

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

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ ОБУЧАЮЩИХСЯ
ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Б1.В.01 Основы научных исследований

Спеальность 10.05.03 Информационная безопасность автоматизированных систем

Специализация Информационная безопасность автоматизированных систем критически
важных объектов

Форма обучения очная

СОДЕРЖАНИЕ

1. Конспект лекций	3
1.1 Лекция № 1 Наука в современном обществе. Организация научно-исследовательской работы в России: история и современность	3
1.2 Лекция № 2 Основы методологии научного исследования	10
1.3 Лекция № 3 Виды научно-исследовательских работ, их особенности, оформление.....	14
1.4 Лекция № 4 Оптимизационные задачи. Основные методы их решения.....	21
1.5 Лекция № 5 Оптимизационные задачи. Основные методы их решения.....	30
1.6 Лекция № 6 Марковские процессы, их приложения к решению инженерных задач.....	39
1.7 Лекция № 7 Теоретические основы обработки экспериментальных данных.....	47
1.8 Лекция № 8 Корреляционно-регрессионный анализ.....	60
2. Методические указания по проведению практических занятий	80
2.1 Практическое занятие № ПЗ-1 Проблемы современной фундаментальной науки	80
2.2 Практическое занятие № ПЗ-2 Проблемы современной фундаментальной науки.....	80
2.3 Практическое занятие № ПЗ-3 Методологическая основа научно-исследовательской работы.....	81
2.4 Практическое занятие № ПЗ-4 Научно-исследовательская работа студентов.....	81
2.5 Практическое занятие № ПЗ-5 Виды научно-исследовательских работ, их особенности, оформление.....	82
2.6 Практическое занятие № ПЗ-6 Математическое моделирование в инженерных исследованиях.....	82
2.7 Практическое занятие № ПЗ-7 Задачи линейной оптимизации. Методы их решения....	83
2.8 Практическое занятие № ПЗ-8 Задачи линейной оптимизации. Методы их решения....	90
2.9 Практическое занятие № ПЗ-9 Задачи нелинейной оптимизации.....	98
2.10 Практическое занятие № ПЗ-10 Задачи нелинейной оптимизации.....	102
2.11 Практическое занятие № ПЗ-11 Основы теории массового обслуживания.....	104
2.12 Практическое занятие № ПЗ-12 Марковские цепи, процессы.....	111
2.13 Практическое занятие № ПЗ-13 Основные понятия и методы математической обработки экспериментальных данных.....	111
2.14 Практическое занятие № ПЗ-14 Основные понятия и методы математической обработки экспериментальных данных.....	119
2.15 Практическое занятие № ПЗ-15 Корреляционно-регрессионный анализ.....	126
2.16 Практическое занятие № ПЗ-16 Корреляционно-регрессионный анализ.....	131

1. КОНСПЕКТ ЛЕКЦИЙ

1. 1 Лекция №1 (2 часа).

Тема: «Наука в современном обществе. Организация научно-исследовательской работы в России: история и современность»

1.1.1 Вопросы лекции:

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



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).

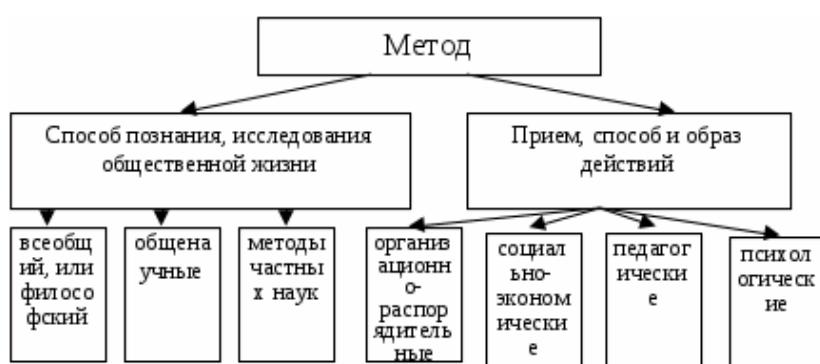


Рис. 2 –Группы научных методов

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*.

3. 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, ИТ, 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).

4. 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.



Рис. 3 – 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». 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». 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. Фонд оказывает финансовую поддержку высокоеффективным 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, которое также занимается Инвестиционным фондом РФ. 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 инфраструктуры 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. 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. Еще одним шагом государства для реализации разработок ИТ-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. 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. 2 Лекция №2 (2 часа).

Тема: «Основы методологии научного исследования»

1.2.1 Вопросы лекции:

1. Научная теория как высшая форма организации знания
2. Методики научных исследований

1.2.2 Краткое содержание вопросов:

1. Научная теория как высшая форма организации знания

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.

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

В частности, *научное творчество* связано 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).

2. Методики научных исследований

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.

1. 3 Лекция №3 (2 часа).

Тема: «Виды научно-исследовательских работ, их особенности, оформление»

1.3.1 Вопросы лекции:

1. Виды научно-исследовательских работ (НИР)
2. Основные принципы построения научно-исследовательской работы студентов (НИРС)

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

1.3.2 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.

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 оправдано.

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

Основными формами НИРС, выполняемой во внеучебное время 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).

Даная 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 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 прийдя 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 развернутые 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 представляют 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 также практическая помощь предприятиям. Работа в такой лаборатории предполагает не только 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 придики. 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 проводится 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 работы является 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 *объективности* 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.

3. Этические аспекты научно-исследовательской деятельности

Этические аспекты научной деятельности. Понятие научного этиоса 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 технико-технологии 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.

1. 4 Лекция №4 (2 часа).

Тема: «Оптимационные задачи. Основные методы их решения»

1.4.1 Вопросы лекции:

1. Оптимизация как цель математического моделирования. Виды оптимационных задач
2. Задача линейного программирования (ЗЛП). Графический метод решения ЗЛП
3. Симплекс-метод решения ЗЛП

1.4.2 Краткое содержание вопросов:

1. Оптимизация как цель математического моделирования. Виды оптимационных задач

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. Ограничеными 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.

2. Задача линейного программирования (ЗЛП). Графический метод решения ЗЛП

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

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

при ограничениях

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

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

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

$$x_j \geq 0 \quad (j = \overline{1, n_1}) \quad (2.5)$$

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

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:

$$\max Z = c\vec{x} \quad (2.7)$$

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

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 \quad (2.9)$$

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 $x_j \geq 0 \quad (j = \overline{1, n})$ оптимальным. В этом случае проблема выбора оптимального решения теряет смысла. 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

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

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 идеи ее решения, делает геометрически наглядными способы решения и пути их практической реализации.

Пусть дана задача

$$\max Z = c_1 x_1 + c_2 x_2 \quad (2.11)$$

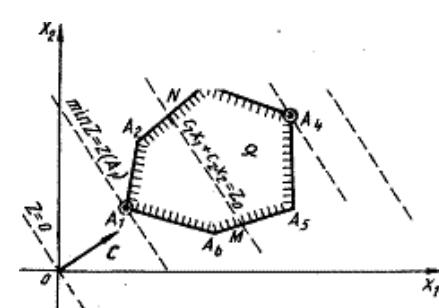
$$\left. \begin{array}{l} a_{11} x_1 + a_{12} x_2 \leq b_1 \\ a_{21} x_1 + a_{22} x_2 \leq b_2 \\ \dots \\ a_{m1} x_1 + a_{m2} x_2 \leq b_m \end{array} \right\} \quad (2.12)$$

$$x_1 \geq 0, x_2 \geq 0 \quad (2.13)$$

Дадим геометрическую интерпретацию элементов этой задачи. Каждое из ограничений (2.12), (2.13) задает на плоскости $x_1 O x_2$ 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

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 :

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

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

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:

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

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

Вектор $\vec{c} = (c_1, c_2)$ перпендикулярен к прямым $Z = \text{const}$ 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 \vec{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 $Z^* = z(\vec{x}^*)$.

3. Симплексный метод решения 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:

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

Говорят, что ограничение ЗЛП имеет предпочтительный вид, если при

неотрицательной правой части ($b_i \geq 0$) левая часть ограничений содержит переменную, входящую с коэффициентом, равным единице, а в остальные ограничения равенства - с коэффициентом, равным нулю.

Пусть система ограничений имеет вид

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

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

$$\begin{aligned} \sum_{j=1}^n a_{ij}x_j + x_{n+1} &= b_i, \quad b_i \geq 0 \quad (i = 1, \dots, m), \\ \vec{x}_o &= (0; \underbrace{0, \dots, 0}_{n}; \underbrace{b_1, b_2, \dots, b_m}_{m}) \end{aligned}$$

которая имеет предпочтительный вид

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

Пусть далее система ограничений имеет вид

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

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

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

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

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

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

Пусть исходная ЗЛП имеет вид

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

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

$$x_j \geq 0 \quad (j = 1, \dots, n), \quad (2.18)$$

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

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

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

$$x_j \geq 0 \quad (j = \overline{1, n}), \quad \omega_i \geq 0, \quad (i = 1, \dots, m) \quad (2.21)$$

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

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

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

Теорема. Если в оптимальном плане

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

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

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

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 $\Delta_i \quad (j = \overline{1, n})$ неотрицательны, то такой план оптimalен.

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

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 ($i = 1, \dots, m$). 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 ($j = \overline{1, n}$) 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:

$$\max Z = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (2.23)$$

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

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

Предположим далее, что с самого начала при изучении вопроса об использовании отходов основного производства на предприятии появилась возможность реализации их некоторой организацией. 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 покупающей 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 $a_{11}y_1 + a_{12}y_2 + \dots + a_{m1}y_m \geq c_1$, где левая часть означает выручку за сырьё идущее на единицу продукции первого вида; правая – её цену.

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

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

По смыслу задачи оценки не должны быть отрицательными:
 $y_1 \geq 0, y_2 \geq 0, \dots, y_m \geq 0$. (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 \leq . Двойственная задача решается на минимум, и ее система ограничений имеет вид неравенств типа \geq .

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

7. Все переменные в обеих задачах неотрицательны.

Основные теоремы двойственности и их экономическое содержание

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

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

— основное неравенство теории двойственности.

Теорема (критерий оптимальности Канторовича).

Если для некоторых допустимых планов \bar{x}^* и \bar{y}^* пары двойственных задач выполняется неравенство $z(\bar{x}^*) = f(\bar{y}^*)$, то \bar{x}^* и \bar{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: $z(\bar{x}^*) = f(\bar{y}^*)$. 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 обуславливают 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 \bar{x}^* и \bar{y}^* пары двойственных задач были оптимальны, необходимо и достаточно выполнение условий:

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

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

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 расхода 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

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

1. 5 Лекция №5 (2 часа).

Тема: «Оптимизационные задачи. Основные методы их решения»

1.5.1 Вопросы лекции:

1. Транспортная задача. Математическая модель
2. Составление опорного плана
3. Методы решения транспортной задачи

1.5.2 Краткое содержание вопросов:

1. Транспортная задача. Математическая модель

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} .

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

Далее, предполагается, что

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

$\sum_{i=1}^m a_i > \sum_{j=1}^n b_j$, то количество продукции, равное $\sum_{i=1}^m a_i - \sum_{j=1}^n b_j$, остается на складах. В этом

случае мы введем "фиктивного" потребителя $n+1$ с потребностью $\sum_{i=1}^m a_i - \sum_{j=1}^n b_j$ и положим транспортные расходы $p_{i,n+1}$ равными 0 для всех i .

$$\sum_{i=1}^m a_i \sum_{j=1}^n b_{ij},$$

3. Если сумма поданных заявок превышает наличные запасы потребность не может быть покрыта. Этую задачу можно свести к обычной транспортной задаче с правильным балансом, если ввести фиктивный пункт

$$\sum_{i=1}^m b_{ij} - \sum_{j=1}^n a_j$$

отправления $m+1$ с запасом и стоимость перевозок из фиктивного пункта отправления во все пункты назначения принять равным нулю. Математическая модель транспортной задачи имеет вид:

$$\begin{aligned} \sum_{j=1}^n x_{ij} = a_j \quad (i = 1, \dots, m); \quad \sum_{i=1}^m \sum_{j=1}^n C_{ij} x_{ij} \rightarrow \min \quad \sum_{i=1}^m x_{ij} = b_j \quad (j = 1, \dots, n); \\ x_{ij} \geq 0 \quad (i = 1, \dots, m; j = 1, \dots, n) \end{aligned}$$

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

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

Ограничения имеют вид

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

2. 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 B₁ подал заявку на 18 единиц груза. Удовлетворим эту заявку за счёт запаса 48, имеющегося в пункте A₁, и запишем перевозку 18 в клетке (1,1). После этого заявка

пункта B₁ удовлетворена, а в пункте A₁ осталось ещё 30 единиц груза.

Удовлетворим за счёт них заявку пункта B₂ (27 единиц), запишем 27 в клетке (1,2); оставшиеся 3 единицы пункта A₁ назначим пункту B₃. В составе заявки пункта B₃ остались неудовлетворёнными 39 единиц. Из них 30 покроем за счёт пункта A₂, чем его запас будет исчерпан, и ещё 9 возьмём из пункта A₃. Из оставшихся 18 единиц пункта A₃ 12 выделим пункту B₄; оставшиеся 6 единиц назначим пункту B₅, что вместе со всеми 20 единицами пункта A₄ 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.

3. Методы решения транспортной задачи

Распределительный метод достижения оптимального плана

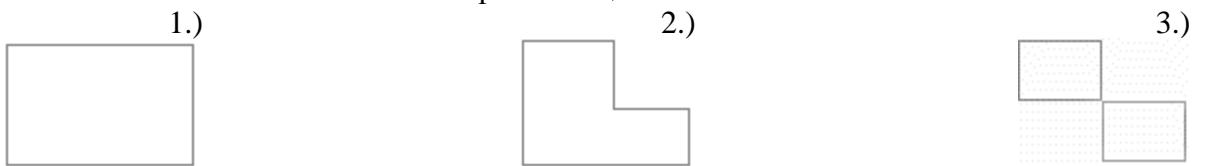
Теперь попробуем улучшить план, составленный способом северо-западного угла. Перенесем, например, 18 единиц из клетки (1,1) в клетку (2,1) и чтобы не нарушить баланса перенесём те же 18 единиц из клетки (2,3) в клетку (1,3). Получим новый план. Подсчитав стоимость опорного плана (она равняется 1039) и стоимость нового плана (она равняется 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	7	8	6	5	30

	18		12			
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 совершают поворот на 90°. Существует несколько вариантов цикла:



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 означеному 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 уменьшаться. 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 увеличится 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.

Метод потенциалов позволяет автоматически выделять циклы с отрицательной ценой и определять их цены.

Пусть имеется транспортная задача с балансовыми условиями

$$\begin{aligned} \sum_{j=1}^n x_{ij} &= a_i \quad (i = 1, \dots, m); \\ \sum_{i=1}^m x_{ij} &= b_j \quad (j = 1, \dots, n); \\ x_{ij} &\geq 0 \quad (i = 1, \dots, m; j = 1, \dots, n) \\ \sum_{i=1}^m a_i &= \sum_{j=1}^n b_j \end{aligned}$$

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 сводится 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 $\alpha_i + \beta_j = \check{c}_{ij}$ ($i=1..m$; $j=1..n$) 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 равны 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 :

$$\check{c}_{ij} = c_{ij} \text{ (для всех базисных клеток)} \quad (2.36)$$

$$\check{c}_{ij} \leq c_{ij} \text{ (для всех свободных клеток)} \quad (2.37)$$

называется **потенциальным** планом, а соответствующие ему платежи (α_i и β_j) — потенциалами пунктов A_i и B_j ($i=1, \dots, m$; $j=1, \dots, n$). 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_i и 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 $\check{c}=7$	8 $\check{c}=6$	5 42	6 6	9 $\check{c}=6$	$\alpha_1=0$
A_2	6 4	7 $\check{c}=5$	8 $\check{c}=4$	6 $\check{c}=5$	5 26	$\alpha_2=-1$
	8	7	10	8	7	$\alpha_3=1$

A_3	$\check{c} = 8$	27	$\check{c} = 6$	$\check{c} = 7$	0	
A_4	7 14	5 $\check{c} = 6$	4 $\check{c} = 5$	6 6	8 $\check{c} = 6$	$\alpha_4 = 0$
β_j	$\beta_1 = 7$	$\beta_2 = 6$	$\beta_3 = 5$	$\beta_4 = 6$	$\beta_5 = 6$	

$$\alpha_4 = 0, \Rightarrow$$

$$\beta_4 = 6, \text{ так как } \alpha_4 + \beta_4 = C_{44} = 6, \Rightarrow \alpha_1 = 0, \text{ так как } \alpha_1 + \beta_4 = C_{14} = 6, \Rightarrow$$

$$\beta_3 = 5, \text{ так как } \alpha_1 + \beta_3 = C_{13} = 5, \Rightarrow \beta_1 = 7, \text{ так как } \alpha_4 + \beta_1 = C_{41} = 7, \Rightarrow$$

$$\alpha_2 = -1, \text{ так как } \alpha_2 + \beta_1 = C_{21} = 6, \Rightarrow \beta_5 = 6, \text{ так как } \alpha_2 + \beta_5 = C_{25} = 5, \Rightarrow$$

$$\alpha_3 = 1, \text{ так как } \alpha_3 + \beta_5 = C_{35} = 7, \Rightarrow \beta_2 = 6, \text{ так как } \alpha_3 + \beta_2 = C_{25} = 7.$$

Если оказалось, что все эти псевдостоимости не превосходят стоимостей $\check{c}_{ij} \leq c_{ij}$, то план потенциален и, значит, оптimalен. 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 равна разности между стоимостью и псевдостоимостью в этой свободной клетке. В таблице № 2.5 мы получили в двух клетках $\check{c}_{ij} \geq c_{ij}$, 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 +	7	8	6	5 -26	-1
A_3	8	7 29	10	8	9 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_i и 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 $(\alpha_i$ и $\beta_j)$ 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 оптимальен.

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 оптимальен, 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 $F_{\min} = 703$.

1. 6 Лекция №6 (2 часа).

Тема: «Марковские процессы, их приложения к решению инженерных задач»

1.6.1 Вопросы лекции:

1. Простейший поток, его свойства. Классификация потоков.
2. Марковские цепи, их свойства
3. Марковские процессы в инженерной практике

1.6.2 Краткое содержание вопросов:

1. Простейший поток, его свойства. Классификация потоков.

2. Марковские цепи, их свойства

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.

Марковские процессы Представление случайных процессов графом состояний

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

число которых конечно (или счетно). Состояния s_1, s_2, \dots 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. Имеются 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 .

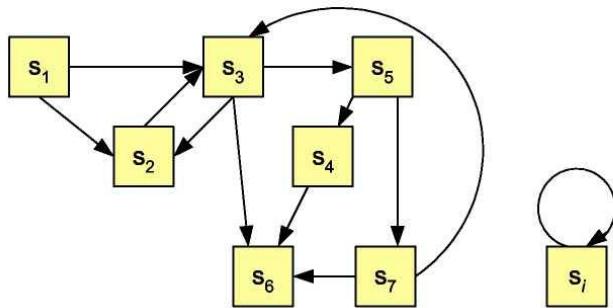


Рисунок 1 – Пример графа состояний

Переход по стрелке, ведущей из состояния s_i в него же, означает задержку системы в состоянии s_i . «Обратные стрелки» можно на графике 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 графике 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 графике 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 $s_i \in V$ 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.

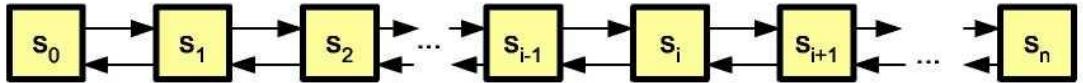


Рисунок 2 - Схема процесса гибели и размножения

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

Если на графике 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: $s_1, s_2, \dots, s_i, \dots, s_n$, а бесконечное (счетное): $s_1, s_2, \dots, s_i, \dots$.

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 $p_i(t)$:

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

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

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

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 $p_i(t)$ ($i = 1, 2, \dots$) остаются постоянными. Обозначим эти постоянные вероятности p_i .

$$p_i = \lim_{t \rightarrow \infty} p_i(t) \quad (4)$$

Вероятности p_i ($i = 1, 2, \dots$), если они существуют, называются финальными (пределными) 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 $s_1, s_2, \dots, s_i, \dots$, 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

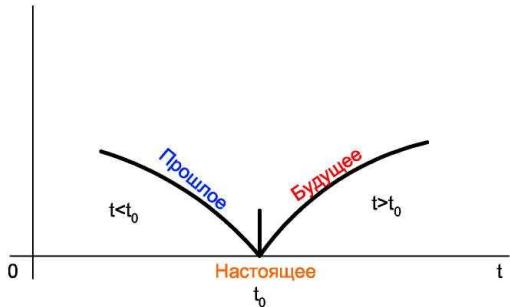


Рисунок 3 – Схема марковского свойства случайного процесса

«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 $s_1, s_2, \dots, s_i, \dots, s_n$. Предположим, что случайные переходы («перескоки») системы из состояния в состояние могут происходить только в определенные моменты времени t_0, t_1, t_2, \dots . 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: $s_1^{(1)}, s_2^{(1)}, \dots, s_i^{(1)}, \dots, s_n^{(1)}$; на втором шаге - $s_1^{(2)}, s_2^{(2)}, \dots, s_i^{(2)}, \dots, s_n^{(2)}$, на k -м шаге

$s_1^{(k)}, s_2^{(k)}, \dots, s_i^{(k)}, \dots, s_n^{(k)}$ (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.

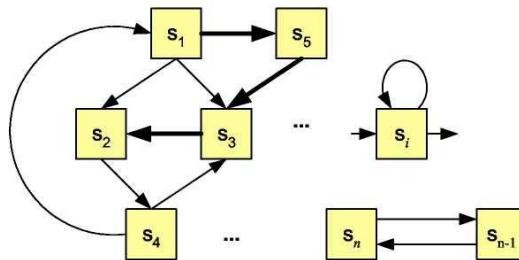


Рисунок 4 – Граф состояний системы S

«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 $s_1, s_2, \dots, s_i, \dots, s_n$, которые она может принимать в последовательности шагов с номерами $0, 1, 2, \dots, k, \dots$.

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

$$\left. \mathbf{P} \left[S(k) = s_i \right] \right\} \quad (i=1,2,\dots,n; k=0,1,2,\dots), \quad (5)$$

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 $t_0, t_1, t_2, \dots, t_k$.

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, ъ, э, ю, я, «пробел», то сразу ясно, что вероятность следующей буквы быть той или другой зависит от того, какова была предыдущая (например, последовательности букв «яы» или «эъ» 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 $p_i(k)$:

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

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_j . Обозначим эту вероятность

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

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 $n \times n$:

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

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

$$p_{11}(k), p_{22}(k), \dots, p_{ii}(k), \dots, p_{nn}(k) \quad (9)$$

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:

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

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

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

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 $t_0 = 0$:

$$p_1(0), p_2(0), \dots, p_i(0), \dots, p_n(0), \quad (12)$$

в сумме образующие единицу:

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

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:

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

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:

$$\|p_{ij}\| = \begin{vmatrix} p_{11} & p_{12} \dots p_{1j} \dots p_{1n} \\ p_{21} & p_{22} \dots p_{2j} \dots p_{2n} \\ \dots & \dots \\ p_{i1} & p_{i2} \dots p_{ij} \dots p_{in} \\ \dots & \dots \\ p_{n1} & p_{n2} \dots p_{nj} \dots p_{nn} \end{vmatrix} \quad (14)$$

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 $p_i(k)$ ($k = 1, 2, \dots$) 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 ($j = 1, 2, \dots, n$)

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

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

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

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

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 $p_{ij}(k)=P\{S(1)=s_j | S(0)=s_i\}$.

По формуле полной вероятности получим:

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

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:

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

По формуле полной вероятности находим

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

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:

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

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):

$$p_1(0)=1, p_2(0)=0; p_1(1)=0,7; p_2(1)=0,3; p_1(2)=0,61; p_2(2)=0,39; p_1(3)=0,583; p_2(3)=0,417; p_1(4)=0,5749; p_2(4)=0,425$$

Ниже мы покажем, что в этом случае $p_1=\lim_{k \rightarrow \infty} p_1(k)=0,4/(0,4+0,3)=0,5714$. 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.

1. 7 Лекция №7 (2 часа).

Тема: «Теоретические основы обработки экспериментальных данных»

1.7.1 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.

1.7.2 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.



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): $P_i = \frac{m_i}{\sum_{i=1}^v m_i}$,

где v – число вариантов. Полагая $n = \sum_{i=1}^v m_i$, 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 вариирования ряда R : $R = x_{\max} - x_{\min}$

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

3. Для более точного определения величины частичного интервала можно

$$\text{воспользоваться формулой Стерджеса: } h = \frac{R}{1 + 3,322 \lg n}.$$

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: $x_{\text{нач}} = x_{\text{мин}} - 0,5h$.

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

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 Бернули, оправдывает целесообразность использования функции $F(x)$ * при больших n 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. $0 \leq F^*(x) \leq 1$
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 ряда. 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 (частоты). 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 оси Ox 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 \bar{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 (её может быть и $M(X)$, 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, \dots, X_n наблюдений величины X . Обозначая буквой f этот алгоритм, запишем $\Theta^*_{(n)}=f(X_1, X_2, \dots, X_n)$. (3)

Подставив в (3) вместо X_1, X_2, \dots, 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, \dots, 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, \dots, 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, несмешённости и эффективности.

Оценка $\Theta^*_{(n)}$ генеральной характеристики Θ называется **состоятельной**, если для любого $\epsilon > 0$ выполняется равенство $\lim_{n \rightarrow \infty} P(|\Theta^*_{(n)} - \Theta| < \epsilon) = 1$. (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 **несмешённой**, если для любого фиксированного числа наблюдений n выполняется равенство $M(\Theta^*_{(n)}) = \Theta$, (5) т.е. математическое ожидание оценки равно неизвестной характеристике.

Несмешённая оценка $\Theta^*_{(n)}$ характеристики Θ называется **несмешённой эффективной**, если она среди всех прочих несмешённых 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 $\ln L$, рассматриваемые как функции параметра λ , достигают максимума при одном и том же значении λ , т.к. $\ln L$ – монотонно возрастающая функция. Поэтому вместо отыскания максимума функции L находят (что удобнее) максимум функции $\ln L$. Функция $\ln L$ называется **логарифмической функцией правдоподобия**.

Для $L(X_1, X_2, \dots, X_n; \lambda) = \frac{\lambda^{x_1} e^{-n\lambda}}{O_1! O_2! \dots O_n!}$ логарифмическая функция правдоподобия имеет вид:

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

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

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

приравняв производную нулю, определим критическую точку – корень полученного уравнения – **уравнения правдоподобия**: $\frac{\sum_{i=1}^n X_i}{\lambda} - n = 0 \rightarrow \lambda_{\text{кр}} = \frac{\sum_{i=1}^n X_i}{n}$;

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

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = \frac{\sum_{i=1}^n X_i}{\lambda^2}; \quad \frac{\partial^2 \ln L(\lambda_{\text{кр}})}{\partial \lambda^2} = \frac{\sum_{i=1}^n X_i}{\lambda_{\text{кр}}^2} = -\frac{n^2}{\sum_{i=1}^n X_i}.$$

Итак, всегда $\lambda_{\text{кр}} = \frac{\sum_{i=1}^n X_i}{n}$ – это точка максимума функции $\ln L$ (или L), поэтому она является оценкой $\lambda_{\text{МП}}$ максимального правдоподобия для неизвестного параметра λ , т.е.

$$\lambda_{\text{кр}}^* = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

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 несмешённости, 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 $\varepsilon > 0$, чтобы выполнялось соотношение

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

Соотношению (8) тождественно соотношению

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

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, \dots, X_n этих наблюдений обладают следующими свойствами:

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

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

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 $\bar{O} = \sum_{i=1}^n X_i / n$ математического ожидания a . Зададимся вероятностью γ и попробуем найти такое число ε , чтобы выполнялось соотношение

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

Интервальная оценка математического ожидания такова:

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

Полученный результат имеет следующий смысл: с вероятностью γ можно быть уверенными 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, \dots, 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. $\alpha = D_{I_0}(H_1)$, (13)

где $D_{I_0}(H_1)$ – 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, $\alpha = 0,05$ 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 $D_{I_0}(H_1) = \alpha$	Правильное решение, его вероятность $D_{I_0}(H_0) = 1 - \alpha$
Гипотеза H_0 неверна, т.е. верна гипотеза H_1	Правильное решение, его вероятность $D_{I_1}(H_1) = 1 - \beta$	Ошибка второго рода, её вероятность $D_{I_1}(H_0) = \beta$

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, \dots, X_n , т.е. для которой справедливо равенство $\phi = \phi(X_1, X_2, \dots, X_n)$;

- её значения позволяют судить о «расхождении выборки с гипотезой H_0 »;

- и она, будучи величиной случайной в силу случайности выборки X_1, X_2, \dots, 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.о.бр., 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.

$$P(\phi \in \omega) = \alpha. \quad (15)$$

Однако критическая область равенством (15) определяется неоднозначно. Действительно, представив себе график функции плотности $f_\phi(x)$ 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 $(x^{\text{kp}}_{\text{пр}, \alpha}, +\infty)$, где точка $x^{\text{kp}}_{\text{пр}, \alpha}$ определяется из условия

$$P(\phi > x^{\text{kp}}_{\text{пр}, \alpha}) = \alpha \quad (16)$$

и называется **правосторонней критической точкой**, отвечающей уровню значимости α ; **левосторонняя критическая область**, состоящая из интервала $(-\infty, x^{\text{kp}}_{\text{лев}, \alpha})$, где точка $x^{\text{kp}}_{\text{лев}, \alpha}$ определяется из условия

$$P(\phi < x^{\text{kp}}_{\text{лев}, \alpha}) = \alpha \quad (17)$$

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

двусторонняя критическая область, состоящая из следующих двух интервалов:

$$(-\infty, x^{\text{kp}}_{\text{лев}, \alpha/2})$$

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, \dots, 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 $\phi = \phi(X_1, X_2, \dots, X_n)$ вместо X_1, X_2, \dots, X_n подставляют конкретные числа, полученные в результате наблюдений, и подсчитывают числовое значение $\phi_{\text{чис}}$ критерия.

Если $\phi_{\text{чис}}$ 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 $\phi_{\text{чис}}$ не попадает в критическую область, гипотеза H_0 не отвергается. Но это вовсе не означает, что H_0 является единственной 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

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

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 -распределение 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 v (см. строки 1 и 2 табл. 7).

Таблица 7.

x_i	x_1	...	x_{v-1}	x_v
m_i	m_1	...	m_{v-1}	m_v
$p_i^{\text{теор}} = P(X = x_i)$	$p_1^{\text{теор}} = P(X = x_1)$...	$p_{v-1}^{\text{теор}} = P(X = x_{v-1})$	$p_v^{\text{теор}} = 1 - p_1^{\text{теор}} - \dots - p_{v-1}^{\text{теор}}$
$m_i^{\text{теор}} = np_i^{\text{теор}}$	$m_1^{\text{теор}} = np_1^{\text{теор}}$...	$m_{v-1}^{\text{теор}} = np_{v-1}^{\text{теор}}$	$m_v^{\text{теор}} = np_v^{\text{теор}}$

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 , $i=1, 2, \dots, v-1$; эти вероятности называют **теоретическими**. Обратим внимание на следующее обстоятельство: т.к. сумма вероятностей ряда распределения должна быть равна единице,

$$\text{т.е. } \sum_i p_i^{\text{теор}} = 1, \quad (19)$$

то полагаем вероятность $p_v^{\text{теор}} = 1 - p_1^{\text{теор}} - p_2^{\text{теор}} - \dots - p_{v-1}^{\text{теор}}$. 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

$$m_i^{\text{теор}} \geq 5, i=1, 2, \dots, v. \quad (20)$$

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 v .

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

$$\varphi = \sum_{i=1}^v \frac{(n_i - m_i^{meop})^2}{m_i^{meop}} \quad (21)$$

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

$$\varphi = \sum_{i=1}^v \frac{(n_i - m_i^{\text{действ}})^2}{m_i^{\text{действ}}} = \chi^2(k = v - l - 1),$$

где v – 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 $(x^{\text{kp}}_{\text{пр}, \alpha}, +\infty)$; найдём числовое значение $\varphi_{\text{чис}}$ критерия (21). Если $\varphi_{\text{чис}}$ попадает в интервал $(x^{\text{kp}}_{\text{пр}, \alpha}, +\infty)$, 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 $\varphi_{\text{чис}}$ не попадает в интервал $(x^{\text{kp}}_{\text{пр}, \alpha}, +\infty)$, то гипотеза H_0 (18) не отвергается.

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

$$\left. \begin{array}{l} \alpha \rightarrow \gamma = 1 - \alpha \\ l, v \rightarrow k = v - l - 1 \end{array} \right\} \longrightarrow \chi^2_{\gamma} \rightarrow x^{\text{kp}}_{\text{пр}, \alpha} = \chi^2_{\gamma}. \quad (22)$$

1. 8 Лекция №8 (2 часа).

Тема: «Теоретические основы обработки экспериментальных данных»

1.8.1 Вопросы лекции:

1. Многомерные СВ, законы их распределения, условные числовые характеристики
2. Функция регрессии, коэффициент детерминации, корреляции, ковариация
3. Виды регрессий, статистическая значимость их параметров. Автокорреляция

1.8.2 Краткое содержание вопросов:

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

Совместное рассмотрение двух или нескольких случайных величин приводит к понятию системы случайных величин. Условимся систему нескольких случайных величин X, Y, \dots, W обозначать (X, Y, \dots, W) . Такая система называется также **многомерной**

случайной величиной. При изучении системы случайных величин недостаточно изучить отдельно случайные величины, составляющие систему, а необходимо учитывать связи или зависимости между этими величинами.

При рассмотрении системы случайных величин удобно пользоваться геометрической интерпретацией системы. Например, систему двух случайных величин (X, Y) можно рассматривать как случайную точку на плоскости XOY с координатами X и Y или как случайный вектор на плоскости со случайными составляющими x и y . По аналогии, систему n случайных величин можно рассматривать как случайную точку в n -мерном пространстве или как n -мерный случайный вектор¹.

В дальнейшем, при изучении системы случайных величин ограничимся подробным рассмотрением системы двух случайных величин.

Закон распределения вероятностей системы случайных величин

Законом распределения вероятностей системы случайных величин называется соответствие, устанавливающее связь между областями возможных значений данной системы случайных величин и вероятностями появления системы в этих областях.

Так же, как и для одной случайной величины, закон распределения системы случайных величин может быть задан в различных формах. Рассмотрим таблицу распределения вероятностей системы двух дискретных случайных величин. Пусть X и Y – дискретные случайные величины, возможные значения которых (x_i, y_j) , где $i = \overline{1, n}$; $j = \overline{1, m}$. Тогда распределение системы таких случайных величин может быть охарактеризовано указанием вероятностей $p_{i,j} = P\{X = x_i, Y = y_j\}$ того, что случайная величина примет значение x_i и одновременно с этим случайная величина примет значение y_j . Вероятности $p_{i,j}$ фиксируются в таблице

y_1	y_2	\dots	y_m	
x_j				
x_1	p_{11}	p_{12}	\dots	p_{1m}
x_2	p_{21}	p_{22}	\dots	p_{2m}
\vdots	\vdots	\vdots	\vdots	\vdots
x_n	p_{n1}	p_{n2}	\dots	p_{nm}

Такая таблица называется таблицей распределения вероятностей системы двух дискретных случайных величин с конечным числом возможных значений. Все возможные события $\{X = x_i, Y = y_j\}$ при составляют полную группу несовместных событий, поэтому $\sum_{i,j} p_{i,j} = \sum_{i,j} P\{X = x_i, Y = y_j\} = 1$.

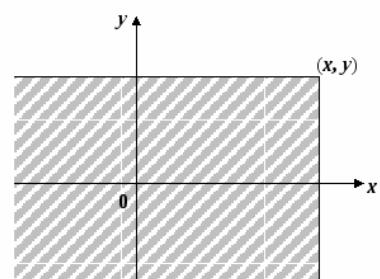
При этом:

$$\sum_j p_{i,j} = \sum_j P\{X = x_i, Y = y_j\} = P\{X = x_i\};$$

$$\sum_i p_{i,j} = \sum_i P\{X = x_i, Y = y_j\} = P\{Y = y_j\}.$$

Функция распределения

Функцией распределения вероятностей системы двух случайных величин называется функция $F(x; y)$ двух



аргументов, равная вероятности совместного выполнения двух неравенств $X < x$ и $Y < y$, то есть $F(x, y) = P\{X < x, Y < y\}$.

Геометрически функцию распределения системы двух случайных величин можно интерпретировать как вероятность попадания случайной точки (X, Y) в левый нижний бесконечный квадрант с вершиной в точке (x, y) плоскости (см. рис.).

Сформулируем основные свойства функции распределения вероятностей системы двух случайных величин (без доказательства):

$$1. \quad 0 \leq F(x, y) \leq 1, \quad \forall x, y$$

$$2. \quad \lim_{y \rightarrow +\infty} F(x, y) = F_1(x); \quad \lim_{x \rightarrow +\infty} F(x, y) = F_2(y) \quad (\text{или } F(x, +\infty) = F_1(x); F(+\infty, y) = F_2(y)).$$

$$3. \quad \lim_{x \rightarrow +\infty} F(x, y) = 1 \quad (\text{или } F(+\infty, +\infty) = 1).$$

$$4. \quad \lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = \lim_{x \rightarrow -\infty} F(x, y) = 0 \quad (F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0).$$

5. Функция распределения является неубывающей функцией по каждому из своих аргументов, то есть:

$$F(x_2, y) \geq F(x_1, y), \quad \text{если } x_2 > x_1;$$

$$F(x, y_2) \geq F(x, y_1), \quad \text{если } y_2 > y_1.$$

6. Вероятность попадания случайной точки (X, Y) в произвольный прямоугольник со сторонами, параллельными координатным осям (см. рис.) вычисляется по формуле:

$$P(x_1 < X < x_2, y_1 < Y < y_2) = F(x_1, y_1) + F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1),$$

где $x_1 < x_2, y_1 < y_2$.

Плотность распределения вероятностей системы двух случайных величин¹

Предположим, что функция распределения $F(x, y)$ всюду непрерывна и дважды дифференцируема² (за исключением, быть может, конечного числа кривых). Тогда, смешанная частная производная функции

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = F''_{xy}(x, y)$$

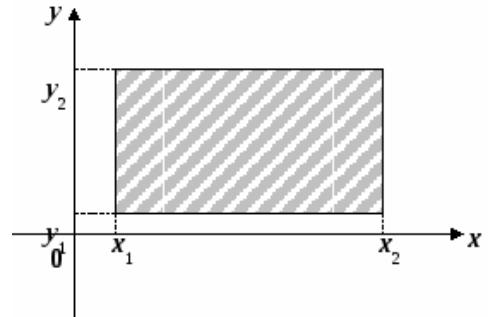
Функция $f(x, y)$ называется плотностью распределения (или, дифференциальной функцией распределения) системы непрерывных случайных величин.

Геометрически эту функцию можно истолковать как поверхность, которую называют *поверхностью распределения*.

Зная плотность распределения, можно определить вероятность попадания случайной точки в произвольную область D :

$$P\{(X, Y) \subset D\} = \iint_D f(x, y) dx dy$$

Используя последнюю формулу, выразим интегральную функцию распределения вероятностей системы двух непрерывных случайных величин через плотность распределения:



$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy$$

Рассмотрим некоторые свойства плотности распределения системы двух непрерывных случайных величин:

$$1. f(x, y) \geq 0, \quad \forall (-\infty < x, y < +\infty).$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$$

2. $\int \int f(x, y) dx dy = 1$, если случайная величина распределена на всей координатной плоскости (если же распределена в некоторой плоской области G , то $\int \int f(x, y) dx dy = 1$)

Условные законы распределения

Пусть известна плотность распределения системы двух случайных величин. Используя свойства функций распределения, можно вывести формулы для нахождения плотности распределения одной величины, входящей в систему:

$$f_1(x) = F_1(x) = \int_{-\infty}^{+\infty} f(x, y) dy; \quad f_2(y) = F_2(y) = \int_{-\infty}^{+\infty} f(x, y) dx \quad (*)$$

Перейдём теперь к решению обратной задачи: по известным законам распределения отдельных случайных величин, входящих в систему, найти закон распределения системы. Легко увидеть, что в общем случае эта задача неразрешима. Действительно, с одной стороны, законы распределения отдельных случайных величин, входящих в систему, характеризуют каждую из случайных величин в отдельности, но ничего не говорят о том, как они взаимосвязаны. С другой стороны, искомый закон распределения системы должен содержать все сведения о случайных величинах системы, в том числе и о характере связей между ними.

Таким образом, если случайные величины X, Y взаимозависимы, то закон распределения системы не может быть выражен через законы распределения отдельных случайных величин, входящих в систему. Это приводит к необходимости введения условных законов распределения.

Распределение одной случайной величины, входящей в систему, найденное при условии, что другая случайная величина, входящая в систему, приняла определённое значение, называется условным законом распределения.

Для дискретных случайных величин условным распределением составляющей при условии, что $Y = y_j$ называется совокупность условных вероятностей $p(x_1 | y_j), p(x_2 | y_j), \dots, p(x_n | y_j)$, вычисленных в предположении, что случайная величина уже приняла значение y_j . Для нахождения $p(x_i | y_j)$ пользуются формулой

$$p(x_i | y_j) = \frac{p(x_i, y_j)}{p(y_j)}, \quad i = \overline{1, n}$$

$$\sum_{i=1}^n p(x_i | y_j) = 1$$

Заметим, что

$$p(y_j | x_i) = \frac{p(x_i, y_j)}{p(x_i)}, \quad j = \overline{1, m}$$

Аналогично находим

Условный закон распределения можно задавать как функцией распределения, так и плотностью распределения. Условная функция распределения обозначается $F(x | y)$; условная плотность распределения обозначается $f(x | y)$.

Плотностью распределения для случайной величины при условии, что случайная величина приняла определённое значение (*условной плотностью распределения*), назовём величину

$$f(x|y) = \frac{f(x,y)}{f_2(y)}$$

Аналогично, плотностью распределения для случайной величины при условии, что случайная величина приняла определённое значение, назовём величину

$$f(y|x) = \frac{f(x,y)}{f_1(x)}$$

Отсюда получаем: $f(x,y) = f_1(x)f(y|x) = f_2(y)f(x|y)$

$$f(x|y) = \frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dx}; \quad f(y|x) = \frac{f(x,y)}{\int_{-\infty}^{+\infty} f(x,y)dy}$$

или, с учётом формул (*)

Условная плотность распределения обладает всеми свойствами безусловной

$$\int_{-\infty}^{+\infty} f(x|y)dx = 1; \quad \int_{-\infty}^{+\infty} f(y|x)dy = 1$$

плотности распределения. В частности,

Числовые характеристики условных законов распределения

Для описания условных законов распределения можно использовать различные характеристики подобно тому, как для одномерных распределений.

Наиболее важной характеристикой является условное математическое ожидание.

Условным математическим ожиданием дискретной случайной величины при $Y = y$ (y – определённое возможное значение случайной величины) называется сумма произведений возможных значений на их условные вероятности:

$$M(X|Y) = \sum_{i=1}^n x_i P(x_i|y)$$

Для непрерывных случайных величин:

$$M(X|Y) = \int_{-\infty}^{+\infty} xf(x|y)dx$$

где – условная плотность распределения случайной величины при .

Аналогично, условным математическим ожиданием дискретной случайной величины при $X = x$ (x – определённое возможное значение случайной величины) называется сумма произведений возможных значений на их условные вероятности:

$$M(Y|X) = \sum_{j=1}^m y_j P(y_j|x)$$

$$M(Y|X) = \int_{-\infty}^{+\infty} yf(y|x)dy$$

Для непрерывных случайных величин:

где $f(y|x)$ – условная плотность распределения случайной величины при .

Аналогично вводятся условные дисперсии и условные моменты более высоких порядков.

Числовые характеристики системы двух случайных величин

Две случайные величины называются независимыми, если закон распределения одной из них не зависит от того, какие возможные значения приняла другая величина. Из этого определения следует, что условные распределения независимых случайных величин равны их безусловным распределениям. Укажем необходимые и достаточные условия независимости случайных величин.

ТЕОРЕМА 1: Для того чтобы случайные величины X и Y были независимыми, необходимо и достаточно, чтобы функция распределения системы (X, Y) была равна произведению функций распределения составляющих:

$$F(x, y) = F_1(x) \cdot F_2(y)$$

ТЕОРЕМА 2: Для того чтобы случайные величины и были независимыми, необходимо и достаточно, чтобы плотность вероятности системы была равна произведению плотностей вероятностей составляющих:

$$f(x, y) = f_1(x) \cdot f_2(y)$$

Для описания системы двух случайных величин кроме математических ожиданий и дисперсий составляющих используют и другие характеристики, к которым относятся корреляционный момент и коэффициент корреляции.

Корреляционным моментом μ_{xy} случайных величин и называют математическое ожидание произведения отклонений этих величин:

$$\mu_{xy} = M((X - M(X)) \cdot (Y - M(Y)))$$

Для вычисления корреляционного момента дискретных величин используют

$$\mu_{xy} = \sum_{i=1}^n \sum_{j=1}^m (x_i - M(X))(y_j - M(Y)) P(x_i, y_j)$$

формулу:

$$\mu_{xy} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - M(X))(y - M(Y)) f(x, y) dx dy$$

а для непрерывных величин:

Корреляционный момент служит для характеристики связи между величинами и .

ТЕОРЕМА 3: Корреляционный момент двух независимых случайных величин и равен нулю.

Замечание: из теоремы 3 следует, что если корреляционный момент двух случайных величин и не равен нулю, то и – зависимые случайные величины.

Коэффициентом корреляции r_{xy} случайных величин и называют отношение корреляционного момента к произведению средних квадратических отклонений этих величин:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}$$

Очевидно, коэффициент корреляции двух независимых случайных величин равен нулю (так как $\mu_{xy} = 0$).

2. Функция регрессии, коэффициент детерминации, корреляции, ковариация

Условимся обозначать через X независимую переменную, а через Y – зависимую переменную.

Зависимость величины Y от X называется **функциональной**, если каждому значению величины X соответствует единственное значение величины Y . С функциональной зависимостью мы встречаемся, например, в математике, при изучении физических законов. Обратим внимание на то, что если X – детерминированная величина (т.е. принимающая вполне определённые значения), то и функционально зависящая от неё величина Y тоже является детерминированной; если же X – случайная величина, то и Y также случайная величина.

Однако гораздо чаще в окружающем нас мире имеет место не функциональная, а **стохастическая, или вероятностная, зависимость**, когда каждому фиксированному значению независимой переменной X соответствует не одно, а множество значений переменной Y , причём сказать заранее, какое именно значение примет величина Y , нельзя. Более частое появление такой зависимости объясняется действием на результирующую переменную не только контролируемого или контролируемых факторов (в данном случае таким контролируемым фактором является переменная X), а и многочисленных

неконтролируемых случайный факторов. В этой ситуации переменная Y является случайной величиной. Переменная же X может быть как детерминированной, так и случайной величиной. Следует заметить, что со стохастической зависимостью мы уже сталкивались в дисперсионном анализе.

Допустим, что существует стохастическая зависимость случайной переменной Y от X . Зафиксируем некоторое значение x переменной X . При $X = x$ переменная Y в силу её стохастической зависимости от X может принять любое значение из некоторого множества, причём какое именно – заранее неизвестно. Среднее этого множества называют **групповым генеральным средним** переменной Y при $X = x$ или **математическим ожиданием случайной величины Y , вычисленным при условии, что $X = x$** ; это **условное математическое ожидание обозначают так: $M(Y/X = x)$** . Если существует стохастическая зависимость Y от X , то прежде всего стараются выяснить, изменяются или нет при изменении x условные математические ожидания $M(Y/X=x)$. Если при изменении x условные математические ожидания $M(Y/X=x)$ изменяются, то говорят, что имеет место **корреляционная зависимость** величины Y от X ; если же условные математические ожидания остаются неизменными, то говорят, что корреляционная зависимость величины Y от X отсутствует.

Функция $\phi(x)=M(Y/X=x)$, описывающая изменение условного математического ожидания случайной переменной Y при изменении значений x переменной X , называется **функцией регрессии**.

Выясним, почему именно при наличии стохастической зависимости интересуются поведением условного математического ожидания.

Рассмотрим пример. Пусть X – уровень квалификации рабочего, Y – его выработка за смену. Ясно, что зависимость Y от X не функциональная, а стохастическая: на выработку помимо квалификации влияет множество других факторов. Зафиксируем значение x уровня квалификации: ему соответствует некоторое множество значений выработки Y . Тогда $M(Y/X = x)$ – средняя выработка рабочего при условии, что его уровень квалификации равен x , или, иначе говоря, $M(Y/X = x)$ – это норматив выработки при уровне квалификации, равном x . Зная зависимость этого норматива от уровня квалификации, можно для любого уровня квалификации рассчитать норматив выработки и, сравнив его с реальной выработкой, оценить работу рабочего.

Обратим внимание на то, что введённые понятия стохастической и корреляционной зависимости относились к генеральной совокупности. Поясним эти понятия числовым примером.

Пример. Допустим, что одновременно изучаются две случайные величины X и Y , или, иначе говоря, двумерная случайная величина (X, Y) , которая задана табл. 1.

Таблица 1.

j	i	1	2	3
		x_i	$x_1 = 2$	$x_2 = 5$
y_i	$y_1 = 0,4$	0,15	0,12	0,03
	$y_2 = 0,8$	0,05	0,30	0,35

Табл. 1 называют **таблицей распределения двумерной величины (X, Y)** ; её следует понимать так. Случайная величина X может принять одно из следующих значений: 2, 5 и 8. Случайная величина Y – значения 0,4 и 0,8. Число 0,15 – это вероятность того, что $X = 2$ и одновременно $Y = 0,4$, или, иначе говоря, вероятность произведения двух событий; события, состоящего в том, что $X = 2$, и события, состоящего в том, что $Y = 0,4$, т.е. $P((X=2)(Y=0,4)) = 0,15$. Аналогично, вероятность $P((X=2)(Y=0,8)) = 0,05$ и т.д. Обратим

внимание на следующее: поскольку в табл. 1 указаны все возможные значения величин X и Y, сумма вероятностей, стоящих в таблице, должна быть равна единице: $0,15 + 0,05 + 0,12 + 0,30 + 0,03 + 0,35 = 1$.

Прежде чем выяснить тип зависимости величины Y от X, найдём:

а) Закон распределения величины X. Он представлен табл. 2.

Таблица 2.

x	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$	
$P(X=x)$	$0,15 + 0,05 = 0,2$	$0,12 + 0,30 = 0,42$	$0,35 + 0,03 = 0,38$	$\Sigma = 1$

$$M(X) = 5,54, D(X) = 4,9284$$

Действительно, например, величина X примет значение, равное 2, только в том случае, когда одновременно с этим величина Y примет значение 0,4 или 0,8, т.е.

$$P(X=2) = P((X=2)(Y=0,4)) + P((X=2)(Y=0,08)) = 0,15 + 0,05 = 0,2.$$

Справа от ряда распределения величины X находятся её математическое ожидание и дисперсия.

б) Закон распределения величины Y. Он имеет вид табл. 3.

Таблица 3.

y	$y_1 = 0,4$	$y_2 = 0,8$	
$P(Y=y)$	$0,15 + 0,12 + 0,03 = 0,30$	$0,05 + 0,30 + 0,35 = 0,7$	$\Sigma = 1$

$$M(Y) = 0,68, D(Y) = 0,0336$$

в) Условные законы распределения величины Y, а именно закон распределения величины Y сначала при условии, что $X = 2$, затем при условии, что $X=5$, и наконец, при условии, что $X = 8$.

Итак, пусть $X = 2$. Тогда условная вероятность

$$P(Y = 0,4/X = 2) = \frac{P(Y = 0,4)(X = 2))}{P(X = 2)} = \frac{0,15}{0,2} = 0,75,$$

а условная вероятность

$$P(Y = 0,8/X = 2) = \frac{P((Y = 0,8)(X = 2))}{P(X = 2)} = \frac{0,05}{0,2} = 0,25.$$

Таким образом, закон распределения величины Y при условии, что $X = 2$, задан табл. 4.

Таблица 4.

y	$y_1 = 0,4$	$y_2 = 0,8$	
$P(Y=y/X = 2)$	0,75	0,25	$\Sigma = 1$

$$M(Y/X = 2) = 0,4*0,75 + 0,8*0,25 = 0,5, D(Y/X = 2) = 0,03$$

Справа помещено условное математическое ожидание и значение условной дисперсии. Покажем, как вычисляется условная дисперсия. Общая формула условной дисперсии имеет вид

$$D(Y/X = x) = M[(Y/X = x) - M(Y/X = x)]^2. \quad (23)$$

Для табл. 4 получаем

$$D(Y/X = 2) = M [(Y/X = 2) - M(Y/X = 2)]^2 = M [(Y/X = 2) - 0,5]^2 = \\ \sum_{i=1}^2 (y_i - 0,5)^2 \cdot P(Y = y_i/X = 2) = (0,4 - 0,5)^2 \cdot 0,75 + (0,8 - 0,5)^2 \cdot 0,25 = 0,03.$$

Пусть $X = 5$. Тогда $P(Y = 0,4/X = 5) = \frac{P((Y=0,4)(X=5))}{P(X=5)} = \frac{0,12}{0,42} = \frac{2}{7}$; $P(Y=0,8/X=5) =$

$$\frac{P((Y=0,8)(X=5))}{P(X=5)} = \frac{0,30}{0,42} = .$$

Таким образом, закон распределения величины Y при условии, что $X = 5$, имеет вид табл. 5.

Таблица 5.

y	0,4	0,8	
$P(Y = y/X = 5)$	2/7	5/7	$\Sigma = 1$

$$M(Y/X = 5) = \frac{24}{35} \approx 0,686, \quad D(Y/X = 5) = 0,03265.$$

И наконец, при $X = 8$ ряд распределения задан таблицей 14.

Таблица 6.

y	0,4	0,8	
$P(Y = y/X = 8)$		$\frac{35}{38}$	$\Sigma = 1$

$$M(Y/X = 8) = \frac{73}{95} \approx 0,768, \quad D(Y/X = 8) = 0,01163$$

Из табл. 4–6 видно, что зависимость Y от X стохастическая, поскольку при каждом фиксированном значении величины X величина Y может быть равной либо 0,4, либо 0,8, причём какому именно из этих чисел она будет равна – сказать заранее нельзя. Ясно прослеживается и корреляционная зависимость величины Y от X , поскольку с изменением значений x величины X меняются и условные математические ожидания $M(Y/X = x)$. Функция регрессии, т.е. зависимость условного математического ожидания $M(Y/X = x)$ от x , задаётся в виде табл. 7.

Таблица 7.

x	2	5	8
$M(Y/X = x)$	0,5	$24/35 \approx 0,686$	$73/95 \approx 0,768$

Выборочный коэффициент корреляции

Выясним, можно ли измерить степень корреляционной и стохастической зависимости величины Y от X . Ответ проиллюстрируем примером 5. Все полученные в примере результаты объединены в табл. 8.

Таблица 8.

x_i	$x_1 = 2$	$x_2 = 5$	$x_3 = 8$
$P(X = x_i)$	0,2	0,42	0,38
$M(Y/X = x_i)$	0,5	0,686	0,768
$D(Y/X = x_i)$	0,03	0,03265	0,01163

$$MY = 0,68 \text{ (см. табл. 3)}, \quad DY = 0,0336 \text{ (см. табл. 3)}$$

Т.к. X – случайная величина, принимающая значения 2, 5 и 8 с вероятностью 0,2; 0,42 и 0,38, то такими же будут вероятности и условных математических ожиданий, и дисперсий. Т.обр., условное математическое ожидание $M(Y/X)$, так же как и условная дисперсия $D(Y/X)$ – случайные величины.

Обратим также внимание на то, что $M(Y)$, найденное в табл. 3, можно вычислить и по табл. 8 следующим образом:

$$M(Y) = M[M(Y/X)] = \sum_{i=1}^3 M(Y/X=x_i)P(X=x_i) = 0,5*0,2 + 0,686*0,42 + 0,768*0,38 = 0,68.$$

Разброс значений величины Y вокруг математического ожидания MY измеряется дисперсией $D(Y)$, или σ_Y^2 :

$$\sigma_Y^2 = D(Y) = M(Y - MY)^2. \quad (24)$$

(По табл. 3: $\sigma_Y^2 = 0,0336$.) Этот разброс может быть вызван:

- зависимостью величины Y от X (эта зависимость может быть обусловлена не только непосредственным влиянием X на Y , но и наличием случайных факторов, действующих на Y через переменную X);

- зависимостью величины Y от случайных факторов, влияющих только на Y и не влияющих на X ; эти факторы называют **остаточными**.

1) Построим показатель разброса значений величины Y , связанного с её зависимостью от фактора X .

Условное математическое ожидание $M(Y/X = x)$ является «представителем игреков», которые имеют место при $X = x$. Характеристикой разброса условных математических ожиданий $M(Y/X = x)$ относительно $M(Y)$ является дисперсия $D[M(Y/X)]$, или

$$\sigma_\varphi^2 = D[M(Y/X)] = M[M(Y/X) - MY]^2 \quad (25)$$

- эта величина и будет показателем разброса значений величины Y , связанного с её зависимостью от фактора X . По таблице 16 найдём:

$$\sigma_\varphi^2 = M[M(Y/X) - MY]^2 = (0,5 - 0,68)^2 * 0,2 + (0,686 - 0,68)^2 * 0,42 + (0,768 - 0,68)^2 * 0,38 = 0,0095.)$$

2) Теперь построим показатель разброса «игреков», связанного с влиянием остаточных факторов.

Зафиксируем какое-либо значение x величины X . Тогда причиной вариации величины Y при $X = x$ будут остаточные факторы, влияющие только на Y и не влияющие на X . Измерителем этой вариации является условная дисперсия $D(Y/X = x)$. При различных же «иксах» характеристикой разброса «игреков», вызванного влиянием на Y остаточных факторов, будет генеральное среднее из условных дисперсий, или, иначе, математическое ожидание условной дисперсии. Эту величину обозначим σ_0^2 . Имеем

$$\sigma_0^2 = M[D(Y/X)], \quad (26)$$

где при $X = x$ дисперсия $D(Y/X = x)$ вычисляется по формуле (23). (По табл. 16 найдём

$$\sigma_0^2 = M[D(Y/X)] = \sum_{i=1}^3 D(Y/X=x_i)P(X=x_i) = 0,03 * 0,2 + 0,03265 * 0,42 + 0,01163 * 0,38 = 0,0241.)$$

Для вычисленных дисперсий справедливо тождество

$$DY = D[M(Y/X)] + M[D(Y/X)]$$

или

$$\sigma_Y^2 = \sigma_\varphi^2 + \sigma_0^2.$$

Степень стохастической зависимости величины Y от X измеряется **генеральным корреляционным отношением**

$$\rho_{Y/X} = + \sqrt{\frac{D[M(Y/X)]}{DY}} = + \sqrt{\frac{\sigma_\varphi^2}{\sigma_Y^2}} = + \sqrt{1 - \frac{\sigma_0^2}{\sigma_Y^2}} = + \sqrt{1 - \frac{M[D(Y/X)]}{DY}}. \quad (28)$$

Квадрат корреляционного отношения

$$\rho_{Y/X}^2 = \frac{\sigma_\varphi^2}{\sigma_Y^2} = \frac{D[M(Y/X)]}{DY} \stackrel{(26), (25)}{=} \frac{M[M(Y/X) - MY]^2}{M(Y - MY)^2} \quad (29)$$

называется **генеральным коэффициентом детерминации**; он показывает, какую долю дисперсии величины Y составляет дисперсия условных математических ожиданий, или, иначе говоря, какая доля дисперсии D(Y) объясняется корреляционной зависимостью Y от X.

(В примере $\sigma_\varphi^2 = 0,0095$, $\sigma_Y^2 = 0,0336$, поэтому $\rho_{Y/X}^2 = \frac{0,0095}{0,0336} = 0,28$, т.е. 28% дисперсии величины Y объясняется её корреляционной зависимостью от X; $\rho_{Y/X} = + \sqrt{0,28} = 0,53$.)

Свойства генерального корреляционного отношения как измерителя степени корреляционной и стохастической зависимости

1. $0 \leq \rho_{Y/X} \leq 1$.

Действительно, согласно (29), $\rho_{Y/X} \geq 0$; с другой стороны, из (28) следует, что $\sigma_\varphi^2 \leq \sigma_Y^2$, поэтому $\rho_{Y/X} \leq 1$.

2. Условие $\rho_{Y/X} = 0$ является необходимым и достаточным для отсутствия корреляционной зависимости Y от X, т.е. для того, чтобы $M(Y/X) = \text{const}$ при любом значении x величины X.

3. Условие $\rho_{Y/X} = 1$ является необходимым и достаточным для функциональной зависимости величины Y от X.

Следствие. Чем ближе $\rho_{Y/X}$ к единице, тем в силу (28) ближе к нулю $M[D(Y/X)]$, а следовательно, и условные дисперсии $D(Y/X = x)$. Это означает, что при каждом допустимом значении x уменьшается разброс «игреков» относительно $M(Y/X = x)$. Т.о.бр., чем ближе $\rho_{Y/X}$ к единице, тем меньше при каждом x отличие «игреков» от постоянного числа, равного $M(Y/X = x)$, или, иначе говоря, тем выше степень стохастической зависимости Y от X. И, наоборот, чем выше степень стохастической зависимости Y от X, тем ближе $\rho_{Y/X}$ к единице.

В практических задачах наибольший интерес представляют следующие вопросы:

- существует корреляционная зависимость Y от X или нет, иначе говоря, отлично ли генеральное корреляционное отношение $\rho_{Y/X}$ от нуля или равно нулю;
- если корреляционная зависимость существует, то какой вид имеет функция регрессии (линейный, параболический или какой-либо другой).

Точно ответить на поставленные вопросы можно лишь только в том случае, когда известен закон распределения двумерной величины (X, Y). В примере 5 этот закон задан табл. 9, в которой даны все возможные значения случайных величин X и Y и вероятности совместного появления этих значений. Обычно такими сведениями не располагают; как правило, имеются лишь наблюдавшиеся значения двумерной величины (X, Y). Покажем как, имея наблюдавшиеся значения, ответить на поставленные выше вопросы.

Выборочное корреляционное отношение. Его значимость

Пусть имеется n наблюдений двумерной величины (X, Y). Наблюдавшиеся «иксы» и «игреки» поместим в табл. 17, которая называется **корреляционной таблицей** и строится следующим образом:

- «иксы» группируются в вариационный ряд, число групп которого обозначим v ; если это дискретный ряд, то x_1, x_2, \dots, x_v – различающиеся между собой результаты наблюдений или варианты; если это интервальный ряд, то x_1, x_2, \dots, x_v – центры интервалов;

- «игреки» группируют в вариационный ряд, число групп которого обозначим q : y_1, y_2, \dots, y_q – это либо варианты, если ряд дискретный, либо середины интервалов, если ряд интервальный;

- подсчитывают числа m_{ji} таких наблюдавшихся пар чисел (x, y) , у которых x попадает в группу x_i , а y – в группу y_j , $i = 1, 2, \dots, v$, $j = 1, 2, \dots, q$; например, m_{12} – число пар чисел (x, y) , у которых x попало в группу x_2 , а y – в группу y_1 . Числа $m_{11}, m_{12}, \dots, m_{qv}$ называются **частотами**.

Прежде чем пояснить остальные элементы этой таблицы, сделаем следующее замечание по поводу схемы построения выборочного корреляционного отношения: от табл. 17, содержащей частоты, можно перейти к таблице частостей (табл. 18). Сравним табл. 18 и табл. 9. Их различие состоит в следующем: табл. 9 относилась к генеральной совокупности, поэтому в ней были указаны все мыслимые значения величин X и Y и вероятности комбинаций этих значений; табл. 18 относится к выборочной совокупности, и в ней приведены наблюдаемые значения величин X и Y и частости, или опытные вероятности комбинаций наблюдаемых значений. Поэтому выборочное корреляционное отношение можно строить по той же схеме, что и генеральное корреляционное отношение, если заменить возможные значения величин X и Y на наблюдаемые, вероятности на частости, математические ожидания на средние, дисперсии на выборочные дисперсии.

Однако чаще выборочное корреляционное отношение строят, используя непосредственно табл. 9, а не табл. 1. В табл. 9 кроме сгруппированных наблюдений и частот содержатся следующие данные:

$$- \text{суммы частот по каждой строке} \quad m_1 = \sum_{i=1}^v m_{1i}, m_2 = \sum_{i=1}^v m_{2i}, \dots, m_q = \sum_{i=1}^v m_{qi}, \quad (30)$$

$$- \text{суммы частот по каждому столбцу} \quad n_1 = \sum_{j=1}^q m_{j1}, n_2 = \sum_{j=1}^q m_{j2}, \dots, n_v = \sum_{j=1}^q m_{jv}, \quad (31)$$

Таблица 9.

j	i	1	2	...	v	
	X	x_1	x_2	...	x_v	Σ
1	y_1	m_{11}	m_{12}	...	m_{1v}	m_1
2...	$y_2\dots$	$m_{21\dots}$	$m_{22\dots}$...	$m_{2v\dots}$	$m_{2\dots}$
q	y_q	m_{q1}	m_{q2}	...	m_{qv}	m_q
Σ		n_1	n_2	...	n_v	$n = \sum_{i=1}^v n_i = \sum_{j=1}^q m_j$
Групповое среднее		$\bar{Y}^{(1)}$	$\bar{Y}^{(2)}$...	$\bar{Y}^{(v)}$	
Групповая выборочная дисперсия		$\sigma^*_{11}^2$	$\sigma^*_{12}^2$...	$\sigma^*_{vv}^2$	

Таблица 10

j	i	1	2	...	v	
	X	x_1	x_2	...	x_v	
1	y_1	p_{11}^*	p_{12}^*	...	p_{1v}^*	
2...	$y_2\dots$	$p_{21\dots}^*$	$p_{22\dots}^*$...	$p_{2v\dots}^*$	$p_{ij}^* = m_{ij}/n, i = 1, 2, \dots, v$

					$j = 1, 2, \dots, q.$
q	y_q	p_{q1}^*	p_{q2}^*	\dots	p_{qv}^*

- групповые средние значения «игреков»

$$\begin{aligned}\bar{Y}^{(1)} &= (y_1 m_{11} + y_2 m_{21} + \dots + y_q m_{q1}) / n_1, \\ \bar{Y}^{(2)} &= (y_1 m_{12} + y_2 m_{22} + \dots + y_q m_{q2}) / n_2, \\ &\dots \\ \bar{Y}^{(v)} &= (y_1 m_{1v} + y_2 m_{2v} + \dots + y_q m_{qv}) / n_v.\end{aligned}\quad (32)$$

Эти средние являются выборочными аналогами соответствующих условных математических ожиданий: $\bar{Y}^{(1)}$ - выборочный аналог математического ожидания величины Y при условии, что $X = x_1$, т.е. аналог величины $M(Y/X=x_1)$; $\bar{Y}^{(2)}$ - аналог $M(Y/X=x_2)$ и т.д.

Групповые выборочные дисперсии:

$$\begin{aligned}\hat{\sigma}_1^2 &= \frac{\left[(y_1 - \bar{Y}^{(1)})^2 m_{11} + (y_2 - \bar{Y}^{(1)})^2 m_{21} + \dots + (y_q - \bar{Y}^{(1)})^2 m_{q1} \right]}{n_1}, \\ \hat{\sigma}_2^2 &= \frac{\left[(y_1 - \bar{Y}^{(2)})^2 m_{12} + (y_2 - \bar{Y}^{(2)})^2 m_{22} + \dots + (y_q - \bar{Y}^{(2)})^2 m_{q2} \right]}{n_2}, \\ &\dots \\ \hat{\sigma}_v^2 &= \left[(y_1 - \bar{Y}^{(v)})^2 m_{1v} + (y_2 - \bar{Y}^{(v)})^2 m_{2v} + \dots + (y_q - \bar{Y}^{(v)})^2 m_{qv} \right].\end{aligned}\quad (33)$$

Эти дисперсии являются выборочными аналогами соответствующих условных дисперсий: $\hat{\sigma}_1^2$ – выборочный аналог условной дисперсии $D(Y/X=x_1)$, $\hat{\sigma}_2^2$ – аналог $D(Y/X=x_2)$ и т.д.

Построим выборочный аналог генерального корреляционного отношения.

Выборочным аналогом генеральной дисперсии $\sigma_y^2 = DY$ является величины $\hat{\sigma}_y^2$. Для того чтобы вычислить $\hat{\sigma}_y^2$, найдем сначала среднее \bar{Y} по данным табл. 17. Это можно сделать по одной из следующих тождественных формул:

$$\bar{Y} = \frac{1}{n} (y_1 m_1 + y_2 m_2 + \dots + y_q m_q) = \frac{1}{n} (\bar{Y}^{(1)} n_1 + \bar{Y}^{(2)} n_2 + \dots + \bar{Y}^{(v)} n_v) \quad (34)$$

Напомним, что \bar{Y} – это выборочный аналог $M(Y)$. Теперь найдем

$$S_Y^2 = (y_1 - \bar{Y})^2 m_1 + (y_2 - \bar{Y})^2 m_2 + \dots + (y_q - \bar{Y})^2 m_q = \sum_{j=1}^q (y_j - \bar{Y})^2 m_j \quad (35)$$

Тогда

$$\hat{\sigma}_y^2 = \frac{S_Y^2}{n} \quad (36)$$

Выборочным аналогом генеральной дисперсии $\sigma_\Phi^2 = D \left[M \left(\frac{Y}{X} \right) \right]$ является $\hat{\sigma}_\Phi^2$

выборочная дисперсия $\hat{\sigma}_y^2$ групповых средних; обозначим ее $\hat{\sigma}_\Phi^2$. Имеем

$$\hat{\sigma}_{\Phi}^2 = \hat{\sigma}_{Y_{(1)}}^2 = \frac{1}{n} \left[(\bar{Y}^{(1)} - \bar{Y})^2 n_1 + (\bar{Y}^{(2)} - \bar{Y})^2 n_2 + \dots + (\bar{Y}^{(v)} - \bar{Y})^2 n_v \right] = \frac{s_{\Phi}^2}{n}$$

,(37)

где

$$s_{\Phi}^2 = (\bar{Y}^{(1)} - \bar{Y})^2 n_1 + (\bar{Y}^{(2)} - \bar{Y})^2 n_2 + \dots + (\bar{Y}^{(v)} - \bar{Y})^2 n_v = \\ = \sum_{i=1}^v (\bar{Y}^{(i)} - \bar{Y})^2 n_i$$

(38)

Выборочным аналогом дисперсии $\sigma_o^2 = D \left[M \left(\frac{Y}{X} \right) \right]$ является средняя групповых выборочных дисперсий. Обозначим $\hat{\sigma}_o^2$ эту среднюю

$$\hat{\sigma}_o^2 = \bar{\sigma}_i^2 = \frac{1}{n} (\hat{\sigma}_1^2 n_1 + \hat{\sigma}_2^2 n_2 + \dots + \hat{\sigma}_v^2 n_v) = \frac{s_o^2}{n},$$

(39)

где $s_o^2 = \hat{\sigma}_1^2 n_1 + \hat{\sigma}_2^2 n_2 + \dots + \hat{\sigma}_v^2 n_v$.

Получаем

$$s_o^2 = (y_1 - \bar{Y}^{(1)})^2 m_{11} + (y_2 - \bar{Y}^{(1)})^2 m_{21} + \dots + (y_q - \bar{Y}^{(1)})^2 m_{q1} + (y_1 - \bar{Y}^{(2)})^2 m_{12} + \\ + (y_2 - \bar{Y}^{(2)})^2 m_{22} + \dots + (y_q - \bar{Y}^{(2)})^2 m_{q2} + \dots + (y_1 - \bar{Y}^{(v)})^2 m_{1v} + (y_2 - \bar{Y}^{(v)})^2 m_{2v} + \dots \\ + (y_q - \bar{Y}^{(v)})^2 m_{qv} = \sum_{i=1}^v \sum_{j=1}^q (y_j - \bar{Y}^{(i)})^2 m_{ji}$$

(40)

Для вычисленных дисперсий справедливо тождество

$$\hat{\sigma}_Y^2 = \sigma_{\Phi}^2 + \hat{\sigma}_o^2$$

$$\hat{\sigma}_Y^2 = \hat{\sigma}_{Y_{(1)}}^2 + \bar{\sigma}_i^2, \text{ или}$$

(41)

аналогичное тождеству, имеющему место в генеральной совокупности.

Выборочный аналог генерального корреляционного отношения вычисляется следующим образом:

$$\hat{\rho}_{\frac{Y}{X}} = + \sqrt{\frac{\hat{\sigma}_{\bar{Y}^{(1)}}^2}{\hat{\sigma}_Y^2}} = + \sqrt{\frac{\hat{\sigma}_{\Phi}^2}{\hat{\sigma}_Y^2}} = + \sqrt{1 - \frac{\hat{\sigma}_o^2}{\hat{\sigma}_Y^2}} = + \sqrt{1 - \frac{\hat{\sigma}_1^2}{\hat{\sigma}_Y^2}}$$

(42)

Величина $\hat{\rho}_{\frac{Y}{X}}$ называется выборочным коэффициентом детерминации. Этот коэффициент показывает, какую долю дисперсии $\hat{\sigma}_Y^2$ составляет выборочная дисперсия групповых средних «игреков» или, иначе говоря, какая доля дисперсии $\hat{\sigma}_Y^2$ объясняется зависимостью Y от X .

$$\hat{\sigma}_{\Phi}$$

Как правило, дисперсии $\hat{\sigma}_Y^2$ и $\hat{\sigma}_{\Phi}^2$ находят не по рассмотренным выше формулам, а по следующим, более удобным для вычислений:

$$\hat{\sigma}_Y^2 = \frac{1}{n} \sum_{j=1}^q y_j^2 m_j - (\bar{Y})^2, \quad S_y^2 = \hat{\sigma}_Y^2 n$$

(43)

$$\hat{\sigma}_i^2 = \frac{1}{ni} \sum_{j=1}^q y_j^2 m_{ji} - (\bar{Y}^{(i)})^2, \quad S_y^2 = \hat{\sigma}_Y^2 n$$

(44)

$$\hat{\sigma}_{\Phi}^2 = \frac{1}{n} \sum_{i=1}^v (\bar{Y}^{(i)})^2 n_i - \bar{Y}^2 , \quad S_{\Phi}^2 = \hat{\sigma}_{\Phi}^2 n . \quad (45)$$

$$H_0: P_{\frac{Y}{X}} = 0 \quad (46)$$

Проверим гипотезу

Предварительно отметим, что в силу свойства 2 корреляционного отношения при выполнении гипотезы имеет место равенство условных, или групповых математических ожиданий величины Y : $M(Y/X=x_1)=M(Y/X=x_2)= \dots =M(Y/X=x_v)$.

Поэтому проверка гипотезы сводится к проверке гипотезы о равенстве групповых математических ожиданий – это задача дисперсионного анализа. Ее можно решить, если выполняются требования, применительно к нашим условиям формулирующиеся следующим образом (*):

- при каждом наблюдаемом значении x_i величины X наблюдения величины Y должны быть независимы и проводиться в одинаковых условиях; наблюдения должны быть независимы и при различных «иксах»;
- при каждом значении x_i величина Y должна иметь нормальный закон с постоянной для различных «иксов» генеральной дисперсией (обозначим эту дисперсию σ_o^2);

$$\sigma_{10}^2 = D(Y/X = x_1) = (Y/X = x_1)^2 = \dots = (Y/X = x_v)$$

Допустим, что эти требования выполняются. Тогда для проверки гипотезы (46) следует заполнить табл. 19.

В заключение заметим, что если гипотеза отвергается, то говорят, что выборочное корреляционное отношение статистически значимо. Если гипотеза не отвергается, то говорят, что выборочное корреляционное отношение незначимо.

Обратим внимание на то, что вычислить выборочное корреляционное отношение, а также проверить его значимость можно только в том случае, когда результаты наблюдений сгруппированы в таблицу типа таблицы.

Допустим, что, располагая выборочными данными, мы пришли к выводу, что корреляционная зависимость Y от X существует, т.е. при изменениих изменяются условные математические ожидания $M(Y/X=x)$. Тогда возникает вопрос: каков вид функции регрессии, т.е. функции $\phi(x) = M(Y/X=x)$?

Располагая только выборочными данными, нельзя дать точный ответ на поставленный вопрос, но высказать гипотезу о виде функции $\phi(x)$ можно; также можно провести статистическую проверку этой гипотезы, т.е. выяснить, противоречит или нет эта гипотеза имеющимся выборочным данным.

3. Виды регрессий, статистическая значимость их параметров. Автокорреляция

Исследование начинается с теории, устанавливающей связь между явлениями. Из всего круга факторов, влияющих на результативный признак, выделяются наиболее существенные факторы. После того, как было выявлено наличие взаимосвязи между изучаемыми признаками, определяется точный вид этой зависимости с помощью регрессионного анализа.

Регрессионный анализ заключается в определении аналитического выражения (в определении функции), в котором изменение одной величины (результативного признака) обусловлено влиянием независимой величины (факторного признака). Количественно оценить данную взаимосвязь можно с помощью построения уравнения регрессии или регрессионной функции.

Базисной регрессионной моделью является модель парной (однофакторной) регрессии. Парная регрессия – уравнение связи двух переменных y и x :

$$y = f(x)$$

где Y – зависимая переменная (результативный признак);

X – независимая, объясняющая переменная (факторный признак).

В зависимости от характера изменения ус изменением х различают линейные и нелинейные регрессии.

$$\text{Линейная регрессия } Y = \beta_0 + \beta_1 X + \varepsilon$$

Данная регрессионная функция называется полиномом первой степени и используется для описания равномерно развивающихся во времени процессов.

Наличие случайного члена ε (ошибки регрессии) связано с воздействием на зависимую переменную других неучтенных в уравнении факторов, с возможной нелинейностью модели, ошибками измерения, следовательно, появление случайной ошибки уравнения регрессии может быть обусловлено следующими объективными причинами:

1) нерепрезентативность выборки. В модель парной регрессии включается фактор, не способный полностью объяснить вариацию результативного признака, который может быть подвержен влиянию многих других факторов (пропущенных переменных) в гораздо большей степени. Например, заработка плата может зависеть, кроме квалификации, от уровня образования, стажа работы, пола и пр.;

2) существует вероятность того, что переменные, участвующие в модели, могут быть измерены с ошибкой. Например, данные по расходам семьи на питание составляются на основании записей участников опросов, которые, как предполагается, тщательно фиксируют свои ежедневные расходы. Разумеется, при этом возможны ошибки.

На основе выборочного наблюдения оценивается выборочное уравнение регрессии (линия регрессии):

$$y_x = a + b x,$$

где a, b – оценки параметров уравнения регрессии (α, β).

Аналитическая форма зависимости между изучаемой парой признаков (регрессионная функция) определяется с помощью следующих методов:

На основе теоретического и логического анализа природы изучаемых явлений, их социально-экономической сущности. Например, если изучается зависимость между доходами населения и размером вкладов населения в банки, то очевидно, что связь прямая.

Графический метод, когда характер связи оценивается визуально.

Эту зависимость можно наглядно увидеть, если построить график, отложив на оси абсцисс значения признака x , а на оси ординат – значения признака y . Нанеся на график точки, соответствующие значениям x и y , получим корреляционное поле:

а) если точки беспорядочно разбросаны по всему полю – это говорит об отсутствии зависимости между этими признаками;

б) если точки концентрируются вокруг оси, идущей от нижнего левого угла в верхний правый – то имеется прямая зависимость между признаками;

в) если точки концентрируются вокруг оси, идущей от верхнего левого угла в нижний правый – то обратная зависимость между признаками.

Если на корреляционном поле соединим точки отрезками прямой, то получим ломаную линию с некоторой тенденцией к росту. Это будет эмпирическая линия связи или эмпирическая линия регрессии. По ее виду можно судить не только о наличии, но и о форме зависимости между изучаемыми признаками.

Построение уравнения парной регрессии

Построение уравнения регрессии сводится к оценке ее параметров. Эти оценки параметров могут быть найдены различными способами. Одним из них является метод наименьших квадратов (МНК). Суть метода состоит в следующем. Каждому значению соответствует эмпирическое (наблюданное) значение y . Построив уравнение регрессии, например уравнение прямой линии, каждому значению будет соответствовать теоретическое (расчетное) значение y_x . Наблюдаемые значения не лежат в точности на

линии регрессии, т.е. не совпадают с . Разность между фактическим и расчетным значениями зависимой переменной называется остатком:

$$e = y - y_x$$

МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака у от теоретических y_x , т.е. сумма квадратов остатков, минимальна:

$$\Sigma(y - y_x)^2 \rightarrow \min$$

Для линейных уравнений и нелинейных, приводимых к линейным, решается следующая система относительно а и b:

$$a n + b \Sigma x = \Sigma y$$

$$a \Sigma x + b \Sigma x^2 = \Sigma yx$$

где n– численность выборки.

Решив систему уравнений, получим значения а и b, что позволяет записать уравнение регрессии(регрессионное уравнение):

$y_x = a + bx$ где x – объясняющая (независимая) переменная;

y_x –объясняемая (зависимая) переменная;

Линия регрессии проходит через точку (\bar{x} , \bar{y}) и выполняются равенства:

$$e = 0, \quad y = y_x$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы уравнений:

$$b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\bar{y} \cdot \bar{x} - \bar{y} \cdot \bar{x}}{\bar{x}^2 - \bar{x}^2},$$

$$a = \bar{y} - b \cdot \bar{x}$$

где – среднее значение зависимого признака;

– среднее значение независимого признака;

$\bar{y} \cdot \bar{x}$ –среднее арифметическое значение произведения зависимого и независимого признаков;

σ_x^2 –дисперсия независимого признака;

$\text{cov}(x, y)$ –ковариация между зависимым и независимым признаками.

Выборочной ковариацией двух переменных x, y называется средняя величина произведения отклонений этих переменных от своих средних

Параметр b при x имеет большое практическое значение и носит название коэффициента регрессии. Коэффициент регрессии показывает, на сколько единиц в среднем изменяется величина y при изменении факторного признака x на 1 единицу своего измерения.

Знак параметра b в уравнении парной регрессии указывает на направление связи:

если $b > 0$, то связь между изучаемыми показателями прямая, т.е. с увеличением факторного признаках увеличивается и результативный признаку, и наоборот;

если $b < 0$, то связь между изучаемыми показателями обратная, т.е. с увеличением факторного признаках результативный признак уменьшается, и наоборот.

Значение параметра a в уравнении парной регрессии в ряде случаев можно трактовать как начальное значение результативного признака y. Такая трактовка параметра a возможна только в том случае, если значение $x = 0$ имеет смысл.

После построения уравнения регрессии, наблюдаемые значения y можно представить как: $y = y_x + e$

Остатки e , как и ошибки , являются случайными величинами, однако они, в отличие от ошибок , наблюдаются. Остаток есть та часть зависимой переменной y, которую невозможно объяснить с помощью уравнения регрессии.

На основании уравнения регрессии могут быть вычислены теоретические значения y для любых значений x .

В экономическом анализе часто используется понятие эластичности функции. Эластичность функции рассчитывается как относительное изменение y к относительному изменению x . Эластичность показывает, на сколько процентов изменяется функция при изменении независимой переменной на 1%.

Поскольку эластичность линейной функции не является постоянной величиной, а зависит от x , то обычно рассчитывается коэффициент эластичности как средний показатель эластичности.

Коэффициент эластичности показывает, на сколько процентов в среднем по совокупности изменится величина результативного признака y при изменении факторного признака на 1% от своего среднего значения:

$$\mathcal{E}_x = b \frac{x}{y}$$

где x, y – средние значения переменных x и y в выборке.

Оценка качества построенной модели регрессии

Качество модели регрессии – адекватность построенной модели исходным (наблюдаемым) данным.

Чтобы измерить тесноту связи, т.е. измерить, насколько она близка к функциональной, нужно определить дисперсию, измеряющую отклонения y от \hat{y} и характеризующую остаточную вариацию, обусловленную прочими факторами. Они лежат в основе показателей, характеризующих качество модели регрессии.

При исследовании регрессии устанавливается однофакторная или многофакторная будущая модель и вид модели (линейный или нелинейный).

Обоснование вида модели состоит в выборе вида функции (некоторого аналитического выражения), с помощью которого можно будет описать изменение исследуемого показателя под воздействием факторов.

К обоснованию вида функции идут двумя путями: Теоретическим (анализируя экономическую природу x_0 и x_j , выдвигается гипотеза о характере изменения показателя под действием фактора) И эмпирическим(закон изменения результативного показателя под действием фактора устанавливается путем анализа совокупности фактических данных по полям корреляции).

Наиболее употребительными выражениями при описании связи одного фактора и исследуемого показателя являются:

- Уравнение прямой - $x_0 = a_0 + a_1 x_1$, - Уравнение параболы - $x_0 = a_0 + a_1 x_1 + a_2 x_1^2$, - Уравнение гиперболы - $x_0 = a_0 + \frac{a_1}{x_1}$.

После обоснования парных взаимосвязей переходят к записи многофакторных моделей. В экономических исследованиях чаще всего применяется линейная многофакторная модель - $x_0 = a_0 + a_1 x_1 + \dots + a_n x_n$.

В качестве нелинейных моделей применяются

- Мультипликативная модель - $x_0 = a_0 x_1^{a_1} x_2^{a_2} x_3^{a_3} \dots$ или $x_0 = a_0 a_1^{x_1} a_2^{x_2} a_3^{x_3} \dots$

Для оценки значений параметров регрессионной модели чаще всего используется Метод наименьших квадратов (МНК).Этот метод можно применить как для линейных моделей, так и для нелинейных, допускающих преобразование их к линейному виду путем замены переменных или дифференцированием.

При использовании МНК делаются определенные предпосылки относительно случайной составляющей ε . В модели $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \varepsilon$ случайная составляющая ε представляет собой ненаблюдаемую величину. Поэтому в задачу

регрессионного анализа входит не только построение самой модели, но и исследование случайных отклонений ε_i , т. е. остаточных величин.

Остатки представляют собой независимые случайные величины, и их среднее значение равно 0; они имеют одинаковую (постоянную) дисперсию и подчиняются нормальному распределению.

Статистические проверки параметров регрессии, показателей корреляции основаны на непроверяемых предпосылках распределения случайной составляющей ε_i . Связано это с тем, что оценки параметров регрессии должны отвечать определенным критериям: быть Несмешенными, состоятельными и эффективными. Эти свойства оценок, полученных по МНК, имеют чрезвычайно важное практическое значение в использовании результатов регрессии и корреляции.

Коэффициенты регрессии, найденные из системы нормальных уравнений, представляют собой выборочные оценки характеристики силы связи. Их несмешенность является желательным свойством, т. к. только в этом случае они могут иметь практическую значимость.

Несмешенность оценки означает, что математическое ожидание остатков равно нулю. Оценки считаются Эффективными, если они характеризуются наименьшей дисперсией. Поэтому несмешенность оценки должна дополняться минимальной дисперсией. Состоятельность оценок характеризует увеличение их точности с увеличением объема выработки.

Указанные критерии оценок (несмешенность, состоятельность, эффективность) обязательно учитываются при разных способах оценивания. Метод наименьших квадратов строит оценки регрессии на основе минимизации суммы квадратов остатков ($y - \hat{y}_x$).

Исследование остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

- случайный характер остатков;
- нулевая средняя величина остатков, не зависящая от x_i ;
- гомоскедастичность – дисперсия каждого отклонения ε_i одинакова для всех значений x ;
- отсутствие автокорреляции остатков, т. е. значения остатков ε_i распределены независимо друг от друга;
- остатки подчиняются нормальному распределению.

С целью проверки случайного характера остатков ε_i строится график зависимости остатков ε_i от теоретических значений результативного признака \hat{y} .

Если на графике нет направленности в расположении точек ε_i , то остатки ε_i представляют собой случайные величины и МНК оправдан. Также возможны следующие случаи: если ε_i зависит от теоретического значения, то:

Вторая предпосылка МНК относительно нулевой средней величины остатков $\sum(y - \hat{y}_x) = 0$ означает, что $\sum \varepsilon_i = 0$. Это выполнимо для линейных моделей и моделей, нелинейных относительно включаемых переменных. Для обеспечения несмешенности оценок коэффициентов регрессии, полученных МНК, необходимо выполнение условий

независимости случайных остатков ε_i и переменных x , что исследуется в рамках соблюдения второй предпосылки МНК. С целью проверки выполнение этой предпосылки строится график зависимости случайных остатков ε от факторов, включенных в регрессию x_i . Если расположение остатков на графике не имеет направленности, то они

независимы от значений x_i . Если же график показывает наличие зависимости ε_i и x_i , то модель неадекватна.

Предпосылка о нормальном распределении остатков позволяет проводить проверку параметров регрессии и корреляции с помощью критериев t и F. Вместе с тем оценки регрессии, найденные с применением МНК, обладают хорошими свойствами даже при отсутствии нормального распределения остатков, т. е. при нарушении пятой предпосылки метода наименьших квадратов.

В соответствии с третьей предпосылкой МНК требуется, чтобы дисперсия остатков была гомоскедастичной. Это означает, что для каждого значения фактора x_i остатки ε_i имеют одинаковую дисперсию. Если это условие применения МНК не соблюдается, то имеет место гетероскедастичность. Используя трехмерной изображение, рассмотрим отличие гомо- и гетероскедастичности.

Наличие гетероскедастичности будет сказываться на уменьшении эффективности оценок b_i , в частности, становится затруднительным использование формулы стандартной ошибки коэффициента регрессии, предполагающей единую дисперсию остатков для любых значений фактора.

Наличие гетероскедастичности в остатках регрессии можно проверить с помощью ранговой корреляции Спирмэна. Суть проверки заключается в том, что в случае гетероскедастичности абсолютные остатки ε_i коррелированы со значениями фактора x_i . Этую корреляцию можно измерять с помощью коэффициента ранговой корреляции

$$\rho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)},$$

Спирмэна:

где ρ – абсолютная разность между рангами значений x_i и $|\varepsilon_i|$.

Статистическую значимость ρ можно определить с помощью t-критерия:

$$t_{\rho} = \frac{\rho}{\sqrt{(1 - \rho^2)}} \sqrt{(n - 2)}$$

Принято считать, что если $t_{\text{расч}} > t_{\text{табл}}$, то корреляция между ε_i и x_i статистически значима, т. е. имеет место гетероскедастичность остатков. В противном случае принимается гипотеза об отсутствии гетероскедастичности остатков.

При построении регрессионных моделей чрезвычайно важно соблюдение четвертой предпосылки МНК – отсутствие автокорреляции остатков, т. е. распределения остатков ε_i и ε_{i-1} независимы. Автокорреляция остатков означает наличие корреляции между остатками текущих и предыдущих (последующих) наблюдений. Находится коэффициент корреляции между ε_i и ε_{i-1} , и если он окажется существенно отличным от нуля, то остатки автокоррелированы и функция плотности вероятности $F(\varepsilon)$ зависит от j-ой точки наблюдения и от распределения значений остатков в других точках наблюдения.

Отсутствие автокорреляции остатков обеспечивает состоятельность и эффективность оценок коэффициентов регрессии.

До сих пор в качестве факторов рассматривались экономические переменные, принимающие количественные значения в некотором интервале. Вместе с тем может оказаться необходимым включить в модель фактор, имеющий два или более качественных уровней. Это могут быть разного рода атрибутивные признаки, такие, например, как профессия, пол, образование, климатические условия, принадлежность к определенному региону. Для того, чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены те или иные цифровые метки, т. е. качественные переменные необходимо

преобразовать в количественные. Такого вида сконструированные переменные в эконометрике принято называть фиктивными переменными.

Качественные признаки могут приводить к неоднородности исследуемой совокупности, что может быть учтено при моделировании двумя путями:

- регрессия строится для каждой качественно отличной группы единиц совокупности, т. е. для каждой группы в отдельности, чтобы преодолеть неоднородность единиц общей совокупности;
- общая регрессионная модель строится для совокупности в целом, учитываяющей неоднородность данных. В этом случае в регрессионную модель вводятся фиктивные переменные, т. е. строится регрессионная модель с переменной структурой, отражающей неоднородность данных.

Качественный фактор может иметь только два состояния, которым будут соответствовать 1 и 0. Если же число градаций качественного признака-фактора превышает два, то в модель вводится несколько фиктивных переменных, число которых должно быть меньше числа качественных градаций. Только при соблюдении этого положения матрица исходных фиктивных переменных не будет линейно зависима и возможна оценка параметров модели.

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории к другой при неизменных значениях остальных параметров. На основе t-критерия Стьюдента делается вывод о значимости влияния фиктивной переменной, существенности расхождения между категориями.

Такая проверка производится с помощью статистических критериев и на их основе делается вывод о статистической надежности построенного уравнения регрессии, о пригодности модели для анализа и прогнозирования исследуемого показателя.

Для проверки надежности модели в целом используется отношение факторной

$$\frac{S_{\text{факт}}^2}{S_{\text{ост}}^2}$$

дисперсии к остаточной $S_{\text{ост}}^2$. Известно, что отношение этих дисперсий подчиняется распределению Фишера (F-распределение). Расчетное значение F-отношения сравнивается с табличным значением, которое определяется для конкретного уровня значимости α . В экономических исследованиях α принимается равным 0,05 (реже 0,01), число степеней свободы $k_1 = p, k_2 = n - p - 1$. Если $F_{\text{расч}} > F_{\text{табл}}$, то построенная модель считается статистически надежной, а следовательно, отражает закон изменения исследуемого показателя под действием факторов.

Для проверки полноту модели используется $R^2 \cdot 100\%$. Этот показатель показывает, на сколько процентов изменится вариация результативного показателя под влиянием факторов, включенных в модель.

Проверку надежности параметров уравнения регрессии проводят с использованием Т-критерия. Расчетное значение вычисляется по

$$t = \frac{|a_j|}{\sigma_{a_j}} \sigma_{a_j} = \frac{S_{\text{ост}}}{\sigma_{x_j} \sqrt{n} \sqrt{1 - R^2_{j,1,2,\dots,j-1,j+1,p}}} \quad \text{Фактическое значение Т-критерия}$$

формуле

сравнивается с табличным и если $t_{\text{факт}} > t_{\text{табл}} (t_{\alpha,k}, \alpha = 0,05(0,01), k = n - p - 1)$, то тогда соответствующий коэффициент регрессии значим, т. е. отличен от нуля, а влияние J-го фактора следует считать сильным. Факторы, оказывающие несущественное влияние на исследуемый показатель, из модели исключают.

На этом этапе разрабатываются рекомендации об использовании результатов моделирования. Анализируется уравнение регрессии в натуральном масштабе: коэффициент регрессии a_j показывает, на сколько своих единиц измерения в среднем изменится исследуемый показатель, при увеличении J-го фактора на единицу своего измерения, при условии, что все остальные факторы находятся на постоянном уровне.

Свободный член уравнения характеризует изменение результативного показателя за счет изменения факторов, неучтенных в модели.

В связи с тем, что факторы имеют различный физический смысл и различные единицы измерения, коэффициенты регрессии нельзя сравнивать между собой и, следовательно, трудно определить, какой из факторов оказывает наибольшее влияние. Для устранения различий в единицах измерения применяют Частные коэффициенты

$\beta_j = \frac{\partial y}{\partial x_j}$ эластичности x_0 характеризующие, на сколько % в среднем изменится x_0 при увеличении J -го Фактора на 1% при фиксированном положении других факторов.

При определении степени влияния отдельных факторов необходим показатель, который бы учитывал влияние анализируемых факторов с учетом различий в уровне их вариации. Таким показателем является Коэффициент регрессии в стандартизованном

масштабе $\beta_j = \frac{\partial y}{\partial z_j}$. Коэффициент β_j показывает,

на какую часть своего среднеквадратического отклонения изменится x_0 при изменении J-го фактора на одно свое среднеквадратическое отклонение при фиксированном значении остальных факторов. Уравнение регрессии в стандартизированном масштабе: $t_{0,1,2,\dots,p} = \sum_{j=1}^p \beta_j t_j$, где $t_j = \frac{x_j - \bar{x}_j}{\sigma_{x_j}}$.

Т. к. в стандартизованном уравнении все факторы и функция измеряются в одинаковых единицах измерения – стандартных отклонениях, то по стандартизованным коэффициентам можно судить о влиянии каждого фактора по сравнению с другими.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1 Практическое занятие № 1 (2 часа).

Тема: «Проблемы современной фундаментальной науки»

2.1.1 Задание для работы:

1. Исследования в области вещества и энергии
2. Исследования в области новых материалов и технологий

2.1.2 Краткое описание проводимого занятия:

1. Исследования в области вещества и энергии – доклады с презентациями по теме, обсуждение
2. Исследования в области новых материалов и технологий – доклады с презентациями по теме, обсуждение

2.1.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с современными направлениями в исследованиях вещества и энергии, новых материалов и технологий;
- выработать навыки по изучению интегративных качеств фундаментальных моделей.

2.2 Практическое занятие № 2 (2 часа).

Тема: «Проблемы современной фундаментальной науки»

2.2.1 Задание для работы:

1. Исследования в области вещества и энергии
2. Исследования в области новых материалов и технологий

2.2.2 Краткое описание проводимого занятия:

1. Исследования в области новых материалов и технологий – доклады с презентациями по теме, обсуждение

2. Построение глобального информационного пространства – доклады с презентациями по теме, обсуждение

2.2.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с современными направлениями в исследованиях вещества и энергии, новых материалов и технологий;
- усвоить основные принципы построения глобального информационного пространства;
- выработать навыки по изучению интегративных качеств фундаментальных моделей.

2.3 Практическое занятие № 3 (2 часа).

Тема: «Методологическая основа научно-исследовательской работы»

2.3.1 Задание для работы:

1. Виды НИР
2. Вопросы методологии
3. Этапы НИР

2.3.2 Краткое описание проводимого занятия:

1. Виды НИР – доклады с презентациями, обсуждение

2. Вопросы методологии - доклады с презентациями, обсуждение

3. Этапы НИР - доклады с презентациями, обсуждение

2.3.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с классификацией НИР, ее особенностями;
- усвоить основные требования к методологии научно-исследовательских работ;
- выработать навыки анализа научной проблемы и построение поэтапного плана ее решения.

2.4 Практическое занятие № 4 (2 часа).

Тема: «Научно-исследовательская работа студентов»

2.4.1 Задание для работы:

1. Основные принципы построения НИРС
2. Работа с информационными источниками. Этические аспекты НИР

2.4.2 Краткое описание проводимого занятия:

1. Основные принципы построения НИРС - доклады с презентациями, обсуждение

2. Работа с информационными источниками. Этические аспекты НИР - доклады с презентациями, обсуждение

2.4.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными принципами построения НИРС;
- усвоить основные правила оформления научно-исследовательских работ; этические нормы и правила ведения научно-исследовательской работы;
- выработать навыки работы с информационными источниками.

2.5 Практическое занятие № 5 (2 часа).

Тема: «Виды научно-исследовательских работ, их особенности, оформление»

2.5.1 Задание для работы:

1. Виды научно-исследовательских работ, их особенности
2. Правила оформления научно-исследовательских работ

2.5.2 Краткое описание проводимого занятия:

- 1. Виды научно-исследовательских работ, их особенности** - доклады с презентациями, обсуждение
- 2. Правила оформления научно-исследовательских работ** - доклады с презентациями, обсуждение

2.5.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными видами научно-исследовательских работ, их особенностями;
- усвоить основные правила оформления научно-исследовательских работ; этические нормы и правила ведения научно-исследовательской работы;
- выработать навыки работы с информационными источниками.

2.6 Практическое занятие № 6 (2 часа).

Тема: «Математическое моделирование в инженерных исследованиях»

2.6.1 Задание для работы:

1. Математическая модель: этапы построения, отличительные особенности.
2. Типовые математические модели в инженерных исследованиях.

2.6.2 Краткое описание проводимого занятия:

- 1. Математическая модель: этапы построения, отличительные особенности** -доклады с презентациями, обсуждение
- 2. Типовые математические модели в инженерных исследованиях-** доклады с презентациями, обсуждение

2.6.3 Результаты и выводы:

В результате проведенного занятия студенты:

- рассмотреть типовые математические модели, применяемые в инженерных приложениях и условия их использования;
- усвоить основные понятия, связанные с математическим моделированием;
- выработать навыки анализа этапов построения математических моделей.

2.7 Практическое занятие № 7 (2 часа).

Тема: «Задачи линейной оптимизации. Методы их решения»

2.7.1 Задание для работы:

1. Графический метод решения ЗЛП.
2. Теоретические основы симплекс- метода
3. Симплекс метод решения ЗЛП

2.7.2 Краткое описание проводимого занятия:

1. Графический метод решения ЗЛП

Пример.

Решить графически данную задачу линейного программирования.

$$F = 4x_1 + 3x_2 - 1 \rightarrow \max$$

$$\begin{cases} x_1 + x_2 \leq 8 \\ -2x_1 + 3x_2 \leq 9 \\ 2x_1 - x_2 \leq 10 \\ x_1 \geq 0, x_2 \geq 0 \end{cases}$$

$$x_1 + x_2 = 8 \quad (1)$$

Решение. Найдем вначале область допустимых решений (ОДР). Решим графически первое неравенство:

$$x_1 + x_2 = 8 \quad (1)$$

Для этого построим вначале прямую линию, соответствующую уравнению:

$$x_1 + x_2 = 8 \quad (1)$$

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: $-2x_1 + 3x_2 \leq 9$. (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.

$$\kappa = \frac{2}{3}$$

Угловой коэффициент этой прямой $\kappa = \frac{2}{3}$. Но угловой коэффициент любой прямой

$$\operatorname{tg}\alpha = \kappa = \frac{2}{3}$$

равен $\operatorname{tg}\alpha$, где α - угол наклона прямой к оси 0x. 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) заменяется на уравнение $x_2 = 2x_1 - 10$.

Ясно, что прямая проходит через т. М5 (5;0), и имеет угловой коэффициент $\kappa = 2$.

При этом самому неравенству $2x_1 - x_2 \leq 10 \Leftrightarrow x_2 \geq 2x_1 - 10$

Соответствует верхняя полуплоскость, отмеченная штрихами вверх от прямой 3.

Тривиальному неравенству $x_1 \geq 0$ 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 $x_2 \geq 0$ 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 ОАВСД.

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 $F(x) = 11$. Она определяется уравнением: $4x_1 + 3x_2 = 12$. (4)

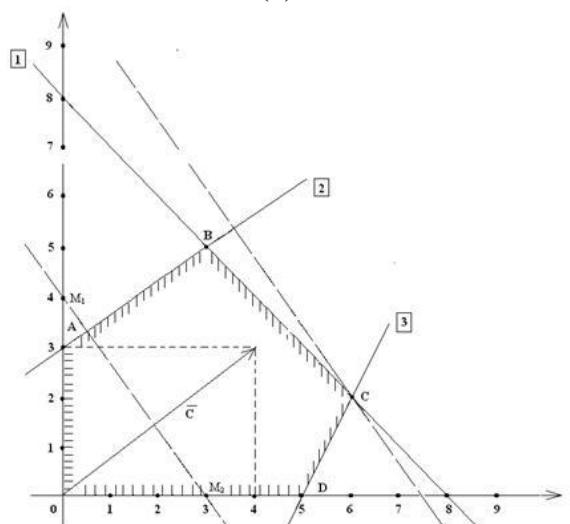
Мы взяли здесь константу $C = 11$, для того чтобы точки пересечения

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} .

Отрезок $[M_1, M_2]$ пересекается с ОДР и в каждой его точке х значение целевой



функции $F(\bar{x})$ равно 11: $F(\bar{x}) = 11, \quad \bar{x} \in [M_1; M_2]$.

Мы знаем, что значение функции F увеличивается в направлении вектора роста \vec{c} . Чтобы найти максимальное значение $F(x)$ на ОДР будем параллельно перемещать линию уровня в направлении вектора роста \vec{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 C найдены: $C(6; 2)$. 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 $C(6; 2)$ 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:

$$F = 4x_1 + 3x_2 - 1 \rightarrow \max$$

$$\begin{cases} x_1 + x_2 \leq 8 \\ -2x_1 + 3x_2 \leq 9 \\ 2x_1 - x_2 \leq 10 \end{cases} \quad (1)$$

$$x_1 \geq 0, \quad x_2 \geq 0$$

Добавив к левым частям системы неравенств соответствующие балансовые переменные W_1, W_2, W_3 , преобразуем задачу (1) в каноническую форму:

$$F = 4x_1 + 3x_2 - 1 \rightarrow \max$$

$$\begin{cases} x_1 + x_2 + W_1 = 8 \\ -2x_1 + 3x_2 + W_2 = 9 \\ 2x_1 - x_2 + W_3 = 10 \end{cases} \quad (2)$$

$$x_1 \geq 0, \quad x_2 \geq 0, \quad W_1 \geq 0, \quad W_2 \geq 0, \quad W_3 \geq 0.$$

Для удобства и единства записи определим целевую функцию F в виде уравнения: $F = 4x_1 + 3x_2 - 1 = 0$ (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)) содержит 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 $F - c_1x_1 - c_2x_2 - \dots - c_nx_n = c_0$, целевой функции F Канонической задачи ЛП неположительные: $c_j \leq 0$, - и функция F зависит только от свободных переменных, то соответствующее опорное решение является оптимальным.

Но условие $c_j \leq 0$ означает, что коэффициенты индексной строки, стоящие в столбцах свободных переменных, должны быть неотрицательны: $-c_j \geq 0$, поскольку индексная строка соответствует уравнению: $F - c_1x_1 - c_2x_2 - \dots - c_nx_n = c_0$, - 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 $\alpha_i = \frac{b_i}{a_{ij}}$,

Где j - номер Ведущего столбца.

В нашем случае $j = 1$ и допустимые отношения $\alpha_1, \alpha_2, \alpha_3$. Соответственно равны:

$$\alpha_1 = \frac{b_1}{a_{11}} = \frac{8}{1} = 8, \quad (a_{11} = 1 > 0),$$

$$\alpha_2 = +\infty, \quad (a_{21} = -2 < 0),$$

$$\alpha_3 = \frac{b_3}{a_{31}} = \frac{10}{2} = 5, \quad (a_{31} = 2 > 0)$$

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 которой 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, $x_1 = 5$, $x_2 = 0$, $W_1 = 3$, $W_2 = 19$, $W_3 = 0$, - 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	$\frac{3}{2}$	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:

$x_1 = 6, x_2 = 2, W_1 = 0, W_2 = 15, W_3 = 0$, и значение целевой функции $F_3 = 29$.

Поскольку индексная строка таблицы (11) не содержит отрицательных элементов,

полученное опорное решение будет оптимальным: $F_{\text{опт}} = F^* = 29$ и $x_1^* = 6, x_2^* = 2$. (12)

При этом $W_1^* = 0, W_2^* = 15, W_3^* = 0$. Здесь мы звездочками помечаем оптимальные значения переменных.

Таким образом, задача (2), а с ней и задача (1) решены.

2.7.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными понятиями линейного, математического программирования, теории двойственности, основными теоремами и свойствами моделей ЗЛП;
- усвоить теоретические основы симплекс метода;
- выработать навыки применения графического метода решения ЗЛП, симплекс- метода.

2.8 Практическое занятие № 8 (2 часа).

Тема: «Задачи линейной оптимизации. Методы их решения»

2.8.1 Задание для работы:

1. Транспортная задача. Первичное распределение поставок
2. Циклы, оценки циклов, оценки клеток
3. Перераспределение поставок
4. Открытая модель транспортной задачи (ТЗ)

2.8.2 Краткое описание проводимого занятия:

1. Первичное распределение поставок

		B ₁	B ₂	B ₃	B ₄	B ₅
		40	80	110	50	90
Пункт отправления	Пункт назначения					
		10	9	8	5	3
A ₁	100	10	9	8	5	3
A ₂	150	4	7	13	6	2
A ₃	90	8	6	9	7	1
A ₄	30	5	4	7	9	2

В табл. 2.2 по строкам матрицы представлены пункты (станции) отправления от A₁ до A₄ и объемы погрузки в тоннах – 100, 150, 90, 30 т, а по столбцам – пункты (станции) назначения от B₁ до B₅ и объемы выгрузки – 40, 80, 110, 50, 90 т. Данная транспортная задача является сбалансированной ($a_i = b_j = 370$ т), поэтому добавлять фиктивного потребителя ФВ или фиктивного поставщика ФА не требуется. На пересечении строк и столбцов в клетках матрицы в маленьких квадратиках записаны показатели критерия оптимальности транспортной задачи, например, затраты на перевозку единицы груза или кратчайшие расстояния между соответствующими пунктами (станциями) погрузки и выгрузки. Расстояние между станцией погрузки A₁ и станцией выгрузки B₁, как следует из матрицы, равно 10 (или 100, 1000 и т. д.) км, потом – 9, 8, 5 км и т. д. Тогда целью, решения задачи явится отыскание совокупности объемов перевозок между всеми пунктами (станциями) погрузки и выгрузки (корреспонденций),

обеспечивающей минимальный объем перевозочной работы (грузооборота) в тонно-километрах. Любую совокупность корреспонденций, обеспечивающую весь объем перевозок, будем называть планом, а совокупность, обеспечивающую минимум грузооборота, – оптимальным планом перевозок.

Алгоритм решения транспортной задачи линейного программирования будем описывать по операциям с присвоением номера и названия.

Операция 1. Построение опорного плана.

Опорным, называется любой допустимый, как правило, не оптимальный план, который является исходным для последующего решения. Для построения опорного плана существует ряд методов. Самый простой из них – метод северо-западного угла.

Метод северо-западного угла

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅
Пункт отправления		40	80	110	50	90
	A ₁	100	40 10	60 9	8	5
A ₂	150	4	20 7	110 13	20 6	2
A ₃	90	8	6	9	30 7	60 1
A ₄	30	5	4	7	9	30 2

Берем «северо-западную» клетку матрицы – это клетка A₁B₁ и записываем в нее максимально возможную поставку – 40 т (объем выгрузки 40 т, ресурсы станции погрузки 100 т). Поскольку ресурсы станции погрузки A₁ не исчерпаны, следуем по первой строке вправо и записываем в клетку A₁B₂ корреспонденцию равную максимально возможной величине – 60 т. Таким образом получается, что ресурсы станции A₁ полностью использованы, однако спрос станции выгрузки B₂ не удовлетворен. Тогда от клетки A₁B₂ опускаемся вниз до клетки A₂B₂ и записываем в нее поставку равную 20 т. Описанным способом следуем далее до последней «юго-западной» клетки матрицы. В результате получаем допустимый план перевозок груза. Грузооборот или функционал транспортной задачи (сумма произведений корреспонденций на расстояние, рассчитанная по табл. 2.3) составит:

$$F_{\text{сев.-зап.}} = 40 \cdot 10 + 60 \cdot 9 + 20 \cdot 7 + 110 \cdot 13 + 20 \cdot 6 + 30 \cdot 7 + 60 \cdot 1 + 30 \cdot 2 = 2960 \text{ ткм.}$$

После расстановки корреспонденции матрица проверяется на вырождение, т. е. должно выполняться условие

$$m + n - 1 = N_{\text{баз.}}, \quad (*)$$

где m – количество строк, n – количество столбцов, N_{баз.} – количество базисных клеток.

Другими словами, количество клеток матрицы, содержащих корреспонденции, должно быть равно сумме строк и столбцов без единицы. В нашем случае это условие соблюдается: 8 = 4 + 5 – 1. План транспортной задачи, отвечающий условию (n + m – 1) называют базисным. Базисными также называются клетки матрицы, содержащие поставки. Клетки, в которых поставки отсутствуют, называются небазисными.

Метод северо-западного угла имеет существенный недостаток. При его использовании не учитываются значения показателей критерия оптимальности в клетках матрицы. Поэтому поставки могут попасть в «дорогие» клетки с заведомо высокой ценой или большим расстоянием. Опорный план, полученный с

использованием данного метода, как правило, далек от оптимального, что обуславливает большой объем последующих расчетов для доведения его до оптимального. Описанный метод обычно не применяется.

Наиболее предпочтительным при ручном решении транспортных задач считается метод минимальной стоимости или, как его еще называют, метод наименьшего элемента в матрице. Суть его в следующем. В транспортной матрице выбирается клетка с минимальной стоимостью (расстоянием). В нашем случае это клетка A_3B_5 . В нее записывается максимально возможная поставка – это 90 т.

Метод наименьшего элемента в матрице

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅				
Пункт отправления		40	80	110	50	90				
	A ₁	100	10	9	50	8	50	5	3	
A ₂	150	40	4	50	7	60	13	6	2	
A ₃	90		8	6		9		7	90	1
A ₄	30		5	30	4		7	9		2

Далее отыскиваются клетка, имеющая следующее по величине расстояние. В нашей матрице это две клетки с расстоянием 2 км в пятом столбце. Однако в эти клетки поставки корреспонденцию грузов ставить нельзя, поскольку спрос станции B₅ полностью удовлетворен со станции A₃. Поэтому столбец 5 из дальнейшего построения плана исключаем. Следующие по величине показателя критерия оптимальности клетки с расстоянием 4 км это клетки A₂B₁ и A₄B₂. Выбираем одну из них, например, A₂B₁ и записываем в нее поставку 40 т. Далее идет клетка A₄B₂ – поставка 30 т, потом A₁B₄ – 50 т, A₂B₂ – 50 т. Все оставшиеся ресурсы по станциям погрузки распределяем между клетками третьего столбца в клетки A₁B₃ и A₂B₃. После составления опорного плана во избежание ошибок целесообразно проверить балансы погрузки, выгрузки и суммы корреспонденций по строкам и по столбцам матрицы. Функционал F плана, составленного методом наименьшей стоимости, равен 2150 ткм. Таким образом, построенный план значительно лучше плана, построенного методом северо-западного угла. Однако число базисных клеток в плане – 7. Это не соответствует условию (*), т. е. меньше требуемого на единицу. Такой план математики назвали вырожденным (случай вырождения). Случай вырождения «лечат» путем введения в матрицу недостающего количества базисных клеток с нулевыми поставками. Нулевую поставку необходимо вводить в матрицу рядом с базисной клеткой, которая обусловила «пропажу» базисной клетки.

Для того чтобы понять, почему «пропадают» поставки, обратимся к методу северо-западного угла. Из табл. Северо-западного угла следует, что как только была заполнена «северо-западная» клетка, рядом с ней сразу появляется соседняя базисная клетка, потом еще одна и т.д. Цепочка базисных клеток без разрыва следует до «юго-восточного угла» матрицы. Однако если бы в этой цепочке появилась клетка, связывающая поставщика и потребителя с равными объемами погрузки и выгрузки, и в нее была бы записана такая же поставка, то это привело бы к пропаже базисной клетки. Описанная ситуация имела место в табл. 2.4, когда в клетку A₃B₅ была введена корреспонденция объемом 90 т, равная объемам погрузки и выгрузки по соответствующим станциям. Поэтому необходимо ввести в план дополнительную базисную клетку с нулевой поставкой. Эта клетка должна стоять рядом с клеткой A₃B₅. Из трех соседних клеток следует выбрать клетку с

минимальным расстоянием, например, A_2B_5 . Записываем в нее корреспонденцию, равную «0».

Добавление нулевой поставки

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅
Пункт отправления		40	80	110	50	90
		10	9	8	5	3
A ₁	100					
A ₂	150	40	4	50	13	6
A ₃	90		8	6	9	7
A ₄	30		5	30	4	7

Таким образом, причиной вырождения плана транспортной задачи является наличие поставщиков и потребителей с равными объемами погрузки и выгрузки или равными объемами сумм погрузки и выгрузки по нескольким станциям в разнообразных комбинациях. Такие случаи необходимо уметь находить для того, чтобы правильно определять места для нулевых поставок. В процессе решения задачи возможны случаи, когда число базисных клеток превышает величину « $n + m - 1$ ». Это означает появление ошибки вследствие того, что при построении опорного плана в какую-то клетку была введена не максимально возможная поставка.

2. Циклы, оценки циклов, оценки клеток. Перераспределение поставок

Проверка плана на оптимальность.

Операция 2.1. Расчет потенциалов.

Проверка плана транспортной задачи в описываемом методе на оптимальность осуществляется с помощью потенциалов. Потенциалы – это такие числа, которые по определенным правилам назначаются каждой строке и каждому столбцу. **Потенциалы строк** обозначим u_i , **потенциалы столбцов** – v_j . Они могут принимать любые значения. Однако удобнее работать с положительными, целыми и относительно небольшими числами. Такой потенциал первоначально назначается любой строке или столбцу. Рекомендуем поступать следующим образом. Выберем базисную клетку с максимальным расстоянием. В нашей матрице это клетка A_2B_3 . Присвоим строке, в которой находится эта клетка, потенциал, равный 0 ($u_2 = 0$). Далее можно рассчитать потенциалы столбцов по

$$v_j = u_i + c_{ij} \quad (**)$$

базисным клеткам строки 3 по формуле

Потенциал первого столбца $v_1 = u_2 + c_{21} = 0 + 4 = 4$;

второго: $v_2 = u_2 + c_{22} = 0 + 7 = 7$;

третьего: $v_3 = u_2 + c_{23} = 0 + 13 = 13$;

пятого: $v_5 = u_2 + c_{25} = 0 + 2 = 2$.

Рассчитанные потенциалы записываем напротив соответствующих столбцов ниже матрицы. Поскольку по всем базисным клеткам строки 2 потенциалы столбов найдены, переходим к расчету потенциалов строк.

Потенциал строки 1 рассчитываем по найденному потенциалу столбца 3 и базисной клетке A_1B_3 по формуле

$$u_i = v_j - c_{ij}, \quad (***)$$

где $u_1 = v_3 - c_{31} = 13 - 8 = 5$.

Для строки 3 потенциал будет равен: $u_3 = v_5 - c_{35} = 2 - 1 = 1$.

Также рассчитываем потенциалы для всех строк и столбцов.

Расстановка потенциалов и перераспределение поставок

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅
Пункт отправления		40	80	110	50	90
A ₁	100	10	9	100 50	8	50 5
A ₂	150	40 4	50 7	10 60	13	* 6 + 50
A ₃	90	8	6	*	9	* 7 90 1
A ₄	30	5	4	*	7	9 2

и
5
0
1
3

Операция 2.2. Проверка небазисных клеток на соответствие их условию оптимальности.

Оптимальный план транспортной задачи должен отвечать **критерию оптимальности**, который выражается в том, соответствуют ли небазисные клетки матрицы условию, формулируемому следующим выражением:

$$v_j - u_i \leq c_{ij}.$$

Если это условие для всех небазисных клеток выполняется, то план является оптимальным, а если нет, хотя бы для одной клетки, то план не оптимален. Иначе говоря, существует некоторый план с меньшим функционалом. Разность потенциалов может интерпретироваться как некоторая условная цена перевозки единицы продукции по маршруту, связывающему соответствующие станции «i» и «j». Если она ниже c_{ij} , значит, использование данного маршрута не улучшит план, а если c_{ij} ниже разности потенциалов, т. е. условие (2.8) не выполняется, следовательно, существует план лучше рассчитанного, который необходимо отыскать.

Проверим условие (2.8) для табл. 2.6.

Клетка A₁B₁: 4 – 5 < 10, условие выполняется.

Клетка A₁B₂: 7 – 5 < 9, условие выполняется и т. д.

Если для всех небазисных клеток условие 3 выполняется, то рассматриваемый план будет оптimalен. Дальнейшие действия переходят по алгоритму к операции 4.

Однако для нашего примера это не так. Не выполняется условие для клетки A₂B₄ (10 – 0 > 6), клетки A₃B₃ (13 – 1 > 9), а также для клеток A₃B₄, A₄B₃, из чего следует, что разработанный опорный план не оптимален. Отметим эти клетки.

Операция 3. Улучшение плана.

Поскольку полученный план не оптимален, дальнейшие действия алгоритма состоят в его преобразовании в лучшую сторону или просто улучшения.

Операция 3.1. Построение цепи (контура, цикла) перераспределения поставок.

Улучшение плана осуществляется по одной из небазисных клеток, для которой условие оптимальности оказалось невыполненным. В нашем плане имеется четыре такие клетки. Выбираем одну из них, для которой условие оптимальности не выполняется в наибольшей степени. В нашем плане это клетка A₂B₄. Для нее условие оптимальности не выполнено на 4 единицы (10 – 0 – 6 = 4). Для этой клетки строим цепь перераспределения поставок. Цепь перераспределения поставок – это такая замкнутая ломаная линия, которая проходит по клеткам матрицы ходом шахматной ладьи. В вершинах контура обязательно лежит одна небазисная клетка (несоответствующая условию оптимальности, найденная ранее), а остальные соответствуют только базисным клеткам. Линии контура могут пересекаться. Для небазисной клетки A₂B₄ цепь будет проходить по клеткам A₁B₄, A₁B₃, A₂B₃.

Возможные варианты построения цикла перераспределения

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅	
Пункт отправления		b ₁	b ₂	b ₃	b ₄	b ₅	
	A ₁	a ₁	C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₅
A ₂	a ₂		C ₂₁	C ₂₂	C ₂₃	C ₂₄	C ₂₅
A ₃	a ₃		C ₃₁	C ₃₂	C ₃₃	C ₃₄	C ₃₅
A ₄	a ₄		C ₄₁	C ₄₂	C ₄₃	C ₄₄	C ₄₅

В нашем случае форма цепи простая. Однако цепь может иметь любую форму, в том числе и причудливую. Её нужно научиться отыскивать, используя эвристические подходы. При этом необходимо учитывать, что каждая небазисная клетка транспортной матрицы обязательно имеет одну цепь перераспределения поставок.

Операция 3.2. Перераспределение поставок.

Перераспределение поставок производится по цепи. Вначале определим объем перераспределения поставок. Для этого присвоим клеткам – вершинам цепи – знаки. В небазисную клетку A₂B₄ ставим «+», поскольку в нее будет вводиться поставка. Далее, чередуя «+» с «-», расставляем знаки по остальным вершинам контура. Величина объема перераспределения поставок принимается равной минимальной поставке в отрицательной клетке. Для нашего случая это 50 единиц груза. Перераспределение заключается в том, что к поставкам в положительных клетках найденный объем прибавляется, а для отрицательных клеток отнимается. Функционал F' нового плана, представленного в табл. 2.6 (выделенные поставки), составляет 1950 ткм, что на 200 ткм меньше значения функционала F предыдущего плана.

Полученный улучшенный план, требует проверки на оптимальность, поэтому необходимо вернуться к операции 2.

Совокупность действий, описанных в операциях 2 и 3, в процессе решения задачи повторяется до тех пор, пока не будет получен оптимальный план. Эта совокупность носит итеративный (циклический) характер, поэтому она называется итерацией. Через определенное число итераций план становится оптимальным. После этого осуществляется переход от второй операции к четвертой.

Повторение операций 2, 3

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅
Пункт отправления		40	80	110	50	90
	A ₁	100	10	9	100	5
A ₂	150	40	4	60 50	7 10 13	6 0 2
A ₃	90		8	6	* 9	7 90 1
A ₄	30	5	20 30	4	* 10 7	9 2

От матрицы к матрице грузооборот (затраты на транспортировку) должны снижаться. Если план не оптимальен, то необходимо произвести повторный расчёт потенциалов, проверить небазисные клетки на соответствие условию оптимальности.

Проверка плана на оптимальность свидетельствует о том, что для двух клеток условия оптимальности не выполняются. После перераспределения поставок по клетке A_4B_3 , получаем новый план.

Оптимальный план поставок

Пункт назначения		B ₁	B ₂	B ₃	B ₄	B ₅
Пункт отправления		40	80	110	50	90
A ₁	100	10	9	100	8	5
A ₂	150	40	4	60	7	13
A ₃	90	8	6	9	7	90
A ₄	30	5	20	4	10	7
	v ₁	4	7	10	6	2

Проверка плана перевозок на оптимальность по условию (2.8) показала, что для всех небазисных клеток матрицы условия оптимальности выполняются. Функционал F'' оптимального плана равен 1920 ткм. Таким образом, получен план перевозок, обеспечивающий минимальный объем перевозочной работы для транспортировки всего груза между станциями погрузки и выгрузки.

3. Открытая модель ТЗ

Стоимость доставки единицы груза из каждого пункта отправления в соответствующие пункты назначения задана матрицей тарифов

	1	2	3	4	5	Запасы
1	3	20	8	13	4	80
2	4	4	18	14	3	60
3	10	4	18	8	6	30
4	7	19	17	10	1	60
Потребности	10	30	40	50	70	

Проверим необходимое и достаточное условие разрешимости задачи.

Как видно, суммарная потребность груза в пунктах назначения превышает запасы груза на базах. Следовательно, модель исходной транспортной задачи является открытой. Чтобы получить закрытую модель, введем дополнительную (фиктивную) базу с запасом груза, равным 30 (230-200). Тарифы перевозки единицы груза из базы во все магазины полагаем равны нулю.

Занесем исходные данные в распределительную таблицу.

	1	2	3	4	5	6	Запасы
1	3	20	8	13	4	0	80
2	4	4	18	14	3	0	60
3	10	4	18	8	6	0	30
4	7	19	17	10	1	0	60
Потребности	10	30	40	50	70	30	

1. Используя метод наименьшей стоимости, построим первый опорный план транспортной задачи.

	1	2	3	4	5	6	Запасы
1	3[10]	20	8[40]	13[20]	4	0[10]	80
2	4	4[30]	18	14	3[10]	0[20]	60
3	10	4	18	8[30]	6	0	30
4	7	19	17	10	1[60]	0	60
Потребности	10	30	40	50	70	30	

В результате получен первый опорный план, который является допустимым, так как все грузы из баз вывезены, потребность магазинов удовлетворена, а план соответствует системе ограничений транспортной задачи.

2. Подсчитаем число занятых клеток таблицы, их 9, а должно быть $m + n - 1 = 9$.

Следовательно, опорный план является невырожденным.

4. Проверим оптимальность опорного плана. Найдем потенциалы u_i, v_j по занятым клеткам таблицы, в которых $u_i + v_j = c_{ij}$, полагая, что $u_1 = 0$.

$$u_1 + v_1 = 3; 0 + v_1 = 3; v_1 = 3$$

$$u_1 + v_3 = 8; 0 + v_3 = 8; v_3 = 8$$

$$u_1 + v_4 = 13; 0 + v_4 = 13; v_4 = 13$$

$$u_3 + v_4 = 8; 13 + u_3 = 8; u_3 = -5$$

$$u_1 + v_6 = 0; 0 + v_6 = 0; v_6 = 0$$

$$u_2 + v_6 = 0; 0 + u_2 = 0; u_2 = 0$$

$$u_2 + v_2 = 4; 0 + v_2 = 4; v_2 = 4$$

$$u_2 + v_5 = 3; 0 + v_5 = 3; v_5 = 3$$

$$u_4 + v_5 = 1; 3 + u_4 = 1; u_4 = -2$$

	$v_1=3$	$v_2=4$	$v_3=8$	$v_4=13$	$v_5=3$	$v_6=0$
$u_1=0$	3[10]	20	8[40]	13[20]	4	0[10]
$u_2=0$	4	4[30]	18	14	3[10]	0[20]
$u_3=-5$	10	4	18	8[30]	6	0
$u_4=-2$	7	19	17	10	1[60]	0

Опорный план не является оптимальным, так как существуют оценки свободных клеток, для которых $u_i + v_j > c_{ij}$

$$(4;4): -2 + 13 > 10; \Delta_{44} = -2 + 13 - 10 = 1$$

Выбираем максимальную оценку свободной клетки (4;4): 10

Для этого в перспективную клетку (4;4) поставим знак "+", а в остальных вершинах многоугольника чередующиеся знаки "-", "+", "-". Цикл приведен в таблице.

	1	2	3	4	5	6	Запасы
1	3[10]	20	8[40]	13[20][-]	4	0[10][+]	80
2	4	4[30]	18	14	3[10][+]	0[20][-]	60
3	10	4	18	8[30]	6	0	30
4	7	19	17	10[+]	1[60][-]	0	60
Потребности	10	30	40	50	70	30	

Из грузов x_{ij} стоящих в минусовых клетках, выбираем наименьшее, т.е. $y = \min(2, 6) = 20$.

Прибавляем 20 к объемам грузов, стоящих в плюсовых клетках и вычитаем 20 из X_{ij} , стоящих в минусовых клетках. В результате получим новый опорный план.

	1	2	3	4	5	6	Запасы
1	3[10]	20	8[40]	13[0]	4	0[30]	80
2	4	4[30]	18	14	3[30]	0	60
3	10	4	18	8[30]	6	0	30
4	7	19	17	10[20]	1[40]	0	60
Потребности	10	30	40	50	70	30	

4. Проверим оптимальность опорного плана. Найдем потенциалы u_i, v_j по занятым клеткам таблицы, в которых $u_i + v_j = c_{ij}$, полагая, что $u_1 = 0$.

$$u_1 + v_1 = 3; 0 + v_1 = 3; v_1 = 3$$

$$u_1 + v_3 = 8; 0 + v_3 = 8; v_3 = 8$$

$$u_1 + v_4 = 13; 0 + v_4 = 13; v_4 = 13$$

$$u_3 + v_4 = 8; 13 + u_3 = 8; u_3 = -5$$

$$u_4 + v_4 = 10; 13 + u_4 = 10; u_4 = -3$$

$$u_4 + v_5 = 1; -3 + v_5 = 1; v_5 = 4$$

$$u_2 + v_5 = 3; 4 + u_2 = 3; u_2 = -1$$

$$u_2 + v_2 = 4; -1 + v_2 = 4; v_2 = 5$$

$$u_1 + v_6 = 0; 0 + v_6 = 0; v_6 = 0$$

	$v_1=3$	$v_2=5$	$v_3=8$	$v_4=13$	$v_5=4$	$v_6=0$
$u_1=0$	3[10]	20	8[40]	13[0]	4	0[30]
$u_2=-1$	4	4[30]	18	14	3[30]	0
$u_3=-5$	10	4	18	8[30]	6	0
$u_4=-3$	7	19	17	10[20]	1[40]	0

Опорный план является оптимальным. Затраты составят:

$$F(x) = 3*10 + 8*40 + 0*30 + 4*30 + 3*30 + 8*30 + 10*20 + 1*40 = 1040 \text{ (усл. ед)}$$

2.8.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными понятиями теории транспортной задачи, установили связь ТЗ с ЗЛП, выявить ее особенности (открытая и закрытая модель ТЗ);
- усвоить основные алгоритмы первичного распределения поставок;
- выработать навыки применения метода перераспределения поставок для получения оптимального решения ТЗ.

2.9 Практическое занятие № 9 (2 часа).

Тема: «Задачи нелинейной оптимизации»

2.9.1 Задание для работы:

1. Условные экстремумы функции многих переменных
2. Метод множителей Лагранжа

2.9.2 Краткое описание проводимого занятия:

1. Условные экстремумы функции многих переменных

Пример. Найти условный экстремум функции двух переменных $z = x^2 + xy + y^2 - 3x - 6y$ при условии, что аргументы этой функции удовлетворяют условию связи $y = x$. Подставляя в выражение функции $z|_{y=x}$, сведем задачу к задаче об отыскании локального экстремума функции одной переменной $z = \Phi(x) = 3x^2 - 9x$.

Так как $\Phi'(x) = 6x - 9 = 0$ при $x = \frac{3}{2}$, то точка M_0 с координатами $x_0 = \frac{3}{2}$, $y_0 = \frac{3}{2}$ является возможной точкой экстремума. Поскольку $\Phi''(x) = 6 > 0$, то это точка условного минимума.

Отметим, что точка условного экстремума не совпадает с точкой безусловного экстремума той же функции (см. пример в предыдущей лекции).

Метод сведения задачи на условный экстремум к определению безусловного экстремума функции одной переменной называется *методом исключения переменной* или *прямым методом определения точек условного экстремума*.

2. Метод множителей Лагранжа

С помощью метода множителей Лагранжа по существу устанавливаются необходимые условия, позволяющие идентифицировать точки оптимума в задачах оптимизации с ограничениями в виде равенств. При этом задача с ограничениями преобразуется в эквивалентную задачу безусловной оптимизации, в которой фигурируют некоторые неизвестные параметры, называемые Множителями Лагранжа.

Для начала рассмотрим случай для функции двух переменных.

Определение. Условным экстремумом функции $Z=F(X, Y)$ называется экстремум этой функции, достигнутый при условии, что переменные X и Y связаны уравнением $\varphi(X, Y)=0$ (уравнение связи).

Отыскание условного экстремума можно свести к исследованию на обычный экстремум так называемой функции Лагранжа $U=F(X, Y)+\lambda\varphi(X, Y)$, где λ - неопределенный постоянный множитель.

Необходимые условия экстремума функции Лагранжа имеют вид

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial f}{\partial x} + \lambda \frac{\partial \varphi}{\partial x} = 0, \\ \frac{\partial u}{\partial y} = \frac{\partial f}{\partial y} + \lambda \frac{\partial \varphi}{\partial y} = 0, \\ \varphi(x, y) = 0. \end{cases}$$

Из этой системы из трех уравнений можно найти неизвестные X, Y, λ .

Для того, чтобы найти наибольшее и наименьшее значения функции в замкнутой области, надо:

- 1) найти стационарные точки, расположенные в данной области, и вычислить значения функции в этих точках;
- 2) найти наибольшее и наименьшее значения функции на линиях, образующих границу области;
- 3) из всех найденных значений выбрать наибольшее и наименьшее.

Рассмотрим задачу минимизации функции N переменных с учетом одного ограничения в виде равенства:

Минимизировать $F(X_1, X_2, \dots, X_N)$

При ограничениях $H_1(X_1, X_2, \dots, X_N) = 0$

В соответствии с методом множителей Лагранжа эта задача преобразуется в следующую задачу безусловной оптимизации:

Минимизировать $L(X; V) = F(X) - Vh_1(X)$.

Функция $L(X; V)$ называется функцией Лагранжа, V – неизвестная постоянная, которая носит название Множителя Лагранжа. На знак V никаких требований не накладывается.

Пусть при заданном значении $V = V_0$ безусловный минимум функции $L(X; V)$ по X достигается в точке $X = X_0$ и X_0 удовлетворяет уравнению $H_1(X_0) = 0$. Тогда X_0 минимизирует $F(X_1, X_2, \dots, X_N)$ с учетом $H_1(X_1, X_2, \dots, X_N) = 0$, поскольку для всех значений X , удовлетворяющих $H_1(X_1, X_2, \dots, X_N) = 0$, $H_1(X) = 0$ И $\text{Min}L(X; V) = \text{Min} F(X)$.

Необходимо подобрать значение $V = V_0$ таким образом, чтобы координата точки безусловного минимума X_0 удовлетворяла равенству $H_1(X_1, X_2, \dots, X_N) = 0$. Это можно сделать, если, рассматривая V как переменную, найти безусловный минимум функции $L(X; V) = F(X) - Vh_1(X)$ в виде функции V , а затем выбрать значение V , при котором выполняется равенство $H_1(X_1, X_2, \dots, X_N) = 0$.

Пример. Минимизировать $f(x) = x_1^2 + x_2^2$

$$h_1(x) = 2x_1 + x_2 - 2 = 0.$$

Решение. Соответствующая задача безусловной оптимизации записывается в следующем виде:

Минимизировать $L(x; v) = x_1^2 + x_2^2 - v(2x_1 + x_2 - 2)$.

Приравняв две компоненты градиента L к нулю, получим

$$\frac{\partial L}{\partial x_1} = 2x_1 - 2v = 0 \rightarrow x_1^0 = v,$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - v = 0 \rightarrow x_2^0 = \frac{v}{2}.$$

Для того чтобы проверить, соответствует ли стационарная точка X_0 Минимуму, вычислим элементы матрицы Гессе функции $L(X; V)$, рассматриваемой как функция X ,

$$H = \begin{bmatrix} \frac{\partial^2 L}{\partial x_1^2} & \frac{\partial^2 L}{\partial x_1 \partial x_2} \\ \frac{\partial^2 L}{\partial x_1 \partial x_2} & \frac{\partial^2 L}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

Которая оказывается положительно определенной. Это означает, что $L(X; V) –$

выпуклая функция X . Следовательно, координаты $x_1^0 = v$ и $x_2^0 = \frac{v}{2}$ определяют точку Глобального минимума. Оптимальное значение V находится путем подстановки значений x_1^0, x_2^0 в уравнение $2x_1 + x_2 = 2$, откуда $2V + V/2 = 2$ или $V_0 = 4/5$. Таким образом, условный минимум достигается при

$$x_1^0 = \frac{4}{5}, \quad x_2^0 = \frac{2}{5} \quad \text{и равен } \min f(x) = \frac{4}{5}.$$

При решении задачи мы рассматривали $L(X; V)$ как функцию двух переменных X_1 и X_2 и, кроме того, предполагали, что значение параметра V выбрано так, чтобы выполнялось ограничение. Если же решение системы

$$\frac{\partial L}{\partial x_j} = 0, \quad j = 1, 2, \dots, N,$$

В виде явных функций V получить нельзя, то значения X и V находятся путем решения следующей системы, состоящей из $N + 1$ уравнений с $N + 1$ неизвестными:

$$\frac{\partial L}{\partial x_j} = 0, \quad j = 1, 2, \dots, N, \quad h_1(x) = 0.$$

Для нахождения всех возможных решений данной системы можно использовать численные методы поиска. Для каждого из решений ($X_0; V_0$) следует вычислить элементы матрицы Гессе функции L , рассматриваемой как функция X , и выяснить, является ли эта Матрица положительно определенной (локальный минимум) или отрицательно определенной (локальный максимум).

Пример . Минимизировать $f(x) = x_1 + x_2$

При ограничении $x_1^2 + x_2^2 = 1$.

Решение.

$$L(x; v) = x_1 + x_2 - v(x_1^2 + x_2^2 - 1), \quad \frac{\partial L}{\partial x_2} = 1 - 2vx_2 = 0,$$

$$\frac{\partial L}{\partial x_1} = 1 - 2vx_1 = 0, \quad h_1(x) = x_1^2 + x_2^2 - 1 = 0.$$

Эта система трех уравнений с тремя неизвестными имеет два решения:

$$(x^{(1)}, v_1) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}; -\frac{1}{\sqrt{2}} \right),$$

$$(x^{(2)}, v_2) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}; \frac{1}{\sqrt{2}} \right).$$

Матрица Гессе функции $L(X; V)$, рассматриваемой как функция X , равна

$$H_L(x, v) = \begin{bmatrix} -2v & 0 \\ 0 & -2v \end{bmatrix}.$$

Вычислив элементы матрицы H для каждого из двух решений, находим, что матрица

$$H_L(x^{(1)}; v_1) = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

положительна определена,

$$H_L(x^{(2)}; v_2) = \begin{bmatrix} -\sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{bmatrix}$$

А матрица

Следовательно, $(X(2); V2)$ соответствует максимуму функции L , рассматриваемой как функция X ; оптимальное решение

$$x_1^0 = x_2^0 = \frac{1}{\sqrt{2}}.$$

Заметим, что $(X(1); V1)$ соответствует минимуму L .

Здесь необходимо подчеркнуть, что если мы рассмотрим L как функцию трех переменных, а именно переменных x_1 , x_2 и V , то точки $(X(1); V1)$ и $(X(2); V2)$ не окажутся точками минимума или максимума L как функции X и V . На самом деле они являются **седловыми точками** функции $L(X; V)$.

2.9.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными понятиями теории нелинейной оптимизации;
- усвоить основные алгоритмы нахождения условных экстремумов функции многих переменных;
- выработать навыки применения метода множителей Лагранжа.

2.10 Практическое занятие № 10 (2 часа).

Тема: «Задачи нелинейной оптимизации»

2.10.1 Задание для работы:

1. Метод градиентного спуска

2.10.2 Краткое описание проводимого занятия:

1. Метод градиентного спуска

Метод градиента в чистом виде формирует шаг по переменным как функцию от градиента $F(x)$ В текущей точке поиска. Простейший алгоритм поиска $\text{Min } F(X)$ записывается в векторной форме следующим образом:

$$x^{i+1} = x^i - h \cdot \text{grad}f(x^i), \quad \text{или в скалярном виде:} \quad x_j^{i+1} = x_j^i - h \cdot \frac{df}{dx_j}, \quad j = 1, \dots, n.$$

Величина рабочего шага в направлении градиента $H \text{Grad } F(x)$ Зависит от величины градиента, который заранее учесть трудно, и от коэффициента пропорциональности шага H , с помощью которого можно управлять эффективностью метода.

Поиск каждой новой точки состоит из двух этапов:

1) оценка градиента $F(X)$ Путем вычисления частных производных от $F(x)$ По каждой переменной X_j ;

2) рабочий шаг по всем переменным одновременно.

Величина H сильно влияет на эффективность метода. Большой эффективностью обладает вариант метода, когда шаг по каждой переменной определяется направляющими косинусами градиента

$$x_j^{i+1} = x_j^i - h \cdot \cos \varphi_j, \quad \text{где} \quad \cos \varphi_j = \frac{(df/dx_j)}{\|\text{grad}f(x)\|}.$$

В этом случае величина рабочего шага не зависит от величины модуля градиента, и ее легче управлять изменением H . В районе оптимума может возникать значительное "рыскание", поэтому используют различные алгоритмы коррекции H .

Наибольшее распространение получили следующие алгоритмы:

1) $h^i = \text{const} = h$ (без коррекции);

$$2) \quad h^i = \begin{cases} \frac{h^{i-1}}{2}, & \text{если } f(x^i) < f(x^{i-1}), \\ h^{i-1}, & \text{если } f(x^i) > f(x^{i-1}), \end{cases}$$

$$3) \quad h^i = \begin{cases} h^{i-1}, & \text{если } \alpha_1 \leq \alpha \leq \alpha_2, \\ 2h^{i-1}, & \text{если } \alpha_1 > \alpha, \\ \frac{h^{i-1}}{3}, & \text{если } \alpha_2 < \alpha, \end{cases}$$

Где α - угол между градиентами на предыдущем и текущем шаге; α_1 и α_2 – заданные пороговые значения выбираются субъективно (например, $\alpha_1=\pi/6$, $\alpha_2=\pi/3$).

Вдали от оптимума направление градиента меняется мало, поэтому шаг можно увеличить (второе выражение), вблизи от оптимума направление резко меняется (угол между градиентами $f(x)$ большой), поэтому h сокращается (третье выражение).

Для оценки частных производных используются разностные методы:

• алгоритм с центральной пробой

$$\frac{df}{dx_i} \approx \frac{f(x_1, \dots, x_i + g_i, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{g_i},$$

• алгоритм с парными пробами

$$\frac{df}{dx_i} \approx \frac{f(x_1, \dots, x_i + g_i, \dots, x_n) - f(x_1, \dots, x_i - g_i, \dots, x_n)}{g_i},$$

Где g_i – пробный шаг по i -й переменной, выбираемый достаточно малым для разностной оценки производной.

Первый алгоритм требует меньших затрат по сравнению со вторым (обычно затраты выражаются количеством вычислений критерия оптимальности), но позволяет получить решение менее точно, чем второй, и эта погрешность зависит от величины пробного шага.

Условием окончания поиска может являться малость модуля градиента $F(X)$, т. е. $|Grad f(X)| < \varepsilon$.

Пример . Требуется найти минимум функции $F(X, Y) = X^3 + 2Y^2 - 3X - 4Y$, завершив вычисления при погрешности $\varepsilon = 0,01$, выбрав начальное приближение $X(0) = -0,5$ и $Y(0) = -1$, коэффициент шага $H = 0,1$.

Решение. Необходимые начальные данные приведены в условии задачи. Для вычислений выберем работу с шагом Без коррекции ($H = Const$). Найдем частные

производные функции: $\frac{\partial f}{\partial x} = 3x^2 - 3$, $\frac{\partial f}{\partial y} = 4y^2 - 4$. Следовательно,

$$\frac{\partial f}{\partial x^i} = 3(x^i)^2 - 3, \quad \frac{\partial f}{\partial y^i} = 4(y^i)^2 - 4.$$

Значит, $|grad f(x^i, y^i)| = \sqrt{\left(\frac{\partial f}{\partial x^i}\right)^2 + \left(\frac{\partial f}{\partial y^i}\right)^2}$. Переменные определяются по формулам:

$$x^{i+1} = x^i - h \cdot \frac{\partial f}{\partial x^i}, \quad y^{i+1} = y^i - h \cdot \frac{\partial f}{\partial y^i}.$$

Результаты вычислений занесем в табл. 1

Таблица 1

X	Y	F(x, y)	Df/dx	Df/dy	grad f	
2	3	4	5	6	7	
-0,500	-1,000	7,3750	-2,2500	-8,0000	8,3104	
-0,275	-0,200	1,6842	-2,7731	-4,8000	5,5435	
0,002	0,280	-0,9701	-3,0000	-2,8800	4,1586	
0,302	0,568	-2,5061	-2,7258	-1,7280	3,2274	
0,575	0,741	-3,4003	-2,0085	-1,0368	2,2603	
0,776	0,844	-3,8120	-1,1947	-0,6221	1,3469	
0,895	0,907	-3,9508	-0,5958	-0,3732	0,7031	
0,955	0,944	-3,9877	-0,2651	-0,2239	0,3471	
0,981	0,966	-3,9967	-0,1111	-0,1344	0,1744	
0	0,992	-3,9990	-0,0453	-0,0806	0,0925	
1	0,997	-3,9997	-0,0183	-0,0484	0,0517	
2	0,999	-3,9999	-0,0073	-0,0290	0,0299	
3	1,000	0,996	-4,0000	-0,0029	-0,0174	0,0177
4	1,000	0,997	-4,0000	-0,0012	-0,0104	0,0105
5	1,000	0,998	-4,0000	-0,0005	-0,0063	0,0063

В последней точке модуль градиента меньше заданной погрешности ($0,0063 < 0,01$), поэтому поиск прекращается.

Итак, $(x^*, y^*) \approx (1, 0.998)$ и $f^* \approx -4$.

2.10.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными понятиями теории нелинейной оптимизации;
- усвоить основные алгоритмы нахождения условных экстремумов функции многих переменных;
- выработать навыки применения метода градиентного спуска.

2.11 Практическое занятие № 11 (2 часа).

Тема: «Основы теории массового обслуживания»

2.11.1 Задание для работы:

1. Системы массового обслуживания (СМО), основные понятия и свойства.
Классификация СМО
2. Практическое применение теории СМО

2.11.2 Краткое описание проводимого занятия:

1. СМО, основные понятия и свойства. Классификация СМО

При исследовании операций часто приходится сталкиваться с системами, предназначенными для многоразового использования при решении однотипных задач. Возникающие при этом процессы получили название процессов обслуживания, а системы – систем массового обслуживания (СМО).

Главная особенность процессов массового обслуживания – случайность. При этом имеются две взаимодействующие стороны – обслуживаемая и обслуживающая.

Примерами процессов этого типа являются:

- 1) обслуживание покупателей в сфере розничной торговли;
- 2) транспортное обслуживание;
- 3) медицинское обслуживание населения;
- 4) ремонт аппаратуры, машин, механизмов, находящихся в эксплуатации;
- 5) обработка документов в системе управления;
- 6) туристическое обслуживание.

Неотъемлемой частью системы массового обслуживания является узел обслуживания, через который осуществляется взаимодействие входного и выходного потоков заявок. В случае транспортного обслуживания каналом может считаться отдельная единица транспортного средства.

Вид графической модели зависит как от числа каналов n , так и от допустимой длины очереди m . По указанным признакам различается ряд типов СМО.

Типы систем массового обслуживания

№ п/п	Параметры СМО		Тип СМО
	n	m	

1	1	0	Одноканальная, без очереди
2	$n > 1$	0	Многоканальная, без очереди
3	1	$1 < m < \infty$	Одноканальная, с ограниченной очередью
4	$n > 1$	$1 < m < \infty$	Многоканальная, с ограниченной очередью
5	1	$m = \infty$	Одноканальная, с неограниченной очередью
6	$n > 1$	$m = \infty$	Многоканальная, с неограниченной очередью

По числу обслуживающих каналов различают одноканальные и многоканальные СМО.

Находящиеся в СМО заявки могут либо ожидать обслуживания, либо находиться под обслуживанием. Часть заявок, ожидающих обслуживания, образует очередь.

В зависимости от целочисленного значения m используются следующие названия в классификации типов СМО:

- 1) $m = 0$ – без очереди;
- 2) $m > 0$ – с очередью.

Если число мест в очереди m является конечным, то в СМО могут происходить отказы в предоставлении обслуживания некоторым заявкам. В связи с этим СМО указанного типа называются системами с отказами. Отклоняются от обслуживания те заявки, в момент прихода которых все места в очереди случайно оказались занятыми, или, если $m = 0$, все каналы оказались занятыми. Считается, что заявка, получившая отказ в обслуживании, навсегда теряется для СМО. Таким образом, пропускная способность СМО этого типа всегда меньше 100 %.

Если m не ограничено, что иногда условно записывают как $m = \infty$, то соответствующая СМО называется системой с ожиданием. В СМО данного типа пришедшая заявка при отсутствии возможности немедленного обслуживания ожидает обслуживания, какой бы длинной ни были очередь и продолжительность времени ожидания.

2. Практическое применение теории СМО

Некоторые обозначения, применяемые в теории массового обслуживания, для формул:

n – число каналов в СМО;

λ – интенсивность входящего потока заявок $\Pi_{\text{вх}}$;

v – интенсивность выходящего потока заявок $\Pi_{\text{вых}}$;

μ – интенсивность потока обслуживания $\Pi_{\text{об}}$;

ρ – показатель нагрузки системы (трафик);

m – максимальное число мест в очереди, ограничивающее длину очереди заявок;

i – число источников заявок;

p_k – вероятность k -го состояния системы;

p_0 – вероятность простояния всей системы, т. е. вероятность того, что все каналы свободны;

$p_{\text{систем}}$ – вероятность принятия заявки в систему;

$p_{\text{отк}}$ – вероятность отказа заявке в принятии ее в систему;

$p_{\text{об}}$ – вероятность того, что заявка будет обслужена;

A – абсолютная пропускная способность системы;

Q – относительная пропускная способность системы;

$\bar{N}_{\text{оч}}$ – среднее число заявок в очереди;

$\bar{N}_{\text{об}}$ – среднее число заявок под обслуживанием;

$\bar{N}_{\text{систем}}$ – среднее число заявок в системе;

$\bar{T}_{\text{оч}}$ – среднее время ожидания заявки в очереди;

$\bar{T}_{\text{об}}$ – среднее время обслуживания заявки, относящееся только к обслуженным заявкам;

$\bar{T}_{\text{сис}}$ – среднее время пребывания заявки в системе;

$\bar{T}_{\text{ож}}$ – среднее время, ограничивающее ожидание заявки в очереди;

K – среднее число занятых каналов.

Абсолютная пропускная способность СМО А – среднее число заявок, которое может обслужить система за единицу времени.

Относительная пропускная способность СМО Q – отношение среднего числа заявок, обслуживаемых системой в единицу времени, к среднему числу поступающих за это время заявок.

При решении задач массового обслуживания необходимо придерживаться нижеприведенной последовательности:

- 1) определение типа СМО по табл. 4.1;
- 2) выбор формул в соответствии с типом СМО;
- 3) решение задачи;
- 4) формулирование выводов по задаче.

Вариант выбирается следующим образом: две последние цифры зачетной книжки студента делятся с остатком на количество вариантов, представленных в таблицах. К остатку от деления прибавляется единица. Полученное число является номером варианта для информации соответствующего вида.

Задача 1. На сортировочную станцию прибывают составы с интенсивностью 0,9 состава в час. Среднее время обслуживания одного состава 0,7 часа. Определить показатели эффективности работы сортировочной станции: интенсивность потока обслуживаний, среднее число заявок в очереди, интенсивность нагрузки канала (трафик), вероятность, что канал свободен, вероятность, что канал занят, среднее число заявок в системе, среднее время пребывания заявки в очереди, среднее время пребывания заявки в системе (табл. 1).

Таблица 1 Исходные данные для решения задачи 1

Показатель	Вариант									
	1	2	3	4	5	6	7	8	9	10
λ	0,5	0,8	0,4	0,6	0,7	0,5	0,7	0,6	0,8	0,4
$\bar{T}_{\text{об}}$	0,3	0,5	0,6	0,9	0,2	0,2	0,4	0,8	0,3	0,5

Решение. Сортировочную станцию можно рассматривать как одноканальную СМО с неограниченным ожиданием (т. е. с очередью). Таким образом, параметры системы: число каналов $n = 1$, число мест в очереди $m = \infty$.

Интенсивность входящего потока $\lambda = 0,9$ состава в час, среднее время обслуживания одной заявки $\bar{T}_{\text{об}} = 0,7$ ч, интенсивность потока обслуживаний

$$\bar{T}_{\text{об}} = \frac{1}{\mu}, \quad (4.1)$$

$\mu = 1/0,7 = 1,429$. Таким образом, нагрузка системы

$$\rho = \frac{\lambda}{\mu} = \lambda \bar{T}_{\text{об}}, \quad (4.2)$$

$$\rho = 0,9/1,429 = 0,63, \text{ или } \rho = 0,9 \cdot 0,7 = 0,63.$$

Среднее число составов, ожидающих обслуживания,

$$\bar{N}_{04} = \frac{\rho^2}{1-\rho}, \quad (4.3)$$

$$\bar{N}_{04} = 0,63^2/(1 - 0,63) = 1,073.$$

Так как $\rho < 1$, то очередь составов на сортировку не может бесконечно возрастать, значит, предельные вероятности существуют. Вероятность того, что станция свободна p_0 , рассчитывается по следующей формуле:

$$p_k = \rho^k(1 - \rho); k = 0, 1, 2, \dots \\ p_0 = 1 - \rho. \quad (4.4)$$

$$p_0 = 1 - 0,63 = 0,37, \text{ тогда вероятность того, что станция занята } p_{\text{зан}} = 1 - 0,37 = 0,63.$$

Среднее число заявок (составов) в системе (на сортировочной станции) рассчитывается по следующей формуле:

$$\bar{N}_{\text{сист}} = \bar{N}_{04} + \bar{N}_{06} = \frac{\rho}{1-\rho}, \quad (4.5)$$

$$\text{где } \bar{N}_{06} = \rho; \bar{N}_{\text{сист}} = 0,63/1 - 0,63 = 1,703 \text{ или } \bar{N}_{\text{сист}} = 0,63 + 1,073 = 1,703.$$

Среднее время пребывания заявки (состава) в очереди (в ожидании сортировки)

$$\bar{T}_{04} = \frac{\bar{N}_{04}}{\lambda} = \frac{\rho^2}{\lambda(1-\rho)} = \frac{\rho}{\mu(1-\rho)}, \quad (4.6)$$

$$\bar{T}_{04} = 1,073/0,63 = 0,63^2/(0,9(1 - 0,63)) = 0,63/(1,429(1 - 0,63)) = 1,19.$$

Среднее время пребывания заявки (состава) в системе (на сортировочной горке под обслуживанием в ожидании обслуживания)

$$\bar{T}_{\text{сист}} = \bar{T}_{04} + \bar{T}_{06} = \frac{1}{\lambda} \cdot \bar{N}_{\text{сист}} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu(1-\rho)}, \quad (4.7)$$

$$\bar{T}_{\text{сист}} = 0,7 + 1,19 = 0,63/(0,9(1 - 0,63)) = 1,703/0,9 = 1/(1,429(1 - 0,63)) = 1,89.$$

Вывод. Очевидно, что скорость обслуживания составов на сортировочной станции невысокая, так как время на ожидание обслуживания (1,19 ч) превышает время на обслуживание (0,7 ч). Для повышения эффективности работы сортировочной горки необходимо уменьшить время обслуживания одного состава или увеличить число сортировочных станций.

Задача 2. Интенсивность потока пассажиров в кассах железнодорожного вокзала составляет $\lambda = 1,35$ чел. в мин. Средняя продолжительность обслуживания кассиром одного пассажира $\bar{T}_{06} = 2$ мин. Определить минимальное количество кассиров $n = n_{\min}$, при котором очередь не будет расти до бесконечности, и соответствующие характеристики обслуживания при $n = n_{\min}$ (вероятность того, что в узле расчета отсутствуют покупатели, вероятность очереди, среднее число заявок находящихся в очереди, среднее время пребывания заявки в очереди, среднее число заявок, находящихся в системе, среднее время пребывания заявки в системе, доля занятых обслуживанием кассиров, абсолютная пропускную способность).

Таблица 2 Исходные данные для решения задачи 2

Показатель	Вариант									
	1	2	3	4	5	6	7	8	9	10

λ	1,37	1,62	1,42	1,83	1,75	1,55	1,4	1,65	1,7	1,3
$\bar{T}_{\text{об}}$	2,3	2	1	2,5	1,5	1,7	1,2	2,6	1	2,5

Указание. Прежде чем использовать формулы предельных вероятностей, необходимо быть уверенным в их существовании, ведь в случае, когда время $t \rightarrow \infty$, очередь может неограниченно возрастать. Доказано, что если $\rho < 1$, т. е. среднее число приходящих заявок меньше среднего числа обслуженных заявок (в единицу времени), то предельные вероятности существуют. Если $\rho \geq 1$, очередь растет до бесконечности. Очередь не будет возрастать до бесконечности при условии $\rho/n < 1$, т. е. при $n > \rho$.

Решение. $n > 1$, $m = \infty$, т. е. имеем многоканальную систему с неограниченной очередью. По условию $\lambda = 1,35$ (1/мин). Показатель нагрузки системы определяется по формуле (4.2): $\rho = 1,35 \cdot 2 = 2,7$.

Очередь не будет возрастать до бесконечности при условии $\rho/n < 1$, т. е. при $n > \rho = 2,7$. Таким образом, минимальное количество контролеров-кассирующих $n_{\min} = 3$.

Найдем характеристики обслуживания СМО при $n_{\min} = 3$.

Вероятность того, что в узле расчета отсутствуют покупатели, определяется по формуле

$$p_0 = \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} + \dots + \frac{\rho^{n+1}}{n!(n-\rho)} \right)^{-1}, \quad (4.8)$$

$p_0 = (1 + 2,7 + 2,7^2/2! + 2,7^3/3! + 2,7^4/3! \cdot (3 - 2,7))^{-1} = 0,025$, т. е. в среднем 2,5 % времени контролеры-кассиры будут простоять.

Вероятность того, что в узле расчета будет очередь, определяется по формуле

$$P_{\text{оч}} = \frac{\rho^{n+1}}{n!(n-\rho)} p_0, \quad (4.9)$$

$$P_{\text{оч}} = (2,7^4/3!(3 - 2,7)) \cdot 0,025 = 0,735.$$

Среднее число покупателей, находящихся в очереди, определяется по формуле

$$\bar{N}_{\text{оч}} = \frac{\rho^{n+1} p_0}{n \cdot n! \left(1 - \frac{\rho}{n}\right)^2}, \quad (4.10)$$

$$\bar{N}_{\text{оч}} = (2,7^4/3 \cdot 3!(1 - 2,7/3)^2) \cdot 0,025 = 7,35.$$

Среднее время ожидания в очереди определяется по формуле

$$\bar{T}_{\text{оч}} = \frac{1}{\lambda} \cdot \bar{N}_{\text{оч}}, \quad (4.11)$$

$$\bar{T}_{\text{оч}} = 7,35/1,35 = 5,44 \text{ мин.}$$

Среднее число покупателей в узле расчета определяется по формуле

$$\bar{N}_{\text{систем}} = \bar{N}_{\text{оч}} + p_0, \quad (4.12)$$

$$\bar{N}_{\text{систем}} = 7,35 + 2,7 = 10,05.$$

Среднее время нахождения покупателей в узле расчета определяется по формуле

$$\bar{T}_{\text{систем}} = \frac{1}{\lambda} \cdot \bar{N}_{\text{систем}}, \quad (4.13)$$

$$\bar{T}_{\text{сис}} = 10,05 / 1,35 = 7,44 \text{ мин.}$$

Среднее число контролеров-кассиров, занятых обслуживанием покупателей, определяется по формуле

$$\bar{K} = \bar{N}_{\text{об}} = \frac{\lambda}{\mu} = p, \quad (4.14)$$

$\bar{K} = 2,7$. Коэффициент (доля) занятых обслуживанием контролеров-кассиров

$\bar{K} = p/n = 2,7/3 = 0,9$. Абсолютная пропускная способность узла расчета $A = 1,35$ (1/мин), или 81 (1/ч), т. е. 81 покупатель в час.

Вывод. Анализ характеристик обслуживания свидетельствует о значительной перегрузке узла расчета при наличии трех кассиров.

Задача 3. На грузовой станции имеется два выгрузочных фронта. Интенсивность подхода составов под выгрузку составляет 0,4 состава в сутки. Среднее время разгрузки одного состава – 2 суток. Приходящий поезд отправляется на другую станцию, если в очереди на разгрузку стоят более трёх составов. Оценить эффективность работы выгрузочных фронтов грузовой станции: вероятность, что выгрузочные фронты свободны, вероятность, что состав останется без разгрузки, относительную пропускную способность, абсолютную пропускную способность, среднее число поездов, ожидающих разгрузки, среднее число заявок в системе, среднее время пребывания заявки в очереди, среднее время пребывания заявки в системе (табл. 3).

Таблица 3 Исходные данные для решения задачи 3

Показатель	Варианты									
	1	2	3	4	5	6	7	8	9	10
λ	0,5	0,9	0,5	0,3	0,6	0,8	0,9	0,4	0,6	0,5
$\bar{T}_{\text{об}}$	2	1	1,5	1,4	1,3	1,2	1,5	2	1,9	1,4

Решение. По условию задачи $n = 2$, $m = 3$, т. е. грузовая станция представляет собой многоканальную систему с ограниченной очередью. Интенсивность потока обслуживаний определяется по формуле (4.1): $\mu = 1/2 = 0,5$.

Интенсивность нагрузки канала (трафик) определяется по формуле (4.2): $p = 0,4 \cdot 2 = 0,8$.

Вероятность того, что выгрузочный фронт свободен, определяется по формуле

$$p_0 = \left(1 + \frac{p}{1!} + \dots + \frac{p^n}{n!} + \dots + \frac{p^{n+1} \left(1 - \frac{p}{n} \right)^m}{n \cdot n! \left(1 - \frac{p}{n} \right)} \right)^{-1}, \quad (4.15)$$

$$p_0 = 0,431.$$

Вероятность того, что состав будет отправлен на другую станцию, определяется по формуле

$$p_{\text{отк}} = p_{n+m} = \frac{p^{n+m}}{n^m \cdot n!} p_0, \quad (4.16)$$

$$p_{\text{отк}} = 0,009.$$

Относительная пропускная способность определяется по формуле

$$Q = 1 - p_{\text{отк}}, \quad (4.17)$$

$$Q = 1 - 0,009 = 0,991.$$

Абсолютная пропускная способность определяется по формуле

$$A = \lambda Q, \quad (4.18)$$

$$A = 0,4 \cdot 0,991 = 0,396, \text{ т. е. в среднем в сутки разгружается } 0,4 \text{ состава.}$$

Среднее число составов, ожидающих разгрузки, определяется по формуле

$$\bar{N}_{\text{оч}} = \frac{p^{n+1} p_0 \left(1 - \left(m + 1 - m \frac{p}{n} \right) \left(\frac{p}{n} \right)^m \right)}{n \cdot n! \left(1 - \frac{p}{n} \right)^2}, \quad (4.19)$$

$$\text{где } \bar{N}_{\text{оч}} = 0,21.$$

$$\text{Среднее время ожидания разгрузки определяется по формуле (4.11): } \bar{T}_{\text{оч}} = 0,21/0,4 = 0,524.$$

Среднее число занятых фронтов (среднее число заявок под обслуживанием) определяется по формуле

$$\bar{k} = p \left(1 - \frac{p^{n+m}}{n^m \cdot n!} p_0 \right), \quad (4.20)$$

$$\bar{k} = 0,77.$$

Среднее число составов, находящихся у разгрузочного фронта определяется по формуле

$$\bar{N}_{\text{системы}} = \bar{N}_{\text{оч}} + \bar{k}, \quad (4.21)$$

$$\bar{N}_{\text{системы}} = 0,21 + 0,77 = 0,98.$$

$$\text{Среднее время пребывания состава у разгрузочного фронта определяется по формуле (4.14): } \bar{T}_{\text{системы}} = 1,564/0,4 = 3,908.$$

Вывод. Среднее время пребывания состава в ожидании разгрузки на другой станции невелико.

2.11.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными понятиями и свойствами СМО, их классификацией;
- усвоить основные алгоритмы анализа СМО;
- выработать навыки построения математических моделей СМО в инженерной практике.

2.12 Практическое занятие № 12 (2 часа).

Тема: «Марковские цепи, процессы»

2.12.1 Задание для работы:

1. Простейший поток, его свойства. Классификация потоков.

2. Марковские цепи, их свойства
3. Марковские процессы в инженерной практике

2.12.2 Краткое описание проводимого занятия:

- 1. Простейший поток, его свойства. Классификация потоков** – доклады с презентациями, обсуждение
- 2. Марковские цепи, их свойства** – доклады с презентациями, обсуждение
- 3. Марковские процессы в инженерной практике** – доклады с презентациями, обсуждение

2.12.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с понятием простейшего потока, его свойствами, классификацией случайных потоков;
- усвоить основные алгоритмы анализа и построения марковских цепей, их особенности и условия применения;
- выработать навыки математического моделирования марковских цепей при решении инженерных задач.

2.13 Практическое занятие № 13 (2 часа).

Тема: «Основные понятия и методы математической обработки экспериментальных данных»

2.13.1 Задание для работы:

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

2.13.2 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 величине, получим вариационный ряд 2, 2, 3, 4, 4, 5, 5, 5, 7, 7, 7, 7, 10, 10. Найдем размах выборки $\omega=10-2=8$. 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 $b \approx \frac{\omega}{k}$. 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 ($i = 1, 2, \dots, k$).

Наряду с частотами одновременно подсчитываются также накопленные частоты $\sum_{j=1}^i n_j$, относительные частоты n_i/n и накопленные относительные частоты $\sum_{j=1}^i n_j/n$, $i = 1, 2, \dots, k$.

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 $\omega = 23,8 - 10,1 = 13,7$; тогда длина интервала группировки будет $b = 13,7/7 \approx 2$. 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 $\sum_{j=1}^i n_j$	Относительная частота n_i/n	Накопленная относительная частота $\sum_{j=1}^i n_j/n$
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 $x \leq 11$.

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.

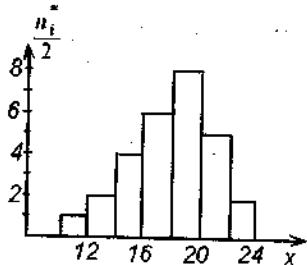


Рис.1

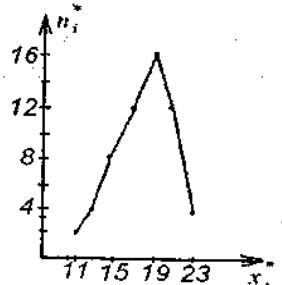


Рис.2

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

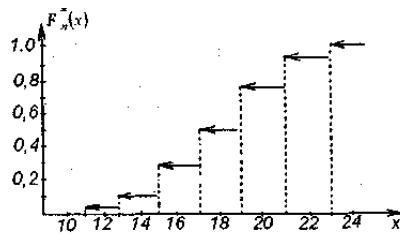


Рис.3

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 $d_x = 1$. Так как $n = 8$, то медиана $h_x = \frac{1}{2}(3+4) = 3,5$.

Итак, $\bar{x} = 3,75$, $d_x = 1$, $h_x = 3,5$.

Для упрощения вычислений выборочных среднего и дисперсии группированной выборки, эту выборку преобразуют так: $u_i = \frac{1}{b}(-d_x^* - d_x)$, $i = 1, 2, \dots, k$, где d_x^* - выборочная мода, а b - длина интервала группировки. Эти соотношения показывают, что в выборку z_1, z_2, \dots, 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, \dots, u_n можно рассматривать как выборку из генеральной совокупности $U = \frac{1}{b}(-d_x^* - d_x)$. 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:

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

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
Σ	-	-	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 $58 + 2 \cdot (-6) + 23 = 99$. Полученный результат показывает, что вычисления выполнены правильно. По формулам, данным выше, находим средние значения U

$$\bar{u} = \frac{-6}{53} \approx -0,113, D_U^* = \frac{58 - (-6)^2 / 53}{53} \approx 1,108. \text{ Далее находим средние данной выборки:}$$

$$\bar{x} \approx (-0,113) \cdot 4 + 148 \approx 147,548, D_X^* \approx 4^2 \cdot 1,103 \approx 17,728.$$

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 $i = 1, 2, \dots, 7$.

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	$n_i u_i^4$	$u_i + 1$	$(u_i + 1)^2$	$n_i (u_i + 1)^2$
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
Σ	-	-	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:

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

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

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

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

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

Варианта X_I	2	5	7	10
Частота N_i	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 *Несмешенной оценкой*. Статистическая оценка, математическое ожидание которой не равно оцениваемому параметру является *Смешенной*.

Несмешенной оценкой генеральной средней (математического ожидания) служит

$$\bar{x}_B = \frac{\sum_{i=1}^k n_i x_i}{n},$$

выборочная средняя

Где X_I – вариант выборки (элемент выборки); N_i – частота варианты X_I (число

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

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

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

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

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

$$D_B = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n}$$

Несмещенной оценкой генеральной дисперсии является «исправленная дисперсия»

$$s^2 = \frac{n}{n-1} D_B \quad \text{или} \quad s^2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n-1}$$

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 Γ -распределения с плотностью

$$f_X(x) = \begin{cases} 0, & x \leq 0 \\ \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, & x > 0 \end{cases}$$

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

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

По выборке x_1, \dots, x_n из генеральной совокупности, имеющей Γ -распределение, находим значения соответствующих выборочных моментов:

$$\alpha_1^* = \bar{x} = \frac{1}{n} \sum x_i, \quad \mu_2^* = D_X^* = \frac{1}{n} \sum (x_i - \bar{x})^2.$$

Приравнивая соответствующие равенства, получаем следующую систему уравнений:

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

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 \bar{x}_B при известном среднем квадратическом отклонении s генеральной совокупности служит доверительный интервал

$\bar{x}_B - t \frac{\sigma}{\sqrt{n}} < a < \bar{x}_B + t \frac{\sigma}{\sqrt{n}}$, где $t \frac{\sigma}{\sqrt{n}}$ – 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, $T = 1,96$.

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

Далее получаем

$$12,04 < a < 15,96$$

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

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

$$\bar{x}_B - t_{\gamma} \frac{s}{\sqrt{n}} < a < \bar{x}_B + t_{\gamma} \frac{s}{\sqrt{n}}. \text{ По таблице приложения, по заданным } N \text{ и } \gamma \text{ находим}$$

$$t_{\gamma} = 3,36.$$

Таким образом

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

$$23,38 < a < 36,82$$

Окончательно получаем

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

Значение признака X_i	-2	1	1	3	4	5
Частота N_i	2	1	2	2	2	1

$$n = \sum_{i=1}^k n_i = 2 + 1 + 2 + 2 + 2 + 1 = 10$$

Решение. Объем данной выборки равен

По данным задачи находим выборочную среднюю:

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

Далее находим исправленное среднее квадратическое отклонение S :

$$S = \sqrt{\frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n-1}} = \sqrt{\frac{2 \cdot 14,44 + 1 \cdot 0,64 + 2 \cdot 0,64 + 2 \cdot 1,44 + 2 \cdot 4,84 + 1 \cdot 10,24}{9}} = \sqrt{\frac{53,6}{9}} \approx 2,44$$

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

$$\bar{x}_B - t_{\gamma} \frac{s}{\sqrt{n}} < a < \bar{x}_B + t_{\gamma} \frac{s}{\sqrt{n}}$$

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 $0,05 < \alpha < 3,55$

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 $\gamma = 0,95$ и $n = 16$: $t_{\gamma} = 2,13$. Подставим в формулу $s = 0,8$ и $t_{\gamma} = 2,13$, вычислим границы доверительного интервала:

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

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

2.13.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с основными понятиями математической статистики, ее предметом и задачами;
- усвоить алгоритмы первичного статистического анализа экспериментальных данных;
- выработать навыки нахождения числовых характеристик статистического ряда, точечных и интервальных оценок параметров статистического распределения.

2.14 Практическое занятие № 14 (2 часа).

Тема: «Основные понятия и методы математической обработки экспериментальных данных»

2.14.1 Задание для работы:

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

2.14.2 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 $\sigma^2 = 4 \text{ л}^2$. 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 $\alpha = 0,05$;
- 3) в качестве статистики критерия используем оценку математического ожидания - выборочное среднее \bar{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 $U = \frac{\bar{X} - 10}{\sqrt{4/25}}$ 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 статистики критерия принадлежит критической области, гипотеза H_0 отклоняется: следует считать, что изменение конструкции двигателя привело к уменьшению расхода топлива.
- Граница \bar{x}_k критической области для исходной статистики X критерия может быть получена из соотношения $\frac{\bar{x}_k - 10}{\sqrt{4/25}} = -1,75$, откуда получаем $\bar{x}_k = 9,342$, т. е. критическая область для статистики X определяется неравенством $\bar{X} < 9,342$.

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: $0 \cdot 427 + 1 \cdot 235 + 2 \cdot 72 + 3 \cdot 21 + 4 \cdot 1 + 5 \cdot 1 = 451$

Таблица 1

Число отказов , k	Количество случаев, в которых наблюдалось k отказов, n_k	$p_k = \frac{0,6^k}{k!} e^{-0,6}$	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:

$$p_k = P[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}, k=0, 1, \dots, \text{при } \alpha=0,01.$$

Решение. Оценка параметра λ равна среднему числу отказов: $\bar{\lambda} = 451/757 \approx 0,6$. По таблице приложений (П3) с $\lambda = 0,6$ 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	$\frac{(n_k - np_k)^2}{np_k}$
0	427	416	0,291
1	235	249	0,787
2	72	75	0,120
≥ 3	23	17	2,118
-	-	-	$\chi_B^2 = 3,316$

Так как по выборке оценивался один параметр λ , то $l = 1$, число степеней свободы равно $4 - 1 = 2$. 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 C . 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 A , B и C .

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 $x_i^* = \frac{x_i + x_{i+1}}{2}$; в качестве частоты n_i варианты x_i^* принимают число вариантов, которые попали в i -й интервал. В итоге получают последовательность равноотстоящих вариантов и соответствующих им частот:

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

причем $\sum n_i = n$.

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

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

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

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 ($\Phi(z)$ 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

$$\bar{x}_B = \frac{\sum_{i=1}^k x_i n_i}{n} = 12,63$$

Решение. 1. Вычислим и выборочное среднее квадратическое отклонение

$$\sigma_B = \sqrt{\bar{x}_B^2 - (\bar{x}_B)^2} = 4,695$$

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

$$\chi^2_{\text{набл}} = \sum_{i=1}^s \frac{(n_i - n_i')^2}{n_i'}$$

которой найдем наблюдаемое значение критерия

i	n_i	n_i'	$ n_i - n_i' $	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
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				$\chi^2_{\text{набл}} = 22,2$

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 $\chi^2_{kp} (0,05; 6) = 12,6$.

Так как $\chi^2_{\text{набл}} = 22,2 > \chi^2_{kp} = 12,6$, 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 s , 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

$$\sum p_i^* = 0,36 + 0,24 + 0,16 + 0,10 + 0,07 + 0,04 + 0,03$$

Гистограмма относительных частот имеет вид:

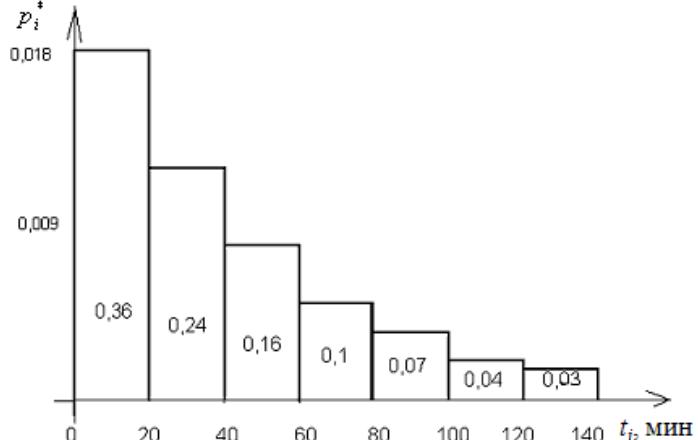
Высоты прямоугольников гистограммы равны:

$$\Delta_1 = \frac{0,36}{20} = 0,018; \quad \Delta_2 = \frac{0,24}{20} = 0,012;$$

$$\Delta_3 = \frac{0,16}{20} = 0,008; \quad \Delta_4 = \frac{0,10}{20} = 0,005;$$

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

$$\Delta_7 = \frac{0,03}{20} = 0,0015.$$



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).

$$\lambda = \frac{1}{m_i}$$

Параметр 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 :

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

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

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

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

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

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 T в каждый из семи интервалов и сравним их с соответствующими статистическими частотами p_i^* :

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

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

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

$$p_4 = P(60 < T < 80) = e^{-\frac{60}{40}} - e^{-\frac{80}{40}} = e^{-1,5} - e^{-2} \approx 0,22 - 0,14 = 0,08;$$

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

$$p_6 = P(100 < T < 120) = e^{-\frac{100}{40}} - e^{-\frac{120}{40}} = e^{-2,5} - e^{-3} \approx 0,08 - 0,05 = 0,03;$$

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

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 T .

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

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

например, на правых концах интервалов:

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

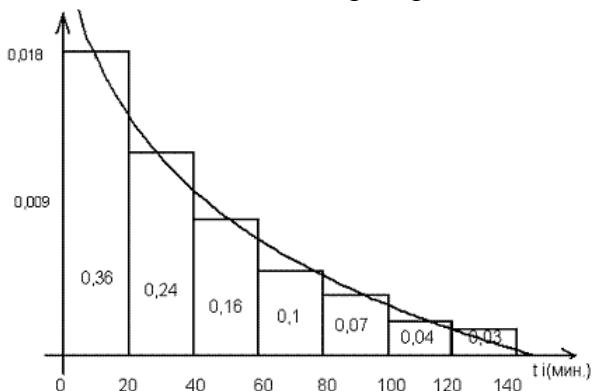
$$f(60) = \frac{1}{40} e^{-\frac{1}{40} \cdot 60} = \frac{1}{40} e^{-1,5} \approx 0,006; \quad f(80) = \frac{1}{40} e^{-\frac{1}{40} \cdot 80} = \frac{1}{40} e^{-2} \approx 0,004;$$

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

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

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.



2.14.3 Результаты и выводы:

В результате проведенного занятия студенты:

- должны ознакомиться с понятиями статистического критерия, его мощности, классификацией статистических критериев;
- усвоить алгоритмы применения различных статистических критериев;
- выработать навыки нахождения оценок параметров неизвестного распределения, выравнивания рядов.

2.15 Практическое занятие № 15 (2 часа).

Тема: «Корреляционно-регрессионный анализ»

2.15.1 Задание для работы:

1. Многомерные СВ, законы их распределения, условные числовые характеристики
2. Функция регрессии, коэффициент детерминации, корреляции, ковариация

2.15.2 Краткое описание проводимого занятия:

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

Пример 1. 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								
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 $n = 12$.

Выборочные средние отсюда находятся по формулам

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

Затем вычисляются суммы квадратов отклонений от среднего и произведений отклонений

$$\text{от средних: } Q_x = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{\sum x_i^2}{n}, \quad Q_y = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{\sum y_i^2}{n},$$

$$Q_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$\text{Отсюда } D_x^* = \frac{1}{n} Q_x, \quad D_y^* = \frac{1}{n} Q_y, \quad r = \frac{\mu_{1,1}^*}{\sqrt{D_x^* D_y^*}} = \frac{Q_{xy}}{\sqrt{Q_x Q_y}}.$$

Предварительно вычислим

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

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

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

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

Окончательно, получаем

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

Диаграмма рассеивания приведена на рис. 1.

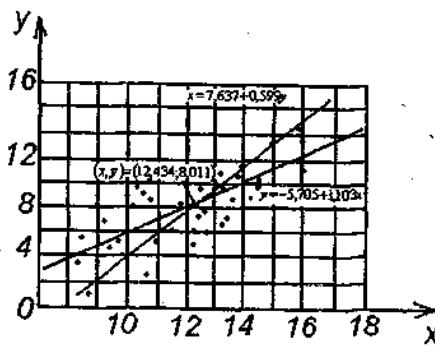


Рис. 1.

Выборочная линейная регрессия Y на X по выборке (x_i, y_i) , $i=1,2,\dots,n$, определяется

$$\text{уравнением } y = \beta_0^* + \beta_1^* x = \bar{y} + r \frac{D_Y^*}{D_X^*} (x - \bar{x}).$$

Коэффициенты β_0^* и β_1^* называются *выборочными коэффициентами регрессии*. Они

$$\text{вычисляются по формулам: } \beta_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \sum x_i^2} = \frac{Q_{xy}}{Q_x}, \quad \beta_0 = \bar{y} - \beta_1 \bar{x}.$$

Аналогично определяется выборочная линейная регрессия X на Y :

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

$$\beta_1^{**} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - \sum y_i^2} = \frac{Q_{xy}}{Q_y}, \quad \beta_0^{**} = \bar{x} - \beta_1^{**} \bar{y}.$$

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 (\bar{x}, \bar{y}) .

Пример 2. 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 $y = -5,705 + 1,103x$.

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

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

Вычисляем следующие суммы:

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

$$\text{находим } Q_u = 215 - \frac{43^2}{42} \approx 170,976, \quad Q_v = 87 - \frac{(-15)^2}{42} \approx 81,643, \quad Q_{uv} = 80 - \frac{43 \cdot (-15)}{42} \approx 95,357.$$

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

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

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

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

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	$\sum=42$	$\sum=-15$	$\sum=87$			
$n_i v_i$		-6	-6	-4	0	8	14	18	4	15	$\sum=43$					
$n_i v_i^2$		18	12	4	0	8	28	54	16	75	$\sum=215$					

2. Функция регрессии, коэффициент детерминации, корреляции, ковариация

Допустим, что в результате лечения 12 больных с артериальной гипертензией в результате суточного мониторирования sistолического 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 Р=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 Р=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:

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

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

2.15.3 Результаты и выводы:

В результате проведенного занятия студенты должны:

- ознакомиться с основными понятиями многомерного статистического анализа, теории корреляции, классификацией регрессий;
- усвоить алгоритмы нахождения условных законов и числовых характеристик многомерных случайных величин, вычисления коэффициента корреляции, детерминации, ковариации;
- выработать навыки нахождения уравнения регрессии, проверки его параметров на статистическую значимость.

2.16 Практическое занятие № 16 (2 часа).

Тема: «Корреляционно-регрессионный анализ»

2.16.1 Задание для работы:

1. Виды регрессий, статистическая значимость их параметров.
- 2 Автокорреляция

2.16.2 Краткое описание проводимого занятия:

1. Виды регрессий, статистическая значимость их параметров.

Задача:

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

<i>k</i>	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

<i>k</i>	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

<i>k</i>	21	22	23	24	25	26
<i>x_k</i>	25.90000	26.30000	26.10000	26.00000	26.40000	25.80000
<i>y_k</i>	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 $\alpha = 0.05$;
- 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:

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

$\text{cov}(X, Y)$ - ковариация случайных величин X и Y

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

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

или по формуле

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

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

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

На практике, для вычисления коэффициента корреляции чаще используется формула (1.4) т.к. она требует меньше вычислений. Однако если предварительно была вычислена ковариация $\text{cov}(X, Y)$, 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

<i>k</i>	<i>x_k</i>	<i>y_k</i>	<i>x_k²</i>	<i>y_k²</i>	<i>x_ky_k</i>
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

$$x_1 + x_2 + \dots + x_{26} = 25.20000 + 26.40000 + \dots + 25.80000 = 669.500000$$

1.2.2. Разделим полученную сумму на число элементов

$$669.50000 / 26 = 25.75000 \quad M_x = 25.75000$$

1.3. Аналогичным образом вычислим M_y .

1.3.1. Сложим последовательно все элементы y_k

$$y_1 + y_2 + \dots + y_{26} = 30.80000 + 29.40000 + \dots + 30.80000 = 793.000000$$

1.3.2. Разделим полученную сумму на число элементов выборки
 $793.00000 / 26 = 30.50000$ $M_y = 30.50000$

1.4. Аналогичным образом вычислим M_{xy} .

1.4.1. Сложим последовательно все элементы 6-го столбца таблицы 1
 $776.16000 + 776.16000 + \dots + 794.64000 = 20412.83000$

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
 $635.04000 + 696.96000 + \dots + 665.64000 = 17256.91000$

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
 $948.64000 + 864.36000 + \dots + 948.64000 = 24191.84000$

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 .

$$S_x^2 S_y^2 = 0.66481 \cdot 0.20538 = 0.136541$$

1.8. Извлечем из 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.).

$$R = (785.10885 - 25.75000 \cdot 30.50000) / 0.36951 = (785.10885 - 785.37500) / 0.36951 = -0.72028$$

ОТВЕТ: $R_{xy} = -0.720279$

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:

$$t = \frac{R_{xy} \sqrt{n - 2}}{\sqrt{1 - R_{xy}^2}}$$

(2.1)

Случайная величина t следует t-распределению Стьюдента и по таблице t-распределения необходимо найти критическое значение критерия ($t_{kp.\alpha}$) при заданном уровне значимости α . Если вычисленное по формуле (2.1) t по модулю окажется меньше чем $t_{kp.\alpha}$, 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 $t_{kp.\alpha}$.
Искомое значение $t_{kp.\alpha}$ 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 $\alpha = 0.05$, что
 соответствует критическому значению критерия $t_{kp.\alpha} = 2.064$ (см. табл. 2)

Таблица 2 **t-распределение**

Число степеней свободы (n - 2)	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	$\alpha = 0.002$	$\alpha = 0.001$
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 $t_{kp.\alpha}$

Абсолютное значение t-критерия не меньше критического $t = 5.08680$, $t_{kp.\alpha} = 2.064$, следовательно **экспериментальные данные, с вероятностью 0.95 ($1 - \alpha$), 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 $Y = a + b \cdot X$ (3.1), где:

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

$$a = M_y - b \cdot M_x \quad (3.3)$$

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 :

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

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

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)

$$b = -0.72028 \cdot 0.55582 = -0.40035$$

3.4 Вычислим коэффициент a по формуле (3.3)

$$a = 30.50000 - (-0.40035 \cdot 25.75000) = 40.80894$$

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 поместились точка $x_{15} = 26.60000$ и отчетливо различались остальные точки.

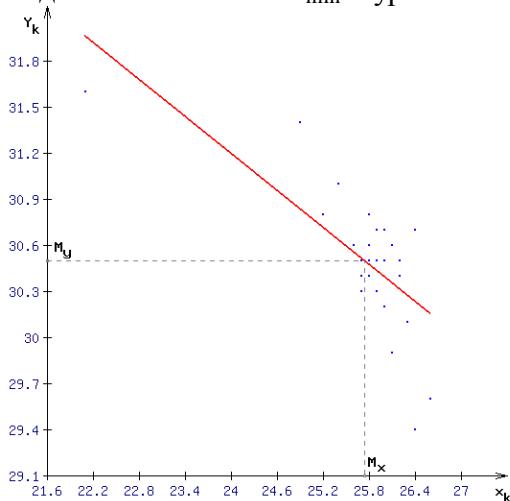
4.4. На оси ординат выбираем начальную точку чуть левее точки $y_2 = 29.40000$, и такой масштаб, чтобы на оси поместились точка $y_{18} = 31.60000$ и отчетливо различались остальные точки.

4.5. На оси абсцисс размещаем значения x_k , а на оси ординат значения y_k .

4.6. Наносим точки $(x_1, y_1), (x_2, y_2), \dots, (x_{26}, y_{26})$ 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: $(26.60000, 30.15970)$.

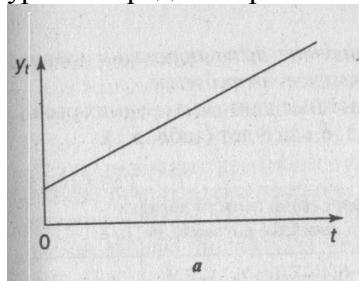
Линия регрессии показана на рисунке ниже красным цветом. Обратите внимание, что линия регрессии всегда проходит через точку средних значений величин X и Y, т.е. с координатами (M_x , M_y).

2. Автокорреляция

Временной ряд (ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

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

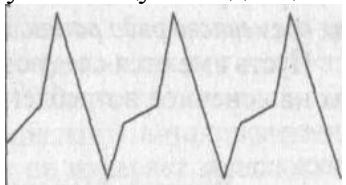
При различных сочетаниях в изучаемом явлении или процессе этих факторов зависимость уровней ряда от времени может принимать различные формы.



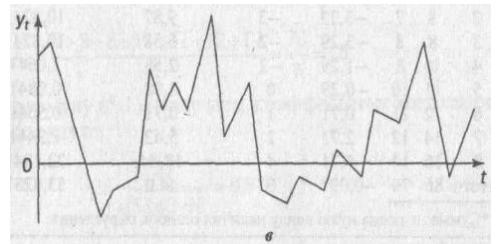
Во-первых, большинство временных рядов экономических показателей имеют тенденцию, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. Очевидно, что эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию.

Во-вторых, изучаемый показатель

может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей экономики зависит от времени года (например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним). При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка, а также с фазой бизнес-цикла, в которой находится экономика страны. Некоторые временные ряды не содержат тенденции и циклической компоненты, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты.



Очевидно, что реальные данные не следуют целиком и полностью из каких-либо описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.



В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется **аддитивной моделью** временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется **мультипликативной моделью** временного ряда. Основная задача отдельного временного ряда – выявление и приданье количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

Автокорреляция уровней временного ряда и выявление его структуры

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют **автокорреляцией уровней ряда**.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени. Рассмотрим пример.

Пример 1. Расчет коэффициентов автокорреляции уровней для временного ряда расходов на конечное потребление.

Пусть имеются следующие условные данные о средних расходах на конечное потребление (y_t , д. е.) за 8 лет (таблица 1).

Таблица 1

Расчет коэффициента автокорреляции первого порядка для временного ряда расходов на конечное потребление, д. е.

t		y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	7						
2	8	7	-3,29	-3,00	9,86	10,7959	9
3	8	8	-3,29	-2,00	6,57	10,7959	4
4	10	8	-1,29	-2,00	2,57	1,6531	4
5	11	10	-0,29	0,00	0,00	0,0816	0
6	12	11	0,71	1,00	0,71	0,5102	1
7	14	12	2,71	2,00	5,43	7,3673	4
8	16	14	4,71	4,00	18,86	22,2245	16
Итого	86	70	0	0	44	53,4286	38

Разумно предположить, что расходы на конечное потребление в текущем году зависят от расходов на конечное потребление предыдущих лет.

Определим коэффициент корреляции между рядами и измерим тесноту связи между расходами на конечное потребление текущего и предыдущего годов. Добавим в табл. 1 временной ряд .

Одна из рабочих формул для расчета коэффициента корреляции имеет вид:

$$r_{xy} = \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\sum (x_j - \bar{x})^2 \cdot \sum (y_j - \bar{y})^2}}$$

В качестве переменной x мы рассмотрим ряд y_2, y_3, \dots, y_8 , в качестве переменной y – ряд y_1, y_2, \dots, y_7 . Тогда приведенная выше формула примет вид

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \text{ где } \bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}, \bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1}$$

Эту величину называют *коэффициентом автокорреляции уровней ряда первого порядка*, так как он измеряет зависимость между соседними уровнями ряда и $t-1$, т. е. при лаге 1.

Для данных примера 1 соотношения (2) составят:

$$\bar{y}_1 = \frac{8+8+10+11+12+14+16}{7} = \frac{79}{7} = 11,29$$

$$\bar{y}_2 = \frac{7+8+8+10+11+12+14}{7} = \frac{70}{7} = 10$$

Используя формулу (1), получаем коэффициент автокорреляции первого порядка:

$$r_1 = \frac{44}{\sqrt{53,4286 \cdot 38}} = 0,977$$

Полученное значение свидетельствует об очень тесной зависимости между расходами на конечное потребление текущего и непосредственно предшествующего годов и, следовательно, о наличии во временном ряде расходов на конечное потребление сильной линейной тенденции.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями и y_{t-2} и определяется по формуле

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3) \cdot (y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \cdot \sum_{t=2}^n (y_{t-2} - \bar{y}_4)^2}}, \text{ где } \bar{y}_3 = \frac{\sum_{t=3}^n y_t}{n-2}, \bar{y}_4 = \frac{\sum_{t=3}^n y_{t-2}}{n-2}$$

Для данных из примера 1 получим:

$$\bar{y}_3 = \frac{8+10+11+12+14+16}{6} = \frac{71}{6} = 11,83$$

$$\bar{y}_4 = \frac{7+8+8+10+11+12}{6} = \frac{56}{6} = 9,33$$

Построим табл. 2.

Полученные результаты еще раз подтверждают вывод о том, что ряд расходов на конечное потребление содержит линейную тенденцию.

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Некоторые авторы считают целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции

использовать правило – максимальный лаг должен быть не больше $\left(\frac{n}{4}\right)$.

Подставив полученные значения в формулу (3), имеем:

$$r_2 = \frac{27,3333}{\sqrt{40,8333 \cdot 19,3333}} = 0,973$$

Таблица 2

Расчет коэффициента автокорреляции второго порядка для временного ряда расходов на конечное потребление, д. е.

			$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	7						
2	8						
3	8	7	-3,83	-2,33	8,9444	14,6944	5,4444
4	10	8	-1,83	-1,33	2,4444	3,3611	1,7778
5	11	8	-0,83	-1,33	1,1111	0,6944	1,7778
6	12	10	0,17	0,67	0,1111	0,0278	0,4444
7	14	11	2,17	1,67	3,6111	4,6944	2,7778
8	16	12	4,17	2,67	11,1111	17,3611	7,1111
Итого	86	56	0	-4E-15	27,3333	40,8333	19,3333

Отметим **два важных свойства коэффициента автокорреляции:**

Во-первых, он строится по аналогии с линейным коэффициентом корреляции и таким образом характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

Во-вторых, по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержит положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию. Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют **автокорреляционной функцией временного ряда**. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется **коррелограммой**.

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а следовательно, и лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т.е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ , ряд содержит циклические колебания с периодичностью в моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты (T) и циклической (сезонной) компоненты (S). Временной ряд расходов на конечное потребление, рассмотренный нами в примере 1, содержит только тенденцию, так как коэффициенты автокорреляции его уровней высокие.

2.16.3 Результаты и выводы:

В результате проведенного занятия студенты должны:

- ознакомиться с основными понятиями многомерного статистического анализа, теории корреляции, классификацией регрессий;
- усвоить алгоритмы нахождения условных законов и числовых характеристик многомерных случайных величин, вычисления коэффициента корреляции, детерминации, ковариации;
- выработать навыки нахождения уравнения регрессии, проверки его параметров на статистическую значимость.