Глава 1. Основы эконометрики

Эконометрика **–** это научная дисциплина, предметом которой является изучение количественной стороны экономических явлений и процессов средствами математического и статистического анализа.

Эконометрика является частью экономической теории, наряду с макро- и микроэкономикой.

Эконометрика дает инструментарий для экономических измерений, а также методологию оценки параметров моделей микро- и макроэкономики. Кроме того, эконометрика активно используется для прогнозирования экономических процессов как в масштабах экономики в целом, так и на уровне отдельных предприятий.

Тема 1.1 Предмет эконометрики

Эконометрика — быстроразвивающаяся отрасль науки, цель которой состоит в том, чтобы придать количественные меры экономическим отношениям.

Термин «эконометрика» был впервые введен бухгалтером П. Цьемпой (Австро-Венгрия, 1910 г.) («эконометрия» — у Цьемпы). Цьемпа считал, что если к данным бухгалтерского учета применить методы алгебры и геометрии, то будет получено новое, более глубокое представление о результатах хозяйственной деятельности. Это употребление термина, как и сама концепция, не прижилось, но название «эконометрика» оказалось весьма удачным для определения нового направления в экономической науке, которое выделилось в 1930 г.

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. «метрон»). Таким образом, сам термин

подчеркивает специфику, содержание эконометрики как науки: количественное выражение тех связей и соотношений, которые раскрыты и обоснованы экономической теорией. Й. Шумпетер (1883—1950), один из первых сторонников выделения этой новой дисциплины, полагал, что в соответствии со своим назначением эта дисциплина должна называться «экономометрика». Советский ученый А. Л. Вайнштейн (1892—1970) считал, что название настоящей науки основывается на греческом слове *метрия* (геометрия, планиметрия и т.д.), соответственно по аналогии — эконометрия. Однако в мировой науке общеупотребимым стал термин «эконометрика». В любом случае, какой бы мы термин ни выбрали, эконометрика является наукой об измерении и анализе экономических явлений.

Зарождение эконометрики является следствием междисциплинарного подхода к изучению экономики. Эта наука возникла в результате взаимодействия и объединения в особый «сплав» трех компонент: *экономической теории, статистических и математических методов.* Впоследствии к ним присоединилось развитие вычислительной техники как условие развития эконометрики.

В журнале «Эконометрика», основанном в 1933 г. Р. Фришем (1895—1973), он дал следующее определение эконометрики: «Эконометрика — это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек — статистика, экономическая теория и математика — необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это — единство всех трех доставляющих. И это единство образует эконометрику».

Таким образом, ***эконометрика — это наука, которая дает количественное***

выражение взаимосвязей экономических явлений и процессов. Нельзя утверждать, что достигнуто однозначное определение эконометрики. Так, Э. Маленво придерживался широкого понимания, интерпретируя эконометрику как «любое приложение математики или статистических методов к изучению экономических явлений».

О. Ланге (1904—1965) писал, что эконометрика занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы. Статистический подход к эконометрическим измерениям стал доминирующим.

Некоторые сведения об истории возникновения эконометрики.

Каждая наука проходит сложный путь зарождения и выделения в самостоятельную область знания. Эконометрика — не исключение. Первоначальные попытки количественных исследований в экономике относятся к XVII в. «Политические арифметики» — В. Петти (1623-1667), Г. Кинг (1648-1712), Ч. Давенант (1656—1714) — вот первая когорта ученых, систематически использовавших цифры и факты в своих исследованиях, прежде всего в расчете национального дохода. Круг их интересов был связан в основном с практическими вопросами: налогообложением, денежным обращением, международной торговлей и финансами. Политическую арифметику можно назвать описательным политико-эконометрическим анализом. Это направление пробудило поиск законов в экономике. Одним из первых был сформулирован так называемый «закон Кинга», в котором на основе соотношения между урожаем зерновых и ценами на зерно была выявлена закономерность спроса. Исследователям хотелось достичь в экономике того, что И. Ньютон достиг в физике. Неопределенная природа экономических закономерностей еще не была осознана. В этот же период все больше учетных данных становятся доступными, создавая основу для измерений.

Существенным толчком явилось развитие статистической теории в трудах Ф. Гальтона (1822—1911), К. Пирсона (1857—1936), Ф. Эджворта (1845—1926). Появились первые применения парной корреляции: при изучении связей между уровнем бедности и формами помощи бедным (Дж. Э. Юл,. 1895, 1896); между уровнем брачности в Великобритании и благосостоянием (Г. Хукер, 1901), в котором использовалось несколько индикаторов благосостояния, к тому же исследовались временное ряды экономических переменных. Это были шаги по созданию современной эконометрики.

Параллельно происходил процесс создания маржиналистской (неоклассической) теории, зарождение которой можно датировать 60-ми годами XIX в. (появление работ С. Джевонса, Л. Вальраса, К. Менгера). С 30-х гг. XIX в. страны с наиболее высоким уровнем развития капитализма стали испытывать спорадические потрясения — упадок деловой активности, возникновение массовой безработицы. Эти явления не находили теоретического объяснения. Быстрая индустриализация выявила огромный диапазон социальных проблем, которые также не согласовывались с теорией. Неоклассическая теория стала восприниматься как слишком удаленная от действительности. Для ее практического значения требовались количественные выражения базовых понятий, таких как «эластичность спроса» или «предельная полезность». Теория спроса могла стать убедительной в том случае, если она смогла бы объяснить и оценить фактические кривые спроса и предложения, продемонстрировать формирование равновесных цен в конкретных условиях.

К этому же времени относится привлечение ученых-экономистов (А. Маршалла, С. Джевонса, К. Менгера) к парламентской деятельности, что подтолкнуло их к анализу макроэкономических проблем на основе временных рядов таких показателей, как, например, валютные курсы и т.п. Это также явилось важным шагом в подготовке развития эконометрики. Многие исследователи признают первой работой, которая могла бы быть названа

эконометрической, книгу американского ученого Г. Мура (1869—1958) «Законы заработной платы: эссе по статистической экономике» (1911). Г. Муром были проведены анализ рынка труда, статистическая проверка теории производительности Дж. Кларка, а также изложены основы стратегии объединения пролетариата и т. д. В это время для США решение этих вопросов было безотлагательным: рабочий класс стремительно рос, возникали такие объединения, как «Индустриальные рабочие мира» и другие радикально настроенные организации. Г. Мур подошел к анализу поставленных проблем с позиций «высшей», как он называл, статистики, используя все достижения теории корреляции, регрессии, анализа динамических рядов. Он стремился показать, что сложные математические построения, наполненные фактическими данными, могли составить основу для разработки социальной стратегии.

К этому же периоду относится первое применение итальянским ученым Р. Бенини (1862—1956) метода множественной регрессии для оценки функции спроса. Значительным вкладом в становление эконометрики явились исследования по цикличности экономики. К. Жюгляр (1819—1905), французский физик, ставший экономистом, первым занялся исследованием экономических временных рядов с целью выделения бизнес-циклов. Им была обнаружена цикличность инвестиций (продолжительность цикла — 7—11 лет). Вслед за ним С. Китчин, С. Кузнец, Н. Кондратьев, автономно занимаясь этой проблемой, выявили цикличность обновления оборотных средств (3—5 лет), циклы в строительстве (15—20 лет), долгосрочные волны, или «большие циклы» Кондратьева, продолжительностью 45—60 лет.

Значительной вехой в формировании эконометрики явилось построение экономических барометров, прежде всего так называемого гарвардского барометра. Большинство экономических барометров, включая названный, основано на следующей идее: в динамике различных элементов экономики

существуют такие показатели, которые в своих изменениях идут впереди других, а потому могут служить предвестниками последних. Гарвардский барометр был создан под руководством У. Персонса (1878—1937) и У. Митчелла (1874—1948). В течение 1903—1914гг. он состоял из пяти групп показателей, которые в дальнейшем были сведены в три отдельные кривые: кривая А характеризовала фондовый рынок; кривая В — товарный рынок; кривая С — денежный рынок. Каждая из этих кривых представляла среднюю арифметическую из рядов входящих в нее нескольких показателей. Эти ряды предварительно статистически обрабатывались путем исключения тенденции, сезонной волны и приведения колебаний отдельных кривых к сравнимому масштабу колеблемости. В основу прогноза гарвардского барометра было положено свойство каждой отдельной кривой повторять движение остальных в определенной последовательности и с определенным отставанием. Так, с 1903 г. и до первой мировой войны поворотные пункты кривой А предшествовали поворотным пунктам кривой В на 6—10 месяцев (в среднем — на 8 месяцев); поворотные пункты кривой В обгоняли аналогичные пункты кривой С на 2—8 месяцев (в среднем на 4 месяца); наконец, колебания кривой С предшествовали колебаниям кривой А следующего цикла на 6—12 месяцев.

Гарвардский барометр представлял собой описание подмеченных эмпирических закономерностей и экстраполяции последних на ближайшие месяцы. Однако в построении гарвардского барометра можно обнаружить и некоторые теоретические предпосылки. Естественно, например, что изменение средних биржевых курсов и показателей фондового рынка (индекс спекуляции А) означало изменение спроса на товары, что влекло за собой, в свою очередь, изменение в том же направлении индекса оптовых цен, объема производства и товарооборота (индекс В). Возрастание, например, объема производства вызывало напряжение на денежном рынке, рост учетной ставки и падение курса ценных бумаг с фиксированным доходом (кривая С). Поэтому максимум кривой А обычно должен был совпадать с минимумом кривой С.

Успех гарвардского барометра породил буквально эпидемию таких построений в других странах (в частности, аналогичный барометр был построен в Великобритании). Несколько лет после первой мировой войны он еще удовлетворительно выполнял свое предназначение. Но затем гарвардский барометр (приблизительно с 1925 г.) потерял чувствительность и сошел со сцены, пережив свою славу. Авторы гарвардского барометра объясняли его крах появлением мощного регулирующего фактора в экономике США. В этих условиях основным методом макроэкономического анализа становится метод «Затраты-выпуск» В. В. Леонтьева (1906-1999).

Что касается экономических барометров, то советский математик-статистик Е. Слуцкий (1880—1948) в работе «Сложение случайных причин как источник циклических процессов» (1927), взяв в качестве случайных рядов последние цифры номеров облигаций из тиражных таблиц выигрышного займа, блестяще доказал, что «сложение случайных причин порождает волнообразные ряды, имеющие тенденцию на протяжении большего или меньшего числа волн имитировать гармонические ряды, сложенные из небольшого числа синусоид». Таким образом, никакой закономерности в любом экономическом барометре могло и не существовать.

В этот же период делались эконометрические построения, использующие методы гармонического анализа и периодограмм-анализа (Г. Мур в США, Бэвэридж в Энстром в Швеции). Это методы перенесены в экономику из области астрономии, метеорологии, физики.

В основе гармонического анализа и периодограмм-анализа лежит теорема Фурье, согласно которой всякая периодическая функция, произвольно данная в некотором промежутке, может быть разложена на ряд простых гармонических колебаний и, в конечном счете, представлена тригонометрическим рядом вида:

$$y = f(t) = A_0 + A_1 \sin(kt + e_1) + A_2 \sin(2kt + e_2) + \cdots.$$

Каждое слагаемое представляет здесь синусоиду — формулу простого гармонического колебания (гармонику), где A_i — полуамплитуда; e_i — фаза колебания, т. е. характеризует точку, в которой ордината соответствующей синусоиды имеет нулевое значение; k — связано с периодом колебания равенством

$$k = \frac{2\pi}{T}$$

Динамика каждого элемента экономики после исключения из нее тенденции представляется в виде волнообразной кривой. Если бы оказалось возможным эту кривую разложить, хотя бы приближенно, на сумму гармоник, то это дало бы базу для прогноза движения интересующего нас элемента. Следовательно, задача сводится к нахождению коэффициентов искомого ряда – полуамплитуд A_i – по наблюденным значениям, если известны периоды отдельных гармоник. Для отыскания периода колебания T или связанного с ним k применяется метод периодограмм-анализа. Он состоит в том, что в качестве первого приближения берутся два первых члена вышеприведенного ряда, т. е. полагают, что

$$y = A_0 + A_1 \sin(kt + e_1)$$

и затем испытывают различные произвольные значения T (целые и дробные). Для каждого из испытываемых периодов вычисляются A_1 и e_1. Затем строится периодо-график или периодограмма, где на оси абсцисс отмечаются периоды, а на оси ординат откладывается A_1 или интенсивность колебания, соответствующая этим периодам. Большей интенсивности колебания отвечает большая вероятность того, что соответствующий ей период колебания не случаен. Затем, выбрав периоды, соответствующие наибольшим интенсивностям, можем представить рассматриваемую волнообразную кривую в виде суммы простых гармоник, имеющих эти периоды, соответствующие A_i.

Эта сумма может сколь угодно близко подойти к исследуемой кривой. К этому нужно добавить, что при применении гармонического метода и периодограмм-анализа не требуется предварительного исключения тенденции.

К 30-м гг. сложились все предпосылки для выделения эконометрики в отдельную науку. Стало ясно, что специалисты, занимающиеся развитием эконометрической науки, должны использовать в той или иной степени математику и статистику. Возникла необходимость появления особого термина, объединяющего все исследования в этом направлении, подобно биометрике – науке, изучающей биологию статистическими методами.

Выделение эконометрики.

В 1912 г. И. Фишер попытался создать группу ученых для стимулирования развития экономической теории путем ее связи со статистикой и математикой. Но тогда эту группу создать не удалось. Тогда Р. Фриш и математик-экономист Ч. Рус обратились с идеей собрать специальный форум экономистов, готовых к использованию математики и статистики.

29 декабря 1930 г. по инициативе И. Фишера (1867—1947), Р. Фриша, Я. Тинбергена (1903—1995), Й. Шумпетера, О. Андерсона (1887—1960) и других ученых на заседании Американской ассоциации развития науки (США, Кливленд, штат Огайо) было создано эконометрическое общество, на котором норвежский ученый Р. Фриш дал новой науке название — «эконометрика».

С самого начала эконометрическое общество было интернациональным. Уже в 1950 г. общество насчитывало почти 1000 членов. С 1933 г. под редакцией Р. Фриша стал издаваться журнал «Эконометрика» («Econometrica»), который и сейчас играет важную роль в развитии эконометрической науки. В 30 – 40-е гг. развитию эконометрики способствовала деятельность Департамента прикладной экономики под руководством Р. Стоуна (Великобритания). В 1941 г. появился первый учебник по эконометрике,

который был создан Я. Тинбергеном (1913—1994).

В эти годы вплоть до 70-х гг. XX в. эконометрика понималась как эмпирическая оценка моделей, разработанных экономической теорией. Р. Фриш определял соотношение между теорией и данными наблюдений следующим образом: теория, абстрактно формулирующая количественные соотношения, должна быть проверена множеством наблюдений. Свежие статистические данные и другие факты должны предотвратить теорию от опасного догматизма. Под влиянием лидеров, таких как Р. Фриш, Т. Хаавелмо, Я. Тинберге, Л. Клейн, экономические модели, построенные в этом периоде, всегда были кейнсианскими.

Все изменилось в 70-е гг. В макроэкономике возникли противоречия между кейнсианцами, монетаристами и марксистами. Формальные методы стали использоваться для доказательства причинности при выборе теоретических концепций. Экономическая теория потеряла свое решающее значение.

Другим важным событием стало появление компьютеров с высоким быстродействием и мощной оперативной памятью. Существенное развитие получил статистический анализ временных рядов.

Г. Бокс и Г. Дженкинс создали ARIMA-модель в 1970 г. (модели авторегрессии и проинтегрированного скользящего среднего). Это – важный класс параметрических моделей позволяет описывать нестационарные ряды. А К. Симс и другие ученые — VAR-модели (модель динамики нескольких временных рядов, в которой текущие значения этих рядов зависят от прошлых значений этих же временных рядов), ставшие популярными в начале 80-х гг. Вершиной этой стадии развития явился метод коинтеграции, развитый С. Йохансеном и др. (1990 г.).

В настоящее время эконометрика располагает огромным разнообразием

типов моделей — от больших макроэкономических моделей, включающих несколько сот, а иногда и тысяч уравнений, до малых коинтеграционных моделей, предназначенных для решения специфических проблем.

Экономическая модель

Основным элементом экономического исследования является анализ и построение взаимосвязей экономических переменных. Математическое выражение таких взаимосвязей называется экономической моделью.

Пример.

$C = \beta 0 + \beta 1 * I,$

I – располагаемый доход семьи; C – потребление.

Построение экономических моделей осложнено следующими факторами:

1. часто эти взаимосвязи не являются строгими, функциональными зависимостями;

2. очень трудно выявлять все факторы, влияющие на данный зависимый экономический показатель;

3. воздействие многих факторов является случайным;

4. экономисты обладают ограниченным набором данных статистических наблюдений, которые к тому же содержат различного рода ошибки.

Если удаётся преодолеть эти трудности, тогда можно построить экономическую модель, выражающую функциональную зависимость некоторой зависимой величины от формирующих её значение факторов.

Особенность функциональной зависимости состоит в том, что по значению независимой величины (переменной) можно однозначно, абсолютно точно вычислить, предсказать значение зависимой величины (например, функциональная зависимость 6=2*3, то есть 6 получится тогда и только тогда, когда 2 умножим именно на 3, а не на другое число. В экономике таких зависимостей практически нет. Например, доллар может стоить 70 руб. при различных значениях курса евро, йены и других валют. Или прибыль 1 млн. руб. фирма может получить не только тогда, когда затраты составят 3 млн., но и при других значениях в зависимости от колебаний цены.)

Эконометрическая модель

Рассмотрим набор реальных статистических данных (Ck, Ik) и изобразим эти данные точками в координатах (C,I).

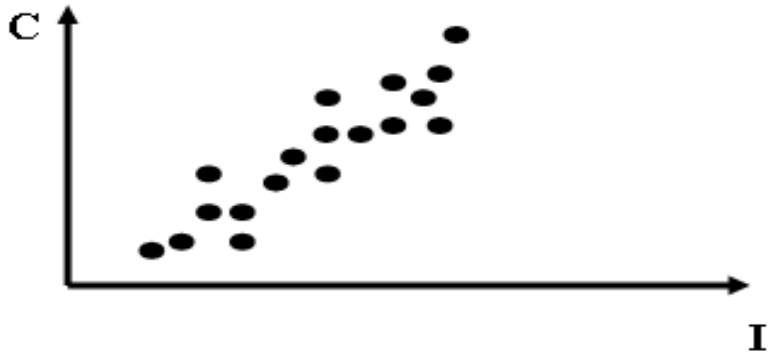


Рис. 1 Набор реальных статистических данных (Ck, Ik)

Как видно из рисунка, зависимость между величинами Ck, Ik не функциональная, а стохастическая, то есть случайная. Но эта случайность не такова, что абсолютно невозможно предсказать или объяснить по величине Ik величину Ck, поскольку видна достаточно устойчивая тенденция роста (в среднем). Другими словами, взаимосвязь между величинами Ck, Ik состоит в следующем: точное значение Ck не вычисляется по значению Ik, однако с

ростом Ik значение Ck в среднем увеличивается. Такой характер зависимости выражается следующим образом:

$Ck = \beta0 + \beta1*Ik + \varepsilon k$, $k = 1, 2,\ldots,N$, где

$\beta0$, $\beta1$ – это числа, которые определяются в ходе обработки статистики о данных Ik и соответствующих им значениях Ck,

εk – определенная погрешность, допускаемая моделью и ее уровнем вероятности.

В общем случае, характер зависимости нелинейный:

$yk = f(xk, \beta0, \beta1,\ldots \beta n) + \varepsilon k$; $k = 1, 2,\ldots,N$.

Приведенные формулы являются *эконометрическими моделями*. Таким образом, ***эконометрическая модель*** – это выражение статистической зависимости между экономическими величинами. Эконометрическая модель строится на основе экономической теории и статистических данных.

Классификация и примеры эконометрических моделей представлены на рис. 2-3.

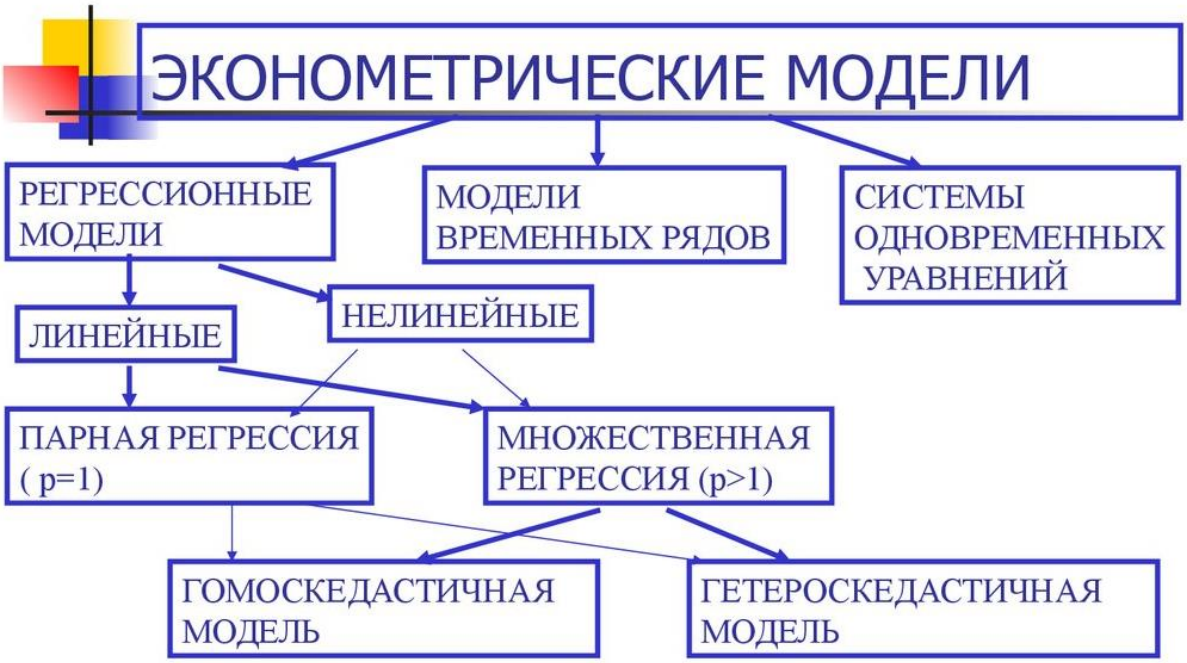


Рис. 2 Классификация эконометрических моделей

Рис. 3 Примеры эконометрических моделей

Элементы эконометрической модели и их свойства

Запишем модель: $y_k = f(x_k, \beta_0, \beta_1, \ldots \beta_n) + \varepsilon_k$; $k = 1, 2, \ldots, N$.

1) Вид функции f называется спецификацией модели. Модель Ck, представленная выше, является частным видом эконометрической модели, в которой спецификация линейная. Спецификация f описывает общий ход экономического процесса, экономическую тенденцию развития, изменения зависимого показателя при изменении независимого.

2) Величина xk – называется независимой или объясняющей переменной.

3) Величина yk называется зависимой или объясняемой переменной. Значение величины yk состоит из двух частей:

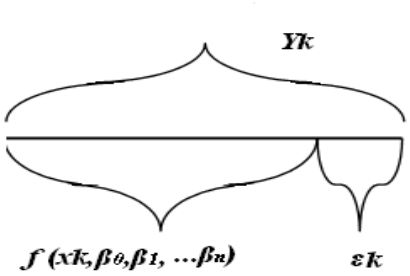


Рис. 4 Состав зависимой (объясняемой) величины yk)

- величина f (xk,β0, β1,… βn) – это часть зависимого показателя, обусловленная или объяснённая экономическими причинами или просто *объясняемая* часть;

- εk – необъяснённая часть, поскольку невозможно описать все случайные факторы.

Например, модель производственной функции имеет вид $y = a_0 \cdot K^{a1} L^{a2}$*, где y – объем производства, K – индекс производства (включает в агрегированном, т.е. безразмерном, виде все производственные затраты: на ресурсы, амортизацию производственных фондов, эксплуатацию машин ит.д.; L – индекс труда, включающий оплату труда, производительность труда, численность рабочих, премии и т.д., a1 и a2 – численное значение коэффициентов, индивидуально определяемое на основе статистики для*

каждого предприятия, a0 – влияние неучтенных факторов, то есть не входящих в K и L, например, изменения законодательства, погода (влияет на строительство) и т.д.

4) εk – величина, выражающая вклад случайных мелких, незначительных факторов, которые отклоняют реальные статистические данные от значений зависимого показателя, обусловленного экономической тенденцией, однако не изменяют эту экономическую тенденцию.

Основные свойства величин εk:

а) эти величины случайные, в противном случае зависимость будет функциональная;

б) εk принимают положительные и отрицательные значения, так как случайные факторы увеличивают или уменьшают величины, обусловленные экономической тенденцией;

в) Абсолютные величины |εk| не должны быть очень большими по сравнению со значениями, вычисляемыми по спецификации. *То есть, необъяснённая часть показателя y должна быть мала по сравнению с объяснённой. Если же отклонения значительны, тогда спецификация fнеточно описывает экономическую тенденцию, её необходимо уточнять и в число объясняющих факторов вводить факторы, вклад которых приводит к большим значениям εk.(например, в предыдущем примере с производственной функцией для строительных предприятий сезонный, то есть погодный фактор, нужно включить в индекс производства)*

5) β0, β1,…, βk – параметры спецификации. Для разных видов функции fколичество этих параметров, их смысл и названия разные.

Например, для линейной спецификации

yk = β0 + β1xk + εk; k = 1, 2,…,N - этих параметров два;

параметр β0 – свободный член,

β1 – угловой коэффициент.

Задачи эконометрики:

1) По имеющимся статистическим данным подбор спецификации, наиболее точно отражающей экономическую тенденцию. Эта задача может быть уже решена экономической или эконометрической теорией.

2) Оценивание параметров спецификации β0, β1,…, βk и определение качества этих оценок и эконометрической модели в целом. (*Насколько хорошо модель объясняет изменение показателя y)*.

3) Формирование прогнозов на основе построенной модели и выработка рекомендаций для эффективных экономических решений.

Классификация переменных эконометрической модели и типов данных

Эконометрические модели отражают статистические закономерности, устанавливаемые экономической наукой и могут применяться как на макро-, так и на микроуровне. Целью их применения является количественный анализ и прогнозирование взаимосвязей показателей, описывающих экономический объект для подготовки и принятия обоснованных экономических решений.

При моделировании экономических процессов используют следующие типы данных:

1) Пространственные (перекрестные) данные (cross-sectional data);
2) Временные ряды (time-series data);
3) Панельные данные (panel data).

Пространственными данными называется совокупность экономической информации, которая характеризует различные объекты, однако полученной за один и то же период или момент времени.

Пространственные данные являются выборочной совокупностью из некоторой генеральной совокупности. Примером пространственных данных может служить комплекс экономической информации по какому-либо предприятию (численность работников, объём производства, размер основных фондов), объёмах потребления продукции определённого вида, данные о ВВП различных стран в каком-либо конкретном году и т. д. (рис. 5).

Временными данными называется совокупность экономической информации, которая характеризует один и тот же объект, но за разные периоды времени.

	Наименование региона	Валовой региональный продукт, всего, млн. руб.	Среднегодовая численность занятых в экономике, тыс. человек	Инвестиции в основной капитал (в фактически действовавших ценах), млн. руб.
1	Тюменская область	4 618 711.00	1 963.00	1 439 576.00
3	Свердловская область	1 484 447.00	2 043.20	341 635.00
4	Республика Татарстан (Татарстан)	1 436 933.00	1 821.80	464 745.00
5	Красноярский край	1 192 648.00	1 439.00	376 090.00
7	Республика Башкортостан	1 154 056.00	1 797.10	232 873.00
8	Самарская область	941 611.00	1 507.30	204 165.00
9	Пермская область	897 598.00	1 298.70	158 314.00
10	Челябинская область	843 340.00	1 672.90	179 728.00
11	Нижегородская область	838 559.00	1 703.20	258 176.00
12	Иркутская область	743 764.00	1 137.00	156 470.00
13	Кемеровская область	717 700.00	1 305.40	264 440.00

Рис. 5 Валовой региональный продукт, всего, млн. руб., 2012

Отдельно взятый временной ряд можно рассматривать как выборку из бесконечного ряда значений показателей во времени. Примером временных данных могут служить данные о динамике индекса потребительских цен, ежедневные обменные курсы валют (рис. 6).

годы	2009	2010	2011	2012	2013	2014
Численность населения (на конец года), млн. человек	142,8	142,9	143,0	143,3	143,7	146,3

Рис. 6 Численность населения России, млн. человек

Отличия временных данных от пространственных данных:

1) единицы временных рядов подвержены явлению автокорреляции (зависимости между прошлыми и текущими наблюдениями временного ряда), т. е. они не являются статистически независимыми в отличие от единиц случайной пространственной выборки;

2) единицы временных рядов не являются одинаково распределёнными величинами;

3) в отличие от пространственных данных временные данные естественным образом упорядочены во времени.

Панельными данными называются данные, содержащие сведения об одном и том же множестве объектов за ряд последовательных периодов времени.

Панельные данные являются обобщением или комбинацией пространственных и временных данных. Примером панельных данных могут служить показатели хозяйственной деятельности совокупности предприятий, которые собираются каждый год. В этом случае мы получим массив данных, в котором содержатся и данные об однородных объектах за один и тот же период

времени, и последовательные значения одной экономической переменной в различные периоды времени. Но если совокупность предприятий из года в год будет различна, то такие данные уже не будут панельными.

Набором признаков называется совокупность экономической информации, которая характеризует изучаемый процесс или объект.

Признаки взаимосвязаны между собой, и при этом они могут выступать в одной из двух ролей:

1) в роли результативного или зависимого признака;

2) в роли факторного или независимого признака.

В эконометрических моделях результативный признак называется объясняемой переменной, а факторный признак называется объясняющей переменной.

В эконометрическом моделировании выделяют следующие виды экономических переменных:

1) экзогенные или независимые переменные (x), значения которых задаются извне. В определённой степени экзогенные переменные поддаются управлению;

2) эндогенные или зависимые переменные (y), значения которых определяются внутри модели;

3) лаговые переменные – это экзогенные или эндогенные переменные, которые относятся к предыдущим моментам времени и находятся в эконометрической модели одновременно с переменными, относящимися к текущему моменту времени. Например, xt-1 – это лаговая экзогенная переменная, а yt-1 – это лаговая эндогенная переменная;

4) предопределённые или объясняющие переменные – это лаговые (xt-1) и текущие (x) экзогенные переменные, а также лаговые эндогенные переменные (yt-1).

5) фиктивные переменные используются в эконометрических моделях для характеристики явления или процесса, в отношении которого нет данных по качественному признаку;

6) переменные-заместители искусственно вводятся в эконометрическую модель для характеристики явления или процесса, который не может быть количественно охарактеризован. При этом переменная-заместитель тесно коррелирует с этим явлением.

В эконометрических исследованиях большое внимание уделяется проблеме данных, т. е. специальным методам работы при наличии данных с пропусками, влиянию агрегирования данных на эконометрические измерения. Зачастую по единицам исследуемой совокупности информация отсутствует, а в наличии имеются данные, характеризующие более крупные единицы (агрегаты). Следует отметить, что при агрегировании временных данных опасность искажения результатов измерений гораздо больше, чем при агрегировании пространных данных, потому что с одной стороны, добавляется эффект автокорреляции, а с другой – происходит погашение случайной компоненты.

Глава 2. Основы статистического анализа.

Тема 2.1 Генеральная совокупность и выборка

При исследовании реальных экономических процессов приходится обрабатывать большие объемы статистических данных по самым разнообразным показателям, которые по своей сути являются случайными величинами. По ходу проводимого анализа часто возникает необходимость оценивания числовых значений различных параметров, неоднократно приходится выдвигать и проверять различные предположения, устанавливать наличие и силу зависимости между разнообразными факторами. На практике мы сталкиваемся с конкретными реализациями рассматриваемых СВ. Количество таких реализаций носит ограниченный характер, что не позволяет применять напрямую теоретические методы анализа. Поэтому здесь в первую очередь используются методы и модели математической статистики (в частности, выборочный метод), позволяющие получить необходимые знания об исследуемом объекте, осуществить направленный анализ и сделать обоснованные выводы.

В статистическом анализе можно выделить 4 основных этапа:

1. Формулирование цели исследований;

2. Наблюдение за исследуемым объектом и сбор данных;

3. Анализ статистических данных;

4. Предсказание.

1. Формулирование целей

Целью статистических исследований – является установление закономерности поведения исследуемых объектов или процессов, чтобы имелась возможность предсказать их поведение в каких-либо условиях.

Исследовать поведение объектов при всех возможных ситуациях не возможно, так как их слишком много; и только установленные закономерности позволяют с некоторой степенью доверия предсказать поведение в ещё не встречавшихся ситуациях. Очевидно, что предсказание абсолютно точно осуществить нельзя, поэтому следует всегда стремиться осуществить такое описание закономерностей, которое дает наименьшую погрешность предсказания.

2. Наблюдение и сбор данных

Наблюдения связаны с регистрацией наблюдаемых событий, процессов, явлений, которые подвергаются изучению. Обычно регистрируют значения отдельных признаков, характеризующих объект исследования.

Признак – это объективная характеристика единицы статистической совокупности, характерная черта или свойство, которое может быть определено или измерено. Признаками, характеризующими промышленное предприятие, является выручка от реализации продукции, прибыль, стоимость основных фондов, численность персонала и др. Признаками человека являются возраст, пол, место жительства, профессия, среднемесячный доход и пр. Для любых окружающих нас объектов и явлений можно выделить достаточно большое число признаков, которые наблюдаются или потенциально могут наблюдаться в процессе статистического исследования.

Возможное значение, которое может принимать признак, называется *вариантом*. Например, существуют всего четыре варианта значений признака «экзаменационная оценка»: «2», «3», «4», «5». Если же учитывать оценки, проставляемые в зачетную книжку бакалавра или магистра, то таких вариантов

остается три, так как неудовлетворительная оценка в зачетку не проставляется. У отдельно взятого учащегося в зачетке могут быть и десять, и двадцать, и более значений признака «экзаменационная оценка», но вариантов будет по-прежнему три, а возможно, два или один, если, например, студент или слушатель учится без троек и четверок.

Признаки подразделяются на количественные и качественные, а последние, в свою очередь, на альтернативные, атрибутивные и порядковые.

Количественным является признак, отдельные варианты которого имеют числовое выражение и отражают размеры, масштабы изучаемого объекта или явления (например, площадь жилого помещения, цена товара, стаж работы).

Альтернативным называется признак, имеющий только два варианта значений. *Атрибутивный* признак имеет более двух вариантов, которые при этом выражаются в виде понятий или наименований. К атрибутивным признакам относятся район проживания, вид продукции, специальность работника, цвет товара. Такие признаки имеют место в различных областях исследования, но в большей степени они характерны для информации, с которой работают маркетологи, социологи, психологи.

Порядковые признаки отличаются от атрибутивных тем, что они имеют несколько ранжированных, т.е. упорядоченных по возрастанию или убыванию, качественных вариантов. Примерами таких признаков являются уровень образования (начальное, общее среднее и т.д.), уровень квалификации, воинское звание, различного рода рейтинги.

Приведенные выше примеры показывают, что изучаемые статистикой признаки, как правило, подвержены вариации.

Вариация – это колеблемость, изменение величины признака в статистической совокупности, т.е. принятие единицами совокупности или их группами разных значений признака.

3. Анализ статистических данных

На этом этапе происходит описание закономерности поведения изучаемых объектов. Это осуществляется с помощью моделей различной природы.

Модель – это описание поведения объектов, отражающее основные их свойства с некоторой точки зрения.

С использованием модели тесно связано понятие *адекватность* – соответствие чему- либо. Принято различать 2 рода адекватности:

• *сильная адекватность* означает – установление законов, которым подчиняется поведение объекта. Примером могут служить физические законы (Ньютона и др.);

• *слабая адекватность* предлагает, что используемая модель позволяет решить прикладную задачу (предсказания) с достаточной для исследователя точностью, при этом такая модель не обязана соответствовать реальным закономерностям поведения объекта.

В статистическом анализе в большинстве случаев используются модели слабой адекватности, особенно это касается исследования экономических процессов.

Принято различать классы моделей:

• информационные;

• физические (вещественные);

• вещественно-математические;

• логико-математические;

• алгоритмические (имитационные).

Информационные модели – описание свойств объектов средствами обычного разговорного языка. Любое понятие или определение – это вербальные модели.

Вещественная модель – уменьшенные копии реальных объектов. Таким образом, можно моделировать только реально существующие объекты.

Вещественно – математические модели – физические объекты иной природы, чем исследуемые, но их поведение описывается одинаковыми математическими зависимостями (структурные, геометрические, графические, цифровые, кибернетические и аналоговые модели). Часто возможность использования этой модели возникает при реализации решений дифференциальных уравнений.

Логико-математические модели – математические зависимости, с помощью которых используются свойства объектов чисто математическими средствами, или с помощью вычислительных экспериментов на компьютере.

Алгоритмические (имитационные) модели. К алгоритмическим моделям относятся такие, в которых критерии и (или) ограничения описываются математическими конструкциями, включающими логические условия, приводящие к разветвлению вычислительного процесса, и так называемые имитационные модели — моделирующие алгоритмы, имитирующие поведение элементов изучаемого объекта и взаимодействие между ними в процессе функционирования.

Также принято различать модели:

• детерминированные;

• вероятностные.

Вероятностные модели описывают поведение объектов как возможность реализации каких-либо событий с указанием вероятности таких исходов.

Использование *детерминированной модели* основывается на уверенности, что модель по значениям одних переменных позволяет точно вычислить значение других, при этом предлагается полный контроль за условиями внешней среды. Детерминированные модели – есть предельный случай вероятностных моделей.

Построение модели – означает выбор вида модели и её уточнение с помощью полученных статистических данных. *Уточнение* – это вычисление значений некоторых параметров модели.

4. Предсказание

Предсказание связано с использованием построенных моделей для получения прогноза развития исследуемых явлений и поведения объектов. При осуществлении предсказания указывается возможное будущее значение предсказываемой величины и интервал, в который случайная прогнозируемая величина попадает с заданной вероятностью. Эта вероятность и соответствующий ей интервал называется *доверительными интервалом и вероятностью.*

Одной из центральных задач математической статистики является выявление закономерностей в статистических данных, на базе чего можно будет строить соответствующие модели для принятия обдуманных решений. *Первая задача математической статистики* – указать способы сбора и

группировки статистических данных, полученных в результате наблюдений или испытаний.

Вторая задача математической статистики – разработать методы анализа статистических данных в зависимости от целей исследования. Сюда относятся: а) оценки неизвестной вероятности события; неизвестной функции распределения; неизвестных параметров известного распределения; зависимости двух или нескольких случайных величин и т. п.;

б) проверка статистических гипотез о виде неизвестного распределения; о величинах параметров известного распределения; о виде и силе зависимости между рассматриваемыми случайными величинами. Таким образом, основная задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Знание методов математической статистики и умение ими оперировать являются необходимой предпосылкой для успешного эконометрического

Пусть изучается совокупность однородных объектов относительно некоторого количественного признака, характеризующего эти объекты. Например, доход населения, количество покупателей в магазине в течение дня, количество качественных товаров в исследуемой партии и т. д.

Генеральной совокупностью называется множество всех возможных значений или реализаций исследуемой случайной величины X при данном реальном комплексе условий.

Выборкой (выборочной совокупностью) называют часть генеральной совокупности, отобранную для изучения.

Число элементов рассматриваемой совокупности называется ее *объемом*.

Изучение всей генеральной совокупности во многих случаях либо невозможно, либо нецелесообразно в силу больших материальных затрат, либо в силу уничтожения или порчи исследуемых объектов. Например, анализ среднего дохода населения г. Москвы формально предполагает наличие достоверной информации о каждом жителе города в конкретный момент времени. Получение такой информации просто невозможно. Проверка качества обуви связана с воздействием на нее различных экстремальных факторов: растяжения, сжатия, влажности, температуры, солнечных лучей, химического воздействия, что приведет к потере товарного вида исследуемой обуви. Поэтому на практике вся генеральная совокупность почти никогда не анализируется. Для осуществления выводов о генеральной совокупности в большинстве случаев используется выборка ограниченного объема. В силу этого задача математической статистики состоит в исследовании свойств выборки и обобщении этих свойств на генеральную совокупность. Полученный при этом вывод называется *статистическим*.

Информация о генеральной совокупности, полученная на основании выборочного наблюдения, практически всегда будет обладать некоторой погрешностью, так как она основывается на изучении только части элементов. Вряд ли средний доход и разброс в доходах, полученных по выборке объема n=1000, будет в точности таким же, что и во всем городе. Это определяет две проблемы, составляющие содержание математической теории выборки:

- как организовать выборочное наблюдение, чтобы полученная информация достаточно полно отражала пропорции генеральной совокупности (*проблема репрезентативности выборки*);
- как использовать результаты выборки для суждения по ним с наибольшей надежностью о свойствах и параметрах генеральной совокупности

(*проблема оценки*).

В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если отбор будет носить случайный характер.

Различают *повторную* и *бесповторную выборки*. В первом случае отобранный объект перед отбором следующего возвращается в генеральную совокупность. Во втором – отобранный в выборку объект не возвращается в генеральную совокупность. Если выборка составляет незначительную часть генеральной совокупности, то различие между повторной и бесповторной выборками стирается.

Случайный отбор может проводиться с помощью датчика таблицы случайных чисел либо обычной жеребьевкой. Однако строгое соблюдение правил случайного отбора не всегда осуществимо, так как оно требует четко ограниченной базы статистического анализа, каковой является генеральная совокупность, перенумеровки всех ее элементов или непосредственного их извлечения при жеребьевке. Так, при проведении обследований дохода населения в масштабах города практически невозможно составить список всех его жителей или семей с последующей организацией выборки с помощью датчика случайных чисел. Аналогично невозможно организовать опросы по изучению покупательного спроса, потребностей населения и т.д. путем образования строго случайной выборки. Поэтому прибегают к различным приемам *неслучайного* отбора, стремясь, однако, приблизиться к условиям случайного. К этим приемам относится *механический* отбор, при котором элементы генеральной совокупности, предварительно упорядоченные, отбираются по заранее установленному правилу, не связанному с вариацией исследуемого признака. Например, можно фиксировать доход каждого сотого, входящего в метро. *Серийным* называют отбор, при котором объекты выбираются из генеральной совокупности не по одному, а «сериями», которые подвергаются сплошному обследованию. Например, о продукции предприятия

можно судить по продукции, выпущенной в какой-то конкретный день. При *типическом* отборе объекты отбираются не из всей генеральной совокупности, а из каждой ее «типической» части. Например, население города можно предварительно классифицировать по социальному статусу (бизнесмены, чиновники, служащие, рабочие и т. д.). Нередко на практике применяется комбинированный отбор, при котором сочетаются описанные выше способы.

Способы представления и обработки статистических данных

Во многих случаях для анализа тех либо других экономических процессов важен порядок получения статистических данных. Но при рассмотрении так называемых перекрестных данных порядок их получения не играет существенной роли. Кроме того, результаты выборочных значений $x_1, x_2, \ldots, x_n$ количественного признака X генеральной совокупности, записанные в порядке их регистрации, обычно труднообозримы и неудобны для дальнейшего анализа. Задачей статистического описания выборки является получение такого ее представления, которое позволит наглядно выявить ее вероятностные характеристики. Для этого применяются различные формы упорядочения данных в выборке – по возрастанию, по совпадающим значениям, по интервалам и т. п.

При анализе какого-то конкретного показателя X за фиксированный момент времени (либо без учета фактора времени) наблюдаемые значения $x_1, x_2, \ldots, x_n$ обычно упорядочивают по неубыванию: $x_1 \le x_2 \le \ldots \le x_n$. Разность между максимальным и минимальным значениями СВ X называется *размахом выборки*. Пусть количество различных значений в выборке равно k ($k \le n$). Для определённости положим $x_1 < x_2 < \ldots < x_k$.

Значения x_i, $i = 1, 2, \ldots, k$ называются *вариантами*.

Пусть значение x_i встретилось в выборке n_i раз, тогда число n_i называется *частотой* значения x_i, а $\omega_i = n_i/n$ – *относительной частотой* значения x_i. Тогда наблюдаемые значения можно сгруппировать в *статистический ряд*, показанный в табл. 5:

Таблица 1

X	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
$\omega_i = n_i/n$	n_1/n	n_2/n	…	n_k/n

$$\sum_{i=1}^{n} n_i = n$$

$$\sum_{i=1}^{n} \frac{n_i}{n} = 1$$

По статистическому ряду можно построить *эмпирическую функцию распределения* $F^*(x)$:

$$F^* = \frac{n_x}{n}$$

где n_x – число значений случайной величины X меньших, чем x; n – объем выборки. По определению $F^*(x)$ обладает следующими свойствами:

1. $0 \le F^*(x) \le 1$;

2. для любых $x_1 < x_2$ $F^*(x_1) \le F^*(x_2)$;

3. $F^*(x) = 0$ при $x \le x_1$; $F^*(x) = 1$ при $x > x_k$.

Эмпирическая функция распределения $F^*(x)$ является оценкой функции распределения $F(x) = P(X < x)$, которая в этом случае называется теоретической функцией распределения.

Пример 2. Анализируется прибыль X (%) предприятий отрасли. Обследованы n = 100 предприятий, данные по которым занесены в следующий статистический ряд.

Таблица 2

X	5	10	15	20	25
n_i	5	20	40	25	10
$\omega_i = n_i/n$	0,05	0,2	0,4	0,25	0,1

Необходимо построить эмпирическую функцию распределения $F^*(x)$ и ее график.

$$F^*(x) = \begin{cases} 0, & x \le 5; \\ 0{,}05, & 5 < x \le 10; \\ 0{,}25, & 10 < x \le 15; \\ 0{,}65, & 15 < x \le 20; \\ 0{,}9, & 20 < x \le 25; \\ 1, & x > 25. \end{cases}$$

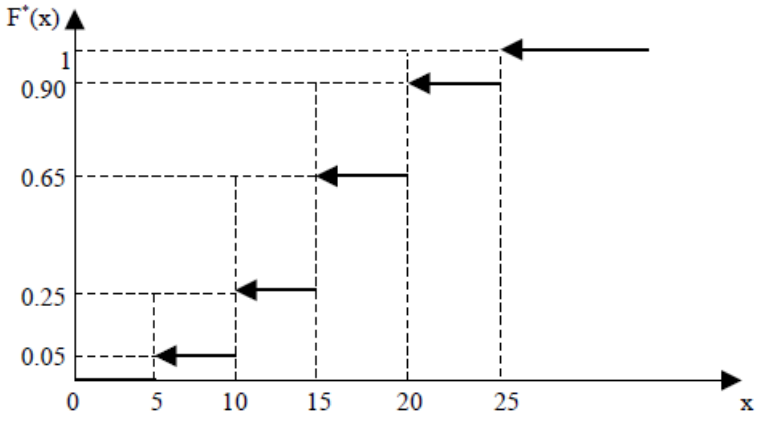


Рис.7 –График функции распределения СВ X для примера 2.

Наглядно статистический ряд может быть представлен в виде *полигона частот* (рис. 26, *а*) или *полигона относительных частот* (рис. 26, *б*):

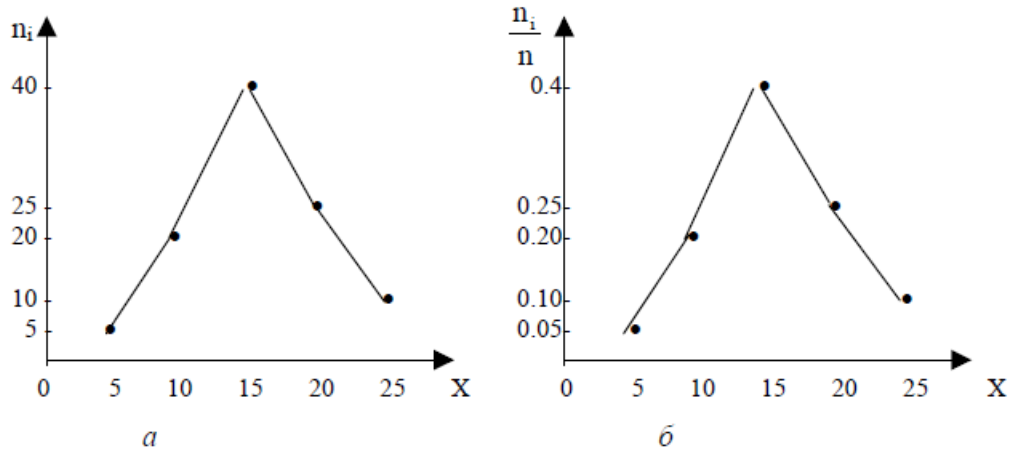


Рис.8 – Полигон частот для примера 2.

При большом объеме выборки ее элементы могут быть сгруппированы в *интервальный статистический ряд*. Для этого все n наблюдаемых значений выборки $x_1, x_2, \ldots, x_n$ разбивают по k непересекающимся подынтервалам равной длины h (h – шаг разбиения). Пусть n_i – количество наблюдаемых значений СВ X, попадающих в i-й подынтервал. $\omega_i = n_i/n$ – относительная частота попадания СВ X в i-й подынтервал. Тогда интервальный статистический ряд имеет вид:

Таблица 3

$[x_{i-1},x_i]$	$[x_0,x_1]$	$[x_1,x_2]$	…	$[x_{k-1},x_k]$
N_i	n_1	n_2	…	n_k
$\omega_i = n_i/n$	n_1/n	n_2/n	…	n_k/n

Интервальный статистический ряд наглядно может быть представлен в виде *гистограммы* – графика, в котором по оси абсцисс откладываются подынтервалы, на i-м из которых строится прямоугольник высотой $\frac{n_i}{nh}$. По виду гистограммы обычно выдвигают предположение о виде закона распределения исследуемой величины, что позволяет придать определенную направленность исследованиям.

Пример 3. Анализируется доход населения, для чего извлечена выборка объема n = 300. По уровню дохода население подразделяется на k = 6 групп. Полученные по выборке данные сгруппированы в следующий интервальный статистический ряд:

Таблица 4

$[x_{i-1},x_i]$	[0,20]	[20,40]	[40,60]	[60,80]	[80,100]	[100,120]
n_i	10	50	80	100	40	20
$\omega_i = n_i/n$	1/30	5/30	8/30	10/30	4/30	2/30

Необходимо построить гистограмму и выдвинуть предположение о виде закона распределения СВ X – дохода населения.

Отметим, что в последнюю группу могут быть включены все субъекты, чей доход превышает 100 у. е. Однако для получения теоретических выводов последний подынтервал полагается той же длины h=20, что и все предыдущие.

Построим гистограмму:

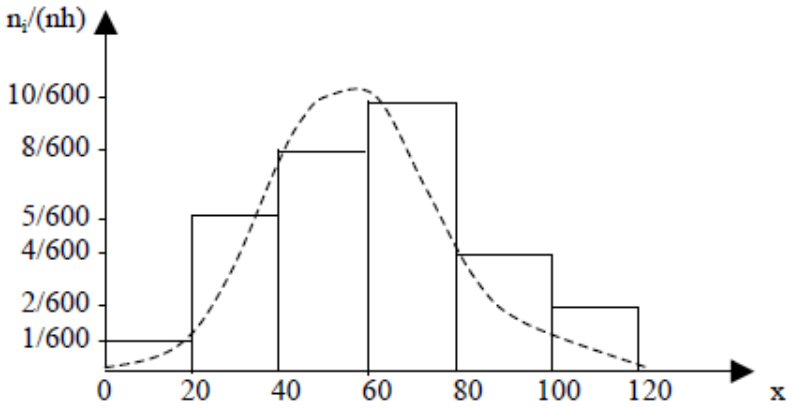


Рис.9 – Гистограмма для примера 3.

Форма гистограммы (рис. 27) в наибольшей степени соответствует нормальному закону распределения. Поэтому естественным является предположение о нормальном распределении СВ X – дохода населения (X~N(m,σ)). Следующим этапом исследования является определение параметров m и σ.

Вычисление выборочных характеристик

Для любой СВ X кроме определения ее функции распределения желательно указать ее числовые характеристики, важнейшими из которых являются математическое ожидание, дисперсия, среднее квадратическое отклонение. Пусть объем генеральной совокупности равен N. Тогда математическим ожиданием СВ X является *генеральное среднее*:

$$\overline{x_\Gamma} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Дисперсией СВ X является *генеральная дисперсия*:

$$D_\Gamma = \frac{1}{N}\sum_{i=1}^{N} (x_i - \overline{x_\Gamma})^2$$

Корень квадратный из генеральной дисперсии называется *генеральным средним квадратическим отклонением*:

$$\sigma_{\Г} = \sqrt{D_{\Г}}$$

Таким образом, для нахождения генеральных числовых характеристик необходим анализ всей генеральной совокупности. В силу того, что в реальности практически всегда имеют дело с выборками, приходится находить оценки указанных выше генеральных характеристик – выборочные числовые характеристики: выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение.

Выборочное среднее – это среднее арифметическое наблюдаемых значений выборки.

$$\overline{x_{\В}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

При задании выборки в виде статистического ряда $\overline{x_{\В}}$ рассчитывается по следующей формуле:

$$\overline{x_{\В}} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i$$

Оценкой генеральной дисперсии является *выборочная дисперсия*:

$$D_{\В} = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - \overline{x_{\В}})^2$$

Зачастую для вычисления $D_{\В}$ применяется следующая формула:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x_{\text{В}}} + (\overline{x_{\text{В}}})^2) = \overline{x^2} - \bar{x}^2$$

При задании выборки в виде статистического ряда имеем:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x_{\text{В}}})^2$$

Корень квадратный из выборочной дисперсии называется *выборочным средним квадратическим отклонением*:

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x_{\text{В}}})^2} = \sqrt{\overline{x^2} - \bar{x}^2}$$

При задании выборки в виде интервального статистического ряда вместо xi рассматривается среднее значение i-го подынтервала: $x_i = \frac{x_{i-1}+x_i}{2}$.

Для примера 2 имеем:

$$\overline{x_{\text{В}}} = \frac{1}{100}\sum_{i=1}^{5} n_i x_i = \frac{1}{100}(5 * 5 + 20 * 10 + 40 * 15 + 25 * 20 + 10 * 25)$$
$$= 15{,}75$$

$$D_{\text{В}} = \frac{1}{100}\sum_{i=1}^{5} n_i(x_i - \overline{x_{\text{В}}})^2$$
$$= \frac{1}{100}(5 * 10{,}75^2 + 20 * 5{,}75^2 + 40 * 0{,}75^2 + 25 * 4{,}25^2 + 10 * 9{,}25^2) = 27{,}76$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{27{,}76} = 5{,}27$$

Для примера 3 имеем:

$$\overline{x_{\text{В}}} = \frac{1}{300}\sum_{i=1}^{6} n_i \overline{x_i}$$

$$= \frac{1}{300}(10 * 10 + 50 * 30 + 80 * 50 + 100 * 70 + 40 * 90 + 20 * 110) = 61{,}33$$

$$D_{\text{В}} = \frac{1}{300}\sum_{i=1}^{6} n_i (\overline{x_i} - \overline{x_{\text{В}}})^2$$

$$= \frac{1}{300}(10 * 51{,}33^2 + 50 * 31{,}33^2 + 80 * 11{,}33^2 + 100 * 8{,}67^2 + 40 * 28{,}67^2 + 20 * 48{,}67^2) = 578{,}22$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{578{,}22} = 24{,}05$$

В дальнейшем для упрощения выкладок $\overline{x_{\text{В}}}$ будем обозначать через $\bar{x}$.

По аналогичной схеме определяются статистические оценки других числовых характеристик СВ.

Выборочный коэффициент вариации V определяется отношением выборочного среднего квадратического отклонения к выборочной средней, выраженным в процентах:

$$V = \frac{\sigma_{\text{в}}}{x} * 100\%$$

Коэффициент вариации – безразмерная величина, удобная для сравнения величин рассеивания двух выборок, имеющих различные размерности.

Меры разброса (дисперсия, среднее квадратическое отклонение. коэффициент вариации) кроме оценивания рассеивания значений случайных величин обычно применяются при изучении риска различных действий со

случайным исходом, в частности при анализе риска инвестирования в ту или иную отрасль, при оценивании различных активов в портфеле и портфеля активов в целом в финансовом анализе и т. д.

Глава 3. Корреляционный анализ

Понятие корреляция появилось в середине XIX века в работах английских статистиков Ф. Гальтона и К.Пирсона. Этот термин произошел от латинского «correlatio» – соотношение, взаимосвязь.

Методы теории корреляции позволяют определять количественную зависимость между различными техническими, технологическими, организационными, экономическими и другими факторами, т. е. строить экономико-статистические модели.

Тема 3.1 Анализ количественной зависимости факторов

Различают функциональную и корреляционную зависимости. Под ***функциональной*** понимается такая зависимость, когда с изменением одного фактора изменяется другой, одному значению независимого фактора обычно соответствует только одно значение зависимого фактора.

Корреляционная зависимость – это такая зависимость, при которой изменение одной случайной величины вызывает изменение среднего значения другой. Конкретных же значений зависимого переменного, соответствующих одному значению независимого, может быть несколько.

Примером функциональной связи является соотношение выпуска и потребления продукции, когда она дефицитна: во сколько раз больше выпуск, во столько раз больше продажа (все распродается, ничего не остается в запасе). Примером корреляционной связи может служить соотношение стажа рабочих и их производительности труда. В среднем, производительность труда рабочих тем выше, чем больше их стаж. Однако бывает, что молодой рабочий (из-за влияния таких дополнительных факторов как образование, здоровье, знание новых компьютерных технологий) работает лучше пожилого. Чем больше влияние этих дополнительных факторов, тем

менее тесна связь между стажем и выработкой.

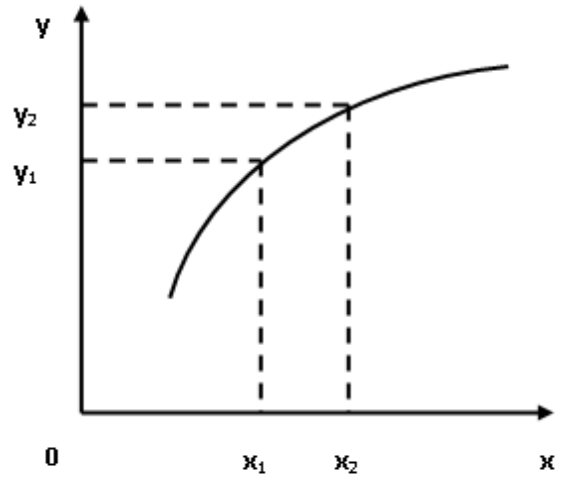


Рис.10 График функциональной зависимости

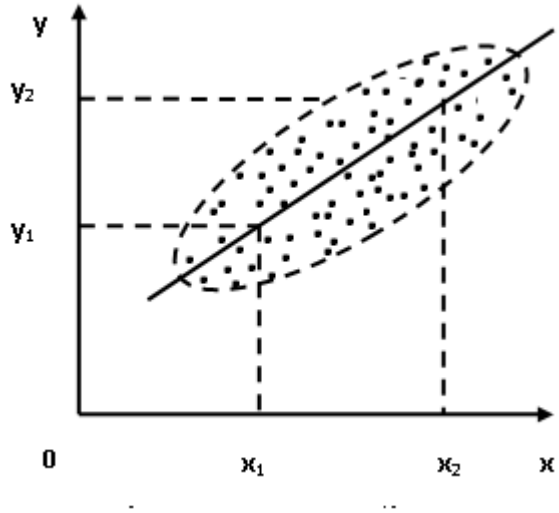


Рис.11 График корреляционной зависимости

Корреляционные зависимости могут быть установлены только при обработке большого количества наблюдений.

При корреляционном анализе решаются следующие задачи:

1. Устанавливается наличие корреляции (связи) между величинами.
2. Устанавливается форма линии связи (линии регрессии).
3. Определяются параметры линии регрессии.
4. Определяются достоверность установленной зависимости и достоверность отдельных параметров.

В простейшем случае корреляционного анализа исследуется связь между двумя показателями, из которых один рассматривается как независимый показатель-фактор (его величина обозначается через *x*), а второй – как зависимая переменная (её величина обозначается через *y*).

Наличие самой зависимости между этими показателями устанавливается,

конечно, не математическим путём, а в результате качественного анализа, позволяющего вскрыть внутреннюю сущность изучаемого явления и порождающих его причин. Сам же корреляционный анализ предназначен для количественного измерения выявленной связи, хотя он нередко способствует и уточнению выводов самого качественного анализа.

Таким образом, ещё до математического расчёта считается установленным, что связь между независимым показателем-фактором *x* и зависимой переменной *y* существует (по меньшей мере, может существовать) и характеризуется функцией y=f(x).

Одной из первых задач корреляционного анализа является установление вида этой функции, то есть отыскание такого корреляционного уравнения (иначе оно называется *уравнением регрессии*), которое наилучшим образом соответствует характеру изучаемой связи.

По форме корреляционная связь может быть прямолинейной или криволинейной. Прямолинейной может быть, например связь между уровнем производительности труда и рентабельностью производства, между количеством тренировок на тренажере и количеством правильно решаемых задач в контрольной сессии. Криволинейной может быть, например, связь между уровнем мотивации и эффективностью выполнения задачи. При повышении мотивации эффективность выполнения задачи сначала возрастает, затем достигается оптимальный уровень мотивации, которому соответствует максимальная эффективность выполнения задачи; дальнейшему повышению мотивации сопутствует уже снижение эффективности.

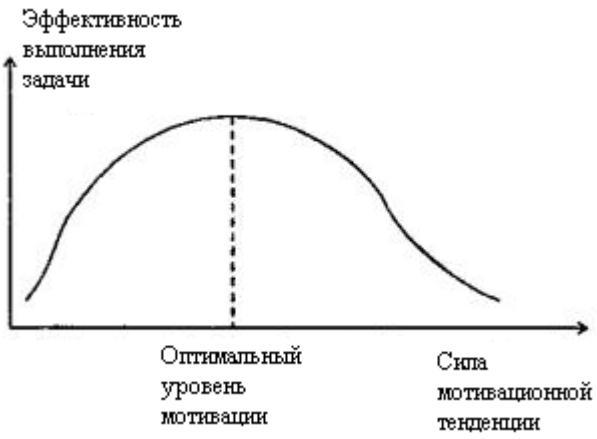


Рис.12 Криволинейная зависимость между уровнем мотивации и эффективностью выполнения задачи

По направлению корреляционная связь может быть положительной («прямой») и отрицательной («обратной»). При положительной прямолинейной корреляции более высоким значениям одного признака соответствуют более высокие значения другого, а более низким значениям одного признака - низкие значения другого.

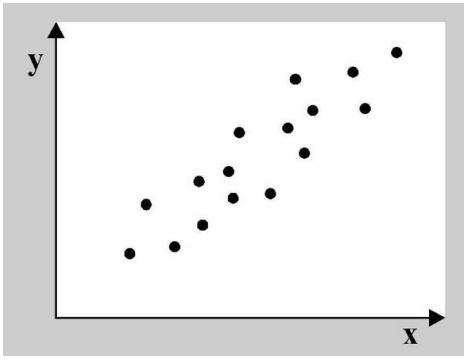


Рис.13 Прямая корреляция

При отрицательной корреляции соотношения обратные. При положительной корреляции коэффициент корреляции имеет положительный знак, при отрицательной корреляции - отрицательный знак.

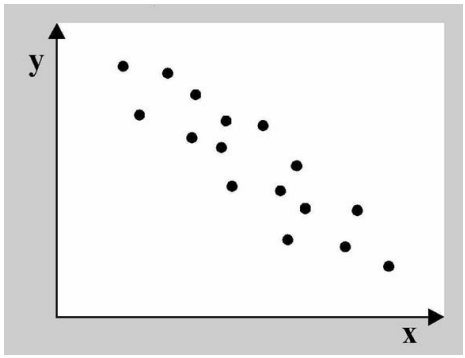


Рис.14 Обратная корреляция

Если значения x и Y разбросаны таким образом, что функцию Y=f(X) можно провести в любом направлении, то линейной связи между переменными нет.

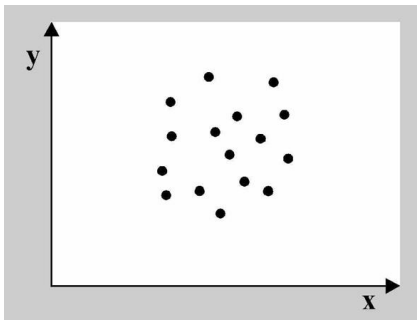


Рис.15 Отсутствие корреляции

Коэффициент корреляции и его свойства

Степень зависимости двух величин X и Y выражает ковариация и коэффициент корреляции.

Ковариацией или корреляционным моментом называется математическое ожидание произведения отклонений случайных величин X и Y от их математических ожиданий.

$$cov_{xy} = M(\big(X - M(X)\big) * \big(Y - M(Y)\big))$$

Раскрыв скобки и преобразовав формулу, получим:

$$cov_{xy} = M(X * Y) - M(X) * M(Y) = \overline{xy} - \bar{x} * \bar{y}$$

Тесноту связи между двумя величинами можно определить при помощи ***коэффициента корреляции***, который определяется по формуле:

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2} * \sqrt{N\sum y^2 - (\sum y)^2}}$$

где x и y—текущие значения наблюдаемых величин; N — число наблюдений.

Существует несколько модификаций данной формулы, наиболее простая из которых имеет вид:

$$r = \frac{cov_{xy}}{\sigma_x * \sigma_y} = \frac{\overline{xy} - \bar{x} * \bar{y}}{\sigma_x * \sigma_y}$$

где $\overline{xy}$ – среднее значение произведения двух коррелируемых величин;

$\bar{x}$ и $\bar{y}$ – средние значения этих величин;

σ_x , σ_y – среднеквадратичные (стандартные) отклонения соответствующих величин.

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2}$$

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2}$$

Отметим основные *свойства коэффициента корреляции:*

1) Коэффициент корреляции принимает значения на отрезке [-1,1], т.е

$$-1 \leq r \leq 1$$

В зависимости от того, насколько $|r|$ приближается к 1, различают связь слабую, умеренную, заметную, достаточно тесную, тесную и весьма тесную.

2) Если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина коэффициента корреляции не изменится.

3) При $r = \pm 1$ корреляционная связь представляет линейную функциональную зависимость. При этом линии регрессии У по Х и Х по Y совпадают и все наблюдаемые значения переменных располагаются на общей прямой.

4) При $r = 0$ линейная корреляционная связь отсутствует. При этом групповые средние переменных совпадают с их общими средними, а линии регрессии Y по Х и Х по Y параллельны осям координат.

Корреляционные поля

Зависимость между экспериментальными данными xi и yi графически отображается в виде геометрического места точек в системе прямоугольных координат. Эту графическую зависимость называются также диаграммой рассеивания или корреляционным полем. Корреляционное поле позволяет дать наглядную графическую интерпретацию коэффициента корреляции.

Если $r = 0$, то значения, xi, yi, полученные из двумерной нормальной совокупности, располагаются на графике в координатах x, y в пределах области, ограниченной окружностью. В этом случае между случайными величинами Х и Y отсутствует корреляция и они называются некоррелированными. Для двумерного нормального распределения некоррелированность означает одновременно и независимость случайных величин Х и Y.

Если r = 1 или r = -1, то между случайными величинами X и Y существует линейная функциональная зависимость (Y =a+bx). В этом случае говорят о полной корреляции. При r = 1 значения xi, yi определяют точки, лежащие на прямой линии, имеющей положительный наклон (с увеличением xi значения yi также увеличиваются), при r = -1 прямая имеет отрицательный наклон.

В промежуточных случаях (-1 <r< 1) точки, соответствующие значениям xi, yi, попадают в область, ограниченную некоторым эллипсом, причем при r> 0 имеет место положительная корреляция (с увеличением xi значения yi имеют тенденцию к возрастанию), при r< 0 корреляция отрицательная. Чем ближе r к ±1, тем уже эллипс и тем теснее экспериментальные значения группируются около прямой линии.

Также следует обратить внимание на то, что линия, вдоль которой группируются точки, может быть не только прямой, а иметь любую другую форму: парабола, гипербола и т. д. В этих случаях рассматривается так называемую, нелинейную (или криволинейную) корреляция.

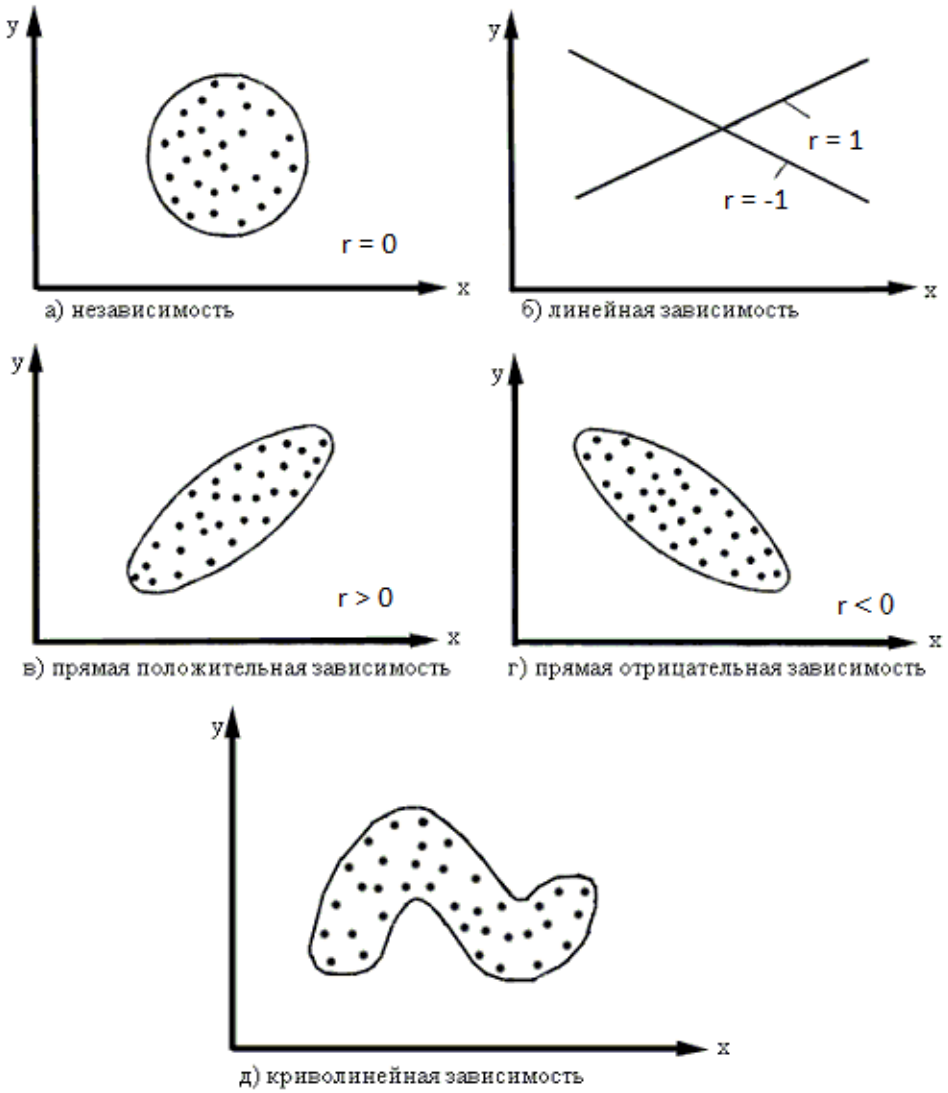


Рис.16 Виды корреляционных полей

Таким образом, визуальный анализ корреляционного поля помогает выявить не только наличия статистической зависимости (линейную или нелинейную) между исследуемыми признаками, но и ее тесноту и форму. Это имеет существенное значение для следующего шага в анализе выбора и вычисления соответствующего коэффициента корреляции.

Глава 4. Регрессионный анализ данных.

Тема 4.1 Линейная парная регрессия

Взаимосвязи экономических переменных

С тех пор, как экономика стала серьезной самостоятельной наукой, исследователи пытаются дать свое представление о возможных путях экономического развития, спрогнозировать ту или иную ситуацию, предвидеть будущие значения экономических показателей, указать инструменты изменения ситуации в желательном направлении. С другой стороны, во многих случаях различные экономисты предлагают разные, а зачастую противоположные методы решения той или иной задачи. Политики либо управляющие производством, выбирая одну из возможных стратегий решения, получают определенный результат. Плох он или хорош, и можно ли было получить лучший результат, проверить весьма затруднительно. Экономическая ситуация практически никогда не повторяется в точности, а следовательно, не позволяет применить две стратегии при одних и тех же условиях с целью сравнения конечного результата. Поэтому одной из центральных задач экономического анализа является предсказание либо прогнозирование развития некоторого экономического объекта при создании тех или иных условий. Поняв глубинные движущие силы исследуемого процесса, можно научиться рационально управлять его развитием. Поведение и значение любого экономического показателя зависят практически от бесконечного количества факторов, и все их учесть нереально. Но в этом и нет необходимости. Обычно среди факторов, воздействующих на исследуемый экономический показатель, существует лишь ограниченное количество тех, влияние которых действительно существенно. Доля оставшихся факторов столь незначительна, что их игнорирование не может привести к существенным отклонениям в поведении исследуемого объекта. Выделение и учет в модели лишь ограниченного числа реально доминирующих факторов и является серьезной

предпосылкой для качественного анализа, прогнозирования и управления ситуацией. Экономическая теория выявила и исследовала значительное число устоявшихся и стабильных связей между различными показателями. Например, хорошо изученными являются зависимости спроса или потребления от уровня дохода и цен на товары; зависимость между уровнями безработицы и инфляции; зависимость объема производства от целого ряда факторов (размера основных фондов, их возраста, качества персонала и т. д.); зависимость между производительностью труда и уровнем механизации, а также многие другие зависимости.

Любая экономическая политика заключается в регулировании экономических переменных, и она должна базироваться на знании того, как эти переменные связаны с другими переменными, ключевыми для принимающего решения политика или предпринимателя. Так, в рыночной экономике нельзя непосредственно регулировать темп инфляции, но на него можно воздействовать средствами фискальной (бюджетно-налоговой) и монетарной (кредитно-денежной) политики. Поэтому, в частности, должна быть изучена зависимость между предложением денег и уровнем цен.

Однако в реальных ситуациях даже устоявшиеся зависимости могут проявляться по-разному. Еще более сложной является задача анализа малоизученных и нестабильных зависимостей, построение моделей которых является краеугольным камнем эконометрики. Здесь следует отметить, что такие экономические модели невозможно строить, проверять и совершенствовать без статистического анализа входящих в них переменных с использованием реальных статистических данных. Инструментарием такого анализа являются методы статистики и эконометрики, в частности регрессионного и корреляционного анализа. Следует сказать, что статистический анализ зависимостей сам по себе не вскрывает существо причинных связей между явлениями, т. е. он не решает вопрос, в силу каких

причин одна переменная влияет на другую. Решение такой задачи лежит в иной плоскости и является результатом качественного (содержательного) изучения связей, которое обязательно должно либо предшествовать статистическому анализу, либо сопровождать его.

В естественных науках большей частью имеют дело со строгими (функциональными) зависимостями, при которых каждому значению одной переменной соответствует единственное значение другой.

Однако в подавляющем большинстве случаев между экономическими переменными таких зависимостей нет. Например, нет строгой зависимости между доходом и потреблением, ценой и спросом, производительностью труда и стажем работы и т. д. Это связано с целым рядом причин и, в частности, с тем, что, во-первых, при анализе влияния одной переменной на другую не учитывается целый ряд других факторов, влияющих на нее; во-вторых, это влияние может быть не прямым, а проявляться через цепочку других факторов; в-третьих, многие такие воздействия носят случайный характер и т. д. Поэтому в экономике говорят не о функциональных, а о *корреляционных*, либо *статистических* зависимостях. Нахождение, оценка и анализ таких зависимостей, построение формул зависимостей и оценка их параметров являются одним из важнейших разделов эконометрики.

Статистической называют зависимость, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой. Такую статистическую зависимость называют *корреляционной*.

Суть регрессионного анализа

Можно указать два варианта рассмотрения взаимосвязей между двумя переменными X и Y. В первом случае обе переменные считаются

равноценными в том смысле, что они не подразделяются на первичную и вторичную (независимую и зависимую) переменные. Основным в этом случае является вопрос о наличии и силе взаимосвязи между этими переменными. Например, между ценой товара и объемом спроса на него, между урожаем картофеля и урожаем зерна, между интенсивностью движения и числом аварий. При исследовании силы линейной зависимости между такими переменными мы попадаем в область корреляционного анализа, основной мерой которого является коэффициент корреляции. Вполне вероятно, что связь в этом случае вообще не носит направленного характера. Например, урожайность картофеля и зерновых обычно изменяется в одном и том же направлении, однако очевидно, что ни одна из этих переменных не является определяющей.

Другой вариант рассмотрения взаимосвязей выделяет одну из величин как *независимую (объясняющую)*, а другую – как *зависимую (объясняемую).* В этом случае изменение первой из них может служить причиной для изменения другой. Например, рост дохода ведет к увеличению потребления. Рост цены – к снижению спроса. Снижение процентной ставки увеличивает инвестиции. Увеличение обменного курса валюты сокращает объем чистого экспорта и т. д. Однако такая зависимость не является однозначной в том смысле, что каждому конкретному значению объясняющей переменной (набору объясняющих переменных) может соответствовать не одно, а множество значений из некоторой области. Другими словами, каждому конкретному значению объясняющей переменной (набору объясняющих переменных) соответствует некоторое вероятностное распределение зависимой переменной (рассматриваемой как СВ). Поэтому анализируют, как объясняющая (-ие) переменная -(ые) влияет (-ют) на зависимую переменную «в среднем». Зависимость такого типа, выражаемая соотношением:

$$M(Y|x) = f(x)$$

и называется *функцией регрессии Y на X*. При этом X называется *независимой (объясняющей) переменной (регрессором)*, Y – *зависимой (объясняемой) переменной*. При рассмотрении зависимости двух СВ говорят о *парной регрессии*. Зависимость нескольких переменных, выражаемая функцией:

$$M(Y|x_1, x_2, \dots, x_m) = f(x_{1,} x_2, \dots, x_m)$$

и называют *множественной регрессией*.

Термин *регрессия* (движение назад, возвращение в прежнее состояние) был введен Фрэнсисом Гальтоном в конце XIX века при анализе зависимости между ростом родителей и ростом детей. Гальтон заметил, что рост детей у очень высоких родителей в среднем меньше, чем средний рост родителей. У очень низких родителей, наоборот, средний рост детей выше. И в том и в другом случае средний рост детей стремится (возвращается) к среднему росту людей в данном регионе. Отсюда и выбор термина, отражающего такую зависимость.

В настоящее время под регрессией понимается функциональная зависимость между объясняющими переменными и условным математическим ожиданием (средним значением) зависимой переменной, которая строится с целью предсказания (прогнозирования) этого среднего значения при фиксированных значениях объясняющих переменных.

Для отражения того факта, что реальные значения зависимой переменной не всегда совпадают с ее условными математическими ожиданиями и могут быть различными при одном и том же значении объясняющей переменной (наборе объясняющих переменных), фактическая зависимость должна быть дополнена некоторым слагаемым ε, которое, по существу, является случайной величиной и указывает на стохастическую суть зависимости. Из этого следует, что связи между зависимой и объясняющей (-ими) переменными выражаются соотношениями:

$$Y = M(Y|x) + \varepsilon$$

$$Y = M(Y|x_1, x_2, \dots, x_m) + \varepsilon$$

и называемыми *регрессионными моделями (уравнениями)*.

Возникает вопрос, в чем причина обязательного присутствия в регрессионных моделях случайного фактора (отклонения). Этому может быть достаточно много объяснений, среди которых выделим наиболее существенные.

1. *Невключение в модель всех объясняющих переменных*. Любая регрессионная (в частности, эконометрическая) модель является упрощением реальной ситуации. Реальная ситуация всегда представляет собой сложнейшее переплетение различных факторов, многие из которых в модели не учитываются, что порождает отклонение реальных значений зависимой переменной от ее модельных значений. Например, спрос (Q) на товар определяется его ценой (P), ценой (P_S) на товары-заменители, ценой (P_C) на дополняющие товары, доходом (I) потребителей, их количеством (N), вкусами (T), ожиданиями (W) и т. д. Безусловно, перечислить все объясняющие переменные здесь практически невозможно. Например, мы не учли такие факторы, как традиции, национальные или религиозные особенности, географическое положение региона, погода и многие другие, влияние которых приведет к некоторым отклонениям реальных наблюдений от модельных, которые выразим через случайный член ε: Q = f(P, P_S, P_C, I, N, T, W, ε). Проблема здесь еще в том, что никогда заранее неизвестно, какие факторы при создавшихся условиях действительно являются определяющими, а какими можно пренебречь. Здесь уместно отметить, что в ряде случаев учесть непосредственно какой-то фактор нельзя в силу невозможности получения по нему статистических данных. Например, величина сбережений домохозяйств может определяться не только доходами его членов, но и, например, их

здоровьем, информация о котором в цивилизованных странах составляет врачебную тайну и не раскрывается. Кроме того, ряд факторов носит принципиально случайный характер (например, погода), что добавляет неоднозначности при рассмотрении некоторых моделей (например, модель, прогнозирующая объем урожая).

2. *Неправильный выбор функциональной формы модели*. Из-за слабой изученности исследуемого процесса, либо из-за его переменчивости может быть неверно подобрана функция, его моделирующая. Это, безусловно, скажется на отклонении модели от реальности, что отразится на величине случайного члена. Например, производственная функция (Y) одного фактора (X) может моделироваться функцией $Y = a + b*X$, хотя скорее должна была использоваться другая модель $Y = a*X^b$ $(0 < b < 1)$, учитывающая закон убывающей эффективности. Кроме того, неверным может быть подбор объясняющих переменных.

3. *Агрегирование переменных*. Во многих моделях рассматриваются зависимости между факторами, которые сами представляют сложную комбинацию других, более простых переменных. Например, при рассмотрении в качестве зависимой переменной совокупного спроса проводится анализ зависимости, в которой объясняемая переменная является сложной композицией индивидуальных спросов, оказывающих на нее определенное влияние помимо факторов, учитываемых в модели. Это может оказаться причиной отклонения реальных значений от модельных.

4. *Ошибки измерений*. Какой бы качественной ни была модель, ошибки измерений переменных отразятся на несоответствии модельных значений эмпирическим данным, что также отразится на величине случайного члена.

5. *Ограниченность статистических данных*. Зачастую строятся модели, выражаемые непрерывными функциями. Но для этого используется набор данных, имеющих дискретную структуру. Это несоответствие находит также свое выражение в случайном отклонении.

6. *Непредсказуемость человеческого фактора*. Эта причина может “испортить” самую качественную модель. Действительно, при правильном выборе формы модели, скрупулезном подборе объясняющих переменных все равно невозможно спрогнозировать поведение каждого индивидуума.

Таким образом, случайный член является отражением влияния всех описанных выше причин и не только их. Этот список может быть дополнен.

Задача построения качественного уравнения регрессии, соответствующего эмпирическим данным и целям исследования, является достаточно сложным и многоступенчатым процессом. Его можно разбить на три этапа:

1) выбор формулы уравнения регрессии;
2) определение параметров выбранного уравнения;
3) анализ качества уравнения и поверка адекватности уравнения эмпирическим данным, совершенствование уравнения.

Выбор формулы связи переменных называется *спецификацией* уравнения регрессии. В случае парной регрессии выбор формулы обычно осуществляется по *корреляционному полю (диаграмме рассеивания)* (рис. 35).

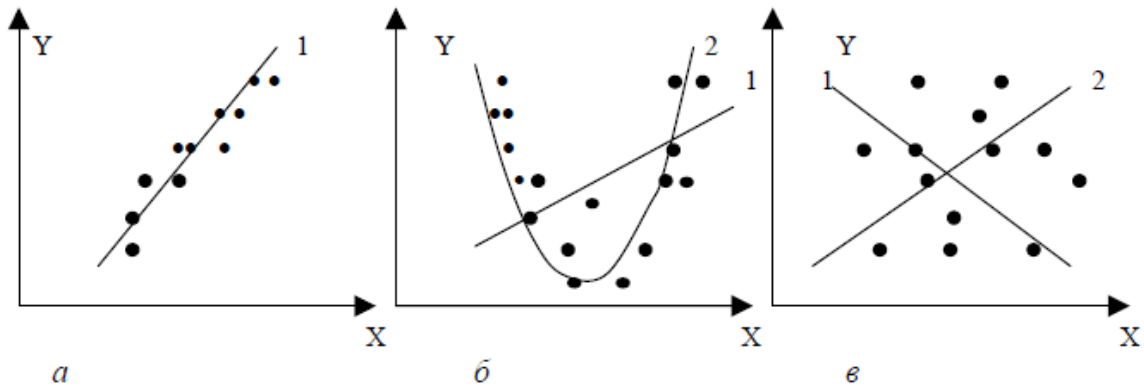


Рис.17 Подбор спецификации по корреляционному полю

На рис. 35 представлены три ситуации.

На графике 35 *а* – взаимосвязь между X и Y близка к линейной, и прямая 1 достаточно хорошо соответствует эмпирическим точкам. Поэтому в данном случае в качестве зависимости между X и Y целесообразно выбрать линейную функцию $Y = b_0 + b_1 X$.

На графике 35 *б* – реальная взаимосвязь между X и Y, скорее всего, описывается квадратичной функцией $Y = aX^2 + bX + c$ (линия 2), и какую бы мы не провели прямую (например, линия 1), отклонения точек наблюдений от нее будут существенными и неслучайными.

На графике 35 *в* – явная взаимосвязь между X и Y отсутствует. Какую бы мы не выбрали форму связи, результаты ее спецификации и параметризации (определение коэффициентов уравнения) будут неудачными. В частности, прямые 1 и 2, проведенные через центр «облака» наблюдений и имеющие противоположный наклон, одинаково плохи для того, чтобы делать выводы об ожидаемых значениях переменной Y по значениям переменной X.

В случае множественной регрессии определение подходящего вида зависимости является более сложной задачей.

Парная линейная регрессия

Если функция регрессии линейна, то речь ведут о *линейной регрессии*. Модель линейной регрессии является наиболее распространенным (и простым) уравнением зависимости между экономическими переменными. Кроме того, построенное линейное уравнение может быть начальной точкой эконометрического анализа.

Например, Кейнсом была предложена формула такого типа для моделирования зависимости частного потребления C от располагаемого дохода I: $C = C_0 + b*I$, где C_0 – величина автономного потребления, b ($0<b\leq1$) – предельная склонность к потреблению. Однако при использовании данной

модели при анализе конкретных данных мы практически всегда будем иметь определенную погрешность, т. к. строгой функциональной зависимости между этими показателями нет. Однако никто не будет отрицать, что люди (домохозяйства) с большим доходом имеют большее в среднем потребление. Данная ситуация наглядно представлена на рис. 36.

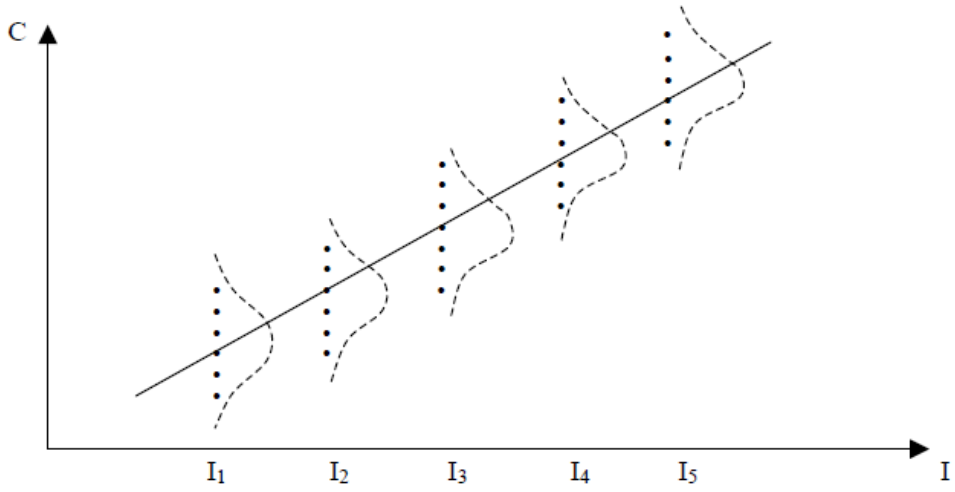


Рис.18 Зависимость потребления от располагаемого дохода

Из предыдущих рассуждений ясно, что *линейная регрессия (теоретическое линейное уравнение регрессии)* представляет собой линейную функцию между условным математическим ожиданием $Y = M(Y|x) = x_i$ зависимой переменной Y и одной объясняющей переменной X.

$$M(Y|x_i) = \beta_0 + \beta_1 x_i$$

Отметим, что принципиальной в данном случае является линейность по параметрам $\beta 0$ и $\beta 1$ уравнения.

Для отражения того факта, что каждое индивидуальное значение yi отклоняется от соответствующего условного математического ожидания, необходимо ввести в соотношение случайное слагаемое εi.

$$y_i = M(Y|X = x_i) + \varepsilon_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Это соотношение называется *теоретической линейной регрессионной моделью*; β_0 и β_1 – *теоретическими параметрами (теоретическими коэффициентами) регрессии*; ε_i – *случайным отклонением*.

Следовательно, индивидуальные значения y_i представляются в виде суммы двух компонент – систематической ($\beta_0 + \beta_1 x_i$) и случайной (ε_i). В общем виде теоретическую линейную регрессионную модель будем представлять в виде:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

При этом зависимую переменную Y называют также функцией отклика, объясняемой, выходной, результирующей, эндогенной переменной, результативным признаком; а независимую переменную X – объясняющей, входной предсказывающей, предикторной, экзогенной переменной, фактором, регрессором, факторным признаком.

Для определения значений теоретических коэффициентов регрессии необходимо знать и использовать все значения переменных X и Y генеральной совокупности, что практически невозможно.

Таким образом, задачи линейного регрессионного анализа состоят в том, чтобы по имеющимся статистическим данным (x_i, y_i), i = 1, 2, …, n, переменных X и Y:

а) получить наилучшие оценки неизвестных параметров β_0 и β_1;

б) проверить статистические гипотезы о параметрах модели;

в) проверить, достаточно ли хорошо модель согласуется со статистическими данными (адекватность модели данным наблюдений).

Следовательно, по выборке ограниченного объема мы сможем построить так называемое *эмпирическое уравнение регрессии:*

$$\hat{y}_i = b_0 + b_1 x_i$$

где $\hat{y}_i$ – оценка условного математического ожидания $Y = M(Y|X = x_i$; b_0 и b_1 – оценки неизвестных параметров β_0 и β_1, называемые *эмпирическими коэффициентами регрессии*. Следовательно, в конкретном случае

$$y_i = b_0 + b_1 x_i + e_i$$

где *отклонение* e_i – оценка теоретического случайного отклонения ε_i.

В силу несовпадения статистической базы для генеральной совокупности и выборки оценки b_0 и b_1 практически всегда отличаются от истинных значений коэффициентов β_0 и β_1, что приводит к несовпадению эмпирической и теоретической линий регрессии. Различные выборки из одной и той же генеральной совокупности обычно приводят к определению отличающихся друг от друга оценок. Возможное соотношение между теоретическим и эмпирическим уравнениями регрессии схематично изображено на рис. 37.

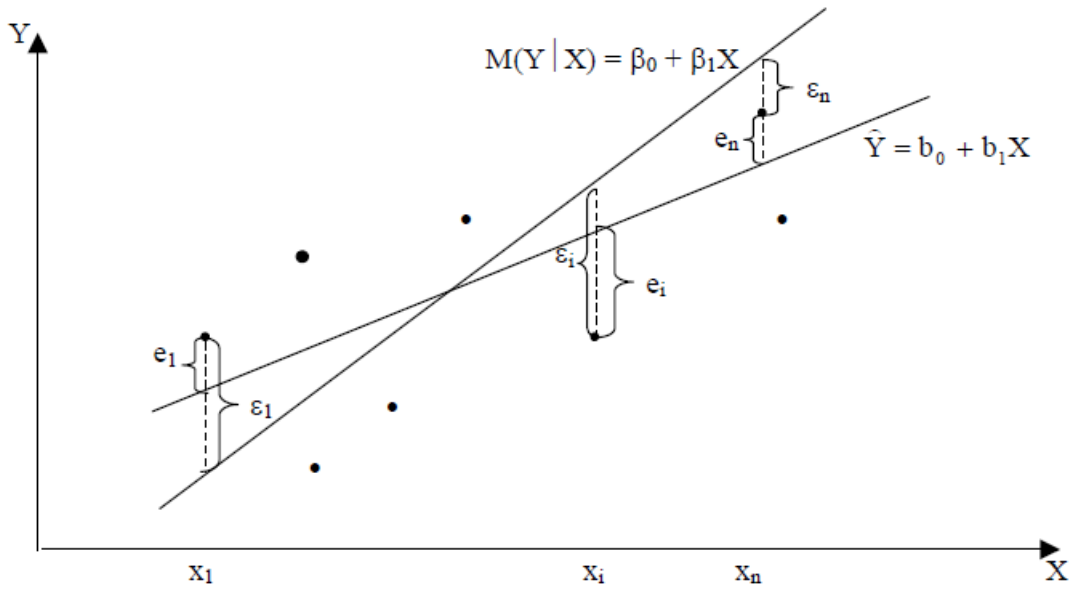


Рис.19 Соотношение между теоретическими и эмпирическими уравнениями регрессии

Задача состоит в том, чтобы по конкретной выборке (x_i, y_i), i = l, 2, … , n, найти оценки b_0 и b_l неизвестных параметров β_0 и β_l так, чтобы построенная линия регрессии являлась бы наилучшей в определенном смысле среди всех других прямых. Другими словами, построенная прямая $\hat{Y} = b_0 + b_1 X$ должна быть «ближайшей» к точкам наблюдений по их совокупности. Мерами качества найденных оценок могут служить определенные композиции отклонений e_i, i = = l, 2, … , n. Например, коэффициенты b_0 и b_l эмпирического уравнения регрессии могут быть оценены, исходя из условия минимизации одной из следующих сумм:

$$1) \quad \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)$$

$$2) \quad \sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - \hat{y}_i| = \sum_{i=1}^{n} |y_i - b_0 - b_1 x_i|$$

$$3) \quad \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Однако первая сумма не может быть мерой качества найденных оценок в силу того, что существует бесчисленное количество прямых (в частности, $Y = \bar{y}$, для которых $\sum_{i=1}^{n} e_i = 0$.

Метод определения оценок коэффициентов из условия минимизации второй суммы называется *методом наименьших модулей (МНМ)*.

Все же самым распространенным и теоретически обоснованным является метод нахождения коэффициентов, при котором минимизируется сумма квадратов отклонений $\sum_{i=1}^{n} e_i^2$. Он получил название *метод наименьших*

квадратов (МНК). Этот метод оценки является наиболее простым с вычислительной точки зрения. Кроме того, оценки коэффициентов регрессии, найденные МНК при определенных предпосылках, обладают рядом оптимальных свойств.

Среди других методов определения оценок коэффициентов регрессии отметим метод моментов (ММ) и метод максимального правдоподобия (ММП).

Метод наименьших квадратов

Пусть по выборке (x_i, y_i), i = l, 2, … , n требуется определить оценки a и b эмпирического уравнения регрессии.

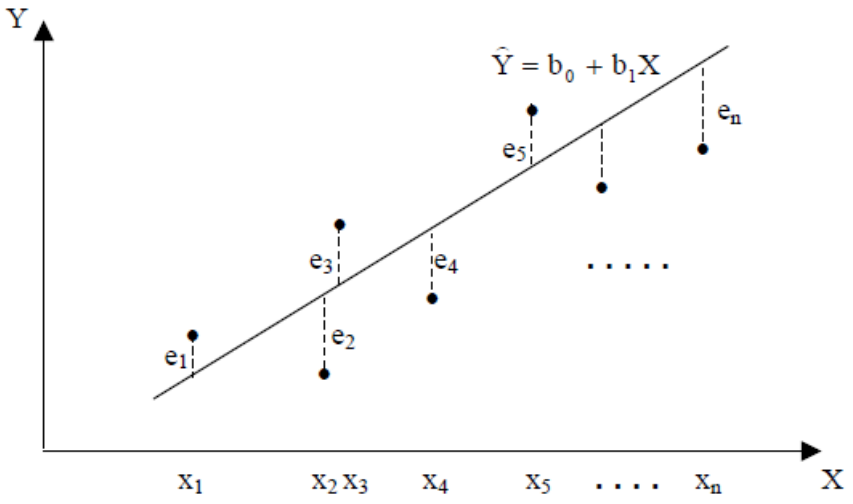


Рис.20 Исходная выборка значений x и y

В этом случае при использовании МНК минимизируется следующая функция:

$$Q(b_0, b_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Нетрудно заметить, что функция Q является квадратичной функцией двух параметров b0 и bl (Q = Q(b0, bl)), поскольку xi, yi – известные данные наблюдений. Так как функция Q непрерывна, выпукла и ограничена снизу (Q≥ 0), то она имеет минимум.

Необходимым условием существования минимума функции двух переменных является равенство нулю ее частных производных по неизвестным параметрам b0 и bl.

$$\begin{cases} \dfrac{\partial Q}{\partial b_0} = -2\sum(y_i - b_0 - b_i x_i) = 0 \\ \dfrac{\partial Q}{\partial b_1} = -2\sum(y_i - b_0 - b_i x_i)x_i = 0 \end{cases}$$

$$\begin{cases} nb_0 + b_1\sum x = \sum y \\ b_0\sum x + b_1\sum x^2 = \sum xy \end{cases}$$

Разделив оба уравнения системы на n, получим:

$$\begin{cases} b_0 + b_1\bar{x} = \bar{y} \\ b_0\bar{x} + b_1\bar{x}^2 = \overline{xy} \end{cases}$$

$$\begin{cases} b_1 = \dfrac{\overline{xy} - \bar{x}*\bar{y}}{\overline{x^2} - \bar{x}^2} \\ b_0 = \bar{y} - b_1\bar{x} \end{cases}$$

Здесь $\bar{x} = \frac{1}{n}\sum x_i$ $\overline{x^2} = \frac{1}{n}\sum x_i^2$ $\bar{y} = \frac{1}{n}\sum y_i$ $\overline{xy} = \frac{1}{n}\sum x_i y_i$

Нетрудно заметить, что b_l можно вычислить по формуле:

$$b_1 = \frac{\sum(x_i - \bar{x}) * (y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{S_{xy}}{S_x^2}$$

Тогда

$$b_1 = \frac{S_{xy}}{S_x^2} = \frac{S_{xy}}{S_x S_y} * \frac{S_y}{S_x} = r_{xy} * \frac{S_x}{S_y}$$

где r_{XY} – выборочный коэффициент корреляции; S_X, S_Y – стандартные отклонения. Таким образом, коэффициент регрессии пропорционален ковариации и коэффициенту корреляции, а коэффициенты пропорциональности служат для соизмерения перечисленных разномерных величин.

Итак, если коэффициент корреляции r_{XY} уже рассчитан, то легко может быть найден коэффициент b_1 парной регрессии.

Если, кроме уравнения регрессии Y на X ($\hat{Y} = b_0 + b_x X$), для тех же эмпирических данных найдено уравнение регрессии X на Y ($\hat{X} = c_0 + b_y Y$), то произведение коэффициентов b_x и b_y равно r^2_{xy}:

$$b_x * b_y = r_{xy} * \frac{S_y}{S_x} * r_{xy} * \frac{S_x}{S_y} = r_{xy}^2$$

Отметим, что коэффициенты c_0 и b_y находятся по формулам, аналогичным формулам нахождения b_0 и b_1:

$$\begin{cases} b_y = \dfrac{\overline{xy} - \bar{x} * \bar{y}}{\overline{y^2} - \bar{y}^2} \\ c_0 = \bar{x} - b_y \bar{y} \end{cases}$$

Проведенные рассуждения и формулы позволяют сделать ряд выводов:

1. Оценки МНК являются функциями от выборки, что позволяет их легко рассчитывать.
2. Оценки МНК являются точечными оценками теоретических коэффициентов регрессии.

3. Эмпирическая прямая регрессии обязательно проходит через точку $(\bar{x},\bar{y})$.

4. Эмпирическое уравнение регрессии построено таким образом, что сумма отклонений $\sum e_i$, а также среднее значение отклонения $\bar{e}$ равны нулю.

5. Случайные отклонения e_i не коррелированны с наблюдаемыми значениями y_i зависимой переменной Y.

6. Случайные отклонения e_i не коррелированны с наблюдаемыми значениями x_i независимой переменной X.

Для иллюстрации МНК рассмотрим следующий пример:

Пример 4. Для анализа зависимости объема потребления Y (у. е.) домохозяйства в зависимости от располагаемого дохода X (у. е.) отобрана выборка объема n=12 (помесячно в течение года), результаты которой приведены в табл. 9. Необходимо определить вид зависимости; по методу наименьших квадратов оценить параметры уравнения регрессии Y на X; оценить силу линейной зависимости между X и Y; спрогнозировать потребление при доходе X=160.

Таблица 5

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Для определения вида зависимости построим корреляционное поле:

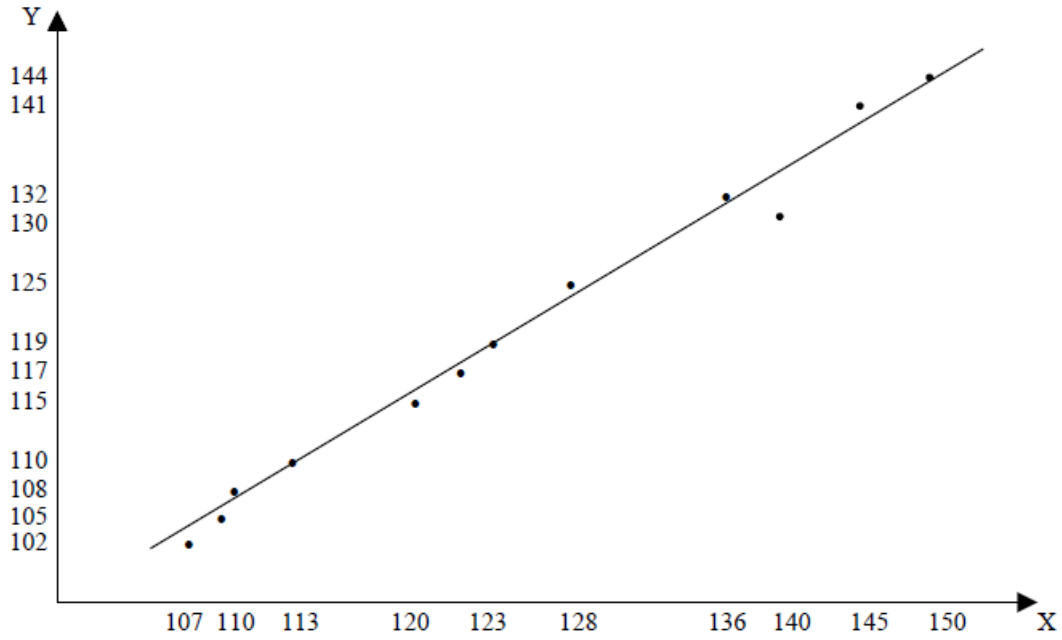


Рис.21 Корреляционное поле по данным примера 4

По расположению точек на корреляционном поле полагаем, что зависимость между X и Y – линейная: $Y= b_0 + b_1 X$.

Для наглядности вычислений по МНК построим следующую таблицу 10.

По МНК имеем:

$$\begin{cases} b_1 = \dfrac{\overline{xy} - \bar{x}*\bar{y}}{\overline{x^2} - \bar{x}^2} = \dfrac{15298{,}08 - 125{,}25 * 120{,}67}{15884{,}75 - (125{,}25)^2} = \dfrac{184{,}1625}{197{,}1875} = 0{,}9339 \\ b_0 = \bar{y} - b_1\bar{x} = 120{,}67 - 0{,}9339 * 125{,}25 = 3{,}699 \end{cases}$$

Таким образом, уравнение парной линейной регрессии имеет вид: Y=3.699+0.9339X. Построим данную регрессию на корреляционном поле.

По этому уравнению рассчитаем yi, а также $e_i = y_i - \hat{y_i}$.

Таблица 6

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\hat{y}_i$	e_i	e_i^2
1	107	102	11449	10914	10404	103.63	–1.63	2.66
2	109	105	11881	11445	11025	105.49	–0.49	0.24
3	110	108	12100	11880	11664	106.43	1.57	2.46
4	113	110	12769	12430	12100	109.23	0.77	0.59
5	120	115	14400	13800	13225	115.77	–0.77	0.59
6	122	117	14884	14274	13689	117.63	–0.63	0.40
7	123	119	15129	14637	14161	118.57	0.43	0.18
8	128	125	16384	16000	15625	123.24	1.76	3.10
9	136	132	18496	17952	17424	130.71	1.29	1.66
10	140	130	19600	18200	16900	134.45	–4.45	19.8
11	145	141	21025	20445	19881	139.11	1.89	3.57
12	150	144	22500	21600	20736	143.78	0.22	0.05
Сумма	1503	1448	190617	183577	176834	–	$^{**}\approx 0$	35.3
Среднее*	125.25	120.67	15884.75	15298.08	14736.17	–	–	–

Для анализа силы линейной зависимости вычислим коэффициент корреляции:

$$r = \frac{\overline{xy} - \bar{x}^*\bar{y}}{\sigma_x{}^*\sigma_y} = \frac{184{,}1625}{14{,}04 * 13{,}23} = 0{,}9914$$

Данное значение коэффициента корреляции позволяет сделать вывод о сильной (прямой) линейной зависимости между рассматриваемыми переменными X и Y. Это также подтверждается расположением точек на корреляционном поле.

Прогнозируемое потребление при располагаемом доходе x=160 по данной модели составит y (160) =153,12.

Построенное уравнение регрессии в любом случае требует определенной интерпретации и анализа. Интерпретация требует словесного описания полученных результатов с трактовкой найденных коэффициентов, с тем чтобы

построенная зависимость стала понятной человеку, не являющемуся специалистом в эконометрическом анализе. В нашем примере коэффициент b_1 может трактоваться как предельная склонность к потреблению (MPC=0,9339). Фактически он показывает, на какую величину изменится объем потребления, если располагаемый доход возрастает на одну единицу. На графике коэффициент b_1 определяет тангенс угла наклона прямой регрессии относительно положительного направления оси абсцисс (объясняющей переменной). Поэтому часто он называется *угловым коэффициентом*.

Свободный член b_0 уравнения регрессии определяет прогнозируемое значение Y при величине располагаемого дохода X, равной нулю (т. е. автономное потребление). Однако здесь необходима определенная осторожность. Очень важно, насколько далеко данные наблюдений за объясняющей переменной отстоят от оси ординат (зависимой переменной), так как даже при удачном подборе уравнения регрессии для интервала наблюдений нет гарантии, что оно останется таковым и вдали от выборки. В нашем случае значение b_0 = 3,699 говорит о том, что при нулевом располагаемом доходе расходы на потребление составят в среднем 3,699 у. е. Это можно объяснить в случае рассмотрения отдельного домохозяйства (оно может тратить накопленные или одолженные средства), но для совокупности домохозяйств это теряет смысл. В любом случае значение коэффициента b_0 определяет точку пересечения прямой регрессии с осью ординат и характеризует сдвиг линии регрессии вдоль оси Y.

Следует помнить, что эмпирические коэффициенты регрессии b_0 и b_1 являются лишь оценками теоретических коэффициентов $\beta 0$ и $\beta 1$, а само уравнение отражает лишь общую тенденцию в поведении рассматриваемых переменных. Индивидуальные значения переменных в силу различных причин могут отклоняться от модельных значений. В нашем примере эти отклонения

выражены через значения e_i. Эти отклонения являются оценками отклонений ε_i для генеральной совокупности.

Однако при определенных условиях уравнение регрессии служит незаменимым и очень качественным инструментом анализа и прогнозирования.

После интерпретации результатов закономерен вопрос о качестве оценок и самого уравнения в целом. Это составит предмет обсуждения следующей главы.

Тема 4.2 Проверка качества уравнения регрессии

Предпосылки МНК (условия Гаусса-Маркова)

Регрессионный анализ позволяет определить оценки коэффициентов регрессии. Но, являясь лишь оценками, они не позволяют сделать вывод, насколько точно эмпирическое уравнение регрессии соответствует уравнению для всей генеральной совокупности, насколько близки оценки b_0 и b_l коэффициентов своим теоретическим прототипам β_0 и β_1, как близко оцененное значение Y к условному математическому ожиданию $M(Y/X = x_i)$, насколько надежны найденные оценки. Для ответа на эти вопросы необходимы определенные дополнительные исследования.

Значения y_i зависят от значений x_i и случайных отклонений ε_i. Следовательно, переменная Y является случайной величиной, напрямую связанной с ε_i. Это означает, что до тех пор, пока не будет определенности в вероятностном поведении ε_i, мы не сможем быть уверенными в качестве оценок.

Доказано, что для получения по МНК наилучших результатов необходимо, чтобы выполнялся ряд предпосылок относительно случайного отклонения.

Предпосылки МНК (условия Гаусса-Маркова)

1. Математическое ожидание случайного отклонения εi равно нулю: M(εi) = 0 для всех наблюдений.

Данное условие означает, что случайное отклонение в среднем не оказывает влияния на зависимую переменную. В каждом конкретном наблюдении случайный член может быть либо положительным, либо отрицательным, но он не должен иметь систематического смещения. Отметим, что выполнимость $M(\varepsilon_i) = 0$ влечет выполнимость

$M(Y/X = x_i) = \beta_0 + \beta_1 * x_i$.

2. Дисперсия случайных отклонений εi постоянна: D(εi) = D(εj) = σ^2 для любых наблюдений i и j.

Данное условие подразумевает, что несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть либо большим, либо меньшим, не должно быть некой априорной причины, вызывающей большую ошибку (отклонение).

Выполнимость данной предпосылки называется *гомоскедастичностью* (постоянством дисперсии отклонений). Невыполнимость данной предпосылки называется *гетероскедастичностью* (непостоянством дисперсий отклонений).

Поскольку $D(\varepsilon_i) = M(\varepsilon_i - M(\varepsilon_i))^2 = M(e^2)$, то данную предпосылку можно переписать в форме: $M(e_i^{\,2}) = \sigma^{\,2}$.

3. Случайные отклонения εi и εj являются независимыми друг от друга для $i \neq j$.

Выполнимость данной предпосылки предполагает, что отсутствует систематическая связь между любыми случайными отклонениями. Другими

словами, величина и определенный знак любого случайного отклонения не должны быть причинами величины и знака любого другого отклонения.

Если данное условие выполняется, то говорят об отсутствии *автокорреляции.* Данную предпосылку можно записать в виде: $M(\varepsilon_i \varepsilon_j) = 0$ $(i \neq j)$.

4. Случайное отклонение должно быть независимо от объясняющих переменных.

Обычно это условие выполняется автоматически при условии, что объясняющие переменные не являются случайными в данной модели.

5. Модель является линейной относительно параметров.

Теорема Гаусса-Маркова. Если предпосылки *1.–5.* выполнены, то оценки, полученные по МНК, обладают следующими свойствами:

1. Оценки являются несмещенными, т. е. $M(b_0) = \beta_0$, $M(b_1) = \beta_1$. Это вытекает из того, что $M(\varepsilon_i) = 0$ и говорит об отсутствии систематической ошибки в определении положения линии регрессии.

2. Оценки состоятельны, т. к. дисперсия оценок параметров при возрастании числа n наблюдений стремится к нулю. Другими словами, при увеличении объема выборки надежность оценок увеличивается (b_0 наверняка близко к β_0, b_1 – близко к β_1).

3. Оценки эффективны, т. е. они имеют наименьшую дисперсию по сравнению с любыми другими оценками данных параметров, линейными относительно величин y_i.

В англоязычной литературе такие оценки называются *BLUE (Best Linear Unbiased Estimators) – наилучшие линейные несмещенные оценки*.

Наряду с выполнимостью указанных предпосылок при построении классических линейных регрессионных моделей делаются еще некоторые предположения. Например:

- объясняющие переменные не являются случайными величинами;
- случайные отклонения имеют нормальное распределение;
- число наблюдений существенно больше числа объясняющих переменных;
- отсутствуют ошибки спецификации;
- отсутствует совершенная мультиколлинеарность.

Проверка качества построенной модели регрессии (коэффициент детерминации, средняя ошибка аппроксимации). Средний коэффициент эластичности

Коэффициент детерминации R.

Общее качество уравнения регрессии оценивается по тому, как хорошо эмпирическое уравнение регрессии согласуется со статистическими данными. Другими словами, насколько широко рассеяны точки наблюдений относительно линии регрессии. Очевидно, если все точки лежат на построенной прямой, то регрессия Y на X «идеально» объясняет поведение зависимой переменной. В реальной жизни такая ситуация практически не встречается. Обычно поведение Y лишь частично объясняется влиянием переменной X. Возможные соотношения между двумя переменными имеют наглядную графическую интерпретацию в виде так называемой диаграммы Венна (рис. 40).

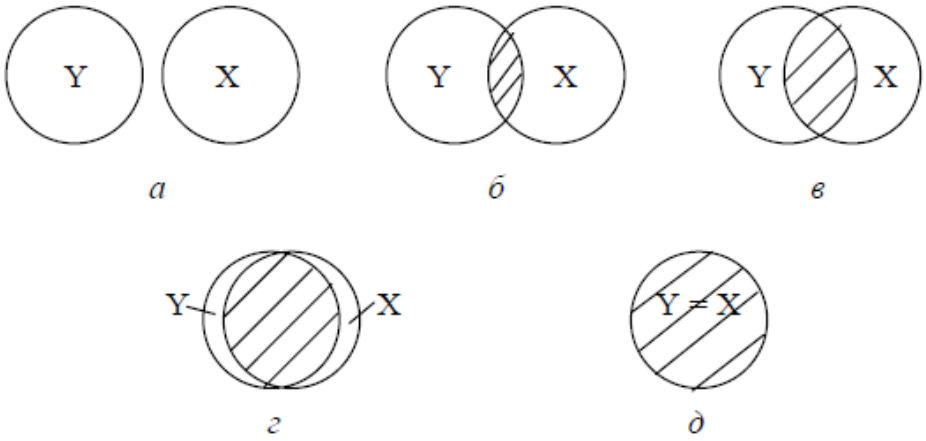


Рис.22 Диаграммы Венна

На рис. 40, *а* X никак не влияет на Y. На каждом следующем рисунке влияние X все усиливается. Наконец, на рис. 40, *д* значения Y целиком определяются значениями X.

Суммарной мерой общего качества уравнения регрессии (соответствия уравнения регрессии статистическим данным) является ***коэффициент детерминации R***. В случае парной регрессии коэффициент детерминации будет совпадать с квадратом коэффициента корреляции.

Коэффициент детерминации характеризует долю дисперсии результативного признака *y*, объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = \frac{\sigma^2_{y\text{объсн}}}{\sigma^2_{y\text{общ}}} = \frac{\sum(\widehat{y_x} - \bar{y})^2}{\sum(y - \bar{y})^2}$$

Соответственно величина (1- r^2) характеризует долю дисперсии y, вызванную влиянием остальных не учтенных в модели факторов.

Например, r^2 =0,98. Следовательно, уравнением регрессии объясняется 98% дисперсии результативного признака, а на долю прочих факторов приходится лишь 2 % ее дисперсии (т. е. остаточная дисперсия). Величина коэффициента детерминации служит одним из критериев оценки качества линейной модели. Чем больше доля объясненной вариации, тем соответственно меньше роль прочих факторов, и, следовательно, линейная модель хорошо аппроксимирует исходные данные и ею можно воспользоваться для прогноза значений результативного признака.

Средняя ошибка аппроксимации (*аппроксимация – «приближение»*) – среднее отклонение расчетных значений от фактических.

Фактические значения результативного признака отличаются от теоретических, рассчитанных по уравнению регрессии, т. е. y и $\widehat{y_x}$. Чем меньше это отличие, тем ближе теоретические значения подходят к эмпирическим данным, лучше качество модели. Величина отклонений фактических и расчетных значений результативного признака ($y - \widehat{y_x}$) по каждому наблюдению представляет собой ошибку аппроксимации. Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации как среднюю арифметическую простую:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right| * 100\%$$

Допустимый предел значений $\bar{A}$ – не более 8–10% (это свидетельствует о хорошем подборе модели к исходным данным).

Средний коэффициент эластичности $\bar{Э} = f'(x)\frac{\bar{x}}{\bar{y}}$

показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины $\bar{y}$ при изменении фактора x на 1% от своего среднего значения:

Оценка значимости уравнения регрессии

Оценка значимости уравнения регрессии в целом дается с помощью F-критерия Фишера. При этом выдвигается нулевая гипотеза, что коэффициент регрессии равен нулю, т. е. b_1=0, и, следовательно, фактор x не оказывает влияния на результат y.

Непосредственному расчету F-критерия предшествует анализ дисперсии. Центральное место в нем занимает разложение общей суммы квадратов отклонений переменной y от среднего значения $\bar{y}$на две части: «объясненную» и «необъясненную»:

$$\sum(y-\bar{y})^2 = \sum(\widehat{y_x}-\bar{y})^2 + \sum(y-\widehat{y_x})^2$$

где $\sum(y-\bar{y})^2$– общая сумма квадратов отклонений;

$\sum(\widehat{y_x}-\bar{y})^2$ – сумма квадратов отклонений, обусловленная регрессией (объясненная, или факторная);

$\sum(y-\widehat{y_x})^2$– остаточная сумма квадратов отклонений (необъясненная).

Если сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии статистически значимо и фактор x оказывает существенное влияние на результат y. Это равносильно тому, что коэффициент r_{xy}^2будет приближаться к единице.

При расчете объясненной суммы квадратов $\sum(\widehat{y_x}-\bar{y})^2$ используются теоретические (расчетные) значения результативного признака $\widehat{y_x}$, найденные по линии регрессии:

$$\widehat{y_x} = b_0 + b_1 x.$$

Сумма квадратов отклонений, обусловленных линейной регрессией, составляет:

$$\sum(\widehat{y_x} - \bar{y})^2 = b^2 * \sum(x - \bar{x})^2$$

Поскольку при заданном объеме наблюдений по *x* и *y* факторная сумма квадратов при линейной регрессии зависит только от одной константы коэффициента регрессии *b*, то данная сумма квадратов имеет одну степень свободы. Число степеней свободы – это число свободы независимого варьирования признака; оно связано с числом единиц совокупности *n* и с числом определяемых по ней констант. Существует равенство между числом степеней свободы общей, факторной и остаточной суммы квадратов. Число степеней свободы остаточной суммы квадратов при линейной регрессии составляет *n*–2.Число степеней для общей суммы квадратов составляет *n*–1, так как для $\sum(y - \bar{y})^2$ требуется *n*–1 независимых отклонений (из *n* единиц после расчета среднего уровня свободно варьируются лишь *n*–1, число отклонений).

Следовательно, имеем два равенства:

$$\sum(y - \bar{y})^2 = \sum(\widehat{y_x} - \bar{y})^2 + \sum(y - \widehat{y_x})^2$$
$$\mathrm{n} - 1 = 1 + (\mathrm{n} - 2)$$

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получим средний квадрат отклонений, или, что тоже самое, дисперсию на одну степень свободы D:

$$D_{\text{общ}} = \frac{\sum(y - \bar{y})^2}{n - 1}$$

$$D_{\text{факт}} = \frac{\sum(\widehat{y_x} - \bar{y})^2}{1}$$

$$D_{\text{ост}} = \frac{\sum(y - \widehat{y_x})^2}{n - 2}$$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F критерия для проверки нулевой гипотезы (H_0: $D_{\text{факт}} = D_{\text{ост}}$):

$$F = \frac{D_{\text{факт}}}{D_{\text{ост}}}$$

Если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Фактическое значение *F*-критерия Фишера сравнивается с табличным значением $F_{\text{табл}}$ (α; k_1; k_2) при уровне значимости α (вероятности не принять гипотезу, при условии, то она верна) и степенях свободы $k_1 = m$ и $k_2 = n – m – 1$.

Табличное значение *F*-критерия – это максимальная величина отношения дисперсий, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы. Вычисленное значение *F*-критерия признается достоверным (отличным от единицы), если оно больше табличного. В этом случае нулевая гипотеза об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи: $F_{\text{факт}} > F_{\text{табл}}$. H_0 отклоняется.

Если же величина окажется меньше табличной $F_{\text{факт}} < F_{\text{табл}}$, то вероятность нулевой гипотезы выше заданного уровня (например,0,05) и она не может быть отклонена без серьезного риска сделать неправильный вывод о наличии связи. В этом случае уравнение регрессии считается статистически незначимым, H_0 не отклоняется.

Величина F-критерия связана с коэффициентом детерминации $R=r^2$. Факторную сумму квадратов отклонений можно представить как

$$\sum(\hat{y} - \bar{y})^2 = r^2 * \sigma_y^2 * n$$

а остаточную сумму квадратов – как

$$\sum(y - \hat{y})^2 = (1 - r^2) * \sigma_y^2 * n$$

Тогда значение F-критерия можно выразить как

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2}(n - 2)$$

где n – число единиц совокупности; m – число параметров при переменных x.

Оценка статистической значимости коэффициентов регрессии и корреляции (*t*-критерии Стьюдента)

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров (по *t*-критерию Стьюдента). С этой целью по каждому из параметров определяется его стандартная ошибка m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется последующей формуле:

$$m_b = \frac{\frac{\sqrt{\sum(y - \widehat{y_x})^2}}{(n-2)}}{\sum(x - \bar{x})^2} = \sqrt{\frac{D_{\text{ост.}}^2}{\sum(x - \bar{x})^2}} = \sqrt{\frac{D_{\text{ост.}}^2}{n * \sigma_x^2}}$$

где $D^2_{\text{ост.}}$ – остаточная дисперсия на одну степень свободы

$$D_{\text{ост}} = \frac{\sum(y - \widehat{y_x})^2}{n-2}$$

Величина стандартной ошибки совместно с t-распределением Стьюдента при n-2 степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительных интервалов.

Для оценки существующего коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т. е. определяется фактическое значение *t*-критерия Стьюдента: $t_b = \frac{b}{m_b}$, которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы (n – 2). Если фактическое значение *t*-критерия превышает табличное, то гипотезу о несущественности коэффициента регрессии можно отклонить.

Можно доказать равенство $t_b{}^2 = F$:

$$t_b{}^2 = \frac{b^2}{m_b^2} = \frac{b^2}{\frac{\sum(y - \widehat{y_x})^2/(n-2)}{\sum(x - \bar{x})^2}} = \frac{b^2\sum(x - \bar{x})^2}{\sum(y - \widehat{y_x})^2/(n-2)} = \frac{\sum(\widehat{y_x} - \bar{y})^2}{\sum(y - \widehat{y_x})^2/(n-2)}$$

$$= \frac{D_{\text{факт}}}{D_{\text{ост}}} = F$$

Стандартная ошибка параметра *a* определяется по следующей формуле:

$$m_a = \sqrt{\frac{\sum(y - \widehat{y_x})^2}{n-2}\frac{\sum x^2}{n\sum(x - \bar{x})^2}} = \sqrt{D^2 * \frac{\sum x^2}{\sum(x - \bar{x})^2}} = \sqrt{D^2_{\text{ост}} * \frac{\sum x^2}{n^2 * \sigma_x^2}}$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии, вычисляется *t*-критерий: $t_a = \frac{a}{m_a}$, его величина сравнивается с табличным значением при n-2 степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Фактическое значение t-критерия определяется как:

$$t_r = \frac{r}{m_r} = \frac{r}{\sqrt{1 - r^2}}\sqrt{n - 2}$$

Данная формула свидетельствует, что в парной линейной регрессии $t_r^2 = F$, так как $F = \frac{r_{xy}^2}{1 - r_{xy}^2}(n - 2)$, кроме того, $t_b^2 = F$, следовательно, $t_r^2 = t_b^2$.

Таким образом, проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

Для расчета доверительных интервалов определяются предельные ошибки каждого показателя:

$$\Delta a = t_{\text{табл}} m_a$$

$$\Delta b = t_{\text{табл}} m_b$$

Формулы для расчета доверительных интервалов имеют следующий вид:

$$a - t_{\text{табл}} m_a \le a \le a + t_{\text{табл}} m_a$$

$$b - t_{\text{табл}} m_b \le b \le b + t_{\text{табл}} m_b$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр

принимается нулевым, так как он не может одновременно принимать и положительное и отрицательное значения.

Прогнозирование с помощью парных линейных регрессий

Прогнозирование с помощью парной линейной регрессии может осуществляться путем подстановки значения объясняющей переменной в полученное уравнение регрессии. Например, получено регрессионное уравнение, связывающее цену на товар (X) и количество проданного товара (Y), **y = 320,73 − 1,04x**. Для определения объема продаж по цене $x_0 = 120$ руб. необходимо подставить x_0в уравнение.

Прогноз продаж: **y (120) = 320,73−1,04· 120=195,9 кг**. Конечно, такой прогноз отвечает на вопрос, каково будет среднее значение результата. Однако зачастую необходимо знать диапазон возможных значений результатов с заданной надежностью, для этого используется оценка доверительных интервалов, позволяющая производить интервальное (а не точечное, как по уравнению регрессии) прогнозирование.

Тема 4.3 Нелинейная регрессия

Основные типы нелинейных регрессий

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций: например, равносторонней гиперболы, параболы второй степени и др.

Различают два класса нелинейных регрессий:

- регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;
- регрессии, нелинейные по оцениваемым параметрам.

Примером нелинейной регрессии по включаемым в нее объясняющим переменным могут служить следующие функции:

- полиномы разных степеней $y=a+b_1x+b_2x^2+\varepsilon$; $y=a+b_1x+b_2x^2+b_3x^3+\varepsilon$.

- равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$.

К нелинейным регрессиям по оцениваемым параметрам относятся функции:

– степенная $y = a * x^b * \varepsilon$;

– показательная $y = a * b^x * \varepsilon$;

– экспоненциальная $y = e^{a+bx} * \varepsilon$.

Примеры графиков нелинейных регрессий:

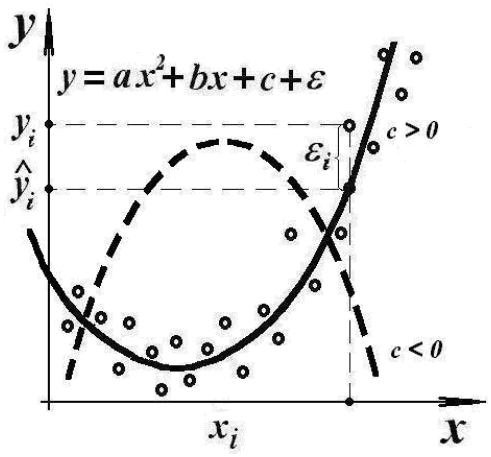


Рис.23 полиномиальные - квадратичные зависимости $y = ax^2 + bx + c + \varepsilon$

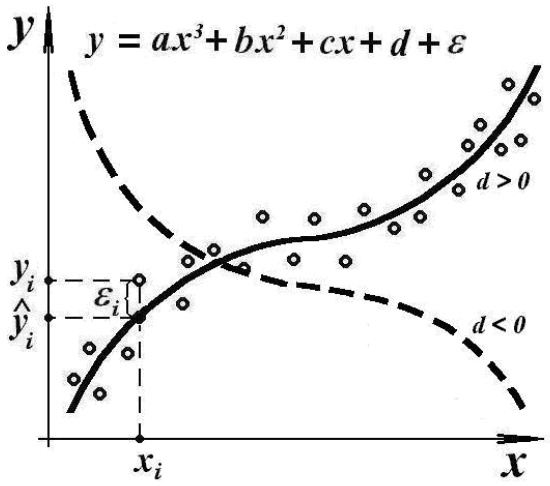


Рис.24 полиномиальные - кубические зависимости $y = ax^3 + bx^2 + cx + d + \varepsilon$

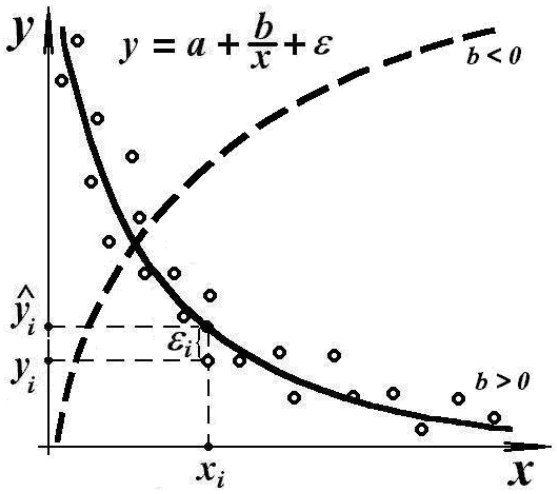


Рис.25 гиперболические зависимости $y = a + \frac{b}{x} + \varepsilon$

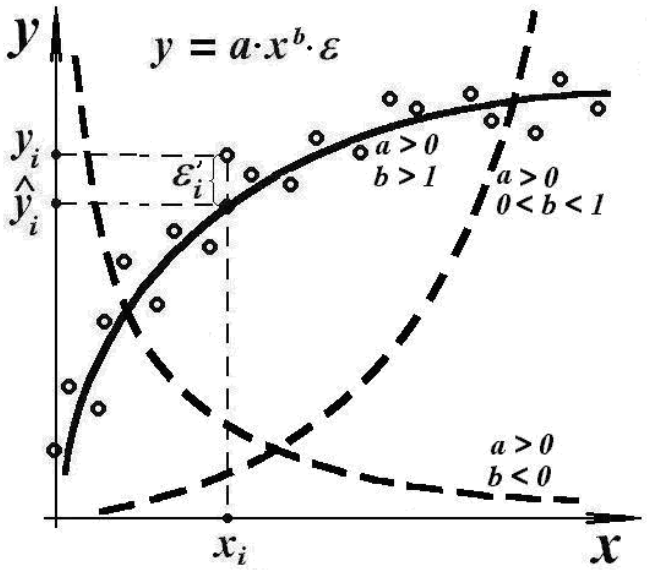


Рис.26 степенные зависимости $y = ax^b\varepsilon$

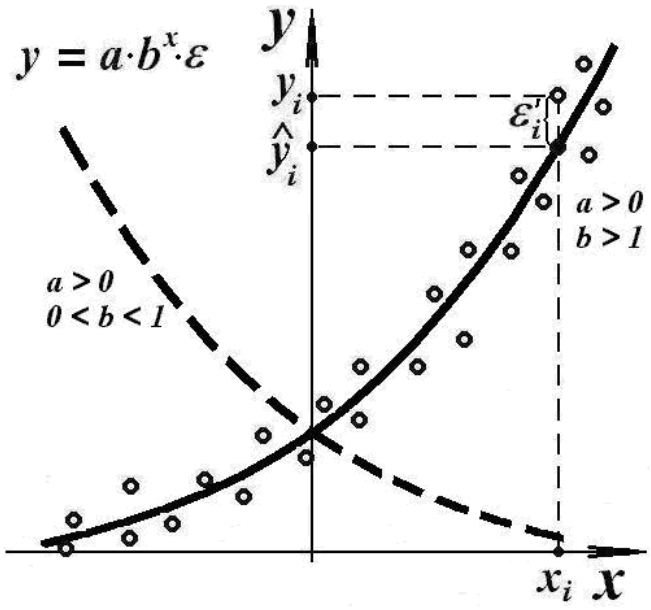


Рис.27 показательные $y = ab^x\varepsilon$, в т. ч. экспоненциальные зависимости $y = e^{a+bx+\varepsilon}$

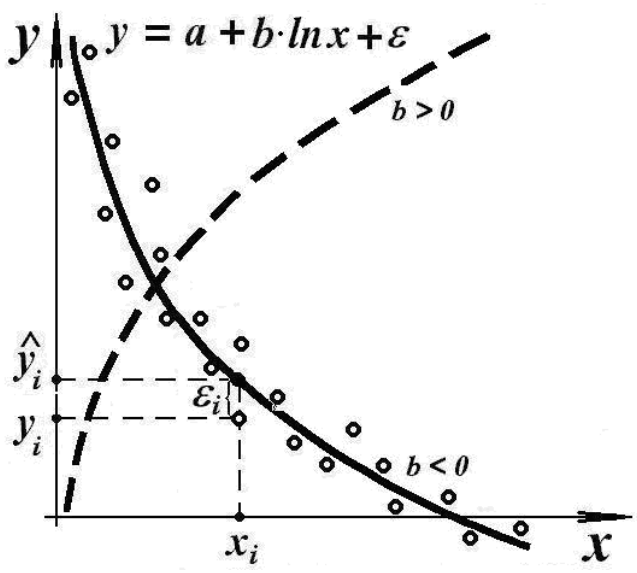


Рис.28 логарифмические зависимости $y = a + b^{*}\ln x + \varepsilon$

Кроме указанных нелинейных регрессий, возможно использование в экономической практике других, в том числе и получающихся из вышеперечисленных путём математических преобразований, например:

$$y = \frac{1}{a + bx + \varepsilon}$$

$$y = \frac{1}{ax^2 + bx + c + \varepsilon}$$

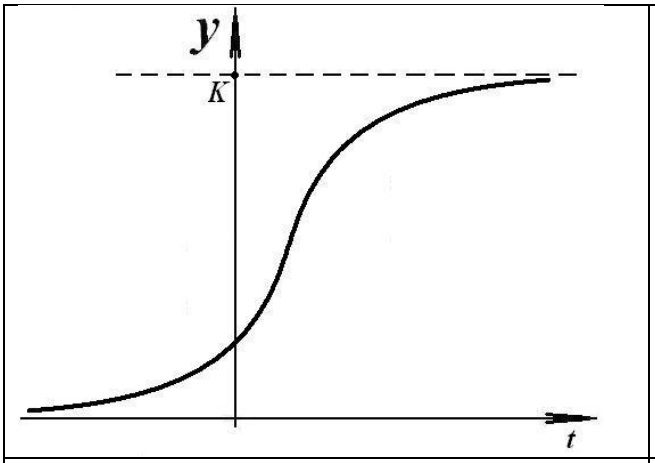
$$\ln y = a + bx + \varepsilon$$

$$\ln \hat{y} = ax^2 + bx + c + \varepsilon$$

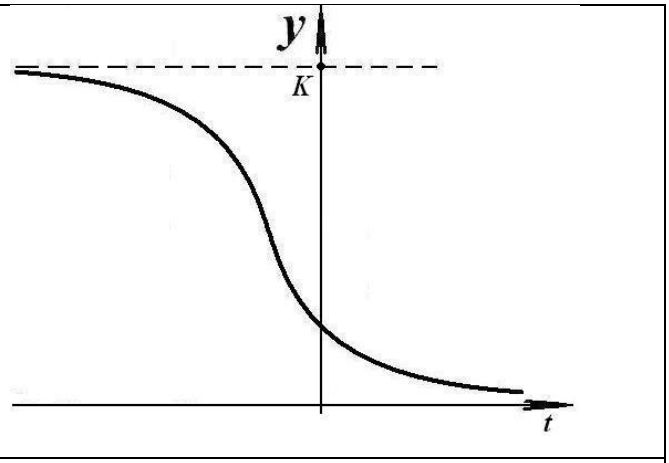
Широкое применение в экономике и социологии получили так называемые *кривые с насыщением*, выходящие на асимптоту при достижении определенных значений показателей. Например, при анализе временных рядов в страховом деле, банковском деле, демографии часто используется *кривая Гомперца*, существенно изменяющая свой вид при различных значениях параметров:

$$y = Kab^t$$

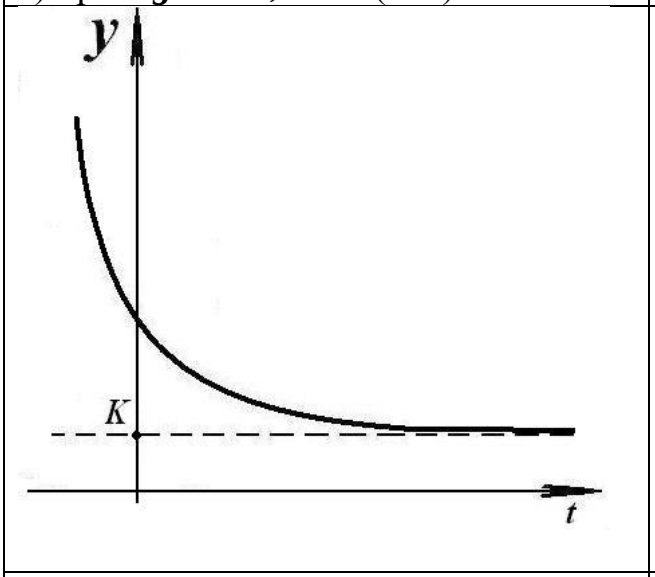
где K, a, b - параметры, t – время (1, 2, …).



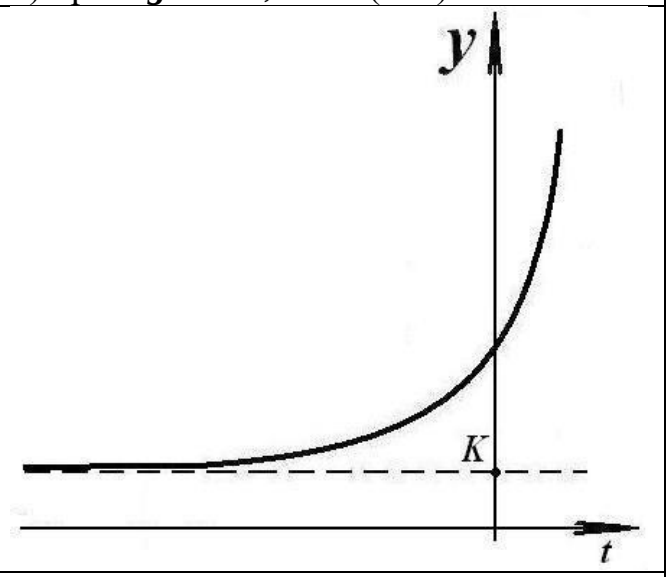
а) при $\log_c a < 0$, $b < 1$ (c>1)



б) при $\log_c a < 0$, $b > 1$ (c>1)



в) при $\log_c a > 0$, $b < 1$ (c>1)



г) при $\log_c a > 0$ $_c$, $b > 1$ (c>1)

Рис.29 Кривая Гомперца $y = Kab^t$ при различных значениях параметров.

Находит применений и *логистическая кривая* $y = \frac{K}{1+be^{-at}}$

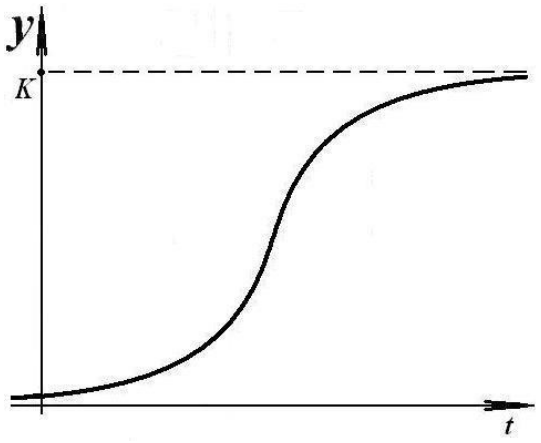


Рис.30 Логистическая кривая

В подобных функциях параметр ***K*** определяется как максимум функции, а в дальнейшем параметры ***a*** и ***b*** определяются методом наименьших квадратов.

В парной регрессии выбор вида функции может осуществляться *следующими методами*:

- аналитическим;
- графическим;
- экспериментальным.

На практике необходимо по возможности использовать наиболее простые зависимости. Попытка получить некую «универсальную» регрессию для всего изучаемого диапазона объясняющей переменной может привести к снижению точности модели.

Для повышения точности и упрощения модели возможно разделение диапазона ***x*** на несколько участков, каждый со своими коэффициентами уравнения регрессии или даже своей спецификацией модели.

Линеаризация некоторых нелинейных регрессий

Метод наименьших квадратов не предназначен для оценки параметров нелинейных регрессий, кроме полиномиальной.

Так, в параболе второй степени:

$$y = a_0 + a_1 * x + a_2 * x^2 + \varepsilon$$

заменив переменные $x = x_1$, $x^2 = x_2$, получим двухфакторное уравнение линейной регрессии:

$$y = a_0 + a_1 * x_1 + a_2 * x_2 + \varepsilon$$

для оценки параметров которого используется МНК.

Соответственно для полинома третьего порядка:

$$y = a_0 + a_1 * x + a_2 * x^2 + a_3 * x^3 + \varepsilon$$

при замене $x = x_1$, $x^2 = x_2$, $x^3 = x_3$ получим трехфакторную модель линейной регрессии:

$$y = a_0 + a_1 * x_1 + a_2 * x_2 + a_3 * x_3 + \varepsilon$$

а для полинома k-го порядка

$$y = a_0 + a_1 * x + a_2 * x^2 + \cdots + a_k * x^k + \varepsilon$$

получим линейную модель множественной регрессии с k объясняющими переменными:

$$y = a_0 + a_1 * x_1 + a_2 * x_2 + \cdots + a_k * x_k + \varepsilon$$

Следовательно, полином любого порядка сводится к линейной регрессии с ее методами оценивания параметров и проверки гипотез. Как показывает опыт большинства исследователей, среди нелинейной полиномиальной регрессии чаще всего используется парабола второй степени; в отдельных случаях – полином третьего порядка. Ограничения в применении полиномов более высоких степеней связаны с требованием однородности исследуемой

совокупности: чем выше порядок полинома, тем больше изгибов имеет кривая и соответственно меньше однородность совокупности по результативному признаку.

Применение МНК для оценки параметров параболы второй степени приводит к следующей системе нормальных уравнений:

$$\begin{cases} \sum y = na + b\sum x + c\sum x^2 \\ \sum xy = a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4 \end{cases}$$

Решение ее возможно методом определителей:

$$a = \frac{\Delta a}{\Delta};\ b = \frac{\Delta b}{\Delta};\ c = \frac{\Delta c}{\Delta};$$

где Δ - определитель системы;

$\Delta a; \Delta b; \Delta c$ - частные определители для каждого из параметров.

Ввиду симметричности кривой парабола второй степени далеко не всегда пригодна в конкретных исследованиях. Чаще всего исследователь имеет дело лишь с отдельными сегментами параболы, а не с полной параболической формой. Кроме того, параметры параболической связи не всегда могут быть логически истолкованы. Поэтому если график зависимости не демонстрирует четко выраженной параболы второго порядка (нет смены направленности связи признаков), то она может быть заменена другой нелинейной функцией, например степенной.

Однако целый класс нелинейных зависимостей можно привести к зависимостям линейным преобразованием независимой и/или зависимой переменной. Такое преобразование называется *линеаризацией*.

Задача построения нелинейной модели регрессии состоит в следующем:

Задана нелинейная спецификация модели $y = f(x,a,b,\varepsilon)$,

где y - зависимая, объясняемая переменная; x - независимая, объясняющая переменная; a, b - параметры модели, для которых должны быть получены оценки; ε – аддитивный (*сложение*) или мультипликативный (*умножение*) случайный фактор.

Требуется

1. Преобразовать исходные данные $x \rightarrow x^*$, $y \rightarrow y^*$ так, чтобы спецификация модифицированной регрессии была линейной: $y^* = a^* + b^* x^*$

2. Методом наименьших квадратов получить оценки параметров a^*, b^*.

3. По оценкам a^*, b^* вычислить искомые оценки параметров a, b исходной регрессии.

Методы линеаризации:

1) Метод логарифмирования – применяется к степенным функциям;
2) Метод обратного преобразования – для дробных функций;
3) Комплексный метод – для дробных и степенных функций.

Таблица 7

Способы преобразования данных и вычисления параметров a, b по оценкам a^*, b^*

Исходная спецификация	Преобразование	Преобразование	Вычисление b по b^*	Вычисление a по a^*
	$x \rightarrow x^*$	$y \rightarrow y^*$		

$y = a + \frac{b}{x} + \varepsilon$	$x^* = \frac{1}{x}$	$y^* = y$	$b = b^*$	$a = a^*$
$y = \frac{1}{a + bx + \varepsilon}$	$x^* = x$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
$y = \frac{x}{a + bx + x\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
$y = ae^{bx+\varepsilon}$	$x^* = x$	$y^* = \ln y$	$b = b^*$	$a = e^*$
$y = ae^{\frac{b}{x}+\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \ln y$	$b = b^*$	$a = e^*$
$y = \frac{1}{a + be^{-x} + \varepsilon}$	$x^* = e^{-x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
$y = ax^b e^{\varepsilon}$	$x^* = \ln x$	$y^* = \ln y$	$b = b^*$	$a = e^*$ $a^* = \ln(a)$

а) Показательная модель $y = ab^x\varepsilon$**;**

Прологарифмируем левую и правую части уравнения и преобразуем полученное равенство:

$$\ln y = \ln(a*b^x*\varepsilon) = \ln a + \ln b^x + \ln \varepsilon = \ln a + x\ln b + \ln \varepsilon$$

Введем промежуточные переменные

$\ln y = Y, \quad \ln a = A, \quad \ln b = B, \quad \ln \varepsilon = E$

С учетом введенных обозначений имеем парную линейную регрессию

$$Y = A + Bx + E$$

Её отличие от исходной в том, что в качестве объясняемой переменной выступает не исходная *y*, а *Y=ln y*. Т.е. в расчетах необходимо брать не исходный массив объясняемой переменной *y*, а массив предварительно полученных ее логарифмов. Но и получены в результате расчетов будут не коэффициенты *a* и *b*, а их логарифмы *ln a=A* и *ln b=B*. Из них требуемые значения можно получить:

$a = e^{\ln a} = e^{A}$, $b = e^{\ln b} = e^{B}$

Остается только подставить полученные значения коэффициентов в уравнение исходной парной нелинейной регрессии $y = ab^{x}\varepsilon$

б) *Гиперболическая модель* $y = a + \frac{b}{x} + \varepsilon$**;**

Обозначим $\frac{1}{x} = X^{*}$

С учетом введенных обозначений имеем парную линейную регрессию

$$y = a + bX + \varepsilon$$

Для расчетов необходимо указать предварительно полученный массив данных, обратных объясняющей переменной. И сразу будут получены готовые коэффициенты ***a*** и ***b*** .

в) *Квадратичная модель* $y = ax^2 + bx + c + \varepsilon$**;**

Она также может быть линеаризована и сведена к линейной множественной (конкретно - к двухфакторной) модели.

Введём промежуточные переменные $x_1 = x^2$, $x_2 = x$

Тогда имеем множественную линейную регрессию

$$y = ax_1 + bx_2 + c + \varepsilon$$

Вычисления её коэффициентов производится, как и у всех множественных линейных регрессий.

Аналогично могут быть линеаризованы и другие нелинейные регрессии.

Корреляция для нелинейной регрессии

Уравнение нелинейной регрессии, так же как ив линейной зависимости дополняется показателем корреляции, а именно индексом корреляции ρ.

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}$$

где σ^2_y – общая дисперсия результативного признака y;

$\sigma^2_{\text{ост}}$ - остаточная дисперсия, определяемая исходя из уравнения регрессии $\hat{y} = f(x)$.

Так как $\sigma^2_y = \frac{1}{n}\sum(y - \bar{y})^2$, а $\sigma^2_{\text{ост}} = \frac{1}{n}\sum(y - \hat{y})^2$, то инндекс корреляции можно выразить как

$$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \widehat{y_x})^2}{\sum(y - \bar{y})^2}}$$

Величина данного показателя находится в границах $0 \leq \rho_{xy} \leq 1$, причем, чем ближе к единице, тем теснее связь рассматриваемых признаков, тем более надежно найденное уравнение регрессии.

Выбор формы модели

Многообразие и сложность экономических процессов предопределяет многообразие моделей, используемых для эконометрического анализа. С другой стороны, это существенно усложняет процесс нахождения максимально адекватной формулы зависимости. Для случая парной регрессии подбор модели обычно осуществляется по виду расположения наблюдаемых точек на корреляционном поле. Однако нередки ситуации, когда расположение точек приблизительно соответствует нескольким функциям и необходимо из них

выявить наилучшую. Например, криволинейные зависимости могут аппроксимироваться полиномиальной, показательной, степенной, логарифмической функциями. Еще более неоднозначна ситуация для множественной регрессии, так как наглядное представление статистических данных в этом случае невозможно.

В данной главе перечислены базовые модели, используемые в эконометрическом моделировании, а также практические задачи, вызывающие необходимость их использования. Правильный выбор вида модели является отправной точкой для качественного анализа экономической модели. Безусловно, на практике неизвестно, какая модель является верной, и зачастую подбирают такую модель, которая наиболее точно соответствует реальным данным. При этом необходимо учитывать, что идеальной модели не существует. Поэтому, чтобы выбрать качественную модель, необходимо ответить на ряд вопросов, возникающих при ее анализе.

1. Каковы признаки «хорошей» (качественной) модели?
2. Какие ошибки спецификации встречаются, и каковы последствия данных ошибок?
3. Как обнаружить ошибку спецификации?
4. Каким образом можно исправить ошибку спецификации и перейти к лучшей (качественной) модели?

Признаки «хорошей» модели

В ряде случаев достаточно очевидно, какая модель лучше. В других случаях для принятия обоснованного решения приходится проводить достаточно кропотливый сравнительный анализ. Для этого необходимо выбрать критерии, которые позволят сделать обоснованный вывод. Обычно для построения «хорошей» работоспособной модели и сравнения ее с другими

возможными моделями необходимо учитывать следующие свойства (критерии).

Скупость (простота). Модель должна быть максимально простой. Данное свойство определяется тем фактом, что модель не отражает действительность идеально, а является ее упрощением. Поэтому из двух моделей, приблизительно одинаково отражающих реальность, предпочтение отдается модели, содержащей меньшее число объясняющих переменных.

Единственность. Для любого набора статистических данных определяемые коэффициенты должны вычисляться однозначно.

Максимальное соответствие. Уравнение тем лучше, чем большую часть разброса зависимой переменной оно может объяснить. Поэтому стремятся построить уравнение с максимально возможным скорректированным коэффициентом детерминации R.

Согласованность с теорией. Никакое уравнение не может быть признано качественным, если оно не соответствует известным теоретическим предпосылкам. Например, если в функции спроса коэффициент при цене положителен, то даже значительная величина коэффициента детерминации R (например, 0,7) не позволит признать уравнение удовлетворительным. Другими словами, модель обязательно должна опираться на теоретический фундамент, т.к. в противном случае результат использования регрессионного уравнения может быть весьма плачевным.

Прогнозные качества. Модель может быть признана качественной, если полученные на ее основе прогнозы подтверждаются реальностью.

Другим критерием прогнозных качеств оцененной модели регрессии может служить следующее отношение:

$$V = \frac{S}{\bar{y}}$$

где $S = \sqrt{\frac{\sum e_i^2}{n-m-1}}$ – стандартная ошибка регрессии, $\bar{y}$ – среднее значение зависимой переменной уравнения регрессии. Если величина V мала (а она определяет относительную ошибку прогноза в процентах) и отсутствует автокорреляция остатков (определяемая по величине статистики DW Дарбина–Уотсона), то прогнозные качества модели высоки.

Если уравнение регрессии используется для прогнозирования, то величина V обычно рассчитывается не для того периода, на котором оценивалось уравнение, а для некоторого следующего за ним временного интервала, для которого известны значения зависимой и объясняющих переменных. Тем самым на практике проверяются прогнозные качества модели. В случае положительного решения, если можно спрогнозировать значения объясняющих переменных на некоторый последующий период, построенная модель обоснованно может быть использована для прогноза значений объясняемой переменной Y. При этом следует помнить, что период прогнозирования должен быть, по крайней мере, в 3 раза короче периода, по которому оценивалось уравнение регрессии.

Поскольку не существует какого-либо единого правила построения регрессионных моделей, анализ перечисленных свойств позволяет строить более качественные эконометрические модели.

Виды ошибок спецификации

Одним из базовых предположений построения качественной модели является правильная (хорошая) спецификация уравнения регрессии. Правильная спецификация уравнения регрессии означает, что оно в целом правильно отражает соотношение между экономическими показателями,

участвующими в модели. Это является необходимой предпосылкой дальнейшего качественного оценивания.

Неправильный выбор функциональной формы или набора объясняющих переменных называется *ошибками спецификации*. Рассмотрим основные типы ошибок спецификации.

1. Отбрасывание значимой переменной

Последствия данной ошибки достаточно серьезны. Оценки, полученные с помощью МНК, являются смещенными и несостоятельными даже при бесконечно большом числе испытаний. Следовательно, возможные интервальные оценки и результаты проверки соответствующих гипотез будут ненадежными.

2. Добавление незначимой переменной

В некоторых случаях в уравнения регрессии включают слишком много объясняющих переменных, причем не всегда обоснованно. Последствия данной ошибки будут не столь серьезными, как в предыдущем случае. Оценки параметров регрессии остаются, как правило, несмещенными и состоятельными. Однако их точность уменьшится, увеличивая при этом стандартные ошибки, т. е. оценки становятся неэффективными, что отразится на их устойчивости.

3. Выбор неправильной функциональной формы

Последствия данной ошибки будут весьма серьезными. Обычно такая ошибка приводит либо к получению смещенных оценок, либо к ухудшению статистических свойств оценок коэффициентов регрессии и других показателей качества уравнения. В первую очередь это вызвано нарушением условий

Гаусса–Маркова для отклонений. Прогнозные качества модели в этом случае очень низки.

Обнаружение и корректировка ошибок спецификации

При построении уравнений регрессии, особенно на начальных этапах, ошибки спецификации весьма нередки. Они допускаются обычно из-за поверхностных знаний об исследуемых экономических процессах, либо из-за недостаточно глубоко проработанной теории, или из-за погрешностей при сборе и обработке статистических данных при построении эмпирического уравнения регрессии. Важно уметь обнаружить и исправить эти ошибки. Сложность процедуры определяется типом ошибки и нашими знаниями об исследуемом объекте.

Если в уравнении регрессии имеется одна несущественная переменная, то она обнаружит себя по низкой t-статистике. В дальнейшем эту переменную исключают из рассмотрения.

Если в уравнении несколько статистически незначимых объясняющих переменных, то следует построить другое уравнение регрессии без этих незначимых переменных. Затем с помощь F-статистики сравниваются коэффициенты детерминации для первоначального и дополнительного уравнений регрессий:

$$F = \frac{r_1^2 - r_2^2}{1 - r_1^2} * \frac{n - m - 1}{k}$$

Здесь n – число наблюдений, m – число объясняющих переменных в первоначальном уравнении, k – число отбрасываемых из первоначального уравнения объясняющих переменных.

При наличии нескольких несущественных переменных, возможно, имеет место мультиколлинеарность.

Однако осуществление указанных проверок имеет смысл лишь при правильном подборе вида (функциональной формы) уравнения регрессии, что можно осуществить, если согласовывать его с теорией.

Отметим, что выбор модели далеко не всегда осуществляется однозначно, и в дальнейшем требуется сравнивать модель как с теоретическими, так и с эмпирическими данными, совершенствовать ее.

Напомним, что при определении качества модели обычно анализируются следующие параметры:

а) скорректированный коэффициент детерминации r^2;

б) t-статистики;

в) статистика Дарбина–Уотсона DW;

г) согласованность знаков коэффициентов с теорией;

д) прогнозные качества (ошибки) модели.

Если все эти показатели удовлетворительны, то данная модель может быть предложена для описания исследуемого реального процесса. Если же какая-либо из описанных выше характеристик не является удовлетворительной, то есть основания сомневаться в качестве данной модели (неправильно выбрана функциональная форма уравнения; не учтена важная объясняющая переменная; имеется объясняющая переменная, не оказывающая значимого влияния на зависимую переменную).

Проблемы спецификации

Итак, стандартная схема анализа зависимостей состоит в осуществлении ряда последовательных процедур.

- Подбор начальной модели. Он осуществляется на основе экономической теории, предыдущих знаний об объекте исследования, опыта исследователя и его интуиции.
- Оценка параметров модели на основе имеющихся статистических данных.
- Осуществление тестов проверки качества модели (обычно используются t-статистики для коэффициентов регрессии, F-статистика для коэффициента детерминации, статистика Дарбина–Уотсона для анализа отклонений и ряд других тестов).
- При наличии хотя бы одного неудовлетворительного ответа по какому-либо тесту модель совершенствуется с целью устранения выявленного недостатка.
- При положительных ответах по всем проведенным тестам модель считается качественной. Она используется для анализа и прогноза объясняемой переменной.

Однако необходимо предостеречь от абсолютизации полученного результата. Проблема заключается в том, что даже качественная модель является подгонкой спецификации модели под имеющийся набор данных. Поэтому вполне реальна картина, когда исследователи, обладающие разными наборами данных, строят разные модели для объяснения одной и той же переменной. Другой проблемой является использование модели для прогнозирования значений объясняемой переменной. Иногда хорошие с точки зрения диагностических тестов модели обладают весьма низкими прогнозными качествами.

Одним из главных направлений эконометрического анализа является постоянное совершенствование моделей. Здесь следует отметить, что какого-то глобального подхода, определяющего заранее возможные пути совершенствования, нет и, скорее всего, быть не может. Исследователь должен помнить, что совершенной модели не существует. В силу постоянно изменяющихся условий протекания экономических процессов не может быть и постоянно качественных моделей. Новые условия требуют пересмотра даже весьма устойчивых моделей.

До сих пор достаточно спорным является вопрос, как строить модели:

а) начинать с самой простой и постоянно усложнять ее;

б) начинать с максимально сложной модели и упрощать ее на основе проводимых исследований.

И тот и другой подход имеют как достоинства, так и недостатки. Например, если следовать схеме а), то происходит обыкновенная подгонка модели под эмпирические данные. При теоретически более оправданном подходе б) поиск возможных направлений совершенствования модели зачастую сводится к полному перебору, что делает проводимый анализ неэффективным. Возможно также на этапах упрощения модели отбрасывание объясняющих переменных, которые были бы весьма полезны в упрощенной модели. В итоге, построение модели является индивидуальным подходом к конкретной ситуации, опирающимся на серьезные знания экономической теории и статистического анализа.

При этом отметим, что при всех недостатках моделей принятие на их основе решений приводит в целом к гораздо более высоким или ожидаемым результатам, чем при принятии решений лишь на основе интуиции и экономической теории.

Тема 4.4 Множественная регрессия и корреляция

Множественная регрессия

На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. Например, спрос на некоторое благо определяется не только его ценой и качеством, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими факторами.

Множественная регрессия – это уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, \dots, x_m) + \varepsilon$$

где y – зависимая переменная (результативный признак); x_1, x_2 , ..., x_m – независимые переменные (признаки-факторы).

Для построения уравнения множественной регрессии чаще используются следующие функции:

- линейная – $y = a + b_1 * x_1 + b_2 x_2 + \dots + b_m x_m + \varepsilon$
- степенная – $y = a * {x_1}^{b_1} * {x_2}^{b_2} * \dots * {x_m}^{b_m} * \varepsilon$
- экспонента – $y = e^{a + b_1 * x_1 + b_2 * x_2 + \dots + b_m * x_m + \varepsilon}$
- гипербола – $y = \frac{1}{a + b_1 * x_1 + b_2 * x_2 + \dots + b_m * x_m + \varepsilon}$

Можно использовать и другие функции, приводимые к линейному виду.

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов* (МНК). Для линейных уравнений

$$y = a + b_1 * x_1 + b_2 x_2 + \dots + b_m x_m + \varepsilon$$

строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\begin{cases} \sum y = na + b_1\sum x_1 + b_2\sum x_2 + \cdots + b_m\sum x_m \\ \sum yx_1 = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 + \cdots + b_m\sum x_1x_m \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \sum yx_m = a\sum x_m + b_1\sum x_1x_m + b_2\sum x_2x_m + \cdots + b_m\sum x_m^2 \end{cases}$$

Для ее решения может быть применен метод определителей:

$$a = \frac{\Delta a}{\Delta}, b_1 = \frac{\Delta b_1}{\Delta}, \dots, b_m = \frac{\Delta b_m}{\Delta},$$

где $\Delta = \begin{pmatrix} n & \sum x_1 & \sum x_2 & \dots & \sum \mathrm{m} \\ \sum x_1 & \sum x_1^2 & \sum x_2x_1 & \dots & \sum x_mx_1 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & \dots & \sum x_mx_2 \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_p & \sum x_1x_p & \sum x_2x_p & \dots & \sum x_m^2 \end{pmatrix}$ – определитель системы.

Для двухфакторной модели данная система будет иметь вид:

$$\begin{cases} \sum y = na + b_1\sum x_1 + b_2\sum x_2 \\ \sum yx_1 = a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 \\ \sum yx_2 = a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 \end{cases}$$

Также можно воспользоваться готовыми формулами, которые являются следствием из этой системы:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} * \frac{r_{yx_1} - r_{yx_2} * r_{x_1x_2}}{1 - r_{x_1x_2}^2}$$

$$b_2 = \frac{\sigma_y}{\sigma_{x_2}} * \frac{r_{yx_2} - r_{yx_1} * r_{x_1x_2}}{1 - r_{x_1x_2}^2}$$

где r_{yxi} и r_{xixj} – коэффициенты парной и межфакторной корреляции.

$$a = \bar{y} - b_1\overline{x_1} - b_2\overline{x_2}$$

В линейной множественной регрессии параметры при *x* называются *коэффициентами «чистой» регрессии*. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне.

Метод наименьших квадратов применим и к уравнению множественной регрессии в *стандартизированном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2 + \dots +} \beta_m t_{x_m} + \varepsilon$$

где $t_y, t_{x_1}, t_{x_2 \dots} t_{x_m}$ – *стандартизированные переменные*:

$$t_y = \frac{y - \bar{y}}{\sigma_y}$$

$$t_{x_i} = \frac{x_i - \overline{x_\iota}}{\sigma_{x_i}}$$

для которых среднее значение равно нулю: $\bar{t}_y = \overline{t_{x_\iota}} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$, β_i – *стандартизированные коэффициенты регрессии*.

В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Применяя МНК к уравнению множественной регрессии в стандартизированном масштабе, получим систему нормальных уравнений вида:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} \dots + \beta_m r_{x_1x_m} \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 + \beta_3 r_{x_1x_3} \dots + \beta_m r_{x_1x_m} \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_1x_m} + \beta_3 r_{x_3x_m} \dots + \beta_m \end{cases}$$

где r_{yx_i} и $r_{x_ix_j}$ – коэффициенты парной и межфакторной корреляции.

Коэффициенты «чистой» регрессии bi связаны со стандартизированными коэффициентами регрессии βi следующим образом:

$$b_i = \beta_i * \frac{\sigma_y}{\sigma_{x_i}}$$

Поэтому можно переходить от уравнения регрессии в стандартизованном масштабе к уравнению регрессии в натуральном масштабе переменных, при этом параметр a определяется как

$$a = \bar{y} - b_1\overline{x_1} - b_2\overline{x_2} - \dots - b_m\overline{x_m}$$

Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсеве факторов – из модели исключаются факторы с наименьшим значением β_i .

Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле:

$$\overline{Э_{yx_j}} = b_j \frac{\overline{x_j}}{\bar{y}}$$

которые показывают на сколько процентов в среднем изменится результат, при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат.

Тесноту совместного влияния факторов на результат оценивает *индекс множественной корреляции:*

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\sigma^2_{y\,\text{ост}}}{\sigma^2_y}}$$

Значение индекса множественной корреляции лежит в пределах от 0 до 1 и должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1x_2\ldots x_m} \geq r_{yx_i} \qquad (\text{i} = \overline{1, m})$$

При линейной зависимости *коэффициент множественной корреляции* можно определить через матрицы парных коэффициентов корреляции:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}$$

Δr – определитель матрицы парных коэффициентов корреляции y, x_1, x_2,…x_k.

Δr_{11} – определитель матрицы коэффициентов межфакторной корреляции x_1, x_2,…x_k.

Для модели, в которой присутствует две независимые переменные, формула упрощается:

$$R_{yx_1x_2} = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2 * r_{yx_1} * r_{yx_2} * r_{x1x2}}{1 - r^2_{x1x2}}}$$

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации. *Коэффициент множественной детерминации* рассчитывается как квадрат индекса множественной корреляции

Для того чтобы не допустить преувеличения тесноты связи, применяется *скорректированный индекс множественной детерминации*, который содержит поправку на число степеней свободы и рассчитывается по формуле

$$\widehat{R^2} = 1 - (1 - R^2) * \frac{n-1}{n-m-1}$$

где n – число наблюдений, m – число факторов. При небольшом числе наблюдений нескорректированная величина коэффициента множественной детерминации $\widehat{R^2}$ имеет тенденцию переоценивать долю вариации результативного признака, связанную с влиянием факторов, включенных в регрессионную модель.

Частные коэффициенты (или индексы) корреляции, измеряющие влияние на y фактора x_i , при элиминировании (исключении влияния) других факторов, при двухфакторной модели можно определить по формулам:

- при множественной корреляции от двух факторов коэффициент частной корреляции первого фактора:

$$r_{y/x1x2} = \frac{r_{yx1} - r_{yx2} * r_{x1x2}}{\sqrt{(1 - r_{yx2}^2) * (1 - r_{x1x2}^2)}}$$

а коэффициент частной корреляции для второго фактора:

$$r_{y/x2x1} = \frac{r_{yx2} - r_{yx1} * r_{x1x2}}{\sqrt{(1 - r_{yx2}^2) * (1 - r_{x1x2}^2)}}$$

Сравнение частных коэффициентов корреляции друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде.

Значимость уравнения множественной регрессии в целом оценивается с помощью F -критерия Фишера:

$$F = \frac{R^2}{1 - R^2} * \frac{n - m - 1}{m}$$

Фактическое значение частного F –критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: $k_1=1$ и $k_2=n-m-1$. Если фактическое значение F превышает $F_{табл}(\alpha, k1, k2)$, то уравнение признается статистически значимым

С помощью частных F-критериев Фишера можно оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 (в случае множественной регрессии от 2-ух факторов) при помощи формул:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1 - R^2_{yx_1x_2}} * \frac{n - m - 1}{1}$$

$$F_{x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1 - R^2_{yx_1x_2}} * \frac{n - m - 1}{1}$$

Если $F_{x1}>F_{табл.}$, то значение частного F-критерия для дополнительно включенного фактора x_1 не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора x_1 является существенным. Фактор x_1 должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора x_2.

Если $F_{x2}<F_{табл.}$, то включение в модель фактора x_2 после того, как в модель включен фактор x_1 статистически нецелесообразно.

Оценка значимости коэффициентов чистой регрессии как и в парной регрессии проводится по t -критерию Стьюдента.

При построении уравнения множественной регрессии может возникнуть проблема *мультиколлинеарности* факторов, их тесной линейной связи. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов.

Считается, что две переменные *коллинеарны*, т. е. находятся между собой в линейной зависимости, если коэффициент корреляции больше или равен 0,7.

Например, $r_{yx_1} = 0{,}97;\ r_{yx2} = 0{,}941;\ r_{x1x2} = 0{,}943$. Эти коэффициенты указывают на весьма сильную связь каждого фактора с результатом, также на высокую межфакторную зависимость (факторы x_1 и x_2 явно коллинеарны, т. к. $r_{x1x2} = 0{,}943 > 0{,}7$). При такой сильной межфакторной зависимости рекомендуется один из факторов исключить из рассмотрения.

Для оценки мультиколлинеарности факторов используется определитель матрицы парных коэффициентов корреляции между факторами.

Если между факторами существует полная линейная зависимость и все коэффициенты корреляции равны 1, то определитель такой матрицы равен 0:

$$Det(R) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 0$$

Если факторы не коррелированны между собой, то матрица коэффициентов корреляции имеет определитель, равный 1.

Оценки параметров регрессии должны отвечать определенным критериям: быть несмещенными, состоятельными и эффективными.

Несмещенность оценки означает, что математическое ожидание остатков равно нулю. Если оценки обладают свойством несмещенности, то их можно сравнивать по разным исследованиям.

Оценки считаются эффективными, если они характеризуются наименьшей дисперсией. В практических исследованиях это означает возможность перехода от точечного оценивания к интервальному.

Состоятельность оценок характеризует увеличение их точности с увеличением объема выборки. Большой практический интерес представляют те результаты регрессии, для которых доверительный интервал ожидаемого значения параметра регрессии b_i имеет предел значений вероятности, равный единице. Иными словами, вероятность получения оценки на заданном расстоянии от истинного значения параметра близка к единице.

Указанные критерии оценок (несмещенность, состоятельность и эффективность) обязательно учитываются при разных способах оценивания. Метод наименьших квадратов строит оценки регрессии на основе минимизации суммы квадратов остатков. Поэтому очень важно исследовать поведение остаточных величин регрессии ε_i. Условия, необходимые для получения несмещенных, состоятельных и эффективных оценок, представляют собой предпосылки МНК, соблюдение которых желательно для получения достоверных результатов регрессии.

Исследования остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

1) случайный характер остатков;

2) нулевая средняя величина остатков, не зависящая от x_i ;

3) гомоскедастичность – дисперсия каждого отклонения ε_i , одинакова для всех значений x;

4) отсутствие автокорреляции остатков – значения остатков ε_i распределены независимо друг от друга;

5) остатки подчиняются нормальному распределению.

Если распределение случайных остатков ε_i не соответствует некоторым предпосылкам МНК, то следует корректировать модель.

Уравнения множественной регрессии могут включать в качестве независимых переменных качественные признаки (например, профессию, пол, образование, климатические условия, отдельные регионы и т. д.). Чтобы ввести такие переменные в регрессионную модель, их необходимо упорядочить и присвоить им те или иные значения, т. е. качественные переменные преобразовать в количественные. Такие переменные в эконометрике принято называть ***фиктивными переменными***. Например: 1 – мужской пол, 0 – женский.

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории (женский пол) к другой (мужской пол) при неизменных значениях остальных параметров. На основе t-критерия Стьюдента делается вывод о значимости влияния фиктивной переменной, существенности расхождения между категориями.

Фиктивные переменные моделей с временными рядами.

Регрессионные модели с временными рядами используют 3 типа фиктивных переменных:

- Переменные индикаторы, которые определяют принадлежность наблюдения к соответствующему периоду;
- Переменные сезонного характера;
- Линейный временной тренд.

Сезонные переменные применяются при моделировании сезонности, они принимают различные значения в соответствии с тем, к какому месяцу или кварталу года (дню недели) относится наблюдение. Линейный временной тренд используется в процессе моделирования постепенных (плавных) структурных сдвигов. Данная переменная отражает промежуток времени, который проходит от определенного (нулевого) момента времени до того времени, к которому отнесли это наблюдение (то есть координаты наблюдения по временной шкале). В случае, когда промежутки времени между последовательными наблюдениями будут равны, то временной тренд составляется из номеров наблюдений. Указанные виды фиктивных переменных можно комбинировать, создавая при этом переменные “взаимодействия” определенных эффектов.

Преимущества использования фиктивных переменных.

Использование фиктивных переменных обладает следующими преимуществами:

1. Не обязательно одинаковые интервалы между наблюдениями, а также наличие пропущенных интервалов между наблюдениями;
2. Легкая интерпретация коэффициентов при фиктивных переменных, легко и наглядно представляющих структуру динамического процесса;
3. В процессе оценки модели не обязательно выходить за рамки классического метода наименьших квадратов.

Тема 4.5 Гомо- и гетероскедастичность остатков

При проведении регрессионного анализа, основанного на методе наименьших квадратов, на практике следует обратить серьезное внимание на проблемы, связанные с выполнимостью свойств случайных отклонений моделей. Как мы отмечали ранее, свойства оценок коэффициентов регрессии напрямую зависят от свойств случайного члена в уравнении регрессии. Для получения качественных оценок необходимо следить за выполнимостью предпосылок МНК (условий Гаусса-Маркова), т. к. при их нарушении МНК может давать оценки с плохими статистическими свойствами. При этом существуют другие методы определения более точных оценок. Одной из ключевых предпосылок МНК является условие постоянства дисперсий случайных отклонений (предпосылка 2):

дисперсия случайных отклонений ε_i *постоянна.* $D(\varepsilon_i)=D(\varepsilon_j) = \sigma^2$ *для любых наблюдений i и j.*

Выполнимость данной предпосылки называется *гомоскедастичностью (постоянством дисперсии отклонений)*. Невыполнимость данной предпосылки называется *гетероскедастичностью (непостоянством дисперсий отклонений)*.

Суть гетероскедастичности

При рассмотрении выборочных данных требование постоянства дисперсии случайных отклонений может вызвать определенное недоумение в силу того, что при каждом i-м наблюдении имеется единственное значение ε_i. Откуда же появляется разброс? Дело в том, что при рассмотрении выборочных данных мы имеем дело с конкретными реализациями зависимой переменной y_i и соответственно с определенными случайными отклонениями ε_i, $i = 1, 2, ..., n$. Но до осуществления выборки эти показатели априори могли принимать произвольные значения на основе некоторых вероятностных распределений.

Одним из требований к этим распределениям является равенство дисперсий. Данное условие подразумевает, что несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть большим либо маленьким, положительным либо отрицательным, не должно быть некой априорной причины, вызывающей большую ошибку (отклонение) при одних наблюдениях и меньшую – при других.

Однако на практике гетероскедастичность не так уж и редка. Зачастую есть основания считать, что вероятностные распределения случайных отклонений ε_i при различных наблюдениях будут различными. Это не означает, что случа
йные отклонения обязательно будут большими при определенных наблюдениях и малыми – при других, но это означает, что априорная вероятность этого велика. Поэтому важно понимать суть этого явления и его последствия.

На рис. 49 приведены два примера линейной регрессии – зависимости потребления C от дохода I: $C = \beta_0 + \beta_1 * I + \varepsilon$.

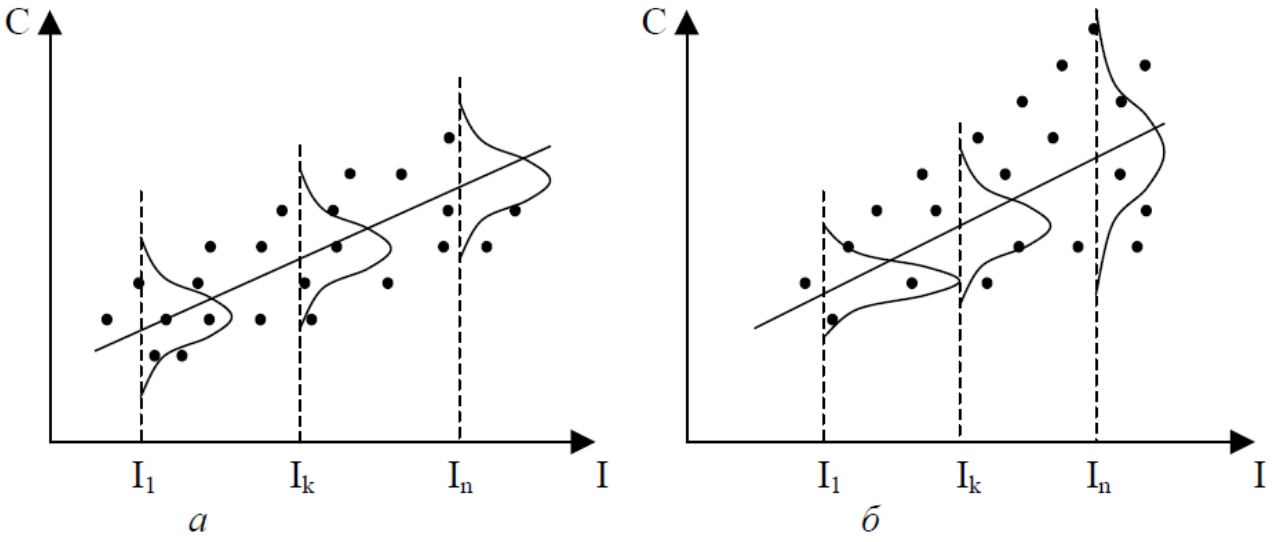


Рис.31 Два примера линейной регрессии зависимости потребления от дохода

В обоих случаях с ростом дохода растет среднее значение потребления. Но если на рис. 49, *а* дисперсия потребления остается одной и той же для различных уровней дохода, то на рис. 49, *б* при аналогичной зависимости среднего потребления от дохода дисперсия потребления не остается постоянной, а увеличивается с ростом дохода. Фактически это означает, что во втором случае субъекты с большим доходом в среднем потребляют больше, чем субъекты с меньшим доходом, и, кроме того, разброс в их потреблении более существенен для большего уровня дохода. Фактически люди с большими доходами имеют больший простор для распределения своего дохода. Реалистичность данной ситуации не вызывает сомнений. Разброс значений потребления вызывает разброс точек наблюдения относительно линии регрессии, что и определяет дисперсию случайных отклонений. Динамика изменения дисперсий (распределений) отклонений для данного примера проиллюстрирована на рис. 50. При гомоскедастичности (рис. 50, *а*) дисперсии ε_i постоянны, а при гетероскедастичности (50, *б*) дисперсии ε_i изменяются (в нашем примере – увеличиваются).

Проблема гетероскедастичности в большей степени характерна для перекрестных данных и довольно редко встречается при рассмотрении временных рядов. Это можно объяснить следующим образом. При перекрестных данных учитываются экономические субъекты (потребители, домохозяйства, фирмы, отрасли, страны и т. п.), имеющие различные доходы, размеры, потребности и т. д. Но в этом случае возможны проблемы, связанные с эффектом масштаба. Во временных рядах обычно рассматриваются одни и те же показатели в различные моменты времени (например, ВНП, чистый экспорт, темпы инфляциии и. т. д. в определенном регионе за определенный период времени). Однако при увеличении (уменьшении) рассматриваемых показателей с течением времени может возникнуть проблема гетероскедастичности.

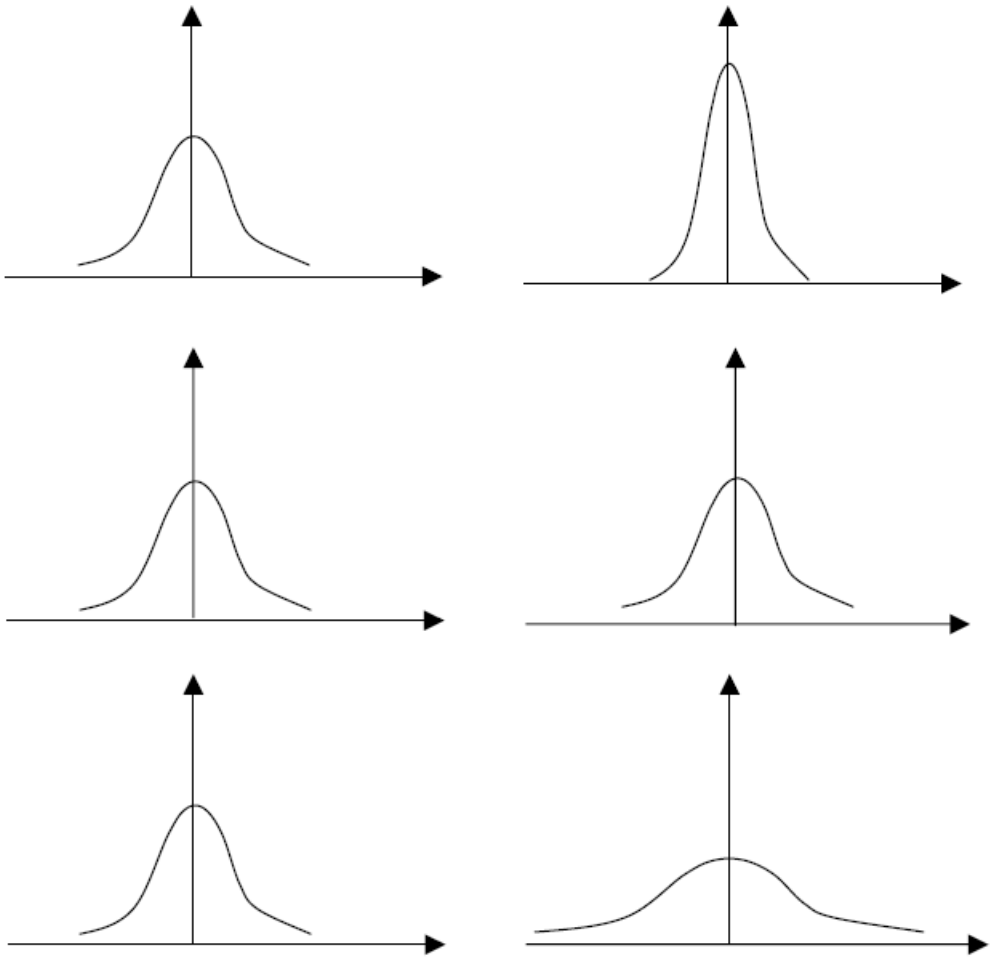


Рис.32 Динамика изменения дисперсий (распределений) отклонений при гомоскедастичности и при гетероскедастичности

Последствия гетероскедастичности

При рассмотрении классической линейной регрессионной модели МНК дает наилучшие линейные несмещенные оценки (BLUE-оценки) лишь при выполнении ряда предпосылок, одной из которых является постоянство дисперсии отклонений (гомоскедастичность): $\sigma^2(\varepsilon_i) = \sigma^2$ для всех наблюдений i, i = 1, 2, …, n.

При невыполнимости данной предпосылки (при гетероскедастичности) последствия применения МНК будут следующими.

1. Оценки коэффициентов по-прежнему остаются несмещенными и линейными.

2. Оценки не будут эффективными (т. е. они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра). Они не будут даже асимптотически эффективными. Увеличение дисперсии оценок снижает вероятность получения максимально точных оценок.

3. Дисперсии оценок будут рассчитываться со смещением. Смещенность появляется вследствие того, что необъясненная уравнением регрессии дисперсия $S^2 = \frac{\sum e_i^2}{n-m-1}$ (m – число объясняющих переменных), которая используется при вычислении оценок дисперсий всех коэффициентов, не является более несмещенной.

4. Вследствие вышесказанного все выводы, получаемые на основе соответствующих t- и F-статистик, а также интервальные оценки будут ненадежными. Следовательно, статистические выводы, получаемые при стандартных проверках качества оценок, могут быть ошибочными и приводить к неверным заключениям по построенной модели. Вполне вероятно, что стандартные ошибки коэффициентов будут занижены, а следовательно, t-статистики будут завышены. Это может привести к признанию статистически значимыми коэффициентов, таковыми на самом деле не являющимися.

Причину неэффективности оценок МНК при гетероскедастичности легко пояснить следующим примером парной регрессии

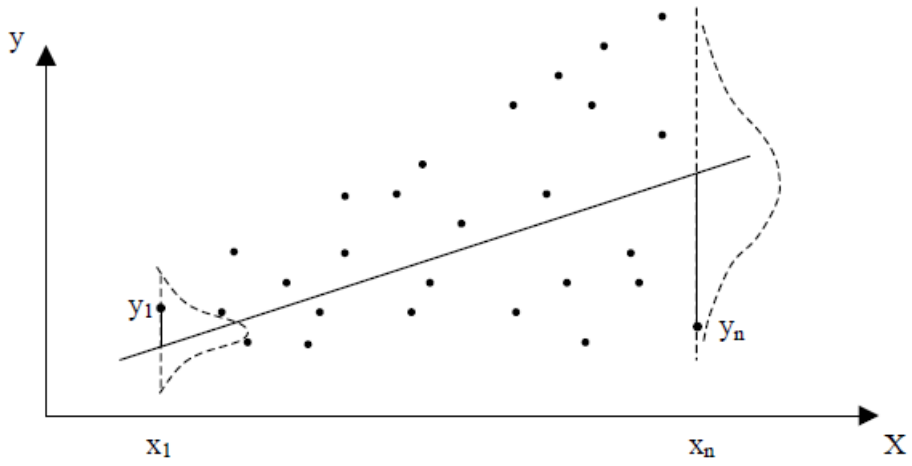


Рис.33 Гетероскедастичность остатков парной регрессии

Из рис. 51 видно, что для каждого конкретного значения x_i СВ X переменная Y принимает значение y_i из некоторого множества, имеющего свое распределение, отличное одно от другого в силу непостоянства дисперсий (сравните распределения для значений y_1 и y_n).

По МНК минимизируется сумма квадратов отклонений:

$$\sum e_i^2 = \sum (y_i - b_0 - b_1 x_1)^2$$

Но в этом случае каждое конкретное значение в данной сумме имеет одинаковый «вес» вне зависимости от того, получено оно из распределения с маленькой дисперсией (например, e_1^2) или с большой (например, e_n^2). Но это противоречит логике, т. к. точка, полученная из распределения с меньшей дисперсией, более точно определяет направление линии регрессии. Поэтому она должна иметь больший «вес», чем точка из распределения с большей дисперсией. Следовательно, методы оценивания, учитывающие «веса» точек наблюдений, позволяют получать более точные (эффективные) оценки. Учет «весов» точек характерен, например, для метода взвешенных наименьших квадратов.

Обнаружение гетероскедастичности

В ряде случаев на базе знаний характера данных появление проблемы гетероскедастичности можно предвидеть и попытаться устранить этот недостаток еще на этапе спецификации. Однако значительно чаще эту проблему приходится решать после построения уравнения регрессии. Обнаружение гетероскедастичности в каждом конкретном случае является довольно сложной задачей, т. к. для знания дисперсий отклонений σ^2 (e_i) необходимо знать распределение СВ Y, соответствующее выбранному значению x_i СВ X. На практике зачастую для каждого конкретного значения x_i определяется единственное значение y_i , что не позволяет оценить дисперсию СВ Y для данного x_i .

Естественно,не существует какого-либо однозначного метода определения гетероскедастичности. Однако к настоящему времени для такой проверки разработано довольно большое число тестов и критериев для них – графический анализ отклонений, тест ранговой корреляции Спирмена, тест Парка, тест Глейзера, тест Голдфелда-Квандта.

Рассмотрим наиболее популярные и наглядные: графический анализ отклонений, тест ранговой корреляции Спирмена.

Графический анализ остатков

Использование графического представления отклонений e_i позволяет определиться с наличием гетероскедастичности. В этом случае по оси абсцисс откладывается объясняющая переменная X (либо линейная комбинация объясняющих переменных $Y = b_0 + b_1X_1 + ... + b_mX_m$), а по оси ординат либо отклонения e_i, либо их квадраты e_i^2 .

Примеры таких графиков приведены на рис. 52.

На рис. 52, *а* все отклонения находятся внутри полуполосы постоянной ширины, параллельной оси абсцисс. Это говорит о независимости дисперсий e_i^2 от значений переменной X и их постоянстве, т.е. в этом случае мы находимся в условиях гомоскедастичности.

На рис. 52, *б* - *д* наблюдаются некие систематические изменения в соотношениях между значениями x_i переменной X и квадратами отклонений e_i^2. Рис. 52, *б* соответствует примеру из рис. 50. На рис. 52, *в* отражена линейная; 52, *г* - квадратичная; 52, *д* - гиперболическая зависимости между квадратами отклонений и значениями объясняющей переменной X. Другими словами, ситуации, представленные на рис. 52, *б-д*, отражают большую вероятность наличия гетероскедастичности для рассматриваемых статистических данных.

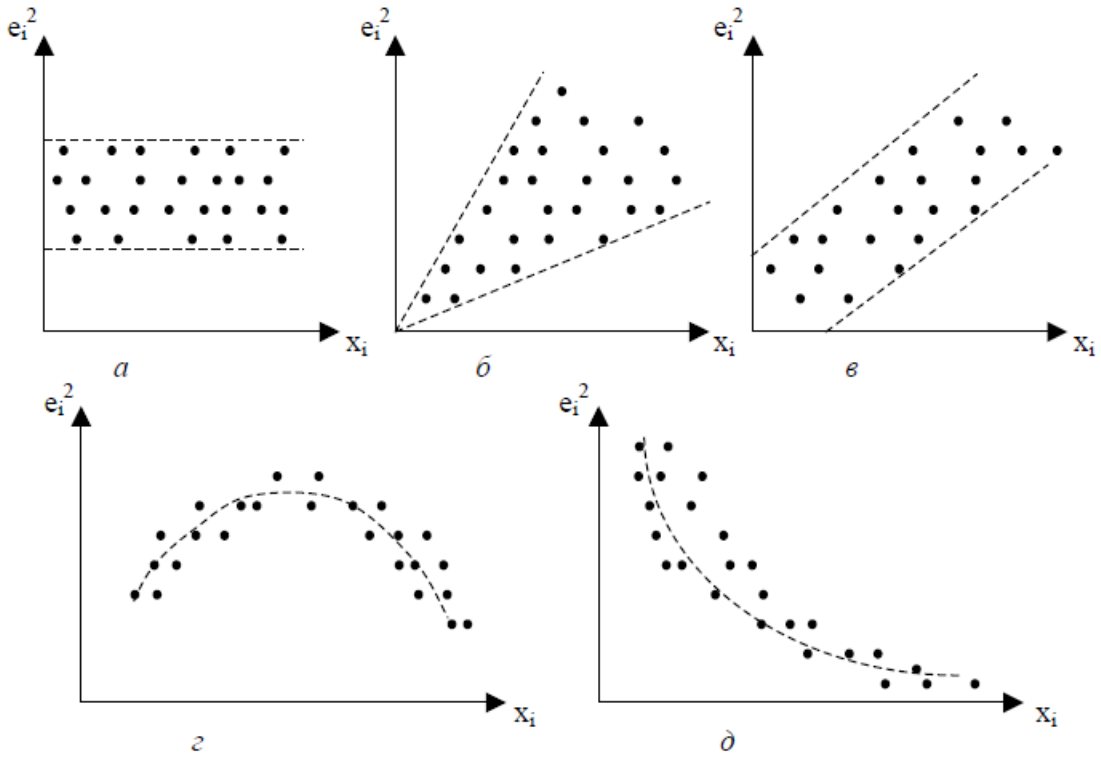


Рис.34 Графический анализ остатков

Отметим, что графический анализ отклонений является удобным и достаточно надежным в случае парной регрессии. При множественной

регрессии графический анализ возможен для каждой из объясняющих переменных X_j , j = 1, 2, …, m отдельно. Чаще же вместо объясняющих переменных X_j по оси абсцисс откладывают значения $\hat{y}_i$, получаемые из эмпирического уравнения регрессии. Поскольку по уравнению множественной линейной регресси $\hat{y}_i$ и является линейной комбинацией x_{ij}, j=1, 2, … , m, то график, отражающий зависимость e_i^2 от $\hat{y}_i$, может указать на наличие гетероскедастичности аналогично ситуациям на рис. 52, *б* - *д*. Такой анализ наиболее целесообразен при большом количестве объясняющих переменных.

Тест ранговой корреляции Спирмена

При использовании данного теста предполагается, что дисперсия отклонения будет либо увеличиваться, либо уменьшаться с увеличением значения X. Поэтому для регрессии, построенной по МНК, абсолютные величины отклонений e_i и значения x_i СВ X будут коррелированы. Значения x_i и e_i ранжируются (упорядочиваются по величинам). Затем определяется коэффициент ранговой корреляции:

$$r_{x,e} = 1 - 6 * \frac{\sum d_i^2}{n(n^2 - 1)}$$

где d_i – разность между рангами x_i и e_i , i = 1, 2, … , n; n – число наблюдений.

Например, если x_{20} является 25-м по величине среди всех наблюдений X; а e_{20} является 32-м, то d_i = 25 – 32= - 7.

Доказано, что если коэффициент корреляции $\rho_{x,e}$ для генеральной совокупности равен нулю, то статистика

$$t = \frac{r_{x,e}\sqrt{n-2}}{\sqrt{1-r_{x,e}^2}}$$

имеет распределение Стьюдента с числом степеней свободы υ = n- 2.

Следовательно, если наблюдаемое значение t-статистики, вычисленное по вышеуказанной формуле, превышает $t_{кр.} = t_{\alpha,n-2}$ (определяемое по таблице критических точек распределения Стьюдента), то необходимо отклонить гипотезу о равенстве нулю коэффициента корреляции $\rho_{x,e}$, а следовательно, и об отсутствии гетероскедастичности. В противном случае гипотеза об отсутствии гетероскедастичности принимается.

Если в модели регрессии больше чем одна объясняющая переменная, то проверка гипотезы может осуществляться с помощью t- статистики для каждой из них отдельно.

Статистика Дарбина-Уотсона (автокорреляция в остатках)

Рассмотрим уравнение регрессии вида:

$$y_t = a + \sum_{j=1}^{k} b_j x_{jt} + \varepsilon_t$$

где k - число независимых переменных модели.

Для каждого момента (периода) времени $t = 1: n$ значение компоненты ε_t, определяется как

$$\varepsilon_t = y_t - \widehat{y_t}$$

или

$$\varepsilon_t = y_t - (a + \sum_{j=1}^{k} b_j x_{jt})$$

Рассматривая последовательность остатков как временной ряд, можно построить график их зависимости от времени (рис. 53).

В соответствии с предпосылками МНК остатки ε_t , должны быть случайными (рис. а). Однако при моделировании временных рядов нередко встречается ситуация, когда остатки содержат тенденцию (б и в) или циклические колебания (г). Это свидетельствует о том, что каждое следующее значение остатков зависит от предшествующих. В этом случае говорят о наличии автокорреляции остатков.

Автокорреляция остатков может быть вызвана несколькими причинами, имеющими различную природу. *Во-первых*, иногда она связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака. *Во-вторых*, в ряде случаев причину автокорреляции остатков следует

искать в формулировке модели. Модель может не включать фактор, оказывающий существенное влияние на результат, влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t. Кроме того, в качестве таких существенных факторов могут выступать лаговые значения переменных, включенных в модель. Либо модель не учитывает несколько второстепенных факторов, совместное влияние которых на результат значительно ввиду совпадения тенденций их изменения или фаз циклических колебаний.

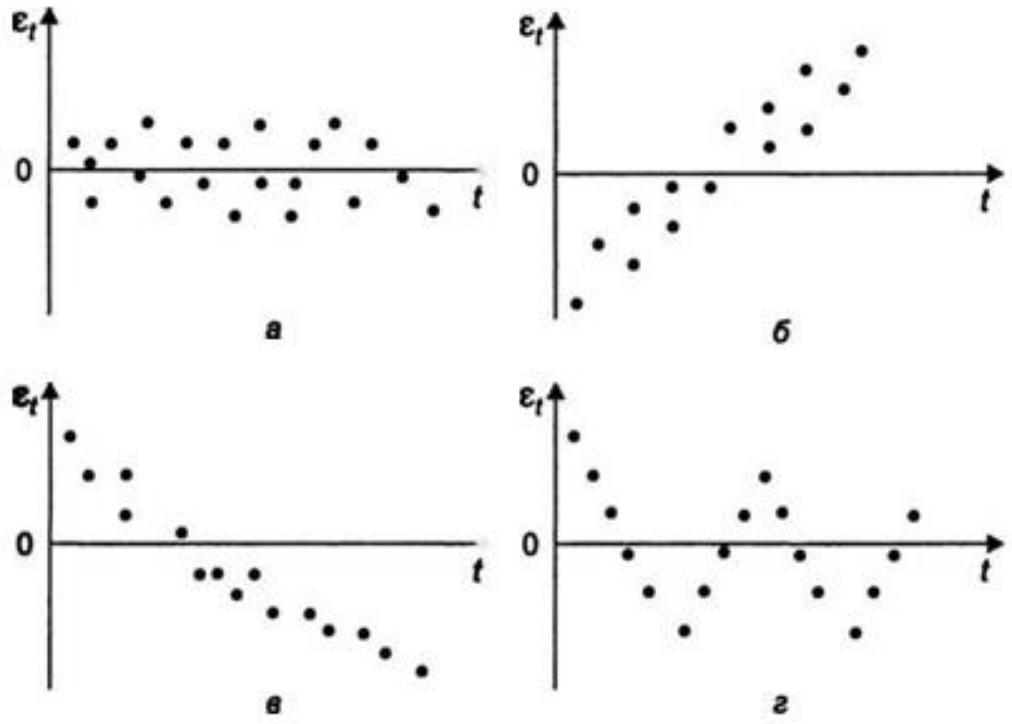


Рис.35 Модели зависимости остатков от времени:

а- случайные остатки;

б - возрастающая тенденция в остатках;

в - убывающая тенденция в остатках;

г - циклические колебания в остатках.

От истинной автокорреляции остатков следует отличать ситуации, когда

причина автокорреляции заключается в неправильной спецификации функциональной формы модели. В этом случае следует изменить форму связи факторных и результативного признаков, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции остатков.

Известны два наиболее распространенных метода определения автокорреляции остатков. Первый метод – это построение графика зависимости остатков от времени и визуальное определение наличия или отсутствия автокорреляции. Второй метод - использование *критерия Дарбина - Уотсона* и расчет величины

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}$$

Таким образом, d – это отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии. Практически во всех статистических пакетах прикладных программ значение критерия Дарбина - Уотсона указывается наряду с коэффициентом детерминации, значениями t- и F-критериев.

Коэффициент автокорреляции остатков первого порядка определяется как

$$r_1^{\varepsilon} = \frac{\sum_{t=2}^{n}(\varepsilon_t - \bar{\varepsilon_1})*(\varepsilon_{t-1} - \bar{\varepsilon_2})}{\sqrt{\sum_{t=2}^{n}(\varepsilon_t - \bar{\varepsilon_1})^2 * \sum_{t=2}^{n}(\varepsilon_{t-1} - \bar{\varepsilon_1})^2}}$$

где

$$\bar{\varepsilon_1} = \frac{\sum_{t=2}^{n}\varepsilon_t}{n-1}$$

$$\overline{\varepsilon_2} = \frac{\sum_{t=2}^{n} \varepsilon_{t-1}}{n-1}$$

Соотношение между критерием Дарбина-Уотсона и коэффициентом автокорреляции остатков первого порядка определяется зависимостью:

$$d = 2(1 - r_1^{\varepsilon})$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1^{\varepsilon} = 1$, то $d = 0$. Если в остатках есть полная отрицательная автокорреляция, то $r_1^{\varepsilon} = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1^{\varepsilon} = 0$ и $d = 2$. Значит,

$$0 \leq d \leq 4$$

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят соответственно в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам определяются критические значения критерия Дарбина-Уотсона d_L и d_U для заданного числа наблюдений n, числа независимых переменных модели κ и уровня значимости α. По этим значениям числовой промежуток [0;4] разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $(1 - \alpha)$ представлено на рисунке:

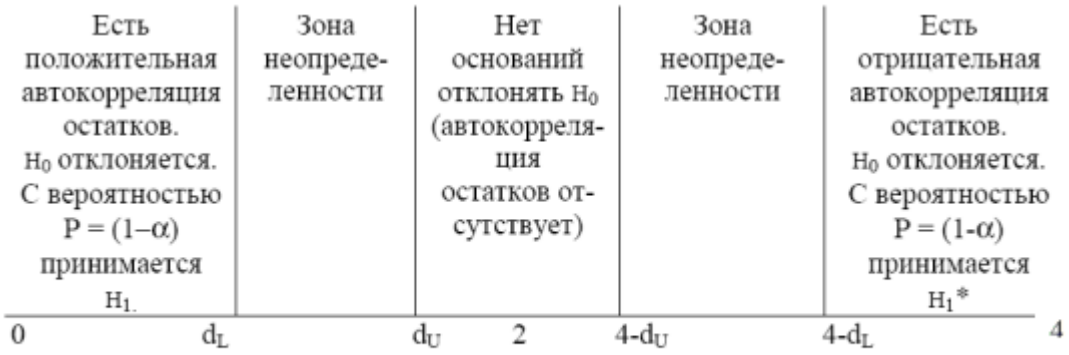


Рис.36 Принятие или отклонение каждой из гипотез (критерий Дарбина-Уотсона)

Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0.

Глава 5 Использование различных функций Microsoft Excel в эконометрике

Тема 5.1 Расчет характеристик модели с помощью Microsoft Excel.

В MS Excel есть несколько инструментов, которыми можно воспользоваться для построения и последующего анализа эконометрических моделей. Рассмотрим надстройку «Анализ данных» и надстройку «Поиск решения».

На основании данных (табл. 8) о десяти случаях ущерба, нанесенного пожаром в жилых помещениях, требуется:

1) построить модель парной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции:

а) с использованием Анализа данных;

б) с использованием Поиска решений;

с) с использованием матричных функций;

в) с использованием функции ЛИНЕЙН.

Дать экономическую интерпретацию параметров модели регрессии.

2) оценить качество построенной модели;

3) оценить стоимость ущерба, нанесенного пожаром, если полагают, что расстояние до ближайшей пожарной станции уменьшится на 10 % от своего среднего уровня;

4) изобразить на графике исходные данные, результаты моделирования и прогнозирования.

Таблица 8

Ущерб, нанесенный пожаром в жилых помещениях

№ п/п	1	2	3	4	5	6	7	8	9	10
Расстояние до ближайшей пожарной станции, км (X)	3,4	1,8	4,6	2,3	3,1	5,5	0,7	3	2,6	4,3
Общая сумма ущерба. тыс. руб. (Y)	262	178	313	231	275	360	141	223	196	313

Решение 1) а) Построение модели линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции с использованием **инструмента Регрессия в пакете Анализ данных Excel.**

Данные записали в таблицу Excel и упорядочили по возрастанию X (процедура необязательная).

а) Выбрали команду на вкладке Данные → команда Анализ данных.

Если в меню «Данные» еще нет команды **Анализ данных**, то необходимо сделать следующее.

Вкладка Файл→Параметры→Надстройки→кнопка «перейти» →в окне «надстройки» установить флажок «Пакет анализа» →ОК (рис. 37).

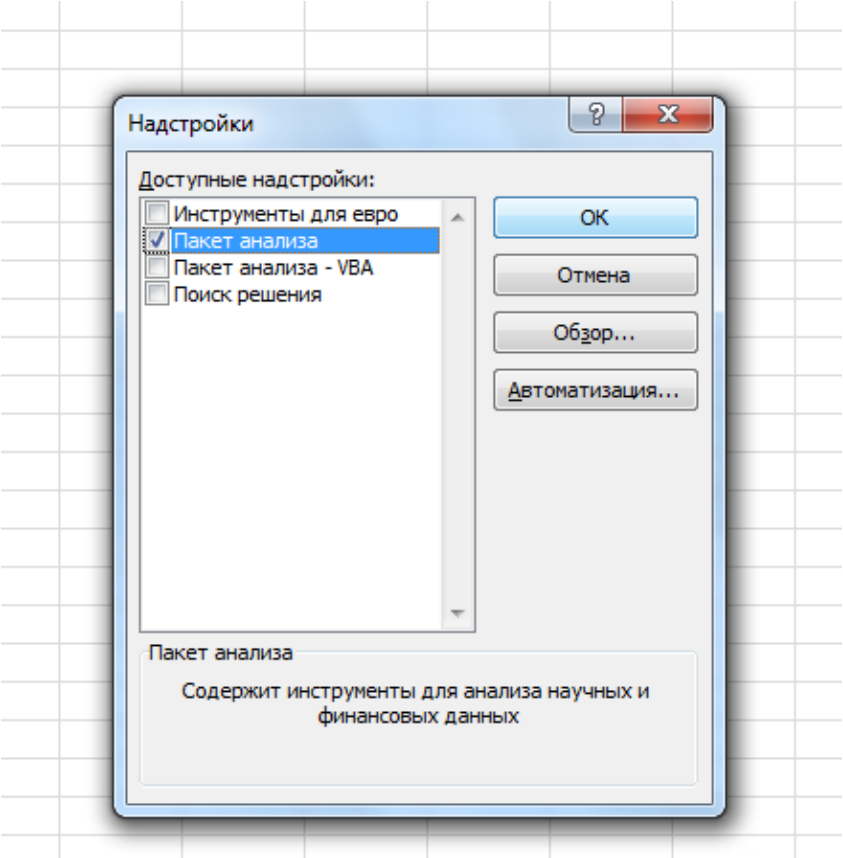


Рис.37 . Добавление команды «Анализ данных»

b) В диалоговом окне Анализ данных выбрали инструмент Регрессия.

c) В диалоговом окне Регрессия в поле Входной интервал У ввели адрес одного диапазона ячеек, который представляет зависимую переменную. В поле Входной интервал X ввели адрес диапазона, который содержит значения независимой переменной (рис. 38).

d) Установили флажок Метки в первой строке для отображения заголовков столбцов.

e) Выбрали параметры вывода. В данном примере Выходной интервал $A$15.

f) В поле Остатки поставили необходимые флажки.

g) ОК.

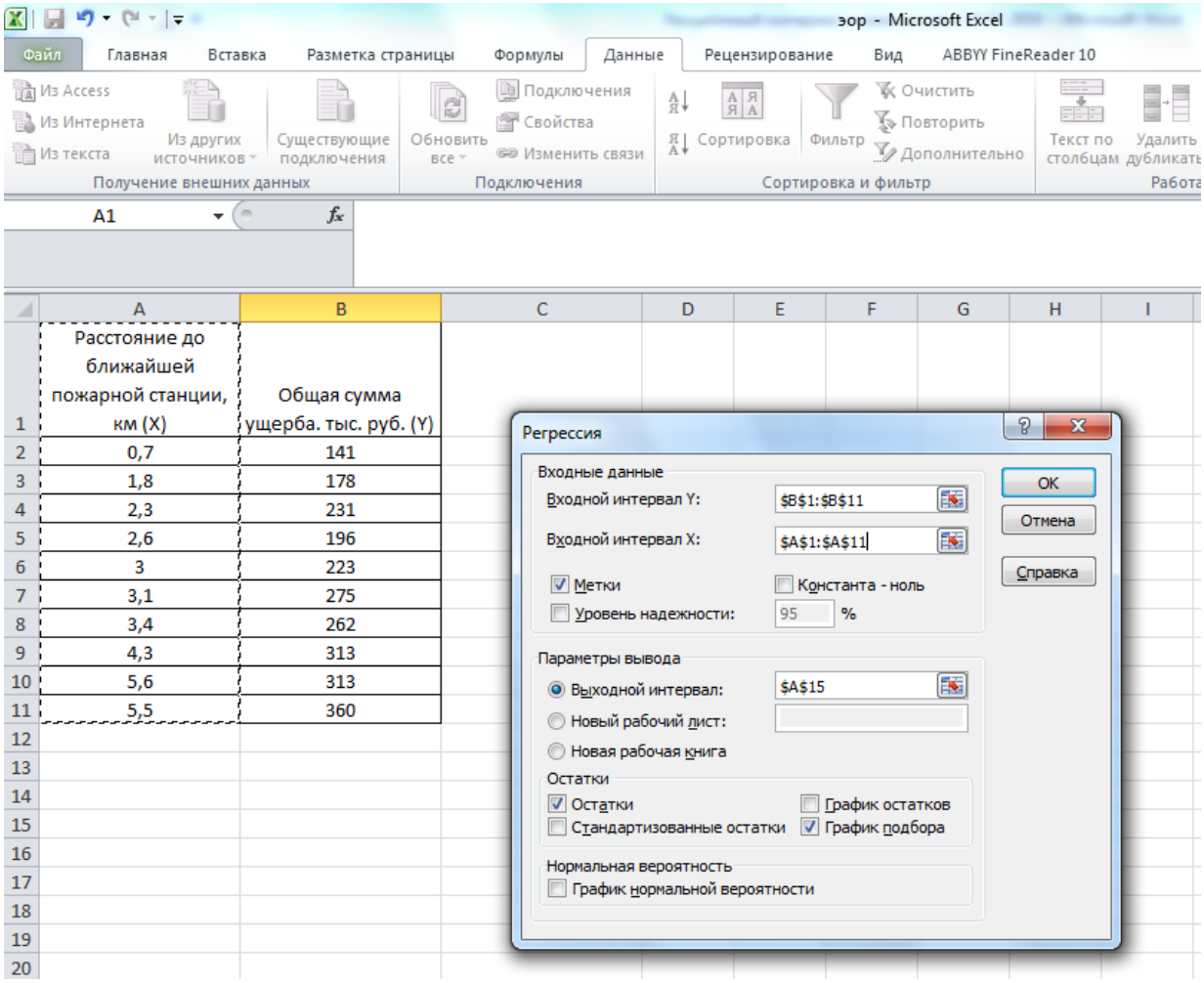


Рис.38 . Диалоговое окно Регрессия подготовлено к построению модели регрессии

Рассмотрим содержание протокола регрессионного анализа по данным рассматриваемого примера. На рис. 39 показаны четыре таблицы протокола регрессионного анализа, в которых отражены основные итоги расчетов.

Регрессионная статистика	
Множественный R	0,968521059
R-квадрат	0,938033042
Нормированный R-кв	0,930287173
Стандартная ошибка	18,00921871
Наблюдения	10

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	39276,94433	39276,94433	121,1010611	4,13578E-06
Остаток	8	2594,655668	324,3319585		
Итого	9	41871,6			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	102,5043342	14,49596026	7,071234493	0,000104936	69,0765899	135,9320785	69,0765899	135,9320785
Расстояние до ближа	46,86762485	4,258914983	11,00459273	4,13578E-06	37,04654929	56,68870042	37,04654929	56,68870042

Рис. 39. Фрагмент протокола выполнения регрессионного анализа

В табл. 9 приведены формулы, по которым выполнены расчеты в протоколе Регрессионная статистика, и результаты вычислений.

Таблица 9

Пояснения к таблице Регрессионная статистика

№	Наименование в отчете EXCEL	Принятые наименования	Формула
1	Множественный R	Коэффициент множественной корреляции, индекс корреляции	$R = \sqrt{R^2} = 0{,}969$
2	R-квадрат	Коэффициент детерминации, R^2	$R_{xy}^2 = \frac{\sigma_{yобъсн}^2}{\sigma_{yобщ}^2} = \frac{\sum(\widehat{y_x} - \bar{y})^2}{\sum(y - \bar{y})^2} = 0{,}938$
3	Нормированный R-квадрат	Скорректированный R^2	$\widehat{R^2} = 1 - (1 - R^2) * \frac{n-1}{n-m-1} = 0{,}93$
4	Стандартная ошибка	Среднеквадратическое отклонение от модели	$\sigma = \sqrt{\frac{\sum(\widehat{y_x} - \bar{y})^2}{n-m-1}} = 18{,}009$
5	Наблюдения	Количество наблюдений n	n=10

В табл. 10 приведены формулы, по которым выполнены расчеты в протоколе Дисперсионный анализ, и результаты вычислений.

Таблица 10

Пояснения к таблице Дисперсионный анализ

	Df – число степеней свободы	SS- сумма квадратов отклонений	MS=SS/Df	F- критерий Фишера= MS (регрессия)/ SS(остаток)	Значимость F
Регрессия	m=1	$\sum(\widehat{y_i} - \bar{y})^2$ 39276,94	$\frac{\sum(\widehat{y_x} - \bar{y})^2}{m}$ 39276,94	$F = \frac{R^2/_k}{(1-R^2)/(n-m-1)}$ 121,101	Значение уровня значимости, соответствующее вычисленной F-статистике 4,13578E-06
Остаток	n-m-1=8	$\sum \widehat{\varepsilon_i}^2$ 2594,66	$\frac{\sum \widehat{\varepsilon_i}^2}{n-m-1}$ 324,33		
Итого	n-1=9	$\sum(y_i - \bar{y})^2$ 41871,6			

В третьей таблице отчета (табл. 11) приводятся значения следующих параметров:

Коэффициенты — значения коэффициентов (параметров) уравнения регрессии.

Стандартная ошибка - стандартные ошибки коэффициентов уравнения регрессии.

t-статистика — расчетные значения t-критерия, используемые для проверки значимости коэффициентов уравнения регрессии.

P-значение — (значимость t) — это значение уровня значимости, соответствующее вычисленной t-статистики.

Если P-значение меньше стандартного уровня значи- мости, то соответствующий коэффициент статистически значим.

Нижние 95% и Верхние 95% - нижние и верхние границы 95% -ных доверительных интервалов для коэффициентов теоретического уравнения регрессии.

Таблица 11

Третья таблица отчета Excel

	Коэффи-циенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	102,504	14,496	7,071	0,000	69,077	135,932
Расстояние до ближайшей пожарной станции, км (X)	46,868	4,259	11,005	0,000	37,047	56,689

В таблице 12 приведены вычисленные (предсказанные) по модели значения зависимой переменной $\widehat{y_i}$, и значения остаточной компоненты $\widehat{\varepsilon_i}$ — Остатки.

Таблица 12

Вывод остатка

Наблюдение	Предсказанное Общая сумма ущерба. тыс. руб. (Y)	Остатки

1	135,31	5,69
2	186,87	-8,87
3	210,30	20,70
4	224,36	-28,36
5	243,11	-20,11
6	247,79	27,21
7	261,85	0,15
8	304,04	8,96
9	318,10	-5,10
10	360,28	-0,28

Уравнение регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции можно записать в следующем виде:

$$\widehat{y_i} = 102{,}5 + 46{,}9 * x_i$$

(14,5) (4,3)

Интерпретация параметров модели: 46,9 показывает, что при увеличении расстояния до ближайшей пожарной станции на 1 км общая сумма ущерба увеличится в среднем на 46,9 тыс. руб.; 102,5 показывает среднюю сумму ущерба, если объект возгорания находится рядом с пожарной станцией (0 км). В скобках указаны стандартные ошибки коэффициентов регрессии. График подбора представлен на рис. 62.

Решение 1) b)

Построение модели парной линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции с использованием надстройки Excel **Поиск решения**.

Согласно принципу метода наименьших квадратов, оценки a_0 и a_1, находятся путем минимизации суммы квадратов отклонений RSS по всем возможным значениям a_0 и a_1, при заданных (наблюдаемых) значениях $x_1,\ldots x_n$,

$y_1,\ldots y_n$. Задача сводится к математической задаче поиска точки минимума функции двух переменных. Задача может быть решена с использованием надстройки Excel **Поиск решения**.

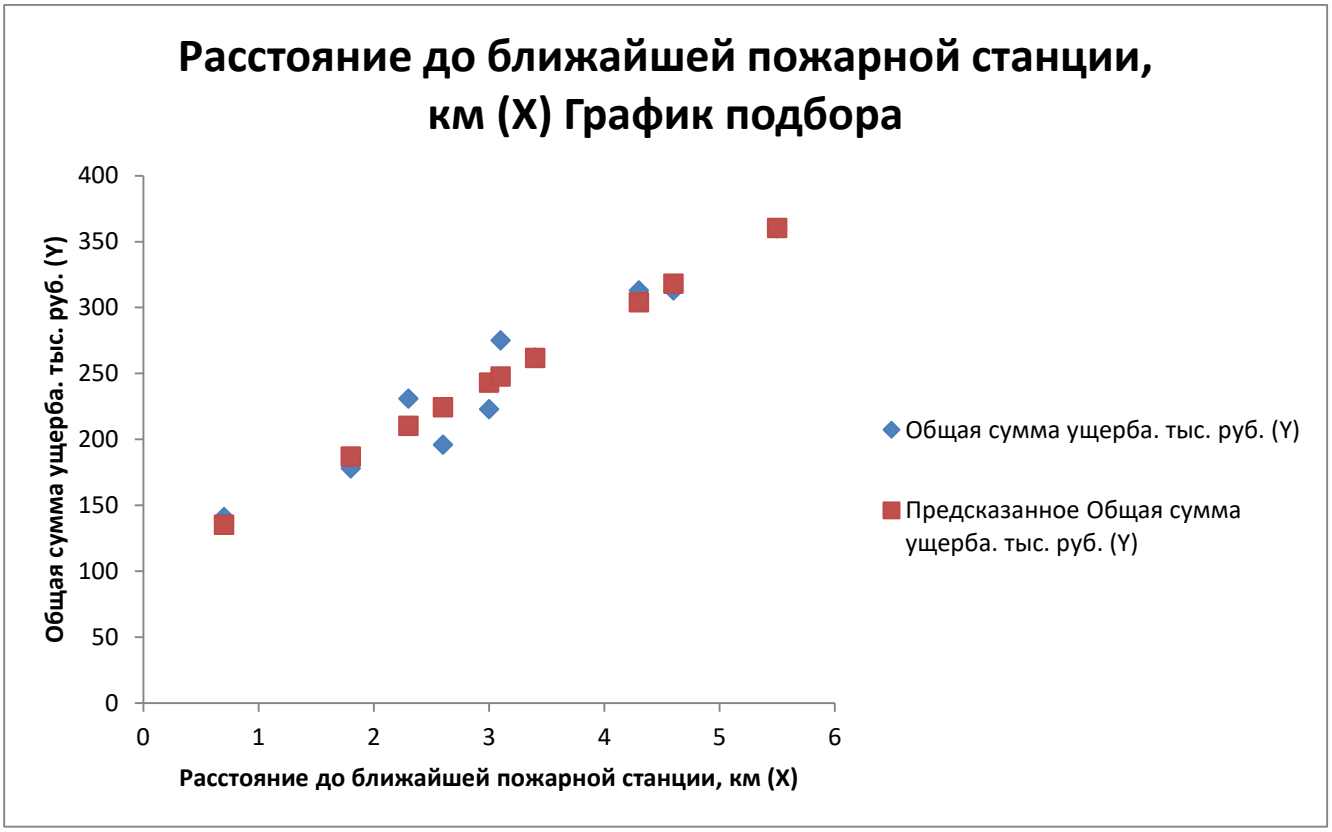


Рис. 40. График подбора

Поиск решения — это надстройка Excel, которая позволяет решать оптимизационные задачи.

В диалоговом окне Поиск решения есть три основных параметра:

- Оптимизировать целевую функцию.
- Изменяя ячейки переменных.
- В соответствии с ограничениями.

Рассмотрим технологию оценки параметров модели линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до

ближайшей пожарной станции на основании данных (табл. 8) с использованием надстройки Excel Поиск решения.

Изменяя ячейки переменных. Здесь указываются ячейки, значения в которых будут изменяться для того, чтобы оптимизировать результат в целевой ячейке. В нашем примере это ячейки E8-E8 (Рис. 41).

	A	B	C	D	E	F	G	H	I	J
1										
2	МНК в Поиске решений									
3										
4	Расстояние до ближайшей пожарной станции, км (X)	Общая сумма ущерба. тыс. руб. (Y)							Yi=A0+A1*Xi	ei*ei
5	0,7	141							=$E$8+$F$8*A5	=(B5-I5)^2
6	1,8	178							=$E$8+$F$8*A6	=(B6-I6)^2
7	2,3	231			A0	A1			=$E$8+$F$8*A7	=(B7-I7)^2
8	2,6	196							=$E$8+$F$8*A8	=(B8-I8)^2
9	3	223							=$E$8+$F$8*A9	=(B9-I9)^2
10	3,1	275							=$E$8+$F$8*A10	=(B10-I10)^2
11	3,4	262							=$E$8+$F$8*A11	=(B11-I11)^2
12	4,3	313							=$E$8+$F$8*A12	=(B12-I12)^2
13	4,6	313							=$E$8+$F$8*A13	=(B13-I13)^2
14	5,5	360							=$E$8+$F$8*A14	=(B14-I14)^2
15										=СУММ(J5:J14)
16										
17										
18										

Рис. 41. Введены формулы для вычисления значения целевой функции (режим «Показать формулы»)

Далее заполним поле Оптимизировать целевую функцию. Целевая ячейка связана с другими ячейками этого рабочего листа с помощью формул. В нашем примере это ячейка J15, в которой в результате введенных формул получим сумму квадратов отклонений расчетных данных от фактических (рис. 42-43,45).

Для запуска Поиска решений надо на вкладке Данные выбрать команду Поиск решения и указать в появившемся меню, адреса целевой функции, изменяемых ячеек и выбрать поиск наименьшего значения.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		МНК в Поиске решений										
3												
4	Расстояние до ближайшей пожарной станции, км (X)	Общая сумма ущерба. тыс. руб. (Y)							Yi=A0+A1	ei*ei		
5	0,7	141							0	19881		
6	1,8	178							0	31684		
7	2,3	231			A0	A1			0	53361		
8	2,6	196							0	38416		
9	3	223							0	49729		
10	3,1	275							0	75625		
11	3,4	262							0	68644		
12	4,3	313							0	97969		
13	4,6	313							0	97969		
14	5,5	360							0	129600		
15										662878		
16												
17												

Значение целевой функции

Рис. 42. Получено начальное значение целевой функции

Если в меню «Данные» еще нет команды **Поиск решений**, то необходимо сделать следующее.

Вкладка Файл→Параметры→Надстройки→кнопка «перейти» →в окне «надстройки» установить флажок «Поиск решений» →ОК (рис. 43).

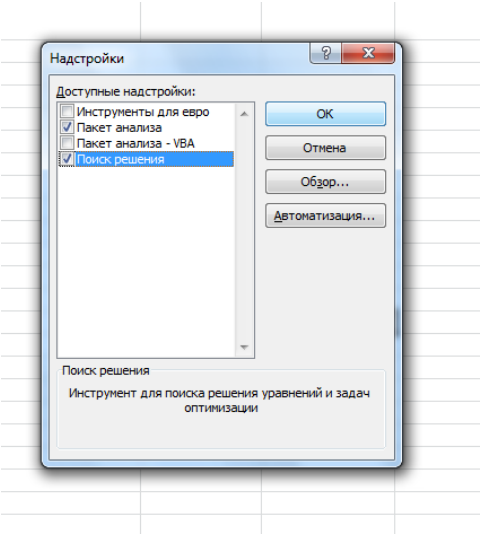


Рис.43 . Добавление команды «Поиск решения»

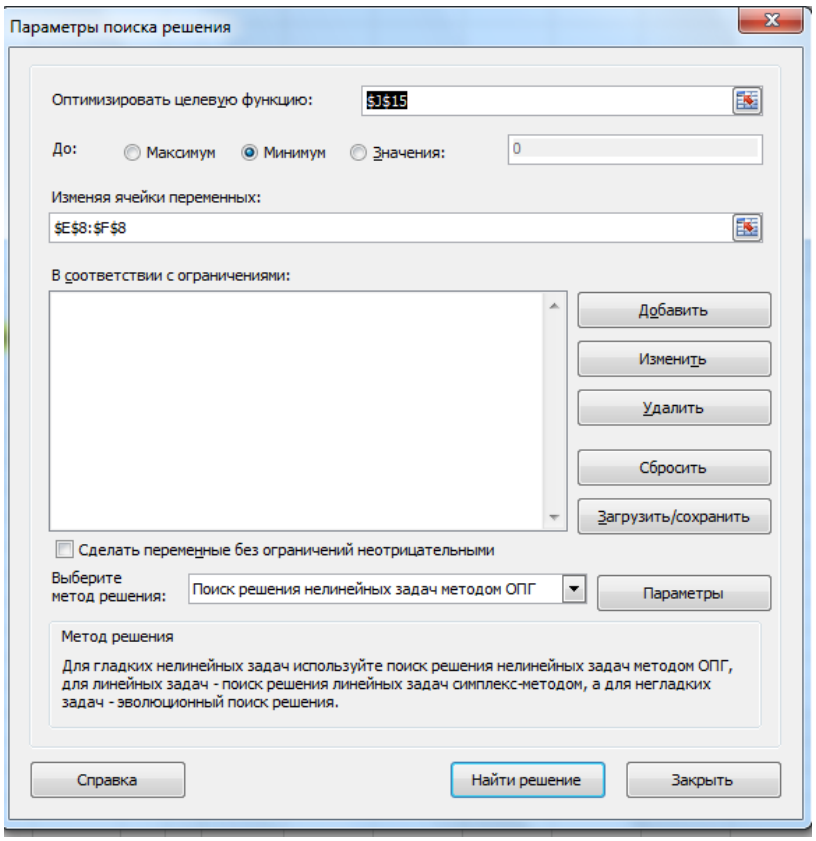


Рис.44 . Диалоговое окно Поиск решения

На рисунках 45-46 показано, что оптимальное решение найдено и можно записать полученную модель:

$$\hat{y}_i = 102{,}5 + 46{,}9 * x_i$$

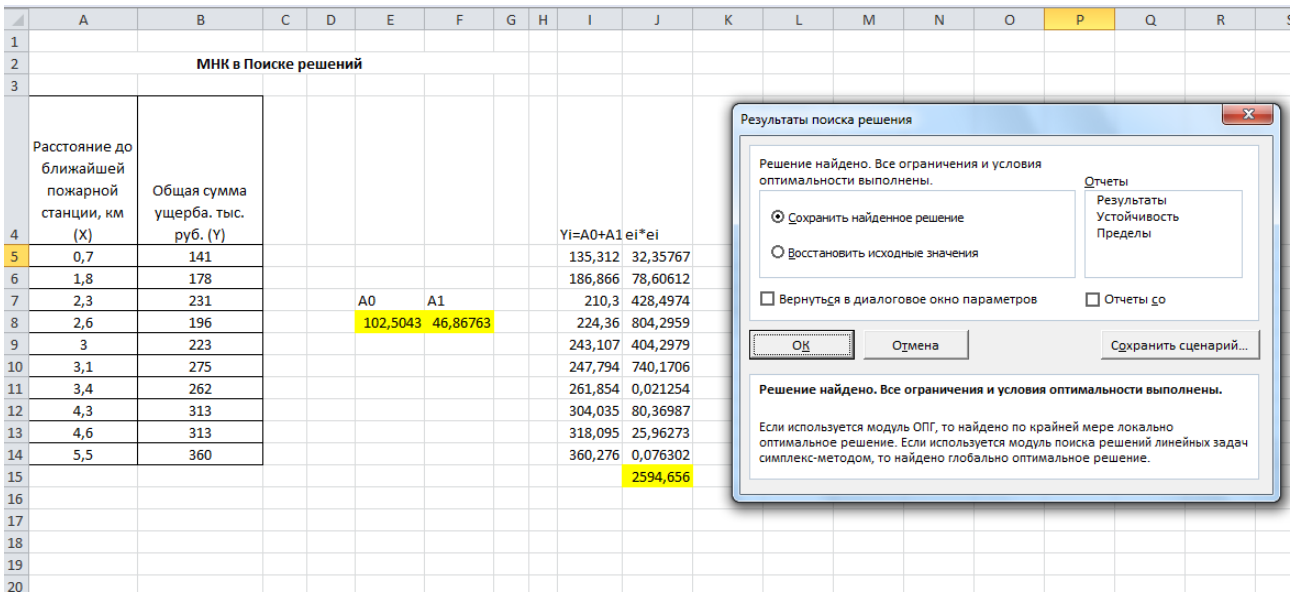
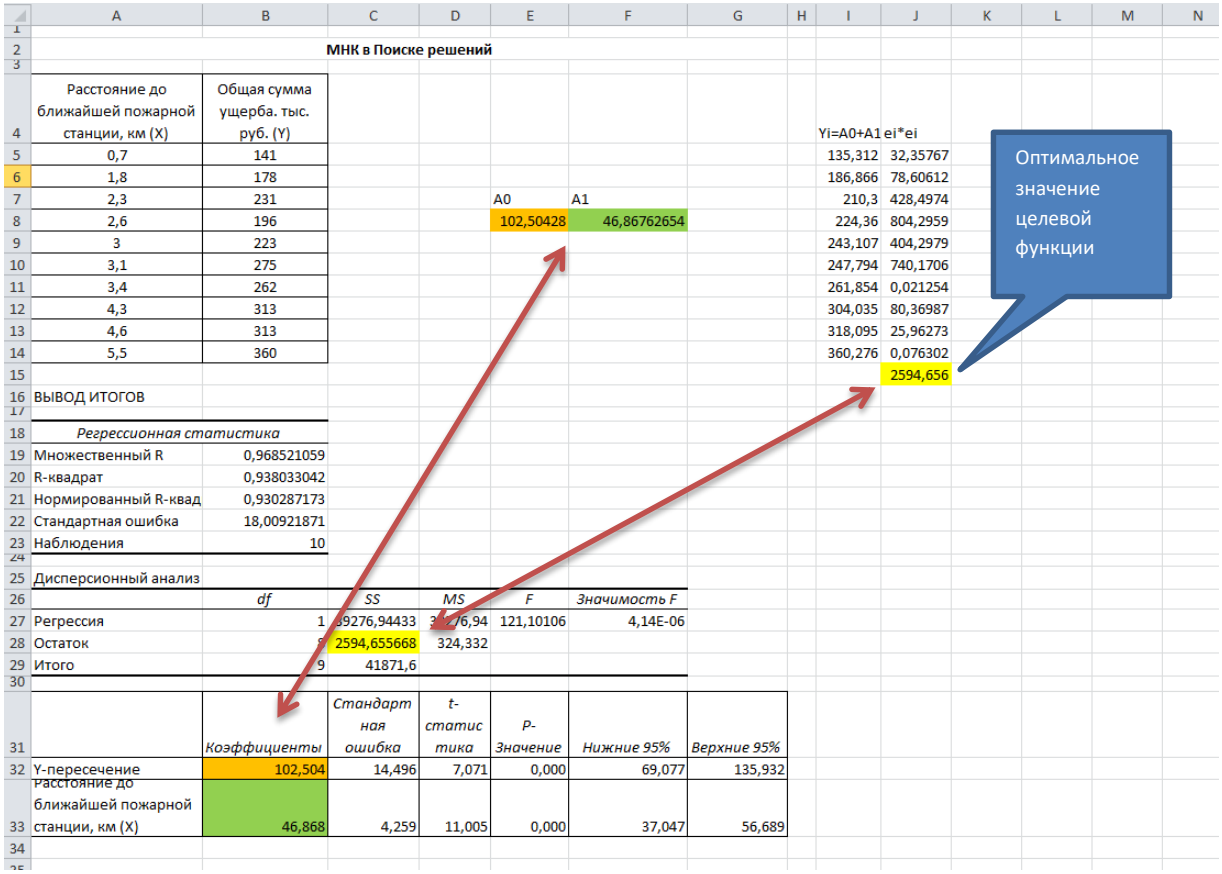


Рис.45 . Оптимальное решение найдено



Расстояние до ближайшей пожарной станции, км (X)	Общая сумма ущерба. тыс. руб. (Y)
0,7	141
1,8	178
2,3	231
2,6	196
3	223
3,1	275
3,4	262
4,3	313
4,6	313
5,5	360

	A0	A1
	102,50428	46,86762654

Yi=A0+A1	ei*ei
135,312	32,35767
186,866	78,60612
210,3	428,4974
224,36	804,2959
243,107	404,2979
247,794	740,1706
261,854	0,021254
304,035	80,36987
318,095	25,96273
360,276	0,076302
	2594,656

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,968521059
R-квадрат	0,938033042
Нормированный R-квад	0,930287173
Стандартная ошибка	18,00921871
Наблюдения	10

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	39276,94433	39276,94	121,10106	4,14E-06
Остаток	8	2594,655668	324,332		
Итого	9	41871,6			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	102,504	14,496	7,071	0,000	69,077	135,932
Расстояние до ближайшей пожарной станции, км (X)	46,868	4,259	11,005	0,000	37,047	56,689

Рис.46 . Сравнение результатов, полученных с помощью надстройки Поиск решения и пакета Анализ данных

Решение 1) с)

Построение модели линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции с использованием матричных функций.

Рассмотрим технологию оценки параметров модели линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции на основании данных (табл. 13) с использованием матричных функций.

В матричной форме расчет параметров линейной модели парной регрессии по приведенной формуле ($A = (X'X)^{-1}X'y$) может быть выполнен с помощью матричных функций МУМНОЖ и МОБР.

При подготовке данных, формируя матрицу X для вычисления свободного члена a_0 необходимо добавить столбец x_0, состоящий из единиц.

1. Транспонируем матрицу X. Это можно выполнить с использованием функции ТРАНСП или путем последовательного копирования и специальной вставки транспонирования.

Итак, копируем матрицу X. А затем, используя, специальную вставку, получаем транспонированную матрицу.

2. Умножаем транспонированную матрицу X^T на матрицу X.

- Выделяем диапазон ячеек для результата умножения матриц - (H8:19) размером 2x2. Результатом является массив с таким же числом строк, как массив транспонированная матрица X^T, т.е. 2 и с таким же числом столбцов, как матрица X, т.е. тоже 2.
- Затем вводится формула умножения матриц = МУМНОЖ(H3:Q4, D4: E13).
- Затем следует нажать клавиши CTRL+SHIFT+ENTER.

3. Вычисляем обратную матрицу.

- Выделяем диапазон для размещения обратной матрицы (H12:I13).
- В категории Математические выбираем функцию вычисления обратной матрицы =МОБР(H8:I9), в качестве массива указываем диапазон ячеек H8:I9, где размещена матрица, к которой надо вычислить обратную.
- Затем следует нажать клавиши CTRL+SHIFT+ENTER.

4. Умножаем обратную матрицу на транспонированную матрицу.

- Выделяем диапазон ячеек для результата умножения матриц (H16:Q17) размером 2 x 10. Результатом является массив с таким же числом строк, как обратная матрица, т.е. 2 и с таким же числом столбцов, как транспонированная матрица , т.е. 10.
- Затем вводится формула умножения матриц = МУМНОЖ(H12:I13, D4: E13).
- Затем следует нажать клавиши CTRL+SHIFT+ENTER.

5. Умножаем матрицу полученную матрицу на y.

- Выделяем диапазон ячеек для результата умножения матриц (H19: H20) размером 2x 1.
- Затем вводится формула умножения матриц = МУМНОЖ(H16:Q17, B4: B13).
- Затем следует нажать клавиши CTRL+SHIFT+ENTER.

В ячейках H19: H20 будут находиться параметры модели линейной регрессии a_0 =102,5 и a_1=46,87.

В матричной форме расчет параметров модели может быть представлен на рис. 47.

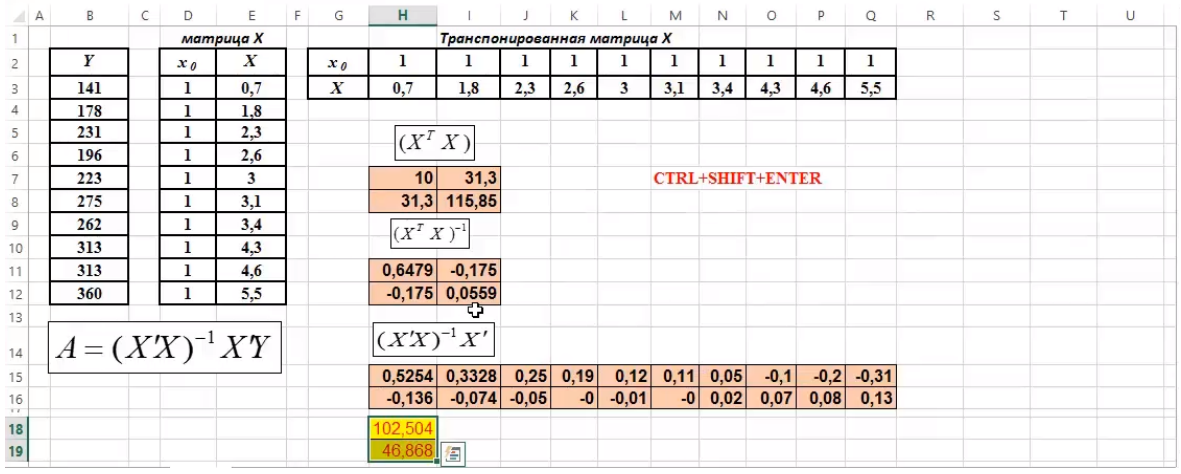


Рис.47 . Вычисление параметров модели с помощью матричных функций МУМНОЖ и МОБР

Решение 1) d)

Построение модели линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции с использованием функции **ЛИНЕЙН** табличного процессора Excel.

Рассмотрим технологию оценки параметров модели линейной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции на основании данных (табл. 13) с использованием с использованием функции ЛИНЕЙН.

Вначале выделяем диапазон для размещения результатов выполнения функции ЛИНЕЙН. На листе выделяется область высотой 5 строк и шириной равной количеству столбцов с данными, в нашем примере их два (D2:E6). Затем вызываем функцию ЛИНЕЙН.

При применении функции ЛИНЕЙН табличного процессора Excel на экран дисплея выдается диалоговое окно (рис. 48) для задания значений (y) и (x).

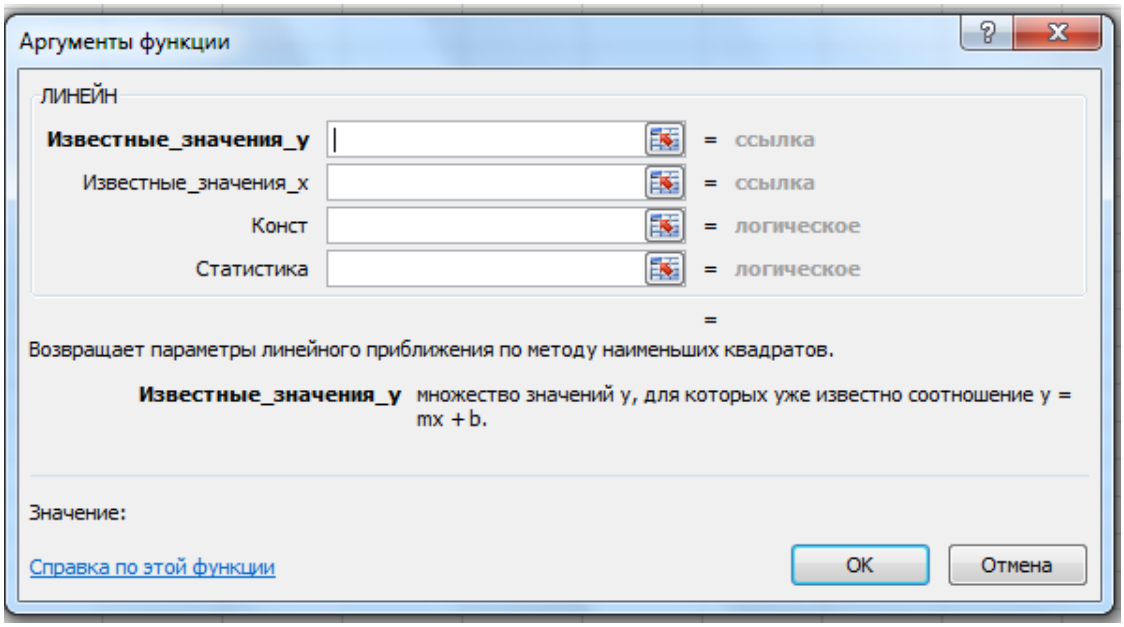


Рис.48 . Диалоговое окно функции ЛИНЕЙН

В данном окне в строке «Известные значения y» следует внести все 10 значений из столбца y, а в строке «Известные значения x» — соответственно все 10 значений из столбца x. Если необходимо выполнить оценку двух параметров — постоянного и регрессионного, то в строке «Конст» следует указать значение 1. Так как при оценке параметров модели нас, в первую очередь, интересуют статистические сведения, то в окне строки «Статистика» следует указать значение 1. После ввода всех значений, как показано на рис. 49; следует нажать клавиши CTRL+SHIFT+ENTER.

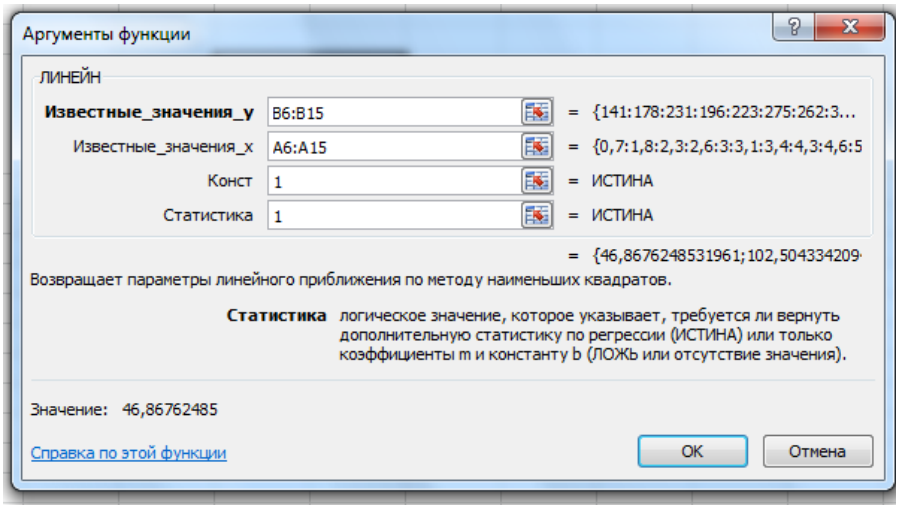


Рис.49 . Диалоговое окно функции ЛИНЕЙН после задания необходимых данных

После этого, будет выдан результат, как показано на рис. 50.

	A	B	C	D	E	F	G	H	I	J	K
1	Расстояние до ближайшей пожарной станции, км (X)	Общая сумма ущерба, тыс.руб. (Y)									
2	0,7	141		46,87	102,5			a1	a0		
3	1,8	178		4,259	14,5			Стандартная ошибка			
4	2,3	231		0,938	18,01			R2	Стандартная ошибка		
5	2,6	196		121,1	8			F	n-k-1		
6	3,0	223		39277	2595			RSS	ESS		
7	3,1	275									
8	3,4	262									
9	4,3	313									
10	4,6	313									
11	5,5	360									
12											
13											

Рис.50 . Результат выполнения функции ЛИНЕЙН

В таблице справа от полученных значений указаны их обозначения (наименования), которые поясняются ниже:

- a_1 и a_0 – оценки параметров полученной парной регрессии (1-я строка);
- стандартные ошибки оцененных параметров (2-я строка);
- R^2 – коэффициент детерминации (3-я строка, 1-ый столбец);
- сстандартная ошибка полученной регрессии (3-я строка, 2-ой столбец);
- F – статистика Фишера полученной парной регрес- сии (4-я строка, 1-ый столбец);
- ν — степень свободы (3-я строка, 2-ой столбец);
- RSS — (5-я строка, 1-ый столбец);
- ESS — квадрат остатков для полученной парной регрессии (5-я строка, 2-ой столбец).

Итак, в первой строке выдаются значения параметров a_1 и a_0, а во второй строке соответственно их стандартные ошибки.

В результате применения всех инструментов Excel были получены одинаковые параметры модели регрессии (рис. 51).

	A	B	C	D	E	F	G	H	I
4									
5	Расстояние до ближайшей пожарной станции, км (X)	Общая сумма ущерба. тыс. руб. (Y)				46,86762	102,5043		
6	0,7	141				4,258915	14,49596		
7	1,8	178				0,938033	18,00922		
8	2,3	231				121,1011	8		
9	2,6	196				39276,94	2594,656		
10	3	223							
11	3,1	275							
12	3,4	262							
13	4,3	313							
14	4,6	313							
15	5,5	360							
16									
17	ВЫВОД ИТОГОВ								
18									
19	*Регрессионная статистика*								
20	Множественный R	0,968521059							
21	R-квадрат	0,938033042							
22	Нормированный R-квадрат	0,930287173							
23	Стандартная ошибка	18,00921871							
24	Наблюдения	10							
25									
26	Дисперсионный анализ								
27		*df*	*SS*	*MS*	*F*	*ачимость F*			
28	Регрессия	1	39276,94	39276,94	121,1011	4,14E-06			
29	Остаток	8	2594,656	324,332					
30	Итого	9	41871,6						
31									
32		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*		
33	Y-пересечение	102,504	14,496	7,071	0,000	69,077	135,932		
34	Расстояние до ближайшей пожарной станции, км (X)	46,868	4,259	11,005	0,000	37,047	56,689		
35									

Рис.51 . Сравнение результатов, полученных с помощью выполнения функции ЛИНЕЙН и пакета Анализ данных

2) **Проверим качество уравнения регрессии.**

Качество модели оценивается коэффициентом детерминации R^2.

Величина R^2= 0,938 означает, что фактором расстояния до ближайшей пожарной станции можно объяснить 93,8% вариации (разброса) стоимости ущерба, нанесенного пожаром.

Оценим значимость уравнения регрессии с помощью критерия Фишера. Расчетное значение F- критерия вы числим по формуле:

$$F = \frac{R^2}{1-R^2} * \frac{n-m-1}{m} = \frac{0{,}938}{1-0{,}938} * \frac{10-2}{1} = 121{,}03$$

Расчетное значение F-критерия Фишера можно найти в таблице *Дисперсионный анализ* протокола Excel].

Уравнение регрессии значимо на уровне α, если расчетное значение F>Fтабл., где Fтабл. — табличное значение F-критерия Фишера

Табличное значение F-критерия можно найти с помощью функции F.ОБР.ПХ.

Табличное значение F-критерия при α=0,05, k_1=m=1 и k_2=n-m-1=10-1-1=8 составляет 5,318.

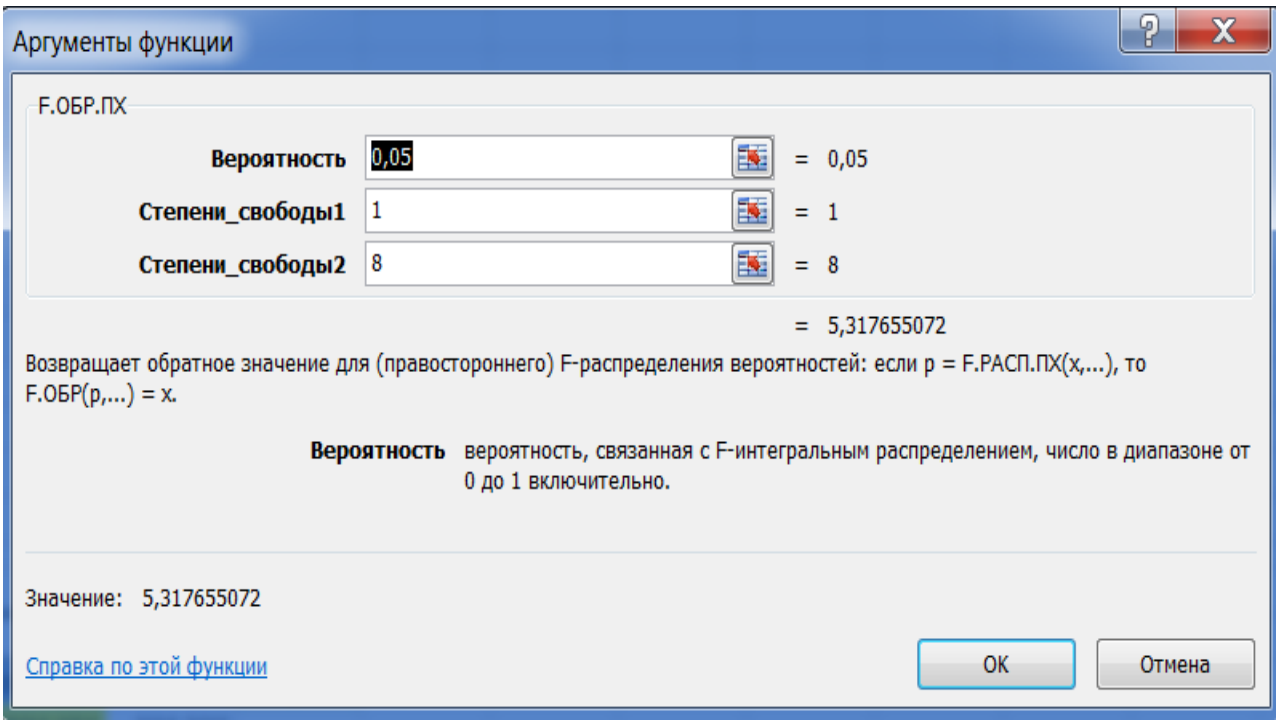


Рис.52 . Функция F.ОБР.ПХ.

Поскольку расчетное значение F-критерия Фишера больше табличного, то уравнение регрессии следует признать значимым.

Оценим значимость коэффициентов полученной модели, используя результаты отчета Excel. Это можно сделать тремя способами.

Коэффициент уравнения регрессии признается значимым в том случае, если:

- Наблюдаемое значение t-статистики Стьюдента для этого коэффициента больше, чем критическое (табличное) значение статистики Стьюдента (для заданного уровня значимости, например, $\alpha = 0,05$ и числа степеней свободы $k=n-m-1$, где n — это число наблюдений, m — число факторов в модели);
- P-значение статистики Стьюдента для этого коэффициента меньше, чем уровень значимости, например, $\alpha = 0,05$;
- Доверительный интервал для этого коэффициента (вычисленный с некоторой доверительной вероятностью, например, 95%) не содержит ноль внутри себя, т.е., если нижняя 95% и верхняя 95% границы доверительного интервала имеют одинаковые знаки,

Значимость коэффициента a1 проверим по второму и третьему способам, используя данные расчетов.

P-значение (a1) = 0,00 < 0,01 <4,05.

Следовательно, коэффициент a1 значим при 1%- ном уровне, а тем более при 5% -ном уровне значимости.

Нижняя 95% граница равна 37,0 и верхняя 95% граница равна 56,7, границы доверительного интервала имеют одинаковые знаки, следовательно, коэффициент a1 значим.

3) Для того, чтобы оценить стоимость ущерба, нанесенного пожаром, если расстояние до ближайшей пожарной станции уменьшится на 10 % от своего среднего уровня, необходимо найти ожидаемое значение фактора x, затем следует подставить значение Хпрогн.=Хсредн.*0,9=3,13*0,9=2,82 в полученную модель:

Упрогн = 102,5+46,9*2,82 = 234,8.

Доверительный интервал для прогнозов индивидуальных значений y, определяется из соотношения :

$$y_i = \hat{y_i} \pm t_\alpha * \hat{\sigma} * \sqrt{1 + \frac{1}{n} + \frac{\left(x_{\text{прогн}} - \bar{x}\right)^2}{\sum(x_i - \bar{x})^2}}$$

$$= 234{,}8 \pm 1{,}86 * 18{,}01 * \sqrt{1 + \frac{1}{10} + \frac{(2{,}82 - 3{,}13)^2}{17{,}881}} = 234{,}8 \pm 35{,}2$$

Коэффициент Стьюдента t (0,05; 8) для m=8 степеней свободы и уровня значимости 0,01 равен 1,86; стандартная ошибка равна 18,01.

Таким образом, прогнозное значение y=234,8 с вероятностью 90% будет находиться между верхней границей, равной 270,0 и нижней границей, равной 199,6.

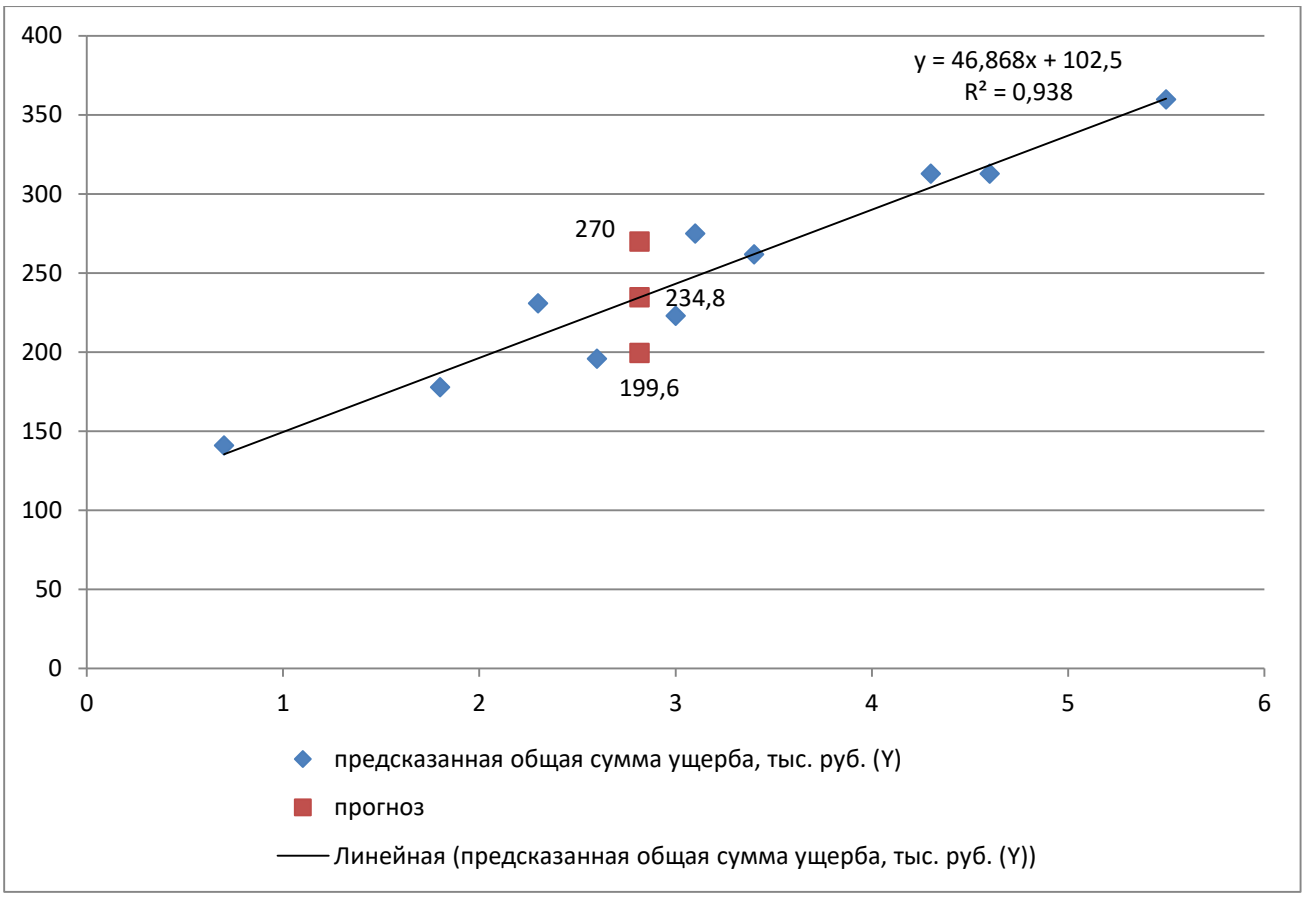


Рис.53 . График модели парной регрессии зависимости стоимости ущерба, нанесенного пожаром от расстояния до ближайшей пожарной станции: точечный и интервальный прогнозы

.

Глава 6. Статистический анализ временных рядов

Тема 6.1 Основы анализа временных рядов

Основные элементы временного ряда

Эконометрическую модель можно построить, используя два типа исходных данных:

- данные, характеризующие совокупность различных объектов в определенный момент (период) времени;
- данные, характеризующие один объект за ряд последовательных моментов (периодов) времени.

Модели, построенные по данным первого типа, называются *пространственными моделями.* Модели, построенные по данным второго типа, называются *моделями временных рядов.*

Временной ряд — это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие циклические колебания ряда;
- случайные факторы.

При различных сочетаниях этих факторов зависимость уровней ряда от времени может принимать разные формы.

Во-первых, большинство временных рядов экономических показателей имеют тенденцию, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. По всей видимости, эти факторы, взятые в отдельности, могут оказывать разнонаправленное

воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию. На рис. 55 показаны компоненты гипотетического временного ряда, содержащего возрастающую тенденцию.

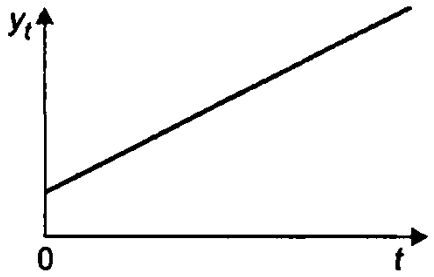


Рис.54 Основные компоненты временного ряда: возрастающая тенденция

Во-вторых, изучаемый показатель может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей зависит от времени года (например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним). При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка, а также с фазой бизнес-цикла, в которой находится экономика страны. На рис. 56 представлен гипотетический временной ряд, содержащий только сезонную компоненту.

Рис.55 Основные компоненты временного ряда: сезонная компонента

Некоторые временные ряды не содержат тенденции и циклическую компоненту, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты. Пример ряда, содержащего только случайную компоненту, приведен на рис. 57.

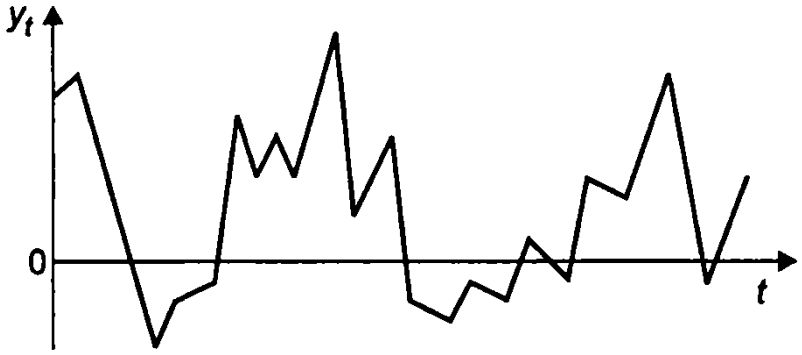


Рис.56 Основные компоненты временного ряд: случайная компонента

Очевидно, что реальные данные не соответствуют полностью ни одной из описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется ***аддитивной моделью*** временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется ***мультипликативной моделью*** временного ряда.

Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент, с тем чтобы использовать полученную

информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

Автокорреляция уровней временного ряда и выявление его структуры

При наличии тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих значений. Корреляционную зависимость между последовательными уровнями временного ряда называют ***автокорреляцией уровней ряда***.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Отметим два важных *свойства коэффициента автокорреляции.* Во-первых, он строится по аналогии с линейным коэффициентом корреляции и, таким образом, характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

Во-вторых, по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом они могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют *автокорреляционной функцией временного ряда.* График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой.*

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, следовательно, лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т. е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, ряд содержит циклические колебания с периодичностью в τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать предположение относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний и имеет структуру, сходную со структурой ряда (см. рис. 6.1 *в),* либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты *T* и циклической (сезонной) компоненты *S.*

Моделирование тенденции временного ряда

Одним из наиболее распространенных способов моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот

способ называют *аналитическим выравниванием временного ряда.*

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются следующие функции:

- линейный тренд $\widehat{y_t} = a + bt$;
- гипербола $\widehat{y_t} = a + \frac{b}{t}$;
- экспоненциальный тренд $\widehat{y_t} = e^{a+bt}$;
- тренд в форме степенной функции $\widehat{y_t} = at^b$;
- парабола второго и более высоких порядков:

$$\widehat{y_t} = a + b_1 t + b_2 t^2 + \cdots + b_k t^k$$

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время t=1,2,...,*n,* а в качестве зависимой переменной – фактические уровни временного ряда y_t. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Известно несколько способов определения типа тенденции; к наиболее распространенным относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени, расчет некоторых основных показателей динамики. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни y_t и y_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то

коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, если ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации R^2 и выбора уравнения тренда с максимальным значением скорректированного коэффициента детерминации. Реализация этого метода относительно проста при компьютерной обработке данных.

Моделирование сезонных и циклических колебаний

Известно несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания.

Простейший подход – расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$Y=T+S+E.$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой T, сезонной S и случайной E компонент. Общий вид мультипликативной модели выглядит так:

$Y = T^* S^* E.$

Данная модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой *T,* сезонной *S* и случайной *E* компонент. Выбор одной из двух моделей проводится на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений *T, S* и *E* для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

Шаг 1. Выравнивание исходного ряда методом скользящей средней.

Шаг 2. Расчет значений сезонной компоненты *S.*

Шаг 3. Устранение сезонной компоненты из исходных уровней ряда и получение выравненных данных *(T+E)* в аддитивной или *(T*E)* в мультипликативной модели.

Шаг 4. Аналитическое выравнивание уровней *(T + E)* или *(T*E)* и расчет значений *T* с использованием полученного уравнения тренда.

Шаг 5. Расчет полученных по модели значений *(T + S)* или *(T*S).*

Ш а г 6. Расчет абсолютных и/или относительных ошибок.

Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок *E* для анализа взаимосвязи исходного ряда и других

временных рядов.

Моделирование тенденции временного ряда при наличии структурных изменений

От сезонных и циклических колебаний следует отличать единовременные изменения характера тенденции временного ряда, вызванные структурными изменениями в экономике или иными факторами. В этом случае, начиная с некоторого момента времени t*, происходит изменение характера динамики изучаемого показателя, что приводит к изменению параметров тренда, описывающего эту динамику. Схематично такая ситуация изображена на рис. 58.

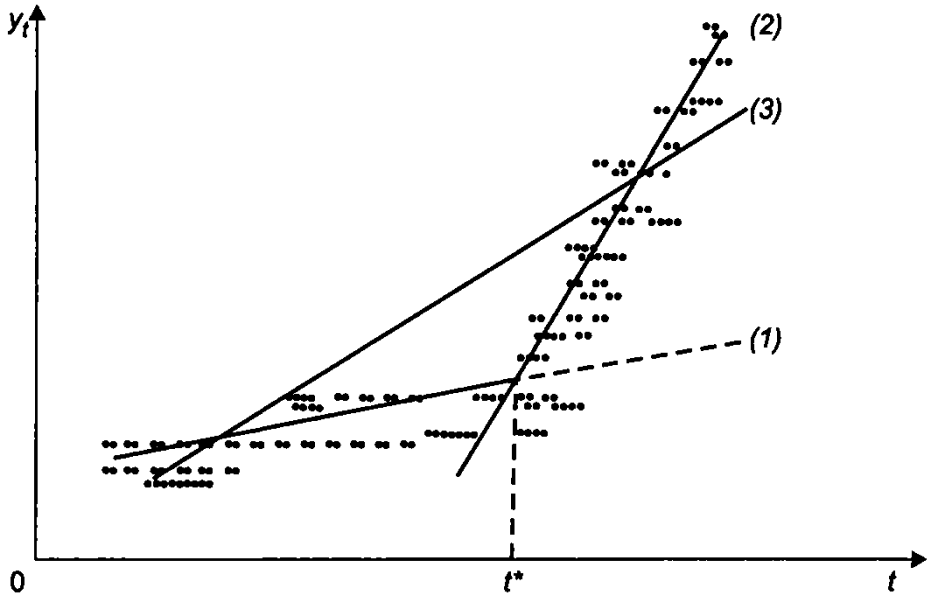


Рис.57 . Изменение характера тенденции временного ряда

Момент (период) времени t^* сопровождается значительными изменениями ряда факторов, оказывающих сильное воздействие на изучаемый показатель y_t. Чаще всего эти изменения вызваны изменениями в общеэкономической ситуации или факторами (событиями) глобального

характера, приведшими к изменению структуры экономики (например, начало крупных экономических реформ, изменение экономического курса, нефтяные кризисы и прочие факторы). Если исследуемый временной ряд включает в себя соответствующий момент (период) времени, то одной из задач его изучения становится выяснение вопроса о том, значимо ли повлияли общие структурные изменения на характер этой тенденции.

Если это влияние значимо, то для моделирования тенденции данного временного ряда следует использовать кусочно-линейные модели регрессии, т.е. разделить исходную совокупность на две подсовокупности (до момента времени t* и после момента t*) и построить отдельно по каждой подсовокупности уравнения линейной регрессии (на рис. 6.8 этим уравнениям соответствуют прямые (1) и (2)). Если структурные изменения незначительно повлияли на характер тенденции ряда y_t, то ее можно описать с помощью единого для всей совокупности данных уравнения тренда (на рис. 6.8 этому уравнению соответствует прямая (*3*)).

Каждый из описанных выше подходов имеет свои положительные и отрицательные стороны. При построении кусочно-линейной модели происходит снижение остаточной суммы квадратов по сравнению с единым для всей совокупности уравнением тренда. Однако разделение исходной совокупности на две части ведет к потере числа наблюдений и, следовательно, к снижению числа степеней свободы в каждом уравнении кусочно-линейной модели. Построение единого для всей совокупности уравнения тренда, напротив, позволяет сохранить число наблюдений n исходной совокупности, однако остаточная сумма квадратов по этому уравнению будет выше по сравнению с кусочно-линейной моделью. Очевидно, что выбор одной из двух моделей (кусочно-линейной или единого уравнения тренда) будет зависеть от соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от единого уравнения регрессии к кусочно-

линейной модели.

Тема 6.2 Фильтрация и сглаживание временного ряда

Сущность фильтрации и сглаживания. Медианная фильтрация.

Распространенным приемом при выявлении тенденции развития является сглаживание временного ряда. Суть различных приемов сглаживания сводится к замене фактических уровней временного ряда расчетными уровнями, которые подвержены колебаниям в меньшей степени. Это способствует более четкому проявлению тенденции развития. Иногда сглаживание применяют как предварительный этап перед использованием других методов выделения тенденции.

Фильтрация – это удаление чего-то, что является ненужным или разделение на компоненты.

Сглаживание **-** это процедура исключения или уменьшения размаха колебаний значений временного ряда.

Медианная фильтрация получила развитие в последние 20 лет.

Говорят, что Ymed – является ***медианой*** подпоследовательности из 2m+1 значений (y_{k-1},…, y_k, y_{k+1}), если выполняется :

$$Ymed = Z(k),$$

где

$$Z(i) \in (y_{k-m}, \dots y_{k+m}),\ i=1 \dots 2m+1,$$

$$Z(1) \le Z(2) \le \dots \le Z(2m+1)$$

То есть последовательность Z(i) представляет собой упорядоченную по возрастанию последовательность значения “y”

Пример: пусть m=1, 2m+1 = 3

0.1, -3, 4

Z1	Z2	Z3
-3	0.1	4

Тогда Ymed = 0.1

В зависимости от значения 2m+1, m=1 – медиана по тройкам, m=2 – медиана по пятеркам и т.д.

В рассмотрение вовлекаются: отрезок исходного временного ряда, содержащий (2m+1) подряд идущих значений. Процедура фильтрации осуществляется для всех значений временного ряда, кроме первого и последнего.

Достоинства медианой фильтрации заключаются в следующем.

Эта фильтрация позволяет исключить выбросы, то есть значения временного ряда, которые «сильно» отличаются от остальных. При этом эти выбросы заменяются соседними значениями той же последовательности. Количество вовлекаемых в рассмотрение значений исходного временного ряда называется ***апертурой фильтра***. При апертуре 2m+1 можно исключить m – подряд идущих выбросов. Медианная фильтрация не искажает монотонно возрастающих или монотонно убывающих отрезков временного ряда.

Метод скользящего среднего

Скользящие средние позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса, и поэтому, являются важным инструментом при фильтрации компонент временного ряда.

Алгоритм сглаживания по простой скользящей средней может быть представлен в виде следующей последовательности шагов:

1. Определяют длину интервала сглаживания g, включающего в себя g последовательных уровней ряда (g<n). При этом надо иметь в виду, что чем шире интервал сглаживания, тем в большей степени взаимопогашаются колебания, и тенденция развития носит более плавный, сглаженный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

2. Разбивают весь период наблюдений на участки, при этом интервал сглаживания как бы скользит по ряду с шагом, равным 1.

3. Рассчитывают арифметические средние из уровней ряда, образующих каждый участок.

4. Заменяют фактические значения ряда, стоящие в центре каждого участка, на соответствующие средние значения.

При этом удобно брать длину интервала сглаживания g в виде нечетного числа: g=2p+1, т.к. в этом случае полученные значения скользящей средней приходятся на средний член интервала.

Наблюдения, которые берутся для расчета среднего значения, называются активным участком сглаживания. При нечетном значении g все уровни активного участка могут быть представлены в виде:

$$y_{t-p}, y_{t-p+1}, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_{t+p-1}, y_{t+p}$$

а скользящая средняя определена по формуле:

$$\widehat{y_t} = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \cdots + y_{t+p-1} + y_{t+p}}{2p+1}$$

где y_i – фактическое значение i-го уровня;

$\widehat{y_t}$– значение скользящей средней в момент t;

2p+1– длина интервала сглаживания.

Процедура сглаживания приводит к полному устранению периодических колебаний во временном ряду, если длина интервала сглаживания берется равной или кратной циклу, периоду колебаний.

Для устранения сезонных колебаний желательно было бы использовать четырех- и двенадцатичленную скользящие средние, но при этом не будет выполняться условие нечетности длины интервала сглаживания. Поэтому при четном числе уровней принято первое и последнее наблюдение на активном участке брать с половинными весами:

$$\widehat{y_t} = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \cdots + y_{t-1} + y_t + y_{t+1} + \cdots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p}$$

$$= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}$$

Тогда для сглаживания сезонных колебаний при работе с временными рядами квартальной или месячной динамики можно использовать следующие скользящие средние:

$$\widehat{y_t} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}$$

$$\widehat{y_t} = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \cdots + y_t + \cdots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}$$

При использовании скользящей средней с длиной активного участка g=2p+1 первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек является существенным недостатком, т.к. для исследователя последние "свежие" данные обладают наибольшей информационной ценностью. Рассмотрим один из приемов, позволяющих восстановить потерянные значения временного ряда. Для этого необходимо:

1) Вычислить средний прирост на последнем активном участке

$$y_{t-p}, y_{t-p+1}, \cdots, y_t, \cdots, y_{t+p-1} + y_{t+p}$$

$$-\Delta y = \frac{y_{t+p} - y_{t-p}}{g - 1}$$

где g – длина активного участка;

y_{t+p}– значение последнего уровня на активном участке;

y_{t-p}– значение первого уровня на активном участке;

$-\Delta y$– средний абсолютный прирост.

2) Получить P сглаженных значений в конце временного ряда путем последовательного прибавления среднего абсолютного прироста к последнему сглаженному значению.

Аналогичную процедуру можно реализовать для оценивания первых уровней временного ряда.

Метод простой скользящей средней применим, если графическое изображение динамического ряда напоминает прямую. Когда тренд выравниваемого ряда имеет изгибы, и для исследователя желательно сохранить мелкие волны, применение простой скользящей средней нецелесообразно.

Если для процесса характерно нелинейное развитие, то простая скользящая средняя может привести к существенным искажениям. В этих случаях более надежным является использование взвешенной скользящей средней.

При сглаживании по взвешенной скользящей средней на каждом участке выравнивание осуществляется по полиномам невысоких порядков. Чаще всего используются полиномы 2-го и 3-его порядка. Так как при простой скользящей средней выравнивание на каждом активном участке производится по прямой (полиному первого порядка), то метод простой скользящей средней может рассматриваться как частный случай метода взвешенной скользящей средней. Простая скользящая средняя учитывает все уровни ряда, входящие в активный участок сглаживания, с равными весами, а взвешенная средняя приписывает каждому уровню вес, зависящий от удаления данного уровня до уровня, стоящего в середине активного участка.

Выравнивание с помощью взвешенной скользящей средней осуществляется следующим образом.

Для каждого активного участка подбирается полином вида

$$\widehat{y_t} = a_0 + a_1 t + a_2 t^2 + \cdots$$

параметры которого оцениваются по методу наименьших квадратов. При этом начало отсчета переносится в середину активного участка. Например, для длины интервала сглаживания g=5, индексы уровней активного участка будут следующими i: -2, -1, 0, 1, 2.

Тогда сглаженным значением для уровня, стоящего в середине активного участка, будет значение параметра a_0 подобранного полинома.

Нет необходимости каждый раз заново вычислять весовые коэффициенты при уровнях ряда, входящих в активный участок сглаживания, т.к. они будут одинаковыми для каждого активного участка. Причем при сглаживании по полиному к-ой нечетной степени весовые коэффициенты будут такими же, как при сглаживании по полиному (к-1) степени. В таблице 2 представлены весовые коэффициенты при сглаживании по полиному 2-го или 3-го порядка (в зависимости от длины интервала сглаживания).

Таблица 13

Весовые коэффициенты при сглаживании по полиномам второго и третьего порядка

Длина интервала сглаживания	Весовые коэффициенты
5	$\frac{1}{35}[-3, +12, +17]$
7	$\frac{1}{21}[-2, +3, +6, +7]$
9	$\frac{1}{231}[-21, +14, +39, +54, +59]$
11	$\frac{1}{429}[-36, +9, +44, +69, +84, +89]$

Так как веса симметричны относительно центрального уровня, то в таблице использована символическая запись: приведены веса для половины уровней активного участка; выделен вес, относящийся к уровню, стоящему в центре участка сглаживания. Для оставшихся уровней веса не приводятся, т. к. они могут быть симметрично отражены.

Например, проиллюстрируем использование таблицы для сглаживания по параболе 2-го порядка по 5-членной взвешенной скользящей средней. Тогда центральное значение на каждом активном участке $y_{t-2}, y_{t-1}, y_t, y_{t+1}, y_{t+2}$ будет оцениваться по формуле:

$$\widehat{y_t} = \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}$$

Отметим важные свойства приведенных весов:

1) Они симметричны относительно центрального уровня.

2) Сумма весов с учетом общего множителя, вынесенного за скобки, равна единице.

3) Наличие как положительных, так и отрицательных весов, позволяет сглаженной кривой сохранять различные изгибы кривой тренда.

Существуют приемы, позволяющие с помощью дополнительных вычислений получить сглаженные значения для P начальных и конечных уровней ряда при длине интервала сглаживания g=2p+1.

Метод экспоненциального сглаживания (модель Брауна)

Метод экспоненциального сглаживания наиболее эффективен при разработке среднесрочных прогнозов. Он приемлем при прогнозировании только на один период вперед. Его основные достоинства простота процедуры вычислений и возможность учета весов исходной информации. Рабочая формула метода экспоненциального сглаживания:

$$U_{t+1} = \alpha^* y_t + (1 - \alpha)^* U_t$$

где t – период, предшествующий прогнозному; t+1 – прогнозный период; U_{t+1} – прогнозируемый показатель; α – параметр сглаживания; $У_t$ – фактическое значение исследуемого показателя за период, предшествующий прогнозному; U_t – экспоненциально взвешенная средняя для периода, предшествующего прогнозному.

При прогнозировании данным методом возникает два затруднения:

- выбор значения параметра сглаживания α;
- определение начального значения Uo.

От величины α зависит, как быстро снижается вес влияния предшествующих наблюдений. Чем больше α, тем меньше сказывается влияние предшествующих лет. Если значение α близко к единице, то это приводит к учету при прогнозе в основном влияния лишь последних наблюдений. Если значение α близко к нулю, то веса, по которым взвешиваются уровни временного ряда, убывают медленно, т.е. при прогнозе учитываются все (или почти все) прошлые наблюдения.

Таким образом, если есть уверенность, что начальные условия, на основании которых разрабатывается прогноз, достоверны, следует использовать небольшую величину параметра сглаживания (α→0). Когда параметр сглаживания мал, то исследуемая функция ведет себя как средняя из большого числа прошлых уровней. Если нет достаточной уверенности в начальных условиях прогнозирования, то следует использовать большую величину α, что приведет к учету при прогнозе в основном влияния последних наблюдений.

Точного метода для выбора оптимальной величины параметра сглаживания α нет. В отдельных случаях автор данного метода профессор

Браун предлагал определять величину α, исходя из длины интервала сглаживания. При этом α вычисляется по формуле:

$$\alpha = \frac{2}{n+1}$$

где n – число наблюдений, входящих в интервал сглаживания.

Задача выбора U_o (экспоненциально взвешенного среднего начального) решается следующими способами:

- если есть данные о развитии явления в прошлом, то можно воспользоваться средней арифметической и приравнять к ней U_o;
- если таких сведений нет, то в качестве U_o используют исходное первое значение базы прогноза $У_1$.

Также можно воспользоваться экспертными оценками.

Отметим, что при изучении экономических временных рядов и прогнозировании экономических процессов метод экспоненциального сглаживания не всегда «срабатывает». Это обусловлено тем, что экономические временные ряды бывают слишком короткими (15-20 наблюдений), и в случае, когда темпы роста и прироста велики, данный метод не «успевает» отразить все изменения.

Анализ временных рядов и прогнозирование в Excel на примере

Анализ временных рядов позволяет изучить показатели во времени. Подобные данные распространены в самых разных сферах человеческой деятельности: ежедневные цены акций, курсов валют, ежеквартальные, годовые объемы продаж, производства и т.д. Типичный временной ряд в метеорологии, например, ежемесячный объем осадков.

Если фиксировать значения какого-то процесса через определенные промежутки времени, то получатся элементы временного ряда. Их изменчивость пытаются разделить на закономерную и случайную составляющие. Закономерные изменения членов ряда, как правило, предсказуемы.

Сделаем анализ временных рядов в Excel. Пример: торговая сеть анализирует данные о продажах товаров магазинами, находящимися в городах с населением менее 50 000 человек. Период – 2012-2015 гг. Задача – выявить основную тенденцию развития.

	A	B	C
1	Год	Квартал	Продажи, млн.руб.
2	2012	1	165
3		2	253
4		3	316
5		4	287
6	2013	1	257
7		2	308
8		3	376
9		4	351
10	2014	1	410
11		2	431
12		3	443
13		4	389
14	2015	1	436
15		2	459
16		3	492
17		4	470

Рис.58 .Данные о реализации

Заходим в меню «Данные» → команда Анализ данных, из предлагаемого списка инструментов для статистического анализа выбираем «Экспоненциальное сглаживание». Этот метод выравнивания подходит для нашего динамического ряда, значения которого сильно колеблются.

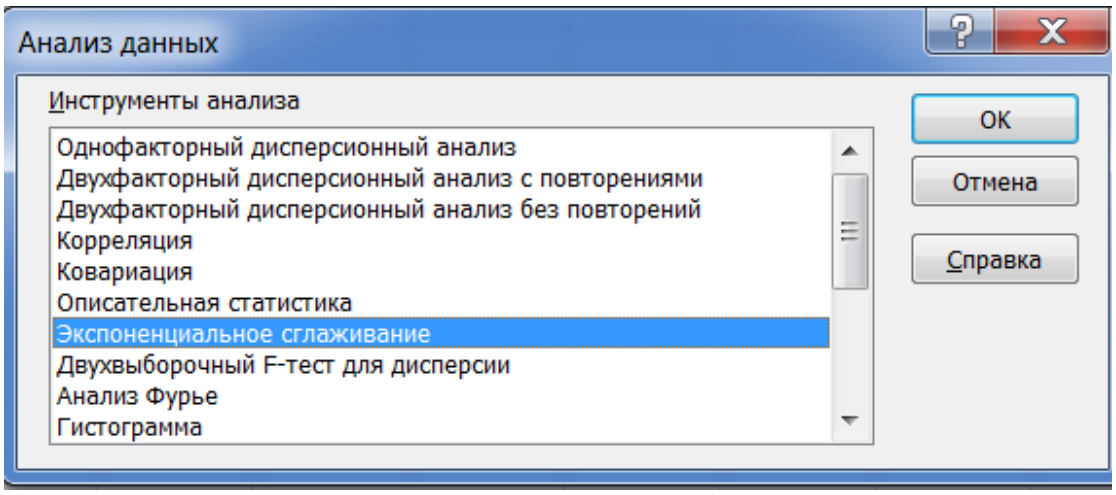


Рис.59 .Окно команды «Анализ данных»

Заполняем диалоговое окно. Входной интервал – диапазон со значениями продаж. Фактор затухания – коэффициент экспоненциального сглаживания (по умолчанию – 0,3). Выходной интервал – ссылка на верхнюю левую ячейку выходного диапазона. Сюда программа поместит сглаженные уровни и размер определит самостоятельно. Ставим галочки «Вывод графика», «Стандартные погрешности».

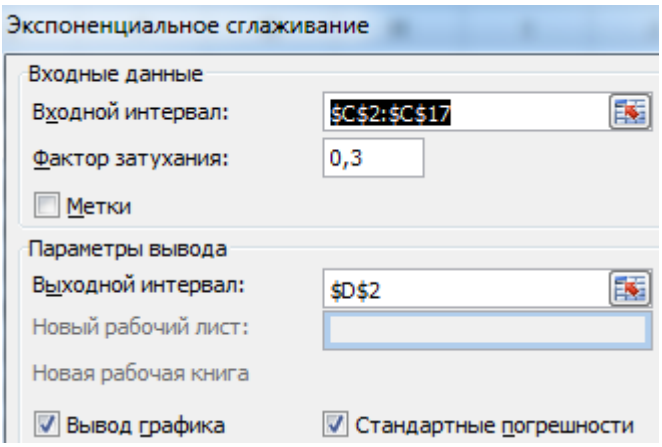


Рис.60 .Окно команды «Экспоненциальное сглаживание»

Закрываем диалоговое окно нажатием ОК.

Год	Квартал	Продажи, млн.руб.	Сглаженные уровни	Стандартные погрешности
2012	1	165	#Н/Д	#Н/Д
	2	253	165	#Н/Д
	3	316	226,6	#Н/Д
	4	287	289,18	#Н/Д
2013	1	257	287,654	72,43643742
	2	308	266,1962	54,57954475
	3	376	295,45886	29,95539913
	4	351	351,837658	55,29948958
2014	1	410	351,2512974	52,39317576
	2	431	392,3753892	57,55862797
	3	443	419,4126168	40,59545248
	4	389	435,923785	42,81602217
2015	1	436	403,0771355	37,63892846
	2	459	426,1231407	35,78696809

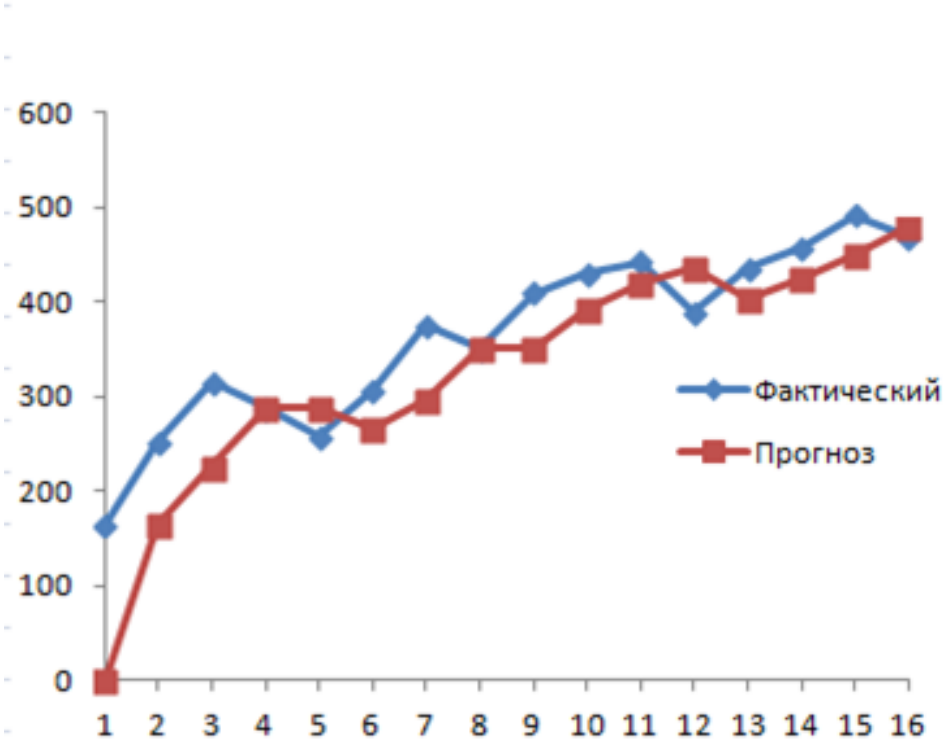


Рис.61 Результаты анализа

Для расчета стандартных погрешностей Excel использует формулу: =КОРЕНЬ(СУММКВРАЗН(‘диапазон фактических значений’; ‘диапазон

прогнозных значений’)/‘размер окна сглаживания’). Например, =КОРЕНЬ(СУММКВРАЗН(C3:C5;D3:D5)/3).

Прогнозирование временного ряда в EXCEL

Составим прогноз продаж, используя данные из предыдущего примера.

На график, отображающий фактические объемы реализации продукции, добавим линию тренда (правая кнопка по графику – «Добавить линию тренда»).

Настраиваем параметры линии тренда:

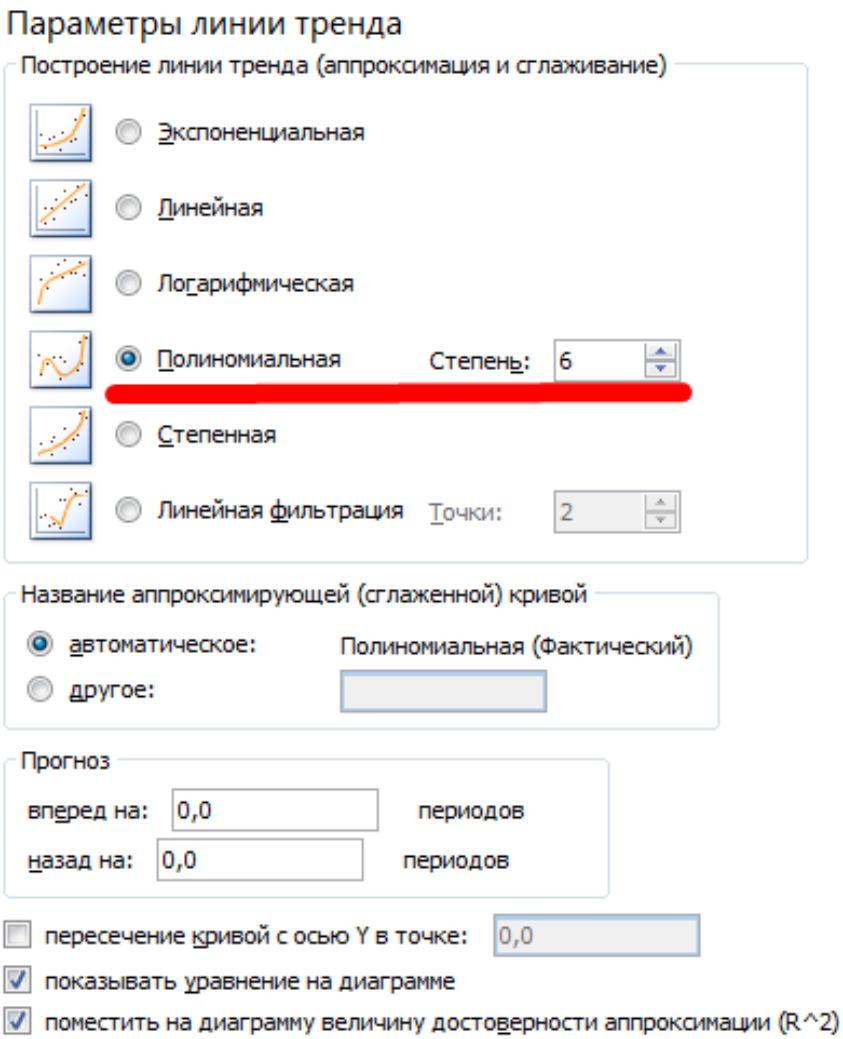


Рис.62 Окно «параметры линии тренда»

Выбираем полиномиальный тренд, что максимально сократить ошибку прогнозной модели.

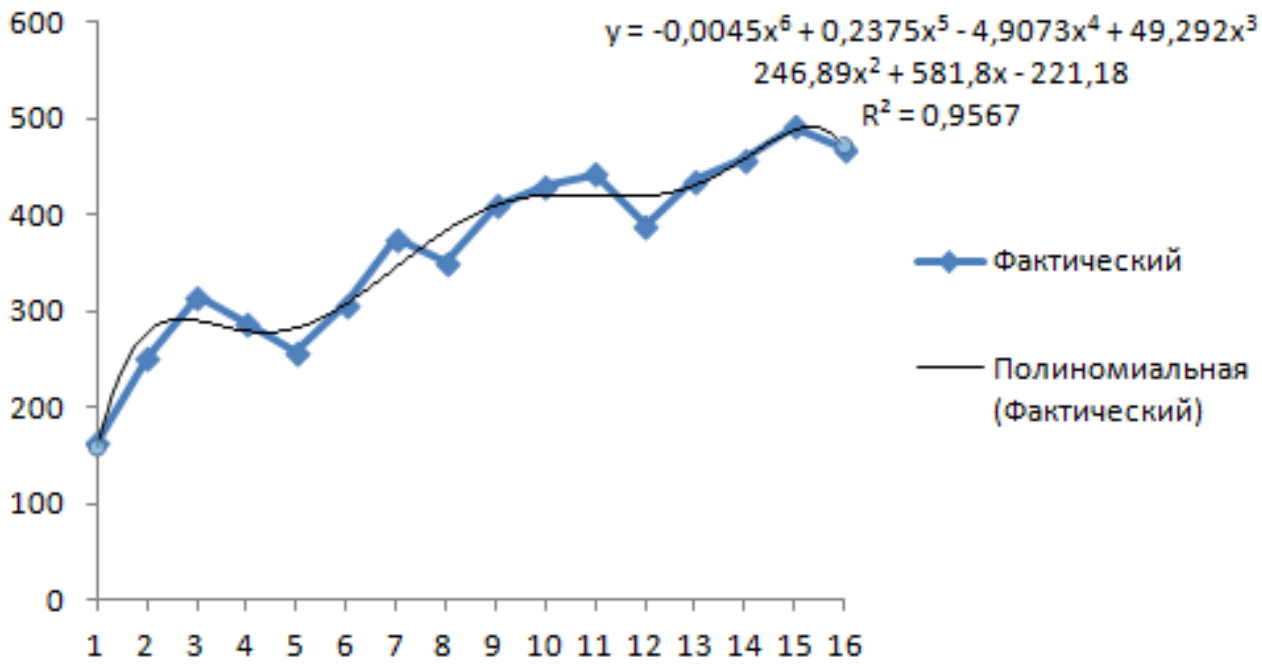


Рис.63 График полиномы

R^2 = 0,9567, что означает: данное отношение объясняет 95,67% изменений объемов продаж с течением времени.

Уравнение тренда – это модель формулы для расчета прогнозных значений.

Большинство авторов для прогнозирования продаж советуют использовать линейную функцию тренда. Чтобы на графике увидеть прогноз, в параметрах необходимо установить количество периодов.

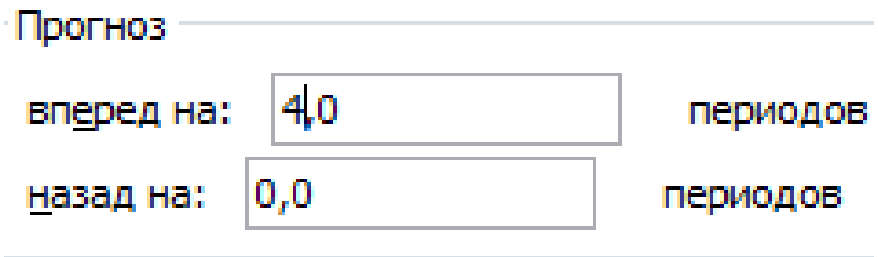


Рис.64 Установка периодов прогноза

Получаем достаточно оптимистичный результат:

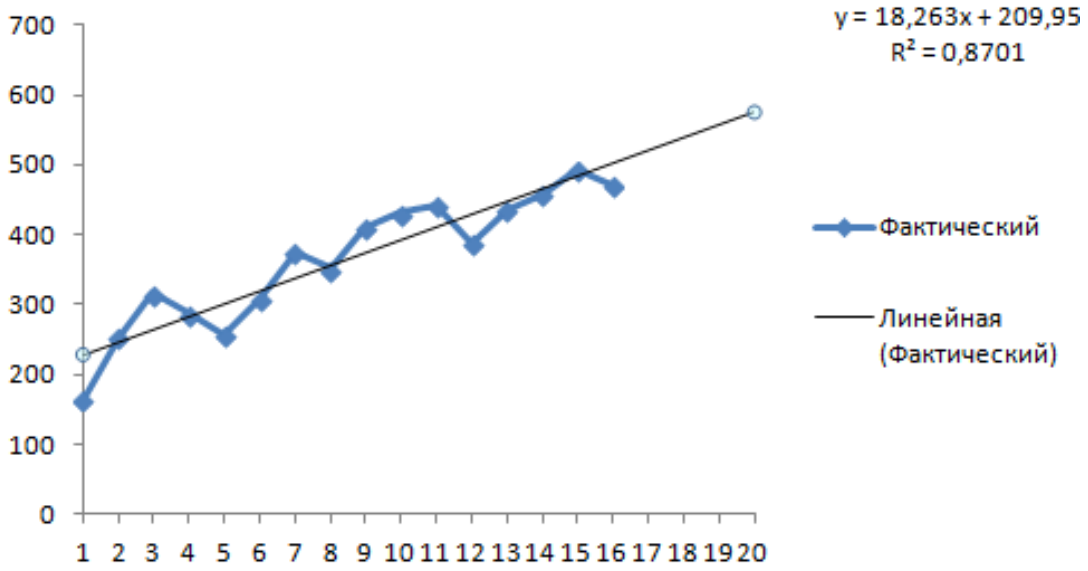


Рис.65 График линейной функции

В нашем примере все-таки экспоненциальная зависимость. Поэтому при построении линейного тренда больше ошибок и неточностей.

Для прогнозирования экспоненциальной зависимости в Excel можно использовать также функцию РОСТ.

=РОСТ($D$2:$D$17;$B$2:$B$17;B2;ИСТИНА)

C	D	E	F	G
Квартал	**Продажи, млн.руб.**	**Сглаженные уровни**	**Стандартные погрешности**	**Прогноз буд.периодов**
1	165	#Н/Д	#Н/Д	232,81
2	253	165	#Н/Д	246,01
3	316	226,6	#Н/Д	259,96
4	287	289,18	#Н/Д	274,70
1	257	287,654	72,43643742	290,27
2	308	266,1962	54,57954475	306,73
3	376	295,45886	29,95539913	324,12
4	351	351,837658	55,29948958	342,50
1	410	351,2512974	52,39317576	361,92
2	431	392,3753892	57,55862797	382,44
3	443	419,4126168	40,59545248	404,12
4	389	435,923785	42,81602217	427,03
1	436	403,0771355	37,63892846	451,25

Рис.66 Функция РОСТ

Для линейной зависимости – ТЕНДЕНЦИЯ.

При составлении прогнозов нельзя использовать какой-то один метод: велика вероятность больших отклонений и неточностей.

Глава 7. Системы эконометрических уравнений

Тема 7.1 Системы одновременных уравнений

Общее понятие о системах уравнений, используемых в эконометрике

Объектом статистического изучения в социальных науках являются сложные системы. Измерение тесноты связей между переменными, построение изолированных уравнений регрессии недостаточны для описания таких систем и объяснения механизма их функционирования. При использовании отдельных уравнений регрессии, например, для экономических расчетов в большинстве случаев предполагается, что аргументы (факторы) можно изменять независимо друг от друга. Однако, это предположение является очень грубым: практически изменение одной переменной, как правило, не может происходить при абсолютной неизменности других. Ее изменение повлечет за собой изменения во всей системе взаимосвязанных признаков.

Следовательно, отдельно взятое уравнение множественной регрессии не может характеризовать истинные влияния отдельных признаков на вариацию результирующей переменной. Именно поэтому в экономических, биометрических и социологических исследованиях важное место заняла проблема описания структуры связей между переменными системы так называемых одновременных уравнений или структурных уравнений. Например, если изучается модель спроса как соотношение цен и количества потребляемых товаров, то одновременно для прогнозирования спроса необходима модель предложения товаров, в которой рассматривается также взаимосвязь между количеством и ценой предлагаемых благ. Это позволяет достичь равновесия между спросом и предложением.

Также при оценке эффективности производства нельзя руководствоваться только моделью рентабельности. Она должна быть дополнена моделью

производительности труда, а также моделью себестоимости единицы продукции.

В еще большей степени возрастает потребность в использовании системы взаимосвязанных уравнений, если мы переходим от исследований на микроуровне к макроэкономическим расчетам. Модель национальной экономики включает в себя следующую систему уравнений: функции потребления, инвестиций заработной платы, тождество доходов и т. д. Это связано с тем, что макроэкономические показатели, являясь обобщающими показателями состояния экономики чаще всего взаимозависимы. Так, расходы на конечное потребление в экономике зависят от валового национального дохода. вместе с тем величина валового национального дохода рассматривается как функция инвестиций.

Система уравнений в эконометрических исследованиях может быть построена по-разному:

Возможна *система независимых уравнений* – когда каждая зависимая переменная *y* рассматривается как функция одного и того же набора факторов *x*:

$$\left(\begin{array}{l} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m + \varepsilon_n. \end{array}\right)$$

Набор факторов x_i в каждом уравнении может варьировать. Например, модель вида

$$y_1 = f(x_1, x_2, x_3, x_4, x_5);$$

$$y_2 = f(x_1, x_3, x_4, x_5);$$

$$y_3 = f(x_2, x_3, x_5);$$

$$y_4 = f(x_3, x_4, x_5).$$

также является системой независимых уравнений с тем лишь отличием, что набор факторов в ней видоизменяется в уравнениях, входящих в систему. Отсутствие того или иного фактора в уравнении системы может быть следствием как экономической нецелесообразности его включения в модель, так и несущественности его воздействия на результативный признак (незначимо значение t-критерия или частного F-критерия для данного фактора).

Для решения этой системы и нахождения ее параметров используется метод наименьших квадратов.

Однако, если зависимая переменная *y* одного уравнения выступает в виде фактора *x* в другом уравнении, то исследователь может строить модель в виде *системы рекурсивных уравнений*:

$$\begin{pmatrix} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \cdots + a_{3m}x_m + \varepsilon_3, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m + \varepsilon_n. \end{pmatrix}$$

В данной системе зависимая переменная y включает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений наряду с набором собственно факторов x. Примером такой системы может служить *модель производительности труда и фондоотдачи* вида:

$$\begin{pmatrix} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2 \end{pmatrix}$$

где y_1 – производительность труда;

где y_2 – фондоотдача;

где x_1 – фондовооруженность труда;

где x_2 – энерговооруженность труда;

где x_3 – квалификация рабочих.

Как и в предыдущей системе, каждое уравнение может рассматриваться самостоятельно, и его параметры определяются методом наименьших квадратов.

Наибольшее распространение в эконометрических исследованиях получила *система взаимосвязанных (совместных) уравнений.* В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других – в правую часть системы:

$$\begin{pmatrix} y_1 = b_{12}y_2 + b_{13}y_3 + \cdots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1 \\ y_2 = b_{21}y_1 + b_{23}y_3 + \cdots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \cdots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m + \varepsilon_n. \end{pmatrix}$$

Система взаимозависимых уравнений получила название *система совместных, одновременных уравнений.* Тем самым подчеркивается, что в системе одни и те же переменные *y* одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других. В эконометрике эта система уравнений называется также ***структурной формой модели***. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

Примером системы одновременных уравнений может служить *модель динамики цены и заработной платы вида*

$$\begin{pmatrix} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2 \end{pmatrix}$$

где y_1 – темп изменения месячной заработной платы;

где y_2 – темп изменения цен;

где x_1 – процент безработных;

где x_2 – темп изменения постоянного капитала;

где x_3 – темп изменения цен на импорт сырья.

Структурная и приведенная формы модели

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные (y) – это зависимые переменные, которые число которых равно числу уравнений в системе.

Экзогенные переменные (x). Это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них.

Простейшая структурная форма модели имеет вид:

$$\begin{pmatrix} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{pmatrix}$$

Классификация переменных на эндогенные и экзогенные зависит от теоретической концепции принятой модели. Экономические переменные могут выступать в одних моделях как эндогенные, а в других – как экзогенные переменные. Внеэкономические переменные (например, климатические условия, социальное положение, пол, возрастная категория) входят в систему только как экзогенные переменные. В качестве экзогенных переменных могут

рассматриваться значения эндогенных переменных за предшествующий период времени (*лаговые переменные*). Так потребление текущего года y_t может зависеть не только от ряда экономических факторов, но и от уровня потребления в предыдущем году (y_{t-1}).

Структурная форма модели в право части содержит при эндогенных и экзогенных переменных коэффициенты b_i и a_j (b_i – коэффициент при эндогенной модели переменной, a_j – коэффициент при экзогенной переменной), которые называются *структурные коэффициенты модели*.

Все переменные в модели выражены в отклонениях от среднего уровня, т.е. под x подразумевается $x - -x$, а под y - соответственно $y - - y$. Поэтому свободный член в каждом уравнении системы отсутствует.

Использование МНК для оценивания структурных коэффициентов модели дает. как принято считать в теории, смещенные и несостоятельные оценки. Поэтому обычно для определения структурных коэффициентов модели структурная форма модели преобразуется в приведенную форму модели.

Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{pmatrix} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \cdots + \delta_{1m}x_m \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \cdots + \delta_{2m}x_m \\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots \\ y_n = \delta_{n1}x_1 + \delta_{n2}x_2 + \cdots + \delta_{nm}x_m \end{pmatrix}$$

где δ_i– коэффициенты приведенной формы модели.

По виду приведенная форма модели ничем не отличается от системы независимых уравнений, параметры которой оцениваются традиционным методом наименьших квадратов. Применяя МНК, можно оценить δ, а затем оценить значения эндогенных переменных через экзогенные. Коэффициенты

приведенной формы модели представляют собой *нелинейные функции коэффициентов структурной формы модели*.

Приведенная форма модели хотя и позволяет получить значения эндогенной переменной через значения экзогенных переменных, аналитически уступает структурной форме модели, так как в ей отсутствуют оценки взаимосвязи между эндогенными переменными.

Проблема идентификации

При переходе от приведенной формы модели к структурной возникает проблема идентификации. *Идентификация* – это единственность соответствия между приведенной и структурной формами модели.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- идентифицируемые;
- неидентифицируемые;
- сверхидентифицируемые.

Модель идентифицируема, если все ее структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели, и модель идентифицируема.

Модель неидентифицируема, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель сверхидентифицируема, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых необходимо проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо.Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Для того, чтобы уравнение было идентифицируемо, нужно, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в j-м уравнении системы через H , а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, – через D , то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

$D + 1 = H$ – уравнение идентифицируемо;

$D + 1 < H$ – уравнение неидентифицируемо;

$D + 1 > H$ – уравнение сверхидентифицируемо.

Предположим, рассматривается следующая система одновременных уравнений:

$$\begin{pmatrix} \widehat{y_1} = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 \\ \widehat{y_2} = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ \widehat{y_3} = b_{31}y_1 + b_{32}y_2 + a_{33}x_3 + a_{34}x_4 \end{pmatrix}$$

Первое уравнение точно идентифицируемо, ибо в нем присутствуют три эндогенные переменные – y_1, y_2, y_3, т. е. Н= 3, и две экзогенные переменные – x_1 и x_2 , число отсутствующих экзогенных переменных равно двум – x_3 и x_4, D=2. Тогда имеем равенство: D + 1 = Н, т. е. 2 + 1 = 3, что означает наличие идентифицируемого уравнения.

Во втором уравнении системы Н= 2 (y_1 и y_2) и D = 1 (x_4). Равенство D+1=Н, т.е. 1 + 1 = 2. Уравнение идентифицируемо.

В третьем уравнении системы Н = 3 (y_1, y_2, y_3), а D=2 (x_1 и x_2). Следовательно, по счетному правилу D+1=Н, и это уравнение идентифицируемо. Таким образом, система (5.6) в целом идентифицируема.

Предположим, что в рассматриваемой модели $a_{21} = 0$ и $a_{33} = 0$. Тогда система примет вид:

$$\begin{pmatrix} \widehat{y_1} = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 \\ \widehat{y_2} = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 \\ \widehat{y_3} = b_{31}y_1 + b_{32}y_2 + a_{34}x_4 \end{pmatrix}$$

Первое уравнение этой системы не изменилось. Система по-прежнему содержит три эндогенные и четыре экзогенные переменные, поэтому для него D = 2 при Н = 3, и оно, как и в предыдущей системе, идентифицируемо. Второе уравнение имеет Н = 2 и D = 2 (x_1, x_4), так как 2 + 1 > 2. Данное уравнение сверхидентифицируемо. Также сверхидентифицируемым оказывается и третье

уравнение системы, где H = 3 (y_1, y_2, y_3) и D = 3 (x_1, x_2, x_3), т.е. счетное правило составляет неравенство: 3 + 1 > 3 или D + 1 >H. Модель в целом является сверхидентифицируемой.

Предположим, что последнее уравнение системы с тремя эндогенными переменными имеет вид:

$$\widehat{y_3} = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + a_{34}x_4$$

т. е. в отличие от предыдущего уравнения в него включены еще две экзогенные переменные, участвующие в системе, – x_1 и x_2. В этом случае уравнение становится неидентифицируемым, ибо при H = 3, D = 1 (отсутствует только x_3) и D + 1 < H, 1 + 1 < 3. Итак, несмотря на то, что первое уравнение идентифицируемо, второе сверхидентифицируемо, вся модель считается неидентифицируемой и не имеет статистического решения.

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели.

Достаточное условие идентификации – определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен 0 и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Целесообразность проверки условия идентификации модели через определитель матрицы коэффициентов, отсутствующих в данном уравнении, но присутствующих в других уравнениях, объясняется тем, что возможна

ситуация, когда для каждого уравнения системы выполнено счетное правило, а определитель матрицы названных коэффициентов равен нулю. В этом случае

Оценивание параметров структурной модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение в литературе получили следующие методы оценивания коэффициентов структурной модели:

- косвенный метод наименьших квадратов (КМНК);
- двухшаговый метод наименьших квадратов (ДМНК);
- трехшаговый метод наименьших квадратов (ТМНК);
- метод максимального правдоподобия с полной информацией (ММПу);
- метод максимального правдоподобия при ограниченной информации (ММП).

Косвенный метод наименьших квадратов применяется для идентифицируемой системы одновременных уравнений, а *двухшаговый метод наименьших квадратов* – для оценки коэффициентов сверхидентифицируемой модели. Перечисленные методы оценивания также используются для сверхидентифицируемых систем уравнений.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к достаточно сложным вычислительным процедурам. Поэтому в качестве модификации используется метод максимального правдоподобия при ограниченной информации (метод наименьшего дисперсионного отношения), разработанный в 1949 г. Т. Андерсоном и Н.

Рубиным.

В отличие от метода максимального правдоподобия в данном методе сняты ограничения на параметры, связанные с функционированием системы в целом. Это делает решение более простым, но трудоемкость вычислений остается достаточно высокой. Несмотря на его популярность, к середине 1960-х годов он был практически вытеснен двухшаговым методом наименьших квадратов в связи с гораздо большей простотой последнего.

Дальнейшим развитием двухшагового метода наименьших квадратов является трехшаговый МНК (ТМНК), предложенный в 1962 г. А. Зельнером и Г. Тейлом. Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным оказывается ДМНК.

Косвенный метод наименьших квадратов используется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы:

- структурная модель преобразовывается в приведенную форму модели;
- для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты (δ_{ij});
- коэффициенты приведенной формы путем алгебраических преобразований трансформируются в параметры структурной модели.

Если система сверхидентифицируема, то КМНК не используется, ибо он не дает однозначных оценок для параметров структурной модели. В этом случае могут применяться разные методы оценивания, среди которых наиболее распространенным и простым является ***двухшаговый метод наименьших квадратов***:

Основная идея ДМНК – на основе приведенной формы модели получить

для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения. Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название «двухшаговый метод наименьших квадратов», ибо МНК используется дважды:

- на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной $\widehat{y_i} = \delta_{i1}x_1 + \delta_{i2}x_2 + \cdots + \delta_{ij}x_j$

- и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

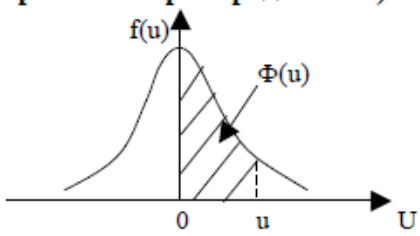
- все уравнения системы сверхидентифицируемы;
- система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Приложение 1

Функция Лапласа (стандартизированное нормальное распределение)

$$\Phi(u) = \frac{1}{\sqrt{2p}} \int_0^u e^{\frac{t^2}{2}}\, dt$$



Пример:
Φ(1.65) = P(0 ≤ U ≤ 1.65) = 0.4505;
P(U > 1.65) = 0.0495.

u	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

3.1 .49903 **3.2** .49931 **3.3** .49952 **3.4** .49966 **3.5** .49977 **3.6** .49984 **3.7** .49989 **3.8** .49993 **3.9** .49995
4.0 .499968
4.5 .49999
5.0 .49999997

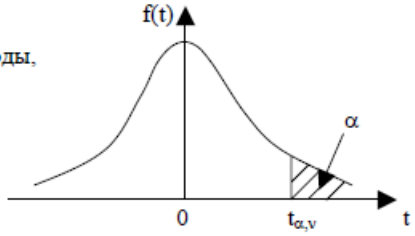
Приложение 2

Распределение Стьюдента (t-распределение)

Пример: $t_{\alpha,\nu} = t_{0.05;20} = 1.725$; ν – число степеней свободы,

$P(T > 1.725) = 0.05$; α – уровень значимости.

$P(|T| > 1.725) = 0.10$.



ν \ α	0.4	0.25	0.10	0.05	0.025	0.01	0.005	0.001	.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.31	636.6
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.6
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.214	12.94
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.859
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.405
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.256	0.684	1.314	1.703	2.050	2.473	2.771	3.421	3.690
28	0.256	0.683	1.313	1.701	2.080	2.467	2.763	3.408	3.674
29	0.256	0.683	1.311	1.699	2.450	2.462	2.756	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.385	2.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.255	0.680	1.296	1.676	2.009	2.403	2.678	3.262	3.495
60	0.255	0.679	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.254	0.679	1.292	1.664	1.990	2.374	2.639	3.195	3.415
100	0.254	0.678	1.290	1.660	1.984	2.365	2.626	3.174	3.389
120	0.254	0.677	1.289	1.658	1.980	2.358	2.467	3.160	3.366
200	0.254	0.676	1.286	1.653	1.972	2.345	2.601	3.131	3.339
500	0.253	0.675	1.283	1.648	1.965	2.334	2.586	3.106	3.310
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.090	3.291

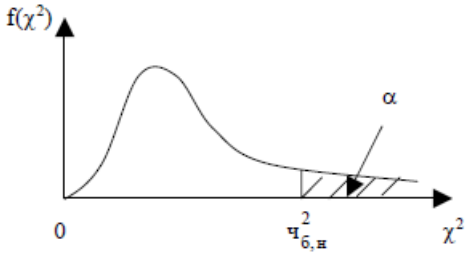
Приложение 3

χ^2-распределение

Пример:

при $\nu = 15$ $P(\chi^2 > 8.55) = 0.9$,
$P(\chi^2 > 22.31) = 0.1$;

при $\nu > 100$ $\sqrt{2\text{ч}^2} - \sqrt{2\text{н} - 1} = U$ $(U \in N(0,1))$.



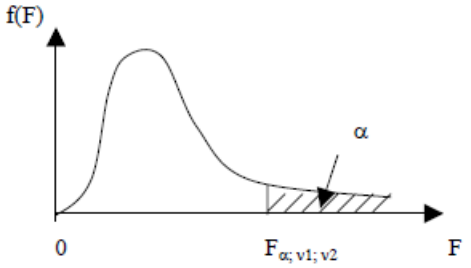
ν \ α	.995	.990	.975	.950	.900	.750	.500	.250	.100	.050	.025	.010	.005
1	$4\cdot10^{-6}$	$2\cdot10^{-5}$	10^{-5}	$4\cdot10^{-4}$	016	101	454	1.32	2.71	3.84	5.02	6.63	7.88
2	.010	.020	.051	.103	.211	.58	1.39	2.77	4.61	5.99	7.38	9.21	10.60
3	.072	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.22	13.37	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.36	24.74	27.69	29.19
14	4.07	4.66	5.63	6.57	7.79	10.1	13.34	17.12	21.06	23.69	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.54	26.30	28.85	32.00	34.27
17	5.68	6.41	7.56	8.67	10.09	12.79	16.34	20.49	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.60	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.72	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.88	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	16.34	20.34	24.93	29.61	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	17.24	21.34	26.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	18.14	22.34	27.14	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	19.04	23.34	28.24	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	19.94	24.34	29.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	20.84	25.34	30.43	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	21.78	26.34	31.53	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	22.66	27.34	32.62	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	23.57	28.34	33.71	39.09	42.56	45.72	49.59	52.34
30	13.78	14.95	16.79	18.49	20.60	24.48	29.34	34.80	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	33.66	39.34	45.62	51.81	55.76	59.34	63.69	66.77
50	27.99	29.70	32.36	34.76	37.69	42.94	49.33	56.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	52.29	59.33	66.98	74.38	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	61.70	69.33	77.58	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	71.14	79.33	88.13	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	80.62	89.33	98.65	107.6	113.1	118.1	124.1	128.3
100	67.32	70.06	74.22	77.93	82.36	90.13	99.33	109.1	118.5	124.3	129.6	135.8	140.2

Приложение 4а

Распределение Фишера (F-распределение)

Пример:

при $\nu_1 = 6$, $\nu_2 = 5$ $P(F > 3.40) = 0.1$;
при $\nu_1 = 6$, $\nu_2 = 5$ $P(F > 4.95) = 0.05$;
при $\nu_1 = 6$, $\nu_2 = 5$ $P(F > 10.7) = 0.01$.



		ν_1 (число степеней свободы)											
ν_2	α	1	2	3	4	5	6	7	8	9	10	11	12
1	.10	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7
	.05	161	200	216	225	230	234	237	239	241	242	243	244
2	.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41
	.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
	.01	98.5	99.2	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
3	.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22
	.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
	.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1
4	.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90
	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
	.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4
5	.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27
	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
	.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89
6	.10	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90
	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
	.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	.10	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67
	.05	5.59	4.74	4.35	4.12	3.97	3.87.	3.79	3.73	3.68	3.64	3.60	3.57
	.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47
8	.10	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50
	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
	.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67
9	.10	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38
	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
	.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11
10	.10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28
	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
	.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71
11	.10	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21
	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40

Приложение 4б

Распределение Фишера (продолжение)

ν_1 (число степеней свободы)													
15	**20**	**24**	**30**	**40**	**50**	**60**	**100**	**120**	**200**	**500**	**∞**	**α**	**ν_2**
61.2	61.7	62.0	62.3	62.5	62.7	62.8	63.0	63.1	63.2	63.3	63.3	.10	**1**
246	248	249	250	251	252	252	253	253	254	254	254	.05	
9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.48	9.49	9.49	9.49	.10	**2**
19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	.05	
99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5.	99.5	995	.01	
5.20	5.18	5.18	5.17	5.16	5.15	5.15	5.14	5.14	5.14	5.14	5.13	.10	**3**
8.70	8.66	8.64	8.62	8.59	8.58	8.57	8.55	8.55	8.54	8.53	8.53	.05	
26.9	26.7	26.6	26.5	26.4	26.4	26.3	26.2	26.2	26:2	26.1	26.1	.01	
3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.78	3.77	3.76	3.76	.10	**4**
5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.66	5.65	5.64	5.63	.05	
14.2	14.0	13.9	13.8	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5	.01	
3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11	3.10	.10	**5**
4.62	4.56	4.53	4.50	4.46	4.44	4.43	4.41	4.40	4.39	4.37	4.36	.05	
9.72	9.55	9.47	9.38	9.29	9.24	9.20	9.13	9.11	9.08	9.04.	9.02	.01	
2.87	2.84	2.82	2.80	2.78	2.77	2.76	2.75	2.74	2.73	2.73	2.72	.10	**6**
3.94	3.87	3.84	3.81	3.77	3.75	3.74	3.71	3.70	3.69	3.68	3.67	.05	
7.56	7.40	7.31	7.23	7.14	7.09	7.06	6.99	6.97	6.93	6.90	6.88	.01	
2.63	2.59	2.58	2.56	2.54	2.52	2.51	2.50	2.49	2.48	2.48	2.47	.10	**7**
3.51	3.44	3.41	3.38	3.34	3.32	3.30	3.27	3.27	3.25	3.24	3.23	.05	
6.31	6.16	6.07	5.99	5.91	5.86	5.82	5.75	5.74	5.70	5.67	5.65	.01	
2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.32	2.31	2.30	2.29	.10	**8**
3.22	3.15	3.12	3.08	3.04	2.02	3.01	2.97	2.97	2.95	2.94	2.93	.05	
5.52	5.36	5.28	5.20	5.12	5.07	5.03	4.96	4.95	4.91	4.88	4.86	.01	
2.34	2.30	2.28	2.25	2.23	2.22	2.21	2.19	2.18	2.17	2.17	2.16	.10	**9**
3.01	2.94	2.90	2.86	2.83	2.80	2.79	2.76	2.75	2.73	2.72	2.71	.05	
4.96	4.81	4.73	4.65	4.57	4.52	4.48	4.42	4.40	4.36	4.33	4.31	.01	
2.24	2.20	2.18	2.16	2.13	2.12	2.11	2.09	2.08	2.07	2.06	2.06	.10	**10**
2.85	2.77	2.74	2.70	2.66	2.64	2.62	2.59	2.58	2.56	2.55	2.54	.05	
4.56	4.41	4.33	4.25	4.17	4.12	4.08	4.01	4.00	3.96	3.93	3.91	.01	
2.17	2.12	2.10	2.08	2.05	2.04	2.03	2.00	2.00	1.99	1.98	1.97	.10	**11**
2.72	2.65	2.61	2.57	2.53	2.51	2.49	2.46	2.45	2.43	2.42	2.40	.05	
4.25	4.10	4.02	3.94	3.86	3.81	3.78	3.71	3.69	3.66	3.62	3.60	.01	

Приложение 4в

Распределение Фишера (продолжение)

ν_2	α	ν_1 (число степеней свободы) 1	2	3	4	5	6	7	8	9	10	11	12
12	.10	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15
	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16
13	.10	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10
	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	.10	3.10	2.73	2.52	239	2.31	2.24	2.19	2.15	2.12	2.10	2.08	2.05
	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02
	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	.10	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99
	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55
17	.10	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96
	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
	.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46
18	.10	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.96	1.93
	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37
19	.10	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.94	1.91
	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	.10	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.92	1.89
	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23
22	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
30	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84

Приложение 4г

Распределение Фишера (продолжение)

ν_1 (число степеней свободы)													
15	**20**	**24**	**30**	**40**	**50**	**60**	**100**	**120**	**200**	**500**	**∞**	**α**	**ν_2**
2.10	2.06	2.04	2.01	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.90	.10	**12**
2.62	2.54	2.51	2.47	2.43	2.40	2.38	2.35	2.34	2.32	2.31	2.30	.05	
4.01	3.86	3.78	3.70	3.62	3.57	3.54	3.47	3.45	3.41	3.38	3.36	.01	
2.05	2.01	1.98	1.96	1.93	1.92	1.90	1.88	1.88	1.86	1.85	1.85	.10	**13**
2.53	2.46	2.42	2.38	2.34	2.31	2.30	2.26	2.25	2.23	2.22	2.21	.05	
3.82	3.66	3.59	3.51	3.43	3.38	3.34	3.27	3.25	3.22	3.19	3.17	.01	
2.01	1.96	1.94	1.91	1.89	1.87	1.86	1.83	1.83	1.82	1.80	1.80	.10	**14**
2.46	2.39	2.35	2.31	2.27	2.24	2.22	2.19	2.18	2.16	2.14	2.13	.05	
3.66	3.51	3.43	3.35	3.27	3.22	3.18	3.11	3.09	3.06	3.03	3.00	.01	
1.97	1.92	1.90	1.87	1.85	1.83	1.82	1.79	1.79	1.77	1.76	1.76	.10	**15**
2.40	2.33	2.29	2.25	2.20	2.18	2.16	2.12	2.11	2.10	2.08	2.07	.05	
3.52	3.37	3.29	3.21	3.13	3.08	3.05	2.98	2.96	2.92	2.89	2.87	.01	
1.94	1.89	1.87	1.84	1.81	1.79	1.78	1.76	1.75	1.74	1.73	1.72	.10	**16**
2.35	2.28	2.24	2.19	2.15	2.12	2.11	2.07	2.06	2.04	2.02	2.01	.05	
3.41	3.26	3.18	3.10	3.02	2.97	2.93	2.86	2.84	2.81	2.78	2.75	.01	
1.91	1.86	1.84	1.81	1.78	1.76	1.75	1.73	1.72	1.71	1.69	1.69	.10	**17**
2.31	2.23	2.19	2.15	2.10	2.08	2.06	2.02	2.01	1.99	1.97	1.96	.05	
3.31	3.16	3.08	3.00	2.92	2.87	2.83	2.76	2.75	2.71	2.68	2.65	.01	
1.89	1.84	1.81	1.78	1.75	1.74	1.72	1.70	1.69	1.68	1.67	1.66	.10	**18**
2.27	2.19	2.15	2.11	2.06	2.04	2.02	1.98	1.97	1.95	1.93	1.92	.05	
3.23	3.08	3.00	2.92	2.84	2.78	2.75	2.68	2.66	2.62	2.59	2.57	.01	
1.86	1.81	1.79	1.76	1.73	1.71	1.70	1.67	1.67	1.65	1.64	1.63	.10	**19**
2.23	2.16	2.11	2.07	2.03	2.00	1.98	1.94	1.93	1.91	1.89	1.88	.05	
3.15	3.00	2.92	2.84	2.76	2.71	2.67	2.60	2.58	2.55	2.51	2.49	.01	
1.84	1.79	1.77	1.74	1.71	1.69	1.68	1.65	1.64	1.63	1.62	1.61	.10	**20**
2.20	2.12	2.08	2.04	1.99	1.97	1.95	1.91	1.90	1.88	1.86	1.84	.05	
3.09	2.94	2.86	2.78	2.69	2.64	2.61	2.54	2.52	2.48	2.44	2.42	.01	
1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.61	1.60	1.39	1.58	1.37	.10	**22**
2.15	2.07	2.03	1.98	1.94	1.91	1.89	1.85	1.84	1.82	1.80	1.78	.05	
2.98	2.83	2.75	2.67	2.58	2.53	2.50	2.42	2.40	2.36	2.33	2.31	.01	
1.78	1.73	1.70	1.67	1.64	1.62	1.61	1.58	1.57	1.56	1.54	1.53	.10	**24**
2.11	2.03	1.98	1.94	1.89	1.86	1.84	1.80	1.79	1.77	1.75	1.73	.05	
2.89	2.74	2.66	2.58	2.49	2.44	2.40	2.33	2.31	2.27	2.24	2.21	.01	
1.76	1.71	1.68	1.65	1.61	1.59	1.58	1.35	1.54	1.53	1.51	1.50	.10	**26**
2.07	1.99	1.95	1.90	1.85	1.82	1.80	1.76	1.75	1.73	1.71	1.69	.05	
2.81	2.66	2.58	2.50	2.42	2.36	2.33	2.25	2.23	2.19	2.16	2.13	.01	
1.74	1.69	1.66	1.63	1.59	1.57	1.56	1.53	1.52	1.50	1.49	1.48	.10	**28**
2.04	1.96	1.91	1.87	1.82	1.79	1.77	1.73	1.71	1.69	1.67	1.65	.05	
2.75	2.60	2.52	2.44	2.35	2.30	2.26	2.19	2.17	2.13	2.09	2.06	.01	
1.72	1.67	1.64	1.61	1.57	1.55	1.54	1.51	1.50	1.48	1.47	1.46	.10	**30**
2.01	1.93	1.89	1.84	1.79	1.76	1.74	1.70	1.68	1.66	1.64	1.62	.05	
2.70	2.55	2.47	2.39	2.30	2.25	2.21	2.13	2.11	2.07	2.03	2.01	.01	

Приложение 4д

Распределение Фишера (продолжение)

		ν_1 (число степеней свободы)											
ν_2	α	1	2	3	4	5	6	7	8	9	10	11	12
40	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
80	.01	2.77	2.37	2.16	2.02	1.93	1.85	1.80	1.75	1.72	1.69	1.65	1.63
	.05	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88
	.01	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41
100	.10	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.70	1.67	1.63	1.61
	.05	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85
	.01	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36
120	.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
∞	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

Приложение 4е

Распределение Фишера (продолжение)

ν_1 (число степеней свободы)													
15	**20**	**24**	**30**	**40**	**50**	**60**	**100**	**120**	**200**	**500**	∞	**α**	**ν_2**
1.66	1.61	1.57	1.54	1.51	1.48	1.47	1.43	1.42	1.41	1.39	1.38	.10	**40**
1.92	1.84	1.79	1.74	1.69	1.66	1.64	1.59	1.58	1.55	1.53	1.51	.05	
2.52	2.37	2.29	2.20	2.11	2.06	2.02	1.94	1.92	1.87	1.83	1.80	.01	
1.60	1.54	1.51	1.48	1.44	1.41	1.40	1.36	1.35	1.33	1.31	1.29	.10	**60**
1.84	1.75	1.70	1.65	1.59	1.56	1.53	1.48	1.47	1.44	1.41	1.39	.05	
2.35	2.20	2.12	2.03	1.94	1.88	1.84	1.75	1.73	1.68	1.63	1.60	.01	
1.58	1.52	1.49	1.45	1.41	1.38	1.36	1.31	1.31	1.29	1.27	1.25	.10	**80**
1.77	1.70	1.65	1.60	1.54	1.51	1.47	1.42	1.40	1.38	1.34	1.32	.05	
2.24	2.11	2.03	1.94	1.84	1.78	1.76	1.65	1.63	1.57	1.52	1.49	.01	
1.56	1.50	1.47	1.43	1.39	1.36	1.34	1.29	1.28	1.26	1.24	1.22	.10	**100**
1.75	1.68	1.63	1.57	1.51	1.48	1.45	1.39	1.38	1.34	1.30	1.28	.05	
2.19	2.06	1.98	1.89	1.79	1.73	1.70	1.59	1.57	1.51	1.46	1.43	.01	
1.55	1.48	1.45	1.41	1.37	1.34	1.32	1.27	1.26	1.24	1.21	1.19	.10	**120**
1.75	1.66	1.61	1.55	1.50	1.46	1.43	1.37	1.35	1.32	1.28	1.25	.05	
2.19	2.03	1.95	1.86	1.76	1.70	1.66	1.56	1.53	1.48	1.42	1.38	.01	
1.52	1.46	1.42	1.38	1.34	1.31	1.28	1.24	1.22	1.20	1.17	1.14	.10	**200**
1.72	1.62	1.57	1.52	1.46	1.41	1.39	1.32	1.29	1.26	1.22	1.19	.05	
2.13	1.97	1.89	1.79	1.69	1.63	1.58	1.48	1.44	1.39	1.33	1.28	.01	
1.49	1.42	1.38	1.34	1.30	1.26	1.24	1.18	1.17	1.13	1.08	1.00	.10	∞
1.67	1.57	1.52	1.46	1.39	1.35	1.32	1.24	1.22	1.17	1.11	1.00	.05	
2.04	1.88	1.79	1.70	1.59	1.52	1.47	1.36	1.32	1.25	1.15	1.00	.01	

Приложение 5а

Распределение Дарбина–Уотсона

Критические точки d_l и d_u при уровне значимости $\alpha = 0.05$
(n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6		m = 7		m = 8		m = 9	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0.610	1.400																
7	0.700	1.356	0.467	1.896														
8	0.763	1.332	0.359	1.777	0.368	2.287												
9	0.824	1.320	0.629	1.699	0.435	2.128	0.296	2.388										
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822								
11	0.927	1.324	0.658	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005						
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832	0.171	3.149				
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692	0.230	2.985	0.147	3266		
14	1.045	1.330	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472	0.343	2.727	0.251	2.979	0.175	3.216
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	1537	0.356	2.757	0.272	2.975
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.667	0.321	2.873
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162	0.595	2.339	0.502	2.521	0.416	2.704
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124	0.637	2.290	0.547	2.460	0.461	2.633
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.751	2.174	0.666	2.318	0.584	2.464
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012	0.784	2.144	0.702	2.280	0.621	2.419
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958	0.874	2.071	0.798	2.188	0.723	2.309
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891	1.015	1.979	0.950	2.069	0.885	2.162
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877	1.053	1.957	0.991	2.041	0.930	2127
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1837	1.369	1.873	1.337	1.910
75	1.598	1.65	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.847	1.608	1.862
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863

Приложение 5б

Распределение Дарбина–Уотсона

Критические точки d_l и d_u при уровне значимости $\alpha = 0.01$
(n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6		m = 7		m = 8		m = 9	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0.390	1.142																
7	0.433	1.036	0.294	1.676														
8	0.497	1.003	0.343	1.489	0.229	2.102												
9	0.554	0.998	0.408	1.389	0.279	1.873	0.183	2.433										
10	0.604	1.001	0.466	1.333	0.340	1.733	0.230	2.193	0.130	2.690								
11	0.633	1.010	0.319	1.297	0.396	1.640	0.286	2.030	0.193	2.433	0.124	2.892						
12	0.697	1.023	0.369	1.274	0.449	1.373	0.339	1.913	0.244	2.280	0.164	2.663	0.103	3.033				
13	0.738	1.038	0.616	1.261	0.499	1.326	0.391	1.826	0.294	2.130	0.211	2.490	0.140	2.838	0.090	3.182		
14	0.776	1.034	0.660	1.234	0.347	1.490	0.441	1.737	0.343	2.049	0.237	2.334	0.183	2.667	0.122	2.981	0.078	3.287
15	0.811	1.070	0.700	1.232	0.391	1.464	0.488	1.704	0.391	1.967	0.303	2.244	0.226	2.330	0.161	2.817	0.107	3.101
16	0.844	1.086	0.737	1.232	0.633	1.446	0.332	1.663	0.437	1.900	0.349	2.133	0.269	2.416	0.200	2.681	0.142	2.944
17	0.874	1.102	0.772	1.233	0.672	1.432	0.374	1.630	0.480	1.847	0.393	2.078	0.313	2.319	0.241	2.366	0.179	2.811
18	0.902	1.118	0.803	1.239	0.708	1.422	0.613	1.604	0.322	1.803	0.433	2.013	0.333	2.238	0.282	2.467	0.216	2.697
19	0.928	1.132	0.833	1.263	0.742	1.413	0.630	1.384	0.361	1.767	0.476	1.963	0.396	2.169	0.322	2.381	0.233	2.397
20	0.932	1.147	0.863	1.271	0.773	1.411	0.683	1.367	0.398	1.737	0.313	1.918	0.436	2.110	0.362	2.308	0.294	2.310
21	0.973	1.161	0.890	1.277	0.803	1.408	0.718	1.334	0.633	1.712	0.332	1.881	0.474	2.039	0.400	2.244	0.331	2.434
22	0.997	1.174	0.914	1.284	0.831	1.407	0.748	1.343	0.667	1.691	0.387	1.849	0.310	2.013	0.437	2.188	0.368	2.367
23	1.018	1.187	0.938	1.291	0.838	1.407	0.777	1.334	0.698	1.673	0.620	1.821	0.343	1.977	0.473	2.140	0.404	2.308
24	1.037	1.199	0.960	1.298	0.882	1.407	0.803	1.328	0.728	1.638	0.632	1.797	0.378	1.944	0.307	2.097	0.439	2.233
25	1.033	1.211	0.981	1.303	0.906	1.409	0.831	1.323	0.736	1.643	0.682	1.776	0.610	1.913	0.340	2.039	0.473	2.209
26	1.072	1.222	1.001	1.312	0.928	1.411	0.833	1.318	0.783	1.633	0.711	1.739	0.640	1.889	0.372	2.026	0.303	2.168
27	1.089	1.233	1.019	1.319	0.949	1.413	0.878	1.313	0.808	1.626	0.738	1.743	0.669	1.867	0.602	1.997	0.336	2.131
28	1.104	1.244	1.037	1.323	0.969	1.413	0.900	1.313	0.832	1.618	0.764	1.729	0.696	1.847	0.630	1.970	0.366	2.098
29	1.119	1.234	1.034	1.332	0.988	1.418	0.921	1.312	0.833	1.611	0.788	1.718	0.723	1.830	0.638	1.947	0.393	2.068
30	1.133	1.263	1.070	1.339	1.006	1.421	0.941	1.311	0.877	1.606	0.812	1.707	0.748	1.814	0.684	1.923	0.622	2.041
31	1.147	1.273	1.083	1.343	1.023	1.423	0.960	1.310	0.897	1.601	0.834	1.698	0.772	1.800	0.710	1.906	0.649	2.017
32	1.160	1.282	1.100	1.332	1.040	1.428	0.979	1.310	0.917	1.397	0.836	1.690	0.794	1.788	0.734	1.889	0.674	1.993
33	1.172	1.291	1.114	1.338	1.033	1.432	0.996	1.310	0.936	1.394	0.876	1.683	0.816	1.776	0.737	1.874	0.698	1.973
34	1.184	1.299	1.128	1.364	1.070	1.433	1.012	1.311	0.934	1.391	0.896	1.677	0.837	1.766	0.779	1.860	0.722	1.937
35	1.193	1.307	1.140	1.370	1.083	1.439	1.028	1.312	0.971	1.389	0.914	1.671	0.837	1.737	0.800	1.847	0.744	1.940
36	1.206	1.313	1.133	1.376	1.098	1.442	1.043	1.313	0.988	1.388	0.932	1.666	0.877	1.749	0.821	1.836	0.766	1.923
37	1.217	1.323	1.163	1.382	1.112	1.446	1.038	1.314	1.004	1.386	0.930	1.662	0.893	1.742	0.841	1.823	0.787	1.911
38	1.227	1.330	1.176	1.388	1.124	1.449	1.072	1.313	1.019	1.383	0.966	1.638	0.913	1.733	0.860	1.816	0.807	1.899
39	1.237	1.337	1.187	1.393	1.137	1.433	1.083	1.317	1.034	1.384	0.982	1.633	0.930	1.729	0.878	1.807	0.826	1.887
40	1.246	1.344	1.198	1.398	1.148	1.437	1.098	1.318	1.048	1.384	0.997	1.632	0.946	1.724	0.893	1.799	0.844	1.876
45	1.288	1.376	1.243	1.423	1.201	1.474	1.136	1.328	1.111	1.384	1.063	1.643	1.019	1.704	0.974	1.768	0.927	1.834
50	1.324	1.403	1.283	1.446	1.243	1.491	1.203	1.338	1.164	1.387	1.123	1.639	1.081	1.692	1.039	1.748	0.997	1.803
55	1.336	1.427	1.320	1.466	1.284	1.306	1.247	1.348	1.209	1.392	1.172	1.638	1.134	1.683	1.093	1.734	1.037	1.783
60	1.383	1.449	1.330	1.484	1.317	1.320	1.283	1.338	1.249	1.398	1.214	1.639	1.179	1.682	1.144	1.726	1.108	1.771
65	1.407	1.468	1.377	1.300	1.346	1.334	1.313	1.368	1.283	1.604	1.231	1.642	1.218	1.680	1.186	1.720	1.133	1.761
70	1.429	1.483	1.400	1.313	1.372	1.346	1.343	1.378	1.313	1.611	1.283	1.643	1.233	1.680	1.223	1.716	1.192	1.734
75	1.448	1.301	1.422	1.329	1.393	1.337	1.368	1.387	1.340	1.617	1.313	1.649	1.284	1.682	1.236	1.714	1.227	1.748
80	1.466	1.313	1.441	1.341	1.416	1.368	1.390	1.393	1.364	1.624	1.338	1.633	1.312	1.683	1.283	1.714	1.239	1.743
85	1.482	1.328	1.438	1.333	1.433	1.378	1.411	1.603	1.386	1.630	1.362	1.637	1.337	1.683	1.312	1.714	1.287	1.743
90	1.496	1.340	1.474	1.363	1.432	1.387	1.429	1.611	1.406	1.636	1.383	1.661	1.360	1.687	1.336	1.714	1.312	1.741
95	1.310	1.332	1.489	1.373	1.468	1.396	1.446	1.618	1.423	1.642	1.403	1.666	1.381	1.690	1.338	1.713	1.336	1.741
100	1.322	1.362	1.303	1.383	1.482	1.604	1.462	1.623	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717	1.337	1.741
150	1.611	1.637	1.398	1.631	1.384	1.663	1.371	1.679	1.337	1.693	1.343	1.708	1.330	1.722	1.313	1.737	1.301	1.732
200	1.664	1.684	1.633	1.693	1.643	1.704	1.633	1.713	1.623	1.723	1.613	1.733	1.603	1.746	1.392	1.737	1.382	1.768

Распределение Дарбина–Уотсона

Критические точки d_l и d_u при уровне значимости $\alpha = 0.05$

(n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6		m = 7		m = 8		m = 9	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0.610	1.400																
7	0.700	1.356	0.467	1.896														
8	0.763	1.332	0.359	1.777	0.368	2.287												
9	0.824	1.320	0.629	1.699	0.435	2.128	0.296	2.388										
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822								
11	0.927	1.324	0.658	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005						
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832	0.171	3.149				
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692	0.230	2.985	0.147	3266		
14	1.045	1.330	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472	0.343	2.727	0.251	2.979	0.175	3.216
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.667	0.321	2.873
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162	0.595	2.339	0.502	2.521	0.416	2.704
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124	0.637	2.290	0.547	2.460	0.461	2.633
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.751	2.174	0.666	2.318	0.584	2.464
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012	0.784	2.144	0.702	2.280	0.621	2.419
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958	0.874	2.071	0.798	2.188	0.723	2.309
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6		m = 7		m = 8		m = 9	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891	1.015	1.979	0.950	2.069	0.885	2.162
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877	1.053	1.957	0.991	2.041	0.930	2.127
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939
□ **65**	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923
□ **70**	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.837	1.369	1.873	1.337	1.910
□ **75**	1.598	1.65	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901
□ **80**	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893
□ **85**	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886
□ **90**	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881
□ **95**	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877
□**100**	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874
□**150**	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.847	1.608	1.862
□**200**	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863
□																		

Распределение Дарбина–Уотсона

Критические точки d_l и d_u при уровне значимости $\alpha = 0.01$

(n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6		m = 7		m = 8		m = 9	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0.390	1.142																
7	0.433	1.036	0.294	1.676														
8	0.497	1.003	0.343	1.489	0.229	2.102												
9	0.554	0.998	0.408	1.389	0.279	1.873	0.183	2.433										
10	0.604	1.001	0.466	1.333	0.340	1.733	0.230	2.193	0.130	2.690								
11	0.633	1.010	0.319	1.297	0.396	1.640	0.286	2.030	0.193	2.433	0.124	2.892						
12	0.697	1.023	0.369	1.274	0.449	1.373	0.339	1.913	0.244	2.280	0.164	2.663	0.103	3.033				
13	0.738	1.038	0.616	1.261	0.499	1.326	0.391	1.826	0.294	2.130	0.211	2.490	0.140	2.838	0.090	3.182		
14	0.776	1.034	0.660	1.234	0.347	1.490	0.441	1.737	0.343	2.049	0.237	2.334	0.183	2.667	0.122	2.981	0.078	3.287
15	0.811	1.070	0.700	1.232	0.391	1.464	0.488	1.704	0.391	1.967	0.303	2.244	0.226	2.330	0.161	2.817	0.107	3.101
16	0.844	1.086	0.737	1.232	0.633	1.446	0.332	1.663	0.437	1.900	0.349	2.133	0.269	2.416	0.200	2.681	0.142	2.944
17	0.874	1.102	0.772	1.233	0.672	1.432	0.374	1.630	0.480	1.847	0.393	2.078	0.313	2.319	0.241	2.366	0.179	2.811
18	0.902	1.118	0.803	1.239	0.708	1.422	0.613	1.604	0.322	1.803	0.433	2.013	0.333	2.238	0.282	2.467	0.216	2.697
19	0.928	1.132	0.833	1.263	0.742	1.413	0.630	1.384	0.361	1.767	0.476	1.963	0.396	2.169	0.322	2.381	0.233	2.397
20	0.932	1.147	0.863	1.271	0.773	1.411	0.683	1.367	0.398	1.737	0.313	1.918	0.436	2.110	0.362	2.308	0.294	2.310
21	0.973	1.161	0.890	1.277	0.803	1.408	0.718	1.334	0.633	1.712	0.332	1.881	0.474	2.039	0.400	2.244	0.331	2.434
22	0.997	1.174	0.914	1.284	0.831	1.407	0.748	1.343	0.667	1.691	0.387	1.849	0.310	2.013	0.437	2.188	0.368	2.367
23	1.018	1.187	0.938	1.291	0.838	1.407	0.777	1.334	0.698	1.673	0.620	1.821	0.343	1.977	0.473	2.140	0.404	2.308
24	1.037	1.199	0.960	1.298	0.882	1.407	0.803	1.328	0.728	1.638	0.632	1.797	0.378	1.944	0.307	2.097	0.439	2.233
25	1.033	1.211	0.981	1.303	0.906	1.409	0.831	1.323	0.736	1.643	0.682	1.776	0.610	1.913	0.340	2.039	0.473	2.209
26	1.072	1.222	1.001	1.312	0.928	1.411	0.833	1.318	0.783	1.633	0.711	1 .739	0.640	1.889	0.372	2.026	0.303	2.168
27	1.089	1.233	1.019	1.319	0.949	1.413	0.878	1.313	0.808	1.626	0.738	1.743	0.669	1.867	0.602	1.997	0.336	2.131
28	1.104	1.244	1.037	1.323	0.969	1.413	0.900	1.313	0 .832	1.618	0.764	1.729	0.696	1.847	0.630	1.970	0.366	2.098
29	1.119	1.234	1.034	1.332	0.988	1.418	0.921	1.312	0.833	1 .611	0.788	1.718	0.723	1.830	0.638	1.947	0.393	2.068
30	1.133	1.263	1.070	1.339	1.006	1.421	0.941	1.311	0.877	1.606	0.812	1.707	0.748	1.814	0.684	1.923	0.622	2.041
31	1.147	1.273	1.083	1.343	1.023	1.423	0.960	1.310	0.897	1.601	0.834	1.698	0.772	1.800	0.710	1.906	0.649	2.017
32	1.160	1.282	1.100	1.332	1.040	1.428	0.979	1.310	0.917	1.397	0.836	1.690	0.794	1.788	0.734	1.889	0.674	1.993

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6		m = 7		m = 8		m = 9	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
33	1.172	1.291	1.114	1.338	1.033	1.432	0.996	1.310	0.936	1.394	0.876	1.683	0.816	1.776	0.737	1.874	0.698	1.973
34	1.184	1.299	1.128	1.364	1.070	1.433	1.012	1.311	0.934	1.391	0.896	1.677	0.837	1.766	0.779	1.860	0.722	1.937
35	1.193	1.307	1.140	1.370	1.083	1.439	1.028	1.312	0.971	1.389	0.914	1.671	0.837	1.737	0.800	1.847	0.744	1.940
36	1.206	1.313	1.133	1.376	1.098	1.442	1.043	1.313	0.988	1.388	0.932	1.666	0.877	1.749	0.821	1.836	0.766	1.923
37	1.217	1.323	1.163	1.382	1.112	1.446	1.038	1.314	1.004	1.386	0.930	1.662	0.893	1.742	0.841	1.823	0.787	1.911
38	1.227	1.330	1.176	1.388	1.124	1.449	1.072	1.313	1.019	1.383	0.966	1.638	0.913	1.733	0.860	1.816	0.807	1.899
39	1.237	1.337	1.187	1.393	1.137	1.433	1.083	1.317	1.034	1.384	0.982	1.633	0.930	1.729	0.878	1.807	0.826	1.887
40	1.246	1.344	1.198	1.398	1.148	1.437	1.098	1.318	1.048	1.384	0.997	1.632	0.946	1.724	0.893	1.799	0.844	1.876
45	1.288	1.376	1.243	1.423	1.201	1.474	1.136	1.328	1.111	1.384	1.063	1.643	1.019	1.704	0.974	1.768	0.927	1.834
50	1.324	1.403	1.283	1.446	1.243	1.491	1.203	1.338	1.164	1.387	1.123	1.639	1.081	1.692	1.039	1.748	0.997	1.803
55	1.336	1.427	1.320	1.466	1.284	1.306	1.247	1.348	1.209	1.392	1.172	1.638	1.134	1.683	1.093	1.734	1.037	1.783
60	1.383	1.449	1.330	1.484	1.317	1.320	1.283	1.338	1.249	1.398	1.214	1.639	1.179	1.682	1.144	1.726	1.108	1.771
65	1.407	1.468	1.377	1.300	1.346	1.334	1.313	1.368	1.283	1.604	1.231	1.642	1.218	1.680	1.186	1.720	1.133	1.761
70	1.429	1.483	1.400	1.313	1.372	1.346	1.343	1.378	1.313	1.611	1.283	1.643	1.233	1.680	1.223	1.716	1.192	1.734
75	1.448	1.301	1.422	1.329	1.393	1.337	1.368	1.387	1.340	1.617	1.313	1.649	1.284	1.682	1.236	1.714	1.227	1.748
80	1.466	1.313	1.441	1.341	1.416	1.368	1.390	1.393	1.364	1.624	1.338	1.633	1.312	1.683	1.283	1.714	1.239	1.743
85	1.482	1.328	1.438	1.333	1.433	1.378	1.411	1.603	1.386	1.630	1.362	1.637	1.337	1.683	1.312	1.714	1.287	1.743
90	1.496	1.340	1.474	1.363	1.432	1.387	1.429	1.611	1.406	1.636	1.383	1.661	1.360	1.687	1.336	1.714	1.312	1.741
95	1 .310	1.332	1.489	1.373	1.468	1.396	1.446	1.618	1.423	1.642	1.403	1.666	1.381	1.690	1.338	1.713	1.336	1.741
100	1. 322	1.362	1.303	1.383	1.482	1.604	1.462	1.623	1.441	1.647	1.421	1.670	1.400	1.693	1.378	1.717	1.337	1.741
150	1.6 11	1.637	1.398	1.631	1.384	1.663	1.371	1.679	1.337	1.693	1.343	1.708	1.330	1.722	1.313	1.737	1.301	1.732
200	1.664	1.684	1.633	1.693	1.643	1.704	1.633	1.713	1.623	1.723	1.613	1.733	1.603	1.746	1.392	1.737	1.382	1.768