КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ В EXCEL: ИНСТРУКЦИЯ ВЫПОЛНЕНИЯ

Регрессионный и корреляционный анализ – статистические методы исследования. Это наиболее распространенные способы показать зависимость какого-либо параметра от одной или нескольких независимых переменных.

Ниже на конкретных практических примерах рассмотрим эти два очень популярные в среде экономистов анализа. А также приведем пример получения результатов при их объединении.

РЕГРЕССИОННЫЙ АНАЛИЗ В EXCEL

Показывает влияние одних значений (самостоятельных, независимых) на зависимую переменную. К примеру, как зависит количество экономически активного населения от числа предприятий, величины заработной платы и др. параметров. Или: как влияют иностранные инвестиции, цены на энергоресурсы и др. на уровень ВВП.

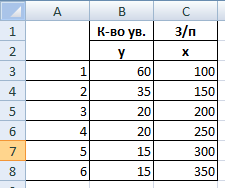
Результат анализа позволяет выделять приоритеты. И основываясь на главных факторах, прогнозировать, планировать развитие приоритетных направлений, принимать управленческие решения.

Регрессия бывает:

* линейной (у = а + bx);
* параболической (y = a + bx + cx2);
* экспоненциальной (y = a \* exp(bx));
* степенной (y = a\*x^b);
* гиперболической (y = b/x + a);
* логарифмической (y = b \* 1n(x) + a);
* показательной (y = a \* b^x).

Рассмотрим на примере построение регрессионной модели в Excel и интерпретацию результатов. Возьмем линейный тип регрессии.

Задача. На 6 предприятиях была проанализирована среднемесячная заработная плата и количество уволившихся сотрудников. Необходимо определить зависимость числа уволившихся сотрудников от средней зарплаты.



Модель линейной регрессии имеет следующий вид:

*У = а0 + а1х1 +…+акхк.*

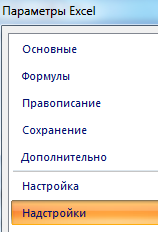
Где а – коэффициенты регрессии, х – влияющие переменные, к – число факторов.

В нашем примере в качестве У выступает показатель уволившихся работников. Влияющий фактор – заработная плата (х).

В Excel существуют встроенные функции, с помощью которых можно рассчитать параметры модели линейной регрессии. Но быстрее это сделает надстройка «Пакет анализа».

Активируем мощный аналитический инструмент:

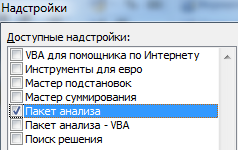
1. Нажимаем кнопку «Офис» и переходим на вкладку «Параметры Excel». «Надстройки».



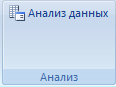
1. Внизу, под выпадающим списком, в поле «Управление» будет надпись «Надстройки Excel» (если ее нет, нажмите на флажок справа и выберите). И кнопка «Перейти». Жмем.

Управление.

1. Открывается список доступных надстроек. Выбираем «Пакет анализа» и нажимаем ОК.

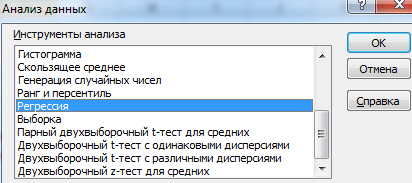


После активации надстройка будет доступна на вкладке «Данные».

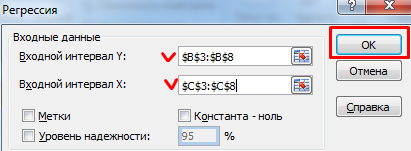


Теперь займемся непосредственно регрессионным анализом.

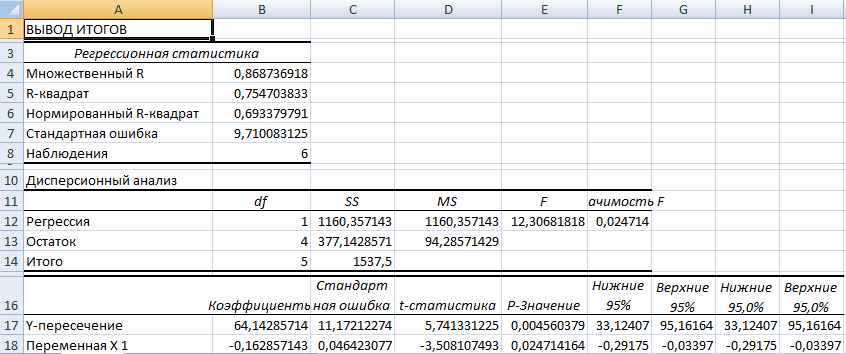
1. Открываем меню инструмента «Анализ данных». Выбираем «Регрессия».



1. Откроется меню для выбора входных значений и параметров вывода (где отобразить результат). В полях для исходных данных указываем диапазон описываемого параметра (У) и влияющего на него фактора (Х). Остальное можно и не заполнять.



1. После нажатия ОК, программа отобразит расчеты на новом листе (можно выбрать интервал для отображения на текущем листе или назначить вывод в новую книгу).



В первую очередь обращаем внимание на R-квадрат и коэффициенты.

R-квадрат – коэффициент детерминации. В нашем примере – 0,755, или 75,5%. Это означает, что расчетные параметры модели на 75,5% объясняют зависимость между изучаемыми параметрами. Чем выше коэффициент детерминации, тем качественнее модель. Хорошо – выше 0,8. Плохо – меньше 0,5 (такой анализ вряд ли можно считать резонным). В нашем примере – «неплохо».

Коэффициент 64,1428 показывает, каким будет Y, если все переменные в рассматриваемой модели будут равны 0. То есть на значение анализируемого параметра влияют и другие факторы, не описанные в модели.

Коэффициент -0,16285 показывает весомость переменной Х на Y. То есть среднемесячная заработная плата в пределах данной модели влияет на количество уволившихся с весом -0,16285 (это небольшая степень влияния). Знак «-» указывает на отрицательное влияние: чем больше зарплата, тем меньше уволившихся. Что справедливо.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ В EXCEL

Корреляционный анализ помогает установить, есть ли между показателями в одной или двух выборках связь. Например, между временем работы станка и стоимостью ремонта, ценой техники и продолжительностью эксплуатации, ростом и весом детей и т.д.

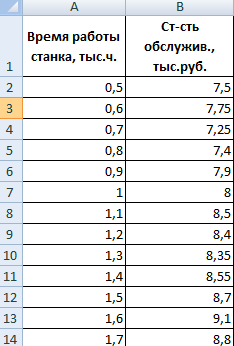
Если связь имеется, то влечет ли увеличение одного параметра повышение (положительная корреляция) либо уменьшение (отрицательная) другого. Корреляционный анализ помогает аналитику определиться, можно ли по величине одного показателя предсказать возможное значение другого.

Коэффициент корреляции обозначается r. Варьируется в пределах от +1 до -1. Классификация корреляционных связей для разных сфер будет отличаться. При значении коэффициента 0 линейной зависимости между выборками не существует.

Рассмотрим, как с помощью средств Excel найти коэффициент корреляции.

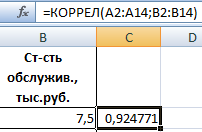
Для нахождения парных коэффициентов применяется функция КОРРЕЛ.

**Задача:** Определить, есть ли взаимосвязь между временем работы токарного станка и стоимостью его обслуживания.



Ставим курсор в любую ячейку и нажимаем кнопку fx.

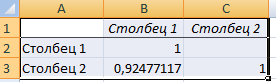
1. В категории «Статистические» выбираем функцию КОРРЕЛ.
2. Аргумент «Массив 1» - первый диапазон значений – время работы станка: А2:А14.
3. Аргумент «Массив 2» - второй диапазон значений – стоимость ремонта: В2:В14. Жмем ОК.



Чтобы определить тип связи, нужно посмотреть абсолютное число коэффициента (для каждой сферы деятельности есть своя шкала).

Для корреляционного анализа нескольких параметров (более 2) удобнее применять «Анализ данных» (надстройка «Пакет анализа»). В списке нужно выбрать корреляцию и обозначить массив. Все.

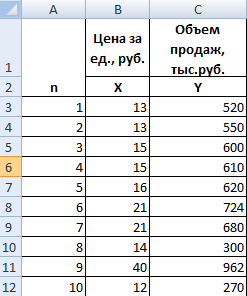
Полученные коэффициенты отобразятся в корреляционной матрице. Наподобие такой:



КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

На практике эти две методики часто применяются вместе.

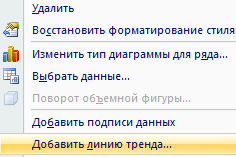
**Пример:**



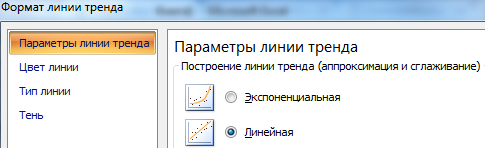
1. Строим корреляционное поле: «Вставка» - «Диаграмма» - «Точечная диаграмма» (дает сравнивать пары). Диапазон значений – все числовые данные таблицы.



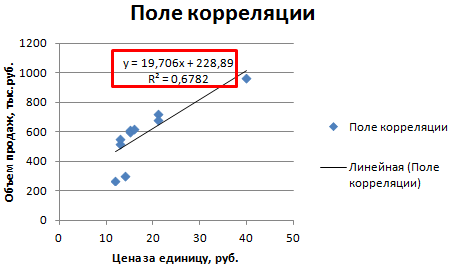
1. Щелкаем левой кнопкой мыши по любой точке на диаграмме. Потом правой. В открывшемся меню выбираем «Добавить линию тренда».



1. Назначаем параметры для линии. Тип – «Линейная». Внизу – «Показать уравнение на диаграмме».



1. Жмем «Закрыть».



Теперь стали видны и данные регрессионного анализа.