

Корреляция

Выполнил: Мацкевич В.А.
Группа: АСОиУб-18-1

Содержание

1. Понятие корреляции
2. Корреляционная связь
3. Виды корреляционной связи
4. Примеры корреляций
5. Ложная корреляция
6. Коэффициент корреляции
7. Нормированный коэффициент корреляции Браве-Пирсона
8. Оценка статистической значимости коэффициента корреляции
9. Расчет коэффициента корреляции средствами MS Excel

Понятие корреляции

Корреляция — статистическая взаимосвязь двух или более случайных величин.

При этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.



Корреляционная связь

Корреляционная связь – это согласованное изменение двух признаков, отражающее тот факт, что изменчивость одного признака находится в соответствии с изменчивостью другого.

Например, связь между ростом и весом человека; посевной площадью и валовым сбором зерна, понижением жизненного уровня и революционной активностью.



Виды корреляционной связи

Корреляционная связь между признаками может быть линейной и криволинейной (нелинейной), положительной и отрицательной.

Положительная корреляция отражает однотипность в изменении признаков: с увеличением значений первого признака увеличиваются значения и другого, или с уменьшением первого уменьшается второй.

Отрицательная корреляция указывает на увеличение первого признака при уменьшении второго или уменьшение первого признака при увеличении второго.



Примеры корреляций

А) - строгая положительная корреляция;

Б) - положительная корреляция;

В) - слабая положительная корреляция;

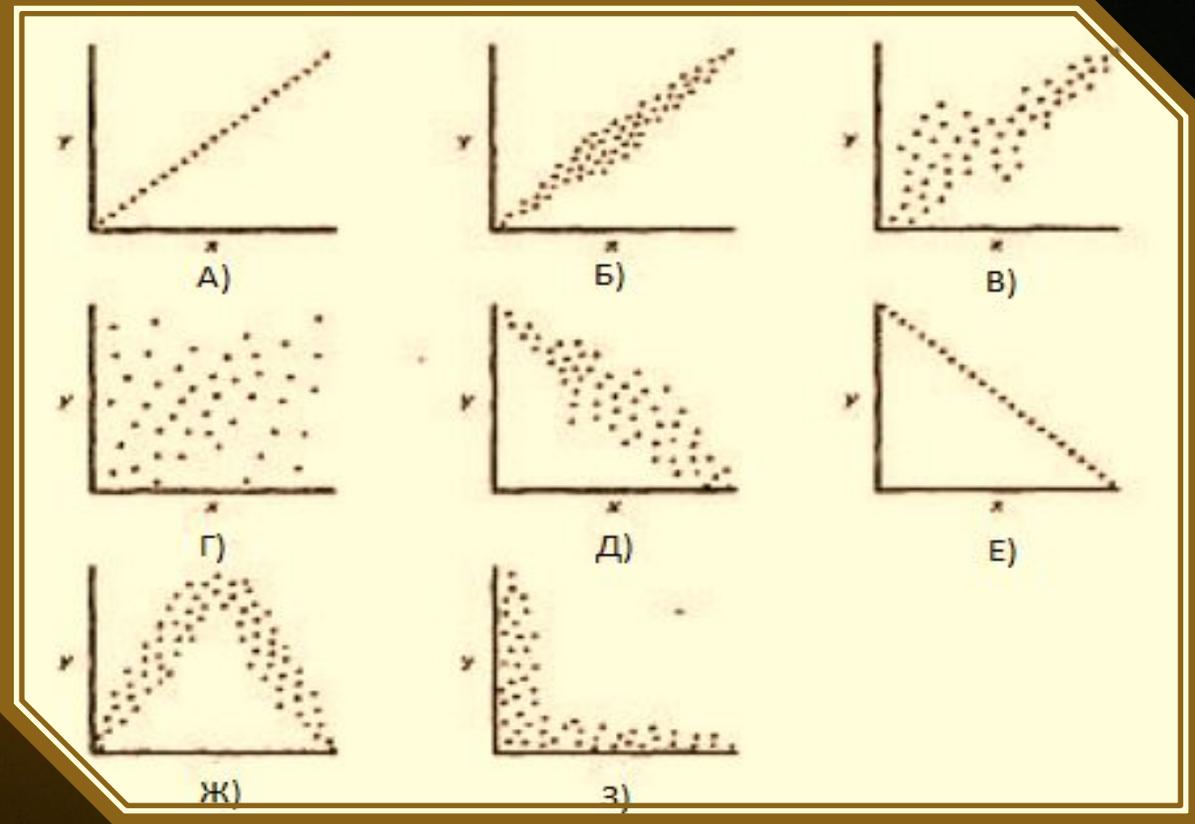
Г) - нулевая корреляция;

Д) - отрицательная корреляция;

Е) - строгая отрицательная корреляция;

Ж) - нелинейная корреляция;

З) - нелинейная корреляция.



Ложная корреляция

Если между двумя исследуемыми величинами установлена тесная зависимость, то из этого еще не следует их причинная взаимообусловленность.

За счет эффектов одновременного влияния неучтенных факторов смысл истинной связи может искажаться. Поэтому такую корреляцию часто называют «ложной».



Коэффициент корреляции

Коэффициент корреляции — это статистический показатель зависимости двух случайных величин.

Коэффициент корреляции может принимать значения от -1 до +1.

При оценке зависимости между показателями используется следующую градацию (по модулю):

- высокая степень взаимосвязи — от 0,7 до 0,99;
- средняя степень взаимосвязи — от 0,5 до 0,69;
- слабая степень взаимосвязи — от 0,2 до 0,49.



Нормированный коэффициент корреляции Браве-Пирсона

Коэффициент корреляции Браве-Пирсона вычисляется по формуле:

$$r_{xy}^P = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$



Оценка статистической значимости коэффициента корреляции

Критическое значение t-критерия определяется из таблицы значений t-распределения для выбранного уровня значимости α и числа степеней свободы $k = n - 2$.

Если $r_{\phi} \geq r_{ст}$ то можно говорить о том, что между признаками наблюдается достоверная взаимосвязь. Если $r_{\phi} \leq r_{ст}$, то между признаками наблюдается недостоверная корреляционная взаимосвязь.

Число степеней свободы	Доверительная вероятность			Число степеней свободы	Доверительная вероятность		
	0,95	0,99	0,999		0,95	0,99	0,999
1.	0,997	1,000	1,000	17.	0,456	0,575	0,693
2.	0,950	0,990	0,999	18.	0,444	0,561	0,679
3.	0,878	0,959	0,992	19.	0,433	0,549	0,665
4.	0,811	0,917	0,974	20.	0,423	0,537	0,652
5.	0,754	0,874	0,951	25.	0,381	0,487	0,597
6.	0,707	0,834	0,925	30.	0,349	0,449	0,554
7.	0,666	0,798	0,898	35.	0,325	0,418	0,519
8.	0,632	0,765	0,872	40.	0,304	0,393	0,490
9.	0,602	0,735	0,847	45.	0,287	0,372	0,465
10.	0,576	0,708	0,823	50.	0,273	0,354	0,443
11.	0,553	0,684	0,801	60.	0,250	0,325	0,408
12.	0,514	0,661	0,780	70.	0,232	0,302	0,380
13.	0,504	0,641	0,760	80.	0,217	0,283	0,357
14.	0,497	0,623	0,742	90.	0,205	0,267	0,337
15.	0,482	0,606	0,725	100.	0,195	0,254	0,321
16.	0,468	0,590	0,708				



Расчет коэффициента корреляции средствами MS Excel

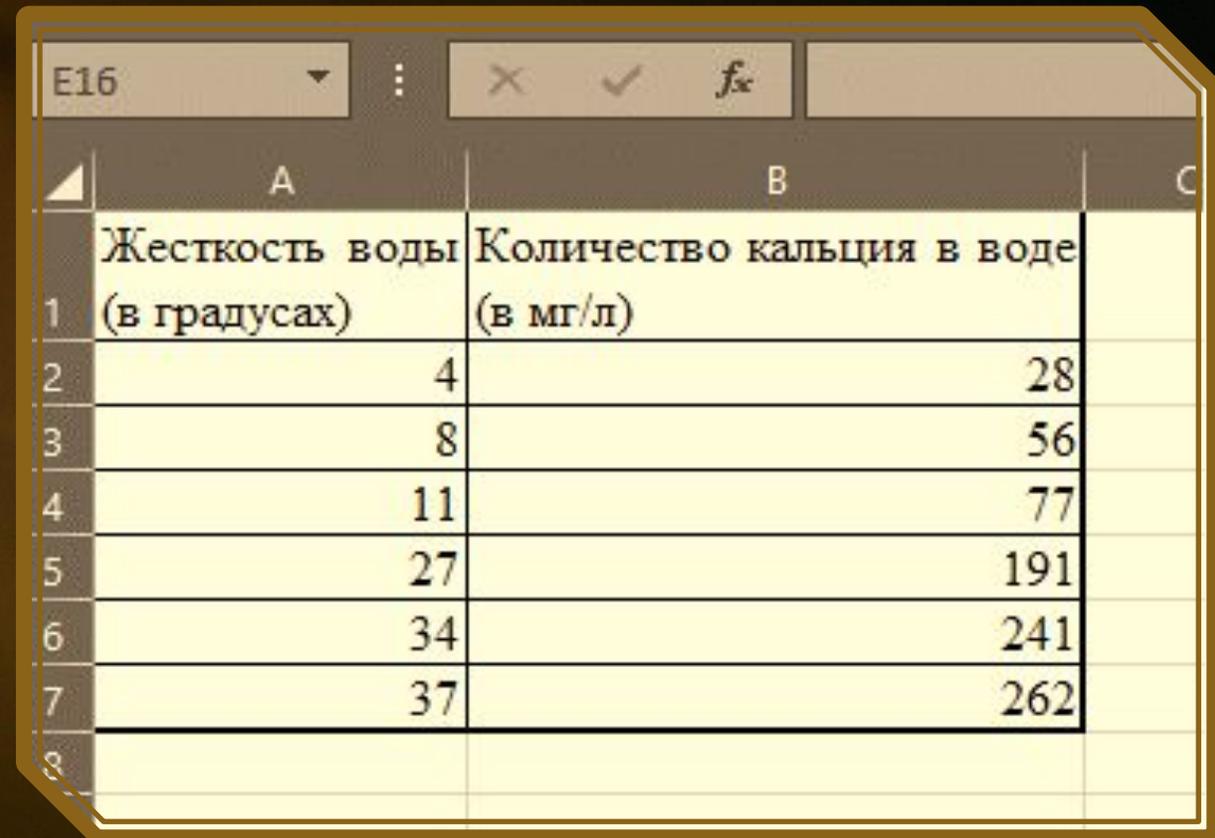
К примеру, нужно вычислить коэффициент корреляции, определить направление и силу связи между количеством кальция в воде и жесткостью воды, если известны следующие данные. Необходимо оценить достоверность связи.

Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)
4	28
8	56
11	77
27	191
34	241
37	262



Расчет коэффициента корреляции средствами MS Excel

Для начала нужно ввести
информацию в Excel.



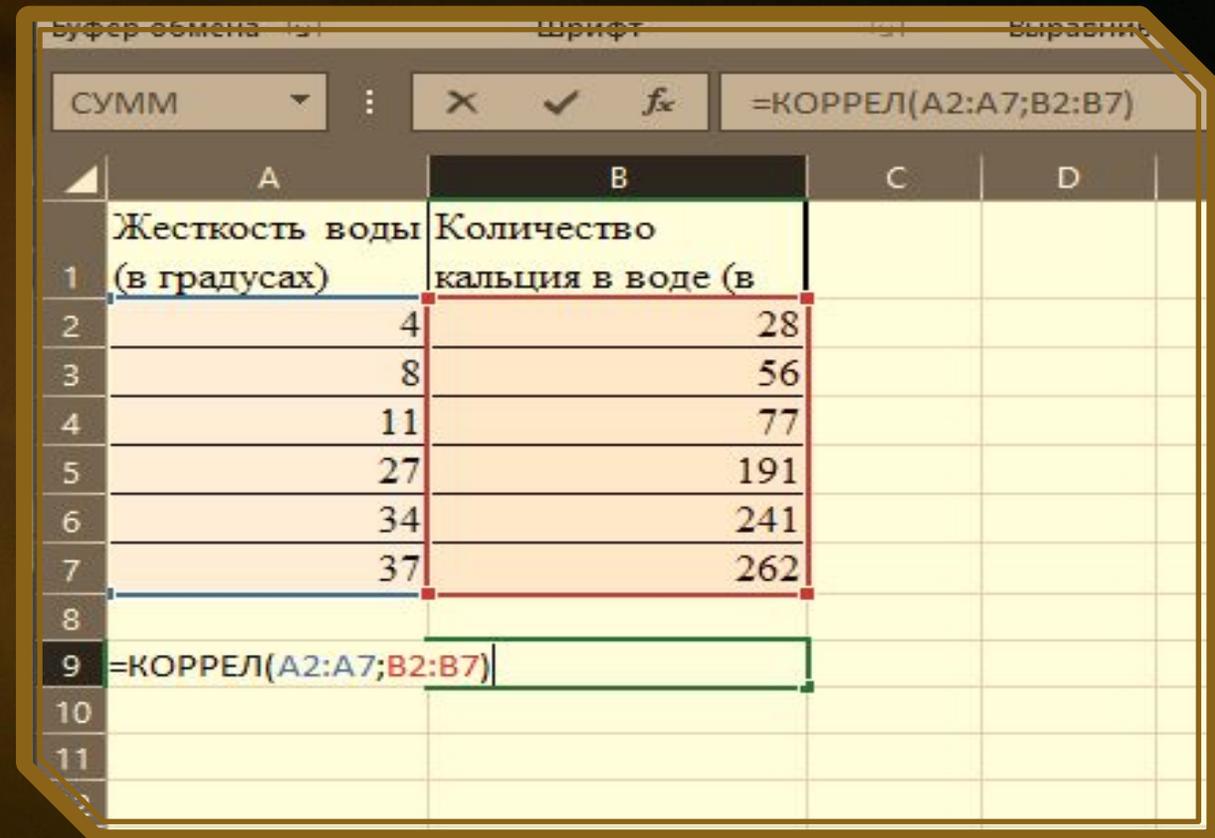
The screenshot shows an Excel spreadsheet with the following data:

	A	B	C
	Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)	
1			
2	4	28	
3	8	56	
4	11	77	
5	27	191	
6	34	241	
7	37	262	
8			



Расчет коэффициента корреляции средствами MS Excel

Далее в пустую ячейку вводим формулу “=КОРРЕЛ()”, аргументами которой являются два массива (массив 1 - ячейки A2:A7; массив 2 - ячейки B2:B7).



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
	Жесткость воды (в градусах)	Количество кальция в воде (в		
1				
2	4	28		
3	8	56		
4	11	77		
5	27	191		
6	34	241		
7	37	262		
8				
9	=КОРРЕЛ(A2:A7;B2:B7)			
10				
11				



Расчет коэффициента корреляции средствами MS Excel

В результате коэффициент корреляции $r_{xy} = +0,99$

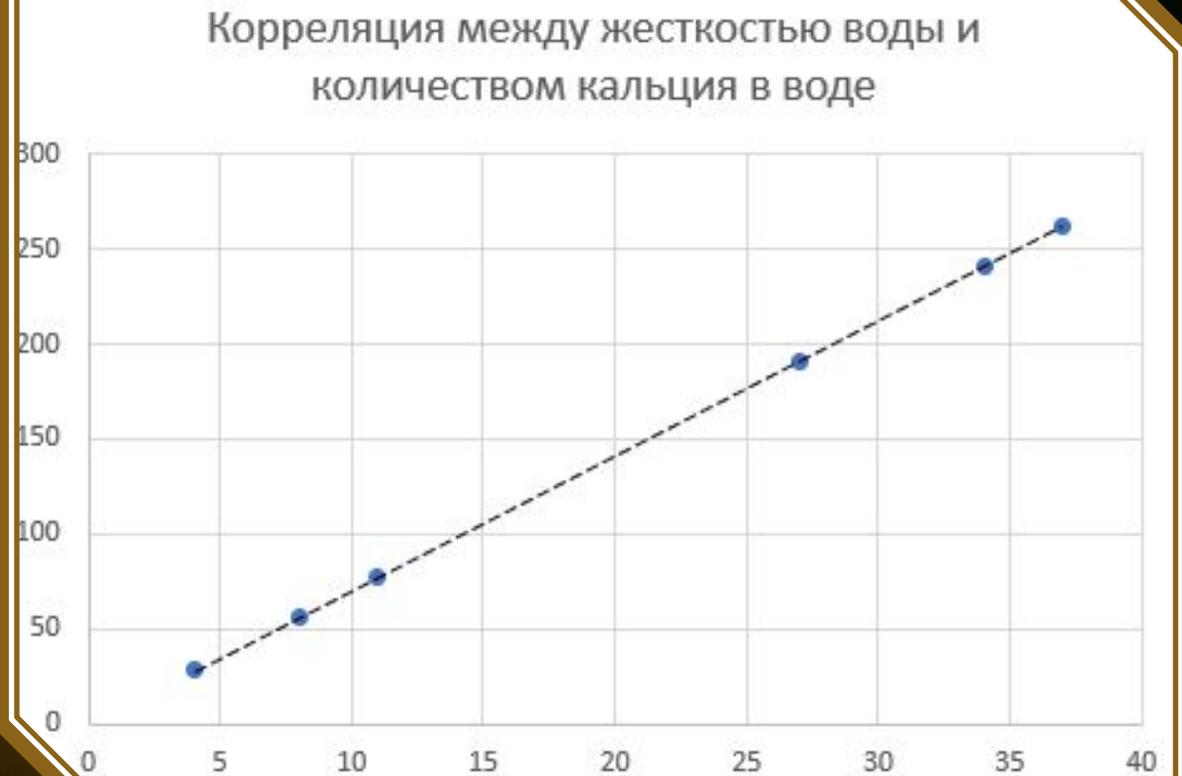
Из этого следует вывод: чем больше кальция в воде, тем она более жесткая. Связь прямая, сильная и достоверная.

Жесткость воды (в градусах)	Количество кальция в воде (в мг/л)
4	28
8	56
11	77
27	191
34	241
37	262
Коэффициент корреляции:	0,999996768



Расчет коэффициента корреляции средствами MS Excel

Построив диаграмму в этом
можно убедиться



Спасибо
внимание

за