

Таблиця 7.6

Купує * Сімейний стан Crosstabulation

			Сімейний стан		
			одружений	неодружений	Total
Купує	Багато	Count	32	12	44
		Expected Count	32,9	11,1	44,0
		% within Сімейний стан	29,9%	33,3%	30,8%
		Count	70	19	89
	Expected Count		66,6	22,4	89,0
	% within Сімейний стан		65,4%	52,8%	62,2%
	Практиче- Count		5	5	10
	ски не Expected Count купує		7,5	2,5	10,0
	% within Сімейний стан		4,7%	13,9%	7,0%
	Total	Count	107	36	143
		Expected Count	107,0	36,0	143,0
		% within Сімейний стан	100,0%	100,0%	100,0%

Таблиця 7.7

Chi - Square Tests

	Value	df	Asymp. Sig. (2 - sided)
Pearson Chi - Square	4,066 ^a	2	,131
Likelihood Ratio	3,649	2	,161
Linear - by - Linear Association	,279	1	,597
N of Valid Cases	143		

a. 1 cells(16.7%) have expected count less than 5.

Аналогічно вирішити завдання про купівлю модного одягу жінками.

Завдання для самостійного виконання

1. За даними таблиці 7.8 визначте, чи існує залежність між віком клієнта і його витратами на літній відпочинок.

Таблиця 7.8

Залежність між віком клієнта і витратами на літній відпочинок

Вік	Рівень витрат на літній відпочинок			Всього
	Високий	Середній	Низький	
Молоді	25	40	31	
Середнього віку	52	63	47	
Літні	12	23	58	
Всього				

2. За даними таблиці 7.9 перевірити припущення про те, що частота звертань до телефонної служби довіри нерівномірно розподіляється по днях

тижня (за допомогою критерію χ^2 порівняти заданий емпіричний розподіл рівномірним).

Таблиця 7.9

Розподіл звертань до телефонної служби довіри по днях тижня

День тижня	Число звернень
Понеділок	9
Вівторок	6
Середа	7
Четвер	6
П'ятниця	9
Субота	15
Неділя	18
Всього	70

Вимоги до звіту:

Звіт повинен містити:

- відповіді на контрольні питання;
- файл з результатами розрахунків.

Контрольні запитання

1. Призначення критерію Пірсона χ^2 .
2. Обмеження критерію Пірсона χ^2 .
3. Що таке емпіричні частоти?
4. Що таке теоретичні частоти?
5. Як обчислюється критерій Пірсона χ^2 в Excel?
6. Як представити дані для розрахунків критерію Пірсона χ^2 за допомогою SPSS?
7. Дайте змістовну інтерпретацію результатам отриманим у вирішенні завдань.

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 8

Тема: *Обчислення коефіцієнта кореляції Пірсона-Браве для метричних шкал*

Мета: навчитися обчислювати коефіцієнт кореляції Пірсона, будувати діаграму розсіювання, визначати значущість отриманого коефіцієнта.

Короткі теоретичні відомості

Для визначення кореляційної залежності між двома змінними, заданими метричними шкалами, слугує *коефіцієнт Пірсона-Браве*. Він характеризує міру лінійного зв'язку між змінними. Він характеризує міру лінійного зв'язку між змінними.

Нехай X і Y дві змінних, що корелюють між собою. Коефіцієнт Пірсона-Браве обчислюється за формулою:

$$r = \frac{\sum_{i=1}^n Z_i^x \cdot Z_i^y}{n-1}, \text{ где} \quad (8.1)$$

$$Z_i^x = \frac{X_i - X_{cp}}{S_x}, \text{ где} \quad (8.2)$$

S_x - стандартне відхилення,

X_i – i -значення ознаки,

X_{cp} - середнє значення ознаки X .

$$Z_i^y = \frac{Y_i - Y_{cp}}{S_y}, \text{ где} \quad (8.3)$$

S_y - стандартне відхилення,

Y_i – i -значення,

Y_{cp} - середнє значення ознаки Y .

Постановка завдання

У таблиці 8.1 наведені дані про кількість машин, що щодня паркуються на відкритій стоянці і в гаражі біля великого університету. Знайдіть і інтерпретуйте кореляцію між цими двома величинами.

Таблиця 8.1

Кількість автомобілів, які паркувалися біля університету

День тижня	Відкрита стоянка	Гараж
Понеділок	140	180
Вівторок	120	200
Середовище	130	190
Четвер	110	210
П'ятниця	160	160
Субота	135	185

Рішення за допомогою Excel

Хід роботи:

1. Ввести дані.
2. Вичислити коефіцієнт Пірсона для цих даних спочатку за формулами, а потім виконати теж саме за допомогою функції КОРРЕЛ (рис. 8.1).

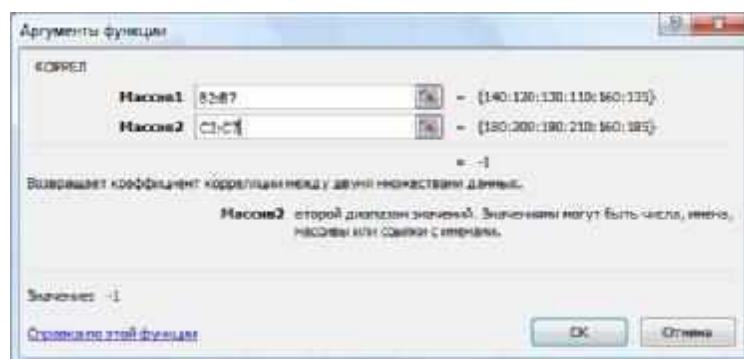


Рис. 8.1. Діалогове вікно «Аргументи функції КОРРЕЛ»

3. Заповнити в зошиті таблицю 8.1 (внести розрахункове і критичне значення коефіцієнта, визначити його значущість по таблиці критичних значень).

Таблиця 8.1

Об'єм вибірки	Емпіричне значення коефіцієнта	Критичне значення	Резюме (означаємо/ незначимий)

4. Побудувати діаграму розсіювання (виділити дані, викликати майстер діаграм, вибрати тип діаграми - «точкова», підписати усі заголовки) (див. рис. 8.2).

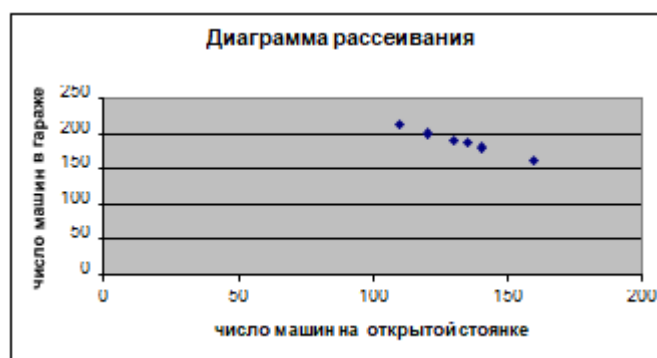


Рис. 8.2. Діаграма розсіювання, побудована за даними таблиці 8.1.

Рішення за допомогою SPSS

Хід роботи:

1. Створити файл даних в SPSS: описати дві змінні (Open - число машин на відкритій стоянці, Garag - число машин в гаражі).
2. Ввести або скопіювати з файлу Excel дані.
3. Вичислити коефіцієнт Пірсона для цих даних: Меню: **Analyze** (Аналіз), **Correlate** (Кореляція), **Bivariate** (Двовимірна).

У вікні **Bivariate Correlation** (рис. 8.3) задати необхідні параметри.

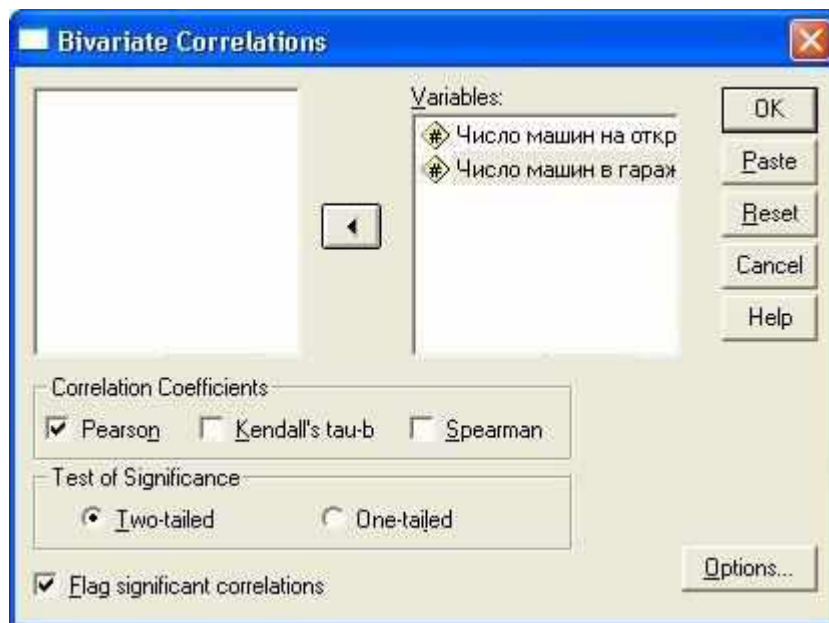


Рис. 8.3. Діалогове вікно «Bivariate Correlation»

4. Ознайомитися з отриманими результатами у вікні перегляду (рис. 8.4). Результати містять: коефіцієнт кореляції (**Pearson Correlation**), *рівень значущості (Sig)*, об'єм вибірки (**N**). Записати їх у зошит. Визначити, чи значимий коефіцієнт кореляції.

Correlations

		Число машин на відкритій стоянці	Число машин в гаражі
Число машин на відкритій стоянці	Pearson Correlation	1	-1,000**
	Sig. (2-tailed)	.	.
	N	6	6
Число машин в гаражі	Pearson Correlation	-1,000**	1
	Sig. (2-tailed)	.	.
	N	6	6

** Correlation is significant at the 0.01 level (2-tailed).

Рис. 8.4. Вікно перегляду для прикладу з парковкою

5. Побудувати діаграму розсіювання:

Меню: **Graphs** (Графіки), **Scatter...**(Діаграма розсіювання), **Simple** (Проста) (рис. 8.5).

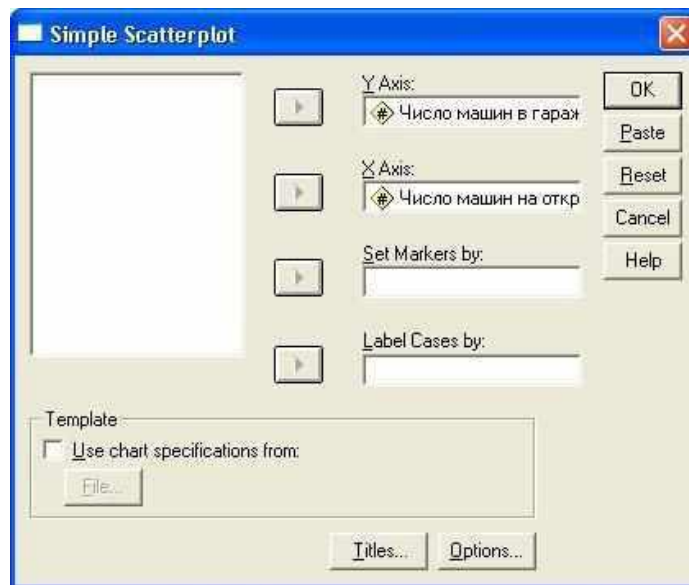


Рис. 8.5. Діалогове вікно «Simple Scatterplot» (проста точкова діаграм)

Розглянути отриману діаграму розсіювання (рис. 8.6).

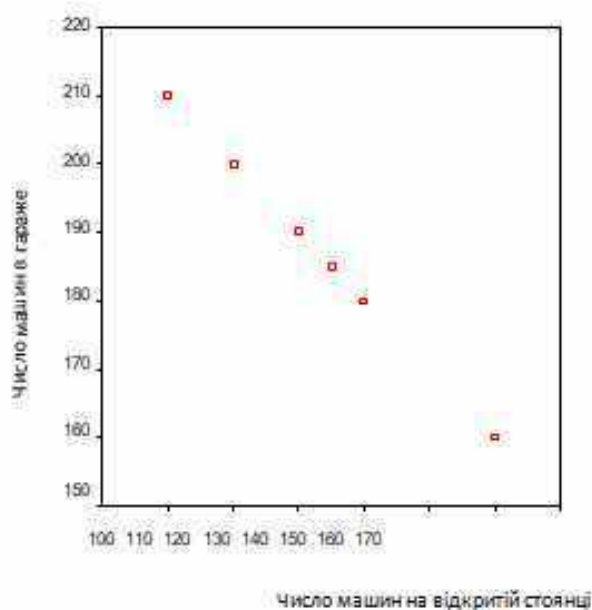


Рис. 8.6. Точкова діаграма для прикладу з парковкою

6. Зберегти дані розрахунків і початкові дані у файлах.
7. Зробити висновок про наявність або відсутність значимої кореляції між двома ознаками.

Завдання для самостійного виконання

1. У таблиці 8.2 містяться дані про щоденний об'єм продажів в доларах лазерних принтерів, картриджів до принтерів і шкільного приладдя. Знайдіть і інтерпретуйте кореляцію між цими величинами. Вирішіть завдання в середовищі Microsoft Excel і в середовищі SPSS.

Таблиця 8.2.

Щоденний об'єм продажів (\$)

День	Лазерні принтери	Картриджі	Шкільне приладдя
1	17291	4379	3618
2	13734	2258	3514
3	18802	4206	3587
4	12171	2137	3007
5	16402	3775	3850
6	19904	4781	3675
7	14023	1991	3120
8	17513	2663	3345
9	17847	4451	2045
10	12718	1648	3292
11	12292	2342	3405
12	11846	2646	2799
13	17088	3216	2417
14	13523	2184	2405

Вимоги до звіту

Звіт повинен містити:

- відповіді на контрольні питання;
- файл з результатами розрахунків.

Контрольні запитання

1. Дайте визначення функціональної залежності.
2. Дайте визначення кореляційної залежності.
3. Які значення можуть приймати коефіцієнти кореляції?
4. Як вичислити коефіцієнт кореляції Пірсона-Браве з допомогою електронних таблиць?
5. Для яких шкал він використовується?
6. Як побудувати діаграму розсіювання?
7. Які параметри задаються функції КОРРЕЛ?
8. Як розрахувати коефіцієнт кореляції Пірсона-Браве і побудувати діаграму розсіювання в SPSS?
9. Як визначити значущість коефіцієнта кореляції в SPSS?
10. Дайте змістовну інтерпретацію результатам розрахунків.

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 9

Тема: *Обчислення коефіцієнта рангової кореляції Спірмена*

Мета: навчитися обчислювати коефіцієнт рангової кореляції Спірмена і визначати його значущість.

Короткі теоретичні відомості

Нехай N об'єктів можуть бути впорядковані як за ознакою X , так і по Y .
Позначимо

R_{ix} - ранг i - об'єкту за ознакою X

R_{iy} - ранг i - об'єкту за ознакою Y

$d_i = R_{ix} - R_{iy}$ - міра неспівпадіння рангів

Для визначення кореляційної залежності між двома змінними, заданими ранговими шкалами служить коефіцієнт рангової кореляції Спірмена, який обчислюється за формулою:

$$\rho = 1 - \frac{6 \sum_{i=1}^N di^2}{N(N^2 - 1)} \quad (9.1)$$

Постановка завдання

Дослідження, проведене на британському ринку, виявило наступну значущість чинників успіху нових продуктів, вироблюваних японськими і британськими фірмами (див. таблиця 9.1, в якій n означає число опитаних фірм).

З таблиці 9.1 видно, що найчастіше ключовим чинником успіху називали міру адаптації до вимог покупців. І тільки відносно невелика доля фірм(1/4), як японських, так і британських, вважає ефективний маркетинг важливим критерієм успіху.

Порівняння оцінок, отриманих від японських і британських фірм, виявляє, що перші з них надають більше значення наявності конкурентної переваги, а другі - більшій адаптації до запитів споживачів.

Таблиця 9.1

Чинники успіху нових продуктів, виявлених на британському ринку

Чинники успіху	Відсоток фірм	
	японських($n=116$)	британських($n=86$)
Хороша адаптація до потреб	69,8	75,6
Перевага над конкурентами за якістю	79,3	59,3
Перевага над конкурентами по дизайну	69,9	45,3
Перевага над конкурентами по співвідношенню гідність/ціна	58,6	61,6
Перевага над конкурентами по конструкції	55,2	48,8
Дуже конкурентна ціна	41,4	27,9
Адаптація	39,7	34,9
Унікальність	36,2	29,1
Ефективний маркетинг	27,6	25,6
Глибокий аналіз ринку	27,6	18,6

Великий об'єм ринку	20,7	16,3
Синергія виробництво/маркетинг	16,4	18,6
Ухилення від ринків з високою конкуренцією і задоволеними покупцями	7,8	10,5
Ухилення від динамічних ринків з частою зміною товару	2,6	4,7

Визначте, чи корелюють між собою оцінки японських і британських фірм.

Рішення задачі за допомогою Excel

Хід роботи

1. Ввести дані в стовпці А, В, С.

2. У колонках D і E записати формули для обчислення рангів.

3. У комірці F3 записати формулу для обчислення різниці рангів.

Скопіювати цю формулу в комірки F4 : F16.

4. У комірку G3 записати формулу для зведення цієї різниці в квадрат.

Скопіювати її в комірки G4 : G16.

5. У комірку G17 підсумувати різниці квадратів.

6. У комірку B18 підрахувати число вимірів(об'єктів) за допомогою функції СЧЕТ (у нашому випадку це число чинників успіху - 14), а в комірку G18 - формулу для обчислення коефіцієнта рангової кореляції Спірмена:

$$=1-6*G17/(B18*(B18*B18-1)) \text{ (див. табл. 9.2).}$$

Таблиця 9.2

Введення формул для розрахунку коефіцієнта рангової кореляції Спірмена

	A	B	C	D	E	F	G
1	Чинники успіху	Відсоток фірм		Ранги		Різниця рангів	Різниця рангів в квадраті
2		японських	британських	японських	британських		
3	Хороша адаптованість до потреб	69,8	75,6	=РАНГ(B3;\$B\$3:\$B\$16)	=РАНГ(C3;C\$3:C\$16)	=D3 - E3	=F3*F3
4	Перевага над конкурентами за	79,3	59,3	=РАНГ(B4;\$B\$3:\$B\$16)	=РАНГ(C4;C\$3:C\$16)	=D4 - E4	=F4*F4

	якістю						
5	Перевага над конкурентами по дизайну	69,9	45,3	=РАНГ(B5;B\$3:\$B\$16)	=РАНГ(C5;C\$3:C\$16)	=D5 - E5	=F5*F5
6	Перевага над конкурентами по співвідношенню гідність/ціна	58,6	61,6	=РАНГ(B6;B\$3:\$B\$16)	=РАНГ(C6;C\$3:C\$16)	=D6 - E6	=F6*F6
7	Перевага над конкурентами по конструкції	55,2	48,8	=РАНГ(B7;B\$3:\$B\$16)	=РАНГ(C7;C\$3:C\$16)	=D7 - E7	=F7*F7
8	Дуже конкурентна ціна	41,4	27,9	=РАНГ(B8;B\$3:\$B\$16)	=РАНГ(C8;C\$3:C\$16)	=D8 - E8	=F8*F8
9	Адаптованість до можливостей фірми	39,7	34,9	=РАНГ(B9;B\$3:\$B\$16)	=РАНГ(C9;C\$3:C\$16)	=D9 - E9	=F9*F9
10	Унікальність	36,2	29,1	=РАНГ(B10;B\$3:\$B\$16)	=РАНГ(C10;C\$3:C\$16)	=D10 - E10	=F10*F10
11	Ефективний маркетинг	27,6	25,6	=РАНГ(B11;B\$3:\$B\$16)	=РАНГ(C11;C\$3:C\$16)	=D11 - E11	=F11*F11
12	Глибокий аналіз ринку	27,6	18,6	=РАНГ(B12;B\$3:\$B\$16)	=РАНГ(C12;C\$3:C\$16)	=D12 - E12	=F12*F12
13	Великий об'єм ринку	20,7	16,3	=РАНГ(B13;B\$3:\$B\$16)	=РАНГ(C13;C\$3:C\$16)	=D13 - E13	=F13*F13
14	Синергія виробництво/маркетинг	16,4	18,6	=РАНГ(B14;B\$3:\$B\$16)	=РАНГ(C14;C\$3:C\$16)	=D14 - E14	=F14*F14
15	Ухилення від ринків з високою конкуренцією і задоволеними покупцями	7,8	10,5	=РАНГ(B15;B\$3:\$B\$16)	=РАНГ(C15;C\$3:C\$16)	=D15 - E15	=F15*F15
16	Ухилення від динамічних ринків з частою зміною товару	2,6	4,7	=РАНГ(B16;B\$3:\$B\$16)	=РАНГ(C16;C\$3:C\$16)	=D16 - E16	=F16*F16
17							=СУМ(G3:G16)
18	Об'єктів	=СЧЕТ(B3:B17)			Коеф.Спірмена		=1-6*G17/(B18*(B18-B18-1))

У результаті обчислень вийде таблиця з результатами (див. таблицю. 9.3).

Таблиця 9.3

Результати обчислення коефіцієнта Спірмена

	A	B	C	D	E	F	G
1	Чинники успіху	Відсоток фірм		Ранги		Різниця рангів	Різниця рангів в квадраті
2		япон- ських	британ- ських	япон- ських	британ- ських		
3	Хороша адаптованість до потреб	69,8	75,6	3	1	2	4
4	Перевага над конкурентами за якістю	79,3	59,3	1	3	-2	4
5	Перевага над конкурентами по дизайну	69,9	45,3	2	5	-3	9
6	Перевага над конкурентами по співвідношенню гідність/ціна	58,6	61,6	4	2	2	4
7	Перевага над конкурентами по конструкції	55,2	48,8	5	4	1	1
8	Дуже конкурентна ціна	41,4	27,9	6	8	-2	4
9	Адаптированность до можливостей фірми	39,7	34,9	7	6	1	1
10	Унікальність	36,2	29,1	8	7	1	1
11	Ефективний маркетинг	27,6	25,6	9	9	0	0
12	Глибокий аналіз ринку	27,6	18,6	9	10	-1	1
13	Великий об'єм ринку	20,7	16,3	11	12	-1	1
14	Синергія виробництво/маркетинг	16,4	18,6	12	10	2	4
15	Ухилення від ринків з високою конкуренцією і задоволеними покупцями	7,8	10,5	13	13	0	0
16	Ухилення від	2,6	4,7	14	14	0	0

	динамічних ринків з частою зміною товару						
17							34
18	Об'єктів	14			Коеф.Спі рмена		0,925274725

Рішення задачі за допомогою SPSS для Windows

Хід роботи

1. Описати в редакторі даних 2 змінних з іменами japan, brit(японські і британські фірми).

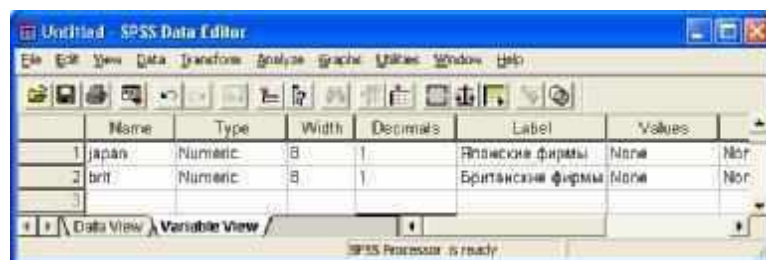


Рис. 9.1. Опис змінних в SPSS

Ввести дані згідно з таблицею, представленою на рис. 9.2.

	japan	brit	var
1	69,8	75,6	
2	79,3	59,3	
3	69,9	45,3	
4	58,6	61,6	
5	55,2	48,8	
6	41,4	27,9	
7	39,7	34,9	
8	36,2	29,1	
9	27,8	25,6	
10	27,8	18,6	
11	20,7	16,3	
12	16,4	18,6	
13	7,8	10,5	
14	2,6	4,7	
15			

Рис. 9.2. Введення даних в SPSS

2. Обчислити ранги для змінних japan, brit:

Меню: **Transform(Перетворити), Rank Cases.**(Ранги) (див. рис. 9.3)

У результаті з'являться 2 нові змінні:

From New

variable variable Label

JAPAN RJAPAN RANK of JAPAN

BRIT RBRIT RANK of BRIT

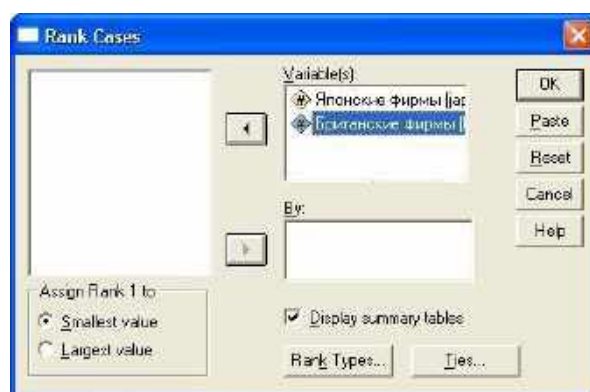


Рис. 9.3. Вікно «Rank Cases»

3. Обчислити коефіцієнти Спірмена і Кенделла для даних, представлених рангами :

Меню: **Analyze(Аналіз), Correlate(Кореляція), Bivariate(Двовимірна).**

У вікні **Bivariate Correlation** задати необхідні параметри (див. рис. 9.4).

Примітка: можна обчислювати коефіцієнти рангової кореляції без попереднього обчислення рангів, оскільки SPSS сам виконає необхідні розрахунки.



Рис. 9.4. Вікно «Bivariate Correlation»

4. Ознайомитися з отриманими результатами у вікні перегляду (рис. 9.5).

			Японські фірми	Британ- ські фірми
Kendall's tau _ b	Японські фірми	Correlation	1,000	,811(**)
		Coefficient		
		Sig. (2 - tailed)	.	,000
		N	14	14
	Британські фірми	Correlation	,811(**)	1,000
		Coefficient		
		Sig. (2 - tailed)	.	,000
		N	14	14
Spearman's rho	Японські фірми	Correlation	1,000	,928(**)
		Coefficient		
		Sig. (2 - tailed)	.	,000

N	14	14
Британські фірми	Correlation Coefficient	,928(**) 1,000
	Sig. (2-tailed)	-
		,000 .
N	14	14

** Correlation is significant at the 0.01 level(2 - tailed).

Рис. 9.5. Результати обчислення коефіцієнтів кореляції

Примітка: коефіцієнти, позначені * -- значимі на рівні 0,05;

** – на рівні 0,01.

Завдання для самостійного виконання

Керівництво компанії Marston Book Services (Великобританія), що займається наданням комплексу послуг з доставки споживачам друкарської продукції (книги, плакати, листівки і тому подібне), розробило план посилення орієнтації діяльності компанії на запити споживачів. На одному з етапів розробки цього плану було проведено вивчення думок співробітників компанії і споживачів відносно важливості окремих показників якості послуг, що надавалися. У таблиці 9.4 наводяться рейтинги окремих показників якості.

Таблиця 9.4

Рейтинг показників якості послуг

Показники якості послуг	Оцінки співробітників фірми	Оцінки споживачів
Швидкість обслуговування	1	7
Надійність	4	1
Неушкодженість упаковки замовлень	3	8
Задоволення термінових замовлень	8	3
Простота оформлення замовлень	7	6

Низький рівень рекламаций	2	4
Надання інформації по запитах	6	5
Витримка термінів виконання замовлень	5	2

Визначте, чи узгоджуються оцінки співробітників компанії і споживачів. Вирішіть завдання в середовищі Excel і в середовищі SPSS для Windows.

Вимоги до звіту

Звіт про роботу повинен містити:

- постановку завдання, файли з початковими даними і результатами, формули для обчислення, значення вчислених коефіцієнтів;
- відповіді на контрольні питання.

Контрольні запитання

1. Які коефіцієнти рангової кореляції ви знаєте?
2. Для чого призначений коефіцієнт кореляції Спирмена?
3. За якою формулою він обчислюється? Яких граничних значень набуває?
4. Як визначити його значущість? Визначте значущість коефіцієнта Спирмена в прикладі «японскиебританские фірми».
5. Як вчислити коефіцієнт Спирмена за допомогою електронних таблиць Excel?
6. Призначення функції «РАХУНОК». Для чого вона використовується в цій лабораторній роботі?
7. Призначення функції «РАНГ». Які параметри передаються цій функції? Для чого використовується у формулі для обчислення рангу абсолютна адресація?
8. Як перетворити дані в ранги? Які типи рангів є в SPSS для Windows?
9. Як вчислити коефіцієнти рангової кореляції в SPSS?
10. Як визначити значущість коефіцієнтів в SPSS?
11. Дайте змістовну інтерпретацію отриманим результатам.

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 10

Тема: *Обчислення коефіцієнтів кореляції та прогноз за допомогою лінійної регресії за допомогою Microsoft Excel*

Мета: навчитися використати електронні таблиці MS Excel для обчислення коефіцієнтів кореляції і прогнозування з використанням лінійної регресії.

Постановка завдання

У таблиці 10.1 наведені результати тестування 20-ти абітурієнтів на вступному іспиті в школу бізнесу (назвемо його «тест 1» - це ознака X) і результати тестування цих же осіб вже на випускному курсі («тест 2», ознака Y). Побудувати діаграму розсіювання для цих даних, вичислити коефіцієнт кореляції Пірсона-Браве і вивести рівняння регресії, що зв'язує ознаки X і Y.

Таблиця 10.1

Результати тестування в школі бізнесу

Тест 1(X)	Тест 2 (Y)
66	94
62	100
70	101
80	102
82	103
74	105
73	104
79	106
83	106
80	109
81	110
84	110
80	111

89	112
88	112
95	114
98	114
97	115
94	117
89	118

Хід роботи

1. Ввести початкові дані згідно з таблицею 10.1 (див. рис. 10.1).

	A	B	C	D	E	F	G
1	Результати тестування в школі бізнесу						
2	№ студента	тест 1	тест 2				
3	1	66	94				
4	2	62	100				
5	3	70	101				
6	4	80	102				
7	5	82	103				
8	6	74	105				
9	7	73	104				
10	8	79	106				
11	9	83	106				
12	10	80	109				
13	11	81	110				
14	12	84	110				
15	13	80	111				
16	14	89	112				
17	15	88	112				

18	16	95	114				
19	17	98	114				
20	18	97	115				
21	19	94	117				
22	20	89	118				

Рис. 10.1. Введення результатів тестування

2. Вичислити коефіцієнт кореляції Пірсона-Браве :

а) меню *Дані, Аналіз даних, Кореляція* (див. рис. 10.2);

б) в діалоговому вікні «*Кореляція*» (рис. 10.3) вивести і записати в зошит довідку;

в) задати параметри: вхідний інтервал (обидва ряди даних), угруповання по стовпцях. Натиснути ОК. Записати в зошит отриманий коефіцієнт кореляції. Визначити його значущість по таблиці.

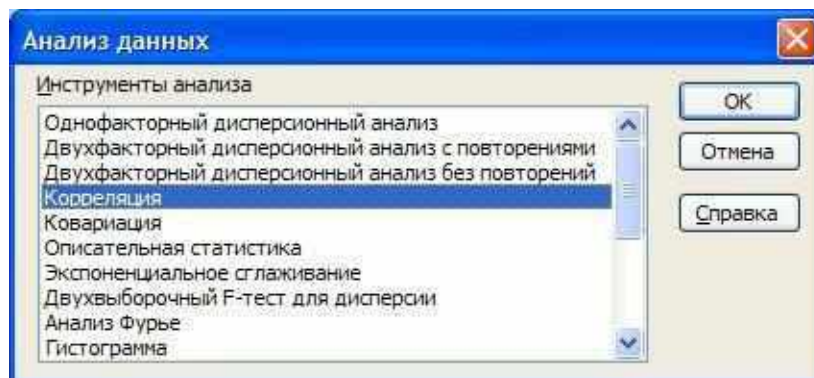


Рис. 10.2. Діалогове вікно «Аналіз даних»

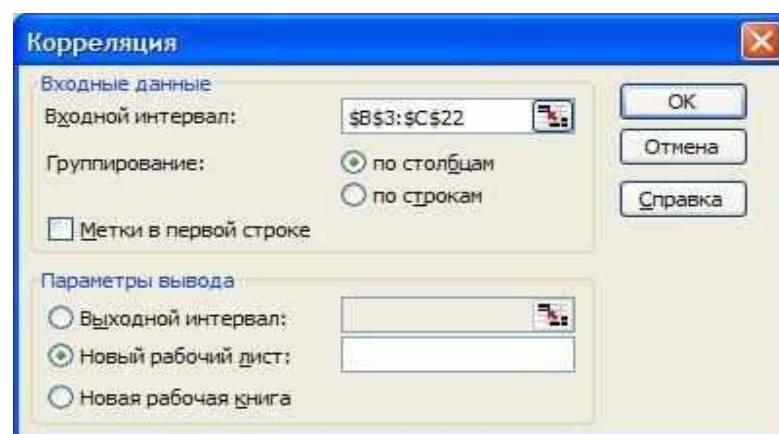


Рис. 10.3. Діалогове вікно «Кореляция»

5. Побудувати діаграму розсіювання за допомогою *Майстра діаграм* (тип «точкова»). Зробити припущення про наявність або відсутність кореляції між X і Y за зовнішнім виглядом діаграми.

6. Побудувати лінію тренда по методу найменших квадратів (лінійна регресія):

- а) клацнути правою кнопкою по точках діаграми;
- б) в контекстному меню вибрати «Додати лінію тренду»;
- в) в діалоговому вікні «Лінія тренду» на вкладці «Тип» (рис. 10.4) *вказіть* - «лінійна», на вкладці «Параметри» - *поставте прапорець у опції* «Показувати рівняння на діаграмі» і «Помістити величину достовірності апроксимації» (див. рис. 10.5).

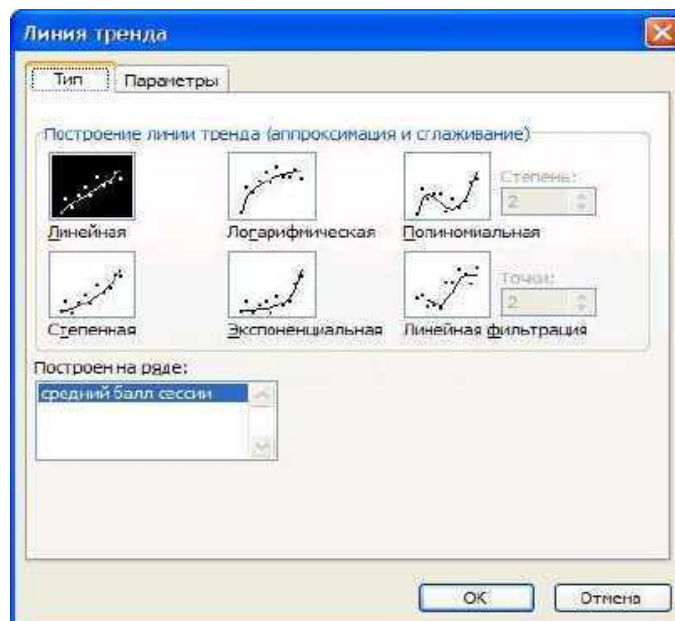


Рис. 10.4. Діалогове вікно «Лінія тренду», вкладка «Тип»

У результаті повинна вийти точкова діаграма з лінією тренду, рівнянням регресії і коефіцієнтом апроксимації, як це показано на рис. 10.6.

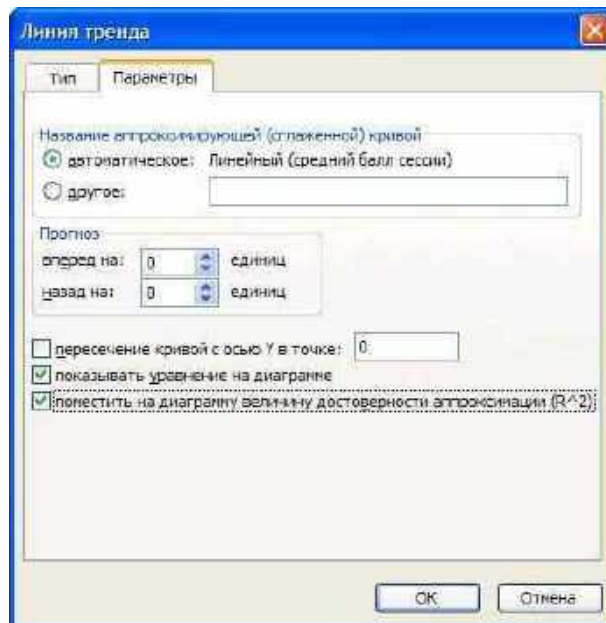


Рис. 10.5. Діалогове вікно «Лінія тренду», вкладка «Параметри»

Рівняння треба записати в зошит. І, користуючись MS Excel, передбачити по ньому значення випускного тіста для осіб, у яких за результатами вступного тіста були наступні бали :

- а) 43;
- б) 75;
- в) 99.

5. Зберегти таблицю в особистій теці.

Діаграма розсіювання і лінія тренду

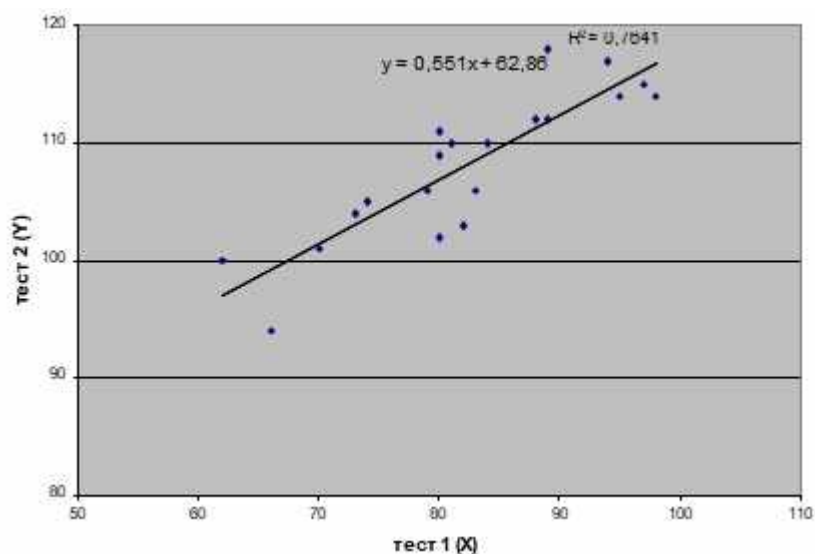


Рис. 10.6. Діаграма розсіювання, лінія тренду і рівняння регресії

Завдання для самостійного виконання

1. У таблиці приведені коефіцієнти інтелекту IQ батьків і дітей десяти сімей. Визначте, чи існує кореляційна залежність між цими ознаками.

IQ батьків	IQ дітей
125	110
120	105
110	95
105	125
105	120
95	105
95	75
90	95
80	90
75	80

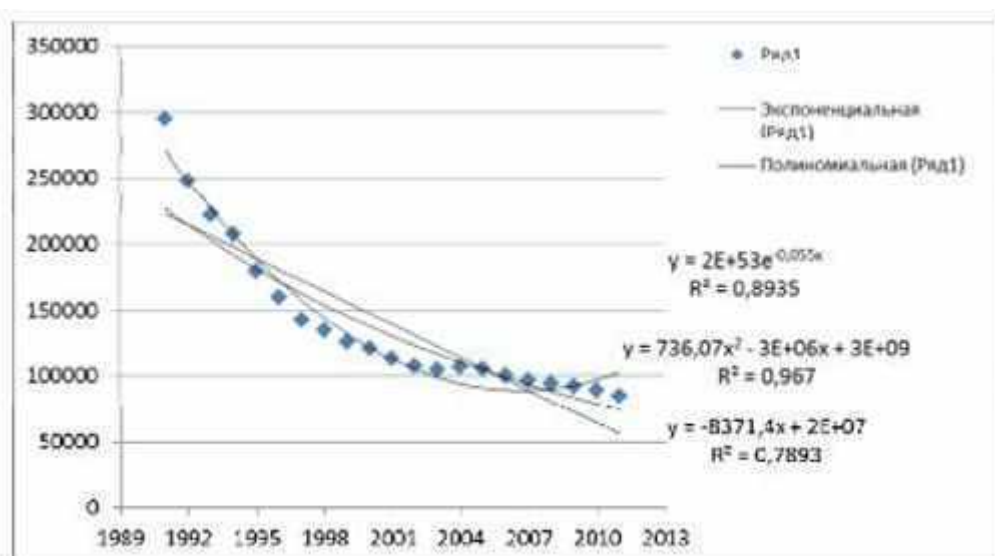
2. Побудуйте модель лінійної регресії між роком і чисельністю науковців в Україні

Наукові кадри і кількість організацій України

Рік	Кількість організацій, які ведуть наукові дослідження і розробки	Чисельність наукових накопичень, чол.	Чисельність докторів наук в економіці України, чол.	Чисельність кандидатів наук в економіці України, чол.
1991	1344	295010	8133	...
1992	1350	248455	8797	...
1993	1406	222127	9224	...
1994	1463	207436	9441	...
1995	1453	179799	9759	57610
1996	1435	160103	9974	58132
1997	1450	142532	10322	59332

1998	1518	134413	10446	59703
1999	1506	126045	10233	59547
2000	1490	120773	10339	58741
2001	1479	113341	10603	60647
2002	1477	107447	11008	62673
2003	1487	104841	11259	64372
2004	1505	106603	11573	65839
2005	1510	105512	12014	68291
2006	1452	100245	12488	71893
2007	1404	96820	12845	74191
2008	1378	94138	13423	77763
2009	1340	92403	13866	81169
2010	1303	89534	14418	84000
2011	1255	84969	14895	84979
2012	1208			

Ці дані можна знайти на сайті Держкомстату України: <http://www.ukrstat.gov.ua/>. Побудуйте також експоненціальну і квадратичну модель. Порівняйте їх (рис. 10.7)



Вимоги до звіту

Звіт про роботу повинен містити:

- постановку завдання, початкові дані, коефіцієнт кореляції, рівняння лінії пророцтва, результати пророцтва;
- файл з результатами.

Контрольні запитання:

1. У чому полягає завдання регресійного аналізу?
2. Що є діаграма розсіювання?
3. Як отримати рівняння лінії пророцтва?
4. Як здійснюються прогнози за допомогою рівняння лінії пророцтва(лінії тренду)?
5. Як вичислити коефіцієнт кореляції Пірсона-Браве?
6. Що показує величина достовірності апроксимації?

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 11

Тема: *Обчислення коефіцієнтів кореляції і прогноз за допомогою лінійної регресії в SPSS для Windows*

Мета: навчитися використати SPSS для обчислення коефіцієнтів кореляції і прогнозування з використанням лінійної регресії.

Постановка завдання

Керівництво швацької фабрики хотіло б мати інформацію про те, скільки суконь кожного фасону кольору і розміру буде продано. З одного боку, не можна допускати надвиробництва, оскільки в цьому випадку доведеться продавати товар за заниженими цінами, з іншого боку, нестача призводить до необхідності випускати більше виробів, ніж було намічене до цього терміну. Якщо мати прогноз через 5 тижнів після надходження першого найменування у продаж, можна забезпечити виготовлення необхідного числа продукції до

наміченого терміну. У таблиці 11.1 приведені дані про сумарний об'єм проданих виробів кожного фасону, кольору і розміру і відповідні цифри за п'ять тижнів, отриманих на початку дослідження. Треба представити графічно дані і отримати оцінки найменших квадратів в моделі $y=b_0+b_1x$.

Таблиця 11.1.

Продано в перші п'ять тижнів	Продано усього
235	392
122	190
34	74
196	307
121	200
127	185
177	291
125	191
159	314
75	140
140	235
127	188
89	142
139	253
70	96
62	96

Рішення за допомогою SPSS

Хід роботи

1. Описати в редакторіві дві змінні «Продано всього» і «Продано в перші п'ять тижнів» і ввести дані таблиці 11.1.
2. Вибрати в меню *Analyze... (Аналіз) Regression...(Регресія) Linear... (Лінійна)*. З'явиться діалогове вікно Linear Regression(Лінійна регресія) (рис. 11.1).
3. Перенести змінну «Продано Усього» в полі для залежних змінних і присвоїти змінній «Продано в перші 5 тижнів» статус незалежної змінної. Почати розрахунок натисненням ОК.

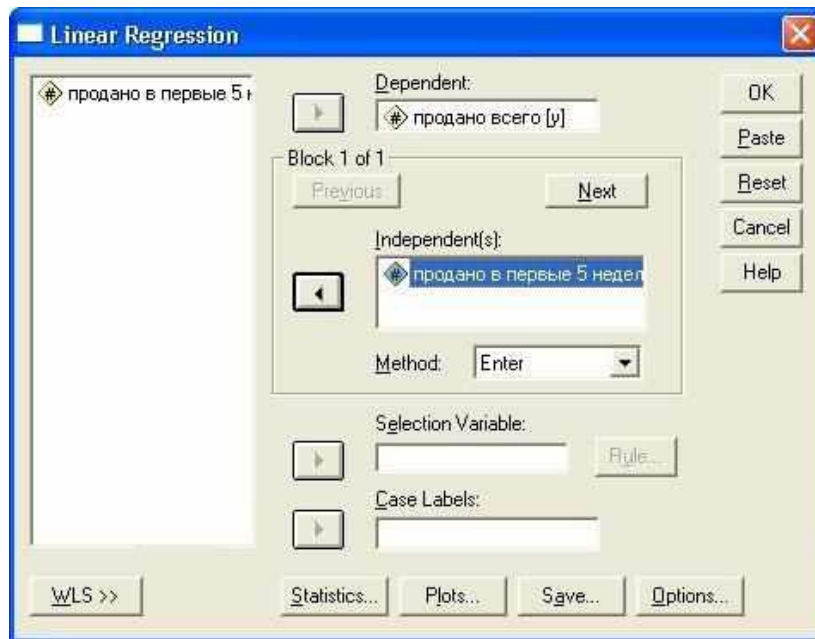


Рис. 11.1. Діалогове вікно «Linear Regression»

4. Виведення основних результатів виглядає так, як представлено на рис. 11.2 і 11.3.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,974 ^a	,949	,946	20,666

a. Predictors:(Constant), продано в перші 5 тижнів

Рис. 11.2. Виведення результатів. Сумарна модель

Сумарна модель регресії (рис. 11.2) включає: значення коефіцієнта кореляції Пірсона $R=0,974$; значення коефіцієнта детерміації $R^2=0,949$; значення уточненого коефіцієнта детерміації і стандартну помилку оцінки.

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	-1,827	13,821	,974	-,132	,897

продано в перші 5 тижнів	1,663	,103		16,202	,000
-----------------------------	-------	------	--	--------	------

а. Dependent Variable : продано усього

Рис. 11.3. Виведення результатів

Коефіцієнти рівняння регресії

У таблиці «Коефіцієнти» (див. рис. 11.3) вікна виведення наводяться значення коефіцієнтів рівняння регресії, їх стандартної помилки, стандартизовані коефіцієнти, а також значення статистики Стюдента t і її значущість Sig.

Коефіцієнт є значимим, якщо $\text{Sig} < 0,05$.

Отримане рівняння регресії : $y = 1,663x - 1,827$.

5. Побудувати лінію регресії. Для цього необхідно виконати наступні кроки:

- Побудувати діаграму розсіювання.
- Клацнути двічі на цьому графіку, щоб перенести його в редактор діаграм.
- Вибрати в редакторові діаграм меню **Chart... (Діаграма) Options... (Опції)**
- Відкриється діалогове вікно **Scatterplot Options** (Опції для діаграми розсіювання) (див. рис. 11.4).

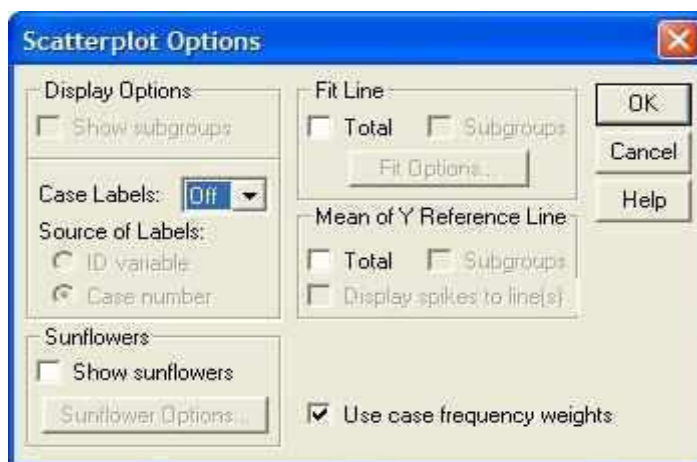


Рис. 11.4. Діалогове вікно «Scatterplot Options» (Опції для діаграми розсіювання)

– В рубриці **Fit Line** (Наближена крива) поставити прапорець навпроти опції **Total** (Цілком для усього файлу даних) і клацнути на кнопці **Fit Options** (Опції для наближення). Відкриється діалогове вікно **Scatterplot Options : Fit Line** (Опції для діаграми розсіювання : наближена крива) (див. рис. 11.5).

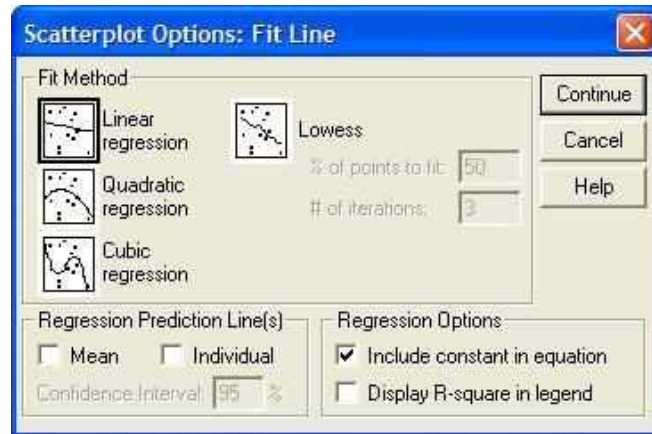


Рис. 11.5. Діалогове вікно «Scatterplot Options : Fit Line»

– Підтвердити попередню установку **Linear Regression** (Лінійна регресія) клацанням **Continue** (Далі) і потім на ОК.

– Закрити редактор діаграм і клацнути один раз де-небудь поза графіком. У вікні перегляду з'явиться діаграма розсіювання з лінією регресії (див. рис. 11.6).

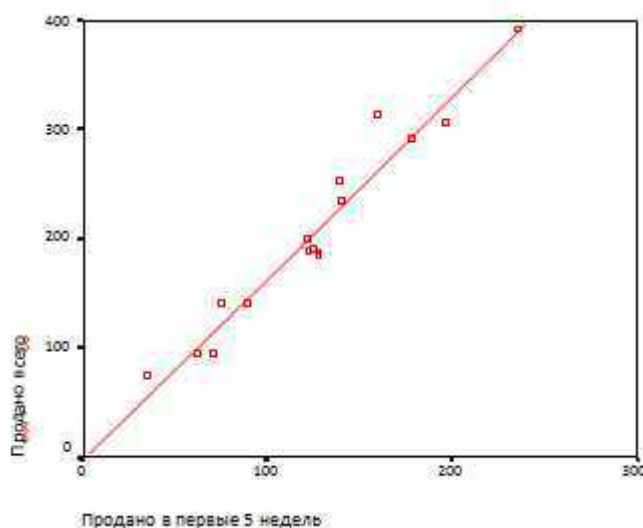


Рис. 11.6. Діаграма розсіювання з лінією регресії

Вимоги до звіту

Звіт про роботу повинен містити:

- постановку завдання, початкові дані, коефіцієнт кореляції, рівняння лінії пророцтва, результати пророцтва;
- файл з результатами.

Контрольні запитання

1. Яким чином ви описали змінну «Продано в перші 5 тижнів» в редакторів даних?
2. Описати елементи діалогового вікна «Scatterplot Options».
3. Описати елементи діалогового вікна «Scatterplot Options : Fit Line».
4. Як отримати рівняння лінії регресії в SPSS? Дайте інтерпретацію отриманим результатам.

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 12

Тема: *Перевірка гіпотез про значущість відмінностей середніх у Microsoft Excel*

Мета: навчитися використати пакет статистичного аналізу MS Excel для перевірки статистичних гіпотез.

Постановка завдання

Фірма найняла 40 чоловік для поширення своєї продукції і організувала їх навчання, розділивши на дві групи по 20 чоловік. У програму навчання першої групи входило знайомство з особливостями продукції, яку учасникам групи належало поширювати, в програму навчання другої групи додатково були введені основи психології. У таблиці 12.1 приведено число продажів, які здійснили представники обох груп в перший місяць роботи. Чи свідчать представлені в таблиці 12.1. дані про відмінність в підготовленості до роботи розповсюджувачів, що навчалися по різних програмах?

Таблиця 12.1

Число продажів залежно від виду попереднього навчання

Число продажів	
Група 1	Група 2
16	9
17	7
5	5
7	6
7	7
8	8
8	4
6	6
6	6
8	8
8	8
15	12
8	4
10	4
10	10
13	5
4	4
4	3
8	8
9	3

Хід роботи

1. Завантажити Excel. Ввести початкові дані згідно з таблицею 12.1.
2. Для кожного стовпця даних вичислити середнє і дисперсію за допомогою функцій **СРЗНАЧ** і **ДИСП** подивитися, чи рівні в обох групах вийшли середні і дисперсії.
3. Запустити пакет аналізу: меню «Дані», «Аналіз даних», «Двохвибірковий *t*- тест з різними дисперсіями» (див. рис. 12.1). Прочитати про нього довідку і записати її в зошит.

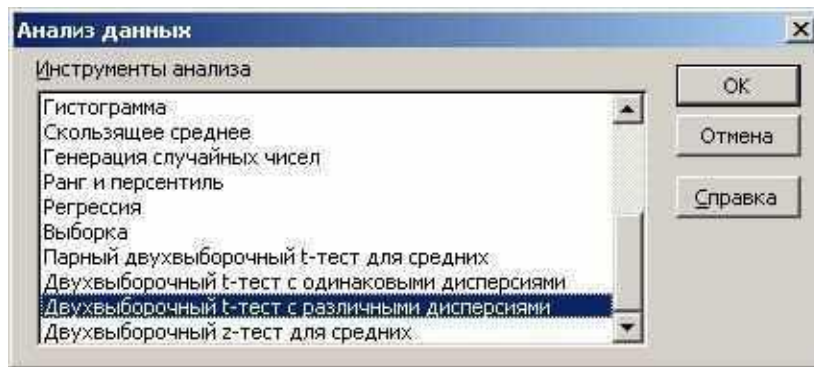


Рис. 12.1. Діалогове вікно «Аналіз даних»

4. У вікні «*Двохвибірковий t- тест*» задати необхідні параметри: інтервал змінної 1, інтервал змінної 2, гіпотетична різниця середніх - 0, рівень значущості 0,05. Натиснути кнопку *ОК* (рис. 12.2).

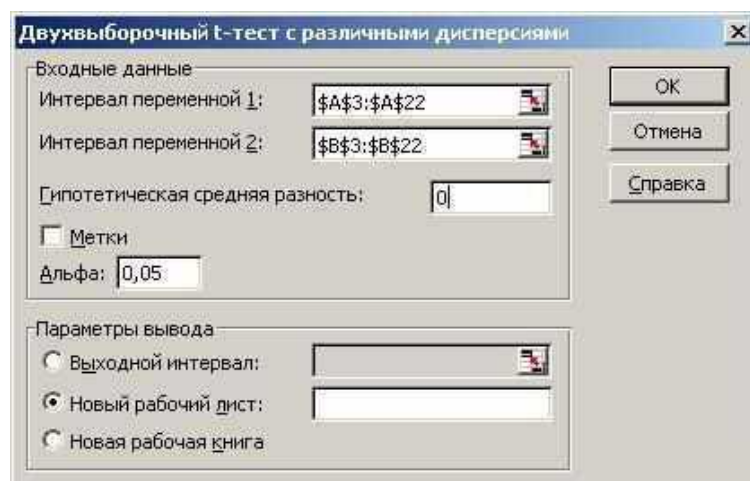


Рис. 12.2. Діалогове вікно «Двохвибірковий t- тест»

5. Виписати таблицю з отриманими результатами (див. рис. 12.3) і зробити висновок відносно значущості відмінностей.

6. Зберегти таблицю в особистій теці.

Вимоги до звіту

Звіт про роботу повинен містити:

- постановку завдання, початкові дані, опис двухвибіркового t- тесту з різними дисперсіями;
- файл з результатами, висновки про істинність або хибність гіпотези.

Двохвибірковий t- тест з різними дисперсіями			
	Змінна 1		Змінна 2
Середнє	8,85		6,35
Дисперсія	13,92368421		5,923684211
Спостереження	20		20
Гіпотетична різниця			
середніх	0	df	33
t- статистика			
2,509594424			
P(T<=t) одностороннє	0,008586329		t критичне
одностороннє	1,692360258	P(T<=t) двостороннє	
	0,017172658	t критичне	двостороннє
	2,034515287		

Рис. 12.3. Результати розрахунку двухвибіркового t- тесту з різними дисперсіями

Контрольні запитання

1. Назвіть види статистичних гіпотез. Наведіть приклади. Сформулюйте нульову і альтернативну гіпотезу для завдання.
2. Охарактеризуйте поняття «Статистичний критерій». У чому полягає відмінність між параметричними і непараметричними критеріями?
3. Які види статистичних критеріїв реалізовані в Excel? Як вони здійснюються?

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 13

Тема: *Перевірка гіпотез про значущість відмінностей дисперсій у MS Excel*

Мета: навчитися використати пакет статистичного аналізу MS Excel для перевірки статистичних гіпотез.

Постановка завдання

В одному з досліджень дітям давалися звичайні арифметичні задачі, а потім одній випадково вибраній половині учнів повідомлялося, що вони не пройшли випробування, а іншим - зворотне. Потім у кожної дитини запитували, скільки секунд йому знадобилося б для вирішення нового завдання. Експериментатор обчислював різницю між очікуваним часом рішення задачі (у сек.), яке назвала дитина, і результатами раніше виконаного завдання.

Гіпотеза полягає в тому, що перевіряється на рівні значущості 0,05 дисперсія сукупності дитячих оцінок постійна незалежно від того, чи повідомлялося дітям про погані результати випробування або ні. Результати експерименту приведені в таблиці 13.1. Групі 1 повідомлялося про позитивний результат, групі 2 - про невдачу. У кожній групі було по 12 дітей.

Таблиця 13.1

Результати експерименту

Різниця між очікуваним і реальним часом рішення задачі, сек.	
Група 1	Група 2
8	40
9	10
10	30
11	20
7	38
6	22
5	30
9	30
12	45
15	26
9	25
6	25

Хід роботи

1. Ввести початкові дані згідно з таблицею 13.1.
2. Для кожного стовпця даних вичислити дисперсію за допомогою функції ДИСП. Подивитися, чи рівні вийшли дисперсії в двох групах.

3. Запустити пакет аналізу: меню «Дані», «Аналіз даних», «Двохвибірковий F- тест для дисперсій» (див. рис. 13.1).

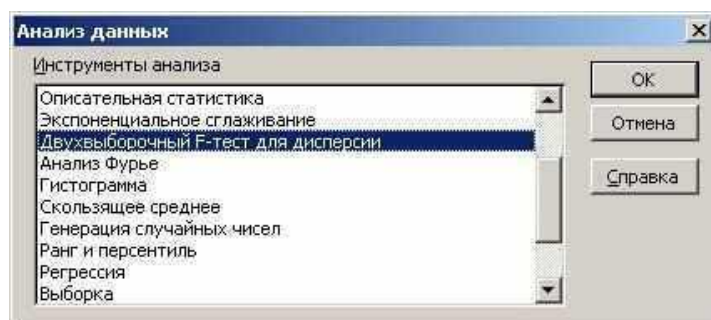


Рис. 13.1. Вікно «Двохвибірковий F- тест для дисперсій»

Прочитати довідку і записати її в зошит.

4. У вікні «Двохвибірковий F- тест» задати необхідні параметри.
5. Виписати таблицю з отриманими результатами (рис. 13.2).

Двохвибірковий F- тест для дисперсії			
	Група 1	Група 2	
Середнє	8,916666667	28,41666667	
Дисперсія	8,083333333	89,90151515	
Спостереження	12		12
Df	11		11
F	0,089913205		
F критичне	0,354869911		
одностороннє			

Рис. 13.2. Результати розрахунку двухвибіркового F- тесту з різними дисперсіями

6. Зробити висновок про підтвердження або спростування гіпотези про рівність дисперсій. (Вказівка: оскільки значення F статистики 0,089 менше нижнього критичного значення 0,354, тобто потрапляє в критичну область, то гіпотеза про рівність дисперсій двох груп даних відхиляється на рівні значущості 0.05)

7. Зберегти таблицю в особистій теці.

Вимоги до звіту

Звіт про роботу повинен містити:

- постановку завдання, початкові дані, опис двохвибіркового F- тесту, результати;
- висновки про істинність або хибність гіпотези;
- файл з результатами.

Контрольні запитання

1. Наведіть приклади відомих вам параметричних і непараметричних статистичних критеріїв.
2. Сформулюйте нульову і альтернативну гіпотезу для завдання з цієї лабораторної роботи.
3. Які види статистичних критеріїв реалізовані в Excel?

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 14

Тема: *Перевірка гіпотез про рівність середніх двох незалежних вибірок в SPSS*

Мета: навчитися робити перевірку гіпотез про рівність середніх двох незалежних вибірок з допомогою Excel і SPSS і інтерпретувати результати.

Постановка завдання

50 студентів університету були випадково розподілені по двох групах. 25 студентів вивчали курс аналізу даних за традиційною методикою ВНЗ, інші 25 - дистанційно. У кінці експерименту був проведений тест на засвоєння знань. Результати його наведені в таблиці 14.1. Чи свідчать результати тесту про відмінність у методиці викладання теми?

Таблиця 14.1

Тест в групі 1	Тест в групі 2
6	6

7	7
5	5
7	7
7	7
8	8
8	4
6	6
6	6
8	8
8	8
15	12
8	4
10	4
10	10
13	5
4	4
4	3
8	8
9	3
8	8
8,25	3
4	3
6	3
8	8

Нульова гіпотеза: Середні бали за результатами тесту в двох групах не відрізняються.

Рішення за допомогою Excel

1. Завантажити Excel. Ввести початкові дані згідно з таблицею 14.1.
2. Для кожного стовпця даних вичислити середнє і дисперсію

допомогою функцій СРЗНАЧ і ДИСП. Подивитися чи рівні вийшли дисперсії.

3. Запустити пакет аналізу: меню «Дані», «Аналіз даних», «Двохвибірковий тест з різними дисперсіями».

Прочитати в ньому довідку і записати її в зошит.

4. У вікні «Двохвибірковий тест...» задати необхідні параметри: інтервал змінної 1 (дані першої групи), інтервал змінної 2 (дані другої групи), гіпотетична різниця середніх - 0, рівень значущості 0,05. Натиснути кнопку ОК.

5. Виписати таблицю з отриманими результатами (див. рис. 14.1)

Двохвибірковий t- тест з різними дисперсіями		
	Змінна 1	Змінна 2
Середнє	7,65	6
Дисперсія	6,5	5,916666667
Спостереження	25	25
Гіпотетична різниця середніх	0	
Df	48	
t- статистика	2,341269661	
t критичне двостороннє	2,01063358	

Рис. 14.1. Таблиця з результатами

6. Порівняйте емпіричне значення t- критерію Стьюдента (t- статистику) з критичним двостороннім.

7. Зробити висновок про прийняття або відкидання нульової гіпотези (якщо емпіричне значення більше за критичне, то нульова гіпотеза відкидається).

8. Зберегти таблицю в особистій теці.

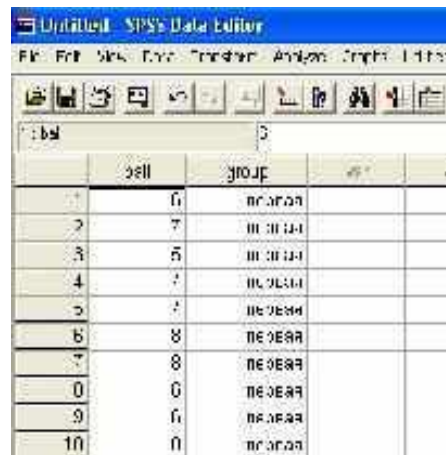
Рішення за допомогою SPSS

Перевірка гіпотези за допомогою SPSS для Windows містить наступні кроки:

1. Завдання двох змінних: залежній змінній **ball** (бал по тесту), незалежною змінної-фактора **group** (група) з градаціями: «1 - перша», «2 -

друга»).

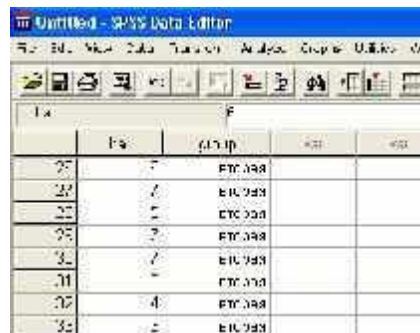
2. Введення початкових даних (представлення даних в SPSS відрізняється від представлення в електронних таблицях) і наводиться на рис. 14.2.



The screenshot shows the 'Untitled - SPSS Data Editor' window. The data is organized into two columns: 'ball' and 'group'. The 'ball' column contains values from 1 to 10, and the 'group' column contains categorical labels: 'перша' for balls 1-5, 'друга' for balls 6-8, and 'третя' for balls 9-10.

	ball	group
1	1	перша
2	2	перша
3	3	перша
4	4	перша
5	5	перша
6	6	друга
7	7	друга
8	8	друга
9	9	третя
10	10	третя

...



This screenshot shows another view of the SPSS Data Editor. It displays a dataset with columns 'ball' and 'group'. The 'ball' column has values 1 through 8, and the 'group' column has labels 'перша', 'друга', and 'третя'.

	ball	group
1	1	перша
2	2	перша
3	3	перша
4	4	перша
5	5	перша
6	6	друга
7	7	друга
8	8	друга

Рис. 14.2. Представлення початкових даних в SPSS для Windows

3. Вибір меню *Analyze, Compare means, Independent-Samples T Test* (для незалежних вибірок)

4. Перенесення змінної **бал по тесту** в список тестованих змінних, а змінну **група** - в групуючу змінну (див. рис. 14.3)

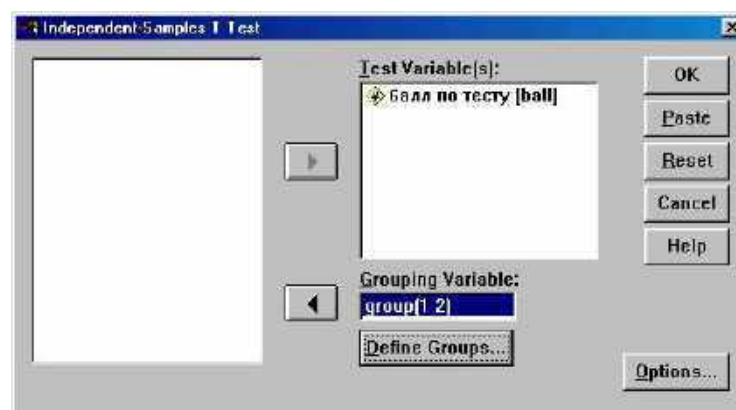


Рис. 14.3. Діалогове вікно Independent Samples T - Test

Клацанням на кнопці «Визначити групи» задається кодування, показана на рис. 14.4.

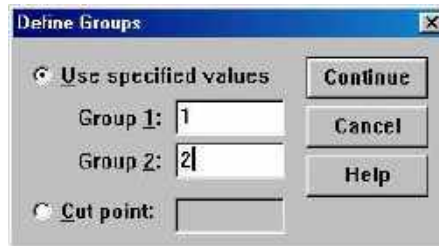


Рис. 14.4. Діалогове вікно «Define Groups»

5. Перегляньте результати у вікні виведення (рис. 14.5).

Group Statistics				
група	N	Mean	Std. Deviation	Std. Error Mean
балл по тесту первая	25	7,6500	2,5495	,5099
вторая	25	6,0000	2,4324	,4865

Independent Samples Test									
	Levene's Test of Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. 2-tailed	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
балл по тесту	Equal variances assumed	,261	,612	2,341	48	,023	1,6500	,7047	,2330 ,0670
	Equal variances not assumed			2,341	7,894	,023	1,6500	,7047	,2329 ,0671

Рис. 14.5. Результати порівняння середніх двох незалежних вибірок

Для нашого прикладу значення критерію Стюдента $t_{\text{емпирическое}}=2,341$ і значущість $p=0,023$. Нульова гіпотеза відхиляється, якщо значущість менше 0,05.

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 15

Тема: Перевірка гіпотез про рівність середніх двох залежних вибірок в SPSS

Мета: навчитися робити перевірку гіпотез про рівність середніх двох залежних вибірок в Excel і SPSS і інтерпретувати результати.

Постановка завдання

Визначте за даними наступної таблиці, чи значущо розрізняються середні процентних доль працюючих жінок в 1968 і 1972 роках в 19 містах США.

Таблиця 15.1.

Відсоток працюючих жінок в 1968 і 1972 роках в 19 містах США

Місто	1968	1972
N, Y	0,42	0,45
L, A	0,5	0,5
Chicago	0,52	0,52
Philadelphia	0,45	0,45
Detroit	0,43	0,46
San Francisco	0,55	0,55
Boston	0,45	0,6
Pitt	0,34	0,49
St, Louis	0,45	0,35
Connecticut	0,54	0,55
Wash,, D, C	0,42	0,52
Cinn	0,51	0,53
Baltimore	0,49	0,57
Newark	0,54	0,53
Minn/St, Paul	0,5	0,59
Buffalo	0,58	0,64
Houston	0,49	0,5
Patterson	0,56	0,57
Dallas	0,63	0,64

Нульова гіпотеза: Середні процентних доль працюючих жінок в 1968 і 1972 роках в 19 містах США значущо не розрізняються.

Альтернативна гіпотеза: Середні процентних доль працюючих жінок в 1968 і 1972 роках в 19 містах США значущо розрізняються.

Рішення задачі за допомогою електронних таблиць Excel

1. Завантажити Excel. Ввести початкові дані згідно з таблицею 15.1.

2. Для кожного стовпця даних вичислити середнє і дисперсію допомогою функцій СРЗНАЧ і ДИСП.

3. Запустити пакет аналізу: меню «Дані», «Аналіз даних», «Парний двохвибірковий t - тест для середніх».

Прочитати в ньому довідку і записати її в зошит.

4. У вікні «Парний двохвибірковий t – тест...» задати необхідні параметри: інтервал змінної 1, інтервал змінної 2, гіпотетична різниця середніх - 0, рівень значущості 0.05. Натиснути кнопку ОК (див. рис. 15.1).

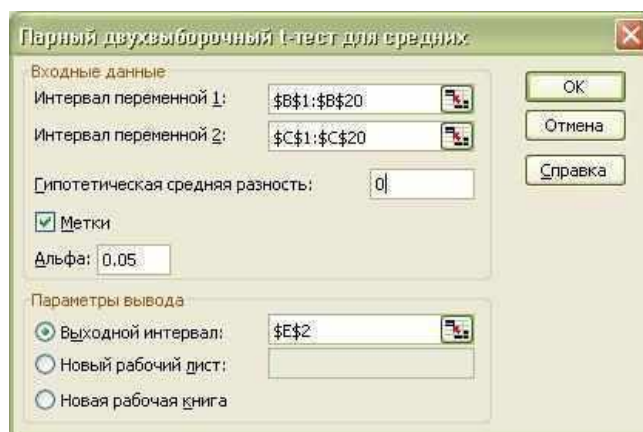


Рис. 15.1. Вікно «Парний двохвибірковий t - тест для середніх»

5. Виписати таблицю з отриманими результатами (див. таблицю 15.2).

Таблиця 15.2.

Результати парного двохвибіркового t - тесту для середніх

Парний двохвибірковий t - тест для середніх		
	1968	1972
Середнє	0.493158	0.526842105
Дисперсія	0.004623	0.005011696
Спостереження	19	19
Кореляція Пірсона	0.630073	
Гіпотетична різниця середніх	0	

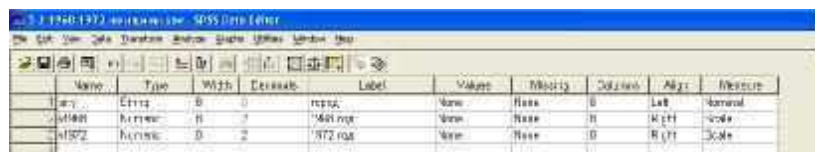
df	18	
t- статистика	-2.4577	
P(T<=t) одностороннє	0.012176	
t критичне одностороннє	1.734064	
P(T<=t) двостороннє	0.024353	
t критичне двостороннє	2.100922	

6. Порівняти емпіричне значення t-статистики Стьюдента з критичним (двостороннім). Зробити висновок про прийняття або відкидання нульової гіпотези.

7. Зберегти таблицю в особистій теці.

Рішення за допомогою SPSS

1. Опишіть три змінні, як показано на рис. 15.2, заповніть даними таблиці 15.1 і збережете у файлі жінки.sav.



	city	x1941	x1942	city	city
1	N.Y.	.42	.45		
2	L.A.	.50	.57		
3	Chicago	.52	.52		
4	Philade.	.45	.45		
5	Detroit	.43	.42		
6	San Francisco	.55	.55		
7	Denver	.45	.51		
8	Phila.	.54	.45		
9	St. Louis	.40	.51		
10	Connect.	.54	.51		
11	Wash., D.C.	.42	.52		
12	Cinn.	.5	.51		
13	Dallas	.49	.52		
14	New York	.54	.51		
15	Minneapolis	.50	.51		
16	DuPont	.50	.54		
17	Houston	.49	.51		
18	Pittsburgh	.50	.57		
19	Boston	.50	.54		

Рис. 15.2. Зміст вкладки «Variable View» і «Data View» для завдання про процентну долю працюючих жінок

2. Виберіть в меню *Analyze, Compare means, Paired Samples T - Test* (тест для парних вибірок).

3. Тепер в полі тестованих змінних треба виділити дві необхідні змінні і цю пару перенести в полі для спарених змінних. У нашому прикладі такими змінними є v1968 і v1972 (рис. 15.3.)

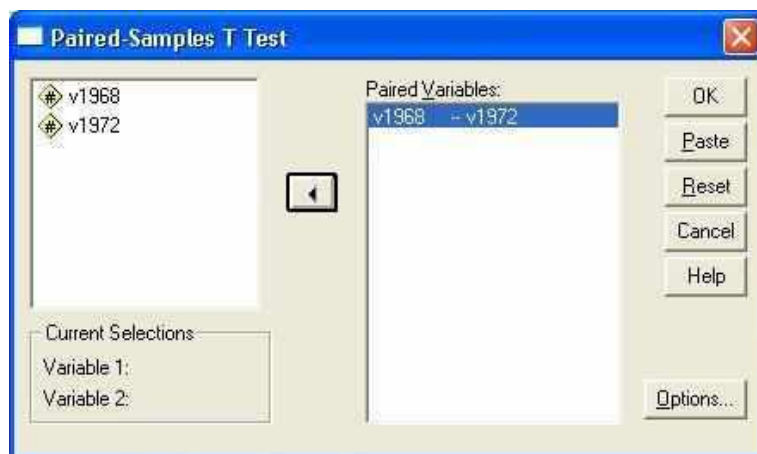


Рис. 15.3. Діалогове вікно «Paired - Samples T Test»

4. Запустіть тест на виконання натисненням клавіші ОК. У вікні перегляду з'являться результати розрахунку (див. рис. 15.4).

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	V1968	,4932	19	,06799	,01560
	V1972	,5268	19	,07079	,01624

Paired Samples Test								
		Paired Differences				t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference			
					Lower	Upper		
Pair 1	V1968 - V1972	-,0337	,05974	,01371	-,0625	-,0049	-2,458	,024

Рис. 15.4. Результати розрахунків парного двохвибіркового тіста для середніх

Зверніть увагу на емпіричне значення статистики Стюдента - 2,458 і її значущість - 0,024. Нульова гіпотеза відкидається при значущості $p < 0,05$.

Отримана в наведеному прикладі величина $p = 0,024$ свідчить про

значиму різницю в середніх.

Завдання для самостійного виконання

1. За результатами оцінки розмірів дебіторської заборгованості, проведеної фінансовим директором фірми, були внесені зміни в її кредитну політику. Після закінчення звітного кварталу було вирішено проаналізувати, чи дійсно зміни в кредитній політиці фірми вплинули на скорочення терміну оплати дебіторської заборгованості. У таблиці 15.3 наведені дані про тривалість збору дебіторської заборгованості (у днях) за старої та нової кредитної політики фірми.

Таблиця 15.3.

Початкові дані

Тип кредитної політики	Тривалість збору дебіторської заборгованості (дні)
Стара	39 42 28 35 38 30 30 29 36 34
Нова	28 31 29 30 24 37 27 22 33 31

2. Майерс у 1976 р. провів цікавий експеримент по перевірці ефективності індивідуального і групового тестування. Випробовувані однієї групи перевіряли програму з 63 операторів, призначену для обробки рядків, індивідуально, причому використали набір тестів, складених після вивчення специфікації програми. Вони не мали в розпорядженні лістингу програми. Випробовувані другої групи мали у розпорядженні ті ж кошти плюс лістинг. Випробовувані третьої групи були розбиті на бригади з трьох чоловік, і від них вимагалось тестувати програми вручну методом перевірки коду. Результати експерименту приведені в таблиці 15.4.

Результати експерименту по тестуванню програм

Характеристика	Індивідуальний перегляд + специфікація + термінал	Індивідуальний перегляд + специфікація + термінал + лістинг	Груповий перегляд + специфікації + лістинг
Середнє число знайдених помилку	4,5	5,4	5,7
Дисперсія	4,8	5,5	3,0
Мінімальне число помилку	1	2	3
Максимальне число помилок	7	9	9
Витрати на помилку, люд.-хв	37	29	75

- Визначте за критерієм Стюдента чи значущо відрізняються середнє число знайдених помилок у цих трьох групах, якщо в кожній з груп було по 25 випробовуваних.
- Визначте, чи значущо відрізняються дисперсії і середні витрати на помилку трьох груп?

Вказівка. При порівнянні середніх двох незалежних вибірок емпіричне значення критерію Стюдента обчислюється за формулою:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ де}$$

X1	Середнє першої вибірки
X2	Середнє другої вибірки
S1	Стандартне відхилення першої вибірки

S2	Стандартне відхилення другої вибірки
n1	Об'єм першої вибірки
n2	Об'єм другої вибірки

і порівнюється з критичним $1 - (\alpha / 2) t_{n_1+n_2-2}$

Якщо емпіричне значення перевищує критичне, то нульова гіпотеза про рівність середніх двох вибірок відкидається на рівні значущості α .

Вимоги до звіту

Звіт про роботу має містити:

- постановку завдання,
- початкові дані,
- файли з результатами,
- висновки про істинність або хибність гіпотези.

Контрольні запитання

1. Назвіть види статистичних гіпотез. Наведіть приклади. Сформулюйте нульову і альтернативну гіпотезу для завдання.
2. Охарактеризуйте поняття «Статистичний критерій». У чому полягає відмінність між параметричними і непараметричними критеріями?
3. Які види статистичних критеріїв реалізовані в Excel? Як вони здійснюються?
4. Як здійснюється перевірка гіпотез про рівність середніх для двох незалежних груп за допомогою SPSS?
5. Поясніть, що виводиться в таблицях «Статистика груп» і «Тест для незалежних вибірок» у вікні виведення?
6. Порівняйте з результатами, отриманими в Excel.
7. Які групи вважаються залежними або парними?
8. Як здійснюється перевірка гіпотез про рівність середніх для двох залежних груп за допомогою SPSS? Яке емпіричне значення критерію Стюдента? Які межі довірчого інтервалу для різниці середніх? Чи потрапляє туди критичне значення?
9. Дайте змістовну інтерпретацію результатам.

ЛАБОРАТОРНЕ ЗАНЯТТЯ № 16

Тема: *Однофакторний дисперсійний аналіз (Microsoft Excel, SPSS для Windows*

Мета: навчитися використовувати пакет статистичного аналізу електронних таблиць Excel і можливості SPSS для Windows для однофакторного дисперсійного аналізу.

Теоретичні відомості:

Дисперсійний аналіз - це один з методів аналізу мінливості ознаки під впливом контрольованих чинників. Автором фундаментальної концепції дисперсійного аналізу є Р.Фишер.

У англomовній літературі терміну «дисперсійний аналіз» відповідає аббревіатура **ANOVA (Analysis of Variance** - аналіз варіативної/дисперсії).

Для того, щоб грамотно здійснити дисперсійний аналіз, явище, що вивчається, треба представити як систему змінних, серед яких слід виділити гіпотетично(імовірно) незалежну змінну(чи контрольований чинник) і гіпотетично залежну змінну(ознака). Незалежна змінна(контрольований чинник) імовірно є причиною варіативної залежної змінної, тобто залежна змінна імовірно змінюється під впливом незалежної змінної(контрольованого чинника). Така модель - це досить грубе спрощення реальності(оскільки на залежну змінну найчастіше впливають і інші чинники, які невідомі дослідникові або вивчення впливу яких не входить в плани дослідника. У останньому випадку вплив таких чинників намагаються або усунути або урівноважити в усіх вибірках), але статистичні методи, зокрема, дисперсійний аналіз, дозволяють нехтувати цим спрощенням і отримати статистично достовірний результат аналізу мінливості залежної змінної під впливом контрольованого чинника.

Метою дисперсійного аналізу є перевірка значущості відмінності між середніми за допомогою порівняння(тобто аналізу) дисперсій. А саме - розділення загальної дисперсії на декілька джерел, дозволяє порівняти