

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

## **БІОМЕТРІЯ**

Методичні вказівки  
до виконання комп'ютерного практикуму  
спеціальність 152 «Метрологія та інформаційно-вимірвальна техніка»

*Рекомендовано вченою радою  
приладобудівного факультету  
(протокол № 6/17 від 21.09.2017 р.)*

Київ  
КПІ ім. Ігоря Сікорського  
2017

**Біометрія:** Методичні вказівки до виконання комп'ютерного практикуму з дисципліни «Біометрія» для студентів спеціальності 152 «Метрологія та інформаційно-вимірювальна техніка» усіх форм навчання [Електронне видання] / Уклад.: С.П. Вислоух, К.С. Барандич, Волошко О.В.. – К.: КПІ ім. Ігоря Сікорського, 2017. – 106 с.

Методичні вказівки призначено для студентів спеціальності 152 «Метрологія та інформаційно-вимірювальна техніка» усіх форм навчання. У методичних вказівках наведено мету, завдання, основні теоретичні відомості практичних робіт, їх зміст та обсяг.

Навчальне електронне мережне видання

## БІОМЕТРІЯ

Методичні вказівки

до виконання комп'ютерного практикуму

для студентів спеціальності

152 «Метрологія та інформаційно-вимірювальна техніка»

Укладачі: *Вислоух Сергій Петрович, к.т.н, доцент*  
*Барандич Катерина Сергіївна, асистент*  
*Волошко Оксана Вячеславівна, асистент*

Відповідальний  
редактор: *Тимчик Г.С., д.т.н., професор*

Рецензент: *Гераїмчук М. Д., д.т.н., професор*

*За редакцією укладачів*

## ЗМІСТ

ЗАГАЛЬНІ ПОЛОЖЕННЯ.....	4
Комп'ютерний практикум №1.....	5
ХАРАКТЕРИСТИКИ ВИПАДКОВИХ ВЕЛИЧИН	
Комп'ютерний практикум №2.....	10
ПЕРЕВІРКА ГІПОТЕЗ ПРО РОЗТАШУВАННЯ ТА РОЗСІЮВАННЯ	
Комп'ютерний практикум №3.....	26
ПЕРЕВІРКА ГІПОТЕЗ ПРО НАЯВНІСТЬ ЗВ'ЯЗКУ МІЖ ЗМІННИМИ .....	
Комп'ютерний практикум №4.....	35
ФАКТОРНИЙ АНАЛІЗ ДАНИХ.....	
Комп'ютерний практикум №5.....	43
КЛАСТЕРНИЙ АНАЛІЗ ДАНИХ.....	
Комп'ютерний практикум №6.....	56
ДИСКРИМІНАТНИЙ АНАЛІЗ ДАНИХ .....	
Комп'ютерний практикум №7.....	65
РЕГРЕСІЙНИЙ АНАЛІЗ ДАНИХ .....	
Комп'ютерний практикум №8.....	83
НЕЙРОМЕРЕЖЕВЕ ДОСЛІДЖЕННЯ ДАНИХ .....	
СПИСОК ЛІТЕРАТУРИ.....	94
ДОДАТОК. ФРАГМЕНТИ СТАТИСТИЧНИХ ТАБЛИЦЬ .....	96

## ЗАГАЛЬНІ ПОЛОЖЕННЯ

Метою виконання комп'ютерного практикуму є отримання практичних навичок розв'язання задач аналізу медико-біологічної інформації з використанням методів багатомірного статистичного аналізу.

Практикум виконуються з використанням популярних систем для статистичних розрахунків, а саме: MS Excel, MathCad, MathLab, SPSS STATISTICS тощо.

Виконання комп'ютерного практикуму включає ознайомлення з метою кожної роботи, його робочим завданням й вивчення теоретичного матеріалу, які проводяться до початку занять. Під час підготовки до виконання кожної із робіт студентам потрібно визначити порядок розв'язання задачі і підготувати початкові (вихідні) дані для виконання розрахунків згідно з індивідуальним завданням.

У ході занять студенти виконують всі пункти робочого завдання та показують викладачеві, який перевіряє правильність їх виконання. За результатами виконання практичної роботи оформляється звіт згідно з вимогами, що зазначені в умові кожної роботи, та надається викладачеві на детальну перевірку на поточному занятті або на наступний день (але не пізніше встановлених строків). Викладач має право повернути недбало оформлену роботу або роботу, що має похибки, або не зарахувати її, якщо теоретична підготовка є недостатньою.

Звіт про виконання роботи має включати мету роботи, умову задачі, що вибрана згідно з варіантом індивідуального завдання, перелік змінних та констант, які використовуються в роботі, результати рішення поставленої задачі, аналіз результатів роботи та висновки за основними пунктами робочого завдання.

Встановлено наступні вимоги до комп'ютерного оформлення матеріалів роботи.

Текст набирається в текстовому редакторі Microsoft Word 2003. Шрифт – Times New Roman. Міжрядковий інтервал – 1,5. Абзац – 1,25 см. Кегль – 14. Вирівнювання основного тексту – по ширині, без переносів. Поля: зліва – 25, справа – 10, зверху та знизу – 20. Рисунки, формули, діаграми, таблиці розмішувати по мірі появи в тексті по центру сторінки. Формули набираються в Microsoft Equation 3.0.

## Комп'ютерний практикум №1

### ХАРАКТЕРИСТИКИ ВИПАДКОВИХ ВЕЛИЧИН

**Мета:** отримання практичних навиків визначення основних характеристик випадкової величини з побудовою відповідних графіків та гістограм в середовищі MS Excel.

#### Теоретичні відомості

Розрахунки виконуються з використанням майстра функцій ( $f_x$ ) MS Excel. Більшість із характеристик випадкових величин, що описані в практичній роботі № 1, можуть бути обчислені за допомогою стандартних функцій MS Excel, які представлені в табл.1.1.

Таблиця 1.1. Опис стандартних функцій MS Excel, що використовуються для розрахунку характеристик випадкових величин.

Характеристики випадкових величин	Назва функції	Опис параметрів
Середнє арифметичне	СРЗНАЧ	Перелік значень або інтервал імен комірок
Середнє геометричне	СРГЕОМ	Перелік значень або інтервал імен комірок
Середнє гармонійне	СРГАРМ	Перелік значень або інтервал імен комірок
Мода	МОДА	Перелік значень або інтервал імен комірок
Медіана	МЕДИАНА	Перелік значень або інтервал імен комірок
Дисперсія	ДИСП	Перелік значень або інтервал імен комірок
Середньоквадратичне відхилення	СТАНДОТКЛОН	Перелік значень або інтервал імен комірок
Напівширина довірчого інтервалу	ДОВЕРИТ	Рівень значимості, середньоквадратичне відхилення, кількість дослідів
Коефіцієнт асиметрії	СКОС	Перелік значень або інтервал імен комірок
Коефіцієнт ексцесу	ЭКСЦЕСС	Перелік значень або інтервал імен комірок

Для варіаційного розмаху, коефіцієнту варіації та довірчого інтервалу немає готових функцій, але це не є проблемою оскільки в рядку формул можна ввести вирази для їх розрахунку.

На рис. 1.1 представлено результати розрахунків характеристик випадкових величин для показників функцій зовнішнього дихання у хлопчиків та дівчаток основної групи спостереження.

Показники функцій зовнішнього дихання у хлопчиків та дівчаток основної групи спостереження																					
N	Год	ст	n	VC		FVC		FEV1		FEV1		PEF		MEF50		MEF25		MVV		Закл	
				b1	%	b2	%	b3	%	b4	%	b5	%	b6	%	b7	%	b8	%		
1	14	1	9	4,76	111	5,4	126	4,08	111	75,5	89,4	7,37	101	5,28	119	2,73	119	43,6	112,7	2	
2	14	2	10	4,76	111	5,4	126	4,08	111	75,5	89,4	7,37	101	5,28	119	2,73	119	51,3	100,5	1	
3	13	2	11	3,2	98,4	3,8	116	3,36	125	88,4	104	5,84	101	5,75	161	4,05	220	55,5	101	1	
4	13	1	6	4,24	91,5	4,92	106	4,32	112	87,8	103	8,13	105	6,87	146	4,74	195	54	67,2	3	
5	11	1	10	2,68	103	3,24	124	2,88	135	88,8	105	4,8	100	3,24	114	2,01	130	57,4	134	1	
6	10	2	5	2,76	110	2,96	118	2,52	123	85,1	100	3,92	85,1	3,52	121	2,24	150	59,4	101	1	
7	14	1	12	4,12	78,6	4,48	85,5	4,36	99,5	97,3	115	7,84	89	5,1	98,5	3,94	146	83,7	91,1	2	
8	8	1	5	1,8	86,5	2,04	98	1,84	109	90,1	106	3,76	95,3	2,31	91,9	1,09	87,7	46,2	138	1	
9	11	1	3	1,88	69,5	1,8	66,6	1,68	74,7	86,3	105	2,7	51,5	2,07	63,2	1,11	66	54	79,1	3	
10	9	1	9	1,96	87,1	2,04	90,6	1,8	96,8	88,2	104	3,37	75,7	2,51	89,2	1,2	83,2	33,1	81,9	2	
11	13	1	3	2,24	85,9	2,44	93,5	2,2	101	90,1	106	3,01	59,3	2,42	76,2	1,87	114	37,8	85,3	2	
12	11	2	4	2,24	92,3	2,24	92,3	2,04	101	91	107	4,07	85,5	2,47	82,5	1,43	93,1	39,6	92,7	1	
Среднє арифметичне				7,25	3,0533	93,7	3,3967	103,5	2,93	108,3	87,01	102,8	5,182	87,45	3,9	106,8	2,428	127	51,3	98,708	1,667
Среднє геометричне				6,508	2,8696	92,8	3,1513	101,9	2,7547	107,2	86,8	102,6	4,826	85,58	3,59	103,3	2,152	120	49,87	96,718	1,513
Среднє гармонійне				5,77	2,7044	91,9	2,9278	100	2,5908	106	86,58	102,3	4,498	83,39	3,32	99,82	1,913	113	48,53	94,82	1,385
Мода				9	4,76	111	5,4	126	4,08	111	75,5	89,4	7,37	101	5,28	119	2,73	119	54	101	1
Медіана				7,5	2,72	91,9	3,1	102	2,7	110	88,3	104,5	4,435	92,15	3,38	106,3	2,125	119	52,65	96,6	1,5
Варіаційний розмах				9	2,96	41,5	3,6	59,4	2,88	60,3	21,8	25,6	5,43	53,5	4,8	97,8	3,65	154	50,6	70,8	2
Дисперсія				10,57	1,2804	177	1,8529	355,7	1,1209	243,8	37,93	51,5	4,084	301,1	2,74	823,1	1,545	2073	174,6	448,4	0,606
Середньоквадратичне відхилення				3,251	1,1315	13,3	1,3612	18,86	1,0587	15,61	6,158	7,176	2,021	17,35	1,65	28,69	1,243	45,5	13,21	21,175	0,778
Коефіцієнт варіації				44,84	37,059	14,2	40,075	18,22	36,134	14,42	7,078	6,98	39	19,84	42,4	26,86	51,19	35,9	25,76	21,452	46,71
Довірчий інтервал від				5,411	2,4131	86,2	2,6265	92,87	2,331	99,42	83,52	98,76	4,038	77,63	2,97	90,56	1,725	101	43,82	86,727	1,226
до				9,089	3,6936	101	4,1668	114,2	3,529	117,1	90,49	106,9	6,325	97,27	4,84	123	3,132	153	58,78	110,69	2,107
Коефіцієнт асиметрії				-0,001	0,5109	-0,2	0,4056	-0,4	0,2645	-0,39	-0,856	-0,85	0,368	-1,15	0,48	0,422	0,701	0,84	1,145	0,698	0,719
Коефіцієнт ексцесу				-1,694	-1,444	-0,7	-1,512	-0,58	-1,792	1,094	1,09	1,163	-1,66	0,384	-1,32	-0,3	-0,71	0,3	2,553	0,0603	-0,79

Рис. 1.1. Характеристик випадкових величин для показників функцій зовнішнього дихання у хлопчиків та дівчаток основної групи спостереження.

Наступним є впорядкування вибірок за зростанням їх елементів та поділ на інтервали (рис.1.2).

Вибірка впорядкована за зростанням					
	N	FVC		FVC	
		%		%	
	1		126		66,6
	2		126		85,5
	3		116		90,6
	4		106		92,3
	5		124		93,5
	6		118		98
	7		85,5		106
	8		98		116
	9		66,6		118
	10		90,6		124
	11		93,5		126
	12		92,3		126
Ранжирувана вибірка					
Параметр	Інтервали				
FVC, %	66-78,1	78,2-92,1	92,2-104,1	104,2-115,1	115,2-126
Частота	1	2	3	1	5
Накопичена частота	1	3	6	7	12

Рис.1.2. Ранжирувана вибірка FVC, %.

Відповідно до отриманих значень частот на кожному із інтервалів будуватиметься діаграма частот випадкової величини вибірки (рис. 1.3).

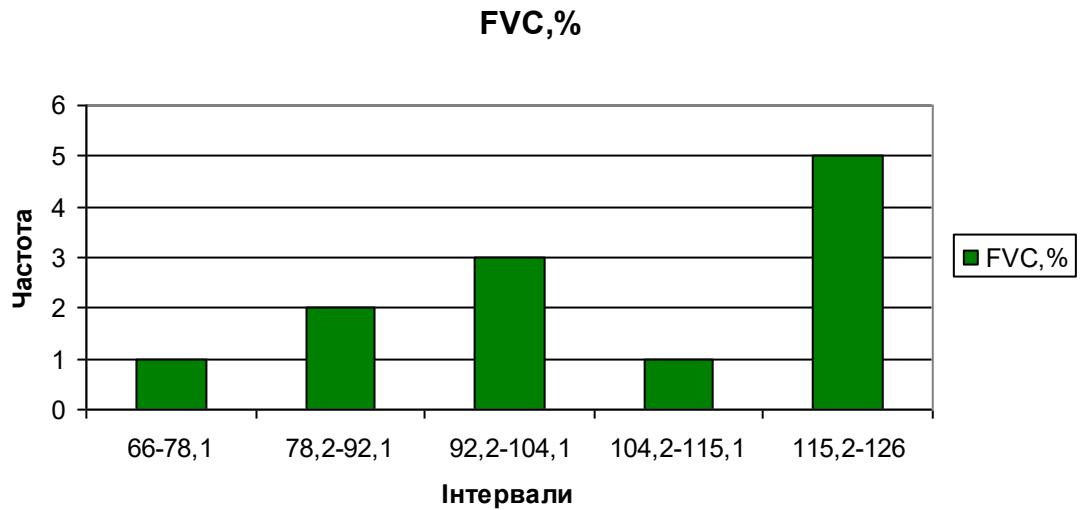


Рис.1.3. Діаграма частот випадкової величини FVC,%.

Для побудови кривої розподілу та функції розподілу випадкових величин вибірки FVC,% створюється таблиця, що показана на рис. 1.4. В стовпчику М представлені розрахунки щільності розподілу змінної FVC,%, а в стовпчику N – розподіл змінної FVC,%.

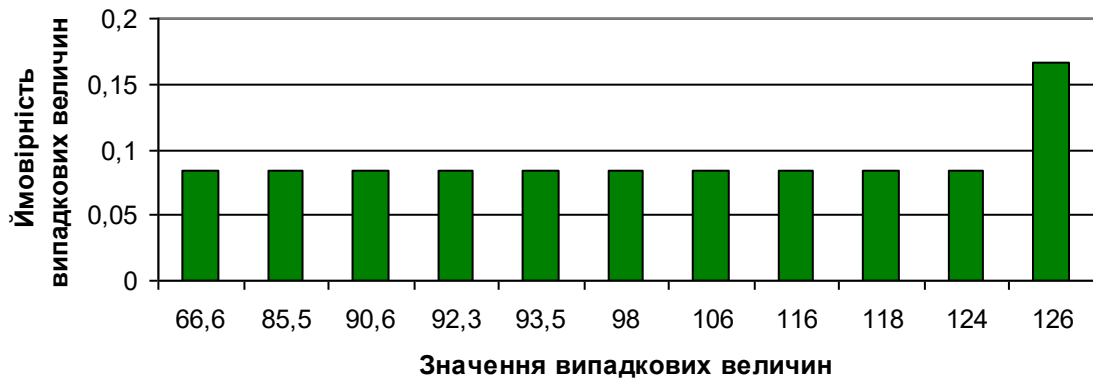
*Побудова гістограм кривої розподілу та функцій розподілу випадкових величин вибірки*

i	FVC,% i	f i	p <sup>i</sup> i=P(X=x i)	p i=P(X<=x i)
1	66,6	1	0,083333333	0,083333333
2	85,5	1	0,083333333	0,166666667
3	90,6	1	0,083333333	0,25
4	92,3	1	0,083333333	0,333333333
5	93,5	1	0,083333333	0,416666667
6	98	1	0,083333333	0,5
7	106	1	0,083333333	0,583333333
8	116	1	0,083333333	0,666666667
9	118	1	0,083333333	0,75
10	124	1	0,083333333	0,833333333
11	126	2	0,166666667	1
	Сума	12	1	

Рис. 1.4. Розрахунки розподілу дискретної величини.

За результатами даних розрахунків будуються відповідні гістограми, як показано на рис. 1.5.

### Крива розподілу випадкових величин (щільність), $p_i$



### Функція розподілу випадкових величин, $P_i$

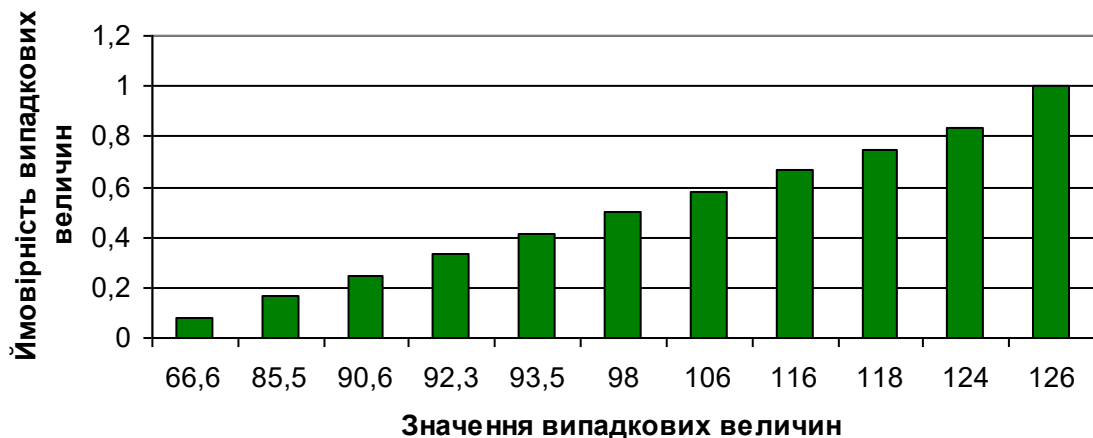


Рис. 1.5. Графіки функцій розподілу дискретної випадкової величини вибірки FVC, %.

### Завдання для виконання комп'ютерного практикуму

Для множини вибірок випадкових величин, що надані в індивідуальному завданні, виконати за допомогою табличного редактора Excel наведені нижче дії для кожної із них.

1. Ранжирувати вибірку за зростанням значень її елементів.
2. Поділити отримана ранжирувану вибірку на ряд приблизно однакових за величиною інтервалів ( не менше 5).
3. Визначити частоти випадкових величин для кожного інтервалу.
4. Визначити накопичені частоти випадкових величин ранжируваної вибірки.
5. Побудувати гістограми кривої розподілу та функції розподілу випадкових величин даної вибірки.
6. Визначити середнє арифметичне, середнє геометричне та середнє гармонійне елементів вибірки.

7. Визначити моду і медіану вибірки.
  8. Визначити варіаційний розмах, дисперсію, середнє квадратичне відхилення та коефіцієнт варіації.
  9. Визначити коефіцієнти асиметрії та ексцесу вибірки.
- Порівняти результати обчислень для кожної вибірки випадкових величин.  
Зробити висновки за результатами обчислень характеристик випадкових величин.

### Контрольні запитання

1. Як здійснюється ранжирування вибірки випадкових величин та поділ її на інтервали?
2. Як визначаються частоти випадкових величин для кожного інтервалу?
3. Як визначити накопичені частоти випадкових величин?
4. Що визначає щільності розподілу випадкових величин? Як її очислити?
5. Як визначаються значення середнього арифметичного, середнього геометричного та середнього гармонійного елементів вибірки? Який статистичний зміст вказаних характеристик випадкових величин?
6. Що характеризує мода і медіана вибірки випадкових величин? Як визначаються мода й медіана?
7. Що таке дисперсія та середньоквадратичне відхилення випадкових величин? Як вони визначаються?
8. Як визначити варіаційних розмах та коефіцієнт варіації?
9. Що таке довірчий інтервал? Як визначити довірчий інтервал для середнього?
10. Яка характеристика показує ступінь несиметричності розподілу ймовірностей дійсної випадкової величини в порівнянні з кривою нормальною розподілу?
11. Що характеризує коефіцієнт ексцесу?
12. Які характеристики випадкових величин можна визначити з використанням майстра функцій ( $f_x$ ) MS Excel?
13. Як розрахувати коефіцієнт варіації та довірчий інтервал в MS Excel?
14. Як побудувати криву розподілу та функції розподілу в MS Excel?

## Комп'ютерний практикум №2

### ПЕРЕВІРКА ГІПОТЕЗ ПРО РОЗТАШУВАННЯ ТА РОЗСІЮВАННЯ

**Мета:** отримання практичних навиків перевірки гіпотез про розташування та розсіювання за допомогою MS Excel та SPSS Statistics.

#### Теоретичні відомості

Розрахунки виконуються з використанням сервісу «Анализ данных» MS Excel та SPSS Statistics. Для підключення сервісу «Анализ данных» для офісу MS Excel 2003 необхідно вибрати команду головного меню Сервис→Настройки... У вікні «Настройки» (рис.2.1) вибираємо «Пакет анализа».

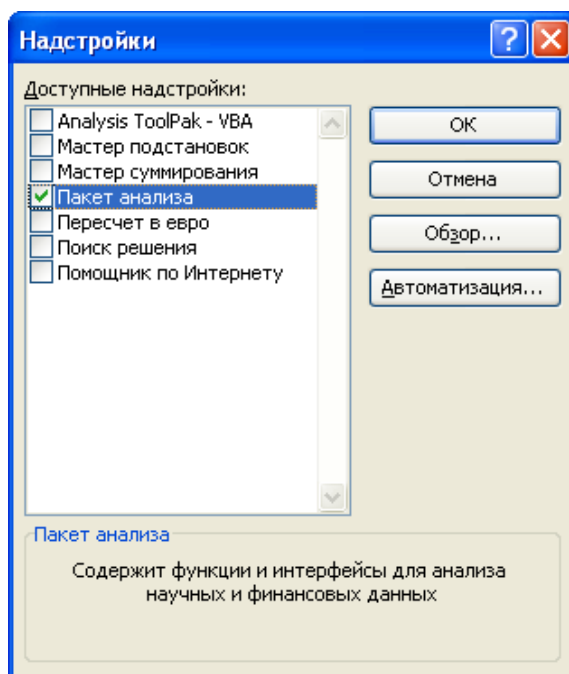


Рис.2.1. Вікно «Настройки» для офісу MS Excel 2003.

Для офісу 2010 необхідно вибрати команду головного меню Файл→Параметры. У вікні «Параметры Excel» вибираємо «Настройки» та у вікні «Настройки» вибираємо «Пакет анализа» (рис.2.2).

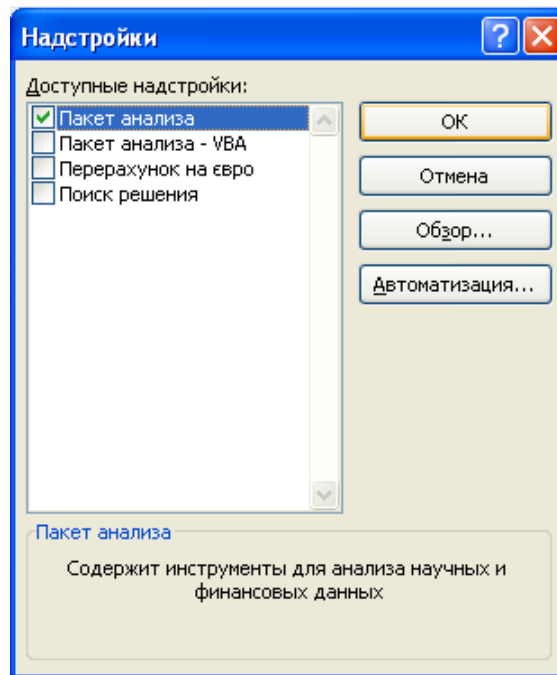


Рис.2.2. Вікно «Настройки» для офісу MS Excel 2010.

Як приклад перевірки гіпотез про положення та розсіювання використаємо вибірку, що складається із 3 груп випадкових величин, які представлено в таблиці 2.1.

Таблица 2.1. Початкові дані для перевірки гіпотез.

X1	X2	X3
84	99	78
90	108	112
92	89	98
77	86	82
69	90	77
94	85	81
89	79	85
99	142	95
75	126	95
78	95	88
62	79	82
79	85	87
92	104	99
69	112	89

Перевірка гіпотез про розташування та розсіювання виконується в такій послідовності.

1. Перевірка гіпотези відповідності вибірок нормальному закону розподілу випадкових величин.

Вказана перевірка виконується відповідно до способу, який представлено в теоретичних відомостях до практичної роботи №2. Результати перевірки

гіпотези про відповідність вибірок нормальному закону розподілу надано на рис. 2.3.

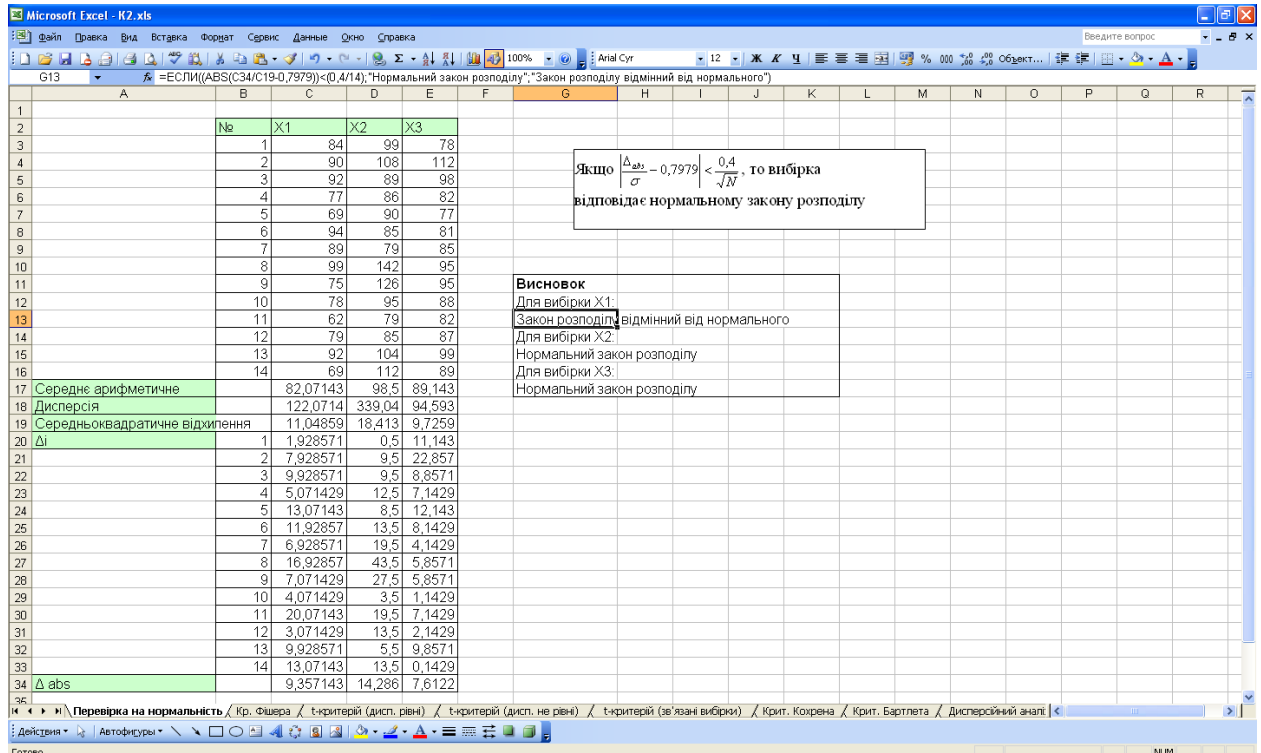


Рис.2.3. Результати перевірки гіпотези відповідності вибірок нормальному закону розподілу.

2. Перевірка гіпотези про рівність двох дисперсій елементів вибірок за критерієм Фішера.

Для виконання цієї перевірки вибирається команда головного меню Сервіс→Анализ данных. У вікні «Анализ данных» використовується підпункт «Двухвыборочный F-тест для дисперсий» (рис.2.4).

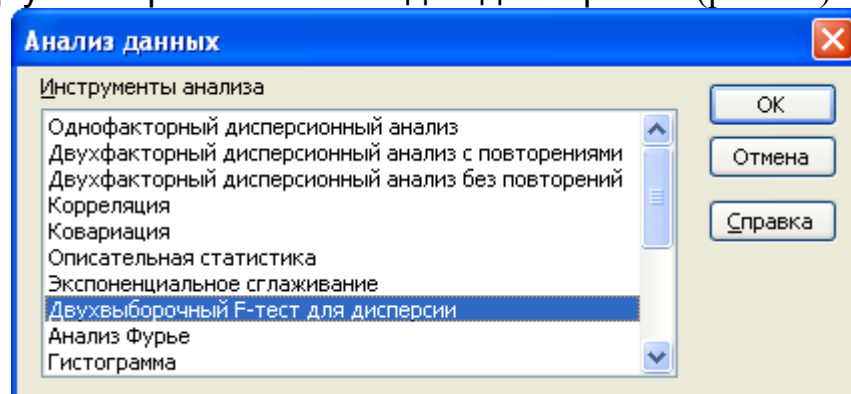


Рис.2.4. Вікно «Анализ данных».

У вікні, що з'явиться (рис. 2.5), задаються інтервали початкових змінних та вихідний інтервал.

Результати перевірки гіпотези про рівність двох дисперсій елементів вибірок за критерієм Фішера з висновками для кожної із пар вибірок представлено на рис. 2.6(а-в).

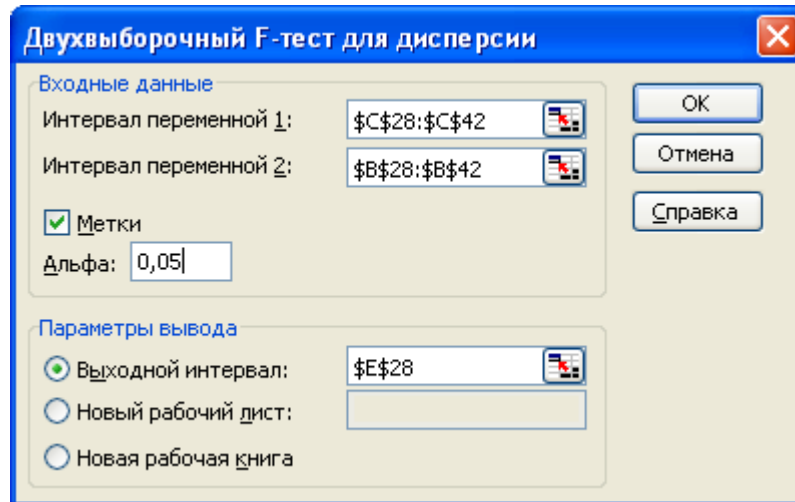
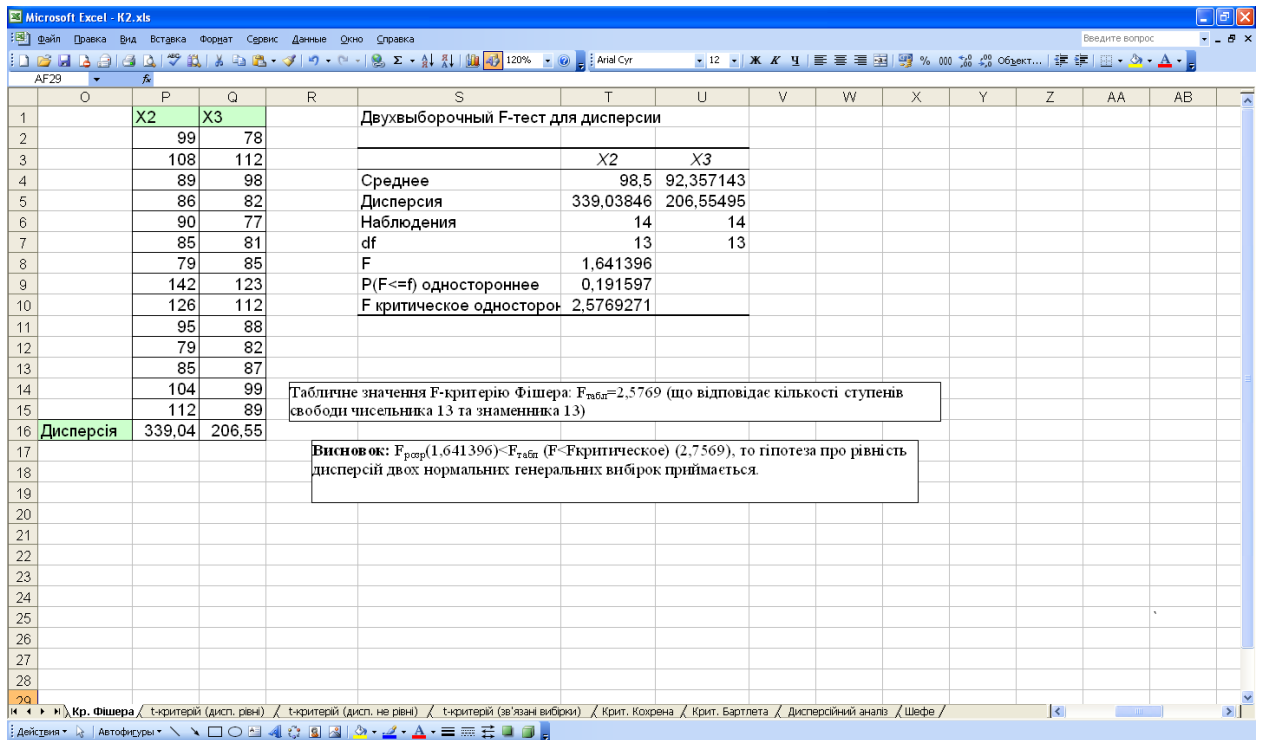


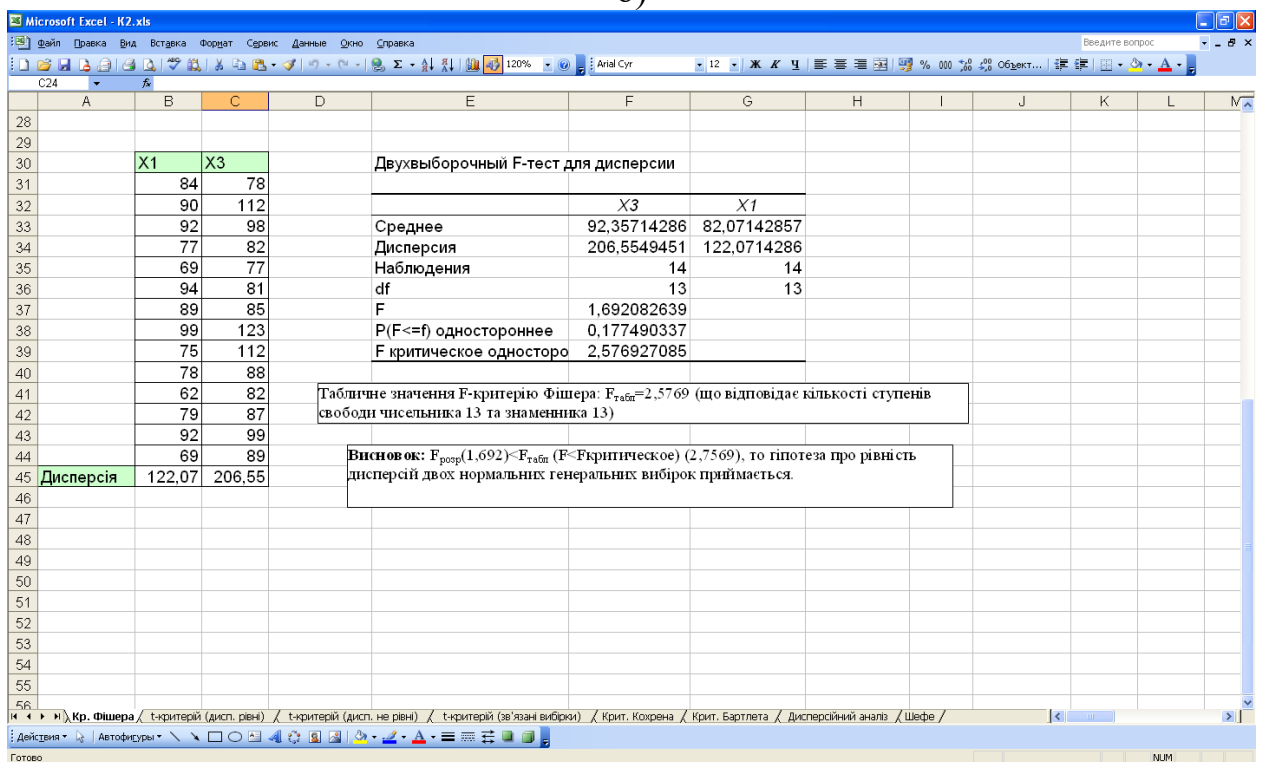
Рис. 2.5. Вікно «Двухвыборочный F-тест для дисперсий».

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		X1	X2		Двухвыборочный F-тест для дисперсии								
2		84	99										
3		90	108			X2	X1						
4		92	89		Среднее	98,5	82,07142857						
5		77	86		Дисперсия	339,0384615	122,0714286						
6		69	90		Наблюдения	14	14						
7		94	85		df	13	13						
8		89	79		F	2,777377684							
9		99	142		P(F<=f) одностороннее	0,038353618							
10		75	126		F критическое односторо	2,576927085							
11		78	95										
12		62	79										
13		79	85		Табличные значения F-критерия Фишера: $F_{табл}=2,5769$ (що відповідає кількості ступенів свободи чисельника 13 та знаменника 13)								
14		92	104										
15		69	112										
16		Дисперсія	122,07	339,04		Висновок: $F_{розра}(2,777) > F_{табл}$ ( $F < F_{критическое}$ ) (2,7569), то гіпотеза про рівність дисперсій двох нормальних генеральних вибірок не приймається.							
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													

а)



б)



в)

Рис. 2.6. Вікна з результатами перевірки гіпотези про рівність двох дисперсій елементів вибірок за критерієм Фішера: а) вибірок змінних X1 та X2, б) вибірок змінних X2 та X3, в) вибірок змінних X1 та X3

3. Перевірка гіпотези про однорідність дисперсій елементів вибірок за критерієм Кохрена.

Перевірка цієї гіпотези виконується шляхом внесення відповідних формул у рядок редактора формул MS Excel.

Результати перевірки гіпотези про однорідність дисперсій елементів 3 вибірок за критерієм Кохрена представлено на рис. 2.7.

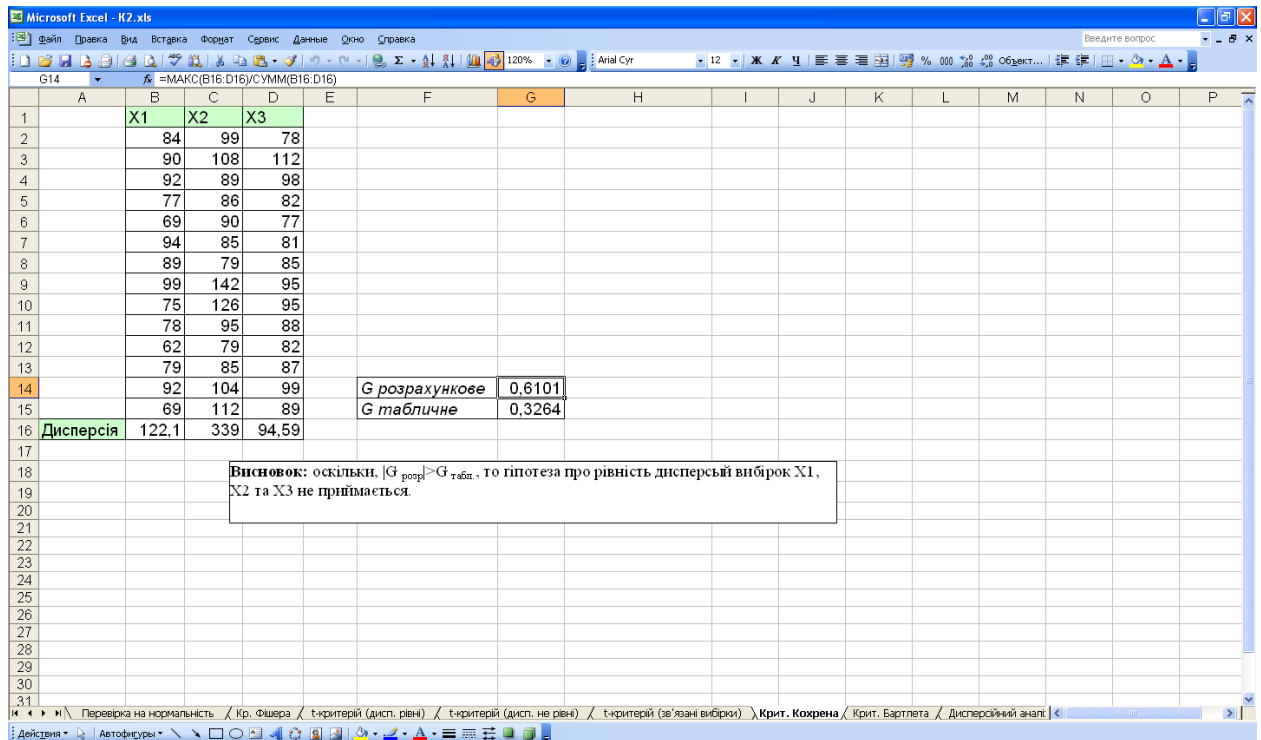


Рис. 2.7. Вікно з результатами перевірки гіпотези про однорідність дисперсій елементів вибірок за критерієм Кохрена.

4. Перевірка гіпотези про однорідність декількох дисперсій елементів вибірок за критерієм Бартлета.

Перевірка гіпотези виконується шляхом внесення відповідних формул у рядок редактора формул MS Excel.

Результати перевірки гіпотези за критерієм Бартлета представлено на рис. 2.8.

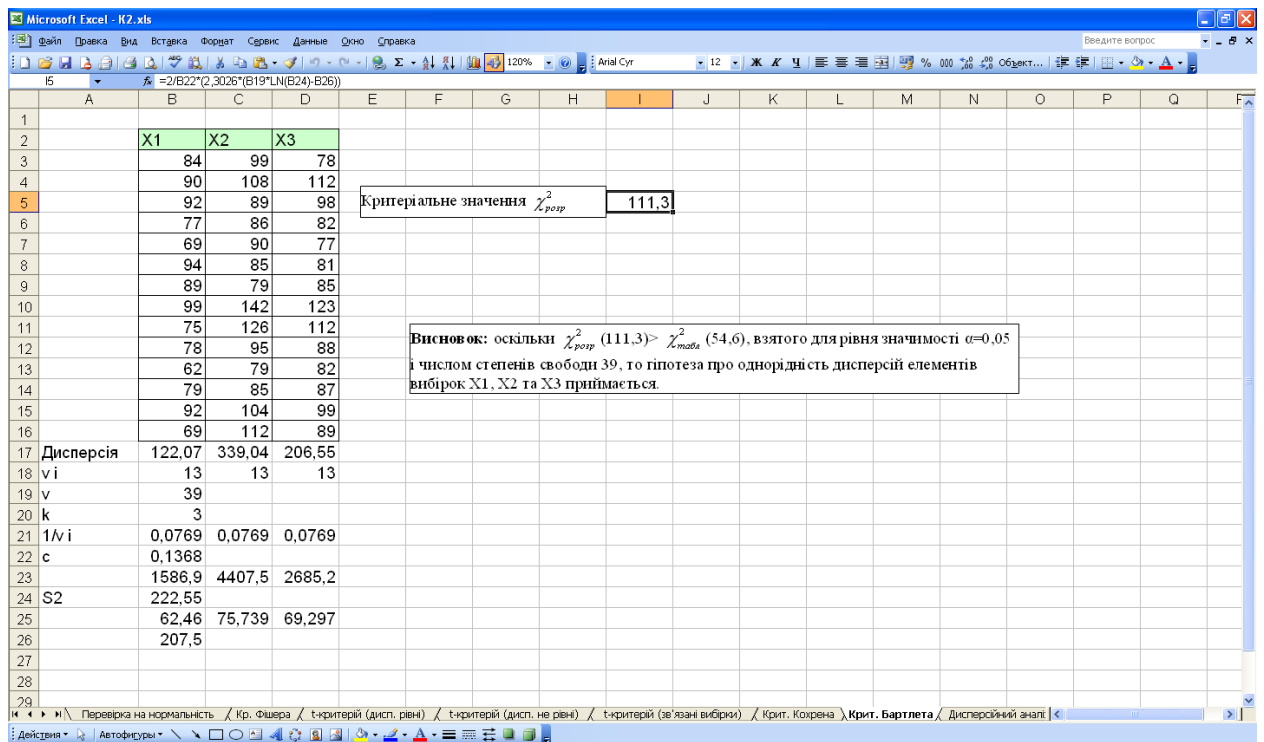
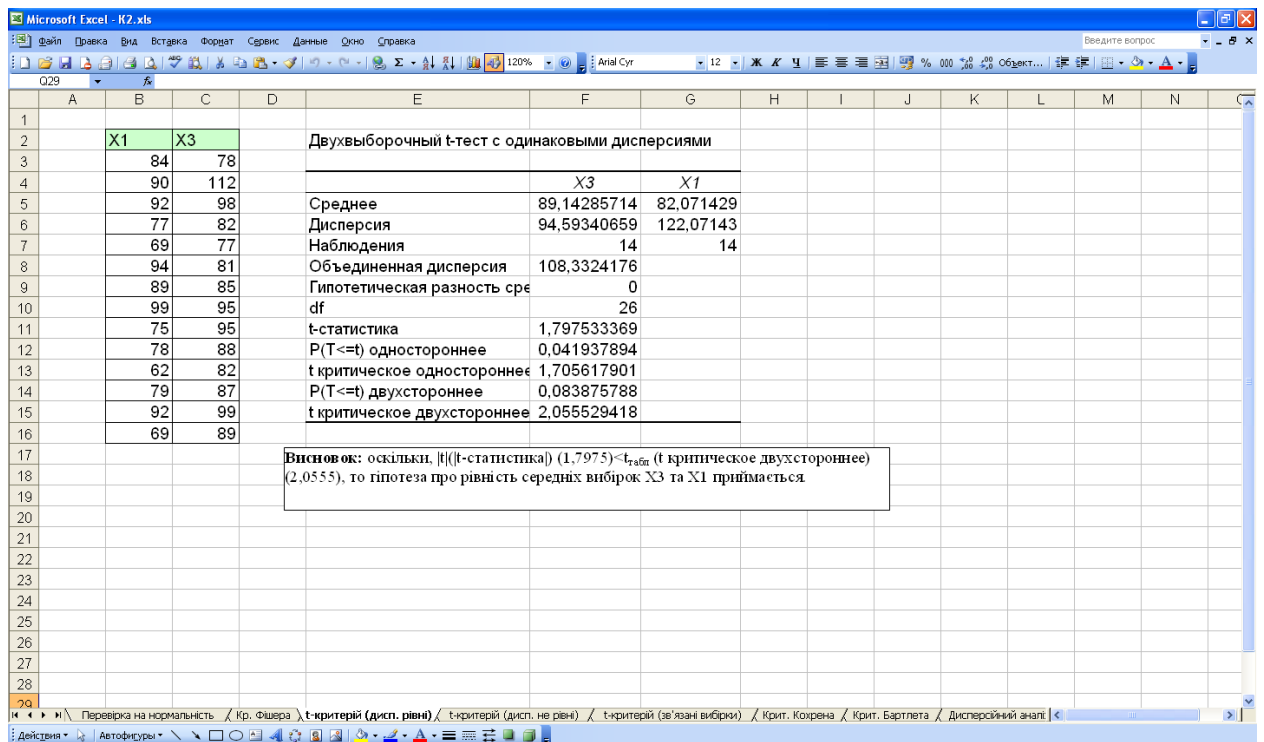


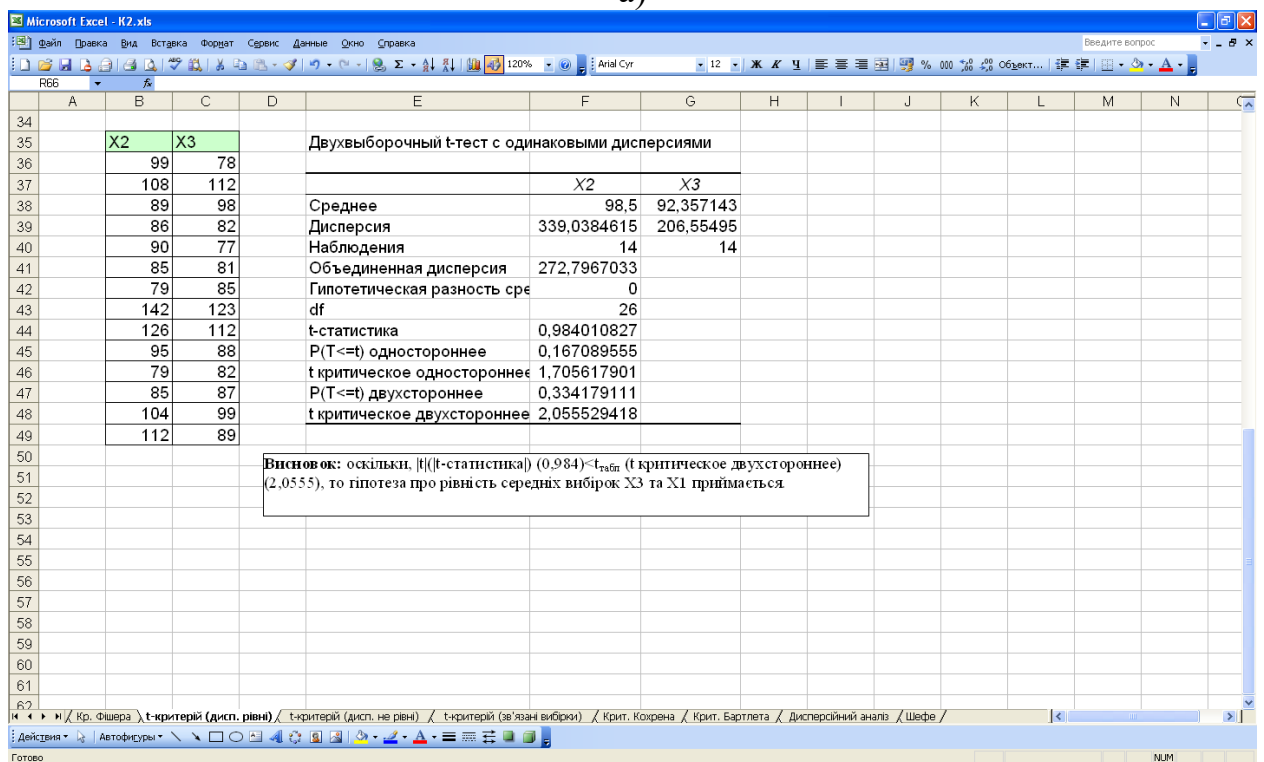
Рис. 2.8. Вікно з результатами перевірки гіпотези про однорідність декількох дисперсій елементів вибірок за критерієм Бартлета.

5. Перевірка гіпотези про рівність середніх значень вибірок при однакових дисперсіях.

Відповідно до отриманих в пункті 2 результатів перевірки гіпотез про рівність двох дисперсій елементів вибірок за критерієм Фішера встановлено, що для вибірок X1 та X3, X2 та X3 гіпотеза приймається (рис. 2.6). Тому в цьому випадку для перевірки вказаної гіпотези про рівність середніх необхідно використати «Двухвиборочний t-тест с однаковими дисперсиями» з підпункту в вікні «Анализ данных». Результати перевірки гіпотез за критерієм Бартлета наведено на рис. 2.9.



а)



б)

Рис. 2.9. Вікна з результатами перевірки гіпотези про рівність середніх значень вибірок при однакових дисперсіях: а) вибірок змінних X1 та X3; б) вибірок змінних X2 та X3

6. Перевірка гіпотези про рівність середніх значень вибірок без припущення про дисперсії.

Для перевірки вказаної гіпотези використовуємо систему SPSS Statistics. При цьому вводяться початкові дані в вигляді стовпчика X та номер групи випадкових величин в стовпчику N (рис 2.10).

The screenshot shows the IBM SPSS Statistics data editor window for a file named 'K2.sav'. The window displays a data grid with two columns: 'N' and 'X'. The data is organized into three groups based on the value of 'N'.

Case	N	X	пер	пер	пер	пер	пер
1	1,00	84,00					
2	1,00	90,00					
3	1,00	92,00					
4	1,00	77,00					
5	1,00	69,00					
6	1,00	94,00					
7	1,00	89,00					
8	1,00	99,00					
9	1,00	75,00					
10	1,00	78,00					
11	1,00	62,00					
12	1,00	79,00					
13	1,00	92,00					
14	1,00	69,00					
15	2,00	99,00					
16	2,00	108,00					
17	2,00	89,00					
18	2,00	86,00					
19	2,00	90,00					
20	2,00	85,00					
21	2,00	79,00					
22	2,00	142,00					
23	2,00	126,00					
24	2,00	95,00					
25	2,00	79,00					
26	2,00	85,00					
27	2,00	104,00					
28	2,00	112,00					
29	3,00	78,00					
30	3,00	112,00					

Рис.2.10. Вікно з початковими даними для перевірки гіпотези про рівність середніх значень.

Далі вибирається команда головного меню **Анализ**→**Сравнение средних**→**Т-критерий для независимых выборок...** У вікні «Т-критерий для независимых выборок» (рис.2.11) вибирається змінна, що перевіряється, задається змінна N для групування та задаються порівнювані групи, як це показано на рис. 2.12.

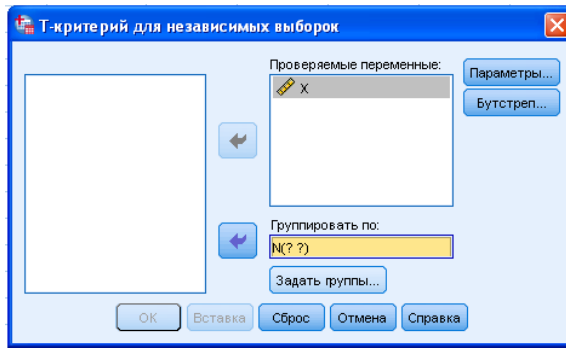


Рис.2.11. Вікно «Т-критерий для независимых выборок».

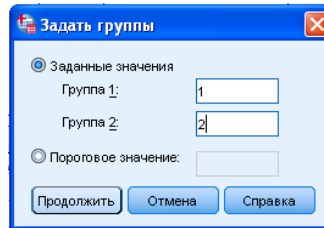


Рис.2.12. Вікно «Задать группы».

Выбором пунктов Продолжить→ОК генерується створення файлу виведення результатів обчислення (рис. 2.13).

**ANOVA**

	Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
Между группами	1901,476	2	950,738	5,133	,011
Внутри групп	7224,143	39	185,234		
Всего	9125,619	41			

T-TEST GROUPS=N(1 2)  
/MISSING=ANALYSIS  
/VARIABLES=X  
/CRITERIA=CI (.95).

**Т-критерий**

**Статистика группы**

	N	N	Среднее значение	Стандартная отклонения	Среднекв. ошибка среднего
X	1,00	14	82,0714	11,04859	2,95286
	2,00	14	98,5000	18,41300	4,92108

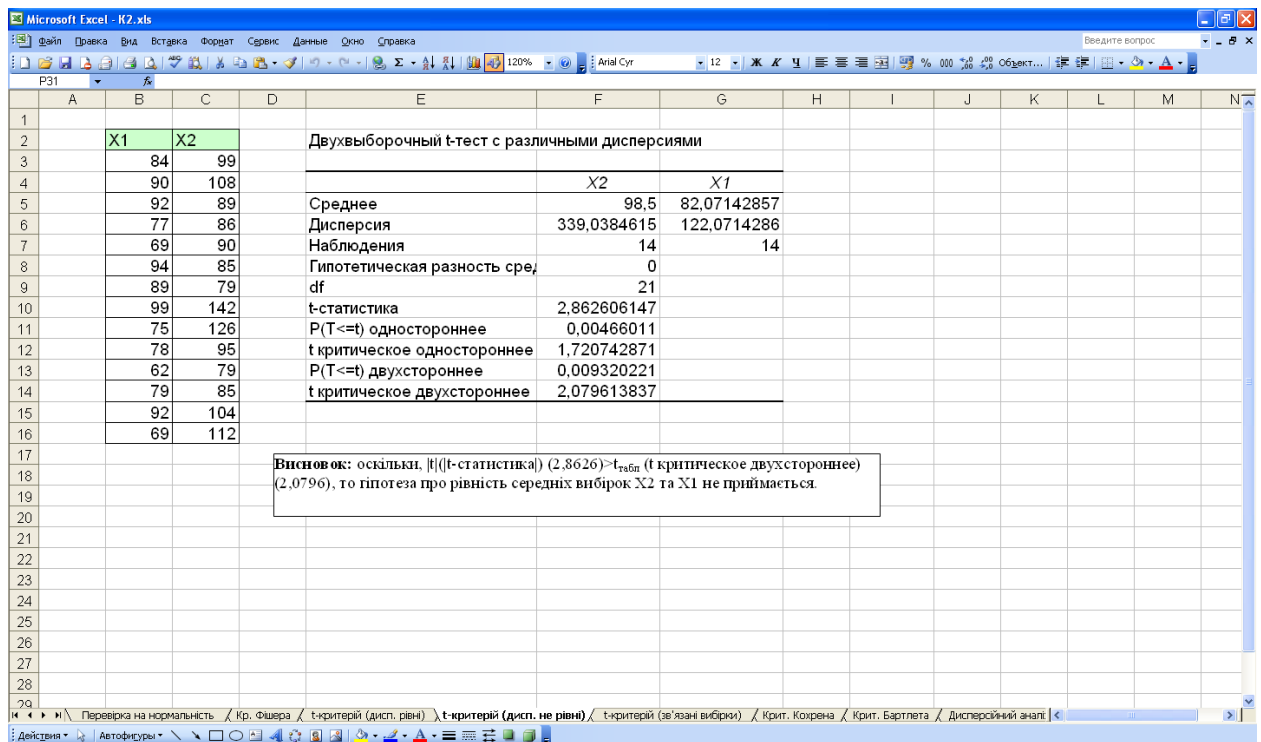
**Критерий для независимых выборок**

	Предполагается равные дисперсии Не предполагается равные дисперсии	Критерий равенства дисперсий Левина		t-критерий для равенства средних						
		F	Знач.	t	ст.св.	Знач. (2-х сторонняя)	Разность средних	Среднекв. атакная ошибка разности	95% доверительный интервал для разности	
									Нижняя	Верхняя
X		2,313	,140	-2,863	26	,008	-16,42857	5,73903	-28,22531	-4,63183
				-2,863	21,287	,009	-16,42857	5,73903	-28,35374	-4,50340

Рис. 2.13. Вікно з результатами перевірки гіпотези про рівність середніх без передбачень про дисперсії.

7. Перевірка гіпотези про рівність середніх значень вибірок при різних дисперсіях.

Відповідно до отриманих в пункті 2 результатів перевірки гіпотези про рівність двох дисперсій елементів вибірок за критерієм Фішера визначено, що для вибірок  $X_1$  та  $X_2$  гіпотеза не приймається. Перевірку рівності середніх значень цих вибірок виконується засобами MS Excel. Для цього вибирається підпункт «Двухвыборочный t-тест с различными дисперсиями» у вікні «Анализ данных». Результати перевірки гіпотези представлено на рис. 2.14.



	X1	X2		X2	X1
84	99	Двухвыборочный t-тест с различными дисперсиями			
90	108				
92	89		Среднее	98,5	82,07142857
77	86		Дисперсия	339,0384615	122,0714286
69	90		Наблюдения	14	14
94	85		Гипотетическая разность сре	0	
89	79		df	21	
99	142		t-статистика	2,862606147	
75	126		P(T<=t) одностороннее	0,00466011	
78	95		t критическое одностороннее	1,720742871	
62	79		P(T<=t) двухстороннее	0,009320221	
79	85		t критическое двухстороннее	2,079613837	
92	104				
69	112				

**Выводок:** оскільки,  $|t|$  (t-статистика) (2,8626) >  $t_{табл}$  (t критическое двухстороннее) (2,0796), то гіпотеза про рівність середніх вибірок  $X_2$  та  $X_1$  не приймається.

Рис. 2.14. Вікно з результатами перевірки гіпотези про рівність середніх значень вибірок при різних дисперсіях.

8. Перевірка гіпотези про рівність середніх значень вибірок заданому значенню  $A$ .

Для перевірки цієї гіпотези використовується система SPSS Statistics. На рис. 2.15 представлено вікно з внесеними початковими даними.

	X1	X2	X3	пер	пер	пер	пер	пер	пер	пер	г
1	84,00	99,00	78,00								
2	90,00	108,00	112,00								
3	92,00	89,00	98,00								
4	77,00	86,00	82,00								
5	89,00	90,00	77,00								
6	94,00	85,00	81,00								
7	89,00	79,00	85,00								
8	99,00	142,00	95,00								
9	75,00	126,00	95,00								
10	78,00	95,00	88,00								
11	62,00	79,00	82,00								
12	79,00	85,00	87,00								
13	92,00	104,00	99,00								
14	89,00	112,00	89,00								
15											

Рис. 2.15. Вікно з початковими даними для перевірки гіпотези про рівність середніх значень вибірок заданому значенню  $A$ .

Після вибору команди головного меню **Анализ**→**Сравнение средних**→**Одновыборочный t-критерий** у вікні «Одновыборочный t-критерий» вибираються вибірки змінних, що підлягають перевірці, та задаються значення числа  $A$  в полі «Проверяемое значение», як це показано на рис. 2.16. Натисканням ОК та отримуємо результати перевірки гіпотези в вигляді файлу виведення (рис.2.17).

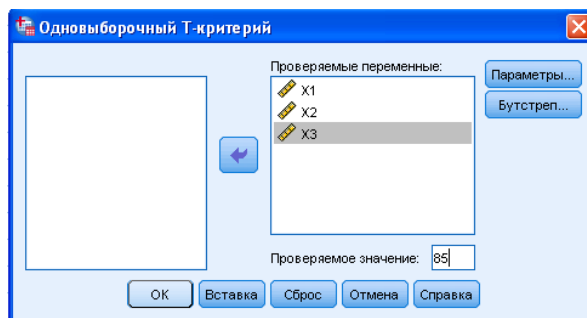


Рис. 2.16. Вікно «Одновыборочный t-критерий».

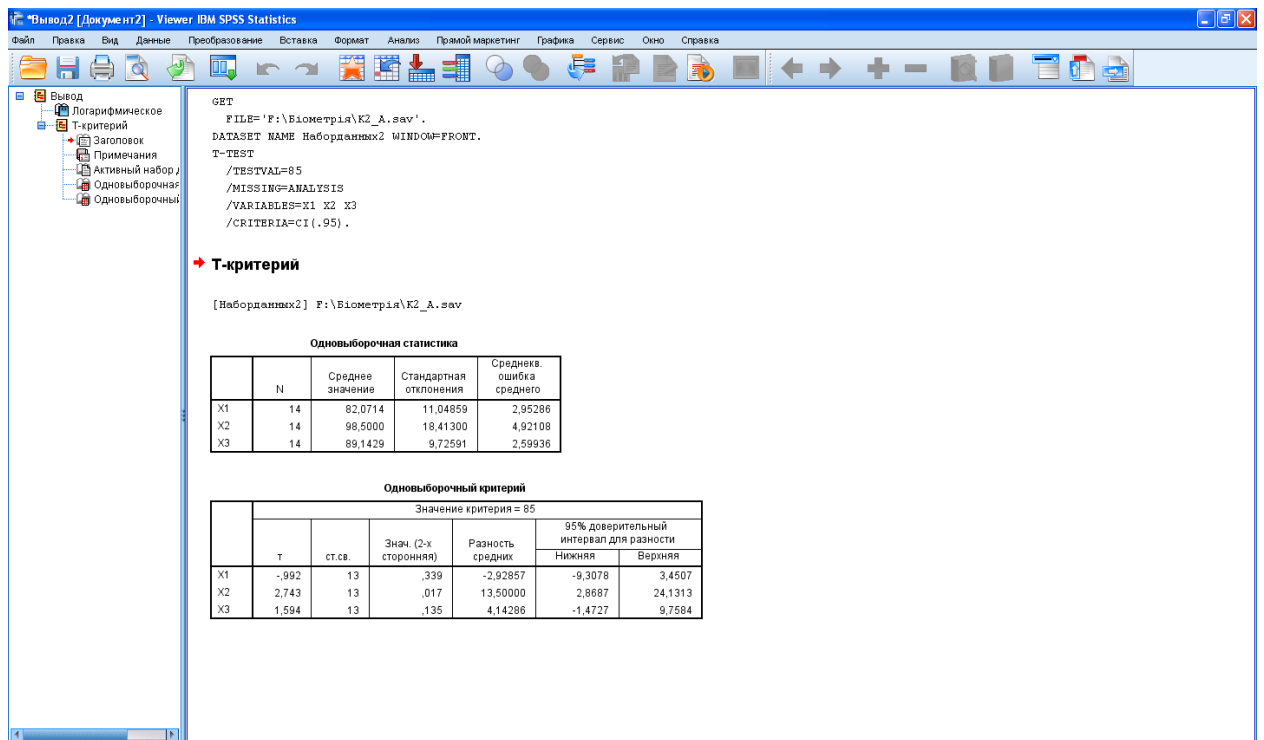


Рис. 2.17. Вікно з файлом виведення результатів обчислення.

9. Перевірка гіпотези про рівність середніх значень для зв'язаних вибірок.

Перевірка цієї гіпотези виконується за допомогою MS Excel. Для цього вибирається «Парный двухвыборочный t-тест для средних» у вікні «Анализ данных». Результати перевірки гіпотези представлено на рис. 2.18.

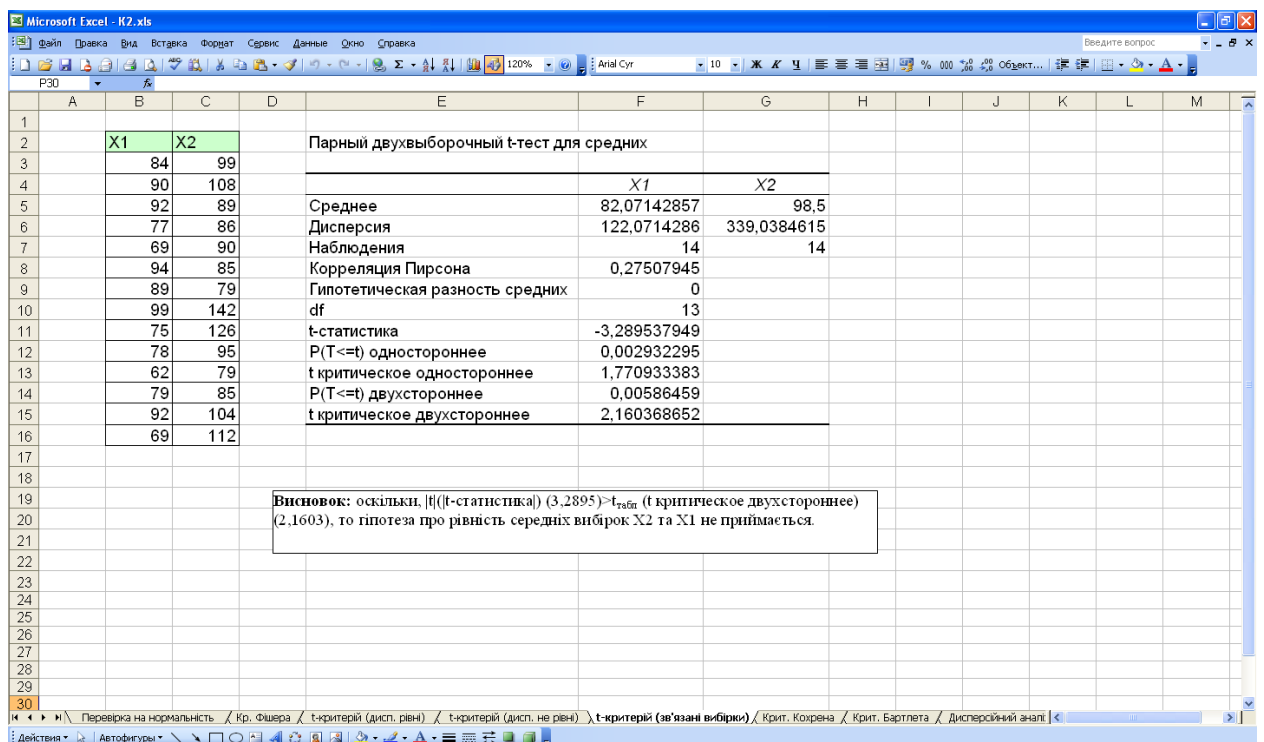
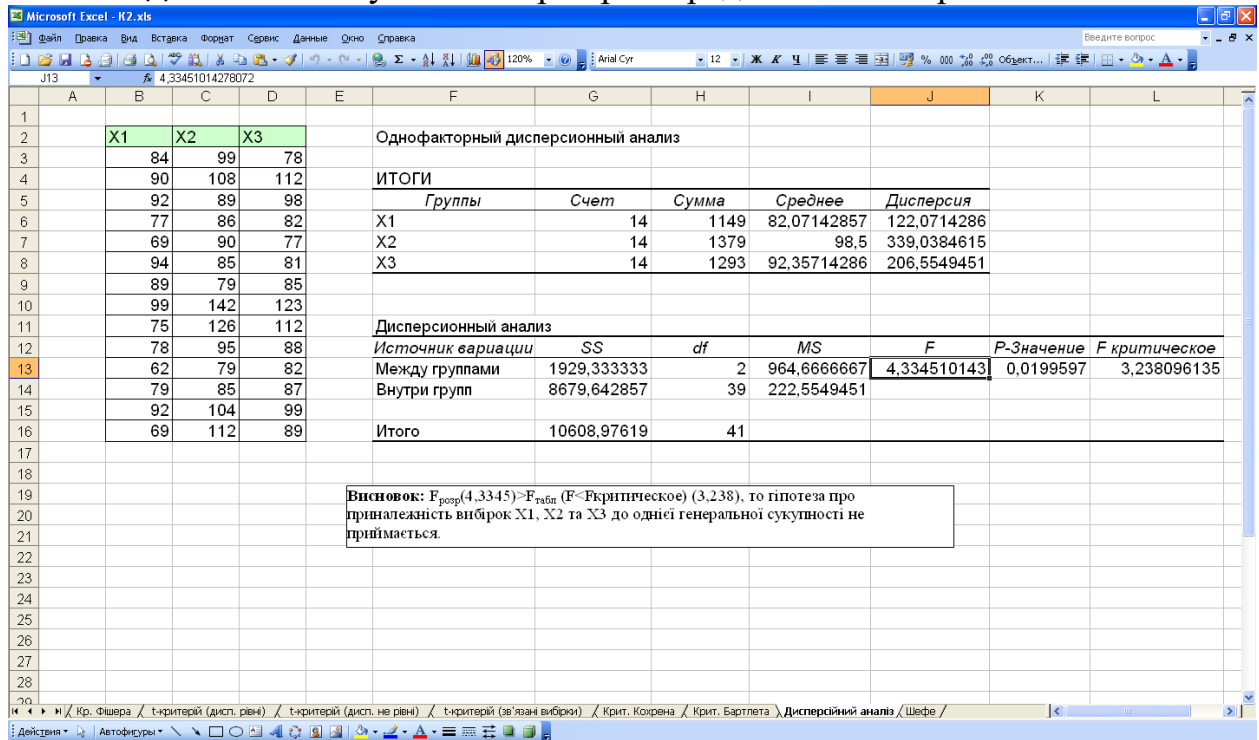


Рис. 2.18. Вікно з результатами перевірки гіпотези про рівність середніх значень для зв'язаних вибірок.

10. Перевірка гіпотези про рівність середніх значень вибірок шляхом дисперсійного аналізу.

Перевірка даної гіпотези виконується в середовищі в MS Excel. Для цього вибирається підпункт «Однофакторный дисперсионный анализ» у вікні «Анализ данных». Результати перевірки представлено на рис. 2.19.



Группы	Счет	Сумма	Среднее	Дисперсия
X1	14	1149	82,07142857	122,0714286
X2	14	1379	98,5	339,0384615
X3	14	1293	92,35714286	206,5549451

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	1929,333333	2	964,6666667	4,334510143	0,0199597	3,238096135
Внутри групп	8679,642857	39	222,5549451			
Итого	10608,97619	41				

**Выводок:**  $F_{\text{розра}}(4,3345) > F_{\text{табл}}(F < F_{\text{критическое}})(3,238)$ , то гіпотеза про принадлежність вибірок X1, X2 та X3 до однієї генеральної сукупності не приймається.

Рис. 2.19. Вікно з результатами перевірки гіпотези про рівність середніх значень вибірок шляхом дисперсійного аналізу.

11. Перевірка гіпотези про рівність середніх значень вибірок методом множинних порівнянь Шефе.

Відповідно до представлених формул в пункті 12 практичної роботи № 2 виконуються розрахунки для трьох вибірок, визначається критеріальне значення зі статистичних таблиць, що наведено в Додатку. Результати перевірки гіпотез представлено на рис. 2.20.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2		X1	X2	X3												
3		84	99	78	3,7194	0,25	206,13									
4		90	108	112	62,862	90,25	385,84									
5		92	89	98	98,577	90,25	31,842									
6		77	86	82	25,719	156,25	107,27									
7		69	90	77	170,86	72,25	235,84									
8		94	85	81	142,29	182,25	128,98									
9		89	79	85	48,005	380,25	54,128									
10		99	142	123	286,58	1892,25	938,98									
11		75	126	112	50,005	756,25	385,84									
12		78	95	88	16,577	12,25	18,985									
13		62	79	82	402,86	380,25	107,27									
14		79	85	87	9,4337	182,25	28,699									
15		92	104	99	98,577	30,25	44,128									
16		69	112	89	170,86	182,25	11,27									
17	Середнє арифметичне	82,071	98,5	92,357	222,55											
18	Сі	1	-0,5	-0,5	82,071	-49,25	-46,18	178,413265								
19					0,0714	0,01786	0,0179	0,10714286								
20																
21				Критеріальне значення S		3,74108										
22																
23				Висновок: оскільки $S(3,74108) > F_{таб}(3,23)$ , взятого для рівня значимості $\alpha=0,05$ , то гіпотеза про рівність середніх вибірок X1, X2 та X3 не приймається.												
24																
25																
26																
27																
28																
29																

Рис. 2.20. Вікно з результатами перевірки гіпотези про рівність середніх значень вибірок методом множинних порівнянь Шефе.

### Завдання для виконання комп'ютерного практикуму

Для випадкових величин, що надані в індивідуальному завданні, кожна з яких складається з трьох вибірок, виконати наведені нижче дії.

1. Виконати перевірку гіпотези відповідності вибірок нормальному закону розподілу.
2. Визначити середнє арифметичне елементів кожної вибірки.
3. Визначити дисперсію елементів кожної вибірки.
4. Виконати перевірку гіпотез про рівність двох дисперсій елементів вибірок за критерієм Фішера.
5. Виконати перевірку гіпотез про однорідність дисперсій елементів вибірок за критерієм Кохрена.
6. Виконати перевірку гіпотез про однорідність декількох дисперсій елементів вибірок за критерієм Бартлета.
7. Виконати перевірку гіпотез про рівність середніх значень вибірок при однакових дисперсіях.
8. Виконати перевірку гіпотез про рівність середніх вибірок без припущення про дисперсії.
9. Виконати перевірку гіпотез про рівність середніх вибірок при різних дисперсіях вибірок.
10. Виконати перевірку гіпотез про рівність середніх вибірок заданому значенню А.

11. Виконати перевірку гіпотези про рівність середніх при пов'язаних вибірках.
12. Виконати перевірку гіпотез про рівність середніх вибірок шляхом дисперсійного аналізу.
13. Виконати перевірку гіпотез про рівність середніх вибірок методом множинних порівнянь Шеффе.  
Порівняти результати обчислень для кожної вибірки випадкових величин.  
Зробити висновки за результатами перевірки гіпотез про розташування та розсіювання.

### Контрольні запитання

1. Що означає «статистична гіпотеза»?
2. Які етапи включає перевірка гіпотез?
3. Як здійснюється вибір критеріїв для перевірки гіпотез про положення і розсіювання?
4. Які є способи перевірки гіпотези відповідності вибірки нормальному закону розподілу?
5. За допомогою якого критерію здійснюється перевірка гіпотези про рівність двох дисперсій елементів вибірок?
6. Як визначається розрахункове значення критерію Фішера?
7. Коли використовується критерій Кохрена, як розраховується критеріальне значення цього критерію?
8. Коли використовується критерій Бартлета, як розраховується критеріальне значення цього критерію?
9. Що таке рівень значущості? Яких значень він може набувати?
10. Як виконується перевірка гіпотези про рівність середніх двох вибірок при рівних їх дисперсіях?
11. Як виконується перевірка гіпотези про рівність середніх двох вибірок при різних їх дисперсіях?
12. Як виконується перевірка гіпотези про рівність середніх двох вибірок без передбачень про їх дисперсії?
13. Як виконується перевірка гіпотези про рівність середнього значення вибірки деякій величині  $A$ ?
14. Як виконується перевірка гіпотези про рівність середніх кількох вибірок шляхом дисперсійного аналізу?
15. Як виконується перевірка гіпотези про рівність середніх двох пов'язаних вибірок?
16. Як виконується перевірка гіпотези про рівність середніх груп вибірок за методом множинних порівнянь Шеффе?
17. Яке призначення  $LSD$ -критерію?
18. Які умови використання непараметричних критеріїв перевірки гіпотез про розташування та розсіювання?
19. Які критерії використовуються для перевірки гіпотез про розташування та розсіювання в непараметричній статистиці?

20. Які допущення використовуються при перевірці гіпотез в непараметричній статистиці?
21. В якій послідовності виконується перевірка гіпотез про розташування та розсіювання в непараметричній статистиці?
22. Як виконати перевірку гіпотез про рівність двох дисперсій елементів вибірок за критерієм Фішера в MS Excel?
23. Як виконати перевірку гіпотез про рівність середніх значень вибірок при однакових дисперсіях в MS Excel?
24. Як виконати перевірку гіпотез про рівність середніх значень вибірок при різних дисперсіях в MS Excel?
25. Як виконати перевірку гіпотез про рівність середніх вибірок без припущення про дисперсії в SPSS Statistics?
26. Як виконати перевірку гіпотез про рівність середніх вибірок заданому значенню A в SPSS Statistics?
27. Як виконати перевірку гіпотез про рівність середніх при зв'язаних вибірках в MS Excel?
28. Як виконати перевірку гіпотез про рівність середніх вибірок шляхом дисперсійного аналізу в MS Excel?

## Комп'ютерний практикум №3

### ПЕРЕВІРКА ГІПОТЕЗ ПРО НАЯВНІСТЬ ЗВ'ЯЗКУ МІЖ ЗМІННИМИ

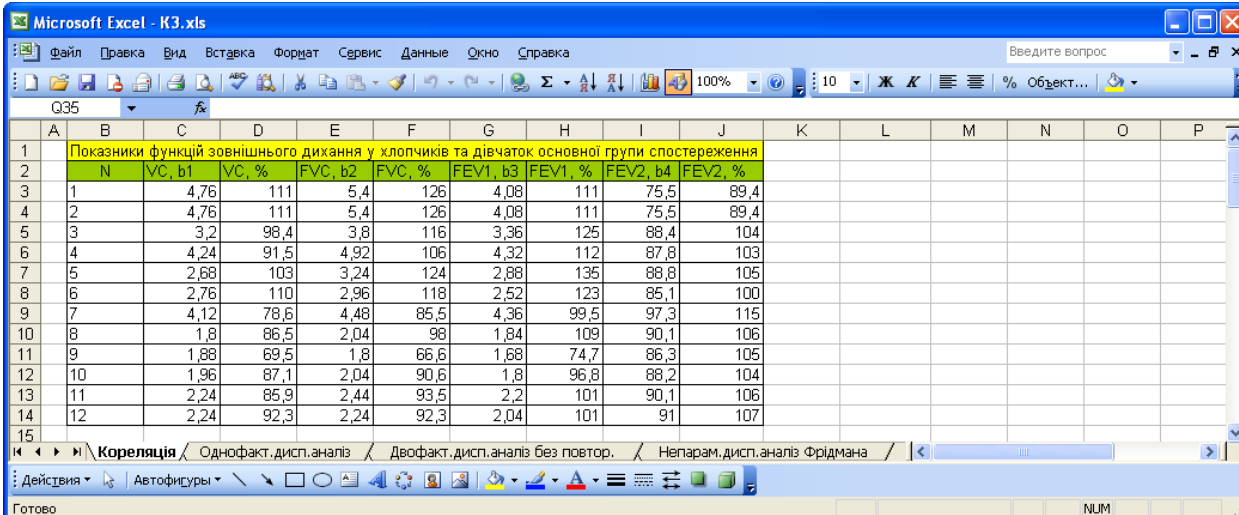
**Мета:** отримання практичних навиків перевірки гіпотез про наявність зв'язку між змінними за допомогою MS Excel.

#### Теоретичні відомості

Розрахунки комп'ютерного практикуму виконуються з використанням сервісу «Анализ данных» MS Excel.

На рис. 3.1 представлено числові значення показників зовнішнього дихання у хлопчиків та дівчаток основної групи спостереження.

Для виконання кореляційного аналізу з визначенням коефіцієнтів кореляції Пірсона необхідно вибрати пункт головного меню Сервис→Анализ данных...→Корреляция→ОК.



The screenshot shows a Microsoft Excel window titled "Microsoft Excel - КЗ.xls". The spreadsheet contains a table with 10 columns of respiratory function indicators and 14 rows of data. The first row is a header, and the second row is a sub-header. The data rows contain numerical values for each indicator.

	N	VС, b1	VС, %	FVC, b2	FVC, %	FEV1, b3	FEV1, %	FEV2, b4	FEV2, %
1	1	4,76	111	5,4	126	4,08	111	75,5	89,4
2	2	4,76	111	5,4	126	4,08	111	75,5	89,4
3	3	3,2	98,4	3,8	116	3,36	125	88,4	104
4	4	4,24	91,5	4,92	106	4,32	112	87,8	103
5	5	2,68	103	3,24	124	2,88	135	88,8	105
6	6	2,76	110	2,96	118	2,52	123	85,1	100
7	7	4,12	78,6	4,48	85,5	4,36	99,5	97,3	115
8	8	1,8	86,5	2,04	98	1,84	109	90,1	106
9	9	1,88	69,5	1,8	66,6	1,68	74,7	86,3	105
10	10	1,96	87,1	2,04	90,6	1,8	96,8	88,2	104
11	11	2,24	85,9	2,44	93,5	2,2	101	90,1	106
12	12	2,24	92,3	2,24	92,3	2,04	101	91	107

Рис.3.1. Вікно з початковими даними для кореляційного аналізу.

У вікні «Корреляция» (рис.3.2) вказуються вхідний та вихідний інтервали, групування, мітки в першому рядку і натисканням ОК генерується створення таблиці значень коефіцієнтів кореляції, як це показано на рис. 3.3.

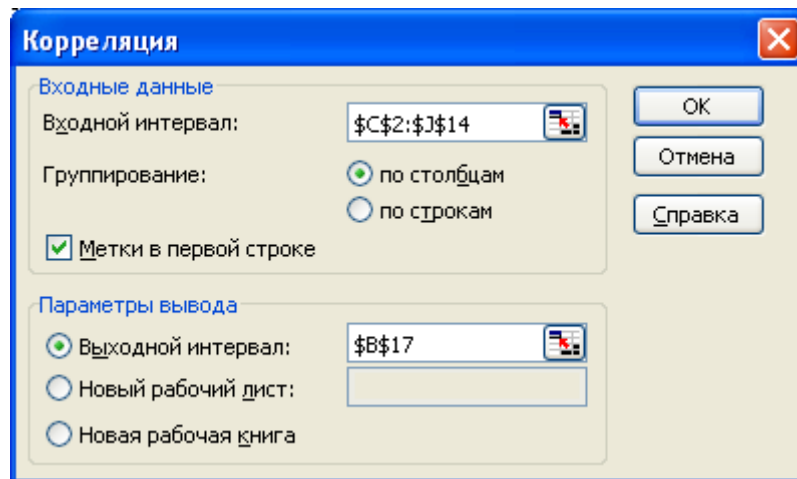


Рис.3.2. Вікно «Корреляция».

Для перевірки значимості даних, згідно пункту 3.1.2. практичної роботи №3, визначається табличне значення критерію Стюдента та за формулою (3.2) обчислюється розрахункове значення критерію Стюдента (рис.3.3).

Значення коефіцієнтів кореляції										
	N	VC, b1	VC, %	FVC, b2	FVC, %	FEV1, b3	FEV1, %	FEV2, b4	FEV2, %	
N	1									
VC, b1	-0,80395	1								
VC, %	-0,6644	0,51228	1							
FVC, b2	-0,84649	0,992907	0,548662	1						
FVC, %	-0,76766	0,56239	0,956515	0,624457	1					
FEV1, b3	-0,77113	0,962819	0,403171	0,969541	0,509224	1				
FEV1, %	-0,55339	0,284972	0,759105	0,371183	0,863561	0,355504	1			
FEV2, b4	0,632753	-0,48761	-0,67187	-0,48687	-0,59212	-0,26493	-0,13822	1		
FEV2, %	0,651733	-0,5176	-0,74495	-0,52307	-0,66888	-0,30164	-0,22785	0,9904669	1	
Розрахункове значення t-критерію Стюдента						tтабл = 2,23 (для α=0,05%)				
	N	VC, b1	VC, %	FVC, b2	FVC, %	FEV1, b3	FEV1, %	FEV2, b4	FEV2, %	
N										
VC, b1	-4,27504									
VC, %	-2,6669	1,789484								
FVC, b2	-4,49703	23,62003	1,856181							
FVC, %	-3,16923	1,923732	8,676212	2,11528						
FEV1, b3	-2,96676	10,08062	1,165623	9,696165	1,449324					
FEV1, %	-1,48563	0,840888	3,085257	0,97916	3,829452	0,85049				
FEV2, b4	1,634276	-1,5797	-2,40001	-1,36533	-1,643	-0,54949	-0,27913			
FEV2, %	1,488346	-1,71103	-2,95443	-1,50331	-2,01199	-0,63275	-0,40531	12,453925		

Рис.3.3. Вікно зі значеннями коефіцієнтів кореляції Пірсона та відповідних їм розрахункових значень критерію Стюдента.

Шляхом порівняння табличного та розрахункових значень критерію Стюдента визначається їх значимість, а напівширину довірчого інтервалу для значимих коефіцієнтів – за формулою

$$\Delta = \frac{t_{n-2,\alpha} (1 - r_{ij}^2)}{\sqrt{N}}$$

де  $N$  – кількість спостережень, за яким розраховується коефіцієнт кореляції;  $r_{ij}$  – значення коефіцієнта кореляції;  $t_{n-2,\alpha}$  – табличне значення критерію Стюдента, що взятий з  $(N - 2)$  степенями свободи.

На рис. 3.4. наведено результати кореляційного аналізу даних, що представлено в вікні на рис. 3.1. В цій таблиці позначено З – значимий, Н – незначимий коефіцієнт кореляції.

	N	VC, b1	VC, %	FVC, b2	FVC, %	FEV1, b3	FEV1, %	FEV2, b4	FEV2, %
40	Значимість коефіцієнтів кореляції								
41	N								
42	N								
43	VC, b1	З							
44	VC, %	З	Н						
45	FVC, b2	З	З	Н					
46	FVC, %	З	Н	З	Н				
47	FEV1, b3	З	З	Н	З	Н			
48	FEV1, %	Н	Н	З	Н	З	Н		
49	FEV2, b4	Н	Н	З	Н	Н	Н	Н	
50	FEV2, %	Н	Н	З	Н	Н	Н	З	
51									
52	де З - значимий коефіцієнт, Н - не значимий коефіцієнт при заданому рівні значимості.								
53									
54	Напівширина довірчого інтервалу								
55	N								
56	N								
57	VC, b1	1,81392							
58	VC, %	1,945837							
59	FVC, b2	1,768727	1,595355						
60	FVC, %	1,850642		1,641024					
61	FEV1, b3	1,847203	1,633235		1,624873				
62	FEV1, %			1,859048		1,749935			
63	FEV2, b4							1,939405	
64	FEV2, %								1,5984697
65									

Рис. 3.4. Вікно з результатами розрахунків коефіцієнтів кореляції.

На рис. 3.5 представлено початкові дані для однофакторного параметричного дисперсійного аналізу даних за F-критерієм Фішера. В наведених даних вважається, що фактор А змінювався три рази ( $A_1, A_2, A_3$ ) і виконувались вимірювання ознаки, на яку впливав даний фактор. Для кожного значення фактору вимірювання ознаки виконували 12 разів.

Для виконання однофакторного параметричного дисперсійного аналізу даних за F-критерієм Фішера необхідно вибрати пункт головного меню Сервіс→Анализ даних...→Однофакторный дисперсионный анализ→ОК.

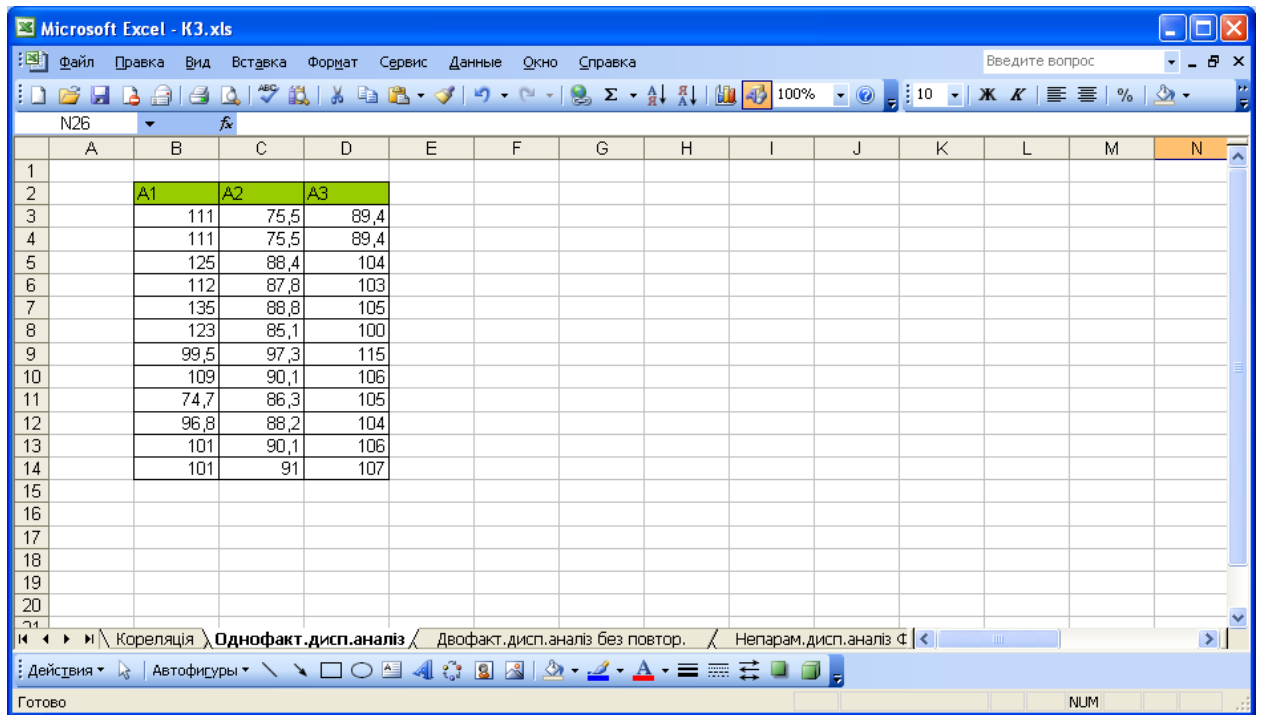


Рис.3.5. Вікно з початковими даними для однофакторного параметричного дисперсійного аналізу даних за F-критерієм Фішера.

У вікні «Однофакторный дисперсионный анализ» (рис.3.6) задається вхідний та вихідний інтервали, вид групування, мітки в першому рядку і натисканням кнопки ОК генерується виведення результату, який наведено на рис. 3.7.

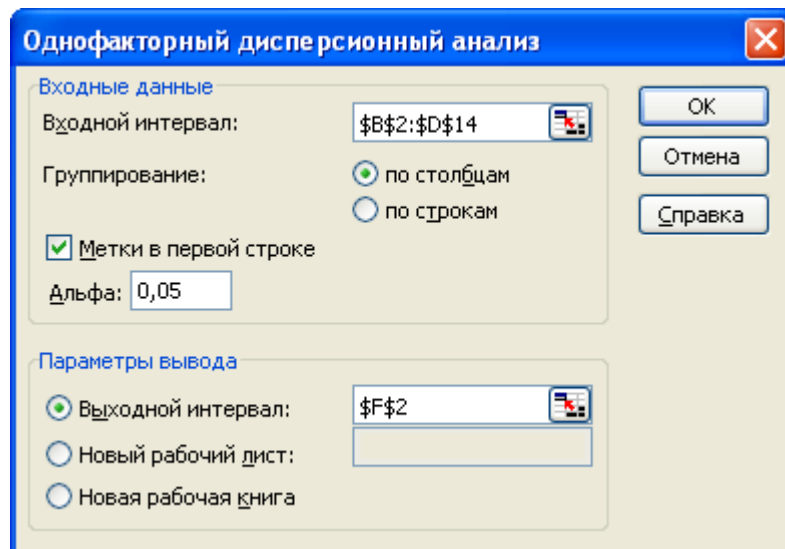


Рис.3.6. Вікно «Однофакторный дисперсионный анализ».

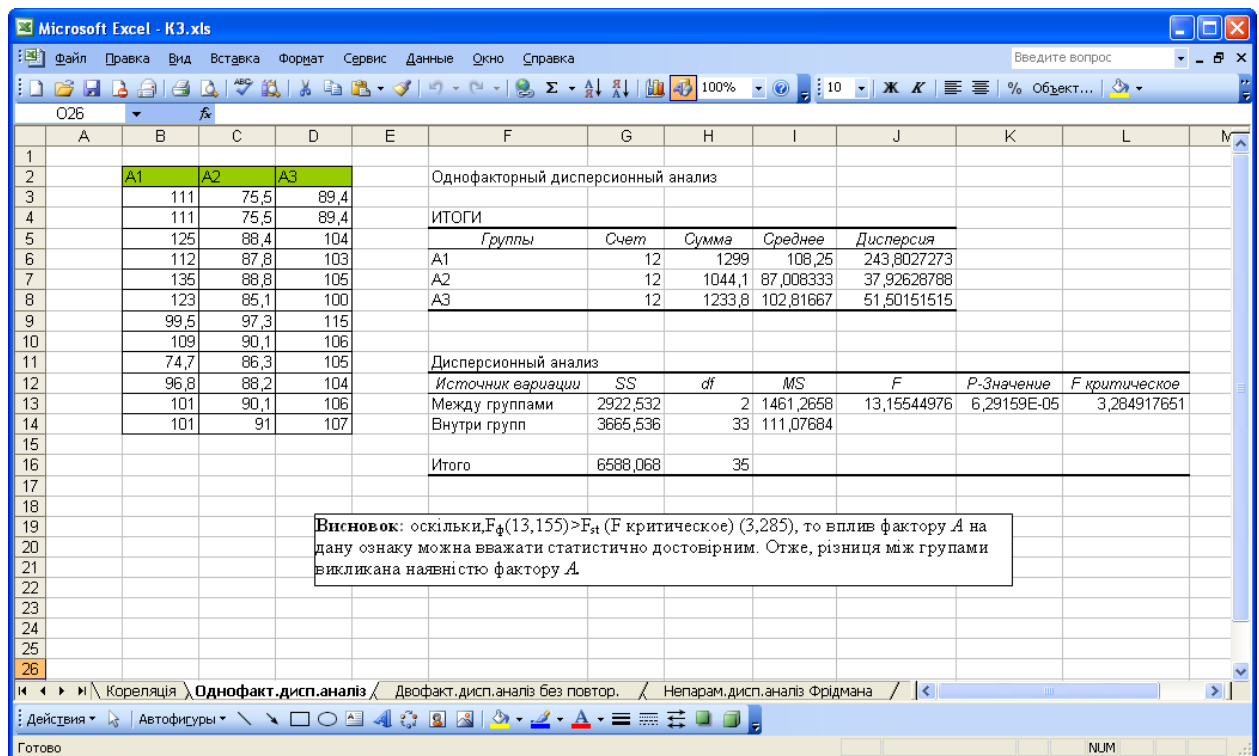


Рис.3.7. Вікно з результатами однофакторного дисперсійного аналізу.

Надалі можна зробити висновок про статистичну достовірність впливу фактору на ознаку. Оскільки, розрахункове значення критерію Фішера  $F_{\phi}$ , що дорівнює 13,155, більше табличного (критичного)  $F_{st} = 3,285$ , то вплив фактору A на дану ознаку можна вважати статистично достовірним. Отже, відмінність між групами викликана наявністю фактору A.

На рис. 3.8 представлено початкові дані для двофакторного параметричного дисперсійного аналізу даних за F-критерієм Фішера. Тут вважається, що фактор A змінювався чотири рази (A1, A2, A3, A4), фактор B змінювався три рази (B1, B2, B3). Виконували вимірювання ознаки, на яку впливали дані фактори. Для кожної пари факторів вимірювання ознаки виконували один раз (що є передумовою двофакторного параметричного дисперсійного аналізу без повторень).

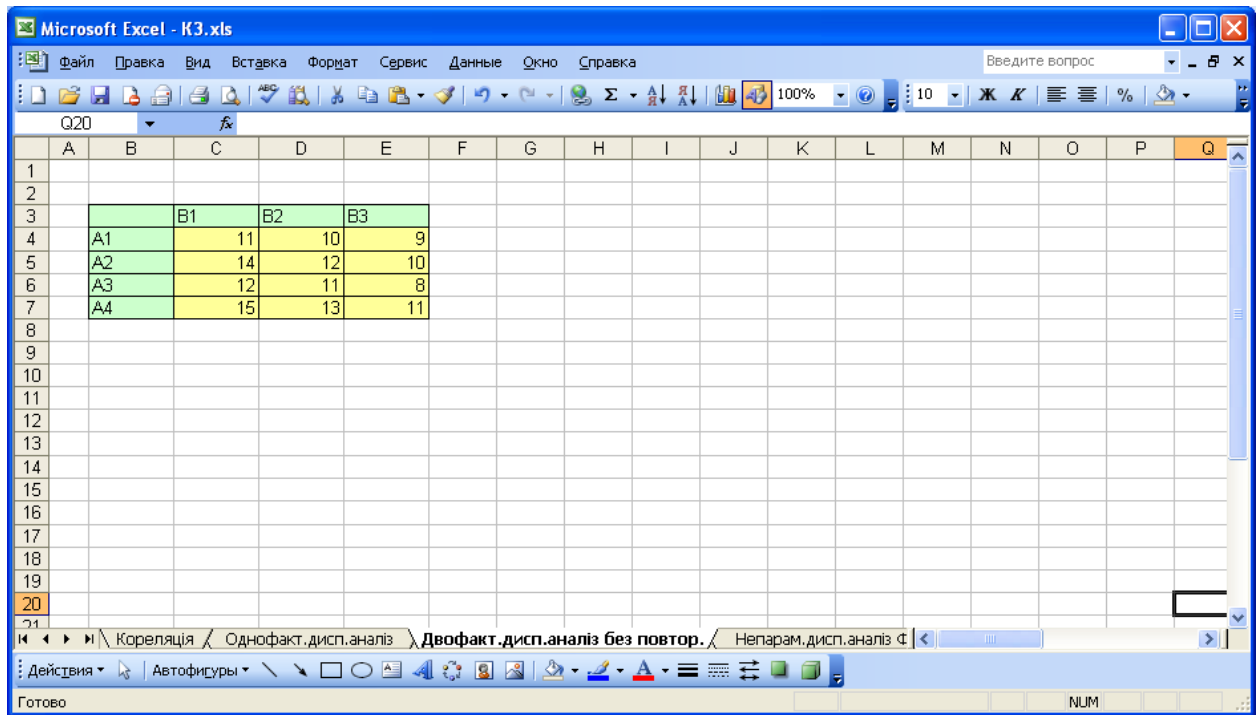


Рис.3.8. Вікно з початковими даними для двофакторного параметричного дисперсійного аналізу даних за F-критерієм Фішера.

Для виконання двофакторного параметричного дисперсійного аналізу даних за F-критерієм Фішера необхідно вибрати пункт головного меню Сервіс→Анализ данных...→Двуфакторный дисперсионный анализ→ОК. У вікні «Двуфакторный дисперсионный анализ» (рис.3.9) вказується вхідний та вихідний інтервали, мітки і натисканням клавиші ОК генерується виведення результату, який наведено на рис. 3.10.

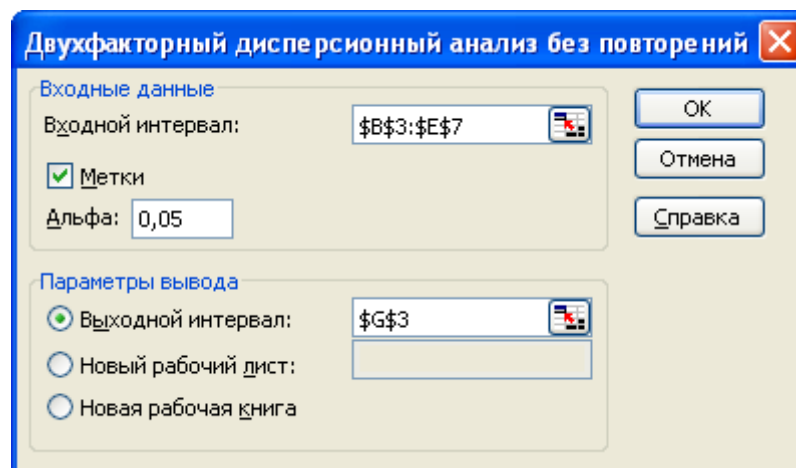


Рис.3.9. Вікно «Двуфакторный дисперсионный анализ».

Відповідно до теоретичних положень методу дисперсійного аналізу формують висновок про статистичну достовірність впливу факторів  $A$  і  $B$  на ознаку  $X$ . Оскільки, для рядків розрахункове значення критерію Фішера  $F$ , що

дорівнює 18 більше табличного (критичного)  $F_{st}$ , що дорівнює 4,757, то вплив фактору  $A$  на дану ознаку можна вважати статистично достовірним. Для стовпців розрахункове значення  $F_f=37$  більше критичного  $F_{st}=5,143$ , то вплив фактору  $B$  на дану ознаку також можна вважати статистично достовірним.

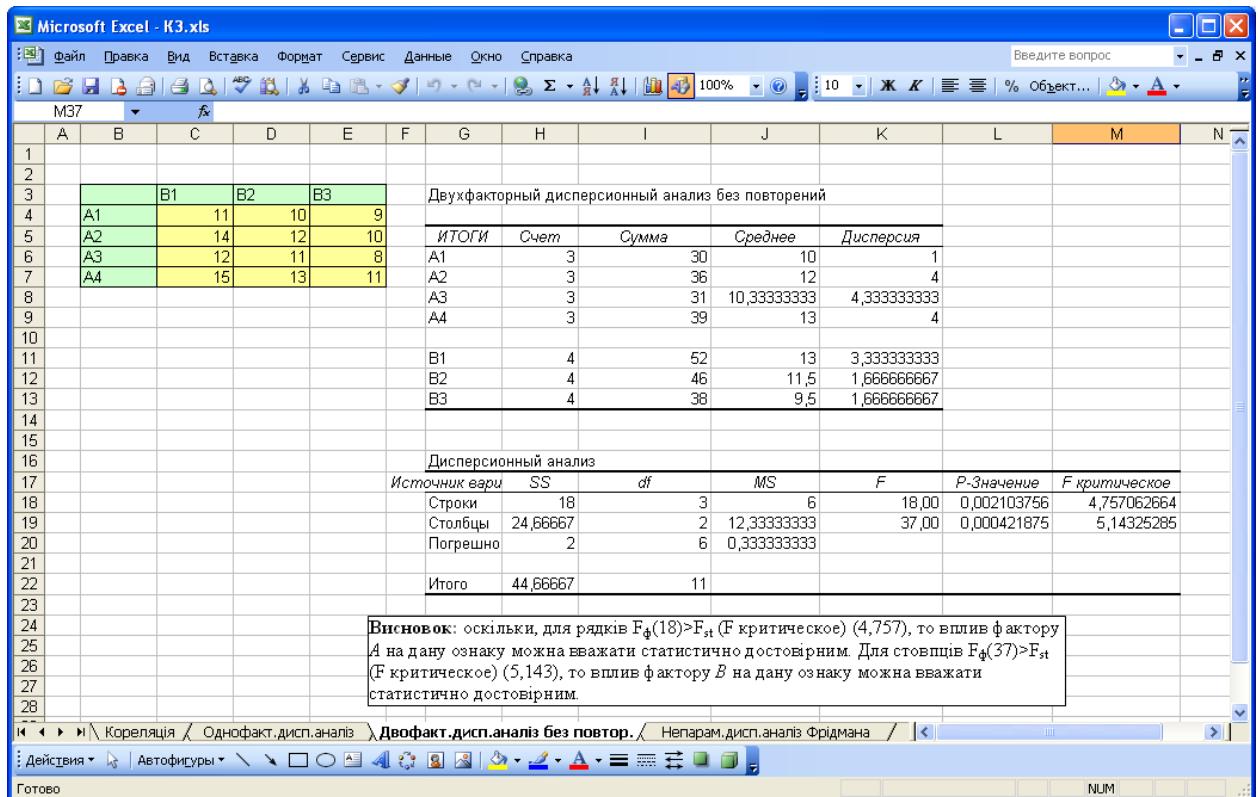


Рис.3.10. Вікно з результатами двофакторного дисперсійного аналізу.

Для непараметричного дисперсійного аналізу Фрідмана початкові дані представляються аналогічно однофакторному дисперсійному аналізу. При цьому необхідно виконати сортування стовпчиків за зростанням за значеннями порядкових їх номерів, як це показано на рис. 3.11. Далі виконуються розрахунки критерію  $\chi^2$ .

При цьому значення  $X$  в кожному стовпчику таблиці замінюються їх рангами, тобто замість значень змінних ставиться їх номер в ряду, що впорядкований за зростанням.

Розрахункове значення критерію визначається за формулою:

$$\chi^2 = \frac{12 \sum_{j=1}^n \left( \sum_{i=1}^m R_{ij} \right)^2}{mn(n+1)} - 3m(n+1),$$

де  $R_{ij}$  – відповідне значення рангів.

Якщо розрахункове значення  $\chi^2$  більше критичного, що взяте з заданим рівнем значимості  $\alpha$  і  $(n-1)$  степенем свободи, то гіпотеза про відмінність між сукупностями випадкових величин приймається.

При розрахунках можна перевірити правильність розстановки рангів і розрахунків, використовуючи таке співвідношення:

$$\sum_{i=1}^m \sum_{j=1}^n R_{ij} = \frac{nm(m+1)}{2}.$$

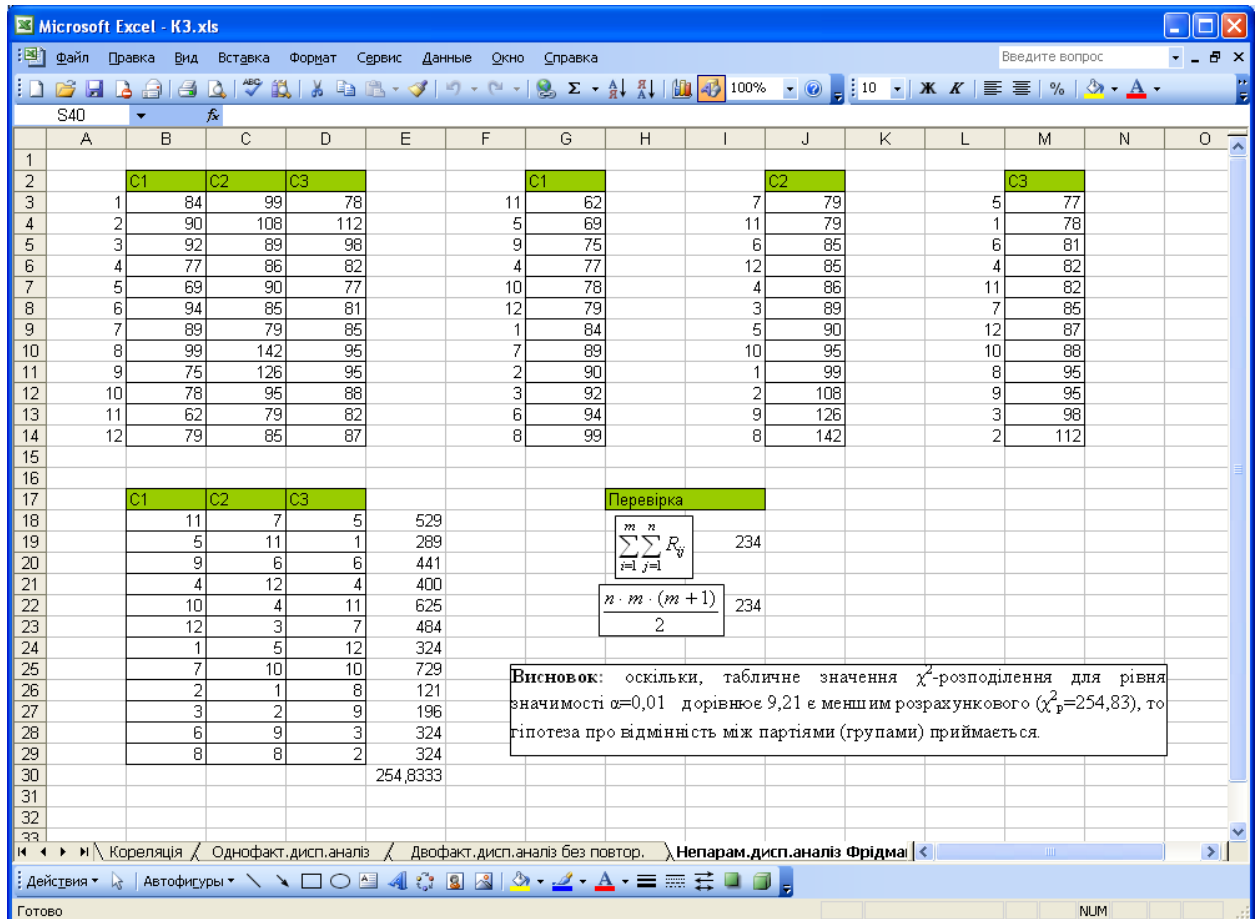


Рис.3.11. Результат непараметричного дисперсійного аналізу Фрідмана.

### Завдання для виконання комп'ютерного практикуму

- Для підмножини вибірок випадкових величин, що надані в індивідуальному завданні до практичної роботи №1, виконати наступні дії:
  - визначити коефіцієнти кореляції Пірсона між змінними вибірок;
  - визначити значимість отриманих коефіцієнтів кореляції;
  - визначити довірчі інтервали для коефіцієнтів кореляції;
  - визначити часткові коефіцієнти кореляції та їх значимість.
- Для підмножини вибірок випадкових величин, що надані в індивідуальному завданні до практичної роботи №2, виконати такі дії:

- розв’язати задачу однофакторного параметричного дисперсійного аналізу даних за  $F$ -критерієм Фішера;
- розв’язати задачу двофакторного параметричного дисперсійного аналізу даних за  $F$ -критерієм Фішера;
- розв’язати задачу непараметричного дисперсійного аналізу даних (Фрідмана) за критерієм  $\chi^2$ -квадрат.

Порівняти результати обчислень для різних вибірок випадкових величин.

Зробити висновки за результатами обчислень параметрів, що характеризують наявність зв’язку між змінними.

### Контрольні запитання

1. Дайте визначення кореляції.
2. Що показує коефіцієнт кореляції Пірсона?
3. Як визначаються коефіцієнти кореляції Пірсона? Як визначається їх значимість?
4. Визначення довірчих інтервалів для коефіцієнтів кореляції.
5. Частковий коефіцієнт кореляції. Його призначення, розрахунок та визначення значимості.
6. Сформулюйте задачу однофакторного параметричного дисперсійного аналізу за  $F$ -критерієм Фішера.
7. Як визначається наявність впливу фактору  $A$  на ознаку  $X$  статистично достовірним?
8. Задача двофакторного параметричного дисперсійного аналізу за  $F$ -критерієм Фішера.
9. Як визначити значимість двох факторів на змінну при розв’язанні двофакторного параметричного дисперсійного аналізу?
10. В чому полягає розв’язання задачі непараметричного дисперсійного аналізу даних (Фрідмана) за критерієм  $\chi^2$ ?
11. Як виконати розрахунок коефіцієнтів кореляції в MS Excel?
12. Як визначити розрахункові значення  $t$ -критерію Стьюдента в MS Excel?
13. Як визначити значимість коефіцієнтів кореляції в MS Excel?
14. Однофакторний параметричний дисперсійний аналіз за  $F$ -критерієм Фішера в MS Excel?
15. Двофакторний параметричний дисперсійний аналіз за  $F$ -критерієм Фішера.

## Комп'ютерний практикум №4

### ФАКТОРНИЙ АНАЛІЗ ДАНИХ

**Мета:** отримання практичних навиків стиснення інформаційних масивів без зниження інформативності шляхом визначення латентних факторів за допомогою програми SPSS Statistics.

#### Теоретичні відомості

З кінця 1950-х років факторний аналіз набув значної популярності в різноманітних дослідженнях. Багато в чому цьому сприяла розробка Раймондом Кет-Теллем знаменитого 16-факторного особистісного опитувальника (16PF). Саме за допомогою факторного аналізу йому вдалося звести близько 4500 найменувань особистісних особливостей до 187 питань, які, в свою чергу, дозволяють виміряти 16 різних властивостей особистості. Факторний аналіз дає можливість кількісно визначити щось, що безпосередньо не вимірюється, виходячи з декількох доступних виміру змінних. Наприклад, характеристики «відвідує розважальні заходи», «багато розмовляє», «охоче йде на контакт з будь-якою незнайомою людиною» можуть служити оцінками якості «товариськість», яке безпосередньо не піддається кількісному виміру. Факторний аналіз дозволяє встановити для великого числа вихідних ознак порівняно невеликий набір «властивостей», що характеризують зв'язок між групами цих ознак і які називаються факторами.

Задача факторного аналізу полягає в тому, щоб представити ознаки, що надано у вигляді матриці  $x_{ij}$  розмірності  $(m \times n)$ , де  $i$  – об'єкт дослідження (наприклад, пацієнт),  $j$  – характеристика (властивість) відповідного об'єкту (результати аналізу), у вигляді схованих гіпотетичних факторів. З метою врахування розмірності властивостей об'єктів дослідження доцільно їх нормувати, тобто представити як  $y_{ij}$ , де  $y_{ij}$  – нормоване значення  $j$ -ї властивості  $i$ -го об'єкту. Найпростішою моделлю для опису кожного з параметрів може служити лінійна модель:

$$y_{ij} = a_{i1}f_1 + a_{i2}f_2 + \dots + a_{ij}f_j + \dots + a_{in}f_n, \quad i=1, 2, 3, \dots, m, \quad 1) \quad (4.)$$

де кожний з параметрів лінійно залежить від  $n$  некорельованих між собою компонентів (латентних факторів)  $f_1, f_2, \dots, f_j, \dots, f_n$ ,  $a_{ij}$  – вклад  $j$ -ої латентної змінної у відповідну компоненту (властивість об'єкту). Важливою властивістю методу є те, що всі компоненти розташовуються в порядку зменшення їх внеску в сумарну дисперсію розходжень параметрів.

При визначенні кількості факторів (латентних змінних) застосовано правило, яке дозволяє залишати фактори з власними числами кореляційної матриці факторів, що мають значення більше одиниці. Таким чином, із (4.1)

отримано залежність початкових властивостей об'єктів від  $k$  некорельованих між собою компонентів (латентних факторів):

$$y_{ij} = a_{i1}f_1 + a_{i2}f_2 + \dots + a_{il}f_l + \dots + a_{ik}f_k; \quad i=1, 2, 3, \dots, m; \quad (4.2)$$

$$j=1, 2, 3, \dots, n; \quad l=1, 2, 3, \dots, k; \quad k < n,$$

де  $a_{il}$  – вклад  $l$ -ої латентної змінної в  $j$ -ту компоненту (властивість об'єкту).

Це простий спосіб визначення кількості латентних змінних добре зарекомендував себе, тому що зазвичай дає результати, які співпадають з тими, що очікує дослідник. Крім того, цей метод був ретельно перевірений на модельних штучних даних.

Отже, у відповідності з методикою факторного аналізу, спочатку необхідно створити масив початкової інформації. Отриманий масив інформації методами факторного аналізу стискається з визначенням латентних факторів, які відображають початковий інформаційний масив, але у зменшеному вигляді.

Для виконання факторного аналізу доцільно використовувати систему статистичного аналізу SPSS Statistics.

Відповідно індивідуального завдання спочатку необхідно створити за допомогою редактора системи масив початкових даних. Вікно редактора даних з масивом початкової інформації представлено на рис. 4.1. При цьому всі значення повинні мати числовий тип.

Year	S	n	VC_b1	VC	FVC_b2	FVC	FEV1_b3	FEV1	FEV2_b4	FEV2	PEF_b5	PEF	MEF50...	MEF50	MEF25...	MEF25	MEF225	MVV_b8	MVV	Conclusion	nep
1	14	1	9	4,76	111,00	5,40	126,00	4,08	111,00	75,50	89,40	7,37	101,00	5,28	119,00	2,73	119,00	43,80	112,70	2	
2	14	2	10	4,76	111,00	5,40	126,00	4,08	111,00	75,50	89,40	7,37	101,00	5,28	119,00	2,73	119,00	51,30	100,50	1	
3	13	2	11	3,20	98,40	3,80	116,00	3,36	125,00	88,40	104,00	5,84	101,00	5,75	161,00	4,05	220,00	55,50	101,00	1	
4	13	1	6	4,24	91,50	4,92	106,00	4,32	112,00	87,80	103,00	8,13	105,00	6,87	146,00	4,74	195,00	54,00	67,20	3	
5	11	1	10	2,88	103,00	3,24	124,00	2,88	135,00	88,80	105,00	4,80	100,00	3,24	114,00	2,01	130,00	57,40	134,00	1	
6	10	2	5	2,76	110,00	2,96	118,00	2,52	123,00	85,10	100,00	3,92	85,10	3,52	121,00	2,24	150,00	58,40	101,00	1	
7	14	1	12	4,12	78,60	4,48	85,50	4,36	99,50	97,30	115,00	7,84	89,00	5,10	98,50	3,84	146,00	83,70	91,10	2	
8	8	1	5	1,80	86,50	2,04	98,00	1,84	109,00	90,10	106,00	3,76	95,30	2,31	91,90	1,09	87,70	46,20	138,00	1	
9	11	1	3	1,88	89,50	1,80	66,60	1,68	74,70	86,30	105,00	2,70	51,50	2,07	63,20	1,11	66,00	54,00	79,10	3	
10	9	1	9	1,96	87,10	2,04	90,60	1,80	96,80	88,20	104,00	3,37	75,70	2,51	89,20	1,20	83,20	33,10	81,90	2	
11	13	1	3	2,24	85,90	2,44	93,50	2,20	101,00	90,10	106,00	3,01	59,30	2,42	76,20	1,87	114,00	37,80	85,30	2	
12	11	2	4	2,24	92,30	2,24	92,30	2,04	101,00	91,00	107,00	4,07	85,50	2,47	82,50	1,43	93,10	39,60	92,70	1	
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					

Рис.4.1. Вікно редактора даних системи з масивом початкової інформації для її факторного аналізу.

Для цього вибирається команда головного меню **Анализ**→**Снижение размерности**→**Факторный анализ**. У вікні «Факторный анализ» (рис.4.2)

вибираються дані із створеного масиву початкової інформації, що підлягають факторному аналізу (рис.4.3).

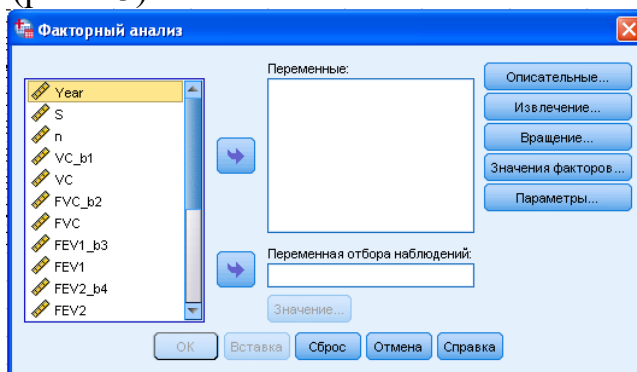


Рис.4.2. Вікно «Факторный анализ».

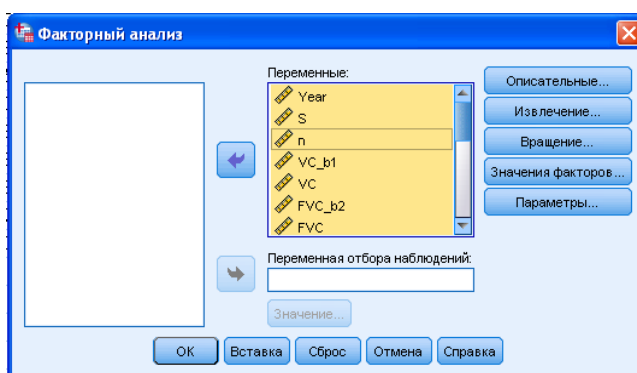


Рис.4.3. Вікно «Факторный анализ» з вибраними даними.

При переході на вкладку «Описательные...» можна вибрати для виведення значення описової статистики початкових даних та їх кореляційну матрицю (рис. 4.4).

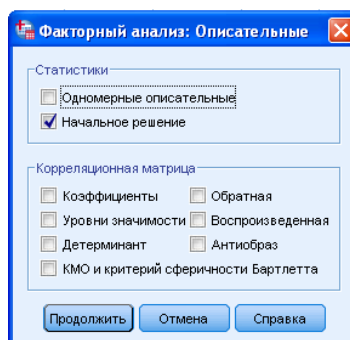


Рис.4.4. Вікно «Факторный анализ: Описательные».

У вікні «Факторный анализ: Выделение факторов» (рис.4.5) можна вибрати метод визначення кількості використовуваних латентних факторів за величиною сумарної дисперсії або завдання граничного власного значення кореляційної матриці факторів.

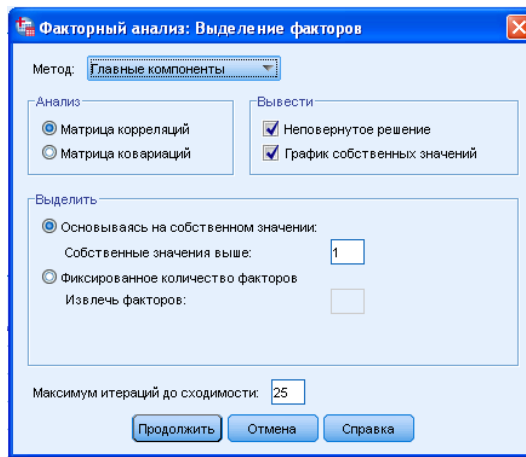


Рис.4.5. Вікно «Факторный анализ: Выделение факторов».

У вікні «Факторный анализ: Значения факторов» (рис.4.6) задається можливість збереження значень факторів як змінні.

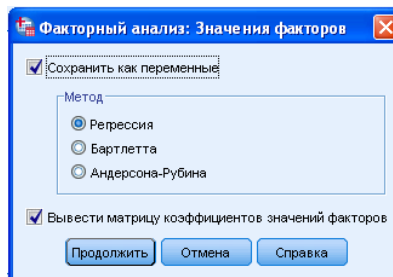


Рис.4.6. Вікно «Факторный анализ: Значения факторов».

Після вибору всіх необхідних параметрів у вікні «Факторный анализ» натисканням кнопки ОК генерується створення файлу результатів факторного аналізу у форматі `srv`. Отриманий файл можна експортувати в зручний для користувача формат. Для цього вибирається команда головного меню **Файл**→**Экспортировать**, а у вікні (рис.4.7) вибирається формат, задається розташування та ім'я файлу й натискається ОК.

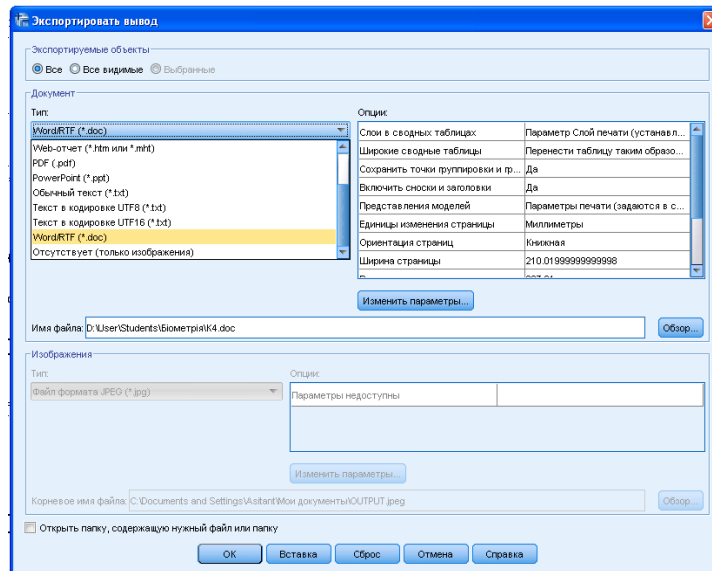


Рис.4.7. Вікно «Экспортировать вывод».

У вікні з введеними початковими даними виводяться розраховані значення латентних факторів (рис.4.8).

	VC_b1	VC	FVC_b2	FVC	FEV1	FEV2_b4	FEV2	PEF_b5	PEF	MEF50...	MEF50	MEF25_b7	MEF25	MVV_b8	MVV	FAC1_1	FAC2_1	FAC3_1
1	4,76	111,00	5,40	126,00	111,00	75,50	89,40	7,37	101,00	5,28	119,00	2,73	119,00	43,80	112,70	1,12841	-1,19202	-1,14745
2	4,76	111,00	5,40	126,00	111,00	75,50	89,40	7,37	101,00	5,28	119,00	2,73	119,00	51,30	100,50	1,14277	-,97526	-1,26623
3	3,20	98,40	3,80	116,00	116,00	125,00	88,40	104,00	5,84	101,00	5,75	161,00	4,05	220,00	55,50	-,99287	-,57319	-,98083
4	4,24	91,50	4,92	106,00	112,00	87,80	103,00	8,13	105,00	6,87	146,00	4,74	195,00	54,00	67,20	1,16291	1,29158	-,58053
5	2,88	103,00	3,24	124,00	135,00	88,80	105,00	4,80	100,00	3,24	114,00	2,01	130,00	57,40	134,00	-,29055	-,66430	1,74400
6	2,76	110,00	2,96	118,00	123,00	85,10	100,00	3,92	85,10	3,52	121,00	2,24	150,00	58,40	101,00	-,25054	-,54477	-,69330
7	4,12	78,60	4,48	85,50	99,50	97,30	115,00	7,84	89,00	5,10	98,50	3,94	146,00	83,70	91,10	-,19809	2,27185	1,2446
8	1,80	86,50	2,04	98,00	109,00	90,10	106,00	3,76	95,30	2,31	91,90	1,09	87,70	46,20	138,00	-,72115	-,63212	1,20500
9	1,88	89,50	1,80	86,60	74,70	86,30	105,00	2,70	51,50	2,07	63,20	1,11	66,00	54,00	79,10	-,16609	-,41717	-1,24885
10	1,96	87,10	2,04	90,60	96,80	88,20	104,00	3,37	75,70	2,51	89,20	1,20	83,20	33,10	81,90	-,99445	-,35613	-,39684
11	2,24	85,90	2,44	93,50	101,00	90,10	106,00	3,01	59,30	2,42	76,20	1,87	114,00	37,80	85,30	-,98099	-,01235	-,23873
12	2,24	92,30	2,24	92,30	101,00	91,00	107,00	4,07	85,50	2,47	82,50	1,43	93,10	39,60	92,70	-,80146	-,17684	-,13105
13																		
14																		
15																		
16																		
17																		
18																		
19																		
20																		
21																		
22																		
23																		
24																		
25																		
26																		
27																		
28																		

Рис.4.8. Вікно з початковою інформацією та визначеними латентними факторами.

В результаті факторного аналізу даних, що складається з 12 об'єктів, 15 початкових властивостей, якими вони характеризуються, замінюються трьома факторами, при цьому зберігається 88,403 % інформативності.

Власні значення кореляційної матриці факторів і відповідні їм дисперсії інформативності наведено в таблиці 4.1.

Таблиця 4.1. Власні значення та відповідні їм дисперсії, що визначають інформативність факторів.

Компонент	Объясненная совокупная дисперсия					
	Начальные собственные значения			Суммы квадратов нагрузок извлечения		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	8,124	54,163	54,163	8,124	54,163	54,163
2	3,159	21,057	75,220	3,159	21,057	75,220
3	1,978	13,183	88,403	1,978	13,183	88,403
4	,959	6,394	94,797			
5	,417	2,777	97,574			
6	,220	1,464	99,038			
7	,109	,725	99,763			
8	,027	,182	99,946			
9	,006	,037	99,983			
10	,002	,011	99,993			
11	,001	,007	100,000			
12	2,958E-16	1,972E-15	100,000			
13	5,259E-17	3,506E-16	100,000			
14	-1,126E-16	-7,504E-16	100,000			
15	-4,190E-16	-2,793E-15	100,000			

Метод выделения факторов: метод главных компонент.

В якості критерію, що визначає кількість латентних змінних, що з високою інформативністю описують досліджуваний процес, можна використовувати власні значення кореляційної матриці факторів. Так в [10] застосовано правило, яке дозволяє залишати фактори з власними числами кореляційної матриці факторів, що мають значення більше одиниці.

На рис. 4.9 наведено графік залежності власних значень кореляційної матриці від номера компоненти.

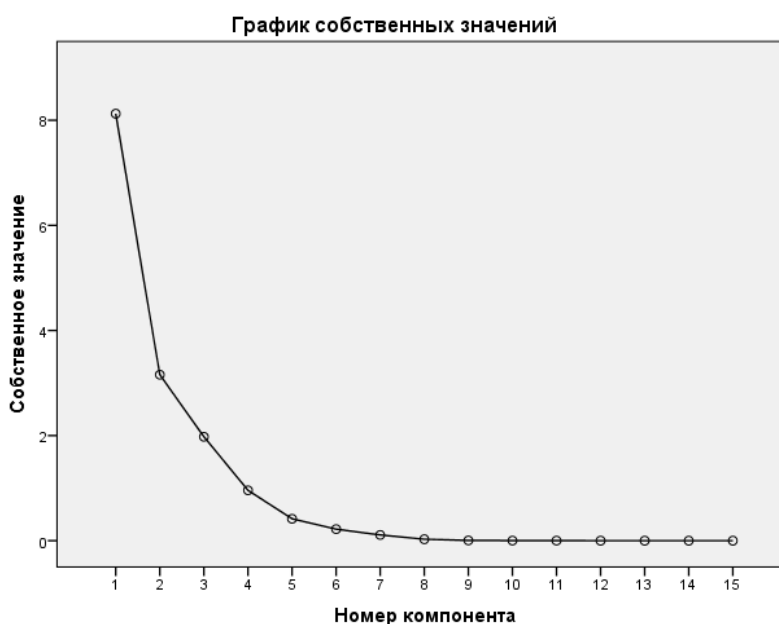


Рис. 4.9. Графік власних значень кореляційної матриці факторів.

В таблицях 4.2 та 4.3 представлено матрицю компонент для кожної властивості досліджуваного об'єкта та матрицю коефіцієнтів значень компонент.

Таблиця 4.2. Матриця компонент.				Таблиця 4.2. Матриця значень коефіцієнтів компонент.			
<b>Матриця компонентів<sup>а</sup></b>				<b>Матриця коефіцієнтів значень компонентів</b>			
	Компонент				Компонент		
	1	2	3		1	2	3
VC_b1	,870	,128	-,394	VC_b1	,107	,041	-,199
VC	,724	-,614	,107	VC	,089	-,194	,054
FVC_b2	,911	,104	-,320	FVC_b2	,112	,033	-,162
FVC	,817	-,511	,220	FVC	,101	-,162	,111
FEV1	,679	-,269	,649	FEV1	,084	-,085	,328
FEV2_b4	-,446	,688	,527	FEV2_b4	-,055	,218	,266
FEV2	-,511	,702	,472	FEV2	-,063	,222	,239
PEF_b5	,861	,322	-,229	PEF_b5	,106	,102	-,116
PEF	,855	-,092	,295	PEF	,105	-,029	,149
MEF50_b6	,910	,345	-,188	MEF50_b6	,112	,109	-,095
MEF50	,892	,083	,222	MEF50	,110	,026	,112
MEF25_b7	,793	,585	-,046	MEF25_b7	,098	,185	-,023
MEF225	,753	,416	,278	MEF225	,093	,132	,141
MVV_b8	,335	,642	,214	MVV_b8	,041	,203	,108
MVV	,137	-,539	,624	MVV	,017	-,171	,315
Метод выделения факторов: метод главных компонент. <sup>а</sup>				Метод выделения факторов: метод главных компонент.			
а. Извлечено компонентов - 3.				Оценки компонентов.			

### Завдання для виконання комп'ютерного практикуму

- Створити файл початкових даних  $X$  у вигляді матриці  $x_{ij}$  розмірності  $(m \times n)$ , де  $i$  – об'єкт дослідження,  $j$  – характеристика відповідного об'єкту, згідно з варіантом індивідуального завдання.
- Виконати попередній аналіз інформації, що підлягає стисненню без втрати інформативності за відповідним методом факторного аналізу даних.
- З метою врахування розмірності властивостей об'єктів дослідження виконати нормалізацію початкових даних з отриманням матриці  $Y = \{y_{ij}\}$ , де  $y_{ij}$  – нормоване значення  $j$ -ї властивості  $i$ -го об'єкту.
- Реалізувати процедуру факторного аналізу даних засобами системи SPSS Statistics в такій послідовності:
  - вибрати пункт меню Редактора даних системи SPSS для введення масиву початкової інформації;
  - визначити типи та формати даних для кожного із параметрів об'єктів дослідження;
  - ввести початкові дані згідно з вказаним форматом;
  - сформувати файл початкових даних для розв'язання задачі факторного аналізу;

- вибрати команду головного меню **Анализ→Снижение размерности→Факторный анализ**;
  - у вікні «Факторный анализ» вибрати дані із створеного масиву інформації, що підлягають факторному аналізу;
  - на вкладці «Описательные...» вказати можливість виведення значень описової статистики початкових даних та їх кореляційну матрицю;
  - у вікні «Факторный анализ: Выделение факторов» вибрати метод визначення кількості латентних факторів шляхом вказання сумарної дисперсії або задати граничне власне значення кореляційної матриці факторів;
  - у вікні «Факторный анализ: Значения факторов» задати можливість збереження значень факторів як змінних;
  - запустити процедуру створення файлу результатів факторного аналізу даних;
  - експортувати отриманий файл результатів за вказаною адресою у відповідному форматі.
5. Представити отримані залежності початкових властивостей об'єктів від  $k$  некорельованих між собою компонентів (латентних факторів) у вигляді формули (4.2).
6. Зробити висновки за результатами факторного аналізу даних.

### **Контрольні запитання**

1. Задача факторного аналізу.
2. Як можна представити модель для опису кожного параметра?
3. Що називається латентним фактором?
4. Які методи використовуються для визначення кількості латентних факторів?
5. Яка послідовність реалізації факторного аналізу даних в системі SPSS Statistics?
6. Яким чином можна отримати значення латентних змінних для сукупності об'єктів початкової матриці даних?
7. Як зберегти значення факторів у вигляді змінних?
8. Яким чином виконується оцінка інформативності отриманих латентних факторів?

## Комп'ютерний практикум №5

### КЛАСТЕРНИЙ АНАЛІЗ ДАНИХ

**Мета:** отримання практичних навиків розв'язання задачі кластеризації за допомогою програми SPSS Statistics.

#### Теоретичні відомості

Термін “кластерний аналіз” (англ. Data clustering) вперше вжив англійський вчений Р.Тріон в 1939 році.

Задача кластеризації належить до широкого класу задач навчання без вчителя і полягає в поділі досліджуваної множини об'єктів на групи «подібних» об'єктів, які називаються кластерами. Слово кластер англійського походження (cluster), перекладається як згусток, пучок, група. Споріднені поняття, що використовуються в літературі – клас, таксон, згущення. Часто рішення задачі розбиття множини елементів на кластери називають кластерним аналізом.

За допомогою кластерного аналізу досліджуваної сукупності об'єктів, що представлена у вигляді матриці „об'єкти-властивості”, розбивають на невелику кількість однорідних груп (їх кількість наперед може бути відома або невідома). Навчання в кластерному аналізі не існує.

Матриця „об'єкти-властивості” має такий вигляд:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix},$$

де  $x_{ij}$  – значення  $j$ -ї властивості для об'єкту номером  $i$ . Тобто маємо  $n$  об'єктів і  $m$  властивостей, що описують ці об'єкти.

Класичним прикладом такого поділу на кластери є якісно різна ефективність лікарського препарату (одужання або погіршення стану) залежно від статі хворого. Розрив простору існування чинника може виникати так само при певній комбінації деякої сукупності незалежних змінних. Проте побудувати якісну і таку, що має фізичний сенс модель для розривної області не можливо по об'єктивних причинах. Потрібно знайти чинник або комбінацію чинників, відповідальних за розрив і будувати моделі для кожної з виділених підобластей

моделі. Тобто одна модель описує вплив препарату на організм чоловіка, інша – на організм жінки і тому подібне

Рішенням задачі класифікації є віднесення кожного з об'єктів даних до одного (або кількох) з заздалегідь визначених класів і побудова в кінцевому рахунку одним з методів класифікації моделі даних, що визначає розбиття множини об'єктів даних на класи.

У задачі кластеризації віднесення кожного з об'єктів даних здійснюється до одного (або кількох) з заздалегідь невизначених класів. Розбиття об'єктів даних по кластерам здійснюється при одночасному їх формуванні. Визначення кластерів і розбиття по ним об'єктів даних виражається в підсумковій моделі даних, яка є рішенням задачі кластеризації.

Рішення задачі кластерного аналізу виконується в наступній послідовності:

- формування вибірки для аналізу;
- вибір сукупності ознак, що характеризують об'єкт;
- вибір міри схожості (відстані) між об'єктами і їх розрахунок;
- формування кластерів;
- аналіз отриманих результатів.

Більшість алгоритмів КА відносять до так званих агломеративних процедур, які спочатку об'єднують в групи найближчі об'єкти, а потім до них приєднують далші.

Існують різні міри схожості (відстані між об'єктами).

Для вибору схожості між об'єктами користуються поняттям відстані  $d_{ij}(O_i, O_j)$  між об'єктами  $O_i$  і  $O_j$ . Чим менше відстань, тим більше схожими вважаються об'єкти. Щоб бути придатною для визначення відстані, пропонується міра повинна відповідати наступним властивостям:

- симетрією  $d_{ij}(O_i, O_j) = d_{ji}(O_j, O_i)$ ;
- мінімальною відстанню об'єкту до самого себе  $d_{ii}(O_i, O_i)=0$ ;
- монотонною зміною  $d_{ij}$  в описуваному просторі;
- змістовною інтерпретацією міри (бажано).

Найчастіше в КА використовують міри, що базуються на узагальненій відстані Махаланобіса, яку задають формулою:

$$d_{ij} = \sqrt{(X_i - X_j)^T \Lambda^{-1} \Sigma^{-1} (X_i - X_j)}, \quad (5.1)$$

де  $X_i$  –  $i$ -й вектор спостережень;  $\Lambda$  – симетрична невід'ємно визначена матриця вагових коефіцієнтів (зазвичай діагональна);  $\Sigma$  – коваріаційна матриця сукупності, з якої вибрані спостереження.

Реально використовуються такі часткові види відстаней:

1. Евклідова відстань. Ця відстань використовується при виконанні наступних умов:

- компоненти  $X$  взаємно незалежні, мають одну і ту ж дисперсію;
- компоненти є однорідними за фізичним змістом.

Якщо різні властивості, що характеризують об'єкт, мають різну важливість і її можна оцінити (хоч би за допомогою експертів) використовують зважену евклідову відстань:

$$d_{ij} = \sqrt{\sum_{k=1}^m \omega_k (x_{ik} - x_{jk})^2}, \quad (5.2)$$

де  $\omega_k$  – ваговий коефіцієнт для  $k$ -ї властивості. При цьому зазвичай приймається  $0 \leq \omega_k \leq 1$  для всіх  $k$ .

1. Хеммінгова відстань. Інколи його називають відстанню міських кварталів (тобто шлях від перехрестя до перехрестя не безпосередньо, а тільки по вулицях). Її визначають за формулою:

$$d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}|; \quad (5.3)$$

2. Відстань між класами. Вищезгадані міри визначають відстань між об'єктами. Для виконання кластерного аналізу необхідно встановити, що вважати за відстань між кластерами. Зазвичай використовують такі міри близькості груп (відстаней між кластерами):

– відстань, що обчислюється за принципом „найближчого сусіда”, є мінімальною відстанню між парою об'єктів, кожен з яких знаходиться в іншому кластері. Його обчислюють за формулою:

$$d(S_l, S_k) = \min d(X_i, X_j) \quad X_i \in S_l, X_j \in S_k \quad (5.4)$$

– відстань, що обчислюється за принципом „далекого сусіда”, є максимальною між парою об'єктів, кожен з яких знаходиться в іншому кластері. Його обчислюють за формулою:

$$d(S_l, S_k) = \max d(X_i, X_j), \quad X_i \in S_l, X_j \in S_k; \quad (5.5)$$

– відстань, що обчислюється за „центрами тяжіння” кластерів за формулою:

$$d(S_l, S_k) = d\left(\bar{X}(l), \bar{X}(k)\right), \quad (5.6)$$

де  $\bar{X}(l)$  – середнє арифметичне векторних спостережень, які входять в кластер  $S_l$ . Таким чином, це відстань між „центрами тяжіння” відповідних кластерів;

– відстань, що обчислюється за принципом „середнього зв'язку”, є арифметичною середньою всіх можливих пар комбінацій між об'єктами, що входять в різні кластери. Ця відстань обчислюється за формулою:

$$d(S_l, S_k) = \frac{1}{n_l n_k} \sum \sum d(X_i, X_j), \quad X_i \in S_l, X_j \in S_k. \quad (5.7)$$

Існує досить велика кількість різних процедур кластерного аналізу. Для порівняння якості розбиття на класи використовується ряд функціоналів якості. Найбільш вживаними з них є:

– сума внутрішньокласових дисперсій відстаней, що визначається за формулою:

$$Q = \sum_{k=1}^p \sum d(X_i, \bar{X}_k), X_j \in S_k \quad (5.8)$$

де  $p$  – кількість кластерів;

– сума попарних внутрішньокласових відстаней, що визначається за формулою :

$$Q = \sum_{k=1}^p \sum_{X_j \in S_k} d(X_j, X_j) \quad (5.9)$$

де  $p$  – кількість кластерів.

Метою кластерного аналізу є пошук існуючих реальних структур даних. Різні процедури кластерного аналізу для одних і тих же даних можуть давати різне розбиття на кластери (як по їх кількості, так і по складу). Більшість методів кластерного аналізу не мають строго статичного обґрунтування.

Тобто, дана множина об'єктів даних  $X$ , кожний з яких представляється набором атрибутів. Необхідно побудувати множину кластерів  $C$  і відображення  $F$  множини  $X$  на множину  $C$  ( $F: X \rightarrow C$ ). Відображення  $F$  задає модель даних, що є рішенням задачі. Якість рішення задачі визначається кількістю вірно класифікованих об'єктів даних.

Множину  $X$  визначимо наступним чином:

$$X = \{x_1, x_2, \dots, x_j, \dots, x_n\}, \quad (5.10)$$

де  $x_j$  – досліджуваний об'єкт.

Прикладом такої множини може бути набір даних про іриса, з якими в середині 30-х років минулого століття працював відомий статистик Р.А. Фішер (ці дані часто називають ірисами Фішера). Він розглянув три класи ірисів: *Iris setosa*, *Iris versicolor* и *Iris virginica*. Для кожного з них було представлено по 50 екземплярів з різними значеннями чотирьох параметрів: довжина і ширина чашолистків, довжина і ширина пелюстки. У табл. 5.1 представлені дані по п'яти екземплярам для кожного класу.

Кожний із об'єктів характеризується набором параметрів. У прикладі з ірисами, як уже зазначалося, такими параметрами є довжина і ширина чашолистків, довжина і ширина пелюстки.

При здійсненні кластеризації важливо знати скільки кластерів має бути побудовано. Припускається, що кластеризація повинна виявити природні локальні скупчення об'єктів. Тому кількість кластерів є параметром, який часто ускладнює вид алгоритму, якщо він наперед невідомий, і суттєво впливає на якість результату, якщо він відомий.

Таблиця 5.1. Іриси Фішера (фрагмент).

№	Довжина чашолистка	Ширина чашолистка	Довжина пелюстки	Ширина пелюстки	Клас
1	5,1	3,5	1,4	0,2	Iris setosa
2	4,9	3,0	1,4	0,2	Iris setosa
3	4,7	3,2	1,3	0,2	Iris setosa
4	4,6	3,1	1,5	0,2	Iris setosa
5	5,0	3,6	1,4	0,2	Iris setosa
51	6,9	3,2	4,4	1,4	Iris versicolor
52	6,4	3,2	4,5	1,5	Iris versicolor
53	6,9	3,1	4,9	1,5	Iris versicolor
54	5,5	2,3	4,0	1,3	Iris versicolor
55	6,5	2,8	4,6	1,5	Iris versicolor
101	6,3	3,3	6,0	2,5	Iris virginica
102	6,8	2,7	5,1	1,9	Iris virginica
103	7,1	3,0	5,9	2,1	Iris virginica
104	6,3	2,9	5,6	2,1	Iris virginica
105	6,5	3,0	5,8	2,2	Iris virginica

Кількість методів розбиття множини на кластери досить велика, але базові методи можна поділити на ієрархічні та неієрархічні. На рис.5.1. представлено узагальнену класифікацію алгоритмів кластерного аналізу.

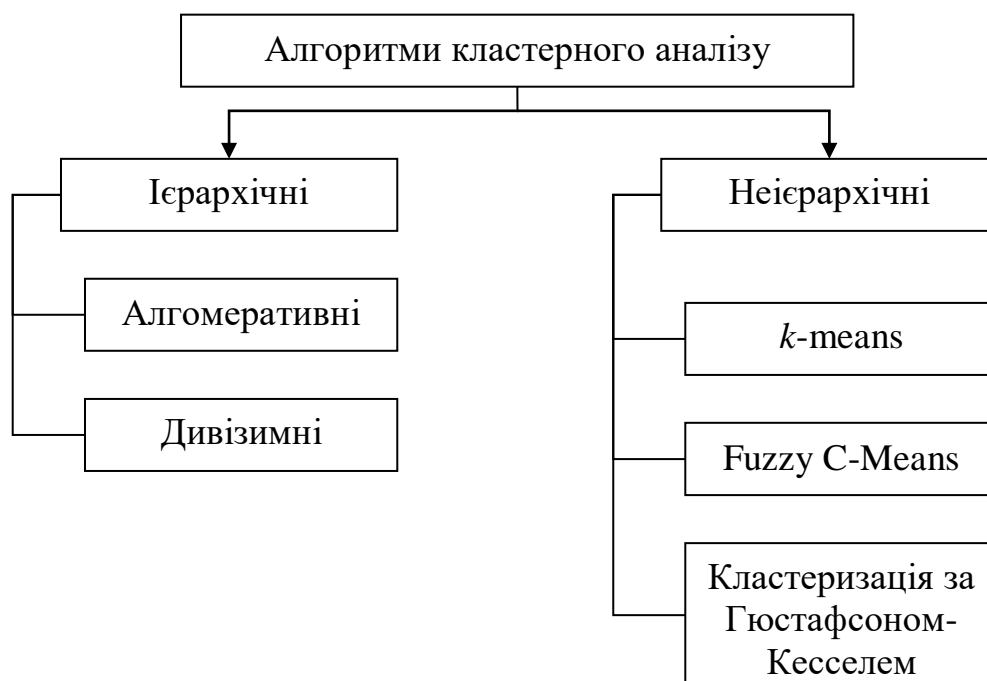


Рис. 5.1. Види алгоритмів кластерного аналізу.

Для виконання кластерного аналізу даних зручно використовувати систему SPSS STATISTICS. В якості початкової інформації можна використати

початковий масив даних із практичної роботи №4. Кластеризацію будемо виконувати двома найбільш популярними методами: ієрархічний алгомеративний та неієрархічний  $k$ -means ( $k$ -середніх).

Як приклад початкових даних для кластеризації розглянемо інформацію, що представлена в таблиці 5.1, а вікно з введеними даними – наведено на рис.5.2.

	Довжина_чашолистка	Ширина_чашолистка	Довжина_пелюстки	Ширина_пелюстки	пер	пі
1	5,10	3,50	1,40	,20		
2	4,90	3,00	1,40	,20		
3	4,70	3,20	1,30	,20		
4	4,60	3,10	1,50	,20		
5	5,00	3,60	1,40	,20		
6	6,90	3,20	4,40	1,40		
7	6,40	3,20	4,50	1,50		
8	6,90	3,10	4,90	1,50		
9	5,50	2,30	4,00	1,30		
10	6,50	2,80	4,60	1,50		
11	6,30	3,30	6,00	2,50		
12	6,80	2,70	5,10	1,90		
13	7,10	3,00	5,90	2,10		
14	6,30	2,90	5,60	2,10		
15	6,50	3,00	5,80	2,20		
16						

Рис.5.2. Вікно з введеними початковими даними для кластерного аналізу.

Для виконання кластерного аналізу даних вибирається команда головного меню **Анализ**→**Классификация**→**Кластеризация К-средними...**. У вікні «Кластерный анализ методом К-средних» (рис.5.3) вибираються змінні, що підлягають кластерному аналізу та задається кількість створюваних кластерів.

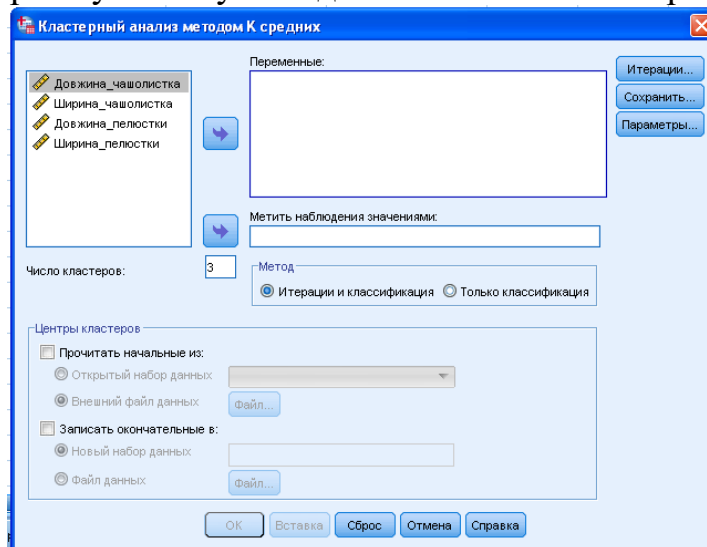


Рис.5.3. Вікно кластерного аналізу даних методом К-середніх.

При переході на вкладку «Параметры...» вибирається значення початкових центрів кластерів для їх виведення та залишкові кластери для кожного спостереження (рис. 5.4).

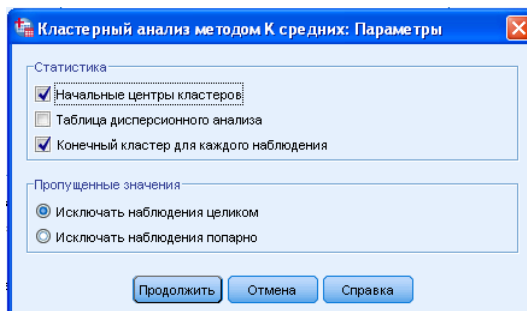


Рис.5.4. Вікно кластерного аналізу методом К-середніх: параметри.

Натисканням клавіші **Продолжить** виконується перехід на вкладку «Сохранить...», а потім у вікні «Кластер К-средних: Сохранить новую переменную» ставиться галочка в пункті «Принадлежность к кластеру» (рис. 5.5).

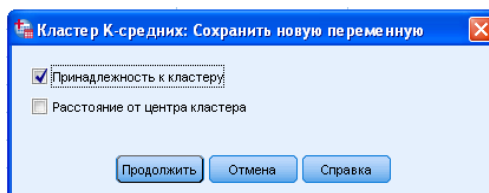


Рис.5.5. Вікно кластерного аналізу методом К-середніх:Зберегти нову змінну.

Натисканням клавіш **Продолжить**→**ОК** генерується створення вихідного файлу **Выводи**, який містить результати, що наведено на рис. 5.6.

### Быстрая кластеризация

Начальные центры кластеров			
	Кластеризовать		
	1	2	3
Довжина_чашолистка	5,00	7,10	5,50
Ширина_чашолистка	3,60	3,00	2,30
Довжина_пелюстки	1,40	5,90	4,00
Ширина_пелюстки	,20	2,10	1,30

Хронология итераций <sup>a</sup>			
Итерация	Изменение центров кластеров		
	1	2	3
1	,349	,676	,878
2	,000	,194	,220
3	,000	,179	,163
4	,000	,000	,000

а. Сходимость достигнута благодаря небольшому изменению центров кластеров или без него.  
 Максимальное изменение абсолютной координаты для любого центра: ,000. Текущая итерация - 4.  
 Минимальное расстояние между начальными центрами: 2,702.

а) початкові центри кластерів

**Принадлежность к кластерам**

Номер наблюдения	Кластеризован	Расстояние
1	1	,326
2	1	,283
3	1	,205
4	1	,332
5	1	,349
6	3	,546
7	3	,290
8	3	,651
9	3	1,232
10	3	,190
11	2	,641
12	2	,723
13	2	,550
14	2	,326
15	2	,162

б) належність точок до кластерів

**Конечные центры кластеров**

	Кластеризовать		
	1	2	3
Довжина_чашолистка	4,86	6,60	6,44
Ширина_чашолистка	3,28	2,98	2,92
Довжина_пелюстки	1,40	5,68	4,48
Ширина_пелюстки	,20	2,16	1,44

**Расстояния между конечными центрами кластеров**

Кластеризовать	1	2	3
1		5,028	3,695
2	5,028		1,410
3	3,695	1,410	

**Число наблюдений в каждом кластере**

Кластеризовать	1	5,000
	2	5,000
	3	5,000
Допустимо		15,000
Пропущенные		,000

в) центри кластерів і відстані між ними

Рис. 5.6. Результаты кластерного анализа методом К-средних.

Також у файлі з початковим масивом інформації з'являється нова змінна, що вказує на приналежність об'єктів до відповідних кластерів (рис. 5.7)

	Довжина_чашолистка	Ширина_чашолистка	Довжина_пелюстки	Ширина_пелюстки	QCL_1	пер	пер
1	5,10	3,50	1,40	,20	1		
2	4,90	3,00	1,40	,20	1		
3	4,70	3,20	1,30	,20	1		
4	4,60	3,10	1,50	,20	1		
5	5,00	3,60	1,40	,20	1		
6	6,90	3,20	4,40	1,40	3		
7	6,40	3,20	4,50	1,50	3		
8	6,90	3,10	4,90	1,50	3		
9	5,50	2,30	4,00	1,30	3		
10	6,50	2,80	4,60	1,50	3		
11	6,30	3,30	6,00	2,50	2		
12	6,80	2,70	5,10	1,90	2		
13	7,10	3,00	5,90	2,10	2		
14	6,30	2,90	5,60	2,10	2		
15	6,50	3,00	5,80	2,20	2		
16							

Рис. 5.7. Вікно файлу з початковими даними та результатами кластеризації.

За результатами кластерного аналізу видно, що масив початкових даних поділено на три кластери, кожний з яких складається з п'яти елементів. При цьому результати кластеризації і класи в початковій таблиці 5.1. співпадають.

Для кластеризації за ієрархічним кластерним аналізом в головному вікні програми, вибирається команда головного меню **Анализ**→**Классификация**→**Иерархическая кластеризация...** У вікні «Иерархический кластерный анализ» (рис.5.8) вибираються змінні, що підлягають кластерному аналізу.

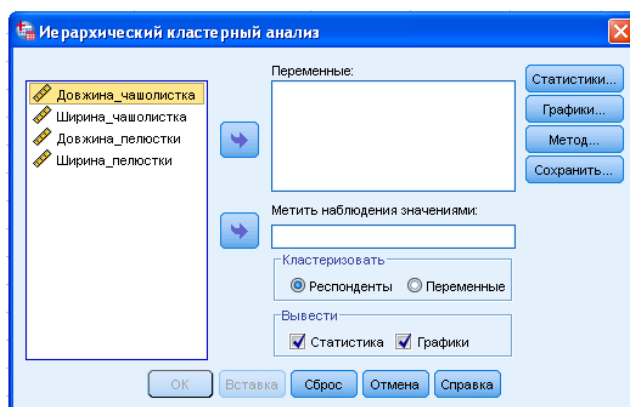


Рис.5.8. Вікно ієрархічного кластерного аналізу.

У цьому вікні вибирається вкладка «Графики» (рис. 5.9) та проставляється галочку для виведення дендрограми.

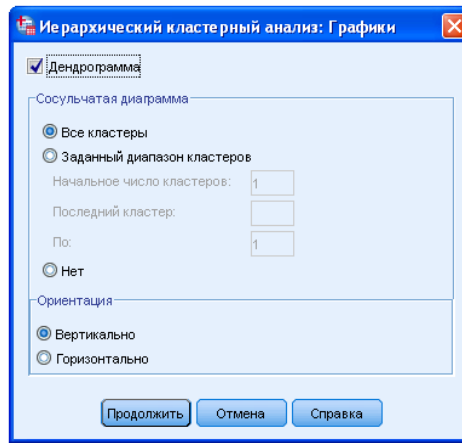


Рис. 5.9. Вікно ієрархічного кластерного аналізу з вибором графіку в вигляді дендограми.

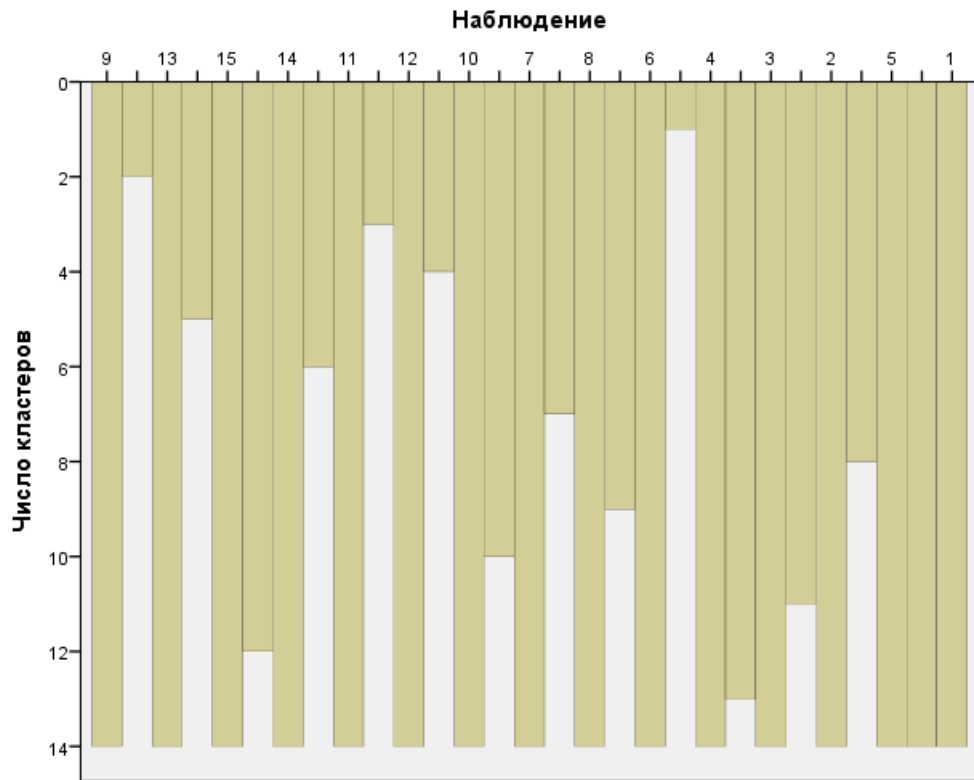
Вибором пунктів Продовжити→ОК генерується створення вихідного файлу Виводи, що містить результати, що наведено на рис. 5.10.

### Метод средней связи (между группами)

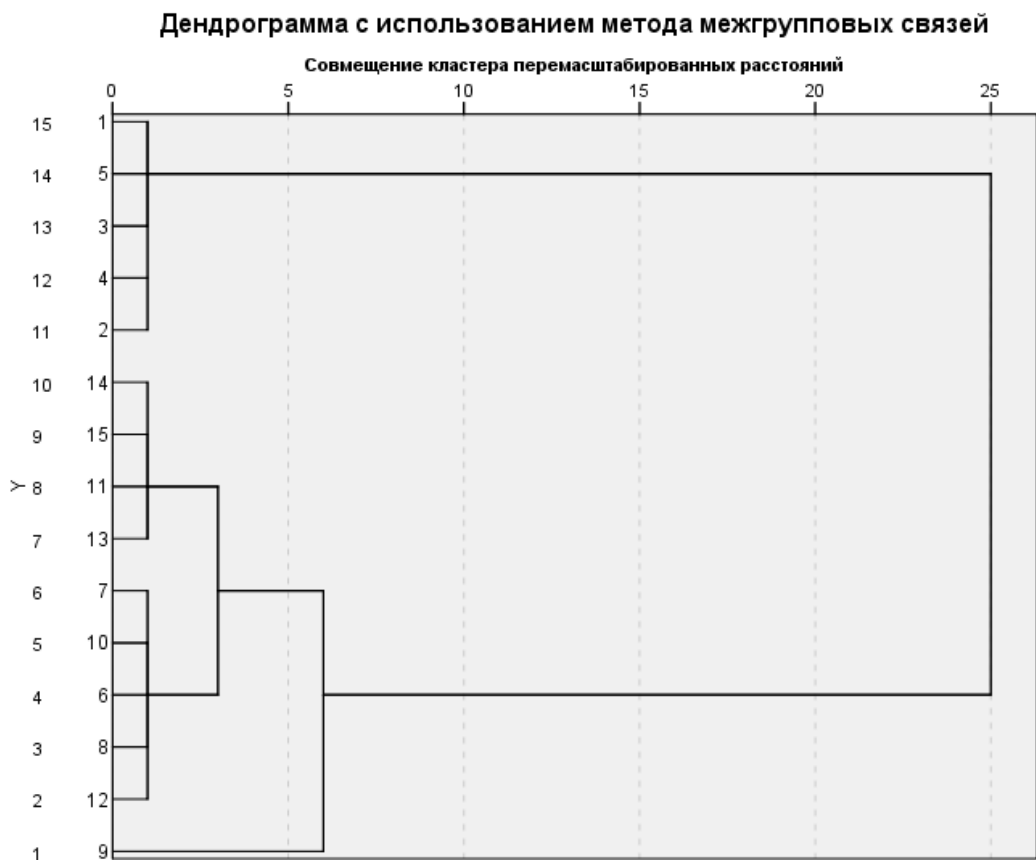
Порядок агломерации (кластеров)

Этап	Объединенный кластер		Коэффициенты	Этап первого появления кластера		Следующий этап
	Кластер 1	Кластер 2		Кластер 1	Кластер 2	
1	1	5	,020	0	0	7
2	3	4	,060	0	0	4
3	14	15	,100	0	0	9
4	2	3	,100	0	2	7
5	7	10	,180	0	0	8
6	6	8	,270	0	0	8
7	1	2	,337	1	4	14
8	6	7	,350	6	5	11
9	11	14	,370	0	3	10
10	11	13	,673	9	0	12
11	6	12	,702	8	0	12
12	6	11	2,098	11	10	13
13	6	9	4,166	12	0	14
14	1	6	19,929	7	13	0

а) етапи утворення кластерів



б) діаграма утворення кластерів



в) дендрограма кластерного аналізу

Рис. 5.10. Вікна з результатами ієрархічного кластерного аналізу.

Отримана дендрограма (рис. 5.10 в) дозволяє визначити послідовне поєднання об'єктів на 4, потім на 3 та на 2 кластери та вказати числові значення відстаней між кластерами.

### **Завдання для виконання комп'ютерного практикуму**

1. Створити файл початкових даних шляхом використання множини об'єктів, які характеризуються рядом характеристик та надані в індивідуальному завданні до комп'ютерного практикуму, для цього:

– здійснити попередній аналіз інформації, що підлягає класифікації за методами кластерного аналізу;

– вибрати пункт меню Редактора даних системи SPSS для введення початкової інформації;

– визначити типи та формати даних для кожного із параметрів об'єктів дослідження;

– ввести початкові дані згідно з вказаним форматом;

– сформувати файл початкових даних для розв'язання задачі кластерного аналізу.

2. Виконати класифікацію об'єктів засобами швидкого кластерного аналізу з використанням сформованого масиву початкових даних. :

– вказати кількість кластерів на які планується класифікувати об'єкти початкових даних;

– в режимі аналізу вказати параметри, за якими буде виконана класифікація об'єктів;

– виконати класифікацію об'єктів швидким кластерним аналізом.

3. Виконати класифікацію об'єктів засобами ієрархічного кластерного аналізу з використанням сформованого масиву початкових даних. :

– вказати кількість кластерів на які планується класифікувати об'єкти початкових даних;

– в режимі аналізу вказати параметри, за якими буде виконана класифікація об'єктів;

– виконати класифікацію об'єктів ієрархічним кластерним аналізом.

4. Порівняти результати класифікації об'єктів обома вказаними методами.

Зробити висновки за результатами кластерного аналізу даних.

### **Контрольні запитання**

1. Які задачі вирішує кластерний аналіз?
2. Чим відрізняються задачі класифікації «з вчителем» та «без вчителя»?
3. Як формулюється задача класифікації?
4. Що таке «кластер»?
5. Які методи розбиття множин об'єктів на кластери?
6. Які види алгоритмів кластерного аналізу?
7. Які є найбільш розповсюджені методи кластерного аналізу?

8. В чому полягає кластеризація ієрархічним алгомеративним методом?
9. Яка послідовність розв'язання задачі кластерного аналізу ієрархічним алгомеративним методом за допомогою SPSS Statistics?
10. В чому полягає кластеризація методом  $k$ -means?
11. Яка послідовність розв'язання задачі кластерного аналізу методом  $k$ -means за допомогою SPSS Statistics?
12. Що називається дендрограмою?
13. Яку інформацію можна отримати з дендрограми?

## Комп'ютерний практикум №6

### ДИСКРИМІНАТНИЙ АНАЛІЗ ДАНИХ

**Мета:** отримання практичних навиків розв'язання задач інтерпретації та класифікації дискримінантного аналізу даних за допомогою системи SPSS Statistics.

#### Теоретичні відомості

Дискримінантний аналіз є статистичним методом, який дозволяє вивчати відмінності між двома і більше групами об'єктів по декільком змінним одночасно. Його розробником є Рональд Фішер (1890-1962), класик біометрії і еволюційної біології.

«Дискримінантний аналіз» – це загальний термін, що стосується кількох тісно пов'язаних статистичних процедур. Їх можна розділити на методи інтерпретації міжгрупових відмінностей і методи класифікації спостережень за групами. Йдеться про інтерпретацію, коли розглядаються відмінності між класами. Іншими словами, при інтерпретації необхідно відповісти на питання: чи можливо, використовуючи даний набір характеристик (змінних), відрізнити один клас від іншого; наскільки добре ці характеристики дозволяють провести розрізнення і які з них найбільш інформативні. Метод, що відноситься до класифікації, пов'язаний з отриманням однієї або декількох функцій, що забезпечують можливість віднести даний об'єкт до однієї з груп. Ці функції, що називаються дискримінантними, залежать від значень характеристик таким чином, що з'являється можливість віднести кожен об'єкт до однієї з груп. Зрозуміло, дискримінантний аналіз необхідний і для інтерпретації, і для класифікації.

Характеристики, що застосовуються для того, щоб відрізнити один клас від іншого, називаються дискримінантними змінними. Ці змінні повинні вимірюватися або за інтервальною шкалою, або за шкалою відношень. Таким чином, стає можливим обчислення математичних очікувань, дисперсій і використання дискримінантних змінних в математичних рівняннях. У загальному випадку, кількість дискримінантних змінних необмежене, але в сумі кількість об'єктів повинно завжди перевищувати кількість змінних принаймні на два.

Виконаємо аналіз природи відмінностей між класами (групами), тобто принципи, що лежать в основі обчислення канонічних дискримінантних функцій.

Канонічна дискримінантна функція є лінійною комбінацією дискримінантних змінних і задовольняє певним умовам. Вона має наступне математичне представлення:

$$f_{km} = u_0 + u_1 X_{1km} + u_2 X_{2km} + \dots + u_p X_{pkm} \quad (6.1)$$

де  $f_{km}$  – значення канонічної дискримінантної функції для  $m$ -го об'єкта в групі  $k$ ;  $X_{ikm}$  – значення дискримінантної змінної  $X_i$  для  $m$ -го об'єкта в групі  $k$ ;  $u_i$  – коефіцієнти, що забезпечують виконання необхідних умов.

Коефіцієнти  $u_i$  для першої функції вибираються таким чином, щоб її середні значення для різних класів якомога більше відрізнялися один від одного. Коефіцієнти другої функції вибираються так само, тобто відповідні середні значення повинні максимально відрізнятися по класах, при цьому накладається додаткова умова, щоб значення другої функції не корелювали зі значеннями першої. Аналогічно третя функція повинна бути некорельованою з першими двома і т. д. Максимальна кількість дискримінантних функцій, яку можна отримати описаним способом, дорівнює кількості класів без одиниці або кількості дискримінантних змінних, в залежності від того, яка з цих величин менша.

Класифікація – це процес, що допомагає досліднику прийняти рішення: вказаний об'єкт «належить до» або «дуже схожий на» дану групу (клас). Таке рішення приймається на основі інформації, що міститься в дискримінантних змінних. Існує декілька способів проведення класифікації. Процедури класифікації можуть використовувати або канонічні дискримінантні функції або самі дискримінантні змінні.

Розглянемо класифікацію, коли дискримінантні змінні використовуються безпосередньо.

Фішер (1936 р.) був першим, хто запропонував, що класифікація може здійснюватися за допомогою лінійної комбінації дискримінантних змінних. Він запропонував застосовувати лінійну комбінацію, яка максимізує відмінності між класами, але мінімізує дисперсію всередині класів. Таким чином, лінійна комбінація дискримінантних змінних або так звана «лінійна класифікаційна функція Фішера», має наступний вигляд:

$$h_k = b_{k0} + b_{k1}X_1 + b_{k2}X_2 + \dots + b_{kp}X_p, \quad (6.2)$$

де  $h_k$  – значення функції для класу  $k$ ;  $b_{ki}$  – коефіцієнти, які необхідно знайти;  $X_i$  – дискримінантні змінні.

Об'єкт відноситься до класу з найбільшим значенням  $h$ .

Розглянемо послідовність розв'язання задачі кластерного аналізу на прикладі.

Нехай відповідно до результатів кластерного аналізу, об'єкти початкової таблиці ірисів Фішера були поділені на три кластери, як це показано на рис. 6.1.

	Довжина_чашолистка	Ширина_чашолистка	Довжина_пелюстки	Ширина_пелюстки	QCL_1
1	5,10	3,50	1,40	,20	1
2	4,90	3,00	1,40	,20	1
3	4,70	3,20	1,30	,20	1
4	4,60	3,10	1,50	,20	1
5	5,00	3,60	1,40	,20	1
6	6,90	3,20	4,40	1,40	3
7	6,40	3,20	4,50	1,50	3
8	6,90	3,10	4,90	1,50	3
9	5,50	2,30	4,00	1,30	3
10	6,50	2,80	4,60	1,50	3
11	6,30	3,30	6,00	2,50	2
12	6,80	2,70	5,10	1,90	2
13	7,10	3,00	5,90	2,10	2
14	6,30	2,90	5,60	2,10	2
15	6,50	3,00	5,80	2,20	2
16					

Рис. 6.1. Вікно, що містить таблицю належності кожного об'єкта таблиці ірисів Фішера до відповідного кластеру.

Для розв'язання даної задачі за допомогою системи SPSS Statistics Далі в головному меню програми, послідовно вибираємо команди **Анализ**→**Классификация**→**Дискриминантный анализ**.... У діалоговому вікні «Дискриминантный анализ» (рис.6.2) поле **Группировать по:** призначене для вказання єдиної залежної змінної, в поле **Независимые** включається будь яка кількість змінних, що приймають участь в дискримінантному аналізі.

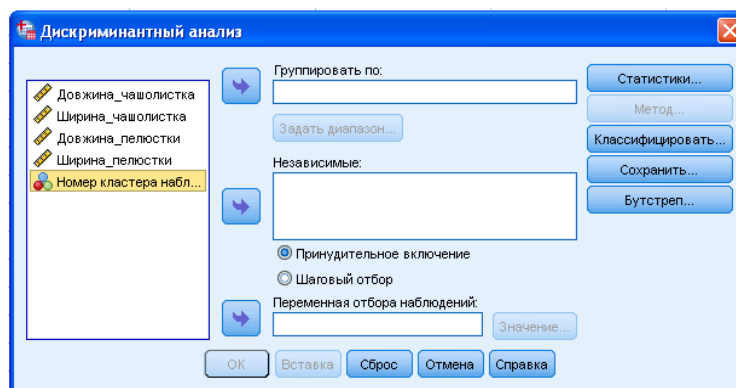


Рис. 6.2. Діалогове вікно «Дискриминантный анализ».

В даній задачі в якості залежної змінної виступає номер кластеру, а в якості незалежних – всі параметри із початкової таблиці (рис. 6.3.). Після введення залежної змінної, кнопка **Задать диапазон** дозволяє відкрити діалогове вікно «Дискриминантный анализ: Задание диапазона». Оскільки

змінна, що виступає в нашому прикладі як залежна, має три рівні: 1, 2, 3, вказуємо 1 як мінімум і 3 як максимум (рис.6.4), натискаємо Продолжить.

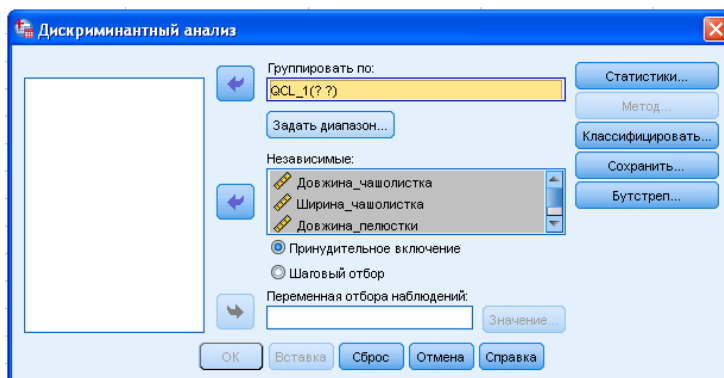


Рис. 6.3. Вікно призначення залежної і незалежних змінних в діалоговому вікні «Дискриминантный анализ».

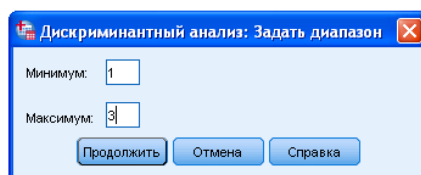


Рис. 6.4. Діалогове вікно «Дискриминантный анализ: Задание диапазона».

Для виведення лінійних дискримінантних функцій Фішера клікаємо по кнопці Статистики..., що генерує відкриття діалогового вікна «Дискриминантный анализ: Статистики», та ставимо галочку для коефіцієнтів Фішера (рис. 6.5), натискаємо Продолжить→ОК.

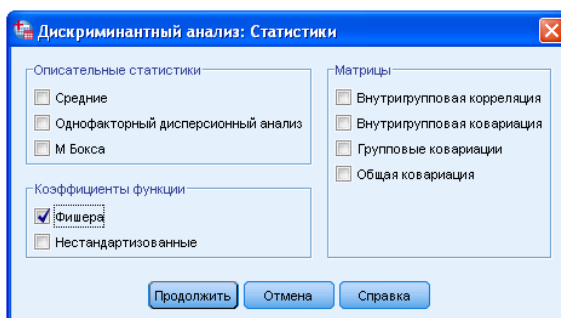


Рис. 6.5. Діалогове вікно «Дискриминантный анализ: Статистики».

У файлі Вывод отримуємо інформацію за результатами дискримінантного аналізу, що наведена на рис.6.6.

## Дискриминантный анализ

### Анализ сводки обработки наблюдений

Невзвешенные наблюдения		N	Проценты
Допустимо		15	100,0
Исключено	Отсутствующие или выходящие за пределы диапазона коды групп	0	,0
	По крайней мере одна дискриминирующая переменная	0	,0
	И отсутствующие или выходящие за пределы диапазоны коды групп, и по крайней мере одна дискриминирующая переменная	0	,0
	Всего	0	,0
Всего		15	100,0

### Статистика группы

Номер кластера наблюдения		N валидных (по списку)	
		Невзвешенных	Взвешенных
1	Довжина_чашолистка	5	5,000
	Ширина_чашолистка	5	5,000
	Довжина_пелюстки	5	5,000
	Ширина_пелюстки	5	5,000
2	Довжина_чашолистка	5	5,000
	Ширина_чашолистка	5	5,000
	Довжина_пелюстки	5	5,000
	Ширина_пелюстки	5	5,000
3	Довжина_чашолистка	5	5,000
	Ширина_чашолистка	5	5,000
	Довжина_пелюстки	5	5,000
	Ширина_пелюстки	5	5,000
Всего	Довжина_чашолистка	15	15,000
	Ширина_чашолистка	15	15,000
	Довжина_пелюстки	15	15,000
	Ширина_пелюстки	15	15,000

а) статистичний аналіз початкової інформації

## Анализ 1

### Сводка канонических дискриминантных функций

Собств. знач				
Функция	Собственное значение	% дисперсии	Суммарный %	Каноническая корреляция
1	156,283 <sup>а</sup>	99,3	99,3	,997
2	1,030 <sup>а</sup>	,7	100,0	,712

а. Для анализа использовались первые 2 из канонических дискриминантных функций.

Лямбда Уилкса				
Критерий для функций	Лямбда Уилкса	Хи-квадрат	ст.св.	Знач.
От 1 до 2	,003	60,543	8	,000
2	,493	7,434	3	,059

#### Коэффициенты стандартизованной канонической дискриминантной функции

	Функция	
	1	2
Довжина_чашолистка	1,151	,068
Ширина_чашолистка	-1,508	,190
Довжина_пелюстки	-,184	-1,450
Ширина_пелюстки	1,599	1,562

#### Матрица структуры

	Функция	
	1	2
Довжина_пелюстки	,571*	-,094
Ширина_пелюстки	,528	,542*
Довжина_чашолистка	,170	-,436*
Ширина_чашолистка	-,043	,253*

Объединенные внутригрупповые корреляции между дискриминирующими переменными и стандартизованными каноническими дискриминантными функциями

Переменные упорядочиваются по абсолютной величине корреляции в функции.

\*. Наибольшая абсолютная корреляция между каждой переменной и любой дискриминантной функцией

б) коефіцієнти канонічної дискримінантної функції

### Функции в центроидах групп

Номер кластера наблюдения	Функция	
	1	2
1	-15,159	,365
2	11,476	,883
3	3,683	-1,248

Нестандартизованные канонические дискриминантные функции, вычисленные в групповых средних

### Статистика классификаций

#### Априорные вероятности для групп

Номер кластера наблюдения	Априорная	Используемые в анализе наблюдения	
		Невзвешенных	Взвешенных
1	,333	5	5,000
2	,333	5	5,000
3	,333	5	5,000
Всего	1,000	15	15,000

#### Коэффициенты функции классификации

Линейные дискриминантные функции Фишера	1 2 3		
	1	2	3
Довжина_чашолистка	17,264	93,132	70,603
Ширина_чашолистка	32,887	-102,956	-64,482
Довжина_пелюстки	-4,938	-24,948	-8,925
Ширина_пелюстки	-20,594	297,065	181,491
(Константа)	-91,468	-405,009	-244,977

в) коефіцієнти класифікаційної функції

Рис 6.6. Вікно з результатами дискримінантного аналізу даних.

За результатами дискримінантного аналізу інформації, що наведено в таблиці 6.1, отримано функції класифікації, коефіцієнти яких надано в таблиці 6.6в).

Отже, перша дискримінантна функція має такий вигляд:

$$F1 = 17,264 * \text{Довжина\_чашолистка} + 32,887 * \text{Ширина\_чашолистка} - 4,938 * \text{Довжина\_пелюстки} - 20,594 * \text{Ширина\_пелюстки} - 91,468$$

друга дискримінантна функція:

$$F2 = 93,132 * \text{Довжина\_чашолистка} - 102,956 * \text{Ширина\_чашолистка} - 24,948 * \text{Довжина\_пелюстки} + 297,065 * \text{Ширина\_пелюстки} - 405,009$$

третя дискримінантна функція:

$$F3 = 70,603 * \text{Довжина\_чашолистка} - 64,482 * \text{Ширина\_чашолистка} - 8,925 * \text{Довжина\_пелюстки} + 181,491 * \text{Ширина\_пелюстки} - 244,977$$

Після розв'язання задачі інтерпретації є можливість розв'язати задачу класифікації.

Нехай є новий об'єкт, характеристики якого наведено в таблиці 6.1.(для нашого прикладу нова квітка ірису)

Треба визначити, до якої із уже існуючих груп він відноситься.

Таблиця 6.1. Характеристики нового об'єкту (нової квітки ірису)

Довжина чашолистка	Ширина чашолистка	Довжина пелюстки	Ширина пелюстки
6,60	3,00	4,20	1,25

Для цього необхідно обчислити значення кожної дискримінантної функції шляхом підстановки в них характеристик нового об'єкту. Розрахунки можна виконати в середовищі MS Excel (рис. 6.6)/

Коефициенты функции классификации				Новый объект	
	Номер кластера наблюдения			Довжина_чашолистка	Ширина_чашолистка
	1	2	3		
Довжина_чашолистка	17,264	93,132	70,603	6,60	
Ширина_чашолистка	32,887	-102,956	-64,482	3,00	
Довжина_пелюстки	-4,938	-24,948	-8,925	4,20	
Ширина_пелюстки	-20,594	297,065	181,491	1,25	
(Константа)	-91,468	-405,009	-244,977		
Сума	74,6633	167,34385	216,93565		

Рис.6.6. Вікно з результатами класифікації нового об'єкта за результатами дискримінантного аналізу даних.

Відповідно до методики дискримінантного аналізу новий об'єкт відноситься до тієї група, для якої значення дискримінантної функції є найбільшим. В нашому випадку це група або кластер 3.

### **Завдання для виконання комп'ютерного практикуму**

1. Створити файл початкових даних шляхом використання інформації, що задана в індивідуальному завданні до комп'ютерного практикуму.

Для цього необхідно:

– виконати попередній аналіз інформації, що підлягає дискримінантному аналізу;

– вибрати пункт меню Редактора даних системи SPSS для введення початкової інформації;

– визначити імена, типи та формати даних для кожного із параметрів об'єктів дослідження;

– ввести початкові дані згідно з вказаним форматом;

– сформувати файли початкових даних для розв'язання задачі дискримінантного аналізу.

2. Розв'язати задачу дискримінації груп об'єктів засобами дискримінантного аналізу з використанням сформованих масивів початкових даних. :

– вказати кількість груп, на які планується класифікувати об'єкти початкових даних;

– виконати дискримінацію об'єктів з отриманням класифікаційних функцій для кожної із груп об'єктів.

3. Виконати класифікацію нового об'єкта засобами дискримінантного аналізу з використанням отриманих класифікаційних функцій:

Здійснити аналіз отриманих результатів та зробити висновки за результатами виконання дискримінантного аналізу даних.

### **Контрольні запитання**

1. Дайте визначення дискримінантному аналізу.
2. Які задачі вирішує дискримінантний аналіз?
3. Які змінні називаються дискримінантними?
4. В чому полягає задача дискримінації?
5. В чому полягає задача інтерпретації між групових відмінностей.
6. В чому полягає задача класифікації?
7. Який вигляд має канонічна дискримінантна функція?
8. Яке призначення канонічної дискримінантної функції?
9. Який вигляд має класифікаційна функція Фішера?
10. Яке призначення класифікаційної функції Фішера?
11. Яка послідовність виконання дискримінантного аналізу за допомогою програми SPSS Statistics?
12. Яка послідовність розв'язання задачі класифікації засобами SPSS Statistics?

## Комп'ютерний практикум №7

### РЕГРЕСІЙНИЙ АНАЛІЗ ДАНИХ

**Мета:** отримання практичних навиків отримання рівнянь регресії засобами системи SPSS Statistics.

#### Теоретичні відомості

Залежність між змінними величинами  $X$  та  $Y$  може бути описана різними способами. Зокрема, будь яку форму зв'язку між ними можна виразити рівнянням загального виду  $y = f(x)$ , де  $Y$  розглядають в якості залежної змінної або функції від іншої незалежної – незалежної змінної величини  $X$ , що називається аргументом. Відповідність між аргументом і функцією може бути задано таблицею, формулою або графіком. Зміна функції в залежності від зміни одного або декількох аргументів називається регресією.

Для представлення регресії слугують кореляційні рівняння або рівняння регресії, емпіричні і теоретичні вираховані ряди регресії, їх графіки, що називаються лініями регресії, а також коефіцієнти лінійної та нелінійної регресії.

Лінійна залежність між змінними  $X$  та  $Y$  описується рівнянням загального вигляду  $\bar{y}_x = a + bx_1 + cx_2 + dx_3 + \dots$ , де  $a, b, c, d, \dots$  – параметри рівняння, що визначають співвідношення між аргументами  $x_1, x_2, x_3, \dots$  і функцією  $\bar{y}_x$ . На практиці враховують не всі можливі, а лише деякі аргументи, в самому простому випадку – всього один:

$$\bar{y}_x = a + bx. \quad (7.1)$$

Визначення параметрів лінійної регресії. Відомо, що сума квадратів відхилень варіант  $x_i$  від їх середнього  $\bar{x}$  має бути найменшою, тобто  $\sum (x_i - \bar{x})^2 \Rightarrow \min$ . Ця теорема складає основу методу найменших квадратів. Даний метод запропонований на початку XIX століття А.М. Лежандром та незалежно від нього К. Гаусом. Він дозволяє найбільш точно вирівнювати емпіричні ряди. По відношенню до лінійної регресії вимогам цієї теореми задовольняє певна система рівнянь, що називаються нормальними:

$$\begin{cases} an + b \sum x = \sum y \\ a \sum x + b \sum x^2 = \sum xy \end{cases} \quad (7.2)$$

де  $n$  – об'єм вибірки.

Загальне розв'язання даної системи рівнянь відносно параметрів  $a$  та  $b$  дозволяє отримати наступні результати:

$$\begin{aligned}
D &= \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix} = n \sum x^2 - (\sum x)^2; \\
A &= \begin{vmatrix} \sum y & \sum x \\ \sum xy & \sum x^2 \end{vmatrix} = \sum y \sum x^2 - \sum x \sum xy; \\
B &= \begin{vmatrix} n & \sum y \\ \sum x & \sum xy \end{vmatrix} = n \sum xy - \sum y \sum x,
\end{aligned} \tag{7.3}$$

Звідси

$$a = \frac{A}{D}; b = \frac{B}{D} \Rightarrow a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}; b = \frac{n \sum xy - \sum y \sum x}{n \sum x^2 - (\sum x)^2}. \tag{7.4}$$

Залежність між декількома змінними величинами прийнято виражати рівнянням регресії, яке може бути лінійним і нелінійним. В найбільш простому випадку множинна лінійна регресія представляється рівнянням з двома невідомими величинами  $(x, z)$ :

$$y = a + bx + cz, \tag{7.5}$$

де  $a$  – вільний член рівняння;  $b$  і  $c$  – параметри рівняння. Для знаходження параметрів даного рівняння (згідно способу найменших квадратів) застосовують наступну систему нормальних рівнянь:

$$\begin{cases} an + b \sum x + c \sum z = \sum y \\ a \sum x + b \sum x^2 + c \sum xz = \sum xy \\ a \sum z + b \sum xz + c \sum z^2 = \sum yz \end{cases} \tag{7.6}$$

Для того щоби за емпіричними даними скласти таку систему, необхідно попередньо розрахувати  $\sum x$ ,  $\sum y$ ,  $\sum yz$ ,  $\sum xz$ ,  $\sum x^2$  та  $\sum z^2$ . Далі система вирішується відносно невідомих  $a$ ,  $b$  і  $c$  аналогічно до системи (7.2).

Наряду з лінійними кореляціями на практиці зустрічаються і нелінійні кореляції між змінними величинами. Часто залежність між змінними величинами  $X$  та  $Y$  описується рівнянням параболи другого порядку:

$$y = a + bx + cx^2. \tag{7.7}$$

Для знаходження параметрів  $a$ ,  $b$  і  $c$  використовується наступна система нормальних рівнянь:

$$\begin{cases} an + b\sum x + c\sum x^2 = \sum y \\ a\sum x + b\sum x^2 + c\sum x^3 = \sum xy \\ a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum ux^2 \end{cases} \quad (7.8)$$

Для розв'язання цієї системи рівнянь відносно параметрів  $a$ ,  $b$  і  $\tilde{n}$  необхідно попередньо розрахувати  $\sum x$ ,  $\sum y$ ,  $\sum xy$ ,  $\sum x^2$ ,  $\sum ux^2$ ,  $\sum x^3$  та  $\sum x^4$ .

Значний інтерес для дослідників представляє логістична залежність між змінними величинами. Графічно ця залежність позначається S-подібною кривою, яка описується рівнянням Ферхюльста:

$$y = \frac{N}{1 + 10^{a+bt}} + c, \quad (7.9)$$

де  $y$  – враховує ознаку;  $t$  – час, що пройшов від початкової або базисної ( $c$ ) величини ознаки, з якого розпочато його вимірювання, до граничної в даних умовах величини  $N$ , якої вона досягає за час  $t$ ;  $a$  і  $b$  – параметри рівняння, що визначають характер логістичної кривої.

Таблиця 7.1. Найбільш широко застосовувані рівняння регресії.

№	Загальний вигляд рівняння	Назва рівняння
1	$y = a + bx$	лінійна регресія
2	$y = a + bx + cz$	множинна лінійна регресія
3	$y = a + bx + cx^2$	рівняння параболи другого порядку
4	$y = a + bx + cx^2 + dx^3$	рівняння параболи третього порядку
5	$y = a + \frac{b}{x}$	рівняння гіперболи першого порядку
6	$y = a + \frac{b}{x^2}$	рівняння гіперболи другого порядку
7	$y = a + \frac{b}{x^3}$	рівняння гіперболи третього порядку
8	$y = a + bx + \frac{c}{x}$	рівняння гіперболи першого порядку з трьома невідомими $a$ , $b$ і $\tilde{n}$
9	$y = ae^{xb}$	рівняння експоненціального типу
10	$y = ab^x$	рівняння показникового типу
11	$y = ax^b$	рівняння степеневого типу
12	$y = \frac{N}{1 + 10^{a+bt}} + c$	рівнянням Ферхюльста

Важливою задачею регресійного аналізу є вибір рівняння, яке б найкращим чином описувало досліджувану закономірність. Для її вирішення доцільно застосовувати аналітичні способи перевірки правильності вибору кореляційних рівнянь.

Одним з цих способів є застосування дисперсійного аналізу. Неадекватність лінії регресії, що знайдена теоретичним способом по відношенню до емпіричної лінії, може бути описана сумою квадратів відхилень

$$D_R^2 = \sum (\bar{y}_{x_i} - \hat{y}_{x_i})^2. \quad (7.10)$$

Випадкова варіація спостережень  $y_{ij}$  по відношенню до умовних середніх  $\bar{y}_{x_i}$ , що знайдені емпірично, визначить девіантну:

$$D_e^2 = \sum_{i=1}^a \sum_{j=1}^{f_x} (y_{ij} - \bar{y}_{x_i})^2. \quad (7.11)$$

Тут  $\bar{y}_{x_i}$  – емпіричні умовні середні;  $\hat{y}_{x_i}$  – теоретичні умовні середні.

Відповідно до кожної девіанти може бути знайдена дисперсія. Так дисперсію неадекватності регресійної моделі знаходять за формулою:

$$s_R^2 = \frac{1}{a - m} D_R^2, \quad (7.12)$$

де  $a$  – число класових інтервалів, для яких знаходили емпіричні середні  $\bar{y}_{x_i}$ ;  $m$  – число параметрів, що визначаються в регресійній моделі. Ця кількість для прямолінійної регресії дорівнює двом, для квадратичної параболи – трьом і т.д.

Залишкову дисперсію можна розрахувати за формулою:

$$s_e^2 = \frac{1}{n - a} D_e^2. \quad (7.13)$$

Порівняння двох дисперсій здійснюється за  $F$ -критерієм Фішера ( $F_\phi = s_R^2 / s_e^2$ ) з числами ступенів свободи  $k_R = a - m$  і  $k_e = n - a$ , де  $n$  – об'єм вибірки. Якщо  $F_\phi$  перевищить табличне значення  $F_{st}$ , що знайдено для  $k_R$  і  $k_e$ , а також вибраного рівня значимості  $\alpha$ , то необхідно бути визнати, що припущення про адекватність отриманої теоретичної регресії треба відхилити і спробувати побудувати іншу модель. В протилежному випадку ( $F_\phi < F_{st}$ ) можна вважати, що адекватність рівняння регресії є доведеною, тому її можна використовувати в подальших розрахунках.

Для знаходження рівнянь регресії та невідомих параметрів рівнянь регресії будемо використовувати систему SPSS STATISTICS.

## Проста лінійна регресія

Для отримання рівняння регресії необхідно ввести початковий масив емпіричних даних в числовому форматі за допомогою редактора даних системи SPSS STATISTICS (рис. 7.1).

	x	y	пер	пер	пер	пер	пер	пер	пер	пер
1	1,70	-1,00								
2	1,90	-1,55								
3	1,91	-1,30								
4	1,98	-1,45								
5	2,00	-1,90								
6	2,08	-2,10								
7	2,14	-1,68								
8	2,22	-1,85								
9	2,30	-1,90								
10	2,38	-2,10								
11	2,46	-1,94								
12	2,54	-1,65								
13	2,62	-,80								
14	2,70	-1,40								
15	2,82	-,80								
16	2,86	-,50								
17	2,94	-,40								
18	3,02	,40								
19	3,10	,10								
20	3,18	,50								
21										
22										

Рис. 7.1. Вікно з введеними емпіричними даними для розв’язання задачі регресійного аналізу.

Після введення масиву початкових даних виконується перехід до головного меню системи шляхом послідовного вибору команд **Анализ**→**Регрессия**→**Линейная...** У діалоговому вікні «Линейная регрессия» (рис.7.2) в полі **Зависимые переменные:** вказується залежна змінна *y* (емпіричне значення функції), в полі **Независимые переменные:** – включаються незалежні змінні. В наведеному прикладі незалежною є одна змінна *x*.

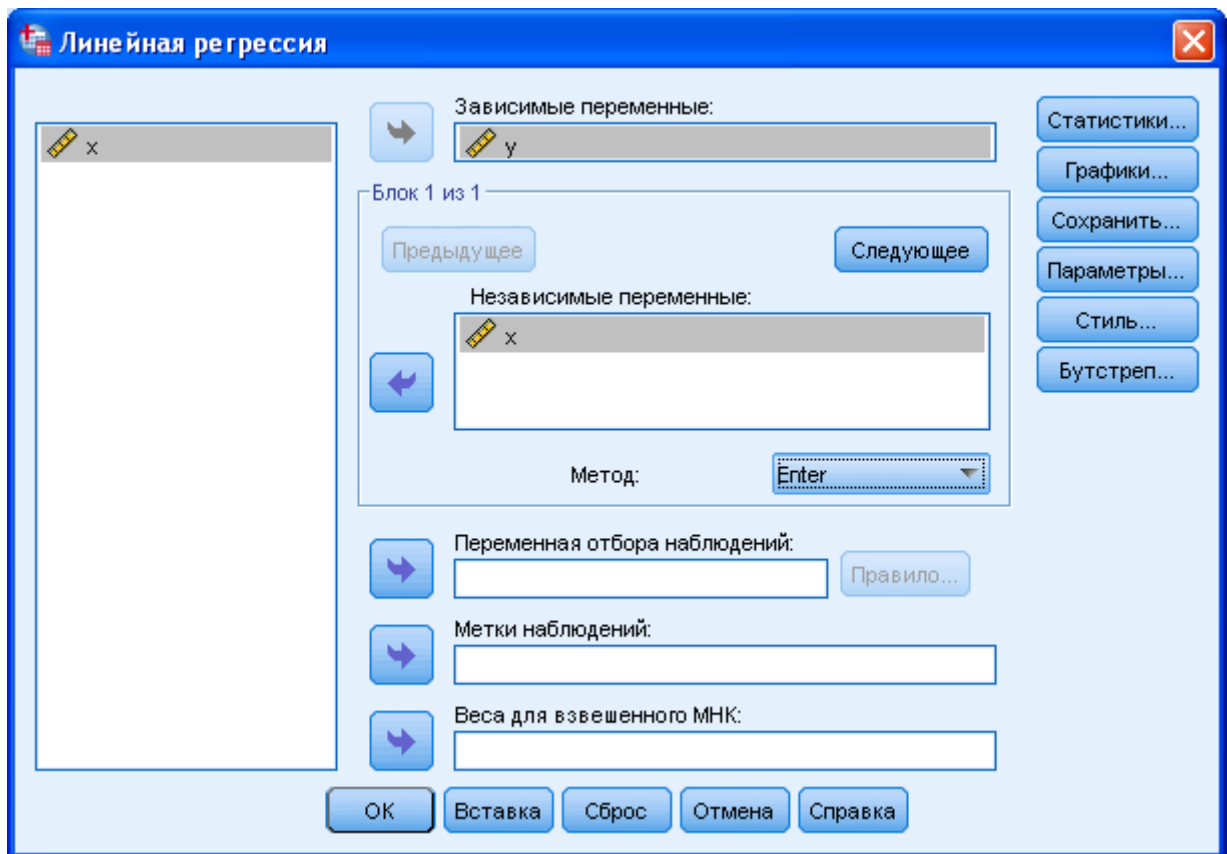


Рис. 7.2. Диалоговое окно «Линейная регрессия».

Натисканням клавiшi ОК генерується створення файлу виведення результатiв регресiйного аналізу, який наведено на рис.7.3.

### Регрессия

#### Введенные/удаленные переменные<sup>a</sup>

Модель	Введенные переменные	Удаленные переменные	Метод
1	x <sup>b</sup>	.	Enter

- a. Зависимая переменная: y
- b. Все требуемые переменные введены.

#### Сводка для модели

Модель	R	R-квадрат	Скорректированный R-квадрат	Стандартная ошибка оценки
1	,734 <sup>a</sup>	,538	,513	,57403

- a. Предикторы: (константа), x

#### ANOVA<sup>a</sup>

Модель		Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
1	Регрессия	6,914	1	6,914	20,981	,000 <sup>b</sup>
	Остаток	5,931	18	,330		
	Всего	12,845	19			

- a. Зависимая переменная: y
- b. Предикторы: (константа), x

Коэффициенты <sup>a</sup>						
Модель		Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
		B	Стандартная Ошибка	Бета		
1	(Константа)	-4,445	,727		-6,112	,000
	x	1,343	,293	,734	4,581	,000

- a. Зависимая переменная: y

Рис.7.3. Результаты одномерного линейного регрессионного анализа.

Відповідно до отриманих результатів (рис. 7.3) лінійна модель регресії матиме наступний вигляд:

$$y = 1.343x - 4.445.$$

Якість отриманої регресійної моделі, окрім критерію Фішера, можна оцінити через коефіцієнт R-квадрат.

Коефіцієнт R-квадрат характеризує долю дисперсії однієї змінної, що обумовлена дією іншої змінної. Величина R-квадрат, що використовується в розрахунках, на практиці виявляється дещо завищеною. Зкорельована величина R-квадрат менш формальна і ближча до реальних результатів. Так, для представленого прикладу зкорельований R-квадрат дорівнює 0,513. Це означає, що 51,3% дисперсії змінної y пояснюється впливом незалежної змінної x. Модель вважається якісною, якщо значення зкорельованого коефіцієнта R-квадрат перевищує 0,65.

### Нелінійна регресія

Для отримання рівняння нелінійної регресії треба послідовно вибрати команди **Анализ**→**Регрессия**→**Нелинейная**.... У діалоговому вікні «Нелинейная регрессия» в полі **Зависимые переменные**: треба вказати залежну змінну y (емпіричне значення функції), в поле **Выражение, задающее модель**: вказати вираз, що буде описувати передбачувану залежність між залежною змінною y та незалежною x (рис.7.4). У вибраному виразі  $a1 \exp(x) + a2$  використовується два параметри a1 та a2, початкові значення яких задаються після натискання кнопки **Параметры...** у діалоговому вікні «Нелинейная регрессия: Параметры» (рис.7.5).

Натисканням кнопок **Продолжить**→**ОК** генерується файл з результатами нелінійного регрессионного аналізу з заданим виразом залежності  $y = f(x)$ .

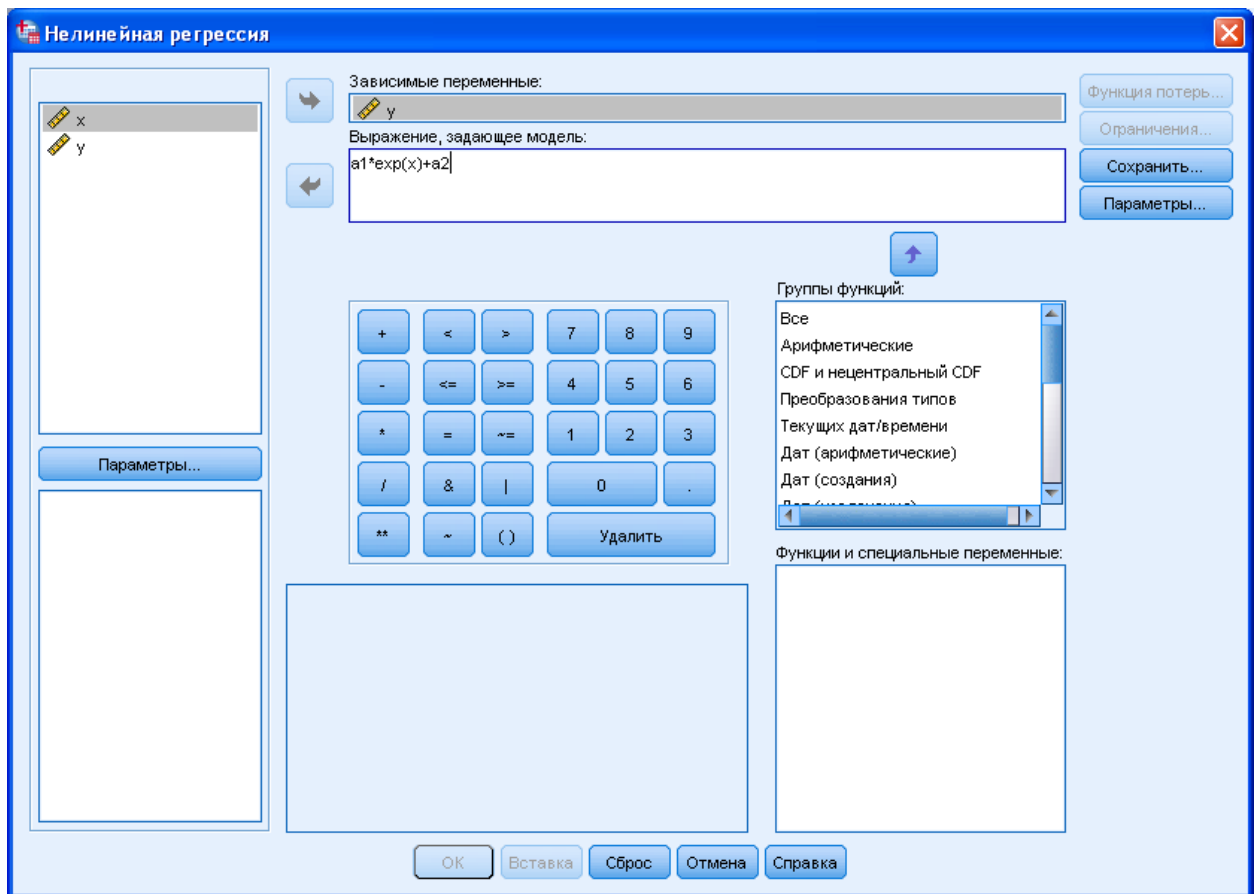


Рис. 7.4. Диалоговое окно «Нелинейная регрессия».

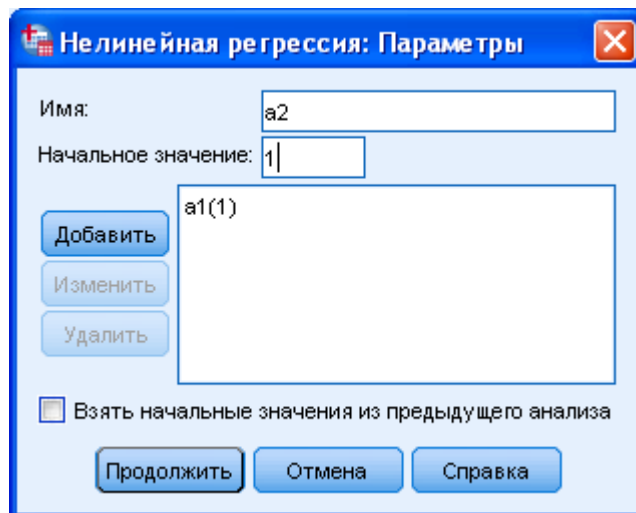


Рис. 7.5. Диалоговое окно «Нелинейная регрессия: Параметры».

Результаты нелинейного регрессионного анализа с указанным видом зависимости  $y = f(x)$  наведено на рис 7.6.

## Анализ нелинейной регрессии

**Хронология итераций<sup>b</sup>**

Номер итерации <sup>a</sup>	Сумма квадратов остатков	Параметр	
		a1	a2
1.0	174,947	1,000	1,000
1.1	4,168	-2,021	,017
2.0	4,168	-2,021	,017
2.1	4,168	-2,021	,017

Производные рассчитываются численно.<sup>b</sup>

a. Номер большой итерации выводится слева от десятичного разделителя, а номер малой итерации - справа от него.

b. Выполнение остановлено после 4 оценок модели и 2 оценок производных, поскольку относительное сдвиг между последовательными остаточными суммами квадратов составляет не более  $SSCON = 1,000E-8$ .

**Оценки параметров**

Параметр	Оценка	Стандартная Ошибка	95% доверительный интервал	
			Нижняя граница	Верхняя граница
a1	-2,021	,330	-2,714	-1,327
a2	,017	,221	-,448	,481

**Корреляции оценок  
параметров**

	a1	a2
a1	1,000	-,874
a2	-,874	1,000

**ANOVA<sup>a</sup>**

Источник	Сумма квадратов	ст. св.	Средние квадраты
Регрессия	35,868	2	17,934
Остаток	4,168	18	,232
Нескорректированный итог	40,036	20	
Скорректированный итог	12,845	19	

Зависимая переменная:  $y^a$

a.  $R$ -квадрат =  $1 - (\text{Остаточная сумма квадратов}) / (\text{Скорректированная сумма квадратов}) = ,676$ .

Рис. 7.6. Результаты нелінійного регресійного аналізу.

Таким чином, отримана нелінійна модель регресії матиме наступний вигляд:

$$y = -2,021e^x + 0.017 .$$

Для отриманого рівняння регресії зкорельований коефіцієнт R-квадрат дорівнює 0,676. Він перевищує значення аналогічного коефіцієнта для рівняння лінійної регресії. Тому можна зробити висновок що отримане рівняння нелінійної регресії краще описує початкові емпіричні дані.

### Підбір кривої

В процесі регресійного аналізу засобами системи SPSS STATISTICS є можливість вибрати вид рівняння регресії, що найкраще описує досліджуваний процес.

Для цього в випадаючому меню **Анализ**→**Регрессия** необхідно вибрати опцію **Подгонка кривых...**. В результаті цього отримуємо діалогове вікно «Подгонка кривых», що представлено на рис.7.7.

В даному вікні аналогічно до попередніх способів введення початкової інформації необхідно задати залежну змінну (**Зависимая**) та незалежну (**Переменная**), а також вибрати види моделей регресії, які необхідно створити.

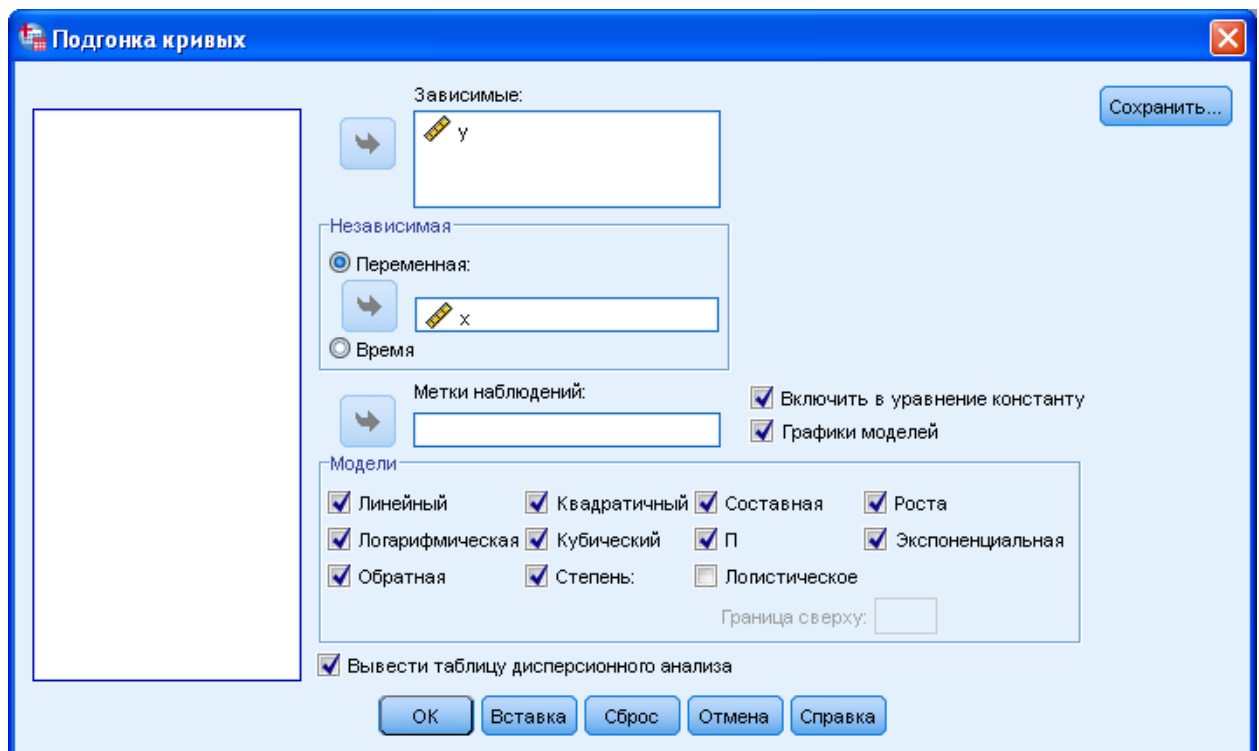


Рис. 7.7. Діалогове вікно «Подгонка кривых».

Для вибору найкращого виду рівняння регресії система пропонує формули, що представлено в таблиці 7.2.

Таблиця 7.2. Види рівнянь регресії, що пропонує для вибору система SPSS STATISTICS.

Назва	Вид рівняння
Лінійна	$y = a + bx$
Логарифмічна	$y = a + b \ln(x)$
Зворотна	$y = a + \frac{b}{x}$
Квадратична	$y = a + bx + cx^2$
Кубічна	$y = a + bx + cx^2 + dx^3$
Степенева	$y = ax^b$
Складова	$y = e^{(a+bx)}$
Показникова	$y = ab^x$
Логістична	$y = \frac{1}{\frac{1}{u} + ab^x}$
Зростання	$y = e^{(a+bx)}$
Експоненціальна	$y = ae^{(bt)}$

Після вибору виду моделей регресії натисканням ОК надається команда генерації вихідного файлу результатів регресійного аналізу. Інформація, що міститься в цьому файлі, представлено на рис 7.8.

### Подбор кривой

Описание модели		
Имя модели		MOD_3
Зависимая переменная	1	y
Уравнение	1	Линейная
	2	Логарифмическая
	3	Обратная
	4	Квадратичная
	5	Кубическая
	6	Составная <sup>a</sup>
	7	Степенная <sup>a</sup>
	8	П <sup>a</sup>
	9	Роста <sup>a</sup>
	10	Экспоненциальная <sup>a</sup>
Независимая переменная		x
Константа		Включено

Переменная, значения которой отмечают наблюдения на графиках	Не задано
Допустимые отклонения при вводе членов в уравнения	,0001

а) таблица вибору рівняння регресії  
**Линейная регрессия**

**Сводка для модели**

R	R-квадрат	Скорректированный R-квадрат	Среднеквадратичная ошибка оценки
,734	,538	,513	,574

Независимая переменная - это x.

**ANOVA**

	Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
Регрессия	6,914	1	6,914	20,981	,000
Остаток	5,931	18	,330		
Всего	12,845	19			

Независимая переменная - это x.

**Коэффициенты**

	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
	B	Стандартная Ошибка	Бета		
x	1,343	,293	,734	4,581	,000
(Константа)	-4,445	,727		-6,112	,000

б) коефіцієнти лінійної регресії

### Логарифмическая регрессия

Сводка для модели

R	R-квадрат	Скорректированный R-квадрат	Среднеквадратичная ошибка оценки
,682	,465	,435	,618

Независимая переменная - это x.

ANOVA

	Сумма квадратов	ст. св.	Средний квадрат	F	Знач.
Регрессия	5,974	1	5,974	15,653	,001
Остаток	6,870	18	,382		
Всего	12,845	19			

Независимая переменная - это x.

Кoeffициенты

	Нестандартизованные коэффициенты	Стандартизованные коэффициенты	t	Знач.

	В	Стандартная Ошибка	Бета		
ln(x)	3,006	,760	,682	3,956	,001
(Константа)	-3,800	,680		-5,589	,000

в) коефіцієнти логарифмічної регресії

### Обратная регрессия

Сводка для модели

R	R-квадрат	Скорректиро- ванный R-квадрат	Среднеквадрат ичная ошибка оценки
,624	,390	,356	,660

Независимая переменная - это x.

### ANOVA

	Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
Регрессия	5,009	1	5,009	11,508	,003
Остаток	7,835	18	,435		
Всего	12,845	19			

Независимая переменная - это x.

### Коэффициенты

	Нестандартизованные коэффициенты		Стандартизова нные коэффициенты	t	Знач.
	В	Стандартная Ошибка	Бета		
1 / x	-6,452	1,902	-,624	-3,392	,003

(Константа)	1,565	,818	1,912	,072
-------------	-------	------	-------	------

г) коефіцієнти оберненої регресії

### Квадратичная регрессия

Сводка для модели

R	R-квадрат	Скорректированный R-квадрат	Среднеквадратическая ошибка оценки
,951	,904	,893	,269

Независимая переменная - это x.

ANOVA

	Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
Регрессия	11,615	2	5,807	80,256	,000
Остаток	1,230	17	,072		
Всего	12,845	19			

Независимая переменная - это x.

Коэффициенты

Нестандартизованные коэффициенты	Стандартизованные коэффициенты	t	Знач.

	B	Стандартная Ошибка	Бета		
x	-12,940	1,777	-7,068	-7,280	,000
x ** 2	2,900	,360	7,825	8,060	,000
(Константа)	12,582	2,140		5,880	,000

д) коефіцієнти квадратичної регресії

### Кубическая регрессия

#### Сводка для модели

R	R-квадрат	Скорректиро- ванный R-квадрат	Среднеквадрат ичная ошибка оценки
,947	,896	,884	,280

Независимая переменная - это x.

#### ANOVA

	Сумма квадратов	ст.св.	Средний квадрат	F	Знач.
Регрессия	11,513	2	5,757	73,478	,000
Остаток	1,332	17	,078		
Всего	12,845	19			

Независимая переменная - это x.

#### Кoeffициенты

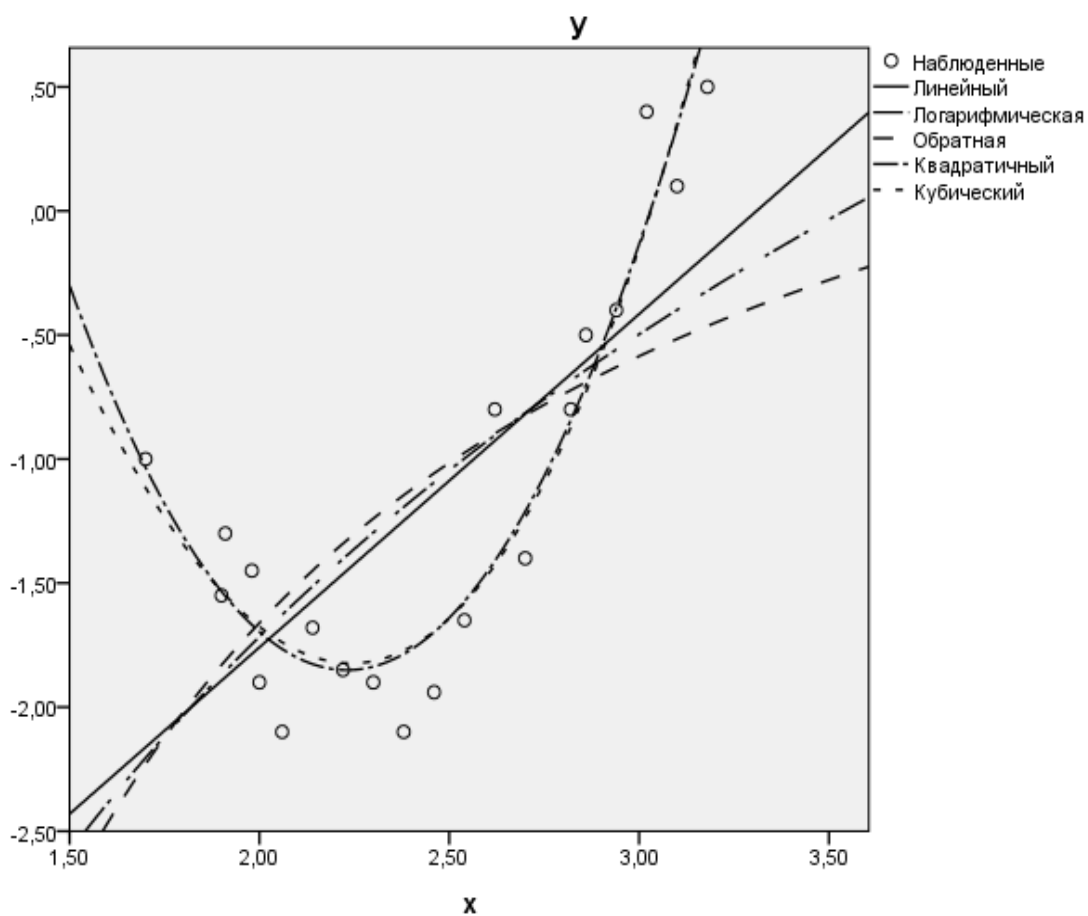
	Нестандартизованные коэффициенты		Стандартизованные коэффициенты	t	Знач.
	B	Стандартная Ошибка	Бета		
x	-5,876	,953	-3,210	-6,166	,000
x ** 3	,390	,051	3,988	7,662	,000
(Константа)	6,959	1,530		4,548	,000

#### Исключенные члены

	Бета-включения	t	Знач.	Частная корреляция	Минимальный допуск
x ** 2 <sup>a</sup>	30,244	1,812	,089	,413	,000

а. Достигнут предел допуска для введения переменных.

є) коефіцієнти кубічної регресії



к) графіки отриманих рівнянь регресивного аналізу

Рис. 7.8. Результати представлення рівнянь регресії, що описують досліджуваний процес.

За результатами регресійного аналізу з побудовою різних видів моделей встановлено, що рівняння регресії не отримано для складової, показникової, логістичної, зростання та експоненціальної, оскільки в початкових даних є від'ємні значення.

Таким чином отримано такі моделі регресії:

- лінійна  $y = 1.343x - 4.445$  ;
- логарифмічна  $y = 3,006 \cdot \ln(x) - 3.8$  ;
- обернена  $y = -6,452 \frac{1}{x} + 1.565$  ;
- квадратична  $y = 2.9x^2 - 12.94x + 12.582$  ;
- кубічна  $y = 0.39x^3 - 5.876x + 6.959$  .

При цьому найкраще описують початкові дані кубічна залежність (зкорельований R-квадрат дорівнює 0,884) та квадратична залежність (зкорельований R-квадрат дорівнює 0,893).

### Завдання для виконання комп'ютерного практикуму

1. Створити файл початкових даних шляхом використання інформації, що задана в індивідуальному завданні до комп'ютерного практикуму.

Для цього необхідно:

- виконати попередній аналіз інформації, що підлягає регресійному аналізу;
- вибрати пункт меню Редактора даних системи SPSS для введення початкової інформації;
- визначити імена, типи та формати даних для кожного із параметрів об'єктів дослідження;
- ввести початкові дані згідно з вказаним форматом;
- сформувати файл початкових даних для розв'язання задачі регресійного аналізу.

2. Розв'язати задачу лінійного регресійного аналізу з використанням сформованого масиву початкових даних. :

- в головному меню системи послідовно вибрати команди **Анализ**→**Регрессия**→**Линейная...**;
- в діалоговому вікні «**Линейная регрессия**» вказати залежні та незалежні змінні;
- запустити створення файлу з результатами лінійної регресії;
- виконати аналіз результатів лінійної регресії – моделі та коефіцієнта R-квадрат.

3. Розв'язати задачу нелінійного регресійного аналізу з використанням сформованого масиву початкових даних, для цього:

– в головному меню системи послідовно вибрати команди **Анализ→Регрессия→Нелинейная...**;

– в полі **Зависимые переменные**: вказати змінну  $y$  (емпіричне значення функції), в поле **Выражение, задающее модель**: вказати вираз, що буде описувати передбачувану залежність між залежною змінною  $y$  та незалежною  $x$ ;

– ввести початкові значення параметрів після натискання кнопки **Параметры...** у діалоговому вікні «Нелинейная регрессия: Параметры»;

– шляхом натискання **Продолжить→ОК** запустити створення файлу з результатами нелінійної регресії;

– виконати аналіз результатів не лінійної регресії – отриманої моделі та коефіцієнта  $R$ -квадрат.

2. Розв'язати задачу вибору рівняння регресії, що найкраще описує залежність вихідної величини від аргументу введених початкових даних, для цього:

– у випадяючому меню **Анализ→Регрессия** вибрати опцію **Подгонка кривых...**;

– в діалоговому вікні «Подгонка кривых» необхідно задати залежну змінну (**Зависимая**) та незалежну (**Переменная**), а також вибрати типи моделей регресії, які необхідно створити;

– шляхом натискання **ОК** запустити створення файлу з результатами регресії з отриманням множини вказаних рівнянь;

– виконати аналіз результатів множини рівнянь регресії – отриманих моделей та коефіцієнтів  $R$ -квадрат та зробити відповідні висновки.

### **Контрольні запитання**

1. Задачі регресійного аналізу даних.
2. Загальний вигляд лінійної моделі регресії.
3. Як визначаються коефіцієнти (параметри) рівняння регресії?.
4. Який метод використовується для визначення коефіцієнтів рівнянь регресії?
5. Які види рівнянь найчастіше використовуються як рівняння регресії?
6. Як виконується вибір рівняння регресії?
7. Яка послідовність реалізації простої лінійної регресії за допомогою SPSS Statistics?
8. Яка послідовність реалізації нелінійної регресії за допомогою SPSS Statistics?
9. Як задаються початкові значення параметрів нелінійної регресії в SPSS Statistics?
10. Як оцінюється якість опису початкових даних регресійною моделлю?
11. Що дозволяє отримати опція «Подгонка кривых» в SPSS Statistics?

## Комп'ютерний практикум №8

### НЕЙРОМЕРЕЖЕВЕ ДОСЛІДЖЕННЯ ДАНИХ

**Мета:** отримання практичних навиків розв'язання задачі нейромережевого дослідження біологічної інформації з використанням нейронної мережі, що реалізована в статистичному пакеті SPSS STATISTICS.

#### Теоретичні відомості

В останні десятиліття бурхливо розвивається нова прикладна область математики, що спеціалізується на штучних нейронних мережах (НМ). Актуальність досліджень в цьому напрямку підтверджується розмаїттям різних застосувань НМ. Це автоматизація процесів розпізнавання образів, адаптивне управління, апроксимація, прогнозування, створення експертних систем, організація асоціативної пам'яті і багато інших можливостей використання.

Широке коло задач, що вирішуються штучними нейронними мережами (ШНМ), не дозволяє в даний час створювати універсальні, могутні мережі, змушуючи використовувати спеціалізовані ШНМ, що функціонують за різними алгоритмами.

Моделі ШНМ можуть бути програмного і апаратного виконання. Надалі розглядатимемо в основному програмну реалізацію нейронних мереж .

Не зважаючи на істотні відмінності, окремі типи ШНМ мають декілька загальних рис.

По-перше, основу кожної НМ складають відносно прості, в більшості випадків – однотипні, елементи (осередки), що імітують роботу нейронів мозку. Далі під нейроном будемо розуміти штучний нейрон, тобто осередок ШНМ. Кожен нейрон характеризується своїм поточним станом за аналогією з нервовими клітинами головного мозку, які можуть бути збуджені або загальмовані. Він має групу синапсів – однонаправлених вхідних зв'язків, сполучених з виходами інших нейронів, а також має аксон – вихідний зв'язок даного нейрона, з яким сигнал (збудження або гальмування) поступає на синапси наступних нейронів. Кожен синапс характеризується величиною синаптичного зв'язку або її вагою  $w_i$ , яка за фізичним сенсом еквівалентна електричній провідності.

Поточний стан нейрона визначається як зважена сума його входів:

$$s = \sum_{i=1}^n x_i \cdot w_i . \quad (8.1)$$

Вихід нейрона є функцією його стану:

$$y = f(s). \quad (8.2)$$

Нелінійна функція  $f$  називається активаційною і може мати різний вигляд. Однією з найбільш розповсюджених є нелінійна функція з насиченням, так звана логістична функція або сигмоїд:

$$f(x) = \frac{1}{1 + e^{-ax}} \quad (8.3)$$

Повертаючись до загальних рис, властивих всім ШНМ, відзначимо, по-друге, принцип паралельної обробки сигналів, який досягається шляхом об'єднання великого числа нейронів в так звані шари і з'єднання певним чином нейронів різних шарів, а також, в деяких конфігураціях, і нейронів одного шару між собою, причому обробка взаємодії всіх нейронів ведеться пошарово.

Теоретично кількість шарів і кількість нейронів в кожному шарі може бути довільною, проте фактично вона обмежена ресурсами комп'ютера або спеціалізованої мікросхеми, на яких зазвичай реалізується НМ. Чим складніша НМ, тим масштабніші завдання, що підвладні їй.

Вибір структури НМ здійснюється відповідно до особливостей і складності завдання. Для вирішення деяких окремих типів завдань вже існують оптимальні, на сьогоднішній день, конфігурації. Якщо ж завдання не може бути зведене ні до одного з відомих типів, розробникові доводиться вирішувати складну проблему синтезу нової конфігурації. При цьому він керується декількома основоположними принципами: можливості мережі зростають із збільшенням кількості осередків мережі, щільністю зв'язків між ними і кількістю виділених шарів; введення зворотних зв'язків разом із збільшенням можливостей мережі піднімає питання про динамічну стійкість мережі; складність алгоритмів функціонування мережі (зокрема, наприклад, введення декількох типів синапсів – збудливих, гальмівних тощо) також сприяє посиленню потужності НМ. Питання про необхідні і достатні властивості мережі для вирішення того або іншого роду завдань є цілим напрямком нейрокомп'ютерної науки. Оскільки проблема синтезу НМ дуже залежить від розв'язуваної задачі, дати загальні докладні рекомендації важко. В більшості випадків оптимальний варіант виходить на основі інтуїтивного підбору.

Вочевидь, що процес функціонування НМ, тобто суть дій, які вона здатна виконувати, залежить від величин синапсичних зв'язків, тому, задавшись певною структурою НМ, що відповідає якому-небудь завданню, розробник мережі повинен знайти оптимальні значення всіх змінних вагових коефіцієнтів (деякі синапсичні зв'язки можуть бути постійними).

Цей етап є навчанням НМ, і від того, наскільки якісно він буде виконаний, залежить здатність мережі вирішувати поставлені перед нею проблеми під час експлуатації. На етапі навчання окрім параметра якості підбору вагів важливу роль відіграє час навчання. Як правило, ці два параметри зв'язані зворотною залежністю і їх доводиться вибирати на основі компромісу.

Навчання НМ може проводитись з вчителем або без нього. У першому випадку мережі пред'являються значення як вхідних, так і бажаних вихідних сигналів, і вона за деяким внутрішнім алгоритмом підстроює ваги своїх синапсичних зв'язків. У другому випадку виходи НМ формуються самостійно, а

ваги змінюються за алгоритмом, що враховує тільки вхідні і похідні від них сигнали.

Існує велика кількість різних алгоритмів навчання, які поділяються на два великі класи: детерміністські і стохастичні. У першому з них підстроювання вагів є жорсткою послідовністю дій, в другому – вона проводиться на основі дій, що підкоряються деякому випадковому процесу.

При класифікації НМ важливо відзначити існування бінарних і аналогових мереж. Перші з них оперують з двійковими сигналами і вихід кожного нейрона може приймати тільки два значення: логічний нуль („загальмований” стан) і логічна одиниця („збуджений” стан). В аналогових мережах вихідні значення нейронів здатні приймати безперервні значення, що могло б мати місце після заміни в активаційній функції нейронів перцептрона на сигмоїд.

Ще одна класифікація ділить НМ на синхронні і асинхронні. У першому випадку в кожен момент часу свій стан міняє лише один нейрон. У другому – стан міняється відразу у цілої групи нейронів, як правило, у всього шару. Алгоритмічно хід часу в НМ задається ітераційним виконанням однотипних дій над нейронами. Надалі будемо розглядати тільки синхронні НМ.

Мережі також можна класифікувати за кількістю шарів. Тут доречно відзначити важливу роль нелінійності активаційної функції, оскільки, якби вона не мала даної властивості або не входила в алгоритм роботи кожного нейрона, результат функціонування будь-якої  $p$ -шарової НМ з ваговими матрицями  $W^{(i)}$ ,  $i=1,2,\dots,p$  для кожного шару  $i$  зводився б до перемножування вхідного вектора сигналів  $X$  на матрицю

$$W^{(\Sigma)} = W^{(1)} \cdot W^{(2)} \cdot \dots \cdot W^{(p)} \quad (8.4)$$

тобто фактично така  $p$ -шарова НМ еквівалентна одношаровій НМ з ваговою матрицею єдиного шару  $W^{(\Sigma)}$ :

$$Y = XW^{(\Sigma)} \quad (8.5)$$

Треба відзначити, що нелінійність іноді вводиться і в синапсичні зв'язки. Більшість відомих на сьогоднішній день НМ використовують для знаходження зваженої суми входів нейрона згідно формули (8.1), проте в деяких випадках НМ корисно ввести інший запис, наприклад:

$$s = \sum_{i=1}^n x_i^2 \cdot w_i \quad (8.6)$$

або навіть

$$s = \sum_{i=1}^n x_i \cdot x_{((i+1) \bmod n)} \cdot w_i \quad (8.7)$$

Введення такого роду нелінійності збільшує обчислювальну потужність мережі, тобто дозволяє з меншою кількістю нейронів з „нелінійними” синапсами сконструювати НМ, що виконує роботу звичайної НМ з великою кількістю стандартних нейронів і складнішої конфігурації.

Робота всіх нейронних мереж зводиться до класифікації (узагальненню) вхідних сигналів, що належать  $n$ -вимірному гіперпростору, по деякій кількості класів. З математичної точки зору це відбувається шляхом розбиття гіперпростору гіперплощинами (запис для випадку одношарового перцептрона)

$$\sum_{i=1}^n x_i \cdot w_{ik} = T_k, k=1, \dots, m \quad (8.9)$$

Кожна отримана область є областю визначення окремого класу. Кількість таких класів для однієї НМ перцептронного типу не перевищує  $2m$ , де  $m$  – кількість виходів мережі. Проте не всі з них можуть бути розділені даною НМ.

Функції, які не реалізуються одношаровою мережею, називаються лінійно нероздільними. Рішення задач, що підпадають під це обмеження, полягає в застосуванні двох і більше шарових мереж або мереж з нелінійними синапсами, проте і тоді існує вірогідність, що коректне розділення деяких вхідних сигналів на класи неможливе.

Алгоритм навчання НМ з вчителем включає наступні етапи.

1. Виконати ініціалізацію елементів вагової матриці (зазвичай невеликими випадковими значеннями).
2. Подати на входи один з вхідних векторів, які мережа повинна навчитися розрізняти, і обчислити її вихід.
3. Якщо вихід правильний, перейти на крок 4. Інакше обчислити різницю між ідеальним і отриманим значеннями виходу:

$$\delta = Y_l - Y$$

та модифікувати ваги відповідно до формули:

$$w_{ij}(t+1) = w_{ij}(t) + \nu \cdot \delta \cdot x_i,$$

де  $t$  і  $t+1$  – номери відповідно поточної і наступної ітерації;

$\nu$  – коефіцієнт швидкості навчання,  $0 < \nu < 1$ ;

$i$  – номер входу;

$j$  – номер нейрона в шарі.

Вочевидь, якщо  $Y_l > Y$  вагові коефіцієнти будуть збільшені то тим самим вони зменшать похибку. Інакше вони будуть зменшені, і  $Y$  теж зменшиться, наближаючись до  $Y_l$ .

4. Повторити цикл, розпочинаючи з кроку 2, поки мережа не перестане помилятися.

На другому кроці на різних ітераціях по черзі у випадковому порядку пред'являються всі можливі вхідні вектори. На жаль, не можна наперед

визначити кількості ітерацій, які потрібно буде виконати, а в деяких випадках і гарантувати повний успіх.

Найбільш доцільним є використання штучних нейронних мереж при розв'язанні біологічних задач з метою прогнозування та моделювання параметрів, класифікації, групування та розпізнавання образів тощо.

Аналіз пакетів програм, що реалізують штучні нейронні мережі показав, що значну гнучкість та функціональність має пакет MatLab. Завдяки достатньо широкому набору архітектур штучних нейронних мереж та видів їх навчання він дозволяє як будувати штучні нейронні мережі з довільною архітектурою, так і використовувати широкий спектр цих архітектур, що представлені функціями програмного пакету, та встановлювати метод їх навчання, які дають можливість вирішувати широке коло задач.

Наряду з цим, серед стандартних програмних систем створення, навчання та використання наявних нейронних мереж для прогнозування, моделювання, класифікації, групування та розпізнавання образів треба виділити систему NeuroSolutions, що має широкий набір різноманітних структур штучних нейронних мереж, надає рекомендації з їх використання, має зручний інтерфейс користувача та дозволяє отримати текст програми на одній із мов програмування, що реалізує налаштовану штучну нейронну мережу.

Ефективні засоби дослідження біологічної інформації реалізовані в статистичному пакеті SPSS STATISTICS. Тому при дослідженні біологічної інформації при виконанні комп'ютерного практикуму використовується вказаний багатофункціональний статистичний пакет.

## **Підготовка до роботи**

Вивчити базові поняття нейротехнологій, ознайомитись з можливостями сучасних нейропакетів.

Отримати у викладача індивідуальне завдання в вигляді таблиці даних для виконання нейромережових досліджень.

Підготувати отриманий інформаційний масив даних до введення в ЕОМ засобами пакету SPSS.

## **Послідовність виконання нейромережевого дослідження засобами SPSS STATISTICS**

Запустити пакет SPSS на виконання. В вікні системи відкрити раніше створений файл, що містить дані індивідуального завдання. Якщо цей файл ще не створено, то вибором відповідного пункту меню ввести початкові дані.

В вікні введення початкових даних вказати ім'я кожної змінної, її тип, ширину тощо. Після цього ввести дані таблиці індивідуального завдання та записати введені в файл з вказаним іменем.

Виконати налаштування пакету для нейромережевого дослідження. Для цього необхідно запустити послідовність: **Анализ** → **Нейронные сети** → **Многослойный перцептрон**.

В вікні налаштування перцептрона (рис.8.1), що відкриється, треба виконати 8 налаштувань:

- Переменные
- Разделение
- Архитектура
- Обучение
- Вывод
- Сохранение
- Экспорт
- Параметры.

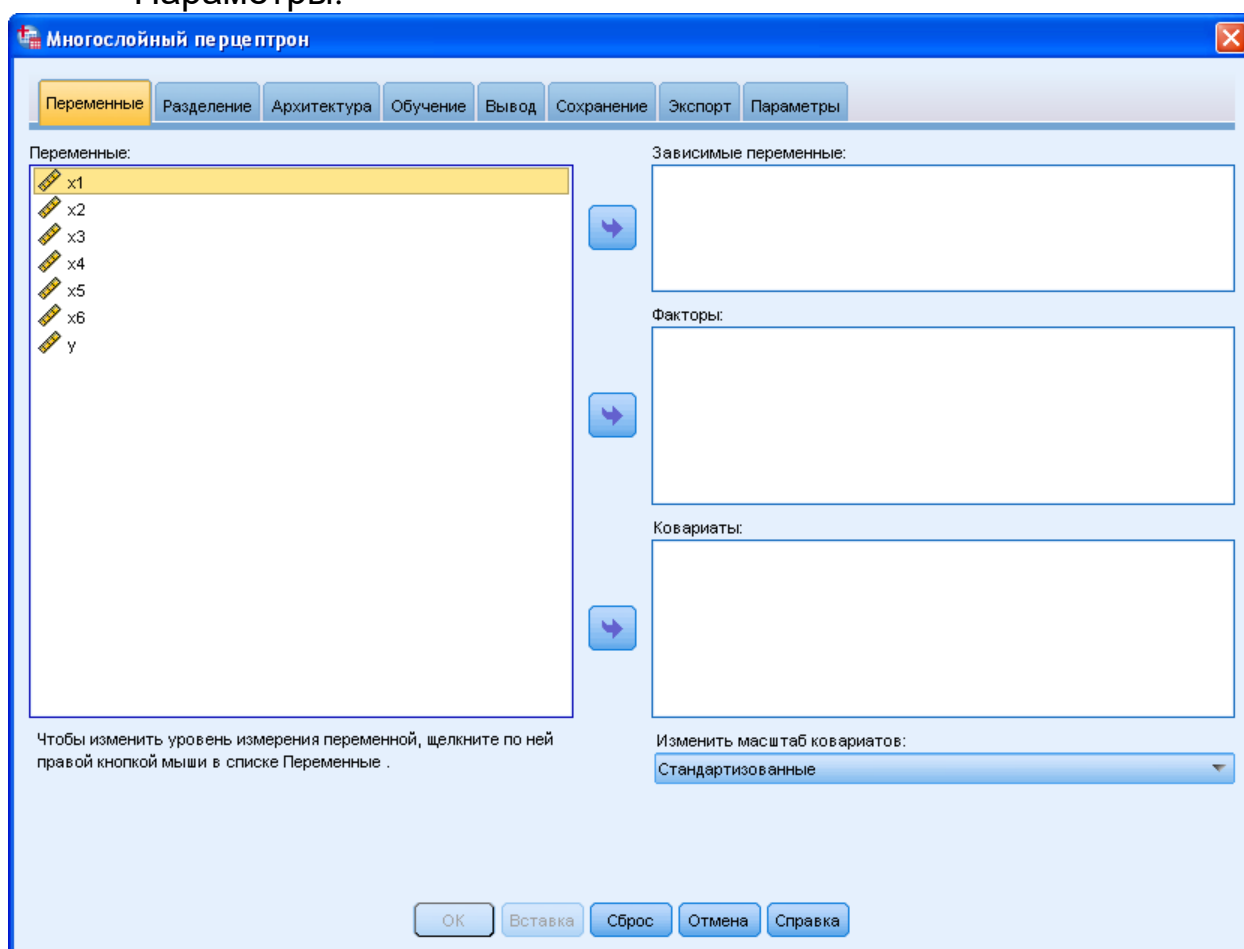


Рис.8.1. Вікно налаштування перцептрона.

**Налаштування змінних** полягає в їх розподілі на залежні змінні, фактори, що є вхідними змінними, та коваріати, а також вказання методу роботи з масштабом коваріатів.

Змінні, значення яких визначається за допомогою вимірювань в ході експерименту є залежними змінними.

Змінні, якими можна керувати при проведенні експерименту є факторами або незалежними змінними.

Приклад розподілу змінних на залежні та фактори наведено на рис. 8.2.

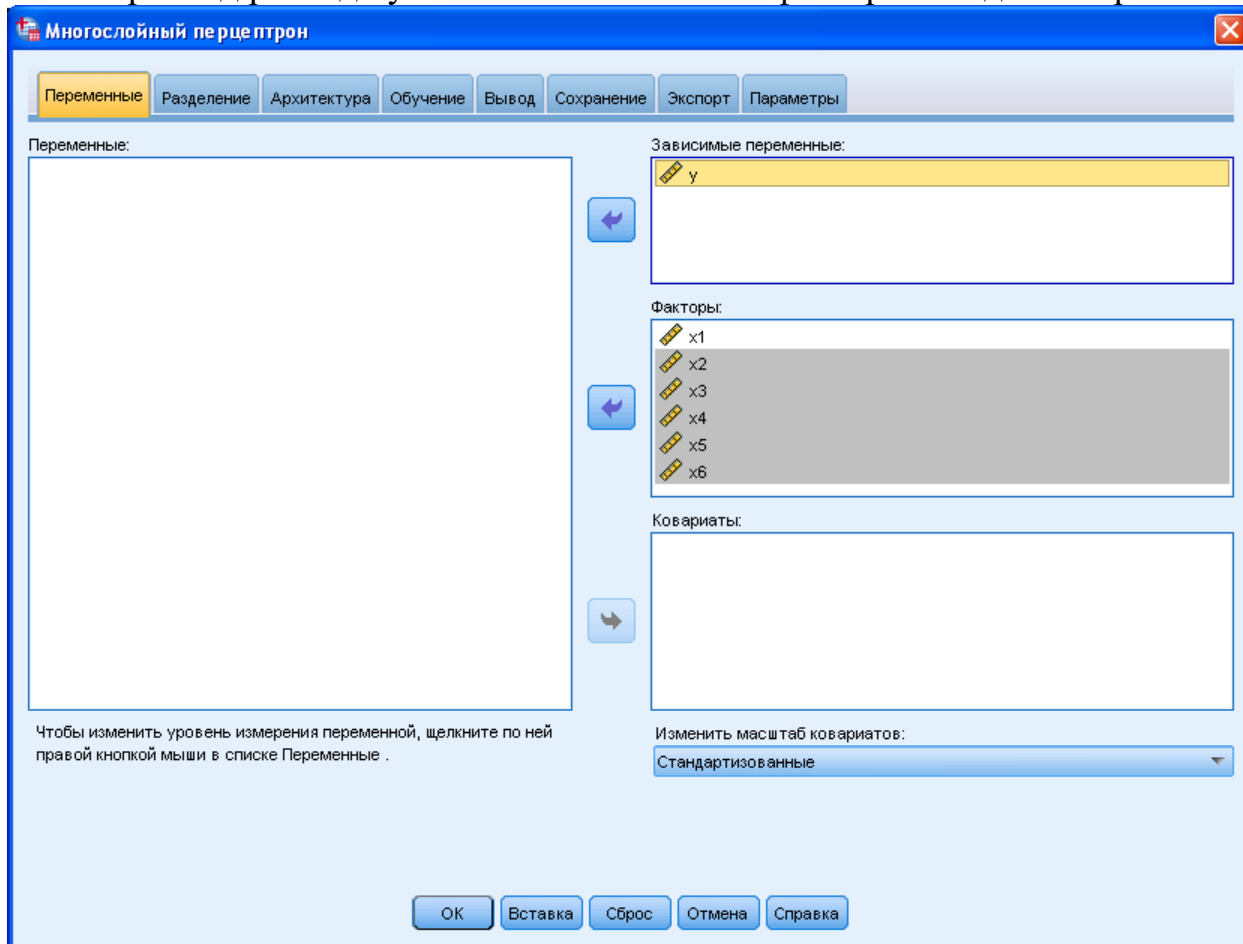


Рис. 8.2. Вікно з результатами розподілу змінних на залежні та незалежні.

**Розподіл змінних.** Всі наявні спостереження поділяються на підмножини: навчальну, контрольну і перевірочну послідовності. Цей розподіл можна виконати або шляхом визначення відносної кількості кожної із них, при цьому буде здійснено випадковим чином розподіл даних на три частини в відповідності з встановленою кількістю процентного складу, або за допомогою спеціальної змінної розподілу.

Приклад розподілу початкових даних на підмножини за процентним складом наведено на рис. 8.3.

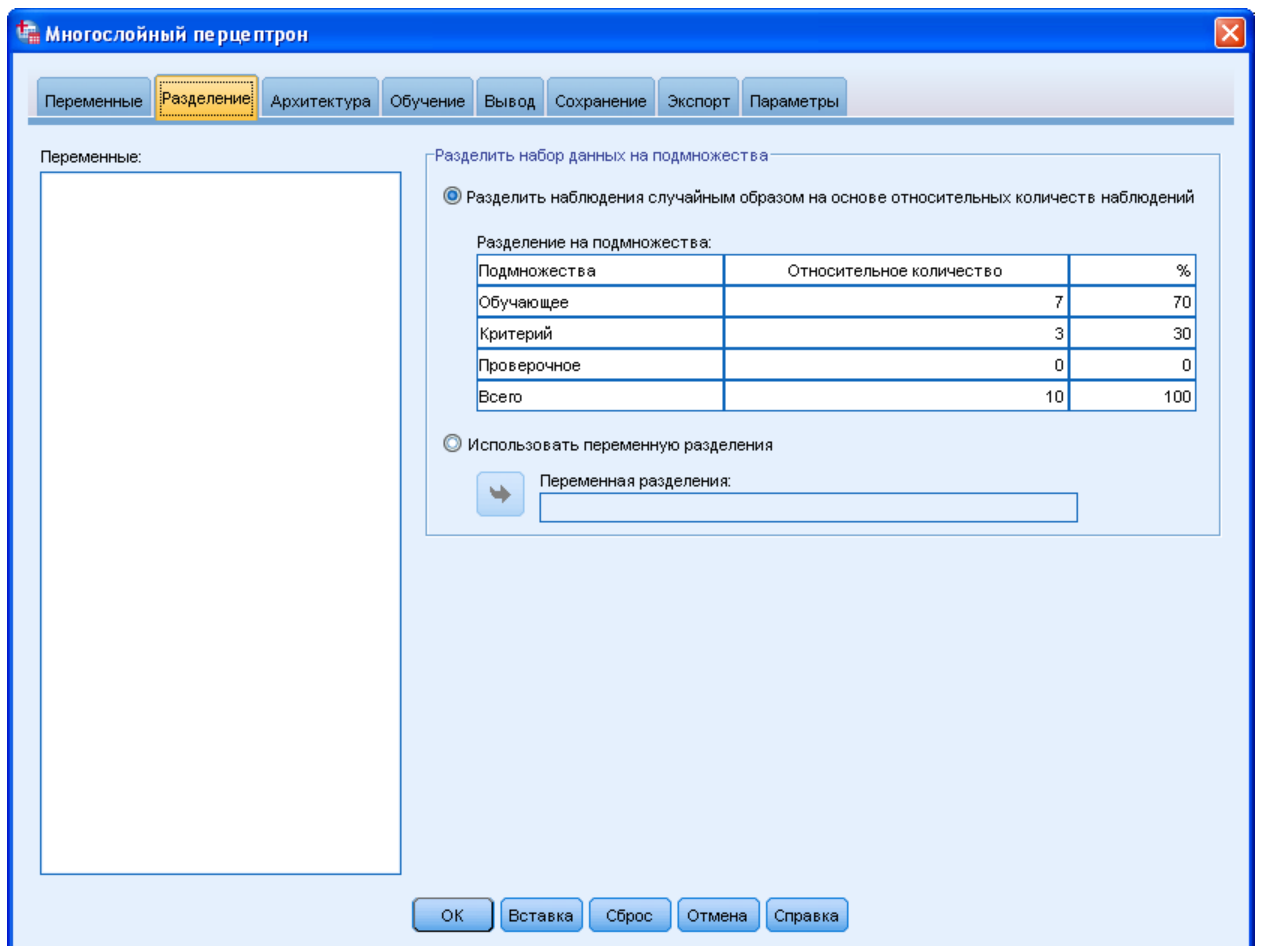


Рис.8.3. Вікно поділу початкової інформації на підмножини за процентним складом.

**Архітектура мережі.** Архітектура нейронної мережі може бути сформована автоматично з визначенням тільки двох параметрів – мінімальної та максимальної кількості нейронів в схованому шарі або шляхом ручного налаштування.

В випадку ручного налаштування визначається кількість схованих шарів, кількість нейронів в кожному схованому шарі (цей параметр може визначатись автоматично) та функції активації: гіперболічний тангенс або сигмоїд.

При визначенні вихідного шару нейронів функцій активації можна використовувати також тотожність, Sogtmax, гіперболічний тангенс або сигмоїд. При цьому вибрана для вихідного шару функція активації визначає доступні методи зміни масштабу кількісних залежних змінних: масштаби стандартизуються, аналізуються (при нормалізації може вводиться поправка), корегуються і нормалізуються (з можливою поправкою) або не змінюються.

**Навчання мережі.** В цьому режимі налаштування мережі виконують вибір типу навчання, алгоритму оптимізації та встановлення параметрів навчання.

В пакеті SPSS доступні три види навчання: пакетне, інтерактивне та мініпакетне, яке можна розраховувати автоматично або налаштувати, визначивши кількість записів.

В якості алгоритмів оптимізації можна вибрати метод масштабованих спряжених градієнтів або метод градієнтного спуску.

Регулюванню піддаються такі параметри навчання: початкове значення лямбди і сігми, центр інтервалу або його зміщення.

**Налаштування виведення інформації.** При виведенні інформації визначають чи потрібний опис, діаграма та виведення ваг нейронів.

Окрім структури мережі визначають:

- чим буде характеризуватись продуктивність мережі;
- чи потрібна узагальнення інформація про модель;
- чи є необхідність в результатах;
- чи треба виводити криву ROC, Кумулятивну діаграму виграшу, діаграми приросту та точності прогнозу;
- чи виводиться діаграма залишків та прогнозів.

**Налаштування збереження результатів.** При виконанні досліджень можна зберегти передбачене значення або категорій кожної залежної змінної, передбачену ймовірність для кожної залежної змінної, вказавши ім'я залежної змінної та імена зберігаємих змінних. Є можливість автоматично створювати унікальні імена незалежних змінних або вказувати налаштовуванні імена.

**Експорт результатів.** Експортувати можна оцінки ваг нейронів в XML-файли та імена залежних змінних у відповідні файли.

**Налаштування додаткових параметрів.** До додаткових параметрів відносяться: особливості обробки пропущених значень користувача, правила зупинки: при збереженні максимальної кількості кроків без зменшення похибки, максимального часу навчання при досягненні максимальної кількості епох, при досягненні мінімальної абсолютної або відносної похибки навчання, при перевищенні максимально допустимої кількості спостережень, що зберігаються в пам'яті.

Після встановлення всіх необхідних параметрів штучної нейронної мережі натисканням клавіші ENTER запустити нейромережеве дослідження.

### **Завдання для виконання комп'ютерного практикуму**

1. Вивчити базові поняття нейротехнологій, ознайомитись з можливостями сучасних нейропакетів.
2. Отримати у викладача індивідуальне завдання в вигляді таблиці даних для виконання нейромережевих досліджень.

3. Підготувати отриманий інформаційний масив даних до введення в ЕОМ засобами пакету SPSS.

3. Створити файл початкових даних шляхом використання інформації, що задана в індивідуальному завданні до комп'ютерного практикуму. Для цього необхідно:

– виконати попередній аналіз інформації, що підлягає регресійному аналізу;

– вибрати пункт меню Редактора даних системи SPSS для введення початкової інформації;

– визначити імена, типи та формати даних для кожного із параметрів об'єктів дослідження;

– ввести початкові дані згідно з вказаним форматом;

– сформувати файл початкових даних для розв'язання задачі нейромережевого дослідження.

4. Виконати налаштування процесу дослідження засобами нейронних мереж послідовністю команд **Анализ** → **Нейронные сети** → **Многослойный перцептрон**.

5. Реалізувати поділ змінних на залежні та незалежні змінні.

6. Виконати поділ початкових даних на навчальну, контрольну і перевірочну послідовності.

7. Вибрати архітектуру штучної нейронної мережі.

8. Вибрати тип, алгоритм та параметри навчання штучної нейронної мережі.

9. Виконати налаштування виведення результатів нейромережевого дослідження інформації та їх збереження.

10. Забезпечити налаштування додаткових параметрів дослідження та експорт їх результатів.

11. Натисканням клавіші ENTER запустити нейромережеве дослідження.

12. Результати дослідження оформити в вигляді звіту з комп'ютерного практикуму.

### **Контрольні запитання**

1. Задачі, що вирішуються за допомогою штучних нейронних мереж.
2. Методи реалізації моделей штучних нейронних мереж.
3. Яка структура штучної нейронної мережі?
4. В чому полягає подібність нейронів мозку і штучної мережі?
5. Що таке синапс?
6. Чим визначається поточний стан нейрону?
7. Що таке сигмоїд?
8. Яким чином утворюється складна ШНМ?
9. Яке призначення навчання ШНМ?
10. Методи навчання ШНМ.

11. Алгоритм навчання ШНМ.
12. Які пакети прикладних програм реалізують дослідження методами ШНМ?
13. Яка послідовність налаштування нейромережевого дослідження засобами пакету SPSS?
14. Які можливості налаштування вікна перцептрона?
15. В чому полягає налаштування змінних?
16. На які послідовності поділяється масив початкових даних?
17. Які методи вибору архітектури ШНМ?
18. В чому полягає ручний вибір архітектури ШНМ?
19. Які типи навчання ШНМ реалізовані в SPSS?
20. В чому полягає налаштування виведення результатів нейромережевого дослідження?
21. В чому полягає налаштування збереження та експорту результатів нейромережевого дослідження?
22. В чому полягає налаштування додаткових параметрів нейромережевого дослідження?

## СПИСОК ЛІТЕРАТУРИ

1. Лакин Г.Ф. Биометрия: Учеб. пособие для биол. спец. вузов – 4-е изд., перераб. и доп. – М.: Высш. шк., 1990. – 352 с.: ил.
2. Плохинский Н.А. Биометрия. – М.: Изд-во МГУ, 1970. – 366 с.
3. Плохинский Н.А. Алгоритмы биометрии. – М.: МГУ, 1980. – 150 с.
4. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей. – М.: Финансы и статистика, 1985. – 487 с.
1. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях. – М.: ЮНИТИ – ДАНА, 2001. – 270 с.
2. Прикладная статистика: Классификация и снижение размерности. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 607 с.
3. Факторный, дискриминантный и кластерный анализ: Пер. с англ. / Дж.О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.; Под. ред. И.С. Енюкова. – М.: Финансы и статистика, 1989. – 215 с.
4. Статистические методы для ЭВМ. / Под ред. К. Энслейна, Э. Релстона, Г.С. Уилфа. – М.: Наука, 1986. – 464 с.
5. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистические методы в медико-биологических исследованиях с использованием Excel. – 2-е изд., перераб. и доп. – К.: МОРИОН, 2001. – 408 с.
10. Дьюк В., Эмануэль В. Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
11. Барсегян А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.: ил.
12. Мінцер О.П. та ін. Інформаційні технології в охороні здоров'я і практичній медицині: У 10 кн. Кн. 5. Оброблення клінічних і експериментальних даних в медицині: Навч. посіб. / О.П. Мінцер, Ю.В. Вороненко, В.В. Власов. – К.: Вищ. шк., 2003. – 350 с.
13. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
14. Петрович М.Л., Давыдович М.И. Статистическое оценивание и проверка гипотез на ЭВМ. – М.: Финансы и статистика, 1989. – 191 с.
15. Каримов Р.Н. Обработка экспериментальной информации. Многомерный анализ. Саратов. техн. ун-т, – Саратов, 2000. – 104 с.
16. Благуш П. Факторный анализ с обобщениями. – М.: Финансы и статистика, 1989. – 248 с.
17. Закс Л. Статистическое оценивание. Пер. с нем. В.Н. Варыгина. Под ред. Ю.П. Адлера, В.Г. Горского. – М.: «Статистика», 1976. – 598 с.

18. Вислоух С.П. Інформаційні технології в задачах технологічної підготовки приладо- та машинобудівного виробництва: моногр. / С.П. Вислоух. – К.: НТУУ «КПІ», 2011. – 488 с.
19. Дюран Б., Оделл П. Кластерный анализ. Пер. с англ. Е.З. Демиденко. Научн. ред. и предисл. А.Я. Боярского. – М.: Статистика, 1977. – 128 с.
20. Айвазян С. А. Методы анализа данных. – М., 1985. – 357 с.
21. Жамбью М. Иерархический кластер-анализ и соответствия: Пер. с фр. – М.: Финансы и статистика, 1988. – 342 с.
22. Наследов А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. – СПб.: Питер, 2013. – 416 с.: ил.
23. Antonyuk V.S. Information technology in deciding of technological problems in instrument making and machine engineering./ V.S. Antonyuk , S.P. Vysloukh // Оптимізація виробничих процесів і технічний контроль у машинобудуванні та приладобудуванні: [збірник наукових праць] / відповідальний редактор З.А. Стоцько. – Львів : Видавництво Львівської політехніки. – 2013. – №760. – С.101-105.
24. Барандич К.С. Методика визначення циклічної довговічності матеріалу деталей, що працюють в умовах змінних навантажень./ К.С. Барандич, С.П. Вислоух // Вісник ЖДТУ. – 2015. – №4 (75). – С. 30–37.
25. Вислоух С.П. Факторный анализ технологической информации. // Вестник Харьковского государственного политехнического университета. Выпуск 100. Машиностроение. – Харьков, 2000. – С. 26-29.
26. Вислоух С.П. Применение методов дискриминантного анализа при технологическом проектировании. // Резание и инструмент в технологических системах. Межвед. Научн.-техн. сборник. – Харьков: ХГПУ, 2001. – Вып. 60. – С. 26-35.
27. Вислоух С.П. Применение методов кластерного анализа при проектировании технологических процессов. // Вестник Сев. ГТУ. Выпуск 36: Автоматизация процессов и управление. Сб. Научн. Трудов Севастоп. Нац. Техн. ун-т. – Севастополь, 2002. – С. 103-108.
28. Вислоух С.П. Нейросетевое моделирование технологических параметров процесса резания. // Резание и инструмент в технологических системах: Межд. научн. – техн. сборник. – Харьков. НТУ «ХПИ», 2005, Вып. 68. – С. 109–116.

## ДОДАТОК. ФРАГМЕНТИ СТАТИСТИЧНИХ ТАБЛИЦЬ

**Таблиця Д1. Значення t-критерію Стьюдента при рівні значимості  $\alpha$ .**

Число ступенів свободи $\nu$	Рівень значимості $\alpha$ (двохстороння критична область)					
	0,1	0,05	0,02	0,01	0,002	0,001
1	6.314	12.706	31.821	63.657	318.309	636.619
2	2.920	4.303	6.965	9.925	22.327	31.599
3	2.353	3.182	4.541	5.841	10.215	12.924
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.869
6	1.943	2.447	3.143	3.707	5.208	5.959
7	1.895	2.365	2.998	3.499	4.785	5.408
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.25	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.552	2.878	3.610	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.768
24	1.711	2.064	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.75	3.385	3.646
31	1.696	2.040	2.453	2.744	3.375	3.633
32	1.694	2.037	2.449	2.738	3.365	3.622
33	1.692	2.035	2.445	2.733	3.356	3.611
34	1.691	2.032	2.441	2.728	3.348	3.601
35	1.690	2.030	2.438	2.724	3.340	3.591
36	1.688	2.028	2.434	2.719	3.333	3.582
37	1.687	2.026	2.431	2.715	3.326	3.574
38	1.686	2.024	2.429	2.712	3.319	3.566
39	1.685	2.023	2.426	2.708	3.313	3.558
40	1.684	2.021	2.423	2.704	3.307	3.551
41	1.683	2.020	2.421	2.701	3.301	3.544
42	1.682	2.018	2.418	2.698	3.296	3.538
43	1.681	2.017	2.416	2.695	3.291	3.532
44	1.680	2.015	2.414	2.692	3.286	3.526
45	1.679	2.014	2.412	2.69	3.281	3.520
46	1.679	2.013	2.410	2.687	3.277	3.515
47	1.678	2.012	2.408	2.685	3.273	3.510
48	1.677	2.011	2.407	2.682	3.269	3.505
49	1.677	2.010	2.405	2.68	3.265	3.500
50	1.676	2.009	2.403	2.678	3.261	3.496

Таблиця Д1. Продовження.

Число ступенів свободи $\nu$	Рівень значимості $\alpha$ (двохстороння критична область)					
	0,1	0,05	0,02	0,01	0,002	0,001
51	1.675	2.008	2.402	2.676	3.258	3.492
52	1.675	2.007	2.400	2.674	3.255	3.488
53	1.674	2.006	2.399	2.672	3.251	3.484
54	1.674	2.005	2.397	2.670	3.248	3.480
55	1.673	2.004	2.396	2.668	3.245	3.476
56	1.673	2.003	2.395	2.667	3.242	3.473
57	1.672	2.002	2.394	2.665	3.239	3.470
58	1.672	2.002	2.392	2.663	3.237	3.466
59	1.671	2.001	2.391	2.662	3.234	3.463
60	1.671	2.000	2.390	2.660	3.232	3.460
61	1.670	2.000	2.389	2.659	3.229	3.457
62	1.670	1.999	2.388	2.657	3.227	3.454
63	1.669	1.998	2.387	2.656	3.225	3.452
64	1.669	1.998	2.386	2.655	3.223	3.449
65	1.669	1.997	2.385	2.654	3.220	3.447
66	1.668	1.997	2.384	2.652	3.218	3.444
67	1.668	1.996	2.383	2.651	3.216	3.442
68	1.668	1.995	2.382	2.650	3.214	3.439
69	1.667	1.995	2.382	2.649	3.213	3.437
70	1.667	1.994	2.381	2.648	3.211	3.435
71	1.667	1.994	2.380	2.647	3.209	3.433
72	1.666	1.993	2.379	2.646	3.207	3.431
73	1.666	1.993	2.379	2.645	3.206	3.429
74	1.666	1.993	2.378	2.644	3.204	3.427
75	1.665	1.992	2.377	2.643	3.202	3.425
76	1.665	1.992	2.376	2.642	3.201	3.423
77	1.665	1.991	2.376	2.641	3.199	3.421
78	1.665	1.991	2.375	2.640	3.198	3.420
79	1.664	1.990	2.374	2.640	3.197	3.418
80	1.664	1.990	2.374	2.639	3.195	3.416
81	1.664	1.990	2.373	2.638	3.194	3.415
82	1.664	1.989	2.373	2.637	3.193	3.413
83	1.663	1.989	2.372	2.636	3.191	3.412
84	1.663	1.989	2.372	2.636	3.190	3.410
85	1.663	1.988	2.371	2.635	3.189	3.409
86	1.663	1.988	2.370	2.634	3.188	3.407
87	1.663	1.988	2.370	2.634	3.187	3.406
88	1.662	1.987	2.369	2.633	3.185	3.405
89	1.662	1.987	2.369	2.632	3.184	3.403
90	1.662	1.987	2.368	2.632	3.183	3.402
91	1.662	1.986	2.368	2.631	3.182	3.401
92	1.662	1.986	2.368	2.630	3.181	3.399
93	1.661	1.986	2.367	2.630	3.180	3.398
94	1.661	1.986	2.367	2.629	3.179	3.397
95	1.661	1.985	2.366	2.629	3.178	3.396
96	1.661	1.985	2.366	2.628	3.177	3.395
97	1.661	1.985	2.365	2.627	3.176	3.394
98	1.661	1.984	2.365	2.627	3.175	3.393
99	1.660	1.984	2.365	2.626	3.175	3.392
100	1.660	1.984	2.364	2.626	3.174	3.39
$\infty$	1,64	1,96	2,33	2,58	3,09	3,29
	<b>0,05</b>	<b>0,025</b>	<b>0,01</b>	<b>0,005</b>	<b>0,001</b>	<b>0,0005</b>
<b>Рівень значимості <math>\alpha</math> (одностороння критична область)</b>						

**Таблиця Д2. Значення F-критерію Фішера для рівня значимості  $\alpha=0,01$ .**

$\nu_2$	$\nu_1$																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	4052.181	4999.5	5403.4	5624.6	5763.7	5858.9	5928.4	5981.1	6022.5	6055.9	6106.3	6157.3	6208.7	6234.6	6260.7	6286.8	6313.0	6339.4	6365.9
2	98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.416	99.433	99.449	99.458	99.466	99.474	99.482	99.491	99.499
3	34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	27.052	26.872	26.690	26.598	26.505	26.411	26.316	26.221	26.125
4	21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.374	14.198	14.020	13.929	13.838	13.745	13.652	13.558	13.463
5	16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.888	9.722	9.553	9.466	9.379	9.291	9.202	9.112	9.020
6	13.745	10.925	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	12.246	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	11.259	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	10.561	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	10.044	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.996	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.744	4.632	4.539	4.397	4.251	4.099	4.021	3.941	3.860	3.776	3.690	3.602
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.780	3.701	3.619	3.535	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.665	3.587	3.507	3.425	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.004	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.959	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.084	3.003	2.920	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.859	2.778	2.695	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.667	2.583	2.495	2.403	2.305
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.702	2.620	2.535	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.403	2.310	2.211
25	7.770	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.131
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.198	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	7.598	5.420	4.538	4.045	3.725	3.499	3.330	3.198	3.092	3.005	2.868	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	7.562	5.390	4.510	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.288	2.203	2.114	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.718	2.632	2.496	2.352	2.198	2.115	2.028	1.936	1.836	1.726	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.950	1.860	1.763	1.656	1.533	1.381
inf	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.791	1.696	1.592	1.473	1.325	1.000

Примітка:  $\nu_1$  і  $\nu_2$  – ступені свободи для дисперсій в чисельнику і знаменнику відповідно.

**Таблиця Д3. Значення F-критерію Фішера для рівня значимості  $\alpha=0,025$ .**

$\nu_2$	$\nu_1$																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	647.7890	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.29	968.63	976.71	984.87	993.10	997.25	1001.4	1005.6	1009.8	1014.0	1018.3
2	38.5063	39.000	39.166	39.248	39.298	39.332	39.355	39.373	39.387	39.398	39.415	39.431	39.448	39.456	39.465	39.473	39.481	39.490	39.498
3	17.4434	16.044	15.439	15.101	14.885	14.735	14.624	14.54	14.473	14.419	14.337	14.253	14.167	14.124	14.081	14.037	13.992	13.947	13.902
4	12.2179	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7512	8.6565	8.5599	8.5109	8.461	8.411	8.360	8.309	8.257
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5245	6.4277	6.3286	6.2780	6.227	6.175	6.123	6.069	6.015
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.3662	5.2687	5.1684	5.1172	5.065	5.012	4.959	4.904	4.849
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.6658	4.5678	4.4667	4.4150	4.362	4.309	4.254	4.199	4.142
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.1997	4.1012	3.9995	3.9472	3.894	3.840	3.784	3.728	3.670
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.8682	3.7694	3.6669	3.6142	3.560	3.505	3.449	3.392	3.333
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6209	3.5217	3.4185	3.3654	3.311	3.255	3.198	3.140	3.080
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4296	3.3299	3.2261	3.1725	3.118	3.061	3.004	2.944	2.883
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.2773	3.1772	3.0728	3.0187	2.963	2.906	2.848	2.787	2.725
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1532	3.0527	2.9477	2.8932	2.837	2.780	2.720	2.659	2.595
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0502	2.9493	2.8437	2.7888	2.732	2.674	2.614	2.552	2.487
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	2.9633	2.8621	2.7559	2.7006	2.644	2.585	2.524	2.461	2.395
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488	2.9862	2.8890	2.7875	2.6808	2.6252	2.568	2.509	2.447	2.383	2.316
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849	2.9222	2.8249	2.7230	2.6158	2.5598	2.502	2.442	2.380	2.315	2.247
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291	2.8664	2.7689	2.6667	2.5590	2.5027	2.445	2.384	2.321	2.256	2.187
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8801	2.8172	2.7196	2.6171	2.5089	2.4523	2.394	2.333	2.270	2.203	2.133
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.6758	2.5731	2.4645	2.4076	2.349	2.287	2.223	2.156	2.085
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	2.8740	2.7977	2.7348	2.6368	2.5338	2.4247	2.3675	2.308	2.246	2.182	2.114	2.042
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628	2.6998	2.6017	2.4984	2.3890	2.3315	2.272	2.210	2.145	2.076	2.003
23	5.7498	4.3492	3.7505	3.4083	3.1835	3.0232	2.9023	2.8077	2.7313	2.6682	2.5699	2.4665	2.3567	2.2989	2.239	2.176	2.111	2.041	1.968
24	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027	2.6396	2.5411	2.4374	2.3273	2.2693	2.209	2.146	2.080	2.010	1.935
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5149	2.4110	2.3005	2.2422	2.182	2.118	2.052	1.981	1.906
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528	2.5896	2.4908	2.3867	2.2759	2.2174	2.157	2.093	2.026	1.954	1.878
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	2.7074	2.6309	2.5676	2.4688	2.3644	2.2533	2.1946	2.133	2.069	2.002	1.930	1.853
28	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	2.6872	2.6106	2.5473	2.4484	2.3438	2.2324	2.1735	2.112	2.048	1.980	1.907	1.829
29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	2.6686	2.5919	2.5286	2.4295	2.3248	2.2131	2.1540	2.092	2.028	1.959	1.886	1.807
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.4120	2.3072	2.1952	2.1359	2.074	2.009	1.940	1.866	1.787
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.2882	2.1819	2.0677	2.0069	1.943	1.875	1.803	1.724	1.637
60	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344	2.2702	2.1692	2.0613	1.9445	1.8817	1.815	1.744	1.667	1.581	1.482
120	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217	2.1570	2.0548	1.9450	1.8249	1.7597	1.690	1.614	1.530	1.433	1.310
inf	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	2.1918	2.1136	2.0483	1.9447	1.8326	1.7085	1.6402	1.566	1.484	1.388	1.268	1.000

Примітка:  $\nu_1$  і  $\nu_2$  – ступені свободи для дисперсій в чисельнику і знаменнику відповідно.

**Таблиця Д4. Значення F-критерію Фішера для рівня значимості  $\alpha=0,05$ .**

$\nu_2$	$\nu_1$																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	161.4476	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05	250.1	251.14	252.2	253.25	254.31
2	18.5128	19.000	19.164	19.247	19.297	19.33	19.353	19.371	19.385	19.396	19.413	19.429	19.446	19.454	19.462	19.471	19.479	19.487	19.496
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5264
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
inf	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

Примітка:  $\nu_1$  і  $\nu_2$  – ступені свободи для дисперсій в чисельнику і знаменнику відповідно.

**Таблиця Д5. Значення F-критерію Фішера для рівня значимості  $\alpha=0,10$ .**

$v_2$	$v_1$																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	INF
1	39.86346	49.5000	53.5932	55.8329	57.2401	58.2044	58.906	59.439	59.8576	60.195	60.7052	61.2203	61.7403	62.0021	62.265	62.5291	62.7943	63.0606	63.3281
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131	9.44962	9.45793	9.46624	9.47456	9.48289	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562	5.20031	5.18448	5.17636	5.16811	5.15972	5.15119	5.14251	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434	3.83099	3.81742	3.80361	3.78957	3.77527	3.76073
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29740	3.26824	3.23801	3.20665	3.19052	3.17408	3.15732	3.14023	3.12279	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446	2.98304	2.95774	2.93693	2.90472	2.87122	2.83634	2.81834	2.79996	2.78117	2.76195	2.74229	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473	2.57533	2.55546	2.53510	2.51422	2.49279	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58935	2.56124	2.53804	2.50196	2.46422	2.42464	2.40410	2.38302	2.36136	2.33910	2.31618	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832	2.27683	2.25472	2.23196	2.20849	2.18427	2.15923
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34731	2.32260	2.28405	2.24351	2.20074	2.17843	2.15543	2.13169	2.10716	2.08176	2.05542
11	3.22520	2.85951	2.66023	2.53619	2.45118	2.38907	2.34157	2.30400	2.27350	2.24823	2.20873	2.16709	2.12305	2.10001	2.07621	2.05161	2.02612	1.99965	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	2.24457	2.21352	2.18776	2.14744	2.10485	2.05968	2.03599	2.01149	1.98610	1.95973	1.93228	1.90361
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.23410	2.19535	2.16382	2.13763	2.09659	2.05316	2.00698	1.98272	1.95757	1.93147	1.90429	1.87591	1.84620
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.19313	2.15390	2.12195	2.09540	2.05371	2.00953	1.96245	1.93766	1.91193	1.88516	1.85723	1.82800	1.79728
15	3.07319	2.69517	2.48979	2.36143	2.27302	2.20808	2.15818	2.11853	2.08621	2.05932	2.01707	1.97222	1.92431	1.89904	1.87277	1.84539	1.81676	1.78672	1.75505
16	3.04811	2.66817	2.46181	2.33274	2.24376	2.17833	2.12800	2.08798	2.05533	2.02815	1.98539	1.93992	1.89127	1.86556	1.83879	1.81084	1.78156	1.75075	1.71817
17	3.02623	2.64464	2.43743	2.30775	2.21825	2.15239	2.10169	2.06134	2.02839	2.00094	1.95772	1.91169	1.86236	1.83624	1.80901	1.78053	1.75063	1.71909	1.68564
18	3.00698	2.62395	2.41601	2.28577	2.19583	2.12958	2.07854	2.03789	2.00467	1.97698	1.93334	1.88681	1.83685	1.81035	1.78269	1.75371	1.72322	1.69099	1.65671
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.05802	2.01710	1.98364	1.95573	1.91170	1.86471	1.81416	1.78731	1.75924	1.72979	1.69876	1.66587	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	1.99853	1.96485	1.93674	1.89236	1.84494	1.79384	1.76667	1.73822	1.70833	1.67678	1.64326	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	1.98186	1.94797	1.91967	1.87497	1.82715	1.77555	1.74807	1.71927	1.68896	1.65691	1.62278	1.58615
22	2.94858	2.56131	2.35117	2.21927	2.12794	2.06050	2.00840	1.96680	1.93273	1.90425	1.85925	1.81106	1.75899	1.73122	1.70208	1.67138	1.63885	1.60415	1.56678
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04723	1.99492	1.95312	1.91888	1.89025	1.84497	1.79643	1.74392	1.71588	1.68643	1.65535	1.62237	1.58711	1.54903
24	2.92712	2.53833	2.32739	2.19488	2.10303	2.03513	1.98263	1.94066	1.90625	1.87748	1.83194	1.78308	1.73015	1.70185	1.67210	1.64067	1.60726	1.57146	1.53270
25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.97138	1.92925	1.89469	1.86578	1.82000	1.77083	1.71752	1.68898	1.65895	1.62718	1.59335	1.55703	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	1.91876	1.88407	1.85503	1.80902	1.75957	1.70589	1.67712	1.64682	1.61472	1.58050	1.54368	1.50360
27	2.90119	2.51061	2.29871	2.16546	2.07298	2.00452	1.95151	1.90909	1.87427	1.84511	1.79889	1.74917	1.69514	1.66616	1.63560	1.60320	1.56859	1.53129	1.49057
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.94270	1.90014	1.86520	1.83593	1.78951	1.73954	1.68519	1.65600	1.62519	1.59250	1.55753	1.51976	1.47841
29	2.88703	2.49548	2.28307	2.14941	2.05658	1.98781	1.93452	1.89184	1.85679	1.82741	1.78081	1.73060	1.67593	1.64655	1.61551	1.58253	1.54721	1.50899	1.46704
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98033	1.92692	1.88412	1.84896	1.81949	1.77270	1.72227	1.66731	1.63774	1.60648	1.57323	1.53757	1.49891	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.87252	1.82886	1.79290	1.76269	1.71456	1.66241	1.60515	1.57411	1.54108	1.50562	1.46716	1.42476	1.37691
60	2.79107	2.39325	2.17741	2.04099	1.94571	1.87472	1.81939	1.77483	1.73802	1.70701	1.65743	1.60337	1.54349	1.51072	1.47554	1.43734	1.39520	1.34757	1.29146
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.76748	1.72196	1.68425	1.65238	1.60120	1.54500	1.48207	1.44723	1.40938	1.36760	1.32034	1.26457	1.19256
inf	2.70554	2.30259	2.08380	1.94486	1.84727	1.77411	1.71672	1.67020	1.63152	1.59872	1.54578	1.48714	1.42060	1.38318	1.34187	1.29513	1.23995	1.16860	1.00000

Примітка:  $v_1$  і  $v_2$  – ступені свободи для дисперсій в чисельнику і знаменнику відповідно.

**Таблиця Дб. Значення коефіцієнту Кохрена (G-критерію)  
для довірчої ймовірності  $p=99\%$  і числі ступенів свободи  $k$ .**

Кількість дослідів, $n$	Число ступенів свободи, $k$													
	1	2	3	4	5	6	7	8	9	10	16	36	144	$\infty$
2	0.9999	0.9950	0.9794	0.9586	0.9373	0.9172	0.8988	0.8823	0.8674	0.8539	0.7949	0.7067	0.6062	0.5000
3	0.9933	0.9423	0.8831	0.8335	0.7933	0.7606	0.7335	0.7107	0.6912	0.6743	0.6059	0.5153	0.4230	0.3333
4	0.9676	0.8643	0.7814	0.7212	0.6761	0.6410	0.6129	0.5897	0.5702	0.5536	0.4884	0.4057	0.3251	0.2500
5	0.9279	0.7885	0.6957	0.6329	0.5875	0.5531	0.5259	0.5037	0.4854	0.4697	0.4094	0.3351	0.2644	0.2000
6	0.8828	0.7218	0.6258	0.5635	0.5195	0.4866	0.4608	0.4401	0.4229	0.4084	0.3529	0.2858	0.2229	0.1667
7	0.8376	0.6644	0.5685	0.5080	0.4659	0.4347	0.4105	0.3911	0.3751	0.3616	0.3105	0.2494	0.1929	0.1429
8	0.7945	0.6152	0.5209	0.4627	0.4226	0.3932	0.3704	0.3522	0.3373	0.3248	0.2779	0.2214	0.1700	0.1250
9	0.7544	0.5727	0.4810	0.4251	0.3870	0.3592	0.3378	0.3207	0.3067	0.2950	0.2514	0.1992	0.1521	0.1111
10	0.7175	0.5358	0.4469	0.3934	0.3572	0.3308	0.3106	0.2945	0.2813	0.2704	0.2297	0.1811	0.1376	0.1000
12	0.6528	0.4751	0.3919	0.3428	0.3099	0.2861	0.2680	0.2535	0.2419	0.2320	0.1961	0.1535	0.1157	0.0833
15	0.5747	0.4069	0.3317	0.2882	0.2593	0.2386	0.2228	0.2104	0.2002	0.1918	0.1612	0.1251	0.0934	0.0667
20	0.4799	0.3297	0.2654	0.2288	0.2048	0.1877	0.1748	0.1646	0.1567	0.1501	0.1248	0.0960	0.0709	0.0500
24	0.4247	0.2871	0.2295	0.1970	0.1759	0.1608	0.1495	0.1406	0.1338	0.1283	0.1060	0.0810	0.0595	0.0417
30	0.3632	0.2412	0.1913	0.1635	0.1454	0.1327	0.1232	0.1157	0.1100	0.1054	0.0867	0.0658	0.0480	0.0333
40	0.2940	0.1915	0.1508	0.1281	0.1135	0.1033	0.0957	0.0898	0.0853	0.0816	0.0668	0.0503	0.0363	0.0250
60	0.2151	0.1371	0.1069	0.0902	0.0796	0.0722	0.0668	0.0625	0.0594	0.0567	0.0461	0.0344	0.0245	0.0167
120	0.1225	0.0759	0.0585	0.0489	0.0429	0.0387	0.0357	0.0334	0.0316	0.0302	0.0242	0.0178	0.0125	0.0083
$\infty$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

**Таблиця Д7. Значення коефіцієнту Кохрена (G-критерію)  
для довірчої ймовірності  $p=95\%$  і числі ступенів свободи  $k$ .**

Кількість дослідів, $n$	Число ступенів свободи, $k$													
	1	2	3	4	5	6	7	8	9	10	16	36	144	$\infty$
<b>2</b>	0.9985	0.9750	0.9392	0.9057	0.8772	0.8534	0.8332	0.8159	0.8010	0.7880	0.7341	0.6602	0.5813	0.5000
<b>3</b>	0.9669	0.8709	0.7977	0.7457	0.7071	0.6771	0.6530	0.6333	0.6167	0.6025	0.5466	0.4748	0.4031	0.3333
<b>4</b>	0.9065	0.7679	0.6841	0.6287	0.5895	0.5598	0.5365	0.5175	0.5017	0.4884	0.4366	0.3720	0.3093	0.2500
<b>5</b>	0.8412	0.6338	0.5981	0.5440	0.5063	0.4783	0.4564	0.4387	0.4241	0.4118	0.3645	0.3066	0.2013	0.2000
<b>6</b>	0.7808	0.6161	0.5321	0.4803	0.4447	0.4184	0.3980	0.3817	0.3682	0.3568	0.3135	0.2612	0.2119	0.1667
<b>7</b>	0.7271	0.5612	0.4800	0.4307	0.3974	0.3726	0.3535	0.3384	0.3259	0.3154	0.2756	0.2278	0.1833	0.1429
<b>8</b>	0.6798	0.5157	0.4377	0.3910	0.3595	0.3362	0.3185	0.3043	0.2926	0.2829	0.2462	0.2022	0.1616	0.1250
<b>9</b>	0.6385	0.4775	0.4027	0.3584	0.3286	0.3067	0.2901	0.2768	0.2659	0.2568	0.2226	0.1820	0.1446	0.1111
<b>10</b>	0.6020	0.4450	0.3733	0.3311	0.3029	0.2823	0.2666	0.2541	0.2439	0.2353	0.2032	0.1655	0.1308	0.1000
<b>12</b>	0.5410	0.3924	0.3624	0.2880	0.2624	0.2439	0.2299	0.2187	0.2098	0.2020	0.1737	0.1403	0.1100	0.0833
<b>15</b>	0.4709	0.3346	0.2758	0.2419	0.2195	0.2034	0.1911	0.1815	0.1736	0.1671	0.1429	0.1144	0.0889	0.0667
<b>20</b>	0.3894	0.2705	0.2205	0.1921	0.1735	0.1602	0.1501	0.1422	0.1357	0.1303	0.1108	0.0879	0.0675	0.0500
<b>24</b>	0.3434	0.2354	0.1907	0.1656	0.1493	0.1374	0.1286	0.1216	0.1160	0.1113	0.0942	0.0743	0.0567	0.0417
<b>30</b>	0.2929	0.1980	0.1593	0.1377	0.1237	0.1137	0.1061	0.1002	0.0958	0.0921	0.0771	0.0604	0.0457	0.0333
<b>40</b>	0.2370	0.1576	0.1259	0.1082	0.0968	0.0887	0.0827	0.0780	0.0745	0.0713	0.0595	0.0462	0.0347	0.0250
<b>60</b>	0.1737	0.1131	0.0895	0.0765	0.0682	0.0623	0.0583	0.0552	0.0520	0.0497	0.0411	0.0316	0.0234	0.0167
<b>120</b>	0.0998	0.0632	0.0495	0.0419	0.0371	0.0337	0.0312	0.0292	0.0279	0.0266	0.0218	0.0165	0.0120	0.0083
$\infty$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

**Таблиця Д8. Значення  $\chi^2$ -розподілення для рівнів значимості  $\alpha < 0,05$  і  $\alpha < 0,01$ .**

<b>k</b>	<b><math>\alpha</math></b>		<b>k</b>	<b><math>\alpha</math></b>		<b>k</b>	<b><math>\alpha</math></b>	
	<b>0,05</b>	<b>0,01</b>		<b>0,05</b>	<b>0,01</b>		<b>0,05</b>	<b>0,01</b>
<b>1</b>	3,84	6,64	<b>31</b>	45,0	52,2	<b>72</b>	92,8	103
<b>2</b>	5,99	9,21	<b>32</b>	46,2	53,5	<b>74</b>	95,1	105
<b>3</b>	7,82	11,3	<b>33</b>	47,4	54,8	<b>76</b>	97,4	108
<b>4</b>	9,49	13,3	<b>34</b>	48,6	56,1	<b>78</b>	99,6	110
<b>5</b>	11,1	15,1	<b>35</b>	49,8	57,3	<b>80</b>	102	112
<b>6</b>	12,6	16,8	<b>36</b>	51,0	58,6	<b>82</b>	104	115
<b>7</b>	14,1	18,5	<b>37</b>	52,2	59,9	<b>84</b>	106	117
<b>8</b>	15,5	20,1	<b>38</b>	53,4	61,2	<b>86</b>	109	119
<b>9</b>	16,9	21,7	<b>39</b>	54,6	62,4	<b>88</b>	111	122
<b>10</b>	18,3	23,2	<b>40</b>	55,8	63,7	<b>90</b>	113	124
<b>11</b>	19,7	24,7	<b>41</b>	56,9	65,0	<b>92</b>	115	126
<b>12</b>	21,0	26,2	<b>42</b>	58,1	66,2	<b>94</b>	118	129
<b>13</b>	22,4	27,7	<b>43</b>	59,3	67,5	<b>96</b>	120	131
<b>14</b>	23,7	29,1	<b>44</b>	60,5	68,7	<b>98</b>	122	133
<b>15</b>	25,0	30,6	<b>45</b>	61,7	70,0	<b>100</b>	124	136
<b>16</b>	26,3	32,0	<b>46</b>	62,8	71,2	<b>110</b>	135	147
<b>17</b>	27,6	33,4	<b>47</b>	64,0	72,4	<b>120</b>	147	159
<b>18</b>	28,9	34,8	<b>48</b>	65,2	73,7	<b>130</b>	158	170
<b>19</b>	30,1	36,2	<b>49</b>	66,3	74,9	<b>140</b>	169	182
<b>20</b>	31,4	37,6	<b>50</b>	67,5	76,2	<b>150</b>	180	193
<b>21</b>	32,7	38,9	<b>52</b>	69,8	78,6	<b>200</b>	234	249
<b>22</b>	33,9	40,3	<b>54</b>	72,2	81,1	<b>250</b>	288	305
<b>23</b>	35,2	41,6	<b>56</b>	74,5	83,5	<b>300</b>	341	360
<b>24</b>	36,4	43,0	<b>58</b>	76,8	86,0	<b>400</b>	448	469
<b>25</b>	37,6	44,3	<b>60</b>	79,1	88,4	<b>500</b>	553	576
<b>26</b>	38,9	45,6	<b>62</b>	81,4	90,8	<b>600</b>	658	683
<b>27</b>	40,1	47,0	<b>64</b>	83,7	93,2	<b>700</b>	763	790
<b>28</b>	41,3	48,3	<b>66</b>	86,0	95,6	<b>800</b>	867	896
<b>29</b>	42,6	49,6	<b>68</b>	88,2	98,0	<b>900</b>	971	1002
<b>30</b>	43,8	50,9	<b>70</b>	90,5	100	<b>1000</b>	1075	1107

*Примітка:*  $k$  – число ступенів свободи.