**Корреляционный анализ для бизнеса**

[**Корреля́ция**](http://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D1%80%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D0%B8%D1%8F)— статистическая взаимосвязь двух или нескольких случайных величин. При этом изменения одной или нескольких из этих величин приводят к систематическому изменению другой или других величин. Математической мерой корреляции двух случайных величин служит коэффициент корреляции – k. k изменяется от -1 до 1. k =-1 устанавливает обратную пропорциональность; k=1 – прямую пропорциональность; k=0 – соответствует отсутствию какой-либо зависимости между случайными величинами; чем ближе k к 1 (или к -1), тем сильнее зависимость; величину k можно интерпретировать следующим образом: «какую часть изменений одной случайной величины определяют изменения другой случайной величины». Например (см. ниже), коэффициент корреляции между бюджетом клуба и его местом в чемпионате страны k=-0,63. То есть, на 63% успех в чемпионате определяется бюджетом.

В своей работе менеджера я постоянно использую корреляционный анализ. Например, недавно я обнаружил, как и предполагал ☺, сильную корреляцию между стажем работы в компании и заработной платой (ЗП);[[1]](#footnote-1) см. также [Excel](Корреляционный%20анализ%20в%20бизнесе.xlsx)-файл:

Коэффициент корреляции между стажем и ЗП составил 0,6, что указывает на значительную корреляцию.

При этом надо понимать, что корреляция не указывает на причинно-следственные связи. Например, на основании полученного значения корреляции можно сказать, что:

1. Чем выше ЗП, тем сотрудники дольше работают в компании… или…
2. Чем сотрудники дольше работают в компании, тем у них выше ЗП… или…
3. Существует какой-то третий фактор, который и является причиной того, что стаж и ЗП связаны.

Сам по себе факт корреляционной зависимости не даёт основания утверждать, какая из переменных предшествует или является причиной изменений, или что переменные вообще причинно связаны между собой, например, ввиду действия третьего фактора. Забавная история на эту тему. Мой сын, готовящийся к сдаче ЕГЭ, сообщил, что в 24-м вопросе по алгебре нужно выбирать 4-й вариант. Я удивленно спросил: «Почему?». «Потому что за последние пять лет правильный ответ на 24-й вопрос всегда был под №4», – ответил сын. ☺

Не являясь доказательством причинно-следственной связи, корреляция, тем не менее, может служить инструментом для выдвижения гипотез о наличии такой связи. Наличие корреляции заставляет искать «общую причину» хорошо коррелированных явлений, или неизвестный третий (четвертый, пятый…) фактор. И, наоборот, отсутствие корреляции, скорее, указывает и на отсутствие причинно-следственных связей.

Вот еще один пример корреляционной зависимости между бюджетом клуба Континентальной хоккейной лиги и местом, занятым этим клубом в регулярном чемпионате.

Коэффициент корреляции –0,63 (отрицательный, поскольку, зависимость обратная: чем больше бюджет, тем меньше номер места).

Следующий пример из области логистики. Современные склады разбивают на зоны (участки):

* например, на основе [АВС-анализа](http://baguzin.ru/wp/?p=310#more-310): чем выше оборачиваемость товаров, тем ближе к зонам приемки и отгрузки они размещаются, или…
* на основе типов товаров (отдельно консервы, стройматериалы, автозапчасти…)

Что делать, если пришел товар, который нельзя разместить в «родной» зоне? Каждый решает эту проблему по-разному. Один из вариантов – положить в «чужую» зону. Со временем товар начинает «расползаться» по складу. Корреляционный анализ позволяет оценить качество стратегий размещения. Если участок **товара** (участок, куда товар должен был разместиться) коррелирует с участком **ячейки** (участок, определенный в системе данному складскому месту), значит стратегии размещения и само разбиение на участки сделаны корректно. Первоначально мы разбили весь номенклатурный ряд на 49 групп (участков), и вот какую картину получили спустя несколько месяцев (см. лист «склад1» Excel-файла):



По вертикали слева указаны номера участков **ячеек**; по горизонтали вверху – номера участков **товаров**; желтым цветом выделены ячейки, для которых эти номера совпадают, то есть товар лёг в «свою» ячейку. Получилось, что только 18% коробок лежит на «родных» местах. Вариабельность объема хранения отдельных категорий товара столь высока, что фиксированное разбиение склада на зоны не позволяет в течение длительного времени поддерживать выбранную стратегию размещения. Зачем иметь столько категорий товара, если не удается размещать его «правильно»? Мы сократили число зон до 12, и спустя некоторое время получили совершенно иную ситуацию (см. лист «склад2» Excel-файла):



На своих местах теперь располагалось 63,5% коробок.

**Вывод.** Корреляционный анализ является весьма полезным в современной бизнес-практике; он позволяет «добыть пищу» для дальнейшего изучения, и принятия качественных продуманных управленческих решений. Однако не следует, в случае наличия корреляции, автоматически считать причинно-следственные связи установленными.

1. Диапазоны не соответствуют истинным цифрам, поскольку это, всё же, коммерческая тайна ☺, но всё остальное – реальный пример. [↑](#footnote-ref-1)