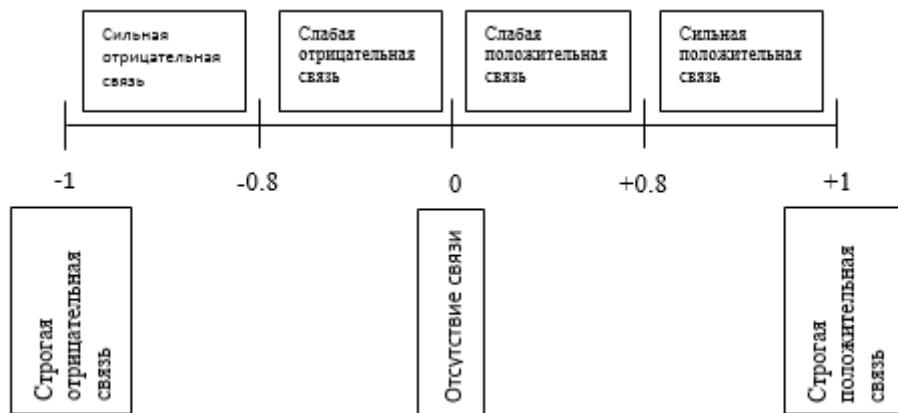


# ФУНКЦИЯ КОРРЕЛ (CORREL)

Корреляция — это статистический показатель, характеризующий силу статистической связи между двумя случайными величинами (наборами наблюдаемых данных).

Коэффициент корреляции любых величин всегда лежит в диапазоне от -1 до 1. На этом промежутке можно выделить следующие точки и интервалы:

- -1 — детерминированная (неслучайная) отрицательная связь. Одна величина растёт — другая падает, и наоборот. Связь строгая, то есть величины связаны напрямую.
- от -1 до -0,8 — сильная отрицательная связь. Вообще, чем больше коэффициент корреляции, тем сильнее связь.
- от -0,8 до 0 — слабая отрицательная связь. Значения ближе к нулю, скорее всего, означают полное отсутствие связи — такая корреляция может возникать случайно.
- 0 — полное отсутствие связи.
- от 0 до 0,8 — слабая положительная связь. Значения ближе к нулю, скорее всего, означают полное отсутствие связи — такая корреляция может возникать случайно.
- от 0,8 до 1 — сильная положительная связь.
- 1 — детерминированная (неслучайная) положительная связь показателей.



Коэффициент корреляции полезен для определения причинно-следственных связей. При этом связь может быть двусторонней (например, привлекательность соцсети и количество зарегистрированных пользователей — так называемый сетевой эффект). Чем привлекательнее соцсеть, тем больше в ней регистрируется людей. Утверждение верно и в обратную сторону: чем больше зарегистрированных пользователей, тем привлекательнее соцсеть.

Часто встречается односторонняя связь, например, сложность пожара и число пожарных, которые участвовали в его тушении. Из наблюдаемой положительной корреляции вряд ли стоит делать вывод, что сложность пожара зависит от числа

пожарных, или пытаться снизить сложность пожара, уменьшая число пожарных в команде. Тем не менее обратная взаимосвязь выглядит разумной. При этом помните, что наличие статистической взаимосвязи необязательно говорит о причинно-следственной связи. Корреляция бывает ложной. Блогер Дмитрий Чернышев в своём «Живом журнале»<sup>[1]</sup> приводит курьёзные примеры, в которых есть статистическая связь, но явно отсутствует причинно-следственная. Так, есть сильная корреляция между:

- потреблением моцареллы и количеством докторских степеней (один из самых известных примеров);
- потреблением сметаны и количеством мотоциклистов, погибших в ДТП;
- средним возрастом «Мисс Америка» и количеством людей, погибших от горячего пара.

Другим примером ложной корреляции может служить корреляция, возникающая из-за наличия общего тренда. Например, положительная корреляция между числом выехавших на отдых за рубеж и количеством произведённых турбин. Очевидно, что причинно-следственной связи между показателями нет. Но можно предположить, что они зависят от экономического роста в стране. Значения коэффициента корреляции, близкие к нулю, необязательно говорят об отсутствии причинно-следственной связи, только об отсутствии линейной зависимости. Взаимосвязь между величинами может быть более сложной.

Таким образом, при анализе статистической взаимосвязи в первую очередь нужно опираться на логическую объяснимость направления (положительная или отрицательная) и силы взаимосвязи. Если эмпирический опыт подтверждается корреляцией, можно смело её использовать. Если статистика не соответствует практике, необходимо использовать её с осторожностью.

Кстати, и наличие корреляции не означает наличия причинно-следственной связи. Возможно, обе величины связаны с какой-то третьей и поэтому коррелируют, не имея при этом причинно-следственной связи между собой.

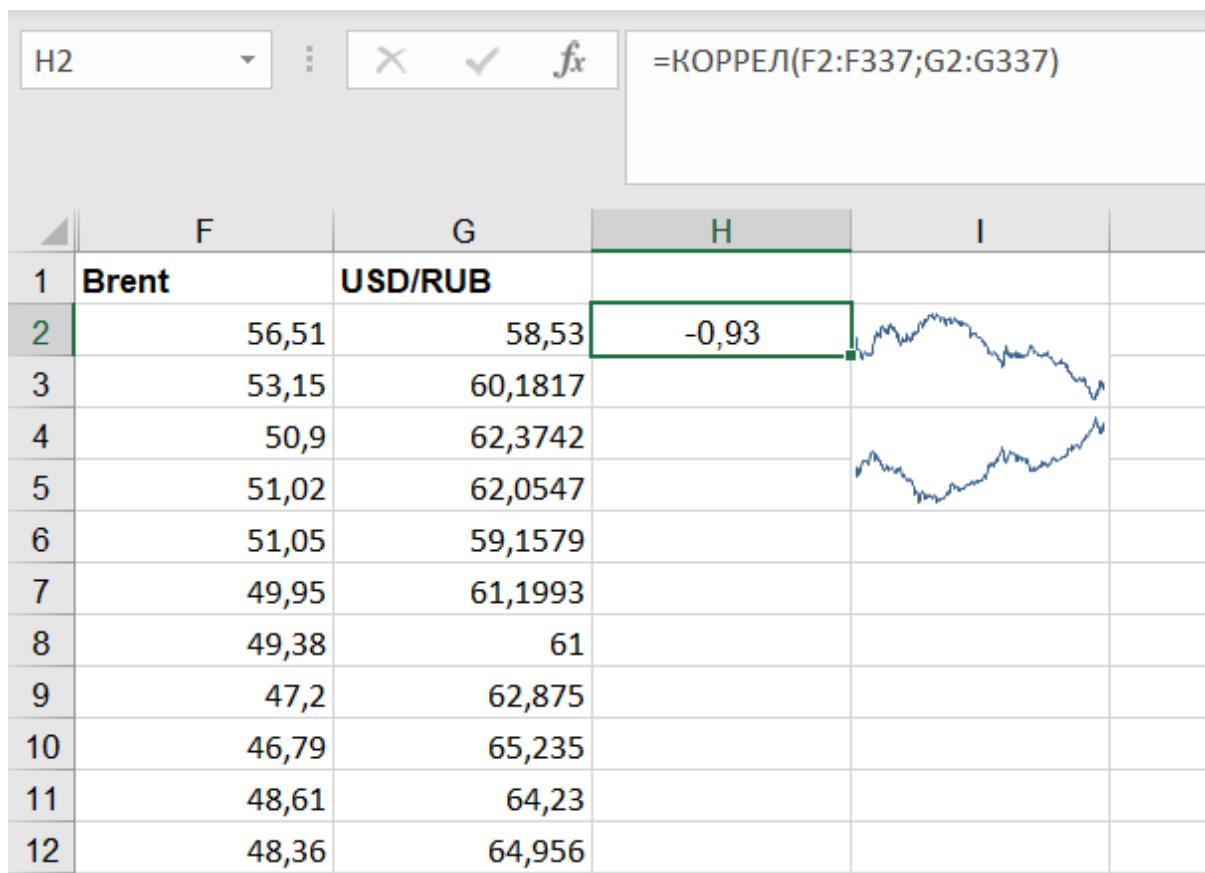
Для расчёта коэффициента корреляции в Google Таблицах есть функция КОРРЕЛ (CORREL). Её аргументы — это диапазоны с наблюдаемыми значениями показателей. В примере в столбцах А и В находятся случайные числа, сгенерированные с помощью функции СЛУЧМЕЖДУ (RANDBETWEEN). Коэффициент корреляции практически нулевой, что вполне естественно. Между двумя массивами случайных чисел связи нет.

C2 : X ✓ fx =КОРРЕЛ(A2:A204;B2:B204)

	A	B	C	D	E
1	Данные 1	Данные 2			
2	21	85	0,03		
3	57	18			
4	48	84			
5	64	79			
6	24	21			
7	47	32			
8	40	52			
9	81	92			
10	45	94			
11	96	24			
12	88	39			
13	23	33			
14	79	81			
15	39	90			
16	99	6			
17	11	41			
18	42	83			
19	90	85			
20	23	86			
21	53	49			
22	41	93			
23	53	69			
24	93	49			

Спарклайны (маленькие графики) добавлены для наглядности — видно, что это два случайных процесса.

В следующем примере мы рассчитываем корреляцию между ценой нефти марки Brent и курсом USD/RUB. Связь весьма сильная — -0,93:



На графиках видно, что показатели «отзеркаливают» друг друга. Когда один из них растёт — другой снижается, и наоборот.

---

[1] <http://mi3ch.livejournal.com/2559227.html>

## Функция СУММПРОИЗВ (SUMPRODUCT)

Функция СУММПРОИЗВ перемножает значения из нескольких столбцов и затем складывает эти произведения, избавляя нас от необходимости производить промежуточные вычисления. Столбцы должны быть одинакового размера.

Допустим, у вас есть таблица с ценами и количеством. Чтобы посчитать общую стоимость, не нужно рассчитывать её в каждой строке, а затем суммировать результаты. Оба этих шага можно произвести с помощью одной функции СУММПРОИЗВ.

C9      :    X    ✓    f<sub>x</sub>    =СУММПРОИЗВ(Таблица2[Цена];Таблица2[Количество])

	A	В	C	D	E	F
1	Наименование	Цена	Количество			
2	Планшет Apple iPad Mini 2019 256Gb Wi-Fi	97 000	1			
3	Acer Swift SF314-55-72FH (NX.H3WER.010)	65 800	2			
4	Монитор 21,5" Acer K222HQLCbid, FHD, IPS	9 600	4			
5	MSI GL63 8RC-469XRU (9S7-16P612-469)	34 200	2			
6	Lenovo Legion Y530-15 (81FV0027RU)	118 500	1			
7	Prestigio SmartBook 141 C2 (PSB141C02CF)	22 700	2			
8						
9			499 300			
10						