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

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.В.ДВ.11.01 Основы научных исследований

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Профиль подготовки: Автоматизированные системы обработки информации и управления

Форма обучения: заочная

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция №1 (2 часа)

Тема: Наука в современном обществе.

1.1.1. Вопросы лекции:

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

1.1.2 Краткое содержание вопросов:

1. Основные тенденции в развитии современной науки

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

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

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

Наука поддерживается и развивается в результате исследовательской деятельности общества.

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

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

Рисунок - 1 - Структура организации научных исследований

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

2. Организация научно-исследовательской работы

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

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

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

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

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

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

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

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

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

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

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

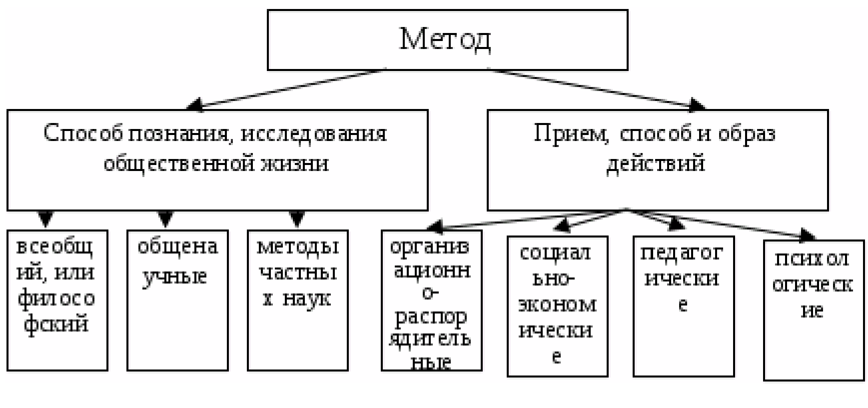


Рисунок 2 –Группы научных методов

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

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

Всеобщий, или философский метод — всеобщий метод материалистической диалектики.

К **общенаучным методам** относятся:

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

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

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

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

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

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

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

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

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

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

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

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

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

Наиболее важным для творчества видом мышления является воображение.

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

Общая схема решения научно-технических задач:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Эксперимент и теория взаимосвязаны:

теория позволяет обосновывать методику эксперимента;

эксперимент позволяет оценить справедливость теории.

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

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

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

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

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

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

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

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

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

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

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

Учебное пособие – учебное издание частично заменяющее или дополняющее учебник.

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

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

4. Научные исследования в России

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

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

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

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

5. Проблемы современной фундаментальной науки

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

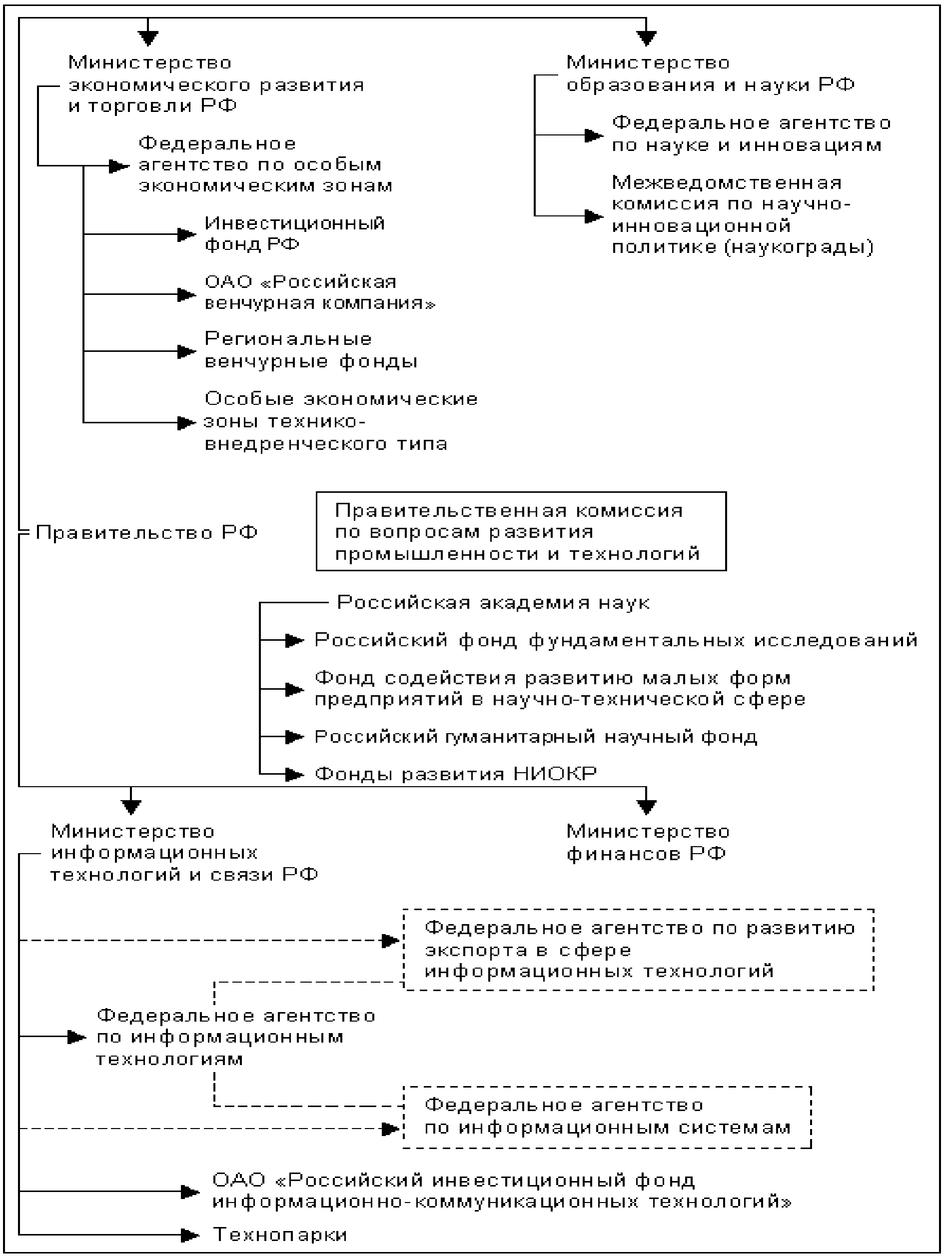


Рисунок 6 – Организационная структура науки в России

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

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

в Дубне — исследования в области ядерных технологий;

в Зеленограде — микроэлектроника;

в Санкт-Петербурге — информационные технологии;

в Томске — новые материалы.

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

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

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

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

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

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

1.2 Лекция №2 (2 часа)Тема: Стохастический метод исследования

1.2.1 Вопросы лекции:

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

1.2.2 Краткое содержание вопросов:

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

Генеральная совокупность и выборка

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

Для получения опытных данных необходимо провести обследование соответствующих объектов.

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

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

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

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

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

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

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

ВЫБОРКА

с возвратом , без возврата

(случайная повторная) (случайная бесповторная)

Вариационные ряды

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

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

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

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

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

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

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

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

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

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

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

Рассмотрим алгоритм построения интервального ряда.

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

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

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

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

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

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

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

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

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

2. Числовые характеристики выборки.

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

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

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

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

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

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

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

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

Такими же свойствами обладает и функция F(x).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Интервальные оценки параметров статистического распределения. Доверительные вероятности

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

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

Соотношению (8) тождественно соотношению

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

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

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

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

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

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

Интервальные оценки параметров нормального распределения

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

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

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

Интервальная оценка математического ожидания нормального распределения при известной дисперсии

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

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

Интервальная оценка математического ожидания такова:

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

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

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

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

Рассмотрим этапы проверки гипотезы и используемые при этом понятия.

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

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

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

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

Гипотезу H_1 называют также **альтернативной**.

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

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

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

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

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

Так вот, уровень значимости α – это вероятность ошибки первого рода, т.е. $\alpha = Đ_{í_0}(H_1)$, (13)

где $Đ_{í_0}(H_1)$ – вероятность того, что будет принята гипотеза H_1, если на самом деле в генеральной совокупности верна гипотеза H_0. Вероятность α задаётся заранее, разумеется, малым числом, поскольку это вероятность ошибочного заключения, при этом обычно используют некоторые стандартные значения: 0,05; 0,01; 0,005; 0,001. Например, α = 0,05 означает следующее: если гипотезу H_0 проверять по каждой из 100 выборок одинакового объёма, то в среднем в 5 случаях из 100 мы совершим ошибку первого рода.

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

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

Таблица 6.

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

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

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

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

Этап 3. Находят величину φ такую, что:

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

Статистический критерий. Мощность критерия

Величину φ называют **критерием**.

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

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

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

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

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

Возможны три вида расположения критической области (в зависимости от вида нулевой и альтернативной гипотез, вида и расположения критерия φ):
правосторонняя критическая область, состоящая из интервала $(x^{кр}_{пр,\ \alpha}, +\infty)$, где точка $x^{кр}_{пр,\ \alpha}$ определяется из условия

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

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

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

и называется **левосторонней критической точкой**, отвечающей уровню значимости α;
двусторонняя критическая область, состоящая из следующих двух интервалов:
$(-\infty, x^{кр}_{лев,\ \alpha/2})$

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

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

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

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

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

5. Выравнивание статистических рядов. Критерий согласия Пирсона

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

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

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

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

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

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

Таблица 7.

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

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

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

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

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

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

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

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

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

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

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

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

Далее поступаем так же, как обычно при проверке гипотез. Задаёмся уровнем значимости α. Зная распределение критерия φ, находим критическую область, как правило, это область правосторонняя, т.е. она имеет вид $(x^{\text{кр}}_{\text{пр}, \alpha}, +\infty)$; найдём числовое значение $\varphi_{\text{чис}}$ критерия (21). Если $\varphi_{\text{чис}}$ попадает в интервал $(x^{\text{кр}}_{\text{пр}, \alpha}, +\infty)$, 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 $(x^{\text{кр}}_{\text{пр}, \alpha}, +\infty)$, то гипотеза H_0 (18) не отвергается.

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

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

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1. Практическое занятие №1 (ПЗ-1) (2 часа)

Тема: Организация научно-исследовательской работы

2.1.1 Задание для работы:

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

2.1.2 Краткое описание проводимого занятия:

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

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

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

2.1.3 Результаты и выводы: В результате проведенного занятия студенты:

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

2.2. Практическое занятие №2 (ПЗ-2) (2 часа)

Тема: Стохастический метод исследования распределения

2.2.1 Задание для работы:

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

2.2.2 Краткое описание проводимого занятия:

1.Первичная обработка статистических данных. Графическое представление статистических рядов. Эмпирическая функция распределения статистических рядов.

Пример. Записать в виде вариационного и статистического рядов выборку 5, 3, 7, 10, 5, 5, 2, 10, 7, 2, 7, 7, 4, 2, 4. Определить размах выборки.

Решение. В данном случае объем выборки $n = 15$. Упорядочим элементы выборки по величине, получим вариационный ряд 2, 2, 3, 4, 4, 5, 5, 5, 7, 7, 7, 7, 10, 10. Найдем размах выборки ω=10-2= 8. Различными в заданной выборке являются элементы $z_1 = 2$, $z_2 =3$, $z_3 = 4$, $z_4 = 5$, $z_5 = 7$, $z_6 = 10$; их частоты соответственно равны $n_1 = 3$, $n_2=1$, $n_3 = 2$, $n_4 = 3$, $n_5 = 4$, $n_6 = 2$. Статистический ряд исходной выборки можно записать в виде следующей таблицы:

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

Для контроля правильности записи находим $\sum n_i = 15$. При большом объеме выборки ее элементы рекомендуется объединять в группы (разряды), представляя результаты опытов в виде *группированного статистического ряда.* В этом случае интервал, содержащий все элементы выборки, разбивается на k непересекающихся интервалов. Вычисления упрощаются, если эти интервалы имеют одинаковую длину $b \approx \frac{\omega}{k}$. 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 ($i = 1,2,...,k$).

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

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

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

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

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

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

Таблица 1

Номер интервала i	Границы интервала	Середина интервала z_i	Частота n_i	Накопленная частота $\sum_{j=1}^{i} n_j$	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 $x \le 11$. Рассуждая аналогично, находим, что $F_n^*(x) = 1$ при $x > 23$. На полуинтервале (11,23] эмпирическую функцию распределения строим по данным третьего и последнего столбцов таблицы 1.

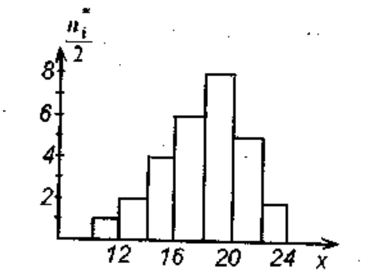


Рис.1

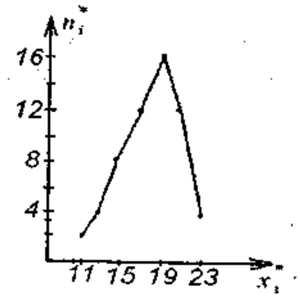


Рис.2

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

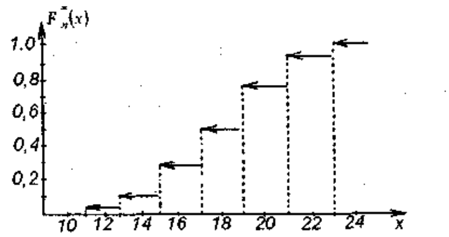


Рис.3

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

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

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

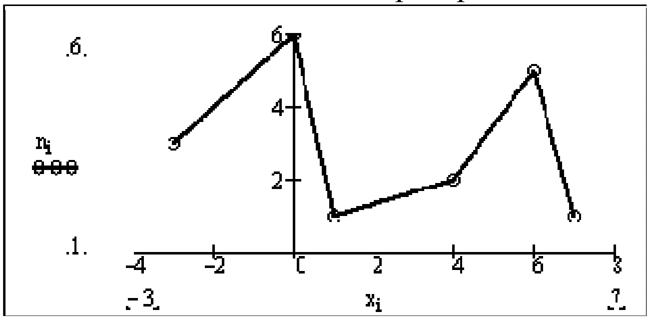


Рис. 4

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

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

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

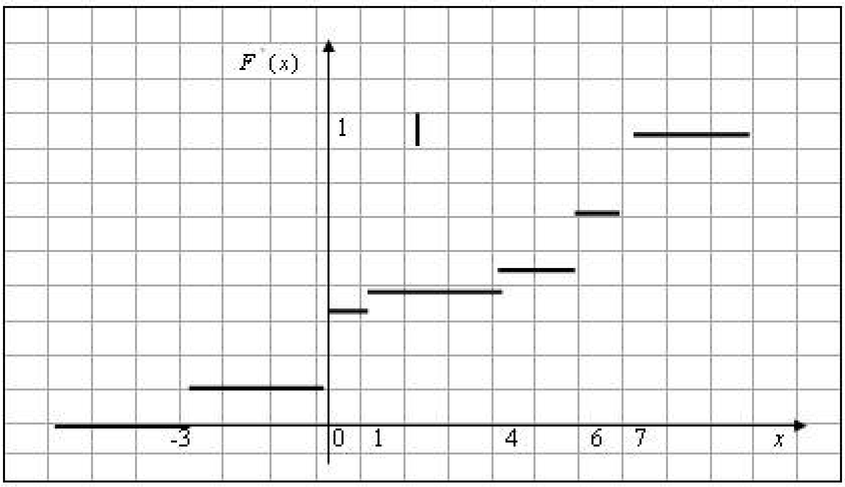
Объем данной выборки равен $n = \sum_{i=1}^{k} n_i$ *=18.*

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

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

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

График полученной эмпирической функции распределения изображен на рисунке .



2. Числовые характеристики статистического ряда, их свойства.

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

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

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

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

Пример. Вычислить среднее и дисперсию группированной выборки

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

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

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

Таблица 2

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

Вычисления сведены в таблицу 2. Последний столбец этой таблицы служит для контроля вычислений при помощи тождества $\sum n_i(u_i+1)^2 = \sum n_i u_i^2 + 2\sum n_i u_i + \sum n_i$.

Выполняя вычисления, получим 58 + 2 · (-6) + 23 = 99 .

Полученный результат показывает, что вычисления выполнены правильно. По формулам, данным выше, находим средние значения U

$\bar{u} = \frac{-6}{53} \approx -0{,}113$, $D_U^* = \frac{58 - (-6)^2/53}{53} \approx 1{,}108$. Далее находим средние данной выборки:

$\bar{x} \approx (-0{,}113) \cdot 4 + 148 \approx 147{,}548$, $D_X^* \approx 4^2 \cdot 1{,}103 \approx 17{,}728$.

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

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

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

Все вычисления оформим в виде таблицы 3.

Таблица 3

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

Контроль вычислений будет $54 + 4\cdot(-32) + 6\cdot 134 + 4\cdot(-278) + 1034 = 653$.
Находим искомые характеристики выборочного распределения:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Решение. Смещенной оценкой генеральной дисперсии служит выборочная дисперсия

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

Несмещенной оценкой генеральной дисперсии является «исправленная дисперсия»

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

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

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

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

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

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

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

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

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

Приравнивая соответствующие равенства, получаем следующую систему уравнений:

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

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

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

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

Или $12{,}04 < a < 15{,}96$

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

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

$\overline{x_B} - t_\gamma\frac{s}{\sqrt{n}} < a < \overline{x_B} + t_\gamma\frac{s}{\sqrt{n}}$. По таблице приложения, по заданным *N* и γ находим t_γ *=3,36.*

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

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

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

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

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

По данным задачи находим выборочную среднюю:

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

Далее находим исправленное среднее квадратическое отклонение *S*:

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

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

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

По таблице 3 приложения по заданным *N* и γ находим t_γ *=2,26.*

Таким образом

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

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

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

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

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

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

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

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

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

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

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

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

2.2.3 Результаты и выводы: В результате проведенного занятия студенты:

- ознакомились с основными понятиями математической статистики, ее предметом и задачами;
- усвоили алгоритмы первичного статистического анализа экспериментальных данных;
- выработали навыки нахождения числовых характеристик статистического ряда, точечных и интервальных оценок параметров статистического распределения.

2.3. Практическое занятие №3 (ПЗ-3) (2 часа)

Тема: Оптимизационные задачи

2.3.1 Задание для работы:

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

2.3.2 Краткое описание проводимого занятия:

1. Статистические критерии и их виды.

Пример(*) . По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива уменьшится. Для проверки проводятся испытания 25 случайно отобранных автомобилей с модернизированным двигателем, причем выборочное среднее расходов топлива на 100 км пробега по результатам испытаний составило $\bar{x}$ = 9,3 л. Предполагается, что выборка расходов топлива получена из нормально распределенной генеральной совокупности со средним m и дисперсией $\sigma^2 = 4$ л^2 . Используя критерий значимости, проверить гипотезу, утверждающую, что изменение конструкции двигателя не повлияло на расход топлива.

Решение. Проверяется гипотеза о среднем (m) нормально распределенной генеральной совокупности. Проверку гипотезы проведем по этапам:

1) проверяемая гипотеза $H_0{:}m = 10$, альтернативная гипотеза H_1: $m < 10$;

2) выберем уровень значимости $\alpha = 0{,}05$;

3) в качестве статистики критерия используем оценку математического ожидания - выборочное среднее $\overline{X}$;

4) так как выборка получена из нормально распределенной генеральной совокупности, выборочное среднее также имеет нормальное распределение с дисперсией $\dfrac{\sigma^2}{n} = \dfrac{4}{25}$. При условии, что верна гипотеза H_0, математическое ожидание этого распределения равно 10. Нормированная статистика критерия $U = \dfrac{\overline{X} - 10}{\sqrt{4/25}}$ имеет нормальное распределение $N(0,1)$.

5) альтернативная гипотеза H_1: $m < 10$ предполагает уменьшение расхода топлива, следовательно, нужно использовать односторонний критерий. Критическая область определяется неравенством $U < u_a$. По таблице приложений П1 находим $u_{0,05} = -u_{0,95} = -1{,}645$;

6) выборочное значение нормированной статистики критерия равно $U = \dfrac{9{,}3 - 10}{\sqrt{4/25}} = -1{,}75$.

7) статистическое решение: так как выборочное значение статистки критерия принадлежит критической области, гипотеза H_0 отклоняется: следует считать, что изменение конструкции
двигателя привело к уменьшению расхода топлива.

Граница $\overline{x_k}$ критической области для исходной статистики X критерия может быть получена из соотношения $\dfrac{\overline{x_k} - 10}{\sqrt{4/25}} = -1{,}75$, откуда получаем $\overline{x_k}$ = 9,342, т. е. критическая область для статистики X определяется неравенством $\overline{X} < 9{,}342$.

2. Критерии согласия.

Пример (**). В первых - двух столбцах таблицы 1 приведены данные об отказах аппаратуры за 10000 часов работы. Общее число обследованных экземпляров аппаратуры n=757, при этом наблюдался отказ: $0\cdot427+ 1\cdot235+ 2\cdot72 + 3\cdot21+ 4\cdot1+ 5\cdot1 = 451$

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, …, при α=0,01.

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

Таблица 2

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

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

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

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

Всего	30	40	60	130

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

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

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

Методика вычисления теоретических частот нормального распределения

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

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

причем $\sum n_i = n$.

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

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

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

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

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

и наконец находят искомые теоретические частоты $n_i' = np_i$.
Задача. Используя критерий Пирсона, при уровне значимости $a = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема $n = 200$.

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение. 1. Вычислим $\overline{x}_B = \dfrac{\sum\limits_{i=1}^{k} x_i n_i}{n} = 12{,}63$ и выборочное среднее квадратическое отклонение

$$\sigma_B = \sqrt{\overline{x_B^2} - \left(\overline{x}_B\right)^2} = 4{,}695.$$

2. Вычислим теоретические частоты учитывая, что n = 200, h = 2, σ_B = 4,695, по формуле

$$n_i = \frac{200 \cdot 2}{4{,}695} \cdot \varphi\left(\frac{x_i - \overline{x}_B}{\sigma_B}\right) = 85{,}2 \cdot \varphi\left(\frac{x_i - \overline{x}_B}{\sigma_B}\right).$$

Составим расчетную таблицу (значения функции $j(x)$ приведены в приложении 1).

i	x_i	$u_i = \dfrac{\left(x_i - \overline{x}_B\right)}{\sigma_B}$	$\varphi(u_i)$	$n_i = \dfrac{n \cdot h}{\sigma_B} \cdot \varphi(u_i)$
1	5	-1,62	0,1074	9,1
2	7	-1,20	0,1942	16,5
3	9	-0,77	0,2966	25,3
4	11	-0,35	0,3752	32,0
5	13	0,08	0,3977	33,9
6	15	0,51	0,3503	29,8
7	17	0,93	0,2589	22,0
8	19	1,36	0,1582	13,5
9	21	1,78	0,0818	7,0

3. Сравним эмпирические и теоретические частоты. Составим расчетную таблицу, из которой найдем наблюдаемое значение критерия $\chi^2_{набл} = \sum\limits_{i=1}^{s} \dfrac{\left(n_i - n_i'\right)^2}{n_i'}$:

i	n_i	n_i	$\left\lvert n_i - n_i' \right\rvert$	$\left(n_i - n_i'\right)^2$	$\dfrac{\left(n_i - n_i'\right)^2}{n_i'}$
1	15	9,1	5,9	34,81	3,8
2	26	16,5	9,5	90,25	5,5
3	25	25,3	0,3	0,09	0,0
4	30	32,0	2,0	4,0	0,1
5	26	33,9	7,9	62,41	1,8
6	21	29,8	8,8	77,44	2,6
7	24	22,0	2,0	4,0	0,2
8	20	13,5	6,5	42,25	3,1
9	13	7,0	6,0	36,0	5,1
Сумма	200				$\chi^2_{набл}$ =22,2

По таблице критических точек распределения χ^2 (приложение 6), по уровню значимости a = 0,05 и числу степеней свободы $k = s - 3 = 9 - 3 = 6$ находим критическую точку правосторонней критической области $\chi^2_{кр}$ (0,05; 6) = 12,6.

Так как $\chi^2_{набл}$ =22,2 > $\chi^2_{кр}$ = 12,6, гипотезу о нормальном распределении генеральной совокупности отвергаем. Другими словами, эмпирические и теоретические частоты различаются значимо.

3. Оценка параметров неизвестного распределения. Выравнивание рядов.

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

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

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

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

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

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

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

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

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

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

Решение

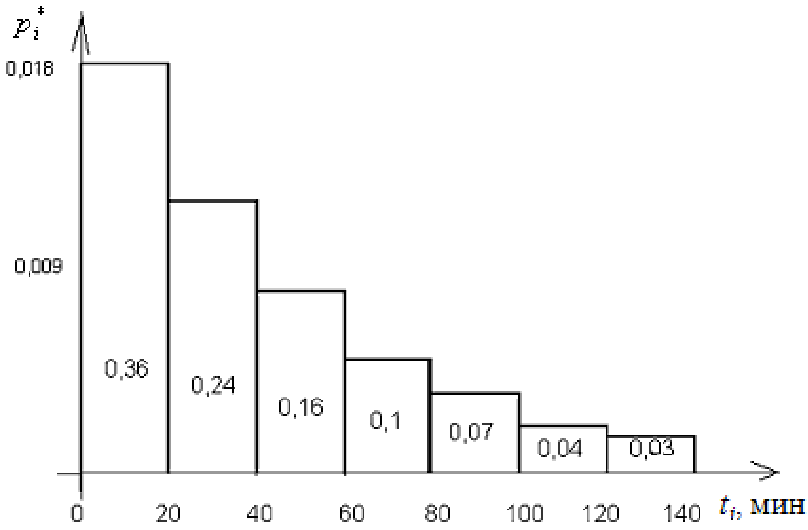
1. По данной таблице абсолютных частот построим таблицу относительных частот и соответствующую ей гистограмму:

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

.

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

Гистограмма относительных частот имеет вид:



Высоты прямоугольников гистограммы равны:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

например, на правых концах интервалов:

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

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

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

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

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

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

2.3.3 Результаты и выводы: В результате проведенного занятия студенты:

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