

Решение типового примера 1. Проведем моделирование линейной парной корреляции между случайными числами в MS Excel. Для этого проведем ряд экспериментов по моделированию наборов значений случайных величин и измерению корреляции между ними.

На рисунке 1 представлен образец расчета в электронных таблицах.

Алгоритм составления расчетной модели:

- 1) В ячейку A1 ввести формулу¹: =слчис()*2-1
- 2) Скопировать формулу в ячейку B1, а затем скопировать формулы протягиванием в диапазон A2:B20. Значения случайных чисел совпадать с образцом не будут.
- 3) Ввести надписи в ячейки D1, F1, G1.
- 4) Ввести формулу² в ячейку D2: =коррел(A1:A20;B1:B20)
- 5) Заполнить номерами ячейки F2:F21.
- 6) Ввести с клавиатуры в ячейку G2 значение коэффициента корреляции между текущими значениями случайных чисел. После нажатия клавиши Enter значения случайных чисел и коэффициент корреляции изменятся. В ячейку G3 вводит с клавиатуры новое значение коэффициента корреляции, нажимаем клавишу Enter. В ячейку G4 вводит с клавиатуры новое значение коэффициента корреляции, нажимаем клавишу Enter и т.д.

¹ Результатом является случайное число, принимающее значения из отрезка [-1;1]. Умножение случайного числа на 2 увеличивает область значений, а вычитание единицы меняет минимальное значение. Исходное случайное число принимает значения из отрезка [0;1].

² Функция КОРРЕЛ в MS Excel вычисляет линейный парный коэффициент корреляции между двумя наборами данных.

один раз из 20, т.е. с частотой 0,05. Таким образом, сильная корреляция с вероятностью 5% может проявиться случайно.

Задание на самостоятельную работу 1. Провести моделирование корреляции между случайными числами для разных объемов выборки и сделать выводы: $n = 5; n = 10; n = 15; n = 20$.

Решение типового примера 2. Проведем корреляционно-регрессионный анализ взаимосвязи между количественными показателями:

x – число спортивных залов в регионах РФ в 2015 году;

y – ожидаемая продолжительность жизни при рождении в регионах РФ в 2015 г., число лет.

Исходные данные и расчетная таблица в MS Excel представлены на рис. 3, в режиме отображения формул на рис. 4.

Рассчитаем линейный парный коэффициент корреляции по формуле:

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \sigma_x = \sqrt{\overline{x^2} - \bar{x}^2}, \sigma_y = \sqrt{\overline{y^2} - \bar{y}^2}.$$

Результаты расчетов представлены на рис. 5 и 6 (в режиме отображения формул).

Линейный парный коэффициент корреляции: $r_{xy} = 0,83$, значит связь между ожидаемой продолжительностью жизни и числом стадионов по данным 2015 года прямая и сильная.

Уравнение линейной парной регрессии имеет вид:

$$y = 69,49 + 0,0018x + \varepsilon.$$

	A	B	C	D	E	F
1	Регион	x	y	yx	y^2	x^2
2	Белгородская область	819	72,61	59467,59	5272,21	670761
3	Брянская область	675	70,36	47493	4950,53	455625
4	Владимирская область	716	69,82	49991,12	4874,83	512656
5	Воронежская область	1428	71,67	102344,76	5136,59	2039184
6	Ивановская область	518	70,62	36581,16	4987,18	268324
7	Калужская область	592	70,73	41872,16	5002,73	350464
8	Костромская область	380	70,38	26744,4	4953,34	144400
9	Курская область	659	70,8	46657,2	5012,64	434281
10	Липецкая область	602	71,07	42784,14	5050,94	362404
11	Московская область	2207	72,26	159477,82	5221,51	4870849
12	Орловская область	586	70,38	41242,68	4953,34	343396
13	Рязанская область	516	71,46	36873,36	5106,53	266256
14	Смоленская область	924	69,74	64439,76	4863,67	853776
15	Тамбовская область	632	71,67	45295,44	5136,59	399424
16	Тверская область	819	69,1	56592,9	4774,81	670761
17	Тульская область	610	70,06	42736,6	4908,40	372100
18	Ярославская область	559	70,98	39677,82	5038,16	312481
19	г. Москва	3617	76,77	277677,09	5893,63	13082689
20	Сумма	16859,00	1280,48	1217949,00	91137,66	26409831,00
21	Среднее	936,61	71,14	67663,83	5063,20	1467212,83

Рис. 3. Исходные данные и расчетная таблица

	A	B	C	D	E	F
1	Регион	x	y	yx	y^2	x^2
2	Белгородская область	=Лист1!C2	=Лист1!B2	=B2*C2	=C2^2	=B2*B2
3	Брянская область	=Лист1!C3	=Лист1!B3	=B3*C3	=C3^2	=B3*B3
4	Владимирская область	=Лист1!C4	=Лист1!B4	=B4*C4	=C4^2	=B4*B4
5	Воронежская область	=Лист1!C5	=Лист1!B5	=B5*C5	=C5^2	=B5*B5
6	Ивановская область	=Лист1!C6	=Лист1!B6	=B6*C6	=C6^2	=B6*B6
7	Калужская область	=Лист1!C7	=Лист1!B7	=B7*C7	=C7^2	=B7*B7
8	Костромская область	=Лист1!C8	=Лист1!B8	=B8*C8	=C8^2	=B8*B8
9	Курская область	=Лист1!C9	=Лист1!B9	=B9*C9	=C9^2	=B9*B9
10	Липецкая область	=Лист1!C10	=Лист1!B10	=B10*C10	=C10^2	=B10*B10
11	Московская область	=Лист1!C11	=Лист1!B11	=B11*C11	=C11^2	=B11*B11
12	Орловская область	=Лист1!C12	=Лист1!B12	=B12*C12	=C12^2	=B12*B12
13	Рязанская область	=Лист1!C13	=Лист1!B13	=B13*C13	=C13^2	=B13*B13
14	Смоленская область	=Лист1!C14	=Лист1!B14	=B14*C14	=C14^2	=B14*B14
15	Тамбовская область	=Лист1!C15	=Лист1!B15	=B15*C15	=C15^2	=B15*B15
16	Тверская область	=Лист1!C16	=Лист1!B16	=B16*C16	=C16^2	=B16*B16
17	Тульская область	=Лист1!C17	=Лист1!B17	=B17*C17	=C17^2	=B17*B17
18	Ярославская область	=Лист1!C18	=Лист1!B18	=B18*C18	=C18^2	=B18*B18
19	г. Москва	=Лист1!C19	=Лист1!B19	=B19*C19	=C19^2	=B19*B19
20	Сумма	=СУММ(B2:B19)	=СУММ(C2:C19)	=СУММ(D2:D19)	=СУММ(E2:E19)	=СУММ(F2:F19)
21	Среднее	=СРЗНАЧ(B2:B19)	=СРЗНАЧ(C2:C19)	=СРЗНАЧ(D2:D19)	=СРЗНАЧ(E2:E19)	=СРЗНАЧ(F2:F19)

Рис. 4. Расчетная таблица в режиме отображения формул

Вывод: коэффициент регрессии $b = 0,0018$, значит при увеличении числа стадионов на единицу ожидаемая продолжительность жизни увеличивается в среднем на 0,0018 лет.

Коэффициент детерминации: $R^2 = 0,69$.

Вывод: 69% вариации результативной переменной – ожидаемой продолжительности жизни объясняется вариацией включенного в модель

фактора – числа стадионов в регионе, а остальные 31% - другими неучтенными в модели факторами.

Н	И
Дисперсия(x)=	589972,46
СКО(x)=	768,10
Дисперсия(y)=	2,62
СКО(y)=	1,62
$r(xy)$ =	0,83
b=	0,0018
a=	69,49
R^2 =	0,69

Рис. 5. Расчет линейного парного коэффициента корреляции и параметров уравнения линейной парной регрессии

Н	И
Дисперсия(x)=	=F21-B21*B21
СКО(x)=	=КОРЕНЬ(I2)
Дисперсия(y)=	=E21-C21*C21
СКО(y)=	=КОРЕНЬ(I4)
$r(xy)$ =	=(D21-C21*B21)/(I3*I5)
b=	=(D21-C21*B21)/I2
a=	=C21-I9*B21
R^2 =	=I7*I7

Рис. 6. Режим отображения формул

Построим по исходным данным точечную диаграмму и добавим на нее линию тренда (рис. 7). Для этого необходимо выделить B2:C19 и добавить точечную диаграмму. Затем выполнить правый щелчок мыши по точкам, в открывшемся контекстном меню выбрать пункт «Добавить линию тренда». В окне диалога «Формат линии тренда» выбрать линейную модель и установить флажки «Показывать уравнение на диаграмме» и «Поместить на диаграмму величину достоверности аппроксимации R^2 ».

Вывод: результаты подробного корреляционно-регрессионного анализа совпадают с результатами расчетов с помощью инструментов «Точечная диаграмма» и «Добавить линию тренда».

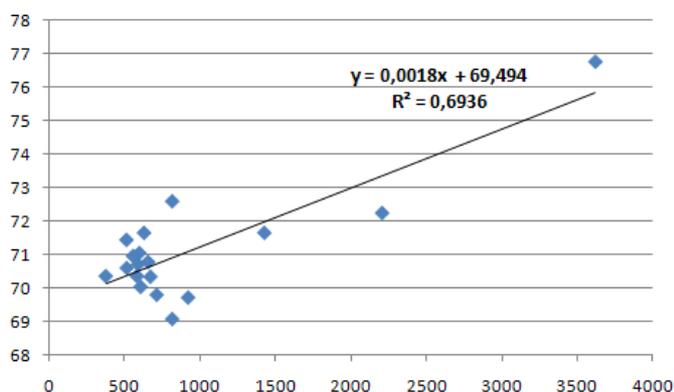


Рис. 7. Корреляционное поле и линейная регрессия

Задание на самостоятельную работу. По данным сайта Федеральной службы государственной статистики - ежегодного статистического справочника выбрать два количественных показателя по регионам РФ и измерить между ними линейную парную корреляцию, найти уравнение линейной парной регрессии, сделать выводы.