**Что такое линейная регрессионная модель и зачем это нужно**

Это наиболее распространенный способ показать зависимость какой-то переменной от других, например, как зависит **уровень ВВП** от величины **иностранных инвестиций** или от **кредитной ставки Нацбанка** или от **цен на ключевые энергоресурсы**.

Моделирование позволяет показать величину этой зависимости (коефициенты), благодаря которым можно делать непосредственно прогноз и осуществлять какое-то планирование, опираясь на эти прогнозы. Также, опираясь на регрессионный анализ, можно принимать управленческие решения направленные на стимулирование приоритетных причин влияющих на конечный результат, собственно модель и поможет выделить эти приоритетные факторы.

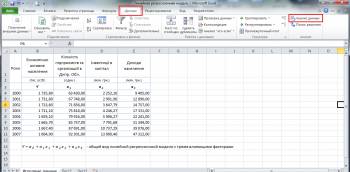
Общий вид модели линейной регрессии:

***Y=a0+a1x1+...+akxk***

где ***a*** — параметры (коэффициенты) регрессии, ***x*** — влияющие факторы, ***k*** — количество факторов модели.

**Исходные данные**

Среди исходных данных нам необходим некий набор данных, который бы представлял из себя несколько последовательных или связанных между собой величин итогового параметра Y (например, ВВП) и такое же количество величин показателей, влияние которых мы изучаем (например, иностранные инвестиции).

[](http://archie-goodwin.net/_ld/3/49502446.png)

На рисунке выше показана таблица с этими самыми исходными данными, в качестве Y выступает показатель экономически активного населения, а количество предприятий, размер инвестиций в капитал и доходов населения - это влияющие факторы, то бишь иксы.

По рисунку также можно сделать ошибочный вывод, что речь в моделировании может идти только о динамических рядах, то есть моментным рядам зафиксированных последовательно во времени, но это не так, с тем же успехом можно моделировать и в разрезе структуры, например, величины указанные в таблице могут быть разбиты не годам, а по областям.

Для построения адекватных линейных моделей желательно чтобы исходные данные не имели сильных перепадов или обвалов, в таких случаях желательно проводить сглаживание, но о сглаживании поговорим в следующий раз.

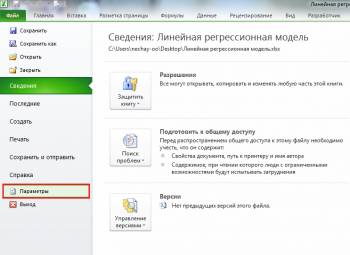
**Пакет анализа**

Параметры модели линейной регрессии можно рассчитать и вручную с помощью Метода наименьших квадратов (МНК), но это довольно затратно по времени. Немного быстрее это можно посчитать по этому же методу с помощью применения формул в Excel, где сами вычисления будет делать программа, но проставлять формулы все равно придется вручную.

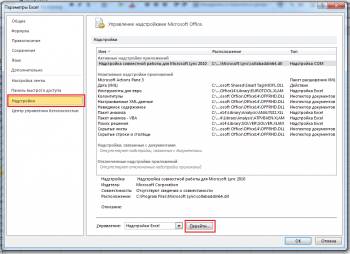
В Excel есть надстройка **Пакет анализа**, который является довольно мощным инструментом в помощь аналитику. Этот инструментарий, помимо всего прочего, умеет рассчитывать параметры регрессии, по тому же МНК, всего в несколько кликов, собственно, о том как этим инструментом пользоваться дальше и пойдет речь.

**Активируем Пакет анализа**

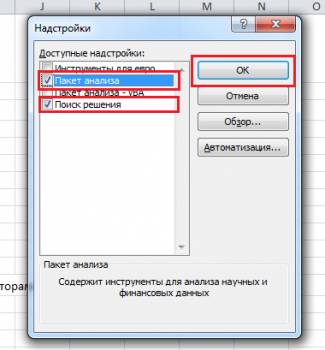
По умолчанию эта надстройка отключена и в меню вкладок вы ее не найдете, поэтому пошагово рассмотрим как ее активировать.

[](http://archie-goodwin.net/_ld/3/50994968.png)

В эксель, слева вверху, активируем вкладку **Файл**, в открывшемся меню ищем пункт **Параметры**и кликаем на него.

[](http://archie-goodwin.net/_ld/3/61235619.png)

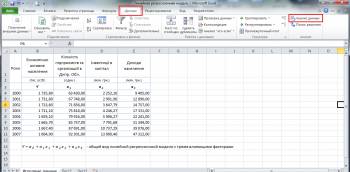
В открывшемся окне, слева, ищем пункт **Надстройки**и активируем его, в этой вкладке внизу будет выпадающий список управления, где по умолчанию будет написано **Надстройки Excel**, справа от выпадающего списка будет кнопка **Перейти**, на нее и нужно нажать.

[](http://archie-goodwin.net/_ld/3/46121208.png)

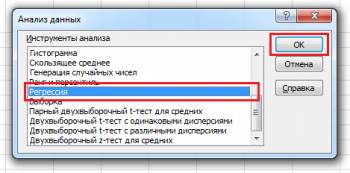
Всплывающее окошко предложит выбрать доступные надстройки, в нем необходимо поставить галочку напротив **Пакет анализа** и заодно, на всякий случай, **Поиск решения** (тоже полезная штука), а затем подтвердить выбор кликнув по кнопочке **ОК**.

**Инструкция по поиску параметров линейной регрессии с помощью Пакета анализа**

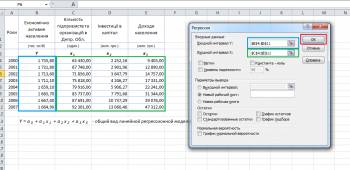
После активации надстройки Пакета анализа она будет всегда доступна во вкладке главного меню **Данные** под ссылкой **Анализ данных**

[](http://archie-goodwin.net/_ld/3/49502446.png)

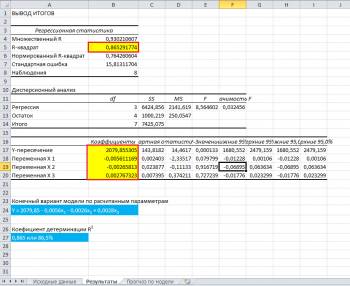
В активном окошке инструмента **Анализа данных** из списка возможностей ищем и выбираем **Регрессия**

[](http://archie-goodwin.net/_ld/3/01911606.png)

Далее откроется окошко для настройки и выбора исходных данных для вычисления параметров регрессионной модели. Здесь нужно указать интервалы исходных данных, а именно описываемого параметра (Y) и влияющих на него факторов (Х), как это на рисунке ниже, остальные параметры, в принципе, необязательны к настройке.

[](http://archie-goodwin.net/_ld/3/68950150.png)

После того как выбрали исходные данные и нажали кнопочку ОК, Excel выдает расчеты на новом листе активной книги (если в настройках не было выставлено иначе), эти расчеты имеют следующий вид:

[](http://archie-goodwin.net/_ld/3/40929298.png)

Ключевые ячейки залил желтым цветом именно на них нужно обращать внимание в первую очередь, остальные параметры значимость также немаловажны, но их детальный разбор требует пожалуй отдельного поста.

Итак, **0,865** - это **R2** - коэффициент детерминации, показывающий что на 86,5% расчетные параметры модели, то есть сама модель, объясняют зависимость и изменения изучаемого параметра - **Y** от исследуемых факторов - **иксов**. Если утрировано, то это показатель качества модели и чем он выше тем лучше. Понятное дело, что он не может быть больше 1 и считается неплохо, когда R2 выше 0,8, а если меньше 0,5, то резонность такой модели можно смело ставить под большой вопрос.

Теперь перейдем к коэффициентам модели:   
**2079,85** - это **a0** - коэффициент который показывает какой будет Y в случае, если все используемые в модели факторы будут равны 0, подразумевается что это зависимость от других неописанных в модели факторов;   
**-0,0056** - **a1** - коэффициент, который показывает весомость влияния фактора x1на Y, то есть количество предприятий в пределах данной модели влияет на показатель экономически активного населения с весом всего -0,0056 (довольно маленькая степень влияния). Знак минус показывает что это влияние отрицательно, то есть чем больше предприятий, тем меньше экономически активного населения, как бы это ни было парадоксальным по смыслу;   
**-0,0026** - **a2** - коэффициент влияния объема инвестиций в капитал на величину экономически активного населения, согласно модели, это влияние также отрицательно;   
**0,0028** - **a3**- коэффициент влияния доходов населения на величину экономически активного населения, здесь влияние позитивное, то есть согласно модели увеличение доходов будет способствовать увеличению величины экономически активного населения.

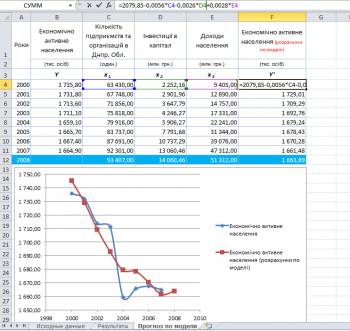
Соберем рассчитанные коэффициенты в модель:

***Y = 2079,85 - 0,0056x1 - 0,0026x2 + 0,0028x3***

Собственно, это и есть линейная регрессионная модель, которая для исходных данных, используемых в примере, выглядит именно так.

**Расчетные значения модели и прогноз**

Как мы уже обсуждали выше, модель строится не только чтобы показать величину зависимостей изучаемого параметра от влияющих факторов, но и чтобы зная эти влияющие факторы можно было делать прогноз. Сделать этот прогноз довольно просто, нужно просто подставить значения влияющих факторов в место соответствующих иксов в полученное уравнение модели. На рисунке ниже эти расчеты сделаны в экселе в отдельном столбце.

[](http://archie-goodwin.net/_ld/3/80185329.png)

Фактические значения (те что имели место в реальности) и расчетные значения по модели на этом же рисунке отображены в виде графиков, чтобы показать разность, а значит погрешность модели.

Повторюсь еще раз, для того чтобы сделать прогноз по модели нужно чтобы были известные влияющие факторы, а если речь идет о временном ряде и соответственно прогнозе на будущее, например, на следующий год или месяц, то далеко не всегда можно узнать какие будут влияющие факторы в этом самом будущем. В таких случаях, нужно еще делать прогноз и для влияющих факторов, чаще всего это делают с помощью авторегрессионной модели - модели, в которой влияющими факторами являются сам исследуемый объект и время, то есть моделируется зависимость показателя от того каким он был в прошлом.

Как строить авторегрессионную модель рассмотрим в следующей статье, а сейчас предположим, что, то какие будут величины влияющих факторов в будущем периоде (в примере 2008 год) нам известно, подставляя эти значения в расчеты мы получим наш прогноз на 2008 год.