

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

**СТАТИСТИЧНЕ МОДЕЛЮВАННЯ
ТА ПРОГНОЗУВАННЯ.
ЗМІСТОВІ МОДУЛІ 3 ТА 4**

**Методичні рекомендації
до лабораторних
та самостійних робіт студентів
спеціальності 051 "Економіка"
першого (бакалаврського) рівня**

**Харків
ХНЕУ ім. С. Кузнеця
2022**

УДК 519.86(07.034)

C78

Укладачі: О. В. Раєвнєва

О. І. Бровко

В. І. Дериховська

Затверджено на засіданні кафедри статистики і економічного прогнозування.

Протокол № 5 від 17.12.2021 р.

Самостійне електронне текстове мережеве видання

Статистичне моделювання та прогнозування. Змістові мо-
C78 дулі 3 та 4 [Електронний ресурс] : методичні рекомендації до ла-
бораторних та самостійних робіт студентів спеціальності 051 "Еко-
номіка" першого (бакалаврського) рівня / уклад. О. В. Раєвнєва,
О. І. Бровко, В. І. Дериховська. – Харків : ХНЕУ ім. С. Кузнеця,
2022. – 105 с.

Подано методичні рекомендації до виконання лабораторних і само-
стійних робіт, метою яких є розширення та поглиблення знань, а також набут-
тя професійних компетентностей щодо прогнозування соціально-економічних
процесів і моделювання складних систем за допомогою статистичних методів
та моделей засобами спеціалізованих аналітичних інструментів.

Рекомендовано для студентів спеціальності 051 "Економіка" першого
(бакалаврського) рівня.

УДК 519.86(07.034)

© Харківський національний економічний
університет імені Семена Кузнеця, 2022

Вступ

Моделювання – це метод дослідження різних явищ і процесів, аналізу та вироблення множини управлінських рішень. Методом моделювання описуються структури об'єкта (статична модель), процес його функціонування і розвитку (динамічна модель). В аналітичній роботі використання моделей дозволяє відтворювати властивості, встановлювати зв'язки, тенденції досліджуваних систем і процесів, що, своєю чергою, дає змогу оцінити їхній стан, побудувати прогноз, прийняти обґрунтоване та ефективне рішення.

Наведені методичні рекомендації призначено для засвоєння студентами теоретичних і практичних знань із дисципліни "Статистичне моделювання та прогнозування", набуття навичок роботи з пакетами прикладних програм Statistica, Eviews і MS Excel, які дозволяють реалізувати спектр економіко-математичних методів і моделей аналізу і прогнозування розвитку соціально-економічних процесів і явищ.

Навчальна дисципліна "Статистичне моделювання та прогнозування" є вибірковою навчальною дисципліною спеціальності 051 "Економіка" спеціалізації "Бізнес-статистика і аналітика" освітнього ступеня "бакалавр".

Метою викладання цієї навчальної дисципліни є розширення та поглиблення теоретичних знань і набуття професійних компетентностей щодо прогнозування соціально-економічних процесів та моделювання складних систем за допомогою статистичних методів і моделей. Ця навчальна дисципліна вивчає інструментарій економіко-математичного моделювання в дослідженні складних соціально-економічних систем.

Вивчення початкової дисципліни "Статистичне моделювання та прогнозування" вимагає, поряд із засвоєнням теоретичних положень, практичної реалізації основних розділів навчальної дисципліни шляхом виконання відповідних лабораторних і самостійних робіт.

У роботі надано авторський переклад українською мовою панелей інструментів (вікон, команд та функцій) усіх програмних продуктів і застосунків, що використовуються.

Змістовий модуль 3

Моделювання часових процесів

Тема 9. Моделювання і прогнозування тенденцій розвитку

Лабораторна робота 5. Побудова моделі декомпозиції часових рядів

Мета: надбання навичок декомпозиції часового ряду в пакеті Statistica 10.0.

Завдання: ознайомитися з модулем Advanced Linear; навчитися визначати вид моделі декомпозиції; виявити всі складові моделі декомпозиції; спрогнозувати трендову складову та здійснити спектральний аналіз циклічної складової; здійснити композицію моделі та перевірити її якість.

Методичні рекомендації

Необхідно сформувати динамічний ряд і подати його у вигляді файла в ППП Statistica 10.0. Припустимо, що нам відомі дані щодо динаміки обсягу ВВП України в кварталному розрізі (в млн грн), які наведено у табл. 1 та візуалізовано на рис. 1.

Таблиця 1

Вихідні дані

Період	ВВП	Період	ВВП	Період	ВВП
2012	217 074	2015	303 753	2018	455 298
	255 545		354 814		535 701
	300 446		398 000		671 456
	306 281		408 631		722 912
2013	258 591	2016	316 905	2019	592 523
	310 277		382 391		665 233
	368 488		440 476		834 287
	362 635		447 143		891 839
2014	292 324	2017	375 991	2020	705 013
	346 005		456 715		810 820
	387 109		566 997		994 850
	379 231		588 841		1 048 023

Data: Spreadsheet2* (2v by 36c)		
	1 t	2 ВВП
1	1	217 074
2	2	255 545
3	3	300 446
4	4	306 281
5	5	258 591
6	6	310 277
7	7	368 488
8	8	362 635
9	9	292 324
10	10	346 005
11	11	387 109
12	12	379 231
13	13	303 753
14	14	354 814
15	15	398 000
16	16	408 631
17	17	316 905
18	18	382 391
19	19	440 476
20	20	447 143
21	21	375991
22	22	456715
23	23	566997
24	24	588841
25	25	455298
26	26	535701
27	27	671456
28	28	722912
29	29	592 523
30	30	665 233
31	31	834 287
32	32	891 839
33	33	705013
34	34	810820
35	35	994850
36	36	1048023

Рис. 1. Вихідні дані

Модель декомпозиції є способом поєднання складових часового ряду: трендової, сезонної, циклічної та випадкової. Існує два способи поєднання цих складових, що дає змогу визначити дві основні моделі декомпозиції:

- 1) адитивна модель – $Y = T + S + C + I$;
- 2) мультиплікативна модель – $Y = T \times S \times C \times I$.

Визначення двох основних моделей декомпозиції часового ряду здійснюється за допомогою графічного аналізу. Існує таке правило визначення типу моделі: якщо початкові дані мають амплітуду коливань значень, яка постійно збільшується чи зменшується, то доцільно використовувати мультиплікативну модель декомпозиції; у випадку якщо амплітуда зміни значень постійна, необхідно використовувати адитивну модель.

Для визначення типу моделі декомпозиції зобразимо вихідний часовий ряд графічно (рис. 2).

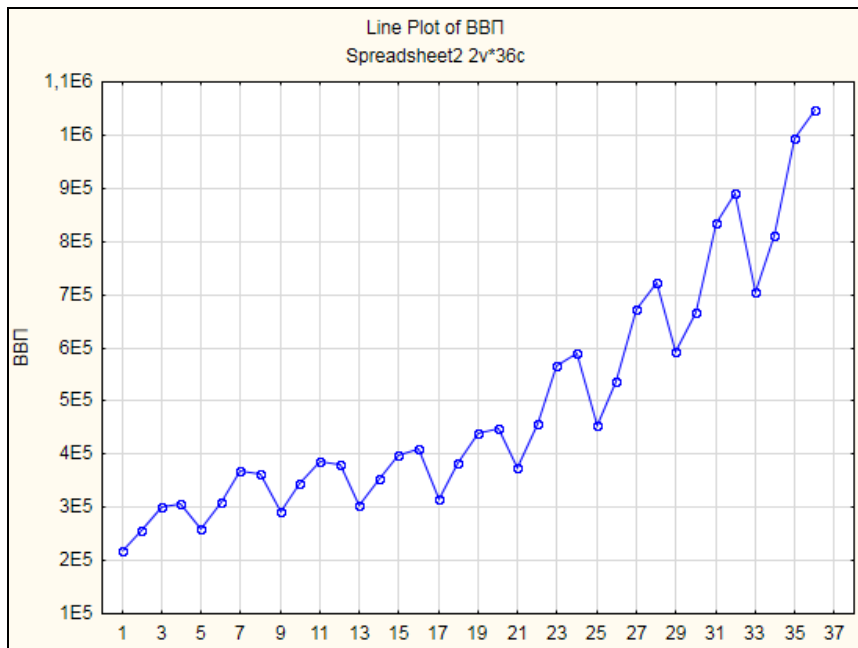


Рис. 2. Графік вихідних даних

Аналіз рис. 2 дозволяє стверджувати про: наявний позитивно спрямований тренд, певну сезонність; мультиплікативний характер взаємодії компонентів, отже раціонально використовувати мультиплікативну модель часового ряду, яку в загальному вигляді наведено далі:

$$Y = T \times S \times C \times I, \quad (1)$$

де T – тренд;

S – сезонна складова;

C – циклічна складова;

I – випадкова складова.

Побудова моделі декомпозиції здійснюється за таким алгоритмом:

1 *етап*. Знаходження згладженої трендциклічної та сезонної складових.

2 *етап*. Побудова регресійної моделі, в якій незалежною змінною є час t (t).

3 *етап*. Знаходження циклічної складової та представлення її у вигляді ряду Фур'є ($C(t)$).

4 *етап*. Знаходження випадкової складової та закону її розподілу.

Згідно з алгоритмом побудови моделі декомпозиції, виконаємо відповідні кроки у ППП Statistica 10.0. Для цього необхідно вибрати вкладку *Seasonal Decomposition (Сезонна декомпозиція)* у стартовій панелі

модуля *Advanced Linear/Nonlinear Models/Time Series/Forecasting* (Поглиблені лінійні / Нелінійні моделі / Часові ряди / Прогнозування), задати змінну для аналізу (рис. 3).

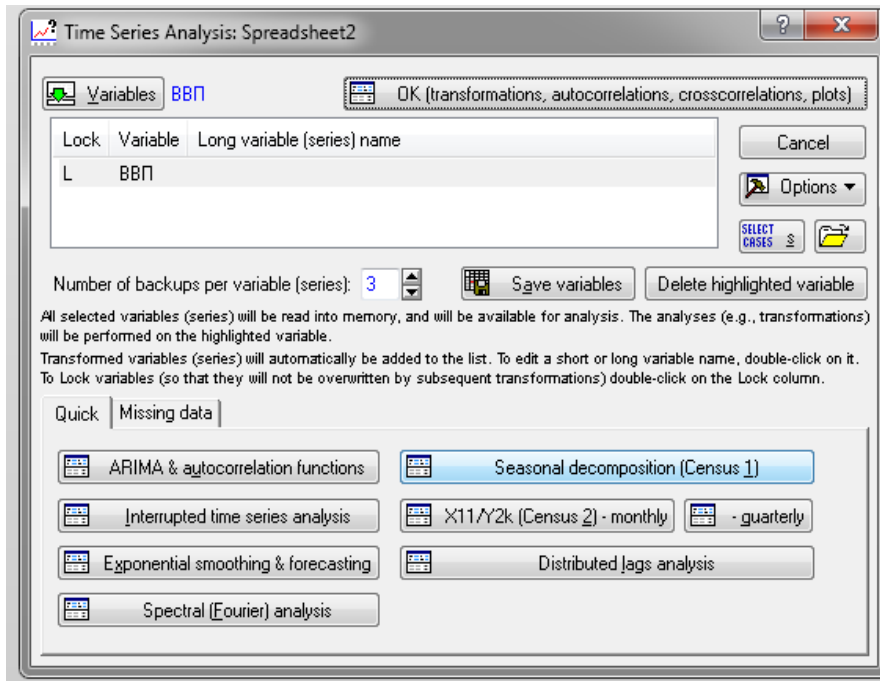


Рис. 3. Запуск модуля *Time Series Analysis*

Потім необхідно провести автокореляційний аналіз, що дозволить виявити наявність сезонності та значення сезонних лагів. Для цього переходимо у вкладку *Autocorrelations* (Автокореляції) та натискаємо відповідну кнопку *Autocorrelations* (рис. 4).

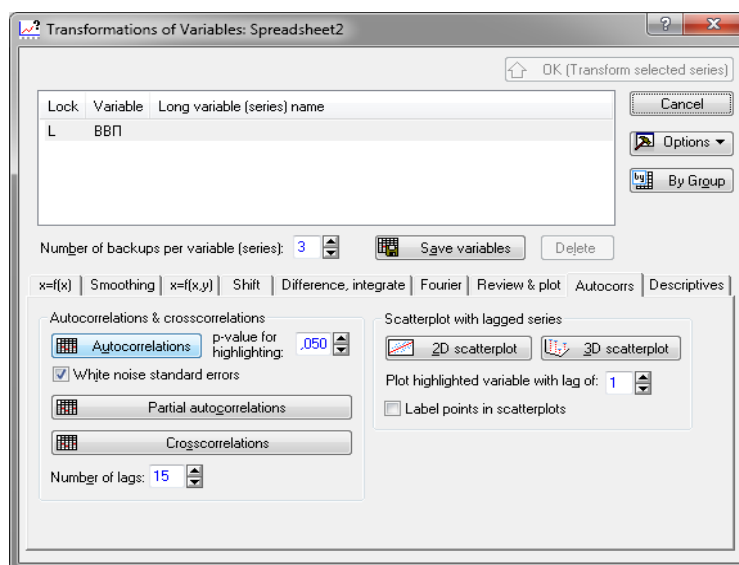


Рис. 4. Вікно запуску автокореляційного аналізу

Відповідно до отриманих результатів (рис. 5) найбільше значення автокореляційного коефіцієнту отримали для першого лагу, потім він поступово знижується та з'являється наступний найбільший локальний максимум для значення у четвертому лазі. Отже, у вихідному ряді присутні тренд та сезонність, а сезонний лаг дорівнює 4.

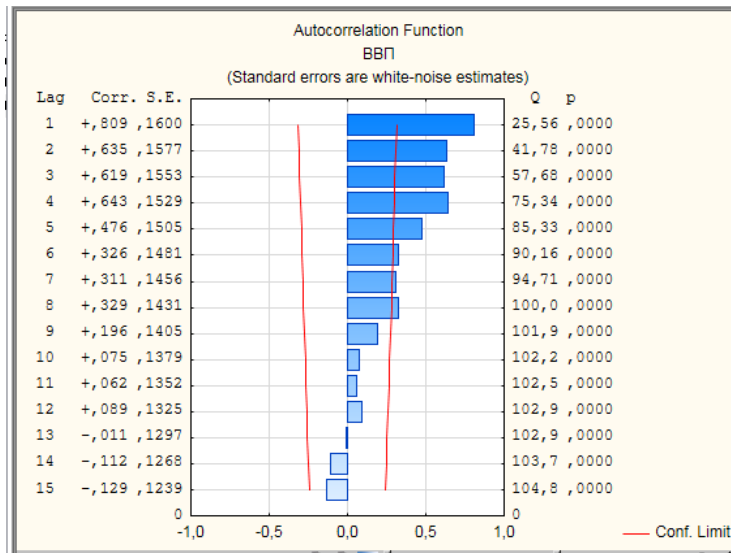


Рис. 5. Результати автокореляційного аналізу

Також для підтвердження наявності сезонності використаємо спектральний аналіз (аналіз Фур'є). При цьому обираємо одномірний аналіз і будуємо графік спектральної щільності (рис. 6 і 7).

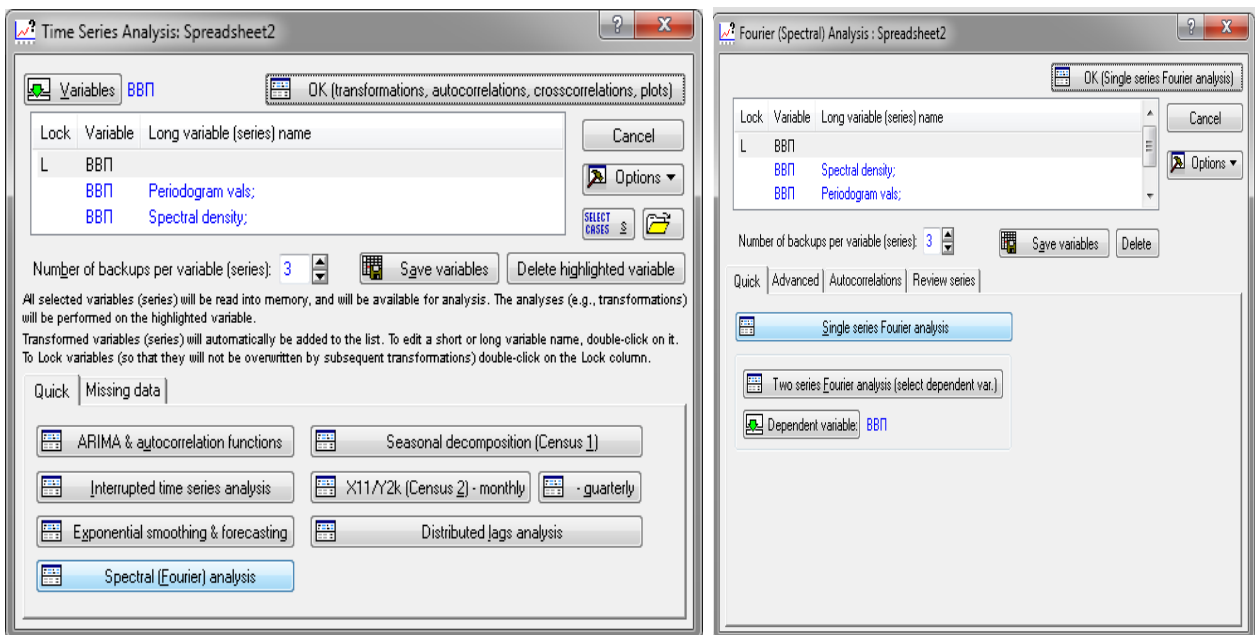


Рис. 6. Етапи виконання спектрального аналізу

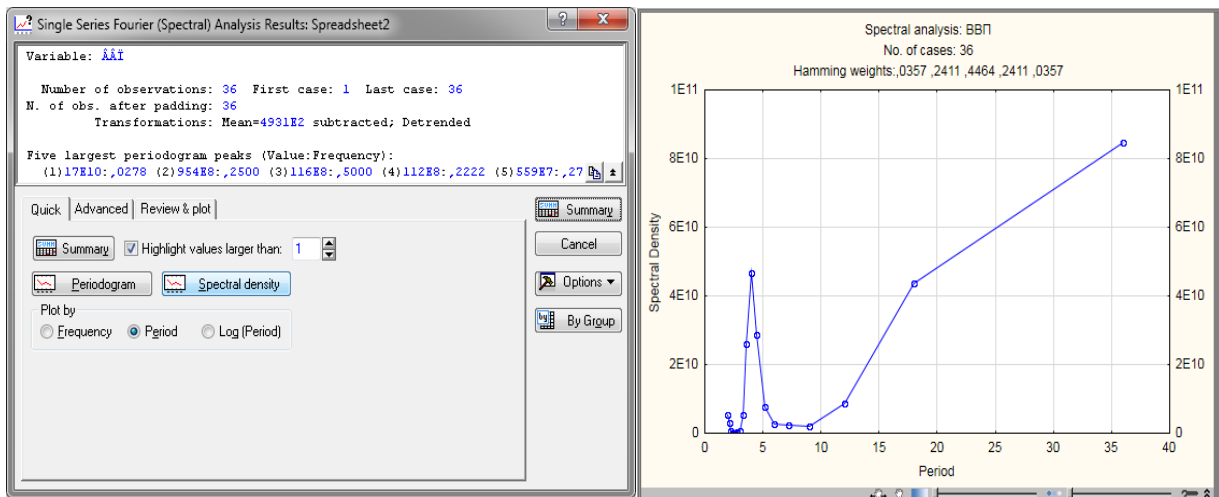


Рис. 7. Результати спектрального аналізу

Перший максимум досягається у точці (періоді) 4, тобто сезонність присутня і з лагом 4 вона впливає на змінну обсягу ВВП.

Тепер повертаємося до вихідного вікна сезонної декомпозиції та задаємо відповідні параметри прогнозування (рис. 8).

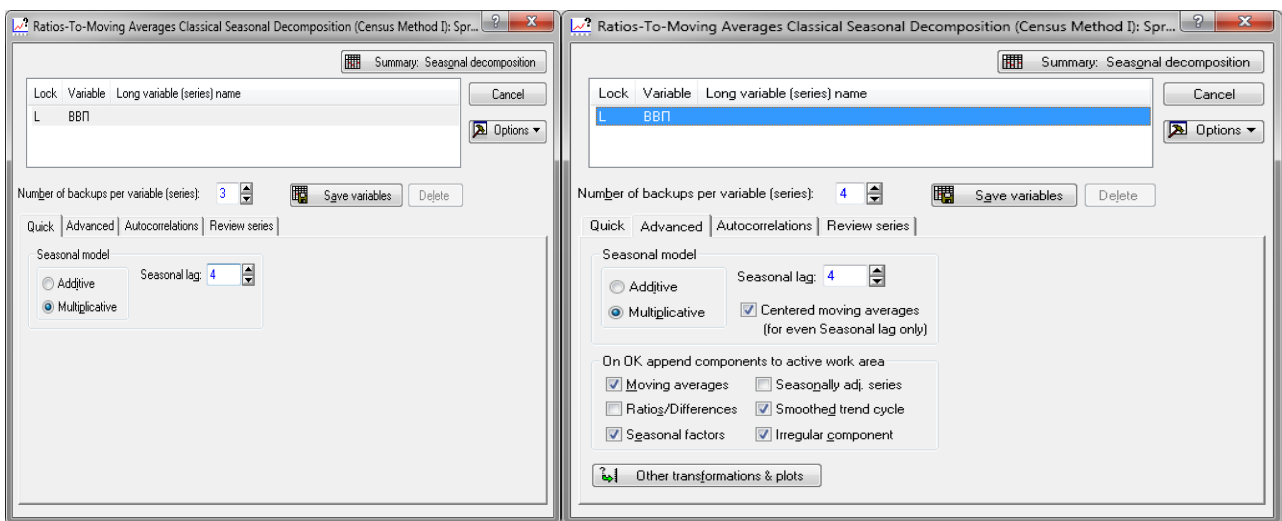


Рис. 8. Параметри моделі сезонної декомпозиції

Кількість копій на змінну (ряд) (Number of backups per variable (series)) користувач може задати самостійно. Зазвичай 3 – 4 копій достатньо.

Результати наведено на рис. 9.

Seasonal Decomposition: Multipl. season (4); Centered means (Spreadsheet2) ВВП							
Case	ВВП	Moving Averages	Ratios	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irreg. Compon.
1	217074			84,7807	256041,8	260310,4	0,983602
2	255545			95,8736	266543,6	265151,3	1,005251
3	300446	275026,1	109,2427	110,1065	272868,6	274833,3	0,992851
4	306281	287057,3	106,6968	109,2392	280376,4	287451,6	0,975386
5	258591	302404,0	85,5118	84,7807	305011,7	303398,2	1,005318
6	310277	317953,5	97,5857	95,8736	323631,3	318065,3	1,017499
7	368488	329214,4	111,9295	110,1065	334665,1	329444,2	1,015848
8	362635	337897,0	107,3212	109,2392	331964,1	337705,7	0,982998
9	292324	344690,6	84,8076	84,7807	344800,2	345151,7	0,998982
10	346005	349092,8	99,1155	95,8736	360897,0	350507,3	1,029642
11	387109	352595,9	109,7883	110,1065	351576,9	352657,7	0,996936
12	379231	355125,6	106,7878	109,2392	347156,5	354685,2	0,978773
13	303753	357588,1	84,9449	84,7807	358280,9	358041,2	1,000669
14	354814	362624,5	97,8461	95,8736	370085,2	363442,2	1,018278
15	398000	367943,5	108,1688	110,1065	361468,3	367198,8	0,984394
16	408631	373034,6	109,5424	109,2392	374069,9	373518,7	1,001476
17	316905	381791,3	83,0048	84,7807	373793,8	380970,3	0,981163
18	382391	391914,8	97,5699	95,8736	398849,1	391957,8	1,017582
19	440476	404114,5	108,9978	110,1065	400045,5	403751,6	0,990821
20	447143	420790,8	106,2626	109,2392	409324,6	421139,9	0,971945
21	375991	445896,4	84,3225	84,7807	443486,6	446316,9	0,993659
22	456715	479423,8	95,2633	95,8736	476372,0	477150,9	0,998367
23	566997	507049,4	111,8228	110,1065	514953,3	506244,1	1,017204
24	588841	526836,0	111,7693	109,2392	539038,1	528467,9	1,020002
25	455298	549766,6	82,8166	84,7807	537030,3	547940,0	0,980090
26	535701	579582,9	92,4287	95,8736	558757,5	574532,2	0,972544
27	671456	613494,9	109,4477	110,1065	609824,2	611827,4	0,996726
28	722912	646839,5	111,7606	109,2392	661769,6	650595,3	1,017176
29	592523	683384,9	86,7041	84,7807	698889,1	686163,2	1,018546
30	665233	724854,6	91,7747	95,8736	693864,6	719218,8	0,964748
31	834287	760031,8	109,7700	110,1065	757709,3	758237,4	0,999303
32	891839	792291,4	112,5645	109,2392	816409,2	796374,8	1,025157
33	705013	830560,1	84,8840	84,7807	831572,6	831135,0	1,000526
34	810820	870153,5	93,1813	95,8736	845717,6	864795,5	0,977939
35	994850			110,1065	903534,5	902878,5	1,000727
36	1048023			109,2392	959383,5	921920,0	1,040636

Рис. 9. Результат побудови сезонної декомпозиції

Примітка:

Moving Averages – ковзні середні;

Ratios – коефіцієнти (без врахування сезонної складової);

Seasonal Factors – сезонні фактори;

Adjusted Series – скориговані значення;

Smoothed Trend-c. – згладжена трендово-циклічна складова;

Irreg. Compon. – випадкова компонента.

Для наочності сприйняття виділених складових необхідно їх візуалізувати. З цією метою обираємо кожну окрему складову та виконуємо команду *Graphs of Block data / Line Plot Entire Columns* (Графіки блочних даних / Лінійний графік / Всі стовбці) (рис. 10).

Case	BBП	Moving Averages	Ratios	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irr. Compon.
4	306281	287057,3	106,6968	109,73071	287057,3	287441,6	0,975386
5	258591	302404,0	85,5118	84	258591	258591	1,005318
6	310277	317953,5	97,5857	98	310277	310277	0,998367
7	368488	329214,4	111,9295	111	368488	368488	1,017204
8	362635	337897,0	107,3212	108	362635	362635	1,020002
9	292324	344690,6	84,8076	84	292324	292324	0,980090
10	346005	349092,8	99,1155	99	346005	346005	0,993659
11	387109	352595,9	109,7883	110	387109	387109	0,998367
12	379231	355125,6	106,7878	106	379231	379231	1,017204
13	303753	357588,1	84,9449	84	303753	303753	1,020002
14	354814	362624,5	97,8461	97	354814	354814	0,975386
15	398000	367943,5	108,1688	110	398000	398000	0,998367
16	408631	373034,6	109,5424	109	408631	408631	1,017204
17	316905	381791,3	83,0048	84	316905	316905	1,020002
18	382391	391914,8	97,5699	97	382391	382391	0,975386
19	440476	404114,5	108,9978	110	440476	440476	0,998367
20	447143	420790,8	106,2626	106	447143	447143	1,017204
21	375991	445896,4	84,3225	84,7807	443486,6	446316,9	0,993659
22	456715	479423,8	95,2633	95,8736	476372,0	477150,9	0,998367
23	566997	507049,4	111,8228	110,1065	514953,3	506244,1	1,017204
24	588841	526836,0	111,7693	109,2392	539038,1	528467,9	1,020002
25	455298	549766,6	82,8166	84,7807	537030,3	547940,0	0,980090

Рис. 10. Побудова графіків

Унаслідок отримаємо візуалізацію складових моделі декомпозиції (рис. 11).

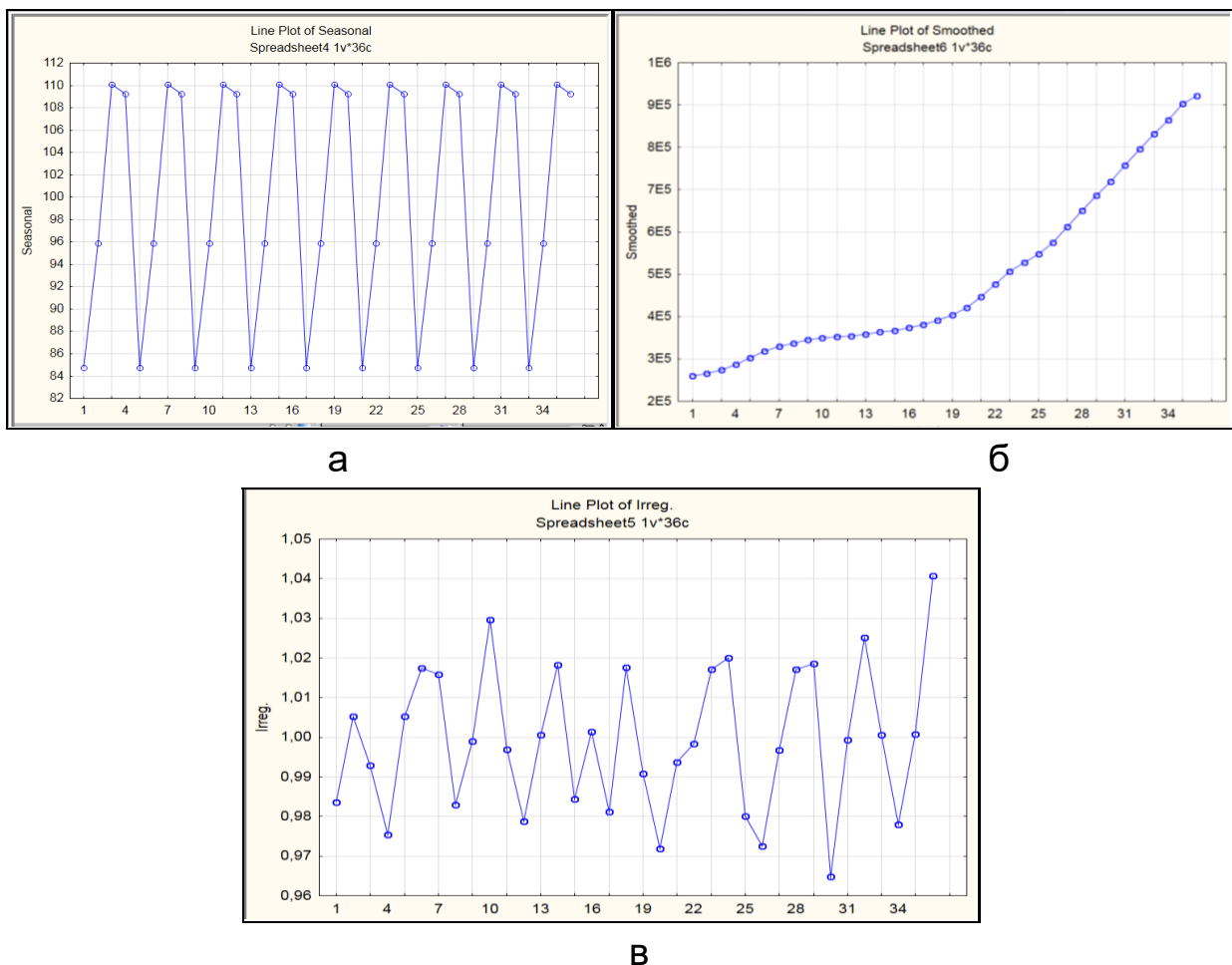


Рис. 11. Візуалізація складових моделі декомпозиції:
 а – графік сезонної складової; б – графік трендово-циклічної складової; в – графік випадкової складової

Графічний аналіз підтверджує наявність відповідних складових у моделі декомпозиції.

На наступному етапі виконання лабораторної роботи необхідно скопіювати результати декомпозиції, а саме трендово-циклічну (STC), сезонну (S) і випадкову (I) складові в вікно з вихідними даними (рис. 12).

	1 t	2 ВВП	3 STC	4 S	5 I
1	1	217 074	260310,4	84,7807	0,983602
2	2	255 545	265151,3	95,8736	1,005251
3	3	300 446	274833,3	110,1065	0,992851
4	4	306 281	287451,6	109,2392	0,975386
5	5	258 591	303398,2	84,7807	1,005318
6	6	310 277	318065,3	95,8736	1,017499
7	7	368 488	329444,2	110,1065	1,015848
8	8	362 635	337705,7	109,2392	0,982998
9	9	292 324	345151,7	84,7807	0,998982
10	10	346 005	350507,3	95,8736	1,029642
11	11	387 109	352657,7	110,1065	0,996936
12	12	379 231	354685,2	109,2392	0,978773
13	13	303 753	358041,2	84,7807	1,000669
14	14	354 814	363442,2	95,8736	1,018278
15	15	398 000	367198,8	110,1065	0,984394
16	16	408 631	373518,7	109,2392	1,001476
17	17	316 905	380970,3	84,7807	0,981163
18	18	382 391	391957,8	95,8736	1,017582
19	19	440 476	403751,6	110,1065	0,990821
20	20	447 143	421139,9	109,2392	0,971945
21	21	375991	446316,9	84,7807	0,993659
22	22	456715	477150,9	95,8736	0,998367
23	23	566997	506244,1	110,1065	1,017204
24	24	588841	528467,9	109,2392	1,020002
25	25	455298	547940,0	84,7807	0,980090
26	26	535701	574532,2	95,8736	0,972544
27	27	671456	611827,4	110,1065	0,996726
28	28	722912	650595,3	109,2392	1,017176
29	29	592 523	686163,2	84,7807	1,018546
30	30	665 233	719218,8	95,8736	0,964748
31	31	834 287	758237,4	110,1065	0,999303
32	32	891 839	796374,8	109,2392	1,025157
33	33	705013	831135,0	84,7807	1,000526
34	34	810820	864795,5	95,8736	0,977939
35	35	994850	902878,5	110,1065	1,000727
36	36	1048023	921920,0	109,2392	1,040636

Рис. 12. Результати декомпозиції

На другому етапі необхідно виділити тренд з трендово-циклічної складової. Для цього потрібно побудувати парну лінійну регресійну модель ($Y = a_0 + a_1 \times x$), у якій залежною змінною є трендово-циклічна складова, а незалежною змінною є час (t):

$$STC = a_0 + a_1 \times t, \quad (2)$$

де STC – трендово-циклічна складова;

a_0 , a_1 – параметри моделі;

t – час (незалежна змінна).

Для цього використовуємо модуль *Statistica (Статистика)* та *Multiple regression (Множинна регресія)* і задаємо параметри, наведені на рис. 13.

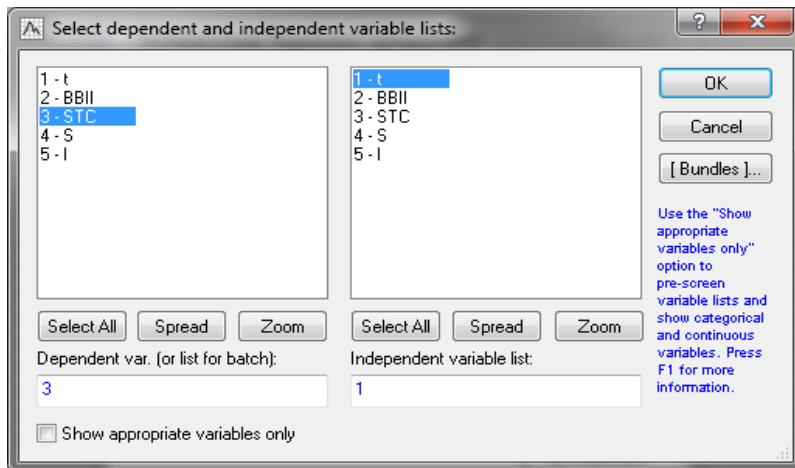
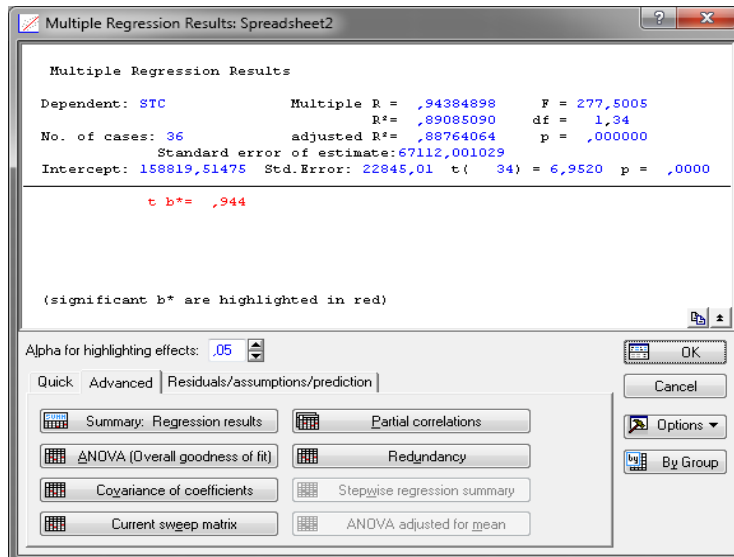


Рис. 13. Встановлення параметрів парної регресії

Результати побудови парної регресії наведено на рис. 14.



Regression Summary for Dependent Variable: STC (Spreadsheet2)						
R= .94384898 RI= .89085090 Adjusted RI= .88764064 F(1,34)=277,50 p<.000000 Std.Error of estimate: 67112,001029						
	b*	Std.Err. of b*	b	Std.Err. of b	t(34)	p-value
N=36						
Intercept			158819,5	22845,01	6,95204	0,000000
t	0,943849	0,056659	17936,5	1076,72	16,65835	0,000000

Рис. 14. Результати моделювання

Аналіз коефіцієнтів парної кореляції та детермінації дозволяють стверджувати, що модель є адекватною та якісною, параметри значимі. Тобто, цю модель можна використовувати для подальших розрахунків.

Наведемо загальний вигляд отриманої моделі: $Y = 158\ 819,5 + 17\ 936,5 \times t$.

Тепер необхідно вести отримане рівняння регресії у вікно параметрів створення нової змінної – *Trend* (Трендова складова) (рис. 15 і 16).

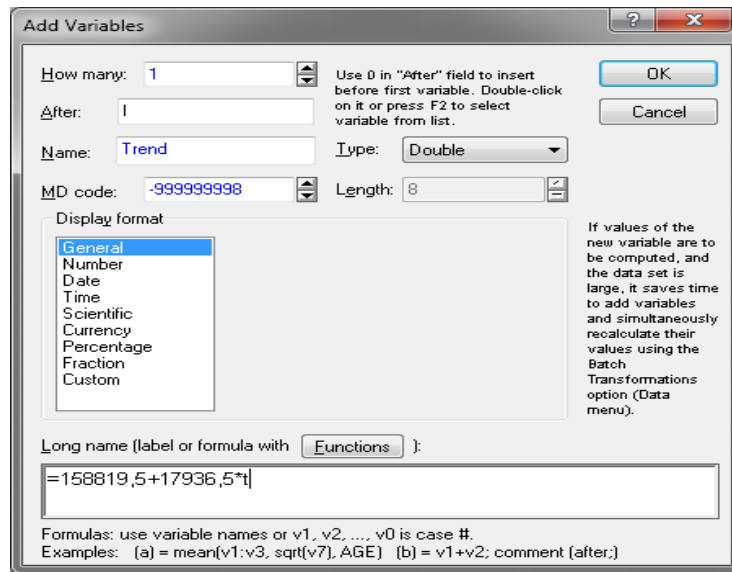


Рис. 15. Створення нової змінної *Trend*

	1 t	2 BBII	3 STC	4 S	5 I	6 Trend
1	1	217 074	260310,4	84,7807	0,983602	176756
2	2	255 545	265151,3	95,8736	1,005251	194692,5
3	3	300 446	274833,3	110,1065	0,992851	212629
4	4	306 281	287451,6	109,2392	0,975386	230565,5
5	5	258 591	303398,2	84,7807	1,005318	248502
6	6	310 277	318065,3	95,8736	1,017499	266438,5
7	7	368 488	329444,2	110,1065	1,015848	284375
8	8	362 635	337705,7	109,2392	0,982998	302311,5
9	9	292 324	345151,7	84,7807	0,998982	320248
10	10	346 005	350507,3	95,8736	1,029642	338184,5
11	11	387 109	352657,7	110,1065	0,996936	356121
12	12	379 231	354685,2	109,2392	0,978773	374057,5
13	13	303 753	358041,2	84,7807	1,000669	391994
14	14	354 814	363442,2	95,8736	1,018278	409930,5
15	15	398 000	367198,8	110,1065	0,984394	427867
16	16	408 631	373518,7	109,2392	1,001476	445803,5
17	17	316 905	380970,3	84,7807	0,981163	463740
18	18	382 391	391957,8	95,8736	1,017582	481676,5
19	19	440 476	403751,6	110,1065	0,990821	499613
20	20	447 143	421139,9	109,2392	0,971945	517549,5
21	21	375991	446316,9	84,7807	0,993659	535486
22	22	456715	477150,9	95,8736	0,998367	553422,5
23	23	566997	506244,1	110,1065	1,017204	571359
24	24	588841	528467,9	109,2392	1,020002	589295,5
25	25	455298	547940,0	84,7807	0,980090	607232
26	26	535701	574532,2	95,8736	0,972544	625168,5
27	27	671456	611827,4	110,1065	0,996726	643105
28	28	722912	650595,3	109,2392	1,017176	661041,5
29	29	592 523	686163,2	84,7807	1,018546	678978
30	30	665 233	719218,8	95,8736	0,964748	696914,5
31	31	834 287	758237,4	110,1065	0,999303	714851
32	32	891 839	796374,8	109,2392	1,025157	732787,5
33	33	705013	831135,0	84,7807	1,000526	750724
34	34	810820	864795,5	95,8736	0,977939	768660,5
35	35	994850	902878,5	110,1065	1,000727	786597
36	36	1048023	921920,0	109,2392	1,040636	804533,5

Рис. 16. Результати створення трендової складової

Необхідно візуалізувати цю складову (рис. 17).

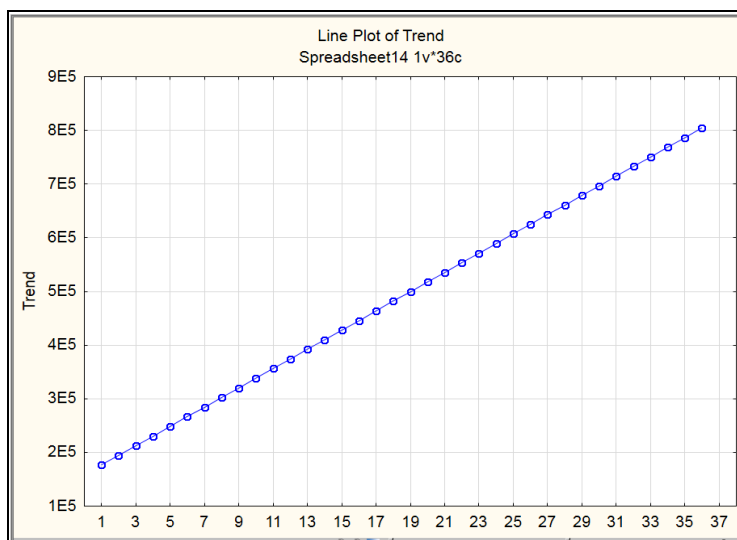


Рис. 17. Графік тренда

На третьому етапі з загальної мультиплікативної моделі декомпозиції розраховується значення циклічної складової у такий спосіб:

$$\text{Cycle} = \text{Smoothed Trend-C} / \text{Trend}. \quad (3)$$

Тобто створюємо нову змінну – *Cycle* (Циклічна складова) (рис. 18 і 19).

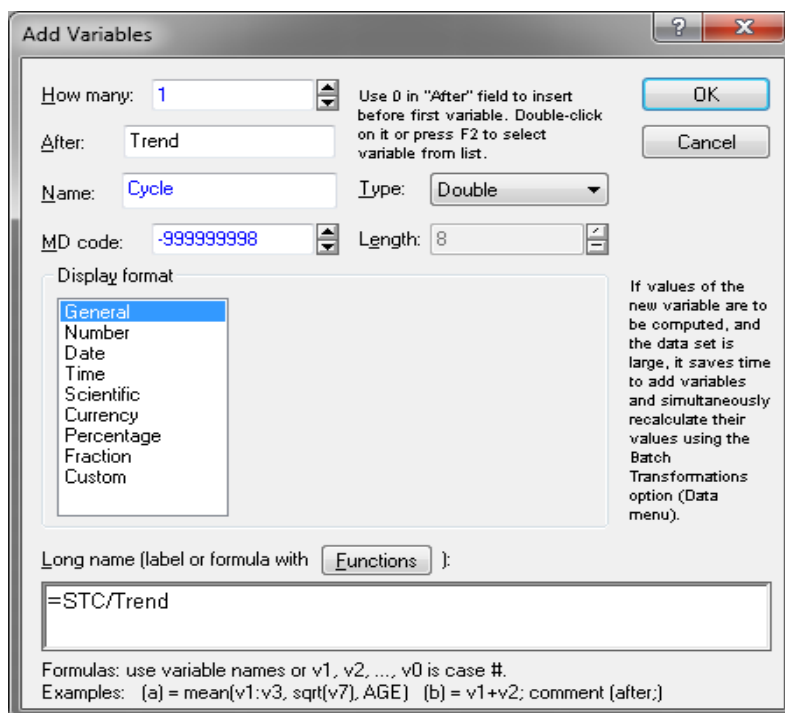


Рис. 18. Вікно створення нової змінної

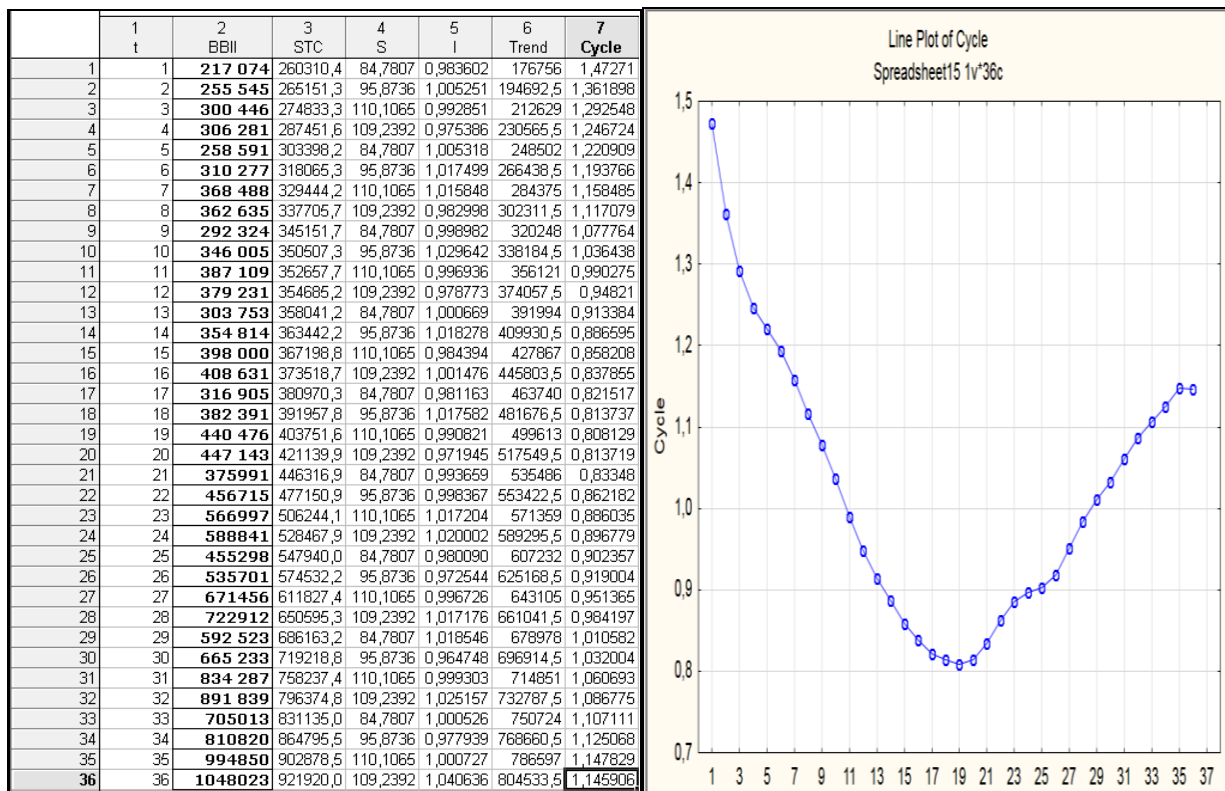


Рис. 19. Результати розрахунку та візуалізації циклічної складової

Перш ніж перейти до прогнозування ВВП на чотири періоди вперед (чотири квартали – один рік) за допомогою моделі декомпозиції часового ряду, необхідно виконати ряд дій:

1) додати чотири спостереження після останнього з наявних в ряді (рис. 20);

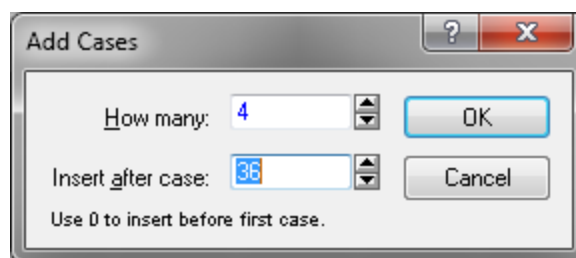


Рис. 20. Додавання чотирьох спостережень до вихідних даних


2) у стовбці даних t (період часу) вписати відповідні порядкові числівники, тим самим продовжити ряд (рис. 21);

3) у стовбці *Seasonal Factors* вписати відповідні значення сезонних складових (див. рис. 21);

	1 t	2 BBII	3 STC	4 S	5 I	6 Trend	7 Cycle
4	4	306 281	287451,6	109,2392	0,975386	230565,5	1,246724
5	5	258 591	303398,2	84,7807	1,005318	248502	1,220909
6	6	310 277	318065,3	95,8736	1,017499	266438,5	1,193766
7	7	368 488	329444,2	110,1065	1,015848	284375	1,158485
8	8	362 635	337705,7	109,2392	0,982998	302311,5	1,117079
9	9	292 324	345151,7	84,7807	0,998982	320248	1,077764
10	10	346 005	350507,3	95,8736	1,029642	338184,5	1,036438
11	11	387 109	352657,7	110,1065	0,996936	356121	0,990275
12	12	379 231	354685,2	109,2392	0,978773	374057,5	0,94821
13	13	303 753	358041,2	84,7807	1,000669	391994	0,913384
14	14	354 814	363442,2	95,8736	1,018278	409930,5	0,886595
15	15	398 000	367198,8	110,1065	0,984394	427867	0,858208
16	16	408 631	373518,7	109,2392	1,001476	445803,5	0,837855
17	17	316 905	380970,3	84,7807	0,981163	463740	0,821517
18	18	382 391	391957,8	95,8736	1,017582	481676,5	0,813737
19	19	440 476	403751,6	110,1065	0,990821	499613	0,808129
20	20	447 143	421139,9	109,2392	0,971945	517549,5	0,813719
21	21	375991	446316,9	84,7807	0,993659	535486	0,83348
22	22	456715	477150,9	95,8736	0,998367	553422,5	0,862182
23	23	566997	506244,1	110,1065	1,017204	571359	0,886035
24	24	588841	528467,9	109,2392	1,020002	589295,5	0,896779
25	25	455298	547940,0	84,7807	0,980090	607232	0,902357
26	26	535701	574532,2	95,8736	0,972544	625168,5	0,919004
27	27	671456	611827,4	110,1065	0,996726	643105	0,951365
28	28	722912	650595,3	109,2392	1,017176	661041,5	0,984197
29	29	592 523	686163,2	84,7807	1,018546	678978	1,010582
30	30	665 233	719218,8	95,8736	0,964748	696914,5	1,032004
31	31	834 287	758237,4	110,1065	0,999303	714851	1,060693
32	32	891 839	796374,8	109,2392	1,025157	732787,5	1,086775
33	33	705013	831135,0	84,7807	1,000526	750724	1,107111
34	34	810820	864795,5	95,8736	0,977939	768660,5	1,125068
35	35	994850	902878,5	110,1065	1,000727	786597	1,147829
36	36	1048023	921920,0	109,2392	1,040636	804533,5	1,145906
37	37			84,7807		822470	1,47271
38	38			95,8736		840406,5	1,361898
39	39			110,1065		858343	1,292548
40	40			109,2392		876279,5	1,246724

Рис. 21. Результати виконання допоміжних кроків

4) у стовбці *Cycle* вписати відповідні значення циклічної складової з урахуванням періоду циклу (див. рис. 11 б);

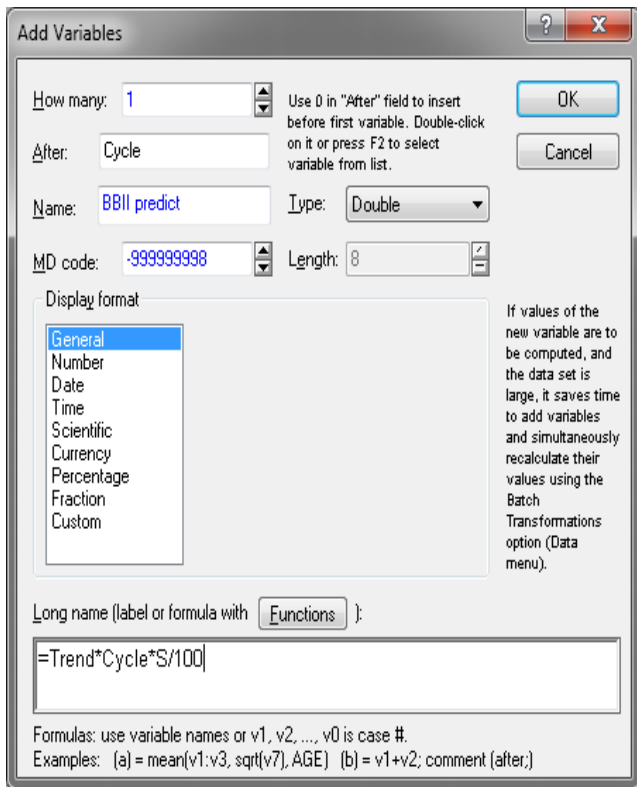
5) у стовбці *Trend* задати перерахунок даних (для цього необхідно виділити весь стовбець змінної *Trend* та натиснути на кнопку *Recalculation of data* (Перерахунок даних) на панелі інструментів , а потім виконати перерахунок (в нашому випадку з 1 по 40);

6) додати нову змінну $ВВП_{predict}$ ($ВВП_{прогноз}$) (рис. 22).

Тепер можна переходити до розрахунку прогнозних значень показника ВВП на 4 кроки вперед за допомогою вказівки моделі виду:

$$ВВП_{predict} = Trend \times Cycle \times Seasonal\ Factors / 100. \quad (4)$$

Ділення на 100 необхідно оскільки під час розрахунку сезонної складової отримані значення було помножено на 100.



1	2	3	4	5	6	7	8
t	BBII	STC	S	I	Trend	Cycle	BBII predict
4	306 281	287451,6	109,2392	0,975386	230665,5	1,246724	314009,917
5	258 591	303398,2	84,7807	1,005318	248502	1,220909	257223,132
6	310 277	318065,3	95,8736	1,017499	266438,5	1,193766	304940,703
7	368 488	329444,2	110,1065	1,015848	284375	1,158485	362738,445
8	362 635	337705,7	109,2392	0,982998	302311,5	1,117079	368907,075
9	292 324	345151,7	84,7807	0,998982	320248	1,077764	292621,977
10	346 005	350507,3	95,8736	1,029642	338184,5	1,036438	336044,011
11	387 109	352657,7	110,1065	0,996936	356121	0,990275	388298,926
12	379 231	354685,2	109,2392	0,978773	374057,5	0,94821	387455,392
13	303 753	358041,2	84,7807	1,000669	391994	0,913384	303549,842
14	354 814	363442,2	95,8736	1,018278	409930,5	0,886595	348445,173
15	398 000	367198,8	110,1065	0,984394	427867	0,858208	404309,721
16	408 631	373518,7	109,2392	1,001476	445803,5	0,837855	408028,885
17	316 905	380970,3	84,7807	0,981163	463740	0,821517	322989,309
18	382 391	391957,8	95,8736	1,017582	481676,5	0,813737	375784,088
19	440 476	403751,6	110,1065	0,990821	499613	0,808129	444556,652
20	447 143	421139,9	109,2392	0,971945	517549,5	0,813719	460049,929
21	375991	446316,9	84,7807	0,993659	535486	0,83348	378390,535
22	456715	477150,9	95,8736	0,998367	553422,5	0,862182	457461,811
23	566997	506244,1	110,1065	1,017204	571359	0,886035	557407,573
24	588841	528467,9	109,2392	1,020002	589295,5	0,896779	577294,203
25	455298	547940,0	84,7807	0,980090	607232	0,902357	464547,309
26	535701	574532,2	95,8736	0,972544	625168,5	0,919004	550824,688
27	671456	611827,4	110,1065	0,996726	643105	0,951365	673661,59
28	722912	650595,3	109,2392	1,017176	661041,5	0,984197	710705,227
29	592 523	686163,2	84,7807	1,018546	678978	1,010582	581733,963
30	665 233	719218,8	95,8736	0,964748	696914,5	1,032004	689540,996
31	834 287	758237,4	110,1065	0,999303	714851	1,060693	834868,549
32	891 839	796374,8	109,2392	1,025157	732787,5	1,086775	869953,643
33	705013	831135,0	84,7807	1,000526	750724	1,107111	704642,035
34	810820	864795,5	95,8736	0,977939	768660,5	1,125068	829110,637
35	994850	902878,5	110,1065	1,000727	786597	1,147629	994127,759
36	1048023	921920,0	109,2392	1,040636	804533,5	1,145906	1007098,25
37		84,7807	822470	1,47271	1026914,67		
38		95,8736	840406,5	1,361898	1097319,41		
39		110,1065	858343	1,292548	1221576,12		
40		109,2392	876279,5	1,246724	1193415,55		

Рис. 22. Вікно створення нової змінної *ВВП predict*

Створення графіку прогнозного значення ВВП наведено на рис. 23, а сам графік – на рис. 24.

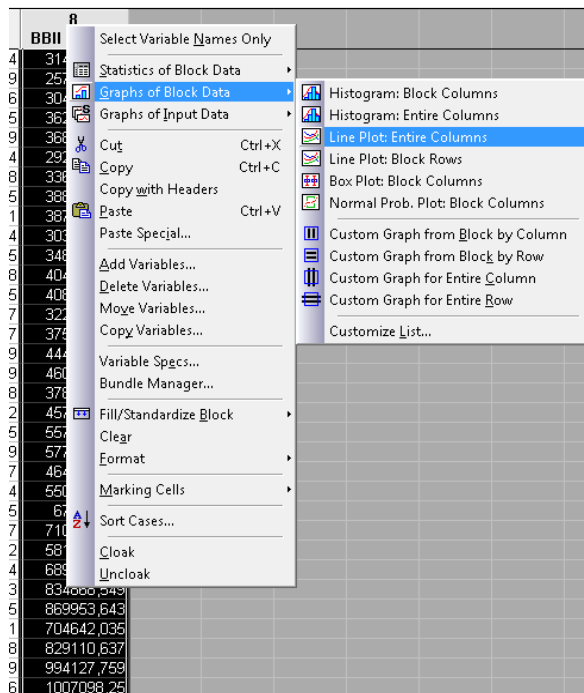


Рис. 23. Створення графіка *ВВП_{predict}*

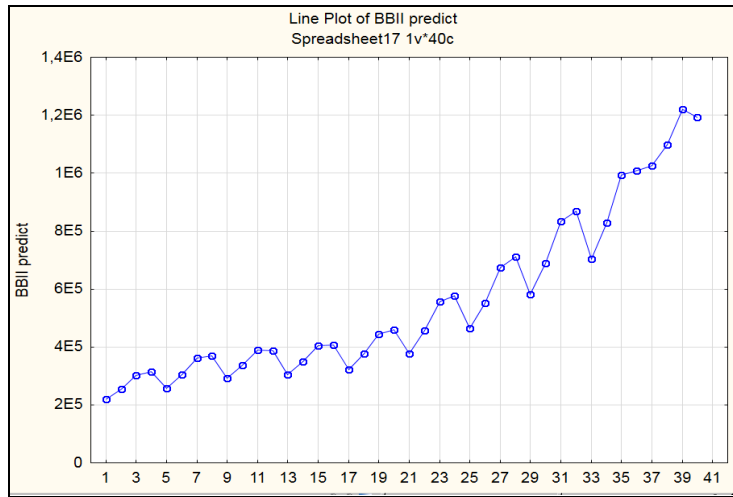


Рис. 24. Графік ВВП_{predict}

Підтвердженням адекватності моделі та точності прогнозу є відповідність (близькість) розподілу випадкової складової нормальному закону розподілу. На рис. 25 наведена гістограма розподілу помилок.

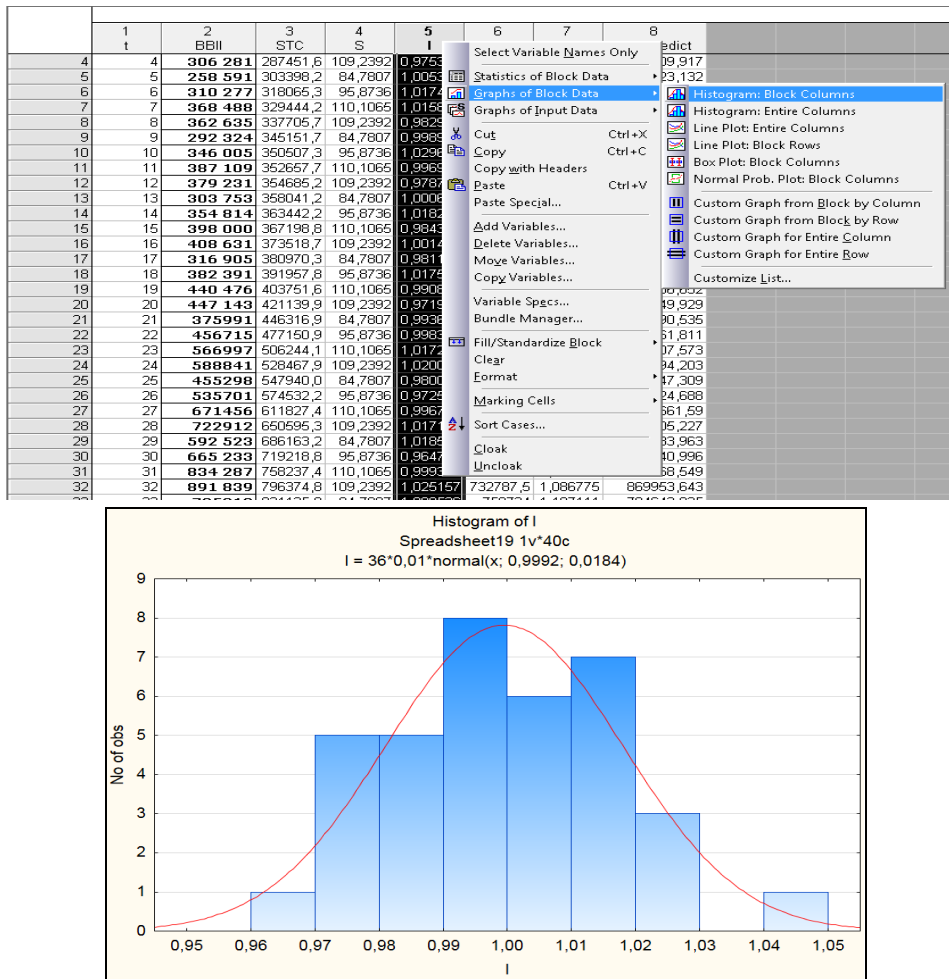


Рис. 25. Гістограма розподілу помилок

Критерієм якості (точності) побудованого прогнозу є розрахунок середньої абсолютної процентної помилки за формулою:

$$\text{MAPE} = \frac{1}{n} \times \sum \frac{|\bar{y} - \hat{y}_t|}{y} \times 100 \%, \quad (5)$$

де n – кількість рівнів ряду;

y – вихідні значення рівнів ряду у моделі;

\hat{y}_t – оцінки рівнів ряду за моделлю.

Розрахунок середньої абсолютної процентної помилки (MAPE) проводимо засобами MS Excel (рис. 26).

B	C	D	E	F
ВВП	ВВП predict	$y_i - y^{\wedge}$	ABS	$ y_i - y^{\wedge} /y$
217 074	220692,948	-3 619	3618,9	0,0166715
255 545	254210,147	1 335	1334,9	0,0052236
300 446	302609,223	-2 163	2163,2	0,0072
306 281	314009,917	-7 729	7728,9	0,0252347
258 591	257223,132	1 368	1367,9	0,0052897
310 277	304940,703	5 336	5336,3	0,0171985
368 488	362739,445	5 749	5748,6	0,0156004
362 635	368907,075	-6 272	6272,1	0,0172958
292 324	292621,977	-298	297,98	0,0010193
346 005	336044,011	9 961	9961	0,0287886
387 109	388298,926	-1 190	1189,9	0,0030739
379 231	387455,392	-8 224	8224,4	0,021687
303 753	303549,842	203	203,16	0,0006688
354 814	348445,173	6 369	6368,8	0,0179498
398 000	404309,721	-6 310	6309,7	0,0158536
408 631	408028,885	602	602,11	0,0014735
316 905	322989,309	-6 084	6084,3	0,0191992
382 391	375784,088	6 607	6606,9	0,0172779
440 476	444556,652	-4 081	4080,7	0,0092642
447 143	460049,929	-12 907	12907	0,0288653
375991	378390,535	-2 400	2399,5	0,0063819
456715	457461,811	-747	746,81	0,0016352
566997	557407,573	9 589	9589,4	0,0169127
588841	577294,203	11 547	11547	0,0196094
465298	464547,309	-9 249	9249,3	0,0203148
535701	550824,688	-15 124	15124	0,0282316
671456	673661,59	-2 206	2205,6	0,0032848
722912	710705,227	12 207	12207	0,0168856
592 523	581733,963	10 789	10789	0,0182086
665 233	689540,996	-24 308	24308	0,0365406
834 287	834868,549	-582	581,55	0,0006971
891 839	869953,643	21 885	21885	0,0245396
705013	704642,035	371	370,97	0,0005262
810820	829110,637	-18 291	18291	0,0225582
994850	994127,759	722	722,24	0,000726
1048023	1007098,25	40 925	40925	0,0390495
			Σ	0,5309368
			MAPE	1,4748245

Рис. 26. Розрахунок MAPE в Excel

Чим менше значення MAPE, тим вища якість побудованого прогнозу. На підставі розрахунку MAPE можна дійти висновку стосовно загального рівня адекватності моделі шляхом їхнього порівняння з табличними значеннями (табл. 2).

Таблиця 2

Залежність точності прогнозу від значень критерію MAPE

MAPE	Точність прогнозу
Менше 10 %	Висока
10 – 20 %	Добра
20 – 40 %	Задовільна
40 – 50 %	Погана
Більше 50 %	Незадовільна

Середня абсолютна процентна помилка дорівнює 1,475 %, що підтверджує точність прогнозу за цією мультиплікативною моделлю часового ряду.

Отже, можна стверджувати, що у 2021 р. ВВП України буде зростати, що матиме позитивний вплив на економічний розвиток держави.

Завдання:

1. Використовуючи власний інформаційний простір (статистичні дані за обраною тематикою), необхідно побудувати модель декомпозиції для прогнозування часового ряду.

- 1.1. Визначте вид моделі декомпозиції.
- 1.2. Виявіть всі складові.
- 1.3. Спрогнозуйте трендову складову.
- 1.4. Здійсніть спектральний аналіз циклічної складової.
- 1.5. Здійсніть композицію моделі та перевірте її якість.
- 1.6. За результатами проведених розрахунків сформулюйте висновки. Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. З метою оцінювання точності прогнозу грошової маси M2 (млн грн) ряд динаміки поділено на: період ретроспекції (7 кварталів) і прогнозний період (три квартали). Складено два варіанти прогнозу (табл. 3). Використовуючи стандартну похибку, оцініть точність прогнозу за кожним варіантом, зробіть висновок.

Дані щодо грошової маси

Прогнозний період	Прогнозний рівень за варіантом		Фактичний рівень
	1	2	
p + 1	8 152	8 330	8 154
p + 2	8 244	8 552	8 245
p + 3	8 658	8 454	8 365

Завдання 2. Відома динаміка макроекономічних показників по певній країні за 3 роки (2019 – 2021 рр.) (табл. 4): ВВП (X1), виробництво товарів та послуг (X2), індекс споживчих цін (X3). Дослідіть нелінійність у динаміці цих показників, проаналізуйте структурні складові нелінійного розвитку та їхній вклад у динаміку показників. Проведіть порівняльний аналіз основних тенденцій розвитку показників національної економіки. Зробіть економічні висновки щодо стану національної економіки на наступний квартал.

Таблиця 4

Вихідні дані

Період (місяць)	X1, млрд грн	X2, млрд грн	X3, %	Період (місяць)	X1, млрд грн	X2, млрд грн	X3, %
1	29,8	10,2	101,2	19	64,2	48,8	101,4
2	30,6	19,4	101,8	20	65,4	50,6	100,6
3	32,6	32,4	99,7	21	73,0	50,1	102,2
4	35,6	36,4	99,6	22	68,6	51,8	102,9
5	37,9	35,0	100,5	23	68,9	54,7	102,2
6	39,5	36,9	100,1	24	72,1	57,2	102,1
7	47,6	38,2	100,9	25	57,7	35,4	102,9
8	47,1	39,9	100,0	26	60,8	18,7	102,7
9	49,8	40,5	102,0	27	71,2	58,0	103,8
10	50,1	40,9	102,6	28	76,8	65,4	103,1
11	52,3	40,7	101,8	29	81,5	68,4	101,3
12	32,3	42,6	100,9	30	81,4	71,6	100,8
13	44,1	20,5	100,5	31	100,8	74,3	99,5
14	43,3	20,4	100,6	32	97,8	76,9	99,9
15	44,9	41,0	100,2	33	86,8	72,5	101,1
16	48,4	48,1	100,0	34	85,1	70,0	101,7
17	51,9	46,8	100,6	35	75,4	58,8	101,5
18	57,4	47,5	102,2	36	80,8	49,0	102,1

Запитання для самоконтролю

1. Чим відрізняється часовий ряд від просторового?
2. Які існують види часових рядів?
3. Назвіть відомі вам критерії перевірки якості побудованих моделей.
4. У чому полягає визначення стаціонарного часового ряду в широкому сенсі?
5. Які існують методи дослідження стаціонарності?
6. У якому випадку використовують адитивну та мультиплікативну модель?
7. Перерахуйте моделі згладжування?
8. Назвіть основні етапи проведення методу Фостера – Стюарта?
9. Які основні кроки побудови моделі декомпозиції?
10. Які методи використовуються для перевірки якості прогнозу?

Тема 10. Моделювання повних циклів і процесів оновлення

Лабораторна робота 6. Оцінювання наявності автокореляції залишків моделі. Усунення автокореляції

Мета: набуття навичок оцінювання наявності автокореляції залишків моделі.

Завдання: побудувати модель впливу соціально-економічних показників на ВВП країни та перевірити модель на наявність автокореляції.

Методичні рекомендації

Вихідні дані для виконання лабораторної роботи наведено у табл. 5.

Вихідні дані

Роки	Випуск продукції тис. грн (X1)	Обсяг роздрібного товарообороту підприємств, млн грн (X2)	Сукупні витрати в середньому за місяць у розрахунку на одне домогосподарство, грн (X3)	Прямі інвестиції (млн дол. США) (X4)	ВВП (млрд грн) (Y)
2007	226 358	19 317	395,6	2 063,6	186,5
2008	356 842	22 151	426,5	2 810,7	192,5
2009	373 893	28 757	541,3	3 281,8	198,9
2010	460 520	34 417	607,0	3 875,0	221,6
2011	504 008	39 691	658,3	4 555,3	225,8
2012	603 704	49 994	736,8	5 471,8	267,3
2013	809 988	67 556	903,5	6 794,4	345,1
2014	995 630	94 332	1 229,4	9 047,0	441,5
2015	1 182 179	129 952	1 442,8	16 890,0	544,2
2016	1 565 055	178 233	1 722,0	21 607,3	720,7
2017	2 072 172	246 903	2 590,4	29 542,7	948,1
2018	1 955 685	230 955	2 754,1	35 616,4	913,3
2019	2 388 289	280 890	3 072,7	40 053,0	1 082,6
2020	2 496 365	350 059	3 456,0	44 806,0	1 316,6

1. Побудова багатofакторної економетричної моделі.

Для побудови та всебічного аналізу множинних лінійних економетричних моделей у пакеті Statistica 10.0 передбачений модуль *Multiple Regression* (множинна регресія). Для того щоб розпочати обчислювальні процедури необхідно увійти в меню *Statistics / Multiple Regression*. У стартовій панелі цього модуля необхідно задати змінні для аналізу. Результати побудови лінійної економетричної моделі представлені в діалоговому вікні (рис. 27).

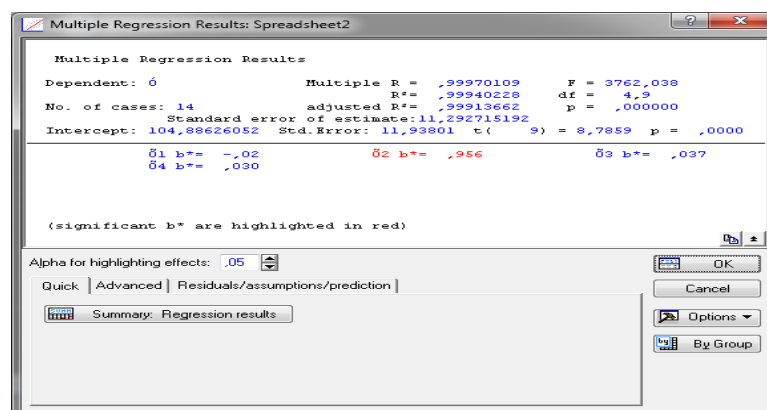


Рис. 27. Вікно результатів регресійного аналізу

У верхній частині вікна міститься основна інформація про модель, в нижній частині знаходяться функціональні кнопки, що дозволяють всебічно розглянути результати аналізу. Ініціювавши кнопку *Summary: Regression results* (результати регресійного аналізу), визначимо найважливіші характеристики моделі та ступінь її адекватності (рис. 28).

Regression Summary for Dependent Variable: Y (Spreadsheet2) R= ,99970109 RI= ,99940228 Adjusted RI= ,99913662 F(4,9)=3762,0 p<,00000 Std.Error of estimate: 11,293						
N=14	b*	Std.Err. of b*	b	Std.Err. of b	t(9)	p-value
Intercept			104,8863	11,93802	8,78590	0,000010
X1	-0,022602	0,073626	-0,0000	0,00004	-0,30699	0,765837
X2	0,955608	0,082814	0,0033	0,00029	11,53916	0,000001
X3	0,036765	0,111982	0,0132	0,04019	0,32831	0,750184
X4	0,030262	0,081871	0,0008	0,00205	0,36963	0,720203

Рис. 28. Результати регресійного аналізу

Аналіз рис. 28 дозволяє стверджувати, що задана модель загалом є адекватною й якісною, але параметри за показників X1, X3 і X4 є не значущими. Відповідно виникає необхідність перевірки моделі на наявність автокореляції.

2. Перевірка моделі на наявність автокореляції.

Важливою передумовою побудови якісної регресійної моделі за допомогою методу найменших квадратів (МНК) є незалежність значень випадкових відхилень ε_i від значень відхилень у всіх інших спостереженнях. Відсутність залежності гарантує відсутність корельованості між будь-якими відхиленнями (при $\sigma(\varepsilon_i, \varepsilon_j) = \text{cov}(\varepsilon_i, \varepsilon_j) = 0$ за $i \neq j$) і, зокрема, між сусідніми відхиленнями ($\sigma(\varepsilon_{i-1}, \varepsilon_i) = 0$), $i = 2, 3, \dots, n$.

Автокореляція (послідовна кореляція) визначається як кореляція між спостережуваними показниками, впорядкованими в часі (часові ряди) або в просторі (перехресні дані). Автокореляція залишків (відхилень) зазвичай зустрічається в регресійному аналізі під час використання даних часових рядів. Під час використання перехресних даних наявність автокореляції (просторової кореляції) вкрай рідка.

Автокореляція не перешкоджає знаходженню зв'язку між досліджуваним показником і факторами, що впливають на нього, як і визначенню

параметрів і статистичних характеристик рівняння, але її наявність не гарантує надійності регресійного рівняння і параметрів та можливості побудови інтервалів довіри. Отже, автокореляція більш "небезпечна" під час побудови прогнозу, ніж при проведенні економічного аналізу.

Серед основних причин, що викликають появу автокореляції, можна виділити: помилки специфікації; інерція (циклічність); ефект павутини (наявність часового лагу); згладжування даних.

Наявність автокореляції визначається за допомогою використання: критерію Дарбіна – Уотсона; критерію фон Неймана; нециклічного критерію автокореляції та циклічного критерію автокореляції.

3. Розрахунок критерію Дарбіна – Уотсона.

Для визначення коефіцієнта Дарбіна – Уотсона з метою оцінювання наявності автокореляції в залишках моделі в меню аналізу помилок ініціюємо кнопку *Perform residual analysis* (рис. 29).

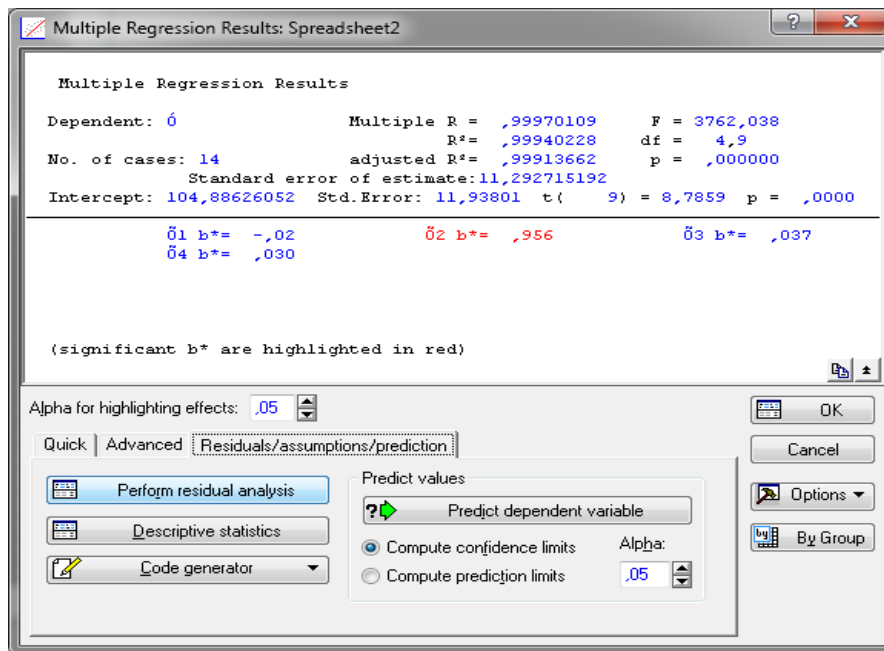


Рис. 29. Меню аналізу помилок

У меню аналізу помилок ініціювавши кнопку *Durbin–Watson statistic* (статистика Дарбіна – Уотсона) (рис. 30) отримуємо значення автокореляції помилок моделі за критерієм Дарбіна – Уотсона і значення нециклічного коефіцієнта автокореляції, що виражає ступінь взаємозв'язку рядів (рис. 31).

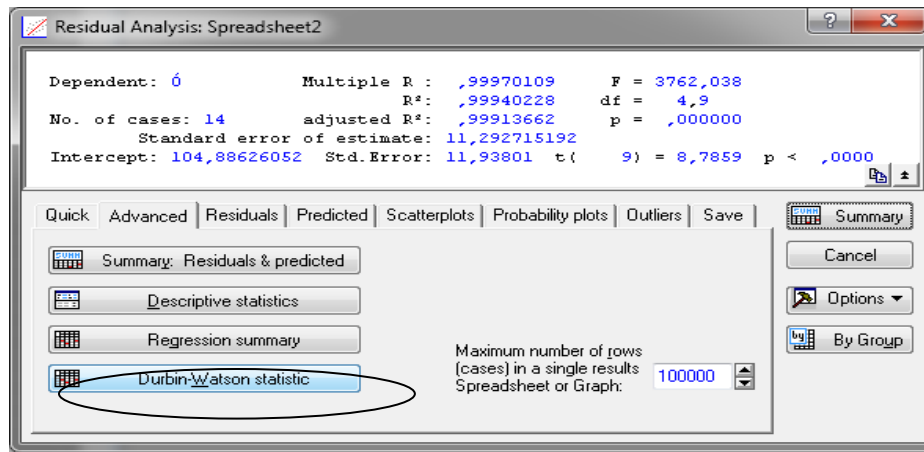


Рис. 30. Модуль аналізу помилок

Durbin-Watson d (Spreadsheet2) and serial correlation of residuals	
	Durbin-Watson d
Estimate	1,595394

Рис. 31. Результат розрахунку критерію Дарбіна – Уотсона за допомогою пакета Statistica

Критичні (табличні) значення критерію при кількості спостережень $n = 14$, ступенях свободи (кількість регресорів) $k = 4$ та рівні значущості $\alpha = 0,05$ дорівнюють $d_l = 0,69$, $d_u = 1,97$.

Формуються висновки за наступною схемою:

- 1) якщо $0 < DW < d_l$, наявною є додатна автокореляція залишків;
- 2) якщо $d_l \leq DW \leq d_u$, висновок про наявність автокореляції не визначається (зона невизначеності);
- 3) якщо $d_u < DW < 4 - d_u$, автокореляція відсутня;
- 4) якщо $4 - d_u \leq DW \leq 4 - d_l$, висновок про наявність автокореляції не визначається;
- 5) якщо $4 - d_l < DW < 4$, наявною є від'ємна автокореляція залишків.

Це графічно зображено на рис. 32.



Рис. 32. Графічна інтерпретація інтервалів автокореляції

Тобто $0,69 < 1,595 < 1,97$ ($d_l < d < d_u$). Це свідчить що розраховане значення критерію перебуває у зоні невизначеності, отже використання одного критерію недостатньо.

Коефіцієнт циклічної кореляції (Serial Corr.) становить 0,124. Фактично обчислене значення циклічного коефіцієнта автокореляції порівнюється з табличним для вибраного рівня значущості α і довжини ряду n . Якщо $|r_{\text{розн.}}| < |r_{5\% \text{табл.}}|$, то приймаємо гіпотезу неавтокорельованості залишків ε_t ; якщо $|r_{\text{розн.}}| > |r_{1\% \text{табл.}}|$ відкидаємо гіпотезу щодо їхньої неавтокорельованості.

Отже, зробити однозначний висновок про наявність автокореляції залишків моделі складно.

4. Усунення автокореляції методом Ейткена.

Коли має місце автокореляція залишків, то визначення параметрів моделі можливо за допомогою метода Ейткена.

Метод Ейткена відрізняється від звичайного МНК тим, що під час оцінювання параметрів моделі використовується матриця Ω , що відображає коригування вихідних даних на мінливість дисперсії. Параметри моделі визначаються за формулами (6), (7) та (8):

$$a = (X^T \times \Omega^{-1} \times X)^{-1} X^T \times \Omega^{-1} \times Y. \quad (6)$$

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & \rho & 0 & 0 & 0 & \dots & 0 \\ \rho & 1+\rho^2 & \rho & 0 & 0 & \dots & 0 \\ 0 & \rho & 1+\rho^2 & \rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \quad (7)$$

На практиці для розрахунку параметру ρ використовується теж співвідношення, що й для визначення циклічного коефіцієнта кореляції:

$$\rho \approx r^0 \approx \frac{\sum_{t=1}^n u_t \times u_{t-1}}{\sum_{t=1}^n (u_t)^2}. \quad (8)$$

Як видно з розрахунків на рис. 31, значення циклічного коефіцієнту кореляції становить 0,124, тобто параметр $\rho = 0,124$.

З метою усунення автокореляції в моделі пропонується використувати метод Ейткена. Подальші розрахунки було проведено в пакеті MS Excel з використанням вбудованих функцій.

Наведено алгоритм розрахунку:

1. Визначаємо матрицю Ω .
2. Обчислюємо обернену матрицю Ω^{-1} .
3. Перемножуємо матрицю X^T на Ω^{-1} , де X^T – матриця, транспонована до матриці незалежних змінних X .
4. Знаходимо добуток $X^T \Omega^{-1} X$.
5. Обчислюємо обернену матрицю $(X^T \Omega^{-1} X)^{-1}$ та матрицю $X^T \Omega^{-1} Y$.
6. Знаходимо матрицю $A = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y$, елементи якої і будуть коефіцієнтами лінійного рівняння. Y – вектор залежної змінної.

Відповідно до наведеного раніше алгоритму спочатку необхідно ввести вихідні дані в робочий аркуш пакета MS Excel (рис. 33).

	A	B	C	D	E
1	x1	x2	x3	x4	y
2	226358	19317	395,6	2063,6	186,5
3	356842	22151	426,5	2810,7	192,5
4	373893	28757	541,3	3281,8	198,9
5	460520	34417	607	3875	221,6
6	504008	39691	658,3	4555,3	225,8
7	603704	49994	736,8	5471,8	267,3
8	809988	67556	903,5	6794,4	345,1
9	995630	94332	1229,4	9047	441,5
10	1182179	129952	1442,8	16890	544,2
11	1565055	178233	1722	21607	720,7
12	2072172	246903	2590,4	29543	948,1
13	1955685	230955	2754,1	35616	913,3
14	2388289	280890	3072,7	40053	1082,6
15	2496365	350059	3456	44806	1316,6

Рис. 33. Вихідні дані для побудови багатофакторної моделі

На наступному етапі проводимо проміжні розрахунки. Для цього запишемо матрицю незалежної змінних X та залежної Y (рис. 34).

	A	B	C	D	E	F	G	H	I	J
16										
17		1	226358	19317	395,6	2063,6				186,5
18		1	356842	22151	426,5	2810,7				192,5
19		1	373893	28757	541,3	3281,8				198,9
20		1	460520	34417	607	3875				221,6
21		1	504008	39691	658,3	4555,3				225,8
22		1	603704	49994	736,8	5471,8				267,3
23	x=	1	809988	67556	903,5	6794,4			y=	345,1
24		1	995630	94332	1229,4	9047				441,5
25		1	1182179	129952	1442,8	16890				544,2
26		1	1565055	178233	1722	21607				720,7
27		1	2072172	246903	2590,4	29543				948,1
28		1	1955685	230955	2754,1	35616				913,3
29		1	2388289	280890	3072,7	40053				1082,6
30		1	2496365	350059	3456	44806				1316,6

Рис. 34. Представлення матриць змінних X та Y

Транспонування в пакеті *MS Excel* проводиться за допомогою функції *ТРАНСП(масив)* (рис. 35). Для цього спочатку виділяємо відповідний масив вільних комірок і вводимо формулу, після натискаємо комбінацію клавіш *Ctrl + Shift + Enter* (рис. 36).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
16															
17		1	226358	19317	395,6	2063,6				186,5					
18		1	356842	22151	426,5	2810,7				192,5					
19		1	373893	28757	541,3	3281,8				198,9					
20		1	460520	34417	607	3875				221,6					
21		1	504008	39691	658,3	4555,3				225,8					
22		1	603704	49994	736,8	5471,8				267,3					
23	x=	1	809988	67556	903,5	6794,4			y=	345,1					
24		1	995630	94332	1229,4	9047				441,5					
25		1	1182179	129952	1442,8	16890				544,2					
26		1	1565055	178233	1722	21607				720,7					
27		1	2072172	246903	2590,4	29543				948,1					
28		1	1955685	230955	2754,1	35616				913,3					
29		1	2388289	280890	3072,7	40053				1082,6					
30		1	2496365	350059	3456	44806				1316,6					

Рис. 35. Вікно функції транспонування

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
33															
34		1	1	1	1	1	1	1	1	1	1	1	1	1	1
35		226358	356842	373893	460520	504008	603704	809988	995630	1182179	1565055	2072172	1955685	2388289	2496365
36	X ^T	19317	22151	28757	34417	39691	49994	67556	94332	129952	178233	246903	230955	280890	350059
37		395,6	426,5	541,3	607	658,3	736,8	903,5	1229,4	1442,8	1722	2590,4	2754,1	3072,7	3456
38		2063,6	2810,7	3281,8	3875	4555,3	5471,8	6794,4	9047	16890	21607,3	29542,7	35616,4	40053	44806

Рис. 36. Значення матриці X^T

З урахуванням значення циклічного коефіцієнта автокореляції та формули (7) було сформовано матрицю Ω^{-1} розмірністю 14×14 (рис. 37).

якої дорівнює кількості рядків першої матриці, а кількість стовбців дорівнює кількості стовбців другої матриці. Тобто, в нашому випадку під час множення матриці X^T розмірністю 5×14 на матрицю Ω^{-1} розмірністю 14×14 отримуємо матрицю розмірністю 5×14 . Але спочатку виділяємо діапазон вільних комірок, де будуть розміщені результати перемноження двох матриць та одночасно натискаємо комбінацію клавіш *Ctrl + Shift + Enter*. Унаслідок отримуємо нову матрицю (рис. 39).

B57		fx {=МУМНОЖ(B34:O38;B41:O54)}													
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
56															
57		0,889679233	0,779362413	0,779362413	0,77936241	0,779362413	0,779362	0,779362	0,779362	0,7793624	0,779362413	0,779362	0,779362	0,889679	
58		184953,2641	292394,6403	282635,9664	364344,77	385726,231	457080,8	633873,8	775842,4	896621,07	1204098,518	1693507,712	1455037	1902213	2234575
59	$X^T \Omega^{-1}$	16829,02863	16788,6229	22531,2612	26871,9278	30300,33697	38049,27	51490,22	72405,03	99685,217	136340,3739	203083,4448	171700,5	216492,8	320151,2
60		348,0656551	321,831965	428,052375	474,886476	509,6287934	563,1266	684,1048	972,316	1116,1774	1267,860066	2107,608114	2126,934	2386,599	3123,003
61		1741,854623	2225,312513	2542,334674	3009,06025	3520,483309	4213,372	5178,179	6346,854	13557,072	16434,64	23258,92665	27964,25	31175,96	40461,55
62															

Рис. 39. Результати використання функції МУМНОЖ

На наступному етапі знаходимо добуток матриць $X^T \Omega^{-1}$ та X за допомогою функції МУМНОЖ і вже отримуємо матрицю розмірністю 5×5 (рис. 40).

МУМНОЖ		fx {=МУМНОЖ(B57:O61;B17:F30)}													
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
58		184953,2641	292394,6403	282635,9664	364344,77	385726,231	457080,8	633873,8	775842,4	896621,07	1204098,518	1693507,712	1455037	1902213	2234575
59	$X^T \Omega^{-1}$	16829,02863	16788,6229	22531,2612	26871,9278	30300,33697	38049,27	51490,22	72405,03	99685,217	136340,3739	203083,4448	171700,5	216492,8	320151,2
60		348,0656551	321,831965	428,052375	474,886476	509,6287934	563,1266	684,1048	972,316	1116,1774	1267,860066	2107,608114	2126,934	2386,599	3123,003
61		1741,854623	2225,312513	2542,334674	3009,06025	3520,483309	4213,372	5178,179	6346,854	13557,072	16434,64	23258,92665	27964,25	31175,96	40461,55
62															
63															
64		=МУМНОЖ(B57:O61;B17:F30)													
65	$X^T \Omega^{-1} X$														
66															
67															
68															
69															

Аргументы функции

МУМНОЖ

Массив1 B57:O61 = {0,889679233;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413;0,779362413}

Массив2 B17:F30 = {1;226358;19317;395,6;2063,6;1;3;...}

= {11,13170742;12762903,33;1422719,272;16430,1945;181629,846}

Возвращает матричное произведение двух массивов; результат имеет то же число строк, что и первый массив, и то же число столбцов, что и второй массив.

Массив2 первый из перемножаемых массивов, число столбцов в нем должно равняться числу строк во втором массиве.

B64		fx {=МУМНОЖ(B57:O61;B17:F30)}					
	A	B	C	D	E	F	
63							
64		11,13170742	12762903,33	1422719,272	16430,1945	181629,846	
65	$X^T \Omega^{-1} X$	12762903,33	2,15965E+13	2,58674E+12	2,8018E+10	3,38454E+11	
66		1422719,272	2,58674E+12	3,15617E+11	3375054625	41371606010	
67		16430,19453	28018149135	3375054625	36596185,8	443500426,8	
68		181629,846	3,38454E+11	41371606010	443500427	5490362228	

Рис. 40. Результати використання функції МУМНОЖ

Для отримання матриці зворотної заданій використовується функція *МОБР(масив)* (рис. 41). Розмірність зворотної матриці буде також 5×5 . Але спочатку виділяємо діапазон вільних комірок, де будуть розміщені результати зворотної матриці та одночасно натискаємо комбінацію клавіш *Ctrl + Shift + Enter*.

The image displays the implementation of the *МОБР* function in Excel. It is divided into three main sections:

- Top Section:** A spreadsheet window titled "МУМНОЖ" showing a 5x5 matrix of values in the range B64:F68. The values are:

63						
64		11,13170742	12762903,33	1422719,272	16430,1945	181629,846
65	ХТ Ω-1 X	12762903,33	2,15965E+13	2,58674E+12	2,8018E+10	3,38454E+11
66		1422719,272	2,58674E+12	3,15617E+11	3375054625	41371606010
67		16430,19453	28018149135	3375054625	36596185,8	443500426,8
68		181629,846	3,38454E+11	41371606010	443500427	5490362228
69						
70						
71		=МОБР(B64:F68)				
72	(ХТ Ω-1 X) ⁻¹					
73						
74						
75						
76						
- Middle Section:** A dialog box titled "Аргументы функции" (Function Arguments) for the *МОБР* function. It shows the array argument as "Массив B64:F68" and its numerical representation: $\{42351759437,5238;65491907698,9...; \dots\}$. The description states: "Возвращает обратную матрицу (матрица хранится в массиве)." (Returns the inverse matrix (the matrix is stored in an array)). The value shown is 14,32884322.
- Bottom Section:** A spreadsheet window titled "B71" showing the resulting inverse matrix in the range B71:F75. The values are:

70						
71		1,156851747	-8,35801E-07	1,26381E-05	-0,0024458	0,000115591
72	(ХТ Ω-1 X) ⁻¹	-8,35801E-07	9,36743E-12	-3,50227E-11	-4,832E-09	1,0439E-10
73		1,26381E-05	-3,50227E-11	6,23548E-10	-2,48E-08	-9,54397E-10
74		-0,00244585	-4,83162E-09	-2,48006E-08	1,2192E-05	-4,19198E-07
75		0,000115591	1,0439E-10	-9,54397E-10	-4,192E-07	3,09767E-08

Рис. 41. Вікно реалізації функції *МОБР*

За відповідною процедурою визначаємо добуток матриці $X^T \Omega^{-1}$ на вектор Y (рис. 42).

B78		fx {=МУМНОЖ(B57:O61;J17:J30)}			
	A	B	C	D	
76					
77					
78		6092,634556			
79	ХТ Ω-1 У	10289905091			
80		1241308071			
81		13405601,96			
82		162272499,1			
83					

Рис. 42. Результати розрахунку матриці $X^T \Omega^{-1} Y$

Визначаємо параметри моделі. Для цього використовуємо наведені функції та розрахунки проводимо відповідно до формули (6). Результати розрахунків наведено на рис. 43.

МУМНОЖ		fx {=МУМНОЖ(B71:F75;B78:B82)}										
	A	B	C	D	E	F	G	H	I	J	K	
70												
71		14,31200329	6,406535675	-22,13954515	-0,8455898	6,232985475						
72	$(X^T \Omega^{-1} X)^{-1}$	4,533456442	2,029330343	-7,012839177	-0,2678783	1,974334452						
73		-21,97577107	-9,837101553	33,99520071	1,29824855	-9,570789399						
74		-1,331480002	-0,596024044	2,059553224	0,07874756	-0,579831662						
75		7,06109867	3,160796048	-10,92317759	-0,417141	3,075269526						
76												
77												
78		6092,634556		=МУМНОЖ(B71:F75;B78:B82)	$a = (X^T * \Omega^{-1} * X)^{-1} X^T * \Omega^{-1} * Y$							
79	ХТ Ω-1 У	10289905091	A									
80		1241308071										
81		13405601,96										
82		162272499,1										
83												

77					
78		6092,634556			104,9388137
79	ХТ Ω-1 У	10289905091	A		-7,48253E-06
80		1241308071			0,003294532
81		13405601,96			0,011566208
82		162272499,1			0,000785971
83					

Рис. 43. Розрахунок параметрів багатofакторної регресійної моделі методом Ейткена

Унаслідок формуємо загальний вигляд багатofакторної регресійної моделі. Загальний вид моделі:

$$Y = 104,945 - 7,48 X_1 + 0,0033 X_2 + 0,0116 X_3 + 0,0008 X_4.$$

5. Перевірка отриманої моделі на наявність автокореляції залишків.

З метою перевірки та усунення автокореляції залишків у моделі доцільно розрахувати коефіцієнт Дарбіна – Уотсона (9):

$$DW = \frac{\sum_{t=2}^T (u_t - u_{t-1})^2}{\sum_{t=1}^T u_t^2}, \quad (9)$$

де u – залишки регресії.

Знайдемо розрахункові значення на основі побудованої економетричної моделі та визначимо залишки (табл. 6).

Таблиця 6

Розрахунок залишків моделі

Роки	Y	Y _t	u _t	u _t ²	(u _t - u _{t-1}) ²
2007	186,5	176,470 543 5	10,029 456 53	100,59	—
2008	192,5	187,728 193 1	4,771 806 87	22,770 14	27,642 879 95
2009	198,9	211,317 528 6	-12,417 528 56	154,195	295,473 252 7
2010	221,6	231,838 907 6	-10,238 907 55	104,835 2	4,746 389 515
2011	225,8	250,667 712 9	-24,867 712 94	618,403 1	214,001 947 1
2012	267,3	286,985 545 6	-19,685 545 63	387,520 7	26,854 858 01
2013	345,1	349,355 257 4	-4,255 257 357	18,107 22	238,093 796 3
2014	441,5	444,498 627 1	-2,998 627 063	8,991 764	1,579 119 697
2015	544,2	571,878321 5	-27,678 321 46	766,089 5	609,087 315 7
2016	720,7	740,743 455 2	-20,043 455 16	401,740 1	58,291 183 54
2017	948,1	987,054 585 2	-38,954 585 22	1517,46	357,630 840 3
2018	913,3	940,308 912 8	-27,008 912 82	729,481 4	142,699 089
2019	1 082,6	1 115,230 38	-32,630 380 33	1064,742	31,600 896 94
2020	1 316,6	1 352,087 602	-35,487 602	1259,37	8,163 715 681
Сума	—	—	—	7 154,295	2 015,865 284

$$DW = \frac{2\ 015,865\ 284\ 2}{7\ 154,295} = 0,282.$$

Критичні (табличні) значення критерію дорівнюють $d_l = 0,69$, $d_u = 1,97$.

Умова $d_u < DW < 4 - d_u$ – d_u свідчить про відсутність автокореляції моделі. В нашому випадку: $0,69 > 0,28 < 4 - 1,97$. Тобто $0 < DW < d_l$, отже наявною є додатна (позитивна) автокореляція залишків.

Отже, модель побудована з використанням методу Ейткена має автокореляції в залишках. Це є одним з підтверджень недостатньої якості моделі.

Існує ряд способів виключення або зменшення автокореляції в часових рядах. Найбільш ефективним є виключення тренда із числового ряду та перехід до випадкової компоненти. Для усунення автокореляції використовуються такі методи:

метод, заснований на включенні часу в рівняння множинної регресії як аргументу (метод Фріша – Боу);

метод кінцевих різниць, коли методом найменших квадратів обробляються самі рівні вихідних рядів, а не їх послідовні різниці між ними;

метод відхилень емпіричних значень від вирівняних по тренду;

метод Кохрейна – Оркатта;

метод Хілдрета – Лу.

Висновок. З метою усунення автокореляції в залишках моделі, в роботі були проведені розрахунки параметрів моделі з використанням методу Ейткена. Відповідно до отриманих результатів, у подальших розрахунках необхідно скористатися одним із методів усунення автокореляції в залишках моделі.

Завдання. Використовуючи власний інформаційний простір (статистичні дані за обраною тематикою), необхідно перевірити модель на наявність автокореляції та, у разі наявності, її усунити.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. Для дослідження економічного розвитку країни в період 2018 – 2020 рр. обрано показник ВВП (X_1) (табл. 7). Проведіть аналіз нелінійного циклічного розвитку економіки та визначити основні структурні характеристики такого розвитку. Зробіть економічні висновки щодо значущості короткострокових і довгострокових циклів у розвитку національної економіки.

Вихідні дані

Період, місяць	X1, млрд грн	X2, млрд грн	Період, місяць	X1, млрд грн	X2, млрд грн
1	29,8	10,2	19	64,2	48,8
2	30,6	19,4	20	65,4	50,6
3	32,6	32,4	21	73,0	50,1
4	35,6	36,4	22	68,6	51,8
5	37,9	35,0	23	68,9	54,7
6	39,5	36,9	24	72,1	57,2
7	47,6	38,2	25	57,7	35,4
8	47,1	39,9	26	60,8	18,7
9	49,8	40,5	27	71,2	58,0
10	50,1	40,9	28	76,8	65,4
11	52,3	40,7	29	81,5	68,4
12	32,3	42,6	30	81,4	71,6
13	44,1	20,5	31	100,8	74,3
14	43,3	20,4	32	97,8	76,9
15	44,9	41,0	33	86,8	72,5
16	48,4	48,1	34	85,1	70,0
17	51,9	46,8	35	75,4	58,8
18	57,4	47,5	36	80,8	49,0

Завдання 2. Необхідно провести розкладання динамічного ряду курсу гривні стосовно долару США (табл. 8) за допомогою аналізу Фур'є.

Вихідні дані

Період	Курс гривні за 100 дол. США	Період	Курс гривні за 100 дол. США	Період	Курс гривні за 100 дол. США
01.2009	787,6	11.2009	798,78	09.2010	793,32
02.2009	810,81	12.2009	799,21	10.2010	795,83
03.2009	797,43	01.2010	808,25	11.2010	796,44
04.2009	794,43	02.2010	805,34	12.2010	797,47
05.2009	763,62	03.2010	796,06	01.2011	799,11
06.2009	768,09	04.2010	789,93	02.2011	795,84
07.2009	785,54	05.2010	791,54	03.2011	795,27
08.2009	815,28	06.2010	791,87	04.2011	796,93
09.2009	815,27	07.2010	788,67	–	–
10.2009	808,79	08.2010	788,79	–	–

Запитання для самоконтролю

1. Що таке автокореляція? Які наслідки вона несе? Перелічіть основні причини, що викликають появу автокореляції.
2. Як використовується метод Дарбіна – Уотсона для дослідження автокореляції?
3. Які критерії визначення автокореляції в ряду залишків вам відомі?
4. Назвіть основні етапи проведення методу Ейткена?
5. У чому полягає основна ідея ряду Фур'є?
6. Яку сутність розкривають цикли Кузнеця?
7. Розкрийте сутність спектрального кросс-аналізу.
8. Чим відрізняються цикли Жугляра та цикли Кондратьєва?
9. Назвіть види економічних циклів.
10. Назвіть основні характеристики гармоніки. Як визначаються значущі гармоніки?

Тема 11. Моделі адаптивного прогнозування та інтегрована модель авторегресії

Лабораторна робота 7. Прогнозування часових рядів за допомогою експоненціального згладжування

Мета: набуття навичок використання методу екстраполяції в рядах динаміки, побудови й аналізу моделей за допомогою пакета *Statistica*.

Завдання: засобами пакета *Statistica* необхідно знайти рівні динамічного ряду, за допомогою методу екстраполяції побудувати прогноз та знайти, в яких межах буде знаходитись прогнозне значення.

Методичні рекомендації

Завдання 1. Використовуючи дані про індекси акцій на біржі Лондона (табл. 9) необхідно:

1) навести графік динаміки показника та проаналізувати характер зміни значень показника;

2) провести згладжування часового ряду методом експоненціального згладжування за різних значень параметрів (досить взяти дві різні

комбінації); навести графіки згладжених даних і відповідних прогнозних значень показника;

3) оцінити якість моделей часових рядів (середня помилка, середня абсолютна помилка, середньоквадратичне відхилення помилок, середня процентна помилка, середня абсолютна процентна помилка); виконати порівняльний аналіз моделей і визначити найбільш адекватну з них;

4) навести економічну інтерпретацію моделей.

Таблиця 9

Вихідні дані

Рік	Квартал	Індекс	Рік	Квартал	Індекс
1966	1	349,3	1969	1	491,0
	2	359,7		2	433,0
	3	320,0		3	378,0
	4	299,9		4	382,6
1967	1	318,5	1970	1	403,4
	2	343,1		2	354,7
	3	360,8		3	343,0
	4	397,8		4	345,4
1968	1	409,1	1971	1	330,4
	2	401,1		2	372,8
	3	491,4		3	409,2
	4	490,5		4	427,6

1. Створюємо файл з вихідними даними в пакеті Statistica. Для цього вказуємо кількість змінних та об'єктів дослідження (випадків) відповідно до вихідних даних (рис. 44).

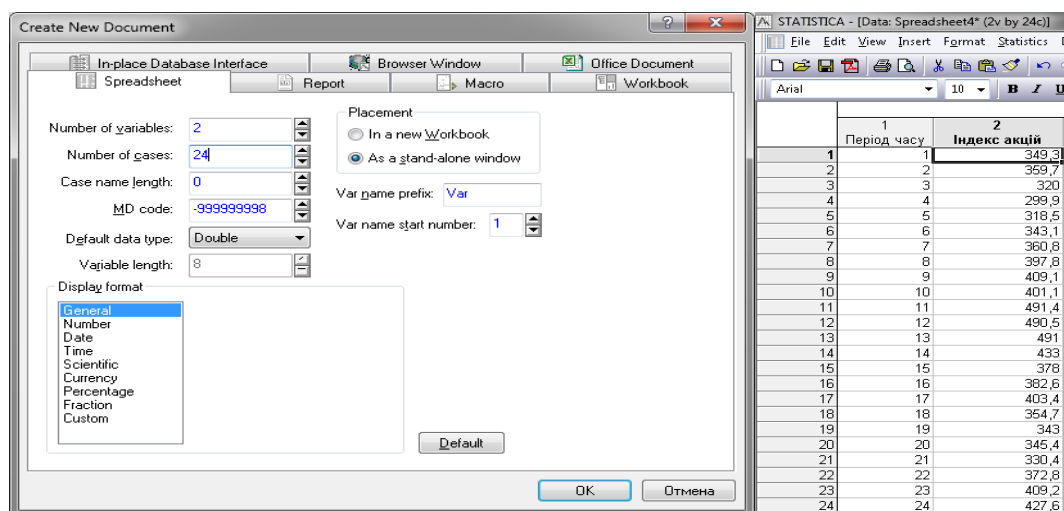


Рис. 44. Створення файлу з вихідними даними в пакеті Statistica

2. Будуємо графік вихідних даних. Для побудови графіка вибираємо в пункті меню *Graphs* підпункт *Scatterplots* (точкові графіки). Під час завдання характеристик графіка вибираємо *Graph type* – *Regular* (простий), заходимо у вкладку *Advanced* (розширені) та вибираємо тип тренду – *Polynomial* (поліноміальний), як змінні, що відображаються на графіку, вибираємо по осі *X* – період часу, а по осі *Y* – індекс акцій. Результат наведено на рис. 45.

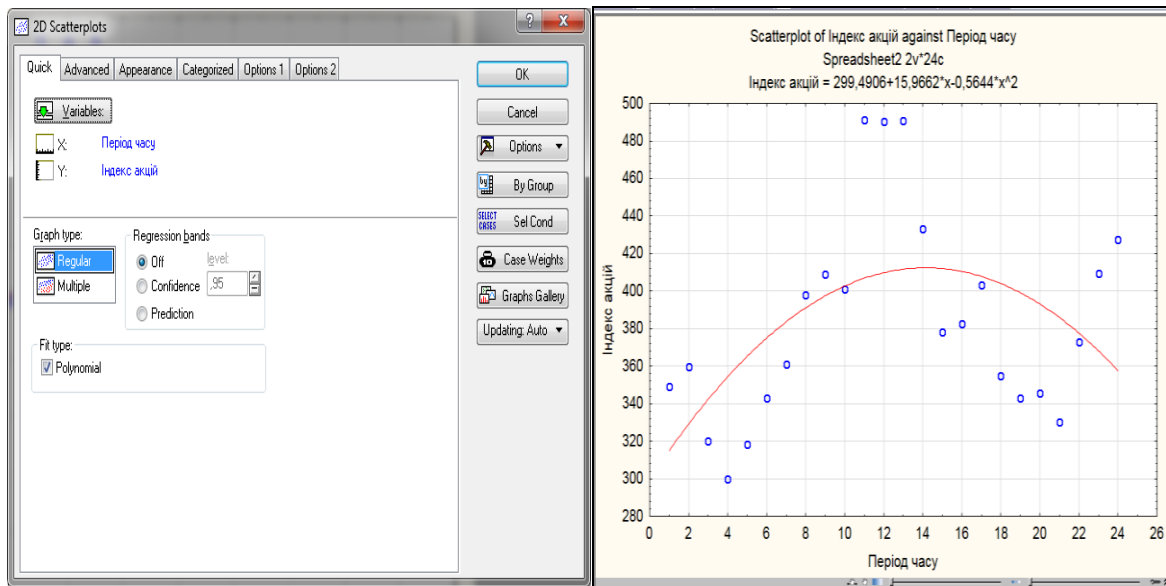


Рис. 45. Графік вхідних даних

Визначення двох основних моделей декомпозиції часового ряду: адитивної чи мультиплікативної, здійснюється за допомогою графічного аналізу. Існує таке правило визначення певної моделі: якщо початкові дані мають амплітуду коливань значень, яка постійно збільшується чи зменшується, то доцільно використовувати мультиплікативну модель декомпозиції. У випадку постійної амплітуди зміни значень доцільно використовувати адитивну модель.

У нашому випадку графічний аналіз рис. 45 дозволяє зробити висновки про адитивний характер залежності між компонентами, тобто використовуємо адитивну модель наступного виду:

$$Y = T + S + C + I, \quad (10)$$

де *T* – трендова компонента;

S – сезонна компонента;

C – циклічна компонента;

I – випадкова компонента.

3. Аналіз часових рядів здійснюється в модулі *Statistics / Advanced Linear / Nonlinear Models / Time Series / Forecasting* (поглиблені лінійні / нелінійні моделі / часові ряди / прогнозування).

Спочатку необхідно провести експоненціальне згладжування вихідних даних. Для цього вибираємо вкладку *Exponential smoothing & forecasting* (експоненціальне згладжування та прогнозування).

У наступному вікні необхідно задати параметри експоненціального згладжування. Так, змінна для аналізу – це індекс, лаг сезонної компоненти (*Seasonal component*) – 4 (так як дані наведено у квартальному розрізі). Вибір виду моделі ґрунтується на попередньому візуальному аналізі графіка вихідних даних. Також вказуємо період прогнозування – 4 (будуємо прогноз на 1 год (4 квартали) вперед) (рис. 46).

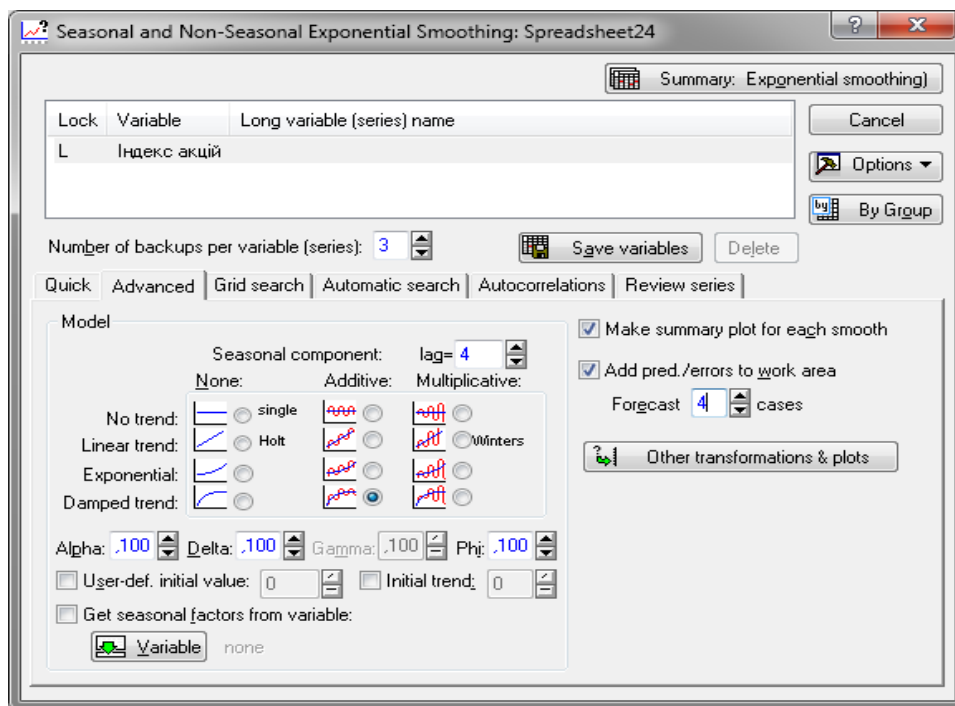


Рис. 46. Встановлення параметрів експоненціального згладжування

У вкладці *Grid search* ("Пошук в сітці") виконуємо процес пошуку та налаштування гіперпараметрів (параметрів згладжування) з метою визначення оптимальних значень для цієї моделі. Це важливо, оскільки продуктивність усієї моделі базується на заданих значеннях гіперпараметрів. Для цього натискаємо кнопку *Perform grid search* (виконати пошук у сітці) (рис. 47).

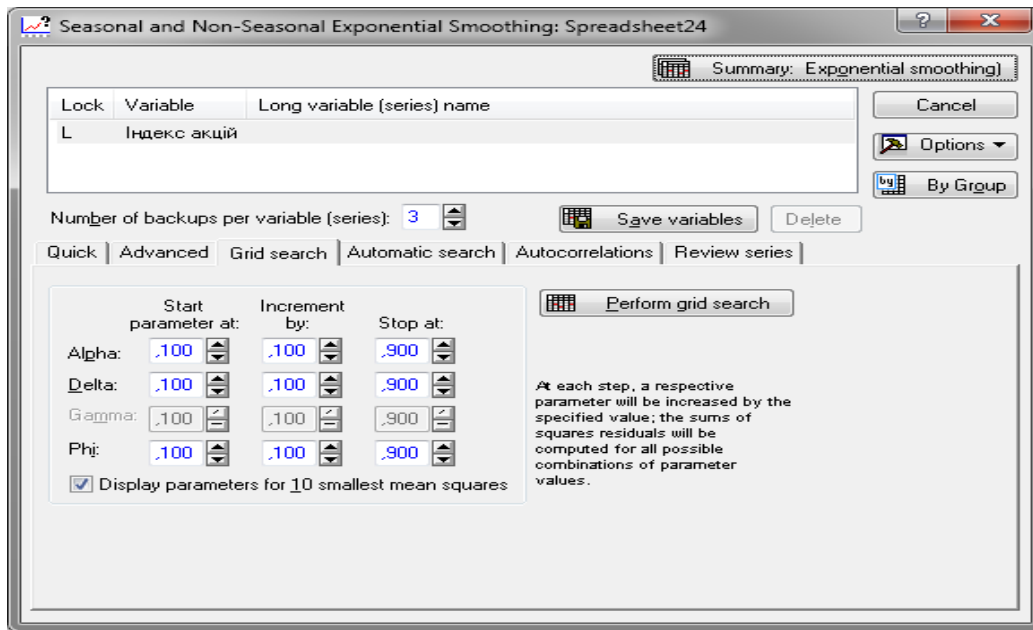


Рис. 47. Визначення гіперпараметрів згладжування

Унаслідок отримали вікно параметрів для заданої моделі (рис. 48).

Parameter grid search (Smallest abs. errors are highlighted) (Spreadsheet24)									
Model: Damped trend, add.season (4); S0=326,4 T0=2,932									
Індекс акцій									
Model Number	Alpha	Delta	Phi	Mean Error	Mean Abs Error	Sums of Squares	Mean Squares	Mean % Error	Mean Abs % Error
570	0,800000	0,100000	0,300000	3,299442	25,08659	25295,64	1053,985	0,692602	6,548562
650	0,900000	0,100000	0,200000	3,757726	25,00429	25361,69	1056,737	0,800064	6,534908
659	0,900000	0,200000	0,200000	3,756257	25,00527	25384,47	1057,686	0,799849	6,534639
579	0,800000	0,200000	0,300000	3,293398	25,08195	25391,64	1057,985	0,691700	6,545055
668	0,900000	0,300000	0,200000	3,754808	25,01024	25407,40	1058,642	0,799638	6,535480
677	0,900000	0,400000	0,200000	3,753381	25,01530	25430,47	1059,603	0,799432	6,536344
686	0,900000	0,500000	0,200000	3,751976	25,02044	25453,70	1060,571	0,799231	6,537231
489	0,700000	0,100000	0,300000	3,630105	25,60374	25460,31	1060,846	0,741194	6,710280
695	0,900000	0,600000	0,200000	3,750591	25,02567	25477,07	1061,544	0,799034	6,538142
588	0,800000	0,300000	0,300000	3,287720	25,07885	25490,09	1062,087	0,690880	6,541984

Рис. 48. Вибір параметрів експоненціального згладжування

Аналітик обирає ту комбінацію параметрів експоненціального згладжування, що має найменше значення середньої абсолютної процентної помилки ($MARE < 15\%$). Отже, в нашому випадку найбільш якісний прогноз буде у комбінації параметрів (коефіцієнтів) моделі № 659, а саме $\alpha = 0,9$; $\Delta = 0,2$; $\Phi = 0,2$.

На наступному кроці задаємо ці параметри у вкладці *Automatic search* (автоматичний пошук), а також вибираємо один із критеріїв вибору найкращих результатів (*Lack of fit indicator*) – *Mean squared error* (середньоквадратична помилка), *Mean absolute error* (середня абсолютна по-

милка), *Mean abs. perc. error* (середня абсолютна процентна помилка). Обираємо *Mean abs. perc. error* (рис. 49).

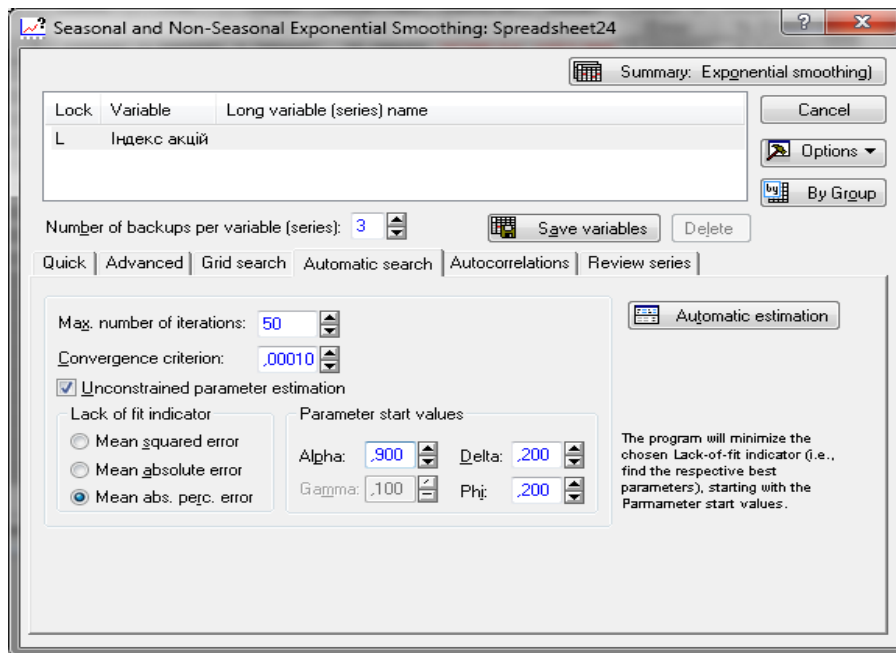


Рис. 49. Встановлення параметрів *Automatic search*

Після натиснення кнопки *Automatic estimation* отримуємо одночасно три групи результатів:

1) графіки вихідних даних, згладжених і прогнозних та залишків моделі (рис. 50);

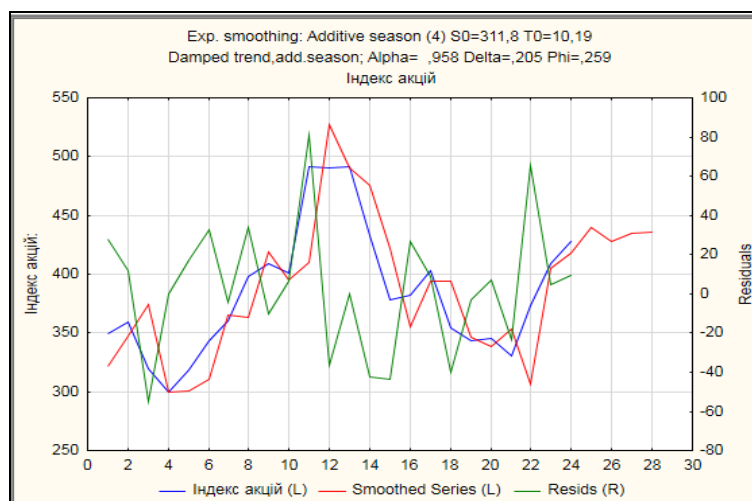


Рис. 50. Графіки результатів експоненціального згладжування

2) таблицю з вихідними та згладженими даними (*Smoothed Series*), залишками (*Resids*) і сезонними компонентами (*Seasonal Factors*) (рис. 51);

Exp. smoothing: Additive season (4) S0=311,8 T0=10,19 (Spreadsheet24) Damped trend, add. season; Alpha= ,958 Delta= ,205 Phi= ,259 Індекс акцій				
Case	Індекс акцій	Smoothed Series	Resids	Seasonal Factors
1	349,3000	321,4504	27,8496	6,96312
2	359,7000	347,6746	12,0254	-7,03687
3	320,0000	374,6985	-54,6985	-0,30437
4	299,9000	299,8667	0,0333	0,37812
5	318,5000	301,0928	17,4072	
6	343,1000	310,4026	32,6974	
7	360,8000	365,0728	-4,2728	
8	397,8000	363,6770	34,1230	
9	409,1000	419,2873	-10,1873	
10	401,1000	394,6943	6,4057	
11	491,4000	410,3744	81,0256	
12	490,5000	526,8013	-36,3013	
13	491,0000	490,8667	0,1333	
14	433,0000	475,4346	-42,4346	
15	378,0000	421,5013	-43,5013	
16	382,6000	355,6736	26,9264	
17	403,4000	394,5252	8,8748	
18	354,7000	394,5079	-39,8079	
19	343,0000	346,0399	-3,0399	
20	345,4000	338,4210	6,9790	
21	330,4000	353,5533	-23,1533	
22	372,8000	307,0607	65,7393	
23	409,2000	404,7095	4,4905	
24	427,6000	418,3382	9,2618	
25		440,2560		
26		427,8385		
27		434,9599		
28		435,7703		

Рис. 51. Результати побудови прогнозу

3) таблицю оцінок якості моделі (рис. 52).

Exp. smoothing: Additive season (4) S0=311,8 T0=10,19 (Spreadsheet24) Damped trend, add. season; Alpha= ,958 Delta= ,205 Phi= ,259 Індекс акцій	
Summary of error	Error
Mean error	3,1906462354
Mean absolute error	24,6403842229
Sums of squares	25597,8030050621
Mean square	1066,5751252109
Mean percentage error	0,6906325497
Mean abs. perc. error	6,4067905316

Рис. 52. Таблиця оцінок якості моделі експоненціального згладжування

Отже, за результатами побудованого прогнозу було отримано прогнозні значення індексу акцій на 4 періоди (1 рік) вперед. Отримані значення свідчать про зростання індексу акцій, що є позитивною тенденцією. Якість прогнозу підтверджується низькими значеннями помилок моделі (див. рис. 52).

Завдання 2. Використовуючи власний інформаційний простір (статистичні дані за обраною тематикою), необхідно побудувати прогноз методом експоненціального згладжування.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. Компанія, що здійснює операції з цінними паперами, зайнялася дослідженням ділової активності фондового ринку за показником обсягу торгів (у помісячному розрізі) (табл. 10). Компанія здійснює свою діяльність у короткостроковому періоді. Для прийняття ефективних управлінських рішень їй необхідно отримати якісний квартальний прогноз і зробити економічні висновки щодо майбутнього стану фондового ринку країни.

Таблиця 10

Вихідні дані

Місяць	Обсяг, млн грн	Місяць	Обсяг, млн грн	Місяць	Обсяг, млн грн	Місяць	Обсяг, млн грн
1	8,405 012	11	44,984 94	21	84,772 43	31	128,048 8
2	11,926 2	12	48,814 89	22	88,938 22	32	132,579 8
3	15,477 11	13	52,677 17	23	93,139 17	33	137,149 1
4	19,058	14	56,572 05	24	97,375 57	34	141,757
5	22,669 11	15	60,499 8	25	101,647 7	35	146,403 8
6	26,310 7	16	64,460 71	26	105,956	36	151,09
7	29,983 03	17	68,455 05	27	110,300 6	–	–
8	33,686 35	18	72,483 11	28	114,681 8	–	–
9	37,420 93	19	76,545 17	29	119,100 1	–	–
10	41,187 04	20	80,641 51	30	123,555 6	–	–

Завдання 2. Динаміка біржових цін акцій ПрАТ "Турбоатом" на торгах минулого тижня характеризується даними, наведеними в табл. 11.

Таблиця 11

Вихідні дані

День	1	2	3	4	5
Ціна акції, грн	122	118	119	121	117

Використовуючи метод експоненційного згладжування (за $\alpha = 0,2$), визначте прогнозний рівень біржової ціни акції на понеділок наступного тижня.

Запитання для самоконтролю

1. У чому полягає сутність методів згладжування?
2. Що таке ітераційні методи згладжування?
3. У чому різниця між експоненційним та простим згладжуванням?
4. Які можливості використання методів Брауна, Холта, Вінтерса?
5. Для чого використовується модель авторегресії?
6. Що таке тест Гренджера?
7. Яка існує класифікація векторних авторегресійних моделей?
8. Які існують способи поєднання складових часового ряду в модель?
9. Для чого використовується імпульсний аналіз?
10. Чим відрізняються ARMA та ARIMA моделі?

Змістовий модуль 4 Моделювання та прогнозування багатомірних процесів

Тема 12. Модель об'єкто-періодів

Лабораторна робота 8. Побудова моделі панельних даних

Мета: оволодіти навичками розроблення та оцінювання параметрів моделі панельних даних у пакеті *Eviews*.

Завдання: використовуючи можливості пакета *Eviews*, необхідно побудувати модель панельних даних для дослідження рівня розвитку країн світу за 4 роки за такими показниками: валовий зовнішній борг, прямі інвестиції, зайнятість, індекс споживчих цін, індекс цін виробників промислової продукції, витрати на кінцеве споживання, ВВП.

Методичні рекомендації

Панельні дані – це дані, які поєднують у собі *просторові дані* (*cross sectional date*) та *часові ряди* (*time series date*), тобто в кожний момент

часу спостерігаються просторові дані (*крос-секції*) за певними економічними об'єктами та для кожного такого об'єкта аналізуються один чи декілька часових рядів економічних показників. Особливістю моделей, побудованих на базі панельних даних, є можливість аналізу та вивчення як просторових даних, так і даних часового типу. Переваги таких моделей полягають у можливості врахування й аналізу індивідуальних особливостей між економічними одиницями, ефективного усунення неспостережної гетерогенності об'єктів на підставі коротких рядів, що неможливо виконати під час побудови стандартних регресійних моделей.

1. Спочатку необхідно створити файл з вихідними даними в середовищі Eviews (рис. 53).

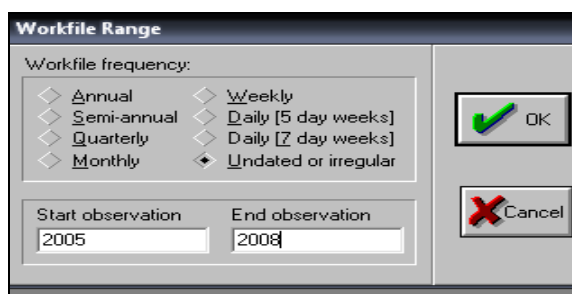


Рис. 53. Створення робочого аркушу в середовищі Eviews 7.0

2. Імпорт показників у середовище Eviews 7.0 (рис. 54).

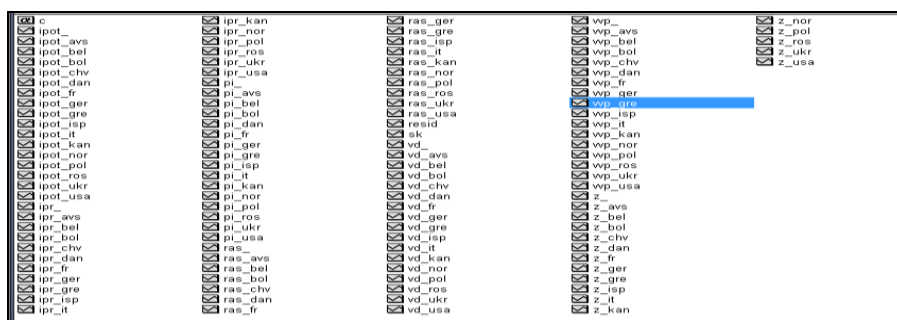


Рис. 54. Робоча область даних в середовищі Eviews 7.0

3. Заповнення об'єкта pool списком ідентифікаторів (рис. 55).

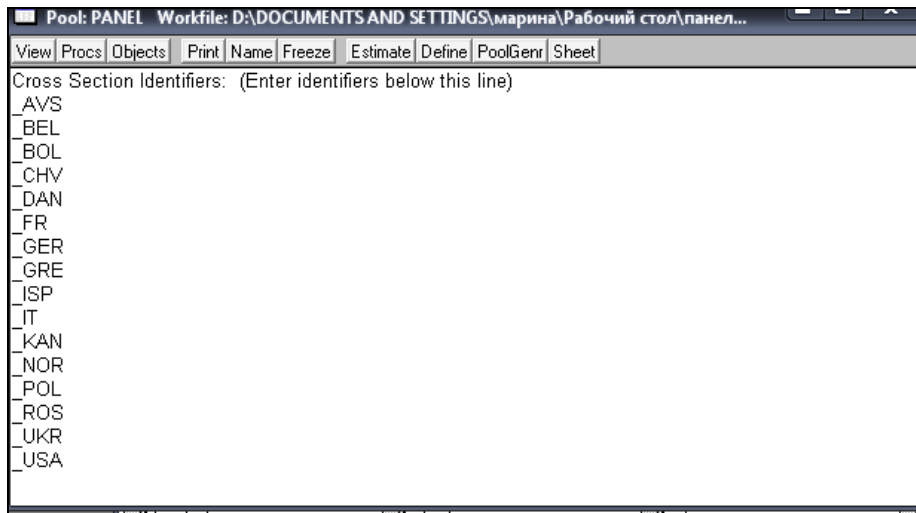


Рис. 55. Заповнення об'єкта pool списком ідентифікаторів

4. На наступному етапі виконуємо встановлення параметрів моделі панельних даних (рис. 56).

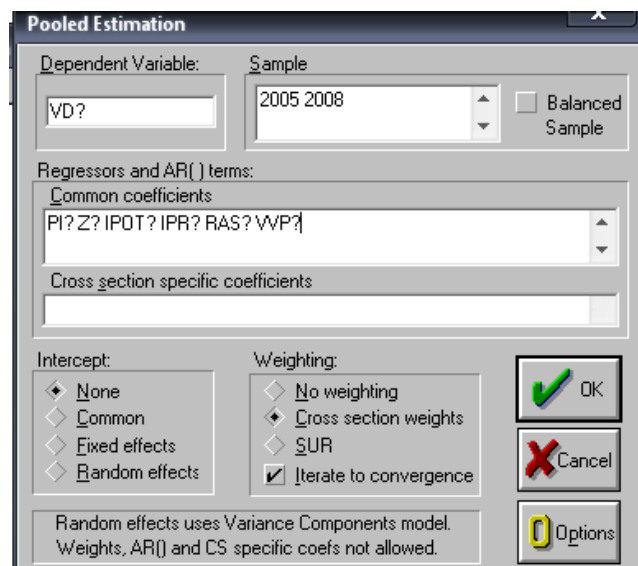


Рис. 56. Оцінка моделі панельних даних

Унаслідок проведених розрахунків у пакеті Eviews 7.0 формується звіт форми, наведеної на рис. 57.

Dependent Variable: VVD?				
Method: GLS (Cross Section Weights)				
Date: 11/30/09 Time: 15:49				
Sample: 2005 2008				
Included observations: 4				
Number of cross-sections used: 16				
Total panel (unbalanced) observations: 62				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
PI?	-20.43562	1.784699	-11.45046	0.0000
Z?	-16.97752	3.050551	-5.565396	0.0000
IPOP?	3172.621	3193.142	0.993573	0.3247
IPR?	-1013.755	3029.295	-0.334650	0.7391
RAS?	1.419541	0.348659	4.071437	0.0001
VVP?	-0.170088	0.334743	-0.508115	0.6134
Weighted Statistics				
R-squared	0.946815	Mean dependent var		1882934.
Adjusted R-squared	0.942067	S.D. dependent var		2018970.
S.E. of regression	485952.5	Sum squared resid		1.32E+13
F-statistic	199.3871	Durbin-Watson stat		1.256380
Prob(F-statistic)	0.000000			
Unweighted Statistics				
R-squared	0.972540	Mean dependent var		1799917.
Adjusted R-squared	0.970089	S.D. dependent var		3032398.
S.E. of regression	524450.7	Sum squared resid		1.54E+13
Durbin-Watson stat	1.179546			

Рис. 57. Результати розрахунків параметрів моделі панельних даних

Аналіз рис. 57 дозволяє зазначити, що найбільший вплив на ВВП країн мають прямі інвестиції та рівень зайнятості населення. Найкращою є модель панельних даних з фіксованими ефектами.

Завдання. Використовуючи власний інформаційний простір (просторові дані та часові ряди, що характеризують багатовимірний процес чи явище), необхідно побудувати регресійну модель на основі панельних даних.

Результати роботи оформити у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. За допомогою використання моделей на основі панельних даних побудуйте модель залежності рівня рентабельності підприємства від показників ліквідності (K_{pa} – коефіцієнт рентабельності всіх активів, що використовуються; K_n – коефіцієнт покриття; $K_{ал}$ – абсолютної ліквідності; $K_{шл}$ – коефіцієнт швидкої ліквідності; $K_{сджз}$ – коефіцієнт співвідношення дебіторської

та кредиторської заборгованості; $K_{\text{ЛПМЗ}}$ – коефіцієнт ліквідності при мобілізації засобів) для чотирнадцяти промислових підприємств Харківського регіону. Вихідні дані наведено у табл. 12.

Таблиця 12

Динаміка показників промислових підприємств Харківського регіону

Роки	Номер підприємства	$K_{\text{ра}}$	Показники оцінювання ліквідності				
			$K_{\text{п}}$	$K_{\text{шп}}$	$K_{\text{ал}}$	$K_{\text{сджз}}$	$K_{\text{ЛПМЗ}}$
1	2	3	4	5	6	7	8
2017	1	0,20	9,648	5,631	2,318	2,680	4,016
	2	0,00	2,374	1,353	0,207	1,812	1,021
	3	0,02	3,310	0,583	0,008	0,863	2,727
	4	-0,03	1,024	0,246	0,043	0,245	0,779
	5	0,05	15,845	2,922	0,166	0,898	12,923
	6	0,08	2,412	0,818	0,118	0,763	1,593
	7	-0,06	2,225	0,388	0,020	0,581	1,837
	8	0,11	1,098	0,574	0,028	0,504	0,524
	9	0,07	1,976	1,078	0,033	1,461	0,899
	10	-0,03	1,448	0,721	0,054	0,873	0,727
	11	0,04	1,073	0,221	0,004	0,259	0,852
	12	-0,08	2,077	0,822	0,004	0,779	1,255
	13	0,03	3,512	1,876	1,258	0,798	1,636
	14	-0,03	1,508	0,355	0,124	0,399	1,153
2018	1	0,21	10,649	6,383	3,143	3,672	4,265
	2	0,00	1,831	0,924	0,106	1,166	0,907
	3	0,00	3,450	0,691	0,042	0,803	2,759
	4	0,00	0,956	0,260	0,000	0,277	0,696
	5	-0,04	9,561	2,208	0,422	0,886	7,353
	6	0,17	3,260	1,454	0,286	1,020	1,806
	7	-0,01	2,484	0,790	0,001	0,611	1,694
	8	0,01	0,849	0,255	0,004	0,463	0,594
	9	0,00	2,130	0,622	0,029	1,601	1,508
	10	0,06	1,146	0,526	0,078	0,620	0,620
	11	0,00	0,970	0,200	0,008	0,268	0,770
	12	0,04	1,322	0,330	0,046	0,349	0,992
	13	0,02	4,224	2,595	1,525	1,092	1,629
	14	-0,05	1,930	0,280	0,001	0,249	1,650
2019	1	0,11	12,790	8,135	3,588	5,494	4,655
	2	0,00	1,641	0,915	0,184	0,825	0,727
	3	0,01	3,434	0,420	0,131	0,604	3,014
	4	-0,01	0,922	0,353	0,022	0,370	0,569

1	2	3	4	5	6	7	8
2019	5	0,05	11,881	4,037	1,065	1,530	7,845
	6	0,19	4,438	1,928	0,889	1,752	2,510
	7	-0,12	1,200	0,304	0,005	0,541	0,896
	8	0,09	1,043	0,478	0,009	0,461	0,566
	9	-0,04	1,517	0,789	0,101	1,009	0,727
	10	-0,10	0,958	0,571	0,026	0,823	0,387
	11	0,03	0,916	0,226	0,022	0,307	0,689
	12	0,13	1,570	0,720	0,156	0,375	0,849
	13	0,02	3,601	1,919	0,999	0,931	1,682
	14	0,00	1,624	0,610	0,002	0,605	1,015
2020	1	0,10	8,595	5,463	2,531	6,598	3,132
	2	0,10	2,153	1,267	0,053	1,183	0,886
	3	0,00	1,630	0,218	0,032	0,149	1,411
	4	0,05	1,005	0,454	0,024	0,484	0,550
	5	0,16	3,582	1,778	0,363	4,191	1,805
	6	0,12	3,720	2,212	1,235	2,360	1,508
	7	0,25	2,141	0,832	0,022	0,585	1,309
	8	0,09	0,792	0,437	0,108	0,492	0,355
	9	-0,01	2,413	1,046	0,379	0,852	1,368
	10	-0,16	0,733	0,325	0,005	0,598	0,408
	11	0,01	0,988	0,252	0,009	0,236	0,735
	12	-0,05	1,467	0,781	0,003	0,667	0,686
	13	0,03	2,508	1,007	0,320	0,728	1,501
	14	0,00	1,932	0,802	0,010	1,094	1,130

Завдання 2. Існує динаміка показників фірми: витрати на рекламу (табл. 13) та прибутку (табл. 14). Визначте, яку модель панельних даних доцільно використовувати. Побудуйте найкращу модель та оцініть її адекватність.

Таблиця 13

Значення витрат на рекламу

Номер фірми	Витрати на рекламу, тис. грн			
	2017	2018	2019	2020
1	41	63	98	143
2	20	29	35	42
3	47	73	80	89
4	13	15	18	21
5	34	44	62	67

Динаміка прибутку фірм

Номер фірми	Прибуток, тис. грн			
	2017	2018	2019	2020
1	496	625	958	1147
2	186	275	296	320
3	387	435	580	593
4	215	240	300	243
5	404	429	513	557

Запитання для самоконтролю

1. У яких випадках використовується модель панельних даних?
2. Коли доцільно використовувати модель з фіксованим, а коли з випадковим ефектами?
3. До чого приводить ситуація, коли кореляція відсутня, а застосовується модель з фіксованими ефектами?
4. До чого зводиться вибір між випадковими та детермінованими ефектами за статистичного підходу?
5. Які тести використовуються для перевірки якості панельних даних?
6. Яка існує ієрархія моделей панельних даних?
7. Які існують види панельних даних?
8. Для чого використовується критерій множників Лагранжа?
9. Які існують критерії перевірки адекватності моделей панельних даних?
10. У якому випадку використовують модель Хаусмана – Тейлора?

Тема 13. Моделювання причинних комплексів і рекурентні моделі

Лабораторна робота 9. Побудова систем структурних рівнянь

Мета: оволодіти навичками розробки систем одночасних рівнянь у середовищі Eviews.

Завдання: розробити систему структурних рівнянь, що дозволяє урахувати вплив внутрішніх і зовнішніх факторів на розвиток підприємства.

Методичні рекомендації

Показники, що відображають розвиток підприємства, та зміни, що відбуваються на національному і галузевому рівні наведено в табл. 15 – 17.

Таблиця 15

Внутрішні показники розвитку підприємства

Період	Обсяг реалізації продукції підприємства за рік	Коефіцієнт маневреності	Оборотність власного капіталу	Продуктивність праці	Рентабельність основних засобів
1	162 007	0,86	38,54	0,058 994	0,363 351
2	173 779,2	0,91	44,11	0,060 307	0,417 698
3	217 694	0,95	57,15	0,073 78	0,579 316
4	202 065	0,93	53,07	0,061 46	0,475 846
5	295 696,1	0,85	79,455 34	0,091 091	0,593 7
6	299 260,6	0,85	86,839 01	0,088 867	0,512 644
7	309 128,5	0,82	89,702 47	0,082 24	0,466 429
8	373 417	0,77	105,096 8	0,100 907	0,585 09
9	552 412	0,69	154,770 6	0,119 622	0,656 643
10	720 162	0,71	205,129 4	0,169 412	0,867 463

Таблиця 16

Показники машинобудівної галузі

Період	Обсяг реалізації продукції галузі за рік	Індекс цін за машинобудівною галуззю	Кредиторська заборгованість підприємств перед банками	Питома вага підприємств, що займалися інноваціями
1	20 078,3	118,1	15,8	2 936,10
2	21 523,1	104,7	15,3	4 255,95
3	24 492,2	102,8	15,6	6 305,25
4	35 133,1	105,8	11,5	10 175,25
5	53 569,3	113,7	10,8	13 286,85
6	59 668,1	105,6	9,2	21 512,70
7	68 730,6	106,0	10,9	36 783,90
8	98 339,9	111,9	11,5	640 29,45
9	121 664,9	122,0	10,8	110 101,50
10	74 240,7	107,7	10,7	108 116,0

Макроекономічні показники

Пе- ріод	ВВП, млн грн	Агрегований показник якості регуляторної політики за розрахунком Світового банку	Інвестиції в основний капітал (млн грн)	Курс національної валюти до євро	Показник структурного безробіття	Чисельність населення із се- редньодушовими загальними доходами у місяць, нижчими прожит- кового мінімуму
1	170 070	26,341 46	23 629	482,02	28,10	80,2
2	204 190	26,097 56	32 573	478,011 7	29,85	82,7
3	225 810	25,853 66	37 178	548,461 7	31,84	83,3
4	267 344	25,365 85	51 011	630,345	33,88	76,2
5	345 113	37,073 17	75 714	669,986 7	34,04	65,6
6	441 452	44,878 05	93 096	619,499 2	29,58	55,3
7	544 153	34,146 34	125 254	640,126 7	24,65	50,9
8	720 731	36,407 77	188 486	694,656 7	23,55	29,3
9	949 864	39,10	233 081	770,80	19,64	18,1
10	914 720	41,58	151 777	987,680 3	14,58	18,1

1. Специфікація моделі. Характер взаємозв'язку між елементами зовнішнього середовища і підприємством обумовлює вибір одночасної системи рівнянь. Виділяємо три рівні: підприємство, галузь, економіка країни, що в системі одночасних рівнянь відображаються окремими рівняннями. Система рівнянь має наступний загальний вигляд:

$$\begin{cases} y_1 = a_0^1 + a_1^1 x_1^1 + a_2^1 x_2^1 + a_3^1 x_3^1 + a_4^1 x_4^1 + a_5^1 x_5^1 \\ y_2 = a_0^2 + a_1^2 x_1^2 + a_2^2 x_2^2 + a_3^2 x_3^2 + a_4^2 y_1 \\ y_3 = a_0^3 + a_1^3 x_1^3 + a_2^3 x_2^3 + a_3^3 x_3^3 + a_4^3 x_4^3 + a_5^3 y_1 + a_6^3 y_2 \end{cases}, \quad (11)$$

де a – параметри рівняння.

Перше рівняння відображає особливості розвитку національної економіки:

y_1 – ВВП України за рік;

x_1^1 – агрегований показник якості регуляторної політики за розрахунком Світового банку;

x_2^1 – інвестиції в основний капітал (млн грн);

x_3^1 – курс національної валюти до Євро;

x_4^1 – показник структурного безробіття (частка структурних безробітних у загальній кількості безробітних);

x_5^1 – чисельність населення із середньодушовими загальними доходами у місяць, нижчими прожиткового мінімуму (% від загальної кількості населення).

Друге рівняння відображає особливості розвитку машинобудівної галузі промисловості:

y_2 – обсяг реалізації продукції галузі за рік;

x_1^2 – індекс цін по машинобудівній галузі;

x_2^2 – кредиторська заборгованість підприємств перед банками;

x_3^2 – питома вага підприємств, що займалися інноваціями.

Третє рівняння відображає розвиток промислового підприємства:

y_3 – обсяг реалізації продукції підприємства за рік;

x_1^3 – коефіцієнт маневреності;

x_2^3 – оборотність власного капіталу;

x_3^3 – продуктивність праці;

x_4^3 – рентабельність основних засобів.

2. Відповідно до запропонованої системи одночасних рівнянь для кожного рівняння необхідно оцінити ступінь його ідентифікованості. Розрахунки за умовою ідентифікації моделі наведено в табл. 18.

Таблиця 18

Перевірка ідентифікованості моделі

Рівняння	H	D	Ідентифікація моделі
$y_1 = a_0^1 + a_1^1 x_1^1 + a_2^1 x_2^1 + a_3^1 x_3^1 + a_4^1 x_4^1 + a_5^1 x_5^1$	5	3	$3 < 5$
$y_2 = a_0^2 + a_1^2 x_1^2 + a_2^2 x_2^2 + a_3^2 x_3^2 + a_4^2 y_1$	3	2	$2 < 3$
$y_3 = a_0^3 + a_1^3 x_1^3 + a_2^3 x_2^3 + a_3^3 x_3^3 + a_4^3 x_4^3 + a_5^3 y_1 + a_6^3 y_2$	4	1	$1 < 4$

З проведених розрахунків видно, що всі рівняння системи є зверх ідентифікованими.

3. За наведеною системою одночасних рівнянь необхідно провести розрахунки параметрів моделі з використанням двокрокового методу найменших квадратів (ДМНК), тому що всі рівняння системи є сверхідентифікованими. Розрахунки проводилися з використаннями можливостей пакета EViews у наступній послідовності:

3.1. Для створення нового файлу необхідно скористатися меню пакета *File / New / Workfile* (файл / новий / робочий файл (рис. 58)).

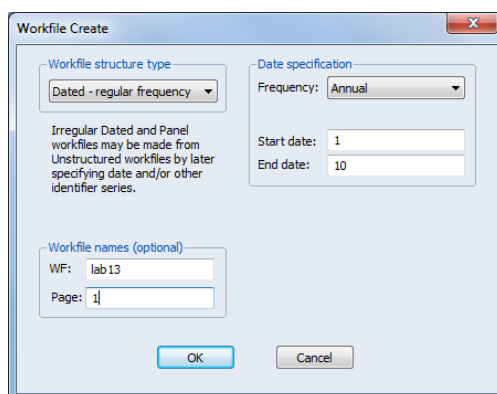


Рис. 58. Діалогове вікно модуля створення нового робочого файлу

У вікні необхідно вказати періодичність даних, а також початковий і кінцевий періоди аналізу. Для річних даних слід вказати номер початкового та кінцевого років.

Вікно нового робочого файлу має вигляд, наведений на рис. 59.

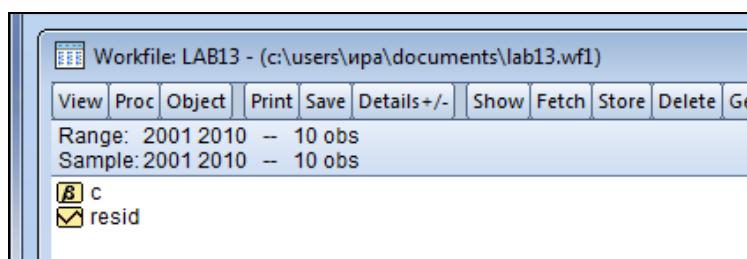


Рис. 59. Вікно нового робочого файлу пакета EViews

3.2. Для введення даних до програми можна обрати один з методів:

1. Введення в ручну (для цього слід обрати меню *Objects / New Object ...*, серед запропонованих варіантів обирається тип *Series* та вказується назва змінної).

2. Імпортування даних з інших джерел.

Найбільш зручним методом є імпортування даних з файлу *MS Excel*. Файл повинен мати коротку назву англійською мовою. Приклад оформлення даних наведено рис. 60.

	N	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1			y3	x31	x34	x32	x33	y2	x21	x22	x23	x11	x12	x13	x14	x15	y1		
2	2000	20880	0.641266	13.08476	0.476909	0.348984	20078.3	118.1	15.8	2936.1	26.34146	23629	482.02	28.1	80.2	170070			
3	2001	13808	0.761515	9.276453	0.205025	0.357723	21523.1	104.7	15.3	4255.95	26.09756	32573	478.0117	29.85	82.7	204190			
4	2002	21456	0.292676	15.08345	0.432811	0.337697	24492.2	102.8	15.6	6305.25	25.85366	37178	548.4617	31.84	83.3	225810			
5	2003	27982.5	0.434984	21.56012	0.506908	0.441508	35133.1	105.8	11.5	10175.25	25.36585	51011	630.345	33.88	76.2	267344			
6	2004	26502.3	0.447838	20.0499	0.437706	0.455031	53569.3	113.7	10.8	13286.85	37.07317	75714	669.9867	34.04	65.6	345113			
7	2005	32574	0.653528	36.96335	0.345411	0.545897	59668.1	105.6	9.2	21512.7	44.87805	93096	619.4992	29.58	55.3	441452			
8	2006	48015.5	1.099819	39.37461	0.4959	0.730434	68730.6	106	10.9	36783.9	34.14634	125254	640.1267	24.65	50.9	544153			
9	2007	54648.3	1.332743	35.37736	0.70078	0.893047	98339.9	111.9	11.5	64029.45	36.40777	188486	694.6567	23.55	29.3	720731			
10	2008	49622	1.845827	43.13081	0.575199	2.29915	121664.9	122	10.8	110101.5	39.1	233081	770.8	19.64	18.1	949864			
11	2009	47277	3.183263	69.40252	0.389778	2.593081	74240.7	107.7	10.7	108116	41.58	151777	987.6803	14.58	18.1	914720			

Рис. 60. Оформлення даних у файлі MS Excel, з якого проводиться імпорт

Для імпорту використовується команда меню *File/Import/Read Text-Lotus-Excel (файл / імпорт / відкрити файл Excel)* та обирається адрес файла. У вікні (рис. 61) вказується порядок розташування даних і перелік назв змінних, що мають бути імпортовані.

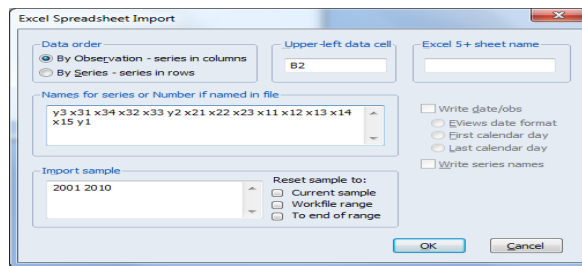


Рис. 61. Діалогове вікно імпорту даних

Унаслідок отримаємо робочий файл, у якому наведено всі вихідні дані моделі (рис. 62).

Series	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
X11		26.34146	26.09756	25.85366	25.36585	37.07317	44.87805	34.14634	39.10000	41.58000

Рис. 62. Вихідні дані моделі у пакеті EViews

3.3. Для створення системи структурних рівнянь необхідно скористатися меню Object / New object... / System (об'єкт / новий об'єкт / система) (рис. 63).

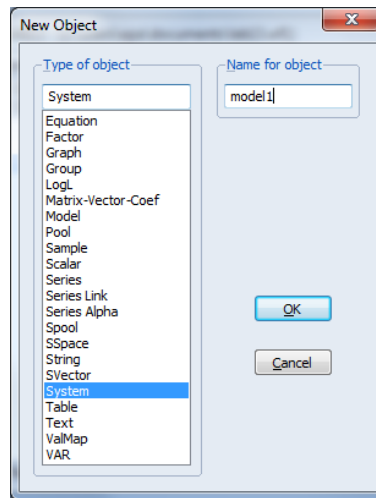


Рис. 63. Діалогове вікно створення нового об'єкта

3.4. Створення формалізованого вигляду моделі (рис. 64).

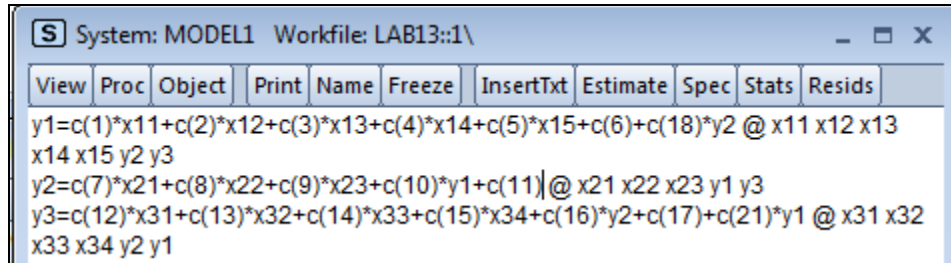


Рис. 64. Створення формалізованого вигляду моделі

Модель задається відповідно до позначень, що використовуються в пакеті. Процедура використання ДМНК потребує визначення інструментальних змінних. Вони наводяться після кожного з рівнянь системи після знаку @. У заданому випадку інструментальними змінними виступають усі змінні, що знаходяться у правій частині рівняння.

3.5. Для розрахунку параметрів моделі необхідно використати кнопку *Estimate* (оцінка). У запропонованому меню (рис. 65) проводиться вибір метода розрахунку параметрів моделі.

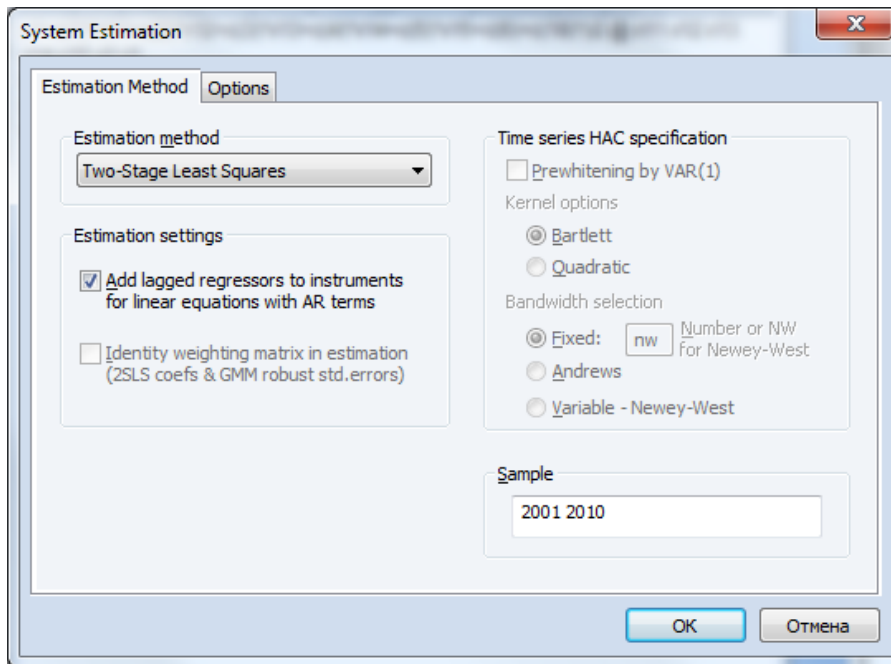


Рис. 65. Вибір методу розрахунку параметрів моделі

На рис. 66 наведено фрагмент отриманих результатів розрахунку параметрів моделі.

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	5291.976	2448.091	2.161674	0.0535
C(2)	4.668307	1.875869	2.488610	0.0301
C(3)	690.2081	135.7708	5.083626	0.0004
C(4)	-13762.70	4671.571	-2.946054	0.0133
C(5)	5307.429	3234.415	1.640924	0.1291
C(6)	-397198.3	310533.0	-1.279086	0.2272
C(18)	-2.615701	3.544241	-0.738014	0.4760
C(7)	1381.547	524.7315	2.632865	0.0233
C(8)	1093.377	2301.738	0.475023	0.6441
C(9)	-1.481605	0.557565	-2.657278	0.0223
C(10)	0.310120	0.087125	3.559494	0.0045
C(11)	-199754.7	74087.54	-2.696199	0.0208
C(12)	3440.590	7170.961	0.479795	0.6408
C(13)	32121.37	11266.52	2.851047	0.0158
C(14)	-12949.87	5689.165	-2.276234	0.0438
C(15)	294.6080	211.0850	1.395684	0.1903
C(16)	0.033572	0.322711	0.104031	0.9190
C(17)	-3726.787	4436.896	-0.839953	0.4188
C(21)	0.042731	0.066340	0.644120	0.5327

Рис. 66. Фрагмент розрахунку параметрів моделі

Унаслідок проведених розрахунків була отримана така система одночасових рівнянь:

$$\begin{cases} y_1 = -397\,198,3 + 5\,291,98x_1^1 + 4,67x_2^1 + 690,21x_3^1 - 13\,762,7x_4^1 + 5\,307,43x_5^1 - 2,61y_2 \\ y_2 = -199\,754,7 + 1\,381,55x_1^2 + 1\,093,38x_2^2 - 1,48x_3^2 + 0,31y_1 \\ y_3 = 39\,033,35 - 440,7x_1^3 - 350\,841,5x_2^3 + 244x_3^3 + 153\,330,5x_4^3 - 0,16y_1 + 1,09y_2 \end{cases}$$

Отримана система одночасових рівнянь дозволяє аналізувати стан розвитку підприємства, відповідно до тенденцій галузевого й економічного розвитку країни, та розробляти прогноз зміни обсягу реалізації продукції машинобудівного підприємства відповідно до змін у довкіллі.

Завдання. Використовуючи власний інформаційний простір, необхідно побудувати економетричну модель на основі системи одночасних структурних рівнянь.

Результати роботи оформити у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. Необхідно оцінити ідентифікованість моделі двома методами та дати рекомендації щодо методу пошуку параметрів моделі. Макроекономічна модель такого виду:

$$\begin{cases} y_1(t) = a_{10} + \beta_{11}y_3(t) + \varepsilon_1 \\ y_2(t) = a_{20} + \beta_{21}y_3(t) + \beta_{22}y_3(t-1) + \varepsilon_1, \\ y_3(t) = y_1(t) + y_2(t) + a_{31}x \end{cases}$$

де $y_1(t)$ – обсяг споживання;

$y_2(t)$ – обсяг інвестицій;

$y_3(t-1)$ – доходи попереднього періоду;

$y_3(t)$ – доходи;

x – державні витрати.

Завдання 2. Необхідно побудувати економетричну модель на основі системи одночасних структурних рівнянь, яка включає два регресійні рівняння: прибутку та інвестицій. Вихідні дані наведено в табл. 19.

Вихідні дані

Місяць	Прибуток, тис. грн	Інвестиції, тис. грн	Основні виробничі фонди, тис. грн	Фонд робочого часу (людино-дні)	Процентна ставка, %
1	44,85	71,30	25,30	120	23,00
2	47,15	74,75	28,75	125	21,85
3	43,70	65,55	19,55	114	25,30
4	48,30	75,90	31,05	131	23,00
5	50,60	79,35	32,20	133	24,15
6	56,35	66,70	23,00	127	26,45
7	50,60	82,80	36,80	137	20,70
8	51,75	80,50	34,50	133	20,125
9	55,20	86,25	39,10	131	19,55
10	58,65	90,85	40,25	138	18,40
11	56,35	88,55	37,95	143	17,825
12	62,10	94,30	42,55	137	16,10
13	63,25	92,00	42,55	148	18,40
14	65,55	86,25	44,85	148	17,825
15	64,40	95,45	43,70	152	16,10
16	62,10	93,15	41,40	150	17,25
17	67,85	100,05	46,00	143	14,95
18	70,15	105,80	48,30	154	13,80
19	71,30	109,25	49,45	158	12,65
20	73,60	111,55	48,30	160	11,50

Запитання для самоконтролю

1. Наведіть основні класи економетричних моделей.
2. Перелічіть етапи побудови системи рівнянь.
3. Перелічіть типи змінних, що входять до системи рівнянь.
4. Наведіть типи систем рівнянь.
5. У чому відмінність між структурною та зведеною формами моделі?
6. У чому полягає ідентифікація моделі?
7. Поясніть, чому система одночасових рівнянь завжди є надідентифікованою?
8. На які групи діляться методи оцінювання параметрів системи рівнянь?
9. Перелічіть етапи застосування непрямого методу найменших квадратів (НМНК).
10. Розкрийте особливість ДМНК.

Тема 14. Модель головних компонент

Лабораторна робота 10. Побудова моделі факторного аналізу

Мета: набуття навичок оброблення даних за допомогою методів факторного аналізу.

Завдання: виконати редукцію інформаційного простору дослідження, використовуючи методи факторного аналізу.

Методичні рекомендації

Факторний аналіз є одним з методів моделювання багатомірних процесів. Він дозволяє виявляти об'єктивно наявні фактори, які впливають на поведінку об'єкта, що досліджується. Ці фактори переважно неможливо визначити безпосередньо, але вони чинять вплив на досліджувані параметри.

Фактор – це скрита змінна, що здійснює вплив на досліджувану характеристику. Фактор є розрахунковою змінною, тобто якоюсь новою характеристикою об'єктів, що вивчаються. Кожен фактор може складатися з ряду елементів, які деталізують його вплив і виступають як самостійні фактори з більшим або меншим ступенем впливу на результативний показник. Чим більша їхня деталізація, тим детальніше вони вивчаються, що дозволяє визначати скриті зв'язки всередині соціально-економічного явища.

Модуль *Factor Analysis* (факторний аналіз) у пакеті *Statistica* містить широкий набір методів, за допомогою яких можливо проводити виділення факторів, тим самим виконувати редукцію (скорочення) вихідного інформаційного простору.

Розглянемо основні етапи проведення факторного аналізу в пакеті *Statistica* на прикладі побудови моделі факторного аналізу, що дозволить виявити показники, що мають найбільш суттєвий вплив на демографічну ситуацію в країні N (табл. 20).

Вихідні дані для проведення факторного аналізу

Роки	Кількість зайнятих (X1)	Рівень безробіття (X2)	Кількість постійного населення (X3)	Природний приріст (скорочення) населення (X4)	Доходи населення (X5)	Середньо-місячна номінальна заробітна плата (X6)
2007	18 570	12,6	48,3	-397,5	10 270	153,9
2008	19 870	12,3	48,5	-385,6	11 480	177,2
2009	20 180	11,6	48,7	-373,0	128 700	230,0
2010	20 170	10,8	48,1	-371,0	169 000	311,8
2011	20 090	9,6	47,8	-364,2	185 100	376,0
2012	20 160	9,1	47,4	-356,8	215 700	462,0
2013	20 300	8,6	47,1	-334,0	274 200	590,0
2014	20 680	7,2	46,7	-355,9	381 400	806,0
2015	20 730	6,8	46,5	-297,7	472 100	1 041,0
2016	20 910	6,4	46,2	-290,2	615 000	1 351,0
2017	20 970	6,4	46,0	-243,9	856 600	1 806,0
2018	20 190	8,8	45,8	-194,2	897 700	1 906,0
2019	20 270	8,1	45,6	-200,5	1 101 000	2 239,0
2020	20 445	7,6	45,1	-198,3	1 112 056	2 356,0

Для запуску модуля факторного аналізу слід використовувати вкладку *Statistics / Multivariate Exploratory Techniques / Factor Analysis* (статистика / багатоваріантні методи дослідження / факторний аналіз). На екрані з'явиться діалогове вікно *Factor Analysis* (рис. 67).

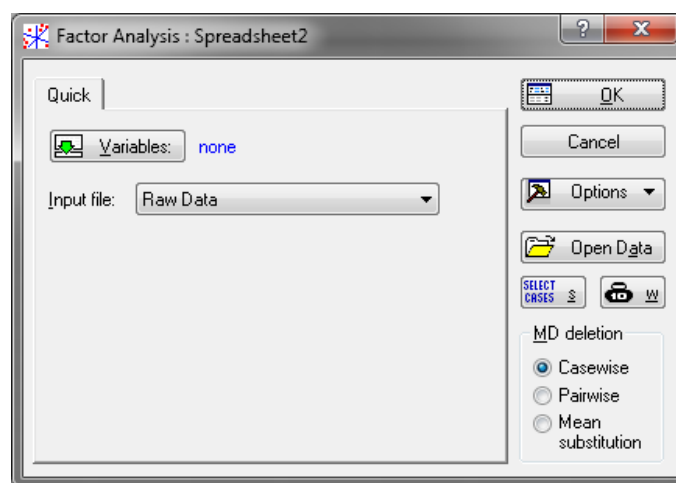


Рис. 67. Діалогове вікно факторного аналізу

Кнопка *Variables* (змінні) дозволяє відібрати всі змінні з файла вихідних даних, які мають бути включені в факторний аналіз (рис. 68). Якщо під час аналізу будуть використовувати всі змінні, то можна скористатися кнопкою *Select All* (виділити все).

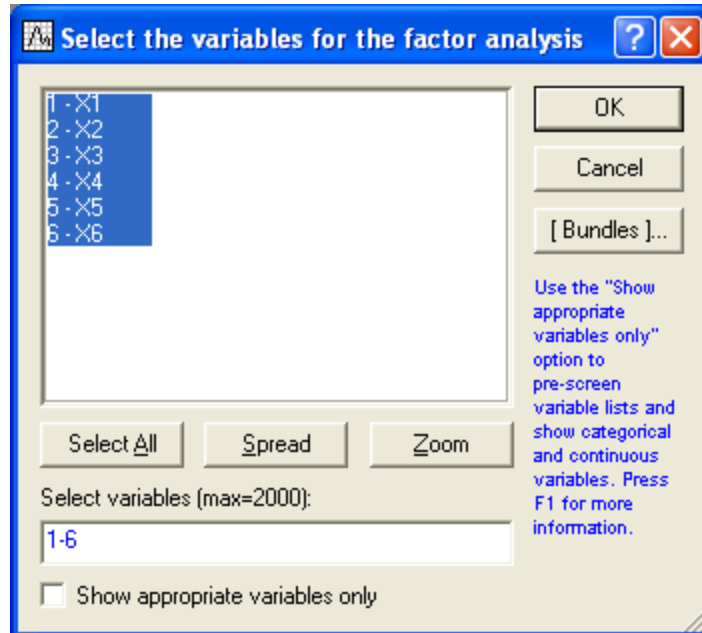


Рис. 68. Вікно вибору змінних

У модулі можливі такі типи вихідних даних:

Correlation Matrix (спосіб представлення даних – кореляційна матриця);

Raw Data (звичайний файл даних, де за рядками записані значення змінних);

MD deletion (спосіб оброблення пропущених значень – заміна пропущених змінних);

Casewise (спосіб виключення пропущених випадків полягає в тому, що в електронній таблиці, яка містить дані, ігноруються всі рядки (випадки), у яких є хоча б одне пропущене значення. У таблиці залишаються тільки випадки, в яких немає жодного пропуску);

Pairwise (парний спосіб виключення пропущених значень) – ігноруються пропущені випадки не для всіх змінних, а лише для вибраної пари. Усі випадки, в яких немає пропусків, використовуються в обробленні, наприклад, під час поелементного обчислення кореляційної матриці, коли послідовно розглядаються всі пари змінних. Очевидно, в способі *Pairwise* залишається більше спостережень для оброблення, ніж у способі *Casewise*;

Mean Substitution (підстановка середнього замість пропущених значень).

Обираємо *Raw Data* та натискаємо в стартовому вікні модуля на кнопку *OK*, розпочинається аналіз обраних змінних. Пакет *Statistica* обробить пропущені значення вказаним способом та обчислить кореляційну матрицю і запропонує на вибір кілька методів факторного аналізу.

У наступному діалоговому вікні (рис. 69) можна отримати таку інформацію: у верхній частині повідомляється, що пропущені значення оброблені методом *Casewise*. Опрацьовано 14 випадків та 14 випадків прийняті для подальших обчислень. Кореляційну матрицю обчислено для шести змінних.

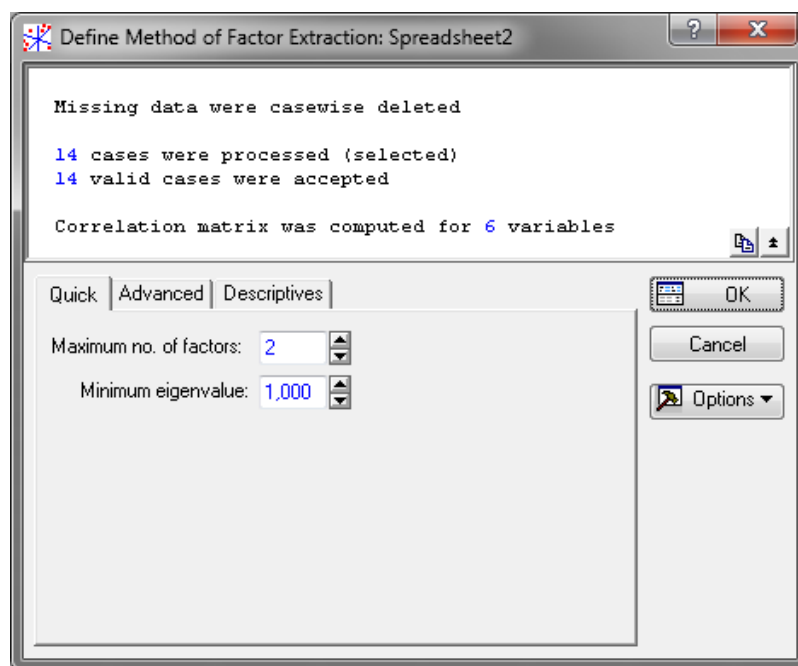
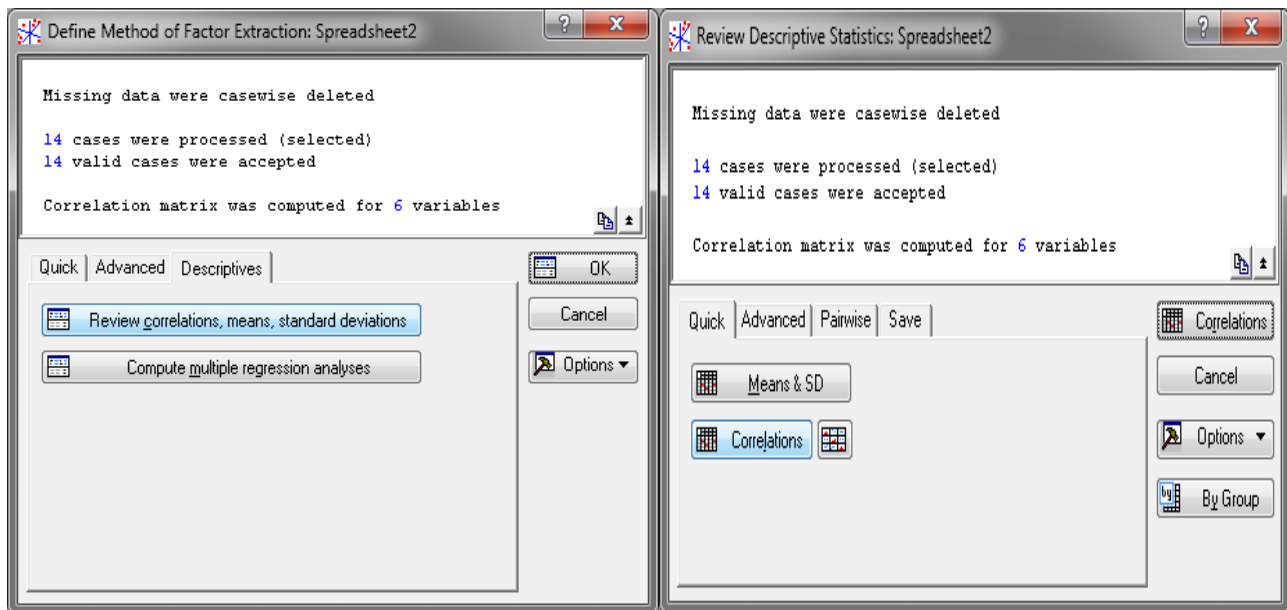


Рис. 69. Вікно встановлення параметрів факторного аналізу

Група опцій, об'єднаних під заголовком *Extraction method* (методи виділення факторів) – дозволяє вибрати метод оброблення.

Обчислення кореляційної матриці (якщо вона не задається відразу) – перший етап факторного аналізу. На вкладці *Descriptives* обираємо команду *Review correlations, means, standard deviations* (переглянути кореляції, середні значення, стандартні відхилення), потім вибираємо *Correlations* (кореляції). Також у цьому вікні можна подивитися середні, стандартні відхилення, кореляції, коваріації, побудувати різні графіки, провести додатковий аналіз поточних даних, перевірити відповідність вибіркового розподілу нормального закону розподілу та існування лінійної кореляції між змінними.

Унаслідок отримаємо кореляційну матрицю, що відображає тісноту зв'язку між досліджуваними показниками (рис. 70).



		Correlations (Spreadsheet2)					
		Casewise deletion of MD					
		N=14					
Variable	X1	X2	X3	X4	X5	X6	
X1	1,00	-0,83	-0,57	0,46	0,51	0,49	
X2	-0,83	1,00	0,83	-0,64	-0,69	-0,69	
X3	-0,57	0,83	1,00	-0,92	-0,95	-0,95	
X4	0,46	-0,64	-0,92	1,00	0,98	0,98	
X5	0,51	-0,69	-0,95	0,98	1,00	1,00	
X6	0,49	-0,69	-0,95	0,98	1,00	1,00	

Рис. 70. Побудова кореляційної матриці

Відповідно до результатів, наведених на рис. 70, між більшістю показників X1 – X6 існує сильний як прямий, так і зворотній зв'язок. Тобто доцільність проведення факторного аналізу підтверджено.

Для продовження аналізу повернемося до вікна *Define Method of Factor Extraction* (визначення методу виділення факторів) та оберемо метод факторного аналізу – *Principal components* (головні компоненти) і натиснемо кнопку *OK*. У цьому прикладі за замовчуванням приймається метод головних компонент, поле *Max. no. of factors* (максимальне число факторів) містить значення 2 і поле *Mini. eigenvalue* (мінімальні власні значення) містить 1 (рис. 71).

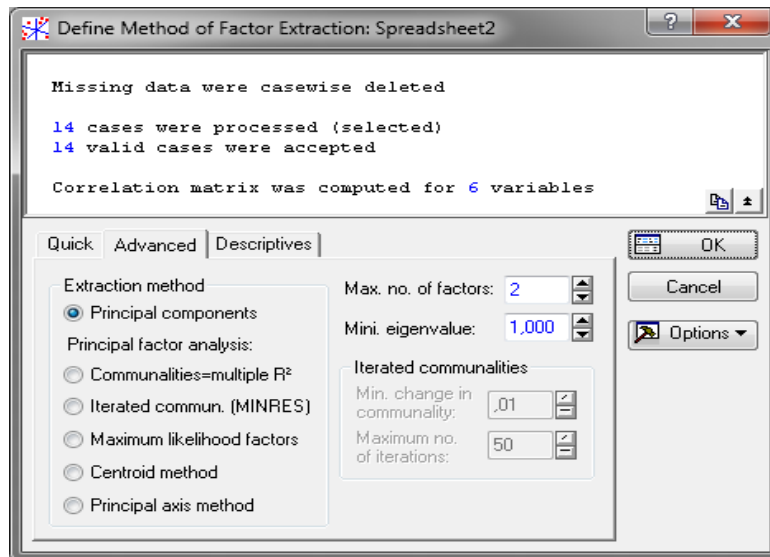


Рис. 71. Вибір параметрів факторного аналізу

Система швидко зробить обчислення та на екрані (рис. 72) з'явиться вікно *Factor Analysis Results* (результати факторного аналізу).

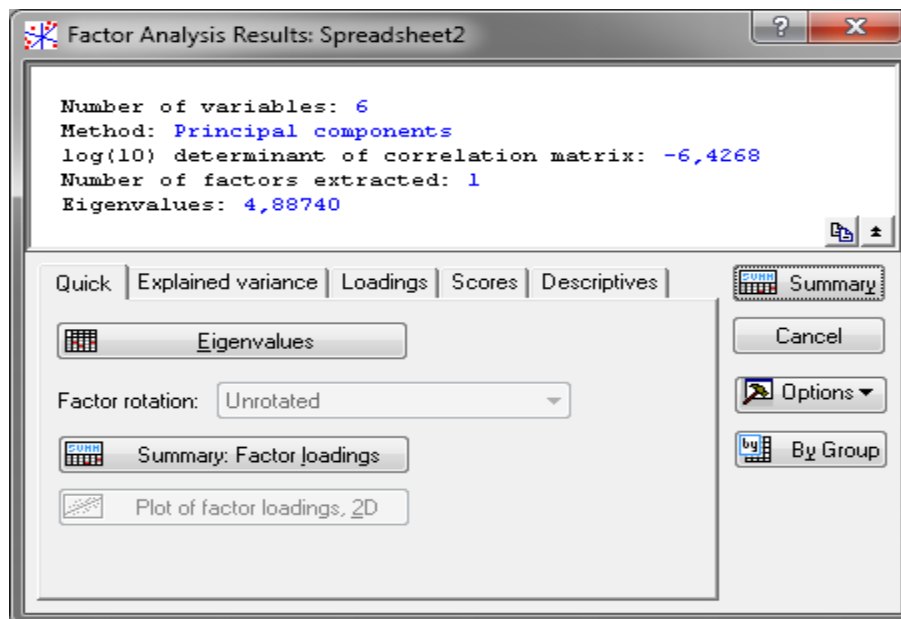


Рис. 72. Результати факторного аналізу

У верхній частині вікна результатів факторного аналізу дається інформаційне повідомлення:

Number of variables (кількість аналізованих змінних) – 6;

Method (метод аналізу) – головні компоненти;

log (10) determination of correlation matrix (десятковий логарифм детермінанта кореляційної матриці) – 6,426 8;

Number of Factor extraction (кількість виділених факторів) – 1;

Eigenvalues (власні значення) – 4,887 40.

Тобто, пакет *Statistica*, проаналізувавши вихідний інформаційний простір, пропонує виділити лише один фактор з власним значенням 4,89.

Доцільно зазначити, що метод головних компонент – це статистичний прийом, що дозволяє перетворювати вихідні змінні в їхню лінійну комбінацію. Мета методу – отримати скорочену систему вихідних даних, яка буде набагато простіша для розуміння та подальшого статистичного оброблення. Цей підхід був запропонований Пірсоном (1901 р.). Цей метод полягає в одночасному виділенні всіх головних компонент, кількість яких дорівнює кількості вихідних ознак. При цьому припускається повне пояснення дисперсії вихідних ознак через виділені латентні фактори.

Під час визначення кількості виділених факторів використовується такий критерій: виділяються тільки змінні з власними значеннями більше зазначеної константи (переважно одиниці). Дисперсії, що виділяються факторами, називаються власними значеннями.

Також як критерії для встановлення кількості факторів використовують: *критерій Кайзера*. Спочатку ви можете відібрати тільки фактори з власними значеннями, більшими 1. По суті, це означає, що якщо фактор не виділяє дисперсію, еквівалентну принаймні дисперсії однієї змінної, то він нівелюється. Цей критерій є, ймовірно, найбільш широко вживаним;

критерій кам'янистого осипу. Цей критерій є графічним методом, вперше запропонованим Кеттелом (1966 р.). Ви можете зобразити власні значення у вигляді простого графіка. Учений запропонував знайти таке місце на графіку, де спадання власних значень зліва направо максимально сповільнюється. Передбачається, що праворуч від цієї точки знаходиться тільки факторіальний осип. Оскільки, система пропонує виділення лише одного фактора, то використання зазначених раніше критеріїв не є доцільним. Однак, для прикладу розглянемо таблицю власних значень, що викликається натисненням кнопки *Eigenvalues* та може мати, наприклад, вигляд, наведений на рис. 73.

Данные: Собственные значения (Factor)*				
Значен.	Собственные значения (Factor) Выделение: Главные компоненты			
	Соб. зн.	% общей дисперс.	Кумулятивн. собст. знач.	Кумулятивн. %
1	6,118369	61,18369	6,11837	61,1837
2	1,800682	18,00682	7,91905	79,1905
3	0,472888	4,72888	8,39194	83,9194
4	0,407996	4,07996	8,79993	87,9993
5	0,317222	3,17222	9,11716	91,1716
6	0,293300	2,93300	9,41046	94,1046
7	0,195808	1,95808	9,60626	96,0626
8	0,170431	1,70431	9,77670	97,7670
9	0,137970	1,37970	9,91467	99,1467
10	0,085334	0,85334	10,00000	100,0000

Рис. 73. Приклад застосування критерію Кайзера

У другому стовбці *Власні значення* таблиці результатів ви можете знайти дисперсію нового, тільки що виділеного фактора. У третьому стовбці для кожного фактора приводиться процент від загальної дисперсії. Четвертий стовпець містить накопичену або кумулятивну дисперсію. Як видно з таблиці, власне значення для першого фактора дорівнює 6,118 369; тобто частка дисперсії, обумовлена першим фактором дорівнює приблизно 61,2 %.

Зауважимо, що сума всіх власних значень дорівнює 10 (кількості змінних, що аналізувалися в іншому прикладі). Другий фактор включає в себе близько 18 % дисперсії. Інші фактори містять не більше 5 % загальної дисперсії. Відповідно до критерію Кайзера необхідно залишити змінні з власними значеннями більше за 1. Оскільки, мета методу полягає в максимізації дисперсії (мінливості) "нової" змінної (фактора) та мінімізації розкиду навколо неї, то з наведеної на рис. 73 таблиці випливає, що доцільно обрати два фактори, що описують 79,2 % загальної дисперсії.

Якщо натиснути на кнопку *Scree plot* (ділянка осипу) у вкладці *Explained variance* (пояснення дисперсії), то також можемо отримати графік власних значень з метою застосування (перевірки) критерію кам'янистого осипу Кеттела (рис. 74).

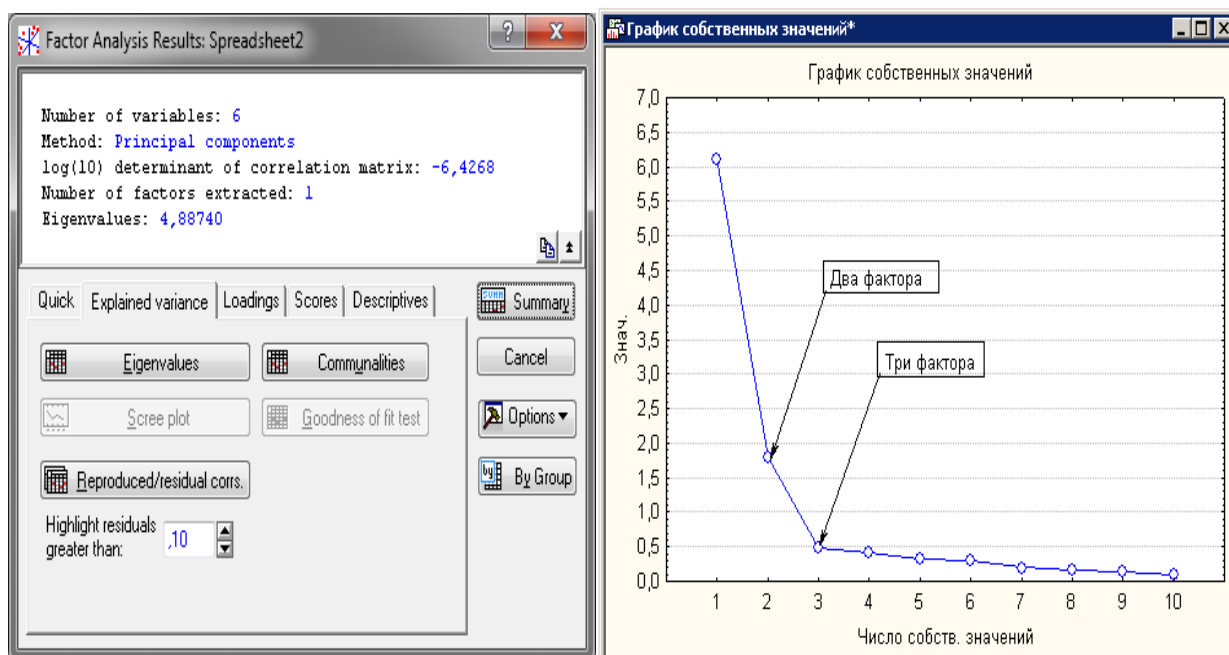


Рис. 74. Приклад застосування критерію кам'янистого осипу

На графіку обираємо таку точку, де безперервне падіння власних значень сповільнюється та після якої рівень інших власних значень відображає тільки випадковий "шум". На графіку, наведеному на рис. 73, ця точка може відповідати фактору 2 або 3. Отже, згідно до двох критеріїв краще виділити два фактори.

Повертаємося до нашого прикладу. Після натиснення кнопки *Eigenvalues* (власні значення) отримаємо власні значення для одного фактора (рис. 75).

Eigenvalues (Spreadsheet2)				
Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	4,887401	81,45669	4,887401	81,45669

Рис. 75. Результати виділення факторів

Тобто, виділення одного фактора дозволить описати 81,5 % загальної дисперсії.

У нижній частині вікна аналізу (рис. 72) знаходяться вкладки, що дозволяють всебічно ознайомитися з результатами факторного аналізу чисельно та графічно:

Plot of loadings, 2D i Plot of loadings, 3D (графіки навантажень) – ці опції побудують графіки факторних навантажень у проєкції на площину будь-яких двох обраних факторів і в проєкції в простір трьох обраних факторів, для чого необхідна наявність як мінімум трьох виділених факторів;

Summary. Factor loadings (підсумок. Факторні навантаження). Ця опція викликає таблицю з поточними факторними навантаженнями (рис. 76), тобто обчисленими для цього методу обертання факторів, який указаний праворуч від відповідної кнопки.

Factor Loadings (Unrotated) (Spreadsheet2)	
Extraction: Principal components (Marked loadings are >,700000)	
Variable	Factor 1
X1	-0,684396
X2	0,853163
X3	0,977557
X4	-0,939809
X5	-0,963751
X6	-0,960958
Expl. Var	4,887401
Prp. Totl	0,814567

Рис. 76. Таблиця факторних навантажень

Основою дослідження факторів є *факторна матриця* (матриця факторних навантажень), яка характеризує ступінь зв'язку між ознаками та факторами. Факторна матриця дозволяє виділити для кожного фактора групу параметрів, найтісніше з ним зв'язаних. Тим самим відкривається можливість зіставити фактори один з одним, дати їм змістовне тлумачення і найменування, тобто навести інтерпретацію факторів. Факторні навантаження можуть інтерпретуватися як кореляції між відповідними змінними і факторами: чим вище навантаження за модулем, тим більше близькість фактора до початкової змінної. Факторні навантаження становлять найбільш важливу інформацію для інтерпретації отриманих факторів. У згенерованій таблиці для полегшення трактування будуть виділені червоним кольором факторні навантаження за абсолютною величиною більше 0,7. Фактори частіше інтерпретують за навантаженням.

За результатами таблиці (див. рис. 76) видно, що фактор корелює з усіма змінними, окрім першої. У такому випадку вдаватися до повороту

осей недоцільно. Знайдене рішення можна інтерпретувати у такий спосіб: отриманий фактор найтісніше пов'язаний з X_2 , X_3 , X_4 , X_5 і X_6 . Результати можливо представити так: $F_1 = 0,853 X_2 + 0,978 X_3 - 0,939 X_4 - 0,964 X_5 - 0,96 X_6$.

Тобто, рівень безробіття, кількість постійного населення, природний приріст (скорочення) населення, доходи населення та середньомісячна номінальна заробітна плата необхідні для подальшого дослідження змінних, які можуть використовуватися для аналізу демографічної ситуації. X_1 (кількість зайнятих) необхідно виключити з аналізу.

У випадку виділення двох і більше факторів рекомендується застосувати метод *обертання факторного рішення*. Реальна орієнтація факторів у факторному просторі довільна, і будь-яке обертання факторів відтворює кореляції так само добре, як й інші обертання. Отже, здається природним повернути фактори у такий спосіб, щоб вибрати найпростішу для інтерпретації факторну структуру. Існують різні методи обертання факторів. Метою цих методів є отримання зрозумілою (легко інтерпретованою) матриці навантажень, тобто факторів, які ясно відзначені високими навантаженнями для деяких змінних і низькими – для інших. Результатом обертання є вторинна структура факторів. Типовими методами обертання є стратегії варімакс, квартімакс і еквімакс.

Найчастіше використовують процедуру *Varimax* (варімакс), яка максимізує варіацію квадратів факторних навантажень для кожної компоненти, збільшуючи великі та зменшуючи малі значення навантажень. Для вибору методу обертання на вкладці *Loadings* необхідно вибрати у полі *Factor rotation* метод *Varimax*. Також доцільність виділення двох і більше факторів підтверджується графічно. Для цього необхідно натиснути на кнопку *Plot of loadings 2D*.

Завдання. Використовуючи власну статистичну базу, необхідно виконати редукцію інформаційного простору за допомогою методу факторного аналізу.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. За результатами компонентного аналізу на семи показниках технічного стану підприємств виділено одну компоненту з дисперсією $\lambda = 5,2$ і властивим вектором $V = (0,90; 0,80; 0,74; 1,00; 0,82; 0,87; 0,96)$.

Визначте факторні навантаження кожного показника та оцініть адекватність моделі головних компонент.

Завдання 2. Фондовий ринок певної країни змінюється під впливом великої кількості макроекономічних факторів, що ускладнюється якісним аналізом усієї сукупності. Аналіз статистичних матеріалів дозволив сформувати перелік з дев'яти макроекономічних показників за 2013 – 2020 рр. (табл. 21), що впливають на фондовий ринок. Здійсніть агрегування факторів та дати їх економічну інтерпретацію.

Таблиця 21

Показники, що характеризують стан фондового ринку

Роки	Гроші поза банками, млн грн	Кредити для резидентів, млн грн	Облікова ставка, %	Імпорт, млн грн	Експорт, млн грн	Прямі інвестиції в Україну, млн дол. США	Обсяг прямих інвестицій з України, млн дол. США	ВВП, млн дол. США	Обсяг реалізованої продукції промисловості, млн дол. США	Індекси споживчих цін
2013	26 434	42 228,0	9,13	21 494	23 351	5 339,0	143,9	225 810	229 634	99,4
2014	33 119	67 892,1	7,00	27 665	28 953	6 657,6	163,5	267 344	289 117	108,2
2015	42 345	88 614,7	7,88	36 313	41 291	8 353,9	175,9	345 113	400 757	112,3
2016	60 231	143 423,0	9,00	43 707	44 378	16 375,2	218,2	441 452	468 563	110,3
2017	74 984	245 230,0	9,00	53 307	50 239	21 186,0	221,5	544 153	551 729	111,6
2018	111 119	426 867,0	10,00	71 877	64 001	29 489,4	6 196,1	720 731	717 077	116,6
2019	154 759	734 022,0	11,00	99 962	85 612	35 723,4	6 198,6	949 864	779 127	122,3
2020	147 890	723 295,0	11,08	56 275	54 253	40 026,8	6 223,3	914 720	668 956	112,3

Запитання для самоконтролю

1. У чому полягає суть методу факторного аналізу?
2. Що є основою для побудови факторної матриці?
3. Які існують етапи побудови методу головних компонент?
4. У чому сутність процедури варімакс?
5. Що таке матриця трансформації та від чого вона залежить?
6. Які існують методи факторного аналізу?
7. Що таке факторні навантаження?
8. У чому полягають попередні процедури факторного аналізу?
9. Які існують групи факторів?
10. Для чого використовується обернення факторів?

Тема 15. Кластерний аналіз

Лабораторна робота 11. Використання кластерного аналізу для дослідження економічних процесів

Мета: надбання навичок використання кластерного аналізу в пакеті *Statistica*.

Завдання: провести класифікацію (групування) країн світу за рівнем енергетичної безпеки, що оцінюється наступними показниками: частка власних джерел у балансі паливно-енергетичних ресурсів держави, % (X1); частка домінуючого паливного ресурсу у споживанні паливно-енергетичних ресурсів, % (X2); енергоємність ВВП, кг умовного палива / грн (X3); обсяг видобутку вугілля, млн тонн (X4); ступінь забезпечення паливно-енергетичними ресурсами (X5).

Вихідні значення за показниками наведено в табл. 22.

Таблиця 22

Вихідні дані

Країни	Показники енергетичної безпеки				
	X1	X2	X3	X4	X5
Австрія	0,512 517	8,139 704	28,086 43	0,657 682	1,229 883
Бельгія	0,441 011	11,184 19	46,610 1	0,423 147	1,198 028
Болгарія	0,390 566	67,683 11	473,796 5	0,728 101	0,629 035
Фінляндія	0,659 931	13,740 58	0	0,644 842	1,181 073
Франція	0,598 438	7,578 963	6,851 12	0,746 463	0,569 163
Германія	0,425 501	9,358 731	2922,118	0,671 615	0,998 937
Італія	0,422 215	9,180 896	4,465 834	0,526 936	1,333 184
Польща	0,447 059	27,105 71	6933,373	1,057 638	1,030 381
Іспанія	0,493 978	10,291 91	4295,369	0,617 82	0,923 633
Швеція	0,828 147	8,865 035	0	0,909 419	0,948 364
Швейцарія	0,694 735	4,621 731	0	0,755 09	0,952 061
Велика Британія	0,389 766	7,828 034	1005,633	1,954 056	2,160 322
Білорусь	0,603 121	80,723 23	0	0,521 061	1,342 974
Росія	0,562 118	121,362 5	3960,709	2,899 916	2,942 098
Україна	0,541 134	179,235 7	2740,338	0,676 241	1,076 723

Методичні рекомендації

Кластерний аналіз – це метод багатомірного статистичного дослідження, до якого належать збір даних, що містять інформацію про вибіркові об'єкти, та упорядкування їх у порівняно однорідні, схожі між собою групи. Мета кластерного аналізу полягає в пошуку наявних структур, що виражається в утворенні груп схожих між собою об'єктів – кластерів. На відміну від інших методів, цей вид аналізу дає можливість класифікувати об'єкти не за однією ознакою, а за декількома одночасно. Для цього вводяться відповідні показники, що характеризують певну міру близькості за всіма класифікаційними параметрами.

Кластер (cluster (англ.) – згусток, пучок) – це термін, що позначає групу об'єктів із однорідними ознаками. Кластерний аналіз вирішує завдання класифікації об'єктів за відсутності апріорної інформації про спостереження в середині класів.

Розглянемо реалізацію методу кластерного аналізу в пакеті *Statistica*.

1. Для побудови кластерних утворень нормуємо значення показників. З цією метою в контекстному меню необхідно обрати *Fill / Standardize Block / Standardize Columns* (заповнити / стандартизувати блок / стандартизувати стовбці) як показано на рис. 77.

	1	2	3	4	5
	X1	X2	X3	X4	X5
Austria	0,512517	8,139704	28,08643	0,657682	1,229883
Belgium	0,441011	11,18419	46,6101	0,423147	1,198028
Bulgaria	0,390566	67,68311	473,7965	0,728101	0,629
Finland	0,659931	13,74058	0	0,644842	1,181
France	0,598438	7,578963	6,85112	0,746463	0,569
Germany	0,425501	9,358731	2922,118	0,671615	0,99
Italy	0,422215	9,180896	4,465834	0,526936	1,333
Poland	0,447059	27,10571	6933,373	1,057638	1,03
Spain	0,493978	10,29191	4295,369	0,61782	0,923
Sweden	0,828147	8,865035	0	0,909419	0,948
Switzerland	0,694735	4,621731	0	0,75509	0,952
UK	0,389766	7,828034	1005,633	1,954056	2,16
Belorussia	0,603121	80,72323	0	0,6521061	1,342
Russia	0,562118	121,3625	3960,709	2,899916	2,942
Ukraine	0,641134	179,2357	2740,338	0,676241	1,076

	1	2	3	4	5
	X1	X2	X3	X4	X5
Austria	-0,17076	-0,56815	-0,67541	-0,39902	-0,00755
Belgium	-0,73874	-0,50982	-0,66688	-0,75668	-0,06092
Bulgaria	-1,13942	0,572676	-0,47012	-0,29163	-1,01414
Finland	1,000142	-0,46084	-0,68834	-0,4186	-0,08932
France	0,511704	-0,57889	-0,68519	-0,26363	-1,11444
Germany	-0,86193	-0,5448	0,657549	-0,37777	-0,39445
Italy	-0,88803	-0,5482	-0,68629	-0,5984	0,165506
Poland	-0,6907	-0,20477	2,505087	0,21091	-0,34177
Spain	-0,31802	-0,52692	1,290053	-0,45981	-0,5206
Sweden	2,336278	-0,55425	-0,68834	-0,01512	-0,47917
Switzerland	1,276589	-0,63555	-0,68834	-0,25047	-0,47298
UK	-1,14577	-0,57412	-0,22516	1,577936	1,551189
Belorussia	0,548901	0,822519	-0,68834	-0,60736	0,181907
Russia	0,223215	1,60115	1,135912	3,020361	2,860878
Ukraine	0,05654	2,709977	0,573823	-0,37072	-0,26414

Рис. 77. Нормування вхідних даних

2. Для проведення кластерного аналізу необхідно увійти до модуля *Cluster analysis* (кластерний аналіз). Для цього необхідно скористатися меню *Statistics / Multivariate Exploratory / Cluster analysis* (рис. 78).

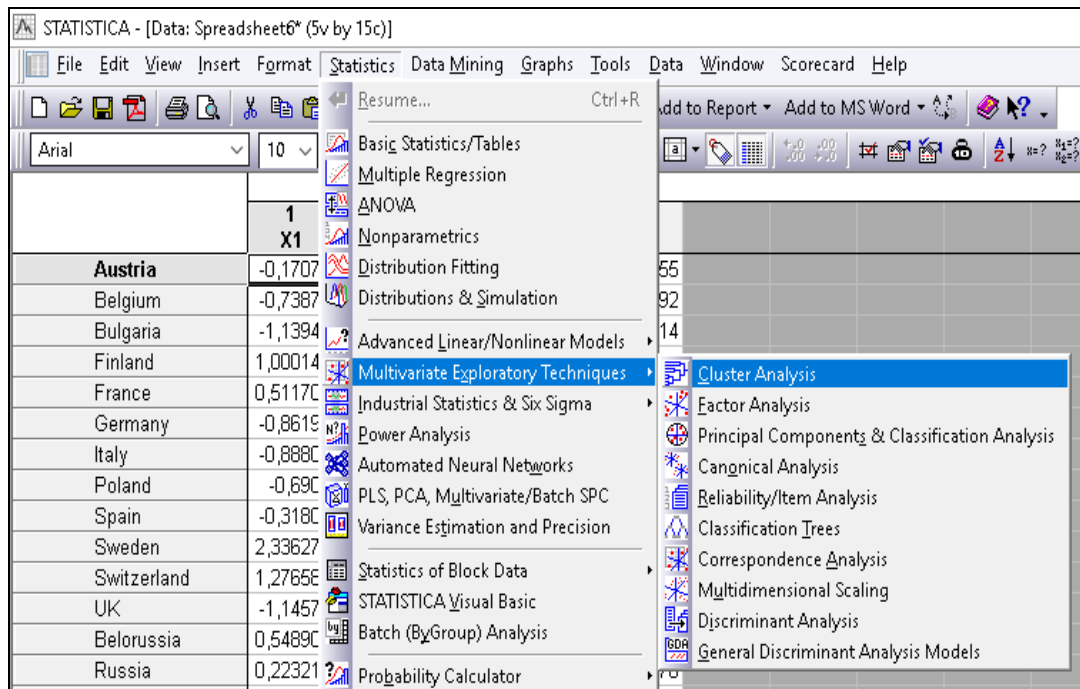


Рис. 78. Запуск модуля кластерного аналізу

Діалогове вікно (рис. 79), що з'явилося, дозволяє використати один з методів кластеризації:

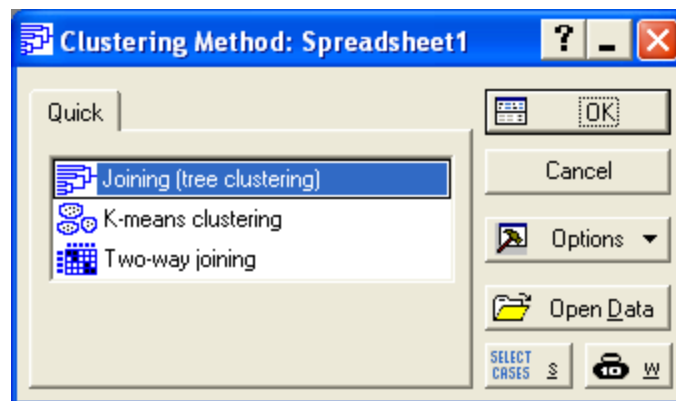


Рис. 79. Діалогове вікно модуля кластерний аналіз

1. *Joining (tree clustering)* – об'єднання (деревовидна кластеризація).
2. *K-means clustering* – кластеризація методом К-середніх.
3. *Two-way joining* – двохходове об'єднання.

Методи кластерного аналізу можна розділити на дві основні групи ієрархічні (природні) та неієрархічні (штучні). Рекомендується застосовувати по черзі обидва методи.

3. З метою визначення кількості кластерів доцільно провести природню (древовидну) кластеризацію. У пакеті *Statistica* цей тип кластеризації передбачає реалізацію декількох етапів:

3.1. Вибір показників, за якими проводиться кластеризація (рис. 80).

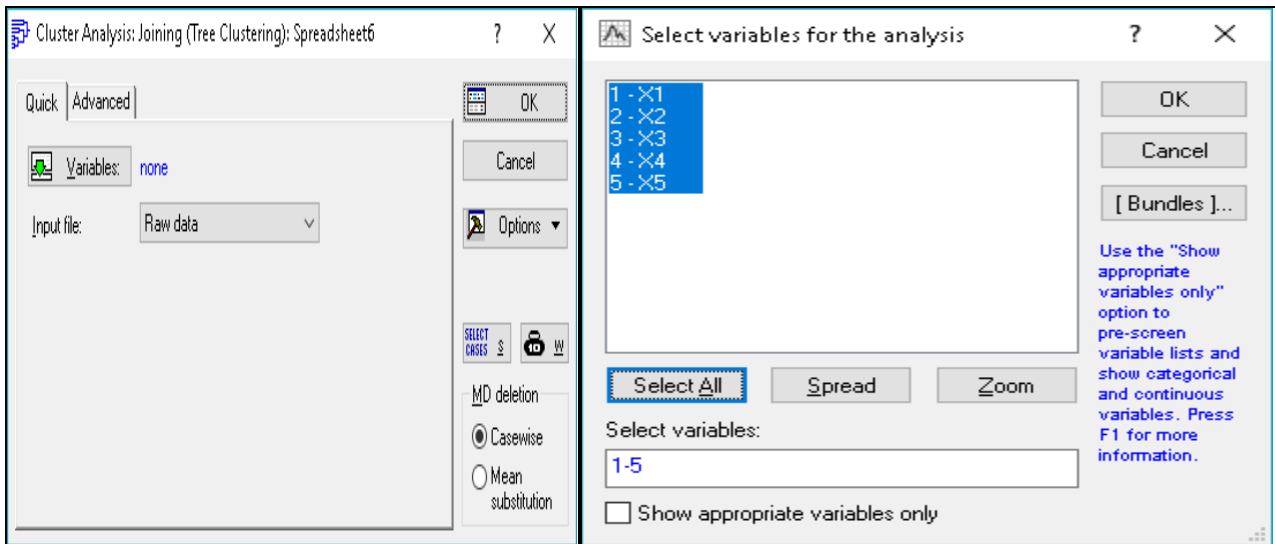


Рис. 80. Діалогове вікно вибору показників, що беруть участь у кластеризації

3.2. Вибір напрямку класифікації, що задається в полі Cluster (кластер) (рис. 81). Доцільно обрати, що саме будемо класифікувати, так під час кластеризації самих змінних позначаються Variables [Columns] (змінні [стовбці]), в заданій задачі Cases [rows] (спостереження [рядки]).

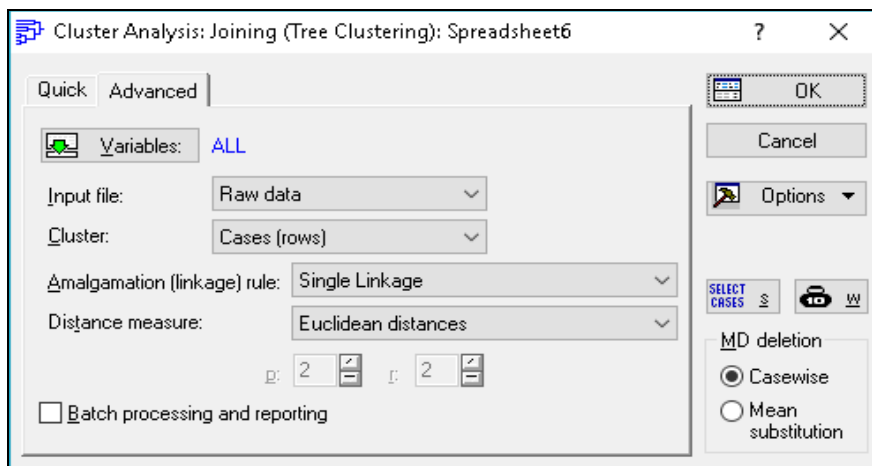


Рис. 81. Діалогове вікно побудови деревовидної кластеризації

3.3. Вибір правила об'єднання об'єктів до групи. З цією метою використовується меню *Amalgamation [linkage] rule* (правило об'єднання [зв'язку]), що дозволяє вибрати одне з наведених правил:

- Single Linkage* (метод одиночного зв'язку "принцип найближчого сусіда");
- Complete Linkage* (метод повного зв'язку "принцип далекого сусіда");
- Unweighted pair-group average* (незважене попарне середнє);
- Weighted pair-group average* (зважене попарне середнє);
- Unweighted pair-group centroid* (незважений центроїдний метод);
- Weighted pair-group centroid* (зважений центроїдний метод);
- Ward's method* (метод Варда).

У роботі пропонується використати метод одиночного зв'язку.

3.4. Вибір типу відстаней, що буде використовуватися в процесі кластеризації. З цією метою у вікні *Distance measure* (міра відстані) необхідно обрати один з типів відстаней, що використовуються в пакеті *Statistica*:

- Squared Euclidean distances* (квадрат Евклідової відстані);
- Euclidean distances* (Евклідова відстань);
- City-block (Manhattan) distance* (відстань міських кварталів (Манхеттенська відстань));
- Chebyshev distance metric* (відстань Чебишева);
- Percent disagreement* (процент незгоди).

У роботі пропонується використовувати Евклідову відстань.

Після встановлення всіх параметрів кластеризації переходимо до вікна її результатів (рис. 82).



Рис. 82. Вікно результатів кластеризації

За допомогою кнопки *Vertical icicle plot* будуємо вертикальну дендрограму (дендрит), а кнопка *Horizontal hierarchical tree plot* дозволяє побудувати горизонтальну дендрограму (рис. 83).

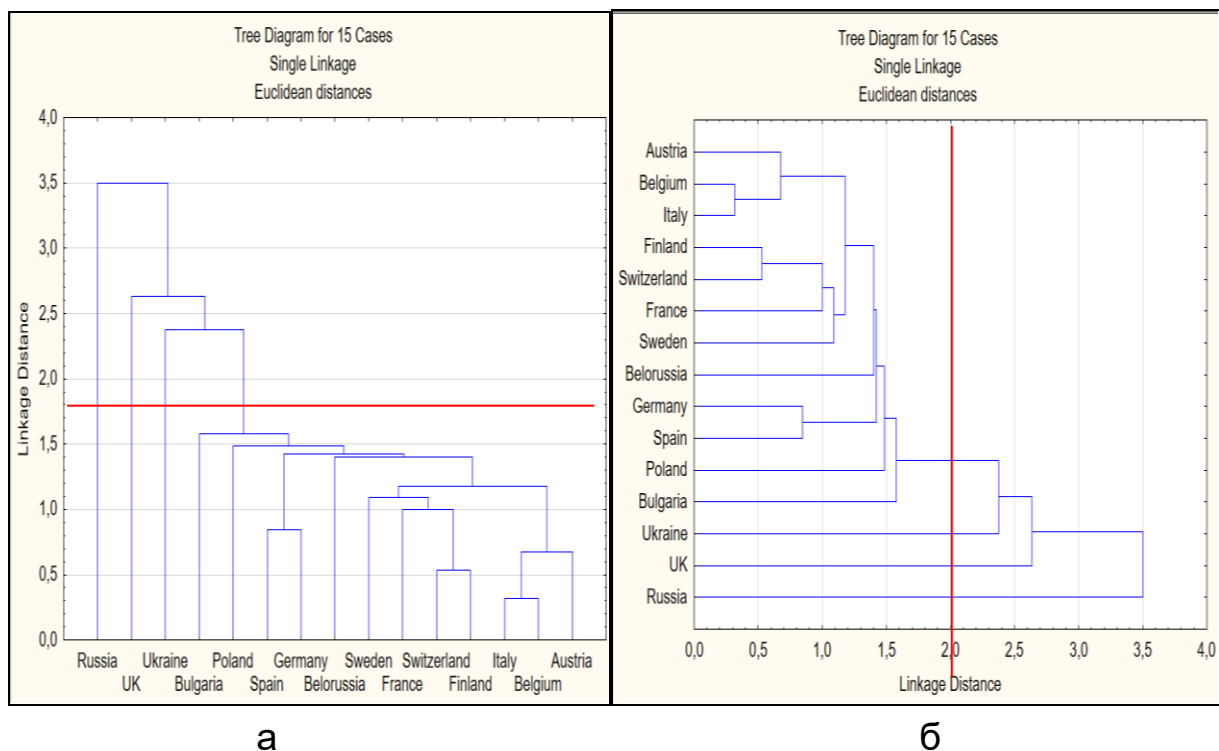
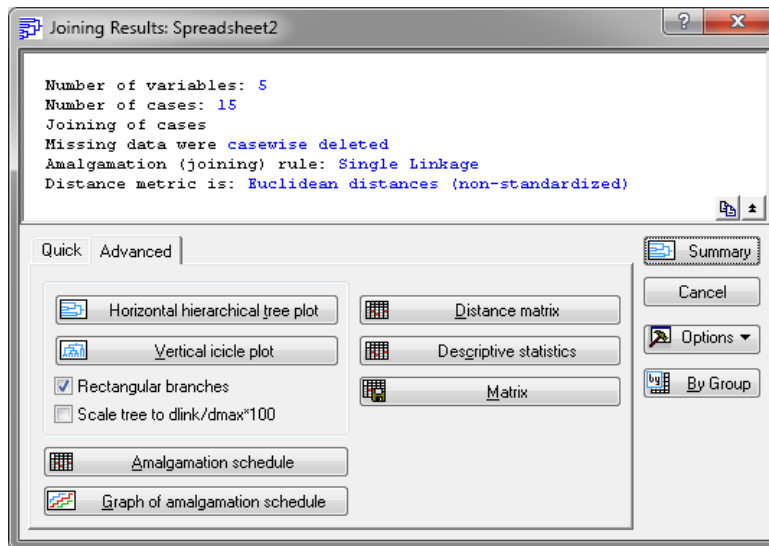


Рис. 83. Візуалізація двох типів дендритів:
а – вертикальна дендрограма;
б – горизонтальна дендрограма

Графічний аналіз визначення кількості кластерних груп на підставі використання дендрограм (див. рис. 83) дозволяє припустити, що найбільш доцільним є розбиття сукупності країн на чотири кластери, оскільки кількість перетинів граничної межі становить чотири рази.

Також одним з доступних у пакеті *Statistica* інструментів для вибору кількості кластерів є графік процесу об'єднання (кнопка *Graph of Amalgamation schedule* (графік об'єднання) і таблиця об'єднання об'єктів (кнопка *Amalgamation schedule*), що наведені на рис. 84 і 85.

Для визначення кількості кластерів на графіку (див. рис. 85) знаходиться точка "перелому" і номер кроку m , на якому стався "перелом"; тоді кількість кластерів дорівнює $n - m$, де n – кількість об'єктів у вибірці; а в стовбці *linkage distance* (відстань зв'язку) таблиці об'єднання (див. рис. 84) знаходиться такий номер кроку m , об'єднання на якому відбулося вже на істотно більшій відстані, ніж на кроці $m - 1$, тоді кількість кластерів дорівнює $n - m$, де n – кількість об'єктів у вибірці.



Amalgamation Schedule (Spreadsheet2)
 Single Linkage
 Euclidean distances

linkage distance	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7	Obj. No. 8	Obj. No. 9	Obj. No. 10	Obj. No. 11	Obj. No. 12	Obj. No. 13	Obj. No. 14	Obj. No. 15
3,3169519	Бельгія	Італія													
5,5314207	Фінляндія	Швейцарія													
6,758969	Австрія	Бельгія	Італія												
8,8478592	Германія	Іспанія													
9,999563	Фінляндія	Швейцарія	Франція												
1,088567	Фінляндія	Швейцарія	Франція	Швеція											
1,178887	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція								
1,399938	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь							
1,423138	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь	Германія	Іспанія					
1,483515	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь	Германія	Іспанія	Польща				
1,579860	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь	Германія	Іспанія	Польща	Болгарія			
2,377592	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь	Германія	Іспанія	Польща	Болгарія	Україна		
2,633688	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь	Германія	Іспанія	Польща	Болгарія	Україна	Велика Британія	
3,500620	Австрія	Бельгія	Італія	Фінляндія	Швейцарія	Франція	Швеція	Білорусь	Германія	Іспанія	Польща	Болгарія	Україна	Велика Британія	Росія

Рис. 84. Таблиця об'єднання об'єктів

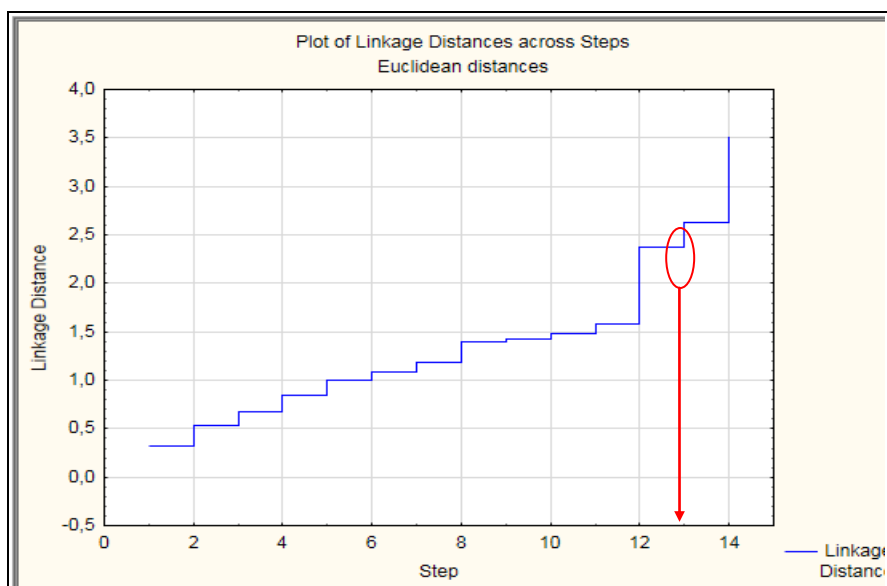


Рис. 85. Графік процесу об'єднання

У нашому випадку як точку перелома можна розглядати крок під номером 12, звідки отримуємо $15 - 12 = 3$ кластера. Однак аналіз таблиці об'єднання свідчить, що на кроці 11 стався стрибок відстані майже на одну одиницю, водночас як на попередніх кроках скачки не перевищували 0,05 одиниць. Отже, за схожістю структури енергетичної безпеки досліджувані країни доцільно розділити на чотири кластери.

4. Побудова кластерів з допомогою метода k-середніх (неієрархічна кластеризація). Проводиться в такі етапи:

4.1. Встановлення основних параметрів кластеризації. Як і в методі деревовидної кластеризації обираються показники, за якими проводиться кластеризація. Враховуючи результати проведеного раніше ієрархічного методу, вказується кількість кластерів, що дорівнює 4 (рис. 86).

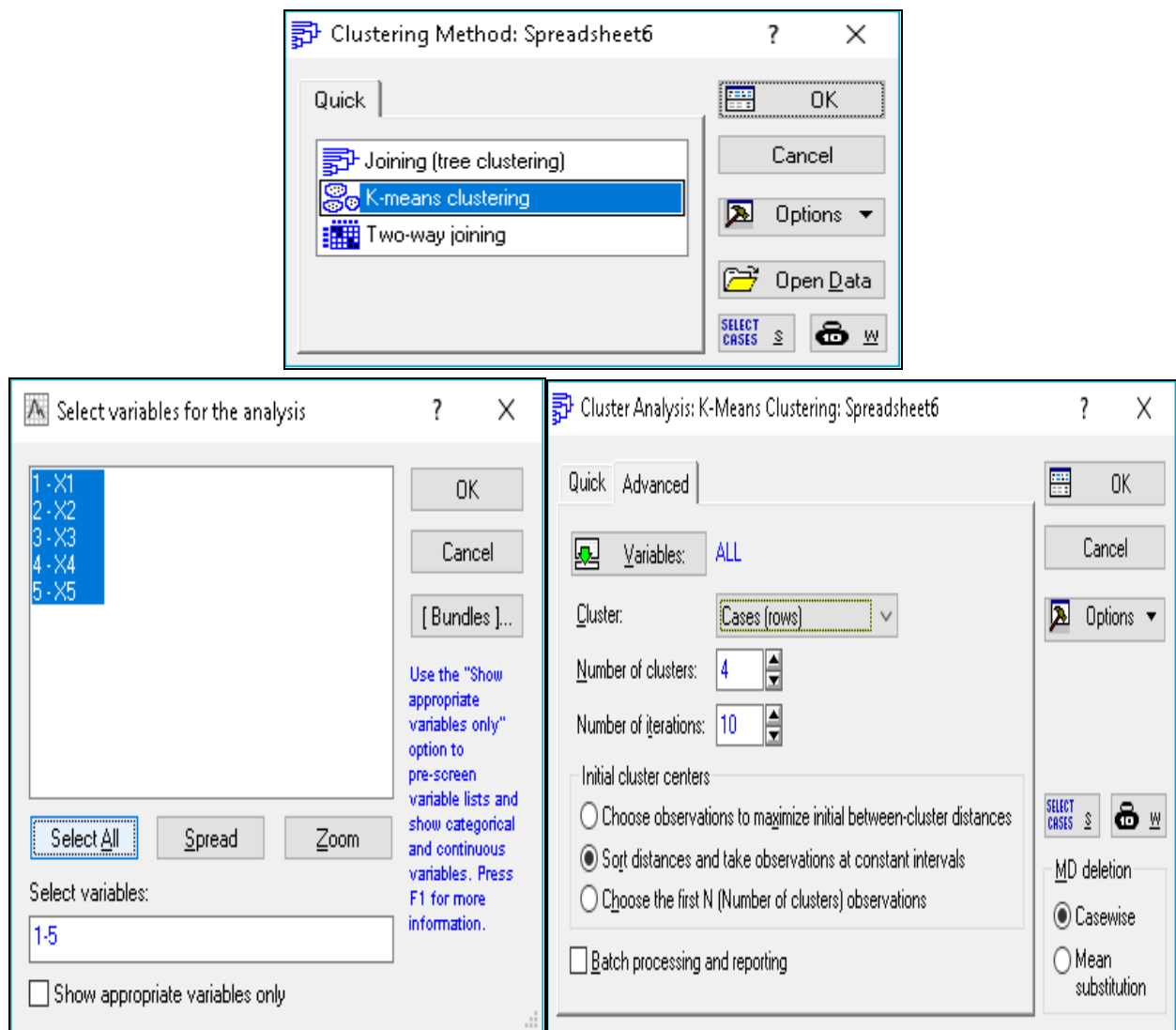


Рис. 86. Діалогове вікно модуля Кластерний аналіз: метод K-means clustering

4.2. У вікні результатів кластеризації (рис. 87) можна вибрати ті розрахунки та звіти з проведеним кластерним аналізом, що необхідно користувачу.

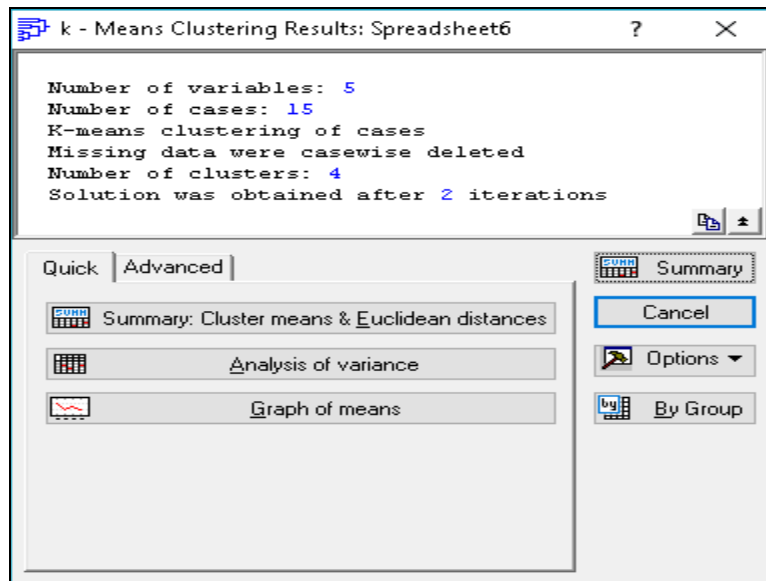


Рис. 87. Вікно результатів кластеризації

4.2.1. Кнопка *Cluster Means & Euclidean Distances* (середні значення в кластерах та евклідові відстані) (рис. 88).

Cluster Number	Euclidean Distances between Clusters (Spreadsheet6)			
	No. 1	No. 2	No. 3	No. 4
No. 1	0,000000	3,060827	0,748075	0,914698
No. 2	1,749522	0,000000	3,612635	2,797337
No. 3	0,864913	1,900693	0,000000	1,520397
No. 4	0,956398	1,672524	1,233044	0,000000

Рис. 88. Евклідові відстані між кластерами

За матрицею відстаней між кластерами можна визначити якість проведеної кластеризації. Чим більше відстань між кластерами і менше відстані між елементами кластерів, тим якісніше проведена кластеризація.

4.2.2. Кнопка *Descriptive Statistics for each cluster* дозволяє визначити описові статистики для кожного з кластерів (рис. 89).

Descriptive Statistics for Cluster 1 (Spreadsheet6) Cluster contains 5 cases			
Variable	Mean	Standard Deviation	Variance
X1	-0,759776	0,359997	0,129598
X2	-0,319659	0,499272	0,249272
X3	-0,368229	0,580369	0,336828
X4	-0,484700	0,189046	0,035738
X5	-0,262310	0,466688	0,217798

Descriptive Statistics for Cluster 2 (Spreadsheet6) Cluster contains 2 cases			
Variable	Mean	Standard Deviation	Variance
X1	-0,461279	0,968021	0,937065
X2	0,513514	1,538150	2,365906
X3	0,455375	0,962425	0,926261
X4	2,299148	1,019948	1,040294
X5	2,206034	0,926090	0,857643

Descriptive Statistics for Cluster 3 (Spreadsheet6) Cluster contains 5 cases			
Variable	Mean	Standard Deviation	Variance
X1	1,134723	0,743860	0,553328
X2	-0,281405	0,620325	0,384803
X3	-0,687714	0,001411	0,000002
X4	-0,311037	0,219579	0,048215
X5	-0,394801	0,489137	0,239255

Descriptive Statistics for Cluster 4 (Spreadsheet6) Cluster contains 3 cases			
Variable	Mean	Standard Deviation	Variance
X1	-0,317392	0,373618	0,139591
X2	0,659430	1,783115	3,179501
X3	1,456321	0,976309	0,953179
X4	-0,206537	0,364254	0,132681
X5	-0,375504	0,131519	0,017297

Рис. 89. Описові статистики для кожного кластера

4.2.3. Перелік країн, що входять до кожного з кластерів, можна отримати з використанням кнопки *Members for each cluster & distances* (члени групи і відстані) (рис. 90).

Members of Cluster Number 1 and Distances from Respective Cluster contains 5 cases		Members of Cluster Number 2 and Distances from Respective Cluster contains 2 cases	
	Distance		Distance
Austria	0,339201	UK	0,782762
Belgium	0,219238	Russia	0,782762
Bulgaria	0,557367		
Germany	0,477962		
Italy	0,270480		

Members of Cluster Number 3 and Distances from Respective Cluster contains 5 cases		Members of Cluster Number 4 and Distances from Respective Cluster contains 3 cases	
	Distance		Distance
Finland	0,176180	Poland	0,657497
France	0,446494	Spain	0,551409
Sweden	0,567955	Ukraine	1,016142
Switzerland	0,176254		
Belorussia	0,629641		

Рис. 90. Перелік країн, що увійшли до кожного кластера

Порівняльний аналіз рис. 88 і 90 дозволив зробити висновок, що побудована кластеризація є якісною, про що свідчить значне перевищення відстані між групами та середніми. З рис. 90 видно, що репрезентантом для першого кластера є Бельгія (найближче розташована країна до центру кластера), другого – Англія, третього кластера – Фінляндія, четвертого – Іспанія.

4.2.4. Для побудови графіка, що відображає характер розбиття країн на кластери, застосуємо кнопку *Graph of means* (графік середніх) (рис. 91).

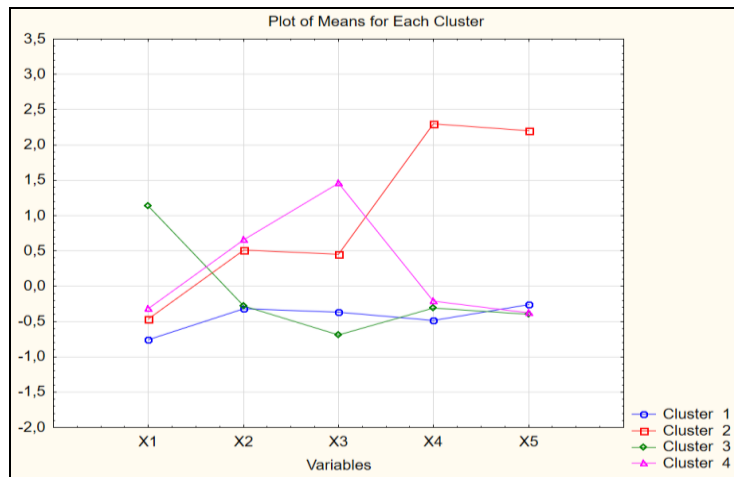


Рис. 91. Графічне середніх значень показників за кластерами

Аналізуючи отримані результати, можна зробити висновки щодо характеристики наведених кластерів (табл. 23).

Таблиця 23

Загальна характеристика кластерів енергозберігальної безпеки

Номер кластера	Перелік країн, що входять до кластера	Основні характеристики класу	Рекомендації
Перший кластер	Австрія, Бельгія, Болгарія, Італія, Германія	Країни з середнім рівнем забезпеченості енергоресурсами та високим рівнем використання енергозберігальних технологій	Диверсифікувати постачальників енергетичних ресурсів
Другий кластер	Великобританія, Росія	Абсолютно енергетично незалежні країни, з великими власними енергетичними запасами, але неефективним використанням енергоресурсів	Підвищити рівень ефективності використання енергоресурсів
Третій кластер	Польща, Іспанія, Україна	Країни з середнім рівнем забезпеченості енергоресурсами та дуже низьким рівнем використання енергозберігальних технологій	Диверсифікувати постачальників енергетичних ресурсів, підвищити рівень ефективності використання енергоресурсів
Четвертий кластер	Фінляндія, Франція, Швеція і Швейцарія, Білорусія	Країни з низьким рівнем забезпеченості енергоресурсами та високим рівнем використання енергозберігальних технологій	Диверсифікувати постачальників енергетичних ресурсів, використовувати нетрадиційні джерела енергетичних ресурсів

Отже, можна зробити висновок, що Україна знаходиться серед країн з середнім рівнем забезпеченості енергоресурсами та дуже низьким рівнем використання енергозберігальних технологій.

Завдання. Використовуючи власну статистичну базу, необхідно виконати групування об'єктів дослідження за допомогою методу кластерного аналізу.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. З метою аналізу фінансового результату діяльності банків міста Києва проводиться їхнє групування за трьома показниками (табл. 24). Необхідно визначити оптимальну кількість кластерів за допомогою методу дендритів, використовуючи різні методи об'єднання в групи. Порівняйте отримані результати.

Таблиця 24

Вихідні дані

№ з/п	Банк	Фінансовий результат, млн дол. США	Прибутковість статутного капіталу, %	Прибутковість активів, %
1	2	3	4	5
1	Інг Банк Україна	170,944	23,375	1,558
2	ВТБ банк	136,412	2,519	0,378
3	Ощадбанк	80,088	0,515	0,108
4	Кліринговий дім	48,058	10,930	1,088
5	Укргазбанк	48,047	0,480	0,291
6	Себ банк	39,626	5,865	1,189
7	КІБ креді агріколь	38,827	22,452	1,004
8	Банк Ренесанс Капітал	33,208	16,489	7,078
9	Сітібанк Україна	20,224	30,412	0,352
10	Сбербанк Росії	11,487	0,384	0,066
11	Платинум банк	8,147	2,141	0,224
12	Кредит Європа банк	6,208	1,229	0,375
13	ЕРДЕ банк	6,053	2,856	0,264
14	Піреус банк МКБ	4,324	0,266	0,103
15	Український професійний банк	4,089	0,818	0,133
16	Альфа-банк	3,969	0,086	0,014
17	Експрес-банк	3,515	1,413	0,122

1	2	3	4	5
18	ЄБРФ	3,019	1,974	1,251
19	Інтеграл банк	3,005	1,878	0,142
20	Правекс-банк	2,907	0,310	0,049
21	Дельта банк	2,812	0,340	0,012
22	Промінвестбанк	2,475	0,047	0,007
23	Унікредит банк	2,458	0,376	0,039
24	Кредитвест банк	2,233	2,536	0,561
25	Хрещатик	2,169	0,298	0,028
26	Астра банк	2,040	0,194	0,117
27	Союз	2,021	0,601	0,057
28	Грін банк	2,020	2,635	1,009
29	Перший інвестиційний банк	1,681	0,731	0,183
30	Банк Кіпру	1,542	0,220	0,059

Завдання 2. Для якісного аналізу фінансового стану підприємств поліграфічної галузі було побудовано інтегральні показники по різних складових фінансового стану (табл. 25). Необхідно дослідити однорідність галузі. У випадку неоднорідності галузі визначити економічну інтерпретацію груп в рамках галузі. Дослідити їх структурні характеристики та зробити висновки щодо якості отриманих результатів.

Таблиця 25

Вихідні дані

Підприємство	Показники			
	Стійкості	Ліквідності	Оборотності	Рентабельності
X1	0,227	0,412	0,228	0,245
X2	0,190	0,240	0,064	0,329
X3	0,352	0,419	0,374	0,653
X4	0,233	0,419	0,229	0,242
X5	0,514	0,414	0,366	0,648
X6	0,212	0,415	0,180	0,487
X7	0,173	0,486	0,232	0,355
X8	0,280	0,913	0,180	0,633
X9	0,280	0,414	0,180	0,463
X10	0,356	0,339	0,308	0,350

Запитання для самоконтролю

1. Чим відрізняється кластеризація від класифікації?
2. Наведіть основні властивості кластера.
3. Перерахуйте основні напрями використання кластерного аналізу.
4. Перелічіть етапи формування матриці спостережень.
5. Наведіть перелік відстаней, що найчастіше застосовуються в багатомірному аналізі.
6. Який вигляд має матриця відстаней?
7. Чим відрізняється розрахунок відстані Чебишева від розрахунку середньої абсолютної різниці значень ознак?
8. Наведіть властивості елементів матриці відстаней.
9. Перерахуйте групи методів кластерного аналізу.
10. Наведіть етапи реалізації методу k-середніх.

Тема 16. Дискримінантний аналіз

Лабораторна робота 12. Вирішення завдання класифікації методом дискримінантного аналізу

Мета: надбання навичок використання дискримінантного аналізу в пакеті *Statistica*.

Завдання: перевірити якість кластеризації методами дискримінантного аналізу та навчитися проводити класифікацію об'єктів за дискримінаційною функцією.

Методичні рекомендації

Дискримінантний аналіз є багатомірним статистичним методом, який дозволяє вивчати відмінності між двома та більше групами об'єктів за декількома змінними одночасно. Тобто, основним завданням дискримінантного аналізу є дослідження групових відмінностей, тобто дискримінація об'єктів за певними ознаками. За допомогою дискримінантного аналізу вирішується два типи завдань:

1. Пошук функції, відповідно до якої об'єкт відноситься до одного з відомих класів.
2. Класифікація нових об'єктів за знайденими правилами.

Розглянемо приклад використання дискримінантного аналізу для вирішення завдання розпізнання об'єктів.

Як вихідні дані використаємо результати кластеризації країн за рівнем енергетичної безпеки (лабораторна робота 11, табл. 22 і 23). Нагадаємо, що досліджувані країни було розбито на чотири кластери (групи) за такими показниками: частка власних джерел у балансі паливно-енергетичних ресурсів держави (X1); частка наявного паливного ресурсу в споживанні паливно-енергетичних ресурсів (X2); енергоемність ВВП (X3); обсяг видобутку вугілля (X4) і ступінь забезпечення паливно-енергетичними ресурсами (X5). Відповідні дані перенесемо у файл пакета Statistica (рис. 92).

	1 X1	2 X2	3 X3	4 X4	5 X5	6 Cluster
Австрія	0,512517	8,139704	28,08643	0,657682	1,229883	1
Бельгія	0,441011	11,18419	46,6101	0,423147	1,198028	1
Болгарія	0,390566	67,68311	473,7965	0,728101	0,629035	1
Фінляндія	0,659931	13,74058	0	0,644842	1,181073	4
Франція	0,598438	7,578963	6,85112	0,746463	0,569163	4
Германія	0,425501	9,358731	2922,118	0,671615	0,998937	1
Італія	0,422215	9,180896	4,465834	0,526936	1,333184	1
Польща	0,447059	27,10571	6933,373	1,057638	1,030381	3
Іспанія	0,493978	10,29191	4295,369	0,61782	0,923633	3
Швеція	0,828147	8,865035	0	0,909419	0,948364	4
Швейцарія	0,694735	4,621731	0	0,75509	0,952061	4
Велика Британія	0,389766	7,828034	1005,633	1,954056	2,160322	2
Білорусь	0,603121	80,72323	0	0,521061	1,342974	4
Росія	0,562118	121,3625	3960,709	2,899916	2,942098	2
Україна	0,541134	179,2357	2740,338	0,676241	1,076723	3

Рис. 92. Результати кластеризації країн за рівнем енергетичної безпеки

Вибір модуля дискримінантного аналізу можливо здійснити через меню *Statistics / Multivariate Exploratory Techniques / Discriminant analysis* (статистика / багатовимірні методи дослідження / дискримінантний аналіз) (рис. 93) або використати перемикач *Module Switcher* (перемикач модулів), який містить перелік усіх доступних модулів і натиснути *Discriminant Analysis*, а потім кнопку *Switch to* (перейти до).

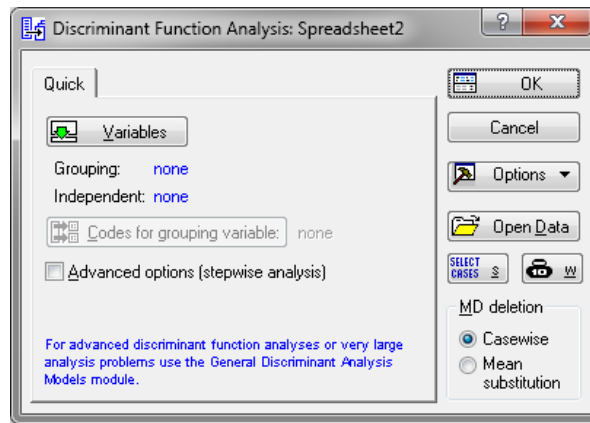


Рис. 93. Запуск модуля *Discriminant Function Analysis*

На екрані з'явиться стартова панель модуля *Discriminant Function Analysis* (аналіз дискримінантних функцій) (див. рис. 93), яка містить такі функціональні кнопки:

Variables (змінні) дозволяє вибрати *Grouping* (групувальні змінні) та *Independent* (незалежні змінні);

Codes for grouping variable (коди для груп змінних) – вказують кількість аналізованих груп об'єктів;

Missing data (пропущені змінні) – дозволяє вибрати *Case wise* (построкове видалення змінних зі списку) або *Mean substitution* (змінити їх на середні значення);

Open Data (відкрити дані) – відкриває файл з даними;

кнопка *Select Cases* (виберіть випадки) задає умови вибору спостережень з бази даних;

кнопка *W* задає ваги змінних, обравши їх зі списку.

Після натискання кнопки *Variables* можливо вибрати групувальну та незалежну змінні (рис. 94).

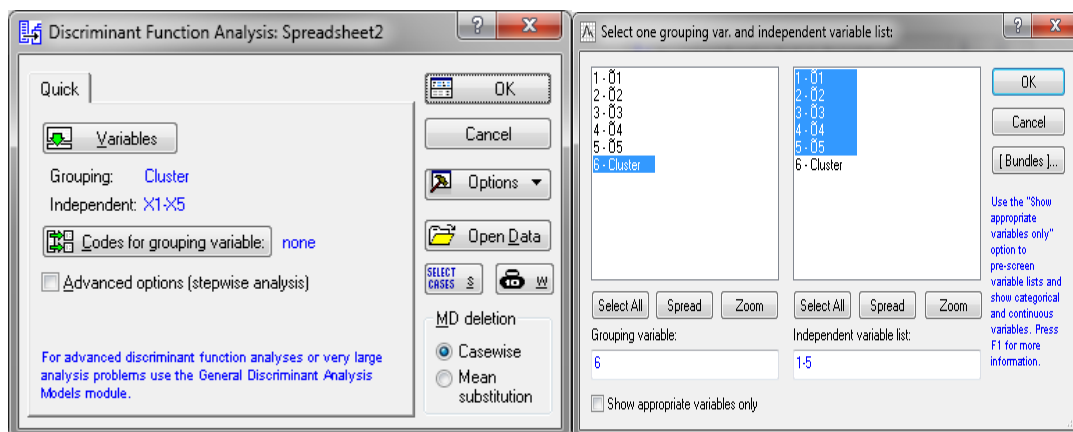


Рис. 94. Вибір змінних для аналізу

Кнопка *Code for grouping variable* дозволяє задати рівні (класи) залежної змінної (рис. 95).

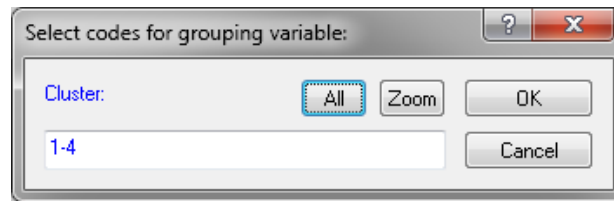


Рис. 95. Вибір рівнів групування для залежної змінної

Додатково обираємо спосіб проведення покрокового дискримінантного аналізу – *Advanced option (stepwise analysis)*. Розширений варіант (покроковий аналіз) (рис. 96).

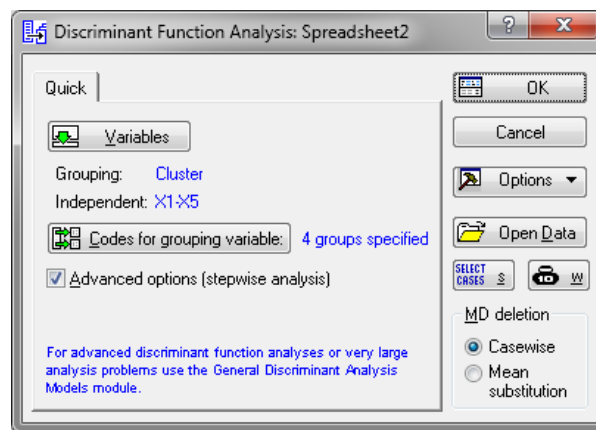


Рис. 96. Встановлення всіх параметрів дискримінантного аналізу

Натискаємо *OK* і переходимо до іншого вікна параметрів. Вигляд діалогового вікна *Model Definition*, яке призначене для вибору моделі, наведено на рис. 97. У вкладці *Advanced* можливо задати метод (*Method*), який буде використовуватися для вибору значущих змінних.

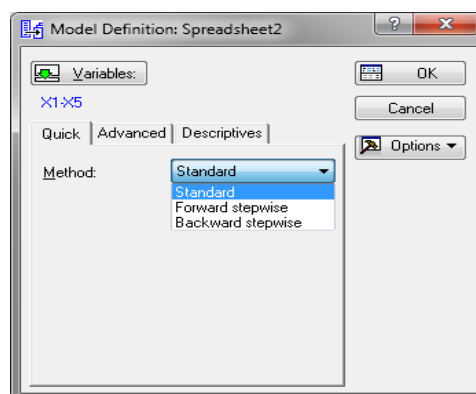


Рис. 97. Вигляд діалогового вікна *Model Definition*

Можливе використання наступних трьох методів:

Standart (стандартний). Усі змінні одночасно включені в модель;

Forward stepwise (покроковий з включенням). На кожному кроці в модель відбирається змінна з максимальним F-значенням. Процедура закінчується, коли всі змінні, значення F яких більше значення вказаного в полі *F to enter* увійшли до моделі;

Backward stepwise (покроковий з виключенням). На кожному кроці в модель відбираються всі змінні, які потім видаляються залежно від величини F-значення. Кроки закінчуються, коли немає змінних F-значення яких, менше певного, визначеного в полі *F to remove*.

Поле *Number of steps* (кількість кроків) визначає максимальну кількість кроків аналізу, після досягнення яких процедура закінчується.

Поле *Tolerance* (толерантність) дозволяє виключити з моделі неінформативні змінні. Якщо толерантність має значення менше ніж значення 0,01, то змінна визнається не інформативною та не включається до моделі.

Як метод аналізу виберемо *Standard*. За результатами, отриманими в ході обчислень, наведеними у вікні *Discriminant Function Analysis Results*, отримали таку інформацію (рис. 98):

кількість змінних в моделі (*Number of variables in the model*) – 5;

значення лямбди Вілкса (*Wilks' Lambda*) – 0,001 39;

приблизне значення F-статистики, яке пов'язане з лямбдою Вілкса (*Approx. F (15, 19)*) = 12,922 18;

рівень значущості F-критерію $p < 0,000 0$ для отриманого значення 12,922 18.

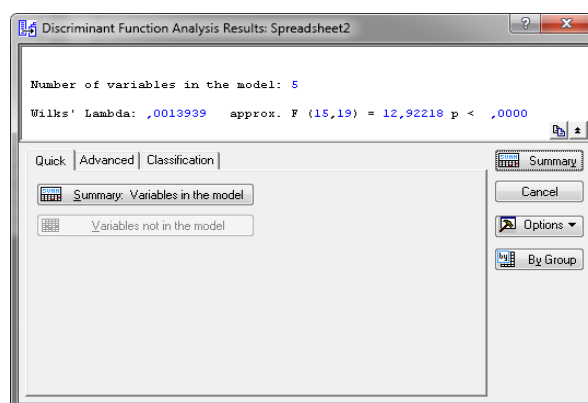


Рис. 98. Результати дискримінантного аналізу

Критерії оцінювання якості класифікації базуються на розрахунку можливих переходів аналізованих об'єктів від одного класу до іншого (рис. 99).

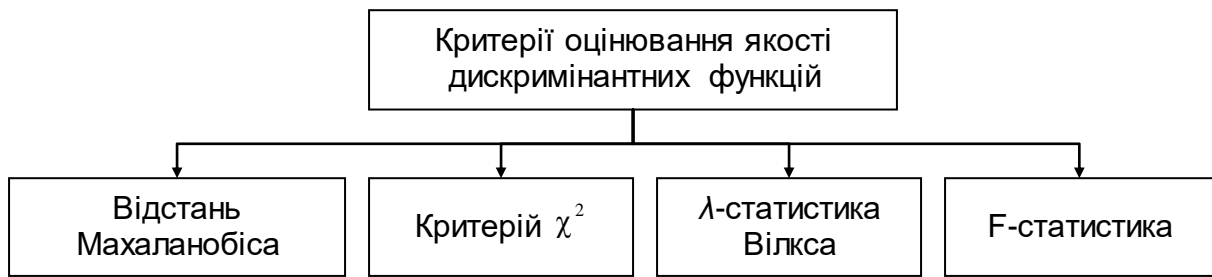


Рис. 99. Критерії оцінювання якості дискримінантних функцій

Для оцінювання спроможності дискримінантної функції розпізнавати класи у багатовимірному ознаковому просторі використовують *λ-статистику Вілкса (Wilks lambda)*. *λ-статистика* враховує як відмінності між класами, так і однорідність кожного класу. Оскільки *λ* розраховується як обернена величина, то чим більше різняться центроїди, тим менше її значення, і навпаки, якщо центроїди збігаються, то *λ* прямує до 1. Отже, близькі значення *λ* до 0 свідчать про високу розпізнавальну спроможність дискримінантної функції. Таким чином, за даними показника Wilks' Lambda, який дорівнює 0,001 39, можливо зробити висновок, що класифікація є коректною. Також значимість *F-статистики* ($F > F_{\text{табл.}}$) підтверджує існування відмінностей між групами. У нашому випадку $F_{\text{табл.}}$ ($\alpha = 0,05, k_1 = 5, k_2 = 15$) = 2,9. Отже, проведена класифікація країн за рівнем енергетичної безпеки є коректною.

Як перевірку коректності досліджуваних вибірок розглянемо результати класифікаційної матриці, натиснувши кнопку *Classification matrix* (класифікаційна матриця), попередньо обравши *Same for all groups* (те саме для всіх груп) у правій частині вікна *Discriminant Function Analysis Results* (результати аналізу дискримінантних функцій) (рис. 100).

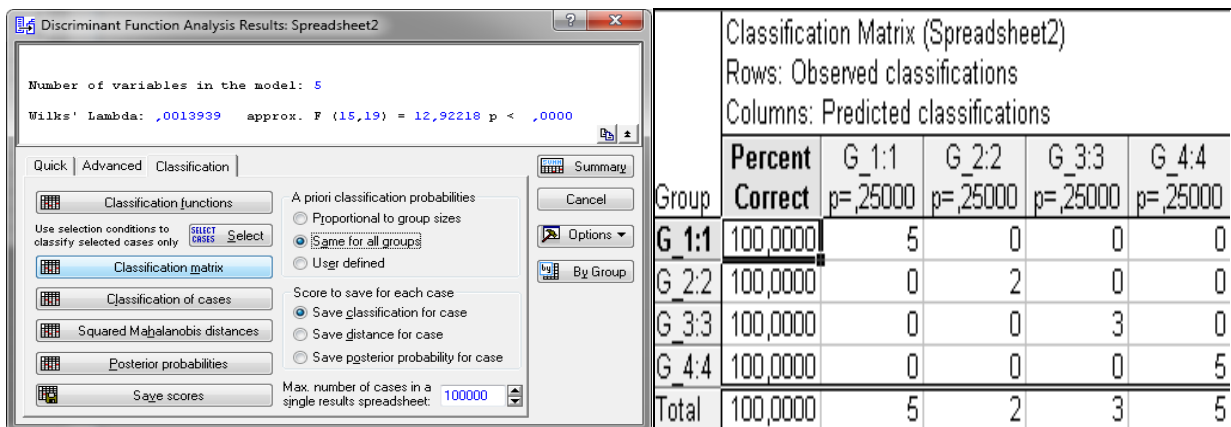


Рис. 100. Класифікаційна матриця

За результатами класифікаційної матриці (див. стовбець *Percent Correct*(процент коректності) – 100 %) (див. рис. 100) можна зробити висновок, що об'єкти розбито вірно на чотири групи за допомогою кластерного аналізу. Якщо є країни, що неправильно віднесені до відповідних груп, можна подивитися *Classification of cases* (класифікація випадків). У таблиці класифікації (рис. 101) випадків некоректно віднесені об'єкти позначаються зірочкою (*).

Case	Classification of Cases (Spreadsheet2) Incorrect classifications are marked with *				
	Observed Classif.	1 p=,25000	2 p=,25000	3 p=,25000	4 p=,25000
Австрія	G_1:1	G_1:1	G_4:4	G_3:3	G_2:2
Бельгія	G_1:1	G_1:1	G_4:4	G_3:3	G_2:2
Болгарія	G_1:1	G_1:1	G_4:4	G_3:3	G_2:2
Фінляндія	G_4:4	G_4:4	G_1:1	G_3:3	G_2:2
Франція	G_4:4	G_4:4	G_1:1	G_3:3	G_2:2
Германія	G_1:1	G_1:1	G_3:3	G_4:4	G_2:2
Італія	G_1:1	G_1:1	G_4:4	G_3:3	G_2:2
Польща	G_3:3	G_3:3	G_1:1	G_4:4	G_2:2
Іспанія	G_3:3	G_3:3	G_4:4	G_1:1	G_2:2
Швеція	G_4:4	G_4:4	G_1:1	G_3:3	G_2:2
Швейцарія	G_4:4	G_4:4	G_1:1	G_3:3	G_2:2
Велика Британія	G_2:2	G_2:2	G_1:1	G_3:3	G_4:4
Білорусь	G_4:4	G_4:4	G_1:1	G_3:3	G_2:2
Росія	G_2:2	G_2:2	G_1:1	G_3:3	G_4:4
Україна	G_3:3	G_3:3	G_4:4	G_1:1	G_2:2

Рис. 101. Класифікація випадків

Класифікаційні функції для кожного класу можливо отримати у вікні *Discriminant Function Analysis Results*, натиснувши кнопку *Classification functions* (класифікаційні функції) (рис. 102).

Variable	Classification Functions; grouping: Cluster			
	G_1:1 p=,25000	G_2:2 p=,25000	G_3:3 p=,25000	G_4:4 p=,25000
X1	84,7159	-168,551	178,1224	165,2260
X2	-0,0120	-0,226	0,0938	0,0089
X3	0,0012	-0,018	0,0105	0,0035
X4	-8,4011	125,805	-54,2853	-25,4289
X5	5,2610	47,406	-8,6650	-3,1706
Constant	-20,5601	-144,775	-47,6582	-46,7314

Рис. 102. Параметри класифікаційних функцій

Дискримінантна функція – це лінійна комбінація певної множини ознак, які називаються класифікаційними та на основі яких ідентифікуються класи однорідних за деякими властивостями об'єктів.

Отримані дискримінантні функції мають такий вигляд:

країни з середнім рівнем забезпеченості енергоресурсами та високим рівнем використання енергозберігальних технологій = $84,7159 \times X_1 - 0,0120 \times X_2 + 0,0012 \times X_3 - 8,4011 \times X_4 + 5,2610 \times X_5 - 20,5601$;

абсолютно енергетично незалежні країни з великими власними енергетичними запасами, але неефективним використанням енергоресурсів = $-168,551 \times X_1 - 0,226 \times X_2 - 0,018 \times X_3 + 125,805 \times X_4 + 47,406 \times X_5 - 144,775$;

країни з середнім рівнем забезпеченості енергоресурсами та дуже низьким рівнем використання енергозберігальних технологій = $178,1224 \times X_1 + 0,0938 \times X_2 + 0,0105 \times X_3 - 54,2853 \times X_4 - 8,6650 \times X_5 - 47,6582$;

країни з низьким рівнем забезпеченості енергоресурсами та високим рівнем використання енергозберігальних технологій = $165,2260 \times X_1 + 0,0089 \times X_2 + 0,0035 \times X_3 - 25,4289 \times X_4 - 3,1706 \times X_5 - 46,7314$.

Унаслідок отримуємо коефіцієнти для кожної змінної та для кожної дискримінантної функції. Вони можуть бути також проінтерпретовані у звичайний спосіб: чим більше стандартизований коефіцієнт, тим більший внесок відповідної змінної в дискримінацію сукупностей.

Для отримання більш детальної інформації можна переглянути результати канонічного аналізу, який можливо проводити, якщо були вибрані принаймні три групи і є хоча б дві змінні в моделі. Для цього необхідно натиснути кнопку *Perform canonical analysis* (рис. 103).

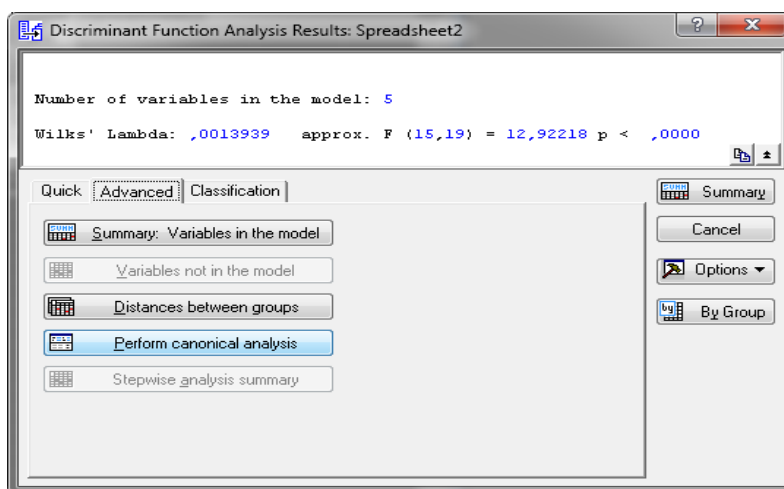


Рис. 103. Запуск канонічного аналізу

З'являється вікно канонічного аналізу (рис. 104), у якому за допомогою опції *Scatterplot of canonical scores* (Діаграма розкиду канонічних партитур) можливо побудувати наступну діаграму розсіювання для значень (рис. 105).

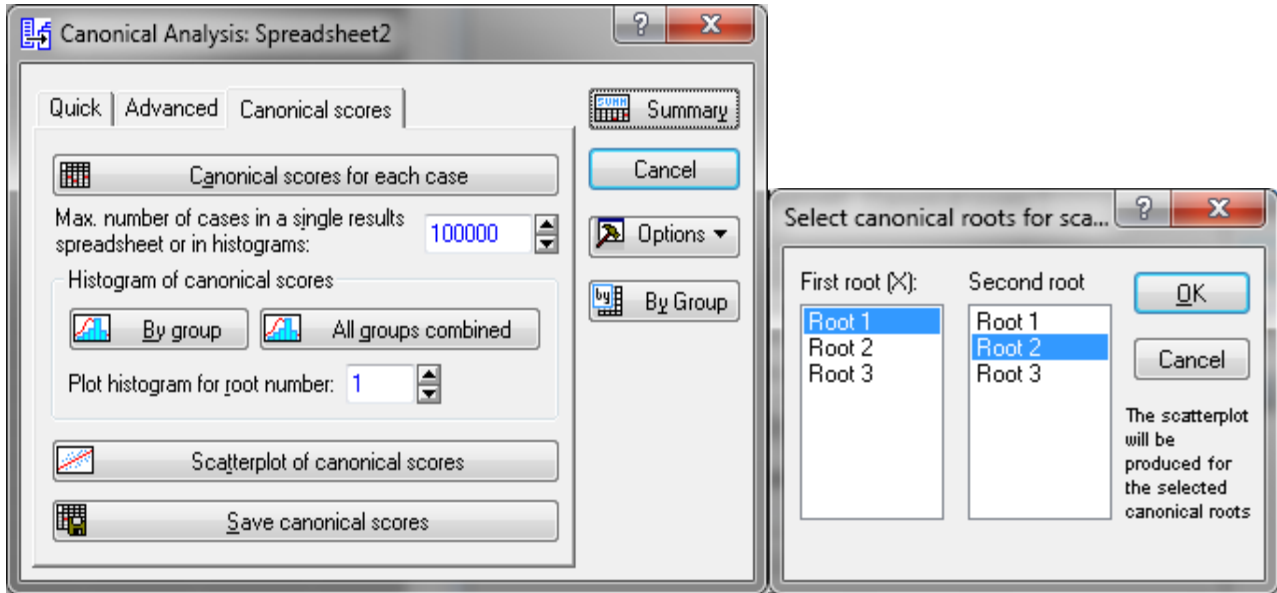


Рис. 104. Діалогові вікна канонічного аналізу

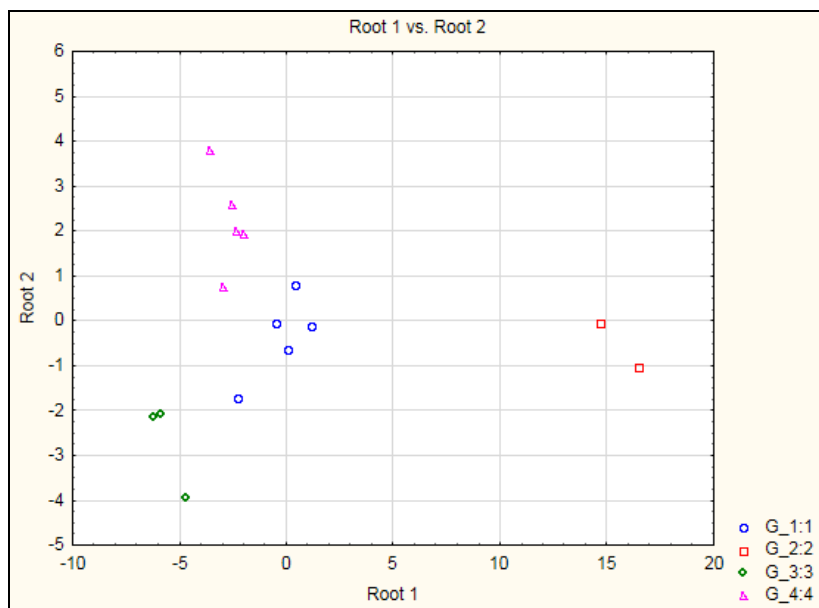


Рис. 105. Графік розсіювання канонічних значень

За допомогою цієї діаграми (див. рис. 105) можливо визначити внесок, що робить кожна дискримінантна функція в розподіл між групами.

Також щоб визначити, чи є побудовані дискримінантні функції статистично значущими. Для цього натиснемо на кнопку Chi square tests of

successive roots (Хі-квадрат критерій) для віддалених коренів з вікна результатів (рис. 106).

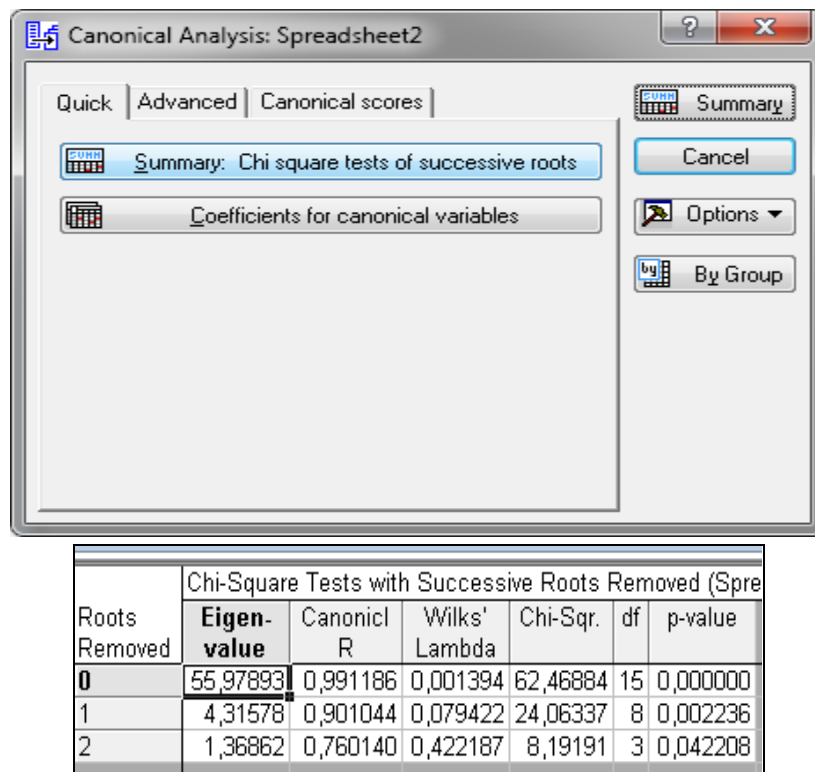


Рис. 106. Перевірка статистичної значущості дискримінантної функції

У цьому прикладі дискримінантні функції статистично значущі, оскільки значення в стовбці p-value менше 0,05.

Отже, проведена класифікація країн за рівнем енергетичної безпеки методом кластерного аналізу є адекватною та коректною. У ході проведення дискримінантного аналізу побудовано функції, які можна використовувати в подальшому для віднесення певної (нової або тієї, що не брали участі у аналізі) країни в один з отриманих класів (кластерів, груп).

Тепер розглянемо приклад використання дискримінантного аналізу для вирішення завдання класифікації нових об'єктів за знайденими правилами.

За умовою завдання необхідно розподілити (віднести) ще дві країни – Молдову та Чехію за вже відомими групами, отриманими в попередньому завданні-прикладі. Дані для класифікації наведено в табл. 26.

Вихідні дані

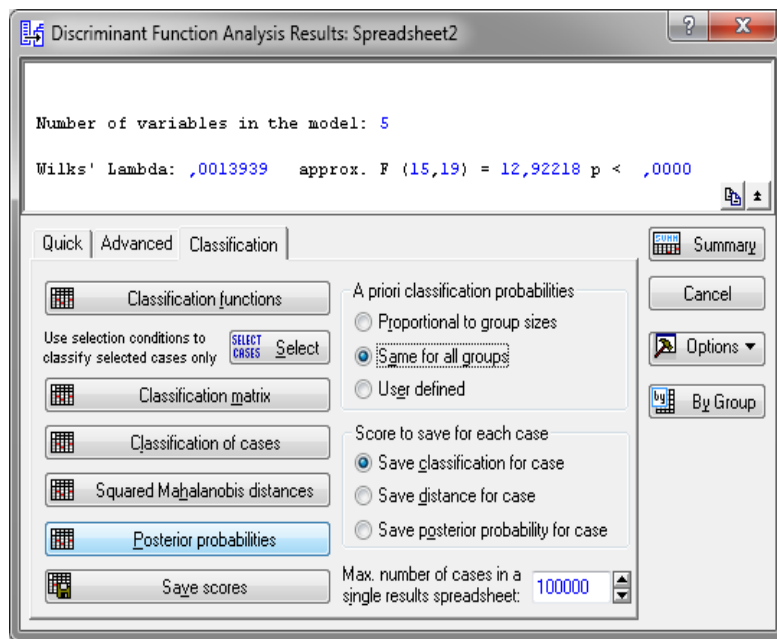
Країна	X1	X2	X3	X4	X5
Молдова	0,437 06	72,437 50	0	0,6123 48	1,039 504
Чехія	0,395 01	13,206 48	7,739 54	0,986 00	1,006 457

Спочатку необхідно внести нові дані до вже створеної таблиці пакета Statistica (рис. 107).

	1 X1	2 X2	3 X3	4 X4	5 X5	6 Cluster
Австрія	0,512517	8,139704	28,08643	0,657682	1,229883	1
Бельгія	0,441011	11,18419	46,6101	0,423147	1,198028	1
Болгарія	0,390566	67,68311	473,7965	0,728101	0,629035	1
Фінляндія	0,659931	13,74058	0	0,644842	1,181073	4
Франція	0,598438	7,578963	6,85112	0,746463	0,569163	4
Германія	0,425501	9,358731	2922,118	0,671615	0,998937	1
Італія	0,422215	9,180896	4,465834	0,526936	1,333184	1
Польща	0,447059	27,10571	6933,373	1,057638	1,030381	3
Іспанія	0,493978	10,29191	4295,369	0,61782	0,923633	3
Швеція	0,828147	8,865035	0	0,909419	0,948364	4
Швейцарія	0,694735	4,621731	0	0,75509	0,952061	4
Велика Британія	0,389766	7,828034	1005,633	1,954056	2,160322	2
Білорусь	0,603121	80,72323	0	0,521061	1,342974	4
Росія	0,562118	121,3625	3960,709	2,899916	2,942098	2
Україна	0,541134	179,2357	2740,338	0,676241	1,076723	3
Молдова	0,43706	72,4375	0	0,612348	1,039504	
Чехія	0,39501	13,20648	7,73954	0,986	1,006457	

Рис. 107. Вихідні дані з новими спостереженнями

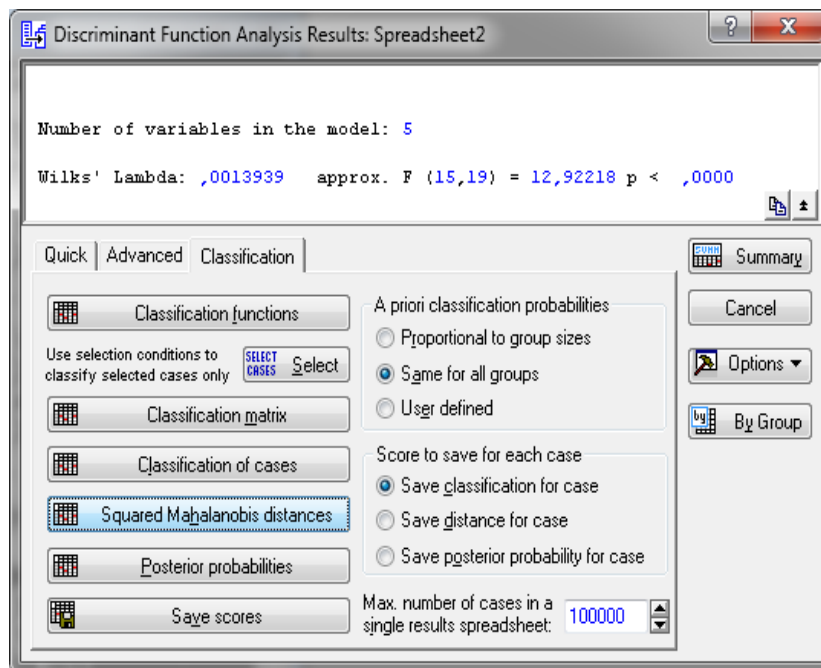
Тепер виконуємо дискримінантний аналіз уже для нової сукупності країн та у вікні результатів обчислимо апостеріорні імовірності, значення яких подано на рис. 108.



Case	Posterior Probabilities (Spreadsheet2)				
	Observed Classif.	G_1:1 p=,25000	G_2:2 p=,25000	G_3:3 p=,25000	G_4:4 p=,25000
Австрія	G_1:1	0,998062	0,000000	0,000000	0,001938
Бельгія	G_1:1	0,999516	0,000000	0,000000	0,000484
Болгарія	G_1:1	0,999952	0,000000	0,000000	0,000048
Фінляндія	G_4:4	0,001817	0,000000	0,000000	0,998183
Франція	G_4:4	0,009255	0,000000	0,000000	0,990745
Германія	G_1:1	0,973026	0,000000	0,020573	0,006401
Італія	G_1:1	0,999995	0,000000	0,000000	0,000005
Польща	G_3:3	0,000000	0,000000	1,000000	0,000000
Іспанія	G_3:3	0,000000	0,000000	0,999999	0,000001
Швеція	G_4:4	0,000000	0,000000	0,000000	1,000000
Швейцарія	G_4:4	0,000127	0,000000	0,000000	0,999873
Велика Британія	G_2:2	0,000000	1,000000	0,000000	0,000000
Білорусь	G_4:4	0,020246	0,000000	0,000161	0,979593
Росія	G_2:2	0,000000	1,000000	0,000000	0,000000
Україна	G_3:3	0,000000	0,000000	1,000000	0,000000
Молдова	---	0,999827	0,000000	0,000000	0,000173
Чехія	---	1,000000	0,000000	0,000000	0,000000

Рис. 108. Таблица апостеріорних імовірностей

Наступним кроком є обчислення відстаней від нових випадків до центрів груп. Для цього обираємо кнопку *Squared Mahalanobis distances* (відстань Махаланобіса), що визначає належність змінної до того чи іншого класу. Результати розрахунків наведено на рис. 109.



Case	Squared Mahalanobis Distances from Group Centroids (Incorrect classifications are marked with *)				
	Observed Classif.	G_1:1 p=,25000	G_2:2 p=,25000	G_3:3 p=,25000	G_4:4 p=,25000
Австрія	G_1:1	2,0729	234,4760	52,8876	14,5610
Бельгія	G_1:1	0,9494	264,7849	41,7228	16,2165
Болгарія	G_1:1	5,3179	250,3143	47,7357	25,2173
Фінляндія	G_4:4	13,4271	329,3805	33,5037	0,8098
Франція	G_4:4	12,9011	319,6329	37,9240	3,5545
Германія	G_1:1	7,1967	322,9463	14,9096	17,2445
Італія	G_1:1	3,0716	215,6587	61,8547	27,4519
Польща	G_3:3	42,8075	424,9053	4,2671	44,3261
Іспанія	G_3:3	39,6876	464,0628	1,7786	29,4089
Швеція	G_4:4	39,9212	387,5931	47,4028	5,5078
Швейцарія	G_4:4	18,2628	337,6522	37,5058	0,3157
Велика Британія	G_2:2	223,2886	4,5423	423,7596	311,0095
Білорусь	G_4:4	13,0294	348,4689	22,6968	5,2710
Росія	G_2:2	293,6248	4,5423	491,3669	380,9826
Україна	G_3:3	49,6542	481,9426	5,8024	36,2577
Молдова	---	1,3132	250,1662	43,6811	18,6321
Чехія	---	25,4861	130,4643	121,4663	64,1878

Рис. 109. Таблиця відстаней від нових випадків до центрів груп

Максимальне значення апостеріорних імовірностей (див. рис. 108) і мінімальна відстань від нового випадку до центрів груп (див. рис. 109) відповідають кластеру 1. Тому досліджувані країни (Молдову та Чехію) доцільно віднести до першого кластера країн з середнім рівнем забезпеченості енергоресурсами та високим рівнем використання енергозберігальних технологій.

Завдання. Використовуючи власну статистичну базу (наприклад, результати проведеної кластеризації), необхідно перевірити якість кластеризації методами дискримінантного аналізу та навчитися проводити класифікацію об'єктів за дискримінаційною функцією *Search for functions* (пошук функції), відповідно до якої об'єкт відноситься до одного з відомих класів.

Результати роботи оформіть у вигляді звіту з лабораторної роботи.

Завдання для самостійного опрацювання

Завдання 1. Для реалізації ефективної фінансової стратегії підприємства, маркетологом було досліджено фінансовий стан підприємств конкурентів шляхом їхнього групування за трьома групами (табл. 27). Необхідно визначити структурні характеристики групування та надати економічну інтерпретацію отриманим групам. Проведіть аналіз якості отриманих груп і визначте імовірнісні зміни підприємствами своєї групи фінансового стану.

Таблиця 27

Вихідні дані

Група	Підприємство	Показники			
		Стійкості	Ліквідності	Оборотності	Рентабельності
3	X1	0,227	0,412	0,228	0,245
3	X2	0,190	0,240	0,064	0,329
1	X3	0,352	0,419	0,374	0,653
3	X4	0,233	0,419	0,229	0,242
1	X5	0,514	0,414	0,366	0,648
3	X6	0,212	0,415	0,180	0,487
3	X7	0,173	0,486	0,232	0,355
2	X8	0,280	0,913	0,180	0,633
3	X9	0,280	0,414	0,180	0,463
2	X10	0,356	0,339	0,308	0,350

Завдання 2. Відповідно до завдання 1 необхідно віднести два підприємства (X11, X12) до трьох раніше утворених груп. Вихідні дані наведено у табл. 28.

Вихідні дані

Підприємство	Показники			
	Стійкості	Ліквідності	Оборотності	Рентабельності
X11	0,318	0,512	0,122	0,198
X12	0,006	0,015	0,124	0,228

Запитання для самоконтролю

1. Сформулюйте визначення дискримінантного аналізу.
2. Які підходи існують до проведення дискримінантного аналізу?
3. Наведіть загальний вигляд функції дискримінації.
4. У чому полягає поняття центроїду?
5. Що характеризує міжкласову варіацію?
6. Які основні завдання дискримінантного аналізу?
7. Назвіть критерії перевірки якості дискримінації.
8. Наведіть визначення дискримінантної функції.
9. У яких межах вимірюється лямбда Вілкса?
10. Чим відрізняються стандартизовані коефіцієнти дискримінантної функції від структурних?

Рекомендована література

Основна

1. Антохонова И. В. Методы прогнозирования социально-экономических процессов : учеб. пособ. / И. В. Антохонова. – Улан-Удэ : Изд-во ВСГТУ, 2004. – 212 с.
2. Бабешко Л. О. Основы эконометрического моделирования : учеб. пособ. / Л. О. Бабешко. – Изд. 3-е. – Москва : Ком книга, 2007. – 432 с.
3. Вітлинський В. В. Моделювання економіки : навч. посіб. / В. В. Вітлинський. – Київ : КНЕУ, 2003. – 408 с.
4. Єріна А. М. Статистичне моделювання та прогнозування : навч. посіб. / А. М. Єріна. – Київ : КНЕУ, 2001. – 170 с.

5. Клебанова Т. С. Эконометрия : учебно-методическое пособие для самостоятельного изучения дисциплины / Т. С. Клебанова, Н. А. Дубовина, Е. В. Раевнева. – Харьков : Изд. Дом "ИНЖЭК", 2003. – 132 с.
6. Когнитивная бизнес-аналитика : учебник / под науч. ред. д. т. н., профессора Н. М. Абдикеева. – Москва : ИНФРА-М, 2011. – 511 с.
7. Мандель И. Д. Кластерный анализ / И. Д. Мандель. – Москва : Финансы и статистика, 1988. – 176 с.
8. Моделі і методи соціально-економічного прогнозування : підручник / В. М. Геєць, Т. С. Клебанова, О. І. Черняк та ін. – 2 вид., виправ. – Харків : ВД "ІНЖЕК", 2008. – 396 с.
9. Присенко Г. В. Прогнозування соціально-економічних процесів : навч. посіб. / Г. В. Присенко, Є. І. Равікович. – Київ : КНЕУ, 2005. – 378 с.
10. Статистика : навч. посіб. / за ред. д-ра. екон. наук, професора О. В. Раєвнєвої. – Харків : Вид. ХНЕУ, 2010. – 520 с.
11. Статистика / С. С. Герасименко, А. В. Головач, А. М. Єріна та ін. – 2-ге вид., перероб. і допов. – Київ : КНЕУ, 2000. – 467 с.
12. Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мюллер, У. Р. Клекка и др. ; пер. с англ. ; под ред. И. С. Енюкова. – Москва : Финансы и статистика, 1989. – 215 с.

Додаткова

13. Буреева Н. Н. Многомерный статистический анализ с использованием ППП "STATISTICA" : учебно-методический материал по программе повышения квалификации "Применение программных средств в научных исследованиях и преподавании математики и механики" / Н. Н. Буреева. – Нижний Новгород : ННГУ, 2007. – 112 с.
14. Моделирование экономики : учеб. пособ. / Т. С. Клебанова, В. А. Забродский, О. Ю. Полякова, В. Л. Петренко. – Харьков : Изд. ХГЭУ, 2001. – 140 с.
15. Орлов А. И. Организационно-экономическое моделирование : учебник / А. И. Орлов. – Москва : Изд-во МГТУ им. Н. Э. Баумана. – 2009. – 254 с.
16. Халафян А. А. STATISTICA 6. Статистический анализ данных / А. А. Халафян. – Москва : ООО "Бином-Пресс", 2008. – 512 с.
17. Христиановский В. В. Экономико-математические методы и модели: теория и практика : учеб. пособ. / В. В. Христиановский, В. П. Щербина. – Донецк : ДонНУ, 2010. – 335 с.

18. Шамилева Л. Л. Статистическое моделирование и прогнозирование: курс лекций : учеб. пособ. / Л. Л. Шамилева. – Донецк : Каштан, 2008. – 310 с.

19. Dickey D. A. Distribution of the estimators for autoregressive time series with a unit root / D. A. Dickey, W. A. Fuller // Journal of the American Statistical Association. – 1979. – Vol. 74. – P. 427–431.

20. Fisher R. A. Statistical methods and scientific induction / R. A. Fisher // Journal of the Royal Statistical Society. – 1955. – No. 17. – P. 69–78.

21. Gencay R. Differentiating intraday seasonalities through wavelet multiscaling / R. Gencay, F. Selcuk, B. Whitcher // Physica A., 2001. – No. 289. – P. 543–556.

22. Granger C. W. J. Forecasting economic time series / C. W. J. Granger, P. Newbold. – 2nd ed. – New York : Academic Press, 1986 – 324 p.

23. Granger C. W. Time series modelling and interpretation / C. W. Granger, M. J. Morris // Journal of the Royal Statistical Society. – 1976. – Ser. A, vol. 139, part 2. – P. 234–256.

24. Lachenbruch P. A. Discriminant Analysis / P. A. Lachenbruch. – New York : Hafner, 1974. – 234 p.

Інформаційні ресурси

25. Офіційний сайт департаменту статистики Організації Об'єднаних Націй. – Режим доступу : <http://unstats.un.org/unsd/default.htm>.

26. Офіційний сайт державної служби статистики України. – Режим доступу : <http://www.ukrstat.gov.ua>.

27. Офіційний сайт Міжнародного валютного фонду. – Режим доступу : <http://www.imf.org>.

28. Офіційний сайт Світового банку. – Режим доступу : <http://web.worldbank.org>.

29. Ратникова Т. А. Введение в эконометрический анализ панельных данных : лекционные и методические материалы [Электронный ресурс] / Т. А. Ратникова. – Режим доступа : http://uisrussia.msu.ru/docs/nov/hse_ejournal/2006/2/10_02_06.pdf.

30. Электронный учебник StatSoft [Электронный ресурс]. – Режим доступа : <http://www.statsoft.ru>.

Зміст

Вступ	3
Змістовий модуль 3. Моделювання часових процесів	4
Тема 9. Моделювання і прогнозування тенденцій розвитку	4
Лабораторна робота 5. Побудова моделі декомпозиції часових рядів	4
Тема 10. Моделювання повних циклів і процесів оновлення	23
Лабораторна робота 6. Оцінювання наявності автокореляції залишків моделі. Усунення автокореляції.....	23
Тема 11. Моделі адаптивного прогнозування та інтегрована модель авторегресії	38
Лабораторна робота 7. Прогнозування часових рядів за допомогою експоненціального згладжування.....	38
Змістовий модуль 4. Моделювання та прогнозування багатомірних процесів	46
Тема 12. Модель об'єкто-періодів.....	46
Лабораторна робота 8. Побудова моделі панельних даних.....	46
Тема 13. Моделювання причинних комплексів і рекурентні моделі.....	52
Лабораторна робота 9. Побудова систем структурних рівнянь.....	52
Тема 14. Модель головних компонент	62
Лабораторна робота 10. Побудова моделі факторного аналізу	62
Тема 15. Кластерний аналіз	74
Лабораторна робота 11. Використання кластерного аналізу для дослідження економічних процесів.....	74
Тема 16. Дискримінантний аналіз	87
Лабораторна робота 12. Вирішення завдання класифікації методом дискримінантного аналізу	87
Рекомендована література	101
Основна.....	101
Додаткова.....	102
Інформаційні ресурси.....	103

НАВЧАЛЬНЕ ВИДАННЯ

**СТАТИСТИЧНЕ МОДЕЛЮВАННЯ
ТА ПРОГНОЗУВАННЯ.
ЗМІСТОВІ МОДУЛІ 3 ТА 4**

**Методичні рекомендації
до лабораторних
та самостійних робіт студентів
спеціальності 051 "Економіка"
першого (бакалаврського) рівня**

Самостійне електронне текстове мережеве видання

Укладачі: **Раєвнєва** Олена Валентинівна
Бровко Ольга Іванівна
Дериховська Вікторія Ігорівна

Відповідальний за видання *О. В. Раєвнєва*

Редактор *А. С. Ширініна*

Коректор *Н. В. Завгородня*

План 2022 р. Поз. № 165 ЕВ. Обсяг 105 с.

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

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