

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПУТЕЙ СООБЩЕНИЯ»**

КАФЕДРА «МАТЕМАТИКА»

М.В. Ишханян

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Учебное пособие

Москва – 2015

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПУТЕЙ СООБЩЕНИЯ»**

КАФЕДРА «МАТЕМАТИКА»

М.В. Ишханян

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Рекомендовано редакционно-издательским советом
университета в качестве учебного пособия
для студентов направления «Строительство» магистерской
программы «Управление стоимостью и девелопмент в ин-
вестиционно-строительном комплексе»

Москва – 2015

УДК – 51
И – 97

Ишханян М.В. Математическое моделирование: Учебное пособие. – М.: МГУПС (МИИТ), 2015. – 150 с.

Учебное пособие предназначено для студентов направления «Строительство» магистерской программы «Управление стоимостью и девелопмент в инвестиционно-строительном комплексе», обучающихся по дисциплине «Математическое моделирование». Содержит основные понятия, определения, положения и подходы математического моделирования, представлена классификация математических моделей. Изложены основные методы и модели в области математического моделирования экономических задач в строительстве, включающих в себя модели прогнозирования на основе временных рядов, эконометрические однофакторные и многофакторные модели. Теоретический материал сопровождается решением типовых практических задач. Каждый из разделов снабжен перечнем заданий для самостоятельной работы. Может быть использовано для проведения практических занятий, а также для организации самостоятельной работы студентов.

Рецензенты:

В.Н. Деснянский, к.ф.-м.н., зав. кафедрой «Математический анализ» МИИТа;
Л.А. Климина, к.ф.-м.н., с.н.с. НИИ механики МГУ имени М.В. Ломоносова.

© МГУПС (МИИТ), 2015

Оглавление

§ 1. Методология экономико-математического моделирования	6
Введение	6
Понятие модели.....	10
Классификация моделей.....	11
Классификация математических моделей экономических систем	15
Классификация математических методов	21
§2. Основы эконометрики.....	23
Понятие эконометрики: предмет, цель, задачи.....	23
Эконометрическая модель	25
Основные этапы эконометрического моделирования.....	28
Типы данных в эконометрическом моделировании	28
Переменные в эконометрическом моделировании	29
§3. Методы и модели корреляционно-регрессионного анализа	31
Понятие корреляционной зависимости	31
Понятие корреляционного анализа. Понятие корреляции	33
Парная корреляция	34
Свойства коэффициента парной корреляции	35
Оценка статистической значимости коэффициента парной корреляции	36
Множественная корреляция.....	37
Проверка статистической значимости коэффициента множественной корреляции.....	38
Коэффициент детерминации.....	39
Проверка статистической значимости коэффициента детерминации R²	40

Понятие регрессионного анализа.....	40
Парная регрессия	43
Требования теоремы Гаусса-Маркова (предпосылки МНК).....	45
Оценка статистической значимости коэффициентов уравнения парной регрессии.....	46
Оценка качества уравнения линейной регрессии	47
Множественная регрессия	49
Требования теоремы Гаусса-Маркова (предпосылки МНК).....	50
Оценка качества уравнения множественной линейной регрессии.....	51
Проверка выполнения предположений регрессионного анализа по остаткам.....	54
Отбор факторных признаков в модель	57
Практические приложения регрессионной модели	61
Эластичность	61
Прогнозирование.....	62
§4. Решение задач корреляционно-регрессионного анализа	64
Пакет «Анализ данных»	64
Ссылки (или имена) как аргументы функций и формул	64
Вставка функций	65
Использование надстройки «Пакет анализа»	65
Форматы чисел в MS Excel	66
Копирование листов Microsoft Excel.....	69
Создание парных диаграмм рассеяния	70
Регрессионная статистика в отчете Excel	72
Дисперсионный анализ в отчете Excel.....	73
Решение типовых задач.....	75
Задачи для самостоятельного решения.....	103

§5. Методы и модели прогнозирования временных рядов экономических показателей	107
Понятие ряда динамики. Понятие временного ряда	107
Проверка гипотезы о наличии тренда.....	109
Модель временного ряда.....	110
Порядок анализа модели	111
Сглаживание ряда динамики	111
Численное сглаживание.....	111
Аналитическое сглаживание ряда динамики .	114
Исследование структуры ряда. Автокорреляция уровней ряда	114
Моделирование сезонных и циклических колебаний	117
Прогнозирование временных рядов.....	118
Мультипликативная модель ряда	119
Аддитивная модель ряда	119
§6. Анализ и прогнозирование временных рядов в Excel.....	120
Анализ временных рядов с помощью инструмента " Мастер диаграмм"	120
Решение типовых задач.....	126
Задачи для самостоятельного решения.....	139
Приложения.....	141
Приложение 1	141
Приложение 2	142
Приложение 3.....	143
Приложение 4.....	145
Приложение 5.....	147
Приложение 6.....	149
Литература.....	150

§ 1. Методология экономико-математического моделирования

Введение

В течение всей своей жизни человек вынужден принимать те или иные решения: какой подарок попросить на день рождения, куда пойти учиться, как лучше потратить лишние (или последние) деньги и т. д.

Если определенная ситуация, требующая принятия решения, повторяется достаточно часто, то решение приходит «само собой», автоматически. Если же ситуация недостаточно знакома или человек не располагает всей необходимой информацией, то принятие решения существенно усложняется. В таких случаях он вынужден, как правило, сравнивать между собой несколько возможных вариантов и выбирать тот, который кажется ему наиболее предпочтительным (или наименее опасным).

Значительно более важные последствия имеют так называемые *управляющие решения*. Из теории исследования операций известно, что количественно принятое решение можно оценить с помощью *показателя эффективности (ПЭ)* функционирования исследуемой (разрабатываемой) системы, в том числе экономической.

В общем случае он отражает *результат функционирования системы*, являющийся функцией трех факторов:

$$Y_{on} = Y(q(u), c(u), t(u))$$

q – полезный эффект операции;

c – затраты ресурсов на проведение операции;

t – затраты времени на ее проведение.

u – стратегия проведения операции u

ПЭ позволяет оценить (точнее, описать) результат операции, полученный при использовании конкретной стратегии. Однако даже если такие оценки будут получены для всего множества допустимых стратегий, этого еще недостаточно, чтобы выбрать одну из них, ту, которая будет реализована.

Например, при оценке загруженности вычислительной сети оказалось, что коэффициент использования равен 0,7. Хорошо это или плохо?

Чтобы ответить на подобный вопрос, необходимо сформулировать правило, позволяющее *лицу, принимающему решение (ЛПР)*, сравнивать между собой стратегии, характеризующиеся различными значениями ПЭ. В одних случаях правило сравнения может быть очень простым, в других же его вообще не удастся найти и придется изменять (уточнять) показатель эффективности.

Скажем, если автомобиль одной и той же марки в двух разных автосалонах продается по разным ценам (при прочих равных условиях), то правило выбора салона напрашивается само. Совсем другое дело, когда автомобили различаются стоимостью, фирмой-изготовителем, дизайном, организацией гарантийного обслуживания и т.д. При этом первый оказывается предпочтительнее по одним показателям, второй - по другим. В такой ситуации покупатель должен сначала определить правило выбора и только после этого сравнивать между собой различные варианты.

В теории принятия решений правило, на основании которого производится выбор стратегии, отвечающей интересам ЛПР, называется *критерием эффективности*.

Таким образом, показатель эффективности и критерий эффективности в совокупности отражают цели, которые преследует ЛПР при проведении данной операции, а также наиболее предпочтительный для него способ достижения этой цели.

Лица, ответственные за принятие решений, касающихся проектирования и создания экономических систем, могут оценить их эффективность одним из трех следующих способов:

- проведение эксперимента с экономической системой;
- использование данных о поведении системы в прошлом;
- использование методов моделирования.

Теоретически всегда есть возможность проводить управляемые эксперименты с экономической системой фирмы, отрасли или страны. Однако принятие неоптимальных решений может причинить ущерб экономической системе. При этом, чем больше масштаб системы, тем ощутимее убытки. Тем не менее, на практике такие эксперименты нередко производились и производятся (в некоторых странах) с неизменно отрицательным результатом.

Даже в случае оптимальных решений, касающихся, например, управления деятельностью фирмы, при проведении натурных экспериментов трудно сохранить постоянство факторов и условий, влияющих на результат, а следовательно, сложно обеспечить надежную оценку различных экономических решений.

Если есть данные о развитии экономической системы за некоторый период времени в прошлом, то можно провести мысленный эксперимент на этих данных. Однако для этого нужно знать точно, какие изменения и каких входных переменных привели к наблюдаемому изменению выходных переменных, характеризующих эффективность экономической системы. Иногда причинами изменений могут оказаться случайные возмущения, или так называемый «шум». Поэтому нельзя слишком доверять оценкам

экономических решений, полученным на основе данных о развитии системы в прошлом.

Всегда можно построить математическую модель рассматриваемой системы, связывающую входные (независимые) переменные с выходными (зависимыми) переменными, а также с экономической стратегией, т.е. со способом управления экономической системой. Если есть основания для того, чтобы считать разработанную математическую модель *адекватной* рассматриваемой экономической системе, то с помощью модели можно производить расчеты или машинные эксперименты (если модель реализована на ЭВМ). По результатам этих экспериментов можно выработать рекомендации по повышению эффективности существующей или проектируемой экономической системы

Условием для разработки модели является наличие так называемой информационной достаточности. Это означает, что разработчик должен иметь достаточное представление о том, что является входными и выходными переменными в исследуемой системе и какие факторы оказывают влияние на процесс ее функционирования. Если уровень информационной достаточности невысок, то создать модель, с помощью которой можно получать новые знания об объекте-оригинале, невозможно. Если же уровень информационной достаточности велик, т.е. система уже хорошо изучена, то вопрос о создании модели теряет смысл, так как новых знаний она также не даст.

Следовательно, *разрабатывать модель имеет смысл только в том случае, если объект-оригинал еще недостаточно изучен или вообще не существует в природе и только проектируется.*

Если объект-оригинал существует, то модель считается *адекватной* ему в том случае, если зависимость выходных переменных от входных параметров в модели и в

объекте-оригинале практически совпадает. При упрощении моделей степень адекватности снижается.

Если же объекта-оригинала еще не существует, то модель считается *адекватной* ему, если она с достаточной степенью приближения на уровне понимания моделируемого процесса исследователем отражает закономерности процесса функционирования реальной системы. Залогом адекватности в этом случае является полнота описания моделируемого процесса, т.е. учет всех факторов, поддающихся формализации.

Понятие модели

Существует множество различных типов моделей, особое место среди которых занимают математические модели.

Что же такое модель?

Модель представляет собой абстрактное описание системы (объекта, процесса, проблемы, понятия) в некоторой форме, отличной от формы их реального существования.

Моделирование представляет собой один из основных методов познания, является формой отражения действительности и заключается в выяснении или воспроизведении тех или иных свойств реальных объектов, предметов и явлений с помощью других объектов, процессов, явлений, либо с помощью абстрактного описания в виде изображения, плана, карты, совокупности уравнений, алгоритмов и программ.

Итак, в процессе моделирования всегда существует ***оригинал*** (объект) и ***модель***, которая воспроизводит (моделирует, описывает, имитирует) некоторые черты объек-

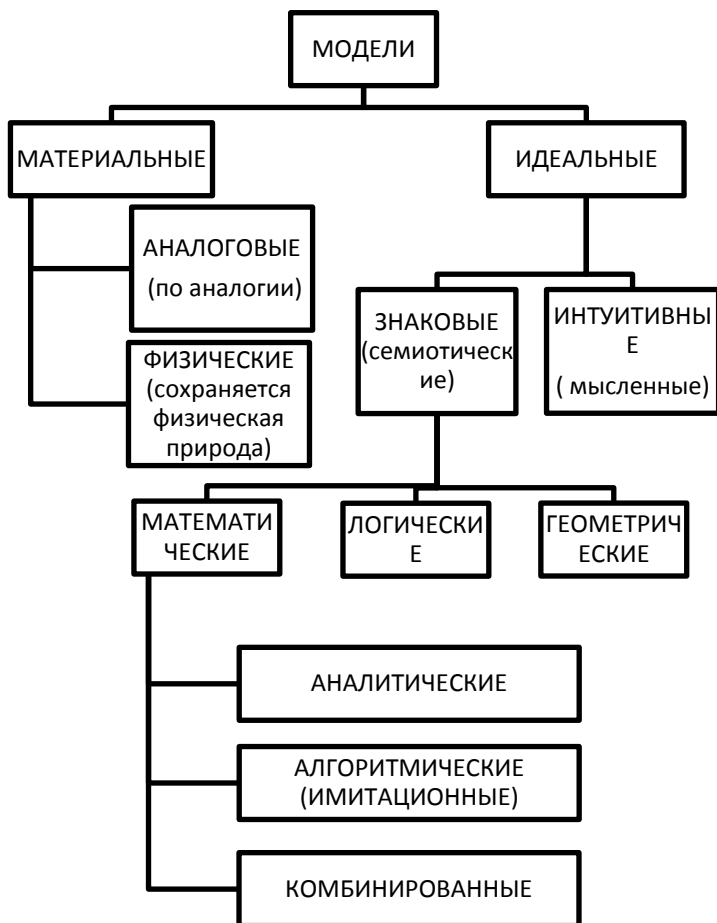
та, т.е. **модель** - объект любой природы, который создается исследователем с целью получения новых знаний об объекте- оригинале и отражает только существенные (с точки зрения разработчика) свойства оригинала.

Анализируя содержание этого определения, можно сделать следующие выводы:

- любая модель **субъективна**, она несет на себе печать индивидуальности исследователя;
- любая модель **гомоморфна**, т.е. в ней отражаются не все, а только существенные свойства объекта-оригинала;
- возможно существование **множества моделей** одного и того же объекта-оригинала, отличающихся целями исследования и степенью адекватности.
- Модель считается **адекватной** объекту-оригиналу, если она с достаточной степенью приближения на уровне понимания моделируемого процесса исследователем отражает закономерности процесса функционирования реальной системы во внешней среде.

Классификация моделей

Приведем общую классификацию основных видов моделей:



1. *Физическая модель* обеспечивает аналогию физической природы и модели (примером может служить аэродинамическая труба).
2. *Аналоговая модель* обеспечивает сходство процессов, протекающих в оригинале и модели (так, с помощью гидроинтегратора моделируется передача тепла).

3. *Интуитивная модель* используется для прогнозирования на основе анализа наблюдений прошлого периода: объема продаж, прибыли и денежного потока. При этом не предпринимаются попытки объяснить причинные взаимосвязи, которые лежат в основе модели.
4. *Логическая модель* - модель, в которой представлены различные варианты выбора действий на основе умозаключений и анализа условий.
5. *Геометрическая модель* - это графические формы и объемные конструкции. Например: рисунок, пиктограмма, чертеж, карта, план, объемное изображение и т. д.
6. Для *аналитического* моделирования характерно то, что для описания процессов функционирования системы используются системы алгебраических, дифференциальных, интегральных или конечно-разностных уравнений. Аналитическая модель может быть исследована следующими методами:
 - *аналитическим*, когда стремятся получить в общем виде явные зависимости для искомым характеристик;
 - *численным*, когда, не умея решать уравнения в общем виде, стремятся получить числовые результаты при конкретных начальных данных;
 - *качественным*, когда, не имея решения в явном виде, можно найти некоторые свойства решения (например, оценить устойчивость решения).

Желая использовать аналитический метод, часто идут на существенные упрощения первоначальной модели, чтобы иметь возможность изучить хотя бы общие свойства системы. Аналитические модели бывают *детерминированные* и *статистические*.

Численный метод проведения аналитических расчетов с помощью датчиков случайных чисел получил название метода статистических испытаний, или *метода Монте-Карло*.

7. При *алгоритмическом (имитационном) моделировании* описывается процесс функционирования системы во времени, причем имитируются элементарные явления, составляющие процесс, с сохранением их логической структуры и последовательности протекания во времени. Имитационные модели также могут быть детерминированными и статистическими. В последнем случае в модели с помощью датчиков случайных чисел имитируется действие неопределенных и случайных факторов. Такой метод моделирования получил название *метода статистического моделирования*. В настоящее время этот метод считается наиболее эффективным методом исследования сложных систем, а часто и единственным практически доступным методом получения информации о поведении гипотетической системы на этапе ее проектирования.
8. *Комбинированное моделирование* позволяет объединить достоинства аналитического и алгоритмического моделирования. При построении комбинированных моделей производится предварительная декомпозиция процесса функционирования модели на составляющие подпроцессы. Для тех из них, где это возможно, используются аналитические модели, а для остальных процессов строятся алгоритмические модели.

Классификация математических моделей экономических систем

Экономические процессы происходят в значительной мере стихийно, неуправляемо. Они плохо поддаются попыткам волевого управления со стороны политических, государственных и хозяйственных руководителей отдельных отраслей и экономики страны в целом. По этой причине экономические системы плохо поддаются изучению и формализованному описанию.

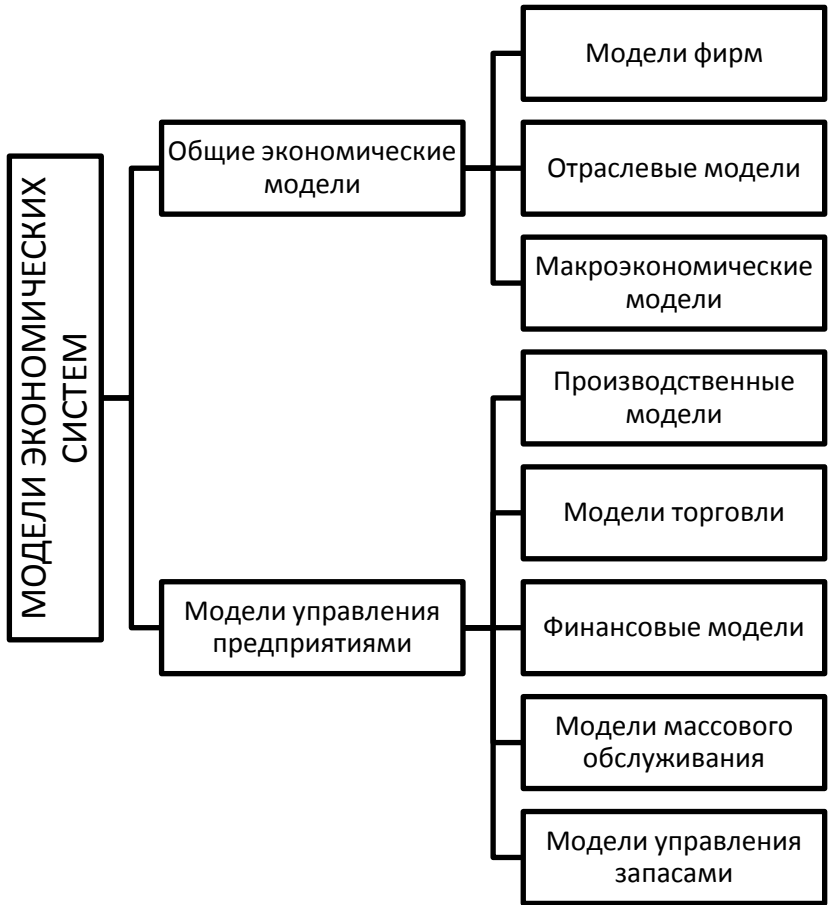
Существующие математические модели, которые принято называть моделями экономических систем, можно условно разделить на две группы: общие экономические модели и модели управления предприятиями.

Общие экономические модели

В основу классификации общих экономических моделей заложен масштаб изучаемой экономической системы. С этой точки зрения модели можно разделить на три большие группы: модели фирм, отраслевые модели и макроэкономические модели.

Разновидности моделей фирм:

- модели отдельных фирм;
- модели конкурентных отраслей;
- модели дуополий (объединений двух фирм);
- модели олигополии (объединений нескольких фирм);
- модели монополий.



К отраслевым моделям относятся комплексные, или агрегированные, модели, описывающие отдельные отрасли народного хозяйства как единое целое.

Макроэкономические модели предназначены для имитации экономических систем крупного масштаба, таких как область или страна в целом.

Модели управления предприятием

Это микроэкономические модели, отличающиеся друг от друга не столько областью применения, сколько тем, какая типовая математическая схема заложена в основу модели и каковы особенности используемого математического аппарата.

К моделям управления предприятиями относятся:

- модели массового обслуживания;
- модели управления запасами;
- производственные модели;
- модели торговли;
- финансовые модели.

Для многих промышленных систем характерен поток входных требований (заявок), поступающих в один или несколько каналов обслуживания и иногда образующих очередь. Заявками могут быть производственные и торговые заказы, заявки на ремонт станков, посадку самолетов в аэропорту и заправку автомобилей на автозаправочной станции. Канал обслуживания может представлять собой совокупность устройств, этап производственного процесса, аэропорт или театральную кассу. Интервалы между последовательными заявками и продолжительность их обслуживания являются случайными величинами.

К *моделям системы массового обслуживания* относятся модель бензоколонки, нотариальной конторы, парикмахерской, столовой самообслуживания, станции автомобильного обслуживания и т.п. Общей для этих моделей является заложенная типовая математическая схема - схема системы массового обслуживания с переменным числом каналов, однородным потоком заявок, без отказов и с ограниченным ожиданием. В задачу моделирования входит установление оптимального числа каналов, которое при определенном соотношении входных параметров (среднего времени между соседними заявками и среднего времени обслуживания) обеспечивает максимальное значение показателя эффективности процесса функционирования системы. Для конкретной экономической системы в качестве критерия эффективности используется условие максимума прибыли.

К *производственным моделям* относится математическую модель производственной фирмы, включающей несколько цехов, которые последовательно участвуют в процессе производства некоторого изделия. Заказы на изготовление изделия поступают нерегулярно (в случайные моменты времени). При оптимальной структуре предприятия (количестве цехов) и оптимальном распределении производственных ресурсов обеспечивается максимум прибыли. Имитация работы предприятия производится с помощью модели одноканальной многофазной системы массового обслуживания, без отказов с неограниченным ожиданием.

К моделям этого же класса относятся модели, условно названные моделями управленческого звена учреждения, фирмы или предприятия, состоящего из начальника (заведующего) и двух его заместителей. Все они принимают участие в процессе приема посетителей или обработки

документации. Часть посетителей, побывавших на приеме у одного из заместителей, затем отправляются на прием к начальнику. При определенном соотношении параметров системы можно обеспечить практически одинаковую занятость каждого из трех должностных лиц. С точки зрения используемого математического аппарата это модель двухканальной двухфазной системы массового обслуживания с двумя приоритетами заявок, без отказов с неограниченным ожиданием.

Обширную группу промышленных систем, при изучении которых эффективна численная имитация, образуют так называемые системы хранения запасов. Большинство задач управления запасами сводится к поиску оптимального распределения поставок в моделируемую систему. Модель должна дать ответ на вопрос: сколько следует фирме заказывать (или производить) и как часто она должна повторять заказы, чтобы минимизировать сумму издержек хранения запаса, издержек, связанных с организацией поставок, и потерь вследствие недостатка продукта на складе?

К *моделям управления запасами* относятся модели системы управления запасами однородного товара на складе. Предполагается, что, когда уровень запаса падает ниже некоторой критической отметки, оформляется заказ на поставку новой партии товара. При отсутствии товара на складе применяются штрафные санкции. При определенном соотношении параметров системы суммарные расходы на содержание склада могут быть минимизированы.

С точки зрения используемого математического аппарата это математическая модель, в которой две входные переменные (дневной спрос и время выполнения заказа) являются случайными величинами, что определяет случайный характер выходной характеристики - суммарных издержек, характеризующих работу склада за определен-

ный период времени. Время между соседними заявками на приобретение товара и время на выдачу товара в модели не фигурируют. Поэтому эту модель нельзя отнести к классу моделей СМО, однако это тоже непрерывно-стохастическая модель.

К *группе моделей торговли* относится так называемая модель фирмы. Примером такой модели может быть модель выездной торговой точки, которая может вести торговлю в различных пунктах с различными условиями при действии случайных факторов. Задача состоит в установлении закономерностей моделируемого процесса и условий, при которых пункты торговли могут считаться эквивалентными по получаемой прибыли.

В качестве типовой математической схемы здесь использована общая непрерывно-стохастическая модель, в которой имитируется влияние дискретных и непрерывных случайных факторов. Такая модель при изменении комплекта исходных данных может использоваться как вариант *транспортной модели*, в которой осуществляется имитация процесса перевозки грузов по нескольким маршрутам в условиях влияния случайных помех и непостоянства скорости движения на различных участках дороги.

С помощью *финансовой модели (модель инвестиций)* определяется объем капиталовложений в условиях неопределенности. Примером такой модели является модель инвестиционной компании, предполагающей вложить свои средства в строительство нового предприятия. Предприятие будет выпускать продукцию, пользующуюся спросом на рынке. Модель должна оценивать минимальную гарантированную прибыль от продажи продукции в условиях конкуренции на рынке. Финансовая модель позволяет оценивать риски инвесторов. Используемая типовая математическая схема представляет собой непрерывно-

стохастическую модель. Для раскрытия неопределенности необходимо выбрать в качестве одной из входных переменных случайную величину, имеющую произвольное дискретно-непрерывное распределение.

Классификация математических методов

Выделим среди экономико-математических методов и моделей следующие разделы:

- ❖ Математическая статистика и эконометрика:
 - выборочный метод,
 - дисперсионный анализ, корреляционный анализ,
 - регрессионный анализ,
 - многомерный статистический анализ,
 - факторный анализ,
 - теория индексов и др.;
- ❖ Математическая экономика и изучающая те же вопросы с количественной стороны эконометрика:
 - теория экономического роста
 - теория производственных функций
 - межотраслевые балансы,
 - национальные счета,
 - анализ спроса и потребления,
 - региональный и пространственный анализ,
 - глобальное моделирование и др.;
- ❖ Исследование операций (или методы оптимальных решений):
 - линейное программирование,
 - нелинейное программирование,
 - динамическое программирование,
 - целочисленное программирование,
 - сетевые методы планирования и управления,

- теория и методы управления запасами,
 - теория массового обслуживания,
 - теория игр,
 - теория и методы принятия решений,
 - теория расписаний и др.;
- ❖ Методы экспериментального изучения экономических явлений:
- методы анализа и планирования экономических экспериментов,
 - имитационное моделирование,
 - деловые игры и др.;
- ❖ Узкопрофильные методы:
- оптимальное планирование,
 - теория оптимального ценообразования,
 - модели материально-технического,
 - модели капиталистического цикла,
 - модели монополии,
 - модели индикативного планирования,
 - модели теории фирмы и т.д.

Учитывая возросший в последнее время интерес к моделям и методам эконометрики, в настоящем курсе мы изучим основные методы и модели эконометрики.

§2. Основы эконометрики

Понятие эконометрики: предмет, цель, задачи

Термин «эконометрика» впервые появился в исследованиях П. Цъемпы(1910), Й. Шумпетера (1923), Р. Фриша (1930) и является соединением двух слов: «Экономика» и «метрика».

В настоящее время не существует единого строгого определения эконометрики. Наиболее часто употребляется следующее понятие:

Эконометрика – это наука, предметом изучения которой является количественное выражение взаимосвязей экономических явлений и процессов.

Ученые–эконометристы, признанные авторитеты в области эконометрических знаний, по-разному подходили к определению эконометрики.

Р.Фриш: «Эконометрика есть единство трех составляющих – статистики, экономической теории и математики».

С. Фишер: «Эконометрика занимается разработкой и применением статистических методов для измерения взаимосвязей между экономическими переменными».

С. Айвазян: «Эконометрика объединяет совокупность методов и моделей, позволяющих придать количественные выражения качественным зависимостям».

Анализ подходов к определению эконометрики, состояние эконометрической науки позволяют сформулировать цель эконометрики, которая достигается решением определенных задач.

Цель эконометрики – это разработка способов моделирования и количественного анализа реальных экономических объектов и систем.

Задачи эконометрики можно разделить на три типа:

- ❖ по конечным прикладным целям
 - Задачи прогноза экономических показателей, характеризующих состояние и развитие изучаемого явления
 - Задачи моделирования возможных вариантов экономического развития системы для определения параметров которые оказывают наиболее сильное влияние на состояние системы в целом
- ❖ по уровню иерархии
 - Задачи макроуровня (страна в целом)
 - Задачи мезоуровня (уровень отраслей, регионов)
 - Задачи микроуровня (уровень фирмы, предприятия, семьи)
- ❖ по области решения проблем изучаемой экономической системы
 - Закономерности в экономике проявляются как взаимосвязи между экономическими показателями.
 - В большинстве случаев их стремятся выразить в сравнительно простой математической форме.

Эконометрическая модель

Основой эконометрического моделирования является эконометрическая модель. Экономический объект в такой модели описывается и изучается с помощью статистических данных.

Эконометрическая модель учитывает реальные условия существования изучаемого объекта и не противоречит общим законам экономики. Ошибка предсказаний по такой модели не превосходит заданной величины.

Общий вид эконометрической модели

$$Y = f(X) + \varepsilon,$$

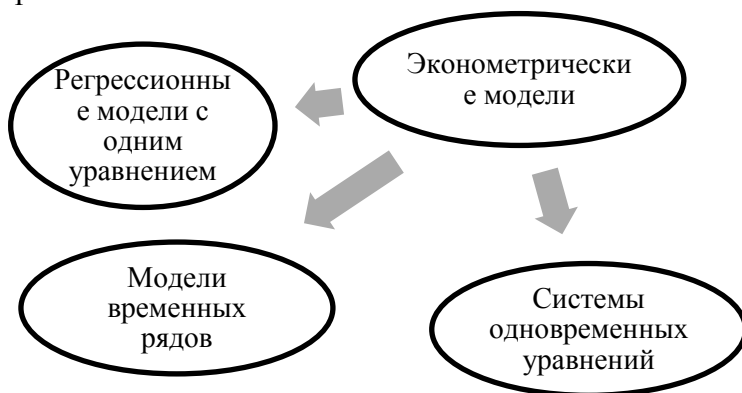
где Y – наблюдаемое значение зависимой переменной (объясняемая переменная, результат); $f(X)$ – объясненная часть зависимой переменной Y , которая зависит от значений объясняющих переменных X (факторов); ε – случайная составляющая (ошибка, возмущение).

Задачи эконометрического моделирования

1. Определение объясненной части, на основе экспериментальных данных
2. Получение оценок параметров распределения случайной составляющей, рассматривая ее как случайную величину

Эконометрическая модель – главный инструмент эконометрики, предназначена для анализа и прогноза экономических объектов, явлений и процессов.

Эконометрические модели можно условно разделить на три класса:



Примеры эконометрических моделей

- ❖ Регрессионные модели с одним уравнением
 - Модель цены от объема поставки
 - Модель спроса от цены на отдельный товар и от реальных доходов потребителей
 - Модель зависимости объема производства от производственных факторов
- ❖ Системы одновременных уравнений
 - Модель спроса и предложения
 - Кейнсианская модель формирования доходов
 - Балансовая модель Леонтьева

❖ Модели временных рядов

- Модель потребления энергоресурсов с учетом сезонности
- Модель объема продаж
- Модель динамики товарооборота и доходов населения

Эконометрические модели отражают **свойства** изучаемых объектов, явлений и процессов:

- свойство **времени** двигаться вперед используется в моделях временных рядов (экономические явления происходят в пространстве и во времени)
- свойство **прошлых, настоящих и будущих значений** переменных влиять на текущее состояние экономического явления реализуется в моделях авторегрессии и автокорреляции, в моделях адаптивного прогноза
- свойство **динамического равновесия** многих экономических явлений применяется в решении систем одновременных уравнений
- свойство **цикличности** большого количества экономических явлений находит место в моделях временных рядов с сезонной составляющей
- свойство **временной задержки (лага)** между причиной и следствием экономического явления проявляется в модели с распределенным лагом

Основные этапы эконометрического моделирования

В процессе эконометрического исследования можно выделить следующие этапы:

- ✓ **1-й этап. Постановочный.** Определение конечных целей моделирования, формирование набора участвующих в модели факторов и их ролей.
- ✓ **2-й этап. Спецификация модели.** Выбор общего вида модели, состава и вида входящих в нее функциональных взаимосвязей.
- ✓ **3-й этап. Информационный.** Сбор необходимой статистической информации, т.е. значений участвующих в модели факторов
- ✓ **4-й этап. Идентификация модели.** Анализ модели, статистическое оценивание параметров модели.
- ✓ **5-й этап. Проверка модели.** Сопоставление реальных и оцененных данных, проверка адекватности модели.

Типы данных в эконометрическом моделировании

В эконометрике применяются два основных типа выборочных данных: пространственные и временные.

Пространственный тип данных - совокупность экономической информации, характеризующие разные объекты за определенный период (момент) времени.

Примеры:

- объем производства предприятий региона

- численность экономически активного населения
- показатели экономической деятельности предприятий

Временной тип данных - совокупность экономической информации, характеризующей определенный объект за различные периоды (моменты) времени.

Примеры:

- индекс потребительских цен
- численность занятых за последние годы
- объем произведенных предприятием товаров

Переменные в эконометрическом моделировании

Переменные в эконометрической модели любого класса условно можно разделить на следующие виды:

Экзогенные (независимые, x) – значения задаются извне модели

Эндогенные (зависимые, y) – значения определяются внутри модели

Лаговые (экзогенные или эндогенные) – значения переменных датируются предыдущими моментами времени и находятся в уравнении с текущими переменными

Предопределенные (лаговые и текущие экзогенные переменные, лаговые эндогенные переменные)

- ✓ Количество независимых переменных(m), включенных в модель, не должно быть слишком большим и должно быть теоретически обоснованным.

- ✓ Качество модели зависит от объема исходных данных (n – объем выборки).
- ✓ Объем выборки и число независимых переменных в модели связаны соотношением $n \geq (6 \div 8) \times t$

Цель эконометрической модели каждого класса – объяснить значения текущих эндогенных (результативных) переменных в зависимости от значений predetermined (независимых) переменных.

§3. Методы и модели корреляционно-регрессионного анализа

Понятие корреляционной зависимости

Экономические явления, обладая большим разнообразием, характеризуются множеством признаков, отражающих те или иные их свойства. Эти признаки изменяются (варьируются) во времени и пространстве. Нередко изменения признаков взаимозависимы и взаимообусловлены. В одних случаях связь (зависимость) между признаками оказывается очень тесной (например, часовая выработка и заработная плата), а в других случаях связь между признаками вовсе не обнаруживается или выражается очень слабо (например, пол студентов и их успеваемость). Чем теснее связь между признаками, тем точнее принимаемые решения и легче управление системами.

Различают два типа зависимости между явлениями и их признаками: *функциональную*, или жестко детерминированную (например, зависимость выработки продукции на одного рабочего от объема выпущенной продукции и численности рабочих), и *статистическую* (например, зависимость между производительностью труда и себестоимостью единицы продукции).

Функциональная зависимость – это связь, при которой каждому значению независимой переменной x соответствует точно *определенное* (единственное) значение зависимой переменной y .

Функциональная зависимость чаще всего встречается в естественных науках. Реже подобные связи наблюдаются в общественной жизни, в частности в экономических процессах.

Для социально-экономических явлений характерно то, что наряду с существенными факторами на них оказывают воздействие многие другие, в том числе случайные факторы. В связи с этим существующая зависимость не проявляется здесь в каждом отдельном случае, как при функциональных связях, а лишь «в общем и среднем» при большом числе наблюдений. В этом случае говорят о статистической зависимости.

Статистическая зависимость – это связь, при которой каждому значению независимой переменной x (факторного признака) соответствует множество значений зависимой переменной y (результативного признака), причем неизвестно заранее, какое именно значение примет y .

Частным случаем статистической зависимости является корреляционная зависимость

Корреляционная зависимость – это связь, при которой каждому значению независимой переменной x соответствует определенное математическое ожидание (среднее значение) зависимой переменной y .

Корреляционная связь является «неполной» зависимостью, которая проявляется не в каждом отдельном слу-

чае, а только в средних величинах при достаточно большом числе случаев.

Известно, например, что повышение квалификации работника ведет к росту производительности труда. Это положение подтверждается в массе явлений и не означает, что у двух или более рабочих одного разряда, занятых аналогичным процессом, будет одинаковая производительность труда. Уровни их выработки будут различаться, хотя и незначительно, так как у этих рабочих могут быть различными стаж работы, техническое состояние станка, состояние здоровья и т.д.

Из этого следует, что ***статистическая зависимость - свойство совокупности в целом, а не отдельных ее единиц.***

Корреляционная связь:

- в зависимости от направления действия бывает прямая и обратная
- по аналитическому выражению может быть прямой (линейной) и криволинейной (нелинейной)
- в зависимости от количества факторных признаков корреляционные связи делят на однофакторные (парные) и многофакторные (множественные)

Понятие корреляционного анализа. Понятие корреляции

Корреляционная зависимость исследуется с помощью методов корреляционного анализа.

Корреляционный анализ – это раздел математической статистики, посвященный изучению взаимосвязей между случайными величинами. Применяется тогда, когда данные наблюдений можно считать случайными и выбранными из генеральной совокупности, распределенной по многомерному

Корреляционный анализ заключается в количественном определении тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Корреляция – это статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Парная корреляция

Мы располагаем n парами выборочных наблюдений над величинами X, Y (т. е. имеем пространственную выборку):

$$\begin{array}{cccccc} x & x_1 & x_2 & \dots & x_n \\ y & y_1 & y_2 & \dots & y_n \end{array}$$

Предварительное представление о зависимости между X и Y можно получить, отобразив элементы выборки как точки на плоскости (x, y) . Такое представление выборки получило название **корреляционное поле (или диаграмма рассеяния)**.

Количественной характеристикой степени линейной зависимости (или еще говорят тесноты связи) между X и Y является **коэффициент корреляции ρ** .

Оценка коэффициента линейной парной корреляции ρ по выборке наблюдений (x_i, y_i) , $i = 1, 2, \dots, n$ вычисляется по формуле:

$$r_{xy} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y},$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$; $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$; $\overline{y \cdot x} = \frac{1}{n} \sum_{i=1}^n (x_i \cdot y_i)$; $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$; $\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$;

$$\sigma_x = \sqrt{\sigma_x^2}; \quad \sigma_y = \sqrt{\sigma_y^2}.$$

Здесь n - объем исследуемой совокупности (объем выборки); \bar{x} , \bar{y} , $\overline{y \cdot x}$ - средние значения; σ_x^2 , σ_y^2 - дисперсии; σ_x , σ_y - среднеквадратические (стандартные) отклонения признаков y , x .

Свойства коэффициента парной корреляции

1. Если признаки x и y независимы, то $r_{xy} = 0$.
2. $r_{xy} = r_{yx}$.
3. При $r_{xy} > 0$ величины x и y одновременно возрастают (прямая зависимость). При $r_{xy} < 0$ с возрастанием величины x (у) величина y (х) убывает (обратная зависимость).
4. Из равенства $r_{xy} = 0$ не следует независимость величин x и y , отрицается лишь их линейная корреляционная зависимость (может существовать нелинейная зависимость).
5. Значения $r_{xy} = -1$, $r_{xy} = 1$ соответствуют практически функциональной линейной связи между величинами x и y .
6. $-1 \leq r_{xy} \leq 1$. Положительное значение коэффициента парной корреляции свидетельствует о наличии прямой связи между признаками, отрицательное - обратной.

Для определения тесноты (силы) линейной корреляционной зависимости можно использовать **шкалу Чеддока**:

Значение $ r_{yx} $	Характер линейной корреляционной связи между признаками у и х
0 - 0,2	Связь практически отсутствует
0,2 - 0,5	Связь слабая
0,5 - 0,7	Связь средняя (умеренная)
0,7 - 0,95	Связь сильная (тесная)
0,95 - 1	Практически функциональная зависимость

Оценка статистической значимости коэффициента парной корреляции

Коэффициент парной корреляции как статистическую величину необходимо оценить на достоверность (*статистическую значимость*).

Для оценки значимости коэффициента парной корреляции используют **t-критерий Стьюдента**, который применяется при **t-распределении**, отличном от нормального, но приближающемся к нормальному при $n > 30$.

Находится расчетное значение (статистика) критерия

$$t_{\text{расч}} = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \cdot \sqrt{n - 2},$$

которое сравнивается с табличным (критическим) значением $t_{\text{табл}}$, найденным по таблице двухсторонних критических значений распределения Стьюдента (табл. Приложения 2) по уровню значимости α и числу степеней свободы $df = n - 2$, или с помощью встроенной функции Excel «СТЮДРАСПОБР». В экономических расчетах значение α обычно принимается равным 0,05.

Если $|t_{расч}| > t_{табл}$, то **коэффициент парной корреляции** признается **статистически значимым**. Если $|t_{расч}| < t_{табл}$, коэффициент корреляции считается незначимым, случайно отличающимся от нуля.

Множественная корреляция

Коэффициент множественной корреляции рассчитывается в случае, когда исследуется взаимосвязь трех и более признаков.

Для трех признаков x, y, z коэффициент множественной корреляции между величиной z и парой величин x, y вычисляется на основе корреляционной матрицы:

	z	x	y
z	1	r_{zx}	r_{zy}
x	r_{xz}	1	r_{xy}
y	r_{yz}	r_{yx}	1

$$R_z = \sqrt{1 - \frac{|r|}{|r_{11}|}}$$

где $|r| = \begin{vmatrix} 1 & r_{zx} & r_{zy} \\ r_{xz} & 1 & r_{xy} \\ r_{yz} & r_{yx} & 1 \end{vmatrix}$ - определитель корреляционной матрицы,

$|r_{11}| = \begin{vmatrix} 1 & r_{xy} \\ r_{yx} & 1 \end{vmatrix}$ - определитель матрицы межфакторных корреляций (корреляций между независимыми переменными).

Коэффициент множественной корреляции является мерой **линейной** связи между одним признаком и остальными (двумя и более).

Коэффициент множественной корреляции принимает значения от 0 до 1.

Теснота связи между признаками устанавливается с помощью таблицы:

Значение коэффициента множественной корреляции R	Характер линейной корреляционной связи между признаками z и x, y
0 - 0,1	Слабая
0,1 - 0,5	Средняя (умеренная)
0,5 - 1	Сильная (тесная)

Для коэффициента множественной корреляции справедливы следующие неравенства

$$R_z \geq |r_{zx}|, R_z \geq |r_{zy}|, R_z \geq |r_{zx|y}|, R_z \geq |r_{zy|x}|$$

Из них следует, что коэффициент множественной корреляции может только увеличиться, если в модель включать дополнительные признаки, и не увеличиться, если из имеющегося набора признаков производить исключение.

Коэффициент множественной корреляции R для произвольного числа признаков x_1, x_2, \dots, x_m ($m > 3$) находят по рекуррентным формулам, что требует большого опыта вычислений. Для них справедливы все изложенные выше свойства и выводы. Для расчетов обычно применяют ППП.

Проверка статистической значимости коэффициента множественной корреляции

Проверка осуществляется с помощью ***F-критерия Фишера***. Рассчитывает расчетное значение критерия

$$F_{\text{расч}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

Здесь m – число факторов в модели, n – объем выборки.

$F_{\text{расч}}$ сравнивается с табличным значением $F_{\text{табл}}$, найденное по таблице критических значений распределения Фишера-Снедекора по уровню значимости α и числу степеней свободы $df_1 = m$ и $df_2 = n - m - 1$ или с помощью встроенной функции Excel «ФРАСПОБР».

Если $F_{\text{расч}} > F_{\text{табл}}(\alpha; m; n - m - 1)$, то R признается статистически значимым.

Коэффициент детерминации

На основе коэффициента корреляции рассчитывается *коэффициент детерминации* R^2 .

Для парной линейной регрессии $R^2 = (r_{xy})^2$

Коэффициент детерминации оценивает долю вариации признака y , обусловленную изменением значений признака x . Чем ближе значение R^2 к единице, тем больше признак x участвует в формировании значений y .

Для множественной линейной регрессии $R^2 = (R)^2$

Коэффициент детерминации оценивает долю вариации признака y , обусловленную изменением значений факторов x_1, x_2, \dots, x_m . Чем ближе значение R^2 к единице, тем больше факторы x_1, x_2, \dots, x_m участвует в формировании значений y .

На практике при большом числе переменных сначала вычисляют коэффициент детерминации, а затем - коэффициент множественной корреляции по формуле

$$R = \sqrt{R^2}$$

Проверка статистической значимости коэффициента детерминации R^2

Проверка осуществляется с помощью ***F-критерия Фишера***. Рассчитывает расчетное значение критерия

$$F_{\text{расч}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

Здесь m – число факторов в модели, n – объем выборки.

$F_{\text{расч}}$ сравнивается с табличным значением $F_{\text{табл}}$, найденное по таблице критических значений распределения Фишера-Снедекора по уровню значимости α и числу степеней свободы $df_1 = m$ и $df_2 = n - m - 1$.

Если $F_{\text{расч}} > F_{\text{табл}}(\alpha; m; n - m - 1)$, то R признается статистически значимым.

Статистическая значимость коэффициента множественной корреляции совпадает со статистической значимостью коэффициента детерминации.

Понятие регрессионного анализа

Рассмотрим некоторый экономический объект и выделим только две переменные, характеризующие этот объект.

Независимая¹ (объясняющая) переменная X оказывает воздействие на значения переменной Y , которая, таким образом, является зависимой переменной.

Функция $f(x)$ называется **регрессией** Y по X , если она описывает изменение условного среднего значения переменной Y в зависимости от значения переменной x :

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

Регрессионная модель имеет вид:

$$y = \hat{y} + \varepsilon \text{ или } y = \hat{y} \cdot \varepsilon,$$

где $\hat{y} = f(x)$ - уравнение парной регрессии, ε - остаток (отклонение).

Модель парной регрессии применяется, если имеется возможность выделить один доминирующий фактор, обуславливающий большую долю изменения изучаемой объясняемой переменной.

В ситуациях, когда из множества факторов, влияющих на изучаемую переменную, нельзя выделить один фактор и необходимо учитывать одновременное влияние нескольких фактор, строится **модель множественной регрессии**:

$$y = \hat{y} + \varepsilon \text{ или } y = \hat{y} \cdot \varepsilon,$$

где $\hat{y} = f(x_1, x_2, \dots, x_m)$ - уравнение множественной регрессии, ε - остаток (отклонение).

¹ Термин «независимая переменная» здесь и далее используется в следующем контексте: переменная, значения которой в процессе определения отклика, могут устанавливаться произвольно, независимо.

Существуют различные регрессионные модели, определяемые видом функции $f(x_1, x_2, \dots, x_m)$:

1. Парная линейная регрессия:

$$y = \beta_0 + \beta_1 x + \varepsilon;$$

2. Множественная линейная регрессия:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon;$$

3. Полиномиальная парная регрессия:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_k x^k + \varepsilon;$$

4. Регрессионная модель общего вида:

$$y = \beta_0 + \beta_1 \varphi(x_1, x_2, \dots, x_m) + \dots + \beta_k \varphi_k(x_1, x_2, \dots, x_m) + \varepsilon,$$

где $\varphi_i(x_1, x_2, \dots, x_m), i = 1, 2, \dots, k$ – заданные функции факторов.

Коэффициенты $\beta_0, \beta_1, \dots, \beta_k$ называются **параметрами регрессии**.

В приведенные выше модели параметры входят линейно. Такие модели принято называть **линейными по параметрам моделями**. Математические методы анализа этих моделей называют **линейным регрессионным анализом**.

Существуют модели, нелинейные по параметрам. Например, $y = \beta_0 e^{\beta_1 x_1}$ – нелинейна по параметрам.

В некоторых случаях нелинейные модели с помощью специальных линеаризующих преобразований могут быть преобразованы в линейные. Рассмотрим несколько примеров.

- Степенная функция $y = b_0 \cdot x^{b_1}$ при $x > 0$ с помощью логарифмирования и замены переменных преобразуется к следующему виду:

$$\ln y = \ln b_0 + b_1 \ln x$$

Введя новые переменные $Y = \ln y, B_0 = \ln b_0, X = \ln x$, получим линейную по параметрам функцию:

$$Y = B_0 + b_1X$$

- Экспоненциальная функция $y = b_0 \cdot e^{b_1x}$ с помощью логарифмирования и замены переменных преобразуется к следующему виду:

$$\ln y = \ln b_0 + b_1x$$

Введя новые переменные $Y = \ln y$, $B_0 = \ln b_0$, получим линейную по параметрам функцию:

$$Y = B_0 + b_1x$$

- Функция $y = \frac{ax}{b+x}$ преобразовывается следующим образом:

$$b + x = a \frac{x}{y} \text{ или } \frac{x}{y} = \frac{b}{a} + \frac{1}{a}x$$

Далее, замена переменных $Y = \frac{x}{y}$; $b_0 = \frac{b}{a}$; $b_1 = \frac{1}{a}$. Получаем, что

$$Y = b_0 + b_1x.$$

- Логистическая функция $y = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}}$ при помощи преобразования $Y = \ln \frac{y}{1-y}$ принимает вид:

$$Y = b_0 + b_1x.$$

После выбора вида регрессионной модели нужно оценить параметры регрессии (на основе имеющихся наблюдений зависимой переменной и факторов), проверить значимость и адекватность модели результатам наблюдений.

Парная регрессия

Уравнение парной линейной регрессии имеет вид

$$\hat{y} = b_0 + b_1x,$$

где b_0, b_1 – параметры уравнения.

b_0 – **свободный член** регрессионного уравнения. Рассматривается как начальное значение результативного признака y , если факторный признак x может принимать нулевое значение ($x = 0$).

b_1 – **коэффициент регрессии** показывает, на какую величину в среднем изменится результативный признак y , если факторный признак x увеличить на единицу своего измерения. Знак при коэффициенте регрессии показывает направление связи между y и x : при $b_1 > 0$ - связь прямая; при $b_1 < 0$ - связь обратная.

Нахождение значений параметров b_0, b_1 производится на основе совокупности наблюдений (выборки, **матрицы наблюдений**). Для разных выборок (даже одного объема) из генеральной совокупности будут найдены разные значения b_0, b_1 .

Поэтому их рассматривают как приближенные значения (**оценки**) истинных параметров регрессионного уравнения. Сама процедура нахождения приближенных значений также называется **оценкой параметров**.

Для получения оценок параметров линейного регрессионного уравнения используют **метод наименьших квадратов** (МНК), который минимизирует функцию

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

Здесь y_i, \hat{y}_i – соответственно фактическое (наблюдаемое), теоретическое значение зависимой переменной y , найденное по уравнению регрессии; x_i – наблюдаемое значение независимой переменной; n – объем выборки.

Из необходимых условий минимума находят формулы для оценок параметров регрессии:

$$\begin{cases} b_1 = \frac{\overline{y\bar{x}} - \bar{y} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2} \\ b_0 = \bar{y} - b_1 \bar{x} \end{cases}$$

Здесь $\overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$.

Требования теоремы Гаусса-Маркова (предпосылки МНК)

Предположим, что выполняются следующие условия:

A – «истинная» зависимость y от x имеет вид

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon$$

B - x – неслучайная переменная (детерминированная)

C - столбцы матрицы наблюдений (статистических данных) с добавлением единичного столбца, линейно независимы (ранг матрицы равен 2)

D – остатки ε_i имеют нулевое математическое ожидание $M(\varepsilon_i) = 0$ и постоянную дисперсию σ_ε^2 , не зависящую от номера наблюдения (свойство гомоскедастичности), т.е. $D(\varepsilon_i) = \sigma_\varepsilon^2 = const$.

E – для разных наблюдений остатки ε_i некоррелированы (независимы), т.е. $cov(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$, $i, j = 1, \dots, n$.

F – остатки подчиняются нормальному закону распределения.

В предположениях A-E оценки, полученные МНК, являются несмещенными и обладают наименьшей дисперсией.

Другими словами, являются наилучшими.

Оценка статистической значимости коэффициентов уравнения парной регрессии

Проверка статистической значимости оценок параметров b_0, b_1 также производится с помощью t-критерия Стьюдента.

Находятся расчетные значения критерия:

$$t_{b_0} = \frac{b_0}{m_{b_0}}; t_{b_1} = \frac{b_1}{m_{b_1}},$$

где средние квадратические ошибки параметров равны:

$$m_{b_0} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - 2} \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}};$$
$$m_{b_1} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{(n - 2) \sum_{i=1}^n (x_i - \bar{x})^2}}.$$

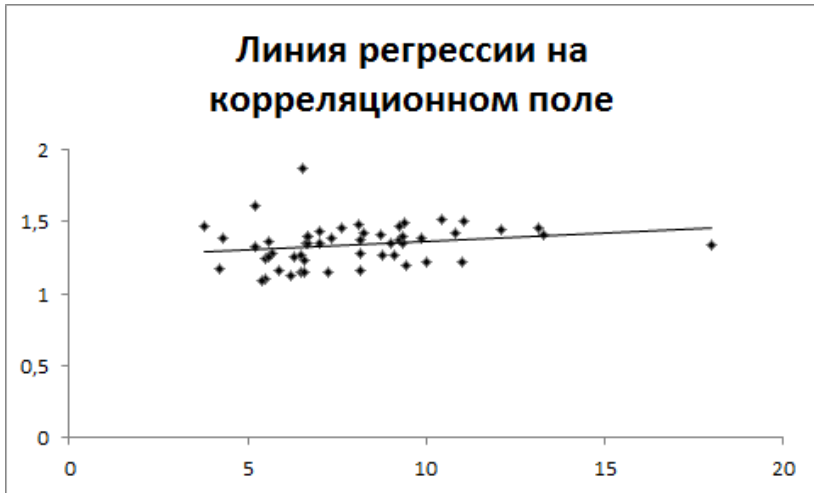
Табличное значение критерия $t_{табл}$ находится по таблице критических значений распределения Стьюдента (табл. Приложения 2) по уровню значимости α и числу степеней свободы $df = n - m - 1$, или с помощью встроенной функции Excel «СТЮДРАСПОБР». Для парной регрессии $m=1$.

Если $|t_{bj}| > t_{табл}$, то с вероятностью ошибки α оценка параметра уравнения регрессии b_j ($j = 0, 1$) признается статистически значимой.

В противном случае ($|t_{bj}| < t_{табл}$) - b_j статистически незначим.

Оценка качества уравнения линейной регрессии

- Найденное уравнение парной линейной регрессии графически изображается прямой линией на корреляционном поле:



Линия регрессии занимает правильное положение на корреляционном поле, если число точек, лежащих выше и ниже линии регрессии примерно равно.

- Если $\bar{y} \approx \widehat{\bar{y}}$, арифметические расчеты выполнены верно.
- Для оценки математической точности уравнения можно воспользоваться **средней относительной ошибкой аппроксимации**

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Для принятия решения о точности уравнения пользуются следующей таблицей:

Значение \bar{A} , %	Точность уравнения
менее 10	высокая
10 - 20	хорошая
20 - 50	удовлетворительная
более 50	неудовлетворительная

В случае, когда уравнение имеет неудовлетворительную точность, необходимо увеличить объем наблюдений (объем выборки) n , скорректировать набор объясняющих переменных x_1, x_2, \dots, x_m , либо взять другое уравнение регрессии (нелинейное).

➤ **Проверка статистической значимости уравнения регрессии в целом** осуществляется с помощью F-критерия Фишера.

Находится расчетное значение (статистика) критерия

$$F_{\text{расч}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (\hat{y}_i - y_i)^2} \cdot \frac{(n - m - 1)}{m}$$

В случае парной линейной регрессии $m = 1$.

Табличное (критическое) значение $F_{\text{табл}}$, находится по таблице критических значений распределения Фишера-Снедекора (F-распределения) (табл. Приложения 3) по уровню значимости α и двум числам степеней свободы $df_1 = m$ и $df_2 = n - m - 1$, или с помощью встроенной функции Excel «FРАСПОБР».

Если $F_{расч} > F_{табл}$, то с вероятностью ошибки α уравнение регрессии признается **в целом статистически значимым** (адекватно описывающим исходные данные).

В противном случае ($F_{расч} < F_{табл}$) уравнение считается незначимым в целом.

Множественная регрессия

Если переменная y зависит от нескольких факторов x_1, x_2, \dots, x_m , то регрессионная модель определяется уравнением множественной регрессии:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m$$

Исходные данные для регрессионного анализа представляют собой результаты наблюдений зависимой переменной y и факторов x_1, x_2, \dots, x_m , записанных в виде матрицы наблюдений:

y	x_1	x_2	...	x_m
y_1	x_{11}	x_{21}	...	x_{m1}
...
y_n	x_{1n}	x_{2n}	...	x_{mn}

Регрессионный анализ парной линейной регрессии обобщается на случай множественной регрессии.

Для получения оценок параметров линейного регрессионного уравнения используется метод наименьших квадратов, дающий наилучшие, согласно **теореме Гаусса-Маркова**, результаты.

Напомним, что сущность МНК заключается в том, что отыскиваются такие значения параметров уравнения регрессии, при которых

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_m x_{mi}))^2 \rightarrow \min$$

Из необходимых условий минимума находят формулы для оценок параметров регрессии.

На практике для нахождения параметров уравнения множественной линейной регрессии обычно пользуются восторженным пакетом прикладных программ (ППП) MS Excel «Анализ данных», инструмент «Регрессия».

Требования теоремы Гаусса-Маркова (предпосылки МНК)

А – «истинная» зависимость между y и x_1, \dots, x_m имеет вид

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon$$

В – факторы x_1, \dots, x_m неслучайные величины, т.е. детерминированные.

С – столбцы матрицы наблюдений с добавлением единичного столбца линейно независимые. (ранг матрицы равен $m + 1$).

Д – остатки ε имеют нулевое математическое ожидание и постоянную дисперсию, не зависящую от номера наблюдения (свойство гомоскедастичности): $M(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma_\varepsilon^2 = \text{const}$

Е – для различных наблюдений остатки ε некоррелированы, т.е. независимые: $\text{cov}(\varepsilon_i; \varepsilon_j) = 0$ при $i \neq j$.

Теорема Гаусса-Маркова. В предположениях А-Е оценки, полученные с помощью МНК являются несмещенными и обладают наименьшей дисперсией среди всех оценок β_j .

Оценка качества уравнения множественной линейной регрессии

Так же, как и в случае парной линейной регрессии необходимо проверить качество построенного регрессионного уравнения и провести анализ остатков.

Оценка качества уравнения линейной регрессии состоит из следующих этапов:

1. Оценка математической точности уравнения. Для этого рассчитывается **средняя относительная ошибка аппроксимации**
2. Проверка статистической значимости уравнения регрессии в целом с помощью **F-критерия Фишера**.
3. Для **линейного уравнения** осуществляется проверка статистической значимости оценок параметров $b_0, b_1, b_2, \dots, b_m$ с помощью **t-критерия Стьюдента**.

Оценка математической точности модели осуществляется с помощью вычисления средней относительной ошибки аппроксимации \bar{A} :

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Для принятия решения о точности уравнения пользуются следующей таблицей:

Значение \bar{A} , %	Точность уравнения
менее 10	высокая
10 - 20	хорошая
20 - 50	удовлетворительная
более 50	неудовлетворительная

В случае, когда уравнение имеет неудовлетворительную точность, необходимо увеличить объем наблюдений, скорректировать набор объясняющих переменных (факторов), либо взять другое уравнение регрессии (нелинейное).

Проверка статистической значимости уравнения регрессии в целом осуществляется с помощью критерия Фишера-Снедекора. Вычисляется расчетное значение критерия:

$$F_{\text{расч}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

Здесь m – число факторов в модели, n – объем выборки, R^2 – коэффициент детерминации.

$F_{\text{расч}}$ сравнивается с табличным значением $F_{\text{табл}}$, найденное по таблице критических значений распределения Фишера-Снедекора по уровню значимости α и числу степеней свободы $df_1 = m$ и $df_2 = n - m - 1$.

Если $F_{\text{расч}} > F_{\text{табл}}$, то уравнение регрессии признается статистически значимым в целом.

Проверка статистической значимости коэффициентов линейного уравнения регрессии осуществляется с помощью t-критерия Стьюдента.

Находятся расчетные значения критерия:

$$t_{b_j} = \frac{b_j}{S_{\varepsilon} \sqrt{a_j}}$$

где $S_\varepsilon = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m-1}}$ - стандартная ошибка уравнения регрессии, a_j - диагональные элементы матрицы $(X^T X)^{-1}$, $X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{bmatrix}$.

Расчетные значения t_{b_j} сравниваются с табличным значением критических точек распределения Стьюдента, вычисляемых по уровню значимости α и числу степеней свободы $df = n - m - 1$.

Если $|t_{b_j}| > t_{\text{табл}}(\alpha; n - m - 1)$, то коэффициент b_j признается статистически значимым. В противном случае, статистически незначимым.

На практике стандартные ошибки коэффициентов регрессии и расчетные значения критерия Стьюдента находятся с помощью пакета MS Excel «Анализ данных», инструмент «Регрессия».

Если уравнение регрессии имеет хорошую математическую точность, статистически значимо в целом и по отдельным параметрам, оно признается качественным.

Проверка выполнения предположений регрессионного анализа по остаткам

Остатки $\varepsilon_i = y_i - \hat{y}_i$ должны удовлетворять требованиям **D**, **E** теоремы Гаусса-Маркова.

Проверка требования D

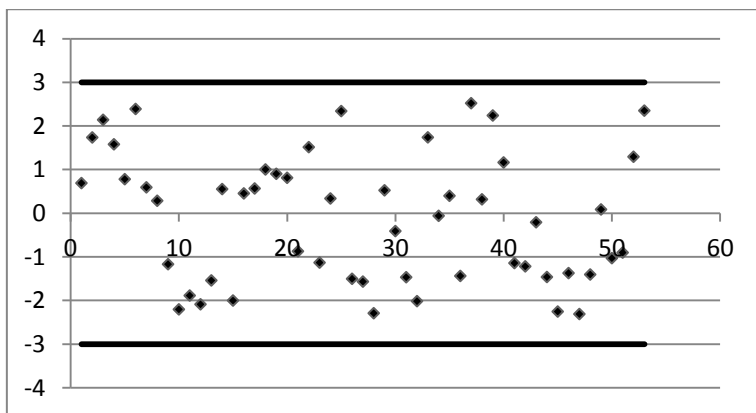
Числовой оценкой математического ожидания $M(\varepsilon_i)$ является среднее значение $\bar{\varepsilon}$. Для выполнения требования $M(\varepsilon_i) = 0$ необходимо, чтобы

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \approx 0$$

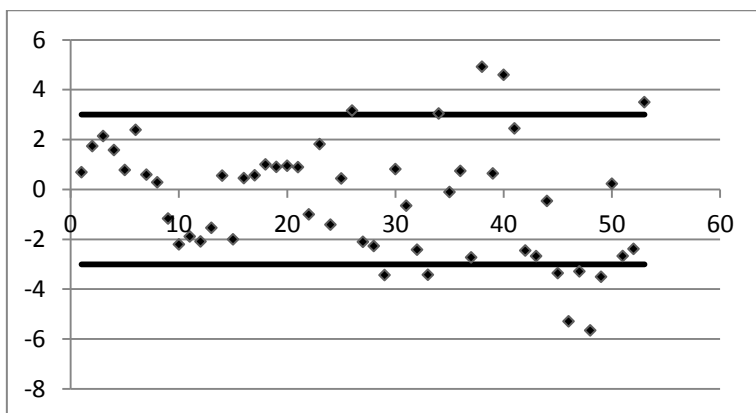
Дисперсия остатков ε_i должна быть одинаковой для всех значений x_i (свойство **гомоскедастичности**). Если это условие не соблюдается, то имеет место **гетероскедастичность** остатков.

Эффективным методом предварительного анализа однородности остатков является визуальный анализ графиков остатков. Как правило, рассматривается график остатков в зависимости от номера наблюдений.

В случае гомоскедастичности точки (остатки) равномерно располагаются внутри горизонтальной полосы, симметричной оси абсцисс:



В случае гетероскедастичности точки (остатки), начиная с некоторого номера наблюдения, устойчиво выходят из горизонтальной полосы, симметричной оси абсцисс :



Для обнаружения эффекта гетероскедастичности используют тесты Голдфелда-Квандта, Глейзера и др., но они достаточно трудоемки, и на практике наиболее часто ограничиваются визуальным анализом.

Проверка требования E

Для разных наблюдений остатки ε_i некоррелированы (независимы).

Наиболее распространенный метод проверки требования о независимости остатков - *критерий Дарбина-Уотсона* (о наличии в остатках автокорреляции первого порядка), в котором рассчитывается статистика

$$d_{расч} = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n \varepsilon_i^2},$$

где $0 \leq d_{расч} \leq 4$.

Для парного уравнения регрессии теоретические значения критерия Дарбина-Уотсона d_U и d_L находятся по таблице критических значений (табл. Приложения 5) по объему выборки n , $df = m$ (m - число параметров уравнения при объясняющих переменных, для парной регрессии равно 1) и уровню значимости α . С помощью критических значений числовой промежуток (0; 4) разбивается на пять отрезков.

Правило принятия решения о зависимости остатков

Есть положительная автокорреляция остатков	Зона неопределенности	Автокорреляция остатков отсутствует	Зона неопределенности	Есть отрицательная автокорреляция остатков
0	d_L	2	$4 - d_U$	$4 - d_L$
	d_U	4	$4 - d_U$	4

На практике в случае, когда статистика критерия $d_{расч}$ попадает в зону неопределенности, обычно признается наличие автокорреляции в остатках.

Отбор факторных признаков в модель

С формальной точки зрения, объясняющие переменные в эконометрической модели должны быть сильно коррелированы с зависимой переменной (имеется в виду теснота связи, т.е. или $r_{x_j y}$, или $R_{x_j y}$).

Во-вторых, факторы должны быть слабо коррелированы между собой.

В-третьих, должны быть сильно коррелированы с другими переменными, не использованными в модели.

Отбор факторов на основе количественного и качественного анализа экономических явлений с использованием статистических и математических методов осуществляется в три этапа:

- ❖ Предварительное определение перечня факторов, исходя из экономической сущности решаемой проблемы, которые оказывают влияние на переменную y .
- ❖ Сравнительный анализ факторов и осуществить отсев на основе проведенного анализа.
- ❖ Окончательный отбор факторов в процессе построения разных вариантов моделей и оценки значимости их параметров.

Факторы, включаемые в множественную регрессию, должны отвечать следующим требованиям:

- ✓ Факторы не должны быть взаимно коррелированы, или еще говорят ***интеркоррелированы***, и тем более находится в функциональной зависимости.
- ✓ Включение в модель фактора должно приводить к существенному увеличению доли объясненной части в общей вариации зависимой переменной. Т.е. к

увеличению значения коэффициента детерминации R^2 .

- ✓ Факторы должны быть количественно измеримы.
- ✓ Переменные не должны быть мультиколлинеарны

Интеркорреляция

Интеркорреляция – это корреляция между объясняющими переменными x_1 и x_2 , для которой одновременно выполняются следующие условия:

$$|r_{x_1x_2}| > 0,75; |r_{yx_1}| < |r_{x_1x_2}|; |r_{yx_2}| < |r_{x_1x_2}|$$

Пример. $r_{x_1x_2} = 0,9$, $r_{yx_1} = 0,87$; $r_{yx_2} = 0,89$. Переменные x_1 и x_2 интеркоррелированы. Так как $0,89 > 0,87$, то в модель нужно включить переменную x_2 .

Включение в модель фактора должно приводить к существенному увеличению доли объясненной части в общей вариации зависимой переменной. Т.е. к увеличению значения коэффициента детерминации R^2 .

Пример. Известно, что $R^2_{yx_1x_2} = 0,65$.

В модель включили переменную x_3 . $R^2_{yx_1x_2x_3} = 0,23$.

Вывод: переменную x_3 в модель включать не следует.

На втором шаге включили фактор x_4 . $R^2_{yx_1x_2x_4} = 0,83$.

Вывод: фактор x_4 следует включить в модель.

Фиктивные переменные

Уравнение множественной регрессии могут включать в качестве независимых переменных качественные признаки (например, пол, профессия, образование, климатические условия, округа, и.т.д.)

Чтобы ввести такие переменные в регрессионную модель, их необходимо упорядочить и присвоить им те или иные значения, т.е. преобразовать в количественные переменные.

Такого вида переменные в эконометрике принято называть фиктивными. Фиктивные переменные могут принимать в качестве значений или «0», или «1». Количество фиктивных переменных на единицу меньше, чем количество уровней качественной переменной.

Например, мы хотим включить в модель фактор «пол». Тогда нам необходимо ввести фиктивную переменную

$$z = \begin{cases} 0 & \text{— женский пол,} \\ 1 & \text{— мужской пол.} \end{cases}$$

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории (например, женского пола) к другой категории (мужскому полу) при неизменных значениях остальных факторных переменных.

На основе t-критерия Стьюдента делается вывод о значимости влияния фиктивной переменной, существенно расхождения между категориями.

Мультиколлинеарность

Под *мультиколлинеарностью* понимается высокая взаимная коррелированность объясняющих переменных (двух и более). Следствием мультиколлинеарностью является линейная зависимость между столбцами матрицы наблюдений.

Проверка наличия мультиколлинеарности основывается на анализе матрицы парных корреляций между факторами:

$$R = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & r_{x_2x_2} & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & \dots & r_{x_mx_m} \end{pmatrix}$$

$r_{x_ix_j}$ - используются для выявления дублирующих факторов

Для оценки мультиколлинеарности можно использовать величину $\det(R)$.

$$\det R = \begin{cases} 1, & \text{то } x_1, \dots, x_m \text{ не коррелируют} \\ 0, & \text{то полная линейная зависимость} \\ & \text{между } x_1, \dots, x_m \\ & \text{чем ближе к 0, тем сильнее} \\ & \text{мультиколлинеарность факторов} \end{cases}$$

Оценка значимости мультиколлинеарности

Проверяется гипотеза о независимости переменных
 $H_0: \det R = 1$.

Вычисляется величина

$$FG_{\text{расч}} = - \left(n - 1 - \frac{1}{6} (2m + 5) \ln(\det R) \right)$$

Значение $FG_{\text{расч}}$ сравнивается с

$$FG_{\text{крит}} = \chi^2(df, \alpha), \text{ где } df = \frac{1}{2} m(m - 1)$$

Если $FG_{\text{расч}} > FG_{\text{крит}}$, то в массиве объясняющих переменных существует мультиколлинеарность

Вопрос: Как найти переменные ответственные за мультиколлинеарность факторов?

Ответ: С помощью коэффициентов множественной детерминации

1. Фактор x_i поочередно рассматривается как результат.
2. Рассчитывается $R_{x_i|x_1x_2,\dots,x_{i-1}x_{i+1},\dots,x_m}^2$.
3. Если $R_{x_i|x_1x_2,\dots,x_{i-1}x_{i+1},\dots,x_m}^2 > 0,6$, то мультиколлинеарность считается установленной.

Практические приложения регрессионной модели

Эластичность

Важную роль в оценке влияния факторов на результат играют коэффициенты регрессионной модели. Однако для непосредственного сравнения степени влияния различных факторов на результат нельзя сравнивать коэффициенты регрессии в виду их различной размерности. Для устранения данных различий вводятся и вычисляются средние коэффициенты эластичности $\bar{\varepsilon}_j$:

$$\bar{\varepsilon}_j = b_j \cdot \frac{\bar{x}_j}{\bar{y}}$$

Здесь b_j – коэффициент регрессии при факторе x_j ; \bar{y} – среднее значение результативного признака; \bar{x}_j – среднее значение факторного признака x_j .

Данный коэффициент также называют средним по совокупности наблюдений коэффициентом эластичности.

$\bar{\varepsilon}_j$ показывает на сколько процентов изменится зависимая переменная y при изменении фактора x_j на 1%, при условии, что все остальные факторы постоянны и закреплены на их среднем уровне.

Пример. По ряду регионов построено уравнение множественной регрессии импорта на определенный товар (y) относительно отечественного производства данного товара (x_1), изменения запасов (x_2) и потребления на внутреннем рынке (x_3):

$$\hat{y} = -66,028 + 0,135x_1 + 0,576x_2 + 0,343x_3$$

$$\bar{y} = 31,5; \bar{x}_1 = 245,7; \bar{x}_2 = 3,7; \bar{x}_3 = 182,5.$$

Для x_1 :

$$\bar{\varepsilon}_1 = b_1 \frac{\bar{x}_1}{\bar{y}} = 0,135 \cdot \frac{245,7}{31,5} = 1,053\%$$

При увеличении отечественного производства товара на 1% импорт товара увеличится на 1,053% (при условии, что изменение запасов $x_2 = 3,7$, а потребление на внутреннем рынке $x_3 = 182,5$).

Для x_2 :

$$\bar{\varepsilon}_2 = b_2 \frac{\bar{x}_2}{\bar{y}} = 0,576 \cdot \frac{3,7}{31,5} = 0,068\%$$

При увеличении изменения запасов на 1% импорт товара увеличится на 0,068% (при условии, что отечественное производство товара составляет 245,7 и потребление на внутреннем рынке составляет 182,5).

Для x_3 :

$$\bar{\varepsilon}_3 = b_3 \frac{\bar{x}_3}{\bar{y}} = 0,343 \cdot \frac{182,5}{31,5} = 1,987\%$$

При увеличении потребления на внутреннем рынке на 1% импорт товара увеличивается на 1,987% (при условии, что отечественное производство товара принимает значение 245,7, а изменение запасов составляет 3,7).

Прогнозирование

В прогнозных расчетах предсказываемое значение $y_{\text{пр}}$ определяется как точечный прогноз $\widehat{y}_{\text{пр}}$ путем подста-

новки в уравнение регрессии заданных значений факторных признаков: $x_{1\text{пр}}, x_{2\text{пр}}, \dots, x_{m\text{пр}}$:

$$\widehat{y}_{\text{пр}} = b_0 + b_1 \cdot x_{1\text{пр}} + b_2 \cdot x_{2\text{пр}} + \dots + b_m \cdot x_{m\text{пр}}$$

Наибольшая точность прогноза достигается в случаях, когда прогнозные значения факторов находятся в центре области наблюдений (незначительно отличаются от средних значений факторов). При удалении от центра ошибка прогноза будет возрастать.

Поэтому помимо точечного прогноза всегда вычисляют интервальную оценку прогноза:

$$\widehat{y}_{\text{пр}} - t_{\text{табл}} \cdot m_{\widehat{y}_{\text{пр}}} < y_{\text{пр}} < \widehat{y}_{\text{пр}} + t_{\text{табл}} \cdot m_{\widehat{y}_{\text{пр}}}$$

Здесь $m_{\widehat{y}_{\text{пр}}}$ – средняя стандартная ошибка прогноза, $t_{\text{табл}}$ – критическое значение распределения Стьюдента, найденная по уровню значимости α и числу степеней свободы $df = n - m - 1$.

Средняя стандартная ошибка прогноза вычисляется по следующим формулам:

А) Парная регрессия:

$$m_{\widehat{y}_{\text{пр}}} = \sqrt{S^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$$

В) Множественная регрессия:

$$m_{\widehat{y}_{\text{пр}}} = \sqrt{S^2 X_{\text{пр}}^T \cdot (X^T X)^{-1} \cdot X_{\text{пр}}}$$

Замечание. В виду трудоемкости вычислений, в случае прогнозирования на основе множественной регрессии на практике ограничиваются вычислением точечного прогноза.

§4. Решение задач корреляционно-регрессионного анализа

Пакет «Анализ данных»

Решение всех задач в настоящем курсе будут выполняться с использованием средств MS Excel. Ниже приводятся некоторые отличия в версиях MS Excel, знание которых облегчит магистру выполнение работы.

Основное отличие MS Excel 2010 от MS Excel 2003 – это полностью переработанный пользовательский интерфейс.

Функциональные возможности версий MS Excel 2003 и MS Excel 2010 практически не отличаются.

Ссылки (или имена) как аргументы функций и формул

Нежелательным является использование конкретных чисел в качестве аргументов функций и формул MS Excel. Этого нужно по возможности избегать.

Например, при вычислении критического значения t -статистики Стьюдента для уровня значимости $\alpha = 0,05$ и числа степеней свободы 38 используется запись:

=СТЮДРАСПОБР(0.05;38).

Данная формула не является универсальной, так как при смене аргументов необходимо будет каждый раз ее редактировать. Правильнее было бы вместо этого использовать три ячейки следующего содержания:

	A	B	C	D	E	F
	уровень значимости	0,05				
1						
2	n	16				
3	m	3				
4	$t_{кр}$	=СТЮДРАСПОБР(B1;B2-B3-1)				
5		СТЮДРАСПОБР(вероятность; степени_свободы)				
6						

Это позволит при необходимости быстро менять аргументы функций, не редактируя саму формулу.

Вставка функций

Кроме ввода формул для выполнения базовых математических операций, таких как сложение, вычитание, умножение и деление, в MS Excel можно использовать большую библиотеку встроенных функций для выполнения других операций. Для вставки функции необходимо:

Excel 2003	Excel 2010
<ol style="list-style-type: none"> 1. В меню Вставка выберите команду Функции... 2. В раскрывшемся диалоговом окне следуйте инструкциям Мастера функций. 	<ol style="list-style-type: none"> 1. На вкладке Формулы в группе Библиотека функций нажмите кнопку Вставить функцию. 2. В раскрывшемся диалоговом окне следуйте инструкциям Мастера функций.

Использование надстройки «Пакет анализа»

При проведении сложного статистического или инженерного анализа можно упростить процесс и сэкономить время, используя надстройку **Пакет анализа**.

Загрузка надстройки **Анализ данных**²

1. Откройте вкладку **Файл** и выберите пункт **Параметры** (в Microsoft Office 2007 щелкните значок



Кнопка Microsoft Office, а затем **Параметры Excel**).

2. Выберите команду **Надстройки**, а затем в поле **Управление** выберите пункт **Надстройки Excel**.
3. Нажмите кнопку **Перейти**.
4. В окне **Доступные надстройки** установите флажок **Пакет анализа** и нажмите кнопку **ОК**.
5. После загрузки надстройки **Пакет анализа** в группе **Анализ** на вкладке **Данные** становится доступна команда **Пакет анализа**.

Для анализа данных с помощью этого пакета следует указать входные данные и выбрать параметры. Расчет будет выполнен с помощью подходящей статистической или инженерной макрофункции, а результат – помещен в выходной диапазон. Некоторые инструменты позволяют представить результаты анализа в графическом виде.

Для выполнения зачетных заданий по эконометрике из **Пакета анализа** необходимо использовать инструменты **Корреляция** и **Регрессия**.

Форматы чисел в MS Excel


Применяя различные числовые форматы, можно изменить формат числа, не изменяя само число. Числовой


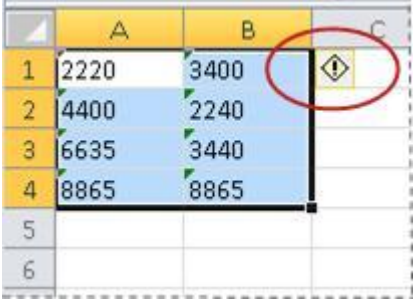
² В более ранних версиях в MS Office для запуска **Анализа данных** необходимо выбрать команду **Сервис⇒Надстройки** и активизировать надстройку **Поиск решения**.

Если после изменения числового формата в ячейке Microsoft Excel отображаются символы #####, то вероятно, что ширина ячейки недостаточна для отображения данных. Чтобы увеличить ширину ячейки, дважды щелкните правую границу столбца, содержащего ячейки с ошибкой #####. Размер столбца автоматически изменится таким образом, чтобы отобразить число. Кроме того, можно перетаскивать правую границу столбца, увеличив его ширину.

Чаще всего числовые данные отображаются правильно независимо от того, вводятся ли они в таблицу вручную или импортируются из базы данных или другого внешнего источника. Однако иногда MS Excel применяет к данным неправильный числовой формат, из-за чего приходится изменять некоторые настройки.

Например, при копировании данных из MS Word к числовым данным может быть применен текстовый формат, что впоследствии приводит к проблемам при вычислениях или нарушению порядка сортировки. Индикатором данной ошибки служит маленький зеленый треугольник в левом верхнем углу ячейки. Для ее исправления:

Excel 2003	Excel 2010
<ol style="list-style-type: none"> 1. В меню Сервис выберите команду Параметры и откройте вкладку Контроль ошибок. 2. Убедитесь, что установлены флажки Включить фоновую проверку ошибок и Число сохранено как текст. 3. Выделите ячейки с зеленым индикатором ошибки в верхнем левом углу . 4. Нажмите на появившуюся 	<ol style="list-style-type: none"> 1. Выделите любую ячейку или диапазон смежных ячеек с индикатором ошибки в верхнем левом углу. 2. Нажмите появившуюся рядом с выделенной ячейкой или диапазоном ячеек кнопку ошибки.

Excel 2003	Excel 2010
<p>рядом с ячейкой кнопку  и выберите команду Преобразовать в число.</p>	 <p>3. Выберите в меню пункт Преобразовать в число.</p>

При выполнении вычислений ошибка также может возникнуть, если при вводе чисел использовать разделитель целой и дробной части, отличный от принятого в системе. Как правило, для разделения целой и дробной частей используются «.» или «,». Определить, какой разделитель используется, можно щелкнув по ярлыку **Язык и региональные стандарты** в окне **Панель управления Windows**.

Копирование листов Microsoft Excel

Лист с исходными данными не следует изменять ни при каких обстоятельствах. Для выполнения различных расчетов или тестов нужно сначала скопировать данные, а затем уже на листе-копии произвести необходимые изменения.

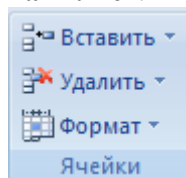
Наиболее удобным способом является копирование целого листа. Для этого:

Excel 2003	Excel 2010
<ol style="list-style-type: none"> 1. В меню Правка выберите команду Переместить/скопировать лист... 2. В диалоговом окне Пере- 	<ol style="list-style-type: none"> 1. На вкладке Главная в группе Ячейки нажмите кнопку Формат и в разделе Упорядочить листы выберите пункт

местить или скопировать в списке **Перед листом** укажите, куда необходимо скопировать лист, установите флажок **Создать копию**.

3. Чтобы закончить и вернуться на лист, нажмите кнопку **ОК**.

Переместить или скопировать лист.



2. В диалоговом окне **Переместить или скопировать** в списке **Перед листом** укажите, куда необходимо скопировать лист, установите флажок **Создать копию**.

3. Чтобы закончить и вернуться на лист, нажмите кнопку **ОК**.

Можно, конечно, не создавать копию целого листа, а скопировать сами данные на новый лист, однако при этом придется заново изменять ширину столбцов, если она была специально настроена на листе с исходными данными, для того чтобы нагляднее отобразить их.

Создание парных диаграмм рассеяния

Важнейшим элементом эконометрического исследования является графический анализ исходных данных. В случае множественной регрессии, то есть когда у нас несколько показателей, необходимо построить парные диаграммы зависимости объясняемой переменной Y от каждой из объясняющих переменных X – диаграммы рассеяния.

В эконометрическом анализе диаграммы, используемые при построении регрессионной модели, имеют тип «Точечная».

Для создания диаграммы рассеяния нужно выделить два столбца данных со значениями показателей, включая

их названия (метки) в первой строке матрицы данных, и выполнить следующие действия:

Excel 2003	Excel 2010
<ol style="list-style-type: none"> 1. В меню Вставка выберите команду Диаграмма.... 2. Следуйте инструкциям в раскрывшемся диалоговом окне Мастер диаграмм. 	<ol style="list-style-type: none"> 1. На вкладке Вставка в группе Диаграммы выберите тип и вид диаграммы, диаграмма добавится на лист. 2. Щелкните в любом месте внедренной диаграммы, чтобы активизировать ее. Откроется панель Работа с диаграммами с дополнительными вкладками Конструктор, Макет и Формат. 3. Выберите необходимые команды для оформления диаграммы.

Нужно помнить, что для того чтобы MS Excel правильно определил переменные, объясняемая переменная Y должна быть расположена в правом из двух выделенных столбцов, а объясняющая переменная X – в левом.

Для добавления данных в уже построенную диаграмму необходимо предварительно внести их в исходную таблицу, а затем выполнить следующие действия:

Excel 2003	Excel 2010
<ol style="list-style-type: none"> 1. Щелкните правой кнопкой мыши на области диаграммы. 2. В раскрывшемся контекстном меню выберите команду Исходные данные. 3. В окне Исходные данные перейдите на вкладку Ряд и укажите новые диапазоны данных для значений Y и X. 	<ol style="list-style-type: none"> 1. Щелкните правой кнопкой мыши на области диаграммы. 2. В раскрывшемся контекстном меню выберите команду Выбрать данные. 3. В окне Выбор источника данных в поле Элементы легенды (ряды) выделите нужный ряд и нажмите кнопку

Excel 2003	Excel 2010
4. Чтобы закончить и вернуться на лист, нажмите кнопку ОК .	Изменить. 4. В окне Изменение ряда укажите новые диапазоны данных для значений Y и X. 5. Чтобы закончить и вернуться на лист, последовательно нажимайте кнопку ОК .

Регрессионная статистика в отчете Excel

Наименование в отчете Excel	Принятое наименование	Формула
Множественный R	Коэффициент множественной корреляции, индекс корреляции	$R = \sqrt{R^2}$
R-квадрат	Коэффициент детерминации	$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Нормированный R-квадрат	Скорректированный коэффициент детерминации	$R^2_{\text{скадр}} = 1 - (1 - R^2) \frac{n - 1}{n - m - 1}$
Стандартная ошибка	Среднеквадратическое отклонение от модели	$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - m - 1}}$

Дисперсионный анализ в отчете Excel

Наименование в отчете Excel	Df – число степеней свободы	SS – сумма квадратов	MS – дисперсия на одну степень свободы	F-критерий Фишера
Регрессия	m	$RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$.	$\frac{RSS}{m}$	$F_{расч} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{(n-m-1)}}$.
Остаток	$n - m - 1$	$ESS = \sum_{i=1}^n e_i^2$.	$\frac{ESS}{n - m - 1}$	
Итого	$n - 1$	$TSS = \sum_{i=1}^n (y_i - \bar{y})^2$		

Названия некоторых функций в Excel 2010 были изменены по сравнению с более ранними версиями.

Чтобы повысить точность работы функций MS Excel, обеспечить их согласованность и привести имена функций в соответствии с их назначением, корпорация Microsoft изменила, переименовала и добавила несколько функций в библиотеку MS Excel 2010.

Для обеспечения обратной совместимости переименованные функции доступны также и по их старым именам.

<i>Название функции в Excel более ранних версий</i>	<i>Название функции в Excel 2010</i>	<i>Примечания</i>
ДИСП(число1,[число2],...)	ДИСП.В(число1,[число2],...)	Оценивает дисперсию по выборке
СТЮДРАС-ПОБР(вероятность; степени_свободы)	СТЪДЕНТ.ОБР.2Х(вероятность, степени_свободы)	Возвращает двустороннее обратное t -распределение Стьюдента
ФРАСПОБР(вероятность; степени_свободы1; степени_свободы2)	Ф.ОБР.ПХ(вероятность, степени_свободы1, степени_свободы2)	Возвращает значение, обратное (правостороннему) F -распределению вероятностей
ХИ2ОБР(вероятность, степени_свободы)	ХИ2.ОБР.ПХ(вероятность, степени_свободы)	Возвращает обратное значение односторонней вероятности распределения хи-квадрат

Решение типовых задач

Задача 1. Имеются выборочные данные показателей «Объем инвестиций» (x , млн руб.) и «Цена за акцию» (y , руб.):

№ наблюдения	Объем инвестиций	Цена за акцию
1	108	12
2	4,4	4
3	3,5	5
4	3,6	6
5	39	13
6	68,4	19
7	7,5	8,5
8	5,5	5
9	37,5	15
10	12	6
11	51	12

Требуется:

- 1) оценить тесноту линейной корреляционно-регрессионной зависимости;
- 2) оценить статистическую значимость коэффициента корреляции.

Решение.

Объем выборки $n = 11$.

- 1) Теснота линейной корреляционно-регрессионной зависимости оценим с помощью коэффициента парной корреляции:

$$r_{xy} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y}.$$

Составим расчетную таблицу в Excel:

	A	B	C	D	E	F	G
№ наблюдения	Объем инвестиций, x	Цена за акцию, y	xу	$(x_i - x_{cp})^2$	$(y_i - y_{cp})^2$	x^2	
1	1,00	108,00	12,00	1296,00	5937,40	5,80	11664,00
2	2,00	4,40	4,00	17,60	704,66	31,26	19,36
3	3,00	3,50	5,00	17,50	753,25	21,08	12,25
4	4,00	3,60	6,00	21,60	747,77	12,89	12,96
5	5,00	39,00	13,00	507,00	64,88	11,62	1521,00
6	6,00	68,40	19,00	1299,60	1402,84	88,53	4678,56
7	7,00	7,50	8,50	63,75	549,69	1,19	56,25
8	8,00	5,50	5,00	27,50	647,47	21,08	30,25
9	9,00	37,50	15,00	562,50	42,96	29,26	1406,25
10	10,00	12,00	6,00	72,00	358,93	12,89	144,00
11	11,00	51,00	12,00	612,00	402,18	5,80	2601,00
12	Сумма	340,40	105,50	4497,05	11612,05	241,41	22145,88
13	Среднее	30,95	9,59	408,82			2013,26
14							
15							
16	σ_x	32,49		r_{xy}	0,74		
17	σ_y	4,68		Критерий Стьюдента для r_{xy}			
18				$ t_{расч} $	3,26		
19				$t_{кр}$	2,26		
20				Вывод	значим		

Введены следующие формулы:

чей ка	Формула	Примечание
2	=B2*C2	Копируем в диапазон D3:D12
13	=СУММ(B2:B12)	Сумма значений x
14	=СРЗНАЧ(B2:B12)	\bar{x}
13	=СУММ(C2:C12)	Сумма значений y
14	=СРЗНАЧ(C2:C12)	\bar{y}

чей ка	Формула	Примечание
2	=(B2-\$B\$14)^2	Копируем в диапазон E3:E12
2	=(C2-\$C\$14)^2	Копируем в диапазон F3:F12
13	=СУММ(E2:E12)	$\sum_{i=1}^n (x_i - \bar{x})^2$.
13	=СУММ(F2:F12)	$\sum_{i=1}^n (y_i - \bar{y})^2$.
16	=(E13/11)^0,5	σ_x
17	=(F13/11)^0,5	σ_y
16	=(D14- B14*C14)/(B16*B17)	Коэффициент парной корреляции

Таким образом, получены следующие результаты

$$\bar{x} = \frac{1}{11} \sum_{i=1}^n x_i = \frac{340,40}{11} = 30,95;$$

$$\bar{y} = \frac{1}{11} \sum_{i=1}^n y_i = \frac{105,50}{11} = 9,59;$$

$$\overline{yx} = \frac{1}{11} \sum_{i=1}^n (x_i \cdot y_i) = \frac{4497,05}{11} = 408,82;$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{1}{11} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{11612,05}{11}} = 32,49;$$

$$\sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \sqrt{\frac{241,41}{11}} = 4,68;$$

$$r_{xy} = \frac{408,82 - 30,95 \cdot 9,59}{32,49 \cdot 4,68} = 0,74.$$

Найдем расчетное значение (статистика) критерия Стьюдента: $t_{\text{расч}} = 3,26$.

Табличное (критическое) значение равно $t_{\text{табл}}(0,05; 11 - 2) = \text{СТЮДРАСПОБР}(0,05;9)=2,26$.

Так как $|t_{\text{расч}}| > t_{\text{табл}}$ ($3,26 > 2,26$) значение коэффициента парной корреляции статистически значимо.

На основе проведенного анализа можно сделать следующий вывод о тесноте линейной корреляционно-регрессионной зависимости:

Величина статистически значимого коэффициента парной корреляции свидетельствует о прямой сильной (тесной) линейной связи между объемом инвестиций и ценой за акцию.

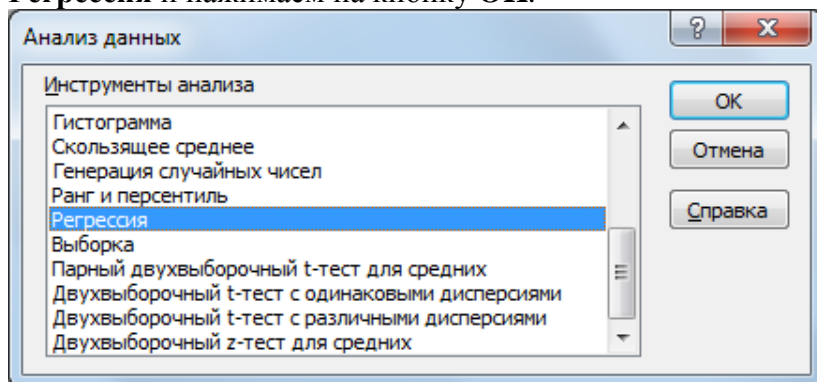
Задача 2. Для данных из задачи 1:

- 1) найти уравнение парной линейной регрессии и построить его на корреляционном поле;
- 2) записать модель парной линейной регрессии;
- 3) оценить качество уравнения регрессии;
- 4) построить уравнение регрессии на корреляционном поле.

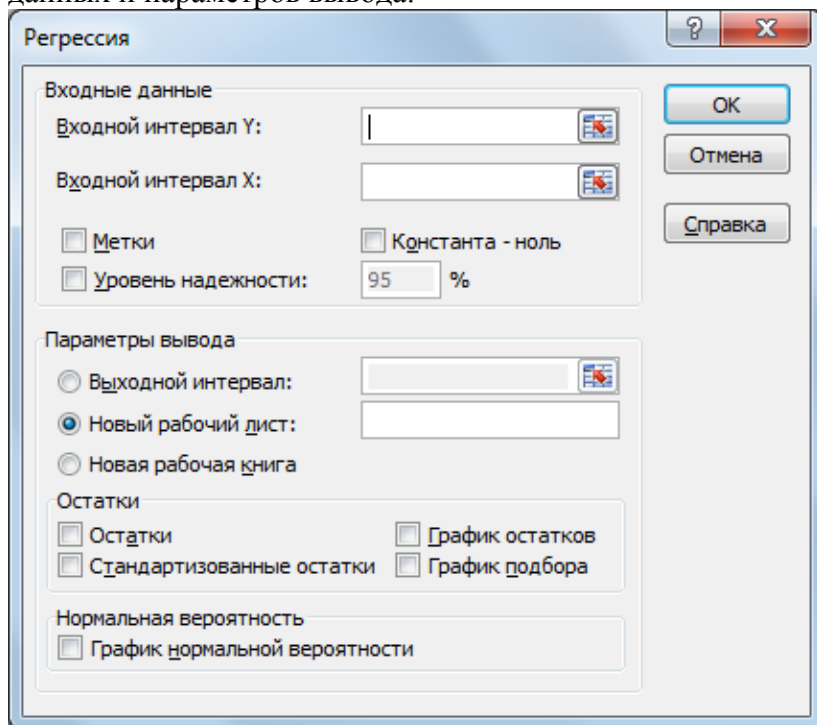
Решение.

1) Найдем уравнение парной линейной регрессии с помощью инструмента анализа данных «Регрессия». Переходим на вкладку **Данные**, вызываем команду **Пакет анализа**.

В появившемся диалоговом окне выбираем режим **Регрессия** и нажимаем на кнопку **ОК**.



Далее необходимо заполнить диалоговое окно ввода данных и параметров вывода:



- ✓ *Входной интервал Y* – диапазон адресов ячеек, содержащих значения резульативного признака (ячейки должны составлять один столбец).
- ✓ *Входной интервал X* – диапазон адресов ячеек, содержащих значения независимых переменных. Значения каждой переменной представляются одним столбцом. Количество независимых переменных $t \leq 16$.
- ✓ *Метки* – флажок, указывающий содержит ли первая строка названия столбцов или нет.
- ✓ *Уровень надежности* – флажок, задающий надежность γ при построении доверительных интервалов. По умолчанию принимается равным 95% (т.е. уровень значимости $\alpha = 0,05$).
- ✓ *Константа-ноль* – флажок, указывающий на наличие или отсутствие ($b_0 = 0$) свободного коэффициента в уравнении.
- ✓ *Выходной интервал* – при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который содержит ячейки с результатами вычислений режима *Регрессия*.
- ✓ *Новый рабочий лист* – при включении этого параметра открывается новый лист, в который начиная с ячейки A1 вставляются результаты работы режима *Регрессия*.
- ✓ *Новая рабочая книга* - при включении этого параметра открывается новая книга на первом листе которой начиная с ячейки A1 вставляются результаты работы режима *Регрессия*.
- ✓ *Остатки* – флажок, при включении которого вычисляется столбец, содержащий ошибки $\varepsilon_i = y_i - \hat{y}_i, i = 1, 2, \dots, n$.

- ✓ *Стандартизованные остатки* – флажок, при включении которого вычисляется столбец, содержащий стандартизованные остатки.
- ✓ *График остатков* – флажок, при включении которого выводятся точечные графики ε_i в зависимости от значений переменных x_j , $j = 1, 2, \dots, t$. Количество графиков равно числу t переменных x_j .
- ✓ *График подбора* – при включении выводятся точечные графики предсказанных по построенной регрессии значений \hat{y}_i от значений переменных x_j , $j = 1, 2, \dots, t$. Количество графиков равно числу t переменных x_j .

Для нашего примера:

	A	B	C	D	E	F	G	H	I	J
	№ наблюдения	Объем инвестиций, x	Цена за акцию, y							
1										
2	1,00	108,00	12,00							
3	2,00	4,40	4,00							
4	3,00	3,50	5,00							
5	4,00	3,60	6,00							
6	5,00	39,00	13,00							
7	6,00	68,40	19,00							
8	7,00	7,50	8,50							
9	8,00	5,50	5,00							
10	9,00	37,50	15,00							
11	10,00	12,00	6,00							
12	11,00	51,00	12,00							
13	Сумма	340,40	105,50							
14	Среднее	30,95	9,59							
15										
16	σ_x	32,49								
17	σ_y	4,68								

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

После нажатия кнопки **ОК**, в книге появится новый рабочий лист с данными регрессионного анализа:

	A	B	C						
1	ВЫВОД ИТОГОВ								
2									
3	<i>Регрессионная статистика</i>								
4	Множественный	0,7360153							
5	R-квадрат	0,541718522							
6	Нормированный								
7	R-квадрат	0,490798358							
8	Стандартная								
9	ошибка	3,506079085							
10	Наблюдения	11							
11									
12	Дисперсионный анализ								
13		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
14	Регрессия	1	130,775776	130,7758	10,63859	0,0098123			
15	Остаток	9	110,6333149	12,29259					
16	Итого	10	241,4090909						
17									
18		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
19	Y-пересечен	6,3068866	1,459880942	4,320138	0,001933	3,0044065	9,6093667	3,00440648	9,60936674
20	Объем								
21	инвестиции								
22	y, x	0,1061229	0,032536236	3,261684	0,009812	0,0325209	0,179725	0,03252086	0,17972502

	A	B	C	D
22	ВЫВОД ОСТАТКА			
23				
24	<i>Наблюдение</i>	<i>Предсказанное Цена за акцию, y</i>	<i>Остатки</i>	
25	1	17,76816366	-5,768163659	
26	2	6,773827526	-2,773827526	
27	3	6,678316884	-1,678316884	
28	4	6,688929178	-0,688929178	
29	5	10,4456811	2,5543189	
30	6	13,56569541	5,434304592	
31	7	7,102808627	1,397191373	
32	8	6,890562756	-1,890562756	
33	9	10,2864967	4,713503304	
34	10	7,580361837	-1,580361837	
35	11	11,71915633	0,280843672	
36				

Для удобства интерпретации полученных значений следует выбрать формат ячеек **числовой**, **число десятичных разрядов задать равным 4**.

Из полученных данных следует, что **уравнение парной линейной регрессии** имеет вид:

$$\hat{y} = 6,3069 + 0,1061 \cdot x$$

Параметр регрессии $b_1 = 0,1061$ показывает, что с увеличением объема инвестиций на 1 млн. руб. цена за акцию в среднем возрастает на 11 коп.

2) Модель парной линейной регрессии имеет вид

$$y = 6,3069 + 0,1061 \cdot x + \varepsilon$$

или

$$y_i = 6,3069 + 0,1061 \cdot x_i + \varepsilon_i$$

3) Оценим качество уравнения регрессии.

3.1) Значение коэффициента детерминации возьмем из таблицы **Регрессионная статистика** протокола Excel (строка R-квадрат):

$$R^2 = 0,5417$$

Коэффициент детерминации показывает долю вариации резульативного признака под воздействием изучаемых факторов.

Следовательно, около 54,17% вариации зависимой переменной учтено в модели и обусловлено влиянием факторов, включенных в модель.

3.2) Найдем среднюю относительную ошибку аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Составим расчетную таблицу:

	A	B	C	D	E	F
	№ наблюдения	Объем инвестиций, x	Цена за акцию, y	y_x	Остатки, ε_i	A_i
1	1,00	108,00	12,00	17,7682	-5,7682	=ABS(E2/C2)
2	2,00	4,40	4,00	6,7738	-2,7738	0,693457
3	3,00	3,50	5,00	6,6783	-1,6783	0,335663
4	4,00	3,60	6,00	6,6889	-0,6889	0,114822
5	5,00	39,00	13,00	10,4457	2,5543	0,196486
6	6,00	68,40	19,00	13,5657	5,4343	0,286016
7	7,00	7,50	8,50	7,1028	1,3972	0,164375
8	8,00	5,50	5,00	6,8906	-1,8906	0,378113
9	9,00	37,50	15,00	10,2865	4,7135	0,314234
10	10,00	12,00	6,00	7,5804	-1,5804	0,263394
11	11,00	51,00	12,00	11,7192	0,2808	0,023404
12	Сумма	340,40	105,50	105,50	0,00	3,25

В столбце y_x будут записаны теоретические значения \hat{y}_i , просчитанные в столбце **Предсказанное y** таблицы **Вывод остатка** протокола Excel. Значения $\varepsilon_i = y_i - \hat{y}_i$ уже

просчитаны в столбце **Остатки** таблицы **Вывод остатка** протокола Excel. Поэтому, скопируем данные столбцы в расчетную таблицу. В столбце A_i посчитаем $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$.

$$\bar{A} = \frac{1}{11} \cdot 3,25 \cdot 100\% = 29,55\%$$

Так как $20\% < \bar{A} < 50\%$, *уравнение имеет удовлетворительную точность*. Другими словами, фактические значения цены на акцию отличаются от расчетных в среднем на 29,55%.

3.3) Проверим статистическую значимость уравнения регрессии в целом с помощью критерия Фишера. Расчетное значения критерия $F_{\text{расч}}$ находим в таблице **Дисперсионный анализ** протокола Excel (столбец F):

$$F_{\text{расч}} = 10,6386$$

Табличное значение с $df_1 = m = 1$ и $df_2 = n - m - 1 = 11 - 1 - 1 = 9$ степенями свободы при уровне значимости $\alpha = 0,05$ найдем с помощью встроенной функции Excel «ФРАСПОБР»:

$$F_{\text{табл}}(0,05, 1, 10) = 5,1174$$

Поскольку $F_{\text{расч}} > F_{\text{табл}}$ ($10,6386 > 5,1174$), *уравнение парной линейной регрессии статистически значимо в целом*, оно адекватно описывает исходные данные.

3.4) Проверим статистическую значимость параметров уравнения регрессии с помощью t -критерия Стьюдента. Расчетные значения критерия возьмем из таблицы **Дисперсионный анализ** протокола Excel (столбец t -статистика):

$$t_{b_0} = 4,3201; t_{b_1} = 3,2617$$

Табличное значение критерия Стьюдента

$$t_{\text{табл}}(0,05; 11 - 2) \text{ СТЬЮДРАСПОБР}(0,05; 9) = 2,26.$$

Так как:

$$|t_{b_0}| > t_{\text{табл}} - \text{параметр } b_0 \text{ статистически значим;}$$

$|t_{b_1}| > t_{\text{табл}}$ - параметр b_1 статистически значим.

Замечание. Есть более простой способ проверить значимость коэффициентов регрессии, не требующий дополнительного вычисления табличного значения критерия Стьюдента.

Если Р-значение t-статистики Стьюдента для параметра регрессии b_i меньше, чем заданный уровень значимости α , то данный параметр статистически значим.

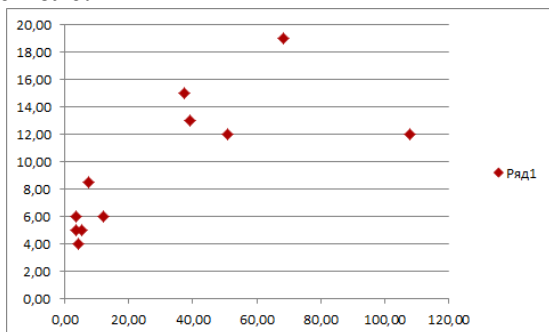
В нашем случае получаем, что

Р-значение (b_0)=0,0019<0,05;

Р-значение (b_1)=0,0098<0,05.

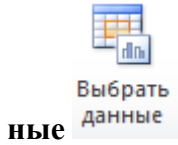
Следовательно, параметры b_0 и b_1 значимы при 5%-ном уровне значимости.

4) Для построения графика уравнения регрессии воспользуемся инструментом **Мастер диаграмм** Excel. Выделим диапазон ячеек **В2:С12** . Вызовем **Мастер диаграмм**. Выберем тип диаграммы **Точечная**. Появится корреляционное поле:

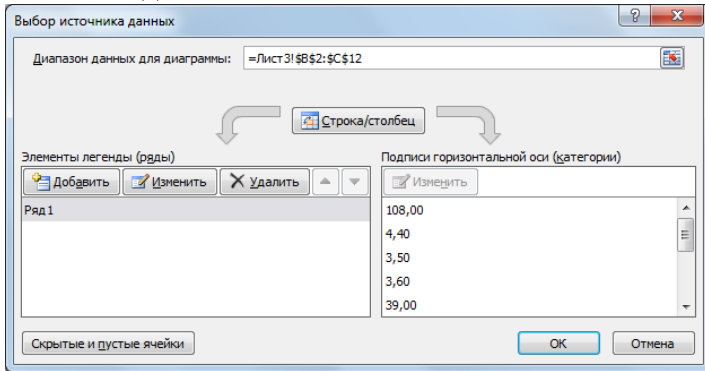


Для добавления на диаграмму уравнения регрессии необходимо выделить данную диаграмму, перейти на

вкладку **Конструктор** и вызвать команду **Выбрать дан-**

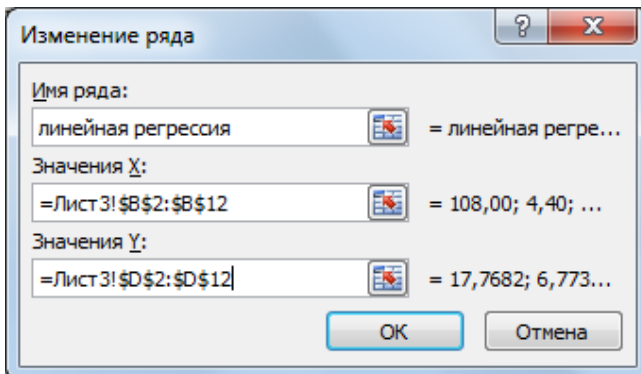


Появится диалоговое окно:

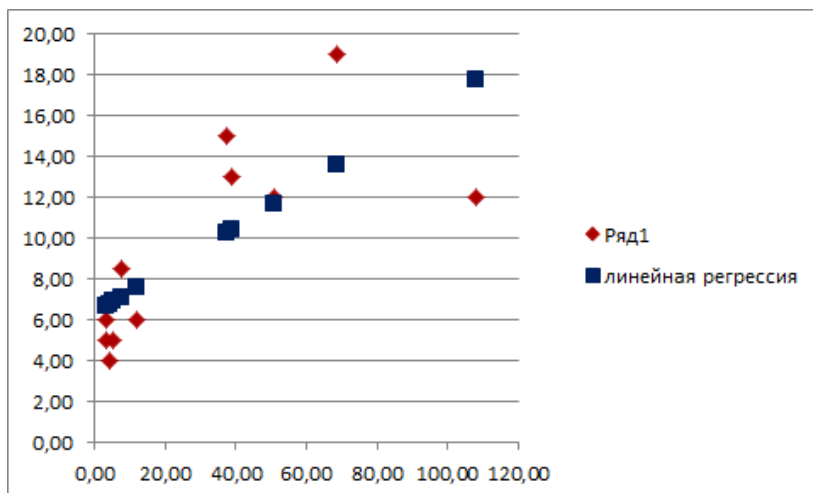


Нажать на кнопку **Добавить** .

Ввести параметры ряда (за значения **Y** взять столбец Y_x).



Нажать кнопку **OK**.



Задача 3. Изучается зависимость величины накладных расходов Y (млн. руб.) строительных организаций от следующих факторов: X_1 – объем выполненных работ, млн руб.; X_2 – численность рабочих, чел.; X_3 – фонд заработной платы, млн руб.

№ компании	Y	X_1	X_2	X_3
1	5,7	26,9	1 276	12,250
2	5,0	24,5	975	10,627
3	4,5	18,4	869	6,865
4	4,0	18,1	757	6,964
5	4,4	18,1	740	7,622
6	3,5	17,9	699	6,291
7	3,5	15,7	840	7,980
8	3,8	14,2	744	6,770
9	5,1	13,3	725	7,105
10	3,4	15,0	670	5,762
11	4,1	14,7	622	6,096
12	4,1	13,3	566	6,056

№ компании	Y	X ₁	X ₂	X ₃
13	3,1	14,6	518	4,921
14	2,8	11,7	510	4,131
15	2,1	10,6	452	4,384
16	2,5	10,0	447	4,157
17	2,0	9,0	497	4,324
18	2,4	9,5	428	4,023
19	2,3	7,0	381	3,315
20	2,4	9,1	385	3,619
21	2,5	6,8	412	3,461
22	2,2	5,5	293	2,139
23	1,6	5,1	284	2,244
24	3,4	12,2	514	3,958
25	2,7	11,0	407	3,337
26	3,2	9,3	577	3,676
27	2,9	5,9	265	2,120
28	4,8	25,9	977	10,649
29	3,7	23,5	724	6,806
30	4,4	19,8	983	9,240
31	3,7	18,8	828	8,860
32	4,8	19,1	766	7,354
33	3,7	18,8	615	5,289
34	3,6	17,4	583	5,830
35	4,0	14,1	591	6,265
36	3,8	13,8	593	5,396
37	3,7	13,7	611	5,194
38	4,1	13,8	562	4,608
39	2,4	13,9	488	5,856
40	2,5	10,6	740	7,326

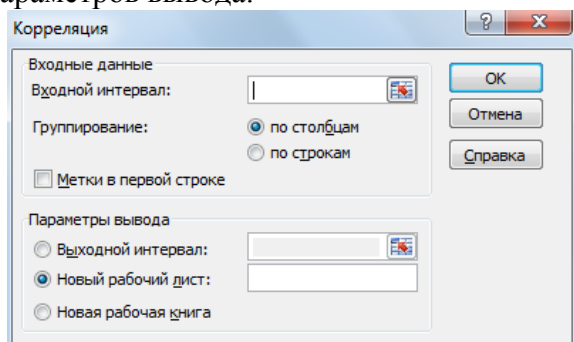
Требуется:

- 1) Построить матрицу парных коэффициентов корреляции. Отобразить факторы для регрессионного анализа.
- 2) Построить уравнение регрессии. Оценить качество уравнения регрессии.
- 3) Рассчитать прогнозное значение результата, если прогнозные значения фактора(ов) составляет(ют) 109% от их среднего уровня.

Решение.

1) Построим матрицу парных коэффициентов корреляции. Переходим на вкладку **Данные**, вызываем команду **Пакет анализа**. В появившемся диалоговом окне выбираем режим **Корреляция** и нажимаем на кнопку **ОК**.

Далее необходимо заполнить диалоговое окно ввода данных и параметров вывода:



- ✓ *Входной интервал* – вводится диапазон адресов ячеек, содержащих значения Y и X_i .
- ✓ *Метки* – флажок, указывающий содержит ли первая строка названия столбцов или нет.
- ✓ *Выходной интервал* – при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который содержит ячейки с результатами вычислений режима *Корреляция*.

- ✓ *Новый рабочий лист* – при включении этого параметра открывается новый лист, в который начиная с ячейки А1 вставляются результаты работы режима *Корреляция*.
- ✓ *Новая рабочая книга* - при включении этого параметра открывается новая книга на первом листе которой начиная с ячейки А1 вставляются результаты работы режима *Корреляция*.
- ✓ В результате будет получена матрица коэффициентов парной корреляции.

Для нашего примера:

	Y	X1	X2	X3
Y	1			
X1	0,814887	1		
X2	0,82489	0,884877	1	
X3	0,795868	0,893143	0,963281	1

Проанализируем полученную матрицу.

Первый столбец матрицы (Y): Имеется тесная связь между Y и X₁ ($r_{YX_1} = 0,814887$), Y и X₂ ($r_{YX_2} = 0,82489$), Y и X₃ ($r_{YX_3} = 0,795868$).

Проанализируем оставшиеся столбцы матрицы для выявления коллинеарных факторов:

$$r_{X_1X_2} = 0,884877; r_{X_1X_3} = 0,893143; r_{X_2X_3} = 0,963281$$

На основе коэффициентов парной корреляции можно сделать вывод, что наиболее тесно связаны друг с другом факторы X₂ и X₃.

Исследуем **интеркорреляцию** переменных X_2 и X_3 .

Напомним, что две переменные x_a и x_b интеркоррелированы, если одновременно выполняются следующие условия:

$$|r_{x_a x_b}| > 0,75; |r_{y x_a}| < |r_{x_a x_b}|; |r_{y x_b}| < |r_{x_a x_b}|.$$

Факторы X_2 и X_3 интеркоррелируемы друг с другом. Из этих двух факторов оставляет X_2 , т.к. - $|r_{Y X_2}| > |r_{Y X_3}|$.

Исследуем теперь **интеркорреляцию** переменных X_2 и X_1 .

$$r_{X_1 X_2} = 0,884877; r_{Y X_1} = 0,814887; r_{Y X_2} = 0,82489.$$

Следовательно, данные переменные интеркоррелированы. Из этих двух факторов оставляет X_2 , т.к. - $|r_{Y X_2}| > |r_{Y X_1}|$.

Таким образом, на основе анализа корреляционной матрицы остается один фактор X_2 – численность рабочих.

Но для полной уверенности в правильности выбора сравним (с помощью инструмента **Регрессия** пакета **Анализ данных** Excel) три модели (одна – по трем факторам (X_1, X_2, X_3), вторая – по двум (X_1, X_2), третья – по одному (X_2)) по $R^2_{\text{скаорр}}$ и $F_{\text{расч}}$.

Модель регрессии по трем факторам:

Вывод итогов				
<i>Регрессионная статистика</i>				
Множественный R	0,846793232			
R-квадрат	0,717058779			
Нормированный R-квадрат	0,693480343			
Стандартная ошибка	0,538209007			
Наблюдения	40			
<i>Дисперсионный анализ</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Регрессия	3	26,42791834	8,809306114	30,41163567
Остаток	36	10,42808166	0,289668935	
Итого	39	36,856		
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	1,04762696	0,285362505	3,671214475	0,000777362
X1	0,07732364	0,035835053	2,15776549	0,037693985
X2	0,002984686	0,001524607	1,957675668	0,058054599
X3	-0,094414971	0,14462092	-0,652844492	0,518004213

Модель регрессии по двум факторам:

Вывод итогов				
<i>Регрессионная статистика</i>				
Множественный R	0,84481301			
R-квадрат	0,713709022			
Нормированный R-квадрат	0,698233834			
Стандартная ошибка	0,534019442			
Наблюдения	40			
<i>Дисперсионный анализ</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Регрессия	2	26,30445972	13,15222986	46,11957044
Остаток	37	10,55154028	0,285176764	
Итого	39	36,856		
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	1,118496648	0,261853833	4,271454174	0,000130147
X1	0,069699818	0,033615161	2,073463751	0,045145502
X2	0,002163018	0,000853743	2,533569666	0,015658404

Модель регрессии по одному фактору:

Вывод итогов				
<i>Регрессионная статистика</i>				
Множественный R	0,824889784			
R-квадрат	0,680443156			
Нормированный R-квадрат	0,672033766			
Стандартная ошибка	0,556719409			
Наблюдения	40			
<i>Дисперсионный анализ</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Регрессия	1	25,07841297	25,07841297	80,91468063
Остаток	38	11,77758703	0,309936501	
Итого	39	36,856		
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	1,137123588	0,272823954	4,167975609	0,000171014
X2	0,003729432	0,0004146	8,995258786	5,94628E-11

Составляем сводную таблицу:

Мо- дель	Перемен- ные	$R^2_{\text{свопп}}$	$F_{\text{расч}}$	$F_{\text{табл}}$
А	X_1, X_2, X_3	0,693480343	30,41163567	2,866265551
В	X_1, X_2	0,698233834	46,11957044	3,251923846
С	X_2	0,672033766	80,91468063	4,098171731

Модель А хуже, чем модель В по обоим сравниваемым параметрам: $R^2_{\text{свопп}}$ увеличилось, следовательно, X_3 не оказывает влияние на Y . $F_{\text{расч}}$ для второй модели также выше, чем для первой. Следовательно, исключение из модели фактора X_3 было правильным.

Сравним теперь модели В и С. Около **69,82%** вариации зависимой переменной учтено в **модели В** и обусловлено влиянием X_1 и X_2 . Для **модели С** – **67,20%** вариации зависимой переменной учтено в модели и обусловлено влиянием фактора X_2 . Таким образом, разница составляет приблизительно 2,62% и может считаться незначительной. Значимость уравнения линейной регрессии (модель С) на порядок выше, чем множественной (модель В).

Дополнительным аргументом за исключение фактора X_1 является проверка значимости параметра b_1 :

P-значение (b_1)=0,045<0,05

Таким образом, параметр является слабо значимым. Делаем вывод, что **модель С** наилучшая.

2) Уравнение регрессии имеет вид:

$$\hat{y} = 1,1371 + 0,0037x_2$$

Интерпретация параметров уравнения: с увеличением численности рабочих в среднем на 1 чел., накладные расходы в среднем возрастают на 0,0037 млн руб.

Коэффициент детерминации $R^2 = 0,6804$ показывает, что около 68,04% вариации зависимой переменной учтено в модели и обусловлено влиянием фактора, включенного в модель.

Математическую точность модель оценим с помощью средней относительной ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% = \frac{1}{40} \cdot 5,438 \cdot 100\% = 13,595\%$$

Так как $10\% < \bar{A} < 20\%$, точность модели хорошая.

Уравнение является статистически значимым в целом ($F_{\text{расч}} = 80,91468063 > F_{\text{табл}} = 4,098171731$).

Параметры уравнения статистически значимы:

P-значение (b_0)=0,00017<0,05;

P-значение (b_1)=0,00000<0,05.

3) Найдем **прогнозное значение признака у** при прогнозном значении X_2 , составляющем 109% от его среднего уровня.

$$x_{\text{пр}} = \frac{109}{100} \cdot \bar{X}_2 = 1,09 \cdot 622,85 = 678,91 \cong 679 \text{ чел.}$$

Прогнозное значение накладных расходов Y :

$$\begin{aligned} \hat{y}_{\text{пр}} &= 1,1371 + 0,0037x_{\text{пр}} = 1,1371 + 0,0037 \cdot 679 \\ &= 3,6494 \text{ млн руб.} \end{aligned}$$

Интервальная оценка прогноза $y_{\text{пр}}$ имеет вид:

$$\hat{y}_{\text{пр}} - t_{\text{табл}} \cdot m_{\hat{y}_{\text{пр}}} \leq y_{\text{пр}} \leq \hat{y}_{\text{пр}} + t_{\text{табл}} \cdot m_{\hat{y}_{\text{пр}}}$$

Средняя стандартная ошибка прогноза равна:

$$m_{\hat{y}_{\text{пр}}} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - 2} \cdot \left(1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}$$

$$= \sqrt{\frac{11,78}{40 - 2} \cdot \left(1 + \frac{1}{40} + \frac{(679 - 622,85)^2}{1803079,10}\right)}$$

$$= 0,564116$$

$$t_{\text{табл}}(0,05; 40 - 2) = 2,02439$$

$$3,6494 - 2,02439 \cdot 0,564116 \leq y_{\text{пр}}$$

$$\leq 3,6494 + 2,02439 \cdot 0,564116$$

$$2,5071 \leq y_{\text{пр}} \leq 4,7914$$

Полученный доверительный интервал узкий – маленькая ошибка прогноза.

Задача 4. По 14 страховым компаниям имеются данные, характеризующие зависимость объема чистой годовой прибыли от годовых объемов собственных средств, страховых резервов, страховых премий и страховых выплат, тыс. руб.:

№ компании	Годовая прибыль, Y	Собственные средства, X_1	Страховые резервы, X_2
1	92	3 444	9 563
2	42	2 658	6 354
3	186	9 723	10 245
4	48	4 526	6 398
5	38	5 369	5 692
6	74	2 248	6 359
7	48	5 671	6 892
8	82	4 312	7 256
9	45	2 226	8 256
10	46	3 654	5 982

№ ком- пании	Годовая прибыль, Y	Собственные средства, X ₁	Страховые резервы, X ₂
11	65	2 635	6 359
12	29	2 463	7 532
13	34	3 265	5 632
14	66	7 546	7 625

Требуется:

- 1) Построить множественную регрессионную модель. Оцените параметры модели.
- 2) Оцените силу связи факторов с результатом с помощью коэффициентов эластичности.

Решение.

1) Уравнение множественной линейной регрессии имеет вид:

$$\hat{y} = b_0 + b_1X_1 + b_2X_2.$$

Оценим параметры регрессии с помощью инструмента **Регрессия** пакета **Анализ данных** Excel:

Вывод итогов				
<i>Регрессионная статистика</i>				
Множественный R	0,836084202			
R-квадрат	0,699036794			
Нормированный R-квадрат	0,644316211			
Стандартная ошибка	23,72857949			
Наблюдения	14			
<i>Дисперсионный анализ</i>				
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Регрессия	2	14385,42824	7192,714122	12,7746591
Остаток	11	6193,500328	563,0454844	
Итого	13	20578,92857		
<i>Коэффициенты Стандартная ошибка t-статистика P-Значение</i>				
Y-пересечение	-83,96177158	34,41557268	-2,439644761	0,032843395
X1	0,007389148	0,003351541	2,204701733	0,049676412
X2	0,016266784	0,005269165	3,087165498	0,010338233

Уравнение регрессии имеет вид:

$$\hat{y} = -83,962 + 0,007X_1 + 0,016X_2.$$

Интерпретация параметров уравнения: При увеличении X_1 на единицу своего измерения \bar{Y} увеличится на 0,007 (при неизменном значении $X_2 = 7153,214$, закрепленном на среднем уровне), а при увеличении X_2 на единицу своего измерения \bar{Y} увеличится на 0,016 (при неизменном значении $X_1 = 4267,143$, закрепленном на среднем уровне).

3) Вычислим средние коэффициенты эластичности $\bar{\epsilon}_j$:

$$\bar{\epsilon}_1 = b_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 0,49; \bar{\epsilon}_2 = b_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 1,82$$

Это значит, что при увеличении собственных средств на 1% (при неизменном объеме страховых резервов), годовая прибыль увеличится на 0,49%.

При увеличении страховых резервов на 1% (при неизменном объеме собственных средств), годовая прибыль увеличится на 1,82%.

Задача 5. Исследуется влияние стажа работы, уровня образования и пола менеджера по продажам на размер дохода от реализации товаров, принесенного торговой фирме за последний год. Имеются сведения по 10 менеджерам:

Менеджер	Доход, млн руб.	Стаж, лет	Образование	Пол
Первая	286	7	высшее	женский
Второй	143	6	среднее	мужской
Третий	187	3	высшее	мужской
Четвертая	110	4	среднее	женский
Пятая	253	7	высшее	женский
Шестой	352	8	высшее	мужской
Седьмая	154	3	высшее	женский
Восьмой	308	5	высшее	мужской

Девятая	187	8	среднее	женский
Десятый	242	8	высшее	мужской

Требуется:

- 1) Построить линейную регрессионную модель дохода с полным набором факторов. Оценить параметры модели.
- 2) Существенна ли разница в размере дохода, принесенного фирме менеджерами-мужчинами и менеджерами-женщинами?
- 3) Существенна ли разница в размере дохода, принесенного фирме менеджерами с высшим образованием и с средним образованием?
- 4) Каков средний доход менеджера с высшим образованием и стажем работы 7 лет?

Решение.

Упорядочим имеющиеся статистические данные по полу менеджера:

Менеджер	Доход, млн руб.	Стаж, лет	Образование	Пол
Первая	286	7	высшее	женский
Четвертая	110	4	среднее	женский
Пятая	253	7	высшее	женский
Седьмая	154	3	высшее	женский
Девятая	187	8	среднее	женский
Второй	143	6	среднее	мужской
Третий	187	3	высшее	мужской
Шестой	352	8	высшее	мужской
Восьмой	308	5	высшее	мужской
Десятый	242	8	высшее	мужской

Введем две фиктивные переменные:

$$z_1 = \begin{cases} 0 & \text{– женский пол,} \\ 1 & \text{– мужской пол.} \end{cases}$$

$$z_2 = \begin{cases} 0 & \text{– среднее,} \\ 1 & \text{– высшее.} \end{cases}$$

Тогда, менеджеру-мужчине с высшим образованием будет соответствовать набор переменных $z_1 = 1, z_2 = 1$.

Получаем окончательную таблицу статистических данных:

Менеджер	Доход, млн руб. (y)	Стаж, лет (x)	Образование (z ₂)	Пол (z ₁)
Первая	286	7	1	0
Четвертая	110	4	0	0
Пятая	253	7	1	0
Седьмая	154	3	1	0
Девятая	187	8	0	0
Второй	143	6	0	1
Третий	187	3	1	1
Шестой	352	8	1	1
Восьмой	308	5	1	1
Десятый	242	8	1	1

С помощью инструмента «Регрессия» пакета «Анализ данных» строим уравнение множественной регрессии вида

$$\hat{y} = b_0 + b_1x + c_1z_1 + c_2z_2$$

ВЫВОД ИТОГОВ						
<i>Регрессионная статистика</i>						
Множественный R	0,8821					
R-квадрат	0,7781					
Нормированный R-квадрат	0,6671					
Стандартная ошибка	45,3532					
Наблюдения	10,0000					
Дисперсионный анализ						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	3,0000	43270,1272	14423,3757	7,0121	0,0218	
Остаток	6,0000	12341,4728	2056,9121			
Итого	9,0000	55611,6000				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
У-пересечение	10,1612	52,3907	0,1940	0,8526	-118,0343	138,3568
Стаж, лет	21,4728	7,4844	2,8690	0,0285	3,1591	39,7864
Образование	105,4946	32,1045	3,2860	0,0167	26,9378	184,0514
Пол	23,0065	29,4471	0,7813	0,4643	-49,0478	95,0609

$$t_{\text{крит}}(0,05; 10 - 3 - 1) = t_{\text{крит}}(0,05; 6) = 2,4469$$

Уравнение регрессии имеет вид

$$\hat{y} = 10,1612 + 21,4728 \cdot x + 23,0065 \cdot z_1 + 105,4946 \cdot z_2,$$

На основе t-критерия Стьюдента получаем, что коэффициент регрессии при z_1 статистически незначим ($0,7813 < 2,4469$); при z_2 – статистически значим ($3,2860 > 2,4469$).

Таким образом, разница в размере дохода, принесенного фирме менеджерами-мужчинами и менеджерами-женщинами незначительна.

А разница, между менеджерами с высшим образованием и со средним образованием существенна.

Исключим переменную z_1 из рассмотрения и построим регрессию вида

$$\hat{y} = b_0 + b_1x + c_2z_2$$

Вывод итогов						
<i>Регрессионная статистика</i>						
Множественный R	0,869195164					
R-квадрат	0,755500233					
Нормированный R-квадрат	0,685643156					
Стандартная ошибка	44,0730291					
Наблюдения	10					
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	2	42014,57674	21007,28837	10,81494	0,00722727	
Остаток	7	13597,02326	1942,431894			
Итого	9	55611,6				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
У-пересечение	15,68992248	50,44537994	0,311027938	0,764837	-103,59445	134,97429
Стаж, лет	21,82945736	7,259589556	3,006982309	0,019746	4,66325584	38,995659
Образование	111,0232558	30,43095648	3,648365634	0,008196	39,0654781	182,98103

$$\hat{y} = 15,6899 + 21,8295x + 111,0234z_2$$

Для менеджеров с высшим образованием:

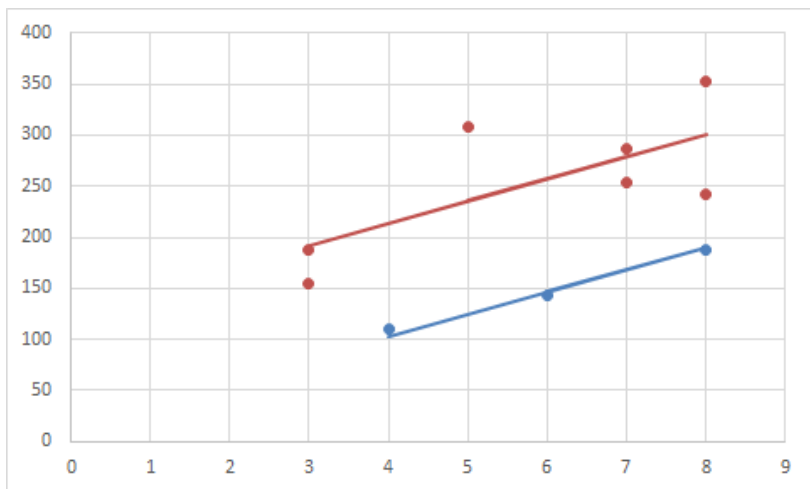
$$\hat{y} = 126,7133 + 21,8295x$$

Для менеджеров со средним образованием:

$$\hat{y} = 15,6899 + 21,8295x$$

Таким образом, разница в размере дохода, принесенного фирме менеджерами-мужчинами и менеджерами-женщинами, в среднем, составляет 111,0234 млн. руб.

Построим корреляционное поле с расслоением выборки на респондентов с высшим (красный цвет) и со средним образованием (синий цвет):



Средний доход менеджера с высшим образованием и стажем работы 7 лет составляет:

$$\hat{y} = 126,7133 + 21,8295 \cdot 7 = 279,5198.$$

Задачи для самостоятельного решения

Задача 1. Исследуется зависимость прибыли банка (y) от объема межбанковских кредитов и депозитов (x). Имеются сведения для 8 банков:

№	x	y	№	x	y
1	2	30,78	5	2,9	34,69
2	2,1	31,71	6	3,3	37,17
3	2,3	32,7	7	3,8	40,38
4	2,4	30,19	8	4,6	44,33

Требуется:

- 1) Построить уравнение регрессии, описывающее зависимость прибыли банка (y) от объема межбанковских кредитов и депозитов (x). Пояснить экономический смысл коэффициентов регрессии.
- 2) Оценить степень тесноты связи между переменными с помощью коэффициента парной корреляции.
- 3) Проверить значимость коэффициентов уравнения регрессии на уровне значимости 0,05
- 4) Найти стандартную ошибку и доверительный интервал для уравнения регрессии в целом, а также для индивидуального прогнозного значения $y_{пр}$ для $x_{пр} = 3$.

Ответ:

- 1) Уравнение регрессии: $\hat{y} = 19,55 + 5,37x$.
- 2) Коэффициент корреляции равен 0,979.
- 3) Расчетные значения t -статистик для оценок коэффициентов уравнения регрессии: $t_{b_0} = 14,34$; $t_{b_1} = 12$.
- 4) $y_{пр} = 35,546$. Доверительный интервал для прогнозного значения: (32,82;38,47).

Задача 2. Приведены данные о стоимости квартиры в зависимости от ее общей площади, площади кухни и расстояния от метро.

№	Цена квартиры, Y	Общая площадь, X_1	Площадь кухни, X_2	Расстояние от метро, X_3
1	39,879	87	13	10
2	50,738	115	12	10
3	49,957	114	10	5
4	40,430	90	8	5

№	Цена квартиры, У	Общая площадь, X1	Площадь кухни, X2	Расстояние от метро, X3
5	49,473	116	10	10
6	45,654	107	9	10
7	39,120	93	8	15
8	72,444	176	10	10
9	42,128	96	9	8
10	40,701	92	8	10
11	73,729	176	25	20
12	31,922	74	6	15
13	44,741	106	8	10
14	40,423	88	8	3
15	32,255	74	8	10
16	33,655	75	8	10
17	49,251	115	9	5
18	38,709	92	8	15
19	47,083	110	9	5

Требуется:

1) Построить уравнение регрессии. оценить значимость уравнения в целом, выписать стандартные ошибки коэффициентов регрессии и их t-статистики. Построить доверительные интервалы на уровне значимости 0,05 для коэффициентов регрессии.

2) Взять в качестве прогнозных значений факторы $x_1 = 120$, $x_2 = 15$, $x_3 = 18$, найти прогнозные значения стоимости квартиры. Дать экономическую интерпретацию оценке прогноза.

Ответ:

1) $\hat{y} = 5,35 + 0,37x_1 + 0,27x_2 - 0,021x_3$; $R^2 = 0,988$;
 $F_{\text{стат}} = 2511,5$; $t_{b_0} = 9,89$; $t_{b_1} = 59,87$; $t_{b_2} = 5,68$;

$t_{b_3} = -6,07$. ДИ: $4,2 < \beta_0 < 6,5$; $0,36 < \beta_1 < 0,39$;
 $0,117 < \beta_2 < 0,38$; $0,29 < \beta_3 < -0,14$.
2) $y_{\text{пр}} = 50,677$.

§5. Методы и модели прогнозирования временных рядов экономических показателей

Понятие ряда динамики. Понятие временного ряда

При анализе многих экономических показателей зачастую используют ежегодные, ежеквартальные, ежемесячные, ежедневные данные. Например, это могут быть годовые данные по ВВП, ВВП, объему чистого экспорта, инфляции и т. д., месячные данные по объему продажи продукции, ежедневные объемы выпуска какой-либо фирмы. Для рационального анализа необходимо систематизировать моменты получения соответствующих статистических данных. В этом случае следует упорядочить данные по времени их получения и построить так называемые **временные ряды**.

Временной ряд – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени.

$$y_1, y_2, \dots, y_t, \dots, y_n$$

t – показатель времени

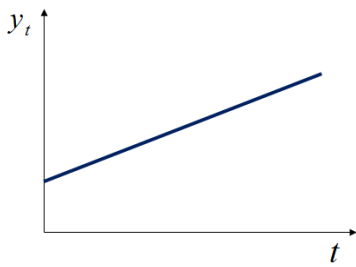
y –уровень ряда

Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

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

Зависимость уровней ряда от времени может принимать различные формы.

Во-первых, большинство временных рядов экономических показателей имеют тенденцию u_t , которая характеризует совокупное долговременное воздействие множества факторов на динамику изучаемого показателя



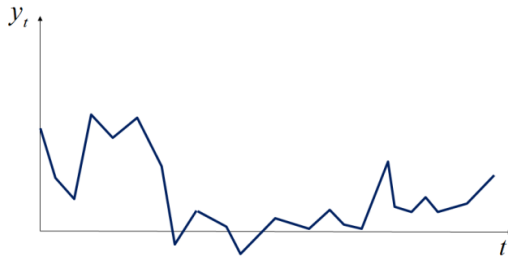
Сезонная компонента характеризует устойчивые внутригодовые колебания уровней (квартальные, месячные, недельные), которые носят периодический характер. Если период колебаний составляет несколько лет, говорят, что во временном ряде присутствует циклическая компонента.

Во-вторых, изучаемый показатель может быть подвержен циклическим (сезонными) колебаниям S_t



В третьих, временные ряды могут не содержать тенденции и циклической компоненты, тогда можно говорить

о некоторой (положительной или отрицательной) случайной компоненте ε_t .



Проверка гипотезы о наличии тренда

Исследование временного ряда начинается с решения вопроса о его стационарности. Одним из возможных вариантов для этого служит проверка гипотезы H_0 о случайности (отсутствии) временного тренда при конкурирующей гипотезе H_1 о неслучайности (наличии) временного тренда, основанная на сравнении средних значений первой и второй половины ряда по статистике (расчетному значению критерия)

$$t_{\text{расч}} = \frac{\bar{y}_{\text{П}} - \bar{y}_{\text{В}}}{\sqrt{(n_{\text{П}} - 1)s_{\text{П}}^2 + (n_{\text{В}} - 1)s_{\text{В}}^2}} \cdot \sqrt{\frac{n_{\text{П}} \cdot n_{\text{В}} \cdot (n_{\text{П}} + n_{\text{В}} - 2)}{n_{\text{П}} + n_{\text{В}}}}$$

Здесь $\bar{y}_{\text{П}}$, $\bar{y}_{\text{В}}$ – средние значения первой и второй половины ряда, имеющих длины $n_{\text{П}}$, $n_{\text{В}}$ ($n_{\text{П}} + n_{\text{В}} = n$); $s_{\text{П}}^2$, $s_{\text{В}}^2$ – исправленные дисперсии первой и второй половины ряда. Средние значения и исправленные дисперсии определяются по следующим формулам:

$$\bar{y}_{\text{П}} = \frac{1}{n_{\text{П}}} \sum_{t=1}^{n_{\text{П}}} y_t, \quad \bar{y}_{\text{В}} = \frac{1}{n_{\text{В}}} \sum_{t=n_{\text{П}}+1}^n y_t,$$

$$s_{\Pi}^2 = \frac{1}{n_{\Pi} - 1} \sum_{t=1}^{n_{\Pi}} (y_t - \bar{y}_{\Pi})^2, \quad \bar{y}_{\text{В}} = \frac{1}{n_{\text{В}} - 1} \sum_{t=n_{\Pi}+1}^n (y_t - \bar{y}_{\text{В}})^2.$$

Значение $t_{\text{расч}}$ сравнивается с табличным значением распределения Стьюдента $t_{\text{табл}}$, найденное по таблице критических значений распределения Стьюдента с $df = n - 2$ степенями свободы и уровнем значимости α (двухсторонняя критическая область). В случае, если $|t_{\text{расч}}| < t_{\text{табл}}$ гипотеза H_0 о случайности временного тренда принимается и ряд считается **стационарным**. В противном случае ($|t_{\text{расч}}| > t_{\text{табл}}$) - гипотеза H_0 отвергается, что свидетельствует о значимости различия средних первой и второй половины ряда и неслучайности (наличии) временного тренда. Другими словами, ряд является **динамическим**.

Модель временного ряда

Временной ряд может быть представлен различными математическими моделями.

Определение. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется **аддитивной моделью** временного ряда.

$$y_t = u_t + S_t + \varepsilon_t$$

Определение. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется **мультипликативной моделью** временного ряда.

$$y_t = u_t S_t \varepsilon_t$$

Мультипликативную модель часто можно свести к аддитивной модели логарифмированием

$$\ln y_t = \ln u_t + \ln S_t + \ln \varepsilon_t$$

Порядок анализа модели

1. Сглаживание исходного ряда методом скользящей средней.
2. Расчет значений периодической компоненты
3. Устранение периодической компоненты.
4. Аналитическое выравнивание ряда
5. Расчет смоделированных уровней ряда
6. Расчет абсолютных и/или относительных ошибок.

Сглаживание ряда динамики

Моделирование временного ряда заключается в анализе его структуры; подборе, построении и проверке качества уравнения тренда; анализе остатков. Построение тренда в теории временных рядов называется *сглаживанием*. Существующие методы выделения тренда можно разделить на две группы: методы численного сглаживания, когда тренд задается численными значениями сглаженных уровней, вычисленных по значениям уровней исходного ряда (метод скользящих средних, экспоненциальное сглаживание и др.), а также методы аналитического выравнивания (построение уравнения тренда).

Численное сглаживание

Сглаживание ряда динамики (временного ряда) – представление тренда в данной точке посредством среднего значения ряда, вычисленного в окрестности данной точки.

Число точек (k), используемых при сглаживании, называется **базой**.

Простое скользящее среднее

Сглаженные значения вычисляют последовательно как среднее арифметическое значение.:

$$\tilde{y}_t = \frac{y_t + y_{t-1} + y_{t+1} + y_{t-2} + y_{t+2} + \dots}{k}, \quad (1)$$

где k – нечетное число.

При сглаживанию по четному числу членов ряда сначала вычисляются средние значения k уровней ряда, которые затем центрируются, т.е. в качестве сглаженных значений принимаются средние значения 2х рядом стоящих средних.

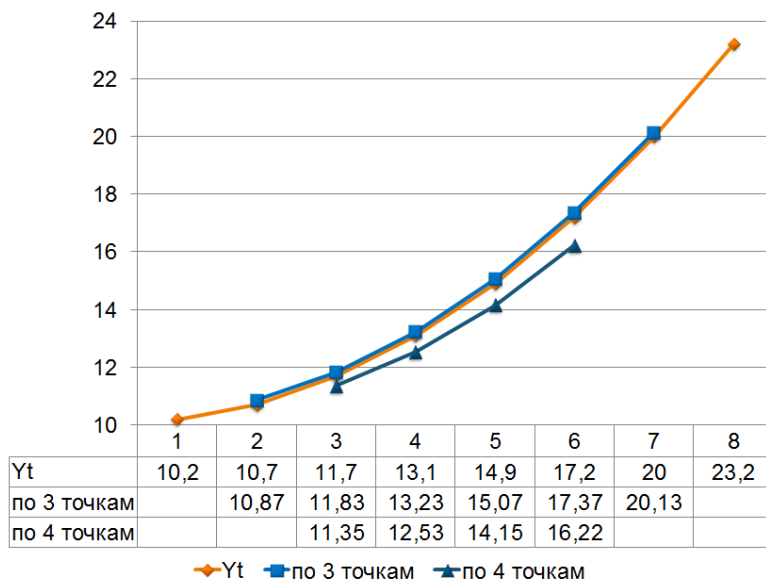
Пример. Сглаживание временного ряда по 3 точкам:

$$\tilde{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}$$

□	□□	Сглаженный ряд
1	10,2	–
2	10,7	$\frac{10,2 + 10,7 + 11,7}{3} = 11,2$
3	11,7	$\frac{10,7 + 11,7 + 13,1}{3} = 11,83$
4	13,1	13,23
5	14,9	15,07
6	17,2	17,37
7	20,0	20,13
8	23,2	–

Пример. Сглаживание временного ряда по 4 точкам:

□	□□	Сглаженный ряд по 4 точкам	Центрирование
1	10,2	—	—
2	10,7	$\frac{10,2+10,7+11,7+13,1}{4} = 11,83.$	—
3	11,7	12,60	$\frac{11,83+12,60}{2} = 12,22.$
4	13,1	14,23	13,41
5	14,9	16,30	15,26
6	17,2	18,83	17,56
7	20,0	—	—
8	23,2	—	—



Аналитическое сглаживание ряда динамики

Построение уравнения тренда для моделирования тенденции временного ряда называют **аналитическим сглаживанием временного ряда**.

Применяют следующие виды тренда:

- Линейный тренд: $\hat{y}_t = b_0 + b_1 t$;
- Степенной тренд: $\hat{y}_t = b_0 t^{b_1}$;
- Экспоненциальный тренд: $\hat{y}_t = b_0 \cdot e^{b_1 t}$;
- Показательный тренд: $\hat{y}_t = b_0 \cdot b_1^t$;
- Гиперболический тренд: $\hat{y}_t = b_0 + \frac{b_1}{t}$;
- Полиномиальный тренд: $\hat{y}_t = b_0 + b_1 t + b_2 t^2 + \dots$;
- Логистическая кривая: $\hat{y}_t = \frac{1}{b_0 + b_1 e^{-t}}$

Критерий отбора наилучшей формы тренда:

наименьшее значение \bar{A} + наибольшее значение

$R_{\text{скорр}}^2$

Выбор вида тренда на практике производится:

1. с учетом выводов экономической теории;
2. результатов визуального анализа графика ряда;
3. результатов исследования структуры ряда.

Анализ качества уравнения тренда

1. Средняя ошибка аппроксимации;
2. Критерии Фишера, Стьюдента.
3. Критерий Дарбина-Уотсона

Исследование структуры ряда. Автокорреляция уровней ряда

При наличии во временном ряде тенденции и циклических колебаний значение каждого последующего уровня ряда зависят от предыдущих. В этом случае говорят, что ряд имеет автокорреляцию.

Автокорреляция элементов временного ряда – корреляционная зависимость между последовательными элементами временного ряда.

Лаг – число периодов, по которым рассчитывается коэффициент автокорреляции между парами элементов ряда.

Автокорреляционная функция временного ряда – последовательность коэффициентов автокорреляции с лагами, равными 1, 2, 3

График зависимости значений функции от величины лага (порядка коэффициента автокорреляции) называется **коррелограммой**.

Коэффициента автокорреляции l -го порядка:

$$r_l = \frac{\sum_{t=l+1}^n (y_t - \bar{y}_t) \cdot (y_{t-l} - \bar{y}_{t-l})}{\sqrt{\sum_{t=l+1}^n (y_t - \bar{y}_t)^2 \cdot \sum_{t=l+1}^n (y_{t-l} - \bar{y}_{t-l})^2}}$$

$$\bar{y}_t = \frac{1}{n-l} \sum_{t=l+1}^n y_t ; \bar{y}_{t-l} = \frac{1}{n-l} \sum_{t=l+1}^n y_{t-l}$$

По знаку коэффициента автокорреляции нельзя сделать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Значимость коэффициентов автокорреляции

Если $|r_l| < r_{\text{крит}}^l(\alpha; n - l - 2)$, то коэффициенты автокорреляции r_l статистически незначимы и выводы, сделанные по его значению, имеют вероятность ошибки, равную $1 - \alpha$.

При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию.

Если $|r_1| > 0,7$, ряд содержит линейную тенденцию.

Если $|r_1| < 0,7$, ряд содержит нелинейную тенденцию.

Если наибольшее абсолютное значение имеет коэффициент автокорреляции порядка $l = \tau$, причем $|r_\tau| > 0,7$ ряд содержит циклические колебания с периодом в τ моментов.

Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений:

1. Ряд не содержит тенденции и циклических колебаний и имеет случайную структуру.

2. Ряд содержит сильную нейтральную тенденцию, для выявления которой нужно провести дополнительный анализ.

Замечание. Возникают ситуации, когда $|r_l| > 0,7$ и $l \geq 2$, но сущность изучаемого процесса, а также вид графика ряда не позволяют сделать вывод о наличии циклическости. В таких случаях динамика описывается авторегрессионным уравнением либо уравнением с распределенным лагом.

Моделирование сезонных и циклических колебаний

Сезонными колебаниями называют изменения уровней ряда, связанные со сменой времени года или с регулярно повторяющимися из года в год событиями, например, связь изменения температуры воздуха с потребительским спросом, объемом товарооборота, энергопотреблением и др.

Циклические колебания уровней обусловлены социальными, юридическими, экономическими, технологическими факторами (изменение тарифов, повышение заработной платы и пенсий и т.п.).

Сезонные колебания, как правило, имеют характер плавных циклов без скачкообразных изменений уровней. Циклические колебания могут иметь резкие скачки уровней, несколько максимумов и минимумов за год.

Наиболее простым подходом к анализу временных рядов, содержащих сезонные или циклические колебания, является расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

1. Построить график временного ряда.
2. Определить вид модели и период сезонных колебаний.
3. Произвести выравнивание исходного ряда методом скользящей средней.
4. Найти значения сезонной компоненты s_t .

5. Устранить сезонную компоненту из исходных уровней ряда и получить выровненные данные модели.
6. Произвести аналитическое выравнивание уровней и рассчитать значения u_t с использованием полученного уравнения тренда.
7. Получить расчетные значения уровней ряда.
8. Получить случайные составляющие ε_t .

Замечание.

1. В аддитивной модели сумма скорректированных сезонных колебаний компонент внутри цикла равна нулю
2. В мультипликативной модели – равна числу моментов времени внутри цикла (лагу)

Прогнозирование временных рядов

Исследование динамики экономических явлений и процессов, выявление и характеристика основного тренда развития и моделей взаимосвязи дают основания для прогнозирования, т.е. определения будущих размеров уровней показателей, описывающих процессы.

Прогнозирование основывается на предположении, что закономерность развития, действующая в прошлом внутри ряда динамики, сохранится и в прогнозируемом будущем.

Мультипликативная модель ряда

$$y_t = u_t S_t \varepsilon_t,$$

где u_t - тренд; S_t - сезонная составляющая; ε_t - случайная составляющая.

Прогнозируемое значение для $t = n + k$, \tilde{y}_{n+k}

$$\tilde{y}_{n+k} = \tilde{u}_{n+k} \cdot S_i,$$

где \tilde{u}_{n+k} - прогнозируемое значение тренда;

S_i - сезонная составляющая в момент времени $t = n + k$

Средняя абсолютная относительная ошибка прогноза:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \cdot 100\%$$

Аддитивная модель ряда

$$y_t = u_t + S_t + \varepsilon_t,$$

где u_t - тренд; S_t - сезонная составляющая; ε_t - случайная составляющая.

Прогнозируемое значение для $t = n + k$, \tilde{y}_{n+k}

$$\tilde{y}_{n+k} = \tilde{u}_{n+k} + S_i,$$

где \tilde{u}_{n+k} - прогнозируемое значение тренда;

S_i - сезонная составляющая в момент времени $t = n + k$;

Средняя абсолютная относительная ошибка прогноза:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \cdot 100\%$$

Отметим, что полное совпадение фактического предсказываемого значения y_{n+k} и прогнозной оценки маловероятно. Возникновение отклонений фактических уровней временного ряда от выравненных по уравнению тренда связано со следующими причинами:

- всегда существует тренд, который дает более точные результаты при описании тенденции, по сравнению с выбранным;
- тренд, выбранный для прогнозирования, содержит случайную компоненту, так как каждый уровень исходных данных обладает случайной компонентой;
- выявленная тенденция характеризует движение среднего уровня временного ряда, следовательно, возможны отклонения от него.

§6. Анализ и прогнозирование временных рядов в Excel

Анализ временных рядов с помощью инструмента "Мастер диаграмм"

При проведении анализа временных рядов широко применяются методы визуального анализа. Это объясняется тем, что график временного ряда, как правило, позволяют понять характер процесса. Так, анализ графика временного ряда позволяет сделать выводы о следующем:

- наличия сезонных и/или циклических колебаний;
- наличия тренда и его характере;
- степени плавности/прерывности изменений последовательных уровней ряда после устранения тренда.

В Excel для анализа временных рядов можно использовать инструмент "Мастер диаграмм".

При анализе временных рядов диаграммы, используемые при построении модели ряда динамики, имеют тип «График».

Работа с Мастером диаграмм состоит из нескольких основных шагов, выполнение которых рассмотрим на следующем примере.

Пример.

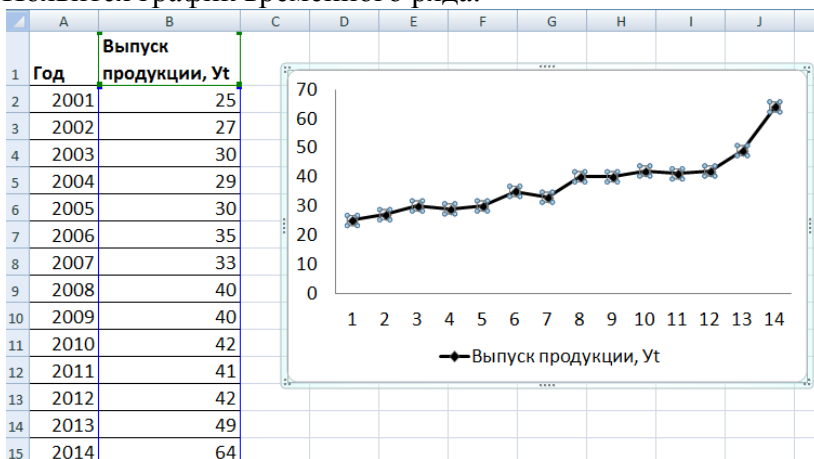
Динамика выпуска продукции некоторой страны характеризуется данными, представленными в таблице:

<i>Год</i>	<i>Выпуск продукции</i>	<i>Год</i>	<i>Выпуск продукции</i>
2001	25	2008	40
2002	27	2009	40
2003	30	2010	42
2004	29	2011	41
2005	30	2012	42
2006	35	2013	49
2007	33	2014	64

Проведем расчет параметров линейного, степенного и полиномиального тренда, выберем наилучшее уравнение тренда, а также построим графики ряда динамики и трендов.

Решение.

Введем данные в Excel. Для построения графика временного ряда воспользуемся инструментом **Мастер диаграмм** Excel. Выделим диапазон ячеек **V1:B15**. Вызовем **Мастер диаграмм**. Выберем тип диаграммы **График**. Появится график временного ряда:



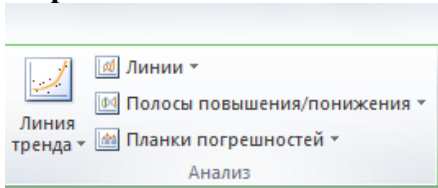
Существует несколько способов определения типа тенденции. К числу наиболее распространённых способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени, расчёт некоторых основных показателей. Выбор наилучшего уравнения тренда принято осуществлять путём перебора основных форм тренда, расчёта по каждому уравнению $R^2_{\text{скаорр}}$ и выбора уравнения тренда с максимальным значением этого коэффициента.

Для осуществления данной процедуры мы воспользуемся опцией **Добавить линию тренда** в **Мастере Диаграмм**, тип **график**. Данная опция позволяет автоматически получить график тренда, уравнение тренда и значение $R^2_{\text{скаорр}}$.

Начнем с построения линейного тренда:

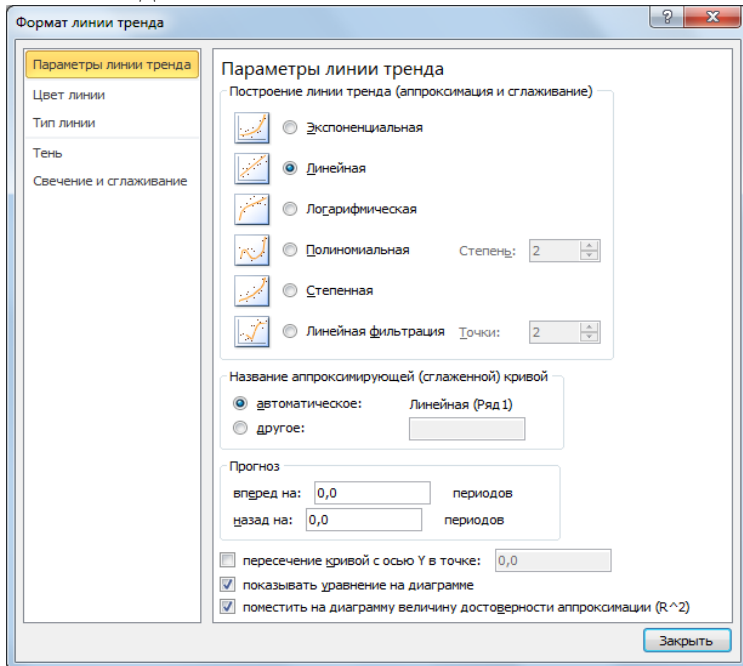
$$\hat{y}_t = b_0 + b_1 t$$

Выделим мышкой график y_t и выберем пункт меню **Добавить линию тренда**. В Excel 2010 после выделения нужно перейти на вкладку **Макет**. Щелкнуть на кнопку **Линия тренда** и



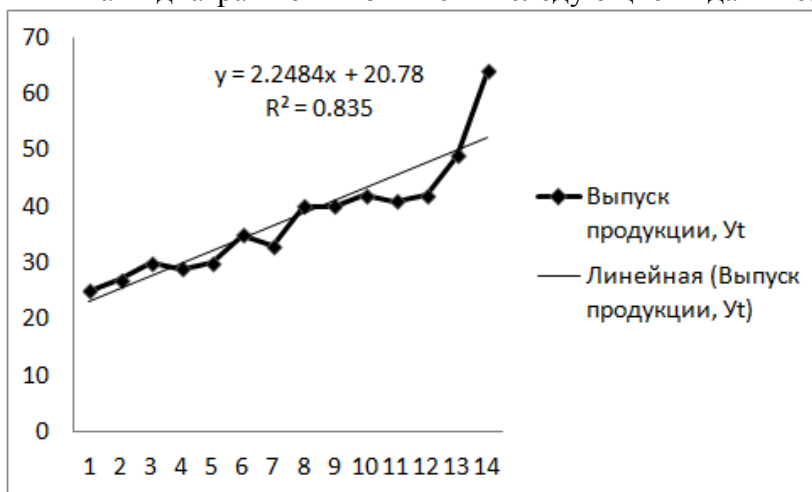
выбрать пункт **дополнительные параметры линии тренда...**

Появится диалоговое окно:

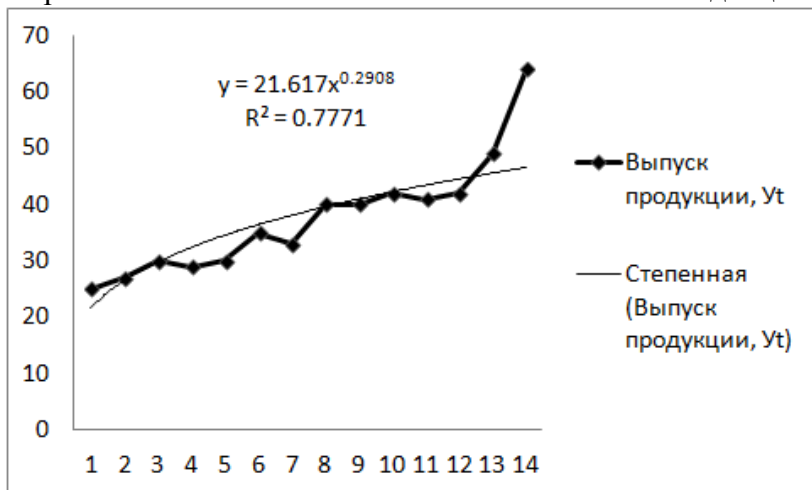


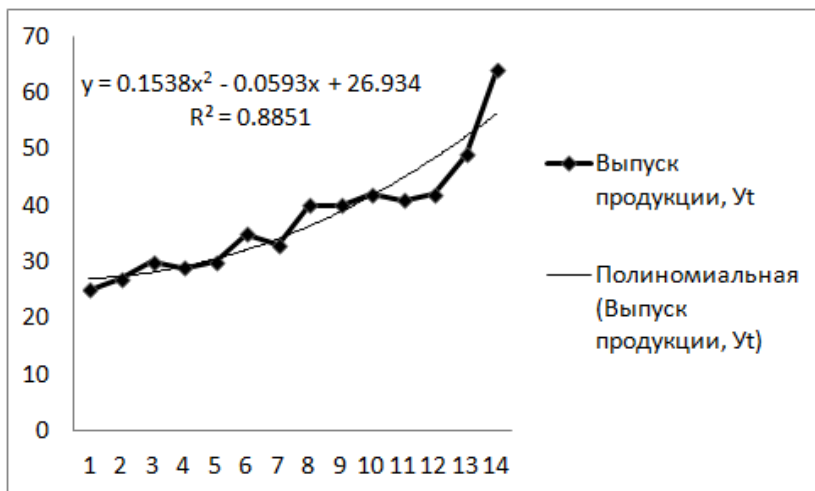
Отметить флажки как на рисунке и нажать на кнопку **закрыть** или **ОК**.

На диаграмме появятся следующие данные:



Далее построить отдельные диаграммы для всех встроенных тенденций:





Сравним полученные значения $R^2_{\text{скорр}}$:

Вид тренда	$R^2_{\text{скорр}}$
Линейный	0,8350
Степенной	0,7771
Полиномиальный 2-й степени	0,8851

В нашем случае исходные данные лучше всего описывает полином 2-ой степени.

Таким образом, уравнение тренда имеет следующий вид:

$$\hat{y}_t = 0,1538t^2 - 0,0593t + 26,934$$

Решение типовых задач

Задача 1. Имеются данные о сводном индексе цен строительной продукции по РФ за период с января 2010 года по октябрь 2014 года.

Исходные данные
на конец периода в %

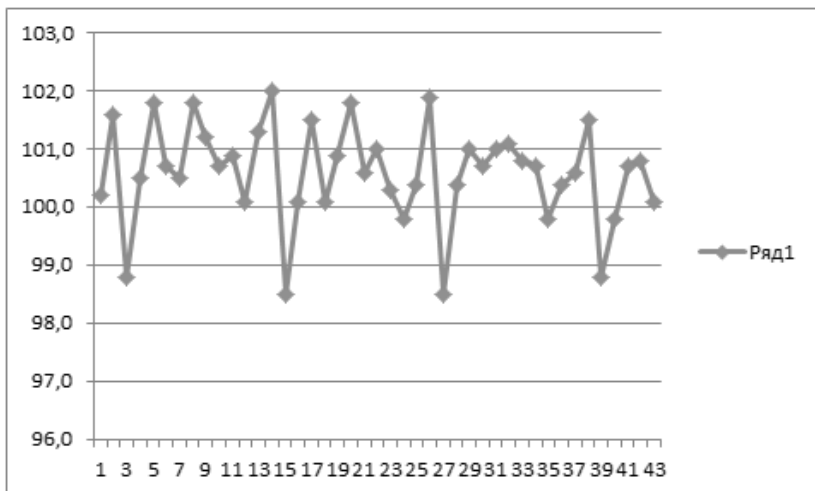
	2010	2011	2012	2013
<i>К предыдущему месяцу</i>				
Январь	100,2	101,3	100,4	100,6
Февраль	101,6	102,0	101,9	101,5
Март	98,8	98,5	98,5	98,8
Апрель	100,5	100,1	100,4	99,8
Май	101,8	101,5	101,0	100,7
Июнь	100,7	100,1	100,7	100,8
Июль	100,5	100,9	101,0	100,1
Август	101,8	101,8	101,1	
Сентябрь	101,2	100,6	100,8	
Октябрь	100,7	101,0	100,7	
Ноябрь	100,9	100,3	99,8	
Декабрь	100,1	99,8	100,4	

Требуется:

- 1) Вычислить коэффициенты автокорреляции. Проверить статистическую значимость коэффициентов автокорреляции. Построить коррелограмму. Сделать вывод.

Решение.

Построим график временного ряда:



Из визуального анализа графика временного ряда следует, что ряд состоит из циклической и случайной составляющих, возможно трендовой.

При наличии во временном ряду тенденции и циклических колебаний значение каждого последующего уровня ряда зависит от значений предыдущих уровней.

Анализ автокорреляционной функции временного ряда позволяет установить степень зависимости последующих членов ряда от предыдущих и временной интервал, в течение которого эта зависимость статистически значима.

Вычислим коэффициенты автокорреляции 1-го, 2-го , ..., 12-го порядка.

Составим вспомогательную таблицу:

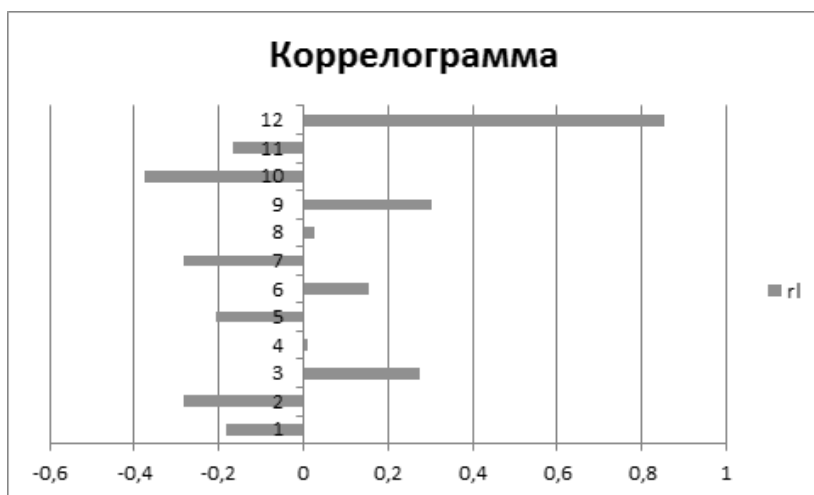
t	Y _t	Y _{t-1}	Y _{t-2}	Y _{t-3}	Y _{t-4}	Y _{t-5}	Y _{t-6}	Y _{t-7}	Y _{t-8}	Y _{t-9}	Y _{t-10}	Y _{t-11}	Y _{t-12}
1	100,2												
2	101,6	100,2											
3	98,8	101,6	100,2										
4	100,5	98,8	101,6	100,2									
5	101,8	100,5	98,8	101,6	100,2								
6	100,7	101,8	100,5	98,8	101,6	100,2							
7	100,5	100,7	101,8	100,5	98,8	101,6	100,2						
8	101,8	100,5	100,7	101,8	100,5	98,8	101,6	100,2					
9	101,2	101,8	100,5	100,7	101,8	100,5	98,8	101,6	100,2				
10	100,7	101,2	101,8	100,5	100,7	101,8	100,5	98,8	101,6	100,2			
11	100,9	100,7	101,2	101,8	100,5	100,7	101,8	100,5	98,8	101,6	100,2		
12	100,1	100,9	100,7	101,2	101,8	100,5	100,7	101,8	100,5	98,8	101,6	100,2	
13	101,3	100,1	100,9	100,7	101,2	101,8	100,5	100,7	101,8	100,5	98,8	101,6	100,2
14	102,0	101,3	100,1	100,9	100,7	101,2	101,8	100,5	100,7	101,8	100,5	98,8	101,6
15	98,5	102,0	101,3	100,1	100,9	100,7	101,2	101,8	100,5	100,7	101,8	100,5	98,8
16	100,1	98,5	102,0	101,3	100,1	100,9	100,7	101,2	101,8	100,5	100,7	101,8	100,5
17	101,5	100,1	98,5	102,0	101,3	100,1	100,9	100,7	101,2	101,8	100,5	100,7	101,8
18	100,1	101,5	100,1	98,5	102,0	101,3	100,1	100,9	100,7	101,2	101,8	100,5	100,7
19	100,9	100,1	101,5	100,1	98,5	102,0	101,3	100,1	100,9	100,7	101,2	101,8	100,5
20	101,8	100,9	100,1	101,5	100,1	98,5	102,0	101,3	100,1	100,9	100,7	101,2	101,8
21	100,6	101,8	100,9	100,1	101,5	100,1	98,5	102,0	101,3	100,1	100,9	100,7	101,2
22	101,0	100,6	101,8	100,9	100,1	101,5	100,1	98,5	102,0	101,3	100,1	100,9	100,7
23	100,3	101,0	100,6	101,8	100,9	100,1	101,5	100,1	98,5	102,0	101,3	100,1	100,9
24	99,8	100,3	101,0	100,6	101,8	100,9	100,1	101,5	100,1	98,5	102,0	101,3	100,1
25	100,4	99,8	100,3	101,0	100,6	101,8	100,9	100,1	101,5	100,1	98,5	102,0	101,3
26	101,9	100,4	99,8	100,3	101,0	100,6	101,8	100,9	100,1	101,5	100,1	98,5	102,0
27	98,5	101,9	100,4	99,8	100,3	101,0	100,6	101,8	100,9	100,1	101,5	100,1	98,5
28	100,4	98,5	101,9	100,4	99,8	100,3	101,0	100,6	101,8	100,9	100,1	101,5	100,1
29	101,0	100,4	98,5	101,9	100,4	99,8	100,3	101,0	100,6	101,8	100,9	100,1	101,5
30	100,7	101,0	100,4	98,5	101,9	100,4	99,8	100,3	101,0	100,6	101,8	100,9	100,1
31	101,0	100,7	101,0	100,4	98,5	101,9	100,4	99,8	100,3	101,0	100,6	101,8	100,9
32	101,1	101,0	100,7	101,0	100,4	98,5	101,9	100,4	99,8	100,3	101,0	100,6	101,8
33	100,8	101,1	101,0	100,7	101,0	100,4	98,5	101,9	100,4	99,8	100,3	101,0	100,6
34	100,7	100,8	101,1	101,0	100,7	101,0	100,4	98,5	101,9	100,4	99,8	100,3	101,0
35	99,8	100,7	100,8	101,1	101,0	100,7	101,0	100,4	98,5	101,9	100,4	99,8	100,3
36	100,4	99,8	100,7	100,8	101,1	101,0	100,7	101,0	100,4	98,5	101,9	100,4	99,8
37	100,6	100,4	99,8	100,7	100,8	101,1	101,0	100,7	101,0	100,4	98,5	101,9	100,4
38	101,5	100,6	100,4	99,8	100,7	100,8	101,1	101,0	100,7	101,0	100,4	98,5	101,9
39	98,8	101,5	100,6	100,4	99,8	100,7	100,8	101,1	101,0	100,7	101,0	100,4	98,5
40	99,8	98,8	101,5	100,6	100,4	99,8	100,7	100,8	101,1	101,0	100,7	101,0	100,4
41	100,7	99,8	98,8	101,5	100,6	100,4	99,8	100,7	100,8	101,1	101,0	100,7	101,0
42	100,8	100,7	99,8	98,8	101,5	100,6	100,4	99,8	100,7	100,8	101,1	101,0	100,7
43	100,1	100,8	100,7	99,8	98,8	101,5	100,6	100,4	99,8	100,7	100,8	101,1	101,0

Коэффициент автокорреляции является коэффициентом парной корреляции между соответствующими столбцами нашей таблицы и находится с помощью встроенной функции Excel **КОРРЕЛ**.

Например, коэффициент автокорреляции седьмого порядка r_7 есть коэффициент парной корреляции между выделенными заливкой столбцами таблицы.

На основе расчетной таблицы вычислим коэффициенты автокорреляции и запишем табличное задание **автокорреляционной функции**:

Лаг, l	Коэффициент автокорреляции r_l
1,0	-0,1837
2,0	-0,2845
3,0	0,27494
4,0	0,00855
5,0	-0,206
6,0	0,15633
7,0	-0,285
8,0	0,02573
9,0	0,30192
10,0	-0,3762
11,0	-0,1664
12	0,85465



Значение $r_{12} = 0,85456$ значительно превышает все остальные.

$$\begin{aligned} r_{\text{крит}}^l(\alpha; n - l - 2) &= r_{\text{крит}}^{12}(0,05; 43 - 12 - 2) \\ &= r_{\text{крит}}^{12}(0,05; 29) = 0,35 \end{aligned}$$

Получаем, что $|r_{12}| > r_{\text{крит}}^{12}$, следовательно, коэффициенты автокорреляции r_{12} статистически значим.

Таким образом, ряд содержит циклические колебания с периодом в 12 месяцев.

На основе коэффициента автокорреляции r_1 можно предположить, что ряд имеет нелинейную тенденцию, т.к. $|r_1| < 0,7$. Проверим значимость $r_1 = -0,183$, для подтверждения или опровержения данного предположения.

Имеем,

$$\begin{aligned} r_{\text{крит}}^l(\alpha; n - l - 2) &= r_{\text{крит}}^1(0,05; 43 - 1 - 2) \\ &= r_{\text{крит}}^1(0,05; 40) = 0,3 \end{aligned}$$

Следовательно, r_1 является статистически незначимым ($|r_1| < r_{\text{крит}}^1$) и предположение о наличии нелинейной тенденции опровергается. То есть, или ряд содержит линейную тенденцию, или не содержит тенденцию.

Задача 2. Для 10 предприятий по данным внутренней отчетности получены данные о квартальных товарооборотах за последние 4 года в условных единицах.

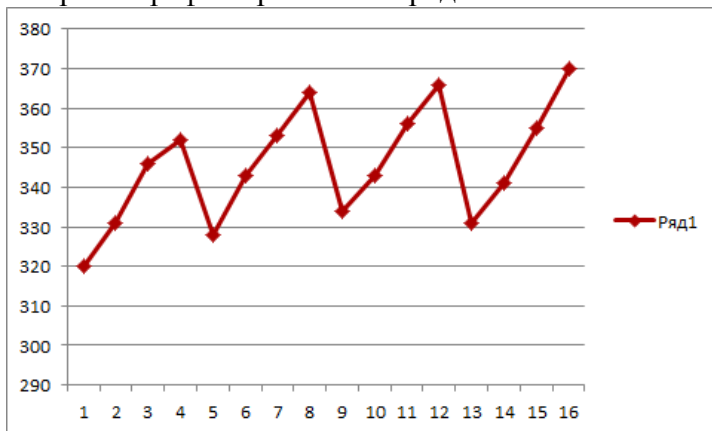
Квартал	Год			
	2011	2012	2013	2014
1	320	328	334	331
2	331	343	343	341
3	346	353	356	355
4	352	364	366	370

Требуется:

1) Найти сезонные компоненты (сезонные индексы) для каждого квартала.

Решение.

Построим график временного ряда:



Из визуального анализа временного ряда следует, что ряд состоит из трендовой, сезонной и случайной составляющих. Период сезонных колебаний равен 4, с увеличением времени не происходит заметного увеличения сезонных колебаний, поэтому построим аддитивную модель ряда.

Шаг 1. Выравнивание исходных уровней ряда методом скользящей средних. Найдем скользящее среднее за 4 квартала и проведем центрирование получившихся уровней.

Шаг 2. Найдем оценки сезонной компоненты S'_t как разность между исходными уровнями ряда и центрированными скользящими средними.

	A	B	C	D	E
1	Время	Временной ряд	Скользящее среднее за 4 квартала	Центрирование	Оценки сезонной
2	1	320			
3	2	331	337,25		
4	3	346	339,25	338,25	7,75
5	4	352	342,25	340,75	11,25
6	5	328	344	343,125	-15,125
7	6	343	347	345,5	-2,5
8	7	353	348,5	347,75	5,25
9	8	364	348,5	348,5	15,5
10	9	334	349,25	348,875	-14,875
11	10	343	349,75	349,5	-6,5
12	11	356	349	349,375	6,625
13	12	366	348,5	348,75	17,25
14	13	331	348,25	348,375	-17,375
15	14	341	349,25	348,75	-7,75
16	15	355			
17	16	370			

Шаг 3. Для вычисления сезонной компоненты сгруппируем данные следующим образом:

		Вычисление сезонной компоненты				
		Номер квартала в году				
		1.00	2.00	3.00	4.00	
	2011.00			7.75	11.25	
	2012.00	-15.13	-2.50	5.25	15.50	
	2013.00	-14.88	-6.50	6.63	17.25	
	2014.00	-17.38	-7.75			
итого		-47.38	-16.75	19.63	44.00	Сумма
Средняя оценка сезонной компоненты <i>Scp</i>		-15.79	-5.58	6.54	14.67	-0.17
Скорректированная сезонная компонента <i>St</i>		-15.75	-5.54	6.58	14.71	0.00

Определим средние оценки сезонной компоненты по каждому кварталу.

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период компенсируют друг друга. Поэтому, в аддитивной модели считается, что сумма значений сезонной компоненты по всем кварталам (или годам, или месяцам, или дням) должна быть равна нулю. В нашем случае:

$$-15,79 - 5,58 + 6,54 + 14,67 = -0,17$$

Определим корректирующий коэффициент $k = \frac{1}{4} \sum_{i=1}^4 S'_{\text{сп}}$.

Определяем скорректированные значения сезонной компоненты как разности между соответствующими средними оценками и корректирующем коэффициентом.

Проверяем условие равенства нулю скорректированных сезонных компонент.

Таким образом, сезонные индексы имеют следующий вид

S1	S2	S3	S4
-15,75	-5,54	6,58	14,71

Задача 3.

Приведен ряд y_t средней цены 1 м² общей площади квартир (все типы квартир) на рынке вторичного жилья города Москвы (руб):

	2011	2012	2013	2014
I квартал	162419.89	166679.88	182005.09	179146.33
II квартал	163426.85	168869.88	183240.08	183080.73
III квартал	162952.13	172260.14	178971.23	184573.86
IV квартал	163202.77	176319.51	177899.02	

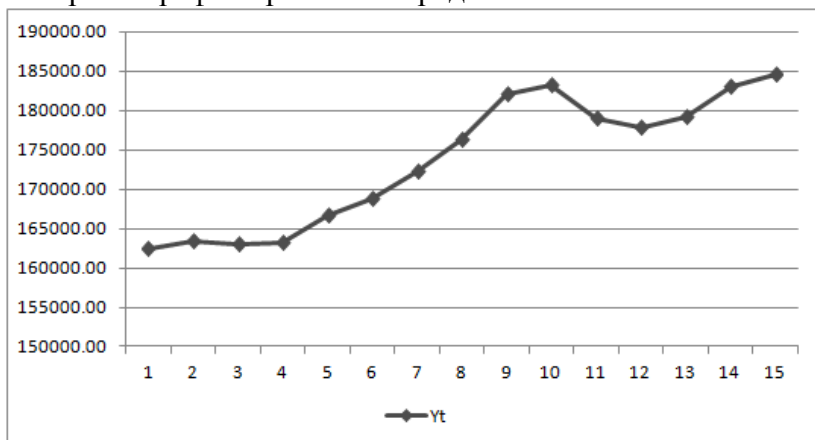
Требуется

- 1) Построить график ряда динамики.
- 2) С помощью автокорреляционной функции определить наличие трендовой и сезонной компоненты.
- 3) Построить модель временного ряда.
- 4) Построить прогноз средней цены 1 м² в первом квартале 2015 г. Оценить точность прогноза.

Решение.

График ряда динамики

Построим график временного ряда



Из визуального анализа графика временного ряда следует, что ряд состоит из трендовой, возможно сезонной, и случайной составляющих.

Автокорреляционная функция

Данные поквартальные, поэтому вычислим коэффициенты автокорреляции с 1-го по 5-ый порядок.

Составим вспомогательную матрицу.

Коэффициент автокорреляции является коэффициентом парной корреляции между соответствующими столб-

цами нашей таблицы и находится с помощью встроенной функции Excel КОРРЕЛ.

$$r_1 = \text{КОРРЕЛ}(D2:D24; E2:E24)$$

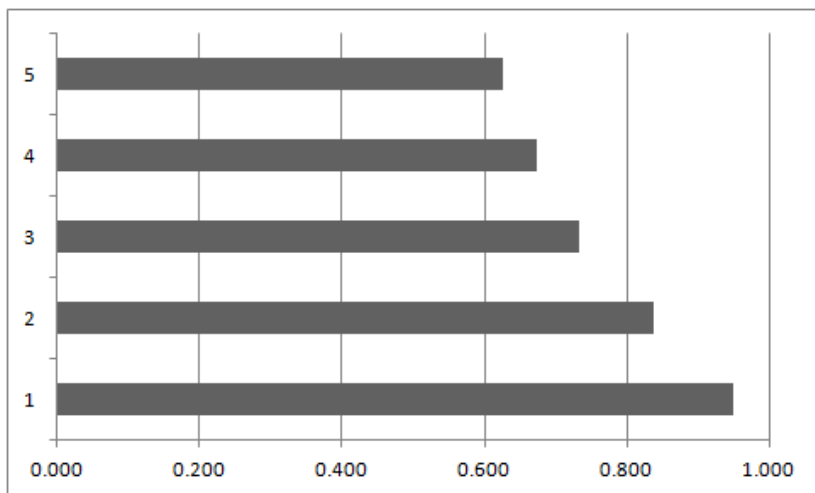
$$r_5 = \text{КОРРЕЛ}(D2:D24; I2:I24)$$

t	Y _t	Y _{t-1}	Y _{t-2}	Y _{t-3}	Y _{t-4}	Y _{t-5}
1.00	162419.89					
2.00	163426.85	162419.89				
3.00	162952.13	163426.85	162419.89			
4.00	163202.77	162952.13	163426.85	162419.89		
5.00	166679.88	163202.77	162952.13	163426.85	162419.89	
6.00	168869.88	166679.88	163202.77	162952.13	163426.85	162419.89
7.00	172260.14	168869.88	166679.88	163202.77	162952.13	163426.85
8.00	176319.51	172260.14	168869.88	166679.88	163202.77	162952.13
9.00	182005.09	176319.51	172260.14	168869.88	166679.88	163202.77
10.00	183240.08	182005.09	176319.51	172260.14	168869.88	166679.88
11.00	178971.23	183240.08	182005.09	176319.51	172260.14	168869.88
12.00	177899.02	178971.23	183240.08	182005.09	176319.51	172260.14
13.00	179146.33	177899.02	178971.23	183240.08	182005.09	176319.51
14.00	183080.73	179146.33	177899.02	178971.23	183240.08	182005.09
15.00	184573.86	183080.73	179146.33	177899.02	178971.23	183240.08

Запишем табличное задание *автокорреляционной функции*:

Лаг, l	Коэффициент автокорреляции r_l	Критическое значение	Вывод
1,0	0,951	0,53	значим
2,0	0,838	0,55	значим
3,0	0,733	0,58	значим
4,0	0,674	0,60	значим
5,0	0,626	0,63	незначим

Коррелограмма имеет вид:



Вывод:

- ряд содержит линейную тенденцию
- ряд не содержит сезонных колебаний.

Построение модели ряда

Определим трендовую компоненту u_t модели.

Для этого построим регрессию $u_t = b_0 + b_1 t$ с помощью функции «Регрессия».

Вывод итогов						
<i>Регрессионная статистика</i>						
Множественный R		0.93				
R-квадрат		0.87				
Нормированный R-квадрат		0.86				
Стандартная ошибка		3167.22				
Наблюдения		15.00				
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	1.00	850920315.30	850920315.30	84.83	0.00	
Остаток	13.00	130406958.96	10031304.54			
Итого	14.00	981327274.27				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	159723.64	1720.94	92.81	0.00	156005.78	163441.50
Переменная X 1	1743.27	189.28	9.21	0.00	1334.36	2152.18

Уравнение тренда имеет вид:

$$u_t = 159723,64 + 1743,27 \cdot t$$

Параметр $b_0 = 159723,64$ интерпретируется как параметр начальных условий (значение переменной y_t в нулевой момент времени, т.е. цена 1 м² в момент $t = 0$, что соответствует трендовому значению средней цены 1 м² общей площади квартиры в VI квартале 2010 г.). $b_1 = 1743,27$ – скорость роста переменной y_t (т.е. в среднем за 2011-2014 г. цена 1 м² общей площади квартиры увеличиваются на 1743,27 руб. в квартал).

Коэффициент парной корреляции $r = 0,93 > 0,7$ показывает, что средняя цена 1 м² общей площади квартиры сильно зависит от времени и на 87% объясняются временным фактором (коэффициент детерминации $R^2 = 0,87$).

Проверка качества уравнения тренда

Проверим статистическую значимость уравнения тренда в целом с помощью F-критерия Фишера. Расчетное значение F-критерия Фишера $F = 84,83$ больше, чем $F_{\text{табл}}(0,05; 1; 15 - 1 - 1) = 4,67$, т.е. уравнение регрессии в целом статистически значимо.

Теперь проверим статистическую значимость параметров уравнения с помощью t-критерия Из результатов *дисперсионного анализа* получаем, что

$$t_{b_0} = 92,81; t_{b_1} = 9,21$$

Табличное значение t-критерия Стьюдента находим с помощью функции СТЬЮДРАСПОБР (Excel): $t_{\text{табл}}(0,05; 15-1-1) = 2,16$. Таким образом, коэффициенты b_0 и b_1 являются статистически значимыми ($|t_{b_i}| > t_{\text{табл}}$).

Дополним анализ вычислением модельной ошибки. Математическую точность модель оценим с помощью среднюю относительную ошибки аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% = \frac{1}{15} \cdot 0,19923 \cdot 100\% = 1,28\%$$

Так как $\bar{A} < 10\%$, точность модели высокая.

Таким образом, уравнение тренда имеет высокое качество.

Модель ряда имеет вид:

$$Y_t = u_t + \varepsilon_t$$

$$u_t = 159723,64 + 1743,27 \cdot t$$

$$r = 0,93; R^2 = 0,87; A = 1,28\%$$

t	$\varepsilon_t = y_t - u_t$	t	ε_t	t	ε_t
1	952.98	6	-1313.40	11	71.58
2	216.66	7	333.59	12	-2743.90
3	-2001.33	8	2649.68	13	-3239.86
4	-3493.96	9	6591.99	14	-1048.73
5	-1760.13	10	6083.71	15	-1298.88

Прогноз средней цены 1 м² общей площади квартиры в I квартале 2015 года

I кварталу 2015 года соответствует момент времени $t = 17$, следовательно

$$y_{17} = 159723,64 + 1743,27 \cdot 17 = 189359,23 \text{ руб.}$$

Средняя абсолютная относительная ошибка прогноза:

$$\begin{aligned}
 MAPE &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \cdot 100\% = \frac{1}{15} \sum_{i=1}^{15} \left| \frac{y_i - u_i}{y_i} \right| \cdot 100\% \\
 &= \frac{0,19923}{15} \cdot 100\% = 1,28\%
 \end{aligned}$$

Таким образом, точность прогноза высокая.

Задачи для самостоятельного решения

Задача 1. Динамика выпуска продукции некоторой страны характеризуется данными (усл. ед.), представленными в таблице:

Год	Выпуск про- дукции	Год	Выпуск продукции	Год	Выпуск продукции
1980	1054	1992	3837	2004	13617
1981	1104	1993	5490	2005	16356
1982	1149	1994	5502	2006	20037
1983	1291	1995	6342	2007	21748
1984	1427	1996	7665	2008	23298
1985	1505	1997	8570	2009	26570
1986	1513	1998	11172	2010	23080
1987	1635	1999	14150	2011	23981
1988	1987	2000	14004	2012	23446
1989	2306	2001	13088	2013	29658
1990	2367	2002	12518	2014	39573
1991	2913	2003	13471	2015	38435

Требуется:

- 1) провести расчет параметров линейного, степенного и экспоненциального трендов;

- 2) построить графики ряда динамики и трендов;
- 3) выбрать наилучший вид трендов на основании графического изображения и значения коэффициента детерминации.

Ответ:

согласно критериям отбора экспоненциальный тренд является наилучшим.

Задача 2. Представлены поквартальные данные о валовом объеме продаж (млн.шт.) за последние четыре года:

Квартал	Год			
	1	2	3	4
I	6	7,2	8	9
II	4,4	4,8	5,6	6,6
III	5	6	6,4	7
IV	9	10	11	10,8

Требуется:

- 1) найти коэффициенты автокорреляции до пятого порядка;
- 2) построить коррелограмму.

Ответ:

Автокорреляционная функция имеет вид:

l	1	2	3	4	5
r_l	0,1652	-0,567	0,1136	0,983	0,1187

Приложения

Приложение 1

Таблица значений z-преобразования Фишера

r	0	1	2	3	4
0,0	0	0,0101	0,02	0,03	0,04
0,1	0,1003	0,1104	0,1206	0,1308	0,1409
0,2	0,2027	0,2132	0,2237	0,2342	0,2448
0,3	0,3095	0,3205	0,3316	0,3428	0,3541
0,4	0,4236	0,4356	0,4477	0,4599	0,4722
0,5	0,5493	0,5627	0,5764	0,5901	0,6042
0,6	0,6932	0,7089	0,725	0,7414	0,7582
0,7	0,8673	0,8872	0,9077	0,9287	0,9505
0,8	1,0986	1,127	1,1568	1,1881	1,2212
0,9	1,4722	1,5275	1,589	1,6584	1,7381
0,99	2,6466	2,6996	2,7587	2,8257	2,9031

r	5	6	7	8	9
0,0	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1511	0,1614	0,1717	0,182	0,1923
0,2	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3654	0,3767	0,3884	0,4001	0,4118
0,4	0,4847	0,4973	0,5101	0,523	0,5361
0,5	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,7753	0,7928	0,8107	0,8291	0,848
0,7	0,973	0,9962	1,0203	1,0454	1,0714
0,8	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,8318	1,9459	2,0923	2,2976	2,6467
0,99	2,9945	3,1063	3,2504	3,4534	3,8002

В качестве примера (для понимания, как получилась таблица) посчитаем значение z для $r=0,997$ (число из строки 0,99 и столбца 7): $z(0,997)=3,2504$.

Приложение 2

Критические значения t-критерия Стьюдента

Число степеней свободы d.f.	Уровень значимости α (двусторонняя критическая область)			Число степеней свободы d.f.	Уровень значимости α (двусторонняя критическая область)		
	0,1	0,05	0,01		0,1	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758
Число степеней свободы d.f.	0,05	0,025	0,005	Число степеней свободы d.f.	0,05	0,025	0,005
	Уровень значимости α (односторонняя критическая область)				Уровень значимости α (односторонняя критическая область)		

Приложение 3

Критические значения распределения Фишера-Снедекора при уровне значимости $\alpha = 0,05$

k1	1	2	3	4	5	6	8	12	24	∞
k2										
1	161,4	199,5	215,7	224,6	230,2	234	238,9	243,9	249	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62

Продолжение прилож.3

k1	1	2	3	4	5	6	8	12	24	∞
k2										
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Приложение 4

Критические значения распределения Пирсона

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6349	5,0239	3,8415	0,0039	0,0010	0,0002
2	9,2103	7,3778	5,9915	0,1026	0,0506	0,0201
3	11,3449	9,3484	7,8147	0,3519	0,2158	0,1148
4	13,2767	11,1433	9,4877	0,7107	0,4844	0,2971
5	15,0863	12,8325	11,0705	1,1455	0,8312	0,5543
6	16,8119	14,4494	12,5916	1,6354	1,2373	0,8721
7	18,4753	16,0128	14,0671	2,1674	1,6899	1,2390
8	20,0902	17,5346	15,5073	2,7326	2,1797	1,6465
9	21,6660	19,0228	16,9190	3,3251	2,7004	2,0879
10	23,2093	20,4832	18,3070	3,9403	3,2470	2,5582
11	24,7250	21,9201	19,6751	4,5748	3,8158	3,0535
12	26,2170	23,3367	21,0261	5,2260	4,4038	3,5706
13	27,6883	24,7356	22,3620	5,8919	5,0088	4,1069
14	29,1412	26,1190	23,6848	6,5706	5,6287	4,6604
15	30,5779	27,4884	24,9958	7,2609	6,2621	5,2294
16	31,9999	28,8454	26,2962	7,9617	6,9077	5,8122
17	33,4087	30,1910	27,5871	8,6718	7,5642	6,4078
18	34,8053	31,5264	28,8693	9,3905	8,2308	7,0149
19	36,1909	32,8523	30,1435	10,1170	8,9065	7,6327
20	37,5662	34,1696	31,4104	10,8508	9,5908	8,2604
21	38,9322	35,4789	32,6706	11,5913	10,2829	8,8972
22	40,2894	36,7807	33,9244	12,3380	10,9823	9,5425
23	41,6384	38,0756	35,1725	13,0905	11,6886	10,1957
24	42,9798	39,3641	36,4150	13,8484	12,4012	10,8564
25	44,3141	40,6465	37,6525	14,6114	13,1197	11,5240
26	45,6417	41,9232	38,8851	15,3792	13,8439	12,1982
27	46,9629	43,1945	40,1133	16,1514	14,5734	12,8785
28	48,2782	44,4608	41,3371	16,9279	15,3079	13,5647
29	49,5879	45,7223	42,5570	17,7084	16,0471	14,2565
30	50,8922	46,9792	43,7730	18,4927	16,7908	14,9535

Продолжение прилож.4

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
31	52,1914	48,2319	44,9853	19,2806	17,5387	15,6555
32	53,4858	49,4804	46,1943	20,0719	18,2908	16,3622
33	54,7755	50,7251	47,3999	20,8665	19,0467	17,0735
34	56,0609	51,9660	48,6024	21,6643	19,8063	17,7892
35	57,3421	53,2034	49,8019	22,4650	20,5694	18,5089
36	58,6192	54,4373	50,9985	23,2686	21,3359	19,2327
37	59,8925	55,6680	52,1923	24,0749	22,1056	19,9602
38	61,1621	56,8955	53,3835	24,8839	22,8785	20,6914
39	62,4281	58,1201	54,5722	25,6954	23,6543	21,4262
40	63,6907	59,3417	55,7585	26,5093	24,4330	22,1643
41	64,9501	60,5606	56,9424	27,3256	25,2145	22,9056
42	66,2062	61,7768	58,1240	28,1441	25,9987	23,6501
43	67,4594	62,9904	59,3035	28,9647	26,7854	24,3976
44	68,7095	64,2015	60,4809	29,7875	27,5746	25,1480
45	69,9568	65,4102	61,6562	30,6123	28,3662	25,9013
46	71,2014	66,6165	62,8296	31,4390	29,1601	26,6572
47	72,4433	67,8207	64,0011	32,2676	29,9562	27,4159
48	73,6826	69,0226	65,1708	33,0981	30,7545	28,1770
49	74,9195	70,2224	66,3387	33,9303	31,5549	28,9407
50	76,1539	71,4202	67,5048	34,7643	32,3574	29,7067

Приложение 5

Критические точки распределения Дарбина-Уотсона

n	k=1		k=2		k=3		k=4		k=5	
	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U
6	0.61	1.40	-	-	-	-	-	-	-	-
7	0.7	1.36	0.47	1.9	-	-	-	-	-	-
8	0.76	1.33	0.56	1.78	0.37	2.29	-	-	-	-
9	0.82	1.32	0.63	1.7	0.46	2.13	-	-	-	-
10	0.88	1.32	0.7	1.64	0.53	2.02	-	-	-	-
11	0.93	1.32	0.66	1.6	0.6	1.93	-	-	-	-
12	0.97	1.33	0.81	1.58	0.66	1.86	-	-	-	-
13	1.01	1.34	0.86	1.56	0.72	1.82	-	-	-	-
14	1.05	1.35	0.91	1.55	0.77	1.78	-	-	-	-
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.1	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.9	1.71	0.78	1.9	0.67	2.1
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.4	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.2	1.41	1.1	1.54	1	1.68	0.9	1.83	0.79	1.99
21	1.2	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.22	1.43	1.15	1.54	1.05	1.66	0.96	1.8	0.86	1.94
23	1.24	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.9	1.92
24	1.26	1.45	1.19	1.55	1.1	1.66	1.01	1.78	0.93	1.9
25	1.27	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.29	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.3	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.32	1.48	1.26	1.56	1.18	1.65	1.1	1.75	1.03	1.85
29	1.33	1.48	1.27	1.56	1.2	1.65	1.12	1.74	1.05	1.84
30	1.34	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.35	1.5	1.3	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.36	1.5	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.37	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.38	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.39	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.8
36	1.4	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.8
37	1.41	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.8

Продолжение прилож.5

N	k=1		k=2		k=3		k=4		k=5	
	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U	d _L	d _U
38	1,42	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,6	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,6	1,34	1,66	1,29	1,72	1,23	1,79
41	1,45	1,55	1,4	1,6	1,35	1,66	1,29	1,72	1,24	1,78
42	1,46	1,55	1,41	1,61	1,36	1,66	1,31	1,72	1,25	1,78
43	1,46	1,56	1,41	1,61	1,37	1,66	1,32	1,72	1,27	1,78
44	1,47	1,56	1,42	1,61	1,37	1,66	1,33	1,72	1,28	1,78
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
46	1,48	1,57	1,44	1,62	1,39	1,67	1,34	1,72	1,3	1,77
47	1,49	1,57	1,44	1,62	1,39	1,67	1,35	1,72	1,31	1,77
48	1,49	1,58	1,45	1,62	1,41	1,67	1,36	1,72	1,32	1,77
49	1,5	1,58	1,46	1,62	1,41	1,67	1,37	1,72	1,32	1,77
50	1,5	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
51	1,51	1,59	1,47	1,63	1,43	1,67	1,38	1,72	1,34	1,77
52	1,51	1,59	1,47	1,63	1,43	1,68	1,39	1,72	1,35	1,77
53	1,52	1,59	1,48	1,63	1,44	1,68	1,4	1,72	1,36	1,77
54	1,52	1,6	1,48	1,64	1,45	1,68	1,41	1,72	1,37	1,77
55	1,53	1,6	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
56	1,53	1,6	1,49	1,64	1,46	1,68	1,42	1,72	1,38	1,77
57	1,54	1,61	1,5	1,64	1,46	1,68	1,43	1,72	1,39	1,77
58	1,54	1,61	1,5	1,65	1,47	1,68	1,43	1,72	1,39	1,77
59	1,54	1,61	1,51	1,65	1,47	1,69	1,44	1,73	1,4	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,5	1,7	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,7	1,49	1,74	1,46	1,77
75	1,6	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,6	1,7	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,7	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,6	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Приложение 6

Критические значения коэффициентов корреляции для уровня значимости 0,05

k	1	2	3	4	5	6	7	8	9	10	11
$r_{кр}$	1,00	1,00	0,88	0,81	0,75	0,71	0,67	0,63	0,60	0,58	0,55
k	12	13	14	15	16	17	18	19	20	25	30
$r_{кр}$	0,53	0,51	0,50	0,48	0,47	0,46	0,44	0,42	0,42	0,38	0,35
k	35	40	45	50	60	70	80	90	100		
$r_{кр}$	0,32	0,30	0,29	0,27	0,25	0,29	0,22	0,21	0,19		

Литература

1. Эконометрика: Учебник/И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. - 2-е изд., перераб. и доп. - М.: Юрайт, 2012.
2. Практикум по эконометрике: Учеб. пособие/И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др.; Под ред. И.И. Елисеевой. - М.: Финансы и статистика, 2005.
3. Вуколов Э.А. Основы статистического анализа.- М., Форум, 2008.
4. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. - М., Дело, 2004..
5. Доугерти К. Введение в эконометрику. /Пер. с англ. - М.: МГУ, ИНФРА-М, 2001.
6. Новак Эдвард. Введение в методы эконометрики. Сборник задач: Пер. с польск./Под ред. И.И. Елисеевой. - М.: Финансы и статистика, 2004. - 248 с.
7. Орлова И.В. Экономико-математические методы и модели. Выполнение расчетов в среде EXCEL/Практикум: учебное пособие для вузов. - М.: ЗАО "Финстатинформ", 2000. - 136 с.

Св. план 2015 г., поз. 182

Ишханян Маргарита Владимировна

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Учебное пособие для направления «Строительство»

Подписано в печать

Формат 60 X 84 / 16

Заказ №

Усл. - печ. л. -

Тираж - 100 экз.
