Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»
(ТвГТУ)

Е.В. Борисова

Экономико-статистические методы и прикладные модели в транспортном строительстве (уровень бакалавриата)

Учебное пособие
Издание первое

Тверь, 2017

Борисова Е.В. Экономико-статистические методы и прикладные модели в транспортном строительстве (уровень бакалавриата): Учебное пособие. 1-е изд. Тверь: ТвГТУ, 2017. 239 с.

В пособии изложены основные сведения использования экономико-статистических методов в области проектирования и строительства транспортных сооружений. Приведены примеры прикладных моделей предметной области. Приложение содержит справочные таблицы, необходимые для выполнения расчетов.

Предназначено для обучающихся по направлению подготовки бакалавров 08.03.01 «Строительство», профиль: автомобильные дороги и аэродромы, а также студентов других направлений и профилей, занимающихся сбором и обработкой статистических данных. Пособие может быть рекомендовано для заочной формы обучения и самообразования.

Рецензенты:

Сова А.Н. – заведующий кафедрой «Транспортные установки» Московского автодорожного института (МАДИ), доктор технических наук, профессор, действительный член (академик) Российской академии космонавтики имени К.Э. Циолковского;

Фомин К.В. – заведующий кафедрой механизации природообустройства и ремонта машин Тверского государственного технического университета (ТвГТУ), доктор технических наук, доцент.

ПРЕДИСЛОВИЕ

Введение широкой бакалаврской программы с последующей специализацией в магистратуре или на производстве больше соответствует быстро меняющемуся рынку труда. Студенты, обучающиеся по такой программе, получают высшее образование в сокращенный срок.

Дисциплина «Экономико-математические методы проектирования транспортных сооружений» изучается бакалаврами на выпускном курсе. Необходимость издания учебного пособия вызвана тем, что имеется множество книг по общей или прикладной математике, рекомендуемых студентам для изучения отдельных тем дисциплины. Это затрудняет работу не только студентов, но и преподавателей. Единственной работой по экономико-математическому моделированию является монография И.А. Золотаря «Экономико-математические методы в дорожном строительстве», изданная в 1974 году, которая сегодня является библиографической редкостью. Монография охватывает значительный круг разнообразных вопросов, представляющих интерес для исследователей. Однако, уровень изложения материала в монографии, не являющейся учебным пособием, не ориентирован на студентов, обучающихся по программе прикладного бакалавриата. В материалах данного пособия использованы некоторые идеи и материалы монографии, которая признана одной из наиболее ярких работ данного направления в дорожной отрасли. Требования к уровню подготовки в прикладном бакаларвиате предполагает, в первую очередь, ознакомление с проблематикой, основными методами, типовыми задачами и алгоритмами, что отражено в незначительном числе зачетных единиц ФГОС и формой итоговой аттестации в форме зачета.

Учебное пособие «Экономико-статистические методы и прикладные модели в проектировании транспортных сооружений (уровень бакалавриата)» подготовлено по результатам многолетнего опыта преподавания автором соответствующей дисциплины. Основной целью пособие ставит формирование у студентов владений основными методами статистической обработки и анализа данных, встречающихся в инженерных и проектных задачах. Особое внимание в пособие уделено примерам и содержательной интерпретации полученных результатов. В соответствии с Государственными требованиями к обязательному минимуму содержания и уровню подготовки выпускников высшей школы, учебное пособие ориентировано на вопросы формирования компетенций в областях обработки и анализа основных статистических показателей, умения использовать их в разработке и оценке моделей профессиональной деятельности, приобретение студентами навыков решения инженерно-экономических задач по специальности. Изложение материала в пособии базируется на знаниях, полученных в процессе изучения естественнонаучных и специальных дисциплин на предыдущих курсах. Приобретенные знания и владения необходимы для выполнения заданий по курсовому проектированию, подготовки выпускной квалификационной работы.

Учебное пособие состоит из семи разделов, содержащих основные экономико-математические методы и прикладные модели в задачах проектирования транспортных сооружений. Приведение технико-экономических и проектных решений к математическому виду существенно расширяет возможности применения операций с абстрактными объектами в области современных методов анализа эксплуатационных характеристик; оценке оптимальности, надежности и экономичности проектируемых транспортных сооружений. В первых шести разделах приведены теоретические положения и практические задачи основных экономико-математические методов. В седьмом разделе, преимущественно, рассмотрено примеры различных моделей. Такое изложение материала, по мнению автора, оправдано тем, что вначале надо усвоить сущность отдельно взятых методов, а затем научиться использовать их в комплексе. Здесь можно провести аналогию с обучением письму ребенка: прежде чем писать самостоятельное сочинение, надо освоить отдельные буквы, слова, фразы, правила. Примеры, иллюстрирующие сущность экономико-математических методов и моделей, касаются дорожного строительства, однако они близки по своей сути любому виду транспортного строительства – железнодорожному, аэродромному и другим.

Автор, при решении типовых задач, сознательно не приводит алгоритмы широко используемого сегодня пакета прикладных программ MS Excel, полагая, что стремительная скорость разработки новых программных инструментов, языков и пакетов программирования уже через несколько лет сделает эти алгоритмы мало актуальными. Знание и глубокое понимание математических основ и методов дисциплины обеспечит их переложение на любой доступный студенту программный продукт. При подготовке пособия, особенно его прикладной части, использованы материалы исследований и научных статей по вопросам моделирования объектов и процессов в строительстве и в дорожном строительстве в частности. В подборе и предварительном обсуждении вариантов моделей активное участие принимали студенты, обучающиеся по направлению подготовки «Автомобильные дороги и аэродромы» Тверского государственного технического университета.

ВВЕДЕНИЕ

Эффективность строительного производства в значительной мере определяется качеством принимаемых проектных и организационно технологических решений, управления и обеспечения строительства. Обоснование решений расценивается как важнейшая функция управления и в конце XX в. оно сформировалось как самостоятельное научное направление. Многочисленные работы отечественных и зарубежных ученых посвящены разработке теории принятия решений: критериям оценки их эффективности, оценке качества решений при наличии многих критериев, методам вероятностного планирования и т. п.

Одним из наиболее эффективных средств обоснования решений считается экономико-математическое моделирование систем и процессов. Экономико-математическое моделирование в строительстве – метод исследования строительных процессов, явлений и систем с применением экономико-математических моделей. Такое моделирование охватывает широкий круг вопросов проектирования, организации строительства, восстановления и эксплуатации транспортных сооружений, управления строительством, мероприятий обеспечения безопасности движения и дорожного сервиса, обоснования организационных структур и другие вопросы.

Практически для любой задачи организации, планирования и управления строительством характерна множественность ее возможных решений, зачастую большая неопределенность и динамичность осуществляемых процессов. В процессе разработки плана возведения объекта строительства, приходится сравнивать между собой огромное количество вариантов и выбирать из них оптимальный, по соответствующим критериям. Для предварительного анализа и поиска эффективных форм организации, а также планирования и управления строительством и используются методы моделирования.

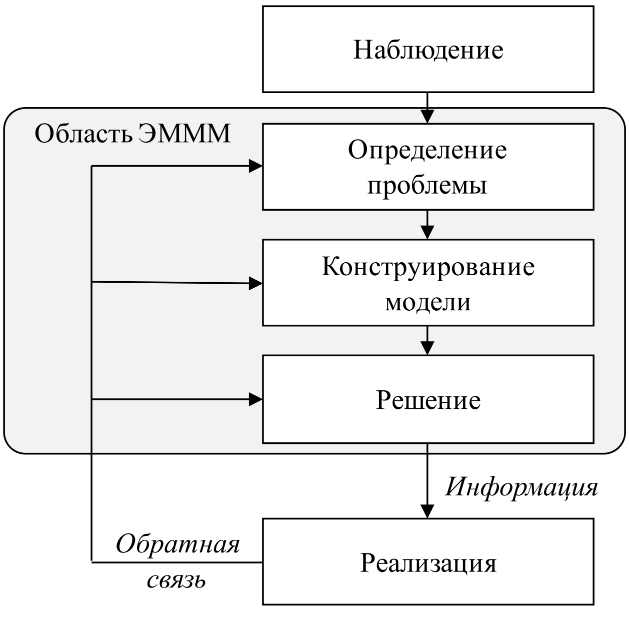
В настоящее время интенсивно ведутся исследования в области совершенствования методов моделирования, однако практика пока располагает моделями с довольно ограниченными возможностями полного адекватного отображения реальных процессов.

Одним из путей решения данной проблемы является построение локальных экономико-математических моделей. Тем не менее, возможно и полезно выделить некоторые черты и вытекающие из них общие подходы к постановке задачи и поиску наилучших решений.

Корректно составленная и предназначенная для практического использования модель должна удовлетворять двум условиям:

- адекватно отражать наиболее существенные черты анализируемого явления, процесса, системы;
- должна быть разрешима, т. е. в описывающей ее системе условий должны отсутствовать математические, экономические, технологические противоречия и иметься эффективные вычислительные алгоритмы для поиска решений. Так как экономико-математическая модель – это всего лишь постановка экономической задачи на математическом языке, то для ее решения необходимо разработать, или подобрать из существующих адекватный метод решения.

Инструментарий экономико-математических моделей и методов (ЭМММ) представляет собой логический системный подход, схематически изображенный на рисунке.



С точки зрения ЭМММ центральным моментом является конструирование модели – абстрактного представления существующей проблемной ситуации. Например, предположим, фирма продает продукт по цене 20$, а его себестоимость – 5$. Полная прибыль фирмы: $z = 20x - 5x$, где x – число проданных единиц продукта, причем x – независимая переменная, z – зависимая переменная; числа 20 и 5 – параметры. Полученное соотношение – модель определения прибыли фирмы. Далее, предположим, что продукт делается из стали и что фирма имеет 100 кг стали в своем распоряжении. На единицу продукта идет 4 кг стали. Следовательно, $4x = 100$ кг.

Теперь модель выглядит так: $\begin{cases} z = 20x - 5x; & (1) \\ 4x = 100. & (2) \end{cases}$. Здесь уравнение (1) – целевая функция, а уравнение ресурсов (2) – ограничение, то есть управленческое решение будет моделироваться так: $max\, z = 20x - 5x$, при $4x = 100$. Итак, если менеджер решает продать 25 единиц продукта ($x = 25$), фирма получит прибыль $z = 375$\$. Эта величина не действительное решение, а скорее информация, которая служит рекомендацией или руководством, помогающим менеджеру принять решение.

Историческая справка

Еще в древней Греции в экономической науке возникли два направления исследований:

- анализ методов рационального управления хозяйством;
- изучение наиболее важных экономических закономерностей.

В дальнейшем первое направление превратилось в науку о рациональном управлении хозяйством; второе направление стало основой политической экономии.

Активное применение математических методов в экономике началось в начале XX века благодаря трудам Минковского (1896 г.) и Фаркаша (1903 г.), разработавших математические основы линейного программирования; Дж. Данциг предложил основной метод решения задач линейного программирования – симплекс-метод; датский ученый Эрланг (1909-1929 гг.), разработал основы теории массового обслуживания; русский академик А.А. Марков, создал теорию ветвящихся процессов и динамического программирования, получивших впоследствии название «марковских процессов» (1909 г.).

В предвоенный период стало формироваться новое направление математического моделирования – «операционные исследования», получившее после второй мировой войны название «исследование операций». Его основоположниками являются фон Нейман и Л.В. Канторович, разработавшие экономические модели линейного программирования (1937-1939 гг.). Наиболее активные работы в области исследования операций проводились в Англии и США. Однако судьба этих работ в указанных странах оказалась различной. В Англии после войны затраты на эти цели существенно сократили, в результате многие операционисты из военной сферы были освобождены и использованы в угольной, металлургической промышленностях, на транспорте. В США наоборот, были расширены исследования в области обороны. Поворот к хозяйственному применению математических методов, в том числе методов исследования операций, был обусловлен началом второй промышленной революции в конце пятидесятых годов.

Внедрение методов исследования операций в управление предприятиями и экономические системы привело к созданию самостоятельного направления в математическом моделировании – экономико-математическое моделирование. Большой вклад в его развитие внесли отечественные ученые Л. Понтрягин, А.Н. Колмогоров, Н.П. Бусленко, Д.Б. Юдин и другие. С 60-х годов прошлого века экономико-математические методы начинают использоваться на транспорте и в транспортном строительстве. Они применялись для решения локальных задач транспортного обеспечения экономики. Разработчиками экономико-математических моделей были: Институт комплексных транспортных проблем Академии наук СССР (ИКТП), транспортные вузы.

Особое место в разработке и внедрении экономико-математических моделей в транспортном строительстве принадлежит профессору Золотарю И.А. и его научной школе, созданной в Военной академии тыла и транспорта в семидесятые-восьмидесятые годы прошлого столетия. Эти модели разрабатывались в интересах планирования и управления дорожным строительством, строительства, восстановления и технического прикрытия военно-автомобильных дорог, эксплуатации дорог и дорожных объектов.

Таким образом, в транспортном строительстве в семидесятых-восьмидесятых годах сложилась система экономико-математических моделей и методическая школа их применения для решения широкого круга задач, которые можно свести к следующим классам:

- распределения (материальных, трудовых, технических ресурсов);
- управления запасами;
- упорядочения;
- массового обслуживания;
- выбора маршрута;
- планирования производства (сетевые и другие модели);
- формирования организационных структур и другие.

Примечателен факт, что практически все первоначально разработанные модели являлись автономными, т. е. моделями, ориентированными на решение какой-либо частной задачи. Применение автономных моделей может быть оправдано тем, что для оптимизации системы надо иметь, как минимум, оптимальность ее элементов. Например, любое предприятие есть система цехов, цехи – система участков. Для того чтобы создать модель предприятия, надо построить частные (автономные) модели участков и цехов. Для построения комплексной модели эти частные модели необходимо агрегировать, однако здесь существует опасность потери некоторых связей между частными моделями, их взаимного влияния.

С другой стороны, опыт применения экономико-математических моделей показал, что стремление к усложнению модели или их агрегирования не всегда оправдано. Надо всегда помнить о том, что затраты на усложнение модели должны соизмеряться с получаемой при этом выгодой. Наибольший расцвет применения автономных экономико-математических моделей приходится на 70-е годы, когда они широко использовались в проектировании транспортного строительства. Тогда пользовались популярностью модели «перемещения земляных масс» (задачи линейного программирования), проектирование «красной линии» на продольном профиле (модели динамического программирования) и другие.

Практика применения экономико-математических моделей в транспортном строительстве в последнюю четверть прошлого века позволяет сделать следующие выводы.

Во-первых, у пользователей создавалось впечатление, что основные трудности экономико-математического моделирования лежат в вычислительной области. В действительности наиболее трудным этапом экономико-математического анализа является переход от вербального описания системы, процесса к математическому. Игнорирование данного положения приводило к описанию сложных процессов упрощенными математическими моделями с использованием какого-либо одного, наиболее известного метода, например, линейного или динамического программирования.

Во-вторых, даже математически корректно построенная модель не всегда может быть хорошей экономико-математической моделью. Необходимо исследовать вопрос о соответствии модели изучаемому экономическому явлению. До сих пор существует мнение, что главная цель экономико-математического моделирования состоит в получении оптимального решения, хотя главным является другой вопрос: можно ли данную математическую модель использовать для анализа изучаемой экономической ситуации. Отсюда и возникает проблема адекватности экономико-математических моделей, которая и сегодня остается весьма актуальной. *Адекватность* (соответствие) модели можно считать понятием условным. Здесь можно согласиться с Р. Акофом и Л.И. Лопатниковым, которые считают, что полного соответствия модели и объекта моделирования просто не может быть. Иначе это была бы не модель, а сам объект. Вместе с тем, автомобиль, например, может иметь несколько моделей в зависимости от поставленных целей:

- модель-тренажер для обучения управлению автомобилем не соответствует его форме и размерам (т. е. в этом смысле неадекватна автомобилю), однако она вполне адекватна ему по процессам управления (руль, педали, кабина полностью соответствуют условиям работы водителя при управлении автомобилем);

- модель автомобиля, построенная для макетного проектирования гаража, только внешне адекватна оригиналу, но не имеет ничего общего с ним по системам управления, питания, электрооборудования и т. п.

Обе модели могут считаться адекватными автомобилю только с позиций поставленной перед моделированием задачи.

Этот пример говорит о том, что путем математической абстракции мы часто отделяем форму решаемой задачи от ее содержания.

В-третьих, проблема критерия, используемого в модели. Вообще, в моделировании проблема критерия считается важнейшей составляющей. Большая часть моделей исследования операций была изначально рассчитана на однокритериальность, то же самое можно сказать и о моделях, применявшихся в транспортном строительстве в семидесятые-восьмидесятые годы. Как правило, это были стоимостные (экономические) критерии (стоимость строительства, перевозок материалов, простоев техники, потерь от ДТП и т. п.). При этом игнорировались организационные, социально-экономические, экологические, технологические и другие факторы, которые могли бы существенно изменить саму модель, равно как и эффективность ее использования.

В-четвертых, недоверие к экономико-математическим моделям в 70-е годы было связано с экстремальным характером моделей.

Экстремальность модели (стремление к минимуму или к максимуму) ведет к получению самого дешевого варианта перевозок (самого дешевого варианта хранения материалов и т. п.), однако в процессе строительства постоянно возникают непредвиденные ситуации, когда приходится платить за простой вагонов на железной дороге, восполнять потери из-за плохих погодных условий (простои из-за низкой температуры или дождя). Для этого необходим резерв сил и средств, который в процессе экономико-математического моделирования исключается.

В-пятых, при экономико-математическом моделировании возникает эффект «частной оптимизации». Его суть можно пояснить на примере моделей массового обслуживания.

При расчете количества самосвалов, работающих совместно с одним экскаватором, используется экстремальная модель, смысл которой установить зависимость потерь от простоев экскаватора от количества самосвалов в подразделении, потерь от простоев самосвалов и суммарных потерь. Решив задачу минимизации стоимости простоев подразделения на модели массового обслуживания, получится оптимальное количество самосвалов, равное 5 единицам. Традиционным расчетом это число равно 9 единицам. Рассмотрим задачу в более широком аспекте и заметим, что при $n_{расч} = 9$ ед. можно потерять от простоев самосвалов C рублей, однако при 9 самосвалах темп вывозки песка будет выше и работы на дороге могут быть завершены досрочно. Эффект от досрочного ввода дороги в эксплуатацию может значительно превысить экономию от оптимальности звена «экскаватор-самосвалы». Таким образом, получив частный эффект от оптимизации количества самосвалов, можно потерять в целом по строительству.

Указанные недостатки экономико-математического моделирования позволили сформулировать основные проблемы и направления дальнейшего использования математических моделей в транспортном строительстве. Их можно свести к четырем основным направлениям:

- переход от разрозненных частных моделей к их комплексному (системному) использованию;
- моделирование не только производственно-технологических процессов, но и организационных структур предприятий, организаций, территориальных транспортных объединений;
- более широкое применение экономико-математических методов в экономическом анализе производственно-хозяйственной деятельности предприятий, организаций, фирм; прогнозировании рынка строительной продукции, спроса и предложения;
- переход от детерминированного экономико-математического моделирования к вероятностному, разработке планов экономического и социального развития с заданным уровнем гарантии их реализации в установленные сроки.

Вернемся к экономико-математическим моделям и методам.

Модели, заметим, не дают рекомендаций по решению задач или проблем. Однако они обеспечивают описательные результаты, то есть описывают моделируемую систему (например, дисперсия продаж некоторых товаров по месяцам в течение года). Специалист по моделям не применяет прямо полученный результат как решение, а сопоставляет его со своими оценками и прогнозами. Если он не использует инструментарий ЭММ, то значит должны быть введены дополнительные ресурсы или усилия при решении проблемы, конструировании модели и ее решения. Результаты моделирования и решения основаны на сравнении путем обратной связи с первоначальной моделью, которая может модифицироваться при испытаниях в различных условиях и будущих решениях специалиста. Результаты могут указывать, что проблема полностью не охвачена и требует изменений или реконструкции первоначальной модели. Иначе, ЭМММ представляют непрерывный процесс, а не одиночное решение одиночной проблемы.

Среди ЭМММ наиболее популярна техника линейного программирования. К ней проводят задачи, связанные с ограничениями (по ресурсам, времени, рабочей силе, энергии, финансам, материалам) и с целевой функцией типа минимизации затрат или максимизации прибыли. При использовании вероятностных процедур, в отличие от линейного программирования, результаты носят вероятностный характер и должны содержать некоторую неопределенность и возможность присутствия альтернативных решений. Процедуры управления запасами специально разработаны для анализа проблем запасов, что характерно для большинства коммерческих фирм. Эта частная функция управления вносит существенный вклад в издержки любого бизнеса. Сетевые модели скорее более диаграммы, чем точные математические соотношения. Они представляют в наглядной форме систему действий для их анализа.

В практических задачах наибольшее значение придается:

- оптимизационным моделям;
- линейному программированию;
- графам (деревьям) решений;
- сетевым моделям;
- теории очередей (задачам массового обслуживания);
- регрессионному анализу;
- корреляционному и дисперсионному анализу.

Следует отметить, что существует определенная переоценка значимости экономико-математических моделей в реальной практике. Это связано со сложностями моделирования процессов в экономико-производственных системах из-за непрерывности изменений продукции, нерегулярности производства, внутренних дестабилизирующих факторов, нерегулярности снабжения, финансирования, сбыта и т. д. Большинство этих факторов носит нестационарный характер, что фактически исключает возможность использования ЭМММ в проектировании, планировании и управлении на реальном производстве.

По цели создания и применения различают модели:

- балансовые;
- эконометрические;
- оптимизационные;
- сетевые;
- систем массового обслуживания;
- имитационные (экспертные).

По учету фактора неопределенности модели подразделяются на:

- детерминированные (с однозначно определенными результатами);
- стохастические (с различными, вероятностными результатами).

По типу математического аппарата различают модели:

- линейного и нелинейного программирования;
- корреляционно-регрессионные;
- матричные;
- сетевые;
- теории игр;
- теории массового обслуживания.

В статических моделях объект (ситуация) описан в статике, применительно к одному определенному моменту времени. Это как бы снимок, срез, фрагмент динамической системы в какой-то момент времени. Динамические модели описывают объект (ситуацию) в развитии.

Проектные задачи могут существенно отличаться по степени сложности расчетов. Например, прочностные расчеты, определяющие степень выносливости несущих слоев, относятся к сложнейшим вычислениям. Однако помимо таких сверхсложных задач существуют и более простые (с точки зрения математики), которые чаще встречаются в практической деятельности.

К таким задачам, имеющим строго прикладной характер можно отнести – определение площади нестандартной фигуры. С этой задачей сталкиваются в основном при определении площади бассейна стока талых вод. Бассейны имеют сложную форму, основанную на сопряжении нескольких геометрических фигур: трапеции и окружности, прямоугольника и треугольника. Просчитать площадь такой фигуры очень сложно, а порой и вовсе невозможно. При этом, используя принцип деления сложной геометрической фигуры на несколько простых, можно быстро добиться нужных результатов. Достаточно вычислить площадь простой геометрической фигуры, а затем добавить или отнять от нее площадь другой фигуры, которая исказила стандартные формы при сопряжении.

Исходя из этого простого примера применения всем известных законов для прикладных целей, можно утверждать, что с помощью аксиом и формул математической области человеческих знаний можно решить теоретическую или практическую задачу.

На сегодняшний момент в транспортном строительстве сложилась система экономико-математических моделей и методическая школа их применения для решения широкого круга задач, которые можно свести к следующим основным классам:

1. Модели экстремального анализа.
2. Обоснование проектных решений с применением моделей линейного программирования.
3. Организация материального обеспечения дорожного строительства с использованием моделей управления запасами.
4. Применение регрессионных моделей.
5. Модели сетевого планирования дорожно-строительных работ.
6. Применение методов статического анализа для оценки качества строительной продукции и надежности транспортных сооружений.
7. Модели проектирования организационных структур.
8. Многокритериальный анализ качества проектных решений.

В следующих разделах рассмотрим подробнее основные методы и модели на простейших и типовых задачах проектирования, организации строительства и анализа транспортных сооружений.

Раздел 1. ПРИКЛАДНЫЕ ВОПРОСЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. Общие теоретические положения и понятия

Происходящие в природе и общественной жизни явления и события могут быть детерминированными или случайными. В детерминированных явлениях и событиях результат (исход) можно с уверенностью предсказать заранее. В качестве примера детерминированных явлений можно указать движение планет солнечной системы. На основе хорошо изученных законов этого движения можно с высокой точностью предсказать положение планет на любой момент времени. Случайное явление при каждом его воспроизведении протекает несколько по-иному. Примером случайного события является производительность какой-либо машины, выполняющей дорожные работы. Вследствие влияния многих факторов (погодных условий, технического состояния машины, квалификации и психологического состояния водителя и др.) конкретная производительность в каждом случае будет несколько различной и заранее точно предсказать ее невозможно. Однако на основе многократного повторения этой работы можно установить долю тех случаев, когда, скажем, выполняются и перевыполняются соответствующие нормативы. При этом следует иметь в виду, что эта вероятность характеризует массовый случайный процесс, дает возможность предсказать средний результат большого числа испытаний, однако исход каждого испытания остается случайным. Так, если была установлена вероятность выполнения нормы при одном испытании, равная 0.80, то это означает, что при производстве данной работы 100 раз можно ожидать 80 случаев, когда нормы будут выполнены. При этом нельзя достоверно предсказать результат любого одиночного испытания (производства работы).

Следует отметить, что на реальном производстве детерминированные процессы встречаются крайне редко, поэтому можно говорить лишь об относительной детерминированности. Поэтому умения и навыки выполнять расчеты в вероятностном пространстве и оценка их точности являются важным инструментом в руках инженера.

В теории вероятностей различают события достоверные, возможные и невозможные. Достоверное событие в результате опыта происходит обязательно. Например, на выполнение какой-либо дорожной работы будет затрачено рабочее время, и это достоверно. Вероятность достоверного события равна единице. Невозможным называют событие, которое в данном опыте не может произойти. Такие события противоположны достоверным событиям. В примере невозможным будет такое событие – выполнение работы без затраты времени. Естественно принять вероятность невозможного события равной нулю. Возможные события (в нашем примере выполнение работы с какой-то конкретной затратой времени) будут характеризоваться вероятностью большей нуля и меньшей единицы. Таким образом, диапазон изменения вероятностей любых событий – от нуля до единицы.

Комбинаторика – это раздел математики, в котором изучают, сколько комбинаций, подчинённых тем или иным условиям, можно составить из данных объектов. Известно, что для оценки вероятности победы в покере или при игре в рулетку надо просчитать, с одной стороны, общее количество возможных ходов, а с другой – сколько из них приведут к выигрышу. Неразумно садиться за карточный стол, не владея даже самой простой техникой, позволяющей вычислить, сколько, например, стрит-флешей можно собрать при колоде в 52 карты. А рассчитать значения вероятности без комбинаторики практически невозможно.

Перестановкой называется определенное линейное расположение всех элементов некоторого множества. Число перестановок из n элементов обозначается P_n и равно $P_n = n!$.

<u>Пример</u>. Найти количество перестановок букв в слове «*зачет*».

$$P_5 = 5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120.$$

Размещением называется любой набор, содержащий m элементов этого множества, взятый в определенном линейном порядке. Число размещений из n элементов по m обозначается A_n^m и равно:

$$A_n^m = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-m+1).$$

<u>Пример</u>. В группе 20 студентов. Найти число способов выбрать среди них старосту и его заместителя.

$$A_{20}^2 = 20 \cdot 19 = 380.$$

Сочетанием называется любой набор, содержащий m элементов этого множества, без учета их порядка. Число размещений из n элементов по m обозначается C_n^m и равно: $C_n^m = \dfrac{A_n^m}{P_m} = \dfrac{n!}{m!(n-m)!}$.

Пример. Найти число способов выбрать из группы в 20 студентов двух дежурных.

$$C_{20}^{2}=\frac{20\cdot 19}{1\cdot 2}=190.$$

Вероятностью события A называется математическая оценка возможности появления этого события в результате опыта. Вероятность события A равна отношению числа, благоприятствующих событию A исходов опыта к общему числу исходов опыта.

$$P(A)=\frac{m}{n}.$$

Замечание. Исход опыта является благоприятствующим событию A, если появление в результате опыта этого исхода влечет за собой появление события A.

В теории вероятностей различают полные группы событий, причем сумма вероятностей событий, составляющих такую группу, равна единице. Отсюда, в частности, следует, что противоположные события составляют полную систему, ибо сумма их вероятностей равна единице.

Допустим, что были проанализированы причины невыполнения норм выработки автомобилями-самосвалами, причем оказалось, что вследствие плохой организации выхода автомобилей из парка на линию наблюдалось 28% таких случаев, ввиду поломок – 36%.

В остальных случаях невыполнение норм обусловливалось плохой организацией погрузки и разгрузки автомобилей-самосвалов по вине производственных подразделений. Все причины невыполнения норм составляют полную группу событий, причем вероятность невыполнения норм из-за плохой работы самого транспортного подразделения составляет $0.28+0.36=0.64$. В этом случае вероятность невыполнения норм по вине производственных подразделений составляет $1-0.64=0.36$.

Большое практическое значение имеют два правила теории вероятностей, а именно: теорема сложения вероятностей и теорема их умножения.

Теорема сложения вероятностей. Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий.

$$P(A+B)=P(A)+P(B).$$

Следствие 1: Если события $A_1, A_2, \ldots, A_n$ образуют полную группу несовместных событий, то сумма их вероятностей равна единице.

$$\sum_{i=1}^{n}P(A_i)=1.$$

Теорема. Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления.

$$P(A+B)=P(A)+P(B)-P(AB).$$

Следствие 2: Сумма вероятностей противоположных событий равна единице: $P(A)+P(\overline{A})=1$.

Событие A называется *независимым* от события B, если вероятность события A не зависит от того, произошло событие B или нет. Событие A называется *зависимым* от события B, если вероятность события A меняется в зависимости от того, произошло событие B или нет.

Вероятность события B, вычисленная при условии, что имело место событие A, называется *условной вероятностью* события B.

$$P_A(B)=P(B/A)=P(AB)/P(A).$$

Теорема умножения вероятностей. Вероятность произведения двух событий (совместного появления этих событий) равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое событие уже наступило.

$$P(AB)=P(A)P(B/A)=P(A)P_A(B).$$

Если события, участвующие в произведении, являются независимыми, то:

$$P(AB)=P(A)\cdot P(B).$$

Пример. В результате статистических наблюдений установлено, что вероятность выхода из строя за месячный период работы бульдозера составляет $p(Б)=0.2$, скрепера $p(С)=0.3$ и экскаватора $p(Э)=0.15$. Найти вероятность выхода из строя в течение месяца какой-либо из этих машин?

В соответствии с правилом сложения она составит:

$$p=0.2+0.3+0.15=0.65.$$

Понятие об условной вероятности поясним на следующем примере.

Пример. Капитальный ремонт двигателей дорожных машин выполняется на двух ремонтных заводах, причем на первом заводе ремонтируется 40%, а на втором 60% всех двигателей, подвергающихся ремонту. Из каждых ста двигателей, выпускаемых первым заводом, 80 отрабатывают положенное им число часов. Для продукции второго завода этот показатель равен 70. Какова вероятность того, что двигатель, устанавливаемый на соответствующую дорожную машину, отрабатывает положенное ему число часов? Легко вычислить, что эта вероятность составит: $p=0.4\cdot 0.8+0.6\cdot 0.7=0.74$.

Поскольку в примере не введено дополнительного условия, характеризующего завод-изготовитель двигателя, то данная вероятность носит название безусловной и характеризует совокупность отремонтированных двигателей в целом. Если же уточнено, что используемый двигатель отремонтирован первым заводом, то вероятность отработки им нормативного числа часов составит 0.80. Подобная вероятность, вычисленная при условии, что имело место другое событие (в нашем примере выпуск двигателя первым заводом), называется условной. Возвращаясь к ранее рассмотренному примеру с выпуском отремонтированных двигателей двумя заводами, применим правило умножения вероятностей для определения вероятности того, что двигатель выпущен вторым заводом и отработает нормативное число часов.

Вероятность выпуска двигателя вторым заводом $p(A)$ по условию примера составляет 0.6. Вероятность кондиционности двигателя, выпущенного вторым заводом, $p(B/A)$ равна 0.7. Тогда вероятность совместного поступления событий:

$$p(AB) = 0.6 \cdot 0.7 = 0.42.$$

Вернемся к примеру, в котором рассматривалась вероятность выхода из строя бульдозера, скрепера и экскаватора. Определим вероятность безотказной работы в течение месяца звена, включающего по одной упомянутой машине. Соответствующие вероятности безотказной работы этих машин найдем по правилу полной группы событий:

$$1 - p(Б) = 1 - 0.2 = 0.8; \quad 1 - p(С) = 1 - 0.3 = 0.7; \quad 1 - p(Э) = 1 - 0.15 = 0.85.$$

Тогда искомая вероятность безотказной работы звена составит:

$$P = 0.8 \cdot 0.7 \cdot 0.85 = 0.48.$$

Важную роль в теории вероятностей играет формула полной вероятности, являющаяся следствием обоих правил – правила сложения и правила умножения вероятностей.

Пусть некоторое событие A может произойти вместе с одним из несовместных событий $H_1, H_2, \ldots, H_n$, составляющих полную группу событий. Пусть известны вероятности этих событий $P(H_1), P(H_2), \ldots, P(H_n)$ и условные вероятности наступления события A при наступлении события H_i – $P(A/H_1), P(A/H_2), \ldots, P(A/H_n)$.

Теорема (формула полной вероятности). Вероятность события A, которое может произойти вместе с одним из событий $H_1, H_2, \ldots, H_n$, равна сумме парных произведений вероятностей каждого из этих событий на соответствующие им условные вероятности наступления события A.

$$P(A) = \sum_{i=1}^{n} P(H_i) P(A/H_i).$$

Применим эту формулу на ранее рассмотренном примере с двигателями. Требуется найти вероятность $p(A)$ кондиционности отремонтированного двигателя, вычисленную без всяких предположений о том, на каком заводе двигатель ремонтировался. Очевидно, что любой кондиционный двигатель выпущен первым или вторым заводом.

Тогда $P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)$, где соответственно вероятности того, что двигатель выпущен первым заводом и он кондиционный или вероятность того, что двигатель выпущен вторым заводом и является кондиционным. Эти вероятности определяются теоремой умножения: $P(A)=0.4\cdot 0.8+0.6\cdot 0.7=0.74$.

Пусть производится некоторое количество испытаний, в результате которых может произойти или не произойти событие A, и вероятность появления этого события в каждом из испытаний не зависит от результатов остальных испытаний, то такие испытания называются независимыми относительно события A. Допустим, что событие A наступает в каждом испытании с вероятностью $p(A)=p$. Определим вероятность $P_{n,m}$ того, что в результате n испытаний событие A наступило ровно m раз. Такую вероятность можно посчитать, используя теоремы сложения и умножения вероятностей. Однако, при достаточно большом количестве испытаний, это приводит к очень сложным вычислениям. Пусть в результате n независимых испытаний, проведенных в одинаковых условиях, событие A наступает с вероятностью $p(A)=p$, а противоположное ему событие $\overline{A}$ с вероятностью $p(\overline{A})=1-p$.

Как найти значение вероятности события, которое состоит из множества однородных исходов? Обозначим A_i – наступление события A в испытании с номером i. Если в результате n опытов событие A наступает ровно m раз, то остальные $n-m$ раз это событие не наступает. Событие A может появиться m раз в n испытаниях в различных комбинациях, число которых равно количеству сочетаний из n элементов по m. Вероятность каждой комбинации равна произведению вероятностей:

$$p^m(1-p)^{n-m}.$$

Применяя теорему сложения вероятностей несовместных событий, получаем формулу Бернулли:

$$P_{m,n}=\frac{n!}{m!(n-m)!}p^m(1-p)^{n-m}.$$

Строгие определения генеральной совокупности и выборки пришли из теории вероятностей, с математическую статистику. Хотя терминология математической статистики отличается от терминологии теории вероятностей, что объясняет трудности использования студентами знаний из курса теории вероятностей в практике прикладной статистики. Так, например, в качестве случайных событий рассматриваются события, каждое их которых состоит в том, что объект обладает определенным сочетанием значений рассматриваемых признаков. Сами признаки служат примерами случайных величин (то есть вместо вероятностей фигурируют относительные частоты). Вместо случайной величины X из теории вероятностей, в математической статистике говорят о генеральной совокупности X. Таким образом, понятие генеральной совокупности тождественно понятию случайной величины, т. е. включает в себя описание области определения (пространства элементарных исходов), множества значений, функциональной зависимости, закона распределения. Вместо эксперимента, в результате которого случайная величина X приняла значение x (в теории вероятностей), в математической статистике говорят о случайной выборке из генеральной совокупности X значения x. Вместо n независимых экспериментов, в результате которых случайная величина X приняла значения $x_1, x_2, ..., x_n$ (как принято в теории вероятностей), в математической статистике говорят о случайной выборке объема n значений $x_1, x_2, ..., x_n$ из генеральной совокупности X.

Генеральной совокупностью назовем всю подлежащую изучению совокупность объектов, относительно некоторого признака A, характеризующего эти объекты.

Выборочной совокупностью (*выборкой*) назовем ту часть объектов генеральной совокупности, которую отобрали для непосредственного изучения признака A.

Параметр генеральной совокупности (*параметр*) – показатель, вычисленный для всей генеральной совокупности (например, среднее квадратичное отклонение случайной величины).

Параметр выборки (*выборочный параметр, статистика*) – некоторый показатель, вычисленный на основе данных выборки.

Параметр – фиксированное, но неизвестное число. Обычно в распоряжении исследователя имеются лишь данные выборки, например значения количественного признака $x_1, x_2, ..., x_n$, полученные в результате n наблюдений. Через эти данные и выражают оцениваемый параметр. Если удалось установить вид распределения, к которому относится признак A, то возникает задача нахождения значений параметров, характеризующих данное распределение. Например, если известно, что изучаемый признак имеет распределение Пуассона, то необходимо оценить параметр a, которым это распределение определяется.

Репрезентативность выборки – это показатель, заключающийся в том, что выборка полно и достоверно отображает признаки той совокупности, частью которой она является. Репрезентативность также можно определять, как свойство выборки наиболее полно представлять характеристики генеральной совокупности, существенные с точки зрения цели исследования.

Существуют различные виды выборок: простая случайная, серийная, типическая, механическая и комбинированная.

Простая случайная выборка состоит в отборе из всей совокупности изучаемых единиц наугад без какой-либо системы.

Механическую выборку применяют когда в генеральной совокупности есть упорядоченность, например, имеется некая последовательность единиц (регистрационные номера работников, избирательные списки, номера телефонов респондентов, номера квартир и домов и другое).

Типический отбор используется в случаях, когда всю совокупность можно разделить на группы по типам. При работе с населением такими могут быть, например, образовательные, возрастные, социальные группы, при исследовании предприятий – отрасль или отдельная организация и др.

Серийный отбор удобен когда единицы объединены в небольшие серии или группы. Такой серией могут быть партии, школьные классы, трудовые коллективы и другие группы.

Кроме того, выделяют качественную и количественную репрезентативность. Случайность, гарантирующая *качественную (структурную) репрезентативность* статистических исследований, достигается выполнением ряда условий формирования выборочных групп (совокупностей):

1. Каждый член генеральной совокупности должен иметь равную вероятность попасть в выборку.
2. Отбор единиц наблюдения из генеральной совокупности необходимо проводить независимо от изучаемого признака. Если отбор проводится целенаправленно, то и при этом необходимо соблюдать условия независимости распределения изучаемого признака.
3. Отбор должен проводиться из однородных групп.

Соблюдение условий, гарантирующих максимальную близость выборочной и генеральной совокупностей, обеспечивается специальными способами отбора. В зависимости от способа формирования различают следующие выборки:

1. Выборки, не требующие разделения генеральной совокупности на части (собственно, случайная повторная или бесповторная выборка).
2. Выборки, требующие разбиения генеральной совокупности на части (механическая, типическая или типологическая выборки, когортная, парно-сопряженная выборки).

Собственно, случайная выборка формируется случайным отбором – наудачу. В основе случайного отбора лежит перемешивание. Например: выбор шара в спортлото после перемешивания всех шаров, выбор выигрышных номеров, случайный выбор карточек больных для исследования и т. п. Иногда используют случайные числа, получаемые из таблиц случайных чисел или с помощью генераторов случайных чисел. Согласно этим числам из заранее пронумерованного массива генеральной совокупности выбираются единицы наблюдения с номерами, соответствующими выпавшим случайным числам. При составлении случайной выборки после того, как объект выбран, и все необходимые данные о нем зарегистрированы, можно поступать двояко: объект можно вернуть, или не вернуть в генеральную совокупность. В соответствии с этим выборку называют повторной (объект возвращается в генеральную совокупность) или бесповторной (объект не возвращается в генеральную совокупность). Поскольку в большинстве статистических исследований разница между повторной и бесповторной выборками практически отсутствует, то априорно принимается условие, что выборка повторная.

Важным свойством выборки является ее достоверность. *Достоверность* – показатель вероятности того, что истинное значение изучаемого параметра генеральной совокупности попадет в доверительный интервал. Чем выше задаваемый уровень достоверности, тем больше должна быть выборка. Размер выборки практически не зависит от размера генеральной совокупности. И в корпорации с персоналом несколько десятков тысяч человек, и в региональной фирме численностью в сотни человек для построения выборки, репрезентативной по одинаковому числу параметров, потребуется опросить одинаковое число респондентов. От чего действительно зависит размер выборки – так это от числа параметров, по которым мы желаем добиться репрезентативности. Задача прикладных методов статистики: описать закон распределения генеральной совокупности; подобрать значения параметров этого закона, оценить числовые характеристики генеральной совокупности. Если имеется несколько выборок, извлеченных из разных генеральных совокупностей, определить, одинаково распределены эти генеральные совокупности или нет; одинаковы ли определенные числовые характеристики этих генеральных совокупностей или нет. Понятно, что все перечисленные вопросы сформулированы на языке теории вероятностей. Построение и анализ выборочного распределения является основным математическим способом исследования реальной случайной величины. Теоретическую основу применимости выборочного метода составляет *закон больших чисел,* согласно которому при неограниченном увеличении объема выборки случайные выборочные характеристики сколь угодно приближаются (сходятся по вероятности) к определенным параметрам генеральной совокупности. Пусть $X_1, X_2, \ldots, X_n$ – выборка, полученная в результате наблюдения из некоторой генеральной совокупности X. По выборкам можно получить только приближенные значения неизвестного параметра θ, которые служат его оценкой. Оценки, как правило, меняются от одной выборки к другой.

1.2. Статистические характеристики распределения и графическое представление выборки случайных величин

Стремление к четкой формулировке того или иного положения практически всегда приводит к возможности выражения его на математическом языке. Собственно, математика начинается там, где можно достаточно четко выделить в реальности интересующие аспекты, абстрагируясь от всех остальных проявлений изучаемого явления. По меткому выражению Уайтхеда «Человек, заметивший аналогию между семью рыбами и семью днями, осуществил значительный сдвиг в истории мышления. Он был первым, кто ввел понятие, относящееся к науке чистой математики».

За каждым математическим методом стоит определенная модель того явления, которое с помощью этого метода изучается. Применяя метод, инженер четко должен представлять себе его сущность, «содержательный» смысл. Он должен давать себе отчет в том, в силу самого применения выбранного метода, что считает истинным, от чего абстрагируется, какие ограничения на реальность накладывает и т. д. Иначе метод перестает играть роль «орудия труда» исследователя. Практически для любого явления оказывается возможной разработка целого ряда моделей, каждая из которых является естественной, но отражает лишь какой-то один его аспект. Складывается своеобразная ситуация объектов – если задачу в принципе оказывается возможным решить с помощью какого-либо формального (чаще всего математического) аппарата, то соответствующих решений, как правило, бывает несколько. И ни одно из них не может считаться «главным». Каждое отвечает определенной стороне реальности.

Пусть случайная величина X задана генеральной совокупностью. Требуется оценить количественные характеристики заданной совокупности и установить функцию распределения случайной величины X. Практически же известны лишь данные выборки.

Статистические характеристики (статистики) представляют собой количественную сторону исследуемых параметров, характеризующих распределение случайной величины, полученную на основе исследования выборочных значений. Статистики используются либо для описания самой выборки, либо, что имеет первостепенное значение в фундаментальных экспериментальных исследованиях, для оценки параметров распределения случайной величины в исследуемой генеральной совокупности.

Разделение понятий «*статистика*» и «*параметр*» является очень важным, так как оно позволяет избежать ряд ошибок, связанных с неверным толкованием данных, получаемых в эксперименте. Дело в том, что, когда мы оцениваем параметры распределения с помощью статистических данных, то получаем величины, лишь в определенной степени близкие к оцениваемым параметрам. Между параметрами и статистиками практически всегда существует какое-то различие, причем, насколько велико это различие, как правило, сказать сложно. Теоретически чем больше выборка, тем ближе оцениваемые параметры оказываются к их выборочным характеристикам. Однако это не означает, что, увеличив объем выборки, мы неминуемо ближе подойдем к оцениваемому параметру, уменьшим разницу между ним и вычисленной статистикой.

Если в теории ожидаемое значение статистики совпадает с оцениваемым параметром, то такую оценку называют *несмещенной.* Оценку, при которой ожидаемое значение оцениваемого параметра отличается от самого параметра на некоторую величину, называют *смещенной.* Также следует различать точечную и интервальную оценки параметров распределения.

Точечной называют оценку, полученную с помощью какого-либо числа. Например, если мы утверждаем, что величина производительности для данного механизма в данных условиях и на данном участке составляет 21.8 единиц, то такая оценка будет точечной. Точно так же точечная оценка имеет место, когда в сводке погоды нам сообщают, что за окном 25°C.

Интервальная оценка предполагает использование в оценке набора или диапазона чисел. Оценивая производительность, можно сказать, что он оказался в диапазоне от 20 до 25 единиц. Аналогичным образом синоптики могут сообщить, что по их прогнозам температура воздуха в ближайшие сутки достигнет значения 22-24°C. Интервальная оценка случайной величины позволяет не только определить искомое значение этой величины, но и задать возможную точность для такой оценки.

Параметры распределения случайной переменной, например, такие как математическое ожидание или дисперсия, асимметрия, эксцесс являются теоретическими величинами, недоступными непосредственному измерению.

Они представляют собой количественную характеристику *генеральной совокупности* и могут быть сами по себе определены как гипотетические величины, поскольку описывают особенности распределения случайной величины в самой генеральной совокупности. Для того чтобы определить их на практике, исследователь, проводящий эксперимент, осуществляет их выборочную оценку.

Случайной величиной называется величина, которая в результате опыта может принимать то или иное значение, причем заранее известно какое именно. Случайные величины можно разделить на две категории.

Дискретной случайной величиной называется такая величина, которая в результате опыта может принимать определенные значения с известной вероятностью, образующие счетное множество (множество, элементы которого могут быть занумерованы). Это множество может быть как конечным, так и бесконечным. Например, количество выстрелов до первого попадания в цель является дискретной случайной величиной, т. к. эта величина может принимать и бесконечное, хотя и счетное количество значений.

Непрерывной случайной величиной называется такая величина, которая может принимать любые значения из некоторого конечного или бесконечного промежутка. Очевидно, что число возможных значений непрерывной случайной величины бесконечно.

Соотношение между возможными значениями случайной величины и их вероятностями называется *законом распределения дискретной* случайной величины. Закон распределения может быть задан аналитически, в виде таблицы или графически.

Таблица соответствия значений случайной величины и их вероятностей называется *рядом распределения*.

X	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

Графическое представление таблицы называется *многоугольником распределения* (рис. 1.1). При этом сумма всех ординат многоугольника распределения представляет собой вероятность всех возможных значений случайной величины, а, следовательно, равна единице. Дискретная случайная величина определяется путем задания значений самой величины и вероятностей этих значений.

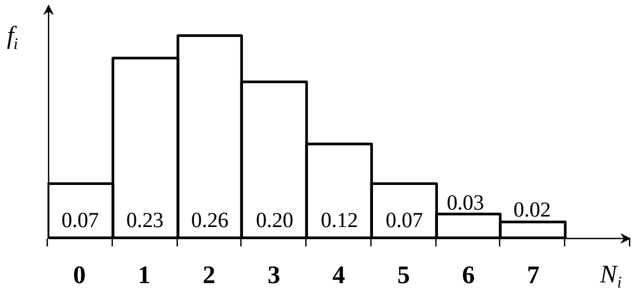


Рис. 1.1. Ряд (многоугольник) распределения

В случае непрерывной случайной величины, ее значения могут заполнять некоторый произвольный интервал. Очевидно, что в этом случае задать все значения случайной величины просто нереально. Пусть X – действительное число. Вероятность события, состоящего в том, что случайная величина X примет значение, меньшее X, т. е. $X < x$, обозначим $F(x)$. *Функцией распределения* называют функцию $F(x)$, определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее x. Функцию $F(x) = P(X < x)$ называют *интегральной функцией*.

Функция распределения существует как для непрерывных, так и для дискретных случайных величин. Она полностью характеризует случайную величину и является одной из форм закона распределения.

Для дискретной случайной величины функция распределения имеет вид:

$$F(x) = \sum_{x_i < x} P(X = x_i).$$

Знак неравенства под знаком суммы показывает, что суммирование распространяется на все те возможные значения случайной величины, которые меньше аргумента x. Функция распределения дискретной случайной величины X разрывна и возрастает скачками при переходе через каждое значение x_i.

Свойства функции распределения:

1. Значения функции распределения принадлежат отрезку $[0;1]$.

$$0 \le F(x) \le 1.$$

2. $F(x)$ – неубывающая функция.

$$F(x_2) \ge F(x_1), \quad \text{при } x_2 \ge x_1.$$

3. Вероятность того, что случайная величина примет значение, заключенное в интервале $(a;b)$, равна приращению функции распределения на этом интервале.

$$P(a \le X < b) = F(b) - F(a).$$

4. На минус бесконечности функция распределения равна нулю, на плюс бесконечности функция распределения равна единице.

5. Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

Таким образом, не имеет смысла говорить о каком-либо конкретном значении случайной величины. Интерес представляет только вероятность попадания случайной величины в заданный интервал, что соответствует большинству практических задач.

Функция распределения полностью характеризует случайную величину, однако, имеет один недостаток. По функции распределения трудно судить о характере распределения случайной величины в небольшой окрестности той или иной точки числовой оси.

Плотностью распределения вероятностей непрерывной случайной величины X называется функция $f(x)$ – первая производная от функции распределения $F(x)$: $f(x) = F'(x)$. Плотность распределения называют *дифференциальной функцией.* Смысл плотности распределения состоит в том, что она показывает, как часто появляется случайная величина X в некоторой окрестности точки x при повторении опытов. Для описания дискретной случайной величины плотность распределения неприемлема.

После введения функций распределения и плотности распределения можно дать следующее определение непрерывной случайной величины.

Случайная величина X называется *непрерывной*, если ее функция распределения $F(x)$ непрерывна на всей оси Ox, а плотность распределения $f(x)$ существует везде, за исключением, может быть, конечного числа точек. Зная плотность распределения, можно вычислить вероятность того, что некоторая случайная величина X примет значение, принадлежащее заданному интервалу.

<u>Теорема.</u> Вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу $(a;b)$, равна определенному интегралу от плотности распределения, взятому в пределах от a до b:

$$P(a < X < b) = \int_a^b f(x)dx.$$

Функция распределения может быть легко найдена, если известна плотность распределения, по формуле: $F(x) = \int_{-\infty}^{x} f(x)dx$.

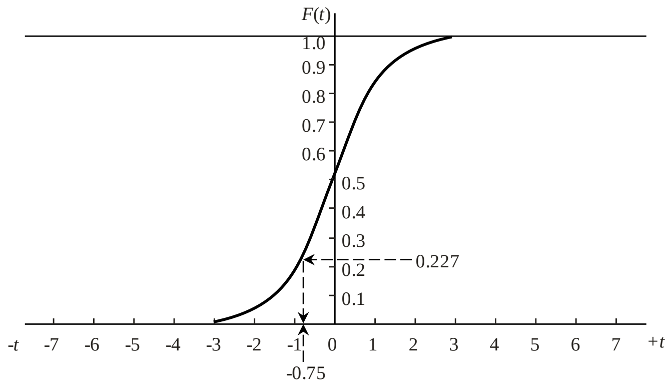


Рис. 1.2. График интегральной функции распределения

Свойства функции плотности распределения:

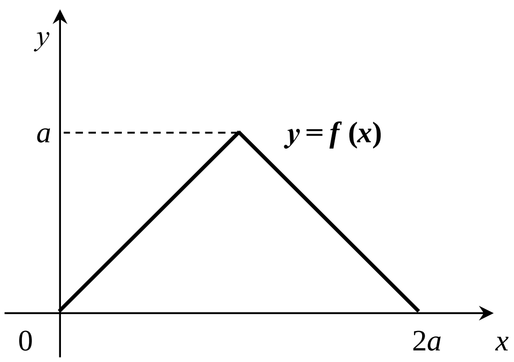
1. Плотность распределения – неотрицательная функция.

$$f(x) \geq 0.$$

2. Несобственный интеграл от плотности распределения в пределах от $-\infty$ до $+\infty$ равен единице.

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

Пример. График плотности распределения вероятностей $f(x)$ случайной величины приведен на рисунке. Найти значение a.



Так как $\int_{-\infty}^{\infty} f(x)dx = 1$, то $\int_{-\infty}^{\infty} f(x)dx = 2\int_{0}^{a} xdx = a^2 \Rightarrow a^2 = 1$. Таким образом, $a = 1$.

Математическим ожиданием дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на их вероятности:

$$m_x = M(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_{n1} = \sum_{i=1}^{n} x_i p_i .$$

Несмещенной оценкой математического ожидания случайной величины (дискретной) называется среднее значение случайной величины:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Пример. Проведено четыре измерения (без систематических ошибок) некоторой случайной величины: 5, 6, 9, 12. Найти несмещенную оценку математического ожидания.

$$\bar{x} = \frac{5+6+9+12}{4} = 8.$$

Пример. Пусть X – дискретная случайная величина, заданная законом распределения вероятностей:

X	-1	3
P	0.4	0.6

Найти математическое ожидание этой случайной величины.

$$M(X) = (-1) \cdot 0.4 + 3 \cdot 0.6 = 1.4.$$

Математическим ожиданием непрерывной случайной величины X, возможные значения которой принадлежат отрезку $[a, b]$, называется определенный интеграл:

$$M(X) = \int_a^b x\, f(x) dx.$$

Если возможные значения случайной величины рассматриваются на всей числовой оси, то математическое ожидание находится по формуле:

$$M(X) = \int_{-\infty}^{\infty} x\, f(x) dx.$$

При этом, конечно, предполагается, что несобственный интеграл сходится.

Модой M_0 случайной величины называется ее наиболее вероятное значение. Для непрерывной случайной величины мода – такое значение случайной величины, при которой плотность распределения имеет максимум:

$$f(M_0) = max.$$

Для дискретной случайной величины модой можно определить самое часто встречающееся значение измеренного признака, которым обладает максимальное число элементов выборки. Числовые значения *моды* можно найти и по формуле:

$$M_0 = x_0 + \delta \frac{n_{M_0} - n^-}{2n_{M_0} - n^- - n^+},$$

где x_0 – нижняя граница модального интервала; δ – величина модального интервала; n_{M_0} – частота модального интервала; n^- – частота интервала, предшествующего модальному; n^+ – частота интервала, следующего за модальным.

Если многоугольник распределения для дискретной случайной величины или кривая распределения для непрерывной случайной величины имеет два или несколько максимумов, то такое распределение называется *бимодальным* или *полимодальным*.

Медианой M_D случайной величины X называется такое ее значение, относительно которого равновероятно получение большего или меньшего значения случайной величины $P(X < M_D) = P(X > M_D)$.

Геометрически медиана – абсцисса точки, в которой площадь, ограниченная кривой распределения делится пополам, иначе справа и слева от которой находится одинаковое число элементов выборки.

Расчетная формула для нахождения медианного значения имеет вид:

$$M_e = x_0 + \delta \frac{\frac{1}{2}n - n_H}{n_{M_e}},$$

где x_0 – нижняя граница медианного интервала, δ – величина медианного интервала, n_{M_e} – частота медианного интервала, n_H – частота, накопленная до медианного интервала. Медиана вычисляется вместе с математическим ожиданием (часто дополнительно к нему) для распределений случайной величины, имеющих значительную асимметрию.

Заметим, что если распределение унимодальное и симметричное, то мода и медиана совпадают с математическим ожиданием.

Средние тенденции помогают определить наиболее типичные значения (одно или несколько), которые наилучшим образом представляют весь комплекс измерений по данной переменной. Целесообразность использования того или иного типа средней величины зависит от цели усреднения, вида распределения, вычислительных соображений. Часто для характеристики рядов распределения оказывается недостаточным указание только средней величины данного признака, поскольку два ряда могут иметь одинаковые средние тенденции.

Кроме средних тенденций существует величина, которая характеризует отклонение значений случайной величины от среднего.

Дисперсией (*рассеиванием*) случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания $D(X) = M[X - M(X)]^2$. Однако, на практике подобный способ вычисления дисперсии неудобен, т. к. приводит при большом количестве значений случайной величины к громоздким расчетам. Поэтому применяется другой способ.

Дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом ее математического ожидания:

$$D(X) = M\left[X - M(X)\right]^2.$$

Средним квадратичным отклонением случайной величины X называется квадратный корень из дисперсии:

$$\sigma(X) = \sqrt{D(X)}.$$

Изменчивость среднеквадратичного отклонения влияет на «растяжение» кривой плотности распределения. Уменьшение значения $\sigma(X)$ отражается увеличением ординаты $f(x)$, то есть «сжатием» кривой.

Дисперсию рассматриваемых распределений можно проинтерпретировать естественным образом, например для нормального закона. Напомним, что нормальное распределение однозначно задается значениями математического ожидания и дисперсии. Рассмотрим рис. 1.3, на котором изображены варианты нормального распределения, отвечающие разным дисперсиям.

Нетрудно понять, что дисперсия говорит об однородности результатов измерений на рассматриваемом объекте. Если множество значений определяется распределением I, то данные измерений, будучи полученными в разное время, примерно с одинаковой вероятностью будет давать совершенно различные, в том числе и весьма отличающиеся от среднего. Так, значения x_1 и x_2 могут встретиться почти с той же вероятностью, что и среднее значение.

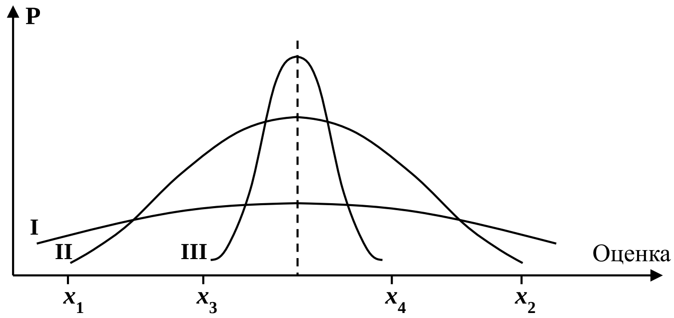


Рис. 1.3. Нормальное распределение данных в *l*-м измерении *i*-го объекта при разных дисперсиях

Если множество измеренных данных определяется распределением III, то, напротив, значения, даже незначительно отличающиеся от среднего, такие, как x_3 и x_4 будут встречаться с меньшей вероятностью, чем само среднее. При реализации распределения II ситуация будет занимать промежуточное положение между двумя описанными выше. Ясно, что близость к среднему может быть объяснена разными факторами: знанием оцениваемых объектов, тонностью и тщательностью измерений и т. д. Характеристики дисперсии и среднеквадратичного отклонения являются абсолютными показателя рассеивания.

Рассмотрим вопрос: должны ли быть схожими, и, если должны, то в какой степени, распределения, отвечающие разным объектам или экспериментам, при определенной однородности изучаемой совокупности. Покажем, что смысл задачи заставляет нас считать равными средние значения соответствующих распределений.

Предположим, что упомянутого равенства нет, то есть в наличии ситуация, отраженная на рис. 1.4 а. Наверное, исследователь при наличии в изучаемой совокупности таких данных, придет к выводу, что измерения относительно рассматриваемого объекта разделились: в одном случае он более однороден, в другом – нет. В такой ситуации, вероятно, разумно разделить всю совокупность на две и для каждой из полученных совокупностей искать среднюю оценку отдельно.

Теперь, представим, что каким-то двум экспериментам отвечают распределения, изображенные на рис. 1.4 б, где σ^1 и σ^2 – средние квадратичные отклонения. Один из исследователей хорошо знает объект и поэтому уверен в своих оценках. Его дисперсия мала, кривая «узка», вероятность погрешности, сильно отличающийся от среднего, практически равна нулю. Напротив, другой имеет об изучаемом объекте весьма смутное представление. Ему более или менее все равно, какие данные измерения получены. Весьма сильно разнящиеся результаты могут встретиться примерно с одинаковой вероятностью. Его кривая «широка», а дисперсия велика.

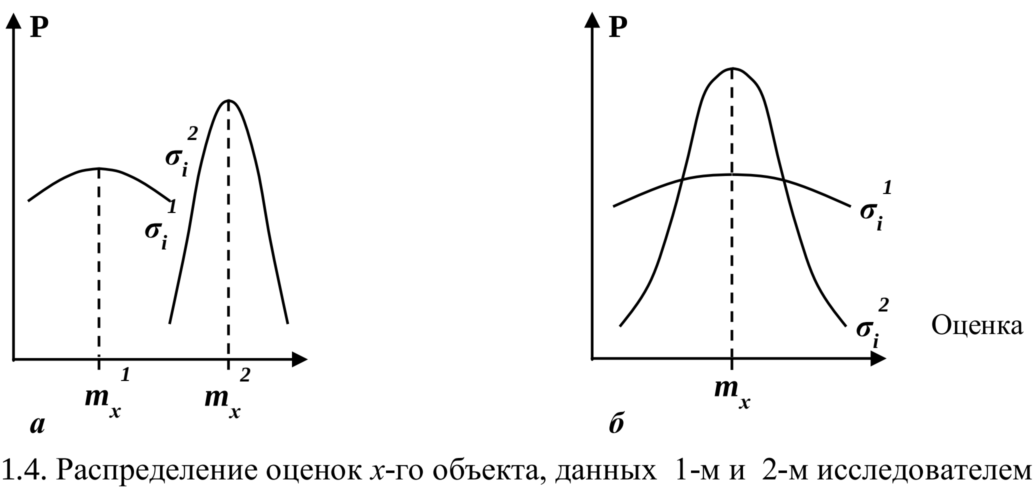


Рис. 1.4. Распределение оценок *x*-го объекта, данных 1-м и 2-м исследователем:
а – с разными средними и разными дисперсиями;
б – с одинаковыми средними, но разными дисперсиями.

Таким образом, степень концентрации (или, наоборот, разброса) значений признаков вокруг средней будет различной. Характеристикой такого разброса служат показатели колебания (вариативности) – разность между максимальным и минимальным значениями признака в некоторой совокупности (вариационный размах), а также другие показатели: среднее абсолютное (линейное) отклонение и т. п.

Пример. На строительстве дороги применяется новый тип экскаваторов нормы выработки которого отсутствуют. Для обоснования этой нормы проведено 100 измерений выработки при проведении земельных работ. Данные измерений приведены в таблице. Рассчитать числовые характеристики и сделать вывод.

Объем выполненных работ, *куб. м/ч*, x_i	40-50	50-60	60-70	70-80	80-90	90-100	100-110
Количество случаев, n_i	4	24	40	20	19	2	0

Решение. Заметим, что проведенное измерение интервальное и дискретное. В качестве расчетного значения случайной величины x_i выбираем середину интервала. (С таким же успехом можно было взять левую или правую границу – на результат это не повлияет!). Найдем значение эмпирического математического ожидания:

$$M(X)=\frac{1}{100}\left(45\cdot 4+55\cdot 24+65\cdot 40+75\cdot 20+85\cdot 10+95\cdot 2\right)=66.4\,куб.м/ч.$$

Рассчитаем дисперсию и среднеквадратичное отклонение. Для расчета дисперсии найдем:

$$M(X^2)=\frac{1}{100}\left(45^2\cdot 4+55^2\cdot 24+65^2\cdot 40+75^2\cdot 20+85^2\cdot 10+95^2\cdot 2\right)=4525,$$

далее вычислим $\left[M(X)\right]^2=66.4^2=4408.96$. Применим формулу $D(X)=M(X^2)-\left[M(X)\right]^2$. Значение дисперсии $D(X)=4525-4408.96=116.04$; $\sigma(X)=\pm\sqrt{116.04}=\pm 10.77\,м^3/ч$.

Полученные числовые характеристики дают вероятностную информацию о производительности экскаватора. Так, полученное среднее $M(X)=66.4\,м^3/ч$ можно приять в качестве часовой нормы выработки, однако среднеквадратичное отклонение дает наивероятнейший диапазон равный: $66.4\pm 10.77\,м^3/ч$, что существенно расширяет среднюю норму. Для проверки достоверности определенного по средней тенденции норматива необходимо установить и проанализировать закон распределения случайной величины. Вернемся к этому примеру позднее.

Кроме средних тенденций и разбросов исследователь может вычислить такие показатели, как асимметрия и эксцесс.

Асимметрия (коэффициент асимметрии, третий момент случайной величины) обозначает соотношение между длинами правого и левого «хвостов» распределения. Существует статистическая оценка отклонения данного распределения от симметричного, или, иначе говоря, его скошенность. Степень скошенности распределения и показывает величина его асимметрии, которая вычисляется по формуле:

$$A = \frac{\sum (x_i - \bar{x})^3}{n \cdot \sigma^3}.$$

Скос влево имеет положительное значение асимметрии, вправо – отрицательное. Показатель асимметрии может быть использован для содержательной интерпретации полученных данных. Если наблюдаемый признак формируется под воздействием большого числа факторов, каждый из которых вносит свой небольшой вклад в величину этого признака, то мы вправе ожидать симметричного распределения. Однако если получена значительная величина асимметрии (большая по абсолютной величине, чем 0.4-0.5), можно предположить, что присутствует значительное влияние одного или группы факторов.

Эксцесс характеризует островершинность распределения. Так, например, величина эксцесса для нормальной (гауссовой) кривой распределения $E_x = 3$. Исходя из целого ряда соображений, заостренность этой кривой принимают за стандарт, поэтому в качестве показателя эксцесса используют величину $y = E_x - 3$. Собственно сам эксцесс может быть вычислен по формуле:

$$E = \frac{\sum (x_i - \bar{x})^4}{n \cdot \sigma^4} - 3.$$

Величина эксцесса может принимать очень большие значения, но он не может быть меньше единицы. Близость к единице свидетельствует о бимодальности (двувершинности) распределения. Бимодальность кривой распределения эмпирических данных может возникать за счет объединения в единую совокупность двух наборов разнородных измерений.

Кривые, полученные в результате графического представления эмпирических данных, могут иметь разнообразную форму. Среди них можно выделить относительно небольшое количество простых типов. Некоторые, наиболее распространенные, приведены на рис. 1.5.

Анализ формы кривых иногда поможет в выявлении внутренней, скрытой структуры исследуемой совокупности. Например, можно предположить, что форма кривой в) обусловлена наложением друг на друга двух кривых а) и б), иначе говоря, существует третья скрытая переменная (или группа переменных), определяющая деление совокупности на две группы.

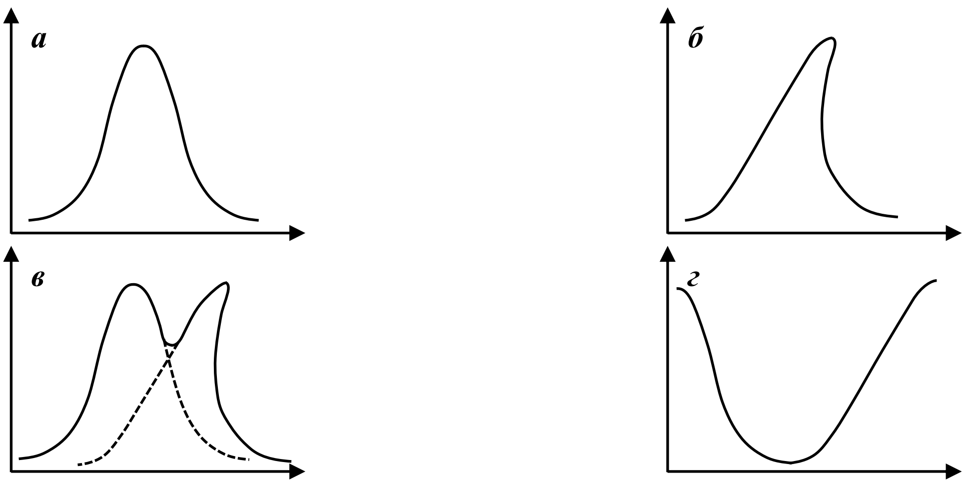


Рис. 1.5. Различие формы кривых распределения

Сгруппированные данные – отдельные значения признаков, объединенные в группы (интервалы). Весь диапазон значений вариант разбивают на разумное число интервалов, как правило, одинаковой ширины h. В этом случае частоты соотносят уже не с каждым отдельным значением признака, а с рядом значений, попадающих в определенный интервал. Во избежание недоразумений при подсчете числа вариантов выборки, попавших в каждый интервал, левая граница интервала считается закрытой, а правая – открытой.

Частотой i-го интервала называется число, равное количеству вариант выборки, попавших в этот интервал. Аналогично вычисляют накопленные и относительные накопленные частоты для правых границ интервалов.

При построении интервальных рядов большое значение имеет выбор размеров интервалов. Общее требование к этому выбору состоит в том, что группировка должна наиболее полно выявлять существенные свойства рядов распределения. Как правило, приходится делать выбор между двумя крайностями: слишком крупные интервалы для данного объема выборки скрадывают многие нюансы в описании явления, а слишком дробные ведут к статистически незначимым частотам внутри интервала. Интервальные ряды распределения могут строиться с равными и неравными интервалами. Неравные интервалы применяются при неравномерном распределении частот значений группировочного признака для выделения качественно отличных типов явления.

Если у исследователя нет предварительной информации о характере распределения по тому или иному признаку, то следует задавать равные интервалы. Равные интервалы также наиболее удобны при использовании методов математической статистики, при этом по каждому из признаков не следует брать более 20 группировочных интервалов. При образовании интервалов необходимо точно обозначить количественные границы группы, избегая таких обозначений границ интервалов, при которых отдельные единицы совокупности могут быть отнесены в две соседние группы.

Пример. Для 50 самосвалов измерили время полной разгрузки (мин) результаты оценили с точностью до сотых. Полученные данные представлены выборкой:

3.7 3.85 3.7 3.78 3.6 4.45 4.2 3.87 3.33 3.76 3.75 4.03 3.75 4.18 3.8 **4.75** 3.25 4.1 3.55 3.35 3.38 3.3 4.15 3.95 3.5 3.88 3.71 3.15 4.15 3.8 4.22 3.75 3.58 3.55 4.08 4.03 3.24 4.05 3.56 **3.05** 3.58 3.98 3.88 3.78 4.05 3.4 3.8 3.06 4.38 4.2

Решение. Для группировки этой выборки выделены наименьшее и наибольшее значения среди данных, которые соответственно равны 3.05 и 4.75. Выборка «упакована» в границы от 3-4.8. Выберем ширину интервала $h = 0.3$ и получим 6 интервалов.

Вычислим накопленные частоты для правых границ и составим таблицу 1.1.

Таблица 1.1. Интервальная таблица накопленных частот

	[3-3.3)	**[3.3-3.6)**	**[3.6-3.9)**	**[3.9-4.2)**	**[4.2-4.5)**	**[4.5-4.8)**
Частоты	5	11	17	11	5	1
Относительные частоты	0.1	0.22	0.34	0.22	0.1	0.02
Накопленные частоты	5	16	33	44	49	50
Относительные накопленные частоты	0.1	0.32	0.66	0.88	0.98	1.0

Табличная форма представления данных

Статистические таблицы – форма рационального и наглядного представления цифровых характеристик исследуемых явлений, дающая возможность характеризовать их размеры, структуру и динамику. Таблицы бывают простые, групповые и комбинационные.

Простая таблица – в ней содержатся сводные показатели, относящиеся к перечню единиц наблюдения, хронологических дат или территориальных подразделений. Соответственно таблицы могут называться простыми перечневыми, хронологическими.

Приведем пример простой таблицы (табл. 1.2).

Таблица 1.2. Динамика уровня безработицы
в Российской Федерации за 1994-1998 гг.

	1994	**1995**	**1996**	**1997**	**1998**
Уровень безработицы, рассчитанный по методологии международной организации труда (МОТ)	7.4	8.8	9.9	11.2	13.3
Уровень официальной безработицы	2.2	3.2	3.4	2.8	2.6

Различные значения уровня безработицы получаются из различий в методологии расчета. По методике, используемой службой занятости населения (уровень официальной безработицы), в основе расчетов лежат вопросы по финансированию мероприятий, связанных с профессиональной переподготовкой и обучением, выплатой пособий по безработице и т. д. Методология МОТ нацелена на определение реальной нагрузки на рынок труда со стороны предложения рабочей силы.

Групповые таблицы – в них статистическая совокупность разделяется на отдельные группы по какому-либо признаку, причем каждая из групп может быть охарактеризована рядом показателей (табл. 1.3).

Таблица 1.3. Распределение занятого населения России по секторам экономики

Занятое население	**млн. чел.**		**в % к итогу**	
	1997 г.	**1998 г.**	**1997 г.**	**1998 г.**
В экономике, всего	64.6	63.6	100	100
В том числе по секторам: государственные и муниципальные предприятия и организации	25.9	24.4	40.1	38.3
Частный сектор	25.8	26.6	39.9	41.9
Общественные организации, фонды	0.4	0.4	0.6	0.6
Совместные предприятия и др.	12.5	12.2	19.4	19.2

Данные таблицы 1.3 свидетельствуют о том, что если до 1997 г. в России наибольшая численность населения была занята в государственном секторе, то в 1998 г. 41.9% общей численности занятых приходится на частный сектор.

Комбинационные таблицы – в них статистическая совокупность разбита на группы не по одному, а по нескольким признакам. Идея их построения состоит в том, что каждую из групп в таблице разбивают на подгруппы по какому-либо признаку, которые в свою очередь дальше могут разделяться по следующему признаку.

Результаты комбинационной группировки по большому количеству признаков даже при небольшом числе интервалов становятся труднообозримыми. Поэтому нецелесообразно составлять комбинационные таблицы по сочетанию более чем трех признаков и более четырех интервалов.

Группировка, осуществляемая одновременно по комплексу признаков, называется *многомерной*. Например, для характеристики технического уровня развития предприятий могут быть использованы различные показатели.

В таблице 1.4 приведены результаты группировки предприятий по уровню технического развития и производительности труда. При выделении однородных по техническому уровню групп предприятий применимы, например, методы кластерного анализа по восьми показателям технического уровня развития. Анализ данных таблицы 1.4 позволяет выделить группы предприятий, добившихся наибольшего эффекта в своей деятельности, и группы предприятий, располагающих резервами роста производства за счет лучшего использования технического потенциала. Это те шесть предприятий, которые имеют техническую оснащенность выше средней по отрасли, но уровень производительности труда здесь ниже среднего по отрасли.

Таблица 1.4. Распределение предприятий по уровням технической оснащенности и производительности труда

Группы предприятий по уровню технической оснащенности	**Группы предприятий по уровню производительности труда**			Итого
	Ниже среднего по отрасли уровня	Среднего по отрасли уровня	Выше среднего по отрасли уровня	
Ниже среднего по отрасли уровня	9	8	8	25
Среднего по отрасли уровня	5	6	1	12
Выше среднего по отрасли уровня	6	3	7	16
Итого	20	17	16	53

Построение таблицы подчинено определенным правилам. Основное ее содержание должно быть отражено в названии (круг рассматриваемых вопросов, географические границы статистической совокупности, время, единицы измерения). Все таблицы должны быть последовательно пронумерованы: сквозная нумерация – табл. 1, табл. 2 – для небольших работ; индексационная поглавная (например, для главы 3) нумерация – табл. 3.1, табл. 3.2 – для работ с несколькими пронумерованными частями. Если таблицы даются наряду с графиками, схемами и другим иллюстративным материалом, то обычно они нумеруются отдельно. В случае если данные охватывают определенный период времени, его границы следует включить в заголовок.

Если таблица частично или полностью составлена по сведениям другого источника, сразу под ней следует дать ссылку.

Хорошо сконструированная таблица позволяет исследователю более четко представить и описать смысл и сущность изучаемого им социального явления. Таким образом, метод группировки и представление материала в виде статистических таблиц уже дают определенные возможности для изучения социологических данных. С другой стороны, они являются совершенно необходимым средством для дальнейшего анализа и применения более тонких статистических методов.

Для сгруппированных данных формула вычисления среднего значения преобразуется в следующую:

$$\bar{x} = \frac{\sum_{i=1}^{h} x_i n_i}{n} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_h n_h}{n},$$

где n_i – частота для i-го значения признака.

Процедуру вычисления среднего значения по сгруппированным данным удобно выполнять по схеме, приведенной в таблице 1.5.

Таблица 1.5. Схема вычисления среднего по сгруппированным данным

Интервал	**Середина интервала** x_i	**Относительная частота** n_i	**Произведения** $x_i\ n_i$
Последовательно выписываются все интервалы	x_1 x_2 … x_k	n_1 n_2 … n_k	$x_1\ n_1$ $x_2\ n_2$ … $x_k\ n_k$
		$\sum_{i=1}^{k} x_i\ n_i = n$	$\sum_{i=1}^{k} x_i\ n_i$

Графически выборку из генеральной совокупности можно представить полигоном частот распределения (многоугольник распределения), кумулятивной кривой, гистограммой, которые являются прообразами законов распределения случайных величин. Гистограмма и полигон частот распределения, построенные на основе эмпирических данных выборки, позволяют выявить приближенную картину реального распределения в генеральной совокупности. При увеличении выборочной совокупности и все большем дроблении величины интервалов эмпирическое распределение в виде гистограммы или полигона частот все более приближается к некоторой кривой, называемой *кривой распределения*.

Полигон частот – ломаная линия, прообраз кривой распределения. Для построения полигона частот величина признака откладывается на оси абсцисс, а частоты или относительные частоты – на оси ординат. Из точек, соответствующих значениям признака, восстанавливаются перпендикуляры, равные по высоте частотам. Вершины перпендикуляров соединяются прямыми линиями. В результате получается ломаная линия.

При построении *полигона частот распределения дискретной случайной величины* по оси абсцисс откладываются значения, которые принимает величина, а по оси ординат – частота появления этого значения в рассматриваемой совокупности. При построении *полигона частот распределения непрерывной случайной величины* или интервального ряда по оси абсцисс располагаются точки, обозначающие интервалы, в которые может попадать случайная величина (это могут быть как серединные значения интервалов, так и другие значения, в зависимости от принимаемой модели), а по оси ординат – частота появления значения признака (рис. 1.6). Площади полученных прямоугольников представляют собой выборочные оценки значений функции плотности распределения вероятностей.

Пример. Данные распределения рабочих в возрасте до 24 лет по тарифным разрядам дают возможность построить полигон частот.

Разряд	I	II	III	IV	V	VI
Численность, %	8.4	22.6	31.9	24.1	6.2	0.3

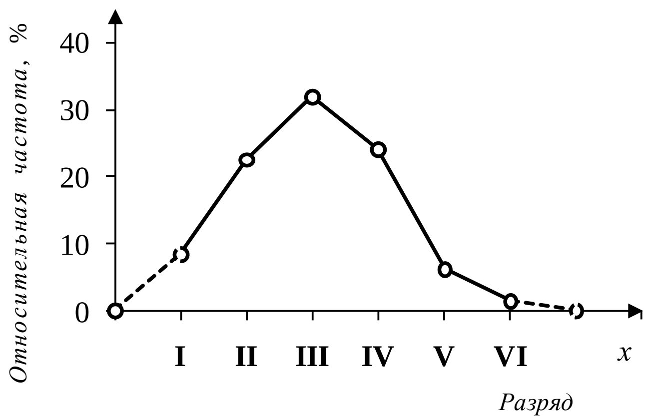


Рис. 1.6. Полигон частот распределения рабочих по тарифным разрядам

Условно принято крайние ординаты признака соединять с серединами примыкающих интервалов (на рисунке эти замыкающие линии нанесены пунктиром). Однако для распределения, где концентрация событий увеличивается на концах полигона, такое изображение может привести к ложным представлениям о сущности явления.

Пример. Из генеральной совокупности извлечена выборка объема $n = 50$, полигон частот которой имеет вид рис. 1.7. Найти число вариант переменной $x_i = 4$ в данной выборке.

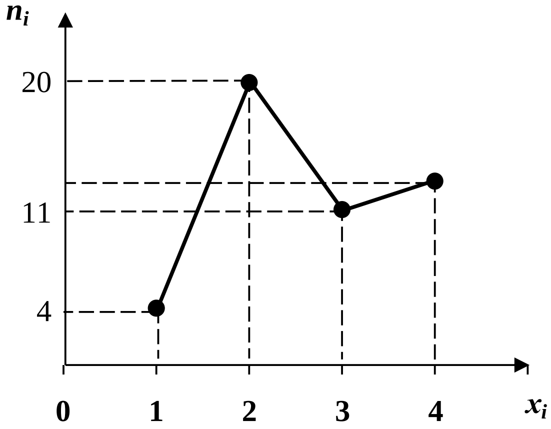


Рис. 1.7. Полигон частот

Решение. Так как объем выборки сумма всех вариант переменной, запишем ее в явном виде $n = n_1 + n_2 + n_3 + n_4 = 4 + 20 + 11 + n_4 = 50$, тогда $n_4 = 15$.

Для графического изображения вариационных рядов используются также кумулятивные кривые (*кумуляты*). При построении кумуляты на оси абсцисс откладываются границы интервалов (либо значения дискретного признака), а на оси ординат – накопленные частоты (либо относительные частоты), соответствующие верхним границам интервалов. Таким образом, на графике кумуляты столбики, пропорциональные частотам, последовательно накладываются один на другой, так что высота последнего столбика является суммой всех высот, составляющей 100%. Кумулята округляет индивидуальные значения признака в пределах интервала и представляет собой возрастающую ломаную линию. Это графическое представление данных позволяет быстро определить процент случаев, находящихся ниже или выше заданной величины признака.

Составим вариационный ряд, для которого построим кумуляту при изучении результатов тестирования на профпригодность соискателей на должность. Наибольшее количество баллов за тест равно 12, сложность заданий возрастает по мере увеличения порядкового номера, объем выборки – 35 соискателей.

Количество баллов	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
Частота	1	1	2	3	4	4	6	5	3	3	2	1
Накопленная частота n	1	2	4	7	11	15	21	26	29	32	34	35

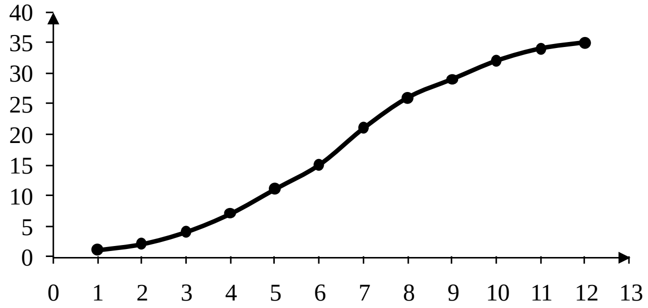


Рис. 1.8. Кумулята распределения числа выполненных заданий в тесте

Гистограммой называется графическое изображение зависимости частоты попадания элементов выборки от соответствующего интервала группировки (диапазона значений показателя). Гистограмма используется только для непрерывных признаков: предполагается, что внутри каждого интервала распределение случайной величины равномерно. По оси абсцисс откладываются крайние значения интервалов, по оси ординат – частота появления значения в рассматриваемой совокупности. Таким образом, концы интервала представляются на графике двумя точками, из которых опускаются проекции на ось абсцисс, после чего полученные четыре точки соединяются в прямоугольники. Гистограмма графическое средство для изображения ряда интервальных переменных. Рассмотрим пример построения гистограммы по данным таблице 1.6.

Таблица 1.6. Распределение сроков нахождения в должности менеджеров строительной корпорации

	Менее года	**1-2 года**	**2-3 года**	**3-4 года**	**4-6 года**	**6-8 лет**	**8-10 лет**	**10 и более**
Относительная частота, %	7.2	14.5	13.2	22.9	16.9	8.4	1.2	15.7

На гистограмме общее число лиц в каждой категории выражается площадью соответствующего прямоугольника, а общая площадь равна численности совокупности (так как гистограмма строится по относительным частотам, то площадь равна единице или 100%). Поэтому для интервалов 4-6; 6-8; 8-10 (по табл. 1.6), длина которых в два раза больше предыдущих, нужно брать высоты прямоугольников в два раза меньше. При нанесении на график последнего открытого интервала «10 лет и более» условно будем считать верхней его границей 20 лет. Тогда ширина интервала равна 10 годам, а плотность распределения – около 0.5%.

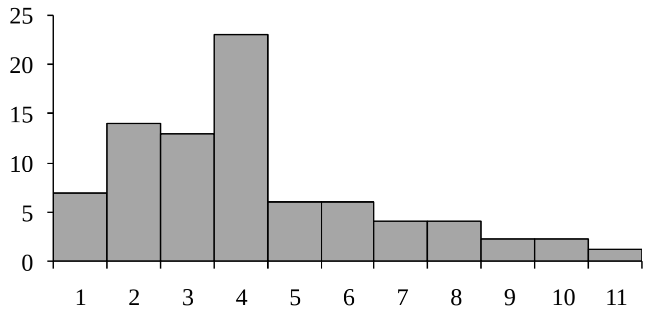


Рис. 1.9. Гистограмма распределения сроков нахождения в должности менеджеров строительной корпорации

Гистограмму следует отличать от столбчатой диаграммы – отличие состоит в том, что столбчатая диаграмма отображает частоту через высоту столбца, а гистограмма – через его площадь. Иными словами, столбцы столбчатой диаграммы всегда имеют одинаковую ширину, а столбцы гистограммы могут иметь разную ширину, в зависимости от интервалов.

Правильно построенные графические изображения эмпирических данных измерений позволяют найти приближенные значения (с точностью до масштаба) средних тенденций для измерений по различным типам шкал. Рассмотрим это на примере.

Пример. Имеются данные о возрастном составе рабочих (лет) некоторого предприятия:

18, 38, 28, 29, 26, 38, 34, 22, 28, 30, 22, 23, 35, 33, 27, 24, 30, 32, 28, 25, 29, 26, 31, 24, 29, 27, 32, 25, 29, 29.

1. Построить интервальный ряд распределения.

2. Построить графическое изображение ряда.

3. Графически определить моду и медиану.

Решение. Объем выборки равен 30 единиц. Определим число групп, на которые надо разделить всю совокупность по формуле Стерджесса:

$$1+3.322\lg 30=6 \text{ групп (не более).}$$

Размах выборки: максимальный возраст – 38, минимальный – 18. Найдем ширину интервала $\dfrac{38-18}{6}=\dfrac{20}{6}=3+1/3$. Так как концы интервалов должны быть целыми числами, разделим совокупность на 5 групп, при этом ширина интервала будет равна 4. Для облегчения подсчетов расположим данные в порядке возрастания:

18, 22, 22, 23, 24, 24, 25, 25, 26, 26, 27, 27, 28, 28, 28,
29, 29, 29, 29, 29, 30, 30, 31, 32, 32, 33, 34, 35, 38, 38.

Составим таблицу распределение возрастного состава рабочих:

№ п/п	Возраст (границы групп), *x*	Частота, *f*	Накопленная частота, *S*
1	18-22	3	3
2	23-26	7	10
3	27-30	12	22
4	31-34	5	27
5	35-38	3	30
	Всего	30	

Изобразим ряд в виде гистограммы и полигона частот (рис. 1.10). Основание столбика гистограммы – ширина интервала. Высота столбика равна частоте.

Полигон (или многоугольник распределения) – график частот. Чтобы его построить по гистограмме, соединяем середины верхних сторон прямоугольников. Многоугольник замыкаем на оси OX на расстояниях, равных половине интервала от крайних значений *x*.

Для определения среднего значения по гистограмме, надо выбрать самый высокий прямоугольник, провести линию от правой вершины этого прямоугольника к правому верхнему углу предыдущего прямоугольника, и от левой вершины модального прямоугольника провести линию к левой вершине последующего прямоугольника. От точки пересечения этих линий провести перпендикуляр к оси OX. Абсцисса и будет средним значением ≈ 27.5 лет – это означает средний возраст рабочих в выборке. Значение моды (M_0) определяем по полигону частот. С этой целью находим абсциссу вершины графика полигона с точностью до масштаба, она равна 29 годам, то есть самым часто встречающимся возрастом в выборке.

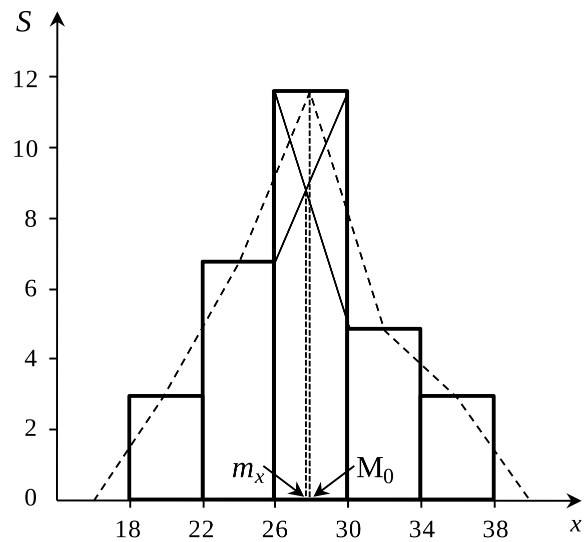


Рис. 1.10. Гистограмма и полигон частот с указанием средних тенденций

Обратим внимание на то обстоятельство, что при внешней симметрии полигона частот (гистограммы) значения математического ожидания m_x и моды (M_0) не совпадают. Медиану находим по кумуляте, графику накопленных частот. Абсциссы этого графика – варианты вариационного ряда. Ординаты – накопленные частоты.

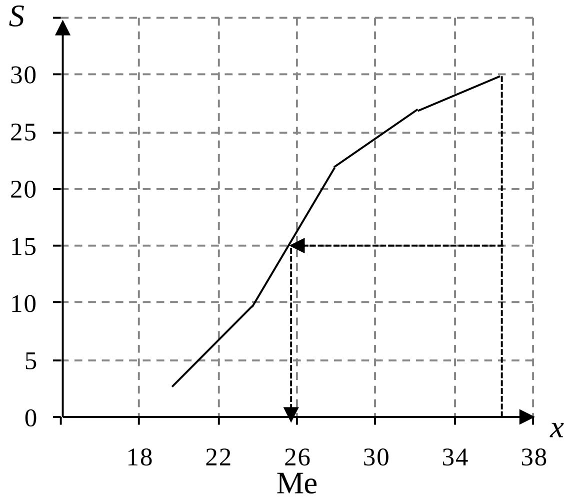


Рис. 1.11. Кумулята с найденным значением медианы

Для определения медианы по кумуляте находим по оси ординат точку, соответствующую 50% накопленных частот (в нашем случае 15), проводим через нее прямую, параллельно оси OX, и от точки ее пересечения с кумулятой проводим перпендикуляр к оси OX. Абсцисса этой точки является медианой: Me $\approx$ 25.9. Это означает, что половина рабочих в данной совокупности имеет возраст менее 26 лет.

В рассмотренных примерах хорошо видно, что владение графическими методами представления эмпирических данных дает возможность провести предварительные оценки средних тенденций без трудоемких расчетов. Получившие широкое распространение в современных условиях пакеты прикладных программ компьютерной графики значительно облегчают задачи исследователя при практическом применении графического представления данных.

1.3. Законы распределений

Рассмотрим работу бетоноотделочной машины при устройстве цементобетонного дорожного покрытия. Очевидно, что она зависит от ряда случайных факторов, влияющих на темп работы. К числу таких факторов относятся: продолжительность доставки бетонной смеси на линию (зависит, в частности, от технического состояния автомобилей-самосвалов), конкретный состав бетонной смеси, также имеющий отклонения от проектного при работе смесителя на ЦБЗ; погодные условия, техническое состояние бетоноотделочной машины в данной смене; физическое состояние членов бригады, обслуживающих машину; время, в которое будут доставлены топливо и смазочные материалы для заправки двигателя машины, и т. д. Однако существуют и другие факторы – продолжительность доставки смеси на линию и техническая исправность бетоноотделочной машины, которые могут оказаться более существенными, по сравнению с уже рассмотренными. Для качественного анализа процесса, носящего вероятностный, или, как говорят, стохастический характер кроме количественных характеристик случайной величины важно знание функции распределения $F(t)$ или плотности распределения $f(t)$.

1.3.1. Основные дискретные законы распределения

Биномиальное распределение (распределение Бернулли)

Рассмотрим ситуацию – исследователя часто интересует вероятность наступления некоторого числа успешных событий. Это может быть количество бракованных изделий в проверяемой партии (1 – бракованная, 0 – годная) или количество простоев (1 – работает, 0 – не работает) и т. д. Количество таких «успехов» будет равно сумме всех значений переменной X, т. е. количеству единичных исходов: $B = x_1 + x_2 + \ldots + x_n$. Случайная величина B называется биномиальной и принимает значения от 0 до n, например, при $B = 0$ – все детали годные, при $B = n$ – все детали бракованные. Предполагается, что все значения x независимы между собой.

Распределение случайной величины X равной количеству наступлений события A в схеме Бернулли из n испытаний называется *биномиальным распределением.*

Для практического использования достаточно просто знать формулу биномиального распределения, а также формулу, с помощью которой рассчитывают количество различных сочетаний из n элементов по k, где n – общее количество элементов, k – количество элементов, варианты расположения которых подсчитываются. Формула сочетания (известна из комбинаторики) из n элементов по k:

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Отсюда, кстати, и название биномиального распределения (формула приведенная выше является коэффициентом в разложении бинома Ньютона). Формулу для определения вероятности легко обобщить на любое количество n и k. В итоге формула биномиального распределения приобретает следующий вид:

$$P(A=k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}.$$

Рассчитаем по этой формуле вероятность выбора 40 бракованных деталей из партии в 100 деталей:

$$P(A=40) = C_{100}^{40} \cdot 0.5^{40} \cdot 0.5^{100-40} = \frac{100!}{40!(100-40)!} \cdot 0.5^{40} \cdot 0.5^{60} = 0.0108$$

или всего 1.08%.

Для сравнения вероятность наступления математического ожидания этого эксперимента, то есть 50 бракованных деталей, равна 7.96%. Максимальная вероятность биномиальной величины принадлежит значению, соответствующему математическому ожиданию.

Рассмотрим основные характеристики биномиальной переменной, то есть установим ее математическое ожидание, дисперсию и среднеквадратичное отклонение. Вспомним, что математическое ожидание суммы величин есть сумма математических ожиданий каждой складываемой величины, а оно у всех одинаковое, поэтому: $m_x = n\,p$. Например, математическое ожидание количества выпавших орлов при 100 подбрасываниях равно $100 \times 0.5 = 50$.

Выведем формулу дисперсии биномиальной переменной. Дисперсия суммы независимых случайных величин есть сумма дисперсий. Отсюда:

$$D_x = np(1-p) = npq.$$

Среднее квадратичное отклонение, соответственно: $\sigma_x = \sqrt{npq}$.

Коэффициент асимметрии: $A_x = \dfrac{q-p}{\sqrt{pq}}$.

Коэффициент эксцесса: $E_x = \dfrac{6p^2 - 6p + 1}{pq}$.

На рисунке 1.12 показан графический образ функции биномиального распределения. Не сложно заметить, что он напоминает нормальный закон.

Да, очень похоже. Еще Муавр (в 1733 г.) говорил, что биномиальное распределение при больших выборках приближается к нормальному закону, но его никто не слушал. Только Гаусс, а затем и Лаплас через 60-70 лет вновь открыли и тщательно изучили нормальной закон распределения. На графике отлично видно, что максимальная вероятность приходится на математическое ожидание, а по мере отклонения от него, резко снижается. Так же, как и у нормального закона. Биномиальное распределение имеет большое практическое значение, встречается довольно часто. Физически распределение Бернулли моделирует многократное проведение независимых реализаций одного и того же случайного эксперимента с двумя исходами: успех и неудача.

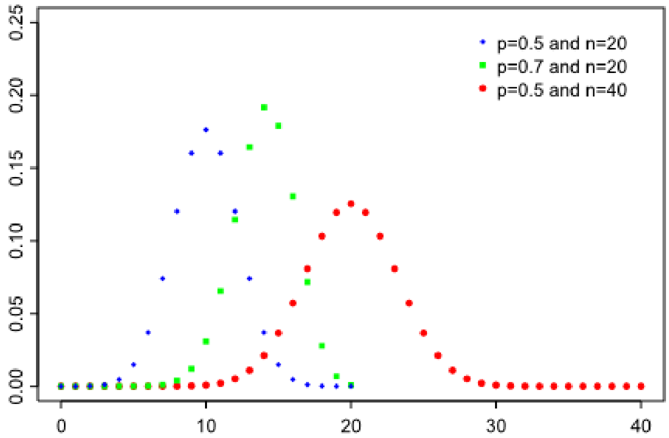


Рис. 1.12. Функция биномиального распределения

Если значения n велики, то непосредственное вычисление вероятностей событий, связанных с данной случайной величиной, технически затруднительно. В этих случаях можно использовать приближения биномиального распределения распределением Пуассона и нормальным (приближение Муавра-Лапласа).

<u>Пример</u>. Агрегат состоит из 7 независимо работающих узлов, отказ каждого равен 0.3. Определить числовые характеристики биномиального распределения числа отказов агрегата.

Решение. По условию задачи $n = 7$, $p = 0.3$, тогда $m_x = np = 7 \cdot 0.3 = 2.1$;

$D_x = npq = 7 \cdot 0.3 \cdot 0.7 = 1.47$; $\sigma_x = \pm\sqrt{1.47} = \pm 1.21$;

$A_x = \dfrac{0.7 - 0.3}{\sqrt{7 \cdot 0.3 \cdot 0.7}} = 0.33$; $E_x = \dfrac{1 - 6 \cdot 0.7 \cdot 0.3}{7 \cdot 0.3 \cdot 0.7} = -0.18$.

Распределение Пуассона получается как предельный случай распределения Бернулли, если устремить p к нулю, а n к бесконечности, но так, чтобы их произведение оставалось постоянным: $np = \lambda$. Формально такой предельный переход приводит к формуле: $p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, параметр распределения: λ (называемый интенсивностью).

Работа Пуассона «Исследования о вероятности приговоров в уголовных и гражданских делах» опубликована в 1837 году. В работе изложена четкая цель теории − вычисление для суда присяжных, долей оправданий и осуждений, которые весьма вероятно будут иметь место, и шанса ошибки судебного решения. Пуассон ввел вероятность ошибки присяжных и судей и, в отличие от Лапласа, предварительную вероятность вины подсудимого, а понятия вины и невиновности он решительно заменил на подлежащие осуждению, оправданию. Пуассон исследовал воздействие изменений в уголовном законодательстве на долю осуждений и на ошибки при осуждениях и оправданиях. Пуассон извлек статистические данные из нескольких источников, тем самым проделав большую работу. Он ни разу не упомянул ни зависимости между решениями отдельных присяжных (судей), ни неизбежных изменений во времени их отношения к делам одного и того же вида. И все-таки многие его заключения были, видимо, более или менее верными.

Распределение Пуассона обладает замечательным свойством, математическое ожидание и дисперсия имеют одинаковое значение, совпадающее с параметром распределения, то есть $m_x = D_x = \lambda$. Асимметрия и эксцесс, также связаны с этим значением: $A_x = \frac{1}{\sqrt{\lambda}}$, $E_x = \frac{1}{\lambda}$. Распределению Пуассона подчиняются очень многие случайные величины, встречающиеся в науке и практической жизни. Оно моделирует случайную величину, равную числу событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга. На рисунке 1.13 представлены варианты функции распределения Пуассона. По графикам видно, что мере возрастания значения параметра λ распределение становиться симметричным, а при $\lambda > 10$ может быть представлено нормальным законом распределения случайной величины. При $\lambda > 5$ оно стремится к Бета-распределению, которые рассматривается далее.

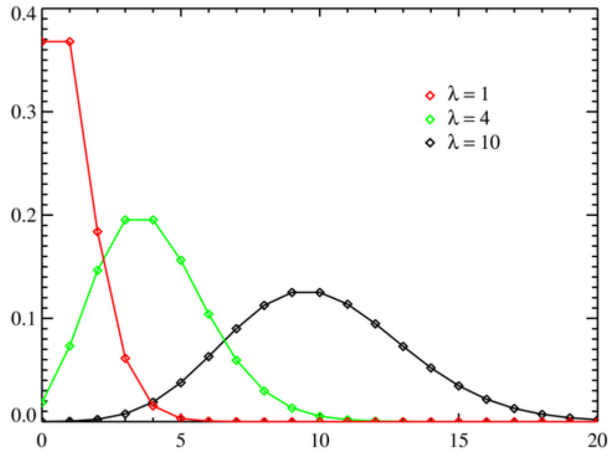


Рис. 1.13. Функция распределения закона Пуассона

Пример. Согласно проведенному хронометражу в карьере в среднем грузится 10 самосвалов в минуту. Определить вероятность погрузки 15 самосвалов в минуту.

Решение. Интенсивность погрузки, иначе $\lambda = 10$, тогда согласно закону Пуассона: $p(15) = \frac{10^{15}}{15!}e^{-10} = 0.021$ (См. приложение для $\lambda = 10$ и $n = 15$). Вероятность достаточно мала. Рассмотрим ситуацию с 12 самосвалами: $p(12) = \frac{10^{12}}{12!}e^{-10} = 0.072$, что примерно в 3.4 раза выше. Значит погрузить 12 машин более вероятно, чем 15.

Рассмотрим задачу из теории массового обслуживания.

Пример. Телефонная станция получает за час в среднем S вызовов. Какова вероятность того, что в данную минуту она получит ровно k вызовов?

Решение. Изучаемый интервал времени (1 мин.) разбиваем на n равных интервалов Δi. Работу станции в эту минуту можно рассматривать как последовательность испытаний (в каждом интервале происходит испытание – будет или нет получен вызов). Задача состоит в том, чтобы вычислить вероятность k событий в n испытаниях, где p – вероятность вызова в интервале Δi.

$$p = \frac{s}{60n}; \qquad \lambda = pn = \frac{s}{60}; \qquad P_n(k) \approx e^{-\frac{s}{60}} \frac{\left(\frac{s}{60}\right)^k}{k!}.$$

Примеры других ситуаций, которые можно смоделировать, применив это распределение: поломки оборудования, длительность исполнения ремонтных работ стабильно работающим сотрудником, ошибка печати, дефекты в длинной ленте или цепи, импульсы счетчика радиоактивного излучения и др.

Геометрическое распределение. Распределение случайной дискретной величины равной количеству испытаний случайного эксперимента до наблюдения первого успеха называется *геометрическим распределением*.

На практике эксперимент или опыт осуществляют до первого появления события A. Число проведенных попыток будет целочисленной случайной величиной $1, 2, \dots$. Вероятность появления события A в каждом опыте не зависит от предыдущих и составляет p, $q = 1 - p$. Вероятности возможных значений случайной величины X определяется зависимостью:

$$p(X = n) = q^n \cdot p.$$

Вычисление числовых характеристик для геометрического закона распределения не так сложны. Математическое ожидание: $m_x = \dfrac{1}{p}$, дисперсия: $D_x = \dfrac{q}{p^2}$, асимметрия: $A_x = \dfrac{2 - p}{\sqrt{q}}$, эксцесс: $D_x = 6 + \dfrac{p^2}{q}$.

Графики функции вероятности геометрического распределения с различными параметрами показаны на рисунке 1.14.

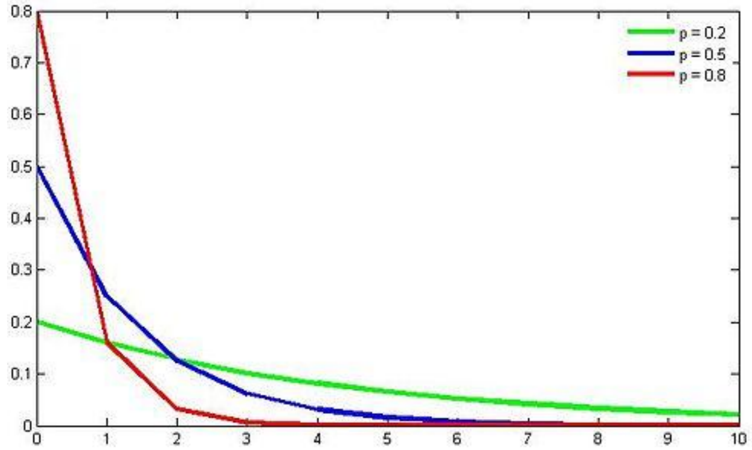


Рис. 1.14. Функция вероятности геометрического распределения

Среди дискретных случайных величин только геометрическому закону дано свойство отсутствия последействия. Это означает, что вероятность появления случайного события в k*-ом* эксперименте не зависит от того, сколько их появилось до k*-го*, и всегда равна p.

1.3.2. Основные законы распределения непрерывных случайных величин

Равномерное распределение. Непрерывная случайная величина имеет равномерное распределение на отрезке $[a,b]$, если на этом отрезке плотность распределения случайной величины постоянна, а вне этого отрезка равна нулю:

$$f(t)=\begin{cases}0, & x<a;\\ C, & a\le x\le b;\\ 0, & x>b.\end{cases}$$

Постоянная величина C может быть определена из условия равенства единице площади, ограниченной кривой распределения. График приведен на рисунке 1.15.

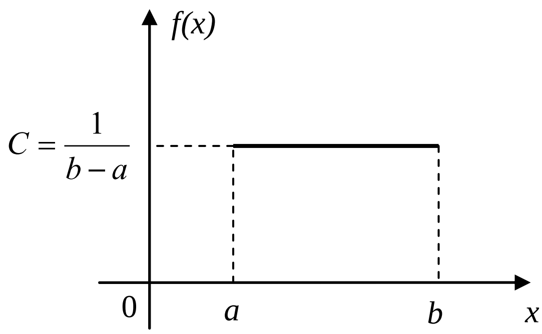
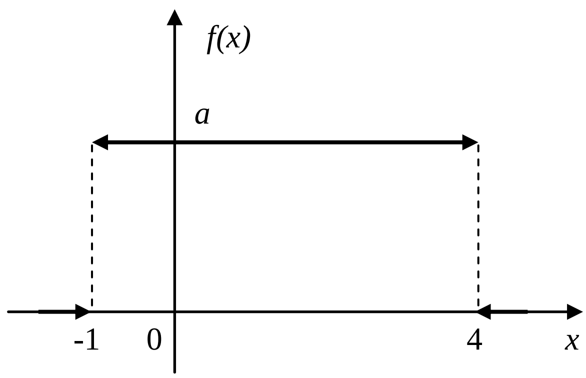


Рис. 1.15. Плотность вероятности равномерного распределения

Математическое ожидание и дисперсия равномерно распределенной непрерывной случайной величины вычисляются по формулам:

$$M(X)=\frac{a+b}{2};\qquad D(X)=\frac{(b-a)^2}{12}.$$

<u>Пример</u>. Задан график плотности распределения вероятностей непрерывной случайной величины X, распределенной равномерно в интервале $(-1;4)$. Найти значение $M(X)$.



$$M(X)=\frac{1}{4-(-)1}=\frac{1}{5}=0.2.$$

Пример. Случайная величина распределена равномерно на интервале $(-10;12)$. Найти ее математическое ожидание и дисперсию.

Параметры распределения $a=10$, $b=12$, тогда $M(X)=11$; $D(X)=\frac{1}{3}$.

Этот закон характеризует случайные величины, все значения которых в пределах заданного интервала равновероятны. Не сложно установить, что вероятность попадания случайной величины на заданный участок $[\alpha,\beta]$, содержащийся внутри интервала $[a,b]$, изменения выражается соотношением:

$$p(\alpha < X < \beta)=\frac{\beta-\alpha}{b-a}.$$

Пример. Автомиксеры на погрузку бетона подходят регулярно с интервалом 2 минуты. Погрузчик заполняется бетоном в случайный момент времени, никак не связанный с подачей автомиксера. Найти характеристики случайной величины X – времени в течение которого миксер будет ожидать погрузки. Вычислить вероятность, что ждать придется не больше полминуты.

Решение. Случайная величина X равномерно изменяется в интервале $(0;2)$ и имеет характеристики: $m_x=\frac{2+0}{2}=1$; $D_x=\frac{(0-2)^2}{12}=\frac{1}{3}$; $\sigma_x=\pm\frac{\sqrt{3}}{3}$; $p\{X<0.5\}=\frac{1}{4}$.

Показательное распределение. Непрерывная случайная величина X имеет *показательный (экспоненциальный)* закон распределения с параметром λ (параметр $\lambda>0$ – оценивается на основе реальных данных), если ее плотность вероятности имеет вид:

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0;\\ 0\quad, & x<0.\end{cases}.$$

Графики функции плотности распределения с различными значениями параметра λ приведены на рисунке 1.16.

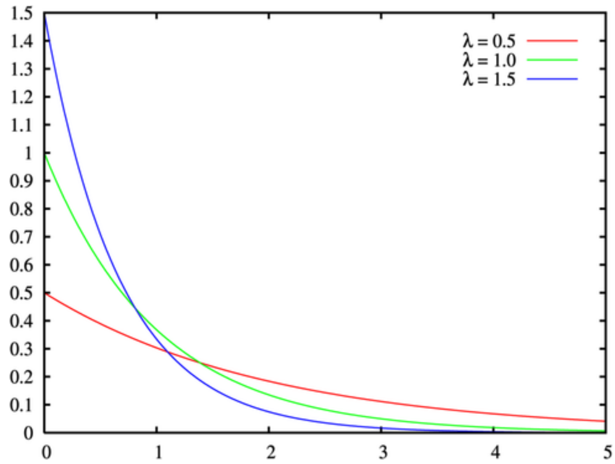


Рис. 1.16. Графики плотности распределения показательного закона

Экспоненциальная случайная величина всегда принимает положительные значения. Математическое ожидание и дисперсия случайной величины распределенной по экспоненциальному закону вычисляются по формулам:

$$M(X)=\frac{1}{\lambda};\quad D(X)=\frac{1}{\lambda^2};\quad \sigma_x=\frac{1}{\lambda}.$$

Заметим, что для данного распределения значение математического ожидания равно значению среднеквадратичного отклонения.

Например, пусть есть автозаправочная станция, на которую время от времени подъезжают автомобили. При определенных допущениях время между появлениями двух последовательных машин будет случайной величиной с экспоненциальным распределением. Параметр распределения λ может быть интерпретирован как среднее число новых автомобилей за единицу времени. Пусть это число равно 5 автомашин в минуту. В таком случае, среднее время ожидания машины на заправке составит 0.2.

Вероятность попадания непрерывной случайной величины в интервал $(\alpha;\beta)$ при показательном распределении вычисляется по следующей формуле:

$$p(\alpha,\beta)=e^{-\lambda\alpha}-e^{-\lambda\beta}.$$

Для нахождения значений функции $y=e^{-\lambda x}$ существуют специальные таблицы.

Замечание. Показательный закон распределения вероятностей встречается во многих задачах, связанных с простейшим потоком событий. Под *потоком событий* понимают последовательность событий, наступающих одно за другим в случайные моменты. Например, поток вызовов на телефонной станции, поток заявок в системе массового обслуживания и др. Также экспоненциальное распределение позволяет моделировать интервалы времени между наступлением событий. Из экспоненциальных величин строятся другие важные величины, например, случайные величины, имеющие распределение Эрланга.

Нормальное распределение. Нормальным называется распределение вероятностей непрерывной случайной величины, которое описывается плотностью вероятности:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{\frac{(x-m_x)^2}{2\sigma_x^2}}.$$

Параметры m_x и σ_x, входящие в плотность распределения являются соответственно математическим ожиданием и средним квадратическим отклонением случайной величины X.

Нормальное распределение (распределение Гаусса) характеризуется тем, что крайние значения признака в нем встречаются достаточно редко, а значения, близкие к средней величине, – достаточно часто. Название это распределение получило, вследствие его распространенности в естественнонаучных исследованиях, что казалось нормой всякого массового случайного проявления признаков. График нормального распределения представляет собой колоколообразную кривую (рис. 1.17).

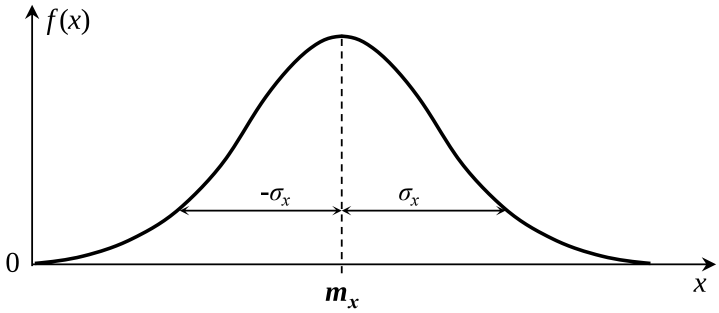


Рис. 1.17. Теоретическая кривая нормального распределения

Нормальное распределение почти всегда имеет место, когда наблюдаемые случайные величины формируются под влиянием большого числа случайных факторов, ни один из которых существенно не превосходит остальные. На рисунке 1.18 приведены графики нормального распределения при $m_x = 0$ и трех возможных значениях среднего квадратичного отклонения $\sigma_x = 1$ (верхняя кривая), $\sigma_x = 2$ (средняя кривая) и $\sigma_x = 7$ (нижняя кривая).

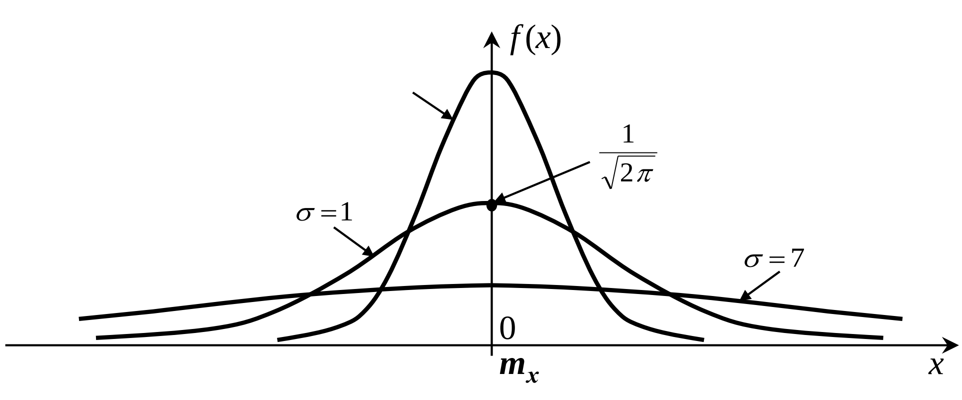


Рис. 1.18. Варианты графиков нормального распределения

Как видно, при увеличении значения среднего квадратичного отклонения график становится более пологим, а максимальное (модальное) значение уменьшается.

Для нормального (гауссова) распределения коэффициент асимметрии равен нулю, а коэффициент эксцесса равен 3.

Широкое распространение этого распределения связано с тем, что оно является бесконечно делимым непрерывным распределением с конечной дисперсией. Поэтому к нему в пределе приближаются некоторые другие, например, биномиальное и пуассоновское. Этим распределением моделируются многие не детерминированные физические процессы.

Нормальное распределение хорошо изучено. Для него существуют статистические таблицы, которые позволяют по каждому значению случайной величины найти вероятность его встречаемости, а по каждой вероятности – значения случайной величины, которые с этой вероятностью встречаются. Однако надо сделать некоторые дополнительные замечания. Дело в том, что известные статистические таблицы разработаны для, так называемого, стандартизованного нормального распределения. Иначе, для таких случайных величин, которым отвечает нулевое среднее $m_x = 0$ и единичная дисперсия, то есть $\sigma_x = 1$ поскольку, нельзя рассчитать таблицы для всех мыслимых нормальных распределений, у которых в качестве математического ожидания и дисперсии – любые положительные действительные числа).

На практике случайные величины обычно не стандартизованы. Признаки могут изменяться в разных пределах (например, доход фирмы может варьироваться от нуля до тысяч и миллионов единиц измерения, а число автомобилей в семье редко превышает три-пять), при сравнении таких признаков их требуется привести к одному виду, к единому стандарту. Наиболее распространенный вариант стандартизации – приведение распределения к стандартизированному виду (z-стандартизация). Это достигается путем вычитания из каждого не стандартизированного значения величины математического ожидания (децентрация) и деления полученной разности на стандартное отклонение (если эти параметры неизвестны, используются их выборочные оценки):

$$x' = \frac{x-\mu}{\sigma} \quad \text{или} \quad x' = \frac{x-x'}{S}.$$

Соответствующая функция распределения носит название функции Лапласа и имеет вид:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i} e^{-z^2/2} dz, \quad \text{где} \quad z = \frac{x-\mu}{\sigma}.$$

Обратим внимание, что функции Лапласа обладает свойством нечетности, иначе $\Phi(-x) = -\Phi(x)$.

Вероятность попадания случайной величины, распределенной по нормальному закону в интервал (α, β) можно вычислить по правилу:

$$p(\alpha < X < \beta) = \frac{1}{2}\left[\Phi\left(\frac{\beta - m_x}{\sigma}\right) - \Phi\left(\frac{\alpha - m_x}{\sigma}\right)\right].$$

Нормальный закон распределения обладает еще одним важным свойством, которое называется «правило трех сигм»: с вероятностью близкой к единице, нормально распределенная случайная величина не выходит за пределы $\pm 3\sigma$.

Это правило позволяет количественно оценивать параметры случайной величины, особенно при недостаточном числе измерений.

<u>Пример.</u> Из наблюдений за устройством гравийного покрытия методом смешения на дороге с вяжущим автогрейдерами было установлено, что выполнение работ на захватке протяжением 0.5 км колебалось в пределах 6-9 ч. Найти вероятность завершения работ на указанной захватке за семичасовую смену.

Решение. На основе «правила трех сигм» сначала найдем ориентировочную величину среднеквадратичного отклонения:

$$\sigma_x = \frac{T_{\max} - T_{\min}}{6} = \frac{9-6}{6} = 0.5\,\textit{часа}.$$

Найдем, математическое ожидание длительности отработки захватки $T = 7.5\,\textit{часа}$. По формуле определим искомую вероятность:

$$p(0 < T < 7) = \frac{1}{2}\left[\Phi\left(\frac{7-7.5}{0.5}\right) - \Phi\left(\frac{0-7.5}{0.5}\right)\right] = 0.5\left[-\Phi(1) + \Phi(15)\right].$$

По табл. найдем соответствующие значения функции Лапласа и вычислим величину: $P(0 < T < 7) = 0.16$.

Вернемся к примеру, рассмотренному в п. 1.2 раздела 1. Оценим степень достоверности норматива часовой выработки экскаватора. Ранее определены числовые характеристики ряда распределения: $\overline{X} = 66.4$ и $\sigma = 10.77$. Согласно симметричности нормального закона и приняв за норматив величину 66.4 м^3/ч, можно утверждать, что фактическая выработка экскаватора с вероятностью $p = 0.5$ будет либо больше, либо меньше указанного норматива. Также можно установить:

1. Если принять норматив ниже $\overline{X} = 66.4$, тогда он будет легко выполним и придется переплачивать за перевыполнение норм и, наоборот, при нормативе более $\overline{X} = 66.4$ он окажется трудновыполнимым. Найдем значение норматива с достоверностью 0.7 ($k = 1$) по формуле: $H = \overline{X} + k\sigma$.

$$H = 66.4 + 10.77 = 77.17 м^3 / ч.$$

2. Вероятность выполнения этого норматива составит:

$$p(H \le 80) = \frac{1}{2}\left[\Phi\left(\frac{80-66.4}{10.77}\right) - \Phi\left(\frac{0-66.4}{10.77}\right)\right]\sigma = 0.5\left[\Phi(1.26) + \Phi(6.17)\right] = 0.9.$$

Вывод. Норматив, рассчитанный с достоверностью равной 0.7 и с учетом среднеквадратичного отклонения выше норматива вычисленного как средняя тенденция по экспериментальным данным, одновременно вероятность его выполнения достаточно высока и составляет 90%.

Если сложить несколько независимых одинаково распределенных величин с конечной дисперсией, то сумма будет распределена *приблизительно* нормально. Например, если сложить 100 независимых стандартно равномерно распределенных случайных величин, то распределение суммы будет приближенно стандартным нормальным. Простейшие приближенные методы моделирования основываются на центральной предельной теореме. В частности, для программного генерирования нормально распределенных псевдослучайных величин предпочтительнее использовать преобразование Бокса-Мюллера. Оно позволяет генерировать одну нормально распределенную величину на базе одной равномерно распределенной.

Семейство распределений Пирсона. Распределение Пирсона *хи-квадрат* (χ^2) с числом степеней свободы n это распределение суммы квадратов n – независимых стандартных нормальных величин. Параметром распределения является значение числа степеней свободы n.

Функция плотности распределения вычисляется по формуле:

$$f(x) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}},$$ где $\Gamma(n)$ – гамма-функция.

Изображение функции плотности распределения Пирсона для разных значений степеней свободы (на рис. обозначены – k) имеет вид (рис. 1.19).

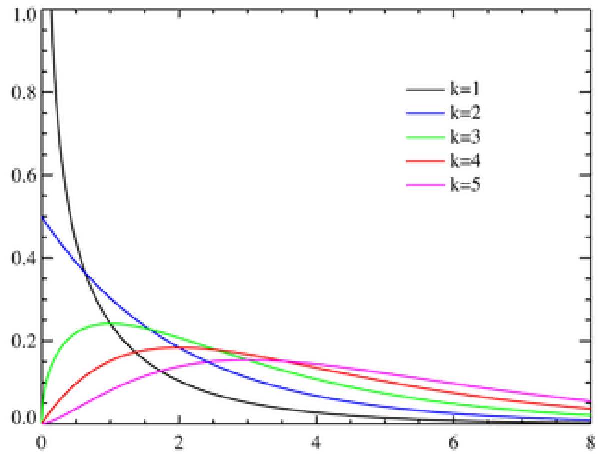


Рис. 1.19. Графики функции плотности распределения Пирсона для разных значений степеней свободы

Числовые характеристики этого закона: математическое ожидание $m_x = n$; дисперсия $D_x = 2n$; среднеквадратичное отклонение $\sigma = \sqrt{2n}$; коэффициент асимметрии $A_x = \sqrt{\dfrac{8}{n}}$; коэффициент эксцесса $E_x = \dfrac{12}{n}$.

Частными случаями распределения Пирсона являются бета-распределение (распределение Пирсона I типа), гамма-распределение (распределение Пирсона III типа), распределение Стьюдента (распределение Пирсона VII типа), показательное распределение (распределение Пирсона X типа), нормальное распределение (распределение Пирсона XI типа).

Распределения χ^2, как уже отмечалось, отличаются друг от друга числом степеней свободы. Соответственно для того чтобы определить вероятность появления некоторого значения случайной величины, требуется предварительно установить, сколько степеней свободы имеет данная величина, т. е. каким из распределений χ^2 она характеризуется. По мере увеличения числа степеней свободы асимметрия и эксцесс распределения уменьшаются. Количество степеней свободы указывает, квадраты скольких величин использовались для получения величины χ^2. Если взять из исходной совокупности по два значения, возвести их в квадрат и сложить, получим величину χ^2 с двумя степенями свободы. Аналогично можно получить кривые распределения для сумм произвольного числа квадратов стандартизированных величин.

Семейство распределений Пирсона характеризуется следующим свойством: *при больших значениях n распределение χ^2 стремится к нормальному*. В приложении приведена таблица значений χ^2 для разного числа степеней свободы и уровней доверительной вероятности. Как ею пользоваться?

Например, что означает число 11.07 для $\alpha = 0.05$ и $\eta = 5$? Это число показывает, что сумма квадратов пяти значений, случайно выбранных из нормального распределения, только в пяти случаях из ста будет превышать величину 11.07. В остальных 95 случаях она будет меньше. Другими словами, вероятность получить значение критерия Пирсона в интервале от 0 до 11.07 равна 0.95. Критическая область для распределения χ^2 всегда односторонняя (рис. 1.20).

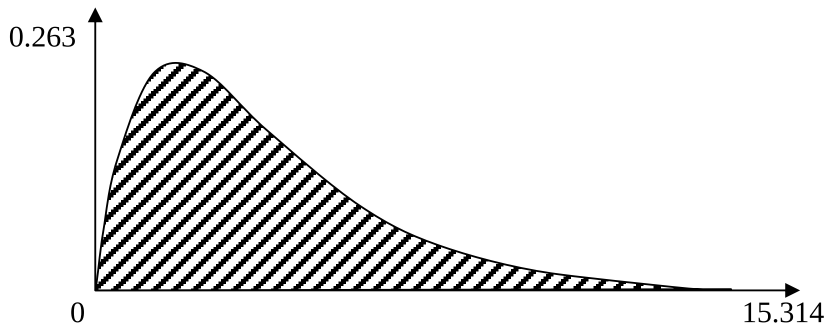


Рис. 1.20. Функция плотности распределения χ^2 (υ = 4)

Распределение Пирсона широко используются в математической статистике при сглаживании распределений эмпирических данных. Для аппроксимации распределения вероятностей опытных данных численными методами вычисляют их первые четыре момента, а затем на их основе вычисляют параметры распределения Пирсона.

Семейство распределений Стьюдента (t-распределения). Это однопараметрическое семейство абсолютно непрерывных распределений. Названо в честь Уильяма Сили Госсета, который первым опубликовал работы, посвященные этому распределению, под псевдонимом «Стьюдент».

Распределение Стьюдента играет важную роль в некоторых широко используемых системах статистического анализа. Пример такой системы, *t*-критерий Стьюдента для оценки статистического значения разницы между двумя выборочными средними, построения доверительных интервалов разницы между двумя доверительными средними, а также в линейном регрессионном анализе. Распределение Стьюдента также появляется в байесовском анализе данных, распределенных по нормальному закону. Распределение Стьюдента может быть использовано для оценки того, насколько вероятно, что истинное среднее находится в каком-либо заданном диапазоне.

График плотности распределения Стьюдента, как и нормального распределения, является симметричным и колоколообразным, но с более тяжелыми хвостами, из-за этого, величины с распределением Стьюдента чаще значительно отличаются от математического ожидания. Критическая область распределения Стьюдента может быть как одно-, так и двусторонней, в зависимости от того, какую вероятность требуется определить (рис. 1.21).

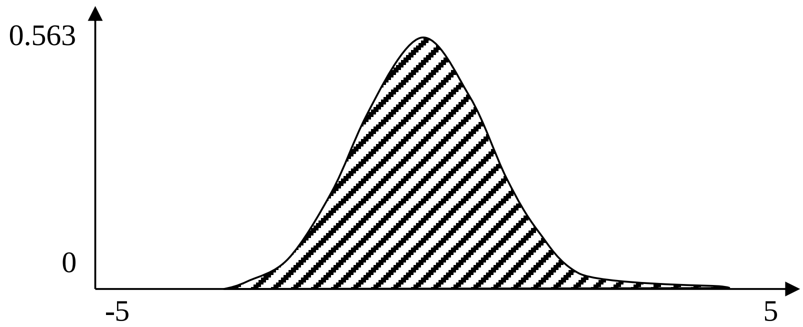


Рис. 1.21. Функция плотности распределения Стьюдента ($\upsilon = 21$)

Семейства *t*-распределения также меняются в зависимости от числа степеней свободы. При большом числе степеней свободы распределение Стьюдента приближается к нормальному распределению. Эксцесс распределения Стьюдента обратно пропорционален числу степеней свободы, а асимметрия при числе степеней свободы больших трех равна нулю. Это важно для понимания статистического поведения определенных типов соотношений случайных величин, в которых отклонение в знаменателе увеличено и может производить отдаленные величины, когда знаменатель соотношения близок к нулю.

Распределение Стьюдента – частный случай обобщенного гиперболического распределения. Оно является результатом комбинации распределения Гаусса (нормального распределения) со средним значением m_x и неизвестной дисперсией, с обратным гамма-распределением. Другими словами, первоначально предполагается, что случайная величина X обладает нормальным распределением с неизвестной дисперсией, распределенной как обратная гамма, а затем дисперсия исключается. Такое свойство полезно из-за того, что обратное гамма-распределение – это сопряженное априорное распределение дисперсии распределения Гаусса, именно поэтому нестандартизированное распределение Стьюдента естественным образом возникает во многих байесовских задачах. Эквивалентно, это распределение является результатом комбинации распределения Гаусса с масштабированным обратным хи-квадрат распределением. Масштабированное обратное хи-квадрат распределение – точно тоже распределение, что и обратное гамма-распределение, но с другой параметризацией.

***Семейство распределений Фишера* (*F-распределения*).** Распределение Фишера существует в виде семейства распределений, которые задаются двумя значениями степеней свободы. Оно представляет собой отношение двух случайных величин, распределенных по закону χ^2, стандартизованных на степени свободы.

Пусть Y_1 и Y_2 – две независимые случайные величины, имеющие распределение хи-квадрат: $Y_i \sim \chi^2(d_i)$, где $d_i \in N$. Тогда распределение случайной величины называется распределением Фишера (распределением Снедекора) со степенями свободы d_1 и d_2:

$$F = \frac{Y_1 / d_1}{Y_2 / d_2}.$$

На этом основании *F*-распределение называют также распределением дисперсионного отношения. Критическая область *F*-распределения может быть односторонней – если проверяется гипотеза о превосходстве какой-либо из двух величин; или двусторонней – если проверяется гипотеза о превосходстве первой величины над второй и второй над первой (соответственно в данном случае вычисляются два критических значения F_1 и F_2, причем $F = 1/F_1$). Асимметрия и эксцесс *F*-распределения зависят от соотношения между степенями свободы (рис. 1.22).

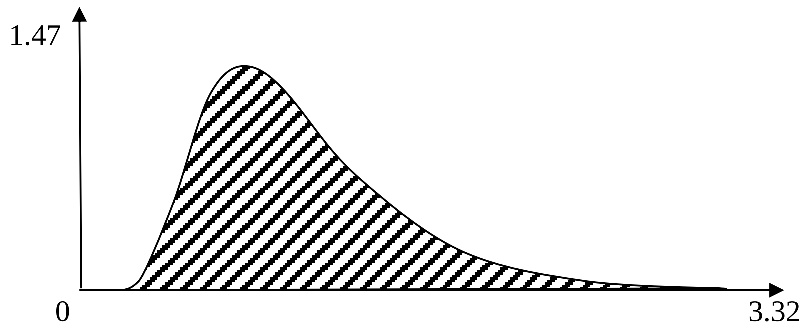


Рис. 1.22. Функция плотности распределения Фишера (d_1 = 22, d_2 = 16)

Практические расчеты с использованием приведенных распределений требуют тщательной группировки данных и достаточно сложных вычислений. Кроме того, их возможности в полной мере проявляются на больших выборках, когда количество измерений превышает тридцать случаев. Одновременно они незаменимы для получения оценок в двух случаях: доказательство неслучайности предпочтений в выборе из нескольких альтернатив; обнаружение точки максимального расхождения между двумя распределениями, которая затем используется для перегруппировки данных.

1.3.3. Распределения, востребованные в строительстве

Бета (β) - распределение. Это распределение характеризуется ограниченным интервалом [0, 1] в пределах которого может находиться случайная величина и возникает в условиях, когда эта величина зависит от большого числа случайных незначительных факторов при наличии нескольких существенных. Нормальный закон в этом смысле является частным случаем β-распределения, когда несколько существенных случайных факторов становятся каждый сам по себе малосущественными.

Функция плотности β-распределения имеет вид:

$$f(x) = \int_0^1 x^{\gamma-1}(1-x)^{\eta-1}\,dx = \frac{\Gamma(\gamma)\Gamma(\eta)}{\Gamma(\gamma+\eta)},$$

где $\Gamma(a)$ – гамма-функции параметров распределения γ и η.

Значения гамма-функции табулированы и приводятся во многих математических справочниках. Функция плотности вероятности β-распределения имеет варианты изображений, представленные на рисунке 1.23.

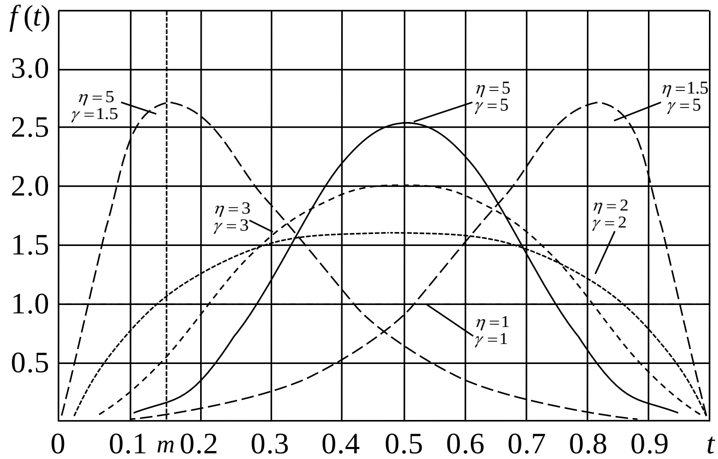


Рис. 1.23. График плотности бета-распределения

Параметры β-распределения (γ и η), определяют форму кривой $f(t)$. Как уже отмечалось, площадь, ограниченная кривой плотности распределения, всегда равна 1. Тогда для определения параметров можно получить соотношения:

$$\eta = \frac{1-\bar{X}}{\sigma^2}\left[\bar{X}\left(1-\bar{X}\right)-\sigma^2\right] \quad \text{и} \quad \gamma = \frac{\bar{X}\cdot\eta}{1-\bar{X}}.$$

Графики плотности β-распределения при различных значениях параметров γ и η показаны на рисунках 1.24 (а, б, в).

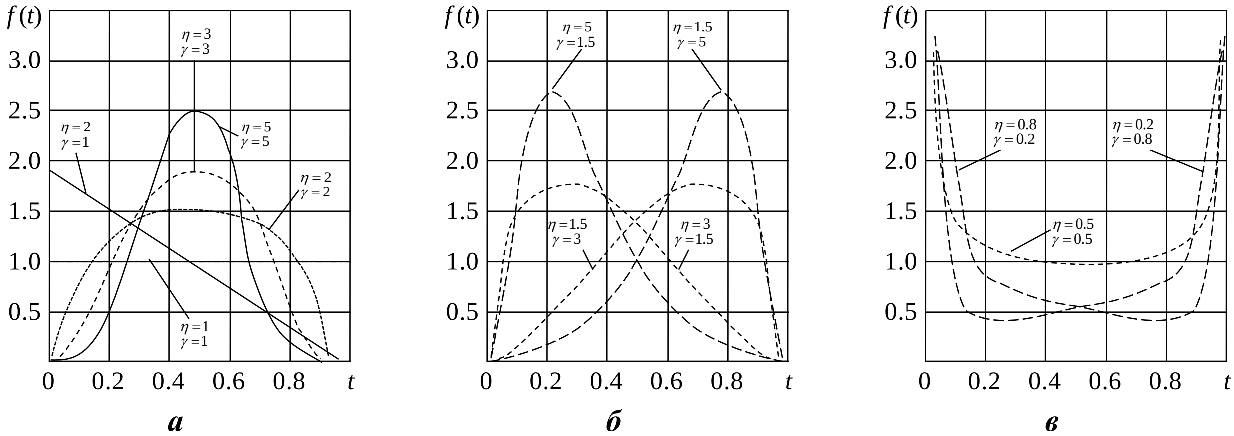


Рис. 1.24. Различные формы кривых плотности β-распределения

При $\gamma = n$ бета-распределение является симметричным и, в частности, при $\gamma = \eta = 1$ переходит в равномерное распределение. В этих условиях распределение является одновершинным, т. е. унимодальным. Поэтому, если найти производную, то легко определяется найти точка, в которой функция достигает максимума. Это дает следующее значение моды:

$$M_0 = \frac{\gamma - 1}{\gamma + \eta - 2}.$$

Например, для кривых, показанных на рис. 1.24 *а*, при $\gamma = \eta = 5$ соответственно получаем: $M_0 = \dfrac{5-1}{5+5-2} = 0.5$, т. е. мода соответствует центру интервала изменений случайной величины, что обязательно для симметричных распределений.

Для количественного анализа вероятностных процессов, соответствующих β-распределению, необходимо знать параметры формы кривой $f(t)$. Они могут быть получены на основе статистического эмпирического материала (рис. 1.25).

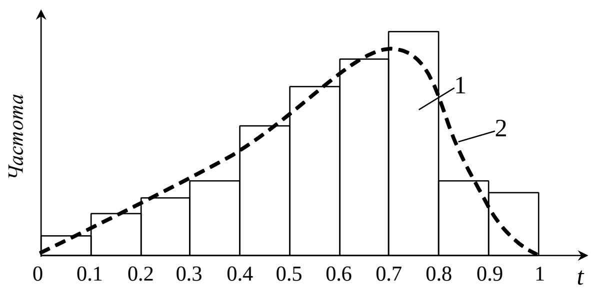


Рис. 1.25.Сопоставление эмпирического распределения (1)
с теоретической кривой (2) плотности β-распределения

Как ясно из приведенных рисунков кривая плотности β-распределения имеет различную форму и в частных случаях дает ряд других распределений. Поэтому β-распределение широко используется для описания большого числа случайных процессов. Так, например, оно применяется для оценки доли дефектных изделий на производственной линии в единицу времени, в теории надежности для определения с любой заранее задаваемой вероятностью периода безотказной работы устройств и их элементов, а также во многих других случаях.

Пример. Рассчитать параметры β-распределения для статистического ряда выработки экскаватора (см. п. 1.2 раздела 1).

Решение. Так как плотность β-распределения определена на интервале [0, 1] примем за ноль – выработку 30 *м*^{2}/*ч*, а за единицу – выработку 110 *м*^{2}/*ч*. Рассчитаем длину каждого равномерного интервала: 1 : 8 = 0.125. Поставим в соответствие каждому реальному интервалу – рассчитанный единичный, а точнее интервалу [30, 40] соответствует значение 0.0625, далее [40, 50] – 0.125 + 0.0625 и так до конца исходного интервала. Рассчитаем статистические характеристики распределения на единичном интервале.

$$\overline{X} = \frac{(0.125+0.0625)\cdot 4+\ldots(0.75+0.0625)\cdot 2}{100} = 0.46; \qquad \sigma^2{}_x = 0.028.$$

Найдем параметры β-распределения:

$$\eta = \frac{1-0.46}{0.028}\left[0.46(1-0.46)-0.028\right] = 4.2; \quad \gamma = \frac{0.46\cdot 4.2}{1-0.46} = 3.6.$$

Параметры распределения близки к цифре 4, что позволяет говорить о практической симметричности статистического ряда и с некоторой долей достоверности может быть представлена также нормальным законом.

Гамма (γ) - распределение. Предположим, есть деньги и задание: «Купить для фирмы несколько одинаковых мониторов». При этом точное количество не уточняется. Предположим также, что известен тот факт, что монитор с диагональю на два дюйма больше стоит вдвое дороже. Тогда количество мониторов, которые можно купить на данную сумму может описать γ-распределение.

Гамма (γ)-распределение описывает неотрицательную случайную величину, формула плотности распределения которой имеет вид:

$$f_k(x) = \frac{\lambda^k \cdot x^{k-1} \cdot e^{-\lambda x}}{\Gamma(k)}, \text{ для } x > 0,$$

где $\lambda > 0$ – параметр формы; $k > 0$ – параметр масштаба; $\Gamma(k)$ – гамма-функция:

$$\Gamma(k) = \int_0^\infty e^{-t} t^{k-1} dt.$$

Если k–целое неотрицательное число, то $\Gamma(k) = k!$, тогда γ-распределение превращается в *распределение Эрланга* k*–го порядка* с функцией плотности распределения вида:

$$f_k(x) = \frac{\lambda(\lambda x)^{k-1} \cdot e^{-\lambda x}}{(k-1)!}.$$

Датский ученый К.А. Эрланг (1878-1929), сотрудник Копенгагенской телефонной компании, изучал функционирование телефонных сетей. С его работ началось развитие теории массового обслуживания. Эта теория занимается вероятностно-статистическим моделированием систем, в которых происходит обслуживание потока заявок, с целью принятия оптимальных решений. Распределения Эрланга используют в тех же прикладных областях, в которых применяют экспоненциальные распределения.

В частном случае γ-распределение при $k = 1$ переходит в *экспоненциальное* распределение с параметром λ. Плотность экспоненциального распределения имеет моду в точке 0, что не всегда удобно для практического применения. Во многих примерах заранее известно, что мода рассматриваемой случайной переменной не равна 0, например, интервалы между подходами самосвалов к экскаватору имеют ярко выраженную моду. Для моделирования таких событий используется γ-распределение.

Стоит также отметить, что предельным случаем γ-распределения является нормальный закон.

Основные характеристики γ-распределения:

математическое ожидание: $m_x = \dfrac{k}{\lambda}$; дисперсия: $D_x = \dfrac{k}{\lambda^2}$.

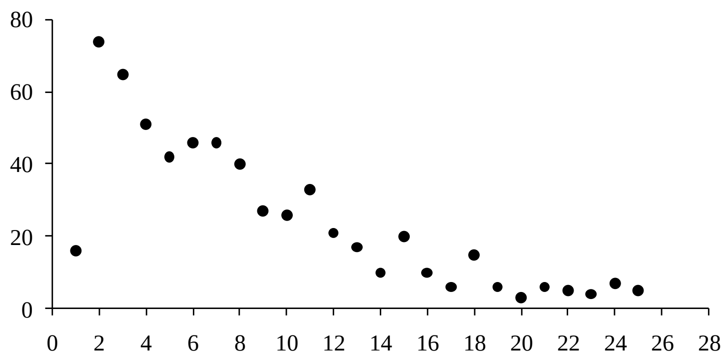
Гамма-распределение и его варианты являются одними из наиболее распространенных в статистической теории вероятностных распределений. В частности, γ-распределение имеют выходные эффекты обработки сигналов при приеме дискретных гауссовских процессов. Или технико-экономическое обоснование результатов проектирования мостового перехода при определении расчетного расхода выполняется по методу, основанному на γ-распределении.

<u>Пример</u>. Рассчитать параметры γ-распределения величины задержки автомобилей на участке дороги, исходя из предположения, что кривая распределения проходит через начало координат.

Исходные данные приведены в таблице.

Интервалы измерения, t_i	Частота, n_i	Интервалы измерения, t_i	Частота, n_i
0.0 – 1.0	16	*13.0 – 14.0*	10
1.0 – 2.0	74	*14.0 – 15.0*	20
2.0 – 3.0	65	*15.0 – 16.0*	10
3.0 – 4.0	51	*16.0 – 17.0*	6
4.0 – 5.0	42	*17.0 – 18.0*	15
5.0 – 6.0	46	*18.0 – 19.0*	6
6.0 – 7.0	46	*19.0 – 20.0*	3
7.0 – 8.0	40	*20.0 – 21.0*	6
8.0 – 9.0	27	*21.0 – 22.0*	5
9.0 – 10.0	26	*22.0 – 23.0*	4
10.0 – 11.0	33	*23.0 – 24.0*	7
11.0 – 12.0	21	*24.0 – 25.0*	5
12.0 – 13.0	17		

Решение. Построим прообраз закона распределения в виде полигона частот.



Используем формулы статистических характеристик γ-распределения математическое ожидание: $m_x = \dfrac{k}{\lambda}$, дисперсия: $D_x = \dfrac{k}{\lambda^2}$, на основании которых найдем $\lambda > 0$ – параметр формы; $k > 0$ – параметр масштаба.

Предварительно рассчитаем $m_x = \bar{t} = \dfrac{1}{\sum n_i} \sum_i t_i \cdot n_i = 7.45$ и дисперсию $D_x = 31.0$.

Отсюда $m_x = \dfrac{k}{\lambda} = 7.45$ и $D_x = \dfrac{k}{\lambda^2} = 31$, решая совместно эти два уравнения находим $k = 1.79$ и $\lambda = 0.24$. Таким образом, γ-распределение случайной величины задержки автомобилей на участке дороги можно приблизить распределением Эрланга с параметром $k = 2$.

Распределение Вейбулла является частным случаем обобщенного распределения экстремальных значений. Распределение Вейбулла хорошо описывает время безотказной работы ряда устройств (электронные лампы, реле, шариковые подшипники, некоторые детали автомобилей и т. п.).

Если величину X принять за наработку до отказа, тогда получается распределение, в котором интенсивность отказов пропорциональна времени. Функция плотности распределения определяется формулой:

$$f(x) = \left(\frac{k}{\lambda}\right)\left(\frac{x}{\lambda}\right)^{(k-1)} \cdot e^{-\left(x/\lambda\right)^k}, \; x \geq 0.$$

λ – коэффициент масштаба кривой; k – коэффициент формы.

Характеристики распределения Вейбулла:

математическое ожидание: $m_x = \lambda \cdot \Gamma\left(1 + \frac{1}{k}\right)$;

дисперсия: $D_x = \lambda^2 \cdot \Gamma\left(1 + \frac{2}{k}\right) - \mu^2$.

Вид функции плотности распределения Вейбулла сильно зависит от значения k. Для $0 < k < 1$ плотность стремится к бесконечности при $x \to 0+$ и строго убывает. Для $k = 1$ плотность стремится к $1/\lambda$ при $x \to 0+$ и также строго убывает. Для $k > 1$ плотность стремится к 0 при $x \to 0+$, возрастает до достижения своей моды и убывает после.

Интересно отметить, что плотность имеет бесконечный отрицательный угловой коэффициент в $x = 0$ при $0 < k < 1$, бесконечный положительный угловой коэффициент в $x = 0$ при $0 < k < 2$, и нулевой угловой коэффициент в $x = 0$ при $k > 2$. При $k = 2$ плотность имеет конечный положительный угловой коэффициент в $x = 0$. Кроме того, коэффициент асимметрии и коэффициент вариации зависят только от коэффициента формы.

На рисунке 1.26 представлены варианты графиков функции плотности распределения Вейбулла.

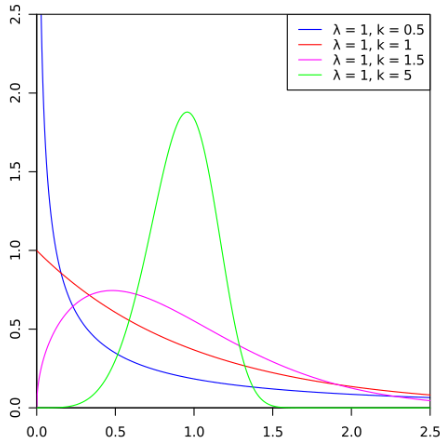


Рис. 1.26. Плотность вероятности распределения Вейбулла

Это распределение применяется для количественной оценки надежности устройств с учетом периода приработки. Частота отказов определяется распределением Вейбулла и в зависимости от значений параметра формы показывает:

для $k<1$ интенсивность отказов уменьшается со временем;

для $k=1$ интенсивность отказов не меняется со временем;

для $k>1$ интенсивность отказов увеличивается со временем.

Если принять $k=1$, т. е. интенсивность отказов постоянна, причем $\frac{1}{\lambda}=a$ соответствует параметру в экспоненциальном распределении. Таким образом, экспоненциальное распределение является частным случаем распределения Вейбулла. При $k=2$ получается распределение Релея, имеющее определенное применение в теории надежности.

Рассмотренные распределения не могут описать изменения частоты $\alpha(t)$ и интенсивности $X(t)$ отказов на всем интервале времени работы устройства. Поэтому часто используется суперпозиция (сложение) нескольких распределений. При этом записывается вначале выражение для $\alpha(t)$, например, суперпозиция двух экспоненциальных распределений, а затем обычным порядком находятся $P(t)$, $X(t)$.

Точное вычисление параметров распределения Вейбулла по данным испытаний весьма сложно. Поэтому обычно применяется приближенный метод. Имея данные испытаний, можно для любых двух значений t вычислить два значения $Q(t)$ и получить легко разрешаемую систему из двух уравнений с двумя неизвестными λ и k.

Раздел 2. ПРИКЛАДНЫЕ МЕТОДЫ СТАТИСТИКИ

2.1. Точечные и интервальные оценки

Точечной называют статистическую оценку $\theta^* = \theta^*(X_1, X_2, ..., X_n)$ параметра θ, определяемую *одним* числом. Точечные оценки зависят от объема выборки и обычно используются в выборках большого объема. Оценки параметров совокупности, полученные по разным выборкам, как правило, отличаются друг от друга.

Ошибкой выборки (оценивания) называют абсолютную разность $\left|\theta^* - \theta\right|$. Выборочная средняя $\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$ есть несмещенная и состоятельная точечная оценка математического ожидания $M(X)$.

Выборочная дисперсия $D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 \cdot n_i$ – это точечная смещенная оценка генеральной дисперсии $D(X)$. Несмещенной оценкой генеральной дисперсии $D(X)$ является «исправленная» выборочная дисперсия $S_в^2 = \frac{n}{n-1} \cdot D_в$.

«Исправленная» выборочная дисперсия является состоятельной оценкой дисперсии $D(X)$. При достаточно больших значениях n выборочная и «исправленная» дисперсии различаются мало.

Если известно математическое ожидание $M(X)$ случайной величины, то выборочная дисперсия $D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - M(X))^2 \cdot n_i$ несмещенная, состоятельная и эффективная оценка генеральной дисперсии $D(X)$. Относительная частота $\frac{n_i}{n}$ – несмещенная и состоятельная оценка вероятности $P(X = x_i)$.

Пример. В таблице приведены данные о средней заработной плате и численности занятых рабочих и сотрудников по строительным предприятиям и фирмам в сравнении по округам.

Федеральный округ	Средний размер заработной платы, x_i, тыс. руб.	Численность занятых, n_i, млн. чел.
Центральный	35.67	20.33
Северо-Западный	32.29	7.15
Южный	22.38	6.41
Северо-Кавказский	19.30	3.95
Приволжский	22.41	14.74
Уральский	34.03	6.08
Сибирский	26.48	9.05
Дальневосточный	37.26	3.19

Необходимо найти несмещенную оценку дисперсии заработной платы.

Решение. Найдем средний размер заработной платы по всем округам:

$$\overline{x}_в = \frac{\sum_{i=1}^{8} x_i \cdot n_i}{\sum_{i=1}^{8} n_i} = (35.67 \cdot 20.33 + 32.29 \cdot 7.15 + 22.38 \cdot 6.41 + 19.30 \cdot 3.95 +$$

$$+ 22.41 \cdot 14.74 + 34.03 \cdot 6.08 + 26.48 \cdot 9.05 + 37.26 \cdot 3.19):(20.33 + 7.15 + 6.41 +$$

$$+ 3.95 + 14.74 + 6.08 + 9.05 + 3.19) = \frac{2071.4646}{70.9} \approx 29.21670804.$$

Для вычисления выборочной дисперсии воспользуемся формулой:

$$D_в = \overline{x^2} - (\overline{x_в})^2,$$

где $\overline{x^2}$ – выборочная средняя квадратов вариантов выборки.

$$\sum_{i=1}^{8} x_i^2 \cdot n_i = (35.67^2 \cdot 20.33 + 32.29^2 \cdot 7.15 + 22.38^2 \cdot 6.41 + 19.30^2 \cdot 3.95 +$$

$$+ 22.41^2 \cdot 14.74 + 34.03^2 \cdot 6.08 + 26.48^2 \cdot 9.05 + 37.26^2 \cdot 3.19) = 63221.54518;$$

$$\overline{x^2} = \frac{\sum_{i=1}^{8} x_i^2 \cdot n_i}{\sum_{i=1}^{8} n_i} = \frac{63221.545186}{70.9} \approx 891.7002142;$$

$$D_в = 891.7002142 - (29.21670804)^2 \approx 38.08418554.$$

Найдем несмещенную оценку дисперсии:

$$S_в^2 = \frac{n}{n-1} \cdot D_в = \frac{70.9}{69.9} \cdot 38.08418554 \approx 38.62902367.$$

При большом объеме выборки выборочное среднее, выборочную и исправленную дисперсии удобно вычислять, используя вспомогательную таблицу:

x_i	n_i	$x_i \cdot n_i$	$x_i - \overline{x_в}$	$(x_i - \overline{x_в})^2$	$(x_i - \overline{x_в})^2 \cdot n_i$
x_1	n_1	$x_1 \cdot n_1$	$x_1 - \overline{x_в}$	$(x_1 - \overline{x_в})^2$	$(x_1 - \overline{x_в})^2 \cdot n_1$
...	...	...	...	...	...
x_k	n_k	$x_k \cdot n_k$	$x_k - \overline{x_в}$	$(x_k - \overline{x_в})^2$	$(x_k - \overline{x_в})^2 \cdot n_k$
Сумма	$n = \sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} x_i \cdot n_i$			$\sum_{i=1}^{k} (x_i - \overline{x_в})^2 \cdot n_i$

Замечание. Для проведения статистической обработки информации используют также табличный процессор Microsoft Excel, включающий в себя программную надстройку «Пакет анализа» и библиотеку статистических функций. В практической деятельности такого набора инструментов бывает вполне достаточно для проведения довольно полного и качественного статистического анализа информации.

Статистическая оценка параметров закона распределения случайной величины X, характеризующаяся двумя числовыми значениями – концами интервала, называется *интервальной*. Интервал, в который с заданной вероятностью (надежностью), попадает оцениваемый параметр, называется *доверительным*. На практике обычно используют два типа доверительных интервалов: *симметричные* и *односторонние*.

Вероятность γ, с которой выполняется неравенство $|\theta^* - \theta| < \varepsilon$, называется *доверительной вероятностью* (*надежностью*) оценки θ для заданного ε. Точностью оценки ε называют число, равное половине длины доверительного интервала. Доверительная вероятность связана не с величиной параметра θ, а лишь с границами интервала, которые меняются при изменении выборки. Вероятность $\alpha = 1 - \gamma$ называется *уровнем значимости* (*вероятностью ошибок*). Общепринятые значения уровня значимости в инженерных исследованиях – 0.1; 0.05; 0.01.

Доверительными границами (*критическими значениями*) называют границы доверительного интервала. Доверительные границы зависят от закона распределения параметра θ^*. Доверительный интервал, определяемый выборкой – случайной величиной, носит случайный характер и относительно длины, и по расположению относительно θ^*, поэтому принято говорить, что доверительный интервал покрывает параметр θ с доверительной вероятностью γ.

Точность оценки ε, доверительная вероятность γ и объем выборки n связаны между собой. Методы построения доверительных интервалов различны. Один из алгоритмов построения доверительного интервала состоит в следующем:

1. Производим выборку объема n случайной величины X из генеральной совокупности с известным распределением $f(x;\theta)$.

2. Находим точечную оценку θ^* неизвестного параметра θ по данным выборки.

3. Задаем доверительную вероятность γ.

4. Определяем границы доверительного интервала $(\theta^* - \varepsilon; \theta^* + \varepsilon)$. Для этого возьмем произвольное значение θ и, используя плотность вероятности, найдем функцию распределения из условия:

$$P\left\{\left|\theta - \theta^*\right| < \varepsilon\right\} = \int_{\theta^*-\varepsilon}^{\theta^*+\varepsilon} f(x,\theta)dx = \gamma .$$

Границы интервала определим из решения уравнений:

$$P(X(\theta) < \theta^* - \varepsilon) = \frac{1-\gamma}{2}; \quad P(X(\theta) > \theta^* + \varepsilon) = \frac{1-\gamma}{2}.$$

Полученный интервал с доверительной вероятностью γ «покрывает» неизвестный параметр θ и является его интервальной оценкой.

Пример. Обследование кадрового состава управленческого звена, насчитывающего 20 сотрудников организации по уровню образования дало следующие результаты.

Номера анкеты	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Число лет получения образования	15	14	15	15	10	13	17	18	15	14	15	15	13	10	18	16	15	14	15	15

В данной организации считается, что в среднем 15 лет (10 лет в школе и 5 лет в вузе) – стандартный уровень образования для сотрудника, а если сотрудник обучался менее 15 лет, его уровень образования недостаточен. Можно ли на уровне значимости α = 0.05 и α = 0.01 говорить о том, что средний уровень образования сотрудников (оцениваемый с помощью среднего арифметического) ниже стандарта? Общее число сотрудников организации составляет 142 человека.

Решение. Найдем число степеней свободы υ = 142 – 1 = 141. Значение *t*-статистики = 1.98 для α = 0.05; υ = 141 и значение *t*-статистики = 2.61 для α = 0.01; υ = 141. Рассчитаем среднее арифметическое и стандартное отклонение: $\bar{x}$ = 14.6, s = 2.06. При α = 0.05 границы доверительного интервала находим по формуле: $\bar{x} \pm 1.98 \frac{S}{\sqrt{n}} = 14.6 \pm 1.98 \frac{2.00}{\sqrt{142}} = 14.6 \pm 0.34$. Отсюда первый интервал: $14.26 \le \mu \le 14.94$. Находим границы доверительного интервала при $\alpha = 0.01$: $\bar{x} \pm 2.58 \frac{S}{\sqrt{n}} = 14.6 \pm 0.45$. Отсюда второй интервал: $14.15 \le \mu \le 15.05$.

Таким образом, на уровне значимости $\alpha = 0.01$ уровень образования отвечает стандарту, а вот на уровне $\alpha = 0.05$ он находится несколько ниже стандарта. Обратим внимание, что во втором случае доверительный интервал меньше, но вероятность ошибки выше. Иными словами, если действительно уровень образования в организации соответствует стандарту, вероятность ошибочно заключить, что это не так, выше при $\alpha = 0.05$, чем при $\alpha = 0.01$.

2.2. Виды статистических гипотез

Развитие научных знаний свидетельствует о том, что на определенной стадии формирования представлений о некотором объекте становится невозможным его описание только лишь путем установления непосредственных зависимостей между эмпирическими данными. И дело здесь не только в достоверности собранных фактов, а в способах их согласования и необходимости поиска гипотезы, позволяющей судить о природе закономерностей, доступных наблюдению.

Гипотеза – это научно обоснованное предположение о структуре объектов, о характере элементов и связей, образующих эти объекты, о механизме их функционирования и развития.

Процесс установления истинности или ложности гипотезы есть процесс ее эмпирического обоснования. В результате такого исследования гипотезы либо опровергаются, либо подтверждаются и становятся положениями теории, истинность которых уже доказана. Общую гипотезу обычно получают в результате предварительного анализа изучаемого объекта. Однако в частном эмпирическом исследовании специалисты предметной области сталкиваются с отдельными сторонами изучаемого объекта, с его элементами и связями. Поэтому в частных эмпирических исследованиях проверяются не сами гипотезы, а их следствия, т. е. частные положения, логически вытекающие из гипотезы и описывающие отдельные элементы и связи.

Нередко одна и та же научная гипотеза подтверждается одними фактами и опровергается другими. Поэтому факты должны быть правильно истолкованы, для того чтобы служить средством проверки. Если установлено, что выведенные следствия ложны и не соответствуют данным, полученным в исследовании, то гипотеза отвергается. По содержанию предположений относительно изучаемого объекта различают описательные и объяснительные гипотезы.

Описательные гипотезы – это предположения о структурных и функциональных связях изучаемого объекта, которые могут относиться и к классификационным характеристикам объекта.

Объяснительные гипотезы – это предположения о причинно-следственных связях в изучаемом объекте.

В прикладной деятельности часто приходится принимать решения по результатам выполненных измерений. Одним из компонентов этого процесса можно считать проверку статистических гипотез. Любое предположение о виде и свойствах наблюдаемых в эксперименте случайных величин называется *статистической гипотезой*. Например, в схеме Бернулли одной из гипотез будет предположение «вероятность успеха равна 0.25», в случае нормального распределения – «теоретическая функция распределения нормальна с дисперсией, не превосходящей квадрата среднего значения». Нахождение точечных или интервальных оценок, как правило – предварительная стадия статистических исследований. На этом шаге обычно формулируется предположение относительно вида функции распределения (*непараметрическая гипотеза*), либо предположения относительно значений параметров функции распределения (дисперсия или математическое ожидание) известного вида (*параметрическая гипотеза*). Проблемы, возникающие при исследовании случайной величины, сводятся к оценке истинности одной или нескольких выдвигаемых гипотез на основе результатов анализа накопленной информации. В ходе решения практических задач выдвигаются следующие гипотезы:

- об общем виде закона распределения исследуемой случайной величины (проверка таких гипотез дает возможность установить закон распределения с точностью до параметров, которые характеризуют неизвестный экспериментатору закон распределения);
- об однородности двух или нескольких выборок (такие гипотезы позволяют сделать вывод о равенстве или различии законов распределения случайной величины, характеризующих изучаемое свойство);

➢ о числовых значениях характеристик исследуемого явления или процесса (используя эти гипотезы, сравнивают значения числовых параметров с заданными значениями);

➢ об общем виде зависимости, существующей между компонентами исследуемого многомерного признака (данные гипотезы позволяют определить характер зависимости между свойствами исследуемого признака);

➢ о независимости и стационарности ряда наблюдений.

Главная задача статистических гипотез – выбор правильного решения из двух альтернативных. *Основную* (выдвинутую) гипотезу называют *нулевой* гипотезой (H_0). Она называется нулевой, потому что выполняется равенство: $x_1 - x_2 = 0$, где x_i – сопоставляемые значения признаков. Обычно нулевые гипотезы учитывают, что различие между сравниваемыми величинами (выборками) отсутствуют, а наблюдаемые отклонения объясняются лишь случайными колебаниями выборки.

Альтернативной (H_1) называется гипотеза, конкурирующая с нулевой гипотезой в том смысле, что если нулевая гипотеза отвергается, то принимается альтернативная. Выбор альтернативной гипотезы зависит от поставленной задачи. Например, H_0: $\sigma = \sigma_0$, тогда H_1: $\sigma \neq \sigma_0$ или $\sigma < \sigma_0$, или $\sigma > \sigma_0$. Такая гипотеза доказывает предположение исследования, поэтому ее называют экспериментальной. Когда нужно убедиться, что выборки не различаются между собой по каким-либо характеристикам, то это означает подтверждение нулевой гипотезы. Чаще требуется доказать значимость различий, ибо они более информативны. Если в процессе исследований было замечено, что в одной из групп экспериментальных данных значения выше, а в другой ниже, то для проверки значимости этих различий лучше сформулировать *направленные гипотезы*. Например, H_0: x_1 не превышает x_2; H_1: x_1 превышает x_2. Если нужно доказать, что различаются формы распределения признака по группам, то формулируются ненаправленные гипотезы: H_0: x_1 не отличается от x_2; H_1: x_1 отличается от x_2.

Правила принятия гипотез

Непосредственно определить истинность предположения нельзя, поэтому проверка статистических гипотез заключается в установлении согласования данных наблюдения с выдвинутой гипотезой. Процесс обоснованного сравнения сформулированной гипотезы с полученными в результате выборки данными называют *статистической проверкой гипотез*.

Если данные наблюдения противоречат высказанной гипотезе, то говорят, что результат сопоставления высказанной гипотезы с выборочными данными отрицателен. В этом случае гипотезу следует *отклонить*. Проверка статистической гипотезы не дает логического доказательства ее верности или неверности и в чем-то похожа на обоснование математических утверждений. Чтобы опровергнуть утверждение, достаточно привести один контрпример, однако для доказательства справедливости недостаточно любого числа примеров. Так и в статистике говорят: «нулевая гипотеза не отвергается, так как полученные данные ей не противоречат». Кроме того, гипотеза может быть отвергнута на основании других выборочных данных или по другим причинам, поэтому принятие нулевой гипотезы нельзя расценивать как точно установленный, содержащийся в ней факт, а только как достаточно правдоподобное, не противоречащее эксперименту утверждение.

Для проверки нулевой гипотезы используют специальным образом подобранную случайную величину, точное или приближенное распределение которой известно, и называют ее статистическим критерием (критерием). В зависимости от вида распределения критерий обозначают по-разному:

- в случае нормального распределения – Z;
- в случае распределения Фишера – F;
- в случае распределения Стьюдента – T;
- если рассматривается закон Пирсона (хи-квадрат) – χ^2.

В общем случае критерий обозначают K. Например, если проверяют гипотезу о равенстве дисперсий двух нормальных генеральных совокупностей, то в качестве критерия K принимают отношение исправленных выборочных дисперсий:

$$K = \frac{S_1^2}{S_2^2}.$$

Под *статистическим критерием* также понимают правило, показывающее, когда статистическую гипотезу на основании опытных значений $x_1, x_2, \ldots, x_n$ величины следует принять, а когда – отвергнуть.

Критерии, применяемые в гипотезах, не требующих знаний о виде функции распределения, аналогично гипотезам называются *непараметрическими*. Непараметрические критерии основаны на оперировании частотами или рангами. Критерии, использующиеся для проверки гипотез о параметрах распределения генеральной совокупности, называются *параметрическими*. Параметрические критерии включают в формулу расчета параметры распределения, то есть средние значения и дисперсии. Каждая группа критериев имеет свои преимущества и недостатки. Параметрические критерии, как правило, оказываются более мощными в случаях измерения признаков по интервальной шкале с нормальным распределением. Непараметрические критерии не требуют сложных расчетов, но они ограничены в том, что с их помощью невозможно оценить взаимодействие двух или более условий или факторов, влияющих на изменение признака.

При проверке гипотезы, как уже отмечено, выборочные данные могут либо согласовываться с основной гипотезой, либо противоречить ей, а значит, всегда есть риск принять неверное решение. Причины расхождения результатов выборки и теоретических характеристик различны. Это и неудачный способ группировки данных, и недостаточный объем выборки, и неправильный выбор распределения и др.

Правило отклонения H_0 и *принятия* H_1. Если эмпирическое значение критерия равняется критическому значению, соответствующему $p < 0.05$, или превышает его, то гипотеза H_0 отклоняется, но мы еще не можем определенно принять гипотезу H_1. Если эмпирическое значение критерия равняется критическому значению, соответствующему $p < 0.01$ или превышает его, то гипотеза H_0 отклоняется и принимается гипотеза H_1.

Ось значимости строится для наглядного представления процесса принятия решения (рис. 2.1).

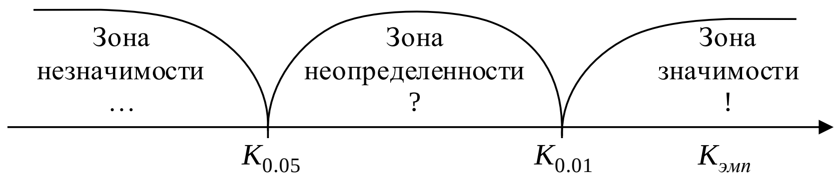


Рис. 2.1. Ось значимости

Критические значения критерия на рисунке обозначены, как $K_{0.05}$ и $K_{0.01}$, а эмпирическое значение критерия как $K_{эмп}$. Вправо от критического значения $K_{0.01}$ располагается зона значимости – в нее попадают все безусловно значимые эмпирические значения. Влево от критического значения $K_{0.05}$ располагается зона незначимости – в нее попадают все безусловно незначимые эмпирические значения критерия. При попадании эмпирического значения в область между $K_{0.05}$ и $K_{0.01}$, то есть зону неопределенности можно отклонить гипотезу о недостоверности различий H_0, но еще нельзя принять гипотезу об их достоверности H_1.

Порядок проверки гипотез состоит из нескольких шагов: сначала по данным выборки определяют частные значения входящих в критерий величин, а затем рассчитывают сам критерий. Значение критерия, вычисленное по выборке, называются *наблюдаемым* (эмпирическим) значением – $K_{набл}$.

Для каждого критерия имеются соответствующие таблицы, по которым, используя данные выборки, находят $K_{крит}$. Для большинства критериев при $K_{набл} \geq K_{крит}$ нулевую гипотезу отвергают, при $K_{набл} < K_{крит}$ нет оснований, чтобы отвергнуть нулевую гипотезу (в этом случае считают, что данные наблюдений согласуются с нулевой гипотезой).

Множество значений критерия, при которых гипотеза H_0 отклоняется и принимается гипотеза H_1, называется *критической областью*. Границы критической области называют *критическими точками* ($K_{крит}$). Критическая область, представляющая собой промежуток $(k_{кр}^{п};\ \infty)$, называется *правосторонней* ($k_{кр}^{п}$ определяется из условия $P(K > k_{кр}^{п}) = \alpha$). Критическая область, представляющая собой промежуток $(-\infty;\ k_{кр}^{л})$, называется *левосторонней* ($k_{кр}^{л}$ определяется из условия $P(K < k_{кр}^{л}) = \alpha$). Критическая область, представляющая собой два промежутка $(-\infty;\ k_{кр}^{л})$ и $(k_{кр}^{п};\ \infty)$, называется *двусторонней* $\Big(k_{кр}^{л}$ и $k_{кр}^{п}$ определяются из условий: $P(K < k_{кр}^{л}) = \dfrac{\alpha}{2}$ и $P(K > k_{кр}^{п}) = \dfrac{\alpha}{2}\Big)$.

Основной *принцип проверки статистических гипотез*: гипотеза, попадающая в критическую область, отвергается, а альтернативная гипотеза принимается как одна из возможных. В практической деятельности по обработке эмпирических данных существует возможность совершить массу различных ошибок. Не является исключением и проверка статистических гипотез, которые принимаются на основе выборочных данных. Ошибки могли возникнуть уже в процессе сбора и предварительной обработки данных. С точки зрения статистической проверки гипотез существует два вида ошибок, называемых *ошибкой I рода* и *ошибкой II рода*.

Ошибкой I рода называется ошибка отклонения верной нулевой гипотезы (H_0), которая на самом деле верна. *Ошибкой II рода* называется ошибка принятия ложной гипотезы (H_0). Вероятность $\alpha = P_{H_0}(H_1)$ совершения ошибки *I* рода называется *уровнем значимости* статистического критерия. Выбор величины уровня значимости α зависит от размера потерь в случае ошибочного решения. Иными словами, уровень статистической значимости – это вероятность признать различия существенными (приняли альтернативную гипотезу и отклонили нулевую), а они в действительности случайные. Например, если указывается, что различия достоверны на 5%-м уровне значимости, то подразумевается вероятность 0.05 того, что они все же недостоверные. Исторически сложилось так, что в инженерных исследованиях используют стандартные значения уровня значимости: α = 0.1; 0.05; 0.01. Наиболее распространенной является величина уровня значимости α = 0.05, которая означает, что высказанная гипотеза будет ошибочно отклонена в среднем в пяти случаях из ста. Если альтернативная гипотеза H_1 единственная, то можно вычислить вероятность ошибки *II* рода – $\beta = P_{H_1}(H_0)$.

Вероятность несовершения ошибки *II* рода, иначе вероятность отклонения неверной гипотезы, называется *мощностью статистического критерия*. Мощность критерия – это его способность выявлять различия, если они есть, то есть отклонить нулевую гипотезу об отсутствии различий, если она неверна. Одни и те же задачи могут быть решены с помощью различных критериев, при этом эмпирически устанавливается, что одни из них позволяют выявить различия там, где другие оказываются неспособными это сделать, или выявляют более высокий уровень значимости различий. Возникает вопрос: зачем использовать менее мощные критерии? Основанием для выбора критерия является не только его мощность, но и другие характеристики: простота расчетов; более широкий диапазон использования; применимость к выборкам разного объема; большая наглядность и информативность результатов.

Уровнем значимости, а значит, и вероятностью ошибки первого рода, можно управлять. Можно установить любую приемлемую степень риска для неправильного вывода по выборочным данным об ошибочности выдвинутой гипотезы. Заметим, что уровень значимости и мощность критерия связаны между собой, причем их связь нелинейная. Поэтому произвольно изменять уровень значимости не следует, так как неоправданное уменьшение ошибки первого рода может привести к существенной потере мощности критерия.

Пример. В ходе исследования рынка труда по группе строительных компаний изучалась дискриминация по возрасту. Работодатели полагают, что ее доля составляет 3%, соискатели – 20%. Достигнута следующая договоренность в трудовой инспекции: если при проверке 10 случайно отобранных соискателей вакансий будет обнаружено не более одного случая дискриминации, то штрафные санкции на группу компаний не накладываются. Сформулируйте нулевую и альтернативную гипотезы с точки зрения соискателя вакансий. Определите критическую область и область принятия нулевой гипотезы. Сформулируйте, в чем состоят ошибки *I* и *II* рода. Найдите их вероятности.

Решение. С точки зрения соискателя нулевой гипотезой H_0 будем считать гипотезу о 20% случаев дискриминации по возрасту, а альтернативной гипотезой H_1 версию работодателей о 3% таких.

Поскольку отбирается 10 человек, а затем фиксируется число случаев дискриминации, то множеством всевозможных результатов испытаний будет: $A=\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ и случаев дискриминации может оказаться 0, 1, …, 10. По условиям задачи гипотеза соискателя отвергается, если $K \le 1$, следовательно, критическая область – $A_{крит}=\{0, 1\}$, а область принятия гипотезы – $\{2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

Ошибка *I* рода: штрафных санкций не будет, даже, если будет выявлено 20% случаев дискриминации по возрасту.

Ошибка *II* рода: штрафных санкций не будет, даже, если будет выявлено 3% случаев дискриминации по возрасту.

Найдем вероятности этих ошибок. Если нулевая гипотеза верна, то вероятность того, что один случайно выбранный человек попал под случай дискриминации по возрасту, составляет 0.2. Ошибка *I* рода произойдет, если из 10 человек попал под случай дискриминации по возрасту не более чем один (0 или 1). В случае истинности нулевой гипотезы H_0 число тех кто попал под дискриминацию по возрасту K является биноминальной случайной величиной $B_i(10; 0.2)$, поэтому $P(K=0)=0.8^{10}\approx 0.107$; $P(K=1)=10\cdot 0.8^9\cdot 0.2\approx 0.268$. Аналогично вычислим вероятность α ошибки *I* рода:

$$\alpha = P(K\le 1| H_0)=P(K=0| H_0)+P(K=1| H_0)=0.107+0.268=0.375.$$

Если верна альтернативная гипотеза H_1, то вероятность того, что один случайно выбранный человек попал под дискриминацию по возрасту составляет 0.03. Ошибка *II* рода произойдет, если из 10 человек под случай дискриминации по возрасту попадут два или более.

В случае истинности альтернативной гипотезы K также является биноминальной случайной величиной, но с другим параметром – $B_i(10; 0.03)$, поэтому $P(K=0)=0.97^{10}\approx 0.737$; $P(K=1)=10\cdot 0.97^9\cdot 0.03\approx 0.228$.

Следовательно, для вероятности β ошибки II рода имеем:

$$\beta = P(K>1| H_a)=1-P(K\le 1| H_a)=1-P(K=0| H_a)-P(K=1| H_a)=0.035.$$

Замечание. Из сравнения вероятностей α и β следует, что оговоренная процедура проверки выгодна скорее работодателю.

Задачи, сводящиеся к оценке истинности нулевой гипотезы H_0 по отношению к конкурирующей гипотезе H_1, можно решить, используя различные статистические критерии.

Несмотря на разнообразие, как самих гипотез, так и применяемых статистических критериев, существует некий алгоритм:

1. На основе выборочных данных $x_1, x_2, \ldots, x_n$ с учетом конкретных условий задачи выдвигают нулевую гипотезу (H_0) и конкурирующую с ней альтернативную гипотезу H_1.

2. Выбирают уровень значимости α.

3. Выбирают (если он не задан) объем выборки n и число степеней свободы r.

4. На основе выборочных данных выбирают критерий (статистику) – некоторую функцию K от результатов наблюдений и условий рассматриваемой статистической задачи, подчиненную некоторому закону распределения.

5. По таблицам, в зависимости от объема выборки и уровня значимости α критических точек ($K_{крит}$), определяют критическую область *W*. В этом случае возможно совершение ошибки *I* рода.

6. Определяют на основе выборочных данных $x_1, x_2, \ldots, x_n$ численную величину критерия $K_{набл}$ по формулам, учитывающим характер проверяемой гипотезы.

7. Вырабатывают решение: если $K_{набл}$ попадает в критическую область, то нулевая гипотеза отклоняется и принимается альтернативная.

Проанализировать выбранное решение можно, используя таблицу:

Выбранное решение	Реальная ситуация	
	H_0 – истинная; H_a – ложная	H_0 – ложная; H_a – истинная
Выбрали H_0 и отвергли H_a	$P_{H_0}(H_0) = 1 - \alpha$ Правильное решение	$P_{H_a}(H_0) = \beta$ Ошибка *II* рода
Выбрали H_a и отвергли H_0	$P_{H_0}(H_a) = \alpha$ Ошибка *I* рода	$P_{H_a}(H_a) = 1 - \beta$ Правильное решение

Для большей уверенности перед окончательным принятием гипотезы, желательно подвергнуть ее проверке с помощью других критериев или повторить эксперимент, увеличив объем выборки.

Пример. Компании необходимо принять решение о вхождении на рынок со своим товаром. В течение 8 месяцев при существующем маркетинге на одном из объектов остатки товара составили 10%, что принято за норму рентабельности. Для проверки была произведена случайная выборка 500 объектов компании, 60 из которых оказались малорентабельными. Согласуются ли данные выборки с утверждением «доля рентабельности (генеральной совокупности) соответствует установленному нормативу»? Уровень значимости при проверке гипотезы принять $a = 0.05$.

Решение. В данной задаче нужно проверить гипотезу о том, что доля признака p равна определенной величине δ. Нулевая гипотеза H_0: $p = \delta = 0.1$, альтернативная гипотеза H_1: $p > 0.1$. С целью проверки нулевой гипотезы используем критерий $K = \frac{m}{n}$ – выборочную долю признака в выборке. Данный критерий имеет биноминальное распределение, но при большом n и не очень маленьком δ ($n > 50,\ n\delta \ge 10$) это распределение приближается к нормальному с центром в точке δ и средним квадратичным отклонением $\sigma\left(\frac{m}{n}\right) = \sqrt{\frac{\delta(1-\delta)}{n}}$.

Следовательно, величина $z = \frac{\frac{m}{n} - \delta}{\sqrt{\frac{\delta(1-\delta)}{n}}}$ распределена по стандартному нормальному закону. Критическую точку определяем из условия:

$$P\left\{\frac{\frac{m}{n} - \delta}{\sqrt{\frac{\delta(1-\delta)}{n}}} \le z\right\} = 0.5 + \Phi(z) = 1 - \alpha .$$

Отсюда $\Phi(z) = 1 - \alpha - 0.5 = 0.45$.

По таблице функций Лапласа (см. прил.) находим $z = 1.65$ и определяем критическую точку

$$K_{крит} = \delta + z\sqrt{\frac{\delta(1-\delta)}{n}} = 0.1 + 1.65 \cdot \sqrt{\frac{0.1(1-0.1)}{500}} = 0.1 + 0.022 = 0.122 .$$

Выборочная доля $K_{набл} = \frac{m}{n} = \frac{60}{500} = 0.12$. Имеем: $K_{набл} < K_{крит}$. Последнее означает, что нулевая гипотеза H_0 не отвергается. Доля нерентабельности объектов в выборочной совокупности равная 12% превысила норматив, однако такое отклонение находится в допустимых пределах и может объясняться случайностью отбора.

Установление теоретического закона распределения случайной величины, характеризующей изучаемый признак по эмпирическому распределению, – основная задача математической статистики. Как бы хорошо ни подобрали теоретический закон распределения, расхождения между эмпирическим и теоретическим распределениями неизбежны. Случайно ли расхождение? Можно ли объяснить его малым числом наблюдений, способом группировки или другими причинами? Возможно, расхождение не случайно и объясняется ложной гипотезой о виде распределения. Ответить на эти вопросы и провести проверку согласия эмпирической функции распределения с предположением относительно теоретической функции распределения $F(x)$ позволяют критерии согласия, в которых гипотеза либо полностью, либо с точностью до небольшого числа параметров определяет закон распределения.

Замечание. Критерии согласия не доказывают справедливость гипотезы, а лишь устанавливают на принятом уровне ее согласие или несогласие с данными наблюдений.

Потребности исследовательской и инженерной практики требуют применения методов количественного описания процессов, позволяющих точно регистрировать не только количественные, но и качественные факторы. С этой целью разработаны несколько групп критериев различной мощности, о которых подробно будет рассказано далее.

При проверке статистических гипотез принципиально существуют четыре возможности:

- гипотеза верна, и она принимается;
- гипотеза верна, но она отвергается (ошибка первого рода);
- гипотеза неверна, и она отвергается;
- гипотеза неверна, но она принимается (ошибка второго рода).

Ошибки первого и второго рода существенно различаются между собой по значимости, и это оказывает большое влияние на всю процедуру проверки статистических гипотез. Необходимо еще раз подчеркнуть, что никакая гипотеза не может быть окончательно принята или отвергнута. Поэтому используемые, довольно категорические, утверждения «принять» и «отвергнуть» являются просто условным сокращением выражений вида «опытные данные не противоречат выдвинутой гипотезе» и «опытные данные противоречат выдвинутой гипотезе».

2.3. Общие положения применения статистических критериев. Критерий Пирсона

Статистический критерий – это решающее правило, которое обеспечивает математически обоснованное принятие истинной и отклонение ложной гипотезы Статистические критерии определяют в практической деятельности метод расчета определенного числа, которое обозначается как эмпирическое значение критерия, например, числа χ^2 для критерия Пирсона.

Соотношение эмпирического и критического значений критерия является основанием для подтверждения или опровержения гипотезы. Согласно назначению статистические гипотезы и статистические критерии делятся на параметрические и непараметрические. Выбор критериев достаточно широк, в чем можно убедиться, ознакомившись с приведенными в списке литературы источниками. Однако, нашей целью является описание статистических критериев, адекватных типовым для инженерных исследований задачам анализа данных.

Применение критериев позволяет устанавливать различия по уровню исследуемого признака между двумя, тремя и более выборками испытуемых. Например, определение различий между технологиями производства строительных работ, финансовыми затратами государственных предприятий и частных фирм и т. д. В результате таких исследований формируется статистически достоверный групповой профиль или усредненный портрет человека той или иной профессии, статуса, например, успешный менеджер, успешная технология. Критерии различий предполагают, что сопоставляются независимые выборки, состоящие из разных испытуемых. Решение о выборе того или иного критерия принимается на основании количества и объема сопоставляемых выборок.

Параметрические критерии используются в задачах проверки параметрических гипотез и включают в свой расчет показатели распределения, например, средние, дисперсии и т. д. Это такие известные классические критерии, как χ^2-критерий Пирсона, t-критерий Стьюдента, F-критерий Фишера.

Параметрические критерии позволяют прямо оценить уровень основных параметров генеральных совокупностей, разности средних и различия в дисперсиях Они способны выявить тенденции изменения признака при переходе от условия к условию, оценить взаимодействие двух и более факторов в воздействии на изменения признака. Параметрические критерии считаются несколько более мощными, чем непараметрические, при условии, что признак измерен в интервальной шкале и его распределение близко к нормальному. Однако в интервальной шкале могут возникнуть определенные проблемы, в частности, если данные, представлены не в стандартизированных оценках. К тому же проверка распределения «на нормальность» требует достаточно сложных расчетов.

Рассмотрим особенности и условия использования *критерия Пирсона*.

Пример. В случайном порядке отобрано 60 студентов. Их оценки по итогам экзамена по курсу высшей математики, представлены выборкой:

4, 3, 5, 3, 4, 4, 3, 3, 4, 2, 3, 3, 3, 5, 3, 2, 3, 4, 5, 5, 4, 5, 4, 4, 3, 4, 2, 4, 3, 4, 4, 4, 3, 3, 4, 4, 3, 2, 4, 4, 2, 3, 3, 2, 5, 3, 2, 5, 3, 4, 3, 2, 3, 3, 5, 3, 4, 3, 3, 3.

Проверь на уровне статистической значимости 0.05 гипотезу о нормальном законе распределения студентов по уровню знаний, продемонстрированных на экзамене по высшей математике.

Решение.

H_0: распределение студентов по уровню знаний продемонстрированных на экзамене по высшей математике (по оценкам) подчинено нормальному закону распределения.

H_1: распределение оценок значимо отличается от нормального распределения.

Чтобы получить предварительное представление о распределении студентов по уровню знаний по высшей математике (по оценкам), построим вариационный ряд. Изучаемый признак (оценка) принимает ограниченное число целых значений, поэтому удобно построить дискретный ряд и подсчитать число студентов, имеющих одинаковую оценку. Сгруппированные данные представим в виде вариационного ряда:

Оценка (x_i)	2	3	4	5
Число студентов (n_i)	8	25	19	8

Проверим длину выборки: $n = \sum_{i=1}^{4} n_i = 8 + 25 + 19 + 8 = 60$.

Найдем параметры распределения: выборочную среднюю (средний балл) $\overline{x}_в$, выборочную дисперсию $D_в$ и среднее квадратичное отклонение выборки S:

$$\overline{x}_в = \frac{2 \cdot 8 + 3 \cdot 25 + 4 \cdot 19 + 5 \cdot 8}{60} = \frac{207}{60} = 3.45;$$

$$\overline{x_в^2} = \frac{2^2 \cdot 8 + 3^2 \cdot 25 + 4^2 \cdot 19 + 5^2 \cdot 8}{60} = \frac{761}{60} \approx 12.68;$$

$$D_в = \overline{x_в^2} - (\bar{x}_в)^2 = 12.68 - 3.45^2 \approx 0.78;$$

$$s = \sqrt{D_в} = \sqrt{0.78} \approx 0.88.$$

Для проверки нулевой гипотезы рассчитаем теоретические частоты нормального распределения с параметрами α и σ. В качестве оценок α и σ возьмем оценки наибольшего правдоподобия, положив $\alpha = \bar{x}_в = 3.45$ и $\sigma \approx s \approx 0.88$.

Расчеты сведем в таблицу:

№	x_i	n_i	$x_i - a$	$z_i = \frac{x_i - a}{\sigma}$	$\varphi(z_i)$	$P_i = \Delta z_i \cdot \varphi(z_i)$	P_i исправленное	$n \cdot P_i$
1	2	8	– 1.45	– 1.65	0.1023	0.117	0.123	7.4
2	3	25	0	0.51	0.3503	0.400	0.400	24.0
1	4	19	0.55	0.625	0.3281	0.374	0.374	22.4
4	5	8	1.55	1.76	0.0848	0.097	0.103	6.2
Σ		60				0.988	1.000	60.0

Замечание: z_i – нормированное значение отклонений x_i от математического ожидания α; $\varphi(z_i)$ – значения локальной функции Лапласа (см. Приложение) $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$; $n \cdot P_i$ – теоретические частоты:

$$\Delta z_i = z_2 - z_1 = z_3 - z_2 = z_4 - z_3 = -0.51 + 1.65 = 1.14.$$

Расчетная сумма вероятностей: $\sum_{i=1}^{4} P_i = 0.988$ отлична от единицы. Чтобы сумма вероятностей равнялась единице, к крайним значениям z_i, т. е. к z_1 и z_4, прибавим $\frac{1 - 0.988}{2} = 0.006$. В предпоследнем столбце таблицы задействованы «исправленные» вероятности P_i.

Если нулевая гипотеза H_0 справедлива, то частоты выборки n_i должны незначительно отличаться от теоретических частот $n \cdot P_i$. Для оценки расхождения эмпирических и теоретических частот используем критерий Пирсона:

$$\sum_{i=1}^{4} \frac{(n_i - nP_i)^2}{nP_i} = \chi^2_{набл.}.$$

Расчеты сведем в таблицу:

№	n_i	$n \cdot P_i$	$n_i - nP_i$	$(n_i - nP_i)^2$	$\dfrac{(m_i - nP_i)^2}{nP_i}$
1	8	7.4	0.6	0.36	0.05
2	25	24.0	1.0	1.00	0.04
1	19	22.4	– 3.4	11.56	0.52
4	8	6.2	1.8	3.24	0.52
$\sum$	60	60.0			$\chi^2_{набл.}=1.13$

Наблюдаемое значение критерия $\chi^2_{набл}=1.13$. Определим число степеней свободы: случайная величина X принимает всего четыре значения ($k=4$), неизвестных параметров – два ($m=2$), поэтому число степеней свободы $\nu=4-3=1$. Из таблиц приложения при уровне значимости $\alpha=0.05$ и числе степеней свободы $\nu=1$ находим: $\chi^2_{крит}=3.84$.

Таким образом, $\chi^2_{набл}=1.13<\chi^2_{крит}=3.84$.

Сформулируем вывод: $\chi^2_{набл.}$ попадает в незначимую область критерия Пирсона. Нет оснований для отказа от гипотезы H_0, т. е. с вероятностью 0.95 можно утверждать, что распределение студентов по уровню знаний продемонстрированных на экзамене по математике согласуется с нормальным распределением.

При работе с критерием Пирсона (χ^2) есть важное ограничение. Если в таблице присутствуют малые теоретические частоты, то статистика χ^2 не будет иметь распределение Пирсона, и соответственно нельзя качественно проверить гипотезу. Однозначного ответа на вопрос, какие частоты считать малыми, нет. Обычно в качестве границы рассматривается число «5», иначе, если есть теоретические частоты, меньшие 5, критерий χ^2 применять нельзя. Эту проблему удается решить за счет объединения категорий переменных.

Пример. Пусть есть переменная «удовлетворенность работой» с возможными значениями «работа совсем не нравится», «работа скорее не нравится, чем нравится», «работа скорее нравится, чем нет», «работа очень нравится», и переменная категория сотрудника» с возможными значениями «руководитель» и «не руководитель». Зависит ли степень удовлетворенности работой от того, является ли человек начальником?

Решение. Построим таблицу сопряженности, чтобы посмотреть, зависит ли степень удовлетворенности работой от того, является ли человек начальником:

Категория сотрудников	*Работа совсем не нравится*	*Работа скорее не нравится, чем нравится*	*Работа скорее нравится, чем нет*	*Работа очень нравится*	*Сумма*
Руководитель	2	4	8	3	17
Не руководитель	3	4	9	2	18
Сумма	5	8	17	5	35

В таблице сопряженности встречаются наблюдаемые частоты, меньшие 5, но, по теории, с числом «5» надо сравнивать теоретические частоты. Найдем значение теоретической частоты для первой ячейки. Получим:

$$\mu_{11} = \frac{17 \cdot 5}{35} = 2.4 .$$

Аналогично найдем остальные теоретические частоты и запишем в таблицу:

Категория сотрудников	*Работа совсем не нравится*	*Работа скорее не нравится, чем нравится*	*Работа скорее нравится, чем нет*	*Работа очень нравится*
Руководитель	2.4	3.9	8.3	2.4
Не руководитель	2.6	4.1	8.7	2.6

Из таблицы видно, что присутствуют теоретические частоты, меньшие 5. Объединим категории переменной «удовлетворенность работой»: категории «работа совсем не нравится» и «работа скорее не нравится, чем нравится» в одну и назовем ее «негативное отношение к работе»; аналогично объединим категории «работа скорее нравится, чем нет» и «работа очень нравится» в одну и назовем ее «позитивное отношение к работе». При укрупнении разрядов часть информации теряется, но это, к сожалению, неизбежно.

В рассматриваемом примере наблюдаемые частоты для соответствующих категорий объединяются и записываются в таблицу:

Категория сотрудников	*Негативное отношение к работе*	*Позитивное отношение к работе*
Руководитель	2 + 4 = 6	8 + 3 = 11
Не руководитель	3 + 4 = 7	9 + 2 = 11

В итоге получаем таблицу сопряженности переменных после объединения категорий:

Категория сотрудников	*Негативное отношение к работе*	*Позитивное отношение к работе*	*Сумма*
Руководитель	6	11	17
Не руководитель	7	11	18
Сумма	13	22	35

Найдем теоретические частоты для «исправленных» данных:

Категория сотрудника	*Негативное отношение к работе*	*Позитивное отношение к работе*
Руководитель	6.3	10.7
Не руководитель	6.7	11.3

Таким образом, в полученной таблице нет теоретических частот меньше 5, значит, можно использовать критерий χ^2.

Рассмотрим схему использования критерия χ^2 для проверки гипотезы о независимости переменных.

<u>Пример</u>. В качестве исходных данных возьмем результаты из таблицы сопряженности переменных после объединения категорий. Гипотеза формулируется следующим образом: переменные «категория сотрудника» и «удовлетворенность работой» – независимы.

Решение. Вычисление теоретических частот уже проведено. Подставляем теоретические и эмпирические частоты в формулу и получим:

$$\chi^2 = \frac{(6.3-6)^2}{6.3} + \frac{(10.7-11)^2}{10.7} + \frac{(6.7-7)^2}{6.7} + \frac{(11.3-11)^2}{11.3} = 0.043.$$

Таким образом, $\chi^2_{выб} = 0.043$. Находим число степеней свободы. В нашем случае $\nu = 1$. В качестве уровня статистической значимости возьмем классическое значение 0.05. Соответствующее табличное значение $\chi^2_{табл} = 3.84$.

Сравним значения $\chi^2_{выб}$ и $\chi^2_{табл}$. Получим $\chi^2_{выб} < \chi^2_{табл}$, следовательно, гипотеза принимается. Таким образом, можно сделать вывод, что рассматриваемые переменные независимы.

С помощью критерия согласия χ^2 можно проверить гипотезу о согласии данных выборки с любым теоретическим законом распределения для любой случайной величины, как непрерывной, так и дискретной.

<u>Пример.</u> Вокруг строительных объектов недалеко друг от друга расположены три АБЗ. Руководители региона решили выяснить, при одинаковой стоимости продукции заводов, потребители закупают

асфальтобетонные смеси одинаково ли на каждом из заводов или имеется различие. Для проверки собрали информацию о количестве закупок в течение недели. На первом заводе это число составило 160 т., во втором – 225, в третьем – 215 т.

Решение. Нулевая гипотеза H_0: вероятности закупок p_i, где $(i = 1, 2, 3)$ у всех АБЗ равны (иначе, распределены равномерно).

В результате испытаний получили: $m_1 = 160$, $m_2 = 225$, $m_3 = 215$, $n = 160 + 225 + 215 = 600$, $np_i = \frac{1}{3} \cdot 600 = 200$.

Вычислим критерий $\chi^2 = 12.25$. $\chi^2_{крит}$ найдем по таблице критических значений при $\nu = 3 - 1 = 2$. При уровне значимости $\alpha = 0.01$ найдем $\chi^2_{крит} = 9.21$. Значит, $\chi^2_{набл.} > \chi^2_{крит}$.

Гипотеза о равномерности не подтвердилась. Разницу в посещаемости АБЗ в течение недели нельзя объяснить случайными колебаниями, это связано с более или менее выгодным расположением баз либо с разным качеством продукции.

Рассмотрим важный элемент применения статистического критерия – *определение числа интервалов* гистограммы, на которые разбивается весь диапазон статистического ряда. Один из приемов основан на расчете оптимальной длины интервала с последующим определением количества интервалов.

Оптимальная длина интервала рассчитывается по зависимости:

$$\Delta x = \frac{x_{max} - x_{min}}{1 + 3.21 \lg n},$$

где x_{max} и x_{min} – максимальное и минимальное значения статистического ряда (границы диапазона); n – объем выборки.

Число интервалов равно:

$$K = \frac{x_{max} - x_{min}}{\Delta x}.$$

Если K число дробное, то за число интервалов принимается ближайшее целое число, не меньше K.

Другая методика предполагает определение числа интервалов по формуле:

$$K \geq \log_2 n + 1.$$

Эта формула, позволяющая получить нижнюю границу (наименьшее значение) K, наиболее точна при больших значениях *n*. Например, при $n = 100$ $K \geq 6$, а при $n = 1000$ $K \geq 9$.

<u>Пример</u>. На заводе строительных конструкций принято решение на изготовление новой балки перекрытия, работающей в условиях динамического нагружения. Для проверки ее работоспособности изготовлено 78 образцов, которые проверялись путем динамического нагружения до появления усталостных трещин. Результаты статистических испытаний приведены в таблице.

Количество нагружений до появления трещин (x_i)	30	40	50	60	70	80	90	100	110	120
Число испытаний, (n_i)	2	6	10	12	18	10	8	6	4	2

Данные таблицы показывают, что в двух случаях трещины появились после 30 нагружений, в шести случаях – после 40 нагружений и т. п. Для удобства расчетов и отображения данных показателей x_i разбит на интервалы через 10 испытаний. Всего интервалов десять. Проверим, достаточно ли такого количества интервалов.

$$\Delta x = \frac{120-30}{1+3.21\lg 78} = \frac{90}{7.074} = 12.72; \qquad K \geq \frac{120-30}{12.72} \geq 7.07.$$

По предварительной оценке для $n=100$ величина $K \geq 6$, следовательно, число интервалов в таблице $K=10$ вполне достаточно.

Подведем итог применения метода статистической проверки гипотез по критерию Пирсона (χ^2). Он реализуется посредством выполнения следующих этапов.

1 этап. Построение статистического ряда (гистограммы). Для этого имеющаяся статистическая выборка разбивается на K интервалов по ранее изложенной методике. В рассмотренном выше примере принято $K=100$; Гистограмма показана на рисунке 2.2.

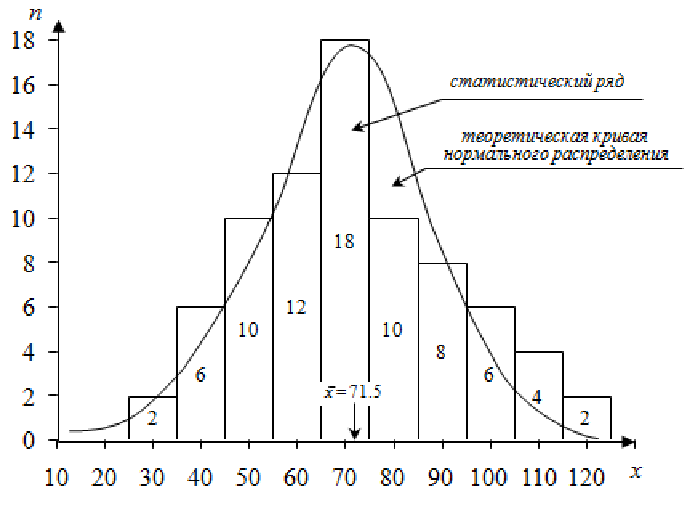


Рис 2.2. Гистограмма и кривая нормального распределения вероятности

2 этап. По виду гистограммы подбирается наиболее близкий закон распределения случайной величины x (в примере в качестве такого принят нормальный закон) и рассматривается эмпирическая вероятность попадания в каждый *j-интервал* гистограммы принятого распределения по формуле:

$f_э = \frac{n_j}{n}$, где n – объем выборки; n_j – количество измерений в j-интервале (например, в первом интервале таблицы $n_j = 2$; $f_{эj} = \frac{2}{78} = 0.026$; во втором $f_{эj} = \frac{6}{78} = 0.077$ и т. п.).

Значения «эмпирических» вероятностей $f_э$ для каждого из K интервалов, приведены в таблице.

Исходные данные для проверки гипотезы о нормальном законе распределения случайной величины x_i по критерию χ^2.

Значения x_i	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*	*110*	*120*
$f_{эj} = \frac{n_j}{n}$	0.026	0.077	0.128	0.154	0.231	0.128	0.102	0.077	0.051	0.026
f_{mj}	0.069	0.094	0.118	0.136	0.128	0.124	0.116	0.070	0.044	0.022

3 этап. Для принятой гипотезы нормального закона распределения и математическим ожиданием $M = 71.5$, дисперсией $D = 407$ и среднеквадратическим отклонением $\sigma = 20.2$ (рассчитанным по экспериментальным данным) определим вероятность попадания x_i в каждый из k интервалов. Для этого воспользуемся известной формулой вероятности попадания случайной величины на участок, ограниченный значениями α и β:

$$P(\alpha < x < \beta) = \frac{1}{2}\left[\Phi\left(\frac{\beta - M}{\sqrt{2}\sigma}\right) - \Phi\left(\frac{\alpha - M}{\sqrt{2}\sigma}\right)\right],$$

где $\Phi(x)$ – интеграл вероятности (функция Лапласа), значения которой приведены в таблице приложений.

Примечание: Функция $\Phi(x)$ – нечетная и в таблице приведены значения $\Phi(x)\cdot 10^4$. Вероятность попасть в интервал от 30 до 40 составляет:

$$f_{m1} = P(30 < x < 40) = \frac{1}{2}\left[\Phi\left(\frac{40 - 71.5}{1.41\cdot 20.2}\right) - \Phi\left(\frac{30 - 71.5}{1.41\cdot 20.2}\right)\right] = \frac{1}{2}\left[-\Phi(1.11) + \Phi(1.46)\right] =$$

$$= \frac{1}{2}(-0.7287 + 0.8664) = 0.0688.$$

Аналогично,

$$f_{m2} = P(40 < x < 50) = \frac{1}{2}\left[\Phi\left(\frac{50-71{,}5}{1.41\cdot 20.2}\right) - \Phi\left(\frac{40-71{,}5}{1.41\cdot 20.2}\right)\right] = 0.0944.$$

Затем, $f_{m3} = 0.118$ и т. д.

4 этап. Имея эмпирические и теоретические вероятности $f_{эj}$ и f_{mj} рассчитаем наблюденное значение критерия Пирсона по зависимости:

$$\chi^2_{набл.} = \sum_{j=1}^{k} \frac{\left(f_{эj} - f_{mj}\right)^2}{f_{mj}}.$$

Расчеты наблюденного значения критерия χ^2 удобно выполнить с помощью таблицы.

j	$f_{эj}$	f_{mj}	$f_{эj} - f_{mj}$	$\left(f_{эj} - f_{mj}\right)^2$	$\frac{\left(f_{эj} - f_{mj}\right)^2}{f_{mj}}$
1	0.026	0.069	– 0.043	0.001849	0.026797
2	0.77	0.094	– 0.017	0.000289	0.003074
3	0.128	0.118	0.010	0.000100	0.000847
4	0.154	0.136	0.018	0.000324	0.002382
5	0.231	0.128	0.103	0.010609	0.082883
6	0.128	0.124	0.004	0.000016	0.000129
7	0.102	0.116	– 0.014	0.000196	0.001689
8	0.077	0.07	0.007	0.000049	0.000700
9	0.051	0.044	0.007	0.000049	0.000962
10	0.026	0.022	0.004	0.000016	0.000727

Из данных таблицы получим $\chi^2 = 0.1202$ (как сумму значений последнего столбца).

5 этап. По таблице критических значений распределения χ^2 с учетом заданного уровня значимости α (в инженерных расчетах принимают $\alpha = 0.05$) и числа степеней свободы $\nu = k - 3$ (k – число интервалов выборки, 3 – число связей для нормального закона распределения) находим критическое значение $\chi^2_{крит}$. Если $\chi^2_{набл} < \chi^2_{крит}$ то гипотеза о нормальности распределения статистической выборки принимается.

Поскольку в примере $\chi^2_{набл} = 0.12$, т. е. значительно меньше $\chi^2_{крит} = 14.1$, то можно утверждать, что полученные экспериментальные данные имеют нормальное распределение.

В экспериментальных исследованиях нередки случаи, когда с помощью критерия χ^2, либо других критериев, доказывается, что одна и та же статистическая выборка одинаково удовлетворяет нескольким законам распределения. Из них экспериментатор должен выбрать один. Но при этом возникают другие проблемы: каковы точность и надежность измерений, почему имеется большой разброс данных и т. д. Эти проблемы особенно актуальны при малом количестве измерений.

2.4. Дополнительные критерии проверки свойств выборок

Существуют методы статистической проверки гипотез, основанные на применении критериев Стьюдента (t); Фишера (F) и других. Проверка гипотез о равенстве числовых характеристик генеральных совокупностей – это, по сути, методы проверки гипотез об однородности выборок. Параметрические критерии обладают большей мощностью. По этой причине, в случаях, когда выборки имеют нормальное распределение, нужно отдавать предпочтение именно параметрическим критериям.

Критерий Стьюдента (t) был разработан Уильямом Госсетом для оценки качества пива в компании Гиннесс. В связи с обязательствами перед компанией по неразглашению коммерческой тайны (руководство Гиннесса считало таковой использование статистического аппарата в своей работе), статья Госсета вышла в 1908 году в журнале «Биометрика» под псевдонимом «Student» (Студент), что обусловило название метода.

Закон Стьюдента положил начало созданию теории «малой выборки». При большом объеме выборки особенность распределения в генеральной совокупности не имеет значения, так как распределение отклонений выборочного показателя от генеральной характеристики при большой выборке всегда оказывается нормальным. В выборках небольшого объема ($n < 30$) на распределении ошибок выборки будет сказываться характер распределения генеральной совокупности. Распределение Стьюдента зависит от двух величин: значения t и числа степеней свободы ν. С увеличением n, т. е. числа наблюдений, это распределение быстро приближается к стандартизированному нормальному (с параметрами $a = 0$ и $\sigma = 1$). Уже при $n \geq 30$ распределение Стьюдента мало отличается от стандартизированного нормального распределения. Для практического использования распределения Стьюдента существуют специальные таблицы, в которых содержатся критические значения t для разных уровней значимости α и чисел степеней свободы ν.

С параметрической точки зрения закон распределения Стьюдента (t) применяется для проверки гипотезы об отличии среднего значения $\bar{x}$ от некоторого известного значения m_x по формуле: $t = \frac{|\bar{x} - m_x|}{\sigma / \sqrt{n}}$, количество степеней свободы определяется соотношением: $\nu = n - 1$.

Замечание 1. Критерий Стьюдента можно применять также и тогда, когда сравниваются не средние величины выборок, а их относительные частоты.

Замечание 2. Если предварительная гипотеза о нормальности распределения попарных разностей окажется отвергнутой, то критерий Стьюдента применять не следует. В таких случаях нужно использовать непараметрические критерии.

Важным вопросом, решаемым при статистической обработке данных, является *доказательство достаточности объема выборки* (числа опытов). Для этого можно воспользоваться неравенством Чебышева:

$$P\left[\left|\bar{x} - M\right| \le \varepsilon\right] \le 1 - \frac{D(x)}{N\varepsilon^2},.$$

где: $\bar{x}$ и M, соответственно, статистическое среднее и математическое ожидание изучаемой случайной величины;

$P\left[\left|\bar{x} - M\right| \le \varepsilon\right]$ – вероятность того, что разница $\bar{x} - M$ не превысит некоторого заданного числа ε (ε – называют точностью измерения и принимают равной 5% или 10% от величины x);

N – количество измерений;

$D(x)$ – дисперсия статистической выборки из N измерений.

Для оценки точности измерений применяют показатели среднеарифметическое значение среднеквадратического отклонения – σ_0 и точность измерения – m.

$$\sigma_0 = \frac{\sigma}{\sqrt{n}}; \quad m = \frac{\sigma_0}{\bar{x}}.$$

Показатель m по своей сути близок к коэффициенту вариации $k_в$ характеризующему однородность измерений:

$$k_в = \frac{\sigma}{\bar{x}}.$$

Указанные показатели позволяют приближенно определить минимально необходимое число измерений в эксперименте (N_{min}) по формуле:

$$N_{min} = \frac{k_в^2 t^2}{m^2},$$

где t – коэффициент гарантии.

Пример. Определить достаточность статистической выборки для исследования работоспособности строительных конструкций с надежностью 0.9 (результаты статистических испытаний приведены в примере предыдущего раздела).

Чтобы воспользоваться неравенством Чебышева, определим характеристики выборки $\bar{x}$ и $D(x)$:

$$\overline{X} = \frac{30 \cdot 2 + 40 \cdot 6 + 50 \cdot 10 + \ldots + 120 \cdot 2}{78} = \frac{5580}{78} = 71.5;$$

$$D(x) = \frac{2(30-71.5)^2 + 6(40-71.5)^2 + \ldots + 2(120-71.5)^2}{78} = 407;$$

$$\sigma = \sqrt{D(x)} = 20.2.$$

При $\bar{x} = 71.5$ показатель $\varepsilon = 0.1$, тогда $x = 7.15$.

Подставив полученные значения в неравенство Чебышева, получим:

$$0.9 \le 1 - \frac{407}{N \cdot 7.15^2}.$$

Решим неравенство относительно N, получим его оценку $N \ge 80$. Имеющийся объем выборки $N = 80$ измерений отличается от требуемого на 2 измерения, т. е. менее, чем на 3%, следовательно, ее можно считать достаточной.

Пример. Определить минимально необходимое количество измерений в эксперименте по оценке работоспособности конструкций (по данным вышеприведенного примера).

Статические характеристики проведенной серии из 78 испытаний равны:

$$\bar{x} = 71.5; \quad \sigma = 20.1, \text{ отсюда:}$$

$$\sigma_0 = \frac{20.1}{\sqrt{78}} = 2.27; \quad m = \frac{2.27}{71.5} = 0.03; \quad k_в = \frac{20.1}{71.5} = 0.28.$$

Минимально необходимое число опытов для $t = 1$ будет равно:

$$N_{min} = \frac{0.28^2 \cdot 1^2}{0.03^2} = \frac{0.0784 \cdot 1}{0.0009} = 87.$$

Полученное значение N_{min} незначительно отличается от ранее вычисленного значения по неравенству Чебышева.

Оценки доверительного интервала по рассмотренным методикам возможны при $n > 30$. Однако стоимость, либо сложность проведения эксперимента, не всегда позволяют выполнить такое количество испытаний.

При малых значениях n для нахождения доверительного интервала рекомендуется распределение Стьюдента.

Критерий Фишера (F). Пусть имеются две выборки из нормальных генеральных совокупностей. Для сравнения дисперсий этих выборок Р. Фишер предложил рассматривать разность их натуральных логарифмов. Позже Д. Снедекор заменил разность логарифмов отношением выборочных дисперсий: $F_{эмп} = \dfrac{\sigma_1^2}{\sigma_2^2}$, где σ_1^2 – большая дисперсия; σ_2^2 – меньшая дисперсия рассматриваемых вариационных рядов.

В этом критерии проверяется гипотеза, состоящая в том, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0 = \{D_x = D_y\}.$$

Критическое значение критерия Фишера (F) следует определять по специальной таблице, исходя из уровня значимости α и степеней свободы числителя ($n_1 - 1$) и знаменателя ($n_2 - 1$). Если вычисленное значение критерия $F_{эмп}$ больше критического для определенного уровня значимости и соответствующих чисел степеней свободы для числителя и знаменателя, то дисперсии считаются различными

Пример. Дисперсия такого показателя, как морозоустойчивость для асфальтобетонной смеси АБЗ-1 составила 6.17 ($n_1 = 32$), а для продукции АБЗ-2 – 4.41 ($n_2 = 33$). Можно ли считать уровень дисперсий примерно одинаковым для данных выборок на уровне значимости 0.05.

Для ответа на поставленный вопрос определим эмпирическое значение критерия: $F_{эмп} = \dfrac{6.17}{4.41} \approx 1.4$. При этом критическое значение критерия $F_{кр}(0.05;\ 31;32) = 2$.

Значит $F_{эмп} = 1.4 < 2 = F_{кр}$, поэтому нулевая гипотеза о равенстве генеральных дисперсий на уровне значимости 0.05 принимается.

Следует заметить, что критерий Фишера весьма чувствителен к отклонениям от нормальности изучаемого признака в рассматриваемых выборках.

Как уже не раз отмечалось, разброс данных измерений характеризуется среднеквадратическим отклонением σ. Другим показателем разброса является коэффициент вариации (или вариация) K_B:

$$K_B = \frac{\sigma}{x}.$$

Чем больше σ и K_B, тем больше разброс показателей. С помощью σ и K_B можно определить интервал, в котором с заданной вероятностью окажутся данные эксперимента.

Интервал называют *доверительным интервалом* μ:

$$\mu = \pm t\,\sigma,$$

где t – коэффициент гарантии (гарантийный коэффициент).

Если обозначить математическое ожидание через M ($M = \lim\limits_{n\to\infty} \bar{x}$), то в интервале $M = \pm\, 0.6745$ находится 50% всех возможных значений случайной величины x. Следовательно, задавшись вероятностью 0.5, получим $t = 0.6745$. С вероятностью 0.9973 доверительный интервал будет равен: $\mu = \pm\, 3\,\sigma$ (т. е. $t = 3$).

Доверительной вероятностью (надежностью) измерения называют вероятность того, что истинное значение измеряемой величины (x_n) попадет в доверительный интервал $\mu = \pm\, t\,\sigma$. Она определяется в долях единицы (или в процентах) по таблице интеграла вероятностей.

Пример. Определить доверительный интервал и доверительную вероятность для измерений, представленных в таблице и на гистограмме (рис. 2.2). При $n = 78$, $\bar{x} = 71.5$ и $\sigma = 20.1$ для различных значений t получим доверительные интервалы, представленные в таблице.

Значения коэффициента гарантии t	*Доверительный интервал* $\mu = \pm\, t\,\sigma$
0.645	$\mu = \pm\, 0.645 \cdot 20.1 = \pm 12.96$
1.00	$\mu = \pm\, 20.1$
2.00	$\mu = \pm\, 2 \cdot 20.1 = \pm 40.2$
3.00	$\mu = \pm\, 3 \cdot 20.1 = \pm 60.3$

Иначе говоря, при $t = 1$ с вероятностью 0.7 истинное значение случайной величины x будет находиться в пределах 71.5 ± 20.1 (или в интервале $51.4 \ldots 91.6$). По таблице значений функции Лапласа: $\Phi(1) = 0.68$, значит из каждой сотни измерений 68 попадут в указанный интервал, а 32 измерения будут за его пределами.

При $t=3$: $\Phi(3)=0.997$. Значит, из 1000 измерений 997 измерений попадут в интервал $71.5\ldots60.3$ и только 3 измерения окажутся вне интервала.

При оценке доверительного интервала, доверительной вероятности и уровня значимости используются статистические характеристики: математическое ожидание M (или статистическое среднее $\bar{x}$), среднеквадратическое отклонение σ. Они зависят от количества и точности измерений.

В практике экспериментальных измерений часто бывает необходимо проверить наличие систематической погрешности вследствие влияния какого-либо фактора. Например, при проведении испытаний мостов в процессе работы погодные условия могут меняться, и это отражается на результатах измерений. Для выявления систематической погрешности весь массив измерений разбивают на группы по принципу различного влияния исследуемого фактора. При проведении измерений некоторого параметра для выяснения, влияет ли на них температура воздуха, можно выделить в отдельные группы утренние, дневные, ночные измерения.

Пусть имеется *n* измерений, разбитых на k групп, и n_i – измерений в *i-ой* группе. В каждой группе должно иметь место нормальное распределение результатов измерений. Это означает, что разброс результатов в каждой группе обусловлен лишь случайными погрешностями. Совокупная характеристика случайных погрешностей в группах может быть выражена средним арифметическим значением дисперсий D_{Xi}, определенных для каждой группы. Эта величина называется *внутригрупповой* дисперсией *DВГ*. Если число измерений в каждой группе одинаково, т. е. $n_i=\frac{n}{k}$, можно записать:

$$D_X ВГ=\frac{1}{n-k}\sum_i^k\sum_j^{n_j}\left(x_{ij}-\bar{x}_i\right)^2 .$$

Поскольку группы подобраны таким образом, что в каждой из них дисперсия результатов определяется только случайными причинами, то и величина *DВГ*, очевидно, отражает только эти случайные причины (случайные погрешности).

Рассеивание средних $\bar{x}$ по различным группам результатов обусловлено не только случайными погрешностями, но и систематическим воздействием фактора, по различным проявлениям которого сформированы группы. Поэтому, если вычислить дисперсию массива средних значений $\{\bar{x}\}$, то она будет отражать различие между группами, обусловленное систематическими причинами.

Такую дисперсию называют *межгрупповой*:

$$D_x M\Gamma = \frac{1}{k-1}\sum_{i}^{k}\left(\bar{x}_i - \bar{x}\right)^2.$$

Отношение $F = \dfrac{D_x M\Gamma}{D_x B\Gamma}$ выражает соотношение влияния на разброс систематической и случайной погрешности. Величина F и есть критерий Фишера.

Пример. В течение суток в разное время проведено 30 измерений провиса металлического пролетного строения, в том числе по 10 измерений – в 5 часов утра, 12 часов дня и в 19 часов вечера. Внутригрупповая дисперсия составила 0.05 см^2. Межгрупповая дисперсия – 0.2 см^2. Выявить, была ли систематическая погрешность, вызванная разным нагревом конструкции на протяжении суток.

Решение. Расчетное значение критерия Фишера:

$$F = \frac{0.2}{0.05} = 4.$$

Число степеней свободы в случае межгрупповых сравнений вычисляем по формулам: $K_1 = k - 1 = 3 - 1 = 2$; $K_2 = n - s = 28$.

Табличное значение критерия Фишера: для уровня значимости $\alpha = 0.05$ – $F_{\alpha=0.05} = 3.37$; для $\alpha = 0.01$ – $F_{\alpha=0.01} = 5.53$. Таким образом, в первом случае, т. е. с вероятностью 0.95, можно говорить о систематической погрешности, а во втором, т. е. с вероятностью 0.99, этого сказать нельзя. Какой именно результат принять – решение принимает экспериментатор.

Исправление выявленных систематических погрешностей при обработке результатов может быть выполнено путем введения поправок, нейтрализующих влияние факторов, вызывающих погрешности.

Критерии Стьюдента и Фишера успешно применяются для задач оптимизации состава цементогрунта с комплексной добавкой на основе кремнийорганических соединений; статистического анализа прочностных характеристик щебня; оценки факторов и показателей рабочего процесса машины для закрепления дорожного полотна и др.

Раздел 3. ОДНОФАКТОРНЫЙ АНАЛИЗ

3.1. Регрессионные модели

Регрессионный анализ – статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную. Независимые переменные иначе называют регрессорами или предикторами, а зависимые переменные критериальными. Терминология *зависимых* и *независимых* переменных отражает лишь математическую зависимость переменных, а не причинно-следственные отношения.

Цели регрессионного анализа:

- определение степени вариации зависимой переменной независимыми переменными;
- предсказание значения зависимой переменной с помощью независимых;
- определение вклада отдельных независимых переменных в вариацию зависимой.

Регрессионный анализ нельзя использовать для определения наличия связи между переменными, поскольку наличие такой связи и есть предпосылка для применения анализа. Проблема регрессионного анализа характерна тем, что о распределениях изучаемых величин нет достаточной информации. Пусть имеются основания предполагать, что случайная величина Y имеет некоторое распределение вероятностей и нужно по результатам наблюдений определить значения параметров этого распределения. В зависимости от природы задачи и целей анализа результаты эксперимента по-разному интерпретируются в отношении переменной x. Довольно часто в практике исследований имеет место ситуация, когда важнейшие переменные, описывающие некоторый процесс, известны заранее, но модель самого процесса неизвестна. В таких случаях возможны разные подходы. Одним из них является построение эмпирических моделей.

Для установления связи между величинами в эксперименте используется модель, основанная на допущениях:

- величина x является контролируемой величиной, значения которой заранее задаются при планировании эксперимента;
- независимые переменные при различных измерениях одинаково распределены с нулевым средним и постоянной дисперсией.

В случае неконтролируемой переменной результаты наблюдений $(x_1, y_1), \ldots, (x_n, y_n)$ представляют собой выборку из некоторой двумерной совокупности. Методы регрессионного анализа одинаковы и в том, и в другом случае, однако интерпретация результатов различается (в последнем случае анализ существенно дополняется методами теории корреляции). Выбор вида функций иногда определяется по расположению экспериментальных значений (x, y) на диаграмме рассеяния, чаще из теоретических соображений. Функциональная зависимость между переменными x и y это правило, которое каждому элементу x из множества $\{X\}$ ставит в соответствие элемент y из множества $\{Y\}$.

Если график функции регрессии $y_x = f(x)$ или $x_y = \varphi(y)$ изображается прямой линией, то регрессию называют *линейной* ($y_x = ax + b$), в противном случае – *нелинейной*.

Для определения вида функции регрессии в системе координат наносят точки $(x; y_x)$, строят *диаграмму рассеяния* (рис. 3.1) и по их расположению делают заключение о примерном виде функции регрессии (выдвигают гипотезу).

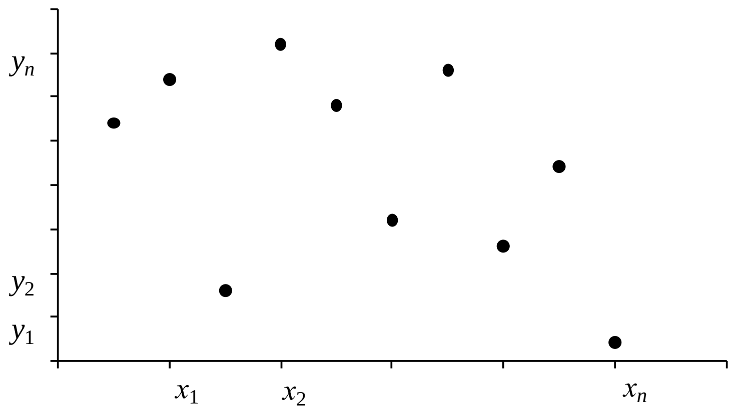


Рис. 3.1. Диаграмма рассеяния

Не существует общего правила для выбора подходящего вида функции. Можно лишь догадываться о форме уравнения регрессии. Однако существуют способы, с помощью которых можно проверить, является ли догадка удачной. Процесс нахождения теоретической линии регрессии заключается в выборе и обосновании типа прямой (кривой) и расчете параметров ее уравнения. Результатом является возможность составить уравнение регрессии и получить количественную оценку влияния факторных признаков на результативный признак.

Регрессионный анализ естественным образом обобщается, когда зависимая переменная зависит не от одной, а от нескольких независимых переменных. Очевидно, что одновременный учет нескольких факторов, связанных с интересующей нас величиной, позволяет построить модель, точнее описывающую имеющиеся данные и лучше прогнозирующую зависимую переменную. Безусловно, процедура оценки наиболее эффективна при правильно спланированном эксперименте, но требуют рассмотрения и те случаи, когда в распоряжении имеются данные, полученные в заранее не спланированном эксперименте. В таком случае следуют некоторым правилам, позволяющим выбрать наиболее подходящую модель с наименьшими возможными затратами:

- регрессионное уравнение должно содержать минимальное число коэффициентов;
- желательно, чтобы уравнение имело под собой содержательное обоснование. Например, демографические изменения, если нет ограничений на пищевые и другие ресурсы, осуществляются по экспоненте, поэтому модель этого процесса должна иметь соответствующую функциональную зависимость.

Лучшая процедура отбора наиболее подходящих моделей – пошаговый регрессионный анализ. Суть его в том, что отдельные переменные последовательно включаются в первоначальную модель и на каждом этапе анализируются, приводит ли добавление переменной к существенному или статистически значимому приближению предсказанных значений к эмпирическим данным.

В значительной мере достоверность полученных оценок зависит от некоторых предположений относительно поведения случайной ошибки:

- случайный характер – отдельные ошибки представляют собой случайные величины;
- нулевое среднее – каждое отклонение ошибки характеризуется нулевым математическим ожиданием и не зависит от значений x_i;
- дисперсия каждого отклонения равна для всех точек и независима от x_i;
- отсутствие взаимосвязи (автокорреляции) ошибок;
- ошибка должна иметь нормальное распределение.

При выполнении этих условий можно приближенно оценить точность предсказания влияния x на y, так как именно предсказание является одной из главных целей регрессионного анализа.

Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК), разработанный К. Гауссом и А. Лежандром. В его основу положена теория исследования на экстремум функции нескольких переменных.

3.2. Метод наименьших квадратов

Метод наименьших квадратов (МНК) позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений y фактических значений результативного признака от расчетных (теоретических) $\overline{y}_x$ минимальна:

$$S=\sum_i (y_i-\overline{y_{x_i}})^2 \to min.$$

Иными словами, из всего множества линий, линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между экспериментальными точками и точками этой линией была минимальной:

$$S=\sum_i (y_i-\overline{y_{x_i}})^2=\sum_i \left(y_i-ax_i-b\right)^2=\sum_i \Delta_i^2 \to min.$$

Процесс выражения опытных данных функциональной зависимостью с помощью МНК состоит из двух этапов: сначала выбирают вид искомой формулы, а затем для данной формулы подбирают параметры.

Рассмотрим в качестве эмпирической формулы линейную зависимость. Если выборочное уравнение регрессии имеет вид $y_x=ax+b$, то параметры a и b находят из системы линейных уравнений, применяя метод Крамера:

$$\begin{cases}\left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot a+\left(\sum\limits_{i=1}^{n} x_i\right)\cdot b=\left(\sum\limits_{i=1}^{n} x_i y_i\right);\\ \left(\sum\limits_{i=1}^{n} x_i\right)\cdot a+n\cdot b=\left(\sum\limits_{i=1}^{n} y_i\right);\end{cases}$$ где n – количество точек измерений.

$$\Delta=\begin{vmatrix}\sum x_i^2 & \sum x_i\\ \sum x_i & n\end{vmatrix}=n\cdot\sum x_i^2-\left(\sum x_i^2\right)^2;$$

$$\Delta a=\begin{vmatrix}\sum y_i x_i & \sum x_i\\ \sum y_i & n\end{vmatrix}=n\cdot\sum y_i x_i-\sum x_i\cdot\sum y_i;$$

$$\Delta b=\begin{vmatrix}\sum x_i^2 & \sum y_i x_i\\ \sum x_i & \sum y_i\end{vmatrix}=\sum x_i^2\sum y_i-\sum x_i y_i\cdot\sum x_i.$$

Вычислим коэффициенты регрессии: $a=\dfrac{\Delta a}{\Delta}$; $b=\dfrac{\Delta b}{\Delta}$

или $a=\dfrac{n\sum y_i x_i-\sum y_i\sum x_i}{n\sum x_i^2-\sum x_i^2}$; $b=\dfrac{\sum x_i^2\sum y_i-\sum x_i\sum y_i x_i}{n\sum x_i^2-(\sum x_i)^2}$.

Критерием точности полученных коэффициентов уравнения регрессии является сумма квадратов отклонений значений расчетной и экспериментальной функций. В результате получим значение критерия, выраженное одним числом $\sum \Delta_i^2 = (y_{p_i} - y_{э_i})^2$.

Параметр a называется коэффициентом регрессии. Его величина показывает среднее изменение признака y на a единиц при изменении фактора x на одну единицу. Знак при коэффициенте a показывает направление связи: при $a > 0$ связь прямая, при $a < 0$ – обратная. Коэффициент b равен значению y при $x = 0$, т. е. характеризует начальное состояние рассматриваемой зависимости – уровень отсчета.

Например, если уравнение регрессии имеет вид $y_x = 2x + 5$, где x – уровень заработной платы персонала частной строительной фирмы, а y_x – полученная средняя покупательная способность работников внебюджетных организаций, то смысл коэффициентов $a = 2$ и $b = 5$ таков: при увеличении заработной платы на 1 условную единицу, покупательная способность (теоретически, в среднем) повышается на 2 две условных единицы. Если заработную плату «заморозить», покупательная способность составит 5 условных единиц.

Уравнение регрессии может быть найдено также в виде $x_y = cy + d$. Процесс нахождения значений коэффициентов уравнения и их суть при этом сохраняются.

Возможность достаточно понятной интерпретации коэффициента линейной регрессии сделала этот вариант регрессионной модели достаточно распространенным в прикладных исследованиях. Параметр b может не иметь прикладного содержания. Интерпретировать можно лишь знак: если $b > 0$, то относительное изменение результата y происходит медленнее, чем изменение фактора x. Иными словами, вариация результата y меньше вариации фактора – коэффициент вариации по фактору x выше коэффициента вариации для результата y: $V_x > V_y$.

Зная уравнение регрессии, по точкам легко построить его наглядное представление – линию регрессии. С помощью полученного уравнения регрессии можно, подставляя задаваемые значения фактора X, в прогнозируемом периоде получить планируемую величину показателя.

Целью экономического анализа обычно является либо минимизация затрат ресурсов для достижения необходимого результата, либо максимизация эффективности использования имеющихся ограниченных ресурсов. Пусть проводится исследование по установлению необходимого типа и рационального радиуса действия производительных предприятий (заводов и баз) по приготовлению вяжущих и смесей для устройства покрытия усовершенствованного типа. По условиям использования в дорожном строительстве различают базы стационарные с радиусом $R_б$ более 100 км, длительность работы (T) на одном месте не менее 10 лет. Полустационарные с радиусом $40 < R_б < 60$ и временем T от 3 до 5 лет и передвижные (полевые) $5 < R_б < 10$. Необходимый тип и рациональный радиус действия базы могут быть определены из соотношения:

$$C = M + C_п + C_{тр},$$

где C – стоимость продукции базы, отнесенная на 1 м^2 устраиваемого покрытия; M – часть этой стоимости, не зависящая от искомого радиуса действия $R_б$ базы; $C_п$ – стоимость, обусловленная расходами по периодической передислокации базы, через каждые $R_б$ километров построенного покрытия; $C_{тр}$ – часть стоимости, обусловленная транспортными затратами на перевозку продукции базы (из расчета на 1 м^2 покрытия) на среднее расстояние $R_б / 2$.

Если рассмотреть выявленное соотношение по составляющим, то в самом общем плане, когда требуется установить как $R_б$, так и тип базы, величина M – отражает стоимость исходных материалов $M_м$ и стоимость процессов их переработки – $M_п$ на базе для выпуска продукции в количестве, требуемом на 1 м^2 покрытия, иначе $M = M_п + M_м$. Величину $M_м$ можно считать независимой от $R_б$. Характер функции $M_п = f(R_б)$ можно выявить на основе статистических данных стоимости приготовления вяжущих (смесей) на базах различных типов. В самом простейшем случае это может быть линейная зависимость вида: $M_п = a \cdot R_б + C_в$, где a – коэффициент связи, $C_в$ – стоимость (энергозатраты) приготовления вяжущего на базе, перемещающейся вместе с фронтом работ. Для стационарных и полустационарных баз величина $M_п$ должна быть меньше, чем для передвижных, поэтому в данной задаче коэффициент связи a должен быть отрицательным ($a < 0$).

Пример. Найти коэффициенты линейной регрессии для функции $M_п = f(R_б)$ на основе экспериментальных данных. Обозначим, для удобства записи $y_э = M_п$ и $x_i = R_б$:

Аргумент, x_i	$x_1 = 1$	$x_2 = 2.2$	$x_3 = 4$	$x_4 = 4.5$	$x_5 = 6$	$x_6 = 9$
Функция, $y_э$	$y_1 = 1$	$y_2 = 1$	$y_3 = 2$	$y_4 = 4$	$y_5 = 4$	$y_6 = 0$

Решение. Выполним предварительные расчеты:

$$\sum_1^6 x_i = 26.7; \quad \sum_1^6 y_i = 12; \quad \sum_1^6 x^2{}_i = 159.09; \quad \sum_1^6 x_i y_i = 53.2; \quad (\sum_1^6 x_i)^2 = 712.89.$$

Найдем значения коэффициентов уравнения регрессии:

$$a = \frac{6 \cdot 53.2 - 26.7 \cdot 12}{6 \cdot 159.09 - 712.89} = \frac{319.2 - 320.4}{954.54 - 712.89} = \frac{-1.2}{241.65} = -0.00497,$$

$$b = \frac{159.09 \cdot 12 - 26.7 \cdot 53.2}{6 \cdot 159.09 - 712.89} = \frac{1909.08 - 1420.44}{954.54 - 712.89} = \frac{488.64}{241.65} = 2.022.$$

Искомое уравнение линейной регрессии имеет вид: $y_p = -0.0045x + 2.022$. Вычислим «модельные» значения зависимой переменной и построим график линейной зависимости (рис. 3.2.).

x_i	$x_1 = 1$	$x_2 = 2.2$	$x_3 = 4$	$x_4 = 4.5$	$x_5 = 6$	$x_6 = 9$
y_p	2.015	2.009	3.5	4.25	5	0

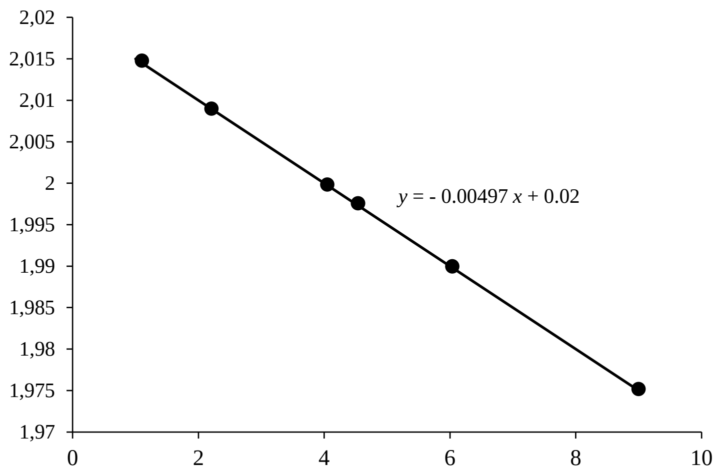


Рис. 3.2. График линейной регрессии

В практических исследованиях всегда имеет место некоторое рассеяние точек относительно линии регрессии. Оно обусловлено влиянием прочих, не учитываемых в уравнении факторов. Имеют место отклонения фактических данных от теоретических $y_э - y_p$. Величина этих отклонений и лежит в основе расчета остаточной дисперсии (суммы квадратов отклонений). Обозначим $\Delta = y_э - y_p$ и рассчитаем критерий оценки построенного уравнения регрессии:

$$\sum_1^6 \Delta_i^2 = \sum_1^6 \left(y_{эi} - y_{pi}\right)^2 = 1.03 + 1.02 + 0 + 4.01 + 3.9 = 14.$$

Чем меньше величина остаточной дисперсии, тем лучше уравнение регрессии подходит к исходным данным. Если остаточная дисперсия оказывается примерно одинаковой для нескольких функций, то на практике предпочтение отдается более простым видам функций, так как они в большей степени поддаются интерпретации и требуют меньшего объема наблюдений.

Результаты многих исследований подтверждают, что число наблюдений должно в 6-7 раз превышать число рассчитываемых параметров при переменной x. Это означает, что искать линейную регрессию, имея менее 6 наблюдений, бессмысленно. Для других видов функциональных зависимостей каждый параметр должен рассчитываться хотя бы по 6 наблюдениям.

3.3. Метод выравнивания для нелинейных зависимостей

Много реальные процессы и явления предполагают нелинейный характер зависимости между переменными. Чаще всего функции регрессии, используемые при количественной оценке связей между переменными имеют вид:

$y_x = ax^2 + bx + c$ – квадратичная;

$y_x = bx^a$ – степенная;

$y_x = be^{ax}$ – экспоненциальная;

$y_x = ba^x$ – показательная;

$y_x = a\ln x + b$ – полулогарифмическая;

$y_x = b + \dfrac{a}{x}$; $\quad y_x = \dfrac{1}{ax+b}$ – гиперболические.

В случае полиномиального вида уравнение регрессии, например, когда уравнение имеет вид $y_x = ax^2 + bx + c$, то параметры a, b и c находят из системы линейных уравнений, тем же методом Крамера:

$$\begin{cases} \left(\sum\limits_{i=1}^{n} x_i^4\right)\cdot a + \left(\sum\limits_{i=1}^{n} x_i^3\right)\cdot b + \left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot c = \left(\sum\limits_{i=1}^{n} x_i^2 y_i\right); \\ \left(\sum\limits_{i=1}^{n} x_i^3\right)\cdot a + \left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot b + \left(\sum\limits_{i=1}^{n} x_i\right)\cdot c = \left(\sum\limits_{i=1}^{n} x_i y_i\right); \\ \left(\sum\limits_{i=1}^{n} x_i^2\right)\cdot a + \left(\sum\limits_{i=1}^{n} x_i\right)\cdot b + n\cdot c = \left(\sum\limits_{i=1}^{n} y_i\right). \end{cases}.$$

В более общих случаях нелинейности применяется аналогичная методика расчета коэффициентов уравнения, только нелинейные функции следует привести к линейному виду, применив метод выравнивания исходных данных.

Приведем пример линеаризации нелинейной зависимости вида $y = ab^x$.

Прологарифмируем обе части уравнения $\lg y = \lg a + x\lg b$ и введем новые переменные $Y = \lg y$, $A = \lg a$, $B = \lg b$. В итоге получим уравнение линейного вида $Y = A + Bx$, коэффициенты которого находим стандартным методом наименьших квадратов. После чего, выполнив обратную замену, получим искомую нелинейную модель.

Другие, вышеприведенные нелинейные функции так же легко приводятся к линейному виду путем введения новых переменных.

Пример. Построим уравнение нелинейной регрессии заданного вида, применив метод выравнивания к закону гиперболического типа $y = \dfrac{1}{ax+b}$.

Аргумент, x_i	$x_1 = 1$	$x_2 = 2$	$x_3 = 3$	$x_4 = 4$	$x_5 = 5$
Значение, $y_э$	$y_1 = 2.3$	$y_2 = 1.8$	$y_3 = 3.8$	$y_4 = 5.3$	$y_5 = 4.3$

Введем новую переменную $Y = \dfrac{1}{y}$ и получим таблицу для расчета коэффициентов регрессионного уравнения по МНК:

x	1	2	3	4	5
y	2.3	1.8	3.8	5.3	4.3
$Y = 1/y$	0.4	0.5	0.3	0.2	0.2

$$\sum_1^5 x_i = 15;\quad \sum_1^5 Y_i = 1.6;\quad \sum_1^5 x^2{}_i = 55;\quad \sum_1^5 x_i Y_i = 4.1;\quad (\sum_1^5 x_i)^2 = 225.$$

Найдем коэффициент регрессии для выравненного закона $Y = A + Bx$.

$$a = \frac{5\cdot 4.1 - 15\cdot 1.6}{5\cdot 55 - 225} = \frac{20.5-24}{50} = \frac{-3.5}{50} = -0.07;$$

$$b = \frac{55\cdot 1.6 - 15\cdot 4.1}{5\cdot 55 - 225} = \frac{88-61.5}{50} = \frac{26.5}{50} = 0.53.$$

Тогда $Y = -0.07x + 0.53$. Вернемся к исходной переменной:

$$\frac{1}{y} = -0.07x + 0.53.$$

Окончательно уравнение нелинейной регрессии примет вид:

$$Y_p = \frac{1}{-0.07x + 0.53}.$$

Рассчитаем «модельные» значения, сумму квадратов отклонений и построим график нелинейной регрессии (рис. 3.3.):

x	1	2	3	4	5
Y_p	2.2	2.6	3.1	4	5.5

$$\sum \Delta_i^2 = 0.01 + 0.64 + 0.49 + 1.69 + 1.11 = 4.27.$$

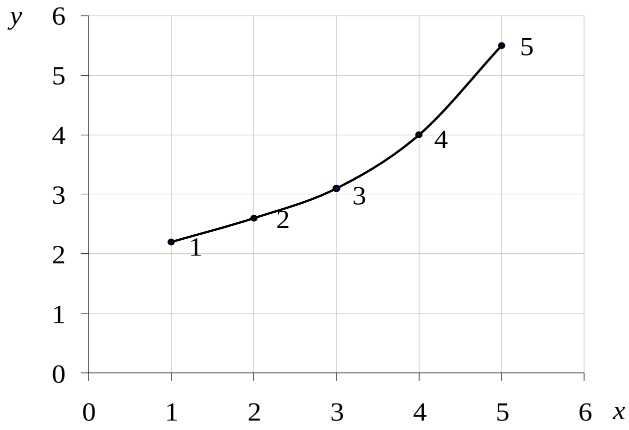


Рис. 3.3. График нелинейной регрессии

Для нелинейных видов функциональных зависимостей каждый параметр, как отмечалось, рассчитывается хотя бы по 6 наблюдениям. Следовательно, для параболы второй степени $y = ax^2 + bx + c$ необходимы не менее 12 данных, а для параболы третьей степени $y = ax^3 + bx^2 + cx + d$ – не менее 18 наблюдений.

3.4. Элементы теории корреляции

Интуитивно ясно, что о взаимозависимости между парой переменных можно говорить в тех случаях, когда уменьшению (увеличению) одной из них будет соответствовать уменьшение (увеличение) другой либо уменьшению (увеличению) первой будет соответствовать увеличение (уменьшение) второй переменной. В первом случае можно говорить о положительной корреляции между переменными (прямая зависимость), во втором – об отрицательной корреляции (обратная зависимость). Если рассмотреть разброс значений переменных относительно их средних, получим положительные и отрицательные разности, и знак их будет также различен.

При однонаправленных изменениях обеих переменных произведение их отклонений положительно, если же изменения переменных разнонаправлены, то произведение отрицательно.

Величина, полученная как отношение суммы произведений отклонений длине выборке без единицы, называется *ковариацией*. Ковариация вычисляется по формуле:

$$S_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n-1}.$$

Признаки x и y, по которым рассчитывается взаимосвязь, могут измеряться в разных единицах, иметь произвольные средние и дисперсии. Поэтому, вычитание соответствующих средних по каждой переменной делает ковариацию независимой от средних. Если разделить ковариацию на произведение стандартных отклонений, получим безразмерный коэффициент связи, который называется коэффициентом корреляции. *Коэффициент корреляции* (*Пирсона*) представляет собой численную меру степени взаимосвязи двух переменных, введенную в статистическую практику К. Пирсоном.

Коэффициент корреляции вычисляется по формуле:

$$r_{xy} = \frac{S_{xy}}{S_x S_y}.$$

Очевидно, что $S_{xy} = S_{yx}$, *следовательно,* $r_{xy} = r_{yx}$, поэтому с помощью коэффициента корреляции можно численно оценить величину и направленность взаимосвязи. Имеются разные модификации формулы линейного коэффициента корреляции, например:

$$r_{xy} = a\frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \overline{y}\cdot\overline{x}}{\sigma_x \sigma_y},$$

где $\overline{xy} = \frac{\sum_{i=1}^{n} x_i\, y_i}{n}$; $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - (\overline{x})^2}$; $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - (\overline{y})^2}$;

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n};\ \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

По поводу интерпретации коэффициента корреляции сделаем два существенных замечания. При выводе расчетной формулы r_{xy} не делалось предположений о характере совместного двумерного распределения величин

x и y. Очевиден вывод о пределах его изменения в диапазоне от -1 до $+1$. Выражения «сильная связь», «слабая связь», «умеренная связь» и т. д. справедливы только в рамках определенной статистической модели. Так, если частотные распределения величин x и y имеют разные значения асимметрии, то есть существенно скошены в разных направлениях, то даже при максимально возможной линейной связи между x и y величина коэффициента корреляции не будет по абсолютной величине превышать значения 0.6-0.7. Эта зависимость максимальной величины коэффициента корреляции от характера распределения x и y приводит к трудностям интерпретации получаемых его конкретных значений. Что означает $r_{xy} = 0.6$? Максимально возможную линейную связь при положительной и отрицательной асимметрии распределений x и y или умеренную связь этих переменных при совместном распределении, подобном двумерному нормальному распределению? Ответы на эти вопросы можно получить из качественного анализа диаграмм рассеяния и гистограмм распределения.

Второе замечание связано со значением коэффициентов, близких к нулю. Равенство нулю коэффициентов корреляции между переменными не всегда свидетельствует об отсутствии статистической связи между x и y. Так может проявляться, например, их нелинейная связь. Возможные варианты проявлений ложной корреляции могут быть еще связаны с появлением в совокупности исходных данных аномальных значений, или за счет неоднородности анализируемого материала, или за счет ошибок при регистрации данных.

Обобщим сказанное. Коэффициент корреляции находится в границах $-1 \le r_{xy} \le 1$. Если величина $r_{xy} \ge 0$, то зависимость между X и Y такова, что возрастание значений одной из переменных приводит к увеличению значений другой переменной. При значениях $r_{xy} < 0$ увеличение одной переменной приводит к уменьшению другой. Чем ближе значение r_{xy} к (± 1), тем сильнее (теснее) переменные X и Y связаны линейной функциональной зависимостью. Если $r_{xy} = \pm 1$, то линии регрессии $\overline{y}_x$ и $\overline{x}_y$ сливаются в одну, что соответствует строгой линейной зависимости между X и Y, в этом случае все наблюдаемые значения располагаются на общей прямой. При $r_{xy} = -1$ имеет место отрицательная линейная зависимость, при $r_{xy} = 1$ – положительная. Если $r_{xy} = 0$, то признаки X и Y некоррелированы. В этом случае линии регрессии $\overline{y}_x$ и $\overline{x}_y$ параллельны координатным осям.

Важно заметить, что величина линейного коэффициента корреляции оценивает тесноту связи рассматриваемых признаков в ее линейной форме. Поэтому близость абсолютной величины линейного коэффициента корреляции к нулю еще не означает, что X и Y независимы, если допускается отклонение этой зависимости от линейной. Следовательно, прямое утверждение $r_{xy} = 0$ не означает независимости исследуемой пары признаков.

В то же время, если X и Y независимы, то верно утверждение $r_{xy} = 0$. Таким образом, при отклонении парной статистической зависимости от линейной коэффициент корреляции теряет свой смысл как характеристика степени тесноты связи. В этом случае следует воспользоваться другим измерителем связи – корреляционным отношением.

Допустим, что выборочный коэффициент корреляции (найденный по выборке) оказался отличным от нуля. Так как выборка случайна, то еще нельзя заключить, что коэффициент корреляции генеральной совокупности также отличен от нуля. В этом случае проверяют гипотезу о значимости (существенности) выборочного коэффициента корреляции. Другая формулировка нулевой гипотезы: коэффициент корреляции генеральной совокупности r_{xy} равен нулю. Для оценки коэффициента корреляции r_{xy} нормально распределенной генеральной совокупности можно воспользоваться формулой:

$$r_{xy} - 3 \cdot \frac{1 - r_{xy}^2}{\sqrt{n}} \le r_{г.с.} \le r_{xy} + 3 \cdot \frac{1 + r_{xy}^2}{\sqrt{n}} \text{ (при } n \ge 50\text{)}.$$

Вернемся к рассматриваемому примеру и выполним расчет коэффициента корреляции по формуле:

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n\sum x_i^2 - \left(\sum x_i\right)^2] \cdot [n\sum y_i^2 - \left(\sum y_i\right)^2]}} =$$

$$= \frac{6 \cdot 53.2 - 26.7 \cdot 12}{\sqrt{(6 \cdot 159.09 - 712.89) \cdot (6 \cdot 38 - 144)}} = \frac{-1.2}{142.47} = -0.00842.$$

Полученные результаты значений и коэффициента a в уравнении регрессии, и коэффициента корреляции r_{xy} согласуются с содержательным анализом условия задачи и подтверждают наличие обратной связи между изучаемыми величинами функции $M_п = f(R_б)$.

Для изучения степени неравномерности распределения определенного суммарного показателя между единицами отдельных групп вариационного ряда часто используют *кривую Лоренца* (или кривая концентрации). Кривая Лоренца была предложена американским экономистом Максом Отто Лоренцем в 1905 году как показатель неравенства в доходах населения. В прямоугольной системе координат кривая Лоренца является выпуклой вниз и проходит под диагональю единичного квадрата, расположенного в I координатной четверти. Каждая точка на кривой Лоренца соответствует утверждению вида: «20 самых бедных процентов населения получают всего 7% дохода». В случае равномерного распределения каждая группа населения имеет доход, пропорциональный своей численности. Такой случай описывается кривой равенства, являющейся прямой, соединяющей начало координат и точку (1; 1). В случае полного неравенства (когда лишь один член общества имеет доход) кривая сначала «прилипает» к оси абсцисс, а потом из точки (1; 0) «взмывает» к точке (1; 1).

Пример. Пусть имеется распределение городов по числу жителей и распределение финансовых потоков в транспортном строительстве в этих городах (графы 1, 2, 3). Построить кривую Лоренца.

Города с числом жителей (тыс. чел.)	*Число городов, в % к итогу* W_i	*Финансирование транспортного строительства % к итогу* Y_i	*Кумулятивные итоги*	
			% городов $cumW_i$	*% финансирования* $cumY_i$
До 3	4.2	0.2	4.2	0.2
3-5	4.6	0.2	8.8	0.5
5-10	13.1	1.7	21.9	2.2
10-20	28.3	6.8	50.2	9.0
20-50	28.7	14.8	78.9	23.8
50-100	9.7	10.3	88.6	34.1
100-500	9.7	33.8	98.3	67.9
Свыше 500	1.7	32.1	100	100
Итого	100.0	100.0	-	-

На координатной плоскости наносим точки ($cumW_i$; $cumY_i$) и по ним строим кривую.

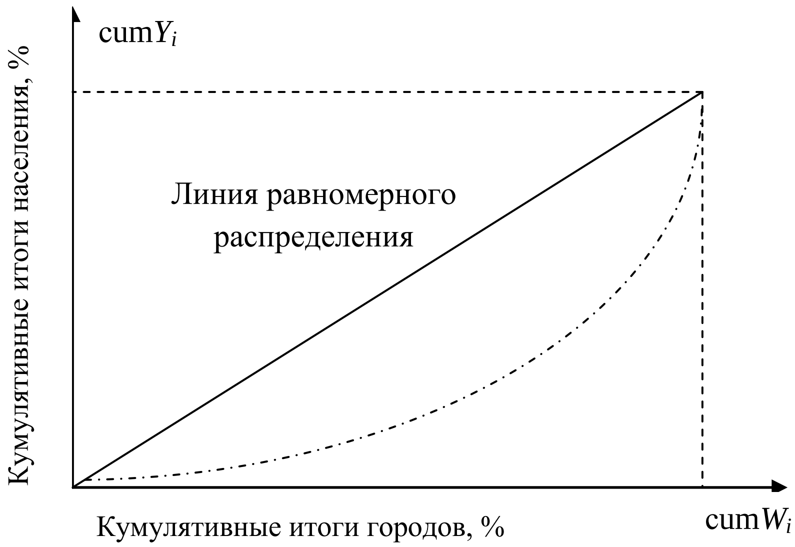


Рис. 3.4. Кривая Лоренца

Замечание, чем больше вогнутость (отличие кривой от линии равномерного распределения), тем выше концентрация (финансирования транспортного строительства) в определенных группах единиц (крупных городах).

Кривые Лоренца применяют для распределений не только доходов, но и имущества домохозяйств, долей рынка для фирм в отрасли, природных ресурсов по государствам. *Концентрация* в производственной сфере представляет собой процесс сосредоточения изготовления продукции на ограниченном числе предприятий и в их производственных подразделениях.

Концентрацию производства в строительной отрасли следует понимать, как процесс сосредоточения средств, предметов и ресурсов труда в крупных производственных звеньях. Материально-технической основой концентрации выступает научно-технический прогресс. Современное производство позволяет наиболее высоко использовать преимущества концентрации в интересах всего общества. Социальное значение концентрации состоит в том, что повышение технического предела производства создает климат для ликвидации тяжелого ручного труда, ускорения процесса ликвидации разницы между работниками умственного и физического труда.

Как форма организации производства концентрация имеет большую экономическую эффективность. Крупные предприятия обеспечивают, как правило, более высокий уровень производительности труда, лучшее качество продукции, более низкую себестоимость. Концентрация неразрывно связана со специализацией производства и комбинированием. Развитие специализированного производства выступает как прогрессивная форма концентрации однородного производства. Комбинирование, осуществляемое на крупных предприятиях, позволяет организовать производство на более высоком научно-техническом уровне. Концентрация производства может осуществляться путем увеличения размеров действующих предприятий, за счет их расширения, реконструкции или технического перевооружения; создания новых крупных предприятий и образования производственных объединений, что ведет к возникновению укрупненных производственных ячеек на базе централизации производства. Так как результатом централизации является расширение размеров предприятий, то ее также принято рассматривать как одну из форм осуществления концентрации. Уровень концентрации определяется системой показателей, основным из которых является объем производимой продукции.

В условиях научно-технического прогресса усиливается значение таких показателей концентрации, как стоимость основных производственных фондов, в том числе активного назначения, мощность энергоустановок и т. п. Для характеристики уровня концентрации применяется, и особенно широко в международных сопоставлениях, показатель численности работников, хотя этот показатель не всегда точно отражает динамику уровня концентрации, поскольку численность работников может сократиться при укрупнении производства в результате механизации и автоматизации трудовых процессов.

Концентрация имеет свои особенности в строительстве. Они вызваны спецификой строительного производства, заключающейся в неподвижности строительной продукции. Поэтому концентрация в строительстве развивается в 2-х направлениях: за счет увеличения строительных организаций и за счет увеличения строительных объектов или увеличения объемов строймонтажных работ, выполняемых на одной площадке строительства.

Наиболее распространен 1-ый тип концентрации – укрупнение строительных организаций. 2-ой тип концентрации – территориальный – полнее всего проявляется при строительстве предприятий и объектов крупными сосредоточенными узлами (промышленный узел, большие жилые массивы и т. п.). Во всех случаях основным показателем для оценки уровня концентрации в строительстве являются годовые объемы строймонтажных работ.

Раздел 4. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

4.1. Основные виды задач линейного программирования

Математические методы и модели применяются для отыскания оптимальных решений в управлении. Всякое управление предполагает наличие цели, ради реализации которой осуществляется управление. Появление достаточно мощной вычислительной техники, способной за короткое время переработать значительные массивы информации, а также развитие теории математических моделей способствовало преобразованию искусства принятия решения в науку, доступную каждому, кто овладел ее принципами и методологией.

Среди методов математического программирования наибольшее распространение получили методы линейного программирования. Термин «программирование» показывает, что они применяются для составления плана (программы), который обеспечивал бы оптимальное использование материальных и трудовых ресурсов. Слово линейное определяет математическую природу этих моделей. Линейность связана с понятиями пропорциональности и возможностью суммирования результатов (аддитивностью). Она отражается в том, что условия задач выражаются системой линейных уравнений или неравенств, содержащих неизвестные только первой степени. Для любых задач линейного программирования характерны три следующих условия (по академику В.С. Немчинову):

- наличие системы взаимосвязанных факторов;
- строгое определение критерия оценки оптимальности;
- точная формулировка условий, ограничивающих использование наличных ресурсов.

Общая схема построения моделей задач принятия решений

1. Выделяют параметры, которые описывают процесс принятия решения. Выбор тех или иных численных значений для параметров набора эквивалентен принятию того или иного решения. Эти параметры принято называть параметрами управления, число их может быть весьма значительным.

2. Выделяют и четко формулируют цель, ради достижения которой принимается то или иное решение. Как правило, целевую установку процесса принятия решения представляют в виде некоторой зависящей от параметров управления функции, значения которой дают оценку качества принятого

решения. Эту функцию называют *целевой функцией* задачи или показателем качества ее решения. Целевая функция дает возможность сравнивать два различных решения: если при одном наборе параметров значение целевой функции меньше, чем при другом (или больше, в зависимости от задачи), то одно решение будет предпочтительнее другого.

3. Поскольку в задачах принятия решений не любой набор параметров управления может быть реализован практически, важно выделить среди всех решений множество возможных решений.

В дорожном строительстве такие задачи достаточно типичны, среди них можно выделить некоторые группы, среди них:

- материальное обеспечение строительства (построение оптимального плана поставок материалов и конструкций с нескольких баз, карьеров на объекты строительства);
- организация заготовительно-транспортных работ:
- перемещение земляных масс (распределение грунтов из m их выемок в n насыпей, а при отсутствии равенства объемов грунта – перемещение из боковых резервов или грунтовых карьеров);
- комплектование специализированных подразделений из m видов работ и n видов техники с целью обеспечения максимальной производительности и темпа строительных работ;
- организация ремонта поврежденной техники, что бы стоимость ремонтов (с учетом доставки техники) была минимальной, либо время возвращения отремонтированной техники на объекты работ было кратчайшим.

Кроме того существует большой класс сетевых задач, позволяющих проектировать рациональную сеть транспортных коммуникаций, выбирать направление обходов барьерных мест. При построении экономико-статистической модели важно уменьшить «размерность» решаемой задачи. Поэтому основным элементом построения модели является выделение доминирующих переменных, параметров и ограничений.

Допустим, что предметом моделирования является процесс поставки щебня из нескольких промышленных карьеров на асфальтобетонный и цементнобетонный заводы, а также для устройства щебеночного основания на линии. На предварительном этапе моделирования выявлен ряд факторов, влияющих на конечные экономические результаты строительства:

- стоимость щебня в карьерах (отпускная цена);
- стоимость транспортировки материала от карьера до приобъектного склада, которая зависит от расстояния перевозки и принятой транспортной схемы;

- возможные потери щебня в процессе перегрузки материала с одного вида транспорта на другой;
- различная сложность организации транспортного процесса и управления при доставке материала из одного или нескольких карьеров;
- степень экономической стабильности поставщиков и их дисциплинированность при выполнении договорных обязательств.

Анализируя приведенные факторы – переменных модели – очевидно, что на первый и пятый стройка влияния не оказывает, их можно только учитывать при выборе альтернатив поставщиков, поэтому их следует отнести к ограничениям модели. Третий и четвертый факторы малозначительны в общей постановке задачи и их можно не учитывать в построенной модели. Наиболее существенным, а главное управляемым фактором, оказывается второй. Его можно включить в критерий эффективности и принять в качестве доминирующего.

Назовем процесс поставки щебня конечной продукцией моделирования. Результат по конечной продукции осуществляется с помощью одного из выбранных способов поставок.

Сформулируем условие задачи. Необходимо составить план, при котором поставки щебня в пределах имеющихся исходных запасов доставляли бы минимум функции затрат.

Формализуем условия задачи. Пусть n – число поставщиков; где j-й карьер $(j=\overline{1,n})$ участник поставок материалов на используемый i-й объект строительства $(i=\overline{1,m})$; m – перечень строительных объектов, на которые необходимо поставить щебень; a_j – запасы материалов на j-м карьере; c_{ij} – стоимость j-го ресурса при доставки его на i-й объект; b_i – потребность i-го объекта в заготавливаемых материалах.

Составим математическую модель задачи:

1. Выбираем параметры управления (то, что можно менять в процессе производства в рамках поставленной задачи). Отметим, что в рамках поставленной задачи нельзя менять b_i, c_{ij}, a_j, m и n (см. выше). В процессе производства можно менять только транспортную схему поставок щебня на строительные объекты, поэтому пусть x_j – объемы поставок материалов на объекты дорожного строительства. Следовательно, $x_1, x_2, \ldots, x_n$ – искомый набор параметров управления, а составленный план производства эквивалентен выбору конкретных значений параметров x_j.

2. Определяем целевую функцию. По условию задачи необходимо составить такую транспортную схему, чтобы прибыль строительной компании

была бы наибольшей. Следовательно, целевая функция должна отражать количество перемещаемого материала и его стоимостные показатели:

$$f(x)=\sum_{j=1}^{n}c_jx_j=c_1x_1+c_2x_2+\ldots+c_nx_n\ .$$

Для нахождения оптимального решения необходимо максимизировать целевую функцию, то есть $f(x)\to max$.

3. Поскольку запасы ресурсов ограничены, то на параметры управления накладываются ограничения. Объемы поставок не должны превышать имеющихся запасов. Необходимо добавить еще ограничения $x_j\ge 0$, так как время поставок ресурсов не может быть отрицательным.

В результате получим математическую модель *общей задачи* линейного программирования задачи:

$$f(x)=\sum_{j=1}^{n}c_jx_j\to max;$$

$$\begin{cases}\sum_{j=1}^{n}a_{1j}x_j\le b_1,\ \sum_{j=1}^{n}a_{2j}x_j\le b_2,\ \ldots,\ \sum_{j=1}^{n}a_{mj}x_j\le b_m.\\ x_j\ge 0,\ j=\overline{1,n}.\end{cases}$$

Целевая функция модели и ограничения линейно зависят от параметров управления. Подобные модели принято называть *задачами линейного программирования* (ЗЛП). Реально зависимости редко бывают строго линейными, однако предположение о линейности дает возможность разрабатывать эффективные методы, использование которых для ряда задач является вполне оправданным. Линейные модели могут содержать в качестве ограничений не только неравенства, но и уравнения.

Стандартная ЗЛП – определение максимального значения целевой функции: $f(x)=\sum_{j=1}^{n}c_jx_j\to max$, при ограничениях в виде неравенств:

$$\begin{cases}\sum_{j=1}^{n}a_{ij}x_j\le b_i, & i=\overline{1,k};\\ \sum_{j=1}^{n}a_{ij}x_j\ge b_i, & i=\overline{k+1,m};\\ x_j\ge 0, & j=\overline{1,n}.\end{cases}$$

Каноническая ЗЛП – определение максимального значения целевой функции:

$f(x)=\sum_{j=1}^{n}c_jx_j\to max$ при ограничениях в виде равенств: $\begin{cases}\sum_{j=1}^{n}a_{ij}x_j=b_i, & i=\overline{1,m};\\ x_j\ge 0, & j=\overline{1,n}.\end{cases}$

Совокупность значений переменных $x_1, x_2, ..., x_n$, удовлетворяющих всем линейным ограничениям, называется допустимым (опорным) планом.

Переход от стандартной модели ЗЛП к канонической осуществляется добавлением новых неотрицательных переменных со знаком «+» для неравенств типа $\le$ и со знаком «–» для неравенств типа $\ge$. Само неравенство заменяется на равенство.

Если в ЗЛП на некоторую переменную x_k не накладывается условие неотрицательности, то добавляют две неотрицательные переменные и делают замену переменных $x_k = x_{k_2} - x_{k_1}, x_{k_2} \ge 0, x_{k_1} \ge 0$.

При переходе от задачи на минимум к задаче на максимум изменяется знак целевой функции.

Задача линейного программирования может принимать оптимальное значение в нескольких точках, такие решения называются *альтернативными оптимальными решениями* или альтернативным оптимумом, причем в каждой из этих точек целевая функция имеет одно и то же оптимальное значение.

Если в ЗЛП с двумя параметрами управления нормальный вектор целевой функции параллелен нормальному вектору прямой одного из ограничений, то ЗЛП может иметь множество оптимальных решений, например:

$$f(x) = 2x_1 + 3x_2 \to max;$$

$$\begin{cases} 4x_1 + 6x_2 \le 15; \\ 7x_1 + 5x_2 \le 20; \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

Вектор целевой функции $\overrightarrow{N}(2;3)$ параллелен нормальному вектору $\overrightarrow{N_1}(4;6)$ прямой первого ограничения: $\overrightarrow{N_1} = 2\overrightarrow{N}$. В этом случае возможно множество оптимальных решений.

Если в ЗЛП с тремя параметрами управления нормальный вектор целевой функции параллелен нормальному вектору плоскости одного из ограничений или перпендикулярен к прямой пересечения двух плоскостей ограничений, то ЗЛП также может иметь множество оптимальных решений.

Для ЗЛП с большим количеством параметров управления сложно определить, имеется ли множество оптимумов по виду составленной математической модели.

Пример. Составить математическую модель задачи, привести ее к каноническому виду.

На автомобильной дороге требуется построить автомобильные мостовые конструкции по одному из трех проектов (МКА). Расходы, связанные с каждой конструкцией, среди прочих, определяются количеством трудозатрат (человеко-часов), электроэнергии (киловатт-часов) и потребляемого материала (т). Определить расходы на производство конструкций мостов, обеспечивающий наибольший доход фирме-производителю.

Расходы ресурсов:	На единицу продукции			Имеется в наличии
	МКА1	МКА2	МКА3	
человеко-часов	10	15	10	320
киловатт-часов	18	16	20	500
материала, т	70	50	60	2000
Доход с единицы продукции, млн. руб.	50	60	70	*max*

Решение.

1. В процессе производства можно менять количество выпускаемой продукции по определенному проекту МКА. Поэтому в качестве параметров управления x_1, x_2 и x_3 выберем количество выпускаемой продукции по соответствующему технологическому проекту МКА1, МКА2 и МКА3.

2. Целевая функция должна максимизировать доход от выпускаемой продукции: $f(x) = 50x_1 + 60x_2 + 70x_3 \to max$.

3. Запасы ресурсов ограничивают количество выпускаемой продукции. В процессе производства нельзя превысить количество имеющихся человеко-часов: $10x_1 + 15x_2 + 10x_3 \le 320$. Аналогично, составим условия для двух других ресурсов: $18x_1 + 16x_2 + 20x_3 \le 500$, $70x_1 + 50x_2 + 60x_3 \le 2000$. Необходимо добавить условие неотрицательности переменных: $x_1 \ge 0$, $x_2 \ge 0$, $x_3 \ge 0$.

В результате математическая модель принимает вид:

$$f(x) = 50x_1 + 60x_2 + 70x_3 \to max;$$

$$\begin{cases} 10x_1 + 15x_2 + 10x_3 \le 320; \\ 18x_1 + 16x_2 + 20x_3 \le 500; \\ 70x_1 + 50x_2 + 60x_3 \le 2\,000; \\ x_j \ge 0, \quad j = \overline{1,3}. \end{cases}$$

Это стандартная ЗЛП, ее решение: $f(x) = 1790$ руб., $x_1 = 0$, $x_2 = 10$, $x_3 = 17$.

Построим каноническую ЗЛП, добавив в первое, второе и третье неравенства неотрицательные переменные x_4, x_5, x_6:

$$f(x) = 50x_1 + 60x_2 + 80x_3 \to max;$$

$$
\begin{cases}
10x_1 + 15x_2 + 10x_3 + x_4 = 320; \\
18x_1 + 16x_2 + 20x_3 + x_5 = 500; \\
70x_1 + 50x_2 + 60x_3 + x_6 = 2\,000; \\
x_j \geq 0, \quad j = \overline{1,6}.
\end{cases}
$$

Заметим, что стоимостные критерии не всегда являются доминирующими. При восстановлении дорог в районах чрезвычайных ситуаций на первый план может выйти критерий «времени восстановления прерванного движения». В этом случае, целевая функция примет вид:

$$L = \sum_i^n \sum_j^m t_{ij} x_{ij} \to min.$$

Математический аппарат линейного программирования позволяет не только получить оптимальный план, но и сделать ряд экономически значимых выводов, основанных на свойствах задачи, двойственной к исходной (прямой).

Двойственная задача по отношению к исходной задаче составляется по следующим правилам:

- в исходной задаче находится максимум целевой функции, а в двойственной – минимум;
- количество неизвестных двойственной задачи равно количеству ограничений исходной задачи;
- количество ограничений двойственной задачи равно количеству переменных исходной задачи;
- ограничению исходной задачи вида «$\geq$» соответствует неотрицательная переменная двойственной задачи, ограничению исходной задачи вида «$\leq$» соответствует неположительная переменная двойственной задачи; ограничению исходной задачи вида «$=$» соответствует переменная двойственной задачи, которая может принимать как положительные, так и отрицательные значения;
- если на переменную исходной задачи наложено (не наложено) условие неотрицательности, то соответствующим ограничением двойственной задачи будет неравенство вида «$\geq$» (или уравнение);
- коэффициенты при неизвестных в целевой функции исходной задачи являются свободными членами в ограничениях двойственной задачи;
- свободные члены ограничений исходной задачи являются коэффициентами при неизвестных в целевой функции двойственной задачи;
- матрица, составленная из коэффициентов при неизвестных в системе ограничений исходной задачи, в двойственной задаче транспонируется.

<u>Пример.</u> Составить двойственную задачу для ЗЛП:

$$f(x) = x_1 + 2x_2 + 3x_3 \to max;$$

$$\begin{cases} x_1 + 5x_2 + 2x_3 = 70; \\ x_1 + x_2 - 4x_3 \le 30; \\ x_1 - 2x_2 + x_3 \ge 40; \\ x_1 \ge 0, \quad x_2 \ge 0. \end{cases}$$

Решение. Исходная задача сформулирована на максимум, значит двойственная задача формулируется на минимум. Исходная задача имеет три ограничения, двойственная задача будет иметь три переменные y_1, y_2 и y_3. Второе ограничение исходной задачи имеет вид «≤», тогда y_2 будет неотрицательным. Третье ограничение исходной задачи имеет вид «≥», тогда y_3 будет неположительным. В исходной задаче $x_1 \ge 0$ и $x_2 \ge 0$, в двойственной задаче первые два ограничения будут записаны в виде неравенств типа «≥». Коэффициенты при переменных целевой функции исходной задачи будут свободными членами ограничений двойственной задачи. Матрица коэффициентов при неизвестных в ограничениях исходной задачи транспонируется в двойственной задаче. Коэффициенты при переменных целевой функции двойственной задачи являются свободными членами ограничений исходной задачи.

Получаем двойственную задачу:

$$g(y) = 70y_1 + 30y_2 + 40y_3 \to min;$$

$$\begin{cases} y_1 + y_2 + y_3 \ge 1; \\ 5y_1 + y_2 - 2y_3 \ge 2; \\ 2y_1 - 4y_2 + y_3 = 3; \\ y_1 \ge 0, \quad y_2 \le 0. \end{cases}$$

Решение исходной задачи: $f(x) = 100$, $x_1 = 10$, $x_2 = 0$, $x_3 = 30$.

Решение двойственной задачи: $g(y) = 100$, $y_1 = 2$, $y_2 = 0$, $y_3 = -1$.

Экономическая интерпретация двойственной задачи заключается в определении оптимальных оценок y_i, называемых неявными или «теневыми» ценами каждого вида ресурса с целью минимизации общей ценности (стоимости) всех ресурсов. Значения переменных y_i в оптимальном решении двойственной задачи представляют собой оценки влияния свободных членов b_i системы ограничений исходной задачи на величину приращения целевой функции $\Delta f(x) = \Delta b_i y_i$.

Рассмотрим следующую производственную задачу.

Предприятие после выпуска основной продукции имеет излишки ресурсов двух типов: R_1 – 10 единиц, R_2 – 8 единиц. Существует два способа распорядиться этими ресурсами:

➢ организовать из них выпуск 3 новых видов продукции: P_1, P_2, P_3.

➢ продать их.

Рассмотрим оба способа и оценим целесообразность их реализации. Исходные данные приведены в таблице:

Ресурсы	*Расход ресурса на единицу продукции*			*Запас ресурсов*
	P_1	P_2	P_3	
R_1	1	2	1	10
R_2	2	1	3	8
Удельная прибыль	\$6	\$4	\$4	

Согласно *первому способу*, надо составить такой план выпуска продукции, который максимизирует суммарную прибыль. Построим математическую модель этой задачи. Пусть x_j – план выпуска продукции P_j. Тогда целевая функция будет выглядеть следующим образом:

$$f(x) = 6x_1 + 4x_2 + 4x_3 \to max;$$

Ограничения по ресурсам:

$$x_1 + 2x_2 + 3x_3 \le 10;$$
$$2x_1 + x_2 + 3x_3 \le 8;$$
$$x_j \ge 0, \quad j = 1, 2, 3.$$

Получили стандартную задачу линейного программирования.

Рассмотрим *второй способ* использования ресурсов, а именно, их продажу.

Интерес предприятия состоит в том, чтобы продать ресурсы по таким ценам, при которых доход от реализации ресурсов будет не меньше прибыли, которую можно получить от реализации продукции, изготовленной из этих ресурсов. В свою очередь покупатель заинтересован в приобретении ресурсов по таким ценам, при которых затраты на покупку будут минимальны.

Задача согласования цен на ресурсы, устраивающих обе стороны может быть описана следующей математической моделью. Пусть y_1 – цена одной единицы ресурса R_1; y_2 – цена одной единицы ресурса R_2. Интерес покупателя будет выражаться целевой функцией, равной суммарной стоимости приобретаемых ресурсов:

$$g(y) = 10y_1 + 8y_2 \to min.$$

Интерес продавца описываться ограничениями:

$$y_1 + 2y_2 \le 6;$$
$$2y_1 + y_2 \le 4;$$
$$y_1 + 3y_2 \le 4,$$

в которых левая часть означает стоимость ресурсов, затраченных на выпуск единицы соответствующей продукции, а правая – удельную прибыль от ее реализации. Дополняя естественные условия неотрицательности цен: $y_1,\ y_2,\ y_3\ \ge 0$, получаем *двойственную задачу ЛП.*

Таким образом, симметричной паре двойственных задач можно придать определенный экономический смысл.

Прямая задача	**Двойственная задача**
Определить такой план выпуска продукции $x = (x_1, x_2, \ldots, x_n)$, используя ограниченные запасы ресурсов, при котором прибыль от реализации продукции будет максимальной.	Установить такой набор цен ресурсов $y = (y_1, y_2, \ldots, y_m)$, при которых стоимость ресурсов, затраченных на выпуск единицы продукции будет не ниже прибыли от ее реализации, но при этом суммарная стоимость затрат будет минимальна.

Двойственные оценки выступают как инструмент балансировки и сопоставления затрат и результатов, они гарантируют рентабельность оптимального плана. Самую высокую ценность имеют те ресурсы, которые в наибольшей степени ограничивают выпуск продукции, прибыль предприятия и на увеличение которых предприятие согласно нести значительные расходы. Ресурс, который предприятие не использует полностью в оптимальном плане, получает нулевую оценку.

Если нужно использовать разнородные ресурсы, например, различные машины, материалы и т. д. для выполнения какой-либо работы, то применяется общий метод линейного программирования, который получил, в соответствии со своей математической основой, название симплекс-метода, предложенного американским ученым Дж. Данцигом.

4.2. Графический метод решения ЗЛП

Графические методы решения в линейном программировании более наглядны, но их применение ограниченно числом неизвестных переменных, которых должно быть не более трех. Рассмотрим методику графического решения основной ЗЛП.

Пример. Для изготовления изделий двух видов имеется 400 кг металла. На одно изделие первого вида расходуется 2 кг металла, а на изделие второго вида – 5 кг. Составить план производства, обеспечивающий наибольшую стоимость выпускаемых изделий, если отпускная цена одного изделия первого вида составляет 4 ед., а изделия второго вида – 3 ед., причем изделий каждого вида требуется изготовить не менее 50 и 20 штук соответственно.

Решение.

1. В качестве параметров управления x_1 и x_2 выберем соответственно количество изготовленных изделий первого и второго вида.

2. Целевая функция должна максимизировать прибыль, получаемую от реализации изделий, поэтому $f(x) = 4x_1 + 3x_2 \to max$.

3. Число выпускаемых изделий в первую очередь ограничено количеством металла (400 кг), которое не должно превысить в процессе производства: $2x_1 + 5x_2 \le 400$. Наличие плана также накладывает ограничение. Изделий первого и второго вида необходимо изготовить не менее 50 и 20 соответственно, т. е. $x_1 \ge 50$, $x_2 \ge 20$. Количество изготовленных изделий принадлежит множеству неотрицательных целых чисел.

Необходимо добавить условие неотрицательности и целочисленности переменных: $x_{1,2} \in N_0$, где N_0 – множество целых неотрицательных чисел, хотя в данной задаче ограничение неотрицательности излишне, так как ограничения $x_1 \ge 50$, $x_2 \ge 20$ являются более строгими. В итоге получим математическую модель ЗЛП:

$$f(x) = 4x_1 + 3x_2 \to max;$$

$$\begin{cases} 2x_1 + 5x_2 \le 400; \\ x_1 \ge 50; \\ x_2 \ge 20; \\ x_{1,2} \in N_0. \end{cases}$$

Аналитическое решение задачи: $f(x) = 660$ ед., $x_1 = 150$, $x_2 = 20$.

Получим решение этой задачи графическим методом.

Определим множество решения первого линейного неравенства $2x_1 + 5x_2 \leq 400$. На плоскости оно отображается полуплоскостью, ограниченной прямой $2x_1 + 5x_2 = 400$. Для построения прямой необходимо найти две точки, принадлежащие прямой. Обычно их находят путем последовательного обнуления каждой из переменных.

Если $x_1 = 0$, то $x_2 = 400/5 = 80$. Если $x_2 = 0$, то $x_1 = 400/2 = 200$. Часто именно точки пересечения прямой с координатными прямыми являются оптимальными решениями.

В системе координат Ox_1x_2 построим прямую по двум точкам (0; 80) и (200; 0). Построенная прямая делит координатную плоскость на две полуплоскости. Для определения искомой полуплоскости достаточно выбрать одну точку, не лежащую на прямой. Пусть это будет начало координат (0; 0). Подставив координаты выбранной точки в неравенство, получим: $0 \leq 400$. Неравенство верное, следовательно, нижняя полуплоскость, которой принадлежит начало координат, является искомой. На рис. 4.1 *а* эта полуплоскость заштрихована.

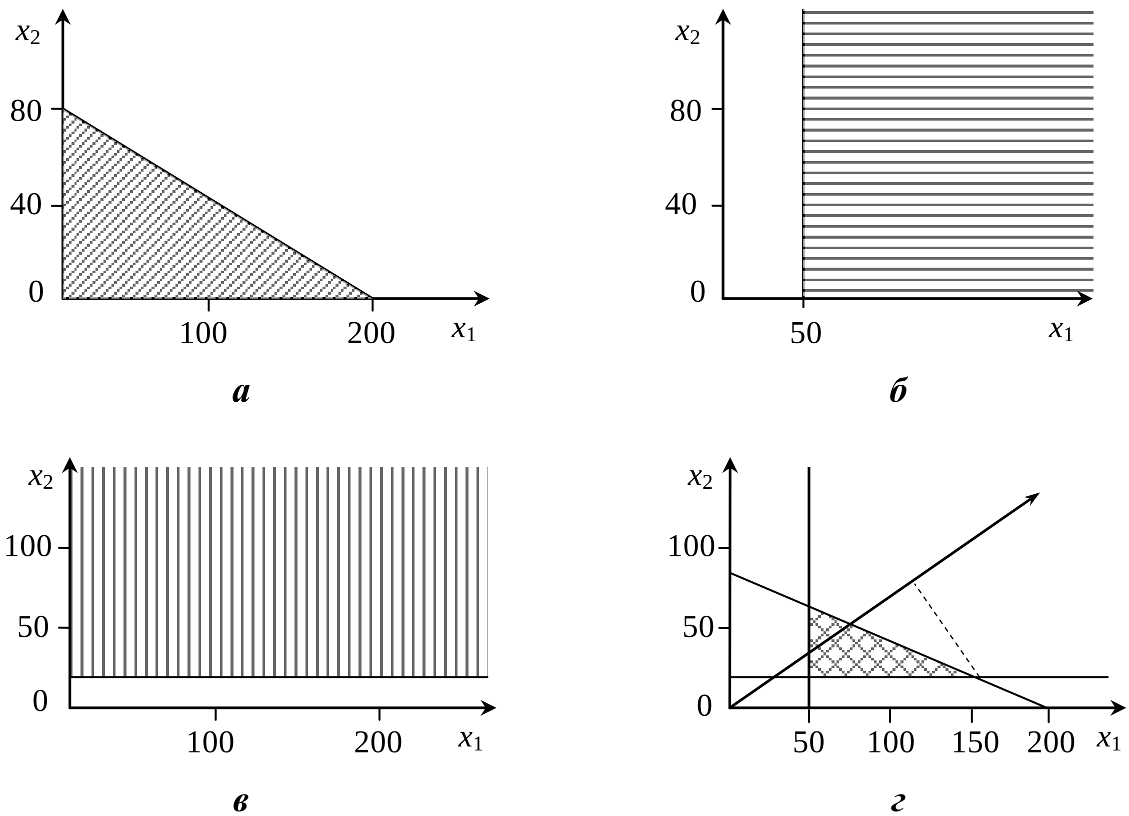


Рис. 4.1. Области допустимых решений

Аналогичная процедура выполняется для второго неравенства. Выберем точку (0; 0) и, подставив во второе неравенство, получим: $0 \geq 50$. Неравенство неверное, следовательно, выбираем правую полуплоскость, которой не принадлежит начало координат (рис. 4.1 *б*).

Для третьего неравенства выбираем точку (0; 0) и, подставив в третье неравенство, получим: $0 \geq 20$. Неравенство неверное, следовательно, выбираем верхнюю полуплоскость, которой не принадлежит начало координат (рис. 4.1 *в*).

В итоге выбираем общую заштрихованную область, точки которой удовлетворяют всем неравенствам задачи (рис. 4.1 *г*). Область допустимых решений, или область *Парето*, изображается в виде замкнутого или незамкнутого многоугольника.

Оптимальное решение ЗЛП находится в одной из вершин многоугольника области Парето.

Возможны два пути нахождения оптимального решения графическим способом:

- находят координаты всех вершин многоугольника области Парето и подставляют их в целевую функцию, после чего выбирают оптимальное значение (наибольшее или наименьшее);
- строят нормальный вектор прямой целевой функции, проецируют область Парето на вектор или его продолжение (см. рис. 4.1 *г*). Наиболее или наименее удаленная от начала координат проекция вершины области Парето на вектор или его продолжение будет оптимальным решением.

Для нахождения трех вершин необходимо решить три системы уравнений:

$$\begin{cases} x_1 = 50; \\ x_2 = 20. \end{cases} \quad \begin{cases} x_1 = 50; \\ 2x_1 + 5x_2 = 400. \end{cases} \quad \begin{cases} x_2 = 50; \\ 2x_1 + 5x_2 = 400. \end{cases}$$

В результате получим точки: $A(50;\,20)$, $B(50;\,60)$, $C(150;\,20)$.

Находим значения целевой функции в вершинах многоугольника Парето:

$$F(50, 20) = 260,\ F(50, 60) = 380,\ F(150, 20) = 660.$$

Наибольшая стоимость выпускаемой продукции составит 660 ед. при изготовлении 150 изделий первого и 20 изделий второго вида.

При графическом решении *вторым способом* находим координаты только одной вершины $C(150;\,20)$, поскольку проекция этой вершины на продолжение вектора целевой функции $\overline{N}(4;3)$ будет наибольшей (см. рис. 4.1 *г*). Второй способ применяется при использовании одинаковых масштабов на осях координат. В противном случае линия проекции на вектор целевой функции визуально не будет составлять 90°.

Если не удается наглядно изобразить область Парето в системе координат с одинаковым масштабом осей координат, то поступают следующим способом. В системе координат с разным масштабом координатных осей рисуют линию уровня целевой функции $f(x)=0$ или $f(x)=const$ и перемещают ее параллельно в направлении вектора целевой функции до пересечения с областью Парето.

В процессе решения задачи использовано понятие «*область и оптимальность по Парето*». Рассмотрим его подробнее.

Пусть функции $f_1, f_2,..., f_K$ достигают максимум в одной и той же точке $X^* \in D$, в таком случае говорят, что задача имеет *идеальное решение*. Случаи существования идеального решения в многокритериальной задаче крайне редки. Поэтому основная проблема при рассмотрении таких задач – формализация *принципа оптимальности*, т. е. определение того, в каком смысле «оптимальное» решение лучше других. В случае отсутствия «идеального решения» в задаче *ищется компромиссное решение.*

Для всякой альтернативы $X \in D$ вектор из значений целевых функций $(f_1(X), f_2(X),..., f_K(X))$ является *векторной оценкой* альтернативы X. Векторная оценка альтернативы содержит полную информацию о ценности (полезности) этой альтернативы для главного конструктора системы, или, как принято говорить в системном анализе, лица, принимающего решение (ЛПР). Сравнение любых двух исходов заменяется сравнением их векторных оценок.

Пусть $X_1, X_2 \in D$. Если для всех критериев $f_1, f_2,..., f_K$ имеют место неравенства $f_k(X_2) \geq f_k(X_1)$, $k = 1, 2,..., K$, причем хотя бы одно неравенство строгое, то говорят, что решение X_2 *предпочтительнее* решения X_1. Условие предпочтительности принято обозначать в виде $X_2 > X_1$.

В задаче точка $X_0 \in D$ называется *оптимальной по Парето*, если не существует другой точки $X \in D$, которая была бы предпочтительнее, чем X_0. Точки, оптимальные по Парето, образуют множество точек, оптимальных по Парето (множество неулучшаемых или эффективных точек) $D_p \subset D$.

Оптимальные решения многокритериальной задачи следует искать только среди элементов множества альтернатив D_p. В этой области ни один критерий не может быть улучшен без ухудшения хотя бы одного из других. Важным свойством множества Парето D_p является возможность «выбраковывать» из множества альтернатив D заведомо неудачные, уступающие другим по всем критериям. Обычно решение многокритериальной задачи должно начинаться с выделения множества D_p. При отсутствии дополнительной информации о системе предпочтений ЛПР должно принимать решение именно из множества Парето D_p.

В векторной оптимизации кроме множества Парето в общем случае нет общих правил, по которому варианту X_2 отдается предпочтение по сравнению с другим вариантом X_1. Часто решение многокритериальной задачи состоит в построении множества Парето-оптимальных точек и дальнейшем выборе одной из них на основе «здравого смысла» или с помощью какого-либо другого критерия.

Во всех случаях задача многокритериальной оптимизации одним из способом сводится к задаче с одним критерием. Существует много способов построения такого окончательного критерия, однако ни одному из них нельзя заранее отдать наибольшее предпочтение. Заметим, что целевые функции отображают множество точек, оптимальных по Парето $D_p \subset D \subset R^n$ в множество $F_p \subset F \subset R^K$, которое называется *множеством Парето*.

Пример. Для возведения дорожного полотна двух видов *A* и *B* используется бульдозеры и скреперы. Затраты времени на подготовку 1 км для каждого вида полотна, общий фонд рабочего времени, а также срок службы эксплуатации одного полотна каждого вида составляют:

Тип строительной техники:	Затраты времени на подготовку полотна вида		Общий фонд рабочего времени оборудования, *ч*
	A	*B*	
бульдозеры	2	5	200
скреперы	4	2	160
Срок службы, ед.	10	24	*max*

Требуется определить, сколько техники каждого вида следует использовать строительному подразделению, чтобы последующий срок службы полотна был максимальным.

Решение.

1. В качестве параметров управления x_1 и x_2 выберем количество используемой техники на обустройстве полотна первого и второго вида соответственно.

2. Целевая функция должна максимизировать срок службы дорожного полотна, поэтому $f(x) = 10x_1 + 24x_2 \to max$.

3. Количество используемых видов техники ограничено общим фондом рабочего времени. В процессе производства нельзя превысить общий фонд

рабочего времени для бульдозеров $2x_1 + 5x_2 \le 200$ и общий фонд рабочего времени для скреперов $4x_1 + 2x_2 \le 160$. Необходимо добавить условие неотрицательности и целочисленности переменных: $x_{1,2} \in N_0$, где N_0 – множество целых неотрицательных чисел.

Получим математическую модель задачи:

$$f(x) = 10x_1 + 24x_2 \to max;$$

$$\begin{cases} 2x_1 + 5x_2 \le 200, & (1); \\ 4x_1 + 2x_2 \le 160, & (2); \\ x_{1,2} \in N_0. \end{cases}$$

Изобразим область допустимых решений в системе координат Ox_1x_2.

Найдем две точки, лежащие на прямой $2x_1 + 5x_2 = 200$. Пусть $x_1 = 0$, тогда из данного уравнения имеем: $x_2 = 200/5 = 40$. Пусть $x_2 = 0$, тогда из уравнения имеем $x_1 = 200/2 = 100$.

В итоге получим координаты двух точек прямой (1): *A*(0; 40), *B*(100; 0). В системе координат Ox_1x_2 построим прямую по точкам *A* и *B* (рис. 4.2).

Аналогично для другой прямой $4x_1 + 2x_2 = 160$. Пусть $x_1 = 0$, тогда $x_2 = 160/2 = 80$. Пусть $x_2 = 0$, тогда $x_1 = 160/4 = 40$.

В итоге имеем координаты двух точек прямой (2): *C*(0; 80), *D*(40; 0). В системе координат Ox_1x_2 построим прямую по точкам *C* и *D*.

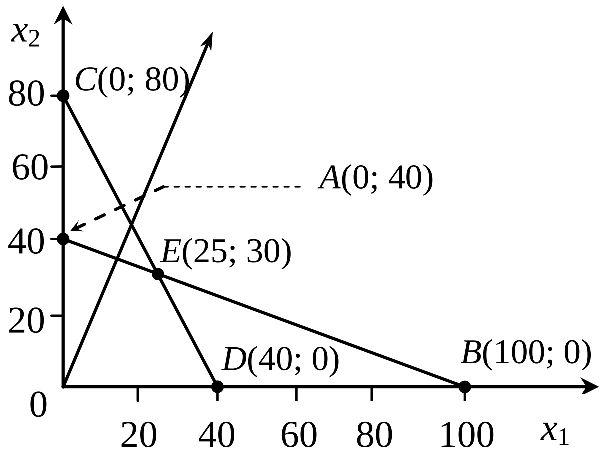


Рис. 4.2. Построение области Парето

В результате получили три замкнутые области и одну незамкнутую. Для определения области допустимых решений возьмем контрольную точку с координатами (0; 0), принадлежащую только четырехугольнику *OAED*. Координаты этой точки удовлетворяют всем ограничениям. Следовательно, выбранный четырехугольник *OAED* является искомой областью Парето.

Контрольной точкой может быть не только вершина многоугольника, но и любая внутренняя точка полученных многоугольников.

Найдем координаты точки пересечения прямых (1) и (2) из системы уравнений:

$$\begin{cases} 2x_1 + 5x_2 = 200; \\ 4x_1 + 2x_2 = 160. \end{cases}$$

Разделим второе уравнение системы на два и вычтем из второго уравнения первое:

$$\begin{cases} 2x_1 + 5x_2 = 200; \\ -4x_2 = -120. \end{cases}$$

Разделим второе уравнение на (– 4):

$$\begin{cases} 2x_1 + 5x_2 = 200; \\ x_2 = 30. \end{cases}$$

Подставим в первое уравнение численное значение $x_2 = 30$. Получим координаты точки E(25; 30) пересечения прямых AB и CD. Координаты точек пересечения построенных прямых с осями координат найдены при построении прямых. Найдем значения целевой функции в вершинах четырехугольника O(0; 0), A(0; 40), E(25; 30) и D(40; 0):

F(0, 0) = 0; F(0, 40) = 960; E(25, 30) = 970; F(40, 0) = 400.

Наибольший срок службы составит 970 ед. при использовании 25 бульдозеров на строительстве полотна типа A и 30 скреперов на строительстве типа B.

Рассмотрим пример нахождения нецелочисленного решения задачи линейного программирования графическим методом:

$$F = 2x_1 - x_2 \to max;$$

$$x_1 - x_2 \geq -3;$$

$$6x_1 + 7x_2 \leq 42;$$

$$2x_1 - 3x_2 \leq 6;$$

$$x_1 + x_2 \geq 4;$$

$$x_1,\ x_2 \geq 0.$$

Шаг № 1. Построим область допустимых решений, т. е. решим графически систему неравенств. Для этого построим каждую прямую и определим полуплоскости, заданные неравенствами, знаки неравенств обозначены штрихом на прямых (рис. 4.3).

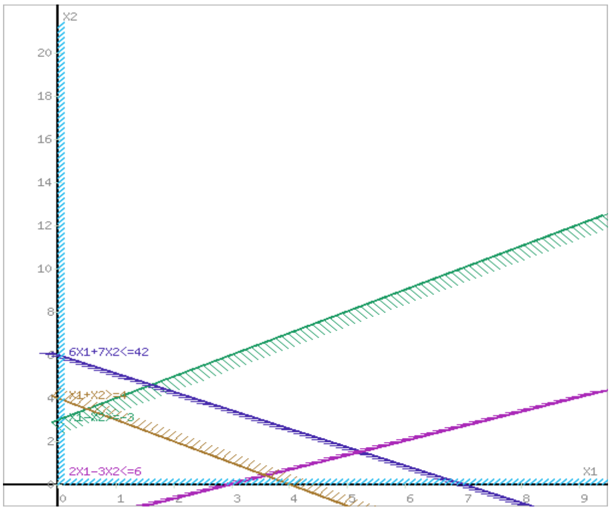


Рис. 4.3. Построение области ограничений

Шаг № 2. Определение границы области допустимых решений. Пересечением полуплоскостей будет являться область, координаты точек которого удовлетворяют условию неравенствам системы ограничений задачи. Обозначим границы области многоугольника решений (рис. 4.4).

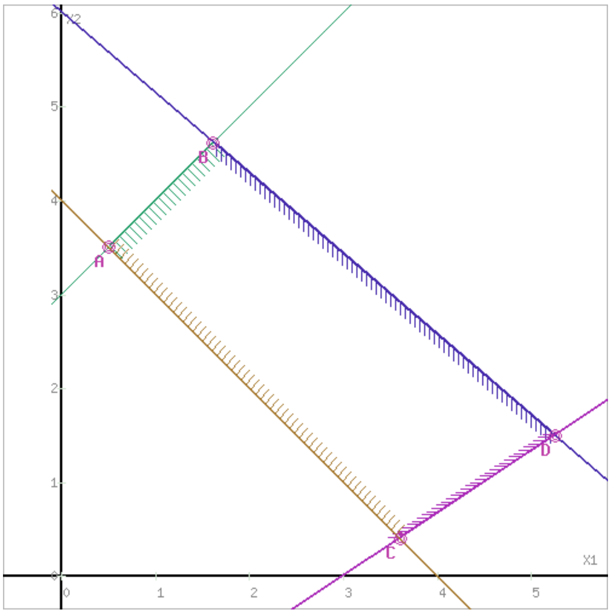


Рис. 4.4. Границы области Парето

Шаг № 3. Рассмотрим целевую функцию задачи $F(x_1, x_2) = 2x_1 - x_2 \to max$. Построим прямую, отвечающую значению функции $F = 0$: $F = 2x_1 - x_2 = 0$. Вектор-градиент, составленный из коэффициентов целевой функции, указывает направление максимума функции $F(x_1, x_2)$. Начало вектора – точка (0; 0), конец – точка (2; – 1). Поскольку по условию задачи интересует максимальное решение, поэтому двигаем прямую до последней точки касания обозначенной области. На графике эта прямая обозначена пунктирной линией (рис. 4.5).

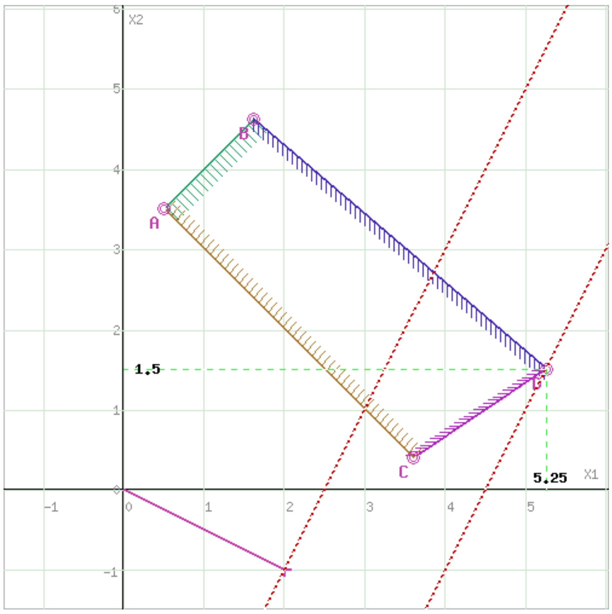


Рис. 4.5. Нахождение «крайней точки» на области решений

Прямая пересекает область в точке D. Так как точка D получена в результате пересечения прямых (2) и (3) из условий ограничений, то ее координаты удовлетворяют уравнениям этих прямых:

$$\begin{cases} 6x_1 + 7x_2 = 42; \\ 2x_1 - 3x_2 = 6. \end{cases}$$

Получим: $x_1 = 5.25$, $x_2 = 1.5$, значение целевой функции: $F(x_1, x_2) = 9$.

4.3. Симплексный метод

Симплексный метод относится к аналитическим методам решения задач линейного программирования. Он представляет собой последовательность шагов (итераций) перехода от одной вершины многогранника области Парето к другой с целью получения максимального значения целевой функции. ЗЛП любого типа могут быть решены симплекс-методом.

Возьмем в качестве примера математическую модель одну из рассмотренных задач без учета ограничений целочисленности переменных:

$$f(x) = 10x_1 + 24x_2 \to max;$$

$$\begin{cases} 2x_1 + 5x_2 \le 200, \quad (1); \\ 4x_1 + 2x_2 \le 160, \quad (2); \\ x_1 \ge 0, \quad x_2 \ge 0. \end{cases}$$

Добавим в 1-е и 2-е ограничения соответственно неотрицательные переменные x_3 и x_4, и запишем целевую функцию в неявном виде $F(x)=0$:

$$f(x)-10x_1-24x_2=0;$$

$$\begin{cases} 2x_1+5x_2+x_3=200; \\ 4x_1+2x_2+x_4=160; \\ x_1 \ge 0, \quad j=\overline{1,4}. \end{cases}$$

Каждый коэффициент целевой функции $F(x)$, если он отрицателен, определяет величину положительного приращения целевой функции $f(x)$, а если положителен – отрицательного приращения целевой функции $f(x)$ при увеличении соответствующей переменной. Если x_1 увеличить на единицу, то целевая функция $f(x)$ увеличится на 10 единиц, а если x_2 увеличить на единицу, то целевая функция $f(x)$ увеличится на 24 единицы. Переменная x_2 сильнее влияет на целевую функцию, чем переменная x_1.

На первом шаге симплексного метода решения канонической ЗЛП необходимо попасть в одну из вершин многогранника области Парето. Выражаем из m ограничений канонической ЗЛП m переменных и разрешаем систему из m уравнений относительно выбранных m переменных. Исключаем эти m переменных из выражения целевой функции. Выбранные m переменные называют базисными или основными, а остальные – небазисными или свободными.

В рассматриваемой задаче в качестве базисных переменных удобно взять переменные x_3 и x_4. Они не входят в целевую функцию и относительно них легко разрешить систему двух уравнений:

$$f(x)-10x_1-24x_2=0;$$

$$\begin{cases} x_3=200-2x_1-5x_2; \\ x_4=160-4x_1-2x_2; \\ x_j \ge 0, \quad j=\overline{1,4}. \end{cases}$$

Первоначальное допустимое решение получается при нулевых значениях свободных переменных $x_1=x_2=0$, $x_3=200$, $x_4=160$, $f(x)=0$.

На втором шаге необходимо перейти в другую вершину многогранника области Парето, в которой целевая функция будет принимать большее значение. Аналитически это означает замену одной базисной переменной на другую, свободную переменную. Выбор новой базисной переменной осуществляют по следующим правилам.

Если в целевой функции имеются свободные переменные, коэффициенты при которых отрицательные, то желательно выбрать из них переменную с наибольшим по модулю значением коэффициента в качестве новой базисной переменной. Если все свободные переменные имеют неотрицательные коэффициенты, то оптимальное решение получено и улучшить его нельзя.

Свободная переменная x_2 имеет наибольший по модулю отрицательный коэффициент в целевой функции. Выберем в качестве новой базисной переменной x_2 и разрешим систему ограничений относительно базисных переменных x_2 и x_4, исключим новую базисную переменную x_2 из целевой функции. Эти преобразования удобно выполнять в виде симплекс-таблицы (табл. 4.1), которая составлена на основе канонической ЗЛП с целевой функцией $F(x)$, записанной в неявном виде.

Таблица 4.1

x_1	x_2	x_3	x_4	Свободные члены	Базис
2	5	1	0	200	x_3
4	2	0	1	160	x_4
– 10	– 24	0	0	0	$F(x)$

Столбцы при переменных x_3 и x_4 являются единичными. Единичный столбец – это столбец, все элементы которого равны нулю, кроме одного, равного единице. В симплекс-таблице количество единичных столбцов должно быть равно количеству базисных переменных. Все свободные члены должны быть неотрицательными. Неотрицательность свободных членов получают при составлении канонической ЗЛП.

Рассмотрим алгоритм ввода новой базисной переменной и исключения старой, реализованный симплекс-таблицей.

1. Выбираем наибольший по модулю отрицательный коэффициент целевой функции – это коэффициент при переменной x_2. Столбец $l = 2$ симплекс-таблицы, которому принадлежит новая базисная переменная x_2, называется ведущим столбцом.

2. Делим элементы столбца свободных членов на соответствующие элементы ведущего столбца, при этом учитываем только положительные элементы. Выбираем наименьшее частное $min\left\{\frac{200}{5}; \frac{160}{2}\right\} = min\{40; 80\}$. Строка $k = 1$, которой принадлежит наименьшее частное, называется ведущей строкой. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{12} = 5$. Первая строка симплекс-таблицы не нумеруется.

3. Делим элементы ведущей строки на ведущий элемент:

$$a'_{kj} = \frac{a_{kj}}{a_{kl}},$$

где k – номер ведущей строки;

l – номер ведущего столбца;

a_{kl} – ведущий элемент.

Все остальные элементы симплекс таблицы пересчитываем по формуле:

$$a'_{ij} = a_{ij} - \frac{a_{kj}a_{il}}{a_{kl}} \quad \text{или} \quad a'_{ij} = a_{ij} - a'_{kj}a_{il}.$$

Элементы ведущего столбца, кроме ведущего элемента, обнуляются:

$$a'_{il} = a_{il} - \frac{a_{kl}a_{il}}{a_{kl}} = 0.$$

Первая строка (ведущая):

$$a'_{11} = \frac{a_{11}}{a_{12}} = \frac{2}{5} = 0.4; \qquad a'_{12} = \frac{a_{12}}{a_{12}} = \frac{5}{5} = 1; \qquad a'_{13} = \frac{a_{13}}{a_{12}} = \frac{1}{5} = 0.2;$$

$$a'_{14} = \frac{a_{14}}{a_{12}} = \frac{0}{5} = 0, \qquad a'_{15} = \frac{a_{15}}{a_{12}} = \frac{200}{5} = 40.$$

Вторая строка:

$$a'_{21} = a_{21} - a'_{11}a_{22} = 4 - 0.4 \cdot 2 = 3.2;$$
$$a'_{22} = 0;$$
$$a'_{23} = a_{23} - a'_{13}a_{22} = 0 - 0.2 \cdot 2 = -0.4;$$
$$a'_{24} = a_{24} - a'_{14}a_{22} = 1 - 0 \cdot 2 = 1;$$
$$a'_{25} = a_{25} - a'_{15}a_{22} = 160 - 40 \cdot 2 = 80.$$

Третья строка:

$$a'_{31} = a_{31} - a'_{11}a_{32} = -10 - 0.4 \cdot (-24) = -4.4;$$
$$a'_{32} = 0;$$
$$a'_{33} = a_{33} - a'_{13}a_{32} = 0 - 0.2 \cdot (-24) = 4.8;$$
$$a'_{34} = a_{34} - a'_{14}a_{32} = 0;$$
$$a'_{35} = a_{35} - a'_{15}a_{32} = 0 - 40 \cdot (-24) = 480.$$

В столбце «Базис» старую базисную переменную ведущей строки x_3 заменяем на новую базисную переменную ведущего столбца x_2 (табл. 4.2).

Таблица 4.2

x_1	x_2	x_3	x_4	Свободные члены	Базис
0.4	1	0.2	0	40	x_2
3.2	0	– 0.4	1	80	x_4
– 4.4	0	4.8	0	480	$F(x)$

Первый итерационный шаг симплексного метода завершен. Получили новую симплекс-таблицу (см. табл. 4.2) и новый опорный план (0; 40; 0; 80), $f(x) = 480$.

Первый шаг соответствует переходу от одной вершины $O(0; 0)$ многоугольника области Парето к другой $A(0; 40)$.

Коэффициент при переменной x_1 в целевой функции отрицателен, т. е. оптимальное значение еще не достигнуто. Повторяем шаги 1–3 еще раз.

1. Выбираем наибольший по модулю отрицательный коэффициент целевой функции – это коэффициент при переменной x_1. Столбец $l = 1$ симплекс-таблицы, которому принадлежит новая базисная переменная x_1, является ведущим столбцом.

2. Делим элементы столбца свободных членов на соответствующие элементы ведущего столбца, при этом учитываем только положительные элементы. Выбираем наименьшее частное $min\left\{\frac{40}{0.4}; \frac{160}{3.2}\right\} = min\{100; 25\}$. Строка $k = 2$, которой принадлежит наименьшее частное, является ведущей строкой. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{21} = 3.2$.

3. Пересчитываем элементы симплекс-таблицы по формулам, приведенным на первом шаге. Начинаем с ведущей строки.

$$a'_{21} = \frac{a_{21}}{a_{21}} = 1; \qquad a'_{22} = \frac{a_{22}}{a_{21}} = \frac{0}{3.2} = 0; \qquad a'_{23} = \frac{a_{23}}{a_{21}} = \frac{-0.4}{3.2} = -0.125;$$

$$a'_{24} = \frac{a_{24}}{a_{21}} = \frac{1}{3.2} = 0.3125; \quad a'_{25} = \frac{a_{25}}{a_{21}} = \frac{80}{3.2} = 25.$$

Первая строка:

$$a'_{11} = 0;$$

$$a'_{12} = a_{12} - a'_{22}a_{11} = 1 - 0 \cdot 0.4 = 1;$$

$$a'_{13} = a_{13} - a'_{23}a_{11} = 0.2 - (-0.125) \cdot 0.4 = 0.25;$$

$$a'_{14} = a_{14} - a'_{24}a_{11} = 0 - 0.3125 \cdot 0.4 = -0.125;$$

$$a'_{15} = a_{15} - a'_{25}a_{11} = 40 - 25 \cdot 0.4 = 30.$$

Третья строка:

$$a'_{31} = 0;$$
$$a'_{32} = a_{32} - a'_{22}a_{13} = 0 - 0 \cdot (-4.4) = 0;$$
$$a'_{33} = a_{33} - a'_{23}a_{13} = 4.8 - (-0.125) \cdot (-4.4) = 4.25;$$
$$a'_{34} = a_{34} - a'_{24}a_{13} = 0 - 0.3125 \cdot (-4.4) = 1.375;$$
$$a'_{35} = a_{35} - a'_{25}a_{13} = 480 - 25 \cdot (-4.4) = 590.$$

В столбце «Базис» прежнюю базисную переменную ведущей строки x_4 заменяем на новую базисную переменную ведущего столбца x_1 (табл. 4.3).

Таблица 4.3

x_1	x_2	x_3	x_4	Свободные члены	Базис
0	1	0.25	– 0.125	30	x_1
1	0	– 0.125	0.3125	25	x_2
0	0	4.25	1.375	590	$F(x)$

Второй шаг симплексного метода завершен. Получили новую симплекс-таблицу (см. табл. 1.3) и новый опорный план (25; 30; 0; 0), $f(x) = 590$. Второй итерационный шаг соответствует переходу от вершины A(0; 40) многоугольника к другой его вершине B(25; 30).

Коэффициенты целевой функции неотрицательны, поэтому новый опорный план (25; 30; 0; 0), $f(x) = 590$ является оптимальным.

Рассмотрим решение второго примера симплексным методом.

Запишем математическую модель в каноническом виде без учета целочисленности переменных x_1 и x_2, добавив три неотрицательные переменные x_3, x_4 и x_5:

$$f(x) - 4x_1 - 3x_2 = 0;$$
$$\begin{cases} 2x_1 + 5x_2 + x_3 = 400; \\ x_1 - x_4 = 50; \\ x_2 - x_5 = 20; \\ x_j \geq 0, \quad j = \overline{1, 5}. \end{cases}$$

Коэффициенты при переменных x_4 и x_5 отрицательны. При составлении симплекс-таблицы они не составят единичных столбцов, поэтому введем две искусственные неотрицательные переменные x_6 и x_7:

$$f(x) - 4x_1 - 3x_2 = 0;$$

$$\begin{cases} 2x_1 + 5x_2 + x_3 = 400; \\ x_1 - x_4 + x_6 = 50; \\ x_2 - x_5 + x_7 = 20; \\ x_j \geq 0,\ j = \overline{1,7}. \end{cases}$$

Искусственные переменные не входят в целевую функцию. Составим симплекс-таблицу (табл. 4.4).

Таблица 4.4

x_1	x_2	x_3	x_4	x_5	x_6	x_7	Свободные члены	Базис
2	5	1	0	0	0	0	400	x_3
1	0	0	– 1	0	1	0	50	x_6
0	1	0	0	– 1	0	1	20	x_7
– 4	– 3	0	0	0	–	–	0	$F(x)$

Выводим пошагово из базиса искусственные переменные x_6 и x_7. Геометрически это означает попадание в область Парето.

1. Выбираем вторую строку $k = 2$ при переменной x_6 в качестве ведущей.

2. Выбираем первый столбец $l = 1$ при переменной x_1 в качестве ведущего. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{21} = 1$.

3. Пересчитываем элементы симплекс-таблицы по формулам, приведенным в предыдущей задаче. Начинаем с ведущей строки.

Вторая строка (ведущая) остается неизменной:

$a'_{21} = 1$; $a'_{22} = 0$; $a'_{23} = 0$; $a'_{24} = -1$; $a'_{25} = 0$; $a'_{26} = 1$; $a'_{27} = 0$; $a'_{28} = 50$.

Первая строка (умножаем ведущую строку на два и результат вычитаем из первой строки):

$a'_{11} = 0$; $a'_{12} = 5$; $a'_{13} = 1$; $a'_{14} = 2$; $a'_{15} = 0$; $a'_{16} = -2$; $a'_{17} = 0$; $a'_{18} = 300$.

Третья строка остается неизменной:

$a'_{31} = 0$; $a'_{32} = 1$; $a'_{33} = 0$; $a'_{34} = 0$; $a'_{35} = -1$; $a'_{36} = 0$; $a'_{37} = 1$; $a'_{38} = 20$.

Четвертая строка (умножаем ведущую строку на четыре и результат прибавляем к четвертой строке):

$a'_{41} = 0$; $a'_{42} = -3$; $a'_{43} = 0$; $a'_{44} = -4$; $a'_{45} = 0$; $a'_{48} = 200$.

В столбце «Базис» прежнюю базисную переменную ведущей строки x_6 заменяем на новую базисную переменную ведущего столбца x_1. Вычеркиваем столбец выведенной искусственной переменной x_6 (табл. 4.5).

Таблица 4.5

x_1	x_2	x_3	x_4	x_5	x_6	x_7	Свободные члены	Базис
0	5	1	2	0		0	300	x_3
1	0	0	– 1	0		0	50	x_1
0	1	0	0	– 1		1	20	x_7
0	– 3	0	– 4	0		–	200	$F(x)$

Выводим из базиса искусственную переменную x_7.

1. Выбираем третью строку $k = 3$ при переменной x_7 в качестве ведущей.

2. Выбираем второй столбец $l = 2$ при переменной x_2 в качестве ведущего. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{32} = 1$.

3. Пересчитываем элементы симплекс-таблицы (см. табл. 4.5).

Третья строка (ведущая) остается неизменной:

$a'_{31} = 0$; $a'_{32} = 1$; $a'_{33} = 0$; $a'_{34} = 0$; $a'_{35} = -1$; $a'_{37} = 1$; $a'_{38} = 20$.

Первая строка (умножаем ведущую строку на пять и результат вычитаем из первой строки):

$a'_{11} = 0$; $a'_{12} = 0$; $a'_{13} = 1$; $a'_{14} = 2$; $a'_{15} = 5$; $a'_{17} = -5$; $a'_{18} = 200$.

Вторая строка остается неизменной:

$a'_{21} = 1$; $a'_{22} = 0$; $a'_{23} = 0$; $a'_{24} = -1$; $a'_{25} = 0$; $a'_{27} = 0$; $a'_{28} = 50$.

Четвертая строка (умножаем ведущую строку на три и результат прибавляем к четвертой строке):

$a'_{41} = 0$; $a'_{42} = 0$; $a'_{43} = 0$; $a'_{44} = -4$; $a'_{45} = -3$; $a'_{48} = 260$.

В столбце «Базис» прежнюю базисную переменную ведущей строки x_7 заменяем на новую базисную переменную ведущего столбца x_2. Вычеркиваем столбец выведенной искусственной переменной x_7 (табл. 4.6).

Таблица 4.6

x_1	x_2	x_3	x_4	x_5	Свободные члены	Базис
0	0	1	2	5	200	x_3
1	0	0	– 1	0	50	x_1
0	1	0	0	– 1	20	x_2
0	0	0	– 4	– 3	260	$F(x)$

Искусственные базисные переменные x_6 и x_7 выведены из базиса. Геометрически это означает попадание в одну из вершин многоугольника области Парето. Если не удается вывести искусственные переменные из базиса или получена неразрешимая задача, то исходная задача также неразрешима.

Далее используем стандартный алгоритм симплексного метода.

1. Выбираем наибольший по модулю отрицательный коэффициент целевой функции – это коэффициент при переменной x_4. Столбец $l=4$ симплекс-таблицы, которому принадлежит новая базисная переменная x_4, является ведущим столбцом.

2. Делим элементы столбца свободных членов на соответствующие элементы ведущего столбца, при этом учитываем только положительные элементы. Выбираем наименьшее частное $min\left\{\frac{200}{2}\right\}=100$. Строка $k=1$, которой принадлежит наименьшее частное, является ведущей строкой. На пересечении ведущего столбца и ведущей строки находится ведущий элемент $a_{14}=2$.

3. Пересчитываем элементы симплекс-таблицы (см. табл. 4.6). Начинаем с ведущей строки.

Первая строка (делим каждый элемент ведущей строки на ведущий элемент):

$a'_{11}=0$; $a'_{12}=0$; $a'_{13}=0.5$; $a'_{14}=1$; $a'_{15}=2.5$; $a'_{16}=100$.

Вторая строка (первую строку a'_{1j} прибавляем ко второй):

$a'_{21}=1$; $a'_{22}=0$; $a'_{23}=0.5$; $a'_{24}=0$; $a'_{25}=2.5$; $a'_{26}=150$.

Третья строка остается неизменной:

$a'_{31}=0$; $a'_{32}=1$; $a'_{33}=0$; $a'_{34}=0$; $a'_{35}=-1$; $a'_{36}=20$.

Четвертая строка (умножаем первую строку a'_{1j} на четыре и результат прибавляем к четвертой строке):

$a'_{41}=0$; $a'_{42}=0$; $a'_{43}=2$; $a'_{44}=0$; $a'_{45}=7$; $a'_{46}=660$.

В столбце «Базис» прежнюю базисную переменную ведущей строки x_3 заменяем на новую базисную переменную ведущего столбца x_4 (табл. 4.7).

Таблица 4.7

x_1	x_2	x_3	x_4	x_5	Свободные члены	Базис
0	0	0.5	1	2.5	100	x_4
1	0	0.5	0	2.5	150	x_1
0	1	0	0	– 1	20	x_2
0	0	2	0	7	660	$F(x)$

Коэффициенты целевой функции неотрицательны, поэтому новый опорный план (150; 20; 0; 100; 0), $f(x) = 660$ является оптимальным.

Рассмотрим симплекс-метод получения нецелочисленного решения ЗЛП. Определим минимальное значение целевой функции $F(x_1, x_2) = 5x_1 - 3x_2$ следующих ограничениях:

$$3x_1 + 2x_2 \geq 6;$$
$$-2x_1 + 3x_2 \leq 6;$$
$$x_1 - x_2 \leq 4;$$
$$4x_1 + 7x_2 \leq 28.$$

Для построения первого опорного плана систему неравенств приведем к системе уравнений путем введения дополнительных переменных (*переход к канонической форме)*.

$$3x_1 + 2x_2 - 1x_3 + 0x_4 + 0x_5 + 0x_6 = 6;$$
$$-2x_1 + 3x_2 - 0x_3 + 1x_4 + 0x_5 + 0x_6 = 6;$$
$$1x_1 - 1x_2 + 0x_3 + 0x_4 + 1x_5 + 0x_6 = 4;$$
$$4x_1 + 7x_2 + 0x_3 + 0x_4 + 0x_5 + 1x_6 = 28.$$

Введем *искусственные переменные* x: в 1-м равенстве вводим переменную x_7; а целевую функцию представим виде:

$$F(X) = Mx_7 \to min.$$

Из уравнений выражаем искусственную переменную: $x_7 = 6 - 3x_1 - 2x_2 + x_3$, которую подставим в целевую функцию: $F(X) = M(6 - 3x_1 - 2x_2 + x_3) \to min$, иначе:

$$F(X) = (-3M)x_1 + (-2M)x_2 + (M)x_3 + (6M) \to min.$$

Введем новую переменную $x_0 = -3x_1 - 2x_2$.

Выразим базисные переменные <7, 4, 5, 6> через небазисные:

$$x_0 = 6 - 3x_1 - 2x_2 + x_3;$$
$$x_7 = 6 - 3x_1 - 2x_2 + x_3;$$
$$x_4 = 6 + 2x_1 - 3x_2;$$
$$x_5 = 4 - x_1 + x_2;$$
$$x_6 = 28 - 4x_1 - 7x_2.$$

Переходим к основному алгоритму симплекс-метода.

Поскольку задача решается на минимум, то переменную для включения в текущий план выбирают по минимальному отрицательному числу в уравнении для x_0.

Проверка критерия оптимальности. В выражении для x_0 присутствуют положительные элементы. Следовательно, полученный план не оптимален.

Определение новой базисной переменной, как $max(-3,-2,1,0,0,0,0)=-3$:

$$x_0 = 6-3x_1-2x_2+x_3;$$
$$x_7 = 6-3x_1-2x_2+x_3;$$
$$x_4 = 6+2x_1-3x_2;$$
$$x_5 = 4-x_1+x_2;$$
$$x_6 = 28-4x_1-7x_2.$$

В качестве новой переменной выбираем x_1. Вычислим значения D_i по всем уравнениям для этой переменной: b_i/a_{i1} из них выберем наименьшее:

$$min(6:3,-,4:1,28:4)=2.$$

Вместо переменной x_7 в план войдет переменная x_1. Выразим переменную x_1 через x_7, то есть $x_1 = 2-\frac{2}{3}x_2+\frac{1}{3}x_3-\frac{1}{3}x_7$ и подставим во все выражения:

$$x_0 = 6-3\left(2-\frac{2}{3}x_2+\frac{1}{3}x_3-\frac{1}{3}x_7\right)-2x_2+x_3;$$
$$x_4 = 6+2\left(2-\frac{2}{3}x_2+\frac{1}{3}x_3-\frac{1}{3}x_7\right)-3x_2;$$
$$x_5 = 4-\left(2-\frac{2}{3}x_2+\frac{1}{3}x_3-\frac{1}{3}x_7\right)+x_2;$$
$$x_6 = 28-4\left(2-\frac{2}{3}x_2+\frac{1}{3}x_3-\frac{1}{3}x_7\right)-7x_2.$$

После приведения всех подобных, получаем новую систему, эквивалентную прежней:

$$x_0 = 0+x_7;$$
$$x_1 = 2-\frac{2}{3}x_2+\frac{1}{3}x_3-\frac{1}{3}x_7;$$
$$x_4 = 10-\frac{13}{3}x_2+\frac{2}{3}x_3-\frac{2}{3}x_7;$$
$$x_5 = 2+\frac{5}{3}x_2-\frac{1}{3}x_3+\frac{1}{3}x_7;$$
$$x_6 = 20-\frac{13}{3}x_2-\frac{4}{3}x_3+\frac{4}{3}x_7.$$

Полагая небазисные переменные $x=(1,4,5,6)$ равными нулю, получим новый допустимый вектор и значение целевой функции:

$$x=(0,0,0,0,0,0,-1),\ x_0 = 0.$$

Выражение для x_0 не содержит отрицательных элементов. Найден оптимальный план:

$$x_0 = 0 + x_7;$$
$$x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3 - \frac{1}{3}x_7;$$
$$x_4 = 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3 - \frac{2}{3}x_7;$$
$$x_5 = 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3 + \frac{1}{3}x_7;$$
$$x_6 = 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3 + \frac{4}{3}x_7.$$

На этом первый этап симплекс-метода завершен. Переходим ко *второму этапу*. Удаляем элементы с искусственными переменными.

$$x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3;$$
$$x_4 = 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3;$$
$$x_5 = 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3;$$
$$x_6 = 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3.$$

Выразим базисные переменные: $x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3$, которые подставим в целевую функцию: $F(X) = 5\left(2 - \frac{2}{3}x_2 + \frac{1}{3}x_3\right) - 3x_2$ или $F(X) = 10 - \frac{19}{3}x_2 + \frac{5}{3}x_3$.

Получаем новую систему переменных.

$$x_0 = 10 - \frac{19}{3}x_2 + \frac{5}{3}x_3;$$
$$x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3;$$
$$x_4 = 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3;$$
$$x_5 = 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3;$$
$$x_6 = 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3.$$

Проверка критерия оптимальности. В выражении для x_0 присутствуют положительные элементы. Следовательно, полученный план еще не оптимален.

Определение новой базисной переменной:

$$max\left(0, -\frac{19}{3}, \frac{5}{3}, 0, 0, 0\right) = -\frac{19}{3};$$

$$x_0 = 10 - \frac{19}{3}x_2 + \frac{5}{3}x_3;$$

$$x_1 = 2 - \frac{2}{3}x_2 + \frac{1}{3}x_3;$$

$$x_4 = 10 - \frac{13}{3}x_2 + \frac{2}{3}x_3;$$

$$x_5 = 2 + \frac{5}{3}x_2 - \frac{1}{3}x_3;$$

$$x_6 = 20 - \frac{13}{3}x_2 - \frac{4}{3}x_3.$$

В качестве новой переменной выбираем x_2. Вычислим значения D_i по всем уравнениям для этой переменной: b_i / a_{i2} из них выберем наименьшее:

$$min\left(2 : \frac{2}{3}, 10 : 4\frac{1}{3}, -, 20 : 4\frac{1}{3}\right) = 2\frac{4}{13}.$$

Вместо переменной x_4 в план войдет переменная x_2. Выразим переменную x_2 через x_4: $x_2 = \frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4$ и подставим во все выражения:

$$x_0 = 10 - 6\frac{1}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4\right) + 1\frac{2}{3}x_3;$$

$$x_4 = 2 - \frac{2}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4\right) + \frac{1}{3}x_3;$$

$$x_5 = 2 + 1\frac{2}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4\right) - \frac{1}{3}x_3;$$

$$x_6 = 20 - 4\frac{1}{3}\left(\frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_{47}\right) - 1\frac{1}{3}x_3.$$

После приведения всех подобных, получаем новую систему, эквивалентную прежней:

$$x_0 = -\frac{60}{13} + \frac{9}{13}x_3 + \frac{19}{13}x_4;$$

$$x_1 = \frac{6}{13} + \frac{3}{13}x_3 + \frac{2}{13}x_4;$$

$$x_2 = \frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4;$$

$$x_5 = \frac{76}{13} - \frac{1}{13}x_3 - \frac{5}{13}x_4;$$

$$x_6 = 10 - 2x_3 + x_4.$$

Полагая небазисные переменные $x = (1, 2, 5, 6)$ равными нулю, получим новый допустимый вектор и значение целевой функции:

$$x = \left(0, 0, -\frac{9}{13}, -\frac{19}{13}, 0, 0\right),\ x_0 = -\frac{60}{13}.$$

Выражение для x_0 не содержит положительных элементов, значит найден оптимальный план. Окончательный вариант системы уравнений:

$$\begin{aligned}
x_0 &= -\frac{60}{13} + \frac{9}{13}x_3 + \frac{19}{13}x_4;\\
x_1 &= \frac{6}{13} + \frac{3}{13}x_3 + \frac{2}{13}x_4;\\
x_2 &= \frac{30}{13} + \frac{2}{13}x_3 - \frac{3}{13}x_4;\\
x_5 &= \frac{76}{13} - \frac{1}{13}x_3 - \frac{5}{13}x_4;\\
x_6 &= 10 - 2x_3 + x_4.
\end{aligned}$$

Так как в оптимальном решении отсутствуют искусственные переменные (они равны нулю), то данное решение является допустимым. Оптимальный план можно записать так: $x_1 = \frac{6}{13}$, $x_2 = 2\frac{4}{13}$.

Симплекс-метод является более громоздким по сравнению с графическим и достаточно трудоемким. При большом числе итераций ручной расчет симплекс-метода займет значительное количество времени, однако его неоспоримое преимущество перед графическим методом – точность расчетов. Разрешить противоречие между скоростью получения решений и их точностью можно путем использования специальных пакетов прикладных программ или широко распространенной компьютерной программы Excel. Она нужна для проведения расчетов, составления таблиц и диаграмм, вычисления простых и сложных функций и входит в состав пакета Microsoft Office.

Раздел 5. МЕТОДЫ ПОИСКА ОПТИМАЛЬНЫХ РЕШЕНИЙ

5.1. Транспортная задача и построение опорного плана перевозок

Транспортная задача – это задача о поиске оптимального распределения поставок однородного товара от поставщиков к потребителям при известных затратах на перевозку (тарифах) между пунктами отправления и назначения. Является задачей линейного программирования специального вида.

В модели транспортной задачи предполагается, что суммарные запасы поставщиков равны суммарным запросам потребителей, т. е.:

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

Такая задача называется задачей с *правильным балансом*, а модель задачи ***закрытой***. Если же это равенство не выполняется, то задача называется задачей с *неправильным балансом*, а модель задачи – *открытой*.

Математическая формулировка транспортной задачи такова: найти переменные задачи $X = (x_{ij})$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, удовлетворяющие системе ограничений, условиям неотрицательности и обеспечивающие минимум целевой функции.

Учитывая неотрицательность объемов перевозок задача выглядит следующим образом: однородный груз сосредоточен у m поставщиков в объемах $a_1, a_2, \ldots, a_m$. Данный груз необходимо доставить n потребителям в объемах $b_1, b_2, \ldots, b_n$. Известны C_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$ – стоимости перевозки единиц груза от каждого i-го поставщика каждому j-му потребителю. Требуется составить такой план перевозок, при котором запасы всех поставщиков вывозятся полностью, запросы всех потребителей удовлетворяются полностью, и суммарные затраты на перевозку всех грузов являются минимальными.

В транспортных задачах под поставщиками и потребителями понимаются различные предприятия, заводы, фабрики, склады, магазины и т. д. Однородными считаются грузы, которые могут быть перевезены одним видом транспорта. Под стоимостью перевозок понимаются тарифы, время, расход топлива и т. п. При решении транспортной задачи выбор критерия оптимальности имеет важное значение. Как известно, оценка экономической эффективности примерного плана может определятся по тому или иному критерию, положенного в основу расчета плана. Этот критерий является экономическим показателем, характеризующим качество плана. До настоящего времени нет общепринятого единого критерия, всесторонне учитывающего экономические факторы. При решении транспортной задачи, в качестве критерия оптимальности в различных случаях используют следующие показатели:

1. Объем работы транспорта (критерий – расстояние в т/км). Минимум пробега удобен для оценки планов перевозок, поскольку расстояние перевозки определяется легко и точно для любого направления. Поэтому критерию нельзя решать транспортные задачи с участием многих видов транспорта.

2. Тарифная плата за перевозку груза (критерий – тарифы провозных плат). Позволяет получить схему перевозок, наилучшую с точки зрения хозрасчетных показателей предприятия. Все надбавки, а также существующие льготные тарифы затрудняют его использование.

3. Эксплуатационные расходы на транспортировку грузов (критерий – себестоимость эксплуатационных расходов). Более верно отражает экономичность перевозок различными видами транспорта. Позволяет делать обоснованные выводы о целесообразности переключения с одного вида транспорта на другой.

4. Сроки доставки грузов (критерий – затраты времени).

5. Приведенные затраты (с учетом эксплуатационных расходов, зависящих от размеров движения и капиталовложения в подвижной состав).

6. Приведенные затраты (с учетом полных эксплуатационных расходов капиталовложений на строительство объектов в подвижной состав):

$$C_{приб} = C_{эи} + K_{эф}\left(K_{к} + \frac{T}{24} \cdot \frac{Ц}{365}\right) руб/т,$$

где $C_{эи}$ – эксплуатационные издержки; $K_{эф}$ – расчетный коэффициент эффективности капиталовложения; $K_{к}$ – капитальные вложения, приходящие на 1 т груза на протяжении участка; T – время следования; $Ц$ – цена одной тоны груза.

Результат решения ТЗ позволяет более полно производить оценку рационализации разных вариантов планов перевозок, с достаточно полной выраженностью количественно-одновременное влияние нескольких экономических факторов.

Сформулируем условие транспортной задачи (ТЗ) в общем виде.

Имеется несколько пунктов производства или сосредоточения какого-либо однородного продукта (например, гравия или песка), а также некоторое количество пунктов, испытывающих потребность в этом продукте (объектов транспортного строительства). Запасы в каждом пункте (карьере) известны. Транспортные издержки, связанные с произвольной перевозкой, заданы. Необходимо составить такой план перевозок материалов из пунктов сосредоточения в пункты потребления, при котором весь объем материалов будет перевезен в пункты потребления, каждый пункт потребления будет полностью удовлетворен и суммарные транспортные затраты окажутся минимальными.

Переведем условие задачи на математический язык. Пусть A_i – пункты производства (их число m); B_j – пункты потребления (их число n); a_i – количество материала в пункте A_i; а b_j – количество материала в пункте B_j. Примем допущение, что стоимость перевозки продукции линейно зависит от количества перевозимой продукции. Пусть $c_{ij}(x)$ – стоимость перевозки x единиц продукта из A_i в B_j. С учетом наших допущений имеем: $c_{ij}(x) = c_{ij}x$. В качестве параметров управления выберем x_{ij} – количество материала, перевозимого из A_i в B_j. Тогда суммарные транспортные издержки (целевую функцию задачи) можно определить как:

$$f(x) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} .$$

Составим ограничения. Необходимо вывести весь материал из каждого пункта производства $\sum_{j=1}^{n} x_{ij} = a_i \quad (i = 1, 2, ..., m)$ и полностью удовлетворить потребность объектов транспортного строительства $\sum_{i=1}^{m} x_{ij} = b_j \quad (j = 1, 2, \ldots, n)$. Поскольку перевозки идут из пунктов производства в пункты потребления, то добавим ограничения $x_{ij} \ge 0$ для любых i и j ($\forall i, j$).

В данной задаче производство полностью удовлетворяет потребление (выполняется условие баланса): $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$.

В результате получим математическую модель ТЗ:

$$f(x)=\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}\to min;$$

$$\begin{cases}\sum\limits_{j=1}^{n}x_{ij}=a_i,\ i=\overline{1,m};\\ \sum\limits_{i=1}^{m}x_{ij}=b_j,\ j=\overline{1,n};\\ x_{ij}\geq 0,\ \forall i,j.\end{cases}$$

Решение *транспортной задачи* состоит из трех этапов:

- составление математической модели;
- нахождение опорного (первоначального) плана;
- нахождение оптимального плана.

Второй этап желателен, но необязателен. Для нахождения опорного плана обычно используют методы:

- «северо-западного угла»;
- метод наименьших стоимостей или минимального элемента.

Пример. Имеются два гравийных карьера $a_1=10$, $a_2=20$ и два объекта дорожного строительства: $b_1=5$, $b_2=25$. Стоимость перевозок единицы груза с карьеров на объекты задана $c_{11}=1$, $c_{12}=2$, $c_{21}=3$, $c_{22}=1$. Необходимо составить математическую модель плана перевозок гравия с карьеров на объекты дорожного строительства, при котором все запасы карьеров будут исчерпаны и каждый объект дорожного строительства будет полностью обеспечен материалом, при этом суммарные транспортные затраты окажутся минимальными.

Решение. Поскольку условие баланса выполняется ($10+20=5+25$), то ТЗ является закрытой и все ограничения записываются в виде равенств.

Пусть x_{ij} – количество продукта, перевозимого из *i*-го пункта производства в *j*-й пункт потребления. Тогда:

$$F(x)=x_{11}+2x_{12}+3x_{21}+x_{22}\to min;$$

$$\begin{cases}x_{11}+x_{12}=10;\\ x_{21}+x_{22}=20;\\ x_{11}+x_{21}=5;\\ x_{12}+x_{22}=25;\\ x_{ij}\geq 0;\ \forall i,j.\end{cases}$$

Первое ограничение показывает, что из первого карьера необходимо вывезти гравий в первый и второй объекты дорожного строительства $x_{11} + x_{12}$ в количестве $a_1 = 10$. Аналогично второе ограничение.

Третье ограничение показывает, что на первый объект строительства необходимо привезти гравий из первого и второго карьеров $x_{11} + x_{21}$ в количестве $b_1 = 5$. Аналогично четвертое ограничение.

Составим первоначальный план перевозок методом наименьших стоимостей с двойным предпочтением. Исходные данные ТЗ представим в виде транспортной таблицы:

a_i / b_j	$b_1 = 5$	$b_2 = 25$
$a_1 = 10$	1^{**}	2
$a_2 = 20$	3	1^{**}

Отмечаем клетки с наименьшими стоимостями перевозок по каждой строке и каждому столбцу. Клетки, имеющие две отметки, заполняем в первую очередь. В клетке $a_1 b_1$ записываем число 5 и вычеркиваем полностью заполненный столбец b_1. В клетке $a_2 b_2$ записываем число 20 и вычеркиваем полностью заполненную строку a_2:

a_i / b_j	$b_1 = 5$	$b_2 = 25$
$a_1 = 10$	**5**	
$a_2 = 20$		**20**

Заполняем клетки с одной отметкой с учетом объемов карьеров и заявок объектов строительства, затем заполняем клетки без отметок. Осталась одна клетка $a_1 b_2$. В нее записываем число 5:

a_i / b_j	$b_1 = 5$	$b_2 = 25$
$a_1 = 10$	**5**	**5**
$a_2 = 20$		**20**

Получили опорный план методом минимизации, значение целевой функции рассчитываем по элементам опорного плана: $f(x) = 35$, $x_{11} = 5$, $x_{12} = 5$, $x_{22} = 20$. В данном примере опорный план совпал с оптимальным решением.

Рассмотрим открытую ТЗ, в которой не выполняется условие баланса.

Пример. Имеются два пункта производства товара, $a_1 = 10$, $a_2 = 25$, и два пункта потребления, $b_1 = 10$, $b_2 = 30$. Стоимость перевозок товара задана $c_{11} = 4$, $c_{12} = 2$, $c_{21} = 3$, $c_{22} = 1$. За недопоставку одной единицы товара первый пункт потребления терпит убыток, равный 1, а второй – 0.5. Необходимо составить математическую модель плана перевозок продукта из пунктов производства в пункты потребления, при котором весь продукт был бы перевезен в пункты потребления, а суммарные транспортные затраты и убытки оказались минимальными.

Решение. Введем обозначения. Пусть x_{ij} – количество товара, перевозимого из i-го пункта производства в j-й пункт потребления, $z_1 = b_1 - x_{11} - x_{21}$, $z_2 = b_2 - x_{12} - x_{22}$ – количество недопоставленного товара соответственно в первый и второй пункты потребления.

Составим математическую модель:

$$F(x) = 4x_{11} + 2x_{12} + 3x_{21} + x_{22} + z_1 + 0.5z_2 \rightarrow min;$$

$$\begin{cases} x_{11} + x_{12} = 10; \\ x_{21} + x_{22} = 25; \\ x_{11} + x_{21} \le 10; \\ x_{12} + x_{22} \le 30; \\ x_{ij} \ge 0; \forall i, j. \end{cases}$$

Прежде чем решать открытую ТЗ, ее необходимо сбалансировать. Вводим дополнительный *фиктивный* пункт производства a_{i+1} (в случае дефицита товара) или фиктивный пункт потребления b_{j+1} (в случае профицита товара). Мощность дополнительно введенного фиктивного пункта равна количеству недопоставленного или избыточного товара. Для нашего случая введем фиктивный пункт производства $a_3 = (30+10)-(25+10) = 5$.

Стоимость перевозок из фиктивного пункта производства равна нулю, поэтому целевая функция не меняется, а система ограничений принимает вид:

$$\begin{cases} x_{11} + x_{12} = 10; \\ x_{21} + x_{22} = 25; \\ x_{31} + x_{32} = 5; \\ x_{11} + x_{21} + x_{31} = 10; \\ x_{12} + x_{22} + x_{32} = 30; \\ x_{ij} \ge 0, \forall i, j. \end{cases}$$

Значение целевой функции $f(x) = 60$, при опорном плане:

$$x_{12} = 10; \quad x_{21} = 5; \quad x_{22} = 20.$$

Рассмотрим задачу большей размерности.

<u>Пример</u>. Завод имеет три цеха *A*, *B*, *C* и четыре склада №№ 1, 2, 3, 4. Цех *A* производит 30 тыс. изделий, цех *B* – 40 тыс. шт., цех *C* – 20 тыс. шт.

за плановое время. Пропускная способность складов за то же время: склад № 1 – 20 тыс. шт., склад № 2 – 30 тыс. шт., склад № 3 – 30 тыс. шт., склад № 4 – 10 тыс. шт. Стоимость перевозок одной тысячи изделий в склады №№ 1, 2, 3, 4 соответственно составляет: из цеха *A* – 2, 3, 3, 4 ед.; из цеха *B* – 3, 2, 5, 1 ед.; из цеха *C* – 4, 3, 2, 6 ед. Составить план перевозок, при котором расходы на перевозку всех изделий минимальны.

Решение. Условие баланса выполняется $\sum_{i=1}^{3} a_i = \sum_{j=1}^{4} b_j = 90$, поэтому ТЗ является замкнутой и все ограничения записываем в виде равенств.

Пусть x_{ij} (тыс. шт.) – количество изделий, перевозимых из *i*-го цеха в *j*-й склад. Тогда математическая модель задачи будет иметь вид:

$$F(x) = 2x_{11} + 3x_{12} + 3x_{13} + 4x_{14} + 3x_{21} + 2x_{22} + 5x_{23} + x_{24} + 4x_{31} + 3x_{32} + 2x_{33} + 6x_{34} \to min;$$

цеха: $\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 30; \\ x_{21} + x_{22} + x_{23} + x_{24} = 40; \\ x_{31} + x_{32} + x_{33} + x_{34} = 20, \end{cases}$ склады: $\begin{cases} x_{11} + x_{21} + x_{31} = 20; \\ x_{12} + x_{22} + x_{32} = 30; \\ x_{13} + x_{23} + x_{33} = 30; \\ x_{14} + x_{24} + x_{34} = 10; \\ x_{ij} \ge 0. \end{cases}$

Составим первоначальный план перевозок методом наименьших стоимостей с двойным предпочтением. Исходные данные ТЗ представим в виде транспортной таблицы. Отметим клетки с наименьшими стоимостями перевозок по каждой строке и каждому столбцу:

a_i / b_j	$b_1 = 20$	$b_2 = 30$	$b_3 = 30$	$b_4 = 10$
$a_1 = 30$	2**	3	3	4
$a_2 = 40$	3	2*	5	1**
$a_3 = 20$	4	3	2**	6

Клетки, имеющие две отметки, заполняем в первую очередь с учетом заявок производителей и потребителей: $x_{11} = 20$ (вычеркиваем первый полностью заполненный столбец), $x_{24} = 10$ (вычеркиваем четвертый столбец), $x_{33} = 20$ (вычеркиваем третью строку).

Затем заполняем клетки с одной отметкой с учетом заявок производителей и потребителей: $x_{22} = 30$ (вычеркиваем второй столбец и вторую строку).

Остается клетка $a_1 b_3 = x_{13} = 10$:

a_i / b_j	$b_1 = 20$	$b_2 = 30$	$b_3 = 30$	$b_4 = 10$
$a_1 = 30$	2** **20**	3	3 **10**	4
$a_2 = 40$	3	2* **30**	5	1** **0**
$a_3 = 20$	4	3	2** **20**	6

Получили начальный план перевозок:

$x_{11} = 20$; $x_{13} = 10$; $x_{22} = 30$; $x_{24} = 10$; $x_{33} = 20$; $f(x) = 180$ ед.

5.2. Методы построения опорных планов решения транспортной задачи

Существует несколько методов построения опорного плана при решении ТЗ:

- метод северо-западного угла;
- метод минимизации по строке (столбцу);
- метод Фогеля.

Рассмотрим основные положения каждого из них.

Метод северо-западного угла

В стандартных условиях ТЗ заданы возможности поставщиков a_{ij} и потребности потребителей b_j. Требуется найти допустимые объемы перевозки от каждого поставщика к каждому потребителю x_{ij}.

Шаг 1. Первая ячейка – с которой начинается распределение – будет «северо-западная» ячейка в левом верхнем углу таблицы (1-й поставщик, 1-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 20 и 30 кг, то есть 20 кг). Поскольку спрос 1-го потребителя полностью удовлетворен, ячейки соответствующего столбца заполняться больше не будут.

Шаг 2. Переходим в следующую «северо-западную» ячейку, не считая уже распределенной области. Этой ячейкой будет x_{12} (1-й поставщик, 2-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 30 и

10 кг, то есть 10 кг). Соответственно, уменьшаем оставшиеся не распределенными объемы поставки и потребления в строке и столбце на 10 кг. Запасы 1-го поставщика (в 1-й – верхней – строке) теперь исчерпаны, распределение по этой строке завершено.

Шаг 3. Переходим в следующую «северо-западную» ячейку, не считая уже распределенной области. Этой ячейкой будет x_{22} (2-й поставщик, 2-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 40 и 20 кг, то есть 20 кг). Соответственно, уменьшаем оставшиеся не распределенными объемы поставки и потребления в строке и столбце на 20 кг. Потребности 2-го потребителя теперь полностью удовлетворены, распределение по этому столбцу завершено.

Шаг 4. Переходим в следующую «северо-западную» ячейку, не считая уже распределенной области. Этой ячейкой будет x_{23} (2-й поставщик, 3-й потребитель). Вписываем в эту ячейку максимальный объем, который позволяет запас поставщика и спрос потребителя (берем минимум между 30 и 20 кг, то есть 20 кг). Соответственно, уменьшаем оставшиеся не распределенными объемы поставки и потребления в строке и столбце на 20 кг. Запасы 2-го поставщика (в 2-й сверху строке) теперь исчерпаны, распределение по этой строке завершено.

Шаг 5. Распределение оставшихся у последнего поставщика 20 кг груза по двум потребителям по 10 кг.

Результаты выполненных шагов, в условиях предыдущей задачи, то есть построенный опорный план – представлены в таблице.

	Потребитель B_1, потребность $20-20=0\,кг$	*Потребитель B_2, Потребность* $30-10-20=0\,кг$	*Потребитель B_3, Потребность* $30-10-20=0\,кг$	*Потребитель B_4, потребность* $10-10=0\,кг$
Поставщик A_1 запас $30-20-10=0\,кг$	$X_{11}=20\,кг$	$X_{12}=10\,кг$		
Поставщик A_2 запас $40-20-20=0\,кг$		$X_{22}=20\,кг$	$X_{23}=20\,кг$	
Поставщик A_3 запас $10-10=0\,кг$			$X_{33}=10\,кг$	$X_{34}=10\,кг$

При построении опорного плана по способу северо-западного угла совершенно не учитываются тарифы, потому план получается весьма далеким от оптимального. Для последующего нахождения оптимального решения ТЗ приходится делать много приближений (шагов).

Метод минимизации по строке

Этот способ учитывает тарифы перевозок и потому позволяет найти план, более близкий к оптимальному Суть метода заключается в том, что из всей таблицы стоимостей выбирают наименьшую, и в клетку, которая ей соответствует, помещают меньшее из чисел a_i, или b_j. Затем, из рассмотрения исключают либо строку, соответствующую поставщику, запасы которого полностью израсходованы, либо столбец, соответствующий потребителю, потребности которого полностью удовлетворены, либо и строку и столбец, если израсходованы запасы поставщика и удовлетворены потребности потребителя.

Из оставшейся части таблицы стоимостей снова выбирают наименьшую стоимость, и процесс распределения запасов продолжают, пока все запасы не будут распределены, а потребности удовлетворены.

Запишем условие ТЗ в таблице и составим опорный план.

Поставщики	***Потребители***					***Запасы***
	b_1	b_2	b_3	b_4	b_5	
a_1	10	7	4	1 100	4	100
a_2	2 200	7 50	10	6	11	250
a_3	8	5	3	2	2 200	200
a_4	11	8 150	12 100	16	13 50	300
Потребности	200	200	100	100	250	850

Выбираем в таблице наименьшую стоимость (это стоимость, помещенная в клетке a_1, b_4) так как $a_1 = b_4 = 100$ ед. груза помещаем в этой клетке и исключаем из рассмотрения первую строку и четвертый столбец.

В оставшейся таблице стоимостей наименьшей является стоимость, расположенная в клетке a_2, b_1 и в клетке a_3, b_5. Заполняем любую из них, например a_2, b_1. Имеем $200 < 250$, следовательно, записываем в нее 200 и исключаем из рассмотрения столбец b_1. В клетку a_3, b_5 записываем 200 ед. и исключаем из рассмотрения строку a_3. В оставшейся таблице стоимостей снова выбираем наименьшую стоимость и продолжаем процесс до тех пор, пока все запасы не будут распределены, а потребности удовлетворены. В результате получен опорный план: $X_{14} = 100$, $X_{21} = 200$, $X_{22} = 50$, $X_{35} = 200$, $X_{42} = 150$, $X_{43} = 100$, $X_{45} = 50$, остальные значения переменных равны нулю.

План не содержит циклов и состоит из семи положительных перевозок, следовательно, является вырожденным опорным планом. Определим его стоимость:

$$Z = 100 \cdot 1 + 200 \cdot 2 + 50 \cdot 7 + 200 \cdot 2 + 150 \cdot 8 + 100 \cdot 12 + 50 \cdot 13 = 4300 \ (ед).$$

Полученный план перевозок значительно ближе к оптимальному плану.

Недостатком метода является то, что при выборе минимального элемента в строке не учитывается стоимостной показатель по столбцу и наоборот, что не всегда экономически оправдано.

Метод наименьшего элемента

Суть метода в том, что в транспортной таблице сначала заполняются ячейки с наименьшими тарифами, а потом уже ячейки с большими тарифами. То есть выбираются перевозки с минимальной стоимостью доставки груза.

Шаг 1. Располагаем все клетки таблицы в очередь по мере возрастания тарифов, начиная с минимального.

Шаг 2. В клетку с минимальным тарифом записываем наибольшую возможную перевозку (исходя из запасов и потребностей), затем заполняем очередную по порядку клетку и т. д., пока не получим опорный план. При этом должен строго соблюдаться баланс по строкам и столбцам.

Пустые клетки прочеркиваем, а не заполняем нулями (чтобы было видно, что они не входят в план).

Если окажется, что есть несколько ячеек с одинаковыми и минимальными тарифами – выбираем любую из них. Он лучше плана, построенного по методу северо-западного угла, и для последующего нахождения оптимального плана потребуется меньше вычислений. Пример применения данного метода приведен в предыдущем подразделе.

Существенным недостатком рассмотренного метода является необходимость анализа всей таблицы исходных данных ТЗ, что очевидно неудобно при решении задач большой размерности.

Метод аппроксимации Фогеля

Первым делом добавляем к транспортной таблице дополнительные строку и столбец. Далее находим для каждой строки и каждого столбца абсолютные разности между двумя минимальными тарифами. Если в строке/столбце две клетки с одинаковыми и минимальными значениями тарифов, то берем именно их. Тогда разность будет равна 0. Найденные разности выписываем в добавочный столбец и добавочную строку. Среди вычисленных разностей (и по строкам, и по столбцам!) выбираем наибольшую. Затем в строке (или столбце), которой соответствует максимальная разность, ищем клетку с минимальным тарифом и заполняем ее. Если клеток с минимальным тарифом несколько, то заполняем ту из них, которой соответствует наибольшая разность. Затем повторяем все действия снова, уже не учитывая заполненные клетки.

Пример. Используя метод аппроксимации Фогеля, найти опорный план транспортной задачи, исходные данные которой приведены в таблице.

Пункты отправления	Пункты назначения				Запасы
	b_1	b_2	b_3	b_4	
a_1	7	8	1	2	160
a_2	4	5	9	8	140
a_3	9	2	3	6	170
Потребности	120	50	190	110	470

Для каждой строки и столбца таблицы условий найдем разности между двумя минимальными тарифами, записанными в данной строке или столбце, и поместим их в соответствующем дополнительном столбце или дополнительной строке таблица ниже.

Пункты отправления	Пункты назначения				Запасы	Разности по строкам					
	b_1	b_2	b_3	b_4							
a_1	7	8	1	2	160	1	6	–	–	–	–
a_2	4	5	9	8	140	1	1	1	1	1	0
a_3	9	2	3	6	170	1	1	1	–	–	–
Потребности	120	50	190	110	470						
Разности по столбцам	3	3	2	4							
	3	3	2	–							
	5	3	6	–							
	5	3	–	–							
	0	0	–	–							
	–	0	–	–							

Так, в строке a_2 минимальный тариф равен 4, а следующий за ним равен 5, разность между ними 5 – 4 = 1. Точно так же разность между минимальными элементами в столбце b_4 равна 6 – 2 = 4. Вычислив все эти разности, видим, что наибольшая из них соответствует столбцу b_4. В этом столбце минимальный тариф записан в клетке, находящейся на пересечении строки a_1 и столбца b_4. Таким образом, эту клетку следует заполнить. Заполнив ее, тем самым мы удовлетворим потребности пункта b_4. Поэтому исключим из рассмотрения столбец b_4 и будем считать запасы пункта a_1 равными 160 – 110 = 50 ед. После этого определим следующую клетку для заполнения. Снова найдем разности между оставшимися двумя минимальными тарифами в каждой из строк и столбцов и запишем их во втором дополнительном столбце и во второй дополнительной строке таблицы. Как видно из этой таблицы, наибольшая указанная разность соответствует строке a_1. Минимальный тариф в этой строке записан в клетке, которая находится на пересечении ее со столбцом b_3. Следовательно, заполняем эту клетку. Поместив в нее число 50, тем самым предполагаем, что запасы в пункте a_1 полностью исчерпаны, а потребности в пункте b_3 стали равными 190 – 50 = 140 ед. Исключим из рассмотрения строку a_1 и определим новую клетку для заполнения. Продолжая итерационный процесс, последовательно заполняем клетки, находящиеся на пересечении строки a_3 и столбца b_3, строки a_3 и столбца b_2, строки a_2 и столбца b_1, строки a_2 и столбца b_2. В результате получим опорный план:

$$X = \begin{pmatrix} 0 & 0 & 50 & 110 \\ 120 & 20 & 0 & 0 \\ 0 & 30 & 140 & 0 \end{pmatrix}.$$

При этом плане общая стоимость перевозок такова:

$$Ц = 1\cdot 50 + 2\cdot 110 + 4\cdot 120 + 5\cdot 20 + 2\cdot 30 + 2\cdot 30 + 3\cdot 140 = 1330.$$

Как правило, применение метода аппроксимации Фогеля позволяет получить либо опорный план, близкий к оптимальному, либо сам оптимальный план. Кстати, найденный выше опорный план транспортной задачи является и оптимальным.

Во всех примерах опорный план содержит не более чем из $k+l-1$ компонент, то есть не вырожденный. Это обусловлено тем, что с заполнением каждой клетки из рассмотрения исключается строка или столбец. Однако встречаются случаи, когда одновременное исключение и строки и столбца происходит не только в конце построения опорного плана, но и на промежуточных этапах. В результате число заполненных клеток становится меньше на одну. Подобные случаи называются *вырождением* плана. Вырожденные опорные планы практически не оптимизируются, так как методы оптимизации «зацикливаются» и происходит бесконечное число итераций, не приводящих к отысканию оптимума.

Подсчитаем число занятых клеток таблицы в примере по методу Фогеля, их 6 и, согласно теории, должно быть $m+n-1=6$. Следовательно, опорный план является невырожденным и ациклическим.

5.3. Создание оптимального плана перевозок

Рассмотрим методы оптимизации построенного опорного плана ТЗ.

С помощью рассмотренных выше методов построения начального плана можно получить вырожденный или невырожденный опорный план. Построенный план транспортной задачи как задачи линейного программирования можно было бы довести до оптимального с помощью симплексного метода. Однако из-за громоздкости симплексных таблиц, содержащих $m \times n$ неизвестных, и большого объема вычислительных работ для получения оптимального плана используют более простые методы. Одним из таких методов и является метод потенциалов.

Метод потенциалов – модификация симплекс-метода решения задачи линейного программирования применительно к транспортной задаче. Он позволяет, отправляясь от некоторого допустимого решения, получить оптимальное решение за конечное число итераций. Метод потенциалов, широко применяется на практике и используется при программировании. Суть метода такова. На первом этапе составляют опорный план перевозок. На втором этапе строят систему потенциалов и проверяют начальный план на оптимальность. Каждому i-му поставщику устанавливается потенциал U_i, который можно интерпретировать как цену продукта в пункте поставщика, а каждому j-му потребителю устанавливается потенциал V_j, который характеризует цену продукта в пункте потребителя. Если опорный план является неоптимальным, то переходят к третьему этапу, суть которого заключается в корректировке плана прикрепления потребителей и поставщиков. Затем снова переходят ко второму этапу и т. д.

Алгоритм метода потенциалов

Для улучшения уже имеющего опорного плана необходимо составить двойственную задачу:

$$u_1 \ldots u_m \text{ и } v_1 \ldots v_n \text{ любые значения;} \tag{1}$$

$$u_i + v_j \le c_{ij}\ ,\ j = 1,\overline{n};\ i = 1,\overline{m}; \tag{2}$$

$$T(u,v) = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \to max; \tag{3}$$

Пусть есть план: $X = \{x_{ij}\}, \left(i = 1,\overline{m};\ j = 1,\overline{n}\right).$ (4)

Воспользуемся теоремой для дальнейшего построения алгоритма:

Теорема (критерий оптимальности): Для того чтобы допустимый план перевозок (4) в транспортной задаче был оптимальным, необходимо и достаточно, чтобы существовали такие числа $u_1 \ldots u_m$ и $v_1 \ldots v_n$, что:

$$u_i + v_j = c_{ij}\ ,\ \textit{если}\ x_{ij} \ge 0; \tag{5}$$

$$u_i + v_j \le c_{ij}\ ,\ \textit{если}\ x_{ij} = 0, \tag{6}$$

где u_i и v_j – потенциалы пункта назначения и отправления.

Сформулированная теорема позволяет построить алгоритм нахождения оптимального решения транспортной задачи:

Для опорного плана, в котором $m + n - 1$ базисных клеток, можно определить потенциалы u_i и v_j так, чтобы выполнялось условие (5). Поскольку система содержит $m + n - 1$ уравнений и $m + n$ неизвестных, то одну из них можно задать произвольно (например, приравнять к нулю). После этого из $m + n - 1$ уравнений определяются остальные потенциалы и для каждой из свободных клеток вычисляются величины $u_i + v_j = c_{ij}$. Если оказалось, что $c'_{ij} \le c_{ij}$, то план оптимален. Если же хотя бы в одной свободной клетке $c'_{ij} > c_{ij}$, то план не является оптимальным и может быть улучшен путем переноса по циклу, соответствующему данной свободной клетке.

Процесс улучшения плана продолжается до тех пор, пока не будут выполнены условия (5) и (6). Рассмотрим вариант применения метода потенциалов.

<u>Пример</u>. Имеется сеть мостов, при построении которых требуются определенное количество металлоконструкций. Также имеется ряд баз поставщиков, где требуемый конструкции хранится, при этом на каждой базе различный объем хранимого товара. Кроме того, известны затраты на перевозку металлоконструкций от каждой базы к каждому мосту. Возникает потребность разработать такой план перевозок, чтобы мосты получили требуемое количество металлоконструкций с наименьшими затратами на

транспортировку. Допустим, что опорный план перевозок уже составлен, но его необходимо улучшить. Именно в такой ситуации и применяется метод потенциалов.

Имеем задачу: улучшить опорный план перевозок, который является невырожденным и представлен в таблице.

Таблица 5.1. Опорный план перевозок

Мосты/Базы	b_1		b_2		b_3	
a_1	–		–		10	
		5		3		1
a_2	15		–		5	
		3		2		4
a_3	–		20		10	
		4		1		2

Для начала вычислим потенциалы для плана перевозок, представленного в таблице. Для этого сопоставим каждому поставщику a_i и каждому потребителю b_j величины u_i и v_j так, чтобы для всех базисных клеток плана было выполнено условие: $u_i + v_j = c_{ij}$. Результат показан в таблице 5.2.

Таблица 5.2. План перевозок с вычисленными потенциалами

Мосты/Базы	b_1		b_2		b_3		***U***
a_1	–		–		10		0
	5	5	3	3	0	1	
a_2	15		–		5		3
	0	3	– **1**	2	0	4	
a_3	–		20		10		1
	3	4	0	1	0	2	
V	0		0		1		

Как видно из данных таблицы – план перевозок не является оптимальным $c_{22} < 0$, поэтому его можно улучшить путем перераспределения поставок.

Для этого найдем ячейку с наибольшей по абсолютной величине отрицательной разностью c_{22} и построим цикл, в котором кроме этой ячейки все остальные являются базисными (такой цикл всегда существует и единственен). Отметим ячейку с отрицательной разностью c_{22} знаком «+», следующую знаком «–», и так далее, поочередно. Затем находим минимальное значение перевозки в ячейках цикла имеющих знак «–» и вписываем его в свободную ячейку со знаком «+». Затем последовательно обходим все ячейки цикла, поочередно вычитая и прибавляя к ним минимальное значение (в соответствии со знаками, которыми эти ячейки помечены: где минус – вычитаем, где плюс – прибавляем). Цикл повторяется до тех пор, пока не будет найдено оптимальное решение. Оптимальное решение, полученное с помощью метода потенциалов, для данной задачи представлено в таблице 5.3.

Таблица 5.3. Оптимальное решение

<table>
<tr><th>Мосты/Базы</th><th colspan="2">b_1</th><th colspan="2">b_2</th><th colspan="2">b_3</th><th>U</th></tr>
<tr><td rowspan="2">a_1</td><td colspan="2">–</td><td colspan="2">–</td><td colspan="2">10</td><td rowspan="2">0</td></tr>
<tr><td>4</td><td>5</td><td>3</td><td>3</td><td>0</td><td>1</td></tr>
<tr><td rowspan="2">a_2</td><td colspan="2">15</td><td colspan="2">5</td><td colspan="2">–</td><td rowspan="2">2</td></tr>
<tr><td>0</td><td>3</td><td>0</td><td>2</td><td>1</td><td>4</td></tr>
<tr><td rowspan="2">a_3</td><td colspan="2">–</td><td colspan="2">15</td><td colspan="2">15</td><td rowspan="2">1</td></tr>
<tr><td>2</td><td>4</td><td>0</td><td>1</td><td>0</td><td>2</td></tr>
<tr><td>V</td><td colspan="2">1</td><td colspan="2">0</td><td colspan="2">1</td></tr>
</table>

Поиск оптимального решения сводится к отысканию такой системы потенциалов, при которой выполняется условие оптимальности плана.

Рассмотрим еще один пример использования метода потенциалов.

Пример. В условиях предыдущей задачи рассмотрим три базы и три моста, но с другими стоимостными показателями, представленными в таблице:

Мосты	Базы		
	b_1	b_2	b_3
a_1	6	3	4
a_2	3	7	2
a_3	2	8	6

Как следует спланировать перевозку, чтобы ее стоимость была минимальной?

Решение. Условие баланса выполняется: $\sum_{i=1}^{3} a_i = \sum_{i=1}^{3} b_j = 50$, поэтому ТЗ является закрытой и все ограничения записываем в виде равенств.

На первом этапе получаем опорный план перевозок методом наименьших стоимостей с двойным предпочтением.

a_1 / b_j	$b_1 = 20$	$b_2 = 12$	$b_3 = 18$	U_i
$a_1 = 15$	6 **3**	3** **12**	4	$U_1 = 0$
$a_2 = 25$	3 **7**	7	2** **18**	$U_2 = 3$
$a_3 = 10$	2** **10**	8	6	$U_3 = 4$
V_j	$V_1 = 6$	$V_2 = 3$	$V_3 = 5$	

Опорный план: $x_{11} = 3$, $x_{12} = 12$, $x_{21} = 7$, $x_{23} = 18$, $x_{31} = 10$, $f(x) = 131$.

На втором этапе построим систему потенциалов U_i и V_j для начального опорного плана.

Для занятых клеток составляем систему линейных уравнений вида $V_j = U_i + C_{ij}$:

$$V_1 = U_1 + 6;\ V_1 = U_2 + 3;\ V_1 = U_3 + 2;\ V_2 = U_1 + 3;\ V_3 = U_2 + 2.$$

Полагаем $U_1 = 0$, тогда $V_1 = 6$; $U_2 = 3$; $U_3 = 4$; $V_2 = 3$; $V_3 = 5$.

Найденные потенциалы указаны в вышеприведенной таблице.

Для оптимальности опорного плана необходимо и достаточно, чтобы система потенциалов U_i и V_j удовлетворяла условию:

$V_j = U_i + C_{ij}$ для $x_{ij} > 0$ (для занятой клетки);

$V_j \le U_i + C_{ij}$ для $x_{ij} = 0$ (для незанятой клетки).

Для незанятых клеток проверим условие оптимальности:

$$V_2 = 3 < U_2 + C_{22} = 3 + 7 = 10; \qquad V_2 = 3 < U_3 + C_{32} = 4 + 8 = 12;$$

$$V_3 = 5 > U_1 + C_{13} = 0 + 4 = 4; \qquad V_3 = 5 < U_3 + C_{33} = 4 + 6 = 10.$$

Третье неравенство, или клетка a_1b_3, не удовлетворяет условию оптимальности, поэтому переходим к третьему этапу.

Для незанятых клеток, не удовлетворяющих условию оптимальности, находим наибольшую величину $\alpha_{ij} = V_j - U_i - C_{ij}$:

$$\alpha_{13} = V_3 - U_1 - C_{13} = 5 - 0 - 4 = 1.$$

Клетка a_1b_3 является положительной вершиной цикла. Строим замкнутый контур, начальная вершина которого находится в выбранной клетке, а остальные вершины контура – в занятых клетках. Линии контура могут быть горизонтальными либо вертикальными отрезками. Число отрезков и вершин будет четным. Первая вершина контура свободной клетки имеет знак «+», а в остальных вершинах контура расставляются поочередно знаки «–» и «+».

a_1/b_j	$b_1=20$	$b_2=12$	$b_3=18$	U_i
$a_1=15$	6 – **3**	3 **12**	4 +	$U_1=0$
$a_2=25$	3 + **7**	7	2 – 18	$U_3=3$
$a_3=10$	2 **10**	8	6	$U_3=4$
V_j	$V_1=6$	$V_2=3$	$V_3=5$	

Выбираем наименьшее из величин в вершинах контура с отрицательным знаком $min\{3{,}18\}=3$. Уменьшаем на три единицы величину каждой вершины со знаком «–» и увеличиваем на три единицы величину каждой вершины со знаком «+». Получим:

a_1/b_j	$b_1=20$	$b_2=12$	$b_3=18$	U_i
$a_1=15$	6	3 **12**	4 **3**	$U_1=0$
$a_2=25$	3 **10**	7	2 **15**	$U_2=2$
$a_3=10$	2 **10**	8	6	$U_3=3$
V_j	$V_1=5$	$V_2=3$	$V_3=4$	

Новый опорный план:

$$x_{12}=12;\ \ x_{13}=3;\ \ x_{21}=10;\ \ x_{23}=15;\ \ x_{31}=10;\ \ f(x)=128.$$

Найдем потенциалы и проверим новый опорный план на оптимальность. Для занятых клеток составляем систему линейных уравнений вида $V_j=U_i+C_{ij}$:

$$V_1=U_2+3;\ \ V_1=U_3+2;\ \ V_2=U_1+3,\ \ V_3=U_1+4,\ \ V_3=U_2+2.$$

Полагаем $U_1=0$, тогда $V_2=3$, $V_3=4$, $U_2=2$, $V_1=5$, $U_3=3$. Найденные потенциалы указаны в вышеприведенной таблице.

Для незанятых клеток проверим условие оптимальности:

$$V_1=5<U_1+C_{11}=0+6=6; \qquad V_2=3<U_2+C_{22}=2+7=9;$$

$$V_2=3<U_3+C_{32}=3+8=11; \qquad V_3=4<U_3+C_{33}=3+6=9.$$

Полученный опорный план $x_{12}=12$, $x_{13}=3$, $x_{21}=10$, $x_{23}=15$, $x_{31}=10$, удовлетворяет всем условиям оптимальности и целевая функция принимает значение $f(x)=128$.

Распределительный метод состоит в последовательном улучшении опорного плана перевозок путем отыскания на каждом шаге выгодных циклов переноса грузов. Опорный план для данного метода можно сформировать, применяя метод «северо-западного» угла. Более подробно рассмотрим процесс формирования очередного цикла переноса на каждом новом шаге алгоритма. Очевидно, что при перемещении x единиц груза по некоторому циклу с ценой g стоимость перевозок изменяется на величину $x \times g$. Тогда, для улучшения текущего плана перевозок имеет смысл перемещать перевозки только по тем циклам, цена которых отрицательна. Если циклов с отрицательной ценой в таблице больше не осталось, это означает, что оптимальный план достигнут. При улучшении плана циклическими переносами пользуются приемом, заимствованным из симплекс-метода: на каждом шаге (цикле) заменяют одну свободную переменную на базисную, т. е. заполняют одну клетку и взамен того освобождают одну из базисных клеток.

Можно доказать, что для любой свободной клетки транспортной таблицы всегда существует цикл (и притом единственный), одна из вершин которого лежит в этой клетке, а все остальные в базисных клетках. Если цена такого цикла, с плюсом в свободной клетке, отрицательна, то план можно улучшить. Количество единиц груза (x), которые можно переместить, определяется минимальным значением перевозок, стоящих в отрицательных вершинах цикла.

Пример. Распределительным методом найти оптимальный план перевозок транспортной задачи, имеющей следующую таблицу издержек:

Сток/ Исток	B_1	B_2	B_3	B_4	*Запасы*
A_1	10	7	6	8	31
A_2	5	6	5	4	48
A_3	8	7	6	7	38
Заявки	22	34	41	20	117

Решение.

1. Методом «северо-западного» угла найдем опорный план перевозок:

Сток / *Исток*	B_1	B_2	B_3	B_4	*Запасы*
A_1	10 22	7 9	6	8	31
A_2	5	6 25	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Опорный план имеет шесть базисных клеток в соответствующей ему транспортной таблице, что позволяет его использовать без модификаций для дальнейшего решения задачи: $n+m-1=4+3-1=6$. Посчитаем стоимость найденного опорного плана: $L=22\cdot 10+9\cdot 7+...=796$.

2. Попытаемся улучшить найденный опорный план перевозок методом циклических переносов. Вычислим цену цикла для каждой свободной клетки. Количество свободных клеток в транспортной таблице данного опорного плана равно: $k=3\cdot 2=6$.

$\gamma_{13}=6-5+6-7=0$; $\gamma_{14}=8-7+6-5+6-7=1$; $\gamma_{21}=5-10+7-6=-4$;

$\gamma_{24}=4-7+6-5=-2$; $\gamma_{31}=8-10+7-6+5-6=-2$; $\gamma_{32}=7-6+5-6=0$.

3. Для всех свободных переменных (клеток) с отрицательной ценой цикла вычислим максимальное количество груза, которое можно перенести по соответствующему циклу. Очевидно, что максимальное количество груза, которое можно переместить по некоторому выбранному циклу будет равно минимальному значению груза среди отрицательных клеток цикла.

$$max\, x_{21}=min(22,25)=22;$$

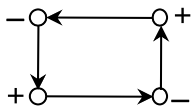
$$max\, x_{24}=min(20,23)=20;$$

$$max\, x_{31}=min(22,25,18)=18.$$

4. Теперь для всех свободных переменных с отрицательной ценой циклов вычислим характеристику $y_{ij}=max\, x_{ij}$.

i j	2; 1	2; 4	3; 1
y	– 4	– 2	– 2
x	22	20	18
y_x	– 88	– 40	– 36

Выберем ту свободную переменную, которой соответствует наименьшее значение величины $y_{ij} = max\, x_{ij}$ и перенесем $max\, x_{ij}$ единиц груза по циклу, соответствующему выбранной переменной, получим:



Сток Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10 22	7 9	6	8	31
A_2	5	6 25	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Таким образом, мы уменьшим значение целевой функции стоимости плана перевозок на 88 единиц.

Новому улучшенному плану перевозок будет соответствовать следующая таблица перевозок:

Сток Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7 31	6	8	31
A_2	5 22	6 3	5 23	4	48
A_3	8	7	6 18	7 20	38
Заявки	22	34	41	20	117

Полученный на данном этапе первой итерации алгоритма новый план перевозок имеет шесть базисных клеток в соответствующей ему транспортной таблице, что позволяет его использовать без модификаций для дальнейшего решения задачи. Стоимость найденного плана перевозок равна:

$L = 31 \cdot 7 + 22 \cdot 5 + \ldots = 708$, что несколько лучше результата опорного плана.

Попытаемся опять улучшить найденный опорный план перевозок. Для этого перейдем к пункту 2 алгоритма.

Вычислим цену цикла для каждой свободной переменной:

$\gamma_{13} = 6 - 5 + 6 - 7 = 0$; $\gamma_{14} = 8 - 7 + 6 - 5 + 6 - 7 = 1$; $\gamma_{11} = 10 - 7 + 6 - 5 = 4$;

$\gamma_{24} = 4 - 7 + 6 - 5 = -2$; $\gamma_{31} = 8 - 5 + 5 - 6 = 2$; $\gamma_{32} = 7 - 6 + 5 - 6 = 0$.

Найдем $max\, x_{24} = min(20, 23) = 20$.

Для единственной свободной переменной рассчитаем значение критерия $y_{ij} = max\, x_{ij} : y_{24} = (max\, x_{24}) = -40$.

Сток / Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7 / 31	6	8	31
A_2	5 / 22	6 / 3	5 / 23	4	48
A_3	8	7	6 / 18	7 / 20	38
Заявки	22	34	41	20	117

Перенесем $max\, x_{24} = 20$ единиц груза по циклу $(2,4)^{+}$, $(3,4)^{-}$, $(3,3)^{+}$, $(2,3)^{-}$, уменьшив этим значение целевой функции на 40 единиц. Новому улучшенному плану будет соответствовать следующая таблица перевозок:

Сток / Исток	B_1	B_2	B_3	B_4	Запасы
A_1	10	7 / 31	6	8	31
A_2	5 / 22	6 / 3	5 / 3	4 / 20	48
A_3	8	7	6 / 38	7	38
Заявки	22	34	41	20	117

Полученный новый план перевозок имеет шесть базисных клеток в соответствующей ему транспортной таблице, что позволяет его использовать без модификаций для дальнейшего решения задачи. Стоимость найденного плана перевозок равна: $L = 31 \cdot 7 + 22 \cdot 5 + \ldots = 668$. Попытаемся опять улучшить найденный опорный план перевозок. Для этого перейдем к пункту 2 алгоритма: вычислим цену цикла для каждой свободной переменной:

$\gamma_{13} = 6 - 5 + 6 - 7 = 0$; $\gamma_{14} = 8 - 4 + 6 - 7 = 3$; $\gamma_{11} = 10 - 7 + 6 - 5 = 4$;

$\gamma_{34} = 7 - 6 + 5 - 4 = 2$; $\gamma_{31} = 8 - 5 + 5 - 6 = 2$; $\gamma_{32} = 7 - 6 + 5 - 6 = 0$.

Так как не существует циклов свободных переменных с отрицательной ценой, полученный план перевозок улучшить нельзя, так как он является оптимальным. Стоимость этого плана перевозок, как было посчитано ранее, составляет 668 единиц.

5.4. Оптимальные назначения

При организации дорожно-строительных работ важно место занимает распределение подразделений по объектам и видам работ. Необходимо отметить, что задача о назначении есть полностью вырожденная транспортная задача. Действительно, при любом назначении машины на работу автоматически «поставки» по строке совпадают со «спросом» по столбцу и вместо $2n-1$ получаем всего n ненулевых значений x_{ij}. Следовательно, нужно дополнить матрицу $(n-1)$ некоторыми, достаточно малыми, но неизвестными величинами для того, чтобы избавиться от вырожденности. Однако это требует громоздких вычислений и гораздо удобнее пользоваться специальными методами, разработанными для такого типа задач. Рассмотрим некоторые из них.

5.4.1. Решение стандартной задачи о назначении по алгоритму Флада

Отличительной особенностью таких задач является то, что ресурсы и объемы работ выражаются в различных единицах измерения (машино-сменами, человеко-днями, кубометрами, тоннами и т. д.). Это не позволяет для их решения использовать рассмотренные выше методы.

Пусть имеются m машин, которые могут выполнять различные m работ. Известна эффективность C_{ij} *i*-й машины при выполнении *j*-й работы. Необходимо определить, какую машину и на какую работу следует назначить, чтобы добиться минимального общего количества машино-смен, необходимых для выполнения всего комплекса работ при условии, что каждая машина может быть назначена только на одну работу.

Для составления математической модели обозначим через x_{ij} назначение *i*-й машины на *j*-ю работу. Так как количество машин равно количеству работ, то каждая из них может быть назначена только на одну работу, поэтому $\sum_{i=1}^{m} x_{ij} = 1$ и $\sum_{j=1}^{m} x_{ij} = 1$. При назначении *i*-й машины на *j*-ю работу эффективность равна $a_{ij}x_{ij}$.

Таким образом, приходим к следующей постановке задачи линейного программирования:

$$x_{i,j} = x_{i,j}^2, \ \ i, j = 1, 2, .. m; \qquad (1)$$

$$\sum_{i=1}^{m} x_{ij} = \sum_{j=1}^{m} x_{ij} = 1; \qquad (2)$$

$$T = \sum_{i,j}^{m} a_{ij}^{(0)} x_{ij} = min. \qquad (3)$$

Условие (1) – один означает, что величина, x_{ij} принимает только два значения: либо 1, если i-я машина назначается для выполнения j-й работы; либо 0, если i-я машина не назначается для выполнения j-й работы.

Условие (2) – отражает вид итоговой матрицы назначений, то есть в каждой строке и в каждом столбце может быть только одна единица.

В условии (3) – индекс (0) означает, что значение стоимостной характеристики a_{ij} взято из исходной матрицы.

Из постановки задачи понятно, что решением является квадратная матрица. Если в исходных данных количество машин больше количества работ, то вводят либо фиктивную работу с нулевой производительностью, либо вторую часть ограничений записывают в виде неравенств.

Алгоритм М. Флада основан на двух теоремах. Первая утверждает, что решение задачи не изменится если к любому столбцу (строке) исходной матрицы добавить (отнять) некую константу. Этим способом ряд величин $a_{ij}^{(0)}$ сводится к нулю, что обеспечивает минимизацию решения. Вторая теорема определяет условие оптимальности решений, получаемых на очередном шаге, состоящее в том, что минимальное число линий содержащих все нули, равно максимальному числу таких нулей. При этом никакие два из них не лежат на одной и той же прямой. Из комбинаторики известно, что для n^2 матрицы возможно получить $n!$ вариантов решения. Алгоритм Флада позволяет получить оптимальное решение задачи о назначении за сравнительно малое число шагов.

Пример. Рассмотрим оценочную таблицу (матрицу) отражающую затраты машино-смен на выполнение отдельных работ.

Исполнители, i	*Работы*, j				
	1	2	3	4	5
	Трудоемкость, машино-смен				
1	2	9	2	7	2
2	6	8	7	6	1
3	4	6	5	3	1
4	4	2	7	3	1
5	5	3	9	5	1

Анализируя данные таблицы, заметим, что есть возможность вычесть из всех строк минимальный элемент равный 1. В результате получим таблицу:

1	8	1	6	1
5	7	6	5	0
3	5	4	2	0
3	1	6	2	0
4	2	8	4	0

Проведем вычислительные действия с новой таблицей: вычтем минимальный элемент равный единице (1) из ненулевых элементов 1-го, 2-го, 3-го и 5-го столбцов, а из 4-го вычтем два (2). В результате получим новую таблицу:

<table>
<tr><td>~~0~~</td><td>~~7~~</td><td>~~0~~</td><td>~~4~~</td><td>|~~0~~</td></tr>
<tr><td>4</td><td>6</td><td>5</td><td>3</td><td>|0</td></tr>
<tr><td>~~2~~</td><td>~~4~~</td><td>~~3~~</td><td>~~0~~</td><td>|~~0~~</td></tr>
<tr><td>~~2~~</td><td>~~0~~</td><td>~~5~~</td><td>~~0~~</td><td>|~~0~~</td></tr>
<tr><td>3</td><td>1</td><td>7</td><td>2</td><td>|0</td></tr>
</table>

Проведем наименьшее из возможных число горизонтальных и вертикальных линий, пересекающих, по крайней мере, один раз все нули. Таких линий получилось четыре (4).

В теоретическом обосновании алгоритма Флада доказано, что если число таких линий меньше n, то полученная матрица не дает оптимального решения.

Для дальнейших действий выбираем наименьший элемент, через который не проведена ни одна линия. В данном случае это единица во втором столбце. Следующим шагом вычитаем это значение из всех элементов, через которые не проведена ни одна линия, и прибавляем его ко всем остальным элементам, через которые проведены две линии. Таким образом, получим:

<table>
<tr><td>~~0~~</td><td>|~~7~~</td><td>~~0~~</td><td>|~~4~~</td><td>|~~1~~</td></tr>
<tr><td>3</td><td>|5</td><td>4</td><td>|2</td><td>|0</td></tr>
<tr><td>2</td><td>|4</td><td>3</td><td>|0</td><td>|1</td></tr>
<tr><td>2</td><td>|0</td><td>5</td><td>|0</td><td>|1</td></tr>
<tr><td>2</td><td>|0</td><td>6</td><td>|1</td><td>|0</td></tr>
</table>

Проверяем полученный вариант на оптимальность, легко заметить, что для этого опять достаточно четырех линий, то есть полученное решение не является оптимальным. Наименьшее из не вычеркнутых значений равно два. Повторим действия, аналогичные предыдущему шагу и получим новый вариант назначений:

~~0~~	~~7~~	***~~0~~***	~~6~~	~~3~~
1	5	2	2	***0***
0	4	1	***0***	1
0	***0***	3	0	1
0	0	4	1	0

Проверим на оптимальность и установим, что для вычеркивания всех нулей в таблице требуется минимально 5 линий, то есть их число равно числу строк (столбцов) матрицы, что определяет условие оптимальности полученного варианта назначения машин на определенные виды работ (в таблице клетки назначений содержат выделенные курсивом элементы – нули). Заметим, что полученное решение не является единственным. Существуют еще три варианта выбора из имеющихся в таблице 11 нулей необходимых пяти так, чтобы в каждом столбце был только один ноль.

Полученное решение приведено в таблице:

Исполнители, i	*Работы, j*				
	1	2	3	4	5
	Трудоемкость, машино-смен				
1			1		
2					1
3				1	
4		1			
5	1				

и дает значение целевой функции: $T = 5 + 2 + 2 + 3 + 1 = 13$ (машино-смен).

При отыскании других вариантов назначений заметим, что в эти решения должны войти нули, стоящие клетках последней таблицы как единственные в третьем столбце и во второй строке.

Если в подобных задачах требуется найти максимум целевой функции, например, наибольшую эффективность или выработку при использовании машин, то оценочная матрица $a_{i\,j}^{(0)}$ преобразуется по правилу:

$$a_{i\,j}^{(0+)} = a_{i\,jmax}^{(0)} - a_{i\,j}^{(0)},$$

где $a_{i\,jmax}^{(0)}$ – наибольшее из значений $a_{i\,j}$ стоящее в исходной оценочной матрице. Далее задача решается в соответствии с приведенным алгоритмом Флада и обеспечит поиск максимума целевой функции.

5.4.2. Решение задачи о назначении венгерским методом

Алгоритм был разработан и опубликован Гарольдом Куном в 1955 г. Сам Кун дал алгоритму название «венгерский», потому что он был в значительной степени основан на более ранних работах двух венгерских математиков: Денеша Кёнига и Эйгена Эгервари. В 1957 г. Джеймс Манкрес показал, что этот алгоритм работает за (строго) полиномиальное время (т. е. за время порядка полинома, не зависящего от величины стоимостей). Поэтому в литературе данный алгоритм известен не только как «венгерский», но и как «алгоритм Куна-Манкреса». Впрочем, в 2006 г. выяснилось, что точно такой же алгоритм был изобретён за 100 лет до Куна немецким математиком Карлом Якоби. Дело в том, что его работа «About the research of the order of a system of arbitrary ordinary differential equations», напечатанная посмертно в 1890 г., содержавшая помимо прочих результатов и полиномиальный алгоритм решения задачи о назначениях, была написана на латыни, а её публикация прошла незамеченной среди математиков.

Пример. По условиям предыдущей задачи, но с иной платежной матрицей решим задачу о назначении *венгерским методом*. Исходная таблица (матрица):

2	4	**1**	3	3
1	5	4	1	2
3	5	**2**	2	4
1	4	3	1	4
3	**2**	5	3	5

Этап 1. В каждой строке ищем минимальный элемент (выделяем жирным в таблице) и отнимаем от всех элементов строки. Получим:

1	3	**0**	2	2
0	4	3	0	**1**
1	3	0	0	2
0	3	2	**0**	3
1	**0**	3	1	3

Теперь проводим аналогичную процедуру для всех столбцов: ищем наименьший элемент по столбцу (выделяем жирным в таблице) и отнимаем его из всех элементов столбца.

Получим:

1	3	0	2	1
0	4	3	0	0
1	3	0	0	1
0	3	2	0	2
1	0	3	1	2

Целью является распределение всех подлежащих назначению единиц в клетки с нулевой стоимостью.

Этап 2. Выбираем строку с одним нулем (строка № 1), выделяем нуль жирным и зачеркиваем оставшиеся нулевые значения этого столбца (столбца № 3).

Далее выбираем строку с одним нулевым значением (строка № 5), выделяем нуль. Опять выбираем строку с одним нулем (строка № 3), выделяем нуль жирным и зачеркиваем (выделено серым) оставшиеся нулевые значения этого столбца (столбца № 4).

Аналогично, выбираем строку с одним нулем (строка № 4), выделяем нуль жирным и зачеркиваем (выделено серым) оставшиеся нулевые значения этого столбца (столбца № 1).

И, наконец, выбираем строку с одним нулевым значением (строка № 2), выделяем нуль, получим:

1	3	**0**	2	1
0	4	3	0	**0**
1	3	0	**0**	1
0	3	2	0	2
1	**0**	3	1	2

Получаем оптимальную матрицу назначений:

		1		
				2
			2	
1				
	2			

Значение целевой функции – стоимость (рациональность, время работ и т. д.) такого назначения составит: $1+1+2+2+2=8$. Как видно венгерский метод позволил получить оптимальное решение с меньшим количеством итераций и без дополнительных проверок на оптимальность.

5.4.3. Задача о нахождении оптимальной последовательности работ

Вопросы об определении последовательности выполняемых работ можно решить распределительными методами, однако более эффективным является специальный метод, носящий название «алгоритм Флада для распределительных задач».

Рассмотрим применение алгоритма на примере конкретной задачи.

Имеется несколько объектов на которых открыт фронт работ. Сначала выполняются земельные работы, затем обустроить дорожные одежды. Время, требуемое на выполнение каждого вида работ на каждом из объектов приведены в таблице.

Объект, i	Время выполнения земельных работ, t_i	Время обустройства дорожных одежд, T_i
1	3	6
2	7	2
3	4	7
4	5	3

Предполагается, что работы можно развернуть на любом объекте. Для простоты положим, что время перемещения техники с объекта на объект несравнимо мало по отношению ко времени выполнения работ, поэтому им можно пренебречь. Заметим, что данное допущение не является обязательном для применения метода по алгоритму Флада.

Теоретически для распределительных задач установлено общее число возможных вариантов последовательностей строительства объектов. Оно равно числу перестановок $n!$. В рассматриваемом примере это значение составляет $4! = 24$ варианта. Какой из них наиболее удачный или оптимальный?

На рис. 5.1 показана последовательность работ в порядке возрастания номера объекта.

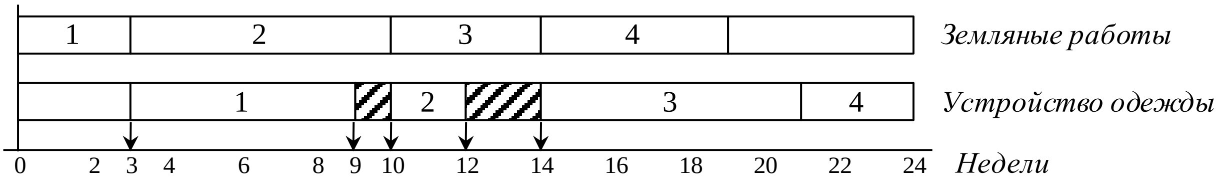


Рис. 5.1. Схема выполнения произвольной последовательности работ

Общее время выполнения всех работ составляет 24 недели и, очевидно, не является оптимальным. Применим алгоритм Флада.

Шаг 1. Находим в таблице исходных данных объект с минимальным временем работы $t_{i\,min}$ или $T_{i\,min}$. В пример это объект № 2 с $T_2 = 2$ недели. Так как это наименьшее время относится ко времени постройки дорожной одежды, то этот объект ставят на последнее место по земельным работам. Смысл такого действия вполне понятен – после завершения на объекте № 2 земельных работ устройство дорожных одежд будет закончено в кратчайший срок. Имеет место минимальный период свертывания работ. Если бы наименьшее время относилось ко времени выполнения земельных работ, то данный объект следовало бы поставить на первое место при разворачивании всех работ. Смысл такого действия также очевиден.

Шаг 2. Продолжаем анализ остальных объектов. Следующие $t_{i\,min}$ или $T_{i\,min}$ со значениями 3 недели относятся к объекту № 1 – время земляных работ и объекту № 4 – время устройства дорожной одежды, можно расставлять любой из них в последовательность. Объект № 1 – на первое место, а объект № 4 на предпоследнее (так как последнее место уже отведено объекту № 2).

Шаг 3. Выполняя аналогичные действия, определяем всю последовательность работ: 1 – 3 – 4 – 2. Результатом является ликвидация простоев подразделений из-за отсутствия фронта работ, что приводит к оптимальному времени в 21 неделю. Экономия составила 3 недели, что существенно. На рисунке 5.2. показана оптимальная последовательность ввода объектов.

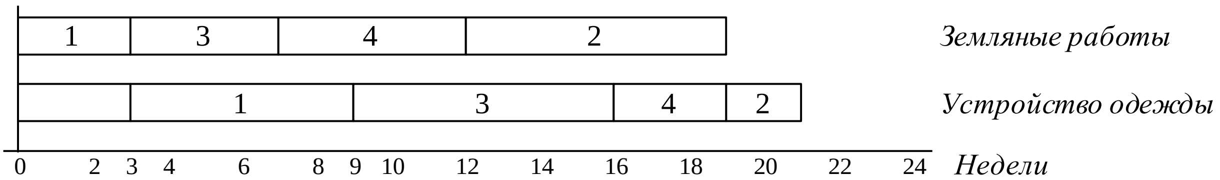


Рис. 5.2. Оптимальная последовательность строительства объектов

Если исключить из постановки задачи допущение о незначительности времени переброски техники с объекта на объект, то следует этот параметр включить (равномерно) во временные показатели t_i и T_i.

Алгоритм Флада может применяться и для более сложных случаев, так если видов работ больше двух, то возможно их различное чередование.

Раздел 6. МЕТОДЫ И МОДЕЛИ ПРИНЯТИЯ ОПТИМАЛЬНЫХ РЕШЕНИЙ

6.1. Элементы теории графов

Теория графов – это область дискретной математики, особенностью которой является геометрический подход к изучению объектов. Обычно эту теорию относят к топологии (потому что во многих случаях рассматриваются лишь топологические свойства графов), однако она пересекается со многими разделами математики: теории множеств, комбинаторной математики, алгебры, геометрии, теории матриц, теории игр, математической логики и многих других математических дисциплин. Основной объект теории – граф и его обобщения.

Исторически задачи теории графов были связаны с решением математических развлекательных задач и головоломок (задача о Кенигсбергских мостах, задача о расстановке ферзей на шахматной доске, задачи о перевозках, задача о кругосветном путешествии и другие). Одним из первых результатов в теории графов явился критерий существования обхода всех ребер графа без повторений, полученный Л. Эйлером при решении задачи о Кенигсбергских мостах.

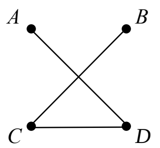
Вот пересказ отрывка из письма Эйлера от 13 марта 1736 году: «Мне была предложена задача об острове, расположенном в городе Кенигсберге и окруженном рекой, через которую перекинуто 7 мостов. Спрашивается, может ли кто-нибудь непрерывно обойти их, проходя только однажды через каждый мост. И тут же мне было сообщено, что никто еще до сих пор не смог это проделать, но никто и не доказал, что это невозможно. Вопрос этот, хотя и банальный, показался мне, однако, достойным внимания тем, что для его решения недостаточны ни геометрия, ни алгебра, ни комбинаторное искусство. После долгих размышлений я нашел легкое правило, основанное на вполне убедительном доказательстве, с помощью которого можно во всех задачах такого рода тотчас же определить, может ли быть совершен такой обход через какое угодно число и как угодно расположенных мостов или не может».

Если на плоскости задать конечное множество V точек и конечный набор линий X, соединяющих некоторые пары из точек V, то полученная совокупность точек и линий будет называться *графом* $G=(V,X)$. При этом элементы множества V называются *вершинами* графа, а элементы множества X – *ребрами*. Во множестве V могут встречаться одинаковые элементы, ребра, соединяющие одинаковые элементы называются *петлями*. Одинаковые пары во множестве X называются *кратными* (или параллельными) ребрами. Количество одинаковых пар (v, w) в X называется *кратностью* ребра (v, w).

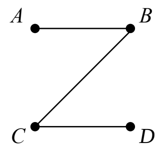
Если пары в наборе X являются упорядоченными, то граф называется ориентированным или *орграфом*. Графу соответствует геометрическая конфигурация. Вершины обозначаются точками (кружочками), а ребра – линиями, соединяющими соответствующие вершины.

<u>Пример</u>. Неориентированные графы имеют множество вершин $\{A, B, C, D\}$. Множество их ребер заданы отношением инцидентности: каждое ребро представлено как пара вершин. Установим соответствие каждому графу его изображение.

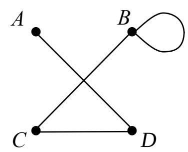
$$\{(A,D),(B,C),(C,D)\}$$



$$\{(A,B),(B,C),(C,D)\}$$



$$\{(A,D),(B,C),(C,D),(B,B)\}$$



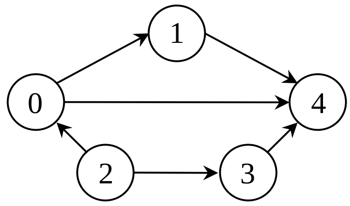
Если $x = \{v, w\}$ – ребро графа, то вершины v, w называются *концами ребра* x. Если $x = (v, w)$ – дуга орграфа, то вершина v – *начало*, а вершина w – *конец дуги* x.

Вершины v, w графа $G = (V, X)$ называются *смежными*, если $\{v, w\} \in X$. Два ребра называются *смежными*, если они имеют общую вершину. *Степенью* вершины графа называется число ребер, которым эта вершина принадлежит. Вершина называется *изолированной*, если ее степень равна единице и *висячей*, если ее степень равна нулю.

Графы $G_1 = (V_1, X_1)$ и $G_2 = (V_2, X_2)$ называются *изоморфными,* если существует взаимно однозначное отображение $\varphi : V_1 \to V_2$, сохраняющее смежность.

Маршрутом (путем) для графа $G = (V, X)$ называется последовательность $v_1\, x_1\, v_2\, x_2\, v_3 \ldots x_k\, v_{k+1}$. Маршрут называется *замкнутым*, если его начальная и конечная точки совпадают. Число ребер (дуг) маршрута (пути) графа называется *длиной* маршрута (пути).

<u>Пример.</u> Для ориентированного графа, изображенного на рисунке:



полный путь может иметь вид L_1: $0 \to 1 \to 4$ или L_2: $0 \to 2 \to 3 \to 4$.

Незамкнутый маршрут (путь) называется *цепью*. Цепь, в которой все вершины попарно различны, называется *простой цепью*. Замкнутый маршрут (путь) называется *циклом (контуром)*. Цикл, в котором все вершины попарно различны, называется *простым циклом.*

Пусть $D=(V,X)$ – орграф, где $V=\{v_1,\ldots,v_n\}$, $X=\{x_1,\ldots,x_m\}$.

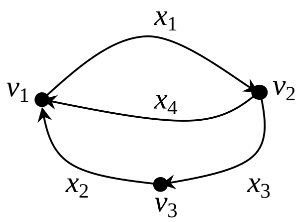
Матрицей смежности орграфа D называется квадратичная матрица $A(D)=(a_{i\,j})$ порядка n, у которой:

$$a_{i\,j}=\begin{cases}1, & \text{если } (v_i,v_j)\in X\\ 0, & \text{если } (v_i,v_j)\notin X\end{cases}$$

Если вершина v является концом ребра x, то говорят, что вершина и ребро v и x – *инциндентны. Матрицей инциндентности* орграфа D называется матрица размерности $n\times m$ $B(D)=(b_{i\,j})$, у которой:

$$b_{i\,j}=\begin{cases}1, & \text{если вершина } v_i \text{ является концом дуги } x_j;\\ -1, & \text{если вершина } v_i \text{ является началом дуги } x_j;\\ 0, & \text{если вершина } v_i \text{ не инцидентна дуге } x_j.\end{cases}$$

Пример. Записать матрицы смежности и инцидентности для графа, изображенного на рисунке.



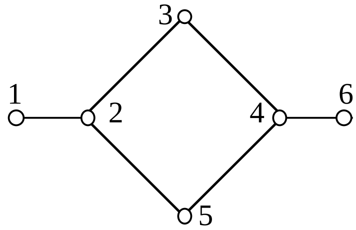
Составим матрицу смежности:

	v_1	v_2	v_3
v_1	0	1	0
v_2	1	0	1
v_3	1	0	0

и матрицу инциндентности: $B(D)=\begin{pmatrix}-1 & 0 & 1 & 1\\ 1 & -1 & 0 & -1\\ 0 & 1 & -1 & 0\end{pmatrix}.$

С помощью матриц смежности и инциндентности всегда можно полностью определить граф и все его компоненты.

Пример. Описать в виде матрицы инцидентности неориентированный граф, представленный на рисунке. Вершины графа пронумерованы и обозначены кружками.



Решение. Решение задачи начинается с построения сетки размером (6, 6), т. е. по числу вершин графа. 1 строка соответствует 1-й вершине графа, которая связана только с вершиной 2 (т. е. $a_{12} = 1$, а остальные элементы строки равны нулю. 2 вершина (2 строка) имеет связи с вершинами 1, 3, 5, поэтому $a_{21} = a_{23} = a_{25} = 1$, остальные элементы второй строки нулевые. Аналогично заполняются другие строки матрицы.

	1	*2*	*3*	*4*	*5*	*6*
1		1	0	0	0	0
2	1		1	0	1	0
3	0	1		1	0	0
4	0	0	1		1	1
5	0	1	0	1		0
6	0	0	0	1	0	

В матрице инцидентности единица означает только наличие связи между двумя вершинами графа. Однако такая матрица позволяет отобразить и интенсивность связи. На рисунке 6.2 показан фрагмент сети автомобильных дорог с указанием протяженности участков.

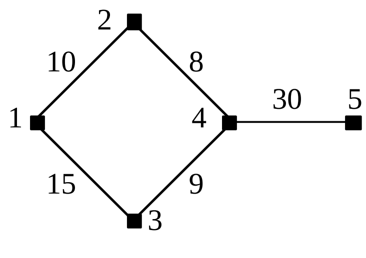


Рис. 6.1. Фрагмент сети автомобильных дорог

Ему соответствует матрица инцидентности, где a_{ij} соответствуют длине участка дороги.

$$\begin{pmatrix} 0 & 10 & 15 & 0 & 0 \\ 10 & 0 & 0 & 8 & 0 \\ 15 & 0 & 0 & 9 & 0 \\ 0 & 8 & 9 & 0 & 30 \\ 0 & 0 & 30 & 30 & 0 \end{pmatrix}$$

Поскольку граф неориентированный, матрица является симметричной. В нем наличие связи отображается не единицей в ненулевой клетке, а цифрой, соответствующей длине участка дороги.

Пример. Имеется 6 специализированных подразделений, которые необходимо объединить в комплексную структуру. Признаками объединения (кластеризации) являются технологическая зависимость ($a_{Tз1k}$) и территориальная близость подразделений ($a_{Tб1k}$). Технологическая зависимость отображена на рисунке 6.2.

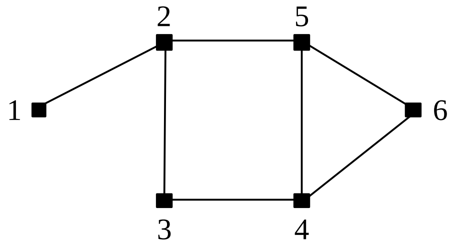


Рис. 6.2. Граф технологической зависимости подразделений

Ему соответствует матрица инцидентности $A_{Tз}$:

$$A_{Tз} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

Территориальная близость (удаление подразделений друг от друга в процессе работ в км) представлена в матрице $Б_{TБ}$ (расстояния указаны в км).

$$Б_{TБ} = \begin{pmatrix} 0 & 2 & 1 & 5 & 5 & 10 \\ 2 & 0 & 6 & 8 & 8 & 10 \\ 1 & 6 & 0 & 4 & 5 & 8 \\ 5 & 8 & 4 & 0 & 6 & 5 \\ 5 & 8 & 5 & 6 & 0 & 4 \\ 10 & 10 & 8 & 5 & 4 & 0 \end{pmatrix}, \qquad Б^{H}_{TБ} = \begin{pmatrix} 0 & 0.2 & 0.1 & 0.5 & 0.5 & 1 \\ 0.2 & 0 & 0.6 & 0.8 & 0.8 & 1 \\ 0.1 & 0.6 & 0 & 0.4 & 0.5 & 0.8 \\ 0.5 & 0.8 & 0.4 & 0 & 0.6 & 0.5 \\ 0.5 & 0.8 & 0.5 & 0.6 & 0 & 0.4 \\ 1 & 1 & 0.8 & 0.5 & 0.4 & 0 \end{pmatrix}$$

Элементы матрицы $Б_{ТБ}$ имеют размерность в км. Для расчетов ее удобно привести к нормированному виду, представив $a_{Тб1k}$ в долях единицы. Для этого все значения $a_{Тб1k}$ следует разделить на максимальное значение удаленности подразделений ($a_{Тб1.6} - a_{Тб2.6} - a_{Тб6.1} - a_{Тб6.2} - 10$ км). В результате образуется нормированная матрица $Б^{Н}_{ТБ}$ ($a_{1.1} = 0$; $a_{1.2} = 2/10 = 0.2$; ... $a_{5.3} = 5/10 = 0.5$ и т. д.).

Поскольку в один кластер надо объединять подразделения, технологически зависимые и расположенные вблизи друг от друга, надо сформировать обобщенный показатель «близости» подразделений ($a_{\sum ik} = a_{ТЗik} + a_{Тбik}$), т. е. надо провести сложение матриц.

Суммой матриц *A* и *B* называется матрица *C*, в которой каждый элемент C_{ik} есть сумма элементов a_{ik} и b_{ik}. Она обозначается: $C = A + B$.

В рассматриваемом примере матрица $C = A_{ТЗ} + Б^{Н}_{ТБ}$ может быть представлена в развернутом виде:

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0.2 & 0.1 & 0.5 & 0.5 & 1 \\ 0.2 & 0 & 0.6 & 0.8 & 0.8 & 1 \\ 0.1 & 0.6 & 0 & 0.4 & 0.5 & 0.8 \\ 0.5 & 0.8 & 0.4 & 0 & 0.6 & 0.5 \\ 0.5 & 0.8 & 0.5 & 0.6 & 0 & 0.4 \\ 1 & 1 & 0.8 & 0.5 & 0.4 & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} 0+0 & 1+0.2 & 0+0.1 & 0+0.5 & 0+0.5 & 0+1 \\ 1+0.2 & 0+0 & 1+0.6 & 0+0.8 & 1+0.8 & 0+1 \\ 0+0.1 & 1+0.6 & 0+0 & 1+0.4 & 1+0.5 & 0+0.8 \\ 0+0.5 & 0+0.8 & 1+0.4 & 0+0 & 0+0.6 & 1+0.5 \\ 0+0.5 & 1+0.8 & 1+0.5 & 0+0.6 & 0+0 & 1+0.4 \\ 0+1 & 0+1 & 0+0.8 & 1+0.5 & 1+0.4 & 0+0 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & 1.2 & 0.1 & 0.5 & 0.5 & 1 \\ 1.2 & 0 & 1.6 & 0.8 & 1.8 & 1 \\ 0.1 & 1.6 & 0 & 1.4 & 1.5 & 0.8 \\ 0.5 & 0.8 & 1.4 & 0 & 0.6 & 0.5 \\ 0.5 & 0.8 & 1.5 & 0.6 & 0 & 1.4 \\ 1 & 1 & 0.8 & 1.5 & 1.4 & 0 \end{pmatrix}.$$

Суммарная матрица отображает «силу связи» подразделений по обобщенному критерию. Например, наиболее связными являются элементы 5 и 2 ($a_{52} = a_{25} = 1.8$). Следовательно, эти подразделения (рис. 6.2) надо в первую очередь объединять в единую структуру. Следующими сильно связными подразделениями являются подразделения 3, 2 ($a_{23} = a_{32} = 1.6$) и т. д. При сложении матриц используются следующие правила:

- суммируемые матрицы должны быть одного типа;
- матрицы разного типа можно привести к однотипным, если в одну из них добавить фиктивную строку или столбец с нулевыми элементами;
- каждый элемент суммируемой матрицы *C* (c_{ik}) и *B* (b_{ik}) получается сложением элементов матриц *A* (a_{ik}) и *B* (b_{ik}) с одинаковыми индексами, т. е. $(a_{ik})_{(m,n)} + (b_{ik})_{(m,n)} = (c_{ik})_{(m,n)}$.

6.2. Динамическое программирование

Динамическое программирование связано в первую очередь с анализом так называемых многошаговых процессов принятия решений. Процесс является одношаговым, если, например, разработана схема оптимального пути из пункта *A* в пункт *B*. Если же вначале известно, как проехать из *A* в первый промежуточный пункт, а там станет известно, как добраться до 2-го промежуточного пункта и т. д., то полностью путь будет известен только по достижении пункта *B* и процесс принятия решений по поводу оптимального пути из *A* в *B* станет многошаговым.

Таким образом, *динамическое программирование* – это метод оптимизации, приспособленный к операциям в которых процесс принятия решений может быть разбит на отдельные этапы.

Как раздел математического программирования динамическое программирование стало развиваться в 50-х годах XX века благодаря работам Р. Белмана и его сотрудников. Впервые такими методами решались задачи оптимального управления запасами. Затем класс задач значительно расширился (многошаговые детерминированные модели задач оптимального распределения ресурсов; расчет развития на перспективу производственной базы стройиндустрии; установление оптимальных режимов замены и изношенного оборудования, производства и хранения продукции во времени, при меняющемся спросе на нее, рациональная загрузка транспортных средств, оптимальное распределение капиталовложений и др.). Большую и практически важную группу моделей динамического программирования составляют задачи календарного планирования строительного производства, оптимизации сроков выполнении этапов работ для минимизации себестоимости их выполнения и т. д.

В основе метода динамического программирования лежит *принцип оптимальности*, который может быть сформулирован следующим образом: чтобы получить оптимальное решение, надо руководствоваться правилом – каков бы ни был путь достижения исследуемой системой некоторого состояния, последующие решения должны принадлежать оптимальной стратегии для остающейся части пути, начинающейся с данного состояния.

Общая постановка задачи динамического программирования. Введем обозначения S_0 – начальное состояние системы; S_K – конечно состояние системы; U – допустимое управление системой (решение по управлению системы); $Z = f(S_0, U)$ – целевая функция.

Требуется определить допустимое управление системой U, приводящее систему из начального состояния S_0 в конечное S_K, при котором целевая функция Z достигает максимума или минимума.

Условия, которым должна удовлетворять задача, описываемая моделью динамического программирования:

1. Задача должна интерпретироваться, как n-шаговый процесс управления, а показатель эффективности процесса должен быть представлен в аддитивной форме, например, как сумма показателей эффективности на любом шаге.

2. Структура задачи (или алгоритм решения) должна быть инвариантна относительно числа шагов n, т. е. должна быть определена для любого n и не зависеть от него.

3. На каждом шаге состояние системы определяется конечным числом переменных состояния и управляется конечным числом переменных уравнения.

4. Выбор управления на k-м шаге не влияет на предшествующие шаги, а состояние в начале этого шага есть функция только предшествующего состояния и выбранного на нем управления (отсутствия действия).

Некоторые задачи естественно распадаются на этапы, в других это деление приходится вводить искусственным путем.

Динамическое программирование есть поэтапное планирование многошагового процесса, при котором на каждом этапе оптимизируется только один шаг. При этом управление на любом шаге должно выбираться с учетом всех его последствий в будущем. Среди всех шагов существует один, который может планироваться попросту, «без оглядки на будущее». Это последний шаг, единственный из всех, который можно планировать так, что бы он как таковой приносил наибольшую выгоду. Спланировав оптимальным образом этот последний шаг, можно к нему «пристраивать» предпоследний, к предпоследнему – еще один шаг и т. д. Процесс динамического программирования всегда разворачивается в обратном по времени направлении: от конца к началу.

Принцип оптимальности Беллмана – решение на любом шаге выбирается таким образом, чтобы обеспечить максимальную эффективность на данном шаге и на всех последовательных шагах.

Решение задач динамического программирования обычно включает два цикла:

- от последнего шага к первому (обратная прогонка, или условная оптимизация);
- от первого шага к последнему (прямая прогонка, или безусловная оптимизация).

В цикле условной оптимизации для любого шага находится множество возможных состояний системы в начале данного шага. Для каждого из этих состояний находится условно оптимальное решение, т. е. решение оптимальное для данного состояния.

Поиск условно оптимальных решений начинается с последнего шага. В цикле безусловной оптимизации для каждого шага определяется безусловно оптимальное решение. Поиск безусловно оптимальных решений начинается с первого шага, т. к. для него известно начальное состояние.

<u>Пример</u>. Необходимо организовать перевозку строительных грузов из пункта 1 в пункт 7, используя дорожную сеть, показанную на рисунке 6.1.

Перевозка будет осуществляться большегрузным транспортом, в связи с чем участки дорог между узловыми пунктами потребуют дооборудования. Время в днях, которое потребуется для дооборудования, показано на схеме (как весовые коэффициенты ребер). Необходимо определить маршрут для перевозки грузов, время дооборудования которого будет наименьшим.

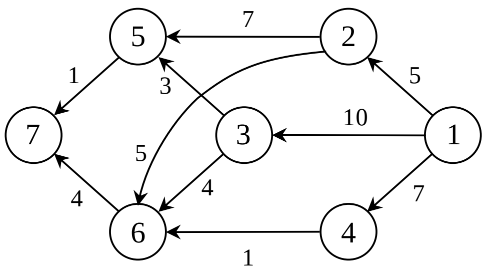


Рис. 6.1. Граф дорожной сети, используемая при перевозке строительных грузов из пункта 1 в пункт 7

Таблица 6.1. Исходные данные к задаче динамического программирования

S	J	C_{ij}	S	J	C_{ij}
1	2	5	3	6	4
	3	10	4	6	1
	4	7	5	7	1
2	5	7	6	7	4
	6	5			
3	5	3			

В таблице 6.2 показана реализация процедур задачи динамического программирования.

Таблица 6.2. Расчет модели динамического программирования

S	J	C_{ij}	$F_{n-1}(s)$	$\int_n (s) = C_{ij} + \int_{n-1}(s)$	$f_n(s)$	$J_n(s)$
$n=1$						
5	7	1	0	1	1	7
6	7	4	0	4	4	7
$n=2$						
2	5	7	1	8	5	5
	6	5	4	9		
3	5	3	1	4	5	5
	6	4	4	8		
4	5	7	1	8	6	6
	6	1	4	5		
$n=3$						
1	2	5	8	13		
	3	10	4	14	12	4
	4	7	5	12		

Из таблицы 6.2 видно, что процесс решения начинается с конечного пункта и заканчивается в начальном, а ответ формируется, начиная с исходного пункта. В рассматриваемом примере ответ имеет вид: $1 \to 4 \to 6 \to 7$, при этом время оборудования этого маршрута составит 12 дней.

Динамическое программирование, позволяя на каждом промежуточном этапе разрабатывать схему дальнейшей реализации программы, является весьма универсальным методом. Однако его применение эффективно в задачах с небольшим числом переменных.

Рассмотрим задачу *сетевого планирования и управления*.

Пусть для некоторого комплекса работ установлены оценки для каждой работы на уровне нормативных продолжительностей и срочного режима, а также указаны стоимости.

	Нормативный режим		*Срочный режим*	
	Продолжительность, дни	Стоимость	Продолжительность, дни	Стоимость
(1,2)	3	6	2	11
(1,3)	5	8	3	12
(1,4)	4	7	8	9
(2,5)	10	25	8	30
(3,5)	8	20	6	24
(3,6)	15	26	12	30
(4,6)	13	24	10	30
(5,7)	3	15	6	25
(6,7)	4	10	3	15

Требуется рассчитать: временные характеристики сетевого графика при нормальном режиме работ; найти критический путь; полные резервы времени; временные характеристики сетевого графика при срочном режиме работ; найти критический путь; полные резервы времени; определить стоимость работ.

Решение. Рассчитаем временные характеристики для нормативного режима. К временным характеристикам относятся ранние и поздние сроки наступления события. Ранний срок наступления события рассчитывается по формуле:

$$t_p(j) = max(t_p(i) + t(i\,j)).$$

где $t_p(j)$ – ранний срок наступления предшествующего J события; $t(i\,j))$ – работа.

Для расчета $t_p(j)$ в данном комплексе будем считать, что ранний срок наступления 1-го события равно $t_p(1) = 0$, тогда для последующих событий будем иметь: $t_p(1) = max(t_p(1) = 0)$; $t_p(2) = max(t_p(1) + t_p(1,2)) = 0 + 3 = 3)$ и так далее. Очевидно, завершающее 7-е событие может наступить через 24 дня от начала выполнения всего комплекса работ.

Поздний срок наступления события определяется по формуле:

$$t_n(i) = min(t_n(j) - t(i\,j)).$$

Для расчета $t_n(i)$ для комплекса будем считать, что самый поздний срок наступления 7-го события равен 24 дня, т. е. раннему сроку наступления 7-го события, тогда получим:

$t_n(7) = min(24) = 24$; $t_n(6) = min(t_n(7) - t(5,7)) = 24 - 4 = 20$ и так далее.

Наконец, $t_n(1) = min(6-3;5-5;7-4) = 0$. Полученный результат говорит о том, что расчеты произведены правильно.

Резервы времени определяем как разность между поздними и ранними сроками по формуле: $P(i) = (t_p(j) - t_n(i))$.

В таком случае получим: $P(1) = 0 - 0 = 0$; $P(2) = 6 - 3 = 3$; $P(3) = 5 - 5 = 0$; $P(4) = 7 - 4 = 3$; $P(5) = 16 - 12 = 4$; $P(6) = 20 - 20 = 0$; $P(7) 24 - 24 = 0$. Резервы времени показывают на какое время можно задержать наступление того или иного события, не вызывая опасности срыва выполнения комплекса работ. Те события, которые не имеют резервов времени, находятся на критическом пути.

Критический путь это наиболее продолжительный путь сетевого графика, который ведет к завершению комплекса работ. Находим пути и их длительности для данного комплекса работ:

1. 1-2-5-7 его стоимость: 3 + 10 + 8 = 21.
2. 1-3-5-7 его стоимость: 5 + 9 + 8 = 22.
3. 1-3-6-7 его стоимость: 5 + 15 + 4 = 24.
4. 1-4-6-7 его стоимость: 4 + 13 + 4 = 21.

Критический путь: (1,3) - (3,6) - (6,7). Резервы времени для работ, находящихся на критическом пути равны нулю. (1,3) = 0; (3,6) = 0; (6,7) = 0.

Рассчитаем временные характеристики сетевого графика при *срочном режиме* работ. Ранний срок наступления события рассчитывается по формуле:

$$t_p(j) = max(t_p(j) + t_p(i\,j)).$$

Для расчета $t_p(j)$ для данного комплекса будем считать, что ранний срок наступления 1-го события равно $t_p(1) = 0$, тогда для последующих событий получим:

$t_p(1) = max(t_p(1)) = 0$; $t_p(2) = max(t_p(1) + t_p(1{,}2)) = 0 + 2 = 2$;

$t_p(3) = max(t_p(1) + t_p(1{,}3)) = 0 + 3 = 3$; $t_p(4) = max(t_p(1) + t_p(1{,}4)) = 0 + 8 = 8$;

$t_p(5) = max(t_p(4) + t_p(4{,}5)) = (2+8);(3+6) = 10$;

$t_p(6) = max(t_p(2) + t_p(2{,}5); t_p(3) + t_p(4{,}6)) = (3+12);(8+10) = 18$;

$t_p(7) = max(t_p(5) + t_p(5{,}7); t_p(6) + t_p(6{,}7)) = (15+3);(18+3) = 21$.

Очевидно, завершающее 7-е событие может наступить через 21 день от начала выполнения всего комплекса работ. Поздний срок (фактическая проверка решения) наступления события определяется по формуле: $t_n(i) = min(t_n(j) - t(ij))$ и равен $t_n(1) = min(15-2; 20-8; 8-8) = 0$. Полученный результат говорит о том, что расчеты произведены правильно.

Резервы времени определяем как разность между поздними и ранними сроками по формуле: $P(i) = (t_p(j) - t_n(i))$.

$P(1) = 0 - 0 = 0$; $P(2) = 7 - 2 = 5$;… $P(6) = 18 - 18 = 0$; $P(7) = 22 - 22 = 0$.

Найдем все пути и их стоимости:

1. 1-2-5-7 его стоимость: 3 + 8 + 6 = 16.
2. 1-3-5-7 его стоимость: 3 + 6 + 6 = 15.
3. 1-3-6-7 его стоимость: 3 + 12 + 3 = 18.
4. 1-4-6-7 его стоимость: 8 + 10 + 3 = 21.

Критический путь (1-3-6-7). Его длительность равна 21. Очевидно, что на критическом пути резервов времени нет.

6.3. Выбор кратчайшего пути

Пусть некоторая сеть задана в виде орграфа (рис. 6.2.), т. е. каждой ориентированной дуге соответствует определенное расстояние. Необходимо найти кратчайший путь из *i*-го узла сети в ее заданный *j*-й узел. К этой задаче, известной в исследовании операций как *задача выбора кратчайшего пути*, сводятся такие практически важные задачи, как *задача о замене оборудования*, *задача о календарном планировании комплекса работ* и т. д.

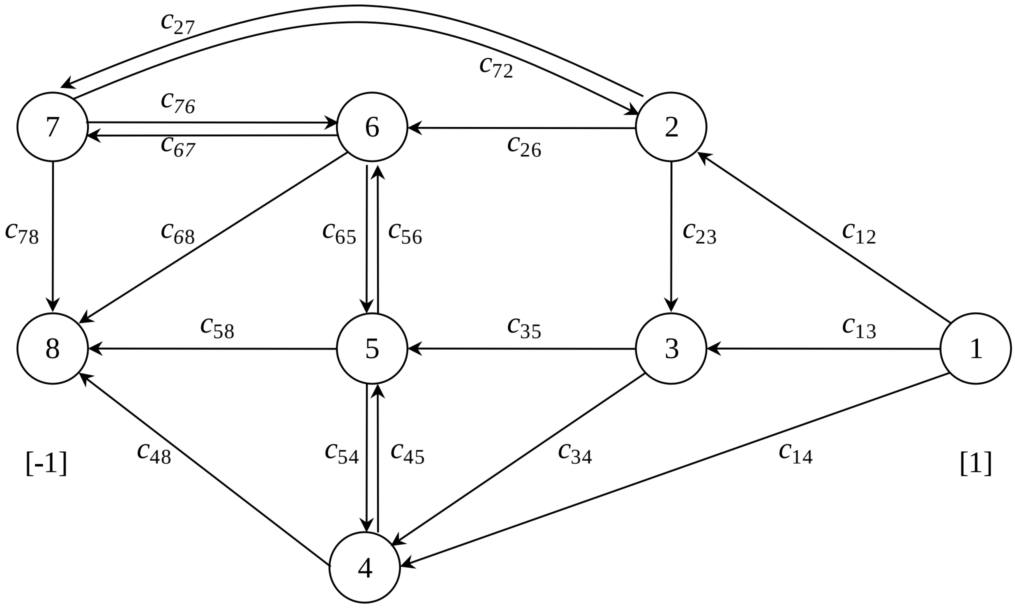


Рис. 6.2. Транспортная сеть

Как правило, в сети выделяют один узел, который является конечным (пункт или станция назначения, *сток*). Задача заключается в отыскании кратчайшего пути в этот конечный узел (на рис. 6.2 конечным является узел с номером 8) из некоторого другого узла сети (например, из первого узла сети на рис. 6.2). Величина c_{ij} определяет расстояние от i-го узла сети до ее j-го узла, которая может измеряться в единицах, отличных от единиц длины.

Так, например, c_{ij} может представлять собой стоимость проезда от i го до j-го узла сети. Тогда задача заключается в отыскании пути минимальной стоимости. Величина c_{ij} может также определять время переезда от i-го до j-го узла сети. При этом необходимо найти путь с минимальной продолжительностью переезда.

При решении прикладных задач, сводящихся к задаче выбора кратчайшего пути, часто встречаются ситуации, когда $c_{ij} \neq c_{ji}$. Кроме того, как правило, не выполняется так называемое неравенство треугольника: $c_{ij} \leq c_{ik} + c_{kj}$ для всех некоторых значений индексов i, j, k.

Существуют сети, содержащие *циклы*, каждый из которых представляет собой замкнутый путь (путь, исходящий из некоторого узла сети и возвращающийся в него же). Так, в сети, представленной на рис. 6.2, много циклов, один из них содержит узлы с номерами 2, 3, 5, 6 и 7. Как правило, в задачах исследования операций значения c_{ij} положительны и общая длина цикла является положительной. Следовательно, решение задачи выбора кратчайшего пути не может содержать циклов.

Итак, для сети необходимо найти кратчайший путь от узла с номером 1 (источник) до узла с номером 8 (сток). Установим связь этой задачи с *классической транспортной задачей*.

Рассмотрим транспортную задачу с промежуточными пунктами, сеть которой представлена на рис. 6.2. При этом предположим, что:

- в узле с номером 1 имеется избыточная единица товара;
- в узле с номером 8 имеется недостаток единицы товара;
- узлы с номерами 2, ..., 7 являются промежуточными пунктами с нулевыми *чистыми запасами* (потребность в дополнительных поставках товара равна нулю).

Необходимо разработать такой план перевозок товара между узлами сети (складами), который при минимальных транспортных затратах позволит на каждом складе поддерживать нулевой чистый запас товара.

Считаем, что каждой ориентированной дуге сети соответствует переменное модели x_{ij}, представляющее собой количество товара, которое должно быть отправлено с i-го склада на j-й.

Для каждого k-го промежуточного пункта вводим переменное x_{kk} с соответствующим ему коэффициентом $c_{kk} = 0$ в целевой функции, а величину обозначаем через T_k. Если множество пар индексов (i, j), соответствующих ориентированным дугам сети, обозначить через J, то рассматриваемую задачу можно записать следующим образом:

$$\begin{cases} \sum\limits_{(i,j)\in J} c_{ij} x_{ij} \to min; \\ \sum\limits_{(i,j)\in J} x_{kj} - \sum\limits_{(i,j)\in J} x_{ik} = T_k; \\ T_1 = 1, T_n = -1, T_k = 0, k = 2, \ldots, n-1; \\ x_{ij} \ge 0, (i, j) \in J. \end{cases}$$

Сформулированная выше задача о нахождении кратчайшего пути эквивалентна классической транспортной задаче.

Вернемся к задаче выбора кратчайшего пути для сети, изображенной на рис. 6.2. Найдем кратчайший путь от узла с номером 1 до узла с номером 8, если $c_{12} = 1$ км, $c_{13} = 4$ км, $c_{14} = 6$ км, $c_{23} = 3$ км, $c_{26} = 5$ км, $c_{27} = 1$ км, $c_{34} = 3$ км, $c_{35} = 5$ км, $c_{45} = 1$ км, $c_{48} = 4$ км, $c_{54} = 1$ км, $c_{56} = 1$ км, $c_{58} = 2$ км, $c_{65} = 1$ км, $c_{67} = 3$ км, $c_{68} = 4$ км, $c_{72} = 1$ км, $c_{76} = 3$ км, $c_{78} = 7$ км.

Здесь мы видим, что кратчайший путь перевозки товара следующий: $1 \to 2 \to 7 \to 6 \to 5 \to 8$. Расстояние перевозки при этом составит 8 км. Аналогично данную задачу можно решить и на максимум, т. е. найти самый длинный путь доставки товара.

<u>Пример.</u> Пусть транспортная сеть состоит из 10 узлов, часть из которых соединены магистралями. Необходимо определить маршрут доставки груза из пункта 1 в пункт 10, обеспечивающий наименьшие транспортные расходы. Модель транспортной сети представлена на рисунке 6.3. Граф сети дорог: вершины обозначены цифрами в кружочках, а стоимости перевозки единицы груза между отдельными пунктами сети, проставлены у соответствующих ребер.

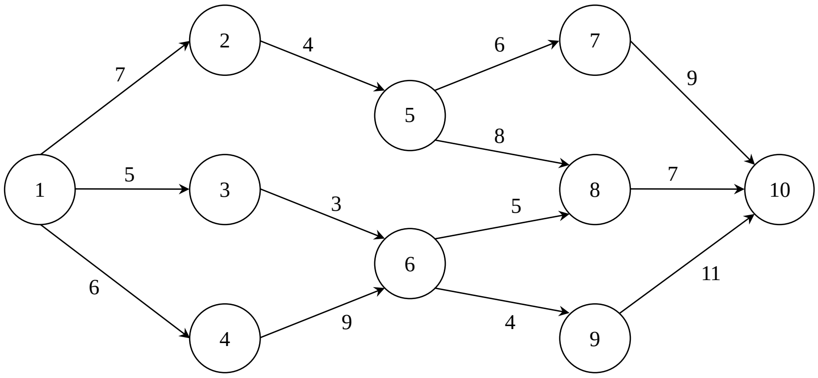


Рис. 6.3. Сеть магистралей

В задаче имеется ограничение – двигаться по стрелкам, изображенным на схеме маршрутам, можно только слева на право, т. е. попав, например, в пункт 7, мы имеем право переместиться только в пункт 10 и не можем возвратиться обратно в 5-й или 6-й. Эта особенность транспортной сети дает право отнести каждый из десяти пунктов к одному из поясов.

Будем считать, что пункт принадлежит k-му поясу, если из него попасть в конечный пункт ровно за k шагов, т. е. с заездом ровно в $(k-1)$-й промежуточный пункт. Таким образом, пункты 7, 8 и 9 принадлежат к первому поясу, 5 и 6 – ко второму, 2, 3 и 4 – к третьему и 1 –к четвертому. Тогда на k-м шаге будем находить оптимальные маршруты перевозки груза из пунктов k-го пояса до конечного пункта. Оптимизацию производим с конца процесса, и потому, дойдя до k-го шага, неизвестно, в каком из пунктов k-го пояса окажется груз, перевозимый из первого пункта.

Обозначим:

k – номер шага (k = 1, 2, 3, 4);

i – пункт, из которого осуществляются перевозки (i = 1, 2, .., 9);

j – пункт, в который доставляется груз (j = 2, 3, .., 10);

$C_{i,j}$ – стоимость перевозки груза из пункта i в пункт j;

$F_k(i)$ – минимальные затраты на перевозку груза на k-м шаге решения задачи из пункта i до конечного пункта.

Очевидно, что минимум затрат на перевозку груза из пунктов k-го пояса до пункта 10 будет зависеть от того, в каком пункте этого пояса мы оказались. Номер i пункта, принадлежащего k-му поясу, будет являться переменной состояния системы на k-м шаге.

Поскольку оптимизация осуществляется с конца процесса, то, находясь в некотором пункте i k-го пояса, принимается решение о перемещении груза в один из пунктов $(k-1)$-го пояса, а направление дальнейшего движения известно из предыдущих шагов. Номер j пункта $(k-1)$-го пояса будет переменной управления на k-м шаге.

Для первого шага управления $(k-1)$ функция Беллмана представляет собой минимальные затраты на перевозку груза из пунктов 1-го пояса в конечный пункт, т. е. $F_1(i) = C_{i\,10}$. Для последующих шагов затраты складываются из двух слагаемых – стоимости перевозки груза $C_{i,j}$ из пункта i k-го пояса в пункт j $(k-1)$-го пояса и минимально возможных затрат на перевозку из пункта j до конечного пункта, т. е. – $F_{k-1}(i)$. Таким образом, функциональное уравнение Беллмана будет иметь вид:

$$F_k(i) = min\{C_{ij} + F_{k-1}(j)\}.$$

Минимум затрат достигается на некотором значении j^*, которое является оптимальным направлением движения из пункта i в конечный пункт.

На четвертом шаге попадаем на 4-й пояс и состояние системы становится определенным $i = 1$. Таким образом, функция $F_4(1)$ представляет собой минимально возможные затраты по перемещению груза из 1-го пункта в 10-й. Оптимальный маршрут определяется в результате анализа всех шагов в обратном порядке, а выбор некоторого управления j на k-м шаге приводит к тому, что состояние системы на $(k – 1)$-м шаге становится определенным.

Приведем решение сформулированной задачи, исходные данные которой приведены на рисунке 6.4 по шагам.

Условная оптимизация

<u>1-й шаг</u>. $k = 1$, $F_1(i) = C_{i10}$. На этом шаге в пункт 10 груз может быть доставлен из пунктов 7, 8 или 9, данные сведем в таблицу:

i, j	*10*	*F₁ (i)*	*j**
7	7	7	10
8	9	9	10
9	11	11	10

<u>2-й шаг</u>: $k = 2$, функциональное уравнение на втором шаге принимает вид:

$$F_2(i) = min\{C_{i\,j} + F_1(j)\}.$$

Все возможные перемещения груза на втором шаге и результаты расчета искомой функции приведены в таблице:

i, j	*7*	*8*	*9*	*F_2 (i)*	*j**
5	6 + 9 = 15	8 + 7 = 15		15	7; 8
8		5 + 7 = 12	4 + 11 = 15	12	7

<u>3-й шаг</u>: $k = 3$, функциональное уравнение на третьем шаге принимает вид:

$$F_3(i) = min\{C_{i\,j} + F_3(j)\}.$$

Возможные перемещения груза на третьем шаге и результаты расчета искомой функции приведены в таблице:

i, j	5	6	*F_3 (i)*	*j**
2	4 + 15 = 19		19	5
3		3 + 12 = 15	15	6
1		9 + 12 = 21	21	6

<u>4-й шаг:</u> $k = 4$: $F_4(i) = min\{C_{i\,j} + F_4(j)\}$.

i, j	2	3	4	F_4 (*i*)	*j**
1	7 + 19 = 26	5 + 15 = 20	6 + 21 = 20	20	3

На этапе условной оптимизации определено, что минимальные затраты на перевозку груза из пункта 1 в пункт 10 составляют $F_4\,(i) = 20$. Данный результат достигается при движении груза из 1-го пункта в 3-й. По данным таблицы из шага 3 необходимо двигаться в пункт 6, затем в пункт 8 и из него в конечный пункт (см. таблицу шага и таблицу шага 1). Таким образом, оптимальный маршрут доставки груза, в виде графа, имеет вид:

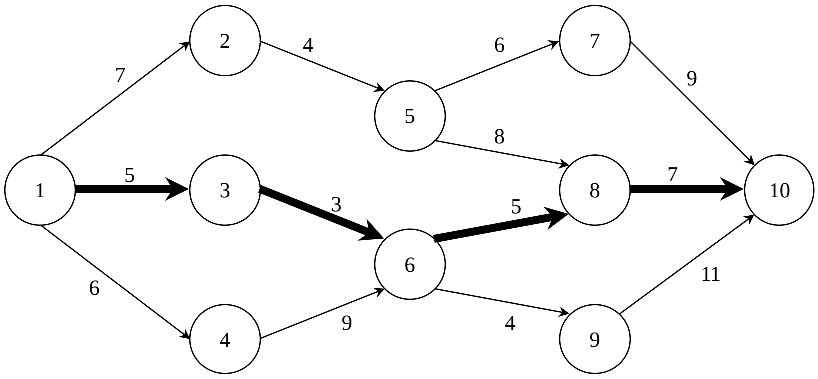


Рис. 6.4. Оптимальный маршрут

Полученный результат свидетельствует о хороших возможностях использования метода графов для решения задач оптимизации.

Раздел 7. МОДЕЛИ В ДОРОЖНО-ТРАНСПОРТНОМ ПЛАНИРОВАНИИ И СТРОИТЕЛЬСТВЕ

7.1. Основные виды и этапы моделирования

Моделирование в научных исследованиях применялось еще в глубокой древности и постепенно охватывало все новые области знаний: техническое конструирование, строительство и архитектуру, астрономию, физику, химию, биологию и, наконец, общественные и экономические науки. Признание практически во всех отраслях современной науки принес методу моделирования двадцатый век. Создание электронно-вычислительных машин, обрабатывающих большие массивы информации, позволило решать задачи, требующие огромного объема вычислений, и разработать численные методы решения новых задач. Появились новые математические дисциплины, в частности *математическое программирование* – раздел математики, занимающийся решением задач нахождения экстремумов функции нескольких переменных с ограничениями на область допустимых значений этих переменных.

Модель (фр. modèle, от лат. modulus – «мера, аналог, образец») – это система, исследование которой служит средством для получения информации о другой системе, это упрощенное представление реального устройства и/или протекающих в нем процессов, явлений. Построение и исследование моделей, то есть моделирование, облегчает изучение имеющихся в реальном устройстве (процессе) свойств и закономерностей. Моделирование является обязательной частью исследований и разработок, неотъемлемой частью жизни, поскольку сложность любого материального объекта и окружающего его мира бесконечна вследствие неисчерпаемости материи и форм ее взаимодействия, – как внутри себя, так и с внешней средой.

Основные этапы моделирования

1. Постановка задачи.

Определение цели анализа и пути ее достижения и выработки общего подхода к исследуемой проблеме. На этом этапе требуется глубокое понимание существа поставленной задачи. Иногда, правильно поставить задачу не менее сложно, чем ее решить. Постановка – процесс не формальный, общих правил нет.

2. Изучение теоретических основ и сбор информации об объекте оригинала.

На этом этапе подбирается или разрабатывается подходящая теория. Если ее нет, устанавливаются причинно-следственные связи между переменными описывающими объект. Определяются входные и выходные данные, принимаются упрощающие предположения.

3. Формализация.

Заключается в выборе системы условных обозначений и с их помощью записывать отношения между составляющими объекта в виде математических выражений. Устанавливается класс задач, к которым может быть отнесена полученная математическая модель объекта. Значения некоторых параметров на этом этапе еще могут быть не конкретизированы.

4. Выбор метода решения.

На этом этапе устанавливаются окончательные параметры моделей с учетом условия функционирования объекта. Для полученной математической задачи выбирается какой-либо метод решения или разрабатывается специальный метод. При выборе метода учитываются знания пользователя, его предпочтения, а также предпочтения разработчика.

5. Реализация модели.

Разработав алгоритм, пишется программа, которая отлаживается, тестируется и получается решение нужной задачи.

6. Анализ полученной информации.

Сопоставляется полученное и предполагаемое решение, проводится контроль погрешности моделирования.

7. Проверка адекватности реальному объекту.

Результаты, полученные по модели сопоставляются либо с имеющейся об объекте информацией или проводится эксперимент и его результаты сопоставляются с расчетными.

Процесс моделирования является итеративным. В случае неудовлетворительных результатов этапов 6 или 7 осуществляется возврат к одному из ранних этапов, который мог привести к разработке неудачной модели. Этот этап и все последующие уточняются и такое уточнение модели происходит до тех пор, пока не будут получены приемлемые результаты.

Требования к моделям

Моделирование всегда предполагает принятие допущений той или иной степени важности. При этом должны удовлетворяться следующие требования к моделям:

➢ адекватность, то есть соответствие модели исходной реальной системе и учет, прежде всего, наиболее важных качеств, связей и характеристик. Оценить адекватность выбранной модели, особенно, например, на начальной стадии проектирования, когда вид создаваемой системы еще неизвестен, очень сложно. В такой ситуации часто полагаются на опыт предшествующих разработок или применяют определенные методы, например, метод последовательных приближений;

- точность, то есть степень совпадения полученных в процессе моделирования результатов с заранее установленными, желаемыми. Здесь важной задачей является оценка потребной точности результатов и имеющейся точности исходных данных, согласование их как между собой, так и с точностью используемой модели;
- универсальность, то есть применимость модели к анализу ряда однотипных систем в одном или нескольких режимах функционирования. Это позволяет расширить область применимости модели для решения большего круга задач;
- целесообразная экономичность, то есть точность получаемых результатов и общность решения задачи должны увязываться с затратами на моделирование. И удачный выбор модели, как показывает практика, – результат компромисса между отпущенными ресурсами и особенностями используемой модели.

По форме представления различают:

- *аналитические* модели. Их решения ищутся в замкнутом виде, в виде функциональных зависимостей. Удобны при анализе сущности описываемого явления или процесса и использовании в других математических моделях, но отыскание их решений бывает весьма затруднено;
- *численные* модели. Их решения – дискретный ряд чисел (таблицы). Эти модели универсальны, удобны для решения сложных задач, но не наглядны и трудоемки при анализе и установлении взаимосвязей между параметрами. В настоящее время такие модели реализуют в виде программных комплексов – пакетов программ для расчета на компьютере. Программные комплексы бывают прикладные, привязанные к предметной области и конкретному объекту, явлению, процессу, и общие, реализующие универсальные математические соотношения (например, расчет системы алгебраических уравнений);
- *формально-логические* информационные модели – это модели, созданные на формальном языке. Например:

модель формальной системы в математике и логике как любая совокупность объектов, свойства которых и отношения между которыми удовлетворяют аксиомам и правилам вывода формальной системы, служащей тем самым совместным (неявным) определением такой совокупности;

модель в теории алгебраических систем как совокупность некоторого множества и заданных на его элементах свойств и отношений;

эталонная модель.

По способу отображения действительности различают три основных вида моделей – эвристические, натурные и математические.

Эвристические модели, как правило, представляют собой образы, рисуемые в воображении человека. Их описание ведется словами естественного языка (например, вербальная информационная модель) и, обычно, неоднозначно и субъективно. Эти модели неформализуемы, то есть не описываются формально-логическими и математическими выражениями, хотя и рождаются на основе представления реальных процессов и явлений.

Натурные модели. Отличительной чертой натурных моделей является их подобие реальным системам (они материальны), а отличие состоит в размерах, числе и материале элементов и т. п. Среди натурных моделей наибольший интерес представляют физические модели. Ими являются реальные изделия, образцы, экспериментальные и натурные модели, когда между параметрами системы и модели одинаковой физической природы существует однозначное соответствие. Выбор размеров таких моделей ведется с соблюдением теории подобия.

Математические модели – формализуемые, представляют собой совокупность взаимосвязанных математических и формально-логических выражений, как правило, отображающих реальные процессы и явления (физические, психические, социальные и т. д.).

Построение математических моделей возможно следующими способами:

- аналитическим путем, то есть выводом из физических законов, математических аксиом или теорем;
- экспериментальным путем, то есть посредством обработки результатов эксперимента и подбора аппроксимирующих (приближенно совпадающих) зависимостей.

Математические модели более универсальны и дешевы, позволяют поставить «чистый» эксперимент (то есть в пределах точности модели исследовать влияние какого-то отдельного параметра при постоянстве других), прогнозировать развитие явления или процесса, отыскать способы управления ими. Результаты математического моделирования нуждаются в обязательном сопоставлении с данными физического моделирования – с целью проверки получаемых данных и для уточнения самой модели. С другой стороны, любая формула – это разновидность модели и, следовательно, не является абсолютной истиной, а всего лишь этап на пути ее познания.

Существующие тенденции в строительстве (усложнение форм и нагрузок, индивидуальность, комбинированность конструкций и материалов), недостаточность нормативной базы, аварийность, принципиальная непредсказуемость разного рода воздействий и их сочетаний, уникальность каждого грунтового основания, геометрии самой конструкции, неопределенность и неполнота знаний о материалах и нагрузках вынуждают инженеров при проектировании проводить все более детальные исследования самих сооружений, их оснований и фундаментов. Разработка и совершенствование современных инженерных методов и математических технологий и их внедрение в сферу проектирования в строительной индустрии проводится с целью повышения надежности, безопасности и экономической эффективности. С помощью высокопроизводительных вычислений стало возможно моделировать поведение грунтовых массивов с использованием моделей грунтов, учитывающих их нелинейные свойства (drucker-prager, mohr-coulomb, cam-clay) на различных этапах строительства в 3-мерной постановке. Безусловно, это все очень вычислительно ресурсоемкие процессы, но их использование насущное требование современной практики.

Зачастую по заказу проектных институтов и коммерческих организаций, методами математического моделирования решаются задачи проектирования зданий, мостов, тоннелей и подземных сооружений. Среди примеров таких работ можно упомянуть экспертные оценки устойчивости моста и расчет напряженно-деформированного состояния подземного комплекса.

Точные расчеты устойчивости, жесткости, прочности, прогрессирующего разрушения транспортных сооружений, аэродинамические расчеты при проектировании мостов и зданий, моделирование устойчивости опор мостов к обтеканию водой требуют учета большого количества неизвестных – до 10 млн. и более. Чем меньше неизвестных учитывает расчет модели, тем менее точен весь расчет.

В условиях нехватки компьютерных ресурсов исследователи вынуждены были ограничивать расчетную схему и представлять многие детали сооружений в упрощенном виде. С другой стороны, именно точность расчета позволяет получать прямую выгоду от использования компьютерного моделирования: не использовать избыточный запас прочности в материалах и сокращать стоимость строительства, прогнозировать долговечность конструкции и планировать профилактические работы, продлевать жизненный цикл сооружений.

В качестве важности построения и исследования моделей в строительных сооружениях приведем в пример историю с Такомским мостом (мост Такома-Нэрроуз).

Крушение произошло в 1940 г. после четырехмесячной эксплуатации. Основная причина аварии – чрезмерные динамические крутильные колебания, вызванные ветром. Для изучения аварии этого моста удачным фактором был тот, что точное поведение моста от начала аварии до момента крушения удалось подробно изучить при помощи киносъемки, зарегистрировавшей колебания и характер разрушения.

Мост висячий (вантовый) трехпролетный. Общая длина 1 662 м, средний пролет 854 м, два боковых – по 335 м, береговой – 137 м, ширина моста 11.9 м. Мост подвешен на двух стальных канатах 0.438 мм каждый. Стрела провеса 70.66 м. Пилоны стальные на бетонных быках. Мост имел очень малую высоту балки жесткости – 2.44 м, что составляло 1/100 пролета, и в связи с этим был подвержен сильным колебаниям.

Обрыв подвесок центрального пролета повлек провисание боковых пролетов и наклон пилонов. Сильные вертикальные и крутильные колебания моста явились следствием чрезмерной гибкости конструкции и относительно малой способности моста поглощать динамические силы. Наблюдались колебания моста при скорости ветра 18.8 м/сек, хотя мост был рассчитан на (статическую) ветровую нагрузку до скорости ветра 50 м/с. Мост был запроектирован и правильно рассчитан на действие статических нагрузок, в том числе и ветровой, но аэродинамическое действие нагрузки не было учтено. Крутильные колебания возникли в результате действия ветра на проезжую часть около горизонтальной оси, параллельной продольной оси моста. Крутильные колебания усиливались вертикальными колебаниями тросов. Опускание троса с одной стороны моста и поднятие его с другой вызвали наклон проезжей части и породили крутильные колебания, которые и привели, в конце концов, к обрушению конструкции.

Результаты производства могут быть значительно улучшены при сбалансированном и адекватном применении экономико-статистических моделей, т. к. данный тип моделирования позволяет выделить причинно-следственные связи, определить какие факторные признаки или показатели производства в большей степени влияют на конечный результат, т. е на результативный признак. Также данный вид моделей используют при ведении оценочных работ при вычислении кадастровой и рыночной стоимости земельных участков.

7.2. Модели экстремального анализа в проектировании дорожного строительства

В транспортном строительстве целью экстремального анализа является минимизация финансовых или материальных затрат, сроков строительства, простоев техники либо максимизация прибыли, темпов строительства, уровня рентабельности производства и т. п. Поиск экстремального значения какого-либо процесса, установление динамики показателей производственной деятельности могут осуществляться различными методами, например методами математической статистики, статистического прогнозирования, дифференциального исчисления и др.

Рассмотрим наиболее распространенные классы задач, решаемых методами экстремального анализа.

1. При формировании комплексных подразделений часто возникает задача равномерной загрузки техники, минимизации потерь от простоев машин. Подобную ситуацию можно проследить на примере комплектования заготовительно-транспортного подразделения, предназначенного для погрузки и транспортирования песка из карьера на объект строительства. Постановка задачи экстремального анализа может быть сформулирована следующим образом: независимо от того, за счет каких простоев строительная организация будет нести потери, суммарные потери должны быть минимальными.

2. При расчете оптимального темпа строительства транспортных объектов приходится решать двойственную задачу. Низкий темп работ требует меньших затрат на строительные машины и заработную плату рабочих (за счет их численности), но создает риск несвоевременного завершения строительства и применения заказчиком штрафных санкций. Наоборот, привлечение дорогостоящей техники (или увеличение ее численности) повышает стоимость строительства, но уменьшает риск, штрафных санкций. Более того, досрочная сдача объекта в эксплуатацию дает экономический эффект (дополнительную прибыль). Отсюда задача выбора оптимального темпа строительства сводится к нахождению такого темпа работ, при котором суммарные издержки строительной организации от уплаты штрафов в дополнительных затрат на строительство будут минимальными. Математическая постановка задачи представляет собой целевую функцию, в которой минимизируются общие издержки строительной организации, представляющие собой сумму затрат на обеспечение принятого темпа работ и штрафов за несвоевременную сдачу объекта.

3. Обеспечение строительства материалами и конструкциями требует создания производственных запасов. Повышенные запасы гарантируют ритмичность строительства, но требует больших затрат на создание и эксплуатацию складского хозяйства. В тоже время при повышенных запасах становятся меньше расходы на их создание, поскольку крупная разовая поставка материалов дешевле, чем частые поставки мелкими партиями. Здесь имеет место та же принципиальная постановка задачи, в которой целевая функция представляет собой сумму частных функций.

Приведенные задачи не исчерпывают всего многообразия экономических ситуаций в транспортном строительстве, оптимизируемых с помощью методов экстремального анализа. Наиболее сложным моментом в решении задач с использованием экстремального анализа является построение математической модели (целевой функции), которую предстоит дифференцировать. Все дальнейшее сводится к выполнению обычной вычислительной процедуры классической математики.

Математическая модель дорожного полотна для оценки собственных частот

Для оценки собственных частот (одной из основных характеристик процесса колебаний) можно построить математическую модель колебательной системы, состоящей из абсолютно твердой пластины, укрепленной на пружинах. Упругие свойства системы «слой асфальтобетона – подушка из щебня» моделируются приведенным коэффициентом жесткости.

Математическая модель представлена уравнениями Лагранжа второго рода в обобщенных координатах. Рассматриваются малые колебания горизонтальной прямоугольной пластины массой m, размерами a и b относительно главных центральных осей. Толщина пластины бесконечно малая. Пластина абсолютно твердая, опирается своими углами на четыре одинаковые пружины жесткости c. Система имеет три степени свободы: поступательное перемещение вдоль вертикальной оси z, вращательные – вокруг осей x и y.

Уравнения Лагранжа второго рода для полученной системы:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{z}}\right)-\left(\frac{\partial T}{\partial z}\right)=Q_z;\quad \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\varphi}_x}\right)-\left(\frac{\partial T}{\partial \varphi_x}\right)=Q_x;\quad \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\varphi}_y}\right)-\left(\frac{\partial T}{\partial \varphi_y}\right)=Q_y,$$

где z, φ_x, φ_y – обобщенные координаты системы;

$\dot{z}$, $\dot{\varphi}_x$, $\dot{\varphi}_y$ – обобщенные скорости; T – кинетическая энергия пластины;

Q_x, Q_y, Q_y – обобщенные силы.

С учетом значений моментов инерции и выражений для обобщенных сил уравнения Лагранжа второго рода применяют следующий вид:

$$\ddot{z}+\frac{4C}{m}z=0;\quad \ddot{\varphi}_x+\frac{12c}{m}\varphi_x=0;\quad \ddot{\varphi}_y+\frac{12c}{m}\varphi_y=0.$$

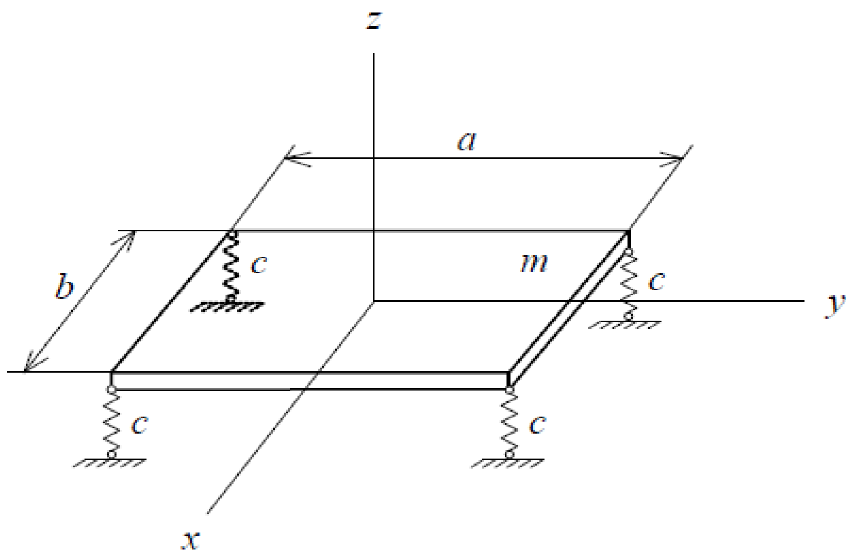


Рис. 7.1. Колебательная система, состоящая из абсолютно твердой пластины, укрепленной на пружинах

Собственные частоты колебаний: $k_z=\sqrt{\frac{4c}{m}}$; $k_x=k_y=\sqrt{\frac{12c}{m}}$.

Математическая модель поверхности дороги при воздействии импульсных возмущений

При дефектах поверхности дороги типа прямого синусоидального коноида имеют место в зависимости от скорости транспорта и длины волн различного рода импульсные возмущения в структуре дороги. При определении движения системы под действием периодических импульсов необходимо учитывать внутреннее трение в системе. Учет частотно-независимого внутреннего трения в задачах о свободных колебаниях диссипативных систем реализуется с использованием гипотезы комплексной жесткости:

$$m\ddot{z}(t)+(a+ib)cz(t)=0,$$

где $z(t)$ – комплексное перемещение; i – мнимая единица.

$$a=\frac{1-\alpha^2}{1+\alpha^2};\quad b=\frac{2}{1+\alpha^2};\quad \alpha=\frac{\gamma}{2},$$

где γ – коэффициент внутреннего трения, связанный с коэффициентом потерь:

$$n=\frac{\gamma}{1-\gamma^2/4}.$$

Подстановкой $z = Ae^{pl}$ получается характеристическое уравнение:

$$p^2 + (a + ib)p_0^2 = 0, \quad \text{где } p_0 = \sqrt{\frac{c}{m}}.$$

Оно дает два корня $p = \pm i(1 + i\alpha)p^*$, где $p^* = (1 + \alpha^2)^{-1/2} p_o$.

Тогда $z = (A - iB)e^{-\alpha pt} e^{ipt}$, откуда вещественное решение $z = Rez$ равно $z = e^{-\frac{\gamma}{2}pt}(A\cos pt + B\sin pt)$.

Круговая частота затухающих колебаний $p < p_0$ и логарифмический декремент δ, не зависящий от частоты, равны соответственно:

$$p = \frac{p_0}{\sqrt{1 + \gamma^2/4}}; \quad \delta = \pi\gamma.$$

При начальных условиях $z(0) = z_0$, $\dot{z}(0) = v_0$.

$$z = e^{-\frac{\gamma}{2}pt}\left[z_0 \cos pt + \left(\frac{v_0}{p} + \frac{\gamma z_0}{2}\right)\sin pt\right].$$

После приложения импульса S_0 к неподвижной системе в момент $t = 0$ при начальных условиях $z_0 = 0$, $v_0 = \frac{S_0}{m}$ получим $z = \frac{S_0}{mp} e^{-\frac{\gamma}{2}pt}$.

Для конечного числа $(n+1)$ периодических импульсов с периодом T_0 решение строится наложением функций с разными началами отсчета времени:

$$z_n = \frac{S_0}{mp}\sum_{r=0}^{n} e^{-\frac{\gamma}{2}p(t - rT_0)} \sin p(t - rT_0),$$

где z_n – перемещение, которое достигается спустя n периодов T_0, так что время t, отсчитываемое от момента приложения первого импульса, заключено в пределах $nT_0 \le t \le (n+1)T_0$, $n = 0$ соответствует одному импульсу.

Вводя относительное время $t^* = \frac{t - nT_0}{T_1}\left(0 \le t^* \le 0\right)$, получаем:

$$z_n = \frac{S_0}{mp} e^{-\gamma\pi t^*}\left(A_n \sin 2\pi t^* + B_n \cos 2\pi t^*\right),$$

где: $A_n \sum_{k=0}^{n} e^{-bk} \cos a k = \frac{e^{-b} - \cos a - e^{-nb}\cos(n+1)a + e^{-(n+1)b}\cos na}{2(chb - \cos a)}$;

$$B_n \sum_{k=0}^{n} e^{-bk} \sin a k = \frac{\sin a - e^{-nb}\sin(n+1)a + e^{-(n+1)b}\sin na}{2(chb - \cos a)};$$

$$a = 2\pi\theta; \quad \theta = \frac{T_0}{T_1}; \quad b = \gamma\pi\theta; \quad n - r = k.$$

При небольших значениях n решение описывает неустановившиеся колебания системы:

$$z_n^{max} = \frac{S_0}{mp_0} e^{-\gamma \pi t_0^*} \sqrt{A_n^2 + B_n^2} \,,$$

где t_0^* – наименьшее положительное значение: $t_0^* = \dfrac{1}{2\pi} arctg \dfrac{2A_n - \gamma B_n}{2B_n + \gamma A_n}$.

При малой диссипации $\gamma \le 0.1$ получаем $z_n^{max} = \dfrac{S_0}{mp_0} \sqrt{A_n^2 + B_n^2}$.

Глобальный максимум устанавливается из n значений z_n^{max}.

При $\dfrac{T_0}{T_1} = N$ наступает импульсный резонанс:

$$z_n^{max} \approx \frac{S_0}{mp_0} \frac{e^{-\gamma \pi N(n+1)}}{e^{-\gamma \pi N}} .$$

Наибольший из $z_n^{\max}$ соответствует $N = 1$.

При достаточно больших значениях n колебания будут практически установившимися.

В случае $n \to \infty$: $z_{рез}^{max} = \dfrac{S_0}{mp_0} \dfrac{1}{1 - e^{-\gamma \pi N}}$.

Собственные частоты колебаний могут быть определены аналитически на основе теории колебаний фундаментальных и континуальных систем.

Полученные значения собственных частот могут использоваться для определения действительной части комплексной жесткости гетерогенной структуры дорожного полотна по Сорокину для исследования резонансных режимов колебаний при непрерывном движении транспортных средств и условий возникновения импульсного резонанса при воздействии мгновенных периодических импульсов в результате торможений и разгонов.

Проектирование мостового перехода при определении расчетного расхода по методу, основанному на гамма-распределении

Для определения величины расчетного расхода можно использовать методики, основанные на гамма-распределении и нормальном законе распределения фактических максимальных расходов. От величины расчетного расхода зависит глубина, общий и местный размыв в реке, величина подпоров перед сооружением и подходными насыпями, величина отверстия моста.

Требуемую величину отверстия моста устанавливали по формулам:

$$l_M = B_{KP} \cdot \frac{Q_p}{Q_{KP}} \cdot \frac{1}{(P_{hpдоп})^{1+x}},$$

где Q_{KP} – расход в коренном русле, м^3/с;

x – показатель степени, зависящий от среднего диаметра частиц грунта, слагающего дно реки;

B_{KP} – ширина коренного русла, м;

$P_{hpдоп}$ – допустимый коэффициент общего размыва.

$$P_{hp} = \left(\frac{Q_p}{Q_{KP}}\right)^{\frac{1}{1+x}}.$$

При данной длине отверстия моста определена длина моста, подходных насыпей и регуляционных сооружений.

Произведены измерения величины отверстия моста, как в сторону уменьшения, так и в сторону увеличения и промеры глубины, коэффициент общего размыва, расход в коренном русле, расходы на поймах.

В результате этих вычислений, установлено, что действительно с уменьшением отверстия моста уменьшалась его стоимость, но увеличивалась стоимость регуляционных сооружений и подходов к мосту, а с увеличением отверстия моста относительно требуемой величины отверстия, стоимость моста намного превышала стоимость подходов и регуляционных сооружений.

Такие зависимости проявили себя как при определении стоимости мостового перехода, определенной при расчетном расходе, с использованием таблицы Фостера-Рыбкина, так и при расходе определенном с использованием метода, основанном на гамма-распределении. Как в том, так и другом методе получена кривая описывающая суммарную стоимость мостового перехода.

Оптимальное отверстие мостового перехода, запроектированного под расчетный расход установленный с использованием таблиц Фостера-Рыбкина равно 136 м. При использовании функции гамма-распределения точка оптимума соответствует 131 м. При этом стоимость строительства мостового перехода при оптимальных длинах отверстий моста имело следующее значение (в относительных единицах): при $l_{опт.} = 136$ м, $S_{\min} = 0.53$; при $l_{опт.} = 131$ м, $S_{\min} = 0.48$.

Результаты сравнений показывают, что не всегда минимальная стоимость оптимального отверстия сооружения соответствует традиционным методам определения расчетного расхода при проектировании мостового перехода. Эти результаты зависят от многих факторов, включая скорости течения воды в коренном русле и на поймах, величин расчетных расходов, которые установлены одним и другим методами и так далее. Поэтому при оценке эффективности капитальных вложений следует учитывать, что все методы дают величину расчетного расхода, с определенной погрешностью. И, следовательно, для повышения надежности предполагаемого результата необходимо устанавливать среднее значение расчетного расхода, используя для этого результаты расчета по тем существующим методам, которые хорошо согласуются с эмпирическим распределением.

Технико-экономическое сравнение строительства мостового перехода показало существование оптимальных длин отверстия моста, при которых стоимость мостового перехода (подходных насыпей, регуляционных сооружений и моста) наименьшая. Наименьшая стоимость мостового перехода, как правило, соответствует среднему значению расчетного расхода, установленному по тем существующим и предлагаемым методам, которые хорошо согласуются с эмпирическим распределением. Средняя величина расчетного расхода в этом случае, как правило, соответствует экономически эффективному отверстию моста.

7.3. Дискретные модели

Цифровое моделирование (метод перебора)

Методы линейного и динамического программирования дают возможность заменить простой перебор возможных вариантов решений упорядоченным и экономным поиском оптимального результата. Однако существует много технико-экономических задач, важных в практическом отношении, для решения которых нужны иные методы. К таким задачам относятся задачи, где оптимальное решение (поведение, стратегию) надо выбирать в условиях неопределенности исходных данных, когда поведение системы случайно и может быть описано лишь в терминах математической статистики (среднее значение, математическое ожидание, дисперсия, спектр, функция корреляции, законы распределения и т. п.). В этих случаях обычно нельзя указать рациональные аналитические методы решения, и поэтому такие задачи решаются методом перебора.

Одним из простейших и, пожалуй, наиболее распространенных методов оптимизации является метод сканирования. Сущность этого метода состоит в следующем.

Пусть в процессе моделирования производственной ситуации, по которой необходимо принять решение, получена символьная модель вида:

$$W = f(c_i, v_j),$$

где W – общий критерий функционирования;

c_i – множество управляемых переменных;

v_j – множество неуправляемых переменных;

f – соотношение, связывающее управляемые и неуправляемые переменные.

Чтобы получить желаемое решение, нужно определить значения управляемых переменных, максимизирующие или минимизирующие критерии функционирования системы W. Обычно для получения решения задачи поступают таким образом. Сначала устанавливают диапазон возможных изменений управляемых переменных c_i. Затем для дальнейших исследований используются управляемые переменные c_i, которые удовлетворяют системе определенных ограничений. Для этих значений вычисляются значения целевой функции W. В качестве решения задачи принимаются значении c_i, при которых целевая функция принимает экстремальные значения. Достоинством метода является не только простота его реализации на ЭВМ, но и принципиальная применимость к решению многих практических задач, возможность получения глобального экстремума. Основной недостаток – большие затраты времени, особенно в связи с возрастанием размерности задачи.

Модель определения среднего числа автомашин, находящихся под загрузкой асфальтобетоном

Рассмотрим асфальтобетонный завод, содержащий N смесителей, каждый из которых может обслуживать только одну машину. Для получения асфальтобетона на завод приезжают машины, загружаются асфальтобетонной смесью и уезжают. Если в момент приезда автомашины все смесители заняты отпуском асфальтобетона, то она ожидает начала обслуживания. В момент освобождения смесителя из очереди на обслуживание заезжает очередной автосамосвал. Будем предполагать, что дисциплина очереди, то есть порядок загрузки автомобилей асфальтобетоном, в рассматриваемом случае роли не играет.

Формальный прогресс отпуска асфальтобетона представляют собой дискретную цепь Маркова с пространством состояний $E = (0, 1, \ldots)$. Такты процесса обслуживания соответствуют шагам цепи Маркова. Состояние $k \in E$

означает наличие k автомобилей под загрузкой на заводе. За каждый такт с вероятностью a приезжает новая машина и с вероятностью $q = min(k,\, n)$ завершается загрузка автомобиля, где $u > 0$ – интенсивность загрузки одним смесителем; k – число автомобилей, стоящих под загрузкой.

Описанная модель является дискретной цепью Маркова размножения и гибели, так как на каждом шаге цепь или остается в том же состоянии k, или переходит в соседние состояние $k+1$ или $k-1$. Поэтому здесь имеет место система уравнений:

$$a_k = \sum_{i=1}^{\infty} a_i p_{i,k}\,,\ k = 1,\, 2,\ldots,$$

которая в данном случае имеет вид:

$a_0 = (1 - p_0)a_0 + q_1 a_1;\quad a_k = (0 - p_k - q_k)a_0 + q_{k+1}a_{k+1} + p_{k-1}a_{k-1};\quad k = 1,\, 2,\ldots,$

где $P_k = a$ для всех $k \in E$.

В частных случаях, для $k = 1,\, 2,\ldots,n$:

$$\frac{p_0 p_1 \cdots p_{k-1}}{q_0 q_1 \ldots q_k} = \alpha^k \frac{1}{\mu \cdot 2\mu \cdot \ldots \cdot k\mu} = \frac{1}{k!}\left(\frac{\alpha}{\mu}\right)^k;$$

для $k = n+1,\, n+2,\ldots,n$:

$$\frac{p_0 p_1 \cdots p_{k-1}}{q_0 q_1 \ldots q_k} = \alpha^k \frac{1}{\mu \cdot 2\mu \cdot \ldots \cdot (n-1)\mu \cdot n\mu \cdot \ldots \cdot n\mu} = \frac{1}{n!}\left(\frac{\alpha}{\mu}\right)^k n^{n-k} = \frac{1}{n!} n^n r^k.$$

Значит,

$$\sum_{k=1}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_0 q_1 \ldots q_k} = \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{\alpha}{\mu}\right)^k + \frac{1}{n!} n^n \sum_{k=n+1}^{\infty} r^k = \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{\alpha}{\mu}\right)^k + \frac{1}{n!}\left(\frac{\alpha}{\mu}\right)^k \frac{r}{1-r}.$$

Если $r < 1$ рассматриваемая сумма конечна. Параметр r назовем коэффициентом загрузки смесителей асфальтобетонного завода. Следовательно, существует стационарный режим, из легко получаемых формул:

$$a_0 = \left(1 + \sum_{k=1}^{\infty} \frac{p_0 p_1 \cdots p_{k-1}}{q_0 q_1 \ldots q_k}\right)^{-1};\quad a_k = \frac{p_0 p_1 \cdots p_{k-1}}{q_0 q_1 \ldots q_k} a_o;\quad k = 1,\, 2,\ldots.$$

Находим выражения для вероятностей состояний:

$$a_0 = \left(1 + \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{\alpha}{\mu}\right)^k + \frac{1}{n!}\left(\frac{\alpha}{\mu}\right)^n \frac{r}{1-r}\right)^{-1},$$

$$a_k = \begin{cases} \frac{1}{k!}\left(\frac{\alpha}{\mu}\right)^k a_0, & k = 1, 2, \ldots, n; \\ \frac{1}{n!}\left(\frac{\alpha}{\mu}\right)^k n^{n-k} a_0, & k = n+1, n+2, \ldots \end{cases}$$

Полученные формулы дают возможность вычислить среднюю длину очереди автомашин, среднее число автомашин, находящихся под загрузкой асфальтобетоном и т. д.

Пример. Пусть $n = 4$; $a = 0.1$; $u = 0.05$. Коэффициент загрузки смесителей асфальтобетона $r = \frac{0.1}{4 \cdot 0.05} = \frac{0.1}{0.2} = \frac{1}{2} = 0.5 < 1$, следовательно, стационарный режим существует. Распределение вероятностей числа автомашин, находящихся на асфальтобетонном заводе под загрузкой и стоящих в очереди, можно представить в виде следующей таблицы:

K	0	1	2	3	4	5	6	7	8	9
a_k	0.130	0.260	0.260	0.173	0.087	0.035	0.012	0.003	0.001	0.0

Модели управления запасами

Модели управления запасами используются при необходимости определения в строительстве объема запаса строительных материалов, конструкций и изделий, характера изменения его в процессе возведения объекта, обновления запаса в связи с поступлением и расходованием ресурсов, с целью обеспечения бесперебойности и надежности строительного процесса при минимальных затратах, связанных с хранением, пополнением, расходованием запаса. Так как уровень спроса неожиданно возникающих потребностей в ресурсах носит чаще всего случайный характер, то модели управления запасами должны быть стохастическими, вероятностными, в упрощенной постановке возможно использование детерминированных моделей. В строительстве чаще всего применяются модели управления складскими запасами. В общем виде экономико-математическая модель управления может быть представлена:

$$З(t) = З_{нач} + P(t) - R(t),$$

где $З(t)$ – текущий уровень запаса материалов на складе в момент времени t;

$З_{нач}$ – начальный запас материалов на складе за время t;

$P(t)$ – поступление материалов на склад за время t;

$R(t)$ – расходование материалов со склада за время t.

Очевидно, что в любой момент запас материалов на складе не может быть отрицательным, то есть: $З(t) > 0$.

Поступление и расходование материалов со склада обычно производится партиями. Обозначив объем поставки через P_i, а объем расходуемой партии R_i, преобразуем исходное соотношение к виду:

$$З(t) = З_{нач} + \sum_{i=1}^{n} P_i - \sum_{j=1}^{m} P_j,$$

где n – количество поставляемых партий стройматериалов;

m – количество расходуемых партий стройматериалов.

Это равенство является базисным в модели управления запасами. В зависимости от того, какие величины (показатели) в нем заданы, а какие являются искомыми, различают разные виды моделей. Часто в модели включают показатели, характеризующие затраты на поставку, хранение, отправку товаров со склада.

Критерием оптимальности моделей управления запасами, как правило, является объем затрат, их минимум (минимум исследуемой функции). В процессе определения экономического содержания затрат учитываются затраты, связанные с заказом каждой новой партии материальных ресурсов; транспортные расходы; расходы на содержание складов и хранение материалов; затраты на складские операции, штрафы и т. д.

Ограничения в задачах управления запасами могут быть самого различного характера. Как правило, они используются для описания предельной величины тех или иных параметров системы (модели). Например, ограничения могут устанавливаться по максимальному объему запасов; максимальной площади, занимаемой складируемыми материалами и конструкциями; максимальной стоимости; средней стоимости числу поставок в заданном интервале времени, максимальному объему и т. д.

7.4. Вероятностные модели

Методы теории вероятностей и математической статистики можно использовать для оценки качества прочности дорожно-строительных материалов, в частности для их исследования на водонасыщение, износ и деформацию. Объективная оценка процесса растрескивания и, на этой основе, прогнозирование срока службы может быть выполнена посредством математического моделирования метеорологических условий, транспортного потока, процесса изменения температуры и влажности конструкции, процессов изменения напряженно-деформированного состояния при проезде транспортных средств и изменения температуры покрытия. Статистической основой для построения вероятностных моделей метеорологических условий могут быть результаты многолетних наблюдений на метеорологических станциях и постах, а также обобщенные данные публикации в справочниках по климату.

Разрушение цементобетонных покрытий является процессом постепенного накопления повреждений. Основными параметрами, определяющими характеристики процесса накопления повреждений являются: параметры транспортного потока, параметры метеорологических условий, параметры процесса влагонакопления и температурного режима конструкции дорожной одежды и земляного полотна.

Анализ экспериментальных исследований прочности цементобетона при циклическом нагружении показывает, что наиболее существенными параметрами, определяющими долговечность материала являются: число циклов нагружения, уровень напряженности и отношение минимального напряжения в цикле к максимальному.

Можно создать трехуровневую математическую модель метеорологических условий температуры воздуха, облачности, количество осадков, влажности воздуха, скорость ветра, позволяющую получать реализации параметров с осреднением на месячном, суточном и часовом уровнях. Возможно применения имитационной модели транспортного потока для получения последовательности транспортных средств, интервалов между ними, загрузки транспортных средств и ее распределение по осям и колесам, скорости и поперечного положения на плите в момент проезда исследуемого сечения.

В качестве примера рассмотрим исследование N образцов асфальтобетонной смеси на водонасыщение. Для определения закона распределения случайной величины ξ (водонасыщения) нужно результаты исследования представить в виде сгруппированного статистического ряда.

Интервалы водонасыщения для частот

Интервал водонасыщения	Частота
$[x_i - 1, x_i)$	k_i
$[x_0, x_1)$	k_1
$[x_1, x_2)$	k_2
…	…
$[x_n - 1, x_n)$	k_n
Сумма $\sum_{i=1}^{n} k_i = N$	

Выдвигаем гипотезу H, состоящую в том, что распределение водонасыщения имеет определенный вероятностный закон распределения. В качестве этого закона чаще всего рассматривают нормальный закон распределения с параметрами a и σ, равными выборочному среднему $\bar{x}$ и выборочному среднему квадратическому отклонению S, так как объем выборки N считается большим, например больше 50.

Чтобы проверить гипотезу H, состоящую в том, что распределение водонасыщения асфальтобетонной смеси имеет нормальный закон распределения с уровнем значимости, равным $\alpha = 0.01$, можно воспользоваться критерием К. Пирсона, или критерием А.Н. Колмогорова. Используем критерий А.Н. Колмогорова.

Для проверки выдвинутой гипотезы H все вспомогательные расчеты, необходимые для вычисления выборочной статистики $\lambda = D\sqrt{N}$, сводят в таблицу:

Вычисление выборочной статистики

Интервалы изменения наблюдаемых значений случайной величины ξ, $[x_i - 1, x_i)$	Частота k_i	Нормированные интервалы $[u_i - 1, u_i)$	Значение выборочной функции распределения для правого конца интервала $[u_i - 1, u_i)$	Теоретическая функция распределения $F(u_i) = 0.5 + 0.5\,\Phi(u_i)$	$\Delta_n = \lvert F^*(u_i) - F(u_i) \rvert$
Менее x_0	0	$(-\infty, u_1)$	0	$F(u_1)$	Δ_1
$[x_0, x_1)$	k_1	$[u_1, u_2)$	$F^*(u_2)$	$F(u_2)$	Δ_2
$[x_1, x_2)$	k_2	$[u_2, u_3)$	$F^*(u_3)$	$F(u_3)$	Δ_3
…	…	…	…	…	…
$[x_n - 1, x_n)$	k_n	$[u_n - 1, u_n)$	$F^*(u_n)$	$F(u_n)$	Δ_n

Сумма $N = \sum_{i=1}^{n} k_i$. Из последнего столбца таблицы определяют значение выборочной статистики: $D = \max_{U_i} \left| F^*(u_i) - F(u_i) \right|$. Вычисляем наблюдаемое значение выборочной статистики: $\lambda_0 = D\sqrt{N}$.

Из таблицы распределения критерия Колмогорова по заданной вероятности α находим критическое значение λ_α. Критерий для проверки гипотезы H формулируется следующим образом:

если $\lambda_0 < \lambda_\alpha$, то нет оснований для отклонения гипотезы о том, что водонасыщение асфальтобетонной смеси имеет нормальное распределение с параметрами $\bar{x}$ и S. В противном случае выдвинутую гипотезу отклоняют.

Как правило, на практике оказывается, что выполняется условие $\lambda_0 < \lambda_\alpha$, поэтому водонасыщение асфальтобетонной смеси считают нормально распределенной случайной величиной ξ с параметрами $\bar{x}$ и S. Аналогичное исследование можно было провести и для других качественных характеристик асфальтобетонной смеси, а также для других дорожно-строительных материалов, например гравия и щебня.

Из вышеизложенного следует, что для качественных характеристик дорожно-строительных материалов можно строить доверительные интервалы с наперед заданными надежностями их оценки. При этом доверительные интервалы, соответствующие одному и тому же коэффициенту доверия, получаются самыми короткими, так как статистические оценки для качественных характеристик дорожно-строительных материалов являются эффективными.

Математическая модель влагопроводности грунта в дорожных конструкциях

Дорожная конструкция, как элемент окружающей среды, находится под постоянным воздействием природно-климатических условий, формирующих ее водно-тепловой режим. Воздействие водно-теплового режима проявляется в попеременном увлажнении и просыхании грунтов земляного полотна, их промерзании и оттаивании, под влиянием которых снижается прочность и повышается деформационная способность дорожной конструкции. Поэтому все инженерные решения главным образом направляются на защиту земляного полотна от влагонакопления. При этом следует отметить, что в условиях резкоконтинентального климата, где годовой перепад температуры воздуха достигает 37.7^{o}C, а перепад температуры покрытий и того выше, защитные свойства капитальных покрытий от поверхностных вод значительно снижается вследствие наличия температурных трещин, которые способствуют к пучинообразования.

Известно, что влажность W земляного полотна находится в функциональной зависимости от суммарного воздействия климата, почво-грунтов, рельефа местности, грунтовых и поверхностных вод. В тоже время, на величину влажности существенное влияние оказывают и конструктивные особенности дорожной одежды и земляного полотна, т. е. водопроницаемость покрытий, пористость нижних слоев дорожной одежды, коэффициент уплотнения земляного полотна и др. Эту зависимость можно выразить формулой (1):

$$W = f\left(\sum Г_i \sum D_i\right), \tag{1}$$

где $Г_i$ – суммарного воздействия географического комплекса;

D_i – суммарный дорожный комплекс.

Таким образом, исследование влагонакопления в грунтах земляного полотна должно основываться на совместном учете геокомплекса и дорожной конструкции. Известно, что расчетная влажность W_P всегда будет находиться в пределах:

$$W_{MAX} \geq W_P \geq W_{ОПТ}; \tag{2}$$

$$W_{MAX} = W_{П.В.} - v, \tag{3}$$

где $W_{ОПТ}$ – оптимальная влажность грунта, %;

$W_{П.В.}$ – полная влагоемкость грунта, %;

v – объем защемленного воздуха в порах грунта, %.

Наиболее важной задачей для принятия оптимального проектного решения являются определение прочностные и деформационные свойства местных грунтов в широком диапазоне плотности, влажности и их фильтрационной способности. Исходя из классификации В.М. Сиденко типов водно-теплового режима (2), для рассматриваемого региона установлены следующие типы водно-теплового режима земляного полотна, аналитическое выражение которой имеет вид:

$$\frac{\partial T}{\partial t} = a\frac{\partial^2 T}{\partial z^2} + в\frac{\partial W}{\partial T}; \tag{4}$$

$$\frac{\partial W}{\partial t} = \pm a_1\frac{\partial^2 W}{\partial z^2} + a_1 в_1\frac{\partial^2 T}{\partial z^2}, \tag{5}$$

где T, W – температура и влажность (на глубине z в момент времени t), град, доли единицы;

a, a_1 – соответственно коэффициенты тепло- и влагопроводности грунта, м^2/сутки;

$в$, $в_1$ – коэффициенты, характеризующие теплообмен при фазовых превращениях и термомиграции влаги, град, 1/ч.

Рассмотрим математическую модель диффузно-инфильтрационного типа водно-теплового режима, представленную системой дифференциальных уравнений (1) и (2), как автомодельную задачу.

В этом случае начальные и граничные условия, полученные на основе экспериментальных исследований, можно записать следующим образом:

$$W(Z;0) = W_H; \quad W(0;Tt) = W_H(1 + m_1 t); \quad W(\infty;T) = W_H; \tag{6}$$

$$t(Z;0) = T_H; \quad T(0;t) = T_H(1 - m_2 t); \quad t(\infty;t) = T_H, \tag{7}$$

где W_H, T_H – начальное распределение влажности и температуры по глубине;

m_1, m_2 – коэффициенты, характеризующие интенсивность изменении влажности и температуры во времени. Здесь $[m_1] = 1/ч$ и $[m_2] = град/ч$:

$$m_1 = \frac{W_K - W_H}{W_H T}; \quad m_2 = \frac{T_H - T_K}{T_H t}, \tag{8}$$

где T_K, W_K – значения температуры и влажности в конце влагонакопления при глубине $Z = 0$;

t – время влагонакопления, сутки.

Для решения системы уравнений (1) и (2) при начальных и граничных условиях (3) и (4) проведем преобразование уравнений в виде:

$$\frac{\partial(T - вW)}{\partial t} = \frac{\partial^2(Ta)}{\partial Z^2}; \quad \frac{\partial W}{\partial t} = \frac{\partial^2(a_1 W)}{\partial Z^2} + \frac{\partial^2(Ta_1 в_1)}{\partial Z^2};$$

$$\frac{\partial W}{\partial t} = a_1 \frac{\partial^2(a_1 W)}{\partial Z^2} + a_1 в_1 \frac{\partial^2 T}{\partial Z^2}$$

или

$$\left.\begin{aligned} \frac{\partial(T - вW)}{\partial t} &= \frac{\partial^2(aT)}{\partial Z^2} \\ \frac{\partial W}{\partial t} &= \frac{\partial^2(a_1 в_1 T + a_1 W)}{\partial Z^2} \end{aligned}\right\}, \tag{9}$$

второе уравнение системы (5) умножим на коэффициент A ($A = Const$):

$$\frac{\partial AW}{\partial t} = \frac{\partial^2}{\partial Z^2}\left[Aa_1 в_1 T + a_1 AW\right]. \tag{10}$$

Коэффициент A определим из условия (10) равенство функций:

$$A = \frac{a_1 + вв_1 a_1 - a \pm \sqrt{D}}{2a_1 в_1},$$

где дискриминант квадратного уравнения будет иметь вид:

$$D = \left(a_1 + вв_1 a_1 - a\right)^2 + 4вa > 0, \text{ т. к. } в > 0; \ a > 0.$$

Так что a, a_1, $в$, $в_1$ положительные величины и поэтому $D>0$ и $D\neq 0$.

Уравнение (8) напишем в безразмерном виде можно представлять таким образом:

$$t=\tau\frac{H^2}{a};\quad Z=H\hat{Z}, \tag{11}$$

или

$$\frac{\partial\psi\left(\hat{Z},\tau\right)}{\partial\tau}=a_0\frac{\partial^2}{\partial\hat{Z}^2}\psi\left(\hat{Z},T\right), \tag{12}$$

где $a_0=\left(1+Aв_1\dfrac{a_1}{a}\right)=\dfrac{a+a_1+a_1вв_1+\sqrt{D}}{2a}$.

Из равенства (10) и (12) находим функцию температуры в виде:

$$T_{(\eta)}=F(\eta)-(A-в)W. \tag{13}$$

Введем функцию:

$$\Phi(\eta)=W^1(\eta). \tag{14}$$

Тогда последнее уравнение примет вид:

$$\frac{a_1в_1(A-в)-a_0}{\nu a_0}\Phi^1(\eta)+\frac{1}{2}\eta\Phi(\eta)=-\frac{a_1в}{\nu a_0}F^{11}(\eta).$$

Введем величины a^* и β^* в виде:

$$a^*=\frac{a_1в_1(A-в)-a_1}{\nu a_0};\quad \beta_0{}^*=\frac{a_1в}{\nu a_0},$$

и разделим уравнение на a^* и тогда получим обыкновенное дифференциальное уравнение Эйлера в виде:

$$\Phi^1(\eta)+\frac{1}{2a^*}\eta\Phi(\eta)=\frac{\beta^*}{a^*}F^{11}(\eta); \tag{15}$$

$$P(\eta)=\int\frac{1}{2a^*}\eta d\eta=\frac{\eta^2}{4a^*};\quad Q(\eta)=-\frac{\beta^*}{a^*}F^{11}(\eta).$$

Решение уравнения (15) напишется в виде:

$$\Phi(\eta)=\exp\left(-\frac{\eta^2}{4a^*}\right)\left\{C_1+\left[F^1(\eta)\exp\left(+\frac{\eta^2}{4a^*}\right)-\frac{\eta}{2a^*}F(\eta)\exp\left(\frac{\eta^2}{4a^*}\right)\right]+\right.$$
$$\left.+\int\left[F(\xi)\left(\frac{\xi}{2a^*}+\frac{\xi^2}{4a^*}\right)\exp\left(\frac{\xi^2}{4a^*}\right)\right]d\xi\right\}. \tag{16}$$

Функции $F(\eta)=\left(\psi_K-\psi_H\right)erf\left(\dfrac{\xi}{2}\right)+\psi_H$ и $F^1(\xi)=\dfrac{1}{\sqrt{\pi}}\exp\left(-\dfrac{\xi^2}{2}\right)(\psi_K+\psi_H)$.

Откуда находим, что $F(0)=\psi_K$; $F^1(0)=\dfrac{(\psi_K-\psi_H)}{\sqrt{\pi}}$.

Тогда для функции $\Phi(\eta)=W^1(\eta)$:

$$\Phi(\eta)=W^1(\eta)=C_W^*+C_1\exp\left(-\frac{\eta^2}{4a^*}\right)+F^1(\eta)+$$

$$+\exp\left(-\frac{\eta^2}{4a^*}\right)\left\{\int_0^n F(\xi)\left[\frac{\xi}{2a^*}+\left(\frac{\xi}{2a^*}\right)^2\right]\exp\left(\frac{\xi^2}{4a^*}\right)d\xi\right\}, \tag{17}$$

где C_W^* – величина обеспечивающая $\left.\frac{dW}{d\eta}\right|_{\eta=0}=0$; $W(0)=W_K$.

Интегрируя полученное равенство (17) по η с граничным условием $W(0)=W_K$ получим распределение влаги в рассматриваемой области (3):

$$W(\eta)=C_W^*\eta+C_1\int_0^\eta\exp\left(\frac{\eta^2}{4a^*}\right)d\eta+F(\eta)-T_K-(A-в)W_K+$$

$$+\int_0^\eta\left\{\exp\left(-\frac{S^2}{4a^*}\right)\int_0^S\left\{F(\xi)\left[\frac{\xi}{2a^*}+\left(\frac{\xi}{2a^*}\right)^2\right]\exp\left(\frac{\xi^2}{4a^*}\right)\right\}d\xi\right\}dS+W_K.$$

При $Z=0$; $\eta=0$ и $W(0)=W_K$; при $\tau=0$; $\eta\to\infty$ и $W_\infty\to W_H$ – согласно начальному условию. Из полученного равенства получаем коэффициент C_1 при $t=0$, в начале процесса, то $\eta\to\infty$. Откуда получим $C_W=\frac{\psi_K-\psi_H}{\sqrt{\pi a^*}}$.

$$W(\infty)=W_H=C_1\int_0^\infty\exp\left(-\frac{\eta^*}{4a^*}\right)d\eta+\psi_H-\psi_K+l_\infty+W_K.$$

Здесь $\psi_H-\psi_K=T_H-T_K+(A-в)(W_H-W_K)$,

где $$l_{(\infty)}=\int_0^\eta\left\{\exp\left(-\frac{S^2}{4a^*}\right)\int_0^S F(\xi)\left[\frac{\xi}{2a^*}+\left(\frac{\xi}{2a^*}\right)^*\right]\exp\left(\frac{\xi^2}{4a^*}\right)d\xi\right\}dS \tag{18}$$

$$C_1=\{W_K-W_H+T_H+(A-в)W_H-T_K-(A-в)W_в+l(\infty)\}\int_0^\infty\exp\left(\frac{\eta^2}{4a^*}\right)d\eta$$

$$\int_0^\infty\exp\left(-\frac{\eta^2}{4a^*}\right)d\eta=\int_0^\infty e^{-\frac{S^2}{2}}\sqrt{2a^*}dS=\sqrt{2a^*}\int_0^\infty e^{-\frac{S^2}{2}}dS=$$

$$=\sqrt{2a^*}\Phi(\infty)\frac{\sqrt{2\pi}}{2}=\int_0^\infty\exp\left(-\frac{\eta^2}{4a^*}\right)d\eta=\frac{\sqrt{\pi}}{\sqrt{a^*}},\quad \text{т. к. } \Phi(\infty)=1.$$

Так что определим коэффициент C_1 из равенства:

$$C_1=\sqrt{\frac{a^*}{\pi}}\left[W_K-W_H+T_H-T_K+(W_K-W_H)(A-в)+l_\infty\right];$$

$$C_W^* = W_K - \frac{(\psi_K - \psi_H)\sqrt{2}}{\sqrt{\pi}}.$$

Откуда получим: $$W_K = \frac{1}{\sqrt{\pi} + A - в}\left[(A - в)W_H - (T_K - T_H)\right]. \qquad (19)$$

Таким образом, для оценки изменения влажности имеем:

$$W(\eta) = R(\eta) + F(\eta) + C_1 \int_0^\eta \exp\left(\frac{\eta^2}{4a^*}\right) d\eta + W_K. \qquad (20)$$

При $t \to 0$, $\eta \to \infty$ и $W(\infty) = W\big|_{t=0} = W_H < \infty$, поэтому $C_W^* = 0$.

$$R(\eta) = \int_0^\eta \exp\left(-\frac{S^2}{4a^*}\right)\left\{\int_0^S F(\xi)\left[\left(\frac{\xi}{2a^*} + \frac{\xi^2}{4a^{*2}}\right)\exp\left(\frac{\xi^2}{4a^*}\right)\right] d\xi\right\} d\eta, \qquad (21)$$

учитывая равенство (9) определим изменение температуры:

$$T(\eta) = \psi(\eta) - (A - в)W(\eta).$$

Откуда из равенств (12), (17), (18) и (19) получим изменение температуры в земляном полотне дороги с учетом тепломассообмена в ее грунтовой теле:

$$T(\eta) = \left[T_H - T_K + (A - в)W_K\right]\sqrt{\frac{\pi}{2}} erf\left(\frac{\eta}{2}\right) -$$

$$-(A - в)\left[R(\eta) + F(\eta) + \sqrt{\frac{\pi a_*}{2}} C_1 erf\left(\frac{\eta}{2\sqrt{a_*}}\right)\sqrt{\frac{\pi}{2}} - (A - в)W_K + T_K\right].$$

здесь $C_1 = \sqrt{\frac{a^*}{\pi}\left[W_K - W_H + T_H - T_K(W_H - W_K)(A - в) + l_\infty\right]}$.

Из условия $C_W^* = 0$ находим: $W_K = \dfrac{T_K - T_H}{\sqrt{\dfrac{\pi}{2}} + в - A} - \dfrac{A - в}{\sqrt{\dfrac{\pi}{2}} + в - A} W_H$.

Таким образом, на автомобильных дорогах пучинообразование прямым образом зависит от содержания влаги в теле земляного полотна и грунтовом оснований дорожной одежды, а также его морозоустойчивости. Рассмотренная модель устанавливает закономерные связи между температурой и естественной влажностью грунта, между влагой и морозоустойчивостью. В конечном счете – эти результаты также могут быть полезным при выборе конструкции основания дорожной одежды из укрепленных различными теплоустойчивыми и прочностными материалами.

ВОПРОСЫ САМОКОНТРОЛЯ
по компетентностной компоненте «Знать»

Основным элементом изучения учебного материала является освоение ключевых понятий, терминов, теоретических положений, методов и методик. Процесс получения знания состоит не только из прослушивания лекций и чтения учебной литературы. Чтобы вникнуть в суть изучаемого материала, следует ознакомиться с теорией и, главное, осознать, как именно она применяется в практических вопросах, примерах и задачах. Однако без четкого оперирования понятиями и определениями все равно не обойтись. С этой целью по каждому из рассмотренных разделов сформулированы контрольные вопросы, ответы на которые продемонстрируют уровень приобретенных знаний перед выходом обучающихся на итоговую аттестацию в форме зачета.

Раздел 1. *Прикладные вопросы теории вероятностей.*

1. Основные описательные характеристики однофакторных выборок.
2. Понятия средних тенденций – «мода», «медиана», «среднее арифметическое».
3. Понятия характеристик вариабельности – «коэффициент вариации», «дисперсия», «среднеквадратичное отклонение».
4. Характеристики симметричности и островершинности.
5. Графические формы представления выборки: полигон частот, кумулята, гистограмма.
6. Закон и функция плотности распределений случайной величины.
7. Дискретные случайные величины и их основные законы.
8. Непрерывные случайные величины и их основные законы.
9. Специальные законы распределений случайной величины.

Раздел 2. *Прикладные методы статистики.*

1. Понятие точечных и интервальных оценок статистических характеристик.
2. Уровень доверительной вероятности.
3. Мощность статистического критерия.
4. Виды статистических гипотез.
5. Правила принятия статистических гипотез.
6. Статистические критерии согласия.
7. Критерий Пирсона.
8. Параметрические критерии проверки свойств выборок.
9. Непараметрические критерии проверки свойств выборок.

Раздел 3. *Однофакторный анализ.*

1. Понятие регрессии.
2. Однофакторная регрессия.
3. Линейная регрессия.
4. Метод наименьших квадратов.
5. Критерий точности корреляционной модели.
6. Нелинейная регрессия.
7. Метод выравнивания.
8. Коэффициент корреляции.
9. Основные понятия корреляционного анализа.

Раздел 4. *Линейное программирование*.

1. Общая и основная задача линейного программирования.
2. Целевая функция ЗЛП.
3. Переход от задачи минимизации к задаче максимизации.
4. Каноническая задача линейного программирования.
5. Двойственная ЗЛП.
6. Геометрический смысл решения ЗЛП.
7. Область допустимых решений. Множество Парето.
8. Графический метод решения ЗЛП.
9. Симплекс-метод решения ЗЛП.

Раздел 5. *Методы поиска оптимальных решений.*

1. Открытая и закрытая транспортная задача.
2. Методы построения опорного плана.
3. Методы улучшения опорного плана.
4. Метод потенциалов оптимизации опорного плана
5. Стандартная задача о назначениях.
6. Оптимальное назначение.
7. Алгоритм Флада в задаче об оптимальном назначении.
8. Венгерский метод оптимальных назначений.
9. Алгоритм Флада оптимальной последовательности работ.

Раздел 6. *Методы и модели принятия решений оптимальных решений.*

1. Основные понятия теории графов.
2. Маршрут, полный и критический путь.
3. Матрица смежности.
4. Матрица инцидентности.
5. Общая задача динамического программирования.
6. Принцип оптимальности Беллмана.
7. Задача сетевого планирования и управления.
8. Построения кратчайшего маршрута.
9. Условная оптимизация транспортной сети.

Раздел 7. *Модели в дорожно-транспортном планировании и строительстве.*

1. Виды моделей.
2. Требования к моделям.
3. Этапы моделирования.
4. Математическое моделирование.
5. Модели экстремального анализа.
6. Методы экстремального анализа.
7. Основные законы распределения в экстремальных моделях.
8. Дискретные модели.
9. Вероятностные модели.

Овладение компонентой «Знать» является фундаментальной основой для формирования и развития компетенций, предусмотренных основными характеристиками образовательных программ. Если фундамент не прочен или отсутствует вовсе, никаких сложных конструкций на нем возвести невозможно. На этом основании автор остановился на сводной подборке вопросов самоконтроля по изложенным в пособии экономико-статистическим методам и моделям в транспортном строительстве для бакалавров.

«Математика сама создает те идеальные образы, над которыми она оперирует…
«чистому» математику нет дела до того, есть ли в природе такие предметы,
к которым его образы относятся, для него важно, что он их создал в своем уме,
приписал им определения, аксиомы и допущения, после чего он с полною логичностью и
строгостью развивает следствия этих аксиом и допущений – до остального ему дела нет.
Инженер должен по своей специальности уметь владеть своим инструментом,
но он вовсе не должен уметь его делать; плотник не должен уметь выковать или направить
топор, но должен уметь отличить хороший топор от плохого;
слесарь не должен уметь сам насекать напильник,
но должен выбрать тот напильник, который ему надо.
Так вот, чистого математика, который создаст новые математические выводы,
можно уподобить некоему воображаемому универсальному инструментальщику,
который готовит на склад инструмент на всякую потребу;
он делает всё, от кувалды до тончайшего микроскопа и точнейшего хронометра.
Он создает методы решения вопросов, не только возникающих вследствие современных
надобностей, но и для будущих, которые возникнут, может быть, через тысячу лет.
***Алексей Николаевич Крылов** (1863-1945) –*
русский и советский кораблестроитель, механик и математик, академик

ЗАКЛЮЧЕНИЕ

Изложенный в пособии материал можно успешно применять для изучения и анализа эмпирических зависимостей, полученных в результате проводимых экспериментов. Однако, следует понимать, что учебное пособие рассматривает не все вопросы получения и обработки эмпирических данных в проектировании, строительстве и управлении транспортными сооружениями.

Методы построения моделей и математической обработки могут оказать исследователю неоценимую помощь, но они лишь средство, которое не должно заслонять собой цель. Достоверная статистическая тенденция – это все же не неоспоримая закономерность, а выпадающие из общей картины индивидуальные значения, которые есть отражение закономерностей более высокого порядка, чем те, что выявлены с помощью экономико-математических методов. В процессе подготовки и проведения инженерных и лабораторных исследований неизбежно возникнут многие важные вопросы, ответы на которые недостаточно отражены в этой книге. Список литературы, приведенный в конце пособия, является возможностью получения дополнительной информации.

Автор надеется, что это учебное пособие будет полезно тем обучающимся, кто хочет разобраться в основах методов статистической обработки данных, их анализа и применять эти знания в своей образовательной и исследовательской деятельности. Можно предположить, что после прочтения учебного пособия у читателя создалось правильное впечатление о том, насколько рассмотренные в ней вопросы, с одной стороны, сложны, а с другой, – неотделимы от содержательных исследовательских представлений об изучаемом явлении, объекте.

Автору хотелось бы, чтобы студенты творчески восприняли главную идею: экономико-математические методы и моделирование реальности, все приемы, подходы о которых шла речь (да и не только они) могут служить лишь некими «кирпичиками», из которых инженеру или бакалавру предстоит построить собственный «дом», для конкретной технико-экономической ситуации.

ЛИТЕРАТУРА

1. Афанасьев М.Ю., Суворов Б.П. Исследование операций в экономике. – М.: ИНФРА – М. 2003.
2. Б.Л. Ван дер Варден. Математическая статистика / пер. с нем.. – М.: Иностранная литература. 1960.
3. Багриновский К.А., Матюшок В.М. Экономико-математические методы и модели (микроэкономика). Учебное пособие. – М.: РУДН, 1999.
4. Баркалов С.А., Мещерякова О.К. и др. Основы научных исследований по организации и управлению строительным производством: Учебное пособие. – Воронеж: ВГАСУ, 2002 (в 2 частях).
5. Бережная Е.В., Бережной В.И. Математические методы моделирования экономических систем. – М.: Финансы и статистика, 2001.
6. Большаков А.С. Моделирование в менеджменте. – М.: «Филинъ», 2000.
7. Вагнер Г. Основы исследования операций / пер. с англ. (в 3 томах). – М.: МИР, 1973.
8. Вентцель Е.С. Исследование операций. – М.: Советское радио, 1972.
9. Военно-экономический анализ: Учебник / под ред. проф. Викулова С.Ф. М: Воениздат, 2001.
10. Воробьев И.Н. Теория игр для экономистов-кибернетиков. М.: Наука, 1985.
11. Гейл Д. Теория линейных экономических моделей. – М.: ИЛ, 1963.
12. Голенко Д.И. Статистические методы сетевого планирования и управления. М.: Наука, 1968.
13. Дувров А.М., Лагоша Б.А., Хрусталев Е.Ю. Моделирование рисковых ситуаций в экономике и бизнесе: Учебное пособие /под ред. Б.А. Лагоши. – М.: Финансы и статистика, 1999.
14. Егорова Н.Е., Хачатрян С.Р., Вороновская О.Е. Моделирование кредитно-инвестиционной политики развития малого бизнеса с учетом рисков / Препринт # WP /99/081. – М.: ЦЭМИ РАН, 1999.
15. Замков О.О., Толстопятенко А.В., Черемных Ю.Н. Математические методы в экономике. – М.: ДИС, 1977.
16. Золотарь И.А. Экономико-математические методы в дорожном строительстве: Монография. М.: Транспорт, 1974.
17. Золотарь И.А., Мальцев Ю.А., Митянин А.А. Экономика и организация дорожного строительства. Учебник. – Л.: ВАТТ, 1975.

18. Иванилов Ю.П. Математические модели и экономике. – М.: Наука, 1999.

19. Иванилов Ю.П., Лотов А.В. Математические модели в экономике. – М.: Наука, 1979.

20. Интриллигатор М. Математические методы оптимизации и экономическая теория. – М.: Прогресс, 1975.

21. Исследование операций в экономике / под ред. проф. Кремера Н.Ш. – М.: Банки и биржи, ЮНИТИ, 1997.

22. Карасев А.И., Кремер Н.Ш., Савельев Т.И. Математические методы и модели в планировании. – М.: Экономика, 1987.

23. Кини Р.Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения / пер. с англ. – М.: Радио и связь, 1981.

24. Клейнер Г.Б. Производственные функции: теория, методы, применения. – М.: Финансы и статистика, 1986.

25. Колемаев В.А. Математическая экономика. – ЮНИТИ, 1998.

26. Лебедев В.В. Математическое моделирование социально-экономических процессов. – М.: Изограф, 1997.

27. Левин М.И., Макаров В.Л., Рубинов А.М. Математические модели экономического взаимодействия. – М.: Физматиздат, 1993.

28. Лотов А.В. Введение в экономико-математическое моделирование. – М.: Наука, 1984.

29. Льюис Р., Райфа Х. Игры и решения. Введение и критический обзор. – М.: ИЛ, 1961.

30. Мальцев Ю.А. Проблемы экономико-математического моделирования при строительстве автомобильных дорог в условиях рыночной экономики. / в юбил. Научно-техн. сб. МВВДИУ, – Балашиха, 1997.

31. Мальцев Ю.А. Основы научных исследований: Учебное пособие – Балашиха: ВТУ Спецстроя России, 2004.

32. Методическое пособие по формированию организационной структуры военно-строительных организаций типа УНР, ХРУ. – М.: 26 ЦНИИ МО РФ, 1998.

33. Митропольский А.К. Элементы математической статистики. – Л.: ЛГУ. 1969.

34. Оуэн Г. Теория игр. – М.: МИР, 1971.

35. Пинегина М.В. Экономико-математические методы и модели. – М.: Экзамен, 2002.

36. Протасов И.Д. Теория игр и исследование операций. – М.: Гелиос АРВ, 2003.

37. Рыжиков Ю.И. Управление запасами.– М.: Наука 1969.

38. Таха Х. Введение в исследование операций (в двух томах). – М.: МИР, 1985.

39. Федосеев В.В., Гармаш А.Н., Дайитбеков Д.М. Экономико-математические методы и прикладные модели. – М.: ЮНИТИ, 2000.

40. Феллер В. Введение в теорию вероятностей и ее приложения (в двух томах). – М.: Мир, 1984.

41. Фон Нейман Дж., Моргенштерн О. Теория игр и экономическое поведение. – М.: Наука, 1970.

42. Хачатрян С.Р.. Пинегина М.В., Буянов В.П. Методы и модели решения экономических задач. – М.: Экзамен, 2005.

43. Четыркин Е.М. Методы финансовых и коммерческих расчетов. – М.: Дело Лдт, 1995.

44. Четыркин Е.М. Статистические методы прогнозирования – М.: Статистика 1997.

45. S.-D. Poisson. Recherches sur la probabilité des jugements en matière criminelle et en matière civile. Paris, 1837. Пуассон С.Д. Исследования о вероятности приговоров в уголовных и гражданских делах / пер. О.Б. Шейнина. Берлин: Oscar Sheynin, 2013.

СОДЕРЖАНИЕ

ПРИЛОЖЕНИЯ.
ТАБЛИЦЫ КРИТИЧЕСКИХ ЗНАЧЕНИЙ

Таблица 1. Критические значения критерия Пирсона χ^2
при разном числе степеней свободы v

	p			*p*			*p*	
v	**0.05**	**0.01**	***v***	**0.05**	**0.01**	***v***	**0.05**	**0.01**
1	3.841	6.635	**35**	49.802	57.342	**69**	89.391	99.227
2	5.991	9.210	**36**	50.998	58.619	**70**	90.631	100.425
3	7.815	11.345	**37**	52.192	59.892	**71**	91.670	101.621
4	9.488	13.277	**38**	53.384	61.162	**72**	92.808	102.816
5	11.070	15.086	**39**	54.572	62.428	**73**	93.945	104.010
6	12.592	16.812	**40**	55.758	63.691	**74**	95.081	105.202
7	14.067	18.475	**41**	56.942	64.950	**75**	96.217	106.393
8	15.507	20.090	**42**	58.124	66.206	**76**	97.351	107.582
9	16.919	21.666	**43**	59.304	67.459	**77**	98.484	108.771
10	18.307	23.209	**44**	60.481	68.709	**78**	99.617	109.958
11	19.675	24.725	**45**	61.656	69.957	**79**	100.749	111.144
12	21.026	26.217	**46**	62.830	71.201	**80**	101.879	112.329
13	22.362	27.688	**47**	64.001	72.443	**81**	103.010	113.512
14	23.685	29.141	**48**	65.171	73.683	**82**	104.139	114.695
15	24.996	30.578	**49**	66.339	74.919	**83**	105.267	115.876
16	26.296	32.000	**50**	67.505	76.154	**84**	106.395	117.057
17	27.587	33.409	**51**	68.669	77.386	**85**	107.522	118.236
18	28.869	34.805	**52**	69.832	78.616	**86**	108.648	119.414
19	30.144	36.191	**53**	70.993	79.84'3	**87**	109.773	120.591
20	31.410	37.566	**54**	72.153	81.069	**88**	110.898	121.767
21	32.671	38.932	**55**	73.311	82.292	**89**	112.022	122.942
22	33.924	40.289	**56**	74.468	83.513	**90**	113.145	124.116
23	35.172	41.638	**57**	75.624	84.733	**91**	114.268	125.289
24	36.415	42.980	**58**	76.778	85.950	**92**	115.390	126.462
25	37.652	44.314	**59**	77.931	87.166	**93**	116.511	127.633
26	38.885	45.642	**60**	79.082	88.379	**94**	117.632	128.803
27	40.113	46.963	**61**	80.232	89.591	**95**	118.752	129.973
28	41.337	48.278	**62**	81.381	90.802	**96**	119.871	131.141
29	42.557	49.588	**63**	82.529	92.010	**97**	120.990	132.309
30	43.773	50.892	**64**	83.675	93.217	**98**	122.108	133.476
31	44.985	52.191	**65**	84.821	94.422	**99**	123.225	134.642
32	46.194	53.486	**66**	85.965	95.626	**100**	124.342	135.807
33	47.400	54.776	**67**	87.108	96.828			
34	48.602	56.061	**68**	88.250	98.028			

Таблица 2. Критические значения модуля максимального расхождения d_{max} при сопоставлении эмпирического распределения с теоретическим

n	Максимальный модуль разности накопленных частностей d_{max}		***n***	Максимальный модуль разности накопленных частностей d_{max}	
	p=0.05	**p =0.01**		**p=0.05**	**p=0.01**
5	0.6074	0.7279	**50**	0.1921	0.2302
10	0.4295	0.5147	**60**	0.1753	0.2101
15	0.3507	0.4202	**70**	0.1623	0.1945
20	0.3037	0.3639	**80**	0.1518	0.1820
25	0.2716	0.3255	**90**	0.1432	
30	0.2480	0.2972	**100**	0.1358	
40	0.2147	0.2574	**>100**	$1{,}36/\sqrt{n}$	$1{,}63/\sqrt{n}$

Таблица 3. Критерий λ Колмогорова-Смирнова для сопоставления эмпирического распределения с теоретическим ($n > 50$) или двух эмпирических распределений между собой ($n > 50$); уровни статистической значимости разных значений $\lambda_{эмп}$

λ	**λ, последний десятичный знак**									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
	р – десятичные знаки («0» опущен)									
0,3	99999	99998	99995	99991	99983	99970	99949	99917	99872	99807
0,4	99719	99603	99452	99262	99027	98741	98400	97998	97532	96998
0,5	96394	95719	94969	94147	93250	92282	91242	90134	88960	87724
0,6	86428	85077	83678	82225	80732	79201	77636	76042	74422	72781
0,7	71124	69453	67774	66089	64402	62717	61036	59363	57700	56050
0,8	54414	52796	51197	49619	48063	46532	45026	43545	42093	40668
0,9	39273	37907	36571	35266	33992	32748	31536	30356	29206	28087
1,0	27000	25943	24917	23922	22957	22021	21114	20236	19387	18566
1,1	17772	17005	16264	15550	14861	14196	13556	12939	12345	11774
1,2	11225	10697	10190	09703	09235	08787	08357	07944	07550	07171
1,3	06809	06463	06132	05815	05513	05224	04949	04686	04435	04196
1,4	03968	03751	03545	03348	03162	02984	02815	02655	02503	02359
1,5	02222	02092	01969	01852	01742	01638	01539	01446	01357	01274
1,6	01195	01121	01051	00985	00922	00864	00808	00756	00707	00661
1,7	00618	00577	00539	00503	00469	00438	00408	00380	00354	00330
1,8	00307	00285	00265	00247	00229	00213	00198	00186	00170	00158
1,9	00146	00136	00126	00116	00108	00100	00092	00085	00079	00073
2,0	00067	00062	00057	00053	00048	00045	00041	00038	00035	00032
2,1	00030	00027	00025	00023	00021	00019	00018	00016	00015	00014
2,2	00013	00011	00010	00010	00009	00008	00007	00007	00006	00006
2,3	00005	00005	00004	00004	00004	00003	00003	00003	00002	00002
2,4	00002	00002	00002	00001	00001	00001	00001	00001	00001	00001

Таблица 4. Критические значения F критерия Фишера
(C_1 число степеней свободы между уровнями, C_2 внутри уровней)

Влияние фактора или взаимодействия факторов достоверно, если $F_{эмп}$, равен или больше критического значения F_{005}и тем более достоверно, если $F_{эмп} > F_{0,01}$

C_1	1	2	3	4	5	6	7	8	9	10	11	12
C_2	P < 0.05											
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28
9	5.12	4.26	3.86	3.63	3.48	3.17	3.29	3.23	3.18	3.13	3.10	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42
C_2	P < 0.01											
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6082	6106
2	98.49	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.41	99.42
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.29	10.15	10.05	9.96	9.89
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.47
8	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67
9	10.56	8.02	6.9	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71
11	9.65	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16
13	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55

Таблица 5. Таблица значений функции Лапласа

Функция Лапласа $\Phi(t)=\frac{2}{\sqrt{2\pi}}\int_{0}^{t} e^{\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}}\int_{-t}^{+t} e^{-\frac{t^2}{2}}dt$.

При разных значениях t; $\Phi(-t)=-\Phi(t)$ (функция нормального распределения).

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0.00	0.00000	**1.00**	0.68269	**2.00**	0.95450	**3.00**	0.99730
0.01	0.00798	**1.01**	0.68750	**2.01**	0.95557	**3.01**	0.99739
0.02	0.01596	**1.02**	0.69227	**2.02**	0.95662	**3.02**	0.99747
0.03	0.02393	**1.03**	0.69699	**2.03**	0.95764	**3.03**	0.99755
0.04	0.03191	**1.04**	0.70166	**2.04**	0.95865	**3.04**	0.99763
0.05	0.03988	**1.05**	0.70628	**2.05**	0.95964	**3.05**	0.99771
0.06	0.04784	**1.06**	0.71086	**2.06**	0.96060	**3.06**	0.99779
0.07	0.05581	**1.07**	0.71538	**2.07**	0.96155	**3.07**	0.99786
0.08	0.06376	**1.08**	0.71986	**2.08**	0.96247	**3.08**	0.99793
0.09	0.07171	**1.09**	0.72429	**2.09**	0.96338	**3.09**	0.99800
0.10	0.07966	**1.10**	0.72867	**2.10**	0.96427	**3.10**	0.99806
0.11	0.08759	**1.11**	0.73300	**2.11**	0.96514	**3.11**	0.99813
0.12	0.09552	**1.12**	0.73729	**2.12**	0.96599	**3.12**	0.99819
0.13	0.10348	**1.13**	0.74152	**2.13**	0.96683	**3.13**	0.99825
0.14	0.11134	**1.14**	0.74571	**2.14**	0.96765	**3.14**	0.99831
0.15	0.11924	**1.15**	0.74986	**2.15**	0.96844	**3.15**	0.99837
0.16	0.12712	**1.16**	0.75395	**2.16**	0.96923	**3.16**	0.99842
0.17	0.13499	**1.17**	0.75800	**2.17**	0.96999	**3.17**	0.99848
0.18	0.14285	**1.18**	0.76200	**2.18**	0.97074	**3.18**	0.99853
0.19	0.15069	**1.19**	0.76595	**2.19**	0.97148	**3.19**	0.99858
0.20	0.15852	**1.20**	0.76986	**2.20**	0.97219	**3.20**	0.99863
0.21	0.16633	**1.21**	0.77372	**2.21**	0.97289	**3.21**	0.99867
0.22	0.17413	**1.22**	0.77754	**2.22**	0.97358	**3.22**	0.99872
0.23	0.18191	**1.23**	0.78130	**2.23**	0.97425	**3.23**	0.99876
0.24	0.18967	**1.24**	0.78502	**2.24**	0.97491	**3.24**	0.99880
0.25	0.19741	**1.25**	0.78870	**2.25**	0.97555	**3.25**	0.99855
0.26	0.20514	**1.26**	0.79233	**2.26**	0.97618	**3.26**	0.99889
0.27	0.21284	**1.27**	0.79592	**2.27**	0.97679	**3.27**	0.99892
0.28	0.22052	**1.28**	0.79945	**2.28**	0.97739	**3.28**	0.99896
0.29	0.22818	**1.29**	0.80295	**2.29**	0.97798	**3.29**	0.99900
0.30	0.23582	**1.30**	0.80640	**2.30**	0.97855	**3.30**	0.99903
0.31	0.24344	**1.31**	0.80980	**2.31**	0.97911	**3.31**	0.99907
0.32	0.25103	**1.32**	0.81316	**2.32**	0.97966	**3.32**	0.99910
0.33	0.25860	**1.33**	0.81648	**2.33**	0.98019	**3.33**	0.99913
0.34	0.26614	**1.34**	0.81975	**2.34**	0.98072	**3.34**	0.99916
0.35	0.27366	**1.35**	0.82298	**2.35**	0.98123	**3.35**	0.99919

Продолжение таблицы 5

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0.36	0.28115	**1.36**	0.82617	**2.36**	0.98172	**3.36**	0.99922
0.37	0.28862	**1.37**	0.82931	**2.37**	0.98221	**3.37**	0.99925
0.38	0.29605	**1.38**	0.83241	**2.38**	0.98269	**3.38**	0.99928
0.39	0.30346	**1.39**	0.83547	**2.39**	0.98315	**3.39**	0.99930
0.40	0.31084	**1.40**	0.83849	**2.40**	0.98360	**3.40**	0.99933
0.41	0.31819	**1.41**	0.84146	**2.41**	0.98405	**3.41**	0.99935
0.42	0.32552	**1.42**	0.84439	**2.42**	0.98448	**3.42**	0.99937
0.43	0.33280	**1.43**	0.84728	**2.43**	0.98490	**3.43**	0.99940
0.44	0.34006	**1.44**	0.85013	**2.44**	0.98531	**3.44**	0.99942
0.45	0.34729	**1.45**	0.85294	**2.45**	0.98571	**3.45**	0.99944
0.46	0.35448	**1.46**	0.85571	**2.46**	0.98611	**3.46**	0.99946
0.47	0.36164	**1.47**	0.85844	**2.47**	0.98649	**3.47**	0.99948
0.48	0.36877	**1.48**	0.86113	**2.48**	0.98686	**3.48**	0.99950
0.49	0.37587	**1.49**	0.86378	**2.49**	0.98723	**3.49**	0.99952
0.50	0.38292	**1.50**	0.86639	**2.50**	0.98758	**3.50**	0.99953
0.51	0.38995	**1.51**	0.86696	**2.51**	0.98793	**3.51**	0.99955
0.52	0.39694	**1.52**	0.87149	**2.52**	0.98826	**3.52**	0.99957
0.53	0.40389	**1.53**	0.87398	**2.53**	0.98859	**3.53**	0.99958
0.54	0.41080	**1.54**	0.87644	**2.54**	0.98891	**3.54**	0.99960
0.55	0.41768	**1.55**	0.87886	**2.55**	0.98923	**3.55**	0.99961
0.56	0.42452	**1.56**	0.88124	**2.56**	0.98953	**3.56**	0.99963
0.57	0.43132	**1.57**	0.88358	**2.57**	0.98983	**3.57**	0.99964
0.58	0.43809	**1.58**	0.88589	**2.58**	0.99012	**3.58**	0.99966
0.59	0.44481	**1.59**	0.88817	**2.59**	0.99040	**3.59**	0.99967
0.60	0.45149	**1.60**	0.89040	**2.60**	0.99068	**3.60**	0.99968
0.61	0.45814	**1.61**	0.89260	**2.61**	0.99095	**3.61**	0.99969
0.62	0.46474	**1.62**	0.89477	**2.62**	0.99121	**3.62**	0.99971
0.63	0.47131	**1.63**	0.89690	**2.63**	0.99146	**3.63**	0.99972
0.64	0.47783	**1.64**	0.89899	**2.64**	0.99171	**3.64**	0.99973
0.65	0.48431	**1.65**	0.90106	**2.65**	0.99195	**3.65**	0.99974
0.66	0.49075	**1.66**	0.90309	**2.66**	0.99219	**3.66**	0.99975
0.67	0.49714	**1.67**	0.90508	**2.67**	0.99241	**3.67**	0.99976
0.68	0.50350	**1.68**	0.90704	**2.68**	0.99263	**3.68**	0.99977
0.69	0.50981	**1.69**	0.90897	**2.69**	0.99285	**3.69**	0.99978
0.70	0.51607	**1.70**	0.91087	**2.70**	0.99307	**3.70**	0.99978

Продолжение таблицы 5

t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$	t	$\Phi(t)$
0.71	0.52230	**1.71**	0.91273	**2.71**	0.99327	**3.71**	0.99979
0.72	0.52848	**1.72**	0.91457	**2.72**	0.99347	**3.72**	0.99980
0.73	0.53461	**1.73**	0.91637	**2.73**	0.99367	**3.73**	0.99981
0.74	0.54070	**1.74**	0.91814	**2.74**	0.99386	**3.74**	0.99982
0.75	0.54675	**1.75**	0.91988	**2.75**	0.99404	**3.75**	0.99982
0.76	0.55275	**1.76**	0.92159	**2.76**	0.99422	**3.76**	0.99983
0.77	0.55870	**1.77**	0.92327	**2.77**	0.99439	**3.77**	0.99984
0.78	0.56461	**1.78**	0.92492	**2.78**	0.99456	**3.78**	0.99984
0.79	0.57047	**1.79**	0.92655	**2.79**	0.99473	**3.79**	0.99985
0.80	0.57629	**1.80**	0.92814	**2.80**	0.99489	**3.80**	0.99986
0.81	0.58206	**1.81**	0.92970	**2.81**	0.99505	**3.81**	0.99986
0.82	0.58778	**1.82**	0.93124	**2.82**	0.99520	**3.82**	0.99987
0.83	0.59346	**1.83**	0.93275	**2.83**	0.99535	**3.83**	0.99987
0.84	0.59909	**1.84**	0.93423	**2.84**	0.99549	**3.84**	0.99988
0.85	0.60468	**1.85**	0.93569	**2.85**	0.99563	**3.85**	0.99988
0.86	0.61021	**1.86**	0.93711	**2.86**	0.99576	**3.86**	0.99989
0.87	0.61570	**1.87**	0.93852	**2.87**	0.99590	**3.87**	0.99989
0.88	0.62114	**1.88**	0.93989	**2.88**	0.99602	**3.88**	0.99990
0.89	0.62653	**1.89**	0.94124	**2.89**	0.99615	**3.89**	0.99990
0.90	0.63188	**1.90**	0.94257	**2.90**	0.99627	**3.90**	0.99990
0.91	0.63718	**1.91**	0.94387	**2.91**	0.99639	**3.91**	0.99991
0.92	0.64243	**1.92**	0.94514	**2.92**	0.99650	**3.92**	0.99991
0.93	0.64763	**1.93**	0.94639	**2.93**	0.99661	**3.93**	0.99992
0.94	0.65278	**1.94**	0.94762	**2.94**	0.99672	**3.94**	0.99992
0.95	0.65789	**1.95**	0.94882	**2.95**	0.99682	**3.95**	0.99992
0.96	0.66294	**1.96**	0.95000	**2.96**	0.99692	**3.96**	0.99992
0.97	0.66795	**1.97**	0.95116	**2.97**	0.99702	**3.97**	0.99993
0.98	0.67291	**1.98**	0.95230	**2.98**	0.99712	**3.98**	0.99993
0.99	0.67783	**1.99**	0.95341	**2.99**	0.99721	**3.99**	0.99993

Таблица 6

Интеграл вероятности (функция Лапласа) $\Phi(x)$ (сокращенный вариант)

X	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.	0	797	1585	2358	3108	3829	4515	5161	5763	6319
1.	6827	7287	7699	8064	8385	8664	8904	9109	9231	9426
2.	9545	9643	9722	9786	9836	9876	9907	9931	9949	9963
3.	9973	9981	9986	9990	9993	19995	9997	9998	9999	9999

Таблица 7. Критические значения распределения Стьюдента

Число степеней свободы $f = n - 1$	***n***	**Доверительная вероятность**			
		0.90	**0.95**	**0.99**	**0.999**
1	2	6.3137515148	12.7062047364	63.6567411629	636.619249432
2	3	2.91998558036	4.30265272991	9.92484320092	31.599054577
3	4	2.3533634348	3.18244630528	5.84090929976	12.9239786366
4	5	2.13184678134	2.7764451052	4.60409487142	8.61030158138
5	6	2.01504837267	2.57058183661	4.03214298356	6.86882663987
6	7	1.94318028039	2.44691184879	3.70742802132	5.95881617993
7	8	1.89457860506	2.36462425101	3.49948329735	5.40788252098
8	9	1.85954803752	2.30600413503	3.35538733133	5.04130543339
9	10	1.83311293265	2.26215716274	3.24983554402	4.78091258593
10	11	1.81246112281	2.22813885196	3.16927266718	4.5868938587
11	12	1.7958848187	2.20098516008	3.10580651322	4.43697933823
12	13	1.78228755565	2.17881282966	3.05453958834	4.31779128361
13	14	1.77093339599	2.16036865646	3.01227583821	4.22083172771
14	15	1.76131013577	2.14478668792	2.97684273411	4.14045411274
15	16	1.75305035569	2.13144954556	2.94671288334	4.0727651959
16	17	1.74588367628	2.11990529922	2.92078162235	4.0149963326
17	18	1.73960672608	2.10981557783	2.89823051963	3.96512626361
18	19	1.73406360662	2.10092204024	2.87844047271	3.92164582001
19	20	1.72913281152	2.09302405441	2.86093460645	3.88340584948
20	21	1.72471824292	2.08596344727	2.84533970978	3.84951627298
21	22	1.72074290281	2.07961384473	2.83135955802	3.81927716303
22	23	1.71714437438	2.0738730679	2.8187560606	3.79213067089
23	24	1.71387152775	2.06865761042	2.80733568377	3.76762680377
24	25	1.71088207991	2.06389856163	2.79693950477	3.74539861893
25	26	1.70814076125	2.05953855275	2.78743581368	3.72514394948
26	27	1.70561791976	2.05552943864	2.77871453333	3.70661174331
27	28	1.70328844572	2.05183051648	2.77068295712	3.68959171334
28	29	1.70113093427	2.0484071418	2.76326245546	3.67390640062
29	30	1.69912702653	2.04522964213	2.75638590367	3.6594050194
30	31	1.69726089436	2.0422724563	2.74999565357	3.645958635
40	41	1.68385101139	2.021075383	2.70445926743	3.55096576086
60	61	1.67064886465	2.00029782106	2.66028303115	3.4602004692
120	121	1.65765089935	1.97993040505	2.61742114477	3.37345376507
999999.0	1000000.0	1.64485515072	1.95996635682	2.57583422011	3.29053646126