

Коэффициенты корреляции позволяют ответить на вопрос:

Влияет ли количественный фактор на количественный отклик?

Коэффициент корреляции (r) – мера линейной связи между количественными факторами и откликом (коэф корреляции измеряется от -1 до 1).

Чем ближе коэффициент корреляции к 1, тем ближе связь.

Если коэффициент корреляции(r):

Меньше 0.25 – корреляция слабая

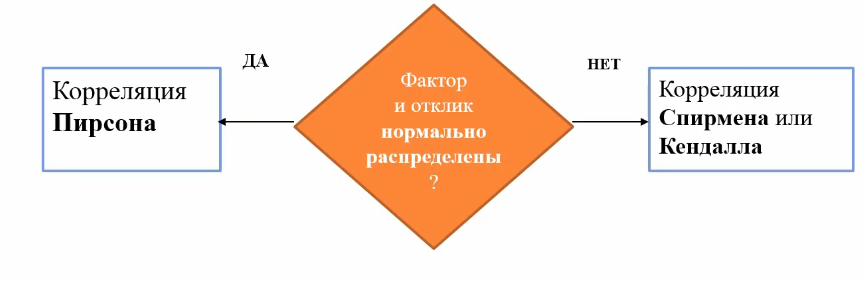
От 0.25 до 0.75 – связь умеренная

Более 0.75 – связь сильная.

Если коэф корреляции больше нуля, то связь прямая, это значит, что с ростом фактора увеличивается и отклик.

Если коэф корреляции меньше нуля, то связь обратная, это значит, с ростом фактора отклик уменьшается.

Если коэф корреляции близок к нулю и уровень значимости (p) больше 0.05, то связи между фактором и откликом нет.



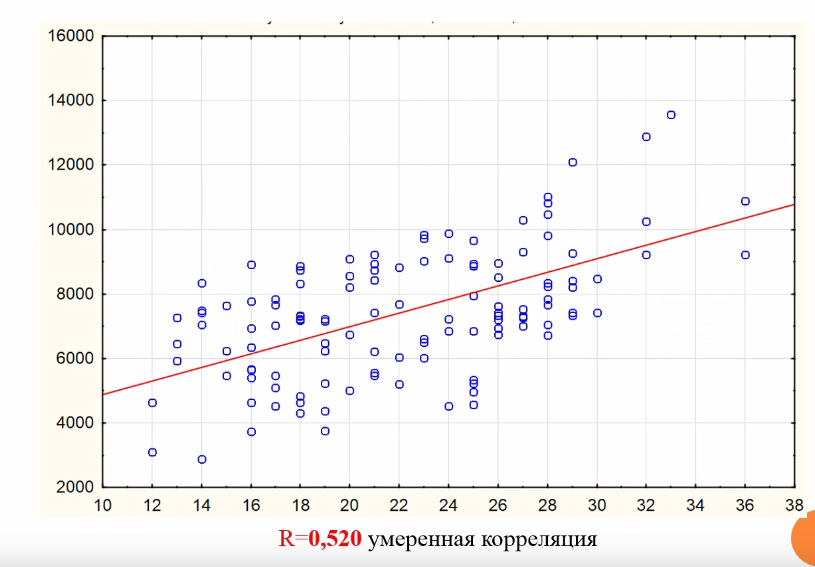
Представлена схема выбора корреляционного исследования.

Для того чтобы выбрать метод, нам нужно ответить на вопрос:

Нормально ли распределены фактор и отклик?

Для использования корреляции Пирсона и фактор и отклик должны быть распределены нормально.

Если хотя бы один из них ненормально распределен, то мы используем либо корреляцию Спирмена или Кендалла (чаще применяют корреляцию Спирмена, метод Кендалла использует, если хотя бы один из факторов порядковый)

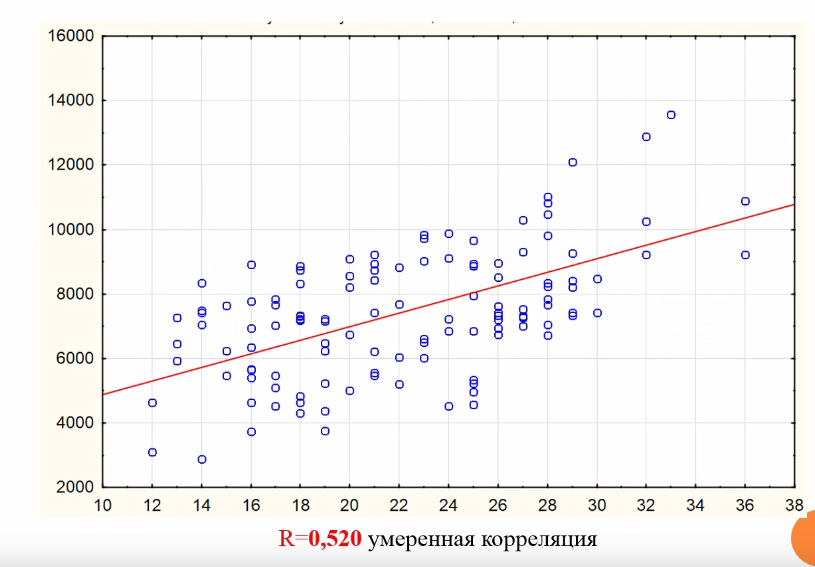


В качестве графического представления используют диаграмму рассеяния с линией регрессии.

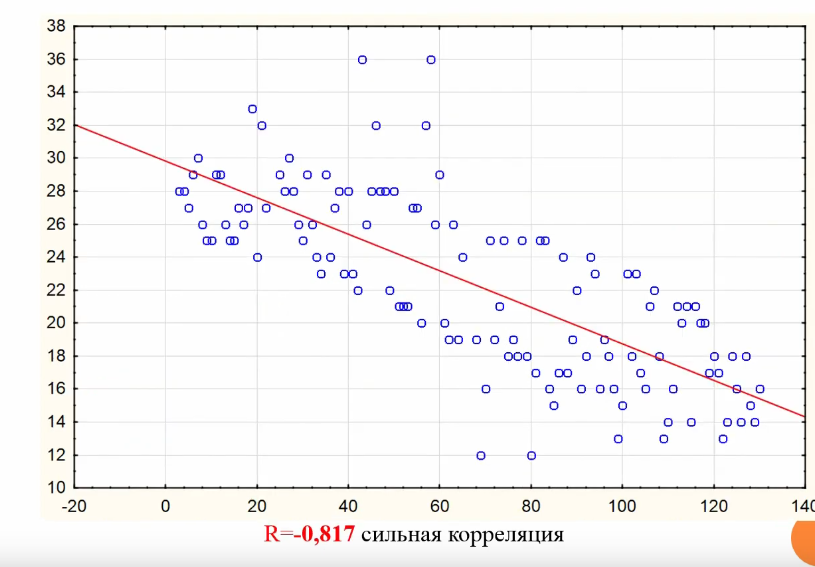
Линия регрессии точнее всего отображает распределение экспериментальных точек на диаграмме рассеяния.

Крутизна ее наклона характеризует зависимость между двумя переменными.

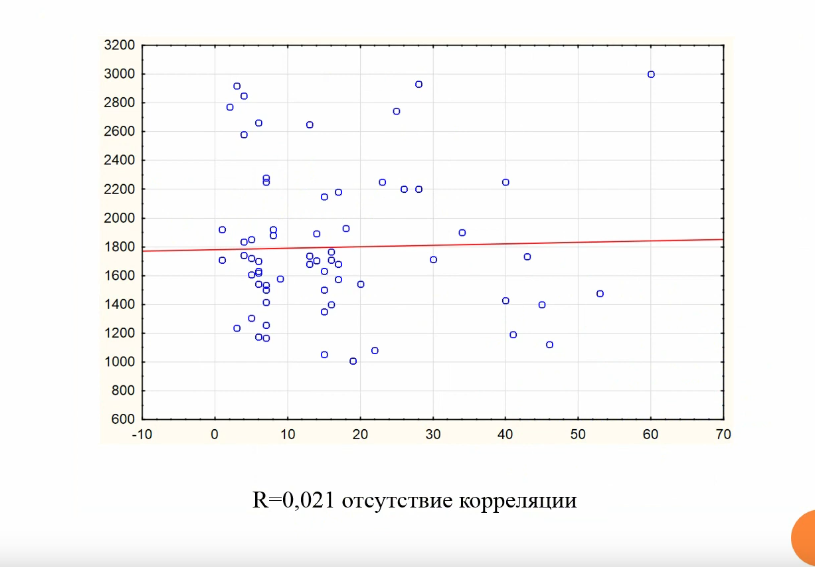
На этом рисунке пример умеренной корреляции с коэффициентом 0.5:



На этой диаграмме пример сильной обратной корреляции с коэфом корреляции -0.8:

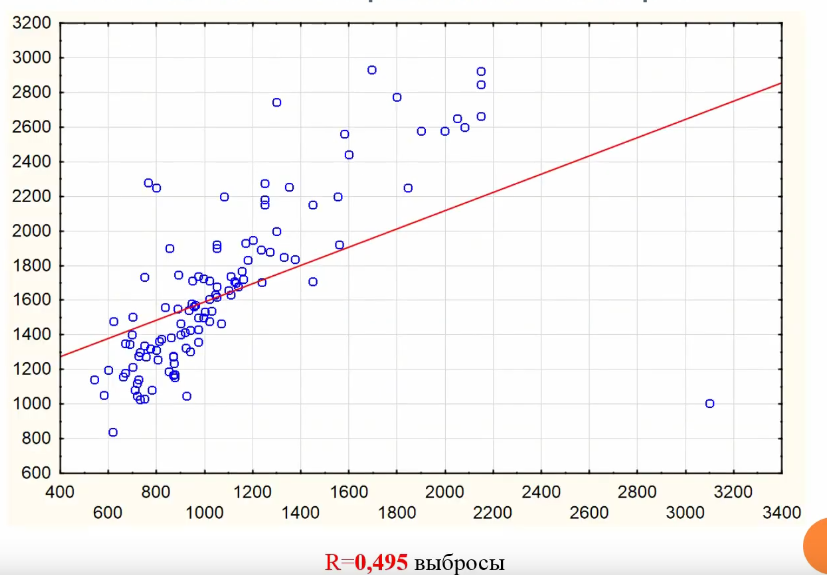


На этой диаграмме пример нулевой корреляции (отсутствием корреляции) с коэф корреляции 0.02:



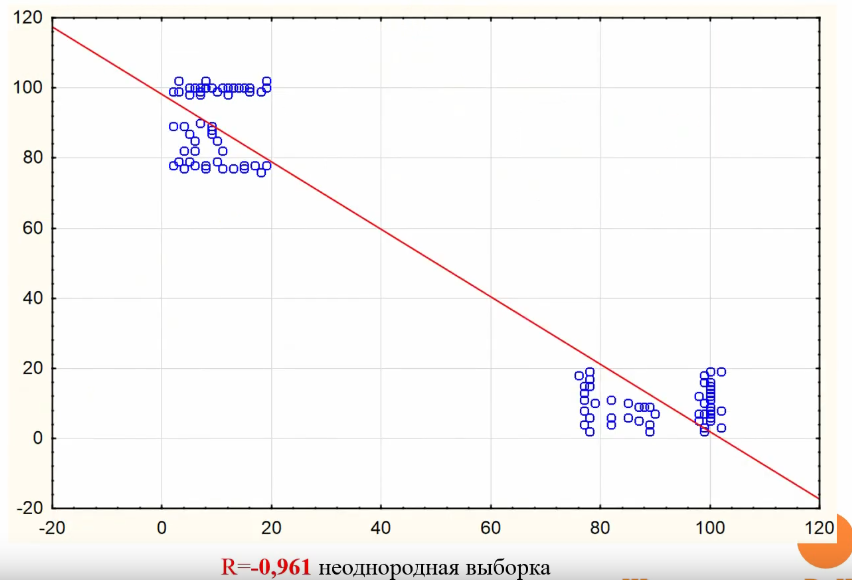
Эта диаграмма иллюстрирует то, как влияют грубые ошибки на линию регрессии:

Мы видим в правой нижней части выброс (скорее даже грубую ошибку) который существенно отклоняет линию регрессии от основного направления эмпирических точек данных. Если бы данного выброса не было, коэф корреляции скорее был бы сильным, а не умеренным.



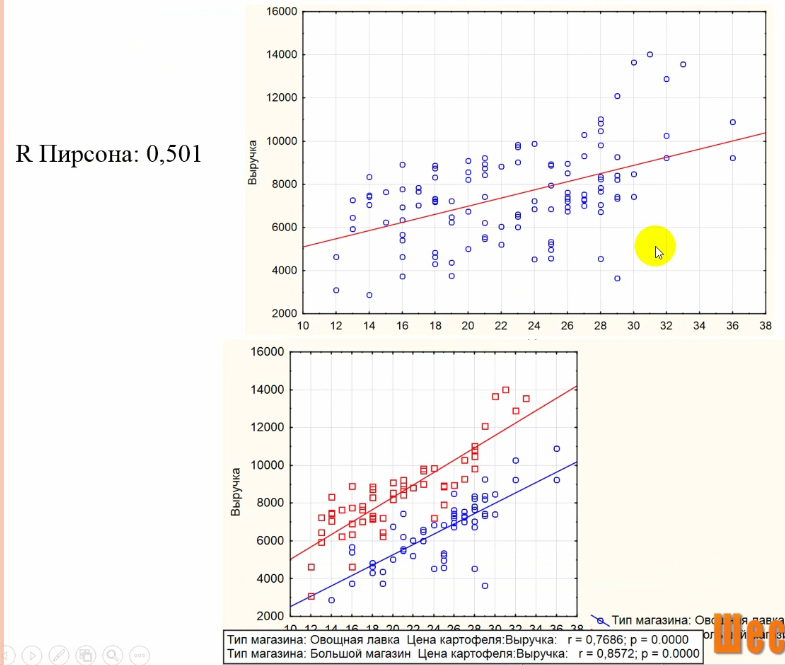
В следующем примере мы видим неоднородную выборку:

Это сильная обратная связь, однако мы видим две изолированные группы данных через которые проведена линия регрессии, однако если бы мы выборку поделили, то корреляция могла бы и не быть.



Приведенные примеры говорят о том, что построение диаграммы рассеяния с линией регрессии – это необходимый этап корреляционного исследования, если ориентироваться только на коэфы корреляции, то легко ошибиться, во первых, во влиянии выбросов, во вторых, не заметить действие третьего фактора.

Пример:



Предположим, нужно выяснить, как влияет цена килограмма картофеля на выручку от его продажи в торговой точке.

На первой диаграмме рассеяния представлена зависимость выручки от цены и рассчитан коэф корреляции пирсона (R)

ВСЕ ДАЛЬШЕ ЛЕНЬ ДОПОЛНЯТЬ ДАВАЙТЕ САМИ