АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ИНСТИТУТ ЭКОНОМИКИ И ИНФОРМАТИКИ»

Е.Н. Надеждин, Е.Е.Смирнова

ЭКОНОМЕТРИКА

Учебное пособие

Тула – 2011

УДК 519.862.6.: 519.233.5
Н 171

Рецензенты: доктор физ.-мат. наук, профессор С.Д. Двоенко,
доктор техн. наук, профессор А.А. Говоров .

Эконометрика: учебное пособие / Е.Н.Надеждин, Е.Е.Смирнова; Под ред. Е.Н.Надеждина.- Тула: АНО ВПО «ИЭУ», 2011.- 176 с.

В учебном пособии нашли отражение вопросы теории и практики эконометрики в объёме, рекомендуемом Государственным образовательным стандартом высшего профессионального образования по специальностям 080105 «Финансы и кредит» и 080801 «Прикладная информатика в экономике». Основное внимание уделено изложению особенностей применения математических методов и современных программных средств для анализа экономических процессов. Пособие предназначено для студентов дневной и заочной формы обучения.

СОДЕРЖАНИЕ:

ВВЕДЕНИЕ

В настоящее время в связи с возрастанием интереса молодёжи к освоению экономических специальностей ощущается острая нехватка учебников и учебных пособий, в которых на доступном уровне рассматриваются вопросы теории и практики эконометрики. Быстрое развитие и освоение вычислительной техники и информационных технологий создаёт объективные предпосылки для формализованного решения традиционно трудных эконометрических задач.

Задачами учебного пособия являются: систематическое изложение базовых понятий и предмета эконометрики, формулировка типовых задач эконометрики и обоснование методических рекомендаций по их решению с использованием ЭВМ. Авторы учебного пособия попытались в некоторой степени восполнить имеющийся пробел в учебной литературе по проблеме идентификации результатов экономических наблюдений на основе применения унифицированных программных средств. Учебное пособие разработано в соответствии с требованиями Государственного образовательного стандарта по дисциплине «Эконометрика» для экономических специальностей вузов. При подготовке материалов авторы полагали, что студенты старших курсов владеют понятийным аппаратом теории вероятностей и математической статистики в объёме курса математики экономического вуза и имеют практические навыки работы с офисными программами WINDOWS.

Учебное пособие содержит пять разделов и четыре приложения. В **первом разделе** изложены теоретические предпосылки эконометрического исследования. Даны краткие сведения из области математической статистики, которая составляет основу математического инструментария эконометрики.

Во **втором разделе** рассмотрены классические модели парной регрессии. На примере линейной формы модели парной регрессии показаны особенности приложений регрессионного анализа к задачам идентификации результатов наблюдений. **Третий раздел** посвящён вопросам построения и анализа моделей множественной регрессии. Для двухфакторной модели получены аналитические соотношения, которые могут служить основой для расчёта и оценки коэффициентов регрессии.

В **четвёртом разделе** изучены особенности использования эконометрических уравнений для анализа взаимосвязи экономических явлений. **Пятый раздел** посвящён изложению вопросов анализа временных рядов в интересах постановки и компьютерного решения задач определения статистических характеристик процессов и задач экономического прогнозирования.

В приложениях представлены: глоссарий, краткие биографические данные об учёных, работы которых определили основные направления эконометрики, фрагменты компьютерных программ применительно к ППП MathCad 2001 и статистические данные, необходимые для решения задач.

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЭКОНОМЕТРИКИ

1.1. Предмет и задачи эконометрики

Термин «***Эконометрика***» /эконометрия/ был впервые введён бухгалтером П.Цьемпой (Австро-Венгрия, 1910 г.). **П.Цьемпа** считал, что если к данным бухгалтерского учёта применить методы алгебры и геометрии, то будет получено новое, более глубокое представление о результатах хозяйственной деятельности. Это употребление термина, как и сама концепция, не прижились, но название «эконометрика» оказалось весьма удачным и было использовано позже (1930 г.) для определения нового направления в экономической науке.

Слово «Эконометрика» представляет собой комбинацию двух слов «экономика» и «метрика» (от греч. «метрон»). Таким образом, сам термин подчёркивает специфику, содержание эконометрики как науки, дающей количественное выражение связей и соотношений, которые раскрыты и обоснованы экономической теорией. Можно считать эконометрику наукой об измерении и анализе экономических явлений.

Зарождение эконометрики можно рассматривать как результат ускорения научно-технического прогресса в первой половине XX-го века, значительного усложнения управления экономическими процессами и необходимости междисциплинарного подхода к изучению экономики. Эта наука возникла в результате взаимодействия и объединения трёх составляющих: экономической теории, статистических и математических методов.

29 декабря 1930 г. по инициативе И.Фишера, Р.Фриша, Я.Тинбергена и других известных учёных на заседании Американской ассоциации развития науки (США, г. Кливленд, штат Огайо) было создано первое эконометрическое общество. Новая наука получила название «эконометрика». Уже в 1950 г. общество насчитывало более 1000 членов. С 1933 г. в США издаётся журнал «Эконометрика» («Econometrica»), котрый и сейчас играет важную роль в развитии эконометрической науки. В 1941 году был опубликован первый учебник по эконометрике, автором которого стал Я.Тинберген.

В эти годы вплоть до 70-х годов 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.

В 70-80-е годы XX-го века на развитие теории и методов эконометрики значительное влияние оказали успехи в области статистического анализа временных рядов. В 1970 г. Г.Бокс и Г.Дженкикинс создали ARIMA-модель, К.Симе и другие учёные – VAR-модели, ставшие популярными в начале 80-х гг. С появление современных ЭВМ с развитым программным обеспечением открылись новые возможности для использования методов и моделей эконометрики.

Свидетельством всемирного признания эконометрики является присуждение во второй половине XX-го века четырёх нобелевских премий по экономике за разработки в этой области:

1969 г. – **Р.Фриш** и **Я.Тинберген** – за разработку математических методов анализа экономических процессов;

1980 г. – **Л.Клейн** – за создание эконометрических моделей и их применение к анализу экономических колебаний и экономической политике;

1989 г. – **Т.Хаавельмо** – за прояснение вероятностных основ эконометрики и анализ одновременных экономических структур;

2000 г.- **Дж.Хекман** - за развитие теории и методов анализа селективных выборок; **Д.Макфадден** – за развитие теории и методов анализа моделей дискретного выбора.

В журнале «Эконометрика», основанном в 1933 г. Р.Фришем (1895-1973), дано 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 трёх отправных точек – статистика, экономическая теория и математика – необходимое, но недостаточное условие для понимания количественных соотношений в современной экономической жизни. Это единство всех трёх составляющих. И это единство образует эконометрику» [17, с.32].

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

С учётом различных точек зрения ведущих учёных-экономистов эконометрику можно считать наукой о связях экономических явлений. *Центральной проблемой* эконометрики является построение эконометрической модели и определение возможностей её эффективного использования для описания, анализа и прогнозирования реальных экономических процессов. Основные определения эконометрики имеют вид:

1. *Эконометрика* - это наука, которая даёт количественное выражение взаимосвязей экономических явлений и процессов.
2. *Эконометрика* - это наука, которая занимается определением и анализом наблюдаемых в экономической жизни конкретных количественных закономерностей на основе применения методов математической статистики.

1.2. Особенности эконометрического метода

Становление и развитие эконометрики происходили на основе известных в математической статистике методах парной и множественной регрессии, парной, частной и множественной корреляции, выделения тренда и других компонент временного ряда, классических методов статистического оценивания.

Сущность эконометрического метода составляют методы статистического анализа эмпирических экономических данных в интересах идентификации эконометрических моделей экономических процессов.

Развитие эконометрического метода происходило в результате поиска и обоснования подходов к преодолению ряда трудностей, проявляющихся при исследовании экономических явлений:

- асимметричность связей;
- мультиколлинеарность объясняющих переменных;
- закрытость механизма связи между переменными в изолированной регрессии;
- эффект гетероскедастичности, т.е. нарушение нормального распределения остатков для регрессионной функции;
- автокорреляция;
- ложная корреляция;
- наличие лагов.

Эконометрическое исследование включает в себя решение следующих проблем:

- качественный анализ связей экономических переменных – выделение зависимых (y_j) и независимых (объясняющих) переменных (x_k);
- подбор данных;
- спецификация формы связи между переменными y и x_k;
- оценка параметров модели;
- проверка ряда гипотез о свойствах распределения вероятностей для случайной компоненты (гипотезы о средней, дисперсии и ковариации);
- анализ мультиколлинеарности объясняющих переменных, оценка её статистической значимости, выделение переменных, ответственных за мультиколлинеарность;
- введение фиктивных переменных;
- выявление автокорреляции, лагов;
- выявление тренда, циклической и случайной компонент;
- проверка остатков на гетероскедастичность;
- анализ структуры связей и построение системы одновременных уравнений;
- проверка условия идентификации;
- оценивание параметров системы одновременных уравнений (двухшаговый и трёхшаговый методы наименьших квадратов, метод максимального правдоподобия);

- моделирование на основе системы временных рядов: проблемы стационарности и коинтеграции;
- построение рекурсивных моделей, ARIMA- и VAR- моделей;
- проблемы идентификации и оценивания параметров.

Эконометрическая модель, как правило, основана на теоретическом предположении (гипотезе) о круге взаимосвязанных переменных и характере связи между ними. На практике часто приоритет отдаётся качественному анализу. Поэтому основными этапами эконометрического исследования являются:

- постановка проблемы;
- получение данных, анализ их качества;
- спецификация модели;
- оценка параметров;
- интерпретация результатов анализа и формирование рекомендаций.

Любые экономические данные представляют собой количественные характеристики (показатели) каких-либо экономических объектов. Они формируются под воздействием множества факторов, не все из которых доступны внешнему контролю. Неконтролируемые факторы могут принимать случайные значения из некоторого множества значений и тем самым обусловливать случайность данных, которые они определяют. Стохастическая природа экономических статистических данных определяет необходимость применения адекватных им математических моделей и статистических методов для их обработки. Эконометрический инструментарий базируется на методах и моделях прикладной математической статистики.

Напомним, что *математическая статистика* – наука, изучающая методы обработки результатов наблюдений массовых случайных явлений, обладающих статистической устойчивостью, закономерностью, с целью выявления этой закономерности. Выводы о закономерностях, которым подчиняются явления, изучаемые в математической статистике, всегда основываются на ограниченном, выборочном числе наблюдений. При большом числе наблюдений выводы могут оказаться иными. Для вынесения более определённого заключения о закономерностях наблюдаемого явления математическая статистика опирается на теорию вероятностей.

Закономерности в экономике выражаются в виде связей и зависимостей экономических показателей, математических моделей их поведения. Такие зависимости и модели могут быть получены только путём обработки реальных статистических данных, с учётом внутренних механизмов связи и случайных факторов. Модель может быть получена и апробирована на основе анализа статистических данных, и измерения о поведении последних стимулируют дальнейшее уточнение и развитие модели. Особенно важен эконометрический анализ в макроэкономике, где взаимосвязи величин нередко неочевидны и изменчивы. Результаты эконометрического анализа позволяют обосновать и уточнить связи в макроэкономических моделях, лучше понять механизмы взаимосвязи макроэкономических показателей.

Ответственным этапом эконометрических исследований являются экономические измерения, которые предполагают получение, сравнение и упорядочение экономической информации. Специфика экономических измерений определяется следующими особенностями экономических данных:

- наличие большого числа разнородных данных – разнородных ресурсов, результатов (например, товаров и услуг);
- мультипликативность свойств (показателей) экономических систем;
- взаимосвязи между показателями могут носить качественный и неоднозначный характер;
- ограниченный набор данных статистических наблюдений;
- наличие неконтролируемых погрешностей наблюдений;
- трудность выявления эмпирических отношений;
- наличие проблемы обобщения (свёртки и агрегирования) данных для представления ненаблюдаемых (латентных) переменных.

В области экономических измерений центральной является проблема точности, которая связана:

- с определением понятия экономической величины;
- с формированием системы принципов, постулатов и других теоретических положений, формирующих базис точности экономических измерений;
- с определением экономических показателей;
- с выбором принципа конструирования измерителей и измерений;
- с обоснованием выбора шкал при конструировании измерителя;
- с разработкой правил формирования систем показателей;
- с выявлением типов и определением методов устранения ошибок экономического измерения;
- с разработкой правил агрегирования и свёртки экономических показателей;
- с выявлением условий сравнимости экономических величин (показателей);
- с разработкой правил и методов измерений.

Существуют различные методы сбора экономических данных, например, метод опроса, анкетирования и интервьюирования, получения официальной статистической отчётности и т.д. В большинстве стран существуют официальные статистические органы, занимающиеся сбором, обработкой, распространением и публикацией важнейших экономических данных. Этой деятельностью занимаются также многочисленные специализированные государственные и частные агенства.

Традиционной базой данных для эконометрических исследований служат данные официальной статистики либо данные бухгалтерского учёта. Используя экономическую теорию, можно определить связь между признаками и показателями, а используя статистику и учёт – ответить, например, на следующие вопросы. Какие показатели следует применять для измерения результатов работы промышленного предприятия (валовая продукция, добавленная стоимость, реализованная продукция или др.)? Как оценить остатки оборотных средств (по стоимости первых или последних поставок или по средней стоимости)?

Основные источники статистических данных в России можно объединить в две группы: внутренние и внешние источники.

К внутренним источникам относятся те виды и формы статистических наблюдений, которые организует, собирает и разрабатывает Госкомстат России: отчётность предприятий; регистр предприятий; переписи и обследования. *К внешним источникам* относят те виды и формы статистического наблюдения, которые организуют другие ведомства: административные источники; денежная и банковская статистика; платёжный баланс; таможенная статистика.

1.3 Основные положения и определения математической статистики

Математическая статистика изучает массовые случайные явления в природе, обществе и технике методами теории вероятностей. Каждое исследование, проводимое методами теории вероятностей, опирается на эксперимент, на опытные данные, полученные в результате специальных наблюдений или экспериментов. Результаты наблюдений - статистические данные - позволяют выявить закономерности, присущие изучаемому случайному явлению.

Математическая статистика как наука возникла и развивалась параллельно с теорией вероятностей, начиная с XVII века. Своим развитием математическая статистика обязана трудам таких выдающихся математиков прошлого, как К.Гаусс, К.Пирсон, А.Кетле.

Ряд важнейших результатов, полученных в конце XIX и начале XX вв., принадлежат учёным Петербургской математической школы В.Я.Буняковскому, П.Л.Чебышеву, А.М.Ляпунову, А.А.Маркову.

Значительный вклад в математическую статистику сделали советские математики В.И. Романовский, Е.Е.Слуцкий, А.Н. Колмогоров, а также Стьюдент, Р.Фишер, Э. Пирсон и американские – Ю.Нейман, А.Вальд, Р.Кендалл.

Специфическая черта статистики состоит в том, что она рассматривает не одно явление (наблюдение) в отдельности, а их совокупность. Появление или исчезновение из статистической совокупности одной ее единицы (одного явления, наблюдения) не уничтожает совокупность как таковую.

В дальнейшем изложении элементами (единицами) статистической совокупности будут считаться результаты наблюдений, обозначенные через $x_1, x_2, ..., x_N$, где N - число имеющихся наблюдений.

Случайной величиной (переменной) называют переменную, которая под воздействием случайных факторов в результате опыта может принимать то или иное численное значение, причём заранее неизвестно, какое именно. Примем, что случайная величина может принимать те или иные значения из некоторого множества чисел с определёнными вероятностями

Обозначаются случайные величины большими буквами X, Y, Z, … . Значения, которые случайная величина может принимать в опыте, обозначаются малыми буквами x,y,.z. Эти значения являются всегда действительными.

С каждым случайным событием можно связать какую-либо случайную величину, принимающую значения из множества действительных чисел.

Выделяют непрерывные и дискретные случайные величины.

Дискретной случайной величиной называется такая случайная величина, которая принимает конечное или счётное множество различных значений. Напомним, что счётным является множество, элементы которого можно пронумеровать числами натурального ряда. Дискретные случайные величины принимают изолированные значения, которые отделены друг от друга.

ПРИМЕР 1.1. Дискретными являются такие случайные величины, как число дефектных изделий в выборке из 10 изделий: X={0;1;2;…10}; число отказов аппаратуры за период испытаний X={0;1;2;…}; число подбрасываний монеты до первого появления цифры X={0;1;2;3…}.

Непрерывной случайной величиной называется такая случайная величина, которая принимает любые значения на некотором интервале действительной числовой оси.

Другими словами значения непрерывной случайной величины могут заполнять конечный или бесконечный интервал числовой оси. Очевидно, что число возможных значений непрерывной случайной величины всегда бесконечно.

ПРИМЕР 1.2. Интервал времени между заказами клиентов в течение рабочего дня есть непрерывная случайная величина: X=$\{0 < X \le 8{,}0\}$.

Годовой доход на одного члена семьи в Тульском регионе является случайной величиной, которая принимает значения в некотором интервале: X=$\{x_{\min} \le x \le x_{\max}\}$.

Непрерывными случайными величинами являются ошибки наблюдений или измерений, например ошибка при взвешивании товара, ошибка в определении месячной прибыли и т.д.

Дискретные и непрерывные случайные величины не следует противопоставлять друг другу. Наоборот, дискретные случайные величины следует рассматривать как частный случай непрерывной случайных величин.

Для полной характеристики случайной величины необходимо знать не только область её возможных значений, но и то, что как часто эти значения встречаются в опыте при его повторении в одних и тех же условиях.

Пусть дискретная случайная величина принимает в опыте одно из своих возможных значений: $X=x_1$, $X=x_2$, …, $X=x_n$. Появление каждого из этих значений свидетельствует о том, произошло одно из полной группы несовместных событий A_1, A_2,…, A_n - событие, состоящее в том, что случайная величина A_i приняла в опыте значение $X=x_i$, где i=1,2,…,n.

Допустим, что вероятности этих событий известны и равны: $P(X=x_1)=p_1$, $P(X=x_2)= p_2$,…, $P(X=x_n)= p_n$, причём сумма этих вероятностей равна единице, так как события несовместны и образуют полную группу событий

$$\sum_{i=1}^{n} p_i = 1.$$

Задавая значения случайной величины x_i и соответствующие им значения p_i вероятностей появления в опыте, мы даём полное описание случайной величины.

В общем случае под *вероятностью* некоторого события понимается доля числа исходов, благоприятствующих данному событию, в общем числе возможных равновероятных исходов.

Соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями , называется законом распределения вероятностей (ЗРВ) случайной величины или просто *законом распределения* случайной величины.

Про случайную величину, для которой известен закон распределения, говорят, что она подчиняется данному закону. Закон распределения вероятностей случайной величины X представляет собой функцию P{x}, которая как и любая функциональная зависимость может быть представлена в форме таблицы, аналитически в виде формулы или графика.

В табл.1.1 представлен в качестве примера закон распределения вероятностей получения числа очков при бросании игрального кубика. Верхняя строка табл.1. содержит возможные значения случайной величины, а нижняя – соответствующие им вероятности. Такая таблица называется рядом распределения и является простейшей формой записи ЗРВ.

Таблица 1.1

x_i	1	2	3	4	5	6
P_i	1/6	1/6	1/6	1/6	1/6	1/6

Очевидно, что $\sum_{i=1}^{n=6} x_i = 1.$

Закон распределения в виде ряда или многоугольника может быть задан только для дискретной случайной величины. Непрерывная случайная величина принимает бесчисленное множество значений, поэтому построить для неё ряд распределения невозможно. Количественной характеристикой распределения вероятностей непрерывной случайной величины служит функция распределения, которая является исчерпывающей характеристикой, пригодной как для непрерывной, так и для дискретной случайной величины.

Функцией распределения F(x) случайной величины X называется функция, существующая для всех действительных значений случайной величины: $-\infty < x < +\infty$, которая определяет вероятность того, что случайная величина примет значение, меньшее x:

$$F(x) = P(X < x) \quad \forall \quad -\infty < x < \infty .$$

Функцию распределения F(x) называют также интегральной функцией распределения или интегральным законом распределения. Она является универсальной формой закона распределения, исчерпывающим образом характеризующей как дискретные, так и непрерывные случайные величины. Если случайную величину рассматривать как случайную точку X на оси 0x, то F(x) есть вероятность того, что точка X находится левее некоторой точки x на оси 0x.

Для непрерывной случайной величины способ её задания с помощью функции распределения не всегда удобен, так как по функции распределения трудно судить о характере распределения случайной величины в небольшой окрестности той или иной точки числовой оси. Более наглядной в этом смысле является дифференциальная функция распределения, называемая также плотностью вероятности или плотностью распределения.

Плотностью распределения f(x) непрерывной случайной величины называется производная от функции распределения F(x):

$$f(x) = \frac{dF(x)}{dx} = F'(x).$$

Свойства плотности распределения

а) Плотность распределения является неотрицательной функцией

$$f(x) \ge 0 \quad \forall -\infty < x < \infty.$$

2. Интеграл от плотности распределения в интервале от - ∞ до x равен функции распределения случайной величины:

$$\int_{-\infty}^{x} f(x) \cdot dx = F(x).$$

3. Вероятность попадания случайной величины X на интервал [a,b] равна интегралу от плотности распределения, взятому по этому участку:

$$P(a \le X \le b) = \int_{a}^{b} f(x) \cdot dx.$$

4. Интеграл в бесконечных пределах от плотности распределения равен единице: $\int_{-\infty}^{\infty} f(x) \cdot dx = 1$.

В теории вероятностей и в математической статистике для общей характеристики случайной величины используют параметры, называемые *числовыми характеристиками* случайной величины, которые в сжатой форме выражают все существенные особенности распределения. Наиболее часто используются следующие числовые характеристики: математическое ожидание, дисперсия, мода, медиана, моменты распределения.

Математическим ожиданием дискретной случайной величины называется число, равное произведению всех возможных значений случайной величины на вероятности этих значений. Математическое ожидание случайной величины обозначается как M[x]=m_x:

$M[x] = \sum_{i=1}^{\infty} x_i \cdot p_i$ или $M[x] = \sum_{i=1}^{N} x_i \cdot p_i$ (где N – конечное число значений случайной величины).

Математическое ожидание дискретной случайной величины – это взвешенное среднее всех её возможных значений, причём в качестве весового коэффициента берётся вероятность соответствующего исхода.

Математическое ожидание непрерывной случайной величины X, имеющей известную плотность распределения $p(x)$ может быть определено по формуле

$$M[x] = \int_{-\infty}^{\infty} x \cdot p(x) \cdot dx .$$

В широком смысле под *статистикой* понимают процесс собирания и математической обработки результатов наблюдений, сведений о фактах для определения статистики по интересующим событиям. В статистике различают два вида совокупностей - генеральную и выборочную.

Генеральная совокупность - бесконечный набор наблюдений $x_{1,}$ x_2 … изучаемой случайной величины, распределение признаков (результатов наблюдений) в котором совпадает с теоретическим распределением вероятностей. *Выборочная совокупность* (выборка) представляет конечное число наблюдений $x_{1,}$ x_2, ..., x_N и является частью генеральной совокупности.

Число объектов выборке или в генеральной совокупности называется её объёмом.

В статистике принят принцип случайного отбора, который означает, что каждый элемент генеральной совокупности имеет равный шанс попасть в выборку. Выборка составляется случайным образом в сопоставимых условиях наблюдения, при этом элементы выборки должны быть попарно независимыми. Выборка должна быть репрезентативной (представительной), т.е. такой, чтобы по ее свойствам можно было судить о генеральной совокупности.

Репрезентативность выборки обеспечивается случайным отбором, при котором вероятность попасть в выборку для всех её объектов одинакова. Репрезентативность выборки зависит также от количества наблюдений N. Пусть $N_{\text{дост}}$ (N достаточное) - такое минимальное число наблюдений, при котором выборка становится уже репрезентативной. Тогда выборку размера N можно считать репрезентативной при выполнении условия $N \geq N_{\text{дост}}$.

Величина $N_{\text{дост}}$ зависит от точности наблюдений ε, от стандартного отклонения S_1 и от вероятности P, с которой она определяется.

В прямом смысле *статистикой* называется любой параметр, зависящий от $x_{1,}$ x_2, ..., $x_{N.}$ Например, среднее значение от результата N наблюдений

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

или второй центральный момент

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 ,$$

или стандартное отклонение

$$S_1 = \sqrt{\frac{N}{N-1}S} .$$

Из приведенного определения следует, что каждая статистика зависит от количества и результатов наблюдений. Так как в общем случае результаты наблюдений являются случайными значениями определенных параметров исследуемого объекта, то статистика также является случайной величиной, имеющей свое распределение вероятностей.

Статистическая оценка параметров распределения. Свойства оценок.

В эконометрике часто осуществляется обработка статистических совокупностей небольших объёмов (N<100). При этом, не ставя перед собой задачу нахождения закона распределения вероятностей, можно ограничиться определением важнейших числовых характеристик случайной величины. В том случае, когда вид закона распределения известен, с помощью найденных статистических характеристик можно определить приближённые значения параметров ЗРВ.

Любое значение параметра, найденное с помощью статистических характеристик распределения, будет приближённой оценкой параметра, содержащей элемент случайности в силу ограниченности объёма наблюдений. Такое приближение называют оценкой параметра.

В задачу математической статистики входит разработка методов, позволяющих сделать вывод о числовых параметрах распределения генеральной совокупности на основании результатов выборочных наблюдений. Пусть имеется случайная величина X, закон распределения которой содержит неизвестный параметр a. Требуется найти приближённое значение a по результатам N независимых опытов, в которых случайная величина X приняла ряд значений $X_1, X_2, \ldots, X_N$.

Оценка $\widehat{a}$ есть функция величин $X_1, X_2, \ldots, X_N$. Найти оценку неизвестного параметра распределения – это получить функцию от наблюдавшихся значений случайной величины, по которой можно всегда установить приближённое числовое значение параметра. Оценка $\widehat{a}$ сама является случайной величиной, её закон распределения зависит от вида ЗРВ величины X и числа опытов N.

Для того, чтобы оценка $\widehat{a}$ давала хорошее приближение искомого параметра а, она должна быть несмещённой, состоятельной и эффективной.

Несмещённость. Оценку $\widehat{a}$ называют несмещённой, если её математическое ожидание равно истинному значению параметра a: $M(\widehat{a}) = a$.

Если оценка смещена, то значение параметра a получается либо завышенным ($M(\widehat{a}) > a$), либо заниженным ($M(\widehat{a}) < a$), что приводит к систематическим ошибкам в оценке параметра.

Состоятельность. Оценку $\widehat{a}$ называют состоятельной, если она стремится по вероятности к параметру a:

$$\lim_{n\to\infty} P(|a - \widehat{a}| < \varepsilon) = 1.$$

Другими словами, состоятельная оценка подчиняется закону больших чисел.

Эффективной называют оценку $\hat{a}$, которая при заданном объёме выборки N имеет наименьшую дисперсию: $D(\hat{a}_{эф}) = \min$.

Асимптотически эффективной называют оценку $\hat{a}$, дисперсия которой при N→ ∞ приближается к дисперсии эффективной оценки, т.е. к минимально возможной.

Кроме размерности N выборка также характеризуется числом степеней свободы К, равным N без числа используемых статистик. На практике чаще используются только две статистики - $\overline{x}$ и $D_x = S^2$ (или S_1). Поэтому в этих случаях К = N-2.

Для определения величины $N_{дост}$ воспользуемся таблицей, входными параметрами для которой являются вероятность Р и $g = \frac{\varepsilon}{S_1}$. Для оперативного определения $N_{дост}$ при Р≥ 0,95 можно использовать табл. 1.2.

Таблица 1.2

g	0,1	0,15	0,2	0,25	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
P	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,95	0,96	0,97	0,97
K	2,0	9,0	55	38	27	118	13	10	8	8	8	7
g	1,5	2,0	2,5	3,0								
P	0,98	0,99	0,995	0,999								
K	6	6	6	6								

ПРИМЕР 1.3. Пусть Р = 0,95, S_1= 0,25, ε = 0,1. Тогда $g = \frac{\varepsilon}{S_1} = \frac{0{,}1}{0{,}25} = 0{,}4$, а К= 8. Следовательно, $N_{дост}$ = К + 2 = 18 + 2 = 20, и поэтому в рассматриваемом случае размерность выборки N должна быть не меньше 20 (N > $N_{дост}$= 20).

В тех случаях, когда размерность выборки достаточно большая (N > 20), считают, что значения статистики S_1 распределены по нормальному закону. Тогда принимается

$$N_{дост} \geq \frac{x_p^2 \cdot S_1^2}{\varepsilon^2 \cdot \bar{x}^2},$$

где x_p— параметр нормального распределения вероятностей,

$$\boxed{\Phi(x_p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{p} e^{-\frac{x^2}{2}} dx}.$$

Значение x_p определяется по таблице функции распределения Ф(х) в зависимости от вероятности Р. Для оперативного определения значения x_p при Р > 0,9 можно использовать табл. 1.3.

Таблица 1.3

Ф(х)	0,9	0,95	0,99
x_p	1,653	1,96	2,576

1.4 Сущность метода наименьших квадратов

Рассмотрим основные идеи основные идеи регрессионного эксперимента на примере линейной модели парной регрессии.

Пусть х и у – одномерные переменные, а функция регрессии у=f(x,C) имеет вид:

$$y_i = a + b_i \cdot x_i + \varepsilon_i\,, \quad i = 1,...,N\,. \tag{1.1}$$

Здесь $x_1, x_2,\ldots, x_N$ - значения фактора; $y_1, y_2,\ldots, y_N$ - наблюдаемые значения зависимой переменной (показателя) у; $\varepsilon_1, \varepsilon_2,\ldots,\varepsilon_N$ – независимые (ненаблюдаемые) одинаково распределённые случайные величины.

При решении рассмотренной задачи используются два основных подхода: непараметрический и гауссовский, которые отличаются характером предположений относительно закона распределения случайных величин ε. Примем в качестве рабочей гауссовскую модель простой линейной регрессии. В ней предполагается, что величины ε_i распределены по нормальному закону с некоторой неизвестной дисперсией σ^2.

При выборе метода определения параметров регрессионной модели у=f(x,C) можно руководствоваться различными подходами. Один из наиболее распространённых состоит в том, что при «хорошем» выборе векторной оценки $\hat{C}$ параметров модели С=(a,b) величины $q_i= y_i- f(x_i, C)$ должны быть в совокупности близки к нулю. В случае линейной парной регрессии $q_i= y_i-(a+b\cdot x_i)=0$. Меру близости совокупности этих величин, называемых остатками, к нулю можно выбирать по-разному (например, максимум модулей, сумму модулей и т.п.). Наиболее простые формулы расчёта можно получить, если в качестве такой меры выбрать *сумму квадратов*:

$$\Delta S^{*}=\Delta y_1^{2}+\Delta y_2^{2}+\ldots+\Delta y_N=\sum_i \Delta y_i^{2}\ . \qquad (1.2)$$

Один из способов построения линии регрессии состоит в минимизации суммы квадратов остатков ΔS^* и выборе параметров уравнения регрессии, отвечающих этому требованию.

Определение. *Методом наименьших квадратов* (МНК) называется способ подбора параметров a и b регрессионной модели исходя из условия минимизации суммы квадратов остатков всех наблюдений.

Метод наименьших квадратов сам по себе не связан с какими-либо предположениями о распределении случайных ошибок $\varepsilon_1, \varepsilon_2,\ldots,\varepsilon_N$, он может применяться тогда и только тогда, когда мы не считаем эти ошибки случайными. В качестве примера укажем задачи сглаживания экспериментальных данных. Далее будем рассматривать МНК применительно к гауссовской модели. Это обусловлено следующими положениями: именно в гауссовской модели метод наименьших квадратов обладает определёнными свойствами оптимальности; в гауссовской модели получаемые на основе МНК оценки неизвестных параметров обладают конкретными статистическими свойствами.

Оценки метода наименьших квадратов

Применительно к линейной модели парной регрессии (см. разд.2.1) МНК заключается в выборе таких коэффициентов a и b, которые обеспечивают наименьшее значение суммы квадратов

$$\Delta S^{*}=\sum_{i=1}^{n}[y_i-(a+b\cdot x_i)]^2\ . \qquad (1.3)$$

Чтобы упростить дальнейшие формулы, перепишем соотношение (1.3) в виде

$$y_i = a + b(x_i - \bar{x}) + \varepsilon_i \qquad i = 1,\ldots,n \qquad (1.3,a)$$

где $\bar{x}=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i$, $a = A + b\cdot\bar{x}$. Этот переход означает перенос начала отсчета на оси абсцисс в точку $\bar{x}$, которая служит центром совокупности (выборки) $x_1, \ldots, x_n$. Для нахождения оценок по методу наименьших квадратов нам надо выяснить, при каких (a , b) достигается минимум выражения

$$\sum_{i=1}^{n}[y_i - a - b(x_i - \bar{x})]^2. \qquad (1.4)$$

Приравнивая нулю частные производные по a и b выражения (1.4), получим систему уравнений относительно неизвестных a и b:

$$\begin{cases}\sum_{i=1}^{n}[y_i - a - b(x_i - \bar{x})] = 0\\ \sum_{i=1}^{n}(x_i - \bar{x})[y_i - a - b(x_i - \bar{x})] = 0\end{cases}$$

Ее решение (*a, b)* легко найти:

$$\widehat{a} = \bar{y} \quad (\textit{где } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i)\,; \tag{1.5}$$

$$\widehat{b} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\cdot(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\quad . \tag{1.6}$$

Величины $\widehat{a}$ *и* $\widehat{b}$ представляют собой полученные по методу наименьших квадратов оценки неизвестных коэффициентов *a* и *b*.

Обобщённый метод наименьших квадратов

В случаях нарушения гомоскедастичности и при наличии автокорреляции ошибок наблюдений на практике применяют обобщенный метод наименьших квадратов – метод GLS (Generalizel Least Squares).

Обобщенный метод наименьших квадратов (ОМНК) применяется к преобразованным данным и позволяет получать оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии. Рассмотрим использование обобщенного МНК для корректировки гетероскедастичности.

Примем допущение, что среднее значение остаточных величин равно нулю. Предположим также, что дисперсия остаточных величин зависит от значений фактора и может быть отражена зависимостью

$$\sigma^2_{\varepsilon i} = \sigma^2 \cdot K_i\,,$$

где $\sigma^2_{\varepsilon i}$ - дисперсия ошибки при конкретном i-м значении фактора;

σ^2 - постоянная дисперсия ошибки при соблюдении предпосылки о гомоскедастичности остатков;

K_i - коэффициент пропорциональности, изменяющийся в зависимости от величины фактора, что обуславливает неоднородность дисперсии.

При этом предполагается, что дисперсия σ^2 неизвестна, а в отношении величины *K* выдвигаются определенные гипотезы, характеризующие структуру гетероскедастичности.

В общем виде для уравнения

$$y_i = a + b\cdot x_i + \varepsilon_i \quad \text{при} \quad \sigma^2_{\varepsilon i} = \sigma^2 \cdot \mathrm{K_i}, \quad \mathrm{i} = \overline{1,N}$$

модель парной регрессии примет вид:

$$y_i = \alpha + \beta_i\cdot x_i + \alpha_i\cdot\varepsilon_i, \qquad \alpha_\mathrm{i} = \sqrt{K_i}\;.$$

Предполагая в них отсутствие автокорреляции, перейдем к уравнению с гомоскедастичными остатками. Для этого все переменные, зафиксированные в процессе *i*-го наблюдения, разделим на $\alpha_i=\sqrt{K_i}$. Тогда дисперсия остатков будет величиной постоянной, т.е. $\sigma^2_{\varepsilon i}=\sigma^2$.

Другими словами, в результате перехода от регрессии y(x) к регрессии на новых переменных $y^*=f(x^*)$, где $y^*=y/\sqrt{K}$; $x^*=x/\sqrt{K}$, уравнение регрессии примет вид:

$$y_i^*=\alpha/\alpha_I+\beta\cdot x_i^*+\varepsilon_I, \quad i=\overline{1,N}$$

Исходные данные для данного уравнения будут иметь вид:

$$y^T=(y_1/\alpha_1;y_2/\alpha_2;...;y_N/\alpha_N);$$

$$x^T=(x_1/\alpha_1;x_2/\alpha_2;...;x_N/\alpha_N).$$

По отношению к обычной регрессии уравнение с новыми, преобразованными переменными представляет собой взвешенную регрессию, в которой переменные *y* и *x* взяты с весами $\alpha_i=(\sqrt{K_i})^{-1}$.

Оценка параметров нового уравнения с преобразованными переменными приводит к взвешенному методу наименьших квадратов, для которого необходимо минимизировать сумму квадратов отклонений вида

$$S=\sum\alpha_i^2\cdot(y_i-a-b\cdot z_i)^2.$$

Соответственно получим следующую систему нормальных уравнений:

$$\begin{cases}\sum\alpha_i^2\cdot y_i=\alpha\cdot\sum\alpha_i^2+b\cdot\sum\alpha_i^2\cdot x_i\\ \\ \sum\alpha_i^2\cdot y_i\cdot x_i=\alpha\cdot\sum\alpha_i^2\cdot x_i+b\cdot\sum\alpha_i^2\cdot x_i^2\end{cases}.$$

Если преобразованные переменные *x* и *y* взять в отклонениях от средних уровней, то коэффициент регрессии *b* можно определить в виде

$$b=\frac{\sum(\alpha^2\cdot x\cdot y)}{\sum\alpha^2\cdot x^2}.$$

Укажем, что при обычном применении МНК к уравнению линейной регрессии для переменных в отклонениях от средних уровней коэффициент регрессии *b* определяется по формуле

$$b=\frac{\sum(x\cdot y)}{\sum x^2}.$$

Таким образом, при использовании обобщенного МНК с целью корректировки гетероскедастичности коэффициент регрессии *b* представляет собой взвешенную сумму по отношению к обычному МНК с весами α^2.

ПРИМЕР 1.4. Пусть при исследовании зависимости сбережений *y* от дохода *x* по первоначальным данным было получено уравнение регрессии

y=a+bx, a= -1,081; b=0,1178

Применяя обобщенный МНК к данной модели в предположении, что ошибки пропорциональны доходу, было получено уравнение для преобразованных данных:

$$\frac{y}{x} = 0{,}1026 - 0{,}8538 \cdot \frac{1}{x} \cdot$$

Коэффициент регрессии 0,1178 первого уравнения сравнивают со свободным членом 0,1026 второго уравнения – оценки параметра в зависимости сбережений от дохода.

Переход к относительным величинам существенно снижает вариацию фактора и соответственно уменьшает дисперсию ошибки. Он представляет собой наиболее простой случай учета гетероскедастичности в регрессионных моделях с помощью обобщенного МНК. Во многих случаях применение обобщенного МНК позволяет получить оценки параметров модели, обладающие меньшей дисперсией.

1.5 Регрессионный и корреляционный анализ экономических процессов

Для количественного описания взаимосвязей между экономическими переменными эконометрика использует математический аппарат регрессионного и корреляционного анализа.

Регрессионный анализ - раздел математической статистики, главная направленность которого состоит в выводе на основании соответствующих выборочных совокупностей уравнения регрессии, устанавливающего связь между значениями зависимой переменной (результирующим показателем) и значениями независимых параметров.

Совокупность методов, определяющих тесноту связи между $\tilde{y}$ и x_j, $j=\overline{1,n}$ составляет другой раздел математической статистики - *корреляционный анализ*. Если связь между переменными y и x является нефункциональной, установлена на основании совместного анализа соответствующих им выборок $y_1,y_2,...,y_N$ и $x_1,x_2,…,x_N$, то считается, что между ними существует корреляционная связь.

В регрессионном анализе методы корреляционного анализа используются в основном для спецификации: определения (выбора) наиболее соответствующего вида зависимости (модели) из числа рассматриваемых (линейной, степенной, параболической, логарифмической и др.) между переменными y и x_j , $j=\overline{1,n}$.

При парной регрессии устанавливается форма линии связи между двумя случайными величинами y и x на основании случайных выборок $x_1,x_2,…,x_N$ и $y_1,y_2,...,y_N$, все элементы (результаты наблюдений) которых получены в сопоставимых условиях. Результаты наблюдений в каждой выборке являются взаимно независимыми и каждому x_i соответствует y_i.

Одна из случайных величин (y) считается зависимой, а другая (x) - независимой.

Уравнение линии регрессии (линии связи) при парной регрессии записывается в виде:

$$\tilde{y} = f(x) . \qquad (1.7)$$

Уравнение (1.7) означает, что между случайными величинами у и х имеется корреляционная (вероятностная) связь.

Если при функциональной зависимости у = f(x) одному значению независимой переменной х обычно соответствует только одно значение зависимой переменной у, то при корреляционной зависимости каждому значению х может соответствовать сколь угодно большое число значений у. Поэтому при корреляционной зависимости изменение значения х вызовет изменение не конкретного значения у, а среднего значения у, и это изменение будет тем больше, чем теснее у и х будут корреляционно зависимы. Методика определения тесноты связи между коррелированными величинами изучается в рамках **корреляционного анализа**.

Для определения корреляционной зависимости необходимо располагать числовыми данными об экономическом процессе. Такие данные определяют путём статистических наблюдений или исследований. Полученный числовой материал группируют в ряды распределения по факториальному x и результативному y признакам (составляют корреляционную матрицу с двумя входами).

При наличии единственного факториального признака разработка гипотезы о форме связи его с функцией упрощается, поскольку данную зависимость можно показать в виде графика.

2 МОДЕЛИ ПАРНОЙ РЕГРЕССИИ

2.1 Спецификация моделей парной регрессии

Регрессионную связь между зависимой переменной и факторами, в отличие от функциональной, будем записывать в виде:

$$\tilde{y} = f(x_1, x_2, ..., x_n), \qquad (2.1)$$

где $\tilde{y}$ - результирующий показатель;

x_j - j-й ($j=\overline{1,n}$) независимый параметр (фактор, воздействующий на результирующий показатель).

Регрессия называется **парной**, если на показатель y действует только один фактор (n=1), и множественной, если число факторов, воздействующих на y, более одного (n > 1).

Во многих случаях форма парной регрессии устанавливается путем выбора из заданного множества стандартных форм зависимостей, к числу которых прежде всего относятся:

- линейная зависимость

$$\tilde{y} = a + bx; \qquad (2.2)$$

- гиперболическая зависимость

$$\tilde{y} = a + {}^{b}\!/_{x}; \qquad (2.3)$$

- степенная зависимость

$$\tilde{y} = a \cdot x^b; \qquad (2.4)$$

- логарифмическая зависимость

$$\tilde{y} = a + b \cdot \ln x; \qquad (2.5)$$

- параболическая зависимость

$$\tilde{y} = a + b \cdot x + c \cdot x^2 + d \cdot x^3 + ...; \qquad (2.6)$$

- тригонометрическая зависимость

$$\boxed{\tilde{y} = a_0 + \sum_{k=1}^{m}(a_k \cdot \cos kx + b_k \cdot \sin kx)}, \qquad (2.7)$$

где m - число гармоник;

a_0, a_k, b_k - неизвестные коэффициенты линии регрессии.

Для определения значений неизвестных параметров a и b при линейной модели (2.2) и параболической зависимости (2.6) в общем случае применяют метод наименьших квадратов.

Рассмотрим основные идеи основные идеи регрессионного эксперимента на примере линейной модели парной регрессии.

Пусть x и y – одномерные переменные, а функция регрессии y=f(x,C) имеет вид:

$$y_i = a + b \cdot x_i + \varepsilon_i, \quad i=1,\dots,N. \qquad (2.8)$$

Здесь $x_1, x_2, \dots, x_N$ - значения фактора; $y_1, y_2, \dots, y_N$ - наблюдаемые значения зависимой переменной (показателя) y; $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ – независимые (ненаблюдаемые) одинаково распределённые случайные величины; N - число наблюдений.

При решении рассмотренной задачи используются два основных подхода: непараметрический и гауссовский, которые отличаются характером предположений относительно закона распределения случайных величин ε. Примем в качестве рабочей гауссовскую модель простой линейной регрессии. В ней предполагается, что величины ε_i распределены по нормальному закону с некоторой неизвестной дисперсией σ^2.

При выборе метода определения параметров регрессионной модели y=f(x,C) можно руководствоваться различными подходами. Один из наиболее распространённых состоит в том, что при «хорошем» выборе векторной оценки $\hat{C}$ параметров модели C=(a,b) величины $q_i = y_i - f(x_i, C)$ должны быть в совокупности близки к нулю. В случае линейной парной регрессии должно выполняться условие $q_i = y_i - (a + b \cdot x_i) = 0$. Меру близости совокупности этих величин, называемых *остатками*, к нулю можно выбирать по-разному (например, максимум модулей, сумму модулей и т.п.). Наиболее простые формулы расчёта можно получить, если в качестве такой меры выбрать *сумму квадратов*:

$$\Delta S^* = \Delta y_1^2 + \Delta y_2^2 + \ldots + \Delta y_N = \sum_i \Delta y_i^2 .$$

Один из способов построения линии регрессии состоит в минимизации суммы квадратов остатков ΔS^* и выборе параметров уравнения регрессии, отвечающих этому требованию.

2.2 Методика оценки коэффициентов парной регрессии

Применительно к линейной модели (2.2) парной регрессии МНК заключается в выборе таких коэффициентов a и b, которые обеспечивают наименьшее значение суммы квадратов

$$\Delta S^* = \sum_{i=1}^{n} [y_i - (a + b \cdot x_i)]^2 . \qquad (2.9)$$

Чтобы упростить дальнейшие формулы, перепишем соотношение (2.9) в виде

$$y_i = a + b(x_i - \bar{x}) + \varepsilon_i \qquad i = 1, \ldots, n \qquad (2.10)$$

где $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$, $a = A + b \cdot \bar{x}$. Этот переход означает перенос начала отсчета на оси абсцисс в точку $\bar{x}$, которая служит центром совокупности (выборки) $x_1, \ldots, x_n$. Для нахождения оценок по методу наименьших квадратов нам надо выяснить, при каких (a, b) достигается минимум выражения

$$\sum_{i=1}^{n} [y_i - a - b(x_i - \bar{x})]^2 . \qquad (2.11)$$

Приравнивая нулю частные производные по a и b выражения (2.11), получим систему уравнений относительно неизвестных a и b:

$$\begin{cases} \sum_{i=1}^{n}[y_i - a - b(x_i - \bar{x})] = 0 \\ \sum_{i=1}^{n}(x_i - \bar{x})[y_i - a - b(x_i - \bar{x})] = 0 \end{cases}$$

Её решение (a, b) легко найти:

$$\hat{a} = \bar{y} \quad (где\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i)\,; \tag{2.12}$$

$$\hat{b} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad . \tag{2.13}$$

Величины $\hat{a}$ *и* $\hat{b}$ будут получены далее по методу наименьших квадратов и являются оценками неизвестных нам величин a и b.

Свойства оценок. Естественно, возникает вопрос: как соотносятся полученные значения $\hat{a}$ и $\hat{b}$ с истинными значениями a и b или, другими словами, каково качество оценок метода наименьших квадратов $\hat{a}$ и $\hat{b}$. Для ответа на этот вопрос укажем некоторые *свойства этих оценок*.

1) $M\hat{a} = a$ и $M\hat{b} = b$;

2) $D\hat{a} = \sigma^2 / n,$ и $D\hat{b} = \sigma^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2$;

3) $\text{cov}(\hat{a}, \hat{b}) = 0;$

4) случайные величины $\hat{a}$ и $\hat{b}$ распределены по нормальному закону;

5) $\hat{a}$ и $\hat{b}$ независимы как случайные величины.

Доказательства утверждений 1-3 могут быть получены прямым вычислением, используя выражения (2.12) и (2.13). Покажем, например, что $M\hat{b} = b$.

$$M\,\hat{b} = M\frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})M(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

поскольку величины $x_1, \ ..., \ x_n$ и $\bar{x}$ не случайны содержащие только их выражения можно вынести из-под знака математического ожидания. Далее, поскольку $M\varepsilon_i = 0$ и $M\bar{\varepsilon} = 0$, то

$$M(y_i - \bar{y}) = My_i - M\bar{y} = a + b(x_i - \bar{x}) - a = b(x_i - \bar{x}).$$

Подставляя это выражение в предыдущую формулу, находим, что $M\hat{b} = b$.

Заметим, что свойства 1-3 не используют предположения о нормальном характере ошибок в модели (2.10). Зато свойство 4 верно только в гауссовском случае. Доказательство свойства 4 следует из вида формул (2.12), (2.13), которые по отношению к $y_1, ... \ y_n$ имеют вид линейных функций $f(\cdot)$ линейные комбинации независимых нормальных случайных величин, как мы отмечали ранее, сами распределены нормально.

Свойство 5 есть следствие нормальности ошибок и свойства 3. Независимость оценок $\hat{a}$ *и* $\hat{b}$ заметно упрощает дальнейший анализ. В первую очередь ради этого модель (2.9) была заменена на (2.10).

В совокупности свойства 1-4 дают важные результаты, характеризующие качество оценок $\widehat{a}$ и $\widehat{b}$:

$$\widehat{a} \sim N\left(a, \frac{\sigma^2 n}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right), \quad \widehat{b} \sim N\left(b, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right). \tag{2.14}$$

Оценка дисперсии. В модели (2.10), кроме коэффициентов *a* и *b,* есть еще один неизвестный параметр – дисперсия σ^2 ошибок наблюдения. Этот параметр явно входит в соотношение (2.14) и тем самым влияет на точность оценок. Поэтому σ^2, в свою очередь, требует оценивания. Ключ к этому дает остаточная сумма квадратов

$$\sum_{i=1}^{n}[y_i - \widehat{a} - \widehat{b}(x_i - \bar{x})]^2 . \tag{2.15}$$

Можно доказать, что в гауссовской модели выражение (2.15) является независимой от $\widehat{a}$ и $\widehat{b}$ случайной величиной, имеющей распределение $\sigma^2\chi^2(n\text{-}2)$, где $\chi^2(n\text{-}2)$ – распределение χ^2 (хи-квадрат) с *n-2* степенями свободы. Благодаря этому свойству мы можем построить для σ^2 несмещенную оценку s^2:

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}[y_i - \widehat{a} - \widehat{b}(x_i - \bar{x})]^2 . \tag{2.16}$$

Поскольку s^2 не зависит от $\widehat{a}$ и $\widehat{b}$, отношения

$$\sqrt{n}\,\frac{\widehat{a} - a}{s} \quad \text{и} \quad \frac{\widehat{b} - b}{s}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} . \tag{2.17}$$

Имеют распределение Стьюдента с *(n-2)* степенями свободы. Это позволяет легко построить для параметров *a* и *b* доверительные интервалы и указать тем самым, каковы статистические свойства погрешности при их оценивании посредством (2.12) и (2.13).

Проверка гипотез о коэффициенте наклона. Наиболее часто в задаче простой линейной регрессии возникает вопрос о равенстве нулю коэффициента наклона. Со статистической точке зрения это означает проверку гипотезы $H\colon b = 0$. Важность этой гипотезы объясняется тем, что в этом случае переменная *y* изменяется чисто случайно, не завися от значения *x*.

Против двусторонних альтернатив $b \neq 0$ гипотезу *H* следует отвергнуть на уровне значимости α, если число *0* не входит в доверительный интервал для *b*, который мы стандартным образом строим с помощью указанного выше отношения Стьюдента.

2.3 Анализ моделей парной регрессии

После того как определено уравнение линейной регрессии, производится оценка значимости уравнения регрессии в целом и его отдельных параметров (коэффициентов).

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

Непосредственному расчёту F–критерия Фишера предшествует анализ дисперсии. Центральное место в нём занимают разложение общей суммы квадратов отклонений переменной y от среднего значения $\bar{y}$ на две части – «объяснённую» и «необъяснённую»:

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

$$\boxed{\mathbf{TSS}} \; \boxed{=} \; \boxed{\mathbf{ESS}} \; \boxed{+} \; \boxed{\mathbf{RSS}}$$

TSS - общая сумма квадратов отклонений;
ESS - объяснённая сумма квадратов отклонений;
RSS - необъяснённая (остаточная) сумма квадратов отклонений;

2.4 Типовые задачи построения модели парной регрессии на основе результатов наблюдений

ЗАДАЧА 2.1

По семи областям РФ за 1999 год известны значения двух признаков (табл.2.1): расходы на покупку продовольственных товаров **y** (в %) и средняя дневная оплата труда **x** (в рублях).

<u>**Требуется**</u>:

1. Для анализа зависимости **y** от **x** рассчитать параметры моделей парной регрессии для трёх случаев: А) линейной $\mathbf{y=a+b\cdot x}$; Б) степенной $\mathbf{y=a\cdot x^b}$; В) показательной регрессии $\mathbf{y=a\cdot b^x}$.

2. Оценить адекватность каждой модели через среднюю ошибку аппроксимации $\bar{A}$ и F-критерий Фишера.

Решение:

Случай А: Для расчёта параметров a и b линейной парной регрессии $y=a+b\cdot x$ составляем и решаем систему нормальных уравнений относительно a и b:

$$\begin{cases} n\cdot a + b\cdot\sum x = \sum y\ , \\ a\cdot\sum \mathbf{x} + b\cdot\sum x^2 = \sum y\cdot x\ . \end{cases}$$

По исходным данным рассчитываем компоненты $\sum y$, $\sum \mathbf{x}$, $\sum y\cdot x$, $\sum x^2$, $\sum y^2$, и результаты заносим в 8-ю строку вспомогательной табл.2.2. Искомые коэффициенты уравнения линейной регрессии вычисляем по формулам:

$$b = \frac{\overline{y\cdot x} - \overline{y}\cdot\overline{x}}{\sigma_x{}^2} = \frac{3166{,}05 - 57{,}89\cdot 54{,}9}{5{,}86^2} = -0{,}35.$$

$$a = \overline{y} - b\cdot\overline{x} = 57{,}89 + 0{,}35\cdot 54{,}9 = 76{,}88.$$

1. Вычисляют линейный коэффициент парной корреляции:

$$r_{xy} = b\cdot\frac{\sigma_x}{\sigma_y} = -\,0{,}35\cdot 5{,}85/5{,}74 = -0{,}357.$$

Полученный результат свидетельствует о наличии умеренной обратной статистической связи между переменными x и y.

2. Определяют коэффициент детерминации:

$$r^2{}_{xy} = (-0{,}35)^2 = 0{,}127;$$

вариация результата на 12,7% объясняется вариацией фактора x.

3. Подставляя в уравнение регрессии фактические значения x, определяют теоретические (расчётные) значения $\hat{y}_x$, которые заносят в 7-й столбец табл.2.2.

4. Вычисляют величину ошибки A_i, $i=1,\ldots,7$ и средней ошибки аппроксимации Ā:

$$Ā = \overline{A} = \frac{1}{n}\cdot\sum A_i = \frac{1}{n}\cdot\sum|y-\hat{y}|\cdot 100\ \% = \frac{56{,}7\cdot 100\ \%}{7} = 8{,}1\ \%.$$

В среднем расчётные значения отклоняются от фактических на 8,1%.

Вычисляют F-критерий: $F_{\text{факт}} = \frac{0{,}127}{0{,}873}\cdot 5 = 0{,}7;$ поскольку $1 \le F \le \infty$, следует рассмотреть значение F^{-1}. Полученное значение указывает на необходимость принять гипотезу H_0 о случайной природе выявленной зависимости и статистической незначимости параметров уравнения и показателя тесноты связи.

Случай Б: Построению степенной модели $\mathbf{y=a\cdot x^b}$ предшествует процедура линеаризации переменных. Воспользуемся приёмом линеаризации на основе логарифмирования обеих частей уравнения: lg y= lg a + lg x.

Пусть lg y=Y, lg x=X, lg a=C. Тогда получим $y=C+b\cdot X$. Для упрощения расчётов воспользуемся табл.2.3.

Таблица 2.1.

№ п/п	Область	Расходы на покупку продовольственных товаров в общих расходах, %, **у**	Средняя дневная заработная плата одного работающего, Руб., **х**
1	Область А	68,8	45,1
2	Область Б	61,2	59,0
3	Область В	59,9	57,2
4	Область Г	56,7	61,8
5	Область Д	55,0	58,8
6	Область Е	54,3	47,2
7	Область Ж	49,3	55,2

Таблица 2.2.

	y	x	y·x	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	A_i
1	2	3	4	5	6	7	8	9
1	68,8	45,1	3102,88	2034,01	4733,44	61,3	7,5	10,9
2	61,2	59,0	3610,80	3481,00	3745,44	56,5	4,7	7,7
3	59,9	57,2	3426,28	3271,84	3588,01	57,1	2,8	4,7
4	56,7	61,8	3504,06	3819,24	3214,89	55,5	1,2	2,1
5	55,0	58,8	3234,00	3457,44	3025,00	56,5	-1,5	2,7
6	54,3	47,2	2562,96	2227,84	2948,49	60,5	-6,2	11,4
7	49,3	55,2	2721,36	3047,04	2430,49	57,8	-8,5	17,2
$\sum$	405,2	384,3	22162,3	21338,4	23685,8	405,2	0,0	56,7
Средн. знач.	57,89	54,90	3166,05	3048,34	3383,68			**8,1**
σ	5,74	5,86						
σ^2	32,92	34,34						

Рассчитаем параметры С и b:

$$b = \frac{\overline{Y \cdot X} - \overline{Y} \cdot \overline{X}}{\sigma_x^{\ 2}} = \frac{3{,}0572 - 1{,}7605 \cdot 1{,}737}{0{,}0484^{\ 2}} = -0{,}298;$$

$$C = \overline{Y} - b \cdot \overline{X} = 1{,}7605 + 0{,}298 \cdot 1{,}7370 = 2{,}278.$$

Получим линейное уравнение: $\widehat{Y} = 2{,}278 - 0{,}298 \cdot X$. Выполнив его потенцирование, определим:

$$\boxed{\widehat{y} = 10^{2{,}278} \cdot x^{-0{,}298} = 189{,}7 \cdot x^{-0{,}298}.}$$

Подставляя в данное уравнение фактические значения x, получаем теоретические значения результата $\widehat{y}_x$. По ним рассчитываем показатели: коэффициент тесноты связи – индекс корреляции ρ_{xy} и среднюю ошибку аппроксимации $\overline{A}_i$:

$$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \widehat{y}_x)^2}{\sum(y - \overline{y})^2}} = \sqrt{1 - \frac{28{,}27}{32{,}92}} = 0{,}3758, \qquad \overline{A} = 8{,}0\ \%.$$

Характеристики степенной модели указывают, что она более точно (по сравнению с линейной моделью) описывает взаимосвязь переменных.

Случай В. Перед построением уравнения показательной кривой $y = a \cdot b^x$ предшествует процедура линеаризации переменных при логарифмировании обеих частей уравнения:

$\lg y = \lg a + \lg b$; $Y = C + B \cdot x$, где Y=lg y, C=lg a, B=lg b.

Для упрощения представления результатов расчётов воспользуемся табл.1.4. Значения параметров регрессии А и В составили:

$$B = \frac{\overline{Y \cdot x} - \overline{Y} \cdot \overline{x}}{\sigma_x^{\ 2}} = \frac{96{,}5711 - 1{,}7605 \cdot 54{,}9}{5{,}86^2} = -\ 0{,}0023\ ,$$

$$A = \overline{Y} - B \cdot \overline{x} = 1{,}7605 + 0{,}0023 \cdot 54{,}9 = 1{,}887.$$

Получено линейное уравнение: $\widehat{Y} = 1{,}887 - 0{,}0023 \cdot x.$

Произведём потенцирование полученного уравнения и запишем его в обычной форме: $\widehat{y} = 10^{1{,}887} \cdot 10^{-0{,}0023 \cdot x} = 77{,}1 \cdot (0{,}9947)^x$.

Тесноту связи оценим через индекс корреляции ρ_{xy}:

$$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \widehat{y}_x)^2}{\sum(y - \overline{y})^2}} = \sqrt{1 - \frac{28{,}68}{34{,}3396}} = 0{,}3589\,.$$

Значение индекс корреляции ρ_{xy}=0,3589 свидетельствует, что связь между переменными умеренная. Средняя ошибка аппроксимации (для случая В) составила $\overline{A} = 8{,}0\ \%$. Это свидетельствует о том, что ошибка находится в допустимых пределах. Показательная функция немного хуже описывает изучаемую зависимость.

Таблица 2.3.

	Y	**X**	**Y·X**	**X^2**	**Y^2**	$\widehat{Y}_x$	$Y-\widehat{Y}_x$	**A_i**
1	2	3	4	5	6	7	8	9
1	1,8376	1,6542	3102,88	2034,01	4733,44	61,3	7,5	10,9
2	1,7868	1,7709	3610,80	3481,00	3745,44	56,5	4,7	7,7
3	1,7774	1,7574	3426,28	3271,84	3588,01	57,1	2,8	4,7
4	1,7536	1,7910	3504,06	3819,24	3214,89	55,5	1,2	2,1
5	1,7404	1,7694	3234,00	3457,44	3025,00	56,5	-1,5	2,7
6	1,7348	1,6739	2562,96	2227,84	2948,49	60,5	-6,2	11,4
7	1,6928	1,7419	2721,36	3047,04	2430,49	57,8	-8,5	17,2
$\sum$	12,3234	12,1587	21,4003	21,7078	21,1355	403,5	1,7	56,3
Средн. знач.	1,7605	1,737	3,0572	3,1011	3,0194		28,27	**8,0**
σ	0,0425	0,0484						
σ^2	0,0018	0,0023						

Таблица 2.4.

	Y	**x**	**Y·x**	**X^2**	**Y^2**	$\widehat{y}_x$	$y-\widehat{y}_x$	**A_i**
1	2	3	4	5	6	7	8	9
1	1,8376	45,1	82,8758	2034,01	3,3768	60,7	8,1	11,8
2	1,7868	59,0	105,4212	3481,00	3,1927	56,4	4,8	7,8
3	1,7774	57,2	101,6673	3271,84	3,1592	56,9	3,0	5,0
4	1,7536	61,8	108,3725	3819,24	3,0751	55,5	1,44	2,1
5	1,7404	58,8	102,3355	3457,44	3,0290	56,4	-1,4	2,5
6	1,7348	47,2	81,8826	2227,84	3,0095	60,0	-5,7	10,5
7	1,6928	55,2	93,4426	3047,04	2,8656	57,5	-8,2	16,6
$\sum$	12,3234	384,3	675,9974	21338,41	21,7078	403,4	-1,8	56,3
Средн. знач.	1,7605	1,737	96,5711	3048,34	3,1011		28,27	**8,0**
σ	0,0425	5,86						
σ^2	0,0018	34,3396						

Таблица 2.5. Результаты решения задачи.

Тип модели парной регрессии	Параметры модели			Средняя ошибка $\overline{A}$, %.
1. Линейная **y=a+b·x**	**a=76,88**	**b=-0,35**	r_{xy}=-0,357	8,1
2. Показательная **y=a·x^{b}**	**a=189,7**	**b=-0,298**	ρ=0,3759	8,0
3. Степенная **y=a·b^{x}**	**a= 77,1**	**b=0,9947**	ρ=0,3589	8,0

ЗАДАЧА 2.2

По двенадцати населённым пунктам области X известны статистические данные за 1998 год (табл.2.6).

Требуется:

1. Для анализа зависимости **y** от **x** построить уравнение модели парной линейной регрессии **y=a+b·x.**

2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации Ā.

3. Оценить статистическую значимость параметров регрессии и корреляции.

4. Выполнить прогноз заработной платы **y** при прогнозном значении среднедушевого прожиточного минимума **x**, составляющем 107 % от среднего уровня.

5. Оценить точность прогноза и его доверительный интервал.

<u>Решение:</u>

1. Для расчёта параметров a и b линейной парной регрессии y=a+b·x составляем и решаем систему нормальных уравнений относительно a и b.

По исходным данным рассчитываем компоненты ∑y, ∑x**,** ∑y·x, ∑x^2, ∑y^2, и результаты заносим в 13-ю строку вспомогательной табл.2.2.

Искомые коэффициенты уравнения линейной регрессии вычисляем по формулам:

$$b=\frac{\overline{y\cdot x}-\overline{y}\cdot\overline{x}}{\sigma_x{}^2}=\frac{13484-85,6\cdot 155,8}{7492,3-(85,6)^2}=0,92;$$

$$a=\overline{y}-b\cdot\overline{x}=155,8-0,92\cdot 85,6=77,0.$$

Получим линейную модель y=77,0+0,92·x . С увеличением среднедушевого прожиточного минимума на 1 рубль среднедневная заработная плата возрастает в среднем на 0,92 руб.

2. Вычислим линейный коэффициент парной корреляции:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0{,}92 \cdot 12{,}95 / 16{,}53 = 0{,}721.$$

Полученный результат свидетельствует о наличии средней прямой статистической связи между переменными x и y.

Определяем коэффициент детерминации:

$$r^2{}_{xy} = (0{,}721)^2 = 0{,}52.$$

Этот результат означает, что 52,0% вариации заработной платы объясняется вариацией фактора x – среднедушевого прожиточного минимума.

Качество модели можно оценить по величине средней ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \cdot \sum A_i = \frac{1}{n} \cdot \sum \left| y - \hat{y} \right| \cdot 100\,\% = \frac{68{,}9\,\%}{12} = 5{,}7\,\%.$$

В среднем расчётные значения отклоняются от фактических на 5,7%, и качество модели оценивается как хорошее, так как средняя величина Ā не превышает 8-10 %.

Таблица 2.6.

№ п/п	Номер населённого пункта	Среднедушевой прожиточный минимум в день одного трудоспособного, руб, **x**	Средняя дневная заработная плата одного работающего, руб., **y**
1	1	78,0	133,0
2	2	82,0	148,0
3	3	87,0	134,0
4	4	79,0	154,0
5	5	89,0	162,0
6	6	106,0	195,0
7	7	67,0	139,0
8	8	88,0	158,0
9	9	73,0	152,0
10	10	87,0	162,0
11	11	76,0	159,0
12	12	115,0	173,0

3. Оценку статистической значимости параметров регрессии выполним с помощью t-статистики Стьюдента и путём расчёта доверительного интервала каждого из показателей.

Выдвигаем гипотезу H_0 о статистически незначимом отличии показателей от нуля: $a=b=r_{xy}=0$.

Число степеней свободы определим как f=n-2=12-2=10. Значение $t_{табл}$ для числа степеней свободы f=10 и α=0,05 составит 2,23.

Определим средние значения ошибок в вычислении коэффициентов:

$$m_a = 12{,}6 \cdot \frac{\sqrt{89907}}{12 \cdot 12{,}95} = 24{,}3; \quad m_b = \frac{12{,}6}{12{,}95 \cdot \sqrt{12}} = 0{,}281; \quad m_r = \sqrt{\frac{1{,}0 - 0{,}52}{12-2}} = 0{,}219.$$

Таблица 2.7.

	x	Y	y·x	x²	y²	$\widehat{y}_x$	$y-\widehat{y}_x$	A_i
1	2	3	4	5	6	7	8	9
1	78	133	10374	6084	17689	149	-16	12,0
2	82	148	12136	6724	21904	152	-4	2,7
3	87	134	11658	7569	17956	157	-23	17,2
4	79	154	12166	6241	23716	150	4	2,6
5	89	162	14418	7921	26244	159	3	1,9
6	106	195	20670	11236	38025	174	21	10,8
7	67	139	9313	4489	19321	139	0	0,0
8	88	158	13904	7744	24964	158	0	0,0
9	73	152	11096	5329	23104	144	8	5,3
10	87	162	14094	7569	26244	157	5	3,1
11	76	159	12084	5776	25281	147	12	7,5
12	115	173	19895	13225	29929	183	-10	5,8
$\sum$	1027	1869	161808	89907	294377	1869	0	68,8
Средн. знач.	85,6	155,8	13484,0	7492,3	24531,4			**5,7**
σ	12,95	16,53						
σ^2	167,7	273,4						

Тогда получим $t_a = {}^{77}\!/_{24,3} = 3{,}2; \quad t_b = {}^{0,92}\!/_{0,281} = 3{,}3; \quad t_r = {}^{0,721}\!/_{0,219} = 3{,}3.$

Фактически значения t-статистики превосходят табличные значения:

$t_\alpha > t_{табл}=2{,}3; \quad t_b > t_{табл}=2{,}3; \quad t_r > t_{табл}=2{,}3,$

поэтому гипотеза H_0 отклоняется, т.е. коэффициенты a, b и r_{xy} статистически значимы.

Рассчитаем доверительные интервалы для коэффициентов a и b. Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a=2{,}23\cdot 24{,}3=54; \qquad \Delta_b=2{,}23\cdot 0{,}281=0{,}62.$$

Доверительные интервалы:

$\delta_a=a\pm\ \Delta_a=77\pm 54;$ $\delta_a^{max}=77+54=131;$ $\delta_a^{min}=77-54=23$

$\delta_b=b\pm\ \Delta_b=0{,}92\pm 0{,}62;$ $\delta_b^{max}=0{,}92+0{,}62=1{,}54;$ $\delta_b^{min}=0{,}92-0{,}62=0{,}30.$

Анализ верхней и нижней границ интервалов приводит к выводу о том, что с вероятностью $p=1-\alpha=0{,}95$ параметры a и b, находясь в указанных границах, не принимают нулевых значений, т.е. не являются статистически незначимыми и существенно отличаются от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогнозирования значения зависимой переменной. Если прогнозное значение прожиточного минимума составит:

$$x_p=\bar{x}\cdot 1{,}07=85{,}6\cdot 1{,}04=91{,}6 \text{ руб.},$$

тогда прогнозное значение средней дневной платы составит:

$$\hat{y}_p=77+0{,}92\cdot 91{,}6=161 \text{ руб.}$$

5. Ошибку прогноза вычислим по формуле:

$$m_y{*}=12{,}6\cdot\sqrt{1+\frac{1}{12}+\frac{(91{,}6-85{,}6)^2}{12\cdot(12{,}95)^2}}=13{,}2 \text{ руб.}$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$\Delta_y=t_{табл}\cdot m_y{*}=2{,}23\cdot 13{,}2=29{,}4.$

Доверительный интервал прогноза: $\delta_y=\hat{y}_p\pm\ \Delta_y=161\pm 29{,}4;$

$\delta_y^{max}=161+29{,}4=190{,}4$ руб.; $\delta_y^{min}=161-29{,}4=131{,}6$ руб.

Выполненный прогноз среднемесячной заработной платы оказался надёжным ($p=1-\alpha=1-0{,}05=0{,}95$), но относительно неточным, так как диапазон отношения верхней и нижней границ доверительного интервала D_y составляет 1,45 раза: $D_y=\delta_y^{max}/\delta_y^{min}=190{,}4/131{,}6=1{,}447$.

Задача решена.

Реализация задачи с помощью ППП Excel

1. Встроенная статистическая функция **ЛИНЕЙН** определяет параметры линейной регрессии $y=a+b\cdot x$. Порядок выполнения следующий:
   1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;
   2) выделите область пустых ячеек 5×2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики или область 1×2 - для получения только оценок коэффициентов регрессии;
   3) активизируйте Мастер функций любым из способов:

   а) в главном меню выберите **Вставка / Функция**;

б) на панели инструментов **Стандартная** щелкните по кнопке **Вставка функции**;

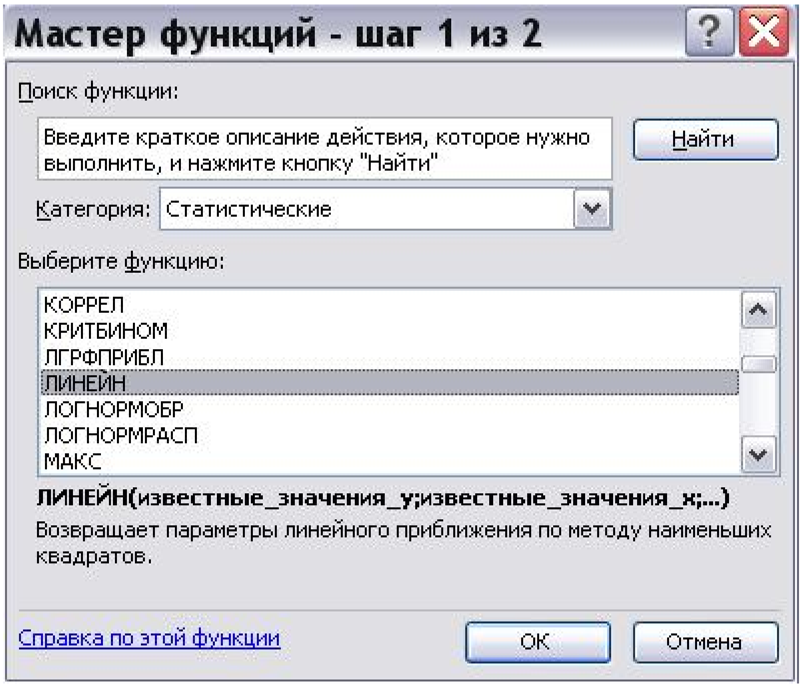
4) в окне Категория (рис. 2.1) выберите **Статистические**, в окне Функция – **ЛИНЕЙН**. Щелкните по кнопке **ОК**;

5) заполните аргументы функции (рис. 2.2):

Известные_значения_y – диапазон, содержащий данные результативного признака;

Известные_значения_x – диапазон, содержащий данные факторов независимого признака;

Константа – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении; если Константа = 1, то свободный член рассчитывается обычным образом, если Константа = 0, то свободный член равен 0;



Мастер функций - шаг 1 из 2

Поиск функции:

Введите краткое описание действия, которое нужно выполнить, и нажмите кнопку "Найти"

Категория: Статистические

Выберите функцию:

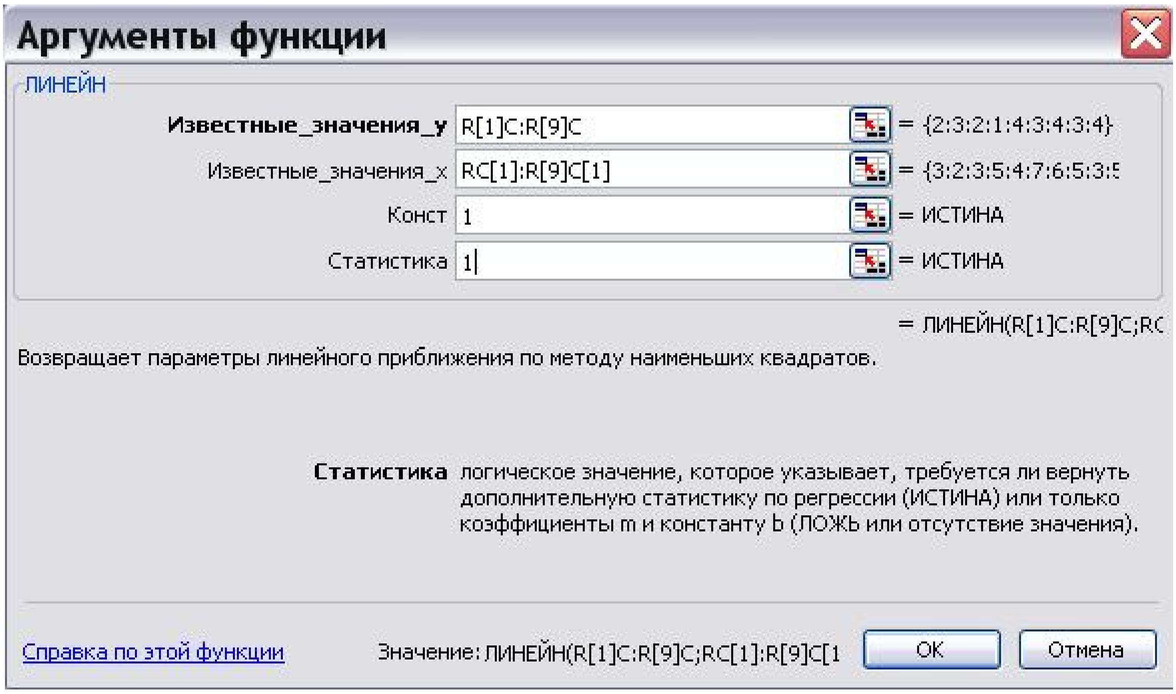
ЛИНЕЙН(известные_значения_y;известные_значения_x;...)

Возвращает параметры линейного приближения по методу наименьших квадратов.

Рисунок 2.1 - Диалоговое окно «Мастер функции»

Статистика – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если Статистика = 1, то дополнительная информация выводится, если Статистика = 0, то выводятся только оценки параметров уравнения.

Щелкните по кнопке **ОК**;



Аргументы функции

ЛИНЕЙН

Известные_значения_y R[1]C:R[9]C = {2:3:2:1:4:3:4:3:4}

Известные_значения_x RC[1]:R[9]C[1] = {3:2:3:5:4:7:6:5:3:5

Конст 1 = ИСТИНА

Статистика 1 = ИСТИНА

= ЛИНЕЙН(R[1]C:R[9]C;RC

Возвращает параметры линейного приближения по методу наименьших квадратов.

Статистика логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии (ИСТИНА) или только коэффициенты m и константу b (ЛОЖЬ или отсутствие значения).

Значение: ЛИНЕЙН(R[1]C:R[9]C;RC[1]:R[9]C[1

Рисунок 2.2 - Диалоговое окно ввода аргумента функции **ЛИНЕЙН**

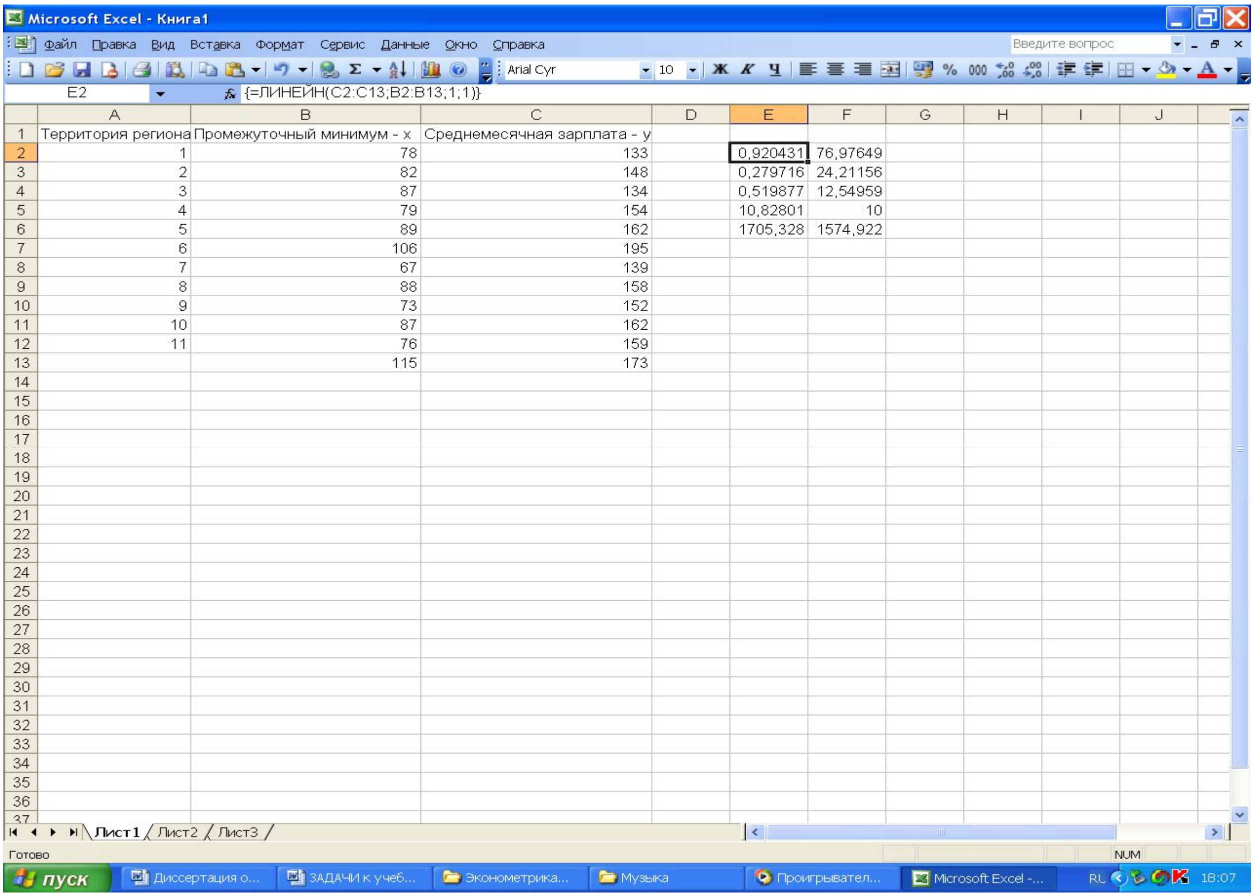
6) в левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу <F2>, а затем – на комбинацию клавиш <CTRL> + <SHIFT> + <ENTER>.

Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Значение коэффициента b	Значение коэффициента a
Среднеквадратическое отклонение b	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	Среднеквадратическое отклонение y
F - статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

Для вычисления параметров экспоненциальной кривой $y = \alpha \cdot \beta^x$ в MS Excel применяется встроенная статистическая функция **ЛГРФПРИБЛ**. Порядок вычисления аналогичен применению функции **ЛИНЕЙН**.

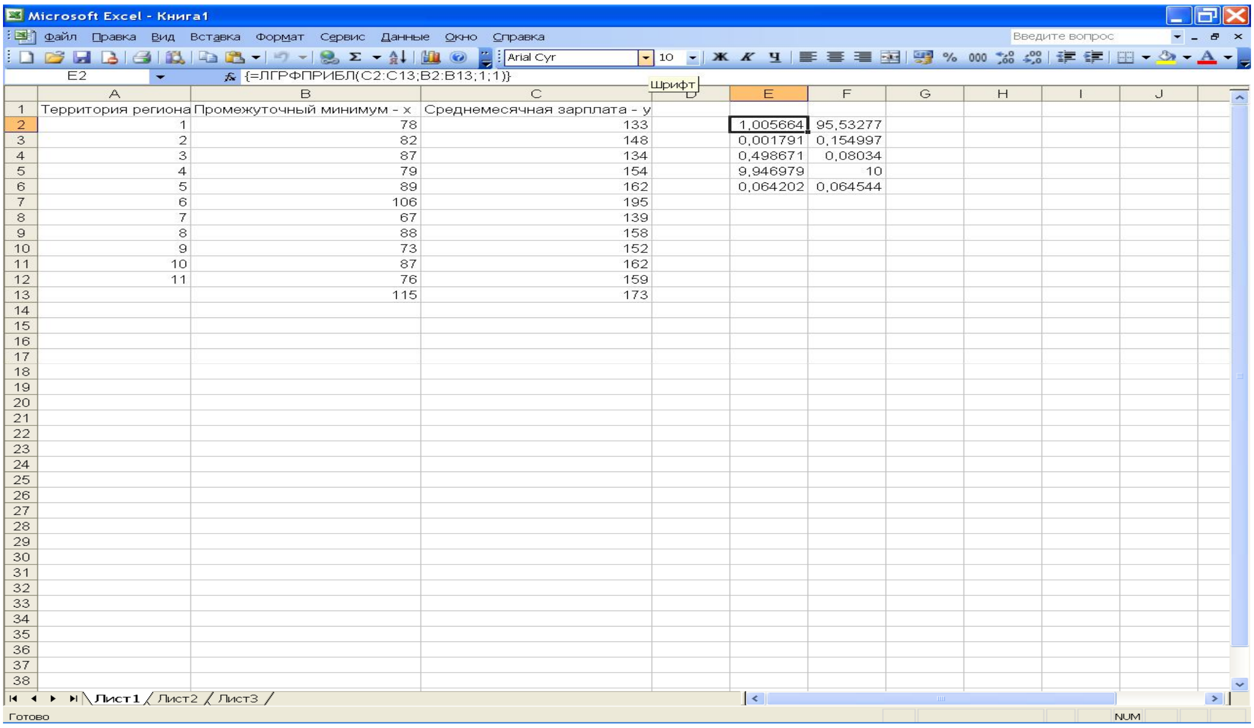
Для данных из примера 2 результат вычисления функции **ЛИНЕЙН** представлен на рис. 2.3, функции **ЛГРФПРИБЛ** на рис. 2.4.



E2 {=ЛИНЕЙН(C2:C13;B2:B13;1;1)}

	A	B	C	D	E	F
1	Территория региона	Промежуточный минимум - x	Среднемесячная зарплата - y			
2	1	78	133		0,920431	76,97649
3	2	82	148		0,279716	24,21156
4	3	87	134		0,519877	12,54959
5	4	79	154		10,82801	10
6	5	89	162		1705,328	1574,922
7	6	106	195			
8	7	67	139			
9	8	88	158			
10	9	73	152			
11	10	87	162			
12	11	76	159			
13		115	173			

Рисунок 2.3. Результат вычисления **ЛИНЕЙН**



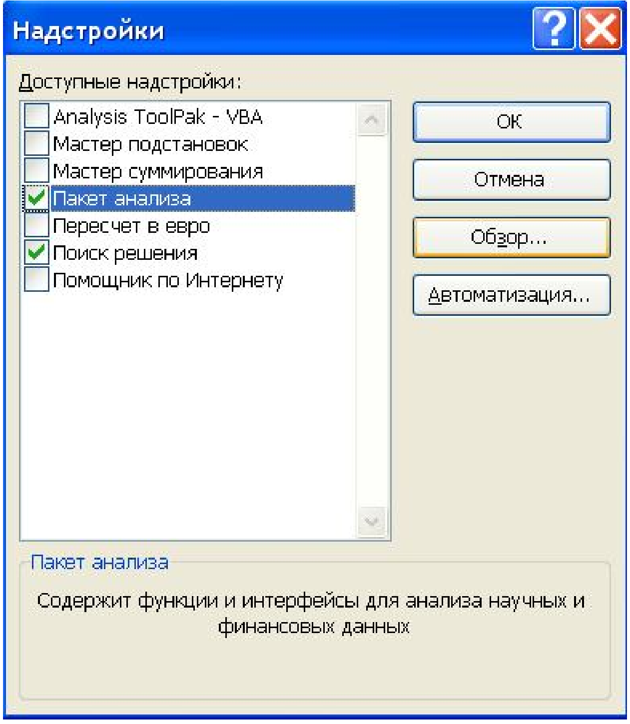
E2 {=ЛГРФПРИБЛ(C2:C13;B2:B13;1;1)}

	A	B	C	D	E	F
1	Территория региона	Промежуточный минимум - x	Среднемесячная зарплата - y			
2	1	78	133		1,005664	95,53277
3	2	82	148		0,001791	0,154997
4	3	87	134		0,498671	0,08034
5	4	79	154		9,946979	10
6	5	89	162		0,064202	0,064544
7	6	106	195			
8	7	67	139			
9	8	88	158			
10	9	73	152			
11	10	87	162			
12	11	76	159			
13		115	173			

Рисунок 2.4. Результат вычисления функции **ЛГРФПРИБЛ**

2. С помощью инструмента анализа данных **Регрессия**, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и графики подбора линии регрессии, остатков и нормальной вероятности. Порядок действий следующий:

1) проверите доступ к пакету анализа. В главном меню последовательно выберите **Сервис/Надстройки**. Установите флажок **Пакет анализа** (рис. 2.5);



☐ Analysis ToolPak - VBA
☐ Мастер подстановок
☐ Мастер суммирования
☑ Пакет анализа
☐ Пересчет в евро
☑ Поиск решения
☐ Помощник по Интернету

Пакет анализа

Содержит функции и интерфейсы для анализа научных и финансовых данных

Рисунок 2.5. Подключение надстройки **Пакет анализа**

2) в главном меню выберите **Сервис/Анализ данных/Регрессия**. Щелкните по кнопке **ОК**;
3) заполните диалоговое окно ввода данных и параметров вывода (рис. 2.6):

Входный интервал Y - диапазон, содержащий данные результативного признака;

Входной интервал X - диапазон, содержащий данные факторов независимого признака;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

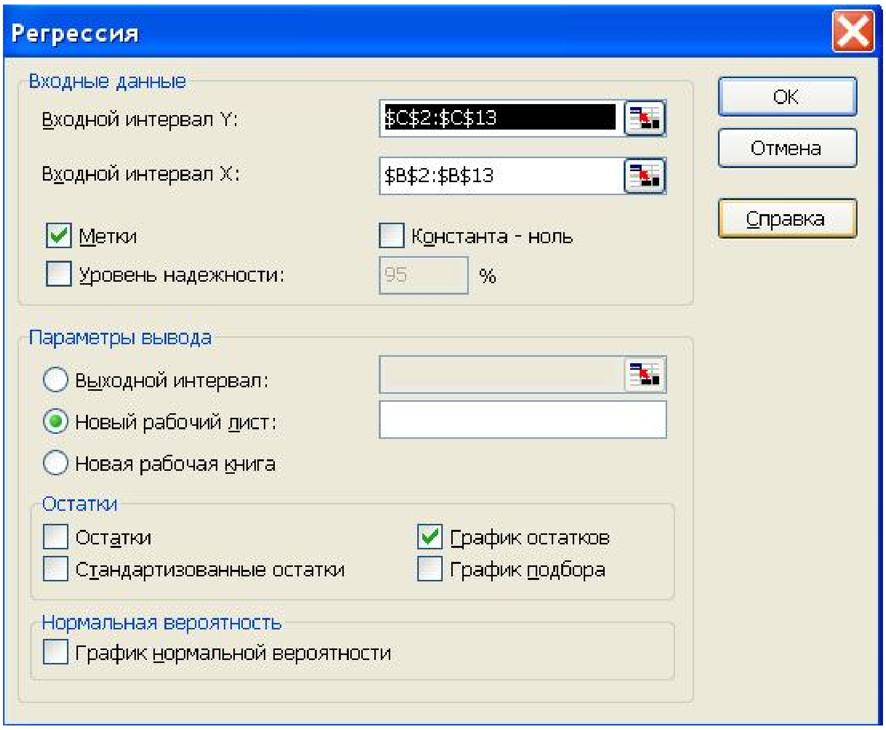
Константа – ноль – флажок, указывающий на наличие или отсутствие свободного члена в управлении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа.

Если необходимо получить информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК**.

Как видим, результаты вычислений вручную и с помощью ППП Excel совпадают.



Входной интервал Y: $C$2:$C$13

Входной интервал X: $B$2:$B$13

☑ Метки

☐ Константа - ноль

☐ Уровень надежности: 95 %

○ Выходной интервал:

● Новый рабочий лист:

○ Новая рабочая книга

☐ Остатки

☑ График остатков

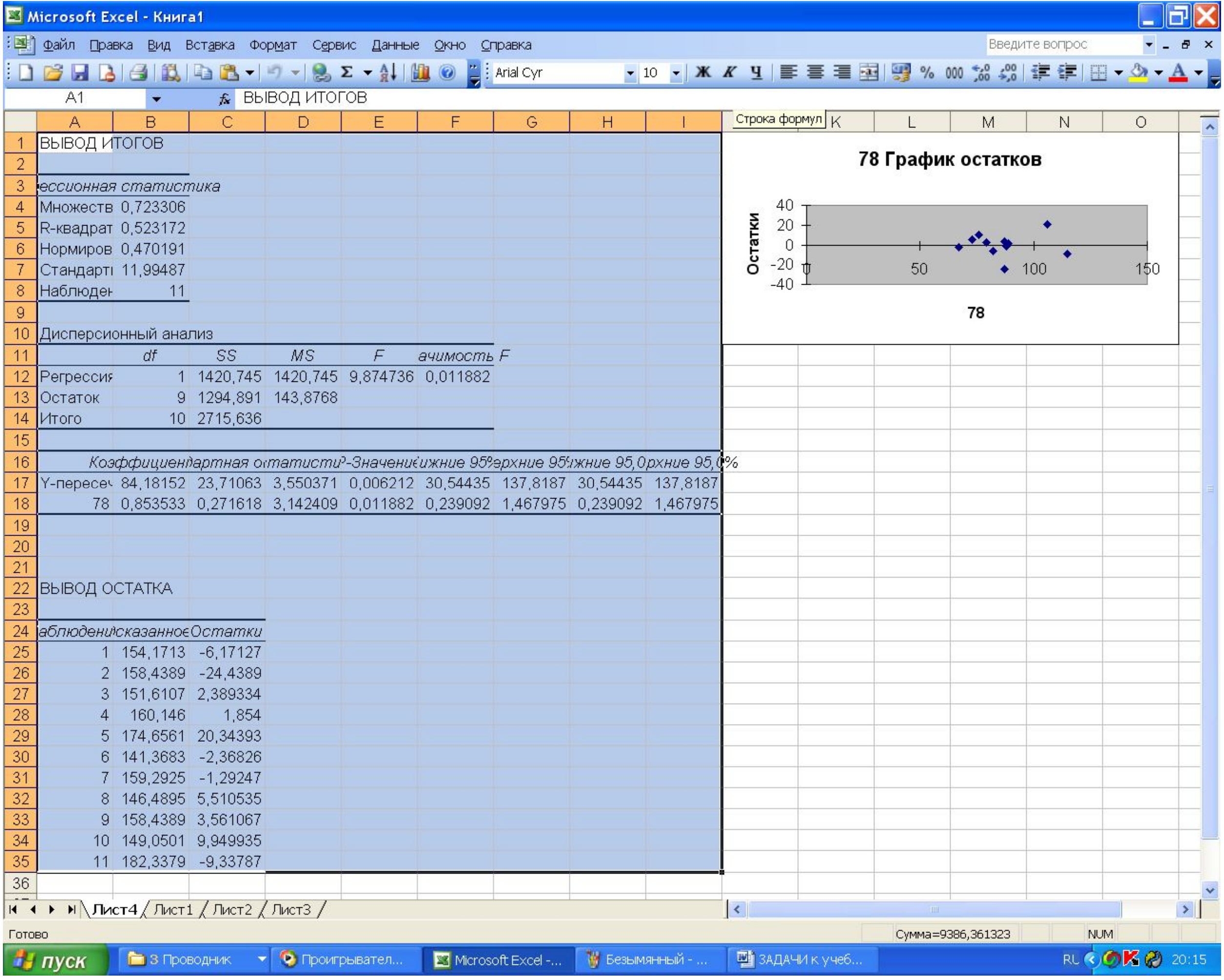
☐ Стандартизованные остатки

☐ График подбора

☐ График нормальной вероятности

Рисунок 2.6. Диалоговое окно ввода параметров инструмента **Регрессия**

Результаты регрессионного анализа для данных из примера 2 представлены на рис. 2.7.



ВЫВОД ИТОГОВ

ессионная статистика	
Множеств	0,723306
R-квадрат	0,523172
Нормиров	0,470191
Стандарті	11,99487
Наблюден	11

Дисперсионный анализ

	df	SS	MS	F	ачимость F
Регрессия	1	1420,745	1420,745	9,874736	0,011882
Остаток	9	1294,891	143,8768		
Итого	10	2715,636			

	Коэффициен	дартная о	статисти	Р-Значени	ижние 95%	ерхние 95%	ижние 95,0	рхние 95,0%
Y-пересеч	84,18152	23,71063	3,550371	0,006212	30,54435	137,8187	30,54435	137,8187
78	0,853533	0,271618	3,142409	0,011882	0,239092	1,467975	0,239092	1,467975

ВЫВОД ОСТАТКА

аблюдени	сказанное	Остатки
1	154,1713	-6,17127
2	158,4389	-24,4389
3	151,6107	2,389334
4	160,146	1,854
5	174,6561	20,34393
6	141,3683	-2,36826
7	159,2925	-1,29247
8	146,4895	5,510535
9	158,4389	3,561067
10	149,0501	9,949935
11	182,3379	-9,33787

Рисунок 2.7. Результат применения инструмента **Регрессия**

2.5 Нестандартные задачи построения моделей парной регрессии

ЗАДАЧА 2.3

Постановка задачи

Каждая ценная бумага – акция, облигация, контракт и другие – в каждый момент времени обладает стоимостью, которая называется курсом и устанавливается рынком (обычно как результат биржевых котировок). Даже обладая всей полнотой информации о выпустившем бумагу эмитенте, однозначно определить ее курс в заданный момент времени в будущем, как правило, невозможно. В этом случае наиболее естественно рассматривать курс ценной бумаги как значение случайной величины X

Пусть X_t - курс ценной бумаги в момент времени t, а X_{t+1} - в момент времени $t+1$ (обычно единица времени – это промежуток между котировками). Обратимся к случаю, когда момент времени t уже наступил, а $t+1$ еще нет. Рассмотрим величину

$$r_t = \frac{X_{t+1} - X_t}{X_t}.$$

Так как X_{t+1} - случайная величина, то и r - тоже случайная величина. Она называется доходностью ценной бумаги.

Очевидно, что значение именно этой величины определяет привлекательность ценной бумаги для инвестора. И одна из главных задач финансового анализа состоит в возможно более точном предсказании значения □ндеек□ны r.

Модели, рассматриваемые в финансовом анализе, связывают случайную величину r с величинами, которые объективно характеризуют финансовый рынок в целом. Такие величины называются факторами. В зависимости от постановки задачи факторы могут считаться как случайными, так и детерминированными, т.е. точно известными величинами.

В самом простом случае выделяется один фактор. Тогда статистическая модель имеет вид:

$$r = a + bF + \varepsilon. \qquad (2.18)$$

Здесь a и b - постоянные (неизвестные параметры), ε - случайная величина, удовлетворяющая условию: $M_F(\varepsilon) = 0$, где $M_F(\varepsilon)$ - условное математическое ожидание случайной величины ε относительно F. Из этого предположения следует, что и безусловное математическое ожидание величины ε также равно нулю. В самом деле:

$$M(\varepsilon) = M(M_F(\varepsilon)) = 0.$$

Отсюда также следует, что если фактор F рассматривается как случайная величина, то ее ковариация с ε равна нулю. Действительно, используя свойства условного математического ожидания, получаем:

$$\operatorname{cov}(F,s) = M(F\varepsilon) - M(F)M(\varepsilon) = M(F\varepsilon) = M(M_F(F\varepsilon)) = M(FM_F(\varepsilon)) = 0.$$

Значения коэффициентов a и b нетрудно выразить через числовые характеристики r и F:

$$\operatorname{cov}(r,F)=b\operatorname{cov}(F,F)+\operatorname{cov}(\varepsilon,F)$$

или $$\operatorname{cov}(r,F)=b\operatorname{cov}(F,F).$$

Отсюда

$$b=\frac{\operatorname{cov}(r,F)}{\operatorname{cov}(F,F)}=\frac{\operatorname{cov}(r,F)}{D(F)}.$$

Перейдя в уравнении модели (2.18) к математическим ожиданиям, получим:

$$M(r)=a+bM(F)+M(\varepsilon).$$

Но $M(\varepsilon)=0$, поэтому

$$a=M(r)-bM(F)=M(r)-\frac{\operatorname{cov}(r,F)}{D(F)}M(F).$$

Коэффициент b называется чувствительностью доходности ценной бумаги к фактору F. Коэффициент a называется сдвигом.

В классическом регрессионном анализе значения факторов F считаются детерминированными величинами, т.е. модель (2.18) имеет вид:

$$r_t=a+bF_t+\varepsilon_t.$$

Здесь $t=1,\ldots,n$ - моменты времени – интерпретируются как номер наблюдения; $F_1,\ldots,F_n$ - известные значения факторов; r_t - наблюдаемые выборочные значения случайной величины r; a и b - неизвестные параметры. Их оценки можно построить методом наименьших квадратов.

Разные модели финансового рынка рассматривают различные величины в качестве фактора F.

Одна из самых распространенных моделей использует в качестве фактора F доходность рыночного индекса r_M. Рыночным индексом называется взвешенная сумма курсов акций наиболее значительных эмитентов финансового рынка. Например, в США наиболее распространены следующие □ндеекксы:

DJ (индекс Доу-Джонса) — рассчитывается по 30 наиболее значимым корпорациям, таким, как Microsoft, Coca Cola, General Motors и т.д.;

индекс S&P 500 (Standard and Poor's) — рассчитывается по 500 наиболее крупным компаниям;

сводный индекс NYSE — для его расчета используются курсы акций, зарегистрированных на Нью-йоркской фондовой бирже.

Очевидно, рыночный индекс в определенной степени отражает состояние экономики в целом. Так что рыночная модель показывает, насколько доходность ценной бумаги соответствует экономической динамике страны (или даже сообщества стран).

Случайная величина е отражает зависимость доходности ценной бумаги от обстоятельств, специфических именно для ее эмитента.

Доходность рыночного индекса по сути представляет собой усредненную доходность различных ценных бумаг. Если рассматривать множество всех ценных бумаг, фигурирующих на рынке, то коэффициент b наудачу выбранной ценной бумаги представляет собой значение случайной величины (обозначим ее той же буквой b). Если ценная бумага включена в индекс, то по самому определению индекса имеет место равенство: $M(b)=1$.

Если конкретное наблюдаемое значение ценной бумаги больше единицы, значит ее доходность растет в среднем быстрее, чем рынок в целом. Такие бумаги называются «агрессивными», бумаги с коэффициентом b меньшим единицы называются «оборонительными».

Исходные данные

Значения доходности акций Widget Manufacturing и доходности индекса даны в таблице 2.8:

Таблица 2.8

Год	Квартал	Доходность WM	Доходность индекса
1	1	-13,38	2,52
	2	16,79	5,45
	3	-1,67	0,76
	4	-3,46	2,36
2	5	10,22	8,56
	6	7,13	8,67
	7	6,71	10,8
	8	7,84	3,33
3	9	2,15	-5,07
	10	7,95	7,1
	11	-8,05	-11,57
	12	7,68	4,65
4	13	4,75	14,59
	14	7,55	2,66
	15	-2,36	3,81
	16	4,98	7,99

Требуется:

Оценить зависимость доходности акций от доходности индекса.

Решение:

Метод наименьших квадратов дает следующие результаты: $b=0{,}63$, $a=0{,}79$. Стандартные ошибки $\sigma_b=0{,}28$, $\sigma_a=2{,}03$, коэффициент детерминации $R^2=0{,}27$. Полученные характеристики свидетельствуют о том, что доходность акций Widget Manufacturing существенно зависит от рыночного индекса. Как видно, акции Widget Manufacturing являются «оборонительными».

На финансовом рынке присутствуют и бумаги с нулевым коэффициентом b. В этом случае имеет место соотношение $r=r_f$, откуда следует, что ожидаемая доходность ценной бумаги фиксированная и не зависит от состояния рынка в целом. Такая ситуация характерна, например, для облигаций.

3. МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

3.1 Спецификация моделей множественной регрессии

Множественная регрессия представляет собой уравнение связи с несколькими независимыми переменными:

$$y = f(x_1,\ x_2,\ ...,\ x_n,,\ \varepsilon\ ),$$

где y – зависимая переменная (результативный признак);

$x_1,\ x_2,\ ...,\ x_n$ - независимые переменные (факторы); n - число значимых факторов.

Для построения уравнений множественной регрессии на практике, как правило, используются следующие аналитические зависимости:

1) линейная $y = a + b_1x_1 + b_2x_2 + ... + b_nx_n + \varepsilon$;
2) степенная $y = ax_1^{b_1} \cdot x_2^{b_2} \cdot ... \cdot x_n^{b_n} \cdot \varepsilon$;
3) экспонента $y = exp(a + b_1x_1 + b_2x_2 + ... + b_nx_n + \varepsilon)$;
4) гипербола $y = (a + b_1x_1 + b_2x_2 + ... + b_nx_n + \varepsilon)^{-1}$.

В экономическом анализе используют и другие функции, приводимые к линейному виду. В модели множественной регрессии за изменение зависимой переменной регрессии y отвечают несколько экономических факторов – объясняющих переменных x. Параметры множественной регрессии b_j, $j=\overline{1,n}$, показывают степень влияния на зависимую переменную y экономических факторов, обозначенных соответствующими объясняющими переменными.

ПРИМЕР 3.1. Пусть $\boxed{P = 2{,}0 \cdot C_1^{\ 0{,}3} \cdot C_2^{\ 0{,}2} \cdot C_3^{\ 0{,}3} \cdot \varepsilon}$ Здесь: P - производственная функция (объём выпуска продукции), у.е.; C_1 - стоимость основных производственных фондов, у.е.; C_2 - затраты труда в человеко-днях; C_3 - затраты на производство.

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов. Оптимальность оценок параметров достигается на основе минимизации суммы квадратов отклонений: $S = \varepsilon_1^{\ 2} + \varepsilon_2^{\ 2} + ... + \varepsilon_N^{\ 2}$; N - число наблюдений.

Здесь ε_i является остатком в наблюдении i, т.е. разностью между фактическим значением y_i в этом наблюдении и значением $\hat{y}(x)$, прогнозируемым по уравнению регрессии после подстановки в него значений факторов $x=(x_{i1}, \ldots, x_{in})$. Для двух факторов ($n = 2$) можно записать:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \varepsilon; \qquad \varepsilon_i = y_i - \hat{y}(x_i) = y_i - a - b_1x_{i1} - b_2x_{i2}\,.$$

При этом сумма квадратов остатков:

$$S=\sum\varepsilon^2=\sum(y_i-a-b_1x_{i1}-b_2x_{i2})^2\,. \tag{3.1}$$

Необходимые условия 1-го порядка достижения минимума функции S имеют вид:

$$\begin{gathered}\{\ \partial S/\partial a=0;\qquad \partial S/\partial b_1=0;\qquad \partial S/\partial b_2=0\}\Rightarrow\\ \partial S/\partial a=-2\sum(y_i-a-b_1x_{i1}-b_2x_{i2})=0;\\ \partial S/\partial b_1=-2\sum x_{i1}(y_i-a-b_1x_{i1}-b_2x_{i2})=0;\\ \partial S/\partial b_2=-2\sum x_{i2}(y_i-a-b_1x_{i1}-b_2x_{i2})=0.\end{gathered} \tag{3.2}$$

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов. Для линейных уравнений и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, аналитическое или численное решение которой позволяет получить оценки параметров регрессии:

$$\begin{cases}\sum y=N\cdot a+b_1\cdot\sum x_1+b_2\cdot\sum x_2+\ldots+b_n\cdot\sum x_n,\\ \sum y\cdot x_1=a\cdot\sum x_1+b_1\cdot\sum x_1^2+b_2\cdot\sum x_1\cdot x_2+\ldots+b_n\cdot\sum x_n\cdot x_1,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \sum y\cdot x_n=a\cdot\sum x_n+b_1\cdot\sum x_1\cdot x_n+b_2\cdot\sum x_2\cdot x_n+\ldots+b_n\cdot\sum x_n^2.\end{cases} \tag{3.3}$$

Для решения системы алгебраических уравнений может быть применен метод определителей, согласно которому:

$$a=\frac{\Delta a}{\Delta},\ b_1=\frac{\Delta b_1}{\Delta},\ldots.,\quad b_n=\frac{\Delta b_n}{\Delta},$$

где $\Delta=\begin{vmatrix}N & \sum x_1 & \sum x_2 & \ldots & \sum x_n\\ \sum x_1 & \sum x_1^2 & \sum x_2x_1 & \ldots & \sum x_nx_1\\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & \ldots & \sum x_nx_2\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ \sum x_n & \sum x_1x_n & \sum x_2x_n & \ldots & \sum x_n^2\end{vmatrix}$ - определитель системы;

$\Delta a,\Delta b_1,\ldots,\Delta b_n$ - частные определители, которые получаются путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Другой вид уравнения множественной регрессии - уравнение регрессии в стандартизованном масштабе $t_y=\beta_1t_{x_1}+\beta_2t_{x_2}+\ldots.+\beta_nt_{x_n}$,

где $t_y=\dfrac{y-\overline{y}}{\sigma_y},t_{x_i}=\dfrac{x_i-\overline{x_i}}{\sigma_{x_i}}$ - стандартизованные переменные;

β_i - стандартизованные коэффициенты регрессии.

К уравнению множественной регрессии в стандартизованном масштабе применим МНК. Стандартизованные коэффициенты регрессии (β - коэффициенты) определяются из следующей системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 \quad + \beta_2 \cdot r_{x_2x_1} + \beta_3 \cdot r_{x_3x_1} + \ldots + \beta_n \cdot r_{x_nx_1}, \\ r_{yx_2} = \beta_1 \cdot r_{x_2x_1} + \beta_2 \quad + \beta_3 \cdot r_{x_3x_2} + \ldots + \beta_n \cdot r_{x_nx_2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yx_n} = \beta_1 \cdot r_{x_nx_1} + \beta_2 \cdot r_{x_nx_2} + \beta_3 \cdot r_{x_3x_n} + \ldots + \beta_n. \end{cases}$$

Применяя МНК к уравнению множественной регрессии в стандартизованном масштабе, после соответствующих преобразований полу-чим систему нормальных уравнений вида

$$\begin{cases} R_{yx_1} = \beta_1 \quad + \beta_2 \cdot R_{x_2x_1} + \beta_3 \cdot R_{x_3x_1} + \ldots + \beta_p \cdot R_{x_px_1}, \\ R_{yx_1} = \beta_1 \cdot R_{x_2x_1} + \beta_2 \quad + \beta_3 \cdot R_{x_3x_2} + \ldots + \beta_p \cdot R_{x_px_2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ R_{yx_p} = \beta_1 \cdot R_{x_px_1} + \beta_2 \cdot R_{x_px_2} + \beta_3 \cdot R_{x_3x_p} + \ldots + \beta_p. \end{cases}$$

Решая систему методом определителей, найдем параметры - стандартизованные коэффициенты регрессии (β-коэффициенты).

Стандартизованные коэффициенты регрессии показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор x_i изменится на одну сигму при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i сравнимы 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.

ПРИМЕР 3.2. Пусть функция издержек производства y (тыс. руб.) характеризуется уравнением вида $\boxed{y = 200 + 1{,}2 \cdot x_1 + 1{,}1 \cdot x_2 + \varepsilon}$, где x_1 - основные производственные фонды (тыс. руб.); x_2 - численность занятых в производстве (чел.).

Анализируя уравнение, отмечаем, что при той же занятости дополнительный рост стоимости основных производственных фондов на 1 тыс. руб. влечет за собой увеличение затрат в среднем на 1,2 тыс. руб., а увеличение численности занятых на одного человека способствует при той же технической оснащенности предприятий росту затрат в среднем на 1,1 тыс. руб. Однако это не означает, что фактор x_1 оказывает более сильное влияние на издержки производства по сравнению с фактором x_2. Такое сравнение возможно, если обратиться к уравнению регрессии в стандартизованном масштабе. Предположим, оно выглядит так: $t_y = 0{,}5 \cdot t_{x_1} + 0{,}8 \cdot t_{x_2}$.

Это означает, что с ростом фактора x_1 на одну сигму при неизменной численности занятых затраты на продукцию увеличиваются в среднем на 0,5 сигмы. Так как $\beta_1 < \beta_2 (0,5 < 0,8)$, то можно заключить, что большее влияние оказывает на производство продукции фактор x_2, а не x_1, как кажется из уравнения регрессии в натуральном масштабе.

В парной зависимости стандартизованный коэффициент регрессии есть не что иное, как линейный коэффициент корреляции r_{yx}. Подобно тому, как в парной зависимости коэффициенты регрессии и корреляции связаны между собой, так и во множественной регрессии коэффициенты «чистой» регрессии b_i, связаны со стандартизованными коэффициентами регрессии β_1, а именно:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \tag{3.4}$$

Это позволяет от уравнения регрессии в стандартизованном масштабе

$$\hat{t}_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \ldots + \beta_p \cdot t_{x_p} \tag{3.5}$$

переходить к уравнению регрессии в натуральном масштабе переменных:

$$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_p \cdot x_p .$$

Параметр a определяется как

$$a = \overline{y} - b_1 \cdot \overline{x_1} - b_2 \cdot \overline{x}_2 - \ldots - b_p \cdot \overline{x}_p . \tag{3.6}$$

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

Компьютерные программы построения уравнения множественной регрессии в зависимости от использованного в них алгоритма решения позволяют получить либо только уравнение регрессии для исходных данных, либо, кроме того, уравнение регрессии в стандартизованном масштабе.

При нелинейной зависимости признаков, приводимой к линейному виду, параметры множественной регрессии также определяются МНК с той лишь разницей, что он используется не к исходной информации, а к преобразованным данным. Так, рассматривая степенную функцию мы преобразовываем ее в линейный вид:

$$lgy = lga + b_1 \cdot lgx_1 + b_2 \cdot lgx_2 + \ldots + b_p \cdot lgx_p + lg\,\varepsilon ,$$

где переменные выражены в логарифмах.

Далее обработка МНК та же, что и описана выше: строится система нормальных уравнений и определяются параметры lga, b_1, b_2,..., b_p. Потенцируя значение lga, найдем параметр a и соответственно общий вид уравнения степенной функции.

Поскольку параметры степенной функции представляют собой коэффициенты эластичности, то они сравнимы по разным факторам.

ПРИМЕР 3.3. При исследовании спроса на масло получено следующее уравнение:

$$\boxed{lgy = -\,1,25 - 0,858 \cdot lgx_1 + 1,126 \cdot lgx_2 + \varepsilon},$$

где y - количество масла на душу населения (кг);

x_1 - цена (руб.); x_2 - доход на душу населения (тыс. руб.).

Анализируя уравнение, отмечаем, что с ростом цены на 1 % при том же доходе спрос снижается в среднем на 0,858 %, а рост дохода на 1 % при неизменных ценах вызывает увеличение спроса в среднем на 1,126 %.

В виде степенной функции данное уравнение примет вид:

$$y = 0{,}056 \cdot x_1 - x_2^{-1{,}126} \cdot \varepsilon \,.$$

При других нелинейных функциях методика оценки параметров МНК осуществляется также. В отличие от предыдущих функций параметры более сложных моделей не имеют четкой экономической интерпретации: они не являются показателями силы связи и ее эластичности. Это не исключает возможности их применения, но делает их менее привлекательными в практических расчетах.

Анализ уравнений множественной регрессии

Связь коэффициентов множественной регрессии b_i со стандартизованными коэффициентами β_i описывается соотношением

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

Параметр a определяется как $a = \overline{y} - b_1 \cdot \overline{x_1} - b_2 \cdot \overline{x}_2 - \ldots - b_p \cdot \overline{x}_n$.

Средние *коэффициенты эластичности* для линейной регрессии рассчитываются по формуле

$$\overline{\text{Э}}_{yx_j} = b_j \frac{\overline{x_j}}{\overline{y}}.$$

Для расчета частных коэффициентов эластичности применяется следующая формула:

$$\text{Э}_{y_{x_i}} = b_i \frac{x_i}{\widehat{y}_{x_i \cdot x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n}}.$$

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1x_2,\ldots,x_p} = \sqrt{1 - \frac{\sigma^2_{y_{ост}}}{\sigma^2_y}}.$$

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

$$R_{yx_1x_2,\ldots,x_n} \geq r_{yx_i} \quad \left(i = \overline{1,n}\right).$$

Индекс множественной корреляции для уравнения в стандартизованном масштабе можно записать в виде

$$R_{yx_1x_2,\ldots,x_n} = \sqrt{\sum \beta_i} \cdot r_{yx_i}.$$

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

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

где $$\Delta r = \begin{vmatrix} 1 & r_{yx1} r_{yx2} & \ldots & r_{yxn} \\ r_{yx1} & 1 & r_{x1x2} \ldots & r_{x1xn} \\ \ldots & \ldots & \ldots & \ldots \\ r_{yxn} & r_{xnx1} & \ldots & 1 \end{vmatrix}$$ - определитель матрицы парных коэффициентов корреляции;

где $$\Delta r = \begin{vmatrix} 1 & r_{yx1} & r_{yx2} & \ldots & r_{yxn} \\ r_{yx1} & 1 & r_{x1x2} & \ldots & r_{x1xn} \\ r_{yx1} & r_{x2x1} & 1 & \ldots & r_{x2xn} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{yxn} & r_{xnx1} & r_{xnx2} & \ldots & 1 \end{vmatrix}$$ - определитель матрицы межфакторной корреляции;

Частные коэффициенты (или индексы) корреляции, изменяющие влияние на y фактора x_i при неизменном уровне других факторов, можно определить по формуле

$$r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_n} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \ldots x_i \ldots x_n}}{1 - R^2_{yx_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_n}}}$$

или по рекуррентной формуле:

$$r_{yx_i \cdot x_1 x_2 \ldots x_n} = \frac{r_{yx_i \cdot x_1 x_2 \ldots x_{n-1}} - r_{yx_n \cdot x_1 x_2 \ldots x_{p-1}} r_{x_i x_p \cdot x_1 x_2 \ldots x_{n-1}}}{\sqrt{\left(1 - r^2_{yx_n \cdot x_1 x_2 \ldots x_{n-1}}\right) \cdot \left(1 - r^2_{x_i x_n \cdot x_1 x_2 \ldots x_{n-1}}\right)}}.$$

Частные коэффициенты корреляции изменяются в пределах от -1 до 1.

Качество построенной модели в целом оценивается с помощью коэффициента (индекса) детерминации. **Коэффициент множественной детерминации** рассчитывается как квадрат индекса множественной корреляции: $R^2_{yx_1 x_2, \ldots, x_n}$.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и вычисляется по формуле

$$\widehat{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{(N-1)}{(N-n-1)},$$

где N - число наблюдений; n - число факторов.

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

$$F = \frac{R^2}{1 - R^2} \cdot \frac{N-n-1}{n}.$$

Частный F- критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде для фактора x_i частный F- критерий определится как

$$F_{частx_i} = \frac{R^2_{yx_1 \ldots x_i \ldots x_n} - R^2_{yx_1 \ldots x_{i-1} x_{t+1} \ldots x_n}}{1 - R^2_{yx_1 \ldots x_i . . x_n}} \cdot \frac{N - n - 1}{1}.$$

Оценка значимости коэффициентов чистой регрессии с помощью *t*-критерия Стьюдента сводится к вычислению значения

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}},$$

где m_{b_i} - средняя квадратичная ошибка коэффициента регрессии b_i, она может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_n}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \ldots x_n}}} \cdot \frac{1}{\sqrt{N - n - 1}}.$$

При построении уравнения множественной регрессии может возникнуть **проблема мультиколлинеарности факторов**, их тесной линейной связанности.

Считается, что *две переменные явно коллинеарны*, т.е. находятся между собой в линейной зависимости, если выполняется условие $r_{x_i x_j} \geq 0{,}7$.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов.

Для оценки мультиколлинеарности факторов может использоваться *определитель матрицы парных коэффициентов корреляции* между факторами.

Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиагональные элементы $r_{x_i x_j}\ (x_i \neq x_j)$ были бы равны нулю. Так, для уравнения

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \varepsilon,$$

включающего три объясняющих переменных, матрица коэффициентов корреляции между факторами будет иметь определитель, равный единице:

$$Det\left|R\right| = \begin{vmatrix} r_{x1x1} & r_{x2x1} & r_{x3x1} \\ r_{x1x2} & r_{x2x2} & r_{x3x2} \\ r_{x1x2} & r_{x2x3} & r_{x3x3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

так как $r_{x_1x_1} = r_{x_2x_2} = r_{x_3x_3} = 1$ и $r_{x_1x_2} = r_{x_1x_3} = r_{x_2x_3} = 0.$

Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю:

$$Det\left|R\right| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и тем менее надежны результаты множественной регрессии. И, наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Проверка мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных H_o: $Det|R|=1$. Доказано, что величина $\left[N-1-\frac{1}{6}(2n+5)\cdot\lg DetR\right]$ имеет приближенное распределение χ^2 *с* $(\frac{1}{2}n(n-1))$ степенями свободы. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi^2_{факт} > \chi^2_{табл(df,\alpha)}$, то гипотеза H_o отклоняется. Это означает, что $Det|R| \neq 1$, недиагональные ненулевые коэффициенты корреляции указывают на коллинеарность факторов. Мультиколлинеарность считается доказанной.

Для применения МНК требуется, чтобы дисперсия остатков была гомоскедастичной. Это значит, что для каждого значения фактора x_j остатки ε_i имеют одинаковую дисперсию. Если это условие не соблюдается, то имеет место гетероскедастичность.

При нарушении гомоскедастичности мы имеем неравенства

$$\sigma^2_{\varepsilon_i} \neq \sigma^2_{\varepsilon_j} \neq \sigma^2, \quad где \quad j \neq i.$$

При малом объеме выборки ($N \leq 12$) для оценки гетероскедастичности может использоваться метод Гольдфельда-Квандта. Основная идея теста Гольдфельда-Квандта состоит в следующем:

1) упорядочение N наблюдений по мере возрастания переменной x;

2) исключение из рассмотрения М центральных наблюдений; при этом $(N - M)$: $2 > p$, где p - число оцениваемых параметров;

3) разделение совокупности из (N - M) наблюдений на две группы (соответственно с малыми и с большими значениями фактора x) и определение по каждой из групп уравнений регрессии;

4) определение остаточной суммы квадратов для первой S_1 и второй S_2 групп и нахождение их отношения: $R = S_1 : S_2$..

При выполнении нулевой гипотезы о гомоскедастичности отношение R будет удовлетворять F-критерию со степенями свободы (N- M - $2p$):2 для каждой остаточной суммы квадратов. Чем больше величина R превышает табличное значение F-критерия, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

Уравнения множественной регрессии могут включать в качестве независимых переменных качественные признаки (например, профессия, 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.

3.2 Линейные модели множественной регрессии

Так же как при парной регрессии, линейная регрессия в данном случае является простейшей формой множественной зависимости. Ее уравнение записывается в виде

$$\tilde{y} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + \varepsilon\,, \qquad (3.7)$$

где $\tilde{y}$ - результирующий показатель (зависимая переменная);

x_j, $j = \overline{1,n}$, - независимые факторы.

Запишем уравнение (3.7) в матричной форме. Пусть *x* - матрица *N* наблюдений по n факторам (nх*N), y* - матрица-столбец наблюдений по результирующему показателю (1х N), а *A* - матрица-столбец неизвестных параметров уравнения регрессии (3.1) (1х(n+1):

$$D = \begin{vmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \ldots & x_{1,n} \\ x_{2,1} & x_{2,2} & x_{2,3} & \ldots & x_{2,n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{N,1} & x_{N,2} & x_{N,3} & \ldots & x_{N,n} \end{vmatrix}; \qquad y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_N \end{pmatrix}; \qquad A = \begin{pmatrix} a_0 \\ a_1 \\ \ldots \\ a_n \end{pmatrix},$$

где x_{ij} - *i-е* $(i = \overline{1,N})$ наблюдение по *j-му* $(j = \overline{1,n})$ фактору.

Ввиду того, что размерность матрицы-столбца *A* есть 1 х (n + 1), матрицу *x* дополним слева столбцом из единиц и обозначим $\bar{x}$:

$$\bar{x} = \begin{vmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{N,1} & x_{N,2} & \ldots & x_{N,n} \end{vmatrix}$$

Тогда в принятых обозначениях уравнение (3.7) может быть представлено в виде следующего матричного уравнения:

$$\bar{x} \cdot A = y\,. \qquad (3.8)$$

Транспонируем матрицу $\bar{x}$*,* для чего её строки сделаем столбцами, а столбцы - строками. После указанного преобразования элемент x_{ij} матрицы $\bar{x}$ становится элементом x_{ji} матрицы $\bar{x}^*$. Умножим левую и правую части уравнения (3.8) слева на матрицу $\bar{x}^*$:

$$\overline{x}*\overline{x}A=\overline{x}*y\,. \tag{3.9}$$

Уравнение (3.9) является системой нормальных уравнений, полученной на основании уравнения регрессии (3.7) и записанной в матричной форме.

Пусть $(\overline{x}*\overline{x})^{-1}$ - матрица, обратная матрице $\overline{x}*\overline{x}$. Тогда, умножив слева на эту матрицу левую и правую части уравнения (3.9), получим

$$A=(\overline{x}*\overline{x})^{-1}\overline{x}*y\,,$$

откуда следует, что коэффициенты уравнения регрессии (3.7) могут быть определены по формуле

$$a_k=\sum_{j=1}^{n}C_{kj}\sum_{i=1}^{N}y_i x_{i,j-1}, \qquad k=\overline{0,n-1}\,,$$

где C_{kj} - элемент обратной матрицы $(\overline{x}*\overline{x})^{-1}$.

Рассмотрим уравнение множественной регрессии для случая двух факторов x_1 и. x_2. Полученные на основе МНК уравнения можно перегруппировать для выражения величины коэффициента a через коэффициенты b_1 и b_2 и данные наблюдений для x и y:

$$a=\overline{y}-b_1\overline{x}_1-b_2\overline{x}_2 \tag{3.10}$$

Выражения для расчёта оценок коэффициентов b_1 и b_2 имеют вид:

$$\begin{cases} \widehat{b}_1=\dfrac{\operatorname{cov}(x_1,y)\operatorname{var}(x_2)-\operatorname{cov}(x_2,y)\operatorname{cov}(x_1,x_2)}{\operatorname{var}(x_1)\operatorname{var}(x_2)-\{\operatorname{cov}(x_1,x_2)\}^2}; \\ \widehat{b}_2=\dfrac{\operatorname{cov}(x_2,y)\operatorname{var}(x_1)-\operatorname{cov}(x_1,y)\operatorname{cov}(x_1,x_2)}{\operatorname{var}(x_1)\operatorname{var}(x_2)-\{\operatorname{cov}(x_1,x_2)\}^2} \end{cases} \tag{3.11}$$

Для оценки параметров модели множественной регрессии без свободного коэффициента также используется МНК:

$$\operatorname{var}(x)=\frac{1}{N}\sum_{i=1}^{N}[x_i-\overline{x}]^2; \qquad \overline{x}=\frac{1}{N}\sum_{i=1}^{N}x_i; \tag{3.12}$$

$$\operatorname{cov}(x,y)=\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})\cdot(y_i-\overline{y}). \tag{3.13}$$

Здесь $x_i - \textit{факторы} \quad \textit{в} \quad i-\textit{м наблюдении},\ i=1,..,N$.

В множественной регрессии в отличие от парной на зависимую случайную переменную (результирующий показатель) y воздействуют одновременно несколько n $(n>1)$ независимых факторов $x_1,x_2,\dots,x_n$.

Если, например, y - производительность производственной организации, то в качестве факторов x могут быть приняты бюджеты, выделенные на механизацию и автоматизацию производства, масштабы решения социальных вопросов, степень загрязненности окружающей среды и др.

Все факторы, рассматриваемые в множественной регрессии, должны иметь количественное выражение. Коэффициент корреляции зависимости между результирующим показателем y и каждым j-м ($j = \overline{1,n}$) фактором x_j должен быть отличен от нуля, и при проверке по нуль-гипотезе она не должна подтверждаться при p=0,9. Факторы $x_1, x_2, ..., x_n$ должны быть попарно независимыми: при проверке коэффициента корреляции зависимости между x_k и x_j ($k, j = \overline{1,n}, k \neq j$) по нуль-гипотезе она должна подтверждаться при p = 0,9.

Очевидно, что множественная регрессия позволяет создавать модели реальных производственных и социально-экономических процессов, имеющих более высокую адекватность моделируемым процессам, чем модели, разрабатываемые на основе парной регрессии. В отличие от парной регрессии, в множественной регрессии отдельно рассматриваются два вида зависимостей: линейная и нелинейная. Модели отличаются только алгоритмами построения уравнений регрессии. Общим для них является способ выбора из заданного множества факторов, попарно независимых.

Пусть на результирующий показатель у воздействуют факторы x_1, x_2, x_3 и x_4. Для каждой пары факторов определяются коэффициенты корреляции, которые примем $r_{1,2} = 0{,}85,\ r_{1,4} = 0{,}64,\ r_{2,4} = 0{,}08,\ r_{1,3} = 0{,}22,\ r_{2,3} = 0{,}75,\ r_{3,4} = 0{,}45$.

В результате проверки значений $r_{k,j}$ по нуль-гипотезе получим $\rho_{1,3} = \rho_{2,4} = 0$, а остальные $\rho_{k,j} \neq 0$. Отсюда следует, что попарно независимыми являются следующие пары факторов: x_1, x_3 и x_2, x_4.

Таким образом, в рассматриваемом случае в качестве независимых факторов могут быть взяты либо x_1 и x_3, либо x_2 и x_4. Какой группе факторов отдать предпочтение, зависит от величины совокупного воздействия каждой из них на результирующий показатель у. Берется та пара, у которой коэффициент корреляции R совокупного воздействия на **у** больше. В случае когда число независимых факторов равно n,

$$R = \sqrt{1 - \frac{D}{D_{11}}} \quad ,$$

где D – определитель вида

$$D = \begin{vmatrix} 1 & r_{yx_2} & r_{yx_2} & ... & r_{x_1x_n} \\ r_{x_1x_2} & 1 & r_{x1x1} & ... & r_{x_1x_n} \\ r_{yx_2} & r_{x_1x_2} & 1 & ... & r_{x_2x_n} \end{vmatrix}.$$

D_{11} - определитель D без первой строки и первого столбца

$$D_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_n} \\ r_{x_1x_2} & 1 & \dots & r_{x_2x_n} \\ \dots & \dots & \dots & \dots \\ r_{x_1x_2} & r_{x_2x_n} & \dots & 1 \end{vmatrix} .$$

Предположим, что в рассматриваемом примере $r_{yx_1} = 0{,}4;\ r_{yx_2} = 0{,}7;\ r_{yx_3} = 0{,}6;\ r_{yx_4} = 0{,}5.$ Тогда для пары x_1, x_3 имеем

$$D = \begin{vmatrix} 1 & 0{,}4 & 0{,}6 \\ 0{,}4 & 1 & 0{,}22 \\ 0{,}6 & 0{,}22 & 1 \end{vmatrix} = 0{,}102;\quad D_{11} = \begin{vmatrix} 1 & 0{,}22 \\ 0{,}22 & 1 \end{vmatrix} = 0{,}516;$$

$$R = \sqrt{1 - \frac{0{,}102}{0{,}516}} = 0{,}283 .$$

Тогда для пары x_2, x_4 выполняя аналогичные действия, находим

$$R = \sqrt{1 - \frac{0{,}312}{0{,}994}} = 0{,}668 .$$

Так как для пары x_2, x_4 коэффициент корреляции R больше, чем для пары x_1, x_3, то в качестве независимых факторов предпочтительнее взять x_2, x_4.

Оценка параметров уравнения линейной регрессии

В случае линейной множественной регрессии оценка значимости параметров $a_0, a_1, \dots, a_n$ производится так же, как в случае парной регрессии: по нуль-гипотезе с помощью t-критерия Стьюдента. Величина t-критерия Стьюдента $t_{расч}$ для параметра a_k, $k = \overline{0, n}$, находится по формуле

$$t_{a_k} = \frac{a_k}{S_{ост}^2 \sqrt{C_{kk}}}, \tag{3.14}$$

где C_{kk} - диагональный элемент обратной матрицы $(\bar{x} * \bar{x})^{-1}$;

$S_{ост}^2$ - остаточная дисперсия, характеризующая степень рассеяния фактических значений $y = (y_1, y_2, \dots, y_N)$ относительно расчетных значений $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N)$:

$$S_{ост}^2 = \frac{\sum_{i=1}^{N} (y_i - \tilde{y}_i)^2}{N - n - 1} . \tag{3.15}$$

Величина $t_{теор}$ находится по специальным статистическим таблицам. Если $t_{расч} \geq t_{теор}$, то нуль-гипотеза отвергается, $a_k \neq 0$, если же $t_{расч} < t_{теор}$, то нуль-гипотеза принимается, $a_k = 0$.

3.3 Нелинейные модели множественной регрессии

Построение уравнений регрессии при множественной нелинейной корреляционной зависимости с помощью аналитических методов математической статистики является сложной проблемой, и поэтому многие практически важные экономические и производственные задачи корректно (строго относительно математической статистики) не могут быть решены. Ввиду этого возникла необходимость в разработке эмпирических методов, более доступных для практического применения. Одним из таких методов и является **метод** американского экономиста **Брандона**. Он имеет достаточно простой алгоритм реализации и высокий уровень спецификации, адекватности моделируемым процессам.

Основная идея метода Брандона заключается в том, что сложная множественная корреляционная зависимость представляется в виде произведения некоторых парных корреляционных зависимостей:

$$\tilde{y}_0 = \overline{y} \cdot \prod_{k=0}^{n-1} \tilde{y}_k, \tag{3.16}$$

где $\tilde{y}_0$ - зависимая переменная в уравнении парной регрессии, построенной для случайных величин $y_{0i} = \frac{y_i}{\overline{y}}$ и x_{i1}, $i = \overline{i, N}$; здесь $\tilde{y}_k$ - зависимая переменная в уравнении парной регрессии, построенной для случайных величин

$$y_{ki} = \frac{y_{k-1,i}}{\tilde{y}_{k-1,i}} \quad \text{и} \quad x_{i,k+1}, \quad i = \overline{1, N}, \quad k = \overline{1, n-1};$$

здесь $x_{i,k+1}$ - i-е наблюдение по (k+1)-му фактору.

Алгоритм метода Брандона

1. Вычисляется среднее значение зависимой случайной переменной *y:*

$$\overline{y} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i, \quad y_i > 0.$$

2. Каждое i-е наблюдение y_i $(i = \overline{i, N})$ преобразуется по формуле

$$y_{0i} = \frac{y_i}{\overline{y}}.$$

3. Для пары переменных y_{0i} и x_{i1} $(i = \overline{1, N})$ так же, как и при парной регрессии, выбирается вид зависимости, имеющей максимальный уровень спецификации по критерию Дарбина-Уотсона и по величине корреляционного отношения η (для линейных зависимостей вместо η берется коэффициент корреляции *r*): $\tilde{y}_0 = f_1(x_1)$.

4. Вычисляются значения $\tilde{y}_{0i}$, $i = \overline{i, N}$, и $y_{1i} = \frac{y_{0,i}}{\tilde{y}_{0,i}}$, $i = \overline{1, N}$.

5. Для пары переменных y_{1i} и x_{i2} $(i=\overline{i,N})$ выбирается вид зависимости, имеющей максимальный уровень спецификации: $\tilde{y}_1 = f_2(x_2)$

Рассмотренный алгоритм определения $\tilde{y}_k$, $\mathrm{k}=\overline{0,n-1}$ выполняется до исчерпания всех n факторов, воздействующих на результирующий показатель.

После определения $\tilde{y}_{n-1} = f_n(x_n)$ строится общая формула множественной регрессии

$$\tilde{y} = \bar{y}\prod_{k=0}^{n-1}\tilde{y}_k = \bar{y}\prod_{k=0}^{n-1} f_k(x_k).$$

Спецификация множественной регрессии

Спецификация множественной линейной и нелинейной регрессии заключается в определении корреляционного отношения η и средней относительной ошибки аппроксимации δ и в оценке автокорреляции остатков $\varepsilon_i = y_i - \tilde{y}_i$ по критерию Дарбина-Уотсона. Указанные операции выполняются так же, как и для парной нелинейной регрессии.

В случае линейной множественной регрессии величина δ может не определяться. Точность аппроксимации для данной регрессии характеризуется корреляционным отношением

$$\eta = \sqrt{1-\frac{\sum\limits_{i=1}^{N}(y_i-\tilde{y}_i)^2}{\sum\limits_{i=1}^{N}(y_i-\bar{y}_i)^2}}\,. \qquad (3.17)$$

Например, если результат $\eta = 0{,}7$ означает, что корреляционная зависимость между результирующим показателем y и факторами $x_1, x_2, ..., x_n$ равна 0,7, а средняя относительная ошибка аппроксимации равна 0,3 (или 30%).

3.4. Множественная регрессия и корреляция

Для оценки параметров уравнения множественной регрессии, как отмечалось выше, применяют метод наименьших квадратов. Для линейных уравнений и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\begin{cases} \sum y = N\cdot a + b_1\cdot\sum x_1 + b_2\cdot\sum x_2 + ... + b_n\cdot\sum x_n, \\ \sum y\cdot x_1 = a\cdot\sum x_1 + b_1\cdot\sum x_1^2 + b_2\cdot\sum x_1\cdot x_2 + ... + b_n\cdot\sum x_n\cdot x_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum y\cdot x_n = a\cdot\sum x_n + b_1\cdot\sum x_1\cdot x_n + b_2\cdot\sum x_2\cdot x_n + ... + b_n\cdot\sum x_n^2. \end{cases} \qquad (3.18)$$

Для её решения может быть применен метод определителей:

$$a = \frac{\Delta a}{\Delta},\ \ b_1 = \frac{\Delta b_1}{\Delta}, \ldots ..,\ b_n = \frac{\Delta b_n}{\Delta}, \tag{3.19}$$

где $\Delta = \begin{vmatrix} N & \sum x_1 & \sum x_2 & \ldots & \sum x_n \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \ldots & \sum x_n x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \ldots & \sum x_n x_2 \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ \sum x_n & \sum x_1 x_n & \sum x_2 x_n & \ldots & \sum x_n^2 \end{vmatrix}$ - определитель системы;

$\Delta a, \Delta b_1, \ldots, \Delta b_n$ - частные определители, которые получаются путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Другой вид уравнения множественной регрессии - уравнение регрессии в стандартизованном масштабе

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots . + \beta_n t_{x_n};$$

где $t_y = \frac{y - \overline{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \overline{x_i}}{\sigma_{x_i}}$ - стандартизованные переменные;

β_i - стандартизованные коэффициенты регрессии.

К уравнению множественной регрессии в стандартизованном масштабе применим МНК. Стандартизованные коэффициенты регрессии (β - коэффициенты) определяются из следующей системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 \qquad + \beta_2 \cdot r_{x_2 x_1} + \beta_3 \cdot r_{x_3 x_1} + \ldots + \beta_n \cdot r_{x_n x_1}, \\ r_{yx_2} = \beta_1 \cdot r_{x_2 x_1} + \beta_2 \qquad + \beta_3 \cdot r_{x_3 x_2} + \ldots + \beta_n \cdot r_{x_n x_2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yx_n} = \beta_1 \cdot r_{x_n x_1} + \beta_2 \cdot r_{x_n x_2} + \beta_3 \cdot r_{x_3 x_n} + \ldots + \beta_n \end{cases}$$

Связь коэффициентов множественной регрессии b_i со стандартизованными коэффициентами β_i описывается соотношением

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \tag{3.20}$$

Параметр a определяется как a=$\overline{y} - b_1 \cdot \overline{x_1} - b_2 \cdot \overline{x_2} - \ldots - b_n \cdot \overline{x_n}$.

Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле

$$\overline{Э}_{yx_j} = b_j \frac{\overline{x_j}}{\overline{y}}. \tag{3.21}$$

Для расчета частных коэффициентов эластичности применяется следующая формула:

$\boxed{Э_{y_{x_i}} = b_i \frac{x_i}{\widehat{y}_z}}$, где $z \equiv x_i \cdot x_1, x_2, \ldots, x_{i-1,} x_{i+1}, \ldots, x_n$ - нижний индекс.

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1x_2,\ldots,x_p} = \sqrt{1 - \frac{\sigma^2_{y_{ост}}}{\sigma^2_y}}. \tag{3.22}$$

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

$$R_{yx_1x_2,\ldots,x_n} \ge r_{yx_i} \quad (i = \overline{1,n}).$$

Индекс множественной корреляции для уравнения в стандартизованном масштабе можно записать в виде

$$R_{yx_1x_2,\ldots,x_n} = \sqrt{\sum \beta_i r_{yx_i}}\,.$$

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

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

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \ldots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & \ldots & r_{x_1x_n} \\ r_{yx_2} & r_{x_2x_1} & 1 & \ldots & r_{x_2x_n} \\ & & & & \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & \ldots & 1 \end{vmatrix}$$
- определитель матрицы парных коэффициентов корреляции;

$$\Delta r = \begin{vmatrix} 1 & r_{x_1x_2} & \ldots & r_{x_1x_n} \\ r_{x_2x_1} & 1 & \ldots & r_{x_2x_n} \\ & & & \\ r_{x_nx_1} & r_{x_nx_2} & \ldots & 1 \end{vmatrix}$$
- определитель матрицы межфакторной корреляции.

Частные коэффициенты (или индексы) корреляции, изменяющие влияние на y фактора x_i при неизменном уровне других факторов, можно определить по формуле

$$r_{yx_i \cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_n} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2\ldots x_i\ldots x_n}}{1 - R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_n}}}$$

или по рекуррентной формуле:

$$r_{yx_i \cdot x_1x_2\ldots x_n} = \frac{r_{yx_i \cdot x_1x_2\ldots x_{n-1}} - r_{yx_p \cdot x_1x_2\ldots x_{n-1}} r_{x_ix_n \cdot x_1x_2\ldots x_{n-1}}}{\sqrt{\left(1 - r^2_{yx_n \cdot x_1x_2\ldots x_{n-1}}\right) \cdot \left(1 - r^2_{x_ix_n \cdot x_1x_2\ldots x_{n-1}}\right)}}.$$

Частные коэффициенты корреляции изменяются в пределах от -1 до 1.

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации. Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции: $R^2_{yx_1x_2,\ldots,x_n}$.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле

$$\hat{R}^2 = 1 - \left(1 - R^2\right)\frac{\left(N-1\right)}{\left(N-m-1\right)},$$

где N - число наблюдений; m - число факторов.

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

$$F = \frac{R^2}{1-R^2} \cdot \frac{N-m-1}{m}.$$

Частный F- критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде для фактора x_i частный F- критерий определится как

$$F_{частx_i} = \frac{R^2_{yx_1 \ldots x_i \ldots x_n} - R^2_{yx_1 \ldots x_{i-1} x_{t+1} \ldots x_n}}{1 - R^2_{yx_1 \ldots x_{i} . . x_p}} \cdot \frac{N-m-1}{1}.$$

Оценка значимости коэффициентов чистой регрессии с помощью t-критерия Стьюдента сводится к вычислению значения

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}}.$$

где m_{b_i} - средняя квадратичная ошибка коэффициента регрессии b_i, она может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_n}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \ldots x_n}}} \cdot \frac{1}{\sqrt{N-m-1}}.$$

3.5 Отбор факторов для моделей множественной регрессии

Включение в уравнение множественной регрессии того или иного набора факторов связано с анализом природы взаимосвязи моделируемого показателя с другими экономическими явлениями. Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям.

1. Факторы должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность, например, осуществить ранжирование.

2. Факторы не должны быть интеркоррелированы и тем более находиться в точной функциональной связи.

Включение в модель факторов с высокой интеркорреляцией, когда $R_{yx1} < R_{x1x2}$ для зависимости $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \varepsilon$ может привести к нежелательным последствиям - система нормальных уравнений может оказаться плохо обусловленной и повлечь за собой неустойчивость и ненадежность оценок коэффициентов регрессии.

Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результативный показатель и параметры уравнения регрессии оказываются неинтерпретируемыми. Так, в уравнении $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \varepsilon$ предполагается, что факторы x_1 и x_2 независимы друг от друга, т. е. $r_{x1x2} = 0$. Тогда можно говорить, что параметр b измеряет силу влияния фактора x_1 на результат y при неизменном значении фактора x_2. Если же $r_{x1x2} = 1$, то с изменением фактора x_1 фактор x_2 не может оставаться неизменным. Отсюда b_1 и b_2 нельзя интерпретировать как показатели раздельного влияния x_1 и x_2 и на y.

ПРИМЕР 3.4. Рассмотрим регрессию себестоимости единицы продукции (руб., y) от заработной платы работника (руб., x) и производительности его труда (единиц в час, z):

$$\boxed{y = 22\,600 - 5 \cdot x - 10 \cdot z + \varepsilon}.$$

Коэффициент регрессии при переменной z показывает, что с ростом производительности труда на 1 ед. себестоимость единицы продукции снижается в среднем на 10 руб. при постоянном уровне оплаты труда. Вместе с тем параметр при x нельзя интерпретировать как снижение себестоимости единицы продукции за счет роста заработной платы. Отрицательное значение коэффициента регрессии при переменной x в данном случае обусловлено высокой корреляцией между x и z: $r_{xz} = 0{,}95$. Поэтому роста заработной платы при неизменности производительности труда (если не брать во внимание проблемы инфляции) быть не может.

Включаемые во множественную регрессию факторы должны объяснить вариацию независимой переменной. Если строится модель с набором p факторов, то для нее рассчитывается показатель детерминации R^2, который фиксирует долю объясненной вариации результативного признака за счет рассматриваемых в регрессии p факторов. Влияние других не учтенных в модели факторов оценивается как $1 - R^2$ с соответствующей остаточной дисперсией S^2.

При дополнительном включении в регрессию $(n+1)$-го фактора коэффициент детерминации должен возрастать, а остаточная дисперсия уменьшаться: $R^2_{n+1} \geq R^2_n$ и $S^2_{n+1} \leq S^2_n$

Если же этого не происходит и данные показатели практически мало отличаются друг от друга, то включаемый в анализ фактор x_{n+1} не улучшает модель и практически является лишним фактором. Так, если для регрессии, включающей пять факторов, коэффициент детерминации составил 0,857 и включение шестого фактора дало коэффициент детерминации 0,858, то дополнительно включать в модель этот фактор нецелесообразно.

Насыщение модели лишними факторами не только не снижает величину остаточной дисперсии и не увеличивает показатель детерминации, но и приводит к статистической незначимости параметров регрессии по t-критерию Стьюдента.

Таким образом, хотя теоретически регрессионная модель позволяет учесть любое число факторов, практически в этом нет необходимости. Отбор факторов производится на основе качественного теоретико-экономического анализа. Однако теоретический анализ часто не позволяет однозначно ответить на вопрос о количественной взаимосвязи рассматриваемых признаков и целесообразности включения фактора в модель. Поэтому отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы исходя из сущности проблемы; на второй - на основе матрицы показателей корреляции определяют t-статистики для параметров регрессии.

Коэффициенты интеркорреляции (т. е. корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменных явно коллинеарны (т.е. находятся между собой в линейной зависимости), если коэффициент $r_{xixj} \geq 0{,}7$.

Поскольку одним из условий построения уравнения множественной регрессии является независимость действия факторов, т. е. $R_{xixj} = 0$, коллинеарность факторов нарушает это условие. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами. В этом требовании проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга.

Пусть, например, при изучении зависимости $y=f(x, z, v)$ матрица парных коэффициентов корреляции оказалась следующей:

Таблица 3.1

	y	x	Z	V
y	1			
x	0,8	1		
z	0,7	0,8	1	
v	0,6	0,5	0,2	1

Очевидно, что факторы x и z дублируют друг друга. В анализ целесообразно включить фактор z, а не x, так как корреляция z с результатом y слабее, чем корреляция фактора x с $y(r_{yz} < r_{yx})$, но зато слабее межфакторная корреляция $r_{zy} < r_{xv}$. Поэтому в данном случае в уравнение множественной регресии включаются факторы z, v.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов, когда более чем два фактора связаны между собой линейной зависимостью, т. е. имеет место совокупное воздействие факторов друг на друга. Наличие мультиколлинеарности факторов может означать, что некоторые факторы будут всегда действовать в унисон. В результате вариация в исходных данных перестает быть полностью независимой, и нельзя оценить воздействие каждого фактора в отдельности. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью МНК.

Если рассматривается регрессия $y = a + b\cdot x + y\cdot z + d\cdot v + \varepsilon$, то для расчета параметров, применяя МНК, предполагается равенство

$$S_y = S_{факт} + S_\varepsilon ,$$

где S_y - общая сумма квадратов отклонений $\sum\left(y_i - \overline{y}\right)^2$; $S_{факт}$ - факторная (объясненная) сумма квадратов отклонений $\sum\left(\widehat{y}_i - \overline{y}\right)^2$; S_ε - остаточная сумма квадратов отклонений $\sum\left(y_i - \widehat{y}_i\right)^2$.

В свою очередь, при независимости факторов друг от друга выполнимо равенство $S_{факт}=S_x+S_z+S_v$, где S_x, S_v, S_r - суммы квадратов отклонений, обусловленные влиянием соответствующих факторов. Если же факторы интеркоррелированы, то данное равенство нарушается.

Включение в модель мулътиколлинеарных факторов нежелательно в силу следующих последствий:

- затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл;
- оценки параметров ненадежны, обнаруживают большие стандартные ошибки и меняются с изменением выборки наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

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

Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиагональные элементы $r_{xixj}(x_i \neq x_j)$ были бы равны нулю. Так, для включающего три объясняющих переменных уравнения $y = a + b_1\cdot x_1 + b_2\cdot x_2 + b_3\cdot x_3 + \varepsilon$ матрица коэффициентов корреляции между факторами имела бы определитель, равный единице.

$$Det|R| = \begin{vmatrix} r_{x1x1} & r_{x2x1} & r_{x3x1} \\ r_{x1x2} & r_{x2x2} & r_{x3x2} \\ r_{x1x3} & r_{x2x3} & r_{x3x3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1,$$

так как $r_{x1x1} = r_{x2x2} = r_{x3x3}$ и $r_{x1x2} = r_{x1x3} = r_{x2x3} = 0$.

Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю:

$$Det\ |R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0 ,$$

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Оценка значимости мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных H_o: $Det|R|=1$. Доказано, что величина $\left[N-1-\frac{1}{6}(2m+5)\lg DetR\right]$ имеет приближенное распределение χ^2 с $\frac{1}{2}n(n-1)$ степеням свободы. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi^2_{факт}>\chi^2_{табл(df,a)}$, то гипотеза H_0 отклоняется. Это означает, что $Det|R|\neq 1$, недиагональные ненулевые коэффициенты корреляции указывают на коллинеарность факторов. Мультиколлинеарность считается доказанной.

Через коэффициенты множественной детерминации можно найти переменные, ответственные за мультиколлинеарность факторов. Для этого в качестве зависимой переменной рассматривается каждый из факторов. Чем ближе значение коэффициента множественной детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов. Сравнивая между собой коэффициенты множественной детерминации факторов

$$\left(R^2_{x_1x_2x_3\ldots x_n};R^2_{x_2x_1x_3\ldots x_n}\right),$$

можно выделить переменные, ответственные за мультиколлинеарность, следовательно, можно решать проблему отбора факторов, оставляя в уравнении факторы с минимальной величиной коэффициента множественной детерминации.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними. Например, при построении модели на основе рядов динамики переходят от первоначальных данных к первым разностям уровней $\Delta t = y_t - y_{t-1}$, чтобы исключить влияние тенденции, или используются такие методы, которые сводят к нулю межфакторную корреляцию, т. е. переходят от исходных переменных к их линейным комбинациям, не коррелированных друг с другом (метод главных компонент).

Одним из путей учета внутренней корреляции факторов является переход к совмещенным уравнениям регрессии, т. е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие. Так, если $y=f(x_1,x_2,x_3)$, то возможно построение следующего совмещенного уравнения:

$$y = a + b_1\cdot x_1 + b_2\cdot x_2 + b_3\cdot x_3 + b_{12}\cdot x_1\cdot x_2 + b_{13}\cdot x_1\cdot x_3 + b_{23}\cdot x_2\cdot x_3 + \varepsilon$$

Рассматриваемое уравнение включает взаимодействие первого порядка (взаимодействие двух факторов). Возможно включение в модель и взаимодействий более высокого порядка, если корреляции могут быть разные. Они приводят построение уравнения множественной регрессии соответственно к разным методикам. В зависимости от принятой методики построения уравнения регрессии меняется и алгоритм решения задачи на ЭВМ.

Наиболее широкое применение получили следующие методы построения уравнения множественной регрессии:

- метод исключения;
- метод включения;
- шаговый регрессионный анализ.

Каждый из этих методов по-своему решает проблему отбора факторов, давая в целом близкие результаты - отсев факторов из полного его набора (метод исключения), дополнительное введение фактора (метод включения), исключение ранее введенного фактора (шаговый регрессионный анализ).

Во многих случаях матрица парных коэффициентов корреляции играет важную роль в отборе факторов. Вместе с тем, вследствие взаимодействия факторов парные коэффициенты корреляции не могут в полной мере решать вопрос о целесообразности включения в модель того или иного фактора. Эту задачу выполняют показатели частной корреляции, оценивающие в чистом виде тесноту связи фактора с результатом. Матрица частных коэффициентов корреляции наиболее широко используется в процедуре отсева факторов. При отборе факторов рекомендуется пользоваться следующим правилом: число включаемых факторов обычно в 6-7 раз меньше объема совокупности, по которой строится регрессия. Если это соотношение нарушено, то число степеней свободы остаточной вариации очень мало. Это приводит к тому, что параметры уравнения регрессии оказываются статистически незначимыми, а *F*-критерий меньше табличного значения.

3.6 Оценка надёжности регрессионных моделей

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

$$F = \frac{D_{факт}}{D_{ост}} = \frac{R^2}{1-R^2} \cdot \frac{N-m-1}{m} \quad ,$$

$D_{факт}$ - факторная сумма квадратов на одну степень свободы;

$D_{ост}$ - остаточная сумма квадратов на одну степень свободы;

R^2 - коэффициент (индекс) множественной детерминации;

m - число параметров при переменных (в линейной регрессии совпадает с числом включенных в модель факторов);

N - число наблюдений.

ПРИМЕР 3.5. Предположим, что модель урожайности пшеницы y (ц/га) от количества внесённых минеральных удобрений x_1 на 1 га (ц) и x_2 осадков (мм) характеризуется следующим уравнением:

$$\boxed{y = -120 + 0{,}2 \cdot x_1 - 0{,}008 \cdot x_1^2 + 0{,}8 \cdot x_2 - 0{,}001 \cdot x_2^2 + \varepsilon}.$$

При этом σ_y=2, n=30, R=0,85. Результаты дисперсионного анализа оказываются следующими (табл.3.2)

Таблица 3.2

Источники информации	Число степеней свободы	Сумма квадратов SS	Дисперсия на одну степень свободы D	$F_{факт}$	$F_{табл}$
Объяснённая за счёт регрессии	4	86,7	21,675	16,27	2,76
Остаточная	25	33,3	1,332	1,00	-
Общая	29	120,0	-	-	-

$$SS_{общ} = n \cdot \sigma_y^2 = 30 \cdot 4 = 120, \quad SS_{факт} = SS_{общ} \cdot R^2 = 120 \cdot 0{,}85^2 = 86{,}7;$$

$$SS_{ость} = SS_{общ} \cdot (1 - R^2) = SS_{общ} - SS_{факт} = 120 - 86{,}7 = 33{,}3.$$

Фактическое значение F–критерия при α=0,05 превышает табличное значение, поэтом уравнение статистически значимо. Такой же результат можно получить, если воспользоваться известной формулой F–критерия:

$$F = \frac{0{,}85^2}{1 - 0{,}85^2} \cdot \frac{30 - 4 - 1}{4} = 16{,}27.$$

Оценивается значимость не только уравнения в целом, но и фактора, который дополнительно включён в регрессионную модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличить долю объяснённой вариации результативного признака. Отметим, что факторы могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного итого же фактора может быть различной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит частный F-критерий - F_{x_i}.

Частный F-критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом.

Предположим, что оценивается значимость влияния x_1 как дополнительно включенного в модель фактора. Для расчётов используют формулу:

$$Fx_1 = \frac{R^2{}_{yx1x2...xn} - R^2{}_{yx2...xn}}{1 - R^2{}_{yx1x2...xn}} \cdot \frac{N - m - 1}{1}, \tag{3.23}$$

где $R^2{}_{yx1x2...xn}$ - коэффициент множественной детерминации для модели с полным набором факторов;

$R^2{}_{yx2...xn}$ - тот же показатель, но без включения в модель фактора x_1.

N - число наблюдений; m - число параметров в модели без свободного члена.

Если оценивают значимость влияния фактора x_n после включения в модель факторов $x_1, x_2, ..., x_{n-1}$, то формула частного F-критерия примет вид:

$$Fx_n = \frac{R^2{}_{yx_1x_2...x_n} - R^2{}_{yx_1x_2...x_{n-1}}}{1 - R^2{}_{yx_1x_2...x_n}} \cdot \frac{N - m - 1}{1}. \qquad (3.24)$$

В общем случае для фактора x_i частный F-критерий определится как

$$Fx_i = \frac{R^2{}_{yx_1...x_i...x_n} - R^2{}_{yx_1...x_{i-1}x_{i+1}...x_n}}{1 - R^2{}_{yx_1x_2...x_n}} \cdot \frac{N - m - 1}{1} \qquad (3.25)$$

В числителе формул (3.23)-(3.25) показан прирост доли объяснённой вариации y за счёт дополнительного включения в модель соответствующего фактора: $R^2{}_{yx_1x_2...x_n} - R^2{}_{yx_2...x_n}$ - прирост за счёт фактора x_1; $R^2{}_{yx_1x_2...x_n} - R^2{}_{yx_1x_2...x_{n-1}}$ - прирост за счёт фактора x_n; $R^2{}_{yx_1...x_i...x_n} - R^2{}_{yx_1...x_{i-1}x_{i+1}...x_n}$ - за счёт фактора x_i.

В знаменателе указанных формул содержится доля остаточной вариации по регрессионной модели, включающей полный набор факторов. Если числитель и знаменатель Fx_i умножить на $\sum(y - \bar{y})^2$ или, что то же самое, на $N \cdot \sigma^2{}_y$, то получим соотношение прироста факторной (объяснённой) суммы квадратов отклонений к остаточной сумме квадратов. Для получения величины критерия необходимо эти суммы квадратов разделить на соответствующее число степеней свободы. Так как прирост факторной суммы квадратов отклонений обусловлен дополнительным включением в модель одного исследуемого фактора (например, x_1 или x_n), то число степеней свободы для него равно: df_1=1. для остаточной суммы квадратов отклонений по регрессионной модели число степеней свободы равно $df_2 = N - m - 1$. Соотношение числа степеней свободы приведено в формуле частного F- критерия в виде дроби: $(N - m - 1)/1$.

Фактическое значение частного F-критерия сравнивается с табличным при 5%-ном или 1%-ном уровнях значимости и числе степеней свободы 1 или ($N - m - 1$). Если фактическое значение Fx_i превышает $F\,табл(\alpha, df_1, df_2)$, то дополнительное включение фактора x_i в модель статистически оправдано и коэффициент чистой регрессии b_i при факторе x_i статистически значим.

Если же фактическое значение Fx_i меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объяснённой вариации признака y, и, следовательно, нецелесообразно его включение в модель. Коэффициент регрессии при этом факторе статистически незначим. С помощью частного F–критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор вводился бы в уравнение множественной регрессии.

3.7 Типовые задачи построения моделей множественной регрессии

ЗАДАЧА 3.1

По данным за 18 месяцев построено уравнение регрессии зависимости прибыли предприятия y (млн. руб) от цен на сырьё x_1 (тыс. руб. за 1 т) и производительности труда x_2 (ед.продукции на 1 работника):

$$y = 200 - 1{,}5 \cdot x_1 + 4{,}0 \cdot x_2 .$$

При анализе остаточных величин были использованы значения, приведённые в табл. 3.3

Таблица 3.3

№	y	x_1	x_2
1	210	800	300
2	720	1000	500
3	300	1500	600
4	950	750	700
5	540	960	450
…	…	…	…
$\sum \varepsilon_e = 10500$		$\sum (\varepsilon_e - \varepsilon_{e-1})^2 = 40000$	

<u>Требуется:</u>

1. По первым четырём позициям рассчитать параметры $\hat{y}_t$, ε_t, ε_{t-1}, $\varepsilon_t^{\ 2}$, $(\varepsilon_t - \varepsilon_{t-1})^2$.
2. Рассчитать критерий Дарбина-Уотсона.
3. Оценить полученный результат при 5% -ном уровне значимости.
4. Оценить пригодность уравнения для прогноза.

<u>Решение:</u>

Параметр $\hat{y}_t$ определяется путём подстановки фактических значений x_1 и x_2 в уравнение регрессии:

$$\boxed{\begin{aligned}
&\hat{y}_1 = 200 - 1{,}5 \cdot 800 + 4{,}0 \cdot 300 = 200;\\
&\hat{y}_2 = 200 - 1{,}5 \cdot 1000 + 4{,}0 \cdot 500 = 700;\\
&\hat{y}_3 = 200 - 1{,}5 \cdot 1500 + 4{,}0 \cdot 600 = 350;\\
&\hat{y}_4 = 200 - 1{,}5 \cdot 870 + 4{,}0 \cdot 530 = 1015;\\
&\hat{y}_5 = 200 - 1{,}5 \cdot 960 + 4{,}0 \cdot 450 = 560;\\
&\ldots \;\ldots.\;\ldots. \qquad\qquad \ldots.
\end{aligned}}$$

Остатки ε_t рассчитываются по формуле $\varepsilon_t = y_t - \hat{y}_t$.

В результате получаем: $\varepsilon_1 = 210 - 200 = 10$; $\varepsilon_2 = 720 - 700 = 20$; $\varepsilon_3 = 300 - 350 = -50$; $\varepsilon_4 = 950 - 1015 = -65$; $\varepsilon_5 = 540 - 560 = -20$; здесь ε_{t-1} - те же значения, что и ε_t, но со сдвигом на 1 месяц. Результаты вычислений оформляют в виде табл. 3.4.

Таблица 3.4

№ п/п	$\hat{y}_t$	ε_t	ε_{t-1}	$(\varepsilon_t - \varepsilon_{t-1})$	$(\varepsilon_t - \varepsilon_{t-1})^2$	$(\varepsilon_t)^2$
1	200	10	-	-	-	100
2	700	20	10	10	100	400
3	350	-50	20	-70	4900	2500
4	1015	-65	-50	-115	13225	4225
5	560	-20	-65	-85	7225	400
…						
$\sum_i$					40 000	10 500

2. Критерий Дарбина –Уотсона вычисляется по формуле

$$d = \frac{\sum (\varepsilon_t - \varepsilon_{t-1})^2}{\sum (\varepsilon_t)^2} = \frac{40000}{10500} = 3{,}81.$$

ЗАДАЧА 3.2

Пусть необходимо для известных статистических данных, представленных в табл.3.5, построить уравнение линейной регрессии индекса человеческого развития вида $\boxed{\{y^*\} = \hat{a} + \hat{b}_1 \cdot x_1 + \hat{b}_2 \cdot x_2 + \varepsilon}$ и оценить его адекватность как математической модели.

Таблица 3.5. Статистические данные для расчёта.

№ п.п.	Страна	Индекс человеческого развития, **у**	Ожидаемая продолжительность жизни при рождении в 1997 г., лет, **x₁**	Суточная калорийность питания населения, Ккал на душу, **x₂**
1	**Аргентина**	**0,27**	72,9	3136
2	**Белоруссия**	**0,63**	68,0	3101
3	**Бельгия**	**0,923**	77,2	3543
4	**Бразилия**	**0,739**	66,8	2938
5	**Великобритания**	**0,918**	77,2	3237
6	**Венгрия**	**0,795**	70,9	3402

7	**Германия**	**0,906**	77,2	3330
8	**Греция**	**0,867**	78,1	3575
9	**Дания**	**0,905**	75,7	3808
10	**Египет**	**0,616**	66,3	3289
11	**Румыния**	**0,752**	69,9	2943
12	**США**	**0,927**	76,6	3642
13	**Турция**	**0,728**	69,0	3568
14	**Украина**	**0,721**	68,8	2753
15	**Финляндия**	**0,913**	76,8	2916
16	**Франция**	**0,918**	78,1	3551
17	**Всего**	**12,528**	1169,5	52732
18	**Среднее значение**	**0,783**	73,09	3295,75

Решение:

1. Определяем средние значения параметров: независимых факторов $\bar{x}_1$ и $\bar{x}_2$ зависимой переменной $\bar{y}$ при заданном числе наблюдений N=16:

$$\bar{x}_1 = \frac{1}{N} \cdot \sum_{i=1}^{N} x_{1i}\ , \quad \bar{x}_1 = \frac{1169{,}5}{16} = 73{,}09; \qquad \bar{x}_2 = \frac{1}{N} \cdot \sum_{i=1}^{N} x_{2i}$$

$$\bar{x}_2 = \frac{52732{,}5}{16} = 3295{,}75\,.$$

$$\bar{y} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i\ , \quad \bar{y} = \frac{12{,}528}{16} = 0{,}78\,.$$

2. Определяем выборочную дисперсию зависимой переменной по следующей формуле:

$$\sigma_y^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2\ , \quad \sigma_y^2 = 0{,}02801.$$

3. Определяем выборочную дисперсию независимых факторов:

$$\sigma_{x1}^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{1i} - \bar{x}_1)^2\ , \quad \sigma_{x1}^2 = 280{,}78; \ \sigma_{x2}^2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{2i} - \bar{x}_2)^2\ , \ \sigma_{x2}^2 = 148465{,}76.$$

Следовательно, связь между факторами представляет неограниченную прямую зависимость.

4. Определяем парную ковариацию факторов:

$$\operatorname{cov}(x_1, x_2) = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{1i} - \bar{x}_1) \cdot (x_{2i} - \bar{x}_2)\,, \qquad \operatorname{cov}(x_1, x_2) = 625{,}04.$$

$$\mathrm{cov}(x_1, y) = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{1i} - \bar{x}_1) \cdot (y_i - \bar{y}), \quad \mathrm{cov}(x_1, y) = 0{,}42.$$

$$\mathrm{cov}(x_2, y) = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{2i} - \bar{x}_2) \cdot (y_i - \bar{y}), \quad \mathrm{cov}(x_1, y) = 19{,}9.$$

5. Выполняем расчёт оценок коэффициентов b_1 и b_2 множественной регрессии по формулам:

$$\widehat{b}_1 = \frac{\mathrm{cov}(x_1, y) \cdot \sigma_{x2}^2 - \mathrm{cov}(x_2, y) \cdot \mathrm{cov}(x_1, x_2)}{\sigma_{x1}^2 \cdot \sigma_{x2}^2 - [\mathrm{cov}(x_1, x_2)]^2};$$

$$\widehat{b}_2 = \frac{\mathrm{cov}(x_2, y) \cdot \sigma_{x1}^2 - \mathrm{cov}(x_1, y) \cdot \mathrm{cov}(x_1, x_2)}{\sigma_{x1}^2 \cdot \sigma_{x2}^2 - [\mathrm{cov}(x_1, x_2)]^2}.$$

$$\widehat{b}_1 = \frac{0{,}42 \cdot 148465{,}7 - 19{,}9 \cdot 625{,}04}{148465{,}7 \cdot 280{,}78 - (625{,}04)^2} = 0{,}0012;$$

$$\widehat{b}_2 = \frac{19{,}9 \cdot 280{,}78 - 0{,}42 \cdot 625{,}04}{280{,}78 \cdot 148465{,}7 - (625{,}04)^2} = 0{,}000135.$$

6. Выполняем расчёт оценки свободного коэффициента $\widehat{a}$ в уравнении множественной регрессии по формуле:

$$\widehat{a} = \widehat{y} - [\widehat{b}_1 \cdot \bar{x}_1 + \widehat{b}_2 \cdot \bar{x}_2]; \quad \widehat{a} = 0{,}78 - (0{,}00012 \cdot 73{,}09 + 0{,}000135 \cdot 3295{,}75) = 0{,}263.$$

7. Выполняем расчёт вектора $\{y^*\}$ значений зависимой переменной y на основе полученного регрессионного уравнения:

$$\{y^*\} = \widehat{a} + \widehat{b}_1 \cdot \{x_1\} + \widehat{b}_2 \cdot \{x_2\}, \quad i = 1, \ldots, N, \quad N = 16.$$

$$y_1^* = 0{,}263 + 0{,}0012 \cdot 72{,}9 + 0{,}00013 \cdot 3136 = 0{,}758;$$
$$y_2^* = 0{,}263 + 0{,}0012 \cdot 68{,}0 + 0{,}00013 \cdot 3101 = 0{,}747;$$
$$y_3^* = 0{,}263 + 0{,}0012 \cdot 77{,}2 + 0{,}00013 \cdot 3543 = 0{,}816;$$
$$y_4^* = 0{,}263 + 0{,}0012 \cdot 66{,}8 + 0{,}00013 \cdot 2938 = 0{,}725.$$
$$y_5^* = 0{,}263 + 0{,}0012 \cdot 77{,}2 + 0{,}00013 \cdot 3237 = 0{,}776.$$
$$y_6^* = 0{,}263 + 0{,}0012 \cdot 70{,}9 + 0{,}00013 \cdot 3402 = 0{,}790.$$
$$y_7^* = 0{,}263 + 0{,}0012 \cdot 77{,}2 + 0{,}00013 \cdot 3330 = 0{,}788.$$
$$y_8^* = 0{,}263 + 0{,}0012 \cdot 78{,}1 + 0{,}00013 \cdot 3575 = 0{,}821.$$
$$y_9^* = 0{,}263 + 0{,}0012 \cdot 75{,}7 + 0{,}00013 \cdot 3808 = 0{,}848.$$
$$y_{10}^* = 0{,}263 + 0{,}0012 \cdot 66{,}3 + 0{,}00013 \cdot 3289 = 0{,}770.$$
$$y_{11}^* = 0{,}263 + 0{,}0012 \cdot 69{,}9 + 0{,}00013 \cdot 2943 = 0{,}729.$$
$$y_{12}^* = 0{,}263 + 0{,}0012 \cdot 76{,}6 + 0{,}00013 \cdot 3642 = 0{,}828.$$
$$y_{13}^* = 0{,}263 + 0{,}0012 \cdot 69{,}0 + 0{,}00013 \cdot 3568 = 0{,}809.$$
$$y_{14}^* = 0{,}263 + 0{,}0012 \cdot 68{,}8 + 0{,}00013 \cdot 2753 = 0{,}703.$$
$$y_{15}^* = 0{,}263 + 0{,}0012 \cdot 76{,}8 + 0{,}00013 \cdot 2916 = 0{,}734.$$
$$y_{16}^* = 0{,}263 + 0{,}0012 \cdot 78{,}1 + 0{,}00013 \cdot 3551 = 0{,}818.$$

8. Определяем среднее значение $\bar{y}*$ зависимой переменной $y*$, $N=16$:

$$\bar{y}* = \frac{1}{N}\cdot\sum_{i=1}^{N}\{y*\};$$

$$\bar{y}* = \frac{0{,}758+0{,}747+0{,}816+0{,}725+0{,}776+0{,}79+0{,}788+0{,}821+0{,}848+0{,}77+0{,}729+0{,}828+}{16}\rightarrow$$

$$\rightarrow\frac{+0{,}809+0{,}703+0{,}734+0{,}818}{1}=0{,}778.$$

9. Определяем общую сумму квадратов отклонений по формуле:

$$TSS=\sum_{i=1}^{N}(y_i-\bar{y});$$

$$TSS=(0{,}27-0{,}78)^2+(0{,}63-0{,}78)^2+(0{,}923-0{,}78)^2+(0{,}739-0{,}78)^2+(0{,}918-0{,}78)^2+$$
$$+(0{,}795-0{,}78)^2+(0{,}906-0{,}78)^2+(0{,}867-0{,}78)^2+(0{,}905-0{,}78)^2+(0{,}616-0{,}78)^2+$$
$$+(0{,}752-0{,}78)^2+(0{,}927-0{,}78)^2+(0{,}728-0{,}78)^2+(0{,}721-0{,}78)^2+(0{,}913-0{,}78)^2+$$
$$+(0{,}918-0{,}78)^2=0{,}889.$$

10. Определяем общую сумму квадратов отклонений для рассчитанной регрессии RSS:

$$RSS=\sum_{i=1}^{N}(y_i-y_i^*);\qquad RSS=(0{,}27-0{,}758)^2+(0{,}63-0{,}747)^2+(0{,}923-0{,}816)^2+$$
$$+(0{,}739-0{,}725)^2+(0{,}918-0{,}776)^2+(0{,}795-0{,}790)^2+(0{,}906-0{,}788)^2+$$
$$+(0{,}867-0{,}821)^2+(0{,}905-0{,}848)^2+(0{,}616-0{,}770)^2+(0{,}752-0{,}729)^2+(0{,}927-0{,}828)^2+$$
$$+(0{,}728-0{,}809)^2+(0{,}721-0{,}703)^2+(0{,}913-0{,}734)^2+(0{,}918-0{,}818)^2=0{,}385786.$$

11. Вычисляем значение необъяснённой суммы квадратов отклонений

$$ESS=\sum_{i=1}^{N}(y_i^*-\bar{y}^*)^2;\qquad ESS=(0{,}758-0{,}778)^2+(0{,}747-0{,}778)^2+(0{,}816-0{,}778)^2+$$
$$+(0{,}725-0{,}778)^2+(0{,}776-0{,}778)^2+(0{,}790-0{,}778)^2+(0{,}788-0{,}788)^2+$$
$$+(0{,}821-0{,}778)^2+(0{,}848-0{,}778)^2+(0{,}770-0{,}778)^2+(0{,}729-0{,}788)^2+$$
$$+(0{,}828-0{,}778)^2+(0{,}809-0{,}778)^2+(0{,}703-0{,}778)^2+$$
$$+(0{,}734-0{,}778)^2+(0{,}818-0{,}778)^2=0{,}0277\ .$$

12. Вычисляем коэффициент детерминации $R^2=\dfrac{ESS}{TSS}$; $R^2=\dfrac{0{,}027}{0{,}889}=0{,}03$.

13. Вычисляем коэффициент корреляции между объясняющими переменными x_1 и x_2:

$$r_{x1,x2}=\frac{\operatorname{cov}(x_1,x_2)}{\sqrt{\sigma_{x1}^2\cdot\sigma_{x2}^2}};\quad r_{x1,x2}=\frac{625{,}04}{\sqrt{280{,}78\cdot148465{,}76}}=0{,}0968.$$

14. Определяем число степеней свободы по формуле:

$$l = n - m - 1;\ l = 16 - 2 - 1 = 13\ .$$

15. Определяем несмещённую оценку дисперсии случайной составляющей регрессии по формуле: $S_\varepsilon^2 = \frac{RSS}{l}$; $S_\varepsilon^2 = \frac{0{,}385786}{13} = 0{,}03$.

16. Вычисляем значения стандартных ошибок $\sigma(b_1)$ и $\sigma(b_1)$ оценок коэффициентов уравнения множественной регрессии:

$$\sigma(b_1) = \sqrt{\frac{S_\varepsilon^2}{(n \cdot \mathrm{var}(x_1) \cdot (1 - r_{x1,x2})}}\ ; \qquad \sigma(b_2) = \sqrt{\frac{S_\varepsilon^2}{(n \cdot \mathrm{var}(x_2) \cdot (1 - r_{x1,x2})}}\ ;$$

$$\sigma(b_1) = \sqrt{\frac{0{,}03}{(16 \cdot 280{,}78) \cdot (1 - 0{,}0968)}} = 0{,}0027\ ;$$

$$\sigma(b_2) = \sqrt{\frac{0{,}03}{(16 \cdot 148465{,}76) \cdot (1 - 0{,}0968)}} = 0{,}00031\ .$$

17. Результаты вычислений сводим в итоговую таблицу 3.6.

ВЫВОДЫ:

В ходе исследований статических данных экономического развития 16-ти государств, выполненным в соответствии с рекомендациями эконометрики, построена двухфакторная линейная модель индекса человеческого развития. Оценка точности регрессионной модели подтвердила её состоятельность и адекватность.

Таблица 3.6.

№ п/п	Наименование параметра	Условное обозначение	Численное значение
1	Выборочное среднее независимой переменной x_1	$\bar{x}_1$	73,09
2	Выборочное среднее независимой переменной x_2	$\bar{x}_2$	3295,75
3	Выборочная дисперсия независимой переменной x_1	$\mathrm{var}(x_1)$	280,78
4	Выборочная дисперсия независимой переменной x_2	$\mathrm{var}(x_2)$	148465,76
5	Выборочная дисперсия зависимой	$\mathrm{var}(y)$	0,028

	переменной y		
6	Выборочное среднее зависимой переменной y	$\bar{y}$	0,078
7	Ковариация переменных x_1 и x_2	$\operatorname{cov}(x_1, x_2)$	625,04
8	Ковариация переменных x_1 и y	$\operatorname{cov}(x_1, y)$	0,42
9	Ковариация переменных x_2 и y	$\operatorname{cov}(x_2, y)$	19,9
10	Оценка коэффициента регрессии b_1	$\hat{b}_1$	0,0012
11	Оценка коэффициента регрессии b_2	$\hat{b}_2$	0,00013
12	Оценка коэффициента регрессии a	$\hat{a}$	0,263
13	Выборочное среднее зависимой переменной y^* по уравнению регрессии	$\bar{y}^*$	0,778
14	Общая сумма квадратов отклонений	TSS	0,889
15	Общая сумма квадратов отклонений для рассчитанной регрессии	RSS	0,385786
16	Необъяснённая сумма квадратов отклонений	ESS	0,0277
17	Коэффициент детерминации	R^2	0,03
18	Коэффициент корреляции факторов	$r_{x1,x2}$	
19	Дисперсия остатков	S_{ε}^2	0,03
20	Стандартное отклонение оценки коэффициента b_1	$\sigma(b_1)$	0,0027
21	Стандартное отклонение оценки коэффициента b_2	$\sigma(b_2)$	0,00031

ЗАДАЧА 3.3

По 20 предприятиям региона (таблица 3.7) изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Таблица 3.7

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7,0	3,9	10,0	11	9,0	6,0	21,0
2	7,0	3,9	14,0	12	11,0	6,4	22,0
3	7,0	3,7	15,0	13	9,0	6,8	22,0
4	7,0	4,0	16,0	14	11,0	7,2	25,0
5	7,0	3,8	17,0	15	12,0	8,0	28,0
6	7,0	4,8	19,0	16	12,0	8,2	29,0
7	8,0	5,4	19,0	17	12,0	8,1	30,0
8	8,0	4,4	20,0	18	12,0	8,5	31,0
9	8,0	5,3	20,0	19	14,0	9,6	32,0
10	10,0	6,8	20,0	20	14,0	9,0	36,0

Требуется:

1. Оценить показатели вариации каждого признака и сделать вывод о возможностях применения МНК для их изучения.

2. Проанализировать линейные коэффициенты парной и частной корреляции.

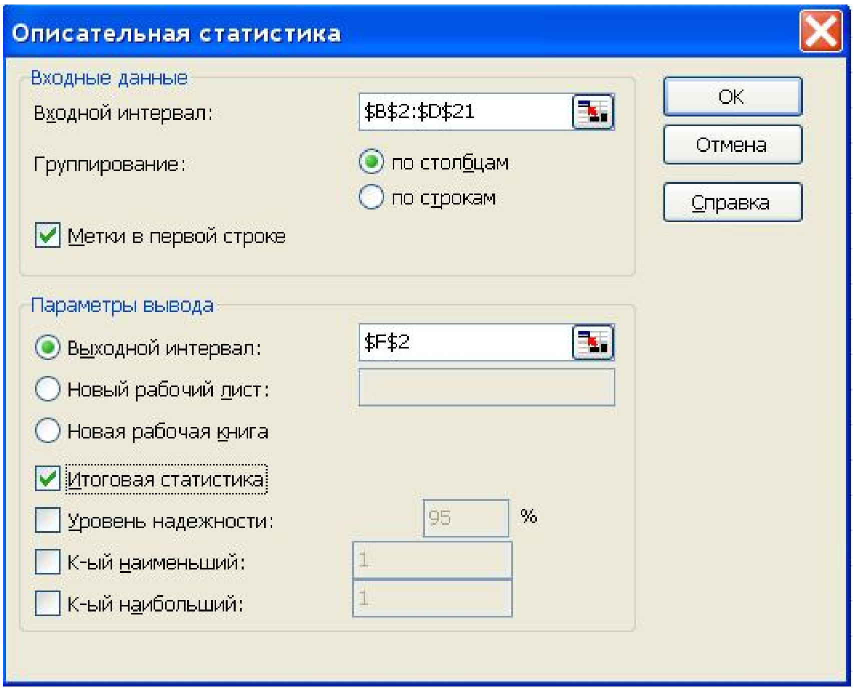
3. Написать уравнение множественной регрессии, оценить значимость его параметров, пояснить их экономический смысл.

4. С помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии и $R^2_{yx_1x_2}$. Сравнить значения скорректированного и нескорректированного линейных коэффициентов множественной детерминации.

Реализация задачи с помощью ППП Excel

1. Сводную таблицу основных статистических характеристик для одного или нескольких массивов данных можно получить с помощью инструмента анализа данных **Описательная статистика**. Для этого выполните следующие шаги:

1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;
2) в главном меню выберите последовательно пункты **Сервис/Анализ данных/Описательная статистика**, после чего щелкните по кнопке **ОК**;



Входной интервал: $B$2:$D$21

Группирование: ● по столбцам ○ по строкам

☑ Метки в первой строке

● Выходной интервал: $F$2

○ Новый рабочий лист:

○ Новая рабочая книга

☑ Итоговая статистика

☐ Уровень надежности: 95 %

☐ К-ый наименьший: 1

☐ К-ый наибольший: 1

Рисунок 3.1. Диалоговое окно ввода параметров инструмента **Описательная статистика**

3) заполните диалоговое окно ввода данных и параметров вывода (рис. 4.1):

Входной интервал – диапозон, содержащий анализируемые данные, это может быть одна или несколько строк (столбцов);

Группирование – по столбцам или по строкам – необходимо указать дополнительно;

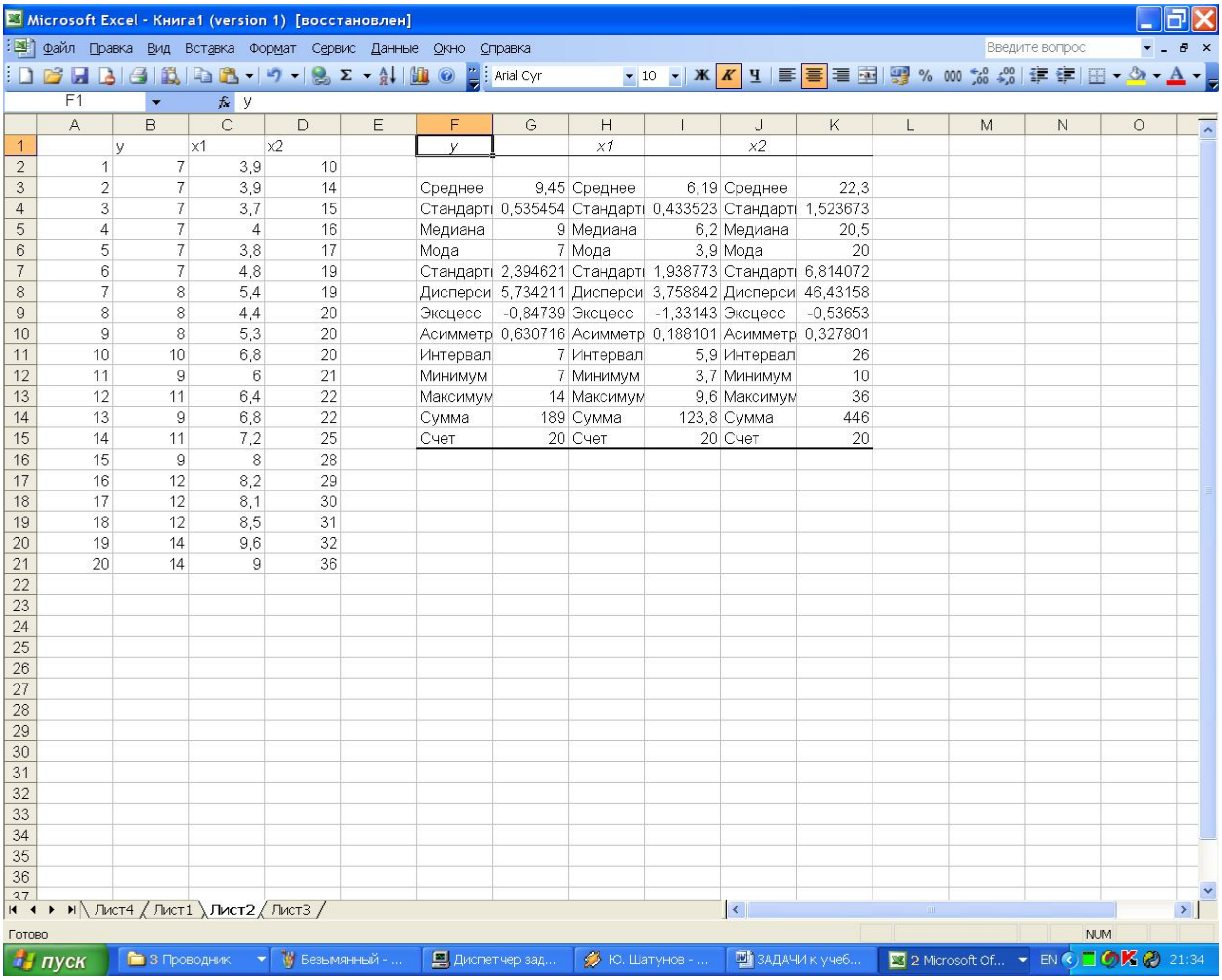
Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно задать произвольное имя нового листа.

Если необходимо получить дополнительную информацию Итоговой статистики, Уровня надежности, к-го наибольшего и наименьшего значений, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК**.

Результаты вычисления соответствующих показателей для каждого признака представлены на рис. 3.2.



	A	B	C	D	E	F	G	H	I	J	K
1		y	x1	x2		*y*		*x1*		*x2*	
2	1	7	3,9	10							
3	2	7	3,9	14		Среднее	9,45	Среднее	6,19	Среднее	22,3
4	3	7	3,7	15		Стандартı	0,535454	Стандартı	0,433523	Стандартı	1,523673
5	4	7	4	16		Медиана	9	Медиана	6,2	Медиана	20,5
6	5	7	3,8	17		Мода	7	Мода	3,9	Мода	20
7	6	7	4,8	19		Стандартı	2,394621	Стандартı	1,938773	Стандартı	6,814072
8	7	8	5,4	19		Дисперси	5,734211	Дисперси	3,758842	Дисперси	46,43158
9	8	8	4,4	20		Эксцесс	-0,84739	Эксцесс	-1,33143	Эксцесс	-0,53653
10	9	8	5,3	20		Асимметр	0,630716	Асимметр	0,188101	Асимметр	0,327801
11	10	10	6,8	20		Интервал	7	Интервал	5,9	Интервал	26
12	11	9	6	21		Минимум	7	Минимум	3,7	Минимум	10
13	12	11	6,4	22		Максимум	14	Максимум	9,6	Максимум	36
14	13	9	6,8	22		Сумма	189	Сумма	123,8	Сумма	446
15	14	11	7,2	25		Счет	20	Счет	20	Счет	20
16	15	9	8	28							
17	16	12	8,2	29							
18	17	12	8,1	30							
19	18	12	8,5	31							
20	19	14	9,6	32							
21	20	14	9	36							

Рисунок 3.2. Результат применения инструмента **Описательная статистика**

Сравнивая значения средних квадратических отклонений и средних величин и определяя коэффициенты вариации:

$$v_y = \frac{\sigma_y}{\bar{y}} \cdot 100\% = \frac{2{,}45807}{9{,}6} \cdot 100\% = 25{,}6\%$$

$$v_{x_1} = \frac{\sigma_{x_1}}{\bar{x}_1} \cdot 100\% = \frac{1{,}93877}{6{,}19} \cdot 100\% = 31{,}3\%$$

$$v_{x_2} = \frac{\sigma_{x_2}}{\bar{x}_2} \cdot 100\% = \frac{6{,}81407}{22{,}3} \cdot 100\% = 30{,}6\%$$

приходим к выводу о повышенном уровне варьирования признаков, хотя и в допустимых пределах, не превышающих 35%. Совокупность предприятий однородна, и для ее изучения могут использоваться метод наименьших квадратов и вероятностные методы оценки статистических гипотез.

2. Значения линейных коэффициентов парной корреляции определяют тесноту попарно связанных переменных, использованных в данном уравне-

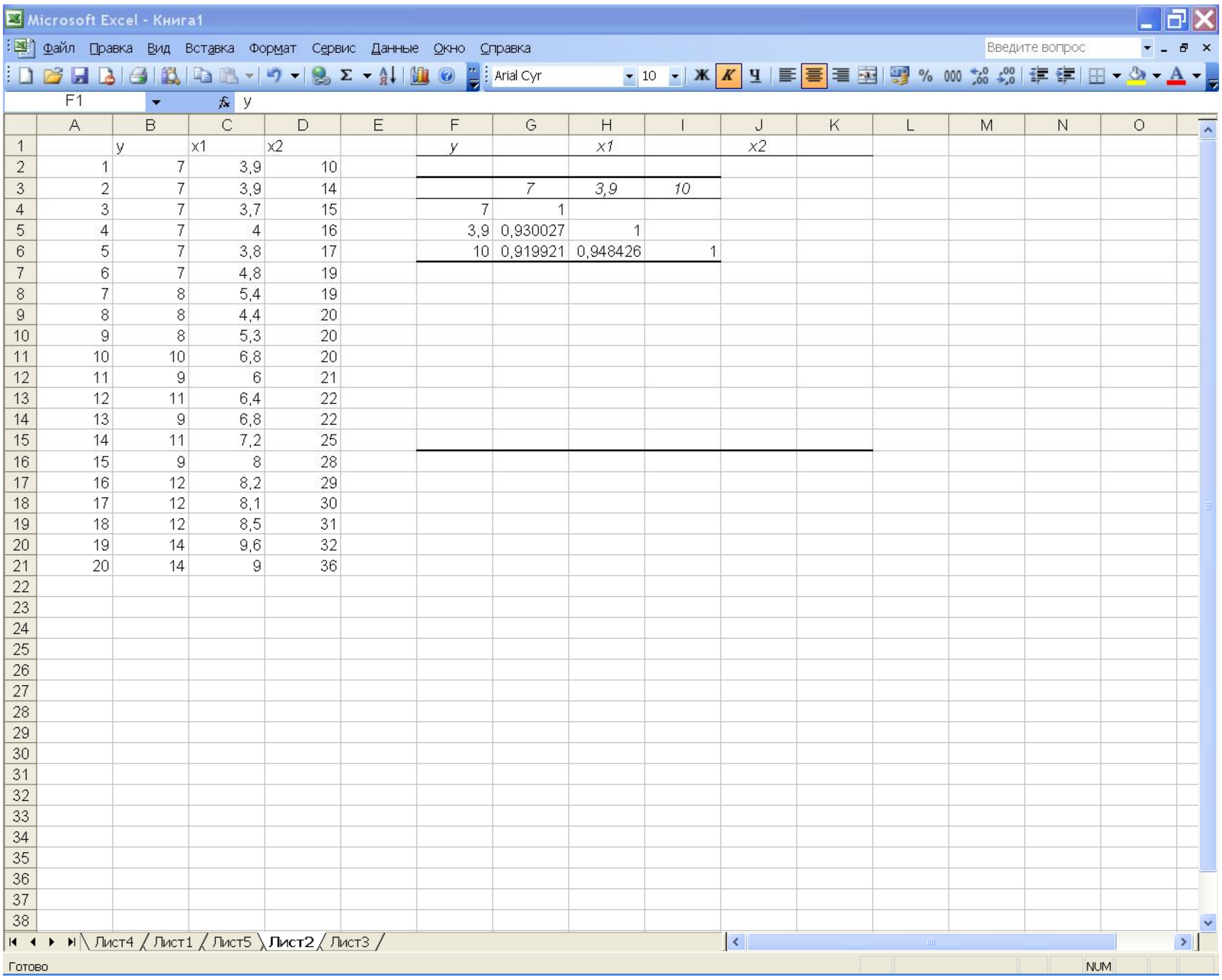
нии множественной регрессии. Линейные коэффициенты частной корреляции оценивают тесноту связи значений двух переменных, исключая влияние всех других переменных, представленных в уравнении множественной регрессии.

К сожалению, в ППП MS Excel нет специального инструмента для расчета линейных коэффициентов частной корреляции. Матрицу парных коэффициентов корреляции переменных можно рассчитать, используя инструмент анализа данных Корреляция. Для этого следует:

1) в главном меню последовательно выберите пункты Сервис/Анализ данных/Корреляция. Щелкните по кнопке ОК;

2) заполните диалоговое окно ввода данных и параметров вывода (рис. 3.1);

3) результаты вычислений – матрица коэффициентов парной корреляции – представлены на рис. 3.3.



	A	B	C	D	E	F	G	H	I	J
1		y	x1	x2		y		x1		x2
2	1	7	3,9	10						
3	2	7	3,9	14			7	3,9	10	
4	3	7	3,7	15		7	1			
5	4	7	4	16		3,9	0,930027	1		
6	5	7	3,8	17		10	0,919921	0,948426	1	
7	6	7	4,8	19						
8	7	8	5,4	19						
9	8	8	4,4	20						
10	9	8	5,3	20						
11	10	10	6,8	20						
12	11	9	6	21						
13	12	11	6,4	22						
14	13	9	6,8	22						
15	14	11	7,2	25						
16	15	9	8	28						
17	16	12	8,2	29						
18	17	12	8,1	30						
19	18	12	8,5	31						
20	19	14	9,6	32						
21	20	14	9	36						

Рисунок 3.3. Матрица коэффициентов парной корреляции

Значения коэффициентов парной корреляции указывают на весьма тесную связь выработки y как с коэффициентом обновления основных фондов - x_1, так и с долей рабочих высокой квалификации - x_2 ($r_{yx_1} = 0{,}9699$ и $r_{yx_2} = 0{,}9408$). Но в то же время межфакторная связь $r_{x_1x_2} = 0{,}9428$ весьма тесная и превышает тесноту связи x_2 с y. В связи с этим для улучшения данной модели можно исключить из нее фактор x_2 как малоинформативный, недостаточно статистически надежный.

Коэффициенты частной корреляции дают более точную характеристику тесноты связи двух признаков, чем коэффициенты парной корреляции, так как очищают парную зависимость от взаимодействия данной пары признаков с другими признаками, представленными в модели. Наиболее тесно связаны y и x_1: $r_{yx_1 \cdot x_2} = 0{,}7335$, связь y и x_2 гораздо слабее: $r_{yx_2 \cdot x_1} = 0{,}3247$, а межфакторная зависимость x_1 и x_2 выше, чем парная y и x_2: $r_{yx_2 \cdot x_1} = 0{,}3247 < r_{x_1 x_2 \cdot y} = 0{,}3679$. Все это приводит к выводу о необходимости исключить фактор x_2 - доля высококвалифицированных рабочих – из правой части уравнения множественной регрессии.

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи:

$$r_{yx_1} = 0{,}9699\,;\ r_{yx_1 \cdot x_2} = 0{,}7335\,;\ r_{yx_2} = 0{,}9408\,;\ r_{yx_2 \cdot x_1} = 0{,}3247\,.$$

Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи.

3. Вычисление параметров линейного уравнения множественной регрессии. Эта операция проводится с помощью инструмента анализа данных Регрессия. Она аналогична расчету параметров парной линейной регрессии, только в отличие от парной регрессии в диалоговом окне при заполнении параметра входной интервал X следует указать не один столбец, а все столбцы, содержащие значения факторных признаков. Результаты анализа представлены на рис. 3.4.

По результатам вычислений составим уравнение множественной регрессии вида

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2\,;$$

$$\hat{y} = 1{,}8353 + 0{,}9459 x_1 + 0{,}0856 x_2\,.$$

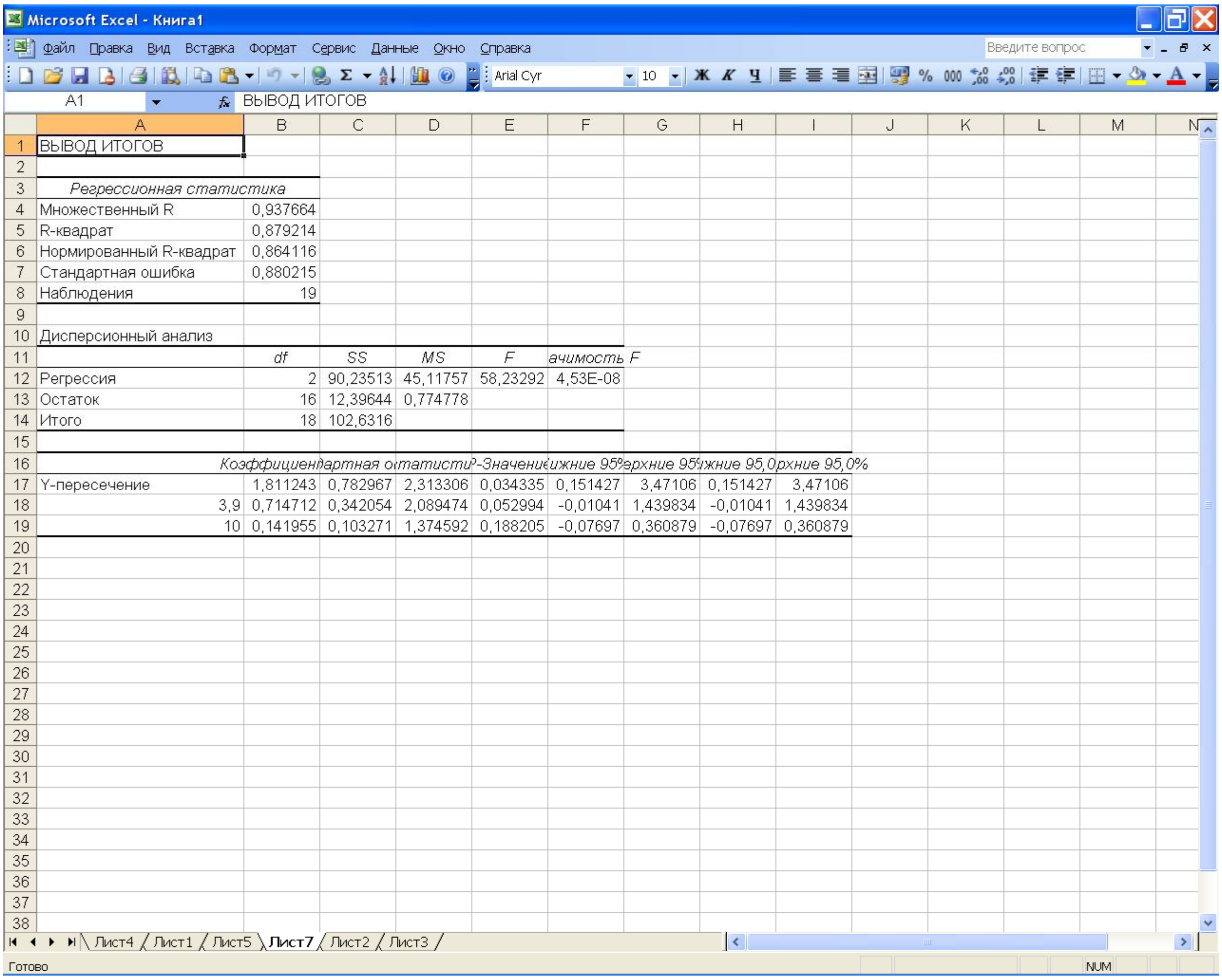
Значения случайных ошибок параметров b_0, b_1 и b_2 с учетом округления:

$$m_{b_0} = 0{,}4711;\quad m_{b_1} = 0{,}2126;\quad m_{b_2} = 0{,}0605\,.$$

Они показывают, какое значение данной характеристики сформировалось под влиянием случайных факторов. Эти значения используются для расчета t-критерия Стьюдента:

$$t_{b_0} = 3{,}90;\quad t_{b_1} = 4{,}45;\quad t_{b_2} = 1{,}42\,.$$

Если значения t-критерия больше 2-3, можно сделать вывод о существенности данного параметра, который формируется под воздействием неслучайных причин. Здесь статистически значимыми являются b_0 и b_1, а величина b_2 сформировалась под воздействием случайных причин, поэтому фактор x_2, силу влияния которого оценивает b_2, можно исключить как несущественно влияющий, неинформативный.



ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,937664
R-квадрат	0,879214
Нормированный R-квадрат	0,864116
Стандартная ошибка	0,880215
Наблюдения	19

Дисперсионный анализ

	df	SS	MS	F	ачимость F
Регрессия	2	90,23513	45,11757	58,23292	4,53E-08
Остаток	16	12,39644	0,774778		
Итого	18	102,6316			

	Коэффициен	дартная о	татисти	Р-Значени	ижние 95%	ерхние 95%	ижние 95,0	рхние 95,0%
Y-пересечение	1,811243	0,782967	2,313306	0,034335	0,151427	3,47106	0,151427	3,47106
3,9	0,714712	0,342054	2,089474	0,052994	-0,01041	1,439834	-0,01041	1,439834
10	0,141955	0,103271	1,374592	0,188205	-0,07697	0,360879	-0,07697	0,360879

Рисунок 3.4. Результат применения инструмента Регрессия

На это же указывает показатель вероятности случайных значений параметров регрессии: если α меньше принятого нами уровня (обычно 10%; 5% или 1%), делают вывод о неслучайной природе данного значения параметра, т.е. о том, что он статистически значим и надежен. В противном случае принимается гипотеза о случайной природе значения коэффициентов уравнения. Здесь $\alpha_{x_2} = 17{,}5\% > 5\%$, что позволяет рассматривать x_2 как неинформативный фактор и удалить его для улучшения данного уравнения.

Величина b_0 оценивает агрегированное влияние прочих (кроме учтенных в модели факторов x_1 и x_2) факторов на результат y.

3.8 Нестандартные задачи построения моделей множественной регрессии

ЗАДАЧА 3.4

Цель задания: провести исследования влияния динамики мировых макроэкономических показателей на индекс РТС.

Постановка задания:

Для эффективного инвестирования денежных средств в фондовый рынок Российской Федерации необходимо верно спрогнозировать направление движения фондового рынка, а для этого надо знать какие макроэкономические факторы имеют наибольшее влияние на Российский рынок ценных бумаг.

Индекс РТС является официальным индикатором Открытого акционерного общества "Фондовая биржа "Российская Торговая Система". Индекс РТС рассчитывается в течение торговой сессии при каждом изменении цены инструмента, включенного в список для его расчета.

Классический рынок РТС – старейший организованный рынок ценных бумаг России - начал работу 5 июля 1995 году.

Цены, формирующиеся на Классическом рынке РТС, являются общепризнанным ориентиром для инвесторов, совершающих операции с российскими акциями и депозитарными расписками на них как через РТС, так и на внебиржевом рынке и на других биржевых площадках России, биржах Европы и США.

Информация о торгах в РТС транслируется огромному числу потребителей в России и за рубежом и является базой расчета главного индикатора фондового рынка России – Индекса РТС.

По данным мировой федерации бирж, Россия занимает 13-е место в мире по капитализации рынка акций.

Кроме того, динамика российского рынка акций оказалась наивысшей среди всех крупных развивающихся рынков. За последние семь лет индекс РТС вырос в 13,4 раза, а индекс ММВБ - в 11,7 раза. По своему объему рынок превысил $1,322 триллиона.

Исследования влияния динамики мировых макроэкономических показателей на индекс РТС актуальна как для частных, так и для корпоративных инвесторов, ведущих свою профессиональную деятельность на фондовом рынке России.

Описание индекса РТС

Индекс рассчитывается в двух значениях – валютном и рублевом. Рублевые значения являются вспомогательными и рассчитываются на основе валютных значений. Характеристики индекса РТС представлены в таблице 3.8.

Индекс рассчитывается при возникновении следующих событий с любой ценной бумагой, входящей в список для его расчета:

а) если совершена сделка, удовлетворяющая требованиям, предъявляемым настоящей методикой, б) если цена лучшей стандартной заявки на покупку превысила последнюю рассчитанную цену по этой ценной бумаге,

в) если цена лучшей стандартной заявки на продажу стала меньше последней рассчитанной цены по этой ценной бумаге.

Таблица 3.8

Характеристики индекса РТС

Краткое наименование индекса	RTSI
Тип индекса	Взвешенный по капитализации с учетом free - float
Количество акций в списке	50 обыкновенных и привилегированных акций по состоянию на 15 июля 2005г.
Периодичность расчета	В режиме реального времени при каждом изменении цены акций
Время расчета	с 10:30 до 18:00 (Московское время)
Индекс открытия	Первое значение индекса за день
Индекс закрытия	Последнее значение индекса за день
Начало расчета	01 сентября 1995г.
Начальное значение	100
Начальная капитализация	$12 666 080 264
Формула расчета индекса	$I_n = Z_n \cdot I_1 \cdot \frac{MC_n}{MC_1}$ I_1 – начальное значение индекса, MC_n , MC_1 – сумма рыночных капитализаций акций, выраженная в долларах США, Z_n – поправочный коэффициент
Формула расчета капитализации	$MC_n = \sum_{i=1}^{N} W_i \cdot P_i \cdot Q_i \cdot C_i$ где W_i – поправочный коэффициент, учитывающий количество ценных бумаг i-го вида в свободном обращении, C_i – коэффициент, ограничивающий долю капитализации ценных бумаг i-го типа; Q_i - количество ценных бумаг соответствующего наименования, выпущенных эмитентом на текущую дату, P_i - цена i-той ценной бумаги в долларах США на расчетное время t , N - число наименований ценных бумаг в списке, по которому рассчитывается индекс.
Ограничение на вес в	15%

индексе акций одного эмитента	
Сроки изменения списка индекса	Изменения вступают в силу 15 марта, 15 июня, 15 сентября, 15 декабря

Методика расчета Индекса РТС

Индекс (I_n) рассчитывается как отношение суммарной рыночной капитализации ценных бумаг (MC_n), включенных в список для расчета индекса, к суммарной рыночной капитализации этих же ценных бумаг на начальную дату (MC_1), умноженное на значение индекса на начальную дату (I_1) и на поправочный коэффициент (Z_n):

$$I_n = Z_n \cdot I_1 \cdot \frac{MC_n}{MC_1},$$

где MC_n - сумма рыночных капитализаций акций на текущее время в долларах США:

$$MC_n = \sum_{i=1}^{N} W_i \cdot P_i \cdot Q_i \cdot C_i,$$

где W_i – поправочный коэффициент, учитывающий количество ценных бумаг i-го вида в свободном обращении,
C_i – коэффициент, ограничивающий долю капитализации ценных бумаг i-го типа;
Q_i - количество ценных бумаг соответствующего наименования, выпущенных эмитентом на текущую дату,
P_i - цена i-ой ценной бумаги в долларах США на расчетное время t,
N - число наименований ценных бумаг в списке, по которому рассчитывается индекс.

Рублевое значение индекса РТС (I_{rn}) определяется как произведение валютного значения индекса на коэффициент, рассчитанный как отношение текущего значения курса рубля к доллару США (K_n) к начальному значению (K_1):

$$I_m = I_n \cdot \frac{K_n}{K_1}.$$

Для ограничения величины влияния на индекс акций отдельных эмитентов установлено требование – доля ценных бумаг каждого эмитента в суммарной капитализации не должна превышать 15% (S_i). Для выполнения ограничения служат коэффициенты C_i.

Коэффициенты C_i пересматриваются по следующей методике. Пусть в n - й день действуют "старые" коэффициенты $C_{n,i,}$ в (n+1) день вводятся новые коэффициенты $C_{n+1,i}$. Новые коэффициенты определяются по итогам торгов дня n следующей итерационной процедурой.

Шаг 1. Рассчитываются доли рыночных капитализаций эмитентов без ограничивающих коэффициентов $C_{n,i}$:

$$S_i^{(1)} = \frac{MCap_i}{\sum_{i=1}^{N} MCap_i}, \text{ где } i = 1,\dots,N,$$

$MCap_i = \sum_k W_{ik} \cdot P_{ik} \cdot Q_{ik}$,где k – число включенных в индекс видов ценных бумаг i -го эмитента

Подсчитывается количество эмитентов, для которых $S_i^{(1)} \geq S = 15\%$. Назовем такие эмитенты ограничиваемыми. Пусть количество таких эмитентов $M^{(1).}$

Шаг 2. Рассчитывается вспомогательная величина $X^{(1)}$ из условия:

$$S = \frac{X^{(1)}}{M^{(1)} \cdot X^{(1)} + \sum_{i=1}^{N-M^{(1)}} MCap_i},$$

где сумма в знаменателе рассчитывается по эмитентам, не вошедшим в число ограничиваемых.

$$X^{(1)} = \frac{S \cdot \sum_{i=1}^{N-M^{(1)}} MCap_i}{1 - S \cdot M^{(1)}},$$

Шаг 3. Определяются доли рыночных капитализаций эмитентов

$$S_{n,i}^{(2)} = \frac{MCap_i}{\sum_{n,i}^{N} MCap_i},$$

при условии, что для каждого из ограничиваемых эмитентов $MCap_\text{i} = X^{(1)}$. Новый список ограничиваемых эмитентов включает эмитенты, для которых $S_{n,i}^{(2)} \geq S = 15\%$. Если новый список не совпадает с предыдущим (содержит больше эмитентов), то повторяем шаг 2 с новым списком ограничиваемых эмитентов. Иначе переходим к шагу 4, при этом по завершении k итераций имеется окончательный список ограничиваемых эмитентов и рассчитанная в последней итерации величина $X = X^{(k)}$.

Шаг 4. Для ограничиваемых эмитентов определяются коэффициенты

$$C_{n+1,i} = \frac{X}{P_i \cdot Q_i \cdot W_i},$$

для остальных $C_{n+1,i} = 1$.

Исходные данные:

Информация о сделках, заключенных в торговой системе во время торговой сессии, с ценными бумагами, входящими в список для расчета индекса, имеющих объем не меньший, чем объем, предъявляемый правилами торговли к заявкам по данным ценным бумагам. При этом цена сделки должна быть не ниже цены лучшей стандартной заявки на покупку и не выше цены лучшей стандартной заявки на продажу.

Определение цены i-й акции (P_i):

1. Если цена последней сделки не ниже цены лучшей стандартной заявки на покупку и не выше цены лучшей стандартной заявки на продажу, то цена ценной бумаги P_i равна цене последней сделки.
2. Если цена лучшей стандартной заявки на покупку стала выше последней рассчитанной цены по этой ценной бумаге, то цена ценной бумаги P_i равна лучшей стандартной заявки на покупку.
3. Если цена лучшей стандартной заявки на продажу стала ниже последней рассчитанной цены по этой ценной бумаге, то цена ценной бумаги P_i равна лучшей стандартной заявки на продажу.

Валюта. Все цены выражаются в долларах США. Если участники торгов устанавливают цену в рублях, то при расчете цены она пересчитывается в доллары США по курсу Центрального Банка Российской Федерации на соответствующий день.

Точность представления информации. Значения индекса рассчитываются с точностью до двух знаков после запятой. Значения цен ценных бумаг рассчитываются с точностью до пяти знаков после запятой, поправочного коэффициента Z_n – до семи знаков после запятой, коэффициентов W_i – до двух знаков после запятой, C_i – до семи знаков после запятой.

Список ценных бумаг. Список ценных бумаг, используемых для расчета индекса (далее - Список), состоит из наиболее ликвидных акций российских компаний, отобранных Информационным комитетом на основе экспертной оценки. Количество ценных бумаг в индексе не должно превышать 50.

Экспертная оценка основана на использовании следующих характеристик:

- объем торгов,
- частота заключения сделок,
- наличие спроса и предложения,
- величина спрэда,
- капитализация с учетом количества акций, находящихся в свободном обращении,

- могут учитываться также и иные факторы, влияющие на ликвидность акций.

При принятии решения о составе Списка исследуется статистическая информация о торгах за три предыдущих месяца по состоянию на даты 15 февраля, 15 мая, 15 августа, 15 ноября. ОАО "РТС" размещает на вэб-сайте следующую информацию о решениях Информационного комитета:

- перечень критериев, используемых для определения Списка,
- состав нового Списка,
- перечень ценных бумаг, являющихся кандидатами на включение в Список при следующем рассмотрении состава Списка,
- перечень ценных бумаг, являющихся кандидатами на исключение из Списка при следующем рассмотрении состава Списка.

Значения поправочных коэффициентов W_i, учитывающих количество акций i-го наименования в свободном обращении, определяются с точностью 0.05 на основании публично доступных сведений с использованием экспертной оценки.

Источниками информации о владельцах ценных бумаг являются информационные агентства, специализирующиеся на раскрытии экономической информации, и эмитенты ценных бумаг.

Коэффициенты W_i рассчитываются по формуле:

$$W_i = \frac{Q_i - Q_i^h}{Q_i},$$

где Q_i - количество выпущенных акций i-го наименования, $Q^\text{h}{}_\text{i}$ - количество акций в собственности государства, контролирующих акционеров, в перекрестном владении, во владении менеджмента и прочих стратегических инвесторов.

При изменении списка акций или поправочных коэффициентов Wi на n-дату с целью предотвращения скачка, обусловленного данными изменениями, производится перерасчет поправочного коэффициента Z_n:

$$Z_{n+1} = Z_n \cdot \frac{MC_n}{MC'_n},$$

где MC'_n – капитализация, рассчитанная по измененному списку ценных бумаг.

Учет корпоративных событий:

а) Дополнительная эмиссия акций, погашение акций

При получении информации от ФСФР России о регистрации итогов размещения новой эмиссии, либо погашения акций какой-либо компании, включенной в Список, производится перерасчет корректирующего коэффициента Z, как указано в статье 8. Сроки учета изменений определены в статье 13 настоящей методики.

б) Сплит, консолидация, выкуп части акций, выплата дивидендов акциями

Поскольку при осуществлении сплита цена акций уменьшается в цене пропорционально отношению нового количества акций к старому количеству, то величина капитализации, вычисленная как произведение текущих цен на новый объем ценных бумаг, не меняется в результате проведения компанией этого корпоративного действия. Таким образом, корректирующий коэффициент Z в этом случае не пересчитывается.

Выплата компанией дивидендов акциями воспринимается рынком также как и сплит, поэтому коэффициент Z в этом случае также не пересчитывается.

Консолидация и выкуп части акций можно интерпретировать как "сплит с обратным коэффициентом", следовательно, корректирующий коэффициент Z также не пересчитывается.

Сроки пересмотра коэффициентов и списка для расчета индекса:

Решение об изменении списка для расчета индекса и коэффициентов принимает Информационный комитет один раз в три месяца.

Расчет коэффициентов C_i осуществляется один раз в три месяца по состоянию на даты 15 февраля, 15 мая, 15 августа, 15 ноября.

Изменения списка для расчета индекса вступают в силу 15 марта, 15 июня, 15 сентября, 15 декабря.

Коэффициенты W_i могут пересматриваться при каждом изменении списка для расчета индекса по мере появления новой информации о структуре собственности эмитента ценных бумаг.

Ценные бумаги могут быть удалены из списка для расчета индекса во внеочередном порядке в случае возникновения следующих событий:

- при исключении ценной бумаги из списка ценных бумаг, допущенных к торгам;
- при появлении информации об изменении структуры собственности в акционерном обществе.

Контроль над процедурой расчета индекса:

В целях защиты индекса от ошибок, допускается перерасчет рассчитанных ранее значений индекса в следующих случаях:

- технический сбой, произошедший при расчете индекса,
- нестандартная ситуация, не связанная с изменением состояния рынка и не предусмотренная настоящей методикой, но оказавшая существенное влияние на индекс.

В случае возникновения технического сбоя перерасчет значений индекса осуществляется биржей в максимально короткие сроки с момента обнаружения сбоя. Могут быть перерассчитаны значения не старше предыдущей торговой сессии.

В случае возникновения нестандартной ситуации перерасчет значений индекса может быть осуществлен по решению Информационного комитета, которое основано на методе экспертной оценки ситуации. При этом могут быть перерассчитаны значения не более чем пяти последних торговых сессий.

При перерасчете значений индекса биржа размещает соответствующее сообщение на главной странице вэб-сайта ОАО "РТС".

Раскрытие информации об индексе осуществляется на вэб-сайте ОАО "РТС" на странице www.rts.ru/rtsindex. Помимо указанной выше информации, раскрывается методика расчета индекса, список ценных бумаг для расчета индекса, значения индекса, информация о доле каждой ценной бумаги в суммарной капитализации всех включенных в него ценных бумаг. Указанная информация за последние два года доступна любому заинтересованному лицу.

Общий контроль и внесение изменений в методику расчета индекса осуществляется Информационным комитетом РТС. Изменения в методику расчета индекса могут вноситься не чаще одного раза в квартал, сообщения о таких изменениях раскрываются на вэб-сайте РТС не позднее, чем за две недели до начала действия изменений.

Формирование писка ценных бумаг для расчета Индекса РТС

В Индекс РТС могут быть включены акции, допущенные к торгам на НП "Фондовая биржа "Российская Торговая Система".

Для анализа используются статистические данные о торгах за 3 месяца, предшествующие дате пересмотра состава индекса.

Формируется список акций, допущенных к торгам, с указанием статистических данных (таблица 3.9).

Из списка исключаются акции, не отвечающие минимальным требованиям ликвидности:

a) среднее количество компаний, чьи котировки на покупку присутствовали в торговой системе по состоянию на конец торговой сессии, не менее двух;

b) среднее количество компаний, чьи котировки на продажу присутствовали в торговой системе по состоянию на конец торговой сессии, не менее двух;

c) средняя величина спрэда между котировками на покупку и продажу (по отношению к цене покупки) по состоянию на конец торговой сессии составляет не более 15%;

d) доля дней с двусторонними котировками в системе по состоянию на конец торговой сессии должна составлять не менее 90%;

e) среднедневное количество сделок не менее 0.1;

f) среднедневной объем торгов по ценной бумаге не менее 3000 долларов США.

Таблица 3.9.

Список акций для расчета Индекса РТС

(действует с 15 декабря 2006 года по 14 марта 2007 года)

№	Код	Наименование	Количество выпущенных акций	Коэффициент учитывающий free-float (Wi)	Коэффициент ограничивающий вес акции (Ci)	Вес акции по сост. на 15.11.2006

1	AFLT	ОАО Аэрофлот	1 110 616 299	0,20	1	0,34%
2	AVAZ	ОАО Автоваз	27 194 624	0,20	1	0,25%
3	BANE	ОАО Башнефть, ао	170 169 754	0,20	1	0,41%
4	BANEP	ОАО Башнефть, ап	34 622 686	1,00	1	0,25%
5	CHMF	ОАО Северсталь, ао	930 784 663	0,15	1	1,10%
6	EESR	ОАО РАО ЕЭС России, ао	41 041 753 984	0,15	1	3,53%
7	EESRP	ОАО РАО ЕЭС России, ап	2 075 149 384	0,90	1	1,01%
8	ENCO	ОАО Сибирьтелеком, ао	12 011 401 829	0,40	1	0,29%
9	ENCOP	ОАО Сибирьтелеком, ап	3 908 420 014	1,00	1	0,18%
10	ESMO	ОАО ЦентрТелеком, ао	1 578 006 833	0,35	1	0,21%
11	GAZP	ОАО Газпром, ао	23 673 512 900	0,40	0,2110301	15,00%
12	GMKN	ОАО ГМК Норильский никель, ао	190 627 747	0,35	1	6,40%
13	IRGZ	Иркутскэнерго, ао	4 766 807 700	0,10	1	0,20%
14	IRKT	ОАО Корпорация ИРКУТ, ао	978 131 612	0,35	1	0,24%
15	LEKZ	ОАО Лебедянский, ао	20 411 300	0,20	1	0,22%
16	LKOH	ОАО ЛУКОЙЛ, ао	850 563 255	0,60	0,4990196	15,00%
17	MGNT	ОАО Магнит, ао	72 000 000	0,20	1	0,33%
18	MSNG	ОАО Мосэнерго, ао	28 249 359 700	0,10	1	0,36%
19	MTLR	ОАО Мечел, ао	416 270 745	0,15	1	0,32%
20	MTSS	ОАО МТС, ао	1 993 326 138	0,45	1	4,78%
21	NLMK	ОАО НЛМК, ао	5 993 227 240	0,15	1	1,33%
22	NNSI	ОАО ВолгаТелеком, ао	245 969 590	0,40	1	0,27%

23	NNSIP	ОАО Волга-Телеком, ап	81 983 404	1,00	1	0,18%
24	NTMK	ОАО НТМК,ао	1 310 002 966	0,10	1	0,20%
25	NVTK	ОАО НОВАТЭК, ао	3 036 306 000	0,25	1	3,10%
26	OGKC	ОАО ОГК-3, ао	29 487 999 252	0,35	1	0,66%
27	PKBA	ОАО Пивоваренная компания Балтика, ао	159 170 667	0,10	1	0,45%
28	PLZL	ОАО Полюс Золото, ао	190 627 747	0,35	1	2,17%
29	RBCI	ОАО РБК Информационные Системы, ао	119 260 000	0,20	1	0,18%
30	RITK	ОАО РИТЭК	99 750 000	0,40	1	0,28%
31	ROSN	ОАО НК Роснефть, ао	9 092 174 000	0,10	1	5,70%
32	RTKM	ОАО Ростелеком, ао	728 696 320	0,20	1	0,56%
33	RTKMP	ОАО Ростелеком, ап	242 831 469	1,00	1	0,35%
34	SBER	ОАО Сбербанк России, ао	19 000 000	0,40	1	12,49%
35	SBERP	ОАО Сбербанк России, ап	50 000 000	1,00	1	1,43%
36	SCON	ОАО Седьмой Континент, ао	75 000 000	0,25	1	0,32%
37	SIBN	ОАО Газпром нефть	4 741 299 639	0,05	1	0,68%
38	SNGS	ОАО Сургутнефтегаз, ао	35 725 994 705	0,25	1	8,21%
39	SNGSP	ОАО Сургутнефтегаз, ап	7 701 998 235	0,70	1	3,47%
40	SPTL	ОАО Северо-Западный Телеком, ао	881 045 433	0,40	1	0,27%
41	SVAV	ОАО Северсталь-авто, ао	34 270 159	0,35	1	0,21%
42	TATN	ОАО Татнефть им.В.Д.Шашина, ао	2 178 690 700	0,30	1	2,16%

43	TATNP	ОАО Татнефть им.В.Д.Шашина, ап	147 508 500	1,00	1	0,29%
44	TRNFP	ОАО АК Транснефть, ап	1 554 875	1,00	1	2,53%
45	UFNC	ОАО Уфанефтехим,ао	275 330 608	0,40	1	0,27%
46	URKA	ОАО Уралкалий, ао	2 124 390 000	0,20	1	0,45%
47	URSI	ОАО Уралсвязьинформ, ао	32 298 782 020	0,35	1	0,31%
48	URSIP	ОАО Уралсвязьинформ, ап	7 835 941 286	1,00	1	0,16%
49	VSMO	ОАО Корпорация ВСМПО-АВИСМА, ао	11 529 538	0,25	1	0,50%
50	WBDF	ОАО Вимм-Билль-Данн, ао	44 000 000	0,35	1	0,41%

Далее список акций ранжируется по величине капитализации, рассчитанной с учетом free - float .

Отбираются 50 акций с наибольшей капитализацией, которые являются кандидатами на включение в список для расчета Индекса РТС.

Из списка кандидатов исключаются акции, которые по экспертному мнению Информационного комитета имеют ограниченные рыночные перспективы.

В список кандидатов на включение в Индекс РТС вместо исключенных Информационным комитетом акций включаются акции из числа оставшихся, так, чтобы снова получить список из 50 акций.

Полученный список сравнивается с существующим списком для расчета индекса, после чего составляются списки:

a) кандидатов на включение в индекс,

b) кандидатов на исключение из состава индекса.

Информационный комитет на основании экспертной оценки принимает решение либо о включении (исключении) акций в состав (из состава) индекса, либо о включении кандидатов в листы ожидания для принятия решения при последующем пересмотре списков.

Решение задания

Для исследования индекса РТС были выбраны следующие макроэкономические показатели:

1. Индекс Dow Jones Industrial Average является одним из главных мировых фондовых индикаторов и оказывает влияние практически на все мировые фондовые рынки.

2. Динамика фьючерсных контрактов на нефть марки Brent является одним из главных индикаторов, который оказывает влияние на Российской фондовый рынок из-за высокой доли акций публичных предприятий нефтегазовой промышленности, обращающихся на рынке.

3. Золото всегда являлось удобным инструментом для хеджирования рисков, в том числе и во время спадов на Фондовых рынках.

4. Курс USD/RUR оказывает влияние не только на фондовый рынок, но и на экономику России. Внушительная часть иностранных инвестиций в долларах это заемные деньги. Соответственно укрепление курса рубля к доллару приводит к удешевлению долларовых кредитов при условии их инвестирования в рублевые активы.

Данные по индексу РТС и валютной паре USD/RUR были взяты с сайта для Хедж менеджеров www.hedging.ru [22]. Месячные котировки нефти Brent взяты с сайта инвестиционной компании Финнам [23]. Данные по индексу Dow Jones Industrial Average скачаны с официального сайта группы Доу Джонс [24]. Данные по золоту взяты с сайта инвестиционной компании Голдман Сахс [25].

В таблице 3.10 представлены макроэкономические данные. X_1, X_2, X_3, X_4 – экзогенные переменные; Y – эндогенная переменная.

Таблица 3.10

Исходные данные

месяц	Индекс РТС, пунктов (У)	Индекс Dow Jones Industrial Average,пунктов. (X1)	Нефть марки Brent, долларов за баррель. (X2)	Золото, долларов за тройскую унцию. (X3)	USD/RUR, рублей за доллар (X4)
31.01.2000	172,3086	11059,7	26,08	283	28,55
29.02.2000	170,9301	10332,1	27,18	293	28,66
31.03.2000	231,877	11244,6	29,35	276	28,46
30.04.2000	226,8656	11005,1	29,89	275	28,50
31.05.2000	190,2074	10692,7	25,74	272	28,25
30.06.2000	171,3954	10626,8	28,78	288	28,7
31.07.2000	194,0904	10727,1	31,83	275	27,8
31.08.2000	239,9919	11416	29,77	277	27,75
30.09.2000	199,0773	10923,2	31,22	274	27,75
31.10.2000	188,9985	11108,8	33,88	264	27,83
30.11.2000	143,4181	10690,2	33	269	27,85
31.12.2000	143,2926	11031	34,4	274	28,16
31.01.2001	173,5274	11072,3	28,46	264	28,37
28.02.2001	164,762	10750,2	29,58	267	28,72
31.03.2001	169,4619	9998,49	29,61	258	28,74
30.04.2001	180,6806	10973,1	27,24	263	28,83

31.05.2001	208,7959	10911,9	27,41	266	29,9
30.06.2001	216,11	10729,2	28,64	271	29,7
31.07.2001	196,12	10639,41	27,6	266	29,27
31.08.2001	205,41	9953,97	26,45	273	29,37
30.09.2001	180,25	8837,45	27,47	294	29,39
31.10.2001	204,04	9075,14	25,88	280	29,7
30.11.2001	226,49	9851,56	22,21	275	29,9
31.12.2001	256,75	10021,57	19,67	276	30,14
31.01.2002	287,53	9920	19,33	282	30,685
28.02.2002	290,75	10106,13	19,67	296	30,9274
31.03.2002	350,75	10408,92	20,74	301	31,1192
30.04.2002	386,1	9946,22	24,42	308	31,1963
31.05.2002	391,26	9925,25	26,27	326	31,3071
30.06.2002	353,79	9243,26	27	318	31,4471
31.07.2002	326,23	8736,59	25,52	305	31,4401
31.08.2002	332,9	8663,5	26,94	314	31,5673
30.09.2002	334,06	7591,93	28,38	324	31,6358
31.10.2002	358,65	8397,03	29,67	317	31,7408
30.11.2002	361,15	8896,09	28,85	319	31,8424
31.12.2002	359,07	8341,63	26,27	348	31,7844
31.01.2003	345,56	8053,81	29,42	368	31,8222
28.02.2003	383,23	7891,08	32,94	348	31,5762
31.03.2003	360,33	7992,13	35,87	335	31,3805
30.04.2003	422,37	8480,09	33,55	336	31,1
31.05.2003	467,1	8850,26	28,25	362	30,709
30.06.2003	503,51	8985,44	28,14	345	30,3483
31.07.2003	457,02	9233,8	30,72	355	30,2596
31.08.2003	530,94	9415,82	30,76	375	30,5036
30.09.2003	566,62	9275,06	31,59	383	30,6119
31.10.2003	506,12	9801,12	28,29	384	29,8584
30.11.2003	529,27	9782,46	30,33	397	29,7387
31.12.2003	567,25	10453,9	31	416	29,4545
31.01.2004	611,1	10488,07	32,15	400	28,4937
29.02.2004	670,14	10583,92	34,27	396	28,5156
31.03.2004	752,66	10357,7	34,74	425	28,4853
30.04.2004	631,11	10225,57	36,76	390	28,8834
31.05.2004	581,07	10188,45	36,69	394	28,985
30.06.2004	583,32	10435,48	40,28	396	29,0274
31.07.2004	540,27	10139,71	38	391	29,1019
31.08.2004	584,65	10173,92	40,69	406	29,2447
30.09.2004	631,65	10080,27	44,94	416	29,2171
31.10.2004	663,67	10027,47	45,95	425	28,7655
30.11.2004	627,98	10428,02	53,13	450	28,2367
31.12.2004	614,11	10783,01	48,46	435	27,7487
31.01.2005	637,21	10489,94	43,33	422	28,0845
28.02.2005	716,42	10766,23	46,84	435	27,7738
31.03.2005	669,07	10503,76	47,97	428	27,8256
30.04.2005	670,36	10192,51	54,31	436	27,7726
31.05.2005	674,44	10467,48	53	415	28,0919
30.06.2005	706,38	10274,97	49,83	437	28,6721

31.07.2005	778,93	10640,91	56,26	429	28,6341
31.08.2005	882,03	10481,6	58,7	434	28,545
30.09.2005	1007,76	10568,7	64,97	473	28,4989
31.10.2005	934,99	10440,07	65,57	470	28,4244
30.11.2005	1037,26	10805,87	62,37	495	28,7312
31.12.2005	1125,6	10717,5	58,3	510	28,7825
31.01.2006	1315,96	10868,3	59,43	570	28,1207
28.02.2006	1453,44	10994	65,51	555	28,1223
31.03.2006	1434,99	11212,69	61,63	581	27,7626
30.04.2006	1657,28	11367,14	71,73	644	27,2739
31.05.2006	1461,22	11164,31	69,48	660	26,984
30.06.2006	1494,63	11166,23	73	615	27,0789
31.07.2006	1551,09	11183	75	639	26,8718
31.08.2006	1626,69	11382	70,27	624	26,7379
30.09.2006	1549,99	11679	62,5	600	26,7799
31.10.2006	1613,57	12075,93	58,85	602	26,7477
30.11.2006	1776,68	12230	64,57	646	26,3147
31.12.2006	1921,92	12463	60,79	631	26,3311

Построение и анализ модели множественной регрессии

Для построения уравнения множественной регрессии используем линейную функцию:

$$y = b_0 + b_1 x_1 + \ldots + b_p x_p + \varepsilon .$$

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов.

Оценки параметров модели в матричной форме определяются выражением: $\widehat{B} = (X^T X)^{-1} X^T Y$,

где $\widehat{B}$ – вектор с компонентами $b_0, b_1 \ldots b_p$; X – матрица значений объясняющих переменных;Y – вектор значений зависимой переменной.

Находим оценки параметров уравнения регрессии (расчеты в таблицах 3.11, 3.12):

$$b_0 = -3224{,}094; \quad b_1 = 0{,}0915; \quad b_2 = 0{,}6941; \quad b_3 = 3{,}8578; b_4 = 47{,}2349 .$$

Уравнение множественной регрессии имеет вид:

$$\widehat{y} = -3224{,}094 + 0{,}0915 x_1 + 0{,}6941 x_2 + 3{,}8578 x_3 + 47{,}2349 x_4 \; (n{=}84, m{=}5, k{=}4).$$

Таблица 3.11

Расчет остатков

месяц	Y^	e^2	(Y-Ycp)^2	e
31.01.2000	246,0076233	5431,546029	185126,62	-73,699023
29.02.2000	223,987752	2815,114437	186314,76	-53,057652
31.03.2000	233,9351195	4,235855756	137414,81	-2,0581195
30.04.2000	210,4332274	270,0228679	141155,32	16,432373
31.05.2000	155,5939773	1198,089031	170044,59	34,613423
30.06.2000	234,6562874	4001,939872	185913,29	-63,260887
31.07.2000	153,2852142	1665,063185	166857,24	40,805186

31.08.2000	220,2264822	390,671742	131464,35	19,765418
30.09.2000	164,580545	1190,026106	162808	34,496755
31.10.2000	148,6052128	1631,617652	171043,07	40,393287
30.11.2000	129,9367315	181,7472955	210822,33	13,481368
31.12.2000	196,0149367	2779,644791	210937,6	-52,722337
31.01.2001	167,011451	42,45759087	184079,3	6,515949
28.02.2001	166,43031	2,783258204	191677,62	-1,66831
31.03.2001	63,91288033	11140,59555	187584,39	105,54902
30.04.2001	174,9606241	32,71812374	177992,38	5,7199759
31.05.2001	231,5950818	519,8026908	155059,64	-22,799182
30.06.2001	225,5783339	89,6493461	149352,9	-9,4683339
31.07.2001	177,0429477	363,9339247	165203,25	19,077052
31.08.2001	145,27228	3616,545368	157737,68	60,13772
30.09.2001	125,8051489	2964,241809	178355,9	54,444851
31.10.2001	107,0778293	9401,662553	158827,78	96,962171
30.11.2001	165,7116439	3694,008568	141437,69	60,778356
31.12.2001	194,6946	3850,872665	119592,88	62,0554
31.01.2002	234,057352	2859,324079	99251,482	53,472648
28.02.2002	316,7786003	677,4880325	97232,98	-26,0286
31.03.2002	373,56764	520,6446965	63414,336	-22,81764
30.04.2002	364,4427088	469,0382627	46860,141	21,657291
31.05.2002	438,4826122	2229,975105	44652,775	-47,222612
30.06.2002	352,3546492	2,060231786	61892,5	1,4353508
31.07.2002	254,4976749	5145,52646	76364,919	71,732325
31.08.2002	289,5258525	1881,316674	72723,005	43,374148
30.09.2002	234,3170467	9948,656726	72098,711	99,742953
31.10.2002	286,8139335	5160,420451	59497,958	71,836067
30.11.2002	344,4110141	280,19365	58284,598	16,738986
31.12.2002	401,0378813	1761,303064	59293,24	-41,967881
31.01.2003	455,8374076	12161,10663	66055,185	-110,27741
28.02.2003	354,6188293	818,5990889	48110,927	28,611171
31.03.2003	306,5006021	2897,604083	58681,202	53,829398
30.04.2003	340,1349789	6762,5987	32472,772	82,235021
31.05.2003	452,151945	223,4443492	18352,671	14,948055
30.06.2003	381,8208635	14808,24595	9813,2862	121,68914
31.07.2003	440,7185834	265,7361847	21185,394	16,301417
31.08.2003	546,078068	229,1611021	5131,148	-15,138068
30.09.2003	569,7560946	9,83508961	1292,5486	-3,1360946
31.10.2003	583,8534015	6042,481707	9302,9945	-77,733401
30.11.2003	628,0598103	9759,426613	5373,1879	-98,78981
31.12.2003	749,8189843	33331,43402	1247,646	-182,56898
31.01.2004	646,6346703	1262,712791	72,726236	-35,53467
29.02.2004	642,4771586	765,2327926	4565,4303	27,662841
31.03.2004	732,555051	404,2089735	22526,398	20,104949
30.04.2004	605,6514216	648,1392133	814,41561	25,458578
31.05.2004	622,4375744	1711,276214	462,33739	-41,367574
30.06.2004	657,244666	5464,856241	370,64074	-73,924666
31.07.2004	612,8366272	5265,915389	3881,5432	-72,566627
31.08.2004	682,4452227	9563,905582	321,19924	-97,795223
30.09.2004	714,1026861	6798,445442	845,52821	-82,452686
31.10.2004	723,3627825	3563,228277	3732,9617	-59,692782
30.11.2004	836,453748	43461,30359	645,56483	-208,47375
31.12.2004	784,7675302	29123,99261	133,1247	-170,65753
31.01.2005	720,1084506	6872,153109	1199,7888	-82,898451

28.02.2005	783,2938115	4472,106671	12961,36	-66,873812
31.03.2005	735,5107617	4414,374819	4421,9797	-66,440762
30.04.2005	739,7984482	4821,698092	4595,2086	-69,438448
31.05.2005	698,1102915	560,2827016	5165,0048	-23,670292
30.06.2005	790,5777094	7089,254271	10776,094	-84,197709
31.07.2005	795,8574539	286,5386943	31102,133	-16,927454
31.08.2005	798,0584571	7051,220022	78096,756	83,971543
30.09.2005	958,6545551	2411,344721	164177,29	49,105445
31.10.2005	932,2121357	7,716530178	110501,71	2,7778643
30.11.2005	1074,389556	1378,603962	188953,63	-37,129556
31.12.2005	1123,771365	3,343907365	273558,26	1,8286354
31.01.2006	1338,558417	510,6884287	508922,39	-22,598417
28.02.2006	1296,485139	24634,82826	723976,3	156,95486
31.03.2006	1397,109473	1434,9343	692919,67	37,880527
30.04.2006	1638,205827	363,8240902	1112408,9	19,074173
31.05.2006	1666,121842	41984,76498	737276,33	-204,90184
30.06.2006	1499,621838	24,91844302	795767,42	-4,9918376
31.07.2006	1585,348986	1173,678144	899686,34	-34,258986
31.08.2006	1536,077807	8210,56949	1048817,6	90,612193
30.09.2006	1467,249622	6845,97012	897600,81	82,740378
31.10.2006	1507,220166	11310,28715	1022116,9	106,34983
30.11.2006	1674,574475	10425,53816	1378529,5	102,10552
31.12.2006	1636,172198	81651,80618	1740679,1	285,7478
	1457,899201	**511498,2736**	**17597971**	**9,397E-09**
	S^2=	**255749,1368**		

Таблица 3.12

Нахождение параметров уравнения регрессии

матрица X

1	11059,7	26,08	283	28,55
1	10332,1	27,18	293	28,66
1	11244,6	29,35	276	28,46
1	11005,1	29,89	275	28,50
1	10692,7	25,74	272	28,25
1	10626,8	28,78	288	28,7
1	10727,1	31,83	275	27,8
1	11416	29,77	277	27,75
1	10923,2	31,22	274	27,75
1	11108,8	33,88	264	27,83
1	10690,2	33	269	27,85
1	11031	34,4	274	28,16
1	11072,3	28,46	264	28,37
1	10750,2	29,58	267	28,72
1	9998,49	29,61	258	28,74
1	10973,1	27,24	263	28,83
1	10911,9	27,41	266	29,9
1	10729,2	28,64	271	29,7
1	10639,41	27,6	266	29,27
1	9953,97	26,45	273	29,37
1	8837,45	27,47	294	29,39
1	9075,14	25,88	280	29,7
1	9851,56	22,21	275	29,9
1	10021,57	19,67	276	30,14
1	9920	19,33	282	30,685

1	10106,13	19,67	296	30,9274
1	10408,92	20,74	301	31,1192
1	9946,22	24,42	308	31,1963
1	9925,25	26,27	326	31,3071
1	9243,26	27	318	31,4471
1	8736,59	25,52	305	31,4401
1	8663,5	26,94	314	31,5673
1	7591,93	28,38	324	31,6358
1	8397,03	29,67	317	31,7408
1	8896,09	28,85	319	31,8424
1	8341,63	26,27	348	31,7844
1	8053,81	29,42	368	31,8222
1	7891,08	32,94	348	31,5762
1	7992,13	35,87	335	31,3805
1	8480,09	33,55	336	31,1
1	8850,26	28,25	362	30,709
1	8985,44	28,14	345	30,3483
1	9233,8	30,72	355	30,2596
1	9415,82	30,76	375	30,5036
1	9275,06	31,59	383	30,6119
1	9801,12	28,29	384	29,8584
1	9782,46	30,33	397	29,7387
1	10453,9	31	416	29,4545
1	10488,07	32,15	400	28,4937
1	10583,92	34,27	396	28,5156
1	10357,7	34,74	425	28,4853
1	10225,57	36,76	390	28,8834
1	10188,45	36,69	394	28,985
1	10435,48	40,28	396	29,0274
1	10139,71	38	391	29,1019
1	10173,92	40,69	406	29,2447
1	10080,27	44,94	416	29,2171
1	10027,47	45,95	425	28,7655
1	10428,02	53,13	450	28,2367
1	10783,01	48,46	435	27,7487
1	10489,94	43,33	422	28,0845
1	10766,23	46,84	435	27,7738
1	10503,76	47,97	428	27,8256
1	10192,51	54,31	436	27,7726
1	10467,48	53	415	28,0919
1	10274,97	49,83	437	28,6721
1	10640,91	56,26	429	28,6341
1	10481,6	58,7	434	28,545
1	10568,7	64,97	473	28,4989
1	10440,07	65,57	470	28,4244
1	10805,87	62,37	495	28,7312
1	10717,5	58,3	510	28,7825
1	10868,3	59,43	570	28,1207
1	10994	65,51	555	28,1223
1	11212,69	61,63	581	27,7626
1	11367,14	71,73	644	27,2739
1	11164,31	69,48	660	26,984
1	11166,23	73	615	27,0789
1	11183	75	639	26,8718

1	11382	70,27	624	26,7379
1	11679	62,5	600	26,7799
1	12075,93	58,85	602	26,7477
1	12230	64,57	646	26,3147
1	12463	60,79	631	26,3311

Для характеристики относительной силы влияния факторов на эндогенную переменную рассчитаем средние коэффициенты эластичности (расчеты в приложении):

$$Э_{yxj} = b_j \frac{\bar{x}_j}{\bar{y}}.$$

$Э_{yx1}$=1,5544; $Э_{yx2}$=0,0452; $Э_{yx3}$=2,4702; $Э_{yx4}$=2,2808.

С ростом индекса DJIA на 1% от его среднего уровня, индекс РТС увеличится на 1,5544% от своего среднего уровня

С ростом цены на нефть марки Brent на 1% от ее среднего уровня, индекс РТС увеличится на 0,0452% от своего среднего уровня.

С ростом цен на золото на 1% от его среднего уровня, индекс РТС увеличится на 2,4702% от своего среднего уровня.

С ростом курса доллара по отношению к рублю на 1% от его среднего уровня, индекс РТС увеличится на 2,2808%.

Анализируя параметры уравнение регрессии можно сделать следующие выводы: изменение индекса РТС происходит пропорционально изменению индекса DJIA, котировок нефти Brent, цены на золото и валютной пары USD/RUR.

Построение матрицы корреляции, вычисление коэффициента (индекса) множественной корреляции.

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

$$R_{yx1x2,...,xp} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

где Δr – определитель матрицы парных коэффициентов корреляции;

Δr_{11} – определитель матрицы межфакторной корреляции.

Матрица коэффициентов парной корреляции представлена в таблице 3.13.

Таблица 3.13

	y	x_1	x_2	x_3	x_4
y	1	0,49336257	0,896846771	0,98035871	-0,5773485
x_1	0,493362567	1	0,510731862	0,43297772	-0,87347753
x_2	0,896846771	0,51073186	1	0,91272134	-0,67762761
x_3	0,980358706	0,43297772	0,91272134	1	-0,56267615
x_4	-0,5773485	-0,8734775	-0,67762761	-0,56267615	1

Матрица межфакторной корреляции представлена в таблице 3.14.

Таблица 3.14

	x_1	x_2	x_3	x_4
x_1	1	0,510731862	0,43297772	-0,87347753
x_2	0,51073186	1	0,91272134	-0,67762761
x_3	0,43297772	0,91272134	1	-0,56267615
x_4	-0,8734775	-0,67762761	-0,56267615	1

Их определители соответственно равны:

$$\Delta r = 0{,}0005626; \quad \Delta r_{11} = 0{,}0193567.$$

Значение коэффициента множественной корреляции лежит в пределах от 0 до 1. $R_{yx1x2x3x4} = 0{,}9855$ означает, что совместное влияние всех факторов на индекс РТС велико.

Оценка уравнения регрессии

Оценку дисперсии ошибок найдем по формуле:

$$\sigma_e^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-m}.$$

Оценки дисперсий параметров $\hat{B}$ расположены на главной диагонали следующей матрицы: $\hat{\sigma}_{\hat{B}}^2 = \sigma_e^2 \cdot \left(X^t X\right)^{-1}$.

Будем иметь: $\sigma_e^2 = 0{,}6438$; $\hat{\sigma}_{b1}^2 = 8{,}0619$; $\hat{\sigma}_{b2}^2 = 0{,}000077$; $\hat{\sigma}_{b3}^2 = 0{,}00048$; $\hat{\sigma}_{b4}^2 = 0{,}000029$; $\hat{\sigma}_{b5}^2 = 0{,}0139$.

Построение доверительных интервалов для коэффициентов модели с выбранным уровнем значимости (α=0.05)

Проверим набор гипотез H_0: $b_0=0$; $b_1=0$; $b_2=0$; $b_3=0$; $b_4=0$.

$$t_{bi} = \left|\frac{b_i}{\sigma_{bi}}\right|.$$

$t_{b0} = 24{,}8808$; $t_{b1} = 5{,}3986$; $t_{b2} = 2{,}6958$; $t_{b3} = 3{,}5532$; $t_{b4} = 1{,}9354$.

$t_{табл.}$ (84-5,0.05)=1,99.

Так как $t_{табл} < t_{факт}$ для параметров *b0, b1, b2, b3*, то гипотеза H_0 не принимается, следовательно, параметры *b0, b1, b2, b3* статистически значимы.

Так как $t_{табл} > t_{факт}$ для параметра *b4* , то он признается статистически незначимыми.

Для построения доверительных интервалов воспользуемся формулами:

$$\Delta_{bi} = t_{табл}\sigma_{bi};$$

$$\gamma_{bi} = b_i \pm \Delta_{bi}.$$

Доверительные интервалы:

$b_0 \in (63{,}6975; 77{,}5927)$; $b_1 \in (0{,}01796; 0{,}24964)$; $b_2 \in (0{,}0383; 0{,}68715)$; $b_3 \in (0{,}0162; 0{,}102)$; $b_4 \in (-0{,}5172; 0{,}0604)$.

Все коэффициенты кроме b_4 статически значимы (ноль не входит в доверительный интервал).

Оценка тесноты связи

Качество построенной модели в целом оценивает коэффициент детерминации. Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции $R^2{}_{yx1x2x3x4} = 0{,}9853$.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле:

$$R_{abj}^2 = 1-(1-R^2)\frac{(n-1)}{(n-m-1)},$$

где n – число наблюдений; m – число факторов.

Расчеты приведены в приложении:

$R_{abj}^2 = 0{,}969 \qquad R^2 = 0{,}9709$.

Коэффициент детерминации показывает, что доля дисперсии, объясненной регрессией, в общей дисперсии результативного признака составляет 97,09%. Это характеризует высокое качество модели.

С добавлением еще одной переменной R^2 обычно увеличивается. Для компенсации такого увеличения рассчитывается скорректированный коэффициент детерминации. Так как скорректированный коэффициент детерминации уменьшился, можно предположить, что увеличение доли объясненной регрессии при добавлении новой переменной мало, и что ее добавлять нецелесообразно.

Оценка статистической надежности уравнения регрессии

Значимость уравнения множественной регрессии в целом оценивают с помощью F–критерия Фишера.

Проверим гипотезу $H_0 : b_0 = b_1 = b_2 = b_3 = b_4 = 0$.

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

$$F_{табл} = F(0.05, m-1, n-m) = F(0.05,4,79) = 2{,}47.$$

$$F_{факт} = 65{,}9.$$

Так как $F_{табл} < F_{факт}$, то H_0 отвергается, т.е. не принимается гипотеза о случайной природе оцениваемых характеристик, признается их статистическая надежность.

Исследование остатков регрессии *e*.

Построим графики зависимости остатков *e* от факторов x_1, x_2, x_3, x_4, y (рис. 3.5 – 3.)

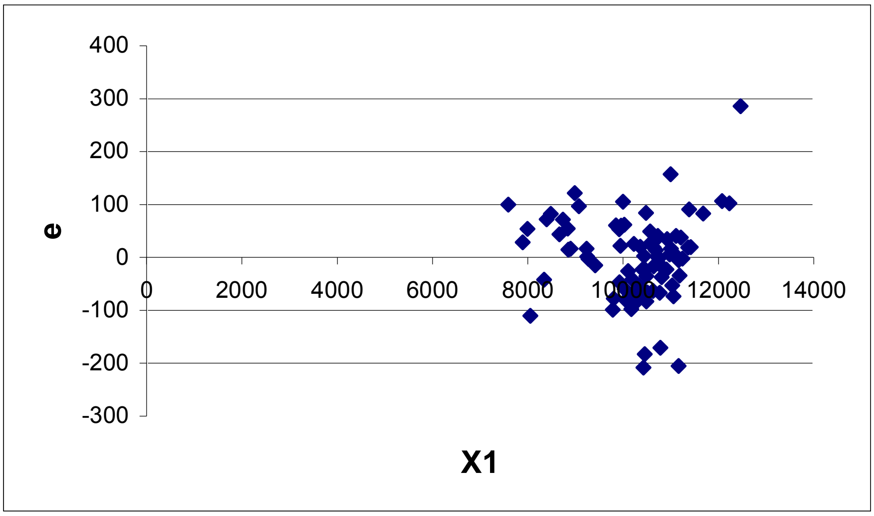


Рис. 3.5. Графическое представление зависимости остатков *e* от индекса DJIA

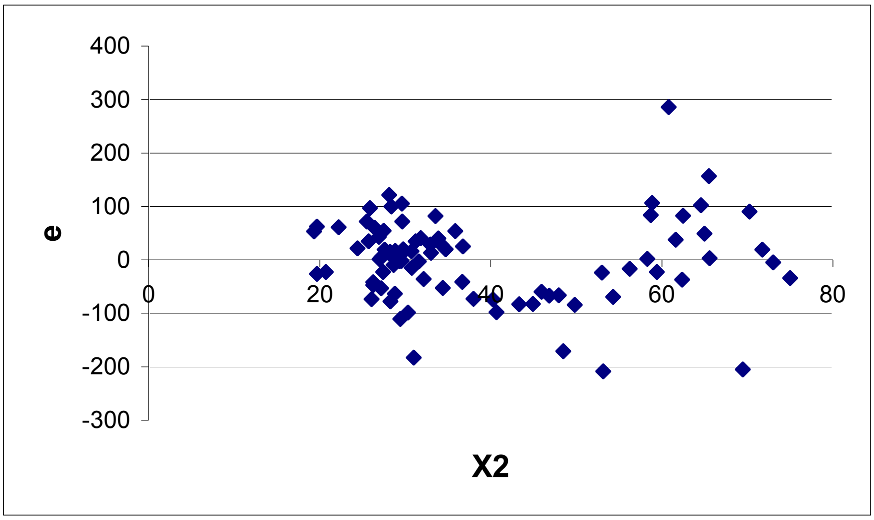


Рис. 3.6. Графическое представление зависимости остатков *e* от нефти марки Brent

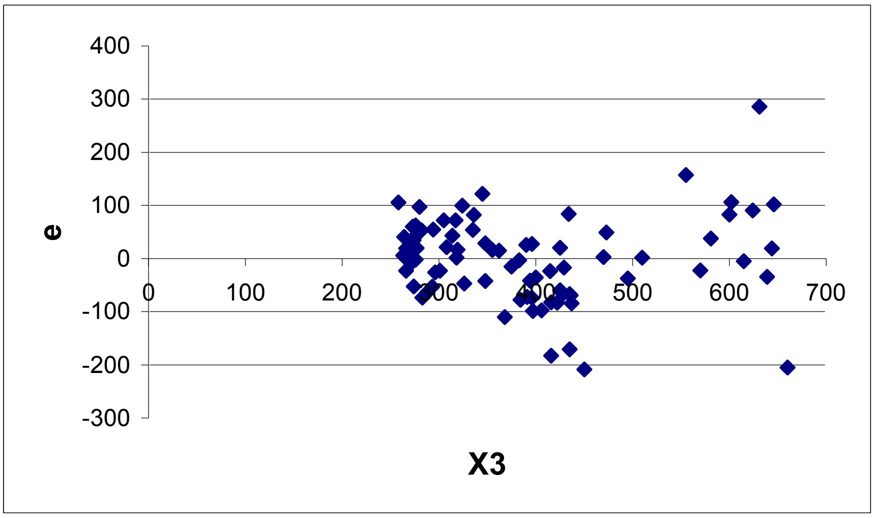


Рис. 3.7. Графическое представление зависимости остатков *e* от золота

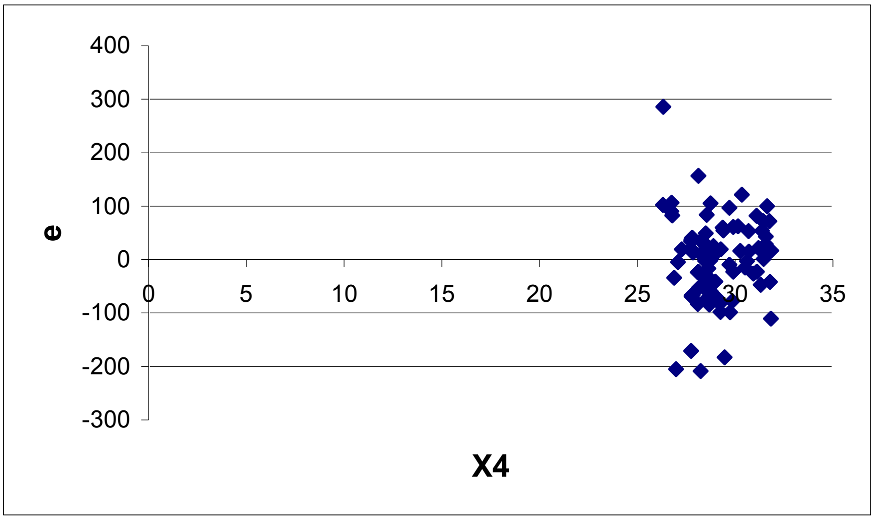


Рис. 3.8. Графическое представление зависимости остатков *e* от курса доллара по отношению к рублю

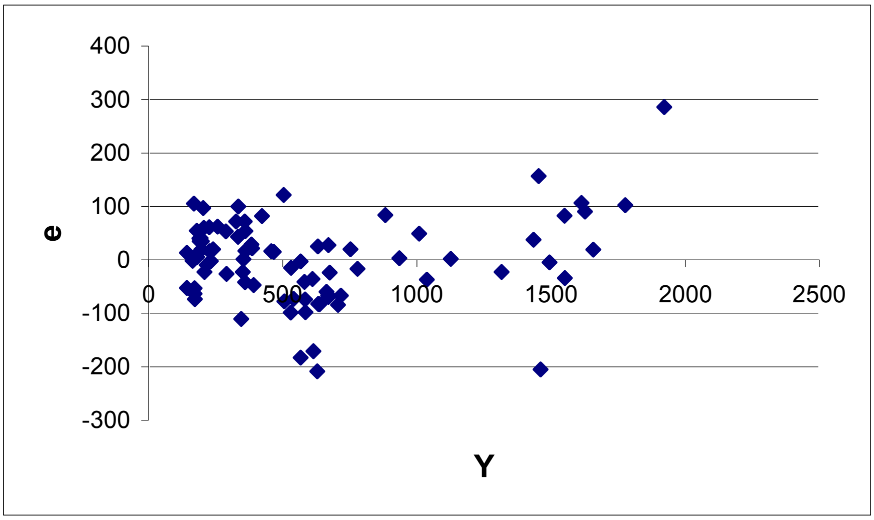


Рис. 3.9. Графическое представление зависимости остатков *e* от индекса РТС

На основании графиков можно сделать вывод о случайном распределении остатков, следовательно, можно предположить, что модель не требует корректировки.

Анализ модели множественной регрессии

Анализируя многофакторную линейную регрессионную модель, получили следующие результаты:

- изменение индекса РТС происходит пропорционально изменению индекса DJIA, котировок нефти Brent, цены на золото и валютной пары USD/RUR.
- данная модель имеет статистически незначимый коэффициент соответствующий фактору – котировки валютной пары USD/RUR;
- значение индекса множественной корреляции $R_{yx1x2x3x4} = 0{,}9853$ означает, что совместное влияние всех факторов на индекс РТС велико.
- коэффициент детерминации показывает, что доля дисперсии, объясненной регрессией, в общей дисперсии результативного признака составляет 97,09%. Это характеризует высокое качество модели;
- модель в целом значима и не требует корректировки.

Итоги исследования оказались в целом ожидаемыми за исключением одного фактора – сильного прямого влияния цены золота на индекс РТС. Поэтому чтобы рассеять сомнения относительно влияния цены золота на индекс РТС, проведем подробное исследование и построим 4 парных модели регрессии: линейную, степенную, показательную и гиперболическую.

Исследование влияния на индекс РТС динамики цен на золото

Данные по золоту взяты с сайта инвестиционной компании Голдман Сахс и представлены в таблице 3.15.

Таблица 3.15

Исходные данные и расчет линейной модели

месяц	Индекс РТС, пунктов (У)	Золото, долларов за унцию (X)	Yтеор	Y-Yт	S^2ост	X*X	(Y-Yт)/Y	mod[(Y-Yт)/Y]
31.01.2000	172,3086	283	223,1705	-50,8619	2586,935	80089	-0,29518	0,295179236
29.02.2000	170,9301	293	260,0653	-89,1352	7945,087	85849	-0,52147	0,521471756
31.03.2000	231,877	276	197,3442	34,53284	1192,517	76176	0,148927	0,148927397
30.04.2000	226,8656	275	193,6547	33,21092	1102,965	75625	0,14639	0,146390276
31.05.2000	190,2074	272	182,5862	7,621157	58,08204	73984	0,040068	0,040067618
30.06.2000	171,3954	288	241,6179	-70,2225	4931,202	82944	-0,40971	0,409710647
31.07.2000	194,0904	275	193,6547	0,435718	0,18985	75625	0,002245	0,002244922
31.08.2000	239,9919	277	201,0336	38,95826	1517,746	76729	0,162332	0,162331554
30.09.2000	199,0773	274	189,9652	9,112098	83,03032	75076	0,045772	0,045771656
31.10.2000	188,9985	264	153,0704	35,9281	1290,828	69696	0,190097	0,190097256
30.11.2000	143,4181	269	171,5178	-28,0997	789,5933	72361	-0,19593	0,195928569
31.12.2000	143,2926	274	189,9652	-46,6726	2178,332	75076	-0,32572	0,325715371
31.01.2001	173,5274	264	153,0704	20,457	418,4887	69696	0,117889	0,117889141
28.02.2001	164,762	267	164,1388	0,623157	0,388324	71289	0,003782	0,003782162
31.03.2001	169,4619	258	130,9335	38,52838	1484,436	66564	0,227357	0,227357154
30.04.2001	180,6806	263	149,3809	31,29968	979,6697	69169	0,173232	0,173232079
31.05.2001	208,7959	266	160,4494	48,34654	2337,388	70756	0,231549	0,231549262
30.06.2001	216,11	271	178,8968	37,21324	1384,825	73441	0,172196	0,172195813
31.07.2001	196,12	266	160,4494	35,67064	1272,394	70756	0,181882	0,181881687
31.08.2001	205,41	273	186,2757	19,13428	366,1206	74529	0,093152	0,093151636
30.09.2001	180,25	294	263,7548	-83,5048	6973,052	86436	-0,46327	0,463272118
31.10.2001	204,04	280	212,1021	-8,06208	64,99716	78400	-0,03951	0,039512259
30.11.2001	226,49	275	193,6547	32,83532	1078,158	75625	0,144975	0,144974691
31.12.2001	256,75	276	197,3442	59,40584	3529,054	76176	0,231376	0,231376195
31.01.2002	287,53	282	219,481	68,04896	4630,661	79524	0,236667	0,236667335
28.02.2002	290,75	296	271,1338	19,61624	384,7969	87616	0,067468	0,067467725
31.03.2002	350,75	301	289,5812	61,16884	3741,627	90601	0,174394	0,174394417
30.04.2002	386,1	308	315,4075	70,69248	4997,427	94864	0,183094	0,183093714
31.05.2002	391,26	326	381,8182	9,441846	89,14845	106276	0,024132	0,024131896
30.06.2002	353,79	318	352,3023	1,487684	2,213205	101124	0,004205	0,004204993
31.07.2002	326,23	305	304,3391	21,89092	479,2125	93025	0,067103	0,067102726
31.08.2002	332,9	314	337,5444	-4,6444	21,57042	98596	-0,01395	0,013951326
30.09.2002	334,06	324	374,4392	-40,3792	1630,479	104976	-0,12087	0,120874079
31.10.2002	358,65	317	348,6128	10,03716	100,7447	100489	0,027986	0,027985959
30.11.2002	361,15	319	355,9918	5,158204	26,60707	101761	0,014283	0,01428272
31.12.2002	359,07	348	462,9867	-103,917	10798,68	121104	-0,28941	0,289405161
31.01.2003	345,56	368	536,7763	-191,216	36563,68	135424	-0,55335	0,553351974
28.02.2003	383,23	348	462,9867	-79,7567	6361,133	121104	-0,20812	0,208117087
31.03.2003	360,33	335	415,0235	-54,6935	2991,376	112225	-0,15179	0,151787176
30.04.2003	422,37	336	418,713	3,657047	13,37399	112896	0,008658	0,008658397
31.05.2003	467,1	362	514,6394	-47,5394	2259,997	131044	-0,10178	0,101775699
30.06.2003	503,51	345	451,9183	51,59173	2661,706	119025	0,102464	0,102464158
31.07.2003	457,02	355	488,8131	-31,7931	1010,799	126025	-0,06957	0,069566037
31.08.2003	530,94	375	562,6027	-31,6627	1002,524	140625	-0,05964	0,059635113
30.09.2003	566,62	383	592,1185	-25,4985	650,1738	146689	-0,045	0,045001069
31.10.2003	506,12	384	595,808	-89,688	8043,935	147456	-0,17721	0,177206958
30.11.2003	529,27	397	643,7712	-114,501	13110,53	157609	-0,21634	0,21633802
31.12.2003	567,25	416	713,8713	-146,621	21497,82	173056	-0,25848	0,258477463

31.01.2004	611,1	400	654,8397	-43,7397	1913,158	160000	-0,07158	0,071575296
29.02.2004	670,14	396	640,0817	30,05826	903,4988	156816	0,044854	0,044853696
31.03.2004	752,66	425	747,0767	5,58334	31,17369	180625	0,007418	0,007418144
30.04.2004	631,11	390	617,9449	13,16514	173,3208	152100	0,02086	0,020860286
31.05.2004	581,07	394	632,7028	-51,6328	2665,944	155236	-0,08886	0,088858114
30.06.2004	583,32	396	640,0817	-56,7617	3221,896	156816	-0,09731	0,097308071
31.07.2004	540,27	391	621,6343	-81,3643	6620,157	152881	-0,1506	0,150599413
31.08.2004	584,65	406	676,9765	-92,3265	8524,19	164836	-0,15792	0,15791763
30.09.2004	631,65	416	713,8713	-82,2213	6760,349	173056	-0,13017	0,130169146
31.10.2004	663,67	425	747,0767	-83,4067	6956,671	180625	-0,12567	0,125674898
30.11.2004	627,98	450	839,3137	-211,334	44661,91	202500	-0,33653	0,336529278
31.12.2004	614,11	435	783,9715	-169,861	28852,91	189225	-0,2766	0,276597773
31.01.2005	637,21	422	736,0082	-98,7982	9761,088	178084	-0,15505	0,155048132
28.02.2005	716,42	435	783,9715	-67,5515	4563,199	189225	-0,09429	0,094290302
31.03.2005	669,07	428	758,1451	-89,0751	7934,373	183184	-0,13313	0,133132705
30.04.2005	670,36	436	787,6609	-117,301	13759,51	190096	-0,17498	0,174982007
31.05.2005	674,44	415	710,1819	-35,7419	1277,481	172225	-0,05299	0,052994871
30.06.2005	706,38	437	791,3504	-84,9704	7219,972	190969	-0,12029	0,120289954
31.07.2005	778,93	429	761,8346	17,09542	292,2534	184041	0,021947	0,021947314
31.08.2005	882,03	434	780,282	101,748	10352,66	188356	0,115357	0,115356645
30.09.2005	1007,76	473	924,1717	83,58831	6987,005	223729	0,082945	0,082944657
31.10.2005	934,99	470	913,1033	21,88675	479,0297	220900	0,023409	0,023408536
30.11.2005	1037,26	495	1005,34	31,91975	1018,871	245025	0,030773	0,030773144
31.12.2005	1125,6	510	1060,682	64,91755	4214,289	260100	0,057674	0,057673732
31.01.2006	1315,96	570	1282,051	33,90876	1149,804	324900	0,025767	0,02576732
28.02.2006	1453,44	555	1226,709	226,731	51406,93	308025	0,155996	0,155996092
31.03.2006	1434,99	581	1322,636	112,3545	12623,53	337561	0,078296	0,078296353
30.04.2006	1657,28	644	1555,073	102,2073	10446,32	414736	0,061672	0,061671687
31.05.2006	1461,22	660	1614,104	-152,884	23373,65	435600	-0,10463	0,10462793
30.06.2006	1494,63	615	1448,078	46,55217	2167,104	378225	0,031146	0,031146283
31.07.2006	1551,09	639	1536,625	14,46465	209,2262	408321	0,009325	0,009325476
31.08.2006	1626,69	624	1481,283	145,4069	21143,15	389376	0,089388	0,089388175
30.09.2006	1549,99	600	1392,736	157,2544	24728,94	360000	0,101455	0,101455085
31.10.2006	1613,57	602	1400,115	213,4554	45563,21	362404	0,132288	0,132287665
30.11.2006	1776,68	646	1562,452	214,2283	45893,76	417316	0,120578	0,120577872
31.12.2006	1921,92	631	1507,11	414,8105	172067,7	398161	0,215831	0,2158313
итого	50616,0507	32410	50616,05		756554	13590476		11,43752664
выб.средн	602,572032	385,83333						
выб.диспер	212023,751	13079,731						
СКО	460,460368	114,36665						
Rxy	0,91637301							
параметр b	3,68947985							
параметр a	-820,95228							
Имеем уравнение линейной регрессии вида **y=-820,95+ 3,7x**								
оц.дисп.ошиб	9226,26802							
оц.дисперсии b	0,00849863	случ.ош b	0,092188					
оц.дисперсии a	1375,00585	случ.ош a	37,08107					
Эср	2,36241351							
коэфф детерм.	0,8397395							
Acred	13,6161031							
Fрасч	429,666942							

Fтаб(0,05;1;9)	3,91							
Fтаб(0,1;1;9)	6,72							
tфакт(b)	40,0212418							
tфакт(a)	-22,139389							
пред.ош b	0,20854778							
пред.ош a	83,8847996							
дов интерв. a	-904,83708	-737,0675						
дов интерв. b	3,48093206	3,8980276						
Tтаб(0,05;9)	2,2622							
			ПРОГНОЗИРОВАНИЕ					
Хпр	424,416667			Упр	744,9245			
дисп.прогноза	10077,6714		СУММА	(Xi-Xcp)^2	1085618			
ССО прогноза	100,387606							
пред.ош. прогн	227,096841							

Исследование поля корреляции, позволит обнаружить зависимость между результирующим показателем Y и фактором X.

Таблица 3.16

месяц	**Y**	**X**	**xi-xcp**	**xi-xcp^2**	**yi-ycp**	**yi-ycp^2**	**(xi-xcp)(yi-ycp)**
31.01.2000	172,3086	283	-102,8	10574,7	-430,3	185126,6	44245,4
29.02.2000	170,9301	293	-92,8	8618,0	-431,6	186314,8	40070,8
31.03.2000	231,877	276	-109,8	12063,4	-370,7	137414,8	40714,7
30.04.2000	226,8656	275	-110,8	12284,0	-375,7	141155,3	41640,8
31.05.2000	190,2074	272	-113,8	12958,0	-412,4	170044,6	46940,8
30.06.2000	171,3954	288	-97,8	9571,4	-431,2	185913,3	42183,4
31.07.2000	194,0904	275	-110,8	12284,0	-408,5	166857,2	45273,4
31.08.2000	239,9919	277	-108,8	11844,7	-362,6	131464,4	39460,8
30.09.2000	199,0773	274	-111,8	12506,7	-403,5	162808,0	45124,2
31.10.2000	188,9985	264	-121,8	14843,4	-413,6	171043,1	50387,0
30.11.2000	143,4181	269	-116,8	13650,0	-459,2	210822,3	53644,5
31.12.2000	143,2926	274	-111,8	12506,7	-459,3	210937,6	51362,7
31.01.2001	173,5274	264	-121,8	14843,4	-429,0	184079,3	52271,9
28.02.2001	164,762	267	-118,8	14121,4	-437,8	191677,6	52026,4
31.03.2001	169,4619	258	-127,8	16341,4	-433,1	187584,4	55365,9
30.04.2001	180,6806	263	-122,8	15088,0	-421,9	177992,4	51822,3
31.05.2001	208,7959	266	-119,8	14360,0	-393,8	155059,6	47187,5
30.06.2001	216,11	271	-114,8	13186,7	-386,5	149352,9	44378,7
31.07.2001	196,12	266	-119,8	14360,0	-406,5	165203,3	48706,5
31.08.2001	205,41	273	-112,8	12731,4	-397,2	157737,7	44813,1
30.09.2001	180,25	294	-91,8	8433,4	-422,3	178355,9	38783,2
31.10.2001	204,04	280	-105,8	11200,7	-398,5	158827,8	42178,0
30.11.2001	226,49	275	-110,8	12284,0	-376,1	141437,7	41682,4
31.12.2001	256,75	276	-109,8	12063,4	-345,8	119592,9	37982,8
31.01.2002	287,53	282	-103,8	10781,4	-315,0	99251,5	32711,9
28.02.2002	290,75	296	-89,8	8070,0	-311,8	97233,0	28012,0

31.03.2002	350,75	301	-84,8	7196,7	-251,8	63414,3	21362,9
30.04.2002	386,1	308	-77,8	6058,0	-216,5	46860,1	16848,7
31.05.2002	391,26	326	-59,8	3580,0	-211,3	44652,8	12643,5
30.06.2002	353,79	318	-67,8	4601,4	-248,8	61892,5	16875,7
31.07.2002	326,23	305	-80,8	6534,0	-276,3	76364,9	22337,6
31.08.2002	332,9	314	-71,8	5160,0	-269,7	72723,0	19371,4
30.09.2002	334,06	324	-61,8	3823,4	-268,5	72098,7	16603,0
31.10.2002	358,65	317	-68,8	4738,0	-243,9	59498,0	16790,0
30.11.2002	361,15	319	-66,8	4466,7	-241,4	58284,6	16135,0
31.12.2002	359,07	348	-37,8	1431,4	-243,5	59293,2	9212,5
31.01.2003	345,56	368	-17,8	318,0	-257,0	66055,2	4583,4
28.02.2003	383,23	348	-37,8	1431,4	-219,3	48110,9	8298,4
31.03.2003	360,33	335	-50,8	2584,0	-242,2	58681,2	12314,0
30.04.2003	422,37	336	-49,8	2483,4	-180,2	32472,8	8980,1
31.05.2003	467,1	362	-23,8	568,0	-135,5	18352,7	3228,8
30.06.2003	503,51	345	-40,8	1667,4	-99,1	9813,3	4045,0
31.07.2003	457,02	355	-30,8	950,7	-145,6	21185,4	4487,9
31.08.2003	530,94	375	-10,8	117,4	-71,6	5131,1	776,0
30.09.2003	566,62	383	-2,8	8,0	-36,0	1292,5	101,9
31.10.2003	506,12	384	-1,8	3,4	-96,5	9303,0	176,8
30.11.2003	529,27	397	11,2	124,7	-73,3	5373,2	-818,5
31.12.2003	567,25	416	30,2	910,0	-35,3	1247,6	-1065,5
31.01.2004	611,1	400	14,2	200,7	8,5	72,7	120,8
29.02.2004	670,14	396	10,2	103,4	67,6	4565,4	686,9
31.03.2004	752,66	425	39,2	1534,0	150,1	22526,4	5878,4
30.04.2004	631,11	390	4,2	17,4	28,5	814,4	118,9
31.05.2004	581,07	394	8,2	66,7	-21,5	462,3	-175,6
30.06.2004	583,32	396	10,2	103,4	-19,3	370,6	-195,7
31.07.2004	540,27	391	5,2	26,7	-62,3	3881,5	-321,9
31.08.2004	584,65	406	20,2	406,7	-17,9	321,2	-361,4
30.09.2004	631,65	416	30,2	910,0	29,1	845,5	877,2
31.10.2004	663,67	425	39,2	1534,0	61,1	3733,0	2393,0
30.11.2004	627,98	450	64,2	4117,4	25,4	645,6	1630,3
31.12.2004	614,11	435	49,2	2417,4	11,5	133,1	567,3
31.01.2005	637,21	422	36,2	1308,0	34,6	1199,8	1252,7
28.02.2005	716,42	435	49,2	2417,4	113,8	12961,4	5597,5
31.03.2005	669,07	428	42,2	1778,0	66,5	4422,0	2804,0
30.04.2005	670,36	436	50,2	2516,7	67,8	4595,2	3400,7
31.05.2005	674,44	415	29,2	850,7	71,9	5165,0	2096,1
30.06.2005	706,38	437	51,2	2618,0	103,8	10776,1	5311,5
31.07.2005	778,93	429	43,2	1863,4	176,4	31102,1	7612,8
31.08.2005	882,03	434	48,2	2320,0	279,5	78096,8	13460,6
30.09.2005	1007,76	473	87,2	7598,0	405,2	164177,3	35318,9
31.10.2005	934,99	470	84,2	7084,0	332,4	110501,7	27978,5
30.11.2005	1037,26	495	109,2	11917,4	434,7	188953,6	47453,4
31.12.2005	1125,6	510	124,2	15417,4	523,0	273558,3	64942,6
31.01.2006	1315,96	570	184,2	33917,4	713,4	508922,4	131382,3
28.02.2006	1453,44	555	169,2	28617,4	850,9	723976,3	143938,5
31.03.2006	1434,99	581	195,2	38090,0	832,4	692919,7	162460,2
30.04.2006	1657,28	644	258,2	66650,0	1054,7	1112409	272290,4
31.05.2006	1461,22	660	274,2	75167,4	858,6	737276,3	235412,7
30.06.2006	1494,63	615	229,2	52517,4	892,1	795767,4	204430,0
31.07.2006	1551,09	639	253,2	64093,4	948,5	899686,3	240133,1
31.08.2006	1626,69	624	238,2	56723,4	1024,1	1048818	243910,8

30.09.2006	1549,99	600	214,2	45867,4	947,4	897600,8	202905,3
31.10.2006	1613,57	602	216,2	46728,0	1011,0	1022117	218544,1
30.11.2006	1776,68	646	260,2	67686,7	1174,1	1378530	305463,8
31.12.2006	1921,92	631	245,2	60106,7	1319,3	1740679	323460,1
итого	50616,051	32410		131198,3		1848964	489685,8099
выб.средн	602,57203	385,83					
выб.диспер	212023,75	13080					
СКО	460,46037	114,37					
var(%)	76,415821	29,641					
cov(x,y)	47682,911	смещ.оц.					
cov(x,y)	48257,404	несмещ.оц					
Rxy	0,916373						
tрасч	6,8672095						
tкрит(0,1;9)	1,293						
tкрит(0,05;9)	1,665						

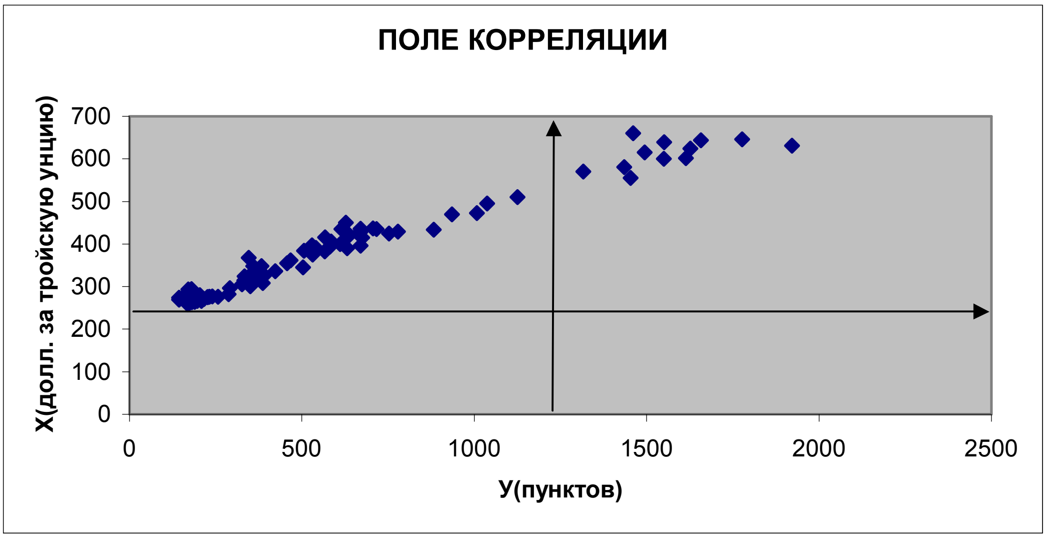


Рис. 3.10. Исследование поля Корреляции

Исследуя поле корреляции (таблица 3.16, рис. 3.10), можно сделать вывод, что между результирующим показателем *У* и фактором *Х* есть четкая зависимость. Для построения уравнения регрессии рассмотрим линейную, степенную, показательную и гиперболическую модели. Построение уравнения регрессии сводится к оценке ее параметров.

Коэффициент вариации *У* свидетельствует о значительном рассеивании фактора *У* по отношению к своему среднему значению. Коэффициент вариации *Х* свидетельствует о рассеивании фактора *Х* по отношению к своему среднему значению.

Линейный коэффициент корреляции показывает тесноту связи между двумя случайными величинами. Получилось, что коэффициент корреляции положительный и равен 0.916, а это означает, что наблюдается сильная связь между показателями.

Сформируем нулевую гипотезу о том, что зависимости между *X* и *У* нет. Для этого вычислим наблюдаемое значение t–распределения Стъюдента по формуле $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ и сравним это значение с критическим значением t–распределения с (n-2) степенями свободы про уровне значимости 0,1 и 0,05. Получим, что линейный коэффициент корреляции статистически значим (отклоняем гипотезу $Ho: r_{xy} = 0$).

Построение линейной зависимости

С помощью МНК находим оценки параметров линейного уравнения регрессии вида $y_i = a + b \cdot x_i + \varepsilon_i$. Для этого следует решить систему линейных уравнений относительно *a* и *b*,

$$\begin{cases} n \cdot a + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x^2{}_i = \sum_{i=1}^{n} x_i \cdot y_i \end{cases}$$

из которой можно определить:

$$\hat{a} = \overline{y} - \hat{b} \cdot \overline{x},$$

$$\hat{b} = \frac{cov(x, y)}{\sigma_x^2} = r_{xy}\frac{\sigma_y}{\sigma_x}.$$

Результаты расчетов для линейной модели представлены в таблице 3.16.

Уравнение регрессии (получено в Excel вручную по указанным выше формулам) будет иметь вид:

$$y=-820,95+3,7x.$$

С увеличением цены на золото в среднем на 1 долл. за унцию, индекс РТС увеличится в среднем на 3*,*7 пункта.

Тесноту связи оценим через индекс корреляции $\rho_{yx} = 0.9164$, т.е. связь сильная и положительная.

Построение степенной зависимости

Построению степенной модели $y_i = a \cdot x_i^b \cdot \varepsilon_i$ предшествует процедура линеаризации переменных, которую осуществим через логарифмирование обеих частей уравнения:

$$\ln y_i = \ln(a \cdot x_i^b \cdot \varepsilon_i);$$

$$\ln y_i = \ln a + b \cdot \ln x_i + \ln \varepsilon_i;$$

Получим $Y = A + b \cdot X + \mathrm{E}$, где $Y = \ln y;\ X = \ln x; A = \ln a; E = \ln \varepsilon$.

Все расчеты осуществим в таблице 3.17.

Параметры уравнения модели находим по следующим формулам:

$$A = \overline{Y} - b\overline{X}\,; \qquad A \approx \text{-}9{,}04;$$

$$b = \frac{\operatorname{cov}(X,Y)}{\sigma_X^2}; \quad b \approx 2{,}57;$$

Для расчета использовали Excel (Сервис – Анализ данных – Регрессия).

Таким образом, линейное уравнение выглядит следующим образом:

$$Y = -9.04 + 2.57 \cdot X\,.$$

Для получения оценки *a* осуществим операцию потенцирования

$$a = e^{-9{,}04} = 0{,}00012.$$

Следовательно, степенная модель будет иметь вид:

$$\widehat{y} = 0{,}00012 \cdot x^{2{,}57}\,.$$

Подставляя в данное уравнение фактические значения *x*, получаем теоретические значения $\widehat{y}$. По ним рассчитаем показатель тесноты связи – индекс корреляции, который находится по формуле:

$$\rho_{yx} = \sqrt{1 - \frac{S^2_{ост}}{S^2_{общ}}} = \sqrt{1 - \frac{\sum\limits_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum\limits_{i=1}^{n} (y_i - \overline{y})^2}} = 0{,}97 \text{ –связь сильная.}$$

Таблица 3.17.

Построение степенной зависимости, расчет коэффициентов и параметров модели в программе MS Excel.

месяц	Y (PTC)	X (Золото)	Ln(Y)	Ln (X)	Yтеор	остатки	S^2ост	(Y-Yт)/Y	mod[(Y-Yт)/Y]
31.01.2000	172,3086	283	5,1493	5,6454	231,48	-59,173	3501,43	-0,34341	0,34341
29.02.2000	170,9301	293	5,1413	5,6802	253,05	-82,119	6743,57	-0,48043	0,48043
31.03.2000	231,877	276	5,4462	5,6204	217,08	14,801	219,071	0,06383	0,06383
30.04.2000	226,8656	275	5,4244	5,6168	215,06	11,8016	139,278	0,05202	0,05202
31.05.2000	190,2074	272	5,2481	5,6058	209,1	-18,889	356,798	-0,09931	0,09931
30.06.2000	171,3954	288	5,144	5,663	242,12	-70,723	5001,81	-0,41263	0,41263
31.07.2000	194,0904	275	5,2683	5,6168	215,06	-20,974	439,891	-0,10806	0,10806
31.08.2000	239,9919	277	5,4806	5,624	219,1	20,8925	436,498	0,08706	0,08706
30.09.2000	199,0773	274	5,2937	5,6131	213,06	-13,986	195,612	-0,07025	0,07025
31.10.2000	188,9985	264	5,2417	5,5759	193,68	-4,6823	21,9243	-0,02477	0,02477
30.11.2000	143,4181	269	4,9658	5,5947	203,23	-59,813	3577,6	-0,41705	0,41705
31.12.2000	143,2926	274	4,9649	5,6131	213,06	-69,771	4867,97	-0,48691	0,48691
31.01.2001	173,5274	264	5,1563	5,5759	193,68	-20,153	406,161	-0,11614	0,11614
28.02.2001	164,762	267	5,1045	5,5872	199,38	-34,615	1198,22	-0,21009	0,21009
31.03.2001	169,4619	258	5,1326	5,553	182,59	-13,127	172,307	-0,07746	0,07746
30.04.2001	180,6806	263	5,1967	5,5722	191,8	-11,124	123,738	-0,06157	0,06157
31.05.2001	208,7959	266	5,3414	5,5835	197,47	11,3286	128,337	0,05426	0,05426
30.06.2001	216,11	271	5,3758	5,6021	207,13	8,97994	80,6393	0,04155	0,04155
31.07.2001	196,12	266	5,2787	5,5835	197,47	-1,3473	1,81528	-0,00687	0,00687
31.08.2001	205,41	273	5,325	5,6095	211,07	-5,6643	32,084	-0,02758	0,02758

30.09.2001	180,25	294	5,1943	5,6836	255,27	-75,021	5628,13	-0,4162	0,41620
31.10.2001	204,04	280	5,3183	5,6348	225,24	-21,199	449,377	-0,10389	0,10389
30.11.2001	226,49	275	5,4227	5,6168	215,06	11,426	130,554	0,05045	0,05045
31.12.2001	256,75	276	5,5481	5,6204	217,08	39,674	1574,03	0,15452	0,15452
31.01.2002	287,53	282	5,6613	5,6419	229,39	58,1411	3380,39	0,20221	0,20221
28.02.2002	290,75	296	5,6725	5,6904	259,75	31,0005	961,032	0,10662	0,10662
31.03.2002	350,75	301	5,8601	5,7071	271,15	79,5952	6335,4	0,22693	0,22693
30.04.2002	386,1	308	5,9561	5,7301	287,63	98,4724	9696,81	0,25504	0,25504
31.05.2002	391,26	326	5,9694	5,7869	332,74	58,5158	3424,1	0,14956	0,14956
30.06.2002	353,79	318	5,8687	5,7621	312,2	41,5929	1729,97	0,11756	0,11756
31.07.2002	326,23	305	5,7876	5,7203	280,5	45,7348	2091,67	0,14019	0,14019
31.08.2002	332,9	314	5,8078	5,7494	302,22	30,6782	941,154	0,09215	0,09215
30.09.2002	334,06	324	5,8113	5,7807	327,53	6,52757	42,6092	0,01954	0,01954
31.10.2002	358,65	317	5,8823	5,7589	309,68	48,9653	2397,6	0,13653	0,13653
30.11.2002	361,15	319	5,8893	5,7652	314,72	46,4282	2155,57	0,12856	0,12856
31.12.2002	359,07	348	5,8835	5,8522	393,43	-34,361	1180,7	-0,0957	0,09570
31.01.2003	345,56	368	5,8452	5,9081	454,07	-108,51	11775,4	-0,31402	0,31402
28.02.2003	383,23	348	5,9486	5,8522	393,43	-10,201	104,066	-0,02662	0,02662
31.03.2003	360,33	335	5,887	5,8141	356,82	3,50787	12,3051	0,00974	0,00974
30.04.2003	422,37	336	6,0459	5,8171	359,56	62,809	3944,97	0,14871	0,14871
31.05.2003	467,1	362	6,1465	5,8916	435,32	31,7765	1009,75	0,06803	0,06803
30.06.2003	503,51	345	6,2216	5,8435	384,79	118,721	14094,7	0,23579	0,23579
31.07.2003	457,02	355	6,1247	5,8721	414,05	42,9659	1846,06	0,09401	0,09401
31.08.2003	530,94	375	6,2746	5,9269	476,56	54,3766	2956,82	0,10242	0,10242
30.09.2003	566,62	383	6,3397	5,948	503,08	63,538	4037,08	0,11214	0,11214
31.10.2003	506,12	384	6,2268	5,9506	506,46	-0,3386	0,11465	-0,00067	0,00067
30.11.2003	529,27	397	6,2715	5,9839	551,62	-22,347	499,38	-0,04222	0,04222
31.12.2003	567,25	416	6,3408	6,0307	621,9	-54,652	2986,84	-0,09635	0,09635
31.01.2004	611,1	400	6,4153	5,9915	562,37	48,7263	2374,26	0,07974	0,07974
29.02.2004	670,14	396	6,5075	5,9814	548,06	122,081	14903,7	0,18217	0,18217
31.03.2004	752,66	425	6,6236	6,0521	657,01	95,655	9149,88	0,12709	0,12709
30.04.2004	631,11	390	6,4475	5,9661	527,01	104,101	10837,1	0,16495	0,16495
31.05.2004	581,07	394	6,3649	5,9764	540,99	40,0835	1606,69	0,06898	0,06898
30.06.2004	583,32	396	6,3687	5,9814	548,06	35,2607	1243,31	0,06045	0,06045
31.07.2004	540,27	391	6,2921	5,9687	530,48	9,78792	95,8033	0,01812	0,01812
31.08.2004	584,65	406	6,371	6,0064	584,27	0,38104	0,14519	0,00065	0,00065
30.09.2004	631,65	416	6,4483	6,0307	621,9	9,74801	95,0237	0,01543	0,01543
31.10.2004	663,67	425	6,4978	6,0521	657,01	6,66498	44,422	0,01004	0,01004
30.11.2004	627,98	450	6,4425	6,1092	760,77	-132,79	17631,9	-0,21145	0,21145
31.12.2004	614,11	435	6,4202	6,0753	697,4	-83,287	6936,67	-0,13562	0,13562
31.01.2005	637,21	422	6,4571	6,045	645,17	-7,9632	63,4133	-0,0125	0,01250
28.02.2005	716,42	435	6,5743	6,0753	697,4	19,0233	361,888	0,02655	0,02655
31.03.2005	669,07	428	6,5059	6,0591	668,97	0,10175	0,01035	0,00015	0,00015
30.04.2005	670,36	436	6,5078	6,0776	701,52	-31,157	970,752	-0,04648	0,04648
31.05.2005	674,44	415	6,5139	6,0283	618,07	56,3659	3177,12	0,08357	0,08357
30.06.2005	706,38	437	6,5602	6,0799	705,65	0,72803	0,53003	0,00103	0,00103
31.07.2005	778,93	429	6,6579	6,0615	672,99	105,945	11224,3	0,13601	0,13601
31.08.2005	882,03	434	6,7822	6,073	693,29	188,739	35622,3	0,21398	0,21398
30.09.2005	1007,76	473	6,9155	6,1591	864,54	143,215	20510,6	0,14211	0,14211
31.10.2005	934,99	470	6,8405	6,1527	850,55	84,4423	7130,5	0,09031	0,09031

30.11.2005	1037,26	495	6,9443	6,2046	971,49	65,7695	4325,63	0,06341	0,06341
31.12.2005	1125,6	510	7,0261	6,2344	1048,8	76,7856	5896,03	0,06822	0,06822
31.01.2006	1315,96	570	7,1823	6,3456	1395,1	-79,182	6269,82	-0,06017	0,06017
28.02.2006	1453,44	555	7,2817	6,319	1302,9	150,554	22666,4	0,10358	0,10358
31.03.2006	1434,99	581	7,2689	6,3648	1465,3	-30,269	916,213	-0,02109	0,02109
30.04.2006	1657,28	644	7,4129	6,4677	1908,1	-250,87	62934,3	-0,15137	0,15137
31.05.2006	1461,22	660	7,287	6,4922	2032,1	-570,92	325951	-0,39072	0,39072
30.06.2006	1494,63	615	7,3096	6,4216	1695,4	-200,79	40315,6	-0,13434	0,13434
31.07.2006	1551,09	639	7,3467	6,4599	1870,4	-319,28	101941	-0,20584	0,20584
31.08.2006	1626,69	624	7,3943	6,4362	1759,8	-133,11	17717,8	-0,08183	0,08183
30.09.2006	1549,99	600	7,346	6,3969	1591,4	-41,361	1710,7	-0,02668	0,02668
31.10.2006	1613,57	602	7,3862	6,4003	1605	8,57578	73,544	0,00531	0,00531
30.11.2006	1776,68	646	7,4825	6,4708	1923,4	-146,71	21522,7	-0,08257	0,08257
31.12.2006	1921,92	631	7,5611	6,4473	1810,9	111,032	12328,2	0,05777	0,05777
итого	50616,0507	32410	515,63	496,94			881580		10,16793024
выб.средн	602,572032	385,8333	6,1385	5,916					
выб.диспер	212023,751	13079,73		0,0766					
S^2общ	17597971,3								
оценка a	0,00011878								
оценка b	2,56537956								
рух	0,97463041								
Эср	2,56537956								
коэфф детерм.	0,94990444								
оц.дисп.ошиб	10750,9783								
Acred	12,1046789								
Fрасч	1554,87163								
Fтаб(0,05;1;9)	3,91	Fт(0,1;1;9)	6,72						
дов интерв. а	5,7524E-05	0,000245							
Ттаб(0,05;9)	2,2622								
					ПРОГНОЗИРОВАНИЕ				
	по линеаризованной моделе							по исходной моделе	
Хпр=LN(хпр)	6,82252413								
Упр	8,46414977						упр=	4741,69	
упр=e^(Yпр)	4741,69415								
дисп.прогноза	35032,5306						дисп.пр.	17743,6	
ССО прогноза	187,169791						ССО пр.	133,205	
пред.ош.прогн	423,415501						пред.ош.пр	301,336	
доверительный интервал для прогнозного значения									
				по линеаризованной моделе					
	4318,27865		5165,1						
				по исходной моделе					
					6E-128		4E+134		

	Коэфф.	*Станд. ош.*	*t-стат*	*P-Знач.*	*Нижн.95%*	*Верхн.95%*	*Нижн.95%*	*Верхн.95%*
Y-пересечение	-9,0382142	0,36449	-24,8	7E-40	-9,7633	-8,3131	-9,7633	-8,31313
Переменная X 1	2,56537956	0,061545	41,683	6E-57	2,4429	2,68781	2,44295	2,68781

Построение показательной зависимости

Построению показательной модели $y_i = a \cdot b^{x_i} \cdot \varepsilon_i$ предшествует процедура линеаризации переменных, которую осуществим через логарифмирование обеих частей уравнения:

$$\ln y_i = \ln(a \cdot b^{x_i} \cdot \varepsilon);$$

$$\ln y_i = \ln a + x_i \ln b + \ln \varepsilon_i;$$

Получим $Y = A + Bx + \nu$,

где $Y = \ln y;\ A = \ln a; B = \ln b;\ \nu = \ln \varepsilon$.

Все расчеты представлены в таблице 3.18.

Параметры уравнения модели находим по следующим формулам:

$$A = \overline{Y} - B\overline{x}; \qquad A \approx 3{,}79;$$

$$B = \frac{\operatorname{cov}(x, Y)}{\sigma_x^2}; \quad B \approx 0{,}006;$$

Для расчета использовали Excel (Сервис – Анализ данных – Регрессия).

Таким образом, линейное уравнение выглядит следующим образом:

$$Y = 3{,}79 + 0.006 \cdot x.$$

Для получения оценок a и b осуществим операцию потенцирования

$$a = e^{3,79} = 44{,}26;$$

$$b = e^{0,006} = 1{,}006.$$

Следовательно, показательная модель будет иметь вид:

$$\hat{y} = 44{,}26 \cdot 1{,}006^{x}.$$

По получившемуся уравнению видно, что $\hat{y}$ почти константа.

Тесноту связи оценим через индекс корреляции $\rho_{yx} = 0.91$ - связь сильная и положительная.

Таблица 3.18.

Построение показательной зависимости, расчет коэффициентов и параметров модели в программе MS Excel.

месяц	Y	X	Ln (У)	Yтеор	остатки	S^2ост	(Y-Yт)/Y	mod[(Y-Yт)/Y]
31.01.2000	172,3086	283	5,149287	247,9831	-75,6745	5726,629	-0,43918	0,439180023
29.02.2000	170,9301	293	5,141255	263,5248	-92,5947	8573,78	-0,541109	0,541710949
31.03.2000	231,877	276	5,446207	237,6525	-5,77555	33,35696	-0,024908	0,024907813
30.04.2000	226,8656	275	5,424358	236,2123	-9,34671	87,36106	-0,041199	0,041199344
31.05.2000	190,2074	272	5,248115	231,9438	-41,7364	1741,924	-0,219425	0,219425567
30.06.2000	171,3954	288	5,143973	255,6359	-84,2405	7096,456	-0,491497	0,491497836
31.07.2000	194,0904	275	5,268324	236,2123	-42,1219	1774,256	-0,217022	0,217022139
31.08.2000	239,9919	277	5,480605	239,1016	0,890335	0,792696	0,0037099	0,003709853
30.09.2000	199,0773	274	5,293693	234,7808	-35,7035	1274,74	-0,179344	0,179344943
31.10.2000	188,9985	264	5,241739	220,9343	-31,9358	1019,896	-0,168973	0,168973877
30.11.2000	143,4181	269	4,965764	227,7524	-84,3343	7112,267	-0,588030	0,588030768
31.12.2000	143,2926	274	4,964889	234,7808	-91,4882	8370,092	-0,638471	0,638471261
31.01.2001	173,5274	264	5,156336	220,9343	-47,4069	2247,415	-0,273195	0,273195526
28.02.2001	164,762	267	5,104502	225,0002	-60,2382	3628,646	-0,365607	0,365607636
31.03.2001	169,4619	258	5,132628	213,0215	-43,5596	1897,442	-0,257046	0,257046767

30.04.2001	180,6806	263	5,196731	219,5954	-38,9148	1514,361	-0,215378	0,215378912
31.05.2001	208,7959	266	5,341357	223,6367	-14,8408	220,2489	-0,071078	0,071077959
30.06.2001	216,11	271	5,375788	230,5381	-14,4281	208,1709	-0,066762	0,06676289
31.07.2001	196,12	266	5,278727	223,6367	-27,5167	757,168	-0,140305	0,140305357
31.08.2001	205,41	273	5,325008	233,358	-27,948	781,0893	-0,136059	0,136059469
30.09.2001	180,25	294	5,194345	265,1316	-84,8816	7204,881	-0,470910	0,470910245
31.10.2001	204,04	280	5,318316	243,5018	-39,4618	1557,237	-0,193402	0,193402471
30.11.2001	226,49	275	5,422701	236,2123	-9,72231	94,52339	-0,042926	0,042926018
31.12.2001	256,75	276	5,548103	237,6525	19,09745	364,7126	0,0743815	0,074381504
31.01.2002	287,53	282	5,661327	246,4803	41,04975	1685,082	0,1427668	0,142766827
28.02.2002	290,75	296	5,672464	268,3746	22,37545	500,6607	0,0769577	0,07695769
31.03.2002	350,75	301	5,860074	276,6566	74,09339	5489,831	0,2112427	0,211242747
30.04.2002	386,1	308	5,956096	288,6826	97,41738	9490,145	0,2523113	0,252311255
31.05.2002	391,26	326	5,969372	322,0621	69,1979	4788,35	0,1768591	0,176859133
30.06.2002	353,79	318	5,868704	306,7751	47,01492	2210,403	0,1328894	0,132889353
31.07.2002	326,23	305	5,787603	283,4659	42,7641	1828,769	0,1310857	0,131085748
31.08.2002	332,9	314	5,807842	299,4059	33,49414	1121,857	0,1006132	0,100613206
30.09.2002	334,06	324	5,811321	318,1704	15,88963	252,4804	0,0475652	0,047565207
31.10.2002	358,65	317	5,882347	304,9159	53,73406	2887,349	0,1498231	0,149823112
30.11.2002	361,15	319	5,889293	308,6455	52,50445	2756,718	0,1453813	0,145381291
31.12.2002	359,07	348	5,883517	368,1445	-9,07453	82,34714	-0,025272	0,025272322
31.01.2003	345,56	368	5,845166	415,7356	-70,1756	4924,614	-0,203077	0,203077888
28.02.2003	383,23	348	5,948635	368,1445	15,08547	227,5713	0,039364	0,039364004
31.03.2003	360,33	335	5,88702	340,1724	20,15758	406,328	0,055942	0,055941999
30.04.2003	422,37	336	6,045882	342,2465	80,12348	6419,772	0,1896997	0,18969974
31.05.2003	467,1	362	6,146543	400,846	66,254	4389,593	0,1418411	0,141841149
30.06.2003	503,51	345	6,221604	361,4919	142,0181	20169,15	0,2820562	0,282056238
31.07.2003	457,02	355	6,124727	384,1474	72,87255	5310,409	0,1594516	0,159451556
31.08.2003	530,94	375	6,274649	433,8073	97,13275	9434,77	0,1829449	0,182944865
30.09.2003	566,62	383	6,339689	455,4245	111,1955	12364,45	0,1962436	0,196243586
31.10.2003	506,12	384	6,226774	458,2013	47,91872	2296,204	0,0946786	0,094678582
30.11.2003	529,27	397	6,271499	495,8788	33,39119	1114,972	0,0630891	0,063089142
31.12.2003	567,25	416	6,3408	556,5888	10,6612	113,6612	0,0187945	0,018794536
31.01.2004	611,1	400	6,415261	505,0047	106,0953	11256,22	0,1736137	0,173613727
29.02.2004	670,14	396	6,507487	492,8737	177,2663	31423,36	0,2645214	0,264521357
31.03.2004	752,66	425	6,623614	587,8871	164,7729	27150,1	0,2189207	0,218920731
30.04.2004	631,11	390	6,44748	475,2214	155,8886	24301,27	0,2470071	0,247007086
31.05.2004	581,07	394	6,364871	486,9179	94,15212	8864,622	0,1620323	0,162032321
30.06.2004	583,32	396	6,368736	492,8737	90,44634	8180,541	0,1550544	0,155054416
31.07.2004	540,27	391	6,292069	478,1189	62,15112	3862,762	0,1150371	0,115037148
31.08.2004	584,65	406	6,371013	523,7633	60,88674	3707,195	0,1041422	0,104142197
30.09.2004	631,65	416	6,448335	556,5888	75,0612	5634,184	0,1188335	0,118833532
31.10.2004	663,67	425	6,497785	587,8871	75,78288	5743,044	0,1141876	0,114187589
30.11.2004	627,98	450	6,442508	684,3724	-56,3924	3180,108	-0,089799	0,08979975
31.12.2004	614,11	435	6,420174	624,7315	-10,6215	112,8154	-0,017295	0,017295692
31.01.2005	637,21	422	6,457099	577,2635	59,94647	3593,58	0,0940765	0,094076478
28.02.2005	716,42	435	6,574267	624,7315	91,68854	8406,789	0,1279816	0,12798155
31.03.2005	669,07	428	6,505889	598,7062	70,36377	4951,06	0,1051665	0,105166533
30.04.2005	670,36	436	6,507815	628,5406	41,81943	1748,864	0,0623835	0,062383535
31.05.2005	674,44	415	6,513883	553,2157	121,2243	14695,32	0,1797406	0,17974063
30.06.2005	706,38	437	6,560153	632,3729	74,00709	5477,049	0,1047695	0,104769508
31.07.2005	778,93	429	6,657921	602,3567	176,5733	31178,14	0,226687	0,226687042
31.08.2005	882,03	434	6,782226	620,9454	261,0846	68165,15	0,2960042	0,296004188
30.09.2005	1007,76	473	6,915485	787,0661	220,6939	48705,8	0,2189945	0,218994516

31.10.2005	934,99	470	6,840536	772,8432	162,1468	26291,6	0,1734209	0,173420928
30.11.2005	1037,26	495	6,944338	899,6839	137,5761	18927,19	0,1326342	0,132634168
31.12.2005	1125,6	510	7,026072	985,5736	140,0264	19607,38	0,1244015	0,124401523
31.01.2006	1315,96	570	7,182322	1419,337	-103,377	10686,89	-0,078556	0,078556644
28.02.2006	1453,44	555	7,281688	1295,646	157,7936	24898,81	0,1085656	0,108565583
31.03.2006	1434,99	581	7,268913	1517,487	-82,4971	6805,777	-0,057489	0,057489692
31.05.2006	1461,22	660	7,287027	2452,904	-991,684	983436,5	-0,678668	0,678668293
30.06.2006	1494,63	615	7,309634	1865,877	-371,247	137824,5	-0,248387	0,248387398
31.07.2006	1551,09	639	7,346713	2158,946	-607,856	369488,4	-0,391889	0,39188931
31.08.2006	1626,69	624	7,394303	1970,8	-344,11	118411,7	-0,21154	0,211540005
30.09.2006	1549,99	600	7,346004	1703,272	-153,282	23495,28	-0,098892	0,098892055
31.10.2006	1613,57	602	7,386204	1724,105	-110,535	12218,08	-0,068503	0,068503637
30.11.2006	1776,68	646	7,482502	2252,793	-476,113	226683,6	-0,267979	0,267979065
31.12.2006	1921,92	631	7,56108	2056,469	-134,549	18103,42	-0,070007	0,070007567
итого	50616,05	32410	515,63			2803725		15,33057141
СРЕДНЕЕ	602,572	385,83333	6,138453		оц.дисп.ош	34191,76		
ДИСП(несм)	212023,8	13079,731						
S^2общ	17597971							
оценка a	44,39339							
оценка b	1,006097							
ryx	0,916886							
Эср	2,345362							
коэфф детерм.	0,840679							
Acred	18,25068							
Fрасч	432,6845							
Fтаб(0,05;1;9)	3,91	Fтаб(0,1;1;9)	6,72					
дов интерв. a	37,73236	52,230306						
дов интерв. b	1,005691	1,0065039						
Tтаб(0,05;9)	2,2622							
		ПРОГНОЗИРОВАНИЕ (по линеаризованной моделе)						
Xпр=LN(xпр)	6,822524							
Уnp	3,834563							
упр=e^(Yпр)	46,27318							
дисп.прогноза	43528,35							
ССО прогноза	208,6345							
пред.ош.прогн	471,973							

	Коэфф.	*Станд. ош.*	*t-стат*	*P-Знач.*	*Нижн.95%*	*Верхн.95%*	*Нижн.95%*	*Верхн.95%*
Y-пересечение	3,793091	0,0817226	46,41421	1,22E-60	3,630518	3,955663	3,6305181	3,955662911
Переменная X 1	0,006079	0,0002032	29,9189	6,92E-46	0,005675	0,006483	0,0056745	0,006482868

Построение гиперболической зависимости

Уравнение равносторонней гиперболы $y_i = a + \frac{b}{x_i} + \varepsilon_i$ линеаризуется при замене: z=1/x. Тогда $y_i = a + b \cdot z_i + \varepsilon_i$.

Параметры уравнения модели находим по следующим формулам:

$$a = \bar{y} - b \cdot \bar{z};$$

$$b = \frac{\operatorname{cov}(z, y)}{\sigma_z^2}$$

Все расчеты осуществим в таблице 3.19. Получим:

a=2251,717; b=-590153.

Для расчета использовали Excel (Сервис – Анализ данных – Регрессия).

Следовательно, гиперболическая модель будет иметь вид:

$$\hat{y} = 2251{,}717 - 590153 / x .$$

Тесноту связи оценим через индекс корреляции $\rho_{yx} = 0.9189$ - связь сильная и положительная.

Анализ индексов корреляции показывает, что степенная модель несколько лучше описывает взаимосвязь между величинами X и $У$, чем показательная, гиперболическая и линейная модели.

Таблица 3.19

Построение гиперболической зависимости, расчет коэффициентов и параметров модели в программе MS Excel.

месяц	Y (РТС)	X (Золото)	Z=1/X	Yтеор	остатки	S^2ост	(Y-Yт)/Y	mod[(Y-Yт)/Y]
31.01.2000	172,31	283	0,003534	166,369	5,939295	35,27522	0,034469	0,0344689
29.02.2000	170,93	293	0,003413	237,542	-66,6115	4437,088	-0,3897	0,3897
31.03.2000	231,88	276	0,003623	113,48	118,3969	14017,83	0,510602	0,5106023
30.04.2000	226,87	275	0,003636	105,705	121,1609	14679,97	0,534065	0,5340649
31.05.2000	190,21	272	0,003676	82,0354	108,172	11701,18	0,568705	0,5687055
30.06.2000	171,4	288	0,003472	202,573	-31,1779	972,0584	-0,18191	0,181906
31.07.2000	194,09	275	0,003636	105,705	88,38574	7812,04	0,455384	0,4553844
31.08.2000	239,99	277	0,00361	121,199	118,7926	14111,68	0,494986	0,4949858
30.09.2000	199,08	274	0,00365	97,8725	101,2048	10242,41	0,508369	0,5083694
31.10.2000	189	264	0,003788	16,2875	172,711	29829,09	0,913822	0,9138221
30.11.2000	143,42	269	0,003717	57,8382	85,57988	7323,916	0,596716	0,596716
31.12.2000	143,29	274	0,00365	97,8725	45,4201	2062,986	0,316975	0,3169745
31.01.2001	173,53	264	0,003788	16,2875	157,2399	24724,39	0,906139	0,9061388
28.02.2001	164,76	267	0,003745	41,4047	123,3573	15217,03	0,7487	0,7487001
31.03.2001	169,46	258	0,003876	-35,6992	205,1611	42091,09	1,210662	1,2106623
30.04.2001	180,68	263	0,003802	7,78776	172,8928	29891,93	0,956898	0,9568976
31.05.2001	208,8	266	0,003759	33,0952	175,7007	30870,73	0,841495	0,8414948
30.06.2001	216,11	271	0,00369	74,0292	142,0808	20186,95	0,657447	0,6574467
31.07.2001	196,12	266	0,003759	33,0952	163,0248	26577,08	0,83125	0,8312501
31.08.2001	205,41	273	0,003663	89,983	115,427	13323,4	0,561935	0,5619349
30.09.2001	180,25	294	0,003401	244,393	-64,1425	4114,261	-0,35585	0,355853
31.10.2001	204,04	280	0,003571	144,026	60,0137	3601,644	0,294127	0,2941271
30.11.2001	226,49	275	0,003636	105,705	120,7853	14589,1	0,533292	0,5332922
31.12.2001	256,75	276	0,003623	113,48	143,2699	20526,28	0,558013	0,5580134
31.01.2002	287,53	282	0,003546	158,974	128,5555	16526,53	0,447103	0,4471031
28.02.2002	290,75	296	0,003378	257,956	32,7945	1075,479	0,112793	0,1127928

31.03.2002	350,75	301	0,003322	291,074	59,67554	3561,17	0,170137	0,170137
30.04.2002	386,1	308	0,003247	335,635	50,46549	2546,766	0,130706	0,1307058
31.05.2002	391,26	326	0,003067	441,43	-50,1705	2517,074	-0,12823	0,1282279
30.06.2002	353,79	318	0,003145	395,889	-42,0987	1772,297	-0,11899	0,1189934
31.07.2002	326,23	305	0,003279	316,788	9,442202	89,15517	0,028943	0,0289434
31.08.2002	332,9	314	0,003185	372,248	-39,3475	1548,229	-0,1182	0,1181963
30.09.2002	334,06	324	0,003086	430,256	-96,1958	9253,641	-0,28796	0,2879598
31.10.2002	358,65	317	0,003155	390,034	-31,3843	984,9752	-0,08751	0,0875068
30.11.2002	361,15	319	0,003135	401,706	-40,5563	1644,814	-0,1123	0,1122977
31.12.2002	359,07	348	0,002874	555,874	-196,804	38731,75	-0,54809	0,5480932
31.01.2003	345,56	368	0,002717	648,039	-302,479	91493,66	-0,87533	0,8753305
28.02.2003	383,23	348	0,002874	555,874	-172,644	29805,89	-0,4505	0,4504966
31.03.2003	360,33	335	0,002985	490,065	-129,735	16831,17	-0,36004	0,360045
30.04.2003	422,37	336	0,002976	495,308	-72,938	5319,954	-0,17269	0,1726875
31.05.2003	467,1	362	0,002762	621,459	-154,359	23826,67	-0,33046	0,3304622
30.06.2003	503,51	345	0,002899	541,127	-37,6174	1415,066	-0,07471	0,0747103
31.07.2003	457,02	355	0,002817	589,313	-132,293	17501,43	-0,28947	0,2894687
31.08.2003	530,94	375	0,002667	677,975	-147,035	21619,15	-0,27693	0,2769324
30.09.2003	566,62	383	0,002611	710,846	-144,226	20801,26	-0,25454	0,2545381
31.10.2003	506,12	384	0,002604	714,859	-208,739	43572,01	-0,41243	0,41243
30.11.2003	529,27	397	0,002519	765,184	-235,914	55655,6	-0,44574	0,4457354
31.12.2003	567,25	416	0,002404	833,079	-265,829	70665	-0,46863	0,4686274
31.01.2004	611,1	400	0,0025	776,333	-165,233	27302,07	-0,27039	0,2703868
29.02.2004	670,14	396	0,002525	761,431	-91,2905	8333,96	-0,13623	0,1362261
31.03.2004	752,66	425	0,002353	863,121	-110,461	12201,55	-0,14676	0,1467603
30.04.2004	631,11	390	0,002564	738,503	-107,393	11533,27	-0,17017	0,1701653
31.05.2004	581,07	394	0,002538	753,866	-172,796	29858,33	-0,29737	0,2973749
30.06.2004	583,32	396	0,002525	761,431	-178,111	31723,36	-0,30534	0,3053393
31.07.2004	540,27	391	0,002558	742,373	-202,103	40845,69	-0,37408	0,3740781
31.08.2004	584,65	406	0,002463	798,137	-213,487	45576,73	-0,36515	0,3651536
30.09.2004	631,65	416	0,002404	833,079	-201,429	40573,6	-0,31889	0,3188932
31.10.2004	663,67	425	0,002353	863,121	-199,451	39780,56	-0,30053	0,3005268
30.11.2004	627,98	450	0,002222	940,265	-312,285	97521,83	-0,49728	0,4972847
31.12.2004	614,11	435	0,002299	895,042	-280,932	78923	-0,45746	0,4574626
31.01.2005	637,21	422	0,00237	853,249	-216,039	46672,89	-0,33904	0,3390391
28.02.2005	716,42	435	0,002299	895,042	-178,622	31905,96	-0,24933	0,2493263
31.03.2005	669,07	428	0,002336	872,854	-203,784	41527,83	-0,30458	0,3045777
30.04.2005	670,36	436	0,002294	898,154	-227,794	51890,12	-0,33981	0,3398085
31.05.2005	674,44	415	0,00241	829,66	-155,22	24093,4	-0,23015	0,2301472
30.06.2005	706,38	437	0,002288	901,251	-194,871	37974,87	-0,27587	0,2758734
31.07.2005	778,93	429	0,002331	876,068	-97,1379	9435,775	-0,12471	0,1247069
31.08.2005	882,03	434	0,002304	891,916	-9,8864	97,741	-0,01121	0,0112087
30.09.2005	1007,8	473	0,002114	1004,04	3,724763	13,87386	0,003696	0,0036961
31.10.2005	934,99	470	0,002128	996,071	-61,0813	3730,927	-0,06533	0,0653283
30.11.2005	1037,3	495	0,00202	1059,49	-22,2277	494,0725	-0,02143	0,0214293
31.12.2005	1125,6	510	0,001961	1094,55	31,0467	963,8978	0,027582	0,0275824
31.01.2006	1316	570	0,001754	1216,36	99,60004	9920,168	0,075686	0,0756862
28.02.2006	1453,4	555	0,001802	1188,38	265,0627	70258,21	0,182369	0,1823692
31.03.2006	1435	581	0,001721	1235,96	199,0278	39612,05	0,138696	0,1386963
30.04.2006	1657,3	644	0,001553	1335,33	321,9505	103652,1	0,194264	0,1942644
31.05.2006	1461,2	660	0,001515	1357,54	103,675	10748,51	0,070951	0,070951
30.06.2006	1494,6	615	0,001626	1292,12	202,5122	41011,21	0,135493	0,1354932
31.07.2006	1551,1	639	0,001565	1328,16	222,931	49698,22	0,143725	0,1437254
31.08.2006	1626,7	624	0,001603	1305,96	320,7319	102868,9	0,197168	0,1971684

30.09.2006	1550	600	0,001667	1268,13	281,8622	79446,3	0,181848	0,1818478
31.10.2006	1613,6	602	0,001661	1271,4	342,1745	117083,4	0,212061	0,2120605
30.11.2006	1776,7	646	0,001548	1338,17	438,5134	192294	0,246816	0,2468162
31.12.2006	1921,9	631	0,001585	1316,45	605,4701	366594	0,315034	0,315034
итого	50616	32410				2737861		28,39042
СРЕДНЕЕ	602,57	385,83333			оц.дисп.ош	33388,54		
ДИСП(несм)	212024	13079,731	5,14E-07					
S^2общ	2E+07							
рух	0,9189							
Эср	2,118							
коэфф детерм.	0,8444							
Acred	33,798							
Fрасч	48,849							
Fтаб(0,05;1;9)	3,91	Fтаб(0,1;1;9)	6,72					
Tтаб(0,05;9)	2,2622							
		ПРОГНОЗИРОВАНИЕ (по линеаризованной моделе)						
Xпр=1/хпр	0,0011							
Упр	1609,1							
дисп.прогноза	108798							
ССО прогноза	329,85							
пред.ош.прогн	746,18							
доверительный интервал для прогнозного значения (по линеаризованной моделе)								
		862,88174		2355,23				

	Коэфф.	*Станд. ош.*	*t-стат*	*P-Знач.*	*Нижн.95%*	*Верхн.95%*	*Нижн.95%*	*Верхн.95%*
Y-пересечение	2251,7	80,673476	27,91149	1,2E-43	2091,231	2412,202	2091,231	2412,2019
Переменная X 1	-6E+05	27973,869	-21,0966	7,1E-35	-645802	-534504	-645802	-534504,3

Анализ полученных зависимостей

Экономическая интерпретация построенных уравнений.

Средний коэффициент эластичности $\overline{Э}$ показывает, насколько процентов в среднем изменится результирующий показатель от своей средней величины при изменении причинного фактора на 1% от своего среднего значения:

$$\overline{Э} = f^{/}(\overline{x}) \cdot \frac{\overline{x}}{\overline{y}}.$$

Для линейной зависимости:

$$y=-820{,}95+3{,}7x.$$

$$\overline{Э} = b\frac{\overline{x}}{a+b\overline{x}} = 2{,}36.$$

$\overline{Э}$ показывает, что при увеличении (уменьшении) цены на золото на 1% индекс РТС увеличивается (уменьшается) на 2,36%.

Коэффициент регрессии $\widehat{b} = 0{,}0085$ показывает, что при изменении фактора X на 1 ед. среднее изменение результирующего фактора составит 0,0085 ед. Так как $\widehat{b} > 0$, то связь между факторами прямая.

Имеет смысл анализировать лишь знак при $\widehat{a} = -820 < 0$, следовательно, относительное изменение результирующего фактора происходит быстрее относительного изменения причинного фактора.

Для степенной зависимости: $\widehat{y} = 0{,}00012 \cdot x^{2{,}57}$.

$$\overline{Э} = f^{/}(\overline{x}) \cdot \frac{\overline{x}}{\overline{y}} = a \cdot b \cdot \overline{x}^{b-1} \cdot \frac{\overline{x}}{a \cdot \overline{x}^{b}} = b = 2{,}57 .$$

$\overline{Э}$ показывает, что при увеличении (уменьшении) цены на золото на 1% индекс РТС увеличивается (уменьшается) на 2,57%.

Для показательной зависимости: $\widehat{y} = 44{,}26 \cdot 1{,}006^{x}$.

$$\overline{Э} = f^{/}(\overline{x}) \cdot \frac{\overline{x}}{\overline{y}} = ab^{x} \ln b \frac{\overline{x}}{ab^{x}} = \overline{x} \ln b = 2{,}35 .$$

$\overline{Э}$ показывает, что при увеличении (уменьшении) цены на золото на 1% индекс РТС увеличивается (уменьшается) на 2,35%.

Для равносторонней гиперболы : $\widehat{y} = 2251{,}717 - 590153 / x$

$$\overline{Э} = -\frac{b}{\overline{x}^{2}} \cdot \frac{\overline{x}}{a + b / \overline{x}} = -\frac{b}{a\overline{x} + b} = 2{,}11.$$

$\overline{Э}$ показывает, что при увеличении (уменьшении) цены на золото на 1% индекс РТС увеличивается (уменьшается) на 2,11%.

Оценка тесноты связи

Коэффициент детерминации R^2 дает оценку качества построенной модели, т.к. характеризует долю дисперсии, объясненной регрессией, в общей дисперсии результативного признака. Чем ближе R^2 к единице, тем лучше качество подгонки, т.е. $\widehat{y}$ более точно аппроксимирует исходные значения y.

$$R^2 = \frac{\sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = r_{xy}^2 .$$

Для линейной зависимости: $R^2 \approx 0{,}84$.

Следовательно, линейным уравнением регрессии объясняется 84% дисперсии результативного признака, а на долю прочих факторов приходится $(1 - R^2) * 100\% = 29{,}6\%$ дисперсии результативного признака. Поэтому качество модели следует признать удовлетворительным, т.е. данная модель пригодна для прогнозирования.

Для степенной модели: $R^2 = \rho_{xy}^2 \approx 0{,}95$.

Зависимость между показателями выше, чем в остальных моделях. Видим, что 95% вариации результирующего показателя объясняется вариацией фактора (цена золота), а на долю прочих факторов приходится 9,75% дисперсии результативного признака.

Для показательной модели: $R^2 \approx 0{,}84$.

Вариация y на 84% объясняется вариацией x.

Для равносторонней гиперболы: $R^2 \approx 0{,}84$.

Вариация y на 84% объясняется вариацией x.

Оценка качества уравнений

Средняя ошибка аппроксимации характеризует среднее отклонение расчетных значений от фактических:

$$\overline{A} = \frac{\sum_{i=1}^{n} \left| \frac{y_i - \widehat{y}_i}{y_i} \right|}{n} \cdot 100\%.$$

Допустимый предел значений $\overline{A}$ – не более 10– 14%.

Для линейной зависимости: $\overline{A}$ = 13,6% .

Для степенной зависимости: $\overline{A}$ = 12,1%.

Для показательной зависимости: $\overline{A}$ = 18,25% .

Для гиперболической зависимости: $\overline{A}$ = 33,8%.

Получилось, что для показательной и гиперболической моделей средние ошибки аппроксимации выше допустимых пределов.

Оценка статистической надежности уравнения регрессии

F–тест оценивает качество уравнения регрессии. При этом проверяется гипотеза H_0 (b=0 и R^2=0), т.е. гипотеза о статистической незначимости уравнения регрессии и показателя тесноты связи.

Для этого проводится сравнение фактического $F_{факт}$ и критического (находим по соответствующей таблице) $F_{табл.}$ значений F–критерия Фишера. $F_{факт}$ рассчитывается по формуле:

$$F_{факт} = \frac{\sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 / m}{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 / (n - m - 1)} = \frac{R^2}{1 - R^2}(n - 2),$$

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

$F_{табл.}$ – это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α (вероятность, что будет отвергнута правильная нулевая гипотеза).

$$F_{табл.} = F(\alpha, 1, n - 2).$$

Найдем $F_{табл.}$ для α =0,1 и α =0,05.

$F_{табл.}(0{,}1;1;82) = 6{,}72$ и $F_{табл.}(0{,}05;1;82) = 3{,}91$.

Если $F_{табл.} < F_{факт}$, то гипотеза H_0 (о случайной природе оцениваемых характеристик) отклоняется и признается их статистическая значимость, в противном случае H_0 – принимается.

Для линейной зависимости: $F_{факт}$ = 429.

Для степенной зависимости: $F_{факт}$ = 1554

Для показательной зависимости: $F_{факт}$ = 432

Для равносторонней гиперболы: $F_{факт} = 48{,}9$

Таким образом, получили, что во всех моделях $F_{табл.} < F_{факт}$, следовательно, гипотеза H_0 откланяется – все модели значимы.

Проверка значимости коэффициентов модели

Для оценки статистической значимости коэффициентов регрессии рассчитываются *t*–критерий Стьюдента и доверительные интервалы каждого из показателей. Выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля. Оценка значимости коэффициентов регрессии с помощью *t*–критерия Стьюдента проводится путем сопоставления их значений с табличным значением $t_{табл.}$ *t*–статистики Стьюдента:

$$t_{факт}(b) = \frac{b}{\sigma_b}; t_{факт}(a) = \frac{a}{\sigma_a}.$$

Случайные ошибки параметров регрессии:

$$\sigma_b = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2/(n-2)}{\sum_{i=1}^{n}(x-\bar{x})^2}} = \sqrt{\frac{S_{ост}^2/(n-2)}{\sum_{i=1}^{n}(x-\bar{x})^2}} = \frac{S_{ост}}{\sigma_x}\cdot\frac{1}{\sqrt{(n-1)(n-2)}};$$

$$\sigma_a = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{(n-2)}\cdot\frac{\sum_{i=1}^{n}x^2}{n\sum_{i=1}^{n}(x-\bar{x})^2}} = \frac{S_{ост}}{\sigma_x}\cdot\sqrt{\frac{\sum_{i=1}^{n}x^2}{n(n-1)(n-2)}}.$$

Сравнивая фактическое и табличное значение *t*–статистики, принимаем или отвергаем гипотезу H_0.

Если $t_{табл} < |\, t_{факт} \,|$, то H_0 отклоняется, т.е. *a* и *b* не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора *x*.

Если $t_{табл} > |\, t_{факт} \,|$, то H_0 принимается и признается случайная природа формирования *a* и *b*.

Для расчета доверительного интервала определяем предельную ошибку Δ для каждого показателя:

$$\Delta_a = t_{табл}\sigma_a, \Delta_b = t_{табл}\sigma_b;$$

$$t_{табл} = t(n-2;0{,}05).$$

Формулы для расчета доверительных интервалов имеют следующий вид:

$$\gamma_a = a \pm \Delta_a, \gamma_b = b \pm \Delta_b,$$

$$t_{табл.}(82{,}0.05)=1{,}669.$$

Для линейной зависимости:

$\sigma_a = 37{,}08; \quad \sigma_b = 0{,}092.$

$t_{факт}(a) = -22{,}14; \quad t_{факт}(b) = 40{,}02.$

Так как $\text{t}_{табл} > |\, t_{факт}(a) \,|$, то гипотеза H_0 принимается, следовательно, параметр *a* статистически незначим.

Так как $t_{табл} < | t_{факт}(b) |$, то параметр b статистически значим.

Доверительные интервалы:

$a \in (-904{,}8; -737)$ и $b \in (3{,}48; 3{,}9)$.

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью 0,95 параметры a и b находясь в указанных границах, не принимают нулевых значений, т. е. являются статистически значимыми,.

Для степенной зависимости:

$\sigma_A = 0{,}365; \quad \sigma_b = 0{,}06$ (случайная или стандартная ошибка).

$t_{факт}(A) = -24{,}8; \quad t_{факт}(b) = 41{,}7$.

Так как $t_{табл} < | t_{факт}(A) |$, то гипотеза H_0 не принимается, параметр A статистически значим.

Так как $t_{табл} < | t_{факт}(b) |$, то параметр b статистически значим.

Доверительные интервалы:

$b \in (2{,}44; 2{,}69)$ и $A \in (-9{,}76; -8{,}31)$.

Делая обратное преобразование, найдем доверительный интервал для параметра а исходной степенной зависимости:

$$a \in (e^{-9,76}; e^{-8,31}) = (0{,}0000575; 0{,}000245).$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью 0,95 параметры a и b находясь в указанных границах, не принимают нулевых значений, т. е. являются статистически значимыми.

Для показательной зависимости:

$\sigma_A = 0{,}0817; \quad \sigma_B = 0{,}000203.$

$t_{факт}(A) = 46{,}41; \quad t_{факт}(B) = 29{,}92$.

Так как $t_{табл} < | t_{факт}(A) |$, то гипотеза H_0 не принимается, параметр A статистически значим.

Так как $t_{табл} < | t_{факт}(B) |$, то параметр B статистически значим.

Доверительные интервалы:

$A \in (3{,}63; 3{,}95)$ и $B \in (0{,}00567; 0{,}0065)$.

Делая обратное преобразование, найдем доверительные интервалы для параметров а и b исходной модели:

$$a \in (e^{3,63}; e^{3,95}) = (37{,}73; 52{,}23);$$
$$b \in (e^{0,00567}; e^{0,0065}) = (1{,}0057; 1{,}0065).$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью 0,95 параметры a и b, находясь в указанных границах, не принимают нулевых значений, т. е. являются статистически значимыми.

Для равносторонней гиперболы:

С помощью пакета Excel имеем:

$\sigma_a = 80{,}67; \quad \sigma_b = 27973$.

$t_{факт}(a) = 27{,}912; \quad t_{факт}(b) = -21{,}097$.

Так как $t_{табл} > | t_{факт}(a) |$, то гипотеза H_0 принимается, следовательно, параметр *a* статистически незначим.

Так как $t_{табл} > | t_{факт}(b) |$, то параметр *b* статистически незначим.

Доверительные интервалы:

$a \in (2091;\ 2412)$ и $b \in (-645802;\ -534504)$.

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью 0,95 параметры а и b, находясь в указанных границах, не принимают нулевых значений, т. е. являются статистически значимыми.

Выводы:

На основании полученных результатов можно сделать следующие выводы.

Индекс Dow Jones Industrial Average оказывает сильное прямое влияние на индекс РТС, что подтверждается коэффициентом эластичности равным 1,55.

Нефть марки Brent, оказывает умеренное влияние на индекс РТС, что подтверждается коэффициентом парной корреляции равным 0,89 и коэффициентом эластичности равным 0,05.

Исследование курса USD/RUR с одной стороны показало сильное прямое влияние на индекс РТС, коэффициент эластичности равен 2,28, в то время как коэффициент парной корреляции, равный -0,58, указывает на обратную связь. Учитывая, что проверка на значимость коэффициента уравнения регрессии соответствующего курсу USD/RUR показала его незначимость, то будем считать, что влияния курса доллара на индекс РТС не обнаружено.

Анализ уравнения множественной регрессии показал, что среди всех факторов золото имеет наибольшее влияние на индекс РТС.

Для золота коэффициенты парной корреляции и эластичности соответственно равны 0,98 и 2,47. Этот результат оказался несколько неожиданным, поскольку среди ведущих аналитиков Российского фондового рынка никто не отмечал золото как фактор, влияющий на котировки российских акций.

Было принято решение провести подробное исследование влияния золота на индекс РТС. Для этого построили 4 модели парной корреляции: линейную, степенную, показательную и гиперболическую. Это исследование подтвердило, что в самом деле существует сильное прямое влияние динамики цены золото на индексом РТС. При этом зависимость лучше всего описывает степенная модель парной корреляции.

Таким образом, в данном задании описывается, как макроэкономические факторы влияют на Российский фондовый рынок, а следовательно отражает риски с которыми сталкиваются инвесторы на Российском фондовом рынке. Фондовый рынок является индикатором здоровья экономики.

Обычно экономическим кризисам предшествует снижение фондового рынка, а факторы, влияющие на фондовый рынок, как правило, влияют и на экономику страны. Поэтому исследования, проведенные в работе, в очередной раз подтверждают сильную зависимость экономики России от внешних рынков.

4. ЭКОНОМЕТРИЧЕСКИЕ УРАВНЕНИЯ

4.1 Системы эконометрических уравнений

В последние годы в экономических исследованиях для описания связей между показателями (зависимыми переменными) используются системы так называемых одновременных уравнений (структурных) уравнений. Такие модели позволяют учесть зависимость между множеством факторов и, следовательно, более полно отвечают реальным экономическим процессам.

Возможны различные <u>формы записи уравнений.</u>

1. <u>Система независимых уравнений</u> имеет вид:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1m}x_m + \varepsilon_1; \\ y_2 = a_{21}x_1 + a_{22}x_2 + ... + a_{2m}x_m + \varepsilon_2; \\ ... \\ y_n = a_{n1}x_1 + a_{n2}x_2 + ... + a_{nm}x_m + \varepsilon_n. \end{cases} \tag{4.1}$$

В системе уравнений (4.1) каждая зависимая переменная (y) рассматривается как функция одного и того же набора факторов (x); при этом набор факторов можно варьировать, например:

$$y_1 = f(x_1, x_2, x_3, x_4, x_5);$$
$$y_2 = f(x_1, x_3, x_4, x_5);$$
$$y_3 = f(x_2, x_3, x_5);$$
$$y_4 = f(x_3, x_4, x_5).$$

В данной системе независимых уравнений набор факторов в уравнениях изменяется. Причинами отсутствия того или иного фактора в уравнении системы могут быть:

а) несущественность его воздействия на результативный признак (незначимо значение t-критерия или частного F-критерия для данного фактора);

б) экономическая нецелесообразность включения фактора в модель.

Каждое уравнение системы независимых уравнений может рассматриваться самостоятельно. Для нахождения его параметров используется метод наименьших квадратов. По существу каждое уравнение этой системы является уравнением регрессии. Поскольку никогда нет уверенности в том, что факторы полностью объясняют зависимые переменные, то в уравнениях присутствует свободный член α_0. Так как фактические значения зависимой переменной отличаются от теоретических на величину случайной ошибки, то в каждом уравнении присутствует величина случайной ошибки.

В итоге система независимых уравнений при трёх зависимых переменных и четырёх факторах примет вид:

$$\begin{cases} y_1 = a_{01} + a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + \varepsilon_1; \\ y_2 = a_{02} + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + \varepsilon_2; \\ y_3 = a_{03} + a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + \varepsilon_3. \end{cases}$$

2. Система рекурсивных уравнений имеет вид:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2; \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \varepsilon_3; \\ \ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots b_{nn-1}y_{n-1} + a_{n1}x_1 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

В данной системе зависимая переменная y_i одного уравнения выступает в качестве фактора x в другом уравнении. Можно отметить, что в указанной системе зависимая переменная y включает в каждое последующее уравнение наряду с набором факторов $x = (x_1,\ldots,x_m)$ все зависимые переменные предшествующих уравнений.

ПРИМЕР 4.1. Модель производительности труда и фондоотдача может быть записана в виде:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2. \end{cases}$$

где y_1 – производительность труда;
y_2 – фондоотдача;
x_1 – фондовооружённость труда;
x_2 – энерговооружённость труда;
x_3 – квалификация рабочих.

Каждое уравнение здесь может рассматриваться отдельно (самостоятельно), а его параметры определяются на основе метода наименьших квадратов.

3. Наибольшее распространение в экономических исследованиях получила система взаимозависимых уравнений:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1; \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + \ldots + a_{2m}x_m + \varepsilon_2; \\ \ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots b_{nn-1}y_{n-1} + a_{n1}x_1 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Система взаимозависимых уравнений получила название системы совместных, одновременных уравнений. Тем самым подчёркивается, что в системе одни и те же переменные y одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других.

В эконометрике система одновременных уравнений называется также структурной формой модели. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно для нахождения его параметров, традиционный МНК неприменим. С этой целью используются специальные методы оценивания.

ПРИМЕР 4.2. Модель динамики цены и заработной платы представлена в виде:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2. \end{cases}$$

где y_1 – темп изменения месячной заработной платы;
y_2 – темп изменения цен; x_1 – процент безработных;
x_2 – темп изменения постоянного капитала;
x_3 – темп изменения цен на импорт сырья.

Структурная и приведённая формы модели

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные обозначены в приведённой ранее системе одновременных уравнений через y. Это зависимые переменные, число которых равно числу уравнений в системе.

Экзогенные переменные обозначаются обычно через x. Это предопределённые переменные, влияющие на эндогенные переменные, но не зависящие от них.

Простейшая структурная форма модели имеет вид:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2. \end{cases}$$

где y_1 и y_2 – эндогенные переменные; x_1 и x_2 – экзогенные переменные.

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. В качестве экзогенных переменных могут выступать, например, климатические условия.

Структурная форма модели в правой части содержит при эндогенных и экзогенных переменных коэффициенты b_i и a_i, которые называются структурными коэффициентами модели.

Использование МНК для оценивания структурных коэффициентов модели даёт, как правило, смещённые и несостоятельные оценки. Поэтому обычно для определения структурных коэффициентов модели структурная форма преобразуется в приведённую форму модели.

Приведённая форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} \hat{y}_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1m}x_m; \\ \hat{y}_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2m}x_m; \\ \ldots \\ \hat{y}_n = \delta_{n1}x_1 + \delta_{n2}x_2 + \ldots + \delta_{nm}x_m. \end{cases}$$

где δ_i – коэффициенты приведённой формы модели.

По своему виду приведённая форма модели ничем не отличается от системы независимых уравнений, параметры которой оцениваются традиционным МНК. Применяя МНК, можно определить оценки параметров, а затем оценить значения эндогенных переменных через экзогенные.

Недостатком приведённой формы модели является отсутствие в явном виде коэффициентов взаимосвязи между эндогенными переменными.

4.2 Модификация метода наименьших квадратов

Переход от приведённой формы модели к структурной форме модели связан с решением задачи идентификации. Идентификация здесь сводится к установлению соответствия между приведённой и структурной формами.

Полный вид структурной модели для случая n=2 и m=3 составляет:

$$\begin{cases} \hat{y}_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + a_{13}x_3; \\ \hat{y}_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3. \end{cases} \tag{4.2}$$

Модель содержит восемь структурных коэффициентов, что соответствует выражению: $n \cdot (n - 1 + m)$.

Приведённая форма модели в полном виде содержит $(n \cdot m)$ параметров. Для данного примера это означает наличие шести коэффициентов приведённой формы модели:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \delta_{13}x_3; \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \delta_{23}x_3. \end{cases} \tag{4.3}$$

На основе шести коэффициентов приведённой формы модели требуется определить восемь структурных коэффициентов структурной формы модели. Следовательно, появляется неоднозначность определения коэффициентов структурной формы модели.

По принципу идентифицируемости структурные модели подразделяют на три вида: а) идентифицируемые; б) неидентифицируемые; в) сверхидентифицируемые.

Модель идентифицируема, если все её структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведённой формы модели, т.е. число параметров структурной форы модели равно числу параметров приведённой формы модели.

Если уменьшить число структурных коэффициентов в модели (4.2), например, приняв $a_{13} = a_{21} = 0$, то она примет вид:

$$\begin{cases} \hat{y}_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2; \\ \hat{y}_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3. \end{cases} \tag{4.4}$$

тогда модель всегда (4.4) будет идентифицируемой.

Модель неидентифицируема, если число приведённых коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведённой формы модели. Структурная модель в полном виде (4.2), содержащая *n* эндогенных и *m* предопределённых переменных в каждом уравнении системы, всегда неидентифицируема.

Модель сверхидентифицируема, если число приведённых коэффициентов больше числа структурных коэффициентов.

Например, если в системе (4.2) принять $a_{13} = a_{21} = a_{22} = 0$, то система уравнений станет сверхидентифицируемой:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2; \\ y_2 = b_{21}y_1 + a_{23}x_3. \end{cases} \tag{4.5}$$

Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решаема, но требует для этого специальных методов расчёта параметров.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся система считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Обозначим число эндогенных переменных в *j*-ом уравнении системы через *H*, а число экзогенных (предопределённых) переменных, которые содержатся в системе, но не входят в данное уравнение, – через *D*, тогда условие идентифицируемости модели можно записать в виде следующего правила:

$D+1=H$ – уравнение идентифицируемо;

$D+1<H$ – уравнение неидентифицируемо;

$D+1>H$ – уравнение сверхидентифицируемо.

ПРИМЕР 4.3. Рассматривается следующая система одновременных уравнений:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2; \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3; \\ y_2 = b_{31}y_1 + b_{32}y_2 + a_{33}x_3 + a_{34}x_4. \end{cases} \qquad (4.6)$$

Первое уравнение идентифицируемо, т.к. в нём присутствуют три эндогенные переменные – y_1, y_2, y_3, т.е. $H=3$, две экзогенные переменные x_1 и x_2, а число отсутствующих экзогенных переменных равно двум – x_3 и x_4, отсюда $D=2$. Тогда имеем равенство $D+1=H=3$, что означает наличие идентифицированного уравнения.

Во втором уравнении $H=2$ (y_1 и y_2) и $D=1$ (x_4), отсюда $D+1=H=2$. Уравнение идентифицируемо.

В третьем уравнении $H=3$ (y_1, y_2, y_3), а $D=2$ (x_1 и x_2). Следовательно $D+1=H=3$, и уравнение идентифицируемо.

В целом система (2.5) идентифицирующаяся.

4.3 Оценивание параметров структурной модели

Коэффициенты структурной модели могут быть оценены различными способами в зависимости от вида системы одновременных уравнений. Наибольшее применение получили следующие методы оценивания:

1. Косвенный метод наименьших квадратов;
2. Двухшаговый метод наименьших квадратов;
3. Трёхшаговый метод наименьших квадратов;
4. Метод максимального правдоподобия с полной информацией;
5. Метод максимального правдоподобия при ограниченной информации.

Косвенный метод наименьших квадратов (КМНК)

Косвенный МНК применяется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы.

1. Структурная модель преобразовывается в приведённую форму модели.
2. Для каждого уравнения приведённой формы модели обычные МНК осуществляются приведённые коэффициенты c_{ij}.
3. Коэффициенты приведённой формы модели трансформируются в параметры структурной модели.

4.4 Типовые задачи построения и анализа эконометрических уравнений

ЗАДАЧА 4.1

Рассмотрим применение КМНК для простейшей идентифицируемой экономической модели с двумя эндогенными и двумя экзогенными переменными:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2. \end{cases}$$

Пусть для построения данной модели мы располагаем некоторой информацией по пяти регионам:

Таблица 4.1.

Регион	y_1	y_2	x_1	x_2
1	2	5	1	3
2	3	6	2	1
3	4	7	3	2
4	5	8	2	5
5	6	5	4	6
Среднее	4	6,2	2,4	3,4

Приведённая форма модели составляет:

$$\begin{cases} y_1 = c_{11}x_1 + c_{12}x_2 + u_1; \\ y_2 = c_{21}x_1 + c_{22}x_2 + u_2. \end{cases}$$

где u_1 и u_2 – случайные ошибки приведённой формы модели

Для каждого уравнения приведённой формы модели применяем традиционный МНК и определяем с-коэффициенты.

Чтобы упростить процедуру расчётов, можно рассчитав с отклонениями от средних уровней, т.е. $y = y - \bar{y}$ и $x = x - \bar{x}$. Тогда для первого уравнения приведённой формы моделей система нормальных уравнений составляет:

$$\begin{cases} \sum y_1 x_1 = c_{11} \cdot \sum x_1^2 + c_{12} \cdot \sum x_1 \cdot x_2; \\ \sum y_1 x_2 = c_{11} \cdot \sum x_1 \cdot x_2 + c_{12} \cdot \sum x_2^2. \end{cases}$$

Применительно к рассматриваемому примеру, используя отклонения от средних уровней, имеем:

$$\begin{cases} 6 = 5{,}2 \cdot c_{11} + 4{,}2 \cdot c_{12}; \\ 10 = 4{,}2 \cdot c_{11} + 17{,}2 \cdot c_{12}. \end{cases}$$

Решая данную систему, получим следующее первое уравнение приведённой формы модели: $y_1 = 0{,}852 \cdot x_1 + 0{,}373 \cdot x_2 + u_1$

Аналогично применяем МНК для второго уравнения приведённой формы модели, получим: $y_2 = c_{21} \cdot x_1 + c_{22} \cdot x_2 + u_2$

Система нормальных уравнений составит:

$$\begin{cases} \sum y_2 x_1 = c_{21} \cdot \sum x_1^2 + c_{22} \cdot \sum x_1 \cdot x_2; \\ \sum y_2 x_2 = c_{21} \cdot \sum x_1 \cdot x_2 + c_{22} \cdot \sum x_2^2. \end{cases}$$

Применительно к примеру 4.3, имеем:

$$\begin{cases} -0{,}4 = 5{,}2 \cdot c_{21} + 4{,}2 \cdot c_{22}; \\ 10 = 4{,}2 \cdot c_{21} + 17{,}2 \cdot c_{22}. \end{cases}$$

Откуда второе приведённое уравнение составит:

$$y_2 = -0{,}072 \cdot x_1 - 0{,}00557 \cdot x_2 + u_2$$

Таким образом, приведённая форма модели имеет вид:

$$\begin{cases} y_1 = 0{,}852 \cdot x_1 + 0{,}373 \cdot x_2 + u_1; \\ y_2 = -0{,}072 \cdot x_1 - 0{,}00557 \cdot x_2 + u_2. \end{cases}$$

Переходим от приведённой формы к структурной форме модели, т.е. к системе уравнений:

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1 + \varepsilon_1; \\ y_2 = b_{21} y_1 + a_{22} x_2 + \varepsilon_2. \end{cases}$$

Для этой цели из первого уравнения приведённой формы модели надо исключить *x*$_2$, выразив его из второго уравнения приведённой формы и подставив в первое:

$$x_2 = \frac{-0{,}072 \cdot x_1 - y_2}{0{,}00557}.$$

Тогда $$\widehat{y}_1 = 0{,}852 \cdot x_1 + 0{,}373 \cdot \left(\frac{-0{,}072 \cdot x_1 - y_2}{0{,}00557} \right);$$

$\text{€}_1 = -66{,}966 \cdot y_2 - 3{,}970 \cdot x_1$ – первое уравнение структурной формы модели.

Чтобы найти второе уравнение структурной модели, обратимся вновь к приведённой форме модели. Для этой цели из второго уравнения приведённой формы модели следует исключить *x*$_1$, выразив его через первое уравнение и подставив во второе:

$$x_1 = \frac{y_1 - 0{,}373 \cdot x_2}{0{,}852}$$

Тогда

$$\hat{y}_2 = -0{,}072 \cdot \left(\frac{y_1 - 0{,}373 \cdot x_2}{0{,}852} \right) - 0{,}00557 \cdot x_2;$$

$\hat{y}_2 = -0{,}085 \cdot y_1 - 0{,}026 \cdot x_2$ – второе уравнение структурной формы модели.

Конечный результат идентификации:

$$\boxed{\begin{cases}\hat{y}_1 = -66{,}966 \cdot y_2 - 3{,}970 \cdot x_1 + \varepsilon_1; \\ \hat{y}_2 = -0{,}085 \cdot y_1 - 0{,}026 \cdot x_2 + \varepsilon_2.\end{cases}}$$

4.5 Нестандартные задачи построения и анализа эконометрических уравнений

ЗАДАЧА 4.2

Изучается модель вида

$$\begin{cases} y = a_1 + b_1(C + D) + \varepsilon_1, \\ C = a_2 + b_2 \cdot y + b_3 \cdot y_{-1} + \varepsilon_2, \end{cases}$$

где y - валовой национальный доход;
y_{-1} - валовой национальный доход предшествующего года;
C - личное потребление;
D - конечный спрос (помимо личного потребления);
ε_1 и ε_2 - случайные составляющие.

информация за девять лет о приростах всех показателей дана в таблице 4.2.

Таблица 4.2

Год	D	y_{-1}	y	C	Год	D	y_{-1}	y	C
1	-6,8	46,7	3,1	7.4	6	44,7	17,8	37,2	8,6
2	22,4	3,1	22,8	30,4	7	23,1	37,2	35,7	30,0
3	-17,3	22,8	7,8	1,3	8	51,2	35,7	46,6	31,4
4	12,0	7,8	21,4	8,7	9	32,3	46,6	56,0	39,1
5	5,9	21,4	17,8	25,8	$\sum$	167,5	239,1	248,4	182,7

Для данной модели была получена система приведённых уравнений:

$$y = 8.219 + 0.6688 \cdot D + 0.2610 \cdot y_{-1},$$

$$C = 8.636 + 0.3384 \cdot D + 0.2020 \cdot y_{-1}.$$

Требуется:

1. Провести идентификацию модели.
2. Рассчитать параметры первого уравнения структурной модели.

Решение:

1. В данной модели две эндогенные переменные (y и C) и две экзогенные переменные (D и y_{-1}). Второе уравнение точно идентифицировано, так как содержит две эндогенные переменные и не содержит одну экзогенную переменную из системы. Иными словами, для второго уравнения имеем по счётному правилу идентификации равенство: 2=1+1.

Первое уравнение сверхидентифицировано, так как в нём на параметры при C и D наложено ограничение: они должны быть равны. В этом уравнении содержится одна эндогенная переменная y. Переменная C в данном уравнении не рассматривается как эндогенная, так как она участвует в уравнении не самостоятельно, а вместе с переменной D. В данном уравнении отсутствует одна экзогенная переменная, имеющаяся в системе. по счётному правилу идентификации получаем: $1+1=2:D+1>H$. Это больше, чем число эндогенных переменных в данном уравнении, следовательно, система сверхидентифицирована.

2. Для определения параметров сверхидентифицированной модели используется двухшаговый метод наименьших квадратов.

Шаг 1. На основе системы приведённых уравнений по точно идентифицированному второму уравнению определим теоретические значения эндогенной переменной C. Для этого в приведённое уравнение

$C = 8.636 + 0.3384 \cdot D + 0.2020 \cdot y_{-1}$

подставим значения D и y_{-1}, имеющиеся в условии задачи. Получим:

$\hat{C}_1 = 15.8$; $\hat{C}_2 = 16.8$; $\hat{C}_3 = 7.4$; $\hat{C}_4 = 14.3$; $\hat{C}_5 = 15.0$;

$\hat{C}_6 = 27.4$; $\hat{C}_7 = 24.0$; $\hat{C}_8 = 33.2$; $\hat{C}_9 = 29.0$.

Шаг 2. По сверхидентифицированному уравнению структурной формы модели заменяем фактические значения C на теоретические $\hat{C}$ и рассчитываем новую переменную $\hat{C}+D$ (таблица 4.3).

Таблица 4.3

Год	D	$\hat{C}$	$\hat{C}+D$	Год	D	$\hat{C}$	$\hat{C}+D$
1	-6,8	15,8	9,0	6	44,7	27,4	72,1
2	22,4	16,8	39,2	7	23,1	24,0	47,1
3	-17,3	7,4	-9,9	8	51,2	33,2	84,4
4	12,0	14,3	26,3	9	32,3	29,0	61,3
5	5,9	15,0	20,9	$\sum$	167,5	182,9	350,4

Далее к сверхидентифицированному уравнению применяется метод наименьших квадратов. Обозначим новую переменную $\hat{C}+D$ через Z. Решаем уравнение

$$y = a_1 + b_1 \cdot Z .$$

Система нормальных уравнений составит:

$$\sum y = n \cdot a_1 + b_1 \cdot \sum Z,$$
$$\sum y \cdot Z = a_1 \cdot \sum Z + b_1 \cdot \sum Z^2,$$

$$248.4 = 9 \cdot a_1 + 350.4 \cdot b_1,$$
$$13508.71 = 350.4 \cdot a_1 + 21142.02 \cdot b_1,$$

$$a_1 = 7.678;\ b_1 = 0.512.$$

итак, первое уравнение структурной модели будет таким:

$$y = 7.678 + 0.512 \cdot (C + D).$$

ЗАДАЧА 4.3

Рассматривается следующая модель:

$C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + U_1$ (функция потребления);

$I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + U_2$ (функция инвестиций);

$r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + U_3$ (функция денежного рынка);

$Y_t = C_t + I_t + G_t$ (тождество дохода),

где C_t - расходы на потребление в период t;

Y_t - совокупный доход в период t;

I_t - инвестиции в период t;

r_t - процентная ставка в период t;

M_t - денежная масса в период t;

G_t - государственные расходы в период t;

C_{t-1} - расходы на потребление в период $t-1$;

I_{t-1} - инвестиции в период $t-1$;

U_1, U_2, U_3 - случайные ошибки.

Требуется:

1. В предположении, что имеются временные ряды данных по всем переменным модели, предложите способ оценки её параметров.
2. Как изменится ваш ответ на вопрос п. 1, если из модели исключить тождество дохода?

Решение:

1. Модель представляет собой систему одновременных уравнений. Для ответа на вопрос о способе оценки параметров модели проверим каждое её уравнение на идентификацию.

Модель включает четыре эндогенные переменные (C_t, I_t, Y_t и r_t) и четыре предопределённые переменные (две экзогенные переменные - M_t и G_t и две лаговые эндогенные переменные - C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для уравнений модели.

I уравнение.

Это уравнение включает две эндогенные переменные (C_t и Y_t) и одну предопределённую переменную (C_{t-1}). Следовательно, число предопределённых переменных. не входящих в это уравнение, плюс 1, больше числа эндогенных переменных, входящих в уравнение: 3+1>2. Уравнение сверхидентифицировано.

II уравнение.

Уравнение II включает две эндогенные переменные (I_t и r_t) и не включает три предопределённые переменные. Как и 1 уравнение, оно сверхидентифицировано.

III уравнение.

Уравнение III тоже включает две эндогенные переменные (Y_t и r_t) и не включает три предопределённые переменные. Это уравнение сверхидентифицировано.

IV уравнение.

Уравнение IV представляет собой тождество, параметры которого известны. Необходимости в его идентификации нет.

Проверим для каждого из уравнений достаточное условие идентификации. для этого составим матрицу коэффициентов при переменных модели:

	C_t	Y_t	C_{t-1}	I_t	r_t	I_{t-1}	M_t	G_t
I уравнение	-1	b_{11}	b_{12}	0	0	0	0	0
II уравнение	0	0	0	-1	b_{21}	b_{22}	0	0
III уравнение	0	b_{31}	0	0	-1	0	b_{32}	0
Тождество	1	-1	0	1	0	0	0	1

В соответствии с достаточным условием идентификации определитель матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, не должен быть равен нулю, а ранг матрицы должен быть равен числу эндогенных переменных модели минус 1, т. е. 4-1=3.

I уравнение.

Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

$$A = \begin{pmatrix} -1 & b_{21} & b_{22} & 0 & 0 \\ 0 & -1 & 0 & b_{32} & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Её ранг равен 3, так как определитель квадратной подматрицы 3×3 этой матрицы не равен нулю:

$$DetA^* = \begin{vmatrix} -1 & b_{21} & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для I уравнения выполняется.

II уравнение.

Выпишем матрицу коэффициентов при переменных, не входящих в уравнение:

$$A = \begin{pmatrix} -1 & b_{11} & b_{12} & 0 & 0 \\ 0 & b_{31} & 0 & b_{32} & 0 \\ 1 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

Её ранг равен трём, так как определитель квадратной подматрицы 3×3 этой матрицы не равен нулю:

$$DetA^* = \begin{vmatrix} -1 & 0 & 0 \\ 0 & b_{32} & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для II уравнения выполняется.

III уравнение.

Выпишем матрицу коэффициентов при переменных, не входящих в уравнение:

$$A = \begin{pmatrix} -1 & b_{12} & 0 & 0 & 0 \\ 0 & 0 & -1 & b_{22} & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Её ранг равен трём, так как определитель квадратной подматрицы 3×3 этой матрицы не равен нулю:

$$DetA^* = \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{vmatrix} \neq 0.$$

Достаточное условие идентификации для III уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицированы. Для оценки параметров каждого из уравнений будем применять двухшаговый МНК.

Шаг 1. Запишем приведённую форму модели в общем виде:

$$C_t = A_1 + A_2 \cdot C_{t-1} + A_3 \cdot I_{t-1} + A_4 \cdot M_t + A_5 \cdot G_t + V_1;$$

$$I_t = B_1 + B_2 \cdot C_{t-1} + B_3 \cdot I_{t-1} + B_4 \cdot M_t + B_5 \cdot G_t + V_2;$$

$$Y_t = D_1 + D_2 \cdot C_{t-1} + D_3 \cdot I_{t-1} + D_4 \cdot M_t + D_5 \cdot G_t + V_3;$$

$$r_t = E_1 + E_2 \cdot C_{t-1} + E_3 \cdot I_{t-1} + E_4 \cdot M_t + E_5 \cdot G_t + V_4;$$

где V_1, V_2, V_3, V_4 - случайные ошибки.

Определим параметры каждого из приведённых выше уранений в отдельности обычным МНК. Затем найдём расчётные значения эндогенных переменных $\hat{Y}_t, \hat{r}_t$, используемых в правой части структурной модели, подставляя в каждое уравнение приведённой формы соответствующее значение предопределённых переменных.

Шаг 2. В исходных структурных уравнениях заменим эндогенные переменные, выступающие в качестве факторных признаков, их расчётными значениями:

$$C_t = a_1 + b_{11} \cdot \hat{Y}_t + b_{12} \cdot C_{t-1} + U_1^*, \text{ где } U_1^* = U_1 + b_{12} \cdot V_1;$$

$$I_t = a_2 + b_{21} \cdot \hat{r}_t + b_{22} \cdot I_{t-1} + U_2^*, \text{ где } U_2^* = U_2 + b_{21} \cdot V_2;$$

$$r_t = a_3 + b_{31} \cdot \hat{Y}_t + b_{32} \cdot M_t + U_3, \text{ где } U_3^* = U_3 + b_{31} \cdot V_3.$$

Применяя к каждому из полученных уравнений в отдельности обычный МНК, определим структурные параметры $a_1, b_{11}, b_{12}, a_2, b_{21}, b_{22}, a_3, b_{31}$ и b_{32}.

2. Если из модели исключить тождество дохода, число предопределённых переменных модели уменьшится на 1 (из модели будет исключена переменная G_t). Число эндогенных переменных модели также снизится на единицу – переменная Y_t станет экзогенной. В правых частях функции потребления и функции денежного рынка будут находится только предопределённые переменные. Функция инвестиций постулирует зависимость эндогенной переменной I_t от эндогенной переменной r_t (которая зависит только от предопределённых переменных) и предопределённой переменной I_{t-1}. Таким образом, мы получим рекурсивную систему. Её параметры можно оценивать обычным МНК, и нет необходимости исследования системы уравнения на идентификацию.

5 ВРЕМЕННЫЕ РЯДЫ В ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЯХ

5.1 Основные понятия и определения временных рядов

Информационной базой для анализа экономических процессов являются динамические и временные ряды. Совокупность наблюдений некоторого явления (показателя), упорядоченная в зависимости от последовательности значений другого явления (признака), называют **динамическим рядом**. Динамические ряды, у которых в качестве признака упорядочения используется время, называют **временными.**

Если в течение длительного времени регулярно фиксировать курсы валют, акций, цены на товары и т.д., то такие данные образуют временные ряды. Сюда относятся данные о выпуске или потреблении различных товаров и услуг по месяцам, кварталам, годам. В производстве временные ряды возникают в случаях:

- измерения количества изделий, выпускаемых подразделениями предприятия за единицу времени (час, смену, декаду);
- оценок количества брака за тот же период времени;
- наблюдения за изменением запасов на складах.

В экономике и в бизнесе временные ряды – это распространённый тип данных. Во временном ряде содержится информация об особенностях и закономерностях протекания процесса, а статистический анализ временных рядов позволяет выявить закономерности и использовать их для оценки характеристик процесса в будущем, т.е. для прогнозирования.

Временной ряд представляет собой набор чисел, привязанный к последовательным, обычно равноотстоящим моментам времени. Числа, составляющие временной ряд и получающиеся в результате наблюдения за ходом некоторого процесса, называются уровнями временного ряда, или *элементами*. Интервал между двумя последовательными моментами времени называют *тактом* (шагом, квантом). Под диной временного ряда понимают количество входящих в него уровней n. Временной ряд обычно обозначают $Y(t)$ или y_t, где $t = 1,2,...,n$.

На рис.5.1 представлены два варианта временных рядов, содержащих 15 уровней (значений) n=15.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или как произведение нескольких компонентов. Модель, в которой временной ряд представляется как сумма трендового, сезонного и случайного компонентов , называется аддитивной моделью временного ряда. Модель, в которой указанные компоненты образуют произведение, называется мультипликативной моделью временного ряда.

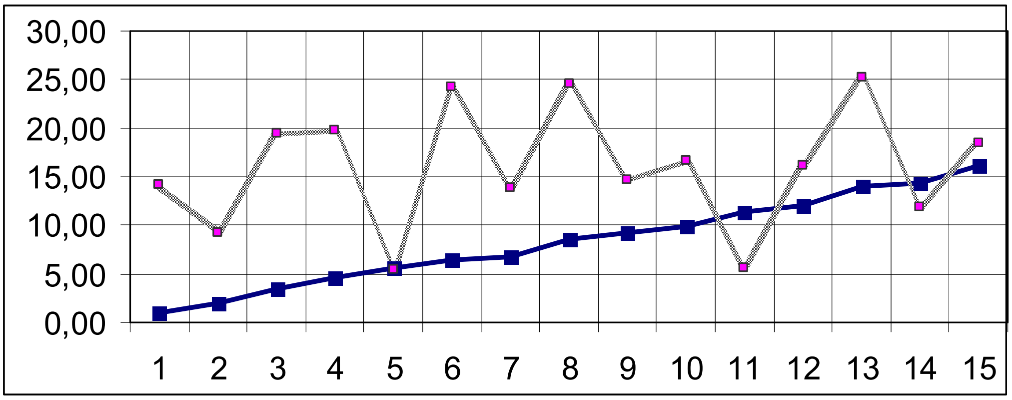


Рис.5.1 - Примеры временных рядов.

Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждому из перечисленных выше компонентов с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

Классические статистические методы исследования исходят из предположения возможности представлять уровни временного ряда в виде суммы нескольких компонентов, отражающих закономерность и случайность развития, в частности, в виде суммы четырёх компонентов:

$$\boxed{Y(t) = f(t) + S(t) + U(t) + E(t)},$$

где $f(t)$ - тренд (долговременная тенденция) развития; $S(t)$ - сезонная составляющая; $U(t)$ - циклическая составляющая; $E(t)$ - остаточный компонент.

В модели временного ряда принято выделять две основные составляющие: детерминированную (систематическую) и случайную (рис. 1.1). Под детерминированной составляющей временного ряда $y_1, y_2, ..., y_n$ понимают числовую последовательность, элементы которой вычисляются по определённому правилу как функция времени t. Исключив детерминированную составляющую из данных, получают колеблющийся около нуля ряд, который может в одном предельном случае представлять собой случайные скачки, а в другом – плавные колебательные движения.

Детерминированная составляющая может содержать следующие структурные компоненты.

1. *Тренд*, или тенденция $f(t)$, представляющая собой устойчивую закономерность, наблюдаемую в течение длительного периода времени В качестве примера таких факторов в экономике можно назвать:

а) изменение демографических характеристик популяции (численности, возрастной структуры);

б) технологическое и экономическое развитие;

в) рост потребления.

Обычно тренд (тенденция) описывается с помощью неслучайной монотонной функции $F_{тр}(t)$, аргументом которой является время t. Эту функцию называют функцией тренда или просто - трендом.

2. *Сезонный компонент* $S(t)$ связан с наличием факторов, действующих с заранее известной периодичностью. Это регулярные колебания, которые носят периодический или близкий к нему характер и заканчиваются в течение года. Типичные примеры сезонного эффекта: изменение загруженности автотрассы в течение суток, по дням недели, временам года, пик продаж товаров для школьников в конце августа – начале сентября. Сезонная составляющая со временем может меняться либо иметь плавающий характер.

3. *Циклический компонент* - неслучайная функция, описывающая длительные периоды (более одного года) относительного подъёма и спада и состоящий из циклов переменной длительности амплитуды. Примером циклического (конъюнктурного) компонента являются волны Кондратьева, демографические «ямы» и т.п. Подобная составляющая весьма характерна для рядов макроэкономических показателей. Здесь циклические изменения обусловлены взаимодействием спроса предложения, а также наложением таких факторов, как истощение ресурсов, погодные условия, изменения в налоговой политике и т.п. Отметим, что циклическую составляющую достаточно трудно идентифицировать формальными методами, исходя только из данных изучаемого ряда.

Случайная составляющая ряда отражает воздействие многочисленных факторов случайного характера и может иметь разнообразную структуру, начиная от простейшей в виде «белого шума» до весьма сложных, описываемых моделями авторегрессии и скользящего среднего.

Основная цель статистического анализа временных рядов – изучение соотношения между закономерностью и случайностью в формировании значений уровней ряда, оценка количественной меры их влияния. Закономерности, объясняющие динамику показателя в прошлом, используются для прогнозирования его значений в будущем, а учёт случайности позволяет определить вероятность отклонения от закономерного развития и его возможную величину.

Требования к исходной информации

Анализ временных рядов, отражающих развитие экономических процессов, начинается с оценки данных. Уровни исследуемого показателя обязательно должны быть сопоставимыми, однородными и устойчивыми, а их число должно быть достаточно велико.

Сопоставимость достигается в результате одинакового подхода к наблюдениям на разных этапах формирования динамического ряда. Уровни во временных рядах должны иметь одинаковые: единицы измерения, шаг наблюдения, интервал времени, методику расчёта, элементы, относящиеся к неизменной совокупности.

Однородность данных означает отсутствие сильных изломов тенденций, а также аномальных (т.е. резко выделяющихся, нетипичных для данного ряда) наблюдений. Аномальные наблюдения проявляются в виде сильного изменения уровня – скачка или спада – с последующим приблизительным восстановлением предыдущего уровня. Наличие аномалии резко снижает качество результатов моделирования. Поэтом аномальные наблюдения ряда необходимо исключить, заменив их расчётными значениями.

Устойчивость характеризуется преобладанием закономерности над случайностью в изменении уровней ряда. На графиках устойчивых временных рядов закономерность прослеживается визуально, на графиках неустойчивых рядов изменения последовательных уровней представляются хаотичными, и поэтому поиск закономерностей в формировании значений уровней таких рядов лишён смысла.

Требование полноты данных обусловливается тем, что закономерность может обнаружиться лишь при наличии минимально допустимого объёма наблюдений.

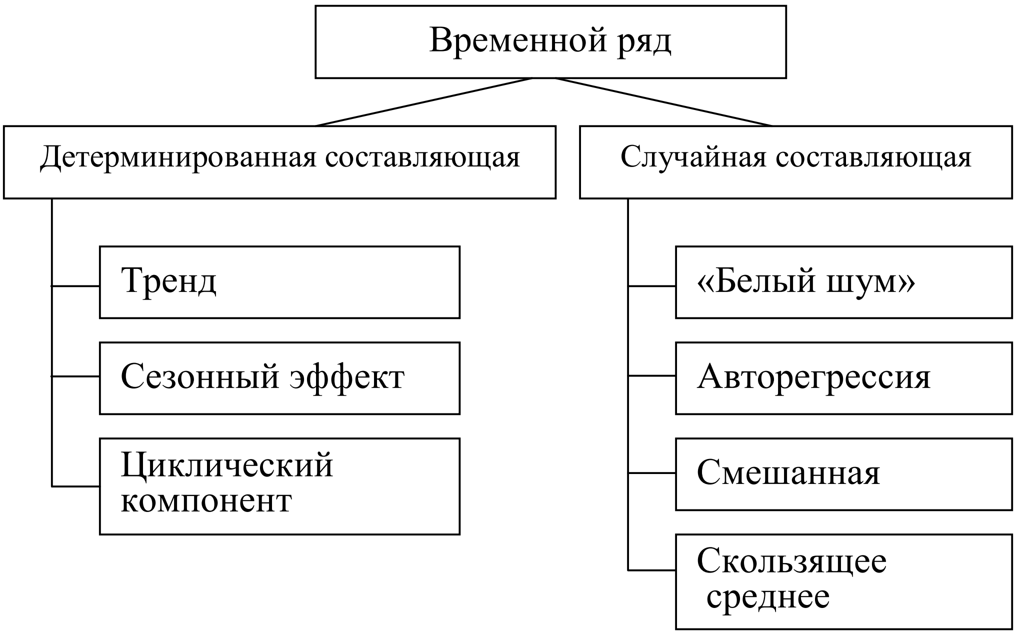


Рисунок 5.2 – Структура временного ряда.

5.2 Анализ структуры и свойств временного ряда

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно её можно оценить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени. Рассмотрим пример.

ПРИМЕР 5.1. Пусть имеются следующие условные данные о средних расходах на конечное потребление y_t за 8 лет (табл.5.1).

Таблица 5.1.

t	y_t	y_{t-1}	$y_t-\bar{y}_1$	$y_{t-1}-\bar{y}_2$	$(y_t-\bar{y}_1)\cdot(y_{t-1}-\bar{y}_2)$	$(y_t-\bar{y}_1)^2$	$(y_{t-1}-\bar{y}_2)^2$
1	7	-	-	-	-	-	-
2	8	7	-3,29	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2,58	1,6641	4
5	11	10	-0,29	0	0	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1
7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16
Итого	86	70	-0,03	0	44,0	53,4287	38

Предположим, что расходы на конечное потребление в текущем году зависят от расходов на конечное потребление предыдущих лет.

Определим коэффициент корреляции между рядами y_t и y_{t-1} и измерим тесноту связи между расходами на конечное потребление текущего и предыдущего годов. Добавим в табл.5.1 временной ряд y_{t-1}.

Одна из рабочих формул для расчёта коэффициента корреляции имеет вид:

$$r_{xy}=\frac{\sum\limits_j (x_j-\bar{x})\cdot(y_j-\bar{y})}{\sqrt{\sum(x_j-\bar{x})^2\cdot\sum(y_j-\bar{y})}}.$$

В качестве переменной x рассмотрим ряд $y_{2,}y_3,...,y_7$. Тогда приведённая выше формула примет вид

$$r_1=\frac{\sum\limits_{t=2}^{n}(y_t-\bar{y}_1)\cdot(y_{t-1}-\bar{y}_2)}{\sqrt{\sum\limits_{t=2}^{n}(y_t-\bar{y}_1)^2\cdot\sum\limits_{t=2}^{n}(y_{t-1}-\bar{y}_2)}}, \quad (5.1) \quad \text{где} \quad \bar{y}_1=\frac{\sum\limits_{t=2}^{n}y_t}{n-1}; \quad \bar{y}_2=\frac{\sum\limits_{t=2}^{n}y_{t-1}}{n-1}. \quad (5.2)$$

Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда t и $t-1$, т.е. при лаге 1.

Для данных примера 5.1 соотношения (5.2) составят:

$$\bar{y}_1=\frac{8+8+10+11+12+14+16}{7}=\frac{79}{7}=11{,}29; \quad \bar{y}_2=\frac{7+8+8+10+11+12+14}{7}=\frac{70}{7}=10{,}0;$$

Используя формулу (5.1), получаем коэффициент автокорреляции первого порядка $r_1 = \frac{44}{\sqrt{53{,}42 \cdot 38}} = 0{,}976$.

Полученное значение коэффициента свидетельствует об очень тесной зависимости между расходами на конечное потребление текущего и непосредственно предшествующего годов и, следовательно, о наличии во временном ряде расходов на конечное потребление сильной линейной тенденции.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-1} и определяется по формуле

$$r_2 = \frac{\sum\limits_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)}{\sqrt{\sum\limits_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum\limits_{t=3}^{n}(y_{t-2} - \bar{y}_4)}}, \text{ где } \bar{y}_3 = \frac{\sum\limits_{t=3}^{n} y_t}{n-2}; \quad \bar{y}_4 = \frac{\sum\limits_{t=3}^{n} y_t}{n-2}. \tag{5.3}$$

Для данных из примера 5.1 получим:

$$\bar{y}_3 = \frac{8+10+11+12+14+16}{6} = \frac{71}{6} = 11{,}83; \quad \bar{y}_4 = \frac{7+8+8+10+11+12}{6} = \frac{56}{6} = 9{,}33;$$

После подстановки численных значений в формулу (5.3) получим:

$$r_2 = \frac{27{,}3334}{\sqrt{40{,}8334 \cdot 19{,}3334}} = 0{,}973.$$

Полученные результаты также подтверждают вывод о том, что ряд расходов на конечное потребление содержит линейную тенденцию.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют лагом. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Как показывает практика, для обеспечения статистической достоверности коэффициентов автокорреляции максимальный лаг не должен превышать величины (n/4).

Выделим два важных свойства коэффициента автокорреляции.

1. Коэффициент автокорреляции 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 тенденции. Дл некоторых временных рядов, имеющих сильную нелинейную тенденцию (например параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержит положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

5.3 Этапы построения прогноза по временным рядам

Формально задача прогнозирования сводится к получению оценок значений ряда для некоторого периода будущего, т.е. к получению значения показателя $Y_{пр}(t)$, где $t = n+1, n+2,...$ При использовании методов экстраполяции исходят из предположения о сохранении закономерностей прошлого развития на период прогнозирования. Во многих случаях при разработке оперативного (до года) и краткосрочного (до 2-х лет) прогноза эти расчёты являются справедливыми.

Экстраполяционное прогнозирование экономических процессов, представленных одномерными временными рядами сводится к выполнению **следующих основных этапов:**

1. Предварительный анализ данных;
2. Построение моделей: формирование набора аппроксимирующих функций (кривых роста) и численное оценивание параметров моделей;
3. Проверка адекватности моделей и оценка их точности; выбор предпочтительной модели;
4. Расчёт точечного и интервального прогнозов.

1. Предварительный анализ данных

На этом этапе производится:

- выявление аномальных наблюдений;
- проверка наличия тренда;
- сглаживание временных рядов;
- расчёт показателей развития динамики экономических процессов.

Так как наличие аномальных наблюдений приводит к искажению результатов моделирования, то необходимо убедиться в отсутствии аномалий данных. В качестве примера аномалии может служить скачок курса доллара, зафиксированный в «чёрный вторник».

Для диагностики аномальных наблюдений разработаны различные подходы, например, метод Ирвина. Для всех или только для подозреваемых в аномальности наблюдений вычисляется величина показателя λ_t :

$$\lambda_t = \frac{|y_t - y_{t-1}|}{S_y}, \qquad (5.3)$$

где $S_y = \sqrt{\dfrac{\sum_{t=1}^{n}(y_t - \bar{y})^2}{n-1}}$ и $\bar{y} = \dfrac{1}{n} \cdot \sum_{t=1}^{n} y_t$.

Если измеряемая величина λ_t превышает табличный уровень (например, для для 10 наблюдений значение критерия Ирвина равно 1,5), то уровень y_t считается аномальным. Аномальные наблюдения необходимо исключить из временного ряда и заменить их расчётными значениями (самый простой способ замены – в качестве нового значения принять среднее из двух соседних значений).

Следующая процедура этапа предварительного анализа данных – *выявление наличия тенденций в развитии исследуемого показателя*. Отметим, что тенденция прослеживается не только в увеличении или уменьшении среднего текущего значения временного ряда, но она присуща и другим его характеристикам: дисперсии, автокорреляции, корреляции с другими показателями и т.д. Тенденцию среднего можно визуально можно определить из графика исходных данных. Наличие тенденции среднего уровня на графике становится более заметным, когда на нём отражены сглаженные значения исходных данных.

Процедура сглаживания необходима при построении некоторых математических моделей и для устранения аномальных наблюдений. Чаще всего для сглаживания применяются методы простого скользящего среднего, взвешенного скользящего среднего и экспоненциального сглаживания.

Традиционными показателями, характеризующими развитие экономических процессов, остаются показатели роста и прироста. Для характеристики динамики изменения экономических показателей на практике используется понятие автокорреляции, которая характеризует не только взаимозависимость уровней одного и то же ряда, относящихся к разным моментам времени наблюдения, но и степень устойчивости развития процесса во времени, величину оптимального периода прогнозирования и т.п.

2 Построение моделей

Для решения задач анализа и моделирования тенденций изменения исследуемого показателя используются модели кривых роста. Плавную кривую (гладкую функцию), аппроксимирующую временной ряд, принято называть кривой роста. Подбор такой кривой является аналитическим (не механическим) выравниванием. Чаще всего используются полиномиальные, экспоненциальные и S-образные кривые роста. Укажем примеры кривых роста: полином первой степени (прямая) $Y(t) = a + b \cdot t$; полином второй степени (парабола) $Y(t) = a + b_1 \cdot t + b_2 \cdot t^2$.

Математические методы позволяют представить прогнозируемую модель в виде полинома любого порядка. Однако без необходимости использование полиномов высокого порядка ($r \geq 3$) представляется излишним.

Параметры кривых роста оцениваются методом наименьших квадратов (МНК), т.е. подбираются таким образом, чтобы график функции «кривой роста» располагался на минимальном удалении от точек исходных данных.

На практике предпочтение отдаётся более простым аналитическим моделям, допускающим содержательную интерпретацию. К числу таких моделей относится линейная модель роста:

$$Y(t) = a + b \cdot t, \tag{5.4}$$

где a и b - параметры (коэффициенты регрессии) модели, $t = 1,2,...,N$.

3 Оценка адекватности регрессионных моделей

Математически критерий оценки параметров модели в соответствии с рекомендациями МНК записывается в виде:

$$S_{\Sigma}(a,b)=\sum_{t=1}^{N}[y_t-(a+b\cdot t)]^2\rightarrow\min\,. \qquad (5.5)$$

Для нахождения минимума функции двух переменных $S_{\Sigma}(a,b)$ следует взять частные производные этой функции по параметрам a и b, а затем приравнять их нулю. В результате получают так называемую систему нормальных уравнений:

$$\begin{cases} a\cdot N+b\cdot\sum\limits_{t=1}^{N}t+\sum\limits_{t=1}^{N}y_t; \\ a\cdot\sum\limits_{t=1}^{N}t+b\cdot\sum\limits_{t=1}^{N}t^2=\sum\limits_{t=1}^{N}y_t\cdot t. \end{cases}$$

Решая систему двух алгебраических уравнений с двумя неизвестными, получим

$$\begin{cases} b=(\sum\limits_{t=1}^{N}(t-\bar{t})\cdot(y_t-\bar{Y}))/\sum\limits_{t=1}^{N}(t-\bar{t})^2; \\ a=\bar{Y}-b\cdot\bar{t} \quad , \end{cases}$$

где $\bar{t}$ и $\bar{Y}$ - средние значения моментов наблюдения и уровней ряда соответственно.

<u>Проверка равенства нулю математического ожидания остатков уровней ряда</u> осуществляется в ходе проверки соответствующей нулевой гипотезы $H_0:|\bar{\varepsilon}|$.

С этой целью строится t- статистика:

$t_{расч}=\dfrac{|\bar{\varepsilon}|}{S_{\varepsilon}}\cdot\sqrt{n}$, где $\bar{\varepsilon}$ - среднее арифметическое значение уровней ряда остатков ε_t;

$$S_{\varepsilon}=\sqrt{\frac{\sum\limits_{t=1}^{N}(\varepsilon_t-\bar{\varepsilon})^2}{n-1}}$$

- среднеквадратическое отклонение для последовательности остатков ε_t, рассчитанное по формуле для малой выборки.

На уровне значимости гипотеза отклоняется, если $t_{расч}>t_{\alpha,\nu}$, где $t_{\alpha,\nu}$ - критерий распределения Стьюдента с доверительной вероятностью $(1-\alpha)$ и степенями свободы $\nu=N-1$.

<u>Проверка условия случайности возникновения отдельных отклонений от тренда.</u> Здесь часто используется критерий, основанный на поворотных точках. Значение случайной переменной считается поворотной точкой, если одновременно больше (меньше) соседних с ним элементов. Если остатки случайны, то поворотная точка приходится примерно на каждые 1,5 наблюдения.

Если их больше, возмущения быстро колеблются, и это не может быть объяснено только случайностью. Если же их меньше, то последовательные значения случайного компонента положительно коррелированны.

Критерий случайности отклонений от тренда при уровне вероятности 0,95 можно представить как

$$h > [\frac{2}{3}\cdot(n-2) - 1{,}96\cdot\sqrt{\frac{16\cdot n-29}{90}}\,], \tag{5.6}$$

где h - фактическое количество поворотных точек в случайном ряду;

1,96 - квантиль нормального распределения для 5% -го уровня значимости.

Квадратные скобки здесь означают, что от результата следует взять только целую часть. Если неравенство (2.6) не выполняется, то ряд остатков нельзя считать случайным, т.е. он содержит случайный компонент и, следовательно, модель не является адекватной.

<u>Наличие (отсутствие) автокорреляции в отклонениях от модели роста проверяют</u> с помощью критерия Дарбина-Уотсона (статистика), в основе которой лежит формула

$$\boxed{d = \left(\sum_{t=2}^{N}(\varepsilon_t - \varepsilon_{t-1})^2\right) / \sum_{t=1}^{N}\varepsilon_t^2} \tag{5.7}$$

При отсутствии автокорреляции показатель $d \approx 2$, а при полной автокорреляции он равен 0 или 4. Следовательно, оценки, получаемые по критерию, являются не точечными, а интервальными. Верхние d_2 и нижние d_1 критические значения, позволяющие принять или опровергнуть гипотезу об отсутствии автокорреляции, зависят от количества уровней динамического ряда и числа независимых переменных модели. Значения этих границ для уровня значимости $\alpha = 0{,}05$ даны в специальных таблицах (см. прилож.4). При сравнении расчётного значения d-статистики (5.7) с табличным могут возникнуть такие ситуации:

а) $d_2 < d < 2$ - ряд остатков не коррелирован;

б) $d < d_1$ - остатки содержат автокорреляцию;

в) $d_1 < d < d_2$ - область неопределённости, кода нет оснований ни принять, ни опровергнуть гипотезу о наличии отрицательной корреляции.

Перед сравнением с табличными значениями d критерий следует преобразовать по формуле $d' = 4 - d$.

Установив наличие автокорреляции остатков изменяют математическую модель процесса. Если же ситуация остаётся неопределённой, применяют другие критерии. В частности, можно воспользоваться первым коэффициентом автокорреляции остатков

$$r(1) = \left(\sum_{t=2}^{n}\varepsilon_{t\cdot}\cdot\varepsilon_{t-1}\right) / \sum_{t=1}^{n}\varepsilon_t^2 \ . \tag{5.8}$$

Для принятия решения о наличии или отсутствии автокорреляции в исследуемом ряду фактическое значение коэффициента автокорреляции $r(1)$ сопоставляется с табличным (критическим) для 5%-ного уровня значимости (вероятности допустить ошибку при принятии нулевой гипотезы о независимости уровней ряда). Если фактическое значение коэффициента автокорреляции меньше табличного, то гипотеза об отсутствии автокорреляции в ряду может быть принята, а если фактическое значение больше табличного – делают вывод о наличии автокорреляции в ряду динамики.

Соответствие ряда остатков нормальному закону распределения вероятностей можно проверить с помощью RS-критерия:

$$RS = (\varepsilon_{\max} - \varepsilon_{\min}) / S_\varepsilon , \qquad (5.9)$$

где $\varepsilon_{\max}$ *и* $\varepsilon_{\min}$ - соответственного максимальный и минимальный уровни ряда остатков; S_ε - среднеквадратическое отклонение ряда остатков

$$\boxed{S_\varepsilon = \sqrt{\sum_{t=1}^{N} \varepsilon_t^2 /(n-1)}}.$$

Если расчётное значение критерия RS попадает между табулированными границами с заданным уровнем вероятности, то гипотеза о нормальном распределении ряда остатков принимается. В этом случае допустимо строить доверительный интервал прогноза. Если все пункты проверки дают положительный результат, то выбранная трендовая модель является адекватной реальному ряду экономической динамики, и, следовательно, её можно использовать для построения прогнозных оценок. В противном случае - модель надо улучшать.

4. Построение точечных и интервальных прогнозов

Идея социально-экономического прогнозирования базируется на предположении, что закономерность развития, действовавшая в прошлом (внутри ряда экономической динамики), сохранится и в прогнозируемом будущем. В этом случае прогноз основан на экстраполяции. Экстраполяция, проводимая в будущее, называется *перспективной*, а в прошлое – *ретроспективной*.

Прогнозирование методом экстраполяции базируется на следующих положениях:

а) развитие исследуемого явления в целом описывается плавной кривой;

б) общая тенденция развития явления в прошлом и настоящем не указывает на серьёзные изменения в будущем;

в) учёт случайности позволяет оценить вероятность отклонения от закономерного развития.

Поэтому надёжность и точность прогноза зависят от того, насколько близкими к действительности окажутся эти предположения и насколько точно удалось охарактеризовать выявленную в прошлом закономерность.

На основе построенной модели рассчитываются точечные и интервальные прогнозы. Точечный прогноз на основе временных моделей получается подстановкой в модель (уравнение тренда) соответствующего значения фактора времени, т.е. $t = n+1, n+2, ..., n+k$.

Точное совпадение фактических данных и прогностических точечных оценок, полученных путём экстраполяции кривых, характеризующих тенденцию, имеет малую вероятность. Возникновение соответствующих отклонений объясняется следующими причинами.

1. Выбранная для прогнозирования кривая не является единственно возможной для описания тенденции. Можно подобрать такую кривую, которая даёт более точные результаты.

2. Прогноз осуществляется на основании ограниченного числа исходных данных. Кроме того, каждый исходный уровень обладает ещё и случайным компонентом. Поэтому и кривая, по которой осуществляется экстраполяция, также будет содержать случайный компонент.

3. Тенденция характеризует движение среднего уровня ряда динамики, поэтому отдельные наблюдения могут от него отклоняться. Если такие отклонения наблюдались в прошлом, то они будут наблюдаться и в будущем.

Интервальные прогнозы строятся на основе точечных прогнозов. *Доверительным интервалом* называется такой интервал, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значение прогнозируемого показателя. Ширина интервала зависит от качества модели, т.е. от степени её близости к фактическим данным, числа наблюдений, горизонта прогнозирования и выбранного пользователем уровня вероятности.

При построении доверительного интервала прогноза рассчитывается величина $U(k)$, которая для линейной модели имеет вид

$$U(k) = S_{\hat{Y}} \cdot t_\alpha \cdot \sqrt{1 + \frac{1}{n} + \frac{(n+k-\bar{t})^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}}\text{, где } S_{\hat{Y}} = \sqrt{\frac{\sum_{t=1}^{n}\varepsilon_t^2}{n-m-1}}. \qquad (5.10;\ 5.11)$$

$S_{\hat{Y}}$ - стандартная ошибка (среднеквадратическое отклонение от модели);

m - количество факторов в модели, для линейной модели $m = 1$.

Коэффициент t_α является табличным значением t-статистики Стьюдента при заданном уровне значимости и числе наблюдений. Если исследователь задаёт уровень вероятности попадания прогнозируемой величины внутрь доверительного интервала, равной 70%, то при $n = 9$ параметр t_α=1,12. При вероятности, равной 95%, t_α=2,36.

Для других моделей величина $U(k)$ рассчитывается аналогичным образом, но имеет более громоздкий вид. Как видно из формулы (5.10), величина U зависит прямо пропорционально от точности модели, коэффициента доверительной вероятности t_α, степени углубления в будущее на k шагов вперёд, т.е. на момент $t = n + k$, и обратно пропорциональна объёму наблюдений. Доверительный интервал прогноза будет иметь следующие границы:

- верхняя граница прогноза $Y_{прог}(n+k) + U(k)$;

- нижняя граница прогноза $Y_{прог}(n+k)-U(k)$.

Если построенная модель адекватна, то с выбранной пользователем вероятностью можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина попадёт в интервал, образованный нижней и верхней границами.

После получения прогнозных оценок необходимо убедиться в их обоснованности и непротиворечивости оценкам, полученным иным способом.

Все задачи эконометрического анализа основываются на исходных данных, которые принято записывать в форме матрицы:

$$P=\left\| x_i^{\,j}(t) \right\|, \qquad i=1,\ldots,n; \;\; j=1,\ldots,m. \qquad (5.13)$$

Здесь $x_i^j(t)$ - это значение j-го анализируемого признака, характеризующего состояние i-го объекта в момент времени t.

Данные образуют пространственно-временную выборку, если время t не фиксировано, пробегает значения t_1, t_2, …,t_N. Данные, представляемые в виде такого упорядоченного набора матриц, называются панельными данными. Хронологическая последовательность значений данного признака, характеризующего состояние данного объекта, т.е. одного элемента матрицы $P(\cdot)$, приводит к понятию одномерного временного ряда.

5.4 Задача проверки наличия тренда

Один из способов обнаружения тренда основан на сравнении средних уровней ряда: временной ряд разбивают на две примерно равные по числу уровней ряда части, каждая из которых рассматривается как некоторая самостоятельная выборочная совокупность, имеющая нормальное распределение. Если временной ряд имеет тенденцию к тренду, то средние, вычисленные для каждой совокупности, должны существенно (значимо) различаться между собой. Если же расхождение незначительно, несущественно (случайно), то временной ряд не имеет тенденции.

Таким образом, проверка наличия тренда в исследуемом ряду сводится к проверке гипотезы о равенстве средних двух нормальных распределённых совокупностей.

Определим наличие основной тенденции (тренда) по статистическим данным, представленным в табл.5.2.

Таблица 5.2

Годы	1	2	3	4	5	6	7	8
Урожайность	14,1	9,3	19,4	19,7	5,4	24,2	13,8	24,5
Годы	9	10	11	12	13	14	15	16
Урожайность	14,7	16,6	5,6	16,2	25,3	11,9	18,5	13,0

Решение

1. Делят исходный временной ряд на две равные по числу уровней части: $n_1 = 8; n_2 = 8;\quad (n_1 + n_2 = n = 16)$.

2. Для каждой из этих частей вычисляют средние значения: $Y_1 = 15{,}13;\quad Y_2 = 16{,}66$ и дисперсии $D_{Y1} = 42{,}15$ и $D_{Y2} = 41{,}22$.

3. Проверяют гипотезу о равенстве (однородности) обеих частей ряда с помощью F-критерия Фишера. Для вычисления F-критерия Фишера большую дисперсию делят на меньшую:

$$F_{\text{расч}} = D_{Y1} / D_{Y2} = 42{,}15 / 41{,}22 = 1{,}022,\quad F_{\text{кр}} = (0{,}05;\ 6{,}67) = 3{,}86.$$

Так как $F_{\text{расч}} < F_{\text{кр}}(0{,}05;\ 7{,}6)$, то нет оснований отвергать нулевую гипотезу. По данным наблюдения дисперсии генеральных совокупностей примерно равны $D_1 = D_2$. Исправленные выборочные дисперсии ($S^2_{y2}\ и\ S^2_{y1}$) различаются незначительно (расхождение между ними величина случайная).

4. На основании полученных данных основную гипотезу о равенстве средних значений можно проверить с использованием t-критерия Стьюдента:

$$t = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{(n_1 - 1)\cdot S_1^2 + (n_2 - 1)\cdot S_2^2}} \cdot \sqrt{\frac{n_1 \cdot n_2 (n_1 + n_2 - 2)}{n_1 + n_2}},$$

подставляя числовые значения, получим

$$t = \frac{15?13 - 16?66}{\sqrt{6 \cdot 42{,}146 + 7 \cdot 41{,}22}} \cdot \sqrt{\frac{7 \cdot 8 \cdot 13}{15}} = -0{,}46.$$

Табличное значение $t_{\text{кр}}(0{,}05;\ 13) = 2{,}16$. Так как $\left| t_{\text{расч}} \right| < t_{\text{кр}}$, то нет оснований отвергать нулевую гипотезу о равенстве средних, расхождение между вычисленными средними незначимо.

ВЫВОД: тренд урожайности ячменя отсутствует.

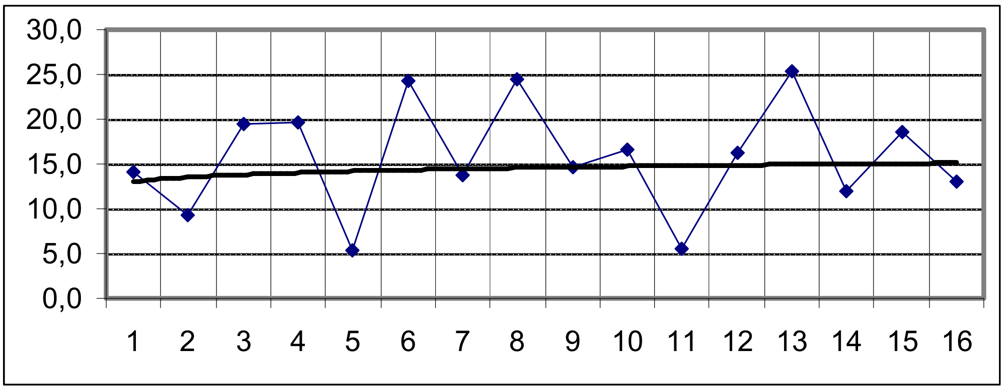


Рисунок 5.3 – График урожайности ячменя

5.5 Задачи сглаживания уровней временного ряда

ЗАДАЧА 5.2

Требуется выполнить сглаживание временного ряда (табл.5.2) по методу скользящего среднего.

Таблица 5.2. Исходные данные для анализа временного ряда.

n	1	2	3	4	5	6
x(n)	99	102,5	104,5	109,0	114,0	117,5
y(n)	29	34	38	43	49	52
n	7	8	9	10	11	12
x(n)	109,5	102,0	108,5	112,0	115,5	119,0
y(n)	53	47	50	52	55	59

Выполним сглаживание временного ряда (табл. 5.2) по методу скользящего среднего. При этом используем формулу:

$$\widehat{x}(k)=\frac{1}{2m+1}\cdot\sum_{i=-m}^{m}x(k+i),\ \text{где}\ k=\overline{(m+1),(N-m)},\quad m=1;\ N=12.$$

Следовательно, расчеты выполняем для значений $k=\overline{2{,}11}$

$$x(2)=\frac{1}{3}\cdot[x(-1+2)+x(0+2)+x(1+2)]=\frac{1}{3}[x(1)+x(2)+x(3)]=$$

$$=\frac{99+102{,}5+104{,}5}{3}=\frac{306}{3}=102;$$

$$x(3)=\frac{1}{3}\cdot[x(2)+x(3)+x(4)]=\frac{102{,}5+104{,}5+109{,}0}{3}=\frac{316}{3}=105{,}33;$$

$$x(4)=\frac{1}{3}\cdot[x(3)+x(4)+x(5)]=\frac{104{,}5+109{,}0+114{,}0}{3}=\frac{327{,}5}{3}=109{,}17;$$

$$x(5)=\frac{1}{3}\cdot[x(4)+x(5)+x(6)]=\frac{109{,}0+114{,}0+117{,}5}{3}=\frac{340{,}5}{3}=113{,}5;$$

$$x(6)=\frac{1}{3}\cdot[x(5)+x(6)+x(7)]=\frac{114{,}0+117{,}5+109{,}5}{3}=\frac{341}{3}=113{,}67;$$

$$x(7)=\frac{1}{3}\cdot[x(6)+x(7)+x(8)]=\frac{117{,}5+109{,}5+102{,}0}{3}=\frac{329}{3}=109{,}67;$$

$$x(8)=\frac{1}{3}\cdot[x(7)+x(8)+x(9)]=\frac{109{,}5+102{,}0+108{,}5}{3}=\frac{320}{3}=106{,}67;$$

$$x(9)=\frac{1}{3}\cdot[x(8)+x(9)+x(10)]=\frac{102{,}0+108{,}5+112{,}0}{3}=\frac{322{,}5}{3}=107{,}5;$$

$$x(10)=\frac{1}{3}\cdot[x(9)+x(10)+x(11)]=\frac{108{,}5+112{,}0+115{,}5}{3}=\frac{336}{3}=112;$$

$$x(11)=\frac{1}{3}\cdot[x(10)+x(11)+x(12)]=\frac{112{,}0+115{,}5+119{,}0}{3}=\frac{346{,}5}{3}=115{,}5.$$

Принимаем дополнительно: $\widehat{x}(1)=x(1)$; $\widehat{x}(12)=x(12)$.
Для расчета значений $\widehat{y}(k)$ используем формулу:

$$\widehat{y}(k)=\frac{1}{2m+1}\cdot\sum_{i=-m}^{m}y(k+i),\quad k=\overline{2{,}11}.$$

$$y(2)=\frac{1}{3}\cdot[y(-1+2)+y(0+2)+y(1+2)]=\frac{1}{3}[y(1)+y(2)+y(3)]=$$

$$=\frac{29+34+38}{3}=\frac{101}{3}=33{,}67;$$

$$y(3)=\frac{1}{3}\cdot[y(2)+y(3)+y(4)]=\frac{34+38+43}{3}=\frac{115}{3}=38{,}33;$$

$$y(4)=\frac{1}{3}\cdot[y(3)+y(4)+y(5)]=\frac{38+43+49}{3}=\frac{130}{3}=43{,}33;$$

$$y(5)=\frac{1}{3}\cdot[y(4)+y(5)+y(6)]=\frac{43+49+52}{3}=\frac{144}{3}=48;$$

$$y(6)=\frac{1}{3}\cdot[y(5)+y(6)+y(7)]=\frac{49+52+53}{3}=\frac{154}{3}=51{,}33;$$

$$y(7)=\frac{1}{3}\cdot[y(6)+y(7)+y(8)]=\frac{52+53+47}{3}=\frac{152}{3}=50{,}67;$$

$$y(8)=\frac{1}{3}\cdot[y(7)+y(8)+y(9)]=\frac{53+47+50}{3}=\frac{150}{3}=50;$$

$$y(9)=\frac{1}{3}\cdot[y(8)+y(9)+y(10)]=\frac{47+50+52}{3}=\frac{149}{3}=49{,}67;$$

$$y(10)=\frac{1}{3}\cdot[y(9)+y(10)+y(11)]=\frac{50+52+55}{3}=\frac{157}{3}=52{,}33;$$

$$y(11) = \frac{1}{3}\cdot[y(10) + y(11) + y(12)] = \frac{52+55+59}{3} = \frac{166}{3} = 55{,}33;$$

Принимаем граничные условия $\hat{y}(1) = y(1)$ и $\hat{y}(12) = y(12)$ и заполняем 4-ю и 5-ю строки таблицы 2.

Таблица 5.3

n	1	2	3	4	5	6
x(n)	99	102,5	104,5	109,0	114,0	117,5
y(n)	29	34	38	43	49	52
$\hat{x}(n)$	99	102	105,33	109,17	113,5	113,67
$\hat{y}(n)$	29	33,67	38,33	43,33	48	51,33
n	7	8	9	10	11	
x(n)	109,5	102,0	108,5	112,0	115,5	
y(n)	53	47	50	52	55	
$\hat{x}(n)$	109,67	106,67	107,5	112	115,5	
$\hat{y}(n)$	50,67	50	49,67	52,33	55,33	

Реализация задачи с помощью ППП Excel

1. Для определения параметров линейного тренда по методу наименьших квадратов используется статистическая функция **ЛИНЕЙН**, для определения экспоненциального тренда – **ЛГРФПРИБЛ**. В качестве зависимой переменной в данной задаче выступает время. Приведем результаты вычисления функции **ЛИНЕЙН** и **ЛГРФПРИБЛ** (рис. 6.1 и 6.2).

Запишем уравнения линейного и экспоненциального тренда, используя данные рис. 6.1 и 6.2:

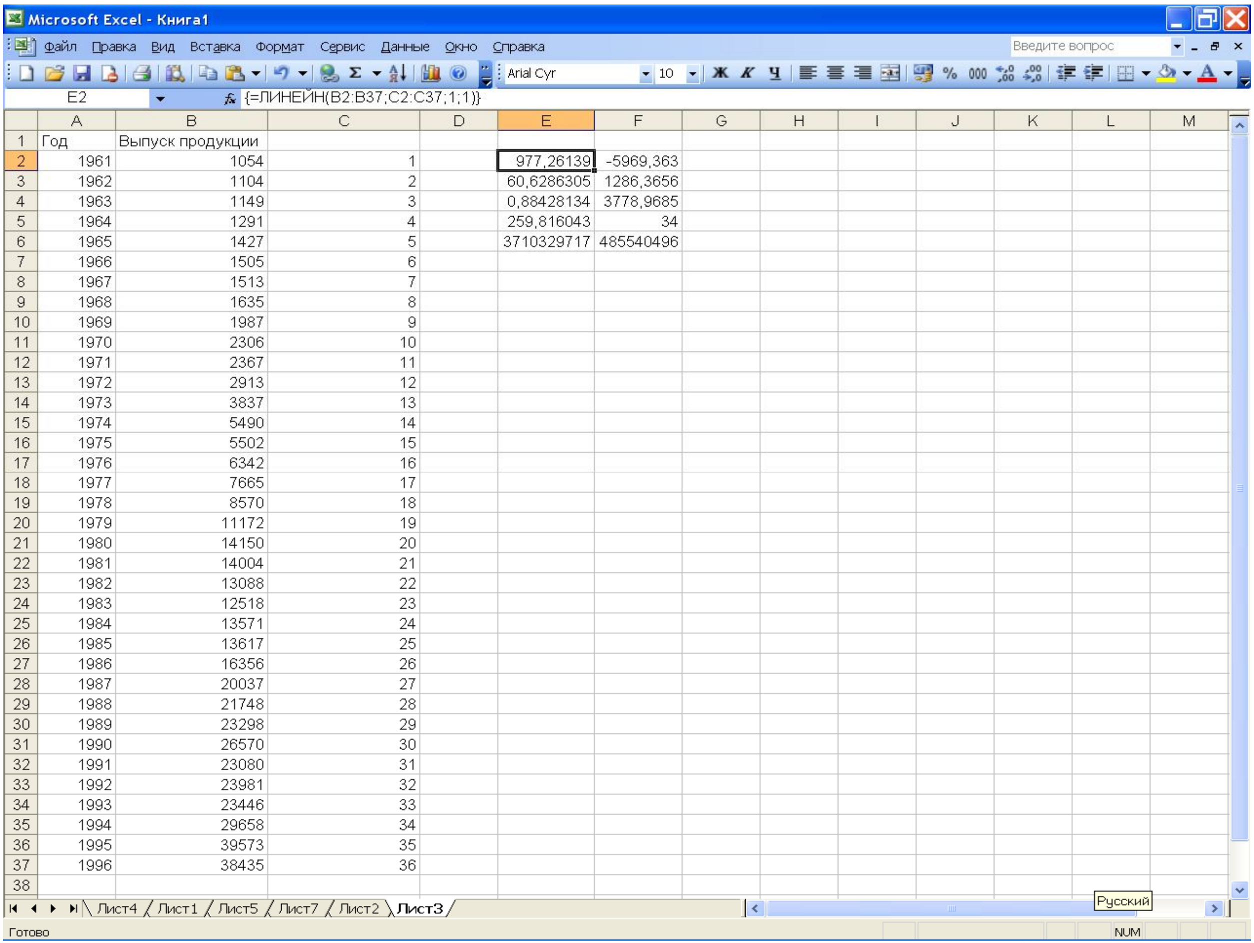
$$\hat{y}_t = -1921124{,}37 + 977{,}12t;$$

$$\hat{y}_t = -1{,}0045^t.$$

2. Построение графиков осуществляется с помощью Мастера диаграмм.

Порядок построения следующий:

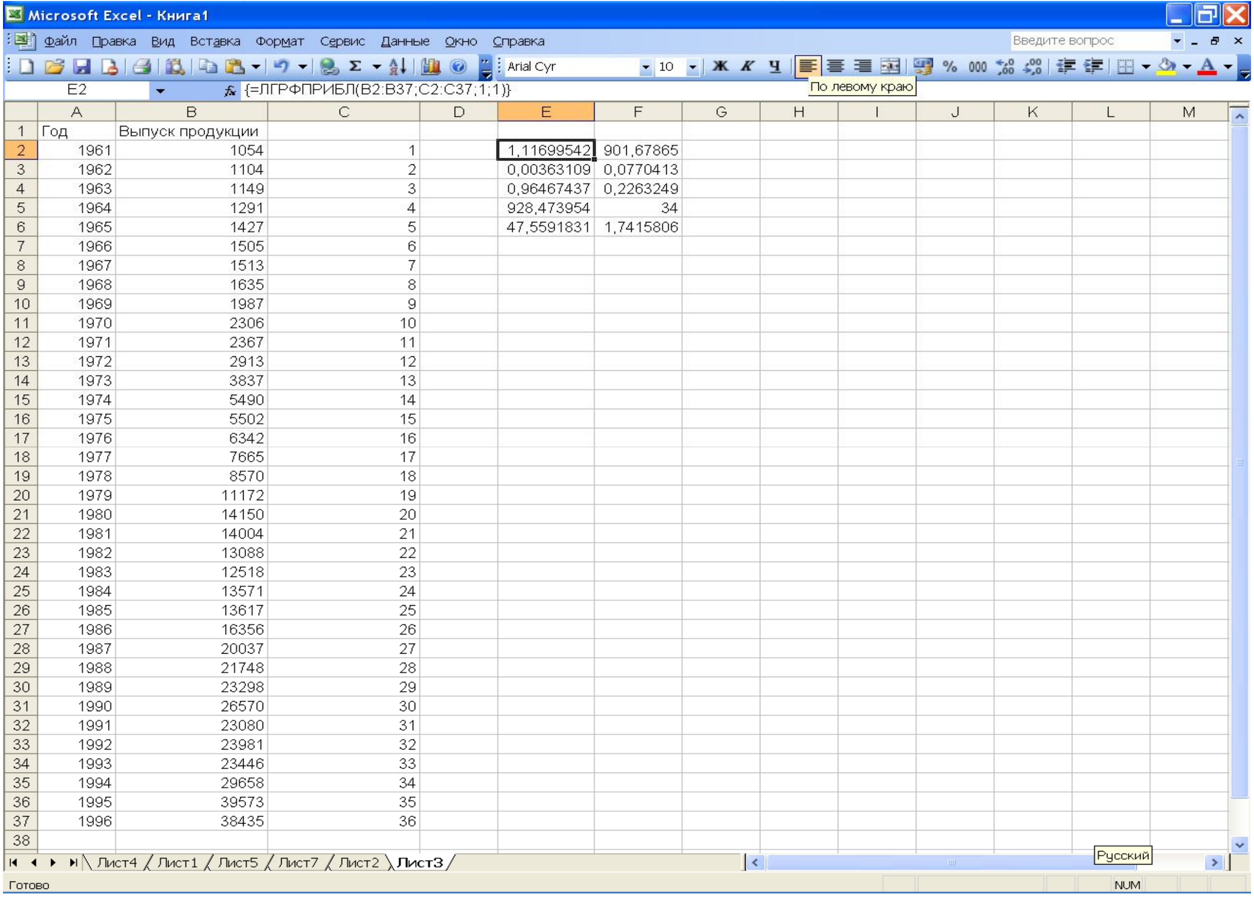
1) введите исходные данные или откройте существующий файл, содержащий анализируемые данные;
2) активизируйте Мастер диаграмм любым из следующих способов:
   а) в главном меню выберите **Вставка/Диаграмма**;
   б) на панели инструментов **Стандартная** щелкните по кнопке **Мастер диаграмм**;
3) в окне Тип выберите **График** (рис. 6.3); вид графика выберите в поле рядом со списком типов. Щелкните по кнопке **Далее**;



E2 {=ЛИНЕЙН(B2:B37;C2:C37;1;1)}

	A	B	C	D	E	F
1	Год	Выпуск продукции				
2	1961	1054	1		977,26139	-5969,363
3	1962	1104	2		60,6286305	1286,3656
4	1963	1149	3		0,88428134	3778,9685
5	1964	1291	4		259,816043	34
6	1965	1427	5		3710329717	485540496
7	1966	1505	6			
8	1967	1513	7			
9	1968	1635	8			
10	1969	1987	9			
11	1970	2306	10			
12	1971	2367	11			
13	1972	2913	12			
14	1973	3837	13			
15	1974	5490	14			
16	1975	5502	15			
17	1976	6342	16			
18	1977	7665	17			
19	1978	8570	18			
20	1979	11172	19			
21	1980	14150	20			
22	1981	14004	21			
23	1982	13088	22			
24	1983	12518	23			
25	1984	13571	24			
26	1985	13617	25			
27	1986	16356	26			
28	1987	20037	27			
29	1988	21748	28			
30	1989	23298	29			
31	1990	26570	30			
32	1991	23080	31			
33	1992	23981	32			
34	1993	23446	33			
35	1994	29658	34			
36	1995	39573	35			
37	1996	38435	36			
38						

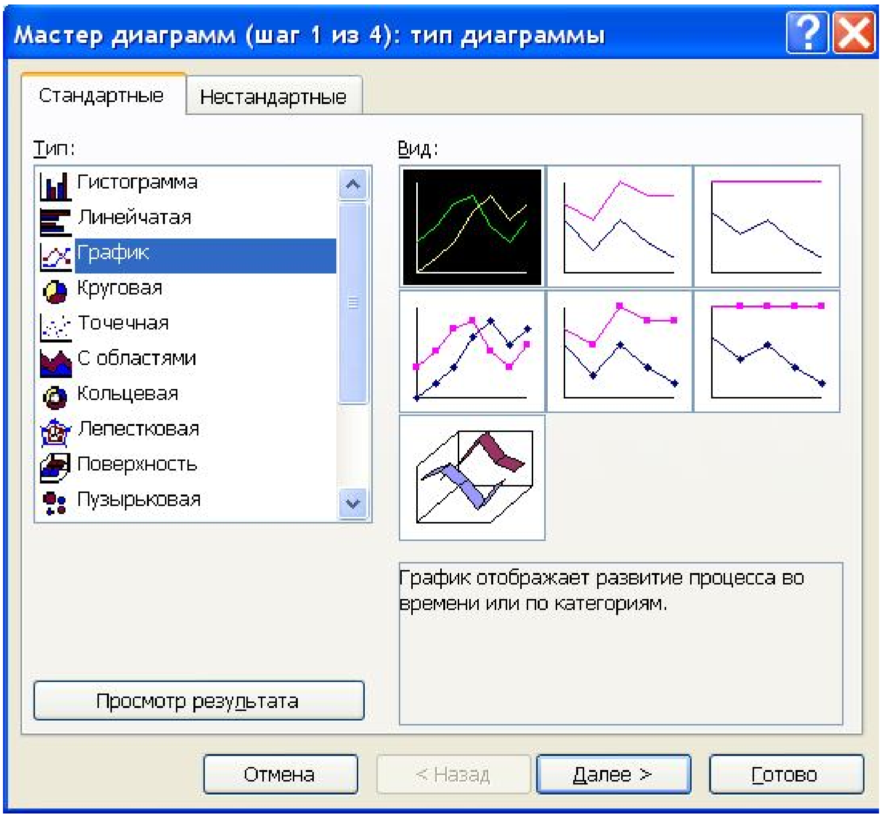
Рисунок 6.1. Результат вычисления функции **ЛИНЕЙН**



E2 {=ЛГРФПРИБЛ(B2:B37;C2:C37;1;1)}

	A	B	C	D	E	F
1	Год	Выпуск продукции				
2	1961	1054	1		1,11699542	901,67865
3	1962	1104	2		0,00363109	0,0770413
4	1963	1149	3		0,96467437	0,2263249
5	1964	1291	4		928,473954	34
6	1965	1427	5		47,5591831	1,7415806
7	1966	1505	6			
8	1967	1513	7			
9	1968	1635	8			
10	1969	1987	9			
11	1970	2306	10			
12	1971	2367	11			
13	1972	2913	12			
14	1973	3837	13			
15	1974	5490	14			
16	1975	5502	15			
17	1976	6342	16			
18	1977	7665	17			
19	1978	8570	18			
20	1979	11172	19			
21	1980	14150	20			
22	1981	14004	21			
23	1982	13088	22			
24	1983	12518	23			
25	1984	13571	24			
26	1985	13617	25			
27	1986	16356	26			
28	1987	20037	27			
29	1988	21748	28			
30	1989	23298	29			
31	1990	26570	30			
32	1991	23080	31			
33	1992	23981	32			
34	1993	23446	33			
35	1994	29658	34			
36	1995	39573	35			
37	1996	38435	36			
38						

Рисунок 6.2. Результат вычисления функции **ЛГРФПРИБЛ**



Тип: График

График отображает развитие процесса во времени или по категориям.

Рисунок 6.3. Диалоговое окно Мастера диаграмм: тип диаграммы

5.6 Задачи экономического прогнозирования

Одной из важнейших задач (этапов) анализа временного ряда является прогнозирование на его основе развития изучаемого процесса. При этом исходят из того, что тенденция развития, установленная в прошлом, может быть распространена (экстраполирована) на будущий период.

Задача прогнозирования может быть сформулирована так: *имеется временной (динамический) ряд* y_t $(t = 1, 2, ..., n)$ *и требуется дать прогноз уровня этого ряда на момент времени* $n + m$. *Здесь* m *- интервал прогноза.*

Если рассматривать временной ряд как регрессионную модель изучемого признака по переменной «время», то к нему могут быть применены методы точечного и интервального прогноза значений зависимой переменной Y. Точечные и интервальные оценки Y определяют с использованием полученных ранее моделей парной и множественной регрессии для значений объясняющих переменных X, расположенных вне пределов изученного диапазона значений X. Вместе с тем напомним, что одна из основных предпосылок регрессионного анализа состоит в том, что возмущения ε_t представляют собой независимые случайные величины с математическим ожиданием (средним значением), равным нулю. При работе с временными рядами такое допущение требует дополнительного обоснования. Ниже будем полагать, что возмущения ε_t, $t = 1,2,...,n$, удовлетворяют предпосылкам регрессионного анализа, т.е. условиям нормальной классической регрессионной модели.

ЗАДАЧА 5.4

Пусть известны статистические данные, определяющие спрос на продукты питания (табл.5.5). Необходимо дать точечную и (с надёжностью 0,95) интервальные оценки прогноза среднего и индивидуального значений спроса на товар на момент времени $t = 9$ (девятый год). Принимаем условие, что тренд линейный, а возмущения удовлетворяют требованиям классической модели.

Таблица 5.5

Год, t	1	2	3	4	5	6	7	8
Спрос, y_t	213	171	291	309	317	362	351	361

Решение

Пусть в соответствии с известными рекомендациями получено уравнение парной регрессии вида $\widehat{y}_t = 181{,}32 + 25{,}679 \cdot t$, согласно которому ежегодный спрос на товар увеличивался в среднем на 25,7 ед. Надо оценить условное математическое ожидание $M(Y)_{t=9} = \overline{y}(9)$. Оценкой $\overline{y}(9)$ является групповая средняя $\widehat{y}_{t=9} = 181{,}32 + 25{,}679 \cdot 9 = 412{,}4$ (ед.).

Определим оценку s^2 дисперсии σ^2:

$$s^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{7059{,}2}{8-2} = 1176{,}5.$$

Вычислим оценку дисперсии групповой средней:

$$s_{\widehat{y}_9}^2 = 1176{,}5 \cdot \left(\frac{1}{8} + \frac{(9-4{,}5)^2}{42}\right) = 714{,}3; \quad s_{\widehat{y}_9} = \sqrt{714{,}3} = 26{,}73\,(ed.);$$

Здесь использовались значения, полученные по формулам:

$$\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{36}{8} = 4{,}5; \qquad \sum_{t=1}^{n}(t-\bar{t})^2 = \sum_{t=1}^{n} t^2 - \frac{(\sum_{t=1}^{8} t)^2}{n} = 204 - \frac{36^2}{8} = 42.$$

По таблице 1 (прилож. 4) определим значение $t(0{,}95;\,6) = 2{,}45$. Интервальная оценка прогноза среднего спроса определяется по известной формуле:

$412{,}4 - 2{,}45 \cdot 26{,}73 \le \overline{y}(9) \le 412{,}4 + 2{,}45 \cdot 26{,}73$, или $346{,}9 \le \overline{y}(9) \le 477{,}9$ (ед.).

Для нахождения интервальной оценки прогноза индивидуального значения $y^*(9)$ вычислим дисперсию его оценки:

$$s_{\widehat{y}_9}^2 = 1176{,}5 \cdot \left(1 + 1/8 + (9-4{,}5)^2/42\right) = 1890{,}8; \quad s_{\widehat{y}_9}^2 = 43{,}48 \text{ (ед.)},$$

а затем саму интервальную оценку для $y^*(9)$:

$412{,}4 - 2{,}45 \cdot 43{,}48 \le y^*(9) \le 412{,}4 + 2{,}45 \cdot 43{,}48$ или $305{,}9 \le y^*(9) \le 518{,}9$ (ед.).

Таким образом, с надёжностью 0,95 среднее значение спроса на товар на девятый год будет находиться в диапазоне от 346,9 до 477,9 (ед.), а его индивидуальное значение – от 305,9 до 518,9 (ед.).

Прогноз развития изучаемого процесса на основе экстраполяции временных рядов может оказаться эффективным, как правило, в рамках краткосрочного или, реже, среднесрочного периода прогнозирования.

Рассмотрим задачу прогнозирования в более общей постановке. Пусть задана информация о случайном процессе y(t), характеризующая прогнозируемый параметр во времени t, а именно $\Omega_n = \{y(t_1), y(t_2), ... y(t_n)\}$. Необходимо на основании изучения статистических свойств процесса y(t), реализованного в виде Ω_n, найти $\Omega_l = \{y(t_{n+1}), y(t_{n+2}), ... y(t_{n+l})\}$ такое, чтобы $\Omega_l \cap \Omega_n \neq 0$ и $\Omega_l \cup \Omega_n = \Omega$, где Ω - информация о всём процессе y(t) в настоящем и будущем.

Анализ известных подходов к решению задач прогнозирования, показал, что наиболее предпочтительными являются регрессионный и статистические методы прогнозирования. Для решения данной задачи используют известные в области математической статистики методы взвешенных отклонений, гармонических весов и экспоненциального сглаживания.

Решение контрольной задачи, заключавшейся в прогнозировании характеристик аппаратуры управления, показало что ошибки прогноза оказались минимальными при реализации алгоритма прогнозирования по методу экспоненциального сглаживания. Вычислительная схема разработанного алгоритма (*алгоритм 1-а*) для случая M=1 состоит в следующем.

Если временной ряд аппроксимируется линейным уравнением регрессии $y_t = a_0 + a_1 \cdot t + \varepsilon_t$, то реализуются следующие шаги алгоритма.

ШАГ 1. Вычисляются оценки параметров a_0 и a_1 по методу наименьших квадратов (МНК):

$$\widehat{a}_0 = \frac{\sum_{t=1}^{n} y_t \sum_{t=1}^{n} t^2 - \sum_{t=1}^{n} y_t \sum_{t=1}^{n} t}{n\sum_{t=1}^{n} t^2 - (\sum_{t=1}^{n} t)^2}; \quad \widehat{a}_1 = \frac{n\sum_{t=1}^{n} y_t \cdot t - \sum_{t=1}^{n} y_t \sum_{t=1}^{n} t}{n\sum_{t=1}^{n} t^2 - (\sum_{t=1}^{n} t)^2} \quad .$$

ШАГ 2. Для заданного параметра сглаживания вычисляются начальные условия:

$$S_1^{[1]}(y) = \widehat{a}_0 - \frac{1-\alpha}{\alpha} \cdot \widehat{a}_1 \ ; \quad S_1^{[2]}(y) = \widehat{a}_0 - \frac{2 \cdot (1-\alpha)}{\alpha} \cdot \widehat{a}_1 \, .$$

При выборе параметра сглаживания α придерживаются рекомендаций:

$1 \leq \alpha \geq 0{,}5$ - для менее чувствительных процессов при высоком уровне помех; $0{,}5 \leq \alpha \geq 0{,}1$ - для наиболее консервативных процессов.

ШАГ 3. На основании рекуррентных формул вычисляются экспоненциальные средние:

$$\widehat{y}_t = \widehat{a}_0 + \widehat{a}_1 \cdot t, \quad t = \overline{2, n+1};$$

$$S_t^{[1]}(y) = \alpha \cdot \widehat{y}_t + (1-\alpha) \cdot S_{t-1}^{[1]}(y), \quad t = \overline{2, n+1};$$

$$S_t^{[2]}(y) = \alpha \cdot S_t^{[1]}(y) + (1-\alpha) \cdot S_{t-1}^{[2]}(y)\,, \quad t = \overline{2, n+1};$$

ШАГ 4. Вычисляются оценки коэффициентов по формулам:

$$\widehat{a}_0(t) = 2 \cdot S_t^{[1]}(y) - S_t^{[2]}(y), \quad t = \overline{2, n+1};$$

$$\widehat{a}_1(t) = \frac{\alpha}{1-\alpha}\{S_t^{[1]}(y) - S_t^{[2]}(y)\}, \quad t = \overline{2, n+1}\ .$$

ШАГ 5. Вычисляются аппроксимированные значения параметра:

$$y_t{*} = \widehat{a}_0(t) + \widehat{a}_1(t) \cdot t, \quad t = \overline{2, n+1}\ .$$

ШАГ 6. Находится абсолютное отклонение ε_t': $\quad y_t - \widehat{y}_t{*},\ t = \overline{2, n}$.

ШАГ 7. Рассчитывается величина среднего квадратичного отклонения σ_ε:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{t=2}^{n} \varepsilon_t^{\,2}}{n-1}}\ .$$

ШАГ 8. Для временного ряда вычисляются значения прогноза $y_{t+l}{*}$:

$$y_{t+l}{*} = \widehat{a}_0(n+1) + \widehat{a}_1(n+1) \cdot l,\ l = \overline{1, k}\ .$$

ШАГ 8. Определяется ошибка прогноза:

$$\sigma_{y(t+l)} = \sigma_\varepsilon \cdot \sqrt{\frac{\alpha}{(2-\alpha)^3} \cdot [1 + 4(1-\alpha) + 5(1-\alpha)^2 + 2\alpha(4-3\alpha)l + 2\alpha^2 l^2]}\,, \quad l = \overline{1, k}\ .$$

ШАГ 10. Устанавливаются верхние и нижние границы доверительного интервала для значений прогноза $y_{t+l}{*}$:

$$\sup_l\{y_{t+l}{*}\} = y_{t+l}{*} + \ \sigma_{y(t+l)}\ , \quad \inf_l\{y_{t+l}{*}\} = y_{t+l}{*} - \sigma_{y(t+l)}\ , \ l = \overline{1, k}\ .$$

Экспериментальная проверка результатов прогнозирования показала, что ошибка прогноза при использовании изложенного алгоритма на основе находится в пределах (8…19)%, т.е. является допустимой.

ЗАКЛЮЧЕНИЕ

Современная эконометрика представляет собой быстроразвивающуюся отрасль экономической науки, основная задача которой состоит в том, чтобы придать количественные меры экономическим отношениям. Иными словами, эконометрика изучает конкретные количественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей. Значение такого подхода для исследования свойств экономических процессов в условиях нестабильности экономического развития России очень велико.

Модели парной регрессии несмотря на введение достаточно жёстких допущений, обусловленных учётом только одного фактора, на практике могут использоваться для приближённого анализа связей и отношений. Важными достоинствами моделей указанной группы являются ясный физический смысл модели и удобство аналитического расчёта и физической интерпретации коэффициентов уравнения регрессии. При исследовании процессов с устойчивым трендом хорошие результаты получить в случае применения однофакторных параболических моделей ($n = \{2; 3\}$).

Модели множественной регрессии обладают гораздо большими разрешающей способностью и прогностическими возможностями. При наличии достаточной мощности выборки наблюдений можно осуществить выбор подходящей формы модели с учётом решаемой задачи анализа. Расчёт коэффициентов линейной многофакторной модели сводится к составлению (согласно МНК) и решению системы алгебраических уравнений. Это легко выполняется на основе вычисления определителей. Для упрощённого анализа количественных связей можно рекомендовать двухфакторные линейные и экспоненциальные модели.

В качестве перспективного подхода к решению эконометрических задач следует рассматривать применение временных рядов. Теория и методы анализа временных рядов к настоящему времени хорошо развиты. Для сглаживания и оценки корреляционных связей уровней ряда могут быть использованы эффективные алгоритмы и программные процедуры, реализованные в унифицированных пакетах прикладных программ.

Список использованной литературы:

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики.- М.: ЮНИТИ, 1998.
2. Айвазян С.А. Основы эконометрики.- М.: ЮНИТИ, 2001. -432 с.
3. Андерсон Т. Статистический анализ временных рядов.- М.: Мир, 1976.- 755 с.
4. Бородич С.А. Эконометрика. – Мн.: Новое Знание, 2001.- 408 с.
5. Бурумкулов Ф.Х., Мировская Е.А. Основы теории вероятностей и математической статистики.- М.: Издательство стандартов, 1981.-164 с.
6. Дьяконов В.П. Mathcad 2000: Учебный курс.-СПб.: Питер, 2001.-592 с.
7. Доугерти К. Введение в эконометрику.- М.: Инфра-М, 2001.-402 с.
8. Дубров А.М. Математико-статистическая оценка эффективности в экономических задачах.- М.: Финансы и статистика, 1982.-176 с.
9. Замков О.О., Черемных Ю.А., Толстопятенко А.В. Математические методы в экономике. Учебник. 2-е изд. Под общ. ред. А.В. Сидоровича.- М.: Изд-во «Дело и Сервис», 1999.- 368 с.
10. Информационные системы и технологии в экономике. Учебник.- 2-е ид., перераб. и доп / Под ред. В.И. Лойко.- М.: Финансы и статистика, 2003.- 416 с.
11. Кремер Н.Ш., Путко Б.А. Экономерика.- М.: ЮНИТИ, 2003.- 311 с.
12. Орехов Н.А. Математические методы и модели в экономике.- М.: ЮНИТИ-ДАНА, 2004.
13. Орлов А.И. Эконометрика.- М.: Экзамен, 2002.
14. Практикум по эконометрике. Учебное пособие. Под ред. И.И. Елисеевой.- М.: Финансы и статистика, 2003.- 192 с.
15. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика.- М.: Экзамен, 2003.- 512 с.
16. Тихонов А.Н., Уфимцев М.В. Статистическая обработка результатов экспериментов: Учебное пособие. – М.: Изд-во Московского ун-та, 1988. - 174 с.
17. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. / Под ред. Ф.Э. Фигурнова- М.: ИНФАМ , 1998.- 528 с.
18. Шапкин А.С., Мазаева Н.М. Математические методы и модели исследования операций: Учебник..- М.: Издательско-торговая корпорация «Дашков и К^0», 2004.- 400 с.
19. Эконометрика. Учебник. Под ред. И.И. Елисеевой.- М.: Финансы и статистика, 2003.- 344 с.
20. Экономико-математические методы и прикладные модели. Учебное пособие для вузов./ Под ред В.В.Федосеева.- М.: ЮНИТИ, 2000.- 391 с.
21. www.rts.ru – Официальный сайт Фондовой биржи «Российская Торговая Система»
22. www.hedging.ru – Сайт для Хедж менеджеров

23. www.finam.ru – Сайт инвестиционной компании Финнам
24. www.dowjones.com – Сайт финансовой группы Доу Джонес
25. www.goldmansachs.com – Сайт инвестиционной компании Голдман Сахс.

ПРИЛОЖЕНИЕ 1. ГЛОССАРИЙ

№ п/п	Новые понятия	Содержание
1	2	3
1	**Эконометрика**	Часть экономической науки, занимающаяся разработкой и применением математических моделей и экономико-статистических методов анализа экономических процессов, а также применением методов обработки статистической экономической информации.
2	**Эконометрические методы**	Методы количественного описания и исследования экономических процессов с позиций экономической статистики.
3	**Выборка**	Некоторое количество наблюдений, отобранных из генеральной совокупности.
4	**Наблюдение**	Наблюдаемое значение случайной величины или набора случайных величин.
5	**Оценка, способ оценивания**	Общее правило (формула) для получения приближённого численного значения какого-либо параметра по данным выборки.
6	**Значение оценки**	Число, полученное в результате применения оценки к конкретной выборке.
7	**Смещение оценки**	Разность между МОЖ оценки и истинным значением оцениваемого параметра.
8	**Несмещённая оценка**	Оценка, имеющая нулевое смещение.
9	**Эффективная оценка**	Несмещённая оценка, имеющая наименьшую дисперсию среди всех несмещённых оценок.
10	**Эксперимент по методу Монте-Карло**	Искусственный, контролируемый (обычно вычислительный) эксперимент, проводимый для проверки и сравнения эффективности различных статистических методов.
11	**Состоятельная оценка**	Оценка, у которой смещение и дисперсия стремятся к нулю при увеличении объёма выборки.
12	**Математическая модель**	1. Количественное отображение на принятом формальном языке в рамках заданных допущений наиболее существенных свойств и отношений экономической системы. 2. Совокупность переменных и связей между ними в форме уравнений, описывающих зависимость между наблюдаемыми переменными.
13	**Модель парной регрессии**	Простейшая линейная модель зависимости между двумя переменными: $\mathbf{y=a+b\cdot x+\varepsilon}$ **.**

1	2	3
14	**Зависимая переменная регрессии**	Переменная величина в модели парной регрессии, которую считают (по экономическим соображениям) зависящей от другой переменной (например, в модели **$y=a+b\cdot x+\varepsilon$** это зависимая переменная y).
15	**Объясняющая переменная регрессии (регрессор)**	Переменная величина в модели парной регрессии, от которой зависит (по экономическим соображениям) зависимая переменная (например, в модели $y=a+b\cdot x+\varepsilon$ объясняющая переменная – это x).
16	**Случайный член регрессии**	Слагаемое в модели $y=a+b\cdot x+\varepsilon$, которое описывает воздействие неконтролируемых случайных факторов.
17	**Уравнение линейной регрессии**	Уравнение вида $y=a+b\cdot x$, где a и b – оценки параметров a и b, полученные в результате оценивания модели регрессии $y_i=a+b\cdot x+\varepsilon$ по данным выборки.
18	**Остаток в наблюдении**	Разность $q_i=y_i-(a+b\cdot x_i)$ между истинным значением переменной в i-м наблюдении (y_i) и значением ($a+b\cdot x_i$) в i-м наблюдении, полученным подстановкой наблюдения x_i в уравнение линейной регрессии.
19	**Метод наименьших квадратов (МНК)**	Метод нахождения оценок параметров регрессии, основанный на минимизации суммы квадратов остатков всех наблюдений $\sum q_i^2$.
20	**Объяснённая дисперсия зависимой переменной**	Выборочная дисперсия расчётных значений величины y: $Var(a+b\cdot x)$.
21	**Общая сумма квадратов отклонений (TSS - Total Sum of Squares)**	Сумма квадратов отклонений величины **y** от своего выборочного среднего $\bar{y}$.
22	**Объяснённая сумма квадратов отклонений (ESS - Explained Sum of Squares)**	Сумма квадратов отклонений величины $a+b\cdot x$ от своего выборочного среднего $a+b\cdot \bar{x}$.
23	**Необъяснённая (остаточная) сумма квадратов отклонений (USS - Unexplained Sum of Squares)**	Сумма квадратов остатков всех наблюдений.

№ п/п	Новые понятия	Содержание
1	2	3
24	**Коэффициент детерминации R^2**	Доля объяснённой дисперсии зависимой переменной во всей выборочной дисперсии y: $R^2=\frac{Var(a+b\cdot x)}{Var(y)}=\frac{ESS}{TSS}$.
25	**Стандартное отклонение случайной величины**	Корень квадратный из теоретической дисперсии случайной величины; среднее ожидаемое расстояние между наблюдениями этой случайной величины и её математическим ожиданием.
26	**Стандартная ошибка случайной величины**	Оценка стандартного отклонения случайной величины, полученная по данным выборки.
29	**Модель множественной регрессии**	Линейная модель зависимости между переменными: $\mathbf{Y=a+b_1x_1+b_2x_2+\ldots+ b_nx_n} +\varepsilon$, содержащая более двух переменных
30	**Доверительный интервал**	такой интервал, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значение прогнозируемого показателя.
31	**Динамический ряд**	Совокупность наблюдений некоторого явления (показателя), упорядоченная в зависимости от последовательности значений другого явления (признака).
32	**Нестрогая линейная зависимость между переменными**	Ситуация, когда теоретическая корреляция двух переменных близка к 1 или -1
33	**Строгая линейная зависимость между переменными**	Ситуация, когда выборочная корреляция двух переменных близка к 1 или -1
34	**Мультиколлинеарность**	Явление, когда нестрогая линейная зависимость между объясняющими переменными в модели множественной регрессии приводит к получению ненадежных оценок регрессии
35	**Полная коллинеарность**	Явление, когда нестрогая линейная зависимость между переменными приводит к невозможности применения МНК
36	**Лишняя переменная**	Объясняющая переменная, включенная в модель множественной регрессии, в то время, как по экономическим причинам ее присутствие в модели не нужно
37	**Отсутствующая переменная**	Необходимая по экономическим причинам объясняющая переменная, отсутствующая в модели
38	**Спецификация переменных**	Выбор необходимых для регрессии переменных и отбрасывание лишних переменных

39	**Замещающая переменная**	Объясняющая переменная, используемая в регрессии вместо трудноизмеримой, но важной переменной
40	**Лаговая переменная**	Наблюдение зависимой переменной в предшествующий момент, используемое как объясняющая переменная
41	**Фиктивная переменная**	Переменная, принимающая в каждом наблюдении только два значения: 1- «да» и 0 – «нет».
42	**Категория**	Событие, про которое для каждого наблюдения можно определённо сказать – произошло оно в этом наблюдении или нет.
43	**Набор категорий**	Конечный набор взаимоисключающих событий, полностью исчерпывающий все возможности.
44	**Совокупность фиктивных переменных**	Некоторое количество фиктивных переменных, предназначенное для описания набора категорий.
45	**Эталонная категория**	Категория, с которой сравниваются другие категории.
46	**Сезонные фиктивные переменные**	Совокупность фиктивных переменных, предназначенная для обозначения различных лет, времён года, месяцев и т.д.

ПРИЛОЖЕНИЕ 2. Биографические данные учёных в области математической статистики и эконометрики

Гаусс Карл Фридрих (30.04.1777-23.02.1855) – немецкий математик, астроном, физик и геодезист. С раннего детства обнаружил выдающиеся математические способности. В 1799 г. защитил докторскую диссертацию, содержащей первое доказательство основной теоремы алгебры. В разносторонней научной деятельности Гаусса органически сочетались исследования по теоретической и прикладной математике. Работы К.Гаусса оказали большое влияние на дальнейшее развитие высшей алгебры, теории чисел, дифференциальной геометрии, классической теории электричества и магнетизма. Именем Гаусса назван закон распределения вероятностей.

Канторович Леонид Витальевич (19.01.1912 - 7.04.1986) – советский математик, один из основателей отечественных школ функционального анализа, вычислительной математики, языков программирования.. Родился в Санкт-Петербурге. В 1930 г. закончил математический факультет Ленинградского ун-та. В 1935 г. Канторовичу Л.В. Присуждена степень доктора физико-математических наук без защиты диссертации. Крупнейшим открытием К. является введение в математическую и экономическую науки понятия «линейное программирование» (1939). Основная заслуга К. заключается в разработке единого подхода к широкому кругу экономических задач о наилучшем использовании ресурсов на базе линейного программирования. В 1975 г. Канторович Л.В. удостоен Нобелевской премии по экономике (совместно с американским учёным Т. Купмансом) за работы по теории оптимизации.

Кетле Адольф Ламберт Жак (22.02.1796-17.02.1874) – бельгийский астроном и математик, основатель математической статистики. Родился в Генте. В 17 лет стал преподавателем математики, грамматики и рисования, в 23 года получил степень доктора математики, а 24 (1820) стал членом Брюссельской АН. Кетле слушал теорию вероятностей у П.Лапласа. В 1825-1839 гг. руководил изданием первого математического и физического журнала в Бельгии. По статистике опубликовал 65 работ. Основной работой являются «Письма по теории вероятностей» (1846 г.).

Леонтьев Василий Васильевич (5.08.1906 - 5.02.1999) - американский экономист. Родился в Санкт-Петербурге. Предложенная им алгебраическая теория анализа «затраты-выпуск» сводилась к системе линейных уравнений, в которых параметрами были коэффициенты затрат на производство продукции. Леонтьев показал, что коэффициенты, выражающие отношения между секторами экономики (коэффициенты текущих материальных затрат) могут быть оценены статистически. При этом они достаточно устойчивы и их можно прогнозировать.. Реалистическая гипотеза и относительная простота измерений определили большие аналитические и прогностические возможности метода «затраты-выпуск». В 1973 г. В.Леонтьев удостоен премии А.Нобеля по экономике «За развитие метода «затраты-выпуск» и его применение к важным экономическим проблемам.

Пирсон Карл (27.03.1857-27.04.36) – английский математик-статистик, биолог и философ-позитивист. Родился в Лондоне. Окончил колледж Кембриджского ун-та (1879) со степенью магистра. С 1884 г. до конца жизни работал профессором прикладной математики и механики в Лондонском ун-те. Основные работы К.Пирсона посвящены: важным вопросам статистики, разработке теории корреляции и её применению к проблемам наследственности в эволюции; разработке тестов математической статистики и критериев согласованности опытных результатов и статистических гипотез; внедрению системы кривых частоты как способа описания явлений; применению метода моментов.

Тинберген Ян (1903-1950) – нидерландский экономист. Основные научные труды посвящены разработке математических методов анализа экономических процессов, моделирования экономического развития. Лауреат Нобелевской премии (1960 г.). Один из создателей теории конвергенции.

Фишер Рональд Айлмер (17.02.1890-29.07.1962) – английский биолог, математик и статистик. Родился в Лондоне. Окончид Кембриджский ун-т. В 1836-1957 гг. работал в Кембридже. В 1957 г. переехал в Австралию. Математические труды относятся к теории вероятностей и математической статистике. В 1912 г. разработал метод максимального правдоподобия. Многие понятия и утверждения в математической статистике связаны с именем К.Фишера.

Фриш Рагнар (1895-1973) – норвежский экономист. Основные труды посвящены разработке экономико-математических методов исследования экономического роста, проблемам моделирования экономических процессов.

ПРИЛОЖЕНИЕ 3.

1. Построение моделей линейной парной регрессии

Программа Р-01а.

Необходимо определить коэффициенты линейной модели парной регрессии для условий задачи 1

$\mathrm{ORIGIN} := \mathbf{1}$

$$PY := \begin{pmatrix} \mathbf{68.8} \\ \mathbf{61.2} \\ \mathbf{59.9} \\ \mathbf{56.7} \\ \mathbf{55.0} \\ \mathbf{54.3} \\ \mathbf{49.3} \end{pmatrix} \qquad PX := \begin{pmatrix} \mathbf{45.1} \\ \mathbf{59.0} \\ \mathbf{57.2} \\ \mathbf{61.8} \\ \mathbf{58.8} \\ \mathbf{47.2} \\ \mathbf{55.2} \end{pmatrix} \qquad \boxed{N := 7}$$

Число наблюдений

$\boxed{1}$ $MX := \mathrm{mean}(PX)$ $\boxed{MX = \mathbf{54.9}}$

$\boxed{2}$ $MY := \mathrm{mean}(PY)$ $\boxed{MY = \mathbf{57.886}}$

$\boxed{3}$ $$\boxed{C1XY := \frac{\sum_{i=1}^{N} (PX_i \cdot PY_i)}{N}} \qquad C1XY = \mathbf{3166.049}$$

$\boxed{4}$ $C2XY := MX \cdot MY$ $C2XY = \mathbf{3177.926}$

$\boxed{5}$ $j := \mathbf{1}..N$ $X2_j := PX_j \cdot PX_j$ $CX1 := \mathrm{mean}(X2)$ $\boxed{CX1 = \mathbf{3048.344}}$

$\boxed{6}$ $CX2 := MX \cdot MX$ $CX2 = \mathbf{3014.01}$

$\boxed{7}$ $$b := \frac{C1XY - C2XY}{CX1 - CX2} \qquad \boxed{b = \mathbf{-0.3459}}$$

$\boxed{8}$ $a := MY - b \cdot MX$ $\boxed{a = \mathbf{76.877}}$

$\boxed{9}$ $Xmin := \min(PX)$ $Xmin = \mathbf{45.1}$ $dX := \dfrac{\max(PX) - \min(PX)}{N}$

$dX = \mathbf{2.386}$ $i := \mathbf{1}..N$ $PX2_i := Xmin + i \cdot dX$ $k := \mathbf{1}..N$ $\boxed{YS2_k := a + b \cdot PX2_k}$

$$PX2^T = (47.486 \quad 49.871 \quad 52.257 \quad 54.643 \quad 57.029 \quad 59.414 \quad 61.8)$$

$$YS2^T = (60.451 \quad 59.625 \quad 58.8 \quad 57.975 \quad 57.149 \quad 56.324 \quad 55.499)$$

$$i := 1 .. 7$$

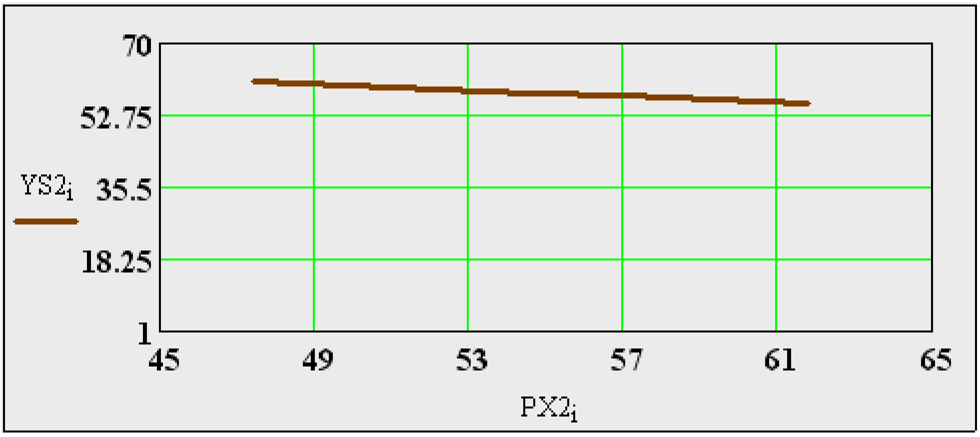


Рисунок 1 – Результаты идентификации регрессионной модели y*=f(X)

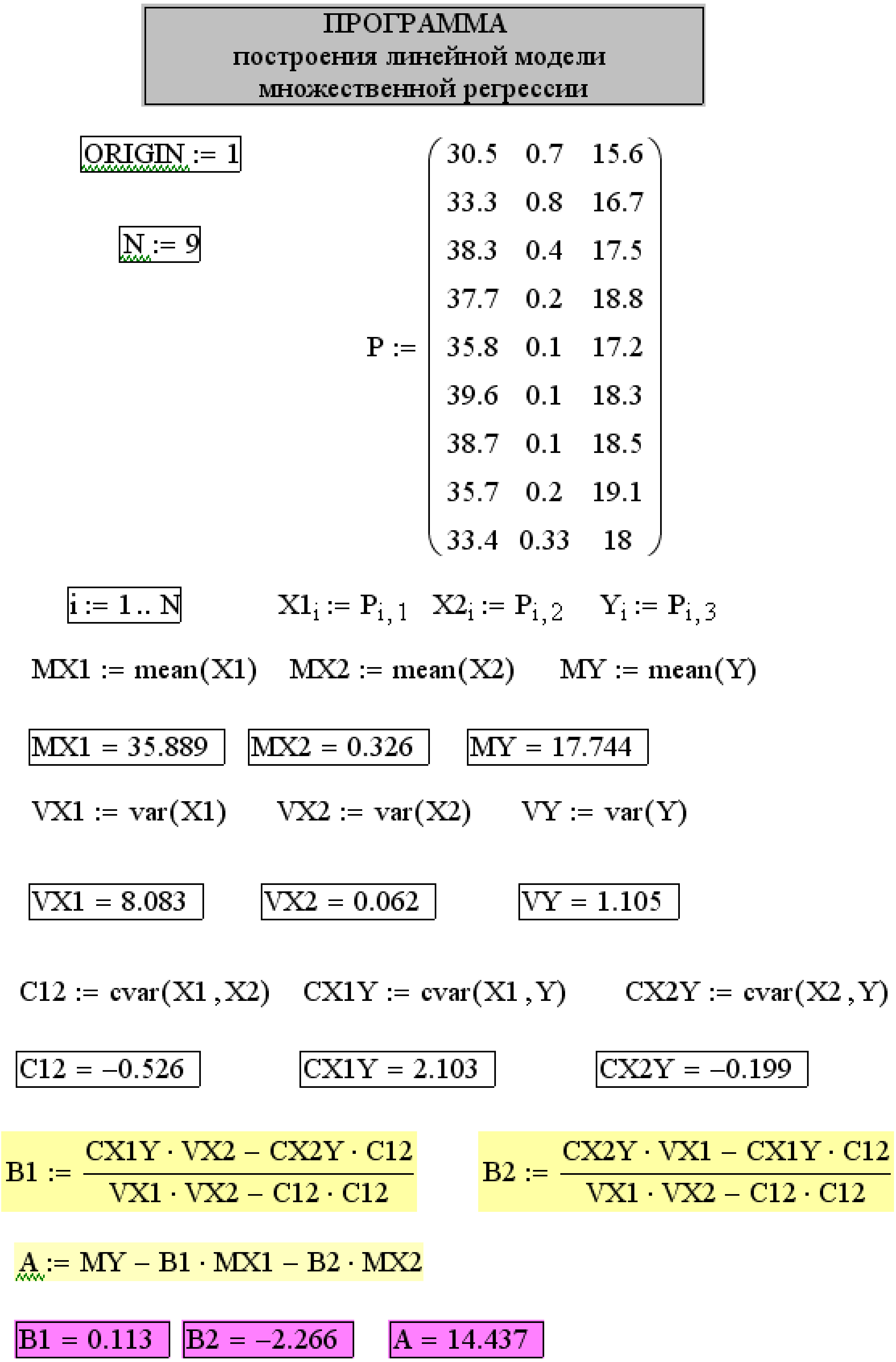
10 Вычисление линейного коэффициента парной корреляции и коэффициента детерминации:

$$Rxy := b \cdot \frac{\sqrt{var(PX)}}{\sqrt{var(PY)}} \qquad Rxy = -0.353 \qquad Rxy \cdot Rxy = 0.125$$

$$k := 1 .. N \qquad YS3_k := a + b \cdot PX_k \qquad var(PX) = 34.334$$

$$PX^T = (45.1 \quad 59 \quad 57.2 \quad 61.8 \quad 58.8 \quad 47.2 \quad 55.2)$$

$$YS3^T = (61.276 \quad 56.467 \quad 57.09 \quad 55.499 \quad 56.537 \quad 60.549 \quad 57.782)$$



ПРОГРАММА
построения линейной модели
множественной регрессии

$\boxed{ORIGIN := 1}$

$\boxed{N := 9}$

$$P := \begin{pmatrix} 30.5 & 0.7 & 15.6 \\ 33.3 & 0.8 & 16.7 \\ 38.3 & 0.4 & 17.5 \\ 37.7 & 0.2 & 18.8 \\ 35.8 & 0.1 & 17.2 \\ 39.6 & 0.1 & 18.3 \\ 38.7 & 0.1 & 18.5 \\ 35.7 & 0.2 & 19.1 \\ 33.4 & 0.33 & 18 \end{pmatrix}$$

$\boxed{i := 1..N}$ $\quad X1_i := P_{i,1} \quad X2_i := P_{i,2} \quad Y_i := P_{i,3}$

$MX1 := mean(X1) \quad MX2 := mean(X2) \quad MY := mean(Y)$

$\boxed{MX1 = 35.889} \quad \boxed{MX2 = 0.326} \quad \boxed{MY = 17.744}$

$VX1 := var(X1) \quad VX2 := var(X2) \quad VY := var(Y)$

$\boxed{VX1 = 8.083} \quad \boxed{VX2 = 0.062} \quad \boxed{VY = 1.105}$

$C12 := cvar(X1, X2) \quad CX1Y := cvar(X1, Y) \quad CX2Y := cvar(X2, Y)$

$\boxed{C12 = -0.526} \quad \boxed{CX1Y = 2.103} \quad \boxed{CX2Y = -0.199}$

$$B1 := \frac{CX1Y \cdot VX2 - CX2Y \cdot C12}{VX1 \cdot VX2 - C12 \cdot C12} \qquad B2 := \frac{CX2Y \cdot VX1 - CX1Y \cdot C12}{VX1 \cdot VX2 - C12 \cdot C12}$$

$A := MY - B1 \cdot MX1 - B2 \cdot MX2$

$\boxed{B1 = 0.113} \quad \boxed{B2 = -2.266} \quad \boxed{A = 14.437}$

$k := 1 .. 9$ $\qquad YZ_k := A + B1 \cdot X1_k + B2 \cdot X2_k$

$$YZ^T = (16.29 \quad 16.38 \quad 17.85 \quad 18.23 \quad 18.25 \quad 18.67 \quad 18.57 \quad 18.01 \quad 17.45)$$

$MYZ := \text{mean}(YZ)$

Методика оценки показателей качества регрессии

1. $TSS := \sum_i (Y_i - MY)^2$ $\qquad TSS = 9.942$

2. $RSS := \sum_i (Y_i - YZ_i)^2$ $\qquad RSS = 3.75$

3. $ESS := \sum_i (YZ_i - MYZ)^2$ $\qquad ESS = 6.192$

4. $R2 := \dfrac{ESS}{TSS}$ $\qquad R2 = 0.623$

5. $Rx1x2 := \dfrac{C12}{\sqrt{VX1 \cdot VX2}}$ $\qquad Rx1x2 = -0.745$

6. $N := 9$ $\qquad m := 2$ $\qquad \boxed{l := N - m - 1}$ $\qquad \boxed{l = 6}$

7. $S2e := \dfrac{RSS}{l}$ $\qquad S2e = 0.625$

8. $Sb1 := \sqrt{\dfrac{S2e}{(N \cdot VX1) \cdot (1 - Rx1x2)}}$ $\qquad \boxed{Sb1 = 0.07}$

$Sb2 := \sqrt{\dfrac{S2e}{(N \cdot VX2) \cdot (1 - Rx1x2)}}$ $\qquad \boxed{Sb2 = 0.803}$

Программа P011
Оценка коэффициентов автокорреляции для временного ряда

$ORIGIN := 1 \qquad N := 8$

Исходный временной ряд

$$YYt := \begin{pmatrix} 1 & 7 \\ 2 & 8 \\ 3 & 8 \\ 4 & 10 \\ 5 & 11 \\ 6 & 12 \\ 7 & 14 \\ 8 & 16 \end{pmatrix}$$

$j := 1 .. N \qquad YY_j := YYt_{j,2} \qquad YY^T = \begin{pmatrix} 7 & 8 & 8 & 10 & 11 & 12 & 14 & 16 \end{pmatrix}$

$j := 1 .. N - 1 \qquad YY2_{j+1} := YYt_{j,2} \qquad YY2^T = \begin{pmatrix} 0 & 7 & 8 & 8 & 10 & 11 & 12 & 14 \end{pmatrix}$

$$My1 := \frac{\sum_{t=2}^{N} YY_t}{N-1} \qquad My2 := \frac{\sum_{t=2}^{N} YY2_t}{N-1} \qquad My1 = 11.286 \qquad My2 = 10$$

$$\boxed{R1 := \frac{\sum_{t=2}^{N} \left[\left(YY_t - My1 \right) \cdot \left(YY2_t - My2 \right) \right]}{\sqrt{\sum_{t=2}^{N} \left(YY_t - My1 \right)^2 \cdot \sum_{t=2}^{N} \left(YY2_t - My2 \right)^2}}} \qquad \boxed{R1 = 0.977}$$

$j := 1 .. N - 2 \qquad YY3_{j+2} := YYt_{j,2} \qquad YY3^T = \begin{pmatrix} 0 & 0 & 7 & 8 & 8 & 10 & 11 & 12 \end{pmatrix}$

$$My3 := \frac{\sum_{t=3}^{N} YY_t}{N-2} \qquad My4 := \frac{\sum_{t=3}^{N} YY3_t}{N-2}$$

$$\boxed{R2 := \frac{\sum_{t=3}^{N} \left[\left(YY_t - My3 \right) \cdot \left(YY3_t - My4 \right) \right]}{\sqrt{\sum_{t=3}^{N} \left(YY_t - My3 \right)^2 \cdot \sum_{t=3}^{N} \left(YY3_t - My4 \right)^2}}} \qquad \boxed{R2 = 0.973}$$

ПРИЛОЖЕНИЕ 4. Статистические таблицы.

Таблица 1. Значения *t*-критерия Стьюдента при фиксированном уровне значимости $\alpha = \{0{,}10;\ 0{,}05;\ 0{,}02;\ 0{,}01\}$ (двухсторонний)

Число степеней свободы *df*	Доверительная вероятность α			
	0,1	0,05	0,02	0,01
1	6,3138	12,706	31,821	63,657
2	2,9200	4,3027	6,965	9,9248
3	2,3534	3,1825	4,541	5,8409
4	2,1318	2,7764	3,747	4,6041
5	2,0150	2,5706	3,365	4,0321
6	1,9432	2,4469	3,143	3,7074
7	1,8946	2,3646	2,998	3,4995
8	1,8595	2,3060	2,896	3,3554
9	1,8331	2,2622	2,821	3,2498
10	1,8125	2,2281	2,764	3,1693
11	1,7959	2,2010	2,718	3,1058
12	1,7823	2,1788	2,681	3,0545
13	1,7709	2,1604	2,650	3,0123
14	1,7613	2,1448	2,624	2,9768
15	1,7530	2,1315	2,602	2,9467
16	1,7459	2,1199	2,583	2,9208
17	1,7396	2,1098	2,567	2,8982
18	1,7341	2,1009	2,552	2,8784
19	1,7291	2,0930	2,539	2,8609
20	1,7247	2,0860	2,528	2,8453

Таблица 2. Данные о критических уровнях при уровне значимости 0,05 для однопараметрической (m=1) и двухпараметрической (m=2) моделей.

Количество наблюдений	Первый коэффициент автокорреляции r(1)	Модель				Граница RS-критерия	
		m=1		m=2		нижняя	верхняя
		d_1	d_2	d_1	d_2		
10	0,360	-	-	-	-	2,67	3,69
15	0,328	1,08	1,36	0,95	1,54	2,96	4,14
20	0,300	1,20	1,41	1,10	1,54	3,18	4,49
25	0,276	1,28	1,45	1,20	1,55	3,34	4,71
30	0,257	1,35	1,49	1,28	1,57	3,47	4,89