МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ АРХИТЕКТУРНО-СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ»
(КазГАСУ)
ИНСТИТУТ ЭКОНОМИКИ И УПРАВЛЕНИЯ В СТРОИТЕЛЬСТВЕ

Учебно- методическое пособие

по дисциплине «Математическое моделирование и информационные технологии в городском хозяйстве» для студентов направления подготовки 08.03.01 «Строительство», профиль «Техническая эксплуатация объектов жилищно- коммунального хозяйства и городской инфраструктуры», «Городское строительство и хозяйство»
очной формы обучения

Кафедра экспертизы и
управления недвижимостью

Казань 2018

УДК 338
ББК
Х20

Х20 «Техническая эксплуатация объектов жилищно- коммунального хозяйства и городской инфраструктуры», «Городское строительство и хозяйство»: Учебно-методическое пособие по дисциплине «Математическое моделирование и информационные технологии в городском хозяйстве» для студентов направления 08.03.01 «Строительство», профиль «Техническая эксплуатация объектов жилищно-коммунального хозяйства и городской инфраструктуры», «Городское строительство и хозяйство» очной формы обучения

Составитель: Л.Н. Устинова. Казань: КГАСУ, 2018.- с.39

Печатается по решению Редакционно- издательского совета Казанского государственного архитектурно- строительного университета.

В данном Учебно- методическом пособии изложены цели и задачи освоения дисциплины, приведена теоретическая база основных понятий и особенностей математического моделирования и практические задания для освоения основных навыков

Рецензент

Кандидат экономических наук, доцент, заведующая кафедрой «Общего менеджмента» ИЭУФ КФУ, **Палей Т.Ф.**

УДК 338
ББК
©Казанский государственный архитектурно- строительный университет, 2018
© Устинова Л.Н., 2018

Содержание

1. Сущность основных понятий 4
1.1 Модели и моделирование 4
1.2.Описательное моделирование 5
1.3.Нормативное моделирование 8
1.4 Формы представления моделей 8
2. Практические занятия 11
2.1 Метод парной корреляции 11
2.2 Регрессионный анализ 17
2.2.1 Проверка на адекватность уравнения регрессии 29
2.2.2 Проверка на адекватность коэффициентов регрессии 33
Заключение 34
Список литературы 35
Приложения 36

1. Сущность основных понятий

1.1 Модели и моделирование.

Исследуя объекты окружающего мира, мы вынуждены как-то отображать результаты исследования для того, чтобы, с одной стороны, представить их и виде, удобном для анализа, а с другой для их хранения и передачи в пространстве или времени. Проектируя, создавая что-то новое, мы первоначально формируем некоторый образ этого нового. Управляя чем-либо, мы, как правило, пытаемся анализировать, к каким последствиям приведет управление. В связи с этим в познавательной и практической деятельности человека большую, если не ведущую, роль играют модели и моделирование. Особенно незаменимо моделирование при работе со сложными объектами (в частности, экономическими). Все это делает моделирование важнейшим инструментом системного анализа.

Модель в широком понимании — это образ (в том числе условный или мысленный) какого-либо объекта или системы объектов, используемый при определенных условиях в качестве их «заместителя» или «представителя». Модель — это упрощенное подобие объекта, которое воспроизводит интересующие нас свойства и характеристики объекта-оригинала или объекта проектирования.

К примеру, моделью Земли служит глобус, а звездного неба — экран планетария. Чучело животного есть его модель, а фотография на паспорте или любой перечень паспортных данных — модель владельца паспорта. Моделирование связано с выяснением или воспроизведением свойств какого-либо реального или создаваемого объекта, процесса или явления с помощью другого объекта, процесса или явления.

Моделирование — это построение, совершенствование, изучение и применение моделей реально существующих или проектируемых объектов (процессов и явлений).

Почему мы прибегаем к использованию моделей вместо попыток «прямого взаимодействия с реальным миром»? Можно назвать три основные причины.

Первая причина — сложность реальных объектов. Число факторов, которые относятся к решаемой проблеме, выходит за пределы человеческих возможностей. Поэтому одним из выходов (а часто единственным) в сложившейся ситуации является упрощение ситуации с помощью моделей, в результате чего уменьшается разнообразие этих факторов до уровня восприимчивости специалиста.

Вторая причина — необходимость проведения экспериментов. На практике

встречается много ситуаций, когда экспериментальное исследование объектов ограничено высокой стоимостью или вовсе невозможно (опасно, вредно, ограниченность науки и техники на современном этапе). Третья причина — необходимость прогнозирования. Важное достоинство моделей состоит в том, что они позволяют «заглянуть в будущее», дать прогноз развития ситуации и определить возможные последствия принимаемых решений.
Среди других причин можно назвать следующие:

• исследуемый объект либо очень велик (модель Солнечной системы), либо очень мал (модель атома);

• процесс протекает очень быстро (модель двигателя внутреннего сгорания) или очень медленно (геологические модели);

• исследование объекта может привести к его разрушению (модель самолета, автомобиля).

1.2.Описательное моделирование

Человек в своей деятельности обычно вынужден решать две задачи — экспертную и конструктивную.
В экспертной задаче на основании имеющейся информации описывается прошлое, настоящее и предсказывается будущее. Суть конструктивной задачи заключается в том, чтобы создать нечто с заданными свойствами. Для решения экспертных задач применяют так называемые описательные модели, а для решения конструктивных — нормативные

Описательные модели (дескриптивные, познавательные) предназначены для описания свойств или поведения реальных (существующих) объектов. Они являются формой представления знаний о действительности.
Примеры. План города, отчет о деятельности фирмы, психологическая характеристика личности.
Можно назвать следующие цели описательного моделирования в зависимости от решаемых задач:
• изучение объекта (научные исследования) — наиболее полно и точно отразить свойства объекта;
• управление — наиболее точно отразить свойства объекта в рабочем диапазоне изменения его параметров;
• прогнозирование — построить модель, способную наиболее точно прогнозировать поведение объекта в будущем;
• обучение — отразить в модели изучаемые свойства объекта. Построение

описательной модели происходит по следующей схеме: наблюдение, кодирование, фиксация (рис. 1).

Рис. 1. Последовательность построения описательной модели.

Модель объекта можно построить, только наблюдая за ним. То, что мы наблюдаем, необходимо закодировать либо с помощью слов, либо символов, в частности, математических, либо графических образов, либо в виде физических предметов, процессов или явлений. И наконец, закодированные результаты наблюдения надо зафиксировать в виде модели. Отражение свойств объекта в модели не является полным в силу разных причин: особенностей восприятия, наличия и точности измерительных приборов, потребности и, наконец, психического состояния субъекта. Если обозначить полную информацию об объекте через $I_о$, а воспринимаемую информацию — $I_в$, то отражение математически можно сформулировать следующим образом:

$$I_В=F(I_о),$$

где $I_в \subset I_о$, или в линейном приближении (рис. 2):

$$I_В=k_с k_и k_ц k_п I_о,$$

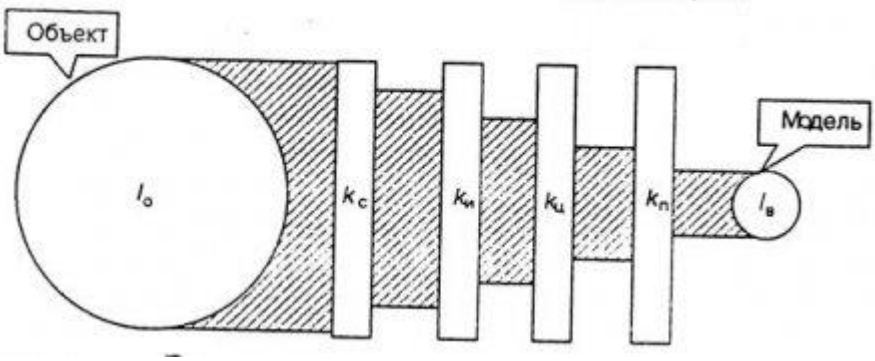


Рис. 2. Фильтрация информации об объекте.

где $k_с$ – информационная проницаемость среды – свойство среды по передаче информации от объекта к субъекту ($0 \leq k_с \leq 1$);
$k_и$ – коэффициент измерительной способности (вооруженности) субъекта – способность субъекта воспринимать (измерять) информацию ($0 < k_и < 1$);
$k_ц$ – целевая избирательность субъекта – связана с потребностью в конкретных свойствах объекта ($0 < k_ц < 1$);
$k_п$ – психологическая избирательность субъекта – связана с его психологическим состоянием ($0 < k_п \leq 1$).

Стоит отметить, что все что ни делает человек, в том числе и построение моделей, он вкладывает свою точку зрения. Это, в частности, может привести к тому, что мы принимаем свою точку зрения за единственную, а карту местности — за саму местность, которую она представляет. Существуют следующие субъективные факторы, влияющие на качество создаваемых моделей.

Избирательность. Модель строится на основании наблюдений за объектом, но человек замечает свойства объекта избирательно. На это влияют образование, мировоззрение, опыт, а также настроение, чувства, заботы и общее самочувствие. В результате формируется модель, не отвечающая целям моделирования.

Конструирование — обратный аналог избирательности: мы начинаем видеть то, чего нет. Мы заполняем пробелы в информации о мире, чтобы он приобрел некий смысл и предстал перед нами в том виде, каким, по нашему мнению, он должен быть. Длительная эволюция воспитала нас дополнять увиденные фрагменты до полного образа: если мы видим из-за дерева голову волка, то мысленно дорисовываем его туловище и хвост. Поэтому когда при исследовании объекта мы получаем неполную информацию о нем, то невольно заполняем информационные «пробелы», исходя из своего опыта. В результате можем получить модель, не адекватную объекту.

Искажение. Искажение проявляется в том, что мы строим модели окружающего мира, выделяя одни его составляющие за счет замалчивания других. В частности, искажение лежит в основе творческих способностей (поэта, художника, композитора) и некоторых болезней, например паранойи.

Обобщения. Пользуясь обобщением, мы создаем мысленные модели, взяв за основу один случай и обобщив его на все возможные случаи. Обобщение является основой статистических выводов, но при условии так называемой репрезентативной (представительной) выборки ситуаций. Опасность обобщения состоит в том, что, взяв какую-либо ситуацию, человек расценивает ее как типичную и распространяет извлеченные из нее выводы на все сходные, по его мнению, ситуации (что, в частности, и является основой суеверия).

Таким образом, не все свойства объекта нам доступны из-за свойств окружающей среды, а из доступных не все мы можем измерить или оценить. Из тех, что можем измерить, не все нам необходимы. Из необходимых свойств мы не все из них адекватно воспринимаем из-за психического состояния (невнимательности, субъективного предпочтения, страха и т. п.). На основании воспринимаемой информации об объекте IB и формируется его

образ, называемый моделью.

1.3.Нормативное моделирование

Моделировать можно не только то, что существует, но и то, чего еще нет. Нормативные модели (прескриптивные, прагматические) предназначены для указания целей деятельности и определенного порядка (алгоритма) действий для их достижения.
Цель — образ желаемого будущего, т. е. модель состояния, на реализацию которого и направлена деятельность.
Алгоритм — образ (модель) будущей деятельности.
При нормативном моделировании обычно не используют слово «модель» — чаще говорят «проект», «план».
Это проекты машин, зданий; планы застройки; законы; уставы организаций и должностные инструкции, бизнес-планы, программы действий, управленческие решения.

Описательные модели отражают существующее, их развитие направлено на приближение модели к реальности (в структурном системном анализе такая модель называется моделью
Нормативные модели показывают не существующее, но желаемое. Здесь решается задача приближения реальности к модели, поскольку модель играет роль стандарта или образца, под который «подгоняются» как сама деятельность, так и ее результаты (в структурном системном анализе она называется моделью «Как должно быть» — «То-Ве»).

Выше была рассмотрена классификация моделей по целевому назначению. Кроме того, познавательные и прагматические модели можно классифицировать по характеру выполняемых функций, форме, зависимости объекта моделирования от времени.

1.4 Формы представления моделей

Применительно к естественно-техническим, социально-экономическим и другим наукам принято различать следующие виды моделирования:
• концептуальное моделирование, при котором с помощью некоторых специальных знаков, символов, операций над ними или с помощью естественного или искусственного языков истолковывается основная мысль (концепция) относительно исследуемого объекта;
• интуитивное моделирование, которое сводится к мысленному эксперименту на основе практического опыта работников (широко применяется в

экономике);

• физическое моделирование, при котором модель и моделируемый объект представляют собой реальные объекты или процессы единой или различной физической природы, причем между процессами в объекте-оригинале и в модели выполняются некоторые соотношения подобия, вытекающие из схожести физических явлений;

• структурно-функциональное моделирование, при котором моделями являются схемы, (блок-схемы), графики, чертежи, диаграммы, таблицы, рисунки, дополненные специальными правилами их объединения и преобразования:

• математическое (логико-математическое) моделирование, при котором моделирование, включая построение модели, осуществляется средствами математики и логики;

• имитационное (программное) моделирование, при котором логико-математическая модель исследуемого объекта представляет собой алгоритм функционирования объекта, реализованный в виде программного комплекса для компьютера.

Перечисленные выше виды моделирования не являются взаимоисключающими и могут применяться при исследовании сложных объектов либо одновременно, либо в некоторой комбинации. Отдельно следует сказать о компьютерном моделировании, являющемся развитием имитационного моделирования.

Компьютерное моделирование. Первоначально под компьютерным моделированием (или, как говорили, моделированием на ЭВМ) понималось лишь имитационное моделирование. Исторически случилось так, что первые работы по компьютерному моделированию были связаны с физикой. Затем разработанные подходы распространились на задачи химии, электроэнергетики, биологии и некоторые другие дисциплины, причем схемы моделирования не слишком отличались друг от друга. Этот вид моделирования все еще широко распространен и в научных, и прикладных исследованиях.

Однако сегодня понятие «компьютерное моделирование» чаще связывают не с фундаментальными дисциплинами, а в первую очередь с системным анализом. Следует заметить, что компьютер может быть весьма полезен при всех видах моделирования (за исключением физического моделирования, где компьютер тоже может использоваться, но, скорее, для целей управления процессом моделирования). Изменилось и понятие компьютерной модели. Раньше под компьютерной моделью чаще всего понимали имитационную модель — отдельную программу, совокупность программ или программный

комплекс, позволяющий с помощью последовательности вычислений и графического отображения их результатов воспроизводить (имитировать) процессы функционирования объекта. В настоящее время под компьютерной моделью чаще всего понимают структурно-функциональную модель — условный образ объекта, описанный с помощью взаимосвязанных компьютерных таблиц, блок-схем, диаграмм, графиков, рисунков, анимационных фрагментов, гипертекстов и отображающий структуру и взаимосвязи между элементами объекта.

Таким образом, мы видим, что понятие «компьютерное моделирование» значительно шире традиционного понятия «моделирование на ЭВМ» и нуждается в уточнении, учитывающем сегодняшние реалии.

Компьютерное моделирование — это метод решения задачи анализа или синтеза объекта на основе использования его компьютерной модели.

Суть компьютерного моделирования заключена в получении количественных и качественных результатов по имеющейся модели. Качественные выводы, получаемые по результатам анализа, позволяют обнаружить неизвестные ранее свойства объекта. Количественные выводы в основном носят характер прогноза некоторых будущих или объяснения прошлых значений переменных, характеризирующих систему.

Предметом компьютерного моделирования могут быть: экономическая деятельность фирмы или банка, промышленное предприятие, информационно-вычислительная сеть, технологический процесс, любой реальный объект или процесс, например процесс инфляции. Цели компьютерного моделирования могут быть различными, однако наиболее часто моделирование является, как уже отмечалось ранее, центральной процедурой системного анализа.

2. Практические занятия

2.1 Метод парной корреляции

Коэффициент корреляции - это статистический показатель зависимости двух случайных величин. Коэффициент корреляции может принимать значения от -1 до +1. При этом, значение -1 будет говорить об обратной корреляции между величинами, 0 - о нулевой корреляции, а +1 - о полной прямой корреляции величин. Т.е., чем ближе значение коэффициента корреляции к +1, тем сильнее связь между двумя случайными величинами.

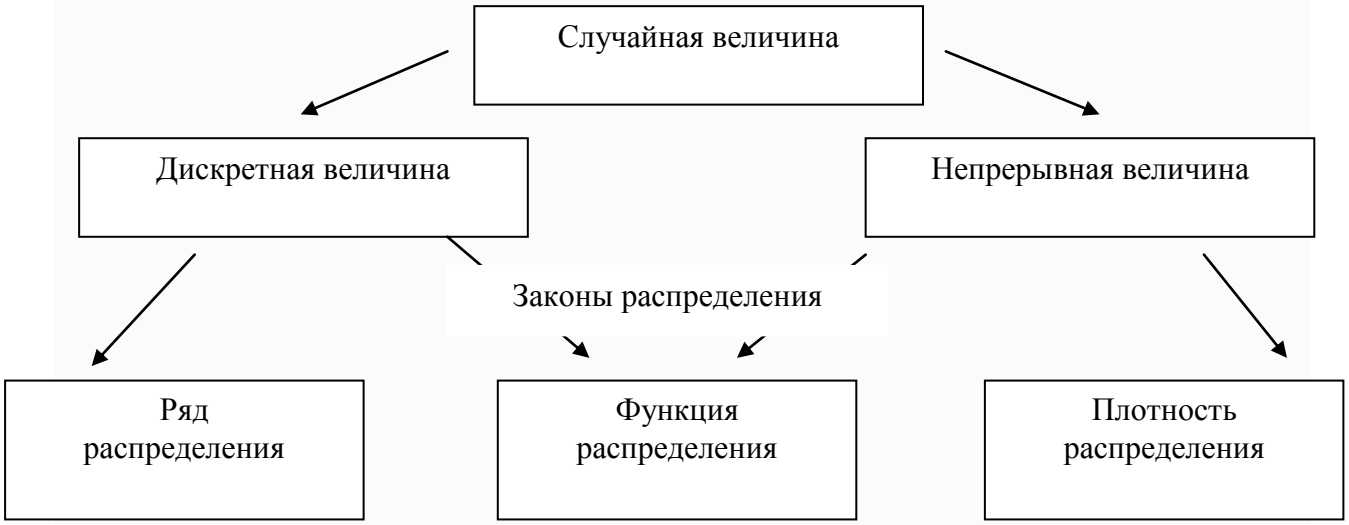


Рис. 3 Виды случайных величин и их законы распределения

Коэффициент корреляции - это корреляционное отношение, математическая мера корреляции двух случайных величин. В случае, если изменение одной случайной величины не ведёт к закономерному изменению другой случайной величины, но приводит к изменению другой статистической характеристики данной случайной величины, то подобная связь не считается корреляционной, хотя и является статистической.

Коэффициент корреляции - это мера линейной зависимости двух случайных величин в теории вероятностей и статистике. Некоторые виды коэффициентов корреляции могут быть положительными или отрицательными. В первом случае предполагается, что мы можем определить только наличие или отсутствие связи, а во втором - также и её направление.

Коэффициенты корреляции могут быть положительными и отрицательными. Если при увеличении значения одной величины происходит уменьшение значений другой величины, то их коэффициент корреляции отрицательный. В случае, когда увеличение значений первого объекта наблюдения приводит к увеличениям значения второго объекта, то можно говорить о положительном коэффициенте.

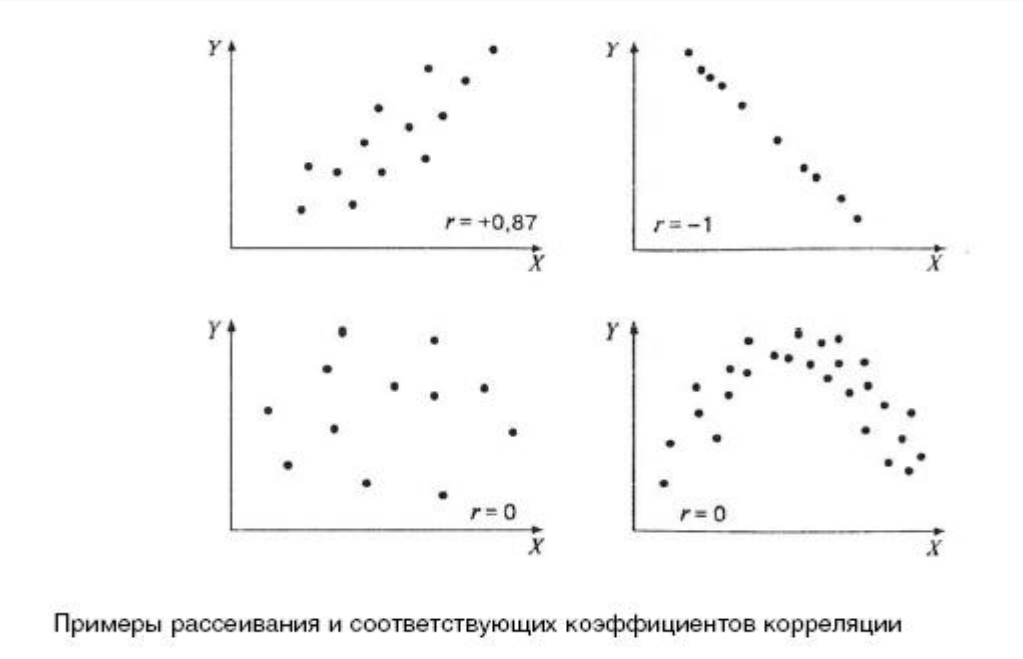


Рис. 4 Примеры рассеивания и соответствующих коэффициентов корреляции.

По форме корреляционная связь может быть линейной или нелинейной. Более удобной для выявления и интерпретации корреляционной связи является линейная форма. Для линейной корреляционной связи можно выделить два основных направления: положительное («прямая связь») и отрицательное («обратная связь»).

Сила связи напрямую указывает, насколько ярко проявляется совместная изменчивость изучаемых переменных. В психологии функциональная взаимосвязь явлений эмпирически может быть выявлена только как вероятностная связь соответствующих признаков. Наглядное представление о характере вероятностной связи дает диаграмма рассеивания – график, оси которого соответствуют значениям двух переменных, а каждый испытуемый представляет собой точку.

В качестве числовой характеристики вероятностной связи используют коэффициенты корреляции, значения которых изменяются в диапазоне от –1 до +1. После проведения расчетов исследователь, как правило, отбирает только наиболее сильные корреляции, которые в дальнейшем интерпретируются (табл. 1).

Коэффициент корреляции +0,6 - пример заметной положительной корреляции

Коэффициент корреляции - это сила и направление связи между независимой и зависимой переменными. Значения r находятся в диапазоне между - 1.0 и + 1.0. Когда r имеет положительное значение, связь между x и y

является положительной, а когда значение r отрицательно, связь также отрицательна. Коэффициент корреляции, близкий к нулевому значению, свидетельствует о том, что между x и y связи не существует.

Таблица 1

Величина коэффициента парной корреляции	Характеристика силы связи
До 0,3	Практически отсутствует
0,3-0,5	слабая
0,5-0,7	заметная
0,7-0,9	сильная
0,9-0,99	Очень сильная

Коэффициент корреляции -0,6 - пример слабой отрицательной корреляции

Коэффициент корреляции - это объективный показатель, свидетельствующий о наличии или отсутствии связи между переменными, и измеряющий выраженность этой связи. Коэффициент корреляции был предложен как инструмент, с помощью которого можно проверить гипотезу о зависимости и измерить силу зависимости двух переменных. Сразу заметим, что коэффициент корреляции оказался не идеальным инструментом, он пригоден лишь для измерения силы линейной зависимости.

<u>Задача №1</u>

Предприятие располагается по адресу:665719 Иркутская обл., Центральный округ, г. Братск, ул. Энгельса 19.

ООО «Жилтрест№1» создано с целью обеспечения сохранности и эксплуатации жилого фонда, санитарного содержания микрорайонов города, удовлетворения потребностей населения в жилищно-коммунальных услугах. Предметом деятельности является техническое обслуживание, ремонт, эксплуатация и санитарное содержание жилищного фонда и придомовых территорий, а так же нежилых, служебных и производственных помещений.

Предприятия часть прибыли используется предприятием в соответствии со сметой, согласованной с собственником имущества. Источниками формирования финансовых ресурсов ООО «Жилтрест№1» являются:

- квартирная плата;
- возмещение за услуги других предприятий;

- возмещение из бюджета на покрытие выпадающих доходов;
- средства на капитальный ремонт жилищного фонда;
- возмещение из бюджета на погашение разницы в тарифах на коммунальные услуги;
- прочие собственные доходы.

Методом парной корреляции определить наиболее воздействующий фактор на формирование прибыли предприятия, определить величину связи этих показателей, оценить ее силу, если известны следующие данные (Табл. 2):

Таблица 2

Прибыль предприятия	Тарифы на коммунальные услуги	Количество заключенных договоров с подрядными организациями	Качество предоставляемых услуг
250	61	68	5
158	53	54	4
136	48	65	5
144	65	66	3
187	77	51	4
216	87	49	4
255	95	40	4

Алгоритм решения задачи:

1. Открываем программу Microsoft Excel и переносим данные таблицы 1 (вместе с шапкой);

2. Ниже создаем таблицу 3:

Таблица 3

Прибыль предприятия	Коэффициент корреляции
Тарифы на коммунальные услуги	
Количество заключенных договоров с подрядными организациями	
Качество предоставляемых услуг	

3. В итоговой таблице активизируем правую верхнюю пустую ячейку (коэффициент корреляции для тарифов на коммунальные услуги);

4. Запускаем Мастер Функций, ищем в инструментальной строке значок f_x и в всплывающем диалоговом окне выбираем категорию - ***СТАТИСТИЧЕСКИЕ,*** функция- ***КОРРЕЛ;***

5. В появившейся панели ***КОРРЕЛ*** нужно заполнить текстовые поля для ***МАССИВ*** 1 (указать диапазон ячеек столбца1 табл. 1 и для ***МАССИВ*** 2 указываем столбец 2 таблицы 1. Таким образом, мы будем исследовать взаимосвязь двух факторов – изменение прибыли предприятия и тарифов на коммунальные услуги.

6. Нажимаем «ОК» и видим полученный коэффициент корреляции;

7. Аналогичным образом находим коэффициент корреляции для двух других факторов.

8. Делается вывод на основе интерпретации результатов.

Интерпретация результатов:

Коэффициент корреляции в общем виде показывает насколько заметна теснота связи между исследуемыми параметрами. Вариативность коэффициента корреляции колеблется между 0 и 1. Чем выше значение коэффициента корреляции r, тем сильнее взаимосвязь между параметрами. Однако данный коэффициент корреляции может принимать и отрицательное значение – это говорит об обратной связи между параметрами.

Таблица 4

Качественная оценка тесноты связи

Величина коэффициента парной корреляции	Характеристика силы связи
До 0,3	Практически отсутствует
0,3-0,5	слабая
0,5-0,7	заметная
0,7-0,9	сильная
0,9-0,99	Очень сильная

Для возможности построения прогнозов необходимо провести статистическую оценку (проверку значимости).

Статистическую оценку проводят путем сравнения его абсолютной величины с табличным (или критическим) показателем $r_{крит}$, значения которого приведены в Приложении 1.

Если окажется, что $|\ r_{расч} \geq r_{крит}\ |$, то с заданной степенью вероятности (обычно 95%) можно утверждать, что между рассматриваемыми числовыми совокупностями существует значимая линейная связь. В случае, если равенство не выполняется, делается вывод об отсутствии значимой связи.

<u>Задача №2 Тема: «Метод парной корреляции</u>»

ООО Промышленная компания — одна из крупнейших независимых компаний, осуществляющая поставки металлопроката на рынки России и стран СНГ. Компания занимается оптовой и розничной торговлей металлопроката, а также развивает производственное направление: осуществляет переработку металла и производит различные профили.

Промышленная компания предлагает полный ассортимент металлопроката, представленный практически всеми металлургическими комбинатами России и ряда зарубежья. На складах Компании всегда в наличии порядка 250 тыс. тонн металлопроката.

Рис. 5 Структура производимой продукции Промышленной компании

Таблица 5

Выручка предприятия, млн.р.	Арматура	Лист Г/К 2-12 мм.	Швеллер	Лист Х/К
2,5	0,52	0,32	0,22	0,02
2,8	0,54	0,33	0,25	0,024
3,2	0,55	0,37	0,26	0,026
2,9	0,51	0,31	0,2	0,019
4	0,56	0,38	0,24	0,025
3,1	0,51	0,35	0,21	0,02
2,8	0,49	0,33	0,2	0,017
2,85	0,495	0,33	0,26	0,02
2,97	0,5	0,34	0,27	0,024
3,3	0,58	0,36	0,29	0,025

Методом парной корреляции определить наиболее воздействующий фактор на формирование прибыли предприятия, определить величину связи этих показателей, оценить ее силу, если известны следующие данные (Табл. 5): Проверку значимости провести, используя данные таблицы 4.

2.2 Регрессионный анализ

В практике статистического исследования весьма часто возникает необходимость определить не только корреляционное соотношение между изучаемыми характеристиками, но и установить определенную обусловленность между ними, представив выявленную связь в строгой аналитической форме. В этом случае результат исследования – экспериментальная зависимость воздействия какого-либо фактора (скажем, производительности труда, уровня образования, практического стажа работы и т.д.) на изменение изучаемого параметра (например, величины прибыли фирмы) − может быть не только представлен в виде графика (что весьма наглядно), но и описан математически с использованием аппроксимирующего выражения. Исследование такой ситуации и является задачей регрессионного анализа, который дает предсказание (прогнозирование) одной переменной на основании другой. Регрессионный анализ четко распределяет роли между изучаемыми характеристиками − одна из них является аргументом, а вторая функцией. Переменная, которая прогнозируется (функция), обозначается как y, а переменная, которая используется для такого прогнозирования (аргумент или фактор), − это x. Таким образом, вести в корреляции дается попытка определения наличия или отсутствия связи между величинами, то целью регрессионного анализа является определение не только силы и характера связи, но и количественное влияние одной переменной на аргумент.

Для определения коэффициентов уравнения регрессии b применяют разные методы (графический, метод средних), однако наибольшее распространение получил метод наименьших квадратов (МНК).

Пусть обсуждается некоторая зависимость $y = f(x)$, которая отражает какой-то процесс, имеющий плавное течение, и поэтому все параметры системы изменяются постепенно, без скачков. В этих случаях экспериментальные точки, нанесенные на графике, должны бы укладываться на некоторую плавную кривую (в частном случае, прямую). Однако на практике определенный разброс экспериментальных точек всегда

наблюдается, что связано с изменчивостью (ошибками) регистрируемых измерений. Понятно, что такого разброса удалось бы избежать, если бы результаты измерений оказались совершенно свободными от ошибок, и тогда точки, отвечающие этим результатам, строго ложились бы на соответствующую плавную кривую, или прямую линию. Поэтому все процессы, которые имеют заведомо плавное течение, принято изображать также плавными кривыми, проводя их не через точки, а так, чтобы кривая проходила по возможности ближе ко всем точкам на графике.

Однако такое указание оставляет при построении кривых определенный произвол. Его частично можно устранить основным положением МНК: сумма квадратов отклонений εi экспериментальных точек от кривой по вертикальному направлению, т.е. сумма квадратов величин εi, должна быть наименьшей (Σεi 2 = минимум).

Довольно часто при описании аппроксимирующей функции ограничиваются простым видом полиноминальной зависимости, полагая ее линейной, т.е. в виде уравнения прямой $y = b_0 + b_1x$. Здесь свободный член b_0 характеризует сдвиг и равен тому значению y, которое получается при $x = 0$, а коэффициент b_1 определяет наклон линии. Отыскание коэффициентов b_0 и b_1 осуществляется по МНК.

Пусть имеется n экспериментальных точек (n пар наблюдений): (x_1, y_1); (x_2, y_2);… (x_n, y_n). Введем следующие обозначения: y_i – это измеренные (экспериментальные) значения изучаемого параметра, а $\hat{y}_i$ – его теоретические (рассчитанные по уравнению) показатели.

Предположим, что экспериментальные точки на графике укладываются так, что по ним вполне возможно провести прямую линию. Значения функции $\hat{y}_i$ в этом случае можно записать в виде линейного уравнения: $\hat{y}_i = b_0 + b_1 x_i$. Расстояние по ординате (вертикали) от точки y_i до прямой составит: $b_0 + b_1 x_i - y_i = \varepsilon_i$, где $b_0 + b_1 x_i = \hat{y}_i$ − рассчитанное (теоретическое) значение функции; y_i − ее измеренное (опытное) значение и ε_i − разница (расстояние) между $\hat{y}_i$ и y_i.

Метод наименьших квадратов. Решение задач с использованием регрессионного анализа в системе «Excel» по показателям экономической деятельности предприятий.

При изучении связи показателей коммерческой деятельности применяются различного вида уравнения прямолинейной и криволинейной связи. Формально могут возникать ситуации двух типов:

1. Вид функциональной зависимости *неизвестен*. В этом случае нужно решить предварительно задачу, направленную на отыскание подходящей функциональной зависимости. Это достаточно сложная задача, но она успешно решается современными средствами информационных технологий (программа Excel).

2. Вид функциональной зависимости *известен* и требуется только найти ее *параметры* (коэффициенты регрессии $b0$, $b1$, $b2$, …).

Термином ***линейный регрессионный анализ*** обозначают такое прогнозирование, которое описывается линейной взаимосвязью между исследуемыми переменными: $y = b0 + b1x$.

В случае *криволинейных* зависимостей применяются математические функции следующего вида:

гиперболическая $y = b0 + b1 /x$;

показательная $y = b0 + b1/x$;

степенная $y = b0x\ b1$;

параболическая $y = b0 + b1x + b2x2$;

логарифмическая $y = b0 + b1\lg x$;

экспоненциальная $y = b0exp\ (b1x)$ и другие.

Решение математических уравнений связи предполагает вычисление по исходным данным их параметров (*свободного члена* $b0$ и *коэффициентов регрессии* $b1$, $b2$, …).

При всем разнообразии эмпирических формул все же имеется вид аналитической зависимости, получивший широкое распространение. Им является уравнение регрессии в виде *многочленов (полинома)*, расположенных по восходящим степеням изучаемого фактора и одновременно линейных ко всем коэффициентам.

Такая формула имеет вид:

$y = f(x) = b0 + b1x + b2x2 + ... + bm\ xm$,

где $b0$, $b1$, $b2$,…, bm − коэффициенты, подлежащие определению.

Предположим, что экспериментальные точки на графике укладываются так, что по ним вполне возможно провести прямую линию (рис.11). Значения функции ŷi в этом случае можно записать в виде линейного уравнения:

ŷi = b0 + b1 xi . Расстояние по ординате (вертикали) от точки yi до прямой составит: b0 + b1 xi − yi = ε

i, где b0 + b1 xi = ŷi − рассчитанное (теоретическое)

значение функции; yi − ее измеренное (опытное) значение и εi − разница (рас стояние) между ŷi и yi.

Задание – определить как взаимосвязаны экономические показатели деятельности коммерческого предприятия и какой фактор в наибольшей

степени влияет на результативный признак, к примеру, на выручку предприятия. Объем выборки составит

Для этого на рабочем листе Excel предварительно создадим таблицу с исходными данными, в которой укажем содержимое табл.4. Причем саму таблицу построим по столбцам и поместим ее в ячейках A1:C9.

Итоговый результат показан на рис.6.

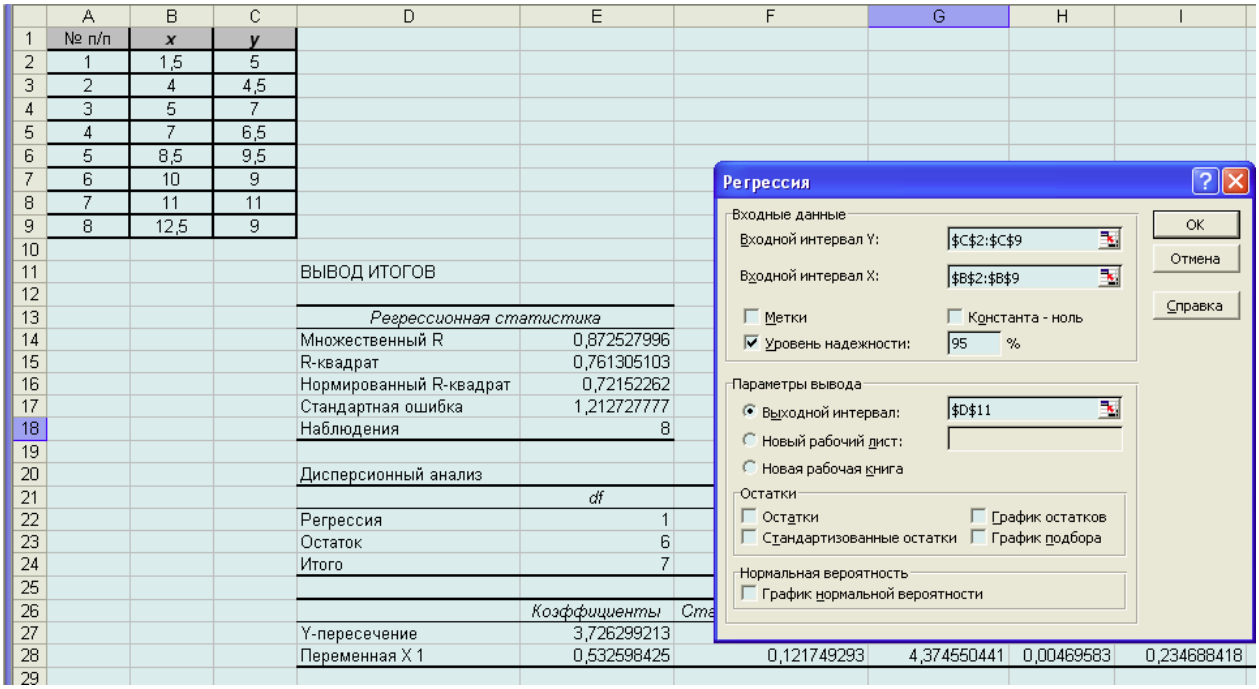


Рис. 6 Лист Excel с результатами расчета коэффициентов регрессии

Далее будем действовать привычным образом:

– в главном меню запустим серию команд Сервис/Анализ данных/Регрессия;

– в появившемся диалоговом окне заполним поля ввода данных для обоих параметров y и x; для этого в каждое окно (Входной интервал Y и Входной интервал X) поместим наши данные, выделив их предварительно в соответствующих столбцах (напомним, что для функции y ее данные «сидят» в третьем столбце C2:C9, а для переменной x – во втором, т.е. B2:B9; при этом выделяются только те ячейки, которые содержат исключительно числовые показатели);

– отметим Уровень надежности (доверительную вероятность), равный 95 %;

– укажем в окне вывода Выходной интервал ту ячейку, от которой будет формироваться весь блок получаемых статистических показателей, это D11;

– после чего нажмем кнопку OK.

На рис.6 в собранном виде представлены все упомянутые элементы – исходная таблица (в верхнем левом углу), заполненное диалоговое окно

Регрессия и, наконец, рассчитанные статистические показатели под заголовком «Вывод итогов».

Таким образом, на примере предложенной задачи мы познакомились с программным продуктом, позволяющим легко и быстро получить итоговый результат. После вычисления коэффициентов полученное уравнение регрессии надлежит подвергнуть проверке на адекватность.

Лабораторная работа №6

Тема: Модель множественной регрессии. Решение задач с использованием регрессионного анализа в системе «Excel» по показателям экономической деятельности отрасли.

До сих пор нами рассматривалась ситуация, когда на зависимую переменную (функцию) воздействовал только один фактор (аргумент). Подобное прогнозирование принято называть простой регрессией. Такие зависимости мы уже рассмотрели ранее.

Однако в подавляющем большинстве случаев приходится иметь дело с экспериментальными данными, касающимися влияния более чем одного фактора. Прогнозирование единственной переменной y на основании нескольких переменных xk называется множественной регрессией. В этом случае математическая модель процесса представляется в виде уравнения регрессии с несколькими переменными величинами, т.е. $y = f(b0, \ldots, xk)$. Общий вид уравнения множественной регрессии обычно стараются представить в форме линейной зависимости:

$y = b0 + b1x1 + b2x2 + \ldots + bkxk$,

где b0 – свободный член (или сдвиг); b1, b2 , …, bk − коэффициенты регрессии, которые подлежат вычислению методом наименьших квадратов.

При анализе уравнения множественной регрессии (как и в случае простой регрессии) используется также такое понятие, как ошибка прогнозирования Δy. Последняя понимается как разность между рассчитанным (теоретическим) значением функции $\hat{y}i$ и ее измеренным (опытным) значением yi,

т.е. $\Delta y = \hat{y}i - yi$.

Статистический вывод о пригодности (значимости) уравнения обычно проверяется в следующей последовательности.

1. Сначала проводится общая проверка методом F-теста, целью которой является выяснение, объясняют ли x-переменные значимую долю вариации y, т.е. превалирует ли влияние факторов xk на изменение функции y над ее колебаниями случайного порядка; если регрессия не является значимой, то говорить больше не о чем.

2. Если регрессия оказывается значимой, то можно продолжить анализ, используя t-тесты для отдельных коэффициентов регрессии; в этом случае пытаются выяснить, насколько значимой является влияние той или иной переменной x на параметр y при условии, что все другие факторы xk остаются неизменными. Построение доверительных интервалов и проверка гипотез на адекватность для отдельного коэффициента регрессии основывается на определении стандартной ошибки. Каждый коэффициент регрессии имеет свою стандартную ошибку Sb1, Sb2,…, Sbk.

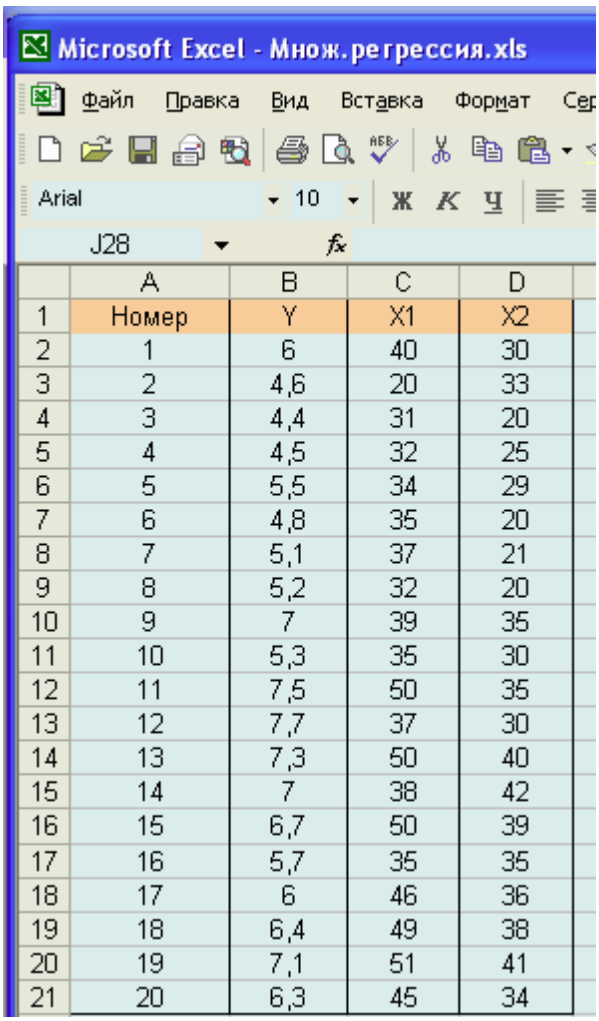
К примеру, на предприятии происходит снижение объемов продаж произведенной продукции, перед руководством стоит задача- определить какие факторы способствуют данному снижению и минимизировать их воздействие. Таким образом стоит задача:

- написать уравнение множественной регрессии;
− оценить статистическую значимость уравнения;
− определить значимость коэффициентов регрессии и пояснить характер влияния исследуемых факторов.

Таблица 6

Годы	Объем выручки, y, т.р.	С/с изгот. Изделий, x1, р.	Ассорт. Продукции, x2, шт	Годы	Объем выручки, У, т.р.	С/с изгот. Изделий, x1, р.	Ассорт. Продукции, x2, шт.
1997	6	40	30	2007	7,5	50	35
1998	4,6	20	33	2008	7,7	37	30
1999	4,4	31	20	2009	7,3	50	40
2000	4,5	32	25	2010	7	38	42
2001	5,5	34	29	2011	6,7	50	39
2002	4,8	35	20	2012	5,7	35	35
2003	5,1	37	21	2013	6	46	36
2004	5,2	32	20	2014	6,4	49	38
2005	7	39	35	2015	7,1	51	41
2006	5,3	35	30	2016	6,3	45	34

Запускаем Excel и воспроизводим в табличной форме имеющиеся исходные результаты.



Microsoft Excel - Множ.регрессия.xls

Файл Правка Вид Вставка Формат Сер

Arial 10 Ж К Ч

J28 fx

	A	B	C	D
1	Номер	Y	X1	X2
2	1	6	40	30
3	2	4,6	20	33
4	3	4,4	31	20
5	4	4,5	32	25
6	5	5,5	34	29
7	6	4,8	35	20
8	7	5,1	37	21
9	8	5,2	32	20
10	9	7	39	35
11	10	5,3	35	30
12	11	7,5	50	35
13	12	7,7	37	30
14	13	7,3	50	40
15	14	7	38	42
16	15	6,7	50	39
17	16	5,7	35	35
18	17	6	46	36
19	18	6,4	49	38
20	19	7,1	51	41
21	20	6,3	45	34

Рис. 7 Данные в программе Excel

В данном случае все экспериментальные данные (по каждой позиции) представляем в виде самостоятельных колонок (рис.7). Размещаем всю таблицу в ячейках от A1 до D21, при этом сами исходные данные (т.е. для y и x1, x2) будут находиться в диапазоне B1: D21. Рис.14. Лист Excel с исходными табличными результатами После этого получим сводную таблицу основных статистических характеристик для функции y. Для этого воспользуемся известным методом анализа данных – программой Описательная статистика.

Предпримем следующие шаги:

− в главном меню выбираем последовательно пункты Сервис/Анализ данных/Описательная статистика, после чего щелкаем по кнопке ОК; − заполняем диалоговое окно для ввода данных и параметров вывода. Чтобы получить их, проделаем следующие манипуляции (рис.15): а) укажем Входной интервал (в виде абсолютных ссылок $B$1:$D$21), т.е. адресуем все ячейки, в которых находятся значения функции y и аргументов x1, x2 ; б) отметим способ Группирования (в нашем случае по столбцам); в) откроем

флажок для Метки, показывающий, что первая строка содержит название столбца;

г) выделим Выходной интервал, для этого достаточно указать левую верхнюю ячейку будущего диапазона ($F$1);

д) установим флажки, показывающие, что нам нужна информация в виде Итоговой статистики, а также Уровень надежности, равный 95 %; после чего нажмем кнопку ОК.

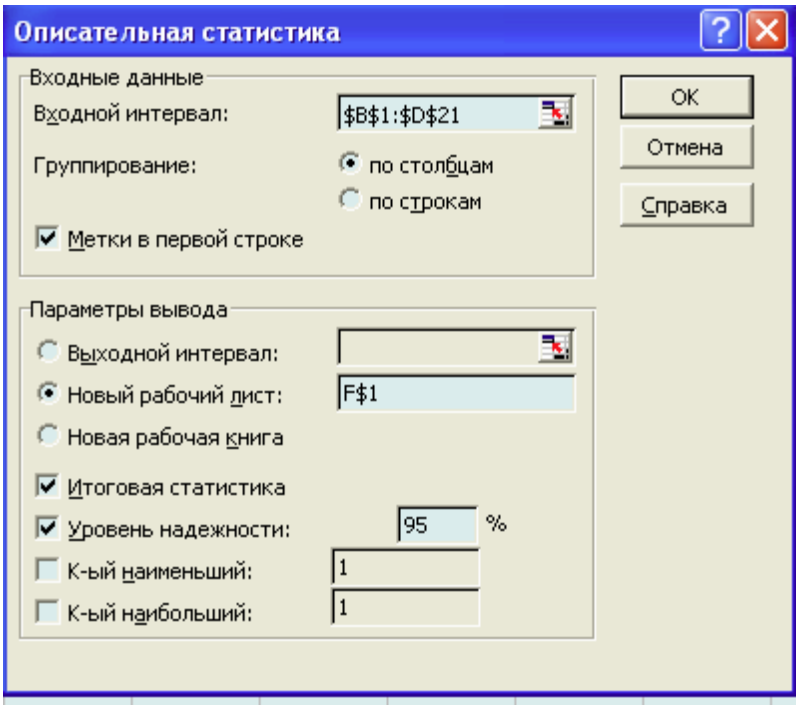


Рис. 8 Диалоговое окно ввода параметров Регрессия

Полученные результаты статистического расчета показаны на рис.9 в виде соответствующего листа Excel.

Из представленного комплекта статистических показателей выберем те, которые нам потребуются для последующего анализа − среднее арифметическое и стандартное отклонение (среднеквадратичное отклонение) Sn.

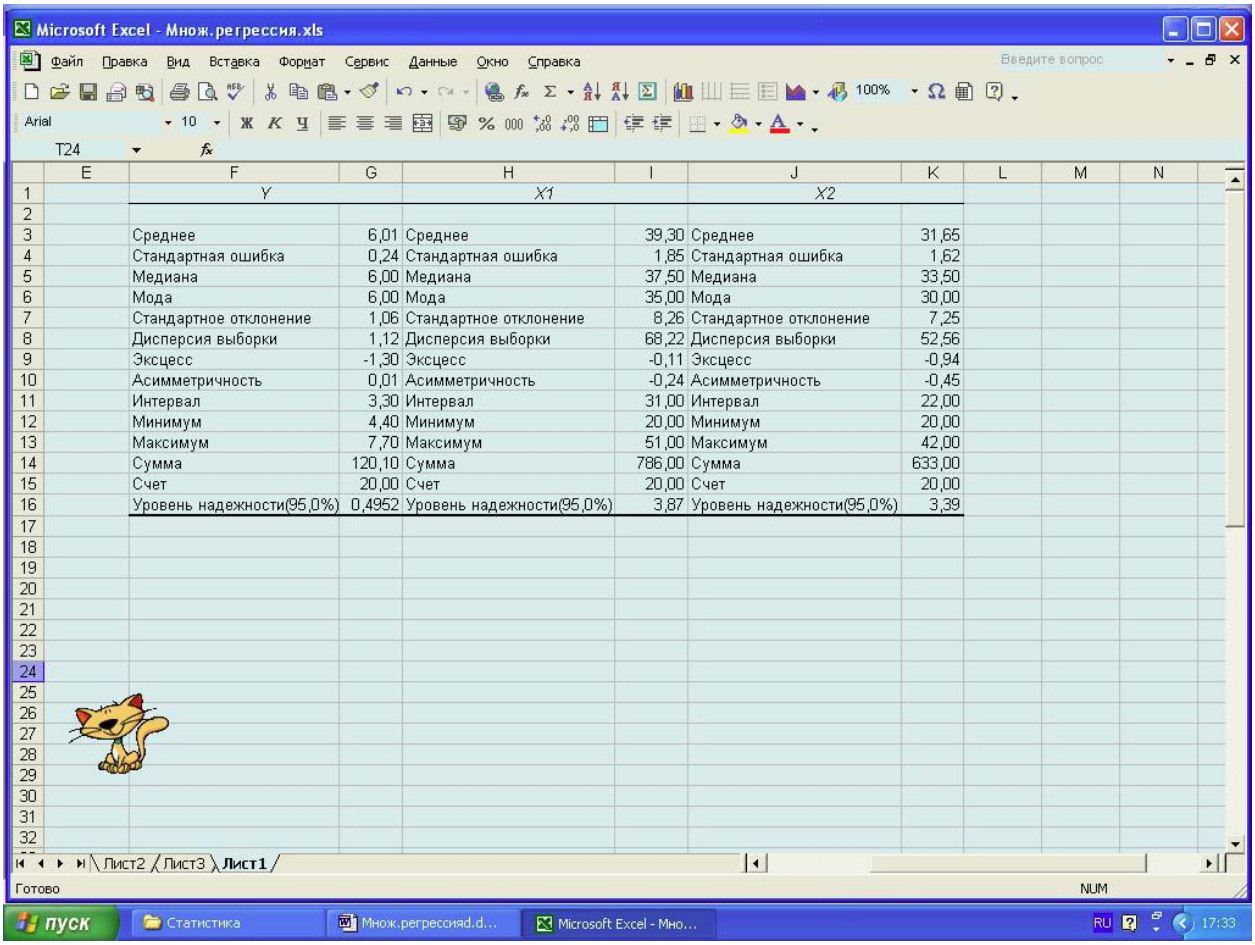
В табл.8 приведены названные статистические показатели для функции y и обеих переменных x1 и x2. Отметим, что для функции y ее среднее арифметическое −y составляет 6,01, а стандартное отклонение (среднеквадратичное отклонение) Sn равно 1,06.

Статистические показатели для функции y и переменных x1 и x2.

Таблица 7

Показатели	y	x1	x2
Среднее	6,01	39,3	31,65
Стандартное	1,06	8,26	7,25

отклонение			



	Y		X1		X2	
Среднее		6,01	Среднее	39,30	Среднее	31,65
Стандартная ошибка		0,24	Стандартная ошибка	1,85	Стандартная ошибка	1,62
Медиана		6,00	Медиана	37,50	Медиана	33,50
Мода		6,00	Мода	35,00	Мода	30,00
Стандартное отклонение		1,06	Стандартное отклонение	8,26	Стандартное отклонение	7,25
Дисперсия выборки		1,12	Дисперсия выборки	68,22	Дисперсия выборки	52,56
Эксцесс		-1,30	Эксцесс	-0,11	Эксцесс	-0,94
Асимметричность		0,01	Асимметричность	-0,24	Асимметричность	-0,45
Интервал		3,30	Интервал	31,00	Интервал	22,00
Минимум		4,40	Минимум	20,00	Минимум	20,00
Максимум		7,70	Максимум	51,00	Максимум	42,00
Сумма		120,10	Сумма	786,00	Сумма	633,00
Счет		20,00	Счет	20,00	Счет	20,00
Уровень надежности(95,0%)		0,4952	Уровень надежности(95,0%)	3,87	Уровень надежности(95,0%)	3,39

Рис.9 Лист Excel с результатами расчета статистических показателей

Расчет показателей регрессии также исполняется по компьютерной программе. Для ее запуска исполним следующие команды:

− в главном меню выберем пункты Сервис/Анализ данных / Регрессия, после чего щелкнем по кнопке ОК;

− заполним диалоговое окно ввода данных для параметра y и обеих характеристик x1 и x2; для этого в каждое окно (Интервал Y и Интервал X) поместим наши данные, выделив их предварительно в соответствующих столбцах (напомним, что для функции y ее данные «сидят» во втором столбце B2:B21, а для переменных x1 и x2 – в третьем и четвертом, т.е. в диапазоне ячеек C2:D21; заметим, что при этом выделяются только те ячейки, которые содержат исключительно числовые показатели);

− выделим в текстовом поле Выходной интервал ту ячейку, от которой будет формироваться весь блок получаемых статистических показателей; при этом укажем другой лист − Лист 2;

– после чего – кнопка ОК.

Заполненное диалоговое окно для программы Регрессия представлено на рис.10.

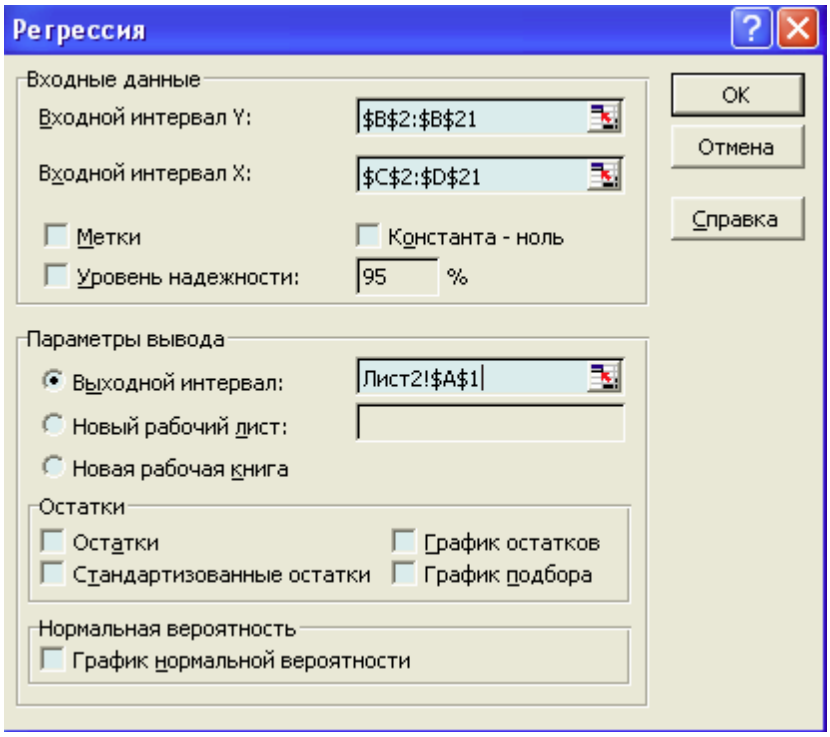


Рис.10 Диалоговое окно ввода параметров Регрессия

Excel выдает богатый набор разнообразных статистических материалов (рис.11). Выберем, однако, из них такие, которые нам потребуются для последующего анализа. Для этого создадим табл.8, в которой поместим расчетные значения коэффициентов регрессии, стандартную ошибку, величины t-критерия и показатели уровня значимости α. Укажем также (ниже таблицы) рассчитанные показатели для самой функции y.

Таблица 8

Данные регрессионной статистики

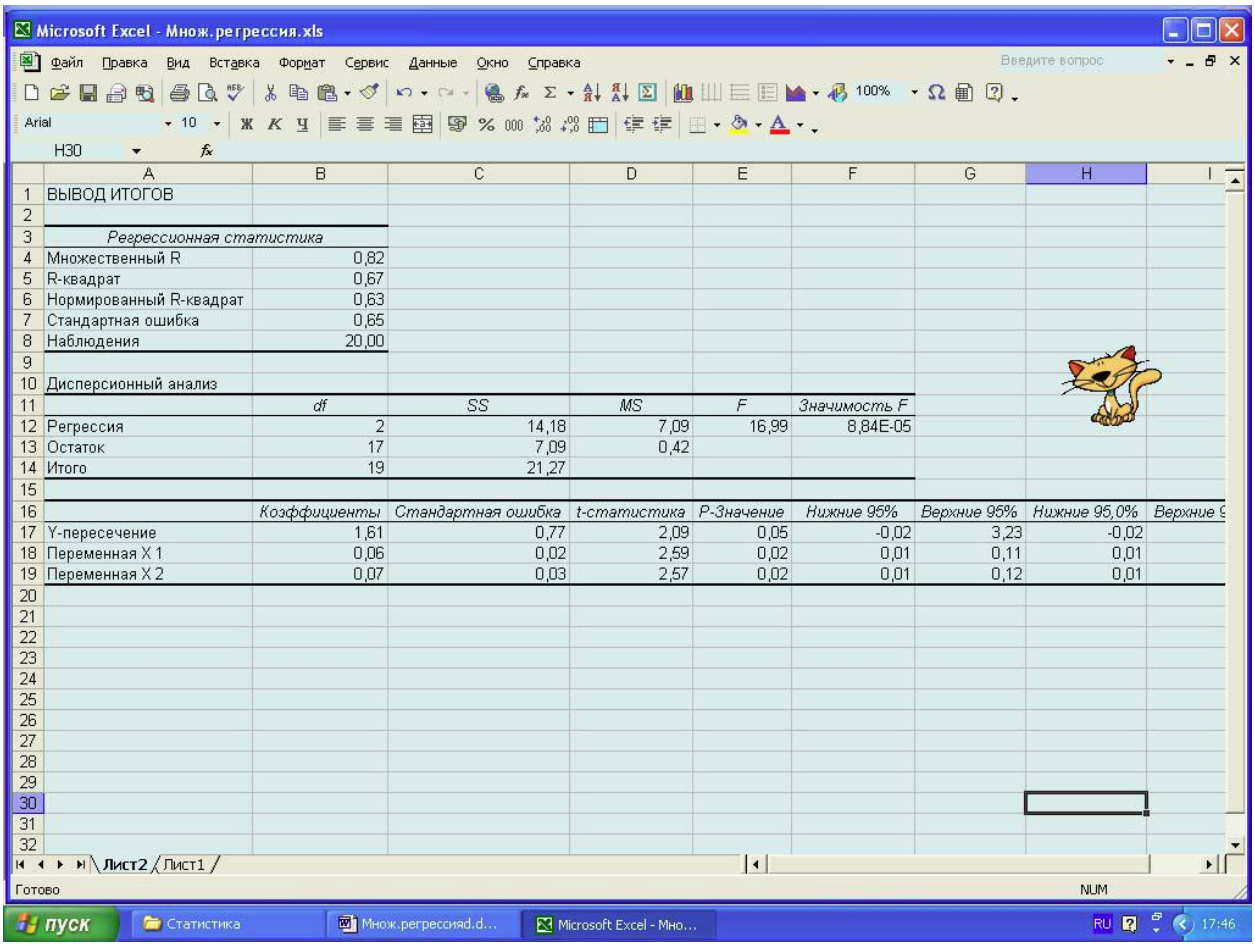
Независимая переменная	Коэффициент	Стандартная ошибка	*t*	*p* (или α)
Свободный член	1,61	0,77	2,09	0,05
*X*1	0,06	0,23	2,59	0,02
*X*2	0,07	0,03	2,57	0,02

Для функции Y: S y = 0,65; R-квадрат = 0,67; R-квадрат (нормир.) = 0,63. Таким образом, для рассматриваемого примера уравнение регрессии (или уравнение прогнозирования) будет иметь следующий вид:

ŷ (Объем выручки) = b0 + b1x1 + b2x2 = 1,61 + 0,06 (себестоимость) + 0,07 (ассортимент продукции).

Запишем полученное уравнение в окончательной редакции:

ŷ = 1,61 + 0,06 x1 + 0,07 x2.



	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	*Регрессионная статистика*								
4	Множественный R	0,82							
5	R-квадрат	0,67							
6	Нормированный R-квадрат	0,63							
7	Стандартная ошибка	0,65							
8	Наблюдения	20,00							
9									
10	Дисперсионный анализ								
11		*df*	*SS*	*MS*	*F*	*Значимость F*			
12	Регрессия	2	14,18	7,09	16,99	8,84E-05			
13	Остаток	17	7,09	0,42					
14	Итого	19	21,27						
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 9*
17	Y-пересечение	1,61	0,77	2,09	0,05	-0,02	3,23	-0,02	
18	Переменная X 1	0,06	0,02	2,59	0,02	0,01	0,11	0,01	
19	Переменная X 2	0,07	0,03	2,57	0,02	0,01	0,12	0,01	

Рис.11. Лист Excel с результатами расчета статистических показателей регрессии

Теперь займемся статистическим анализом этого уравнения регрессии.

Свободный член (сдвиг) b_0, равный 1,61, формально надлежит понимать следующим образом: объем продажи на предприятии, когда себестоимость изготовления продукции и ассортимент продукции не оказывают влияние на результирующий признак. Однако мы полагаем, что в указанной совокупности исходных данных нет подобных примеров (всегда себестоимость и ассортимент продукции влияет на объем продаж и чистую прибыль предприятия). Поэтому сдвиг b_0 следует обсуждать как вспомогательную величину, необходимую для получения оптимальных прогнозов, и не истолковывать ее столь буквально.

Коэффициенты регрессии b_1 и b_2 следует рассматривать как степень влияния каждой из переменных (себестоимость и ассортимент) на размер объема продажи, если все другие независимые переменные остаются неизменными. Так, коэффициент b_1, равный 0,06, указывает, что (при прочих равных условиях) повышение себестоимости на 1 % приводит к возрастанию объемов продаж на 0,06 тыс. руб. Анализируя коэффициент b2, можно заметить, что увеличение ассортиментного ряда продукции на 1 % приводит также к повышению продаж и этот прирост составляет почти такую же величину − 0,07 тыс. руб.

Еще раз заметим, что все названные коэффициенты регрессии отражают влияние на исследуемый параметр y только какой-то одной переменной x при непременном условии, что все другие переменные (факторы) не меняются. Например, применительно к коэффициенту b2 это нужно понимать так: указанное влияние ассортиментного ряда проявляется при условии, когда сохраняется статично себестоимостью изготовления продукции.

Можно воспользоваться двумя приемами для оценки добротности выполненного нами регрессионного анализа. В статистике для этого используют:

− стандартную ошибку (S−y), которая дает представление о приблизительной величине ошибки прогнозирования;

− коэффициент детерминации (R^2), указывающий, какой процент вариации функции y объясняется воздействием факторов xk.

Рассмотрим оба подхода более подробно.

1. Результаты статистического расчета показывают, что стандартная ошибка для функции составляет 0,65. Этот результат применительно к нашему примеру следует рассматривать следующим образом: фактическая величина объема продаж отличается от прогнозируемых показателей не более чем на 0,65 тыс.руб. Однако ценность этого показателя невелика, если не надежность этого утверждения. При условии сохранения нормального распределения можно полагать, что примерно 2/3 фактических данных будут находиться в пределах S− y от прогнозируемых показателей; примерно 95 %.

Эта стандартная ошибка S−y, равная 0,65, указывает отклонение фактических данных от прогнозируемых на основании использования воздействующих факторов x1 и x2. В то же время мы располагаем обычным стандартным отклонением Sn, равным 1,06, которое было рассчитано для одной переменной, а именно: сами текущие значения yi и величина среднего арифметического− y, которое равно 6,01. Легко видеть, что S−y < Sn; следовательно, ошибки прогнозирования, как правило, оказываются

меньшими, если использовать уравнение регрессии (учитывается вклад факторов x1 и x2), а не ограничиваться только значением– y. Сказанное можно истолковать следующим образом. Если бы нам ничего не было известно про переменные x1 и x2, то в качестве оптимальной приблизительной величины среднего уровня продаж пришлось бы использовать показатель –y = 6,01 т.р. и полагать, что наши прогнозы дают ошибку Sn, равную 1,06 т.р.

Такое сокращение погрешности прогнозирования с 1,06 до 0,65 и является одним из преимуществ использования регрессионного анализа.

Если вновь обратиться к нашему примеру, то коэффициент детерминации R^2 (на рис.11 Excel его подает как R-квадрат) равен 0,67, что составляет 67 %. Этот результат следует толковать так: все исследуемые воздействующие факторы (себестоимость и ассортимент продукции) объясняют 67 % вариации анализируемой функции. Остальное же (33 %, что весьма прилично!) остается необъясненным и может быть связано с влиянием других, неучтенных факторов.

Для нашего примера показатель R^2 (67 %) считается умеренным, поэтому можно полагать, что именно эти два фактора в данном конкретном случае оказывают наиболее значительное влияние на $\bar{y}$. Итак, нами получено уравнение множественной регрессии, коэффициенты которого bi формально показывают, как и в каком направлении действуют (пока лишь вероятно!) исследуемые факторы xk i и какой процент изменчивости функции y объясняется влиянием именно этих факторов. Теперь нам надлежит определить статистическую значимость полученного аналитического выражения.

2.2.1 Проверка на адекватность уравнения регрессии

При проверке значимости модели принято придерживаться следующей последовательности действий:

1. Сначала выполняется общая проверка полученного уравнения на пригодность.
2. Если результат оказался положительным (уравнение значимо), то проверяют на значимость уже каждый коэффициент уравнения регрессии bi.
3. Дается сравнительная оценка степени влияния каждого из анализируемых факторов xk.

Статистическую оценку полученного уравнения (так называемый статистический вывод) принято начинать с проведения F-теста, целью которого является выяснение способности исследуемых факторов xk

объяснять значимую часть колебания функции y. Этот тест используется как своеобразные «входные ворота» в статистический вывод: если результат теста значим, то связь существует, значит приступать к ее исследованию и объяснению. Если проверка указывает на незначимость связи, то заключение лишь одно: мы имеем дело с набором случайных чисел, никак не связанных между собой. И больше делать нечего, так как нет предмета для анализа.

Заметим при этом, что сам формальный факт отсутствия значимости на деле может и не соответствовать отсутствию взаимосвязи как таковой. Просто в указанных обстоятельствах у нас не хватило экспериментальных данных доказать, что такая связь вообще-то есть. Иначе говоря, она может и быть, но из-за малого размера выборки или какой-либо случайности нам не удалось ее доказать на основании тех опытных данных, которые были в нашем распоряжении.

Использование так называемой нулевой гипотезы для F-теста означает, что между переменными xk и y значимая связь отсутствует. Следовательно, признается, что параметр y является чисто случайной величиной, поэтому значения переменных xk не оказывают на него никакого систематического влияния. Применительно к уравнению регрессии это утверждение можно трактовать как случай, когда все коэффициенты уравнения равны нулю.

С другой стороны, альтернативная гипотеза F-теста говорит о том, что между параметром y и переменными xk существует определенная прогнозирующая взаимосвязь. Следовательно, параметр y уже не является чисто случайной величиной и должен зависеть хотя бы от одной из переменных xk. Тем самым альтернативная гипотеза настаивает на том, что по крайней мере один из коэффициентов регрессии отличен от нуля. Как видно, здесь принимается во внимание следующее обстоятельство: совершенно необязательно, чтобы каждая x-переменная влияла на параметр y, вполне достаточно, чтобы влияла хотя бы одна из них.

Для выполнения F-теста воспользуемся результатами компьютерного расчета, который исполнил замечательный Excel. Здесь обычно рекомендуются следующие приемы.

1. Решение принимается на основе критерия Фишера.

Это достаточно традиционный способ, им привычно пользуются при статистических анализах, хотя по удобству и простоте он может уступать другим методам.

Обычно F-тест проводится путем сопоставления вычисленного значения F-критерия с эталонным (табличным) показателем Fтабл для

соответствующего уровня значимости. Если выполняется неравенство Fрасч < Fтабл, то уверенностью, например на 95 %, можно утверждать, что рассматриваемая зависимость y = b0 + b1x1 + b2x2 +…+ bkxk является статистически значимой. В противном случае наоборот.

2. Решение принимается на основе уровня значимости α.

Для этого обратим внимание на представленные значения уровня значимости α (в интерпретации Excel это показатель p). Если p-значение больше, чем 0,05, то полученный результат нужно трактовать как незначимый (для 95-процентной вероятности). В том случае, когда величина p оказывается меньше 0,05, то вывод такой: это значимое уравнение с вероятностью 95%.

Если же p < 0,01, то полученный результат является высоко значимым, (степень риска ошибиться в нашем утверждении оказывается меньше 1 %, т.е. степень надежности составляет 99 %)

3. Решение принимается на основе коэффициента детерминации R_2.

В этом случае имеющуюся расчетную величину R_2 расчет (это то, что нам выдал Excel, см. рис.18) необходимо сравнить с табличными (критическими) значениями $R_{2\ крит}$ для соответствующего уровня значимости (повторим еще раз, обычно это 0,05). Если окажется, что $R_{2расч} > R2_{крит}$, то с упомянутой степенью вероятности (95 %) можно утверждать, что анализируемая регрессия является значимой.

Теперь проанализируем наше уравнение с использованием рассмотренных статистических критериев.

1. Проведем проверку по F-критерию. Компьютерная распечатка выдала нам величину Fрасч, равную 16,99. С учетом сделанных замечаний для анализа уравнения будем пользоваться величиной $F_{расч}$, обратной представленной Excel. Она составит 1:16,99 = 0,06.

Отыщем по эталонной таблице (прил.3) критическую величину $F_{крит}$ при условии, что для числителя степень свободы f1 = k, т.е. составит 2 (число воздействующих факторов равно 2), а для знаменателя f2 = n − k− 1 = 20 − 2 − 1= 17. Тогда будем иметь следующие значения для $F_{крит}$: 3,6 (для α = 0,05), 6,2 (α = 0,01) и 10,5 (α = 0,001). Понятно, что для всех рассмотренных вероятностей выполняется соотношение Fрасч < Fкрит, поэтому уверенно можно говорить о высокой степени адекватности анализируемого уравнения.

2. Теперь выполним проверку с использованием уровня значимости α (еще раз напомним, что Excel этот показатель именует как p). На рис.11 , где

дано изображение листа Excel, находим позицию «Значимость F». Там указана величина 8,84E-5, т.е. это число 8,84, перед которым стоит 5 нулей. Фактически можно признать, что α = 0,000. Это говорит о том, что действительно обнаруживается устойчивая зависимость рассматриваемой функции y от воздействующих факторов x1 и x2, т.е. обе переменные не является чисто случайной величиной. Правда, нам пока неизвестно, какие именно факторы (оба x1 и x2 или какой-то один из них) реально участвует в прогнозировании, но нам доподлинно понятно, что по крайней мере один из них влияет непременно.

3. Напомним, что, по нашим расчетам, коэффициент детерминации $R_{2расч}$ составляет 0,67, или 67 %. Таблица для тестирования на уровне значимости 5 % в случае выборки n = 20 и числа переменных k = 2 дает критическое значение $R_{2крит}$ = 0,297 (прил.4). Поскольку выполняется соотношение $R_{2расч} > R_{2крит}$, то с вероятностью 95 % можно утверждать о наличии значимости данного уравнения регрессии.

Кстати заметим, что для наших обстоятельств (n = 20, k = 2) можно оценить критическое значение $R_{2крит}$ для α=0,01 (высокая значимость) и α = 0,001 (высшая степень значимости). В этом случае $R_{2крит}$ составляет соответственно 0,384 и 0,517, что, как видно, все равно остается меньше расчетного показателя $R_{2расч}$, т.е. 0,67. Из чего следует заключить, что обсуждаемое нами уравнение действительно характеризуется очень высокой степенью значимости.

Как видно, все три рассмотренных приема статистической проверки дают одинаковый результат. В этом примере мы воспользовались подобным разнообразием способов анализа только с одной целью – дать представление о существующих методах такой проверки. На практике же нет нужды проводить статистическую оценку с использованием всех указанных вариантов.

Вполне разумно (да и экономично) ограничиться каким-то одним методом. Каким именно? Более распространенным методом считается выполнение проверки по F-критерию.

Итак, нами проведена проверка на значимость самого уравнения, т.е. мы понимаем, что существует взаимосвязь между параметром y и переменными xk. Однако нам пока неясно, каково влияние конкретных факторов x1 и x2 на исследуемую функцию y: действуют ли оба фактора или только какой-то один из них. Поэтому предстоит определить значимость отдельных коэффициентов регрессии b1 и b2. Для этой цели используется так называемый t-тест.

2.2.2 Проверка на адекватность коэффициентов регрессии

Проверку на адекватность коэффициентов регрессии рекомендуется проводить по следующим эквивалентным методам.

1. Использование t-критерия. Необходимые расчеты делает исполнительный Excel, который выдает соответствующую компьютерную распечатку с обозначением значений показателя t. Анализируемый коэффициент считается значимым, если его t-критерий по абсолютной величине превышает 2,00 (точнее 1,96), что соответствует уровню значимости 0,05. В нашем примере имеем для коэффициентов b0, b1 и b2 следующие показатели критерия Стьюдента: tb0 = 2,09; tb1 = 2,59 и tb2 = 2,57. Из всего вышесказанного следует, что значимыми оказываются все коэффициенты нашего уравнения.

2. Использование уровня значимости. В этом случае оценка проводится путем анализа показателя p, т.е. уровня значимости α. Коэффициент признается значимым, если рассчитанное для него p-значение (эти данные выдает Excel) меньше (или равно) 0,05 (т.е. для 95 %-ной доверительной вероятности). Показатель p составляет для коэффициентов b0, b1 и b2 следующие величины: p b0 = 0,05; p b1= 0,02 и p b2 = 0,02.

Эти данные позволяют также заключить, что все рассмотренные коэффициенты статистически значимы. Иначе говоря, можно сделать вывод о неслучайном характере влияния всех изученных параметров.

Таким образом, проверка обоими методами дает вполне согласованные результаты. Поэтому в окончательном виде наше уравнение регрессии (для уровня значимости 0,05) следует записать так: $\hat{y} = 1{,}61 + 0{,}06\ x1 + 0{,}07\ x2$.

Заключение

В настоящее время в век информационных технологий и IT достижений высокая значимость математического моделирования очевидна, поскольку любые разработки и технические изобретения – это прежде всего моделирование.

Сегодня общепризнанная точка зрения по поводу места математики и экономико-математического моделирования в экономических исследованиях определяется следующим:

1. Математика и экономико-математическое моделирование один из способов получения принципиально новых знаний в экономической науке;
2. Математическая формализация есть средство для превращения многих экономических процессов и явлений в наглядные и понятные;
3. Экономико-математическое моделирования и ЭВМ есть мощный инструмент, обеспечивающий возможность быстрых и многовариантных расчетов экономических процессов и явлений.

Экономико-математическое моделирование в экономической науке заключается в совершенствовании системы экономической информации. Математические методы позволяют упорядочить систему экономической информации, выявлять недостатки в имеющейся информации и вырабатывать требования для подготовки новой информации или ее корректировки. Кроме того, формализация экономических задач и применение ЭВМ многократно ускоряют типовые, массовые расчеты, повышают точность и сокращают трудоемкость, позволяют проводить многовариантные экономические обоснования сложных мероприятий, недоступные при господстве «ручной» технологии. Благодаря применению метода моделирования значительно усиливаются возможности конкретного количественного анализа: изучение взаимодействия многих факторов, оказывающих влияние на экономические процессы, количественная оценка последствий изменения условий развития экономических объектов.

Список литературы

1. Аверченков В.И. Основы математического моделирования технических систем [Электронный ресурс] : учебное пособие / В.И. Аверченков, В.П. Федоров, М.Л. Хейфец. — Электрон. текстовые данные. — Брянск: Брянский государственный технический университет, 2012. — 271 с. — 5-89838-126-0. — Режим доступа: http://www.iprbookshop.ru/7003.html
2. Ахмадиев Ф. Г. Математическое моделирование и методы оптимизации [Текст] : учеб.пособие / КГАСУ. - Казань : КГАСУ, 2017. - 178с. - ISBN 978-5-7829-0534-7 : 200.00.
1. Бабак С.В. Экономико-математическое моделирование экономических систем при прогнозировании развития и функционирования нефтяного сектора России и зарубежных стран [Электронный ресурс] / С.В. Бабак. — Электрон. текстовые данные. — М. : Геоинформмарк, Геоинформ, 2006. — 76 с. — 5-98877-019-3. — Режим доступа: http://www.iprbookshop.ru/16882.html
2. Бараз В.Р. Корреляционно-регрессионный анализ связи показателей коммерческой деятельности с использованием программы Excel : учебное пособие / В.Р. БАРАЗ. – Екатеринбург : ГОУ ВПО «УГТУ–УПИ», 2005. – 102 с.
3. Беляев Н.Н. Математическое моделирование в задачах экологической безопасности и мониторинга чрезвычайных ситуаций [Текст] : монография. - Днепропетровск : Акцент ПП, 2013. - 159с. - ISBN 978-966-2607-97-0 : 100.00.
4. Федосеев В.В. Математическое моделирование в экономике и социологии труда. Методы, модели, задачи [Электронный ресурс] : учебное пособие / В.В. Федосеев. — Электрон. текстовые данные. — М. : ЮНИТИ-ДАНА, 2012. — 167 с. — 5-238-01114-8. — Режим доступа: http://www.iprbookshop.ru/15390.html

Приложение 1

Критические значения корреляции $\tau_{крит}$
для уровня значимости α и степени свободы *f*

f \ α	0,1	0,05	0,01
1	0,988	0,997	0,999
2	0,9	0,950	0,990
3	0,805	0,878	0,959
4	0,729	0,811	0,917
5	0,669	0,754	0,874
6	0,622	0,707	0,834
7	0,582	0,666	0,798
8	0,549	0,632	0,765
9	0,521	0,602	0,735
10	0,497	0,576	0,708
11	0,476	0,553	0,684
12	0,457	0,532	0,661
13	0,441	0,514	0,641
14	0,426	0,497	0,623
15	0,412	0,482	0,606
16	0,400	0,468	0,590
17	0,389	0,455	0,575
18	0,378	0,444	0,561
19	0,369	0,433	0,549
20	0,360	0,423	0,537
25	0,323	0,381	0,487
30	0,296	0,349	0,449
35	0,275	0,325	0,418
40	0,257	0,304	0,393
45	0,243	0,287	0,372
50	0,231	0,273	0,354
60	0,211	0,250	0,325
70	0,195	0,232	0,302
80	0,183	0,217	0,283
90	0,173	0,205	0,267
100	0,164	0,196	0,254

Приложение 2

Значения F-критерия для уровня значимости α=0,05 и числа степени свободы *f*

Знаменатель: степени свободы *f*	Числитель: степени свободы *f*									
	1	2	3	4	5	6	8	12	20	30
1	161,45	199,500	215,710	224,58	230,16	234,00	238,90	243,91	248,01	250,10
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,46
3	10,13	9,55	9,28	9,12	9,01	8,94	8,86	8,74	8,66	8,62
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,80	5,75
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,56	4,50
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,87	3,81
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,15	3,08
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,77	2,70
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,54	2,47
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,12	2,04
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,93	1,84

Приложение 3

Критические значения R^2 для уровня значимости α, числа переменных (аргуметов) χ и количество опытов n

Уровень значимости		**α=0,1**			**α=0,05**			**α=0,01**		
Число переменных χ		1	2	3	1	2	3	1	2	3
Число опытов n	3	0,976			0,994			1,000		
	4	0,810	0,990		0,902	0,997		0,980	1,000	
	5	0,649	0,900	0,994	0,771	0,95	0,998	0,919	0,990	1,000
	6	0,532	0,785	0,932	0,658	0,864	0,966	0,841	0,954	0,993
	7	0,448	0,684	0,844	0,569	0,776	0,903	0,765	0,900	0,967
	8	0,386	0,602	0,759	0,499	0,698	0,832	0,696	0,842	0,926
	9	0,339	0,536	0,685	0,444	0,632	0,764	0,636	0,785	0,879
	10	0,302	0,482	0,622	0,399	0,575	0,704	0,585	0,732	0,830
	11	0,272	0,438	0,568	0,362	0,527	0,651	0,540	0,684	0,784
	12	0,247	0,401	0,523	0,332	0,486	0,604	0,501	0,641	0,740
	13	0,227	0,369	0,484	0,306	0,451	0,563	0,467	0,602	0,700
	14	0,209	0,342	0,450	0,283	0,420	0,527	0,437	0,567	0,663
	15	0,194	0,319	0,420	0,264	0,393	0,495	0,411	0,536	0,629
	16	0,181	0,298	0,394	0,247	0,369	0,466	0,388	0,508	0,598
	18	0,160	0,264	0,351	0,219	0,329	0,417	0,348	0,459	0,544
	20	0,143	0,237	0,316	0,197	0,297	0,378	0,315	0,418	0,498
	22	0,129	0,215	0,287	0,179	0,270	0,345	0,288	0,384	0,459
	24	0,118	0,197	0,263	0,164	0,248	0,317	0,265	0,355	0,426
	26	0,109	0,181	0,243	0,151	0,229	0,294	0,246	0,330	0,396
	28	0,101	0,168	0,2225	0,140	0,213	0,273	0,229	0,308	0,371
	30	0,094	0,157	0,210	0,130	0,199	0,256	0,214	0,289	0,349

Приложение 4

Значения коэффициента Стьюдента (τ-критерия) для уровня значимости α и числа измерений n

n \ α	0,1	0,05	0,01
2	6,314	12,706	63,657
3	2,920	4,303	9,925
4	2,353	3,182	5,841
5	2,132	2,776	4,604
6	2,015	2,571	4,032
7	1,943	2,447	3,707
8	1,895	2,365	3,499
9	1,860	2,306	3,355
10	1,833	2,262	3,250
11	1,812	2,228	3,169
12	1,796	2,201	3,106
13	1,782	2,179	3,055
14	1,771	2,160	3,012
15	1,761	2,145	2,977
16	1,753	2,131	2,947
17	1,746	2,120	2,921
18	1,740	2,110	2,898
19	1,734	2,101	2,878
20	1,729	2,093	2,861
21	1,725	2,086	2,845
22	1,721	2,080	2,831
23	1,717	2,074	2,819
25	1,711	2,064	2,797
27	1,706	2,056	2,779
29	1,701	2,048	2,763
31	1,697	2,042	2,750
40	1,684	2,021	2,704
60	1,671	2,000	2,660
120	1,658	1,980	2,617
∞	1,645	1,960	2,576

Редакционно-издательский отдел
Казанский государственный архитектурно-строительный университет

Подписано в печать		Формат 60x84/16
Заказ №	Печать ризографическая	Усл. печ. л. 2,4
Тираж 50 экз.	Бумага офсетная №1	Уч. –изд. л.

Издательство КГАСУ
420043, Казань, Зеленая, 1