Министерство образования Республики Беларусь
Учреждение образования
«Витебский государственный технологический университет»

В.Л. Шарстнев, Е.Ю. Вардомацкая

КОМПЬЮТЕРНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ.

ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ МОДЕЛИРОВАНИЯ И АНАЛИЗА ЗАДАЧ ЭКОНОМИКИ.

Витебск
2007

УДК 004.9
ББК 32.973
Ш 26

Р е ц е н з е н т ы : кафедра «Экономика» УО «Витебский государственный технологический университет» (зав. кафедрой кандидат технических наук, доцент Касаева Т.В.),

заведующий кафедрой информатики и информационных технологий УО «Витебский государственный университет им. П.М. Машерова» кандидат физико-математических наук, доцент А.И. Бочкин.

Рекомендовано в качестве пособия редакционно-издательским советом УО «ВГТУ», протокол №7 от 20 декабря 2006 г.

Шарстнев В.Л.

Ш 26 Компьютерные информационные технологии: пакеты прикладных программ для моделирования и анализа задач экономики: пособие / В.Л Шарстнев, Е.Ю. Вардомацкая. – Витебск: УО «ВГТУ», 2007. – 138 стр.

ISBN 985-481-069-0

Рассмотрены возможности (ТП MS Excel, ИС Statistica, СКМ Maple) для моделирования и анализа задач экономики и управления. На примерах анализа фондовых показателей предприятий легкой промышленности изложены рекомендации и методики использования утилит и библиотек этих пакетов для построения моделей экономической и управленческой деятельности предприятий и поиска оптимальных решений.

Пособие может быть использовано студентами экономического факультета при изучении курса «Компьютерные информационные технологии», курсовом и дипломном проектировании.

УДК 004.9
ББК 32.973

ISBN 985-481-069-0

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Программа курса "Компьютерные информационные технологии" предполагает изучение раздела, связанного с использованием современных пакетов прикладных программ (ППП) в предметной области.

Для студентов экономического профиля является обязательным знание и умение работать с программными продуктами, позволяющими анализировать сформулированные задачи экономического содержания. Для этого студенту необходимо знать:

- основные типы экономических задач, которые могут быть решены с использованием ППП;
- основные понятия и этапы моделирования экономических задач;
- пакеты прикладных программ, позволяющие выполнить решение и анализ без использования навыков программирования:
- стандартные офисные программные продукты;
- системы компьютерной алгебры;
- системы управления и оптимального планирования производства;
- CASE – технологии (Computer Aided System/Software Engineering);
- Специализированные статистические пакеты.

Изучение такого вида пакетов прикладных программ позволит:

- провести статистическую обработку данных для получения различного рода зависимостей и оценки их адекватности;
- осуществить графо – аналитическое исследование полученных моделей;
- отыскать оптимальное решение экономической задачи;
- оптимизировать структуру управления и планирования производства по различным критериям.

Работа современного специалиста экономического профиля не представляется возможной без использования информационных технологий. Поэтому изучение раздела, посвященного пакетам прикладных программ в предметной области, является чрезвычайно актуальным и необходимым.

Изложенный материал может быть полезен для изучения не только студентам экономического профиля, но и магистрантам, аспирантам и преподавателям.

ГЛАВА 1. ОСНОВНЫЕ ТИПЫ ЭКОНОМИЧЕСКИХ ЗАДАЧ И МЕТОДЫ ИХ РЕШЕНИЯ

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

Модель в общем смысле (обобщенная модель) есть создаваемый с целью получения и (или) хранения информации специфический объект (в форме мысленного образа, описания знаковыми средствами либо материальной системы), отражающий свойства, характеристики и связи объекта-оригинала произвольной природы, существенные для задачи, решаемой субъектом. Для принятия решений наиболее полезны модели, которые выражаются словами или формулами, алгоритмами и иными математическими средствами.

В процессе исследования современных сложных систем можно выделить различные классы моделей. В основе этих систем лежат модели различных типов: семантические, логические, математические и т.п.

1.1. Математические модели принятия решений

Анализ различного вида моделей показал, что при принятии решений в области экономики и управления чаще всего используются:

- модели технологических процессов (прежде всего модели контроля и управления);
- модели обеспечения качества продукции (в частности, модели оценки и контроля надежности);
- модели массового обслуживания;
- модели управления запасами (модели логистики);
- имитационные и эконометрические модели деятельности предприятия в целом и др.

Основные термины математического моделирования.

В качестве основных терминов, относящихся к разделу математического моделирования, можно предложить:

- компоненты системы - части системы, которые могут быть вычленены из нее и рассмотрены отдельно;
- независимые переменные – они могут изменяться, но это внешние величины, не зависящие от проходящих в системе процессов;
- зависимые переменные - значения этих переменных есть результат (функция) воздействия на систему независимых внешних переменных;
- управляемые (управляющие) переменные - те, значения которых могут изменяться исследователем;

- внутренние переменные – их значения определяются в ходе деятельности компонент системы (т.е. "внутри" системы);

- внешние переменные - определяются либо исследователем, либо извне, т.е. в любом случае действуют на систему извне.

При построении любой модели процесса управления желательно придерживаться следующего плана действий:

- сформулировать цели изучения системы;
- выбрать те факторы, компоненты и переменные, которые являются наиболее существенными для данной задачи;
- учесть тем или иным способом посторонние, не включенные в модель факторы;
- осуществить оценку результатов, проверку модели, оценку полноты модели.

При количественном моделировании бизнес-среды необходимо описывать взаимодействия многих переменных. Для этого нужно сформулировать математическую модель. В реальном мире обычно не существует единственно верного способа построения модели. Различные модели могут дать различные представления об одной и той же ситуации.

Этапы моделирования

Анализ различных литературных источников позволил выделить следующие этапы моделирования:

- постановка управленческой (экономической) проблемы, ее качественный анализ;
- построение математической модели;
- математический анализ модели;
- подготовка исходной информации;
- численное решение;
- анализ численных результатов и их применение.

Кратко охарактеризуем каждый из этапов.

Постановка экономической проблемы и ее качественный анализ

На этом этапе требуется сформулировать сущность проблемы, принимаемые предпосылки и допущения. Необходимо выделить важнейшие черты и свойства моделируемого объекта, изучить его структуру и взаимосвязь его элементов, хотя бы предварительно сформулировать гипотезы, объясняющие поведение и развитие объекта.

Построение математической модели

Это этап формализации экономической проблемы, т. е. выражения ее в виде конкретных математических зависимостей (функций, уравнений, неравенств и др.). Для некоторых сложных объектов целесообразно строить несколько разноаспектных моделей; при этом каждая модель выделяет лишь

некоторые стороны объекта, а другие стороны учитываются укрупненно и приближенно.

Математический анализ модели

На этом этапе чисто математическими приемами исследования выявляются общие свойства модели и ее решений. В частности, важным моментом является доказательство существования решения сформулированной задачи. При аналитическом исследовании выясняется, единственно ли решение, какие переменные могут входить в решение, в каких пределах они изменяются, каковы тенденции их изменения и т. д. Однако модели сложных экономических объектов с большим трудом поддаются аналитическому исследованию; в таких случаях используют численные методы исследования.

Подготовка исходной информации

В экономических задачах это, как правило, наиболее трудоемкий этап моделирования. Математическое моделирование предъявляет жесткие требования к системе информации; при этом надо принимать во внимание не только принципиальную возможность подготовки информации требуемого качества, но и затраты на подготовку информационных массивов. В процессе подготовки информации используются методы теории вероятностей, теоретической и математической статистики для организации выборочных обследований, оценки достоверности данных и т.д.

Численное решение

Этот этап включает разработку алгоритмов численного решения задачи, подготовку программ на ЭВМ, определение необходимых пакетов прикладных программ и непосредственное проведение расчетов. При этом значительные трудности вызываются большой размерностью экономических задач. Обычно расчеты на основе экономико-математической модели носят многовариантный характер. Многочисленные модельные эксперименты, изучение поведения модели при различных условиях возможно проводить благодаря высокому быстродействию современных ЭВМ. Численное решение существенно дополняет результаты аналитического исследования, а для многих моделей является единственно возможным.

Анализ численных результатов и их применение

На этом этапе прежде всего решается важнейший вопрос о правильности и полноте результатов моделирования и применимости их как в практической деятельности, так и в целях усовершенствования модели. Поэтому в первую очередь должна быть проведена проверка адекватности модели по тем свойствам, которые выбраны в качестве существенных.

Перечисленные этапы моделирования находятся в тесной взаимосвязи, в частности, могут иметь место возвратные связи этапов. Так, на этапе построения модели может выясниться, что постановка задачи или противоречива,

или приводит к слишком сложной математической модели; в этом случае исходная постановка задачи должна быть скорректирована. Наиболее часто необходимость возврата к предшествующим этапам моделирования возникает на этапе подготовки исходной информации. Если необходимая информация отсутствует или затраты на ее подготовку слишком велики, приходится возвращаться к этапам постановки задачи и ее формализации, чтобы приспособиться к доступной исследователю информации.

Недостатки, которые не удается исправить на тех или иных этапах моделирования, устраняются в последующих циклах. Однако результаты каждого цикла имеют и вполне самостоятельное значение. Начав исследование с построения простой модели, можно получить полезные результаты, а затем перейти к созданию более сложной и более совершенной модели, включающей в себя новые условия и более точные математические зависимости.

1.2. Методы оптимизации

В настоящее время специалист любого профиля может использовать при принятии решения различные компьютерные и математические средства. В памяти компьютеров может содержаться масса информации, организованная с помощью баз данных и других программных продуктов, позволяющих оперативно ею пользоваться. Экономико-математические и эконометрические модели позволяют просчитывать последствия тех или иных решений, прогнозировать развитие событий.

Сформулируем основные понятия, используемые в задачах оптимизации:

Управляемые переменные $x_1, x_2, \ldots, x_n$ – переменные, значения которых можно выбирать в определенных допустимых пределах.

ЛПР (лицо принимающее решение) – человек или группа людей, которые занимаются анализом и выбором значений управляемых переменных, обеспечивающих оптимальное решение.

Эффективное решение – набор значений управляемых переменных, который по некоторым соображениям ЛПР считает наиболее предпочтительными среди всех возможных решений.

Целевая функция задачи оптимизации – количественная мера оптимальности процесса.

Ограничения задачи оптимизации – совокупность условий (равенств, неравенств и т.п.), связывающих характеристики процесса и ограничивающих область изменения управляемых переменных.

Неуправляемые параметры – неизменяемые параметры процесса, значения которых известны.

Случайные факторы – факторы процесса, для которых ввиду их случайности неизвестны точные значения, но известен закон распределения вероятностей этих значений.

Неопределенные факторы – это факторы процесса, значения которых неизвестны.

Математическая модель оптимизации процесса – целевая функция и совокупность ограничений, зависящие от значений управляемых переменных, неуправляемых параметров, случайных и неопределенных факторов.

Допустимое решение – набор значений управляемых переменных, который удовлетворяет одновременно всем ограничениям задачи оптимизации.

Оптимальное решение - набор значений управляемых переменных, который не только удовлетворяет одновременно всем ограничениям задачи оптимизации, но и дает экстремальное значение целевой функции.

В зависимости от вида целевой функции, ограничений и присутствия случайных и неопределенных факторов оптимизационные модели можно в общем случае разделить на следующие классы:

- задачи математического программирования;
- задачи параметрического программирования;
- задачи стохастического программирования;
- оптимизационные задачи массового облуживания;
- задачи статистических игр.

Можно выделить несколько основных типов оптимизационных задач:

- задачи управления запасами;
- задачи распределения ресурсов;
- задачи ремонта и замены оборудования;
- сетевые оптимизационные задачи;
- задачи составления оптимальных расписаний;
- задачи оптимизации систем обслуживания;
- комбинированные задачи, объединяющие в себе черты задач разных типов.

Наиболее часто используются оптимизационные модели принятия решений. Их общий вид таков:

$F(X) \rightarrow \max (\min)$

$X \in A$.

Здесь X - параметр, который экономист может выбирать (управляющий параметр). Он может иметь различную природу - число, вектор, множество и т.п. Цель экономиста - максимизировать (минимизировать) целевую функцию $F(X)$, выбрав соответствующий X. При этом он должен учитывать ограничения $X \in A$ на возможные значения управляющего параметра X - он должен лежать в множестве A. Приведем основные виды оптимизационных задач экономики и управления.

1.2.1. Линейное программирование (ЛП)

Среди оптимизационных задач управления наиболее известны задачи линейного программирования, в которых максимизируемая (минимизируемая) функция $F(X)$ является линейной, а ограничения A задаются линейными неравенствами.

Линейное программирование как научно-практическая дисциплина.

Из всех задач оптимизации задачи линейного программирования выделяются тем, что в них ограничения - системы линейных неравенств или равенств. Ограничения задают выпуклые линейные многогранники в конечном линейном пространстве. Целевые функции также линейны. То есть:

- показатель оптимальности L(X) представляет собой ***линейную*** функцию от элементов решения $X = (x_1, x_2, ..., x_n)$;
- ограничительные условия, налагаемые на возможные решения, имеют вид ***линейных*** равенств или неравенств.

Общая форма записи модели задачи ЛП

Целевая функция (ЦФ)

$$L(X) = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max(\min),$$

при ограничениях

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le (\ge, =) b_1, \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le (\ge, =) b_2, \\ ... \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le (\ge, =) b_m, \\ x_1, x_2, ... x_k \ge 0 \, (k \le n) \end{cases}$$

Допустимое решение – это совокупность чисел (план) $X = (x_1, x_2, ..., x_n)$, удовлетворяющих ограничениям задачи.

Оптимальное решение – это план, при котором целевая функция (ЦФ) принимает свое максимальное (минимальное) значение.

Методы решения задач линейного программирования. Методы решения задач линейного программирования относятся к вычислительной математике, а не к экономике и менеджменту. Однако инженеру, менеджеру и экономисту необходимо знать о свойствах программного продукта, с которым он работает.

С ростом мощности компьютеров необходимость применения сложных математических методов снижается, поскольку во многих случаях время счета перестает быть лимитирующим фактором. Приведем пример некоторых из методов отыскания оптимума.

Простой перебор. Берется некоторый многомерный параллелепипед, в котором лежит многогранник, задаваемый ограничениями. Затем выполняется перебор точек параллелепипеда с заданным шагом, вычисляя значения целевой функции и проверяя выполнение ограничений. Из всех точек, удовлетворяющих ограничениям, возьмем ту, в которой целевая функция максимальна.

Направленный перебор. Берется точка, удовлетворяющая ограничениям. Затем последовательно или случайно меняем ее координаты на определенную величину, каждый раз в точке с более высоким значением целевой функции (если разыскивается максимум целевой функции). Вначале движение осуществляется по плоскости ограничений, затем по ребру ограничений и, наконец, отыскание вершины, где находится оптимум.

Симплекс-метод. Это один из первых специализированных методов оптимизации, нацеленный на решение задач линейного программирования, в то время как методы простого и направленного перебора могут быть применены для решения практически любой задачи оптимизации. Основная его идея состоит в продвижении по выпуклому многограннику ограничений от вершины к вершине, при котором на каждом шаге значение целевой функции улучшается до тех пор, пока не будет достигнут оптимум.

Сформулируем некоторые типы задач, сводящихся к задачам линейного программирования.

Транспортная задача.

Имеются склады, запасы на которых известны. Известны потребители и объемы их потребностей. Необходимо доставить товар со складов потребителям. Можно по-разному организовать "прикрепление" потребителей к складам, т.е. установить, с какого склада какому потребителю и сколько везти. Кроме того, известна стоимость доставки единицы товара с определенного склада определенному потребителю. Требуется минимизировать издержки по перевозке.

$$L(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}c_{ij}x_{ij}\rightarrow \min;$$

$$\begin{cases}\sum_{j=1}^{m}x_{ij}=a_i,\ i=\overline{1,n},\\ \sum_{i=1}^{n}x_{ij}=b_j,\ j=\overline{1,m},\\ x_{ij}\geq 0\ \left(i=\overline{1,n};\ j=\overline{1,m}\right).\end{cases}$$

Целевая функция (ЦФ) представляет собой общие транспортные расходы на осуществление всех перевозок в целом. Первая группа ограничений указывает, что запас продукции в любом пункте отправления должен быть равен суммарному объему перевозок продукции из этого пункта. Вторая группа ограничений указывает, что суммарные перевозки продукции в некоторый пункт потребления должны полностью удовлетворить спрос на продукцию в этом пункте. Наглядной формой представления модели транспортной задачи (ТЗ) является транспортная матрица.

Общий вид транспортной матрицы

Пункты отправления, A_i	Пункты потребления, B_j				Запасы, ед. продукции
	B_1	B_2	…	B_m	
A_1	c_{11}, [руб./ед. прод.]	c_{12}		c_{1m}	a_1
A_2	c_{21}	c_{22}		c_{2m}	a_2
…	…	…	…	…	…
A_n	c_{n1}	c_{n2}	…	c_{nm}	a_n
Потребность ед. продукции	b_1	b_2	…	b_m	$\sum_{i=1}^{n}a_i=\sum_{j=1}^{m}$

Сумма запасов продукции во всех пунктах отправления должна равняться суммарной потребности во всех пунктах потребления, т.е.

$$\sum_{i=1}^{n}a_i=\sum_{j=1}^{m}b_j.$$

Если условие выполняется, то ТЗ называется **сбалансированной** (закрытой), в противном случае – **несбалансированной** (открытой). В случае, когда *суммарные запасы превышают суммарные потребности*, необходим дополнительный **фиктивный** (реально не существующий) пункт потребления, который будет формально потреблять существующий излишек запасов, т.е.

$$b_{ф} = \sum_{i=1}^{n} a_i - \sum_{j=1}^{m} b_j .$$

Если *суммарные потребности превышают суммарные запасы*, то необходим дополнительный **фиктивный** пункт отправления, формально восполняющий существующий недостаток продукции в пунктах отправления:

$$a_{ф} = \sum_{j=1}^{m} b_j - \sum_{i=1}^{n} a_i .$$

Для фиктивных перевозок вводятся **фиктивные** тарифы $c^{ф}$, величина которых обычно приравнивается к нулю: $c^{ф} = 0$. Но в некоторых ситуациях величину фиктивного тарифа можно интерпретировать как **штраф**, которым облагается каждая единица недопоставленной продукции. В этом случае величина $c^{ф}$ может быть любым положительным числом.

Задача о назначениях – частный случай ТЗ. В задаче о назначениях количество пунктов отправления равно количеству пунктов назначения. Объемы потребности и предложения в каждом из пунктов назначения и отправления равны 1 (единице). Примером типичной задачи о назначениях является распределение работников по различным видам работ, минимизирующее суммарное время выполнения работ.

Переменные задачи о назначениях определяются следующим образом:

$$x_{ij} = \begin{cases} 1, & \text{если } i-\text{й рабочий работает на } j-\text{м станке,} \\ 0, & \text{в противном случае.} \end{cases}$$

Количество переменных и ограничений в транспортной задаче таково, что для ее решения не обойтись без компьютера и соответствующего программного продукта.

Общая распределительная задача ЛП

Общая распределительная задача ЛП – это распределительная задача (РЗ), в которой работы и ресурсы (исполнители) выражаются в различных единицах измерения. Типичным примером такой задачи является организация выпуска разнородной продукции на оборудовании различных типов.

Исходные параметры модели РЗ:

n – количество исполнителей;

m – количество видов выполняемых работ;

a_i – запас рабочего ресурса исполнителя A_i $(i = \overline{1,n})$ [ед.ресурса];

b_j – план по выполнению работы B_j $(j = \overline{1,m})$ [ед. работ];

c_{ij} – стоимость выполнения работы B_j исполнителем A_i [руб./ед. работ];

λ_{ij} – интенсивность выполнения работы B_j исполнителем A_i [ед. работ/ед. ресурса];

x_{ij} – планируемая загрузка исполнителя A_i при выполнении работ B_j [ед. ресурса];

$x_{ij}^{к}$ – количество работ B_j, которые должен будет произвести исполнитель A_i [ед. работ];

$L(X)$ – общие расходы на выполнение всего запланированного объема работ [руб.].

Этапы построения модели

- Определение переменных.
- Построение распределительной матрицы.
- Задание целевой функции (ЦФ).
- Задание ограничений.

Общий вид распределительной матрицы

Исполнители, A_i	Работы, B_j				Запас ресурса, ед.ресурса
	B_1	B_2		B_m	
A_1	λ_{11} c_{11}	λ_{12} c_{12}		λ_{1m} c_{1m}	a_1
A_2	λ_{21} c_{21}	λ_{22} c_{22}	...	λ_{2m} c_{2m}	a_2
...	...	...	...	...	...
A_n	λ_{n1} c_{n1}	λ_{n2} c_{n2}	...	λ_{nm} c_{nm}	a_n
План, ед.работы	b_1	b_2	...	b_m	

Модель РЗ

$$L(X)=\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}\left(\lambda_{ij}x_{ij}\right)\to \min;$$

$$\begin{cases} \sum_{j=1}^{m} x_{ij} = a_i,\ i=\overline{1,n}, \\ \sum_{i=1}^{n} \lambda_{ij}x_{ij} = b_j,\ j=\overline{1,m}, \\ x_{ij}\geq 0\ \left(i=\overline{1,n};\ j=\overline{1,m}\right), \end{cases}$$

где $(\lambda_{ij}x_{ij})$ – это количество работ j-го вида, выполненных i-м исполнителем.

Целочисленное программирование

Задачи оптимизации, в которых переменные принимают целочисленные значения, относятся к целочисленному программированию. Обозначим некоторые из таких задач.

Задача о выборе оборудования.

Задача отличается от задачи линейного программирования только условием целочисленности, поскольку численность оборудования не может выражаться дробным числом.

Задача о ранце

Общий вес ранца заранее ограничен. Какие предметы положить в ранец, чтобы общая полезность отобранных предметов была максимальна? Вес каждого предмета известен.

С точки зрения экономики предприятия и организации производства более актуальна интерпретация задачи о ранце, в которой в качестве "предметов" рассматриваются заказы (или варианты выпуска партий тех или иных товаров), в качестве полезности – прибыль от выполнения того или иного заказа, а в качестве веса – себестоимость заказа.

В отличие от предыдущих задач, управляющие параметры принимают значения из множества, содержащего два элемента - 0 и 1(то есть заказ принят или нет).

К целочисленному программированию относятся задачи размещения (производственных объектов), теории расписаний, календарного и оперативного планирования, назначения персонала и т.д.

В качестве наиболее распространенных методов решения задач целочисленного программирования можно назвать: метод приближения непрерывными задачами и метод направленного перебора.

Модели сетевого планирования и управления

Сетевой моделью (другие названия: сетевой график, сеть) называется экономико-математическая модель, отражающая комплекс работ (операций) и событий, связанных с реализацией некоторого проекта (научно-исследовательского, производственного и др.) в их логической и технологической последовательности и связи. Анализ сетевой модели, пред-

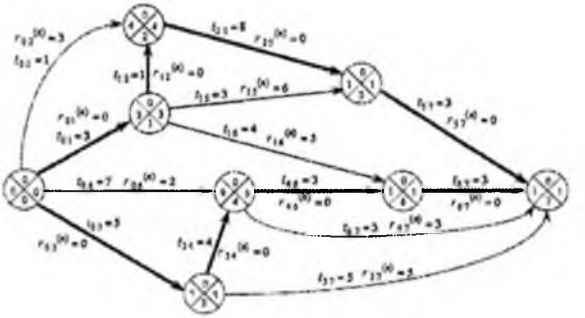
ставленной в графической или табличной (матричной) форме, позволяет, во-первых, более четко выявить взаимосвязи этапов реализации проекта и, во-вторых, определить наиболее оптимальный порядок выполнения этих этапов в целях, например, сокращения сроков выполнения всего комплекса работ. Таким образом, методы сетевого моделирования относятся к методам принятия оптимальных решений.

Математический аппарат сетевых моделей базируется на теории графов. Графом называется совокупность двух конечных множеств: множества точек, которые называются вершинами, и множества пар вершин, которые называются ребрами. Если рассматриваемые пары вершин являются упорядоченными, т. е. на каждом ребре задается направление, то граф называется ориентированным; в противном случае — неориентированным. Последовательность неповторяющихся ребер, ведущая от некоторой вершины к другой, образует путь. Граф называется связным, если для любых двух его вершин существует путь, их соединяющий; в противном случае граф называется несвязным. В экономике чаще всего используются два вида графов: дерево и сеть. Дерево представляет собой связный граф без циклов, имеющий исходную вершину (корень) и крайние вершины; пути от исходной вершины к крайним вершинам называются ветвями. Сеть — это ориентированный конечный связный граф, имеющий начальную вершину (источник) и конечную вершину (сток). Таким образом, сетевая модель представляет собой граф вида «сеть».

В экономических исследованиях сетевые модели возникают при моделировании экономических процессов методами сетевого планирования и управления (СПУ).

Объектом управления в системах сетевого планирования и управления являются коллективы исполнителей, располагающих определенными ресурсами и выполняющих определенный комплекс операций, который призван обеспечить достижение намеченной цели, например, разработку нового изделия, строительства объекта и т.п.

Основой СПУ является сетевая модель (СМ), в которой моделируется совокупность взаимосвязанных работ и событий, отображающих процесс достижения определенной цели. Она может быть представлена в виде графика или таблицы. Пример сетевого графика приведен ниже.



Ориентированный граф был бы полезен, например, для иллюстрации организации перевозок в транспортной задаче. В экономике дугам ориентированного или обычного графа часто приписывают числа, например, стоимость проезда или перевозки груза из пункта А (начальная вершина дуги) в пункт Б (конечная вершина дуги).

Некоторые, наиболее типичные задачи принятия решений, связанных с оптимизацией на графах.

Задача коммивояжера

Требуется посетить все вершины графа и вернуться в исходную вершину, минимизировав затраты на проезд (или минимизировав время).

Исходные данные - это граф, дугам которого приписаны положительные числа - затраты на проезд или время, необходимое для продвижения из одной вершины в другую. В общем случае граф является ориентированным, и каждые две вершины соединяют две дуги - туда и обратно. Действительно, если пункт А расположен на горе, а пункт Б - в низине, то время на проезд из А в Б, очевидно, меньше времени на обратный проезд из Б в А.

Многие постановки экономического содержания сводятся к задаче коммивояжера. Например:

- составить наиболее выгодный маршрут обхода наладчика в цехе (контролера, охранника, милиционера), отвечающего за должное функционирование заданного множества объектов (каждый из этих объектов моделируется вершиной графа);
- составить наиболее выгодный маршрут доставки деталей рабочим или хлеба с хлебозавода по заданному числу булочных и других торговых точек (парковка у хлебозавода).

Задача о кратчайшем пути

Как кратчайшим путем попасть из одной вершины графа в другую (и, следовательно, с наименьшим расходом топлива и времени, наиболее дешево попасть из пункта А в пункт Б)? Для решения этой задачи каждой дуге ориентированного графа должно быть сопоставлено число - время движения по этой дуге от начальной вершины до конечной.

Оптимизационные задачи на графах, возникающие при подготовке управленческих решений, весьма многообразны.

Задача о максимальном потоке

Как (т.е. по каким маршрутам) послать максимально возможное количество грузов из начального пункта в конечный пункт, если пропускная способность путей между пунктами ограничена?

Для решения этой задачи каждой дуге ориентированного графа, соответствующего транспортной системе, должно быть сопоставлено число - пропускная способность этой дуги.

В различных проблемах принятия решений возникают самые разнообразные задачи оптимизации. Для их решения применяются те или иные методы, точные или приближенные. Задачи оптимизации часто используются в теоретико-экономических исследованиях. Например, задачи определения оптимального объема выпуска по функции издержек при фиксированной цене или минимизации издержек при заданном объеме выпуска путем выбора оптимального соотношения факторов производства.

Конкретные виды задач оптимизации и методы их решения рассматриваются в соответствующей литературе.

1.2.2. Регрессионный и корреляционный анализ

Регрессионный и корреляционный анализ позволяет установить и оценить зависимость изучаемой случайной величины Y от одной или нескольких других величин X и делать прогнозы значений Y. Параметр Y, значение которого нужно предсказывать, является **зависимой** переменной. Параметр X, значения которого нам известны заранее и который влияет на значения Y, называется **независимой** переменной. Например, X – величина затрат компании на рекламу своего товара, Y – объем продаж этого товара и т.д.

Корреляционная зависимость Y от X – это функциональная зависимость вида

$$\bar{y}_x = f(x),$$

где $\bar{y}_x$ – среднее арифметическое (**условное среднее**) всех возможных значений параметра Y, которые соответствуют значению $X = x$. Уравнение называется **уравнением регрессии** Y на X, функция $f(x)$ – **регрессией** Y на X, а ее график – **линией регрессии** Y на X.

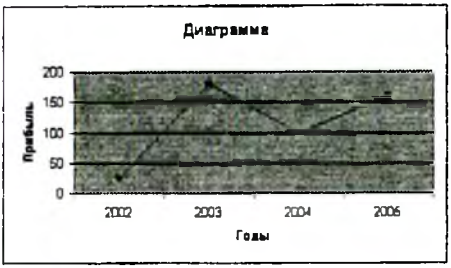
Основная задача регрессионного анализа – установление ***формы*** корреляционной связи, т.е. вида функции регрессии (линейная, квадратичная, показательная и т.д.).

Метод наименьших квадратов

Метод наименьших квадратов позволяет определить коэффициенты уравнения регрессии таким образом, чтобы точки, построенные по исходным данным (x_i, y_i), лежали как можно ближе к точкам линии регрессии. Формально это записывается как минимизация суммы квадратов отклонений (ошибок) функции регрессии и исходных точек:

$$S = \sum_{i=1}^{n} \left(y_i^p - y_i \right)^2 \rightarrow \min,$$

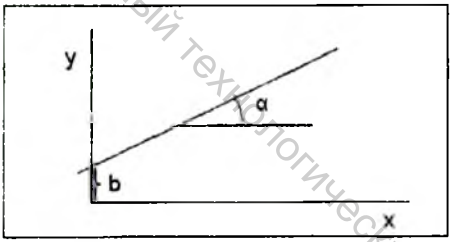
где y_i^p – значение, вычисленное по уравнению регрессии; $\left(y_i^p - y_i\right)$ **отклонение** ε (ошибка, остаток); n – количество пар исходных данных.



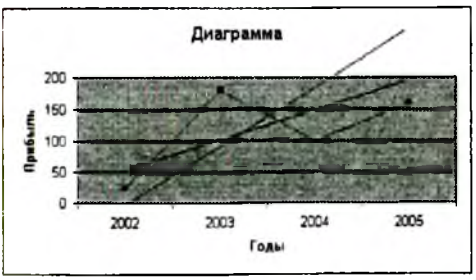
Простейший вариант модели - прямая линия на плоскости.

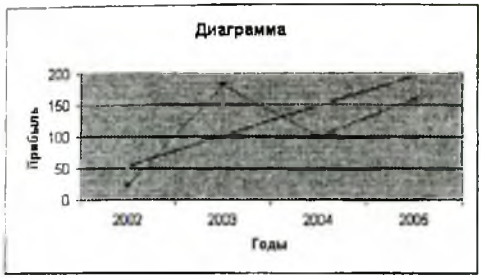
$$y = a * x + b$$

где b - значение y при x=0;
a=tg(α) - тангенс угла наклона прямой по отношению к оси x.



Возможные варианты модели представлены ниже.





Первый шаг

$$\Delta_1 = t_1 - p_1;$$
$$\Delta_2 = t_2 - p_2;$$
$$\ldots\ldots\ldots\ldots\ldots;$$
$$\Delta_n = t_n - p_n;$$

Второй шаг

$$\sum_{i=1}^{n} \Delta = \sum_{i=1}^{n} (t_i - p_i)$$

Третий шаг

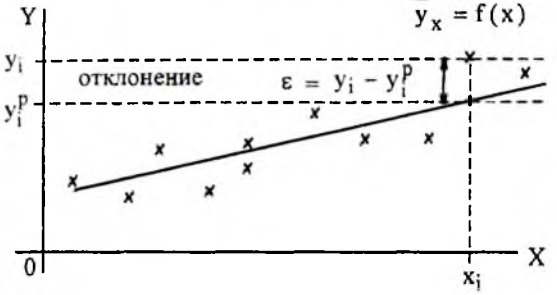
$$\sum_{i=1}^{n} \Delta^2 = \sum_{i=1}^{n} (t_i - p_i)^2$$

Четвертый шаг

$$\Delta S^2 = \frac{\sum_{i=1}^{n} \Delta^2}{n} = \frac{\sum_{i=1}^{n} (t_i - p_i)^2}{n}$$

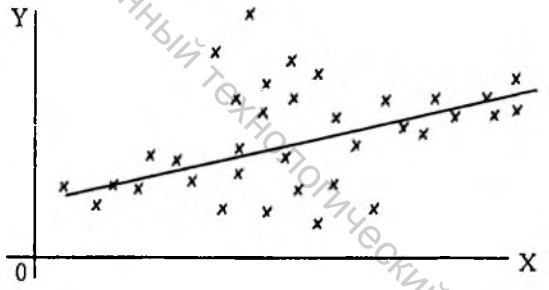
Пятый шаг

$$\Delta S^2 = \frac{\sum_{i=1}^{n} \Delta^2}{n} = \frac{\sum_{i=1}^{n} (t_i - p_i)^2}{n} = \min(0)$$



Понятие отклонения ε для случая линейной регрессии

В регрессионном анализе предполагается, что математическое ожидание случайной величины ε равно нулю и ее дисперсия одинакова для всех наблюдаемых значений Y. Отсюда следует, что рассеяние данных возле линии регрессии должно быть одинаково при всех значениях параметра X. В случае, показанном на рисунке, приведенном ниже, данные распределяются вдоль линии регрессии неравномерно, поэтому метод наименьших квадратов в этом случае неприменим.



Неравномерное распределение исходных точек вдоль линии регрессии

Основная задача корреляционного анализа – оценка ***тесноты*** (силы) корреляционной связи. Теснота корреляционной зависимости Y от X оценивается по величине рассеяния значений параметра Y вокруг условного среднего $\bar{y}_x$. Большое рассеяние говорит о слабой зависимости Y от X либо об ее отсутствии, и, наоборот, малое рассеяние указывает на наличие достаточно сильной зависимости.

Коэффициент детерминации (по-другому – детерминированности) r^2 показывает, на сколько процентов ($r^2 \cdot 100\%$) найденная функция регрессии описывает связь между исходными значениями параметров X и Y:

$$r^2 = \frac{\sum_{i=1}^{n}\left(y_i^p - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2},$$

где $\left(y_i^p - \bar{y}\right)^2$ – объясненная вариация; $\left(y_i - \bar{y}\right)^2$ – общая вариация.

Графическая интерпретация коэффициента детерминации для случая линейной регрессии

Соответственно, величина $(1-r^2)\cdot 100\%$ показывает, сколько процентов вариации параметра Y обусловлены факторами, не включенными в регрессионную модель. При высоком ($r^2 \geq 75\%$) значении коэффициента детерминации можно делать прогноз $y^* = f(x^*)$ для конкретного значения x^*.

Линейная регрессия

Коэффициенты **линейной** регрессии $y = a_0 + a_1 x$ вычисляются по следующим формулам (все суммы берутся по n парам исходных данных):

$$a_1 = \frac{n\left(\sum y_i x_i\right) - \sum y_i \sum x_i}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2};$$

$$a_0 = \frac{1}{n}\left(\sum y_i - a_1 \sum x_i\right).$$

Нелинейная регрессия

Рассмотрим наиболее простые случаи ***нелинейной*** регрессии: гиперболу, экспоненту и параболу. При нахождении коэффициентов гиперболы и экспоненты используют прием приведения нелинейной регрессионной

зависимости к линейному виду. Это позволяет использовать для вычисления коэффициентов функций регрессии формулы линейной зависимости.

Гипербола

При нахождении **гиперболы** $y = a_0 + \frac{a_1}{x}$ вводят новую переменную $z = \frac{1}{x}$, тогда уравнение гиперболы принимает линейный вид $y = a_0 + a_1 z$. После этого используют формулы для нахождений линейной функции, но вместо значений x_i используются значения $z_i = \frac{1}{x_i}$:

$$a_1 = \frac{n\left(\sum y_i z_i\right) - \sum y_i \sum z_i}{n\left(\sum z_i^2\right) - \left(\sum z_i\right)^2}; \qquad a_0 = \frac{1}{n}\left(\sum y_i - a_1 \sum z_i\right).$$

Экспонента

Для приведения к линейному виду **экспоненты** $y = a_0 e^{a_1 x}$ проводят логарифмирование:

$\ln y = \ln\left(a_0 e^{a_1 x}\right)$;

$\ln y = \ln a_0 + \ln\left(e^{a_1 x}\right)$;

$\ln y = \ln a_0 + a_1 x$.

Введя переменные $b_0 = \ln a_0$ и $b_1 = a_1$, получаем $\ln y = b_0 + b_1 x$, откуда следует, что можно применять формулы линейной зависимости, в которых вместо значений y_i надо использовать $\ln y_i$:

$$b_1 = \frac{n\left(\sum \left[\ln y_i\right] x_i\right) - \sum \ln y_i \sum x_i}{n\left(\sum x_i^2\right) - \left(\sum x_i\right)^2}; \qquad b_0 = \frac{1}{n}\left(\sum \ln y_i - b_1 \sum x_i\right).$$

При этом получаем численные значения коэффициентов b_0 и b_1, от которых надо перейти к a_0 и a_1, используемых в модели экспоненты. Исходя из введенных обозначений и определения логарифма, получаем

$a_0 = e^{b_0}$, $a_1 = b_1$.

Парабола

Для нахождения коэффициентов **параболы** $y = a_0 + a_1 x + a_2 x^2$ необходимо решить линейную систему из трех уравнений:

$$\begin{cases} n \cdot a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i, \\ \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum \left(y_i x_i\right), \\ \left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum \left(y_i x_i^2\right). \end{cases}$$

При вычислении коэффициента детерминации экспоненты все значения параметра Y (исходные, регрессионные, среднее) необходимо заменить на их логарифмы, например, y_i^p – на $\ln\left(y_i^p\right)$ и т.д.

ГЛАВА 2. ОБЗОР СОВРЕМЕННЫХ ПАКЕТОВ ПРИКЛАДНЫХ ПРОГРАММ, ПОЗВОЛЯЮЩИХ МОДЕЛИРОВАТЬ И РЕШАТЬ ЗАДАЧИ В ОБЛАСТИ ЭКОНОМИКИ

Информационная поддержка процессов моделирования, анализа и управления может осуществляться с использованием самых разнообразных программных средств. Назовем и охарактеризуем некоторые из них:

- Универсальные и специализированные языки программирования.
- Стандартные офисные программные продукты.
- Системы компьютерной математики.
- Системы управления проектами.
- CASE-технологии.
- Специализированные статистические пакеты.

2.1. Универсальные и специализированные языки программирования

Существующие языки программирования, безусловно, позволяют осуществить построение модели любого вида и любой сложности. Однако для этого от экономиста требуются профессиональные знания и навыки программирования. В случае разработки собственного программного средства, безусловно, целесообразнее возложить исполнение этой задачи на профессионального программиста.

2.2. Стандартные офисные программные продукты

К наиболее известным программным продуктам, позволяющим моделировать процессы управления можно отнести:

- Microsoft Office.
- Star Office.
- Lotus.
- Open Office и т.д.

Наиболее известной компонентой является Microsoft Excel, в состав которой входят функции для построения различного вида моделей. Помимо этого, имеется возможность отыскания оптимального решения при заданных ограничениях.

2.3. Системы компьютерной математики

Достоинства систем компьютерной математики в разрезе построения моделей заключается в том, что возможно не численное, а символьное

представление уравнения. Анализ символьной модели значительно нагляднее, точнее и привычнее.

Данное направление моделирования очень перспективно и все более широко используется во всем мире.

Назовем некоторые из систем компьютерной математики (систем символьной алгебры):

- Gauss 6.0;
- Maple 10.03;
- Mathematica 5.2.0;
- MathLab 6.5;
- MuPad 4.0;
- SciLab 4.0;
- Maxima 5.9.1 и т.д.

Каждая из систем имеет свои достоинства и недостатки. Рассмотрение работы каждого из этих продуктов не представляется возможным. Ограничимся демонстрацией возможностей системы Maple (см. главу 5).

2.4. Системы управления проектами

Первые программные средства для управления проектами были разработаны достаточно давно: почти сорок лет назад. В основе данных систем лежали алгоритмы сетевого планирования и расчета временных параметров проекта по методу критического пути. Первые системы позволяли представить проект в виде сети, рассчитать ранние и поздние даты начала и окончания работ проекта и отобразить работы на временной оси в виде диаграммы Ганта. Позже в системы были добавлены возможности ресурсного и стоимостного планирования, средства контроля за ходом выполнения работ.

Использование систем долгое время ограничивалось традиционными областями - крупными строительными, инженерными или оборонными проектами – и требовало профессиональных знаний. Однако за последнее десятилетие ситуация в области использования ПО календарного планирования резко изменилась.

Благодаря повышению мощности и снижению стоимости персональных компьютеров, а также при участии таких корпораций, как Microsoft и Symantec, буквально заваливших рынок дешевыми системами для управления проектами, программное обеспечение и методики управления, доступные раньше только состоятельным организациям, пришли на рабочие столы и вошли в повседневную практику менеджеров и сотрудников средних и малых компаний.

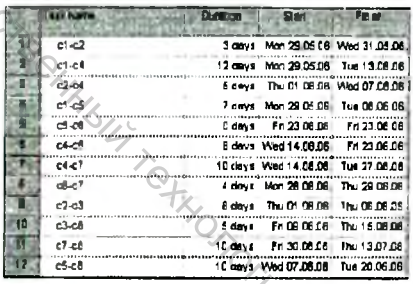
В настоящее время на рынке представлено значительное количество универсальных программных пакетов для персональных компьютеров, автоматизирующих функции планирования и контроля календарного графика выполнения работ. Среди наиболее популярных можно привести следующие:

- Primavera Project Planner (P3) (Primavera);

- Microsoft Project (Microsoft);
- Time Line (Time Line Solutions);
- Open Plan (Welcome Software);
- Artemis Views (Artemis Management Systems);
- CA-Super Project (Computer Associates International Inc.);
- Project Scheduler (Scitor Corp.);
- TurboProject (IMSI);
- Project Workbench (Applied Business Technology);
- Spider Project (Технологии управления Спайдер);
- PlanBee;
- Rillsoft Project и т.д.

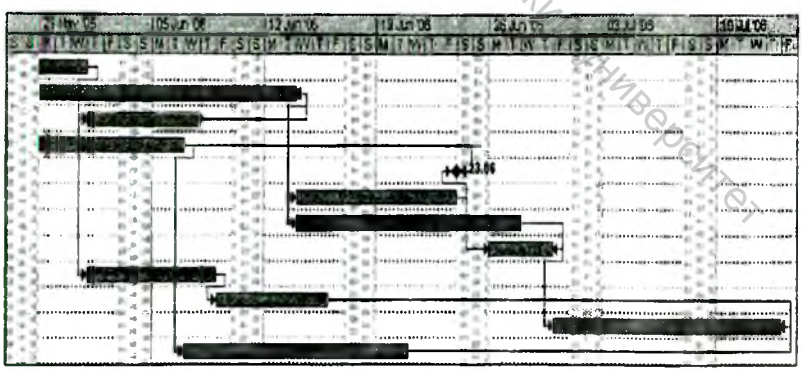
Покажем, как выглядит пример использования программы Microsoft Office Project 2003 для управления проектами.

Исходные данные

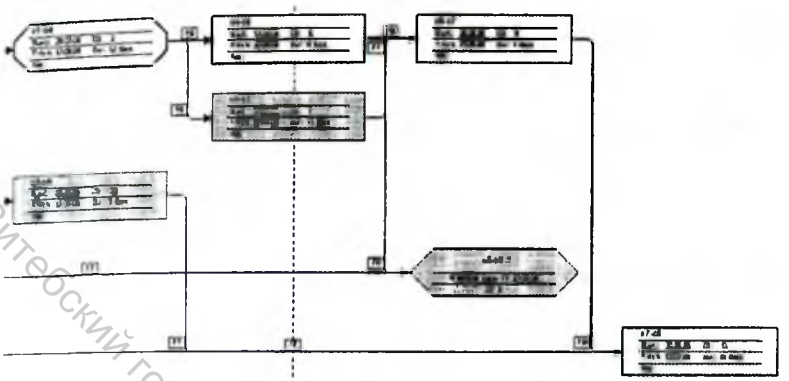


	Task Name	Duration	Start	Finish
1	c1-c2	3 days	Mon 29.05.06	Wed 31.05.06
2	c1-c4	12 days	Mon 29.05.06	Tue 13.06.06
3	c2-c4	5 days	Thu 01.06.06	Wed 07.06.06
4	c1-c5	7 days	Mon 29.05.06	Tue 06.06.06
5	c5-c8	5 days	Fri 23.06.06	Fri 23.06.06
6	c4-c8	8 days	Wed 14.06.06	Fri 23.06.06
7	c4-c7	10 days	Wed 14.06.06	Tue 27.06.06
8	c6-c7	4 days	Mon 26.06.06	Thu 29.06.06
9	c2-c3	8 days	Thu 01.06.06	Thu 08.06.06
10	c3-c6	5 days	Fri 09.06.06	Thu 15.06.06
11	c7-c8	1 days	Fri 30.06.06	Thu 13.07.06
12	c5-c8	10 days	Wed 07.06.06	Tue 20.06.06

Диаграмма Ганта



Сетевая диаграмма



2.5. CASE-технологии (Computer Aided System/Software Engineering)

К настоящему моменту наиболее интенсивное развитие получили два главных направления применения CASE-средств:

1. реорганизация (перепроектирование) бизнес-процессов организации;
2. системный анализ и проектирование, включающий функциональное, информационное и событийное моделирование как вновь создаваемой, так и существующей системы.

Необходимо отметить, что такое разбиение является весьма условным, поскольку при анализе организации и разработке проекта ее автоматизации используются элементы перепроектирования, в то же время необходимым этапом перепроектирования является, по крайней мере, создание функциональной модели бизнес-процесса.

В таблице приведен перечень некоторых известных CASE-средств и поддерживаемые ими виды проектной деятельности.

Название	Фирма	Реорганизация	Функции	Данные	События
BPWin	Logic Works	+	+	-	-
CASE-Аналитик	Эйтэкс	-	+	+	+
CASE/4/0	MicroTOOL	-	+	+	+
Database Designer	Oracle	-	-	+	-
Design/IDEF	Meta Software	+	+	+	-
Designer/2000	Oracle	+	+	+	-
EasyCASE	Evergreen	-	+	+	+

	CASE Tools				
ERWin	Logic Works	-	-	+	-
I-CASE Yourdon	CAYENNE	-	+	+	+
Prokit* WORKBENCH	MDIS	-	+	+	-
S-Designor	Sybase/Powersoft	-	+	+	-
SILVERRUN	CSA	-	+	+	+
Visible Analyst Workbench	Visible Systems	-	+	+	-

Средства реорганизации бизнес-процессов

Для моделирования бизнес-процессов обычно используется методология SADT - Structured Analysis and Design Technique (точнее, ее подмножество IDEFO), поддерживаемая пакетами BPWin и Design/IDEF. Однако статическая SADT-модель не обеспечивает полного решения задач перепроектирования, необходимо иметь возможность исследования динамических характеристик бизнес-процессов. Одним из решений является использование системы динамического моделирования Design/CPN. Фактически Design/IDEF и Design/CPN являются компонентами интегрированной методологии перепроектирования: статические SADT-диаграммы автоматически преобразуются в прообраз динамической модели, которая дорабатывается вручную и затем исполняется в различных режимах с целью получения соответствующих оценок.

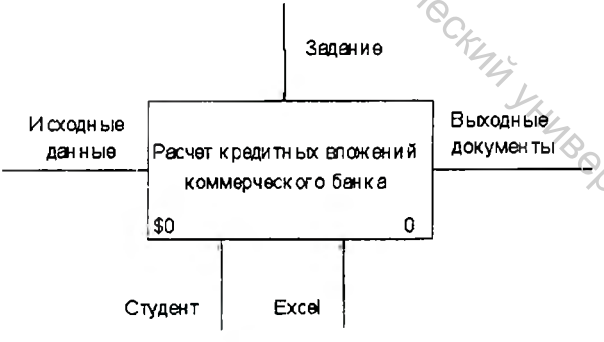
Другой возможный подход реализуется пакетом Designer/2000: моделирование бизнес-процессов является первым этапом разработки системы, а соответствующая модель является основой для разработки концептуальных моделей и проектирования системы. Нотация для моделирования бизнес-процессов включает следующие элементы: базовый процесс, шаг процесса, хранилище, поток, событие и организационная единица. Для каждого элемента можно задать разнообразные количественные параметры (временные затраты, ресурсы и т.п.), а затем с помощью специальной процедуры анимации проследить поведение модели в динамике с учетом введенных параметров. Следует отметить, что не существует принципиальных ограничений в использовании в качестве средства построения статических моделей бизнес-процессов и традиционных DFD - диаграмм потоков данных. Более того, в настоящий момент за рубежом доступен ряд продуктов динамического моделирования (INCOME Mobile, CPN-AMI и др.), интегрируемых с DFD-моделью, которые позволяют успешно решать задачи перепроектирования.

Средства функционального моделирования

Для решения задачи функционального моделирования на базе структурного анализа традиционно применяются два типа моделей: SADT-диаграммы и диаграммы потоков данных. В случае наличия в моделируемой системе программной/программируемой части (т.е. практически всегда) предпочтение, как правило, отдается DFD по следующим соображениям: DFD с самого начала создавались как средство проектирования программных систем (тогда как SADT -как средство проектирования систем вообще) и имеют более богатый набор элементов, адекватно отражающих их специфику (например, хранилища данных являются прообразами файлов или баз данных).

Охарактеризуем наиболее известные программные проекты фирмы Logic Works (BPWin, ERWin).

Пакет BPWin основан на методологии IDEF0 и предназначен для функционального моделирования и анализа деятельности предприятия. Модель в BPWin представляет собой совокупность SADT-диаграмм, каждая из которых описывает отдельный процесс в виде разбиения его на шаги и подпроцессы. С помощью соединяющих дуг описываются объекты, данные и ресурсы, необходимые для выполнения функций. Имеется возможность для любого процесса указать стоимость, время и частоту его выполнения. Эти характеристики в дальнейшем могут быть просуммированы с целью вычисления общей стоимости затрат - таким образом выявляются узкие места технологических цепочек, определяются затратные центры. BPWin может импортировать фрагменты информационной модели из описываемого ниже средства проектирования баз данных ERWin (при этом сущности и атрибуты информационной модели ставятся в соответствие дугам SADT-диаграммы). Генерация отчетов по модели может осуществляться в формате MS Word и MS Excel.



Семейство продуктов ERWin предназначено для моделирования и создания баз данных произвольной сложности на основе диаграмм "сущность-связь". В настоящее время ERWin является наиболее популярным пакетом

моделирования данных благодаря поддержке широкого спектра СУБД самых различных классов: SQL-серверов (Oracle, Informix, Sybase SQL Server, MS SQL Server, Progress, DB2, SQLBase, Ingress, Rdb и др.) и "настольных" СУБД типа XBase (Clipper, dBASE, FoxPro, MS Access, Paradox и др.). Информационная модель представляется в виде диаграмм "сущность-связь", отражающих основные объекты предметной области и связи между ними. Дополнительно определяются атрибуты сущностей, характеристики связей, индексы и бизнес-правила, описывающие ограничения и закономерности предметной области. После создания ER-диаграммы пакет автоматически генерирует SQL-код для создания таблиц, индексов и других объектов базы данных. По заданным бизнес-правилам формируются стандартные триггеры БД для поддержки целостности данных, для сложных бизнес-правил можно создавать собственные триггеры, используя библиотеку шаблонов.

Пакет может осуществлять реинжиниринг существующих БД: по SQL-текстам автоматически генерируются ER-диаграммы. Таким образом, пакет полностью поддерживает технологию FRE (forward and reverse engineering), последовательность этапов которой приведена ниже:

- импорт с сервера существующей БД;
- автоматическая генерация модели БД;
- модификация модели;
- автоматическая генерация новой схемы и построение физической БД на том же самом или любом другом сервере.

Для разработки клиентской части приложения имеются специальные версии пакета, обеспечивающие интеграцию с такими инструментами, как SQLWindows, PowerBuilder, Visual Basic, Delphi.

Для коллективной разработки модели БД предназначен специальный продукт ModelMart, позволяющий контролировать версии модели, гибко распределять права доступа между членами группы, строить библиотеки моделей, осуществлять объединение моделей и т.п. Продукт построен в архитектуре "клиент-сервер", репозитарий использует одну из трех СУБД - Oracle, Sybase, MS SQL Server.

2.6. Специализированные статистические пакеты

Одной из важнейших задач при работе с экономической информацией является статистический анализ данных. Сейчас на рынке имеется большое количество компьютерных программ, которые позволяют проводить такой анализ. Обилие систем, создатели которых утверждают, что их программа является наилучшей для обработки данных, а также отсутствие у большинства врачей достаточного времени для освоения нескольких пакетов, приводит к усложнению процесса выбора. Попытаемся сравнить несколько систем статистического анализа, акцентируя внимание на их полноте и простоте для начинающих.

Для сравнения были выбраны универсальные пакеты статистических программ, работающие под управлением ОС Windows, такие, как:

1. SAS for Windows (SAS Institute Inc.);
2. SPSS (SPSS Inc.);
3. S-Plus (Mathworks) ;
4. Systat (SPSS Inc.) ;
5. NCSS (NCSS);
6. STATA (Stata corp.);
7. Statistica (Statsoft Inc.);
8. Statgraphics Plus (Manguistics, Inc).

Из перечисленных выше статистических систем лишь две (Statistica и Statgraphics) изначально создавались для IBM-совместимых ПЭВМ, причем первая из них была разработана уже с расчетом на графическую среду Windows. Stata, Systat и S-plus - относительно "молодые" многоплатформенные системы (DOS, UNIX и/или Macintosh), поддерживающие широкий набор графических команд, однако из-за совместимости с текстовой средой UNIX в значительной степени полагающиеся на командную строку при управлении системой. SAS и SPSS были разработаны в эпоху больших ЭВМ и поэтому их ядро продолжает носить отпечаток тех лет.

Краткая характеристика систем приведена ниже.

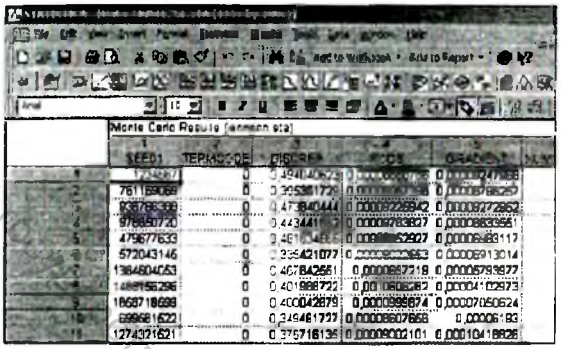
Программы	Доступ к файлам других форматов / число поддерживаемых форматов (различные форматы одной программы считаются за один)		Возможности расширения системы / наличие готовых дополнительных подпрограмм	Графика / экспорт результатов/ графики
SPSS	SPSS*, SYLK, Excel, dBASE, Lotus, Systat, ASCII (с разделением символом табуляции), любых текстовых форматов с описанием	7	Макросы, развитый командный язык/ ++	+++/RTF, текст/cgm, pict, eps, tiff, wmf, bmp
SAS	SAS*, dBASE, Excel, Lotus, ASCII (с разделением символом табуляции, запятыми (csv) или пробелами),	5	Макросы, развитый командный язык/ +++	++ /RTF, текст/bmp. emf, jpeg, gif, tiff, ps, eps, pbm, dib
Statistica	Lotus, Quattro, Excel, Paradox, ASCII (с разделением символом табуляции, запятыми (csv), точкой с запятой	5	Командный язык, макросы/ -	++++/ текст/wmf, bmp

	или пробелами)			
NCSS97	Access, ASCII, BMDP, Clone, dBASE, Excel, Gauss, Lotus, NCSS*, Paradox, Quattro, SAS, SigmaPlot, SOLO, S-Plus, SPSS, STATA, Symphony, Systat	19	-	+++/rtf/wmf
S-Plus	ASCII, Excel, Lotus, dBASE, Paradox, Quattro, SigmaPlot, Systat, SAS, SPSS, STATA, MS Access, Gauss, MatLab, S-Plus*	15	Развитый командный язык, скомпилированные дополнительные модули/ +++	++++/rtf, текст/bmp, eps, wmf, gif, GEM, jpeg, HPGL, pcx, tiff, targa
STATA/ StatTransfer	Lotus, Access, dBASE, ASCII, Excel, Gauss, MatLab, Paradox, Quattro, SAS, S-Plus, SPSS, STATA, Systat	14	Развитый командный язык / +++	++/текст/-
SYSTAT	Excel, DIF, Dbase, Lotus, BMDP, Symphony, ASCII, SPSS	9	Развитый командный язык / +	+++/ текст, rtf (версия 8)/eps, jpeg, cgm. bmp, pict, wmf
MINITAB	Minitab*, Excel, Quattro, Lotus, Symphony, dBase, ASCII	7	Командный язык, кросы/ +	+++/ текст/jpeg, tiff, bmp,
STAT GRAPHICS+	Statgraphics*, Execustat, dBASE, DIF, Lotus, Excel	6	-	+++/текст/wmf

Statistica. Данная система задумывалась как полная статистическая система для пользователей персональных компьютеров, не привыкших к работающим в пакетном режиме ранних версий SAS или SPSS.

С самого начала эта программа обладала развитым графическим интерфейсом и опиралась на поддержку высококачественной графики для анализа данных. Система состоит из ряда модулей, работающих независимо. Каждый модуль включает определенный класс процедур (например, для кластерного анализа или анализа выживаемости). Графики в данной системе строятся как из общего меню, так и из подменю процедур, что очень облегчает начинающим выбор адекватного графического представления данных. Почти все процедуры являются интерактивными, т.е. для запуска обработки необходимо выбрать из меню переменные и ответить на ряд вопросов системы.

Это очень удобно для начинающего пользователя, однако резко замедляет деятельность опытного и не позволяет эффективно повторять одну и ту же процедуру несколько раз. Автоматизация возможна при помощи командного языка (Statistica Command Language), интерпретатор которого доступен из любого модуля. Мастер команд (Command Wizard) резко облегчает составление программы, сводя программирование к выбору соответствующих пунктов меню.

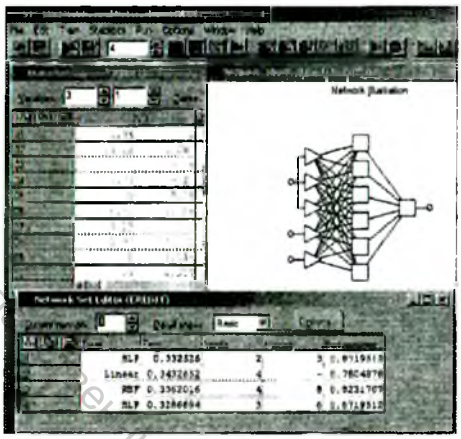


Однако на самом деле интерпретатор работает с языком макропоследовательностей, т.е. он просто автоматизирует действия пользователя, а не обращается напрямую к ядру системы, поэтому после запуска на выполнение вызываются отдельные модули и на экране начинают мелькать отдельные окна и подменю. Кроме того, необходимо (в отличие от других систем) включить сохранение результатов, иначе они пропадут. Встроенный язык программирования (Statistica Basic), несмотря на название, по своей структуре похож на Паскаль и не получил значительного распространения, поэтому дополнительные подпрограммы, созданные третьими сторонами, практически отсутствуют. Жесткая структура также не позволяет использование дополнительных модулей.

Statistica является относительно достаточно небольшой программой и обладает одной из наилучших систем подсказки. Возможности экспорта и импорта данных развиты достаточно, но без особых дополнений (табл.2). Работать с графикой в этой программе удобно и легко, смена названий и подписей проходит без проблем.

Statistica обладает очень широкой палитрой статистических методов. Так, начиная с версии 5.11, появились три дополнительных модуля, один из которых - дендрологическое моделирование – является оболочкой для хорошо известной программы QUEST, выполняющей моделирование по алгоритмам дискриминантного разделения и CART. Для сравнения можно отметить, что продукт фирмы SPSS для дендрологического моделирования никогда не входил в комплект полной поставки системы (приобретался отдельно) и до недавнего времени базировался на устаревшем алгоритме (CHAID).

Начиная с версии 5.11 (в настоящее время выпущена Statistica 7.0), можно приобрести интегрируемый в нее пакет нейросететвого моделирования Statistica Neural Networks. Этот пакет по своим возможностям явно превосходит аналогичные модули, доступные для S-Plus или SPSS (NeuroSolutions), однако его стоимость практически равна стоимости всей остальной системы.



Таким образом, Statistica является одной из наиболее простых для неподготовленного пользователя систем с наименьшим периодом освоения ее возможностей. К недостаткам системы можно отнести ее малую расширяемость, отсутствие модулей третьих фирм, а также недостаточно эффективный командный язык.

STATA. Stata является весьма развитой системой статистической обработки данных, существующей на всех основных операционных системах – MS DOS, Windows и UNIX . По своей сути эта программа является не чем иным, как интерпретатором языка программирования статистических задач. Отсюда проистекают все положительные и отрицательные стороны системы. К явно положительным относятся расширяемость, наличие большого количества программ, написанных пользователями системы, полная совместимость процедур, созданных на разных платформах, и легкость программирования собственных статистических программ. Понятно, что все эти достоинства необходимы в первую очередь профессионалам в области статистической обработки данных, но вряд ли произведут большое впечатление на начинающих.

Надо отметить, что оригинальная версия даже не имеет пользовательского интерфейса, а полностью управляется при помощи командного языка. Для того, чтобы облегчить использование STATA, студентами была разработана оболочка StataQuest, которая добавляет к системе

меню и диалоговые окна, позволяющие осуществлять простой доступ к ряду статистических процедур. Однако, поскольку StataQuest разрабатывалась для студентов, она включила доступ лишь к основным процедурам (правда, включая основные виды множественной регрессии, дисперсионного анализа, непараметрической статистики и корреляционного анализа). Кроме того, в STATA встроены достаточно полные графические возможности. Следует отметить, что графики высокого разрешения можно сохранять только в одном из двух форматов – собственном формате STATA и в формате WMF. Последний позволяет использовать изображения в других программах, например, MS Word или MS PowerPoint.

STATA позволяет использовать в командной строке условия, например, рассчитывать суммарные статистики не по всей анализируемой группе, а по определенному поднабору данных. Полно представлены различные методики регрессионного анализа, анализ выживаемости и факторный анализ.

Несколько удивительным является отсутствие среди реализованных алгоритмов кластерного анализа.

В целом STATA ориентирована на пользователей, обладающих некоторыми знаниями как в области статистической обработки данных, так и в программной реализации статистических алгоритмов. Для этой категории пользователей она представляет мощный, быстрый и компактный инструмент.

Statgraphics Centurion XV.I. Данная система была разработана еще для персональных компьютеров, работающих под управлением MS DOS. В те времена она открыла перед пользователями, уставшими от командной строки SAS и SPSS, систему меню, четкую графику высокого разрешения, большие возможности по экспорту графических изображений в сочетании с достаточно полным набором статистических алгоритмов.

Однако на компьютерах, оснащенных операционной системой Windows, Statgraphics уступил свои позиции в качестве "статистической системы №1 для начинающих" пакету Statistica. Вместе с тем до сих пор Statgraphics сохраняет свою приверженность ориентировке на начинающих пользователей в сочетании с мощными возможностями по визуализации данных.

При этом методики параметрической и непараметрической статистик обычно находятся в одном пункте меню и могут быть использованы при просмотре опций данного типа анализа. После каждого анализа идет краткий комментарий того, что было получено, и даются предложения по использованию дополнительных методик. Активно используются опции, вызываемые нажатием правой кнопки мыши.

Если исследователь привык работать с другими программами, которые задают вопросы до тех пор, пока не смогут однозначно выполнить поставленную перед ними задачу, работа в Statgraphics может показаться несколько неуклюжей. Однако для тех, кто начинает работу с этой программы, данный подход может показаться естественным – выбрать тип анализа, указать

переменные, затем получить комментарий по поводу данных и первоначальны результаты, а после этого выбрать уточняющие методики анализа.

Также надо указать, что одной из наиболее сильных сторон Statgraphi являются его возможности по визуализации данных.

S-plus . S-plus является, наверное, одной из самых развитых систе статистического анализа, находящихся на рынке. Основой S-plus является язы с аналогичным именем, разработанный более десяти лет назад в лаборатори ATT. Это язык, специально предназначенный для анализа и исследовательско работы в области статистики.

Современные версии S-plus базируются на ядре языка (реализованном виде динамической библиотеки) и графическом пакете Axum, который так отвечает за графический интерфейс пользователя. S-plus является наиболе всеобъемлющим пакетом.

Именно он наряду со стандартными методами анализа включает дендрологическое моделирование, нечеткий кластерный анализ и ряд други дополнительных возможностей. S-plus позволяет подключать дополнительны модули уже в скомпилированном виде, поэтому расширение системы н составляет труда и достаточно широко доступны дополнительные модули дл робастной статистики, нейросетевого моделирования и ряда других типо анализа. Конечно, работа с дополнительными модулями не столь удобна, как со встроенными, однако возможность их бесплатного получения из Интернет является очень привлекательной.

Графика всегда была одной из сильных сторон S-plus, предоставля пользователю широкий выбор различных высококачественных диаграмм и позволяя достаточно легко манипулировать ими. Не случайно поэтому S-plus широко используется в Северной Америке для обучения студентов статистике.

Существует и ряд недостатков. Отсутствие в меню возможности расчета непараметрических коэффициентов корреляции вызывает некоторое удивление. Кроме того, вывод данных в S-plus не всегда удобен для интерпретации. Так например, команда корреляционного анализа выдает столбцы значений коэффициентов с точностью до восьмого знака после запятой, однако без расчета достоверности коэффициентов или вспомогательных статистик.

Использование командной строки (а значит, и полностью всех возможностей S-plus) требует изучения языка. Надо заметить, что использование командной строки несколько сложнее, чем в SAS, SPSS или Stata.

S-plus является системой, рассчитанной в основном на профессионалов в области статистической обработки данных, исследовательской работы в области статистики и обучение студентов-статистиков. Поэтому данная система является очень мощной, настраиваемой и расширяемой. Вместе с тем значительные требования к аппаратной части компьютера, сложность в использовании командной строки (для доступа к специальным возможностям)

делают S-plus менее привлекательным для начинающих пользователей-непрофессионалов.

Minitab 14.20. Одна из старейших систем обработки данных для персональных компьютеров - Minitab - явно утрачивает свои лидирующие позиции. Несмотря на достаточно большой набор статистических процедур, отсутствуют те из них, которые являются стандартом и необходимы для анализа биомедицинских данных. Так, например, отсутствие непараметрических коэффициентов корреляции вообще трудно объяснимо.

Графика в этой системе достаточно развитая, однако диалоговые окна, через которые необходимо пройти, чтобы вызвать график, никак нельзя назвать интуитивными. Вообще пользовательский интерфейс Minitab является одной из наиболее слабых его сторон. Наползающие друг на друга поля, "съеденные" метки и т.п. придают программе незаконченный вид.

Minitab имеет свой командный язык и ряд макросов, расширяющих возможности системы. Вместе с тем среди этих макросов нет тех, которые кардинально бы расширяли возможности системы, и их число явно уступает таковому для Stata или SAS.

В целом можно отметить, что Minitab расчитан на начинающих пользователей и может ими успешно использоваться, однако он явно уступает по своим возможностям и тщательности проработки Statistica.

SAS . Одна из старейших и наиболее часто используемых систем статистической обработки данных - SAS - начинала свой путь на больших ЭВМ и до сих пор имеет наиболее широкий охват различных компьютерных платформ. SAS имеет программно-модульную структуру, что означает, что существуют специализированные модули обработки данных (статистика - STAT, поддержка принятия решений - OR, графика и т.п.), а внутри модуля имеются программы выполняющие эту обработку. SAS практически не попала под влияние среды Windows, и поэтому версии для этой операционной системы выглядят так же, как и для других сред. Более того, основным способом общения с системой является командная строка.

Графический интерфейс пользователя поставляется в отдельном модуле - SAS/ASSIST - и не предоставляет доступа не только что ко всем возможностям системы, но даже к их большей части. Дело в том, что этот модуль является оболочкой не для всей системы, а лишь для блока общего анализа данных, и поэтому полнота охвата процедур ограничена. Справедливости ради, следует отметить, что на рынке существует распространяемая бесплатно оболочка для SAS - Overstat, которая превращает эту программу в систему, более привычную для пользователя Windows, с полностью настраиваемыми меню и диалоговыми окнами. Данная программа рассчитана на работу с более ранними версиями SAS (6.03-6.04), однако может использоваться и для новых версий.

Сильной стороной SAS являются ее возможности по обработке данных и полнота представленных процедур. Графики, предлагаемые системой,

достаточно впечатляющи, однако не могут сравниться с генерируемыми S-Pl или Statistica.

Еще одной сильной стороной SAS является ее расширяемость. Систе включает командный язык, язык работы с матрицами (IML) и поддержку макр Неудивительно поэтому, что на рынке можно найти достаточно больш количество готовых подпрограмм и макросов для решения различн статистических задач.

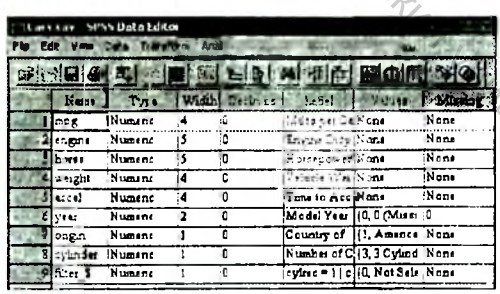
Каждая процедура SAS имеет множество опций, которые позволя выполнять дополнительные тесты и специфицировать дополнительные модел Естественно, полное и гибкое использование всех этих возможностей требуе знания командного языка или создания детальных меню в Overstat.

SAS продолжает нести с собой наследие старых компьютерных систем том, что результаты анализа обычно очень многословны. Система старается рассчитать и распечатать все известные ей тесты, поскольку при работе н старых машинах не могло быть большего разочарования для аналитика, ка получив долгожданную распечатку, обнаружить, что он забыл заказать то самый тест, который ему был нужнее других. Чтобы избежать этого, SA выводит результаты всех тестов.

В целом следует отметить, что SAS является наиболее гибкой и развитой системой обработки данных, которая особенно хорошо подходит для профессионалов в области анализа данных, однако может быть использована и начинающими, если они воспользуются оболочкой Overstat или SAS/ASSIST.

SAS была одной из двух систем, представленных в нашем обзоре, которая смогла достаточно просто справиться с задачей последовательного использования процедур импорта файла, корреляционного анализа и факторного анализа полученной корреляционной матрицы.

SPSS 14.0 (Statistical Package for the Social Science). SPSS является наряду с SAS одной из старейших систем статистического анализа данных.



Однако, в отличие от SAS, компания, производящая эту программу, всегда была в значительной степени ориентирована на непрофессионалов, и поэтому уже с ранних версий для персональных компьютеров (SPSS PC+) программа полагалась на разветвленную систему меню. При этом система

меню была лишь оболочкой (front-end) для ядра программы, управляемого командным языком. Использование меню приводило к автоматическому формированию команд для SPSS, которые затем необходимо было передать командному процессору (тот же принцип используется и Overstat).

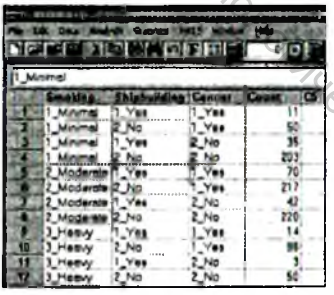
В ходе эволюции SPSS принцип "общения" оболочки с ядром оставался прежним, однако становился все более и более скрытым от пользователя. Так, в последних версиях SPSS для того, чтобы увидеть команды, сформированные оболочкой, необходимо специально сообщить системе о своем желании.

Длительное время ядро SPSS оставалось без изменений, однако в версии 7.5 один из основных компонентов - общая линейная модель (GLM) был переписан.

Кроме того, фирма-производитель отказалась от развития SPSS на платформах, отличных от Windows, что привело к возможности создать систему, имеющую столь привычную для пользователей Windows, сокращая, таким образом, время на обучение пользователей.

Сейчас SPSS включает большое количество статистических процедур, возможности по манипуляции данными и создания графиков. Проработка статистических алгоритмов чрезвычайно тщательная и позволяет хорошо контролировать процесс обработки данных. Большинство опций доступно из меню и диалоговых окон, что выгодно отличает SPSS от оболочек SAS.

NCSS 2006. Система NCSS является относительно молодой, однако это с лихвой компенсируется полнотой охвата статистических процедур внутри одной программы (без дополнительных модулей).



По количеству предлагаемых процедур NCSS напоминает S-plus, предлагая большое количество процедур кластерного анализа, детальную описательную статистику, графики и многие другие статистические методики. Внешний вид программы также напоминает S-plus, особенно диалоговыми окнами с закладками для выбора параметров процедур. Явным достоинством системы является то, что все ее возможности доступны из ниспадающих меню, а сопровождающая программу система подсказки содержит набор пошаговых инструкций с примерами, позволяющий быстро овладеть ее основными

возможностями. Результаты, генерируемые программой, автоматически сохраняются в rtf-файле, который затем легко прочитать и редактировать любым современным текстовым редактором. Сами результаты организованы таким образом, что их легко просматривать и анализировать. Аналогично SAS система по умолчанию рассчитывает большое количество статистик, позволяя охватывать их всех одним взглядом.

Недостатки системы частично кроются в том, что она создавалась при помощи Visual Basic - отнюдь не самого быстрого языка программирования. Следствием оказывается очень большое время загрузки программы и исполнения на относительно маломощных компьютерах. Кроме того, генерируемые программой графики не могут редактироваться. Однако в остальном NCSS является весьма привлекательной системой для начинающих пользователей.

SYSTAT 11.0. Данная программа позиционируется как полномасштабная статистическая система для исследователей. SYSTAT существует в версиях для пользователей обоих основных типов персональных компьютеров - на платформе Windows и Macintosh. Структура SYSTAT очень похожа на структуру всех остальных программ, базирующихся на ядре командного интерпретатора с оболочкой в виде меню и панели кнопок.

Вместе с тем по целому ряду параметров SYSTAT действительно является очень удобной для исследователей системой. Во-первых, эта программа предлагает весьма широкий набор статистических процедур в рамках одного модуля, достаточно компактного и быстрого. Данная система была одной из первых, включивших дендрологическое моделирование в структуру встроенных команд. Кроме того, SYSTAT поддерживает специальный набор процедур для статистического распознавания сигнала. Анализ опросных данных позволяет проводить классический анализ и логистический анализ. Имеется и множество других примеров мощности системы SYSTAT.

В целом система SYSTAT является мощным и удобным инструментом для начинающих пользователей и пользователей среднего звена.

2.7. Итоги сравнения

Сравнивая различные статистические программы, следует учитывать, что практически все они обладают набором стандартных процедур. Алгоритмы, используемые программами, по большей части, стандартные, и различий при использовании той или иной программы нет (было бы удивительно, если бы они существовали). Поэтому на первое место выходят различия в пользовательском интерфейсе, полнота охвата современных статистических методов, программируемость, наличие дополнительных модулей расширения и легкость использования полученных графиков и таблиц в других программах.

Не все перечисленные выше требования могут выполняться одновременно. Так, например, программируемость и расширяемость обычно плохо сочетаются с полнотой пользовательского интерфейса. Вообще достаточно четко проявляется правило - чем более разработан пользовательский интерфейс и графическая подсистема, тем "тяжелее" приложение. На одном полюсе находятся в этом отношении STATA и SAS - управляемые преимущественно из командной строки, но зато с большим количеством легко подключаемых и используемых дополнительных модулей. На другом - Statgraphics+, NCSS и Statistica, которые имеют чрезвычайно привлекательный интерфейс, полный и удобный для начинающих, однако при почти полном отсутствии дополнительных (бесплатных) модулей и подпрограмм.

Те, кто по ходу своей работы сталкивается с необходимостью применять необычные статистические подходы или хочет воспользоваться идеями, созданными и реализованными другими людьми, должен обратить свое внимание на системы, управляемые из командной строки. SAS является наиболее разработанной программой этого класса с большим количеством готовых подпрограмм, широким охватом процедур и языком манипулирования матрицами, что чрезвычайно удобно для написания статистических программ.

Наличие оболочек для SAS (собственный ASSISST и бесплатный OverStat) позволяет легко создавать сложные задания для обработки данных и анализировать их достаточно быстро (все системы, работающие в пакетном режиме, позволяют обсчитывать данные быстрее, чем системы, основанные на меню). Stata несколько отстает от SAS в отношении удобства программирования (субъективный взгляд автора), обладает не столь широкими возможностями, однако это с лихвой компенсируется наличием большого количества дополнительных модулей и подписным листом, в котором пользователи приводят свои программы, рассчитанные на анализ различных статистических моделей.

Естественно, эти две программы вряд ли могут быть рекомендованы для новичков в области обработки данных, которые не собираются часто прибегать к нестандартным статистическим процедурам. Это инструменты лиц, часто занимающихся анализом больших массивов данных, использующих различные статистические подходы и желающих иметь контроль над процессом обработки результатов.

Промежуточную позицию занимает SPSS, являясь одновременно и системой с мощным языком программирования и достаточно дружелюбным к пользователю интерфейсом. Вообще ряд возможностей, предлагаемых SPSS, особенно в области факторного анализа, является самым широким среди всех описанных систем.

Язык SPSS достаточно прост и позволяет автоматизировать часто повторяющиеся задания. В целом SPSS может быть рекомендована пользователям, которые хотят иметь систему с простым, интуитивным интерфейсом, относительно развитой графикой и периодически используют язык программирования для автоматизации более сложных заданий.

Если же речь заходит о начинающих пользователях, то им следует обратить свое внимание на Statistica или Statgraphics. Для людей, относительно ориентирующихся в статистических методиках или начинающих изучение статистики, наиболее адекватной будет использование системы Statistica. Дружелюбный интерфейс, развитая система подсказки и полнота представленных статистических процедур позволяют рекомендовать эту систему начинающим пользователям и непрофессионалам, часто использующим в своей работе статистические методы анализа.

С целью демонстрации возможностей некоторых из перечисленных выше программ (ИС Statistica, пакета NCSS Statistical and Data Analysis и табличного процессора MS EXCEL) в Приложении 3 приведен пример построения и анализа регрессионной модели, оценивающей влияния фондовых показателей деятельности предприятия на уровень рентабельности. с использованием

ГЛАВА 3. ТАБЛИЧНЫЙ ПРОЦЕССОР MICROSOFT EXCEL

Как упоминалось выше, в настоящее время существует большое количество пакетов прикладных программ специального назначения для статистического анализа данных. Эти пакеты обладают дружественным пользователю интерфейсом и позволяют получить результаты в виде высококачественных таблиц и диаграмм.

Табличный процессор (ТП) Excel, не являясь узкоспециализированным пакетом, также поддерживает возможности статистического анализа данных: во-первых, предлагает пользователю стандартные функции статистического анализа, объединенные в категорию «**Статистические**», во-вторых, имеет специальную надстройку «**Пакет Анализа**», значительно упрощающую расчеты.

Ключевым моментом статистического анализа является корреляционно-регрессионный анализ данных и построение математической модели, на основании которой можно выполнять прогнозы рассматриваемых процессов или явлений. Расчет модели и построение прогнозов является процессом достаточно трудоемким, основанным на большом количестве вычислений. Поэтому использование стандартных функций, входящих в пакеты прикладных и специальных программ, значительно облегчает эту процедуру.

Следует отметить, что встроенные функции категории «Статистические» ТП Excel обеспечивают расчет и оценку экономико-математических моделей линейной и **экспоненциальной** структуры, надстройка «Пакет анализа» – только моделей **линейной** структуры. Некоторые другие виды однофакторных экономико-математических моделей (полиномиальная, логарифмическая, степенная) могут быть получены графическим способом.

В практической работе наибольшее распространение получили модели линейной зависимости.

Линейная *однофакторная* модель – это уравнение прямой на плоскости, которая может быть описана уравнением

$$Y = m \cdot x + b, \qquad (3.1)$$

где m – коэффициент уравнения регрессии, представляющий собой тангенс угла наклона прямой $Y = f(x)$ к оси OX;

b – свободный член, равный значению точки пересечения прямой $Y=f(x)$ с осью OY.

Линейная *многофакторная* модель представляет собой уравнение прямой в многомерном пространстве и имеет вид

$$Y = b + m_1x_1 + m_2x_2 + \ldots + m_nx_n, \qquad (3.2)$$

где $x_1, \ldots, x_n$ – независимые переменные (факторы),

Y – зависимая переменная,

M_1, ... , m_n – коэффициенты уравнения регрессии.

Экспоненциальная *однофакторная* модель имеет вид

$$Y = b*m^x, \quad (3)$$

Экспоненциальная *многофакторная* модель имеет вид

$$Y = b*m_1^{x1}*m_2^{x2}* \ . \ . \ .*m_n^{xn}, \quad (4)$$

3.1. Статистические функции Excel

В последних версиях Excel категория «статистические» насчитыва более восьмидесяти функций. Эти функции служат как для расче статистических характеристик объекта – ВЕРОЯТНОСТЬ, ДИСП, КОРРЕ КОВАР, FРАСПРОБР и др., так и для предсказания поведения объекта н основе заданного вида аппроксимации – НАКЛОН, ОТРЕЗОК, ЛИНЕЙ ТЕНДЕНЦИЯ, ПРЕДСКАЗ, ЛГРФПРИБЛ, РОСТ и др.

Рассмотрим использование перечисленных функций для построения оценки математических моделей экономических процессов.

Функции НАКЛОН и ОТРЕЗОК

Параметры *линейной однофакторной модели* вида

$$Y = m \cdot x + b$$

проще всего определить с помощью функций **НАКЛОН** и **ОТРЕЗОК.**

Функция **НАКЛОН** определяет коэффициент наклона линии линейной регрессии m к оси OX.

Формат:

НАКЛОН(известные_значения_Y; известные_значения_X)

известные_начения_Y – это массив зависимого множества наблюдаемой величины.

известные_значения_X – это массив независимого множества наблюдаемой величины. Если аргумент **известные_значения_X** опущен, то предполагают, что это массив значений 1, 2, 3, ... , n той же длины, что и аргумент *известные значения Y.*

Функция **ОТРЕЗОК** вычисляет точку пересечения b линии линейной регрессии с осью OY.

Формат:

ОТРЕЗОК(известные_значения_Y; известные_значения_X)

Аргументы аналогичны аргументам функции **НАКЛОН**.

Функция ЛИНЕЙН

Эта функция рассчитывает статистику ряда с применением метода наименьших квадратов для вычисления уравнения прямой линии, которое наилучшим образом описывает исходные данные. Результатом работы функции является массив, который описывает полученную теоретическую прямую.

Функция ЛИНЕЙН является поистине самой универсальной для расчета параметров линейных моделей. Во-первых, она может использоваться как для расчета однофакторных (3.1), так и многофакторных (3.2) моделей (что определяется размером массива независимых переменных x_i), во-вторых, при желании в качестве результата, кроме коэффициентов уравнения регрессии, можно получить и статистические характеристики, характеризующие построенную модель.

Формат:

ЛИНЕЙН(известные_значения_Y,известные_значения_X, константа; статистика)

Аргумент **известные_значения_Y** – это известные значения Y, для которых параметры X по уравнению определены.

Аргумент **известные_значения_X** – это известные значения независимой переменной *X*. Массив **известные_значения_X** может быть многомерным в отличие от массива **известные_значения_Y**, который является одномерным. Массив *X* может быть опущен, тогда значения *X* устанавливаются автоматически как предварительный ряд чисел, начиная с 1. Но обязательно должно быть соответствие между размерностями массива *X* и *Y*, если массив *X* задан.

Константа – это логическое значение, которое указывает функции, каким образом должен быть определен коэффициент b. Если логическое значение ИСТИНА или оно опущено, то b определяется в обычном порядке. Если константа равна ЛОЖЬ, то коэффициенты подбираются таким образом, чтобы выполнялось равенство `y =m·x (b=0)`.

Статистика – логическое значение, которое может принимать значение ИСТИНА или ЛОЖЬ. Если статистика имеет значение ИСТИНА, то будет представлена дополнительная регрессионная статистика по регрессии, если ЛОЖЬ или опущено, то выходным массивом будет основная статистика, т.е. коэффициенты $m_1, m_2, \ldots, m_n$ и b.

В зависимости от количества независимых переменных уравнение получаемой теоретической прямой может иметь вид (3.1) для однофакторной модели или (3.2) для многофакторной модели.

В качестве результата функция **ЛИНЕЙН** возвращает массив коэффициентов уравнения регрессии и дополнительную статистику по регрессии, как показано в таблице 3.1.

Таблица 3.1. Результаты, возвращаемые функцией **ЛИНЕЙН**

m_n	m_{n-1}	...	m_2	m_1	b
Se_n	Se_{n-1}	...	Se_2	Se_1	Se_b
r^2	Se_y				
F	d_f				
SS_{reg}	SS_{resid}				

Здесь m_1, m_2 ... m_n, b – коэффициенты уравнения регрессии.

Все остальное – дополнительная статистика по регрессии:

Se_1, Se_2, Se_n – стандартные ошибки для коэффициентов m_1, m_2 ... m_n ;

Se_b – стандартная ошибка для свободного члена b;

r^2 – коэффициент детерминированности, который показывает, как близко теоретическое уравнение описывает исходные данные.

Se_y – стандартная ошибка для Y;

F – критерий Фишера используется для определения того, является ли наблюдаемая взаимосвязь между зависимой и независимой переменными случайной или нет;

Df – степень свободы системы (уравнение надежности). Степени свободы полезны для нахождения F-критических значений в статистической таблице. Для определения уровня надежности модели нужно сравнить значения в таблице с F-статистикой, возвращаемой функцией ЛИНЕЙН;

SSreg - регрессионная сумма квадратов;

SSresid – остаточная сумма квадратов.

Функция ПРЕДСКАЗ

Теоретическое (прогнозируемое) значение *Y* для фиксированного значения *X* можно вычислить, не рассчитывая уравнения регрессии, с помощью функции **ПРЕДСКАЗ.** Эта функция на основании линейного тренда вычисляет или предсказывает будущее значение зависимой переменной *Y*, соответствующее заданному *X*-значению, по существующим *X*- и *Y*-значениям.

Формат:

ПРЕДСКАЗ(X; известные_значения_Y; известные_значения_X)

Аргумент **X** – это точка данных, для которой предсказывается значение.

Аргумент **известные_значения_Y** – это зависимый массив или интервал данных.

Аргумент **известные_значения_X** – это независимый массив или интервал данных.

Функция ТЕНДЕНЦИЯ

Эта функция на основании линейного тренда вычисляет или предсказывает будущее значение зависимой переменной *Y*, соответствующее заданному массиву *X*-значений по существующим *X*- и *Y*-значениям.

В отличие от функции **ПРЕДСКАЗ**, функция **ТЕНДЕНЦИЯ** позволяет рассчитать теоретические значения *Y* для массива новых значений *X*.

Формат:

ТЕНДЕНЦИЯ(известные_значения_Y; известные_значения_X; новые_значения_X; константа)

Аргумент **новые_значения_X** – массив (или интервал данных), который должен содержать столбец (или строку) для каждой независимой переменной, как и аргумент **известные_значения_X**. Если аргумент **новые_значения_X** опущен, то предполагается, что он совпадает с аргументом **известные_значения_X.**

Аргумент **константа** – логическое значение, которое указывает, требуется ли, чтобы константа b была равна 0.

Функция ЛГРФПРИБЛ

Функция **ЛГРФПРИБЛ** в регрессионном анализе вычисляет *экспоненциальную* кривую, аппроксимирующую данные, и возвращает массив значений, описывающий эту кривую. Она является универсальной для расчета параметров *экспоненциальных* моделей, так как, кроме коэффициентов уравнения регрессии, может возвращать и дополнительную статистику по регрессии.

Формат:

ЛГРФПРИБЛ(известные_значения_Y;известные_значения_X; константа; статистика)

В отличие от функции **ЛИНЕЙН** аргумент ***константа*** — это логическое значение, которое указывает, требуется ли, чтобы константа b была равна 1.

Если константа имеет значение ИСТИНА или опущено, то b вычисляется обычным образом.

Если константа имеет значение ЛОЖЬ, то b полагается равным 1 и значения m подбираются так, чтобы удовлетворить соотношению $y = m^x$.

Способ использования функции **ЛГРФПРИБЛ** аналогичен функции **ЛИНЕЙН**, с той лишь разницей, что исходный набор данных аппроксимируется не прямой линией, а экспонентой. В соответствии с этим и количеством независимых переменных уравнение регрессии соответствует виду (3.3) или виду (3.4).

Функция РОСТ

Функция **РОСТ** аналогична функции **ТЕНДЕНЦИЯ** и используется для расчета прогнозируемого экспоненциального роста на основании имеющихся данных. Она возвращает значения Y для последовательности новых значений X, задаваемых с помощью существующих X- и Y-значений.

Формат:

РОСТ(известные_значения_Y;известные_значения_X; новые_значения_X; константа)

Функция КОРРЕЛ

Адекватность модели можно оценить с помощью коэффициента детерминированности R^2, который равен квадрату коэффициента корреляции и показывает, насколько точно уравнение, полученное с помощью регрессионного анализа, объясняет взаимосвязи между переменными.

Коэффициент корреляции R, в свою очередь, рассчитывается при помощи функции **КОРРЕЛ** и должен находиться в пределах от -1 до +1. Его положительное значение свидетельствует о прямой связи, отрицательное – об обратной, т.е. при увеличении одной переменной другая переменная уменьшается. Чем ближе его абсолютное значение к единице, тем теснее связь. Связь считается достаточно сильной, если коэффициент корреляции по абсолютной величине превышает 0,7, и слабой, если меньше 0,4.

Формат:

КОРРЕЛ(масив1; массив2)

Аргументы **масив1; массив2** – анализируемые массивы или интервалы значений.

Функция FРАСПРОБР

Функция **FРАСПРОБР** служит для вычисления критерия Фишера - F-статистики, используемой для определения того, является ли наблюдаемая взаимосвязь между зависимой и независимой переменными случайной или нет.

Формат:

FРАСПРОБР(вероятность;степени_свободы1; степени_свободы2)

Вероятность — это вероятность, связанная с F-распределением.
Степени_свободы1 — это числитель степеней свободы.
Степени_свободы2 — это знаменатель степеней свободы.

Функция СТЬЮДРАСПОБР

Функция **СТЬЮДРАСПОБР** Возвращает t-значение распределения Стьюдента как функцию вероятности и числа степеней свободы.

Формат:

СТЬЮДРАСПОБР(вероятность; степени_свободы

Вероятность — вероятность, соответствующая двустороннему распределению Стьюдента.
Степени_свободы — число степеней свободы, характеризующее распределение.

Рассмотрим использование статистических функций для построения регрессионных моделей.

3.1.1. Линейная регрессия

Пример 3.1. В таблице П1 приложения 1 приведены показатели уровня жизни по территориям регионов республики Беларусь за 200Хг. Провести анализ зависимости среднедневной заработной платы, руб. (Y) от среднедушевого прожиточного минимума в день одного трудоспособного, руб. (x).

Разместим таблицу с исходными данными в ячейках A3:C21 рабочего листа Excel. В ячейки B22-B24 внесем значения среднедушевого прожиточного минимума, для которых требуется выполнить прогноз уровня среднедневной заработной платы (см. рис.3.1).

Чтобы выполнить анализ зависимости среднедневной заработной платы, руб. (Y) от среднедушевого прожиточного минимума в день одного трудоспособного, руб. (X) следует построить однофакторную регрессионную модель вида `y= m·x`.

Рассчитаем линейную регрессионную однофакторную модель (см. рис.3.1 и рис.3.2), для чего в ячейки E5:F9 введем функцию ЛИНЕЙН в формате: =ЛИНЕЙН(C6:C21;B6:B21;1;1).

	A	B	C	D	E	F
1	**Пример 3.1**					
2						
3	Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб.	Среднедневная заработная плата, руб.			
4					Функция **ЛИНЕЙН**	
5		X	Y		1,411024962	3816,154305
6	1	2639	7855		0,135696895	359,8045022
7	2	2750	7963		0,885364164	310,2561187
8	3	2610	7785		108,1258598	14
9	4	2440	7580		10408071,91	1347624,028
10	5	2350	7420			
11	6	2510	7550		Функция **ТЕНДЕНЦИЯ**	
12	7	3500	8500		**9460,3**	
13	8	3800	9100		**10165,8**	
14	9	3620	8560		**10871,3**	
15	10	2296	6800			
16	11	2158	6600		Функция **ПРЕДСКАЗ**	
17	12	2064	6600		**9460,3**	
18	13	2688	7700			
19	14	2330	7500		Функция **КОРРЕЛ**	
20	15	1890	6000		**0,885364164**	
21	16	1788	6000			
22	17	*4000*	*9460,3*		Функция **FРАСПОБР**	
23	18	*4500*	*10165,8*		**4,600**	
24	19	*5000*	*10871,3*			

Рис.3.1. Расчет однофакторной регрессионной модели. Результаты

Результатом работы функции является массив значений:

ячейки E5, F5 – коэффициенты уравнения регрессии `m=1.4111` и `b=3816,154`;

ячейка E7 – коэффициент детерминированности R^2 `=0.885`;

ячейка E8 – критерий Фишера `F=108.12`.

Таким образом, однофакторная регрессионная модель, оценивающая влияние среднедушевого прожиточного минимума в день одного трудоспособного на величину среднедневной заработной платы имеет вид

$$Y = 1.411 \cdot x + 3816.154$$

Вывод: поскольку коэффициент детерминированности $R^2=0.885$ лежит в пределах 0,75 – 1 (см. главу 1), расчетное значение критерия Фишера F=108,12 больше табличного (FРАСПОБР(0,05;1;F8)= 4,6), модель следует признать адекватной и использовать для прогнозирования.

В ячейке С22 рассчитаем прогнозное значение среднедневной заработной платы по формуле =E5*B22+F5. В ячейках С23, С24 расчет аналогичен (см. рис.3.2).

В ячейках Е12:Е14 рассчитаем прогнозное значение среднедневной заработной платы для среднедушевого прожиточного минимума, равного 4000 руб., 4500 руб., 5000 руб. (ячейки В22–В24) с использованием функции ТЕНДЕНЦИЯ:

{=ТЕНДЕНЦИЯ(C6:C21;B6:B21;B22:B24;1)}.

	A	B	C	D	E	F
1	Пример 3					
3	Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб.	Среднедневная заработная плата, руб.			
4				Функция ЛИНЕЙН		
6		X	Y	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	
5	11	2633	7865	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	
7	2	2750	7963	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	
8	3	2610	7785	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	
9	4	2440	7580	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	=ЛИНЕЙН(C6:C21;B6:B21;1;1)	
10	5	2350	7420			
11	6	2610	7550	Функция ТЕНДЕНЦИЯ		
12	7	3500	8500	=ТЕНДЕНЦИЯ(C6:C21;B6:B21;B22:B24;1)		
13	8	3900	8700	=ТЕНДЕНЦИЯ(C6:C21;B6:B21;B22:B24;1)		
14	9	3620	8550	=ТЕНДЕНЦИЯ(C6:C21;B6:B21;B22:B24;1)		
15	10	2295	6800			
16	11	2168	6800	Функция ПРЕДСКАЗ		
17	12	2084	6800	=ПРЕДСКАЗ(B22;C6:C21;B6:B21)		
18	13	2688	7700			
19	14	2430	7500	Функция КОРРЕЛ		
20	15	1890	6000	=КОРРЕЛ(C6:C21;B6:B21)^2		
21	16	1788	6000			
22	17	4000	=$E$5*B22+$F$5	Функция ГРАСПОБР		
23	18	4500	=$E$5*B23+$F$5	=ГРАСПОБР(0,05;1;F8)		
24	19	5000	=$E$5*B24+$F$5			

Рис. 3.2. Расчет однофакторной регрессионной модели. Формулы

В ячейке Е15 рассчитаем прогнозное значение среднедневной заработной платы с использованием функции ПРЕДСКАЗ:

=ПРЕДСКАЗ(B22;C6:C21;B6:B21).

В ячейке Е18 рассчитаем значение коэффициента корреляции R^2:

=КОРРЕЛ(C6:C21;B6:B21)^2.

В ячейке Е12 рассчитаем табличное значение критерия Фишера:

=FРАСПОБР(0,05;1;F8).

Полученные значения совпали с результатами, возвращенными функцией ЛИНЕЙН.

Пример 3.2. В *таблице П2 Приложения 2* приведены экономические показатели деятельности предприятия:

Y1 – Производительность труда, млн. руб.
Y2 – Индекс снижения себестоимости продукции, %.
Y3 – Рентабельность, %.
X1 – Трудоёмкость единицы продукции, час.
X2 – Удельный вес рабочих в составе ППП.
X3 – Удельный вес комплектующих изделий.
X4 – Коэффициент сменности оборудования.
X5 – Премии и вознаграждения на одного рабочего, млн. руб.
X6 – Удельный вес потерь от брака в себестоимости продукции, %.
X7 – Фондоотдача, руб.
X8 – Среднесписочная численность ППП.
X9 – Среднегодовая стоимость ОПС, млрд. руб.
X10 – Фонд з/п ППП, млн. руб.
X11 – Фондовооруженность труда, млн. руб.
X12 – Оборачиваемость нормируемых оборотных средств.
X13 – Оборачиваемость ненормируемых оборотных средств.
X14 – Непроизводительные расходы.

Провести анализ влияния трудоемкости единицы продукции X_1, удельного веса комплектующих изделий X_3, коэффициента сменности оборудования X_4, премий и вознаграждений на одного рабочего X_5, среднегодовой стоимости ОПС X_9, и фондовооруженности труда X_{11} на уровень производительности труда Y.

Разместим таблицу и сходными данными в ячейках A3:H33 рабочего листа Excel (см. рис. 3.3). В ячейки C34:H34 введем планируемые показатели деятельности предприятия, по которым требуется выполнить прогноз.

Рассчитаем многофакторную регрессионную модель вида (3.2), для чего в ячейки J3:P7 введем функцию ЛИНЕЙН в формате: {=ЛИНЕЙН(B4:B33;C4:H33;1;1)}. Результатом работы функции является массив значений:

ячейки J3:P3 – коэффициенты уравнения регрессии m_1 ... m_6 и b.
ячейка J5 – коэффициент детерминированности R^2 =0.82686;
ячейка J6 – критерий Фишера F=18.30. (см. рис.3.4).

	A	B	C	D	E	F	G	H
1	Пример 3.2							
2								
3	№	Y	x1	x3	x4	x5	x9	x11
4	1	9,28	0,23	0,40	1,37	1,23	167,69	6,40
5	2	9,36	0,24	0,26	1,49	1,04	186,10	7,80
6	3	12,11	0,19	0,40	1,44	1,80	220,45	9,78
7	4	10,81	0,17	0,50	1,42	0,43	169,30	7,90
8	5	9,35	0,23	0,40	1,35	0,88	39,53	5,35
9	6	9,87	0,43	0,19	1,39	0,57	40,41	9,90
10	7	8,17	0,31	0,25	1,16	1,72	102,96	4,50
11	8	9,12	0,26	0,44	1,27	1,70	37,02	4,88
12	9	5,88	0,49	0,17	1,16	0,84	45,74	3,46
13	10	6,30	0,36	0,39	1,25	0,60	40,07	3,60
33	30	8,16	0,29	0,4	1,28	0,99	91,43	5,23
34	31	?	0,35	0,35	1,25	1,35	95	7,5

Рис. 3.3. Таблица с исходными данными для анализа

J3 fx {=ЛИНЕЙН(B4:B33;C4:H33;1;1)}

	K	L	M	N	O	P
	m6	m5	m4	m3	m2	m1	b
2	0,30029	-0,00041	0,8092	2,80799	6,70493	-8,14536	2,57486
3	0,11663	0,0031	0,44394	2,31778	2,20753	3,96435	4,35314
5	0,82686	0,97701	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
6	18,3072	23	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
7	104,851	21,9547	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д

Рис. 3.4. Результаты, возвращенные функцией ЛИНЕЙН

Таким образом, многофакторная регрессионная модель, оценивающая влияние трудоемкости единицы продукции X_1, удельного веса комплектующих изделий X_3, коэффициента сменности оборудования X_4, премий и вознаграждений на одного рабочего X_5, среднегодовой стоимости ОПС X_9 и фондовооруженности труда X_{11} на уровень производительности труда Y имеет вид

$$Y = -8.14 \cdot X_1 + 6.704 \cdot X_3 + 2.607 \cdot X_4 + 0.809 \cdot X_5 - 0.00041 \cdot X_9 + 0.3 \cdot X_{11} + 2.57.$$

Оценив значения коэффициента детерминированности R^2 и критерия Фишера F, полученное уравнение регрессии следует признать адекватным. Следовательно, его можно использовать для прогнозирования

В ячейке B34 рассчитаем прогноз значения уровня производительности труда по формуле =O3*C34+N3*D34+M3*E34+L3*F34+K3*G34+J3*H34+P3.

В ячейке B35 рассчитаем прогноз значения уровня производительности труда с использованием функции ТЕНДЕНЦИЯ:

=ТЕНДЕНЦИЯ(B4:B33;C4:H33;C34:H34;1).

Прогнозируемое значение уровня производительности труда в обоих случаях получилось одинаковым и равным 8,636 ед.

3.1.2. Экспоненциальная регрессия

Пример 3.3. Обратимся к условию примера 3.1 и построим экспоненциальную однофакторную регрессионную модель, оценивающую влияние среднедушевого прожиточного минимума в день одного трудоспособного (*X*) на величину среднедневной заработной платы (*Y*).

Воспользуемся данными, размещенными в ячейках A3:C21 рабочего листа Excel. (см. рис. 3.1).

Рассчитаем экспоненциальную регрессионную однофакторную модель вида

$$y = b \cdot m^x ,$$

для чего в ячейки H5:I9 введем функцию ЛГРФПРИБЛ в формате {=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)} (см. рис.3.5).

Результатом работы функции является массив значений:

ячейки H5, I5 – коэффициенты уравнения регрессии m=1.0001 и b=4553.49;

ячейка H7 – коэффициент детерминированности R^2 =0.8501;

ячейка H8 – критерий Фишера F=79.42.

Таким образом, экспоненциальная однофакторная регрессионная модель, оценивающая влияние среднедушевого прожиточного минимума в день одного трудоспособного на величину среднедневной заработной платы, имеет вид

$$Y = 4553.49 \cdot 1.0001^x.$$

	H	I
3	Пример 3.3	
4	Функция ЛГРФПРИБЛ	
5	1,000188572	4553,493087
6	2,11582E-05	0,056101546
7	0,850134808	0,046375848
8	79,41728921	14
9	0,185854144	0,032763118
10		
11	Функция РОСТ	
12	9680,47	
13	10637,52	
14	11689,19	
15		
16	Функция КОРРЕЛ	
17	0,9409	
18		
19	Прогноз	
20	9680,47	
21	10637,52	
22	11689,19	

	H	I
3	Пример 3.3	
4	Функция ЛГРФПРИБЛ	
5	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)
6	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)
7	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)
8	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)
9	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)	=ЛГРФПРИБЛ(C6:C21;B6:B21;1;1)
10		
11	Функция РОСТ	
12	=РОСТ(C6:C21;B6:B21;B22:B24;1)	
13	=РОСТ(C6:C21;B6:B21 B22:B24,1)	
14	=РОСТ(C6:C21;B6:B21,B22:B24;1)	
15		
16	Функция КОРРЕЛ	
17	=КОРРЕЛ(C6:C21;B6:B21)	
18		
19	Прогноз	
20	=$I$5*$H$5^B22	
21	=$I$5*$H$5^B23	
22	=$I$5*$H$5^B24	

Рис.3.5. Расчет экспоненциальной однофакторной модели

Вывод: поскольку коэффициент детерминированности R^2=0.8501 лежит в пределах 0,75 – 1 (см. главу 1), расчетное значение критерия Фишера 79,41 больше табличного (4.6), модель следует признать адекватной, и, следовательно, ее можно использовать для прогнозирования.

В ячейке H20 рассчитаем прогнозное значение среднедневной заработной платы при уровне среднедушевого прожиточного минимума в день, равном 4000 руб. (ячейка B22), по формуле `=I5*H5^B22`.

В ячейках H12 – H14 рассчитаем прогнозное значение среднедневной заработной платы с использованием функции РОСТ:

`{=РОСТ(C6:C21;B6:B21;B22:B24;1)}`.

В ячейке H15 рассчитаем значение коэффициента корреляции R:

`=КОРРЕЛ(C6:C21;B6:B21)`.

Полученные значения совпали с результатами, возвращенными функцией ЛГРФПРИБЛ.

Пример 3.4. Обратимся к условию примера 3.2 и построим экспоненциальную многофакторную регрессионную модель, оценивающую влияние трудоемкости единицы продукции X_1, удельного веса комплектующих изделий X_3, коэффициента сменности оборудования X_4, премий и вознаграждений на одного рабочего X_5, среднегодовой стоимости ОПС X_9 и фондовооруженности труда X_{11} на уровень производительности труда Y.

Воспользуемся данными, размещенными в ячейках A3:H33 рабочего листа Excel. (см. рис. 3.3).

В ячейки C34:H34 введены планируемые показатели деятельности предприятия, по которым требуется выполнить прогноз.

Рассчитаем экспоненциальную многофакторную регрессионную модель вида (3.4), для чего в ячейки J10:P14 введем функцию ЛГРФПРИБЛ в формате {=ЛГРФПРИБЛ(B4:B33;C4:H33;1;1)}. Результатом работы функции является массив значений:

ячейки J10:P10 – коэффициенты уравнения регрессии m_1 ... m_6 и b.

ячейка J12 – коэффициент детерминированности R^2 =0.7787;

ячейка J13 – критерий Фишера F=13.48 (см. рис.3.6).

P14 (=ЛГРФПРИБЛ(B4:B33;C4:H33;1;1))

	J	K	L	M	N	O	P
9							
10	m6	m5	m4	m3	m2	m1	b
11	1,03856	0,98972	1,11956	1,33168	2,10645	0,25981	4,96877
12	0,01662	0,00044	0,06328	0,33036	0,31464	0,56505	0,62046
13	0,7787	0,13926	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
14	13,4886	23	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
15	1,58944	0,44602	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д

Рис.3.6. Результаты, возвращенные функцией ЛГРФПРИБЛ

Таким образом, многофакторная регрессионная модель, оценивающая влияние трудоемкости единицы продукции X_1, удельного веса комплектующих изделий X_3, коэффициента сменности оборудования X_4, премий и вознаграждений на одного работника X_5, среднегодовой стоимости ОПС X_9 и фондовооруженности труда X_{11} на уровень производительности труда Y имеет вид

$$Y = 4{,}968 \cdot 0{,}259^{x1} \cdot 2.106^{x3} \cdot 1.331^{x4} \cdot 1.119^{x5} \cdot 0.999^{x9} \cdot 1.0305^{x11}.$$

Оценив значения коэффициента детерминированности R^2 и критерия Фишера F, полученное уравнение регрессии следует признать адекватным. Следовательно, его можно использовать для прогнозирования.

В ячейке B36 рассчитаем прогноз значения уровня производительности труда по формуле

=P10*O10^C34*N10^D34*M10^E34*L10^F34*K10^G34*J10^H34.

В ячейке B37 рассчитаем прогноз значения уровня производительности труда с использованием функции РОСТ:

{=РОСТ(B4:B33;C4:H33;C34:H34;1)}.

Прогнозное значение уровня производительности труда в обоих случаях получилось одинаковым и равным 8,183 ед.

3.2. Надстройка «Пакет анализа»

Самым мощным средством ТП Excel для решения сложных статистических и инженерных задач вообще и для проведения полного корреляционно-регрессионного анализа в частности является специальная надстройка **Пакет Анализа**, которая значительно упрощает процесс расчета.

Для проведения анализа данных с помощью этого инструмента необходимо указать входные данные и выбрать параметры. Анализ будет проведен с использованием подходящей статистической или инженерной макрофункции, а результаты будут помещены в выходной диапазон. Кроме того, в этой надстройке имеются средства, которые позволяют представить результаты анализа в графическом виде. В состав пакета анализа входят, в частности, дисперсионный и корреляционный анализ, ковариационный анализ, описательная статистика, экспоненциальное сглаживание и др.

Основными этапами построения и анализа регрессионной модели с помощью пакета анализа являются:

- построение системы показателей (факторов);
- сбор и предварительный анализ исходных данных. Построение матрицы коэффициентов парной корреляции;
- выбор вида модели и численная оценка ее параметров;
- проверка качества модели;
- оценка влияния отдельных факторов на основе модели;

- прогнозирование на основе модели регрессии.

Выбор факторов, влияющих на исследуемый показатель, производится на основании качественного и количественного анализа исследуемых явлений.

Построим и проанализируем многофакторную регрессионную модель при помощи надстройки «пакет анализа».

Пример 3.5. Построить линейную многофакторную регрессионную модель, и провести анализ зависимости производительности труда *(Y)* от следующих факторов:

X1 – трудоемкость единицы продукции;

X2 – удельный вес рабочих в составе ППП;

X3 – удельный вес комплектующих изделий;

X4 – коэффициент сменности оборудования;

X5 – премии и вознаграждения на одного работника;

X 6 – удельный вес потерь от брака;

X7 – фондоотдача;

X8 – среднегодовая численность ППП;

X9 – среднегодовая стоимость ОПС;

X10 – среднегодовой фонд заработной платы;

X11 – фондовооруженность труда;

X12 – оборачиваемость нормированных оборотных средств;

X13 – оборачиваемость ненормированных оборотных средств;

X14 – непроизводительные расходы.

В качестве исходных показателей используем данные из таблицы П2 Приложения 2.

Для проведения корреляционного анализа необходимо выполнить следующие действия:

- расположить в смежных диапазонах ячеек данные для корреляционного анализа (ячейки A1:P32, см. рис. 3.6);

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
	№	Производительность труда	Трудоемкость единицы продукции	Удельный вес рабочих в составе ППП	Удельный вес покупных изделий	Коэффициент сменного оборудования	Премии и вознаграждения на одного рабочего	Удельный вес потерь от брака	Фондоотдача	Среднегодовая численность ППП	Среднегодовая стоимость ОПФ	Среднегодовой фонд зар ППП	Фондовооруженность труда	Оборачиваемость нормируемых оборотных средств	Оборачиваемость ненормируемых оборотных средств	Непроизводственные расходы
2		Y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
3	1	9,26	0,23	0,76	0,40	1,37	1,23	0,23	1,45	26006,00	167,69	47750,00	6,40	168,32	10,08	17,72
4	2	9,38	0,24	0,75	0,26	1,49	1,04	0,39	1,30	23936,00	186,10	50391,00	7,80	92,88	14,76	18,39
5	3	12,11	0,19	0,68	0,40	1,44	1,80	0,43	1,37	22589,00	220,45	43149,00	9,76	159,04	6,46	26,46
6	4	10,81	0,17	0,70	0,50	1,42	0,43	0,18	1,65	21220,00	169,30	41088,00	7,90	93,96	21,96	22,37
7	5	9,35	0,23	0,62	0,40	1,36	0,88	0,15	1,91	7394,00	39,53	14257,00	5,36	173,86	11,88	28,13
8	6	9,87	0,43	0,76	0,19	1,39	0,57	0,34	1,68	11586,00	40,41	22661,00	9,90	162,30	12,60	17,55
9	7	8,17	0,31	0,73	0,25	1,16	1,72	0,38	1,94	26609,00	102,96	52509,00	4,50	86,56	11,52	21,92
10	8	9,12	0,26	0,71	0,44	1,27	1,70	0,09	1,89	7801,00	37,02	14903,00	4,88	101,16	8,28	19,52

Рис.3.6. Фрагмент таблицы с исходными данными

- выбрать команду *Сервис-->Анализ данных*;
- в диалоговом окне *Анализ данных* выбрать инструмент *Корреляция*;
- в диалоговом окне *Корреляция* в поле *«Входной интервал»* ввести диапазон ячеек, содержащих исходные данные (B2:P29). Если выделены и заголовки столбцов (а они выделены), то установить флажок *«Метки в первой строке»*;
- установить переключатель *новый рабочий лист* или *выходной диапазон*;
- ОК (см. рис. 3.7).

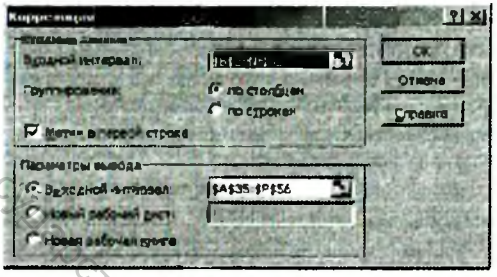


Рис. 3.7. Окно Корреляция

В результате в ячейках, указанных в поле «Выходной интервал», получим матрицу коэффициентов парной корреляции (см. рис. 3.8).

	A	B	C	D	E	F	G	H	I	J	K	M	N	O	P
34															
35		Y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14
36	Y	1														
37	x1	-0,82	1													
38	x2	0,096	-0,14	1												
39	x3	0,642	-0,66	-0,05	1											
40	x4	0,508	-0,54	0,392	0,055	1										
41	x5	0,408	-0,28	0,091	0,063	0,2082	1									
42	x6	-0,16	0,01	0,354	-0,43	0,1971	-0,05	1								
43	x7	0,294	-0,23	0,055	0,6	-0,16	-0,1	-0,27	1							
44	x8	0,556	-0,43	0,135	0,48	0,2398	0,435	-0,18	0,11	1						
45	x9	0,646	-0,52	0,066	0,431	0,3648	0,489	-0,14	-0,13	0,895775	1					
46	x10	0,567	-0,43	0,13	0,47	0,2407	0,445	-0,18	0,11	0,997008	0,902	1				
47	x11	0,419	-0,23	-0,07	-0,14	0,4393	0,255	0,08	-0,66	0,204716	0,496	0,21165	1			
48	x12	0,128	0,17	0,179	0,248	0,1259	-0,04	-0,15	0,29	0,178373	0,021	0,15734	-0,14	1		
49	x13	-0,16	0,031	0,032	0,143	-0,115	-0,59	0,071	0,2	-0,20137	-0,207	-0,20249	-0,23	0,0308	1	
50	x14	-0,01	0,095	0,95	0,177	-0,418	-0,06	-0,39	0,03	0,031177	0,059	0,0282	0,089	-0,132	-0,024	1

Рис. 3.8. Матрица коэффициентов парной корреляции

Полученная матрица коэффициентов парной корреляции показывает, что зависимая переменная Y – производительность труда - имеет *тесную* обратную связь с фактором X_1 – трудоемкость единицы продукции (r_{yx1}= -0,82) и *среднюю* прямую связь с факторами

X_3 – удельный вес комплектующих изделий (r_{yx3} = 0,642),
X_4 – коэффициент сменности оборудования (r_{yx4} = 0.508),
X_5 – премии и вознаграждения на одного работника (r_{yx5} = 0.408),
X_8 – среднегодовая численность ППП (r_{yx8} = 0.556),
X_9 – среднегодовая стоимость ОПС (r_{yx9} = 0.646),
X_{10} – среднегодовой фонд заработной платы (r_{yx10} = 0.567),
X_{11} – фондовооруженность труда (r_{yx11} = 0.419).
Влияние остальных факторов незначительно.

Кроме того, факторы X_8, X_9, X_{10} мультиколлинеарны, о чем свидетельствуют парные коэффициенты корреляции (r_{x8x9} = 0.896, r_{x9x10} =0.902 и r_{x8x10} = 0,997). Следовательно, из этих трех признаков следует оставить один *X9*, как наиболее коррелирующий с результирующим признаком *Y*.

Таким образом, регрессионную модель будем строить по следующим независимым факторам-признакам: X_1 – трудоемкость единицы продукции, X_3 – удельный вес покупных изделий, X_4 – коэффициент сменности оборудования, X_5 – премии и вознаграждения на одного работника, X_9 – среднегодовая стоимость ОПС, X_{11} – фондовооруженность труда.

Данные для уточненной модели (столбцы значений *Y*, X_1, X_3, X_4, X_5, X_9, X_{11}) разместим в ячейках S2:Y32. Затем

- выберем команду *Сервис-->Анализ данных*;
- в диалоговом окне *Анализ данных* выберем инструмент *Регрессия* (см. рис. 3.9);

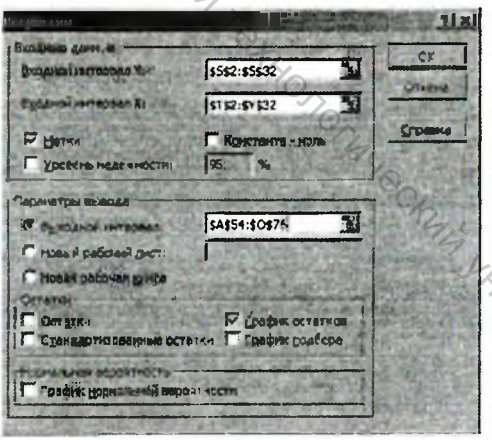


Рис.3.9. Окно Регрессия

- в диалоговом окне *Регрессия* в поле *«Входной интервал Y»* введем диапазон ячеек, содержащих значения производительности труда Y (ячейки S2:S32);

- в поле *«Входной интервал X»* введем диапазон ячеек, содержащих значения факторов X_1, X_3, X_4, X_5, X_9, X_{11} (ячейки T2:Y32). Если выделены и заголовки столбцов (а они выделены), то следует установить флажок *«Метки»;*
- установим переключатель *выходной диапазон и укажем, в какие ячейки текущего рабочего листа Excel (A54:O76) выводить результаты регрессии*, OK.

Результаты регрессии, оформленные в виде трех таблиц, будут представлены на текущем рабочем листе.

Первая таблица *«Регрессионная статистика»* (см. рис.3.10) содержит показатели регрессии.

	A	B
53		
54	ВЫВОД ИТОГОВ	
55		
56	*Регрессионная статистика*	
57	Множественный R	0,909320565
58	R-квадрат	0,82686389
59	Нормированный R-квадрат	0,781697948
60	Стандартная ошибка	0,977010643
61	Наблюдения	30

Рис. 3.10. Регрессионная статистика

Коэффициент детерминации (R^2 = 0,8268), размещенный в ячейке B58, показывает долю вариации результативного признака Y под действием изучаемых факторов X_1, X_3, X_4, X_5, X_9, X_{11}. Следовательно, около 83% вариации зависимой переменной учтено в модели и обусловлено влиянием изучаемых факторов.

Вторая таблица *«Дисперсионный анализ»* (см. рис. 3.11) позволяет осуществить проверку значимости уравнения регрессии.

	A	B	C	D	E	F
62						
63	Дисперсионный анализ					
64		*df*	*SS*	*MS*	*F*	*Значимость F*
65	Регрессия	6	104,8510957	17,4751826	18,307	1,04009E-07
66	Остаток	23	21,95465432	0,95455019		
67	Итого	29	126,80575			

Рис.3.11. Дисперсионный анализ

Регрессионная и остаточная суммы квадратов показаны в ячейках C65 и C66 соответственно. Регрессионная сумма квадратов довольно существенно превосходит остаточную. Это говорит о том, что большая часть вариации производительности труда Y связана с факторами X_1, X_3, X_4, X_5, X_9, X_{11}.

В ячейке E65 показано расчетное значение F-критерия Фишера, а в ячейке B66 число степеней свободы, которое определяется как количество наблюдений (30) за вычетом параметров линейной регрессии (6) и единицы.

Проведем оценку значимости связи, сравнив табличное и расчетное значения F-критерия, который показывает, является ли наблюдаемая взаимосвязь между зависимой и независимой переменными случайной или нет. Для этого можно использовать функцию FРАСПОБР(0.05; 6; 23).

При уровне значимости 0,05 и числе степеней свободы 6 и 23 табличное значение F-критерия составляет около 2,52. Расчетное значение 18,37 гораздо больше, что свидетельствует о статистической значимости связи, и, значит, уравнение регрессии следует считать адекватным.

Третья таблица «Регрессионная модель» (см. рис. 3.12) показывает коэффициенты уравнения регрессии – b, m_1, ... m_6 (ячейки B70:B76 соответственно) и позволяет оценить из значимость.

	A	B	C	D	E	F	G	H
68								
69		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*
70	Y-пересечение	2,5746634	4,353143038	0,591495245	0,559958	-6,43028681	11,58001363	-6,430286806
71	x1	-8,1453571	3,964353034	**-2,054649774**	0,051445	-16,346235	0,055520645	-16,34623499
72	x3	6,704935	2,207530943	**3,037300562**	0,005836	2,138315502	11,27155443	2,138315502
73	x4	2,5079982	2,317782529	1,125208307	0,272106	-2,186703789	7,202699986	-2,186703786
74	x5	0,8092029	0,443940278	1,822774238	0,081364	-0,109156628	1,727562085	-0,109156282
75	x9	-0,0004095	0,003102692	-0,131976258	0,896151	-0,006627788	0,006008917	-0,006627881
76	x11	0,3002944	0,116627499	**2,574816165**	0,016936	0,059032335	0,541556405	0,059032335

Рис.3.12. Регрессионная модель

Оценка значимости этих коэффициентов производится с использованием t-критерия Стьюдента (ячейки D70:D76) и уровня значимости $p < 0.05$ (ячейки E70:E76). Табличное значение t-критерия при уровне значимости 5% и степенях свободы n = 23 (функция =СТЬЮДРАСПОБР(0,05;23)) составляет 2,068, что меньше фактических значений t-критерия Стьюдента коэффициентов при X_3 и X_{11}, с натяжкой соответствует t-критерию коэффициента при X_1 и больше значений t-критерия коэффициентов при X_4, X_5, X_9, т. е. не все коэффициенты существенны.

Формально уравнение регрессии может быть представлено в виде

$$Y = -8.145\ X1 + 6{,}7 \cdot X3 + 2.607 \cdot X4 + 0.809 \cdot X5 - 0.0004 \cdot X9 + 0.3 \cdot X11 + 2.575 \quad (3.5)$$

Но произведенный анализ наблюдений показывает, что производительность труда *(Y)* весьма тесно связана только с тремя факторами: трудоемкостью единицы продукции (X_1), удельным весом покупных изделий (X_3) и фондовооруженностью (X_{11}), и следовательно, факторы X_4, X_5 и X_9 из модели должны быть исключены. Просто удалить соответствующие слагаемые из уравнения 3.5 нельзя, так как результат будет искажен.

Поэтому, пользуясь приведенной выше методикой либо с помощью статистической функции ЛИНЕЙН, (см. п.3.1.1) рассчитаем уравнение регрессии, отражающее влияние трудоемкости единицы продукции *(X_1)*, удельного веса покупных изделий (X_3) и фондовооруженности *(X_{11})* на производительность труда *(Y)*.

Разместим исходные данные (значения факторов *Y*, *X_1*, X_3 и *X11*) в ячейках AA2:AD32 рабочего листа Excel (см. рис. 3.13).

	AA	AB	AC	AD
1	Производительность труда	Трудоёмкость единицы продукции	Удельный вес покупных изделий	Фондовооруженность труда
2	Y	x1	x3	x11
3	9,26	0,23	0,40	6,40
4	9,38	0,24	0,26	7,80
5	12,11	0,19	0,40	9,75
6	10,81	0,17	0,50	7,90
7	9,35	0,23	0,40	5,35
8	9,87	0,43	0,19	9,90
9	8,17	0,31	0,25	4,50
10	9,12	0,26	0,44	4,88

Рис.3.13. Исходные данные для модели

В ячейки J69:M73 рабочего листа Excel введем функцию ЛИНЕЙН в формате =ЛИНЕЙН(AA3:AA32;AB3:AD32;1;1) и получим результат (см. рис.3.14). При этом способе расчета будет представлена не вся дополнительная статистика по регрессии Здесь верхняя строка (ячейки J69:M69) – коэффициенты уравнения регрессии, ячейка J71 – коэффициент детерминированности $R^2 = 0.789$, ячейка J72 – критерий Фишера F = 32.47.

Эта модель отличается большей устойчивостью..

J69 {=ЛИНЕЙН(AA3:AA32;AB3:AD32,1;1)}

	K	L	M	N	O	P
68						
69	0,348477	5,436	-12,1768	8,085		
70	0,098498	1,874	3,24752	1,785		
71	0,789342	1,014	#Н/Д	#Н/Д		
72	32,47419	26	#Н/Д	#Н/Д		
73	100,0931	26,71	#Н/Д	#Н/Д		

Рис.3.14. Результаты, возвращенные функцией ЛИНЕЙН

Таким образом, уравнение регрессии зависимости производительности труда Y от трудоемкости единицы продукции (X_1), удельного веса покупных изделий (X_3) и фондовооруженности (X_{11}) имеет вид

$$Y = 12{,}176 \cdot X1 + 5{,}436 \cdot X3 + 0{,}348 \cdot X11 + 8.085. \quad (3.6)$$

Рассчитаем ошибку вычислений по уравнению 3.5 и 3.6. Для этого в качестве параметров X1 –X11 подставим результаты первого наблюдения.

Значение производительности труда Y, рассчитанное по уравнению 3.5, составляет

$Y = 8.145 \cdot 0{,}23 + 6{,}7 \cdot 0{,}4 + 2.6 \cdot 1{,}37 + 0.809 \cdot 1{,}23 - 0.0004 \cdot 167{,}69 + 0.3 \cdot 6{,}4 + 2.575 =$ **9,804.**

Значение производительности труда Y, рассчитанное по уравнению 3.6, составляет

$Y = 12{,}176 \cdot 0{,}23 + 5{,}436 \cdot 0{,}4 + 0{,}348 \cdot 6{,}4 + 8.085 =$ **9,63.**

В первом случае ошибка прогноза = (9,804-9,26)/9,26 = 5,9%.

Во втором случае ошибка прогноза = (9,63-9,26)/9,26 = 4,8%.

Очевидно, что уравнение множественной регрессии (3.6) следует признать адекватным и использовать для выполнения прогнозов.

3.3. Утилита «Поиск решения»

Прикладной программный продукт Excel фирмы Microsoft содержит в своем составе достаточно мощное средство для решения задач оптимизации с учетом ограничений. Это так называемая утилита "Поиск решения" (см. рис. 3.15). Прокомментируем некоторые аспекты работы с этой утилитой.

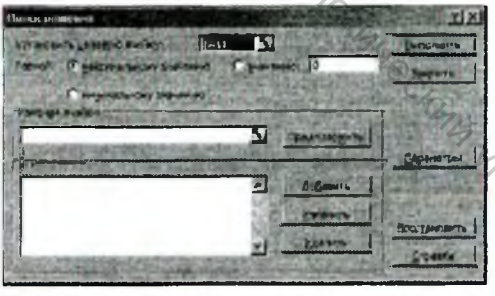


Рис.3.15. Окно утилиты Поиск решения

Искомые переменные - ячейки рабочего листа Excel - называются регулируемыми ячейками.

Целевая функция $F(x_1, x_2, \ldots, x_n)$, называемая иногда просто целью, должна задаваться в виде формулы в ячейке рабочего листа. Эта формула может содержать функции, определенные пользователем, и должна зависеть

(ссылаться) от регулируемых ячеек. В момент постановки задачи определяется, что делать с целевой функцией. Возможен выбор одного из вариантов:

- найти максимум целевой функции $F(x_1, x_2, \ldots, x_n)$;
- найти минимум целевой функции $F(x_1, x_2, \ldots, x_n)$;
- добиться того, чтобы целевая функция $F(x_1, x_2, \ldots, x_n)$ имела фиксированное значение: $F(x_1, x_2, \ldots, x_n) = a$. (см. рис 3.16).

Рис.3.16. Определение целевой функции в окне утилиты Поиск решения

Функции $G(x_1, x_2, \ldots, x_n)$ называются ограничениями. Их можно задать как в виде равенств, так и неравенств. На регулируемые ячейки можно наложить дополнительные ограничения: неотрицательности и/или целочисленности, тогда искомое решение ищется в области положительных и/или целых чисел (см. рис.3.17)..

Рис. 3.17. Определение ограничений

Под эту постановку попадает самый широкий круг задач оптимизации, в том числе решение различных уравнений и систем уравнений, задачи линейного и нелинейного программирования.

Управление диалоговым окном поиска решения (см. рис. 3.15).

Установить целевую ячейку

Служит для указания целевой ячейки, значение которой необходимо максимизировать, минимизировать или установить равным заданному числу. Эта ячейка должна содержать формулу для вычисления целевой функции.

Равной

Служит для выбора варианта оптимизации значения целевой ячейки (максимизация, минимизация или подбор заданного числа). Чтобы установить число, введите его в поле.

Изменяя ячейки

Служит для указания ячеек, значения которых изменяются в процессе поиска решения до тех пор, пока не будут выполнены наложенные ограничения и условие оптимизации значения ячейки, указанной в поле Установить целевую ячейку. В этих ячейках должны содержаться переменные оптимизационной модели.

Ограничения

Служит для отображения списка граничных условий поставленной задачи.

Выполнить

Служит для запуска поиска решения поставленной задачи.

Закрыть

Служит для выхода из окна диалога без запуска поиска решения поставленной задачи. При этом сохраняются установки, сделанные в окнах диалога.

Параметры

Служит для отображения диалогового окна Параметры поиска решения, в котором можно загрузить или сохранить оптимизируемую модель и указать предусмотренные варианты поиска решения.

Восстановить

Служит для очистки полей окна диалога и восстановления значений параметров поиска решения, используемых по умолчанию.

Диалоговое окно "Параметры поиска решения"

Можно изменять условия и варианты поиска решения для линейных и нелинейных задач, а также загружать и сохранять оптимизируемые модели. Значения и состояния элементов управления, используемые по умолчанию, подходят для решения большинства задач (см. рис.3.18).

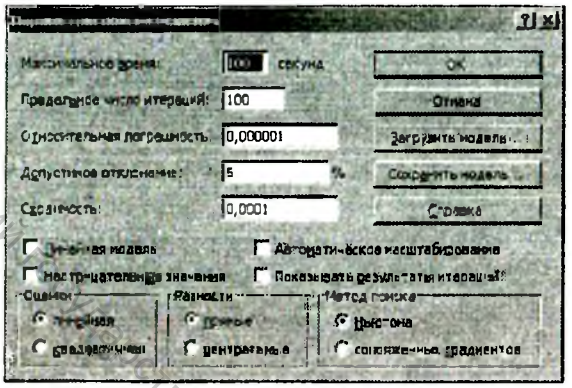


Рис.3.18. Диалоговое окно "Параметры поиска решения"

Максимальное время

Служит для ограничения времени, отпускаемого на поиск решения задачи. В поле можно ввести время (в секундах), не превышающее 32767; значение 100, используемое по умолчанию, подходит для решения большинства простых задач.

Итерации

Служит для управления временем решения задачи путем ограничения числа промежуточных вычислений. В поле можно ввести значение, не превышающее 32767; значение 100, используемое по умолчанию, подходит для решения большинства простых задач.

Точность

Служит для задания точности, с которой определяется соответствие ячейки целевому значению или приближение к указанным границам. Поле должно содержать десятичную дробь от 0 (нуля) до 1. Чем больше десятичных знаков в задаваемом числе, тем выше точность — например, число 0,0001 представлено с более высокой точностью, чем 0,01.

Допустимое отклонение

Служит для задания допуска на отклонение от оптимального решения, если множество значений влияющей ячейки ограничено множеством целых

чисел. При указании большего допуска поиск решения заканчивается быстрее.

Сходимость

Когда относительное изменение значения в целевой ячейке за последние пять итераций становится меньше числа, указанного в поле Сходимость, поиск прекращается. Сходимость применяется только к нелинейным задачам, условием служит дробь из интервала от 0 (нуля) до 1. Лучшую сходимость характеризует большее количество десятичных знаков — например, 0,0001 соответствует меньшему относительному изменению по сравнению с 0,01. Лучшая сходимость требует больше времени на поиск оптимального решения.

Линейная модель

Служит для ускорения поиска решения линейной задачи оптимизации.

Показывать результаты итераций

Служит для приостановки поиска решения для просмотра результатов отдельных итераций.

Автоматическое масштабирование

Служит для включения автоматической нормализации входных и выходных значений, качественно различающихся по величине — например, максимизация прибыли в процентах по отношению к вложениям, исчисляемым в миллионах рублей.

Значения не отрицательны

Позволяет установить нулевую нижнюю границу для тех влияющих ячеек, для которых она не была указана в поле **Ограничение** диалогового окна.

Добавить ограничение

Оценка

Служит для указания метода экстраполяции — линейная или квадратичная — используемого для получения исходных оценок значений переменных в каждом одномерном поиске.

Линейная – служит для использования линейной экстраполяции вдоль касательного вектора.

Квадратичная – служит для использования квадратичной экстраполяции, которая дает лучшие результаты при решении нелинейных задач.

Производные

Служит для указания метода численного дифференцирования — прямые или центральные производные — который используется для вычисления частных производных целевых и ограничивающих функций.

Прямые – используется в большинстве задач, где скорость изменения ограничений относительно невысока.

Центральные – используется для функций, имеющих разрывную производную. Данный способ требует больше вычислений, однако его применение может быть оправданным, если выдается сообщение о том, что получить более точное решение не удается.

Метод

Служит для выбора алгоритма оптимизации — метод Ньютона или сопряженных градиентов — для указания направление поиска.

Метод Ньютона. Реализация квазиньютоновского метода, в котором запрашивается больше памяти, но выполняется меньше итераций, чем в методе сопряженных градиентов.

Поиск решения позволяет представлять результаты в виде трех отчетов: **Результаты**, **Устойчивость** и **Пределы**.

Для генерации одного или нескольких отчетов выделяются их названия в окне диалога Результаты поиска решения.

Отчет по устойчивости содержит информацию о том, насколько целевая ячейка чувствительна к изменениям ограничений и переменных. Этот отчет имеет два раздела: один для изменяемых ячеек, а второй для ограничений. Раздел для изменяемых ячеек содержит значение нормированного градиента, которое показывает, как целая ячейка реагирует на увеличение значения в соответствующей изменяемой ячейке на одну единицу. Подобным образом, множитель Лагранжа в разделе для ограничений показывает, как целевая ячейка реагирует на увеличение соответствующего значения ограничения на одну единицу.

Отчет по результатам содержит три таблицы: в первой приведены сведения о целевой функции до начала вычисления, во второй - значения искомых переменных, полученные в результате решения задачи, в третьей - результаты оптимального решения для ограничений. Этот отчет также содержит информацию о таких параметрах каждого ограничения, как статус и разница. Статус может принимать три состояния: связанное, несвязанное или невыполненное. Значение разницы - это разность между значением, выводимым в ячейке ограничения при получении решения, и числом, заданным в правой части формулы ограничения.

Отчет по пределам содержит информацию о том, в каких пределах значения изменяемых ячеек могут быть увеличены или уменьшены без нарушения ограничений задачи. Для каждой изменяемой ячейки этот отчет содержит оптимальное значение, а также наименьшие значения, которые ячейка может принимать без нарушения ограничений.

Сформулируем простую экономическую задачу для определения номенклатуры и количества выпускаемой продукции с целью максимизации прибыли.

Пример 3.6. Фирма производит три вида продукции (A, B, C), для выпуска каждого требуется определенное время обработки на всех четырех устройствах

Вид продукции	Время обработки, ч.				Прибыль, у.е.
	I	II	III	IV	
A	1	3	1	2	3
B	6	1	3	3	6
C	3	3	2	4	4

Пусть время работы на устройствах I, II, III, IV составляет соответственно 84, 42, 21 и 42 часа.

Требуется рассчитать план производства, обеспечивающий максимальную прибыль.

Разместим таблицу с исходными данными в ячейках A1:G9 Рабочего листа Excel и выполним необходимые предварительные расчеты (см. рис.3.19).

	A	B	C	D	E	F	G
1	Вид продукции			Время обработки, ч.		Прибыль, у.е	Кол-во
2		I	II	III	IV		
3	A	1	3	1	2	3	0
4	B	6	1	3	3	6	0
5	C	3	3	2	4	4	0
6							0
7	Время	0	0	0	0		Прибыль
8							
9	Огранич	84	42	21	42		

	A	B	C	D	E	F	G
1	Вид продукции			Время обработки, ч.		Прибыль, у.е	Кол-во
2		I	II	III	IV		
3	A	1	3	1	2	3	0
4	B	6	1	3	3	6	0
5	C	3	3	2	4	4	0
6							=СУММПРОИЗВ(F3:F5;G3:G5)
7	Время	=СУММПРОИЗВ(B3:B5;G3:G5)	СУММ	=СУММПРОИЗ	=СУМ		Прибыль
8							
9	Огранич	84	42	21	42		

Рис. 3.19. Исходные данные оптимизационной задачи

Отыскать решение задачи, приняв следующие условия:

1. Общая итоговая прибыль (G6) => max.
2. Количество изделий (G3:G5)- целое и неотрицательное число.

3. Баланс времени по каждому устройству (B7:E7) <= (B9:E9).
4. Изменению подлежат: количество изделий (G3:G5).
Окончательный вид формулировки задачи представлен на рис. 3.20.

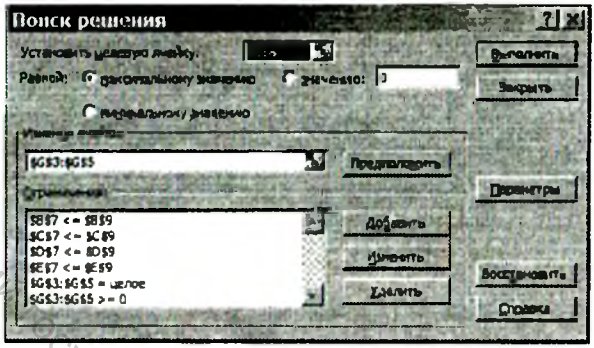


Рис.3.20. Формулировка задачи в терминах рабочего листа Excel

Итоговый результат представлен на рис.3.21.

	A	B	C	D	E	F	G
1	Вид продукции			Время обработки, ч.		Прибыль, у.е	Кол-во
2		I	II	III	IV		
3	A	1	3	1	2	3	12
4	B	6	1	3	3	6	3
5	C	3	3	2	4	4	0
6							54
7	Время	30	39	21	33		Прибыль
8							
9	Огранич.	84	42	21	42		

Рис.3.21. Результат оптимизации

Анализ решения показывает, что все без исключения требования задачи оптимизации выполнены. При этом видно, что для получения максимальной прибыли нецелесообразно выпускать изделие С.

Результаты расчетов представлены в отчете по результатам (рис.3.22).

	A	B	C	D	E	F	G
1	Microsoft Excel 10.0 Отчет по результатам						
2	Рабочий лист: [Книга1 (version 2).xls]Лист1						
3	Отчет создан: 13.11.2006 20:44:25						
4							
5							
6	Целевая ячейка (Максимум)						
7		Ячейка	Имя	Исходное значение	Результат		
8		$G$6	Прибыль	0	54		
9							
10							
11	Изменяемые ячейки						
12		Ячейка	Имя	Исходное значение	Результат		
13		$G$3	A Кол-во	0	12		
14		$G$4	B Кол-во	0	3		
15		$G$5	C Кол-во	0	0		
16							
17							
18	Ограничения						
19		Ячейка	Имя	Значение	Формула	Статус	Разница
20		$B$7	Время I	30	$B$7<=$B$9	не связан.	54
21		$C$7	Время II	39	$C$7<=$C$9	не связан.	3
22		$D$7	Время III	21	$D$7<=$D$9	связанное	0
23		$E$7	Время IV	33	$E$7<=$E$9	не связан.	9
24		$G$3	A Кол-во	12	$G$3=целое	связанное	0
25		$G$4	B Кол-во	3	$G$4=целое	связанное	0
26		$G$5	C Кол-во	0	$G$5=целое	связанное	0
27		$G$3	A Кол-во	12	$G$3>=0	не связан.	12
28		$G$4	B Кол-во	3	$G$4>=0	не связан.	3
29		$G$5	C Кол-во	0	$G$5>=0	связанное	0

Рис.3.22. Отчет по результатам

Утилита Поиск решения может использоваться и для решения более сложных задач оптимизации (см., раздел 1.2.1).

Рассмотрим пример расчета и оптимизации сетевого графика с помощью надстройки Поиск решения.

Пример 3.7. Имеется сетевая модель проведения капитального ремонта шлихтовальной машины, используемой в прядильном производстве. представленная таблицей 3.2, в которой учтены соотношения всех видов работ.

Таблица 3.2. Список работ проведения капитального ремонта

№ работы	Код работы	Наименование работы	Норматив времени t_{ij} на выполнение работы
1	0 – 1	Составление ведомости дефектов.	2
2	1 – 2	Ремонт сушильной камеры.	10
3	1 – 3	Выполнение слесарных работ.	6

Окончание таблицы 3.2.

4	1 – 4	Замена электродвигателя и ремонт привода.	30
5	1 – 6	Ремонт клеильного аппарата	60
6	2 - 5	Ремонт навивающего устройства	55
7	3 – 6	Согласование спецификации деталей с клеильным аппаратом	0
8	4 – 6	Согласование спецификации привода с клеильным аппаратом	0
9	5 – 6	Наладка мерильно-ленточного механизма	5
10	6 – 7	Испытание машины и принятие ее из ремонта	2

Графическое изображение сети приведено на рисунке 3.23, причем над дугами-работами записана продолжительность этих работ.

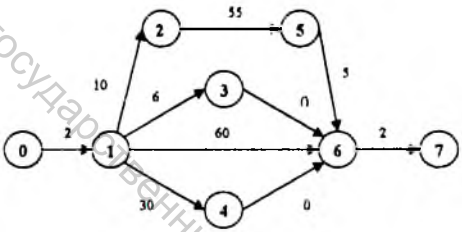


Рисунок 3.23. Сетевой график проведения капитального ремонта

Требуется рассчитать критический путь, затем оптимизировать сетевой график (рисунок 3.23) по ресурсам (например, времени) в соответствии с параметрами из таблицы 3.3.

Таблица 3.3. Параметры оптимизации сетевого графика

Параметры		Работы									
		0-1	1-2	1-3	1-4	1-6	2-5	3-6	4-6	5-6	6-7
Продолжительность работы	t_{ij}	2	10	6	30	60	55	0	0	5	2
Минимальное время	d_{ij}	1	8	5	20	40	40	0	0	4	1
Технологические коэффициенты	k_{ij}	0,05	0,1	0,1	0,3	0,4	0,3	0	0	0,2	0,05

Технологические коэффициенты k_{ij} рассчитываются как отношение продолжительности соответствующей работы к длительности критического пути. Конкретные значения k_{ij} принимаются не менее расчетных.

Для сокращения срока реализации проекта выделено 100 денежных единиц. Вложение дополнительных средств x_{ij} в работу (i,j) сокращает время ее выполнения до величины

$$t^{*}_{ij} = t_{ij} - k_{ij} \cdot x_{ij} \qquad (3.7)$$

где t_{ij} - продолжительность работы;

k_{ij} - технологические коэффициенты использования дополнительных средств;

x_{ij} – дополнительные средства, вложенные в работу.

Расчет критического пути осуществляется по следующим этапам:

1. Разместим таблицу исходных данных (рисунок 3.24) в ячейках A1:N14:

   а) в ячейки A4:A11 внесем номера вершин – от 0-ой до 7-ой;

   б) в ячейках B3:K3 обозначим ребра графика ;

   в) в ячейки B1:K1 внесем длины ребер, представляющие собой продолжительности работ ;

   г) в ячейках B4:K11 разместим матрицу состояния: все связи от исходного события к рассматриваемому (или входящие потоки) будем обозначать +1, обратные им связи – от рассматриваемого события к исходному обозначим -1.

2. В столбец L4:L11 внесем суммарные показатели потока. Поток 0→1, как начальный, входящий в работу 1, определим равным (+1). Поток 7→, как исходящий из последней работы, определим равным (-1). При расчете всех остальных потоков будем исходить из предпосылки, что количество входных и выходных потоков в любое событие, кроме первого и последнего, должно быть равным, а значит, сумма входных и выходных потоков в любое событие, кроме первого и последнего, должна равняться нулю.

3. В ячейках B14:K14 определим показатели критичности или некритичности пути (ребра) сетевого графика. При этом показателем критичности каждого пути (ребра) будет являться значение (1). Показателем некритичности пути (ребра) будет равняться значение (0). Поскольку критический путь заранее неизвестен, первоначально в ячейки B14:K14 можно занести нули или единицы.

4. Рассчитаем ограничения на показатели потока, перемножив соответствующую строку матрицы состояния на массив ребер. В ячейки M4:M11 внесем формулы:

M4: «=СУММПРОИЗВ(B4:K4;$B$14:$K$14)»
M5: «=СУММПРОИЗВ(B5:K5;$B$14:$K$14)»
.....
M11: «=СУММПРОИЗВ(B11:K11;$B$14:$K$14)».

5. Определим целевую функцию, представляющую собой длину критического пути, как сумму произведений продолжительности каждой работы (ячейки B1:K1) на значения критичности пути (ячейки B14:K14). В ячейку N14 внесем формулу «=СУММПРОИЗВ(B1:K1;B14:K14)».

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Длительность	2	10	6	30	60	55	5	0	0	2	Показа		
2	Вершины	Дуги										тели		
3		0-1	1-2	1-3	1-4	1-6	2-5	5-6	3-6	4-6	6-7	потока	Ограничения	
4	0	1										1	0	
5	1	-1	1	1	1	1						0	0	
6	2		-1				1					0	0	
7	3			-1					1			0	0	
8	4				-1					1		0	0	
9	5						-1	1				0	0	
10	6					-1		-1	-1	-1	1	0	0	
11	7										-1	-1	0	
12														
13		x01	x12	x13	x14	x16	x25	x56	x36	x46	x67			
14		0	0	0	0	0	0	0	0	0	0		Функция	0

Рисунок 3.24. Матрица нахождения критического пути сетевого графика

Теперь сформулируем постановку задачи в терминах рабочего листа Excel.

Добиться максимально возможного значения в ячейке N14, изменяя значения ячеек B14:K14 так, чтобы значения в ячейках L4:L11 равнялись значениям в ячейках M4:M11, а значения в ячейках B14:K14 были целыми и большими нуля. Реализуем решение средствами утилиты «Поиск решения».

Окно «Поиск решения» представлено на рисунке 3.25.

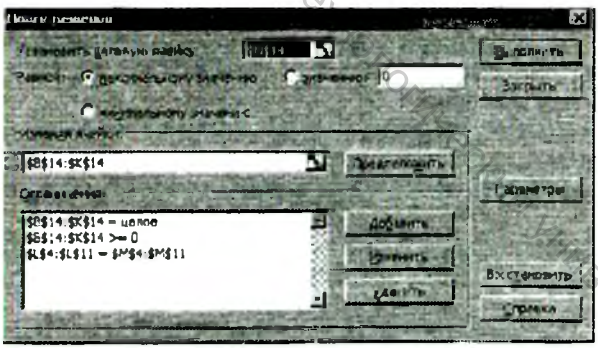


Рисунок 3.25. Окно «Поиск решения»

В результате будет найдена длина критического пути, равная 74 дням, а содержимое ячеек B14:K14 покажет, по каким ребрам этот путь пройдет (см. таблицу 3.4)

Таблица 3.4. Критический путь

x01	x12	x13	x14	x16	x25	x56	x36	x46	x67
1	1	0	0	0	1	1	0	0	1

Таким образом, критическим путем является путь 1-2-5-6-7, и его длительность составляет 74 дня.

После того, как критический путь рассчитан, оптимизируем сетевой график по времени.

Для формулировки экономико-математической модели оптимизации по времени сетевого графика, представленного на рисунке 3.23, добавим фиктивную работу (7-8), как показано на рисунке 3.26.

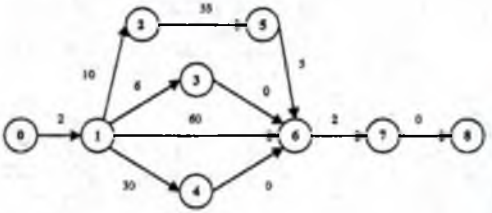


Рисунок 3.26. Сетевой график с фиктивной работой

Тогда целевая функция будет иметь вид

$$t_{кр} = t^{o}_{78} \rightarrow \min. \qquad (3.8)$$

Ограничения задачи могут быть представлены в следующем виде:

1. сумма вложенных во все работы средств не должна превышать 100 ден. ед.:

$$x_{01} + x_{12} + x_{13} + x_{14} + x_{16} + x_{25} + x_{56} + x_{67} \le 100; \qquad (3.9)$$

2. продолжительность выполнения каждой работы должна быть не менее минимально возможного времени d_{ij} (см. таблицу 3.3):

$t^{o}_{01} - t^{н}_{01} \ge 1;$	$t^{o}_{14} - t^{н}_{14} \ge 20;$	$t^{o}_{56} - t^{н}_{56} \ge 4;$
$t^{o}_{12} - t^{н}_{12} \ge 8;$	$t^{o}_{16} - t^{н}_{16} \ge 40;$	$t^{o}_{67} - t^{н}_{67} \ge 1;$
$t^{o}_{13} - t^{н}_{13} \ge 5;$	$t^{o}_{25} - t^{н}_{25} \ge 40;$	$t^{o}_{78} - t^{н}_{78} = 0;$

3. на основании формулы (3.7) продолжительность работы зависит от вложенных средств:

$t^{o}_{01} - t^{н}_{01} = 2 - 0.05 \cdot x_{01};$	$t^{o}_{14} - t^{н}_{14} = 30 - 0.3 \cdot x_{14};$	$t^{o}_{56} - t^{н}_{56} = 5 - 0.2 \cdot x_{56};$
$t^{o}_{12} - t^{н}_{12} = 10 - 0.1 \cdot x_{12};$	$t^{o}_{16} - t^{н}_{16} = 60 - 0.4 \cdot x_{16};$	$t^{o}_{67} - t^{н}_{67} = 2 - 0.05 \cdot x_{67};$
$t^{o}_{13} - t^{н}_{13} = 6 - 0.1 \cdot x_{13};$	$t^{o}_{25} - t^{н}_{25} = 55 - 0.3 \cdot x_{25};$	

4. время начала каждой работы должно быть не менее времени окончания непосредственно предшествующей ей работы:

$t^{н}_{12} \ge t^{o}_{01};$	$t^{н}_{25} \ge t^{o}_{12};$	$t^{н}_{67} \ge t^{o}_{16};$
$t^{н}_{13} \ge t^{o}_{01};$	$t^{н}_{36} \ge t^{o}_{13};$	$t^{н}_{67} \ge t^{o}_{36};$
$t^{н}_{14} \ge t^{o}_{01};$	$t^{н}_{46} \ge t^{o}_{14};$	$t^{н}_{67} \ge t^{o}_{46};$
$t^{н}_{16} \ge t^{o}_{01};$	$t^{н}_{56} \ge t^{o}_{25};$	$t^{н}_{67} \ge t^{o}_{56};$
		$t^{н}_{78} \ge t^{o}_{67};$

5. условие неотрицательности неизвестных: $t^н_{ij} \geq 0$; $t^о_{ij} \geq 0$; $x_{ij} \geq 0$;
6. условие целочисленности неизвестных: $t^н_{ij}$, $t^о_{ij}$, x_{ij} – целое.

Фрагмент табличной записи математической модели в Excel представлен на рисунке 3.27 (полная табличная запись математической модели представлена в Приложениях 4-6).

	A	B	C	D	E	F	G	H	I
1	№	x01	x12	x13	x14	x18	x25	x58	x87
2		X1	X2	X3	X4	X5	X6	X7	X8
3		0	0	0	0	0	0	0	0
4	ЦФ	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1
6	2								
7	3								
8	4								
9	5								
10	6								
11	7								
12	8								
13	9								
14	10								
15	11								
16	12								
17	13	0,05							
18	14		0,1						
19	15			0,1					
20	16				0,3				
21	17					0,4			
22	18						0,3		
23	19							0,2	
24	20								0,05

	AE	AF	AG
1	Знач	Вид	
2	огр	огр	
3			
4	0	min	
5	0	<=	100
6	0	>=	1
7	0	>=	8
8	0	>=	5
9	0	>=	20
10	0	>=	40
11	0	=	0
12	0	=	0
13	0	>=	40
14	0	>=	4
15	0	>=	1
16	0	=	0
17	0	=	2
18	0	=	10
19	0	=	6
20	0	=	30
21	0	=	60
22	0	=	55
23	0	=	5
24	0	=	2

Рисунок 3.27. Табличная запись математической модели в Excel

В ячейки B5:AD37 внесены исходные данные в соответствии с условием задачи:

а) показатели сбалансированности потоков;

б) технологические коэффициенты.

В ячейках B3:AD3 будут представлены рассчитанные значения неизвестных:

а) в ячейках B3:I3 – суммы средств, вкладываемых в каждую работу;

б) в ячейках I4: AD3 – время начала и окончания работ.

В ячейку AE4 внесена формула для вычисления целевой функции =СУММПРОИЗВ(B3:AD3,B4:AD4). Значение целевой функции равно времени окончания последней работы (7-8). Поэтому в ячейку AD4 внесено значение 1, в то время как ячейки B4:AC4 могут быть заполнены нулями или просто оставлены пустыми.

В столбце AG5:AG37 размещены заданные по условию значения ограничений, а в столбце AF5:AF37 показаны их знаки.

Соответственно, в столбец AE5:AE37 внесены формулы для расчета реальных значения ограничений:

в ячейку AE5: =СУММПРОИЗВ($B$3:$AD$3;B5:AD5);

в ячейку AE6: =СУММПРОИЗВ($B$3:$AD$3;B6:AD6);

.....

в ячейку AE37: =СУММПРОИЗВ($B$3:$AD$3;B37:AD37).

Теперь в терминах ячеек рабочего листа Excel задача может быть сформулирована следующим образом: добиться минимально возможного значения в ячейке AE3, изменяя значения ячеек B3:AD3 при условии выполнения ограничений в ячейках AE5:AG37, целочисленности и неотрицательности значений ячеек B3:AD3.

Окно Поиска решения с постановкой задачи представлено на рисунке 3.28. Результат оптимизации представлен на рисунке 3.29.

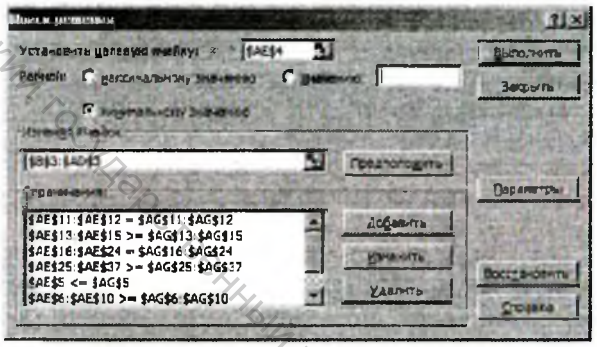


Рисунок 3.28. Окно Поиска решения

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	№	x01	x12	x13	x14	x16	x25	x56	x67	to01	tH12	to12	tH13
2		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
3		0	20	0	0	20	50	5	0	2	2	10	2
4	ЦФ	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1				

	AE	AF	AG
1	Знач	Вид	
2	ср	ор	
3			
4	66	min	
5	95	<=	100

Рисунок 3.29. Результат оптимизации

Таким образом, при дополнительном вложении 100 денежных единиц проект может быть выполнен за 56 единиц времени. При этом средства распределяются следующим образом: 20 денежных единиц в работу (1-2), 20 денежных единиц в работу (1-6), 50 денежных единиц – в работу (2-5) и 5 денежных единиц – в работу (5-6). Сокращение срока реализации проекта за счет вложения дополнительных средств составит 18 единиц времени. Рациональность рассмотренных методик заключается в том, что они позволяют найти критический и наикратчайший пути сетевого графика без перебора всех возможных вариантов. Последнее позволяет в короткие сроки осуществить решение двух основных задач сетевого планирования: задачу анализа оптимальности уже готового сетевого графика и задачу его оптимизации по длительности, в случае если сетевой график оказывается не оптимальным.

ГЛАВА 4. ИНТЕГРИРОВАННАЯ СИСТЕМА *STATISTICA*

Интегрированная система *STATISTICA* представляет собой программный продукт, предназначенный для выполнения статистического анализа результатов исследования. Несомненным преимуществом этого пакета является его полная интеграция в среду Windows, полноценный набор статистических методов, необходимых для экономиста, менеджера, социолога, наличие русскоязычной версии пакета и качественной литературы на русском языке.

4.1.Основные статистические модули пакета *STATISTICA*

Статистические процедуры системы ***STATISTICA*** сгруппированы в нескольких специализированных статистических модулях. В каждом модуле можно выполнить определенный способ обработки данных, не обращаясь к процедурам из других модулей.

Ниже приводится краткое описание основных статистических процедур, реализованных в конкретных статистических модулях.

Модуль «Основные статистики и таблицы» (Basic Statistics/Tables)

Обычно с модуля **Basic Statistics/Tables** (Основные статистики/Таблицы) начинается работа в системе. С его помощью можно провести предварительную обработку данных, осуществить разведочный анализ данных, определить зависимости между ними, разбить их различными способами на группы, просмотреть эти группы визуально и определить взаимосвязи между данными. Этот статистический модуль включает в себя приведенные ниже группы статистических процедур.

Описательные статистики, группировки, разведочный анализ

STATISTICA предлагает широкий выбор методов разведочного статистического анализа. Система может вычислить практически все описательные статистики, включая медиану, моду, квартили, средние и стандартные отклонения, доверительные интервалы для среднего, коэффициенты асимметрии, эксцесса (с их стандартными ошибками), гармоническое и геометрическое среднее, а также многие другие описательные статистики. Широкий выбор графиков помогает проведению разведочного анализа данных.

Корреляции. Этот подраздел включает большое количество средств, позволяющих исследовать зависимости между переменными. Возможно вычисление практически всех общих мер зависимости, включая коэффициент корреляции Пирсона, коэффициент ранговой корреляции Спирмена, коэффициент сопряженности признаков и многие другие. Корреляционные матрицы могут быть вычислены и для данных с пропусками, используя специальные методы обработки пропущенных значений.

***t*-критерии (и другие критерии для групповых различий).** В этом

подразделе представлены *t*-критерии для зависимых и независимых выборок, а также статистика Хоттелинга. Данные критерии могут быть вычислены и в модуле *ANOVA/MANOVA.*

Таблицы частот и таблицы кросс-табуляций. В данном подразделе содержится обширный набор процедур, обеспечивающих табулирование непрерывных, категориальных, дихотомических переменных, а также переменных, полученных в результате многовариативных опросов. Вычисляются как кумулятивные, так и относительные частоты. Доступны тесты для кросс-табулированных частот. Вычисляются статистики Пирсона, максимального правдоподобия, Йетс-коррекция, хи-квадрат, статистики Фишера и многие другие. Каскады сложных графиков для многократно классифицированных данных могут быть просмотрены интерактивно.

Модуль «Множественная регрессия» (Multiple Regression)

Основное назначение данного модуля — построение зависимостей между многомерными переменными, подбор простой линейной модели и оценка ее адекватности.

Модуль «Множественная регрессия» включает в себя исчерпывающий набор средств множественной линейной и фиксированной нелинейной (в частности, полиномиальной, экспоненциальной, логарифмической и др.) регрессии, включая пошаговые, иерархические и другие методы, а также ридж-регрессию.

Модуль Nonlinear Estimation

Предоставляет возможность пользователю создавать собственные функции для аппроксимации исследуемых данных.

Модуль «Дисперсионный анализ» (ANOVA/MANOVA–модуль)

Этот модуль позволяет дать оценку степени воздействия известных факторов на измеряемые данные. *ANOVA/MANOVA*–модуль представляет собой набор процедур общего одномерного и многомерного дисперсионного и ковариационного анализа.

В модуле доступны решения многих задач в наиболее прямых "функциональных" терминах. Даже пользователи, имеющие малый опыт работы с *ANOVA*, могут анализировать достаточно сложные проекты с помощью системы *STATISTICA*.

Модуль «Дискриминантный анализ» (Discriminant Analysis)

Этот модуль предназначен для решения задач, в которых по результатам измерений необходимо отнести объект к одному из нескольких классов, например, в медицине при обследовании больных, или в геологии при оценке перспективности месторождений, или в банковской деятельности .

Модуль «Непараметрическая статистика и подгонка распределений» (Nonparamrtrics)

Данный модуль предназначен для проверки различных гипотез о характере распределения имеющихся данных. Модуль содержит обширный набор непараметрических критериев согласия, в частности критерий Колмогорова–Смирнова, ранговые критерии Манна–Уитни, Вальда–Вольфовица, Вилкоксона и многие другие.

Модуль «Факторный анализ» (Factor Analysis)

Этот модуль помогает выделить основные общие факторы качественного характера, влияющие на наблюдаемые характеристики сложного объекта и связи между ними, например основные социально-экономические факторы или факторы, влияющие на результаты голосования. Модуль содержит широкий набор методов и опций, снабжающих пользователя исчерпывающими средствами факторного анализа.

Модуль «Многомерное шкалирование» (Multidimensional Scaling)

Модуль содержит инструментарий для выполнения многомерного шкалирования. Матрицы подобия, различия и корреляции могут быть вычислены для большого числа переменных (до 90 переменных) с размерностью до 9 компонент. Начальная конфигурация может быть вычислена автоматически с помощью анализа главных компонент либо задана пользователем. Доступны всевозможные метрики.

Модуль «Кластерный анализ» (Claster Analysis)

Предназначение данного модуля — произвести сложную иерархическую классификацию данных или выделить в них кластеры.

Модуль «Нелинейное оценивание» (Advanced Nonlinear Models)

Данный модуль предназначен для определения нелинейных зависимостей в данных, подгонки к ним функциональных кривых. Модуль предоставляет возможность осуществить подгонку к наблюдаемым данным кривой, по существу, любого типа.

Модуль «Каноническая корреляция» (Canonical Analysis)

Модуль включает в себя широкий набор процедур для выполнения канонического корреляционного анализа, исследования связи между двумя множествами переменных. Модуль может обрабатывать векторные данные или корреляционные матрицы и вычислять все стандартные канонические корреляционные статистики.

Модуль «Анализ временных рядов и прогнозирование» (Time Series /Forecasting)

Модуль предлагает широкий набор методов анализа. и состоит из нескольких общих процедур, интегрированных вокруг динамического графического представления временных рядов и их сглаживающих/моделирующих преобразований. Возможно моделирование одновременно нескольких рядов и выполнение интерактивного "что-если" анализа, наблюдая ряд на графике. Методы преобразования рядов включают следующие процедуры: исключение среднего, тренда, взвешенное скользящее среднее, медианное сглаживание, фильтрацию, взятие разностей с любым сдвигом и многое другое.

Модуль «Моделирование структурными уравнениями (SEPATH)»

Данный модуль помогает построить и тестировать различные модели, объясняющие структуру связей между наблюдаемыми переменными. Моделирование структурными уравнениями — мощное средство многомерного статистического анализа, развитое в последние годы и имеющее целью соединить статистические методы с методами теории систем.

4.2.Корреляционный анализ экономической информации

Совокупность методов оценки корреляционных характеристик и проверка статистических гипотез о них по выборочным данным называется корреляционным анализом. В корреляционном анализе используются следующие основные приемы:

1. построение корреляционного поля (диаграммы рассеяния) для двух экономических показателей или двумерных сечений, если речь идет о большем их количестве;
2. определение выборочных коэффициентов корреляции или составление корреляционных матриц;
3. проверка статических гипотез о значимости связи между показателями.

Коэффициент корреляции оценивает тесноту связи между исследуемыми переменными и является мерой линейной зависимости величин.

Если оценивается связь между двумя любыми количественными переменными, используется *парный коэффициент корреляции*. Для оценки тесноты связи между результирующим показателем и совокупностью входных факторов используется *множественный коэффициент корреляции*. Для оценки тесноты связи между качественными (порядковыми) переменными используется *ранговый коэффициент корреляции*.

Этот коэффициент, всегда обозначаемый латинской буквой **r**, может принимать значения между -1 и +1, причём если значение находится ближе к 1, то это означает наличие сильной связи, а если ближе к 0, то слабой.

Если коэффициент корреляции отрицательный, это означает наличие

противоположной связи: чем выше значение одной переменной, тем ниже значение другой. Сила связи характеризуется также и абсолютной величиной коэффициента корреляции. Для словесного описания величины коэффициента корреляции используются следующие градации:

Значение	Интерпретация
до 0,2	Очень слабая корреляция
до 0,5	Слабая корреляция
до 0,7	Средняя корреляция
до 0,9	Высокая корреляция
свыше 0,9	Очень высокая корреляция

Ориентировочно определить величину коэффициента корреляции можно, анализируя диаграмму рассеяния. Чем теснее расположены точки относительно некоторой прямой, тем больше по абсолютной величине он стремится к единице, и наоборот, чем более расплывчата диаграмма рассеяния, тем ближе к нулю коэффициент корреляции.

Пример 4.1. Проанализировать показатели хозяйственной деятельности предприятий легкой промышленности [2]. Построить корреляционную матрицу, дать графическую интерпретацию результатов. В качестве исходных данных использовать данные из таблицы П2 Приложения 2 (файл *analiz.sta)*:

Y1 – Производительность труда, млн. руб.
X1 – Трудоёмкость единицы продукции, час.
X2 – Удельный вес рабочих в составе ППП.
X3 – Удельный вес комплектующих изделий.
X4 – Коэффициент сменности оборудования.
X5 – Премии и вознаграждения на одного рабочего, млн. руб.
X6 – Удельный вес потерь от брака в себестоимости продукции, %.
X7 – Фондоотдача, руб.
X8 – Среднесписочная численность ППП.
X9 – Среднегодовая стоимость ОПС, млрд. руб.
X10 – Фонд з/п ППП, млн. руб.
X11 – Фондовооруженность труда, млн. руб.
X12 – Оборачиваемость нормируемых оборотных средств.
X13 – Оборачиваемость ненормируемых оборотных средств.
X14 – Непроизводительные расходы.

Проведем корреляционный анализ переменных X1, X3, X7.

Для проведения корреляционного анализа необходимо выбрать модуль **Basic Statiatics/Tables** и далее пункт **Correlation matrices** (Корреляционные матрицы).

В появившемся окне **Pearson Product-Moment Correlation (Корреляция Пирсона)** (рис.4.1), нажав кнопку **Two lists (Два списка),** следует определить,

какие переменные будут находиться в строках (**First variable list**) –

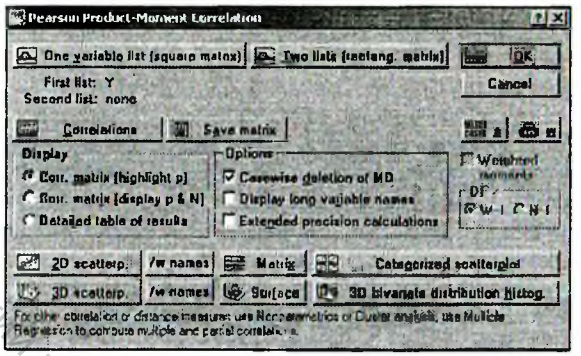


Рис. 4.1. Окно Pearson Product-Moment Correlation

это переменные X1, X3, X7 и в столбцах (**Second variable list**) – это переменная Y (рис.4.2). Для подтверждения выбора и возврата в окно **Pearson Product-Moment Correlation** нажать OK.

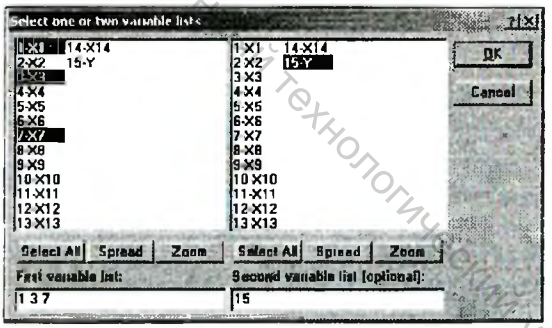


Рис.4.2. Окно Select one or two variable list

В окне **Pearson Product-Moment Correlation** также следует нажать кнопку OK. На экране появится корреляционная матрица (рис.4.3).

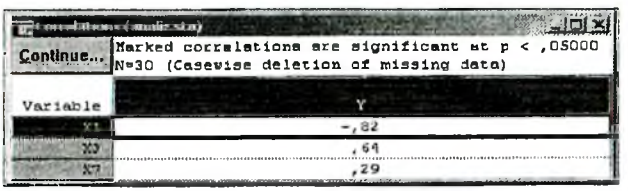


Рис.4.3. Корреляционная матрица

В этой матрице имеется только один столбец, так как во втором списке мы выбрали только одну переменную. В столбце даны коэффициенты корреляции между переменными Y и X1, X3, X7. Красным цветом автоматически выделены коэффициенты, значимые на уровне p<0,05. Именно на эти коэффициенты следует обратить наибольшее внимание. В рассматриваемом примере переменная Y наиболее зависима от переменных X1 (коэффициент корреляции -0,82) и X3 (коэффициент корреляции 0,64).

Чтобы построить корреляционную матрицу, отражающую тесноту связи между всеми переменными, следует в окне **Pearson Product-Moment Correlation** нажать кнопку **One variable list** и выбрать переменные Y, X1, X3 и X7. Затем вернуться в окно **Pearson Product-Moment Correlation** и нажать кнопку **ОК.** В результате на экране появится корреляционная матрица, отражающая все парные коэффициенты корреляции для рассматриваемых переменных (см. рис.4.4). Красным цветом в ней автоматически выделены коэффициенты, значимые на уровне p<0,05.

Correlations (analiz.sta)

Continue...

Marked correlations are significant at p < ,05000
N=30 (Casewise deletion of missing data)

Variable	X1	X3	X7	Y
X1	1,00	-,65	-,23	-,82
X3	-,65	1,00	,60	,64
X7	-,23	,60	1,00	,29
Y	-,82	,64	,29	1,00

Рис.4.4. Корреляционная матрица

Для построения графического отображения корреляционной взаимосвязи любых двух переменных необходимо нажать кнопку **2D scatterplot** (диаграмма рассеяния) в окне **Pearson Product-Moment Correlation** и определить, какие переменные будут расположены по горизонтальной и вертикальной оси. Построим диаграмму рассеяния для переменных Y и X1 (см. рис. 4.5).

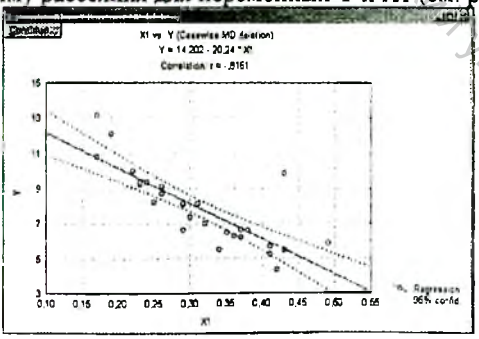


Рис. 4.5. Диаграмма рассеяния для переменных Y и X1

Для сравнения аналогичным образом построим диаграмму рассеяния для переменных Y и X7 (см. рис. 4.6).

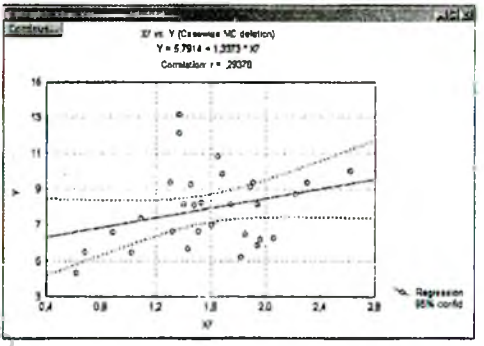


Рис. 4.6. Диаграмма рассеяния для переменных Y и X7

В этом случае из графика видно, что зависимость нельзя признать линейной. Можно попробовать графически подобрать вид зависимости. Для этого в меню **Graphs** следует выбрать команду **Stats 2D Graph**, затем команду **Scatterplots**, после чего в окне **Scatterplots** (см. рис. 4.7) кнопкой **Variables** определить исследуемые переменные и выбрать вид зависимости, например полиномиальную.

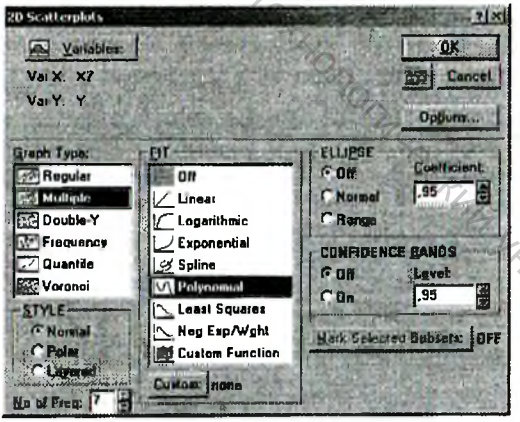


Рис. 4.7. Окно Scatterplots

В результате на экране появится диаграмма рассеяния (см. рис. 4.8), построенная на базе полинома пятой степени. Интересным является факт, что степень полинома (максимальной является пятая степень) **STATISTICA** определяет сама в зависимости от исходного набора данных.

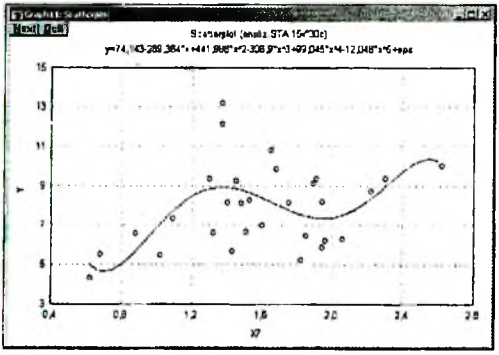


Рис. 4.8. Диаграмма рассеяния для переменных Y и X7 при аппроксимации полиномом

Как и на рисунке 4.5, на этом графике (рис. 4.8) представлены функция, отражающая взаимосвязь переменных Y и X7, и корреляционное поле точек. Из графика отчетливо видно, что и полином пятой степени не описывает должным образом исходные данные. В этом случае следует воспользоваться упомянутым выше модулем **Nonlinear Estimation,** предоставляющим возможность пользователю создавать собственные функции для аппроксимации исследуемых данных.

Чтобы оценить все связи визуально, можно представить корреляционную матрицу в графическом виде (см. рис. 4.9). Для этого в окне **Pearson Product-Moment Correlation** следует нажать кнопку Matrix, затем выбрать все переменные нажатием кнопки Select All и подтвердить выбор нажатием OK.

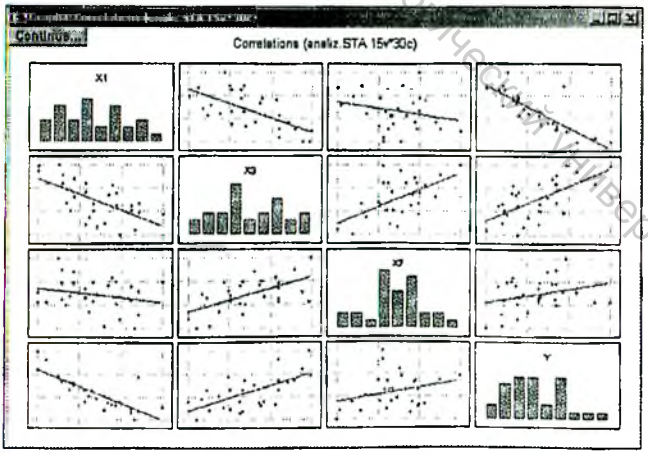


Рис. 4.9. Корреляционная матрица в графическом представлении

Для получения числовых характеристик (описательной статистики) изучаемых признаков в стартовой панели модуля **Basic Statistics** необходимо выбрать раздел **Decriptive Statistics** и нажать OK. Кнопкой **Variables** выбрать переменные, для которых требуется получить числовые характеристики (рис. 4.10), затем, нажав кнопку **More statistics**, определить, какие характеристики должны быть рассчитаны, и нажать OK.

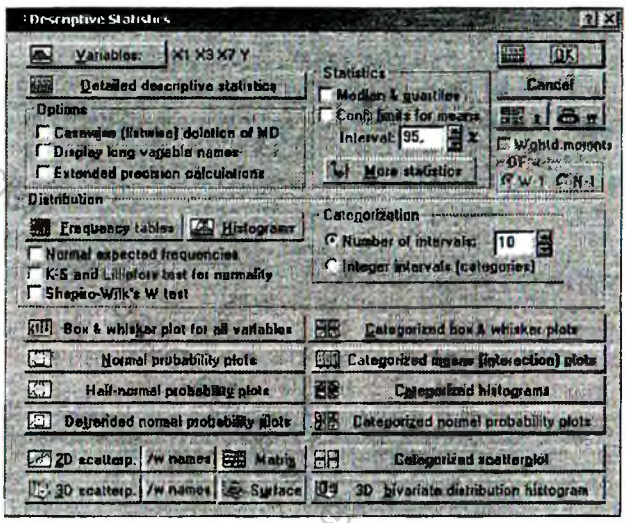


Рис. 4.10. Окно Decriptive Satistics

Выберем числовые характеристики, характеризующие распределение переменных: среднее значение (Mean), медиану (Median), дисперсию (Variance), коэффициенты асимметрии (Skewness) и эксцесса (Kurtosis). Далее следует нажать кнопку **Detailed decriptive statistics** и получить результат (см. рис. 4.11).

Descriptive Statistics (analiz.sta)

Continue...	Valid N	Mean	Median	Minimum	Maximum	Variance	Skewness	Kurtosis
X1	30	,310667	,305000	,170000	,49000	,007110	,177709	-,755800
X3	30	,282667	,255000	,020000	,54000	,020579	,071751	-,861896
X7	30	1,568000	1,565000	,620000	2,62000	,211037	-,130824	,159484
Y	30	7,915000	8,125000	4,320000	13,17000	4,372612	,571071	,108261

Рис.4.11. Описательная статистика переменных X1, X3, X7 и Y

Как видно из таблицы на рис. 4.11, распределение переменных близко к нормальному.

4.3. Регрессионный анализ экономической информации

При исследовании экономических процессов часто возникает необходимость установить связь и определить количественную зависимость между анализируемыми факторами. При этом одна из величин выделяется как зависимая Y, остальные – как независимые ($x_1, x_2, \ldots x_n$).

Если расчёт корреляции характеризует силу связи между двумя переменными, то регрессионный анализ служит для определения вида этой связи и дает возможность для прогнозирования значения одной (зависимой) переменной, отталкиваясь от значения другой (независимой) переменной.

Методы регрессионного анализа, позволяющие моделировать статистические зависимости между двумя или несколькими переменными, представлены в *STATISTICA* модулем **Multiple Regression** (Множественная регрессия). В нем реализованы различные методы множественной линейной, пошаговой и фиксированной нелинейной регрессии (в частности, полиномиальной, экспоненциальной, логарифмической). *STATISTICA* позволяет вычислить все необходимые статистики и оценить адекватность построенной модели. Анализ остатков и выбросов может быть проведен при помощи широкого набора графиков, включая разнообразные точечные графики, графики частичных корреляций и многие другие. Система прогноза позволяет пользователю выполнять анализ "что - если". Методы общей нелинейной регрессии реализованы в модуле **Nonlinear Estimation** (Нелинейное оценивание). Он позволяет строить произвольную регрессионную модель, задаваемую некоторой алгебраической формулой, которая может быть нелинейной как по переменным, так и по параметрам.

При работе с моделями множественной регрессии необходимо провести предварительный анализ целесообразности включения выбранных переменных в регрессионную модель.

Установка флажка в поле **Review descr. stats, corr. Matrix** *(Обзор описательных статистик, корреляционная матрица)* позволит провести предварительный анализ исходных переменных и построить корреляционную матрицу, анализ которой дает возможность сделать важные выводы о структуре связей между выбранными переменными.

Не рекомендуется включать в модель переменные, слабо связанные с результативным признаком, а также переменные, тесно связанные друг с другом. Когда между объясняющими переменными существует ощутимая линейная зависимость, говорят о существовании мультиколлинеарности. В этом случае решение становится неустойчивым, незначительное изменение состава выборки (значений признаков) или состава объясняемых переменных может вызвать кардинальное изменение модели, что делает ее использование малопригодным в практических целях. Наиболее распространенные в таких случаях приемы: исключение одной из двух сильно связанных переменных, использование гребневой регрессии, переход от первоначальных переменных к их главным компонентам.

Если сбросить флажок в поле **Perform default analysis** (Метод анализа по умолчанию), то появляется доступ к диалоговому окну **Model Definition**, открывающему возможность дополнительного выбора методов анализа, среди которых имеются методы пошаговой (**Stepwise**) и гребневой (**Ridge**) регрессии.

Методы пошаговой регрессии позволяют из множества независимых переменных отобрать только те, которые наиболее значимы для адекватного описания многопараметрической регрессии. В модуле реализованы две процедуры отбора переменных, каждая из которых может давать различный конечный набор переменных: последовательное включение (**Forward stepwise**) и последовательное исключение (**Backward stepwise**).

Гребневая регрессия используется для получения более устойчивых оценок параметров регрессионной модели в условиях мультиколлинеарности переменных.

При изучении линейного регрессионного анализа проведем различие между простым анализом (одна независимая переменная) и множественным анализом (несколько независимых переменных). Собственно говоря, никаких принципиальных отличий между этими видами регрессии нет, однако простая линейная регрессия является простейшей и применяется чаще всех остальных видов.

4.3.1. Простая линейная регрессия

В практической работе наибольшее распространение получили модели линейной зависимости. Этот вид регрессии лучше всего подходит для того, чтобы продемонстрировать основополагающие принципы регрессионного анализа.

Линейная *однофакторная* модель (4.1) – это уравнение прямой на плоскости:

$$Y = m \cdot x + b, \qquad (4.1)$$

где `m` – коэффициент уравнения регрессии, представляющий собой тангенс угла наклона прямой `Y = f(x)` к оси OX;

`b` – свободный член, равный значению точки пересечения прямой `Y=f(x)` с осью OY.

Пример 4.2. В таблице П1 Приложения 1 (файл данных *zarplata.sta*) приведены показатели уровня жизни по территориям регионов республики Беларусь за 200Хг. Провести анализ зависимости среднедневной заработной платы, руб. (Y) от среднедушевого прожиточного минимума в день одного трудоспособного, руб. (x).

Последовательность проведения регрессионного анализа:

1) открыть модуль **Multiple regression** (Множественная регрессия);
2) создать или открыть файл данных (*zarplata.sta)*;

3)идентифицировать переменные – выбрать список зависимых и независимых переменных;

4) выбрать вид модели;

4) провести оценивание параметров модели;

5) проверить качество полученных оценок параметров;

6) провести анализ адекватности модели.

В пакете *STATISTICA* откройте модуль **Multiple regression** (Множественная регрессия). В стартовой панели модуля нажмите кнопку **Open Data** (Открыть данные) и откройте файл данных *zarplata.sta*, в котором находятся исходные данные, либо выполните команду **File/New Data** и введите исходные данные для переменных **X, Y** в столбцы **Var1** и **Var2.** При этом лишние столбцы **Var3-Var10** можно удалить командой **Variables→ Delete** меню **Edit**, строки – добавить командой **Cases→ Add** меню **Edit** (см. рис.4.12).

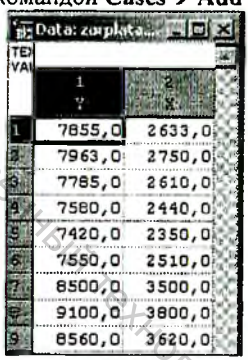


Рис.4.12. Исходные данные для построения модели

Сделайте активным окно с таблицей данных и в меню **Analisis** выберите команду **Resume Analisis**. На экране появится окно **Multiple regression.** Далее с помощью кнопки **Variables** (Переменные) перейдите в окно **Select dependent and independent variable list** (Выбрать списки зависимых и независимых переменных) и выберите переменные для анализа. Зависимую переменную *Y* – среднедневная заработная плата, руб. – внесите в строку **Dependent variable list** (Список зависимых переменных), независимую переменную *X* – среднедушевой прожиточный минимум в день одного трудоспособного, руб. – внесите в строку **Independent variable list** (Список независимых переменных) и нажмите кнопку **ОК.** Вы вновь окажетесь в стартовой панели модуля (см.рис. 4.13).

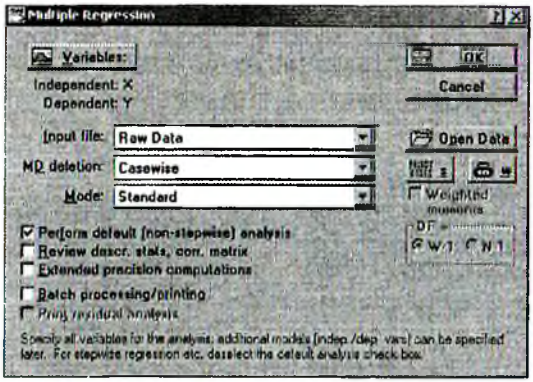


Рис.4.13. Окно Multiple regression

Переменные для анализа выбраны. Никаких дополнительных установок в стартовой панели в данном случае производить не нужно. Нажмите кнопку **OK.** На экране перед вами появится диалоговое окно **Multiple Regression Results** (Результаты множественной регрессии).

В диалоговом окне **Multiple Regression Results** (*Результаты множественной регрессии)* можно просмотреть результаты оценивания, которые представлены в численном и графическом виде(рис. 4.14).

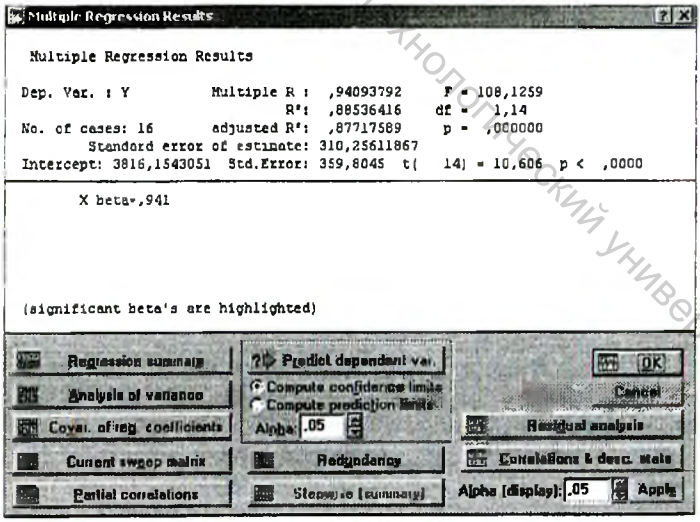


Рис. 4.14 – Окно Результаты множественной регрессии

Окно результатов анализа имеет следующую структуру: верх окна —

информационный. Он состоит из двух частей: в первой части содержится основная информация о результатах оценивания, во второй высвечиваются значимые регрессионные коэффициенты. Внизу окна **Результаты множественной регрессии** находятся функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа.

Рассмотрим вначале информационную часть окна. В ней содержатся краткие сведения о результатах анализа, а именно:

• **Dep. Var.** *(Имя зависимой переменной)* - в данном случае Y;

• **No. of Cases** *(Число наблюдений, по которым построена регрессия)* - в примере это число равно 16;

• **Multiple R** *(Коэффициент множественной корреляции)*;

• **R-square** (R^2 - *коэффициент детерминации, квадрат коэффициента множественной корреляции*) используется для статистической оценки тесноты связи между результативным и объясняющими показателями. Он выражает долю объясненной изучаемыми факторами дисперсии результативного признака и служит важной характеристикой качества построенной модели. Этот коэффициент может принимать значения от 0 до 1. Несмещенной оценкой $\mathbf{R^2}$ служит скорректированный на потерю степеней свободы коэффициент множественной детерминации (**Adjusted** $\mathbf{R^2}$);

• **Adjusted R-square** *(Скорректированный коэффициент детерминации),* определяемый как

$$Adjusted\ R\text{-}square = 1 - (1 - R^2) \cdot (n/(n-p)),$$

где n — число наблюдений в модели, p — число параметров модели (число независимых переменных плюс 1, так как в модель включен свободный член);

• **Std. Error of estimate** *(Стандартная ошибка оценки)* - эта статистика является мерой рассеяния наблюдаемых значений относительно регрессионной прямой;

• **Intercept** *(Оценка свободного члена регрессии)* - значение коэффициента B_0 в уравнении регрессии;

• **Std. Error** *(Стандартная ошибка оценки свободного члена)* - стандартная ошибка коэффициента B_0 в уравнении регрессии;

• **t(df) and p-value** *(Значение t-критерия и уровень p)* - *t*-критерий используется для проверки гипотезы о равенстве 0 свободного члена регрессии (см. прил. 2);

• **F**– критерий Фишера, определяющий значимость полученной модели. Оценивает вероятность случайного отклонения от нуля коэффициента детерминации при отсутствии связи между элементами совокупности. Желательно, чтобы полученный минимальный уровень значимости F-критерия **(p-level)** был меньше 0,05 (см. прил. 2);

• **df** — число степеней свободы F-критерия;

• **p** — уровень значимости.

В информационной части посмотрим прежде всего на значения коэффициента детерминации, которые лежат в пределах от 0 до 1. В нашем

примере R^2 = 0,885. Это значение показывает, что построенная регрессия объясняет более 88,5% разброса значений переменной Y относительно среднего.

Далее посмотрите на значение *F*-критерия и уровень значимости *p*. *F*-критерий используется для проверки гипотезы о значимости регрессии. В данном случае для проверки гипотезы, утверждающей, что между зависимой переменной Y и независимой переменной X нет линейной зависимости, т. е. $B_1 = 0$, против альтернативы B_1 не равен 0. В данном примере мы имеем большое значение *F*-критерия — 108,12, а представленный в окне уровень значимости $p = 0{,}00$ показывает, что построенная регрессия высоко значима.

Рассмотрим вторую часть информационного окна. В ней представлена информация о значимых и незначимых оценках регрессионных коэффициентов. При этом высвечивается строка: *x beta* = 0,941 и приводится пояснение **Significant beta's are highlighted** *(Значимые beta высвечены)*. Отметим, что в данном случае *beta* есть стандартизованный коэффициент B_1, т. е. коэффициент при независимой переменной x.

Перейдем в функциональную часть окна результатов.

Прежде всего нажмите кнопку **Regression summary** *(Итоговый результат регрессии)*. На экране появится **Spreadsheet** *(Электронная таблица вывода)*, в которой представлены итоговые результаты оценивания регрессионной модели (рис. 4.16).

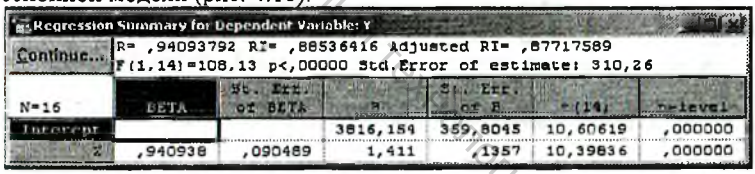


Рис. 4.16. Итоговая таблица регрессии

В первом столбце таблицы даны значения коэффициентов *beta* (стандартизованные коэффициенты регрессионного уравнения), во втором — стандартные ошибки этих коэффициентов, в третьем — точечные оценки параметров модели:

- свободный член B_0= 3816,154;
- коэффициент B_1 (при независимой переменной X) = 1,411.

Далее представлены стандартные ошибки для B_0, B_1, значения статистик t-критерия и т.д. По итоговой таблице регрессии можно построить модель, которая имеет вид

$$Y = 1{,}411 \cdot X + 3816.154.$$

Оценка адекватности модели - важный элемент анализа. После того как доказана адекватность модели, полученные результаты можно уверенно использовать для дальнейших действий.

Анализ адекватности основывается на анализе остатков. Остатки

представляют собой разности между наблюдаемыми значениями и модельными, т.е. значениями, подсчитанными по модели с оцененными параметрами. Графики зависимости регрессионных остатков от экспериментальных значений исходных переменных позволяют проверить предположения об однородности и независимости ошибок, являющихся предпосылками применения метода наименьших квадратов, и локализовать выбросы. Если указанные допущения выполняются, графики будут представлять собой симметричное, случайное и равномерное распределения точек. Графики эмпирической функции распределения остатков на нормальной вероятностной бумаге (**Probability plots**) и гистограммы позволяют проверить справедливость предположения о нормальном распределении остатков.

Кроме этого, имеется возможность вычислить статистику Дарбина–Уотсона (**Darbin-Watson Stat**) для проверки наличия автокорреляции в остатках, вывести на экран (**Display residuals&pred**) и сохранить в файле (**Save residuals**) информацию о наблюдаемых и подобранных по модели значениях результативного показателя и остатках. Рекомендуется также построить график линейной зависимости предсказанных (подобранных по модели) значений зависимой переменной от наблюдаемых (**Predicted & Observed**), что позволяет наглядно изобразить результаты регрессионного анализа.

В модуле **Множественная регрессия** имеется специальное диалоговое окно, в котором проводится всесторонний анализ остатков. Нажав кнопку **Residual Analysis** (Анализ остатков) в окне **Multiple Regression Results,** можно перейти в окно анализа остатков **Residual Analysis** (Анализ остатков) (см. рис. 4.17).

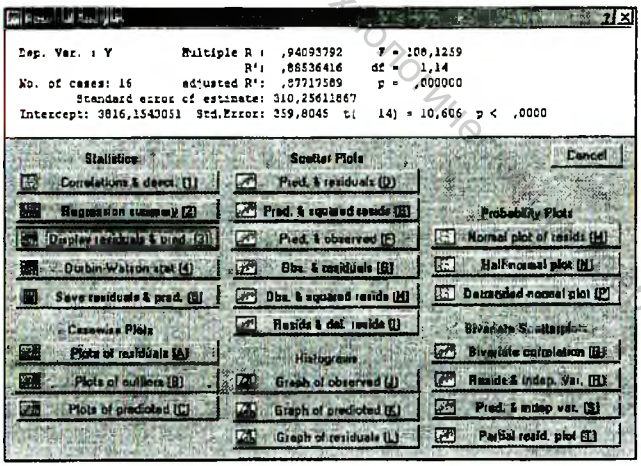


Рис. 4.17. Окно Residual Analysis

Нажмите в этом окне, например, кнопку **Obs&residuals**, на экране появится график (рис.4.18), который говорит о достаточной адекватности модели.

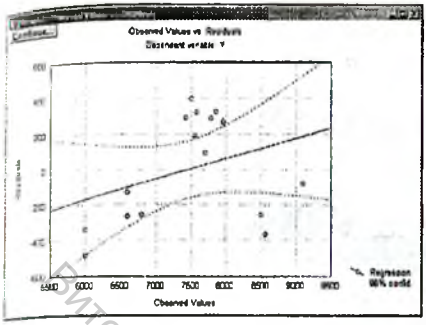


Рис.4.18. График наблюдаемых переменных-остатков

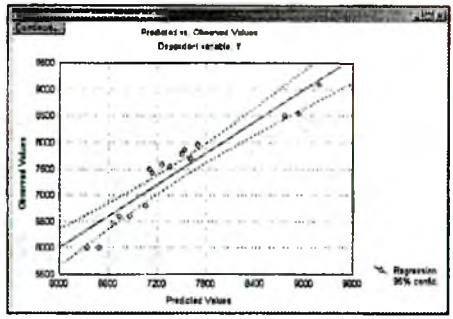


Рис. 4.19. График наблюдаемых и предсказанных значений

Нажав кнопку Predicted & Observed), можно наглядно изобразить результаты регрессионного анализа: график линейной зависимости предсказанных (подобранных по модели) значений зависимой переменной от наблюдаемых (см. рис. 4.19).

Часто, если остатки не распределены по нормальному закону, а также для стабилизации дисперсии данных, применяют преобразования зависимых и независимых переменных, например, логарифмическое преобразование зависимых переменных или извлечение квадратного корня.

Для получения описательной статистики в окне **Multiple Regression Results** нужно нажать кнопку **Correlations & desc. stats,** после чего на экране появится окно **Review Descriptive Statistics,** из которого следует выбрать необходимые для анализа статистики (см. рис. 4.20).

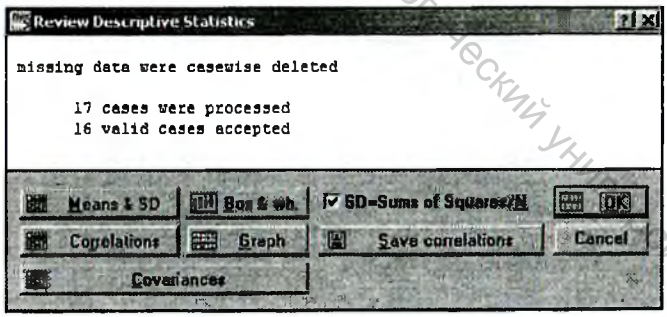


Рис.4.20. Окно Review Descriptive Statistics

Чтобы получить прогноз значения зависимой переменной Y, в окне **Multiple Regression Results** следует нажать кнопку **Predict dependent var** и в появившееся на экране окно **Specify values for indep. vars** ввести новое значение *X*, например 5000 (см. рис. 4.21), и нажать OK. В результате в окне

Predicting values for (см. рис.4.22) на основании полученного ранее уравнения регрессии *Y= 1,411·X +3816.154* будет рассчитано прогнозируемое значение Y (в данном случае 10871,28).

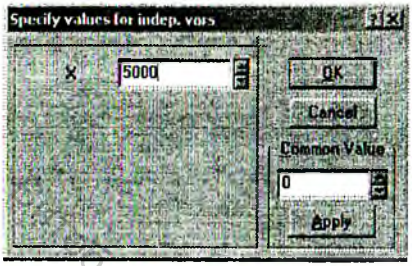


Рис.4.21. Окно Specify values for indep. vars

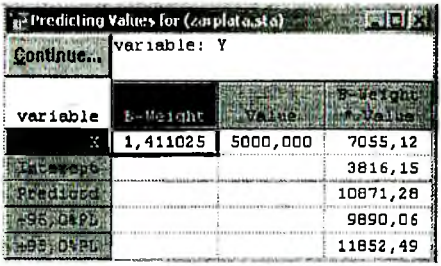


Рис.4.22. Окно Predicting values for

Кроме аналитического расчета регрессионной модели, ***STATISTICA*** позволяет построить модель графическим способом.

Графическое построение и представление модели. Для расчета модели графическим способом необходимо в стартовой панели программы *STATISTICA* выбрать модуль **«Основные статистики» («Basic Statistica»).** В меню **Graphs → Stats 2D Graphs** выбрать команду **Scatterplots (диаграммы рассеяния)**. На экране появится диалоговое окно для построения **2D** диаграмм рассеяния разного вида (линейного, логарифмического, экспоненциального, полиномиального и т.д.), отражающих зависимость переменных X и Y (см. рис.4.23).

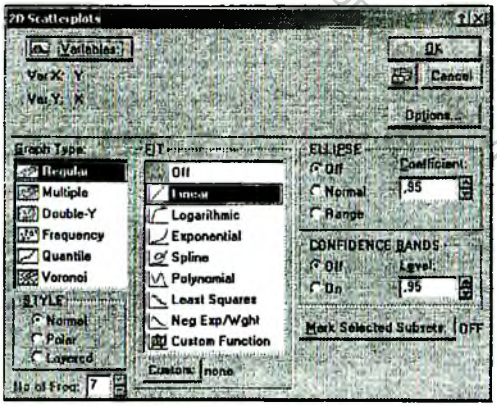


Рис.4.23. Окно 2D Scatterplots.

Если выбрать **Линейную (Linear)** зависимость и нажать OK, на экране

появится окно диаграммы рассеяния (рис.4.24) для выбранных переменных, причем в верхней части окна будет представлено уравнение регрессии.

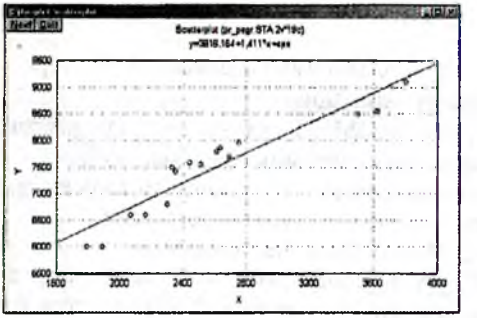


Рис.4.24. Диаграмма рассеяния. Линейная зависимость.

Из окна **2D Scatterplots** щелчком правой клавиши мыши по линии тренда можно легко построить диаграмму рассеяния любого из перечисленных видов, причем в верхней части окна будет выводиться и соответствующее уравнение регрессии (рис. 4.25 и рис.4.26). Если выбрать полиномиальную зависимость, то очевидно, что, чем выше степень полинома, тем точнее линия тренда «ложится» на данные.

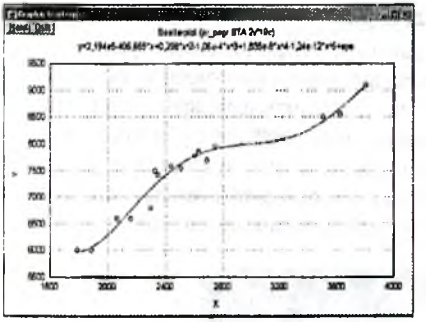


Рис.4.25. Диаграмма рассеяния Полиномиальная зависимость.

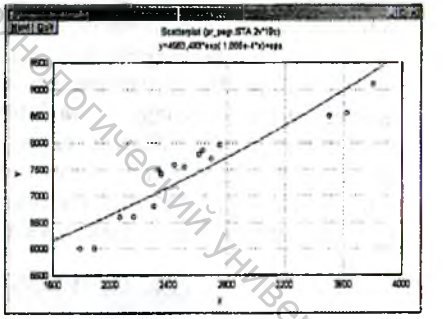


Рис.4.26. Диаграмма рассеяния. Экспоненциальная зависимость.

4.3.2. Множественная линейная регрессия

В общем случае в регрессионный анализ вовлекаются несколько независимых переменных. Это, конечно же, наносит ущерб наглядности получаемых результатов, так как подобные множественные связи в конце концов становится невозможно представить графически.

Линейная *многофакторная* модель (4.2) представляет собой уравнение прямой в многомерном пространстве и имеет вид

$$Y = b + m_1x_1 + m_2x_2 + ... + m_nx_n, \quad (4.2)$$

где x_1, ... , x_n – независимые переменные (факторы);

Y – зависимая переменная;

m_0, ... , m_n – коэффициенты уравнения регрессии;

n — количество независимых переменных.

В случае множественного регрессионного анализа необходимо оценить коэффициенты уравнения множественной регрессии.

Переменные, объявленные независимыми, могут сами коррелировать между собой; этот факт, называемый мультиколлинеарностью, необходимо обязательно учитывать при определении коэффициентов уравнения регрессии для того, чтобы избежать ложных корреляций.

Пример 4.3. Используя данные из таблицы П2 приложения 2 (файл *analiz.sta*), построить линейную многофакторную регрессионную модель и провести анализ зависимости производительности труда *(Y)* от трудоемкости единицы продукции *(X1)*, удельного веса комплектующих изделий *(X3)* и фондоотдачи (*X7)*.

Основные действия те же, что и в предыдущем примере. В данном примере независимой переменной является Y — производительности труда, зависимыми – трудоемкость единицы продукции *(X1)*, удельный вес комплектующих изделий *(X3)* и фондоотдача (*X7)*.

Сначала следует открыть файл исходных данных *(analiz.sta)*, затем переключиться в модуль **Multiple Regression,** сделать соответствующие установки в окне **Select dependent and independent variable list** и установить флажок в поле **Review descr. stats, corr. Matrix** (Обзор описательных статистик, корреляционная матрица), что позволит провести предварительный анализ исходных переменных и построить корреляционную матрицу, анализ которой дает возможность сделать важные выводы о структуре связей между выбранными переменными (см. рис. 4.27).

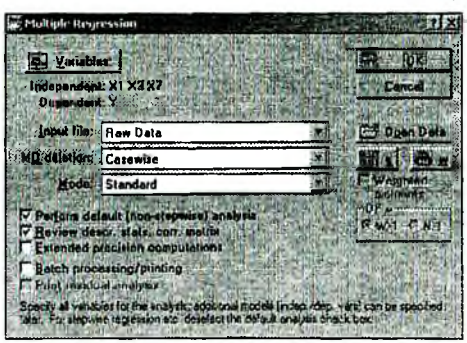


Рис. 4.27. Окно Multiple Regression

После того, как будет нажата кнопка **ОК**, на экране появится окно **Correlations** (рис.4.28), в котором представлены значения коэффициентов парной корреляции. Не рекомендуется включать в модель переменные, слабо связанные с результативным признаком – это фактор X7 (r_{yx7}=0.293).

Correlations (analiz.sta)

Continue...	X1	X3	X7	Y
X1	1,000000	-,648699	-,234445	-,816090
X3	-,648699	1,000000	,600464	,642082
X7	-,234445	,600464	1,000000	,293780
Y	-,816090	,642082	,293780	1,000000

Рис.4.28. Окно Correlations

Наиболее тесную связь с результирующим признаком Y имеют факторы X1 (r_{yx1}=0,816) и X3 (r_{yx3}=0,64). Их и нужно оставить для построения модели.

Далее следует вернуться в окно **Select dependent and independent variable list,** определить в качестве независимых переменных факторы *X1* и *X3* и нажать кнопку **ОК**. Система произведет вычисления, и на экране появится следующее окно результатов (см. рис. 4.29).

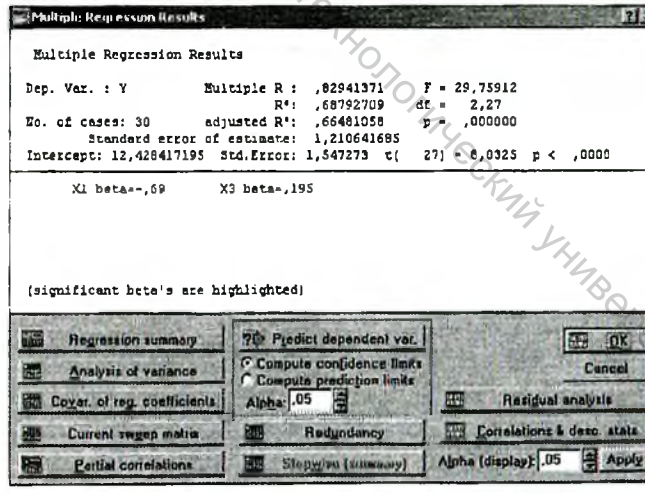


Рис.4.29. Окно Multiple Regression Results

В информационной части окна содержатся краткие сведения о результатах анализа, а именно:

коэффициент детерминации R^2 = 0,688. Это значение показывает, что построенная регрессия объясняет более 68,8% разброса значений переменной Y относительно среднего;

значение F-критерия Фишера и уровень значимости p. В данном примере мы имеем достаточно высокое значение F-критерия — 29,795, а представленный в окне уровень значимости p = 0,00 показывает, что построенная регрессия высоко значима.

Рассмотрим вторую часть информационного окна. В ней представлена информация о значимых и незначимых оценках регрессионных коэффициентов. При этом высвечивается строка

x1 beta = -0,69, x3 beta = 0.195

и приводится пояснение **Significant beta's are highlighted** *(Значимые beta высвечены)*. Отметим, что в данном случае *beta* есть стандартизованные коэффициенты B_1, т. е. коэффициент при независимой переменной *X1* и *B3*, т. е. коэффициент при независимой переменной *X3*.

Перейдем в функциональную часть окна результатов.

Нажав кнопку **Regression summary** *(Итоговый результат регрессии)*, получим на экране **Spreadsheet** *(Электронная таблица вывода)* электронную таблицу с численными результатами оценивания регрессионной модели (см. рис. 4.30).

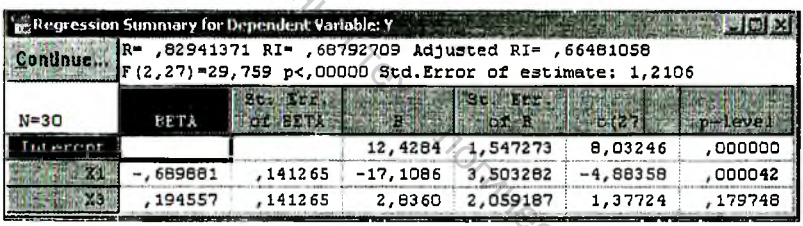


Рис.4.30. Параметры модели множественной регрессии

Верхняя часть окна – информационная, в нижней части находятся параметры модели. В столбце B, например, коэффициенты b_0 = 12,428, b_1 =-17,108, b3 =2,836.

Таким образом, полученное уравнение множественной регрессии имеет вид:

$$Y = -17,108 X_1 + 2,836 X_3 + 12,428.$$

Значения критериев Стьюдента (**t**) позволяют оценить значимость коэффициентов уравнения регрессии, критерий Фишера (***F***=29,759) и скорректированный коэффициент детерминации (***Adjusted R1***=*0,6648)* – значимость построенной модели.

Для получения описательной статистики следует вернуться в окно **Multiple Regression Results,** нажать кнопку **Correlations & desc. stats,** после чего на экране появится окно **Review Descriptive Statistics** (см. рис. 4.20), из

которого следует выбрать необходимые для анализа статистики:

кнопкой **Means & SD** (поставив флажок в поле SD=Sums of Squares/N) смещенные среднеквадратичные отклонения;

кнопкой **Correlations –** коэффициенты корреляции.

Из окна **Multiple Regression Results,** нажав кнопку **Analysis of variance,** можно получить таблицу адекватности – значения общей суммы квадратов, регрессионной суммы квадратов, сумму квадратов остатков, критерий Фишера, число степеней свободы, уровень значимости (см. рис.4.31).

Analysis of Variance; DV: Y (analiz.sta)

MULTIPLE REGRESS.	Sums of Squares	df	Mean Squares	F	p-level
Regress.	87,2331	2	43,61656	29,75912	,000000
Residual	39,5726	27	1,46565		
Total	126,8058				

Рис.4.31. Таблица адекватности

Чтобы посмотреть, как связаны остатки с наблюдаемыми значениями (см. рис.4.32), в окне **Multiple Regression Results** следует нажать кнопку ***Residual Analysis*** *(Анализ остатков)* и в появившемся окне выбрать команду ***Obs & Residuals.***

Чтобы посмотреть, как наблюдаемые значения связаны с предсказанными с помощью построенной модели (см. рис. 4.33), следует нажать кнопку ***Pred & observed(F)***.

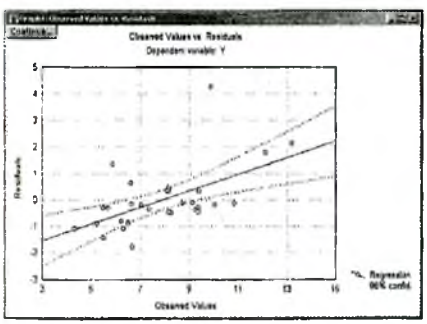


Рис.4.32. График наблюдаемых переменных остатков

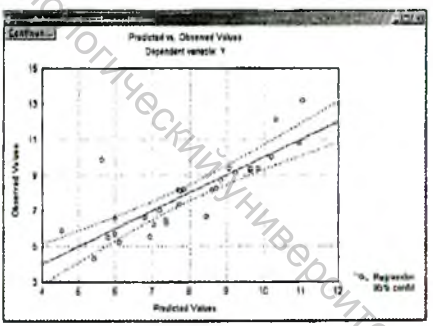


Рис.4.33. График наблюдаемых и предсказанных значений

Из графиков на рис. 4.32 и рис. 4.33 видно, что модель достаточно адекватно описывает данные. Следовательно, с ее помощью можно делать достаточно точные выводы о зависимости производительности труда от трудоемкости единицы продукции и удельного веса комплектующих изделий.

Чтобы получить прогноз значения зависимой переменной Y, в окне **Multiple**

Regression Results следует нажать кнопку **Predict dependent var** и в появившееся на экране окно **Specify values for indep. vars** ввести новые значения *X1*, например 0,18 и X3 например 0,55 и нажать OK (см. рис. 4.34).

Рис. 4.34. Окно Specify values for indep. Vars

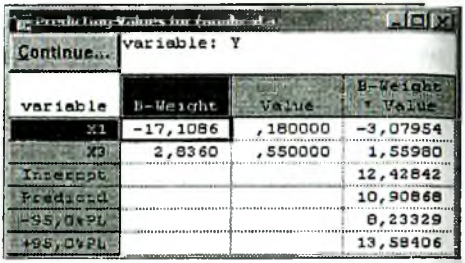


Рис. 4.35. Окно Predicting values for

В результате в окне **Predicting values for** (см. рис.4.35) на основании полученного ранее уравнения регрессии $Y = -17{,}108 \cdot X_1 + 2{,}836 \cdot X_3 + 12{,}428$ будет рассчитано прогнозируемое значение производительности труда *Y* (10,908) при снижении трудоемкости единицы продукции *X1* до 0,18 и уровне удельного веса комплектующих изделий *X3* равном значению 0,55.

4.3.3. Нелинейная регрессия

Кроме линейного регрессионного анализа, *STATISTICA* предоставляет возможность проведения нелинейного регрессионного анализа. Для этой цели служит модуль **Nonlinear Estimation** (Нелинейное оценивание). Он позволяет строить произвольную регрессионную модель, задаваемую некоторой алгебраической формулой, которая может быть нелинейной как по переменным, так и по параметрам. Для расчета модели могут использоваться различные итерационные алгоритмы минимизации. Программа осуществляет полный контроль за всеми аспектами вычислительных процедур (начальное значение, размер шага, критерий сходимости и т.д.). Большинство обычных нелинейных регрессионных моделей задано в модуле и может быть просто выбрано из меню.

Для переключения в этот модуль следует в переключателе разделов **(Statistica Module Switcher)** выбрать раздел **Nonlinear Estimation,** после чего на экране появится окно с перечнем доступных пользователю нелинейных функций для построения регрессионной модели (см. рис. 4.36). Особый интерес вызывает раздел **«Функции, определенные пользователем» (User-specified regression).** Здесь пользователь сам может математически задать вид уравнения регрессии и рассчитать и оценить его.

Обратимся к условию **примера 4.3** и построим нелинейную модель, отражающую зависимость производительности труда *Y* от трудоемкости

единицы продукции *X1* и удельного веса комплектующих изделий *X3*.

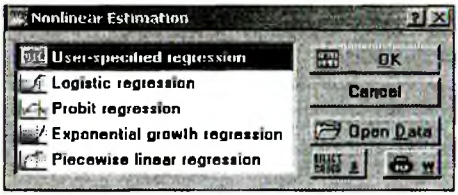


Рис.4.36. Окно Nonlinear Estimation

Для этого в окне **User-Specified Regression Function** (рис.4.37) нужно нажать **Function to be estimated & loss function.**

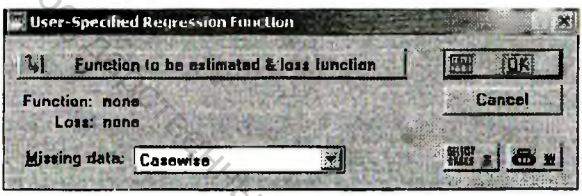


Рис.4.37. Окно User-Specified Regression Function

и в поле **Estimated function** (рис. 4.38) определить уравнение регрессии, которое требуется рассчитать и оценить.

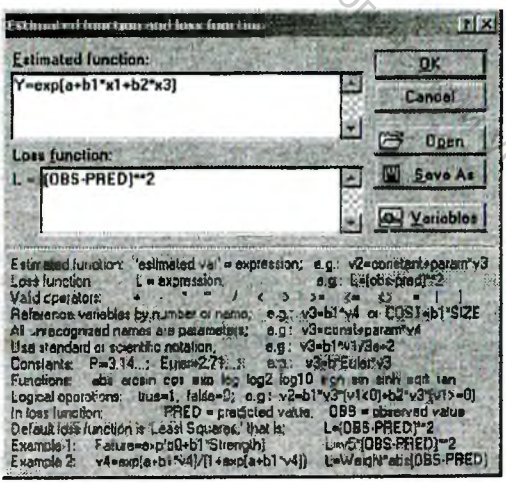


Рис.4.38. Окно Estimated function & loss function

В нижней части окна приведены допускаемые в формулах арифметические операторы и стандартные функции, а также примеры их использования для записи выражений.

Созданную функцию можно сохранить для дальнейшего использования, для чего следует нажать Save As.

Для оценки отклонений между расчетным и фактическим значениями результирующего параметра *(Y)* в поле **Loss function** по умолчанию находится функция **(OBS-PRED)**2** . Сюда также при необходимости можно ввести другую функцию.

Далее следует нажать ОК, перейти в окно **Model estimation** (см. рис.4.39), определить метод (**Estimation metod**), при помощи которого будут рассчитываться коэффициенты уравнения регрессии, число итераций и точность вычислений. Кроме того, обозначить флажком поле **Asymptotic standart errors** для включения в итоговый отчет оценок стандартных ошибок и уровней значимости.

После проведения расчетов результаты, по которым можно оценить адекватность модели, будут находиться в таблице окна **Model** (см. рис. 4.40).

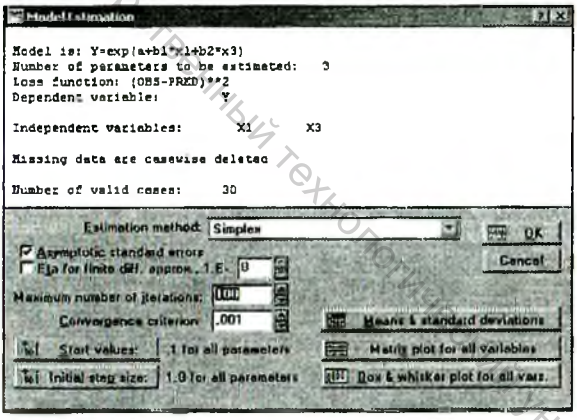


Рис.4.39. Окно Model estimation

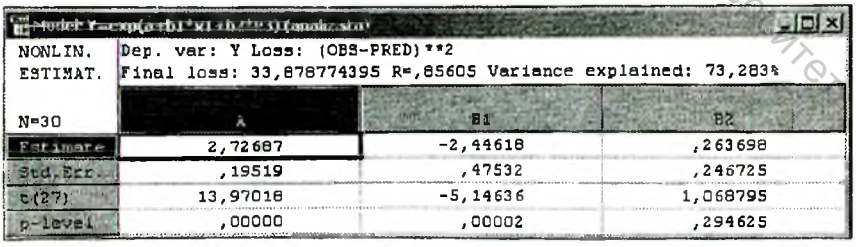


Рис.4.40. Окно Model

4.3. Факторный анализ экономической информации

Факторный анализ – это процедура, с помощью которой большое число показателей-признаков, относящихся к имеющимся наблюдениям, которыми характеризуется экономический процесс, сводится к меньшему количеству искусственно построенных независимых влияющих величин, называемых факторами. При этом в один фактор объединяются переменные, сильно коррелирующие между собой. Переменные из разных факторов слабо коррелируют между собой. Таким образом, целью факторного анализа является нахождение таких комплексных факторов, которые как можно более полно объясняют наблюдаемые связи между переменными, имеющимися в наличии. Следует отметить, что факторы не являются реальными наблюдениями, они присутствуют гипотетически.

На первом шаге процедуры факторного анализа происходит стандартизация заданных значений переменных (z-преобразование); затем при помощи стандартизированных значений рассчитывают корреляционные коэффициенты между рассматриваемыми переменными.

Исходным элементом для дальнейших расчётов является корреляционная матрица. Для построенной корреляционной матрицы определяются так называемые собственные значения и соответствующие им собственные векторы, для определения которых используются оценочные значения диагональных элементов матрицы (так называемые относительные дисперсии простых факторов).

Собственные значения сортируются в порядке убывания, для чего обычно отбирается столько факторов, сколько имеется собственных значений, превосходящих по величине единицу. Собственные векторы, соответствующие этим собственным значениям, образуют факторы. Элементы собственных векторов получили название факторной нагрузки. Их можно понимать как коэффициенты корреляции между соответствующими переменными и факторами. Для решения задачи определения факторов были разработаны многочисленные методы, наиболее часто употребляемым из которых является метод определения главных факторов (компонентов).

Описанные выше шаги расчёта ещё не дают однозначного решения задачи определения факторов. Основываясь на геометрическом представлении рассматриваемой задачи, поиск однозначного решения называют задачей вращения факторов. И здесь имеется большое количество методов, наиболее часто употребляемым из которых является ортогональное вращение по так называемому методу варимакса. Факторные нагрузки повёрнутой матрицы могут рассматриваться как результат выполнения процедуры факторного анализа. Кроме того, на основании значений этих нагрузок необходимо попытаться дать толкование отдельным факторам.

Если факторы найдены и истолкованы, то на последнем шаге факторного анализа отдельным наблюдениям можно присвоить значения этих факторов, так называемые факторные значения. Таким образом, для каждого наблюдения

значения большого количества переменных можно перевести в значения небольшого количества искусственно выявленных факторов.

Любой метод факторного анализа имеет одну главную задачу: представить результирующий фактор в виде линейной комбинации некоторого числа общих факторов и одного характерного фактора.

Пример 4.4. Выполнить в системе *STATISTICA* факторный анализ влияния показателей экономической деятельности предприятий на уровень производительности труда.

В качестве исходных данных используем данные из файла *analiz.sta* (см. условие примера 4.1).

Для переключения в модуль Факторный анализ следует в переключателе разделов **(Statistica Module Switcher)** нужно выбрать раздел **Factor Analysis,** после чего на экране появится стартовая панель модуля, в которой выбрать переменные для проведения факторного анализа (см. рис. 4.41) и нажать OK.

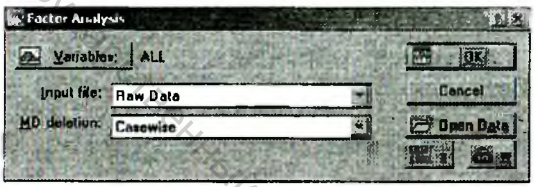


Рис.4.41. Стартовая панель модуля Factor Analysis

Первый этап факторного анализа – это вычисление корреляционной матрицы и числовых характеристик изучаемых признаков. В окне **Define Method of Factor Extraction** (рис.4.42) можно нажать **Review corrs/means/SD,** провести анализ исходных данных и построить корреляционную матрицу, содержащую коэффициенты парной корреляции всех признаков (рис. 4.43).

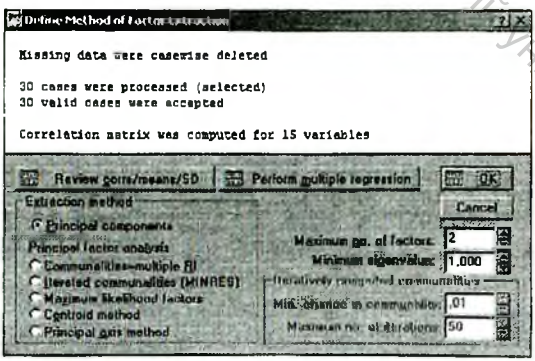
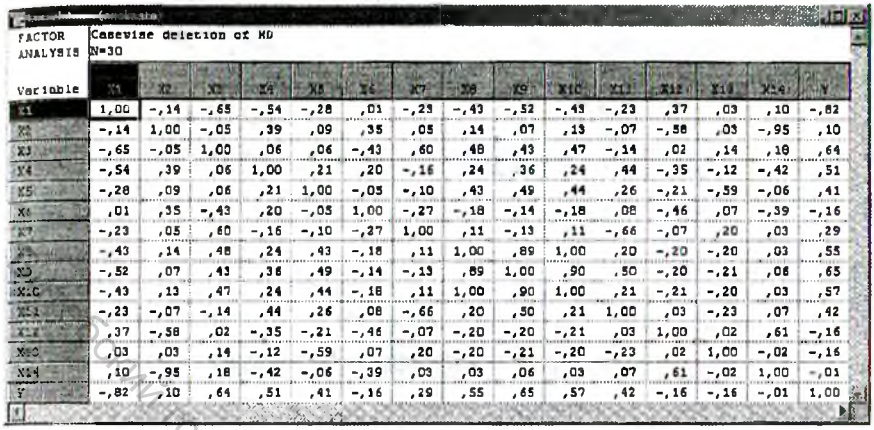


Рис.4.42. Окно Define Method of Factor Extraction



FACTOR ANALYSIS — Casewise deletion of MD, N=30

Variable	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	Y
X1	1,00	-,14	-,65	-,54	-,28	,01	-,23	-,43	-,52	-,43	-,23	,37	,03	,10	-,82
X2	-,14	1,00	-,05	,39	,09	,35	,05	,14	,07	,13	-,07	-,58	,03	-,95	,10
X3	-,65	-,05	1,00	,06	,06	-,43	,60	,48	,43	,47	-,14	,02	,14	,16	,64
X4	-,54	,39	,06	1,00	,21	,20	-,16	,24	,36	,24	,44	-,35	-,12	-,42	,51
X5	-,28	,09	,06	,21	1,00	-,05	-,10	,43	,49	,44	,26	-,21	-,59	-,06	,41
X6	,01	,35	-,43	,20	-,05	1,00	-,27	-,18	-,14	-,18	,08	-,46	,07	-,39	-,16
X7	-,23	,05	,60	-,16	-,10	-,27	1,00	,11	-,13	,11	-,66	-,07	,20	,03	,29
X8	-,43	,14	,48	,24	,43	-,18	,11	1,00	,89	1,00	,20	-,20	-,20	,03	,55
X9	-,52	,07	,43	,36	,49	-,14	-,13	,89	1,00	,90	,50	-,20	-,21	,06	,65
X10	-,43	,13	,47	,24	,44	-,18	,11	1,00	,90	1,00	,21	-,21	-,20	,03	,57
X11	-,23	-,07	-,14	,44	,26	,08	-,66	,20	,50	,21	1,00	,03	-,23	,07	,42
X12	,37	-,58	,02	-,35	-,21	-,46	-,07	-,20	-,20	-,21	,03	1,00	,02	,61	-,16
X13	,03	,03	,14	-,12	-,59	,07	,20	-,20	-,21	-,20	-,23	,02	1,00	-,02	-,16
X14	,10	-,95	,16	-,42	-,06	-,39	,03	,03	,06	,03	,07	,61	-,02	1,00	-,01
Y	-,82	,10	,64	,51	,41	-,16	,29	,55	,65	,57	,42	-,16	-,16	-,01	1,00

Рис. 4.43. Корреляционная матрица

Из корреляционной матрицы (рис.4.43) следует, что уровень производительности труда *Y* на предприятии имеет значимую корреляционную связь не со всеми признаками.

Поэтому признаки, имеющие не значимую корреляционную связь с уровнем производительности труда *Y*, следует исключить из исходного набора независимых переменных *X1-X14*. Это признаки *X2*, *X6*, *X7*, *X12-X14*, имеющие коэффициенты парной корреляции с результирующим признаком *Y* 0.1, -0.16, 0.29, -0.16, -0.16, -0.1 соответственно. Кроме того, признаки *X8*, *X9* и *X10* мультиколлинеарны, то есть в значительной степени коррелируют между собой, поэтому из перечисленных трех признаков следует оставить только один, имеющий больший коэффициент корреляции с результирующим признаком *Y*. Это независимый признак X9 ($r_{x8y} = 0.55$, $r_{x9y} = 0.65$, $r_{x10y} = 0.57$).

Таким образом, для проведения факторного анализа влияния показателей деятельности предприятия на уровень производительности труда рассмотрим следующие признаки:

Y – уровень производительности труда;
X1 – трудоемкость единицы продукции;
X3 – удельный вес комплектующих изделий;
X4 – коэффициент сменности оборудования;
X5 – премии и вознаграждения на одного работника;
X9 – среднегодовая стоимость ОПС;
X11 – фондовооруженность труда.

Вернувшись в стартовую панель модуля **Factor Analysis,** именно эти переменные следует указать в качестве исходных для анализа (см. рис. 4.44).

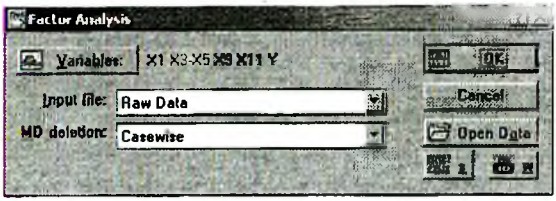


Рис.4.44. Стартовая панель модуля Factor Analysis

Далее необходимо получить и оценить числовые характеристики изучаемых признаков. Это можно сделать в окне выбора **Define Method of Factor Extraction,** нажав **Review corrs/means/SD и Correlations** (см. рис. 4.45).

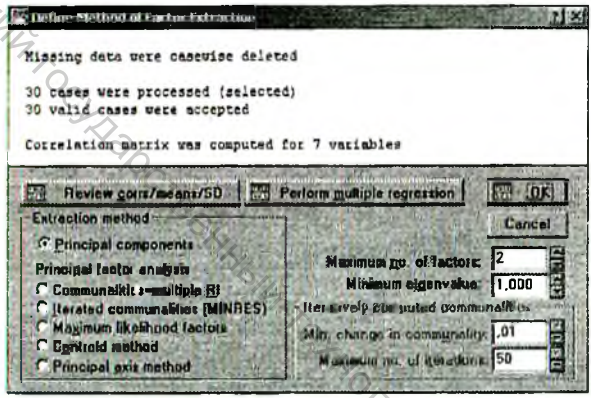
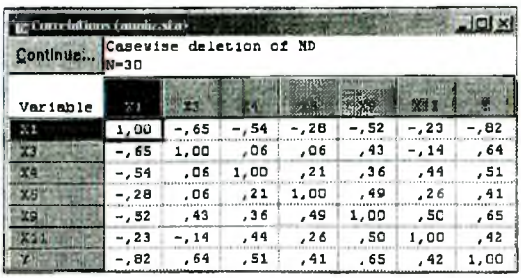


Рис.4.45. Окно Define Method of Factor Extraction

Поскольку числовые характеристики изучаемых признаков близки к нормальным (см. п. 4.2) и уровень производительности труда имеет значимую корреляционную связь со всеми признаками (см. рис.4.46), препятствий к проведению факторного анализа внутренней структуры корреляционной матрицы нет.



Variable	X1	X3	X5	X9	X11	Y
X1	1,00	-,65	-,54	-,28	-,52	-,23	-,82
X3	-,65	1,00	,06	,06	,43	-,14	,64
X5	-,54	,06	1,00	,21	,36	,44	,51
X9	-,28	,06	,21	1,00	,49	,26	,41
X11	-,52	,43	,36	,49	1,00	,50	,65
Y	-,23	-,14	,44	,26	,50	1,00	,42
	-,82	,64	,51	,41	,65	,42	1,00

Рис. 4.46. Корреляционная матрица

Используем метод главных компонент, который дает возможность путем привлечения всех факторных признаков разложить корреляционную матрицу на такое же число ортогональных компонент и определить наибольшее число главных из них.

В окне **Define Method of Factor Extraction** активизируем этот метод и определим максимальное количество факторов (главных компонент) – **Maximum no. of factors** – зададим 2 фактора, кроме того, определим величину минимальных собственных значений главных компонент – 1. Изменяя эти параметры, можно проводить факторный анализ по разным вариантам и выбирать наилучший. Нажав OK, перейдем в следующее окно **Factor Analysis Results** (см. рис. 4.47).

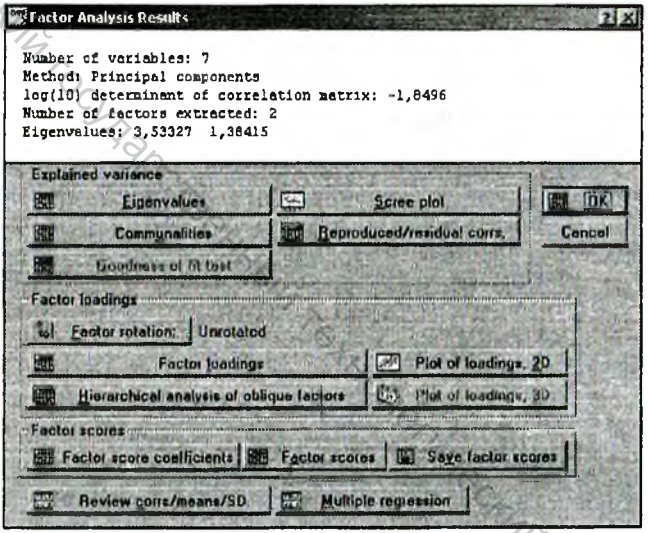


Рис.4.47. Окно Factor Analysis Results

В верхней части окна результатов факторного анализа дается информационное сообщение:

Number of variables (число анализируемых переменных) – 7;

Method (метод анализа – главные компоненты);

Log(10) determination of correlation matrix (десятичный логарифм детерминанта корреляционной матрицы) –1,8496;

Number of factor extraction (число выделенных факторов) – 2;

Eigenvalues (собственные значения) – 3,53327; 2,13761.

В нижней части окна находятся функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа численно и графически.

Нажмем **Eigenvalues** и получим таблицу, в которой будут представлены значения главных компонент (eigenvalues) относительно величины вклада (% total Variance) и накопленный вклад (Cumul. %) главных компонент в

объяснение дисперсии всех признаков (см. рис. 4.48).

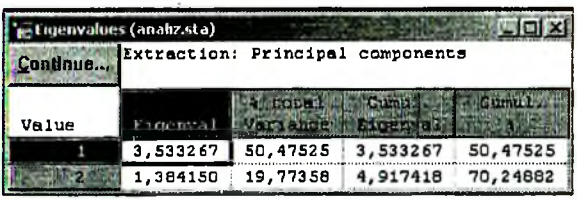


Рис. 4.48. Значения главных компонент

Накопленный вклад двух главных компонент составляет 70,24 %.

Затем из окна **Factor Analysis Results,** нажав **Scree plot**[1] можно получить график с отображением значений всех главных компонент (их максимальное количество равно числу переменных) (см. рис.4.49).

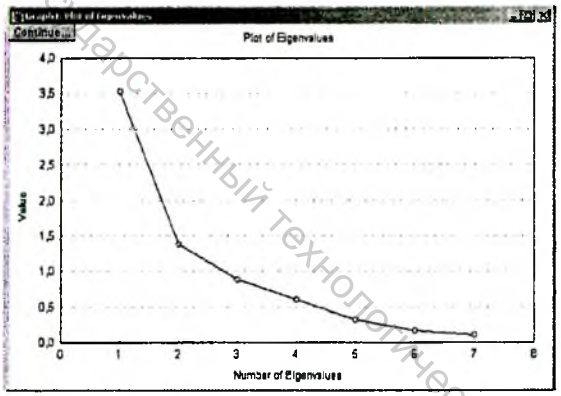


Рис.4.49. Критерий каменистой осыпи

Из графика видно, что в точках с ординатами 1 и 2 осыпание замедляется наиболее существенно, следовательно эти значимые факторы на графике образовывают в своего рода склон, то есть ту часть линии, которая характеризуется крутым подъёмом, поэтому можно ограничиваться одним или двумя факторами.

Для получения матрицы нагрузок (корреляций) трех главных компонент на исходные признаки в окне **Factor Analysis Results** нажмем ОК (см. рис. 4.50).

[1] Слово Screeplot (**критерий каменистой осыпи**), употребляемое для обозначения этого графика, состоит из двух частей: английского слова **scree**, что означает щебень, и слова **plot**, что в английском соответствует графическому представлению. Такая диаграмма служит для того, чтобы маловажные факторы — щебень — можно было отделить от самых значимых факторов.

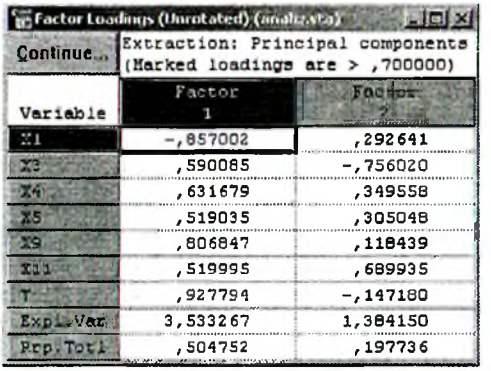


Рис.4.50. Матрица нагрузок

Для облегчения интерпретации главных компонент необходима более тесная привязка их к определенным группам наблюдавшихся признаков. Это достигается путем вращения осей главных компонент в пространстве признаков до нахождения оптимального решения по критерию **Varimax.** В таком оптимальном решении вновь полученные два главных фактора будут иметь тот же накопленный вклад 70,24 % в объяснение дисперсии всех признаков, но, кроме того, более выраженные нагрузки (корреляции) с определенными группами признаков.

В окне **Factor Analysis Results** следует нажать **Factor rotation** и выбрать метод **Varimax normalized.** Система произведет вращение факторов методом нормализованного варимакса, и на экране появится повернутая матрица нагрузок (см. рис.4.51).

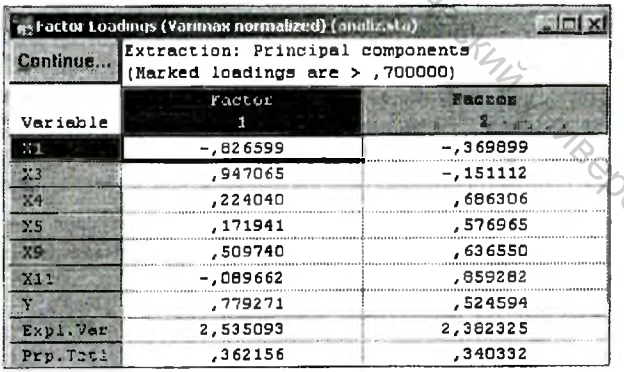


Рис.4.51. Повернутая матрица нагрузок

Объясним отобранные факторы. Эти факторные нагрузки следует

понимать как корреляционные коэффициенты между переменными и факторами. Так, переменные *X1* и *X3* сильнее всего коррелируют с фактором 1, а именно: величина корреляции составляет 0,826 и 0,947 соответственно, переменная X11 коррелирует с фактором 2 (0,859). В большинстве случаев включение отдельной переменной в один фактор, осуществляемое на основе коэффициентов корреляции, является однозначным. В исключительных случаях переменная может относиться к двум факторам одновременно (в рассматриваемом примере X9). Могут быть также и переменные, которыми нельзя нагрузить ни один из отобранных факторов.

Таким образом, первый главный фактор (2,53 по уровню или 36.21% от общей дисперсии), прямо связанный с X3 и обратно с X1, можно определить как показатель живого и прошлого труда.

Второй фактор (2,38 по уровню или 34.03% от общей дисперсии), и сильно связанный с X11 и в средней степени с X4 и X9, можно определить как эффективность использования оборудования.

На основе полученной факторной матрицы (рис. 4.51) можно построить ряд моделей в нормированных значениях.

Например, построим модель зависимости производительности труда от главных факторов f1 и f2:

$$Y_{норм} = 0,7792*f1 + 0,5246* f2 .$$

Натуральное значение производительности труда $Y_{нат}$ можно рассчитать по формуле

$$Yнат = Yср + Yнорм*S\{Y\}, \text{ где}$$

Yср – среднее значение производительности труда;

S{Y} – среднее квадратическое отклонение производительности труда.

В использовании, безусловно, удобнее регрессионные модели в натуральных значениях признаков. С определенной долей достоверности вместо двух факторов f1 и f2 можно использовать наиболее информативные признаки *X3* и *X11*, так как именно на них приходится наибольшая нагрузка (0,947 и 0,859 соответственно). Остальные признаки не учитываются, так как они находятся в сильной корреляционной зависимости с вышеупомянутыми.

Для построения регрессионной модели в окне **Factor Analysis Results** следует нажать **Multiple regression,** затем определить зависимую *(Y)* и независимые *(X3, X11)* переменные и нажать ОК. На экране появится таблица с характеристиками и коэффициентами уравнения регрессии (см. рис.4.52).

Regression Summary for Dependent Variable: Y (analiz.sta)

Continue...

R=,821845 RI=,675429 Adjusted RI=,651387
F(2,27)=28,093 p<,00000 Std.Error of estimate:1,2346

N=30	Beta	St. Err. of Beta	B	St. Err. of B	t(27)	p-level
Intercept			1,95225	,870367	2,243016	,033304
X3	,713938	,110711	10,40687	1,613809	6,448643	,000001
X11	,517998	,110711	,50701	,108362	4,678821	,000072

Рис.4.52. Характеристики регрессионной модели

Все коэффициенты модели (столбец B) значимы ($p < 0.05$), что свидетельствует об истинности сделанных ранее выводов.

Искомая модель имеет вид

$$Y = 10{,}406*X3 + 0{,}507*X11 + 1{,}95.$$

Чтобы рассчитать прогноз, следует переключиться в модуль **Multiple regression** и повторить действия, описанные в примере 4.3 (рис.4.53 и 4.54).

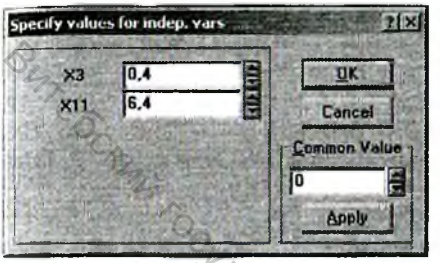


Рис. 4.53. Окно Specify values for indep. Vars

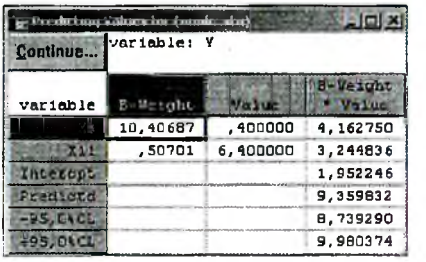


Рис.4.54. Окно Predicting values for

Предсказанное значение производительности труда Y= 9.35 при значении удельного веса комплектующих изделий X3 = 0.4 и значении фондовооруженности труда $X11$ = 6.4 (наблюдаемое значение Y при тех же значениях $X3$ и $X11$ составляет 9.26). Ошибка прогноза – 0,97%.

Следует отметить, что предварительный анализ корреляционной матрицы (рис. 4.43) и исключение незначимых факторов не является обязательным. При проведении факторного анализа Statistica сама будет стремиться оценить влияние каждого фактора (см. матрицу нагрузок, рис. 4.50 и рис.4.51) и исключить факторы, имеющие слабую связь с результирующим признаком.

ГЛАВА 5. СИСТЕМА КОМПЬЮТЕРНОЙ МАТЕМАТИКИ MAPLE

5.1. Характеристика пакета

MAPLE - это пакет для аналитических вычислений на компьютере, содержащий более двух тысяч команд, которые позволяют решать задачи алгебры, геометрии, математического анализа, дифференциальных уравнений, статистики, математической физики, финансов, экономики и т.д.

Когда программа *Maple* запускается, она не имеет ни одной команды, полностью загруженной в память. Большая часть команд имеет указатели их нахождения, и при вызове они загружаются автоматически. Другие команды находятся в стандартной библиотеке и перед выполнением обязательно должны быть вызваны командой

```
readlib(command),
```

где **command** – имя вызываемой команды.

Остальная часть процедур *Maple* содержится в специальных библиотеках подпрограмм, называемых пакетами. Пакеты необходимо подгружать при каждом запуске файла с командами из этих библиотек. Имеется два способа вызова команды из пакета:

1) можно загрузить весь пакет командой `with(package)`,
где **package** – имя пакета;
2) вызов какой-нибудь одной команды **command** из любого пакета **package** можно осуществить, если набрать команду в специальном формате:

```
> package[command](options);
```

где вначале записывается название пакета **package**, из которого надо вызвать команду, а затем в квадратных скобках набирается имя самой команды **command**, и после чего в круглых скобках следуют параметры **options** данной команды.

В таблице 5.1 перечислены библиотеки подпрограмм *Maple* .

Таблица 5.1. Библиотеки Maple

Название	Назначение
DEtools	Дифференциальные уравнения
geom3d	Стереометрия
geometry	Планиметрия
linalg, LinearAlgebra	Линейная алгебра
logic	Математическая логика

Окончание таблицы 5.1

networks	Теория графов
plots	Графический пакет
simplex	Линейная оптимизация
stats	Статистика
student	Пошаговое обучение
finance	Финансовая математика
plot	Двухмерная графика
plot3d	Трехмерная графика
Optimization	Поиск оптимального решения

5.2 Построение и анализ регрессионных моделей в MAPLE

При использовании стандартного офисного приложения Excel для построения экономико-математической модели имеется возможность получения только двух видов моделей: линейной или логарифмической, то есть выбор модели предопределен разработчиками пакета. Так как эти виды моделей не всегда являются оптимальными, возник интерес к возможности использования других пакетов, в частности, системы символьной математики Maple для решения сформулированной выше задачи. СКА Maple позволяет пользователю произвольно, по своему усмотрению, задавать вид модели и при необходимости оперативно его изменять.

Построить и оценить регрессионную модель в MAPLE можно с помощью модулей и функций (таблица 5.2) статистической библиотеки **stats**. Подключение этой библиотеки осуществляется командой

```
>with(stats);
```

Таблица 5.2. Функции библиотеки stats

Функции	Описание
importdata	импорт данных из файла
anova	вариационный анализ
describe	статистические данные
fit	аппроксимация
random	случайные значения
statevalf	численная оценка

Окончание таблицы 5.2

`statplots`	графика
`transform`	преобразования данных

Функция `Fit`

Эта функция предназначена для нахождения корреляционных отношений и для аппроксимации данных выбранными зависимостями с использованием метода наименьших квадратов.

Формат:

```
fit[leastsquare[[x,y]]]([[dataX], [dataY]]);
```

По умолчанию система приближает зависимость к уравнению прямой линии.

Для того, чтобы объединить построение и исследование экономико-математической модели в единую структуру, можно предложить методику, включающую в себя следующие элементы:

- построение модели по экспериментальному набору данных;
- проверка модели на адекватность с расчетом корреляционно - регрессионной статистики;
- графическое отражение реальной и расчетной зависимостей между зависимыми и независимыми переменными с выводом уравнения регрессии на графике.

В качестве объекта исследования используем набор исходных данных, состоящий из двух массивов: массива независимых переменных X и массива зависимых переменных Y.

Пример 5.1. Обратимся к условию примера 3.1 и проведем анализ зависимости среднедневной заработной платы, руб. (Y) от среднедушевого прожиточного минимума в день одного трудоспособного, руб. (x). Таблица П1 с исходными данными приведена в Приложении 1.

Сеанс работы в Maple:

На первом этапе эти массивы следует оформить типом `statsdata` для возможности обработки процедурами и функциями библиотеки `stats` СКМ Maple:

```
> restart;
> with(stats);
```

[*anova*, *describe*, *fit*, *importdata*, *random*, *statevalf*, *statplots*, *transform*]

```
>X:=[2633,2750,2610,2440,2350,2510,3500,3800,3620,2296,21
58,2064,2688,2330,1890,1788];
```

$$X := [2633, 2750, 2610, 2440, 2350, 2510, 3500, 3800, 3620, 2296, 2158, 2064, 2688, 2330, 1890, 1788]$$

```
>Y:=[7855,7963,7785,7580,7420,7550,8500,9100,8560,6800,66
00,6600,7700,7500,6000,6000];
```

$$Y := [7855, 7963, 7785, 7580, 7420, 7550, 8500, 9100, 8560, 6800, 6600, 6600, 7700, 7500, 6000, 6000]$$

Для расчета функциональной зависимости между экспериментальными данными X и Y и возможности ее графического отображения определим функцию пользователя `spisok=f(x)` с использованием функционального оператора →:

```
> spisok:=(x,y)->[x,y];
```

Далее, задав вид модели (например, линейная), рассчитаем коэффициенты уравнения регрессии:

```
> fit[leastsquare[[x, y]]]([X,Y]);
```

$$y = \frac{319188179584}{83641319} + \frac{118019989}{83641319}x$$

приведем полученное уравнение к численному виду:

```
> evalf(%,4);
```

$$y = 3816. + 1.411x$$

Следует отметить, что при определенной ранее функциональной зависимости между независимой и зависимой переменными функция `leastsquare[[x,y],y=f(x)]` позволяет задавать вид модели по усмотрению разработчика. Если задать, например, квадратичную или кубическую зависимость, можно рассчитать и такие модели:

```
> eq:=y=z*x^2+b*x+c;
```

$$eq := y = z x^2 + b x + c$$

```
> evalf(fit[leastsquare[[x, y], eq]]([X,Y]),4);
```

$$y = -.0006551 x^2 + 5.115 x - 1169.$$

```
>f:=evalf(fit[leastsquare[[x,y],y=a*x^3+b*x^2+c*x+d]]([X,
Y]),7);
```

$$f := y = .4493261\ 10^{-6} x^3 - .004330596 x^2 + 14.79227 x - 9393.114$$

Для того, чтобы графически отобразить экспериментальные данные и построить линию тренда, значения X и Y сначала следует сгруппировать попарно функцией zip:

```
> k:=zip(spisok,X,Y);
```

затем на основании полученного уравнения регрессии рассчитать линию тренда:

```
> fun:=rhs(f);
```

$$fun := 1.411025\,x + 3816.154$$

```
> for i from 1 to nops(X) do
  Y1[i]:=evalf(subs({x=X[i]},fun))
end do:
> Y1:=convert(Y1,list);
```

Здесь стандартная функция rhs библиотеки stats выделяет правую часть полученной функциональной зависимости для расчета линии тренда Y1, функция nops в цикле for подсчитывает количество значений X, функция subs осуществляет подстановку значений аргумента из массива X[i] в уравнение регрессии, а функция convert преобразовывает полученный массив Y1 в данные типа list (список) для возможности использования их в функции построения графика plot:

```
> k1:=zip(spisok,X,Y1):
> plot([k,k1], thickness=2, labels=["Независимая
переменная X","Зависимая переменная Y"],
labeldirections=[horizontal, vertical], legend=["Исходные
данные", "Теоретическая модель"], title=cat("Модель
y=",convert(evalf(fun,7),string)));
```

Заданные в функции plot параметры позволяют не только построить реальную и расчетную зависимости, применив различные графические стили и комментарии, но и вывести на графике уравнение регрессии (см. рис.5.1).

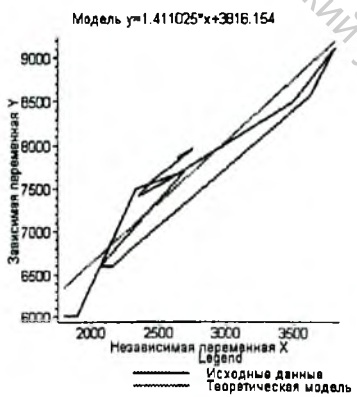


Рис.5.1 – Графическое представление модели

После расчета уравнения регрессии необходимо определить адекватность модели, для чего рассчитывается статистика по регрессии: коэффициент корреляции, коэффициент детерминированности, ошибки по X и по Y, остаточная сумма квадратов, регрессионная сумма квадратов, критерий Фишера и некоторые другие. Для возможности использования стандартных процедур и функций библиотеки `stats` значения `Y` и `Y1` вначале необходимо преобразовать к символьному типу (`array`) и только затем обрабатывать:

```
> y:=convert(Y,array):
> n:=nops(Y):
> sr:=evalf((sum(y[j],j=1..n)/n),7);
```

$$sr := 7469.562$$

```
> Q:=evalf((sum((y[j]-sr)^2,j=1..n)),12);
>
```

$$Q := .117556959375\ 10^{8}$$

```
> y1:=convert(Y1,array):
> Qe:=evalf((sum((y[j]-y1[j])^2,j=1..n)),12);
```

$$Qe := .134762402842\ 10^{7}$$

```
> Qr:=evalf(sum((y1[j]-sr)^2,j=1..n),12);
>
```

$$Qr := .104080724633\ 10^{8}$$

```
> R:=evalf(Qr/(Qr+Qe),4);
```

$$R := .8852$$

```
> k:=1;
```

$$k := 1$$

```
> S:=Qe/(n-k-1);
```

$$S := 96258.85914$$

```
> F:=evalf(Qr/S,4);
```

$$F := 108.1$$

Результаты выводятся функцией `printf`:

```
> printf("Коэффициент корреляции
=>%20.6f\nКоэффициент детерминированности
=>%18.4f\nРегрессионная сумма квадратов
=>%15.1f\nОстаточная сумма квадратов      =>%15.1f\nОбщая
сумма квадратов           =>%15.1f\nКритерий Фишера,
=>%16.2f\n", correl,R,Qr,Qe,Q,F);
Коэффициент корреляции          =>          .941100
Коэффициент детерминированности =>          .8852
Регрессионная сумма квадратов   =>    10408072.5
```

```
Остаточная сумма квадратов       =>      1347624.0
Общая сумма квадратов            =>     11755695.9
Критерий Фишера,                 =>         108.10
```

Чтобы сделать прогноз на последующие периоды, нужно задать новые значения X (xn) и применить описанную выше методику:

```
> xn:=[4000,4500,5000];
> for i from 1 to nops(xn) do
  Yn[i]:=evalf(subs({x=xn[i]},fun))
end do:
> Yn:=evalf(convert(Yn,list),5);
```

$$Yn := [9460.3,\ 10166.,\ 10871.]$$

```
> A:=[correl,R,F,Qr,Qe];
```

$$A := [.9411,\ .9583,\ 323.2,\ .112675414336\ 10^{8},\ 488248.774515]$$

```
> for i from 1 to nops(A) do
> printf("%14.2f  \n",A[i])end do;
          .94
          .89
       108.10
  10408072.46
   1347624.03
```

Пример 5.2. Обратимся к условию примера 3.2. (таблица П2 с исходными данными приведена в Приложении 2) и решим задачу по отысканию вида функциональной зависимости производительности труда (Y) от трудоемкости единицы продукции (x_1), удельного веса комплектующих изделий (x_3), коэффициента сменности оборудования (x_4), премии и вознаграждения на одного рабочего (x_5), среднегодовой стоимости ОПС (x_9) и фондовооруженности труда (x_{11}).

Предположим линейную функциональную зависимость вида

$$y = m_1x_1 + m_2x_2 + m_3x_3 + m_4x_4 + m_5x_5 + m_6x_6 + b.$$

Введем исходные данные в виде массивов (**X1**, **X2**, **X3**, **X4**, **X5**, **X6**, **Y**), где

X1 – трудоемкость единицы продукции;
X2 – удельный вес комплектующих изделий;
X3 – коэффициент сменности оборудования;
X4 – премии и вознаграждения на одного рабочего;
X5 – среднегодовой стоимости ОПС;
X6 – фондовооруженность труда;
Y – производительность труда.

```
>evalf(fit[leastsquare[[x1,x2,x3,x4,x5,y]]]([X1,X2,X3,X4,X
5,Y]),4);
```

$$Y := -8.14\,X1 + 6.704\,X2 + 2.607\,X3 + .809\,X4 - .00041\,X5 + .3\,X6 + 2.57$$

Результаты расчетов примеров 5.1 и 5.2, полученные в Maple и Excel, совпадают, что говорит об их достоверности и возможности использования любого из этих программных продуктов.

Describe

Эта функция позволяет вычислять широкий спектр описательных характеристик, используемых при анализе статистических данных. Выделим некоторые наиболее употребляемые параметры этой функции (см. таблицу 5.3).

Таблица 5.3. Параметры функции ***Describe***

Параметры	Описание
coefficientofvariation	коэффициент вариации
count	число элементов
covariance	линейная ковариация
geometricmean	среднее геометрическое
linearcorrelation	линейная корреляция
mean	среднее арифметическое
median	медиана
quadraticmean	квадратичное среднее арифметическое
standarddeviation	стандартное отклонение
variance	дисперсия

Покажем на примерах использование функции **Describe** с описанными выше параметрами:

```
> restart:with(stats);
> describe[linearcorrelation](X,Y):evalf(%,4);
                                .9411

> describe[coefficientofvariation](X):evalf(%,4);
                                .2208

> describe[count](Y);
                                16

> describe[covariance](X,Y):evalf(%);
```

461015.5820

```
> describe[geometricmean](X):evalf(%,5);
                                2530.9

> describe[mean](X):evalf(%,5);
                                2589.2

> describe[median](X):evalf(%,5);
                                2475.

> describe[quadraticmean](X):evalf(%,5);
                                2651.5

> describe[variance](X): evalf(%,7);
                                326723.9
```

5.3. Поиск оптимальных решений

Рассмотрим использование стандартной библиотеки Optimization, которая позволяет отыскивать оптимальные решения для задач следующего вида:

- линейного программирования LPSolve;
- квадратического программирования QPSolve;
- нелинейного программирования NLPSolve;
- среднеквадратического отклонения LSSolve.

Решим задачу линейного программирования, сформулированную нами для разбора примера отыскания оптимального решения в электронных таблицах Excel (см. пример 3.6).

Фирма производит три вида продукции (A, B, C), для выпуска каждого требуется определенное время обработки на всех четырех устройствах.

Вид продукции	Время обработки, ч.				Прибыль, у.е.
	I	II	III	IV	
A	1	3	1	2	3
B	6	1	3	3	6
C	3	3	2	4	4

Пусть время работы на устройствах I, II, III, IV составляет соответственно 84, 42, 21 и 42 часа.

Требуется рассчитать план производства, обеспечивающий максимальную прибыль.

Целевая функция: `func:=3*A+6*B+4*C.`

Ограничения: 1*A+6*B+3*C<=84
3*A+1*B+3*C<=42
1*A+3*B+2*C<=21
2*A+3*B+4*C<=42
A, B, C – неотрицательные (> =0)
A, B, C – целочисленные.

Приведем решение в системе Maple:

```
>restart;
>with(Optimization);
```

[ImportMPS, Interactive, LPSolve, LSSolve, Maximize, Minimize, NLPSolve, QPSolve]

```
>func:=3*A+6*B+4*C;
```

$$3A+6B+4C$$

```
>organ:={1*A+6*B+3*C<=84, 3*A+1*B+3*C<=42,
1*A+3*B+2*C<=21, 2*A+3*B+4*C<=42};
```

{1A+6B+3C<=84, 3A+1B+3C<=42, 1A+3B+2C<=21, 2A+3B+4C<=42};

```
>LPSolve(func, organ, assume={nonnegative, integer},
maximize);
```

[54., [A=12, B=3, C=0]]

Результаты полностью совпадают с результатами, полученными в Excel.

Литература

1. Экономико–математические методы и модели. Компьютерные технологии решения : учеб. пособие / И. Л. Акулич [и др.]. – Минск : БГЭУ, 2003. – 348 с.
2. Экономико–математические методы и модели : учеб. пособие / Н. И. Холод [и др.] ; под общ. ред. А. В. Кузнецова. – Минск : БГЭУ, 1999. – 413 с.
3. Бронштейн, И. Н. Справочник по математике для инженеров и учащихся втузов / И. Н. Бронштейн, К. А. Семендяев. – Москва : Наука, 1986. – 544 с.
4. Шуп, Т. Решение инженерных задач на ЭВМ / Т. Шуп. – Москва : Мир, 1982. – 238 с.
5. Поляк, Б. Т. Введение в оптимизацию / Б. Г. Поляк, – Москва : Наука, 1983. – 384 с.
6. Лычкина, Н. Н. Имитационное моделирование экономических процессов : учеб. пособие для слушателей программы eMBI. / Н. Н. Лычкина. – Москва : Академия АйТи, 2005. – 164 с.
7. Мур, Дж. Экономическое моделирование в Microsoft Excel : пер. с англ. / Дж. Мур, Л. Уэдерфорд. – 6-е изд. – Москва : Издательский дом "Вильямс", 2004. – 1024 с.
8. Неуймин, Я. Г. Модели в науке и технике. История, теория, практика / Я. Г. Неуймин. – Ленинград : Наука, 1984. – 190 с.
9. Болтянский, В. Г. Математические методы оптимального управления / В. Г. Болтянский. – Москва : Наука, 1969. – 325 с.
10. Орлов, А. И. Эконометрика / А. И. Орлов. – Москва : Экзамен, 2003. – 576 с.
11. Нейлор, Т. Машинные имитационные эксперименты с моделями экономических систем / Т. Нейлор. – Москва : Мир, 1975. – 500 с.
12. Белов, В. В. Теория графов / В. В. Белов, Е. М. Воробьев, В. Е. Шаталов. – Москва : Высшая школа, 1976. – 392 с.
13. Долженков, В. А. Microsoft Excel 2002 / В. А. Долженков, Ю. В. Колесников. – Санкт-Петербург : БХВ – Петербург, 2002. – 1072 с.
14. Дьяконов, В. Maple 7 : учебный курс / В. Дьяконов. – Санкт–Петербург. : Питер, 2002. – 672 с.
15. Костевич, Л. С. Теория игр. Исследование операций / Л. С. Костевич, А. А. Лапко. – Минск : Выш. школа, 1982. – 375 с.
16. Айвазян, С. А. Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – Москва : Статистика, 1974. – 285 с.
17. Балашевич, В. А. Экономико-математическое моделирование производственных систем : учеб. пособие для высш. учеб. заведений / В. А. Балашевич, А. М. Андронов. – Мінск : Універсітэцкае, 1995. – 185 с.

18. Дубров, А. М. Моделирование рисковых ситуаций в экономике и бизнесе : учеб. пособие для высш. учеб. заведений / А. М. Дубров, Б. А. Лагоша, Е. Ю. Хрусталев ; под общ. ред. Б. А. Лагоша. – Москва : Финасы и статистика, 1999. – 245 с.
19. Дубров, А. М. Многомерные статистические методы / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. – Москва : Финансы и статистика, 1998. – 168 с.
20. Исследование операций в экономике : учеб. пособие для высш. учеб. заведений / Н. В. Кремер [и др.] ; под общ. ред. Н. В. Кремер. – Москва : Банки и биржи, ЮНИТИ, 1997. – 225 с.
21. Калянов, Г. Н. Case структурный системный анализ (автоматизация и применение) / Г. Н. Калянов. – Москва : Изд-во "Лори", 1996. – 228 с.
22. Карлберг, К. Бизнес-анализ с помощью Excel : пер. с англ. / К. Карлберг. – Киев : Диалектика, 1997. – 345 с.
23. Романов, А. Н. Компьютеризация финансово-экономического анализа коммерческой деятельности предприятий, корпораций, фирм / А. Н. Романов, И. Я. Лукасевич, Г. А. Титоренко. – Москва : Интер-пракс, 1994. – 345 с.
24. Колемаев, В. А. Математическая экономика : учеб. пособие для эконом. высш. учеб. заведений / В. А. Колемаев. – Москва : ЮНИТИ, 1998. – 325 с.

Таблица П1. Показатели уровня жизни по территориям регионов республики Беларусь за 200Хг

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб.	Среднедневная заработная плата, руб.
№	X	Y
1	2633	7855
2	2750	7963
3	2610	7785
4	2440	7580
5	2350	7420
6	2510	7550
7	3500	8500
8	3800	9100
9	3620	8560
10	2296	6800
11	2158	6600
12	2064	6600
13	2688	7700
14	2330	7500
15	1890	6000
16	1788	6000

Таблица П2. Показатели экономической деятельности предприятия

	Y1	Y2	Y3	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14
1	9,26	204,20	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45	26006,00	167,69	47750,00	6,40	166,32	10,08	17,72
2	9,38	209,60	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30	23935,00	186,10	50391,00	7,80	92,88	14,76	18,39
3	12,11	222,60	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37	22589,00	220,45	43149,00	9,76	158,04	6,48	26,46
4	10,81	236,70	12,85	0,17	0,70	0,50	1,42	0,43	0,18	1,65	21220,00	169,30	41089,00	7,90	93,96	21,96	22,37
5	9,35	62,00	10,63	0,23	0,62	0,40	1,35	0,88	0,15	1,91	7394,00	39,53	14257,00	5,35	173,88	11,88	28,13
6	9,87	53,10	9,12	0,43	0,76	0,19	1,39	0,57	0,34	1,68	11586,00	40,41	22661,00	9,90	162,30	12,60	17,55
7	8,17	172,10	25,83	0,31	0,73	0,25	1,16	1,72	0,38	1,94	26609,00	102,96	52509,00	4,50	88,56	11,52	21,92
8	9,12	56,50	23,39	0,26	0,71	0,44	1,27	1,70	0,09	1,89	7801,00	37,02	14903,00	4,88	101,16	8,28	19,52
9	5,88	52,60	14,68	0,49	0,69	0,17	1,16	0,84	0,14	1,94	11587,00	45,74	25587,00	3,46	166,32	11,52	23,99
10	6,30	46,60	10,05	0,36	0,73	0,39	1,25	0,60	0,21	2,06	9475,00	40,07	16821,00	3,60	140,76	32,40	21,76
11	6,22	53,20	13,99	0,37	0,68	0,33	1,13	0,82	0,42	1,96	10811,00	45,44	19459,00	3,56	128,52	11,52	25,68
12	5,49	30,10	9,68	0,43	0,74	0,25	1,10	0,84	0,05	1,02	6371,00	41,08	12973,00	5,65	177,84	17,28	18,13
13	6,50	146,40	10,03	0,35	0,66	0,32	1,15	0,67	0,29	1,85	26761,00	136,14	50907,00	4,28	114,48	16,20	25,74
14	6,61	18,10	9,13	0,38	0,72	0,02	1,23	1,04	0,48	0,88	4210,00	42,39	6920,00	8,85	93,24	13,32	21,21
15	4,32	13,60	5,37	0,42	0,68	0,06	1,39	0,66	0,41	0,62	3557,00	37,39	5736,00	8,52	126,72	17,28	22,97
16	7,37	89,80	9,86	0,30	0,77	0,15	1,38	0,86	0,62	1,09	14148,00	101,78	26705,00	7,19	91,80	9,72	16,38
17	7,02	62,50	12,62	0,32	0,78	0,08	1,35	0,79	0,56	1,60	9872,00	47,55	20068,00	4,82	69,12	16,20	13,21
18	8,25	46,30	5,02	0,25	0,78	0,20	1,42	0,34	1,76	1,53	5975,00	32,61	11487,00	5,46	66,24	24,88	14,48
19	8,15	103,50	21,18	0,31	0,81	0,20	1,37	1,60	1,31	1,40	16662,00	103,25	32029,00	6,20	67,68	14,76	13,38
20	8,72	73,30	25,17	0,26	0,79	0,30	1,41	1,46	0,45	2,22	9166,00	38,95	18946,00	4,25	50,40	7,56	13,69
21	6,64	76,60	19,40	0,37	0,77	0,24	1,35	1,27	0,50	1,32	15118,00	81,32	28025,00	5,38	70,56	8,64	16,66
22	8,10	73,01	21,00	0,29	0,78	0,10	1,48	1,58	0,77	1,48	11429,00	67,26	20968,00	5,88	72,00	8,64	15,06
23	5,52	32,30	6,57	0,34	0,72	0,11	1,24	0,68	1,20	0,68	6462,00	59,92	11049,00	9,27	97,20	9,00	20,09
24	9,37	199,60	14,19	0,23	0,79	0,47	1,40	0,86	0,21	2,30	24628,00	107,34	45893,00	4,36	8028,00	14,76	15,98
25	13,17	598,10	15,81	0,17	0,77	0,53	1,45	1,98	0,25	1,37	49727,00	512,60	99400,00	10,31	51,48	10,08	18,27
26	6,67	71,20	5,23	0,29	0,80	0,34	1,40	0,33	0,15	1,51	11470,00	53,81	20719,00	4,69	105,12	14,76	14,42
27	5,68	90,80	7,99	0,41	0,71	0,20	1,28	0,45	0,66	1,43	19448,00	80,83	36813,00	4,16	128,52	10,44	22,76
28	5,22	82,10	17,50	0,41	0,79	0,24	1,33	0,74	0,74	1,82	18363,00	59,42	33956,00	3,13	94,68	14,76	15,41
29	10,02	76,20	17,16	0,22	0,76	0,54	1,22	0,03	0,32	2,62	9185,00	36,96	17016,00	4,02	85,32	20,52	19,35
30	8,16	119,5	14,54	0,29	0,78	0,4	1,28	0,99	0,89	1,75	17478	91,43	34873	5,23	76,32	14,4	16,83

Y1 – Производительность труда, млн. руб.
Y2 – Индекс снижения себестоимости продукции, %.
Y3 – Рентабельность, %.
X1 – Трудоёмкость единицы продукции, час.
X2 – Удельный вес рабочих в составе ППП
X3 – Удельный вес комплектующих изделий
X4 – Коэффициент сменности оборудования
X5 – Премии и вознаграждения на одного рабочего, млн. руб.
X6 – Удельный вес потерь от брака в себестоимости продукции, %.
X7 – Фондоотдача, руб.
X8 – Среднесписочная численность ППП
X9 – Среднегодовая стоимость ОПС, млрд. руб.
X10 – Фонд з/п ППП, млн. руб.
X11 – Фондовооруженность труда, млн. руб.
X12 – Оборачиваемость нормируемых оборотных средств
X13 – Оборачиваемость ненормируемых оборотных средств
X14 – Непроизводительные расходы

Пример анализа влияния фондовых показателей деятельности предприятия на уровень рентабельности. Исходные данные размещены в таблице П3.

Таблица П3. Исходные данные

	1 Рентабельность	2 Уд.вес рабочих в ППП	3 Уд.вес покуп. изделий	4 Коэф.сменности оборуд.	5 Среднегод.числен. ППП	6 Среднегод. стоимость ОПФ
1	13,26	0,78	0,40	1,37	26006	167,89
2	10,16	0,75	0,26	1,49	23935	186,10
3	13,72	0,68	0,40	1,44	22589	220,45
4	12,85	0,70	0,50	1,42	21220	189,30
5	10,63	0,62	0,40	1,35	7394	39,53
6	7,12	0,78	0,19	1,39	11586	40,41
7	25,83	0,73	0,25	1,16	26809	102,96
8	23,39	0,71	0,44	1,27	7801	37,02
9	14,88	0,69	0,17	1,16	11567	45,74
10	10,05	0,73	0,39	1,25	9475	40,07
11	13,99	0,68	0,33	1,13	10811	45,44
12	9,68	0,74	0,25	1,10	6371	41,09
13	10,09	0,68	0,32	1,15	26781	136,14
14	3,13	0,77	0,02	1,23	4210	42,39
15	5,37	0,68	0,06	1,39	3557	37,39
16	9,86	0,77	0,15	1,36	14146	101,79
17	12,82	0,78	0,06	1,36	9872	47,55
18	5,82	0,78	0,20	1,42	5976	32,61
19	21,19	0,81	0,20	1,37	19862	103,25
20	25,17	0,79	0,30	1,41	9186	39,95
21	19,40	0,77	0,24	1,36	15118	81,32
22	21,00	0,78	0,10	1,48	11429	67,26
23	9,57	0,72	0,11	1,24	8462	59,90
24	14,19	0,79	0,47	1,40	24628	107,34
25	15,81	0,77	0,63	1,45	49727	512,80
26	5,23	0,80	0,34	1,40	11470	53,81
27	7,99	0,71	0,20	1,28	19448	80,83
28	17,50	0,79	0,24	1,33	18963	59,42
29	17,16	0,76	0,54	1,22	9186	36,96
30	14,54	0,78	0,40	1,26	17479	91,43

В качестве зависимой переменной будет выступать рентабельность (столбец 1), в качестве независимых переменных – столбцы 2-6:

удельный вес рабочих в составе ППП, удельный вес покупных изделий, коэффициент сменности оборудования, среднегодовая численность ППП, среднегодовая стоимость ОПФ.

Интегрированная система Statistica 6.0

Регрессия

N=30	Regression Summary for Dependent Variable: Рентабельность (Р R= ,43472544 R?= ,18896521 Adjusted R?= ,02002500 F(5,24)=1,1185 p<,37708 Std Error of estimate: 5,7239					
	Beta	Std Err. of Beta	B	Std Err. of B	t(24)	p-level
Intercept			-8,51757	18,56223	-0,351120	0,728582
Уд.вес рабочих в ППП	0,247676	0,210508	30,00301	25,50027	1,176576	0,250903
Уд.вес покуп. изделий	0,195870	0,211187	7,89479	8,51219	0,927469	0,362917
Коэф.сменности оборуд.	-0,095366	0,225848	-5,03739	11,92984	-0,422252	0,676600
Среднегод.числен. ППП	0,500562	0,460543	0,00031	0,00027	1,154046	0,259845
Среднегод. стоимость ОПФ	-0,442532	0,456257	-0,02726	0,02823	-0,965686	0,343832

Описательная статистика

Correlations of Regression Coefficients B; DV: Рентабельность (Rentabeln sta)

Variable	Уд.вес рабочих в ППП	Уд.вес покуп. изделий	Коэф.сменности оборуд.	Среднегод.числен. ППП	Среднегод. стоимость ОПФ
Уд.вес рабочих в ППП	1,000000	0,106251	-0,443457	-0,300895	0,293022
Уд.вес покуп. изделий	0,106251	1,000000	0,021348	-0,233172	-0,000633
Коэф.сменности оборуд.	-0,443457	0,021348	1,000000	0,292336	-0,427718
Среднегод.числен. ППП	-0,300895	-0,233172	0,292336	1,000000	-0,878615
Среднегод. стоимость ОПФ	0,293022	-0,000633	-0,427718	-0,878615	1,000000

Covariances of Regression Coefficients B; DV: Рентабельность (Rentabeln sta)

Variable	Уд.вес рабочих в ППП	Уд.вес покуп. изделий	Коэф.сменности оборуд.	Среднегод.числен. ППП	Среднегод. стоимость ОПФ
Уд.вес рабочих в ППП	661,2641	23,06329	-134,909	-0,002077	0,210504
Уд.вес покуп. изделий	23,063	72,46744	2,166	-0,000537	-0,000126
Коэф.сменности оборуд.	-134,909	2,16785	142,321	0,000944	-0,144023
Среднегод.числен. ППП	-0,002	-0,00054	0,001	0,000000	-0,000007
Среднегод. стоимость ОПФ	0,211	-0,00013	-0,144	-0,000007	0,000797

NCSS Statistical and Data Analysis

Multiple Regression Report

Page/Date/Time 1 13.12.2005 16:28:45
Database
Dependent Y

Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Y	Rows Processed	30
Number Ind. Variables	5	Rows Filtered Out	0
Weight Variable	None	Rows with X's Missing	0
R2	0,19	Rows with Weight Missing	0
Adj R2	0,02	Rows with Y Missing	0
Coefficient of Variation	0,42	Rows Used in Estimation	30
Mean Square Error	32,76	Sum of Weights	30,00
Square Root of MSE	5,72	Completion Status	Normal Completion
Ave Abs Pct Error	36,25		

Descriptive Statistics Section

Variable	Count	Mean	Standard Deviation	Minimum	Maximum
X1	30	0,74	0,05	0,62	0,81
X2	30	0,28	0,14	0,02	0,54
X3	30	1,32	0,11	1,1	1,49
X4	30	15321,43	9626,87	3557	49727
X5	30	94,22	93,88	32,61	512,6

Y	30	13,50	5,78	5,02	25,83

Regression Equation Section

Independent Variable	Regression Coefficien b(i)	Standard Error to test Sb(i)	Prob H0:B(i)=0	Level	Reject H0 at 5,0%?	Power of Test at 5,0%
Intercept	-6,52	18,56	-0,35	0,73	No	0,06
X1	30,00	25,50	1,18	0,25	No	0,20
X2	7,89	8,51	0,93	0,36	No	0,14
X3	-5,04	11,93	-0,42	0,68	No	0,07
X4	0,00	0,00	1,15	0,26	No	0,20
X5	-0,03	0,03	-0,97	0,34	No	0,15

Estimated Model

-6.51757324242446+ 30.0030134781668*X1+ 7.89479437491242*X2-
5.03739438502802*X3+
3.12359142056359E-04*X4-2.72569392544667E-02*X5

Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95,0% C.L.	Upper 95,0% C.L.	Standardized Coefficient
Intercept	-6,52	18,56	-44,83	31,79	0,00
X1	30,00	25,50	-22,63	82,63	0,25
X2	7,89	8,51	-9,67	25,46	0,20
X3	-5,04	11,93	-29,66	19,58	-0,10
X4	0,00	0,00	0,00	0,00	0,52
X5	-0,03	0,03	-0,09	0,03	-0,44

Note: The T-Value used to calculate these confidence limits was 2,06.

Multiple Regression Report

Page/Date/Time 2 13.12.2005 16:28:45
Database
Dependent Y

Analysis of Variance Section

Source	DF	R2	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5,0%)
Intercept	1		5471,01	5471,01			
Model	5	0,19	183,23	36,65	1,12	0,38	0,33
Error	24	0,81	786,31	32,76			
Total(Adjusted)	29	1,00	969,54	33,43			

Analysis of Variance Detail Section

Model Term (5,0%)	DF	R2	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power
Intercept	1		5471,01	5471,01			
Model	5	0,19	183,23	36,65	1,12	0,38	0,33
C2	1	0,05	45,35	45,35	1,38	0,25	0,20
C3	1	0,03	28,18	28,18	0,86	0,36	0,14
C4	1	0,01	5,84	5,84	0,18	0,68	0,07
C5	1	0,05	43,63	43,63	1,33	0,26	0,20
C6	1	0,03	30,55	30,55	0,93	0,34	0,15
Error	24	0,81	786,31	32,76			
Total(Adjusted)	29	1,00	969,54	33,43			

Табличный процессор MS EXCEL

Модель

m5	m4	m3	m2	m1	b
-0,027256939	0,000312369	-5,037394385	7,884794375	30,00301348	-6,517573242
0,026225467	0,000270664	11,92983938	8,51219368	25,50027101	18,56222677
0,186986207	5,723893899	#Н/Д	#Н/Д	#Н/Д	#Н/Д
1,118618329	24	#Н/Д	#Н/Д	#Н/Д	#Н/Д
183,2296539	786,3110727	#Н/Д	#Н/Д	#Н/Д	#Н/Д

y=m1*x1+m2*x2+m3*x3+m4*x4+m5*x5+b

Рентабельность=m1*(Уд.вес рабочих в ППП)+m2*(Уд.вес покуп. изделий)+m3*(Коэф.сменности сборуд.)+
m4*(Среднегод.числен. ППП)+m5*(Среднегод. стоимость ОПФ)+b

Рентабельность=30,00*(Уд.вес рабочих в ППП)+7,89*(Уд.вес покуп. изделий)-5,04*(Коэф.сменности оборуд.)+
0,0003*(Среднегод.числен. ППП)-0,027*(Среднегод. стоимость ОПФ)-6,52

Статистика

A	B	C	D	E	F	G	H	I	J	K	L
Рентабельность		Уд.вес рабочих в ППП		Уд.вес покуп. изделий		Коэф.сменности сборуд.		Среднегод.числен. ППП		Среднегод.стоимость ОПФ	
Среднее	13,50433333	Среднее	0,741	Среднее	0,282666667	Среднее	1,322	Среднее	15321,43333	Среднее	84,22466667
Стандартная ош	1,055859501	Стандартная ош	0,009714501	Стандартная ош	0,02819087	Стандартная ош	0,019985063	Стандартная ош	1757,617526	Стандартная ош	17,13923942
Медиана	13,055	Медиана	0,755	Медиана	0,255	Медиана	1,35	Медиана	11586,5	Медиана	59,87
Мода	#Н/Д	Мода	0,78	Мода	0,4	Мода	1,95	Мода	#Н/Д	Мода	#Н/Д
Стандартное отк	5,782078736	Стандартное отк	0,047731266	Стандартное отк	0,143453305	Стандартное отк	0,109482632	Стандартное отк	9626,887678	Стандартное отк	93,87547503
Дисперсия выбо	33,43144809	Дисперсия выбо	0,002278276	Дисперсия выбо	0,020578961	Дисперсия выбо	0,011992066	Дисперсия выбо	92675581,23	Дисперсия выбо	8812,804912
Эксцесс	-0,368466844	Эксцесс	-0,257652868	Эксцесс	0,861895817	Эксцесс	-0,804267608	Эксцесс	4,234154707	Эксцесс	13,87488818
Асимметричнос	0,573058367	Асимметричнос	-0,67787701	Асимметричнос	0,071750666	Асимметричнос	-0,49813949	Асимметричнос	1,648268297	Асимметричнос	3,335725447
Интервал	20,81	Интервал	0,19	Интервал	0,52	Интервал	0,39	Интервал	46170	Интервал	479,89
Минимум	5,02	Минимум	0,62	Минимум	0,02	Минимум	1,1	Минимум	3557	Минимум	32,61
Максимум	25,83	Максимум	0,81	Максимум	0,54	Максимум	1,49	Максимум	49727	Максимум	512,5
Сумма	405,13	Сумма	22,23	Сумма	8,48	Сумма	39,66	Сумма	459643	Сумма	2526,74
Счет	30	Счет	30	Счет	30	Счет	30	Счет	30	Счет	30

Корреляция

A	B	C	D	E	F	G
	Рентабельность	Уд.вес рабочих в ППП	Уд.вес покуп. изделий	Коэф.сменности оборуд.	Среднегод.числен. ППП	Среднегод. стоимость ОПФ
Рентабельность	1					
Уд.вес рабочих в ППП	0,243161909	1				
Уд.вес покуп. изделий	0,237273774	-0,045734366	1			
Коэф.сменности оборуд.	-0,024242284	0,391834065	0,056426172	1		
Среднегод.числен. ППП	0,229051998	0,136726532	0,478622234	0,23979545	1	
Среднегод. стоимость ОПФ	0,088461342	0,066823108	0,430824228	0,364810983	0,894308133	1

Ковариация

A	B	C	D	E	F	G
	Рентабельность	Уд.вес рабочих в ППП	Уд.вес покуп. изделий	Коэф.сменности оборуд.	Среднегод.числен. ППП	Среднегод. стоимость ОПФ
Рентабельность	32,31803122					
Уд.вес рабочих в ППП	0,064872333	0,002202333				
Уд.вес покуп. изделий	0,190248444	-0,000309333	0,019892889			
Коэф.сменности оборуд.	-0,014832	0,001978	0,000841333	0,011582667		
Среднегод.числен. ППП	12324,80078	60,7409	636,9608444	244,2691333	89587361,91	
Среднегод. стоимость ОПФ	46,41566978	0,266108667	5,808407556	3,623787333	781266,888	8610,861318

Дисперсия

A	B	C	D	E	F	G
Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*		
Рентабельность	30	405,13	13,50433333	33,43244609		
Уд.вес рабочих в ППП	30	22,23	0,741	0,002278276		
Уд.вес покуп. изделий	30	8,48	0,282666667	0,020578851		
Коэф.сменности оборуд.	30	39,66	1,322	0,011982069		
Среднегод.числен. ППП	30	459643	15321,43333	92676581,29		
Среднегод. стоимость ОПФ	30	2826,74	94,22466667	8812,604812		
Дисперсионный анализ						
Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	6962004277	5	1170400855	75,75501126	7,48899E-42	2,266052883
Внутри групп	2687877393	174	15447571,23			
Итого	8539881670	179				

ПРИЛОЖЕНИЕ 4

Матрица с исходными данными

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	№	x01	x12	x13	x14	x16	x25	x56	x67	to01	tн12	to12	tн13	to13	tн14	to14	tн16	to16	tн36	to36	tн46	to46	tн25	to25	tн56	to56	tн67	to67	tн78	to78	Знач	Вид	
2		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	огр	огр	
3		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
4	Ц/Ф	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	min	
5	1	1	1	1	1	1	1	1	1																						0	=	100
6	2									1																					0	=	1
7	3										-1	1																			0	=	8
8	4												-1	1																	0	=	5
9	5														-1	1															0	=	20
10	6																-1	1													0	=	40
11	7																		-1	1											0	=	0
12	8																				-1	1									0	=	0
13	9																						-1	1							0	=	40
14	10																								-1	1					0	=	4
15	11																										-1	1			0	=	1
16	12																												-1	1	0	=	0
17	13	0,05								1																					0	=	2
18	14		0,1								-1	1																			0	=	10
19	15			0,1									-1	1																	0	=	6
20	16				0,3										-1	1															0	=	30
21	17					0,4											-1	1													0	=	60
22	18						0,3																-1	1							0	=	55
23	19							0,2																		-1	1				0	=	5
24	20								0,05																		-1	1			0	=	2
25	21									-1	1																				0	X	0
26	22											-1											1								0	X	0
27	23									-1			1																		0	X	0
28	24													-1					1												0	X	0
29	25									-1					1																0	X	0
30	26															-1					1										0	X	0
31	27									-1							1														0	X	0
32	28																	-1									1				0	X	0
33	29																			-1							1				0	X	0
34	30																					-1					1				0	X	0
35	31																								-1	1					0	X	0
36	32																									-1	1				0	X	0
37	33																											-1	1		0	X	0

Результат оптимизации по критерию минимизации времени с вложением 100 ден. ед.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	№	x01	x12	x13	x14	x15	x25	x56	x67	to01	th12	to12	th13	to13	th14	to14	th16	to16	th36	to36	th46	to46	th25	to25	th56	to56	th67	to67	th78	to78	Значч	Вид	
2		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23	X24	X25	X26	X27	X28	X29	огр	огр	
3		0	20	0	0	20	50	5	0	2	2	10	2	8	6	36	2	54	8	8	40	40	10	50	50	54	54	56	56	56			
4	ЦФ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	56	min	
5	1	1	1	1	1	1	1	1	1																						95	<=	100
6	2									1																					2	>=	1
7	3										-1	1																			8	>=	8
8	4												-1	1																	6	>=	5
9	5														-1	1															30	>=	20
10	6																-1	1													52	>=	40
11	7																		-1	1											0	=	0
12	8																				-1	1									0	=	0
13	9																						-1	1							40	>=	40
14	10																								-1	1					4	>=	4
15	11																										-1	1			2	>=	1
16	12																												-1	1	0	=	0
17	13	0,05								1																					2	=	2
18	14		0,10								-1	1																			10	=	10
19	15			0,10									-1	1																	6	=	6
20	16				0,3										-1	1															30	=	30
21	17					0,4											-1	1													60	=	60
22	18						0,3																-1	1							55	=	55
23	19							0,20																	-1	1					5	=	5
24	20								0,05																		-1	1			2	=	2
25	21									-1	1																				0	>=	0
26	22											-1											1								0	>=	0
27	23									-1			1																		0	>=	0
28	24													-1					1												0	>=	0
29	25									-1					1																4	>=	0
30	26															-1					1										4	>=	0
31	27									-1							1														-0	>=	0
32	28																	-1									1				0	>=	0
33	29																			-1							1				46	>=	0
34	30																					-1					1				14	>=	0
35	31																							-1	1						0	>=	0
36	32																									-1	1				0	>=	0
37	33																											-1	1		0	>=	0

ПРИЛОЖЕНИЕ 6

Матрица исходных данных в режиме отображения формул

	Z	AA	AB	AC	AD	AE	AF	AG
1	to56	tн67	to67	tн78	to78	Знач	Вид	
2	X25	X26	X27	X28	X29	опр	опр	
3	0	0	0	0	0			
4	0	0	0	0	1	**=СУММПРОИЗВ(B3:AD3;B4:AD4)**	min	
5						=СУММПРОИЗВ($B$3:$AD$3;B5:AD5)	<=	100
6						=СУММПРОИЗВ($B$3:$AD$3;B6:AD6)	>=	1
7						=СУММПРОИЗВ($B$3:$AD$3;B7:AD7)	>=	8
8						=СУММПРОИЗВ($B$3:$AD$3;B8:AD8)	>=	5
9						=СУММПРОИЗВ($B$3:$AD$3;B9:AD9)	>=	20
10						=СУММПРОИЗВ($B$3:$AD$3;B10:AD10)	>=	40
11						=СУММПРОИЗВ($B$3:$AD$3;B11:AD11)	=	0
12						=СУММПРОИЗВ($B$3:$AD$3;B12:AD12)	=	0
13						=СУММПРОИЗВ($B$3:$AD$3;B13:AD13)	>=	40
14	1					=СУММПРОИЗВ($B$3:$AD$3;B14:AD14)	>=	4
15		-1	1			=СУММПРОИЗВ($B$3:$AD$3;B15:AD15)	>=	1
16				-1	1	=СУММПРОИЗВ($B$3:$AD$3;B16:AD16)	=	0
17						=СУММПРОИЗВ($B$3:$AD$3;B17:AD17)	=	2
18						=СУММПРОИЗВ($B$3:$AD$3;B18:AD18)	=	10
19						=СУММПРОИЗВ($B$3:$AD$3;B19:AD19)	=	6
20						=СУММПРОИЗВ($B$3:$AD$3;B20:AD20)	=	30
21						=СУММПРОИЗВ($B$3:$AD$3;B21:AD21)	=	60
22						=СУММПРОИЗВ($B$3:$AD$3;B22:AD22)	=	55
23	1					=СУММПРОИЗВ($B$3:$AD$3;B23:AD23)	=	5
24		-1	1			=СУММПРОИЗВ($B$3:$AD$3;B24:AD24)	=	2
25						=СУММПРОИЗВ($B$3:$AD$3;B25:AD25)	>=	0
26						=СУММПРОИЗВ($B$3:$AD$3;B26:AD26)	>=	0
27						=СУММПРОИЗВ($B$3:$AD$3;B27:AD27)	>=	0
28						=СУММПРОИЗВ($B$3:$AD$3;B28:AD28)	>=	0
29						=СУММПРОИЗВ($B$3:$AD$3;B29:AD29)	>=	0
30						=СУММПРОИЗВ($B$3:$AD$3;B30:AD30)	>=	0
31						=СУММПРОИЗВ($B$3:$AD$3;B31:AD31)	>=	0
32		1				=СУММПРОИЗВ($B$3:$AD$3;B32:AD32)	>=	0
33		1				=СУММПРОИЗВ($B$3:$AD$3;B33:AD33)	>=	0
34		1				=СУММПРОИЗВ($B$3:$AD$3;B34:AD34)	>=	0
35						=СУММПРОИЗВ($B$3:$AD$3;B35:AD35)	>=	0
36	-1	1				=СУММПРОИЗВ($B$3:$AD$3;B36:AD36)	>=	0
37			-1	1		=СУММПРОИЗВ($B$3:$AD$3;B37:AD37)	>=	0

Учебное издание

Шарстнев Владимир Леонидович,
Вардомацкая Елена Юрьевна

КОМПЬЮТЕРНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
ПАКЕТЫ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ МОДЕЛИРОВАНИЯ И АНАЛИЗА ЗАДАЧ ЭКОНОМИКИ

Редактор В.Е.Казаков
Технический редактор Е.П. Андрушкевич
Корректор И.П.Лабусова
Компьютерная верстка И.В. Соколов

Подписано к печати 13.03.07 . Формат 60x84 1/16. Бумага офсетная №1. Гарнитура «Таймс». Усл.-печ. листов 9,1 . Уч.-издат. листов 8,5 . Тираж 163 экз. Зак. № 128 .

Учреждение образования «Витебский государственный технологический университет» 210035, г. Витебск, Московский пр-т, 72.

Отпечатано на ризографе учреждения образования «Витебский государственный технологический университет»
Лицензия №02330/0133005 от 1 апреля 2004 г.