ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ ОБУЧАЮЩИХСЯ

ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.В.ОД.3 Математические методы и модели в прикладных научных исследованиях

Направление подготовки 36.06.01 «Ветеринария и зоотехния» (уровень подготовки кадров высшей квалификации по программе подготовки научно-педагогических кадров в аспирантуре)

Профиль подготовки: 06.02.10 –«Частная зоотехния, технология производства продуктов животноводства»

Квалификация (степень) выпускника Исследователь. Преподаватель-исследователь

Форма обучения очная

СОДЕРЖАНИЕ

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция 1 (2 ч.)

Тема: Программа курса. Общие подходы к построению программы исследований. Методология исследования.

1.1.1. Вопросы лекции:

1. Общенаучная и частнонаучная методология.

2. Методика научного исследования. Структура научного исследования.

3. Общелогические методы научного исследования. Методы эмпирического исследования

1.1.2 Краткое содержание вопросов:

Общенаучная и частнонаучная методология

Общенаучная методология представляет собой совокупность знаний о принципах и методах, применяемых в любой научной дисциплине. Она выступает своего рода «промежуточной методологией» между философией и фундаментальными теоретико-методологическими положениями специальных наук. К общенаучным относят такие понятия, как «система», «структура», «элемент», «функция» и т.д.

На основе общенаучных понятий и категорий формулируются соответствующие методы познания, которые обеспечивают оптимальное взаимодействие философии с конкретно-научным знанием и его методами.

Общенаучные методы разделяют на:

1) общелогические, применяемые в любом акте познания и на любом уровне. Это анализ и синтез, индукция и дедукция, обобщение, аналогия, абстрагирование;

2) методы эмпирического исследования, применяемые на эмпирическом уровне исследования (наблюдение, эксперимент, описание, измерение, сравнение);

3) методы теоретического исследования, применяемые на теоретическом уровне исследования (идеализация, формализация, аксиоматический, гипотетико-дедуктивный и т.д.);

4) методы систематизации научных знаний (типологизация, классификация).

Характерные черты общенаучных понятий и методов:

- соединение в их содержании элементов философских категорий и понятий ряда частных наук;

- возможность формализации и уточнения математическими средствами.

На уровне общенаучной методологии формируется общенаучная картина мира.

Частнонаучная методология представляет собой совокупность знаний о принципах и методах, применяемых в той или иной специальной научной дисциплине. В ее рамках формируются специальные научные картины мира. Каждая наука имеет свой специфический набор методологических средств. В то же время методы одних наук могут транслироваться в другие науки. Возникают междисциплинарные научные методы.

Методика научного исследования. Структура научного исследования

Главное внимание в рамках методологии науки направлено на научное исследование как вид деятельности, в котором находит свое воплощение применение различных научных методов.

Научное исследование – деятельность, направленная на получение истинного знания об объективной реальности.

Знания, применяемые на предметно-чувственном уровне некоторого научного исследования, составляют базу его методики. В эмпирическом исследовании методика обеспечивает сбор

и первичную обработку опытных данных, регулирует практику научно-исследовательской работы – экспериментально-производственную деятельность. Теоретическая работа тоже требует своей методики. Здесь ее предписания относятся к деятельности с объектами, выраженными в знаковой форме. Например, существуют методики различного рода вычислений, расшифровки текстов, проведения мысленных экспериментов и т.д. На современном этапе развития науки как на ее эмпирическом, так и на теоретическом уровне исключительно важная роль принадлежит компьютерной технике. Без нее немыслимы современный эксперимент, моделирование ситуаций, различные вычислительные процедуры.

Всякая методика создается на основе более высоких уровней знаний, но представляет собой совокупность узкоспециализированных установок, включающую в себя достаточно жесткие ограничения – инструкции, проекты, стандарты, технические условия и т.д. На уровне методики установки, существующие идеально, в мыслях человека, как бы смыкаются с практическими операциями, завершая образование метода. Без них метод представляет собой нечто умозрительное и не получает выхода во внешний мир. В свою очередь, практика исследования невозможна без управления со стороны идеальных установок. Хорошее владение методикой – показатель высокого профессионализма ученого.

Научное исследование содержит в своей структуре ряд элементов.

Объект исследования — фрагмент реальности, на который направлена познавательная деятельность субъекта, и который существует вне и независимо от сознания познающего субъекта.

Объекты исследования могут быть как материальными, так и нематериальными по своей природе. Их независимость от сознания заключается в том, что они существуют вне зависимости от того известно или неизвестно о них что-либо людям.

Предметом исследования является часть объекта, непосредственно задействованная в исследовании; это главные, наиболее существенные признаки объекта с точки зрения того или иного исследования.

Специфика предмета научного исследования заключается в том, что вначале он задается в общих, неопределенных чертах, предвосхищается и прогнозируется в незначительной степени. Окончательно он «вырисовывается» в конце исследования. Приступая к нему, ученый не может представить его в чертежах и расчетах. Что нужно «вырвать» из объекта и синтезировать в продукте исследования – об этом исследователь имеет поверхностное, одностороннее, не исчерпывающее знание. Поэтому формой фиксации предмета исследования является вопрос, проблема. Постепенно преобразуясь в продукт исследования, предмет обогащается и развивается за счет неизвестных вначале признаков и условий его существования.

Каждый из предметов создает свой понятийный аппарат, свои специфические методы исследования, свой язык. Цель исследования – идеальное, мысленное предвосхищение результата, ради которого предпринимаются научно-познавательные действия. Особенности предмета исследования прямо сказываются на его цели. Последняя, заключая в себе образ предмета исследования, отличается свойственной предмету неопределенностью в начале процесса исследования. Она конкретизируется по мере приближения к конечному результату.

Задачи исследования формулируют вопросы, на которые должен быть получен ответ для реализации целей исследования. Цели и задачи исследования образуют взаимосвязанные цепочки, в которых каждое звено служит средством удержания других звеньев. Конечная цель исследования может быть названа его общей задачей, а частные задачи, выступающие в качестве средств решения основной, можно назвать промежуточными целями, или целями второго порядка. Выделяют также основные и дополнительные задачи исследования: Основные задачи отвечают его целевой установке, дополнительные — ставятся для подготовки будущих исследований, проверки побочных (возможно, весьма актуальных), не связанных с данной проблемой гипотез, для решения каких-то методических вопросов и т.п.

Способы достижения цели: Если основная цель формулируется как теоретическая, то при разработке программы главное внимание уделяется изучению научной литературы по данному вопросу, четкой интерпретации исходных понятий, построению гипотетической общей

концепции предмета исследования, выделению научной проблемы и логическому анализу рабочих гипотез. Иная логика управляет действиями исследователя, если он ставит перед собой непосредственно практическую цель. Он начинает работу, исходя из специфики данного объекта и уяснения практических задач, подлежащих решению. Только после этого он обращается к литературе в поисках ответа на вопрос: имеется ли "типовое" решение возникших задач, т. е. специальная теория, относящаяся к предмету? Если "типового" решения нет, дальнейшая работа развертывается по схеме теоретического исследования. Если же такое решение имеется, гипотезы прикладного исследования строятся как различные варианты "прочтения" типовых решений применительно к конкретным условиям. Очень важно иметь в виду, что любое исследование, ориентированное на решение теоретических задач, можно продолжить как прикладное. На первом этапе мы получаем некоторое типовое решение проблемы, а затем переводим его в конкретные условия.

Также элементом структуры научного исследования выступают средства научно-познавательной деятельности. К ним относятся: - материальные средства; - теоретические объекты (идеальные конструкты); - методы исследования и другие идеальные регулятивы исследования: нормы, образцы, идеалы научной деятельности.

Средства научного поиска находятся в постоянном изменении и развитии. То, что некоторые из них успешно применяются на одном этапе развития науки, не является достаточным гарантом их согласования с новыми сферами реальности и потому требуют усовершенствования или замены.

Общелогические методы научного исследования. Методы эмпирического исследования. Методы теоретического исследования

Анализ – расчленение целостного предмета на составляющие части (признаки, свойства, отношения) с целью их всестороннего изучения.

Синтез – соединение ранее выделенных частей (сторон, признаков, свойств, отношений) предмета в единое целое.

Абстрагирование – мысленное отвлечение от ряда признаков, свойств и отношений изучаемого объекта при одновременном выделении для рассмотрения тех из них, которые интересуют исследователя. В результате появляются «абстрактные предметы», которыми являются как отдельно взятые понятия и категории, так и их системы.

Обобщение – установление общих свойств и признаков объектов. Общее – философская категория, отражающая сходные, повторяющиеся признаки, черты, которые принадлежат единичным явлениям или всем предметам данного класса.

Различают два вида общего: - абстрактно-общее (простая одинаковость, внешнее сходство, подобие ряда единичных предметов); - конкретно-общее (внутренняя, глубинная, повторяющаяся у группы сходных явлений основа – сущность).

В соответствии с этим выделяют два вида обобщений:

- выделение любых признаков и свойств объектов;
- выделение существенных признаков и свойств объектов.

По другому основанию обобщения разделяют на:

- индуктивные (от отдельных фактов и событий к их выражению в мыслях);
- логические (от одной мысли к другой, более общей).

Метод, противоположный обобщению – ограничение (переход от более общего понятия к менее общему).

Индукция – метод исследования, в котором общий вывод строится на основе частных посылок.

Дедукция – метод исследования, посредством которого из общих посылок следует заключение частного характера.

Аналогия – метод познания, при котором на основе сходства объектов в одних признаках заключают об их сходстве и в других признаках.

Моделирование – изучение объекта путем создания и исследования его копии (модели), замещающей оригинал с определенных сторон, интересующих познание.

Методы эмпирического исследования. На эмпирическом уровне применяются такие методы, как наблюдение, описание, сравнение, измерение, эксперимент.

Наблюдение – это систематическое и целенаправленное восприятие явлений, в ходе которого мы получаем знание о внешних сторонах, свойствах и отношениях изучаемых объектов. Наблюдение всегда носит не созерцательный, а активный, деятельный характер. Оно подчинено решению конкретной научной задачи и поэтому отличается целенаправленностью, избирательностью и систематичностью. Основные требования к научному наблюдению: однозначность замысла, наличие строго определенных средств (в технических науках – приборов), объективность результатов. Объективность обеспечивается возможностью контроля путем либо повторного наблюдения, либо применения других методов исследования, в частности, эксперимента. Обычно наблюдение включается в качестве составной части в процедуру эксперимента. Важным моментом наблюдения является интерпретация его результатов – расшифровка показаний приборов и т.д. Научное наблюдение всегда опосредуется теоретическим знанием, поскольку именно последнее определяет объект и предмет наблюдения, цель наблюдения и способ его реализации. В ходе наблюдения исследователь всегда руководствуется определенной идеей, концепцией или гипотезой. Он не просто регистрирует любые факты, а сознательно отбирает те из них, которые либо подтверждают, либо опровергают его идеи. При этом очень важно отобрать наиболее репрезентативную группу фактов в их взаимосвязи. Интерпретация наблюдения также всегда осуществляется с помощью определенных теоретических положений.

Осуществление развитых форм наблюдения предполагает использование особых средств – и в первую очередь приборов, разработка и воплощение которых также требует привлечения теоретических представлений науки. В общественных науках формой наблюдения является опрос; для формирования средств опроса (анкетирование, интервьюирование) также требует специальных теоретических знаний.

Описание – фиксация средствами естественного или искусственного языка результатов опыта (данных наблюдения или эксперимента) с помощью определенных систем обозначения, принятых в науке (схемы, графики, рисунки, таблицы, диаграммы и т.д.). В ходе описания проводится сравнение и измерение явлений.

Сравнение – метод, выявляющий сходство или различие объектов (либо ступеней развития одного и того же объекта), т.е. их тождество и различия. Но данный метод имеет смысл только в совокупности однородных предметов, образующих класс. Сравнение предметов в классе осуществляется по признакам, существенным для данного рассмотрения. При этом признаки, сравниваемые по одному признаку, могут быть несравнимы по другому.

Измерение – метод исследования, при котором устанавливается отношение одной величины к другой, служащей эталоном, стандартом. Наиболее широкое применение измерение находит в естественных и технических науках, но с 20 – 30-х годов XX в. оно входит в употребление и в социальных исследованиях. Измерение предполагает наличие: объекта, над которым проводится некоторая операция; свойства этого объекта, которое поддается восприятию, и величина которого устанавливается с помощью данной операции; инструмента, посредством которого эта операция производится. Общей целью любых измерений является получение числовых данных, позволяющих судить не столько о качестве, сколько о количестве некоторых состояний. При этом значение получаемой величины должно быть настолько близким к истинному, что для данной цели его можно использовать вместо истинного. Возможны погрешности результатов измерений (систематические и случайные). Различают прямые и косвенные процедуры измерения. К последним относятся измерения объектов, которые удалены от нас или непосредственно не воспринимаются. Значение измеряемой величины устанавливается при этом опосредованно. Косвенные измерения осуществимы тогда, когда известна общая зависимость между величинами, которая позволяет вывести искомый результат из уже известных величин.

Эксперимент – метод исследования, при помощи которого происходит активное и целенаправленное восприятие определенного объекта в контролируемых и управляемых условиях.

Основные особенности эксперимента:

1) активное отношение к объекту вплоть до его изменения и преобразования;

2) многократная воспроизводимость изучаемого объекта по желанию исследователя;

3) возможность обнаружения таких свойств явлений, которые не наблюдаются в естественных условиях;

4) возможность рассмотрения явления «в чистом виде» путем изоляции его от внешний влияний, или путем изменения условий эксперимента;

5) возможность контроля за «поведением» объекта и проверки результатов. Можно сказать, что эксперимент – идеализированный опыт.

Он дает возможность следить за ходом изменения явления, активно воздействовать на него, воссоздавать, если в этом есть необходимость, прежде чем сравнивать полученные результаты. Поэтому эксперимент является методом более сильным и действенным, чем наблюдение или измерение, где исследуемое явление остается неизменным. Это высшая форма эмпирического исследования. Эксперимент применяется либо для создания ситуации, позволяющей исследовать объект в чистом виде, либо для проверки уже существующих гипотез и теорий, либо для формулировки новых гипотез и теоретических представлений. Всякий эксперимент всегда направляется какой-либо теоретической идей, концепцией, гипотезой. Данные эксперимента, также как и наблюдения, всегда теоретически нагружены – от его постановки до интерпретации результатов.

Стадии проведения эксперимента:

1) планирование и построение (его цель, тип, средства и т.п.);

2) контроль;

3) интерпретация результатов.

Структура эксперимента:

1) объект исследования;

2) создание необходимых условий (материальные факторы воздействия на объект исследования, устранение нежелательных воздействий – помех);

3) методика проведения эксперимента;

4) гипотеза или теория, которую нужно проверить.

Как правило, экспериментирование связано с использованием более простых практических методов – наблюдений, сравнений и измерений. Поскольку эксперимент не проводится, как правило, без наблюдений и измерений, то он должен отвечать их методическим требованиям.

В зависимости от задач эксперимента выделяют исследовательские (задача – формирование новых научных теорий), проверочные эксперименты (проверка существующих гипотез и теорий), решающие (подтверждение одной и опровержение другой из соперничающих теорий).

В зависимости от характера объектов выделяют физические, химические, биологические, социальные и др. эксперименты.

Выделяют также качественные эксперименты, имеющие целью установить наличие или отсутствие предполагаемого явления, и измерительные эксперименты, выявляющие количественную определенность некоторого свойства.

Методы теоретического исследования. На теоретическом этапе используются мысленный эксперимент, идеализация, формализация, аксиоматический, гипотетико-дедуктивный методы, метод восхождения от абстрактного к конкретному, а также методы исторического и логического анализа.

Идеализация – метод исследования, состоящий в мысленном конструировании представления об объекте путем исключения условий, необходимых для его реального существования. По сути, идеализация представляет собой разновидность процедуры абстрагирования, конкретизированной с учетом потребностей теоретического исследования. Результатами такого

конструирования являются идеализированные объекты. Формирование идеализаций может идти разными путями:

- последовательно осуществляемое многоступеньчатое абстрагирование (так, получаются объекты математики – плоскость, прямая, точка и т.д.);

- вычленение и фиксация некоего свойства изучаемого объекта в отрыве от всех других (идеальные объекты естественных наук).

Идеализированные объекты гораздо проще реальных объектов, что позволяет применить к ним математические методы описания. Благодаря идеализации процессы рассматриваются в их наиболее чистом виде, без случайных привнесений извне, что открывает пути к выявлению законов, по которым эти процессы протекают. Идеализированный предмет в отличие от реального характеризуется не бесконечным, а вполне определенным числом свойств и потому исследователь получает возможность полного интеллектуального контроля над ним. Идеализированные предметы моделируют наиболее существенные отношения в реальных предметах.

Моделирование (метод, тесно связанный с идеализацией) - это метод исследования теоретических моделей, т.е. аналогов (схем, структур, знаковых систем) определенных фрагментов действительности, которые называются оригиналами. Исследователь, преобразуя эти аналоги и управляя ими, расширяет и углубляет знания об оригиналах. Моделирование – это метод опосредованного оперирования объектом, в ходе которого исследуется непосредственно не сам интересующий нас объект, а некоторая промежуточная система (естественная или искусственная), которая:

- находится в некотором объективном соответствии с познаваемым объектом (модель – это, прежде всего, то, с чем сравнивают - необходимо, чтобы между моделью и оригиналом было сходство в каких-то физических характеристиках, или в структуре, или в функциях);

- способна в ходе познания на известных этапах замещать в определенных случаях изучаемый объект (в процессе исследования временное замещение оригинала моделью и работа с нею позволяет во многих случаях не только обнаружить, но и предсказать его новые свойства);

- давать в процессе ее исследования в конечном счете информацию об интересующем нас объекте. Логической основой метода моделирования являются выводы по аналогии.

Существуют различные виды моделирования. Основные:

- Предметное (прямое) – моделирование, в ходе которого исследование ведется на модели, воспроизводящей определенные физические, геометрические и пр. характеристики оригинала. Предметное моделирование используется как практический метод познания.

- Знаковое моделирование (моделями служат схемы, чертежи, формулы, предложения естественного или искусственного языка и т.д.). Поскольку действия со знаками есть одновременно действия с некоторыми мыслями, постольку всякое знаковое моделирование по своей сути является моделированием мысленным.

В исторических исследованиях выделяют отражательно-измерительные модели («как было») и имитационно-прогностические («как могло быть»). Мысленный эксперимент – метод исследования, основанный на комбинации образов, материальная реализация которых невозможна. Данный метод формируется на основе идеализации и моделирования. Модель при этом оказывается воображаемым объектом, преобразуемым в соответствии с правилами, пригодными для данной ситуации. Недоступные практическому эксперименту состояния раскрываются с помощью его продолжения – мысленного эксперимента.

В последнее время для осуществления моделирования и проведения мысленного эксперимента все чаще применяется вычислительный эксперимент. Главное преимущество компьютера состоит в том, что с его помощью при исследовании весьма сложных систем удается глубоко проанализировать не только их наличные, но и возможные, в том числе будущие состояния. Сущность вычислительного эксперимента состоит в том, что проводится эксперимент над некоторой математической моделью объекта при помощи компьютера. По одним параметрам модели вычисляются другие ее характеристики и на этой основе делаются выводы о свойствах явлений, представленных математической моделью. Основные этапы вычислительного эксперимента:

1) построение математической модели изучаемого объекта в тех или иных условиях (как правило, она представлена системой уравнений высокого порядка);

2) определение вычислительного алгоритма решения базовой системы уравнений;

3) построение программы реализации поставленной задачи для ЭВМ. Вычислительный эксперимент на основе накопленного опыта математического моделирования, банка вычислительных алгоритмов и программного обеспечения позволяет быстро и эффективно решать задачи практически в любой области математизированного научного знания. Обращение к вычислительному эксперименту в ряде случаев позволяет резко снизить стоимость научных разработок и интенсифицировать процесс научного поиска, что обеспечивается многовариантностью выполняемых расчетов и простотой модификаций для имитации тех или иных условий эксперимента.

Формализация – метод исследования, в основе которого лежит отображение содержательного знания в знаково-символическом виде (формализованном языке). Последний создается для точного выражения мыслей с целью исключения возможности для неоднозначного понимания. При формализации рассуждения об объектах переносятся в плоскость оперирования со знаками (формулами), что связано с построением искусственных языков. Использование специальной символики позволяет устранить многозначность и неточность, образность слов естественного языка. В формализованных рассуждениях каждый символ строго однозначен. Формализация служит основой для процессов алгоритмизации и программирования вычислительных устройств, а тем самым и компьютеризации знания. Главное в процессе формализации состоит в том, что над формулами искусственных языков можно производить операции, получать из них новые формулы и соотношения. Тем самым операции с мыслями заменяются действиями со знаками и символами (границы метода). Метод формализации открывает возможности для использования более сложных методов теоретического исследования, например метода математической гипотезы, где в качестве гипотезы выступают некоторые уравнения, представляющие модификацию ранее известных и проверенных состояний. Изменяя последние, составляют новое уравнение, выражающее гипотезу, которая относится к новым явлениям. Часто исходная математическая формула заимствуется из смежной и даже не смежной области знания, в нее подставляются значения, иной природы, а затем проверяют, совпадение рассчитанного и реального поведения объекта. Разумеется, применимость этого метода ограничена теми дисциплинами, которые уже накопили, достаточно богатый математический арсенал.

Аксиоматический метод – способ построении научной теории, при котором за ее основу принимаются некоторые положения, не требующие специального доказательства (аксиомы или постулаты), из которых все остальные положения выводятся при помощи формально-логических доказательств. Совокупность аксиом и выведенных на их основе положений образует аксиоматически построенную теорию, включающую в себя абстрактные знаковые модели. Такая теория может быть использована для модельного представления не одного, а нескольких классов явлений, для характеристики не одной, а нескольких предметных областей. Для вывода положений из аксиом формулируются особые правила вывода – положения математической логики. Отыскание правил соотнесения аксиом формально построенной системы знания с определенной предметной областью называют интерпретацией. В современном естествознании примерами формальных аксиоматических теорий являются фундаментальные физические теории, что влечет за собой ряд специфических проблем их интерпретации и обоснования (особенно для теоретических построений неклассической и постнеклассической науки). В силу специфики аксиоматически построенных систем теоретического знания для их обоснования особое значение приобретают внутритеоретические критерии истинности: требование непротиворечивости и полноты теории и требование достаточных оснований для доказательства или опровержения любого положения, сформулированного в рамках такой теории. Данный метод широко применяется в математике, а также в тех естественных науках, где применяется метод формализации. (Ограниченность метода).

Гипотетико-дедуктивный метод – способ построения научной теории, в основе которого лежит создание системы взаимосвязанных гипотез, из которых затем путем дедуктивного

развертывания выводится система частных гипотез, подлежащая опытной проверке. Тем самым этот метод основан на дедукции (выведении) заключений из гипотез и других посылок, истинное значение которых неизвестно. А это значит, что заключение, полученное на основе данного метода, неизбежно будет иметь вероятностный характер. Структура гипотетико-дедуктивного метода:

1) выдвижение гипотезы о причинах и закономерностях данных явлений с помощью разнообразных логических приемов;

2) оценка основательности гипотез и выбор из их множества наиболее вероятной;

3) выведение из гипотезы дедуктивным путем следствий с уточнением ее содержания;

4) экспериментальная проверка выведенных из гипотезы следствий.

Тут гипотеза или получает экспериментальное подтверждение или опровергается. Однако подтверждение отдельных следствий не гарантирует ее истинности или ложности в целом. Лучшая по результатам проверки гипотеза переходит в теорию.

Метод восхождения от абстрактного к конкретному – метод, заключающийся в том, что первоначально находится исходная абстракция (главная связь (отношение) изучаемого объекта), а затем, шаг за шагом, через последовательные этапы углубления и расширения познания, прослеживается, как она видоизменяется в различных условиях, открываются новые связи, устанавливаются их взаимодействия и, таким образом, отображается во всей полноте сущность изучаемого объекта.

Метод исторического и логического анализа. Исторический метод требует описания фактической истории объекта во всем разнообразии его существования.

Логический метод – это мысленная реконструкция истории объекта, очищенная от всего случайного, несущественного и сосредоточенная на выявлении сущности. Единство логического и исторического анализа. Логические процедуры обоснования научных знаний

Все конкретные методы, как эмпирические, так и теоретические, сопровождаются проведением логических процедур. Эффективность эмпирических и теоретических методов находится в прямой зависимости от того, насколько правильно с точки зрения логики строятся соответствующие научные рассуждения.

Обоснование – логическая процедура, связанная с оценкой некоторого продукта познания в качестве компонента системы научного знания с точки зрения его соответствия функциям, целям и задачам этой системы.

Основные виды обоснования:

Доказательство – логическая процедура, при которой выражение с неизвестным пока значением выводится из высказываний, истинность которых уже установлена. Это позволяет исключить всякие сомнения и признать истинность данного выражения.

Структура доказательства:

- тезис (выражение, истинность, которого устанавливается);
- доводы, аргументы (высказывания, с помощью которых устанавливается истинность тезиса);
- добавочные допущения (выражения вспомогательного характера, вводимые в структуру доказательства и устраняемые при переходе к окончательному результату);
- демонстрация (логическая форма данной процедуры).

Типичный пример доказательства – любое математическое рассуждение, по результатам которого принимается некоторая новая теорема. В нем эта теорема выступает в качестве тезиса, ранее доказанные теоремы и аксиомы – в качестве аргументов, демонстрация представляет собой форму дедукции.

Виды доказательств:

- прямое (тезис непосредственно вытекает из доводов);
- косвенное (тезис доказывается косвенным путем):
- апагогическое (доказательство от противного – установление ложности антитезиса: допускается, что антитезис истинен, и из него выводятся следствия, если хотя бы одно из

полученных следствий вступает в противоречие с наличными истинными суждениями, то следствие признается ложным, а вслед за ним и сам антитезис – признается истинность тезиса);

- разделительное (истинность тезиса устанавливается путем исключения всех противостоящих ему альтернатив).

1.2 Лекция 2 (2 ч.)

Тема: Математическая модель и этапы ее построения. Математические методы планирования эксперимента.

1.2.1. Вопросы лекции:

1. Математическая модель и этапы ее построения
2. Планирование эксперимента. Основы математического планирования эксперимента

1.2.2 Краткое содержание вопросов:

1. **Математическая модель и этапы ее построения**

При всем многообразии содержания конкретных работ по моделированию систем каждое моделирование требует последовательного выполнения следующих шести этапов:
- постановка задачи;
- построение математической модели;
- составление программы для компьютера;
- оценка адекватности модели;
- планирование эксперимента;
- интерпретация результатов моделирования.

Рассмотрим каждый этап отдельно.

Постановка задачи. Как и всякое исследование, компьютерное моделирование должно начинаться с постановки задачи моделирования, т.е. с ясного изложения целей моделирования и ограничений, которые необходимо учитывать при построении моделей. Цели обычно формируются в виде либо вопросов, на которые надо ответить; либо гипотез, которые надо проверить; либо воздействий, которые надо оценить . Например, компьютерное моделирование можно использовать для разрешения следующих вопросов: как повлияет предполагаемый алгоритм опроса датчиков на функционирование автоматизированной системы управления технологическим процессом сложной установки или каково влияние конкретной процедуры оперативного планирования на производственные затраты? Кроме того, надо не только поставить вопросы перед моделированием, но и сформулировать критерии оценки возможных ответов на них.

Целью моделирования может быть также проверка одной или нескольких гипотез относительно поведения некоторой сложной системы. Как повлияет изменение автобусного маршрута на загруженность машин? Не приведет ли к нарушению экологического равновесия в заповеднике искусственная миграция некоторых видов животных? В каждом случае проверяемая гипотеза, а также и критерии решающие «принять» или «отвергнуть» ее, должны быть сформулированы явно.

Наконец, компьютерное моделирование может быть предпринято, чтобы оценить воздействие некоторой переменной управления на входные или зависимые переменные, описывающие поведение системы. Например, необходимо оценить влияние процентного содержания кислорода в дутьевом воздухе на содержание SO_2 в отходящих газах металлургических заводов.

Для определения ограничений задачи моделирования необходимо выявить характеристики системы, подлежащей изучению. Первым шагом на этом пути является установление граничных условий, т.е. того, что является или не является частью изучаемой системы. Далее

определяются существенные параметры и переменные системы, которые характеризуют объект изучения и позволяют установить основные ограничения задачи моделирования.

Построение модели. Под математической моделью будем понимать совокупность соотношений, которые связывают во времени характеристики процесса, протекающего в системе с ее параметрами, входными сигналами и начальными условиями. Разнородность назначения элементов сложных систем, функционирование в условиях воздействия случайных факторов приводят к разнообразию математических схем, применяемых для описания сложных систем и их элементов. Среди математических схем, используемых при исследовании сложных систем, особое место занимают дифференциальные и разностные уравнения, марковские процессы, системы массового обслуживания, динамические системы, агрегативные системы, вероятностные автоматы.

Эти схемы можно назвать типовыми математическими схемами, поскольку они широко используются при исследовании сложных систем. Наличие разработанных математических методов исследования этих схем значительно повышает их ценность при использовании в качестве моделей элементов сложных систем. Однако не следует упускать из виду, что любая математическая модель никогда не бывает адекватной изучаемому процессу, а отражает лишь основные его черты с учетом задач, стоящих перед исследователем. В связи с этим возникает вопрос о сложности математической модели . С одной стороны, можно утверждать, что реальные системы крайне сложны, и поэтому математические модели, претендующие на описание поведения этих систем, по необходимости должны быть достаточно сложными. Но это верно лишь до некоторой степени, так как строить модели, настолько сложные, что реализация их потребует непомерно больших затрат времени, не имеет смысла. Надо строить такие математические модели, которые обеспечивали бы достаточно точное описание поведения системы и не требовали бы при этом слишком много времени на программирование и вычисления.

При компьютерном моделировании сложных систем математическая модель преобразуется в моделирующий алгоритм, с помощью которого имитируются элементарные явления, составляющие исследуемый процесс. При этом в алгоритме сохраняются логическая структура, последовательность протекания во времени, характер и состав информации о состояниях процесса.

Оценка адекватности модели. В сложных системах, к которым относятся объекты компьютерного моделирования, любая модель лишь частично отражает реальный процесс. Модель считается хорошей, если, несмотря на свою неполноту, может точно предсказать влияние изменений в системе на общую эффективность системы. Поэтому необходима проверка степени соответствия (адекватности) модели и реального процесса. Существуют три подхода к проверке на адекватность. Применяя первый из них, мы должны убедиться, что модель верна в первом приближении. Например, следует поставить вопрос: не будет ли модель давать абсурдные ответы, если ее параметры будут принимать предельные значения.

Второй подход к оценке адекватности модели состоит в проверке исходных предположений. Например, какие параметры и переменные модели можно считать существенными и охвачены ли моделью все существенные параметры объекта. Характер переменной, т.е. является ли она существенной или нет, обычно определяется влиянием этой переменной на критерий эффективности системы. Для выяснения степени охвата в модели всех существенных параметров объекта используются современные методы статистического анализа, такие как анализ дисперсии критериев эффективности.

Третий подход к оценке адекватности модели состоит в проверке преобразований информации от входа к выходу. Этот подход основан на использовании статистических выборок для оценки средних значений и дисперсии, дисперсионного анализа, регрессионного анализа, факторного анализа, спектрального анализа, автокорреляции, проверок с помощью критериев согласия и др.

Наконец, всегда следует помнить о потребителе информации, получаемой с помощью компьютерного моделирования. Нельзя оправдать разработку компьютерной модели, если она не приносит пользу потребителю.

Приняв во внимание все это, сформулируем конкретные критерии, которым должна удовлетворять хорошая модель . Такая модель должна быть:

- простой и понятной пользователю;
- удобной в управлении;
- надежной в смысле гарантии от абсурдных решений;
- полной с точки зрения возможностей решения главных задач;
- адаптивной, позволяющей легко переходить к другим модификациям или обновлять данные.

2. Планирование эксперимента. Основы математического планирования эксперимента

После завершения этапа оценки пригодности модели необходимо осуществить прогон (реализацию) модели с целью получения желаемой информации. Результаты моделирования, полученные при воспроизведении единственной реализации процесса, описываемого моделью в силу действия случайных факторов, не могут объективно характеризовать процесс функционирования модели. Поэтому искомые величины при исследовании процессов методом компьютерного моделирования получают как средние значения по данным большого числа реализаций процесса. Исключение составляют так называемые эргодические процессы, для которых искомые величины можно получить усреднением по времени результатов единственной реализации процесса.

Если количество реализаций N достаточно велико, то, в силу закона больших чисел, получаемые оценки становятся устойчивыми и могут служить приближенными значениями искомых случайных величин.

После экспериментирования с моделью надо обработать его результаты. Для сложных систем и большого количества реализаций, воспроизводимых при моделировании, объем информации о состоянии системы может быть настолько значительным, что запоминание ее в памяти компьютера, обработка и последующий анализ оказываются практически невозможными, или, во всяком случае, очень трудоемкими. Поэтому необходимо таким образом организовать фиксацию и обработку результатов моделирования, чтобы оценки искомых величин формировались постепенно по ходу моделирования, без специального запоминания всей информации о состояниях системы.

Если при моделировании некоторой системы учитываются случайные факторы, то и среди результатов моделирования присутствуют случайные величины. В такой ситуации в качестве оценок для искомых величин используются средние значения, дисперсии и другие вероятностные характеристики соответствующих случайных величин, полученные по результатам многократного моделирования.

Рассмотренные выше этапы моделирования в принципе необходимы и при любых других исследованиях, например, при системном анализе и исследовании операций. Однако реализация отдельных этапов при компьютерном моделировании имеет принципиальные отличия от других методов исследования систем. К таким этапам относятся этапы построения математической модели, планирования эксперимента и интерпретации результатов моделирования. Знакомство с этими особенностями компьютерного моделирования начнем с этапа построения математической модели, в котором такой особенностью является необходимость разработки моделирующих алгоритмов.

Основы математического планирования эксперимента

Планирование эксперимента - это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

Задачи, для решения которых может использоваться планирование эксперимента, чрезвычайно разнообразны. К ним относятся: поиск оптимальных условий, построение интерполяционных формул, выбор существенных факторов, оценка и уточнение констант теоретических

моделей, выбор наиболее приемлемых из некоторого множества гипотез о механизме явлений, исследование диаграмм состав - свойство и т.д. Одной из главных задач эксперимента является получение и проверка математической модели объекта, описывающей в количественной форме взаимосвязи между входными и выходными параметрами объекта.

Общая последовательность при планировании эксперимента с целью получения математической модели такова:

1) Определение объекта исследований, параметров оптимизации, факторов, интервалов и уровней варьирования.

2) Выбор зависимости (линейная, квадратичная и т.д.) и полинома для построения модели.

3) Составление матрицы планирования для проведения эксперимента.

4) Проведение эксперимента.

5) Математическая обработка полученных данных: поиск коэффициентов регрессии и составление математической модели.

6) Проверка адекватности модели.

Объект исследования. В теории планирования эксперимента объект исследований принято представлять в виде «черного ящика» (рисунок 1). Стрелки справа изображают численные характеристики целей исследования. Мы их обозначаем буквой***y***и называем*параметрами оптимизации.* В литературе встречаются другие названия: критерий оптимизации, целевая функция, выход «черного ящика» и т.д.

Для проведения эксперимента необходимо иметь возможность воздействовать на наведение «черного ящика». Все способы такого воздействия мы обозначаем буквой ***x***и называем*факторами.* Их называют также входами «черного ящика».

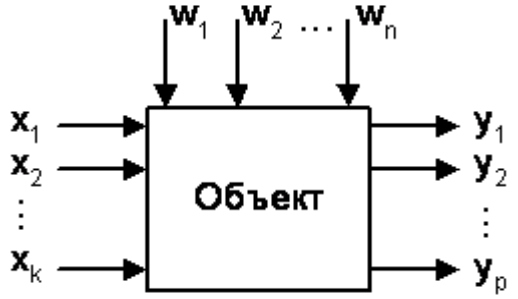
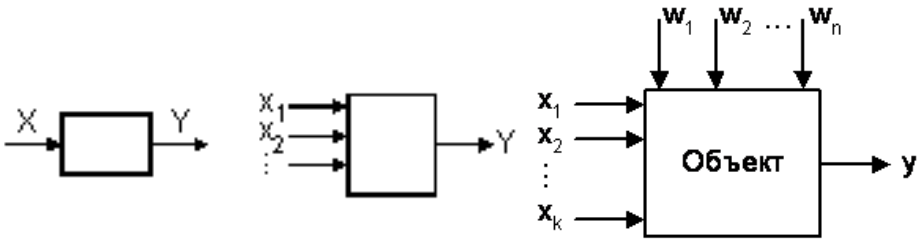


Рисунок 1. Информационная модель процесса

Различные виды экспериментов схематично представлены на рисунке 2.

Однофакторный пассивный эксперимент проводится путем выполнения *n*пар измерений в дискретные моменты времени единственного входного параметра *x* и соответствующих значений выходного параметра *y* (рисунок 2,а). Аналитическая зависимость между этими параметрами вследствие случайного характера возмущающих воздействий рассматривается в виде зависимости математического ожидания *y* от значения *x*, носящей название регрессионной. Целью однофакторного пассивного эксперимента является построение *регрессионной модели*-установление зависимости *y = f(x).*



а б в

Рисунок 2. Виды экспериментов

Многофакторный пассивный эксперимент проводится при контроле значений нескольких входных параметров x_i (рисунок 2,б) и его целью является установление зависимости выходного параметра от двух или более переменных $y=F(x_1, x_2, ...)$.

Полный факторный эксперимент предполагает возможность управлять объектом по одному или нескольким независимым каналам (рисунок 2,в).

В общем случае, схема эксперимента может быть представлена в виде, представленном на рисунке 1. В схеме используются следующие группы параметров:

1. *управляющие* (входные x_i)
2. *параметры состояния* (выходные Y)
3. *возмущающие воздействия* (W_i)

При многофакторном и полном факторном эксперименте выходных параметров может быть несколько.

Под ***параметром оптимизации***(критерий оптимизации) понимают характеристику цели, заданную количественно. Параметр оптимизации является реакцией (откликом) на воздействие факторов, которые определяют поведение выбранной системы.

Он должен быть *количественным,* задаваться числом. Множество значений, которые может принимать параметр оптимизации, называется областью его определения. Количественная оценка параметра оптимизации на практике не всегда возможна. В таких случаях пользуются приемом, называемым ранжированием. При этом параметрам оптимизации присваиваются оценки - ранги по заранее выбранной шкале: двухбалльной, пятибалльной и т.д.

Параметр оптимизации должен соответствовать следующим требованиям:

*1)*должен быть *количественным.*

2) выражаться *одним числом.*

3) должен обладать *однозначностью* в статистическом смысле. Заданному набору значений факторов должно соответствовать одно значение параметра оптимизации, при этом обратное неверно: одному и тому же значению параметра могут соответствовать разные наборы значений факторов.

*4)*должен давать *возможность действительно эффективной оценки функционирования системы.* Представление об объекте не остается постоянным в ходе исследования. Оно меняется по мере накопления информации и в зависимости от достигнутых результатов. Это приводит к последовательному подходу при выборе параметра оптимизации. Так, например, на первых стадиях исследования технологических процессов в качестве параметра оптимизации часто используется выход продукта. Однако в дальнейшем, когда возможность повышения выхода исчерпана, начинают интересоваться такими параметрами, как себестоимость, чистота продукта и т.д.

5) требование универсальности или *полноты.* Под универсальностью параметра оптимизации понимают его способность всесторонне охарактеризовать объект. В частности, технологические параметры недостаточно универсальны: они не учитывают экономику. Универсальностью обладают, например, обобщенные параметры оптимизации, которые строятся как функции от нескольких частных параметров.

6) желательно, чтобы параметр оптимизации имел *физический смысл, был простым и легко вычисляемым.*

После выбора объекта исследования и параметра оптимизации нужно рассмотреть все ***факторы***, которые могут влиять на процесс. Если какой-либо существенный фактор окажется неучтенным и принимал произвольные значения, не контролируемые экспериментатором, то это значительно увеличит ошибку опыта. При поддержании этого фактора на определенном уровне может быть получено ложное представление об оптимуме, т.к. нет гарантии, что полученный уровень является оптимальным.

С другой стороны большое число факторов увеличивает число опытов и размерность

Фактором называется измеряемая переменная величина, принимающая в некоторый момент времени определенное значение и влияющая на объект исследования. В практических

задачах области определения факторов имеют ограничения, которые носят либо принципиальный, либо технический характер.

Факторы разделяются на количественные и качественные.

К количественным относятся те факторы, которые можно измерять, взвешивать и т.д.

Качественные факторы - это различные вещества, технологические способы, приборы, исполнители и т.п.

Хотя качественным факторам не соответствует числовая шкала, но при планировании эксперимента к ним применяют условную порядковую шкалу в соответствии с уровнями, т.е. производится кодирование.

Факторы должны быть *управляемыми*, это значит, что выбранное нужное значение фактора можно поддерживать постоянным в течение всего опыта. Планировать эксперимент можно только в том случае, если уровни факторов подчиняются воле экспериментатора. Например, экспериментальная установка смонтирована на открытой площадке. Здесь температурой воздуха мы не можем управлять, ее можно только контролировать, и потому при выполнении опытов температуру, как фактор, мы не можем учитывать.

Точность замеров факторов должна быть возможно более высокой. Степень точности определяется диапазоном изменения факторов. В длительных процессах, измеряемых многими часами, минуты можно не учитывать, а в быстрых процессах приходится учитывать доли секунды.

Факторы должны быть *однозначны*. Трудно управлять фактором, который является функцией других факторов. Но в планировании могут участвовать другие факторы, такие, как соотношения между компонентами, их логарифмы и т.п.

При планировании эксперимента одновременно изменяют несколько факторов, поэтому необходимо знать требования к совокупности факторов. Прежде всего выдвигается требование *совместимости*. Совместимость факторов означает, что все их комбинации осуществимы и безопасны. Несовместимость факторов наблюдается на границах областей их определения. Избавиться от нее можно сокращением областей. Положение усложняется, если несовместимость проявляется внутри областей определения. Одно из возможных решений - разбиение на подобласти и решение двух отдельных задач.

При планировании эксперимента важна *независимость* факторов, т.е. возможность установления фактора на любом уровне вне зависимости от уровней других факторов. Если это условие невыполнимо, то невозможно планировать эксперимент.

Фактор считается заданным, если указаны его название и область определения. В выбранной области определения он может иметь несколько значений, которые соответствуют числу его различных состояний. Выбранные для эксперимента количественные или качественные состояния фактора носят название ***уровней фактора***. Минимальное число уровней, обычно применяемое на первой стадии работы, равно 2. Это верхний и нижний уровни, обозначаемые в кодированных координатах через +1 и -1. Но такое число уровней недостаточно для построения моделей второго порядка (ведь фактор принимает только два значения, а через две точки можно провести множество линий различной кривизны).

Выбор уровней варьирования может осуществляться следующим образом. Предположим, в некоторой задаче фактор (температура) мог изменяться от 140 до 180°С. Естественно, за нулевой уровень было принято среднее значение фактора, соответствующее 160°С. Тогда при трех уровнях варьирования значение фактора на верхнем уровне (+1) будет равно 180°С, а на нижнем 140°С. Интервал варьирования будет равен 20°С.

При решении задачи будем использовать ***математические модели исследования***. Под математической моделью мы понимаем уравнение, связывающее параметр оптимизации с факторами. Это уравнение в общем виде можно записать так:

$$y=f(x_1,x_2,\ldots,x_k).$$

где символ $f(\ldots)$, как обычно в математике, заменяет слова: «функция от». Такая функция называется *функцией отклика.* Наглядное, удобное воспринимаемое представление о функции отклика дает ее геометрический аналог - поверхность отклика (рис. 3).

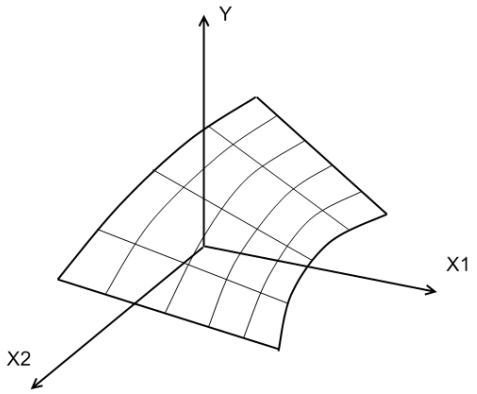


Рис. 3. Поверхность отклика

Наиболее часто в качестве моделей применяются приведенные ниже полиномы.

Полином первой степени:

$$y = в_o + \sum_1^k в_i x_i + \sum_1^k в_{ij} x_i x_j$$

Полином второй степени:

$$y = в_o + \sum_1^k в_i x_i + \sum_1^k в_{ij} x_i x_j + \sum_1^k в_{ii} x_i^2 .$$

Полиномы третьей степени:

$$y = в_o + \sum_1^k в_i x_i + \sum_1^k в_{ij} x_i x_j + \sum_1^k в_{iij} x_i^2 x_j +$$

$$+ \sum_1^k в_{ijj} x_i x_j^2 + \sum_1^k в_{iii} x_i^3 .$$

Здесь в этих уравнениях:

y - значения критерия; $в_i$- линейные коэффициенты регрессии;

$в_{ij}$ - коэффициенты двойного взаимодействия; x_i - кодированные значения факторов.

Модель должна быть *адекватной,* т.е. с достаточной точностью описывать изменение реального процесса. *Проверка адекватности модели* выполняется при помощи специальных статистических методов.

После определения факторов, их уровней и интервалов варьирования, параметров оптимизации и построения информационной модели необходимо заполнить матрицу планирования, по которой в дальнейшем будет проводиться эксперимент.

Число возможных опытов определяют по выражению

$$N=p^k,$$

где N - число опытов; р - число уровней; к - число факторов.

Примеры матриц планирования для 2-х и 3-х факторов на 2-х уровнях варьирования представлены в таблицах 1 и 2.

Таблица 1

Матрица проведения эксперимента 2^2

Номер опыта	Кодовое обозначение		Натуральные значения		У
	x_1	x_2			
1	+1	+1			y_1
2	-1	+1			y_2
3	+1	-1			y_3
4	-1	-1			y_4

Таблица 2

Матрица проведения эксперимента 2^3

Номер опыта	Кодовое обозначение			Натуральные значения			У
	x_1	x_2	x_3				
1	+1	+1	+1				y_1
2	-1	+1	+1				y_2
3	+1	-1	-1				y_3
4	-1	-1	-1				y_4
5	+1	+1	-1				y_5
6	-1	+1	-1				y_6
7	+1	-1	+1				y_7
8	-1	-1	+1				y_8

Расчет коэффициентов регрессии

Построив матрицу планирования осуществляют эксперимент. Получив экспериментальные данные рассчитывают значения коэффициентов регрессии.

Их можно рассчитать следующим образом. Значение свободного члена *($в_о$)* берут как среднее арифметическое всех значений параметра оптимизации в матрице:

$$в_о = \frac{\sum_1^N y_u}{N},$$

где y_u - значения параметра оптимизации вu-м опыте; N - число опытов в матрице.

Линейные коэффициенты регрессии рассчитывают по формуле

$$в_i = \frac{\sum_1^N x_{iu} y_u}{\sum_1^N x_{iu}^2} = \frac{\sum_1^N x_{iu} y_u}{N},$$

где x_{iu} - кодированное значение фактора x_i в u-м опыте.

$$в_{ij} = \frac{\sum_1^N x_{iu} x_{ju} y_u}{\sum_1^N x_{iu}^2} = \frac{\sum_1^N x_{iu} x_{ju} y_u}{N}.$$

Коэффициенты регрессии, характеризующие парное взаимодействие факторов, находят по формуле:

Полное число всех возможных коэффициентов регрессии, включая $в_0$, линейные коэффициенты и коэффициенты взаимодействий всех порядков, равно числу опытов полного факторного эксперимента.

Для планирования эксперимента и математической обработки полученных данных (в т.ч. расчета коэффициентов регрессии и получения математической модели) рекомендуется использование различных прикладных пакетов программ (например, TATGRAPHICS,STATISTIKAи др.)

1.3 Лекция 3 -5 (6 ч.)

Тема: Основы статистической обработки результатов наблюдения. Элементы теории ошибок. Обоснование числа измерений. Использование надстроек Microsoft Excel. Проверка статистических гипотез. Уровень значимости. Критерии. Примеры. Оценка чувствительности критерия при проверке значимости различий. Двухвыборочный t - тест в Excel. Оценка тесноты связи. Корреляция. Дисперсионный анализ с использованием таблиц Excel. Анализ таблиц сопряженности.

1.3.1. Вопросы лекции:

1. Статистическая обработка результатов измерений
2. Элементы теории ошибок.
3. Обоснование числа измерений.
4. Использование надстроек Microsoft Excel.
5. Проверка статистических гипотез. Уровень значимости. Критерии. Примеры. Оценка чувствительности критерия при проверке значимости различий. Двухвыборочный t - тест в Excel
6. Оценка тесноты связи. Корреляция. Дисперсионный анализ с использованием таблиц Excel. Анализ таблиц сопряженности.

1.3.2 Краткое содержание вопросов:

Статистическая обработка результатов измерений

Статистическая обработка результатов измерений – обработка измерительной информации с целью получения достоверных данных. Разнообразие задач, решаемых с помощью измерений, определяет и разнообразие видов статистической обработки их результатов.

Задача статистической обработки результатов многократных измерений заключается в нахождении оценки измеряемой величины и доверительного интервала, в котором находится истинное значение.

Статистическая обработка используется для повышения точности измерений с многократными наблюдениями, а также определения статистических характеристик случайной погрешности.

Для прямых однократных измерений статистическая обработка менее сложна и громоздка, что значительно упрощает оценку погрешностей.

Статистическую обработку результатов косвенных измерений производят, как правило, методами, основанными на раздельной обработке аргументов и их погрешностей, и методом линеаризации.

Наиболее распространенные совместные измерения обрабатываются разными статистическими методами. Среди них широко известен и часто применяется метод наименьших квадратов.

Прямые измерения с многократными наблюдениями

Необходимость в многократных наблюдениях некоторой физической величины возникает при наличии в процессе измерений значительных случайных погрешностей. При этом задача обработки состоит в том, чтобы по результатам наблюдений определить наилучшую (оптимальную) оценку измеряемой величины и интервал, в котором она находится с заданной вероятностью. Данная задача может быть решена способом статистической обработки результатов наблюдений, основанным на гипотезе о распределении погрешностей результатов по нормальному закону.

Порядок такой обработки должен соответствовать государственному стандарту и рекомендациям по метрологии.

Итак, рассмотрим группу из n независимых результатов наблюдений случайной величины x, подчиняющейся нормальному распределению. Оценка рассеяния единичных результатов наблюдений в группе относительно их среднего значения вычисляется по формуле:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m_x)^2}{n-1}}$$

Поскольку число наблюдений в группе, на основании результатов которых выполнено вычисление среднего арифметического, ограничено, то, повторив заново серию наблюдений этой же величины, мы получили бы новое значение среднего арифметического. Повторив многократно наблюдения и вычисляя каждый раз их среднее арифметическое значение, принимаемое за результат наблюдений (измерений), обнаружим рассеяние среднего арифметического значения.

Характеристикой этого рассеяния является средний квадрат отклонения среднего арифметического:

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m_x)^2}{n(n-1)}} = \frac{\sigma}{\sqrt{n}}$$

Среднее квадратичное отклонение среднего арифметического используется для оценки погрешности результата измерений с многократными наблюдениями.

Теория показывает, что если рассеяние результатов наблюдения в группе подчиняется нормальному закону, то и их среднее арифметическое тоже подчиняется нормальному закону распределения при достаточно большом числе наблюдений (n>50). Отсюда при одинаковой доверительной вероятности доверительный интервал среднего арифметического в $\sqrt{n}$ ùже, чем доверительный интервал результата наблюдений. Теоретически случайную погрешность результата измерений можно было бы свести к 0, однако практически это невозможно, да и не имеет смысла, так как при уменьшении значения случайной погрешности определяющим в суммарной погрешности становится значение неисключенных остатков систематической погрешности.

При нормальном законе распределения плотности вероятностей результатов наблюдений и небольшом числе измерений среднее арифметическое подчиняется закону распределения Стьюдента с тем же средним арифметическим m_x. Особенностью этого распределения является то, что доверительный интервал с уменьшением числа наблюдений расширяемся по сравнению с нормальным законом распределения при этой же доверительной вероятности. В формуле для оценки доверительных границ случайной погрешности это отражается введением коэффициента tq вместо t: $\Delta x(P) = t\sigma = tq\sigma$

Коэффициент распределения Стьюдента зависит от числа наблюдений и выбранной доверительной вероятности и находится по таблице. Например, для n=4 и $P_Д$ =0,95 tq=3,182; n=5 при $P_Д$ =0,95 tq=2,776; для n=10 tq=2,262; n=15tq=2,145 при той же $P_Д$ =0,05.

Правила обработки результатов измерения с многократными наблюдениями учитывают следующие факторы:

обрабатывается группа из n наблюдений (то есть группа ограничена);

результаты наблюдений могут содержать систематическую погрешность;

в группе наблюдений могут встречаться грубые погрешности;

распределение случайных погрешностей может отличаться от нормального.

Обработка результатов наблюдения производится в следующей последовательности:

1) Исключить известные систематические погрешности из результатов наблюдения (введением поправки);

2)Вычислить среднее арифметическое исправленных результатов наблюдений, принимаемое за результат наблюдений: $$X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

3)Вычислить оценку среднего квадратичного отклонения результата наблюдения:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - X)^2}{n-1}}$$

Определив σ, целесообразно проверить наличие в группе наблюдений грубых погрешностей, помня, что при нормальном законе распределения ни одна случайная погрешность $x_i - X$, с вероятностью, практически равной 1, не может выйти за пределы $\pm 3\sigma$. Наблюдения, содержащие грубые погрешности, исключают из группы и заново повторяют вычисления X и σ.

4) Вычислить оценку среднего квадратичного отклонения среднего арифметического $S_{\bar{x}}$ по формуле:

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - m_x)^2}{n(n-1)}} = \frac{\sigma}{\sqrt{n}}$$

5) Проверить гипотезу о том, что результаты измерений принадлежат нормальному распределению.

Приближенно о характере распределения можно судить, построив гистограмму. Существуют и строгие методы проверки гипотез о том или ином характере распределения случайной величины с использованием специальных критериев. Об этом подробнее можно узнать в книге П. В. Новицкий, И. А. Зограф «Оценка погрешностей результатов измерений».

При числе наблюдений n<15 принадлежность их к нормальному распределению не проверяют, а доверительные границы случайной погрешности результата определяют лишь в том случае, если достоверно известно, что результаты наблюдений принадлежат нормальному закону.

6) Вычислить доверительные границы ε случайной погрешности результата измерения при заданной вероятности P: $\varepsilon = t_g S_{\bar{x}}$, где t_g - коэффициенты Стьюдента

7) Вычислить границы суммарной неисключенной систематической погрешности (НСП) результата измерения.

НСП результата измерений образуется из неисключенных остатков измерений, погрешностей, поправок и т. д.

При суммировании эти составляющие рассматриваются как случайные величины. При отсутствии данных о виде распределений НСП, их распределения принимают за равномерные.

При равномерном распределении НСП границы НСП вычисляют по формуле:

$$\theta = k\sqrt{\sum_{i=1}^{m}\theta_i^2},$$

где θ_i - граница i-той НСП, k – коэффициент, определяемый принятой доверительной вероятностью (при $P_Д$ =0,95 k =1,1); m – число неисключенных составляющих систематической погрешности.

Доверительную вероятность для вычисления границ НСП принимают той же, что при вычислении границ случайной погрешности результата измерений.

8) Вычислить доверительные границы погрешности результата измерения.

Анализ соотношения между НСП и случайной погрешностью показывает, что если $\frac{\theta}{S_{\bar{x}}} < 0{,}8$, то НСП можно пренебречь и принять границы погрешности результата $\Delta = \pm\varepsilon$.

Если $\frac{\theta}{S_{\bar{x}}} > 8$, то случайной погрешностью можно пренебречь и принять $\Delta = \pm\theta$.

Если оба неравенства не выполнены, вычисляют среднее квадратичное отклонение результата как сумму НСП и случайной погрешности в следующем виде:

$$S_\Sigma = \sqrt{\sum_{i=1}^{m} \frac{\theta_i^2}{3} + S_{\bar{x}}^2},$$

а границы погрешности результата измерения в этом случае вычисляют по формуле: $\Delta = \pm k S_\Sigma$, где k – коэффициент, определяемый как

$$k = \frac{\varepsilon + \theta}{S_{\bar{x}} + \sqrt{\sum_{i=1}^{m} \frac{\theta_i^2}{3}}}$$

9) Записать результат измерения в регламентированной стандартом форме:

а) при симметричном доверительном интервале погрешности результата измерения $x \pm \Delta, P$, где x – результат измерения;

б) при отсутствии данных о виде функции распределения составляющих погрешности результата или при необходимости использования данных для дальнейшей обработки результатов, результат представляют в форме: X, S_x, n, θ

Из условия, что при $\frac{\theta}{S_{\bar{x}}} > 8$ случайной погрешностью можно пренебречь, следует оценка максимального целесообразного числа наблюдений в эксперименте:

$\frac{\theta}{S_{\bar{x}}} = 8 \Rightarrow n_{\max} = (\frac{8\sigma}{\theta})^2$, где $\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - X)^2}{n-1}}$

Теория ошибок

Для оценки истинности данных эксперимента следует рассмотреть возможные причины ошибок и степень их влияния на измеряемую величину.

Приборные погрешности. Эта погрешность равна той доле шкалы прибора, до которой с уверенностью можно производить отсчет, что определяется конструкцией и ценой деления шкалы прибора.

Допустим, мы измеряем объем колбы мерным цилиндром с ценой деления 1 мл и получаем значение 242 мл. С какой точностью оно получено? По цене деления: 242±0,5 мл. Пусть мы проделали серию измерений и рассчитали среднее значение 242,837569. Можем ли мы утверждать, что определили объем колбы с точностью до десятимиллионной? Конечно нет. Точность нашего определения не может превосходить точности отдельного измерения, и мы вправе лишь записать: 242,8±0,5 мл. При этом последняя цифра является недостоверной, а все отброшенные цифры были незначащими.

Это надо учитывать при записи результатов наблюдений и рассчитанных данных. Объем воды 10 мл, отмеренный мерной пипеткой с ценой деления 0,01 мл правильно указать 10,00, а мерным цилиндром просто 10. Число значащих цифр в результате вычислений не может быть больше, чем в наименее точном исходном числе.

Приборные ошибки можно уменьшить, используя более совершенные приборы, но это связано, как правило, с большими материальными затратами. Кроме того, наряду с приборными

ошибками можно выделить еще две группы ошибок, которые возникают в процессе эксперимента: систематические и случайные.

Систематические ошибки вызываются неправильной конструкцией приборов, их неисправностью, недостаточно продуманной методикой эксперимента, наличием неучтенных факторов, влияющих на измеряемую величину Например: школьник решил изучить влияние освещенности на рост побегов. этой целью он разместил одно растение на подоконнике, другое в темном углу кабинета. Но при этом он не учел, что температура воздуха вблизи окна 12 градусов, а в темном углу – 24 градуса. О том, что систематические ошибки могут иметь очень серьезные последствия свидетельствует и роман Жюль Верна, в котором отважный 15-летний капитан и его команда из-за систематической ошибки в показаниях компаса попали вместо одного материка на другой. Но характерной особенностью систематических ошибок является их принципиальная устранимость или возможность коррекции.

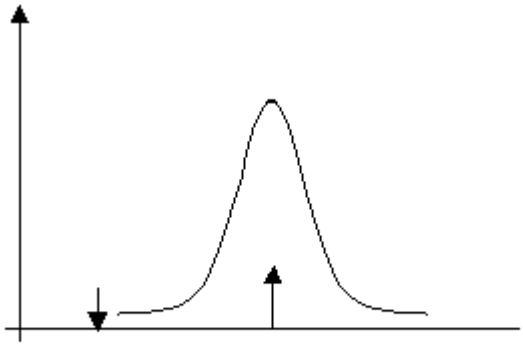


Рис. 1. Нормальное распределение

Случайные ошибки устранить нельзя, а также нельзя вывести никакой формулы для исправления полученного результата. В тоже время, влияние случайных ошибок может быть уменьшено проведением повторных измерений и статистической обработкой полученных данных.

Как показано на большом числе экспериментальных данных распределение результатов измерения некоторой постоянной величины описывается некоторым математическим выражением, которое называется «нормальным распределением» или гауссовым распределением», «распределением вероятности» и т.п. Графически оно выражается следующей кривой (рис 1):

Из графика видно, что существует вероятность, пусть и очень маленькая, что наше единичное измерение покажет результат, сколь угодно далеко отстоящий от истинного значения. Выходом из положения является проведение серии измерений. Если на разброс данных действительно влияет случай, то в результате нескольких измерений мы скорее всего получим следующее (рис 2):

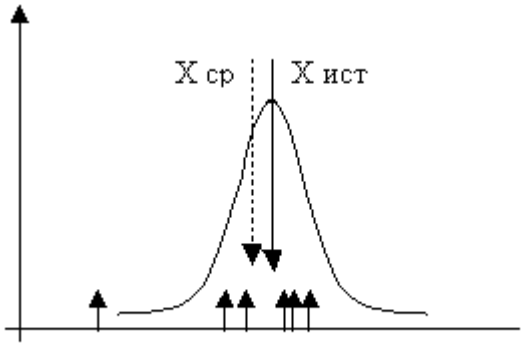


Рис. 2. Серия измерений

Будет ли рассчитанное среднее значение нескольких измерений совпадать с истинным? Как правило – нет. Но по теории вероятности, чем больше сделано измерений, тем ближе найденное среднее значение к истинному. На языке математики это можно записать так:

$$\lim_{n\to\infty} \overline{x}_{ср} = x_{ист}$$

Но с бесконечностью у всех дело обстоит неважно. Поэтому на практике мы имеем дело не со всеми возможными результатами измерений, а с некоторой выборкой из этого бесконечного множества. Сколько же реально следует делать измерений? Наверное, до тех пор, пока полученное среднее значение не будет отличаться от истинного меньше чем точность отдельного измерения.

Рис. 3. Доверительный интервал

Следовательно, когда наше среднее значение (рис. 2) отличается от истинного меньше чем погрешность измерений, дальнейшее увеличение числа опытов бессмысленно. Однако на практике мы не знаем истинного значения! Значит, получив среднее по результатам серии опытов, мы должны определить, какова вероятность того, что истинное значение

находится внутри заданного интервала ошибки. Или каков тот доверительный интервал, в который с заданной надежностью попадет истинное значение (рис 3).

Рассмотрим некоторый условный эксперимент, где в серии измерений получены некоторые значения величины X (см. табл. 1). Рассчитаем среднее значение и, чтобы оценить разброс данных найдем величины DX = X – Xср

Таблица 1. Данные измерений и их обработка

№	X	X ср	ΔX	ΔX^2	s^2	s
1	130		-	182,	420	20,5
2	162	143,5 ≈ 144	18,5	342,		
3	160		16,5	272,	$s^2_{ср}$	$s_{ср}$
4	122		-	462,	105	10,2

Ясно, что величины DX как-то характеризуют разброс данных. На практике для усредненной характеристики разброса серии измерений используется дисперсия выборки:

$$s^2 = \frac{\Sigma(\Delta x)^2}{n-1} = \frac{\Sigma(x-\bar{x})^2}{n-1} = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$$

и среднеквадратичное или стандартное отклонение выборки:

$$s = \sqrt{s^2} = \sqrt{\frac{\Sigma(\Delta x)^2}{n-1}} = \sqrt{\frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}}$$

Последнее показывает, что каждое измерение в данной серии (в данной выборке) отличается от другого в среднем на ± s.

Понятно, что каждое отдельное значение оказывает влияние на средний результат. Но это влияние тем меньше, чем больше измерений в нашей выборке. Поэтому дисперсия и стандартное отклонение среднего значения, будет определяться по формулам:

$$s^2_{ср} = \frac{\Sigma(\Delta x)^2}{n(n-1)}; \quad s_{ср} = \sqrt{\frac{\Sigma(\Delta x)^2}{n(n-1)}}$$

Можем ли мы теперь определить вероятность того, что истинное значение попадет в указанный интервал среднего? Или наоборот, рассчитать тот доверительный интервал в который истинное значение попадет с заданной вероятностью (95%)? Поскольку кривая на наших графиках это распределение вероятностей, то площадь под кривой, попадающая в указанный интервал и будет равна этой вероятности (доля площади, в процентах). А площади математики научились рассчитывать хорошо, знать бы только уравнение этой кривой.

И здесь мы сталкиваемся еще с одной сложностью. Кривая, которая описывает распределение вероятности для выборки, для ограниченного числа измерений, уже не будет кривой нормального распределения. Ее форма будет зависеть не только от дисперсии (разброса данных) но и от степени свободы для выборки (от числа независимых измерений) (рис 4):

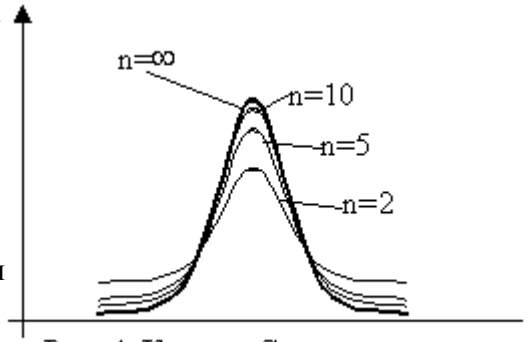


Рис. 4. Кривые Стьюдента

Уравнения этих кривых впервые были предложены в 1908 году английским математиком и химиком Госсетом, который опубликовал их под псевдонимом Student (студент), откуда пошло хорошо известные термины «коэффициент Стьюдента» и аналогичные. Коэффициенты Стьюдента получены на основе обсчета этих кривых для разных степеней свободы (f = n-1) и уровней надежности (P) и сведены в специальные таблицы. Для получения доверительного интервала необходимо

умножить уже найденное стандартное отклонение среднего на соответствующий коэффициент Стьюдента. ДИ = scp*tf, P

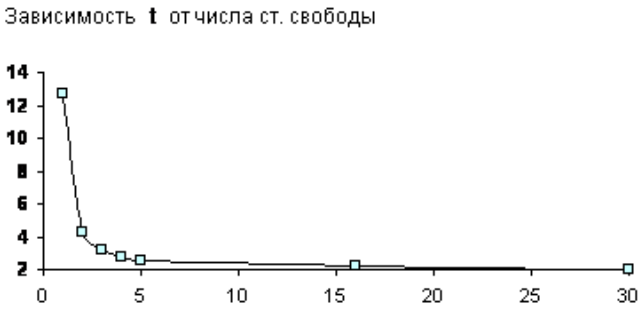
Проанализируем, как меняется доверительный интервал при изменении требований к надежности результата и числа измерений в серии. Данные в таблице 2 показывают, что чем больше требование к надежности, тем больше будет коэффициент Стьюдента и, следовательно, доверительный интервал. В большинстве случаев, приемлемым считают значение P=95%

Таблица 2. Коэффициент Стьюдента для различных уровней надежности.

P	0,9	0,95	0,99	0,999
t5, P	2,02	2,57	4,03	6,87

Таблица 3. Коэффициент Стьюдента для различных степеней свободы.

f= n-1	1	2	3	4	5	16	30
tf, 0,95	12,7	4,3	3,18	2,78	2,57	2,23	2,04



Из таблицы 3 и графика видно, что чем больше число измерений, тем меньше коэффициент и доверительный интервал для данного уровня надежности. Особенно значительное падение происходит при переходе от степени свободы 1 (два измерения) к 2 (три измерения). Отсюда следует, что имеет смысл ставить не менее трех параллельных опытов, проводить не менее трех измерений.

Окончательно для измеряемой величины X получаем значение Xсред±scp*tf,P. В нашем случае получаем: f=3; t=3,18; ДИ = 3,18*10,2 = 32,6; X = 143,5 ±32,6

Как правило, значение доверительного интервала округляется до одной значащей цифры, а значение измеряемой величины – в соответствии с округлением доверительного интервала. Поэтому для нашей серии окончательно имеем: X = 140 ±30

Найденная нами погрешность является абсолютной погрешностью и ничего не говорит еще о точности измерений. Она свидетельствует о точности измерений только в сравнении с измеряемой величиной. Отсюда представление об относительной ошибке:

$$\varepsilon_{отн} = \frac{\Delta x_{абс}}{x_{ср}} = \frac{32,6}{143,5} = 0,227 = 23\%$$

Исследуемая величина рассчитывается в этом случае с помощью математических формул по другим величинам, которые были измерены непосредственно. В этом случае для расчета ошибок можно использовать соотношения, приведенные в таблице 4.

Таблица 4. Формулы для расчета абсолютных и относительных ошибок.

Формула	Абсолютная	Относительная
x = a ± b	Dx = Da+Db	e = (Da+Db) /(a±b)
x = a* b; x = a* k	Dx = bDa+aDb; Dx = kDa	e = Da/a+Db/b + e b
x = a / b	Dx = (bDa+aDb) / b2	e = Da/a+Db/b + e b
x = a*k; (x = a / k)	Dx = Da*k; (Dx = Da/k)	e = ea
x = a2	Dx = 2aDa	e = 2Da/a = 2ea
x = Öa	Dx = Da/(2Öa)	e = Da/2a = ea/2

Из таблицы видно, что относительная ошибка и точность определения не изменяются при умножении (делении) на некоторый постоянный коэффициент. Особенно сильно относительная ошибка может возрасти при вычитании близких величин, так как при этом абсолютные ошибки суммируются, а значение X может уменьшиться на порядки.

Пусть например, нам необходимо определить объем проволочки. Если диаметр проволочки измерен с погрешностью 0,01 мм (микрометром) и равен 4 мм, то относительная погрешность составит 0,25% (приборная). Если длину проволочки (200 мм) мы измерим линейкой с погрешностью 0,5 мм, то относительная погрешность также составит 0,25%. Объем можно рассчитать по формуле: V=(pd2/4)*L. Посмотрим, как будут меняться ошибки по мере проведения расчетов (табл. 5):

Таблица 5. Расчет абсолютных и относительных ошибок

.

Величи-	Значе-	Абсолютная	Относительная
d2	16	Dx = 2*4*0,01=0,08	e = 0,5%
pd2 *)	50,27	Dx = 0,08*3,14+0,0016*16 =0,28	e = 0,55%
pd2/4	12,57	Dx = 0,28/4 = 0,07	e = 0,55%
(pd2/4)*	2513	Dx = 12,57*0,5+200*0,07=20	e = 0,8%

*) Если мы возьмем привычное p=3,14, то Dp=0,0016 то ep = 0,05%, но если используем более точное значение, то Dp и ep можно будет пренебречь

Окончательный результат V=2510±20 (мм3) e =0,8%. Чтобы повысить точность косвенного определения, нужно в первую очередь повышать точность измерения той величины, которая вносит больший вклад в ошибку (в данном случае – точность измерения диаметра проволочки).

План проведения измерений:

1. Знакомство с методикой, подготовка прибора, оценка приборной погрешности **δ.** Оценка возможных причин систематических ошибок, их исключение.
2. Проведение серии измерений. Если получены совпадающие результаты, можно считать что случайная ошибка равна 0, **ΔX** = **δ**. Переходим к пункту 7.
3. Исключение промахов – результатов значительно отличающихся по своей величине от остальных.
4. Расчет среднего значения $\mathbf{X_{ср}}$, и стандартного отклонение среднего значения $\boldsymbol{s_{ср}}$
5. Задание значения уровня надежности **P**, определение коэффициента Стьюдента **t** и нахождение доверительного интервала **ДИ= t*$s_{ср}$**
6. Сравнение случайной и приборной погрешности, при этом возможны варианты:

- **ДИ << δ,** можно считать, что **ΔX = δ**, повысить точность измерения можно, применив более точный прибор
- **ДИ >> δ,** можно считать, что **ΔX = ДИ**, повысить точность можно, уменьшая случайную ошибку, повышая число измерений в серии, снижая требования к надежности.
- **ДИ ≈ δ,** в этом случае расчитываем ошибку по формуле **ΔX** = $\sqrt{\mathbf{ДИ}^2 + \boldsymbol{\delta}^2}$

7. Записывается окончательный результат $\mathbf{X = X_{ср} \pm \Delta X}$.

Оценивается относительная ошибка измерения $\boldsymbol{\varepsilon = \Delta X / X_{ср}}$

Если проводится несколько однотипных измерений (один прибор, исследователь, порядок измеряемой величины, условия) то подобную работу можно проводить один раз. В дальнейшем можно считать **ΔX** постоянной и ограничиться минимальным числом измерений (два-три измерения должны отличаться не более, чем на **ΔX**)

Для косвенных измерений необходимо провести обработку данных измерения каждой величины. При этом желательно использовать приборы, имеющие близкие относительные

погрешности и задавать одинаковую надежность для расчета доверительного интервала. На основании полученных значений **Δa, Δb,** определяется **ΔX** для результирующей величины (см табл. 4). Для повышения точности надо совершенствовать измерение той величины, вклад ошибки которой в **ΔX** наиболее существенен.

Изучение зависимостей.

Частым вариантом экспериментальной работы является измерение различных величин с целью установления зависимостей. Характер этих зависимостей может быть различен: линейный, квадратичный, экспоненциальный, логарифмический, гиперболический. Для выявления зависимостей широко используется построение графиков.

Частым вариантом экспериментальной работы является измерение различных величин с целью установления зависимостей. Характер этих зависимостей может быть различен: линейный, квадратичный, экспоненциальный, логарифмический, гиперболический. Для выявления зависимостей широко используется построение графиков.

При построении графиков вручную важно правильно выбрать оси, величины, масштаб, шкалы. Следует предупредить школьников, что шкалы должны иметь равномерный характер, нежелательна как слишком детальная, так и слишком грубая их разметка. Точки должны заполнять всю площадь графика, их расположение в одном углу, или «прижатыми» к одной из осей, говорит о неправильно выбранном масштабе и затрудняет определение характера зависимости. При проведении линии по точкам надо использовать теоретические представление о характере зависимости: является она непрерывной или прерывистой, возможно ли ее прохождение через начало координат, отрицательные значения, максимумы и минимумы.

Наиболее легко проводится и анализируется прямая линия. Поэтому часто при изучении более сложных зависимостей часто используется линеаризация зависимостей, которая достигается подходящей заменой переменных. Например:

Зависимость $\alpha = \sqrt{\frac{K}{C}} = \frac{\sqrt{K}}{\sqrt{C}}$. Вводя новую переменную $x = \frac{1}{\sqrt{C}}$ **и** $b = \sqrt{K}$, получаем уравнение a = bx, которое будет изображаться на графике прямой линией. Наклон этой прямой позволяет рассчитать константу диссоциации.

Разумеется и в этом случае полученные в эксперименте данные включают в себя различные ошибки, и точки редко лежат строго на прямой. Возникает вопрос, как с наибольшей точностью провести прямую по экспериментальным точкам, каковы ошибки в определении параметров.

Математическая статистика показывает, что наилучшим приближением будет такая линия, для которой дисперсия (разброс) точек относительно ее будет минимальным. А дисперсия определяется как средний квадрат отклонений наблюдаемого положения точки от рассчитанного:

$$s^2 = \frac{\Sigma(\Delta y)^2}{n} = \frac{\Sigma(y - y_{рас})^2}{n}$$

Отсюда название этого метода – метод наименьших квадратов. Задавая условие, чтобы величина s2 принимала минимальное значение, получают формулы для коэффициентов a и b в уравнении прямой y = a + bx:

$$b = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}; \quad a = \frac{\Sigma y - b\Sigma x}{n}$$

и формулы для расчета соответствующих ошибок

$$s_b^2 = \frac{n\Sigma y^2 - (\Sigma y)^2}{(n-2)(n\Sigma x^2 - (\Sigma x)^2)}; \quad \Delta b = \pm t_{P,f} s_b$$

Обоснование числа измерений

Для проведения опытов с заданной точностью и достоверностью крайне важно знать то количество измерений, при котором экспериментатор уверен в положительном исходе. В связи с этим одной из первоочередных задач при статических методах оценки является установление минимального, но достаточного числа измерений для данных условий. Задача сводится к установлению минимального объема выборки (числа измерений) $N_{\min}$ при заданных значениях доверительного интервала 2μ и доверительной вероятности. При выполнении измерений крайне важно знать их точность: $\Delta = \sigma_o / \bar{x},$

где σ_o - среднеарифметическое значение среднеквадратичного отклонения σ, равное $\sigma_o / \sqrt{n}.$

Значение σ_o часто называют *средней ошибкой.* Доверительный интервал ошибки измерения Δ определяется аналогично для измерений $\mu = t\sigma_o$. С помощью *t* легко определить доверительную вероятность ошибки измерений из таблиц.

В исследованиях часто по заданной точности Δ и доверительной вероятности измерения определяют минимальное количество измерений, гарантирующих требуемые значения Δ и $p_д$.

Можно получить

$\mu = \sigma \arg \varphi(p_д) = \sigma_o / \sqrt{n}t.$ При $N_{\min} = n$ получаем

$$N_{\min} = \sigma^2 t^2 / \sigma_o^2 = k_в^2 t^2 / \Delta^2,$$

здесь $k_в$ - коэффициент вариации (изменчивости), %; Δ - точность измерений, %.

Для определения $N_{\min}$ может быть принята такая последовательность вычислений:
1) проводится предварительный эксперимент с количеством измерений *n,* к составляет в зависимости от трудоемкости опыта от 20 до 50;

2) вычисляется среднеквадратичное отклонение σ;

3) в соответствии с поставленными задачами эксперимента устанавливается требуемая точность измерений Δ, которая не должна превышать точности прибора;
4) устанавливается нормированное отклонение *t,* значение которого обычно задается (зависит также от точности метода);

5) определяют $N_{\min}$ и тогда в дальнейшем в процессе эксперимента число измерений не должно быть меньше $N_{\min}$.

Критерий Стьюдента (Госсета).

Оценки измерений с помощью σ и $\sigma_{ш}$ по приведенным методам справедливы при *n* > 30. Стоит сказать, что для нахождения границы доверительного интервала при малых значениях применяют метод, предложенный английским математиком В.С.Госсетом (псевдоним Стьюдент). Кривые распределения Стьюдента в случае $n \to \infty$ (практически при n >20) переходят в кривые нормального распределения (рис.1).

1 - $n \to \infty$; 2 - $n = 10$; 3 - $n = 2$

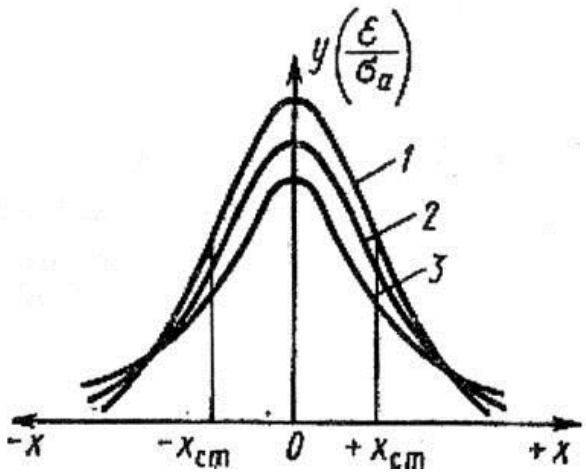


Рис.1. Кривые распределения Стьюдента для различных значений:

Для малой выборки доверительный интервал $\mu_{cm} = \sigma_o \alpha_{cm}$,

где α_{cm} – коэффициент Стьюдента, принимаемый по табл.10.2 в зависимости от значения доверительной вероятности $p_д$.

Зная μ_{cm}, можно вычислить действительное значение изучаемой величины для малой выборки $x_д = \bar{x} \pm \mu_{cm}$.

Возможна и иная постановка задачи. По *n* известных измерений малой выборки крайне важно определить доверительную вероятность $p_д$ при условии, что погрешность среднего значения не выйдет за пределы $\pm \mu_{cm}$. Задачу решают в такой последовательности: вначале вычисляется среднее значение $\bar{x}, \sigma_o$ и $\alpha_{cm} = \mu_{cm} / \sigma_o$. С помощью величины α_{cm}, известного *n* и табл. определяют доверительную вероятность.

В процессе обработки экспериментальных данных следует исключать грубые ошибки ряда. Появление этих ошибок вполне вероятно, а наличие их ощутимо влияет на результат измерении. При этом прежде чем исключить то или иное измерение, крайне важно убедиться, что это действительно грубая ошибка, а не отклонение вследствие статистического разброса. Известно несколько методов определения грубых ошибок статистического ряда. Наиболее простым способом исключения из ряда резко выделяющегося измерения является правило трех сигм: разброс случайных величин от среднего значения не должен превышать $x_{\max,\min} = \bar{x} \pm 3\sigma$.

Более достоверными являются методы, базируемые на использовании доверительного интервала. Пусть имеется статистический ряд малой выборки, подчиняющийся закону нормального распредел□ения. При наличии грубых ошибок критерии их появления вычисляются по формулам

$$\beta_1 = (x_{\max} - \bar{x}) / \sigma\sqrt{(n-1)/n}; \quad \beta_2 = (\bar{x} - x_{\min}) / \sigma\sqrt{(n-1)/n},$$

где $x_{\max}, x_{\min}$ – наибольшее и наименьшее значения из *n* измерений.

В табл. приведены в зависимости от доверительной вероятности максимальные значения $\beta_{\max}$, возникающие вследствие статистического разброса. В случае если $\beta_1 > \beta_{\max}$, то значение $x_{\max}$ крайне важно исключить из статистического ряда как грубую погрешность. При $\beta_2 < \beta_{\max}$ исключается величина $x_{\min}$. После исключения грубых ошибок определяют новые значения x и σ из $(n-1)$ или $(n-2)$ измерений.

Таблица: Критерий появления грубых ошибок

n	β_{max} при $p_д$			n	β_{max} при $p_д$		
	0,90	0,95	0,99		0,90	0,95	0,99
3	1,41	1,41	1,41	15	2,33	2,49	2,80
4	1,64	1,69	1,72	16	2,35	2,52	2,84
5	1,79	1,87	1,96	17	2,38	2,55	2,87
6	1,89	2,00	2,13	18	2,40	2,58	2,90
7	1,97	2,09	2,26	19	2,43	2,60	2,93
8	2,04	2,17	2,37	20	2,45	2,62	2,96
9	2,10	2,24	2,46	25	2,54	2,72	3,07
10	2,15	2,29	2,54	30	2,61	2,79	3,16
11	2,19	2,34	2,61	35	2,67	2,85	3,22
12	2,23	2,39	2,66	40	2,72	2,90	3,28
13	2,26	2,43	2,71	45	2,76	2,95	3,33
14	2,30	2,46	2,76	50	2,80	2,99	3,37

Использование надстроек Microsoft Excel.

1. Установка надстройки «Пакет анализа»

При создании новой или открытии существующей книги Microsoft Excel появится окно активного рабочего листа. Для того чтобы отыскать команду вызова надстройки *Пакет анализа,* **необходимо воспользоваться меню *Сервис* (рис. 1.1). Здесь возможны две ситуации, в которых нужно действовать следующим образом:**

1. В меню *Сервис* присутствует команда *Анализ данных*. Это идеальный случай - достаточно щелкнуть указателем мыши по данной команде, чтобы попасть в окно надстройки.

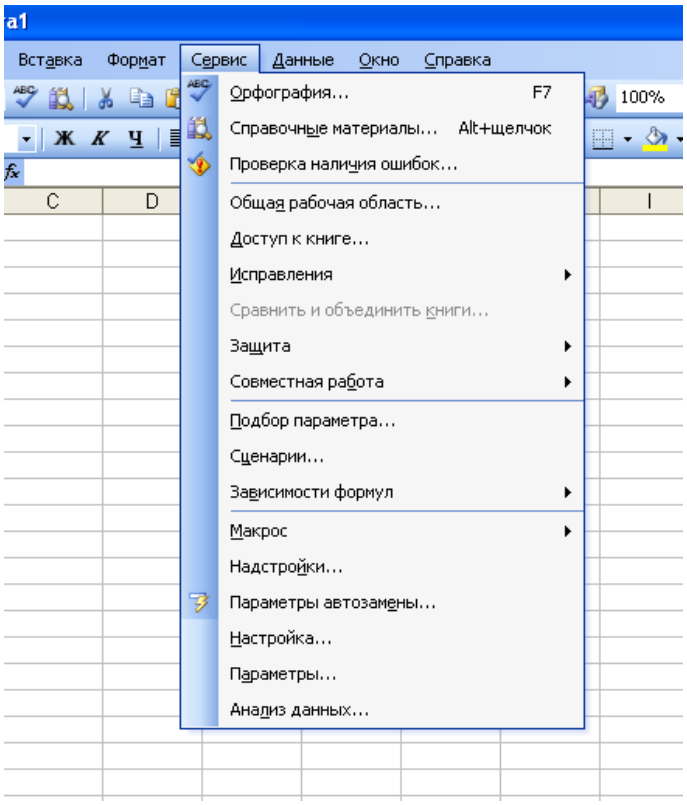


Рис. 1

2. В меню *Сервис* отсутствует команда *Анализ данных*. В этом случае необходимо в том же меню выполнить команду *Надстройки*.... Раскроется одноименное окно (рис. 2) со списком доступных надстроек. В этом списке нужно найти элемент *Пакет анали-*

за, поставить рядом с ним «галку» и щелкнуть по кнопке ОК. После этого в меню *Сервис* появится команда *Анализ данных*. Эта ситуация наиболее типична, так как надстройка *Пакет анализа* инсталлируется при стандартной установке.

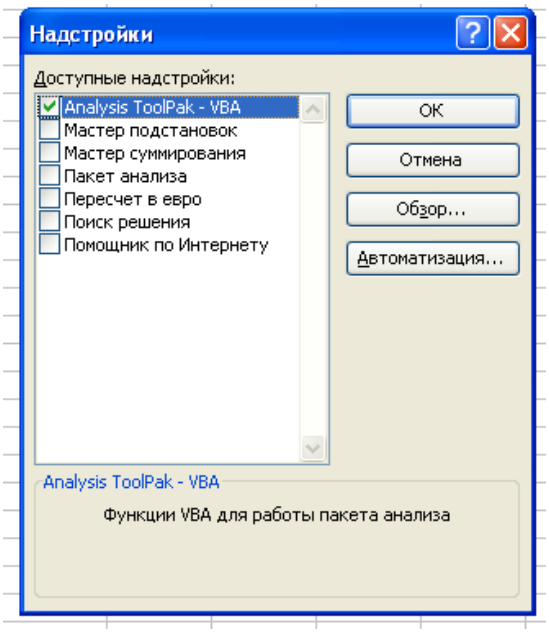


Рис. 2

2. Технология работы в режиме «Анализ данных»

Выберем в меню *Сервис* пункт *Анализ данных*..., появится окно с одноименным названием (рис. 3). Это окно - по существу «центр управления» надстройки *Пакет анализа*, главным элементом которого является область Инструменты анализа. В данной области представлен список реализованных в Microsoft Excel методов статистической обработки данных:

• «Гистограмма»;
• «Выборка»;
• «Описательная статистика»;
• «Ранг и персентиль»;
• «Генерация случайных чисел»;
• «Двухвыборочный г-тест для средних»;
• «Двухвыборочный /-тест с одинаковыми дисперсиями»;
• «Двухвыборочный /-тест с различными дисперсиями»;
• «Двухвыборочный F-тест для дисперсий»;
• «Парный двухвыборочный /-тест для средних»;
• «Однофакторный дисперсионный анализ»;
• «Двухфакторный дисперсионный анализ без повторений»;
• «Двухфакторный дисперсионный анализ с повторениями»;
• «Ковариация»;
• «Корреляция»;
• «Регрессия»;
• «Скользящее среднее»;
• «Экспоненциальное сглаживание»;
• «Анализ Фурье».

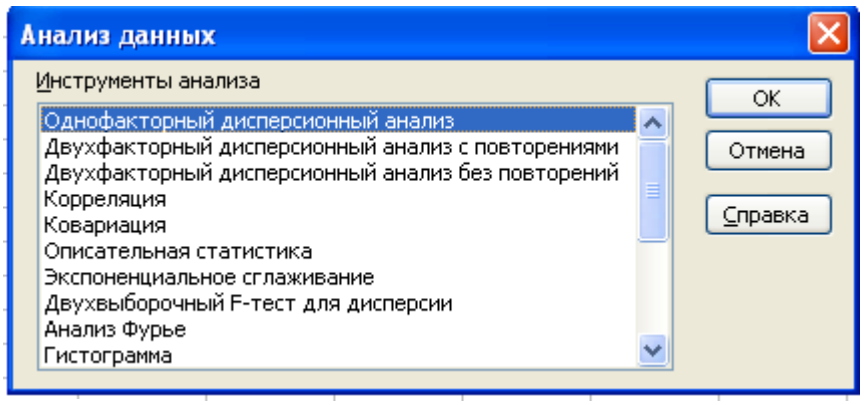


Рис. 3

Каждый из перечисленных методов реализован в виде отдельного режима работы, для активизации которого необходимо выделить соответствующий метод указателем мыши и щелкнуть по кнопке ОК. После появления диалогового окна вызванного режима можно приступать к работе.

Диалоговое окно каждого режима включает в себя элементы управления (поля ввода, раскрывающиеся списки, флажки, переключатели и т п.), которые задают определенные параметры выполнения режима (рис. 4).

Одна часть параметров является специфической и присуща только одному (или малой группе) режиму работы. Назначение таких параметров будет рассмотрено при изучении технологий работы с соответствующими режимами. Другая часть параметров универсальна и присуща всем (или подавляющему большинству) режимам работы. Элементами управления, задающими такие параметры, являются:

1. Поле *Входной интервал* — вводится ссылка на ячейки, содержащие анализируемые данные.

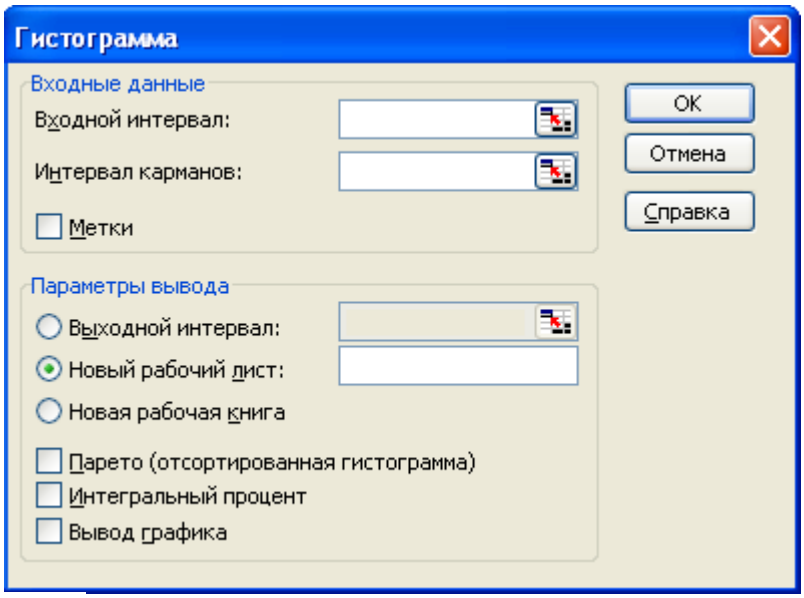


Рис. 4

2. Переключатель *Группирование* - устанавливается в положение По столбцам или По строкам в зависимости от расположения данных во входном диапазоне.

3. Флажок *Метки* — устанавливается в активное состояние, если первая строка (столбец) во входном диапазоне содержит заголовки. Если заголовки отсутствуют, флажок следует деактивизировать. В этом случае будут автоматически созданы стандартные названия для данных выходного диапазона.

4.Переключатель *Выходной интервал*/Новый рабочий лист/Новая рабочая книга. В положении Выходной интервал активизируется поле, в которое необходимо ввести ссылку на левую верхнюю ячейку выходного диапазона. Размер выходного диапазона будет опреде-

лен автоматически, и на экране появится сообщение в случае возможного наложения выходного диапазона на исходные данные.
В положении Новый рабочий лист открывается новый лист, в который начиная с ячейки A1 вставляются результаты анализа. Если необходимо задать имя открываемого нового рабочего листа, введите его имя в поле, расположенное напротив соответствующего положения переключателя. В положении Новая рабочая книга открывается новая книга, на первом листе которой начиная с ячейки A1 вставляются результаты анализа.

Работа с мастером функций

Наряду с надстройкой *Пакет анализа* в практике статистической обработки могут широко применяться статистические функции Microsoft Excel. В состав Excel входит библиотека, содержащая 78 статистических функций, ориентированных на решение самых рааличных задач прикладного статистического анализа.
Причем одну часть статистических функций можно рассматривать как своего рода элементарные составляющие того или иного режима надстройки Пакет анализа, другую часть – как уникальные функции, не дублирующиеся в надстройке Пакет анализа.
Тем не менее, функции, входящие и в первую часть, и во вторую часть, имеют самостоятельное значение и могут применяться автономно при решении конкретных статистических задач.
Работать со статистическими функциями Excel, как, впрочем, и с функциями из других категорий, удобнее всего с помощью мастера функций. При работе с мастером функций необходимо сначала выбрать саму функцию, а затем задать ее отдельные аргументы. Запустить мастер функций можно командой Функция... из меню Вставка, или щелчком по кнопке вызова мастера функций f_x, или активизацией комбинации клавиш Shift+F3
Для упрощения работы с мастером отдельные функции сгруппированы по тематическому признаку. Тематические категории представлены в области Категория (рис. 5). В категории Полный алфавитный перечень содержится список всех доступных Б программе функций, К категории 10 недавно использовавшихся относятся десять применявшихся последними функций. Поскольку пользователь во время работы применяет ограниченное число функций, то с помощью этой категории можно получить быстрый доступ к тем из них, которые необходимы в повседневной работе.
Чтобы задать статистическую функцию, сначала необходимо выбрать категорию Статистические, При перемещении строки выделения по списку функций под областями Категория и Функция будет представлен пример, иллюстрирующий способ задания выбранной статистической функции с краткой информацией о ней.
Если краткой информации недостаточно, щелкните в диалоговом окне по кнопке Справка. На экране появится помощник и предложит помощь. Щелкните по кнопке Справка по выделен

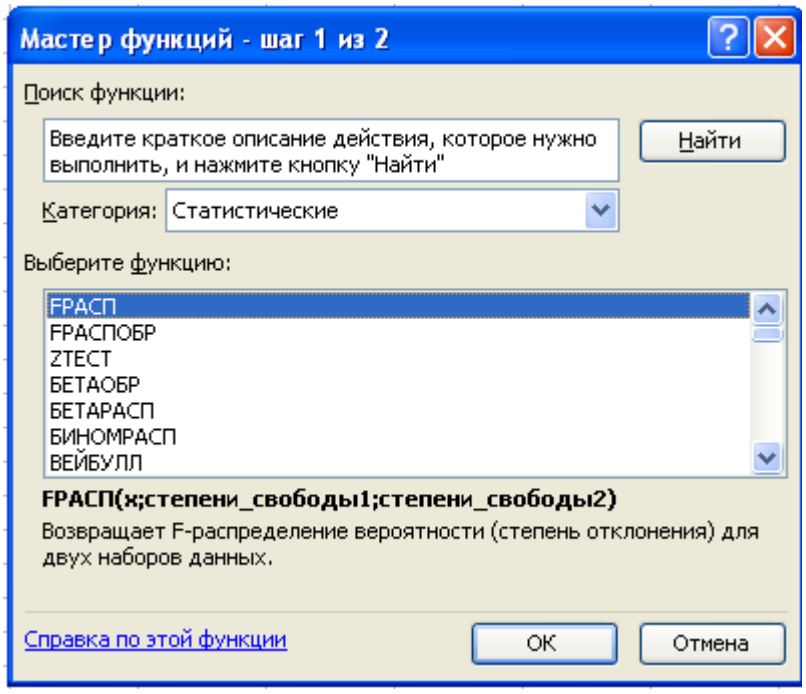


Рис. 5.

ной функции, и на экране будет представлена соответствующая страница справочной подсистемы.
После выбора функции щелкните по кнопке ОК для перехода в следующее диалоговое окно мастера функций, в котором должны быть заданы аргументы. В этом диалоговом окне мастер подсказывает пользователю, какие аргументы следует указать обязательно (обязательные аргументы), а какие - опционально (необязательные аргументы). Задать аргументы можно различными способами, наиболее удобные из них предлагает помощник.
После задания всех аргументов функции щелкните по кнопке ОК, чтобы в ячейке появились результаты выполнения функции.

2. Определение характера распределения и формирование выборки

Теоретические основы группировки

Результаты сводки и группировки материалов статистического наблюдения оформляются в виде таблиц и статистических рядов распределения.
Группировка – объединение единиц статистической совокупности в количественные однородные группы в соответствии со значениями одного или нескольких признаков.
Статистический ряд распределения представляет собой упорядоченное распределение единиц изучаемой совокупности по определенному варьирующему признаку. Он характеризует состояние (структуру) исследуемого явления, позволяет судить об однородности совокупности, единицах ее изменения, закономерностях развития наблюдаемого объекта.
Построение рядов распределения является составной частью сводной обработки статистической информации.
В зависимости от признака, положенного в основу образования ряда распределения, различают *атрибутивные* и *вариационные* ряды распределения. Последние, в свою очередь, в зависимости от характера вариации признака делятся
на *дискретные* ***(прерывные)*** **и** ***интервальные (непрерывные)*** **ряды распределения.**
Группировка осуществляется поэтапно. Вначале определяется примерное число групп, затем величина интервала. Строится первый вариант группировки, который при необходимости уточняется. Для определения числа групп может применяться формула Стерджесса: $r = 1 + 2,30259 \lg N$,где N – численность совокупности, r – число групп.
Величина интервала определяется по формуле: $c = \frac{X_{\max} - X_{\min}}{r}$,
где x_{max}, x_{min} – соответствующие максимальное и минимальное значения признаков совокупности, r – величина интервала. Полученный результат округляется.
Равные интервалы группировки применяются для однородных совокупностей, а для социально-экономических явлений чаще применяются неравноинтервальные группировки. Если крайнее значение единиц совокупности значительно отличается по величине от остальных, применяются группировки с открытыми границами интервалов.
Первый интервал с открытой нижней границей, последний интервал с открытой верхней границей. Величина первого интервала принимается равной величине следующего за ним интервала (не более чем). Величина последнего интервала с открытой верхней границей принимается равной величине предпоследнего интервала.
Различают абсолютные и относительные частотные характеристики. Абсолютная характеристика – **частота**, показывает, сколько раз встречается в совокупности данный вариант ряда. Достоинство частоты – простота, недостаток – невозможность сравнительного анализа рядов распределения разной численности.
Для подобных сравнений применяют относительные частоты или **частости**, которые рассчитываются по формуле: $di = \frac{f_i}{\sum f_i}$, $\sum f_i = N$, где N – численность совокупности.

Это относительная величина структуры (по форме). Сумма частостей равна 1. $\Sigma d_i = \Sigma \frac{f_i}{\Sigma f_i} = 1$ $\Sigma d_i = \Sigma \frac{f_i}{\Sigma f_i} = -$ Если частости выражены в процентах или в промилях их суммы равны соответственно 100 или 1000.

В неравных интервальных рядах распределения частотные характеристики зависят не только от распределения вариантов ряда, но и от величины интервала при прочих равных условиях расширение границ интервала приводит к увеличению наполненности групп.

Для анализа рядов распределения с неравными интервалами используют показатели плотности: **Абсолютная плотность:** $p_i = \frac{f_i}{c_i}$

где f_i – частота, c_i - величина интервала – показывает, сколько единиц в совокупности приходится на единицу величины соответствующего интервала. Абсолютная плотность позволяет сопоставлять между собой насыщенность различных по величине интервалов ряда. Абсолютные плотности не позволяют, однако, сравнивать ряды распределения разной численности. Для подобных сравнений применяются **относительные плотности:** $p_i' = \frac{d_i}{c_i}$, где d_i – частости (доли), c_i - величины соответствующих интервалов – показывает, какая часть (доля) совокупности приходится на единицу величины соответствующего интервала. Удобнее всего ряды распределения анализировать с помощью их графического изображения, позволяющего судить о форме распределения. Наглядное представление о характере изменения частот вариационного ряда дают **полигон** и **гистограмма.**

Полигон используется для изображения *дискретных* вариационных рядов. При построении полигона в прямоугольной системе координат по оси абсцисс в одинаковом масштабе откладываются ранжированные значения варьирующего признака, а по оси ординат наносится шкала частот, т. е. число случаев, в которых встретилось то или иное значение признака. Полученные на пересечении абсцисс и ординат точки соединяют прямыми линиями, в результате чего получают ломаную линию, называемую полигоном частот. Например, на рис. 6. приведено распределение числа студентов по успеваемости и полигон частот для данного распределения. Для построения полигона воспользуемся мастером диаграмм Microsoft Excel (режим «График»).

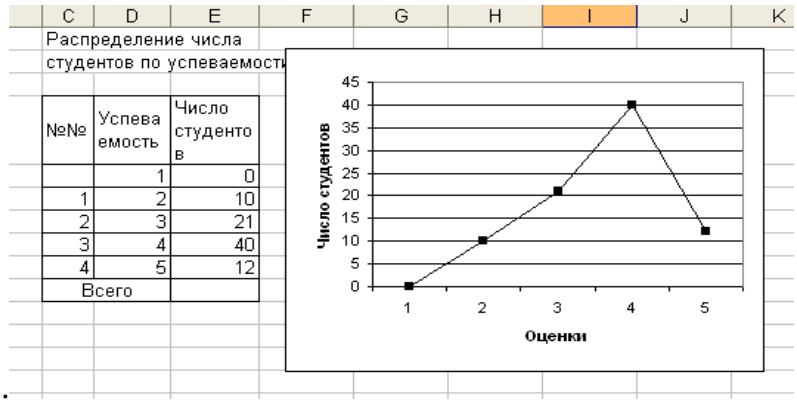


Рис. 6

Для изображения *интервальных* вариационных рядов распределений применяются *гистограммы.* При этом на оси абсцисс откладываются значения интервалов, а частоты изображаются прямоугольниками, построенными на соответствующих интервалах. В результате получается гистограмма – график, на котором ряд распределения представлен в виде смежных друг с другом областей. Для характеристики рядов распределения применяют так же графики накопленных частот или **кумуляты.**

Кумулята позволяет определить, какая часть совокупности обладает значениями изучаемого признака не превышающими заданного предела, а какая часть – наоборот – превышает этот предел.

Теоретические основы формирования выборки

Методология исследования массовых статистических явлений в зависимости от полноты охвата изучаемого объекта (явления) различает *сплошное* и *не сплошное* наблюдение. Разновидностью не сплошного наблюдения является выборочное, которое все более широкое применение.

Под *выборочным* наблюдением понимается метод статистического исследования, при котором обобщающие показатели изучаемой совокупности устанавливаются по некоторой ее части на основе положений случайного отбора. При выборочном методе обследованию подвергается сравнительно небольшая часть всей изучаемой совокупности, получившая название *выборочной совокупности* или просто *выборки.*

Выборка должна быть *представительной (репрезентативной),* чтобы по ней можно было судить о генеральной совокупности. Репрезентативность означает, что объекты выборки достаточно хорошо представляют генеральную совокупность. Заметим, что при отборе объектов могут сыграть роль личные мотивы или психологические факторы, о которых исследователь, проводящий выборку, и не подозревает. При этом выборка, как правило, не будет репрезентативной.

Предупреждение систематических (тенденциозных) ошибок выборочного обследования достигается в результате применения научно обоснованных способов формирования выборочной совокупности, в зависимости от которых выборка может быть: собственно-случайной; механической; типической; серийной; комбинированной.

В табличном процессоре Microsoft Excel реализована собственно-случайная выборка. ***Собственно-случайная выборка*** **состоит в том, что выборочная** совокупность образуется в результате случайного (непреднамеренного) отбора отдельных единиц из генеральной совокупности. Именно принцип случайности попадания любой единицы генеральной совокупности в выборку предупреждает возникновение систематических (тенденциозных) ошибок выборки. Собственно-случайная выборка может быть осуществлена по схемам *повторного* и *бесповторного* отбора. Повторный отбор предполагает возможность включения в выборку одного и того же элемента генеральной совокупности два раза и более. Бесповторный отбор исключает такую возможность. В Microsoft Excel **реализована схема*повторного отбора.*** На практике, особенно при большом объеме генеральной совокупности, дляорганизации собственно-случайной выборки часто используют таблицу случайных чисел или генератор случайных чисел. В Microsoft Excel выборка формируется на основе *генератора случайных чисел.*

Выборочный метод, обладая несомненным достоинством, состоящим в возможности значительно сократить время на контроль и получение основных статистических характеристик, приводит к появлению ошибки и уменьшению гарантии получения истинных характеристик генеральной совокупности. Данное обстоятельство особенно важно учитывать при формировании так называемых *малых выборок.* При этом достаточно сложной проблемой является определение необходимого (оптимального) объема выборки. В математической статистике доказывается, что необходимая численность *собственно-случайной повторной* выборки определяется выражением: $n = \frac{t^2\sigma^2}{\Delta_x^2}$

Δ_x^2 – предельная ошибка выборки;

σ^2 – дисперсия генеральной совокупности;

t – коэффициент доверия (определяется в зависимости от того, с какой доверительной вероятностью надо гарантировать результаты выборочного обследования).

Затруднительным моментом применения приведенной формулы на практике является расчет генеральной дисперсии σ^2 . Для ее оценки пользуются или материалами предыду-

щих исследований, или производственно-техническими нормативами, или, если предыдущие варианты неосуществимы, проводят пробное обследование. По результатам пробного обследования оценивают значение генеральной дисперсии для последующего обоснования необходимого объема выборки.

Технология работы при построении гистограммы

Режим «Гистограмма» служит для вычисления частот попадания данных в указанные границы интервалов, а также для построения гистограммы *интервального* вариационного ряда распределения.

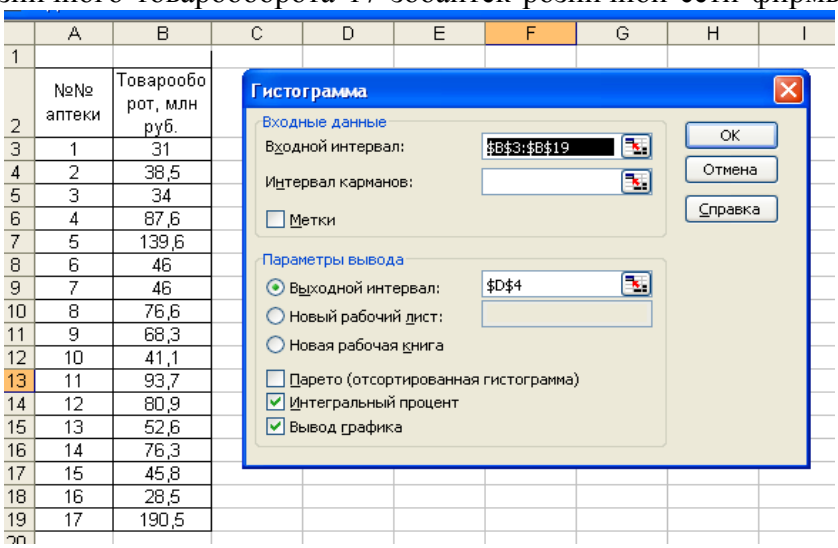
В диалоговом окне данного режима (рис. 7) задаются следующие параметры:

1. *Входной интервал*
2. *Интервал карманов* (необязательный параметр) – вводится ссылка на ячейки, содержащие набор граничных значений, определяющих интервалы (карманы). Эти значения должны быть введены в возрастающем порядке. В Microsoft Excel вычисляется число попаданий данных в сформированные интервалы, причем границы интервалов являются строгими нижними границами и нестрогими верхними: $a < x \leq b$

Если диапазон карманов не был введен, то набор интервалов, равномерно распределенных между минимальным и максимальным значениями данных, будет создан автоматически.

3. *Метки*
4. *Выходной интервал/Новый рабочий лист/Новая рабочая книга*
5. *Парето {отсортированная гистограмма)* — устанавливается в активное состояние, чтобы представить данные в порядке убывания частоты. Если флажок снят, то данные в выходном диапазоне будут приведены в порядке следования интервалов.
6. *Интегральный процент* - устанавливается в активное состояние для расчета выраженных в процентах *накопленных* частот *(накопленных частостей)* и включения в гистограмму графика куммуляты.
7. *Вывод графика* - устанавливается в активное состояние для автоматического создания встроенной диаграммы на листе, содержащем выходной диапазон.

Пример: Известен объем розничного товарооборота 17 зооаптек розничной сети фирмы «Однако» за 2007 г. Необходимо провести группировку аптек по уровню товарооборота, для этого воспользуемся режимом «Гистограмма» пакета *Анализ данных*. Исходные данные и значения параметров, установленных в диалоговом окне Гистограмма изображены на рис. 7. По набору этих данных необходимо построить гистограмму и куммуляту.



№№ аптеки	Товарооборот, млн руб.
1	31
2	38,5
3	34
4	87,6
5	139,6
6	46
7	46
8	76,6
9	68,3
10	41,1
11	93,7
12	80,9
13	52,6
14	76,3
15	45,8
16	28,5
17	190,5

Рис. 7

Частоты и накопленные частоты, построенные гистограмма и кумулята изображены на рис. 8

Карман	Частота	Интегральный %
28,5	1	5,88%
69	9	58,82%
109,5	5	88,24%
150	1	94,12%
Еще	1	100,00%

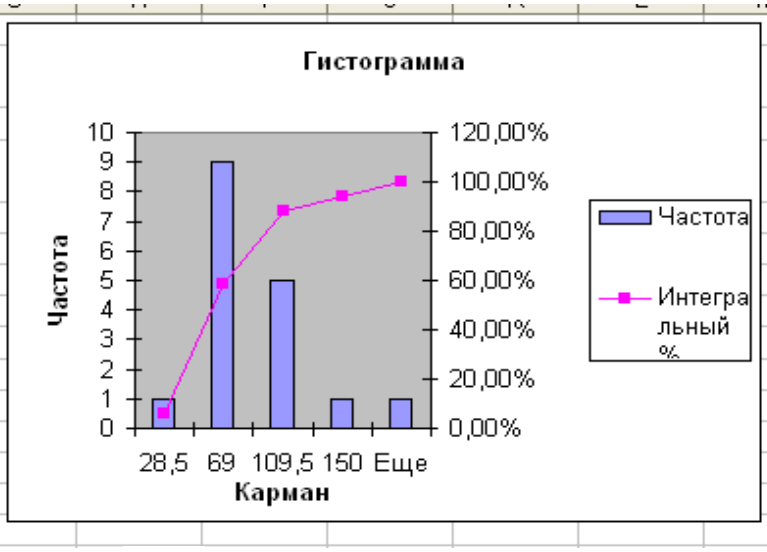


Рис. 8

Как правило, гистограммы изображаются в виде *смежных* прямоугольных областей, поэтому столбики гистограммы на рис. 8 целесообразно расширить до соприкосновения друг с другом. Для этого на панели инструментов Диаграмма необходимо в раскрывающемся списке элементов диаграммы выбрать элемент *Ряд 'Частота',* после чего щелкнуть по кнопке **Формат рядов данных.** В появившемся одноименном диалоговом окне необходимо активизировать вкладку ***Параметры*** и в поле *Ширина зазора* установить значение 0. После указанных преобразований гистограмма примет стандартный вид (рис. 9).

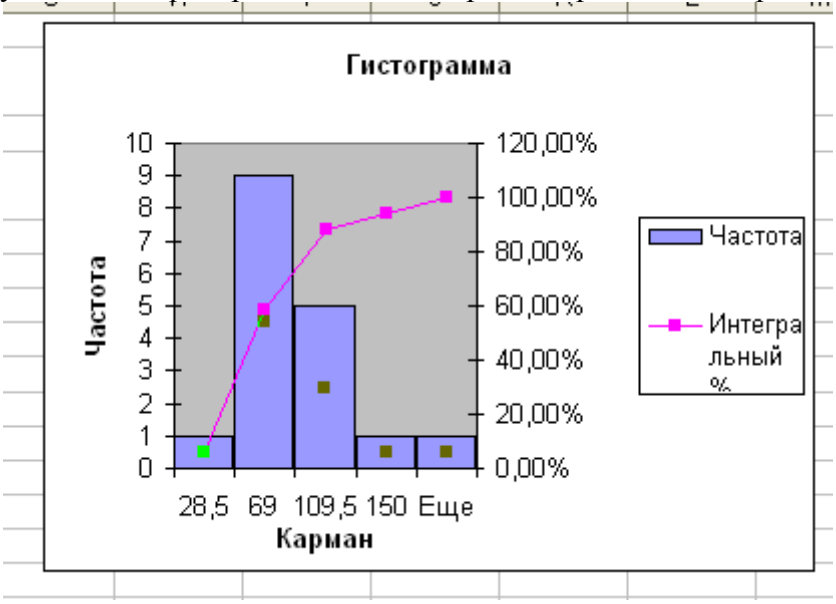


Рис. 9

Технология работы при формировании выборки

Пример. Оптовая фирма, торгующая изделиями медицинского назначения, решила для посетителей своего Web-сайта организовать лотерею по рассылке каталогов новой продукции. Для этого на сайте фирмы реализован счетчик посещений и предлагается (по желанию пользователя) заполнить электронный бланк с указанием своего почтового адреса. Отбор посетителей производится на основе показаний счетчика посещений за неделю. Для этого случайным образом отбираются пять показаний счетчика и проверяются соответствующие им регистрации посетителей. Если посетитель не указал своего адреса –каталог не высылается, в противном случае – высылается. При этом если одно и то же показание

счетчика попало в выигрышную выборку несколько раз или несколько «выигрышных визитов» на сайт осуществил один и тот же посетитель, каталог высылается по одному и тому же адресу в соответствующем количестве экземпляров. Рассмотрим следующую ситуацию. За последнюю неделю на сайте фирмы было зарегистрировано 25 посещений (показания счетчика увеличились с 360 до 385). Исходная информация, параметры расчетов и результаты приведены на рис. 10.

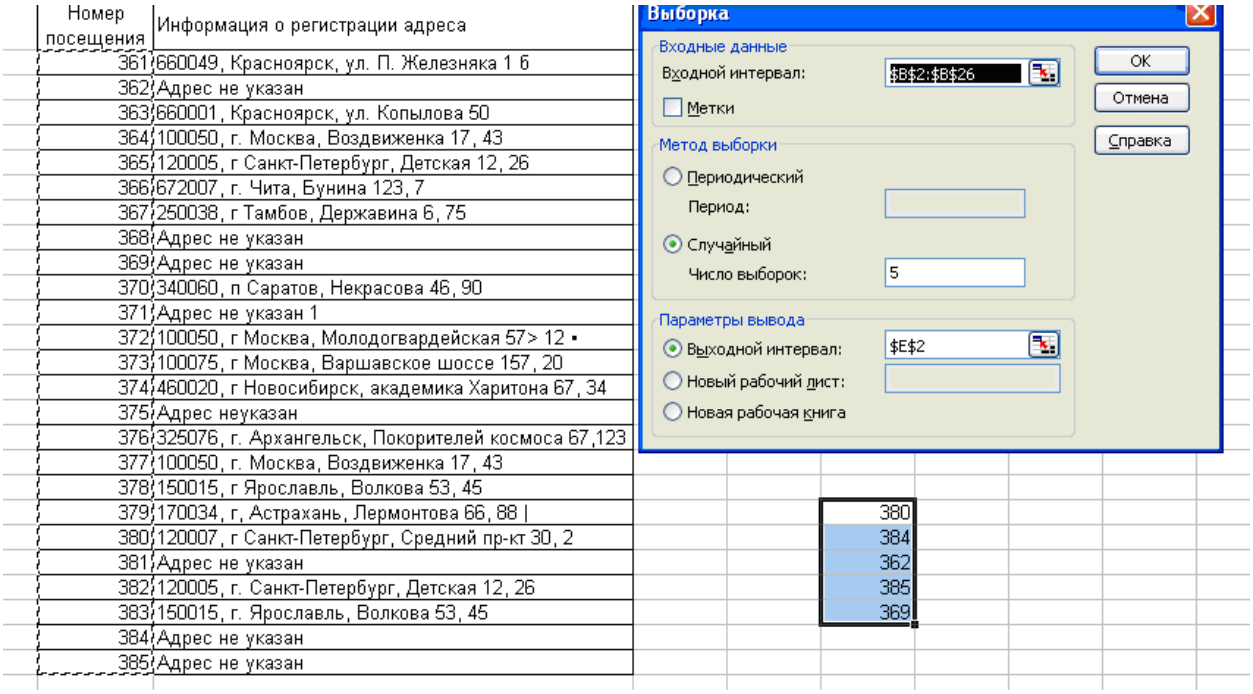


Рис. 10

Как видно из рисунка за последнюю неделю выигрышными оказались 362, 369, 380, 384, 385 посещения. Но каталог будет выслан только номеру 380 по указанному адресу. В остальных случаях рассылка проводиться не будет, т.к. клиентом не был указан адрес.

Пример. Предприятием «Импульс» за месяц было выпущено 150 медицинских приборов, которым были присвоены заводские номера с 10001-го по 10150-й включительно. Все приборы выпускаются на основании технической документации, в соответствии с которой дисперсия чувствительности приборов не превышает 25 мкВ^2/м^2. Необходимо на основе схемы повторного собственно-случайного отбора сформировать контрольную выборку, чтобы с уровнем надежности не менее 95 % предельная ошибка выборки не превышала 3 мкВ^2/м^2.

В примере 3.3.2, в отличие от примера 3.3.1, важным является момент определения необходимого объема выборки, чтобы она была репрезентативной. Для определения величины объема выборки воспользуемся формулой:

$$n = \frac{t^2\sigma^2}{\Delta_x^2}$$

Подставляя в нее исходные данные, получаем: $n = \frac{1.98^2 * 25}{3^2} \approx 11$

Примечание: В расчете необходимого объема выборки используется коэффициент дове-

рия *t*, для вычисления которого в Microsoft Excel предусмотрена функция СТЬЮДРАСПОБР. Коэффициент доверия *t* рассчитывается по формуле =СТЬЮДРАСПОБР(0,05;149), где 0,05 = 1-0,95 - требуемый уровень значимости, 149 = 150-1 – число степеней свободы. Таким образом, минимально допустимый объем выборки составляет 11 приборов. При меньшем объеме выборка не будет репрезентативной. Последующая технология решения задачи аналогична технологии решения задачи в примере 3.3.1. При этом в поле *Число выборок* вводится рассчитанное значение необходимого объема выборки г =11 (рис. 3.3.2). Для быстрого ввода исходных данных рекомендуем воспользоваться таким техническим приемом, как копирование ячеек с помощью меню Правка, Заполнить, Прогрессия *арифметической прогрессии с шагом 1.*

	A	B	C	D	E	F	G	H	I	J	K	L
1	10001	10016	10031	10046	10061	10076	10091	10106	10121	10136		
2	10002	10017	10032	10047	10062	10077	10092	10107	10122	10137		
3	10003	10018	10033	10048	10063	10078	10093	10108	10123	10138		
4	10004	10019	10034	10049	10064	10079	10094	10109	10124	10139		10014
5	10005	10020	10035	10050	10065	10080	10095	10110	10125	10140		10065
6	10006	10021	10036	10051	10066	10081	10096	10111	10126	10141		10025
7	10007	10022	10037	10052	10067	10082	10097	10112	10127	10142		10144
8	10008	10023	10038	10053	10068	10083	10098	10113	10128	10143		10027
9	10009	10024	10039	10054	10069	10084	10099	10114	10129	10144		10028
10	10010	10025	10040	10055	10070	10085	10100	10115	10130	10145		10033
11	10011	10026	10041	10056	10071	10086	10101	10116	10131	10146		10008
12	10012	10027	10042	10057	10072	10087	10102	10117	10132	10147		10138
13	10013	10028	10043	10058	10073	10088	10103	10118	10133	10148		10030
14	10014	10029	10044	10059	10074	10089	10104	10119	10134	10149		10051
15	10015	10030	10045	10060	10075	10090	10105	10120	10135	10150		

Рис. 11

Результатом решения задачи явилась выборка из 11 приборов с заводскими номерами: 10008, 10014, 10025, 10027, 10028, 10030, 10033, 10051, 10065, 10138, 10144. Так как в выборке номера приборов не повторяются, то каждый прибор подвергается проверке только один раз.

3. Характеристика статистической совокупности. Средние величины. Меры рассеяния

Статистическая информация представляется совокупностью данных, для характеристики которых используются разнообразные показатели, называемые показателями *описательной статистики.*

Уровень образования, прожиточный минимум, дифференциация доходов населения, среднее число детей в семье, средний курс доллара и мера его колебания за определенный интервал времени, таблицы продолжительности жизни.

Показатели описательной статистики можно разбить на несколько групп:

1. *Показатели положения* описывают положение данных на числовой оси. Примеры таких показателей - минимальный и максимальный элементы выборки (первый и последний члены вариационного ряда), верхний и нижний квартили (ограничивают зону, в которую попадают 50% центральных элементов выборки). Наконец, сведения о середине совокупности могут дать средняя арифметическая, средняя гармоническая, медиана и другие характеристики.

2. Показатели разброса описывают степень разброса данных относительно своего центра. К ним в первую очередь относятся: дисперсия, стандартное отклонение, размах выборки (разность между максимальным и минимальным элементами), межквартильный размах (разность между верхней и нижней квартилью), эксцесс и т. п. Эти показатели определяют, насколько кучно основная масса данных группируется около центра.

3. *Показатели асимметрии* характеризуют симметрию распределения данных около своего центра. К ним можно отнести коэффициент асимметрии, положение медианы относительно среднего и т. п.
4. *Показатели, описывающие закон распределения***,** дают представление о законе распределения данных. Сюда относятся таблицы частот, таблицы частостей, полигоны, кумуляты, гистограммы.
На практике чаще всего используются следующие показатели: средняя арифметическая, медиана, дисперсия, стандартное отклонение. Однако для получения более точных и достоверных выводов необходимо учитывать и другие из перечисленных выше характеристик, а также обращать внимание на условия получения выборочных совокупностей. Наличие выбросов, т. е. грубых ошибочных наблюдений, может не только сильно исказить значения выборочных показателей (выборочного среднего, дисперсии, стандартного отклонения и т. д.), но и привести ко многим другим ошибочным выводам. Наиболее распространенной формой статистических показателей, используемой в социально – экономических явлениях, является средняя величина, представляющая собой обобщенную количественную характеристику признака в статистической совокупности в конкретных условиях места и времени.
Сущность средней состоит в том, что в ней взаимопогащаются отклонения значений признака у отдельных единиц совокупности, обусловленные действием случайных факторов. Наиболее распространенным видом средних величин является **средняя арифметическая,** которая как и все средние, в зависимости от характера имеющихся данных может быть **простой или взвешенной**.
Средняя арифметическая простая (не взвешенная). Эта форма средней используется в тех случаях, когда расчет осуществляется по не сгруппированным данным. Зависимость для определения простой средней арифметической имеет вид: $\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{\sum x_i}{n}$
Средняя арифметическая взвешенная. При расчете средних величин отдельные значения осредняемого признака могут повторяться (встречаться по несколько раз). В подобных случаях расчет средней производится по сгруппированным данным или вариационным рядам Зависимость для определения средней арифметической взвешенной **для дискретного вариационного ряда** имеет вид: $\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i}$, где f_i – частота i-го признака
Наряду со средней арифметической, в санитарной статистике применяются, хотя и реже, такие виды средних, как медиана и мода.
Медиана (обозначаемая буквами Me) — это серединная, центральная варианта, делящая вариационный ряд пополам, на две равные части.
Мода (обозначаемая Mo) — чаще всего встречающаяся или наиболее часто повторяющаяся величина, соответствующая при графическом изображении максимальной ординате, т. е. наивысшей точке графической кривой. Таким образом, при приближенном нахождении моды в простом (несгруппированном) ряду она определяется как наиболее насыщенная или частая величина, как варианта с наибольшим количеством частот. Средняя арифметическая (M) является результативной суммой всех влияний. В ее формировании принимают участие все без исключения варианты, в том числе и крайние варианты. Медиана и мода, в отличие от средней арифметической, не зависят от величины всех индивидуальных значений, т. е. всех членов вариационного ряда, а обусловливаются относительным расположением или распределением вариант. Поэтому медиану и моду также называют описательными или позиционными средними, т. к. они характеризуют главнейшие свойства данного распределения. Особенно это касается медианы, являющейся в известном смысле, непараметрической величиной. M характеризует всю массу наблюде-

ний, а Me и Mo — основную массу, без учета воздействия крайних вариант, т. е. исключая крайние значения, зависящие иногда от случайных причин.
Средние арифметические величины, взятые сами по себе без дополнительных приемов оценки, имеют подчас ограниченное значение, т. к. они не отражают степени разброса (или рассеяния) ряда.
Одинаковые по размеру средние могут быть получены из рядов с различной степенью рассеяния. Средние — это величины, вокруг которых рассеяны различные варианты. Понятно, что чем ближе друг к другу отдельные варианты, (значит меньше рассеяние, колеблемость ряда), тем типичнее его средняя.

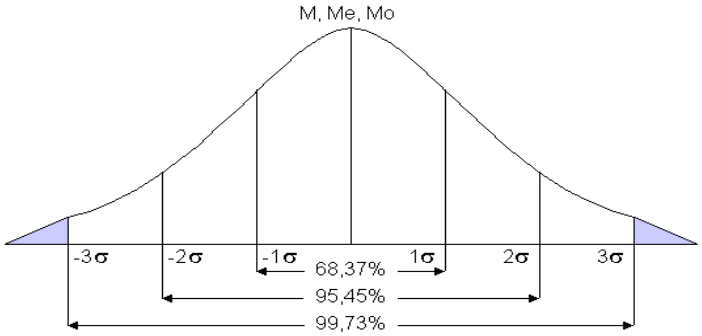


Рис. 12 Средние величины в ряду с нормальным распределением

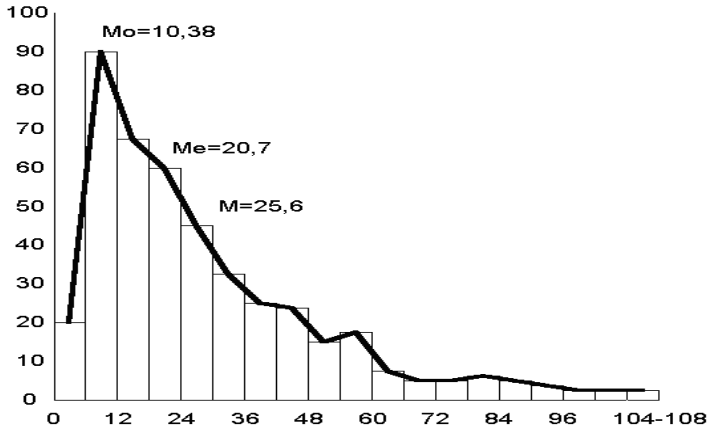


Рис. 13 Средние величины в ряду с ассиметричным распределением

Меры разброса (рассеяния). Средние арифметические величины, взятые сами по себе без дополнительных приемов оценки, имеют подчас ограниченное значение, т. к. они не отражают степени разброса (или рассеяния) ряда.
Одинаковые по размеру средние могут быть получены из рядов с различной степенью рассеяния. Средние — это величины, вокруг которых рассеяны различные варианты. Понятно, что чем ближе друг к другу отдельные варианты, (значит меньше рассеяние, колеблемость ряда), тем типичнее его средняя. Существует ряд показателей, с помощью которых как раз и оценивается мера разброса или рассеяния ряда.
Размах – разность между наибольшими и наименьшими значениями в ряде или распределении. Размах учитывает только экстремальные значения и поэтому не дает информации о разбросе отдельных элементов
Дисперсия – средний квадрат отклонения значений от их арифметического среднего
Среднеквадратичное отклонение – положительный корень из дисперсии. Это показатель разброса данных около арифметического среднего

Коэффициент вариации – это среднеквадратичное отклонение, деленное на арифметическое среднее, выраженное в процентах.

Технология работы по описательной статистике

Режим «Описательная статистика» служит для генерации одномерного статистического отчета по основным показателям положения, разброса и асимметрии выборочной совокупности. В диалоговом окне данного режима (рис. 14) задаются следующие параметры:

1. *Входной интервал*
2. *Группирование*
3. *Метки в первой строке/Метки в первом столбце*
4. *Выходной интервал/Новый рабочий лист/Новая рабочая книга*
5. *Итоговая статистика* – установите в активное состояние, если в выходном диапазоне необходимо получить по одному полю для каждого из следующих показателей описательной статистики: средняя арифметическая выборки ($\tilde{x}$), средняя ошибка выборки ($\mu_{\tilde{x}}$), медиана (Me), мода (Mo), оценка стандартного отклонения по выборке (σ), оценка дисперсии по выборке (D), оценка эксцесса по выборке (E_K), оценка коэффициента асимметрии по выборке (A_K), размах вариации выборки (R), минимальный и максимальный элементы выборки, сумма элементов выборки, количество элементов в выборке, k-й наибольший и k-й наименьший элементы выборки, предельная ошибка выборки ($\Delta_{\tilde{x}}$).

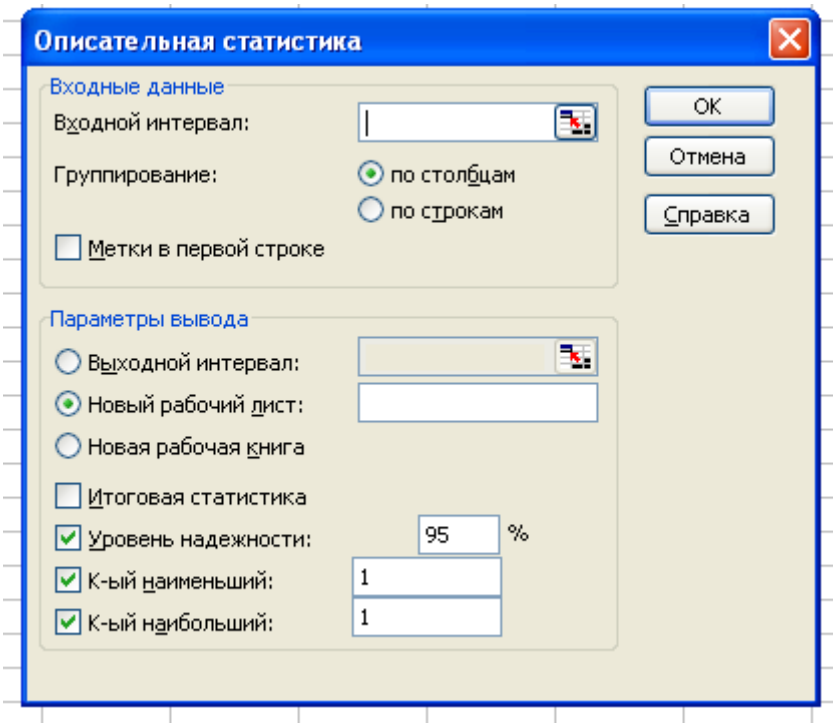


Рис. 14

6. *Уровень надежности* - установите в активное состояние, если в выходную таблицу необходимо включить строку для предельной ошибки выборки ($\Delta_{\tilde{x}}$) при установленном уровне надежности. В поле, расположенном напротив флажка, введите требуемое значение уровня надежности (например, значение уровня надежности 95 % равносильно доверительной вероятности γ = 0,95 или уровню значимости α = 0,05).
7. *К-й наибольший* - установите в активное состояние, если в выходную таблицу необходимо включить строку для *k-го* наибольшего (начиная с максимума x_{max}) значения элемента выборки. В поле, расположенное напротив флажка, введите число *k*. Еслиk = 1, то строка будет содержать максимальное значение элемента выборки.

8. *К-й наименьший* — установите в активное состояние, если в выходную таблицу необходимо включить строку для *k-mo* наименьшего (начиная с минимума x_{min}) значения элемента выборки. В поле, расположенное напротив флажка, введите число *k*. Если $k = 1$, то строка будет содержать минимальное значение элемента выборки.

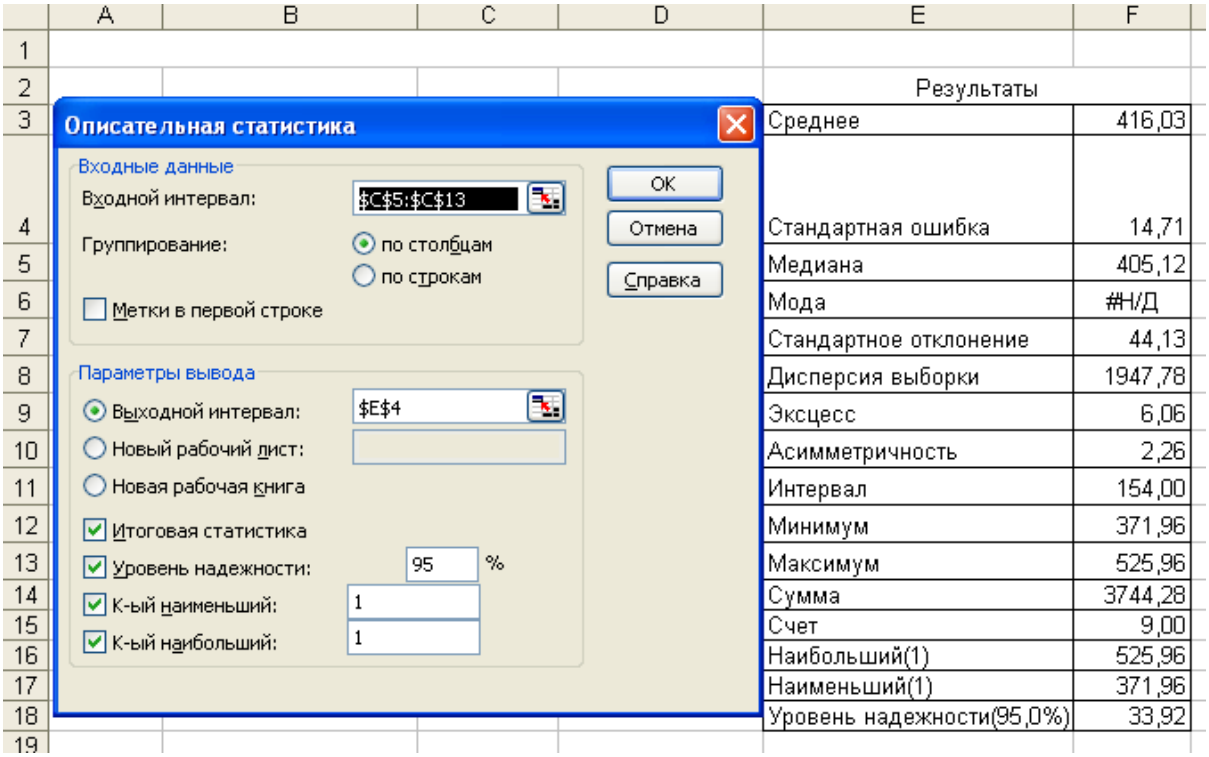
Пример. На рабочем листе Microsoft Excel сформирована таблица, отражающая условную стоимость лечения заболевания X в некоторых городах Красноярского края (рис. 15). Необходимо рассчитать основные показатели описательной статистики и сделать соответствующие выводы. Для решения задачи используем режим работы «Описательная статистика». Значения параметров, установленных в одноименном диалоговом окне представлены и показатели, рассчитанные в данном режиме отображены на рис. 16 (результаты округлены до двух значащих цифр).

Стоимость лечения заболевания X

Города	Стоимость лечения (руб.)
Красноярск	389,04
Ачинск	417,78
Назарово	394
Сосновоборск	371,96
Норильск	525,96
Дивногорск	405,12
Заозерный	419,52
Канск	401,93
Шарыпово	418,97

Рис. 15

На основании проведенного выборочного обследования и рассчитанных по данной выборке показателей описательной статистики с уровнем надежности 95% можно пред-



Результаты	
Среднее	416,03
Стандартная ошибка	14,71
Медиана	405,12
Мода	#Н/Д
Стандартное отклонение	44,13
Дисперсия выборки	1947,78
Эксцесс	6,06
Асимметричность	2,26
Интервал	154,00
Минимум	371,96
Максимум	525,96
Сумма	3744,28
Счет	9,00
Наибольший(1)	525,96
Наименьший(1)	371,96
Уровень надежности(95,0%)	33,92

Рис 16

положить, что средняя стоимость лечения заболевания X в целом по всем городам Красноярского края находилась в пределах от 382,11 до 449,95 руб.

Поясним, на основании каких показателей описательной статистики был сформулирован соответствующий вывод. Такими показателями являются: средняя арифметическая выборки $\tilde{x}$ и предельная ошибка выборки $\Delta\tilde{x}$

Из выражения для доверительного интервала: $\tilde{x} - \Delta\tilde{x} \le \tilde{x} \le \tilde{x} + \Delta\tilde{x}$ находим: 416,03–33, 92 = 382,11 – *левая граница;* 416,03 + 33,92= 449,95 – *правая граница.*

Коэффициент вариации $\nu = \frac{\sigma}{\tilde{x}}$ существенно меньше 40 %, что свидетельствует о небольшой колеблемости признака в исследованной выборочной совокупности. Надежность средней в выборке подтверждается также и ее незначительным отклонением от медианы: 416,03 - 405,12 = 10,91. Значительные положительные значения коэффициентов асимметрии *{A_S)* и эксцесса *(E_K)* позволяют говорить о том, что данное эмпирическое распределение существенно отличается от нормального, имеет правостороннюю асимметрию и характеризуется скоплением членов ряда в центре распределения.

4. Методы проверки статистических гипотез

Понятие статистической гипотезы

Проверка статистических гипотез является одним из самых важных статистических методов, применяемых в медицинских исследованиях.

Под статистической гипотезой понимают всякое высказывание о генеральной совокупности (случайной величине), проверяемое по выборке (по результатам наблюдений). Процедуру сопоставления высказанной гипотезы с выборочными данными называют проверкой статистической гипотезы.

Проверяемую статистическую гипотезу принято называть основной (или нулевой) гипотезой (обозначается H_0), а противоречащую ей гипотезу - альтернативной (или конкурирующей) гипотезой (обозначается H_1). Нулевая гипотеза обычно заключается в том, что изучаемое вмешательство не оказывает значимого воздействия на генеральные параметры распределения, и альтернативная, исследовательская гипотеза. Изначально предполагается, что вмешательство не влияет. Любые различия между изучаемыми группами объясняются случайностью, и становится задача опровергнуть это предположение

Поскольку при проверке статистических гипотез приходится иметь дело со статистическим материалом, то, отвергая или принимая нулевую гипотезу, всегда рискуем совершить ошибку. Ошибку, заключающуюся в том, что нулевая гипотеза отвергается, тогда как она в действительности верна, называют ошибкой первого рода. Ошибку, состоящую в том, что нулевая гипотеза не отвергается, тогда как она в действительности неверна, называют ошибкой второго рода.

Проверка статистических гипотез осуществляется с помощью различных статистических критериев. В качестве критерия используется некоторая случайная величина, значения которой могут быть вычислены на основе имеющихся данных. В множестве возможных значений критерия выбирается подмножество, называемое критической областью. Если вычисленное значение критерия принадлежит критической области, то нулевая гипотеза отвергается.

Критическая область выбирается таким образом, чтобы вероятность совершить ошибку первого рода не превосходила некоторого заранее определенного положительного числа α. Это число α называют уровнем значимости и говорят: «нулевая гипотеза отвергается на уровне значимости α. В качестве α обычно берут одно из чисел: 0,05; 0,01; 0,001.

Вероятность совершить ошибку второго рода обозначается β. Величина 1-β называется мощностью критерия; она равна вероятности отвергнуть неверную гипотезу Чаще всего множество возможных значений критерия принадлежит некоторому интервалу. Интерва-

лом является и критическая область. Граничные точки критической области называются критическими точками. Критические точки выбираются таким образом, чтобы при выбранном уровне значимости, а мощность критерия (1 - β) была наибольшей.

Двухвыборочный t-тест для средних. Технология работы

Часто в биологических исследованиях требуется оценить достоверность различий между двумя выборками. Это могут быть выборки из разных совокупностей (например, сравнение опытной и контрольной групп), а также выборки из одной совокупности (например, исследование какого-либо параметра в одной выборке до и после проведения эксперимента).

Двухвыборочный t-тест проверяет равенство средних значений генеральной совокупности по каждой выборке. Расчет проводится по формуле: В пакете статистического анализа Excel реализовано три варианта использования t-теста: равные дисперсии генерального распределения, дисперсии генеральной совокупности не равны, а также представление двух выборок до и после наблюдения по одному и тому же субъекту.

Режимы работы **«Двухвыборочный t-тест с одинаковыми дисперсиями»** и **«Двухвыборочный t-тест с различными дисперсиями»** служат для проверки гипотез о различии между средними (математическими ожиданиями) двух нормальных распределений соответственно с неизвестными, но равными дисперсиями и с неизвестными дисперсиями, равенство которых не предполагается.

Параметры, задаваемые для расчетов, отображены на рисунке 17.

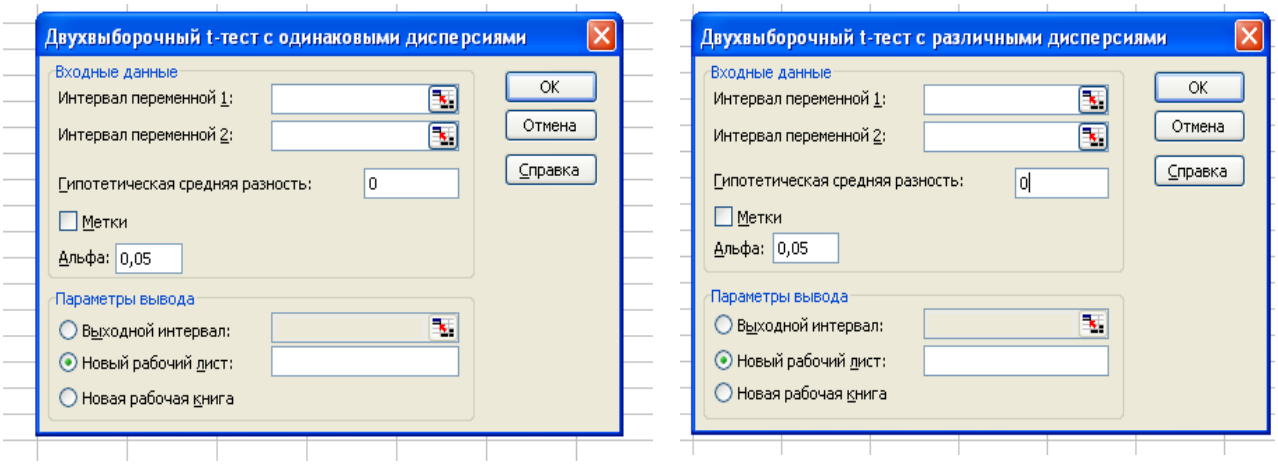


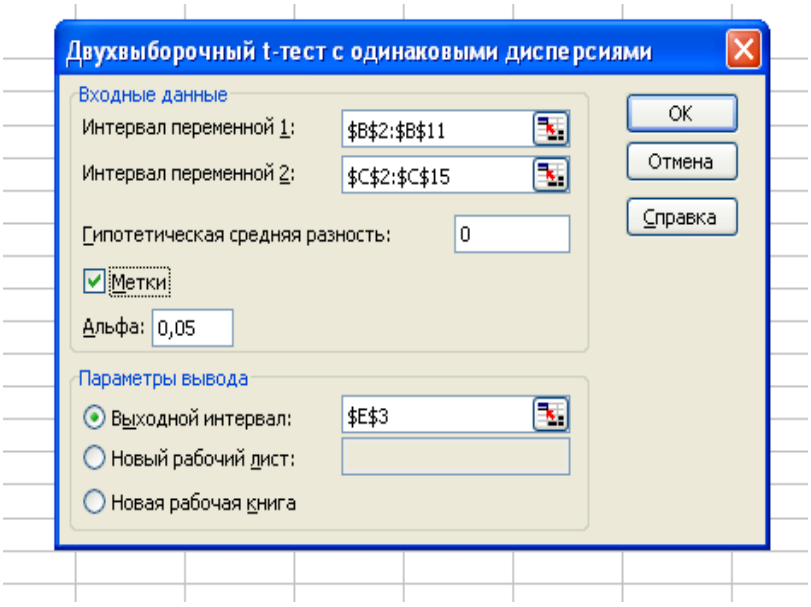
Рис. 17

Пример. На рабочем листе Microsoft Excel сформирована таблица, отражающая расход в граммах (на 100 м^2 площади) на дезинфекцию отделений с использованием старого и нового дезинфицирующего препарата (рис. 18).

	A	B	C
1	расходы на дезинфекцию		
2	Номер отделения	Старый препарат	Новый препарат
3	1	308	308
4	2	308	304
5	3	307	306
6	4	304	306
7	5	307	304
8	6	307	304
9	7	308	304
10	8	307	306
11	9	307	304
12	10		304
13	11		303
14	12		304
15	13		303
16			

Рис. 18

При уровне значимости α = 0,05 требуется проверить гипотезу H_0: $a_x = a_y$, предположив, что соответствующие генеральные совокупности *X и Y* имеют нормальные распределения: 1) с одинаковыми дисперсиями σ_x^2 и σ_y^2; 2) *с* различными дисперсиями σ_x^2 и σ_y^2. Для проверки предположения 1 используем режим работы «Двухвыборочный t-тест с одинаковыми дисперсиями», а для проверки предположения 2 - «Двухвыборочный t-тест с различными дисперсиями». Значения параметров, установленных в одноименных диалоговых окнах, представлены на рис. 19, а рассчитанные в этих режимах показатели — в табл. 20 соответственно.



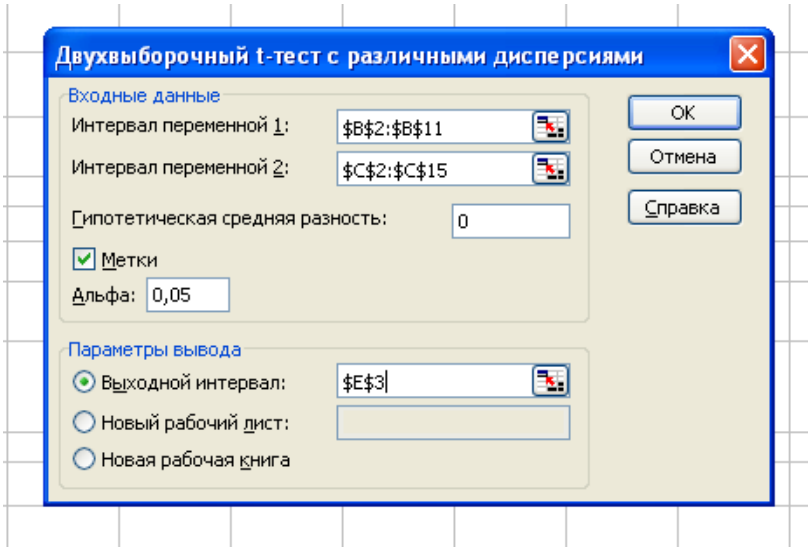


Рис. 19

Двухвыборочный t-тест с одинаковыми дисперсиями		
	Старый препарат	*Новый препарат*
Среднее	307	304,62
Дисперсия	1,5	2,09
Наблюдения	9	13
Объединенная дисперсия	1,85	
Гипотетическая разность средних	0	
df	20	
t-статистика	4,04	
P(T<=t) одностороннее	0,0003	
t критическое одностороннее	1,72	
P(T<=t) двухстороннее	0,001	
t критическое двухстороннее	2,09	

Двухвыборочный t-тест с различными дисперсиями		
	Старый препарат	*Новый препарат*
Среднее	307	304,62
Дисперсия	1,5	2,09
Наблюдения	9	13
Гипотетическая разность	0	
df	19	
t-статистика	4,17	
P(T<=t) одностороннее	0,0003	
t критическое одностороннее	1,73	
P(T<=t) двухстороннее	0,0005	
t критическое двухстороннее	2,09	

Рис. 20

Так как и в первом, и во втором варианте расчета показатели *t-статистики* = 4,04 и 4,17 соответственно превышали *t* - критическое = 2,09 то гипотезу H_0: $a_x=a_y$ отвергаем, т. е. при переходе на новый дезинфицирующий препарат происходит уменьшение его среднего расхода на обработку 100 м^2.

Рассмотренные выше процедуры сравнения двух выборок часто применяются для обнаружения результата какого-либо воздействия либо, напротив, для подтверждения его отсутствия. Чем более однородными окажутся выбранные для эксперимента объекты (для контроля и воздействия), чем меньше их случайные различия, тем точнее можно будет

дать ответ на поставленный вопрос. Ясно, что различие между объектами, выбранными для воздействия и для контроля (или для двух разных воздействий, если интерес представляет их сопоставление), будет наименьшим, если в обоих качествах выступает один и тот же объект.

Если это возможно, то далее обычным порядком составляется группа экспериментальных объектов и затем для каждого объекта измеряются два значения интересующей нас характеристики (например, до воздействия и после или при двух разных воздействиях). Так возникают пары наблюдений или парные данные.

Пусть x_i и y_i — результаты измерений для объекта номер i, i= 1, ..., n, где n - численность экспериментальной группы (число объектов). Тогда совокупность пар случайных величин$(x_1, y_1)..., (x_n \ldots y_n)$ образует парные данные. Как обычно, все наблюдения будем считать реализациями случайных величин и предполагать, что методика эксперимента обеспечивает их независимость для разных объектов. Но наблюдения, входящие в одну пару, нельзя считать независимыми, поскольку они относятся к одному и тому же объекту. Эти два наблюдения отражают свойства общего для них индивидуального объекта и потому могут зависеть друг от друга. Для пар наблюдений (x_1, y_1) введем величину z_i =y,- x_1, которую будем считать независимой и нормально распределенной. Тем самым задача о парных данных сводится к задаче об одной нормальной выборке при неизвестной дисперсии.

Режим работы «Парный двухвыборочный t-тест для средних» служит для проверки гипотезы о различии между средними (математическими ожиданиями) двух нормальных распределений на основе парных выборочных данных. При этом равенство дисперсий генеральных совокупностей не предполагается ($\sigma_x^2 \neq \sigma_y^2$). В диалоговом окне данного режима задаются параметры, аналогичные параметрам, задаваемым выше.

Пример. Проводили клинический эксперимент. Измеряли артериальное давление у группы пациентов в начале лечения препаратом и в конце. Исходные данные и результаты расчетов приведены в таблице, сформированной в MS Excel (рис. 21). Требовалось проверить, повлиял ли препарат на уровень артериального давления.

Как видно из результатов расчета показатель *t-статистики* =0,96 превышал *t* - критическое = 2,26, поэтому гипотезу H_0:a_x=a_y принимаем, т. е. действие препарата не привело к уменьшению средних параметров давления.

Номер истории болезни	Уровень АД в начале эксперимента	Уровень АД в конце эксперимента
1	145	140
2	140	140
3	151	150
4	175	165
5	135	130
6	150	145
7	165	160
8	170	175
9	150	155
10	140	145

	Уровень АД в начале эксперимента	Уровень АД в конце эксперимента
Среднее	152,1	150,5
Дисперсия	184,1	180,3
Наблюдения	10	10
Корреляция Пирсона	0,92	
Гипотетическая разность средних	0	
df	9	
t-статистика	0,96	
P(T<=t) одностороннее	0,18	
t критическое одностороннее	1,83	
P(T<=t) двухстороннее	0,36	
t критическое двухстороннее	2,26	

Рис. 21

5. Статистические методы изучения взаимосвязей явлений и процессов

Корреляционный анализ. Краткие сведения из теории статистики

В экономических исследованиях одной из важных задач является анализ зависимостей между изучаемыми переменными. Зависимость между переменными может быть либо функциональной, либо статистической. Для оценки тесноты и направления связи между изучаемыми переменными при их стохастической зависимости пользуются показателями корреляции.

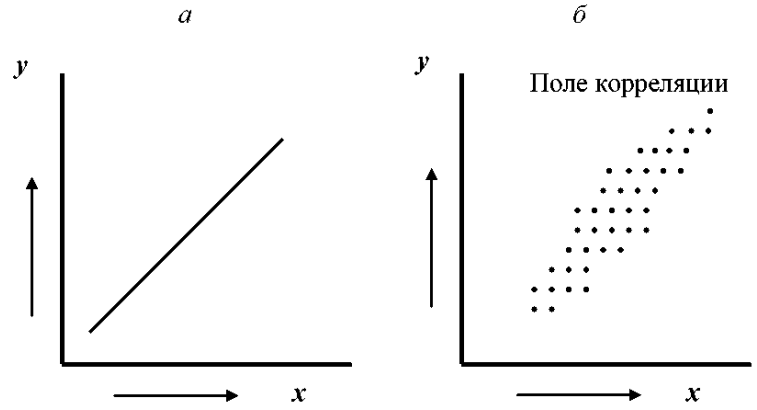


Рис. 22. Зависимость функциональная (а) и статистическая (б)

В 1889 г. Ф. Голтон высказал мысль о коэффициенте, который мог бы измерить тесноту связи между двумя коррелируемыми признаками. В начале 90-х гг. XIX в. Пирсон, Эджворт и Велдон получили формулу линейного коэффициента корреляции: $r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y}$

Линейный коэффициент корреляции характеризует степень тесноты не всякой, а только линейной зависимости. Линейная вероятностная зависимость случайных величин заключается в том, что при возрастании одной случайной величины другая имеет тенденцию возрастать (или убывать) по линейному закону. Эта тенденция к линейной зависимости может быть более или менее ярко выраженной, т. е. более или менее приближаться к функциональной. Если случайные величины *X* и *Y* связаны точной линейной функциональной зависимостью $y=ax+b$ то $r_{xy}=\pm 1$. В общем случае, когда величины *X* и *Y* связаны произвольной вероятностной зависимостью, линейный коэффициент корреляции принимает значение в пределах $-1 < r_{xy} < 1$, тогда качественная оценка тесноты связи величин *X* и *Y* может быть выявлена на основе шкалы Чеддока (табл. 1).

Таблица 1

Теснота связи	Значение коэффициента корреляции при наличии:	
	прямой связи	обратной связи
Слабая	0,1-0,3	(-0,1)-(-0,3)
Умеренная	0,3 - 0,5	(-0,3)-(-0,5)
Заметная	0,5 - 0,7	(-0,5)-(-0,7)
Сильная	0,7 - 0,9	(-0,7) - (-0,9)
Очень сильная	0,9 - 0,99	(-0,9) - (-0,99)

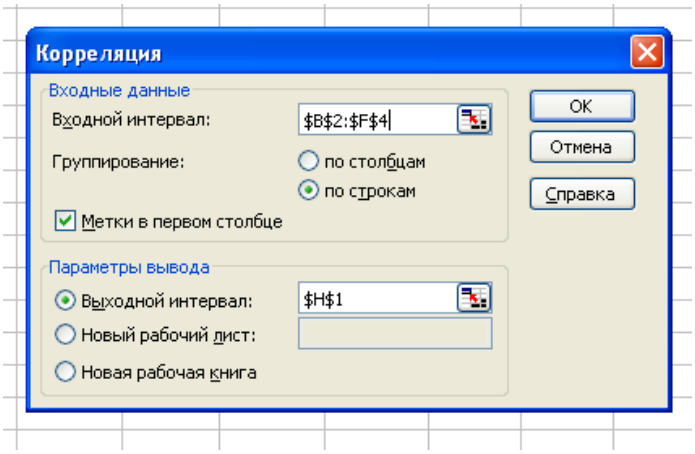
Расчет коэффициента корреляции. Технология работы.

Режим работы «Корреляция» предназначен для расчета *генерального* и *выборочного* коэффициентов корреляции соответственно на основе генеральных и выборочных данных.**Пример**: На рабочем листе Excel сформирована таблица (рис. 23). Изучалась зависимость между распространенностью наркомании, преступностью и уровнем образования в некоторых городах России. По выборочным данным, представленным в таблице, требуется установить наличие взаимосвязи между показателями

. Рис. 23

Значение параметров в диалоговом окне и полученная корреляционная матрица выглядят следующим образом (рис. 24).

	A	B	C	D
1	Города	Распространенность наркомании	Преступность связанная с наркотиками	Уровень образования
2	А	313,5	125,2	735
3	Б	345,5	146	730
4	В	434,6	205,3	715
5	Г	509,4	185,8	725
6	Д	501,4	205,2	705
7	Е	270,6	114,4	730
8	Ж	300,6	117,8	730
9				



	Распространенность наркомании	Преступность связанная с наркотиками	Уровень образования
Распространенность наркомании	1,0		
Преступность связанная с наркотиками	0,93	1,0	
Уровень образования	-0,76	-0,86	1,0

Рис. 24

Как видно из корреляционной матрицы на рис. 24, между парами всех исследуемых показателей существуют стохастические связи. Характер всех выявленных связей различен и состоит в следующем:

Связь «преступность, связанная с оборотом наркотиков» и «распространенность наркомании» является положительной и очень сильной (r_{xy} = 0,93), связь «уровень образования» – «уровень преступности» является сильной и обратной (r_{xy} = -0,76), т. е. с повышением уровня образования уровень преступности уменьшается; связь «уровень образования» - «распространенность наркомании» также является сильной и обратной (r_{xy} = - 0,86), т. е. с увеличением уровня образования распространенность наркомании уменьшается.

Статистическую оценку коэффициента парной корреляции проводят путем сравнения его абсолютной величины с табличным (или критическим) показателем r_{xy} крит, значения которого отыскиваются из специальной таблицы. Если окажется, что r_{xy} расч $\geq r_{xy}$ крит, то с заданной степенью вероятности (обычно 95 %) можно утверждать, что между рассматриваемыми числовыми совокупностями существует значимая линейная связь. Или по-другому − гипотеза о значимости линейной связи не отвергается. В случае же обратного соотношения, т.е. при r_{xy} расч $< r_{xy}$ крит, делается заключение об отсутствии значимой связи. Так, в нашем примере r_{xy} крит = 0,754, поэтому с вероятностью 95% мы можем говорить о достоверной линейной связи между изучаемыми показателями.

Часто корреляцию и причинную обусловленность считают синонимами. Этот тезис имеет определенные основания, поскольку если н*е*что является причиной чего-либо другого, то можно говорить о связи первого и второго и, следовательно, об их коррелированности (например, действие и результат, проверка и качество, капиталовложения и прибыль, окружающая среда и прибыль).

Однако корреляция может быть и без причинной обусловленности. Это можно представить так: корреляция − лишь число, которое указывает на то, что б*о*льшим значениям одной переменной соответствуют б*о*льшие (или же м*е*ньшие) значения другой переменной. Корреляция *не* может объяснить, *почему* эти две переменные связаны между собой. Так, корреляция не объясняет, почему капиталовложения порождают прибыль (или наоборот). Корреляция просто констатирует, что между этими величинами существует определенное соответствие. И не более того.

Одним из возможных оснований для существования «корреляции без причинной обусловленности» является наличие некоторого скрытого, ненаблюдаемого, *третьего фактора*, который «маскируется» под другую переменную. В результате фиксируется так называемая «*ложная корреляция*».

В качестве статистического показателя может быть использован также *коэффициент (индекс) детерминации* (*причинности*) R^2, который равен квадрату коэффициента корреляции (r^2). Он показывает, в какой мере изменчивость y (результативного признака) объясняется поведением x (факторного признака), или иначе: какая часть общей изменчивости y вызвана собственно влиянием x. Этот показатель вычисляется путем простого возведения в квадрат коэффициента корреляции. Тем самым доля изменчивости y, определяемая выражением $1-R^2$, оказывается необъясненной. Коэффициент детерминации R^2 для нашего примера равен $0{,}93^2$=0,865 или 86,5%. Величина показателя свидетельствует о том, что 86,5% (вариации) распространенности наркомании объясняются уровнем преступности, связанной с наркотиками. Остальные 13,5% объясняются какими-то другими причинами.

Величина этого коэффициента меняется в пределах от 0 до 1. Чем ближе он к единице, тем, следовательно, меньше в нашей модели процесса влияние неучтенных факторов.

1.4 Лекция 6 (2 ч.)

Тема: Экспертные оценки в прикладных исследованиях. Ранговый коэффициент корреляции. Коэффициент конкордации для оценки согласия экспертов. Метод парных сравнений в условиях иерархии.

1.4.1. Вопросы лекции:

1. Конкордация
2. Ранговая корреляция.
3. Теория и практика экспертных оценок

1.4.2 Краткое содержание вопросов:

Конкордация

Английским статистиком Кенделлом был предложен коэффициент (коэффициент конкордации) позволяющий оценивать степень согласованности мнений экспертов. Будучи теоретически обоснованным, он позволяет с достаточной степенью уверенности судить о согласованности мнений экспертов в некоторой группе.

Пусть даны результаты опроса M экспертов относительно Nобъектов каждому, из которых они приписывают определенный ранг.

эксперты	Объекты					
	1	2	.....	j	.....	N
1	x_{11}	x_{12}	.....	x_{1j}	.....	x_{1N}
2	x_{21}	x_{22}	.....	x_{2j}	.....	x_{2N}
⋮	.....	.....	.....	.....	.....	.....
i	x_{i1}	x_{i2}	.....	x_{ij}	.....	x_{iN}
⋮	.....	.....	.....	.....	.....	.....
M	x_{M1}	x_{M2}	.....	x_{Mj}	.....	x_{MN}

$W = \frac{12S}{M^2(N^3 - N)}$, где $S = \sum_{j=1}^{N}[\sum_{i=1}^{M} x_{ij} - \frac{1}{2}M(N+1)]^2$ W- коэффициент конкордации

W=1 при полном согласии экспертов

W=0 при рассогласованности мнений экспертов.

Кендаллом был найден закон распределения W как случайной величины, т.е. в предположении, что ранги всем объектам приписываются случайным равновероятным способом, т.е. что любая комбинация рангов из возможных n! равновероятна. Тогда как это принято в мат. статистике вычислив $W_{наблюденное}$ можно по специальным таблицам найти вероятность события P ($W \geq W_{набл}$). Если эта вероятность мала ($P(W \geq W_{набл}) < \alpha$, где α- наперед заданное число, обычно=0,5), то считают, что согласованность достаточно высокая, а в противном случае считается, что нет оснований говорить о хорошей согласованности экспертов. Распределение W табулировано только для небольших значений N(1-10), при больших N случайная величина W*M(N-1) имеет распределение асимптотически близкое к распределению χ^2 с числом степеней свободы υ=N-1.

Замечание. Иногда эксперт затрудняется отдать предпочтение одному объекту перед другими и приписывает двум или более объектам одинаковые ранги. В таких случаях формула для вычисления W принимает

$$W = \frac{12S}{M^2(N^3-N) - 12M\sum_{i=1}^{M} T_i}, \qquad T_i = \frac{1}{12}\sum_{t_j^{(i)}} \left(t_j^{(i)3} - t_j^{(i)}\right)$$

вид:

$t^{(i)}_j$– число одинаковых рангов поставленных i-ым экспертом.

Пример: От i-го эксперта получены следующие ранги дляN=10, 14444488810,

тогда T_i= 1/12(5^3-5+3^3-3)=144/12=12

В случае совпадения рангов и больших N случайная величина аппроксимирующая W распределения χ^2будет рассчитываться тоже несколько иначе:

$$\chi^2 = \frac{S}{\frac{1}{2}MN(N+1) - \frac{1}{N-1}\sum_{i=1}^{M} T_i}$$

Пример на оценку согласованности:

Пусть 5 экспертам было предложено проранжировать 7 факторов влияющих на технологический процесс. Результаты представлены в таблицы:

эксперты	Факторы						
	1	2	3	4	5	6	7
1	1	2	6	4	7	3	5
2	1	2	7	6	3	5	4
3	7	1	6	4	2	5	3
4	3	1	5	6	4	7	2
5	1	2	6	4	5	7	3

M=5 N=7

S=$(13-20)^2$+$(8-20)^2$+$(30-20)^2$+$(24-20)^2$+$(21-20)^2$+$(27-20)^2$+$(14-20)^2$=368

$W_{íàáë} = \frac{12*368}{257*(49-1)} = 0{,}526$

$S_{кр}$=343,8 при α=0,01

Если $S_{набл} > S_{кр}$, то мнения экспертов согласованы удовлетворительно.

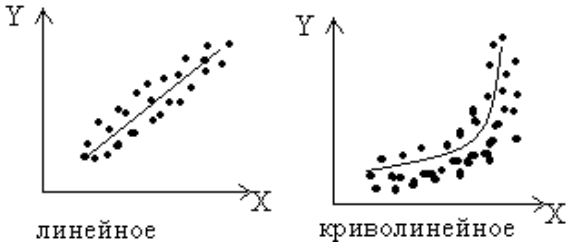
Ранговая корреляция. (Коэффициент корреляции Кендалла.)

При неудовлетворительной согласованности ответов экспертов необходим более подробный анализ. В частности, следует выяснить какие группы экспертов сильно расходятся во мнениях, какие нет. Здесь, например можно исследовать корреляции между суждениями экспертов в каждой паре. Таких пар будет C^2_N, где N- число экспертов. Напомним что обычный выборочный коэффициент корреляции между двумя случайными величинами X иY представленными выборкой (x_1,y_1);(x_2,y_2)…(x_n,y_n) вычисляется по формуле:

$$R = \frac{\frac{1}{n} * \sum_{i=1}^{n} x_i y_i - \bar{x} * \bar{y}}{\sigma_x * \sigma_y}$$ (1), гдеx$_i$,y$_i$- элементы выборки из случайных величин X и Y имеющих нормальное распределение, $\bar{x}, \bar{y}$ - средние выборочные $\bar{x} = \frac{1}{n}\sum_{i-1}^{n} x_i$; $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$,

а σ_x и σ_y– стандартные отклонения выборки $\sigma_x = [\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2]^{\frac{1}{2}}$; $\sigma_y = [\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2]^{\frac{1}{2}}$

В математической статистике доказывается, что коэффициент корреляции R измеряет силу (тесноту) корреляционной связи (частный случай стохастической зависимости) между случайными величинами X и Y когда с изменением одной случайной величины меняется условное мат. ожидание другой случайной величины.
Кроме того: 0≤|R|≤1: R>0- положительная корреляция; R<0- отрицательная корреляция
R=0-XиYне коррелированны; R=±1-XиYсвязаны функционально
Заметим что функциональная зависимость, есть частный случай стохастической. В общем случае, чем больше, |R| тем больше сила корреляционной связи между X и Y. В математической статистике существуют критерии по проверке значимости величины R. При значимо большом R строят уравнения регрессии и уточняют его тип (линейное, криволинейное и т.д.)



Подчеркнем, что все это справедливо касается данных полученных в абсолютной шкале измерений, и строго говоря, неприменимо к другим шкалам.
Кенделл сконструировал свой ранговый коэффициент корреляции, используя понятие инверсии в перестановке рангов. Назовем любое взаимное расположение N– натуральных чисел, перестановкой. Например: N=5 21534 или 15423
Будем называть перестановку 123... N- полным порядком. Будем говорить, что любая пара чисел ($r_i r_j$) в которой $r_i > r_j$ нарушает порядок или дает инверсию. В любой перестановке можно подсчитать число инверсий U.
Например: N=8 23154867U=5
Естественно назвать перестановку N, (N-1), (N-2)…. 321 полным беспорядком. Число инверсий в нем равно $U = U_{MAX} = \frac{N(N-1)}{2}$
Кендалл предложил следующие выражение для рангового коэффициента корреляции:

$$R_K = 1 - \frac{4U}{N(N-1)}.$$

Все замечания относительно области изменения R_K, аналогичны случаю R_S.

Оценка значимости r_k.

При N≤10, значимость оценивают по специальным таблицам. При N>10,R_Kхорошо аппроксимируется нормальным распределением и оценивается следующим образом, вычис-

$T_{KP} = Z_{KP}\sqrt{\frac{2(2N+5)}{9N(N-1)}}$ ляется , где Z_{KP} – односторонняя критическая точка стандартного нормального распределения для уровня значимости α, т.е. $\Phi(Z_{KP}) = \frac{1-2\alpha}{z}$,

$\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_0^z e^{\frac{-t^2}{2}} dt$ а - функция Лапласа.

Если $R_{Kнабл} > T_{KP}$ – согласованность хорошая.

Если $R_{Kнабл} < T_{KP}$ – согласованность плохая.

Пример:

10 объектов отданы на экспертизу 2-м экспертам.

Объекты	Ранги 1-го	Ранги 2-го
1	1	3
2	2	4
3	3	1
4	4	2
5	5	8
6	6	5
7	7	10
8	8	6
9	9	9
10	10	7

$R_k = 1 - \frac{4U}{N(N-1)}$ $R_k = 1 - \frac{4*11}{10*9} = 0.511$

U=2+2+3+3+1=11

α=0.05 согласие удовлетворительно; α=0,01 согласия нет

Теория и практика экспертных оценок

Получение группового мнения экспертов это довольно широкая и непростая программа мероприятий, требующая от организаторов высокой компетентности и обычно немалых затрат времени и средств. Поэтому неправильно организованная программа экспертизы может, поглотив средства и время, не достичь поставленной цели.

Программа экспертного опроса требует комплексного системного подхода и предполагает следующую последовательность работ: формулирование цели экспертизы и определение набора альтернативных вариантов оцениваемых объектом событий или фактов; формирование экспертной группы; формулирование правил работы экспертной группы с организаторами экспертизы; формулирование правил выработки группового (коллективного) суждения экспертной группы. Совокупность всех этих этапов, связанных в единое целое и образует систему экспертных оценок.

На первом этапе формулируется цель (или цели), а также возможные альтернативные варианты средств для достижения их цели, критерии, по которым можно судить о степени достижения цели. И всё это должно быть доведено до сведений экспертной группы после её формирования.

Принципы формирования экспертной группы

Очевидно критериями для выбора того или иного эксперта являются такие, как сфера деятельности, учёная степень, стаж работы, непредвзятость, готовность искренне сотрудничать, умение отстаивать свою точку зрения, проявлять достаточную гибкость, не быть упрямым. Есть несколько методов формирования экспертной группы.

Метод «снежного кома»

Пример: Применения метода для формирования экспертной группы.

В таблице 1 представлены результаты опроса 10 экспертов, каждый из которых должен был назвать 5 специалистов.

Кого назвали	Кто назвал										Сколько раз назвали	Какие места занимает
	1	2	3	4	5	6	7	8	9	10		
1				1	1	1		1		1	5	4,5,6,7
2	1		1	1		1	1	1	1	1	8	1,2
3		1			1						2	9,0
4	1	1					1	1		1	5	4,5,6,7
5	1	1	1	1		1	1		1		7	3
6			1		1		1	1	1		5	4,5,6,7
7	1	1	1	1	1	1			1	1	8	1,2
8				1						1	2	9,10
9		1	1		1	1	1				5	4,5,6,7
10	1							1	1		3	8

1. эга… метод («всё равно»); весовой метод

Кого назвали	Кто назвал										Сколько раз назвали	Какие места занимает	Веса
	1	2	3	4	5	6	7	8	9	10			
1		2		5	7	5		2		3	22	7	0,088
2	5		2	5		5	8	2	5	3	35	3	0,140
3		8			7						15	8	0,06
4	5	8					8	2		3	26	5	0,104
5	5	8	2	5		5	8		5		38	2	0,152
6			2		7		8	2	5		24	6	0,096
7	5	8	2	5	7	5			5	3	40	1	0,160
8				5						3	8	10	0,032
9		8	2		7	5	8				30	4	0,120
10	5							2	5		12	9	0,048
											Σ = 250		Σ = 1,0

Методы экспертного опроса.

Опросы экспертов обычно осуществляются в виде специальных анкет.

Анкета (опросный лист) – структурно-организованный набор вопросов, каждый из которых в какой-то мере связан с основной задачей экспертизы.

Вопросы делятся на 3 группы:

1. Объективные анкетные данные о самом эксперте
2. Характеристики, касающиеся мотивов, которыми руководствуется эксперт при оценке проблем
3. Основные вопросы, касающиеся существа исследуемой проблемы.

По форме бывают:

а) открытые, закрытые

б) прямые, косвенные

а) это такие, на которые можно отвечать в свободной форме - открытые. Закрытые – множество ответов на которые конечно.

б) Прямо относящиеся к теме – прямой

Сами опросы бывают очные и заочные.

По характеру взаимодействия экспертов с экспертной группой опросы бывают трех типов:

1. круглый стол или консилиум
2. мозговая атака
3. метод Делфи

Метод Дельфе. Основные черты: анонимность, регулируемая обратная связь, сходящийся групповой ответ, многотуровость. Анонимно рассылаются анкеты, эксперты их заполняют, далее руководитель экспертизы выясняет согласованность ответов, выявляются особые мнения. Во втором туре эксперты узнают результаты первого тура. Эксперт задумывается, почему его ответы не совпали с мнением большинства или наоборот совпали. После этого эксперт принимает ответ большинства или настаивает на своем. После многих туров руководитель экспертизы принимает решение. Недостаток этого метода заключается в его дороговизне.

Поэтому более предпочтительным способом получить групповое мнение является так называемый метод золотой середины и суть его состоит в том, что за коллективное мнение принимают медиану μе множества суждений, т.е. такое число μе, что $n_{<}=n_{>}$, где$n_{<}(n_{>})$ – количество оценок меньших (больших) числа μ.

Медиана выгодно отличается от среднего тем, что не требует для своего применения каких-либо серьезных теоретических предпосылок и кроме того, медиана более устойчива к изменениям индивидуальных оценок.

Измерение выполненных в шкале отношений

Когда эксперт работает в шкале отношений, он решает вопрос во сколько раз объект по іважнее объекта поj. Если в качестве базовой оценки принятьVj=1 (а это надо делать обязательно во избежание парадоксов), то шкала отношений вырождается в бальную.

Бальная же шкала всегда может быть поставлена во взаимно однозначное соответствие с ранговой (шкалой порядка).

N	**1**	**2**	**3**	**4**	**5**
баллы	20	100	70	50	10
ранги	4	1	2	3	5

Шкала интервалов

Если в шкале порядка оценки от экспертов получены в виде рангов, отражающих систему предпочтений, то в качестве групповой оценки вычисляют либо средний ранг для каждого объекта, либо медиану рангов, а затем уже в соответствии с результатом усреднения снова ранжируют объекты.

Если нужно вычислить какой из двух объектов A или B предпочтительней (A>B(B>A)), то общее мнение часто вырабатывается путем голосования, используя принцип большинства. По сути дела человечество ничего лучшего не придумало, кроме как принцип большинства при голосовании.

Пример:

Пытаются выявить силу предпочтения эксперта, как сильно A важнее B. В криминалистической практике экспертам предлагается наряду с объектами A и B упорядочить еще несколько альтернативных объектов C,D,E.

1 этап (C, D, A, B, E) A>B

2 этап (B, C, D, E, A) B>A

Но при использовании принципа большинства при голосовании нас подстерегают трудности более глубокого и принципиального характера. Речь идет о так называемом парадоксе голосования.

Пример:

пусть 3 эксперта сравнивают 3 объекта A,B,C и результат сравнения таков:

1 этап A>B>C 2 из 3 A>B

2 этап B>C>A 2 из 3 B>C

3 этап C>A>B

A>B и B>C => A>C

Здесь налицо нарушение транзитивности, если использовать принципы большинства при голосовании.

Многие из рассмотренных выше парадоксов оказываются носят принципиальный характер, т.е. их в принципе нельзя избежать. В свое время это доказал американский специалист в области математической экономики К. Дж. Эрроу в теореме «Парадокс Эрроу».

Приведем смысл теоремы:

Предъявим несколько разумных требований, которые должно удовлетворять всякое правило согласования индивидуальных мнений:

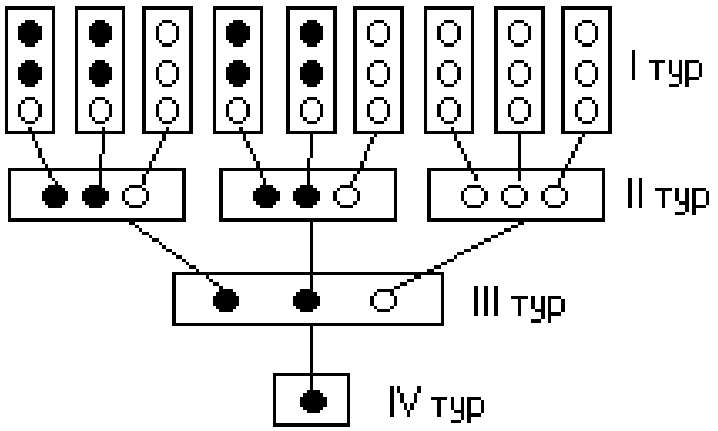
1. оно должно быть универсальным, т.е. применимо к любым совокупностям индивидуальных оценок.
2. введение некоторых объектов не должно изменять порядок предпочтений в предыдущей совокупности объектов – условие монопольности.
3. если все индивидуальные оценки имеют вид A>B, то групповая оценка имеет вид A>B(условие суверенности)
4. среди экспертов не должно быть эксперта, чьё мнение всегда принимается за групповое независимо от мнения других экспертов (условие отсутствия диктатора)

Так вот теорема утверждает, что не существует такого правила согласия мнений, которое бы удовлетворяло одновременно всем четырем требованиям. Парадокс состоит в том, что в эту теорему очень трудно поверить, но она строго доказана и поэтому верна.

В заключении отметим еще одни парадокс процедуры голосования, по причине большинства, связанный на этот раз с вмешательством коалиций.

Пример: при многоступенчатом голосовании по правилу большинства коалиция, находящаяся в большинстве может добиться принятия своего решения в ущерб большинству.

Пример: пусть в одном туре голосования серые оказались в меньшинстве против белых в соотношении 8:19



3й тур 2:1
2й тур 4:5
1й тур 8:19

В голосовании во втором туре решено организовать коалицию по 3 человека по правилу большинства (если в коалиции больше серых, то коалиция будет голосовать за серого и наоборот)

Наиболее распространенным является метод анализа иерархий (МАИ), который применяется для согласования результатов, полученных с использованием различных подходов и методов оценки [4].

МАИ, опирающийся на многокритериальное описание проблемы, был предложен и детально описан Т. Л. Саати в работе «Принятие решений: метод анализа иерархий». В методе используется дерево критериев, в котором общие критерии разделяются на критерии частного характера. Для каждой группы критериев определяются коэффициенты важности. Альтернативы также сравниваются между собой по отдельным критериям с целью определения критериальной ценности каждой из них. Средством определения коэффициентов важности критериев либо критериальной ценности альтернатив является попарное сравнение. Результат сравнения оценивается по балльной шкале. На основе таких сравнений вычисляются коэффициенты важности критериев, оценки альтернатив и находится общая оценка как взвешенная сумма оценок критериев.

Общая идея данного метода заключается в декомпозиции проблемы выбора на более простые составляющие части и обработку суждений лица, принимающего решение. В результате определяется относительная значимость исследуемых альтернатив по всем критериям, находящимся в иерархии.

Первым шагом МАИ является структурирование (декомпозиция) проблемы, согласование результатов в виде иерархии (рис. 2). В наиболее простом виде иерархия строится с вершины, представляющей цель проблемы через промежуточные уровни, обычно являющиеся критерием сравнения к самому нижнему уровню, который в общем случае является набором альтернатив.

Верхний уровень - **цель** (например, определение рыночной стоимости); промежуточный уровень - **критерии согласования** (например, возможность отразить действительные намерения потенциального инвестора и продавца; тип, качество, обширность данных, на основе которых проводится анализ; способность параметров используемых методов учитывать конъюнктурные колебания; способность учитывать специфические особенности объекта, влияющие на его стоимость (местонахождение, размер, потенциальная доходность)); нижний уровень -**набор альтернатив** (например, результаты, полученные различными методами оценки).

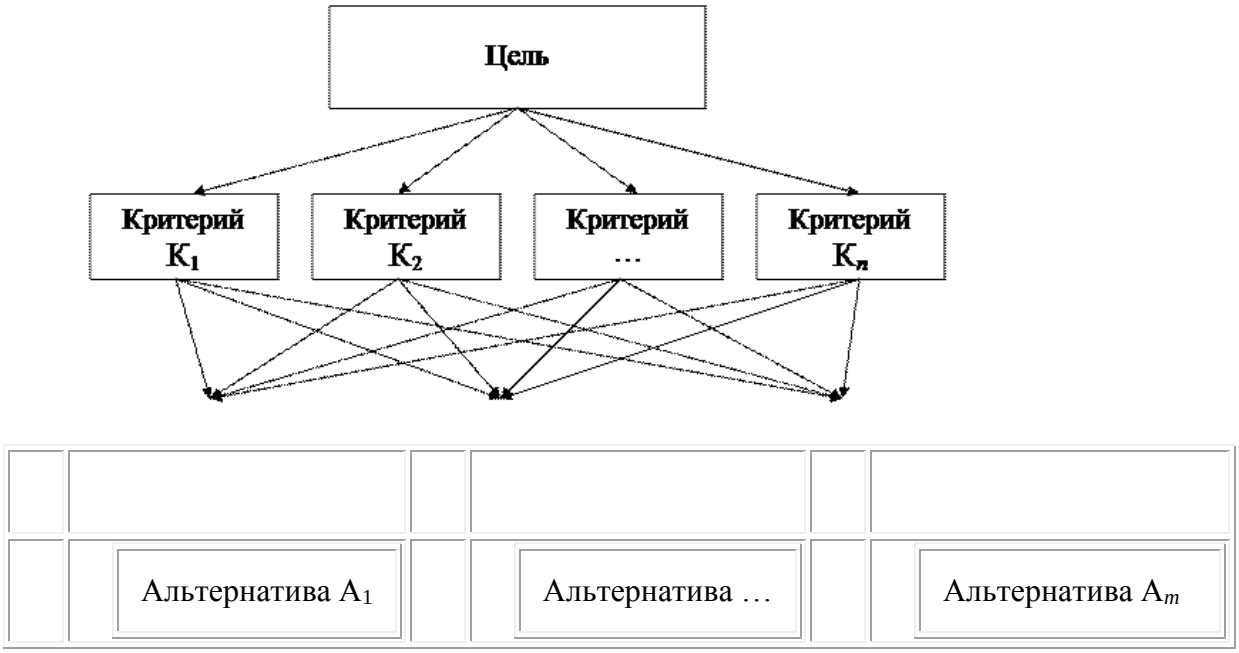


Рис.2 Структура иерархии

Очень часто в практике оценки собственности используется следующий набор критериев:

- тип, качество, обширность данных, на основе которых проводится анализ;
- способность параметров используемых методов учитывать конъюнктурные колебания рынка;
- способность учитывать специфические особенности объекта, влияющие на его стоимость;
- возможность отразить действительные намерения продавца или покупателя.

Конечно же, можно указывать более развернутый перечень факторов, влияющих на достоверность определения рыночной стоимости, но это приведет к значительному повышению трудоемкости последующих вычислительных процедур.

После иерархического воспроизведения проблемы строится матрица сравнения критериев и рассчитывается значение приоритетов критериев.

Для этого попарно сравниваются критерии по отношению к их воздействию на общую цель. Система парных сравнений приводит к результату, который может быть представлен в виде обратно симметричной матрицы, на главной диагонали которой находятся единицы.

Парные сравнения проводятся в терминах доминирования одной альтернативы над другой. Полученные суждения выражаются в целых числах с учетом девятибалльной шкалы. Преимущества именно этой шкалы качественных оценок отмечают многие исследователи, так как она позволяет наилучшим образом учесть степень отличия и имеет наименьшее среднеквадратическое отклонение.

Элементом матрицы a(i, j) является интенсивность проявления элемента иерархии i относительно элемента иерархии j, оцениваемая по шкале интенсивности от 1 до 9.

Заполняя матрицы парных сравнений, оценщик руководствуется следующими соображениями:

- если сравниваемые подходы по анализируемому критерию одинаково предпочтительны, то соответствующий элемент матрицы равен 1;
- если один из подходов имеет незначительное превосходство над другим по анализируемому критерию, то соответствующий элемент матрицы равен 3 или 4;
- если один из подходов имеет значительное превосходство над другим по анализируемому критерию, то соответствующий элемент матрицы равен 5 или 6;

- если один из подходов имеет явное превосходство над другим по анализируемому критерию, то соответствующий элемент матрицы равен 7 или 8;
- если один из подходов имеет абсолютное преимущество по сравнению с другим по анализируемому критерию, то соответствующий элемент матрицы равен 9.

Приведенное описание шкалы представлено в табл. 7.1.

Таблица 1

Шкала отношений интенсивности критериев оценки

Важность параметра оценки	1-9
Одинаковая важность	
Незначительное преимущество	
Значительное преимущество	
Явное преимущество	
Абсолютное преимущество	
2, 4, 6, 8 – промежуточные значения	

Таким образом, перед оценщиком ставится задача проанализировать при парном сравнении преимущества каждого из подходов по выделенным критериям второго уровня. Процесс заполнения матриц парных сравнений, несмотря на кажущуюся тривиальность, является довольно трудоемкой процедурой. Но сложности на этом этапе заключаются не в громоздких расчетах, а в многообразии анализируемой информации. Например, при анализе полноты исходной информации, использованной в рамках каждого из подходов, оценщику необходимо ответить на вопрос: исходные данные какого из подходов были наиболее полными? Естественно, что понятие полноты для различных подходов определяется различными требованиями.

Полнота исходных данных для затратного подхода определяется наличием, например, следующей информации:
- объемно-планировочные и конструктивно-планировочные характеристики объекта оценки;
- данные о проводимых капитальных ремонтах;
- технические заключения об авариях на объекте;
- объемно-планировочные и конструктивно-планировочные характеристики объектов-аналогов.

Полнота исходных данных для сравнительного подхода определяется наличием, например, следующей информации:
- данные по основным ценоформирующим факторам объектов-аналогов;
- информация о предпочтениях потребителей на данном рынке объектов недвижимости.

Полнота исходных данных для доходного подхода определяется наличием, например, следующей информации:
- информация о состоянии рынка аренды для аналогичных объектов недвижимости;
- информация о предпочтениях потребителей на данном рынке объектов недвижимости.

Конечно, при заполнении матриц парных сравнений определенная доля субъективизма присутствует. Но объективность конечного результата определения весовых коэффициентов обеспечивается за счет полностью формализованных процедур, а именно расчета отношения согласованности суждений.

Пусть $К_1 \ldots К_n$ - множество критериев из n элементов, тогда $W_1 \ldots W_n$ - интенсивности проявления этих элементов. Оценка весов критериев происходит по схеме, представленной в табл. 7.2.

При определении интенсивностей проявления критериев необходимо, чтобы соблюдались следующие обязательные условия:

$W_{k-1}/W_k + W_k/W_{k+1} > W_{k-1}/W_k$;

$W_{k-1}/W_k + W_k/W_{k+1} > W_k/W_{k+1}$,
где k Î [2; n].
Таблица 7.2
Схема оценки весов критериев

Критерий	K_1	K_2	…	K_n	Расчет	Вес критериев
K_1		W_1/W_2	...	W_1/W_n	$P_1 = \sqrt[n]{1 \cdot \frac{W_1}{W_2} \cdot \ldots \cdot \frac{W_1}{W_n}}$	B_{K1}
K_2	W_2/W_1		…	W_2/W_n	$P_2 = \sqrt[n]{\frac{W_2}{W_1} \cdot 1 \cdot \ldots \cdot \frac{W_2}{W_n}}$	B_{K2}
…	…	…	…	…	…	…
K_n	W_n/W_1	W_n/W_2	…		$P_n = \sqrt[n]{\frac{W_n}{W_1} \cdot \frac{W_n}{W_2} \cdot \ldots \cdot \frac{W_n}{W_n}}$	B_{Kn}
Сумма	S_{K1}	S_{K2}		S_{Kn}	$S = \sum_{i=1}^{n} P_i$	

Информацию о степени отклонения от согласованности матрицы дает отношение согласованности.
Обычно в качестве допустимого используется значение отношения согласованности на уровне 10 %. Если для какой-либо матрицы парных сравнений это отношение превышает 0,1, то это свидетельствует о существенном нарушении логичности суждений, допущенном оценщиком при заполнении матрицы, поэтому оценщику предлагается пересмотреть данные, использованные для построения матрицы, чтобы улучшить согласованность. Такая процедура предполагает заранее неизвестное число итераций пересмотра и изменения значений в матрицах парных сравнений с повторной проверкой на согласованность - до тех пор, пока не будет достигнут допустимый уровень согласованности оценок.

$$ИС = \frac{\lambda_{max} - n}{n-1}; \quad \lambda_{max} = \sum_{i=1}^{n} S_{Ki} B_{Ki},$$

где ИС - индекс согласованности; λ_{max} - максимальное собственное число матрицы; S_{Ki} - сумма интенсивностей проявления критериев по отношению к критерию K_i; B_{Ki} - вес критерия K_i.

1.5 Лекция 7 (2 ч.)

Тема: Регрессионные математические модели. Методы построения и статистической оценки. Оценка значимости коэффициентов, адекватности модели и ошибки прогнозирования. Задачи многофакторного моделирования.

1.5.1. Вопросы лекции:

1. Регрессионные модели прогнозирования
2. Расчет параметров уравнение регрессии. Технология работы
3. Интерпретация коэффициентов регрессии

1.5.2 Краткое содержание вопросов:

Регрессионные модели прогнозирования

В экономических исследованиях часто изучаются связи между случайными и неслучайными величинами. Такие связи называют *регрессионными*, а метод их изучения - *регрессионным анализом*.

Математически задача формулируется следующим образом. Требуется найти аналитическое выражение зависимости экономического явления (например, производительности труда) от определяющих его факторов; т.е. ищется функция $y=f(x_1,x_2,...,x_n)$, отражающая зависимость, по которой можно найти приближенное значение зависимого показателя y. В качестве функции в регрессионном анализе принимается случайная переменная, а аргументами являются неслучайные переменные.

Примерами возможного применения регрессионного анализа в экономике являются исследование влияния на производительность труда и себестоимость таких факторов, как величина основных производственных фондов, заработная плата и др.; влияние безработицы на изменение заработной платы на рынках труда (кривые Филипса); зависимость структуры расходов от уровня доходов (кривые Энгеля); функции потребления и спроса и многие другие.

При выборе вида регрессионной зависимости руководствуются следующим: он должен согласовываться с профессионально-логическими соображениями относительно природы и характера исследуемых связей; по возможности используют простые зависимости, не требующие сложных расчетов, легко экономически интерпретируемые и практически применимые.

Практика регрессионного анализа говорит о том, что уравнение линейной регрессии часто достаточно хорошо выражает зависимость между показателями даже тогда, когда на самом деле они оказываются более сложными. Это объясняется тем, что в пределах исследуемых величин самые сложные зависимости могут носить приближенно линейный характер.

В общей форме прямолинейное уравнение регрессии имеет вид

$$y=a_0+b_1*x_1+b_2*x_2+.....+b_m*x_m,$$

где y - результативный признак, исследуемая переменная;

x_i - обозначение фактора (независимая переменная);

m - общее число факторов;

a_0 - постоянный (свободный) член уравнения;

b_i - коэффициент регрессии при факторе.

Увеличение результативного признака y при изменении фактора x_i на единицу равно коэффициенту регрессии b_i (с положительным знаком); уменьшение - (с отрицательным знаком).

Очевидная экономическая интерпретация результатов линейной регрессии одна из основных причин ее применения в исследовании и прогнозировании экономических процессов.

В зависимости от числа факторов, влияющих на результативный показатель, различают парную и множественную регрессии.

Кратко изложим основные положения по разработке и использованию в прогнозировании множественных линейных регрессионных моделей (парная регрессия может быть рассмотрена как частный случай множественной). Экономические явления определяются, как правило, большим числом совокупно действующих факторов. В связи с этим часто возникает задача исследования зависимости одной переменной Y от нескольких объясняющих переменных $X_1,X_2,...X_n$. Эта задача решается с помощью множественного регрессионного анализа. Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели, включающего отбор факторов и выбор вида уравнения регрессии. Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям: они должны быть количественно измеримы (качественным факторам необ-

ходимо придать количественную определенность); между факторами не должно быть высокой корреляционной, а тем более функциональной зависимости, т.е. наличия мультиколлинеарности.

Включение в модель мультиколлинеарных факторов может привести к следующим последствиям: затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом виде», поскольку факторы связаны между собой; параметры линейной регрессии теряют экономический смысл; оценки параметров ненадежны, имеют большие стандартные ошибки и меняются с изменением объема наблюдений.

Пусть $Y = (y_1, y_2, \ldots, y_n)^T$ - матрица – столбец значений зависимой переменной размера n (значок «*T*» означает транспонирование);

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots x_{1m} \\ 1 & x_{21} & x_{22} & \ldots x_{2m} \\ \ldots & \ldots & \ldots & \ldots \ldots \\ 1 & x_{n1} & x_{n2} & \ldots x_{nm} \end{bmatrix}$$ - матрица объясняющих переменных;

$b = (b_0, b_1, \ldots, b_m)^T$ - матрица – столбец (вектор) параметров размера *m+1*;

$\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ - матрица – столбец (вектор) остатков размера *n*.

Тогда в матричной форме модель множественной линейной регрессии запишется следующим образом:

$$Y = Xb + \varepsilon$$

При оценке параметров уравнения регрессии (вектора b) применяется метод наименьших квадратов (МНК). При этом делаются определенные предпосылки.

1. В модели (5.2) ε – случайный вектор, X - неслучайная (детерминированная) матрица.
2. Математическое ожидание величины остатков равно нулю. $M(\varepsilon) = 0_n$.
3. Дисперсия остатков ε_i постоянна для любого *i* (условие гомоскедастичности), остатки ε_i и ε_j при $i \neq j$ не коррелированны: $M(\varepsilon\varepsilon^T) = \sigma^2 E_n$.
4. ε – нормально распределенный случайный вектор.
5. $r(X)=m+1<n$. Столбцы матрицы X должны быть линейно независимыми (ранг матрицы X максимальный, а число наблюдений n превосходит ранг матрицы).

Модель (5.2), в которой зависимая переменная, остатки и объясняющие переменные удовлетворяют предпосылкам 1-5 (предпосылки перечислены выше) называется классической нормальной линейной моделью множественной регрессии (КНЛММР). Если не выполняется только предпосылка 4, то модель называется классической линейной моделью множественной регрессии (КЛММР).

Согласно методу наименьших квадратов неизвестные параметры выбираются таким образом, чтобы сумма квадратов отклонений фактических значений от значений, найденных по уравнению регрессии, была минимальной:

$$S = \sum_{i=1}^{n} (y_i - y_{xi})^2 = (Y - Xb)(Y - Xb)^T \to \min \qquad (5.3)$$

Решением этой задачи является вектор $b = (X^T X)^{-1} X^T Y$.

Оценка качества регрессионного уравнения осуществляется по совокупности критериев, проверяющих адекватность модели фактическим условиям и статистической достоверности регрессии.

Одной из наиболее эффективных оценок адекватности модели является *коэффициент детерминации* R^2, определяемый по формуле

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_{xi})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(\hat{y}_{xi} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{RSS}{TSS},$$

где y_i – фактическое значение результирующего признака;

$\hat{y}_{xi}$ - значение результирующего признака, рассчитанное по полученной модели регрессии;

$\bar{y}$ - среднее значение признака;

RSS – объясненная сумма квадратов;

TSS – общая сумма квадратов.

R^2 характеризует долю вариации зависимой переменной, обусловленной регрессией или изменчивостью объясняющих переменных. Чем ближе R^2 к единице, тем лучше построенная регрессионная модель описывает зависимость между объясняющими и зависимой переменной. В случае $R^2 = 1$ изучаемую связь можно трактовать как функциональную (а не статистическую), что требует дополнительных качественных и количественных сведений и изменений в процессе исследования.

Следует иметь в виду, что при включении в модель новой объясняющей переменной, коэффициент детерминации увеличивается, хотя это и не обязательно означает улучшение качества регрессионной модели. В этой связи лучше использовать скорректированный (поправленный) коэффициент детерминации R^2, пересчитываемый по формуле:

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

где *n* – число наблюдений, *m* – число параметров при переменных *x*.

Таким образом, скорректированный коэффициент детерминации может уменьшаться при добавлении в модель новой объясняющей переменной, не оказывающей существенного влияния на результативный признак.

Средняя относительная ошибка аппроксимации рассчитывается по формуле:

$$\overline{A} = \frac{1}{n} \cdot \sum \left| \frac{y_i - \hat{y}_{xi}}{y_i} \right| \cdot 100\%.$$

Большинство авторов рекомендуют считать модель регрессии адекватной, если средняя относительная ошибка аппроксимации не превышает 12%.

Проверку значимости вида регрессионной зависимости можно осуществлять с применением дисперсионного анализа. Основной идеей этого анализа является разложение общей суммы квадратов отклонений результативной переменной *y* от среднего значения y на «объясненную» и «остаточную»:

$\sum (y - \bar{y})^2$	=	$\sum (\hat{y}_x - \bar{y})^2$	+	$\sum (y - \hat{y}_x)^2$.
общая сумма квадратов отклонений		*Сумма квадратов отклонений, объясненная регрессией*		*Остаточная сумма квадратов отклонений*

Для приведения дисперсий к сопоставимому виду, определяют дисперсии на одну степень свободы. Результаты вычислений заносят в специальную таблицу дисперсионного анализа В данной таблице *n* – число наблюдений, *m* – число параметров при переменных *x*. Сравнивая полученные оценки объясненной и остаточной дисперсии на одну степень свободы,

определяют значение *F*-критерия Фишера, используемого для оценки значимости уравнения регрессии:

$$F = \frac{\sigma_R^2}{\sigma_x^2}.$$

С помощью *F* – критерия проверяется нулевая гипотеза о равенстве дисперсий H_0: $\sigma_R^2=\sigma_x^2$. Если нулевая гипотеза справедлива, то объясненная и остаточная дисперсии не отличаются друг от друга.

Таблица 1 - Результаты дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Оценка дисперсии на одну степень свободы
Общая	$\sum(y-\bar{y})^2$	*n-1*	$\sigma^2 = \sum(y-\bar{y})^2 / n-1$
Объясненная	$\sum(\hat{y}_x-\bar{y})^2$	*n*	$\sigma_R^2 = \sum(\hat{y}_x-\bar{y})^2 / n$
Остаточная	$\sum(y-\hat{y}_x)^2$	*n-m-1*	$\sigma_x^2 = \sum(y-\hat{y}_x)^2 /(n-m-1)$

Для того, чтобы уравнение регрессии было значимо в целом (гипотеза H_0 была опровергнута) необходимо, чтобы объясненная дисперсия превышала остаточную в несколько раз. Критическое значение *F* – критерия определяется по таблице Фишера – Снедекора (приложение 1). $F_{табл}$ – максимально возможное значение критерия под влиянием случайных факторов при степенях свободы $k_1 = m$, $k_2 = n–m–1$ (для линейной регрессии $m = 1$) и уровне значимости α. Уровень значимости α – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно величина α принимается равной 0,05 или 0,01. Расчетное значение сравнивается с табличным: если оно превышает табличное ($F_{расч}>F_{табл}$), то гипотеза H_0 отвергается, и уравнение регрессии признается значимым. Если $F_{расч}<F_{табл}$, то уравнение регрессии считается статистически незначимым. Нулевая гипотеза H_0 не может быть отклонена.

Расчетное значение *F*-критерия связано с коэффициентом детерминации R^2 следующим соотношением:

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

где *m* –число параметров при переменных *x*;

n – число наблюдений.

Для оценки статистической значимости коэффициентов регрессии и коэффициента корреляции *r* ($r=\sqrt{R^2}$) применяется *t*-критерий Стьюдента.

Оценка значимости коэффициентов регрессии сводится к проверке гипотезы о равенстве нулю коэффициента регрессии при соответствующем факторном признаке, т.е. гипотезы: H_0: $b_i=0$ (5.10)

Проверка нулевой статистической гипотезы проводится с помощью *t* – критерия Стьюдента:

$$t_{bi} = \frac{b_i}{m_{b_i}},$$

где b_i – коэффициент регрессии при x_i,

m_{bi} – средняя квадратическая ошибка коэффициента регрессии b_i.

Средняя квадратическая ошибка коэффициента регрессии может быть определена по формуле:

$$m_{b_i} = \frac{\sigma_y \sqrt{1-R^2_{yx_1,\ldots x_m}}}{\sigma_{x_i}\sqrt{1-R^2_{x_i x_1,\ldots x_m}}} \cdot \frac{1}{\sqrt{n-m-1}}$$

где σ_y - среднее квадратическое отклонение для признака *y*;

σ_{x_i} - среднее квадратическое отклонение для признака x_i;

$R^2_{yx_1\ldots x_m}$ - коэффициент детерминации для уравнения множественной регрессии;

$R^2_{x_i x_1\ldots x_m}$ - коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии;

n-m-1 - число степеней свободы для остаточной суммы квадратов отклонений.

Использование данной формулы для расчета средней квадратической ошибки коэффициента регрессии предполагает расчет по матрице межфакторной корреляции соответствующих коэффициентов детерминации. Поэтому иногда рекомендуется использовать для определения средней квадратической ошибки коэффициента регрессии m_{bi}частные критерии Фишера.

Расчетное значение критерия Стьюдента сравнивается с табличным $t_{табл}$ при заданном уровне значимости $\alpha = 0{,}05$ (для экономических процессов и явлений) и числе степеней свободы, равном *n-2*. Если расчетное значение превышает табличное, то гипотезу о несущественности коэффициента регрессии b_i можно отклонить.

В линейной модели множественной регрессии $\hat{y}_x = b_0 + b_1 x_1 + \ldots + b_m \cdot \lambda_m$ коэффициенты регрессии b_i характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Значимость коэффициента корреляции *r* проверяется также на основе *t*-критерия Стьюдента. При этом выдвигается и проверяется гипотеза о равенстве коэффициента корреляции нулю (H_0: $r = 0$). При проверке этой гипотезы используется *t*-статистика:

$$t_p = \frac{|r|}{\sqrt{1-r^2}}\sqrt{n-2}.$$

При выполнении H_0 *t*-статистика имеет распределение Стьюдента с входными параметрами: $\alpha=0{,}05$; $k=n-2$. Если расчетное значение больше табличного, то гипотеза H_0 отвергается.

На практике часто бывает необходимо сравнить влияние на зависимую переменную различных объясняющих переменных, когда последние выражаются разными единицами измерения. В этом случае используют стандартизованные коэффициенты регрессии β_i и коэффициенты эластичности $Э_i$ ($i=1,2,\ldots,m$).

Уравнение регрессии в стандартизованной форме обычно представляют в виде:

$t_y = \beta \cdot t_{x1} + \beta_2 \cdot t_{x2} + \ldots + \beta_m \cdot t_{xm} + \varepsilon$, где $t_y = \frac{y-\bar{y}}{\sigma_y}, t_x = \frac{x_i - \bar{x}_i}{\sigma_{xi}}$ - стандартизованные переменные.

Заменив значения «*y*» на t_y, а значения «*x*» на t_x получаем нормированные или стандартизованные переменные. В результате такого нормирования средние значения всех стандартизованных переменных равны нулю, а дисперсии равны единице,

т.е. $\bar{t}_y = \bar{t}_{x1} = \ldots = \bar{t}_{xm} = 0,\ \sigma_1 = \sigma_{xi} = \ldots = \sigma_{xm} = 1.$

Коэффициенты обычной («чистой») регрессии связаны со стандартизованными коэффициентами следующим соотношением:

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

Стандартизованные коэффициенты могут принимать значения от -1 до +1 и показывают, на сколько стандартных отклонений (сигм) изменится в среднем результат, если соответствующий фактор x_i изменится на одно стандартное отклонение (одну сигму) при неизменном среднем уровне других факторов. Данные коэффициенты сохраняют свою величину при изменении масштаба шкалы. Сравнивая стандартизованные коэффициенты друг с другом, можно ранжировать факторы по силе их воздействия на результат.

Регрессия. Краткие сведения из теории статистики

В практике статистического исследования весьма часто возникает необходимость определить не только корреляционное соотношение между изучаемыми характеристиками, но и установить определенную обусловленность между ними, представив выявленную связь в строгой аналитической форме.

В этом случае результат исследования – экспериментальная зависимость воздействия какого-либо фактора (скажем, производительности труда, уровня образования, практического стажа работы и т.д.) на изменение изучаемого параметра (например, величины прибыли фирмы) − может быть не только представлен в виде графика (что весьма наглядно), но и описан математически с использованием аппроксимирующего выражения (эмпирической формулы).

Исследование такой ситуации и является задачей *регрессионного* анализа, который дает *предсказание (прогнозирование)* одной переменной на основании другой. Регрессионный анализ четко распределяет роли между изучаемыми характеристиками − одна из них является *аргументом*, а вторая *функцией*. Переменная, которая прогнозируется (функция), обозначается как *y*, а переменная, которая используется для такого прогнозирования (аргумент или фактор), − это *x*.

Таким образом, в случае выявления корреляции дается попытка ответить на вопрос: «Существует ли связь?» Целью регрессионного анализа является поиск ответа на уже более сложный вопрос: «Каков вид этой связи? Что на что влияет?» Однако в последнем случае речь не идет о выяснении механизма причинности обнаруженной связи, т.е. не ставится вопрос «Почему существует связь?» Это уже считается проблемой специального исследования, касающегося выявления физической (или социальной) природы изучаемого процесса.

Форма связи результативного признака *Y с* факторами $X_1, X_2, \ldots X_m$ получила название *уравнения регрессии.* В зависимости от типа выбранного уравнения *различают линейную* и*нелинейную* регрессию (в последнем случае возможно дальнейшее уточнение: квадратичная, экспоненциальная, логарифмическая и т. д.).

В зависимости от числа взаимосвязанных признаков различают *парную* и *множественную* регрессию» Если исследуется связь между двумя признаками (результативным и факторным), то регрессия называется *парной,* если между тремя и более признаками - *множественной (многофакторной)* регрессией.

При изучении регрессии следует придерживаться определенной последовательности этапов:

1. Задание аналитической формы уравнения регрессии и определение параметров регрессии.
2. Определение в регрессии степени стохастической взаимосвязи результативного признака и факторов, проверка общего качества уравнения регрессии.
3. Проверка статистической значимости каждого коэффициента уравнения регрессии и определение их доверительных интервалов. Основное содержание выделенных этапов рассмотрим на примере множественной линейной регрессии, реализованной в режиме «Регрессия» надстройки Пакет анализа Microsoft Excel.

Расчет параметров уравнение регрессии. Технология работы

Для расчета параметров уравнения регрессии воспользуемся следующим **примером**. В аптеке продается новый препарат для профилактики гриппа. Необходимо выяснить как объем продаж *y* (число упаковок в день) зависит от а) числа покупателей, которые слышали рекламу этого препарата (их доля от общего числа покупателей x_1, %) и работе в торговом зале врача-консультанта (относительное время x_2, когда он работал, %). Исходные данные представлены в таблице.

Таблица 2

День продажи	*y*, *уп/день*	x_1, %	x_2, %	День продажи	*y*, *уп/день*	x_1, %	x_2, %
1	6	40	30	11	8	50	35
2	5	20	33	12	8	37	30
3	4	31	20	13	7	50	40
4	5	32	25	14	7	38	42
5	6	34	29	15	7	50	39
6	5	35	20	16	6	35	35
7	5	37	21	17	6	46	36
8	5	32	20	18	6	49	38
9	7	39	35	19	7	51	41
10	5	35	30	20	6	45	34

Разместим исходные данные на лист табличного редактора в следующем виде (рис. 25).

	A	B	C	D
1	*День продажи*	*y, уп/день*	x_1, %	x_2, %
2	1	6	40	30
3	2	5	20	33
4	3	4	31	20
5	4	5	32	25
6	5	6	34	29
7	6	5	35	20
8	7	5	37	21
9	8	5	32	20
10	9	7	39	35
11	10	5	35	30
12	11	8	50	35
13	12	8	37	30
14	13	7	50	40
15	14	7	38	42
16	15	7	50	39
17	16	6	35	35
18	17	6	46	36
19	18	6	49	38
20	19	7	51	41
21	20	6	45	34
22				

Рис. 25

Режим работы «Регрессия» служит для расчета параметров *уравнения линейной* регрессии

и проверки его адекватности исследуемому процессу. В диалоговом окне данного режима (рис. 26) задаются следующие параметры:

1. *Входной интервал Y* – вводится ссылка на ячейки, содержащие данные по результативному признаку Диапазон должен состоять из одного столбца.

2. *Входной интервал X* – вводится ссылка на ячейки, содержащие факторные признаки. Максимальное число входных диапазонов (столбцов) равно 16.

3. *Метки в первой строке/Метки в первом столбце*

4. *Уровень надежности* – установите данный флажок в активное состояние, если в поле, расположенное напротив флажка, необходимо ввести уровень надежности, отличный от уровня 95 %, применяемого по умолчанию. Установленный уровень надежности используется для проверки значимости коэффициента детерминации R^2и коэффициентов регрессии a_i.

5. Выходной интервал/Новый рабочий лист/Новая рабочая книга

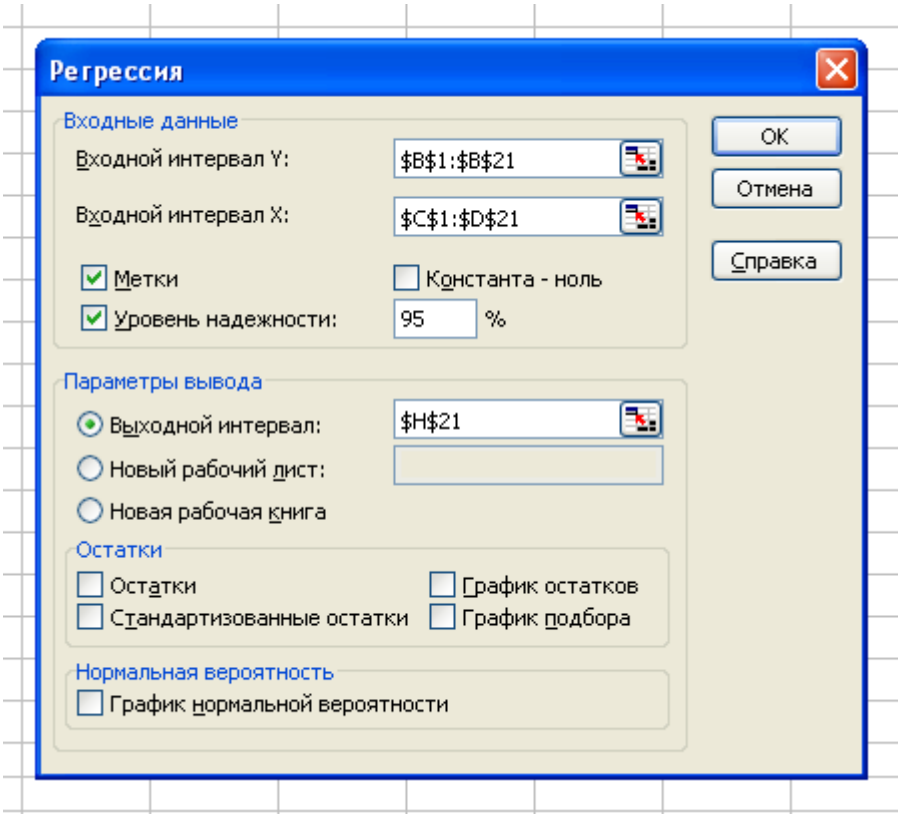


Рис. 26

После того, как мы нажмем на ОК, получим результат, содержащий большое количество информации (рис. 27). Но выберем только те из них, которые потребуются для последующего анализа. Для этого создадим таблицу, в которой поместим *расчетные* значения коэффициентов регрессии, стандартную ошибку, величины *t*-критерия и показатели уровня значимости *p*. Укажем также (ниже таблицы) рассчитанные показатели для самой функции *y*.

Регрессионная статистика	
Множественный R	0,75
R-квадрат	0,56
Нормированный R-квадрат	0,51
Стандартная ошибка	0,77
Наблюдения	20

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	12,92	6,46	10,95	0,001
Остаток	17	10,03	0,59		
Итого	19	22,95			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	1,91	0,92	2,09	0,05	-0,02	3,84	-0,02	3,84
x1, %	0,04	0,03	1,64	0,12	-0,01	0,10	-0,01	0,10
x2, %	0,08	0,03	2,48	0,02	0,01	0,14	0,01	0,14

Рис. 27

Таблица 3

Данные регрессионной статистики

Независимая переменная	Коэффициент	Стандартная ошибка	*Критерий t*	*Уровень значимости p*
Свободный член	1,91	0,92	2,09	0,05
Доля покупателей, слышавших рекламу, %	0,04	0,03	1,64	0,12
Время работы консультанта, %	0,08	0,03	2,48	0,02

Для функции *Y*: стандартная ошибка $S_{\bar{y}}$ = 0,77; R^2-квадрат = 0,56; R^2 (нормированный) = 0,51. Таким образом, для рассматриваемого примера уравнение регрессии (или уравнение прогнозирования) будет иметь следующий вид: *y*(*объем продаж, уп/день*) $=b_0+b_1x_1+b_2x_2$=1,91+0,04 (*доля покупателей, слышавших рекламу, %*)+1,62(*относительное время работы консультанта, %*). Запишем полученное уравнение в окончательной редакции: y=1,91+0,04x_1+0,08x_2

Интерпретация коэффициентов регрессии

Свободный член (сдвиг) b_0, равный 1,91, формально надлежит понимать следующим образом: объем продажи препарата, когда среди покупателей нет тех, кто слышал рекламу, и нет врача консультанта, составляет 1,91 упаковки в день. Однако мы полагаем, что в указанной совокупности исходных данных нет подобных примеров. Поэтому сдвиг b_0 следует обсуждать как вспомогательную величину, необходимую для получения оптимальных прогнозов, и не истолковывать ее буквально. Коэффициенты регрессии b_1 и b_2 следует рассматривать как степень влияния каждой из переменных на уровень продаж препарата, если все другие независимые переменные остаются неизменными. Так, коэффициент b_1, равный 0,04, указывает, что (при прочих равных условиях) повышение доли покупателей, слышавших о рекламе препарата на 1 %, приводит к возрастанию его продажи на 0,04 упаковок в день. Анализируя коэффициент b_2, можно заметить, что увеличение относительного участия врача-консультанта на 1 % приводит также к повышению продажи, этот

прирост составляет почти такую же величину − 0,08 уп/день.
Еще раз заметим, что все названные коэффициенты регрессии отражают влияние на исследуемый параметр *y* только какой-то одной переменной *x*при непременном условии, что все другие переменные (факторы) не меняются. Например, применительно к коэффициенту b_2 это нужно понимать так: указанное влияние коммерческой помощи консультанта проявляется при условии, что среди покупателей сохраняется неизменной доля тех, кто слышал рекламу.

Ошибки прогнозирования (определение качества регрессионного анализа)

Для оценки выполненного регрессионного анализа используют *коэффициент детерминации* (R^2), указывающий, какой процент вариации функции *y* объясняется воздействием факторов x_k.
В нашем примере коэффициент детерминации R^2 равен 0,56, что составляет 56 %. Этот результат следует толковать так: все исследуемые воздействующие факторы объясняют 56 % вариации анализируемой функции (объема проданных препаратов). Остальное же (44 %, что очень существенно) остается необъясненным и может быть связано с влиянием других, неучтенных факторов.
Итак, нами получено уравнение множественной регрессии, коэффициенты которого *b*i формально показывают, как и в каком направлении действуют (пока лишь вероятно!) исследуемые факторы x_{ki}и какой процент изменчивости функции *y* объясняется влиянием именно этих факторов.
Теперь нам надлежит определить статистическую значимость полученного аналитического выражения.

Проверка значимости модели

Решение принимается на основе коэффициента детерминации R^2. В этом случае имеющуюся расчетную величину R^2расч (рис.27) необходимо сравнить с табличными (критическими) значениями R^2крит для соответствующего уровня значимости (повторим еще раз, обычно это 0,05). Если окажется, что R^2расч> R^2крит, то с упомянутой степенью вероятности (95 %) можно утверждать, что анализируемая регрессия является значимой.
Напомним, что, по нашим расчетам, коэффициент детерминации R^2расч составляет 0,56, или 56 %. Таблица для тестирования на уровне значимости 5 % в случае выборки $n = 20$ и числа переменных $k = 2$ дает критическое значение R^2крит = 0,297. Поскольку выполняется соотношение R^2расч> R^2крит, то с вероятностью 95 % можно утверждать о наличии значимости данного уравнения регрессии.

Проверка на адекватность коэффициентов регрессии

Проверку на адекватность коэффициентов регрессии рекомендуется проводить по следующим эквивалентным методам.
1. *Использование t-критерия*. Анализируемый коэффициент считается значимым, если его *t-критерий* по абсолютной величине превышает 2,00 (точнее 1,96), что соответствует уровню значимости 0,05. В нашем примере имеем для коэффициентов b_0, b_1 и b_2 следующие показатели критерия Стьюдента: $t\ b_0 = 2{,}09$; $t\ b_1 = 1.64$ и $t\ b_2 = 2.48$. Из всего вышесказанного следует, что значимыми оказываются коэффициенты b_0 и b_2 нашего уравнения.
2. *Использование уровня значимости*. В этом случае оценка проводится путем анализа показателя *p.* Коэффициент признается значимым, если рассчитанное для него *p*-значение (эти данные выдает Excel) меньше (или равно) 0,05 (т.е. для 95 %-ной доверительной вероятности). Показатель *p* составляет для коэффициентов следующие величины: $p\ b_0 = 0{,}05$; $p\ b_1 = 0{,}12$ и $p\ b_2 = 0{,}02$. Эти данные позволяют также заключить, что из рассмотренных коэффициентов статистически значимыми являются первый и третий.
В целом же проведенный регрессионный анализ дает нам основание сделать вывод о том, что влияние на продажи работы врача-консультанта значимо. Вместе с тем необходимо

учитывать, что нами явно не приняты во внимание все факторы (вспомним про 44 %, приходящихся на неучтенные причины), что требует проведения дополнительных исследований.

Поиск закономерностей для качественных данных. Анализ «хи-квадрат»

Критерий *хи-квадрат* используют для проверки гипотез о качественных данных, представленных *не* числами, а категориями. Здесь принято оперировать подсчетом *частоты*(поскольку ранжирование или арифметические действия выполнять невозможно).

Критерий (тест) «хи-квадрат» основан на частотах, которые представляют собой количество объектов выборки, попадающих в ту или иную категорию. Суть показателя *хи-квадрат*(χ^2):

он измеряет *разницу* между *наблюдаемыми* (экспериментальными) *частотами* $f_Э$ и *ожидаемыми* (теоретическими) *частотами* f_T. Конкретно он рассчитывается как сумма квадратов разности этих частот, выраженная в долях частоты теоретической. Это утверждение можно записать следующим образом: $\chi^2 = \sum \frac{(f_э - f_т)}{f_т}$

Использование этого статистического подхода рассмотрим на следующем примере. Мы решили провести маркетинговое исследование, чтобы уяснить, какую марку минеральной воды предпочитают мужчины и женщины. Для каждой покупки фиксировались две качественные переменные – марка воды и пол покупателя. В качестве продаваемой воды фигурировали «Нарзан», «Ессентуки» и «Тагарская».

Полученные данные статистического опроса представлены в табличной форме (табл.4), в которой для каждого вида минеральной воды указано количество совершаемых покупок тем или иным покупателем.

Необходимо дать заключение по итогам статистической проверки по критерию «хи-квадрат», т.е. сформулировать вывод и пояснить результат с практической точки зрения – определить, какую рыночную стратегию необходимо принять, т.е., на какого покупателя и на какую марку минеральной воды необходимо ориентироваться.

Таблица 4.– Экспериментальные данные о результатах опроса посетителей аптеки

Марка воды	Частота предпочтений		
	Мужчины	Женщины	Итого
Нарзан	38	15	53
Ессентуки	24	31	55
Тагарская	18	27	45
Итого	80	73	153

Чисто визуально трудно ответить, есть ли взаимосвязь между этими признаками: разными категориями покупателей и марками минеральной воды. Поэтому необходимо дать анализ распределения частот в таблице по строкам и графам.

При этом исходят из следующего положения. Если признак, положенный в основу группировки по строкам (*марка минеральной воды*), *не* зависит от признака, положенного в основу группировки по столбцам (пол *покупателя*), то в *каждой* строке (столбце) распределение частот должно быть пропорционально распределению их в *итоговой* строке (столбце). Такое распределение можно рассматривать как *теоретическое* (ожидаемое), частоты которого рассчитаны в предположении *отсутствия* связи между изучаемыми

совокупностями.
Рассчитаем *ожидаемые* частоты *внутри* таблицы пропорционально распределению частот в *итоговой* строке.
Так, *«Нарзан»* как одна из марок минеральной воды в зависимости от поведения посетителей аптеки по частоте попадания в категории «Мужчины» и «Женщины» имеет следующие показатели:

$$f_{11} = \frac{53 \times 80}{153} = 27,7 \; ; \; f_{12} = \frac{53 \times 73}{153} = 25,3$$

Для второй строки, т.е. для воды «Ессентуки», эти показатели имеют следующие значения:

$$f_{12} = \frac{55 \times 80}{153} = 28,8 \; ; \; f_{12} = \frac{55 \times 73}{153} = 26,2$$

Для третей строки – категория «Тагарская»:

$$f_{13} = \frac{45 \times 80}{153} = 23,5 \; ; \; f_{13} = \frac{45 \times 73}{153} = 21,5$$

Полученные результаты поместим в таблицу 5.
Таблица 5. – Теоретические данные о результатах опроса посетителей аптеки

Марка воды	Частота предпочтений		
	Мужчины	Женщины	Итого
Нарзан	27,7	25,3	53
Ессентуки	28,8	26,2	55
Тагарская	23,5	21,5	45
Итого	80	73	153

Анализ «хи-квадрат. Технология расчета

Перенесем данные обеих таблиц в рабочий лист Excel (рис.28).
Анализ *хи-квадрат* выполним с помощью функции **ХИ^{2}тест**. Для этого используем **Мастер функций/ Статистические / ХИ^{2}тест.**
В ячейку B11 поместим результаты теста. При заполнении диалогового окна в текстовом поле *фактического интервала* укажем адрес ячеек C4:D6, в которых находятся экспериментальные данные по частотам (табл.4). Соответственно в текстовом поле *ождаемого интервала* укажем диапазон H4:I6, содержимое которого отражает теоретические значения частот (табл.5). В окончательном виде в ячейке B11 будет находиться следующий показатель, а именно: 0,002.
Как же следует трактовать полученный результат? Тезис о независимости обсуждаемых параметров (вид минеральной воды и пол покупателя) можно было бы принять, если бы уровень значимости α был бы меньше 0,002. Но для 95 %−ной вероятности (даже 99-процентной) установленные значения α (0,05 и 0,01) превышают 0,002. Это говорит о высокой степени значимости, следовательно, указанные качественные переменные являются зависимыми друг от друга. Другими словами, предпочтение в выборе той или иной марки минеральной воды определяется полом покупателя.

Марка воды	Частота предпочтений		Итого
	Мужчины	Женщины	
Нарзан	38	15	53
Ессентуки	24	31	55
Тагарская	18	27	45
Итого	80	73	153

Марка воды	Частота предпочтений		Итого
	Мужчины	Женщины	
Нарзан	27,7	25,3	53
Ессентуки	28,8	26,2	55
Тагарская	23,5	21,5	45
Итого	80	73	153

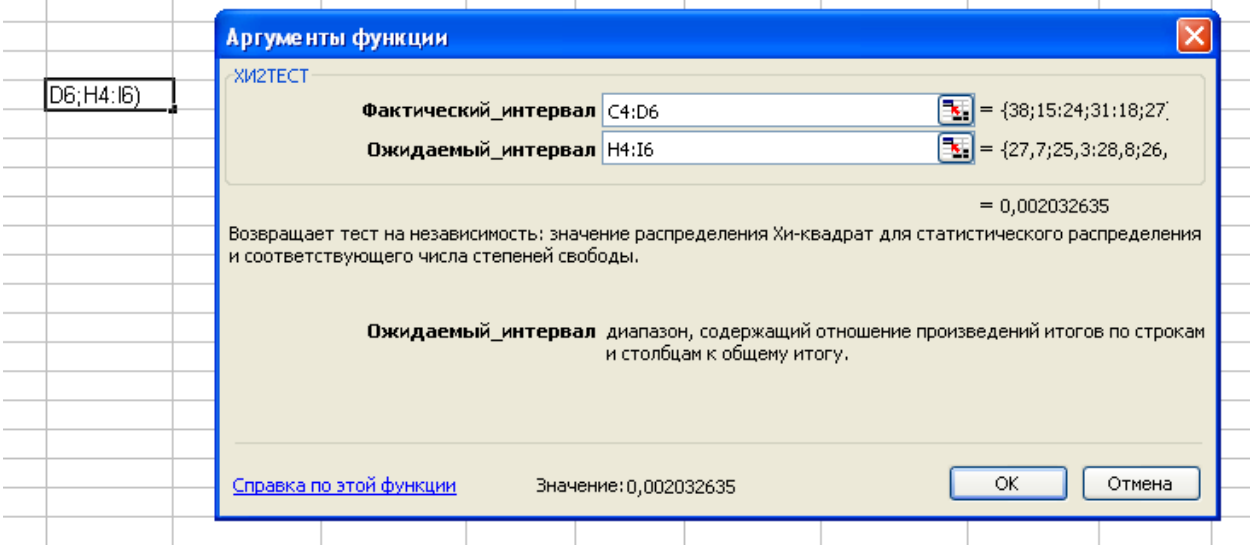


Рис.28

1.6 Лекция 8 (2 ч.)

Тема: Методы оптимизации. Использование надстроек Microsoft Excel

1.6.1. Вопросы лекции:

1. ЗЛП. Графический метод решения ЗЛП, симплекс-метод
2. Постановка задачи линейного программирования и свойства ее решений
3. Симплексный метод решение ЗЛП
4. Теория двойственности
5. Основные виды задач, сводящихся к ЗЛП
6. Транспортная задача, методы ее решения.

1.6.2 Краткое содержание вопросов

ЗЛП. Графический метод решения ЗЛП

Многие задачи, с которыми приходится иметь дело в повседневной практике, являются многовариантными. Среди множества возможных вариантов в условиях рыночных отношений приходится отыскивать наилучшие в некотором смысле при ограничениях, налагаемых на природные, экономические и технологические возможности. В связи с этим возникла необходимость применять для анализа и синтеза экономических ситуаций и систем математические методы и современную вычислительную технику? Такие методы объединяются под общим названием — математическое программирование.

Основные понятия

Математическое программирование — область математики, разрабатывающая теорию и численные методы решения многомерных экстремальных задач с ограничениями, т. е.

задач на экстремум функции многих переменных с ограничениями на область изменения этих переменных.

Функцию, экстремальное значение которой нужно найти в условиях экономических возможностей, называют *целевой, показателем эффективности* или *критерием оптимальности.* Экономические возможности формализуются в виде *системы ограничений.* Все это составляет математическую модель. *Математическая модель* задачи — это отражение оригинала в виде функций, уравнений, неравенств, цифр и т. д. Модель задачи математического программирования включает:

1) совокупность неизвестных величин, действуя на которые, систему можно совершенствовать. Их называют *планом задачи* (вектором управления, решением, управлением, стратегией, поведением и др.);

2) целевую функцию (функцию цели, показатель эффективности, критерий оптимальности, функционал задачи и др.). Целевая функция позволяет выбирать наилучший вариант - из множества возможных. Наилучший вариант доставляет целевой функции экстремальное значение. Это может быть прибыль, объем выпуска или реализации, затраты производства, издержки обращения, уровень обслуживания или дефицитности, число комплектов, отходы и т. д.;

Эти условия следуют из ограниченности ресурсов, которыми располагает общество в любой момент времени, из необходимости удовлетворения насущных потребностей, из условий производственных и технологических процессов. Ограниченными являются не только материальные, финансовые и трудовые ресурсы. Таковыми могут быть возможности технического, технологического и вообще научного потенциала. Нередко потребности превышают возможности их удовлетворения. Математически ограничения выражаются в виде уравнений и неравенств. Их совокупность образует *область допустимых решений (область экономических возможностей).* План, удовлетворяющий системе ограничений задачи, называется *допустимым*. Допустимый план, доставляющий функции цели экстремальное значение, называется *оптимальным.* Оптимальное решение, вообще говоря, не обязательно единственно, возможны случаи, когда оно не существует, имеется конечное или бесчисленное множество оптимальных решений.

Один из разделов математического программирования - *линейным программированием*. Методы и модели линейного программирования широко применяются при оптимизации процессов во всех отраслях народного хозяйства: при разработке производственной программы предприятия, распределении ее по исполнителям, при размещении заказов между исполнителями и по временным интервалам, при определении наилучшего ассортимента выпускаемой продукции, в задачах перспективного, текущего и оперативного планирования и управления; при планировании грузопотоков, определении плана товарооборота и его распределении; в задачах развития и размещения производительных сил, баз и складов систем обращения материальных ресурсов и т. д. Особенно широкое применение методы и модели линейного программирования получили при решении задач экономии ресурсов (выбор ресурсосберегающих технологий, составление смесей, раскрой материалов), производственно-транспортных и других задач.

Начало линейному программированию было положено в 1939 г. советским математиком-экономистом Л. В. Канторовичем в работе «Математические методы организации и планирования производства». Появление этой работы открыло новый этап в применении математики в экономике. Спустя десять лет американский математик Дж. Данциг разработал эффективный метод решения данного класса задач — симплекс-метод. Общая идея *симплексного метода (метода последовательного улучшения плана)* для решения ЗЛП состоит в следующем:

1) умение находить начальный опорный план;

2) наличие признака оптимальности опорного плана;

3) умение переходить к нехудшему опорному плану.

Постановка задачи линейного программирования и свойства ее решений

Линейное программирование — раздел математического программирования, применяемый при разработке методов отыскания экстремума линейных функций нескольких переменных при линейных дополнительных ограничениях, налагаемых на переменные. По типу решаемых задач его методы разделяются на универсальные и специальные. С помощью универсальных методов могут решаться любые *задачи линейного программирования* (ЗЛП). Специальные методы учитывают особенности модели задачи, ее целевой функции и системы ограничений.

Особенностью задач линейного программирования является то, что экстремума целевая функция достигает на границе области допустимых решений. Классические же методы дифференциального исчисления связаны с нахождением экстремумов функции во внутренней точке области допустимых значений. Отсюда — необходимость разработки новых методов.

Формы записи задачи линейного программирования:

Общей задачей линейного программирования называют задачу

$$\max(\min)\ Z = \sum_{j=1}^{n} c_j x_j \qquad (2.1)$$

при ограничениях

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad (i = 1,\dots,m_1) \qquad (2.2)$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i (i = m_1 + 1,\dots, m_2) \qquad (2.3)$$

$$\sum_{j=1}^{n} a_{ij} x_j \ge b_i (i = m_2 + 1,.., m) \qquad (2.4)$$

$$x_j \ge 0 \qquad (j = \overline{1, n_1}) \qquad (2.5)$$

x_j - произвольные $(j = n_1 + 1,\dots, n)$ (2.6)

где c_j, a_{ij}, b_i - заданные действительные числа; (2.1) – целевая функция; (2.1) – (2.6) – ограничения; $\vec{x} = (x_1;\dots; x_n)$ - план задачи.

Пусть ЗЛП представлена в следующей записи:

$$\max Z = cx \qquad (2.7)$$

$$A_1 x_1 + A_2 x_2 + \dots + A_n x_n = A_0 \qquad (2.8)$$

$$x_1 \ge 0, x_2 \ge 0,\dots, x_n \ge 0 \qquad (2.9)$$

Чтобы задача (2.7) – (2.8) имела решение, системе её ограничений (2.8) должна быть совместной. Это возможно, если r этой системы не больше числа неизвестных n. Случай $r>n$ вообще невозможен. При $r=n$ система имеет единственное решение, которое будет при $x_j \ge 0 \ \ (j = \overline{1, n})$ оптимальным. В этом случае проблема выбора оптимального решения теряет смысл. Выясним структуру координат угловой точки многогранных решений. Пусть $r<n$. В этом случае система векторов $A_1, A_2,\dots, A_n$ содержит базис — максимальную линейно независимую подсистему векторов, через которую любой вектор системы может быть выражен как ее линейная комбинация. Базисов, вообще говоря, может быть несколько, но не более C_n^r. Каждый из них состоит точно из r векторов. Переменные ЗЛП, соответствующие r векторам базиса, называют, как известно, *базисными* и обозначают БП. Остальные $n - r$ переменных будут *свободными,* их обозначают СП. Не ограничивая общности, будем считать, что базис составляют первые m векторов $A_1, A_2,\dots, A_m$. Этому базису соответствуют базисные переменные $x_1, x_2,\dots, x_m$, а свободными будут переменные $x_{m+1}, x_{m+2},\dots, x_n$.

Если свободные переменные приравнять нулю, а базисные переменные при этом примут неотрицательные значения, то полученное частное решение системы (8) называют *опорным решением (планом).*

Теорема. Если система векторов $A_1, A_2, \ldots, A_n$ содержит m линейно независимых векторов $A_1, A_2, \ldots, A_m$, то допустимый план

$$\vec{x} = (x_1; x_2; \ldots; x_m; \underbrace{0;0;\ldots;0}_{n-m}) \qquad (2.\ 10)$$

является крайней точкой многогранника планов.

Теорема. Если ЗЛП имеет решение, то целевая функция достигает экстремального значения хотя бы в одной из крайних точек многогранника решений. Если же целевая функция достигает экстремального значения более чем в одной крайней точке, то она достигает того же значения в любой точке, являющейся их выпуклой линейной комбинацией.

Графический способ решения ЗЛП

Геометрическая интерпретация экономических задач дает возможность наглядно представить, их структуру, выявить особенности и открывает пути исследования более сложных свойств. ЗЛП с двумя переменными всегда можно решить графически. Однако уже в трехмерном пространстве такое решение усложняется, а в пространствах, размерность которых больше трех, графическое решение, вообще говоря, невозможно. Случай двух переменных не имеет особого практического значения, однако его рассмотрение проясняет свойства ОЗЛП, приводит к идее ее решения, делает геометрически наглядными способы решения и пути их практической реализации.

Пусть дана задача

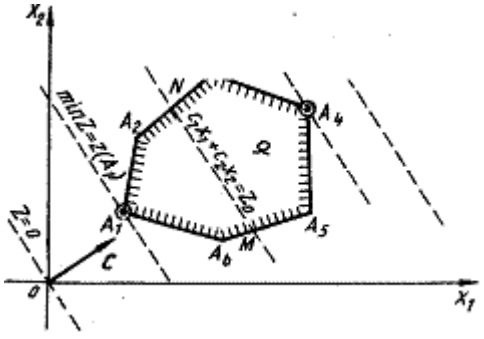
$$\max Z = c_1x_1 + c_2x_2 \qquad (2.11)$$

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 \le b_1 \\ a_{21}x_1 + a_{22}x_2 \le b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 \le b_m \end{array}\right\} \qquad (2.12)$$

$$x_1 \ge 0, x_2 \ge 0 \qquad (2.13)$$

Дадим геометрическую интерпретацию элементов этой задачи. Каждое из ограничений (2.12), (2.13) задает на плоскости x_1Ox_2 некоторую полуплоскость. Полуплоскость — выпуклое множество. Но пересечение любого числа выпуклых множеств является выпуклым множеством. Отсюда следует, что область допустимых решений задачи (2.11) — (2.13) есть выпуклое множество.

Перейдем к геометрической интерпретации целевой функции. Пусть область допустимых решений ЗЛП — непустое множество, например многоугольник $A_1A_2A_3A_4A_5A_6$.



Выберем произвольное значение целевой функции $Z = Z_0$. Получим $c_1x_1 + c_2x_2 = Z_0$. Это уравнение прямой линии. В точках прямой *NM* целевая функция сохраняет одно и то же постоянное значение Z_0. Считая в равенстве (2.11) Z параметром, получим уравнение семейства параллельных прямых, называемых линиями уровня целевой функции (линиями постоянного значения).

Найдём частные производные целевой функции по x_1 и x_2:

$$\frac{\partial Z}{\partial x_1} = c_1, \qquad (2.14)$$

$$\frac{\partial Z}{\partial x_2} = c_2. \qquad (2.15)$$

Частная производная (2.14) (так же как и (2.15)) функции показывает скорость ее возрастания вдоль данной оси. Следовательно, c_1 и c_2 — скорости возрастания Z соответственно вдоль осей Ox_1 и Ox_2. Вектор $\vec{c} = (c_1, c_2)$ называется градиентом функции. Он показывает направление наискорейшего возрастания целевой функции:

$$\vec{c} = (\partial Z / \partial x_1, \partial Z / \partial x_2)$$

Вектор $(-\vec{c})$ указывает направление наискорейшего убывания целевой функции. Его называют антиградиентом.

Вектор $\vec{c} = (c_1, c_2)$ перпендикулярен к прямым $Z = const$ семейства $c_1 x_1 + c_2 x_2 = Z$.

Из геометрической интерпретации элементов ЗЛП вытекает следующий порядок ее графического решения.

1. С учетом системы ограничений строим область допустимых решений Ω.

2. Строим вектор $\vec{c} = (c_1, c_2)$ наискорейшего возрастания целевой функции — вектор градиентного направления.

3. Проводим произвольную линию уровня $Z = Z_0$.

4. При решении задачи на максимум перемещаем линию уровня $Z = Z_0$ в направлении вектора c так, чтобы она касалась области допустимых решений в ее крайнем положении (крайней точке). В случае решения задачи на минимум линию уровня $Z = Z_0$ перемещают в антиградиентном направлении.

5. Определяем оптимальный план $\vec{x}^* = (x_1^*, x_2^*)$ и экстремальное значение целевой функции $Z^* = z(\vec{x}^*)$.

Симплексный метод решение ЗЛП

Общая идея симплексного метода (метода последовательного улучшения плана) для решения ЗЛП состоит

1) умение находить начальный опорный план;

2) наличие признака оптимальности опорного плана;

3) умение переходить к нехудшему опорному плану.

Пусть ЗЛП представлена системой ограничений в каноническом виде:

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad b_i \geq 0 \qquad (i = 1, \ldots, m).$$

Говорят, что ограничение ЗЛП имеет предпочтительный вид, если при неотрицательной правой части $(b_i \geq 0)$ левая часть ограничений содержит переменную, входящую с коэффициентом, равным единице, а в остальные ограничения равенства - с коэффициентом, равным нулю.

Пусть система ограничений имеет вид

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i, \qquad b_i \geq 0 \qquad (i = 1, \ldots, m)$$

Сведем задачу к каноническому виду. Для этого прибавим к левым частям неравенств дополнительные переменные $x_{n+1} \geq 0 \ \ (i = 1,..,m)$. Получим систему, эквивалентную исходной:

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+1} = b_i, \qquad b_i \geq 0 \qquad (i = 1,..,m),$$

$$\vec{x}_o = (0; \underbrace{0;...;0}_{n}; \underbrace{b_1; b_2;...; b_m}_{m})$$

которая имеет предпочтительный вид .

В целевую функцию дополнительные переменные вводятся с коэффициентами, равными нулю $c_{n+1} = 0 \ \ (i = 1,..,m)$.

Пусть далее система ограничений имеет вид

$$\sum_{j=1}^{n} a_{ij} x_j \geq b_i, \qquad b_i \geq 0 \qquad (i = 1,..,m).$$

Сведём её к эквивалентной вычитанием дополнительных переменных $x_{n+1} \geq 0$ $(i = 1,..,m)$ из левых частей неравенств системы. Получим систему

$$\sum_{j=1}^{n} a_{ij} x_j - x_{n+1} = b_i, \qquad b_i \geq 0 \qquad (i = 1,...,m).$$

Однако теперь система ограничений не имеет предпочтительного вида, так как дополнительные переменные x_{n+1} входят в левую часть (при $b_i \geq 0$) с коэффициентами, равными –1.

$$\vec{x}_o = (\underbrace{0;...;0;}_{n} -b_1; -b_2;...; -b_m)$$

Поэтому, вообще говоря, базисный план не является допустимым. В этом случае вводится так называемый искусственный базис. К левым частям ограничений-равенств, не имеющих предпочтительного вида, добавляют искусственные переменные ω_i. В целевую функцию переменные ω_i, вводят с коэффициентом *M* в случае решения задачи на минимум и с коэффициентом - *M* для задачи на максимум, где *M* - большое положительное число. Полученная задача называется *M*-задачей, соответствующей исходной. Она всегда имеет предпочтительный вид.

Пусть исходная ЗЛП имеет вид

$$\max(\min) \ \ Z = \sum_{j=1}^{n} c_j x_j, \tag{2.16}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad b_i \geq 0 \qquad (i = 1,..,m), \tag{2.17}$$

$$x_j \geq 0 \ \ (j = 1,..,n), \tag{2.18}$$

причём ни одно из ограничений не имеет предпочтительной переменной. М-задача запишется так:

$$\max(\min) \ \ \overline{Z} = \sum_{j=1}^{n} c_j x_j - (+) \sum_{i=1}^{m} M \omega_i \tag{2.19}$$

$$\sum_{j=1}^{n} a_{ij} x_j + \omega_i = b_i, \qquad (i = 1,..,m) \tag{2.20}$$

$$x_j \geq 0 \ \ (j = \overline{1,n}), \qquad \omega_i \geq 0, \qquad (i = 1,..,m) \tag{2.21}$$

Задача (2.19) - (2.21) имеет предпочтительный план. Её начальный опорный план имеет вид

$$\vec{x}_o = (\underbrace{0;0;...;0;}_{n} b_1; b_2;...; b_m)$$

Если некоторые из уравнений (2.17) имеют предпочтительный вид, то в них не следует вводить искусственные переменные.

Теорема. Если в оптимальном плане

$$\vec{x} = (x_1; x_2; \ldots; x_n; \omega_1; \omega_2; \ldots; \omega_m) \qquad (2.22)$$

М-задачи (2.19) - (2.21) все искусственные переменные $\omega_i = 0 \ \ (i = 1,..,m)$, то план $\vec{x} = (x_1; x_2; \ldots; x_n)$ является оптимальным планом исходной задачи (2.16) - (2.18).

Для того чтобы решить задачу с ограничениями, не имеющими предпочтительного вида, вводят искусственный базис и решают расширенную М - задачу, которая имеет начальный опорный план

$$\vec{x}_o = (\underbrace{0;0;\ldots;0}_{n};b_1; b_2; \ldots; b_m)$$

Решение исходной задачи симплексным методом путем введения искусственных переменных ω_i называется симплексным методом с искусственным базисом.

Если в результате применения симплексного метода к расширенной задаче получен оптимальный план, в котором все искусственные переменные $\omega_i^* = 0$, то его первые n компоненты дают оптимальный план исходной задачи.

Теорема. Если в оптимальном плане М-задачи хотя бы одна из искусственных переменных отлична от нуля, то исходная задача не имеет допустимых планов, т. е. ее условия несовместны.

Теорема. Пусть исходная задача решается на максимум. Если для некоторого опорного плана все оценки $\Delta_i \ \ (j = \overline{1,n})$ неотрицательны, то такой план оптимален.

Теорема. Если исходная задача решается на минимум и для некоторого опорного плана все оценки $\Delta_i \ \ (j = \overline{1,n})$ неположительны, то такой план оптимален.

Теория двойственности

Понятие двойственности рассмотрим на примере задачи оптимального использования сырья. Пусть на предприятии решили рационально использовать отходы основного производства. В плановом периоде появились отходы сырья m видов в объемах b_i единиц $(i = 1,\ldots,m)$. Из этих отходов, учитывая специализацию предприятия, можно наладить выпуск n видов неосновной продукции. Обозначим через a_{ij} норму расхода сырья i-го вида на единицу j-й $(j = 1,\ldots,n)$ продукции, c_j - цена реализации единицы j-й продукции (реализация обеспечена). Неизвестные величины задачи: x_j — объемы выпуска *j*-й продукции, обеспечивающие предприятию максимум выручки.

Математическая модель задачи:

$$\max Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \qquad (2.23)$$

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + a_{1n}x_n \le b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{2n}x_n \le b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + a_{mn}x_n \le b_m \end{array}\right\} \qquad (2.24)$$

$$x_j \ge 0 \ \ (j = \overline{1,n}) \qquad (2.25)$$

Предположим далее, что с самого начала при изучении вопроса об использовании отходов основного производства на предприятии появилась возможность реализации их некоторой организации. Необходимо установить прикидочные оценки (цены) на эти отходы.

Обозначим их $y_1, y_2, \ldots, y_m$.

Оценки должны быть установлены исходя из следующих требований, отражающих несовпадающие интересы предприятия и организации:

1) общую стоимость отходов сырья покупающая организация стремится минимизировать;

2) предприятие согласно уступить отходы только по таким ценам, при которых оно получит за них выручку, не меньшую той, что могло бы получить, организовав собственное производство.

Эти требования формализуются в виде следующей ЗЛП.

Требование 1 покупающей организации – минимизация покупки:

$$\min\ f = b_1y_1 + b_2y_2 + \ldots + b_my_m \tag{2.26}$$

Требование 2 предприятия, реализующего отходы сырья, можно сформулировать в виде системы ограничений. Предприятие откажется от выпуска каждой единицы продукции первого вида, если $a_{11}y_1 + a_{12}y_2 + \ldots + a_{m1}y_m \geq c_1$, где левая часть означает выручку за сырьё идущее на единицу продукции первого вида; правая – её цену.

Аналогичные рассуждения логично провести в отношении выпуска продукции каждого вида. Поэтому требование предприятия, реализующего отходы сырья, можно формализовать в виде сл. системы ограничений:

$$\left.\begin{array}{l} a_{11}y_1 + a_{12}y_2 + \ldots + a_{m1}y_m \geq c_1 \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_m \geq c_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \geq c_n \end{array}\right\} \tag{2.27}$$

По смыслу задачи оценки не должны быть отрицательными:

$$y_1 \geq 0, y_2 \geq 0, \ldots, y_m \geq 0. \tag{2.28}$$

Переменные $y_i\ \ (i = 1, \ldots, m)$ называют двойственными оценками или объективно обусловленными оценками.

Задачи (2.23) - (2.25) и (2.26) - (2.28) называют парой взаимно двойственных ЗЛП.

Между прямой и двойственной задачами можно установить следующую взаимосвязь:

1. Если прямая задача на максимум, то двойственная к ней — на минимум, и наоборот.

2. Коэффициенты c_i целевой функции прямой задачи являются свободными членами ограничений двойственной задачи.

3. Свободные члены b_i ограничений прямой задачи являются коэффициентами целевой функции двойственной.

4. Матрицы ограничений прямой и двойственной задач являются транспонированными друг к другу.

5. Если прямая задача на максимум, то ее система ограничений представляется в виде неравенств типа $\leq$. Двойственная задача решается на минимум, и ее система ограничений имеет вид неравенств типа $\geq$.

6. Число ограничений прямой задачи равно числу переменных двойственной, а число ограничений двойственной — числу переменных прямой.

7. Все переменные в обеих задачах неотрицательны.

Основные теоремы двойственности и их экономическое содержание

Теорема. Для любых допустимых планов $\vec{x} = (x_1, \ldots, x_n)$ и $\vec{y} = (y_1, \ldots, y_m)$ прямой и двойственной ЗЛП справедливо неравенство $z(\vec{x}) \leq f(\vec{y})$, т.е.

$$\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i \tag{2.29}$$

– основное неравенство теории двойственности.

Теорема (критерий оптимальности Канторовича).

Если для некоторых допустимых планов $\vec{x}^*$ и $\vec{y}^*$ пары двойственных задач выполняется неравенство $z(\vec{x}^*) = f(\vec{y}^*)$, то $\vec{x}^*$ и $\vec{y}^*$ являются оптимальными планами соответствующих задач.

Теорема (малая теорема двойственности).

Для существования оптимального плана любой из пары двойственных задач необходимо и достаточно существование допустимого плана для каждой из них.

Теорема. Если одна из двойственных задач имеет оптимальное решение, то и другая имеет оптимальное решение, причем экстремальные значения целевых функций равны: $z(\vec{x}^*) = f(\vec{y}^*)$. Если одна из двойственных задач неразрешима вследствие неограниченности целевой функции на множестве допустимых решений, то система ограничений другой задачи противоречива.

Экономическое содержание первой теоремы двойственности состоит в следующем: если задача определения оптимального плана, максимизирующего выпуск продукции, разрешима, то разрешима и задача определения оценок ресурсов. Причем цена продукции, полученной при реализации оптимального плана, совпадает с суммарной оценкой ресурсов. Совпадение значений целевых функций для соответствующих планов пары двойственных задач достаточно для того, чтобы эти планы были оптимальными. Это значит, что план производства и вектор оценок ресурсов являются оптимальными тогда и только тогда, когда цена произведенной продукции и суммарная оценка ресурсов совпадают. Оценки выступают как инструмент балансирования затрат и результатов. Двойственные оценки, обладают тем свойством, что они гарантируют рентабельность оптимального плана, т. е. равенство общей оценки продукции и ресурсов, и обусловливают убыточность всякого другого плана, отличного от оптимального. Двойственные оценки позволяют сопоставить и сбалансировать затраты и результаты системы.

Теорема (о дополняющей нежесткости)

Для того, чтобы планы $\vec{x}^*$ и $\vec{y}^*$ пары двойственных задач были оптимальны, необходимо и достаточно выполнение условий:

$$x_j^*\left(\sum_{i=1}^{m} a_{ij} y_i^* - c_j\right) = 0; \qquad (j = \overline{1,n}), \tag{2.30}$$

$$y_i^*\left(\sum_{j=1}^{n} a_{ij} x_j^* - b_i\right) = 0; \qquad (i = \overline{1,m}), \tag{2.31}$$

Условия (2.30), (2.31) называются условиями дополняющей нежесткости. Из них следует: если какое-либо ограничение одной из задач ее оптимальным планом обращается в строгое неравенство, то соответствующая компонента оптимального плана двойственной задачи должна равняться нулю; если же какая-либо компонента оптимального плана одной из задач положительна, то соответствующее ограничение в двойственной задаче ее оптимальным планом должно обращаться в строгое равенство.

Экономически это означает, что если по некоторому оптимальному плану $\vec{x}^*$ производства расход *i*-го ресурса строго меньше его запаса b_i, то в оптимальном плане соответствующая двойственная оценка единицы этого ресурса равна нулю. Если же в некотором оптимальном плане оценок его *i*-я компонента строго больше нуля, то в оптимальном плане производства расход соответствующего ресурса равен его запасу. Отсюда следует вывод: двойственные оценки могут служить мерой дефицитности ресурсов. Дефицитный ресурс (полностью используемый по оптимальному плану производства) имеет положительную оценку, а ресурс избыточный (используемый не полностью) имеет нулевую оценку.

Теорема (об оценках). Двойственные оценки показывают приращение функции цели,

вызванное малым изменением свободного члена соответствующего ограничения задачи математического программирования, точнее

$$\frac{\partial z(\vec{x}^*)}{\partial b_i} = y_i^* \qquad (i = \overline{1,m}) \tag{2.32}$$

Основные виды задач, сводящихся к ЗЛП

Задача о наилучшем использовании ресурсов. Пусть некоторая производственная единица (цех, завод, объединение и т. д.), исходя из конъюнктуры рынка, технических или технологических возможностей и имеющихся ресурсов, может выпускать n различных видов продукции (товаров), известных под номерами, обозначаемыми индексом *j* $(j = \overline{1,n})$. Будем обозначать эту продукцию Π_j. Предприятие при производстве этих видов продукции должно ограничиваться имеющимися видами ресурсов, технологий, других производственных факторов (сырья, полуфабрикатов, рабочей силы, оборудования, электроэнергии и т. д.). Все эти виды ограничивающих факторов называют ингредиентами R_i. Пусть их число равно *m*; припишем им индекс *i* $(i = \overline{1,m})$. Они ограничены, и их количества равны соответственно $b_1, b_2, \ldots, b_m$ условных единиц. Таким образом, $\vec{b} = (b_1; \ldots; b_i; \ldots; b_m)$ - вектор ресурсов. Известна экономическая выгода (мера полезности) производства продукции каждого вида, исчисляемая, скажем, по отпускной цене товара, его прибыльности, издержкам производства, степени удовлетворения потребностей и т. д. Примем в качестве такой меры, например, цену реализации c_j $(j = \overline{1,n})$, т.е. $\vec{c} = (c_1; c_2; \ldots; c_j; \ldots; c_n)$ — вектор цен. Известны также технологические коэффициенты a_{ij}, которые указывают, сколько единиц *i*–го ресурса требуется для производства единицы продукции >*j*-го вида. Матрицу коэффициентов a_{ij} называют технологической и обозначают буквой *A*. Имеем $A = [a_{ij}]$. Обозначим через $\vec{x} = (x_1; \ldots; x_j; \ldots; x_n)$ план производства, показывающий, какие виды товаров $\Pi_1, \ldots, \Pi_j, \ldots, \Pi_n$ нужно производить и в каких количествах, чтобы обеспечить предприятию максимум объема реализации при имеющихся ресурсах.

Так как c_i - цена реализации единицы *j*-й продукции, цена реализованных x_j единиц будет равна $x_j c_j$, а общий объем реализации $Z = c_1 x_1 + \ldots + c_n x_n$.

Это выражение — целевая функция, которую нужно максимизировать.

Так как $a_{ij} x_j$ - расход *i*-го ресурса на производство x_j единиц *j*-й продукции, то, просуммировав расход *i*-го ресурса на выпуск всех *n* видов продукции, получим общий расход этого ресурса, который не должен превосходить b_i $(i = \overline{1,m})$ единиц: $a_{i1} x_1 + \ldots + a_{ij} x_j + \ldots a_{in} x_n \le b_i$

Чтобы искомый план $\vec{x} = (x_1; x_2; \ldots x_j; \ldots x_n)$ был реализован, наряду с ограничениями на ресурсы нужно наложить условие неотрицательности на объёмы x_j выпуска продукции: $x_j \ge 0$ $(j = \overline{1,n})$.

Таким образом, модель задачи о наилучшем использовании ресурсов примет вид:

$$\max Z = \sum_{j=1}^{n} c_j x_j \tag{2.33}$$

при ограничениях:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \qquad (i = \overline{1,m}) \tag{2.34}$$

$$x_j \ge 0 \qquad (j = \overline{1,n}) \tag{2.35}$$

Так как переменные x_j входят в функцию $z(\vec{x})$ и систему ограничений только в первой степени, а показатели a_{ij}, b_i, c_j являются постоянными в планируемый период, то (2.33)-(2.35) – задача линейного программирования.

Задача о смесях

В различных отраслях народного хозяйства возникает проблема составления таких рабочих смесей на основе исходных материалов, которые обеспечивали бы получение конечного продукта, обладающего определенными свойствами. К этой группе задач относятся задачи о выборе диеты, составлении кормового рациона в животноводстве, шихт в металлургии, горючих и смазочных смесей в нефтеперерабатывающей промышленности, смесей для получения бетона в строительстве и т. д. Высокий уровень затрат на исходные сырьевые материалы и необходимость повышения эффективности производства выдвигает на первый план следующую задачу: получить продукцию с заданными свойствами при наименьших затратах на исходные сырьевые материалы.

Пример. Для откорма животных используется три вида комбикорма: А, В и С. Каждому животному в сутки требуется не менее 800 г. жиров, 700 г. белков и 900 г. углеводов. Содержание в 1 кг. каждого вида комбикорма жиров белков и углеводов (граммы) приведено в таблице:

Содержание в 1 кг.	Комбикорм		
	A	B	C
Жиры	320	240	300
Белки	170	130	110
Углеводы	380	440	450
Стоимость 1 кг	31	23	20

Сколько килограммов каждого вида комбикорма нужно каждому животному, чтобы полученная смесь имела минимальную стоимость?

Математическая модель задачи есть:

x_1, x_2, x_3 - количество комбикорма А,В и С. Стоимость смеси есть:

$31x_1 + 23x_2 + 20x_3 \to \min$

Ограничения на количество ингредиентов:

$$\begin{cases} 320x_1 + 240x_2 + 300x_3 \ge 800; \\ 170x_1 + 130x_2 + 110x_3 \ge 700; \\ 380x_1 + 440x_2 + 450x_3 \ge 900, \end{cases}$$

$$x_{1,2,3} \ge 0.$$

Задача о раскрое материалов

Сущность задачи об оптимальном раскрое состоит в разработке таких технологически допустимых планов раскроя, при которых получается необходимый комплект заготовок, а отходы (по длине, площади, объему, массе или стоимости) сводятся к минимуму. Рассматривается простейшая модель раскроя по одному измерению. Более сложные постановки ведут к задачам целочисленного программирования.

Задача о назначениях

Речь идет о задаче распределения заказа (загрузки взаимозаменяемых групп оборудования) между предприятиями (цехами, станками, исполнителями) с различными произ-

водственными и технологическими характеристиками, но взаимозаменяемыми в смысле выполнения заказа. Требуется составить план размещения заказа (загрузки оборудования), при котором с имеющимися производственными возможностями заказ был бы выполнен, а показатель эффективности достигал экстремального значения.

Пример. Цеху металлообработки нужно выполнить срочный заказ на производство деталей. Каждая деталь обрабатывается на 4-х станках С1, С2, С3 и С4. На каждом станке может работать любой из четырех рабочих Р1, Р2, Р3, Р4, однако, каждый из них имеет на каждом станке различный процент брака. Из документации ОТК имеются данные о проценте брака каждого рабочего на каждом станке:

Рабочие	Станки			
	С1	С2	С3	С4
Р1	2,3	1,9	2,2	2,7
Р2	1,8	2,2	2,0	1,8
Р3	2,5	2,0	2,2	3,0
Р4	2,0	2,4	2,4	2,8

Необходимо так распределить рабочих по станкам, чтобы суммарный процент брака (который равен сумме процентов брака всех 4-х рабочих) был минимален. Чему равен этот процент?

Обозначим за $x_{ij};\ i=1,2,3,4;\ j=1,2,3,4$ - переменные, которые принимают значения 1, если *i*-й рабочий работает на *j*-м станке. Если данное условие не выполняется, то $x_{ij}=0$. Целевая функция есть:

$$2{,}3x_{11}+1{,}9x_{12}+2{,}2x_{13}+2{,}7x_{14}+1{,}8x_{21}+2{,}2x_{22}+2x_{23}+1{,}8x_{24}+$$
$$+2{,}5x_{31}+2x_{32}+2{,}2x_{33}+3x_{34}+2x_{41}+2{,}4x_{42}+2{,}4x_{43}+2{,}8x_{44}\to\min.$$

Вводим ограничения. Каждый рабочий может работать только на одном станке, то есть

$$\begin{cases} x_{11}+x_{12}+x_{13}+x_{14}=1;\\ x_{21}+x_{22}+x_{23}+x_{24}=1;\\ x_{31}+x_{32}+x_{33}+x_{34}=1;\\ x_{41}+x_{42}+x_{43}+x_{44}=1. \end{cases}$$

Кроме того, каждый станок обслуживает только один рабочий:

$$\begin{cases} x_{11}+x_{21}+x_{31}+x_{41}=1;\\ x_{12}+x_{22}+x_{32}+x_{42}=1;\\ x_{13}+x_{23}+x_{33}+x_{43}=1;\\ x_{14}+x_{24}+x_{34}+x_{44}=1. \end{cases}$$

Кроме того, все переменные должны быть целыми и неотрицательными: $x_{ij}\ge 0,\ x_{ij}-целые$.

Транспортная задача, методы ее решения.

Математическая модель задачи

Линейные транспортные задачи составляют особый класс задач линейного программирования. Задача заключается в отыскании такого плана перевозок продукции с *m* складов в пункт назначения *n* который, потребовал бы минимальных затрат. Если потребитель *j* получает единицу продукции (по прямой дороге) со склада *i*, то возникают издержки C_{ij}. Предполагается, что транспортные расходы пропорциональны перевозимому количеству продукции, т.е. перевозка *k* единиц продукции вызывает расходы kC_{ij}.

Далее, предполагается, что $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_i,$

где a_i есть количество продукции, находящееся на складе i, и b_j – потребность потребителя $>j$. Такая транспортная задача называется закрытой. Однако, если данное равенство не выполняется, то получаем открытую транспортную задачу, которая сводится к закрытой по следующим правилам:

1. Если сумма запасов в пунктах отправления превышает сумму поданных заявок $\sum_{i=1}^{m} a_i \rangle \sum_{j=1}^{n} b_i,$ то количество продукции, равное $\sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_i,$ остается на складах. В этом случае мы введем "фиктивного" потребителя n+1 с потребностью $\sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_i$ и положим транспортные расходы $p_{i,n+1}$ равными 0 для всех i.

2. Если сумма поданных заявок превышает наличные запасы $\sum_{i=1}^{m} a_i \langle \sum_{j=1}^{n} b_i,$ то потребность не может быть покрыта. Эту задачу можно свести к обычной транспортной задаче с правильным балансом, если ввести фиктивный пункт отправления m+1 с запасом $\sum_{i=1}^{m} b_i - \sum_{j=1}^{n} a_i$ и стоимость перевозок из фиктивного пункта отправления во все пункты назначения принять равным нулю.

Математическая модель транспортной задачи имеет вид:

$$\sum_{j=1}^{n} x_{ij} = a_i \ (i = 1,\ldots,m);$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}x_{ij} \to \min$$

$$\sum_{i=1}^{m} x_{ij} = b_i \ \ (j = 1,\ldots,n);$$

$$x_{ij} \geq 0 \quad (i = 1,\ldots,m;\ j = 1,\ldots,n)$$

где x_{ij} количество продукции, поставляемое со склада i потребителю j, а C_{ij} издержки (стоимость перевозок со склада i потребителю j).

ПРИМЕР. Компания «Стройгранит» производит добычу строительной щебенки и имеет на территории региона три карьера. Запасы щебенки на карьерах соответственно равны 800, 900 и 600 тыс. тонн. Четыре строительные организации , проводящие строительные работы на разных объектах этого же региона дали заказ на поставку соответственно 300, 600, 650 и 750 тыс. тонн щебенки. Стоимость перевозки 1 тыс. тонн щебенки с каждого карьера на каждый объект приведены в таблице:

Карьер	*Строительный объект*			
	1	*2*	*3*	*4*
1	8	4	1	7
2	3	6	7	3
3	6	5	11	8

Необходимо составить такой план перевозки (количество щебенки, перевозимой с каждого карьера на каждый строительный объект), чтобы суммарные затраты на перевозку были минимальными.

Данная транспортная задача является закрытой, так как запасы поставщиков 800+900+600=2300 равны спросу потребителей 300+600+650+750=2300. Математическая модель ЗЛП в данном случае имеет вид:

$x_{ij};\ i=1,2,3;\ j=1,2,3,4$ - количество щебенки, перевозимой с *i*–го карьера на *j*–й объект. Тогда целевая функция равна

$$8x_{11}+4x_{12}+x_{13}+7x_{14}+3x_{21}+6x_{22}+7x_{23}+3x_{24}+$$
$$+6x_{31}+5x_{32}+11x_{33}+8x_{34}\to\min.$$

Ограничения имеют вид

$$\begin{cases} x_{11}+x_{12}+x_{13}+x_{14}=800; \\ x_{21}+x_{22}+x_{23}+x_{24}=900; \\ x_{31}+x_{32}+x_{33}+x_{34}=600; \\ x_{11}+x_{21}+x_{31}=300; \\ x_{12}+x_{22}+x_{32}=600; \\ x_{13}+x_{23}+x_{33}=650; \\ x_{14}+x_{24}+x_{34}=750; \end{cases}$$

$x_{ij}\geq 0.$

Составление опорного плана

Решение транспортной задачи начинается с нахождения опорного плана. Для этого существуют различные способы. Например, способ северо-западного угла, способ минимальной стоимости по строке, способ минимальной стоимости по столбцу и способ минимальной стоимости таблицы.

Рассмотрим простейший, так называемый ***способ северо-западного угла***. Пояснить его проще всего будет на конкретном примере:

Условия транспортной задачи заданы транспортной таблицей.

Таблица № 2.1

	B_1	B_2	B_3	B_4	B_5	Запасы $a_{>i}$
A_1	10	8	5	6	9	48
A_2	6	7	8	6	5	30
A_3	8	7	10	8	7	27
A_4	7	5	4	6	8	20
Заявки b_j	18	27	42	12	26	125

Будем заполнять таблицу перевозками постепенно начиная с левой верхней ячейки ("северо-западного угла" таблицы). Будем рассуждать при этом следующим образом. Пункт B_1 подал заявку на 18 единиц груза. Удовлетворим эту заявку за счёт запаса 48, имеющегося в пункте A_1, и запишем перевозку 18 в клетке (1,1). После этого заявка пункта B_1 удовлетворена, а в пункте A_1 осталось ещё 30 единиц груза. Удовлетворим за счёт них заявку пункта B_2 (27 единиц), запишем 27 в клетке (1,2); оставшиеся 3 единицы пункта A_1 назначим пункту B_3. В составе заявки пункта B_3 остались неудовлетворённными 39 единиц. Из них 30 покроем за счёт пункта A_2, чем его запас будет исчерпан, и ещё 9 возьмём из пункта A_3. Из оставшихся 18 единиц пункта A_3 12 выделим пункту B_4; оставшиеся 6 единиц назначим пункту B_5, что вместе со всеми 20 единицами пункта A_4 покроет его заявку. На этом распределение запасов закончено; каждый пункт назначения получил груз, согласно своей заявки. Это выражается в том, что сумма перевозок в каждой строке равна соответствующему запасу, а в столбце - заявке. Таким образом, нами сразу же составлен план перевозок,

удовлетворяющий балансовым условиям. Полученное решение является ***опорным решением транспортной задачи:***

Таблица № 2.2

	B_1	B_2	B_3	B_4	B_5	Запасы a_i
A_1	10 18	8 27	5 3	6	9	48
A_2	6	7	8 30	6	5	30
A_3	8	7	10 9	8 12	7 6	27
A_4	7	5	4	6	8 20	20
Заявки b_j	18	27	42	12	26	125

Составленный нами план перевозок, не является оптимальным по стоимости, так как при его построении мы совсем не учитывали стоимость перевозок C_{ij} . Другой способ - способ минимальной стоимости по строке - основан на том, что мы распределяем продукцию от пункта A_i не в любой из пунктов B_j, а в тот, к которому стоимость перевозки минимальна. Если в этом пункте заявка полностью удовлетворена, то мы убираем его из расчетов и находим минимальную стоимость перевозки из оставшихся пунктов B_j. Во всем остальном этот метод схож с методом северо-западного угла. В результате, опорный план, составленный способом минимальной стоимости по строке выглядит, так как показано в таблице № 2.3.

При этом методе может получиться, что стоимости перевозок C_{ij} и C_{ik} от пункта A_i к пунктам B_j и B_k равны. В этом случае, с экономической точки зрения, выгоднее распределить продукцию в тот пункт, в котором заявка больше. Так, например, в строке 2: $C_{21} = C_{24}$, но заявка b_1 больше заявки b_4, поэтому 4 единицы продукции мы распределим в клетку (2,1).

Таблица № 2.3

	B_1	B_2	B_3	B_4	B_5	Запасы a_i
A_1	10	8	5 42	6 6	9	48
A_2	6 4	7	8	6	5 26	30
A_3	8	7 27	10	8	7 0	27
A_4	7 14	5	4	6 6	8	20
Заявки b_j	18	27	42	12	26	125

Способ минимальной стоимости по столбцу аналогичен предыдущему способу. Их отличие состоит в том, что во втором способе мы распределяем продукцию от пунктов B_i к пунктам A_j по минимальной стоимости C_{ji}.

Опорный план, составленный способами минимальных стоимостей, обычно более близок к оптимальному решению. Так в нашем примере общие затраты на транспортировку по плану, составленному первым способом $F_0 = 1039$, а по второму $F_0 = 723$.

Клетки таблицы, в которых стоят ненулевые перевозки, являются **базисными**. Их число должно равняться $m + n - 1$. Необходимо отметить также, что встречаются такие ситуации, когда количество базисных клеток меньше чем $m + n - 1$. В этом случае распределительная задача называется вырожденной. И следует в одной из свободных клеток поставить количество перевозок равное нулю. Так, например, в таблице

№ 2.3: $m + n - 1 = 4 + 5 - 1 = 8$, а базисных клеток 7, поэтому нужно в одну из клеток строки 3 или столбца 2 поставить значение “0”. Например в клетку (3,5).

Составляя план по способам минимальных стоимостей в отличии от плана по способу северо-западного угла мы учитываем стоимости перевозок C_{ij}, но все же не можем утверждать, что составленный нами план является оптимальным.

Распределительный метод достижения оптимального плана

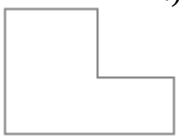
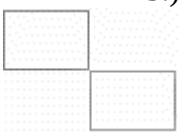
Теперь попробуем улучшить план, составленный способом северо-западного угла. Перенесем, например, 18 единиц из клетки (1,1) в клетку (2,1) и чтобы не нарушить баланса перенесём те же 18 единиц из клетки (2,3) в клетку (1,3). Получим новый план. Подсчитав стоимость опорного плана (она ровняется 1039) и стоимость нового плана (она ровняется 913) нетрудно убедиться, что стоимость нового плана на 126 единиц меньше. Таким образом, за счёт циклической перестановки 18 единиц груза из одних клеток в другие нам удалось понизить стоимость плана:

Таблица № 2.4

	B_1	B_2	B_3	B_4	B_5	Запасы a_i
A_1	10	8 27	5 21	6	9	48
A_2	6 18	7	8 12	6	5	30
A_3	8	7	10 9	8 12	7 6	27
A_4	7	5	4	6	8 20	20
Заявки b_j	18	27	42	12	26	125

На этом способе уменьшения стоимости в дальнейшем и будет основан алгоритм оптимизации плана перевозок. **Циклом** в транспортной задаче мы будем называть несколько занятых клеток, соединённых замкнутой, ломанной линией, которая в каждой клетке совершает поворот на 90°. Существует несколько вариантов цикла:

1.) 2.) 3.)

Нетрудно убедиться, что каждый цикл имеет чётное число вершин и значит, чётное число звеньев (стрелок). Условимся отмечать знаком + те вершины цикла, в которых перевозки необходимо увеличить, а знаком - , те вершины , в которых перевозки необходимо уменьшить. Цикл с отмеченными вершинами будем называть означенным. Перенести какое-то количество единиц груза по означенному циклу, это значит увеличить перевозки, стоящие в положительных вершинах цикла, на это количество единиц, а перевозки, стоящие в отрицательных вершинах уменьшить на то же количество. Очевидно, при переносе любого числа единиц по циклу равновесие между запасами и заявками не меняется: по прежнему сумма перевозок в каждой строке равна запасам этой строки, а сумма перевозок в каждом столбце - заявке этого столбца. Таким образом, при любом циклическом переносе, оставляющем перевозки неотрицательными допустимый план остаётся допустимым. Стоимость же плана при этом может меняться: увеличиваться или уменьшатся. Назовём ценой цикла увеличение стоимости перевозок при перемещении одной единицы груза по означенному циклу. Очевидно, цена цикла ровна алгебраической сумме стоимостей, стоящих в вершинах цикла, причём стоящие в положительных вершинах берутся со знаком +, а в отрицательных со знаком -. Обозначим цену цикла через γ. При перемещении одной единицы груза по циклу стоимость перевозок увеличивается на величину γ. При переме-

щении по нему k единиц груза стоимость перевозок увеличиться на ky. Очевидно, для улучшения плана имеет смысл перемещать перевозки только по тем циклам, цена которых отрицательна. Каждый раз, когда нам удаётся совершить такое перемещение, стоимость плана уменьшается на соответствующую величину ky. Так как перевозки не могут быть отрицательными, мы будем пользоваться только такими циклами, отрицательные вершины которых лежат в базисных клетках таблицы, где стоят положительные перевозки. Если циклов с отрицательной ценой в таблице больше не осталось, это означает, что дальнейшее улучшение плана невозможно, то есть оптимальный план достигнут.

Метод последовательного улучшения плана перевозок и состоит в том, что в таблице отыскиваются циклы с отрицательной ценой, по ним перемещаются перевозки, и план улучшается до тех пор, пока циклов с отрицательной ценой уже не останется. При улучшении плана циклическими переносами, как правило, пользуются приёмом, заимствованным из симплекс-метода: при каждом шаге (цикле) заменяют одну свободную переменную на базисную, то есть заполняют одну свободную клетку и взамен того освобождают одну из базисных клеток. При этом общее число базисных клеток остаётся неизменным и равным $m+n-1$. Этот метод удобен тем, что для него легче находить подходящие циклы.

Можно доказать, что для любой свободной клетке транспортной таблице всегда существует цикл и притом единственный, одна из вершин которого лежит в этой свободной клетке, а все остальные в базисных клетках. Если цена такого цикла, с плюсом в свободной клетке, отрицательна, то план можно улучшить перемещением перевозок по данному циклу. Количество единиц груза k, которое можно переместить, определяется минимальным значением перевозок, стоящих в отрицательных вершинах цикла (если переместить большее число единиц груза, возникнут отрицательные перевозки).

Применённый выше метод отыскания оптимального решения транспортной задачи называется распределённым; он состоит в непосредственном отыскании свободных клеток с отрицательной ценой цикла и в перемещении перевозок по этому циклу.

Распределительный метод решения транспортной задачи, с которым мы познакомились, обладает одним недостатком: нужно отыскивать циклы для всех свободных клеток и находить их цены. От этой трудоёмкой работы нас избавляет специальный метод решения транспортной задачи, который называется методом потенциалов.

Решение транспортной задачи методом потенциалов

Этот метод позволяет автоматически выделять циклы с отрицательной ценой и определять их цены.

Пусть имеется транспортная задача с балансовыми условиями

$$\sum_{j=1}^{n} x_{ij} = a_i \ (i = 1,\dots,m);$$

$$\sum_{i=1}^{m} x_{ij} = b_j \ (j = 1,\dots,n);$$

$$x_{ij} \geq 0 \quad (i = 1,\dots,m;\ j = 1,\dots,n)$$

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

Стоимость перевозки единицы груза из A_i в B_j равна C_{ij}; таблица стоимостей задана. Требуется найти план перевозок x_{ij}, который удовлетворял бы балансовым условиям и при этом стоимость всех перевозок бала минимальна.

Идея метода потенциалов для решения транспортной задачи сводиться к следующему. Представим себе что каждый из пунктов отправления A_i вносит за перевозку

единицы груза (всё равно куда) какую-то сумму α_i ; в свою очередь каждый из пунктов назначения B_j также вносит за перевозку груза (куда угодно) сумму β_j. Эти платежи передаются некоторому третьему лицу (“перевозчику“). Обозначим $a_i + b_j = \check{c}_{ij}$ $(i=1..m; j=1..n)$ и будем называть величину $\check{c}_{ij}$ “псевдостоимостью” перевозки единицы груза из A_i в B_j. Заметим, что платежи α_i и β_j не обязательно должны быть положительными; не исключено, что “перевозчик” сам платит тому или другому пункту какую-то премию за перевозку. Также надо отметить, что суммарная псевдостоимость любого допустимого плана перевозок при заданных платежах (α_i и β_j) одна и та же и от плана к плану не меняется.

До сих пор мы никак не связывали платежи (α_i и β_j) и псевдостоимости $\check{c}_{ij}$ с истинными стоимостями перевозок C_{ij}. Теперь мы установим между ними связь. Предположим, что план x_{ij} невырожденный (число базисных клеток в таблице перевозок ровно m + n -1). Для всех этих клеток $x_{ij}>0$. Определим платежи (α_i и β_j) так, чтобы во всех базисных клетках псевдостоимости были ровны стоимостям: $\check{c}_{ij}=\alpha_i+\beta_j=c_{ij}$, при $x_{ij}>0$.

Что касается свободных клеток (где $x_{ij}=0$), то в них соотношение между псевдостоимостями и стоимостями может быть, какое угодно.

Оказывается соотношение между псевдостоимостями и стоимостями в свободных клетках показывает, является ли план оптимальным или же он может быть улучшен. Существует специальная теорема: Если для всех базисных клеток плана $x_{ij}>0$, $\alpha_i+\beta_j=\check{c}_{ij}=c_{ij}$, а для всех свободных клеток $x_{ij}=0$, $\alpha_i+\beta_j=\check{c}_{ij}\le c_{ij}$, то план является ***оптимальным*** и никакими способами улучшен быть не может. Нетрудно показать, что это теорема справедлива также для вырожденного плана, и некоторые из базисных переменных равны нулю. План обладающий свойством :

$\check{c}_{ij} = c_{ij}$ (для всех базисных клеток) (2.36)

$\check{c}_{ij} \le c_{ij}$ (для всех свободных клеток) (2.37)

называется ***потенциальным*** планом, а соответствующие ему платежи (α_iи β_j) — потенциалами пунктов A_i и B_j ($i=1,...,m; j=1,..., n$). Пользуясь этой терминологией вышеупомянутую теорему можно сформулировать так:

Всякий потенциальный план является оптимальным.

Итак, для решения транспортной задачи нам нужно одно - построить потенциальный план. Оказывается его можно построить методом последовательных приближений, задаваясь сначала какой-то произвольной системой платежей, удовлетворяющей условию (2.36). При этом в каждой базисной клетке получиться сумма платежей, равная стоимости перевозок в данной клетке; затем, улучшая план следует одновременно менять систему платежей. Так, что они приближаются к потенциалам. При улучшении плана нам помогает следующее свойство платежей и псевдостоимостей: какова бы ни была система платежей (α_iи β_j) удовлетворяющая условию (2.36), для каждой свободной клетки цена цикла пересчёта равна разности между стоимостью и псевдостоимостью в данной клетке: $\gamma_{i,j}=c_{i,j} - \check{c}_{i,j}$.

Таким образом, при пользовании методом потенциалов для решения транспортной задачи отпадает наиболее трудоёмкий элемент распределительного метода: поиски циклов с отрицательной ценой.

Процедура построения потенциального (оптимального) плана состоит в следующем.

В качестве первого приближения к оптимальному плану берётся любой допустимый план (например, построенный способом минимальной стоимости по строке). В этом плане $m+n-1$ базисных клеток, где m - число строк, n - число столбцов транспортной таблицы. Для этого плана можно определить платежи (α_i и β_j), так, чтобы в каждой базисной клетке выполнялось условие: $\alpha_i+\beta_j=c_{ij}$ (2.38)

Уравнений (2.38) всего $m+n-1$, а число неизвестных равно $m+n$. Следовательно, одну из этих неизвестных можно задать произвольно (например, равной нулю). После

этого из $m+n-1$ уравнений (2.38) можно найти остальные платежи α_i, β_j, а по ним вычислить псевдостоимости, $\check{c}_{i,j}=\alpha_i+\beta_j$ для каждой свободной клетки.

Таблица № 2.5

	B_1	B_2	B_3	B_4	B_5	α_i
A_1	10 č= 7	8 č= 6	5 42	6 6	9 č= 6	α_1= 0
A_2	6 4	7 č= 5	8 č= 4	6 č= 5	5 26	α_2= -1
A_3	8 č= 8	7 27	10 č= 6	8 č= 7	7 0	α_3= 1
A_4	7 14	5 č= 6	4 č= 5	6 6	8 č= 6	α_4= 0
β_j	β_1= 7	β_2= 6	β_3= 5	β_4= 6	β_5= 6	

$\alpha_4 = 0$, =>

β_4=6, так как $\alpha_4+\beta_4=C_{44}$=6, =>

α_1=0, так как $\alpha_1+\beta_4=C_{14}$=6, =>

β_3=5, так как $\alpha_1+\beta_3=C_{13}$=5, =>

β_1= 7, так как $\alpha_4+\beta_1=C_{41}$=7, =>

α_2=-1, так как $\alpha_2+\beta_1= C_{21}$=6, =>

β_5=6, так как $\alpha_2+\beta_5=C_{25}$=5, =>

α_3= 1, так как $\alpha_3+\beta_5=C_{35}$=7, =>

β_2=6, так как $\alpha_3+\beta_2=C_{25}$=7.

Если оказалось, что все эти псевдостоимости не превосходят стоимостей $\check{c}_{ij} \le c_{ij}$, то план потенциален и, значит, оптимален. Если же хотя бы в одной свободной клетке псевдостоимость больше стоимости (как в нашем примере), то план не является оптимальным и может быть улучшен переносом перевозок по циклу, соответствующему данной свободной клетке. Цена этого цикла ровна разности между стоимостью и псевдостоимостью в этой свободной клетке. В таблице № 2.5 мы получили в двух клетках $\check{c}_{ij} \ge c_{ij}$, теперь можно построить цикл в любой из этих двух клеток. Выгоднее всего строить цикл в той клетке, в которой разность $\check{c}_{ij}-c_{ij}$ максимальна. В нашем случае в обоих клетках разность одинакова (равна 1), поэтому, для построения цикла выберем, например, клетку (4,2):

Таблица № 2.6

	B_1	B_2	B_3	B_4	B_5	α_i
A_1	10	8	5 42	6 6	9	0
A_2	6 + 4	7	8	6	5 - 26	-1
A_3	8	7 27	10	8	7 + 0	1
A_4	7 - 14	5 +	4	6 6	8	0
β_i	7	6	5	6	6	

Теперь будем перемещать по циклу число 14, так как оно является минимальным из чисел, стоящих в клетках, помеченных знаком - . При перемещении мы будем вычитать 14 из клеток со знаком - и прибавлять к клеткам со знаком + .

После этого необходимо подсчитать потенциалы α_iи β_j и цикл расчетов повторяется.

Итак, мы приходим к следующему ***алгоритму решения транспортной задачи методом потенциалов:***

1. Взять любой опорный план перевозок, в котором отмечены $m+n-1$ базисных клеток (остальные клетки свободные).

2. Определить для этого плана платежи (α_i и β_j) исходя из условия, чтобы в любой базисной клетке псевдостоимости были равны стоимостям. Один из платежей можно назначить произвольно, например, положить равным нулю.

3. Подсчитать псевдостоимости $\check{c}_{i,j}=\alpha_i+\beta_j$ для всех свободных клеток. Если окажется, что все они не превышают стоимостей, то план оптимален.

4. Если хотя бы в одной свободной клетке псевдостоимость превышает стоимость, следует приступить к улучшению плана путём переброски перевозок по циклу, соответствующему любой свободной клетке с отрицательной ценой (для которой псевдостоимость больше стоимости).

5. После этого заново подсчитываются платежи и псевдостоимости, и, если план ещё не оптимален, процедура улучшения продолжается до тех пор, пока не будет найден оптимальный план.

Так в нашем примере после 2 циклов расчетов получим оптимальный план. При этом стоимость всей перевозки изменялась следующим образом: $F_0=723$, $F_1=709$, $F_2=F_{min}=703$.

Следует отметить так же, что оптимальный план может иметь и другой вид, но его стоимость останется такой же $F_{min} = 703$.

Решение задач оптимизации в Excel

Оптимизационные модели применяются в экономической и технической сфере. Их цель – подобрать сбалансированное решение, оптимальное в конкретных условиях (количество продаж для получения определенной выручки, лучшее меню, число рейсов и т.п.).

В Excel для решения задач оптимизации используются следующие команды:

Подбор параметров («Данные» - «Работа с данными» - «Анализ «что-если»» - «Подбор параметра») – находит значения, которые обеспечат нужный результат.

Поиск решения (надстройка Microsoft Excel; «Данные» - «Анализ») – рассчитывает оптимальную величину, учитывая переменные и ограничения. Перейдите по ссылке и узнайте как подключить настройку «**Поиск решения**».

Диспетчер сценариев («Данные» - «Работа с данными» - «Анализ «что-если»» - «Диспетчер сценариев») – анализирует несколько вариантов исходных значений, создает и оценивает наборы сценариев.

1.7 Лекция 9 (2 ч.)

Тема: Основные понятия теории графов. Классификация графов, их свойства. Деревья, сети. Основы сетевого анализа

1.7.1. Вопросы лекции:

1. Основные понятия и определения теории графов.
2. Модели и алгоритмы сетевого анализа

1.7.2 Краткое содержание вопросов:

Основные понятия и определения теории графов.

Прежде чем приступить к точным определениям, поясним их наглядно. Любое конечное множество точек (*вершин*), некоторые из которых соединены попарно стрелками (в теории

графов эти стрелки называются *дугами*), можно рассматривать как ***граф*** (рис. 1). Например, граф может изображать сеть улиц в городе, вершины графа — перекрестки, стрелками обозначены улицы с разрешенным направлением движения. (Улицы могут быть с односторонним и двусторонним движением.) Отметим здесь два момента. Во-первых, не любой граф можно изобразить на плоскости так, чтобы дуги пересекались только в вершинах (рис. 2). Во-вторых, две вершины могут быть соединены несколькими дугами, идущими в одном направлении (например, таким свойством может обладать схема железных дорог). Такие дуги будем называть *кратными*, а граф, содержащий кратные дуги, часто называют***мультиграфом*** (рис. 3). Если подчеркивается, что в данном графе нет кратных дуг, то он называется ***графом без кратных дуг***.

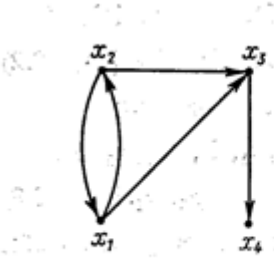


Рис. 2

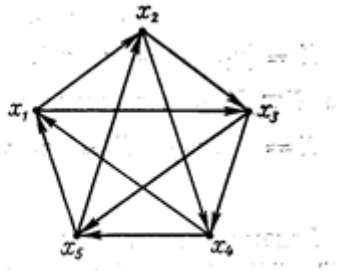


Рис. 3

Определение 1:Мультиграфом или просто ***графом*** G называется пара объектов: $G=\{X,\Gamma\}$, где X - конечное множество вершин, а Γ - конечное подмножество прямого произведения $X\times X\times z_+$. При этом X называется *множеством вершин*, а Γ - *множеством дуг* графа G. Число z_+ указывает на количество дуг (кратность).

Рис. 4

Покажем, каким образом введенное определение формализует понятие графа, интуитивно описанное выше.

Дугу (стрелку), соединяющую вершины x_i, x_j, мы обозначим через (x_i,x_j,n), где x_i - начало дуги, x_j - ее конец, а n - номер дуги. Например, граф, изображенный на рис. *1*, может быть аналитически описан соотношениями:

$$X=\{x_1,x_2,x_3,x_4\},\ \Gamma=\{(x_1,x_2,1),(x_2,x_1,1),(x_2,x_3,1),(x_1,x_3,1),(x_3,x_4,1)\}\ .\ (*)$$

Отметим, что при нумерации дуг не требуется, чтобы были использованы все номера, начиная с единицы и до числа дуг между данными вершинами. Например, граф на рис. 1 может быть описан так:

$$X=\{x_1,x_2,x_3,x_4\},\ \Gamma=\{(x_1,x_2,101),(x_2,x_1,1),(x_2,x_3,4),(x_3,x_4,5),(x_1,x_3,10^5)\}\ .\ (**)$$

Графы, заданные соотношениями (*) и (**), будем считать различными. Для формализации «идентичности» этих графов вводится ниже понятие изоморфизма.

Как уже сказано в определении *1*, элементы множества Γ называются ***дугами***.

Определение 2: Дуга вида (x_i,x_i,n) называется ***петлей***.

На рис. 3 присутствует петля $(x_1,x_1,1)$. Несколько сложнее понятие *ребра*. Иногда на графе в какой-либо прикладной задаче нет необходимости задавать направление связи между вершинами x_i и x_j. При изображении в этом случае связь рисуют без стрелки и говорят о ребре, соединяющем x_i и x_j (рис. 4, а). Будем считать такие ребра объединением дуг (x_i,x_j,n) и (x_j,x_i,n) с одинаковыми номерами.

Определение 3:Ребром называется подмножество вида $\{(x_i, x_j, n), (x_j, x_i, n)\} \subset \Gamma$.

Иллюстрацией может служить рис. 4, б. Видно, что если на графе две вершины связываются дугами с противоположными направлениями, то эти две дуги могут быть заменены одним ребром.

Определение 3: Граф $G = \{X, \Gamma\}$ называется ***симметрическим***, если для любых номеров i, j из того, что дуга $(x_i, x_j, n) \in \Gamma$ следует, что дуга $(x_j, x_i, n) \in \Gamma$.

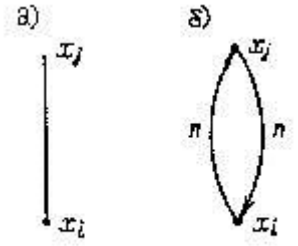


Рис. 5

Из определения видно, что множества вида $\{(x_i, x_j, n), (x_j, x_i, n)\}$ для симметрического графа образуют разбиение множества Γ. Множество классов этого разбиения обозначается $\dot{\Gamma}$ и называется множеством ребер симметрического графа G. Подчеркнем, что $\dot{\Gamma}$ определяется только для симметрических графов.

В дальнейшем, допуская вольность речи в случаях, где это не может привести к разночтениям. Вместо слов «ребро $\{(x_i, x_j, n), (x_j, x_i, n)\}$» будем говорить «ребро (x_i, x_j, n)».

Определение 4: Полностью неориентированным графом называется пара объектов $G = \{X, \overline{\Gamma}\}$, где X - конечное множество вершин, а $\overline{\Gamma}$ - конечное подмножество прямого произведения $2^X \times z_+$, состоящее из элементов вида (A, n), где A содержит ровно два элемента.

У полностью неориентированных графов все дуги – есть ребра (нет направляющих стрелок ни на одной дуге).

Дадим формальное определение графа без кратных дуг.

Определение 5:Графом без кратных дуг называется граф $G = \{X, \Gamma\}$, удовлетворяющий условию: для любых i, j существует самое большее один элемент $n \in z_+$ такой, что $(x_i, x_j, n) \in \Gamma$.

Это означает, что каждая дуга, соединяющая ребра, встречается только один раз. Для описания графов без кратных дуг можно воспользоваться упрощающей символикой - дугу (x_i, x_j, n) графа G обозначать (x_i, x_j). Таким образом, можно дать другое определение графа без кратных дуг, эквивалентное приведенному выше.

Определение 5*: Графом без кратных дуг называется пара объектов $G = \{X, \Gamma\}$, где X - конечное множество, а $\Gamma \subset X \times X$.

Покажем, что граф без кратных дуг определяет некоторое отображение (которое также будем обозначать буквой Γ) $\Gamma : X \to 2^X$. Каждому элементу $x_i \in X$ сопоставим подмножество $\Gamma(x_i) \in 2^X$ следующим образом. Считаем, что $x_j \in \Gamma(x_i)$ тогда и только тогда, когда $(x_i, x_j) \in \Gamma$. Верно и обратное: любое отображение $\Gamma : X \to 2^X$ определяет граф. Можно дать ещё одно эквивалентное определение.

Определение 5:** Графом без кратных дуг называется пара объектов $G = \{X, \Gamma\}$, где X — конечное множество, а $\Gamma : X \to 2^X$ — отображение.

В дальнейшем будут рассматриваться и полностью неориентированные графы без кратных ребер. Ребра такого графа будем часто обозначать (x_i, x_j) или, эквивалентно (x_j, x_i) и говорить о неупорядоченной паре вершин (x_i, x_j).

Определение 6: Дугу $u \in \Gamma$ и вершину $x \in X$ назовем ***инцидентными***, если $x = Пр_1 u$ или $x = Пр_2 u$ (вершина x является либо первой, либо второй проекцией дуги u). Ребро $\bar{u} \in \Gamma$ и вершину $x \in X$ назовем ***инцидентными***, если инцидентны дуга $u \in \bar{u}$ и x. Дугу u, такую, что $Пр_1 u = x$, назовем исходящей из x, а если $Пр_2 u = x$, то входящей в x. Две вершины x_i и x_j называются ***смежными***, если существует дуга, соединяющая либо x_i с x_j, либо x_j с x_i.

Определение 7:Путём в графе $G = \{X, \Gamma\}$ называется такая конечная последовательность дуг $\{u_1, u_2, \ldots, u_m\}$, $u_i \in \Gamma$, что $Пр_2 u_i = Пр_1 u_{i+1}$. Если, кроме того, $Пр_2 u_m = Пр_1 u_1$, то путь $\{u_1, \ldots, u_m\}$ называется ***контуром***. Путь называется ***простым***, если из неравенства $i \neq j$ следует $u_i \neq u_j$, и ***элементарным***, если последовательность вершин $\{Пр_1 u_1, Пр_2 u_1, \ldots, Пр_2 u_m\}$ не содержит совпадающих элементов, кроме, может быть, крайних.

Интерпретация пути (и контура) в графе очевидна. Путь является простым, если при движении вдоль него одна и та же дуга никогда не приходится дважды, и элементарным, если при этом никакая вершина не встречается дважды.

Пусть теперь $G = \{X, \Gamma\}$ - симметрический граф.

Определение 8:Цепью в симметрическом графе G называется такая последовательность $\{\bar{u}_1, \ldots, \bar{u}_m\}$, $\bar{u}_i \in \bar{\Gamma}$ ребер графа G, что существует путь $\{u_1, \ldots, u_m\}$ в G, удовлетворяющий условию $u_i \in \bar{u}_i$ для $i = 1, 2, \ldots, m$. Цепь называется ***циклом***, если соответствующий путь есть контур. Цепь называется ***простой*** (элементарной), если соответствующий путь простой (элементарный).

Очевидно, что для любой цепи $\{\bar{u}_1, \ldots, \bar{u}_m\}$, $m \geq 2$ существует ровно один путь $\{u_1, \ldots, u_2\}$, такой, что $u_i \in \bar{u}_i$. Это доказывает корректность определений простой и элементарной цепи.

Определение 9: Граф $G' = (X', \Gamma')$ называется ***суграфом*** графа $G = (X, \Gamma)$, если $X' = X, \Gamma' \subset \Gamma$. Граф G' называется ***подграфом*** графа G, если $X' \subset X, \Gamma' = \Gamma \cap (X' \times X')$. Наконец, граф G' называется ***частью*** графа G, если $X' \subset X, \Gamma' \subset \Gamma$.

Таким образом, суграф получается из графа удалением некоторого числа дуг (с сохранением вершин). Подграф получается из графа удалением некоторого числа вершин вместе со всеми дугами, инцидентными удаленной вершине. Часть графа получается из графа применением конечного числа обеих описанных операций.

Определение 10: Наименьший симметрический граф $G_{сим}$ такой, что G является суграфом $G_{сим}$, называется ***симметризованным*** графом графа $G = \{X, \Gamma\}$.

Определение 11: Граф $G=(X,\Gamma)$ называется ***сильно связным***, если для любых двух вершин $x_i, x_j \in X$, $x_i \neq x_j$ существует путь $\{u_1, u_2, \ldots, u_n\}$ такой, что $x_i = Пр_1 u_1, x_j = Пр_2 u_n$.

Определение 12: Симметрический граф $G=(X,\Gamma)$ называется **связным**, если он сильно связен. Произвольный граф называется связным, если связен его симметризованный граф $G_{сим}$.

Из этих определений прямо следует, что понятия связности и сильной связности совпадают для симметрических графов.

Определение 13: Подграф G' графа G называется **компонентой связности** графа G, если а) G' связен, б) G' обладает свойством максимальности, т. е. если G'' - некоторый другой связный подграф G и $G' \subset G''$, то графы G' и G'' совпадают.

Определение 14: Будем говорить, что ребро $\bar{u} \in \bar{\Gamma}$ является ***перешейком*** в симметрическом графе $G=(X,\Gamma)$, если при удалении этого ребра количество компонент связности графа увеличивается на единицу.

Очевидно, что ребро является перешейком тогда и только тогда, когда не существует простого цикла, содержащего это ребро.

Определение 15: Полустепенью исхода p вершины x называется число исходящих из нее дуг. **Полустепенью захода** q вершины x называется число входящих в нее дуг.**Степенью вершины** $st\,x$ называется число $p+q$.

Замечание: Иногда в симметрических графах степенью вершины x будет называться число инцидентных ей ребер. Из контекста всегда будет ясно значение числа $st\,x$ в каждом случае.

Определение 16: Ребро $\bar{u}$ симметрического графа $G=(X,\Gamma)$ называется **тупиком**, если степень одной из его вершин равна единице (здесь степень вершины определяется как число инцидентных ей ребер).

Определение 17: Пусть $G=(X,\Gamma)$ — граф, $n=|X|$. Матрицей смежности G называется $n \times n$ матрица $A=\|a_{ij}\|$, в которой элемент a_{ij} равен количеству дуг из Γ вида (x_i, x_j, k).

Определение 18: Графы $G=(X,\Gamma)$ и $G'=(X',\Gamma')$ называются **изоморфными**, если существуют взаимно однозначные отображения $\varphi: X \to X', \psi: \Gamma \to \Gamma'$, такие, что $Пр_1 \circ \psi = \varphi \circ Пр_1, Пр_2 \circ \psi = \varphi \circ Пр_2$.

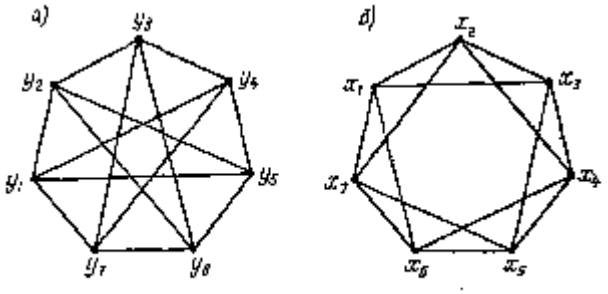


Рис. 5

Пример. Графы G_1 и G_2, изображенные на рис. 5, изоморфны; их изоморфизм определяется подстановкой второго порядка:

$$\varphi: \begin{pmatrix} x_1 x_2 x_3 x_4 x_5 x_6 x_7 \\ y_1 y_4 y_7 y_3 y_6 y_2 y_5 \end{pmatrix}.$$

У изоморфных графов матрицы смежности очень похожи. Их можно получить одну из другой с помощью конечного числа перестановок строк и столбцов.

Географические сети

Важным объектом исследований в географии являются различные *Географические сети,* представляющие собой совокупности линейных фрагментов природного (например, речные, орографические, тектонические) и антропогенного (например, дорожные, электрические, коммуникационные) характера. В общем случае в понятие "географическая сеть" включаются все пространственные (территориальные) связи и отношения, существенные для изучения пространственной организации природных и социально-экономических систем. В этом случае географическая реальность может быть представлена в виде суперпозиции (объединения, наложение) большого количества разнообразных пространственных отношений и связей (транспортных, технологических, экологических, миграционных, информационных и др.) между различными геообъектами (населенными пунктами, предприятиями, административными и экономическими районами, экосистемами и др.). При этом географичность данных отношений состоит в том, что в указанную суперпозицию всегда включается отношение взаимного размещения, которое и придает всему комплексу территориальный, географический характер.

Целью изучения географических сетей является выявление закономерностей их строения, формирования и развития, а также мониторинг, оптимизация и управление (например, в случае транспортных и коммуникационных сетей). ГИС-технология обеспечивает возможность компьютерного представления, моделирования и анализа, сколь угодно больших по числу вершин и ребер сетевых объектов, в сочетании с автоматизированным тематическим картографированием, интерактивным редактированием и визуализацией (включая мультимедиа) соответствующих сетевых моделей.

Модели и алгоритмы сетевого анализа

В моделировании и анализе географических сетей широко применяются методы теории графов. Как известно, любое картографическое изображение территориальных отношений содержит метрические и топологические атрибуты. Графовые модели акцентируют внимание именно на топологические свойства сетей: порядок соединения вершин, наличие циклов, степень связности и др.

Реальные территориальные отношения и связи можно формализовать и изобразить в виде многомерных графов-картосхем. Однако методика анализа таких графов еще недостаточно разработана. Поэтому при изучении географических сетей чаще всего используются относительно простые графовые модели, методика анализа которых разработана до уровня алгоритмов и программ.

Если в качестве свойств графов рассматривать *Помеченность* вершин, а также помеченность и *Направленность* ребер, то можно выделить 8 типов графовых моделей сетей (Рис. 3.4). (Помеченным называется граф, ребрам и/или вершинам которого приписано значение некоторого признака - числового, порядкового, классификационного). Рассмотрим кратко основные типы выделенных графовых моделей:

1. Непомеченные неориентированные графы. С помощью этого типа

Ребра →	Непомеченные		Помеченные	
Вершины ↓	Неориентированные	Ориентированные	Неориентированные	Ориентированные
Непомеченные				
Помеченные				

Рис. 3.4. Типы графовых моделей сетей по свойствам ребер и вершин

Графовых моделей изучаются территориальные связи и отношения, для которых не известны (или не важны) интенсивность и направление: коммуникационные сети и сообщения (в оба конца), производственные связи, маятниковые миграционные потоки и т. п. Анализ таких графов используется при решении следующих задач:

— исследование связности графа: выявление несвязных подграфов (Рис. 3.5а), критических вершин и ребер (т. е. таких, при удалении которых граф перестает быть связным);

— общая характеристика структуры и формы графа с помощью различных показателей;

— оценка вершин графа по их положению в структуре графа;

— нахождение максимальных полных подграфов (клик) и анализ структуры их соединения в исходном графе (Рис. 3.5б);

— поиск кратчайших путей между вершинами графа, решение оптимизационных задач (задача "коммивояжера" и т. п.);

— преобразование (генерализация) исходного графа в более простой, удобный для анализа и картографирования вид.

Моделирование и анализ рассмотренных графов представляет собой быстро развивающийся раздел комбинаторики, имеющий приложения в различных областях науки и техники. Разработаны и разрабатываются специальные алгоритмические языки и пакеты прикладных

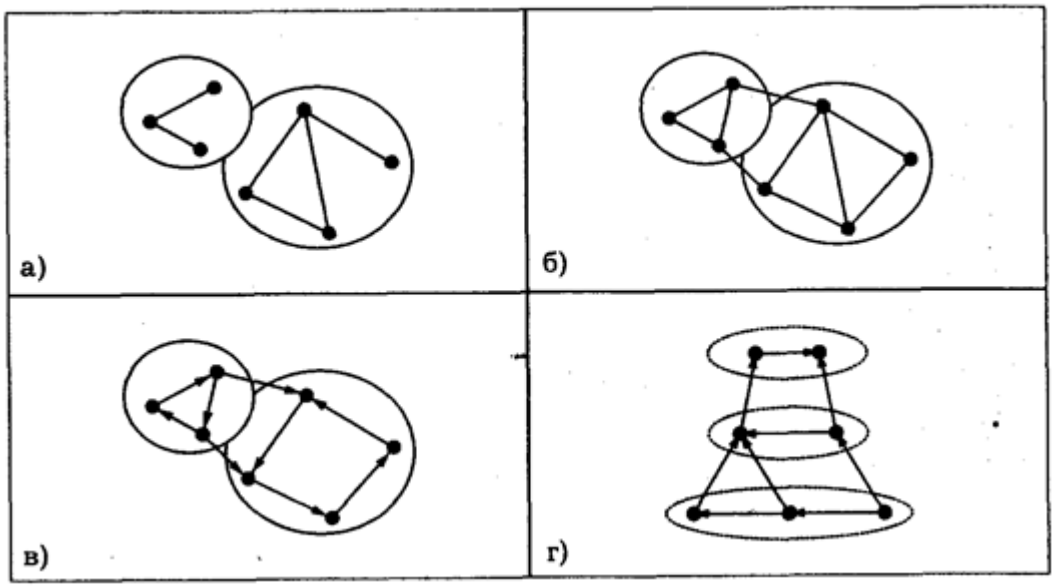


Рис. 3.5. Способы кластеризации графов;

А) изолированные подграфы; б) клики (максимально связные подграфы); В) сильно связные компоненты орграфа; г) слои поточного иерархического графа программ для представления в ЭВМ и анализа графов, в том числе в ГИС.

II. *Помеченные неориентированные графы*. Этот тип графовых моделей целесообразно использовать в случае, когда известна интенсивность территориальных связей между геообъектами, измеренная в числовой или порядковой шкале. Ключевыми понятиями для анализа помеченных (по ребрам) неориентированных графов являются понятия порога и *Устойчивого разбиения*. Устойчивое разбиение, например, можно рассматривать как оптимальное расчленение (районирование) географической сети по данному виду связей.

Основные алгоритмы анализа помеченных неориентированных графов следующие: 1) алгоритм получения пороговых матриц смежности вершин графа при различных значениях порога; 2) алгоритм подсчета несвязных компонент графа, задаваемого соответствующей пороговой матрицей.

III. *Непомеченные ориентированные графы*. Изучая сети, географ иногда располагает информацией лишь о *Направлении* связей между геообъектами, а их интенсивность ему не известна (или не важна в контексте исследования). В этом случае в качестве моделей географических сетей используются непомеченные ориентированные графы - *Орграфы.* Для географических приложений важны такие понятия теории орграфов, как *Сильно связная компонента* и *Поточная иерархическая структура.*

В географическом аспекте сильно связные компоненты можно рассматривать как функциональные районы, а процесс их выделения - как функциональное районирование географической сети (Рис. 3.5в).

Поточная иерархическая структура - это такое представление орграфа, при котором: 1) вершины разбиты на упорядоченные группы (иерархические слои); 2) вершины, находящиеся в одном иерархическом слое, могут иметь связи только между собой и с вершинами более высоких иерархических слоев (Рис. 3.5г). В географическом плане представление и анализ ориентированных графов в виде поточных иерархических структур дает возможность решить следующие задачи: 1) определить общую иерархическую структуру подчинения геообъектов по данному виду связей (например, населенных пунктов по миграционным связям); 2) оценить значение каждого геообъекта в иерархической структуре территориальных связей (например, ландшафта определенного таксономического ранга); 3) осуществить иерархическое районирование географической сети (например, сети поселений различного ранга по степени миграционной привлекательности).

Алгоритм сокращения поточной иерархической структуры орграфа, являющийся основой иерархического районирования, представляет собой циклическую процедуру поиска максимального элемента в строках матрицы интенсивности связей и не предполагает использования специального программного обеспечения.

IV. *Помеченные ориентированные* графы. Графовые модели этого типа используются в случае, когда известны интенсивность связей между геообъектами и их направление. Методика анализа помеченных ориентированных графов содержит в качестве этапов нахождение пороговых матриц для различных значений порога и подсчет для каждой из них числа сильно связных компонент, интерпретируемых в качестве функциональных районов.

1.8 Лекция 10 (2 ч.)

Тема: Оптимизационные модели в сельском хозяйстве

1.8.1. Вопросы лекции:

1. Оптимизация производства сельскохозяйственного предприятия
2. Оптимальный план структуры производства сельскохозяйственного предприятия

1.8.2 Краткое содержание вопросов:

Оптимизация производства сельскохозяйственного предприятия

Постановка экономико-математической задачи оптимизации структуры производства сельскохозяйственного предприятия

В хозяйстве имеется 3500 га пашни, 449 га естественных сенокосов и 657 га естественных пастбищ. Ресурсы труда составляют 1782 тыс.чел.-час. В хозяйстве необходимо произвести не менее 6000 ц молока, 5000 ц прироста молодняка крупного рогатого скота, 200 ц прироста свиней и реализовать не менее 8000 ц озимой пшеницы и 20000 ц овощей. Многолетних трав на семена необходимо иметь в хозяйстве не менее 30 га, зернобобовых - не менее 100 га, а кукурузы на зерно - не более 400 га.
Среднегодовой удой молока на корову 4200 кг, привес на 1 гол молодняка крупного рогатого скота 160 кг, свиней - 130 кг. На содержание 1 коровы требуется 133,6 чел-час труда и 719,5 ден.ед материально-денежных затрат, на содержание 1 гол молодняка крупного рогатого скота - 55,0 и 335,2, 1 гол свиней – 32,9 чел-час и 193,2 соответственно.
Критерий оптимальности задачи - максимальное количество прибыли, получаемое при реализации озимой пшеницы – 192,2 ден.ед., овощей – 83,9, продукции скотоводства – 436,1 и 121,7 и свиноводства -77,2 ден.ед [10].

Таблица 1
Урожайность с.-х. культур и затраты производственных ресурсов.

Культуры	Урожайность, ц/га	Затраты на 1 га	
		труда, чел.-час.	мат.-ден. ср-в, ден.ед
1.Озимые зерновые	37,8	11,4	243,1
2.Яровые зерновые	30,2	9,1	195,9
3.Зернобобовые	18,5	18,5	218,5
4.Кукуруза на зерно	44,5	47,7	495,4
5.Овощи	148,3	387,3	2029,7
6.Кормовые корнеплоды	503,5	278,1	1331,3
7.Многолетние травы на сено	80	27,6	265,2
8.Многолетние травы на зел.корм	274,7	23,9	235,5
9.Многолетние травы на семена	2	8,9	200
10.Однолетние травы на зел.корм	243,5	33,7	198,9
11.Кукуруза на силос и зел.корм	533,4	108,9	572,2
12.Естественные сенокосы	15	12,9	58
13.Естественные пастбища	50	1,5	50,2

Таблица 2 Распределение продукции растениеводства

Культуры	Урожайность продукции, ц/га		Использование, ц		
			реали-зация	на корм скоту	
	основн.	побоч.		ос-новн.	побоч.
Озимые	37,8	37,8	20	12,8	35,0
Яровые	30,2	30,2		25,2	30,2
Зернобобовые	18,5	18,5		16,5	18,5
Кукуруза на зерно	44,5	53,4		44,5	53,4
Овощи	148,3		140	8,3	
Кормовые корнеплоды	503,5	201,4		503,5	201,4
Мн. травы на сено	80			80	
Мн. травы на з/к	274,7			274,7	
Мн. травы на семена	2				
Одн. травы на з/к	243,5			243,5	
Кук. на силос и з/к	533,4			533,4	
Сенокосы	15			15	
Пастбища	50			50	

Нормативно-справочная информация для составления моделей
Таблица 3 Нормативы расхода кормов на одну корову, ц

Удой, кг	Норматив расхода		Удой, кг	Норматив расхода	
	корм.ед	перев. прот.		корм.ед	перев. прот.
2000	31,1	3,11	3300	42,7	4,36
2100	32,1	3,21	3400	43,7	4,43
2200	33,1	3,31	3500	44,1	4,54
2300	34,1	3,41	3600	44,8	4,61

2400	35,1	3,51	3700	45,5	4,69
2500	36,1	3,64	3800	46,2	4,76
2600	37,0	3,74	3900	46,9	4,83
2700	37,9	3,83	4000	47,6	4,95
2800	38,8	3,92	4100	48,2	5,00
2900	39,7	4,01	4200	48,8	5,02
3000	40,6	4,10	4300	49,4	5,14
3100	41,3	4,21	4400	50,0	5,20
3200	42,0	4,28	4500	50,6	5,31

Таблица 4 Структура расхода кормов на одну корову, %

Виды кормов	Удой, ц						
	20-21	22-23	24-25	26-27	28-29	30-32	33-35
Конц. корма	20	21	22	23	24	25	26
Грубые -всего	24	24	24	24	24	23	23
в т.ч. сено	11	11	11	11	11	10	10
солома	6	6	5	5	5	5	4
Сочные - всего	24	23	23	23	23	22	21
в т.ч. силос	19	18	18	18	17	16	15
корм. корн.	5	5	5	5	5	6	6
Зеленые - всего	32	32	31	30	30	30	30
в т.ч. пастбища	5	5	5	5	5	5	5
Всего	100	100	100	100	100	100	100

Таблица 5 Нормативы расхода кормов на одну головы молодняка крупного рогатого скота, ц

Привес на 1 гол., кг	Норматив расхода		Привес на 1 гол., кг	Норматив расхода	
	корм.ед	перев. прот.		корм.ед	перев. прот.
101-110	15,0	1,41	191-200	19,9	2,03
111-120	15,5	1,47	201-210	20,6	2,12
121-130	15,9	1,51	211-220	21,4	2,23
131-140	16,5	1,58	221-230	22,0	2,31
141-150	17,0	1,65	231-240	22,7	2,41
151-160	17,5	1,72	241-250	23,4	2,50
161-170	18,1	1,79	251-260	24,1	2,60
171-180	18,6	1,86	261-270	24,9	2,71
181-190	19,3	1,95	271 и выше	25,6	2,82

Таблица 6 Структура расхода кормов на молодняк крупного рогатого скота, %

Виды кормов	Привес, кг						
	100-120	121-140	141-160	161-180	181-200	201-220	221 и>
Конц. корма	20	21	22	23	24	25	26
Грубые -всего	21	21	20	20	20	20	20
в т.ч. сено	6	5	5	5	5	5	5
солома	13	12	11	11	10	10	9
Сочные - всего	33	32	31	31	30	29	28
в т.ч. силос	32	31	29	29	28	26	25
корм. корн.	1	1	2	2	2	3	3
Зеленые - всего	29	23	23	22	22	21	21
в т.ч. пастбища	1	1	1	1	1	1	1
Молочные	3	3	4	4	4	5	5
Всего	100	100	100	100	100	100	100

Таблица 7 Нормативы расхода кормов на одну голову поголовья свиней, ц

Привес на 1 гол., кг	Норматив расхода		Привес на 1 гол., кг	Норматив расхода	
	корм.ед	перев. прот.		корм.ед	перев. прот.
До 90	7,0	0,66	121-130	9,0	0,95
91-100	7,5	0,75	131-140	9,5	1,03
101-110	8,0	0,82	1441-150	9,7	1,07
111-120	8,5	0,88	151 и выше	10,5	1,20

Таблица 8 Структура расхода кормов на поголовье свиней и птицы, %

Виды кормов	Вид животных	
	свиньи	птица
Конц. корма	83	96
Сочные - всего	10	2
в т.ч. силос	5	
Зеленые - всего	5	2
Молочные	2	
Всего	100	100

Таблица 9 Выход питательных веществ с 1 ц кормовых культур

Культуры	Всего ц к.ед.	в том числе в продукции		Перев. прот., ц
		основной	побочной	
Озимые зерновые	43	36	7	2,9
Яровые зерновые	40	30,3	9,7	2,4
Зернобобовые	42	42		2,8
Кукуруза на зерно	58,5	45,6	12,9	3,35
Кукуруза на силос	42	42	-	3,0
Кукуруза на зеленый корм	30	30	-	2,4

Культуры	Всего ц к.ед.	в том числе в продукции		Перев. прот., ц
		основной	побочной	
Озимые зерновые	43	36	7	2,9
Корнеплоды	38	38	-	2,6
Озимые на зеленый корм	21	21	-	0,9
Однолетн. травы на зел. корм	28	28	-	1,4
Однолетние травы на сено	26	26	-	2,4
Многолет. травы на зел. корм	46	46	-	7
Многолетние травы на сено	30	30	-	6,6
Сенокосы	3,5	3,5	-	1,3
Пастбища	60	60	-	4,5

2.2 Методика подготовки технико-экономических коэффициентов и объектов ограничений матрицы задачи

Система переменных:
x_1 – площадь посева под озимые зерновые, га;
x_2 - площадь посева под яровые зерновые, га;
x_3 - площадь посева под зернобобовые, га;
x_4 - площадь посева под кукурузу на зерно, га;
x_5 - площадь посева под овощи, га;
x_6 - площадь посева под кормовые корнеплоды, га;
x_7 - площадь посева под многолетние травы на сено, га;
x_8 - площадь посева под многолетние травы на зеленый корм, га;
x_9 - площадь посева под многолетние травы на семена, га;
x_{10} - площадь посева под однолетние травы на зеленый корм, га;
x_{11} - площадь посева под кукуруза на силос и зеленый корм, га;
x_{12} - площадь естественных сенокосов, га;
x_{13} - площадь естественных пастбищ, га;
y_1 – количество голов КРС;
y_2 - количество голов молодняка КРС;
y_3 – количество голов свиней;
y_4 – объем производства озимых зерновых на реализацию, ц.;
y_5 – площадь посева озимых зерновых на собственные нужды, га.;
Система ограничений:
1. По площади посевов:
$X_1+X_2+X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10}+X_{11}\leq 3500$
$X_{12}\leq 449$
$X_{13}\leq 657$
2. По планам производства:

$Y_1 \geq 6000/42$
$Y_2 \geq 500/1,6$
$Y_3 \geq 200/1,3$
$Y_4 \geq 8000$
$148,3X_5 \geq 20000$

3. По площади посевов отдельных культур:

$X_9 \geq 30$
$X_3 \geq 100$
$X_4 \leq 400$

4. По затратам труда:

$11,4X_1+9,1X_2+18,5X_3+47,7X_4+387,3X_5+278,1X_6+27,6X_7+23,9X_8+$
$+8,9X_9+33,7X_{10}+108,9X_{11}+12,9X_{12}+1,5X_{13}+133,6Y_1+55,0Y_2+32,9Y_3 \leq 1782000$

5. По кормовым единицам:
5.1 По концентрированным кормам:

$48,8*26/100Y_1 +17,5*22/100Y_2+9*83/100Y3 \leq 37,8*36Y_5+30,2*30,3X_2+44,5*45,6X_4;$

5.2 По грубым кормам:

$48,8*23/100Y_1+17,5*20/100Y_2 \leq 18,5*42X_3+80*30X_7+15*3,5X_{12};$

5.3 По сочным кормам (силос):

$48,8*15/100Y_1+17,5*29/100Y_2+9*5/100Y_3 \leq 533,4*42X_{11};$

5.4 По сочным кормам (корнеплоды):

$48,8*6/100Y_1+17,5*2/100Y_2+9*5/100Y_3 \leq 503,5*38X_6;$

5.5 По зеленым кормам:

$48,8*30/100Y_1+17,5*23/100Y_2+9*5/100Y_3 \leq$
$274,7*46X_8+243,5*28X_{10}+50*60X_{13};$

6. По перевираемому протеину:
6.1 По концентрированным кормам:

$5,02*26/100Y_1+1,72*22/100Y_2+0,95*83/100Y_3 \leq 37,8*2,9Y_5+30,2*2,4X_2+$
$44,5*3,35X_4;$

6.2 По грубым кормам:

$5,02*23/100Y_1+1,72*20/100Y_2 \leq 18,5*2,8X_3+80*2,4X_7+15*1,3X_{12};$

6.3 По сочным кормам (силос):

$5,02*15/100Y_1+1,72*2/100Y_2+0,95*5/100Y_3 \leq 533,4*3X_{11}$;

6.4 По сочным кормам (корнеплоды):

$5,02*6/100Y_1+1,72*2/100Y_2+0,95*5/100Y_3 \leq 503,5*2,6X_6$;

6.5 По зеленым кормам:

$5,02*30/100Y_1+1,72*23/100Y_2+0,95*5/100Y_3 \leq 274,7*7X_8+243,5*1,4X_{10}+ 50*4,5X_{13}$;

7. Дополнительные переменные:
7.1 Значение всех переменных должно быть больше нуля;

$X_1 \geq 0$; $X_2 \geq 0$; $X_3 \geq 0$; $X_4 \geq 0$; $X_5 \geq 0$; $X_6 \geq 0$; $X_7 \geq 0$; $X_8 \geq 0$; $X_9 \geq 0$; $X_{10} \geq 0$; $X_{11} \geq 0$; $X_{12} \geq 0$; $X_{13} \geq 0$; $Y_1 \geq 0$; $Y_2 \geq 0$; $Y_3 \geq 0$; $Y_4 \geq 0$;

7.2 Для более облегчения подсчета мы разбили X_1 на две части – Y_1 и Y_2, следовательно

$X_1=Y_4/37,8+Y_5$.

Целевая функция задачи:

$Z=192,2Y_4+148,3*83,9X_5+436,1*4,2Y_1+121,7*1,6Y_2+77,2*1,3Y_3-243,1X_1-195,9X_2-218,5X_3-495,4X_4-2029,7X_5-1331,3X_6-265,2X_7-235,5X_8-200X_9-198,9X_{10}-572,2X_{11}-58X_{12}-50,2X_{13}-719,5Y_1-335,2Y_2-193,2Y_3$;

Экономико-математическая модель оптимизации структуры производства сельскохозяйственного предприятия

Таблица 10

№	Переменные															Ограничения				
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	Y5		Число	Y1	Y2	Y3	Y4
1.	1	1	1	1	1	1	1	1	1	1	1				≤	3500				
1.												1			≤	449				
1.													1		≤	657				
2.															≥	142,86	1			
2.															≥	312,5		1		
2.															≥	153,85			1	
2.															≥	8000				1
2.					1										≥	134,86				
3.									1						≥	30				
3.			1												≥	100				
3.				1											≤	400				
4.	11,4	9,1	18,5	47,7	387,3	278,1	27,6	23,9	8,9	33,7	108,9	12,9	1,5		≤	1782000	133,6	55	32,9	
5.1		915,06		2029,2										1360,8	≥		12,688	3,85	7,47	
5.2			777				2400					52,5			≥		11,224	3,5		
5.3											22403				≥		7,32	5,075	0,45	

5.4						19133									≥		2,928	0,35	0,45	
5.5								12636		6818			3000		≥		14,64	4,025	0,45	
6.1		72,48		149,08										109,62	≥		1,3052	0,3784	0,7885	
6.2			51,8				192					19,5			≥		1,1546	0,344		
6.3											1600,2				≥		0,753	0,0344	0,0475	
6.4						1309,1									≥		0,3012	0,0344	0,0475	
6.5								1922,9		340,9			225		≥		1,506	0,3956	0,0475	
7.1	1	1	1	1	1	1	1	1	1	1	1	1	1		≥	0	1	1	1	
7.2	1													1	=					0,026
Z	-243,1	-195,9	-218,5	-495,4	10412,67	-1331,3	-265,2	-235,5	-200	-198,9	-572,2	-58	-50,2		=		1112,12	-140,5	-92,84	192,2
	211,64	0	100	37,34	3118,13	0,95	0	0	30	0	1,93	0	27,90	0		38384427	4081	313	154	8000,00

Оптимальный план структуры производства сельскохозяйственного предприятия

Из таблицы 10 видно, что предприятие получит прибыль в 38,4 млн. руб. При этом будет использована вся пашня предприятия, и практически все трудовые ресурсы предприятия. Основная часть прибыли будет получена за счет реализации продукции овощеводства (32,5 млн. руб.), а площадь посева овощей составит 3118,13 га. Большую часть в структуре прибыли составляет реализация молока (4,5 млн. руб.), при этом выручка от реализации молока составляет 7,5 млн. руб. 1,5 млн. руб. предприятие получит за счет реализации озимых зерновых.

Математические модели в сельском хозяйстве дают возможность не только определить взаимосвязи между изучаемыми явлениями, но и установить вид вычислительной техники, количество и точность требуемой для решения информации. Полученные при реализации моделей данные анализируют, в случае необходимости корректируют применительно к конкретным природно-экономическим условиям и используют для целей проектирования и обоснования принятых решений.

Задача *Для одного их предприятий-монополистов по производству кукурузной крупы зависимость объема спроса на продукцию q (единиц в месяц) от ее цены p (тыс.руб.) задается формулой q=40-5p. Определите максимальный уровень p цены (в тыс. руб.), при котором значение выручки предприятия за месяц r=q·p составит 75 тыс.руб.*

Решение: $(40-5p).p=75$; $-5p^2+40p-75=0$; $p^2-8p+15=0$; $p_1=5$, $p_2=3$.

Так как выручка составляет не менее 75 тыс. руб., то максимальный уровень цены 5 тыс.руб.

Ответ. 5 тыс. руб.

Задача. *Ферма занимается возделыванием только двух культур – зерновых и картофеля – и располагает следующими ресурсами: пашня – 5 000 га, труд – 300 000 чел.-час, возможный объем тракторных работ – 28 000 условных га. Найти оптимальное сочетание посевных площадей культур*.

Культуры	Затраты на 1 га посева		Стоимость валовой продукции с 1 га, р.
	Труда, чел.-час	Тракторных работ, усл. га	
Зерновые	30	4	400
Картофель	150	12	1000

Решение. Критерием оптимальности является максимум стоимости валовой продукции. Этот максимум должен достигаться в условиях использования ограниченных ресурсов пашни, труда и механизированных работ. В задаче имеется множество допустимых вариантов сочетания посевных площадей двух культур, но не все из них равнозначны с точки зрения требования оптимальности.

Для поиска оптимального решения задачи обозначим через x_1 га площадь, отводимую под зерновые, а через x_2 га — площадь, отводимую под картофель. Тогда стоимость зерновых составит $400\,x_1$ р., а стоимость картофеля — $1000\,x_2$ р. Отсюда стоимость всей валовой продукции составит $(400\,x_1 + 1000\,x_2)$ р. Обозначим это выражение через y: $y = 400\,x_1 + 1000\,x_2$

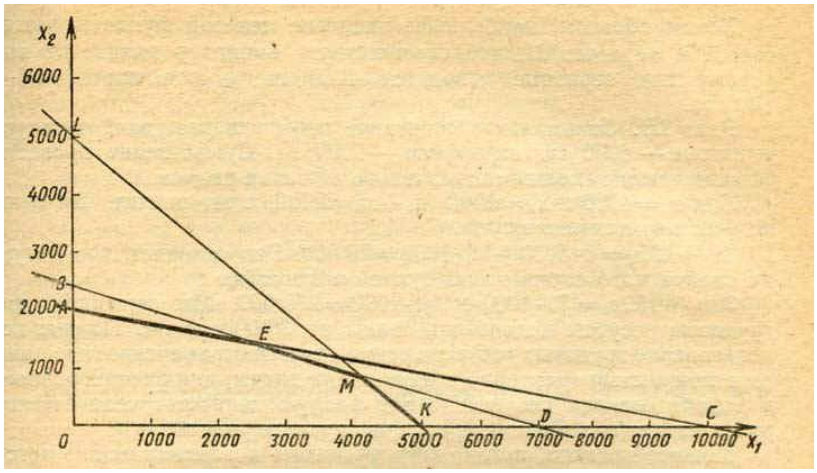
Нам надо найти максимум этой целевой функции при соблюдении следующих условий:

а) общая площадь зерновых и картофеля не должна превышать 5000 га, т. е. $x_1+x_2<5000$;
б) общие затраты труда не должны превосходить 300 тыс. человеко-часов, т. е. $30\,x_1 +150x_2 < 300\,000$ (или $x_1 +5x_2 < 10\,000$);

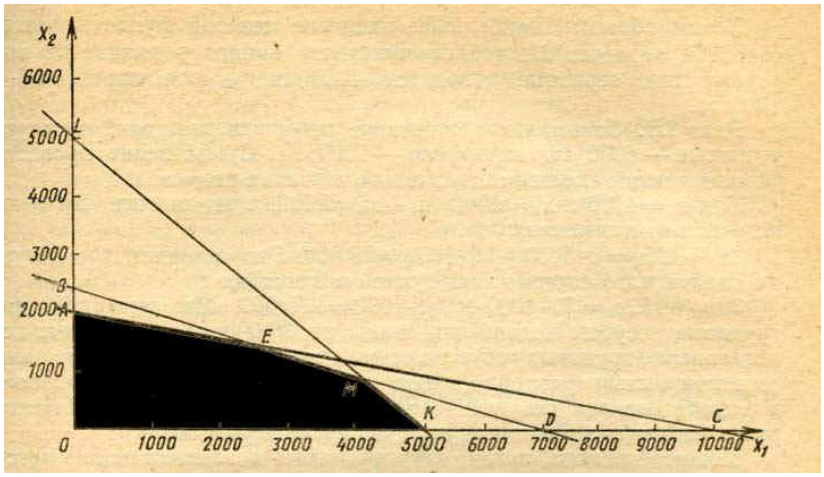
в) общий объем механизированных работ не должен превосходить 28 000 усл. га, т. е. 4 x_1 + 12x_2<28 000 (или x_1 + 3x_2<7 000);
г) площади, отводимые под зерновые и картофель, могут принимать только неотрицательные значения: x_1 >0; x_2>0
Таким образом, условия задачи выражаются следующей системой неравенств:

$$\begin{cases} x_1 + x_2 \le 5000 \\ x_1 + 5x_2 \le 10000 \\ x_1 + 3x_2 \le 7000 \\ x_1 \ge 0 \\ x_2 \ge 0 \end{cases}$$

Требуется найти такие значения x_1 и x_2, при которых функция y = 400 x_1 + 1000 x_2 принимает наибольшее значение.
Решение задачи было выполнено графическим способом: на координатной плоскости X_1X_2 были построены прямые $x_1 + x_2 = 5000$, $x_1 + 5x_2 = 10000$, $x_1 + 3x_2 = 7000$.



Затем была выделена область, состоящая из точек плоскости, координаты которых удовлетворяют системе. Этой областью являлся пятиугольник.



Для нахождения наибольшего значения функции y = 400 x_1 + 1000 x_2 были найдены ее значения в вершинах пятиугольника. Наибольшее значение функции было при x_1 = 4000, x_2 = 1000.

Таким образом, оптимальное сочетание посевных площадей культур: зерновые — 4000 га, картофель – 1000 га.
На конечном этапе решения задачи был проведен экономический анализ оптимального ее решения.
При x_1 =4000 и x_2 =1000: x_1 + x_2 = 5000, а это значит, что пашня используется полностью.

$4x_1 + 12x_2 = 4*4000 + 12*1000 = 28\ 000$. Это означает, что ресурсы тракторного парка используются полностью.
$30x_1 + 150x_2 = 30*4000 + 150*1000 = 270\ 000$. Это значит, что трудовые ресурсы недоиспользованы на 30000 чел.-ч. Полное использование трудовых ресурсов сдерживается ограниченностью пашни и мощностью тракторного парка.
Таким образом, для рассмотренной в задаче фермы ресурсы имеют разную ценность: человеческих рук в избытке, а механизированный труд дефицитен.
Можно придти к выводу, что в сельском хозяйстве также необходимо применение математических расчетов, методов и моделей для достижения наилучшего результата.

2. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1 Практическое занятие № 2 (2часа).

Тема: Математическая модель и этапы ее построения. Математические методы планирования эксперимента.

2.1.1 Задание для работы:

1. Математическая модель и этапы ее построения
2. Планирование эксперимента. Основы математического планирования эксперимента

2.1.2 Краткое описание проводимого занятия:

1. Общее понятие модели и моделирования.

Модель – это аналог чего-либо, образ исследуемого объекта, отражающий основные характеристики этого объекта.

Геометрическая (материальные)	←	МОДЕЛЬ ↓ Физическая (материальные)	→	Математическая (Знаковые)

Геометрическая модель – это некий объект, геометрически подобный своему прототипу (оригиналу). Основная особенность – дает внешнее представление об оригинале и цель демонстрации, (макет, репродукции и копии картин). Основная роль – это геометрическое подобие объектов, а не процессов которые протекают в них (топографо-геодезический макет местности не говорит о кругообороте воды в природе, а модель почвенного разреза, о физико-химических процессах протекающих в данном типе почвы).

Физическая модель – отражает подобие между оригиналом и моделью, с точки зрения подобия происходящих в них процессах. Процессы, протекающие в модели и ее аналоге имеют, одинаковую природу (исследователи предполагают проверку гидротехнических сооружений путем проведения испытаний, аналогичных объектов значительно меньших размеров, специально сконструированных для целей моделирования). В модели по сравнению с оригиналом меняются не только геометрические свойства, но и физические.

Математическая модель – описание объектов, явлений или процессов, с помощью знаков, символов, т.е. с использованием математического языка. Имеет вид некоторой совокупности уравнений и неравенств, таблиц, формул и т.д. Математическая структура отображает свойства объекта, проявляющиеся им в условном его существовании и развитии. Любая математическая модель подразумевает наличие определенных количественных показателей, характеристик объекта.

Основными числовыми характеристиками проекта землеустройства – это площадь (участки, контуры участка) или длина, при оговоренной ширине.

2. Типы и свойства моделей, модели и моделирование в землеустройстве

Целесообразность применения математических методов:

1. Математические модели позволяют принимать наиболее целесообразные решения по перераспределению, использованию и охране земельных ресурсов, от конкретных с/х предприятий до народного хозяйства в целом.

2. Оптимальные планы использования производственных ресурсов связанных с землей, способствует достижения заданных объема производства, при минимальных затратах труда и средств. В результате будет увеличиваться производительности труда.

3. Создаются наилучшие организационно-производственные условия, следовательно, повышение урожайности с/х культур, повышение плодородия, прекращение процессов эрозии, высоко-производственное использование техники.

4. Улучшение качества подготовки информации и ее использование, и землеустроительная наука получает возможность стать точной.

5. Улучшение экономических показателей, экологических, социальных, технических, проекта землеустройства.

6. Математические методы позволяют с большой точностью проверять и оценивать реальную значимость для теоретических моделей и концентрацию развития землевладения и землепользования на перспективу.

7. Это связующее звено между землеустройством, естественными и техническими науками, изучающими сельское хозяйство, как с природоохранительной, так и с экономической и социальной сторон.

8. Внедрение математических методов и вычислительной техники в землеустройство позволяет перестроить всю систему землеустроительного проектирования, организации планирования землеустроительных работ, освобождает значительное количество квалифицированных работников от малопродуктивного труда.

3. Общая характеристика экономико-математических методов и областей их применения при решении земельно-кадастровых задач.

Для решения землеустроительных задач различных классов, используют разнообразные виды экономических, математических моделей, позволяющих проводить анализ использования земельных ресурсов, выявить определенные тенденции и находить оптимальные варианты устройства территории.

Классификация математических моделей в землеустройстве

Классификационные признаки	Тип, вид, класс модели
Вид проектной документации	1 – графические (цифровые) 2 – экономические (числовые)
Степень определения информации	1 – детерминированные 2 – стохастические
Вид (форма) землеустройства или землеустроительного действия	1 – межотраслевые 2 – межхозяйственные 3 – внутрихозяйственные 4 – рабочего проекта
Математические методы лежащие в основе модели	1 – аналитические (дифференциал исчисления) 2 – экономико-статистические 3 – оптимизационные (математические программирование) 4 – межотраслевого баланса 5 – сетевого планирования и управления
Класс проекта землеустройства	1 – распределяется по 37 классам проекта землеустройства

В отличие от других экономических решений, землеустроительные решения всегда могут идентифицироваться на местности виде определенной пространственной ор-

ганизации территории, представленной системой севооборотов, полей, рабочих участков, дорог и составляет «скелет» хозяйства. Землеустроительная документация состоит из двух частей.

Землеустроительная документация

Графическая	Текстовая
графическая модель землеустроительной документации характеризует в числовом виде лишь условия производства (цифровая модель местности – рельеф)	основа расчетная связь, экономическая модель землеустроительной документации, позволяющая определить экономическую эффективность организации производства и территории

Графические математические модели дают характеристику различным элементам проекта землеустройства или их совокупности, которые показываются на проектном плане, к ним относится площадь, линейные и точечные объекты.

Площадные объекты – отдельные землевладения или землепользования, севообороты, их поля или рабочие участки, загоны очередного стравливания; гуртовые, отарные пастбища; сенокосы и т.д. Они характеризуются площадью, координатами поворотных точек и центром тяжести, что позволяет определить местоположение этих участков, их форму и другие параметры.

Линейные объекты – это линейные объекты организации территории, полевые и магистральные дороги, лесополосы, инженерные коммуникации (газопровод, ЛЭП), отдельные границы участков, зон и т.д. Эти объекты могут размещаться в виде прямых и ломаных линий, а также кривых. Они характеризуются протяженностью, а также шириной, координатами начальных, конечных и промежуточных точек.

Точечные объекты – позволяют определить на местности местоположение отдельных инженерных сооружений (колодцы, родники, буровые вышки и т.д.) их размещение, характеристики местоположения.

Экономические (числовые) – это обличенные в математический вид различные расчеты проектов землеустройства (математические модели агроэкономического обоснования проектов внутрихозяйственного землеустройства, технико-экономическое обоснование проектов межхозяйственного землеустройства, сметно-хозяйственный расчет рабочих проектов).

Детерминированные модели основываются на абсолютно точной информации, либо на сведениях, которые считаются точными.

Стохастический основываются на информации имеющей вероятностные характеристики, описывающие процесс, которые зависят от случайных величин, подчиняются законам теории вероятности (планирование урожайности с/х культур от будущих климатических условий задается в модели с определенной вероятностью, в зависимости от случайности изменяемых факторов).

Класс межотраслевых математических моделей, обеспечивает решение задач по прогнозированию и оптимизации планирования использования земельных ресурсов и их охране на уровне регионов: по стране в целом, в субъекте федерации, на территории местной администрации. Модели данного класса позволяют установить распределение земель по категориям земельного фонда страны (земли с/х назначения, промышленности, связи, обороны и иного специального назначения, лесного фонда запаса и т.д.), решить задачи по развитию агропромышленного производства в регионах, планирование и осуществление природоохранных мероприятий и т.д. Основным видом землеустроительных работ включающих модели этого класса, является разработка генеральных схем пользования и охраны земель страны.

Класс моделей межхозяйственного землеустройства, позволяет реализовать задачи по перераспределению земель между хозяйствами, по образованию и упорядочиванию землевладений и землепользований с/х и не с/х назначения, по установлению границ административно территориальных образований, границ населенных пунктов и т.д. К данному классу относят задачи по определению оптимальных размеров землепользо-

ваний и рациональному размещению производства на территории по наиболее целесообразным, ликвидным, недостаточным в использовании земельным ресурсам.

Класс моделей внутрихозяйственного землеустройства предназначен для решения вопросов наиболее полного рационального и эффективного использования земель и организации производства в конкретных с/х. предприятиях.

Основные задачи: установление оптимального сочетания отраслей, состава и площадей угодий; определение видов, количество и площадей севооборотов и их размещение; рациональная организация кормопроизводства; планирование грузоперевозок и комплекса мелиоративных работ; установление оптимальных размеров производства и т.д.

Класс моделей рабочего проектирования, обеспечивает решение различных задач землеустроительного проектирования, связанных с землеустройством конкретных участков и инвестиций в эти земли (задачи по созданию орошаемых культурных пастбищ на территории отдельных хозяйств, выглаживание оврагов, трансформация и мелиорация земельных участков, строительство дорог и дорожных сооружений, закладка многолетних сооружений и т.д.).

Сложность математических моделей каждого класса зависит от числа учитываемых факторов и характеристик, взаимосвязанных между ними, от наличия точности и достоверности исходной информации и непосредственно от сложности изучаемого процесса.

Для построения и использования этих моделей необходимо знать стандарт экономико-математического метода (методы математической статистики и методы математического программирования). Вместе с тем ряд землеустроительных математико-экономических моделей требуют разработки своего нестандартного математического решения, своих индивидуальных математических методов.

Каждый вид данного экономико-математического моделирования имеют свои стадии (этапы построения и использования).

Аналитические методы в землеустройстве основываются на применении математических методов алгебры и геометрии, дифференциального и интегрального исчисления, и т.д., имеющих функциональный характер, т.е. каждому набору значений факторов независимых переменных соответствует строго определенное значение результатов.

Экономико-статистические модели базируются на использовании теории вероятностей и методов математической статистики (корреляционный, регрессионный, дисперсионный, теорию выборок и т.д.). Главное место среди них занимает производственная функция, представляющая собой уравнение связей зависимой переменной (результатом) и факторов (аргументов). С помощью этих моделей рассчитывается прогноз урожайности культур, продуктивности животноводства, а также некоторые параметры организации территорий (распаханность, облесенность, освоенность). При помощи экономико-математической моделей, осуществляют также анализ уровня использования земель, подготавливают информацию для применения оптимизационных методов, производственно обосновывают землеустроительные проектные решения.

Функциональные ЭСМ Корреляционные

ЭСМ аналогичны аналитическим моделям, но только основаны на статистической информации, т.к. наличие функций, т.е. однозначной связи между результатом и фактором в экономике чрезвычайно редкое явление, то данная разновидность модели в землеустройстве встречается не часто

Базируются на корреляционном уравнении связи между факторами, а также между фактором и результатом. Неопределенность либо обуславливается как природой моделирующего процесса (объект причины), так и погрешностью исходной информации о процессе (субъект причины)

Оптимизационные ЭСМ основаны главным образом на методах математического программирования, позволяющие находить экстремальные значения целевой функции при заданных условиях.

Комбинированная Оптимизационные модели в землеустройстве → ← Дифференцированная

Все вопросы землеустроительного проекта решаются комплексно, в их взаимообусловленности и взаимозависимости. Этот вид модели является наиболее полным, но приводит к громоздким задачам, решение которых затруднено

Заключается в последовательном моделировании некоторых частичных задач проекта. Модели получаются значительно меньшего объема и их применение значительно облегчается.

Применение дифференцированных моделей в землеустройстве объясняется сложностью и многообразием решаемых вопросов, что приводит к построению упрощенных моделей. Дифференцирование моделей связано с аппроксимацией комбинированных моделей.

Аппроксимация реализуется в следующих видах:

– либо модель отражает часть сложной системы без учета всех других ее сторон – частичная аппроксимация;

– либо модель упрощается, чтобы быть в дальнейшем запрограммированной с последующим наращиванием информации – полная аппроксимация.

2.1.3 Результаты и выводы:

Стадии экономико-статистических моделей.

1) Экономический анализ производства, определение зависимой переменной и выявление факторов влияющих на её значение.

2) Получение статистических данных и их обработка.

3) Определение математической формы связи независимых и зависимых переменных.

4) Получение (определение) параметров ЭСМ.

5) Оценка степени соответствия ЭСМ изучаемому процессу.

6) Экономическая интерпретация модели, её использование, для конкретных землеустроительных целей.

Методы получения статистических данных:

а) Экспериментальный метод – информация полученная в результате проведения опытов условия которых позволяют контролировать процесс получения данных. Редко применяется в землеустройстве.

б) Основан на использовании статистической информации:

Сплошные. Выборочные.

2.2 Практическое занятие № 2-3 (4часа).

Тема: Основы статистической обработки результатов наблюдения. Элементы теории ошибок. Обоснование числа измерений. Использование надстроек Excel.

2.2.1 Задание для работы:

1. Статистическая обработка результатов измерений
2. Элементы теории ошибок.
3. Обоснование числа измерений.
4. Использование надстроек Microsoft Excel.

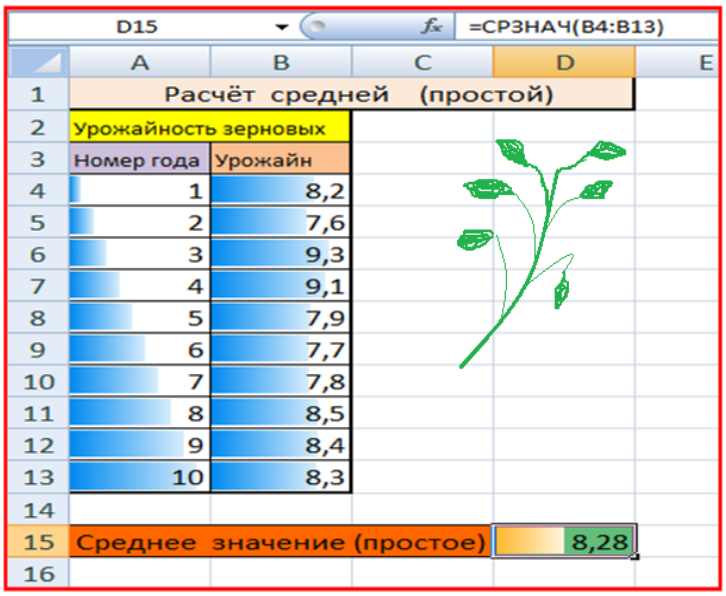
2.2.2 Краткое описание проводимого занятия:

«Расчёт средних и показателей вариации с Excel»

1. Расчёт средней арифметической величины (простой) производится по формуле

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad \overline{x} = 8,28.$$

Номер значения: x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
Величина x_i	8,2	7,6	9,3	9,1	7,9	7,7	7,8	8,5	8,4	8,3

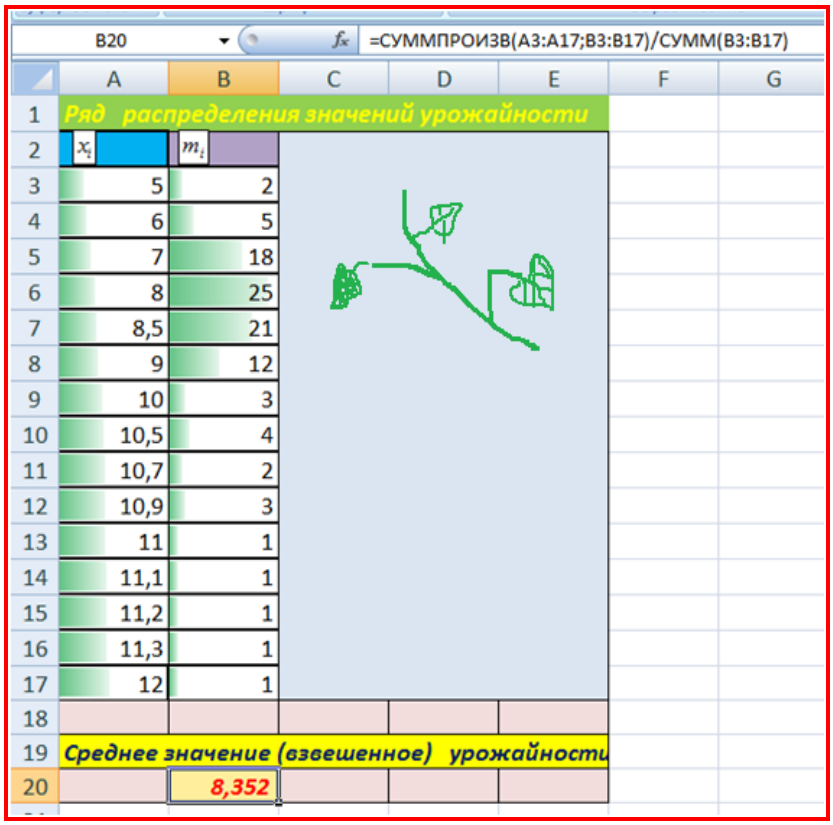


2. Расчёт средней арифметической (взвешенной) производится по формуле

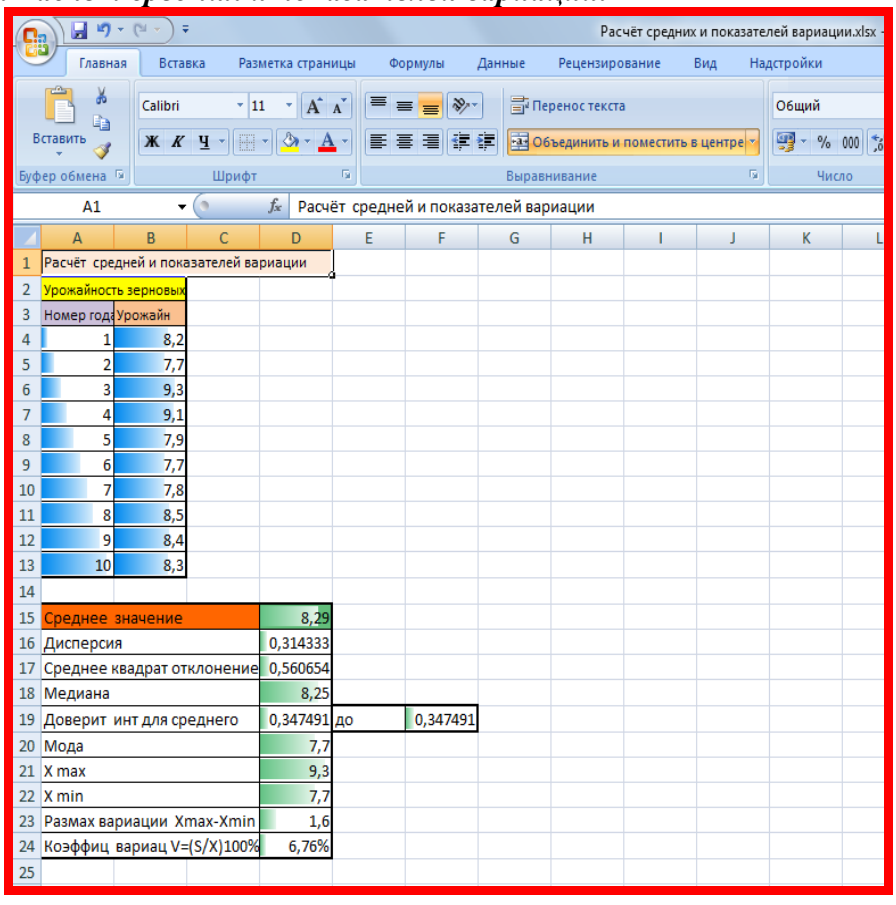
$$\overline{x} = \frac{x_1 \cdot m_1 + x_2 \cdot m_2 + x_3 \cdot m_3 + \ldots + x_n \cdot m_n}{n} = \frac{\sum_{i=1}^{n} x_i \cdot m_i}{n}, \qquad \overline{x} = 8,28.$$

x_i	5	6	7	8	8,5	9	10	10,5	10,7	10,9	11	11,1	11,2	11,3	12
m_i	2	5	18	25	21	12	3	4	2	3	1	1	1	1	1

$n = 100$



3. Расчёт средних и показателей вариации.



Понятие ошибки выборки

Задача выборочного наблюдения - дать верное представление о сводных показателях всей совокупности факторов на основе некоторой их части, подвергнутой обследованию, т.е. определение характеристик генеральной совокупности по выборочным данным. Чаще других при выборочном наблюдении исследуется либо среднее значение того или иного признака у единиц совокупности (например, средняя урожайность, средняя заработная плата и т.д.), либо доля единиц обладающих тем или иным признаком, т.е. удельный вес определённых единиц в совокупности (например, доля орошаемых земель, доля отдельных пород деревьев в лесном массиве и т.д.).

Поскольку речь идёт о варьирующих признаках и изучают не всю совокупность единиц, а только их часть, то можно заранее сказать, что сводные показатели по этим признакам у части единиц совокупности почти никогда не будут абсолютно совпадать со сводными показателями всей статистической совокупности. Выборочные показатели, как правило, не совпадают с соответствующими показателями генеральной совокупности, а несколько отличаются от них в одну или другую сторону, т.е. при выборочном наблюдении всегда могут возникнуть ошибки, которые можно подразделить на ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации при выборочном наблюдении, как и при сплошном, могут возникнуть по разным причинам: и по вине того, кто проводит наблюдение, и по вине отвечающего на те или иные вопросы, и от способа наблюдения. Но если тщательно провести подготовку кадров и продумать организацию проведения наблюдения, то в силу ограниченности выборочной совокупности (по сравнению с генеральной совокупностью) ошибки регистрации можно свести к минимуму или, во всяком случае, уменьшить их по сравнению с ошибками регистрации сплошного наблюдения.

Ошибка репрезентативности (представительства) свойственна лишь выборочному наблюдению и представляет собой величину возможных расхождений между показателями выборочной и генеральной совокупности.

Ошибки репрезентативности в свою очередь могут иметь случайный характер и систематический.

Систематическая ошибка - это ошибка, тенденциозно искажающая величину исследуемого признака в сторону её увеличения или уменьшения. Возникает она главным образом в результате нарушения случайности отбора.

Случайная ошибка - это ошибка, имеющая одинаковую величину вероятности в сторону уменьшения или увеличения изучаемого показателя; это ошибка, появление которой возможно в результате сущности содержания самого выборочного (не сплошного) наблюдения, в силу того, что исследуется часть, а не вся статистическая совокупность.

Определение величины случайных ошибок репрезентативности и является одной из главных задач теории выборочного метода. Их фиксирование позволяет судить о точности выборки, о возможности распространения выборочных характеристик на генеральную совокупность.

Случайные ошибки выборки определяются по формулам, разработанным на основе теории вероятностей и носят вероятностный характер.

3.2 Методы определения ошибки выборки

Возможные расхождения между характеристиками выборочной и генеральной совокупности измеряются средней ошибкой выборки u &. В математической статистике, которая лежит в основе всех расчётов показателей выборочных совокупностей, доказывается, что значения средней ошибки выборки определяются по формуле:

$$\mu = \sqrt{\frac{\sigma_0^2}{n}}$$

где: m- средняя ошибка выборки; s^2 генеральная дисперсия; n - численность единиц выборочной совокупности.

Использование данной формулы предполагает, что известна генеральная дисперсия. Но при проведении выборочных исследований эти показатели, как правило, неизвестны.

Применение выборочного метода как раз и предполагает определение характеристик генеральной совокупности.
На практике для определения средней ошибки выборки обычно используются дисперсии выборочной совокупности. Эта замена основана на том, что при соблюдении принципа случайного отбора дисперсия достаточно большого объёма выборки стремиться отобразить дисперсию в генеральной совокупности.
В математической статистике доказано следующее соотношение между дисперсиями в генеральной и выборочной совокупностях:

$$\sigma_0^2 = \sigma^2\left(\frac{n}{n-1}\right)$$

Из приведённой формулы видно, что дисперсия выборочной совокупности меньше дисперсии в генеральной совокупности на величину определяемую отношением:

$$\frac{n}{n-1}$$

Если n достаточно велико, то данное отношение близко к единице.
Например, при n = 100 оно равно 1,01, а при n = 500 оно равно 1,002. Поэтому с определённой долей погрешности формулу расчёта средней ошибки выборки можно представить в следующем виде.

$$\mu \approx \sqrt{\frac{\sigma^2}{n}}$$

Однако следует иметь в виду, что данная формула применяется для определения средней ошибки выборки лишь при повторном отборе. Поскольку при бесповторном отборе численность генеральной совокупности N в ходе выборки сокращается, то в формулу для расчёта n средней ошибки выборки включают дополнительный множитель. Формула средней ошибки выборки принимает следующий вид:

$$\mu \approx \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}$$

Для практики выборочных обследований важно, что средняя ошибка выборки применяется для установления предела отклонений характеристик выборки из соответствующих показателей генеральной совокупности. Лишь с определённой степенью вероятности можно утверждать, что эти отклонения не превысят величины t u, которая в статистике называется предельной ошибкой выборки.
Предельная ошибка выборки связана со средней ошибкой выборки u отношением:

$$\Delta = t\mu$$

При этом t как коэффициент кратности средней ошибки выборки зависит от вероятности, с которой гарантируется величина предельной ошибки выборки. Обычно в практике экономических исследований обычно ограничиваются значением t не превышающим двух трёх единиц.

Обоснование числа измерений

Для проведения опытов с заданной точностью и достоверностью крайне важно знать то количество измерений, при котором экспериментатор уверен в положительном исходе. В связи с этим одной из первоочередных задач при статических методах оценки является установление минимального, но достаточного числа измерений для данных условий. Задача сводится к установлению минимального объема выборки (числа измерений) $N_{\min}$ при заданных значениях доверительного интервала 2μ и доверительной вероятности. При выполнении измерений крайне важно знать их точность:

$$\Delta = \sigma_o / \bar{x},$$

где σ_o - среднеарифметическое значение среднеквадратичного отклонения σ, равное $\sigma_o / \sqrt{n}.$

Значение σ_o часто называют *средней ошибкой.* Доверительный интервал ошибки измерения Δ определяется аналогично для измерений $\mu = t\sigma_o$. С помощью *t* легко определить доверительную вероятность ошибки измерений из таблиц.

В исследованиях часто по заданной точности Δ и доверительной вероятности измерения определяют минимальное количество измерений, гарантирующих требуемые значения Δ и $p_д$.

Можно получить $\mu = \sigma \arg \varphi(p_д) = \sigma_o / \sqrt{n}t.$

При $N_{\min} = n$ получаем $N_{\min} = \sigma^2 t^2 / \sigma_o^2 = k_в^2 t^2 / \Delta^2,$

здесь $k_в$ – коэффициент вариации (изменчивости), %; Δ - точность измерений, %.

Для определения $N_{\min}$ может быть принята такая последовательность вычислений:
1) проводится предварительный эксперимент с количеством измерений *n,* составляет в зависимости от трудоемкости опыта от 20 до 50;
2) вычисляется среднеквадратичное отклонение σ;
3) в соответствии с поставленными задачами эксперимента устанавливается требуемая точность измерений Δ, которая не должна превышать точности прибора;
4) устанавливается нормированное отклонение *t,* значение которого обычно задается (зависит также от точности метода);

5) определяют $N_{\min}$ и тогда в дальнейшем в процессе эксперимента число измерений не должно быть меньше $N_{\min}$.

17. Критерий Стьюдента (Госсета).

Оценки измерений с помощью σ и $\sigma_{ш}$ по приведенным методам справедливы при *n* > 30. Стоит сказать, что для нахождения границы доверительного интервала при малых значениях применяют метод, предложенный английским математиком В.С .Госсетом (псевдоним Стьюдент). Кривые распределения Стьюдента в случае $n \to \infty$ (практически при n >20) переходят в кривые нормального распределения (рис.1).

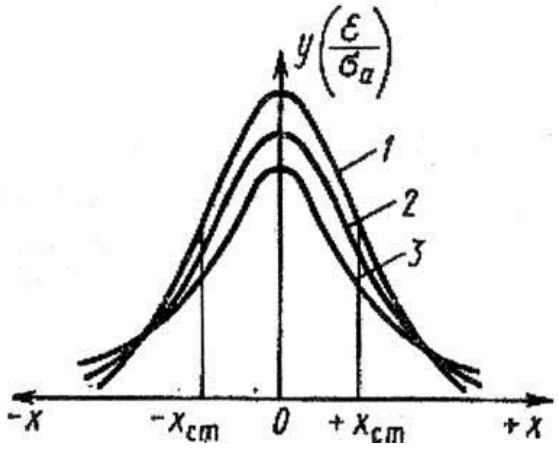


Рис.1. Кривые распределения Стьюдента для различных значений:
1 - $n \to \infty$; 2 - $n = 10$; 3 - $n = 2$

Для малой выборки доверительный интервал

$$\mu_{cm} = \sigma_o \alpha_{cm},$$

где α_{cm} – коэффициент Стьюдента, принимаемый по табл.10.2 в зависимости от значения доверительной вероятности $p_д$.

Зная μ_{cm}, можно вычислить действительное значение изучаемой величины для малой выборки

$$x_{\partial} = \bar{x} \pm \mu_{cm}.$$

Возможна и иная постановка задачи. По *n* известных измерений малой выборки крайне важно определить доверительную вероятность p_{∂} при условии, что погрешность среднего значения не выйдет за пределы $\pm \mu_{cm}$. Задачу решают в такой последовательности: вначале вычисляется среднее значение $\bar{x}, \sigma_o$ и $\alpha_{cm} = \mu_{cm} / \sigma_o$. С помощью величины α_{cm}, известного *n* и табл.8.2 определяют доверительную вероятность.

В процессе обработки экспериментальных данных следует исключать грубые ошибки ряда. Появление этих ошибок вполне вероятно, а наличие их ощутимо влияет на результат измерении. При этом прежде чем исключить то или иное измерение, крайне важно убедиться, что это действительно грубая ошибка, а не отклонение вследствие статистического разброса. Известно несколько методов определ□ения грубых ошибок статистического ряда. Наиболее простым способом исключения из ряда резко выделяющегося измерения является правило трех сигм: разброс случайных величин от среднего значения не должен превышать

$$x_{\max,\min} = \bar{x} \pm 3\sigma.$$

Более достоверными являются методы, базируемые на использовании доверительного интервала. Пусть имеется статистический ряд малой выборки, подчиняющийся закону нормального распредел□ения. При наличии грубых ошибок критерии их появления вычисляются по формулам

$$\beta_1 = (x_{\max} - \bar{x}) / \sigma\sqrt{(n-1)/n}; \quad \beta_2 = (\bar{x} - x_{\min}) / \sigma\sqrt{(n-1)/n},$$

где $x_{\max}, x_{\min}$ – наибольшее и наименьшее значения из *n* измерений.

В табл. приведены в зависимости от доверительной вероятности максимальные значения $\beta_{\max}$, возникающие вследствие статистического разброса. В случае если $\beta_1 > \beta_{\max}$, то значение $x_{\max}$ крайне важно исключить из статистического ряда как грубую погрешность. При $\beta_2 < \beta_{\max}$ исключается величина $x_{\min}$. После исключения грубых ошибок определяют новые значения x и σ из $(n-1)$ или $(n-2)$ измерений.

2.2.3 Результаты и выводы:

Таблица

Критерий появления грубых ошибок

n	$\beta_{\max}$ при p_{∂}			*n*	$\beta_{\max}$ при p_{∂}		
	0,90	0,95	0,99		0,90	0,95	0,99
3	1,41	1,41	1,41	15	2,33	2,49	2,80
4	1,64	1,69	1,72	16	2,35	2,52	2,84
5	1,79	1,87	1,96	17	2,38	2,55	2,87
6	1,89	2,00	2,13	18	2,40	2,58	2,90
7	1,97	2,09	2,26	19	2,43	2,60	2,93
8	2,04	2,17	2,37	20	2,45	2,62	2,96
9	2,10	2,24	2,46	25	2,54	2,72	3,07
10	2,15	2,29	2,54	30	2,61	2,79	3,16
11	2,19	2,34	2,61	35	2,67	2,85	3,22
12	2,23	2,39	2,66	40	2,72	2,90	3,28
13	2,26	2,43	2,71	45	2,76	2,95	3,33
14	2,30	2,46	2,76	50	2,80	2,99	3,37

2.3 Практическое занятие № 4-6 (6часа)

Тема: Проверка статистических гипотез. Уровень значимости. Критерии. Примеры. Оценка чувствительности критерия при проверке значимости различий. Двухвыборочный t - тест в Excel.

2.3.1 Задание для работы:

1. Проверка статистических гипотез. Уровень значимости. Критерии. Примеры.
2. Оценка чувствительности критерия при проверке значимости различий. Двухвыборочный t - тест в Excel

2.3.2 Краткое описание проводимого занятия:

1. Статистические гипотезы и их виды. Критерии согласия.

Пример. Для подготовки к зачету преподаватель сформулировал 100 вопросов (генеральная совокупность) и считает, что студенту можно поставить «зачтено», если тот знает 60 % вопросов (критерий). Преподаватель задает студенту 5 вопросов (выборка из генеральной совокупности) и ставит «зачтено», если правильных ответов не меньше трех. Гипотеза H_0: «студент курс усвоил», а множество $\{3, 4, 5\}$ — область принятия этой гипотезы. Критической областью является множество $\{0, 1, 2\}$ — правильных ответов меньше трех, в этом случае основная гипотеза отвергается в пользу альтернативной H_1 «студент курс не усвоил, знает меньше 60 % вопросов».

Студент А выучил 70 вопросов из 100, но ответил правильно только на два из пяти, предложенных преподавателем, — зачет не сдан. В этом случае преподаватель совершает ошибку первого рода.

Студент Б выучил 50 вопросов из 100, но ему повезло, и он ответил правильно на 3 вопроса — зачет сдан, но совершена ошибка второго рода.

Преподаватель может уменьшить вероятность этих ошибок, увеличив количество задаваемых на зачете вопросов.

Алгоритм проверки статистических гипотез сводится к следующему:

1) сформулировать основную H_0 и альтернативную H_1 гипотезы;

2) выбрать уровень значимости α;

3) в соответствии с видом гипотезы H_0 выбрать статистический критерий для ее проверки, т. е. случайную величину К, распределение которой известно;

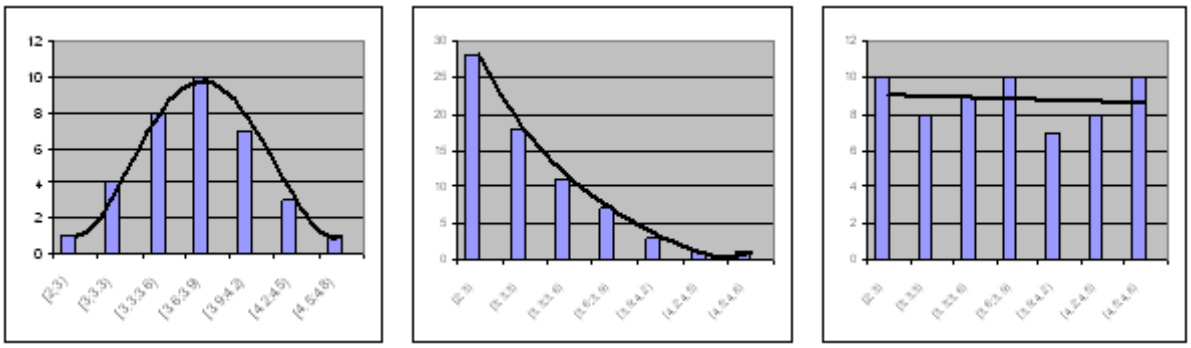
4) по таблицам распределения случайной величины К найти границу критической области $K_{кр}$ (вид критической области определить по виду альтернативной гипотезы H_1);

5) по выборочным данным вычислить наблюдаемое значение критерия $K_{набл}$;

6) принять статистическое решение: если $K_{\text{набл}}$ попадает в критическую область — отклонить гипотезу H_0 в пользу альтернативной H_1; если $K_{\text{набл}}$ попадает в область допустимых значений, то нет оснований отклонять основную гипотезу.

1. Критерии согласия. Критерии однородности.

Одной из важных задач математической статистики является установление теоретического закона распределения случайной величины, характеризующей изучаемый признак по эмпирическому распределению, представляющему вариационный ряд. Предположение о виде закона распределения можно сделать по гистограмме или полигону (Рис. 1)



A) B) C)

Рис.1. Возможные виды гистограмм:
а) нормального, б) показательного, в) равномерного распределений

Например, по гистограмме (рис. 1, а)) можно сделать предположение о том, что генеральная совокупность распределена по нормальному закону.

Для проверки гипотез о виде распределения служат специальные критерии — *Критерии согласия*. Они отвечают на вопрос: согласуются ли результаты экспериментов с предположением о том, что генеральная совокупность имеет заданное распределение.

Проверим это предположение с помощью ***Критерия согласия Пирсона***. В этом критерии мерой расхождения между гипотетическим (предполагаемым) и эмпирическим распределением служит статистика

$$K = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j},$$

где N — объем выборки; K — количество интервалов (групп наблюдений); n_j — количество наблюдений, попавших в J-й интервал; p_j — вероятность попадания в J-й интервал случайной величины, распределенной по гипотетическому закону.

Если предположение о виде закона распределения справедливо, то статистика Пирсона распределена по закону «хи-квадрат» с числом степеней свободы $k - r - 1$ (R — число параметров распределения, оцениваемых по выборке): $K \sim \chi^2_{(k-r-1)}$.

Оцениваются неизвестные параметры с использованием теории точечных оценок, некоторые оценки приведены в табл. 1.

Таблица 1. Оцениваемые параметры и их точечные оценки

Вид распределения	Оцениваемые параметры	Точечные оценки параметров
$f_X(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\ge 0\end{cases}$	$\lambda>0$	$\overline{\lambda}=\frac{1}{\overline{x}_{в}}$
$f_X(x)=\begin{cases}0, & x<\alpha\\ 1/(\beta-\alpha), & x\in[\alpha,\beta]\\ 0, & x>\beta\end{cases}$	$\alpha,\ \beta$	$\alpha=\overline{x}_{в}-\sqrt{3}\cdot\sqrt{D_{в}}$ $\beta=\overline{x}_{в}+\sqrt{3}\cdot\sqrt{D_{в}}$
$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-m)^2}{2\sigma^2}}$	$m,\ \sigma^2$	$m=\overline{x}_{в},\quad \sigma^2=\sqrt{D_{в}}$

Здесь $\overline{x}_{в}=\frac{\sum_{i=1}^{m}x_i\cdot n_i}{n}$, $D_{в}=\frac{\sum_{i=1}^{m}(x_i-\overline{x}_{в})^2\cdot n_i}{n}$

Количество интервалов *K* рекомендуется рассчитывать по формуле Стерджеса $k=1+3.3\cdot\lg n$, где *N* — объем выборки. Длину *I*-го интервала принимают равной $h=\frac{x_{(n)}-x_{(1)}}{k}$, где $x_{(n)}$—наибольшее, а $x_{(1)}$ — наименьшее значение в вариационном ряду.

Пример 1. Для среднего балла среди 30-ти групп (с точностью до сотых долей балла) получили выборку x_i:

3.7, 3.85, 3.7, 3.78, 3.6, 4.45, 4.2, 3.87, 3.33, 3.76, 3.75, 4.03, 3.8, 4.75, 3.25, 4.1, 3.55, 3.35, 3.38, 3.05, 3.56, 4.05, 3.24, 4.08, 3.58, 3.98, 3.4, 3.8, 3.06, 4.38. Проверить гипотезу о нормальном распределении среднего балла на уровне значимости $\alpha=0.025$.

Решение. Сгруппируем эту выборку. Наименьший средний балл равен 3.05, наибольший — 4.75. Интервал [3; 4.8] разобьем на 6 частей длиной $h=0.3$, применяя формулу Стерджеса ($k=5.875\approx 6$). Подсчитаем частоту n_i (относительную частоту $\frac{n_i}{n}$) для каждого интервала и получим сгруппированный статистический ряд (табл. 2).

Таблица 2. Статистический ряд

Интервалы	[3;3.3)	[3.3;3.6)	[3.6;3.9)	[3.9;4.2)	[4.2;4.5)	[4.5;4.8)
Частоты n_i	4	7	10	5	3	1
Относительные частоты $\frac{n_i}{n}$	0.133	0.233	0.3	0.167	0.1	0.033

Сформулируем основную и альтернативную гипотезы.

$H_0: X\sim N(\overline{a},\ \overline{\sigma})$ — случайная величина *X* (средний балл) подчиняется нормальному закону с параметрами $\overline{a},\ \overline{\sigma}$. Так как истинных значений параметров $a,\ \sigma$ мы не знаем, возьмем их оценки, рассчитанные по выборке: $\overline{a}=3.746,\ \overline{\sigma}=0.399$.

H_1: случайная величина *X* не подчиняется нормальному закону с данными параметрами. Рассчитаем наблюдаемое значение $K_{\text{набл}}$ статистики Пирсона. Эмпирические частоты n_j уже известны (табл. 2), а для вычисления вероятностей p_j (в предположении, что гипотеза H_0 справедлива) применим уже известную формулу (свойство **В**):

И таблицу функции Лапласа. Полученные результаты сведем в таблицу (табл. 3). Наблюдаемое значение статистики Пирсона равно $K_{\text{набл}} = 0.978.$

Определим границу критической области. Так как статистика Пирсона измеряет разницу между эмпирическим и теоретическим распределениями, то чем больше ее наблюдаемое значение $K_{\text{набл}}$, тем сильнее довод против основной гипотезы. Поэтому критическая область для этой статистики всегда правосторонняя: $[K_{\text{кр}};\ +\infty).$ Её границу $K_{\text{кр}} = \chi^2_{(k-r-1;\,\alpha)}$ находим по таблицам распределения «хи-квадрат» и заданным значениям $\alpha = 0.025,\ k = 6$ (число интервалов), $r = 2$ (параметры a и σ оценены по выборке): $K_{\text{кр}} = \chi^2(6-2-1;\ 0.025) = \chi^2(3;\ 0.025) = 9.4.$

Наблюдаемое значение статистики Пирсона не попадает в критическую область: $K_{\text{набл}} < K_{\text{кр}}$, поэтому н*ет оснований отвергать основную гипотезу*.

Вывод: на уровне значимости 0.025 справедливо предположение о том, что средний балл имеет нормальное распределение

Таблица 3. Сравнение наблюдаемых и ожидаемых частот

№ п/п	Интервалы группировки $[a_j; a_{j+1})$	Наблюдаемая частота n_j	Вероятность p_j попадания в *J*-й интервал	Ожидаемая частота $n \cdot p_j$	Слагаемые статистики Пирсона $\dfrac{(n_j - np_j)^2}{np_j}$
1.	[3; 3.3)	4	0.101	3.032	0.309
2.	[3.3; 3.6)	7	0.225	6.761	0.008
3.	[3.6; 3.9)	10	0.295	8.79	0.166
4.	[3.9; 4.2)	5	0.222	6.665	0.416
5.	[4.2; 4.5)	3	0.098	2.946	0.001
6.	[4.5; 4.8)	1	0.025	0.758	0.077
$\sum$	—	30	0.965	28.95	$K_{\text{набл}} = 0.978.$

Пример 2. Проверить с помощью критерия χ2 при уровне значимости 0,05 гипотезу о том, что выборка объема $n = 50$, представленная интервальным вариационным рядом в таблице 4, извлечена из нормальной генеральной совокупности.
Таблица 4

Номер интервала I	Границы интервала	Частота m_i
1	0 – 2	5
2	2 – 4	11
3	4 –6	17
4	6 – 8	10
5	8 – 10	7

Решение.

1. Сформулируем нулевую и альтернативную гипотезы: H_0 – эмпирическое распределение соответствует нормальному; H_1 - эмпирическое распределение не соответствует нормальному.

Для проверки нулевой гипотезы необходимо рассчитать наблюдаемое значение критерия $\chi 2$набл по формуле и сравнить его с критическим значением $\chi 2$кр.

2. Определим параметры предполагаемого (теоретического) нормального закона распределения.

Найдем середины интервалов $\overline{x_i} = \frac{z_{i-1} + z_i}{2}$ и относительные частоты $p_i^* = \frac{m_i}{n}$. Получим следующие значения:

$\overline{x_i}$	1	3	5	6	7
p_i^*	$\frac{5}{50}$	$\frac{11}{50}$	$\frac{17}{50}$	$\frac{10}{50}$	$\frac{7}{50}$

Найдем оценку математического ожидания:

$$m^* = \left(1 \cdot \frac{5}{50} + 3 \cdot \frac{11}{50} + 5 \cdot \frac{17}{50} + 7 \cdot \frac{10}{50} + 9 \cdot \frac{7}{50}\right) = \frac{1}{50}(5 + 33 + 85 + 70 + 63) = \frac{256}{50} = 5{,}12.$$

Вычислим оценки дисперсии и стандартного отклонения по формулам:

$$s^2 = \left((1 - 5{,}12)^2 \cdot 5 + (3 - 5{,}12)^2 \cdot 11 + (5 - 5{,}12)^2 \cdot 17 + (7 - 5{,}12)^2 \cdot 10 + (9 - 5{,}12)^2 \cdot 7\right) \cdot \frac{1}{49} = \frac{1}{49} \cdot 275{,}27 = 5{,}62;$$

$$s = \sqrt{5{,}62} = 2{,}37.$$

3. Выполним расчет теоретических частот m_i^T.

$$p_1 = \Phi\left(\frac{2 - 5{,}12}{2{,}37}\right) - \Phi(-\infty) = \Phi(-1{,}3) - 0 = 1 - \Phi(1{,}3) = 1 - 0{,}9 = 0{,}1$$

и

$$m_1^T = 50 \cdot 0{,}1 = 5;$$

Для интервала $(2,4)$ находим

$$p_2 = \Phi\left(\frac{4-5{,}12}{2{,}37}\right) - \Phi\left(\frac{2-5{,}12}{2{,}37}\right) = \Phi(-0{,}5) - \Phi(-1{,}3) = 0{,}90 - 0{,}69 = 0{,}21$$

и $m_2^T = 50 \cdot 0{,}21 = 10{,}5$;

Для интервала (4,6) соответственно:

$$p_3 = \Phi\left(\frac{6 - 5{,}12}{2{,}37}\right) - \Phi\left(\frac{4 - 5{,}12}{2{,}37}\right) = \Phi(0{,}4) - \Phi(-0{,}5) = 0{,}66 - 0{,}31 = 0{,}35;$$

$$m_3^T = 50 \cdot 0{,}35 = 17{,}5;$$

Для интервала (6,8):

$$p_4 = \Phi\left(\frac{8 - 5{,}12}{2{,}37}\right) - \Phi\left(\frac{6 - 5{,}12}{2{,}37}\right) = \Phi(1{,}2) - \Phi(0{,}4) = 0{,}88 - 0{,}66 = 0{,}22$$

И $m_4^T = 50 \cdot 0{,}22 = 11$;

Для интервала $(8,\infty)$ вычислим

$$p_5 = \Phi(\infty) - \Phi\left(\frac{8 - 5{,}12}{2{,}37}\right) = \Phi(\infty) - \Phi(1{,}2) = 1 - 0{,}88 = 0{,}12;$$

$$m_5^T = 50 \cdot 0{,}12 = 6.$$

4. По формуле (4.8) найдем значение $\chi^2_{набл}$:

$$\chi^2_{набл} = \sum_{i=1}^{5} \frac{\left(m_i - m_i^T\right)^2}{m_i^T} = \frac{(5-5)^2}{5} + \frac{(11-10,5)^2}{10,5} +$$

$$+ \frac{(17-17,5)^2}{17,5} + \frac{(10-11)^2}{11} + \frac{(7-6)^2}{6} = 0,29$$

.

5. По таблице квантилей распределения χ2 с числом степеней свободы $r = k - 3 = 5 - 3 = 2$ находим, что χ2кр = 6,0 для $\alpha = 0,05$.

Поскольку $\chi^2_{набл} < \chi^2_{кр}$ ($0,29 < 6,0$), то можно считать, что гипотеза о нормальном распределении генеральной совокупности не противоречит опытным данным.

Пример 3: Используя критерий Пирсона, при уровне значимости а = 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема n = 200.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение.

1. Вычислим $\bar{x}_B = \frac{\sum_{i=1}^{k} x_i n_i}{n} = 12,63$ и выборочное среднее квадратическое отклонение $\sigma_B = \sqrt{\overline{x_B^2} - \left(\bar{x}_B\right)^2} = 4,695$.

2. Вычислим теоретические частоты учитывая, что n = 200, h = 2, σ_B = 4,695, по формуле

$$n_i = \frac{200 \cdot 2}{4,695} \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right) = 85,2 \cdot \varphi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right).$$

Составим расчетную таблицу

i	x_i	$u_i = \frac{\left(x_i - \bar{x}_B\right)}{\sigma_B}$	$\varphi(u_i)$	$n_i = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i)$
1	5	-1,62	0,1074	9,1
2	7	-1,20	0,1942	16,5
3	9	-0,77	0,2966	25,3
4	11	-0,35	0,3752	32,0
5	13	0,08	0,3977	33,9
6	15	0,51	0,3503	29,8
7	17	0,93	0,2589	22,0
8	19	1,36	0,1582	13,5
9	21	1,78	0,0818	7,0

3. Сравним эмпирические и теоретические частоты. Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия $\chi^2_{набл} = \sum_{i=1}^{s} \frac{\left(n_i - n_i'\right)^2}{n_i'}$:

i	n_i	n_i	$\left\|n_i - n_i'\right\|$	$\left(n_i - n_i'\right)^2$	$\frac{\left(n_i - n_i'\right)^2}{n_i'}$
1	15	9,1	5,9	34,81	3,8
2	26	16,5	9,5	90,25	5,5
3	25	25,3	0,3	0,09	0,0
4	30	32,0	2,0	4,0	0,1
5	26	33,9	7,9	62,41	1,8
6	21	29,8	8,8	77,44	2,6
7	24	22,0	2,0	4,0	0,2
8	20	13,5	6,5	42,25	3,1
9	13	7,0	6,0	36,0	5,1
Сумма	200				$\chi^2_{набл}$ =22,2

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости a = 0,05 и числу степеней свободы k = s – 3 = 9 – 3 = 6 находим критическую точку правосторонней критической области $\chi^2_{кр}$ (0,05; 6) = 12,6.

Так как $\chi^2_{набл}$ =22,2 > $\chi^2_{кр}$ = 12,6, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо.

Пример 4: Представлены статистические данные.
Результаты измерений диаметров n = 200 валков после шлифовки обобщены в табл. (мм):
Таблица Частотный вариационный ряд диаметров валков

i	1	2	3	4	5	6	7	8
xi, мм	6,68	6,69	6,7	6,71	6,72	6,73	6,74	6,75
ni	2	3	12	6	11	14	30	25

i	9	10	11	12	13	14	15	16
xi, мм	6,76	6,77	6,78	6,79	6,8	6,81	6,82	6,83
ni	27	31	14	8	5	6	5	1

Требуется:
1) составить дискретный вариационный ряд, при необходимости упорядочив его;
2) определить основные числовые характеристики ряда;
3) дать графическое представление ряда в виде полигона (гистограммы) распределения;
4) построить теоретическую кривую нормального распределения и проверить соответствие эмпирического и теоретического распределений по критерию Пирсона. При проверке статистической гипотезы о виде распределения принять уровень значимости a = 0,05

Решение: Основные числовые характеристики данного вариационного ряда найдем по определению. Средний диаметр валков равен (мм):

$x_{ср} = \frac{1}{n}\sum_{k=1}^{16} n_i \cdot x_k$ = 6,753; исправленная дисперсия (мм2): D = $\frac{1}{n-1}\sum_{k=1}^{16} n_i \cdot (x_k - x_{ср})^2$ = 0,0009166;

исправленное среднее квадратическое (стандартное) отклонение (мм):s = $\sqrt{D}$ = 0,03028.

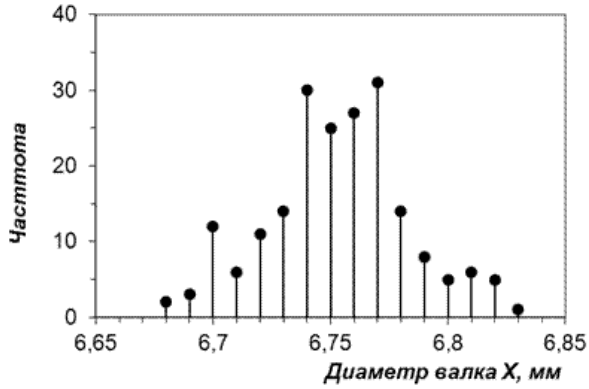


Рис. Частотное распределение диаметров валков

Исходное («сырое») частотное распределение вариационного ряда, т.е. соответствие ni(xi), отличается довольное большим разбросом значений ni относительно некоторой гипотетической «усредняющей» кривой (рис.). В этом случае предпочтительно построить и анализировать интервальный вариационный ряд, объединяя частоты для диаметров, попадающих в соответствующие интервалы.

Число интервальных групп K определим по формуле Стерджесса:

$K = 1 + \log_2 n = 1 + 3{,}322 \lg n$, где $n = 200$ – объем выборки. В нашем случае

$K = 1 + 3{,}322 \times \lg 200 = 1 + 3{,}322 \times 2{,}301 = 8{,}644 \approx 8$.

Ширина интервала равна $(6{,}83 - 6{,}68)/8 = 0{,}01875 \approx 0{,}02$ мм.

Интервальный вариационный ряд представлен в табл.

Таблица Частотный интервальный вариационный ряд диаметров валков.

k	1	2	3	4	5	6	7	8
xk, мм	6,68 – 6,70	6,70 – 6,72	6,72 – 6,74	6,74 – 6,76	6,76 – 6,78	6,78 – 6,80	6,80 – 6,82	6,82 – 6,84
nk	5	18	25	55	58	22	11	6

Интервальный ряд может быть наглядно представлен в виде гистограммы частотного распределения.

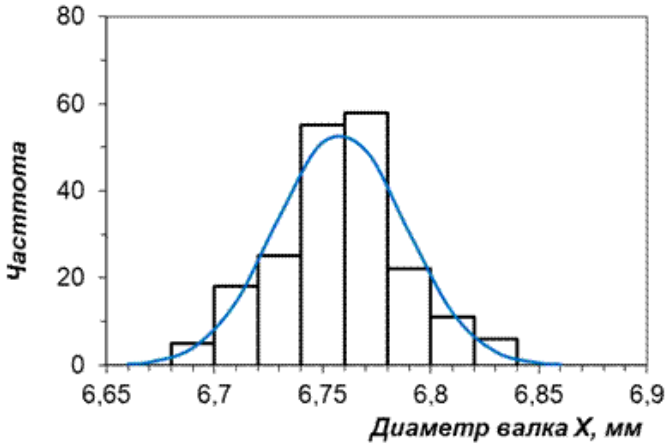


Рис. Частотное распределение диаметров валков. Сплошная линия – сглаживающая нормальная кривая.

Вид гистограммы позволяет сделать предположение о том, что распределение диаметров валков подчиняется нормальному закону, согласно которому теоретические частоты могут быть найдены как

$nk, теор = n \times N(a; s; x_k) \times D_{xk}$, где, в свою очередь, сглаживающая гауссова кривая нормального распределения определяется выражением: $N(a; s; xk)$

$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_k - a)^2}{2\sigma^2}\right\}$.

В этих выражениях xk – центры интервалов в частотном интервальном вариационном ряде.

Например, x1 = (6,68 + 6,70)/2 = 6,69. В качестве оценок центра a и параметра s гауссовой кривой можно принять: $a = x_{ср}$.

Из рис. видно, что гауссова кривая нормального распределения в целом соответствует эмпирическому интервальному распределению. Однако следует удостовериться в статистической значимости этого соответствия. Используем для проверки соответствия эмпирического распределения эмпирическому критерий согласия Пирсона. Для этого следует вычислить эмпирическое значение критерия как сумму $\chi^2 = \sum_{k=1}^{8} \frac{(n_k - n_{k,теор})^2}{n_{k,теор}}$, где n_k и $n_{k,теор}$ – эмпирические и теоретические (нормальные) частоты, соответственно. Результаты расчетов удобно представить в табличном виде:

Таблица Вычисления критерия Пирсона

$[x_k, x_{k+1})$, мм	x_k, мм	n_k	$n_{k,теор}$	$\frac{(n_k - n_{k,теор})^2}{n_{k,теор}}$
6,68 – 6,70	6,69	5	4,00	0,25
6,70 – 6,72	6,71	18	14,57	0,81
6,72 – 6,74	6,73	25	34,09	2,42
6,74 – 6,76	6,75	55	51,15	0,29
6,76 – 6,78	6,77	58	49,26	1,55
6,78 – 6,80	6,79	22	30,44	2,34
6,80 – 6,82	6,81	11	12,07	0,09
6,82 – 6,84	6,83	6	3,07	2,80
			c2эмп	10,55

Критическое значение критерия найдем по таблице Пирсона [2, 3] для уровня значимости a = 0,05 и числа степеней свободы d.f. = K – 1 – r, где K = 8 – число интервалов интервального вариационного ряда; r = 2 – число параметров теоретического распределения, оцененных на основании данных выборки (в данном случае, – параметры a и s).

Таким образом, d.f. = 5. Критическое значение критерия Пирсона есть χ^2 крит(a; d.f.) = 11,1. Так как c2эмп < c2крит, заключаем, что согласие между эмпирическим и теоретическим нормальным распределением является статистическим значимым. Иными словами, теоретическое нормальное распределение удовлетворительно описывает эмпирические данные.

Пример5: Коробки с шоколадом упаковываются автоматически. По схеме собственно-случайной бесповторной выборки взято 130 из 2000 упаковок, содержащихся в партии, и получены следующие данные об их весе:

Вес упаковки (гр.)	Менее 975	975-1000	1000-1025	1025-1050	Более 1050	Всего
Число упаковок	6	38	44	34	8	130

Требуется используя критерий χ^2 – Пирсона при уровне значимости a=0,05 проверить гипотезу о том, что случайная величина X – вес упаковок – распределена по нормальному закону. Построить на одном графике гистограмму эмпирического распределения и соответствующую нормальную кривую.

Решение

$\bar{x}$ = 1012,5 s^2 = 615,3846

Примечание:

В принципе в качестве дисперсии нормального закона распределения следует взять исправленную выборочную дисперсию. Но т.к. количество наблюдений – 130 до-

статочно велико, то подойдет и «обычная» s^2.
Таким образом, теоретическое нормальное распределение имеет вид:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}$$

подставляем a = 1012,5 σ^2 = 615,3846 σ = 24,8069

$$f(x)=0{,}0160819\cdot e^{-0{,}000812[x-1012{,}5]^2}$$

Для расчета вероятностей p_i попадания случайной величины в интервал $[x_i ; x_{i+1}]$ используем функцию Лапласа:

$$p_i(x_i\le X\le x_{i+1})=\frac{1}{2}\left[\Phi\left(\frac{x_{i+1}-a}{\sigma}\right)-\Phi\left(\frac{x_i-a}{\sigma}\right)\right]$$

$$p_i(x_i\le X\le x_{i+1})\approx\frac{1}{2}\left[\Phi\left(\frac{x_{i+1}-1012{,}5}{24{,}8069}\right)-\Phi\left(\frac{x_i-1012{,}5}{24{,}8069}\right)\right]$$

в нашем случае получаем:

$$p_i(950\le X\le 975)\approx\frac{1}{2}[\Phi(-1{,}51)-\Phi(-2{,}52)]=\frac{1}{2}[0{,}9883-0{,}869]=0{,}0597$$

$$p_i(975\le X\le 1000)\approx\frac{1}{2}[\Phi(-0{,}5)-\Phi(-1{,}51)]=\frac{1}{2}[0{,}869-0{,}3829]=0{,}2431$$

$$p_i(1000\le X\le 1025)\approx\frac{1}{2}[\Phi(0{,}5)-\Phi(-0{,}5)]=\frac{1}{2}[0{,}3829+0{,}3829]=0{,}3829$$

$$p_i(1025\le X\le 1050)\approx\frac{1}{2}[\Phi(1{,}51)-\Phi(0{,}5)]=\frac{1}{2}[0{,}869-0{,}3829]=0{,}2431$$

$$p_i(1050\le X\le 1075)\approx\frac{1}{2}[\Phi(2{,}52)-\Phi(1{,}51)]=\frac{1}{2}[0{,}9883-0{,}869]=0{,}0597$$

Примечание: Такие симметричные вероятности получились из-за того, что по нашим начальным условиям выборочная средняя попала точно в середину среднего интервала выборки.
Составим таблицу:

i	Интервал [xi ; xi+1]	Эмпирические частоты ni	Вероятности pi	Теоретические частоты npi	(ni-npi)2	$\frac{(n_i-np_i)^2}{np_i}$
1	Менее 975	6	0,0597	7,761	3,101	0,3996
2	975-1000	38	0,2431	31,603	40,922	1,2949
3	1000-1025	44	0,3829	49,777	33,374	0,6705
4	1025-1050	34	0,2431	31,603	5,746	0,1818
5	Более 1050	8	0,0597	7,761	0,057	0,0073
Σ		130	0,9885	128,5		$\chi^2=2{,}55$

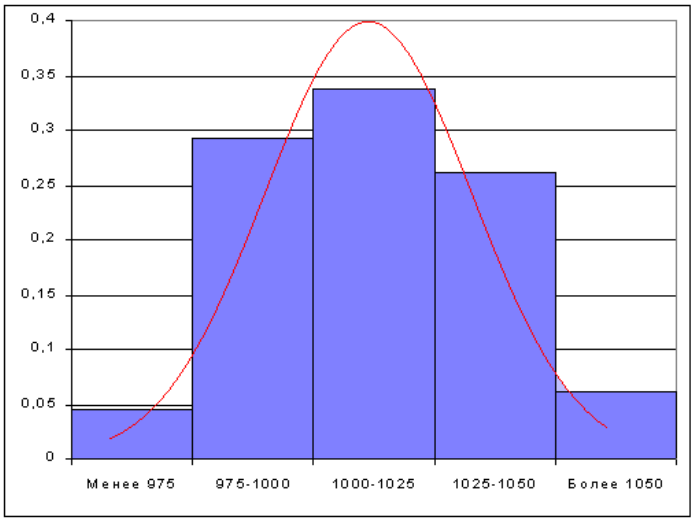
Итого, значение статистики $\chi^2=2{,}55$.
Определим количество степеней свободы по формуле: $k=m-r-1$.
m – число интервалов (m = 5), r– число параметров закона распределения (в нормальном распределении r = 2) Т.е. k = 2.

Соответствующее критическое значение статистики

$\chi^2_{0,05;2} = 5{,}99$

Поскольку $\chi^2_{0,05;2} > \chi^2$, гипотеза о нормальном распределении с параметрами N(1012,5; 615,3846) согласуется с опытными данными.

Ниже показана гистограмма эмпирического распределения и соответствующая нормальная кривая.



3. Оценка параметров неизвестного распределения. Выравнивание рядов

Пример. С целью исследования закона распределения ошибки измерения дальности с помощью радиодальномера произведено 400 измерений дальности. Результаты опытов представлены в виде статистического ряда:

$l_i\ (м)$	20; 30	30; 40	40; 50	50; 60	60; 70	70; 80	80; 90	90; 100
m_i	21	72	66	38	51	56	64	32
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080

Выровнять статистический ряд с помощью закона равномерной плотности.

Решение. Закон равномерной плотности выражается формулой

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{при } \alpha < x < \beta, \\ 0 & \text{при } x < \alpha \text{ или } x > \beta \end{cases}$$

и зависит от двух параметров α и β. Эти параметры следует выбрать так, чтобы сохранить первые два момента статистического распределения – математическое ожидание m_x^* и дисперсию D_x^*. Имеем выражения математического ожидания и дисперсии для закона равномерной плотности:

$$m_x = \frac{\alpha + \beta}{2};$$

$$D_x = \frac{(\beta - \alpha)^2}{12}.$$

Для того, чтобы упростить вычисления, связанные с определением статистических моментов, перенесем начало отсчета в точку $x_0 = 60$ и примем за представителя его разряда его середину. Ряд распределения имеет вид:

$\tilde{x}_i'$	−35	−25	−15	−5	5	15	25	35
p_i^*	0,052	0,180	0,165	0,095	0,128	0,140	0,160	0,080

где $\tilde{x}_i'$ - среднее для разряда значение ошибки радиодальномера X' при новом начале отсчета.

Приближенное значение статистического среднего ошибки X' равно:

$$m^*_{x'} = \sum_{i=1}^{k} \tilde{x}'_i p_i^* = 0,26$$

Второй статистический момент величины X' равен: $\alpha_2^* = \sum_{i=1}^{k} (\tilde{x}'_i)^2 p_i^* = 447,8$,

откуда статистическая дисперсия:

$$D^*_{x'} = \alpha_2^* - \left(m^*_{x'}\right)^2 = 447,7.$$

Переходя к прежнему началу отсчета, получим новое статистическое среднее:

$m^*_x = m^*_{x'} + 60 = 60,26$ в ту же статистическую дисперсию:

$D^*_x = D^*_{x'} = 447,7$.

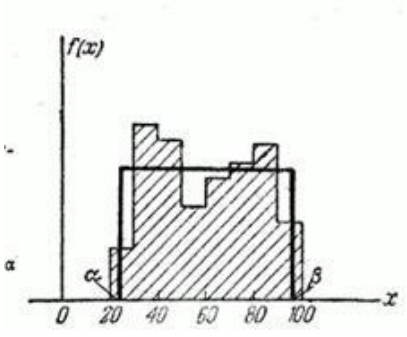


Рис. 1

Параметры закона равномерной плотности определяются уравнениями:

$$\frac{\alpha+\beta}{2} = 60,26; \quad \frac{(\beta-\alpha)^2}{12} = 447,7.$$

Решая эти уравнения относительно α и β, имеем: $\alpha \approx 23,6; \quad \beta \approx 96,9$,

откуда

$$f(x) = \frac{1}{\beta-\alpha} = \frac{1}{73,3} \approx 0,0136.$$

На рис. 1. показаны гистограмма и выравнивающий ее закон равномерной плотности $f(x)$.

2.3.3 Результаты и выводы: В результате проведенного занятия студенты:
- освоили основные понятия теории проверки статистических гипотез;
- усвоили алгоритмы применения статистических критериев;
- выработали навыки по применению критерия Пирсона;
- усвоили основные методы и алгоритмы выравнивания статистических рядов.

2.4 Практическое занятие № 7-9 (6часа)

Тема: Оценка тесноты связи. Корреляция. Дисперсионный анализ с использованием таблиц Excel. Анализ таблиц сопряженности.

2.4.1 Задание для работы:

1. Оценка тесноты связи. Корреляция.
2. Дисперсионный анализ с использованием таблиц Excel. Анализ таблиц сопряженности.

2.4.2 Краткое описание проводимого занятия:

Определение тесноты связи между результатирующим фактором и факторами, влияющими на него, а также тесноты связи между самими влияющими факторами

Таблица 11. Исходные данные

№ п/п	Урожайность естеств. сенокосов, ц/га	Влияющие факторы							
		Производств. затраты, руб./га	Основные произв. фонды, руб./га	Затраты мин. удобрений, ц/га	Энергетические мощности, л. с.	Удельный вес залес. и закуст. сенокосов, %	Удельный вес заболоченных сенокосов, %	Удельный вес улучшенных сенокосов, %	Балл оценки по совокупным свойствам почв
1	12,2	54,0	600	0,81	2,00	20,0	1,6	11,6	51
2	11,4	70,8	400	0,50	1,40	38,0	9,6	12,3	60
3	11,1	160,0	602	2,25	3,10	22,0	3,5	6,0	55
4	21,1	110,0	680	1,50	1,75	9,6	3,0	42,0	86
5	10,8	71,0	450	0,76	1,68	40,0	26,5	8,0	55
6	11,1	75,0	420	0,65	1,10	32,0	13,0	26,1	61
7	13,9	60,0	380	2,14	1,80	25,0	5,2	7,9	72
8	9,0	64,4	450	0,80	1,90	30,0	5,0	22,3	50
9	17,0	120,0	715	1,31	2,55	7,0	0,5	40,0	92
10	11,7	64,0	350	0,69	1,56	31,0	8,0	35,0	45
11	10,6	70,0	410	1,12	1,80	26,4	14,2	15,2	61
12	12,7	62,5	500	1,58	1,78	21,5	24,0	20,0	84
13	14,0	55,0	620	1,05	1,40	33,6	6,1	12,4	78
14	12,5	60,0	550	0,90	1,70	19,0	60,0	15,0	72
15	12,1	85,0	550	0,70	1,60	40,0	9,0	2,0	76
16	12,0	70,0	560	0,75	1,86	20,0	1,0	13,0	60
17	15,8	108,0	420	0,74	1,23	18,5	4,7	25,0	86
18	12,6	85,0	680	0,90	2,31	32,6	8,4	17,4	81
19	27,3	147,0	621	0,70	3,75	1,58	0,5	9,9	92
20	18,9	78,0	480	1,12	2,68	40,0	12,8	8,6	90
21	14,3	55,6	568	0,88	1,74	18,8	2,5	6,0	96
22	8,8	45,4	340	0,68	1,01	26,0	48,4	12,5	54
23	13,5	68,0	508	1,32	2,14	42,4	11,0	10,6	74

Теснота и направление парной линейной корреляционной зависимости переменных X и Y определяется коэффициентом корреляции. Он принимает значения от –1 до +1. При $|r| \geq 0{,}75$ связь тесная, фактор, оказывающий влияние на результирующий показатель достоверен. При $|r| \geq 0{,}25$ связь практически отсутствует и рассматриваемый фактор следует исключить.

Связь между результирующим и влияющими факторами отражается уравнением множественной линейной регрессии:

$$Y=A_o + A_1X_1 + A_2X_2 +\ldots+ A_nX_n ,$$

где A_o – свободный член уравнения, экономической интерпретации не имеет;

$A_1,A_2,\ldots,A_n$ – коэффициенты уравнения, показывающие на сколько изменится результирующий фактор при изменении влияющего на единицу;

$X_1, X_2,\ldots,X_n$ – значения влияющих факторов.

В результате решения задачи с помощью “Regma” были получены следующие коэффициенты уравнения множественной линейной регрессии:

A[0]=	3.3854	A[3]=	-1.7198	A[6]=	-0.0252
A[1]=	0.0101	A[4]=	2.9394	A[7]=	0.0501
A[2]=	-0.0076	A[5]=	-0.0764	A[8]=	0.1559

Приведенное значение среднего квадратического отклонения фактических значений результирующего показателя от его вычисленных значений = 0.1376.

Коэффициент множественной корреляции = 0.89.

Коэффициент детерминации = 0.79.

Первоначально при расчете используются все факторы, которые могут влиять. В полученных результатах отражается теснота связи между результирующим фактором и факторами, влияющими на него (I матрица результатов), а также связь между самими влияющими факторами (II матрица результатов).

Таблица 12

Характеристики рядов исходной матрицы (I)

Ряд	среднее	Среднее квадратич. отклонение	энтропия	эластичность	Коэф. вариации	Бета-коэф.
1	13,67	4,07	1,41	3,39	0,30	3,39
2	79,94	29,09	2,39	0,06	0,36	0,07
3	515,39	107,77	3,05	-0,29	0,21	-0,20
4	1,04	0,45	0,31	-0,13	0,44	-0,19
5	1,91	0,62	0,47	0,41	0,33	0,45
6	25,87	10,78	1,90	0,14	0,42	-0,20
7	12,11	14,68	2,05	-0,02	1,21	-0,09
8	16,47	10,56	1,89	0,06	0,64	0,13
9	70,91	15,37	2,07	0,81	0,22	0,59

Таблица 13 **Характеристики рядов исходной матрицы (II)**

Ряд	Макс. значение	Мин. значение	энтропия
1	27,30	8,80	4,21
2	160,00	45,40	6,84
3	715,00	340,00	8,55
4	2,25	0,50	0,81
5	3,75	1,01	1,45
6	42,40	1,58	5,35
7	60,00	0,50	5,89
8	42,00	2,00	5,32
9	96,00	45,00	5,67

Таблица 14. Таблица парных коэффициентов корреляции

пара	Коэф. корреляции	Оценка существ.	энтропия
1-2	0,5627	3,1928	19,9089
1-3	0,4762	2,5400	16,6867
1-4	0,0935	0,4407	7,9087
1-5	0,6006	3,5230	8,4706
1-6	-0,5608	-3,1774	12,2834
1-7	-0,3411	-1,7018	11,3714
1-8	0,1771	0,8439	11,8814
1-9	0,7180	4,8378	13,4880
2-3	0,4725	2,5148	19,2380
2-4	0,3262	1,6187	10,3819

2-5	0,6947	4,5305	10,8659
2-6	-0,4871	-2,6162	14,9084
2-7	-0,3975	-2,0319	13,8846
2-8	0,1661	0,7900	14,4323
2-9	0,3056	1,5056	16,4879
3-4	0,2068	0,9917	13,1201
3-5	0,5333	2,9570	13,7885
3-6	-0,4547	-2,3948	17,6253
3-7	-0,3327	-1,6546	16,6127
3-8	0,1326	0,6277	17,1282
3-9	0,5129	2,8400	19,0220
4-5	0,3471	1,7361	4,9801
4-6	-0,1836	-0,8759	8,8106
4-7	-0,1560	-0,7407	7,7223
4-8	-0,0148	-0,0694	8,1837
4-9	0,1656	0,7875	10,2701
5-6	-0,3767	-1,9075	9,6031
5-7	-0,3500	-1,7527	8,5241
5-8	-0,1596	-0,7585	-,0435
5-9	0,3196	1,5821	11,0907
6-7	0,1558	0,7399	12,3632
6-8	-0,3928	-2,0037	12,7037
6-9	-0,3666	-1,8484	14,8268
7-8	-0,1351	-0,6395	11,7162
7-9	-0,1905	-0,9100	13,8091
8-9	0,0661	0,3107	14,2763

В I матрице отбраковываются факторы, не влияющие или мало влияющие на результирующий ($|r| \le 0{,}25$), а во II матрице исключается мультикоррелярность, означающая, что факторы являются результатом друг друга ($|r| \ge 0{,}75$). Для исключения одного из двух влияющих факторов необходимо определить, какой из них имеет меньшую тесноту связи с результирующим (рассматривается матрица I).

В I матрице исключаются 4 и 8 факторы (т. к. 1 фактором является урожайность, следовательно, исключаются X_3 и X_7). Во второй исключать ничего не пришлось. После исключения малозначащих и мультикорреляционных факторов снова производится обработка исходной числовой матрицы.

A[0]= 4.4290
A[1]= 0.0114
A[2]= -0.0069
A[4]= 2.1302
A[5]= -0.0967
A[6]= -0.0297
A[8]= 0.1508

Приведенное значение среднего квадратического отклонения фактических значений результирующего показателя от его вычисленных значений = 0.1508

Коэффициент множественной корреляции = 0.86

Коэффициент детерминации = 0.74

Таблица 15. Характеристики рядов исходной матрицы (I)

Ряд	среднее	Среднее квадратич. отклонение	энтропия	эластичность	Коэф. вариации	Бета-коэф.
1	13,67	4,07	1,41	4,43	0,30	4,43
2	79,94	29,09	2,39	0,07	0,36	0,08
3	515,39	107,77	3,05	-0,26	0,21	-0,18
5	1,91	0,62	0,47	0,30	0,33	0,33
6	25,87	10,78	1,90	-0,18	0,42	-0,26
7	12,11	14,68	2,05	-0,03	1,21	-0,11
9	70,91	15,37	2,07	0,78	0,22	0,57

Таблица 16. Таблица парных коэффициентов корреляции

пара	Коэф. корреляции	Оценка существ.	энтропия
1-2	0,5627	3,1928	19,9089
1-3	0,4762	2,5400	16,6867
1-5	0,6006	3,5230	8,4706
1-6	-0,5608	-3,1774	12,2834
1-7	-0,3411	-1,7018	11,3714
1-9	0,7180	4,8378	13,4880
2-3	0,4725	2,5148	19,2380
2-5	0,6947	4,5305	10,8659
2-6	-0,4871	-2,6162	14,9084
2-7	-0,3975	-2,0319	13,8846
2-9	0,3056	1,5056	16,4879
3-5	0,5333	2,9570	13,7885
3-6	-0,4547	-2,3948	17,6253
3-7	-0,3327	-1,6546	16,6127
3-9	0,5129	2,8400	19,0220
5-6	-0,3767	-1,9075	9,6031
5-7	-0,3500	-1,7527	8,5241
5-9	0,3196	1,5821	11,0907
6-7	0,1558	0,7399	12,3632
6-9	-0,3666	-1,8484	14,8268
7-9	-0,1905	-0,9100	13,8091

Графическое отображение связи между результирующим фактором и фактором, в наибольшей степени на него влияющим

По результатам повторной обработки исходной числовой матрицы определяется фактор, имеющий наибольшее влияние на результирующий (величина коэффициента корреляции близка к 1). В данной матрице это 9 фактор. Пакет программных средств “Coreg” позволяет наглядно отразить связь между результирующим фактором и фактором в наибольшей степени на него влияющим.

График зависимости между результирующим фактором и фактором, в наибольшей степени на него влияющим

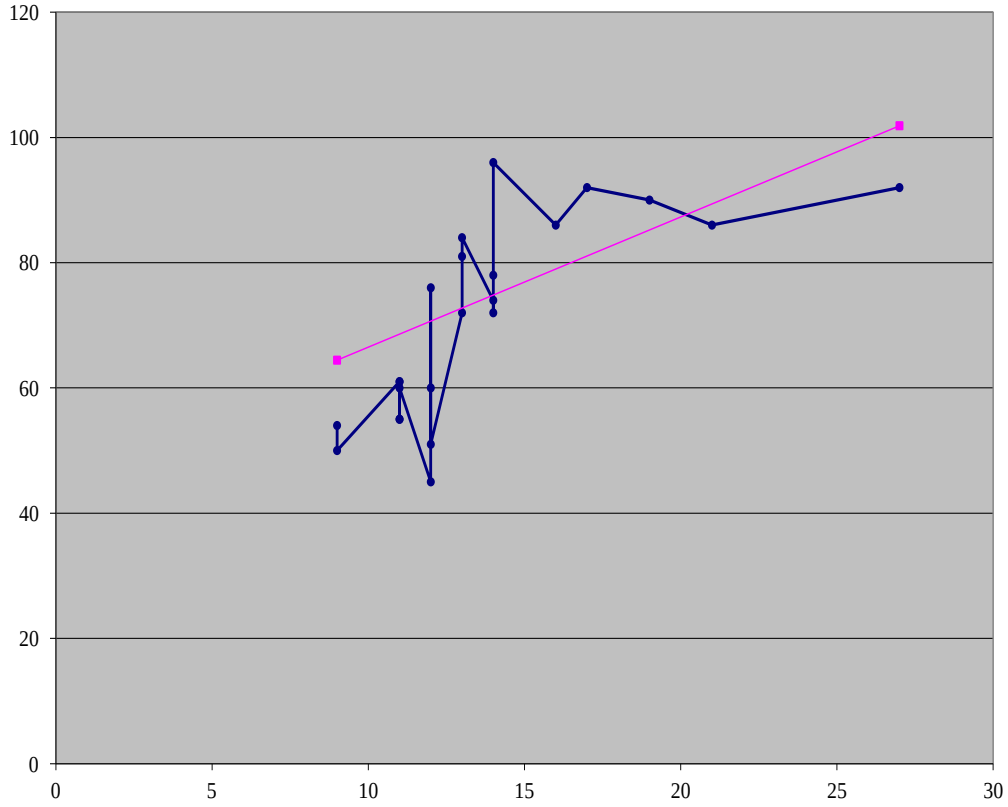


Таблица 6.1. Расчет исходных данных для проверки нормальности распределения вариационного ряда урожайности ячменя по затратам удобрений (X)

№№ по порядку	Урожайность ячменя с 1 га, У	Затраты удобрений на 1 га посевов ц. д. в., X	Урожайность в расчете на 1 ц удобрений, ц с 1 га, $У^t$	$(У-Ў)^2$	$(У^t-Ў^t)$	$(У^t-Ў^t)^2$	$(У^t-Ў^t)^3$	$(У^t-Ў^t)^4$	$t=\frac{(Y^t-\overline{Y^t})}{\sigma_{ост факторов}}$	Ордината нормальной кривой F(t)
1	25,2	3,34	7,54	58,68	-0,35	0,12	-0,04	0,01	-0,35	0,3614
2	23,0	2,89	7,96	29,81	0,07	0,00	0,00	0,00	0,07	0,3975
3	21,5	2,79	7,71	15,68	-0,19	0,03	-0,01	0,00	-0,19	0,3878
4	20,4	2,49	8,19	8,18	0,30	0,09	0,03	0,01	0,30	0,3702
5	20,9	2,39	8,74	11,29	0,85	0,73	0,62	0,53	0,85	0,2190
6	19,4	2,35	8,26	3,46	0,36	0,13	0,05	0,02	0,36	0,3577
7	19,4	2,35	8,26	3,46	0,36	0,13	0,05	0,02	0,36	0,3577
8	18,6	2,18	8,53	1,12	0,64	0,41	0,26	0,17	0,64	0,2845
9	18,2	2,18	8,35	0,44	0,46	0,21	0,10	0,04	0,46	0,3358
10	17,4	2,04	8,53	0,02	0,64	0,41	0,26	0,17	0,64	0,2853
11	17,4	2,04	8,53	0,02	0,64	0,41	0,26	0,17	0,64	0,2853
12	16,8	2,00	8,40	0,55	0,51	0,26	0,13	0,07	0,51	0,3224
13	16,8	2,00	8,40	0,55	0,51	0,26	0,13	0,07	0,51	0,3224
14	16,0	1,93	8,29	2,37	0,40	0,16	0,06	0,03	0,40	0,3500
15	15,1	1,93	7,82	5,95	-0,07	0,00	0,00	0,00	-0,07	0,3974
16	15,1	1,93	7,82	5,95	-0,07	0,00	0,00	0,00	-0,07	0,3974
17	14,1	1,90	7,42	11,83	-0,47	0,22	-0,10	0,05	-0,47	0,3325
18	13,0	1,90	6,84	20,61	-1,05	1,10	-1,16	1,21	-1,05	0,1611
19	12,2	1,84	6,63	28,52	-1,26	1,59	-2,00	2,53	-1,26	0,1077
20	10,3	1,84	5,60	52,42	-2,29	5,26	-12,06	27,67	-2,29	0,0052
сумма	350,8	44,31	157,83	260,91	0,00	11,53	-13,43	32,75	x	x

$\sigma_{ост. факторов} = 0,78$; $\sigma^2_{ост. факторов} = 0,61$; $\sigma^2_{общ} = 13,73$; $\sigma^2_{уд} = 13,12$. $A_s = -1,419$; $E_x = 1,443$.

Распределение урожайности ячменя по затратам удобрений

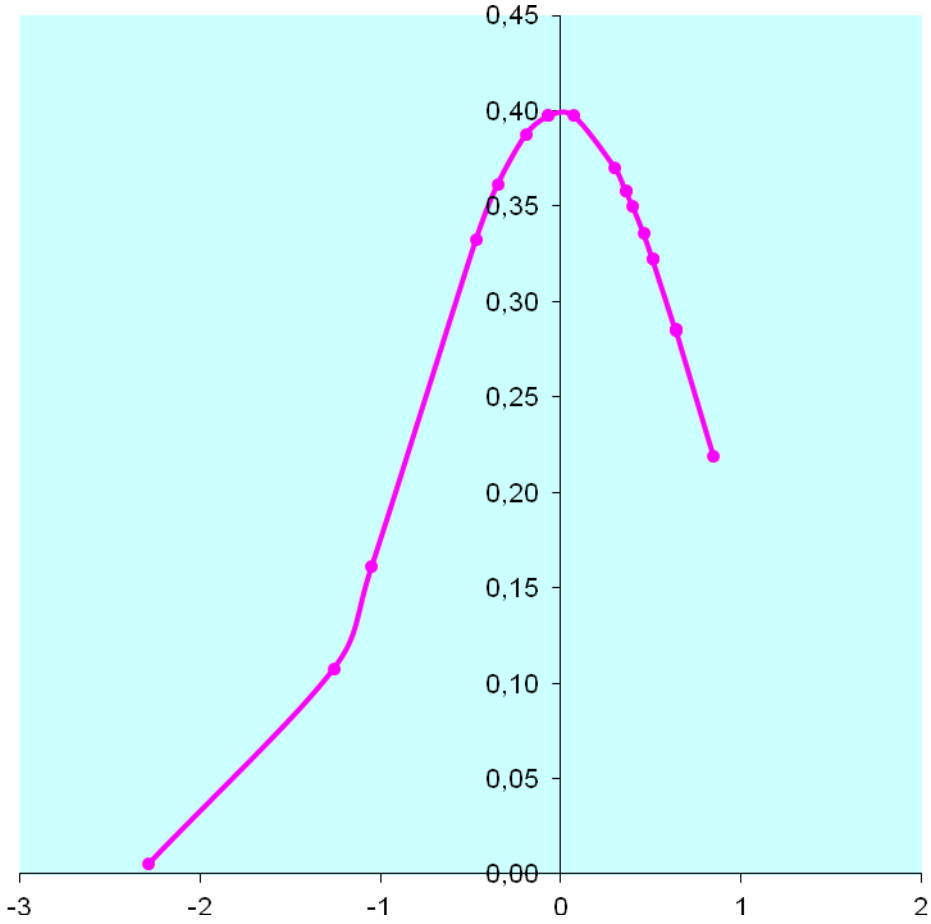


Таблица 6.2.Расчет исходных данных для проверки нормальности распределения вариационного ряда урожайности овса по затратам удобрений (Х)

№№ по порядку	Урожайность овса с 1 га, У	Затраты удобрений на 1 га посевов ц. д. в., Х	Урожайность в расчете на 1 ц удобрений, ц с 1 га, $У^t$	$(У-\breve{У})^2$	$(У^t-\breve{У}^t)$	$(У^t-\breve{У}^t)^2$	$(У^t-\breve{У}^t)^3$	$(У^t-\breve{У}^t)^4$	$t=\frac{(Y^t-\overline{Y^t})}{\sigma_{\text{ост факторов}}}$	Ордината нормальной кривой F(t)
	24,2	3,34	7,25	21,67	-1,61	2,60	-4,19	6,76	-1,61	0,1214
	24,2	2,89	8,37	21,67	-0,48	0,23	-0,11	0,06	-0,48	0,3583
	23,3	2,79	8,35	14,10	-0,51	0,26	-0,13	0,07	-0,51	0,3547
	22,2	2,49	8,92	7,05	0,06	0,00	0,00	0,00	0,06	0,3983
	22,2	2,39	9,29	7,05	0,43	0,19	0,08	0,03	0,43	0,3665
	21,7	2,35	9,23	4,64	0,38	0,14	0,05	0,02	0,38	0,3740
	21,2	2,35	9,02	2,74	0,16	0,03	0,00	0,00	0,16	0,3941
	20,8	2,18	9,54	1,58	0,68	0,47	0,32	0,22	0,68	0,3222
	20,8	2,18	9,54	1,58	0,68	0,47	0,32	0,22	0,68	0,3222
0	20,1	2,04	9,85	0,31	0,99	0,99	0,98	0,98	0,99	0,2537
1	20,1	2,04	9,85	0,31	0,99	0,99	0,98	0,98	0,99	0,2537
2	19,4	2,00	9,70	0,02	0,84	0,71	0,60	0,50	0,84	0,2885
3	19,4	2,00	9,70	0,02	0,84	0,71	0,60	0,50	0,84	0,2885
4	18,7	1,93	9,69	0,71	0,83	0,69	0,57	0,48	0,83	0,2909
5	18,2	1,93	9,43	1,81	0,57	0,33	0,19	0,11	0,57	0,3435
6	18,6	1,93	9,64	0,89	0,78	0,61	0,47	0,37	0,78	0,3022

7	16,2	1,90	8,53	11,19	-0,33	0,11	-0,04	0,01	-0,33	0,3793
8	15,2	1,90	8,00	18,88	-0,86	0,74	-0,63	0,54	-0,86	0,2848
9	13,2	1,84	7,17	40,26	-1,68	2,84	-4,78	8,05	-1,68	0,1090
0	11,2	1,84	6,09	69,64	-2,77	7,68	-21,28	58,97	-2,77	0,0119
умма	390,9	44,31	177,16	226,11	0,00	20,77	-25,99	78,86	x	x

$\sigma_{ост.\ факторов} = 1,05$; $\sigma^2_{ост.\ факторов} = 1,09$; $\sigma^2_{общ} = 11,90$; $\sigma^2_{уд} = 10,81$. $A_s = -1,137$; $E_x = 0,301$.

2.4.3 Результаты и выводы:

1. Любому производству присуща в некотором смысле противоречивая особенность - с помощью одних и тех же производственных ресурсов на равноценных по плодородию землях достигать различного уровня эффекта. Это требует многовариантного подхода к изучению способов организации производства. В новых экономических условиях хозяйствования, все более усугубляющейся проблемы оптимального распределения ограниченных ресурсов, обеспечения конкурентоспособности предприятия, многовариантность и системность становятся основными принципами в экономических исследованиях и организации производства.

Умение математически сформулировать экономическую задачу, решить и провести анализ, корректировку оптимального плана, обосновать управленческие решения на всех уровнях хозяйственной иерархии, становятся обязательным в рыночной экономике.

2. Большинство явлений и процессов в экономике находятся в постоянной взаимной и всеохватывающей объективной связи. Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами играет большую роль в экономике. Оно дает возможность глубже понять сложный механизм причинно-следственных отношений между явлениями. Для исследования интенсивности, вида и формы зависимостей широко применяется корреляционно-регрессионный анализ, который является методическим инструментарием при решении задач прогнозирования, планирования и анализа хозяйственной деятельности предприятий.

3. За последние десять лет, наблюдается рост производительности труда, но пока он меньше роста зарплаты.

4. Фондообеспеченность и фондовооруженность постепенно, но неуклонно возрастает, что объясняется ростом стоимости ОПФ, увеличением площади сельскохозяйственных угодий и снижением численности работников.

5. В результате проведенного корреляционно-регрессионного анализа, связи между производительностью и фондообеспеченностью на 100 га площади сельскохозяйственных угодий, фондовооруженностью на одного работника и урожайностью выявили, что критерий Фишера равен 5,12,это означает что во всех моделях построенные уравнения регрессии не значимы и надежны.

6. Степень точности описания моделью процесса R - квадрат (коэффициент детерминаций) равен 0,6972 можно говорить о средней точности аппроксимации, то есть модель хорошо описывает явление, и введения новых независимых переменных в модель не требуется.

7. Коэффициент детерминации показывает, что на производительность влияют фондообеспеченность, фондовооруженность и урожайность на 69,72%.

2.5 Практическое занятие № 10-11 (4часа)

Тема: Экспертные оценки в прикладных исследованиях. Ранговый коэффициент корреляции. Коэффициент конкордации для оценки согласия экспертов. Метод парных сравнений в условиях иерархии.

2.5.1 Задание для работы:

1. Конкордация
2. Ранговая корреляция.
3. Теория и практика экспертных оценок

2.5.2 Краткое описание проводимого занятия:

Существует множество процессов и явлений, количественная информация для характеристики которых отсутствует или очень быстро изменяется.

В этом случае используются методы экспертных оценок, сущность которых заключается в том, что в основу прогноза закладывается мнение специалиста, основанное на профессиональном, научном и практическом опыте.

Метод экспертных оценок применяется для сравнения каких-то параметров объектов (напр., комфортность самолета, сравнение автомобилей и др.), находящихся в одном "классе", одинаковой категории, и относится к разновидности мозгового штурма.

Рассмотрим пример метода экспертных оценок:

1. Сравниваем самолеты Боинг, Туполев, аэробус, бомбардир, альбатрос .

2. Выбираем параметры сравнения.

Параметров желательно выбирать не менее 4 и не более 7, т.к. большее количество параметров влечет расфокусировку и отсутствие четкого понимания результата. То же самое и с количеством сравниваемых объектов - от 4 до 7.

№	Параметр	Вес	А (Боинг)	Б (Туполев)	В (Аэробус)	Г (Бомбардир)	Д (Альбатрос)	Е
1	Магистральная дальность							
2	Вместимость							
3	Расход топлива V							
4	Комфорт							
5	Стоимость эксплуатации							
Сумма		1						

3. Далее распределяем "Вес" между параметрами таким образом, чтобы в сумме он был равен 1. Наиболее приоритетным параметрам мы выделяем большее значение. причем каждый параметр варьируется в диапазоне от 0,015 до 0.3.

№	Параметр	Вес	А (Боинг)	Б (Туполев)	В (Аэробус)	Г (Бомбардир)	Д (Альбатрос)	Е
1	Магистральная дальность	0,2						
2	Вместимость	0,15						
3	Расход топлива V	0,25						
4	Комфорт	0,15						
5	Стоимость эксплуатации	0,25						
Сумма		1						

4. Задаем описательную сравнительную часть для объектов (самолетов):

Объекты сравнения по параметру "Максимальная дальность" сравниваются по 10-тибальной шкале. Максимальное значение 10 баллов присваивается, если (возможные варианты):

а) самолет пролетает без дозаправки более 5 000 км

б) самолет с дозаправкой пролетает более 10 000 км

в) если данный класс самолета имеет максимальную дальность беспосадочного перелета в своем классе.

При уменьшении расстояния перелета по данным параметрам на каждые 500 км снимается 1 балл.

Параметр "Стоимость эксплуатации" (в нашем случае это самый важный параметр, т.к. ему присвоен максимальный вес):
Самолету присваивается 10 баллов, если (возможные варианты):
а) стоимость эксплуатации самолета за квартал не более 100 000 долларов
б) стоимость эксплуатации не превышает 10 % от номинальной стоимости самолета
И так далее по всем параметрам.
В итоге наша таблица выглядит следующим образом:

№	Параметр	Вес	А (Боинг)	Б (Туполев)	В (Аэробус)	Г (Бомбардир)	Д (Альбатрос)	Е
1	Магистральная дальность	0,2	9	6	10	8	7	
2	Вместимость	0,15	7	10	10	6	8	
3	Расход топлива V	0,25	8	7	9	9	8	
4	Максимальная скорость	0,15	10	8	9	8	7	
5	Стоимость эксплуатации	0,25	6	8	9	10	8	
Сумма		1						

Далее необходимо наши баллы умножить на вес данного параметра. В последний столбец "Е" ставится максимальное значение получившихся чисел. В строке "Сумма" складываем сумму "весов" параметров для каждого самолета.

№	Параметр	Вес	А (Боинг)	Б (Туполев)	В (Аэробус)	Г (Бомбардир)	Д (Альбатрос)	Е
1	Магистральная дальность	0,2	9*0,2 = 1,8	6 *0,2 = 1,2	10*0,2 = 2	8*0,2 = 1,6	7*0,2 = 1,4	2
2	Вместимость	0,15	7*0,15 = 1,1	10*0,15 = 1,5	10*0,15 = 1,5	6*0,15 = 0,9	8*0,15 = 1,2	1,5
3	Расход топлива V	0,25	8*0,25 = 2	7*0,25 = 1,75	9*0,25 = 2,25	9*0,25 = 2,25	8*0,25 = 2	2,25
4	Максимальная скорость	0,15	10*0,15 = 1,5	8*0,15 = 1,2	9*0,15 = 1,35	8*0,15 = 1,2	7*0,15 = 1,1	1,5
5	Стоимость эксплуатации	0,25	6*0,25 = 1,5	8*0,25 = 2	9*0,25 = 2,25	10*0,25= **2,5**	8*0,25 = 2	2,5
Сумма		**1**	**7,9**	**7,65**	**9,35**	**8,46**	**7,7**	

5. Таким образом, самолет "В" оказался самым эффективным. Далее мы уже анализируем, надо ли нам покупать самолет В, оказавшийся самым эффективным, либо , например, склониться в сторону самолета Г с самыми маленькими эксплуатационными расходами, так как именно этот параметр ключевой.

Метод экспертных оценок

Заполняем следующую таблицу:

№	Параметр	Вес	А	Б	В	Г	Д	Е
1								
2								
3								
4								
5								
Сумма		1						

Шаги заполнения таблицы:

1. Выбираем объект для экспертной оценки (А, Б, В, Г, Д - самолеты);
2. Выбираем параметры для сравнения (от 3 до 7 параметров);
3. Определяем вес каждого параметра (В сумме равен 1. Самому важному параметру задаем максимальный вес);
4. Задаем сравнительную шкалу для сравниваемых объектов (по 10-тибальной шкале);
5. Сравниваем.

Пример (сравниваем самолеты):

1. Сравниваемы объекты - самолеты.
2. Описываем параметры для сравнения.
3. Определяем вес каждого параметра. В нашем случае самый важный параметр "Стоимость эксплуатации" имеет максимальный вес.

4. Оцениваем по 10-тибальной шкале каждый параметр самолетов.

№	Параметр	Вес	А (Боинг)	Б (Туполев)	В (Аэробус)	Г (Бомбар-дир)	Д (Альбатрос)	Е
1	Магистральная дальность	0,2	9	6	10	8	7	
2	Вместимость	0,15	7	10	10	6	8	
3	Расход топлива V	0,25	8	7	9	9	8	
4	Максимальная скорость	0,15	10	8	9	8	7	
5	Стоимость эксплуатации	0,25	6	8	9	10	8	
Сумма		1						

Умножаем наши значения параметров на вес.

№	Параметр	Вес	А (Боинг)	Б (Туполев)	В (Аэробус)	Г (Бомбар-дир)	Д (Альбатрос)	Е
1	Магистральная дальность	0,2	9*0,2 = **1,8**	6 *0,2 = **1,2**	10*0,2 = **2**	8*0,2 = **1,6**	7*0,2 = **1,4**	2
2	Вместимость	0,15	7*0,15 = **1,1**	10*0,15 = **1,5**	10*0,15 = **1,5**	6*0,15 = **0,9**	8*0,15 = **1,2**	1,5
3	Расход топлива V	0,25	8*0,25 = **2**	7*0,25 = **1,75**	9*0,25 = **2,25**	9*0,25 = **2,25**	8*0,25 = **2**	2,25
4	Максимальная скорость	0,15	10*0,15 = **1,5**	8*0,15 = **1,2**	9*0,15 = **1,35**	8*0,15 = **1,2**	7*0,15 = **1,1**	1,5
5	Стоимость эксплуатации	0,25	6*0,25 = **1,5**	8*0,25 = **2**	9*0,25 = **2,25**	10*0,25= **2,5**	8*0,25 = **2**	2,5
Сумма		1	**7,9**	**7,65**	**9,35**	**8,46**	**7,7**	

5. Сравниваем показатели.

В нашем случае самым эффективным оказался самолет В (Аэробус), но лучшие показатели по стоимости эксплуатации имеет самолет Г (Бомбардир).

Коэффициент конкордации Кендалла

Коэффициент конкордации Кендалла или по-другому Коэффициент множественной ранговой корреляции нужен для того, чтобы выявить согласованность мнений экспертов по нескольким факторам.

Например, провели вы исследование, в котором попросили 5 человек проранжировать по важности 4 разных фактора. Они вам расставили ранги от 1 до 4 и вам теперь надо это анализировать. Вот тут то и может понадобиться **Коэффициент конкордации Кендалла.**

Способ расчета Коэфициента конкордации Кендалла.

$$W = \frac{12S}{m^2(n^3 - n)}$$ (1), где *m* - число экспертов в группе, *n* - число факторов, *S* - сумма квадратов разностей рангов (отклонений от среднего).

Разберем на примере, в котором 5 экспертов оценили 4 фактора.

	A	B	C	D	E
1		фак 1	фак 2	фак 3	фак 4
2	экс 1	1	3	2	4
3	экс 2	3	2	1	4
4	экс 3	4	3	1	2
5	экс 4	2	3	4	1
6	экс 5	2	4	1	3

m = 5, n = 4

Остается найти сумма квадратов разностей рангов (S).

Её можно найти по любой из следующих формул:

$$S = \sum_{i=1}^{n} (\sum_{j=1}^{m} R_{ij})^2 - \frac{(\sum_{i=1}^{n} \sum_{j=1}^{m} R_{ij})^2}{n}$$ **(2)**

$$S = \sum_{j-1}^{n} (\sum_{i-1}^{m} Aij - \frac{1}{2} m(n+1))^2$$ **(3)**

Для вычисления нам нужно добавить пару новых строк в исходную таблицу (Сумма по столбцу и Квадрат этой суммы).

	A	B	C	D	E	F
1		фак 1	фак 2	фак 3	фак 4	
2	экс 1	1	3	2	4	
3	экс 2	3	2	1	4	
4	экс 3	4	3	1	2	
5	экс 4	2	3	4	1	
6	экс 5	2	4	1	3	
7	Сумма	12	15	9	14	50
8	Квадрат с	144	225	81	196	646

По формуле (2) получаем:

S = 646 - 50 ^{2} / 4 = 21

По формуле (3) получаем:

S = (12 - 12,5) ^{2} + (15 - 12,5) ^{2} + (9- 12,5) ^{2} + (14 - 12,5) ^{2} = 21

Теперь считаем, собственно говоря, сам коэфициента конкордации Кендалла по формуле (1):

W = (12 * 21) / (25 * (64 - 4)) = 0,168

.

Если W < 0.2 - 0.4, значит слабая согласованность экспертов, если W > 0.6 - 0.8, то согласованность экспертов сильная.

Слабая согласованность обычно является следствием следующих причин:

- в рассматриваемой группе экспертов действительно отсутствует общность мнений;
- внутри группы существуют коалиции с высокой согласованностью мнений, однако, обобщенные мнения коалиций противоположны.

РАНГОВАЯ КОРРЕЛЯЦИЯ

Метод линейной корреляции применяется для определения меры соответствия двух признаков, выраженных количественно, т.е. для числовых величин. Это параметрический метод, который требует соответствия распределения данного исследуемого признака закону нормального распределения. В отличие от этого метода, метод ранговой корреляции (корреляция Спирмена) применим к любым количественно измеренным или ранжированным данным. Этот метод способен, в отличие от других, измерять согласованность изменения разных признаков у одного испытуемого или выявлять совпадения индивидуальных ранговых показателей у двух испытуемых; или у испытуемого и усредненный показатель некой группы; или какие-либо показатели в сравнении двух групп. Мощность коэффициента ранговой корреляции Спирмена несколько уступает мощности параметрического коэффициента корреляции.

Коэффициент ранговой корреляции Спирмена - это непараметрический метод, который используется с целью статистического изучения связи между явлениями.

Данный метод может быть использован не только для количественно выраженных данных, но также и в случаях, когда регистрируемые значения определяются описательными признаками различной интенсивности. Для подсчета ранговой корреляции необходимо располагать двумя рядами значений, которые могут быть проранжированы.

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента равные 0,3 и менее, показателями слабой тесноты связи; значения более 0,4, но менее 0,7 - показателями умеренной тесноты связи, а значения 0,7 и более - показателями высокой тесноты связи.

Ограничения метода ранговой корреляции. По каждой переменной должно быть представлено не менее 5 наблюдений. Верхняя граница выборки – меньше или равна 40. Помимо этих ограничений, следует так же помнить об ограничениях корреляционного метода вообще – невозможность обнаружения причинной связи между явлениями.

Коэффициент ранговой корреляции Спирмена.

Присвоим ранги признаку Y и фактору X. Найдем сумму разности квадратов d^2.

По формуле вычислим коэффициент ранговой корреляции Спирмена.

$$p = 1 - 6\frac{\sum d^2}{n^3-n}$$

X	Y	ранг X, d_x	ранг Y, d_y	$(d_x - d_y)^2$
8	13	4	5	1
3	11	1	3	4
12	2	5	2	9
6	12	3	4	1
4	1	2	1	1
0	0	0	0	16

$$p = 1 - 6\frac{16}{5^3 - 5} = 0.2$$

Связь между признаком Y фактором X слабая и прямая

Оценка коэффициента ранговой корреляции Спирмена.

Значимость коэффициента ранговой корреляции Спирмена

$$T_{набл} = p_{xy}\frac{\sqrt{n-2}}{\sqrt{1 - p^2_{xy}}} = 0.2\frac{\sqrt{3}}{\sqrt{1 - 0.2^2}} = 0.35$$

По таблице Стьюдента находим $t_{табл}$:

$t_{табл}$ (n-m-1;α/2) = (3;0.05/2) = 3.182

Поскольку Тнабл < ttабл , то принимаем гипотезу о равенстве 0 коэффициента ранговой корреляции. Другими словами, коэффициент ранговой корреляции статистически - не значим.

Интервальная оценка для коэффициента корреляции (доверительный интервал).

$$(p - t_{табл}\frac{1-p^2}{\sqrt{n}}; p + t_{табл}\frac{1-p^2}{\sqrt{n}})$$

Доверительный интервал для коэффициента ранговой корреляции r(-1.1661;1.5661)

ПРИМЕР. Зависимость между объемом промышленной продукции и инвестициями в основной капитал по 10 областям одного из федеральных округов РФ в 2003 году характеризуется следующими данными:

Область	Объем промышленной продукции, млрд руб.	Инвестиции в основной капитал, млрд руб.
Белгородская	64,6	10,22
Брянская	21,5	4,12
Владимирская	51,1	8,58
Воронежская	54,4	14,79
Ивановская	20,6	2,88
Калужская	35,7	7,24
Костромская	18,4	5,57
Курская	37,1	9,67
Липецкая	90,6	10,45
Смоленская	39,8	10,48

Вычислите ранговые коэффициенты корреляции Спирмена и Кендэла. Проверить их значимость при α=0,05. Сформулируйте вывод о зависимости между объемом промышленной продукции и инвестициями в основной капитал по рассматриваемым областям РФ.

Решение. Присвоим ранги признаку Y и фактору X.
Расположим объекты так, чтобы их ранги по X представили натуральный ряд. Так как оценки, приписываемые каждой паре этого ряда, положительные, значения «+1», входя-

щие в P, будут порождаться только теми парами, ранги которых по Y образуют прямой порядок.
Их легко подсчитать, сопоставляя последовательно ранги каждого объекта в ряду Y с стальными.
Упорядочим данные по X.
В ряду Y справа от 3 расположено 7 рангов, превосходящих 3, следовательно, 3 породит в P слагаемое 7.
Справа от 1 стоят 8 ранга, превосходящих 1 (это 2, 4, 6, 9, 5, 10, 7, 8), т.е. в P войдет 8 и т.д. В итоге P = 37 и с использованием формул имеем:

X	Y	ранг X, d_x	ранг Y, d_y	P	Q
18.4	5.57	1	3	7	2
20.6	2.88	2	1	8	0
21.5	4.12	3	2	7	0
35.7	7.24	4	4	6	0
37.1	9.67	5	6	4	1
39.8	10.48	6	9	1	3
51.1	8.58	7	5	3	0
54.4	14.79	8	10	0	2
64.6	10.22	9	7	1	0
90.6	10.45	10	8	0	0
				37	8

$$\tau = \frac{37-8}{\frac{1}{2}10(10-1)} = 0.64$$

По упрощенным формулам:

$$\tau = 1 - \frac{4 \cdot 8}{10(10-1)} = 0.64$$

$$\tau = \frac{4 \cdot 37}{10(10-1)} - 1 = 0.64$$

Для того чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента ранговой корреляции Кендалла при конкурирующей гипотезе H_1: $\tau \neq 0$,надо вычислить критическую точку:

$$T_{kp} = z_{kp}\sqrt{\frac{2(2n+5)}{9n(n-1)}}$$

где n - объем выборки; z_{kp} - критическая точка двусторонней критической области, которую находят по таблице функции Лапласа по равенству $\Phi(z_{kp})=(1 — \alpha)/2$.
Если $|\tau| < T_{kp}$ — нет оснований отвергнуть нулевую гипотезу. Ранговая корреляционная связь между качественными признаками незначима. Если $|\tau| > T_{kp}$ — нулевую гипотезу отвергают. Между качественными признаками существует значимая ранговая корреляционная связь.
Найдем критическую точку z_{kp}
$\Phi(z_{kp}) = (1 - \alpha)/2 = (1 - 0.05)/2 = 0.475$
По таблице Лапласа находим $z_{kp} = 1.96$
Найдем критическую точку:

$$T_{kp} = 1.96\sqrt{\frac{2(2 \cdot 10+5)}{9 \cdot 10(10-1)}} = 0.49$$

2.5.3 Результаты и выводы:
Так как $\tau > T_{kp}$ — отвергаем нулевую гипотезу; ранговая корреляционная связь между оценками по двум тестам значимая.

2.6 Практическое занятие № 12-13 (4часа)

Тема: Регрессионные математические модели. Методы построения и статистической оценки. Оценка значимости коэффициентов, адекватности модели и ошибки прогнозирования. Задачи многофакторного моделирования.

2.6.1 Задание для работы:

1. Регрессионные модели прогнозирования
2. Расчет параметров уравнение регрессии. Технология работы
3. Интерпретация коэффициентов регрессии

2.6.2 Краткое описание проводимого занятия:

Оценка степени влияния производственных факторов на результат производства

Большинство явлений и процессов в экономике находятся в постоянной взаимной и всеохватывающей объективной связи. Исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами играет большую роль в экономике. Оно дает возможность глубже понять сложный механизм причинно-следственных отношений между явлениями. Для исследования интенсивности, вида и формы зависимостей широко применяется корреляционно-регрессионный анализ, который является методическим инструментарием при решении задач прогнозирования, планирования и анализа хозяйственной деятельности предприятий.

Различают два вида зависимостей между экономическими явлениями и процессами:

- функциональную;
- стохастическую (вероятностную, статистическую).

В случае функциональной зависимости имеется однозначное отображение множества А на множество В. Множество А называют областью определения функции, а множество В - множеством значений функции.

Функциональная зависимость встречается редко. В большинстве случаев функция (У) или аргумент (Х) - случайные величины. Х и У подвержены действию различных случайных факторов, среди которых могут быть факторы, общие для двух случайных величин.

Если на случайную величину А действуют факторы Z1,Z2, ... , V1,V2 а на У – Z0, Z1,Vx, Кз ... , то наличие двух общих факторов Z2 и Vx позволит говорить о вероятностной или статистической зависимости между Х и У.

Статистической называется зависимость между случайными величинами, при которой изменение одной из величин влечет за собой изменение закона распределения другой величины.

В частном случае статистическая зависимость проявляется в том, что при изменении одной из величин изменяется математическое ожидание другой. В этом случае говорят о корреляции или корреляционной зависимости.

Статистическая зависимость проявляется только в массовом процессе, при большом числе единиц совокупности.

При стохастической закономерности для заданных значений зависимой переменной можно указать ряд значений объясняющей переменной, случайно рассеянных в интервале. Каждому фиксированному значению аргумента соответствует определенное статистическое распределение значений функции. Это обусловливается тем, что зависимая переменная, кроме выделенной переменной, подвержена влиянию ряда неконтролируемых или неучтенных факторов. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть предсказаны с достаточной точностью, а только указаны с определенной вероятностью.

В экономике приходится иметь дело со многими явлениями, имеющими вероятностный характер. Например, к числу случайных величин можно отнести стоимость продукции, доходы предприятия, межремонтный проб е г автомобилей, время ремонта оборудования и т. д.

Односторонняя вероятностная зависимость между случайными величинами есть регрессия. Она устанавливает соответствие между этими величинами.

Односторонняя стохастическая зависимость выражается с помощью функции, которая называется регрессией.

Виды регрессий

1. Регрессия относительно числа переменных:

·простая регрессия - регрессия между двумя переменными;

·множественная регрессия - регрессия между зависимой переменной y и несколькими объясняющими переменными x_1, x_2 ... , $x_т$.

2. Регрессия относительно формы зависимости:

· линейная регрессия, выражаемая линейной функцией;

· нелинейная регрессия, выражаемая нелинейной функцией.

3. В зависимости от характера регрессии различаются следующие ее виды:

· положительная регрессия: она имеет место, если с увеличением (уменьшением) объясняющей переменной значения зависимой переменной также соответственно увеличиваются (уменьшаются);

· отрицательная регрессия: в этом случае с увеличением или уменьшением объясняющей переменной зависимая переменная уменьшается или увеличивается.

4. Относительно типа соединения явлений различаются:

· непосредственная регрессия: в этом случае зависимая и объясняющая переменные связаны непосредственно друг с другом;

· косвенная регрессия: в этом случае объясняющая переменная действует на зависимую через ряд других переменных;

· ложная регрессия: она возникает при формальном подходе к исследуемым явлениям без уяснения того, какие причины обусловливают данную связь.

Регрессия тесно связана с корреляцией. Корреляция в широком смысле слова означает связь, соотношение между объективно существующими явлениями. Связи между явлениями могут быть различны по силе. При измерении тесноты связи говорят о корреляции в узком смысле слова. Если случайные переменные причинно обусловлены, и можно в вероятностном смысле высказаться об их связи, то имеется корреляция.

Понятия «корреляция» и «регрессия» тесно связаны между собой. В корреляционном анализе оценивается сила связи, а в регресс ионном анализе исследуется ее форма. Корреляция в широком смысле объединяет корреляцию в узком смысле и регрессию. Корреляция, как и регрессия, имеет различные виды, так различают:

1) относительно характера – положительную и отрицательную

2) относительно числа переменных - простую, множественную, частную;

3) относительно формы связи – линейную, нелинейную;

4) относительно типа соединения – непосредственную, косвенную, ложную.

Любое причинное влияние может выражаться либо функциональной, либо корреляционной связью. Но не каждая функция или корреляция соответствует причинной зависимости между явлениями. Поэтому требуется обязательное исследование причинно-следственных связей.

Исследование корреляционных связей мы называем корреляционным анализом, а исследование односторонних стохастических зависимостей - регрессионным анализом.

Корреляционный анализ проводится между величинами, не имеющими причинно-следственного характера отклонений. Цель корреляционного анализа - количественное определение тесноты связи между признаками. Если исследуются только 2 признака, то

говорят о простой (парной) корреляции. Если исследуется связь между тремя и более признаками, то имеет место множественная корреляция. [

При оценке корреляционной связи не ставится вопрос о характере причинно-следственных соотношений между признаками. Оценивается только степень тесноты между ними. Оценку связи делают с помощью коэффициента корреляции г, который изменяется в пределах -1 > r > + 1.

Коэффициент корреляции - мера тесноты связи между признаками. Положительное значение коэффициента корреляции означает, что с увеличением одного показателя возрастает и другой, и наоборот, отрицательное значение определяет уменьшение величины одного показателя при возрастании другого.

Коэффициент корреляции свыше 0,8 означает тесную причинно- следственную связь.

Коэффициент множественной корреляции определяют по усложненной формуле, которая используется в статистике. В научных исследованиях и практической деятельности пользуются готовыми компьютерными программами для персонального компьютера.

Чтобы составить объективное представление о том, сильна или слаба связь между показателями, используют так называемый коэффициент детерминации, на основании которого можно сделать вывод о количестве случаев изменения одного показателя (признака) под влиянием другого. Коэффициент детерминации представляет собой квадрат коэффициента корреляции г, или то же, но выраженное в процентах (r2 * 100)

Итак, коэффициент корреляции, возведенный в квадрат и умноженный на 100 %, называется коэффициентом детерминации, который показывает, насколько результативный признак зависит от анализируемых одно- и двухфакторных признаков.

Корреляционный анализ позволяет установить связь между признаками и показывает форму этой связи, но он не дает представления об изменении одного показателя ряда в зависимости от изменения другого. Для исследования же нередко необходимо знать, насколько (в среднем) изменяется один признак при изменении другого на единицу. Эти важные и более глубокие свойства связи раскрывает регрессионный анализ, который для исследований различных вопросов экономики представляет большой интерес.

Применяя регрессионный анализ, можно, например, установить по некоторым показателям значения показателей совсем других размерностей, зная лишь о связи между ними и не затрачивая времени и средств на их непосредственное экспериментальное измерение или определение.

Регрессионный анализ - метод статистической обработки наблюдений, в результате которой оказывается возможным составить уравнение регрессии и получить количественную оценку влияния факторных признаков x на результативный y.

Уравнение $y\ i = b_{10}\ x1j + b_{20}\ x_{20} + ... + b\ mn\ x\ mj$,

где $b_{10}, b_{20},\ldots,b_{mn}$- средняя квадратическая оценка случайных факторов;

$x_{1i}, x_{20},\ldots, x_{mn}$ - значения непрерывных переменных x_1, x_2; называют уравнением регрессии.

Регрессионные зависимости могут быть самыми различными, важно лишь установить их экспериментально и сделать правильное математическое описание соответствующими формулами.

Регрессионный анализ проводят для установления связи между величинами, которые можно рассматривать как функции и аргументы, т. е. когда четко выражен характер причинно-следственных отношений между исследуемыми признаками.

Коэффициент регрессии показывает, на сколько изменится в среднем значение результативного признака y при увеличении факторного x.

Характеристикой относительного изменения прироста функции y=J{x) при малых относительных изменениях прироста аргумента x является эластичность функции.

Корреляционный и регрессионный анализ имеют свои задачи .

Задачи корреляционного анализа

1. Измерение степени связности (тесноты, силы) двух и более явлений. Здесь речь идет в основном о подтверждении уже известных связей.

2. Отбор факторов, оказывающих наиболее существенное влияние на результативный признак на основе измерения тесноты связи между явлениями.

3. Обнаружение неизвестных причинных связей. Корреляция непосредственно не выявляет причинных связей между явлениями, но устанавливает степень необходимости этих связей и достоверность суждений об их наличии. Причинный характер связей выясняется с помощью логически-профессиональных рассуждений, раскрывающих механизм связей.

Задачи регрессионного анализа

1. Установление формы зависимости (линейная или нелинейная; положительная или отрицательная и т. д.).

2. Определение функции регрессии и установление влияния факторов на зависимую переменную. Важно не только определить форму регрессии, указать общую тенденцию изменения зависимой переменной, но и выяснить, каково было бы действие на зависимую переменную главных факторов, если бы прочие не изменялись и если бы были исключены случайные элементы. для этого определяют функцию регрессии в виде математического уравнения того или иного типа.

3. Оценка неизвестных значений зависимой переменной, т. е. решение задач экстраполяции и интерполяции. В ходе экстраполяции распространяются тенденции, установленные в прошлом, на будущий период. Экстраполяция широко используется в прогнозировании. В ходе интерполяции определяют недостающие значения, соответствующие моментам времени между известными моментами, т. е. определяют значения зависимой переменной внутри интервала заданных значений факторов.

Однако экономические исследования многоаспектны, и, как правило, здесь применяют в комплексе все методы анализа. Так, например, цель регрессионного анализа - определение формы связи (уравнение регрессии), количественное определение коэффициентов уравнения (оценка коэффициентов регрессии) и определение изменения тесноты связи между признаками. Так же как и корреляционный, регрессионный анализ может проводиться по двум признакам y =f(x) (x - аргумент, y - функция), и тогда имеет место простой регрессионный анализ. Если же одновременно рассматривать несколько аргументов, т.е. y=f(xu x2,… ,x,,), то имеет место множественная регрессия.

В реальных условиях исследования прикладных задач, рассматриваемые признаки, как правило, взаимосвязаны, поэтому всегда приходится выявлять тесноту связи между ними и изучать формы связи.

Таким образом, более правильно говорить о комплексном корреляционно-регрессионном анализе.

Этапы проведения корреляционно-регрессионного анализа

Области его применения.

Корреляционно-регрессионный анализ проводят поэтапно в определенной логической последовательности.

1) анализ существа происходящих в исследуемой системе процессов и выявление причин возникновения взаимосвязей между признаками, характеризующими эти процессы;

2)выбор наиболее существенных признаков для исследования их на предмет включения в корреляционно-регрессионные модели, дифференциация признаков на факторные и результативные;

3)предварительный расчет и анализ парных коэффициентов корреляции, построение матрицы коэффициентов парной корреляции и оценка возможных вариантов группировки признаков для построения корреляционно регрессионных моделей;

4)выявление причинно-следственных соотношений между признаками и логическая оценка возможных вариантов формы связи, т.е. предварительная оценка формы уравнения регрессии;

5)решение уравнения регрессии - вычисление коэффициентов регрессии по уравнениям связи и их смысловая интерпретация с учетом прикладных задач исследуемой предметной области;

6)расчет теоретически ожидаемых (воспроизведенных по уравнению регрессии) значений результативного признака (функции);

7)определение и сравнительный анализ дисперсий: общей факторной (воспроизводственной) и остаточной; оценка тесноты связи между признаками, включенными в регрессионную модель;

8)общая оценка качества модели, отсев несущественных (или включение дополнительных) факторов, построение и решение новой модели (т. е. повторение п.п. 1-7, получение достаточно хорошей модели нередко требует ряда таких интерпретаций);

9)статистическая оценка достоверности параметров уравнения регрессии, построение доверительных границ для теоретически ожидаемых значений функции;

10) практические выводы из анализа.

В области экономико-математического моделирования и анализа экономических объектов и систем требуется переход от исследования отдельных процессов к изучению взаимодействия их совокупностей.

В рыночных условиях для получения исходно статистической информации используют методы маркетинговых исследований и управленческого учета. Для подготовки решений, ориентированных на перспективу, необходимо использование методов прогноза для обработки маркетинговой и учетной информации.

Целесообразно использование на предприятиях АПК корреляционно регрессионного анализа для определения:

доли рынка при различных ценах на продукцию;

культур в зависимости от погодных условий (температуры воздуха, температуры и влажности почвы в определенное время), от норм внесения минеральных удобрений;

продуктивности животных в зависимости от питательности кормов и живой массы.

Разработка числовой экономико-математической модели задачи

Статистические методы являются составной частью эконометрики науки, изучающей экономические явления с количественной точки зрения. Эконометрика устанавливает и исследует количественные закономерности в экономике на основе методов теории вероятности и математической статистики, адаптированных к обработке экономических данных.

Закономерности в экономике выражаются в виде связей и зависимостей экономических показателей, математических моделей их поведения. Такие зависимости и модели могут быть получены только путем обработки реальных статистических данных, с учетом внутренних механизмов связи и случайных факторов. Модель может быть получена и апробирована на основе анализа статистических данных, и изменения в поведении последних говорят о необходимости уточнения и развития модели.

Любое эконометрическое исследование всегда предполагает объединение теории (экономической модели) и практики (статистических данных). Мы используем теоретические модели для описания и объяснения наблюдаемых процессов и собираем статистические данные с целью эмпирического построения и обоснования моделей.

Введем переменные. Независимый показатель: у - производительность, руб./чел.-час. Факторные показатели:

X_1 - фондообеспеченность на 100 га площади сельскохозяйственных угодий, тыс. руб.

X_2 - фондовооруженность на одного работника, тыс. руб.

X_3 - урожайность, ц/га. (см. таблица 8).

Предлагается проанализировать степень влияния на производительность следующих факторов: x_1 - ФО, x_2 - ФВ, x_3 - урожайность.

Анализ результатов решения

Рассмотрим производительность при помощи корреляционно-регрессионного анализа. Для этого определим:

1. От какого фактора может зависеть производительность. Рассмотрим, например такие показатели как фондообеспеченность на 100га площади сельскохозяйственных угодий, фондовооруженность на 1-го работника и урожайность. Используя пакет прикладных программ Excel, рассчитаем коэффициенты корреляции и определим наиболее близкие к единице коэффициенты, которые будут свидетельствовать о тесноте связи между факторным и результативным признаком (табл.10), но для этого необходимо сгруппировать предполагаемые факторные показатели в таблицу (табл.7)

Таблица 7- Исходные данные для определения матрицы парных коэффициентов корреляции

Годы A	Производительность, руб./чел.-час 1	Фондообеспеченность на 100 га площади сельскохозяйственных угодий, тыс. руб 2	Фондовооруженность на одного работ. тыс. руб. 3	Урожайность, ц/га 4
12567	793	74	7,6	
15782	823	97	8,4	
18865	836	105	9,3	
18689	868	138	10,8	
19851	902	174	13,9	
18321	939	193	12,5	
17814	1021	201	9,4	
23068	1210	258	7,1	
27017	1292	300	5,9	
27331	2639	608	11,4	

Таблица 8 -Матрица парных коэффициентов корреляции

	Y	X1	X2	X3
Y	1			
X1	0,7513264	1		
X2	0,83496389	0,980173642	1	
X3	-0,0642293	0,096956747	0,116972555	1

Т.к. коэффициент корреляции rx2y = 0,835 связь между x2 и y считается тесной; прямой т.е. при увеличении факторного признака фондовооруженности значение результативного признака производительности увеличивается.

Из расчетов следует, что для последующего анализа факторным признаком будет являться такой показатель как фондовооруженность.

2. Следующим этапом анализа производительности является установление формы зависимости между переменными, для этого рассмотрим несколько моделей и выберем наиболее лучшую из них, на основе, которой будет составлен прогноз.

Составим и проанализируем следующие модели: линейную, степенную, показательную и гиперболическую.

Для того чтобы рассмотреть линейную модель, необходимо составить уравнение линейной регрессии (y^= a+b*x), что предполагает вычисление параметров a и b. Данные параметры определим при помощи пакета прикладных программ Excel (выбираем меню «Вставка» далее «Функция», «Статистические», «Линейн», заполняем диалоговое окно и нажимаем F2 и комбинацию клавиш Ctrl+Shift+Enter).

Для рассмотрения степенной, показательной и гиперболической моделей, необходимо составить уравнение степенной, показательной и гиперболической регрессии (y^=a*xb , y^=a*bx и y^=a+b/x), что предполагает линеаризацию данных моделей путем логарифмирования для степенной и показательной модели, а для гиперболической замену переменной. Коэффициенты a и b вычисляются также как и для линейной модели, только с преобразованными переменными. (расчет см. табл. 10, 11, 12)

Проведенные расчеты показывают, что рассматриваемые модели имеют следующий вид:

- Линейная – y^ = 25,05 * X2+14549,06;
- Степенная – y^= 3287,99 * X20,34;
- Показательная – y^ = 14943,67 * 1,001X2 ;
- Гиперболическая – y^ = 27253,29 – 1120538,5/X2

Таблица 9 - Определение параметров a и b уравнения линейной регрессии

b	a
25,0532708	14549,05742
5,83787784	1521,333492
0,69716469	2723,966562
18,4169988	8
136654018	59359950,66

Таблица 10 - Определение параметров a и b уравнения степенной регрессии

b	a
0,3427381	8,0980319
0,0602364	0,3143198
0,8018563	0,1114492
32,374732	8
0,402124	0,0993674

Таблица 11 - Определение параметров a и b уравнения показательной регрессии

b	a
0,00120956	9,615383336
0,00032411	0,084461915
0,63515948	0,15123011
13,9273891	8
0,318527	0,18296437

Таблица 12 - Определение параметров a и b уравнения гиперболической регрессии

b	a
-1120538,5	27253,28815
223169,262	1648,398149
0,75911375	2429,430767
25,2106952	8
148796898	47217070,8

Проанализируем коэффициенты регрессии:

Линейной модели. Коэффициент регрессии b = 25,05 показывает, что при увеличении фондовооруженности на 1 пункт производительность увеличивается на 25,05 руб./чел.-час.

Степенной модели. Коэффициент регрессии b = 0,343 показывает, что при увеличении фондовооруженности на 1 пункт производительность увеличивается на 0,343 руб./чел.-час.

Показательная модель. Коэффициент регрессии b = 0,001 показывает, что при увеличении фондовооруженности на 1 пункт производительность увеличивается на 0,001 руб./чел.-час.

Гиперболическая модель. Коэффициент регрессии b = -1120538,5 показывает, что при увеличении фондовооруженности на 1 пункт производительность уменьшается на -1120538,5 руб./чел.-час.

3. Рассчитаем и проанализируем коэффициенты, оценивающие построенные модели (табл.13).

Таблица 13- Сводная таблица показателей

Модель	A	R2	Fрасч	r
линейная	21,4787	0,6972	18,4170	0,8350
степенная	17,1118	0,8019	32,3747	0,8905
показательная	3985898,1393	0,6352	13,9274	0,7818
гиперболическая	16,5784	0,7591	25,2107	0,8713

Ошибка аппроксимации (А) показывает, что превышено допустимое значение (8-10%) среднего отклонения расчетных данных от фактических во всех моделях.

Коэффициент детерминации:

Линейная модель. R2 равный 0,6972 показывает, что вариация получения производительности на 69,72% объясняется вариацией фондовооруженности на 30,28% зависит от других, не учтённых факторов.

Степенной модели. R2 равный 0,8019 показывает, что вариация получения производительности на 80,19% объясняется вариацией фондовооруженности на 19,81% зависит от других, не учтённых факторов.

Показательная модель. R2 равный 0,6353 показывает, что вариация получения производительности на 63,52% объясняется вариацией фондовооруженности на 36,48% зависит от других, не учтённых факторов.

Гиперболическая модель. R2 равный 0,7591 показывает, что вариация получения производительности на 75,91% объясняется вариацией фондовооруженности на 24,09% зависит от других, не учтённых факторов.

F-критерий Фишера позволяет оценить значимость и надежность уравнения, и т.к. Fрасч < Fтабл (5,12) во всех моделях значит построенные уравнения регрессии не значимы и надежны.

Индекс корреляции (r): во всех связь тесная, т.к. индекс корреляции этих моделей превышает 0,7.

2.6.3 Результаты и выводы:

Таким образом, из проведенного анализа следует, что наиболее лучшей моделью отражающей зависимость получения производительности от коэффициента фондовооруженности является линейная модель.

Проанализировав взаимосвязь получения производительности и коэффициента фондовооруженности, следует отметить, что в рассматриваемом периоде производительность

имеет возрастающую тенденцию, при этом производительность увеличивается на 25,05 руб./ чел.-час. при увеличении коэффициента фондовооруженности на 1 пункт.

Из графика следует, что в прогнозируемом периоде производительность будет увеличиваться, т.к. имеет возрастающий тренд и производительность составит 30330,56 руб./ чел.-час. при коэффициенте фондовооруженности равного 630.

Таким образом, в результате анализа выявлено, что между производительностью и фондообеспеченностью на 100 га с/х угодий, фондовооруженностью и урожайностью средняя степень прямой линейной зависимости.

2.7 Практическое занятие № 14-17 (8часа)

Тема: Методы оптимизации. Использование надстроек Microsoft Excel

2.7.1 Задание для работы:

1. ЗЛП. Графический метод решения ЗЛП, симплекс-метод
2. Постановка задачи линейного программирования и свойства ее решений
3. Симплексный метод решение ЗЛП
4. Теория двойственности
5. Основные виды задач, сводящихся к ЗЛП
6. Транспортная задача, методы ее решения.

2.7.2 Краткое описание проводимого занятия:

Методы решения ЗЛП

1) Решить графическим методом типовую задачу оптимизации.

На имеющихся у фермера 400 га земли он планирует посеять кукурузу и сою. Сев и уборка кукурузы требуют на каждый гектар 200 ден. ед. затрат, а сои – 100 ден. ед. На покрытие расходов, связанных с севом и уборкой, фермер получил ссуду в 60 тыс. ден. ед. Каждый гектар, засеянный кукурузой, принесет 30 центнеров, а каждый гектар, засеянный соей, - 60 центнеров. Фермер заключил договор на продажу, по которому каждый центнер кукурузы принесет ему 3 ден. ед., а каждый центнер сои – 6 ден. ед. Однако согласно этому договору фермер обязан хранить убранное зерно в течение нескольких месяцев на складе, максимальная вместимость которого равна 21 тыс. центнеров.

Фермеру хотелось бы знать, сколько гектаров нужно засеять каждой из этих культур, чтобы получить максимальную прибыль.

Построить экономико-математическую модель задачи, дать необходимые комментарии к ее элементам и получить решение графическим методом. Что произойдет, если решать задачу на минимум, и почему?

Решение.

Обозначим через x_1 сколько гектаров нужно засеять кукурузы, через x_2 – сои. Так как у фермера всего имеется 400 га земли, то первое ограничение задачи имеет вид: $x_1 + x_2 ? 400$. Найдем общие затраты на сев и уборку кукурузы и сои: $(200x_1+100x_2)$ ден. ед. Фермер получил на расходы ссуду в 60 тыс. ден., поэтому следующее ограничение имеет вид: $200x_1+100x_2 ? 60\,000$. Найдем, сколько центнеров зерна соберет фермер: $(30x_1+60x_2)$ ц. Вместимость склада составляет 21 тыс. центнеров, поэтому следующее ограничение имеет вид: $30x_1+60x_2 ? 21\,000$. Выясним сколько ден. ед. получит фермер по договору за собранное зерно: $(30x_1 \cdot 3+60x_2 \cdot 6)$ ден. ед.

Построим экономико-математическую модель задачи:

$$\max f(X) = 90x_1+120x_2$$

$x_1 + x_2 \ ? \ 400$

$200x_1 + 100x_2 \ ? \ 60\,000$

$30x_1 + 60x_2 \ ? \ 21\,000$

$x_{1,2} \geq 0$

Решим задачу графическим методом.
Последнее ограничение означает, что область решений будет лежать в первой четверти декартовой системы координат.
Этап 1. Определим множество решений первого неравенства. Оно состоит из решения уравнения и строгого неравенства. Решением уравнения служат точки прямой $x_1+x_2-400=0$. Построим прямую по двум точкам (0;400) и (400;0), которые легко получить в результате последовательного обнуления одной из переменных. На рисунке обозначим ее цифрой I.
Множество решений строгого неравенства — одна из полуплоскостей, на которую делит плоскость построенная прямая. Какая из них является искомой, можно выяснить при помощи одной контрольной точки. Если в произвольно взятой точке, не принадлежащей прямой, неравенство выполняется, то оно выполняется и во всех точках той полуплоскости, которой принадлежит контрольная точка, и не выполняется во всех точках другой полуплоскости. В качестве такой точки удобно брать начало координат. Подставим координаты (0; 0) в неравенство $x_1+x_2-400<0$, получим $-400 < 0$, т.е. оно выполняется. Следовательно, областью решения неравенства служит нижняя полуплоскость.
Аналогичным образом построим области решения двух других неравенств

$200x_1+100x_2-60\,000=0$; $x_1 = 0$, $x_2 = 600$

$x_1 = 300$, $x_2 = 0$ (на рисунке прямая II);

$200x_1+100x_2-60\,000<0$ при $x_1 = x_2 = 0$, $-60\,000<0$ выполняется, берется левая полуплоскость.

$30x_1+60x_2-21\,000=0$ $x_1 = 0$, $x_2 = 350$

$x_1 = 700$, $x_2 = 0$ (на рисунке прямая III);

$30x_1+60x_2-21\,000<0$ при $x_1 = x_2 = 0$, $-21\,000<0$ выполняется, берется нижняя полуплоскость.
Заштрихуем общую область для всех неравенств. Обозначим вершины области латинскими буквами и определим их координаты, решая систему уравнений двух пересекающихся соответствующих прямых. Например, определим координаты точки B, являющейся точкой пересечения первой и третьей прямой:

$x_1+x_2-400=0$, $x_1 = 100$; $x_2 = 300$

$30x_1+60x_2-21\,000=0$.

Вычислим значение целевой функции в этой точке:

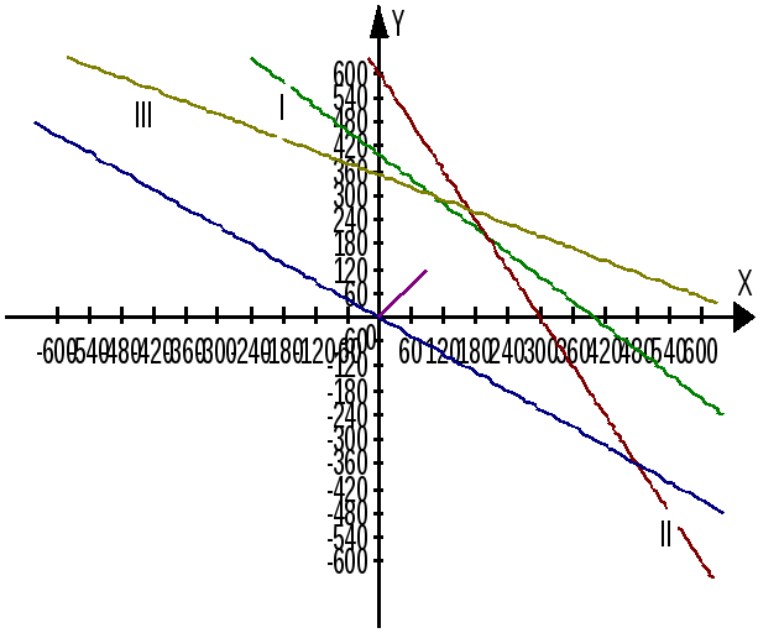
$f(X)= 90x_1+120x_2=90\cdot100 + 120\cdot300 = 45000$.
Аналогично поступим для других точек, являющихся вершинами области ABCDO, представляющей собой область допустимых решений рассматриваемой ЗЛП. Координаты этих

вершин имеют следующие значения: т. A(0;350), т. B(100;300), т.C(200;200), т. D(300;0), т. O(0;0).

Этап 2. Приравняем целевую функцию постоянной величине a: $90x_1+120x_2 = a$.
Это уравнение является множеством точек, в котором целевая функция принимает значение, равное a. Меняя значение *a,* получим семейство параллельных прямых, каждая из которых называется *линией уровня.* Пусть a=0, вычислим координаты двух точек, удовлетворяющих соответствующему уравнению $90x_1+120x_2 = 0$. В качестве одной из этих точек удобно взять точку O(0;0), а так как при $x_1 = 4$, $x_2 = -3$ то в качестве второй точки возьмем точку G(4;-3). Через эти две точки проведем линию уровня $f(X)= 90x_1+120x_2 = 0$.

Этап 3. Для определения направления движения к оптимуму построим вектор-градиент , координаты которого являются частными производными функции *f(X),* т.е. =(90;120) Чтобы построить этот вектор, нужно соединить точку (90;120) с началом координат. При максимизации целевой функции необходимо двигаться в направлении вектора-градиента, а при минимизации — в противоположном направлении.

В нашем случае движение линии уровня будем осуществлять до ее пересечения с точкой B, далее она выходит из области допустимых решений. Следовательно, именно в этой точке достигается максимум целевой функции. Отсюда легко записать решение исходной ЗЛП: max f(X) =45000 и достигается при $x_1 = 100$, $x_2=300$. Следовательно, чтобы получить максимальную прибыль, фермер должен засеять 100 га земли кукурузой, 300 га – соей. При этом он получит 45 тыс. ден. ед. при реализации зерна по договору. Если поставить задачу минимизации функции $f(X) = 90x_1+120x_2$ при тех же ограничениях, линию уровня необходимо смещать параллельно самой себе в направлении, противоположном вектору-градиенту. В нашем случае минимум функции будет в точке O(0;0). Это означает, что фермер не получит ни чего, если не засеет поле зерновыми культурами.



Задача 2 Предприятие выпускает два вида продукции используя три вида ресурсов. Приняты обозначения:

A – матрица норм затрат сырья;

$$A=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

В – запасы ресурсов; $B = (в_1 \quad в_2 \quad в_3)$

С – прибыль на единицу продукции $C = (c_1 \quad c_2)$

С помощью следующих данных составить математическую модель. Определить план выпуска изделий, обеспечивающих максимальную прибыль с помощью графического метода.

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \\ 1 & 1 \end{bmatrix}; \qquad B = \begin{bmatrix} 90 \\ 120 \\ 40 \end{bmatrix}; \qquad C = (5;2)$$

Решение задачи.

Обозначим через x1 количество единиц продукции первого вида, а через x2 – количество единиц продукции второго. Тогда, учитывая количество единиц сырья, расходуемое на изготовление продукции, а так же запасы сырья, получим систему ограничений:

1*x1+3*x2<=90
4*x1+2*x2<=120
1*x1+1*x2<=40
x1,x2 >=0; - условие неотрицательности переменных.

Конечную цель решаемой задачи – получение максимальной прибыли при реализации продукции – выразим как функцию двух переменных x1 и x2. Реализация x1 единиц продукции первого вида и x2 единиц продукции второго дает соответственно 5x1 и 2x2 ден. ед. прибыли, суммарная прибыль C = 5x1 + 2x2. Условиями не оговорена неделимость единицы продукции, поэтому x1 и x2 (план выпуска продукции) могут быть и дробными числами. Требуется найти такие x1 и x2, при которых функция C достигает максимум, т.е. найти максимальное значение линейной функции C = 5x1 + 2x2 при ограничениях.

Математическая модель задачи: Cmax = 5x1 + 2x2

Система ограничений:

1*x1+3*x2<=90
4*x1+2*x2<=120
1*x1+1*x2<=40
x1,x2 >=0; - условие неотрицательности переменных.

Решение задачи с использованием графического симплекс-метода.

Построим систему координат и проведем прямые ограничивающие область допустимых решений (ОДР), построив их, соответственно, по неравенствам системы ограничений. Чтобы построить прямую нужно знать координаты двух точек. Координаты точек прямых соответствующих неравенствам:

Неравенство	11	21	12	22
1*x1+3*x2 <=90	0			0
4*x1+2*x2 <=120	0			0
1*x1+1*x2 <=40	0			0

Построим вектор целевой функции C(5;2). Система координат с областью допустимых решений OABCD и вектором целевой функции C приведена на рис.

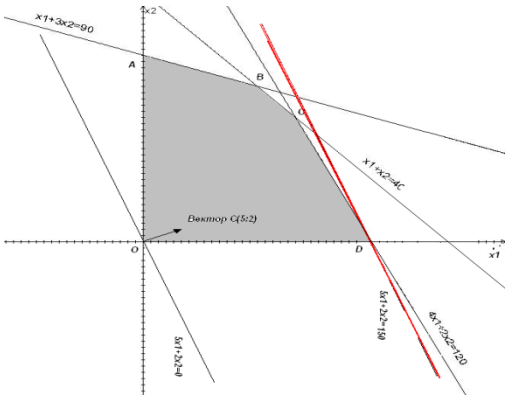


Рис. График области допустимых решений.

Построим линию уровня 5x1+2x2 = 0, проходящую через начало координат и перпендикулярную вектору C (5;2). Будем передвигать ее в направлении вектора C, в результате чего находим точку, в которой функция принимает максимальное значение – точку D. При дальнейшем перемещении она уже не будет иметь общих точек с областью допустимых решений OABCD. Точка D имеет координаты (30;0). Cmax = 5*30+2*0=150

Ответ: Для того чтобы получить максимальную прибыль в размере 150 ден. ед., необходимо запланировать производство 30 ед. продукции первого вида, а продукцию второго вида не выпускать совсем.

Решение задачи с использованием метода симплекс-таблиц.

Приведем математическую модель задачи к каноническому виду, избавившись от неравенств посредством ввода дополнительных переменных:

Целевая функция:

C max = 5*x1+2*x2+0*x3+0*x4+0*x5

Система ограничений:

1*x1+3*x2+x3=90
4*x1+2*x2+x4=120
1*x1+1*x2+x5=40

Проведем векторный анализ системы ограничений. Выберем единичные вектора, позволяющие получить систему координат и указать в ней координаты одной из вершин симплекса.

P0 - вектор свободных коэффициентов

Pi - вектор коэффициентов при переменной xi

Расширенная целевая функция:

C max = 5*x1+2*x2+0*x3+0*x4+0*x5

Вектора:

0	**P 1(x1)**	**P 2(x2)**	**P 3(x3)**	**P 4(x4)**	**P 5(x5)**
0	1	3	1	0	0
20	4	2	0	1	0
	1	1	0	0	1

0					

Базисными могут быть только единичные вектора. Базис:
Базисный вектор №1: P3(x3)
Базисный вектор №2: P4(x4)
Базисный вектор №3: P5(x5)
Заполним первую таблицу:

	Базис	Коэффициенты при базисе	**P0**		2	0	0	0
				1	**P2**	**P3**	**P4**	**P5**
	P3	0	90		3	1	0	0
	P4	0	120		2	0	1	0
	P5	0	40		1	0	0	1
C max =			**0**	-5	-2	0	0	0

При просмотре последней (индексной) строки среди коэффициентов этой строки (исключая столбец свободных членов) находим наименьшее отрицательное число: -5 (первый столбец - ключевой).

Просматривая первый столбец таблицы (ключевой) выбираем среди положительных коэффициентов столбца тот, для которого абсолютная величина отношения соответствующего свободного члена (находящегося в столбце свободных членов) к этому элементу минимальна – 4. Этот коэффициент называется разрешающим, а строка, в которой он находится ключевой;

Замещаемый базисный вектор: P4 (2-я строка)

Новый базисный вектор: P1 (1-й столбец)

Заменяем базисный вектор P4 на P1.

Строим новую таблицу, содержащую новые названия базисных переменных, для этого:

- разделим каждый элемент ключевой строки (исключая столбец свободных членов) на разрешающий элемент и полученные значения запишем в строку с измененной базисной переменной новой симплекс таблицы.

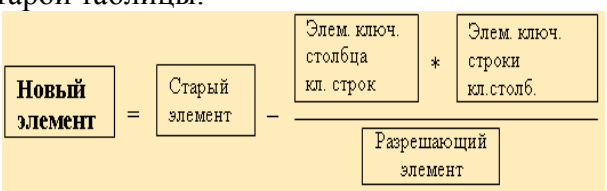
- строка разрешающего элемента делится на этот элемент и полученная строка записывается в новую таблицу на то же место.

- в новой таблице все элементы ключевого столбца = 0, кроме разрезающего, он всегда равен 1.

- столбец, у которого в ключевой строке имеется 0,в новой таблице будет таким же.

- строка, у которой в ключевом столбце имеется 0,в новой таблице будет такой же.

- в остальные клетки новой таблицы записывается результат преобразования элементов старой таблицы:



В результате получили новую симплекс-таблицу, отвечающую новому базисному решению:

№	Базис	Коэффициенты при базисе	**P0**	5	2	0	0	
				P1	**P2**	**P3**	**P4**	**5**
1	P3	0	60	0	2.5	1	-0.25	
2	P1	5	30	1	0.5	0	0.25	
3	P5	0	10	0	0.5	0	-0.25	
C max =		**150**	0	0.5	0	1.25	0	

Просматривая строку целевой функции (индексную), видим, что в ней нет отрицательных значений, значит, оптимальное решение получено.

Из таблицы получим значения переменных целевой функции:

1	**2**	**3**	**4**	**5**
0		0		0

Целевая функция: C max = 5*30+2*0

И в результате: **Ответ:** Для того чтобы получить максимальную прибыль в размере 150 ден. ед., необходимо запланировать производство 30 ед. продукции первого вида, а продукцию второго вида не выпускать совсем (ответ совпадает с ответом, полученным графическим методом).

Методы решения ЗЛП

1. Решение ЗЛП симплекс-методом
2. Двойственные задачи.
3. Программное обеспечение решения задач линейного программирования на ПК.

Задача. *С.\ х. предприятие производит и продаёт продукцию двух видов: «1 Продукт» и «2 Продукт». Для производства продукции используются ресурсы двух категорий: А и В. Расходы ресурсов А и В на производство единицы продукции каждого вида, запасы продуктов и цены продуктов приведены в таблице 1.*

Таблица 1

Ресурсы	*Расход ресурсов на ед. продукции*		*Запасы ресурсов*
	1 Продукт	*2 Продукт*	
А	*1*	*2*	*3*
В	*3*	*1*	*3*
Количество продукции	x_1	x_2	
Цены	*2(ден. ед.)*	*1(ден. ед.)*	

Выяснить, какое количество продукции каждого вида надо производить предприятию(составить план производства), чтобы получить максимум прибыли.

Задание.

1. *Составить математическую модель задачи.*

2. *Решить задачу в Excel.*

Решение. 1. Составить математическую модель задачи. Для составления математической модели задачи прежде всего ***введём переменные (неизвестные) задачи***: x_1 - количество продукции 1-го вида, а x_2 - количество продукции 2-го вида, производимые предприятием.

Ограниченность запасов ресурсов приводит ***к ограничениям на*** x_1 ***и*** x_2 ***:*** ограничения на расход ресурса А $x_1 + 2 \cdot x_2 \le 3$,

ограничения на расход ресурса В $3 \cdot x_1 + x_2 \le 3$.

Кроме того, $x_1, x_2 \ge 0$.

Качество решения задачи определяется с помощью ***целевой функции задачи*** $Z(x_1, x_2)$ - функции, определяющей доход предприятия от продажи продукции: $Z = 2 \cdot x_1 + x_2$.

Задача об определении плана производства продукции свелась к следующей математической задаче: *найти вектор* (x_1, x_2) *(план производства), координаты которого удовлетворяют системе ограничений*

$$\begin{cases} x_1 + 2 \cdot x_2 \le 3 \\ 3 \cdot x_1 + x_2 \le 3 \end{cases}$$

и условиям неотрицательности $x_1, x_2 \ge 0$,

который доставляет максимум целевой функции $Z = 2 \cdot x_1 + x_2$.

Эту математическую задачу принято записывать в виде

$$Z = 2 \cdot x_1 + x_2 \to \max \quad (1)$$

$$\begin{cases} x_1 + 2 \cdot x_2 \le 3 \\ 3 \cdot x_1 + x_2 \le 3 \end{cases} \quad (2)$$

$$x_1, x_2 \ge 0. \quad (3)$$

и называть ***математической моделью*** данной производственной задачи.

Подобные задачи называются ***задачами линейного программирования***. Они изучаются в разделе математики, называемом ***математическим программированием***. Так как переменные x_1 и x_2 входят в систему ограничений (2) и целевую функцию Z (1) линейно, то эту задачу математического программирования называют ***задачей линейного программирования***.

Множество точек декартовой плоскости (x_1, x_2), координаты которых удовлетворяют системе ограничений (2) и условиям неотрицательности (3), называется областью допустимых решений задачи линейного программирования(областью допустимых планов). В данной задаче она представляет собой выпуклый четырёхугольник. Значения x_1^* и x_2^* из области допустимых планов, при которых Z принимает наибольшее значение в этой области, называются ***оптимальными(оптимальный план),*** а соответствующее наибольшее значение $Z^* = 2 \cdot x_1^* + x_2^*$ является **оптимальным значением прибыли.** Таким образом, задача о распределении ресурсов является задачей оптимизации и её математической моделью служит задача линейного программирования, заключающаяся в поиске оптимального плана и оптимального значения целевой функции.

Задачей оптимизации может быть поиск наименьшего значения.

2. Решение задачу в Excel.

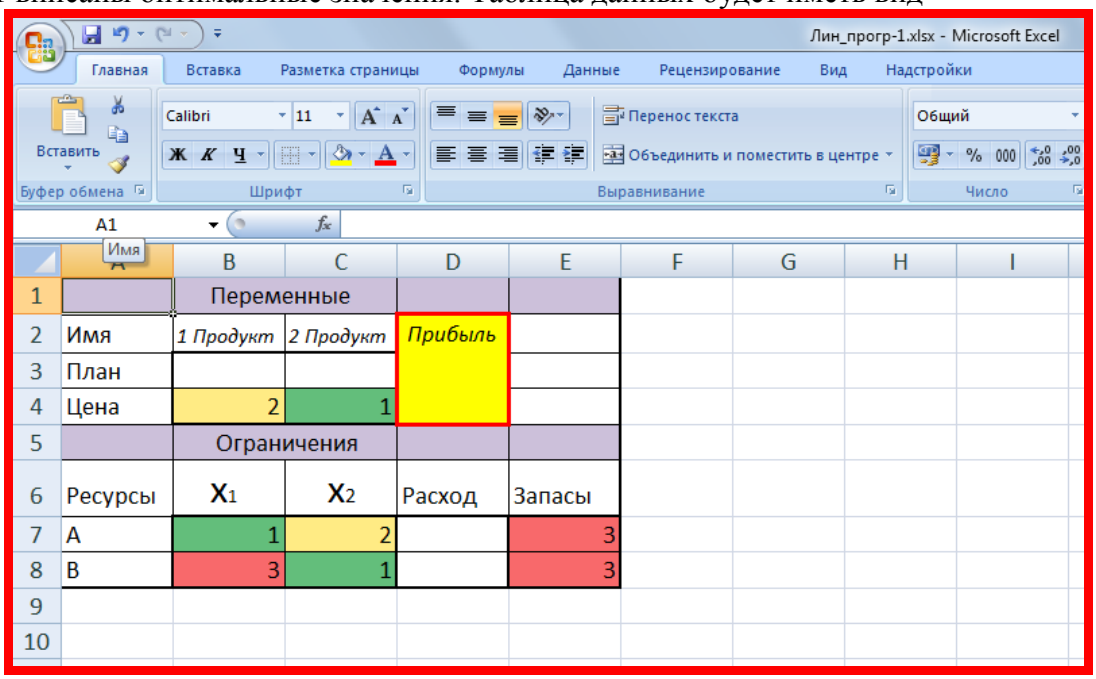
2.1. Ввод данных и формул в таблицу Excel. Открыть Книгу ***Excel,*** Лист1.

-Объединим ячейки B1 и C1. Для этого выделить ячейки, нажать правую кнопку мыши. В появившемся окне вызвать «Формат ячеек», затем «Выравнивание» и поставить галочку против опции «объединение ячеек», нажать ОК. В объединённые ячейки впишем заголовок «Переменные».

-В ячейку A2 вписать «Имя», в A3- «План», в ячейку A4 «Цена», в B2- «1 Продукт», в C2- «2 Продукт», в D2 « Прибыль».

-В ячейки B4 и C4 заносятся значения цен на продукцию.

-Для переменных x_1 и x_2 отводятся ячейки B3 и C3. Это изменяемые(рабочие) ячейки, В них исходные данные не заносятся и в результате решения задачи в эти ячейки будут вписаны оптимальные значения. Таблица данных будет иметь вид

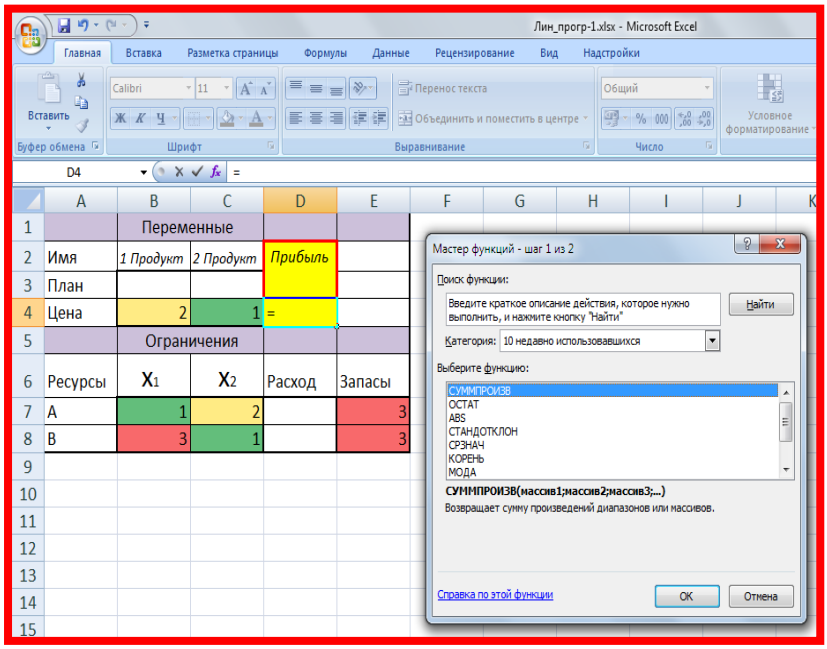


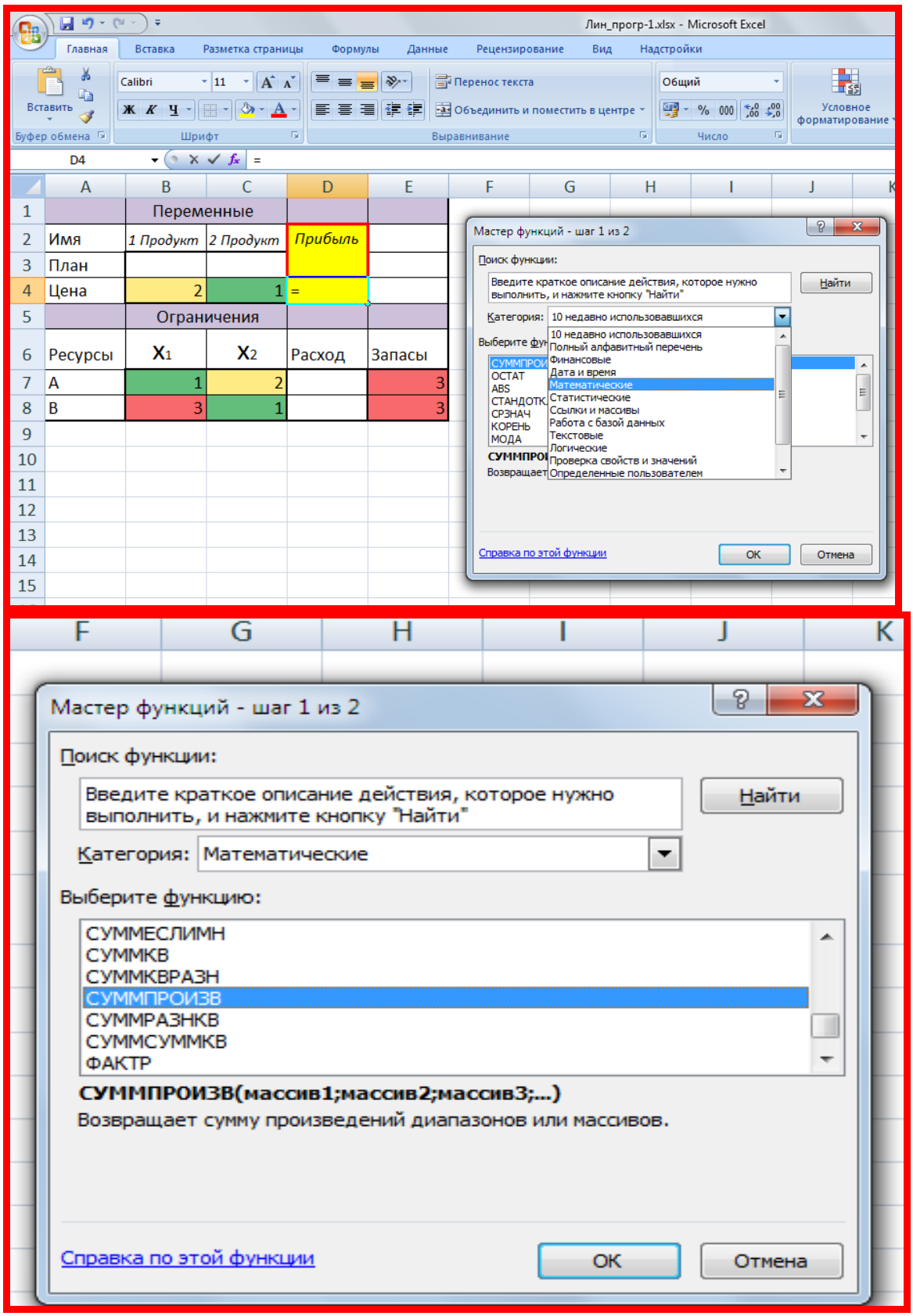
	A	B	C	D	E
1		Переменные			
2	Имя	*1 Продукт*	*2 Продукт*	*Прибыль*	
3	План				
4	Цена	2	1		
5		Ограничения			
6	Ресурсы	X_1	X_2	Расход	Запасы
7	A	1	2		3
8	B	3	1		3
9					
10					

-В ячейке D4 после окончания решения задачи будет указана оптимальное значение прибыли(целевая ячейка). С этой целью в ячейку D4 вводится формула для вычисления значений целевой функции $Z = 2 \cdot x_1 + x_2$. Для этого надо выполнить следующие операции:

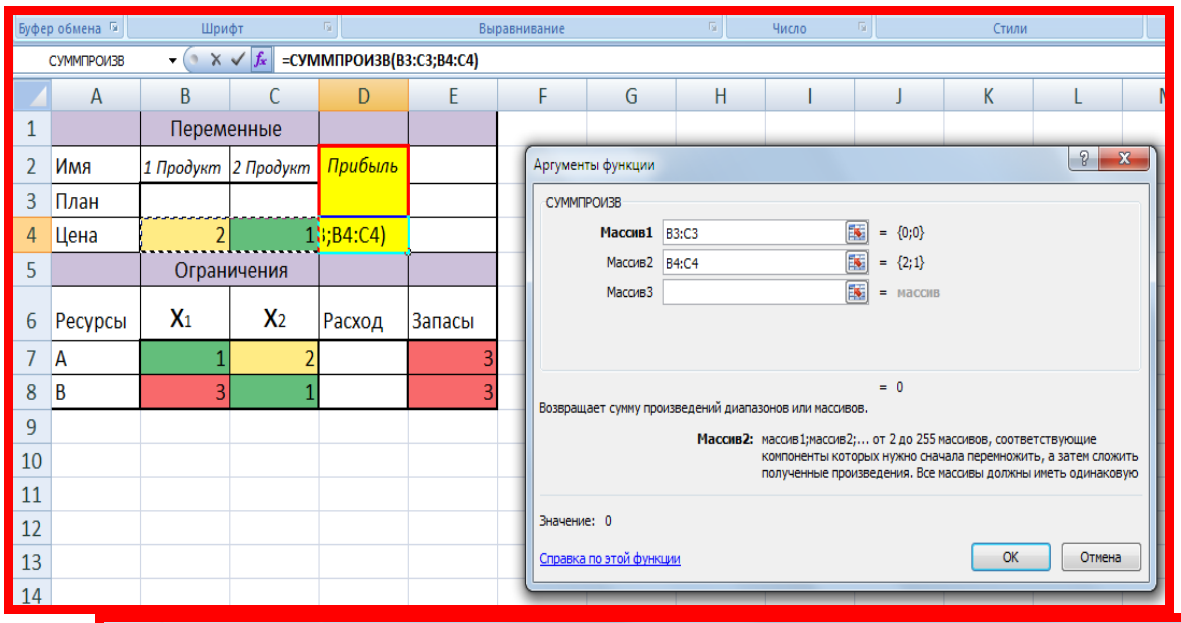
1) курсор в D4, выделить эту ячейку,

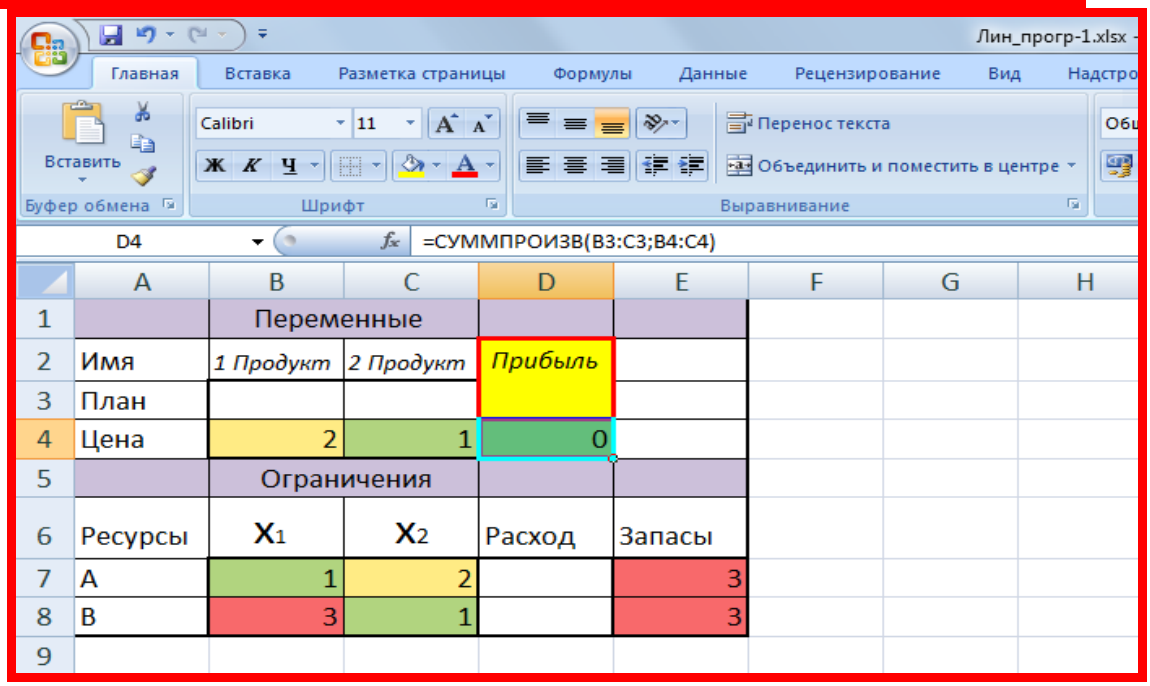
2) щёлкнув по кнопке f_x вызвать Мастера функций, в открывшемся окне в категории «10 недавно использовавшихся» выбрать «Математические», а затем «СУММПРОИЗВ», ОК.





В появившемся окне «Аргументы функции» в поле «Массив 1» ввести адреса изменяемых ячеек B3:C3(протаскивая курсор мыши по ячейкам), в поле «Массив 2» вводятся адреса ячеек с ценами на продукцию B4:C4, «Массив 3» игнорируется. Нажать ОК. В ячейке D4 появится число 0.





-Объединить ячейки B5 и C5 и вписать «Ограничения», в A6- «Ресурсы», в B6 и C6 x_1 и x_2, в D6 «Расход», в E6 «Запасы», A7 и A8 значки ресурсов,

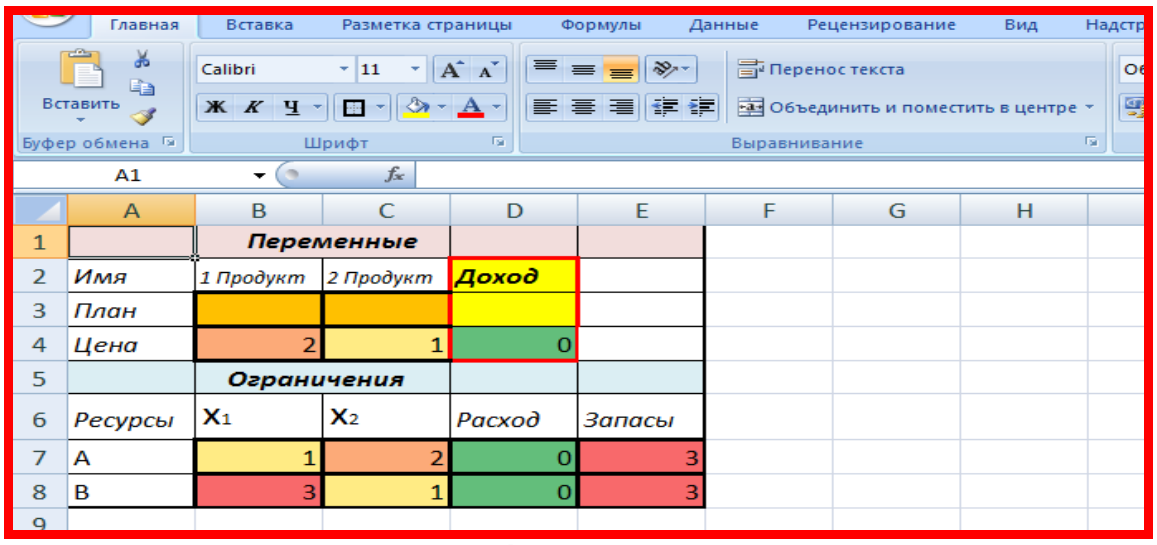
в поле B7:C8- нормы расхода ресурсов.

-В ячейку D7 вводится формула вычисления израсходованного ресурса A $x_1 + 2 \cdot x_2$, в ячейку D8- формула израсходованного ресурса B $3 \cdot x_1 + x_2$

(также, как и формула целевой функции).

- В ячейки E7 и E8 вносим размеры запасов ресурсов.

Данные и формулы введены. Интерфейс задачи будет иметь вид

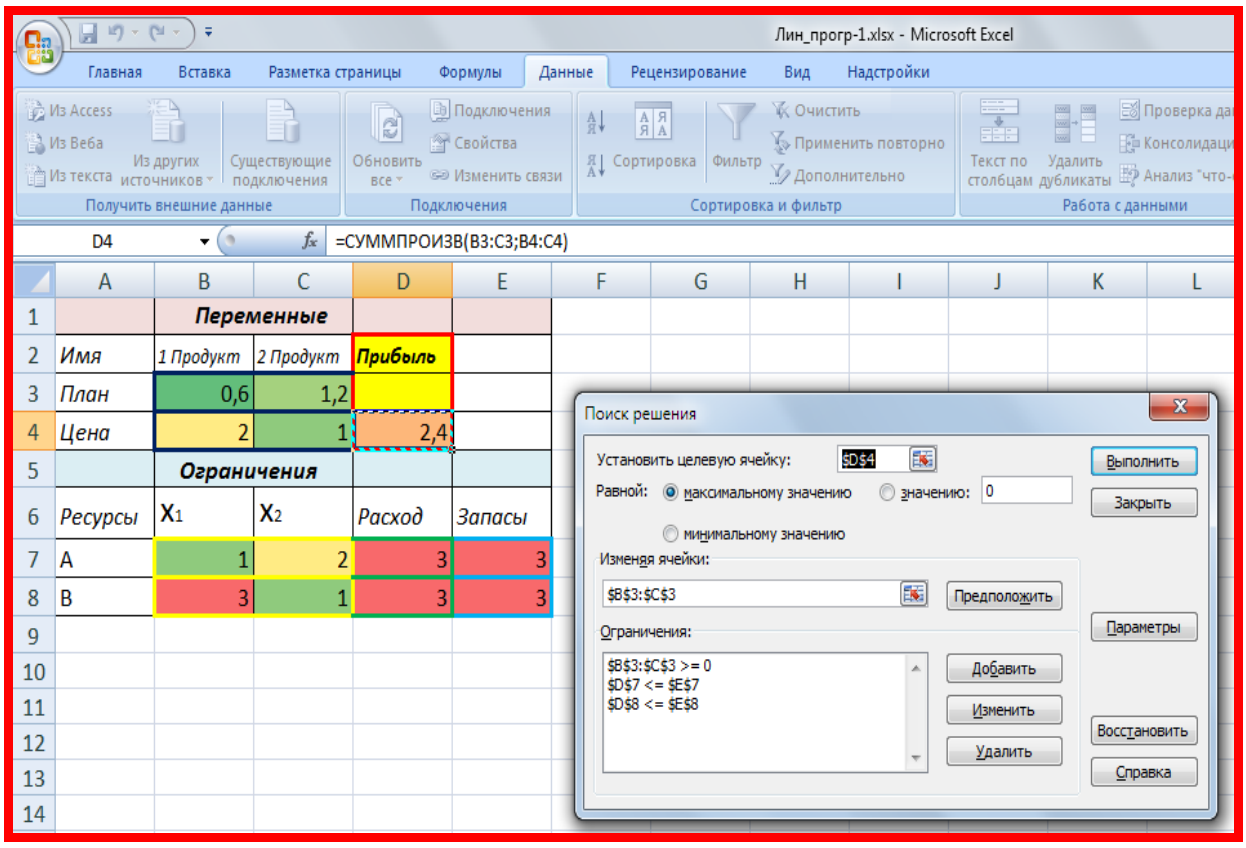


	A	B	C	D	E	F	G	H
1		Переменные						
2	Имя	1 Продукт	2 Продукт	Доход				
3	План							
4	Цена	2	1	0				
5		Ограничения						
6	Ресурсы	X_1	X_2	Расход	Запасы			
7	A	1	2	0	3			
8	B	3	1	0	3			
9								

2.2. Использование надстройки Excel «Поиск решения».

Надстройка Excel *«Поиск решения»* при первом использовании должна быть предварительно активирована. Открыв Excel, нажать кнопки «Office» → «Параметры Excel» → «Надстройки» → «Неактивные надстройки приложений» → выделить строку *«Поиск решения»* → «Управление: надстройки Excel» → «перейти» → ОК.

Щёлкнув на ленте кнопку «Данные», затем «Поиск решений» откроем окно «Поиск решений».



	A	B	C	D	E
1		Переменные			
2	Имя	1 Продукт	2 Продукт	Прибыль	
3	План	0,6	1,2		
4	Цена	2	1	2,4	
5		Ограничения			
6	Ресурсы	X_1	X_2	Расход	Запасы
7	A	1	2	3	3
8	B	3	1	3	3
9					
10					
11					
12					
13					
14					

-В поле «Установить целевую ячейку» ввести адрес целевой ячейки D4, щёлкнув по ней курсором мыши.

-Выбрать «равной максимальному значению».

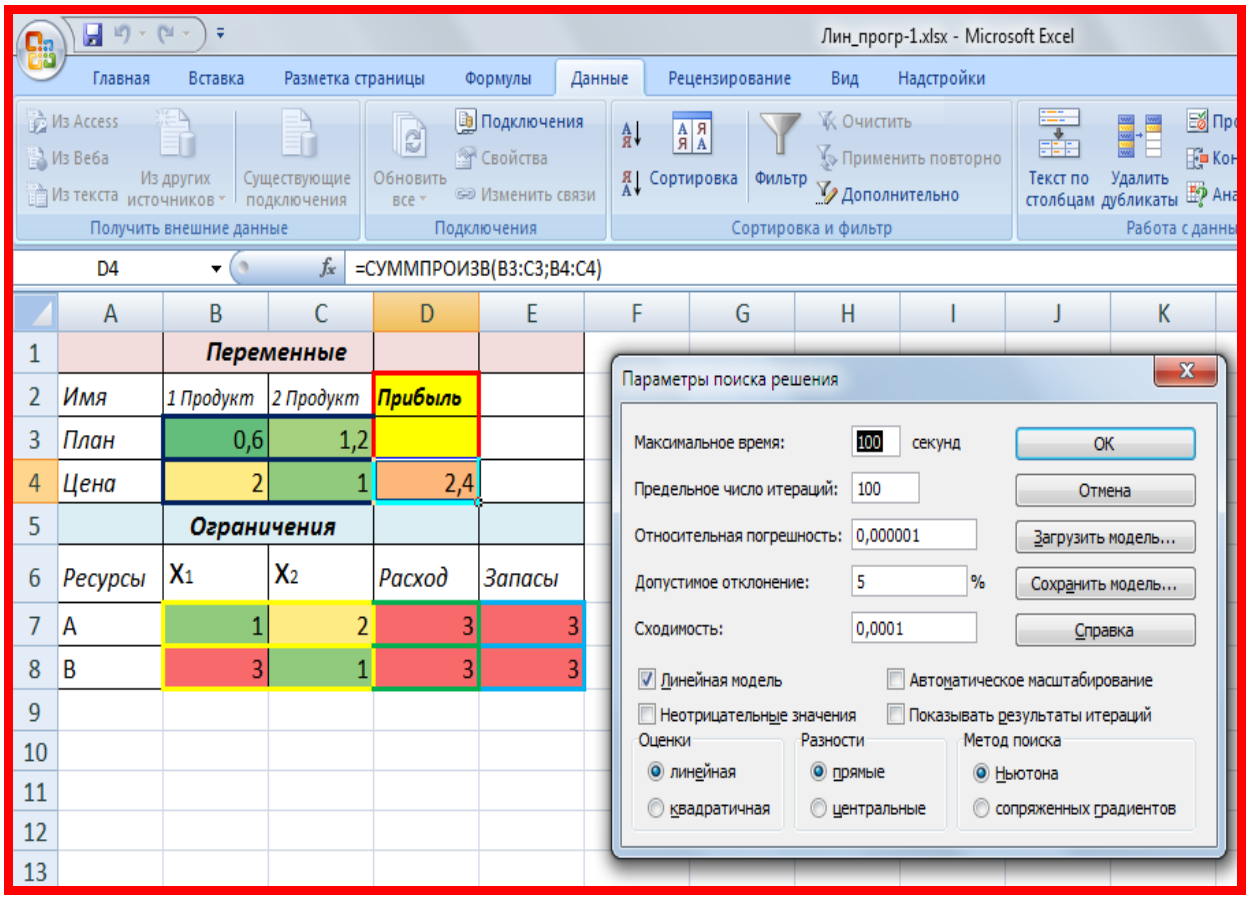
-В поле «изменяя ячейки» указать адреса B3:C3.

-В поле «Ограничения» щёлкнуть «Добавить». После появления поля «Добавление ограничения» в поле «Ссылка на ячейку:» сделать ссылку на ячейку D7, выбрать знак <=, в поле «Ограничение:» ввести адрес ячейки с запасом ресурса A- E7. Вновь выбрать

«Добавить» провести ввод ограничения по ресурсу В, затем по ограничению $x_1, x_2 \geq 0$. После этого нажать ОК.

2.3. Настройка параметров решения задачи.

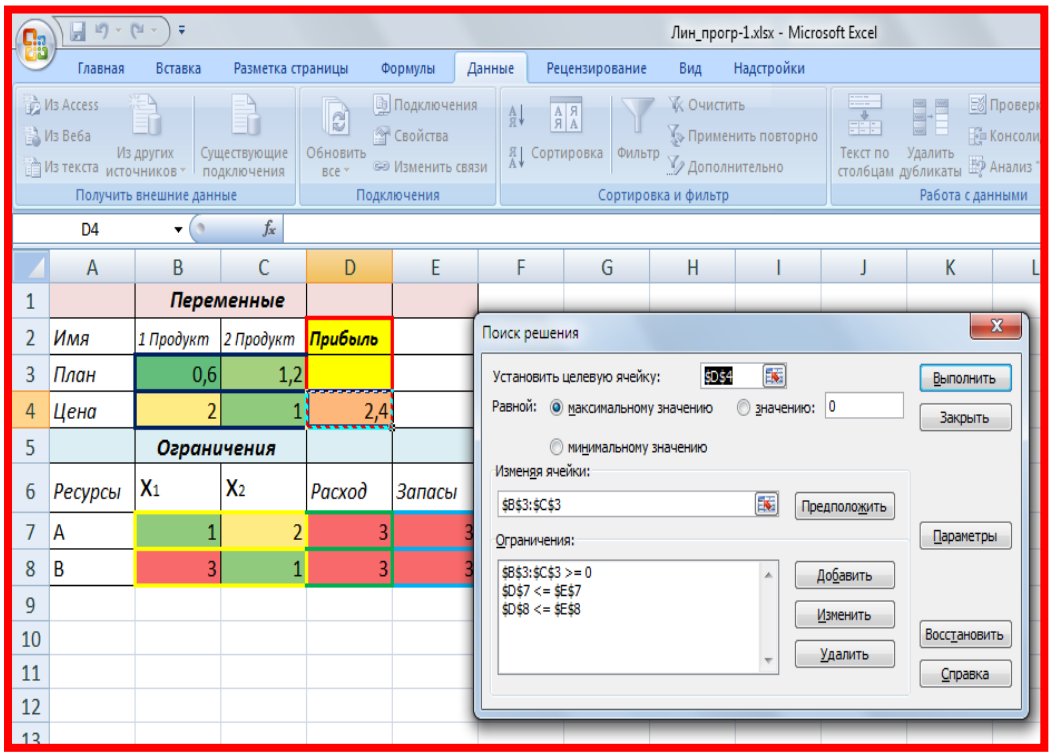
Выбрав в окне «Поиск решений» опцию «Параметры» в появившемся окне «Параметры поиска решения»установить флажок в поле «Линейная модель». При таком выборе при решении задачи будет использоваться симплекс-метод. Остальные значения можно оставить без изменения. Нажать ОК.

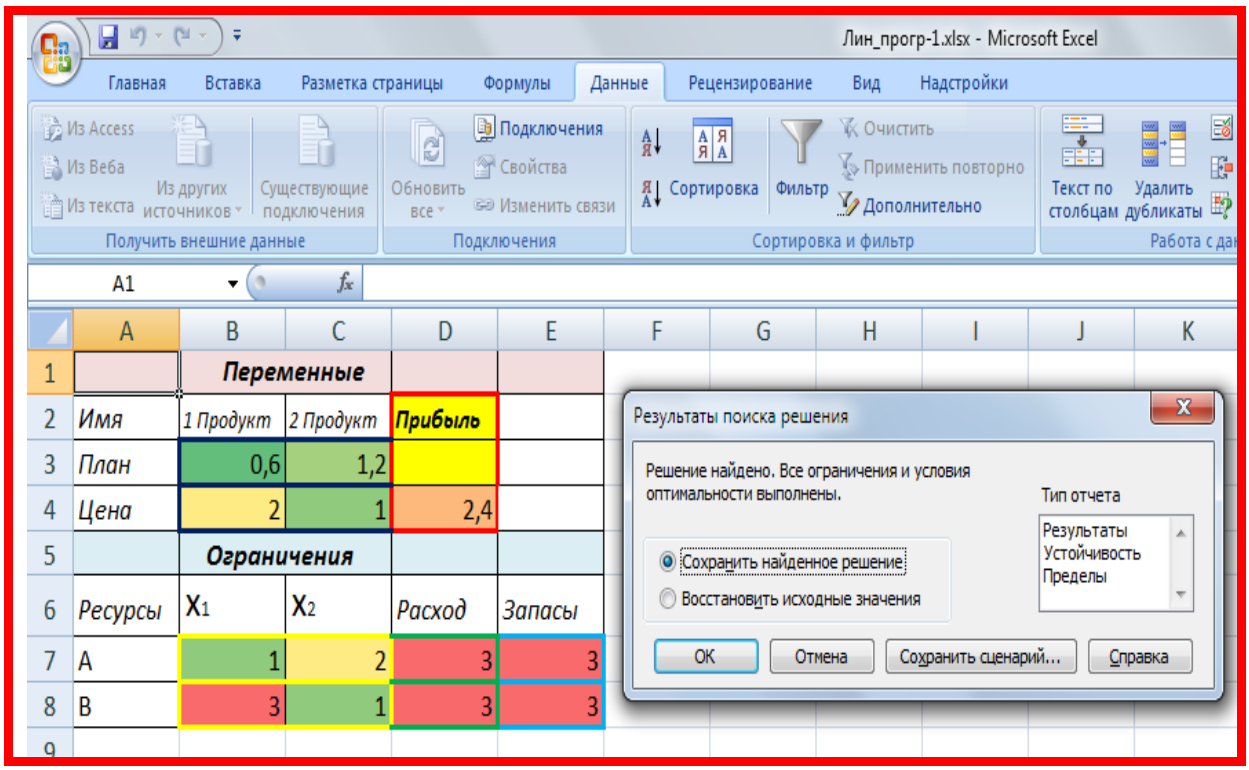


2.3. Завершение решения задачи и просмотр результатов.

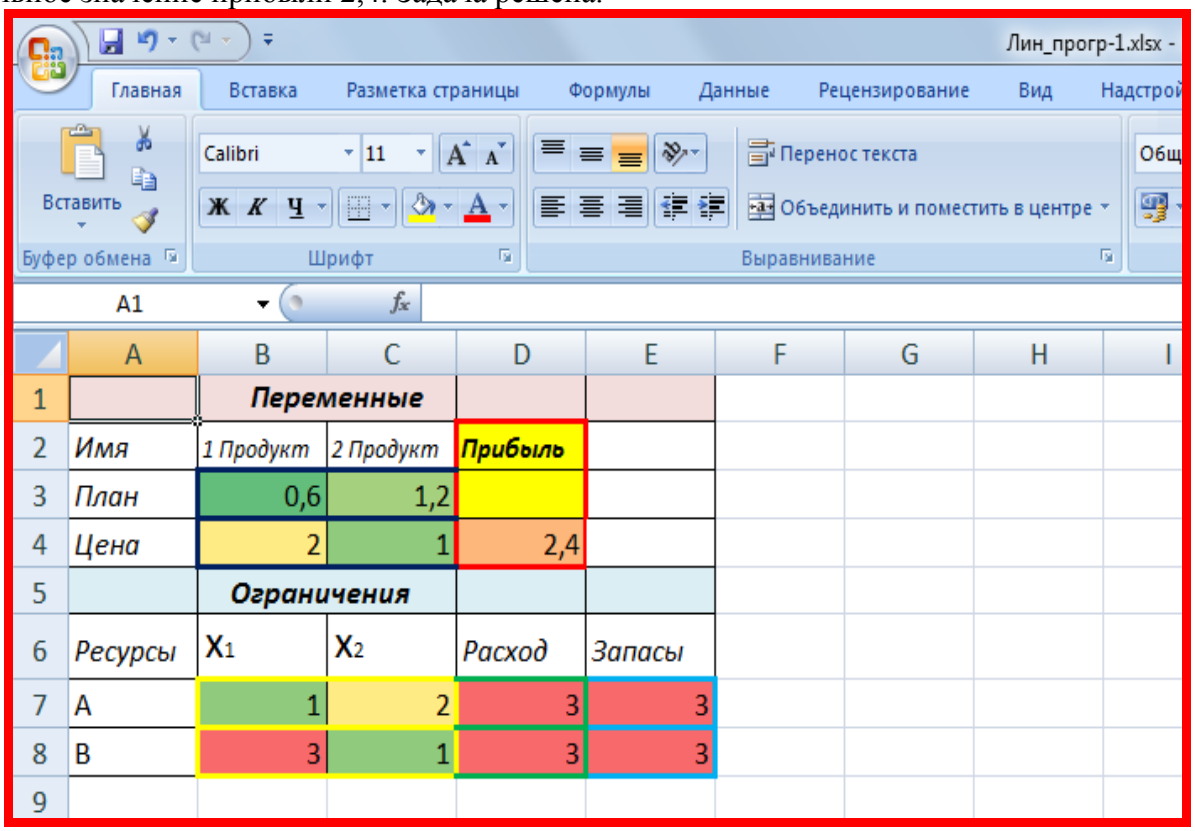
В окне «Поиск решений» нажимаем кнопку «Выполнить». Появляется окно «Результаты поиска решения». Можно выбрать тип отчёта, сохранить найденное решение или восстановить исходные значения,

ОК.





В ячейках B3 и C3 появятся оптимальные значения плана 0,6 и 1,2, а в ячейке D4 оптимальное значение прибыли 2,4. Задача решена.



«Решения типовой задачи линейного программирования с помощью симплекс-метода в Excel: задача о выборе оптимального рациона кормления животных»

Задача. Составляется комбинированный корм из трёх злаков: кукурузы, овса и ржи. Калорийность и содержание витамина С в одном килограмме каждого злака, а также цена одного кг каждого злака указаны в таблице:

	Кукуруза	Овёс	Рожь
Калорийность(ккал)	200	175	100
Содержание С (в Гр)	5	1	3
Цена (руб.)	6	4	1

Составить наиболее дешёвый комбинированный корм, 1кг которого содержал бы не менее 125 ккал и не менее 2г витамина С.

Решение.

1. Составление математической модели задачи. Обозначим содержание кукурузы, овса и ржи в 1кг комбикорма через x_1, x_2, x_3 соответственно, а стоимость злаков в комбикорме Z. Тогда математическая модель задачи об оптимизации производства комбикорма формулируется следующим образом: найти вектор $x=(x_1;x_2;x_3)$ (называемый планом), доставляющий минимум целевой функции задачи (функции затрат)

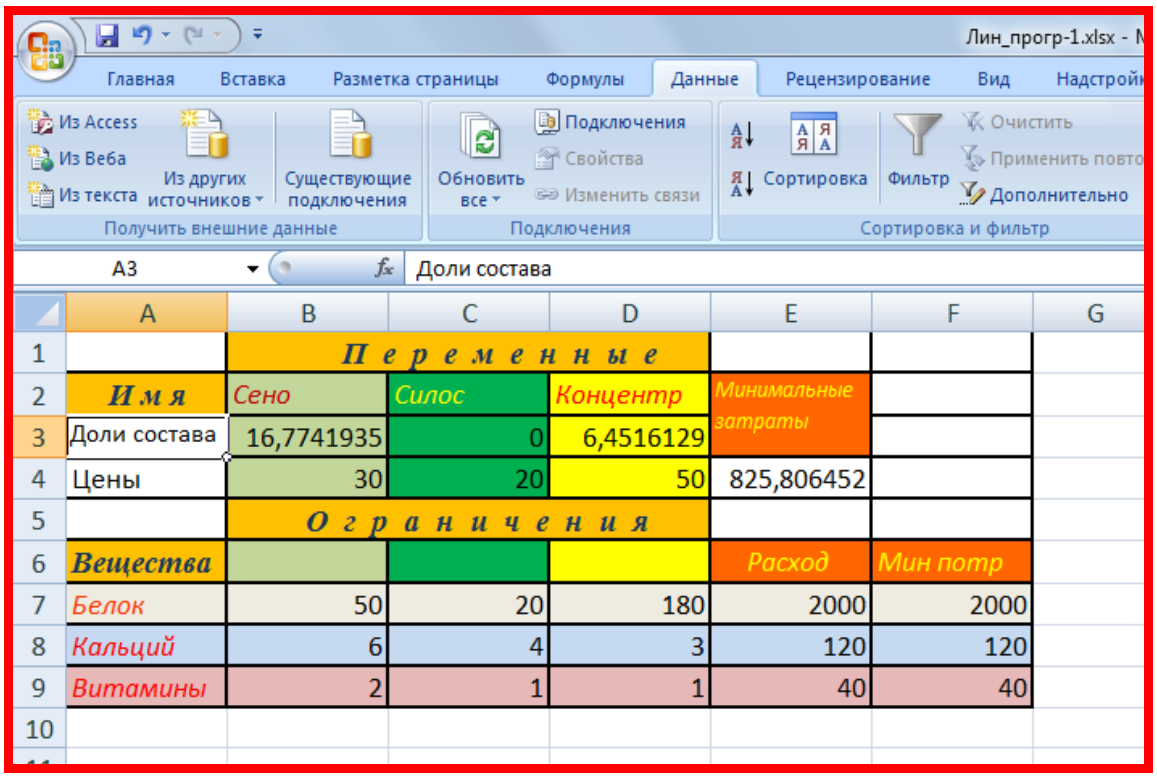
$$Z=6\cdot x_1+4\cdot x_2+x_3\to\min,$$

координаты которого x_1, x_2, x_3 удовлетворяют системе ограничений

$$\begin{cases}200\cdot x_1+175\cdot x_2+100\cdot x_3\ge 125,\\ 5\cdot x_1+x_2+3\cdot x_3\ge 2,\\ x_1+x_2+x_3=1,\end{cases}$$

$$x_1\ge 0.\ x_2\ge 0.\ x_3\ge 0.$$

2. Решение задачи в Excel.



	A	B	C	D	E	F	G
1		Переменные					
2	Имя	Сено	Силос	Концентр	Минимальные затраты		
3	Доли состава	16,7741935	0	6,4516129			
4	Цены	30	20	50	825,806452		
5		Ограничения					
6	Вещества				Расход	Мин потр	
7	Белок	50	20	180	2000	2000	
8	Кальций	6	4	3	120	120	
9	Витамины	2	1	1	40	40	
10							

Первичное распределение поставок

1. Метод северо-западного угла
2. Метод учета наименьшей поставки.
3. Циклы, оценки циклов, оценки клеток.

В регионе расположено несколько НГДУ, обеспечивающих определённые объёмы добычи нефти, которая поступает в НПЗ, расположенные в различных регионах страны и имеющие различные производственные мощности. В силу разноудалённости потребителей от НГДУ затраты на транспортировку нефти различаются.

В задаче необходимо составить план закрепления поставщиков за потребителями, который учитывает, по возможности, наиболее полное удовлетворение потребителей НПЗ и при этом обеспечивает минимальные затраты на транспортировку нефти.

Введены условные обозначения:

i – индекс НГДУ, i=1,m

m – общее число НГДУ в регионе

j – индекс НПЗ, j=1,n

n – общее число НПЗ.

Известно:

a_i - объёмы добычи нефти в i-ом НГДУ, тыс.т.;

b_j - потребность j-го НПЗ в нефти, тыс.т.;

c_{ij} - издержки на транспортировку 1000 т. нефти, тыс. руб.

b_j a_i	180	190	110	210	200	120
490	5	7	8	4	6	9
270	7	2	5	8	6	7
380	5	4	7	6	9	8

Модель задачи. В качестве неизвестных задачи принимаются переменные x_{ij}, означающие объём перевозок нефти i-го НГДУ к j-му НПЗ. В качестве коэффициентов целевой функции выступают издержки на перевозку 1000 т. нефти. Целевая функция минимизируется. Модель задачи записывается в общем виде, при этом необходимо учесть, что по исходным данным задача является открытой.

Имеем транспортную задачу с избытком запасов:

$\sum ai > \sum bj$ (где i=1..m ; j=1..n).

490+270+380>180+190+110+210+200+120

1140>1010 C max = 150;

Требуется найти такой план перевозок (X), при котором все заявки будут выполнены, а общая стоимость перевозок минимальна. Очевидно, при этой постановке задачи некоторые условия-равенства транспортной задачи превращаются в условия-неравенства, а некоторые — остаются равенствами.

$$\sum_{j=1}^{n} X_{i,j} \; ? \; a_i \; (i=1, \ldots, m);$$

$$\sum_{i=1}^{m} Xi,j = bj \; (j=1, \ldots, n).$$

Мы получаем следующую задачу:

$x_{11}+x_{12}+x_{13}+x_{14}+x_{15}+x_{16}$? 490

$x_{21}+x_{22}+x_{23}+x_{24}+x_{25}+x_{26}$? 270

$x_{31}+x_{32}+x_{33}+x_{34}+x_{35}+x_{36}$? 380

$x_{11}+x_{21}+x_{31} = 180$

$x_{13}+x_{23}+x_{33} = 190$

$x_{14}+x_{24}+x_{34} = 110$

$x_{12}+x_{22}+x_{32} = 210$

$x_{15}+x_{25}+x_{35} = 200$

$x_{16}+x_{26}+x_{36} = 120$

$x_{ij} \geq 0$ для i = 1,2,3; j = 1,2,3,4,5,6;

$K_{min}=5x_{11}+7x_{12}+8x_{13}+4x_{14}+6x_{15}+9x_{16}+7x_{21}+2x_{22}+5x_{23}+8x_{24}+6x_{25}+7x_{26}+5x_{31}+4x_{32}+7x_{33}+$
$+6x_{34}+9x_{35}+8x_{36}$;

Решение задачи.

Данную транспортную задачу необходимо решить *методом потенциалов*. Поскольку по исходным данным имеем открытую задачу, то до начала её решения следует получить закрытую модель.

Для этого, сверх имеющихся n пунктов назначения B_1, B_2, ... , B_n, введём ещё один, фиктивный, пункт назначения B_{n+1}, которому припишем фиктивную заявку, равную избытку запасов над заявками

$и_{т+1} = \sum a_ш - \sum$ и (где ш=1бююююбь ж о=1бююютбт) б

$b_7 = 1140 – 1010= 130$,

а стоимость перевозок из всех пунктов отправления в фиктивный пункт назначения b_7 будем считать равным нулю. Введением фиктивного пункта

назначения B_{n+1} с его заявкой b_{n+1} мы сравняли баланс транспортной задачи и теперь его можно решать как обычную транспортную задачу с правильным балансом.

Первоначальный опорный план поставок построим на основе *метода северо-западного угла:*

$b_j a_i$	180	190	110	210	200	120	130
490	5 **180**	7 **190**	8 **110**	4 **10**	6	9	0
270	7	2	5	8 **200**	6 **70**	7	0
380	5	4	7	6	9 **130**	8 **120**	0 **130**

Стоимость перевозок по данному плану составляет: 7300 тыс. руб.

Решим задачу с применением *метода потенциалов*.

Для этого плана можно определить платежи (α_i и β_j), так, чтобы в каждой базисной клетке выполнялось условие : $\alpha_i + \beta_j = \mathbf{c}_{i,j}$ (*)

Уравнений (*) всего m + n - 1, а число неизвестных равно m + n. Следовательно, одну из этих неизвестных можно задать произвольно (например, равной нулю). После этого из m + n - 1 уравнений (*) можно найти остальные платежи α_i , β_j , а по ним вычислить псевдостоимости: $u_{i,j}= \alpha_i + \beta_j$ для каждой свободной клетки.

Если оказалось, что все эти псевдостоимости не превосходят стоимостей $u_{i,j}$ □ $\mathbf{c}_{i,j}$,

то план потенциален и, значит, оптимален. Если же хотя бы в одной свободной клетке псевдостоимость больше стоимости (как в нашем примере), то план не является оптимальным и может быть улучшен переносом перевозок по циклу, соответствующему данной свободной клетке. Цена этого цикла ровна разности между стоимостью и псевдостоимостью в этой свободной клетке.

$_j a_i$	180	190	110	210	200	120	130	α_i
90	**80**	**90**	**10**	**0**			7	
70		1	2	**00**	**0**		3	
80	2	4	5	1	**30**	**20**	**30**	
$_i$	5	7	8	4	2	1	-7	

Мы получили в семи клетках $u_{i,j} - c_{i,j}$, теперь можно построить цикл в любой из этих клеток. Выгоднее всего строить цикл в той клетке, в которой разность $u_{i,j} - c_{i,j}$ максимальна. В нашем случае для построения цикла берем клетку (3,2):

$_ja_i$	18 0	19 0	11 0	21 0	2 00	1 20	1 30	$_i$
90	**80**	130 **90**	**10**	130 **0**			7	
70		1	2	130 **00**	130 **0**		3	
80	2	130 4	5	1	130 **30**	**20**	**30**	
$_i$	5	7	8	4	2	1	- 7	

Теперь будем перемещать по циклу число 130, так как оно является минимальным из чисел, стоящих в клетках, помеченных знаком -. При перемещении мы будем вычитать 130 из клеток со знаком - и прибавлять к клеткам со знаком + .

После этого необходимо подсчитать потенциалы α_i и β_j и цикл расчетов повторяется:

Стоимость перевозок по данному плану составляет: 6000 тыс. руб.

$_ja_i$	180	190	110	210	200	120	130	α_i
90	**80**	60 **0**	**10**	60 **40**		1		
70		1	60 2	60 **0**	**00**	5		
80		**30**			1	**20**	**30**	3
$_i$	5	7	8	4	2	11	3	

$_ja_i$	180	190	110	210	200	120	130	α_i
90	**80**	2	10 **10**	10 **00**			6	
70		**0**	10 2	10 **0**	**00**		2	
80	1	**30**	4	0		**20**	**30**	
$_i$	5	-2	8	4	2	2	-6	

Стоимость перевозок по данному плану составляет: 5460 тыс. руб.

$_j a_i$	1 80	1 90	1 10	2 10	2 00	1 20	1 30	α_i
90	**80**		100 **00**	**10**	100			0
70		**0**	100 **0**		100 **00**		2	-3
80		**30**				**20**	**30**	-1
$_i$	5	5	8	4	9	9	1	

Стоимость перевозок по данному плану составляет: 5390 тыс. руб.

$_j a_i$	1 80	1 90	1 10	2 10	2 00	1 20	1 30	α_i
90	100 **80**			**10**	100 **00**		2	0
70		100 **0**	**10**		100 **00**		2	0
80	100	100 **30**				**20**	**30**	2
$_i$	5	2	5	4	6	6	-2	

Стоимость перевозок по данному плану составляет: 5090 тыс. руб.

$_j a_i$	1 80	1 90	1 10	2 10	2 00	1 20	1 30	α_i
90	**0**			**10**	**00**			0
70		**60**	**10**				2	-2
80	**00**	**0**				**20**	**30**	0
$_i$	5	4	7	4	6	8	0	

Стоимость перевозок по данному плану составляет: 4890 тыс. руб. Псевдостоимости $u_{i,j} = \alpha_i + \beta_j$ для всех свободных клеток не превышают стоимостей, план оптимален. K_{min}=4890

Ответ: план закрепления поставщиков за потребителями, который учитывает, по возможности, наиболее полное удовлетворение потребителей НПЗ и при этом обеспечивает минимальные затраты на транспортировку нефти представлен ниже (Стоимость перевозок по данному плану составляет 4890 тыс. руб.):

$_jа_i$	180	190	110	210	200	120	130
90	5 **80**	7	8	4 **210**	6 **200**	9	0
70	7	2 **160**	5 **110**	8	6	7	0
80	5 **100**	4 **30**	7	6	9	8 **120**	0 **130**

Используя данные предыдущей задачи, решить транспортную задачу, построив первоначальный опорный план поставок *методом минимальной стоимости*.

Решение задачи.

Первоначальный опорный план поставок построим на основе *метода минимальной стоимости*.

$_jа_i$	180	190	110	210	200	120	130
90				**10**	**50**		**30**
70		**90**	**0**				
80	**80**		**0**		**0**	**20**	

Стоимость перевозок по данному плану составляет: 5040 тыс. руб.

Применяем *метод потенциалов*.

$_jа_i$	180	190	110	210	200	120	130	α_i
90				**10**	50 **50**		50 **30**	
70		**90**	**0**					
80	**80**		**0**		50 **0**	**20**	50	
$_i$	2	1	4	4	6	5	0	

$_j a_i$	180	190	110	210	200	120	130	α_i
90				**10**	**00**		**0**	0
70		**90**	**0**				2	-2
80	**80**		**0**			**20**	**0**	0
$_i$	5	4	7	4	6	8	0	

Стоимость перевозок по данному плану составляет: 4890 тыс. руб. Псевдостоимости $u_{i,j} = \alpha_i + \beta_j$ для всех свободных клеток не превышают стоимостей, план оптимален (стоимость совпадает с полученной стоимостью задачи №3, но план перевозок альтернативен).

Ответ: план закрепления поставщиков за потребителями, который учитывает, по возможности, наиболее полное удовлетворение потребителей НПЗ и при этом обеспечивает минимальные затраты на транспортировку нефти представлен ниже (Стоимость перевозок по данному плану составляет 4890 тыс. руб.):

$_j a_i$	180	190	110	210	200	120	130
90				**10**	**00**		**0**
70		**90**	**0**				
80	**80**		**0**			**20**	**0**

Оптимальное распределение поставок

1. Перераспределение поставок в цикле.
2. Методика получения оптимального распределения поставок.
3. Программное обеспечение решения задач транспортного типа на ПК.

Задача об оптимизации перевозок. Четыре отделения сельхозпредприятия B_1, B_2, B_3, B_4 закупают корма у трёх поставщиков A_1, A_2, A_3. Запасы кормов у поставщиков, потребности сельхозпредприятия в кормах и стоимость перевозки единицы продукта от поставщика к потребителю даны в таблице.

Значительную часть расходов с /х предприятия составляют именно транспортные расходы. Минимизировать расходы предприятия: составить такой план перевозок, при котором суммарные транспортные расходы будут минимальными, все запасы поставщиков будут вывезены, все потребности отделений с.\ х. предприятия будут удовлетворены.

Потре-бители / *Поставщики*	*Потребители*				*Запасы кормов у поставщиков*
	B_1	B_2	B_3	B_4	
A_1	10 x_{11}	0 x_{12}	20 x_{13}	11 x_{14}	$a_1 = 15$
A_2	12 x_{21}	7 x_{22}	9 x_{23}	20 x_{24}	$a_2 = 25$
A_3	0 x_{31}	14 x_{32}	16 x_{33}	18 x_{34}	$a_3 = 5$
	$b_1 =$ 5	$b_2 =$ 15	$b_3 =$ 15	$b_4 =$ 10	45
	Потребность в кормах				

1. Математическая модель задачи. Обозначим через x_{ij} количество кормов, перевозимое от поставщика A_i к потребителю B_j, c_{ij} - стоимость перевозок по этому маршруту, $i = 1,2,3;\ j = 1,2,3,4$.

Целевая функция: транспортные расходы на перевозку кормов вычисляются по формуле

$$Z = \sum_{i=1}^{3}\sum_{j=1}^{4} c_{ij} \cdot x_{ij} = 10 \cdot x_{11} + 0 \cdot x_{12} + 20 \cdot x_{13} + 11 \cdot x_{14} +$$

$$+ 12 \cdot x_{21} + 7 \cdot x_{22} + 9 \cdot x_{23} + 20 \cdot x_{24} +$$

$$+ 0 \cdot x_{31} + 14 \cdot x_{32} + 16 \cdot x_{33} + 18 \cdot x_{34} \rightarrow \min.$$

Ограничения на переменные задачи.

Ограничения вывоза: из A_1 $x_{11} + x_{12} + x_{13} + x_{14} = 15$,

из A_2 $x_{21} + x_{22} + x_{13} + x_{14} = 25$,

из A_3 $x_{31} + x_{32} + x_{33} + x_{34} = 5$.

Ограничения ввоза: *ввоз в* B_1 $x_{11} + x_{21} + x_{31} = 5$,

ввоз в B_2 $x_{12} + x_{22} + x_{32} = 15$,

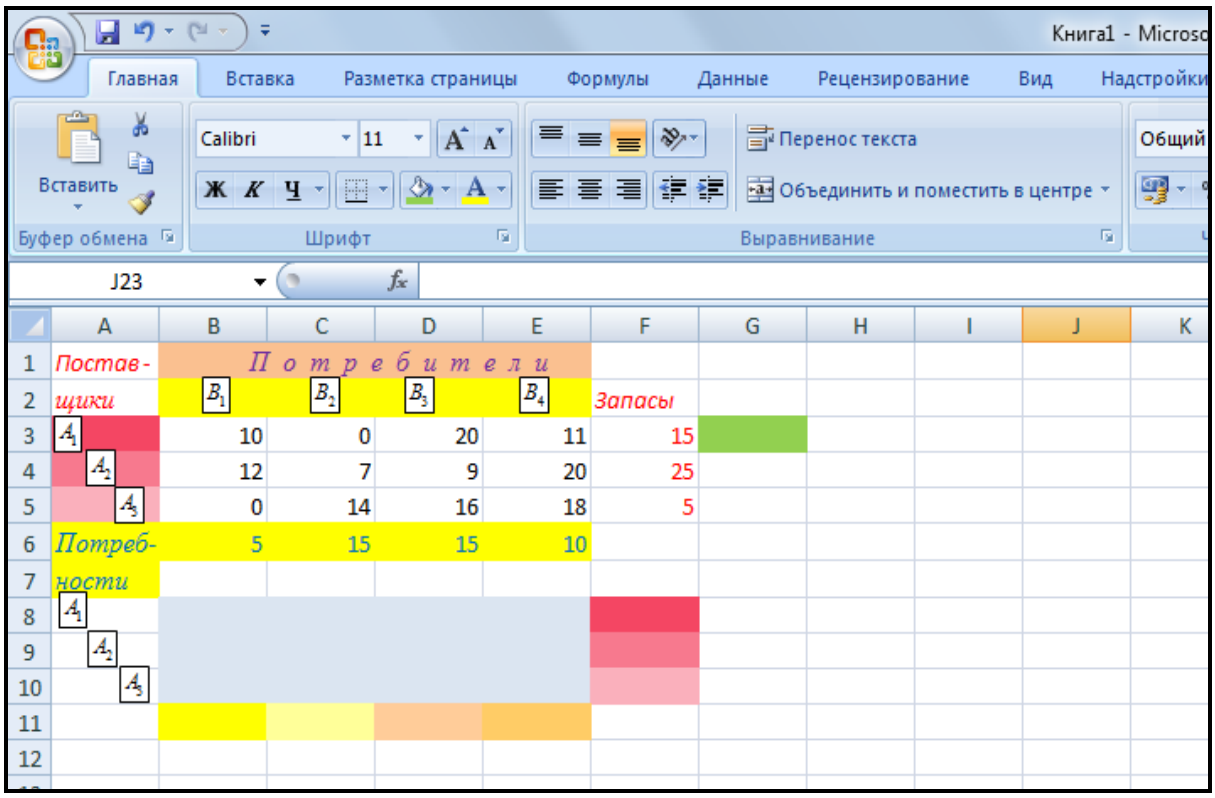
ввоз в B_3 $x_{13} + x_{23} + x_{33} = 15$,

ввоз в B_4 $x_{14} + x_{24} + x_{34} = 10$.

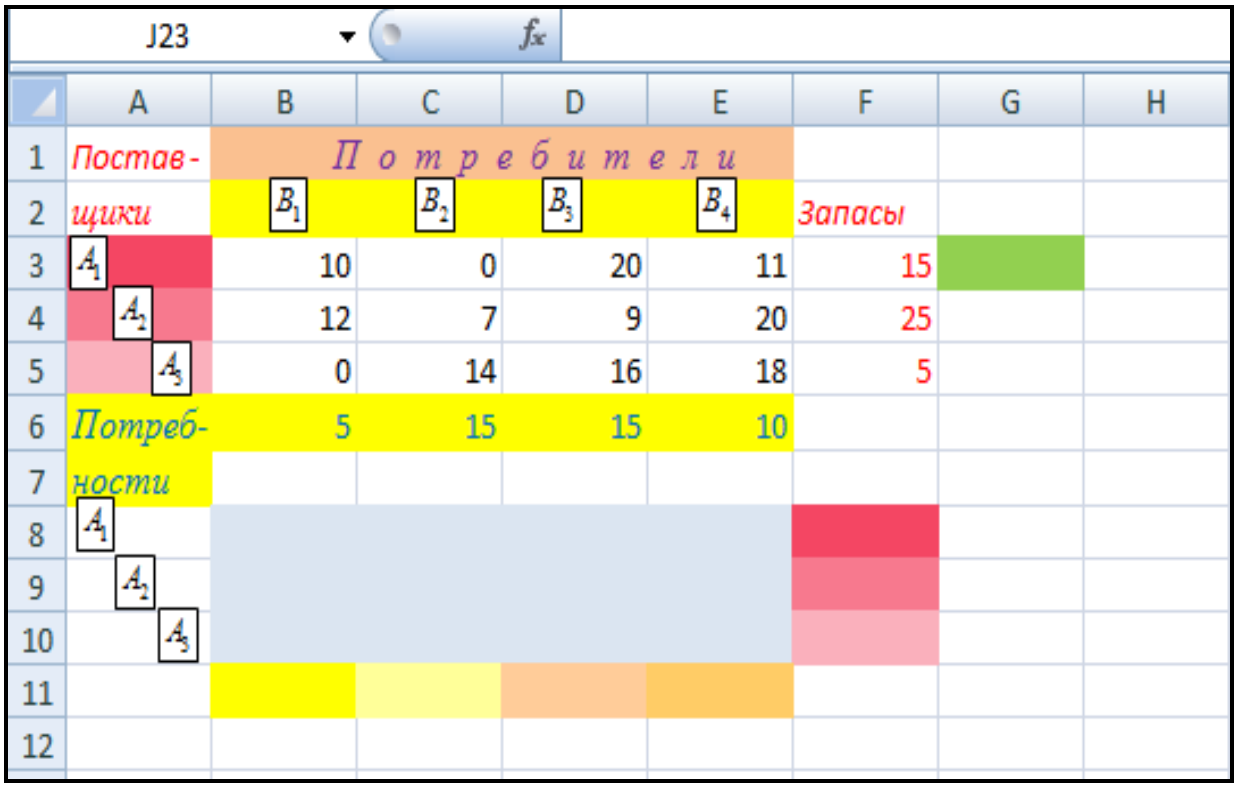
Ограничения на знаки(значения) переменных: $x_{ij} \geq 0$.

Задача закрытого типа : $a_1 + a_2 + a_3 = b_1 + b_2 + b_3 + b_4 = 45$.

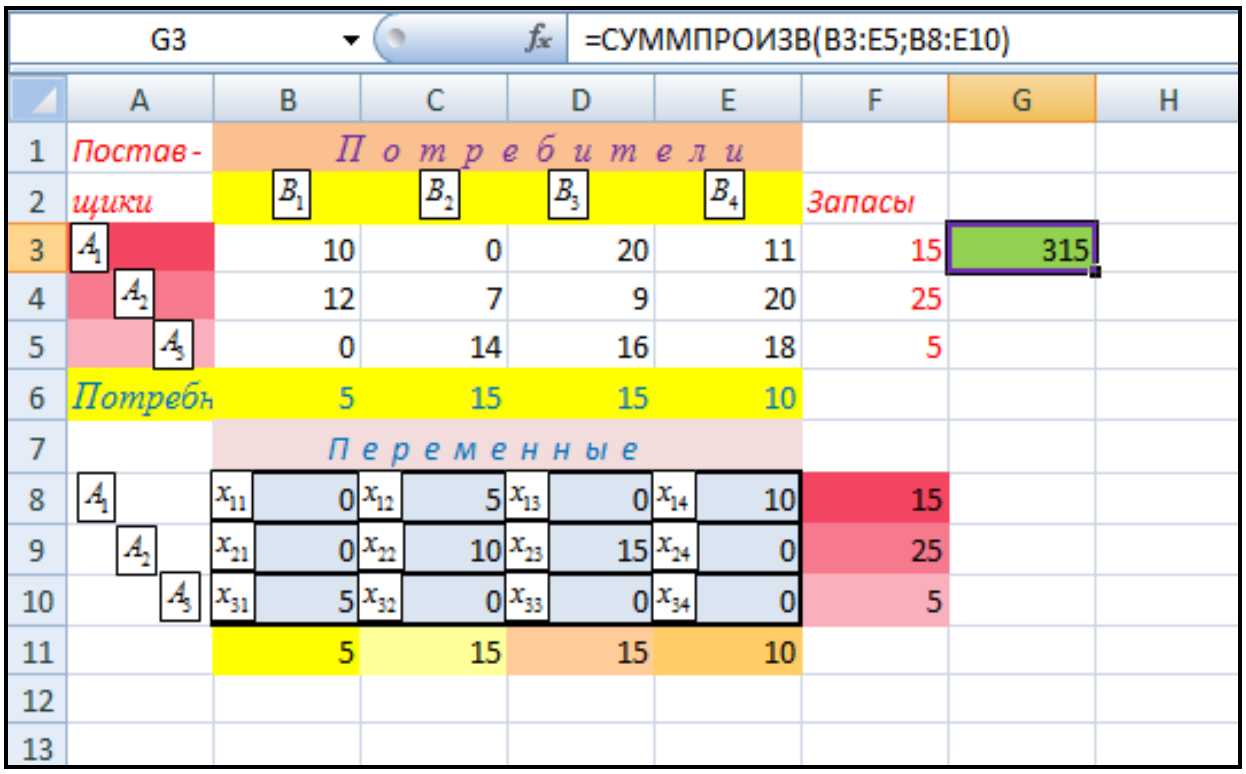
2. Решение задачи в Excel.



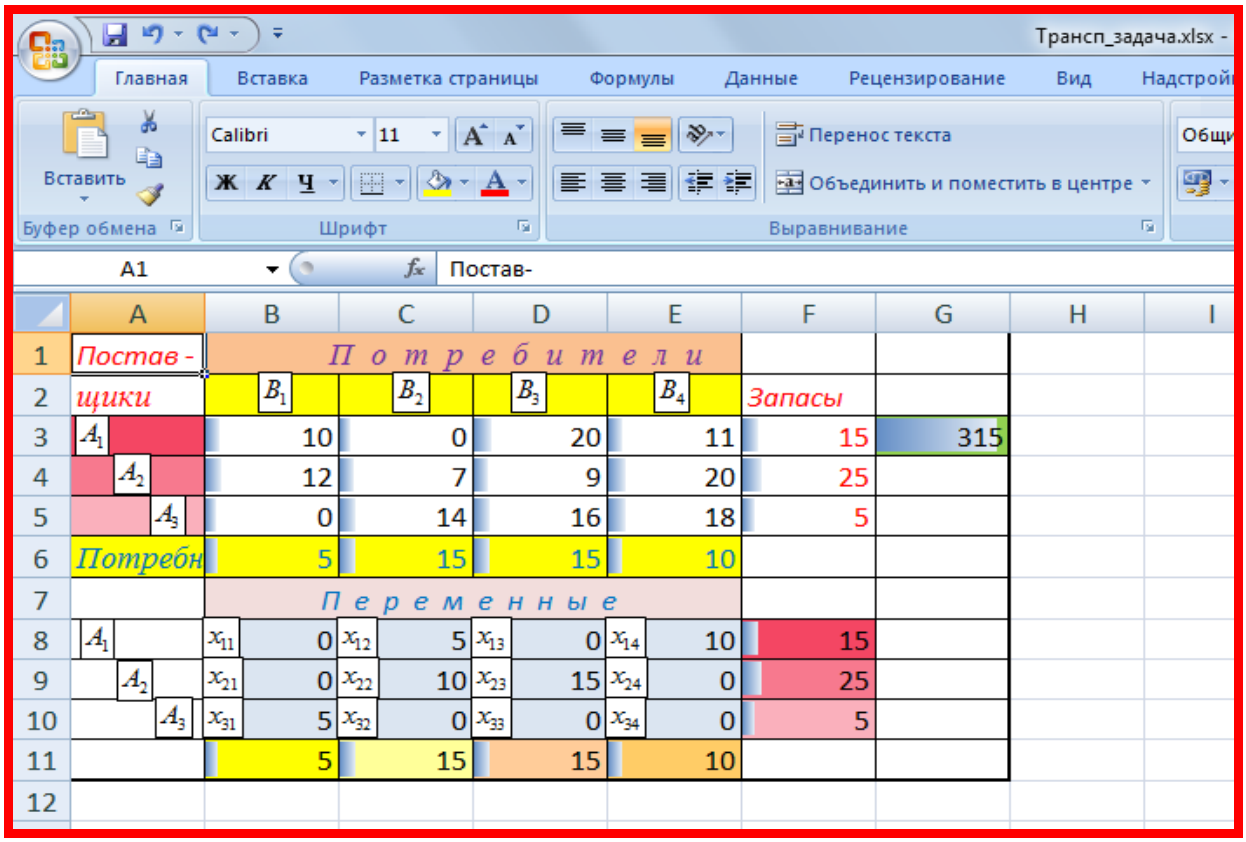
	A	B	C	D	E	F	G	H	I	J	K
1	Постав-	Потребители									
2	щики	B_1	B_2	B_3	B_4	Запасы					
3	A_1	10	0	20	11	15					
4	A_2	12	7	9	20	25					
5	A_3	0	14	16	18	5					
6	Потреб-	5	15	15	10						
7	ности										
8	A_1										
9	A_2										
10	A_3										
11											
12											



	A	B	C	D	E	F	G	H
1	Постав-	Потребители						
2	щики	B_1	B_2	B_3	B_4	Запасы		
3	A_1	10	0	20	11	15		
4	A_2	12	7	9	20	25		
5	A_3	0	14	16	18	5		
6	Потреб-	5	15	15	10			
7	ности							
8	A_1							
9	A_2							
10	A_3							
11								
12								



G3	=СУММПРОИЗВ(B3:E5;B8:E10)					
	A	B	C	D	E	F	G	H
1	Постав-	Потребители					
2	щики	B_1	B_2	B_3	B_4	Запасы	
3	A_1	10	0	20	11	15	315
4	A_2	12	7	9	20	25	
5	A_3	0	14	16	18	5	
6	Потребн	5	15	15	10		
7		Переменные					
8	A_1	x_{11} 0	x_{12} 5	x_{13} 0	x_{14} 10	15	
9	A_2	x_{21} 0	x_{22} 10	x_{23} 15	x_{24} 0	25	
10	A_3	x_{31} 5	x_{32} 0	x_{33} 0	x_{34} 0	5	
11		5	15	15	10		
12							
13							



	A	B	C	D	E	F	G	H	I
1	Постав-	Потребители							
2	щики	B_1	B_2	B_3	B_4	Запасы			
3	A_1	10	0	20	11	15	315		
4	A_2	12	7	9	20	25			
5	A_3	0	14	16	18	5			
6	Потребн	5	15	15	10				
7		Переменные							
8	A_1	x_{11} 0	x_{12} 5	x_{13} 0	x_{14} 10	15			
9	A_2	x_{21} 0	x_{22} 10	x_{23} 15	x_{24} 0	25			
10	A_3	x_{31} 5	x_{32} 0	x_{33} 0	x_{34} 0	5			
11		5	15	15	10				
12									

«Решение задачи о назначениях с Excel»

Задача о назначениях. Производится классификация 7 болезней животных (птицы) по 5 категориям. Результаты тестирования каждого заболевания уровень опасности заболевания) по каждой категории выражены в баллах по 10-балльной шкале и представлены матрицей

$C =$

*	P_1	P_2	P_3	P_4	P_5
S_1	7	5	7	6	7
S_2	6	4	8	4	9
S_3	8	6	4	3	8
S_4	7	7	8	5	7
S_5	5	9	7	9	5
S_6	6	8	6	4	7
S_7	7	7	8	6	4

Определить самые опасные заболевания в каждой из 5 категорий так, чтобы сумма баллов выбранных заболеваний была наибольшей (суммарный уровень опасности выбранных заболеваний был наибольшим). Каждое заболевание может быть самым опасным только в одной категории и все категории должны быть заняты.

Решение.

1. Математическая модель задачи. Обозначим через x_{ij} переменные задачи:

$x_{ij} = 1$, если заболевание S_i выбирается самым опасным в категории P_j;

$x_{ij} = 0$, если заболевание S_i не является самым опасным в категории P_j;

c_{ij} - уровень опасности заболевания S_i (количество баллов по результатам тестирования) в категории P_j,

$i = 1,2,3,4,5,6,7;\ j = 1,2,3,4,5.$

Целевая функция: суммарный уровень опасности всех заболеваний в баллах вычисляется по формуле

$$Z = \sum_{i=1}^{7}\sum_{j=1}^{5} c_{ij} \cdot x_{ij} = 7 \cdot x_{11} + 5 \cdot x_{12} + 7 \cdot x_{13} + 6 \cdot x_{14} + 7 \cdot x_{15} +$$
$$+ 6 \cdot x_{21} + 4 \cdot x_{22} + 8 \cdot x_{23} + 4 \cdot x_{24} + 9 \cdot x_{25} +$$
$$+ 8 \cdot x_{31} + 6 \cdot x_{32} + 4 \cdot x_{33} + 3 \cdot x_{34} + 8 \cdot x_{35} +$$
$$+ 7 \cdot x_{41} + 7 \cdot x_{42} + 8 \cdot x_{43} + 5 \cdot x_{44} + 7 \cdot x_{45} +$$
$$+ 5 \cdot x_{51} + 9 \cdot x_{52} + 7 \cdot x_{53} + 9 \cdot x_{54} + 5 \cdot x_{55} +$$
$$+ 6 \cdot x_{61} + 8 \cdot x_{62} + 6 \cdot x_{63} + 4 \cdot x_{64} + 7 \cdot x_{65} +$$
$$+ 7 \cdot x_{71} + 7 \cdot x_{72} + 8 \cdot x_{73} + 6 \cdot x_{74} + 4 \cdot x_{75} \to \max.$$

Ограничения на переменные задачи.

Ограничения на лидерство одного заболевания: каждое заболевание может быть самым опасным только в одной категории

$$\text{для } S_1 \quad x_{11} + x_{12} + x_{13} + x_{14} + x_{15} = 1,$$
$$\text{для } S_2 \quad x_{21} + x_{22} + x_{23} + x_{24} + x_{25} = 1,$$
$$\text{для } S_3 \quad x_{31} + x_{32} + x_{33} + x_{34} + x_{35} = 1,$$
$$\text{для } S_4 \quad x_{41} + x_{42} + x_{43} + x_{44} + x_{45} = 1,$$
$$\text{для } S_5 \quad x_{51} + x_{52} + x_{53} + x_{54} + x_{55} = 1,$$
$$\text{для } S_6 \quad x_{61} + x_{62} + x_{63} + x_{64} + x_{65} = 1,$$
$$\text{для } S_7 \quad x_{71} + x_{72} + x_{73} + x_{74} + x_{75} = 1.$$

Ограничения по занятости категорий: в каждой категории может быть только один лидер

$$\text{для } P_1 \quad x_{11}+x_{21}+x_{31}+x_{41}+x_{51}+x_{61}+x_{71}=1,$$
$$\text{для } P_2 \quad x_{11}+x_{21}+x_{31}+x_{41}+x_{51}+x_{61}+x_{71}=1,$$
$$\text{для } P_3 \quad x_{13}+x_{23}+x_{33}=15,$$
$$\text{для } P_4 \quad x_{14}+x_{24}+x_{34}=10,$$
$$\text{для } P_2 \quad x_{11}+x_{21}+x_{31}+x_{41}+x_{51}+x_{61}+x_{71}=1,$$

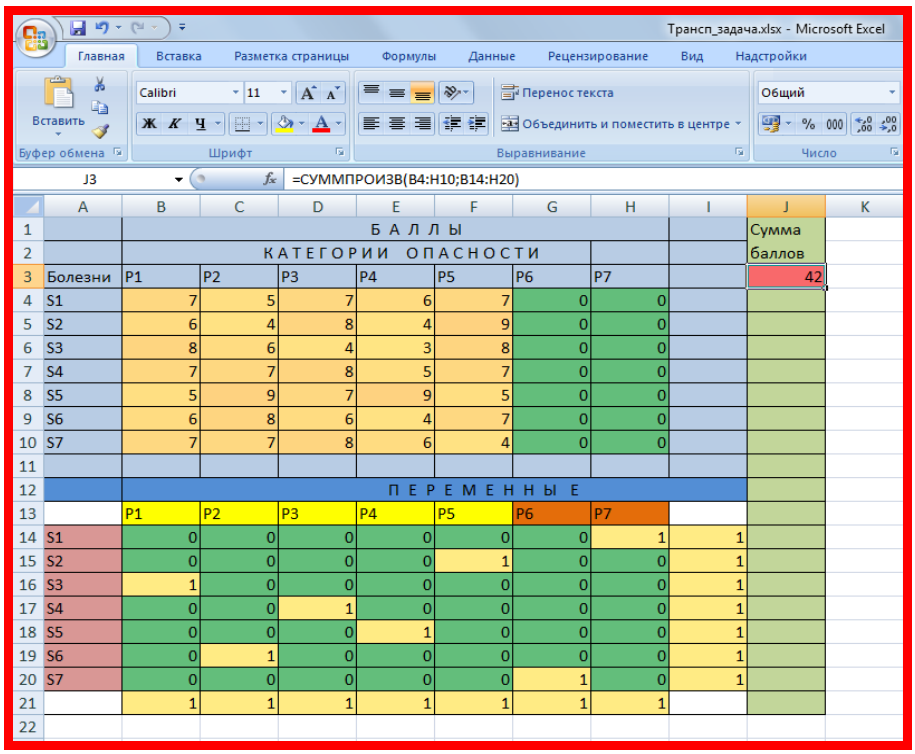
Ограничения на знаки(значения) переменных: $x_{ij} \geq 0$, x_{ij} - двоичные числа.

Задача открытого типа: вводятся две фиктивные категории с нулевыми столбцами баллов. Матрица C^* становится квадратной.

$C^* =$

*	P_1	P_2	P_3	P_4	P_5	P_6	P_7
S_1	7	5	7	6	7	0	0
S_2	6	4	8	4	9	0	0
S_3	8	6	4	3	8	0	0
S_4	7	7	8	5	7	0	0
S_5	5	9	7	9	5	0	0
S_6	6	8	6	4	7	0	0
S_7	7	7	8	6	4	0	0

2.7.3 Результаты и выводы: ***Решение задачи в Excel.***



	A	B	C	D	E	F	G	H	I	J	K
1		Б А Л Л Ы								Сумма	
2		К А Т Е Г О Р И И О П А С Н О С Т И								баллов	
3	Болезни	P1	P2	P3	P4	P5	P6	P7		42	
4	S1	7	5	7	6	7	0	0			
5	S2	6	4	8	4	9	0	0			
6	S3	8	6	4	3	8	0	0			
7	S4	7	7	8	5	7	0	0			
8	S5	5	9	7	9	5	0	0			
9	S6	6	8	6	4	7	0	0			
10	S7	7	7	8	6	4	0	0			
11											
12		П Е Р Е М Е Н Н Ы Е									
13		P1	P2	P3	P4	P5	P6	P7			
14	S1	0	0	0	0	0	0	1	1		
15	S2	0	0	0	0	1	0	0	1		
16	S3	1	0	0	0	0	0	0	1		
17	S4	0	0	1	0	0	0	0	1		
18	S5	0	0	0	1	0	0	0	1		
19	S6	0	1	0	0	0	0	0	1		
20	S7	0	0	0	0	0	1	0	1		
21		1	1	1	1	1	1	1			
22											

2.8 Практическое занятие № 18-19 (4 часа)

Тема: Основные понятия теории графов. Классификация графов, их свойства. Деревья, сети. Основы сетевого анализа

2.8.1 Задание для работы:

1. Основные понятия и определения теории графов.
2. Модели и алгоритмы сетевого анализа

2.8.2 Краткое описание проводимого занятия:

Примеры приложений теории графов

1. «Транспортные» задачи, в которых вершинами графа являются пункты, а ребрами - дороги (автомобильные, железные и др.) и / или другие транспортные (например, авиационные) маршруты. Другой пример - сети снабжения (энергоснабжения, газоснабжения, снабжения товарами и т.д.), в которых вершинами являются пункты производства и потребления, а ребрами - возможные маршруты перемещения (линии электропередач, газопроводы, дороги и т.д.). Соответствующий класс задач оптимизации потоков грузов, размещения пунктов производства и потребления и т.д., иногда называется задачами обеспечения или задачами о размещении. Их подклассом являются задачи о грузоперевозках.

2. «Технологические задачи», в которых вершины отражают производственные элементы (заводы, цеха, станки и т.д.), а дуги потоки сырья, материалов и продукции между ними, заключаются в определении оптимальной загрузки производственных элементов и обеспечивающих эту загрузку потоков.

3. Обменные схемы, являющиеся моделями таких явлений как бартер, взаимозачеты и т.д. Вершины графа при этом описывают участников обменной схемы (цепочки), а дуги - потоки материальных и финансовых ресурсов между ними. Задача заключается в определении цепочки обменов, оптимальной с точки зрения, например, организатора обмена и согласованной с интересами участников цепочки и существующими ограничениями

4. Управление проектами. (Управление проектами - раздел теории управления, изучающий методы и механизмы управления изменениями (проектом называется целенаправленное изменение некоторой системы, осуществляемое в рамках ограничений на время и используемые ресурсы; характерной чертой любого проекта является его уникальность, то есть нерегулярность соответствующих изменений.)). С точки зрения теории графов проект - совокупность операций и зависимостей между ними. Хрестоматийным примером является проект строительства некоторого объекта. Совокупность моделей и методов, использующих язык и результаты теории графов и ориентированных на решение задач управления проектами, получила название календарно-сетевого планирования и управления (КСПУ). В рамках КСПУ решаются задачи определения последовательности выполнения операций и распределения ресурсов между ними, оптимальных с точки зрения тех или иных критериев (времени проекта, затрат, и др.).

5. Модели коллективов и групп, используемые в социологии, основываются на представлении людей или их групп в виде вершин, а отношений между ними (например, отноше-

ний знакомства, доверия, симпатии и т.д.) - в виде ребер или дуг. В рамках подобного описания решаются задачи исследования структуры социальных групп, их сравнения, определения агрегированных показателей, отражающих степень напряженности, согласованности, взаимодействия и др.

6. Модели организационных структур, в которых вершинами являются элементы организационной системы, а ребрами или дугами - связи (информационные, управляющие, технологические и др.) между ними.

При исследовании, анализе и решении управленческих проблем, моделировании экономических объектов широко используются методы формализированного представления, являющегося предметом рассмотрения в дискретной математике. Широкое применение при решении таких задач получила теория графов благодаря своей наглядности и универсальности. Исследования показывают, что не менее 70% реальных задач математического программирования можно представить в виде сетевых моделей [3]. Приведём несколько конкретных примеров [1, 2].

1. Проектирование газопровода, соединяющего буровые скважины морского базирования, с находящейся на берегу приёмной станцией. Целевая функция соответствующей модели должна минимизировать стоимость строительства газопровода.
2. Поиск кратчайшего маршрута между городами по соответствующей сети дорог.
3. Определение максимальной пропускной способности трубопровода для транспортировки угольной пульпы от угольных шахт к электростанциям.
4. Определение схемы транспортировки нефти от пунктов нефтедобычи к нефтеперерабатывающим заводам с минимальной стоимостью транспортировки.
5. Составление временного графика строительных работ (определение дат начала и завершения отдельных этапов работ).

Чаще всего **метод дерева решений** используют в сложных, но поддающихся классификации задачах принятия решений, когда перед нами есть несколько альтернативных "решений" (проектов, выходов, стратегий), каждое из которых в зависимости от наших действий или действий других лиц (а также глобальных сил, вроде рынка, природы и т.п.) может давать разные последствия (результаты).

Задача состоит в том, чтобы правильно отобразить все возможные варианты развития ситуации (ветви дерева) и конечные результаты, вычислить некоторые показатели (например, ожидаемая прибыльность проекта, затраты и т.п.) и на основе полученных данных принять решение и выборе нужной линии поведения.

Принятие решений с помощью дерева возможных вариантов производится поэтапно:

1. **Построение дерева решений** (графа без циклов). Дерево строится по определенным правилам: вершины альтернативных решений, вершины событий, дуги решений, конечные решения - листья вводятся и обозначаются определенным образом в нужном порядке.

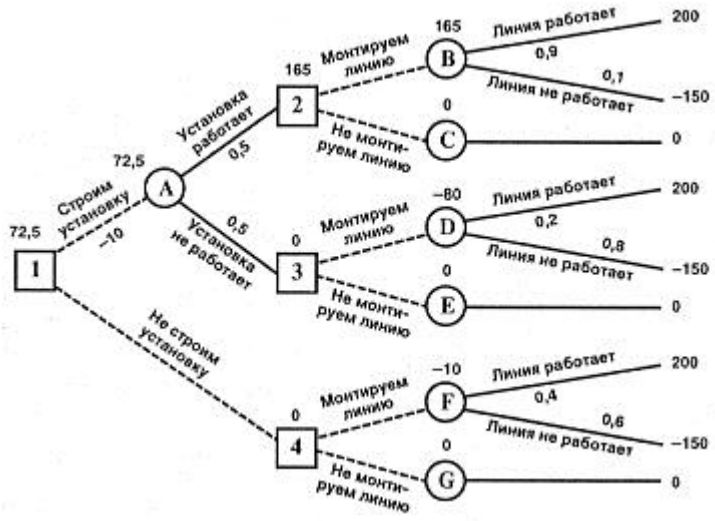
2. **Анализ дерева решений**: подсчет вероятностей и математических ожиданий (стоимостных оценок решения, EMV), расчет оптимистического и пессимистического прогноза, выбор оптимального решения.

Пример

Главному инженеру компании надо решить, монтировать или нет новую производственную линию, использующую новейшую технологию. Если новая линия будет работать безотказно, компания получит прибыль 200 млн. рублей. Если же она откажет, компания может потерять 150 млн. рублей. По оценкам главного инженера, существует 60% шансов, что новая производственная линия откажет. Можно создать экспериментальную установку, а затем уже решать, монтировать или нет производственную линию. Эксперимент обойдется в 10 млн. рублей.

Главный инженер считает, что существует 50% шансов, что экспериментальная установка будет работать. Если экспериментальная установка будет работать, то 90% шансов за то, что смонтированная производственная линия также будет работать. Если же экспериментальная установка не будет работать, то только 20% шансов за то, что производственная линия заработает. Следует ли строить экспериментальную установку? Следует ли монтировать производственную линию? Какова ожидаемая стоимостная оценка наилучшего решения?

Рисунок 1 (нажмите для увеличения).



В узле F возможны исходы «линия работает» с вероятностью 0,4 (что приносит прибыль 200) и «линия не работает» с вероятностью 0,6 (что приносит убыток — 150) => оценка узла F. EMV(F) = 0,4 x 200 + 0,6 x (-150) = -10. Это число мы пишем над узлом F.

EMV(G) = 0.

В узле 4 мы выбираем между решением «монтируем линию» (оценка этого решения EMV(F) = -10) и решением «не монтируем линию» (оценка этого решения EMV(G) = 0): EMV(4) = max {EMV(F), EMV(G)} = max {-10, 0} = 0 = EMV(G). Эту оценку мы пишем над узлом 4, а решение «монтируем линию» отбрасываем и зачеркиваем.

Аналогично:

EMV(B) = 0,9 x 200 + 0,1 x (-150) = 180 — 15 = 165.

EMV(C) = 0.

EMV(2) = max {EMV(B), EMV(C} = max {165, 0} = 165 = EMV(5). Поэтому в узле 2 отбрасываем возможное решение «не монтируем линию».

EMV(D) = 0,2 x 200 + 0,8 x (-150) = 40 — 120 = -80.

EMV(E) = 0.

EMV(3) = max {EMV(D), EMV(E)} = max {-80, 0} = 0 = EMV(E). Поэтому в узле 3 отбрасываем возможное решение «монтируем линию».

EMV(A) = 0,5 x 165 + 0,5 x 0 — 10 = 72,5.

EMV(l) = max {EMV(A), EMV(4)} = max {72,5; 0} = 72,5 = EMV(A). Поэтому в узле 1 отбрасываем возможное решение «не строим установку».

Ожидаемая стоимостная оценка наилучшего решения равна 72,5 млн. рублей. Строим установку. Если установка работает, то монтируем линию. Если установка не работает, то линию монтировать не надо.

Пример

Компания рассматривает вопрос о строительстве завода. Возможны три варианта действий.

А. Построить большой завод стоимостью $M_1 = 700$ тысяч долларов. При этом варианте возможны большой спрос (годовой доход в размере $R_1 = 280$ тысяч долларов в течение следующих 5 лет) с вероятностью $p_1 = 0,8$ и низкий спрос (ежегодные убытки $R_2 = 80$ тысяч долларов) с вероятностью $p_2 = 0,2$.

Б. Построить маленький завод стоимостью $M_2 = 300$ тысяч долларов. При этом варианте возможны большой спрос (годовой доход в размере $T_1 = 180$ тысяч долларов в течение следующих 5 лет) с вероятностью $p_1 = 0,8$ и низкий спрос (ежегодные убытки $T_2 = 55$ тысяч долларов) с вероятностью $p_2 = 0,2$.

В. Отложить строительство завода на один год для сбора дополнительной информации, которая может быть позитивной или негативной с вероятностью $p_3 = 0,7$ и $p_4 = 0,3$ соответственно. В случае позитивной информации можно построить заводы по указанным выше расценкам, а вероятности большого и низкого спроса меняются на $p_5 = 0,9$ и $p_6 = 0,1$ соответственно. Доходы на последующие четыре года остаются прежними. В случае негативной информации компания заводы строить не будет.

Все расчеты выражены в текущих ценах и не должны дисконтироваться. Нарисовав дерево решений, определим наиболее эффективную последовательность действий, основываясь на ожидаемых доходах.

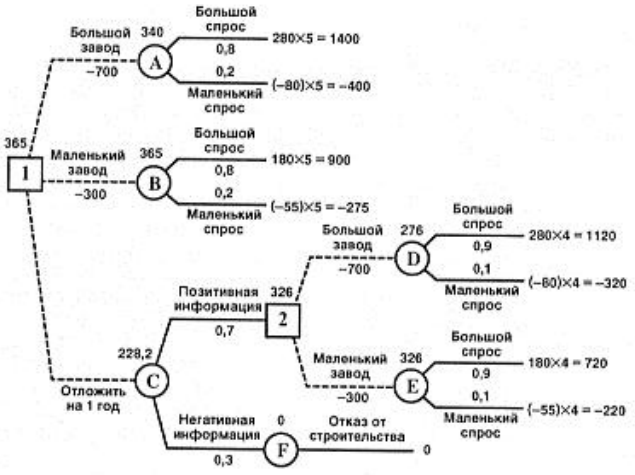


Рисунок 2

Ожидаемая стоимостная оценка узла A равна EMV(A) = 0,8 x 1400 + 0,2 x (-400) — 700 = 340.

EMV(B) = 0,8 x 900 + 0,2 x (-275) — 300 = 365.

EMV(D) = 0,9 x 1120 + 0,1 x (-320) — 700 = 276.

EMV(E) = 0,9 x 720 + 0,1 x (-220) — 300 = 326.

EMV(2) = max {EMV(D), EMV(E)} = max {276, 326} = 326 = EMV(E). Поэтому в узле 2 отбрасываем возможное решение «большой завод».

EMV(C) = 0,7 x 326 + 0,3 x 0 = 228,2.

EMV(1) = max {EMV(A), EMV(B), EMV(C)} = max {340; 365; 228,2} = 365 = EMV(B). Поэтому в узле 1 выбираем решение «маленький завод». Исследование проводить не нужно. Строим маленький завод. Ожидаемая стоимостная оценка этого наилучшего решения равна 365 тысяч долларов.

Модели сетевого планирования и управления

Поиск более эффективных способов планирования сложных процессов привели к необходимости использования моделей сетевого планирования и управления (СПУ). СПУ основано на моделировании процесса с помощью сетевого графика (сетевой модели). Сетевая модель и её основные элементы.

Сетевая модель представляет план выполнения некоторого комплекса работ.

Главными элементами сетевого графика является события и работа.

События – это завершение, какого либо процесса, отражающий отдельный этап выполнения проекта. На сетевом графике событие изображается кружком. Временные параметры сетевых графиков, коэффициенты напряжённости работы, анализ и оптимизация сетевого графика

ЗАДАЧА

Пусть для некоторого комплекса работ установлены оценки для каждой работы на уровне нормативных продолжительностей и срочного режима, а также даны стоимости. Информация представлена в таблице.

Таблица 1.

	Нормативный режим		Срочный режим	
	Продолжительность, дни	Стоимость, м/р	Продолжительность, дни	Стоимость, м/р
(1,2)	3	6	2	11
(1,3)	5	8	3	12
(1,4)	4	7	8	9
(2,5)	10	25	8	30
(3,5)	8	20	6	24
(3,6)	15	26	12	30
(4,6)	13	24	10	30
(5,7)	3	15	6	25
(6,7)	4	10	3	15

Построить график данного комплекса работ.

Требуется рассчитать:

- временные характеристики сетевого графика при нормальном режиме работ;
- найти критический путь;
- полные резервы времени;
- временные характеристики сетевого графика при срочном режиме работ;
- найти критический путь;

- полные резервы времени;
- определить стоимость работ.

Решение:

Рассчитаем временные характеристики для нормативного режима.

К временным характеристикам относятся ранние и поздние сроки наступления события.

Ранний срок наступления события рассчитывается по формуле:

tp(j) =mac ((t p (i) +t (ij)), где

t p (j) –ранний срок наступления предшествующего I события.

t (ij)- работа.

Для расчёта t p (j) для данного комплекса будем считать, что ранний срок наступления 1-го события равно tp(1)=0, тогда для последующих событий будем иметь:

t p(1)= ma x (tp(1)=0)

t p(2)= ma x (tp(1)+ tp(1,2)) =0+3=3

t p(3)= ma x ((tp(1)+ tp(1,3))=0+5=5

t p(4)= ma x (tp(1)+ tp(1,4))=0+4=4

t p(5)= ma x ((tp(4)+ tp(4,5)) =(2+10);(5+9)=14

t p(6)= ma x (tp(4)+ tp(4,6); tp(3)+ tp(3,6))=(4+13):(5+15)=20

t p(7)= ma x (tp(5)+ tp(5,7); tp(6)+

tp(6,7)=(14+8)(20+4)=24.

Очевидно, завершающее 7-е событие может наступить через 24 дня от начала выполнения всего комплекса работ. Поздний срок наступления события определяется по формуле:

tп(i)=min (tп(j)-t(ij))

Для расчёта t п(i) для комплекса будем считать, что самый поздний срок наступления 7-го события равен 24 дня, т.е. раннему сроку наступления 7-го события, тогда будем иметь:

tп(7)=min(24) =24

tп(6)=min(tп(7) - t(5,7)=(24-4)=20

tп(5)=min(24-4)=20

tп(4)=min(20-13)=7

tп(3)=min((16-9);(20-15) =5

tп(2)= min (16-10)=6

tп(1)= min (6-3; 5-5;7-4)=0

Полученный результат говорит о том, что расчёты произведены правильно.

Резервы времени определяем как разность между поздними и ранними сроками по формуле:

P(i)=tp(j) - t п (i)

P(1) =0-0=0

P(2)=6-3=3

P(3)=5-5=0

P(4)=7-4=3

P(5)=16-12=2

P(6)=20-20=0

P(7)=24-24=0

Полученные резервы времени показывают на какое время можно задержать наступление того или иного события, не вызывая опасности срыва выполнения комплекса работ. Те события, которые не имеют резервов времени, находятся на критическом пути.

Критический путь это наиболее продолжительный путь сетевого графика, который ведёт к завершению комплекса работ.

Находим пути и их длительности для данного комплекса работ:

1) 1-2-5-7 его стоимость: 3+10+8=21.

2) 1-3-5-7 его стоимость 5+9+8=22

3) 1-3-6-7. его стоимость: 5+15+4=24

4) 1-4-6-7. его стоимость: 4+13+4=21.

Критический путь: (1,3)-(3,6)-(6,7)
Резервы времени для работ, находящихся на критическом пути равны нулю.
(1,3)=0; (3,6)=0; (6,7)=0,
Рассчитаем временные характеристики сетевого графика при срочном режиме работ. Ранний срок наступления события рассчитывается по формуле:
tp(j) =max((tp(i) +t(ij)), где
tp(j) –ранний срок наступления предшествующего I события.
t(ij)- работа.
Для расчёта t p (j) для данного комплекса будем считать, что ранний срок наступления 1-го события равно tp(1)=0, тогда для последующих событий будем иметь:
tp(1)= max (t p(1)=0
tp(2)= max (tp(1)+ tp(1,2)) =0+2=2
tp(3)= max ((tp(1)+ tp(1,3))=0+3=3
tp(4)= max (tp(1)+ tp(1,4))=0+8=8
tp(5)= max ((tp(4)+ tp(4,5)) =(2+8);(3+6)=10
tp(6)= max (tp(2)+ tp(2,5); tp(3)+ t p(4,6))=(3+12):(8+10)=18
tp(7)= max (tp(5)+ tp(5,7); tp(6)+ t p(6,7)=(15+3);(18+3)=21.

Очевидно, завершающее 7-е событие может наступить через 21 день от начала выполнения всего комплекса работ.
Поздний срок наступления события определяется по формуле:
tп(7)=min(22) =24
tп(6)=min(t п (7)- t(5,7)=(21-3)=18
tп(5)=min(21-6)=15
tп(4)=min(18-10)=8
tп(3)=min((16-6);(19-15) =4
tп(2)= min (15-8)=7
tп(1)= min (15-2; 20-8;8-8)=0
Полученный результат говорит о том, что расчёты произведены правильно.
Резервы времени определяем как разность между поздними и ранними сроками по формуле:
P(i) =t p(j) - t п (i)
P(1) =0-0 =0
P(2)=7-2=5
P(3)=8-8=14
P(4)=8-8 =0
P(5)=12-8=4
P(6)=18-18=0
P(7)=22-22= 0
Найдём все пути: и их длительности.
1) 1-2-5-7 его стоимость: 3+8+6=16.
2) 1-3-5-7 его стоимость 3+6+6=15
3) 1-3-6-7. его стоимость:3+12+3=18
4) 1-4-6-7. его стоимость: 8+10+3=21.
Очевидно, что на критическом пути резервов времени нет.
Критический путь (1-3-6-7). Его длительность равна 21

2.9 Практическое занятие № 20 (2 часа)

Тема: Оптимизационные модели в сельском хозяйстве

2.9.1 Задание для работы:

1. Оптимизация производства сельскохозяйственного предприятия
2. Оптимальный план структуры производства сельскохозяйственного предприятия

2.9.2 Краткое описание проводимого занятия:

Постановка и экономико-математическая модель задачи оптимизации структуры производства и территории на примере крестьянского (фермерского) хозяйства.

Размеры крестьянских (фермерских) хозяйств и их структура: состав и площади земельных угодий, сочетание и размеры основных и дополнительных отраслей, структура посевов находится под влиянием множества природных и экономических факторов. Причем, для одного и того же хозяйства, находящегося в одних и тех же природных условиях и имеющего разные ресурсы денежно-материальных средств и труда, могут намечаться различные варианты организации производства территории , которые будут иметь и неодинаковую производственно-экономическую эффективность. Поэтому задача состоит в том, чтобы из всех возможных вариантов производства и территории крестьянского хозяйства выбрать ту производственную модель, которая, с одной стороны, удовлетворяла бы интересы крестьянина и государства, а с другой стороны – при наличии лимитированных ресурсов давала максимальный эффект. Решение данной задачи возможно с использованием оптимизационных экономико-математических методов моделирования ЭВМ.

Задача по организации производства и территории крестьянского (фермерского) хозяйства может иметь две основные постановки. Первая заключается в том, чтобы определить при известной площади крестьянского (фермерского) хозяйства его структуру, состав и площади земельных угодий, оптимальные размеры производства различных видов продукции. Такая постановка ничем не отличается от экономико-математической задачи по установлению специализации хозяйства, оптимальных размеров и сочетания его отраслей и хорошо известна в землеустройстве.

Более сложно устанавливать одновременно общую площадь и структуру крестьянского хозяйства и оптимизировать его производство исходя из размера крестьянской семьи, ее финансовых возможностей и конкретной экономической ситуации. Варьируя при этом ресурсами хозяйства, ценами, качественными характеристиками закрепленных земель и другими условиями, можно подобрать любой оптимальный вариант развития крестьянского (фермерского) хозяйства с его параметрами и характеристикой ожидаемых экономических результатов.

Вторая постановка задачи является общей по отношению к первой, поэтому с ее использованием сформулируем экономико-математическую модель.

Разделим все основные переменные задачи (x_j) на следующие совокупности:

$x_j(j \in Q_1)$ – площади с/х культур, возделываемых в крестьянском хозяйстве, га;

$x_j(j \in Q_2)$ – площади с/х угодий (кроме пашни), га;

$x_j(j \in Q_3)$ – поголовье животных, голов.

В качественных дополнительных переменных в задаче выступают следующие:

x_n- общая площадь пашни в хозяйстве, га;

x_y- потребное количество дополнительно приобретаемых органических удобрений, необходимых для поддержания бездефицитного баланса гумуса в почве, тонн;

x_{kk}- общее количество приобретаемых комбикормов, ц;

x_0- общая площадь с/х угодий крестьянского (фермерского) хозяйства, га;

x_N, x_P, x_{Kn} – потребность соответственно в азотных, калийных, фосфорных удобрениях, кг действующего вещества;

x_3- общие производственные затраты, руб;

$x_j(j \in Q_4)$ - переменные, характеризующие основные направления использования капиталовложений в хозяйстве, руб.

К числу их отнесены следующие:

- на производственное строительство(здания и сооружения, включая приобретение оборудования для животноводства);
- па покупку с/х техники;
- на использование или приобретения автотранспорта;
- на покупку скота;

В современных условиях хозяйство должно непрерывно развиваться, поэтому направления капиталовложений рассматриваются как направляющие и основные элементы такого развития. Главным условием такого развития является определение и изменение специализации, развитие структуры производства.

Дополнительно к названным в задачу могут включаться переменные, характеризующие размер капиталовложений на мелиорацию земель, осуществление комплекса противоэрозионных мероприятий, закладку многолетних насаждений и др.

x_k- общий размер капиталовложений, необходимых или вкладываемых в развитие хозяйства;

x_l^t – привлекаемые трудовые ресурсы в напряженные периоды времени;

$x_{lj}(j \in Q_h)$ – обьемы производства товарной продукции растениеводства и животноводства, тонн;

На неизвестные накладываются следующие ограничения:

1. По общей площади с/х угодий (S_0)

$$\Sigma_{j \in Ql} x_j + \Sigma_{j \in Q2} x_j - x_0 = 0 \quad (x_0 \le S_0)$$

2. По площади пашни(S_n)

$$\Sigma_{i \in Ql} x_j - x_n = 0 \quad (x_n \le S_n)$$

3. По трудовым ресурсам

$$\Sigma_{j \in Q} t_{ij} x_j - x^t_i \le T_i, \quad i \in M_1,$$

где $Q = Q_1 V Q_2 V Q_3$;

t_{ij} – норма затрат труда на 1 га площади или голову скота в период времени i (в среднем за год, на период уборочных работ.), чел.-ч; T_i – общий объём трудовых ресурсов в i-й период; $(i \in M_1)$- число выделенных периодов работ.

4. По поддержанию бездефицитного баланса гумуса в почве с целью создания условий для воспроизводства почвенного плодородия $\Sigma_{j \in Q1VQ2} a_j x_j - \Sigma_{j \in Q3} \beta_j x_j - x_y \le 0$

где a_j- норма минерализации (накопления) гумуса под посевами с/х культур и угодья, тонн\га (вводится со знаком (+) в случае выноса гумуса и со знаком (-) при его образовании); β_j – коэффициент, учитывающий образование гумуса за счет разложения органических удобрений, получаемых с одной головы скота, тонн\голову.

5. По балансу минеральных удобрений

- Азотных

$$\Sigma_{j \in QlVQ2} y_{nj} x_j - x_n = 0,$$

✓ Фосфорных

$$\Sigma_{j \in Q1VQ2} y_{pj} x_j - x_p = 0,$$

✓ Калийных

$$\Sigma_{j \in Q1VQ2} y_{kj} x_j - x_k = 0,$$

где y_{nj}, y_{pj}, y_{kj} – соответственно норма внесения азотных, фосфорных и калийных удобрений в расчёте на 1 га площади, кг действующего вещества.

6. По расчету ежегодных производственных затрат хозяйства (без оплаты собственного труда)

$$\Sigma_{j\in Q}\,\Delta j x_j + \Delta K_k x_k + \Delta y x_y + \Delta t x_t - x_3 = 0, \quad x^{\Sigma}_t = \Sigma_i\, x^t_i$$

где Δj, ΔK_j, Δy, Δt - соответственно производственные затраты на возделывание культур, уход за угодьями и скотом, стоимость приобретения единицы комбикормов; одной тонны органических удобрений; оплата одного человеко-часа работы привлекаемых трудоспособных.

7. По общему размеру капиталовложений, необходимых для организации крестьянского хозяйства и их структуре

✓ общие капиталовложения

$\Sigma_{j\in Q4}\ x_j - x_k = 0,\ x_k \le D,\ \Sigma_{j\in Q4\setminus Q4i}\ x_j = x_n \Sigma_r \eta^n_r,$

✓ виды единовременных затрат

$\Sigma_{j\in Q}\ \eta_{rj}\ x_j - x_r = 0;\quad i\in M_2;\ r\in Q_r;\ Q_r = Q_4;$

где x_n- расчетная площадь пашни

η_{rj} - норма затрат капиталовложений r-го вида на одно скотоместо, голову приобретаемого скота, гектар пашни или площади с/х угодий;

$i,\ i\in M_2$ – число видов производственных затрат.

8. По кормам, кроме зеленых кормов (в кормовых единицах, переваримом протеине, натуральных центнерах)

$-\Sigma_{j\in Q1 \vee Q2}\ y_{ij}x_j + \Sigma_{j\in Q3}\ v_{ij}x_j \le 0;\ i\in M_3$

в т.ч по концентратам

$-\Sigma_{j\in Q1 \vee Q2}\ y_1 x_j + \Sigma_{j\in Q3}\ v_{ij}x_j - x_{Kk} \le 0;\ i\in M_5,$

где y_{ij} – урожайность кормовых культур и продуктивность угодий, ц/га, по i-му виду корма.

В этой же группе ставятся ограничения по обьему полученного комбикорма за счет сдачи продукции растениеводства и животноводства в переводе на соответствующие эквиваленты.

V_{ij} – потребность в i-ом корме на одну голову скота в год;

$i,\ i\in M_5$ – виды кормов, кроме зеленых.

9. По схеме зеленого конвейера по месяцам пастбищного периода

- $\Sigma_{j\in Q1 \vee Q2}\ y^3_{ij}x_j + \Sigma_{j\in Q3}\ d_{ij}\ v_{ij} \le 0\ ;\ i\in M_4,\ З_j = \Sigma_j y^3_{ij}$

где y_{ij} – урожайность культур на зеленый корм и выход кормов с пастбищ в i-ый месяц пастбищного периода, ц/га;

d_{ij}- удельный вес потребности животных в зеленом корме в i-й месяц пастбищного периода, ц;

10. По агротехническим требованиям, предъявляемым к возделываемым культурам и их рекомендуемому удельному весу в структуре посевных площадей

$$x_j - d_j x_n \le 0,\ j\in Q_6,$$
$$\Sigma_{j\in Qn} x_j - d_j x_n \le 0,$$

где d_j- рекомендуемый удельный вес культур или групп по верхней границе в структуре посевных площадей;

$x_j,\ j\in Q_6$ – площади с/х культур одинаковых агрогрупп

Дополнительно в этой группе формируются ограничения по предшественникам озимых культур, по площади многолетних трав, идущих на семена, для обеспечения потребностей в семенах многолетних трав за счет собственного производства.

11. По расчету объёмов производства товарной продукции крестьянского хозяйства (x_j)

$\Sigma_{j\in Q}\ q_{ij}x_j - x_i = 0,\ i\in M_5,$

где q_{ij} – выход продукции i-го вида с гектара площади или одной головы скота;

i, $i \in M_5$, - виды товарной продукции.

12. По гарантированным объёмам производства товарной продукции с/х

$\Sigma_{j \in Q} q_{ij} x_{j, j \in M5} \geq Q_i$,

где Q_i – гарантированное производство продукции i-го вида.

В данной задаче могут ставиться также условия, ограничивающие или фиксирующие на данной величине поголовье скота, а также площади отдельных с/х культур и другие ограничения, учитывающие специфику природных и экономических условий хозяйства.

13. Условия неотрицательности переменных:

$$x_j \geq 0;\ x_{ij} \geq 0;\ x_0 \geq 0;\ x_n \geq 0;$$
$$x_y \geq 0;\ x_{kk} \geq 0;\ x\ldots \geq 0;\ x\ldots. \geq 0;$$
$$x_{kc} \geq 0;\ x_з \geq 0.$$

В связи с тем что, крестьяне в ходе свое деятельности заинтересованы произвести большое количество товарной продукции с меньшими материально-денежными затратами, в качестве критерия оптимальности данной задачи наиболее целесообразно использовать максимум хозяйственной прибыли.

$Z = \Sigma_{j \in Q} c_j xj - x_З \rightarrow \max$,

Где c_j – стоимость единицы товарной продукции хозяйства, руб.

. $x_з$ – общие денежные затраты, руб.

Максимизация прибыли – основная цель. При оптимальной структуре производства может быть получена реальная максимальная прибыль. Показатель Z при заданной структуре издержек можно определить как условно-чистый доход.

2. ПОДГОТОВКА ИСХОДНОЙ ИНФОРМАЦИИ И ПОСТРОЕНИЕ МАТРИЦЫ ЭКОНОМИКО-МАТЕМАТИЧЕСКОЙ МОДЕЛИ.

2.1. Подготовка исходной информации

Решение землеустроительных и экономических задач требует наличия информации, которая наполняет экономическую модель конкретным содержанием. Экономическая и землеустроительная информация представляет совокупность сведений о моделируемом объекте или процессе, необходимых для решения конкретной задачи.

Информация имеет определенное материальное воплощение. При подготовке исходной информации изучают следующие условия:

- экономические (перечень отраслей, виды производственных ресурсов, гарантированные объёмы производства или размеры отдельных отраслей, условия реализации продукции и др.);
- технологические (особенности агротехники возделывания с/х культур, ветеринарные и зоотехнические требования к размещению животных, условия их размещения);
- землеустроительные (размеры землепользований, землевладений, структура угодий, возможности трансформации, условия организации севооборотов, их видов, вопросы устройства угодий и др.);
- социальные и экологические условия.

Характер информации определяется содержанием задачи и математическим методом, с помощью которого ее будут решать. При подготовке информации используется отчетная, плановая и нормативная информация.

К отчетной информации относится:

1. Земельно-учетные данные, экспликация земель, структура угодий.
2. Размеры с/х отраслей (основных и дополнительных).
3. Наличие трудовых ресурсов
4. Виды и размеры производственных ресурсов (денежно-материальных, виды органических и минеральных удобрений, наличие с/х техники и др.)

Нормативная информация - ее основным источником являются технологические карты, справочники, она представляет собой нормы затрат определенного ресурса на единицу переменной.

Плановая информация - перспективные исходные данные, характеризующие наличие и распределение основных производственных ресурсов, урожайность с/х культур, продуктивность животных, цены реализации основных видов продукции, чистый доход и т.д.

Результаты решения задачи по оптимизации размера землепользования (землевладения) и структуры производства крестьянского (фермерского) хозяйства во многом определяется и имеющейся базой данных, т. е. той информацией, которая будет вводиться в ЭВМ в качестве технолого-экономических коэффициентов и объёмов ограничений.

Учитывая это, а также отсутствие достоверных фактических сведений по вновь организуемым крестьянским хозяйствам, основным источником информации должны служить нормативные данные, специально разрабатываемые для этих целей на основе технологических карт, типовых проектов или анализов.

Основным источником информации по нормативам затрат труда и денежно- материальных средств, необходимых для производства единицы продукции (на 1 га площади с/х культур, кормовых угодий, многолетних насаждений или голову скота) по крестьянским хозяйствам должны служить технологические карты, включающие основные виды работ, состав машинно-тракторного парка хозяйства, рассчитанные на определенную техническую оснащенность и фондовооруженность предприятия. При этом основой для вычисления производственных затрат по видам продукции должна служить калькуляция себестоимости.

При решении задачи и разработке технолого-экономических коэффициентов следует использовать данные по затратам труда, нормам высева, минерализация гумуса, внесению удобрений, выходу кормовых единиц и переваримого протеина, ориентировочной урожайности с/х культур, схеме зеленого конвейера.

2.2. Подготовка исходной матрицы задачи, решаемой симплексным методом на примере крестьянского (фермерского) хозяйства.

Крестьянскому хозяйству выделено 127 га пашни, 10 га сенокосов и 12 га пастбищ. Общее число трудоспособных составляет 5 человек. Необходимо определить оптимальную структуру землевладения и производства крестьянского хозяйства при заданном отраслевом составе.

При разработке матрицы выделено 23 основные переменные. К ним отнесены следующие группы неизвестных.

1.Зерновые и зернобобовые культуры, идущие на товарные цели и характеризующиеся площадями посева, га

x1 - яровая пшеница
x2 – зернобобовые
x3 – овес

2.Зерновые, используемые в качестве концентрированных кормов.

x4 – кукуруза
x5 – зернобобовые

3 Технические культуры

x6 – овощи
x7 – кукуруза на силос

4.Кормовые культуры и угодья

x8 – многолетние травы

Зеленые корма:

x9 – кукуруза на зеленый корм

x10 – многолетние травы на сенаж

Кроме того, введены переменные, характеризующие площади: x11 - сенокосов; x12 – пастбищ.

К числу переменных, описывающих отрасли животноводства, отнесены следующие:
x13 – поголовье КРС, гол.
x14 – поголовье свиней, гол.

Кроме названных в задачу включены следующие неизвестные:
X20 – общие ежегодные производственные затраты, тыс. руб.
x17 – потребность в азотных минеральных удобрениях, кг д. в.
x18– потребность в фосфорных минеральных удобрениях, кг д. в.
x19 – потребность в калийных минеральных удобрений, кг д. в.
x15- объем приобретаемых комбикормов (концентратов), ц
x16 – количество приобретаемых органических удобрений, т
x21- капиталовложения в основные фонды с/х назначения, тыс. руб.
В число переменных включены также:
x22 – расчетная площадь пашни, га
x23 - размер привлекаемых трудовых ресурсов в напряженные периоды полевых работ, чел.-час.

Целевая функция:
$Z = 6x1 + 8x2 + 3x3 + 50x6 + 7x8 + 21x12 + 10x14 – x21 \rightarrow \max$

На неизвестные наложено 25 ограничений, т. е. размер матрицы задачи составляет 23*25. Ограничения включают в себя следующие :
1. По имеющейся площади пашни
$x22 \leq 65$
2. По расчету площади пашни
$x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 – x22 = 0$
3. По имеющейся площади сенокосов
$x11 = 22$
4. По имеющейся площади пастбищ
$x12 = 28$
5. По трудовым ресурсам
$15x1 + 10.5x2 + 27x3 + 27x4 + 10.5x5 + 1800x6 + 32.3x7 + 12x8 + 32.3x9 + 15.1x10 +13.3x11 +10x12 +100x13 + 22x14 \leq 14190$
6. По трудовым ресурсам в напряженный период
$8.1x1 + 9.8x2 + 26x3 + 26x4 + 9.8x5 + 1350x6 + 30x7 + 8x8 + 30x9 + 3.9x10 + 8.4x11 + 1.7x12 +18.3x13 + 3.5x14 – x23 \leq 3360$
Величина трудовых ресурсов на 1 человека в крестьянском хозяйстве должна составлять не менее 2838 чел.-ч в год ,в том числе, в напряженный период (июль-август) – не менее 672 чел.-ч. При наличии в хозяйстве 5 трудоспособных, общая величина трудовых затрат за год не должна превышать 14190 чел.-час, в том числе ,в напряженный период - 3360 чел.-час.
7. По поддержанию бездефицитного баланса гумуса в почве с целью создания условий для воспроизводства почвенного плодородия
$0.029x1 + 0.023x2 + 0.6x3 + 1.57x4 + 0.6x5 + 1.57x6 + 1.57x7 + 0.8x8 + 1.57x9 – 0.8x14 – 0.8x15 – 15.2x10 – 3.1x11 – x27 \leq 0$

8. По балансу азотных минеральных удобрений

$60x1 + 60x2 + 60x3 + 120x4 + 60x5 + 120x6 + 26x7 + 60x8 + 120x9 – x16 = 0$

9. По балансу фосфорных минеральных удобрений

$43x1 + 43x2 + 50x3 + 86x4 + 50x5 + 86x6 + 43x7 + 50x8 + 89x9 – x17 = 0$

10. По балансу калийных удобрений

$60x2 + 60x3 + 60x5 + 60x8 – x18 = 0$

11. По расчету производственных затрат

$7.42x1 + 4.42x2 + 2.72x3 + 4.86x4 + 2.72x5 + 4.86x6 + 3.52x7 + 8.33x8 + 2.86x9 +$
$10.5x10 + 5x11 + 0.5x12 – x13 + 2.5x14 + 2x15 + x26 + 0.05x28 = 0$

12. По расчету общих капиталовложений

$x19 + x23 – x24 + 8.85x25 = 0$

13. По расчету капиталовложений на с/х технику, автотранспорт, мелиорацию и природоохрану

$– x22 + 8.85x25 = 0$

14. По расчету капиталовложений на здания, сооружения и приобретение скота

$26.5x10 + 8x11 + 0.25x12 – x19 – x23 = 0$

15. По ощему размеру капиталовложений, необходимых для организации

$x24 \leq 150000$

16. По кормам в кормовых единицах

$-41.1x5 – 49.6x6 – 44x7 + 36.5x10 + 6.5x11 + 0.5x12 – 4.2x14 – 9x15 – x26 \leq 0$

17. По концентратам

$-41.1x5 – 49.6x6 + 7.4x10 + 6x11 + 0.35x12 \leq 0$

18. По кормам в перевариваемом протеине

$-2.8x5 – 2.9x6 –44x7 + 3.68x10 + 0.65x11 + 0.04x12 – 0.5x14 – 0.8x15 –0.1x26 \leq 0$

19. По зеленым кормам за пастбищный период

$-220x9 + 27x10 + 0.32x11 – 33x15 \leq 0$

20. По агротехническим требованиям возделывания зерновых культур

$x1 + x2 + x3 + x4 –0.35x25 \leq 0$

На Дальнем Востоке рекомендуемый удельный вес зерновых культур в структуре посевных площаей составляет 35 %.В районах с более благоприятными природными условиями удельный вес зерновых может приниматься равным 55% и более.

21. По агротехническим требованиям возделывания кукурузы на силос

$x7 – 0.2x25 \leq 0$

22. По агротехническим требованиям возделывания сои

$x8 – 0.2x25 \leq 0$

23. По агротехническим требованиям возделывания кукурузы на зеленый корм

$x9 – 0.2x25 \leq 0$

На Дальнем Востоке рекомендуемый удельный вес в структуре посевных площадей кукурузы на силос, сои и кукурузы на зеленый корм составляет 20%.

24. По гарантированным обьемам производства зерна (по госзаказу)

$0.32x1 + 28x2 +13x3 + 29x4 \geq 100$

25. По гарантированным обьемам производства мяса (по госзаказу)

$0.2x10 + 0.1x11 + 0.001x12 \geq 1$

Матрица экономико-математической модели задачи по оптимизации структуры землевладения и проиводства крестьянского хозяйства представлена в таблице 1.

3. АНАЛИЗ ПОЛУЧЕННОГО РЕШЕНИЯ

а) Эктремальное значение целевой функции – чистый доход хозяства:

$Z_{max} = 2541.5$ тыс. руб.

б) Основные переменные, попавшие в базис:

площадь посева кукурузы на зерно	x4 = 3.45 га
площадь посева ярового ячменя на концентраты	x5 = 82.4 га
площадь посева кукурузы на силос	x7 = 2.74 га
поголовье свиней	x11 = 562 гол
площадь сенокосов	x14 = 10 га
площадь пастбищ	x15 = 12 га

в) Основные переменные, не попавшие в базис:

x1 - площадь посева яровой пшеницы, га
x2 - площадь посева ярового ячменя на зерно, га
x3 - площадь посева зернобобовых, га
x6 - площадь посева кукурузы на концентраты, га
x8 - площадь посева сои, га
x9 - площадь посева кукурузы на зеленый корм, га
x10 -поголовье КРС, гол
x12 - поголовье птицы, гол

г) Вспомогательные переменные, попавшие в базис:

общие ежегодные производственные затраты	x13 = 3113 тыс.руб.
потребность в азотных минеральных удобрениях	x16 = 5429.6 кг д. в.
потребность в фосфорных минеральных удобрениях	x17 = 4534.8 кг д. в.
потребность в калийных минеральных удобрениях	x18 = 4944.7 кг д. в.
капиталовложения в строительство зданий и сооружений	x19 = 4501.тыс.руб .
капиталовложения на покупку с/х техники	x20 = 784.1 тыс. руб.
суммарные затраты	x24 = 5285.6 тыс.руб.
расчетная площадь пашни	x25 = 88.6 га

д) Впомогательные переменные, не попавшие в базис:

x21 - капиталовложения на приобрктение автотранспорта, тыс. руб
x22 - капиталовложения на мелиорацию и проведение природоохранных мероприятий, тыс.руб
x23 – капиталовложения на покупку продуктивного скота, тыс. руб.
x26 – обьем приобретаемых комбикормов (концентратов), ц.
x27- количество приобретаемых органических удобрений, т.
x28 – размер привлекаемых трудовых ресурсов в напряженные периоды полевых работ,чел-час

е) Дополнительные переменные, попавшие в базис:

остаток площади пашни	x29 = 38.4 га
недоиспользованные трудовые ресурсы в напряженный период полевых работ	x30 = 183.2 чел.-час.
остаток органические удобрения	x31 = 1702.8 кг д. в.
недоиспользованные денежные ресурсы	x32 =149740тыс.руб
остаток концентрированных кормов	x33 = 10.9 ц.к.е.
остаток зеленых кормов	x34 = 215.9 ц.
остаток площади пашни, предусмотренной под зерновые	x35 = 27.6 га
остаток площади пашни, предусмотренной под кукурузу на силос	x36 = 15.0 га
остаток площади пашни, предусмотренной под сою	x37 = 17.7 га
остаток площади пашни, предусмотренной под кукурузу на зеленый корм	x38 = 17.7 га
объем производства мяса	x39 = 55.3 т

ж) Дополнительные переменные не папавшие в базис:

x40 – остаток трудовых ресурсов, чел.-час
x41 – остаток кормов, соответствующий общему балансовому уравнению, ц.к.е.

x42 – остаток кормов, соответствующий общему балансовому уравнению (переваримый протеин), ц

x43 – объем производства зерна, ц

Анализируя полученное решение, можно сказать следующее:

Эффективными отраслями хозяйства, т. е. которые при заданных ресурсных ограничениях целесообразно развивать, являются свиноводство и производство ярового ячменя на концентрированный корм. Небольшие площади пашни целесообразно отвести под посевы кукурузы на зерно и кукурузы на силос.

Недефицитными ресурсами, т. е. ресурсами, которых в хозяйстве избыток, являются пашня, трудовые ресурсы в напряженный период полевых работ, денежные ресурсы, органические удобрения, концентрированные корма, зеленые корма, полученные с сенокосов и пастбищ.

К ресурсам, которые в оптимуме исчерпываются полностью, т. е. дефицитным, относятся общие трудовые ресурсы и корма (кукуруза на силос). Ограниченность этих ресурсов сдерживает дальнейший рост производства. Именно за счет их увеличения можно повысить доход хозяйства.

Увеличение трудовых ресурсов возможно за счет повышения производительности труда путем разделения рабочего дня на две смены. Продолжительность всего рабочего дня изменяется по месяцам. Средняя продолжительность рабочей смены принимается равной 5 часов, а в выходые дни – 2.5 час Общая величина трудовых ресурсов за год, в этом случае, составит 15095 чел.- час, в том числе, в напряженный период полевых работ (июль- август) – 3430 чел.- час.

Увеличение площади посева кукурузы на силос возможно за счет недоиспользованной площади пашни.

Анализ вариантного решения

Результаты вариантного решения представлены в таблицах 5 и 6.

а) Экстремальное значение целевой функции – чистый доход хозяйства:

Z_{max} = 2776 тыс. руб.

б) Основные переменные, попавшие в базис:

площадь посева кукурузы на зерно	x4 = 7.2 га
площадь посева ярового ячменя на концентраты	x5 = 92.1 га
площадь посева кукузузы на силос	x7 = 3 га
поголовье свиней	x11 = 603 гол

в) Основные переменные, не попавшие в базис:

x1 – площадь посева яровой пшеницы, га
x2 – площадь посева ярового ячменя на зерно, га
x3 – площадь посева зернобобовых, га
x6 – площадь посева кукурузы на концентраты, га
x8 – площадь посева сои, га
x9 – площадь посева кукурузы на зеленый корм, га
x10 – поголовье КРС, гол
x12 – поголовье птицы, гол
x14 – площадь сенокосов, га
x15 – площадь пастбищ, га

г) Вспомогательные переменные, попавшие в базис:

общие ежегодные производственные затраты	x13 = 3310.8 тыс. руб.
потребность в азотных минеральных удобрениях	x16 = 6471.7 кг д. в.
потребность в фосфорных минеральных удобрениях	x17 = 5356.9 кг д.
потребность в калийных минеральных удобрениях	x18 = 5525.5 кг д. в.
капиталовложения на строительство зданий и. сосоружений	x19 = 4823.1 тыс. руб
капиталовложения на покупку с/х техники	x20 = 905.9 тыс. руб.

суммарные затраты x24 = 5729.0 тыс руб.
расчетная площадь пашни x25 = 102.4 га

д) Вспомогательные переменные, не попавшие в базис:

x21 - капиталовложения нп приобретение автотранспорта, тыс. руб.
x22 – капиталов ложения на мелиорацию и проведение природоохранных мероприятий, тыс. рубю
x23 - капиталовложения на покупку продуктивного скота, тыс. руб.
x26 - объем приобретаемых комбикормов (концентратов), тыс. руб.
x27 - количество приобретаемых органических удобрений, т
x28 – размер привлекаемых трудовых ресурсов в напряженный период полевых работ, чел.-час.

е) Дополнительные переменные, попавшие в базис:

остаток органических удобрений x31 = 1797.6 кг д. в.
недоиспользованные денежные ресур x32 = 1494300 тыс. руб.
остаток концентрированных кормов x33 = 167.4 ц.к.е.
остаток площади пашни, предусмотренной под зерновые x35 = 28.6 га
остаток площади пашни, предусмотренной под кукузуру на силос x36= 17.4 га
остаток площади пашни, предусмотрнной под x37 = 20.5 га
остаток площади пашни, предусмотренной под кукурузу на зеленый корм x38 = 20.5 га
объем производства мяса x39 = 59.3 т
объем производства зерна (кукурузуза) x43 = 109.6 ц

ж) Дополнительные переменные, не попавшие в базис:

x41- остаток кормов, соответствующий общему балансовому уравнению, ц.к.е.
x42 - остаток кормов, соответствующий общему балансовому уравнению (переваримый протеин), ц
x44 - остаток площади сенокосов, га
x45 - остаток площади пастбищ, га

Анализируя полученный результат вариантного решения, можно сказать следующее: Специализация хозяйства – свиноводство. Максимальный чистый доход хозяйства – 2776 тыс. руб. Максимальное количество свиней, которое можно содержать при заданных ресурсных ограничениях равно 603 головы.

При использовании всех имеющихся в хозяйстве трудовых ресурсов оптимальная площадь пашни, при которой достигается максимальный чистый доход, составляет 102.3 га. Увеличение площади пашни нецелесообразно, так как это приведет к уменьшению дохода хозяйства. Сенокосы и пастбища в данном хозяйстве использовать также нецелесообразно.

Эффективной отраслью хозяйства является производство ярового ячменя, используемого в качестве концентрированного корма для свиней. На корм свиньям также используется кукуруза на силос.

В хозяйстве возможно производство кукурузы на зерно, при использовании ее в товарных целях.

К недефицитным ресурсам хозяйства относятся пашня, денежные ресурсы. Кроме того в хозяйстве имеится остаток концентрированных кормов и органических удобрений, прибыль от их реализации будет дополнительным доходом к основному доходу, получаемому от продажи мяса.

Дефицитным ресурсом хозяйства, сдерживающим дальнейший рост производства , являются трудовые ресурсы. Увеличить величину трудовых ресурсов возможно за счет

увеличении продолжительности рабочей смены. В данном случае средняя ее величина составляет 5 часов.

Кроме того, используя оставшиеся в хозяйстве денежные средства, можно применять наемный .труд.

Увеличение трудовых ресурсов, позволит использовать всю имеющуюся в хозяйстве площадь пашни (127 га), что приведет к повышению чистого дохода хозяйства.

2.9.3 Результаты и выводы: Представленная экономико- математическая модель оптимизации землевладения и производства хозяйства, является рабочей. Полученное решение приемлемо. Основные отрасли хозяйства – свиноводство и кормопроизводство. Дополнительный доход может быть получен за счет реализации остатка концентрированных кормов и органических удобрений. При полученном оптимальном сочетании отраслей хозяйство является прибыльным и имеет перспективы развития.