

Методы статистического анализа в Ms Excel для анализа социологических данных

Среднее арифметическое

Среднее арифметическое набора чисел определяется как их сумма, деленная на их количество.

Медиана

Медиана (от лат. *mediāna* — середина) 50-й перцентиль или квантиль 0,5 — статистика, которая делит ранжированную совокупность (вариационный ряд выборки) на две равные части: 50 % «нижних» членов ряда данных будут иметь значение признака не больше, чем медиана, а «верхние» 50 % — значения признака не меньше, чем медиана.

В Excel функция **Медиана**

Стандартное отклонение

Стандартное отклонение показывает, как распределены значения относительно среднего в выборке

$$STD = \sqrt{[(\sum (x - \bar{x})^2) / n]}$$

Если количество элементов в выборке превышает 30, то знаменатель дроби под корнем принимает значение $n-1$. Иначе используется n .

В Excel функция Стандоткл (количество элементов в выборке более 30)

Пример. Количество участников в общественном мероприятии

						Среднее
33	31	32	36	31	31	32,3
22	34	58	52	10	21	32,8

Пример:

						Среднее	Станд откл
33	31	32	36	31	31	32,3	2,0
22	34	58	52	10	21	32,8	18,9

Пошагово вычисление стандартног о отклонения:

ВЫЧИСЛЯЕМ СРЕДНЕЕ
АРИФМЕТИЧЕСКОЕ ВЫБОРКИ
ДАННЫХ

ОТНИМАЕМ ЭТО СРЕДНЕЕ ОТ
КАЖДОГО ЭЛЕМЕНТА
ВЫБОРКИ

ВСЕ ПОЛУЧЕННЫЕ РАЗНИЦЫ
ВОЗВОДИМ В КВАДРАТ

СУММИРУЕМ ВСЕ
ПОЛУЧЕННЫЕ КВАДРАТЫ

ДЕЛИМ ПОЛУЧЕННУЮ
СУММУ НА КОЛИЧЕСТВО
ЭЛЕМЕНТОВ В ВЫБОРКЕ (ИЛИ
НА $n-1$, ЕСЛИ $n > 30$)

ВЫЧИСЛЯЕМ КВАДРАТНЫЙ
КОРЕНЬ ИЗ ПОЛУЧЕННОГО
ЧАСТНОГО (ИМЕНУЕМОГО
ДИСПЕРСИЕЙ)

Дисперсия

Мера рассеивания, отклонения от среднего.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Ранги и ранжирование

Рангом наблюдения называют тот номер, который получит это наблюдение в упорядоченной совокупности всех данных — после их упорядочения по определенному правилу (например, от меньших значений к большим или наоборот).

Пример

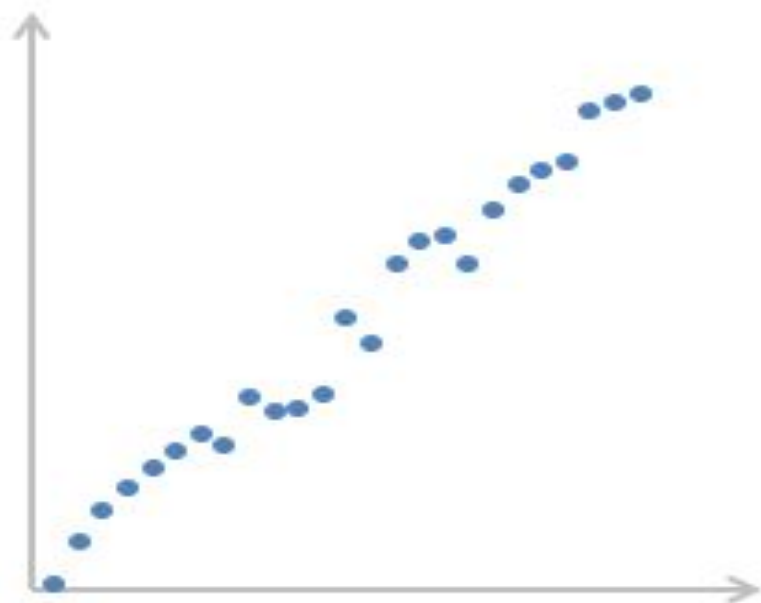
6, 17, 14, 5, 12

Процедура перехода от совокупности наблюдений к последовательности их рангов называется ранжированием. Результат ранжирования называется ранжировкой.

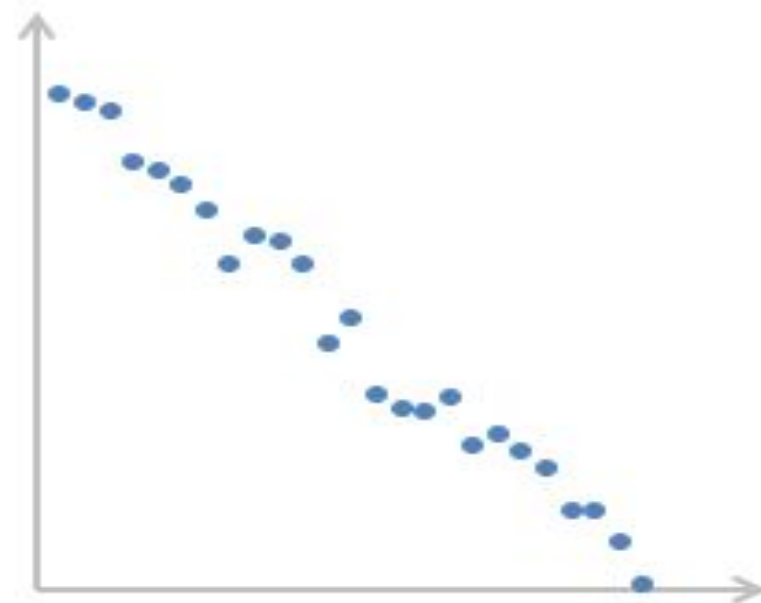
2, 5, 4, 1, 3

Коэффициент корреляции

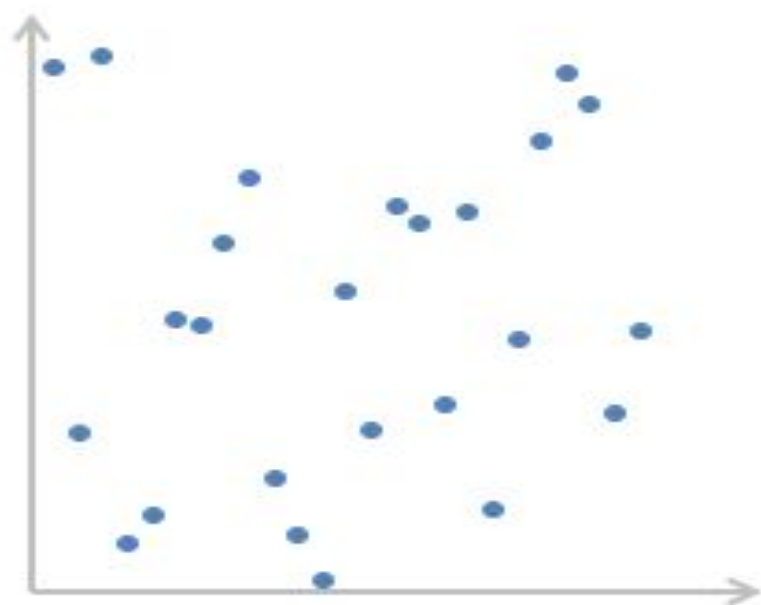
Корреляция измеряет мощность и направление связи между x и y .



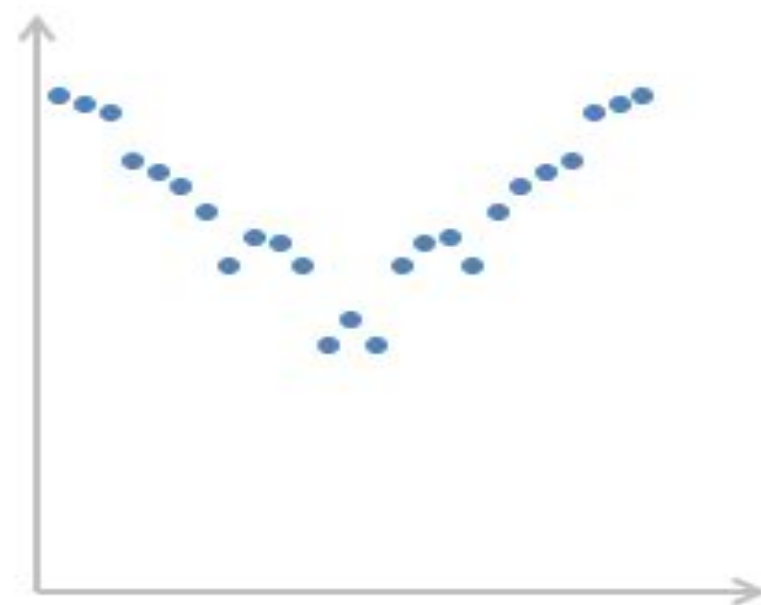
(A) Положительная
линейная корреляция



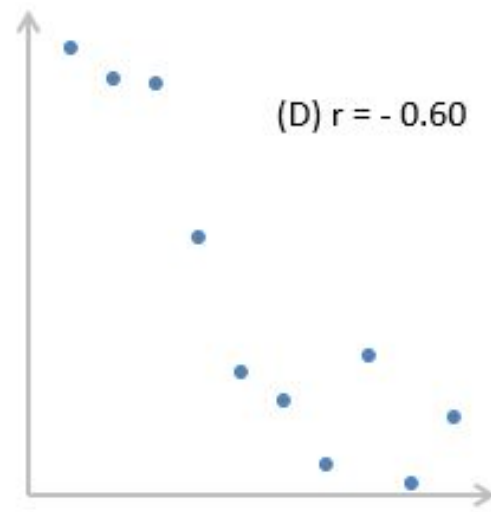
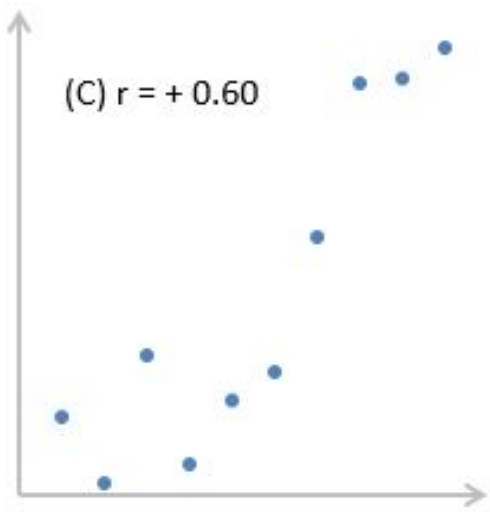
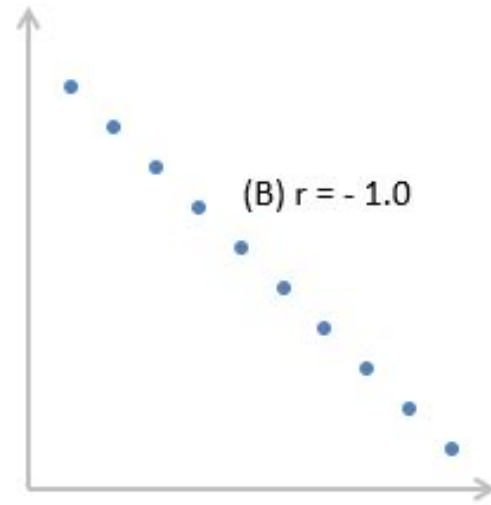
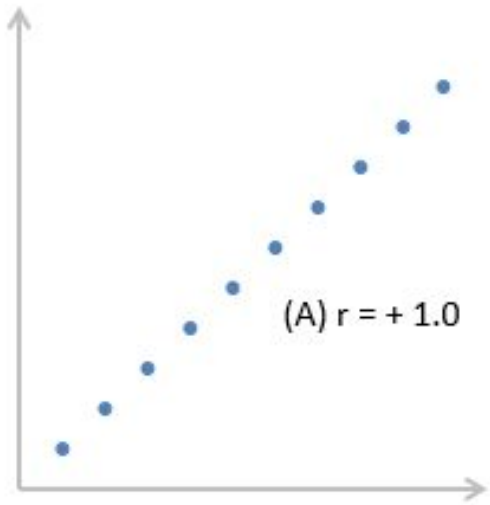
(B) Отрицательная
линейная корреляция




(C) Отсутствие корреляции



(D) Нелинейная корреляция



Коэффициент корреляции


$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

Коэффициент корреляции

	A	B	C	D	E
	Часы изучения	Экзамен	Оценка	x^2	y^2
1	x	y	xy	x^2	y^2
2	3	86	258	9	7396
3	5	95	475	25	9025
4	4	92	368	16	8464
5	4	83	332	16	6889
6	2	78	156	4	6084
7	3	82	246	9	6724
8	$\Sigma x=21$	$\Sigma y=516$	$\Sigma xy=1835$	$\Sigma x^2=79$	$\Sigma y^2=44582$

$$r = \frac{6(1835) - (21)(516)}{\sqrt{[6(79) - (21)^2][6(44582) - (516)^2]}}$$

$$r = \frac{174}{\sqrt{[33][1236]}} = 0,862$$