ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.В.ДВ.11.01 Основы научных исследований

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Профиль подготовки: Автоматизированные системы обработки информации и управления

Форма обучения: очная

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция 1 (2 ч.)
Тема: Наука в современном обществе.

1.1.1. Вопросы лекции:

1. Основные тенденции в развитии современной науки
2. Организация научно-исследовательской работы
3. Научные исследования в СССР, России
4. Проблемы современной фундаментальной науки

1.1.2 Краткое содержание вопросов:

1. Основные тенденции в развитии современной науки

Наука – это непрерывно развивающаяся система знаний объективных законов природы, общества и мышления, получаемых и превращаемых в непосредственную производительную силу общества в результате социально-экономической деятельности.

Это синтез организованной особым образом познавательной деятельности и ее результатов. Под **особым образом познавательной деятельности** понимается методологические и мировоззренческие принципы, обеспечивающие научный подход к выбору, постановке и реализации исследования. Термин наука применяется также и для обозначения отдельной области знаний.

Основная **цель науки** – познание объективного мира (теоретическое отражение действительности) и воздействие на окружающую среду с целью получения полезных обществу результатов.

Наука поддерживается и развивается в результате исследовательской деятельности общества.

Научное исследование – это форма существования и развития науки. Структуру организации научных исследований целесообразно представить в виде четырех компонентов (рис.1.):

- первый - общие вопросы научных исследований (теория, методология и методы);
- второй – процессы научных исследований (формы, методы и средства познания);
- третий – методика научных исследований (выбор конкретных форм, методов и средств, эффективных для соответствующей области науки или отрасли профессиональной деятельности);
- четвертый – технология научных исследований

Рисунок - 1 - Структура организации научных исследований

(совокупность знаний о процессах научных исследований и методике их выполнения);

2. Организация научно-исследовательской работы

Научная теория – это высшая форма организации теоретического знания, представляющая собой совокупность объединенных в единую систему основных элементов теории (подтвержденных гипотез, понятий, суждений) в соответствующей отрасли (в данном случае в информатике). Критерием истинности теории является ее практическое подтверждение.

Основой любой науки и, в частности, науковедения является **методология**, которая представляет собой учение о структуре, логической организации, методах и средствах деятельности.

В научной литературе под **методологией** обычно понимается, прежде всего, система научного познания, т.е. учение о принципах построения, формах и способах научно-познавательной деятельности.

Методология может быть **специально-научная и философская**.

Специально-научная методология разделяется на несколько уровней: общенаучные методологические концепции и направления, методология отдельных специальных наук, методика и технология исследований.

Философская методология определяет систему философских знаний. Частным способом реализации методологии на практике является метод, как система действий в различных видах человеческой деятельности направленных на достижение поставленной задачи.

Научный метод – это система правил и предписаний, направляющих человеческую деятельность (производственную, политическую, культурную, научную, образовательную и т.д.) к достижению поставленной цели.

Если методология – это стратегия научных исследований, обеспечивающих достижение цели, сформулированной в гипотезе предполагаемых научных результатов (генеральный путь познания), то метод – это тактика, показывающая как лучше всего идти этим путем.

Метод (гр. methodos) — 1) способ познания, исследования явлений природы и общественной жизни; 2) прием, способ и образ действий.

Метод — путь исследования, способ достижения какой-либо цели, решения конкретных задач. Это совокупность подходов, приемов, операций практического или теоретического освоения действительности.

Из определения метода вытекает, что существуют **две большие группы методов**: познания (исследования) и практического действия (преобразовательные методы) (рис.2).

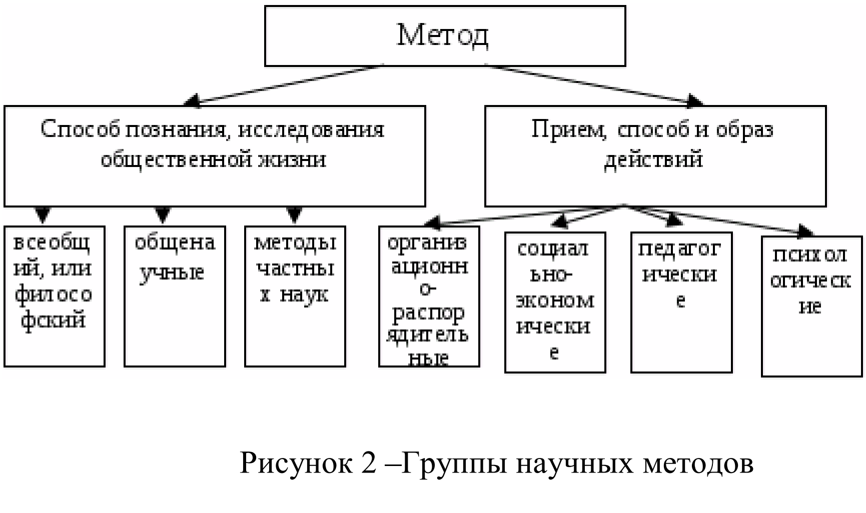


Рисунок 2 –Группы научных методов

1) Методы исследования — приемы, процедуры и операции эмпирического и теоретического познания и изучения явлений действительности. С помощью этой группы методов получают достоверные сведения, используемые для построения научных теорий и выработки практических рекомендаций. Система методов исследования определяется исходной концепцией исследователя: его представлениями о сущности и структуре изучаемого, общей методологической ориентации, целей и задач конкретного исследования. Методы подразделяются на следующие:

- всеобщий, или философский, общенаучные и методы частных наук;
- констатирующие и преобразующие;
- эмпирические и теоретические;
- качественные и количественные;
- содержательные и формальные;
- методы сбора эмпирических данных, проверки и опровержения гипотез и теории;
- описания, объяснения и прогноза;
- обработки результатов исследования.

Всеобщий, или философский метод — всеобщий метод материалистической диалектики.

К **общенаучным методам** относятся:

- Наблюдение – это способ познания объективного мира, основанный на непосредственном восприятии предметов и явлений при помощи органов чувств без вмешательства в процесс со стороны исследователя.
- Сравнение - это установление различия между объектами материального мира или нахождение в них общего; осуществляется как при помощи органов чувств, так и при помощи специальных устройств.
- Счет – это нахождение числа, определяющего количественное соотношение однотипных объектов или их параметров, характеризующих те или иные свойства.
- Измерение – это физический процесс определения численного значения некоторой величины путем сравнения ее с эталоном.
- Эксперимент – одна из сфер человеческого практики, в которой подвергается проверке истинность выдвигаемых гипотез или выявляются закономерности объективного мира.
- Обобщение – определение общего понятия, в котором находит отражение главное, основное, характеризующее объекты данного класса.
- Абстрагирование – это мысленное отвлечение от несущественных свойств, связей, отношений предметов и выделение нескольких сторон, интересующих исследователя.
- Формализация – отображение объекта или явления в знаковой форме какого-либо искусственного языка (математики, химии и т.д.).
- Аксиоматический метод – способ построения научной теории, при котором некоторые утверждения принимаются без доказательств.
- Анализ – метод познания при помощи расчленения или разложения предметов исследования на составные части.
- Синтез – соединение отдельных сторон предмета в единое целое.
- Индукция – умозаключение от фактов к некоторой гипотезе (общему утверждению).
- Дедукция – умозаключение, в котором вывод о некотором элементе множества делается на основании знания общих свойств всего множества.
- Аналогия – метод, посредством которого достигается знание о предметах и явлениях на основании того, что они имеют сходство с другими.
- Гипотетический метод познания предполагает разработку научной гипотезы на основе изучения физической, химической и т.п., сущности исследуемого явления, формулирование гипотезы, составление расчетной схемы алгоритма (модели), ее изучение, анализ, разработка теоретических положений.
- Исторический метод познания предполагает исследование возникновения, формирования и развития объектов в хронологической последовательности.
- Идеализация - это мысленное конструирование объектов, которые практически неосуществимы.

- Системные методы: исследование операций, теория массового обслуживания, теория управления, теория множеств и др.

Методы частных наук — специфические способы познания и преобразования отдельных областей реального мира, присущие той или иной конкретной системе знаний (социология — социометрия; психология — психодиагностика).

2) Методы как прием, способ и образ действий (методы практической деятельности) включают в себя способы воздействия, совокупность приемов, операций и процедур подготовки и принятия решения, организации его выполнения.

Для выбора методов на каждом этапе необходимо знать общие и конкретные возможности каждого метода, его место в системе исследовательских процедур. Задача исследователя состоит в том, чтобы для каждого этапа исследования определить оптимальный комплекс методов.

Разнообразные **методы** научного познания условно подразделяются на ряд **уровней**: *эмпирический, экспериментально-теоретический, теоретический и метатеоретический.*

Методы эмпирического уровня: *наблюдение, сравнение, счет, измерение, анкетный опрос, собеседование, тесты, метод проб и ошибок и т.д.*

Методы экспериментально-теоретического уровня: *эксперимент, анализ и синтез, индукция и дедукция, моделирование, гипотетический, исторический и логический методы.*

Методы теоретического уровня: *абстрагирование, идеализация, формализация, анализ и синтез, индукция и дедукция, аксиоматика, обобщение и т.д.*

К **методам метатеоретического уровня** относятся *диалектический* и метод *системного анализа.*

Творчество – мышление в его высшей форме, выходящее за пределы известного, а также деятельность, порождающая нечто качественно новое.

В частности, *научное творчество* связанно с познанием окружающего мира. *Научно-техническое творчество* имеет прикладные цели и направлено на удовлетворение практических потребностей человека.

Одной из проблем творчества является его мотивационная структура. **Мотивации** (побуждения) связаны с потребностями, которые делятся на три группы: *биологические, социальные и идеальные (подсознательные).*

Наиболее важным для творчества видом мышления является воображение.

Творческая личность обладает рядом особенностей и прежде всего умением сосредоточить внимание и долго удерживать его на каком-либо вопросе или проблеме.

Общая схема решения научно-технических задач:

- анализ систем задач и выбор конкретной задачи;
- анализ технической системы и разработка ее модели;
- анализ и формулировка условий технической задачи;
- анализ и формулировка условий изобретательской задачи;
- поиск идей решения (принципа действия);
- синтез нового технического решения.

Цель научного исследования – всестороннее, достоверное изучение объекта, процесса или явления; их структуры, связей и отношений на основе разработанных в науке принципов и методов познания, а также получение и внедрение в производство (практику) полезных для человека результатов.

Любое научное исследование имеет свой *объект и предмет*. **Объектом** научного исследования является материальная или идеальная система. **Предмет** – это структура системы, закономерности взаимодействия элементов внутри системы, закономерности развития, различные свойства, качества и т.д.

Научные исследования классифицируются *по видам связи с производством и степени важности для него; целевому назначению; источникам финансирования и длительности ведения.*

Каждую НИР можно отнести к определённому направлению. **Под научным направлением** понимается наука или комплекс наук, в области которых ведутся исследования (например, техническое, социальное и др.).

Структурными единицами научного направления являются *комплексные проблемы, темы и научные вопросы.*

Проблема – это совокупность сложных теоретических и практических задач, решения которых назрели в обществе (противоречие между знанием и незнанием). Она возникает тогда, когда человеческая практика встречает затруднения или даже наталкивается на «невозможность» достижения цели.

Тема научного исследования является составной частью проблемы. В результате исследований по теме получают ответы на определённый круг научных вопросов, охватывающих часть проблемы. **Под научными вопросами** понимается мелкие научные задачи, относящиеся к конкретной теме научного исследования.

Выбор направления, проблемы, темы научного исследования и постановка научных вопросов является чрезвычайно ответственной задачей.

При выборе проблемы и темы научного исследования вначале на основе анализа противоречий исследуемого направления формулируется сама проблема и определяются в общих чертах ожидаемые результаты, затем разрабатывается структура проблемы, выделяются темы, вопросы, исполнители, устанавливается их актуальность.

Выбору темы должно предшествовать тщательное ознакомление с отечественными и зарубежными литературными источниками данной и смежной специальностей.

К **процессам научных исследований** относят формы, средства и методы познания, совокупность которых составляет методику исследований конкретной научной области знаний, представляющий собой один из уровней специальной научной методологии.

Научные исследования начинаются с постановки проблемы на основе обнаружения имеющихся противоречий между потребностью научных знаний об объекте и фактическими знаниями об объекте (процессе, явлении) которыми располагает наука на данный период ее развития.

Постановка проблемы определяет выбор темы исследования, уточняет ее название и обеспечивает обоснование актуальности разработки.

Для уточнения задач исследования осуществляется информационный поиск и также проводится научный поиск, обеспечивающий получение научных результатов.

Решающее значение для научных исследований имеют интеллектуальные способности исследователя, его научное мировоззрение, широта научных знаний, системное мышление, ассоциативное восприятие, информационная культура, творческая активность, толерантность. Научные работники должны хорошо владеть психологией научной работы и грамотной организацией научных исследований.

Таким образом, что процесс научных исследований состоит из четырех последовательных и взаимосвязанных этапов (подпроцессов).

Методика научных исследований это совокупность конкретных форм, методов и средств теоретических и прикладных исследований в определенной области знаний (направления профессиональной деятельности исследователя).

Методика научных исследований выбирается для решения научной задачи в соответствии со сформулированной целью изучения конкретного объекта исследований (структуры, характеристики, информационные связи и другие свойства объекта) с помощью научных принципов и методов познания для получения запланированных результатов, определяющих целесообразную деятельность для достижения определенного эффекта при дальнейшем использовании научных результатов в теории и практике (внедрение в производство, науку, образование и т.п.).

Как ранее указывалось научные исследования начинаются с постановки проблемы, поэтому методика должна позволить вскрыть противоречия между имеющимися знаниями об объекте исследования, которые необходимы для практического решения задачи, т.е. на лицо недостаточность теоретических сведений об объекте исследования для получения необходимого результата .

Постановка проблемы позволяет выбрать тему исследования на основе методики формулирования темы и обоснования ее актуальности для решения конкретной задачи исследования.

Выбор темы, ее формулирование и обоснование актуальности разработки позволяет перейти к следующему этапу – информационному поиску путей решения проблемы на основе методики анализа литературных источников для обобщения имеющихся научных результатов в данной области знаний (обзор литературных источников и использование информационных ресурсов Internet). Результатом будет являться план проведения научных исследований по поставленной проблеме.

Методика научного поиска обычно формируется на основе выбора из уже имеющихся методик, которые ранее применялись для других объектов (процессов, явлений) в смежных областях или если прототип такой методики отсутствует, то разрабатывается новая авторская методика для решения задачи, поставленной в теме.

Методики теоретических исследований определяют общую структуру теоретического исследования и методики решения главной и вспомогательной задач в соответствии с названием темы и поставленной проблемой.

Теоретические исследования являются творческими, направленными на создание новых научных гипотез, глубокое объяснение неизученных явлений или процессов, обобщение отдельных явлений или процессов, обоснование стратегии и тактики научных исследований, а также решении других подобных задач.

Научные исследования базируются на интеллектуальной деятельности (мышлении) человека – исследователя. Важнейшим элементом теоретического исследования является умственный труд. Существует большое количество методик теоретического исследования, поэтому выбор можно делать только в соответствии с конкретной научной проблемой.

Отметим некоторые принципы научного труда, в котором теоретические исследования составляют базисный компонент научного результата:

1. Постоянно думать о предмете исследования. Так И.Ньютон на вопрос о том, как он сумел открыть законы небесной механики, ответил: «Очень просто, я все время думал о них». Из этого принципа следует два практических вывода: нельзя заниматься научной работой только на работе, человек должен думать о предмете своего исследования постоянно.

2. Не работать без плана. При научном исследовании сначала пишется укрупненный план, а затем в процессе теоретических исследований его детализируют и корректируют.

3. Контролировать ход работы в процессе теоретических исследований. По результатам постоянного контроля хода исследований осуществляется корректировка работ и выполняется анализ научных результатов.

Методики экспериментальных исследований – это общая структура, последовательность и приемы выполнения экспериментальных исследований. Экспериментальные исследования подтверждают теоретические понятия, законы, принципы на практике и являются базой для подтверждения достоверности полученных научных результатов сформулированных в гипотезе научных исследований по выбранной теме.

Эксперимент и теория взаимосвязаны:

теория позволяет обосновывать методику эксперимента;

эксперимент позволяет оценить справедливость теории.

Экспериментальные исследования состоят из трех этапов: планирование, эксперимент и анализ (обработка результатов).

В подавляющем большинстве случаев эксперимент является многофакторным опытом. Многофакторность эксперимента дает возможность изложения его стратегии после очередного этапа. Многофакторный эксперимент базируется на общематематическом аппарате, основы которого были заложены в трудах Р.Фишера.

Приступая к эксперименту необходимо: составить программу, обосновать методику, выбрать измерительную аппаратуру, произвести оценку измерений, определить последовательность и составить календарный план.

Математическая теория эксперимента и его планирование, предусматривающее изменение всех исследуемых факторов (измеряемых параметров) по определенному плану и учитывающее их взаимодействие – качественно новый подход к исследованию с применением ЭВМ для обработки результатов факторного эксперимента. Это направление в экспериментальных исследованиях получило название «вычислительный эксперимент».

Важным разделом методики экспериментальных исследований является обработка и анализ данных. Особое внимание в подборе методики эксперимента должно быть уделено математическим методам обработки и удобным формам записи результатов в виде таблиц, графиков, формул, диаграмм и т.п.

Методика оформления научных результатов в виде научного положения, которое является заключающим этапом решения научной проблемы. Формами научной продукции являются: научно-технический отчет; доклад; тезисы; статья; монография; учебное пособие; выпускная квалификационная работа.

Новые научные результаты, имеющие важное теоретическое значение и имеют практическое применение, публикуются в монографиях, статьях, научных отчетах, а учебные материалы в учебниках, учебных пособиях, методических рекомендациях.

Монография – научное издание в виде книги, содержащее всестороннее исследование одной проблемы.

Доклад – краткое изложение содержания основных научных положений, сформулированных автором, выводы и предложения. При подготовке доклада необходимо составить краткие тезисы на 1-2 страницах с изложением цели и содержания идей.

Статья – материал, предоставленный в виде информации для специалистов, которые могут использовать результаты в своей работе.

Учебник – учебное издание в виде книги, содержащее систематическое изложение определенной учебной дисциплины, соответствующее учебной программе, утвержденной официальными органами.

Учебное пособие – учебное издание частично заменяющее или дополняющее учебник.

Выпускная квалификационная работа – результат научных исследований выпускника высшего учебного заведения. ВКР классифицируется как специальная, публично защищаемая квалификационная работа.

Для проведения научных исследований необходимо выбрать оптимальную методику для данной темы (задачи) из имеющихся в науке или разработать новую. Причем необходимо обратить особое внимание на три взаимосвязанных научных понятия: методология, метод, методика, значение которых носит принципиальный характер для бакалавра, выполняющего исследования по теме ВКР.

4. Научные исследования в России

С 2005 года заметно усилилось внимание органов государственной власти к научно-технической и инновационной сфере. 14 сентября 2006 года Постановлением Правительства РФ № 563 создана Правительственная комиссия по вопросам развития промышленности и технологий. Появление данного органа вполне логично ввиду проведенных за последние 2 года масштабных изменений, главным образом, в плане организации инновационных процессов в РФ (появление государственных и смешанных фондов (венчурных, инвестиционных), способствующих внедрению научных разработок, создание особых экономических зон технико-внедренческого типа и т.п.). Главной задачей новой комиссии является «обеспечение взаимодействия органов исполнительной власти по разработке и реализации основных направлений государственной политики по вопросам, касающимся увеличения темпов экономического роста, диверсификации структуры промышленного производства, повышения конкурентоспособности отечественной продукции, развития научно-технического и инновационного потенциала страны, качественного изменения структуры экспорта» .

Создание комиссии, а также широкий круг вопросов, касающихся сферы науки и инноваций, входящий в ее компетенцию, свидетельствует о намерении Правительства качественно изменить структуру российской экономики, сделав развитие высокотехнологичных отраслей основой экономического роста государства. «По замыслу Минэкономразвития, доля «новой экономики» (связь, электроника, IT, точное машиностроение, космические разработки, авиа- и судостроение) должна вырасти с нынешних 5,6% ВВП до 8-10% в 2009-2010 годах». На сегодняшний день основную долю в ВВП России составляют такие отрасли, как топливная промышленность, черная и цветная металлургия, химия и нефтехимия, металлообработка. При этом главным фактором экономического роста стали цены на нефть, которые росли в течение последних трех с половиной лет. Рекордные цены на нефть гарантируют нам высокие показатели экономического роста, однако не позволяют реально судить о его качестве. В этом смысле формируемый Стабилизационный фонд есть не что иное, как инструмент, сдерживающий инфляционные процессы в стране. С другой стороны, именно высокие цены на энергоносители сегодня дают возможность изменить структуру российской экономики, сделав акцент на развитии высокотехнологичных отраслей. Для этого на государственном уровне необходимо принимать меры, которые бы способствовали коммерциализации научных разработок. Именно этап внедрения является в России сегодня наиболее проблематичным. Возможная причина этого кроется в организационной структуре современной российской науки.

На сегодняшний день организационная структура сферы науки и инноваций может быть представлена следующим образом.

Как уже было отмечено, организационным ядром структуры является Правительственная комиссия по вопросам развития промышленности и технологий, которая является координатором мероприятий, проводимых государственными органами исполнительной власти в области науки и инноваций, представленными Министерством образования и науки РФ, Министерством экономического развития и торговли РФ, Министерством информационных технологий и связи. При этом особую роль при проведении научных исследований и реализации разработок играет Российская академия наук (РАН).

5. Проблемы современной фундаментальной науки

Российская академия наук является независимой некоммерческой организацией, имеющей государственный статус. Главным образом РАН занимается проведением фундаментальных исследований в различных областях знаний. При этом при РАН существуют фонды, содействующие реализации наиболее перспективных научных разработок. Это Российский фонд фундаментальных исследований (РФФИ), Российский гуманитарный научный фонд (РГНФ), Фонд содействия развитию малых форм предприятий в научно-технической сфере. В условиях необходимости сохранения целостности государства и стабилизации экономики в первой половине 90-х годов XX века создание этих фондов явилось единственной мерой, предпринятой для поддержки проводимых научных исследований и для содействия внедрению их результатов.

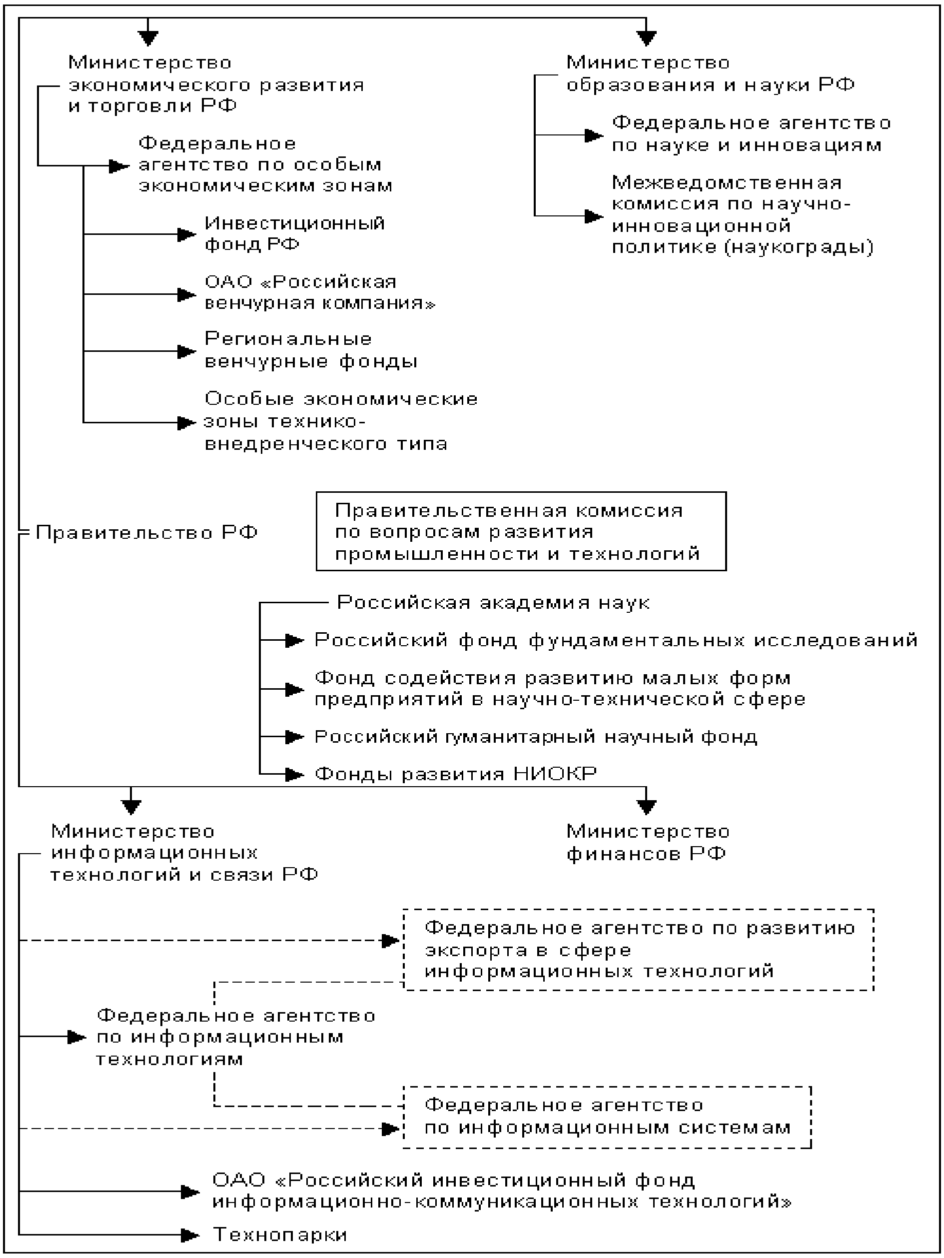


Рисунок 6 – Организационная структура науки в России

РФФИ был образован Указом Президента РФ от 27 апреля 1992 года № 426 «О неотложных мерах по сохранению научно-технического потенциала РФ». Фонд «финансируется из государственного бюджета и поддерживает ученых на безвозвратной основе» [4]. Одним из важных направлений в работе РФФИ является создание баз данных по научным разработкам и предоставление информации о них заинтересованным сторонам. РГНФ выделился из состава РФФИ в 1994 году. Главные задачи фонда — «поддержка гуманитарных научных исследований и распространение гуманитарных научных знаний об обществе»[5]. Финансируется РГНФ за счет ассигнований в размере 0,5% от средств из федерального бюджета, направляемых на развитие науки. Фонд содействия развитию малых форм предприятий в научно-технической сфере образован 3 февраля 1994 года. Начиная с 2001 года, его размер финансирования вырос с 0,5 до 1,5% средств, направляемых на науку из федерального бюджета [6]. Фонд оказывает финансовую поддержку высокоэффективным наукоемким проектам, разрабатываемым малыми предприятиями. Финансирование проектов осуществляется на паритетной основе с малыми инновационными предприятиями. Отбор проектов, поддерживаемых фондами РАН, проводится на конкурсной основе.

Другим не менее важным органом сферы науки и инноваций ввиду последних изменений является Министерство экономического развития и торговли (МЭРТ), которое сосредотачивает внимание на этапе внедрения разработок, осуществляя инвестирование в инновационные проекты. В рамках МЭРТ недавно образовано Федеральное агентство по управлению особыми экономическими зонами [7], которое также занимается Инвестиционным фондом РФ [8]. Среди уже созданных и создаваемых типов особых экономических зон (ОЭЗ) в рамках рассматриваемой нами темы важно выделить технико-внедренческие ОЭЗ. К настоящему моменту созданы четыре таких зоны в различных субъектах РФ, имеющие свою специализацию:

в Дубне — исследования в области ядерных технологий;

в Зеленограде — микроэлектроника;

в Санкт-Петербурге — информационные технологии;

в Томске — новые материалы.

Целью создания ОЭЗ технико-внедренческого типа является государственная поддержка инновационных предприятий путем предоставления резидентам ОЭЗ налоговых льгот и упрощения таможенного режима. При этом государство берет на себя обязательство по строительству инфрастуктуры ОЭЗ. Порядок финансирования создания ОЭЗ устанавливается Соглашением между Правительством РФ в лице МЭРТ, субъектом РФ и администрацией города, на территории которого создана ОЭЗ. Необходимо отметить, что срок действия ОЭЗ составляет 20 лет [9]. Основное требование, которое предъявляется к компаниям, которые желают стать резидентами технико-внедренческой ОЭЗ, — технико-внедренческий характер их деятельности на территории такой ОЭЗ.

Одна из отраслей, на которую Правительство делает ставку, создавая «новую» экономику, — отрасль информационных технологий. Это понятно, ввиду темпов роста, демонстрируемых в последнее время как мировой, так и отечественной IT-отраслью.

Еще одним шагом государства для реализации разработок IT-компаний стала одобренная Правительством государственная программа «Создание в Российской Федерации технопарков в сфере высоких технологий». Действующие до настоящего времени технопарки были созданы в разных отраслях экономики благодаря частным инициативам. Несмотря на то, что Министерство экономического развития и торговли и Министерство информационных технологий и связи РФ обладают достаточно широким кругом полномочий при реализации государственной политики в научно-технической и инновационной сфере, основным органом, разрабатывающим и реализующим политику государства в этой сфере, является Министерство образования и науки РФ и, в частности, Федеральное агентство по науке и инновациям.

Важнейшим условием реализации эффективной государственной научно-технической политики является концентрация научного потенциала, финансовых и материально-технических ресурсов на приоритетных направлениях развития науки и техники.

Под приоритетными направлениями развития науки и техники понимаются основные области исследований и разработок, реализация которых должна обеспечить значительный вклад в социально-экономическое и научно-техническое развитие страны и в достижение за счет этого национальных социально-экономических целей.

В каждом из приоритетных направлений развития науки и техники можно выделить некоторую совокупность критических технологий. Под критическими технологиями понимаются такие технологии, которые носят межотраслевой характер, создают существенные предпосылки для развития многих технологических областей или направлений исследований и разработок и дают в совокупности главный вклад в решение ключевых проблем реализации приоритетных направлений развития науки и техники.

1.2Лекция 2,3,4 (6ч.)**Тема:** Стохастический метод исследования

1.2.1 Вопросы лекции:

1. Статистический материал и его первичная обработка. Эмпирические законы распределения. Полигон частот, гистограмма.
2. Числовые характеристики выборки. Точечные оценки выборочных характеристик. Интервальные оценки, их свойства.
3. Метод доверительных интервалов при заданных условиях. Метод моментов.
4. Статистические гипотезы, ошибки первого и второго рода.. Статистические критерии, их виды, мощность критерия.
5.Выравнивание статистических рядов.

1.2.2 Краткое содержание вопросов:

1. Статистический материал и его первичная обработка. Эмпирические законы распределения. Полигон частот, гистограмма.

Генеральная совокупность и выборка

Предметом математической статистики является изучение случайных величин (или случайных событий) по результатам наблюдений.

Для получения опытных данных необходимо провести обследование соответствующих объектов.

Совокупность всех мысленно возможных объектов данного вида, над которыми проводятся наблюдения с целью получения конкретных значений определённой случайной величины, называется **генеральной совокупностью**.

Генеральную совокупность будем называть **конечной** или **бесконечной** в зависимости от того, конечна или бесконечна совокупность составляющих её элементов.

Часть отобранных объектов из генеральной совокупности называется **выборочной совокупностью** или **выборкой**.

Число N объектов генеральной совокупности и число n объектов выборочной совокупности будем называть **объёмами генеральной и выборочной совокупности** соответственно.

Для того чтобы по выборке можно было достаточно уверенно судить о случайной величине, выборка должна быть **представительной (репрезентативной)**. Репрезентативность выборки означает, что объекты выборки достаточно хорошо представляют генеральную совокупность. Она обеспечивается случайностью отбора.

Существуют несколько способов отбора, обеспечивающих репрезентативность выборки. Рассмотрим некоторые из них.

После того как сделана выборка, все объекты этой совокупности обследуются по отношению к определённой случайной величине и получают наблюдаемые данные.

<u>ВЫБОРКА</u>

<u>с возвратом , без возврата</u>

(случайная повторная) (случайная бесповторная)

Вариационные ряды

Для изучения закономерностей варьирования значений случайной величины опытные данные подвергают обработке.

Операция, заключающаяся в том, что результаты наблюдений над случайной величиной, т.е. наблюдаемые значения случайной величины, располагают в порядке неубывания, называется **ранжированием опытных данных**.

После операции ранжирования опытные данные объединяют в группы так, чтобы в каждой отдельной группе значения случайной величины будут одинаковы.

Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется вариантом (x_i) (**вариантой**), а изменение этого значения – **варьированием**.

Численность отдельной группы сгруппированного ряда наблюдаемых данных называется **частотой** или **весом**(m_i) соответствующей **варианты**.

Отношение частоты данного варианта к общей сумме частот всех вариантов называется **частостью** или долей этой **варианты** (p_i): $p_i = \frac{m_i}{\sum_{i=1}^{\upsilon} m_i}$,

где υ – число вариант. Полагая $n = \sum_{i=1}^{\upsilon} m_i$, где n – объём выборки, имеем: $p_i = \frac{m_i}{n}$.

Заметим, что частость p_i – статистическая вероятность появления варианта x_i.

Дискретным вариационным рядом распределения называется ранжированная совокупность вариантов x_i, с соответствующими им частотами m_i или частостями p_i .

Если изучаемая случайная величина является непрерывной, то ранжирование и группировка наблюдаемых значений зачастую не позволяют выделить характерные черты варьирования её значений. Это объясняется тем, что отдельные значения случайной величины могут как угодно мало отличаться друг от друга и поэтому в совокупности наблюдаемых данных одинаковые значения случайной величины могут встречаться редко, а частоты вариантов мало отличаются друг от друга.

Интервальным вариационным рядом называется упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частостями попаданий в каждый из них значений величины.

Рассмотрим алгоритм построения интервального ряда.

1. Для построения интервального ряда необходимо определить величину частичных интервалов, на которые разбивается весь интервал варьирования наблюдаемых значений случайной величины. Считая, что все частичные интервалы имеют одну и ту же длину, для каждого интервала следует установить его верхнюю и нижнюю границы, а затем в соответствии с полученной упорядоченной совокупностью частичных интервалов сгруппировать результаты наблюдений. Длину частичного интервала h следует выбрать так, чтобы построенный ряд не был громоздким и в то же время позволил выявить характерные черты изменения значений случайной величины.

2. Найдём размах варьирования ряда R: $R = x_{наиб} - x_{наим}$

Выберем число интервалов υ (обычно от 7 до 11).

3. Для более точного определения величины частичного интервала можно воспользоваться **формулой Стерджеса**: $h = \frac{R}{1+3{,}322\lg n}$.

Если h – дробное, то за длину частичного интервала следует брать ближайшее целое число, либо ближайшую простую дробь.

4. За начало первого интервала следует брать величину: $x_{нач} = x_{наим} - 0{,}5h$.

5. Конец последнего интервала ($x_{кон}$) должен удовлетворить условию: $x_{кон} - h \le x_{наиб} < x_{кон}$.

6. Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала h.

7. Определим, сколько значений признака попало в каждый конкретный интервал. При этом в интервал включают значения случайной величины, большие или равные нижней границе и меньшие верхней границы. Иногда интервальный вариационный ряд для простоты исследования условно заменяют дискретным. В этом случае серединное значение i-го интервала принимают за вариант x_i, а соответствующую интервальную частоту m_i – за частоту этой варианты.

2. Числовые характеристики выборки.

Закон распределения (или просто распределение) случайной величины можно задать различными способами. Например, дискретную случайную величину можно задать с помощью или ряда распределения, или интегральной функции, а непрерывную случайную величину – с помощью или интегральной, или дифференциальной функции. Рассмотрим выборочные аналоги этих двух функций.

В теории вероятностей для характеристики распределения случайной величины X служит интегральная функция распределения $F(x) = P(X<x)$. В дальнейшем, если величина X распределена по некоторому закону $F(x)$, будем говорить, что и генеральная совокупность распределена по закону$F(x)$. Введём выборочный аналог функции $F(x)$.

Пусть имеется выборочная совокупность значений некоторой случайной величины X объёма n и каждому варианту из этой совокупности поставлена в соответствие его частость. Пусть, далее, x – некоторое действительное число, а m_x– число выборочных значений случайной величины X, меньших x. Тогда число m_x/n является частостью наблюдаемых в выборке значений величины X, меньших x, т.е. частостью появления события $X<x$. При изменении x в общем случае будет изменяться и величина m_x/n. Это означает, что относительная частота m_x/n является функцией аргумента x. А т.к. эта функция находится по выборочным данным, полученным в результате опытов, то её называют **выборочной** или **эмпирической.**

Выборочной функцией распределения (или **функцией распределения выборки**) называется функция $F(x)$ *, задающая для каждого значения x относительную частоту события $X<x$.

Итак, по определению, $F(x)$ *= m_x/n, где n – объём выборки, m_x – число выборочных значений случайной величины X, меньших x. В отличие от выборочной функции $F(x)$ * интегральную функцию $F(x)$ генеральной совокупности называют **теоретической функцией распределения**. Главное различие функций $F(x)$ и $F(x)$ * состоит в том, что теоретическая функция распределения $F(x)$ определяет вероятность события$X<x$, а выборочная функция – относительную частоту этого события.

Свойство статистической устойчивости частоты, обоснованное теоремой Бернулли, оправдывает целесообразность использования функции $F(x)*$ при большихn в качестве приближённого значения неизвестной функции $F(x)$.

В заключение отметим, что функция $F(x)$ и её выборочный аналог $F(x)*$ обладают одинаковыми свойствами. Действительно, из определения функции $F(x)*$ имеем следующие свойства:

1. $0 \le F^*(x) \le 1$ 2. $F^*(x)$ – неубывающая функция. 3. $F^*(-\infty) = 0$, $F(\infty) = 1$.

Такими же свойствами обладает и функция F(x).

Наблюдаемые данные, представленные в виде вариационного ряда, можно изобразить графически, используя не только функцию $F^*(x)$. К наиболее распространённым видам графического изображения вариационных рядов относятся **полигон** и **гистограмма.** Графическое изображение рядов с помощью полигона или гистограммы позволяет получить наглядное представление о закономерности варьирования наблюдаемых значений случайной величины.

Полигон обычно используют для изображения дискретного вариационного ряд. Для его построения в прямоугольной системе координат наносят точки с координатами (x_i; m_i) или (x_i; p^*_i), где x_i – значение i-го варианта, а m_i (p^*_i) – соответствующие частоты (частости). Затем отмеченные точки соединяют отрезками прямой линии. Полученная ломаная называется **полигоном.**

Если полигон частостей построен по дискретному вариационному ряду дискретной случайной величины, то его называют **многоугольником распределения частостей,** который является выборочным аналогом многоугольника распределения вероятностей. Заметим, что сумма ординат многоугольника распределения частостей, как и у многоугольника распределения вероятностей, равна 1, т.к. $\sum p^*_i=1$.

Гистограмма служит только для изображения интервальных вариационных рядов. Для её построения в прямоугольной системе координат на осиOx откладывают отрезки, изображающие частичные интервалы варьирования, и на этих отрезках, как на основаниях, строят прямоугольники с высотами, равными частотам или частостям соответствующих интервалов. В результате такой операции получают ступенчатую фигуру, состоящую из прямоугольников, которую называют **гистограммой.**

Для графического изображения интервального вариационного ряда можно использовать полигон, если этот ряд преобразовать в дискретный. В этом случае интервалы заменяют их серединными значениями и ставят им в соответствие интервальные частоты (частости). Для полученного дискретного ряда строят полигон.

Построив вариационный ряд и изобразив его графически, можно получить первоначальное представление о закономерностях, имеющих место в ряду наблюдений. Однако на практике зачастую этого недостаточно. Такая ситуация возникает, когда следует уточнить те или иные сведения о ряде распределения или когда имеется необходимость сравнить два ряда и более. При этом следует сравнивать однотипные вариационные ряды, т.е. такие ряды, которые получены при обработке сравнимых статистических данных.

Сравниваемые распределения могут существенно отличаться друг от друга. Они могут иметь различные средние значения случайной величины, вокруг которых группируются в основном остальные значения, или различаться рассеиванием данных наблюдений вокруг указанных значений и т.д. Поэтому для дальнейшего изучения изменения значений случайной величины используют числовые характеристики вариационных рядов. Поскольку эти характеристики вычисляются по статистическим данным (данным, полученным в результате наблюдений), их обычно называют **статистическими характеристиками** или **оценками.**

Пусть собранный и обработанный статистический материал представлен в виде вариационного ряда. Теперь результаты наблюдений над случайной величиной следует подвергнуть анализу и выявить характерные особенности поведения случайной величины. Для этого удобнее всего выделить некоторые постоянные, которые представляли бы вариационный ряд в целом и отражали присущие изучаемой совокупности закономерности.

Некоторые из этих постоянных отличаются тем, что вокруг них концентрируются остальные результаты наблюдений. Такие величины называются **средними величинами.** К ним относятся среднее арифметическое (среднее выборочное), среднее геометрическое, среднее гармоническое и т.д. Однако эти характеристики не отражают «величину изменчивости» наблюдаемых данных, например величину разброса значений признака вокруг среднего арифметического. Другими словами, упомянутые средние величины не отражают вариацию.

Для характеристики изменчивости случайной величины, т.е. вариации, служат показатели вариации. К ним относятся размах варьирования R, среднее квадратическое отклонение, дисперсия и т.д.

2. Точечные оценки выборочных характеристик. Интервальные оценки, их свойства. Метод доверительных интервалов при заданных условиях. Метод моментов.

Выборочная характеристика, используемая в качестве приближённого значения неизвестной генеральной характеристики, называется её **точечной статистической оценкой**.

Среднее арифметическое $\overline{\tilde{O}}$ - это точечная статистическая оценка математического ожидания M(X); D*(X)– оценка дисперсии D(X).

«Точечная» означает, что оценка представляет собой число или точку на числовой оси. «Статистическая» означает, что оценка рассчитывается по результатам наблюдений, т.е. по собранной исследователем статистике. Далее слово «статистическая» будет опускаться.

Обозначим через Θ («тэта») некоторую генеральную характеристику (ею может быть и MX, и любая другая числовая характеристика случайной величины X). Её числовое значение неизвестно, однако предложен некоторый алгоритм или формула вычисления точечной оценки $\Theta_{(n)}$ этой характеристики по результатам $X_1, X_2, \ldots, X_n$ наблюдений величины X. Обозначая буквой f этот алгоритм, запишем $\Theta^*_{(n)}=f(X_1, X_2, \ldots, X_n)$. (3)

Подставив в (3) вместо $X_1, X_2, \ldots, X_n$ конкретные результаты наблюдений (конкретные числа), получим число, которое и принимают за приближённое значение неизвестной генеральной характеристики Θ. Найти погрешность этого приближения нельзя, поскольку числовое значение характеристики Θ неизвестно. Чтобы ответить на вопрос, хорошо или нет найденное приближение, рассмотрим оценку $\Theta^*_{(n)}$ с других позиций.

Пусть в формуле (3) $X_1, X_2, \ldots, X_n$ – не конкретные числа, а лишь обозначения тех результатов наблюдений, которые мы хотели бы получить. Но результат каждого отдельного наблюдения случайной величины случаен, т.е. $X_1, X_2, \ldots, X_n$– это случайные величины, поэтому и оценка $\Theta^*_{(n)}$ также величина случайная; следовательно, можно говорить о её математическом ожидании ($M(\Theta^*_{(n)})$), дисперсии ($D(\Theta^*_{(n)})$) и законе распределения. Интерпретация оценки $\Theta^*_{(n)}$ как случайной величины позволяет сформулировать свойства, которыми должна была обладать оценка, чтобы её можно было считать хорошим приближением к неизвестной генеральной характеристике. Это свойства состоятельности, несмещённости и эффективности.

Оценка $\Theta^*_{(n)}$ генеральной характеристики Θ называется **состоятельной**, если для любого ε>0 выполняется равенство $\lim_{n\to\infty} P(\left|\Theta^*_{(n)} - \Theta\right| < \varepsilon)=1$. (4)

Поясним смысл равенства (4). Пусть ε – очень малое положительное число. Тогда равенство (4) означает, что чем больше число наблюдений n, тем больше уверенность (вероятность) в незначительном по абсолютной величине отклонении оценки $\Theta^*_{(n)}$ от неизвестной характеристики Θ или короче: чем больше объём исходной информации, тем «ближе мы к истине». Если это так, то $\Theta^*_{(n)}$ – состоятельная оценка.

«Хорошая» оценка обязательно должна обладать свойством состоятельности. В противном случае оценка не имеет практического смысла: увеличение объёма исходной информации не будет «приближать нас к истине». Поэтому свойство состоятельности следует проверять в первую очередь.

Оценка $\Theta^*_{(n)}$ генеральной характеристики Θ называется **несмещённой**, если для любого фиксированного числа наблюдений n выполняется равенство $M(\Theta^*_{(n)})=\Theta$, (5)
т.е. математическое ожидание оценки равно неизвестной характеристике.

Несмещённая оценка $\Theta^*_{(n)}$ характеристики Θ называется **несмещённой эффективной**, если она среди всех прочих несмещённых оценок той же самой характеристики обладает наименьшей дисперсией.

Метод нахождения оценки неизвестного параметра, основанный на требовании максимизации функции правдоподобия, называется **методом максимального правдоподобия**, а найденная этим методом оценка – **оценкой максимального правдоподобия**.

Функции L и lnL, рассматриваемые как функции параметра λ, достигают максимума при одном и том же значении λ, т.к. lnL – монотонно возрастающая функция. Поэтому вместо отыскания максимума функции L находят (что удобнее) максимум функции *lnL*. Функция *lnL* называется **логарифмической функцией правдоподобия**.

Для $L(X_1, X_2, \ldots X_n; \lambda) = \dfrac{\lambda^{i=1} e^{-n\lambda}}{\tilde{O}_1!\,\tilde{O}_2!\ldots\tilde{O}_n!}$ логарифмическая функция правдоподобия имеет вид:

$$lnL(X_1, X_2, \ldots X_n; \lambda) = ln\frac{\lambda^{\sum\limits_{i=1}^{n} x_1} e^{-n\lambda}}{X_1!\,X_2!\ldots X_n!} = \left(\sum_{i=1}^{n} X_i\right)\ln\lambda - n\lambda - \ln(X_1!) - \ln(X_2!) - \ldots - - -\ln(X_n!).$$

Найдём точку максимума этой функции, рассматривая её как функцию параметра λ. Для этого:

найдём производную функции ln$L$по λ: $\dfrac{\partial \ln L}{\partial \lambda} = -\dfrac{\sum\limits_{i=1}^{n} X_i}{\lambda} n$;

приравняв производную нулю, определим критическую точку – корень полученного уравнения – **уравнения правдоподобия**: $\dfrac{\sum\limits_{i=1}^{n} X_i}{\lambda} - n = 0 \to \lambda_{\hat{e}\eth} = \sum\limits_{i=1}^{n} X_i / n$;

найдём вторую производную функции ln$L$ и её значение в точке $\lambda_{кр}$:

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = \frac{\sum\limits_{i=1}^{n} X_i}{\lambda^2};\ \frac{\partial^2 \ln L(\lambda_{\hat{e}\eth})}{\partial \lambda^2} = \frac{\sum\limits_{i=1}^{n} X_i}{\lambda_{\hat{e}\eth}{}^2} = -\frac{n^2}{\sum\limits_{i=1}^{n} X_i}.$$

Итак, всегда $\lambda_{кр} = \sum\limits_{i=1}^{n} X_i / n$ - это точка масимума функции ln$L$ (или L), поэтому она и является оценкой $\lambda_{мп}$ максимального правдоподобия для неизвестного параметра λ, т.е.

$$\lambda^*_{\grave{\imath}\ddot{\imath}} = \sum_{i=1}^{n} X_i / n = \overline{X}.$$

Интервальные оценки параметров статистического распределения. Доверительные вероятности

Вычисляя на основании результатов наблюдений точечную характеристику Θ* неизвестной числовой характеристики Θ, мы понимаем, что величина Θ* является лишь приближённым значением характеристики Θ. Если для большого числа наблюдений точность приближения бывает достаточной для практических выводов (в силу несмещённости, состоятельности и эффективности «хороших» оценок), то для выборок небольшого объёма вопрос о точности оценок очень важен. В математической статистике он решается следующим образом. По сделанной выборке находится точечная оценка Θ* неизвестной характеристики Θ, затем задаются вероятностью γ и по определённым правилам находят такое число ε>0, чтобы выполнялось соотношение

$$P(\underline{\Theta^*-\varepsilon} < \Theta < \underline{\Theta^*+\varepsilon}) = \gamma. \qquad (8)$$

Соотношению (8) тождественно соотношению

$$P(\left|\Theta^* - \Theta\right| < \varepsilon) = \gamma, \qquad (9)$$

из которого видно, что абсолютная погрешность оценки Θ не превосходит числа ε. Это верно с вероятностью, равной γ. Число ε называется **точностью оценки** Θ* (чем меньше ε, тем выше точность оценки), числа Θ_1 и Θ_2 называются **доверительными границами**, интервал (Θ_1, Θ_2) – **доверительным интервалом** или **интервальной оценкой** характеристики Θ, вероятность γ называется **доверительной вероятностью** или **надёжностью** интервальной оценки.

В соотношении (8) случайными величинами являются доверительные границы Θ_1 и Θ_2: во-первых, эти границы могут изменяться при переходе от одной выборки к другой хотя бы потому, что при этом изменяется значение оценки Θ^*; во-вторых, при фиксированной выборке границы Θ_1 и Θ_2 изменяются при изменении вероятности γ, поскольку ε выбирается в зависимости от γ. Генеральная же характеристики Θ – постоянная величина. Поэтому соотношение (8) следует читать так: «вероятность того, что интервал (Θ_1, Θ_2) накроет характеристику Θ, равна γ»; именно «интервал накроет характеристику», а не «характеристика попадёт в интервал».

Надёжность γ принято выбирать равной 0,95; 0,99; 0,999. Тогда событие, состоящее в том, что интервал (Θ_1, Θ_2) накроет характеристику Θ, будет практически достоверным. Также практически достоверным является событие, состоящее в том, что погрешность оценки Θ^* меньше ε, или, иначе, точность оценки Θ^* больше ε.

В соотношении (8) границы Θ_1 и Θ_2 симметричны относительно точечной оценки Θ^*. Обратим внимание на то, что не всегда удаётся построить границы с таким свойством.

Поскольку довольно часто встречаются нормально распределённые случайные величины, построим интервальные оценки для параметров нормального распределения – математического ожидания a и среднего квадратического отклонения σ.

Интервальные оценки параметров нормального распределения

Обозначим через X случайную величину, имеющую нормальный закон распределения с параметрами a и σ, т.е. $X = N(a, \sigma)$. Будем предполагать, что наблюдения над этой величиной независимы и проводятся в одинаковых условиях, т.е. возможные результаты X_1, X_2, …, X_n этих наблюдений обладают следующими свойствами:

$X_1, X_2, \ldots, X_n$ – независимые случайные величины;
закон распределения любой из величин $X_1, X_2, \ldots, X_n$ совпадает с законом распределения величины X, т.е.

$$X_1 = N(a, \sigma), X_2 = N(a, \sigma), \ldots, X_n = N(a, \sigma). \quad (10)$$

Интервальная оценка математического ожидания нормального распределения при известной дисперсии

Итак, $X = N(a, \sigma)$, причём математическое ожидание a неизвестно, а дисперсия σ^2 известна. По наблюдениям найдём точечную оценку $\overline{\tilde{O}} = \sum_{i=1}^{n} X_i / n$ математического ожидания a. Зададимся вероятностью γ и попробуем найти такое число ε, чтобы выполнялось соотношение

$$P(\overline{X} - \varepsilon < a < \overline{X} + \varepsilon) = \gamma. \quad (11)$$

Интервальная оценка математического ожидания такова:

$$(\overline{X} - u_\gamma \sigma / \sqrt{n}\ ,\ \overline{X} + u_\gamma \sigma / \sqrt{n}\ ). \quad (12)$$

Полученный результат имеет следующий смысл: с вероятностью γ можно быть уверенным в том, что интервал (12) накроет среднее математическое ожидание.

4. Статистические гипотезы, ошибки первого и второго рода. Статистические критерии, их виды, мощность критерия.

Под **статистической гипотезой** понимают всякое высказывание о генеральной совокупности (случайной величине), проверяемое по выборке (по результатам наблюдений). Примером статистических гипотез являются следующие высказывания: генеральная совокупность, о которой мы располагаем лишь выборочными сведениями, имеет нормальный закон распределения или генеральная средняя (математическое ожидание случайной величины) равна 5. Не располагая сведениями о всей генеральной совокупности, высказанную гипотезу сопоставляют, по определённым правилам, с выборочными сведениями, и делают вывод о том, можно принять гипотезу или нет. Процедура сопоставления высказанной гипотезы с выборочными данными называется **проверкой гипотезы**.

Рассмотрим этапы проверки гипотезы и используемые при этом понятия.

Этап 1. Располагая выборочными данными X_1, X_2, …, X_n и руководствуясь конкретными условиями рассматриваемой задачи, формулируют гипотезу H_0, которую называют **основной** или **нулевой**, и гипотезу H_1, **конкурирующую** с гипотезой H_0.

Термин «конкурирующая» означает, что являются противоположными следующие два события:

- по выборке будет принято решение о справедливости для генеральной совокупности гипотезы H_0;

- по выборке будет принято решение о справедливости для генеральной совокупности гипотезы H_1.

Гипотезу H_1 называют также **альтернативной**.

Например, если нулевая гипотеза такова: математическое ожидание равно 5, – то альтернативная гипотеза может быть следующей: математическое ожидание меньше 5, что записывается следующим образом: $H_0 : M(X) = 5$; $H_1 : M(X) < 5$.

Этап 2. Задаются вероятностью α («альфа»), которую называют **уровнем значимости**. Поясним её смысл:

Решение о том, можно ли считать высказывание H_0 справедливым для генеральной совокупности, принимается по выборочным данным, т.е. по ограниченному ряду наблюдений; следовательно, это решение может быть ошибочным. При этом может иметь место ошибка двух родов:

- отвергают гипотезу H_0, или, иначе, принимают альтернативную гипотезу H_1, тогда как на самом деле гипотеза H_0 верна – это **ошибка первого рода**;

- принимают гипотезу H_0, тогда как на самом деле высказывание H_0 неверно, т.е верной является гипотеза H_1 – это **ошибка второго рода**.

Так вот, уровень значимости α – это вероятность ошибки первого рода, т.е. $\alpha = P_{H_0}(H_1)$, (13)

где $P_{H_0}(H_1)$ – вероятность того, что будет принята гипотеза H_1, если на самом деле в генеральной совокупности верна гипотеза H_0. Вероятность α задаётся заранее, разумеется, малым числом, поскольку это вероятность ошибочного заключения, при этом обычно используют некоторые стандартные значения: 0,05; 0,01; 0,005; 0,001. Например, α = 0,05 означает следующее: если гипотезу H_0 проверять по каждой из 100 выборок одинакового объёма, то в среднем в 5 случаях из 100 мы совершим ошибку первого рода.

Вероятность ошибки второго рода обозначают β, т.е. $\beta = P_{H_1}(H_0)$, (14)

где $P_{H_1}(H_0)$ – вероятность того, что будет принята гипотеза H_0, если на самом деле верна гипотеза H_1. Зная α, можно найти вероятность β. Сказанное иллюстрирует табл. 6.

Таблица 6.

Решение, принимаемое о гипотезе H_0 по выборке. Верна ли гипотеза H_0 или нет?	Гипотеза отвергается, т.е принимается гипотеза H_1	Гипотеза H_0 принимается
Гипотеза H_0 верна	Ошибка первого рода, её вероятность $\text{Đ}_{i\ 0}\ (H_1) = \alpha$	Правильное решение, его вероятность $\text{Đ}_{i\ 0}\ (H_0) = 1 - \alpha$
Гипотеза H_0 неверна, т.е. верна гипотеза H_1	Правильное решение, его вероятность $\text{Đ}_{i\ 1}\ (H_1) = 1 - \beta$	Ошибка второго рода, её вероятность $\text{Đ}_{i\ 1}\ (H_0) = \beta$

Обратим внимание на то, что в результате проверки гипотезы относительно гипотезы H_0 может быть принято и правильное решение. Существует правильное решение двух следующих видов:

- принимают гипотезу H_0, тогда как и в действительности, в генеральной совокупности, она имеет место; вероятность этого решения P_{H_0} $(H_0) = 1 - \alpha$;

- не принимают гипотезу P_{H_0} $(H_0) = 1 - \alpha$ (т.е. принимают гипотезу H_1), тогда как на самом деле гипотеза H_0 неверна (т.е. верна гипотеза H_1); вероятность этого решения P_{H_1} $(H_1) = 1 - \beta$.

Этап 3. Находят величину φ такую, что:

- её значения зависят от выборочных данных X_1, X_2, …, X_n, т.е. для которой справедливо равенство $\varphi = \varphi(X_1, X_2, \ldots, X_n)$;
- её значения позволяют судить о «расхождении выборки с гипотезой H_0»;
- и она, будучи величиной случайной в силу случайности выборки X_1, X_2, …, X_n, подчиняется при выполнении гипотезы H_0 некоторому известному, затабулированному закону распределения.

Статистический критерий. Мощность критерия

Величину φ называют **критерием**.

Отметим, что в основе метода построения критерия лежит понятие функции правдоподобия.

Этап 4. Далее рассуждают так. Т.к. значения критерия позволяют судить о «расхождении выборки с гипотезой H_0», то из области допустимых значений критерия φ следует выделить подобласть ω таких значений, которые свидетельствовали бы о существенном расхождении выборки с гипотезой H_0, и, следовательно, о невозможности принять гипотезу H_0. Подобласть ω называют **критической областью**. Допустим, что критическая область выделена. Тогда руководствуются следующим правилом: если вычисленное по выборке значение критерия φ попадает в критическую область, то гипотеза H_0 отвергается и принимается гипотеза H_1. При этом следует понимать, что такое решение может оказаться ошибочным: на самом деле гипотеза H_0 может быть справедливой. Т.обр., ориентируясь на критическую область, можно совершить ошибку первого рода, вероятность которой задана заранее и равна α. Отсюда вытекает следующее требование к критической области ω:

вероятность того, что критерий φ примет значение из критической области ω, должна быть равна заданному числу α, т.е.

$$P(\varphi \in \omega) = \alpha. \qquad (15)$$

Однако критическая область равенством (15) определяется неоднозначно. Действительно, представив себе график функции плотности $f_\varphi(x)$ критерия φ, нетрудно понять, что на оси абсцисс существует бесчисленное множество областей-интервалов таких, что площади построенных на них криволинейных трапеций равны α, т.е. областей, удовлетворяющих требованию (15). Поэтому кроме требования (15) выдвигается следующее требование: критическая область ω должна быть расположена так, чтобы при заданной вероятности α ошибки первого рода вероятность β ошибки второго рода была минимальной.

Возможны три вида расположения критической области (в зависимости от вида нулевой и альтернативной гипотез, вида и расположения критерия φ):

правосторонняя критическая область, состоящая из интервала $(x^{кр}_{пр,\,\alpha}, +\infty)$, где точка $x^{кр}_{пр,\,\alpha}$ определяется из условия

$$P(\varphi > x^{кр}_{пр,\,\alpha}) = \alpha \qquad (16)$$

и называется **правосторонней критической точкой**, отвечающей уровню значимости α;

левосторонняя критическая область, состоящая из интервала $(-\infty, x^{кр}_{лев,\,\alpha})$, где точка $x^{кр}_{лев,\,\alpha}$ определяется из условия

$$P(\varphi < x^{кр}_{лев,\,\alpha}) = \alpha \qquad (17)$$

и называется **левосторонней критической точкой**, отвечающей уровню значимости α;

двусторонняя критическая область, состоящая из следующих двух интервалов:
$(-\infty, x^{кр}_{лев,\,\alpha/2})$

По значению критерия φ судят о «расхождении выборочных данных с гипотезой H_0». Естественно, что гипотеза H_0 должна быть отвергнута, если расхождения велики; именно этим объясняется включение в критическую область больших значений критерия φ (больше, чем критическая точка).

Включение же в ряде случаев в критическую область малых значений критерия φ (меньше, чем критическая точка) на первый взгляд противоречит смыслу этой величины. Однако не следует забывать, что φ – случайная величина (она зависит от результатов наблюдений $X_1, X_2, \ldots, X_n$, которые случайны), поэтому маловероятно появление не только слишком больших, но и слишком малых её значений и их следует включить в критическую область.

Этап 5. В формулу критерия $\varphi = \varphi(X_1, X_2, \ldots, X_n)$ вместо $X_1, X_2, \ldots, X_n$ подставляют конкретные числа, полученные в результате наблюдений, и подсчитывают числовое значение $\varphi_{чис}$ критерия.

Если $\varphi_{чис}$ попадает в критическую область ω, то гипотеза H_0 отвергается и принимается гипотеза H_1. Поступая таким образом, следует понимать, что можно допустить ошибку с вероятностью α.

Если $\varphi_{чис}$ не попадает в критическую область, гипотеза H_0 не отвергается. Но это вовсе не означает, что H_0 является единственно подходящей гипотезой: просто расхождение между выборочными данными и гипотезой H_0 невелико, или, иначе, H_0 не противоречит результатами наблюдений; однако таким же свойством наряду с H_0 могут обладать и другие гипотезы.

5. Выравнивание статистических рядов. Критерий согласия Пирсона

Выше рассматривались гипотезы, относящиеся к отдельным параметрам распределения случайных величин, причём модели законов распределения этих величин представлялись известными. Однако во многих практических задачах модель закона распределения заранее не известна и возникает задача выбора модели, согласующейся с результатами наблюдений над случайной величиной.

Пусть высказано предположение, что неизвестная функция распределения $F_X(x)$ исследуемой случайной величины X имеет вполне определённую модель $F_{\text{теор}}(x)$, т.е. высказана гипотеза

$$H_0 : F_X(x) = F_{\text{теор}}(x). \qquad (18)$$

В качестве теоретической модели $F_{\text{теор}}(x)$ может быть рассмотрена нормальная, биномиальная или какая-либо другая модель. Это определяется сущностью изучаемого явления, а также результатом предварительной обработки наблюдений над случайной величиной (формой графика вариационного ряда, соотношениями между выборочными характеристиками и т.д.).

Критерии, с помощью которых проверяется гипотеза (19), называются **критериями согласия**. Рассмотрим лишь один из них, использующий χ^2-рас-пределение и получивший название **критерия согласия Пирсона**.

Критерий предполагает, что результаты наблюдений сгруппированы в вариационный ряд. Для определённости положим, что это дискретный вариационный ряд с числом групп, равным ν (см. строки 1 и 2 табл. 7).

Таблица 7.

x_i	x_1	…	$x_{\nu-1}$	x_ν
m_i	m_1	…	$m_{\nu-1}$	m_ν
$p_i^{\text{теор}} = P(X = x_i)$	$p_1^{\text{теор}} = P(X = x_1)$	…	$p_{\nu-1}^{\text{теор}} = P(X = x_{\nu-1})$	$p_\nu^{\text{теор}} = 1 - p_1^{\text{теор}} - \ldots - p_{\nu-1}^{\text{теор}}$
$m_i^{\text{теор}} = np_i^{\text{теор}}$	$m_1^{\text{теор}} = np_1^{\text{теор}}$	…	$m_{\nu-1}^{\text{теор}} = np_{\nu-1}^{\text{теор}}$	$m_\nu^{\text{теор}} = np_\nu^{\text{теор}}$

Однако, прежде чем рассматривать сам критерий Пирсона, вспомним параметрическое оценивание закона распределения. Последовательность оценивания такая: формулируют гипотезу о модели закона распределения случайной величины; по результатам наблюдений находят оценки неизвестных параметров этой модели (допустим, что число неизвестных параметров равно l); вместо неизвестных параметров подставляют в модель найденные оценки. В результате предполагаемая модель закона оказывается полностью определённой и, используя её, рассчитывают вероятности $p_i^{\text{теор}} = P(X = x_i)$ того, что случайная величина X примет зафиксированные в наблюдениях значения x_i, i=1, 2, …, ν – 1; эти вероятности называют **теоретическими**. Обратим внимание на следующее обстоятельство: т.к. сумма вероятностей ряда распределения должна быть равна единице, т.е.

$$\sum_i p_i^{\text{òåîð}} = 1, \qquad (19)$$

то полагаем вероятность $p_\nu^{\text{теор}} = 1 - p_1^{\text{теор}} - p_2^{\text{теор}} - \ldots - p_{\nu-1}^{\text{теор}}$. Теоретические вероятности записаны в строке 3 табл. 7. Теперь найдём теоретические частоты $m_i^{\text{теор}} = np_i^{\text{теор}}$; они записаны в строке 4 табл. 7.

Обратим внимание на следующее: критерий согласия Пирсона можно использовать только в том случае, когда

$$m_i^{\text{теор}} \geq 5,\ i=1, 2, \ldots, \nu. \qquad (20)$$

Поэтому ту группу вариационного ряда, для которой это условие не выполняется, объединяют с соседней и соответственно уменьшают число групп; так поступают до тех пор, пока для каждой новой группы $m_i^{\text{теор}}$ будет не меньше 5. Новое число групп, как и прежде, обозначим символом ν.

Оказывается, что если предполагаемая модель закона распределения действительно имеет место, т.е. верна гипотеза (18), и если к тому же выполняются условия (19) и (20), то величина

$$\varphi = \sum_{i=1}^{\nu} \frac{\left(m_i - m_i^{\text{теор}}\right)^2}{m_i^{\text{теор}}} \qquad (21)$$

будет иметь χ^2-распределение с числом степеней свободы $k = \nu - l - 1$, т.е.

$$\varphi = \sum_{i=1}^{\nu} \frac{\left(m_i - m_i^{\text{òíîð}}\right)^2}{m_i^{\text{òåîð}}} = \chi^2(k = \nu - l - 1),$$

где ν – число (новое) групп вариационного ряда; l – число неизвестных параметров предполагаемой модели, оцениваемых по результатам наблюдений (если все параметры предполагаемого закона известны точно, то l = 0). Величину (21) и называют **критерием согласия χ^2** или **критерием согласия Пирсона**.

Далее поступаем так же, как обычно при проверке гипотез. Задаёмся уровнем значимости α. Зная распределение критерия φ, находим критическую область, как правило, это область правосторонняя, т.е. она имеет вид $(x^{кр}_{пр,\,\alpha}, +\infty)$; найдём числовое значение $\varphi_{чис}$ критерия (21). Если $\varphi_{чис}$ попадает в интервал $(x^{кр}_{пр,\,\alpha}, +\infty)$, то делаем вывод о неправомерности гипотезы H_0 (18); при этом не следует забывать, что этот вывод может оказаться ошибочным (на самом деле в генеральной совокупности гипотеза H_0 (18) имеет место) и вероятность того, что вывод ошибочен, равна α.

Если $\varphi_{чис}$ не попадает в интервал $(x^{кр}_{пр,\,\alpha}, +\infty)$, то гипотеза H_0 (18) не отвергается.

В заключение приведём схему определения точки $x^{кр}_{пр,\,\alpha}$:

$$\left.\begin{aligned} \alpha \to \gamma = 1 - \alpha \\ l, \nu \to k = \nu - l - 1 \end{aligned}\right\} \longrightarrow \quad \chi^2_{\gamma} \to x^{кр}_{пр,\,\alpha} = \chi^2_{\gamma}. \qquad (22)$$

1.3 Лекция 5-6 (4 ч.)

Тема: Оптимизационные задачи.

1.3.1 Вопросы лекции:

1. ЗЛП. Графический метод решения ЗЛП, симплекс-метод
2. Транспортная задача, методы ее решения.
3. Задачи нелинейного программирования и методы их решения

1.3.2 Краткое содержание вопросов:

1. ЗЛП. Графический метод решения ЗЛП, симплекс-метод

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

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

при ограничениях

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

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

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

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

$$x_j \text{ - } \qquad \text{произвольные} \qquad (j = n_1 + 1,\dots,n) \tag{2.6}$$

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:

$$\max\ Z = cx \tag{2.7}$$

$$A_1 x_1 + A_2 x_2 + \dots + A_n x_n = A_0 \tag{2.8}$$

$$x_1 \ge 0, x_2 \ge 0,\dots, x_n \ge 0 \tag{2.9}$$

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 $A_1, A_2, \dots, A_n$ содержит m линейно независимых векторов $A_1, A_2, \dots, A_m$, то допустимый план

$$\bar{x} = (x_1; x_2; \dots; x_m; \underbrace{0; 0; \dots; 0}_{n-m}) \qquad (2.10)$$

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.

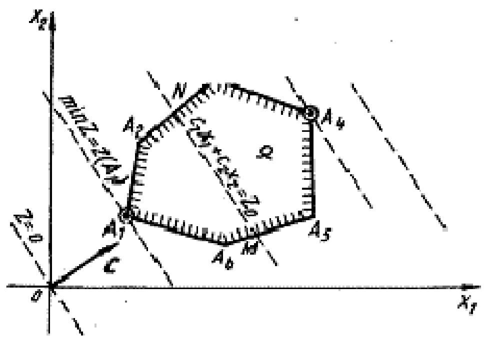
Пусть дана задача

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

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

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

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 $c_1x_1 + c_2x_2 = Z$.

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

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

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

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

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

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

Пусть далее система ограничений имеет вид

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

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

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

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

Поэтому, вообще говоря, базисный план $\vec{x}_o = (\underbrace{0;\ldots;0}_{n};-b_1;-b_2;\ldots;-b_m)$ 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

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

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

$$x_j \ge 0 \quad (j = 1,..,n), \qquad (2.18)$$

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

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

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

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

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

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

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

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

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.

Признаки оптимальности

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 $(i = 1,...,m)$. 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 $(j = 1,...,n)$ 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:

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

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

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

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 $y_1, y_2, \dots, y_m$.

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:

$$\min\ f = b_1y_1 + b_2y_2 + \dots + b_my_m \quad (2.26)$$

Требование 2 предприятия, реализующего отходы сырья, можно сформулировать в виде системы ограничений. Предприятие откажется от выпуска каждой единицы продукции первого вида, если $a_{11}y_1 + a_{12}y_2 + \dots + a_{m1}y_m \ge c_1$, 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:

$$\left.\begin{array}{l} a_{11}y_1 + a_{12}y_2 + \dots + a_{m1}y_m \ge c_1 \\ a_{12}y_1 + a_{22}y_2 + \dots + a_{m2}y_m \ge c_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}y_1 + a_{2n}y_2 + \dots + a_{mn}y_m \ge c_n \end{array}\right\} \quad (2.27)$$

По смыслу задачи оценки не должны быть отрицательными:

$$y_1 \ge 0, y_2 \ge 0, \dots, y_m \ge 0. \quad (2.28)$$

Переменные $y_i\ (i = 1, \dots, m)$ 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 $\le$. Двойственная задача решается на минимум, и ее система ограничений имеет вид неравенств типа $\ge$.

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 $\vec{x}=(x_1,\dots,x_n)$ и $\vec{y}=(y_1,\dots,y_m)$ прямой и двойственной ЗЛП справедливо неравенство $z(\vec{x})\le f(\vec{y})$, т.е.

$$\sum_{j=1}^{n} c_j x_j \le \sum_{i=1}^{m} b_i y_i \qquad (2.29)$$

– 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:

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

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

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 экономических 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 $Z = c_1x_1 + \ldots + c_nx_n$. 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: $x_j \ge 0 \quad (j=\overline{1,n})$.

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

$$\max Z = \sum_{j=1}^{n} c_j x_j \qquad (2.33)$$

при ограничениях:

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

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

Так как переменные x_j входят в функцию $z(\overline{x})$ 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		
	А	В	С
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 производственными 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: $x_{ij} \geq 0, \quad x_{ij} - целые$.

2. 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:

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

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

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1,\dots, n);$$

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

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

$$\begin{cases} x_{11}+x_{12}+x_{13}+x_{14}=800; \\ x_{21}+x_{22}+x_{23}+x_{24}=900; \\ x_{31}+x_{32}+x_{33}+x_{34}=600; \\ x_{11}+x_{21}+x_{31}=300; \\ x_{12}+x_{22}+x_{32}=600; \\ x_{13}+x_{23}+x_{33}=650; \\ x_{14}+x_{24}+x_{34}=750; \end{cases}$$
$$x_{ij}\ge 0.$$

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 неудовлетворёнными 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:

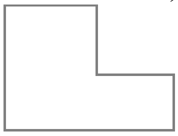
Таблица № 2.4

	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

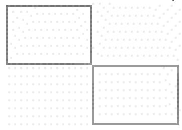
На этом способе уменьшения стоимости в дальнейшем и будет основан алгоритм оптимизации плана перевозок. **Циклом** в транспортной задаче мы будем называть несколько занятых клеток, соединённых замкнутой, ломанной линией, которая в каждой клетке совершает поворот на 90°. Существует несколько вариантов 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 перемещении по нему k единиц груза стоимость перевозок увеличиться на $k\gamma$. 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 $k\gamma$. 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 $x_{ij}=0$, $\alpha_i+\beta_j=\check{c}_{ij}\leq c_{ij}$, 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 + x	4	6 6	8	0
β_j	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_iи 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.

3. Задачи нелинейного программирования и методы их решения

Метод неопределенных множителей Лагранжа является классическим методом решения задач математического программирования. При практическом применении метода могут встретиться значительные вычислительные трудности, сужающие область его исполь-

зования. Метод Лагранжа является аппаратом, используемым для обоснования различных современных численных методов.
Рассмотрим задачу оптимизации
max (min) $z = f(x)$; (1)
$\varphi_i(x) = b_i$ $(i = 1,...,m)$, $x = (x_1, x_2,..., x_n)$. (2)
Последовательность решения данной задачи методом неопределенных множителей Лагранжа:
1) составить функцию Лагранжа
$L(x_1,..., x_n, \lambda_1,..., \lambda_m) = f(x_1,..., x_n) + \sum_{i=1}^{m} \lambda_i \left(b_i - \varphi_i(x_1, \dots, x_n)\right)$; (3)
$\lambda_1,..., \lambda_m$ – множители Лагранжа;
2) найти частные производные функции Лагранжа по всем переменным $x_1, x_2,..., x_n, \lambda_1,..., \lambda_m$

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial \varphi_i}{\partial x_j} = 0 \quad (j = 1, 2, \dots, n)$$

(4)

$$\frac{\partial L}{\partial \lambda_i} = b_i - \varphi_i = 0 \quad (i = 1, 2, \dots, m)$$

Получили систему уравнений(6.6), состоящую из $n + m$ уравнений. Решить полученную систему (если это окажется возможным!) и найти таким образом все стационарные точки функции Лагранжа;
3) из стационарных точек, взятых без координат $\lambda_1,..., \lambda_m$, выбрать точки, в которых функция $f(x)$ имеет условные локальные экстремумы при наличии ограничений (4).

Признаком существования минимума функции $f(x)$ в стационарнойточке x^* является выполнение в ней достаточных условий минимума – выпуклости функции в окрестности этой точки, что может быть представлено следующим образом [*]:

$$\frac{\partial^2 f(x)}{\partial x_1^2}\Big|_{x=x^*} > 0;$$

$$\begin{vmatrix} \frac{\partial^2 f(x)}{\partial x_1^2}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2}\Big|_{x=x^*} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_2^2}\Big|_{x=x^*} \end{vmatrix} > 0;$$

$$\begin{vmatrix} \frac{\partial^2 f(x)}{\partial x_1^2}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_3}\Big|_{x=x^*} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_2^2}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3}\Big|_{x=x^*} \\ \frac{\partial^2 f(x)}{\partial x_3 \partial x_1}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_3 \partial x_2}\Big|_{x=x^*} & \frac{\partial^2 f(x)}{\partial x_3^2}\Big|_{x=x^*} \end{vmatrix} > 0,$$

т.е. все определители должны быть положительны.
Этот метод можно обобщить и на случай, когда переменные не отрицательны и некоторые ограничения заданы в форме неравенств.
Методы спуска. Численные (поисковые) методы играют существенную роль при решении многих прикладных задач, в том числе электроэнергетики. Это обусловлено рядом причин, среди которых главное место занимает разнообразие целевых функций и ограничений, а также форм их задания.

Допустим, что рассматривается задача безусловной минимизации целевой функции F(x). Сущность всех методов решения этой задачи, о которых далее пойдет речь, состоит в построении последовательности точек $x^{(0)}, x^{(1)}, ..., x^{(p)},...$, монотонно уменьшающих значение целевой функции, т.е.

$F(x^{(0)}) \geq F(x^{(1)}) \geq F(x^{(2)}) \geq \ldots \geq F(x^{(p)}) \geq \ldots$.

Такие методы (алгоритмы) называют методами спуска (или подъема при максимизации целевой функции). Их важнейшей характеристикой является сходимость, которая состоит в том, что при беспредельном увеличении последовательность $x^{(0)}, x^{(1)}, ..., x^{(p)},...$ сходится в точке глобального (локального) минимума.

Использование всех методов спуска сводится к следующей схеме. Пусть на p – й итерации (p – м шаге) имеется точка $x^{(p)}$. Тогда выбирают направление спуска – единичный вектор $l^{(p)}$, определяющий это направление, а также длину шага $\alpha_p > 0$ вдоль этого направления. Очередную точку вычисляют по следующей формуле:

$x^{(p+1)} = x^{(p)} + \alpha_p\, l^{(p)},\ p = 0,1,...$.

Различные методы спуска отличаются друг от друга подходом к выбору длины шага α_p (при $\alpha_p = \alpha$спуск осуществляется с постоянным шагом) и вектора $l^{(p)}$. Все методы спуска можно разделить на три группы в зависимости от того, какие характеристики целевой функции F(x) используются для определения α_p и $l^{(p)}$. К первой группе относят методы, требующие только вычисления значений целевой функции. Эти методы называются *методами нулевого порядка. Методы первого порядка* требуют вычисления значений функции и ее первых производных. Наконец, *методы второго порядка* предполагают исследование и вторых производных.

С помощью методов нулевого порядка можно решать задачи более широкого класса, чем с помощью методов первого и второго порядка. В частности, на их основе могут минимизироваться не дифференцируемые функции, задаваемые таблично или алгоритмически.

Широкое распространение в энергетических расчетах нашел метод нулевого порядка – *метод покоординатного спуска.* В соответствии с этим методом направление спуска выбирают параллельно координатным осям. Сначала проводят спуск вдоль оси Оx_1, затем – вдоль оси Оx_2 и т.д. вплоть до оси Оx_n.

Градиентные методы. Градиентным методом можно решать, вообще говоря, любую нелинейную задачу. Однако при этом находится лишь локальный экстремум. Поэтому применять этот метод рационально при решении задач выпуклого программирования, в которых любой локальный экстремум является одновременно и глобальным.

Задачи без ограничений. Рассмотрим задачу максимизации нелинейной дифференцируемой функции f(X). Суть градиентного поиска точки максимума X* очень проста: надо взять произвольную точку X_0 и с помощью градиента $\nabla f(X_0)$, вычисленного в этой точке, определить направление, в котором возрастает с наибольшей скоростью, а затем, сделав небольшой шаг в найденном направлении, перейти в новую точку X_1. Потом снова определить наилучшее направление $\nabla f(X_1)$ для перехода в очередную точку X_2 и т.д. Таким образом, надо построить последовательность точек $X_0, X_1, X_2, \ldots, X_к, \ldots$так, чтобы она сходилась к точке X*, т.е. для точек последовательности выполнялись условия

$f(X_0) < f(X_1) < f(X_2) < \ldots < f(X_к) < \ldots$.

Градиентные методы позволяют получать точное решение за бесконечное число шагов и только в некоторых случаях – за конечное. В связи с этим градиентные методы относят к приближенным методам решения.

Движение из точки $X_к$ в новую точку $X_{к+1}$ осуществляется по прямой, проходящей через точку $X_к$ и имеющей уравнение

$X_{к+1} = X_к + \alpha_к \nabla f(X_к)$, (5)

где $\alpha_к$ – числовой параметр, от которого зависит длина шага.
Градиентные методы отличаются друг от друга способом выбора величины шага – значения параметра $\alpha_к$. Можно, например, двигаться из точки в точку с постоянным шагом $\alpha_к = \alpha$, т.е. при любом *к*

$X_{к+1} = X_к + \alpha\nabla f(X_к)$.

Если при этом окажется, что $f(X_{к+1}) < f(X_к)$, то следует возвратиться в точку $X_к$ и уменьшить значение параметра α, например до $\alpha/2$.
Если ищется приближенное решение, то поиск можно прекратить, основываясь на следующих соображениях. После каждой серии из определенного числа шагов сравнивают достигнутые значения целевой функции $f(X)$. Если после очередной серии изменение $f(X)$ не превышает некоторого наперед заданного малого числа ε, поиск прекращают и достигнутое значение $f(X)$ рассматривают как искомый приближенный максимум, а соответствующее ему X принимают за X^*.
Если целевая функция $f(X)$ вогнутая (выпуклая), то необходимым и достаточным условием оптимальности точки X^* является равенство нулю градиента функции в этой точке.
Распространенным является вариант градиентного поиска, называемый *методом наискорейшего подъема* (наискорейшего спуска – если решается задача минимизации). Суть его в следующем. После определения $\nabla f(X_к)$ в точке $X_к$ движение вдоль прямой $X = X_к + \alpha_к\nabla f(X_к)$ производится до точки $X_{к+1}$, в которой достигается максимальное значение функции $f(X)$ в направлении градиента $\nabla f(X_к)$. затем в этой точке вновь определяется градиент, и движение совершается по прямой $X = X_{к+1} + \alpha_{к+1}\nabla f(X_{к+1})$ в направлении нового градиента $\nabla f(X_{к+1})$ до точки $X_{к+2}$, в которой достигается максимальное в этом направлении значение $f(X)$. Движение продолжается до тех пор, пока не будет достигнута точка X^*, соответствующая наибольшему значению целевой функции $f(X)$. На рис.х.х приведена схема движения к оптимальной точке X^* методом наискорейшего подъема. В данном случае направление градиента $\nabla f(X_к)$ в точке $X_к$ является касательным к линии уровня поверхности $f(X)$ в точке $X_{к+1}$, следовательно, градиент $\nabla f(X_{к+1})$ в точке $X_{к+1}$ ортогонален градиенту $f(X_к)$.

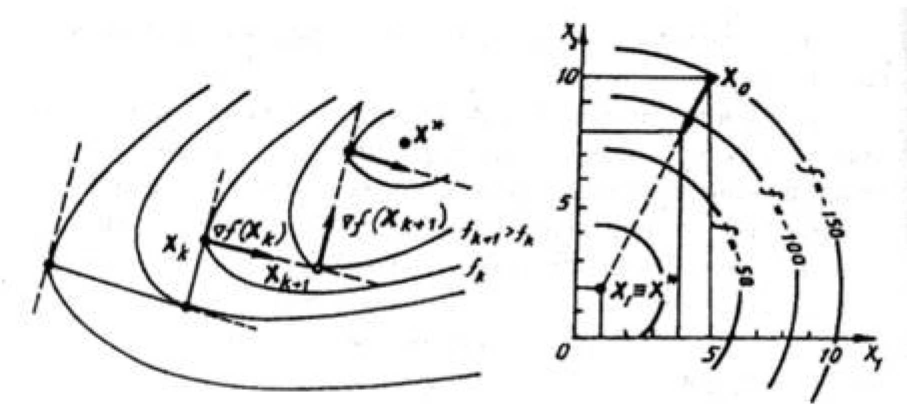


Рисунок - 1

Перемещение из точки $X_к$ в точку $X_{к+1} = X_к + \alpha_к\nabla f(X_к)$ сопровождается возрастанием функции $f(X)$ на величину

$\Delta f(X) = f(X_{к+1}) - f(X_к) = f(X_{к+1,1};\ldots;X_{к+1,n}) - f(X_{к,1};\ldots;X_{к,n}) =$

$= f\left\{x_{k,n} + \alpha_k \frac{\partial f(x_k)}{\partial x_1}; \dots; \; x_{k,n} + \alpha_k \frac{\partial f(x_k)}{\partial x_n}\right\} - f\left(x_{k,1}; \dots; x_{k,n}\right)$. (6)

Из выражения (6.8) видно, что Δf является функцией переменной $\alpha_к$, т.е. $\Delta f = \Delta f(\alpha_к)$. При прохождении максимума функции $f(x)$ в направлении градиента $\nabla f(X_к)$ необходимо выбирать шаг перемещения (множитель $\alpha_к)$, обеспечивающий наибольшее возрастание приращению функции, именно функции $\Delta f(\alpha_к)$. Величина $\alpha_к$, при которой достигается наибольшее значение $\Delta f(\alpha_к)$, может быть определена из необходимого условия экстремума функции $\Delta f(\alpha_к)$:

$d(\Delta f(\alpha_к))/d\alpha_k = 0$. (6.9)

Найдем выражение для производной, дифференцируя равенство (6.9) по $\alpha_к$ как сложную функцию:

$$\frac{d(\Delta f(\alpha_к))}{d\alpha k} = \frac{\partial f(x_{k+1})}{\partial x_1}\frac{\partial f(x_k)}{\partial x_1} + \dots + \frac{\partial f(x_{k+1})}{\partial x_n}\frac{\partial f(x_k)}{\partial x_n} = \nabla f(X_{к+1}) \nabla f(X_к).$$

Подставляя этот результат в равенство (x.x1), получаем

$d(\Delta f(\alpha_к))/d\alpha_k = \nabla f(X_{к+1}) \nabla f(X_к)$.

Это равенство имеет простое геометрическое истолкование: градиент в очередной точке $X_{к+1}$ ортогонален градиенту в предыдущей точке $X_к$.

Задачи с линейными ограничениями. Рассмотрим задачу выпуклого программирования с линейными ограничениями:

$max\ Z = f(x)$; (7) $ax \le a_{i0}$ $(i = 1, 2,...m)$ (8) $x \ge 0$, (9)

где $x = (x_1; x_2;...; x_n)$; $a_i = (a_{i1}; a_{i2};...; a_{in})$.

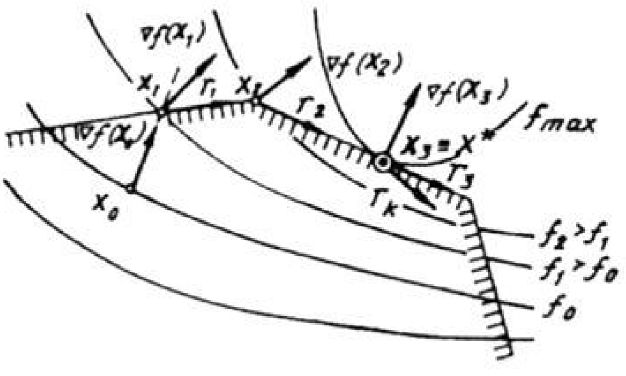


Рисунок - 2

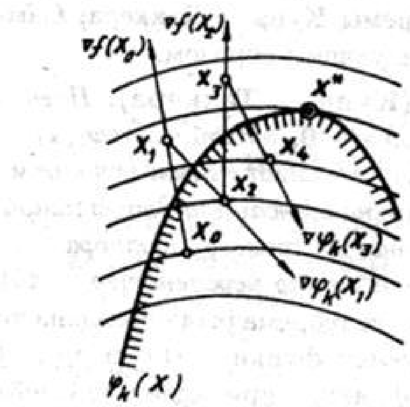
Предполагается, что $f(x)$ является вогнутой функцией и имеются непрерывные частные производные в каждой точке допустимой области.

Начнем с геометрической иллюстрации процесса решения задачи Пусть начальная точка X_0 расположена внутри допустимой области. Из точки X_0 можно двигаться в направлении градиента $\nabla f(X_0)$, пока $f(X)$ не достигнет максимума. В нашем случае $f(X)$ все время возрастает, поэтому остановиться надо в точке X_1 на граничной прямой. Как видно из рисунка, дальше двигаться в направлении градиента $\nabla f(X_1)$ нельзя, т.к. выйдем из области допустимых решений. Поэтому надо найти другое направление перемещения, которое, с одной стороны, не выходит из допустимой области, а с другой –

обеспечивает наибольшее возрастание $f(X)$. Такое направление определяет вектор r_1, составляющий с вектором $\nabla f(X_1)$ наименьший острый угол по сравнению с любым другим вектором, выходящим из точки X_1 и лежащим в допустимой области. Аналитически такой вектор найдется из условия максимизации скалярного произведения $\nabla f(X_1)\ r_1 > 0$. В данном случае вектор r_1, указывающий наивыгоднейшее направление, совпадает с граничной прямой.

Таким образом, на следующем шаге двигаться надо по граничной прямой до тех пор, пока возрастает $f(X)$; в нашем случае – до точки X_2. Из рисунка видно, что далее следует перемещаться в направлении вектора r_2, который находится из условия максимизации скалярного произведения $\nabla f(X_2)\ r_2 > 0$, т.е. по граничной прямой. Движение заканчивается в точке X_3, поскольку в этой точке завершается оптимизационный поиск, т.к. в ней функция $f(X)$ имеет локальный максимум. Ввиду вогнутости оптимизационной функции, в этой точке $f(X)$ достигает глобального максимума в допустимой области. Градиент в точке максимума $X_3 = X^*$ составляет тупой угол с любым вектором $r_к$ из допустимой области, проходящей через X_3, поэтому скалярное произведение $\nabla f(X_3)\ r_к$ будет отрицательным для любого допустимого $r_к$, кроме r_3, направленного по граничной прямой. Для него скалярное произведение $\nabla f(X_3)\ r_3 = 0$, т.к. $\nabla f(X_3)$ и r_3 взаимно перпендикулярны (граничная прямая касается линии уровня поверхности $f(X)$, проходящей через точку максимума X^*). Это равенство и служит аналитическим признаком того, что в точке X_3 функция $f(X)$ достигла максимума.

Аналитическое решение задачи (x.xx1) – (x.xx3) требует, чтобы при выборе $\alpha_к$ определение очередной точки $X_{к+1} = X_к + \alpha_к \nabla f(X_к)$ осталось в допустимой области. Существуют различные методы оптимизации длины шага ($\alpha_к$) и проверки принадлежности очередной точки $X_к$ допустимой области.



<u>*Задачи с нелинейными ограничениями.*</u> Если в задачах с линейными ограничениями движение по граничным прямым оказываетя возможным и даже целесообразным, то при нелинейных ограничениях, определяющих выпуклую область, любое сколь угодно малое перемещение из граничной точки может сразу вывести за пределы области допустимых решений, и возникает необходимость в возвращении в допустимую область (рисунок 6.4). Подобная ситуация характерна для задач, в которых экстремум функции $f(X)$ достигает на границе области. В связи с этим применяются различные способы перемещения, обеспечивающие построение последовательности точек, расположенных вблизи границы и внутри допустимой области, или зигзагообразное движение вдоль границы с пересечением последней. Как видно из рисунка, возврат из точки X_1 в допустимую область следует осуществлять вдоль градиента той граничной функции $\varphi_к(X)$, которая оказалась нарушенной. Это обеспечит отклонение очередной точки X_2 в сторону тоски экстремума X^*. Признаком экстремума в подобном случае будет коллинеарность векторов ∇f и $\nabla \varphi_к$.

1.4 Лекция 7-8 (4 ч.)

Тема: Марковские процессы, их приложения к решению инженерных задач. Системы массового обслуживания

1.4.1 Вопросы лекции:

1. Простейший поток, его свойства. Классификация потоков.

2. Марковские цепи, их свойства
3. Марковские процессы в инженерной практике
4. СМО, основные понятия и свойства. Классификация СМО
5. Практическое применение теории СМО

1.4.2 Краткое содержание вопросов:

Аппарат теории марковских процессов с дискретными состояниями и цепей Маркова широко используют в теории систем, в исследовании операций и других прикладных дисциплинах. Это обусловлено многими причинами, среди которых отметим следующие:

1) многие реальные технические системы имеют конечные множества возможных состояний, а их поведение в процессе функционирования адекватно моделируется марковскими процессами,

2) теория марковских процессов с дискретными состояниями и цепей Маркова разработана настолько глубоко, что позволяет решать широкий класс прикладных задач.

2. Марковские цепи, их свойства

Марковские процессы Представление случайных процессов графом состояний

Рассмотрим физическую систему *S*, в которой протекает случайный процесс с дискретными состояниями: $s_1, s_2, ..., s_i$, (1)

число которых конечно (или счетно). Состояния $s_1, s_2, ...$ могут быть качественными (т. е. описываться словами) или же каждое из них характеризуется случайной величиной (либо случайным вектором).

Прежде всего, рассмотрим множество состояний (1) с точки зрения его структуры - возможности системы *S* переходить из состояния s_j в данное состояние s_i - непосредственно или через другие состояния. Для этого удобно пользоваться наглядной схемой, так называемым графом состояний. Здесь и далее мы будем отчасти пользоваться терминологией теории графов. Имеется две основные разновидности графов: неориентированные и ориентированные.

Неориентированный граф - совокупность точек (вершин графа) с соединяющими некоторые из них отрезками (ребрами графа).

Ориентированный граф - это совокупность точек (вершин) с соединяющими некоторые из них ориентированными отрезками (стрелками).

При изложении теории случайных процессов с дискретными состояниями мы будем пользоваться только ориентированными графами. Вершины графа будут соответствовать состояниям системы. Вершину будем изображать прямоугольником, в который вписано обозначение состояния; стрелка, ведущая из вершины s_j в вершину s_i , будет обозначать возможность перехода системы *S* из состояния s_j в состояние s_i - непосредственно, минуя другие состояния. Стрелки графа могут изображаться не только прямолинейными, но и криволинейными отрезками (рис. 1). Сам граф системы *S* будем обозначать буквой *G*.

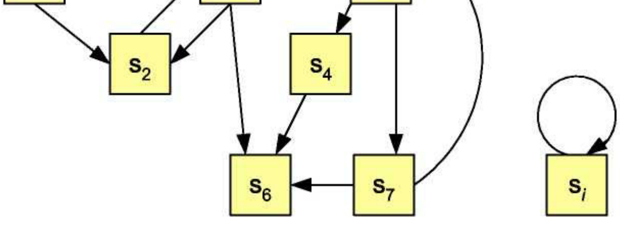


Рисунок 1 – Пример графа состояний

Переход по стрелке, ведущей из состояния s_i в него же, означает задержку системы в состоянии s_i. «Обратные стрелки» можно на графе не проставлять, так как все расчеты можно вести и без них.

Проведем некоторую необходимую для дальнейшего классификацию состояний. Состояние s_i называется источником, если система S может выйти из этого состояния, но попасть в него обратно уже не может, т. е. на графе G состояний в состояние s_i не ведет ни одна стрелка. На рисунке 1 состояние s_1 является источником.

Состояние s_i называется концевым (или поглощающим), если система S может попасть в это состояние, но выйти из него уже не может. Для графа состояний это означает, что из состояния s_i не ведет ни одна стрелка (для графа, изображенного на рисунке 1, состояние s_6 поглощающее).

Если система S может непосредственно перейти из состояния s_i в состояние s_j то состояние s_j - называется соседним по отношению к состоянию s_i.

Состояние s_i называется транзитивным, если система S может войти в это состояние и выйти из него, т. е. на графе состояний есть хотя бы одна стрелка, ведущая в s_i и хотя бы одна стрелка, ведущая из s_i. На рисунке 1 все состояния, кроме s_1 и s_6, являются транзитивными.

Для полноты картины можно рассматривать также и «изолированные» состояния. Состояние s_i называется изолированным, если из него нельзя попасть ни в одно из других состояний и в него нельзя попасть ни из какого другого состояния.

Наряду с отдельными состояниями системы S в ряде задач практически бывает нужно рассматривать подмножества ее состояний.

Обозначим W множество всех состояний системы S (конечное или бесконечное, но счетное) и рассмотрим его подмножество $V \subset W$. Подмножество V называется замкнутым (концевым), если система S, попав в одно (или находясь в одном) из состояний $s_i \in V$, не может выйти из этого подмножества состояний. Концевое подмножество состояний может включать в себя поглощающее состояние, а может и не включать.

Подмножество состояний $V \subset W$ называется связным или эргодическим, если из любого состояния, входящего в него, можно попасть в любое другое состояние, принадлежащее этому подмножеству. Эргодическим может быть и все множество W состояний системы S. В эргодическом множестве состояний нет ни источников, ни поглощающих состояний.

Подмножество состояний V называется транзитивным, если система S может войти в это подмножество и выйти из него, т. е. из любого состояния $s_i \subset V$ можно (за то или другое число перескоков) выйти из этого подмножества.

Случайный процесс, протекающий в системе S, можно трактовать как процесс блуждания системы по множеству состояний W. Если подмножество $V \subset W$ является концевым, то, попав в него, система будет продолжать блуждание уже по этому подмножеству состояний V. Если все множество эргодично, то блуждание будет происходить по всем его состояниям.

На практике очень часто встречаются системы, состояния которых образуют цепь (рисунок 2), в которой каждое состояние s_i (кроме двух крайних s_0 и s_n) связано прямой и обратной связью с двумя соседними s_{i-1}, s_{i+1}, а каждое из двух крайних связано прямой и обратной связью только с одним соседним.

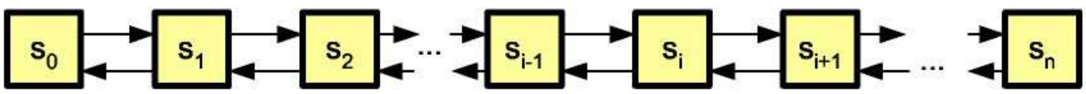


Рисунок 2 - Схема процесса гибели и размножения

Такая схема случайного процесса называется схемой гибели и размножения, а сам процесс — процессом гибели и размножения.

Если на графе состояний системы *S* стрелки, ведущие справа налево, отсутствуют, то говорят о процессе «чистого размножения», в противоположном случае — о процессе «чистой гибели».

Процесс гибели и размножения может в некоторых случаях иметь не конечное число состояний: $s_1, s_2, \ldots, s_i, \ldots, s_n$, а бесконечное (счетное): $s_1, s_2, \ldots, s_i, \ldots$.

При анализе случайных процессов, протекающих в системах с дискретными состояниями, важную роль играют вероятности состояний.

Обозначим *S(t)* состояние системы *S* в момент *t.* Вероятностью *i*-го состояния в момент *t* называется вероятность события, состоящего в том, что в момент *t* система *S* будет в состоянии s_i. Обозначим ее $p_i(t)$:

$$p_i(t) = P\{S(t) = s_i\}, \qquad (2)$$

где *S(t)* - случайное состояние системы *S* в момент *t*. Очевидно, что для системы с дискретными состояниями $s_1, s_2, \ldots, s_i, \ldots,$ в любой момент *t* сумма вероятностей состояний равна единице:

$$\sum_i p_i(t) = 1, \qquad (3)$$

как сумма вероятностей полной группы несовместных событий.

В ряде задач практики нас интересует так называемый установившийся или стационарный режим работы системы, который в ней устанавливается, когда от начала процесса прошло достаточно большое время *t*. Например, процесс изменения напряжения в сети питания технического устройства, пройдя сразу после включения через ряд колебаний, по прошествии времени, устанавливается. Аналогично этому и в некоторых случайных процессах по прошествии достаточно большого времени *t* устанавливается стационарный режим, во время которого состояния системы хотя и меняются случайным образом, но их вероятности $p_i(t)$ $(i = 1, 2, \ldots)$ остаются постоянными. Обозначим эти постоянные вероятности p_i.

$$p_i = \lim p_i(t) \qquad (4)$$

Вероятности $p_i (i = 1, 2, \ldots)$, если они существуют, называются финальными (предельными) вероятностями состояний. Финальную вероятность p_i можно истолковать как среднюю долю времени, которую в стационарном режиме проводит система *S* в состоянии s_i. В дальнейшем будет показано, при каких условиях финальные вероятности существуют и какими они могут быть для разных состояний и подмножеств состояний.

Введем очень важное для дальнейшего понятие марковского случайного процесса.

Случайный процесс, протекающий в системе *S* с дискретными состояниями $s_1, s_2, \ldots, s_i, \ldots,$ называется марковским, если для любого момента времени t_0 вероятность каждого из состояний системы в будущем (при $t>t_0$) зависит только от ее состояния в настоящем (при $t = t_0$) и не зависит от того, когда и как она пришла в это состояние; т. е. не зависит от ее поведения в прошлом (при $t < t_0$).

Не надо понимать марковское свойство случайного процесса как полную независимость «будущего» от «прошлого»; в общем случае «будущее» зависит от «настоящего», т. е. вероятности $p_i(t)$ при $t > t_0$ зависят от того, в каком состоянии s_i находится система в настоящем (при $t=t_0$); само же это «настоящее» зависит от «прошлого», от того, как вела себя система *S* при $t < t_0$. Это можно сформулировать следующим образом: для марковского случайного процесса «будущее» зависит от «прошлого» только через «настоящее» (рисунок 3). При фиксированном «настоящем» условные вероятности всех состояний системы в «будущем» не зависят от предыстории процесса, т. е. от того, когда и как система S к моменту t_0 пришла в состояние

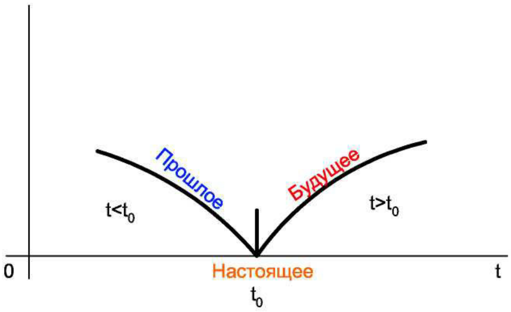


Рисунок 3 – Схема марковского свойства случайного процесса

«Настоящее» может быть задано не одним каким-то состоянием s_i, а целым подмножеством состояний $V \subset W$, где W - множество всех возможных состояний системы.

Подчеркнем также, что «настоящее» может быть задано не только одним состоянием системы S в момент t_0; в него при желании можно включить и те элементы из «прошлого», от которых, при заданном «настоящем», зависит будущее. Например, вероятности состояний в «будущем» могут зависеть не только от состояния s_i системы в настоящем, но и от того, из какого состояния s_i система перешла к моменту t_0 в состояние s_i; в этом случае настоящее характеризуется не только состоянием s_i , в которое система перешла к моменту t_0, но и состоянием s_j, из которого она перешла в s_i. Вводя в состав параметров, характеризующих настоящее состояние системы, те параметры из прошлого, от которых зависит будущее, можно, как говорится, «марковизировать» многие немарковские случайные процессы, но, как правило, это приводит к сильному усложнению математического аппарата.

3. Марковские процессы в инженерной практике

Марковские случайные процессы с дискретными состояниями и дискретным временем

Пусть имеется система S с дискретными состояниями $s_1, s_2, ..., s_i, ..., s_n$. Предположим, что случайные переходы («перескоки») системы из состояния в состояние могут происходить только в определенные моменты времени $t_0, t_1, t_2, ...$. Эти моменты мы будем называть шагами процесса; $t_0=0$ - его началом. Сам процесс представляет собой случайное блуждание системы S по состояниям. После первого шага система может оказаться в одном (и только в одном) из своих возможных состояний: $s_1^{(1)}, s_2^{(1)}, ..., s_i^{(1)}, ..., s_n^{(1)}$; на втором шаге - $s_1^{(2)}, s_2^{(2)}, ..., s_i^{(2)}, ..., s_n^{(2)}$, на k-м шаге $s_1^{(k)}, s_2^{(k)}, ..., s_i^{(k)}, ..., s_n^{(k)}$ (число состояний в общем случае может быть бесконечным, но счетным. Здесь же для простоты ограничимся конечным числом n состояний).

Предположим, что граф состояний системы S имеет вид, представленный на рисунке 4. Процесс блуждания системы S по состояниям можно представить как последовательность или «цепь» событий, состоящих в том, что в начальный момент $t_0=0$ система находится в одном из состояний (например, в состоянии $s_1^{(0)}$), в момент первого шага перешла из него скачком в состояние $s_5^{(1)}$, из которого на втором шаге перешла в $s_3^{(2)}$, на третьем шаге перешла в $s_2^{(3)}$ и т. д. «Траектория» системы, блуждающей по состояниям s_1, s_5, s_3, s_2 показана на рисунке 4 жирными линиями. На каких-то шагах система может задерживаться в том или другом из своих состояний, $s_i^{(k)} = s_i^{(k+1)}$ (это показано «возвратной стрелкой» на рисунке 4) или же вернуться в него после ряда шагов.

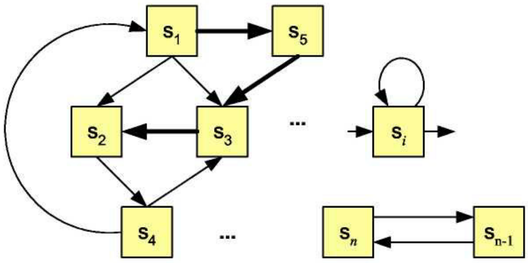


Рисунок 4 – Граф состояний системы S

«Траектория» блуждания системы по графу состояний, изображенная на рисунке 4 жирными линиями, представляет собой не что иное, как реализацию случайного процесса, полученную в результате одного опыта. При повторении опыта, естественно, реализации в общем случае не совпадают.

Рассмотрим общий случай. Пусть происходит случайный процесс в системе S с дискретными состояниями $s_1, s_2, ..., s_i, ..., s_n$, которые она может принимать в последовательности шагов с номерами *0, 1, 2, ..., k, ...*.

Случайный процесс представляет собой последовательность событий вида $\{S(k) = s_i\}$ $(i = 1,2,... ... n;\ k = 0,1,2,...)$. Наиболее важной ее характеристикой являются вероятности состояний системы

$$\mathbf{P}\{S(k)= s_i\} \quad (i = 1,2,......n;\ k = 0,1,2,...), \qquad (5)$$

где $P\{S(k)=s_i\}$ - вероятность того, что на k-м шаге система S будет находиться в состоянии s_i.

Распределение вероятностей (5) представляет собой не что иное, как одномерный закон распределения случайного процесса $S(t)$, протекающего в системе S с «качественными» дискретными состояниями и дискретным временем $t_0, t_1, t_2, ..., t_k$.

Процесс, протекающий в такой системе S, называется марковским процессом с дискретными состояниями и дискретным временем (или, короче, марковской цепью), если выполняется условие: для любого фиксированного момента времени (любого шага k_0) условные вероятности состояний системы в будущем (при $k > k_0$) зависят только от состояния системы в настоящем (при $k = k_0$) и не зависят от того, когда (на каком шаге, при $k < k_0$) и откуда система пришла в это состояние. Марковская цепь представляет собой разновидность марковского процесса, в котором будущее зависит от прошлого только через настоящее.

Цепь, в которой условные вероятности состояний в будущем зависят только от состояния на данном, последнем, шаге и не зависят от предыдущих, иногда называют простой цепью Маркова, в отличие от такой, где будущее зависит от состояний системы не только в настоящем на данном шаге, но и от ее состояний на нескольких предыдущих шагах; такую цепь называют сложной цепью Маркова. Сам А. А. Марков рассматривал сложные цепи, построенные на материале буквенных последовательностей, взятых из текста пушкинского «Евгения Онегина».

Если в качестве системы, в которой происходит случайный процесс, рассмотреть букву, входящую в текст, которой могут быть: а, б, в, ... щ, ъ, ы, ь, э, ю, я, «пробел», то сразу ясно, что вероятность последующей буквы быть той или другой зависит от того, какова была предыдущая (например, последовательности букв «яы» или «эь» в русском языке исключены); не так очевидно, но все же ясно, что эта вероятность зависит не только от предыдущей буквы, но и от других, ей предшествовавших (например, последовательность букв «ттт» в русском языке если не исключена, то практически невозможна, тогда как последовательность «тт» встречается довольно часто). Мы в данном элементарном изложении будем рассматривать только простые цепи Маркова и вычислять для них вероятности состояний.

Из определения марковской цепи следует, что для нее вероятность перехода системы S в состояние s_i на $(k+1)$-м шаге зависит только от того, в каком состоянии s_i находилась система на предыдущем k-м шаге и не зависит от того, как она вела себя до этого k-го шага.

Основной задачей исследования марковской цепи является нахождение безусловных вероятностей нахождения системы S на любом k-м шаге в состоянии s_i; обозначим эту вероятность $p_i(k)$:

$$p_i(k) = P\{S(k) = s_i\} \ (i = 1,2,......n;\ k = 0,1,2,...). \qquad (6)$$

Для нахождения этих вероятностей необходимо знать условные вероятности перехода системы S на k-м шаге в состояние s_i, если известно, что на предыдущем $(k - 1)$-м шаге она была в состоянии s_i. Обозначим эту вероятность

$$p_{ij}(k) = P\{S(k) = s_i \mid S(k-1) = s_j\} \quad (i, j = 1,2,......n;\ k = 0,1,2,...). \qquad (7)$$

Вероятности $p_{ij}(k)$ называются переходными вероятностями марковской цепи на k-м шаге. Вероятность $p_{ij}(k)$ есть вероятность того, что на k-м шаге система задержится (останется) в состоянии s_i.

Переходные вероятности $p_{ij}(k)$ можно записать в виде квадратной таблицы (матрицы) размерности n x n:

$$\|p_{ij}(k)\| = \begin{Vmatrix} p_{11}(k) & p_{12}(k) & \ldots & p_{1j}(k) & \ldots & p_{1n}(k) \\ p_{21}(k) & p_{22}(k) & \ldots & p_{2j}(k) & \ldots & p_{2n}(k) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ p_{i1}(k) & p_{i2}(k) & \ldots & p_{ij}(k) & \ldots & p_{in}(k) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ p_{n1}(k) & p_{n2}(k) & \ldots & p_{nj}(k) & \ldots & p_{nn}(k) \end{Vmatrix} \quad (k = 0, 1, 2,...). \qquad (8)$$

По главной диагонали матрицы (8) стоят вероятности задержки системы в данном состоянии $s_j \ (j = 1,..., n)$ на k-м шаге.

$$p_{11}(k), p_{22}(k),..., p_{ii}(k),..., p_{nn}(k). \qquad (9)$$

Так как на каждом шаге система S может находиться только в одном из взаимно исключающих состояний, то для любой k-й строки матрицы (8) сумма всех стоящих в ней вероятностей равна единице:

$$\sum_{j=1}^{n} p_{ij}(k) = 1. \qquad (10)$$

Матрица, обладающая таким свойством, называется стохастической. Естественно, что все элементы стохастической матрицы отвечают условию $0 \le p_{ij}(k) \le 1$. В силу условия (10) можно в матрице (8) не задавать вероятности задержки, а получать их как дополнения до единицы всех остальных членов строки:

$$p_{ii}(k) = 1 - \sum_{j=1}^{n} p_{ij}(k). \qquad (11)$$

Чтобы найти безусловные вероятности $p_i(k)$, недостаточно знать матрицу переходных вероятностей (8); нужно еще знать начальное распределение вероятностей, т. е. вероятности состояний $p_i(0)$, соответствующие началу процесса - моменту $t_0 = 0$:

$$p_1(0), p_2(0), ..., p_i(0), ..., p_n(0), \qquad (12)$$

в сумме образующие единицу:

$$\sum_{i=1}^{n} p_i(0) = 1 \qquad (13)$$

Если известно, что в начальный момент система S находится во вполне определенном состоянии s_i, то вероятность $p_i(0)$ этого состояния в формуле (13) равна единице, а все остальные - нулю:

$$p_i(0), p_1(0) = p_2(0) = ... = p_{i-1}(0) = p_{i+1}(0) = ... = p_i(0) = 0. \qquad (14)$$

Цепь Маркова называется однородной, если переходные вероятности $p_{ij}(k)$ не зависят от номера шага k: $p_{ij}(k) = p_{ij}$. Матрица переходных вероятностей для однородной цепи Маркова имеет вид:

$$\|p_{ij}\| = \left\| \begin{matrix} p_{11} & p_{12} \cdots p_{1j} \cdots p_{1n} \\ p_{21} & p_{22} \cdots p_{2j} \cdots p_{2n} \\ \cdots & \cdots \\ p_{i1} & p_{i2} \cdots p_{ij} \cdots p_{in} \\ \cdots & \cdots \\ p_{n1} & p_{n2} \cdots p_{nj} \cdots p_{nn} \end{matrix} \right\| \qquad (14)$$

При выводе формул для вероятностей состояний, в целях простоты записи, будем рассматривать только однородные цепи Маркова (в случае, когда цепь неоднородна, можно все переходные вероятности в формулах просто положить зависящими от номера шага k).

При нахождении вероятностей состояний марковской цепи на k-м шаге $p_i(k)$ $(k = 1,2,...)$ удобно бывает пользоваться так называемым размеченным графом состояний системы S, где возле каждой стрелки, ведущей из состояния s_i в состояние s_j, проставлена переходная вероятность p_{ij}; вероятности задержки на размеченном графе не проставляются, а просто получаются дополнением до единицы суммы вероятностей, стоящих у всех стрелок, ведущих из данного состояния s_i.

Теперь покажем, как найти для однородной цепи Маркова безусловную вероятность нахождения системы S на k-м шаге в состоянии s_j ($j=1, 2, ..., n$)

$$p_j(k) = \mathbf{P}\{S(k)=s_j\}, \qquad (15)$$

если задана матрица переходных вероятностей $\|p_{ij}\|$ (или, что равнозначно, размеченный граф состояний) и начальное распределение вероятностей

$$p_i(0) \quad (i = 1,2,...,n).$$

$$\sum_{j=1}^{n} p_{ij}(k) = 1 \qquad (16)$$

Сделаем гипотезу, состоящую в том, что в начальный момент ($k=0$) система находилась в состоянии s_i. Вероятность этой гипотезы известна из (16) и равна $p_i(0)=P\{S(0)=s_i\}$. В предположении, что эта гипотеза имеет место, условная вероятность того, что система S на первом шаге будет в состоянии s_j, равна переходной вероятности $p_{ij}(k) = \mathbf{P}\{S(1) = s_j \mid S(0) = s_i\}$.

По формуле полной вероятности получим:

$$p_j(1) = \sum_{i=1}^{n} P\{S(1) = s_j | S(0) = s_i\} \cdot P\{S(0) = s_i\} = \sum_{i=1}^{n} p_{ij} p_i(0), \quad (j = 1,...,n) \quad (17)$$

Таким образом, мы нашли распределение вероятностей системы S на первом шаге. Теперь у нас есть все необходимое для того, чтобы найти распределение вероятностей на втором шаге, которое для цепи Маркова зависит только от распределения вероятностей на первом шаге и матрицы переходных вероятностей.

Опять сделаем гипотезу, состоящую в том, что на первом шаге система находится в состоянии s_i вероятность этой гипотезы нам уже известна и равна $p_i(1) = P\{S(1) = s_i\}$. При этой гипотезе условная вероятность того, что на втором шаге система S будет в состоянии s_i, равна:

$$p_{ij}(k) = P\{S(2) = s_j | S(1) = s_i\}$$

По формуле полной вероятности находим

$$p_j(2) = \sum_{i=1}^{n} p_i(1) p_{ij}, (j = 1,2,...,n) \quad (18)$$

Таким образом, мы выразили распределение вероятностей (18) на втором шаге через распределение вероятностей на первом шаге и матрицу $\|p_{ij}\|$. Переходя таким же способом от $k = 2$ к $k = 3$ и т. д., получим рекуррентную формулу:

$$p_j(k) = \sum_{i=1}^{n} p_i(k-1) p_{ij}, (k = 1,2,...,n; j = 1,2,...,n) \quad (19)$$

При некоторых условиях в цепи Маркова с возрастанием k (номера шага) устанавливается стационарный режим, в котором система S продолжает блуждать по состояниям, но вероятности этих состояний уже от номера шага не зависят. Такие вероятности называются предельными (или финальными) вероятностями цепи Маркова.

Например, если рассматривать ЭВМ в двух состояниях: s_1 - исправна, s_2 - не исправна, то имеет место следующая динамика изменения вероятностей (при начальных условиях):

$$p_1(0) = 1, p_2(0) = 0: p_1(1) = 0,7; p_1(2) = 0,61; p_1(3) = 0,583; p_1(4) = 0,5749 .$$

Ниже мы покажем, что в этом случае $p_1 = \lim_{k \to \infty} p_1(k) = 0,4/(0,4+0,3) = 0,5714$. Таким образом, в рассматриваемой системе стационарный режим наступит практически через четыре шага.

Можно убедиться в том, что в этом примере финальные вероятности не зависят от начальных условий.

Сформулируем условия существования стационарного режима для системы S с конечным числом состояний n, в которой протекает марковский случайный процесс с дискретными состояниями и дискретным временем (цепь Маркова):

1. Множество всех состояний W системы S должно быть эргодическим.
2. Цепь Маркова должна быть однородной:

$$p_{ij}(k) = p_{ij} \quad (20)$$

3. Цепь Маркова должна быть «достаточно хорошо перемешиваемой» (не должна быть «циклической»).

Цепи Маркова, отвечающие этим условиям, будем называть эргодическими цепями Маркова.

4. СМО, основные понятия и свойства. Классификация СМО.

Для эффективной работы банка необходимы, во-первых, постоянное изучение и прогнозирование состояния рынка банковских услуг и, во-вторых, всестороннее планирование банковской деятельности и оперативное управление финансовыми ресурсами банка.

Банковские учреждения оказывают клиентам множество услуг, вступают в сложные взаимоотношения между собой и другими субъектами хозяйственной жизни, выполняют разнообразные функции. Для выживания в условиях обострившейся конкуренции банки должны искать пути совершенствования базовых технологий, внедрять новые банковские инструменты, поддерживать свою работу автоматизированной информационной системой управления и обработки данных, соответствующей международным требованиям и стандартам.

Моделирование банковских операций является одним из важных этапов анализа и оценки деятельности банка, а также разработки проекта автоматизированной банковской системы (АБС).

Отличительная черта российского финансового рынка — его субъективизм, крайняя зависимость от внеэкономических факторов и, как следствие, высокая степень неопределенности, которая затрудняет принятие обоснованных финансовых решений. Применение традиционных средств поддержки управленческих решений и прогнозирования в этих условиях затруднено, и тем ценнее возможность использования метода имитационного моделирования, повышенный интерес к которому проявляется сегодня в развитых странах. Этот метод воспринимается сегодня как мощный и перспективный инструмент конструирования и последующего исследования сложных бизнес-процессов и систем, в которых велико число переменных, трудоемок, а зачастую и невозможен математический анализ зависимостей, высок уровень неопределенности имитируемых ситуаций.

Распространению подобных моделей способствовал также коммерческий успех ряда аналитических программных продуктов, успешно используемых в банках, промышленных и торговых фирмах, государственных учреждениях, страховых компаниях и т.д. Практика применения имитационных моделей открыла новые возможности по концептуальному анализу проблем управления бизнесом, сокращению сроков разработки перспективных пилотных проектов, организации эффективного сопровождения сложных корпоративных приложений.

Сегодня подходы и методы имитационного моделирования могут оказаться чрезвычайно плодотворными в отечественных условиях перманентной экономической неустойчивости и риска. Прежде всего это касается перспективных, динамично развивающихся и находящихся в стадии становления активных секторов отечественного бизнеса, таких, как банковская и страховая деятельность, рынок информационных технологий, торгово-посреднический бизнес, а также рынок ценных бумаг.

Имитационная модель предназначена для имитации процесса функционирования реальных систем массового обслуживания. Системы массового обслуживания (СМО) представляют собой системы специального вида, реализующие многократное выполнение однотипных задач. Подобные системы играют важную роль во многих областях экономики, финансов, производства и быта. В качестве примеров СМО в финансово-экономической ; сфере можно привести банки различных типов (коммерческие, инвестиционные, ипотечные, инновационные, сберегательные), страховые организации, государственные акционерные общества, компании, фирмы, ассоциации, кооперативы, налоговые инспекции, аудиторские службы, различные системы связи (в том числе телефонные станции), погрузочно-разгрузочные комплексы (порты, товарные станции), автозаправочные станции, различные предприятия и организации сферы обслуживания (магазины, справочные бюро, парикмахерские, билетные кассы, пункты по обмену валюты, ремонтные мастерские, больницы). Такие системы, как компьютерные сети, системы сбора, хранения и обработки информации, транспортные системы, автоматизированные производственные участки, поточные линии, различные военные системы, в частности системы противовоздушной или противоракетной обороны, также могут ? рассматриваться как своеобразные СМО

Каждая СМО включает в свою структуру некоторое ; число обслуживающих устройств, которые называют каналами (приборами, линиями) обслуживания. Роль каналов могут играть различные приборы, лица, выполняющие те или иные операции (кассиры, операторы, парикмахеры, продавцы), линии связи, автомашины, краны, ремонтные бригады, железнодорожные пути, бензоколонки и т.д.

Системы массового обслуживания могут быть одноканальными или многоканальными.

Каждая СМО предназначена для обслуживания (выполнения) некоторого потока заявок (требований), поступающих на вход системы большей частью не регулярно, а случайные моменты времени. Обслуживание заявок, в этом случае, также длится не постоянное, заранее известное время, а случайное время, которое зависит от многих случайных, порой неизвестных нам, причин. После обслуживания заявки канал освобождается и готов к приему следующей заявки. Случайный характер потоказаявок и времени их обслуживания приводит к неравномерной загруженности СМО: в иное время на входе СМО могут скапливаться необслуженные заявки, что приводит к перегрузке СМО, а иногда при свободных каналах на входе СМО заявки не будет, что приводит к недогрузке СМО, т.е. к простаиванию ее каналов. Заявки, скапливающиеся на входе СМО, либо «становятся» в очередь, либо по причине невозможности дальнейшего пребывания в очереди покидают СМО необслуженными. Схема СМО изображена на рис.1. ъ

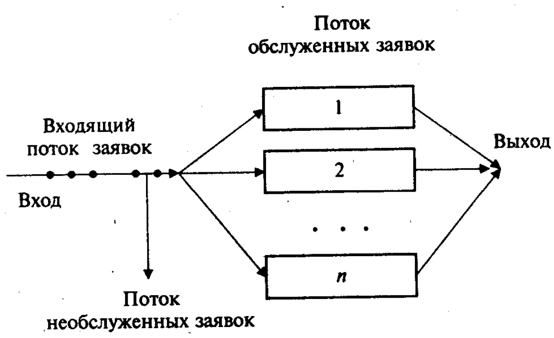


Рис.1. Схема СМО

Таким образом, во всякой СМО можно выделить следующие основные элементы:

1) входящий поток заявок;
2) очередь;
3) каналы обслуживания;
4) выходящий поток обслуженных заявок.

Каждая СМО в зависимости от своих параметров — характера потока заявок, числа каналов обслуживания и их производительности, а также от правил организации работы обладает определенной эффективностью функционирования (пропускной способностью), позволяющей ей более или менее успешно справляться с потоком заявок.

Предметом изучения теории массового обслуживания •является СМО.

Цель теории массового обслуживания — выработка рекомендаций по рациональному построению СМО, рациональной организации их работы и регулированию потока заявок для обеспечения высокой эффективности функционирования СМО.

Для достижения этой цели решаются задачи теории массового обслуживания, состоящие в установлении зависимостей эффективности функционирования СМО I от ее организации (параметров): характера потока заявок, числа каналов и их производительности и правил работы СМО.

В качестве характеристик эффективности функционирования СМО можно, выбрать три основные группы (обычно средних) показателей:

1. Показатели эффективности использования СМО

• Абсолютная пропускная способность СМО — среднее число заявок, которое может обслужить СМО в единицу времени.

• Относительная пропускная способность СМО — отношение среднего числа заявок, обслуживаемых СМО в единицу времени, к среднему числу поступивших заявок за это же время.

• Средняя продолжительность периода занятости СМО.

• Коэффициент использования СМО — средняя доля времени, в течение которого СМО занята обслуживанием заявок, и т.п.

2. Показатели качества обслуживания заявок

• Среднее время ожидания заявки в очереди.

• Среднее время пребывания заявки в СМО.

• Вероятность отказа заявке в обслуживании без ожидания.

• Вероятность того, что поступившая заявка немедленно будет принята к обслуживанию.

• Закон распределения времени ожидания заявки в очереди.
• Закон распределения времени пребывания заявки в СМО.
• Среднее число заявок, находящихся в очереди.
• Среднее число заявок, находящихся в СМО, и т.п.

3. Показатели эффективности функционирования пары «СМО — потребитель», где под потребителем понимают всю совокупность заявок или некий их источник (например, средний доход, приносимый СМО в единицу времени, и т.п.)

Отметим, что третья группа показателей оказывается полезной в тех случаях, когда некоторый доход, получаемый от обслуживания заявок, и затраты на обслуживание измеряются в одних и тех же единицах. Эти показатели обычно носят вполне конкретный характер и определяются спецификой СМО, обслуживаемых заявок и дисциплиной обслуживания.

5. Практическое применение теории СМО

Случайный характер потока заявок и длительности их обслуживания порождает в СМО случайный процесс.

Случайным процессом (или случайной функцией) называется соответствие, при котором каждому значению аргумента (в данном случае — моменту из промежутка времени проводимого опыта) ставится в соответствие случайная величина (в данном случае — состояние СМО).

Случайной величиной называется величина, которая в результате опыта может принять одно, но неизвестное заранее, какое именно, числовое значение из данного числового множества.

Для решения задач теории массового обслуживания необходимо этот случайный процесс изучить, т. е. построить и проанализировать его математическую модель.

Математическое изучение функционирования СМО значительно упрощается, если протекающий в ней случайный процесс является марковским. Случайный процесс, протекающий в СМО, называется марковским (или процессом без последействия, или процессом без памяти), если вероятность любого состояния СМО в будущем зависит только от ее состояния в настоящем и не зависит от.ее состояний в прошлом. В этом случае работа СМО сравнительно легко описывается с помощью аппарата конечных систем обыкновенных линейных дифференциальных уравнений первого порядка, а в предельном режиме (при достаточно длительном функционировании СМО) — с помощью аппарата конечных систем линейных алгебраических уравнений, и в результате удается выразить в явном виде основные характеристики эффективности функционирования СМО через параметры СМО, потока заявок и дисциплины работы СМО.

Чтобы случайный процесс был марковским, необходимо и достаточно, чтобы все потоки событий, под воздействием которых происходят переходы системы из состояния в состояние, были пуассоновскими, т.е. обладающими свойствами отсутствия последействия (для любых двух непересекающихся промежутков времени число событий, наступающих за один из них, не зависит от числа событий, наступающих за другой) и ординарностью (вероятность наступления за элементарный — малый промежуток времени более одного событий пренебрежимо мала по сравнению с вероятностью наступления за этот промежуток времени одного события). В СМО потоками событий являются потоки заявок, потоки «обслуживания» заявок и т.д.

1.5 Лекция 9 (2 ч.)

Тема: Учебно-научные работы студента вуза

1.5.1 Вопросы лекции:

1. Виды НИР

2. Основные принципы построения НИРС

3.Работа с источниками. Этические аспекты научно-исследовательской деятельности

1.5.2 Краткое содержание вопросов:

1. Виды НИР

Существует и применяется два основных вида научно-исследовательской работы студентов (НИРС).

1. Учебная научно-исследовательская работа студентов, предусмотренная действующими учебными планами. К этому виду НИРС можно отнести курсовые работы, выполняемые в течение всего срока обучения в ВУЗе, а так же дипломную работу, выполняемую на пятом курсе.

Во время выполнения курсовых работ студент делает первые шаги к самостоятельному научному творчеству. Он учится работать с научной литературой (если это необходимо, то и с иностранной), приобретает навыки критического отбора и анализа необходимой информации. Если на первом курсе требования к курсовой работе минимальны, и написание её не представляет большого труда для студента, то уже на следующий год требования заметно повышаются, и написание работы превращается в действительно творческий процесс. Так, повышая с каждым годом требования к курсовой работе, ВУЗ способствует развитию студента, как исследователя, делая это практически незаметно и ненавязчиво для него самого.

Выполнение дипломной работы имеет своей целью дальнейшее развитие творческой и познавательной способности студента, и как заключительный этап обучения студента в ВУЗе направлено на закрепление и расширение теоретических знаний и углубленное изучение выбранной темы. На старших курсах многие студенты уже работают по специальности, и, выбирая тему для курсовой работы это чаще всего учитывается. В данном случае, кроме анализа литературы, в дипломную работу может быть включён собственный практический опыт по данному вопросу, что только увеличивает научную ценность работы.

К НИРС, предусмотренной действующим учебным планом, можно отнести и написание рефератов по темам практических занятий. При этом следует сказать о том, что чаще всего реферат является или переписанной статьёй, или, что ещё хуже, конспектом главы какого-то учебника. Назвать это научной работой можно с большим сомнением. Но некоторые рефераты, написанные на основе нескольких десятков статей и источников, по праву можно назвать научными трудами и включение их в список видов НИРС вполне оправданно.

2. Исследовательская работа сверх тех требований, которые предъявляются учебными планами.

Основными формами НИРС, выполняемой во внеучебное являются:

·Предметные кружки;

·Проблемные кружки;

·Проблемные студенческие лаборатории;

·Участие в научных и научно-практических конференциях;

·участие во внутривузовских и республиканских конкурсах.

2. Основные принципы построения НИРС

Предметные кружки. Проблемные кружки. Проблемные студенческие лаборатории (ПСТ).

Данная форма НИРС чаще всего используется при работе со студентами младших курсов. Руководителями выступают общенаучные и общетеоретические кафедры. Научный кружок является самым первым шагом в НИРС, и цели перед его участниками ставятся несложные. Чаще всего, это подготовка докладов и рефератов, которые потом заслушиваются на заседаниях кружка или на научной конференции. Кружок может объединять как членов группы, курса, факультета, а иногда - и всего института. Последний вариант чаще всего встречается в кружках, изучающих проблемы общественных и гуманитарных наук, так как в технических и естественных кружках научные исследования студента пятого курса скорее всего будут малопонятны студентам первого, и у них может пропасть интерес к кружку, как таковому.

Работа кружков, как правило, выглядит следующим образом:

На организационном собрании, проходящем приблизительно в октябре, происходит распределение тем докладов и рефератов выборным путём, после чего преподаватель указывает на наличие для каждой темы основной и дополнительной литературы и рекомендует в ближайшее время продумать план работы. Некоторые преподаватели считают, что выборное распределение докладов не является необходимым, так как студент концентрируется на одной теме, не уделяя большого внимания другим. С одной стороны, принудительное распределение тем может ликвидировать такую «зацикленность», но, с другой стороны, такой подход может не найти поддержки у самих студентов. Представим себе первокурсника, который впервые пришёл на заседание кружка, где, как он как он считает, к нему должны относиться почти, как к равному, и вдруг он получает для работы тему, которая его интересует очень мало, а тема, которую ему хотелось развить в своей работе, досталась другому. Конечно, студент обидится, и его присутствие на остальных заседаниях кружка ставится под сомнение.

Таким образом, распределение тем должно быть исключительно выборным, тем более что к началу обучения в ВУЗе человек уже достаточно развит, чтобы иметь собственные интересы и пристрастия.

После распределения тем начинается главная и основная работа кружка. На первых порах основная роль принадлежит его руководителю. Именно от его опыта, таланта и терпения зависит, сменит ли первоначальный пыл юных исследователей вдумчивая работа, или всё так и останется в зачаточной стадии. Необходимо наблюдать за каждым студентом, стараться предсказать проблемы, которые могут возникнуть у него в процессе работы. Может случиться так, что молодой человек постесняется задать вопрос, считая себя достаточно взрослым для его самостоятельного решения, а затем, так и не придя к ответу, откажется от исследования вообще, приняв решение о собственной научной несостоятельности. Такие психологические проблемы часто встают перед студентами младших курсов. Причиной является сложившийся стереотип, что студент-это уже полностью сложившийся человек, и сам должен решать свои проблемы. На самом деле же, мышление студентов младших курсов ещё несёт в себе большой отпечаток школьного и, говоря откровенно, просто детского. Поэтому конфликт между «взрослой» моделью поведения и юношеским мышлением может перечеркнуть усилия самого талантливого, но недостаточно чуткого педагога. Поэтому будет не лишним прочитать студентам две-три лекции о методах и способах научного исследования, о сборе материала, о работе над литературой, о пользовании научным аппаратом, а так же ознакомит студентов с научными направлениями преподавателей кафедры, чтобы студенты знали, к кому можно обратиться для более детальной консультации по некоторым вопросам.

Если начальный период работы кружка прошёл успешно, и большая часть тем принята в работу, то составляется график выступлений, и начинается заслушивание готовых докладов. Как правило, на одном заседании кружка заслушивается не более двух выступлений, так как только в данном случае можно подробно обсудить каждый доклад, задать вопросы и получить развёрнутые ответы на них. Кроме этого, большое количество докладов трудно для восприятия, и может снизиться активность и заинтересованность членов кружка.

Формами подведения итогов работы кружка могут стать конкурс докладов, участие в научных конференциях и предметных олимпиадах, проведение круглых столов, встречи с учёными, а так же публикация тезисов лучших работ в научных сборниках ВУЗов.

Всё сказанное о научных кружках можно отнести и к проблемным, но следует учесть некоторые отличия.

* Проблемный кружок может объединять собой студентов разных факультетов и курсов, а так же, если при ВУЗе имеется таковые, колледжей и лицеев. Во главу угла может быть поставлена проблема, которой занимается научный руководитель кружка, или любая другая по его выбору. Большим достоинством данной формы НИРС является возможность рассмотрения выбранной темы наиболее глубоко и с разных ракурсов. Так, например тему «Безработица в России» может быть рассмотрена с экономической (влияние безработицы на ВНП, государственная политика в отношении безработицы и т. д.), социальной (социальный состав безработных, социальные последствия безработицы и т. д.), культурной (безработица и культура, народный фольклор о безработице и т. д.), и даже литературной (безработица в произведениях российских писателей) точек зрения. Это придаёт заседаниям кружка большую разносторонность и привлекает в него новых членов. Кроме того, что немаловажно, это способствует укреплению связей между студентами разных возрастов и специальностей, поддерживает чувство единого коллектива.

* Проблемные кружки представляю собой «облегчённую» форму НИРС, и поэтому на их базе возможно организация встреч с людьми, которые сталкиваются с проблемами, выбранными кружком для рассмотрения, на работе и в быту, проведение различных викторин и КВН.

* Проблемный кружок может сочетать в себе элементы научного кружка, лаборатории и т. д.

ПСТ относятся к следующей ступени сложности НИРС. В них принимают участие студенты второго курса и старше. Лаборатория не является школой научной работы, занятия в ней предполагают определённый запас знаний и навыков. В рамках ПСТ осуществляются различные виды моделирования, изучение и анализ реальных документов, программ, деловых игр, а так же практическая помощь предприятиям. Работа в такой лаборатории предполагает не столько изучение и анализ литературы, сколько постановку эксперимента, создание чего-то нового. ПСТ, скорее всего, будут не столь многочисленны, как научные и проблемные кружки. Происходит отсев студентов, когда из способных выбираются ещё более способные.

Ещё одним отличием ПСТ от кружка является большее значение способности студента к коллективной работе. Если в кружке каждый студент отвечает, как правило, только за себя, то в ПСТ, где темы исследований гораздо более глобальные, одной самостоятельной работой обойтись практически невозможно. Руководитель лаборатории должен помочь студентам разделить тему на отдельные вопросы, решение которых приведёт к решению главной проблемы. Важно внимание к интересам каждого студента, к его склонностям и возможностям. Опыт коллективной работы приходит не сразу, и разрешение споров и конфликтов, возникающих в процессе работы, так же во многом лежит на плечах преподавателя.

Работа в ПСТ воспитывает «не мальчика, но мужа». В процессе этой работы студент может полученные за время учёбы и работы в кружках знания реализовать в исследованиях, имеющих практическое значение. Так, некоторые предприятия города Саратова обращаются к кафедрам бухгалтерского учёта и аудита СГЭА с просьбой проанализировать их годовую отчётность. Студенческие лаборатории при кафедрах подключаются к этой работе, получая от неё не только моральное, но иногда и материальное вознаграждение. Кроме того, заинтересовавшие предприятия студенты в последствии могут быть приглашены на работу в них, что во время отсутствия государственного распределения является ценным результатом.

Таким образом, работа в ПСТ - следующий важный шаг к полноценной научно-исследовательской работе и ценный опыт для дальнейшей научной и практической деятельности.

Каждый из указанных выше типов конкуренции является итогом проделанной работы: научных исследований, работы в лаборатории, практики по специальности.

На конференции молодые исследователи получают возможность выступить со своей работой перед широкой аудиторией. Это заставляет студентов более тщательно прорабатывать будущее выступление, оттачивает его ораторские способности. Кроме того, каждый может сравнить, как его работа выглядит на общем уровне и сделать соответствующие выводы. Это является очень полезным результатом научной конференции, так как на раннем этапе многие студенты считают собственные суждения непогрешимыми, а свою работу - самой глубокой и самой ценной в научном плане. Часто даже замечания преподавателя воспринимаются как простые придирки. Но слушая доклады других студентов, каждый не может не заметить недостатков своей работы, если таковые имеются, а так же выделить для себя свои сильные стороны.

Кроме того, если в рамках конференции проводится творческое обсуждение прослушанных докладов, то из вопросов и выступлений каждый докладчик может почерпнуть оригинальные идеи, о развитии которых в рамках выбранной им темы он даже не задумывался. Включается своеобразный механизм, когда одна мысль порождает несколько новых.

Научно-практические конференции, уже исходя из самого названия, включают в себя не только и не столько теоретические научные доклады, сколько обсуждение путей решения практических задач. Очень часто они проводятся вне стен ВУЗа, а на территории завода, фабрики, колхоза, фермерского хозяйства, управляющего органа, с которыми ВУЗ поддерживает отношения. Например, научно-практическая конференция может проводиться по результатам летней практики студентов, когда последние, столкнувшись с определёнными проблемами, могут с помощью работников предприятия и преподавателей попытаться найти пути их решения. Такие конференции способствуют установлению тесных дружеских связей между ВУЗом и предприятиями, а также помогают студентам учиться применять изученную теорию на практике. Отличительной чертой научно-практической конференции является сложность её слаженной организации, так, чтобы участие в ней было одинаково полезно и интересно и студентам, и работникам предприятия. Разработка и проведение такой конференции требует от организаторов и участников большого внимания и терпения.

Формы и методы привлечения студентов к научному творчеству

Современное **понятие «научно-исследовательская работа студентов»** (НИРС) включает в себя два взаимосвязанных элемента:

- обучение студентов элементам исследовательского труда, привитие им навыков этого труда;
- собственно научные исследования, проводимые студентами под руководством профессоров и преподавателей.

Формы и методы привлечения студентов к научному творчеству условно подразделяются на НИР, включенную в учебный процесс, а также НИР, выполняемую студентами во внеучебное время.

Учебно-исследовательская работа выполняется в отведенное расписанием занятий учебное время по специальному заданию в обязательном порядке каждым студентом. Основной задачей УИРС является обучение студентов навыкам самостоятельной теоретической и экспериментальной работы, ознакомление с реальными условиями труда в лаборатории, в научном коллективе.

К таким занятиям относятся:
- лекции по дисциплине «НИР»;
- практические занятия с элементами научных исследований по дисциплине;
- курсовое и дипломное проектирование с элементами научных исследований.

Основной **формой научной работы студентов**, выполняемой **во внеучебное время**, является участие студентов в научных исследованиях, проводимых преподавателями кафедр и сотрудниками научных учреждений вуза по госбюджету и хоздоговорной тематике.

Формы творческой работы студентов: *студенческие КБ, проектные, технологические, исследовательские бюро (СКБ), научные и вычислительные центры, научно-производственные отряды*.

Формы научно-исследовательской работы студентов, проводимой на кафедре достаточно разнообразны:
- подготовка докладов по экономической и методической проблематике для ежегодных научно-теоретических конференций;
- проблемные группы;
- написание курсовых работ по методике и экономике;
- написание выпускных квалификационных (дипломных) работ;
- написание рефератов по теоретическим дисциплинам;
- написание научных статей, тезисов.

3. Работа с источниками. Этические аспекты научно-исследовательской деятельности

От степени серьёзности подхода к изучению литературных и электронных источников информации во многом зависит общий итог, теоретическая и практическая ценность результатов, полученных при написании диссертации. Как правило, изучение следует начать с общетеоретических работ, которые, как правило, известны и используются для формирования теоретического фундамента работы и выявления тех проблем и тем, которые прямо или опосредовано связаны с темой диссертации аспиранта/соискателя.
С целью проникновения в глубинный смысл научных книг и сборников статей необходимо работать с карандашом, выписывая или подчёркивая (в собственном экземпляре книги или в ксерокопии) наиболее важные моменты, которые можно использовать в диссертации. Поиск нового материала становится возможным тогда, когда у аспиранта/соискателя в сознании сложится общая картина состояния изучаемой проблемы на данный момент. В процессе работы с литературой можно выделить несколько этапов и важных моментов, на которые следует обратить внимание, в частности:
• общий просмотр содержания работы в целом;
• выявление наиболее близких к теме диссертации аспиранта/соискателя глав;
• последовательное прочтение книги (статьи, доклада и т.п.) с выписыванием наиболее значимых, на взгляд аспиранта, положений;
• анализ собранных записей по данной публикации, аккумуляция наиболее значимых мест и редактирование для применения в тексте собственной диссертации с указанием ссылок на источник;
• письменная фиксация собственных умозаключений и выводов, приходящих по мере изучения научных источников;
• обучение в процессе работы с чужими исследованиями концентрации внимания на информации, непосредственно связанной с темой диссертации;
• поиск способов использования прочитанной информации для получения собственной точки зрения на проблему и практического применения ее при написании диссертации;
• обретение навыков отбора нужной информации, чтобы «не утонуть» в обилии материала.

• дифференцирование основной и второстепенной выбранной научной информации, анализ и самостоятельные обобщения.

Изучение каждого информационного источника необходимо заканчивать подробным обобщением - анализом полученной информации с выделением наиболее значимых фактов и данных.

В ходе творческого исследования литературных источников и библиографических описаний следует также помнить о необходимости поиска связи узкого вопроса с темой в целом, и, наоборот, в широкой проблеме уметь выделить наиболее значимые детали и особенности, имеющие непосредственное отношение к теме собственной диссертации.

При этом целью при сборе информации для будущей работыявляется поиск научных фактов, а не просто фактов в широком понимании слова. Для научного факта характерны такие признаки, как новизна, точность, достоверность и объективность. Научный факт должен вписываться в концепцию научного познания мира, в основе которого находится отражение объективных процессов, явлений, свойств, позволяющих прослеживать закономерности и формировать законы, по которым происходит развитие окружающей действительности. При этом *новизна* научного факта подразумевает получение нового знания при применении определённых подходов при изучении каких-либо процессов и явлений или получении новых данных и показателей при изменении условий существования и развития процессов и явлений. Критерии *точности* научного факта основаны на доказанной объективности методов и средств изучения свойств и признаков явлений и предметов. *Достоверность* научных фактов не всегда может быть доказана на момент их обнаружения, некоторые научные факты находят подтверждение через достаточно длительный промежуток времени, после их выявления. Распространённым на сегодняшний день подтверждением достоверности научного факта является существование аналогичных явлений и процессов. Или же достоверным может считаться научная информация, взятая из достоверных источников, к которым относят официальную информацию, идущую от государственных научно-исследовательских организаций или ведущих учебных учреждений, ведомств, занимающихся государственными проектами, законотворчеством и т.п. Достоверность научно-исследовательской информации может быть подтверждена экспериментально-практической деятельностью, проводимой на базе научно-исследовательского учреждения.

В большой степени достоверность научного факта зависит от первоисточника, на базе разработок или заключений которого выводится новое знание. Понятие научной о*бъективности* вызывает наибольшее разночтение. В точных науках, где утверждения могут быть доказаны, объективность новых научных знаний заслуживает большего доверия. В гуманитарных науках объективность подтвердить сложнее – чаще всего она представляется выявленной, если научные факты опять-таки, подтверждены признанными объективными первоисточниками. В остальных случаях в значительной мере на те или иные утверждения накладываются субъективные знания и черты личности, проявляющиеся процессе научного творчества.

Этические аспекты научно-исследовательской деятельности

Этические аспекты научной деятельности. Понятие научного этоса и проблема его современного расширения. Этика науки изучает нравственные основы научной деятельности, совокупность ценностных принципов, принятых в научном сообществе, и концентрирует в себе социальный и гуманистический аспекты науки. Этические проблемы современной науки являются чрезвычайно актуальными и значимыми. Противоречие между этическими нормами и необходимостью технического бытия человека ведет за собой обширный класс этических проблем мира искусственного. Современная техника помещает человека в условия, далеко отстоящие от его прежнего состояния. Небывалое расширение технических возможностей общества сопровождается тем, что в ряде исследований объектом становится сам человек. А это в свою очередь создает определенную угрозу его здоровью и существованию. Первыми столкнулись с этими проблемами физики-ядерники. Ныне угрозы затрагивают и область молекулярной биологии, генетики, медицины и т. д. Ситуация, связанная с созданием атомной бомбы, а также новейших смертоносных видов вооружения, ставит задачи гуманного контроля над наукой в качестве приоритетных и первостепенных. Другая сторона этических проблем относится к вопросам авторства научных открытий, плагиата, компетентности и фальсификации научных открытий. Здесь довольно жесткие морально-этические санкции, не говоря уже о юридических нормативах по охране интеллектуальной собственности. На страже этических принципов стоит институт ссылок, как академическая составляющая науки. Этос науки – правило деятельности ученого, отвечает следующим требованиям: 1) универсализм (неличностный характер научного знания, его объективность, деятельность в области науки не может иметь никаких национальных и расовых ограничений, а также соц. и имущественных) 2 коллективизм (научные результаты не должны быть скрываемыми, научное познание всегда есть процесс коллективного творчества) 3) бескорыстие (ученый должен в своей деятельности руководствоваться принципом поиска истинности) 4) критицизм (критическое отношение к себе самому, к своим предшественникам и современникам). Этос науки направлен на защиту науки от лженауки. Ученый может ошибаться, но не может фальсифицировать. Научное сообщество отторгает исследователей, занимающихся плагиатом, бойкотирует их. Весьма значимыми становятся этические проблемы, исходящие из увеличения технизации медицины и появления принципиально иных, новых медицинских технологий и препаратов, которые расширяют возможности воздействия на человека. Нужны жесткие критерии, допускающие экспериментирование на человеке. Важно исключить опасность разрушения исходной биогенетической основы человека, сейчас есть возможность вмешиваться в генетический код человека, изменять его целенаправленно. Революционная ситуация в генетике породила этическую проблему клонирования. Одно дело целесообразен ли запрет на клонирование животного мира: приобретение элитных коров, пушных зверей и т. д. А в вопросе клонирования человека возникает медицинский, этический, философский, религиозный, экономический и прочие аспекты. Клонирование человека преступно и аморально. Этическое регулирование науки и появление высокого уровня этической культуры, оцениваемые сегодня как жизненная необходимость, являются важной предпосылкой будущего развития науки. Это будет способствовать обеспечению качества моральности современной науки. Ученый должен проникнуться сознанием своей ответственности за судьбу человечества.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1. Практическое занятие 1 (ПЗ-1) (2 часа)

Тема: Организация научно-исследовательской работы

2.1.1 Задание для работы:

1. Исследования в области вещества и энергии
2. Исследования в области новых материалов и технологий
3. Построение глобального информационного пространства

2.1.2 Краткое описание проводимого занятия:

1. Исследования в области вещества и энергии – доклады с презентациями по теме

2. Исследования в области новых материалов и технологий – доклады с презентациями по теме

3. Построение глобального информационного пространства – доклады с презентациями по теме

2.1.3 Результаты и выводы: В результате проведенного занятия студенты:
- ознакомились с современными направлениями в исследованиях вещества и энергии, новых материалов и технологий;
- усвоили основные принципы построения глобального информационного пространства;
- выработали навыки по изучению интегративных качеств фундаментальных моделей.

2.2. Практическое занятие 2,3,4,5,6,7,8 (ПЗ 2,3,4,5,6,7,8) (14 часов)

Тема: Стохастический метод исследования распределения

2.2.1 Задание для работы:

1.Первичная обработка статистических данных.Графическое представление статистических рядов. Эмпирическая функция распределения статистических рядов.
2.Чсловые характеристики статистического ряда, их свойства.
3. Точечные оценки параметров статистического распределения. Оценки параметров генеральной совокупности. Метод моментов.
4.Метод доверительных интервалов.

2.2.2 Краткое описание проводимого занятия:

1.Первичная обработка статистических данных. Графическое представление статистических рядов. Эмпирическая функция распределения статистических рядов.
Пример. Записать в виде вариационного и статистического рядов выборку 5, 3, 7, 10, 5, 5, 2, 10, 7, 2, 7, 7, 4, 2, 4. Определить размах выборки.
Решение. В данном случае объем выборки $n = 15$. Упорядочим элементы выборки по ве-

личине, получим вариационный ряд 2, 2, 3, 4, 4, 5, 5, 5, 7, 7, 7, 7, 10, 10. Найдем размах выборки ω=10-2= 8. Различными в заданной выборке являются элементы $z_1 = 2$, $z_2 =3$, $z_3 = 4$, $z_4 = 5$, $z_5 = 7$, $z_6 = 10$; их частоты соответственно равны $n_1 = 3$, $n_2=1$, $n_3 = 2$, $n_4 = 3$, $n_5 = 4$, $n_6 = 2$. Статистический ряд исходной выборки можно записать в виде следующей таблицы:

z_i	2	3	4	5	7	10
n_i	3	1	2	3	4	2

Для контроля правильности записи находим $\sum n_i = 15$. При большом объеме выборки ее элементы рекомендуется объединять в группы (разряды), представляя результаты опытов в виде *группированного статистического ряда.* В этом случае интервал, содержащий все элементы выборки, разбивается на k непересекающихся интервалов. Вычисления упрощаются, если эти интервалы имеют одинаковую длину $b \approx \dfrac{\omega}{k}$. В дальнейшем рассматривается именно этот случай. После того как частичные интервалы выбраны, определяют частоты - количество n_i элементов выборки, попавших в i-й интервал (элемент, совпадающий с верхней границей интервала, относится к следующему интервалу). Получающийся статистический ряд в верхней строке содержит середины z_i интервалов группировки, а в нижней — частоты n_i ($i = 1,2,...,k$).

Наряду с частотами одновременно подсчитываются также накопленные частоты $\sum\limits_{j=1}^{i} n_j$,

относительные частоты n_i /n и *накопленные относительные частоты* $\sum\limits_{j=1}^{i} n_j / n$, $i = 1,2,...,k$.

Полученные результаты сводятся в таблицу, называемую *таблицей частот группированной выборки.* Следует помнить, что группировка выборки вносит погрешность в дальнейшие вычисления, которая растет с уменьшением числа интервалов.

Пример. Представить выборку 55 наблюдений в виде таблицы частот, разбив имеющиеся данные выборки на семь интервалов группировки. Выборка:

20,3	15,4	17,2	19,2	23,3	18,1	21,9
15,3	16,8	13,2	20,4	16,5	19,7	20,5
14,3	20,1	16,8	14,7	20,8	19,5	15,3
19,3	17,8	16,2	15,7	22,8	21,9	12,5
10,1	21,1	18,3	14,7	14,5	18,1	18,4
13,9	19,1	18,5	20,2	23,8	16,7	20,4
19,5	17,2	19,6	17,8	21,3	17,5	19,4
17,8	13,5	17,8	11,8	18,6	19,1	

В данном случае размах выборки ω=23,8 -10,1 = 13,7; тогда длина интервала группировки будет b = 13,7/7≈2 . В качестве первого интервала возьмем интервал 10 - 12. Результаты группировки сведем в таблицу 1

Таблица 1

Номер интервала i	Границы интервала	Середина интервала z_i	Частота n_i	Накопленная частота $\sum\limits_{j=1}^{i} n_j$	Относительная частота n_i /n	Накопленная относительная частота $\sum\limits_{j=1}^{i} n_j / n$
1	10-12	11	*2*	2	0,0364	0 0364
2	12-14	13	4	6	0,0727	0 1091
3	14-16	15	8	14	0,1455	0 2546
4	16-18	17	12	26	0,2182	0,4728
5	18-20	19	16	42	0,2909	0 7637

6	20-22	21	10	52	0,1818	0,9455
7	22-24	23	3	55	0,0545	1,0000

Пример. Построить гистограмму и полигон частот, а также график эмпирической функции распределения группированной выборки из примера 29.

Решение. По результатам группировки (см. таблицу 1.) строим гистограмму частот (рис. 1). Соединяя отрезками ломаной середины верхних оснований прямоугольников, из которых состоит полученная гистограмма, получаем соответствующий полигон частот (рис. 2).

Так как середина первого интервала группировки $z_1 = 11$, то $F_n^*(x) = 0$ при $x \le 11$. Рассуждая аналогично, находим, что $F_n^*(x) = 1$ при $x > 23$. На полуинтервале (11,23] эмпирическую функцию распределения строим по данным третьего и последнего столбцов таблицы 1.

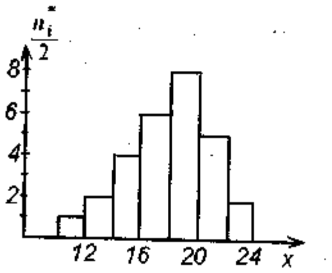


Рис.1

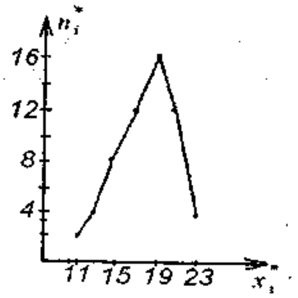


Рис.2

$F_n^*(x)$ имеет скачки в точках, соответствующих серединам интервалов группировки. В результате получаем график $F_n^*(x)$, изображенный на рис. 3.

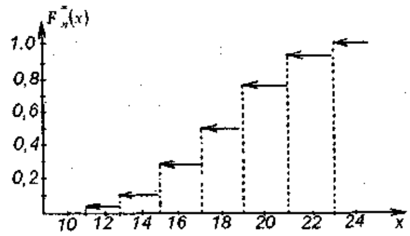


Рис.3

Задача. Построить полигон частот и эмпирическую функцию по данному распределению выборки:

Варианты *XI*	-3	0	1	4	6	7
Частоты *Ni*	3	6	1	2	5	1

Решение. *Полигоном частот* называют ломаную, отрезки которой соединяют точки (x_1, n_1); (x_2, n_2);…;(x_k, n_k), где *XI* – варианты выборки, *Ni* – соответствующие им частоты. Полигон частот для данного распределения изображен на рисунке 4.

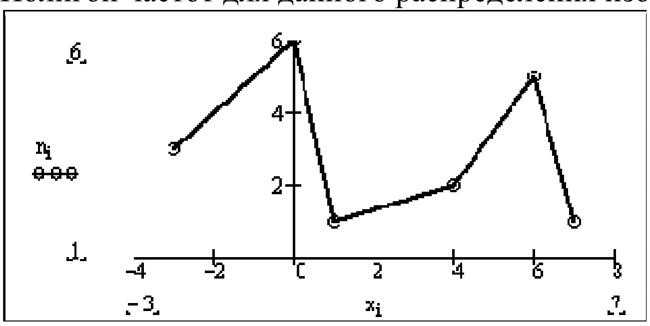


Рис. 4

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения *X* относительную частоту события $X < x$: $F^*(x) = \frac{n_x}{n}$,

где n_x – число вариант, меньших *X*; *N* – объем выборки.

Из определения следует, что $0 \le F^*(x) \le 1$.
Найдем эмпирическую функцию распределения.

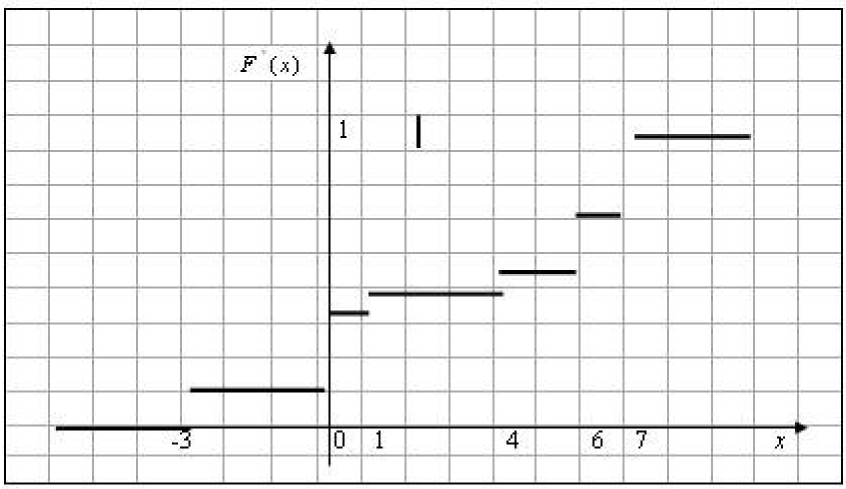
Объем данной выборки равен $n = \sum_{i=1}^{k} n_i$ *=18.*

Если $x \le -3$, то $F^*(x)$ *=0* (так как *-3* – наименьшая варианта). Если $-3 < x \le 0$, то значение $X < 0$, а именно $x_1 = -3$ наблюдалось 3 раза, следовательно, $F^*(x) = \frac{3}{18}$. При $0 < x \le 1$ значения $X < 1$, а именно $x_1 = -3$ и $x_2 = 0$ наблюдались 3+6=9 раз, следовательно, $F^*(x) = \frac{9}{18}$.

Аналогично получаем, что при $1 < x \le 4$ функция распределения $F^*(x) = \frac{10}{18}$; при $4 < x \le 6$ функция распределения $F^*(x) = \frac{12}{18}$; при $6 < x \le 7$ функция распределения $F^*(x) = \frac{17}{18}$. Далее, если $x > 7$, то $F^*(x) = 1$ (так как 7 – наибольшая варианта).
Таким образом, эмпирическая функция распределения равна:

$$F^*(x) = \begin{cases} 0 & \text{при } x \le -3 \\ 3/18 & \text{при } -3 < x \le 0 \\ 9/18 & \text{при } 0 < x \le 1 \\ 10/18 & \text{при } 1 < x \le 4 \\ 12/18 & \text{при } 4 < x \le 6 \\ 17/18 & \text{при } 6 < x \le 7 \\ 1 & \text{при } x > 7 \end{cases}$$

График полученной эмпирической функции распределения изображен на рисунке .



2. Числовые характеристики статистического ряда, их свойства.

Пример. Определить среднее, моду и медиану для выборки 5, 6, 8, 2, 3, 1, 1, 4.
Решение. Представим данные в виде вариационного ряда: 1, 1, 2, 3, 4, 5, 6, 8. Выборочное

среднее $\bar{x} = \frac{1}{8}(1+1+2+3+4+5+6+8) = 3{,}75$. Все элементы входят в выборку по одному разу, кроме 1, следовательно, мода $\tilde{d}_X = 1$. Так как n = 8, то медиана $\tilde{h}_X = \frac{1}{2}(3+4) = 3{,}5$.

Итак, $\bar{x} = 3{,}75$, $\tilde{d}_X = 1$, $\tilde{h}_X = 3{,}5$.

Для упрощения вычислений выборочных среднего и дисперсии группированной выборки, эту выборку преобразуют так: $u_i = \frac{1}{b}\left(z_i - d_X^*\right)$, i= 1,2,...,k, где d_X^* - выборочная мода, а b - длина интервала группировки. Эти соотношения показывают, что в выборку $z_1, z_2, ..., z_n$ внесена систематическая ошибка d_X^*, а результат подвергнут преобразованию масштаба с коэффициентом $k = 1/b$. Полученный в результате набор чисел $u_1, u_2, ..., u_n$ можно рассматривать как выборку из генеральной совокупности $U = \frac{1}{b}\left(x - d_X^*\right)$. Тогда выборочные среднее $\bar{x}$ и дисперсия D_X^* исходных данных связаны со средним $\bar{u}$ и дисперсией D_U^* преобразованных данных следующими соотношениями: $\bar{x} = b\bar{u} + d_X^*$, $D_X^* = b^2 D_U^*$.

Пример. Вычислить среднее и дисперсию группированной выборки

Границы интервалов	134-138	138-142	142-146	146-150	150-154	154-158
Частоты	1	3	15	18	14	2

Решение. Длина интервала группировки b = 4, значение середины интервала, встречающегося с наибольшей частотой $d_X^* = 148$. Преобразование последовательности середин интервалов выполняется по формуле:

$$u_i = \frac{z_i - 148}{4}, \text{ где } i = 1,2,...,6.$$

Таблица 2

i	z_i	u_i	n_i	$n_i u_i$	$n_i u_i^2$	$n_i (u_i + 1)^2$
1	136	-3	1	-3	9	4
2	140	-2	3	-6	12	3
3	144	-1	15	-15	15	0
4	148	0	18	0	0	18
5	152	1	14	14	14	56
6	156	2	2	4	8	18
$\sum$	-	-	53	-6	58	99

Вычисления сведены в таблицу 2. Последний столбец этой таблицы служит для контроля вычислений при помощи тождества $\sum n_i(u_i+1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + \sum n_i$.
Выполняя вычисления, получим 58 + 2 · (-6) + 23 = 99 .
Полученный результат показывает, что вычисления выполнены правильно. По формулам, данным выше, находим средние значения U

$\bar{u} = \frac{-6}{53} \approx -0{,}113$, $D_U^* = \frac{58 - (-6)^2/53}{53} \approx 1{,}108$. Далее находим средние данной выборки:
$\bar{x} \approx (-0{,}113) \cdot 4 + 148 \approx 147{,}548$, $D_X^* \approx 4^2 \cdot 1{,}103 \approx 17{,}728$.

Пример. Вычислить среднее, дисперсию, коэффициенты асимметрии и эксцесса для следующей группированной выборки:

Границы интервалов	10-12	12-14	14-16	16-18	18-20	20-22	22-24
Частоты	2	4	8	12	16	10	3

Длина интервала группировки $b = 2$. Значение z_i , встречающееся с наибольшей частотой, $d_X^* = 19$. Поэтому преобразование имеет вид $u_i = \frac{z_i - 19}{2}$, где $i = 1,2,...,7$.

Все вычисления оформим в виде таблицы 3.

Таблица 3

i	z_i	u_i	n_i	$n_i u_i$	u_i^2	$n_i u_i^2$	u_i^3	$n_i u_i^3$	u_i^4	$n_i u_i^4$	u_i+1	$(u_i+1)^4$	$n_i(u_i+1)^4$
1	11	-4	2	-8	16	32	-64	-128	256	512	-3	81	162
2	13	-3	4	-12	9	36	-27	-108	81	324	-2	16	64
3	15	-2	8	-16	4	32	-8	-64	16	128	-1	1	8
4	17	-1	12	-12	1	12	-1	-12	1	12	0	0	0
5	19	0	16	0	0	0	0	0	0	0	1	1	16
6	21	1	10	10	1	10	1	10	1	10	2	16	160
7	23	2	3	6	4	12	8	24	16	48	3	81	243
$\sum$	-	-	55	-32	-	134	-	-278	-	1034	-	-	653
$\sum \frac{n_i u_i'}{n}$			-	-0,582	-	2,436	-	-5,054	-	18,8	-	-	-

Контроль вычислений будет $54 + 4\cdot(-32) + 6\cdot 134 + 4\cdot(-278) + 1034 = 653$.

Находим искомые характеристики выборочного распределения:

$$\bar{x} = 19 + 2\cdot\frac{-32}{55} \approx 17{,}8,\quad D_U^* = \frac{134 - (-32)^2/55}{55} \approx 2{,}10,\quad D_X^* = 2^2\cdot 2{,}10 = 8{,}40,$$

$$a_X^* = \frac{1}{2{,}10^{3/2}}\left[-5{,}054 - 3\cdot(-0{,}582)\cdot 2{,}436 + 2(-0{,}582)^3\right] \approx -0{,}393,$$

$$e_X^* = \frac{1}{2{,}10^2}\left[18{,}8 - 4\cdot(-0{,}582)\cdot(-5{,}054) + 6\cdot(-0{,}582)^2\cdot 2{,}436 - 3(-0{,}582)^4\right] - 3 \approx -0{,}303.$$

3. Точечные оценки параметров статистического распределения. Оценки параметров генеральной совокупности. Метод моментов.

Задача. Из генеральной совокупности извлечена выборка объема N, заданная вариантами XI и соответствующими им частотами. Найти несмещенную оценку генеральной средней.

Варианта XI	2	5	7	10
Частота Ni	16	12	8	14

Решение. Множество всех объектов, подлежащих изучению, называется *Генеральной совокупностью*. Множество случайно отобранных объектов называется выборочной совокупностью или *Выборкой*.

Для оценки неизвестных параметров теоретического распределения служат статистические оценки. Статистическая оценка, определяемая одним числом, называется *Точечной оценкой.*

Точечная статистическая оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки, называется *Несмещенной оценкой.* Статистическая оценка, математическое ожидание которой не равно оцениваемому параметру является *Смещенной*.

Несмещенной оценкой генеральной средней (математического ожидания) служит выборочная средняя $\overline{x_B} = \frac{\sum_{i=1}^{k} n_i x_i}{n}$ (1),

Где X_I – варианта выборки (элемент выборки); Ni – частота варианты XI (число наблюдений варианты XI); $n = \sum_{i=1}^{k} n_i$ – объем выборки (число элементов совокупности).

Объем данной выборки равен $n = 16 + 12 + 8 + 14 = 50$.

Далее по формуле (1) вычисляем несмещенную оценку генеральной средней:

$$\overline{x}_B = \frac{2 \cdot 16 + 5 \cdot 12 + 7 \cdot 8 + 10 \cdot 14}{50} = 5{,}76$$

Задача. По выборке объема $N=41$ найдена смещенная оценка генеральной дисперсии $D_B = 3$. Найти несмещенную оценку дисперсии генеральной совокупности.

Решение. Смещенной оценкой генеральной дисперсии служит выборочная дисперсия

$$D_B = \frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{x_B}\right)^2}{n}$$

Несмещенной оценкой генеральной дисперсии является «исправленная дисперсия»

$$s^2 = \frac{n}{n-1} D_B \text{ или } s^2 = \frac{\sum_{i=1}^{k} n_i \left(x_i - \overline{x_B}\right)^2}{n-1}$$

Таким образом, мы получаем искомую несмещенную оценку дисперсии генеральной совокупности:

$$s^2 = \frac{41}{41-1} \cdot 3 = 3{,}075$$

Пример. Методом моментов найти оценки неизвестных параметров a и b для Г - распределения с плотностью

$$f_X(x) = \begin{cases} 0, \ x \le 0 \\ \dfrac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \ x > 0 \end{cases}.$$

Решение. Для нахождения оценок параметров a и b по методу моментов воспользуемся начальным моментом первого порядка (математическим ожиданием) и центральным моментом второго порядка (дисперсией):

$$\alpha_1(a,b) = m = \frac{a}{b}, \ \mu_2(a,b) = \sigma^2 = \frac{a}{\sigma^2}.$$

По выборке $x_1,...,x_n$ из генеральной совокупности, имеющей Г-распределение, находим значения соответствующих выборочных моментов:

$$\alpha_1^* = \overline{x} = \frac{1}{n}\sum x_i, \ \mu_1^* = D_X^* = \frac{1}{n}\sum (x_i - \overline{x})^2.$$

Приравнивая соответствующие равенства, получаем следующую систему уравнений:

$\frac{a}{b} = \overline{x}$, $\frac{a}{b^2} = D_X^*$. Решая ее, находим $\tilde{a} = \frac{\overline{x^2}}{D_X^*}$, $\tilde{b} = \frac{\overline{x}}{D_X^*}$.

4. Метод доверительных интервалов.

Задача. Найти доверительный интервал для оценки с надежностью *P=0,95* неизвестного математического ожидания *A* нормально распределенного признака *X* генеральной совокупности, если даны генеральное среднее квадратическое отклонение *S=5*, выборочная средняя = 14, а объем выборки *N=25*.

Решение. *Интервальной оценкой* называется интервал, покрывающий оцениваемый параметр. *Доверительным интервалом* является интервал, который с данной надежностью покрывает оцениваемый параметр.

Для оценки математического ожидания *A* нормально распределенного количественного признака *X* по выборочной средней $\overline{x_B}$ при известном среднем квадратическом отклонении s генеральной совокупности служит доверительный интервал

$\overline{x_B} - t\frac{\sigma}{\sqrt{n}} < a < \overline{x_B} + t\frac{\sigma}{\sqrt{n}}$, где $t\frac{\sigma}{\sqrt{n}} = \delta$ – точность оценки, *T* – значение аргумента функции Лапласа В данной задаче *T* находим из условия $2\Phi(t) = 0{,}95$. По таблице 2 определяем $\Phi(t) = 0{,}475$. Таким образом, *T=1,96.*

Далее получаем
$$14 - 1{,}96 \cdot \frac{5}{\sqrt{25}} < a < 14 + 1{,}96 \cdot \frac{5}{\sqrt{25}}$$

Или $12{,}04 < a < 15{,}96$

Задача. По данным *N=9* независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\overline{x_B} = 30{,}1$ и исправленное среднее квадратическое отклонение *S=6.* Оценить истинное значение измеряемой величины при помощи доверительного интервала с надежностью γ *=0,99.*

Решение. Оценкой математического ожидания *A* нормально распределенного количественного признака *X* в случае неизвестного среднего квадратического отклонения является доверительный интервал

$\overline{x_B} - t_\gamma\frac{s}{\sqrt{n}} < a < \overline{x_B} + t_\gamma\frac{s}{\sqrt{n}}$. По таблице приложения, по заданным *N* и γ находим

t_γ *=3,36.*

Таким образом
$$30{,}1 - 3{,}36 \cdot \frac{6}{\sqrt{9}} < a < 30{,}1 + 3{,}36 \cdot \frac{6}{\sqrt{9}}$$

Окончательно получаем $23{,}38 < a < 36{,}82$

Задача. Из генеральной совокупности извлечена выборка объема *N.* Оценить с надежностью γ *=0,95* математическое ожидание *A* нормально распределенного признака *X* генеральной совокупности по выборочной средней с помощью доверительного интервала.

Значение признака *XI*	-2	1	1	3	4	5
Частота *Ni*	2	1	2	2	2	1

Решение. Объем данной выборки равен $n = \sum_{i=1}^{k} n_i = 2+1+2+2+2+1=10$

По данным задачи находим выборочную среднюю:

$$\overline{x_B} = \frac{\sum_{i=1}^{k} n_i x_i}{n} = \frac{-2\cdot 2+1\cdot 1+1\cdot 2+3\cdot 2+4\cdot 2+5\cdot 1}{10} = 1{,}8$$

Далее находим исправленное среднее квадратическое отклонение S:

$$s = \sqrt{\frac{\sum n_i \left(x_i - \overline{x_B}\right)^2}{n-1}} = \sqrt{\frac{2\cdot 14{,}44+1\cdot 0{,}64+2\cdot 0{,}64+2\cdot 1{,}44+2\cdot 4{,}84+1\cdot 10{,}24}{9}} = \sqrt{\frac{53{,}6}{9}} \approx 2{,}44$$

Для оценки математического ожидания A нормально распределенного количественного признака X в случае неизвестного среднего квадратического отклонения служит доверительный интервал

$$\overline{x_B} - t_\gamma \frac{s}{\sqrt{n}} < a < \overline{x_B} + t_\gamma \frac{s}{\sqrt{n}}.$$

По таблице 3 приложения по заданным N и γ находим $t_\gamma = 2{,}26$.

Таким образом

$$1{,}8 - 2{,}26\cdot\frac{2{,}44}{\sqrt{10}} < a < 1{,}8 + 2{,}26\cdot\frac{2{,}44}{\sqrt{10}}$$

Окончательно получаем $0{,}05 < a < 3{,}55$

Пример. Найти доверительный интервал для оценки математического ожидания с надежностью 0,95, если объем выборки n =16, среднее выборочное и исправленная дисперсия соответственно равны 20,2 и 0,8.

По таблице приложения найдем t_γ по заданной надежности γ =0,95 и n= 16: t_γ =2,13. Подставим в формулу s =0,8 и t_γ=2,13 , вычислим границы доверительного интевала:

$20{,}2 - \frac{2{,}13\cdot 0{,}8}{4} < a < 20{,}2 + \frac{2{,}13\cdot 0{,}8}{4}$, откуда получим доверительный интервал (19,774; 20,626)

Смысл полученного результата: если взять 100 различных выборок, то в 95 из них математическое ожидание будет находится в пределах данного интервала, а в 5 из них- нет.

Пример. Измеряют диаметры 25 корпусов электродвигателей. Получены выборочные характеристики $\overline{x} = 100мм,\ s = 16мм.$

Необходимо найти вероятность (надежность) того, что $(0{,}9\overline{x};1{,}1\overline{x})$ - является доверительным интервалом оценки математического ожидания при нормальном распределении.

Из условия задачи найдем точность δ, составив и решив систему:

$$\begin{cases} 0{,}9\overline{x} = \overline{x} - \delta \\ 1{,}1\overline{x} = \overline{x} + \delta \end{cases}$$

Откуда δ =10. Из равенства $\delta = \frac{t_\gamma s}{\sqrt{n}}$ выразим $t_\gamma = \frac{\delta\sqrt{n}}{s}$, откуда t_γ =3,125.

По таблице для найденного t_γ и n= 25 находим γ =0,99.

3.1.3 Результаты и выводы: В результате проведенного занятия студенты:
- ознакомились с основными понятиями математической статистики, ее предметом и задачами;
- усвоили алгоритмы первичного статистического анализа экспериментальных данных;
- выработали навыки нахождения числовых характеристик статистического ряда, точечных и интервальных оценок параметров статистического распределения.

2.3. Практическое занятие 9, 10, 11, 12, 13 (ПЗ-9, 10, 11, 12, 13) (10 часов)
Тема: Оптимизационные задачи

2.3.1 Задание для работы:

1. Статистические критерии и их виды.
2. Критерии согласия.
3. Оценка параметров неизвестного распределения. Выравнивание рядов.

2.3.2 Краткое описание проводимого занятия:

1. Статистические критерии и их виды.

Пример(*) . По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива уменьшится. Для проверки проводятся испытания 25 случайно отобранных автомобилей с модернизированным двигателем, причем выборочное среднее расходов топлива на 100 км пробега по результатам испытаний составило $\bar{x}$ = 9,3 л. Предполагается, что выборка расходов топлива получена из нормально распределенной генеральной совокупности со средним m и дисперсией σ^2 = 4 л^2 . Используя критерий значимости, проверить гипотезу, утверждающую, что изменение конструкции двигателя не повлияло на расход топлива.
Решение. Проверяется гипотеза о среднем (m) нормально распределенной генеральной совокупности. Проверку гипотезы проведем по этапам:
1) проверяемая гипотеза H_0:m = 10, альтернативная гипотеза H_1: m < 10;
2) выберем уровень значимости α = 0,05;
3) в качестве статистики критерия используем оценку математического ожидания - выборочное среднее $\overline{X}$;
4) так как выборка получена из нормально распределенной генеральной совокупности, выборочное среднее также имеет нормальное распределение с дисперсией $\frac{\sigma^2}{n} = \frac{4}{25}$. При условии, что верна гипотеза H_0, математическое ожидание этого распределения равно 10. Нормированная статистика критерия $U = \frac{\overline{X} - 10}{\sqrt{4/25}}$ имеет нормальное распределение N(0,1).
5) альтернативная гипотеза H_1: m < 10 предполагает уменьшение расхода топлива, следовательно, нужно использовать односторонний критерий. Критическая область определяется неравенством $U < u_a$. По таблице приложений П1 находим $u_{0,05}=-u_{0,95}=-1{,}645$;
6) выборочное значение нормированной статистики критерия равно $U = \frac{9{,}3 - 10}{\sqrt{4/25}} = -1{,}75$.
7) статистическое решение: так как выборочное значение статистки критерия принадлежит критической области, гипотеза H_0 отклоняется: следует считать, что изменение конструкции
двигателя привело к уменьшению расхода топлива.
Граница $\overline{x_k}$ критической области для исходной статистики X критерия может быть получена из соотношения $\frac{\overline{x_k} - 10}{\sqrt{4/25}} = -1{,}75$, откуда получаем $\overline{x_k}$ = 9,342, т. е. критическая об-

ласть для статистики X определяется неравенством $\overline{X} < 9{,}342$.

2. Критерии согласия.

Пример (**). В первых - двух столбцах таблицы 1 приведены данные об отказах аппаратуры за 10000 часов работы. Общее число обследованных экземпляров аппаратуры n=757, при этом наблюдался отказ: 0·427+ 1·235+ 2·72 + 3·21+ 4·1+ 5·1 = 451

Таблица 1

Число отказов, k	Количество случаев, в которых наблюдалось k отказов, n_k	$p_k = \dfrac{0{,}6^k}{k!} e^{-0{,}6}$	Ожидаемое число случаев с k отказами, np_k
0	427	0,54881	416
1	235	0,32929	249
2	72	0,09879	75
3	21	0,01976	15
4	1	0,00296	2
5	1	0,00036	0
≥6	0	0,00004	0
Сумма	757	-	-

Проверить гипотезу о том, что число отказов имеет распределение Пуассона:

$p_k = P[X = k] = \dfrac{\lambda^k}{k!} e^{-\lambda}$, k=0, 1, …, при α=0,01.

Решение. Оценка параметра λ равна среднему числу отказов: $\overline{\lambda} = 451/757 \approx 0{,}6$. По таблице приложений (П3) с λ = 0,6 находим вероятности p_k и ожидаемое число случаев с k отказами (третий и четвертый столбцы таблицы 2).
Для k = 4,5 и 6 значения np_k < 5, поэтому объединяем эти строки со строкой для k= 3 .
Итак, получаем значения, приведенные в таблице 1.

Таблица 2

k	n_k	np_k	$\dfrac{(n_k - np_k)^2}{np_k}$
0	*427*	416	0,291
1	235	249	0,787
2	72	75	0,120
≥3	23	17	2,118
-	-	-	$\chi^2_B = 3{,}316$

Так как по выборке оценивался один параметр λ, то l = 1, число степеней свободы равно 4-1-1 = 2. По таблице приложений находим $\chi^2_{0,99}(2)$ = 9,21, гипотеза о распределении числа отказов по закону Пуассона принимается.

Пример . Комплектующие изделия одного наименования поступают с трех предприятий A, B и С. Результаты проверки изделий следующие:

Результаты проверки	Поставщики			Всего
	A	B	C	
Годные	29	38	53	120
Негодные	1	2	7	10

Всего	30	40	60	130

Можно ли считать, что качество изделий не зависит от поставщика? Принять $\alpha = 0{,}10$.
Решение. Проверим гипотезу о независимости двух признаков: качества изделия и места его изготовления. Находим

$$\chi_B^2 = 130 \cdot \left(\frac{29}{30 \cdot 120} + \frac{38^2}{40 \cdot 120} + \frac{53^2}{60 \cdot 120} + \frac{1^2}{30 \cdot 10} + \frac{2^2}{40 \cdot 10} + \frac{7^2}{60 \cdot 10} - 1 \right) \approx 2{,}546\,;$$

число степеней свободы (2 —1)(3 — 4) = 2. Так как по таблице приложений (П5) $\chi^2_{0,90}(2) = 4{,}605$, то это означает, что качество изделий не зависит от поставщика.
Заметим, что утверждение о том, что качество изделий не зависит от поставщика, можно трактовать как проверку гипотезы об однородности трех выборок изделий объемом 30, 40 и 60, полученных соответственно от поставщиков *А, В и С*.

Методика вычисления теоретических частот нормального распределения

Сущность критерия Пирсона состоит в сравнении эмпирических и теоретических частот. Эмпирические частоты находят из опыта. Как найти теоретические частоты, если предполагается, что генеральная совокупность распределена нормально?
Укажем один из способов решения этой задачи.
1. Весь интервал наблюдаемых значений X делят на s частичных интервалов $(x_i\,;\, x_{i+1})$ одинаковой длины. Находят середины частичных интервалов $x_i^* = \dfrac{x_i + x_{i+1}}{2}$; в качестве частоты n_i варианты x_i^* принимают число вариант, которые попали в i-й интервал. В итоге получают последовательность равноотстоящих вариант и соответствующих им частот:

$$x_1^*, x_2^*, \ldots, x_s^*,\ n_1, n_2, \ldots, n_s,$$

причем $\sum n_i = n$.

2. Вычисляют, например, методом произведений или сумм, выборочную среднюю $\overline{x}^*$ и выборочное среднее квадратическое отклонение σ^*.

3. Нормируют случайную величину X, т.е. переходят к величине: $z = \dfrac{x - \overline{x}^*}{\sigma^*}$, и вычисляют концы интервалов $(z_i\,;\, z_{i+1})$:

$$z_i = \frac{x_i - \overline{x}^*}{\sigma^*},\quad z_{i+1} = \frac{x_{i+1} - \overline{x}^*}{\sigma^*},$$

причем наименьшее значение, т.е. z_1 полагают равным $-\infty$, а наибольшее, т.е. z_s полагают равным ∞.
4. Вычисляют теоретические вероятности p_i попадания x в интервалы $(x_i\,;\, x_{i+1})$ по равенству (*Ф(z)* функция Лапласа)

$$p_i = \Phi(z_{i+1}) - \Phi(z_i)$$

и наконец находят искомые теоретические частоты $n_i' = np_i$.
Задача. Используя критерий Пирсона, при уровне значимости $a = 0{,}05$ 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 $n = 200$, $h = 2$, $\sigma_B = 4{,}695$, 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 (значения функции $j(x)$ приведены в приложении 1).

i	x_i	$u_i = \frac{(x_i - \bar{x}_B)}{\sigma_B}$	$\varphi(u_i)$	$n_i = \frac{n \cdot h}{\sigma_B} \cdot \varphi(u_i)$
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.

3. Оценка параметров неизвестного распределения. Выравнивание рядов.

Рассмотрим задачу «выравнивания» статистического распределения. Порядок решения этой задачи может быть следующим.

1. На основании статистических данных, оформленных в виде интервальной таблицы частот p^*, строят полигон или гистограмму и по внешнему виду этих графиков выдвигают гипотезу (делают предположение) о возможном теоретическом законе распределения случайной величины (кривой распределения).

Замечание. В некоторых случаях вид теоретической кривой распределения выбирается заранее из соображений, связанных с существом задачи.

2. Выясняют, от каких параметров зависит аналитическое выражение выбранной кривой распределения, и находят статистические оценки этих параметров. В этом случае задача выравнивания статистического распределения переходит в задачу рационального выбора тех значений параметров, при которых соответствие между статистическим и теоретическим распределениями оказывается наилучшим.

Например, если выдвигается гипотеза о нормальном законе распределения $X \sim N(a; \sigma)$, то он зависит только от двух параметров: математического ожидания a и среднего квадратического отклонения σ. Их наилучшими статистическими оценками будут соответственно среднее выборочное $\bar{x}$ и выборочное среднее квадратическое отклонение $\tilde{\sigma}$, т. е.

$$a \approx \bar{x}, \ \sigma \approx \tilde{\sigma}.$$

3. С учетом выдвинутой гипотезы о законе распределения случайной величины находят вероятности p_i попадания случайной величины в каждый из интервалов, указанных в статистической таблице распределения; записывают их в третьей строке таблицы и сравнивают полученные значения вероятностей p_i с соответствующими заданными частотами p_i^* (для наглядности можно изобразить графически). Проводя такое сравнение, делается приблизительная оценка степени согласования статистического и теоретического распределений. На этом первый этап решения задачи по определению закона распределения случайной величины заканчивается.

Пример. Для разумного планирования и организации работы ремонтных мастерских специальной техники оказалось необходимым изучить длительность ремонтных операций, производимых мастерскими.
Результаты (сгруппированные по интервалам) соответствующего статистического обследования (фиксированы длительности операций в
100 случаях) представлены в таблице:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
n_i	36	24	16	10	7	4	3

Требуется выровнять это статистическое распределение с помощью показательного закона $f(t)=\lambda e^{-\lambda t}$ (при $t \geq 0$), где λ – длительность операции в единицу времени.

Решение

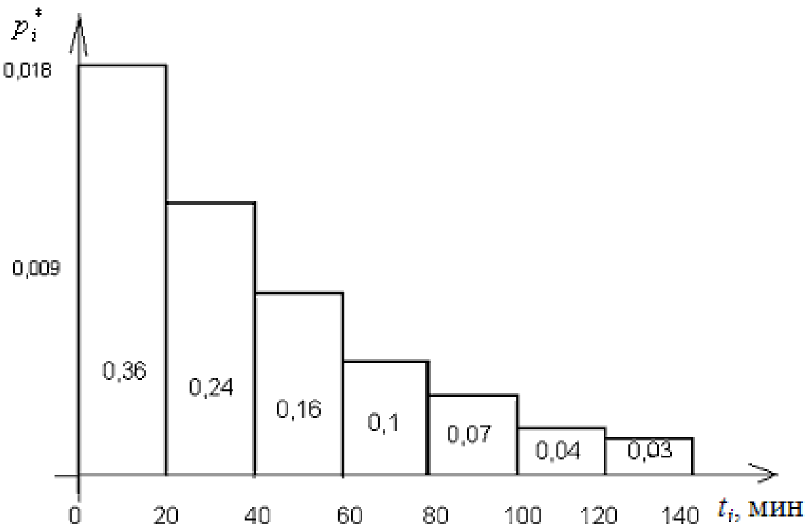
1. По данной таблице абсолютных частот построим таблицу относительных частот и соответствующую ей гистограмму:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
p_i*	0,36	0,24	0,16	0,10	0,07	0,04	0,03

.

$$\sum p_i^* = 0,36+0,24+0,16+0,1+0,07+0,04+0,03$$

Гистограмма относительных частот имеет вид:



Высоты прямоугольников гистограммы равны:

$$\Delta_1 = \frac{0,36}{20} = 0,018;$$

$$\Delta_2 = \frac{0,24}{20} = 0,012;$$

$$\Delta_3 = \frac{0,16}{20} = 0,008;$$

$$\Delta_4 = \frac{0,10}{20} = 0,005;$$

$$\Delta_5 = \frac{0,07}{20} = 0,0035; \quad \Delta_6 = \frac{0,04}{20} = 0,002; \quad \Delta_7 = \frac{0,03}{20} = 0,0015.$$

2. По внешнему виду гистограммы выдвигаем гипотезу, что случайная величина T (время ремонта) подчиняется показательному закону

$$f(t)=\lambda e^{-\lambda t},$$

который зависит только от одного параметра λ (длительность операции в единицу времени).

Параметр $\lambda = \frac{1}{m_i}$, где m_i – математическое ожидание (среднее время ремонта) случайной величины T.

Следовательно, для выравнивания статистического распределения с помощью кривой показательного распределения найдем статистическую оценку параметра m_i:

$$m_i \approx \bar{t} = 10\cdot 0,36+30\cdot 0,24+50\cdot 0,16+70\cdot 0,1+90\cdot 0,07+110\cdot 0,04+130\cdot 0,03=40$$

(числа 10, 30, 50, 70, 90, 110, 130 – это середины интервалов).

Тогда параметр $\lambda = \frac{1}{40}$.

3. Запишем теоретический закон распределения в виде функции плотности вероятности с учетом значения $\lambda = \frac{1}{40}$:

$$f(t) = \frac{1}{40} e^{-\frac{1}{40}t}.$$

По формуле вероятности попадания случайной величины (распределенной по показательному закону) на заданный интервал (α, β)

$$P(\alpha < X < \beta) = e^{-\lambda\alpha} - e^{-\lambda\beta}$$

найдем теоретические вероятности p_i, попадания случайной величины Т в каждый из семи интервалов и сравним их с соответствующими статистическими частотами p_i^*:

$$p_1 = P(0 < T < 20) = e^{-\frac{0}{40}} - e^{-\frac{20}{40}} = e^0 - e^{-0,5} \approx 1 - 0,6 = 0,4;$$

$$p_2 = P(20 < T < 40) = e^{-\frac{20}{40}} - e^{-\frac{40}{40}} = e^{-0,5} - e^{-1} \approx 0,6 - 0,37 = 0,23;$$

$$p_3 = P(40 < T < 60) = e^{-\frac{40}{40}} - e^{-\frac{60}{40}} = e^{-1} - e^{-1,5} \approx 0,37 - 0,22 = 0,15;$$

$$p_5 = P(60 < T < 80) = e^{-\frac{60}{40}} - e^{-\frac{80}{40}} = e^{-1,5} - e^{-2} \approx 0,22 - 0,14 = 0,08;$$

$$p_5 = P(80 < T < 100) = e^{-\frac{80}{40}} - e^{-\frac{100}{40}} = e^{-2} - e^{-2,5} \approx 0,14 - 0,08 = 0,06;$$

$$p_5 = P(100 < T < 120) = e^{-\frac{100}{40}} - e^{-\frac{120}{40}} = e^{-2,5} - e^{-3} \approx 0,08 - 0,05 = 0,03;$$

$$p_7 = P(120 < T < 140) = e^{-\frac{120}{40}} - e^{-\frac{140}{40}} = e^{-3} - e^{-3,5} \approx 0,05 - 0,03 = 0,02.$$

Для удобства сравнения теоретических вероятностей p_i с частотами p_i^* запишем полученные вероятности p_i в третью строку таблицы:

l_i	0–20	20–40	40–60	60–80	80–100	100–120	120–140
p_i^*	0,36	0,24	0,16	0,10	0,07	0,04	0,03
p_i	0,40	0,23	0,15	0,08	0,06	0,03	0,02

Замечаем, что расхождение между опытными частотами p_i^* и теоретическими вероятностями p_i незначительны. Следовательно, вполне допустима гипотеза о показательном законе распределения изучаемой случайной величины Т.

4. Построим на одном графике с гистограммой выравнивающую ее кривую распределения f(t). Для этого вычислим значения

$$f(t) = \frac{1}{40} e^{-\frac{1}{40}t}$$

например, на правых концах интервалов:

$$f(20) = \frac{1}{40} e^{-\frac{1}{40}\cdot 20} = \frac{1}{40} e^{-0,5} \approx 0,015; \quad f(40) = \frac{1}{40} e^{-\frac{1}{40}\cdot 40} = \frac{1}{40} e^{-1} \approx 0,009;$$

$$f(60)=\frac{1}{40}e^{\frac{1}{40}\cdot 60}=\frac{1}{40}e^{-1,5}\approx 0,006;\quad f(80)=\frac{1}{40}e^{-\frac{1}{40}\cdot 80}=\frac{1}{40}e^{-2}\approx 0,004;$$

$$f(100)=\frac{1}{40}e^{-\frac{1}{40}\cdot 100}=\frac{1}{40}e^{-2,5}\approx 0,002;\quad f(120)=\frac{1}{40}e^{-\frac{1}{40}\cdot 120}=\frac{1}{40}e^{-3}\approx 0,0013;$$

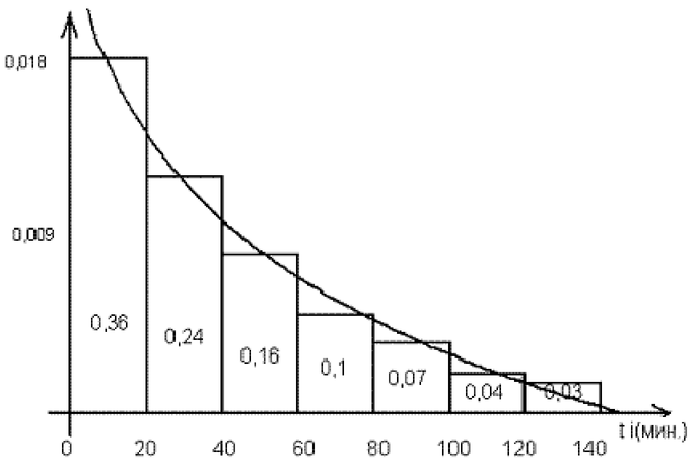
$$f(140)=\frac{1}{40}e^{-\frac{1}{40}\cdot 140}=\frac{1}{40}e^{-3,5}\approx 0,0008.$$

Построим график полученной кривой распределения f(t), в той же системе координат, что и гистограмма относительных частот.

Из рисунка видно, что теоретическая кривая f(t) сохраняет в основном существенные особенности статистического распределения.

3.3.3 Результаты и выводы: В результате проведенного занятия студенты:

- ознакомились с понятиями статистического критерия, его мощности, классификацией статистических критериев;
- усвоили алгоритмы применения различных статистических критериев;
- выработали навыки нахождения оценок параметров неизвестного распределения, выравнивания рядов.



2.4. Практическое занятие 14,15,16 (ПЗ-14,15,16) (6 часов)
Тема: Марковские процессы. Системы массового обслуживания

2.4.1 Задание для работы:

1. Многомерные СВ, законы их распределения, условные числовые характеристики
2. Функция регрессии, коэффициент детерминации, корреляции, ковариация
3. Виды регрессий, статистическая значимость их параметров.
1. Графический метод решения ЗЛП.
2. Теоретические основы симплекс- метода
3. Симплекс метод решения ЗЛП
1. Простейший поток, его свойства. Классификация потоков.
2. Марковские цепи, их свойства
3. Культура в инженерной практике

2.4.2 Краткое описание проводимого занятия:

1. Многомерные СВ, законы их распределения, условные числовые характеристики

Пример (*). Найти выборочные средние, дисперсии и коэффициент корреляции для выборки, приведенной в таблицей. Построить диаграмму рассеивания.

Решение. Вычисление указанных выборочных характеристик удобно выполнять в следующей последовательности. Сначала вычисляют суммы $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum y_i^2$, $\sum x_i y_i$, $\sum (x_i + y_i)^2$.Для контроля правильности вычислений используется тождество

$$\sum (x_i + y_i)^2 = \sum x_i^{\,2} + 2\sum x_i y_i + \sum y_i^2 .$$

Таблица 1

x	*y*	*x*	*y*	*x*	*y*	*x*	*y*	*x*	*y*
8,35	3,50	10,50	6,00	11,35	9,50	12,15	6,00	12,85	9,50
8,74	1,49	10,75	2,50	11,50	6,00	12,25	8,05	13,15	9,02
9,25	6,40	10,76	5,74	11,50	9,00	12,35	5,01	13,25	6,49
9,50	4,50	11,00	8,50	11,62	8,50	12,50	7,03	13,26	10,50
9,75	5,00	11,00	5,26	11,75	10,00	12,76	7,53	13,40	7,51
10,24	7,00	11,25	8,00	12,00	9,00	12,85	6,01	13,50	10,00
13,65	9,50	14,50	10,00	13,75	8,51	14,75	12,00	14,00	11,00
15,25	12,50	14,23	8,40	16,00	11,50	14,26	10,00	16,00	13,00
14,51	9,50	16,25	12,00						

Объем выборки $n = 42$.

Выборочные средние отсюда находятся по формулам

$$\bar{x} = \alpha_{1,0}^{*} = \frac{1}{n}\sum x_i \,,\ \bar{y} = \alpha_{0,1}^{*} = \frac{1}{n}\sum y_i$$

Затем вычисляются суммы квадратов отклонений от среднего и произведений отклонений от средних: $Q_x = \sum \left(x_i - \bar{x}\right)^2 = \sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}$, $Q_y = \sum \left(y_i - \bar{y}\right)^2 = \sum y_i^2 - \dfrac{\left(\sum y_i\right)^2}{n}$,

$$Q_x = \sum \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \sum x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n}$$

Отсюда $D_X^* = \dfrac{1}{n} Q_x$, $D_Y^* = \dfrac{1}{n} Q_y$, $r = \dfrac{\mu_{1,1}^{*}}{\sqrt{D_X^* D_Y^*}} = \dfrac{Q_{xy}}{\sqrt{Q_x Q_y}}$.

Предварительно вычислим

$$\sum x_i = 522{,}23 \,,\ \sum y_i = 336{,}41 \,,\ \sum x_i^2 = 6652{,}25 \,,\ \sum y_i^2 = 2987{,}80 \,,\ \sum x_i y_i = 4358{,}626 .$$

Тогда найдем $\bar{x} = 12{,}434$, $\bar{y} = 8{,}011$.

Далее находим $Q_x = 6652{,}25 - \dfrac{522{,}23^2}{42} \approx 158{,}8182$, $Q_y = 2987{,}805 - \dfrac{336{,}41^2}{42} \approx 292{,}5958$,

$$Q_{xy} = 4358{,}626 - \frac{522{,}23 \cdot 336{,}41}{42} \approx 175{,}1912 .$$

Окончательно, получаем

$$D_X^* = \frac{158{,}8182}{42} \approx 3{,}7814 \,,\ D_Y^* = \frac{292{,}5958}{42} \approx 6{,}9666 \,,\ r = \frac{175{,}1912}{\sqrt{158{,}8182 \cdot 292{,}5958}} \approx 0{,}813 .$$

Диаграмма рассеивания приведена на рис. 1.

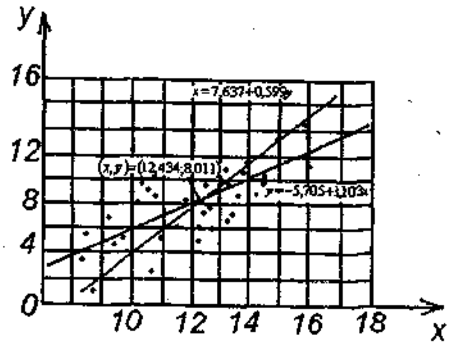


Рис. 1.

Выборочная линейная регрессия Y на X по выборке (x_i,y_i), i=1,2,...,n, определяется уравнением $y = \beta_0^* + \beta_1^* x = \overline{y} + r\frac{D_Y^*}{D_X^*}(x - \overline{x})$.

Коэффициенты β_0^* и β_1^* называются *выборочными коэффициентами регрессии.* Они вычисляются по формулам: $\beta_1 = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{Q_{xy}}{Q_x}$, $\beta_0 = \overline{y} - \beta_1 \overline{x}$.

Аналогично определяется выборочная линейная регрессия X на Y:

$x = \beta_0^{'*} + \beta_1^{'*} y = \overline{x} + r\frac{D_X^*}{D_Y^*}(y - \overline{y})$ коэффициенты $\beta_0^{'*}$ и $\beta_1^{'*}$ которой находятся по формулам

$$\beta_1^{'*} = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{n\sum y_i^2 - \left(\sum y_i\right)^2} = \frac{Q_{xy}}{Q_y},\quad \beta_0^{'*} = \overline{x} - \beta_1^{'*}\overline{y}.$$

Для контроля правильности расчетов используют соотношение $\sqrt{\beta_1^* \beta_1^{'*}} = |r|$.

Прямые $y = \beta_0^* + \beta_1^* x$, $x = \beta_0^{'*} + \beta_1^{'*} y$ пересекаются в точке с координатами $\left(\overline{x}, \overline{y}\right)$.

Пример. Вычислить выборочные коэффициенты линейной регрессии X на Y и Y на X по выборке из предыдущего примера . Нанести прямые регрессии на диаграмму рассеивания.
Решение. Воспользуемся результатами вычислений в предыдущем примере . По формулам находим

$$\beta_1^* = \frac{175{,}1912}{158{,}8182} \approx 1{,}103,\quad \beta_0^* = 8{,}011 - 1{,}103 \cdot 12{,}434 \approx -5{,}705.$$

Таким образом, прямая регрессии Y на X имеет уравнение y=-5,705+1,103x.

Аналогично находим $\beta_1^{'*} = \frac{175{,}1912}{292{,}5958} \approx 0{,}599$, $\beta_0^{'*} = 12{,}434 - 0{,}599 \cdot 8{,}011 \approx 7{,}637$.

Отсюда прямая регрессии X на Y имеет уравнение $x = 7{,}637 + 0{,}599y$.

Проверка показывает $\sqrt{1{,}103 \cdot 0{,}599} \approx 0{,}813$, что полученный результат совпадает со значением r, вычисленным в примере (*). Прямые регрессии нанесены на диаграмму рассеивания на рис.1.

Пример. Используя группировку выборки, заданной таблицей в примере (*), вычислить выборочные средние, дисперсии, коэффициент корреляции, а также выборочные коэффициенты линейной регрессии X на Y и Y на X.
Решение. Выберем b_x=1, b_y=2. Прямоугольная сетка, соответствующая этим значениям, нанесена на диаграмму рассеивания (рис. 1). Непосредственно по диаграмме строим корреляционную таблицу (таблица 2). Находим $d_X^* = 11{,}5$, $d_Y^* = 9$ и вычисляем значения u_i и v_j

по формулам $u_i = \frac{\hat{x}_i - 11{,}5}{1}$, i=1,2,...,9, $v_j = \frac{\hat{y}_j - 9}{2}$, j=1,2,...,7.

Вычисляем следующие суммы:

$\sum n_i u_i = 43$, $\sum n_j v_j = -15$, $\sum n_i u_i^2 = 215$, $\sum n_j v_j^2 = 87$, $\sum\limits_{i=1}^{9} \sum\limits_{j=1}^{7} n_{ij} u_i v_j = 80$. По формулам находим $Q_u = 215 - \dfrac{43^2}{42} \approx 170{,}976$, $Q_v = 85 - \dfrac{(-15)^2}{42} \approx 81{,}643$, $Q_{uv} = 80 - \dfrac{43 \cdot (-15)}{42} \approx 95{,}357$.

Далее получаем $\bar{x} = 1\dfrac{43}{42} + 11{,}5 \approx 12{,}52$, $\bar{y} = 2\dfrac{(-15)}{42} + 9 \approx 8{,}28$, $D_X^* = 1^2\dfrac{170{,}976}{42} \approx 4{,}071$, $D_Y^* = 2^2\dfrac{81{,}643}{42} \approx 7{,}775$, $r = \dfrac{95{,}357}{\sqrt{170{,}976 \cdot 81{,}643}} \approx 0{,}807$. Находим выборочные коэффициенты регрессии: $\beta_1^* = \dfrac{2}{1}\dfrac{95{,}357}{170{,}976} \approx 1{,}12$, $\beta_1^{'*} = \dfrac{1}{2}\dfrac{95{,}357}{81{,}643} \approx 0{,}58$, $\beta_0^* = 8{,}28 - 1{,}12 \cdot 12{,}52 \approx -5{,}74$, $\beta_0^{'*} = 12{,}52 - 0{,}58 \cdot 8{,}28 \approx 7{,}72$.

Окончательно получим, что уравнение линейной регрессии Y на X имеет вид $y = -5{,}74 + 1{,}12x$, а уравнение линейной регрессии X на Y имеет вид $x = 7{,}72 + 0{,}58y$.

Расхождение полученных результатов с результатами выше рассмотренных примеров обусловлено группировкой.

Таблица 2.Корреляционная таблица для диаграммы рассеивания

Границы и средины интервалов для y	v_j	Границы и середины интервалов для x									n_j	$n_j v_j$	$n_j v_j^2$
		8-9 8,5	9-10 9,5	10-11 10,5	11-12 11,5	12-13 12,5	13-14 13,5	14-15 14,5	15-16 15,5	16-17 16,5			
		u_i											
		-3	-2	-1	0	1	2	3	4	5			
0-2 1	-4	1	0	0	0	0	0	0	0	0	1	-4	16
2-4 3	-3	1	0	1	0	0	0	0	0	0	2	-6	18
4-6 5	-2	0	2	1	1	1	0	0	0	0	5	-10	20
6-8 7	-1	0	1	2	1	4	2	0	0	0	10	-10	10
8-10 9	0	0	0	0	5	3	3	2	0	0	13	0	0
10-12 11	1	0	0	0	1	0	2	3	0	1	7	7	7
12-14 13	2	0	0	0	0	0	0	1	1	2	4	8	16
n_i		2	3	4	8	8	7	6	1	3	∑=42	∑=-15	∑=87
$n_i v_i$		-6	-6	-4	0	8	14	18	4	15	∑=43		
$n_i v_i^2$		18	12	4	0	8	28	54	16	75	∑=215		

2. Функция регрессии, коэффициент детерминации, корреляции, ковариация

Допустим, что в результате лечения 12 больных с артериальной гипертензией в результате суточного мониторирования систолического артериального давления (САД) до лечения и после месячного лечения были получены следующие результаты:

№	САД до (x_i)	САД после (y_i)	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$
			-9,6	-7,5	
			-19,6	-17,5	
			-14,6	-12,5	182,5
			-4,6	7,5	-34,5
			0,4	12,5	
			5,4	12,5	67,5
			-9,6	-7,5	
			10,4	7,5	
			15,4	12,5	192,5
			0,4	-7,5	-3
			5,4	-2,5	-13,5
			20,4	2,5	
	$\bar{x}=169{,}6$	$\bar{y}=137{,}5$			$\sum=1012{,}5$
	$\sigma_x=12{,}1$	$\sigma_y=10{,}6$			

$$r_{x,y}=\frac{\frac{1}{12-1}\cdot 1012{,}5}{12{,}1\cdot 10{,}6}=0{,}718$$

Итак, коэффициент корреляции получился равным 0,718.
Определим, достоверно ли он отличается от нуля. Для этого используем Таблицу 10 приложения. У нас 12 пар измерений, поэтому входим в Таблицу по 12 строке. На пересечении 12 строки и столбца P=0,05 стоит число 0,576. Полученный коэффициент корреляции (0,718) больше этого числа. Следовательно, на этом уровне коэффициент корреляции достоверно отличается от нуля, то есть связь есть. На пересечении этой же строки и столбца P=0,01 стоит число 0,708. Поскольку коэффициент корреляции больше и этого числа, следовательно, мы можем говорить, что связь существует и на этом более значимом уровне.
Итак, ответ на первый вопрос таков: существование связи высоко достоверно. Далее, поскольку получено положительное значение коэффициента корреляции, мы заключаем, что связь прямая. Используя Таблицу 2 данного раздела, мы приходим к заключению, что связь сильная.

Найдем коэффициент детерминации:

$$d=r^2\times 100=0{,}718^2\times 100=0{,}516\times 100=51{,}6\ (\%)$$

Таким образом, систолическое артериальное давление после лечения на 51,6 % определяется систолическим артериальным давлением до лечения, а на 48,4 % другими факторами.

Формы проявления взаимосвязей явлений и процессов весьма разнообразны. Из них в самом общем виде выделяют *функциональную* (полную) и *корреляционную* (неполную) связи.

Математически ковариация представляет собой меру линейной зависимости двух случайных величин.

Коэффициент корреляции - это математическая мера корреляции двух величин. Коэффициенты корреляции могут быть положительными и отрицательными. Иногда показателям тесноты связи можно дать качественную оценку (шкала Чеддока):

Количественная мера тесноты связи	**Качественная характеристика силы связи**
0,1 - 0,3	Слабая
0,3 - 0,5	Умеренная
0,5 - 0,7	Заметная
0,7 - 0,9	Высокая
0,9 - 0,99	Весьма высокая

3. Виды регрессий, статистическая значимость их параметров.

Пример. Используя группировку выборки, заданной таблицей в примере (*), вычислить выборочные средние, дисперсии, коэффициент корреляции, а также выборочные коэффициенты линейной регрессии X на Y и Y на X.

Решение. Выберем b_x=1, b_y=2. Прямоугольная сетка, соответствующая этим значениям, нанесена на диаграмму рассеивания (рис. 1). Непосредственно по диаграмме строим корреляционную таблицу (таблица 2). Находим $d_X^* = 11{,}5$, $d_Y^* = 9$ и вычисляем значения u_i и v_j

по формулам $u_i = \dfrac{\hat{x}_i - 11{,}5}{1}$, i=1,2,...,9, $\nu_j = \dfrac{\hat{y}_j - 9}{2}$, j=1,2,...,7.

Вычисляем следующие суммы:

$\sum n_i u_i = 43$, $\sum n_j \nu_j = -15$, $\sum n_i u_i^2 = 215$, $\sum n_j \nu_j^2 = 87$, $\sum_{i=1}^{9} \sum_{j=1}^{7} n_{ij} u_i \nu_j = 80$. По формулам

находим $Q_u = 215 - \dfrac{43^2}{42} \approx 170{,}976$, $Q_v = 85 - \dfrac{(-15)^2}{42} \approx 81{,}643$, $Q_{uv} = 80 - \dfrac{43 \cdot (-15)}{42} \approx 95{,}357$.

Далее получаем $\bar{x} = 1\dfrac{43}{42} + 11{,}5 \approx 12{,}52$, $\bar{y} = 2\dfrac{(-15)}{42} + 9 \approx 8{,}28$, $D_X^* = 1^2 \dfrac{170{,}976}{42} \approx 4{,}071$,

$D_Y^* = 2^2 \dfrac{81{,}643}{42} \approx 7{,}775$, $r = \dfrac{95{,}357}{\sqrt{170{,}976 \cdot 81{,}643}} \approx 0{,}807$. Находим выборочные коэффициенты

регрессии: $\beta_1^* = \dfrac{2}{1}\dfrac{95{,}357}{170{,}976} \approx 1{,}12$, $\beta_1^{'*} = \dfrac{1}{2}\dfrac{95{,}357}{81{,}643} \approx 0{,}58$, $\beta_0^* = 8{,}28 - 1{,}12 \cdot 12{,}52 \approx -5{,}74$,

$\beta_0^{'*} = 12{,}52 - 0{,}58 \cdot 8{,}28 \approx 7{,}72$.

Окончательно получим, что уравнение линейной регрессии Y на X имеет вид y = -5,74 + 1,12x, а уравнение линейной регрессии X на Y имеет вид x = 7,72 + 0,58y.

Расхождение полученных результатов с результатами выше рассмотренных примеров обусловлено группировкой.

Таблица 3. Корреляционная таблица для диаграммы рассеивания.

Границы и средины интервалов для y	v_j	Границы и середины интервалов для x									n_j	$n_j v_j$	$n_j v_j^2$
		8-9 8,5	9-10 9,5	10-11 10,5	11-12 11,5	12-13 12,5	13-14 13,5	14-15 14,5	15-16 15,5	16-17 16,5			
		u_i											
		-3	-2	-1	0	1	2	3	4	5			
0-2 1	-4	1	0	0	0	0	0	0	0	0	1	-4	16

2-4 3	-3	1	0	1	0	0	0	0	0	0	2	-6	18
4-6 5	-2	0	2	1	1	1	0	0	0	0	5	-10	20
6-8 7	-1	0	1	2	1	4	2	0	0	0	10	-10	10
8-10 9	0	0	0	0	5	3	3	2	0	0	13	0	0
10-12 11	1	0	0	0	1	0	2	3	0	1	7	7	7
12-14 13	2	0	0	0	0	0	0	1	1	2	4	8	16
n_i		2	3	4	8	8	7	6	1	3	∑=42	∑=-15	∑=87
$n_i v_i$		-6	-6	-4	0	8	14	18	4	15	∑=43		
$n_i v_i^2$		18	12	4	0	8	28	54	16	75	∑=215		

Задача:
Имеется связанная выборка из 26 пар значений (x_k,y_k):

k	1	2	3	4	5	6	7	8	9	10
x_k	25.20000	26.40000	26.00000	25.80000	24.90000	25.70000	25.70000	25.70000	26.10000	25.80000
y_k	30.80000	29.40000	30.20000	30.50000	31.40000	30.30000	30.40000	30.50000	29.90000	30.40000

k	11	12	13	14	15	16	17	18	19	20
x_k	25.90000	26.20000	25.60000	25.40000	26.60000	26.20000	26.00000	22.10000	25.90000	25.80000
y_k	30.30000	30.50000	30.60000	31.00000	29.60000	30.40000	30.70000	31.60000	30.50000	30.60000

k	21	22	23	24	25	26
x_k	25.90000	26.30000	26.10000	26.00000	26.40000	25.80000
y_k	30.70000	30.10000	30.60000	30.50000	30.70000	30.80000

Требуется вычислить/построить:

- коэффициент корреляции;
- проверить гипотезу зависимости случайных величин X и Y, при уровне значимости α = 0.05;
- коэффициенты уравнения линейной регрессии;
- диаграмму рассеяния (корреляционное поле) и график линии регрессии;

РЕШЕНИЕ:

1. Вычисляем коэффициент корреляции.

Коэффициент корреляции — это показатель взаимного вероятностного влияния двух случайных величин. Коэффициент корреляции **R** может принимать значения от **-1** до **+1**. Если абсолютное значение находится ближе к **1**, то это свидетельство сильной связи между величинами, а если ближе к **0** — то, это говорит о слабой связи или ее отсутствии. Если абсолютное значение **R** равно единице, то можно говорить о функциональной связи между величинами, то есть одну величину можно выразить через другую посредством математической функции.

Вычислить коэффициент корреляции можно по следующим формулам:

$$R_{x,y} = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y} \quad (1.1),$$ где:

cov(X,Y) - ковариация случайных величин X и Y

$$\sigma_x^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - M_x)^2, \qquad \sigma_y^2 = \frac{1}{n}\sum_{k=1}^{n}(y_k - M_y)^2 \quad (1.2),$$ - оценки дисперсий случайных величин X и Y соответственно.

$$M_x = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad M_y = \frac{1}{n}\sum_{k=1}^{n} y_k \quad (1.3),$$ - оценки математического ожидания случайных величин X и Y соответственно.

или по формуле

$$R_{x,y} = \frac{M_{xy} - M_x M_y}{S_x S_y} \quad (1.4),$$ где:

$$M_x = \frac{1}{n}\sum_{k=1}^{n} x_k, \quad M_y = \frac{1}{n}\sum_{k=1}^{n} y_k, \quad M_{xy} = \frac{1}{n}\sum_{k=1}^{n} x_k y_k \quad (1.5)$$

$$S_x^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - M_x^2, \quad S_y^2 = \frac{1}{n}\sum_{k=1}^{n} y_k^2 - M_y^2 \quad (1.6)$$

На практике, для вычисления коэффициента корреляции чаще используется формула (1.4) т.к. она требует меньше вычислений. Однако если предварительно была вычислена ковариация **cov(X,Y)**, то выгоднее использовать формулу (1.1), т.к. кроме собственно значения ковариации можно воспользоваться и результатами промежуточных вычислений.

1.1 Вычислим коэффициент корреляции по формуле (1.4), для этого вычислим значения x_k^2, y_k^2 и $x_k y_k$ и занесем их в таблицу 1.

Таблица 1

k	x_k	y_k	x_k^2	y_k^2	$x_k y_k$
1	2	3	4	5	6
1	25.2	30.8	635.04000	948.64000	**776.16000**
2	26.4	29.4	696.96000	864.36000	**776.16000**
3	26.0	30.2	676.00000	912.04000	**785.20000**
4	25.8	30.5	665.64000	930.25000	**786.90000**

5	24.9	31.4	620.01000	985.96000	**781.86000**
6	25.7	30.3	660.49000	918.09000	**778.71000**
7	25.7	30.4	660.49000	924.16000	**781.28000**
8	25.7	30.5	660.49000	930.25000	**783.85000**
9	26.1	29.9	681.21000	894.01000	**780.39000**
10	25.8	30.4	665.64000	924.16000	**784.32000**
11	25.9	30.3	670.81000	918.09000	**784.77000**
12	26.2	30.5	686.44000	930.25000	**799.10000**
13	25.6	30.6	655.36000	936.36000	**783.36000**
14	25.4	31	645.16000	961.00000	**787.40000**
15	26.6	29.6	707.56000	876.16000	**787.36000**
16	26.2	30.4	686.44000	924.16000	**796.48000**
17	26	30.7	676.00000	942.49000	**798.20000**
18	22.1	31.6	488.41000	998.56000	**698.36000**
19	25.9	30.5	670.81000	930.25000	**789.95000**
20	25.8	30.6	665.64000	936.36000	**789.48000**
21	25.9	30.7	670.81000	942.49000	**795.13000**
22	26.3	30.1	691.69000	906.01000	**791.63000**
23	26.1	30.6	681.21000	936.36000	**798.66000**
24	26	30.5	676.00000	930.25000	**793.00000**
25	26.4	30.7	696.96000	942.49000	**810.48000**
26	25.8	30.8	665.64000	948.64000	**794.64000**

1.2. Вычислим M_x по формуле (1.5).
1.2.1. Сложим последовательно все элементы x_k
$x_1 + x_2 + \ldots + x_{26} =$ 25.20000 + 26.40000 + ... + 25.80000 = 669.500000
1.2.2. Разделим полученную сумму на число элементов
669.50000 / 26 = 25.75000 **M_x = 25.750000**
1.3. Аналогичным образом вычислим M_y.
1.3.1. Сложим последовательно все элементы y_k
$y_1 + y_2 + \ldots + y_{26} =$ 30.80000 + 29.40000 + ... + 30.80000 = 793.000000
1.3.2. Разделим полученную сумму на число элементов выборки
793.00000 / 26 = 30.50000 **M_y = 30.500000**
1.4. Аналогичным образом вычислим M_{xy}.
1.4.1. Сложим последовательно все элементы 6-го столбца таблицы 1
776.16000 + 776.16000 + ... + 794.64000 = 20412.830000
1.4.2. Разделим полученную сумму на число элементов
20412.83000 / 26 = 785.10885 **M_{xy} = 785.108846**
1.5. Вычислим значение S_x^2 по формуле (1.6.).
1.5.1. Сложим последовательно все элементы 4-го столбца таблицы 1
635.04000 + 696.96000 + ... + 665.64000 = 17256.910000
1.5.2. Разделим полученную сумму на число элементов
17256.91000 / 26 = 663.72731
1.5.3. Вычтем из последнего числа квадрат величины M_x получим значение для S_x^2
S_x^2 = 663.72731 - 25.75000^2 = 663.72731 - 663.06250 = **0.66481**

1.6. Вычислим значение S_y^2 по формуле (1.6.).
1.6.1. Сложим последовательно все элементы 5-го столбца таблицы 1
948.64000 + 864.36000 + ... + 948.64000 = 24191.840000
1.6.2. Разделим полученную сумму на число элементов
24191.84000 / 26 = 930.45538
1.6.3. Вычтем из последнего числа квадрат величины M_y получим значение для S_y^2
S_y^2 = 930.45538 - 30.50000^2 = 930.45538 - 930.25000 = **0.20538**
1.7. Вычислим произведение величин S_x^2 и S_y^2.
$S_x^2 S_y^2$ = 0.66481 • 0.20538 = 0.136541
1.8. Извлечем и последнего числа квадратный корень, получим значение $S_x S_y$.
$S_x S_y$ = 0.36951
1.9. Вычислим значение коэффициента корреляции по формуле (1.4.).
R = (785.10885 - 25.75000 • 30.50000) / 0.36951 = (785.10885 - 785.37500) / 0.36951 = -0.72028
ОТВЕТ: $R_{x,y}$ = -0.720279

2. Проверяем значимость коэффициента корреляции (проверяем гипотезу зависимости).
Поскольку оценка коэффициента корреляции вычислена на конечной выборке, и поэтому может отклоняться от своего генерального значения, необходимо проверить значимость коэффициента корреляции. Проверка производится с помощью t-критерия:

$$t = \frac{R_{x,y}\sqrt{n-2}}{\sqrt{1-R^2_{x,y}}} \qquad (2.1)$$

Случайная величина **t** следует t-распределению Стьюдента и по таблице t-распределения необходимо найти критическое значение критерия ($t_{кр.\alpha}$) при заданном уровне значимости α. Если вычисленное по формуле (2.1) t по модулю окажется меньше чем $t_{кр.\alpha}$, то зависимости между случайными величинами X и Y нет. В противном случае, экспериментальные данные не противоречат гипотезе о зависимости случайных величин.

2.1. Вычислим значение t-критерия по формуле (2.1) получим:

$$t = \frac{-0.72028\sqrt{26-2}}{\sqrt{1-(-0.72028)^2}} = -5.08680$$

2.2. Определим по таблице t-распределения критическое значение параметра $t_{кр.\alpha}$
Искомое значение $t_{кр.\alpha}$ располагается на пересечении строки соответствующей числу степеней свободы и столбца соответствующего заданному уровню значимости α.
В нашем случае число степеней свободы есть n - 2 = 26 - 2 = **24** и α = **0.05** , что соответствует критическому значению критерия $t_{кр.\alpha}$ = **2.064** (см. табл. 2)

Таблица 2 **t-распределение**

Число степеней свободы (n - 2)	**α = 0.1**	**α = 0.05**	**α = 0.02**	**α = 0.01**	**α = 0.002**	**α = 0.001**
1	6.314	12.706	31.821	63.657	318.31	636.62
2	2.920	4.303	6.965	9.925	22.327	31.598
3	2.353	3.182	4.541	5.841	10.214	12.924
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.869
6	1.943	2.447	3.143	3.707	5.208	5.959

7	1.895	2.365	2.998	3.499	4.785	5.408
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.250	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.552	2.878	3.610	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.767
24	1.711	**2.064**	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.750	3.385	3.646
40	1.684	2.021	2.423	2.704	3.307	3.551
60	1.671	2.000	2.390	2.660	3.232	3.460
120	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.645	1.960	2.326	2.576	3.090	3.291

2.2. Сравним абсолютное значение t-критерия и $t_{кр.\alpha}$
Абсолютное значение t-критерия не меньше критического t = 5.08680, $t_{кр.\alpha}$ = 2.064, следовательно **экспериментальные данные, с вероятностью 0.95** (1 - α), **не противоречат гипотезе** о зависимости случайных величин X и Y.

3. Вычисляем коэффициенты уравнения линейной регрессии.
Уравнение линейной регрессии представляет собой уравнение прямой, аппроксимирующей (приблизительно описывающей) зависимость между случайными величинами X и Y. Если считать, что величина X свободная, а Y зависимая от X, то уравнение регрессии запишется следующим образом $Y = a + b\bullet X$ (3.1), где:

$$b = R_{x,y}\frac{\sigma_y}{\sigma_x} = R_{x,y}\frac{S_y}{S_x} \qquad (3.2),$$

$$a = M_y - b\bullet M_x \qquad (3.3)$$

Рассчитанный по формуле (3.2) коэффициент **b** называют коэффициентом линейной регрессии. В некоторых источниках **a** называют постоянным коэффициентом регрессии и **b** соответственно переменным.

Погрешности предсказания Y по заданному значению X вычисляются по формулам :

$\sigma_{y/x} = \sigma_y \sqrt{1-R^2_{x,y}} = S_y \sqrt{1-R^2_{x,y}}$ (3.4) - абсолютная погрешность,

$\delta_{y/x} = \frac{\sigma_{y/x}}{M_y} 100\%$ (3.5) - относительная погрешность

Величину $\sigma_{y/x}$ (формула 3.4) еще называют **остаточным средним квадратическим отклонением**, оно характеризует уход величины Y от линии регрессии, описываемой уравнением (3.1), при фиксированном (заданном) значении X.

3.1. Вычислим отношение $\frac{S_y^2}{S_x^2}$.

$S_y^2 / S_x^2 = 0.20538 / 0.66481 = 0.30894$

3.2. Вычислим отношение $\frac{S_y}{S_x}$.

Извлечем из последнего числа квадратный корень - получим:

$S_y / S_x = 0.55582$

3.3 Вычислим коэффициент b по формуле (3.2)

b = -0.72028 • 0.55582 = **-0.40035**

3.4 Вычислим коэффициент a по формуле (3.3)

a = 30.50000 - (-0.40035 • 25.75000) = **40.80894**

3.5 Оценим погрешности уравнения регрессии.

3.5.1 Извлечем из S_y^2 квадратный корень получим:

$S_y = \sqrt{0.20538} = 0.45319$;

3.5.2 Возведем в квадрат $R_{x,y}$ получим: $R^2_{x,y} = -0.72028^2 = 0.51880$

3.5.3 Вычислим абсолютную погрешность (остаточное среднее квадратическое отклонение) по формуле (3.4)

$\sigma_{y/x} = 0.45319 \sqrt{1 - 0.51880} = 0.31437$

3.5.4 Вычислим относительную погрешность по формуле (3.5)

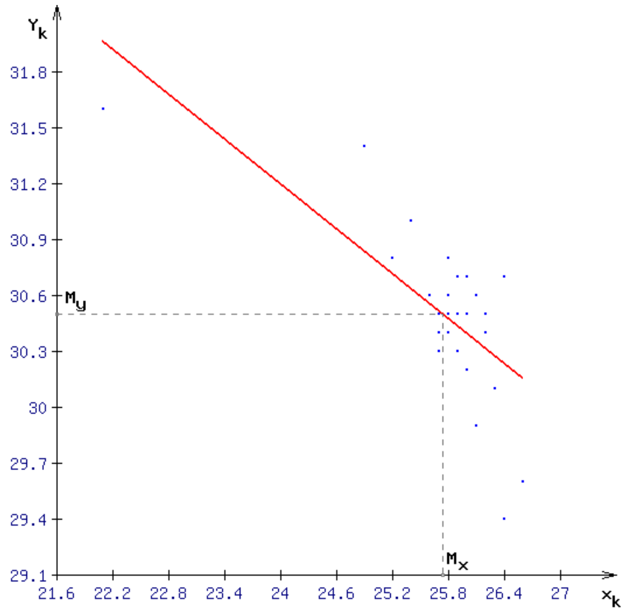
$\delta_{y/x} = (0.31437 / 30.50000)100\% = 1.03073\%$

ОТВЕТ: Уравнение линейной регрессии имеет вид: **Y = 40.80894 -0.40035 X** (3.6)

Погрешности уравнения: $\sigma_{y/x} = 0.31437$; $\delta_{y/x} = 1.03073$**%**

4. Строим диаграмму рассеяния (корреляционное поле) и график линии регрессии.

Диаграмма рассеяния — это графическое изображение соответствующих пар (x_k , y_k) в виде точек плоскости, в прямоугольных координатах с осями X и Y. Корреляционное поле является одним из графических представлений связанной (парной) выборки. В той же системе координат строится и график линии регрессии. Следует тщательно выбрать масштабы и начальные точки на осях, чтобы диаграмма была максимально наглядной.



4.1. Находим минимальный и максимальный элемент выборки X это 18-й и 15-й элементы соответственно, x_{min} = 22.10000 и x_{max} = 26.60000.

4.2. Находим минимальный и максимальный элемент выборки Y это 2-й и 18-й элементы соответственно, y_{min} = 29.40000 и y_{max} = 31.60000.

4.3. На оси абсцисс выбираем начальную точку чуть левее точки x_{18} = 22.10000, и такой масштаб, чтобы на оси поместилась точка x_{15} = 26.60000 и отчетливо различались остальные точки.

4.4. На оси ординат выбираем начальную точку чуть левее точки y_2 = 29.40000, и такой масштаб, чтобы на оси поместилась точка y_{18} = 31.60000 и отчетливо различались остальные точки.

4.5. На оси абсцисс размещаем значения x_k, а на оси ординат значения y_k.

4.6. Наносим точки (x_1, y_1), (x_2, y_2),…,(x_{26}, y_{26}) на координатную плоскость. Получаем диаграмму рассеяния (корреляционное поле), изображенное на рисунке ниже.

4.7. Начертим линию регрессии.

Для этого найдем две различные точки с координатами (x_{r1} , y_{r1}) и (x_{r2} , y_{r2}) удовлетворяющие уравнению (3.6), нанесем их на координатную плоскость и проведем через них прямую. В качестве абсциссы первой точки возьмем значение x_{min} = 22.10000. Подставим значение x_{min} в уравнение (3.6), получим ординату первой точки. Таким образом имеем точку с координатами (22.10000, 31.96127). Аналогичным образом получим координаты второй точки, положив в качестве абсциссы значение x_{max} = 26.60000. Вторая точка будет: (26.60000, 30.15970).

Линия регрессии показана на рисунке ниже красным цветом. Обратите внимание, что линия регрессии всегда проходит через точку средних значений величин X и Y, т.е. с координатами (M_x , M_y).

1. Графический метод решения ЗЛП.

Пример.

Решить графически данную задачу линейного программирования.

$$F = 4x_1 + 3x_2 - 1 \rightarrow \max$$

$$\begin{cases} x_1 + x_2 \le 8 \\ -2x_1 + 3x_2 \le 9 \\ 2x_1 - x_2 \le 10 \end{cases}$$

$$x_1 \ge 0,\ x_2 \ge 0$$

Решение. Найдем вначале область допустимых решений (ОДР). Решим графически первое неравенство:

$x_1 + x_2 \le 8$ (1)

Для этого построим вначале прямую линию, соответствующую уравнению:

$x_1 + x_2 = 8$. (11)

Поскольку, если $x_1 = 0,$ то $x_2 = 8,$ то прямая (11) проходит через точку M1(0;8). Аналогично, если $x_1 = 8,$ то $x_2 = 0,$ и прямая (11) проходит также через точку M2 (8;0). Проведем через эти две точки прямую линию и отметим ее с помощью 1 (см. рис. 1). Эта линия делит плоскость на две полуплоскости, которые мы условно назовем верхней и нижней полуплоскостями. Так как координаты точки (0;0) удовлетворяют неравенству (1), то этому неравенству соответствует нижняя полуплоскость, которая содержит эту точку. Этот факт мы изобразим на рис. 1 штрихами, направленными вниз от линии 1 .

Теперь решим графически второе неравенство: $-2x_1 + 3x_2 \le 9.$ (2)

Ему соответствует прямая, заданная уравнением: $-2x_1 + 3x_2 = 9$ (21)

Ее мы построим несколько иначе. Перепишем уравнение (21) в виде:

$$x_2 = \frac{2}{3}x_1 + 3$$

Тогда при $x_1 = 0$ оказывается $x_2 = 3$, что дает точку М3 (0;3) искомой прямой. Угловой коэффициент этой прямой $к = \frac{2}{3}$. Но угловой коэффициент любой прямой равен $tg\alpha$, где α- угол наклона прямой к оси 0х: $tg\alpha = к = \frac{2}{3}$. Если теперь мы отложим три единицы вправо от точки М3 (0;3) и затем две единицы вверх, то получим другую точку М4 (3;5) которая также лежит на прямой (21). Через точки М3 и М4 мы проводим прямую 2 (рис.1). Начало координат (0;0) удовлетворяет (2) и лежит ниже графика линии 2 , поэтому соответствующая полуплоскость является «нижней», что мы и отмечаем штрихами, направленными вниз от прямой 2 (рис.1). Аналогично строим прямую 3.

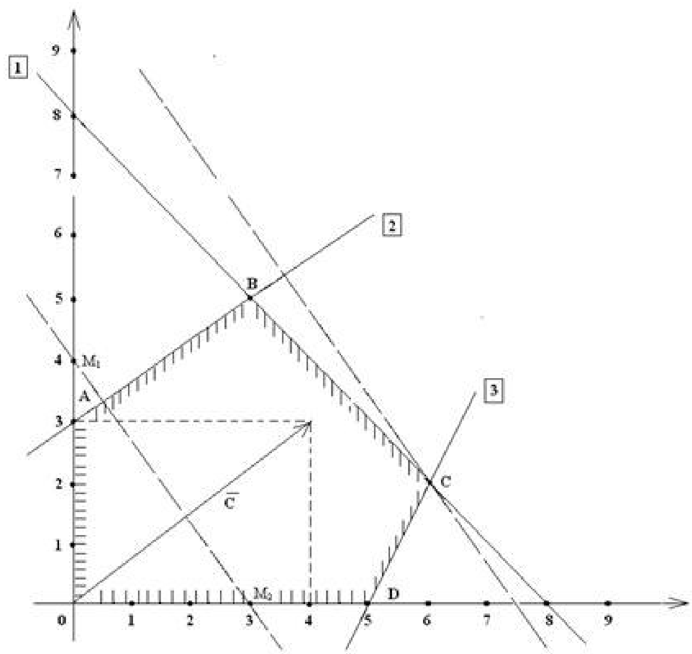
Уравнение $2x_1 - x_2 = 10$ (31) заменяем на уравнение $x_2 = 2x_1 - 10$.

Ясно, что прямая проходит через т. М5 (5;0), и имеет угловой коэффициент к = 2. При этом самому неравенству $2x_1 - x_2 \le 10 \Leftrightarrow x_2 \ge 2x_1 - 10$

Соответствует верхняя полуплоскость, отмеченная штрихами вверх от прямой 3.

Тривиальному неравенству $x_1 \ge 0$ соответствует правая полуплоскость координатной плоскости, то есть полуплоскость, лежащая справа от вертикальной оси $0x_2$. Ее отмечаем штрихами, направленными вправо от оси $0x_2$. Наконец неравенству $x_2 \ge 0$ соответствует верхняя полуплоскость координатной плоскости, отмеченная штрихами, направленными вверх от оси $0x_1$. Пересечение всех указанных полуплоскостей определяет ОДР данной задачи. На рисунке 1 это область, ограниченная выпуклым пятиугольником OABCD.

Изобразим на рисунке 1 вектор роста $\bar{c}$ целевой функции F. Это вектор $\bar{c}$ началом в т. (0;0) и концом в точке М (4;3), поскольку $\bar{c} = (4;3)$. Построим теперь линию уровня $F(\bar{x}) = 11$. Она определяется уравнением: $4x_1 + 3x_2 = 12$. (4)



Мы взяли здесь константу С =11, для того чтобы точки пересечения прямой (4) с осями x_1 и x_2 имели целые координаты. Действительно, если $x_1 = 0$, то $x_2 = 4$, и, если $x_2 = 0$, То $x_1 = 3$, Что дает две точки М1 (0;4) и М2 (3;0) линии уровня (4). Через них проводим пунктиром соответствующую линию уровня (рис. 1). Она оказывается перпендикулярна вектору роста $\bar{c}$. зок $[M_1;M_2]$ пересекается с ОДР и в каждой его точке х значение целевой функции $F(\bar{x})$ равно 11:

$F(\bar{x}) = 11, \quad x \in [M_1; M_2]$.

Мы знаем, что значение функции F увеличивается в направлении вектора роста $\bar{c}$. Чтобы найти максимальное значение $F(x)$ на ОДР будем параллельно перемещать линию уровня в направлении вектора роста $\bar{c}$ До тех пор, пока, она будет иметь хотя бы одну точку пересечения с ОДР задачи. Из рисунка 1 ясно, что последнее пересечение смещенной линии уровня (4) будет точка C

На этой линии очевидно и будет достигаться максимальное значение целевой функции F в ОДР, поскольку при дальнейшем движении линии уровня в направлении вектора роста, она перестает пересекаться с ОДР. Итак, максимальное значение функция F(x) имеет в точке C. Так как точка C является пересечением прямых 1 и 3, то ее координаты находятся: Рисунок 1. Графическое решение задачи ЛП.

из системы

$$\begin{cases} x_1 + x_2 = 8 \\ 2x_1 - x_2 = 10 \end{cases} \quad (5)$$

Чтобы решить эту систему, сложим оба уравнения. Тогда получим, что $3x_1 = 18$ или $x_1 = 6$.

Из первого уравнения находим, что $x_2 = 8 - x_1 = 8 - 6 = 2$

Итак, координаты точки С найдены : С (6;2). Найдем максимальное значение функции: $F_{max} = F(C) = F(6;2) = 4 \times 6 + 3 \times 2 - 1 = 29$.

Задача решена.

Ответ: максимальное значение целевой функции F достигается в точке С (6;2) и равно 29:

$F_{max} = F(6;2) = 29$.

2. Теоретические основы симплекс- метода. Симплекс метод решения ЗЛП

Пример.

Решить задачу линейного программирования симплекс-методом.

Рассмотрим однородную задачу ЛП:

$$F = 4x_1 + 3x_2 - 1 \to \max$$

$$\begin{cases} x_1 + x_2 \le 8 \\ -2x_1 + 3x_2 \le 9 \\ 2x_1 - x_2 \le 10 \end{cases} \quad (1)$$

$x_1 \ge 0, \quad x_2 \ge 0$

Добавив к левым частям системы неравенств соответствующие балансовые переменные $W_1, W_2, W_3,$ преобразуем задачу (1) в каноническую форму:

$$F = 4x_1 + 3x_2 - 1 \to \max$$

$$\begin{cases} x_1 + x_2 + W_1 = 8 \\ -2x_1 + 3x_2 + W_2 = 9 \\ 2x_1 - x_2 + W_3 = 10 \end{cases} \quad (2)$$

$x_1 \ge 0, \quad x_2 \ge 0, \quad W_1 \ge 0, \quad W_2 \ge 0, \quad W_3 \ge 0.$

Для удобства и единообразия запишем определение целевой функции F в виде уравнения: $F-4x_1-3x_2=-1$ (3)
Запишем (2) и (3) в виде первой симплекс таблицы:

F	x_1	x_2	W_1	W_2	W_3	b	
0	1	1	1	0	0	8	
0	- 2	3	0	1	0	9	(4)
0	2	- 1	0	0	1	10	
1	- 4	- 3	0	0	0	- 1	

Первые три строки таблицы (4) содержат, по сути, расширенную матрицу системы линейных уравнений (2), к которой слева приписан столбец переменной F. Последняя строка, называемая индексной, содержит уравнение (3). Буквой b, как обычно, обозначен столбец свободных членов. Отметим, что таким образом составленная таблица (4) называется симплексной, поскольку задача (2) имеет симплексную форму. Напомним, что это означает, что, во-первых, матрица системы (и таблица (4)) содержат т базисных столбцов (столбцы W_1, W_2, W_3), где т - число уравнений (в данном случае $m=3$); во-вторых, все элементы столбца свободных членов неотрицательны (это числа 8, 9 и 10), кроме, возможно, элемента индексной строки; в - третьих, целевая функция F зависит только от свободных переменных (x_1 И x_2). Последнее верно, поскольку в базисных столбцах (W_1, W_2 и W_3) в индексной строке находятся только нули. Первая симплекс-таблица (4) определяет первое опорное решение. Напомним, что опорное решение является допустимым базисным решением, и, следовательно, свободные переменные x_1 и x_2 равны нулю: $x_1=0$ и $x_2=0$.

Далее, переменная W_1 определяется первой строкой таблицы (4), которая является сокращённой записью первого уравнения системы (2). При $x_1=0, x_2=0$ оно принимает вид:
$1\cdot 0+1\cdot 0+W_1=8 \Rightarrow W_1=8$

Вторая строка таблицы определяет переменную W_2:
$-2\cdot 0+3\cdot 0+W_2=9 \Rightarrow W_2=9$

Третья строка определяет W_3: $2\cdot 0-1\cdot 0+W_3=10 \Rightarrow$
$W_3=10$ Значение целевой функции определяем по индексной строке: $F_1=-1$.
В дальнейшем мы покажем, что оптимальное решение канонической задачи ЛП является опорным, и, следовательно, его следует искать среди опорных решений. Симплекс-таблица (4) и дает одно из таких решений. Как проверить, является ли оно оптимальным? Оказывается, просто. Как мы увидим далее, если коэффициенты $F-c_1x_1-c_2x_2-....-c_nx_n=c_0$, целевой функции F Канонической задачи ЛП неположительные: $c_j\le 0$,- и функция F зависит только от свободных переменных, то соответствующее опорное решение является оптимальным.

Но условие $c_j\le 0$ означает, что коэффициенты индексной строки, стоящие в столбцах свободных переменных, должны быть неотрицательны: $-c_j\ge 0$, поскольку индексная строка соответствует уравнению: $F-c_1x_1-c_2x_2-....-c_nx_n=c_0$, - и содержит коэффициенты c_j с противоположным знаком.

Мы видим, что в таблице (4) условие неотрицательности всех элементов индексной строки (разумеется, кроме правой части c_0, стоящей в столбце свободных членов) не выполнено. Более того, оба столбца свободных переменных x_1 и x_2 содержат в индексной строке отрицательные элементы: - 4 и -3, - соответственно.
Выберем любой из этих столбцов, например, первый и назовем его Ведущим. Определим для каждого $i = 1,\dots,m$ так называемое допустимое отношение α_i следующим образом. Если в i - ой строке Ведущего столбца стоит неположительный элемент, то положим $\alpha_i = +\infty,$ если же этот элемент $a_{ij} > 0,$ то положим $\alpha_i = \dfrac{b_i}{a_{ij}}$,
Где j - номер Ведущего столбца.
В нашем случае $j = 1$ и допустимые отношения $\alpha_1, \alpha_2, \alpha_3$ Соответственно равны:

$$\alpha_1 = \frac{b_1}{a_{11}} = \frac{8}{1} = 8, \qquad (a_{11} = 1 > 0),$$

$$\alpha_2 = +\infty, \qquad (a_{21} = -2 < 0),$$

$$\alpha_3 = \frac{b_3}{a_{31}} = \frac{10}{2} = 5, \qquad (a_{31} = 2 > 0)$$

Добавим к симплекс-таблице (4) столбец α:

F	x_1	x_2	W_1	W_2	W_3	b	α	
0	1	1	1	0	0	8	8	
0	- 2	3	0	1	0	9	$+\infty$	(5)
0	2	- 1	0	0	1	10	5	
1	- 4	- 3	0	0	0	- 1		

В таблице (5) отрицательный элемент – 4 ведущего столбца взят в рамочку, для того чтобы выделить Ведущий столбец. Можно, разумеется, выделить этот столбец и любым другим разумным образом: цветом, шрифтом и т. п.

Среди всех допустимых отношений α_i найдем наименьшее: $\alpha_{\min} = \min\ \alpha_i = \alpha_3 = 5$. Наименьшее допустимое отношение соответствует третьей строке таблицы, которую мы теперь объявляем Ведущей строкой. На пересечении Ведущей строки и Ведущего столбца стоит Ключевой элемент таблицы. В нашем случае это $a_{31} = 2.$ Выделим в таблице (5) минимальное допустимое отношение и Ключевой элемент, рамочкой:

F	x_1	x_2	W_1	W_2	W_3	b	α	
0	1	1	1	0	0	8	8	
0	- 2	3	0	1	0	9	$+\infty$	(6)
0	2	- 1	0	0	1	10	5 = $\min\ \alpha_i$	
1	- 4	- 3	0	0	0	- 1		

Дальнейшая наша цель состоит в том, чтобы преобразовать методом Гаусса таблицу (6) в новую симплекс-таблицу, первый столбец который стал бы базисным, содержащим число 1 в Ведущей (третьей) строке.
Вначале разделим ведущую строку на Ключевой элемент:

F	x_1	x_2	W_1	W_2	W_3	b	α	
0	1	1	1	0	0	8		
0	- 2	3	0	1	0	9		(7)
0	1	$-\frac{1}{2}$	0	0	$\frac{1}{2}$	5		
1	- 4	- 3	0	0	0	- 1		

В таблице (7) мы не заполняем столбец α, поскольку он нужен, только для того, чтобы определить Ведущую строку, что мы уже сделали. Мы выделили только Ключевой элемент, так как он определяет одновременно и Ведущую строку (третью) и Ведущий столбец (первый).

Проделаем теперь следующие преобразования Гаусса:

1) вычтем из первой строки Ведущую (третью) строку;
2) прибавим ко второй строке Ведущую, умноженную на 2;
3) прибавим к индексной строке Ведущую, умноженную на 4.

В итоге получим новую таблицу:

F	x_1	x_2	W_1	W_2	W_3	b	α	
0	0	$\frac{3}{2}$	1	0	$-\frac{1}{2}$	3		
0	0	2	0	1	1	19		(8)
0	1	$-\frac{1}{2}$	0	0	$\frac{1}{2}$	5		
1	0	- 5	0	0	2	19		

Нетрудно видеть, что мы получили симплекс-таблицу. Действительно, в таблице (8) после перестановки столбца x_1 со столбцом W_3 в последних трех столбцах получается единичная матрица; столбец свободных членов неотрицателен; целевая функция F зависит только от свободных переменных x_2 И W_3

Отметим, что столбец x_1, став базисным, вытеснил «из базиса» столбец W_3. В силу этого обстоятельства проделанный процесс называют операцией однократного замещения. В данном случае эта операция состояла из последовательности элементарных преобразований Гаусса 1), 2) и 3).

Таким образом, получена вторая симплекс-таблица (8), которой соответствует второе опорное решение. Переменные x_1 и W_3 - свободные и, следовательно, $x_2 = 0$ и $W_3 = 0$. Поскольку первое уравнение имеет вид.

$$0 \cdot x_1 + \frac{3}{2} \cdot 0 + W_1 + 0 \cdot W_2 - \frac{1}{2} \cdot 0 = 3,$$

То значение базисной переменной W_1 Равно 3: $W_1 = 3$. Базисная переменная W_2 определяется вторым уравнением:

$$0 \cdot x_1 + 2 \cdot 0 + 0 \cdot W_1 + W_2 + 1 \cdot 0 = 19 \Rightarrow W_2 = 19,$$

А базисная переменная x_1 определяется третьим уравнением (так как в столбце x_1 единица стоит в третьей строке):

$$x_1 - \frac{1}{2} \cdot 0 + 0 \cdot W_1 + 0 \cdot W_2 + \frac{1}{2} \cdot 0 = 5 \Rightarrow x_1 = 5$$

Итак, $x_1 = 5,\quad x_2 = 0,\quad W_1 = 3,\quad W_2 = 19,\quad W_3 = 0$, - второе опорное решение. Новое значение целевой функции F_2 определяется индексной строкой: $F_2 = 19$.
Это опорное решение также не является оптимальным, что следует из того, что в индексной строке таблицы (8) имеется отрицательный элемент (-5) во втором столбце, который мы выберем теперь в качестве Ведущего столбца. Затем найдем ведущую строку с минимальным допустимым отношением и Ключевой элемент:

F	x_1	x_2	W_1	W_2	W_3	b	α	
0	0	$\boxed{\frac{3}{2}}$	1	0	$-\frac{1}{2}$	3	2 $= \min \alpha_i$	
0	0	2	0	1	1	19	9,5	(9)
0	1	$-\frac{1}{2}$	0	0	$\frac{1}{2}$	5	$+\infty$	
1	0	- 5	0	0	2	19		

Разделим ведущую строку (первую) на Ключевой элемент $\left(\frac{3}{2}\right)$:

F	x_1	x_2	W_1	W_2	W_3	b	
0	0	1	$\frac{2}{3}$	0	$-\frac{1}{3}$	2	
0	0	2	0	1	1	19	(10)
0	1	$-\frac{1}{2}$	0	0	$\frac{1}{2}$	5	
1	0	- 5	0	0	2	19	

Выполним теперь следующие преобразования Гаусса:
1) вычтем из второй строки первую, умноженную на 2;
2) прибавим к третьей строке первую, умноженную на $\frac{1}{2}$;
3) прибавим к индексной строке первую, умноженную на 5. В результате получим третью симплекс-таблицу:

F	x_1	x_2	W_1	W_2	W_3	b	
0	0	1	$\frac{2}{3}$	0	$-\frac{1}{3}$	2	
0	0	0	$-\frac{4}{3}$	1	$\frac{5}{3}$	15	(11)
0	1	0	$\frac{1}{3}$	0	$\frac{1}{3}$	6	
1	0	0	$\frac{10}{3}$	0	$\frac{1}{3}$	29	

Ей соответствует третье опорное решение:
$x_1 = 6,\quad x_2 = 2,\quad W_1 = 0,\quad W_2 = 15,\quad W_3 = 0,$
- и значение целевой функции $F_3 = 29$.

Поскольку индексная строка таблицы (11) не содержит отрицательных элементов, полученное опорное решение будет оптимальным: $F_{опт} = F^* = 29$ и $x_1^* = 6, \quad x_2^* = 2.$ (12) При этом $W_1^* = 0, \quad W_2^* = 15, \quad W_3^* = 0.$ Здесь мы звездочками помечаем оптимальные значения переменных.
Таким образом, задача (2), а с ней и задача (1) решены.

3.5.3 Результаты и выводы: В результате проведенного занятия студенты:
- ознакомились с основными понятиями линейного, математического программирования, теории двойственности, основными теоремами и свойствами моделей ЗЛП;
- усвоили теоретические основы симплекс метода;
- выработали навыки применения графического метода решения ЗЛП, симплекс- метода..

2.4.3 Результаты и выводы: В результате проведенного занятия студенты:
- ознакомились с понятием простейшего потока, его свойствами, классификацией случайных потоков;
- усвоили основные алгоритмы анализа и построения марковских цепей, их особенности и условия применения;
- выработали навыки математического моделирования марковских цепей при решении инженерных задач.

2.5. Практическое занятие 17 (ПЗ-17) (2 часа)
Тема: Культура и мастерство исследователя

2.5.1 Задание для работы:

1. Виды НИР
2. Вопросы методологии
3. Этапы НИР.
1. Основные принципы построения НИРС
2. Правила оформления научно-исследовательских работ. Работа с информационными источниками. Этические аспекты НИР

2.5.2 Краткое описание проводимого занятия:

1. Виды НИР – доклады с презентациями

2. Вопросы методологии - доклады с презентациями

3. Этапы НИР - доклады с презентациями
4. Основные принципы построения НИРС. Правила оформления научно-исследовательских работ. Работа с информационными источниками. Этические аспекты НИР - доклады с презентациями

2.5.3 Результаты и выводы: В результате проведенного занятия студенты:
- ознакомились с классификацией НИР, ее особенностями;
- усвоили основные требования к методологии научно-исследовательских работ;

- выработали навыки анализа научной проблемы и построение поэтапного плана ее решения;
- усвоили основные правила оформления научно-исследовательских работ; этические нормы и правила ведения научно-исследовательской работы;
- выработали навыки работы с информационными источниками.