**ФИНАНСОВАЯ МАТЕМАТИКА**

**ЛЕКЦИЯ 5. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ**

**Основная цель лекции:**

* 1. Познакомиться с понятиями корреляционного и регрессионного анализа;
  2. Изучить практическое применение формул и функций Excel для расчета корреляции и регрессии;
  3. Познакомиться с основами анализа финансовых временных рядов

**На этой лекции мы**

* Познакомимся с понятиями линейной корреляции
* Изучим применение корреляционного анализа в экономике и финансах
* Разберемся с теорией регрессионного анализа
* Изучим применение регрессий в экономике и финансах для прогнозирования
* Научимся рассчитывать корреляции и определять регрессионные коэффициенты в Excel
* Познакомимся с основами анализа временных рядов

**Термины**

**Эконометрика** – наука, изучающая количественные и качественные экономические взаимосвязи с помощью статистических и других математических методов и моделей.

**Статистическая** **зависимость** – означает, что распределение вероятностей одной случайной величины зависит от значения другой случайной величины.

**Корреляционный** **анализ** — метод обработки статистических данных, с помощью которого измеряется теснота связи между случайными величинами.

**Линейная** **корреляция** - мера линейной связи между двумя случайными величинами.

**Регрессия** - поиск уравнения зависимости одной случайной величины (называемой зависимой) от другой (называемой независимой), коэффициенты которого подбираются по некоторому критерию, позволяющему наилучшим образом объяснить изменение зависимой величины.

**Множественная** **регрессия** - поиск коэффициентов линейной зависимости одной, зависимой случайной величины от нескольких независимых случайных величин.

**Коэффициент детерминации** – показатель качества регрессии, который равен доле вариации зависимой случайной величины, которую удалось объяснить с помощью регрессии.

**Условная гетероскедастичность** - зависимость ошибки регрессии от величины независимой величины.

**Автокорреляция** - когда ошибка регрессии зависит от предыдущих значений ошибки.

**Мультиколлинеарность** - когда две или более независимых величин оказываются связаны (коррелированы) между собой.

**Анализ временных рядов** — совокупность математических методов анализа, предназначенных для прогнозирования временного ряда (значений случайной величины во времени).

**План лекции:**

Введение

Эконометрика и ее роль в экономике и финансах

Корреляционный анализ

Регрессионный анализ

Анализ финансовых временных рядов

**ВВЕДЕНИЕ**

Для анализа экономических и финансовых явлений используется большое количество математических методов и инструментов. Часть из них используются в неизменном виде, как и в других приложениях, а другая часть приобрела некоторые черты, свойственные экономическим наукам. Примером такой трансформации и стала эконометрика.

С одной стороны, это просто направление в рамках классической математической статистики. С другой стороны, арсенал методов, инструментов и подходов адаптирован под конкретные прикладные экономические задачи, и экономическая интерпретация результатов является неотъемлемой частью этой науки.

**ЭКОНОМЕТРИКА И ЕЕ РОЛЬ В ЭКОНОМИКЕ И ФИНАНСАХ**

**Эконометрика** - наука, изучающая количественные и качественные экономические взаимосвязи с помощью статистических и других математических методов и моделей.

Упрощенно, эконометрика применяет существующий инструментарий современной математики (в т.ч. математической статистики) к анализу экономических явлений. Зачем же нужно было создавать отдельную науку?

Существует ряд причин, по которым эконометрику выделяют в отдельную науку, несмотря на то, что практически все инструменты и методы взяты из других областей математики:

* **сложность** интерпретации результатов эконометрических расчетов без четкого понимания предметной (экономической) области. Математические методы не включают в себя никакой интерпретации результатов. А в эконометрических задачах самое главное – это интерпретация, объяснение и использование результатов расчетов;
* малое количество данных и **ограниченный** **размер** выборок (зачастую количество данных ограничено тем объемом, который был предоставлен провайдером, типа Bloomberg или Reuters, содержится в аналитическом отчете или находится в открытом доступе);
* **нерегулярный** **характер** данных и наличие пропусков/”выбросов” (Выбросами называются значения в выборке данных, которые по величине очень сильно отличаются от остальных. Выбросы появляются как погрешность или ошибка, поскольку, в отличие от физических экспериментов, в экономике точность данных и их полнота ничем не гарантированы);
* **косвенный** **характер** измерений экономических величин (даже вопрос измерения инфляции, как мы обсуждали на одной из предыдущих лекций – очень нетривиален, и зависит от субъективного выбора потребительской корзины);
* **невозможность** проведения дополнительных экспериментов или наблюдений.

С точки зрения используемого инструментария, эконометрику можно разбить на несколько областей, отличающихся функционалом и целями:

* корреляционный анализ -> предназначен для качественного поиска зависимостей между экономическими случайными величинами (например, между доходностью инвестиционного фонда и доходностью рыночного индекса);
* регрессионный анализ -> предназначен для количественного определения аналитической зависимости между двумя экономическими величинами (например, поиск уравнения зависимости между ценой золота и ценой акции золотодобывающей компании);
* анализ временных рядов -> используется для прогнозирования будущих значений экономической величины на основе ее исторических значений (например, прогнозирование будущей волатильности индекса на основе его исторической динамики);
* панельный анализ -> предназначен для одновременного изучения статистических данных и во времени, и между экономическими величинами (например, анализ данных по бедности или преступности). Этот вид анализа мы не будем рассматривать, поскольку он достаточно редко встречается в типовых задачах финансового аналитика и выходит за рамки нашего курса.

**КОРРЕЛЯЦИОННЫЙ АНАЛИЗ**

Корреляционный анализ является одним из инструментов для поиска и оценки статистической зависимости.

**Статистическая зависимость** означает, что распределение вероятностей одной случайной величины зависит от значения другой случайной величины. Если случайные величины статистически зависимы, то изменения значений одной приводят к систематическому изменению значений другой величины. Например, мы все понимаем, что цена акции нефтедобывающей компании и цена нефти должны быть статистически (и экономически!) взаимосвязаны.

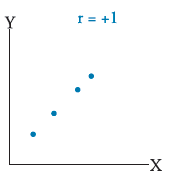
**Корреляционный анализ** — метод обработки статистических данных, с помощью которого измеряется теснота связи между случайными величинами.

При этом функциональная форма статистической зависимости может быть любая. Обычно ее классифицируют на **линейную** и **нелинейную** зависимости. Подавляющее большинство экономических величин связаны именно **линейной** зависимостью, поэтому ей посвящена существенная часть прикладной эконометрики. Например, цена грузового железнодорожного вагона (который производится из стали) и цена стали связаны линейно: чем дороже сталь, тем дороже вагон.

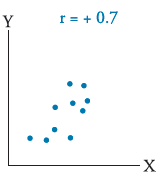
**Линейная корреляция** - мера линейной связи между двумя случайными величинами X и Y, измеряется коэффициентом линейной корреляции Пирсона, обладающего следующим свойством:

***-1 <= r <= +1***

Если **r = +1** - это совершенная положительная корреляция, т.е. все точки (Х,Y) располагаются на одной прямой:



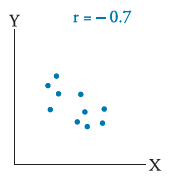
А если же коэффициент корреляции положителен, но меньше 1, т.е. **0 < r < +1** - несовершенная положительная корреляция, когда точки уже не лежат на одной прямой, но можно визуально определить, что чем больше одна величина, тем обычно больше другая:



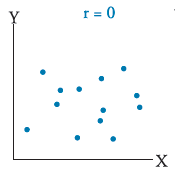
Если **r = -1** - это совершенная отрицательная корреляция, точки (Х,Y) также лежат на одной прямой, только эта прямая имеет отрицательный наклон:



Если же **-1 < r < 0** - это несовершенная отрицательная корреляция, точки уже не будут лежать на одной прямой, но общая зависимость визуально прослеживается:



Наконец, если **r = 0**, то это значит, что линейная зависимость отсутствует, и визуально никакой ясной картины нет, точки расположены хаотично:



Отсюда следует **интерпретация** коэффициента линейной корреляции: чем больше значение коэффициента корреляции по модулю (чем он сильнее отличается от 0), тем сильнее линейная взаимосвязь между исследуемыми величинами.

Существуют следующие **ограничения корреляционного анализа**, которые надо понимать в процессе использования этого инструмента:

* требуется достаточное количество данных (если у вас всего 5 точек, то этого явно недостаточно для качественного анализа);
* он измеряет только линейную взаимосвязь, а статистическая взаимосвязь может быть нелинейной, и коэффициент корреляции ее не сможет найти (к счастью, в подавляющем большинстве эконометрических задач взаимосвязь между переменными является линейной или квази-линейной, т.е. «почти» или «примерно» линейной);
* он не позволяет выделять причинно-следственные связи (и это огромная проблема с точки зрения интерпретации результатов: мы можем определить, что у величин X и Y есть статистическая линейная взаимосвязь. Но это нам не говорит, зависит ли X от Y, или, может быть, Y от X, или, возможно, они обе зависят от некой третьей величины Z). Например, цена грузового вагона и цена пассажирского вагона обладают существенной корреляцией. Однако на самом деле они обе зависят от цены на сталь;
* большое влияние оказывают “выбросы” в данных (если по какой-то причине в данных возникла ошибка, то она очень сильно исказит результаты анализа);
* существует т.н. ложная корреляция (то есть, ситуация, когда математически корреляция есть, а экономически взаимосвязи нет).

**РЕГРЕССИОННЫЙ АНАЛИЗ**

По результатам корреляционного анализа мы можем получить информацию о наличии взаимосвязи. Однако этого мало. Нам же нужно понимать, как выглядит эта зависимость, нужна формула, чтобы мы могли ее использовать для расчетов. Именно этим инструментом и является регрессионный анализ.

**Регрессия** - поиск уравнения зависимости одной случайной величины (называемой **зависимой**) от другой (называемой **независимой**) величины (или нескольких), коэффициенты которого подбираются по некоторому критерию, позволяющему наилучшим образом объяснить изменение зависимой величины.

В основе регрессии лежат определенные статистические допущения, которые выходят за рамки нашего курса, но которые желательно знать при выполнении серьезных эконометрических расчетов. Главный аспект, который нужно всегда помнить - при поиске зависимости величины Y от величины X нужно быть уверенным, что на величину Y не оказывает существенного влияния некая другая величина Z.

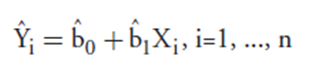
Для базовой работы на уровне встроенных функций Excel и получения количественных результатов материалов данной лекции хватит. Однако для профессиональной работы с регрессионным анализом рекомендуется поглубже изучить литературу по данной теме.

Если предполагаемое уравнение зависимости величин имеет линейный вид, то такая регрессия будет называться **линейной**. Формула линейной регрессии имеет вид:

https://lh4.googleusercontent.com/Qbsobd_jJvzG3iaWDIvU08WYT0mzNlHartbkk_IQYyuAUzoR3_T7gR859hi1br91ddrJmbkHFrvKKFHLQ6JEVfGVORLIwwh5ADdTHea6E1ihc8vUXYvsqeSUahVtogV7M2vZ9_czSDDdTZWUtVoDtA

где Х - независимая величина (переменная), например, цена золота; Y - зависимая величина, например, цена акции золотодобывающей компании, а ε – случайная ошибка.

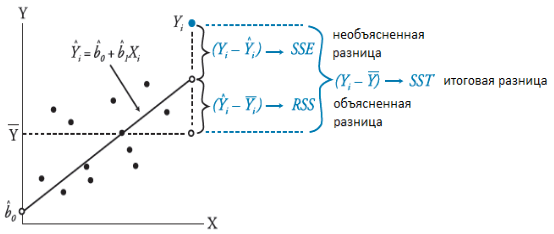
**Целью** расчета регрессии является определение коэффициентов b0 и b1. Однако, точного значения этих коэффициентов мы никогда не узнаем, поскольку всегда имеем дело не с генеральными совокупностями, а с ограниченными выборками. Поэтому вместо точных значений коэффициентов мы рассчитываем их оценки, которые используем для записи уравнения:



Теперь, зная это уравнение, можно прогнозировать значения величины Y в будущем для разных прогнозных значений величины X. А это и есть главная задача поиска регрессии в эконометрике!

Поиск коэффициентов линейной регрессии может осуществляться по-разному, однако в подавляющем большинстве случаев используется **метод наименьших квадратов**: подбираются такие коэффициенты, чтобы сумма квадратов отклонений реального значения величины Y от ее прогнозного значения согласно уравнению регрессии для всех значений величины X была минимальной.

Запоминать это не нужно, поскольку при текущем уровне развития компьютерных технологий никто не считает регрессию самостоятельно, все делает Excel или другие программные пакеты. Однако для лучшего понимания, что такое регрессия и как она работает, можно предложить следующую визуализацию:



Из графика видно, что отклонение точки от среднего значения частично было учтено в нашей линии регрессии («объясненная разница»), а частично нет («необъясненная разница»). Так вот целью метода наименьших квадратов является подбор такого уравнения линии регрессии, чтобы сумма квадратов этих необъясненных разниц была минимальной…

Мы изучили линейную регрессию в виде зависимости от одной независимой переменной. Однако бывают случаи, когда на искомую величину влияют сразу несколько переменных. Тогда нам нужна множественная регрессия.

**Множественная регрессия** - поиск коэффициентов линейной зависимости одной случайной величины Y от нескольких случайных величин Xi, например, зависимость цены продукции предприятия от цен на различные ресурсы, которые используются в производстве.

Общая формула имеет вид:

https://lh5.googleusercontent.com/i7GcMKAkC_8b_8EZebtuqfTWXzWVf7Wau0S9YSi9SP-Jrgh1ecCvQissBNBiDmepfwJ9o6g9k2LkwST1yBr2cl5lqSHe7AmxZzofzv9TmMBmdr1f-8twS_VJ9gpGV5E6sH9cWoWPp5FlDZMUHRXQqA

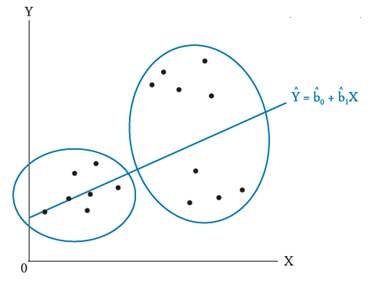
**Целью** является поиск коэффициентов b1, b2, ... , bk, чтобы иметь возможность прогнозировать будущие значения величины Y (аналогично классической линейной регрессии, разобранной выше).

После расчета регрессии, прежде чем пользоваться ее результатами,  необходимо оценить ее **качество**. Главным показателем качества регрессии является **коэффициент детерминации (R-квадрат),** который показывает, какую долю вариации величины Y при разных значениях X удалось объяснить с помощью регрессии. Для большинства прикладных задач желательно, чтобы этот коэффициент был не ниже **70-80%.**

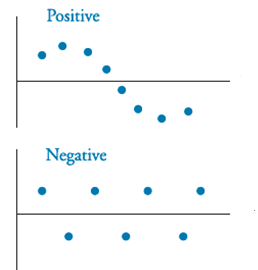
Этот показатель во всех статистических программных пакетах рассчитывается автоматически. А для обычной линейной регрессии с одним параметром можно просто возвести в квадрат коэффициент корреляции, чтобы его получить (отсюда и название «R-квадрат»).

Для оценки применимости регрессии к прогнозированию реальных данных, следует обратить внимание на следующие потенциальные проблемы (при их наличии формулы регрессии перестают корректно работать):

* **условная гетероскедастичность** - зависимость ошибки регрессии от величины независимой величины (Х):



* **автокорреляция** - когда ошибка регрессии зависит от предыдущих значений ошибки:



* **мультиколлинеарность** - когда две или более независимых величин (Х) оказываются связаны (коррелированы) между собой.

Рассмотрим задачу поиска корреляции и регрессии с помощью Excel.

**Задача**. Есть исторические данные о величине условно-постоянных затрат промышленного предприятия, объемах производства и инфляции. Банк-кредитор в рамках анализа финансовой модели компании попросил сделать расчет, насколько эти затраты зависят от уровня производства. И если зависят, то какое уравнение описывает эту зависимость (чтобы ее можно было интегрировать в финансовую модель).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год |  | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| Производство | ед. | 1 250 | 2 800 | 7 700 | 11 050 | 13 250 | 16 050 | 17 450 |
| Постоянные издержки | млн.руб. | 2 780 | 3 960 | 5 210 | 7 040 | 8 600 | 9 100 | 9 250 |
| Инфляция | % | 8% | 10% | 15% | 10% | 9% | 8% | 6% |

Решение. Найдем коэффициент корреляции с помощью функции КОРРЕЛ(): 98,8%. Это очень высокая величина, которая говорит о том, что линейная зависимость есть и ей можно пользоваться для расчетов.

Тогда можно воспользоваться встроенной функцией поиска коэффициентов регрессии – ЛИНЕЙН(). Обратите внимание, что эта функция должна выдать не один, а два коэффициента. С другой стороны, мы все знаем, что в одной ячейке Excel может быть только одно значение, поэтому нам надо воспользоваться т.н. «формулой массива», т.е. выделить две ячейки, набрать функцию ЛИНЕЙН(), передать туда нужные параметры, а зачем нажать Ctrl+Shift+Enter. Если все сделано правильно, то формула в строке формул окажется в фигурных скобках {}, а функция заполнит обе ячейки требуемыми коэффициентами.

Мы получили выражение вида: Y = 2485 + 0.410\*X, то есть

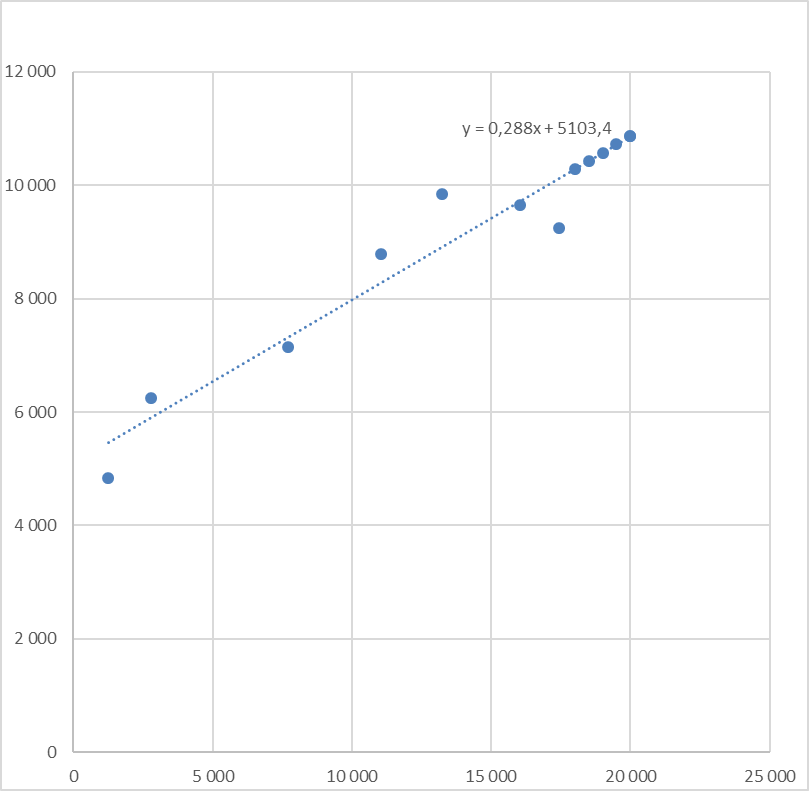
***Издержки [млн руб.] = 2485 [млн руб.] + 0,410 \* Объем\_производства [ед]***

Однако можно заметить, что в наших расчетах есть методологическая ошибка. Дело в том, что мы взяли достаточно большой временной диапазон , где цены в силу временной стоимости денег, мало сравнимы между собой. То есть 100 млн руб. издержек в 2012 году и 100 млн руб. издержек в 2018 – это совсем разные суммы. Давайте скорректируем на инфляцию, приведя все к ценам 2018 года:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Год |  | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| Производство | ед. | 1 250 | 2 800 | 7 700 | 11 050 | 13 250 | 16 050 | 17 450 |
| Постоянные издержки | млн.руб. | 2 780 | 3 960 | 5 210 | 7 040 | 8 600 | 9 100 | 9 250 |
| Инфляция | % | 8% | 10% | 15% | 10% | 9% | 8% | 6% |
| Индекс для пересчета цен | % | 108% | 110% | 115% | 110% | 109% | 108% | 106% |
| Накопленный индекс | % | 174% | 158% | 137% | 125% | 114% | 106% | 100% |
| Скорректированные издержки | млн.руб. | 4 827 | 6 251 | 7 151 | 8 785 | 9 845 | 9 646 | 9 250 |

Теперь мы можем еще раз рассчитать корреляцию (94,3%) и регрессию: Y = 5103 + 0.288\*X. Заметьте, как после очистки от инфляции поменялись коэффициенты!

Теперь эту формулу мы можем применить для прогнозирования условно-постоянных издержек на будущее для разных значений прогнозных объемов производства (по горизонтально оси – объем производства в ед., по вертикальной – издержки в млн руб.):



**АНАЛИЗ ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДОВ**

**Анализ временных** **рядов** — совокупность математических методов анализа, предназначенных для прогнозирования временного ряда (значений случайной величины во времени).

При этом в отличие от вышеизложенного не проводится оценки зависимости от других случайных величин.

В рамках анализа временных рядов выделяют две группы методов, в зависимости от того, что используется в качестве зависимой переменной:

* трендовые модели (от времени);
* авторегрессионные модели (от прошлых значений величины).

**Трендовые** **модели** - изучают зависимость некоторой случайной величины от времени (например, зависимость процентной ставки от времени):

https://lh5.googleusercontent.com/5lQ2UWLZTYOCbHQTmG3xMp7o0Upn40AAJ-Yrl1fgXADxEEBYAtUHy8ifU8EWuxi8Zj4Blo__ZJtLsT1bG4tOnu2o6qpO1y1ki_jZ6zMpTvzCEnYx4q9iVn7syUKuVg0v_5ttKKCAr_bfOqaERvwndw,

где t – порядковый номер периода времени, а b0 и b1 – искомые коэффициенты трендовой модели.

В этом случае может использоваться классический регрессионный аппарат, изученный ранее, где в качестве независимой переменной выступает время (номер точки данных).

**Авторегрессионные** **модели** - изучают зависимость случайной величины от ее предыдущих значений

https://lh3.googleusercontent.com/_mCohPN_VimELKpbDcm8G9oKuIPu2IZwNmDek3f8-0iUq86lEp_hs7yLaBm-bw2lPIeHflxDjRrvG1IJRgVFNHKTe1M5JbX2jpgAu5KGzm8h_6L1ks_szKFSL53Wz7tk0d_kkQqk14ypS-DW9t9oSA,

где xt – текущее значение случайной величины, а xt-1, xt-2, … - это ее прошлые значения.

В этом случае формулы для стандартной регрессии также применимы, однако существуют более строгие технические требования к характеристикам данных.

Самое частое применение авторегрессионных моделей в финансах - прогнозирование цен и доходностей акций, процентных ставок и волатильности финансовых активов.

Применимость эконометрического аппарата анализа временных рядов сталкивается с несколькими **практическими проблемами**:

* **нестационарность** - изменение свойств финансового временного ряда во времени, например, в связи с изменением государственного регулирования экономики, политической обстановки, развития или стагнации рынка/отрасли и т.д.;
* **сезонность** - изменения параметров финансового временного ряда в связи с объективными экономическими циклами;
* **условная гетероскедастичность** - изменение волатильности случайной величины во времени (была рассмотрена выше).

**На этой лекции мы**

* Познакомились с понятиями линейной корреляции
* Изучили применение корреляционного анализа в экономике и финансах
* Разобрались с теорией регрессионного анализа
* Изучили применение регрессий в экономике и финансах для прогнозирования
* Научились рассчитывать корреляции и определять регрессионные коэффициенты в Excel
* Познакомились с основами анализа временных рядов

**Используемые и рекомендуемые источники:**

И.И.Елисеева. Эконометрика. Учебник для бакалавриата и магистратуры.

В.Носко. Эконометрика.