

---

## ЛАБОРАТОРНЕ ЗАНЯТТЯ 4

**Обчислення коваріації, лінійного коефіцієнту кореляції та парного коефіцієнту детермінації за означеннями та за допомогою вбудованих функцій та пакетів *MS Excel*. Побудова функції лінійної парної регресії за допомогою вбудованих функцій *MS Excel*.**

---

**Мета:** Навчитись обчислювати коваріацію, лінійний коефіцієнт кореляції та парний коефіцієнт детермінації за означеннями та за допомогою вбудованих функцій та пакетів *MS Excel*.

---

Пригадаємо теоретичні відомості з лекції.

**Означення.** *Кореляційний аналіз* – це розділ математичної статистики, який вивчає взаємозв'язки між випадковими величинами, та полягає у кількісному визначенні тісноти взаємозв'язку між ознаками.

**Означення.** *Кореляція* – це статистична залежність між випадковими величинами, при якій зміна однієї випадкової величини приводить до зміни математичного сподівання другої випадкової величини.

Ми часто чуємо з новин, що «спостерігається кореляційний зв'язок» або «величини корельовані», коли представляють економічні дані. Це лише вказує на те, що випадкові величини, такі, наприклад, як ціни на паливо та ціни на комунальні платежі, є взаємопов'язані. Кореляційний аналіз дозволяє нам визначити, наскільки тісний цей зв'язок, та чи є він прямим або зворотним, тобто випадкові величини зростають або спадають одночасно, чи зростання однієї спостерігається при спаданні іншої.

Методи кореляційного аналізу застосовуються в математичній статистиці при дослідженні та виявленні взаємозв'язку між декількома випадковими величинами. Найпростіші дослідження – це аналіз взаємозв'язку між двома

випадковими величинами, і саме таке дослідження ми проведемо на цьому лабораторному занятті. Так, наприклад, збираючи статистичні дані про дохід населення та купівельну спроможність, навіть не застосовуючи математичні методи, можна побачити, що зростання доходу одночасно спостерігається зі зростанням купівельної спроможності, і так само зі зниженням – ці дві випадкові величини також знижуються одночасно. Цей простий приклад ілюструє прямий зв'язок, коли ж зростання однієї випадкової величини є одночасним зі зростанням другої. Якщо ж навпаки – зі зниженням першої одночасно зростає друга, то говорять про зворотній зв'язок. Але для того, щоб можна було проводити більш точний математичний аналіз, необхідно визначити зручний коефіцієнт, який би вказував як на наявність або відсутність зв'язку, так і визначав тісноту цього зв'язку, якщо він є.

Нехай досліджується можливий взаємозв'язок двох ознак  $X$  та  $Y$  на основі отриманих відповідних спостережень – вибірок однакового об'єму  $n$ :

$X$	$x_1$	$x_2$	...	$x_n$
$Y$	$y_1$	$y_2$	...	$y_n$

Тіснота *взаємозв'язку* ознак  $X$  та  $Y$  кількісно виражається *величиною коефіцієнта кореляції*, який пов'язаний з обчисленням суми добутків відхилень значень  $x_i$  та  $y_i$ , що спостерігаються при випробуваннях, від їх середніх  $\bar{x}$  та  $\bar{y}$ :

$$\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y}).$$

Сума таких добутків, розділена на об'єм вибірки  $n$ , називається *коваріацією*:

$$\text{cov}(y, x) = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n},$$

де  $n$  – об'єм вибірки,

$x_i$  та  $y_i$  – значення ознак  $X$  та  $Y$ , що спостерігаються при випробуваннях (дані вибірок),

$\bar{x}$  та  $\bar{y}$  – вибіркові середні ознак  $X$  та  $Y$ .

На практиці часто використовують наступну формулу для обчислення коваріації:

$$\text{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y},$$

де

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i,$$

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n y_i,$$

$$\overline{xy} = \frac{x_1 y_1 + x_2 y_2 + \dots + x_n y_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i y_i.$$

Коваріація характеризує спряженість варіацій двох ознак та являє собою статистичну міру взаємодії двох випадкових величин. При наявності *прямого зв'язку* коваріація є додатно визначеною, *зворотного зв'язку* – від'ємне, а при *відсутності зв'язку* – коваріація за значенням буде близька до нуля. Але розмірність коваріації залежить від величин значень ознак  $X$  та  $Y$ , що є недостатньо зручним при дослідженні. Для зручності використовують відносну характеристику зв'язку – для цього коваріацію ділять на максимально можливе значення, яке дорівнює добутку середньоквадратичних відхилень двох ознак  $\sigma_x$  та  $\sigma_y$ , і в результаті отримують *лінійний коефіцієнт кореляції*:

$$r_{yx} = \frac{\text{cov}(y, x)}{\sigma_x \cdot \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y},$$

де

$$\sigma_x = \sqrt{d_x} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{x^2 - (\bar{x})^2},$$

$$\sigma_y = \sqrt{d_y} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{y^2 - (\bar{y})^2}.$$

На практиці часто використовують наступну формулу для обчислення коефіцієнту кореляції:

$$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{x^2 - (\bar{x})^2} \cdot \sqrt{y^2 - (\bar{y})^2}}.$$

Коефіцієнт кореляції  $r_{yx}$  приймає значення від  $-1$  до  $+1$ :

$$-1 \leq r_{yx} \leq 1.$$

Додатне значення коефіцієнта говорить про наявність прямого зв'язку, від'ємне значення – про наявність зворотного. Якщо  $r_{yx} = -1$  або  $r_{yx} = 1$ , кореляційний зв'язок представляється лінійною функціональною залежністю, при  $r_{yx} = 0$  лінійний кореляційний зв'язок відсутній.

### Якісні характеристики кореляційного зв'язку

Значення $r_{yx}$	Характер зв'язку
від 0 до $ \pm 0,3 $	практично відсутній
від $ \pm 0,3 $ до $ \pm 0,5 $	слабий
від $ \pm 0,5 $ до $ \pm 0,7 $	помірний
від $ \pm 0,7 $ до $ \pm 1 $	сильний

Знаючи лінійний коефіцієнт кореляції, можна визначити *парний коефіцієнт детермінації*, який дорівнює  $(r_{yx})^2$  та показує, яка доля варіації змінної  $y$  обумовлена впливом на неї зміною  $x$ .

В лекції 10 наведений приклад з розрахункам вручну. На лабораторному занятті проведемо всі розрахунки у *MS Excel*.

---

**Задача.** За даними двох вибірок, а саме даних про середню заробітню платню на підприємстві та кількість звільнень на цьому ж підприємстві знайти коваріацію, лінійний коефіцієнт кореляції та парний коефіцієнт детермінації за означеннями та за допомогою вбудованих функцій та пакетів *MS Excel*.

Заробітня платня (тис. грн.)	Кількість звільнень (люд.)
6	32
6,5	30
7	26
7,5	25
8	21
8,5	18
9	16
10	14
10,5	11
11	10
11,5	7
12	5

### Вказівки до виконання роботи

1. Для обчислення *коваріації* використаємо наступну формулу:

$$\text{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y},$$

де

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i,$$

$$\bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n y_i,$$

$$\overline{xy} = \frac{x_1y_1 + x_2y_2 + \dots + x_ny_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_iy_i.$$

Обчислимо спочатку всі необхідні вибіркові середні:  $\bar{x}$  в комірці E3 за формулою =СУММ(A2:A13)/СЧЁТ(A2:A13)

The screenshot shows an Excel spreadsheet with two columns: 'Заробітна платня (тис. грн.)' (Salary) in column A and 'Кількість звільнень (люд.)' (Number of layoffs) in column B. The data points are: (6, 32), (6.5, 30), (7, 26), (7.5, 25), (8, 21), (8.5, 18), (9, 16), (10, 14), (10.5, 11), (11, 10), (11.5, 7), (12, 5). The spreadsheet is in the '1 шаг (по формулам)' (1 step - by formulas) mode. A dialog box titled 'Аргументы функции' (Function Arguments) is open, showing the 'СЧЕТ' (COUNT) function. The 'Значение1' (Argument1) is set to 'A2:A13', and the 'Значение2' (Argument2) is empty. The dialog indicates that the function counts the number of cells in the range that contain numbers, and the result is 12.

$\bar{y}$  в комірці E4 за формулою =СУММ(B2:B13)/СЧЁТ(B2:B13)

This screenshot shows the same Excel spreadsheet as above, but now in the '2 шаг (встроенные функции)' (2 step - built-in functions) mode. The formula bar shows '=СУММ(B2:B13)/СЧЁТ(B2:B13)'. The result of the calculation, 17.92, is displayed in cell E4. The dialog box is no longer visible.

$\overline{xy}$  в комірці E5 за формулою  $=\text{СУММПРОИЗВ}(A2:A13;B2:B13)/\text{СЧЁТ}(B2:B13)$

1	Заробітня платня (тис. грн.)	Кількість звільнень (люд.)		1 крок	(за формулами)
2	6	32			
3	6,5	30		$\overline{x}$	8,96
4	7	26		$\overline{y}$	17,92
5	7,5	25		$\overline{xy}$	143,96
6	8	21	$\text{cov}(y, x) = \overline{xy} - \overline{x} \cdot \overline{y} =$		-16,55

І після чого знаходимо *коваріацію*  $\text{cov}(y, x) = \overline{xy} - \overline{x} \cdot \overline{y}$  в комірці E6 за формулою  $=E5-E3*E4$

E6		fx		=E5-E3*E4	
	A	B	C	D	E
1	Заробітня платня (тис. грн.)	Кількість звільнень (люд.)		1 крок	
2	6	32			
3	6,5	30		$\overline{x}$	8,96
4	7	26		$\overline{y}$	17,92
5	7,5	25		$\overline{xy}$	143,96
6	8	21	$\text{cov}(y, x) = \overline{xy} - \overline{x} \cdot \overline{y} =$		-16,55
7	8,5	18		Висновок	зворотний зв'язок
8	9	16			

Робимо висновок про існування зворотного зв'язку, бо коваріація приймає від'ємне значення.

Переходимо до обчислення коефіцієнту кореляції, який нам надасть інформацію про тісноту зв'язку. Для обчислення *коефіцієнту кореляції* використаємо формулу:

$$r_{yx} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{x^2 - (\overline{x})^2} \cdot \sqrt{y^2 - (\overline{y})^2}}$$

Зробимо ще допоміжні розрахунки, а саме знайдемо  $\overline{x^2}$  та  $\overline{y^2}$  комірках E10 та E11 за формулами  $=\text{СУММПРОИЗВ}(A2:A13;A2:A13)/\text{СЧЁТ}(A2:A13)$  та  $=\text{СУММПРОИЗВ}(B2:B13;B2:B13)/\text{СЧЁТ}(B2:B13)$

E11		fx = СУММПРОИЗВ(B2:B13;B2:B13)/СЧЁТ(B2:B13)			
	A	B	C	D	E
1	Заробітня платня (тис. грн.)	Кількість звільнень (люд.)		1 крок	(за формулами)
2	6	32			
3	6,5	30		$\bar{x}$	8,96
4	7	26		$\bar{y}$	17,92
5	7,5	25		$\overline{xy}$	143,96
6	8	21	$\text{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y} =$		-16,55
7	8,5	18		Висновок	
8	9	16			
9	10	14			
10	10,5	11		$\overline{x^2}$	84,02
11	11	10		$\overline{y^2}$	394,75
12	11,5	7		$\overline{xy} - \bar{x} \cdot \bar{y}$	
13	12	5	$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}} =$		
14					
15					
16				Висновок	
17					
18				$(r_{yx})^2 =$	

І сам коефіцієнт кореляції обчислимо у комірці E13 за формулою  $=(E5-E3 \cdot E4)/(\text{КОРЕНЬ}(E10-E3^2) \cdot \text{КОРЕНЬ}(E11-E4^2))$

E13		fx = E6/((КОРЕНЬ(E10-E3^2))*(КОРЕНЬ(E11-E4^2)))				
	A	B	C	D	E	F
1	Заробітня платня (тис. грн.)	Кількість звільнень (люд.)		1 крок		
2	6	32				
3	6,5	30		$\bar{x}$	8,96	
4	7	26		$\bar{y}$	17,92	
5	7,5	25		$\overline{xy}$	143,96	
6	8	21	$\text{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y} =$		-16,55	
7	8,5	18		Висновок	зворотний зв'язок	
8	9	16				
9	10	14				
10	10,5	11		$\overline{x^2}$	84,02	
11	11	10		$\overline{y^2}$	394,75	
12	11,5	7		$\overline{xy} - \bar{x} \cdot \bar{y}$		
13	12	5	$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}} =$		-0,99	
14						
15						
16				Висновок	сильний зв'язок	
17						

Отриманий коефіцієнт кореляції за модулем близький до 1, що говорить про сильний зв'язок випадкових величин. Від'ємний знак говорить про зворотний зв'язок. Отже величина коефіцієнта кореляції говорить про сильний зворотний зв'язок між середньою заробітною платнею та кількістю звільнень працівників.

Обчислимо парний коефіцієнт детермінації  $(r_{yx})^2$ .



	A	B	C	D	E
1	Заробітна платня (тис. грн.)	Кількість звільнень (люд.)		1 крок	
2	6	32			
3	6,5	30		$\bar{x}$	8,96
4	7	26		$\bar{y}$	17,92
5	7,5	25		$\overline{xy}$	143,96
6	8	21	$\text{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y} =$		-16,55
7	8,5	18		Висновок	зворотний зв'язок
8	9	16			
9	10	14			
10	10,5	11		$\overline{x^2}$	84,02
11	11	10		$\overline{y^2}$	394,75
12	11,5	7		$\overline{xy} - \bar{x} \cdot \bar{y}$	
13	12	5	$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}} =$		-0,99
14					
15					
16				Висновок	сильний зв'язок
17					
18				$(r_{yx})^2 =$	0,98
19					

Отримане значення парного коефіцієнту детермінації

$$(r_{yx})^2 \approx 0,98$$

показує, що на 98% зміна числа звільнень працівників пояснюється зміною середньою заробітною платнею.

2. Застосуємо вбудовані функції *MS Excel* для знаходження коваріації та коефіцієнту кореляції.

Для обчислення коваріації в статистичних функціях оберемо КОВАРИАЦИЯ.Г, де у Массив 1 та Массив 2 внесемо дані наших вибірок.

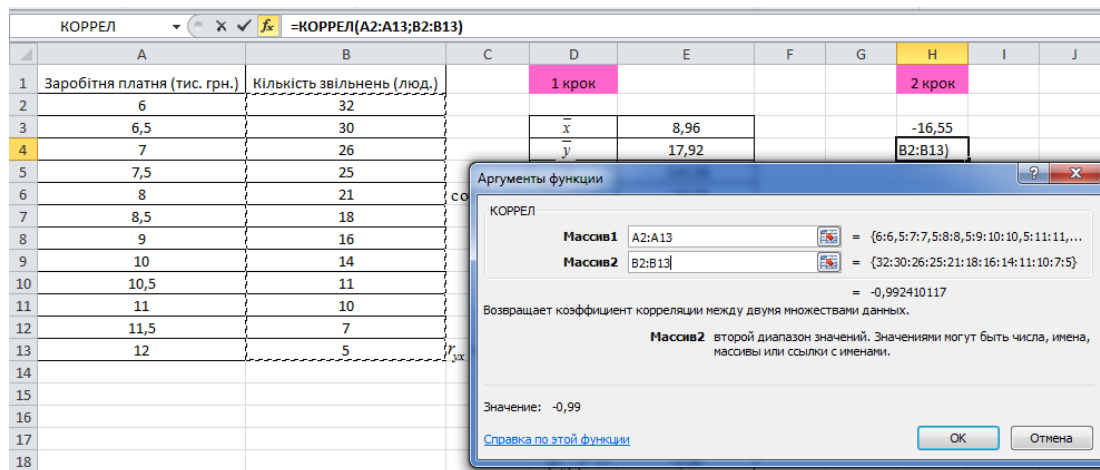
The screenshot shows an Excel spreadsheet with the following data in columns A and B:

	A	B
1	Заробітна платня (тис. грн.)	Кількість звільнень (люд.)
2	6	32
3	6,5	30
4	7	26
5	7,5	25
6	8	21
7	8,5	18
8	9	16
9	10	14
10	10,5	11
11	11	10
12	11,5	7
13	12	5

The 'Аргументы функции' dialog box shows:

- Функция: КОВАРИАЦИЯ.Г
- Массив1: A2:A13 = {6;6,5;7;7,5;8;8,5;9;10;10,5;11;11;11,5;12}
- Массив2: B2:B13 = {32;30;26;25;21;18;16;14;11;10;7;5}
- Возвращает ковариацию генеральной совокупности, среднее попарных произведений отклонений.
- Массив2: второй диапазон целых чисел - числа, массивы или ссылки на ячейки, содержащие числа.
- Значение: -16,54513889

Для знаходження коефіцієнту кореляції в статистичних функціях оберемо КОРРЕЛ, де у Массив 1 та Массив 2 внесемо дані наших вибірок.



Отримані значення співпадають з обчисленими за формулами.

	A	B	C	D	E	F	G	H
1	Заробітна платня (тис. грн.)	Кількість звільнень (люд.)		1 крок	(за формулами)		2 крок	(вбудовані функції)
2	6	32						
3	6,5	30		$\bar{x}$	8,96		коваріація	-16,55
4	7	26		$\bar{y}$	17,92		кореляція	-0,99
5	7,5	25		$\overline{xy}$	143,96			
6	8	21	$\text{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y} =$		-16,55			
7	8,5	18		Висновок	зворотний зв'язок			
8	9	16						
9	10	14						
10	10,5	11		$\overline{x^2}$	84,02			
11	11	10		$\overline{y^2}$	394,75			
12	11,5	7		$\overline{xy} - \bar{x} \cdot \bar{y}$				
13	12	5	$r_{yx} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}} =$		-0,99			
14				Висновок	сильний зв'язок			
15								
16				$(r_{yx})^2 =$	0,98			
17								
18								
19								

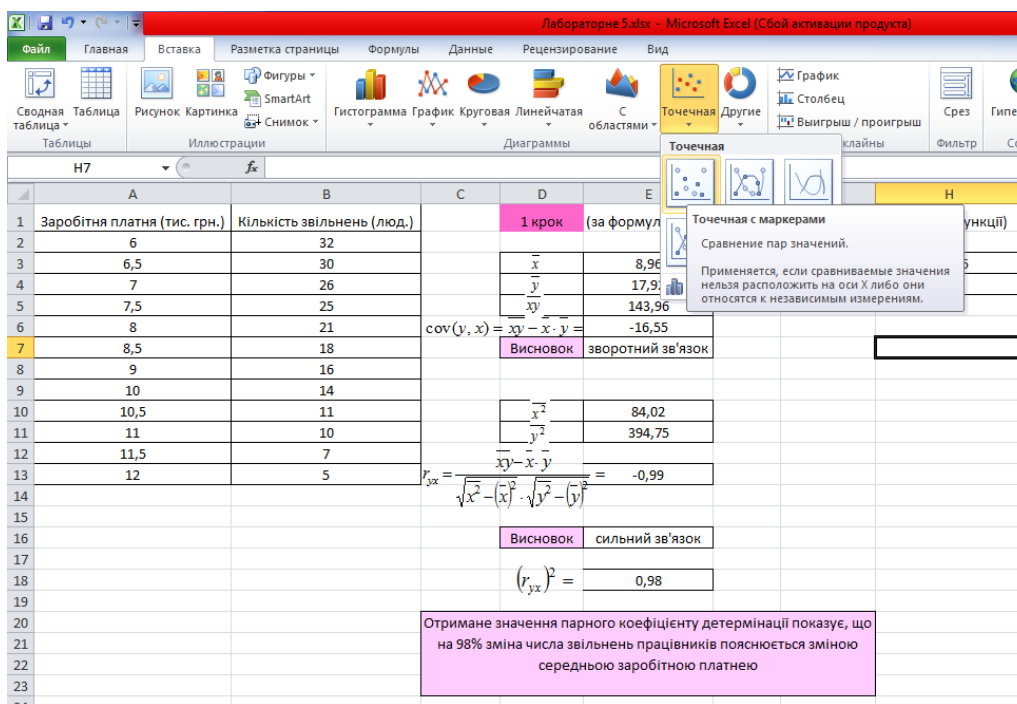
3. Останнім кроком дослідження є побудова функції лінійної парної регресії за допомогою вбудованих функції *MS Excel*. Окрім кореляційного аналізу, при якому досліджується тіснота взаємозв'язку між ознаками, в економічних дослідженнях часто звертаються до *регресійного аналізу*, який надає методи побудови математичних моделей цього взаємозв'язку між ознаками для подальшого прогнозування варіації ознак.

Регресійний аналіз проводиться на основі побудованого рівняння регресії, яке у нашому випадку є рівнянням *парної регресії*  $\hat{y}_x = f(x)$ , бо характеризує зв'язок між двома ознаками. Результативною ознакою, тобто залежною від

іншої, в нашому прикладі виступає «Кількість звільнень», а факторною – «Заробітна платня».

Не кожний існуючий зв'язок між двома випадковими величинами можна виразити за допомогою лінійної функції, але можна побачити, коли це є можливим вже на графічному представленні їх розподілів, яке називається *діаграмою розсіяння (кореляційним полем)*. Для цього необхідно в декартовій системі координат відмітити точки з координатами  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , і якщо розташування цих точок буде таким, що можливо провести деяку пряму так, що всі вони будуть розсіяні навколо неї, то є можливість, що зв'язок двох ознак  $X$  та  $Y$  може бути виражений за допомогою лінійної функції. У міру того, як зростає сила лінійного зв'язку, точки на графіку будуть лежати більш близько до прямої.

За допомогою вбудованих функції *MS Excel* представимо графічно розподіли представлених випадкових величин. Для цього оберемо Вставка→Точечная діаграма→Точечная с маркерами



Та оберемо Макет 9

Лабораторне 5.xlsx - Microsoft Excel (Свій активації продукту)

Работа с диаграммами

Конструктор Макет Формат

Изменить тип диаграммы как шаблон Сохранить шаблон Строка/столбец Выбрать данные

Тип Данные

Диаграмма 5

	A	B	D	E	F	G	H
1	Заробітня платня (тис. грн.)	Кількість звільнень	1 крок (за формулами)			2 крок (вбудовані функції)	
2	6	32					
3	6,5	30		8,96		коваріація	-16,55
4	7	26		17,92		кореляція	-0,99
5	7,5	25		143,96			
6	8	21					
7	8,5	18					
8	9	16					
9	10	14					
10	10,5	11					
11	11	10					
12	11,5	7					
13	12	5					

Отримане на 98% зміна числа звільнень працівників пояснюється зміною середньою заробітною платнею

І задамо дані: в значення x – стовпець із значеннями заробітної платні, в значення y – стовпець із значеннями кількості звільнень.

Excel (Свій активації продукту)

Работа с диаграммами

Конструктор Макет Формат

Данные Рецензирование Вид

Стили диаграмм

C	D	E	F	G	H	I	J	K	L	M
	1 крок (за формулами)			2 крок (вбудовані функції)						
	$\bar{x}$	8,96		коваріація	-16,55					
	$\bar{y}$	17,92		кореляція	-0,99					
	$\overline{xy}$	143,96								
	$\text{cov}(y, x)$	-16,55								

Выбор источника данных

Диапазон данных для диаграммы:

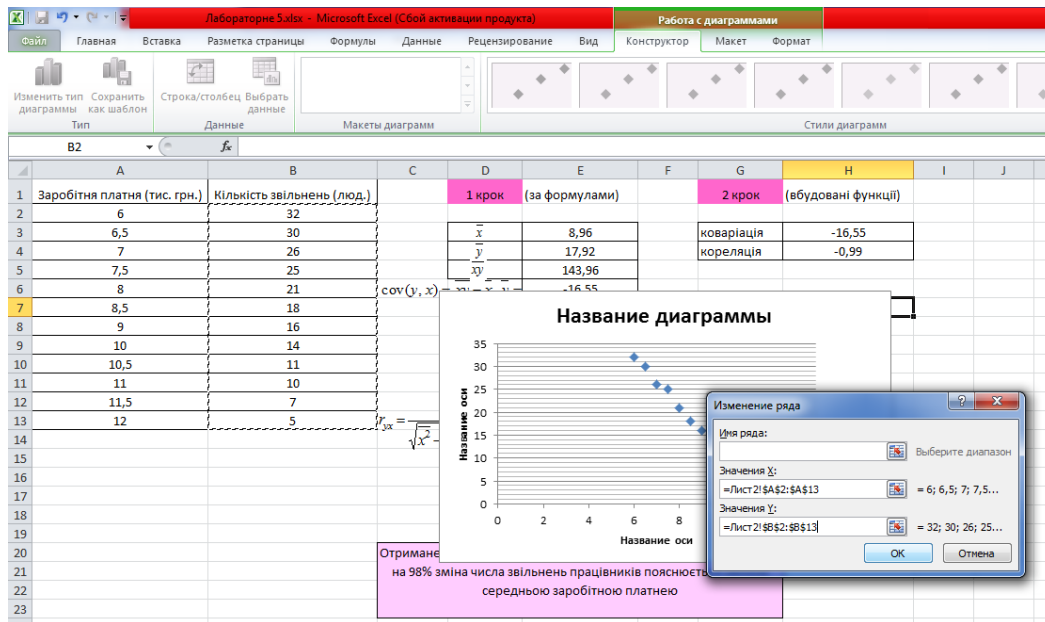
Строка/столбец

Элементы легенды (ряды)

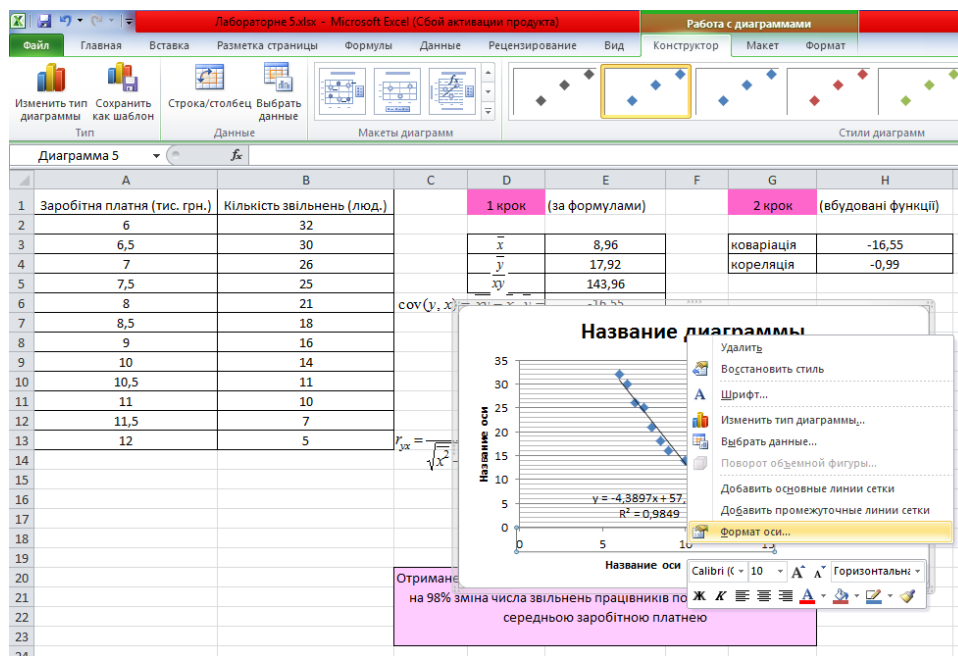
Подписки горизонтальной оси (категории)

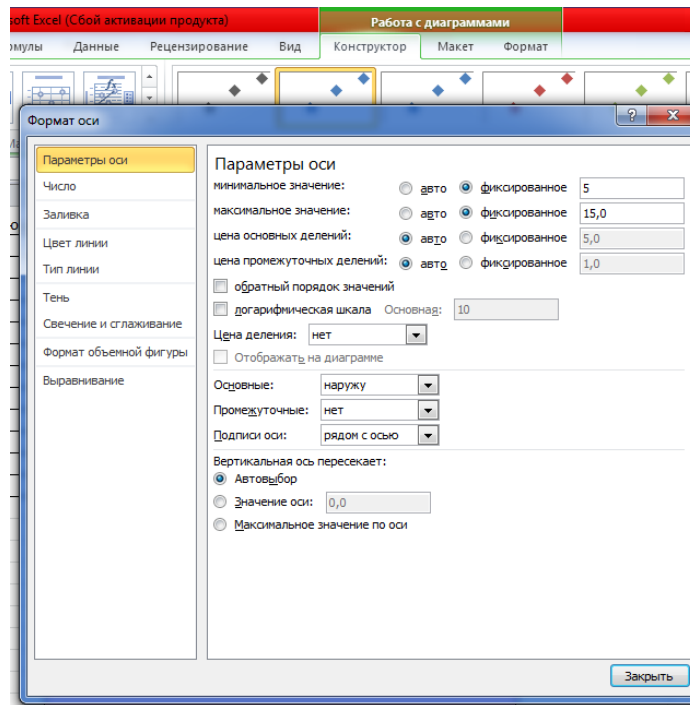
Скрытые и пустые ячейки

OK Отмена

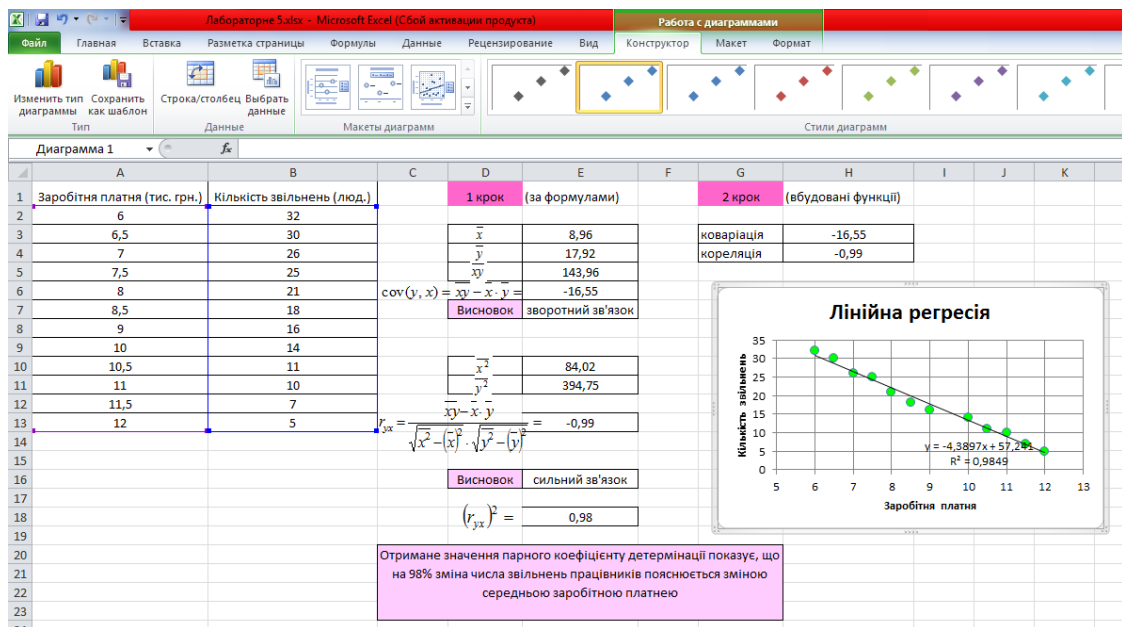


Після чого підправимо формат осі Ох від 5 до 15





Та дамо назви осям і самому рисунку.



Нахил прямої залежить від характеру зв'язку: прямий зв'язок відповідає додатному кутовому коефіцієнту, зворотний – від'ємному.

## Лабораторна робота №4

Тема: Обчислення коваріації, лінійного коефіцієнту кореляції та парного коефіцієнту детермінації за означеннями та за допомогою вбудованих

**функцій та пакетів *MS Excel*. Побудова функції лінійної парної регресії за допомогою вбудованих функцій *MS Excel*.**

**Звіт студента(-ки) групи \_\_\_\_\_**

**ПІБ**

**Для вибірових даних, наданих у Домашнє (файл Шаблон 6), обчислити коваріацію та коефіцієнт кореляції за означеннями та за допомогою вбудованих функцій та пакетів *MS Excel*. Побудувати функцію лінійної парної регресії за допомогою вбудованих функцій *MS Excel***

---