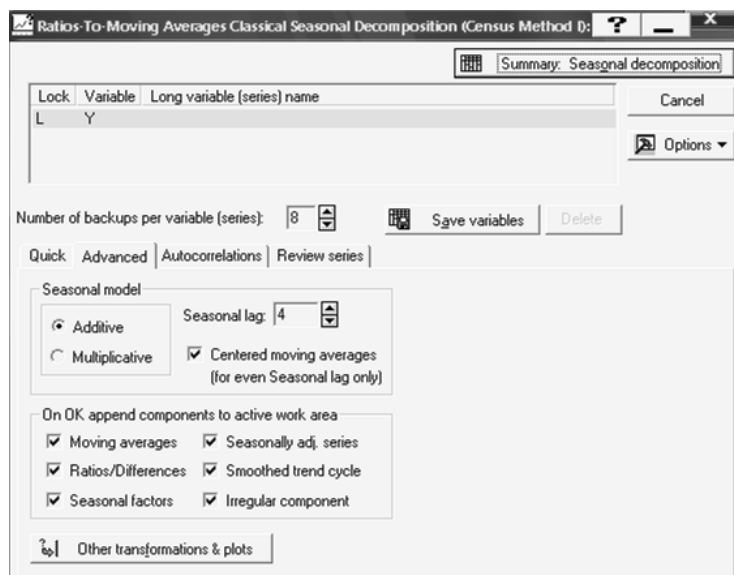


Рис. 8.16. Параметри моделі адитивної сезонної декомпозиції

На рис. 8.17 наведені результати моделі адитивної сезонної декомпозиції.

Seasonal Decomposition: Additive season (4) (Spreadsheet1)							
Y							
Case	Y	Moving Averages	Diffnrcs	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irreg. Compon.
1	100.8000			0.38734	100.4127	100.6942	-0.281535
2	98.1000			-2.61203	100.7120	100.9804	-0.268417
3	99.3600	101.2500	-1.89000	-2.45666	101.8167	101.5530	0.263694
4	106.7400	101.7900	4.95000	4.68134	102.0587	102.0395	0.018806
5	102.9600	102.2850	0.67500	0.38734	102.5727	102.5170	0.055694
6	100.0800	102.7350	-2.65500	-2.61203	102.6920	102.9502	-0.258194
7	101.1600	103.2300	-2.07000	-2.45666	103.6167	103.5130	0.103694
8	108.7200	103.7700	4.95000	4.68134	104.0387	104.0595	-0.021194
9	105.1200	104.3550	0.76500	0.38734	104.7327	104.5770	0.155694
10	102.4200	104.7600	-2.34000	-2.61203	105.0320	105.0102	0.021806
11	102.7800	105.2550	-2.47500	-2.45666	105.2367	105.4330	-0.196306
12	110.7000	105.6600	5.04000	4.68134	106.0187	105.9595	0.058806
13	106.7400	106.2450	0.49500	0.38734	106.3527	106.4970	-0.144306
14	104.7600	106.7850	-2.02500	-2.61203	107.3720	107.0702	0.301806
15	104.9400	107.2800	-2.34000	-2.45666	107.3967	107.5341	-0.137417
16	112.6800	107.8225	4.85750	4.68134	107.9987	108.0254	-0.026750
17	108.9100	108.2350	0.67500	0.38734	108.5227	108.5081	0.014583
18	106.4100	108.7775	-2.36750	-2.61203	109.0220	109.0336	-0.011528
19	107.1100	109.2925	-2.18250	-2.45666	109.5667	109.5507	0.015917
20	114.7400	109.8050	4.93500	4.68134	110.0587	110.0660	-0.007333
21	110.9600			0.38734	110.5727	110.3236	0.249042

Рис. 8.17. Результат адитивної сезонної декомпозиції

Аналіз побудованої моделі, виділення трендової складової, побудову графіків основних компонент слід проводити аналогічно розглянутій мультиплікативній моделі. Параметри моделі тренда наведені на рис. 8.18.

Regression Summary for Dependent Variable Adjusted (add)						
R= .99846200 R ² = .99692637 Adjusted R ² = .99676460						
F(1,19)=6162.6 p<0.0000 Std.Error of estimate: .17814						
N=21	Beta	Std.Err. of Beta	B	Std.Err. of B	t(19)	p-level
Intercept			99.9709	0.080608	1240.21	0.000000
t	0.998462	0.012719	0.50395	0.006420	78.502	0.000000

Рис. 8.18. Оцінки параметрів лінійного тренда

Розрахунок прогнозних значень часового ряду можливий шляхом задавання формули розрахунку $Prognoz = 99,97 + 0,50395 * t + Seasonal\ factors$ в області специфікації змінної *Long name*. Графік розрахункових значень обсягів продажу наведений на рис. 8.19.

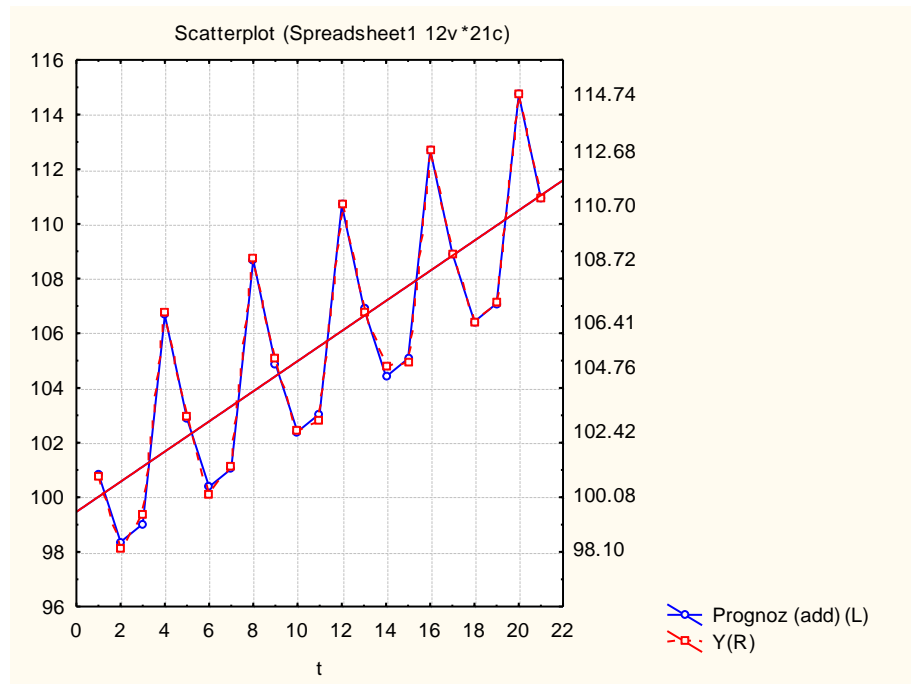


Рис. 8.19. Графік розрахункових та фактичних значень обсягів продажу

3. Моделювання періодичної складової часового ряду за допомогою *dummy*-змінних. Фіктивні або бінарні змінні (*dummy variables*) приймають тільки значення 0 або 1. Оскільки часовий ряд має квартальну структуру даних та містить трендову складову, яка може бути подана лінійною функцією, то доцільним є розгляд такої моделі:

$$y_j = a + \beta_1 d_{j1} + \beta_2 d_{j2} + \beta_3 d_{j3} + \varepsilon_j,$$

$$d_{j1} = \begin{cases} 1, \text{ якщо } j \text{ відповідає першому кварталу,} \\ 0, \text{ в інших випадках} \end{cases}$$

$$d_{j2} = \begin{cases} 1, \text{ якщо } j \text{ відповідає другому кварталу,} \\ 0, \text{ в інших випадках} \end{cases}$$

$$d_{j3} = \begin{cases} 1, \text{ якщо } j \text{ відповідає третьому кварталу,} \\ 0, \text{ в інших випадках.} \end{cases}$$

Вихідні дані для побудови регресійної моделі з *dummy*-змінними наведені на рис. 8.20.

	1	2	3	4	5	6	7
	Year	Quater	t	Y	d1	d2	d3
1	1	I	1	100.8	1	0	0
2		II	2	98.1	0	1	0
3		III	3	99.36	0	0	1
4		IV	4	106.74	0	0	0
5	2	I	5	102.96	1	0	0
6		II	6	100.08	0	1	0
7		III	7	101.16	0	0	1
8		IV	8	108.72	0	0	0
9	3	I	9	105.12	1	0	0
10		II	10	102.42	0	1	0
11		III	11	102.78	0	0	1
12		IV	12	110.7	0	0	0
13	4	I	13	106.74	1	0	0
14		II	14	104.76	0	1	0
15		III	15	104.94	0	0	1
16		IV	16	112.68	0	0	0
17	5	I	17	108.91	1	0	0
18		II	18	106.41	0	1	0
19		III	19	107.11	0	0	1
20		IV	20	114.74	0	0	0
21	6	I	21	110.96	1	0	0

Рис. 8.20. Вихідні дані

Щоб почати обчислювальні процедури, необхідно увійти в позицію меню *Statistics/Multiple Regression*. Після підтвердження вибору модуля з'явиться стартова панель даного модуля, де необхідно задати змінні для аналізу.

Ініціювати кнопку *Variables (Змінні)* і у вікні, що з'явилося, указати *Dependent* (залежну) і *Independent* (незалежну) змінні для побудови багатofакторної регресійної моделі. Вибір змінних поданий на рис. 8.21. Після вказівки змінних слід підтвердити свій вибір натисканням кнопки *OK*.

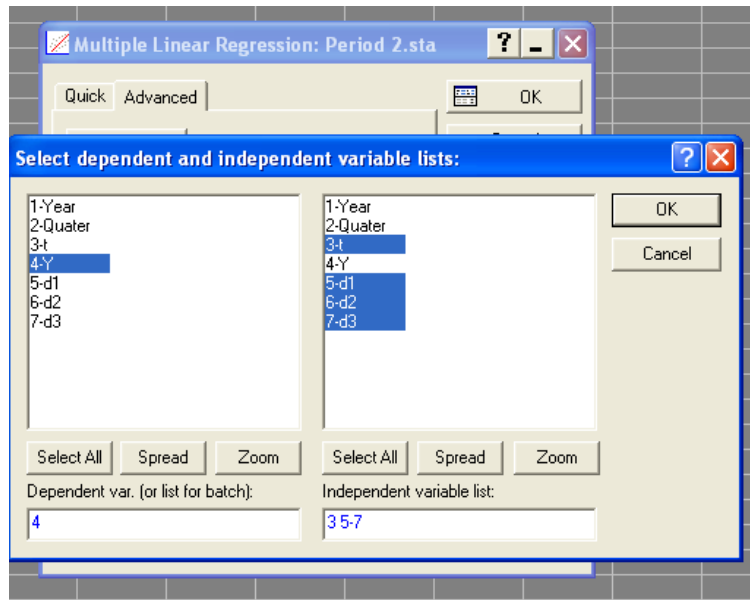


Рис. 8.21. Вибір змінних для аналізу

Слід побудувати лінійну багатофакторну модель і визначити всі її характеристики. Перевірити статистичну значимість параметрів моделі та адекватність моделі за критерієм Фішера. Результати побудови моделі наведені на рис. 8.22.

Regression Summary for Dependent Variable (Period 2.sta)						
R= .99927934 R ² = .99855920 Adjusted R ² = .99819900						
F(4, 16)=2772.2 p<0.0000 Std.Error of estimate: .19200						
N=21	Beta	Std.Err. of Beta	B	Std.Err. of B	t(16)	p-level
Intercept			104.673	0.11982	873.585	0.00000
t	0.69062	0.00955	0.5036	0.00696	72.303	0.00000
d1	-0.43971	0.01191	-4.297	0.11646	-36.898	0.00000
d2	-0.70951	0.01179	-7.354	0.12222	-60.175	0.00000
d3	-0.68902	0.01173	-7.142	0.12162	-58.723	0.00000

Рис. 8.22. Результати багатофакторного регресійного аналізу

Для перевірки гіпотези щодо значимості регресійної моделі використовується дисперсійний аналіз. Для цього необхідно ініціювати кнопку *Advanced/ANOVA*. Результати дисперсійного аналізу для досліджуваної моделі наведені на рис. 8.23. У даній таблиці наведено суму квадратів відхилень за регресією (*Sums of Squares Regress*), суму квадратів відхилень похибок (*Sums of Squares Residual*), дисперсію похибок (*Mean Squares Residual*) та критерій Фішера (*F*).

Effect	Analysis of Variance; DV: Y (Period 2.sta)				
	Sums of Squares	df	Mean Squares	F	p-level
Regress.	408.7646	4	102.1912	2772.235	0.000000
Residual	0.5898	16	0.0369		
Total	409.3544				

Рис. 8.23. Таблиця дисперсійного аналізу

Теоретичні значення обсягу продажів за побудованою моделлю можна визначити ініціювавши клавішу аналізу помилок *Summary: Residuals & Predicted*, теоретичні значення залежної змінної (*Predicted value*) (рис. 8.24).

Case No.	Predicted & Residual Values (Period 2.sta)							
	Observed Value	Predicted Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalanobis Distance	Deleted Residual
1	100.8000	100.8790	-0.079460	-1.02930	-0.41390	0.104850	5.012530	-0.113240
2	98.1000	98.3250	-0.225570	-1.59420	-1.17490	0.102350	4.731830	-0.315140
3	99.3600	99.0410	0.318420	-1.43580	1.65840	0.102350	4.731830	0.444850
4	106.7400	106.6870	0.052420	0.25540	0.27300	0.102350	4.731830	0.073230
5	102.9600	102.8930	0.066310	-0.58370	0.34540	0.088820	3.328320	0.084370
6	100.0800	100.3390	-0.259780	-1.14860	-1.35300	0.090260	3.468670	-0.333510
7	101.1600	101.0550	0.104210	-0.99030	0.54280	0.090260	3.468670	0.133790
8	108.7200	108.7010	0.018210	0.70090	0.09480	0.090260	3.468670	0.023370
9	105.1200	104.9070	0.212100	-0.13820	1.10470	0.079610	2.486210	0.256140
10	102.4200	102.3540	0.066000	-0.70310	0.34370	0.085860	3.047610	0.082500
11	102.7800	103.0700	-0.290000	-0.54470	-1.51040	0.085860	3.047610	-0.362500
12	110.7000	110.7160	-0.016000	1.14640	-0.08330	0.085860	3.047610	-0.020000
13	106.7400	106.9220	-0.182100	0.30730	-0.94840	0.079610	2.486210	-0.219910
14	104.7600	104.3680	0.391790	-0.25760	2.04060	0.090260	3.468670	0.502970
15	104.9400	105.0840	-0.144210	-0.09920	-0.75110	0.090260	3.468670	-0.185130
16	112.6800	112.7300	-0.050200	1.59200	-0.26150	0.090260	3.468670	-0.064450
17	108.9100	108.9360	-0.026310	0.75280	-0.13700	0.088820	3.328320	-0.033480
18	106.4100	106.3820	0.027580	0.18790	0.14360	0.102350	4.731830	0.038530
19	107.1100	107.0980	0.011580	0.34630	0.06030	0.102350	4.731830	0.016180
20	114.7400	114.7440	-0.004420	2.03750	-0.02300	0.102350	4.731830	-0.006180
21	110.9600	110.9500	0.009470	1.19830	0.04930	0.104850	5.012530	0.013500
Minimum	98.1000	98.3250	-0.290000	-1.59420	-1.51040	0.079610	2.486210	-0.362500
Maximum	114.7400	114.7440	0.391790	2.03750	2.04060	0.104850	5.012530	0.502970
Mean	105.5320	105.5320	0.000000	0.00000	0.00000	0.093320	3.809520	0.000750
Median	105.1200	105.0840	0.009470	-0.09920	0.04930	0.090260	3.468670	0.013500

Рис. 8.24. Розрахункові та фактичні значення обсягу продажів

Результати порівняльного аналізу розрахункових та фактичних значень обсягів продажу підприємства (рис. 8.25) дозволяють зробити висновок про добру якість апроксимації.

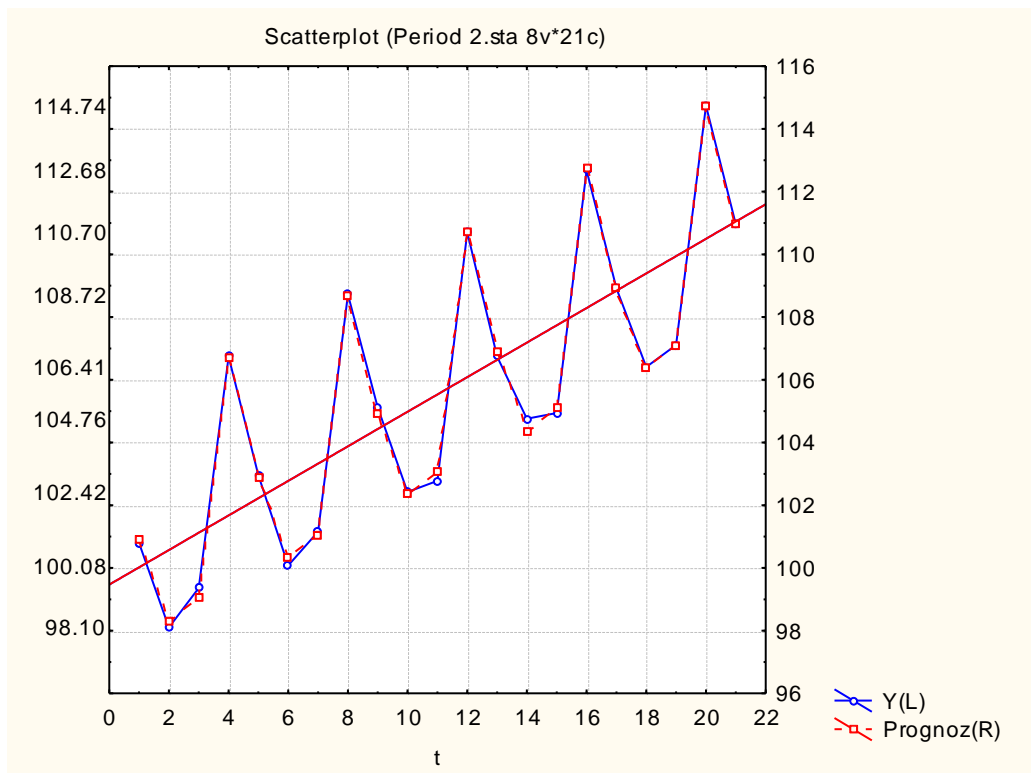


Рис. 8.25. Графік розрахункових та фактичних значень обсягів продажу

4. Дослідження періодичних складових на підставі спектрального аналізу. Визначити рівні детрендованого ряду. Знайти параметри лінійного тренду. Оцінювання параметрів тренда за допомогою МНК здійснюється у модулі *Multiple Regression (Множинна регресія)*. У якості залежної змінної необхідно задати Обсяг продажів (Y), а незалежної – *Період часу* (t). Результат побудови моделі тренда наведено на рис. 8.26.

Regression Summary for Dependent Variable: Y (Period 3.sta)						
R= .75612364 R ² = .57172296 Adjusted R ² = .54918206						
F(1,19)=25.364 p<.00007 Std.Error of estimate: 3.0376						
N=21	Beta	Std.Err. of Beta	B	Std.Err. of B	t(19)	p-level
Intercept			99.46843	1.374548	72.36446	0.000000
t	0.756124	0.150136	0.55131	0.109469	5.03625	0.000073

Рис. 8.26. Оцінювання параметрів лінійного тренда

Таким чином, модель тренда буде мати такий вигляд:

$$y_t = 99,46843 + 0,55131t.$$

Розрахунок значень детрендованого часового ряду можливий шляхом задавання формули розрахунку $P_R = Y - 99,46843 - 0,55131 * t$ в області специфікації змінної *Long name*. Графік значень першого залишкового ряду наведений на рис. 8.27.

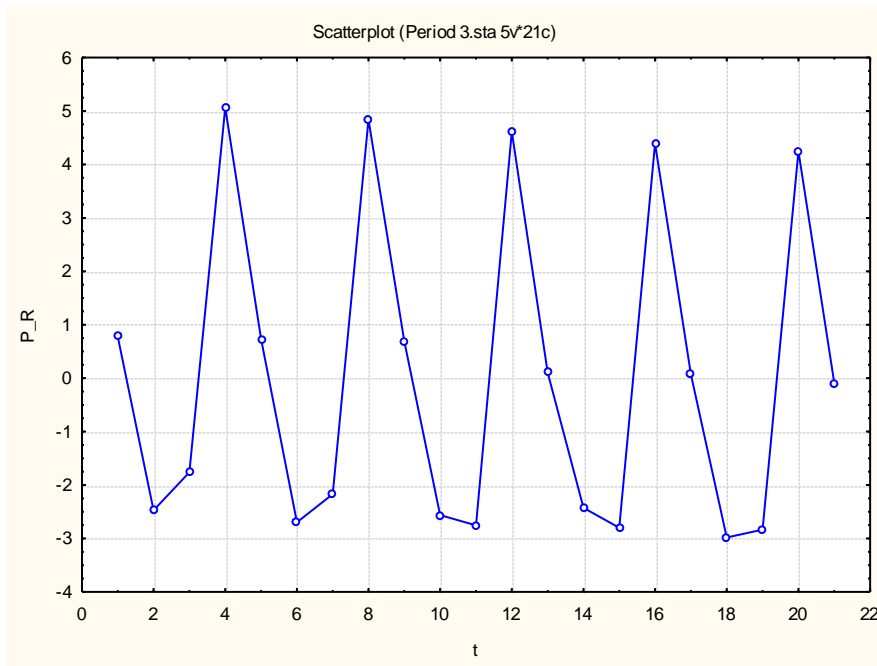


Рис. 8.27. Графік детрендованого ряду

Слід визначити найбільш значущі періодичні складові за допомогою спектрального аналізу. Для цього вибрати опцію *Spectral (Fourier) analysis* – моделі спектрального аналізу Фур'є на стартовій панелі модуля (див. рис. 8.4). Ініціювавши дану опцію, з'явиться вікно для вибору типу аналізу (рис. 8.28).

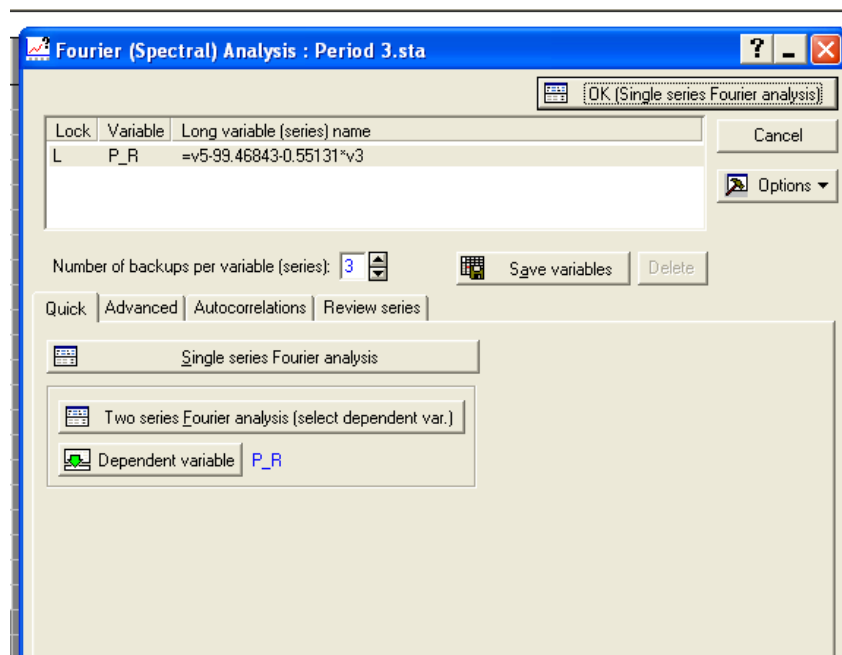


Рис. 8.28. Вибір типу аналізу

У даному модулі реалізовані такі методи:

однофакторний Фур'є-аналіз – *Single series Fourier analysis*;

крос-спектральний аналіз – *Two series Fourier analysis*.

Після ініціації опції *Single series Fourier analysis* (однофакторний Фур'є-аналіз), визначення залежної змінної *Dependent variable* – *P_R* (*перший залишковий ряд*), буде отримане наступне вікно результатів аналізу (рис. 8.29).

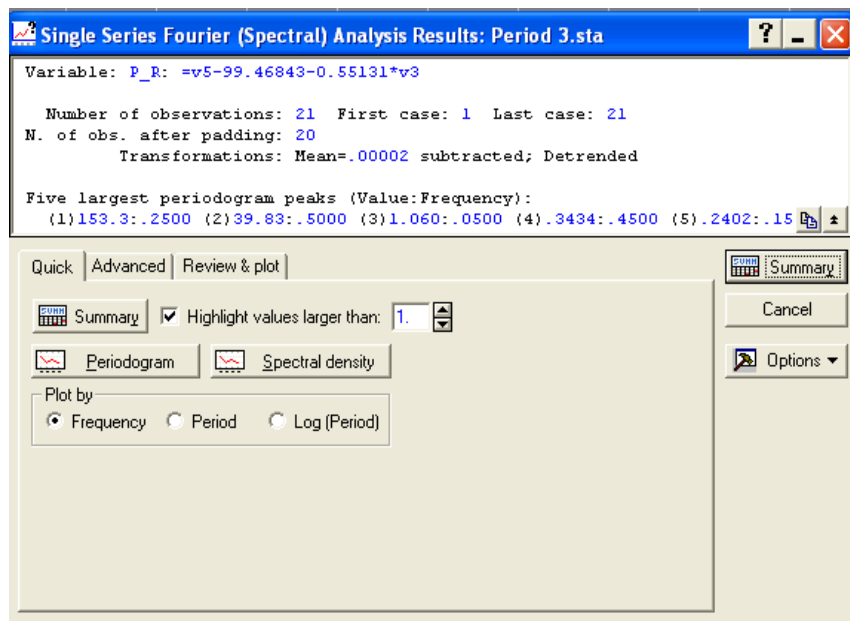


Рис. 8.29. Вікно результатів аналізу

Ініціювавши опцію *Periodogram* (*періодограма*) буде отримано графік для часового ряду, що розглядається (рис. 8.30). На рисунку розглянуті два варіанти побудови періодограми. На першому з двох наведених графіків побудована залежність інтенсивності від частоти (*frequency*). На другому – інтенсивності від часу (*period*).

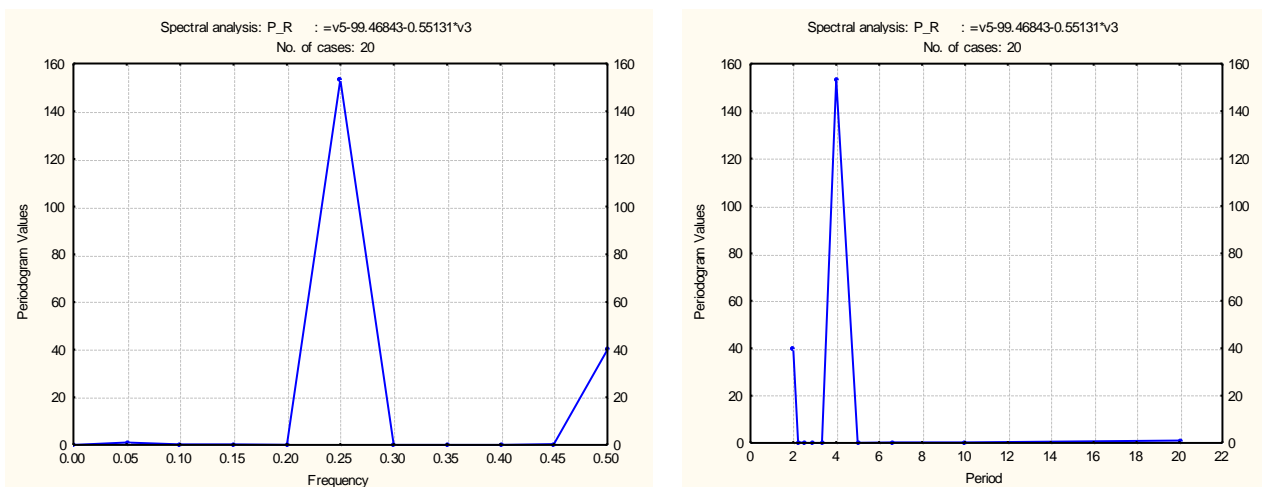


Рис. 8.30. Періодограма часового ряду

Як видно з рисунка, часовий ряд має періодичну складову з періодом коливань, що дорівнює 4 кварталам. Періодограма для цього значення має гострий пік. Цьому періоду відповідає частота, яка дорівнює 0.25.

Як періодограма, так і спектрограма розглядають розподіл гармонік у часовому ряді та використовуються для визначення та оцінювання періодичних компонент невідомої частоти та визначення даної частоти. Ініціювавши опцію *Spectral density* буде отримано графік для часового ряду (рис. 8.31).

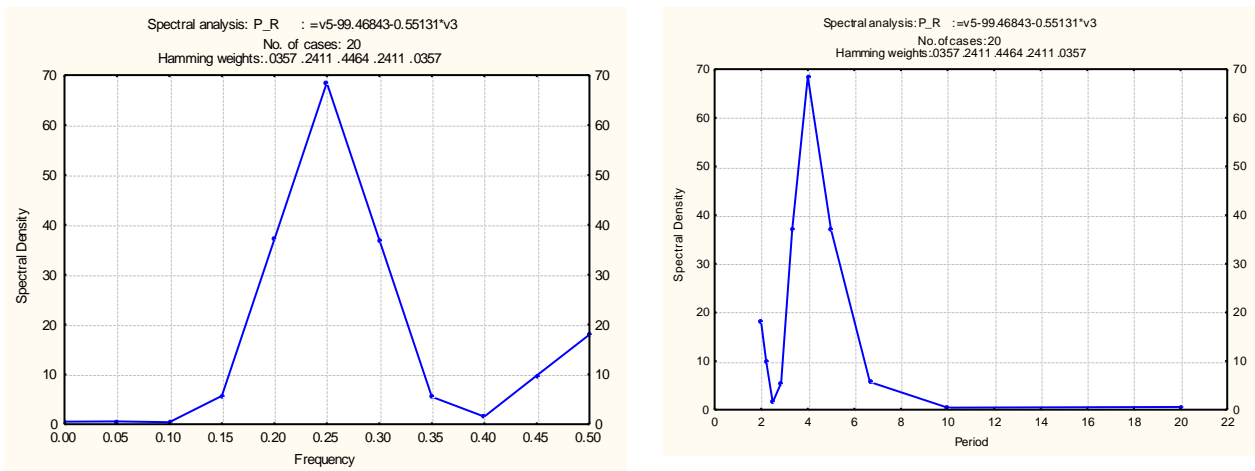


Рис. 8.31. Спектрограма часового ряду

Після ініціювання опції *Summary* будуть отримані коефіцієнти складових ряду Фур'є, які наведені на рис. 8.32.

Spectral analysis: P_R :=v5-99.46843-0.55131*v3 (Period 3.sta)							
No. of cases: 20							
	Frequency	Period	Cosine Coeffs	Sine Coeffs	Periodog	Density	Hamming Weights
0	0.000000		0.00860	0.00000	0.0007	0.52753	0.035714
1	0.050000	20.00000	0.04953	0.32171	1.0595	0.57484	0.241071
2	0.100000	10.00000	0.02323	0.14959	0.2292	0.41905	0.446429
3	0.150000	6.66667	0.06800	0.13928	0.2402	5.69942	0.241071
4	0.200000	5.00000	-0.09452	0.02317	0.0947	37.07456	0.035714
5	0.250000	4.00000	1.46931	-3.62969	153.3351	68.49518	
6	0.300000	3.33333	0.05789	-0.02723	0.0409	36.99236	
7	0.350000	2.85714	0.03814	-0.02337	0.0200	5.51513	
8	0.400000	2.50000	0.02064	0.05306	0.0324	1.52595	
9	0.450000	2.22222	0.14158	-0.11955	0.3434	9.77541	
10	0.500000	2.00000	-1.99569	0.00000	39.8277	17.94810	

Рис. 8.32. Результати спектрального аналізу

У таблиці, наведеній на рисунку, розраховані всі коефіцієнти Фур'є. Також у таблиці зазначені значення періоду та частоти для кожної гармоніки. З даної таблиці видно, що сама повільна хвиля має період, який дорівнює довжині ряду (20 кварталів), а найшвидший період рівний 2 кварталам. Крім того, у таблиці, наведеній на рисунку, вказуються ваги так званого *спектрального вікна Хеммінга*. Наявність даних ваг припускає, що здійснюється згладжування вибіркового спектра у спектральному вікні. Розмір даного вікна рівний 5. Згладжування використовується внаслідок того, що існує зсув оцінки спектра. Для усунення зсуву оцінки можна використовувати фільтри згладжування. Крім ваг вікна *Хеммінга* використовуються також вікна *Даніеля*, *Тьюки*, *Парзена*, *Барлетта*. Визначити вікна згладжування можна за допомогою відповідної опції у вкладці *Advanced* (розширений аналіз) (рис. 8.33).

За винятком вікна Даніеля, усі функції приписують більшу вагу спостереженню, що згладжується та перебуває в центрі вікна й менші ваги значенням у міру видалення від центру. У багатьох випадках усі ці вікна даних дають дуже схожі результати (рис. 8.34).

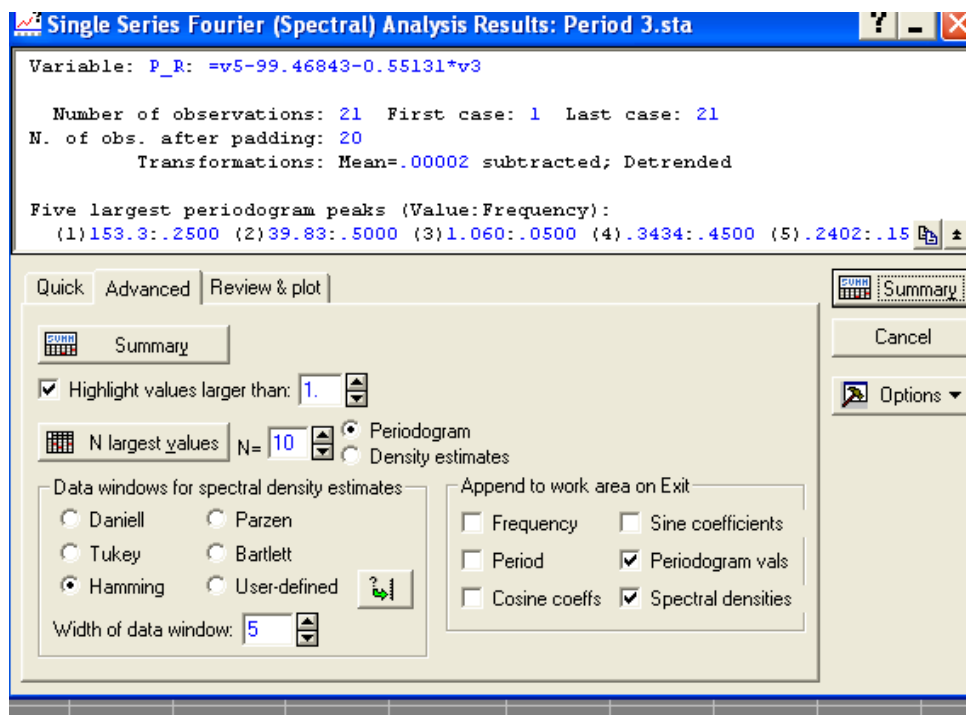


Рис. 8.33. Визначення вікна згладжування

Spectral analysis: P_R : =v5-99.46843-0.55131*v3 (Period 3.sta)							
No. of cases: 20							
	Frequency	Period	Cosine Coeffs	Sine Coeffs	Periodog	Density	Parzen Weights
0	0.000000		0.00860	0.00000	0.0007	0.3537	0.000000
1	0.050000	20.00000	0.04953	0.32171	1.0595	0.7447	0.166667
2	0.100000	10.00000	0.02323	0.14959	0.2292	0.3694	0.666667
3	0.150000	6.66667	0.06800	0.13928	0.2402	0.2141	0.166667
4	0.200000	5.00000	-0.09452	0.02317	0.0947	25.6590	0.000000
5	0.250000	4.00000	1.46931	-3.62969	153.3351	102.2460	
6	0.300000	3.33333	0.05789	-0.02723	0.0409	25.5865	
7	0.350000	2.85714	0.03814	-0.02337	0.0200	0.0256	
8	0.400000	2.50000	0.02064	0.05306	0.0324	0.0822	
9	0.450000	2.22222	0.14158	-0.11955	0.3434	6.8723	
10	0.500000	2.00000	-1.99569	0.00000	39.8277	26.6663	

а) спектральне вікно Парзена

Spectral analysis: P_R : =v5-99.46843-0.55131*v3 (Period 3.sta)							
No. of cases: 20							
	Frequency	Period	Cosine Coeffs	Sine Coeffs	Periodog	Density	Daniell Weights
0	0.000000		0.00860	0.00000	0.0007	0.51562	0.200000
1	0.050000	20.00000	0.04953	0.32171	1.0595	0.51783	0.200000
2	0.100000	10.00000	0.02323	0.14959	0.2292	0.32487	0.200000
3	0.150000	6.66667	0.06800	0.13928	0.2402	30.99174	0.200000
4	0.200000	5.00000	-0.09452	0.02317	0.0947	30.78804	0.200000
5	0.250000	4.00000	1.46931	-3.62969	153.335	30.74620	
6	0.300000	3.33333	0.05789	-0.02723	0.0409	30.70464	
7	0.350000	2.85714	0.03814	-0.02337	0.0200	30.75437	
8	0.400000	2.50000	0.02064	0.05306	0.0324	8.05289	
9	0.450000	2.22222	0.14158	-0.11955	0.3434	8.11338	
10	0.500000	2.00000	-1.99569	0.00000	39.8277	8.11586	

б) спектральне вікно Данієля

Рис. 8.34. Результати спектрального аналізу

Таким чином, модель періодичної складової часового ряду має вид:

$$p_t = 1,46931 \cos \frac{2\pi}{4}(t-1) - 3,62969 \sin \frac{2\pi}{4}(t-1) + \varepsilon_t.$$

Знайти розрахункові значення рівнів ряду з урахуванням тренда та періодичної (сезонної) складової на підставі комбінованої моделі, яка має вид:

$$y_t = 99,46843 + 0,55131t + 1,46931 \cos \frac{2\pi}{4}(t-1) - 3,62969 \sin \frac{2\pi}{4}(t-1) + \varepsilon_t.$$

Результати порівняльного аналізу розрахункових та фактичних значень обсягів продажу підприємства (рис. 8.35) дозволяє зробити висновок про задовільну якість апроксимації.

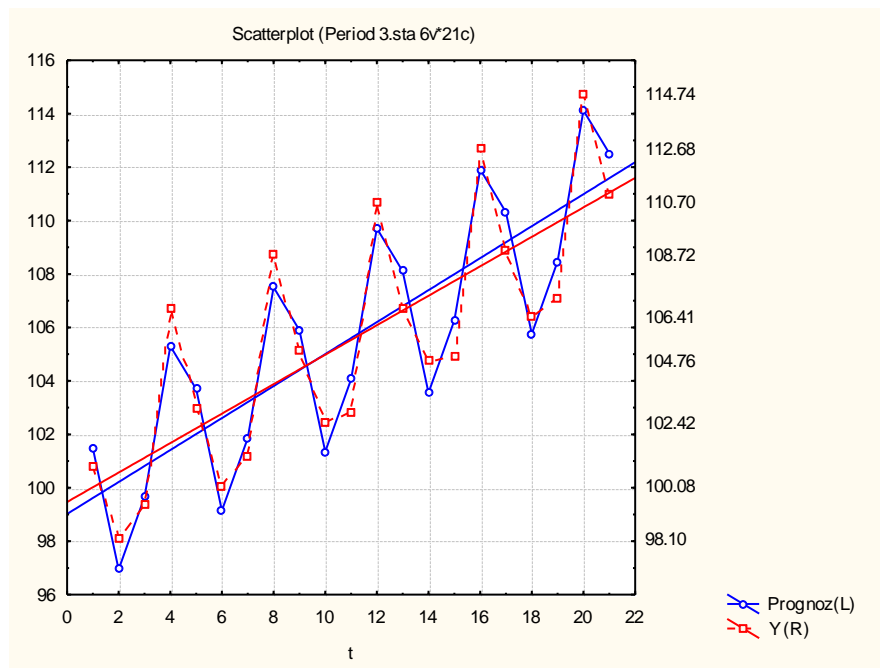


Рис. 8.35. Графік розрахункових та фактичних значень обсягів продажу

5. Оцінювання якості прогнозу. Для прогнозування динаміки об'ємів продажів необхідно вибрати з побудованих моделей найбільш адекватну. Оцінювання якості моделей часових рядів здійснюється за такими критеріями адекватності: середня помилка, середня абсолютна помилка, середнє квадратичне відхилення помилок, середня процентна помилка, середня абсолютна процентна помилка. Порівняння даних моделей за цими критеріями наведено у табл. 8.2.

Таблиця 8.2

Критерії якості моделей

Критерії якості прогнозу / Модель	Мультиплікативна тренд-сезонна модель	Адитивна тренд-сезонна модель	Регресійна модель із <i>dummy</i> -змінними	Комбінована модель
середня помилка	0.000299	-0.000962	-0.000002	0.069950
середня абсолютна помилка	0.153678	0.122393	0.121724	1.024388
сума квадратів помилок	0.835731	0.602937	0.589802	24.39331
середньоквадратична помилка	0.19949	0.16944	0.16759	1.07777
середньовідсоткова помилка	0.00247	-0.00024	0.00030	0.06673
середня абсолютна процентна помилка	0.148286	0.119181	0.118324	0.969248

Отже, на основі оцінювання та аналізу моделей часового ряду можна зробити висновок, що побудовані моделі є адекватними та забезпечують високу точність прогнозу. Але найкращу якість апроксимації забезпечує регресійна модель з *dummy*-змінними та адитивна тренд-сезонна модель.

Щоб розрахувати прогнозні значення на підставі регресійної моделі з *dummy*-змінними, у нижній частині вікна результатів регресійного аналізу є опція *Predict dependent variable* (Прогнозування залежної змінної), ініціювавши дану опцію необхідно вказати значення фіктивних змінних та фактора часу, для яких необхідно спрогнозувати залежну величину.

Результати прогнозування наведені на рис. 8.36.

Predicting Values for (Period 2 variable: Y)			
Variable	B-Weight	Value	B-Weight * Value
t	0.50355	22.00000	11.0782
d1	-4.29745	0.00000	0.0000
d2	-7.35489	1.00000	-7.3549
d3	-7.14245	0.00000	0.0000
Intercept			104.6734
Predicted			108.3966
-95.0%CL			108.1426
+95.0%CL			108.6506

а) прогнозне значення обсягу продажів на II квартал

Predicting Values for (Period 2 variable: Y)			
Variable	B-Weight	Value	B-Weight * Value
t	0.50355	23.00000	11.5817
d1	-4.29745	0.00000	0.0000
d2	-7.35489	0.00000	0.0000
d3	-7.14245	1.00000	-7.1424
Intercept			104.6734
Predicted			109.1126
-95.0%CL			108.8586
+95.0%CL			109.3666

б) прогнозне значення обсягу продажів на III квартал

Predicting Values for (Period 2 variable: Y)			
Variable	B-Weight	Value	B-Weight * Value
t	0.50355	24.00000	12.0853
d1	-4.29745	0.00000	0.0000
d2	-7.35489	0.00000	0.0000
d3	-7.14245	0.00000	0.0000
Intercept			104.6734
Predicted			116.7586
-95.0%CL			116.5046
+95.0%CL			117.0126

в) прогнозне значення обсягу продажів на IV квартал

Рис. 8.36. Прогнозні значення обсягу продажів

Для того щоб знайти прогнозні значення об'єму продажів на 3 квартали уперед за допомогою адитивної тренд-сезонної моделі часового ряду, необхідно виконати ряд дій:

- 1) додати 3 нових спостереження після останнього;
- 2) у стовпці даних *Період часу (t)* вписати відповідні порядкові числівники (продовжуючи ряд);
- 3) у стовпці *Seasonal Factors* вписати відповідні значення сезонних індексів для відповідних номерів кварталу року;
- 4) у стовпці прогнозних значень (*Prognoz*) задати перерахування даних (*Vars/Recalculate*) відповідно до моделі часового ряду $Prognoz_t = 99,97 + 0,50395 * t + Seasonal\ factors$.

Результати розрахунку прогнозних значень показника об'єму продажів наведено на рис. 8.37. На рис. 8.38 подано графік динаміки прогнозних значень об'ємів продажів.

	1 Year	2 Quater	3 t	4 Prognoz (add)	5 Seasonal (add)
1	1	I	1	100.861	0.38734
2		II	2	98.366	-2.61203
3		III	3	99.025	-2.45666
4		IV	4	106.667	4.68134
5	2	I	5	102.877	0.38734
6		II	6	100.382	-2.61203
7		III	7	101.041	-2.45666
8		IV	8	108.683	4.68134
9	3	I	9	104.893	0.38734
10		II	10	102.397	-2.61203
11		III	11	103.057	-2.45666
12		IV	12	110.699	4.68134
13	4	I	13	106.909	0.38734
14		II	14	104.413	-2.61203
15		III	15	105.073	-2.45666
16		IV	16	112.715	4.68134
17	5	I	17	108.924	0.38734
18		II	18	106.429	-2.61203
19		III	19	107.088	-2.45666
20		IV	20	114.730	4.68134
21	6	I	21	110.940	0.38734
22		II	22	108.445	-2.61203
23		III	23	109.104	-2.45666
24		IV	24	116.746	4.68134

Рис. 8.37. Таблиця розрахунку прогнозних значень

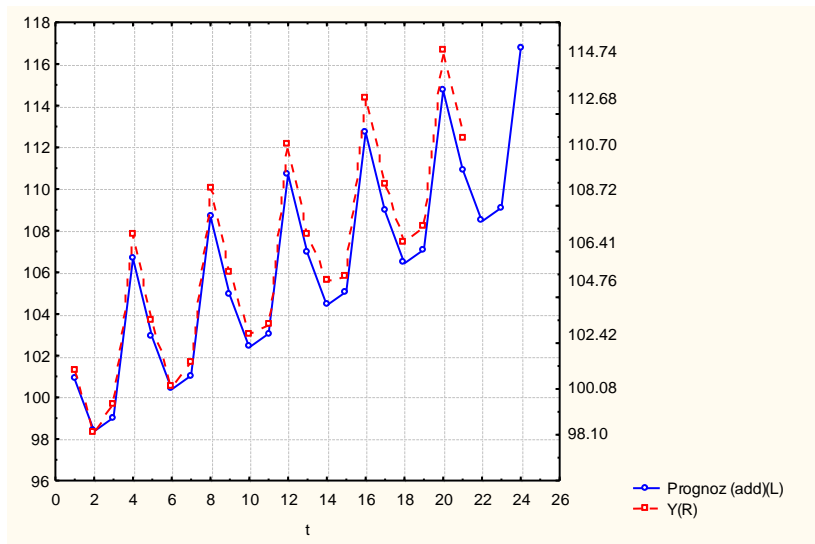


Рис. 8.38. Графік прогнозних значень об'ємів продажів

Отже, можна зробити висновок, що регресійна модель із *dumtmy*-змінними та адитивна тренд-сезонна модель показують близькі за значеннями точкові прогнози обсягу продажів. Однак побудована регресійна модель забезпечує більш високу точність, тому прогноз, отриманий за цією моделлю, можна сприйняти як базовий.

Лабораторна робота 9

Реалізація моделей часових рядів за допомогою моделей авторегресії і ковзного середнього

Мета – закріплення теоретичного та практичного матеріалу, формування навичок побудови моделі авторегресії – проінтегрованого ковзного середнього у модулі *Time Series/Forecasting – Часові ряди/Прогнозування* ППП *Statistica*.

Умови завдання

У таблиці 9.1 наведені дані, що відбивають динаміку запасів підприємства (y_t).

Вихідні дані, тис. шт.

t	Y _t	t	Y _t	T	Y _t	t	Y _t
1	2	3	4	5	6	7	8
1	267	14	241	27	258	40	279
2	267	15	244	28	259	41	287
3	268	16	254	29	268	42	276
4	264	17	262	30	276	43	273
5	263	18	261	31	285	44	270
6	260	19	265	32	288	45	264
7	256	20	261	33	295	46	261
8	256	21	261	34	297	47	268
9	252	22	257	35	292	48	270
10	245	23	268	36	299	49	276
11	243	24	270	37	294	50	274
12	240	25	266	38	284	51	284
13	238	26	259	39	277	52	304

Необхідно

1. Побудувати графік вихідних даних.
2. Розрахувати коефіцієнти автокореляцій, часткових автокореляцій. Проаналізувати дану інформацію для розробки адекватної прогнозної моделі.
3. Перевірити ряд на стаціонарність. У випадку нестаціонарності даних провести їх корегування.
4. Побудувати *ARIMA*-модель і провести тестування її адекватності.

Методичні рекомендації до виконання завдання

Ідентифікацію моделі *ARIMA* (тобто ідентифікацію порядку різниці й ідентифікацію стаціонарного процесу) у середовищі *Statistica* можна здійснити в діалоговому вікні *Transformations, autocorrelations, plots* (Перетворення змінних, автокореляції, графіку), що відкривається зі стартової панелі модуля *Time series/forecasting* (Часові ряди/Прогнозування) або в діалозі *ARIMA&autocorrelation functions* (APIКC і автокореляційні функції) (рис. 9.1).

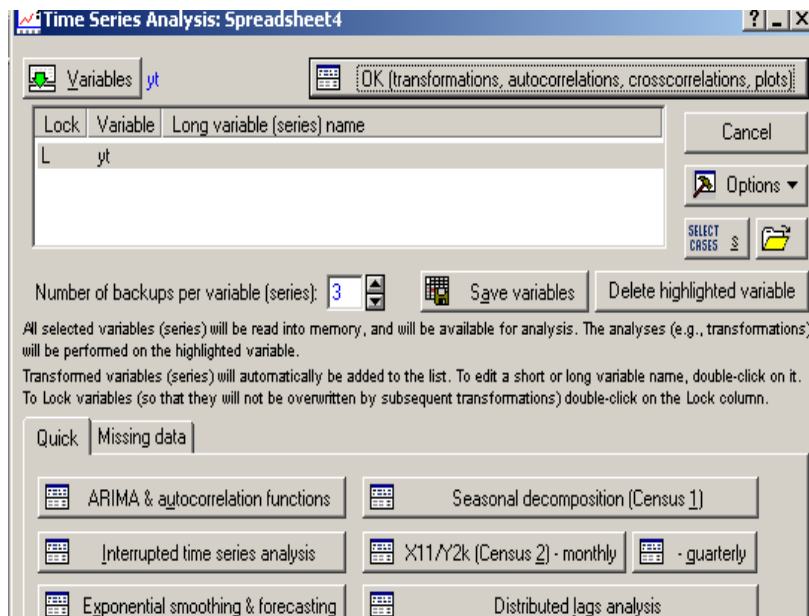


Рис. 9.1. Стартова панель модуля *Time series/forecasting*

Перед аналізом необхідно визначити – є ряд стаціонарним, чи ні. Графічне подання ряду можна одержати за допомогою натискання кнопки *Review&plot* (Огляд і графіки) (рис. 9.2).

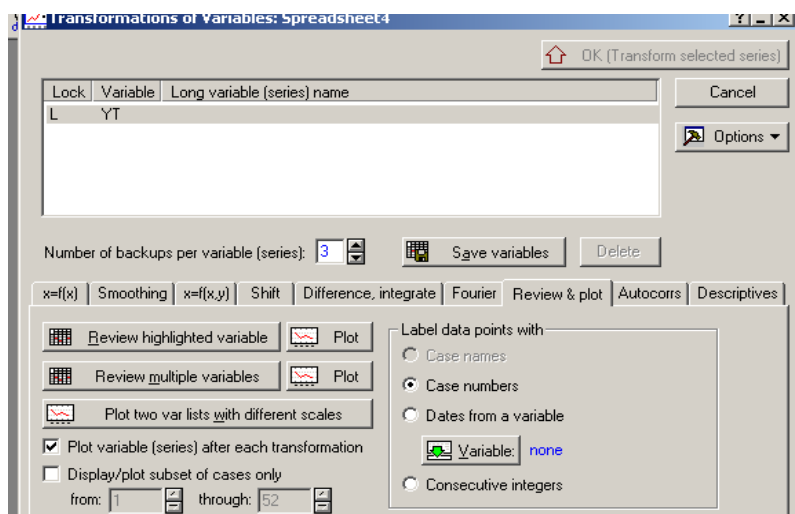


Рис. 9.2. Панель діалогу *Transformations, autocorrelations, plots* – перетворення, автокореляції, графіки

Якщо тренд не є чітко вираженим і немає інших особливостей ряду, що вказують на нестационарність, то варто розглянути вибірккову автокореляційну функцію, тобто оцінку автокореляційної функції. У випадку, коли автокореляційна функція не має тенденції до загасання, можна зробити висновок про нестационарність ряду. Якщо ряд є стаціонарним, то можна

перейти до визначення параметрів авторегресійної складової і ковзного середнього. У протилежному випадку варто привести ряд до стаціонарного вигляду. Для переходу до стаціонарного ряду традиційно застосовують оператор послідовних різниць. Якщо необхідний для стаціонарності ряду ступінь різниці досягнутий, то автокореляційна функція буде мати тенденцію до загасання. Графік досліджуваного часового ряду наведений на рис. 9.3.

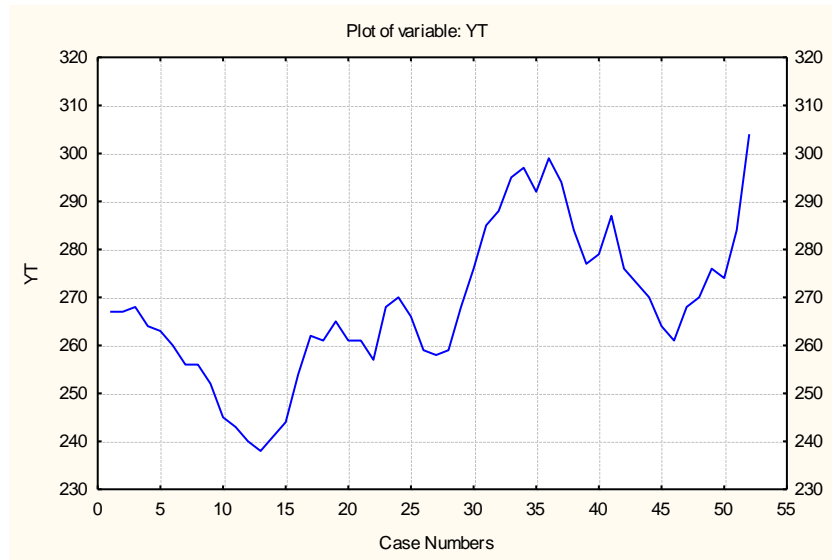


Рис. 9.3. Динаміка запасів підприємства

Як видно з рис. 9.3, досліджуваний часовий ряд має тенденцію до росту, тобто він не є стаціонарним. Для переходу до стаціонарного ряду застосовується оператор узяття послідовних різниць. У *Statistica* розрахунок перших різниць здійснюється в діалоговому вікні *Перетворення змінних* за допомогою опції *Differencing* (процедура дискретного диференціювання) (рис. 9.4).

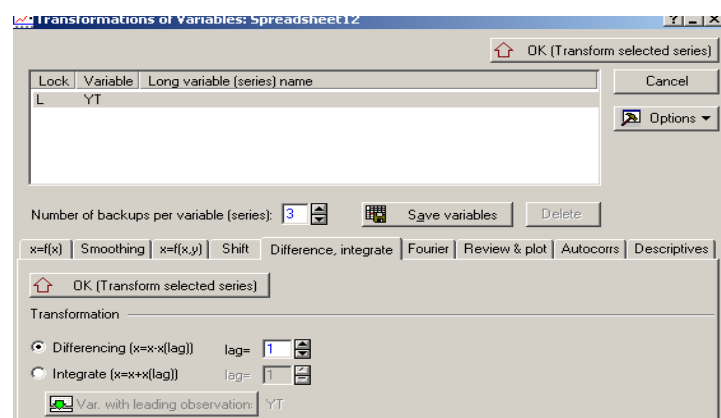


Рис. 9.4. Панель діалогу *Transformations, autocorrelations, plots* – перетворення, автокореляції, графіки

Для отриманого ряду перших різниць застосовують критерій стаціонарності. Якщо ряд перших різниць нестационарний, то беруть знову його різниці першого порядку та застосовують критерій стаціонарності. Процедура закінчується на кроці k , якщо ряд став стаціонарним. При цьому порядок різниці в $ARIMA$ -моделі приймається рівним k . Графік перших різниць ряду наведений на рис. 9.5.

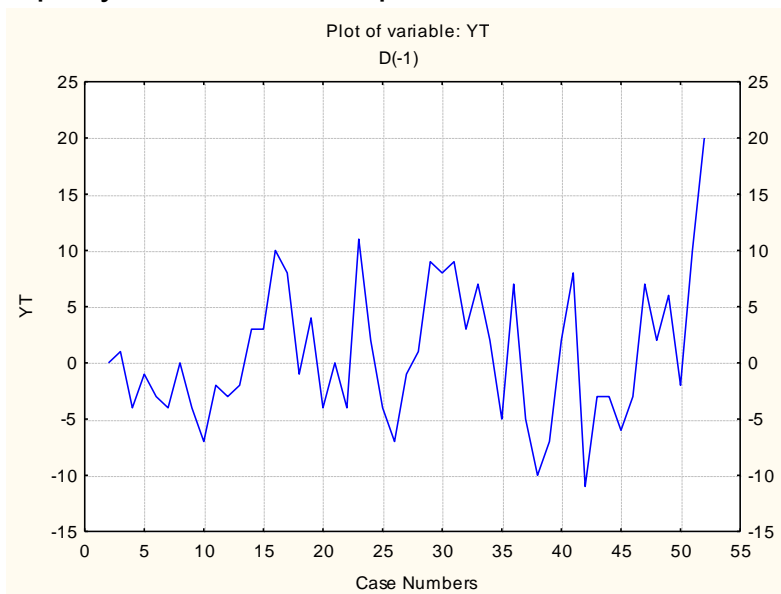


Рис. 9.5. Графік перших різниць

Натискання кнопки *Autocorrelations* (автокореляції) запускає процедуру оцінювання автокореляцій ряду для різних лагів (максимальне число лагів вказується в опції *Number of lags* – число лагів). Аналогічне натискання кнопки *Partial autocorrelations* (часні автокореляції) запускає процедуру оцінювання часткових автокореляцій для різних лагів (рис. 9.6).

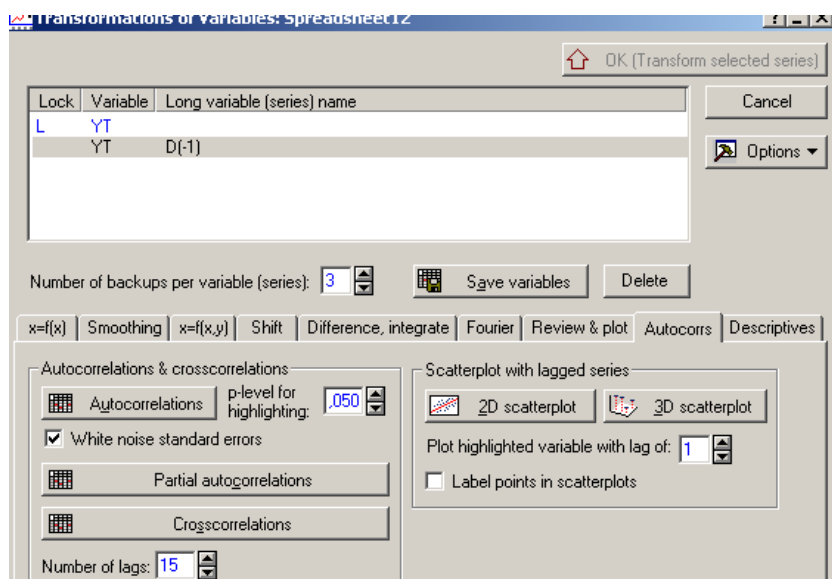


Рис. 9.6. Панель діалогу *Transformations, autocorrelations, plots* – перетворення, автокореляції, графіки

Автокореляційна функція та часткова автокореляційна функція процесу наведені на рис. 9.7 і 9.8.

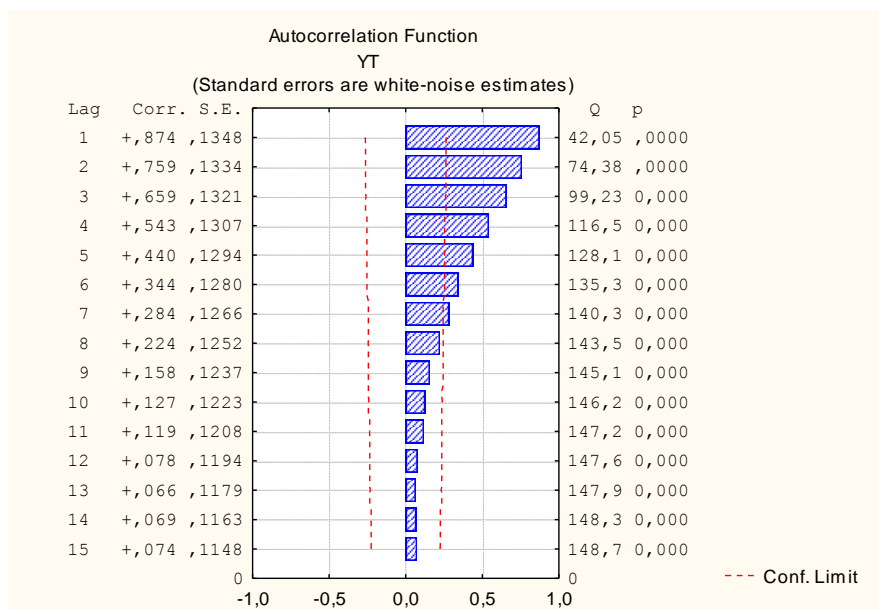


Рис. 9.7. Автокореляційна функція процесу

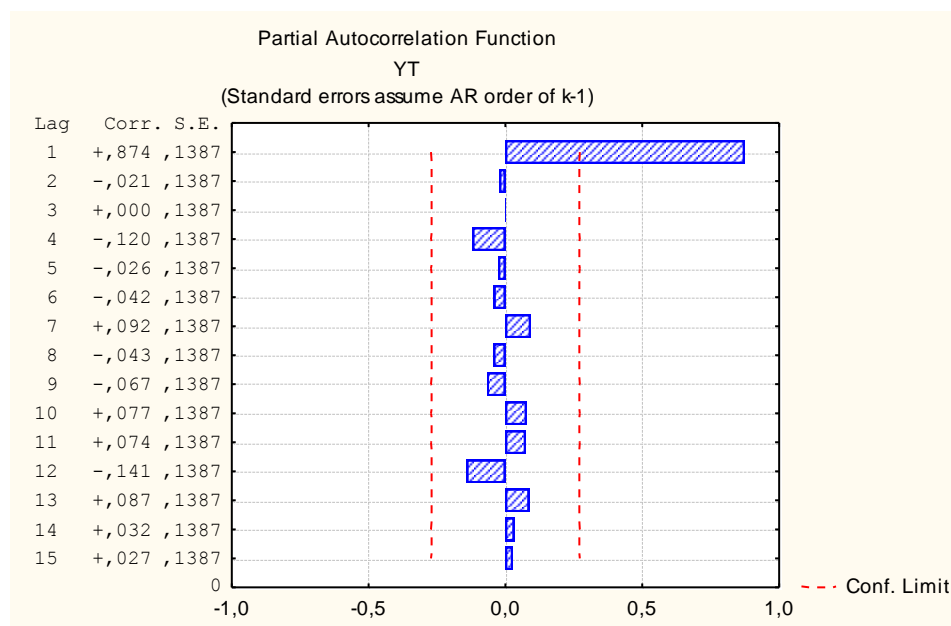


Рис. 9.8. Часткова автокореляційна функція процесу

Як видно з рис. 9.7 і 9.8, поведження автокореляційної функції (плавне зменшення коефіцієнтів автокореляції) і часткової автокореляційної функції (обрив після першого значення) характерне для моделі авторегресії першого порядку.

Для оцінювання параметрів *ARIMA*-моделі в середовищі *Statistica* необхідно вибрати вкладку *ARIMA&autocorrelation functions* (*APIKC* і автокореляційні функції) на стартовій панелі модуля *Time Series/Forecasting* (*Часові ряди/Прогнозування*). На екрані з'явиться стартова панель діалогу *ARIMA&autocorrelation functions* (рис. 9.9).

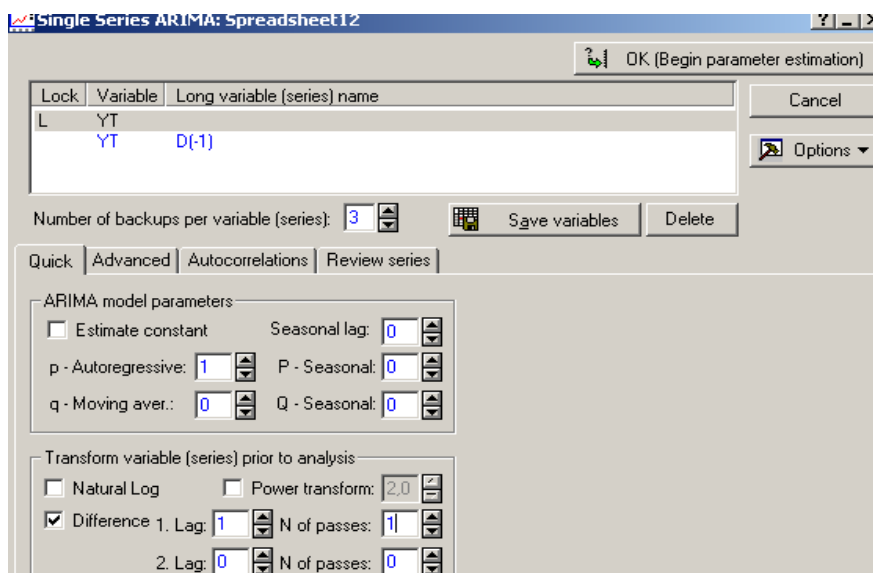


Рис. 9.9. Стартова панель діалогу *ARIMA&autocorrelation functions* – *APIKC* і автокореляційні функції

Група опцій *ARIMA model parameters* – *Параметри моделі авторегресії* – проінтегрованого ковзного середнього дозволяє задати параметри для повної сезонної моделі *ARIMA*:

p – *Autoregressive* – порядок авторегресії;

q – *Moving aver* – порядок ковзного середнього;

Seasonal lag – сезонний лаг;

P – *Seasonal* – сезонний параметр авторегресії;

Q – *Seasonal* – сезонний параметр ковзного середнього;

Estimate constant – включити константу в модель.

Група опцій *Transform variable* (*Перетворення змінної*) дозволяє вибрати вид необхідного перетворення:

Natural log – взяти натуральний логарифм;

Difference – обчислити різницю;

Power transform – підносити до степеня.

Після запуску обчислювальної процедури можна переглянути результати оцінювання (рис. 9.10).

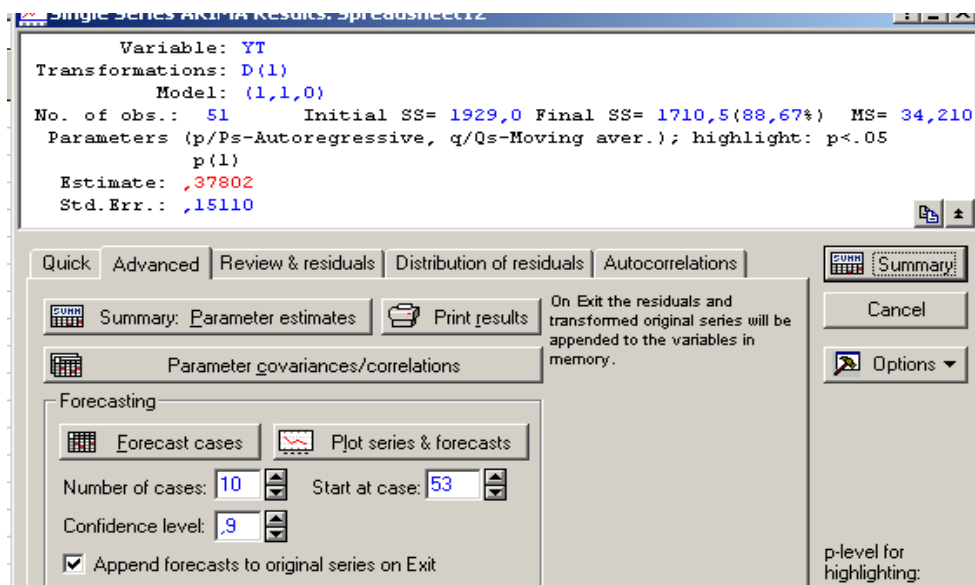


Рис. 9.10. Вікно результатів оцінювання

Верхня частина вікна результатів оцінювання – інформаційна – містить основну інформацію про модель, нижня частина вікна – функціональна – дозволяє вибрати такі напрямки аналізу:

1. *Summary: Parameter estimates* – перегляд основних характеристик моделі;
2. *Autocorrelations* – оцінювання автокореляцій залишків моделі;
3. *Forecast cases* – прогнозування.

Вибір першого напрямку дозволяє переглянути таблицю з основними характеристиками моделі (рис. 9.11), що містить: *Param.* – оцінки коефіцієнтів моделі; *Std. error* – оцінки стандартної помилки; *t* – критерій Стьюдента; *p-level* – рівень значимості критерію Стьюдента; *Lower 95 % Con*, *Upper 95 % Con* – довірчий інтервал для оцінок параметрів моделі.

	Input: YT (Spreadsheet12)					
	Transformations: D(1)					
	Model:(1,1,0) MS Residual=34,210					
Paramet.	Param.	Asympt. Std.Err.	Asympt. t(50)	p	Lower 95% Conf	Upper 95% Conf
p(1)	0,378025	0,151098	2,501845	0,015672	0,074535	0,681515

Рис. 9.11. Основні характеристики моделі

Вибір другого напрямку (рис. 9.10) дозволяє переглянути значення автокореляційної (**autocorrelations**) і часткової автокореляційної (**partial autocorrelations**) функцій залишків моделі (рис. 9.12, 9.13).

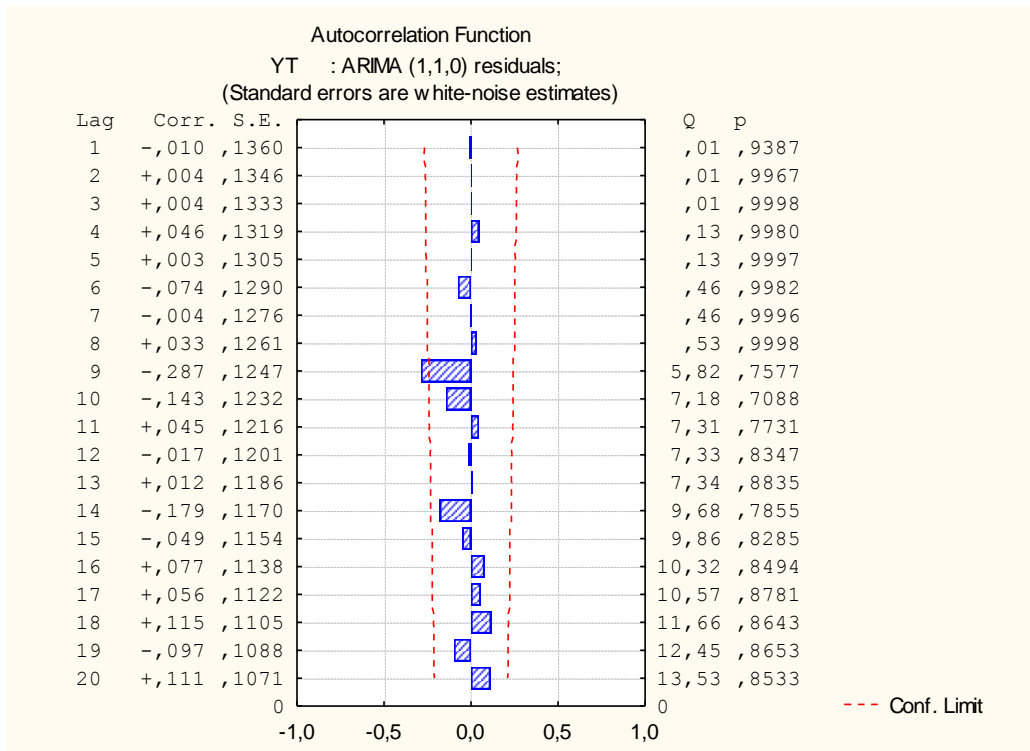


Рис. 9.12. Автокореляційна функція залишків моделі

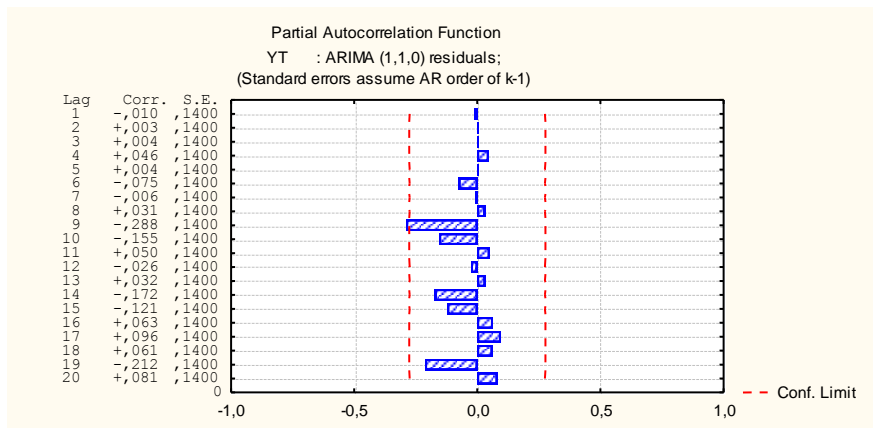


Рис. 9.13. Часткова автокореляційна функція залишків моделі

Як видно з рис. 9.12, Q-Статистика Бокса-Пірса дорівнює 13,53, імовірність того, що залишки становлять білий шум, дорівнює 0,85. При виборі рівня значимості $\alpha = 0,05$ параметр рівняння буде статистично значимий ($p < 0,05$). У цілому, запропонована модель адекватна досліджуваному процесу.

Вибір третього напрямку (рис. 9.10) дає можливість побудувати прогноз майбутніх значень ряду. У рядку *Number of cases* (Число випадків) задається число випадків, на які необхідно побудувати прогноз. Опція *Start at case* (Почати з випадку) дозволяє задати номер випадку, з якого починає будуватися прогноз. Функціональна клавіша *Plot series&forecasts*

(Графік ряду й прогнозу) дає можливість побудувати графік ряду із прогнозом. Перед натисканням функціональної клавіші встановлюється опція *Append forecasts to original series on Exit* (Приєднати прогнози до кінця вихідної серії).

Прогнозні значення запасів досліджуваного підприємства наведені на рис. 9.14.

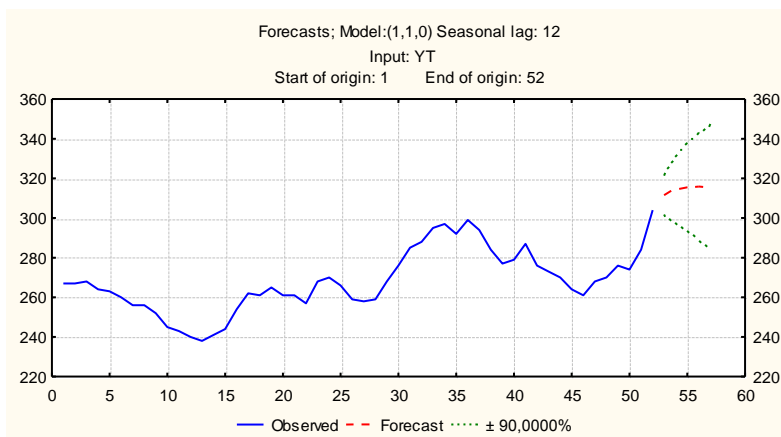


Рис. 9.14. Результати прогнозування

Лабораторна робота 10

Реалізація експертних методів прогнозування, розрахунки непараметричних статистик, аналіз відповідностей

Мета – закріплення теоретичного та практичного матеріалу, формування навичок обробки результатів експертного опитування у модулі *Nonparametrics – Непараметрична статистика* ППП *Statistica*.

Умови завдання 1. Коефіцієнт рангової кореляції Спірмена, тау Кендала, гама (*Spearman rank R, Kendall Tau, Gamma*)

У табл. 10.1 наведені дані, що відображають результати соціологічної анкети 10 працівників організації (бальні оцінки якостей працівників) за трьома властивостями (успішність кар'єри, рівень лідерських якостей, моральний клімат у колективі). Кожна оцінка подана натуральним числом, 1 – привласнюється максимальній оцінці, n – мінімальній.

Таблиця 10.1

Бальні оцінки якостей працівників

Змінні/№ спостереження	1	2	3	4	5	6	7	8	9	10
Успішність кар'єри	2	2	5	3	4	6	1	7	7	1
Рівень лідерських якостей	1	1	3	2	3	5	3	6	5	2
Моральний клімат у колективі	5	7	2	3	1	2	4	2	3	6

Необхідно

1. Розрахувати коефіцієнти рангової кореляції Спірмена, тау Кендала, гама.
2. Оцінити їх статистичну значущість за відповідними критеріями.
3. Зробити висновки щодо тісноти зв'язку між властивостями якостей працівників.

Методичні рекомендації до виконання завдання

Слід розглянути розрахунок коефіцієнтів рангової кореляції у пакеті прикладних програм *Statistica*. Аналіз непараметричних критеріїв здійснюється в модулі *Nonparametrics/Correlations* (*Spearman*, *Kendall tau*, *gamma*) – *Непараметрична статистика/Рангові кореляції*. Вибір модуля та методу дослідження наведені на рис. 10.1 і 10.2.

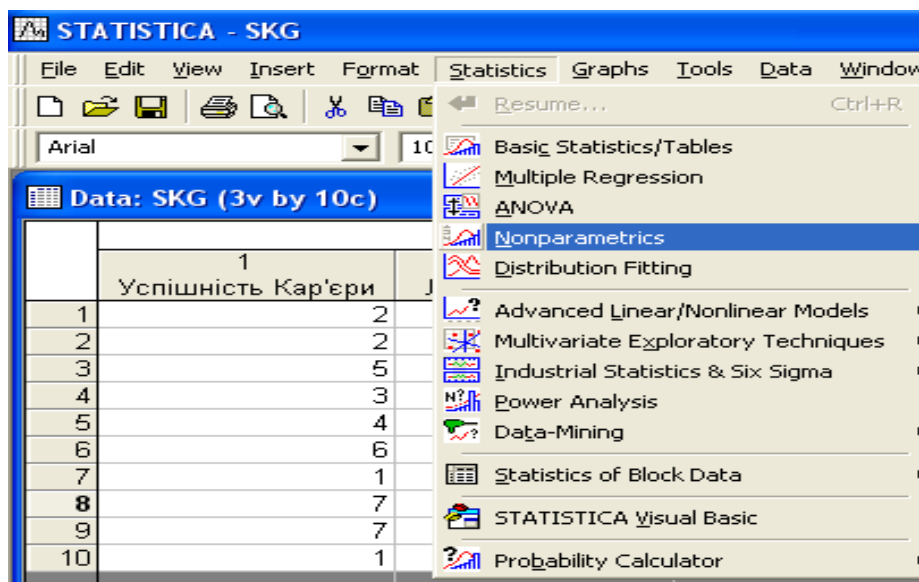


Рис. 10.1. Вибір модуля

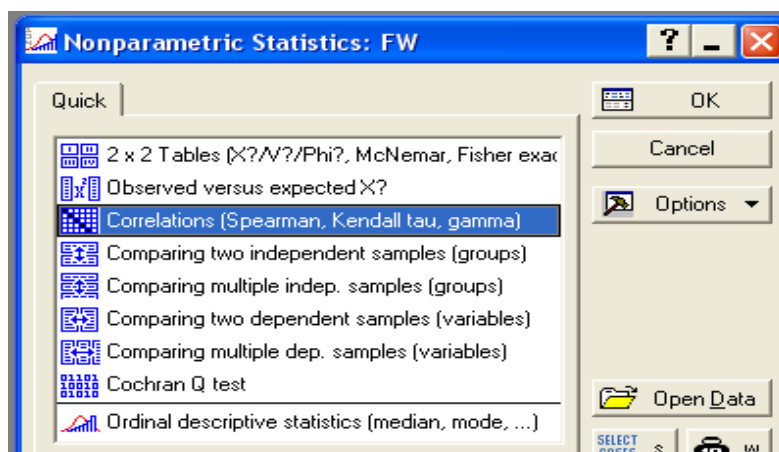


Рис. 10.2. Вибір методу рішення

Для рішення задачі дані вводяться таким чином: k – змінні (vars) за якими оцінюються респонденти, кожна з яких має n – значень (cases), записаних у бальній шкалі. Вибір змінних для аналізу та вибір виду подання інформації (*Detailed report* – Детальний звіт) наведені на рис. 10.3 і 10.4.

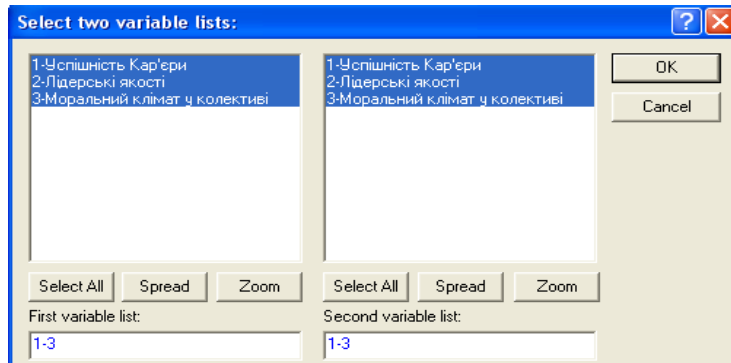


Рис. 10.3. Вибір змінних

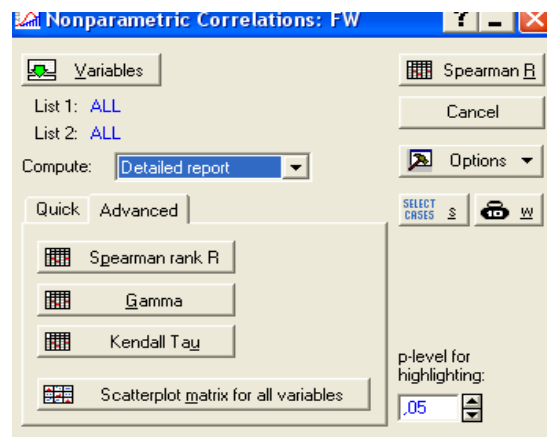


Рис. 10.4. Вибір напрямку дослідження та виду подання інформації

Ініціювавши клавішу *Spearman rank R* будуть отримані коефіцієнти рангової кореляції Спірмена між досліджуваними змінними, їх статистичну значущість та рівень довірчої ймовірності (рис. 10.5).

Pair of Variables	Spearman Rank Order Correlations (Spearman R)			
	Valid N	Spearman R	t(N-2)	p-level
Успішність Кар'єри & Лідерські якості	10	0,76881	3,40047	0,00935
Успішність Кар'єри & Моральний клімат у колективі	10	0,82972	4,20447	0,00297
Лідерські якості & Моральний клімат у колективі	10	0,84329	4,43802	0,00217

Рис. 10.5. Коефіцієнт рангової кореляції Спірмена

Аналіз коефіцієнтів дозволяє зробити такі висновки: для досліджуваних респондентів є досить суттєвий прямий зв'язок між успішністю кар'єри, лідерськими якостями та моральним кліматом у колективі. Всі оцінки статистично значимі за t-критерієм. Найбільш взаємозалежні лідерські якості працівника та моральний клімат у колективі, моральний клімат суттєво впливає на успішність кар'єри, та в останню чергу на успішність кар'єри впливають лідерські якості працівника.

Ініціювавши клавішу *Kendall Tau*, будуть отримані коефіцієнти рангової кореляції Кендала між досліджуваними характеристиками, їх статистичну значущість за Z-критерієм та рівень довірчої ймовірності (рис. 10.6). Даний коефіцієнт також вказує на суттєвий зв'язок між чинниками, що підтверджує Z-критерій. Однак, за критерієм Кендала моральний клімат однаково сильно впливає як на успішність кар'єри, так і на лідерські якості працівників.

Pair of Variables	Kendall Tau Correlations (SKG) MD pairwise deleted Marked correlations are significant at p <			
	Valid N	Kendall Tau	Z	p-level
Успішність Кар'єри & Лідерські якості	10	0,667124	2,685124	0,007250
Успішність Кар'єри & Моральний клімат у колективі	10	0,722944	2,909794	0,003617
Лідерські якості & Моральний клімат у колективі	10	0,725227	2,918981	0,003512

Рис. 10.6. Коефіцієнт рангової кореляції Кендала

Ініціювавши клавішу *Gamma* будуть отримані коефіцієнти гама-кореляції між досліджуваними чинниками у колективі, їх статистична значущість за Z-критерієм та рівень довірчої ймовірності (рис. 10.7). Даний коефіцієнт також вказує на суттєвий зв'язок між всіма чинниками та їх значущість за критерієм Z.

Pair of Variables	Gamma Correlations (SKG) MD pairwise deleted Marked correlations are significant at p <			
	Valid N	Gamma	Z	p-level
Успішність Кар'єри & Лідерські якості	10	0,729730	2,685124	0,007250
Успішність Кар'єри & Моральний клімат у колективі	10	0,789474	2,909794	0,003617
Лідерські якості & Моральний клімат у колективі	10	0,828577	2,918981	0,003512

Рис. 10.7. Коефіцієнт гама-кореляції

Отже, коефіцієнти рангових кореляцій трохи відрізняються один від одного, але їх змістове значення та зв'язок відбивають одні й тіж результати. За всіма критеріями спостерігаємо значний вплив усіх чинників один на одного.

Умови завдання 2. Двофакторний аналіз Фрідмана і коефіцієнт конкордації Кендала (*Friedman ANOVA & Kendall's concordance*).

У таблиці 10.2 наведені дані, що відображають результати експертного оцінювання п'ятьма експертами 9 показників (ранжування), за ступенем їх впливу на продуктивність праці робітників підприємства (X1 – коефіцієнт напруженості норм; X2 – умови праці робітника; X3 – стаж роботи за спеціальністю; X4 – стан тарифної системи; X5 – стан обладнання; X6 – коефіцієнт ритмічності поставок; X7 – збут продукції; X8 – соціально-психологічні умови; X9 – організація роботи цеху). Чиннику, що найбільше впливає на продуктивність праці, присвоюється 1-й ранг, наступному – 2-й ранг і т. д.

Таблиця 10.2

Результати ранжування чинників

	X1	X2	X3	X4	X5	X6	X7	X8	X9
1	1	2	3	2	4	1	1	6	2
2	3	4	5	6	3	2	1	7	4
3	2	3	4	3	3	1	2	5	6
4	1	1	2	4	4	1	1	5	3
5	4	5	6	7	5	3	1	4	2

Необхідно

1. Розрахувати статистику Фрідмана (F).
2. Розрахувати коефіцієнт конкордації Кендала (W), його значущість за критерієм χ^2 .
3. Зробити висновки щодо оцінок експертів та їх погодженості.

Методичні рекомендації до виконання завдання

Слід розглянути розрахунок статистики Фрідмана та коефіцієнта конкордації Кендалла у пакеті прикладних програм *Statistica*. Аналіз даного непараметричного критерію здійснюється в модулі *Nonparametric/*

Comparing multiple dep. Samples (variables)/Friedman ANOVA & Kendall's concordance – Непараметрична статистика/Зрівняння (оцінка) множинних змінних. Вибір модуля та методу дослідження наведені на рис. 10.8.

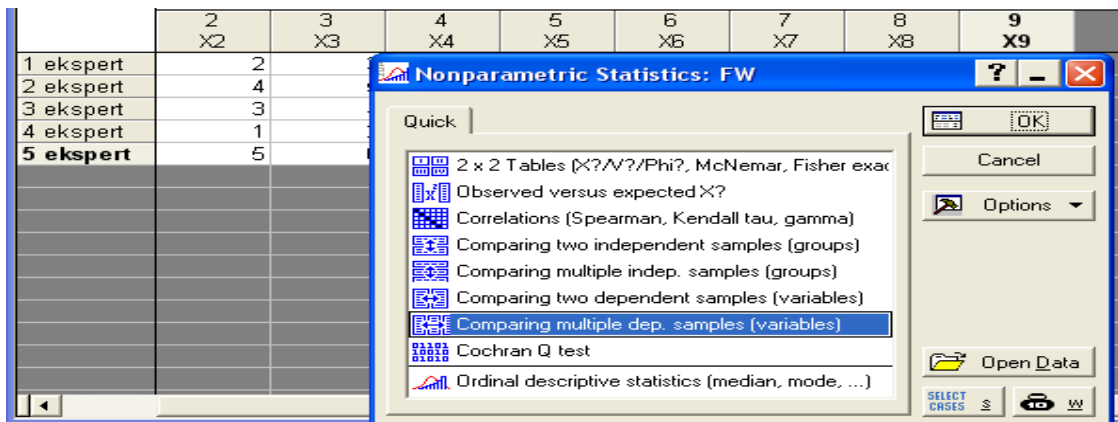


Рис. 10.8. Вибір методу дослідження

Для рішення даної задачі дані вводяться таким чином: k – змінних (vars), які оцінюють експерти (cases), записаних у бальній шкалі. Для вибору змінних для аналізу необхідно ініціювати кнопку (*Variables/Змінні*) і вказати змінні, які аналізуються (рис. 10.9).



Рис. 10.9. Вибір змінних

Після ініціювання кнопки *Summary: Friedman ANOVA & Kendall's concordance* будуть отримані характеристики для кожної змінної: *Average Rank/Стандартизований ранг*, *Sum of Ranks/Сума рангів*, *Persent Mean/Середня оцінка*, *Std. Dev./Середнєквадратичне відхилення*, а також вибіркове значення статистики Фрідмана з кількістю ступеней свободи та рівнем значущості, та коефіцієнт конкордації Кендала (рис. 10.10).

Variable	Friedman ANOVA and Kendall Coeff. of ANOVA Chi Sqr. (N = 5, df = 8) = 27,538 Coeff. of Concordance = ,68846 Aver. ra			
	Average Rank	Sum of Ranks	Percent Mean	Std.Dev.
X1	3,00000	15,0000	2,20000	1,30384
X2	4,90000	24,5000	3,00000	1,58113
X3	6,80000	34,0000	4,00000	1,58113
X4	6,90000	34,5000	4,40000	2,07364
X5	6,10000	30,5000	3,80000	0,83666
X6	2,10000	10,5000	1,60000	0,89442
X7	1,80000	9,0000	1,20000	0,44721
X8	7,90000	39,5000	5,40000	1,14017
X9	5,50000	27,5000	3,40000	1,67332

Рис. 10.10. Таблиця результатів

На рис. 10.11 наведена діаграма розмаху оцінок за кожним із показників від найменшої до найбільшої, а також медіана оцінки.

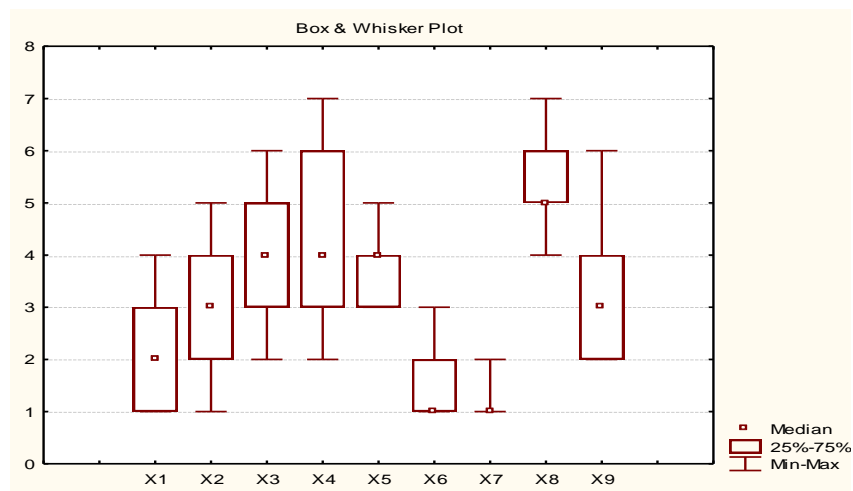


Рис. 10.11. Діаграма розмаху

Значення статистики Фрідмана дорівнює: $F_B = 27,53$ Табличне значення критерію χ^2 на рівні значущості $\alpha = 0,05$ дорівнює $\chi_{0,05}^2 8 = 15,51$. Отже, на рівні значущості $\alpha = 0,05$ приймаємо гіпотезу H_0 і робимо висновок, що немає значущих розбіжностей.

Коефіцієнт Конкордації дорівнює $W = 0,69$, отже гіпотеза про випадковість думок експертів відкидається. З імовірністю 95 % можна стверджувати, що існує певна погодженість думок експертів щодо чинників, що впливають на продуктивність праці працівників підприємства.

Лабораторна робота 11

Прогнозування на підставі нейронних мереж

Мета – формування у студентів навичок побудови моделей нейронних мереж, оцінювання якості моделей нейронних мереж, вибору моделей для розробки прогнозів різних типів у модулі *Statistica NeuralNetworks* ППП *Statistica*.

Умови завдання

Необхідно побудувати модель класифікації підприємств і провести розпізнавання для вибіркового даних у модулі *Statistica NeuralNetworks* ППП *Statistica*:

1. Побудувати модель нейронної мережі за допомогою *Майстра рішення задач (Intelligent Problem Solver)*.
2. Оцінити якість побудованих моделей різних типів нейронних мереж. Побудувати архітектуру мережі, провести аналіз чуттєвості, проаналізувати похибки моделі розпізнавання.
3. Проаналізувати результати класифікації (статистика та коректність класифікації), навести теоретичну класифікацію за нейронною моделлю.
4. Зробити висновки та прогнози за побудованими моделями.

Методичні рекомендації до виконання завдання

Для рішення й аналізу задач даного типу в ППП *Statistica* передбачений модуль *Neural Networks (Нейронні мережі)*. Для рішення поставленої задачі варто скористатися *Майстром рішення задач (Intelligent Problem Solver)*. Слід розглянути порядок роботи в даному модулі.

У прикладі розглядається модель розпізнавання фінансового стану підприємства, що дозволяє визначити поточний стан підприємства шляхом віднесення його до одного із трьох сформованих класів фінансових ситуацій: кризового, нормального або успішного.

Першим етапом побудови нейронної моделі є ретельний відбір вхідних даних, що впливають на очікуваний результат. У якості вхідних даних скористатися такими показниками: (коефіцієнт абсолютної ліквідності (x_1); коефіцієнт термінової ліквідності (x_2); коефіцієнт загальної поточної ліквідності (x_3); рентабельність сукупного капіталу (x_4); рентабельність власного капіталу (x_5); валова рентабельність продажів (x_6);

операційна рентабельність продажів (x_7); чиста рентабельність продажів (x_8). Вихідна вибірка для побудови нейромережі складається зі значень фінансових показників 87 підприємств, що розділені на три класи.

Для рішення поставленої задачі необхідно скористатися *Майстром рішення задач (Intelligent Problem Solver)*.

У першому вікні вибрати режим *Advanced* (рис. 11.1) і задати тип задачі *Standard*, за допомогою якої будуються нейронні мережі класифікації (тип задачі *Time Series* передбачає побудову моделей часових рядів (рис. 11.2)).

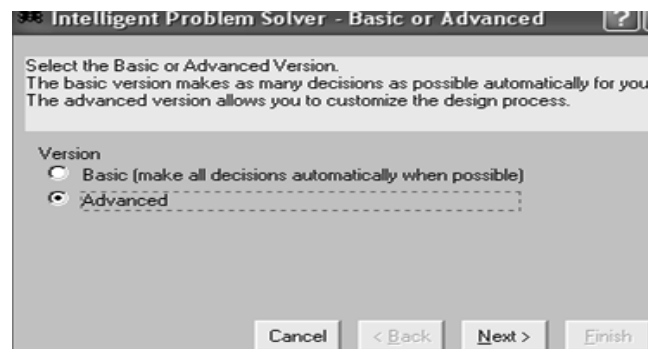


Рис. 11.1. Вибір режиму моделювання

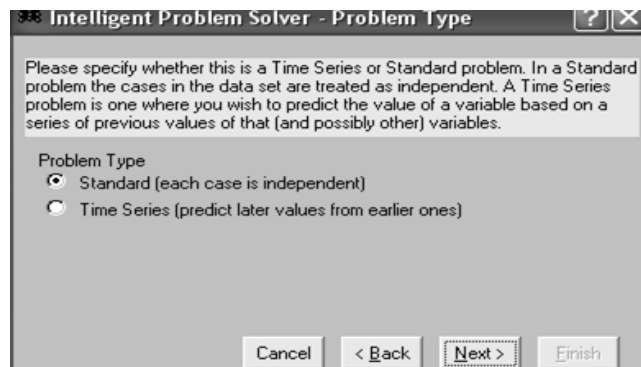


Рис. 11.2. Вибір типу задачі

Далі необхідно задати залежні й незалежні змінні моделі (рис. 11.3), залежною змінною є змінна *Class*. У якості залежної змінної моделі будуть виступати класи фінансових ситуацій підприємства: кризовий, нормальний і успішний.

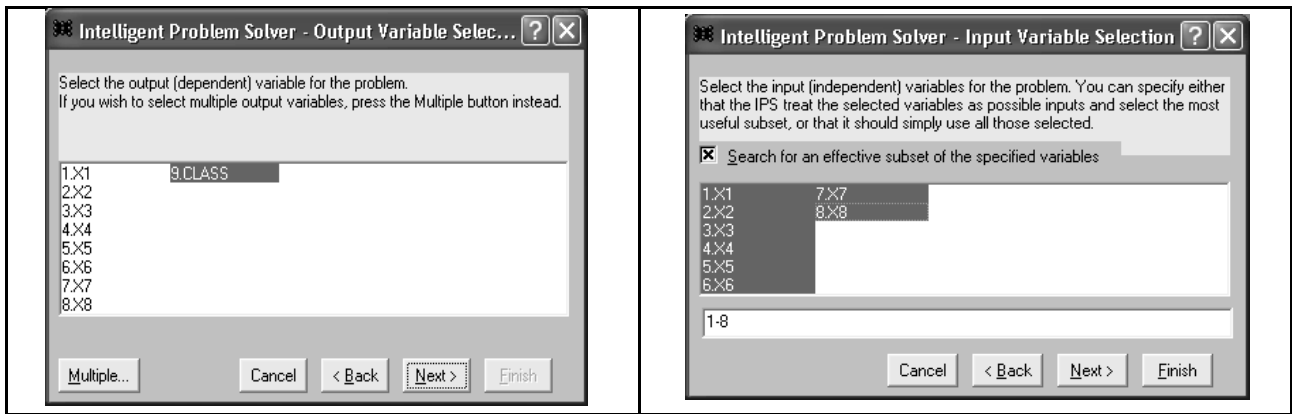


Рис. 11.3. Вибір змінних моделі

Наступним кроком є визначення навчальної, контрольної і тестової множини. Загальна ідея завдання навчальної вибірки полягає в такому: спочатку на вхід мережі подається навчальна вибірка з відомими результатами, величини X і спостерігаються відгуки $Y = F(X)$. Змінюючи ваги $\omega(i; j)$ та значення порога активації для кожного нейрона, настроїти мережу (знайти як можна більш точне наближення функції F).

Далі на тестовій вибірці екзамнується побудована мережа або мережі, якщо їх кілька. Якщо мережа пройшла іспит, можна використати її для аналізу даних, побудувати прогноз або провести класифікацію.

Після натискання клавіші *Next*, *Майстер рішення задач* розіб'є вибірку на навчальну (чорні кольори значень), контрольну (сині кольори) та тестову вибірки (червоні кольори).

Також варто задати типи мереж, серед яких буде організовано пошук. Для даної моделі вибрати лінійну модель, що дозволяє зрівняти якість побудованих мереж і визначити параметри, що дають абсолютний мінімум помилок; задати багатoshаровий персептрон, за допомогою якого можливий більш точний результат моделі (рис. 11.4).

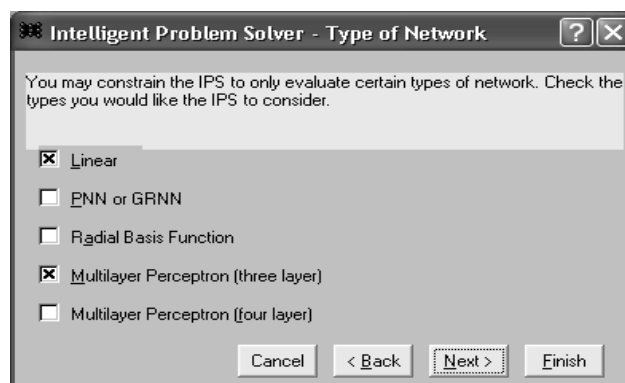


Рис. 11.4. Вибір типів мереж

У наступному діалоговому вікні необхідно задати кількість нейронів у прихованому шарі тришарового персептрону. Визначення числа прихованих шарів і числа нейронів у кожному шарі для конкретної задачі є неформальною проблемою, під час рішення якої можна використати евристичне правило: число нейронів у наступному шарі у два рази менше, ніж у попередньому (рис. 11.5).

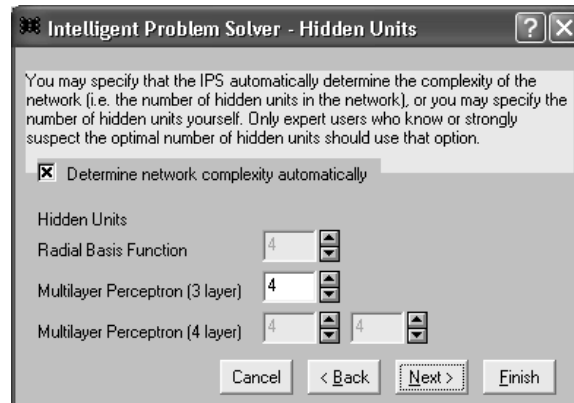


Рис. 11.5. Визначення кількості нейронів у прихованому шарі

У наступному діалоговому вікні вибрати тривалість пошуку із запропонованих: швидкий, середній, повний або обмежений у часі (рис. 11.6).

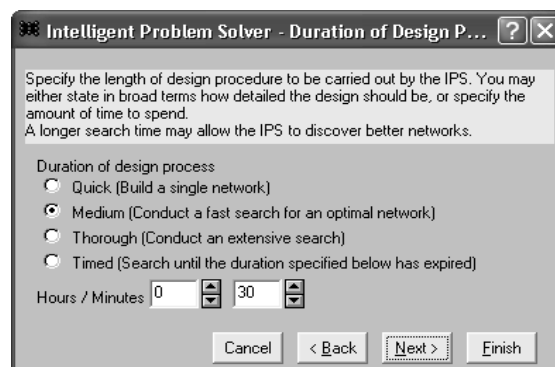


Рис. 11.6. Вибір тривалості пошуку

Далі на екрані з'явиться вікно *Saving Network – Збереження мереж*. У цьому вікні можна задати способи збереження мереж, наприклад, максимальне число мереж, зберегти мережі з найкращою якістю рішення і т. д. (рис. 11.7).

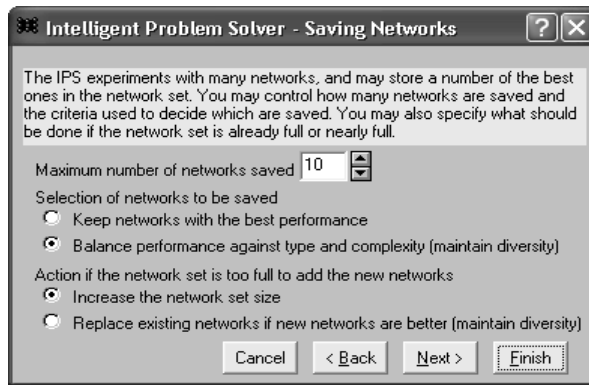


Рис. 11.7. Вибір способу збереження мереж

Потім відкриється вікно, у якому зазначені опції подання результатів: результати за кожною моделлю, загальна сумарна статистика й аналіз чутливості кращої нейронної мережі (рис. 11.8).

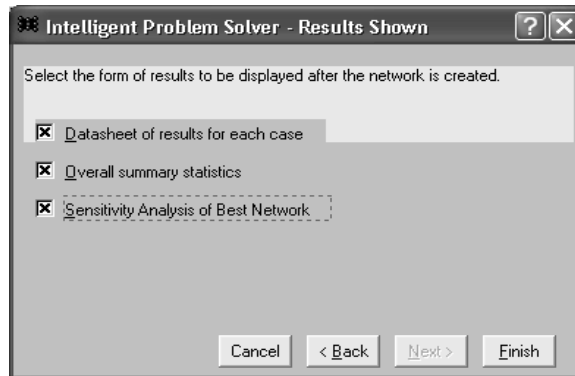


Рис. 11.8. Вибір опцій подання результатів

Ініціювавши клавішу *Finish* з'явиться побудована нейронна мережа, архітектура якої наведена на рис. 11.9.

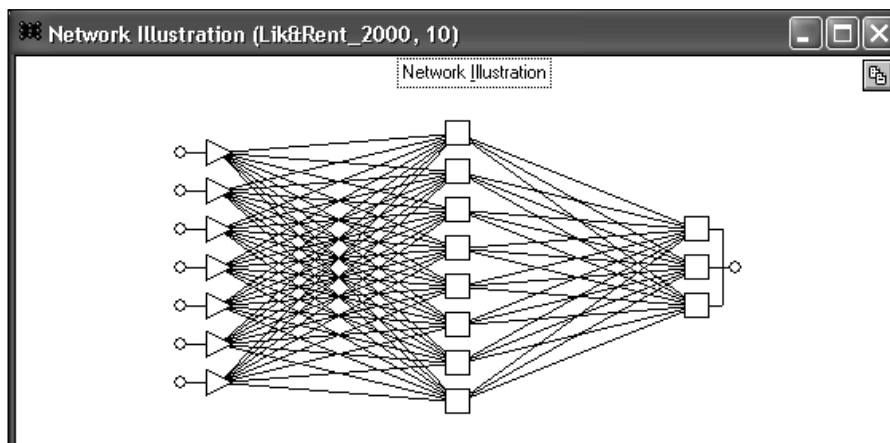


Рис. 11.9. Архітектура мережі

Рис. 11.9 відображає, що в багатошарових мережах встановлюються зв'язки тільки між нейронами сусідніх шарів. Кожен елемент може бути з'єднаний модифікованим зв'язком із будь-яким нейроном сусідніх шарів, але між елементами одного шару зв'язку немає. У ході розпізнавання образів вхідний вектор відповідає набору ознак, а вихідний – розпізнавальним образам. Схований шар призначений для відбиття специфіки знань.

Результати розрахунків під час побудови мережі подані у вигляді таблиці (рис. 11.10).

	Type	Error	Inputs	Hidden	Performance
01	Linear	0.3668725	1	-	0.8636364
02	Linear	0.3191385	2	-	0.9090909
03	Linear	0.2598321	3	-	1
04	Linear	0.2432871	5	-	0.9545455
05	Linear	0.2429893	4	-	0.9545455
06	MLP	0.2270141	1	1	0.9090909
07	MLP	0.07213	5	7	1
08	MLP	0.05428	7	8	1
09	MLP	0.03473	6	6	1
10*	MLP	0.03238	7	8	1

Рис. 11.10. Результати роботи радника

У даній таблиці наведено 10 найкращих мереж, знайдених радником. У стовпці *Type* – зазначений тип мереж: *Linear* – лінійний, *MLP* – багатошаровий перцептрон. Далі в таблиці результатів ідуть стовпці:

Error – помилка мережі, отримана на контрольній підмножині, що обчислюється для всіх контрольних спостережень. Чим менше значення помилки, тим краще якість мережі (у більшості побудованих мереж помилка менше нуля, що говорить про їх якість);

Inputs – число вхідних змінних, що використовуються нейронною мережею;

Hidden – число схованих елементів мережі (лінійні мережі не мають схованих елементів, тому для них у цьому стовпці зазначений пропуск);

Performance – якість мережі, що визначається за контрольною множиною. Для даної задачі класифікації якість – це частка правильно класифікованих спостережень.

Найкраща мережа відзначена (*) – (це мережа з номером 10).

Результати побудови моделі можна побачити у вікні *Data Set Editor* (рис. 11.11) у якому наведені вхідні змінні, що відбивають найбільш вагомий вплив на вихідну змінну *CLASS*. Засвітлені значення в таблиці говорять про те, що ці дані були виключені з моделі.

Variables	1	Cases	44	22	21							
	x1	x2	x3	x4	x5	x6	x7	x8	CLASS			
01	0	0.1483441	2.849778	0.01768	0.01984	-0.08689	0.06981	0.3961242	normal			
02	0.0007888	0.3263334	1.129027	-0.02617	-0.03744	-0.2205	-0.2144	-0.130537	crisis			
03	0.002979	0.07703	0.8263084	-0.079153	-0.1048	-0.4181	-0.4204	-0.4274	crisis			
04	0.1075425	0.6747426	3.391053	0.1439242	0.1654825	0.1569391	0.1648381	0.1433027	normal			
05	0.2366903	0.5790463	1.283792	0.07741	0.1188413	0.128639	0.128639	0.1002515	normal			
06	0.0006454	0.2415051	2.288941	0.01359	0.0166107	0.0694716	0.09518	0.06982	normal			
07	0	0.1395349	1.79327	0.01395	0.01884	0.2462532	0.2565891	0.4490956	normal			
08	0.0004936	0.1735234	2.635544	0.008042	0.009132	0.0907519	0.1136212	0.1136212	normal			
09	0.0273	1.652357	2.063027	0.00682	0.007102	-0.002937	0.07485	0.06556	normal			
10	0	0.01094	0.4384494	-0.01157	-0.01405	-0.6382	-0.2167	-0.2167	crisis			
11	0	1.244272	2.303523	0.04407	0.04978	-0.4115	0.1893256	0.2692308	normal			
12	0.001731	0.1216106	0.7131607	-0.09496	-0.14	-0.3103	-0.2799	-0.3163	crisis			
13	0.000214	0.04923	0.7815711	-0.04492	-0.05985	-0.3125	-0.2883	-0.271505	crisis			
14	0.0005618	0.4694298	1.909959	-0.03176	-0.03542	-0.07034	-0.05433	-0.1886	normal			
15	0.000267	0.1984326	1.246771	-0.04979	-0.05736	-0.4213	-0.3957	-0.4148	crisis			
16	1.624068	5.920128	13.70234	0.1346219	0.1415071	0.2045901	0.2045901	0.1854588	success			
17	0.007761	0.420154	1.160767	0.0429825	0.05379	0.1348979	0.1676509	0.08923	normal			
18	0	0.135638	0.5177443	-0.0433	-0.05593	-1.218332	-1.218332	-1.218332	crisis			
19	0	0.3662129	1.764831	-0.04579	-0.08026	-0.1934	-0.1888	-0.2629	crisis			
20	6.848e-05	0.4947868	1.354654	0.005578	0.006369	0.02266	0.09441	0.1126106	normal			
21	0.001185	0.055511	1.013611	0.02712	0.06648	0.4845	0.3377	1.305308	crisis			

Рис. 11.11. Результати побудови моделі

Для дослідження важливості вхідних змінних мережі проводиться аналіз чутливості. У вікні *Sensitivity Analysis* відображується чутливість мережі стосовно кожної змінної (рис. 11.12). У таблиці наведено три показники: Ранг, Помилка та Відношення. Показники чутливості розраховуються окремо для навчального (перші три рядки) і контрольного набору спостережень (стовпці таблиці – це змінні вихідного файлу даних).

Sensitivity Analysis									
Update		Baseline errors 0.01602		0.03238					
Prune inputs with low sensitivity ratio									
Threshold 1.05		Prune							
	x1	x3	x4	x5	x6	x7	x8		
Rank	1	5	7	6	3	4	2		
Error	0.2533835	0.02519	0.01548	0.01644	0.1001701	0.08222	0.1329085		
Ratio	15.81864	1.572581	0.9667057	1.026316	6.253584	5.132666	8.297433		
Rank	1	7	5	6	4	2	3		
Error	0.1766347	0.02631	0.03781	0.03403	0.03784	0.08974	0.06722		
Ratio	5.454451	0.8123519	1.167567	1.050787	1.168538	2.771221	2.075639		

Рис. 11.12. Діалогове вікно Аналізу чутливості

Для кожної змінної значення *Error* показує, якою буде якість мережі, якщо дану змінну виключити із числа вхідних змінних. Більш важливим для класифікації змінним відповідають більші значення помилок.

Відношення – *Ratio* – це відношення між значенням у рядку *Помилка* – *Error* та основною помилкою (*Baseline Error*). *Baseline Error* обчислюється для мережі з усіма вхідними змінними. Якщо *Відношення* – *Ratio*

менше одиниці, то виключення даної змінної поліпшує якість роботи мережі (значення *Відношення – Ratio* більше одиниці говорить про значний ступінь впливу змінних на підсумковий результат класифікації).

У рядку *Rank* змінні проранжовані у порядку убавання.

За ходом навчання можна стежити у вікні *Training Error Graph* (Графік помилки навчання), де на графіку відображується середньоквадратична помилка на навчальній множині на даному етапі (рис. 11.13).

Помилки мережі (під час і за результатами навчання) можна спостерігати у вікні *Case Errors*, де виводиться діаграма помилок для окремих спостережень. Визначивши опцію *Real-time Update* (Перерахувати по ходу), можна стежити за зміною помилок після кожного етапу навчання (рис. 11.13).

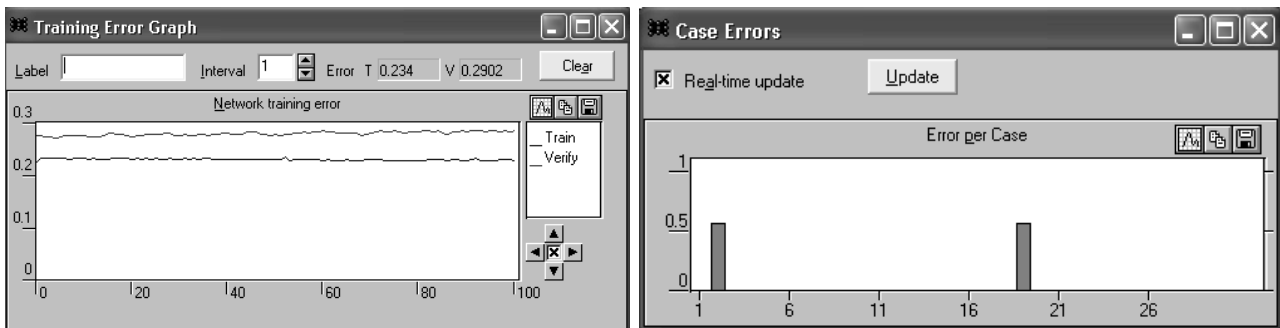


Рис. 11.13. Графік помилок навчання

У загальному вигляді результати побудови моделі відбиваються у вікні повідомлень (рис. 11.14), де описані основні характеристики нейронної моделі: зазначено тип мереж, найкраща мережа, оцінка якості класифікації (1,0000) і загальні помилки (0,032384). На підставі цих даних можна зробити висновок про якість побудованої мережі.

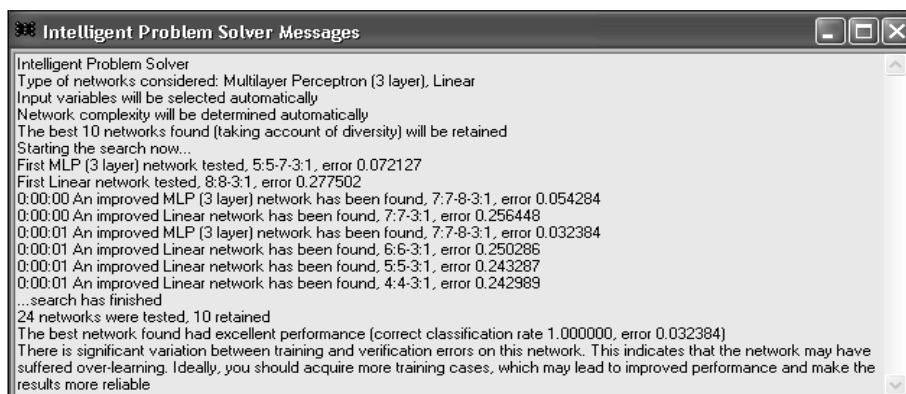


Рис. 11.14. Вікно повідомлень

Вікно *Classification Statistics (Статистика класифікації)* характеризує якість проведеної класифікації (рис. 11.15). Кожен стовпчик відповідає одному із трьох класів. Стовпчики розділені на три сектори: чорний – навчальна множина, червоний – контрольна та синій – тестова. У першому рядку таблиці виводиться загальна кількість підприємств за кожним класом, що були обрані із загальної вибірки. У другому рядку відбита кількість підприємств, що були правильно визначені в результаті процесу класифікації, а в третьому – кількість неправильно визначених підприємств. За даним прикладом було неправильно визначено два підприємства з тестової множини й одне з контрольної, які необхідно було віднести до кризових підприємств.

Classification Statistics																		
Variable: [] Run																		
	normal			crisis			success			normal			crisis			success		
Total	21	18	5	15	6	1	8	10	3	21	18	5	15	6	1	8	10	3
Correct	21	18	5	14	6	1	6	10	3	21	18	5	14	6	1	6	10	3
Wrong	0	0	0	1	0	0	2	0	0	0	0	0	1	0	0	2	0	0
Unknown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
normal	21	0	0	14	0	0	6	0	0	21	0	0	14	0	0	6	0	0
crisis	0	18	0	1	6	0	2	10	0	0	18	0	1	6	0	2	10	0
success	0	0	5	0	0	1	0	0	3	0	0	5	0	0	1	0	0	3

Рис. 11.15. Вікно статистики класифікації

У рядку *"Unknown"* наводиться кількість підприємств, які не були класифіковані через відсутність повноти інформації. Під рядком *"Unknown"* обкреслена чорна горизонтальна лінія, нижче якої подана зведена матриця. Вона показує, яка кількість спостережень початкової класифікації відповідає відтвореній класифікації за допомогою нейронної мережі. Ця матриця більш детально характеризує помилки, що виникли в результаті обчислень.

В вікні *Run Data Set* (рис. 11.16) наведено фрагмент класифікації за нейронною мережею. У дійсності нейронні мережі працюють винятково із цифровою інформацією та виконують прогнози змінної за допомогою вбудованих функцій, що супроводжують кожну нейронну мережу в *SNN*. Прогнози здійснюються на підставі обробки дійсного рівня уведеного нейрона та порівняння його з деякими пороговими значеннями.

RMS Error Train 0.01602 Verify 0.03238 Test 0.1315					
	CLASS	T. CLASS	E. CLASS	Error	
01	normal	normal	Right	0.01948	▲
02	crisis	crisis	Right	0.07768	▾
03	crisis	crisis	Right	3.267e-05	
04	normal	normal	Right	0.01839	
05	normal	normal	Right	0.01134	
06	normal	normal	Right	0.008784	
07	normal	normal	Right	0.02497	
08	normal	normal	Right	0.01089	
09	normal	normal	Right	0.007954	
10	crisis	crisis	Right	0.0001402	
11	normal	normal	Right	0.01798	
12	crisis	crisis	Right	0.0003124	
13	crisis	crisis	Right	0.0004185	
14	normal	normal	Right	0.02282	
15	crisis	crisis	Right	3.922e-05	
16	success	success	Right	0.0170152	
17	normal	normal	Right	0.009081	
18	crisis	crisis	Right	7.503e-06	
19	crisis	crisis	Right	0.02641	
20	normal	normal	Right	0.007984	
21	crisis	crisis	Right	1.333e-05	
22	normal	normal	Right	0.0205867	
23	normal	normal	Right	0.006378	
24	normal	normal	Right	0.0309195	
25	success	success	Right	0.006164	

Рис. 11.16. Результати класифікації

У першому стовпчику наведено класи підприємств, визначені в результаті роботи нейронної мережі, в другому – вихідні класи підприємств. Третій стовпчик показує помилку класифікації.

Після оцінки якості моделі можна визначити фінансовий стан досліджуваного підприємства (рис. 11.17).

Run Single Case								
Case No	25		Error	0.006164				Run
	X1	X3	X4	X5	X6	X7	X8	
Input	1.79227	6.723995	0.09164	0.09987	0.3405201	0.3423167	0.4862469	
Outputs Shown Variables								
	CLASS							
Output	success							
Target	success							
Error	Right							

Рис. 11.17. Визначення фінансового стану підприємства

Таким чином, можна зробити висновок, що наведене підприємство належить до класу успішних підприємств.

Щоб спрогнозувати фінансовий стан підприємства в наступні періоди, необхідно скористатися опцією *Run One-off Case*. У даному вікні знаходяться дві таблиці. В першу вводимо прогнозовані значення коефіцієнтів, а в другій таблиці після натискання кнопки *Run* з'явиться результат, тобто визначиться прогнозований клас підприємства (рис. 11.18).

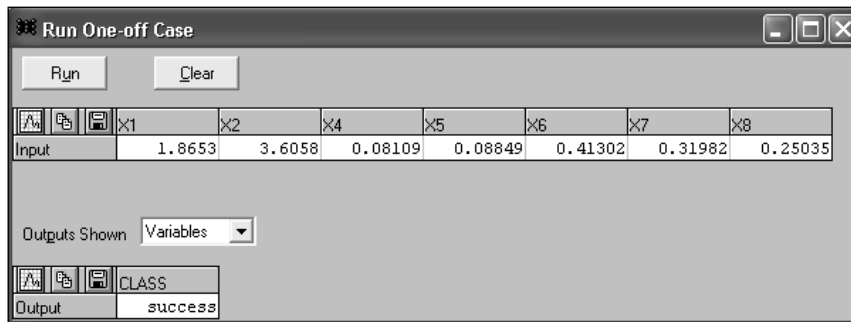


Рис. 11.18. Прогноз фінансового стану підприємства

Таким чином, для прогнозованих значень фінансових показників досліджуване підприємство буде належати до класу підприємств з ефективною фінансовою діяльністю.

Рекомендована література

Основна

Боровиков В. П. Прогнозирование в системе STATISTICA в среде Windows / В. П. Боровиков, Г. И. Ивченко. – М. : Финансы и статистика, 1997. – 268 с.

Иванов В. В. Анализ временных рядов и прогнозирование экономических показателей / В. В. Иванов. – Х. : ХНУ, 1999. – 230 с.

Клебанова Т. С. Методы прогнозирования : учеб. пособ. / Т. С. Клебанова, В. В. Иванов, Н. А. Дубровина. – Х. : Изд. ХГЭУ, 2002. – 372 с.

Методы и модели прогнозирования социально-экономических процессов учебное пособие / Т. С. Клебанова, В. А. Курзенов, В. Н. Наумов и др. – СПб. : Изд. СЗИУ РАНХ и ГС, 2012. – 566 с.

Моделі і методи соціально-економічного прогнозування : підручник / В. М. Геєць, Т. С. Клебанова, О. І. Черняк та ін. – Х. : ВД "ІНЖЕК", 2005. – 396 с.

Эконометрия на персональном компьютере / Клебанова Т. С., Дуброва Н. А., Милов А. В. и др. – Х. : Изд. ХГЭУ, 2002. – 208 с.

Додаткова

Басовский Л. Е. Прогнозирование и планирование в условиях рынка / Л. Е. Басовский. – М. : ИНФРА-М, 2001. – 260 с.

Бокс Дж. Анализ временных рядов / Дж. Бокс, Г. Дженкинс // Прогноз и управление. – Вып.1. – М. : Мир, 1974.

Боровиков В. П. Популярное введение в программу STATISTICA / В. П. Боровиков. – М. : Компьютер Пресс, 1998. – 194 с.

Боровиков В. П. STATISTICA : искусство анализа данных на компьютере. Для профессионалов / В. П. Боровиков. – СПб. : Питер, 2001. – 656 с.

Грабовецький Б. Є. Економічне прогнозування і планування : навч. посіб. / Б. Є. Грабовецький. – К. : Центр навчальної літератури, 2003. – 188 с.

Дуброва Т. А. Статистические методы прогнозирования : учеб. пособ. для вузов / Т. А. Дуброва. – М. : ЮНИТИ – ДАНА, 2003. – 206 с.

Ермаков С. М. Курс статистического моделирования / С. М. Ермаков, Г. А. Михайлов. – М. : Наука, 1976. – 320 с.

Клебанова Т. С. Нечітка логіка та нейронні мережі в управлінні підприємством / Т. С. Клебанова, Л. О. Чаговец, О.В. Панасенко. – Х. : ВД "ІНЖЕК", 2011. – 240 с.

Кэнделл М. Временные ряды / М. Кэнделл. – М. : Финансы и статистика, 1981. – 580 с.

Лук'яненко І. Сучасні економетричні методи у фінансах : навч. посіб. / І. Лук'яненко, Ю. Городніченко. – К. : Літера ЛТД, 2002. – 352 с.

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования / Ю. П. Лукашин. – М. : Статистика, 1979. – 420 с.

Магнус Я. Р. Эконометрика. Начальный курс : учебник. / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – 5-е изд. – М. : Дело, 2001. – 400 с.

Морозова Т. Г. Прогнозирование и планирование в условиях рынка : учеб. пособ. для вузов / Т. Г. Морозова, А. В. Пикулькина. – М. : ЮНИТИ-ДАНА, 1999. – 318 с.

Орлов. А. Н. Эконометрика : учеб. пособ. для вузов / А. Н. Орлов. – М. : Изд. "Экзамен", 2002. – 576 с.

Пашута М. Т. Прогнозування та макроекономічне планування : навч. посіб. / М. Т. Пашута. – К. : МАУП, 1998. – 192 с.

Уотшем Т. Количественные методы в финансах / Т. Уотшем, К. Паррамоу. – М. : Финансы, ЮНИТИ, 1999. – 528 с.

Цыгичко В. Н. Руководителю о принятии решений / В. Н. Цыгичко. – М. : Инфра-М, 1996. – 272 с.

Черняк О. І. Динамічна економетрика / О. І. Черняк, А. В. Ставицький. – К. : КВІЦ, 2000. – 120 с.

Чураков Е. П. Математические методы обработки экспериментальных данных в экономике : учеб. пособ. / Е. П. Чураков. – М. : Финансы и статистика, 2004. – 240 с.

Інформаційні ресурси

Сайт Государственного комитета статистики Украины. – Режим доступа : www.ukrstat.gov.ua.

Сайт Национального банка Украины. – Режим доступа : www.bank.gov.ua.

Сайт ПФТС. – Режим доступа: <http://pfts.com>.

Зміст

Вступ.....	3
Змістовий модуль 1. Методи та моделі короткострокового прогнозування..	3
Лабораторна робота 1. Критерії відповідності для перевірки гіпотез про нормальний закон розподілу.....	3
Лабораторна робота 2. Оцінювання якості прогнозів.....	14
Лабораторна робота 3. Перевірка стаціонарності часових рядів	19
Лабораторна робота 4. Рішення та аналіз моделей простої екстраполяції й інтерполяції одномірних процесів.....	30
Лабораторна робота 5. Методи розв'язання адаптивних моделей прогнозування.....	37
Змістовий модуль 2. Методи та моделі середньострокового, довгострокового прогнозування.....	48
Лабораторна робота 6. Реалізація моделей виділення та вирівнювання трендів.....	48
Лабораторна робота 7. Моделі прогнозування багатомірних процесів...	57
Лабораторна робота 8. Моделювання періодичних складових часового ряду.....	74
Лабораторна робота 9. Реалізація моделей часових рядів за допомогою моделей авторегресії і ковзного середнього.....	100
Лабораторна робота 10. Реалізація експертних методів прогнозування, розрахунки непараметричних статистик, аналіз відповідностей.....	109
Лабораторна робота 11. Прогнозування на підставі нейронних мереж...	116
Рекомендована література.....	126
Основна.....	126
Додаткова.....	127
Інформаційні ресурси.....	128

НАВЧАЛЬНЕ ВИДАННЯ

**Методичні рекомендації
до виконання лабораторних робіт
з навчальної дисципліни
"ПРОГНОЗУВАННЯ
СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ"**

**для студентів напряму підготовки
6.030502 "Економічна кібернетика"
денної форми навчання**

Самостійне електронне текстове мережне видання

Укладачі: **Гур'янова** Лідія Семенівна
Сергієнко Олена Андріанівна

Відповідальний за випуск *Т. С. Клебанова*

Редактор *М. А. Ковальчук*

Коректор *М. А. Ковальчук*

План 2015 р. Поз. № 105 ЕВ. Обсяг 130 с.

Видавець і виготівник – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Леніна, 9-А

*Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.*