**15. Корреляция. Корреляционный анализ.**

**Корреляционный анализ** — это метод обработки статистических данных, заключающийся в изучении коэффициентов корреляции между переменными.

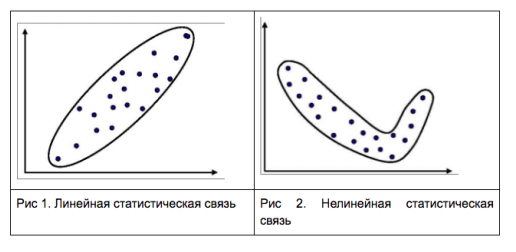
**Проще говоря:** связь между результативными и факторными переменными (урожайностью какой-либо культуры и количеством осадков, ростом и весом человека в однородных группах по полу и возрасту, частотой пульса и температурой тела и т.д.).

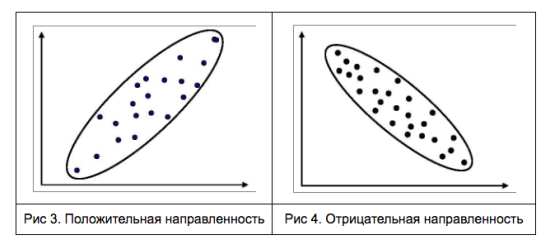
**Ложная корреляция —** корреляция по отношению к признакам абсурдных по отношению друг к другу (например: уровень IQ и размер ноги).

**Задачи корреляционного анализа:**

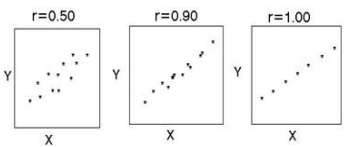
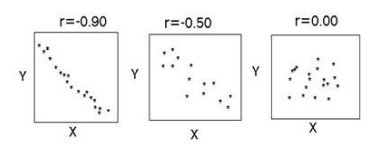
* получение информации об одной из искомых переменных с помощью другой
* определение тесноты связи между исследуемыми переменными
* выявление факторов, оказывающих наибольшее влияние на результативный признак
* выявление неизученных ранее причин связей
* построение корреляционной модели с ее параметрическим анализом
* исследование значимости параметров связи и их интервальная оценка.

**Корреляционное поле** (или диаграмма рассеяния) является графической зависимостью между результатами измерений двух признаков. (Ось х и у. Например: х - курение, у - смерность)





r - коэффициент корреляции

****

*Обратная корреляция Прямая корреляция*

Корреляции нет - точки расположены хаотично

Средняя степень связи - точки более-менее равномерно удалены от медианы.

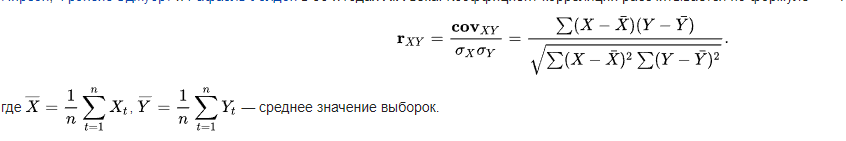
Сильная связь - стремится к прямой и при r=1, ровная линия

Знак коэффициента корреляции определяет направленность взаимосвязи: минус – отрицательная, плюс – положительная

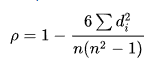


Если распределение является нормальным законом — результатом корреляционного анализа выступают коэффициенты корреляции Пирсона, либо, в случае, если признаки не подчиняются этому закону — коэффициент ранговой корреляции Спирмена.

**Коэффициент корреляции Бравэ-Пирсона** —линейный коэффициент корреляции



**Коэффициент корреляции рангов Спирмена (**rs**)** — ранговый коэффициент корреляции.



Значение коэффициента меняется от −1 (последовательности рангов полностью противоположны) до +1 (последовательности рангов полностью совпадают). Нулевое значение показывает, что признаки независимы.

**Ограничения использования коэффициента корреляции:**

* Нелинейность связи. Решения:

1. Найти точку перегиба и разделить на две части

2. Отказаться от использования коэффициента корреляции.

3. Если выявленная связь является монотонной, то целесообразно использовать ранговые коэффициенты корреляции.

* Дисперсионные выбросы, асимметричность распределения (значения r разбросаны от -1 до 1).Решения:

1. Исключать дисперсионные выбросы из выборки

2. Использовать ранговые коэффициенты корреляции

* Влияние третьей переменной.

**В Exel:**

1. КОРРЕЛ [CORREL] (массив1; массив2) - Коэффициент корреляции. ПИРСОН (PEARSON) с теми же массивами - Линейный коэффициент корреляции

2. СТЬЮДРАСПОБР (вероятность; степени\_свободы)

Оценка значимости коэффициента парной корреляции с использованием t критерия Стьюдента. Рассчитанное значение t-критерия сравнивается с табличной (критической) величиной данного показателя из соответствующей таблицы значений рассматриваемого параметра с учетом заданного уровня значимости и числа степеней свободы.

3. Матрица коэффициентов парной корреляции. Анализ осуществляется с помощью средства «Анализ данных», в котором выбирается «Корреляция».