

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
ДАЛЬНЕВОСТОЧНЫЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ
УНИВЕРСИТЕТ

ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

С.Б. Пастушенко

Статистика

Учебное пособие

Благовещенск
Издательство ДальГАУ
2014

УДК 31(075)

Учебное пособие по дисциплине «Статистика» для студентов, обучающихся по направлению «Экономика» подготовлено: кандидатом экономических наук, доцентом кафедры «Бухгалтерский учет, статистика, анализ и аудит» Пастушенко С.Б.

Учебное пособие содержит, теоретические аспекты по каждой теме дисциплины «Статистика» (первая часть – теория статистики), примеры применения теоретических аспектов в практических целях, с возможностью применения программы Microsoft Excel, контрольные вопросы по каждой теме.

Содержание учебного пособия соответствует Федеральному государственному образовательному стандарту высшего профессионального образования (ФГОС ВПО) бакалавриата по направлению подготовки «Экономика» профилей подготовки «Экономика предприятий и организаций», «Бухгалтерский учет, анализ и аудит», «Финансы и кредит», «Страхование» и рабочим программам дисциплины «Статистика» для очной и заочной форм обучения.

Включенные в учебное пособие теоретические аспекты и примеры практического применения статистических методов, способов и приемов в анализе массовых общественных явлений максимально приближены к практике статистических работ на микро- и макроуровне и соответствуют современным требованиям подготовки бакалавров экономики высшей квалификации.

Рецензент:

к.э.н., доцент кафедры экономики и организации ФГБОУ ВПО ДальГАУ
Реймер В.В.

Благовещенск
Издательство ДальГАУ

2014

ВВЕДЕНИЕ

В современных условиях в РФ остро стоят вопросы качества высшего профессионального образования. В утвержденной распоряжением Правительства РФ, «Концепции долгосрочного социально-экономического развития Российской Федерации на период до 2020 года» [4], определены основные цели развития высшего профессионального образования в РФ, которые заключаются в основном в повышении его качества, доступности и эффективности. Те же принципы отражены в Федеральном законе от 29.12.2012г. N 273-ФЗ «Об образовании в Российской Федерации» [1].

На основе материалов 55-й сессии Международного статистического института (г.Берлин) разработаны международные варианты образовательных стандартов по экономическим направлениям государственных и негосударственных вузов, для двухуровневого образования – бакалавриата и магистратуры. Основой стандартов является принцип фундаментальности образования, который предполагает, что человек, получивший соответствующее образование, способен самостоятельно работать, выполняя определенные функции, и одновременно получает способность и возможность к самообучению на основе полученной фундаментальной образовательной базы в рамках системы непрерывного профессионального образования.

Комплексный подход в подготовке высококвалифицированных бакалавров экономики предусматривает глубокое изучение современных методов сбора, обработки и анализа цифровых данных на различных уровнях управления экономикой. Это реальное направление повышения интеллектуальности российского экономиста, бухгалтера и финансиста высшей квалификации. В современных предприятиях нередко отсутствуют специалисты, владеющие возможностями статистико-экономического анализа. Подготовка таких специалистов, повышение квалификации практических работников позволят поднять на более высокий уровень не только аналитические разработки, но и научную обоснованность принимаемых решений, обоснование социально-

экономической политики отдельных предприятий, отраслей, регионов и государства в целом.

В условиях комплексного подхода к подготовке высококвалифицированных бакалавров экономики роль статистического образования возрастает во всем мире, что обусловлено возрастающей потребностью в профессиональном использовании информации для принятия решений в экономике, планировании, прогнозировании, управлении, политике и любых социально-экономических процессах, как на уровне хозяйствующих субъектов, так и на уровне отраслей, регионов, секторов экономики и экономической системы в целом. Органам управления, как на государственном и региональном уровне, так и на уровне предприятий и организаций, необходимы квалифицированные специалисты, владеющие статистической методологией, способные адекватно оценивать результаты деятельности объекта и его подразделений, его развития, анализировать состояние внутренней и внешней экономической среды, его конкурентоспособность, выявлять резервы повышения эффективности его функционирования и прогнозировать его развитие на перспективу. В результате изучения дисциплины «Статистика» у обучающихся формируется целостное представление о применении комплекса статистических методов, способов и приемов в оценке функционирования и состояния объекта исследования, влияния на его функционирование внутренних и внешних факторов, в выявлении тенденций и прогнозировании развития объекта на перспективу.

Учебное пособие по дисциплине «Статистика» разработано также с учетом реализации в РФ Федеральной целевой программы «Развитие государственной статистики в 2007-2011 годах» (с изменениями и дополнениями) в редакции от 21.03.2012г. [3] В Программе указывается, что на современном этапе в условиях осуществления в России социально-экономических перемен, административной реформы, реформирования бюджетного процесса, перехода на международные стандарты финансовой отчетности, необходимы новые подходы к статистической деятельности, формированию гибкой, быстро

реагирующей на изменение жизни общества системы информационно-статистического обеспечения. Поэтому, дисциплина «Статистика» ориентирована на подготовку и переподготовку бакалавров высшей квалификации экономических направлений с целью овладения методологией сбора и обработки всех видов информации, а также всестороннего статистического анализа состояния и развития массовых общественных явлений в рыночной экономике.

Целью учебного пособия «Статистика» является систематизация изложения вопросов, отражающих методологию применения комплекса статистических методов в анализе массовых общественных явлений в условиях рыночной экономики в соответствии с Федеральным государственным образовательным стандартом (ФГОС ВПО).

В учебное пособие включены основные теоретические вопросы и практические примеры по определению статистических показателей и применению комплекса статистических методов, способов и приемов анализа массовых общественных явлений, которые максимально приближены к практике статистических работ и соответствуют современным требованиям перехода России на международные стандарты статистики и учета, а также требованиям подготовки специалистов высшего профессионального образования для двухуровневого образования – бакалавриата и магистратуры. Практические примеры предполагают применение общедоступных прикладных компьютерных программ Microsoft Word и Microsoft Excel.

Теоретические аспекты, практические примеры и контрольные вопросы охватывают 9 разделов дисциплины «Статистика», включающие 10 тем в рамках данных разделов.

Представленный в учебном пособии теоретический и практический материал направлен на формирование следующих компетенций подготовки бакалавриата направления «Экономика»: ОК-4 - способен анализировать социально-значимые проблемы и процессы, происходящие в обществе, и прогнозировать возможное их развитие в будущем; ПК-4 - способен

осуществлять сбор, анализ и обработку данных, необходимых для решения поставленных экономических задач; ПК-5 - способен выбрать инструментальные средства для обработки экономических данных в соответствии с поставленной задачей, проанализировать результаты расчетов и обосновать полученные выводы; ПК-8 - способен анализировать и интерпретировать данные отечественной и зарубежной статистики о социально-экономических процессах и явлениях, выявлять тенденции изменения социально-экономических показателей; ПК-10 - способен использовать для решения аналитических и исследовательских задач современные технические средства и информационные технологии; ПК-12 - способен использовать для решения коммуникативных задач современные технические средства и информационные технологии.

1 ПРЕДМЕТ, МЕТОД И ЗАДАЧИ СТАТИСТИКИ

1.1 Значение и основное содержание статистики

Возникновение статистики было обусловлено объективными практическими потребностями общества, когда развитие хозяйствования достигло уровня, при котором появилась необходимость подсчета отдельных его элементов: населения, скота, земельных угодий, имущества и др. По мере развития общественного производства, внутренней и внешней торговли увеличивалась потребность в статистической информации. Это расширило сферу деятельности практической статистики, привело к совершенствованию ее приемов и методов. Многообразная практическая статистическая деятельность и практика учетно-статистических работ стала подвергаться теоретическим обобщениям, сформировалась группа ученых, непосредственно занимающихся развитием теории, методологии и практики статистики, вследствие чего началось формирование статистики как науки, возникновение которой относят к середине XVIII века.

Термин «статистика» произошел от латинского слова «status», что означает состояние дел, положение вещей. Термин «статистика» был введен в употребление в 1746г. немецким философом, историком, экономистом, юристом и педагогом Готфридом Ахенвалем (1719-1772г.г.), который впервые стал преподавать «Статистику» как учебную дисциплину в Марбургском университете Германии с 1747г.

В современном обществе статистика является одним из важнейших инструментов управления обществом. Она обеспечивает сбор, обработку, систематизацию, анализ и представление цифровой информации об уровне и возможностях развития массовых общественных явлений и процессов.

Эффективное развитие страны, государственное управление и регулирование связаны с необходимостью своевременного получения и анализа

полной, достоверной, научно обоснованной статистической информации о социальных, экономических, демографических и других общественных явлениях. Поэтому статистическая информация является базой для разработки объективной и научно обоснованной социально-экономической политики государства. Развитие государственной статистики в РФ способствует достижению стратегических целей развития страны, информационно-статистическому обеспечению реализации программ социально-экономического развития, нормативных правовых актов Президента РФ и Правительства РФ, а также выполнению международных информационных обязательств в целях более полной международной интеграции экономики РФ в мировое сообщество.

В современных условиях понятие «статистика» рассматривается в четырех основных значениях (рис. 1).

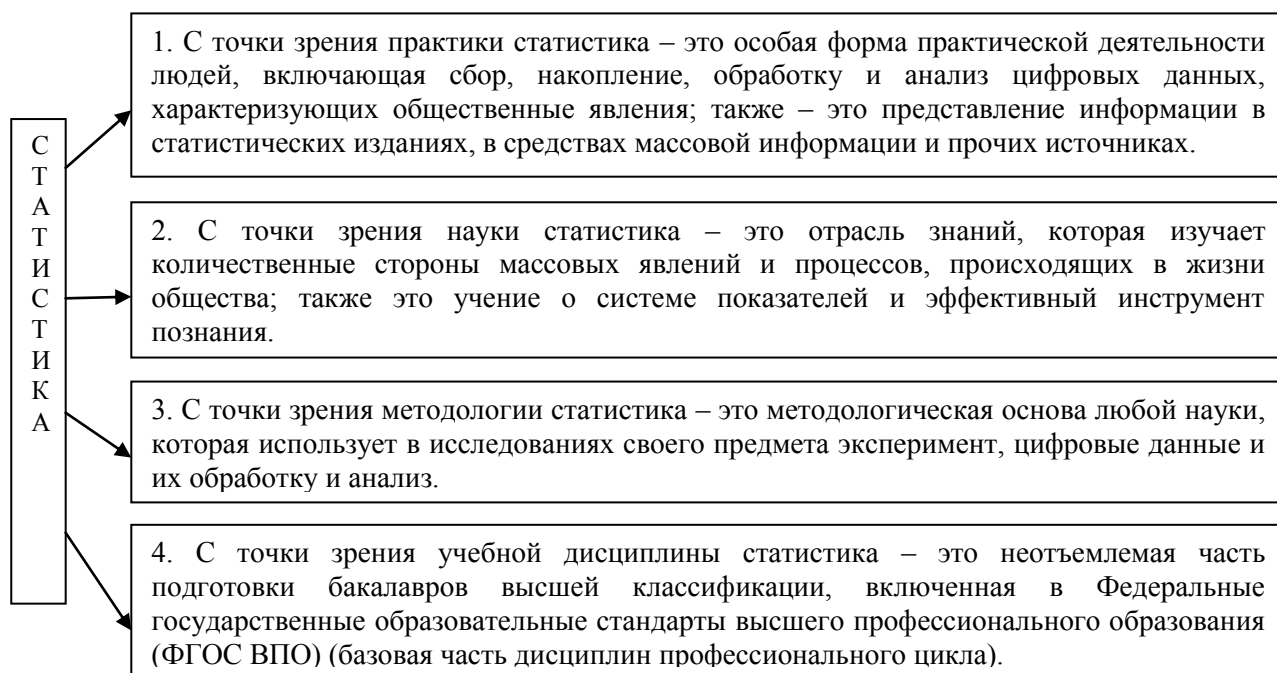


Рисунок 1 – Значение статистики

В этих условиях особое место отводится отрасли статистической науки - общей теории статистики, которая является инструментом, обеспечивающим

теоретическую и методологическую подготовку экономистов, финансистов, коммерсантов, бухгалтеров, аудиторов и других профессионалов высшей квалификации.

1.2 Объект изучения, теоретические основы и предмет статистики

Статистика решает информационные, познавательные и аналитические задачи, изучает закономерности развития общества в целом и отдельных его элементов. Следовательно, объектом изучения статистики является общество, состоящее из совокупности общественных явлений и процессов.

Общественные явления и процессы многообразны, сложны и изменчивы. Они претерпевают количественные и качественные изменения. Особенно изменчива их количественная сторона. При этом, общественные явления и процессы находятся во взаимосвязи и взаимообусловленности. Статистика изучает эту количественную сторону в процессе ее изменения и взаимосвязи.

Поэтому предметом изучения статистики является количественная характеристика массовых общественных явлений в неразрывной связи с их качественной характеристикой в конкретных условиях места и времени.

В определении предмета статистики можно выделить 4 основных особенности, которые отличают статистику от других наук:

- 1) статистика изучает массовые общественные явления, то есть имеет дело с совокупностями явлений и единиц;
- 2) статистика имеет дело с количественными характеристиками, то есть с величинами явлений;
- 3) статистика рассматривает количественные характеристики в неразрывной связи с качественными характеристиками общественных явлений, т.к. количество и качество есть неразрывные стороны явления, не существующие отдельно друг от друга и от самого явления;

4) статистика изучает общественные явления в их исторической определенности, то есть в конкретных условиях места и времени, т.к. любое явление существует лишь в конкретной временной и пространственной определенности, и при изменении одной из них, меняет свою сущность и само явление.

Теоретической основой статистики являются диалектика, которая формирует законы общественного развития, и экономическая теория, которая формирует экономические законы развития общества и раскрывает сущность социально-экономических явлений и процессов.

Исходя из законов и принципов диалектики и экономической теории, статистика дает конкретную цифровую характеристику состоянию и закономерностям развития общественных явлений.

То есть, статистика в соответствии с законами диалектики и экономической теории: 1) изучает все общественные явления и процессы в их взаимосвязи; 2) изучает все общественные явления и процессы в движении и изменении; 3) выделяет различные типы и формы общественных явлений и процессов; 4) устанавливает новое и прогрессивное, что зарождается в существующей действительности и определяет направление его развития; 5) на основе изучения количественных характеристик и анализа их изменений оценивает качественное изменение общественных явлений и процессов; 6) переносит свойства одних общественных явлений и процессов на другие в соответствии со сходством закономерностей их развития; 7) на основе изучения закономерностей развития общественных явлений и процессов определяет причинно-следственные связи.

1.3 Основные категории статистической науки

При изучении своего объекта статистика использует ряд особых понятий и терминов, составляющих ее специфический язык. Комплекс этих особых

понятий и терминов и составляет основные категории статистической науки. К основным категориям статистики относятся: (рис. 2).



Рисунок 2 – Основные категории статистической науки

Статистическая совокупность – это множество явлений, объединенных по одному признаку и различных между собой по другим признакам, существование которого ограничено во времени и в пространстве. Статистическая совокупность обладает следующими свойствами:

- массовость - вытекает из основной особенности статистики – изучения массовых общественных явлений, так как закономерность развития изучаемых явлений и процессов проявляется лишь в массе наблюдений;
- однородность – подчиненность элементов, составляющих совокупность, общему закону развития. В соответствии с этим статистическая совокупность может быть однородной (если существенные признаки для каждой ее единицы в основном одинаковы) и разнородной (если статистическая совокупность объединяет различные типы явлений). Причем, одна и та же статистическая совокупность может быть одновременно и однородной (по одному признаку) и разнородной (по всем другим признакам);

- стабильность – сохранение состава совокупности неизменным в течение некоторого периода времени. В соответствии с этим статистическая совокупность может быть стабильной (если ее состав сохраняется в течение определенного периода времени) и динамичной (если ее состав постоянно изменяется);
- неделимость – возможность исключения из совокупности отдельных единиц, так как это не приведет к потере ее качественного единства.

Единица статистической совокупности – первичный элемент, который является непосредственным носителем признака, в единстве составляет совокупность и выражает ее качественную определенность.

Признак – свойство или характерная черта единицы статистической совокупности, которые могут быть наблюдаемы или измерены в процессе их изучения. Признаки классифицируются на следующие виды (рис. 3).

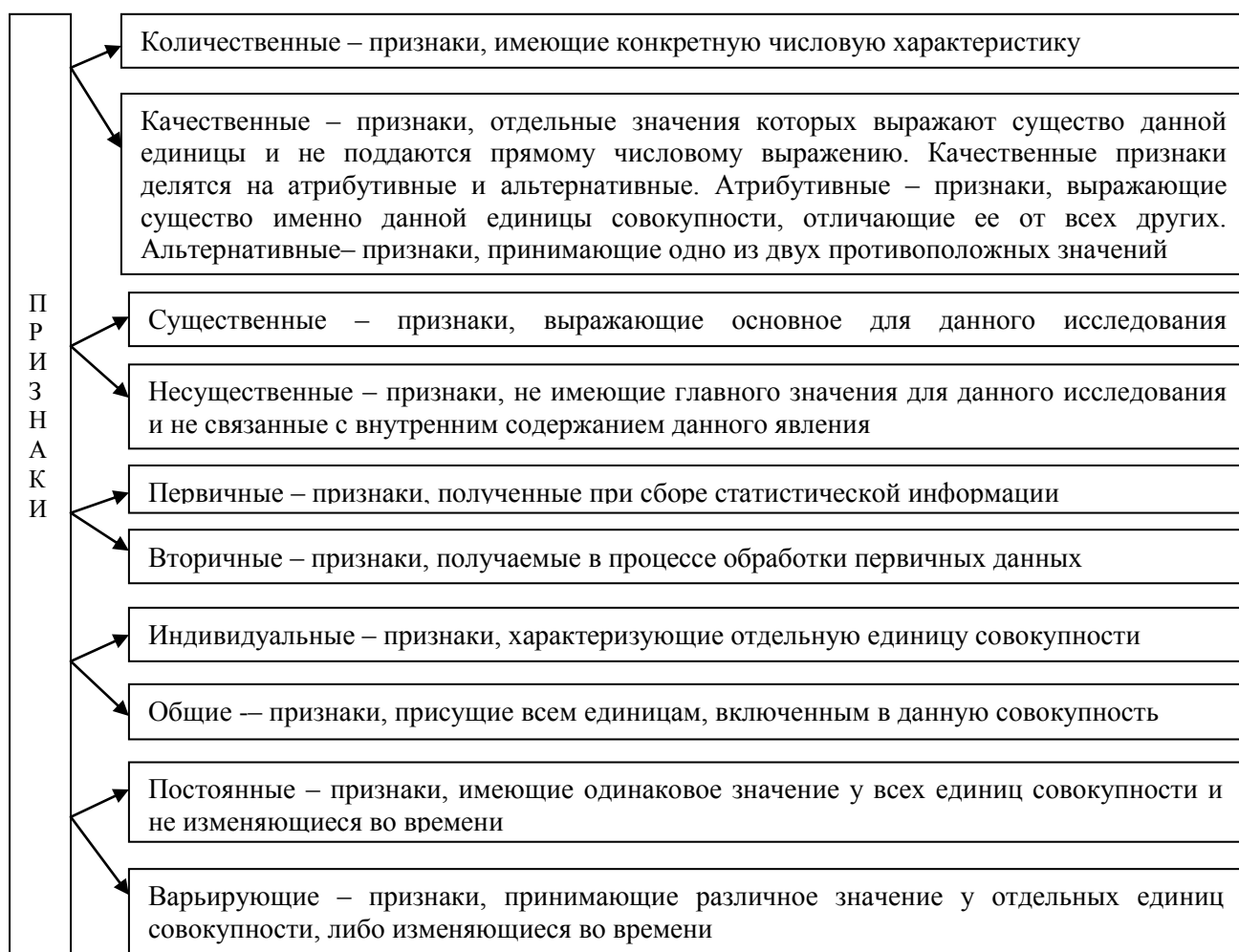


Рисунок 3 – Виды признаков

Вариация – это изменчивость, колеблемость и разнообразие величины признака. Пределы, в которых возможно изменение признака, называются границами вариации: верхняя граница – максимальное значение признака, нижняя граница – минимальное значение признака. Отдельное значение признака у каждой единицы совокупности называется вариантом.

Статистический показатель – обобщенная количественная характеристика массовых общественных явлений во взаимосвязи с их качественной характеристикой в конкретных условиях места и времени. Статистические показатели бывают индивидуальные (характеризуют отдельные явления) и обобщающие (характеризуют массу явлений). Обобщающие статистические показатели выступают в форме абсолютных, относительных и средних величин.

Система статистических показателей – это множество взаимосвязанных показателей, характеризующих общественное явление со всех сторон. Поскольку общественные явления многообразны, изменчивы и взаимосвязаны между собой, то для полного их изучения недостаточно одной характеристики, а необходимо несколько взаимосвязанных характеристик, образующих систему статистических показателей.

Статистическая закономерность – это объективная количественная закономерность массового процесса, которая возникает под действием объективных законов и обнаруживается в результате массового статистического наблюдения.

1.4 Методологическая основа и метод статистики

Методологической основой статистики является диалектическая теория познания, определяющая научный подход к изучению общественных явлений. Это означает, что статистика изучает общественные явления и процессы в двух основных аспектах: 1) не изолированно, а в их взаимодействии и взаимосвязи; 2) не в состоянии покоя и неизменности, а в движении и развитии.

На основе общенаучного диалектического метода познания статистика разработала свой специфический метод.

Метод статистики – это совокупность приемов и способов, с помощью которых статистика изучает свой предмет.

Специфика статистического метода заключается в том, что он выражается в трех стадиях статистического исследования (рис. 4).

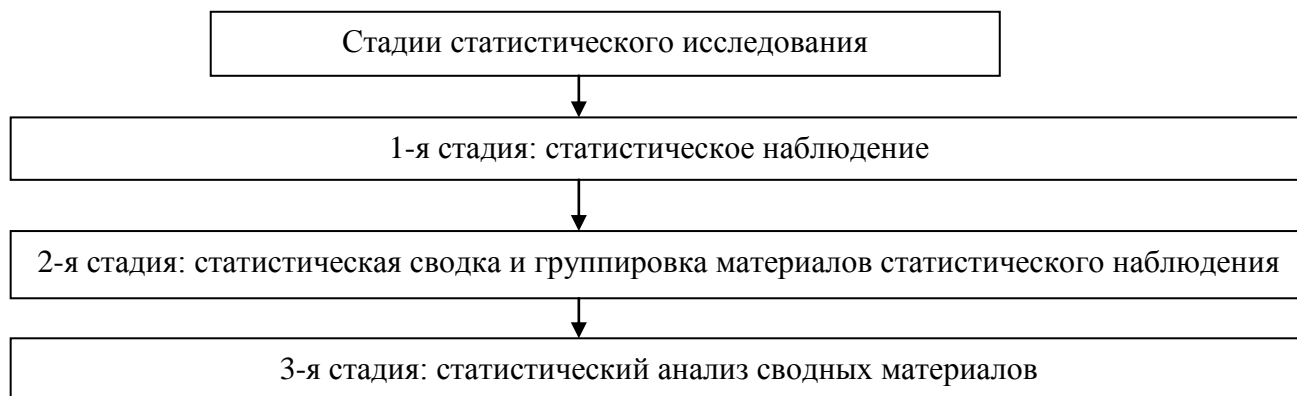


Рисунок 4 – Основные стадии статистического исследования

Первая стадия - статистическое наблюдение – это планомерный научно организованный и систематический процесс сбора массовых цифровых данных об изучаемом явлении с целью его дальнейшего исследования. Характерной чертой для этой стадии является массовость наблюдений, основанная на основном принципе статистики. На данной стадии получают первичную информацию об отдельных единицах изучаемой статистической совокупности. Статистическое наблюдение предоставляет исходный материал для последующих обобщений и формулировки объективных научно обоснованных выводов о состоянии и развитии изучаемого явления. Получение достоверных сведений о достаточно большом числе единиц на данной стадии дает возможность освободиться от влияния случайных причин и установить закономерности развития явления и процесса.

Вторая стадия - сводка и группировка материалов наблюдения. На данной стадии производится обобщение единичных фактов с целью выявления

типичных черт и закономерностей, присущих уже не отдельным единицам, а явлению в целом.

Статистическая группировка предполагает разделение совокупности на однородные группы по существенному варьирующему признаку, дающее возможность выделить из состава совокупности единицы, объединенные общим законом развития, показать особенности явлений.

Статистическая сводка предполагает подсчет общих и групповых итогов и выявление закономерностей развития явления и процесса. Результаты статистической сводки и группировки обобщаются в виде статистических таблиц и графиков, которые являются наиболее рациональной, систематизированной и наглядной формой выражения массовых данных.

Третья стадия - анализ сводных материалов. Анализ производится по следующим этапам: 1) констатация фактов и их оценка; 2) установление характерных черт явления и причинно-следственных связей; 3) сопоставление явления с другими, принятыми за базу сравнения; 4) формулировка гипотез; 5) статистическая проверка выдвинутых гипотез; 6) формулировка выводов и предложений по результатам анализа.

Характерным для данной стадии является применение обобщающих показателей: абсолютных, относительных, средних величин и индексных систем. На данной стадии:

- определяется состав явления при помощи относительных величин структуры;
- определяется изменение явления во времени при помощи построения и анализа рядов динамики;
- выявляются причинно-следственные связи при помощи корреляционно-регрессионного анализа и многомерного анализа;
- производится увязка отдельных показателей в систему при помощи балансового метода;
- производится наглядное представление результатов исследования при помощи графического и табличного метода.

1.5 Общность и различие статистических дисциплин. Связь статистики с другими науками

Общество в процессе своего развития ставит перед статистикой новые задачи, что приводит к выделению отдельных отраслей в единой статистической науке (рис. 5). Каждая из отраслей статистики имеет свой объект исследования, выясняет сущность определенной системы показателей, разрабатывает правила и методы их получения и использования в научной и практической деятельности.

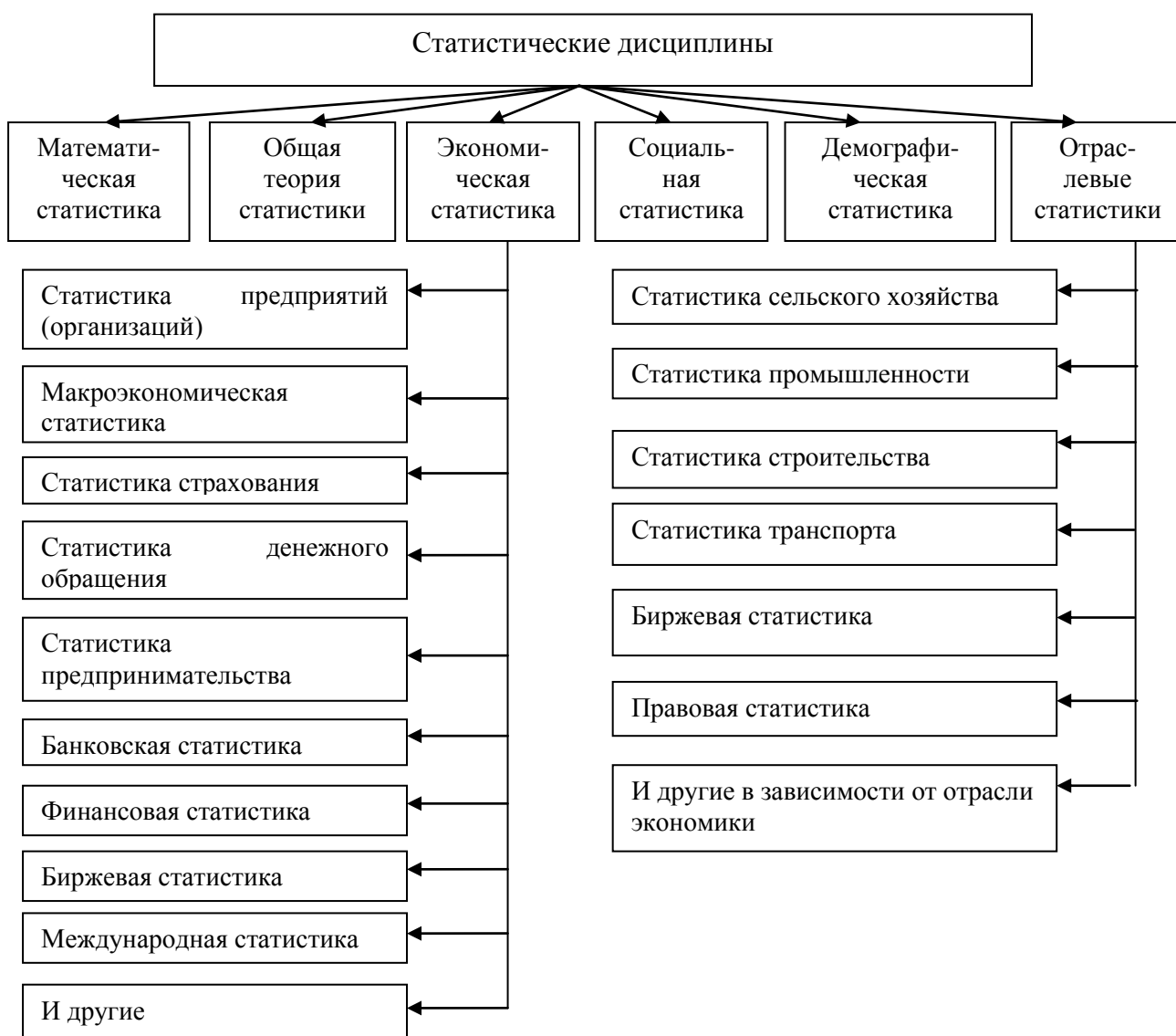


Рисунок 5 – Основные статистические дисциплины

Математическая статистика предполагает широкое использование различных математических методов в процессе выявления и анализа статистических закономерностей: дисперсионный анализ, регрессионный анализ, корреляционный анализ, анализ параллельных рядов, проверка статистических гипотез, выборочный метод и другие. На основе этих методов и теории вероятности математическая статистика исследует закономерности случайных явлений и процессов, проявляющихся в массе наблюдений. Характерной особенностью математической статистики является получение сводных количественных характеристик массовых совокупностей. Большое применение здесь находят математические теоремы Закона больших чисел. Закон больших чисел – «это общий принцип, в силу которого совокупное действие большого числа случайных факторов приводит при некоторых общих условиях к результату, почти не зависящему от случая». Это означает, что при достаточно большом числе наблюдений случайное влияние факторов устраняется и выявляется закономерность развития явления.

Общая теория статистики разрабатывает общие понятия, категории статистической науки, которые имеют общестатистический смысл и формирует методы, приемы и способы статистического исследования. Общая теория статистики показывает сущность статистических терминов, категорий и понятий, разрабатывает научную методологию их построения. Категориями, показателями и методами общей теории статистики пользуются все остальные статистические дисциплины и большинство других наук. Общая теория статистики является методологической основой большинства наук.

Экономическая статистика изучает явления и процессы в области экономики: структуру, пропорции, взаимосвязи элементов экономической системы. Дает количественные характеристики общественного воспроизводства, оценивает его эффективность, измеряет степень влияния факторов на развитие экономики и выявляет резервы повышения ее эффективности. В современных условиях экономическая статистика делится на статистику предприятий (организаций), изучающую процесс общественного

воспроизводства на микроуровне; макроэкономическую статистику, изучающую процесс общественного воспроизводства на макроуровне; статистику страхования, денежного обращения, предпринимательства, банковскую статистику, финансовую статистику и другие.

Социальная статистика изучает социальные явления и процессы, которые характеризуют условия жизнедеятельности людей, их взаимоотношения в процессе труда и внетрудовой деятельности, социально-экономическое положение и уровень жизни населения.

Демографическая статистика изучает население, демографические процессы, происходящие в совокупности населения страны, его изменение, факторы, влияющие на демографическую ситуацию, разрабатывает основные демографические прогнозы.

Отраслевые статистики изучают количественные характеристики массовых явлений в отдельных отраслях экономики: статистика промышленности; сельскохозяйственная статистика; статистика строительства; статистика транспорта; медицинская статистика; уголовно-правовая статистика и другие.

Статистика тесно связана со многими науками, так как одни науки являются ее основой, а другие сами основаны на методах, разработанных статистикой, на информации, собранной и опубликованной статистикой и пр.:

- 1) теоретической основой статистики является диалектика и экономическая теория, формирующие общие и экономические законы развития общества;
- 2) методологической основой статистики является диалектика, формирующая законы познания;
- 3) математика формирует методы, используемые при анализе массовых статистических данных;
- 4) теория вероятности формирует методы проверки статистических гипотез, прогнозирования развития общественных явлений;

5) фундаментальные науки (физика, химия, биология, астрономия другие) используют при проведении экспериментов и анализе полученных результатов статистические методы обработки цифровых данных;

6) экономика использует систему показателей, разработанную экономической статистикой;

7) менеджмент использует статистические методы в разработке управленческих решений и оценке их эффективности;

8) маркетинг использует статистические методы в анализе рынков сбыта и оценке эффективности маркетинговой деятельности;

9) социология и демография основаны на применении статистической информации, полученной при переписи населения и статистические методы ее обработки;

8) наиболее существенная взаимосвязь прослеживается между статистикой и учетом в соответствии с реализацией в РФ международных стандартов учета и статистики. Выделяется четыре уровня учета по степени обобщения информации (рис. 6):

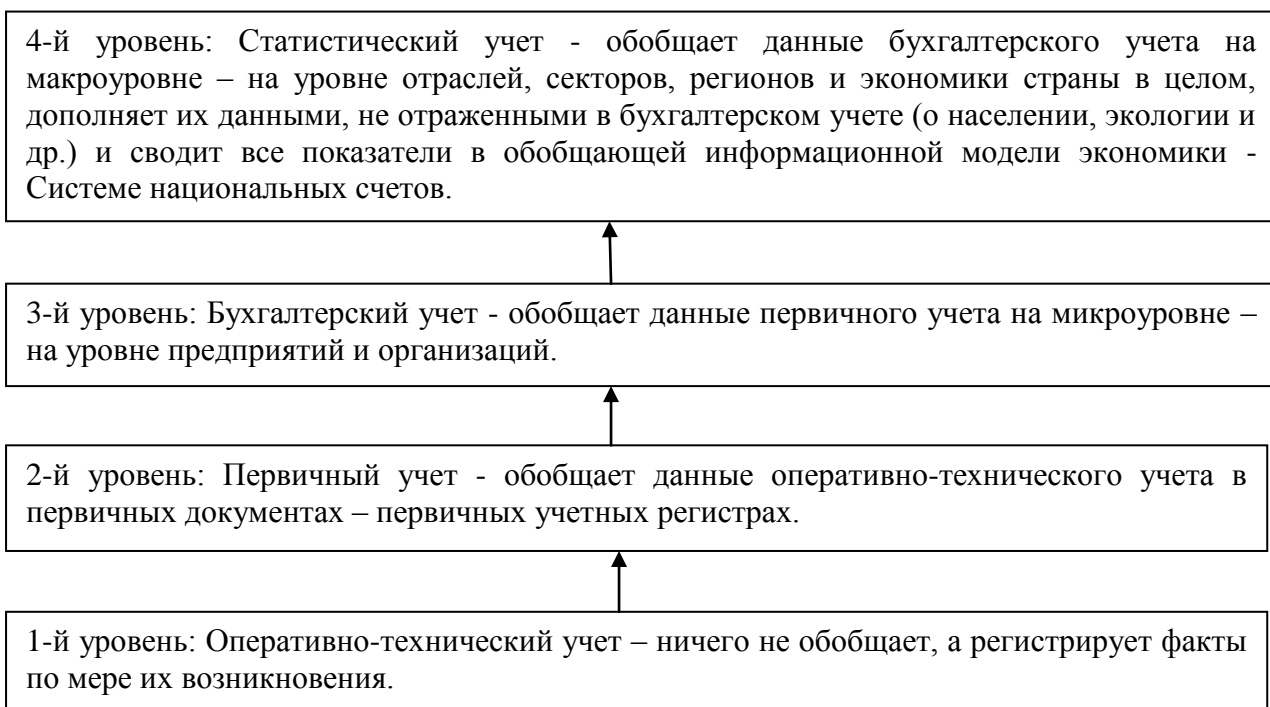


Рисунок 6 – Уровни учета по степени обобщения информации

1.6 Задачи статистики

В соответствии с реализацией Федеральной целевой программы «Развитие государственной статистики в России в 2007-2011 годах» (с изменениями и дополнениями в ред. 2012г.) определено, что на современном этапе государственная статистика развивается в условиях осуществления в России социально-экономических перемен, административной реформы, реформирования бюджетного процесса, перехода на международные стандарты финансовой отчетности, что определяет задачи поиска и реализации новых подходов к официальной статистической деятельности, формированию гибкой, быстро реагирующей на изменение жизни общества системы информационно-статистического обеспечения. [3]

В данных условиях перед статистикой ставятся следующие задачи, отвечающие современным требованиям:

1. Всестороннее, достоверное и оперативное освещение социально-экономического положения страны, происходящих изменений, связанных с функционированием рыночной экономики.
2. Удовлетворение потребностей органов власти и управления, средств массовой информации, общественности, предпринимателей и международных организаций в объективной и полной статистической информации.
3. Реформирование методологии, теории и практики статистической деятельности в соответствии с международными стандартами учета и статистики (МСФО).
4. Построение системы показателей, отражающих современное состояние страны и соответствующей международным стандартам; полный переход на Систему национальных счетов (СНС), принятую в международной практике учета и статистики и обобщающую показатели развития экономики страны на макроуровне.
5. Совершенствование системы сбора и обработки статистической информации и организации статистического наблюдения - полный переход на формы

наблюдения - регистры (регистры населения, регистры налогового учета, регистры агентов рынка), цензы и другие; создание постоянно действующих мониторингов - специально организованных систематических наблюдений.

6. Обеспечение компьютеризации статистики в соответствии с программой информатизации РФ, обеспечение программной и технической совместимости различных автоматизированных систем на федеральном и региональном уровнях. Использование программ: 1) STform «Статистические формы» (предприятие), разработанной НИПИСтатинформ Росстата для предоставления статистической отчетности в электронном виде; 2) «Программное обеспечение «Формы статотчетности» (StFormsInstaller)», определяющее порядок представления статистической информации для проведения государственных статистических наблюдений, бланки форм федерального государственного статистического наблюдения, табель форм статистической отчетности, документацию, программное обеспечение, XML-шаблоны форм статистической отчетности в электронном виде; 3) программное обеспечение «СТАТИСТИКА», разработанное специально для Федеральной службы государственной статистики РФ и для всех предприятий, подающих статистическую отчетность, в рамках проекта ФСС РФ по автоматизации представления первичной статистической информации от предприятий в территориальные органы статистики для упрощения процесса представления статистической отчетности за счет встроенных в программу формул контроля, готового списка бланков и представления отчетности в электронном виде для заполнения пользовательских форм отчетности; 4) программа СТАТИСТИКА beta - версия для организации статистического наблюдения за выполнением федеральных целевых программ; 5) комплексная автоматизированная система «АС ВПН» для обработки материалов переписи населения в два этапа: на региональном уровне и на федеральном уровне и другие программы ЭВМ.

7. Полное внедрение информационно-телекоммуникационной системы статистики (ИТСС), строящейся на основе создания локальных вычислительных сетей (ЛВС) во всех органах государственной статистики,

которая позволит перейти к информационной технологии, максимально автоматизирующей разработку статистической отчетности и информационно-справочное обслуживание органов власти и управления, предприятий и населения, полный переход к комплексам регламентной обработки информации (КРО), основой которой является автоматизация формирования статистической информации.

8. Расширение доступности сводной статистической информации при сохранении принципа конфиденциальности индивидуальных данных; доступности открытых статистических данных и прозрачности методологии их формирования для всех пользователей;

9. Разработка и внедрение методов обеспечения конфиденциальности первичных статистических данных, полученных от респондентов.

10. Применение международных статистических стандартов и классификаций при определении национальной системы статистических показателей и методов их формирования.

11. Снижение нагрузки на респондентов путем расширения практики использования в статистических целях административных данных, то есть документированной информации, полученной федеральными органами государственной власти, органами государственной власти субъектов РФ, органами местного самоуправления и государственными организациями в связи с осуществлением ими своих полномочий.

12. Повышение качества официальной статистической информации; создание и внедрение корпоративной системы документооборота. [3]

Кроме этого, статистика как наука и как специфическая практическая деятельность решает комплекс информационных, аналитических и методологических задач в различных областях функционирования экономики, государства и общества в целом.

1.7 Организация статистики в Российской Федерации

Изучением социально-экономического развития страны, отдельных ее регионов, отраслей, предприятий занимаются специально созданные органы, совокупность которых называется статистической службой. В Российской Федерации функции статистической службы выполняют органы государственной статистики РФ и органы ведомственной статистики.

В соответствии с Федеральным законом от 29.11.2007г. N 282-ФЗ «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» (с изменениями и дополнениями от: 19.10.2011г., 16.10.2012г., 2.07.2013г.) [2] в основу организации статистической работы в РФ положены следующие основные принципы: 1) централизованное руководство; 2) единое организационное строение и методология; 3) неразрывная связь статистических органов с органами государственного управления.

Принципами официального статистического учета и системы государственной статистики в РФ являются

1) полнота, достоверность, научная обоснованность, своевременность предоставления и общедоступность официальной статистической информации (за исключением информации, доступ к которой ограничен федеральными законами);

2) применение научно обоснованной статистической методологии, соответствующей международным стандартам и принципам официальной статистики, а также законодательству РФ, ее открытость и доступность;

3) рациональный выбор источников в целях формирования официальной статистической информации для обеспечения ее полноты, достоверности и своевременности предоставления, а также в целях снижения нагрузки на респондентов;

4) обеспечение возможности формирования официальной статистической информации по Российской Федерации в целом, по субъектам Российской Федерации, по муниципальным образованиям;

5) обеспечение конфиденциальности первичных статистических данных при осуществлении официального статистического учета и их использование в целях формирования официальной статистической информации;

6) согласованность действий субъектов официального статистического учета;

7) применение единых стандартов при использовании информационных технологий и общероссийских классификаторов технико-экономической и социальной информации для создания и эксплуатации системы государственной статистики в целях ее совместимости с другими государственными информационными системами;

8) обеспечение сохранности и безопасности официальной статистической информации, первичных статистических и административных данных. [2]

Главным учетно-статистическим органом в РФ является Федеральная служба государственной статистики РФ (ФСС) - Росстат. Он осуществляет руководство российской статистикой в соответствии со статьей 71 Конституции РФ, Федеральным Законом «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» от 29.11.2007г. №282-ФЗ [2], Постановлением Правительства РФ «О федеральной службе государственной статистики» от 2.06.2008г. № 420 [5] и Указом Президента РФ от 12.05.2008г. №724 «Вопросы системы и структуры федеральных органов исполнительной власти» [6].

Росстат является федеральным органом исполнительной власти. В его задачи входит 1) представление официальной статистической информации Президенту РФ, Правительству РФ, Федеральному собранию РФ, федеральным органам исполнительной власти РФ, общественным и международным организациям; 2) разработка научно обоснованной статистической методологии; 3) координация статистической деятельности федеральных и региональных органов исполнительной власти; 4) анализ экономико-статистической информации; 5) составление национальных счетов и балансовых расчетов; 6) разработка форм статистической и бухгалтерской

отчетности; 7) формирование бухгалтерских и статистических показателей в соответствии с требованиями международного бухгалтерского учета и системы национальных счетов; 8) обеспечение контактов с международными статистическими службами ООН, в частности с ее Статистической комиссией.

Методы сбора и обработки статистических данных, методология исчисления статистических и бухгалтерских показателей, установленные Росстатом, являются официальными стандартами РФ. Управляется Росстат Председателем. При Председателе действует Совет руководителей органов государственной статистики, созданный для координации деятельности региональных органов государственной статистики.

Для целей исполнения Федерального закона используются следующие основные понятия [2]:

1) официальный статистический учет - деятельность, направленная на проведение в соответствии с официальной статистической методологией федеральных статистических наблюдений и обработку данных, полученных в результате этих наблюдений, и осуществляемая в целях формирования официальной статистической информации;

2) система государственной статистики - государственная федеральная информационная статистическая система, представляющая собой совокупность позволяющих осуществлять официальный статистический учет первичных статистических данных и административных данных, формируемой на их основе в соответствии с официальной статистической методологией, официальной статистической информацией и обеспечивающих формирование такой информации информационных технологий и технических средств;

3) официальная статистическая информация - сводная агрегированная документированная информация о количественной стороне социальных, экономических, демографических, экологических и других общественных процессов в РФ, формируемая субъектами официального статистического учета в соответствии с официальной статистической методологией;

4) субъекты официального статистического учета - федеральные органы государственной власти, иные федеральные государственные органы, осуществляющие формирование официальной статистической информации в установленной сфере деятельности в соответствии с законодательством РФ;

5) федеральное статистическое наблюдение - сбор первичных статистических данных и административных данных субъектами официального статистического учета;

6) административные данные - используемая при формировании официальной статистической информации документированная информация, получаемая федеральными органами государственной власти, иными федеральными государственными органами, органами государственной власти субъектов РФ, иными государственными органами субъектов РФ, органами местного самоуправления, государственными организациями в связи с осуществлением ими разрешительных, регистрационных, контрольно-надзорных и других административных функций, а также иными организациями, на которые осуществление указанных функций возложено законодательством РФ;

7) первичные статистические данные - документированная информация по формам федерального статистического наблюдения, получаемая от респондентов, или информация, документируемая непосредственно в ходе федерального статистического наблюдения;

8) пользователи официальной статистической информацией - государственные органы, органы местного самоуправления, юридические и физические лица, обращающиеся к системе государственной статистики или субъектам официального статистического учета за получением необходимой им статистической информации и (или) пользующиеся такой информацией. [2]

В состав Росстата входят следующие управления (рис. 7). Росстат имеет подведомственные органы: 1) Федеральное государственное унитарное предприятие Главный межрегиональный центр обработки и распространения статистической информации ФСС (ФГУП ГМЦ Росстата); 2) Федеральное

государственное унитарное предприятие Научно-исследовательский институт проблем социально-экономической статистики ФСС (ФГУП НИИ статистики Росстата); 3) Федеральное государственное унитарное предприятие Научно-исследовательский и проектно-технологический институт статистической информационной системы ФСС (ФГУП НИПИ статинформ Росстата); 4) Автономная некоммерческая организация информационно-издательский центр «Статистика России»; 5) Федеральное государственное унитарное предприятие «Управление по эксплуатации зданий ФСС» (ФГУП УЭЗ Росстата).

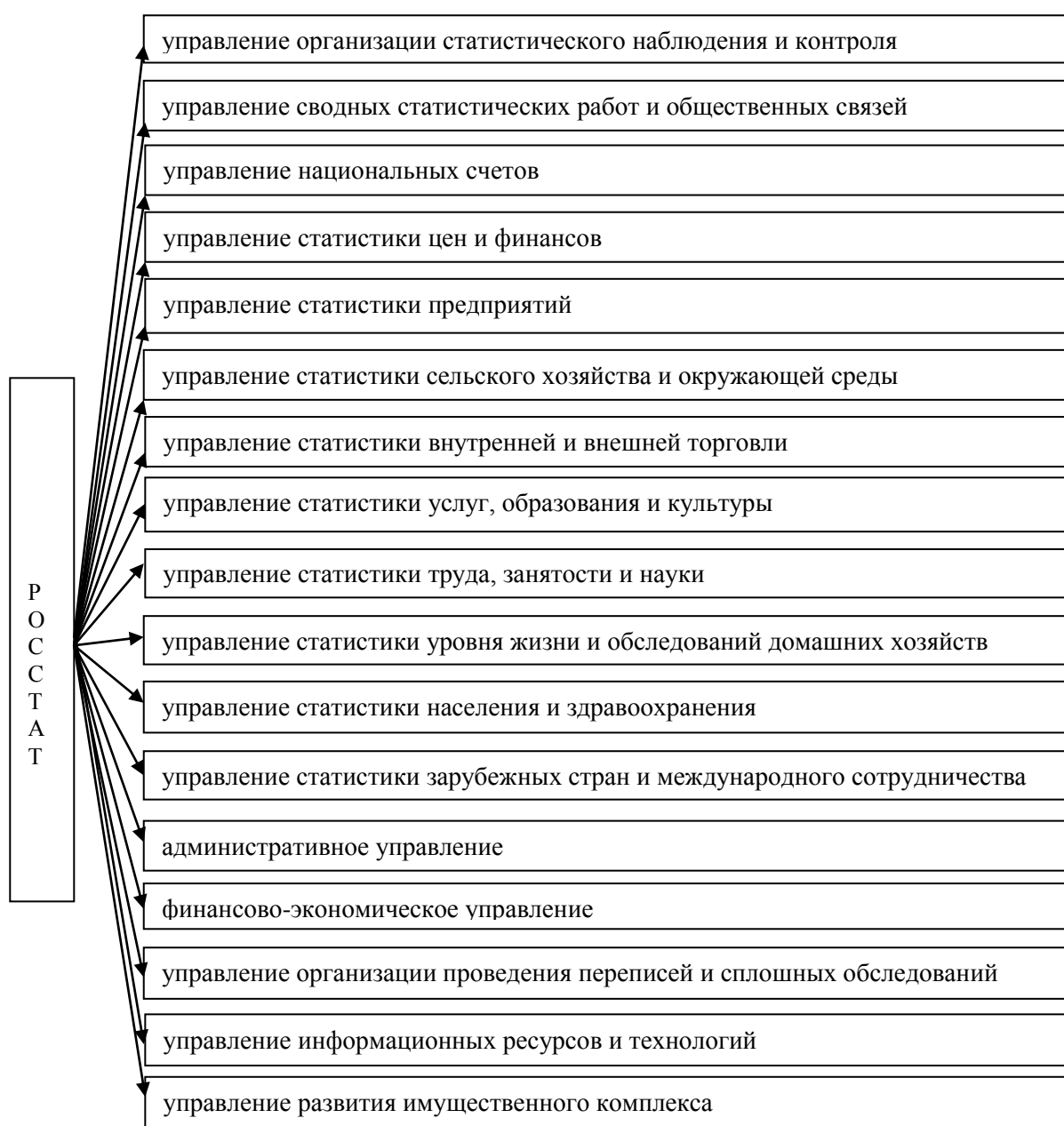


Рисунок 7 – Структура Росстата

Система государственной статистики находится в ведении Правительства РФ и ему подотчетна, входит в структуру и подотчетна Министерству экономического развития РФ; имеет иерархическую структуру, создана по административно-территориальному делению и включает федеральный, республиканский, краевой, областной, окружной, городской и районный уровни. Каждый административно-территориальный субъект имеет свой Территориальный орган федеральной службы государственной статистики РФ с иерархическим подчинением. Структура территориальных органов ФСС на территориальном уровне в основном повторяет структуру Росстата.

Территориальный орган Федеральной службы государственной статистики по Амурской области включает следующие подразделения (рис. 8):

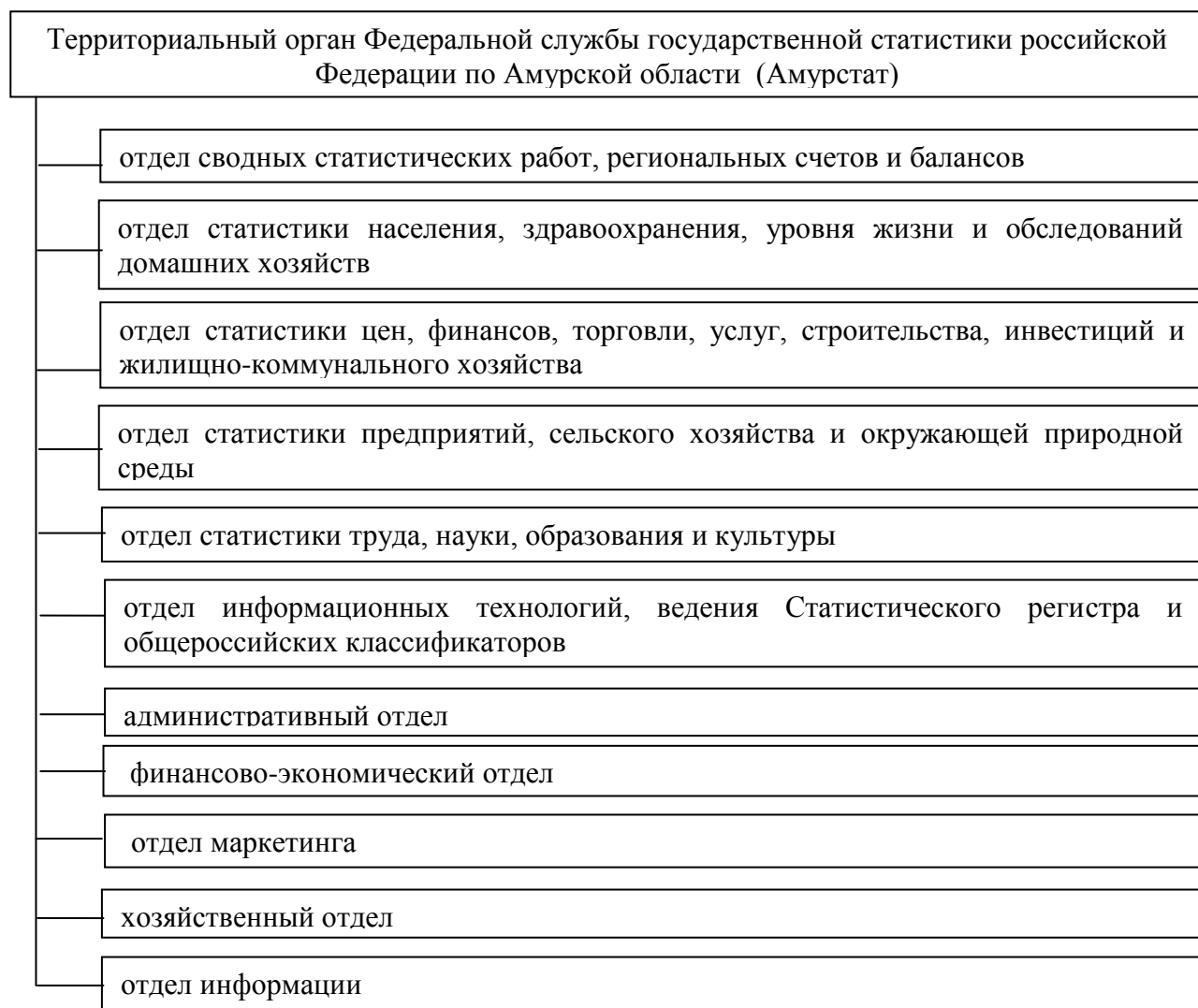


Рисунок 8 – Структура Территориального органа Федеральной службы государственной статистики РФ по Амурской области

2 СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ МАССОВЫХ ОБЩЕСТВЕННЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ

2.1 Сущность, значение и задачи статистического наблюдения

Статистическое исследование начинается со сбора сведений об изучаемом массовом общественном явлении и процессе. Такой сбор сведений называется статистическим наблюдением. Статистическое наблюдение – это первая и важнейшая стадия статистического исследования, на которой формируется статистический материал – первичная статистическая информация.

Важность и значение данной стадии заключается в том, что даже небольшие неточности, допущенные в процессе статистического наблюдения, приводят к значительным искажениям результатов всего статистического исследования.

Статистическое наблюдение – это планомерный, научно организованный систематический процесс сбора данных об изучаемых массовых общественных явлениях с целью их дальнейшего исследования.

Полученный в результате статистического наблюдения статистический материал регистрируется в специальных учетных документах – регистрах для последующего его обобщения и анализа. Занесенный в учетные регистры статистический материал называется статистической информацией, которая является фундаментом статистического исследования.

Статистическая информация – это первичный статистический материал, который формируется в процессе статистического наблюдения и впоследствии подвергается обработке, систематизации, обобщению и анализу.

Задачи статистического наблюдения.

1. Обеспечение полноты охвата фактов, относящихся к изучаемому массовому общественному явлению и процессу.
2. Обеспечение достоверности статистической информации.

3. Выбор массовых общественных явлений, имеющих научную и практическую значимость.
4. Обеспечение оперативности получения статистической информации.
5. Обеспечение сопоставимости полученной статистической информации во времени и пространстве для возможности ее сравнения.
6. Учет факторов и событий, под влиянием которых происходит изменение изучаемого явления и процесса.
7. Обеспечение возможности изучения закономерности развития изучаемого явления и процесса.

2.2 Этапы проведения статистического наблюдения

Поскольку статистическое наблюдение, это научно организованный и планомерный процесс, то оно осуществляется последовательно в несколько этапов (рис. 9):



Рисунок 9 – Этапы проведения статистического наблюдения

Первый этап - подготовка к проведению статистического наблюдения. Заключается в комплексе подготовительных мероприятий, предшествующих непосредственному сбору статистической информации. Это важнейший этап, от которого зависит успех проведения всего статистического наблюдения и достоверность полученных результатов. Подготовительный этап включает:

- 1) определение цели и задач статистического наблюдения;

- 2) выбор объекта статистического наблюдения и единиц его составляющих в зависимости от цели и задач;
- 3) составление организационного плана и программы статистического наблюдения;
- 4) подготовка специальных работников – регистраторов для непосредственного проведения статистического наблюдения;
- 5) подготовка специальных учетных регистров - формуляров, в которые будет заноситься статистическая информация;
- 6) подготовка технических средств для проведения статистического наблюдения.

Второй этап - непосредственный сбор статистической информации. Заключается в том, что подготовленные работники-регистраторы выходят на объект статистического наблюдения и регистрируют данные о его состоянии в специальных формулярах – учетных регистрах в соответствии с целью, задачами и программой статистического наблюдения.

Третий этап - контроль материалов статистического наблюдения. Заключается в проверке наличия ошибок статистического наблюдения путем контроля его материалов.

Контроль проводится двумя способами:

- 1) логический контроль – проверка соответствия значения показателя, полученного при статистическом наблюдении, его значению в действительности, либо сопоставление ответов на один вопрос формуляра с ответами на другие вопросы;
- 2) арифметический контроль – проверка счетом, если имеются взаимосвязанные показатели, которые можно проверить определенным математическим действием.

2.3 Основные организационные формы статистического наблюдения

Организационные формы статистического наблюдения выделяются на основе их общих организационных особенностей. В статистике выделяются две формы статистического наблюдения (рис. 10):



Рисунок 10 – Основные организационные формы статистического наблюдения

Статистическая отчетность – это, регламентируемая законодательством, система регулярного сбора статистической информации о деятельности хозяйствующих субъектов любой формы собственности и вида деятельности.

Статистическая отчетность является такой формой проведения статистического наблюдения, при которой статистическая информация поступает от самих единиц статистического наблюдения в статистическую службу в виде обязательных отчетов в установленные сроки и по утвержденным формам. При этом источником сведений обычно являются данные первичного и бухгалтерского учета.

Статистическая отчетность представляется организациями и иными лицами в соответствии со следующими нормативно-правовыми актами:

1. Федеральным Законом «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» от 29 ноября 2007 года №282-ФЗ ((в ред. от 23.07.2013) [2].
2. Постановлением Правительства РФ «О федеральной службе государственной статистики» от 2 июня 2008г. № 420 ((в ред. от 02.11.2013) [5].
3. Положением об условиях предоставления в обязательном порядке первичных статистических данных и административных данных субъектам официального статистического учета, утвержденным Постановлением Правительства РФ от 18.08.2008 № 620 [7].
4. Положением «О порядке представления статистической информации, необходимой для проведения государственных статистических наблюдений», утвержденным постановлением Госкомстата России от 15.07.2002г. № 154 (данный документ применяется в части, не противоречащей указанному выше Постановлению Правительства РФ) [8].
5. Постановлением Правительства РФ от 16.02.2008 № 79 «О порядке проведения выборочных статистических наблюдений за деятельностью субъектов малого и среднего предпринимательства» [9].

Статистическая отчетность является основной формой статистического наблюдения и обладает следующими свойствами:

- 1) обязательность предоставления – предприятия (организации) обязаны, в соответствии с указанными нормативно-правовыми актами, предоставлять статистическую отчетность в статистическую службу;
- 2) срочность предоставления – предприятия (организации) обязаны в соответствии с указанными нормативно-правовыми актами, предоставлять статистическую отчетность в статистическую службу в строго установленные сроки;
- 3) ответственность за достоверность информации – предприятие (организация) несет ответственность за достоверность данных, отраженных в

статистической отчетности в соответствии с указанными нормативно-правовыми актами.

На предприятиях и в организациях составлением статистической отчетности занимаются определенные работники, которые наряду с руководителем предприятия несут ответственность за обязательность предоставления статистической отчетности, за достоверность представляемой информации и за соблюдение сроков предоставления отчетности. При нарушении данных условий, предприятие (организация) несет административную ответственность в виде штрафов, в соответствии с указанными нормативно-правовыми актами.

Типовая – это статистическая отчетность, имеющая единую форму и содержание для всех хозяйствующих субъектов независимо от их вида деятельности и формы собственности (отчет по труду, отчет о наличии и движении основных фондов и др.).

Специализированная – это статистическая отчетность, отражающая специфические стороны хозяйствующего субъекта в зависимости от его вида деятельности (отчет об экспорте-импорте для предприятий внешнеэкономической деятельности, отчет о товарообороте для торговых предприятий, отчет о ходе сева для сельскохозяйственных предприятий и т.д.).

Годовая – это отчетность, содержащая обобщающие сведения в целом за год и предоставляемая в статистическую службу один раз в год не позднее 1 апреля следующего за отчетным года.

Квартальная – это отчетность, содержащая обобщающие сведения в целом за каждый квартал и предоставляемая в статистическую службу четыре раза в год не позднее 20 числа месяца, следующего за отчетным кварталом.

Месячная – это отчетность, содержащая сведения за каждый месяц и предоставляемая в статистическую службу ежемесячно не позднее 10 числа месяца, следующего за отчетным.

Формы статистической отчетности разрабатываются Росстатом и утверждаются руководителем Федеральной службы государственной

статистики РФ. В соответствии с Приказом от 29.08.2012г. №470 «Об утверждении статистического инструментария для организации федерального статистического наблюдения за деятельностью предприятий» [10] утверждены следующие основные формы статистической отчетности (рис. 11):

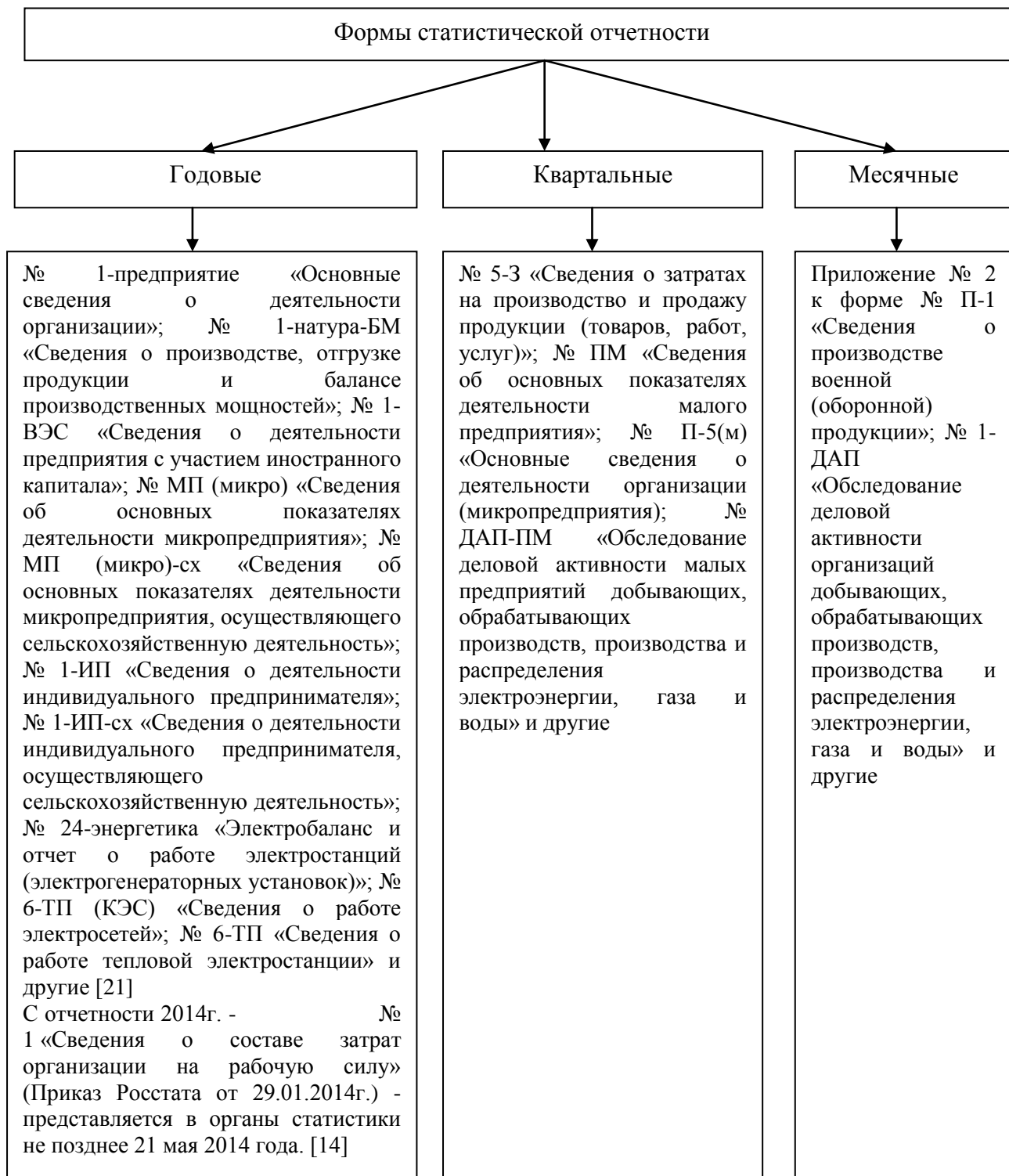


Рисунок 11 – Примеры форм статистической отчетности

Специально организованное – это статистическое наблюдение, которое организуется с определенной целью, проводится, как правило, прерывно через определенные промежутки времени или одновременно.

Цели специально организованного статистического наблюдения:

- 1) получение данных, которые не содержатся в статистической отчетности, так как статистическая отчетность составляется только предприятиями и организациями (данные о населении, о личных подсобных хозяйствах населения и другие);
- 2) проверка данных статистической отчетности (инвентаризация на складе при смене материально ответственного лица, выборочные налоговые проверки и другие).

Специально организованное наблюдение бывает двух видов:

- 1) обследование (социологическое обследование, обследование общественного мнения и так далее);
- 2) перепись (населения, жилого фонда, сельскохозяйственная перепись и другие).

2.4 Виды статистического наблюдения

Виды статистического наблюдения выделяются по трем классификационным признакам: 1) по степени охвата изучаемой совокупности; 2) по времени и систематичности проведения; 3) по источнику сведений (рис. 12).

Сплошное – это статистическое наблюдение, при котором обследованию подвергаются все без исключения единицы изучаемой совокупности;

Несплошное – это статистическое наблюдение, при котором обследованию подвергается не вся изучаемая совокупность, а ее часть, и по этой обследуемой части с определенным уровнем допущения делаются выводы обо всей совокупности.

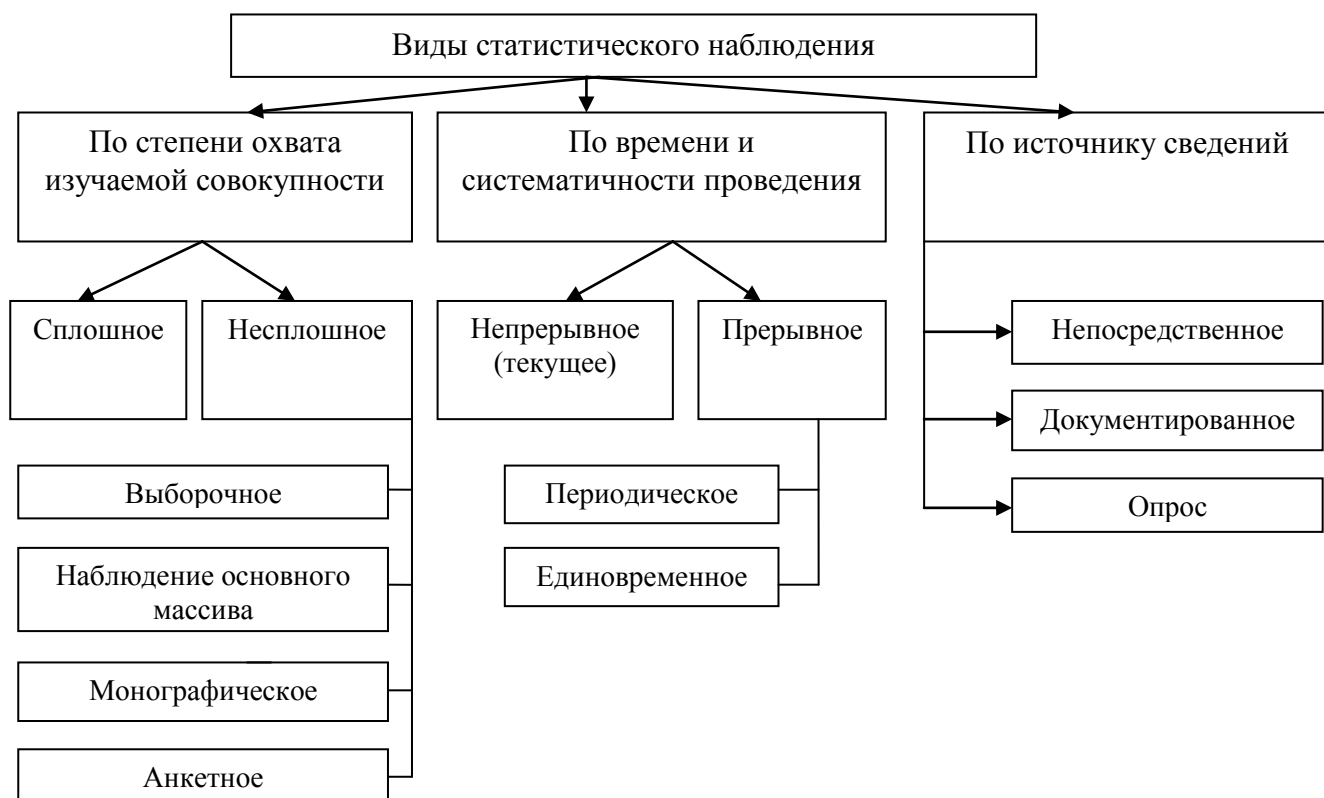


Рисунок 12 – Основные виды статистического наблюдения

Несплошное статистическое наблюдение по способу отбора единиц в выборочную совокупность делится на четыре вида:

- 1) выборочное – это несплошное статистическое наблюдение, при котором отбор части изучаемой совокупности производится в порядке случайного и беспристрастного для каждой единицы отбора отбор части изучаемой совокупности производится в порядке случайного и беспристрастного для каждой единицы отбора;
- 2) наблюдение основного массива – это несплошное статистическое наблюдение, при котором отбор части изучаемой совокупности производится на основе пристрастного отбора, то есть в выборочную совокупность отбираются наиболее крупные и показательные единицы, в которых сосредоточена основная часть всех подлежащих изучению фактов;
- 3) монографическое наблюдение – это несплошное статистическое наблюдение, при котором отбор в выборочную совокупность производится пристрастно, то есть в выборочную совокупность отбираются наиболее характерные единицы –

самые лучшие или самые худшие, которые подвергаются детальному статистическому описанию;

4) анкетное – это несплошное статистическое наблюдение, при котором заполнение и возвращение анкет является добровольным и, как правило, число заполненных и возвращенных анкет намного меньше разосланных.

Наиболее достоверным из перечисленных видов несплошного наблюдения является выборочный, так как он основан на случайном и беспристрастном отборе единиц.

Непрерывное (текущее) - это статистическое наблюдение, при котором сбор данных производится постоянно, а факты, подлежащие регистрации, фиксируются по мере их возникновения (факты рождения, смертности, брака в ЗАГСах и др.). Основная особенность непрерывного (текущего) наблюдения заключается в том, что нельзя допускать разрыва во времени между моментом возникновения факта и моментом его регистрации.

Прерывное – это статистическое наблюдение, при котором сбор данных производится с равными или неравными перерывами во времени, либо вообще единовременно.

Периодическое – это статистическое наблюдение, которое проводится регулярно через равные промежутки времени (перепись населения, статистическая отчетность и др.).

Единовременное – это статистическое наблюдение, которое проводится по мере необходимости, без соблюдения строгой периодичности во времени, либо вообще один раз. (инвентаризации на складе при смене материально ответственного лица, перепись жилого фонда и др.).

Непосредственное – это статистическое наблюдение, при котором факты, подлежащие регистрации, устанавливаются специальными работниками-регистраторами, которые непосредственно проводят статистическое наблюдение путем замера и подсчета и заносят полученные данные в специальные формуляры (перепись населения, сельскохозяйственная перепись и другие).

Документированное – это статистическое наблюдение, при котором источником сведений являются соответствующие документы – статистическая отчетность, данные первичного учета и др.

Опрос – это статистическое наблюдение, при котором источником сведений является опрашиваемое лицо, со слов которого полученная информация регистраторами заносится в специальные формуляры.

2.5 Способы проведения статистического наблюдения

Способы сбора статистической информации – проведения статистического наблюдения различаются по их организационным принципам (рис. 13).

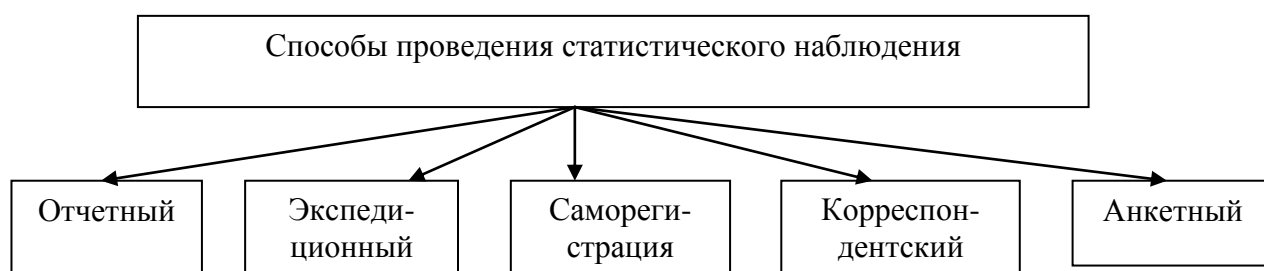


Рисунок 13 – Способы проведения статистического наблюдения

Отчетный - это способ, при котором сведения поступают в статистические органы в порядке обязательного и систематического представления хозяйствующими субъектами установленных для них форм статистической отчетности со сведениями об их деятельности.

Экспедиционный – это способ, при котором сведения собираются специально уполномоченными работниками-регистраторами, которые вступают в непосредственный контакт с объектом и единицей статистического наблюдения, заносят полученную статистическую информацию в специальные формуляры и доставляют заполненные формуляры в статистические органы (перепись населения, сельскохозяйственная перепись и другие).

Саморегистрация – это способ, при котором регистраторы вступают в контакт с объектом и единицей статистического наблюдения, разъясняя им инструкцию по заполнении формуляров, а формуляры заполняются самими опрашиваемыми лицами. Регистраторы собирают заполненные формуляры, проверяют их и предоставляют в статистические органы (изучение рейтингов телевизионных программ, маркетинговые исследования и другие). Саморегистрация может быть осуществлена явочным способом, когда опрашиваемые лица сами являются в статистические органы и сообщают необходимые сведения, которые регистрируются работниками органов статистики в специальные формуляры.

Корреспондентский – это способ, при котором информацию собирают, либо сообщают лица, добровольно взявшие на себя эту обязанность.

Анкетный – это способ, при котором опрашиваемые лица добровольно заполняют анкеты и возвращают их в статистические органы.

2.6 Методологические аспекты проведения статистического наблюдения

Методологические аспекты статистического наблюдения включают комплекс мероприятий по правильному и успешному его проведению. Методологические аспекты включают:

1. Объект статистического наблюдения.
2. Единица статистического наблюдения.
3. Программа статистического наблюдения.
4. Время и период статистического наблюдения.
5. Организационный план статистического наблюдения.

Объект статистического наблюдения – это совокупность массовых общественных явлений и процессов, сведения о которых необходимо собрать в процессе данного статистического наблюдения.

Для обеспечения успешности проведения статистического наблюдения выбор объекта должен соответствовать следующим требованиям: 1) соответствие цели и задачам статистического наблюдения; 2) указание существенных характерных признаков объекта, отличающих его от других объектов; 3) установление материальных, территориальных и временных границ объекта наблюдения; 4) обеспечение однородности исследуемой совокупности.

Для выполнения указанных требований устанавливается ценз – ограничительный признак, которому должны соответствовать все единицы обследуемой совокупности.

Единица статистического наблюдения – это составная часть, элемент объекта статистического наблюдения, который является первичным носителем признаков, подлежащих регистрации и основой ведущегося при обследовании учета. При определении единиц статистического наблюдения также используется ценз.

Программа статистического наблюдения – это перечень показателей, по которым необходимо собрать данные, либо перечень вопросов, на которые нужно получить ответы в процессе статистического наблюдения. При составлении программы статистического наблюдения необходимо соблюдать следующие основные требования:

- 1) вопросы должны быть сформулированы ясно и кратко;
- 2) в программу должны включаться только те вопросы, которые соответствуют цели и задачам статистического наблюдения;
- 3) в программу должны быть включены вопросы, на которые могут быть получены достоверные ответы;
- 4) в программу нельзя включать вопросы компрометирующего характера для опрашиваемого лица;
- 5) в программу должны быть включены вопросы контрольного характера;
- 6) в программу должны быть включены вопросы, обеспечивающие сопоставимость информации во времени и пространстве.

Время наблюдения – это время, к которому относятся данные собранной информации. Период наблюдения – это период между моментом начала и окончания проведения статистического наблюдения.

Выбор времени и периода статистического наблюдения зависит от следующих факторов: 1) особенности объекта наблюдения, его размеры и территориальная протяженность; 2) объем программы статистического наблюдения; 3) степень динамичности объекта наблюдения. Время наблюдения выбирается на основании критерия минимальной мобильности – это время, в которое объект наблюдения наименее подвержен перемещению в пространстве; 4) наличие достаточного количества специально подготовленных работников и уровень их квалификации; 5) наличие технических средств проведения статистического наблюдения; 6) установление критического момента статистического наблюдения (при некоторых динамичных объектах). Критический момент статистического наблюдения – это момент времени, по состоянию на который регистрируются сведения (как правило, при переписи населения устанавливается критический момент на 00 часов определенной даты. Это свидетельствует о том, что на данный момент времени сосчитаны и переписаны все лица, живущие в стране. Умершие в ночь после 00 часов – включены в численность населения, а родившиеся в эту же ночь после 00 часов – не включены).

Поскольку статистическое наблюдение - это научно организованный и планомерный процесс, то для его проведения составляется организационный план.

Организационный план статистического наблюдения это - основной официальный документ, в котором отражаются все организационные мероприятия, необходимые для проведения статистического наблюдения. В организационном плане отражается все статистическое наблюдение от его начала до конца. Организационный план включает 1) указание цели и задач статистического наблюдения; 2) определение объекта статистического наблюдения и характеристика единиц его составляющих; 3) указание

программы статистического наблюдения; 4) перечень подготовительных мероприятий к проведению статистического наблюдения; 5) указание органов статистики, проводящих наблюдение, и лиц, занимающихся проведением наблюдения; 6) указание времени и периода наблюдения – даты его начала и окончания и при необходимости критического момента статистического наблюдения; 7) указание порядка проведения статистического наблюдения, порядка приема и сдачи материалов наблюдения; 8) указание формы, способа и вида статистического наблюдения.

2.7 Ошибки статистического наблюдения и меры борьбы с ними

Основная задача статистического наблюдения – это обеспечение достоверности полученной статистической информации, то есть – обеспечение ее точности. Точность статистической информации – это уровень соответствия величины показателя в действительности его величине, полученной при статистическом наблюдении.

В ходе статистического наблюдения возникают погрешности, которые являются ошибками статистического наблюдения. Ошибка статистического наблюдения – это отклонение величины показателя, полученной при статистическом наблюдении, от величины данного показателя в действительности.

Ошибки статистического наблюдения происходят в результате разных причин и оказывают различное влияние на конечные результаты статистического наблюдения. В зависимости от этого ошибки наблюдения бывают двух видов: ошибки регистрации; ошибки репрезентативности (рис. 14).

Ошибки регистрации возникают в результате неправильного установления фактов в процессе статистического наблюдения или неправильной их записи.

Ошибки репрезентативности возникают только при несплошном статистическом наблюдении в результате того, что состав выборочной совокупности недостаточно полно соответствует составу основной изучаемой совокупности.



Рисунок 14 – Виды ошибок статистического наблюдения

Случайные ошибки регистрации возникают в результате неточного ответа опрашиваемого лица или неточной записи регистратора. Систематические ошибки регистрации возникают под действием отдельных постоянных причин (неточность измерительных приборов, округление чисел в одну и ту же сторону и других).

Непреднамеренные ошибки регистрации возникают в результате несовершенства методологического и организационного обеспечения статистического наблюдения (недостаточная подготовка кадров, недостаточное материально-техническое обеспечение, недостаточная трудовая и исполнительная дисциплина и другие). Преднамеренные ошибки регистрации

возникают в результате сознательного искажения фактов регистратором с целью получения выгоды.

Случайные ошибки репрезентативности возникают при несплошном наблюдении в результате неполного воспроизводства выборочной совокупности состава основной изучаемой совокупности. Эти ошибки неизбежны. Их невозможно устранить. Возможно только оценить их величину и сделать поправку на эту величину. Систематические ошибки репрезентативности возникают в результате нарушения принципов отбора единиц изучаемой совокупности. Размеры систематической ошибки репрезентативности не подлежат количественной оценке, но их можно устранить путем соблюдения правил отбора.

Отнесение ошибок статистического наблюдения к устраняемым, либо к не устраняемым зависит от момента их обнаружения. При наличии времени и условий для внесения исправлений в документацию ошибки могут относиться к устраняемым, в противном случае – к не устраняемым.

Меры борьбы с ошибками статистического наблюдения направлены на предупреждение или уменьшение размеров ошибок. Основные меры борьбы с ошибками статистического наблюдения:

- 1) включение в организационный план проведения статистического наблюдения ряда контрольных мероприятий;
- 2) осуществление контроля полноты проведения наблюдения – проверка того, насколько полно охвачен объект наблюдения и все ли единицы наблюдения включены в обследование;
- 3) осуществление контроля качества материалов статистического наблюдения путем логического и арифметического контроля.

3 ОБОБЩАЮЩИЕ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

3.1 Обобщающие статистические показатели, их свойства, виды и формы

Статистический показатель является одной из основных категорий статистической науки. Статистический показатель – это особый объективный признак состояния и развития массовых общественных явлений.

Поскольку массовые общественные явления, которые изучаются статистикой, многообразны, а также многообразны их свойства и признаки, постольку статистические показатели имеют множество форм и видов. Формы и виды статистических показателей подразделяются по трем классификационным признакам (рис. 15).



Рисунок 15 – Виды и формы объективных статистических показателей

В связи с многообразием статистических показателей и их сложным содержанием они обладают специфическими свойствами:

- 1) статистический показатель представляет собой единство качественной и количественной сторон массовых общественных явлений, так как он имеет определенное качественное содержание и соответствующую этому содержанию величину;
- 2) статистический показатель является общим, так как, как правило, относится не к единичному явлению, а охватывает всю совокупность массовых общественных явлений;
- 3) статистический показатель имеет временную и пространственную определенность, историческую конкретность, так как любое явление общественной жизни происходит в определенных границах места и времени;
- 4) статистический показатель является существенным, так как выражает основные характеристики состояния массовых общественных явлений и выявляет основные закономерности их развития;
- 5) статистический показатель является объективной стороной отражаемого массового общественного явления, так как не может существовать вне характеризуемого им явления;

На основании указанных свойств обобщающий статистический показатель в широком смысле – это признак, особое свойство массовых общественных явлений, их количественная характеристика в неразрывной связи с их качественной характеристикой в определенных условиях места и времени.

Все массовые общественные явления имеют качественную и количественную определенность. Статистика измеряет и выражает количественную определенность массовых общественных явлений при помощи особых категорий – статистических величин.

Количественная определенность массовых общественных явлений имеет два вида: 1) абсолютный размер явления; 2) относительный размер явления.

Абсолютный размер явления – это его величина, рассматриваемая независимо от размеров других явлений. Относительный размер явления – это соотношение величины данного явления с величиной какого-либо другого явления, либо с величиной этого же явления, но взятой за другое время или в другом пространстве.

Любой статистический показатель принимает одну из этих трех форм: 1) абсолютные величины; 2) относительные величины; 3) средние величины.

3.2 Абсолютные величины, их виды и единицы измерения

Абсолютная величина – это количественная характеристика размера массового общественного явления, независимо от размеров других явлений в определенных границах места и времени.

Абсолютная величина является первоначальным видом любого обобщающего статистического показателя, который получают в процессе статистического наблюдения и в процессе сводки и группировки статистических данных. Абсолютные величины подразделяются на виды по трем классификационным признакам (рис. 16).



Рисунок 16 – Виды абсолютных величин

Абсолютные величины всегда имеют определенную размерность и единицы измерения, то есть являются именованными. В зависимости от характера изучаемого явления, абсолютные величины имеют четыре вида единиц измерения: 1) натуральные; 2) условно-натуральные; 3) стоимостные (денежные); 4) трудовые.

Натуральные единицы измерения применяются для количественной оценки только одного элемента и выражают размер явления в физических мерах в соответствии с его физическими свойствами – веса, объема, протяженности и др. (площадь сельскохозяйственных угодий в гектарах, объем производства отдельного вида продукции в тоннах и т.д.).

Условно-натуральные единицы измерения используются для обобщения группы однородных элементов и выражают размер группы однородных явлений в единицах какого-либо эталона. Их получают путем пересчета натуральных единиц в единицы установленного эталона при помощи специальных коэффициентов (условный эталонный трактор, условное топливо, условные головы сельскохозяйственных животных и др.).

Стоимостные (денежные) единицы измерения выражают размер группы разнородных явлений в единицах универсального эталона – деньгах. Денежные единицы могут быть использованы для определения общего размера разнородных элементов (объем производства различных видов продукции, объем продаж различных видов товаров и др.).

Трудовые – специфические единицы измерения, которые выражают объем затрат труда на производство определенных видов продукции (в человеко-часах) и используются для определения уровня производительности труда работников и уровня трудоемкости производства продукции).

Анализ общественных явлений заключается в сравнении и сопоставлении статистических данных, поэтому абсолютных величин недостаточно для проведения статистического анализа. В целях анализа абсолютные величины дополняются относительными и средними величинами, которые являются производными от абсолютных величин.

3.3 Относительные величины, их виды и единицы измерения

Относительные величины выражают относительный размер общественных явлений, рассматриваемый в зависимости от размера других явлений, либо этого же явления, принятого в другом времени или в другом пространстве.

Следовательно, относительная величина – это количественная характеристика соотношения между двумя абсолютными величинами.

Формально относительную величину (ОВ) можно выразить:

$$ОВ = \frac{\text{Сравниваемый показатель}}{\text{База сравнения}}, \quad (1)$$

где сравниваемый показатель – абсолютная величина, характеризующая общественное явление, которое изучается;

база сравнения – абсолютная величина, характеризующая общественное явление, принимаемого за основание, с которым производится сравнение.

Единицы измерения относительных величин зависят от варианта их расчета и вида (рис. 17).

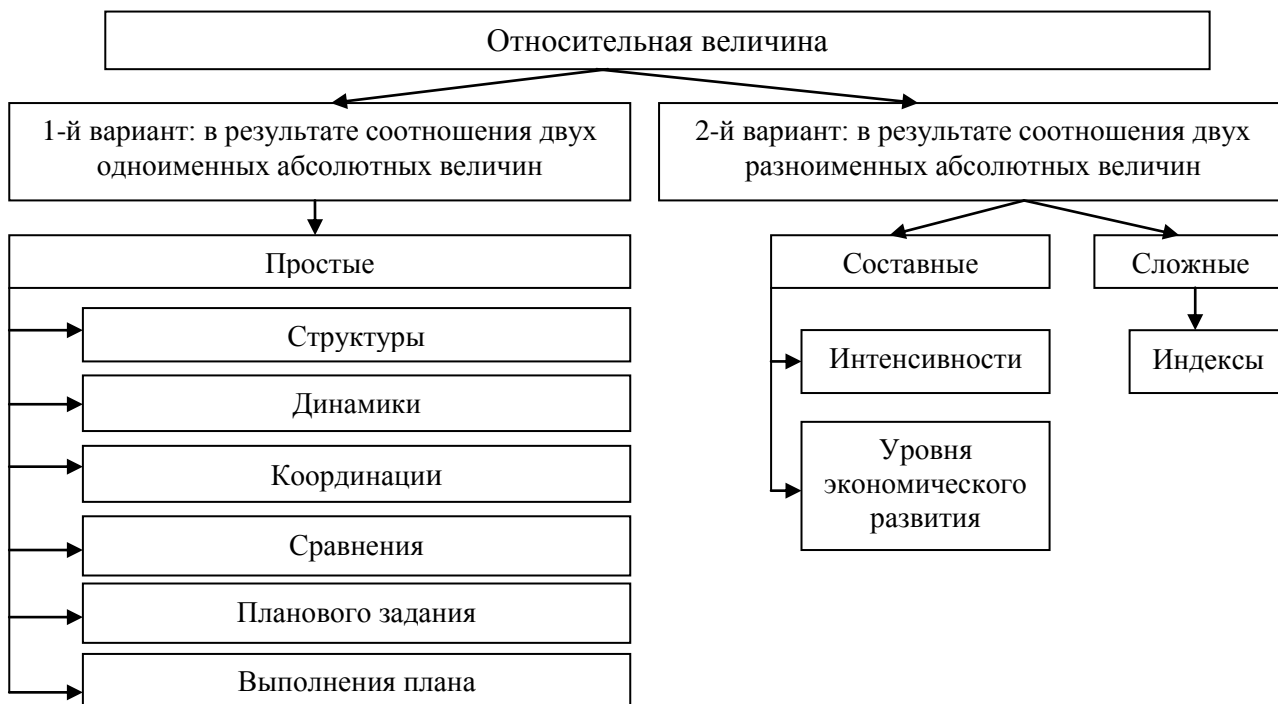


Рисунок 17 – Виды относительных величин

Единицы измерения простых относительных величин, полученных при первом варианте расчета зависят от размерности базы сравнения, который может приниматься за 1, 100, 1000 и т.д. Следовательно, единицами измерения простых относительных величин являются:

1) если значение базы сравнения принимается за единицу, то полученная относительная величина измеряется в коэффициентах (или размах) и показывает, во сколько раз сравниваемый показатель больше базы сравнения. Данная единица измерения применяется в том случае, когда сравниваемый показатель намного больше базы сравнения;

2) если значение базы сравнения принимается за 100, то полученная относительная величина измеряется в процентах и показывает, сколько процентов составляет сравниваемый показатель к базе сравнения. Данная единица измерения применяется в том случае, когда сравниваемый показатель ненамного больше или меньше базы сравнения;

3) если значение базы сравнения принимается за 1000, то полученная относительная величина измеряется в промилле и показывает, сколько единиц сравниваемого показателя приходится на 1000 единиц базы сравнения. Данная единица измерения применяется в том случае, когда сравниваемый показатель намного меньше базы сравнения (демографические коэффициенты – рождаемости, смертности, естественного прироста и др.);

4) при условии, что сравниваемый показатель намного меньше базы сравнения, ее значение может приниматься за 10000. Полученная относительная величина в данном случае измеряется в продецемилле и показывает, сколько единиц сравниваемого показателя приходится на 10000 единиц базы сравнения (социальные показатели условий жизни – число больничных коек на 10000 чел. населения, обеспеченность предметами длительного пользования на 10000 чел. населения и др.).

Единицы измерения составных относительных величин, полученных при втором варианте расчета образуются путем сочетания единиц измерения исходных абсолютных величин – сравниваемого показателя и базы сравнения

(урожайность сельскохозяйственных культур – в центнерах с гектара, средняя заработная плата работников – в рублях на человека и др.).

Относительная величина структуры характеризует состав изучаемого явления и показывает, какую долю или удельный вес занимают отдельные части явления в общем его объеме.

Формально расчет относительной величины структуры можно представить следующим образом:

$$ОВ_{структуры} = \frac{\text{Размер части явления}}{\text{Общий размер явления}} \times 100. \quad (2)$$

При исчислении относительной величины структуры, общий размер изучаемого явления принимается за базу сравнения – за 100% и отдельные части этого явления соотносятся с общим его размером (табл. 1).

Таблица 1

Расчет относительных величин структуры

Наименование культур	Посевная площадь, га		Относительная величина структуры в % к итогу		Изменение структуры (+;-), процентных пункта
	Базисный период	Отчетный период	Базисный период	Отчетный период	
Зерновые	2215	540	13,37	4,76	-8,61
Соя	6670	5411	40,26	47,72	+7,46
Многолетние травы	5922	4174	35,74	36,81	+1,07
Однолетние травы	1204	851	7,27	7,51	+0,24
Кукуруза	557	362	3,36	3,19	-0,17
Всего посевов	16568	11338	100,00	100,00	-

Из таблицы 1 видно, что наибольшую долю в составе посевных площадей занимает соя как в базисном, так и в отчетном периодах – соответственно 40,26 и 47,72% и многолетние травы – 35,74% в базисном периоде и 36,81% в отчетном периоде. Наименьшую долю в составе посевов занимают кукуруза на силос и зеленый корм – в базисном периоде 3,36% и в отчетном периоде 3,19%.

При сопоставлении структуры явления за несколько периодов определяются структурные сдвиги в его составе – изменение относительных

величин структуры в отчетном периоде по сравнению с базисным, в чем проявляется взаимосвязь относительных величин структуры и динамики.

Относительная величина динамики характеризует изменение изучаемого явления во времени, выявляет направление его развития и показывает, во сколько раз или на сколько процентов размер явления в изучаемом периоде больше либо меньше его размера в предыдущем периоде.

Формально расчет относительной величины динамики можно выразить следующим образом:

$$ОВ\ динамике = \frac{\text{Величина показателя в изучаемом (отчетном) периоде}}{\text{Величина показателя в предыдущем (базисном) периоде}} \times 100. \quad (3)$$

Относительные величины динамики выражаются в виде коэффициентов роста и темпов роста.

На основании данных таблицы 1 (размер посевных площадей в базисном периоде 16568 га; размер посевных площадей в отчетном периоде 11338 га) относительная величина динамики размера посевных площадей составит

$$ОВ_{динамики} = \frac{11338}{16568} \times 100 = 0,684 \times 100 = 68,4\%.$$

Расчет показал, что в отчетном периоде по сравнению с базисным периодом размер посевных площадей уменьшился на 31,6%.

Относительная величина сравнения (наглядности) характеризует соотношение двух одноименных абсолютных величин, которые относятся к одному периоду, но к разным объектам или территориям и показывает, на сколько процентов или во сколько раз показатель одного объекта больше либо меньше этого же показателя, относящегося к другому объекту.

Формально расчет относительной величины сравнения можно выразить следующим образом:

$$ОВ\ сравнения = \frac{\text{Величина показателя изучаемого объекта}}{\text{Величина показателя объекта, с которым сравнивают}} \times 100 \quad (4)$$

На основании данных таблицы 1 (размер посевных площадей в отчетном периоде первого хозяйства 11338 га; размер посевных площадей в отчетном

периоде второго хозяйства 15500 га) относительная величина сравнения составит

$$ОВ\ сравнения = \frac{11338}{15500} \times 100 = 0,731 \times 100 = 73,1\% .$$

Расчет показал, что размер посевных площадей в изучаемом сельскохозяйственном предприятии на 26,9% меньше размера посевных площадей предприятия, принятого за базу сравнения.

Относительная величина координации характеризует соотношение между отдельными составными частями изучаемого явления и показывает, на сколько процентов или во сколько раз все части изучаемого явления больше либо меньше одной его части, принятой за базу сравнения.

Относительная величина координации представляет собой разновидность относительных величин структуры и сравнения. Формально расчет относительной величины координации можно выразить следующим образом:

$$ОВ\ координации = \frac{\text{Величина отдельных частей явления}}{\text{Величина части явления, принятая за базу сравнения}} \times 100 \quad (5)$$

На основании данных таблицы 1 относительные величины координации составят (табл. 2):

Таблица 2

Расчет относительных величин координации (база сравнения – посевная площадь сои)

Наименование культур	Посевная площадь, га		Относительная величина координации, %	
	Базисный период	Отчетный период	Базисный период	Отчетный период
Зерновые	2215	540	33,2	10,0
Соя	6670	5411	-	-
Многолетние травы	5922	4174	88,8	77,1
Однолетние травы	1204	851	18,0	15,7
Кукуруза	557	362	8,3	6,7
Всего посевов	16568	11338	-	-

Из таблицы 2 видно, что в базисном периоде на 100 га посевов сои приходится 33,2 га посевов зерновых культур, 88,8 га посевов многолетних трав, 18га посевов однолетних трав и 8,3 га посевов кукурузы. В отчетном

периоде на 100 га посевов сои приходится 10 га посевов зерновых культур, 77,1 га посевов многолетних трав, 15,7 га посевов однолетних трав и 6,7 га посевов кукурузы.

Относительная величина планового задания характеризует степень увеличения или уменьшения изучаемого показателя по плану на следующий период по сравнению с фактическим размером данного показателя в отчетном периоде. Показывает, на сколько процентов или во сколько раз размер изучаемого явления, предусмотренный планом на будущий период, больше или меньше фактического размера данного показателя в отчетном периоде.

Формально, расчет относительной величины планового задания можно представить следующим образом:

$$ОВ \text{ планового задания} = \frac{\text{Размер явления, предусмотренный планом на будущий период}}{\text{Фактический размер явления отчетного периода}} \times 100 \quad (6)$$

На основании данных таблицы 1 (фактический размер посевных площадей в отчетном периоде 11338 га; запланированный на будущий период размер посевных площадей 12500 га), относительная величина планового задания составит

$$ОВ \text{ планового задания} = \frac{12500}{11338} \times 100 = 110,2\%$$

Расчет показал, что планом предусмотрено увеличить фактический размер посевных площадей сельскохозяйственных культур в будущем периоде на 10,2%.

Относительная величина выполнения плана (договора) характеризует степень выполнения планового задания или договорных обязательств и показывает, на сколько процентов или во сколько раз фактический размер изучаемого явления больше или меньше предусмотренного планом или договором.

Формально, расчет относительной величины выполнения плана или договора можно представить следующим образом:

$$ОВ \text{ выполнения плана (договора)} = \frac{\text{Фактический размер явления}}{\text{Размер явления, предусмотренный планом (договором)}} \times 100 \quad (7)$$

На основании данных таблицы 1 (фактический размер посевных площадей в отчетном периоде 11338 га; размер посевных площадей по плану на отчетный период составлял 10550 га), относительная величина выполнения плана составит

$$ОВ \text{ выполнения плана (договора)} = \frac{11338}{10550} \times 100 = 107,5\%$$

Расчет показал, что в отчетном периоде фактический размер посевных площадей сельскохозяйственных культур был выше планового на 7,5%.

Между относительными величинами динамики, планового задания и выполнения плана существует следующая зависимость

$$ОВ \text{ динамики} = ОВ \text{ планового задания} \times ОВ \text{ выполнения плана (договора)} \quad (8)$$

Относительная величина интенсивности характеризует степень распространения или развития изучаемого явления в определенной среде и показывает, сколько единиц изучаемого явления приходится на единицу среды его распространения.

Формально расчет относительной величины интенсивности можно представить следующим образом:

$$ОВ \text{ интенсивности} = \frac{\text{Размер изучаемого явления}}{\text{Размер среды распространения изучаемого явления}} \quad (9)$$

Относительная величина интенсивности может рассчитываться как с умножением на 100, так и без умножения на 100.

Пример расчета относительных величин интенсивности представлен в таблице 3.

Расчет относительных величин интенсивности

Отрасль	Среднегодовая численность работников, человек		Стоимость валовой продукции, тыс. рублей		Производительность труда, тыс.рублей на человека (относительная величина интенсивности)	
	Базисный период	Отчетный период	Базисный период	Отчетный период	Базисный период	Отчетный период
Растениеводство	91	86	68126	79994	748,6	930,2
Животноводство	463	429	47848	68126	103,3	158,8
В целом по предприятию	554	515	115974	148120	209,3	287,6

Из таблицы 3 видно, что в базисном периоде на одного работника предприятия приходилось в отрасли растениеводства 748,6 тыс.р. валовой продукции, в отрасли животноводства – 103,3 тыс.р. валовой продукции и в целом по предприятию – 209,3 тыс.р. валовой продукции. В отчетном периоде на одного работника предприятия приходилось в отрасли растениеводства 930,2 тыс.р. валовой продукции, в отрасли животноводства – 158,8 тыс.р. валовой продукции и в целом по предприятия – 287,6 тыс.р. валовой продукции.

Относительная величина уровня экономического развития характеризует размер производства или потребления отдельных продуктов и товаров на душу населения. Формально, расчет относительной величины уровня экономического развития можно представить следующим образом:

$$ОВ \text{ уровня экономического развития} = \frac{\text{Размер производства (потребления) отдельных продуктов и товаров}}{\text{Численность населения региона или страны}} \quad (10)$$

Сложная относительная величина характеризует изменение во времени и пространстве сложных социально-экономических явлений.

К сложным относительным величинам в статистике относятся индексы. Индекс – это обобщающий относительный показатель сравнения двух сложных совокупностей, отдельные элементы которых не поддаются непосредственному суммированию.

4 СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАЦИИ

4.1 Сущность, значение и свойства средней величины

На второй стадии статистического исследования производится обобщение единичных массовых данных, полученных при статистическом наблюдении путем расчета относительных и средних величин. Следовательно, статистическая средняя величина – это обобщающая характеристика совокупности однотипных массовых явлений по количественному варьирующему признаку.

Значение средней величины можно определить следующим.

1. Статистическая средняя является одним из самых распространенных приемов обобщения, который используется при переходе от характеристик отдельных единиц совокупности к характеристике групп и совокупности в целом.

2. Статистическая средняя (согласно диалектическим понятиям «случайное и необходимое» и «единичное и общее») через единичное и случайное позволяет выявить общую закономерность развития массовых общественных явлений.

3. В статистической средней выражается действие общих условий развития массовых общественных явлений.

4. Только с помощью статистической средней величины можно охарактеризовать всю совокупность в целом по определенному количественному признаку.

В связи с большим значением средней величины в статистических исследованиях она обладает следующими особенностями:

1. Статистическая средняя величина рассчитывается на основе массовых данных, полученных при статистическом наблюдении.

2. Статистическая средняя величина может быть достоверной, только если она рассчитывается для качественно однородной совокупности.

3. При помощи статистической средней величины происходит сглаживание различий в величине признака у отдельных единиц совокупности, которые возникают под действием случайных факторов.

4. Статистическая средняя – это величина абстрактная, потому что характеризует не конкретную единицу совокупности, а совокупность в целом и абстрагируется от разнообразия признака у отдельных единиц.

5. В статистической средней величине отражается характерный, типичный уровень изучаемого явления, его объективное свойство.

6. Статистическая средняя величина является результатом совокупного действия всех единиц и проявляется в массе повторяющихся случайностей.

7. Статистическая средняя величина является отражением значений изучаемого признака и, следовательно, измеряется в тех же единицах, что и сам признак.

8. Статистическая средняя величина может быть рассчитана только по количественному признаку.

9. Статистическая средняя величина может быть рассчитана только по варьирующему признаку и обобщает количественную вариацию признака.

10. Статистическая средняя величина непосредственно связана с проявлением Закона больших чисел, которое заключается в том, что в изменении средней величины проявляется общая основная тенденция, под влиянием которой складывается процесс развития явлений в целом.

Статистические средние величины делятся на две группы:

- 1) степенные средние;
- 2) структурные средние.

4.2 Степенные средние величины

Степенная средняя величина характеризует типичную величину изучаемого явления в статистической совокупности и изменяется в зависимости от изменения показателя степени.

Для того, чтобы вычислить статистическую степенную среднюю величину, сначала необходимо построить ее логическую формулу: 1) определить сущность и содержание изучаемого явления; 2) определить сущность и содержание средней величины в данном конкретном случае; 3) определить исходные показатели, результатом соотношения которых является средняя величина в данном конкретном случае; 4) оценить предполагаемый результат расчета средней величины.

В статистике для расчета средних величин используются следующие понятия и условные обозначения:

- 1) признак, по которому рассчитывается средняя величина, называется осредняемым признаком. Средняя величина обозначается \bar{X} ;
- 2) величина осредняемого признака у каждой единицы совокупности называется индивидуальным значением признака или вариантой. Варианты обозначаются: $X_1, X_2, X_3, \dots, X_n$;
- 3) количество единиц совокупности называется ее численностью или объемом. Объем совокупности обозначается n (или $\sum f$);
- 4) показатель степени для степенной средней обозначается m (или k);
- 5) число повторений значений осредняемого признака (варианты) в изучаемой совокупности называется частотой или статистическим весом. Частота (статистический вес) обозначается f .

В зависимости от числа повторений одного и того же значения осредняемого признака (варианты) в данной совокупности (наличия, либо отсутствия частоты) степенные средние принимают две формы:

- 1) простая - применяется в том случае, когда совокупность представлена абсолютными величинами или когда каждое значение осредняемого признака

встречается в данной совокупности один раз (в формуле отсутствует частота или статический вес);

2) взвешенная - применяется в том случае, когда совокупность представлена относительными величинами или когда каждое значение осредняемого признака встречается в данной совокупности несколько раз (в формуле присутствует частота или статический вес).

В зависимости от изменения показателя степени и простая и взвешенная статистические средние величины принимают четыре основных вида (табл. 4):

1) средняя арифметическая; 2) средняя гармоническая; 3) средняя геометрическая; 4) средняя квадратическая.

Таблица 4

Формы и виды степенных средних величин

Название средней и показатель степени	Средняя величина		Значение
	Простая	Взвешенная	
1	2	3	4
Общая m	$\bar{X} = \sqrt[m]{\frac{\sum X^m}{n}}$	$\bar{X} = \sqrt[m]{\frac{\sum X^m \times f}{\sum f}}$	Является исходной для расчета всех видов степенных средних
Средняя арифметическая $m = 1$	$\bar{X} = \frac{\sum X}{n}$	$\bar{X} = \frac{\sum X \times f}{\sum f}$	Является наиболее распространенным и часто используемым видом средней величины. Используется в рядах распределения, состоящих из абсолютных величин, и в тех случаях, когда объем осредняемого признака равен сумме его значений
Средняя гармоническая $m = -1$	$\bar{X} = \frac{n}{\sum \frac{1}{X}}$	$\bar{X} = \frac{\sum f}{\sum \frac{f}{X}}$	Является величиной, обратной средней арифметической. Используется в рядах распределения, в которых значение осредняемого признака и частота скрыты в произведении, и в тех случаях, когда варианта не умножается, а делится на частоту

Продолжение таблицы 4

1	2	3	4
Средняя геометрическая $m = 0$	$\bar{X} = \sqrt[n]{X_1 \times X_2 \times X_3 \dots \times X_n}$	$\bar{X} = \sqrt[\sum f]{X_1^{f_1} \times X_2^{f_2} \times X_3^{f_3} \dots X_n^{f_n}}$	Используется в рядах распределения, состоящих из относительных величин и представленных в виде геометрической прогрессии. По средней геометрической определяются средние показатели динамики.
Средняя квадратическая $m = 2$	$\bar{X} = \sqrt{\frac{\sum X^2}{n}}$	$\bar{X} = \sqrt{\frac{\sum X^2 \times f}{\sum f}}$	Используется в рядах распределения, представленных абсолютными отклонениями индивидуальных значений признака от средней величины. По средней квадратической определяются показатели вариации

Рассмотренные степенные средние подчиняются правилу мажорантности, которое заключается в том, что чем больше показатель степени, тем больше само значение средней величины, рассчитанной по одним и тем же исходным данным. Формально, правило мажорантности можно выразить следующим образом:

$$\bar{X}_{\text{гарм}} \leq \bar{X}_{\text{геом}} \leq \bar{X}_{\text{арифм}} \leq \bar{X}_{\text{квадр}} \quad (9)$$

Таблица 5

Расчет различных видов средних величин

Номер единицы	Значение признака, X	Квадрат признака, X ²	Обратное значение признака, 1/X	Произведение значений признака,
1	2	4	0,5	2 × 3 × 5
2	3	9	0,33	-
3	5	25	0,2	-
Σ	10	38	1,03	30

Степенные средние:

- средняя гармоническая $\bar{X}_{\text{гарм}} = \frac{n}{\sum \frac{1}{X}} = \frac{3}{1,03} = 2,9;$

- средняя геометрическая $\bar{X}_{geom} = \sqrt[n]{X_1 \times X_2 \times X_3 \dots \times X_n} = \sqrt[3]{30} = 3,2;$

- средняя арифметическая $\bar{X}_{арифм} = \frac{\Sigma X}{n} = \frac{10}{3} = 3,3;$

- средняя квадратическая $\bar{X}_{квадр} = \sqrt{\frac{\Sigma X^2}{n}} = \sqrt{\frac{38}{3}} = 3,6.$

Степенные средние имеют ряд свойств. Поскольку наиболее часто используется средняя арифметическая, то рассмотрим свойства средней арифметической взвешенной:

1) произведение средней на сумму частот равно сумме произведений вариантов на частоты

$$\bar{X} \times \Sigma f = \Sigma X \times f; \quad (10)$$

2) если от каждой варианты отнять или к каждой варианте прибавить произвольное число, то полученная средняя уменьшится или увеличится на это же число

$$\frac{\Sigma (X + -a) \times f}{\Sigma f} = \bar{X} + -a; \quad (11)$$

3) если каждую варианту разделить на произвольное число, то полученная средняя уменьшится во столько же раз

$$\frac{\Sigma \frac{X}{a} \times f}{\Sigma f} = \frac{\bar{X}}{a}; \quad (12)$$

4) если каждую варианту умножить на произвольное число, то полученная средняя увеличится во столько же раз

$$\frac{\Sigma (X \times a) \times f}{\Sigma f} = \bar{X} \times a; \quad (13)$$

5) если все частоты разделить или умножить на произвольное число, то полученная средняя от этого не изменится

$$\frac{\Sigma X \times (f \times a)}{\Sigma (f \times a)} = \bar{X}; \quad (14)$$

6) сумма отклонений вариантов от средней величины всегда равна 0

$$\Sigma (X - \bar{X}) = 0. \quad (15)$$

4.3 Структурные средние величины

Структурные средние характеризуют состав изучаемой совокупности.

К структурным средним относятся: 1) мода (M_o); 2) медиана (M_e); 3) дециль (d); 4) квартиль (Q); 5) перцентиль (P).

Мода в статистике это - значение признака, которое соответствует максимальной точке теоретической кривой распределения. То есть мода – это значение варианты, которое чаще всего встречается в данной совокупности, или значение варианты, которому соответствует наибольшая частота. Мода показывает, какое значение признака имеет наибольшее число единиц данной совокупности.

Мода определяется в двух вариантах в зависимости от вида ряда распределения, по которому она рассчитывается: 1) в дискретном ряду распределения; 2) в интервальном ряду распределения.

В дискретном ряду распределения мода определяется визуально по наибольшей частоте (то есть визуально отыскивается варианта, которой соответствует наибольшая частота, это значение признака и является модой). Если в данном ряду распределения несколько вариантов имеющих одинаково наибольшую частоту, то такой ряд распределения имеет несколько значений моды и называется двумодальным, трехмодальным и т.д.

В интервальном ряду распределения определение моды производится в два этапа. На первом этапе определяется модальный интервал - интервал, в котором заключена мода. Модальный интервал определяется визуально путем определения интервала, которому соответствует наибольшая частота. На втором этапе, после определения модального интервала, значение моды определяется по формуле

$$M_o = X_0 + i \times \frac{f_{M_o} - f_{M_{o-1}}}{(f_{M_o} - f_{M_{o-1}}) + (f_{M_o} - f_{M_{o+1}})}, \quad (16)$$

где M_o – значение моды;

X_0 – нижняя граница модального интервала;

i – величина модального интервала;

f_{M_0} – частота модального интервала;

f_{M_0-1} – частота интервала, предшествующего модальному;

f_{M_0+1} – частота интервала, следующего за модальным.

Медиана в статистике это - значение признака у варианты, которая расположена в середине ранжированного ряда распределения. То есть, медиана – это варианта, которая делит ранжированный ряд распределения на две равные части.

Медиана определяется в двух вариантах в зависимости от вида ряда распределения, по которому она рассчитывается: 1) в дискретном ряду распределения; 2) в интервальном ряду распределения.

В дискретном ряду распределения для определения значения медианы находится номер единицы, расположенной в середине ранжированного ряда,

$$№(M_e) = \frac{n+1}{2}, \quad (17)$$

где $№(M_e)$ - номер единицы, соответствующей значению медианы.

n – численность единиц совокупности.

В интервальном ряду распределения медиана определяется в два этапа. На первом этапе определяется медианный интервал – интервал, в котором заключена медиана. Медианный интервал определяется по сумме накопленных частот. Накопленные частоты рассчитываются путем последовательного суммирования всех частот. Медианным является тот интервал, сумма накопленных частот которого первая превышает половину объема совокупности. На втором этапе, после определения медианного интервала, медиана определяется по формуле

$$M_e = X_0 + i \times \frac{1/2 \sum f - S_{M_e}}{f_{M_e}}, \quad (18)$$

где M_e – медиана;

X_0 – нижняя граница медианного интервала;

i – величина медианного интервала;

Σf - сумма частот (объем совокупности);

S_{Me} – сумма накопленных частот до медианного интервала;

f_{Me} – частота медианного интервала.

Медиана показывает, какое предельное значение имеет признак у половины (50%) единиц изучаемой совокупности.

Значения моды и медианы количественно сходны со степенной средней, рассчитанной по одному и тому же ряду распределения. При нормальном распределении все эти три характеристики равны.

Дециль в статистике – это значение варианты, которая делит ранжированный ряд распределения на десять равных частей. Первый дециль (d_1) делит совокупность в соотношении 1/10 и 9/10, второй дециль (d_2) делит совокупность в соотношении 2/10 и 8/10 и т.д. Определяется дециль по той же схеме, что и медиана, по формуле

$$d_1 = X_0 + i \times \frac{1/10 \Sigma f - S_{d_1}}{f_{d_1}}, \quad (19)$$

где d – дециль;

X_0 – нижняя граница децильного интервала;

i – величина децильного интервала;

Σf - сумма частот (объем совокупности);

S_{d1} – сумма накопленных частот до децильного интервала;

f_{d1} – частота децильного интервала.

Дециль показывает, какое предельное значение имеет признак у 10% единиц изучаемой совокупности.

Квартиль в статистике - это значение признака у варианты, которая делит ранжированный ряд на четыре равные части. Различаются верхний квартиль, который показывает 1/4 часть совокупности с наибольшим значением изучаемого признака, и нижний квартиль, который показывает 1/4 часть совокупности с наименьшим значением признака. Определяется квартиль по той же схеме, что и медиана и дециль по формуле

$$Q_1 = X_0 + i \times \frac{1/4 \Sigma f - S_{Q_1}}{f_{Q_1}}, \quad (20)$$

где Q – квартиль;

X_0 – нижняя граница квартильного интервала;

i – величина квартильного интервала;

Σf – сумма частот (объем совокупности);

S_{Q_1} – сумма накопленных частот до квартильного интервала;

f_{Q_1} – частота квартильного интервала.

Квартиль показывает, какое предельное значение имеет признак у 25% единиц изучаемой совокупности.

Перцентиль в статистике – это значение признака у варианты, которая делит ранжированный ряд распределения на сто равных частей. Определяется перцентиль по той же схеме, что и медиана, и дециль, и квартиль по формуле:

$$p_1 = X_0 + i \times \frac{1/100 \Sigma f - S_{p_1}}{f_{p_1}}, \quad (21)$$

где p – перцентиль;

X_0 – нижняя граница перцентильного интервала;

i – величина перцентильного интервала;

Σf – сумма частот (объем совокупности);

S_{p_1} – сумма накопленных частот до перцентильного интервала;

f_{p_1} – частота перцентильного интервала.

Перцентиль показывает, какое предельное значение имеет признак у 1% единиц изучаемой совокупности.

4.4 Понятие вариации признаков. Показатели вариации

Вариация – это различие индивидуальных значений варьирующего признака внутри изучаемой совокупности. Вариация возникает в результате того, что индивидуальные значения признака складываются под совокупным

влиянием различных факторов, которые по-разному сочетаются в каждом отдельном единичном случае.

Индивидуальные значения признака складываются под влиянием двух групп факторов: 1) систематические факторы; 2) случайные (остаточные) факторы. Влияние систематических факторов формируют статистическую закономерность и находят отражение в значении средней величины. Влияние случайных (остаточных) факторов находят отражение в отклонениях значений отдельных вариантов от значения средней величины и формирует вариацию признака.

Средняя величина дает обобщающую характеристику всей совокупности, но не характеризует, насколько индивидуальные значения признака сосредоточены близко к средней, либо отклоняются от нее. В одних случаях индивидуальные значения признака незначительно отличаются от значения средней величины. В таком случае признак считается слабо варьирующим, а средняя - типичной и достоверной характеристикой совокупности. В других случаях отдельные значения признака имеют значительные отличия (отклонения) от средней величины. В таком случае признак считается сильно варьирующим, а средняя величина - нетипичной и недостоверной характеристикой совокупности.

Таблица 6

Выработка рабочих двух бригад за неделю

№ бригады	Количество произведенных изделий по дням недели, единиц							Средняя выработка
	понедельник	вторник	среда	четверг	пятница	суббота	воскресенье	
1	4	4	5	5	5	6	6	$\bar{X} = 5$
2	1	2	7	2	2	10	11	$\bar{X} = 5$

В результате расчета (табл. 6) видно, что средние выработки двух бригад одинаковы. Но первая бригада работала равномерно в течение недели, а вторая бригада работала скачкообразно. Поэтому, в первом случае средняя является типичной и достоверной характеристикой среднечасовой выработки, а во

втором случае – нетипичной и недостоверной, поскольку в первом случае признак варьирует слабо, а во втором – сильно.

Поэтому для оценки достоверности средней величины, однородности изучаемой совокупности и степени варьирования признака определяются показатели вариации.

Показатели вариации характеризуют уровень отклонений индивидуальных значений признака от средней величины, оценивают силу варьирования изучаемого признака, степень однородности изучаемой совокупности, степень типичности и достоверности средней величины.

Показатели вариации бывают двух видов: абсолютные и относительные (рис.18) .

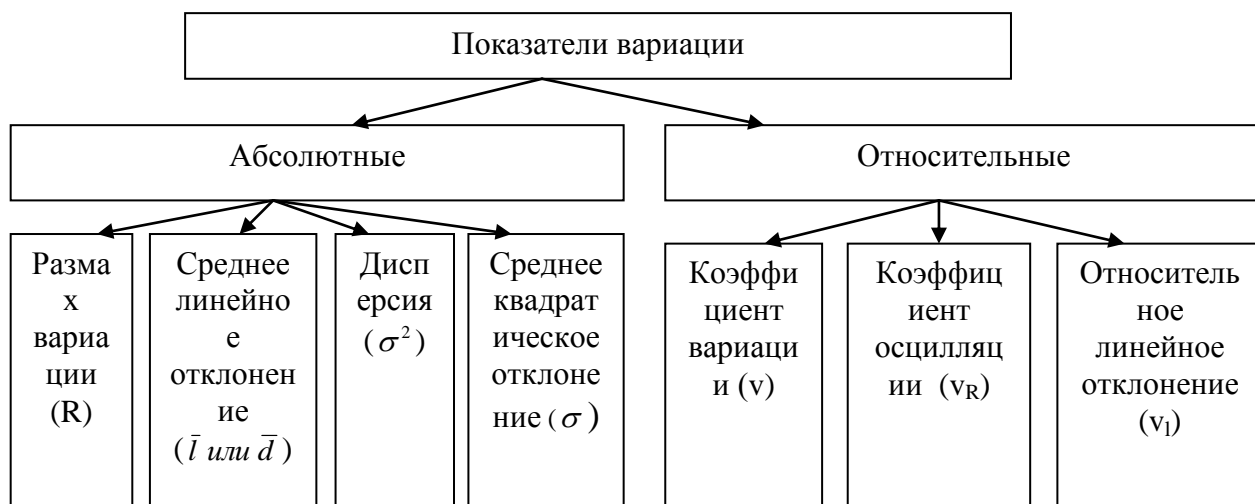


Рисунок 18 – Основные виды показателей вариации

Размах вариации (R) – характеризует разницу между наибольшим и наименьшим значением признака в изучаемой совокупности. Определяется по формуле

$$R = X_{\max} - X_{\min}, \quad (22)$$

где X_{\max} – максимальное значение изучаемого признака;

X_{\min} – минимальное значение изучаемого признака.

Размах вариации показывает, на сколько единиц отклоняется максимальное значение признака от минимального его значения в данной совокупности.

Применяется данный показатель редко - в основном в контроле качества продукции. Недостатком данного показателя является то, что при его расчете учитываются только крайние значения изучаемого признака (границы вариации) и не учитываются отклонения каждой варианты внутри совокупности.

Среднее линейное отклонение (l или d) характеризует отклонения индивидуальных значений признака от средней величины первой степени, то есть без учета знака отклонений. Определяется в зависимости от того, простая или взвешенная средняя положена в основу расчета, по следующим формулам:

- простое:
$$\bar{l} = \frac{\sum |X - \bar{X}|}{n}; \quad (23)$$

- взвешенное:
$$\bar{l} = \frac{\sum |X - \bar{X}| \times f}{\sum f}. \quad (24)$$

Среднее линейное отклонение показывает, на сколько единиц в среднем отклоняются индивидуальные значения признака от средней величины без учета знака отклонений. Поскольку данный показатель не учитывает знак отклонений, то он является недостаточно аналитичным и используется редко.

Дисперсия характеризует отклонения индивидуальных значений признака от средней величины второй степени. То есть при расчете дисперсии отклонения вариантов от средней величины возводятся во вторую степень (в квадрат), чем достигается положительное значение всех отклонений и учитывается их знак. Определяется в зависимости от того, простая или взвешенная средняя положена в основу расчета, по следующим формулам:

- простая:
$$\sigma^2 = \frac{\sum (X - \bar{X})^2}{n}; \quad (25)$$

- взвешенная:
$$\sigma^2 = \frac{\sum (X - \bar{X})^2 \times f}{\sum f}. \quad (26)$$

Имеется упрощенная формула расчета дисперсии:

- простая:
$$\sigma^2 = \frac{\sum X^2}{n} - (\bar{X})^2; \quad (27)$$

- взвешенная:
$$\sigma^2 = \frac{\sum X^2 \times f}{\sum f} - (\bar{X})^2. \quad (28)$$

Дисперсия показывает квадрат отклонений, поэтому смысловой нагрузки не несет, а является промежуточным показателем вариации.

Расчет дисперсии может быть упрощен. При равных интервалах в вариационном ряду распределения используется способ отсчета от условного нуля (способ моментов). Для использования данного способа необходимо знать следующие свойства дисперсии:

1) дисперсия постоянной величины равна нулю

$$\sigma^2(a \rightarrow const) = 0; \quad (29)$$

2) при уменьшении всех значений признака на одну и ту же величину A , значение дисперсии не изменяется

$$\sigma^2(X - A) = \sigma^2(X); \quad (30)$$

3) при уменьшении всех значений признака в A раз, значение дисперсии уменьшится в A^2 раз

$$\sigma^2(X / A) = \sigma^2(X) / A^2; \quad (31)$$

4) при исчислении среднего квадрата отклонений от любой величины A , отличающейся от средней величины, данный квадрат отклонений всегда будет больше среднего квадрата отклонений, исчисленного от средней. То есть, дисперсия от средней всегда меньше дисперсий, исчисленных от любых других величин, имеет свойство минимальности

$$\sigma^2(A) \leq \sigma^2(X^2). \quad (32)$$

Дисперсия определяется также для характеристики вариации альтернативного признака при условии, что альтернативный признак принимает всего два значения: ноль и единица с весами соответственно q и p . При этом значение признака принимается равным нулю у тех единиц, у

которых он отсутствует, и равным единице у тех единиц, у которых он присутствует.

Среднее значение альтернативного признака определяется по формуле

$$\bar{X} = \frac{1 \times p + 0 \times q}{p + q} = p, \quad (33)$$

где p – доля единиц совокупности, обладающих изучаемым признаком;

q – доля единиц совокупности, не обладающих изучаемым признаком.

Дисперсия альтернативного признака определяется по формуле

$$\sigma^2(p) = p \times q, \quad (34)$$

Предельное значение дисперсии альтернативного признака равно 0,25 при $p = 0,5$.

Среднее квадратическое отклонение характеризует степень отклонения индивидуальных значений признака от средней величины с учетом знака отклонений. Определяется как квадратный корень из дисперсии по формуле

$$\sigma = \sqrt{\sigma^2}. \quad (35)$$

Среднее квадратическое отклонение является наиболее аналитичным из абсолютных показателей вариации, измеряется в тех же единицах, что и осредняемый признак. Показывает, на сколько единиц в среднем отклоняется каждое индивидуальное значение изучаемого признака от средней величины в ту или другую сторону.

Между средним линейным и средним квадратическим отклонениями, при условии, что фактическое распределение близко к нормальному, существует следующее примерное соотношение

$$\sigma = 1,25 \times \bar{d}(\bar{l}). \quad (36)$$

Исчисление среднего квадратического отклонения для явно несимметричных распределений не имеет смысла, поскольку согласно правилу мажорантности средних величин, среднее квадратическое отклонение всегда больше среднего линейного отклонения.

Среднее квадратическое отклонение имеет важное значение при анализе рядов распределения. При условии нормального распределения существует

следующая зависимость между величиной среднего квадратического отклонения и числом единиц наблюдения:

1) в пределах $\bar{X} \pm 1\sigma$ располагается 0,683 или 68,3% числа единиц наблюдения;

2) в пределах $\bar{X} \pm 2\sigma$ располагается 0,954 или 95,4% числа единиц наблюдения;

3) в пределах $\bar{X} \pm 3\sigma$ располагается 0,997 или 99,7% числа единиц наблюдения. На практике почти не встречаются отклонения, которые превышают $\pm 3\sigma$, поэтому отклонение $\pm 3\sigma$ считается максимально возможным. Это положение называется «правилом трех сигм».

Коэффициент вариации характеризует степень отклонения индивидуальных значений признака от средней величины в относительных величинах. Поскольку значение среднего квадратического отклонения зависит не только от степени варьирования признака, но и от абсолютных значений вариант, сравнивать средние квадратические отклонения с разными абсолютными значениями и единицами измерения невозможно. Для достижения возможности сравнения, различные средние квадратические отклонения приводятся в сопоставимый вид путем преобразования их в относительные показатели, выраженные в одинаковых относительных величинах, то есть путем расчета коэффициента вариации. Коэффициент вариации дает относительную оценку вариации признака и позволяет сравнивать степень вариации различных признаков с разными единицами измерения.

Коэффициент вариации определяется по формуле

$$v = \frac{\sigma}{\bar{X}} \times 100. \quad (37)$$

Коэффициент вариации показывает, на сколько процентов отклоняется в среднем каждое индивидуальное значение признака от средней величины в ту или другую сторону.

Коэффициент вариации является критерием однородности изучаемой совокупности, степени варьирования изучаемого признака, степени типичности и достоверности средней величины. Если коэффициент вариации меньше 33%, то изучаемую совокупность можно считать однородной по анализируемому признаку, признак – слабо варьирующим, а рассчитанную среднюю величину типичной и достоверной характеристикой данной совокупности. Если коэффициент вариации больше 33%, то изучаемую совокупность нельзя считать однородной, в результате сильной вариации анализируемого признака, а полученную среднюю величину можно считать нетипичной и недостоверной характеристикой данной совокупности.

Коэффициент осцилляции является разновидностью коэффициента вариации и характеризует отношение размаха вариации к средней величине. Определяется по формуле

$$\nu_R = \frac{R}{\bar{X}} \times 100. \quad (38)$$

Коэффициент осцилляции показывает, на сколько процентов отклоняется в среднем каждое индивидуальное значение признака от максимального (минимального) значения признака в данной совокупности.

Относительное линейное отклонение (линейный коэффициент вариации) является разновидностью коэффициента вариации и характеризует отношение среднего линейного отклонения к средней величине или медиане. Определяется по формулам

$$\nu_l = \frac{l}{\bar{X}} \times 100 \quad (39)$$

или

$$\nu_l = \frac{l}{M_e} \times 100. \quad (40)$$

Линейный коэффициент вариации показывает, на сколько процентов отклоняется в среднем каждое индивидуальное значение признака от средней величины или медианы.

Пример расчета средних величин и показателей вариации в дискретном ряду распределения.

Для расчета средних величин и показателей вариации строится вспомогательная таблица 7.

Таблица 7

Расчет средних величин и показателей вариации в дискретном ряду
распределения

Число детей, чел., X	Колич ество семей, f	$X \times f$	Сумм а накоп ленны х частот , S	$X - \bar{X}$	$ X - \bar{X} $	$ X - \bar{X} \times f$	$(X - \bar{X})^2$	$(X - \bar{X})^2 \times f$
1	3500	3500	3500	-1	1	3500	1	3500
2	4000	8000	7500	0	0	0	0	0
3	1700	5100	9200	1	1	1700	1	1700
4	600	2400	9800	2	2	1200	4	2400
5	200	1000	10000	3	3	600	9	1800
Итого	10000	20000	-	-	-	7000	-	9400

Среднее количество детей в семье определяется по формуле средней арифметической взвешенной. Среднее число детей: $\bar{X} = \frac{\sum X \times f}{\sum f} = \frac{20000}{10000} = 2$ чел.

Расчет показал, что в среднем по данной совокупности на одну семью приходится по 2 ребенка.

Мода в дискретном ряду распределения определяется по наибольшей частоте, т.е. мода – это варианта, имеющая наибольшую частоту. Наибольшая частота $f = 4000$, значит мода: $M_0 = 2$ чел.

Расчет показал, что наибольшее число семей данной совокупности имеет по 2 ребенка.

Медиана в дискретном ряду распределения – это варианта, находящаяся в середине ранжированного ряда. Определяется по сумме накопленных частот (S). Медианой является та варианта, сумма накопленных частот которой первая превышает половину объема совокупности. Объем совокупности: $\sum f = 10000$.

Половина объема совокупности $1 / 2 \Sigma f = 10000 / 2 = 5000$. Следовательно, накопленная частота, первая, превышающая половину объема совокупности: $S_2 = 8000$. Следовательно, медиана $M_e = 2$ чел.

Расчет показал, что половина семей данной совокупности имеет меньше двух детей, вторая половина семей данной совокупности имеет более двух детей.

Размах вариации: $R = X_{\max} - X_{\min} = 5 - 1 = 4$ чел.

Расчет показал, что разница между наибольшим и наименьшим числом детей в данной совокупности составляет 4 ребенка.

Среднее линейное отклонение: $l = \frac{\Sigma |X - \bar{X}| \times f}{\Sigma f} = \frac{7000}{10000} = 0,7$ чел.

Расчет показал, что число детей в каждой семье данной совокупности отклоняется от общего среднего числа детей по всей совокупности в ту или другую сторону в среднем на 0,7 человек.

Дисперсия: $\sigma^2 = \frac{\Sigma (X - \bar{X})^2 \times f}{\Sigma f} = \frac{9400}{10000} = 0,94$.

Среднее квадратическое отклонение: $\sigma = \sqrt{\frac{\Sigma X^2 \times f}{\Sigma f} - (\bar{X}^2)} = \sqrt{\sigma^2} = \sqrt{0,94} = 0,97$ чел.

Расчет показал, что число детей в каждой семье данной совокупности отклоняется от общего среднего числа детей по всей совокупности семей в ту или другую сторону в среднем на 0,97 человек.

Коэффициент вариации: $v = \frac{\sigma}{\bar{X}} \times 100 = \frac{0,97}{2} \times 100 = 48,5\%$

Расчет показал, что число детей в каждой семье данной совокупности отклоняется от общего среднего числа детей по всем семьям совокупности в ту или другую сторону в среднем на 48,5%. Т.к. коэффициент вариации больше 33 %, то можно сказать, что признак – число детей в данной совокупности семей варьирует сильно. Следовательно, изучаемую совокупность семей можно считать неоднородной по числу детей, а рассчитанную среднюю нетипичной и недостоверной характеристикой данной совокупности.

Пример расчета средних величин и показателей вариации в интервальном ряду распределения.

Расчет средних величин и показателей вариации возможен только по дискретному ряду распределения. Поэтому, с целью расчета средних величин и показателей вариации интервальный ряд распределения преобразуется в дискретный путем расчета средин интервалов. Середина каждого интервала определяются как полусумма его верхней и нижней границ. (табл. 8).

Таблица 8

Расчет средних величин и показателей вариации в интервальном ряду
распределения

Среднесписочная численность работников предприятий, чел.	Число предприятий, f	Середина интервала, X	$X \times f$	Сумма накопленных частот, S	$X - \bar{X}$	$ X - \bar{X} $	$ X - \bar{X} \times f$	$(X - \bar{X})^2$	$(X - \bar{X})^2 \times f$
20-100	12	60	720	12	-82,7	82,7	992,4	6839,3	82071,6
100-180	9	140	1260	21	-2,7	2,7	24,3	7,3	65,7
180-260	5	220	1100	26	+77,3	77,3	386,5	5975,3	29876,5
260-340	4	300	1200	30	+157,3	157,3	629,2	24743,3	98973,2
Итого	30	-	4280	-	-	-	2032,4	-	210987,0

Среднесписочная численность работников по всей совокупности предприятий определяется по формуле средней арифметической взвешенной

$$\text{Среднесписочная численность работников } \bar{X} = \frac{\sum X \times f}{\sum f} = \frac{4280}{30} = 142,7 \text{ чел.}$$

Расчет показал, что в среднем по всей совокупности предприятий среднесписочная численность работников составила 142,7 человек на одно предприятие.

Для расчета моды в интервальном ряду распределения определяется модальный интервал по наибольшей частоте. Наибольшая частота $f = 12$, поэтому модальным является первый интервал «от 20 чел. до 100 чел.»

$$\text{Мода } M_o = X_0 + i \times \frac{f_{M_o} - f_{M_{o-1}}}{(f_{M_o} - f_{M_{o-1}}) + (f_{M_o} - f_{M_{o+1}})} = 20 + 80 \times \frac{12 - 0}{(12 - 0) + (12 - 9)} = 84 \text{ чел.}$$

Расчет показал, что наибольшее число предприятий данной совокупности имеет численность работников 84 человек.

Для расчета медианы в интервальном ряду распределения определяется медианный интервал по сумме накопленных частот. Объем совокупности: $\Sigma f = 30$. Половина объема совокупности: $1/2 \Sigma f = 30/2 = 15$. Следовательно, первая накопленная частота, превышающая половину объема совокупности, это $S_2 = 21$. Поэтому медианным является второй интервал «от 100 чел. до 180 чел.». Следовательно, медиана равна

$$\text{Медиана } M_e = X_0 + i \times \frac{1/2 \Sigma f - S_{M_e}}{f_{M_e}} = 100 + 80 \times \frac{15 - 12}{9} = 126,7 \text{ чел.}$$

Расчет показал, что половина предприятий данной совокупности имеет среднесписочную численность работников меньше 126,7 человек, а вторая половина предприятий – больше 126,7 человек.

$$\text{Размах вариации: } R = X_{\max} - X_{\min} = 340 - 20 = 320 \text{ чел.}$$

Расчет показал, что разница между наибольшей и наименьшей среднесписочной численностью работников предприятий данной совокупности составляет 320 человек.

$$\text{Среднее линейное отклонение: } l = \frac{\Sigma |X - \bar{X}| \times f}{\Sigma f} = \frac{2032,4}{30} = 67,7 \text{ чел.}$$

Расчет показал, что среднесписочная численность работников каждого предприятия данной совокупности отклоняется в среднем от общей среднесписочной численности в ту или другую сторону на 67,7 человек.

$$\text{Дисперсия: } \sigma^2 = \frac{\Sigma (X - \bar{X})^2 \times f}{\Sigma f} = \frac{210987,0}{30} = 7032,9.$$

$$\text{Среднее квадратическое отклонение: } \sigma = \sqrt{\sigma^2} = \sqrt{7032,9} = 83,9 \text{ чел.}$$

Расчет показал, что среднесписочная численность работников каждого предприятия данной совокупности отклоняется от общей средней численности работников по совокупности в ту или другую сторону в среднем на 83,9 человек.

$$\text{Коэффициент вариации: } \nu = \frac{\sigma}{\bar{X}} \times 100 = \frac{83,9}{142,7} \times 100 = 58,8\%.$$

Расчет показал, что среднесписочная численность работников каждого предприятия данной совокупности отклоняется от общей средней численности работников по совокупности в ту или другую сторону в среднем на 58,8%.

Поскольку, коэффициент вариации превышает 33%, следовательно, изучаемую совокупность предприятий можно считать неоднородной по среднесписочной численности работников, признак можно считать варьирующим сильно, поэтому рассчитанную среднюю можно считать нетипичной и недостоверной характеристикой данной совокупности.

4.5 Вариация сгруппированных данных. Виды дисперсий, правило их сложения

Вариация сгруппированных данных характеризуется следующими видами вариаций и дисперсий:

1. Общая вариация – общая дисперсия.
2. Межгрупповая вариация – межгрупповая дисперсия.
3. Внутригрупповая вариация – внутригрупповая дисперсия.

Общая дисперсия характеризует вариацию признака под влиянием всех, обуславливающих ее факторов и определяется по формуле

$$\sigma^2 = \frac{\sum (X_i - \bar{X})^2 \times f_i}{\sum f_i}, \quad (41)$$

где X_i – индивидуальное значение признака;

\bar{X} - среднее значение признака по совокупности;

f_i - частота.

Межгрупповая дисперсия характеризует систематическую вариацию – различия в величине признака, возникающие под влиянием признака-фактора, положенного в основу группировки и определяется по формуле

$$\delta^2 = \frac{\sum (\bar{X}_i - \bar{X})^2 \times n_i}{\sum n_i}, \quad (42)$$

где \bar{X}_i - групповые средние;

n_i – численность единиц в группах.

Внутригрупповая дисперсия характеризует случайную вариацию – часть вариации, возникающую под влиянием остаточных неучтенных (случайных) факторов, которая не зависит от влияния группировочного признака и определяется по формуле

$$\sigma_i^2 = \frac{\sum (X_i - \bar{X})^2 \times n_i}{\sum n_i}. \quad (43)$$

Внутригрупповых дисперсий, как правило, несколько (в зависимости от числа групп), поэтому определяется средняя из внутригрупповых дисперсий по формуле

$$\bar{\sigma}_i^2 = \frac{\sum \sigma_i^2 \times n_i}{\sum n_i}. \quad (44)$$

Существует закон, который связывает все три вида дисперсий, который называется правилом сложения (разложения) дисперсий и заключается в том, что общая дисперсия равна сумме из внутригрупповых и межгрупповых дисперсий

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2. \quad (45)$$

Согласно правилу сложения дисперсий, общая дисперсия, возникающая под действием всех факторов, равна сумме дисперсии, возникающей под влиянием прочих факторов и дисперсии, возникающей под влиянием группировочного признака. Правило сложения (разложения) дисперсий применяется при расчете показателей тесноты связи по сгруппированным данным в дисперсионном анализе. При этом определяются показатели тесноты связи: 1) эмпирический коэффициент детерминации; 2) эмпирическое корреляционное отношение.

Эмпирический коэффициент детерминации характеризует долю межгрупповой дисперсии в общей дисперсии, и определяется по формуле

$$\eta^2 = \frac{\delta^2}{\sigma^2}. \quad (46)$$

Эмпирический коэффициент детерминации показывает долю (удельный вес) общей вариации, обусловленную группировочным признаком.

Эмпирическое корреляционное отношение характеризует степень влияния группировочного признака на вариацию результативного признака и определяется по формуле

$$\eta = \sqrt{\eta^2} = \sqrt{\frac{\delta^2}{\sigma^2}}. \quad (47)$$

Эмпирическое корреляционное отношение принимает значения в пределах от 0 до 1. Если $\eta = 0$, то группировочный признак не влияет на результативный признак; если $\eta = 1$, то на результативный признак оказывает влияние только группировочный признак, а влияние прочих остаточных факторов равно нулю. Следовательно, чем ближе значение η к нулю, тем связь между результативным признаком и факторным признаком, положенным в основу группировки, слабее, и наоборот, чем ближе значение η к единице, тем связь между признаками сильнее.

Пример дисперсионного анализа.

Дисперсионный анализ проводится на основании результатов статистической группировки, построенной в главе 5 (табл. 9).

Таблица 9

Характеристика групп

Интервалы по качеству почв, баллы	Число хозяйств в группе	Средняя балльная оценка почв, балл	Средняя урожайность овощей, т/га
До 65	8	50,25	13,19
65 – 75	3	71,00	17,13
75 – 85	4	80,75	20,27
85 – 95	5	89,80	26,66
Всего и в среднем по совокупности	20	69,35	18,50

Аналитическая группировка (построенная в главе 6) показала, что между средней балльной оценкой почв и средней урожайностью овощей наблюдается прямая зависимость, то есть, при увеличении средней балльной оценки почв от первой к четвертой группе, средняя урожайность овощей также увеличивается.

Для расчета межгрупповой дисперсии строится вспомогательная таблица 10.

Таблица 10

Расчет межгрупповой дисперсии

Интервалы по качеству почв, баллы	Число хозяйств в группе, f	Средняя урожайность овощей, т/га, Y	$y - \bar{y}$	$(y - \bar{y})^2$	$(y - \bar{y})^2 \times f$
До 65	8	13,19	-5,31	28,20	225,60
65 – 75	3	17,13	-1,37	18,77	56,31
75 – 85	4	20,27	1,77	3,13	12,52
85 – 95	5	26,66	8,16	66,58	332,90
В среднем по совокупности	20	18,50	-	-	627,33

Межгрупповая дисперсия определяется по формуле 42:

$$\sigma^2 = \frac{\sum(\bar{Y}_{gp} - \bar{Y}_o)^2 \times f}{\sum f} = \frac{627,33}{20} = 31,37.$$

Общая дисперсия в данном случае определяется по индивидуальным значениям признака по формуле

$$\sigma^2 = \frac{\sum Y^2}{n} - (\bar{Y})^2, \quad (50)$$

где $\bar{Y} = \frac{\sum Y}{n} = \frac{374,0}{20} = 18,7$ т/га.

Для определения общей дисперсии строится вспомогательная таблица 11.

Таблица 11

Расчет общей дисперсии

№ п/п	Y	Y ²	№ п/п	Y	Y ²
1	16,0	256,0	11	18,0	324,0
2	18,0	324,0	12	21,5	462,25
3	15,0	225,0	13	23,0	529,0
4	12,0	144,0	14	22,0	484,0
5	23,0	529,0	15	16,0	256,0
6	20,0	400,0	16	18,0	324,0
7	9,0	81,0	17	29,0	841,0
8	13,0	169,0	18	13,0	169,0
9	31,5	992,25	19	14,0	196,0
10	14,0	196,0	20	28,0	784,0
Итого	-	-	-	374,0	7685,5

Общая дисперсия: $\sigma^2 = \frac{7685,5}{20} - (18,7)^2 = 384,275 - 349,69 = 34,585.$

Коэффициент детерминации определяется по формуле 46:

$$\eta^2 = \frac{\delta^2}{\sigma^2} = \frac{31,37}{34,585} = 0,907 \text{ (90,7\%)}.$$

Коэффициент детерминации показывает, что 90,7% вариации урожайности овощей обусловлено изменением качества почв, а остальные 9,3% вариации урожайности овощей обусловлены изменением остаточных, неучтенных в данном случае факторов.

Эмпирическое корреляционное отношение определяется по формуле 47:

$$\eta = \sqrt{0,907} = 0,9524.$$

Эмпирическое корреляционное отношение показывает, что между урожайностью овощей и качеством почв наблюдается сильная связь, так как значение эмпирического корреляционного приближено к единице.

Внутригрупповая дисперсия определяется по правилу сложения (разложения) дисперсий (43): $\bar{\sigma}_i^2 = \sigma^2 - \delta^2 = 34,585 - 31,37 = 3,215$.

Согласно правилу сложения (разложения) дисперсий, получим:

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2 = 3,215 + 31,37 = 34,585.$$

При этом, если выразить значения дисперсий в относительных величинах – процентах, приняв значение общей дисперсии за 100%, получим:

$$\sigma^2 = \bar{\sigma}_i^2 + \delta^2 = (3,215 \div 34,585 \times 100) + (31,37 \div 34,585 \times 100) = 9,29\% + 90,71\% = 100\%$$

Расчет показал, что на влияние систематического (основного) фактора, положенного в основу группировки – качества почв, приходится 90,71% вариации результативного признака – урожайности овощей. А на влияние остаточных, неучтенных в данном случае факторов приходится 9,29% вариации результативного признака – урожайности овощей. Следовательно, влияние систематического фактора существенно.

5 СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

5.1 Сущность, виды и правила построения статистической сводки

На первой стадии статистического исследования в результате статистического наблюдения получают материалы, которые содержат информацию о каждой отдельной единице изучаемой статистической совокупности.

Задачи статистики на второй стадии статистического исследования следующие: 1) упорядочить материалы статистического наблюдения; 2) систематизировать полученные характеристики единиц статистической совокупности; 3) дать сводную характеристику всей статистической совокупности и отдельных ее частей при помощи обобщающих статистических показателей.

Указанные задачи выполняются при помощи статистической сводки. Следовательно, статистическая сводка – это вторая стадия статистического исследования, в процессе которой первичный статистический материал обобщается, образуя статистические совокупности, которые характеризуются итоговыми сводными показателями.

На стадии статистической сводки осуществляется переход от характеристики каждой отдельной единицы совокупности к характеристике групп и совокупности в целом, то есть – от характеристики индивидуальных значений варьирующего признака у отдельных единиц к характеристике общего их проявления в массе.

Статистическая сводка разделяется на виды по следующим классификационным признакам: по глубине обработки материалов; по форме обработки материалов. (рис. 19).

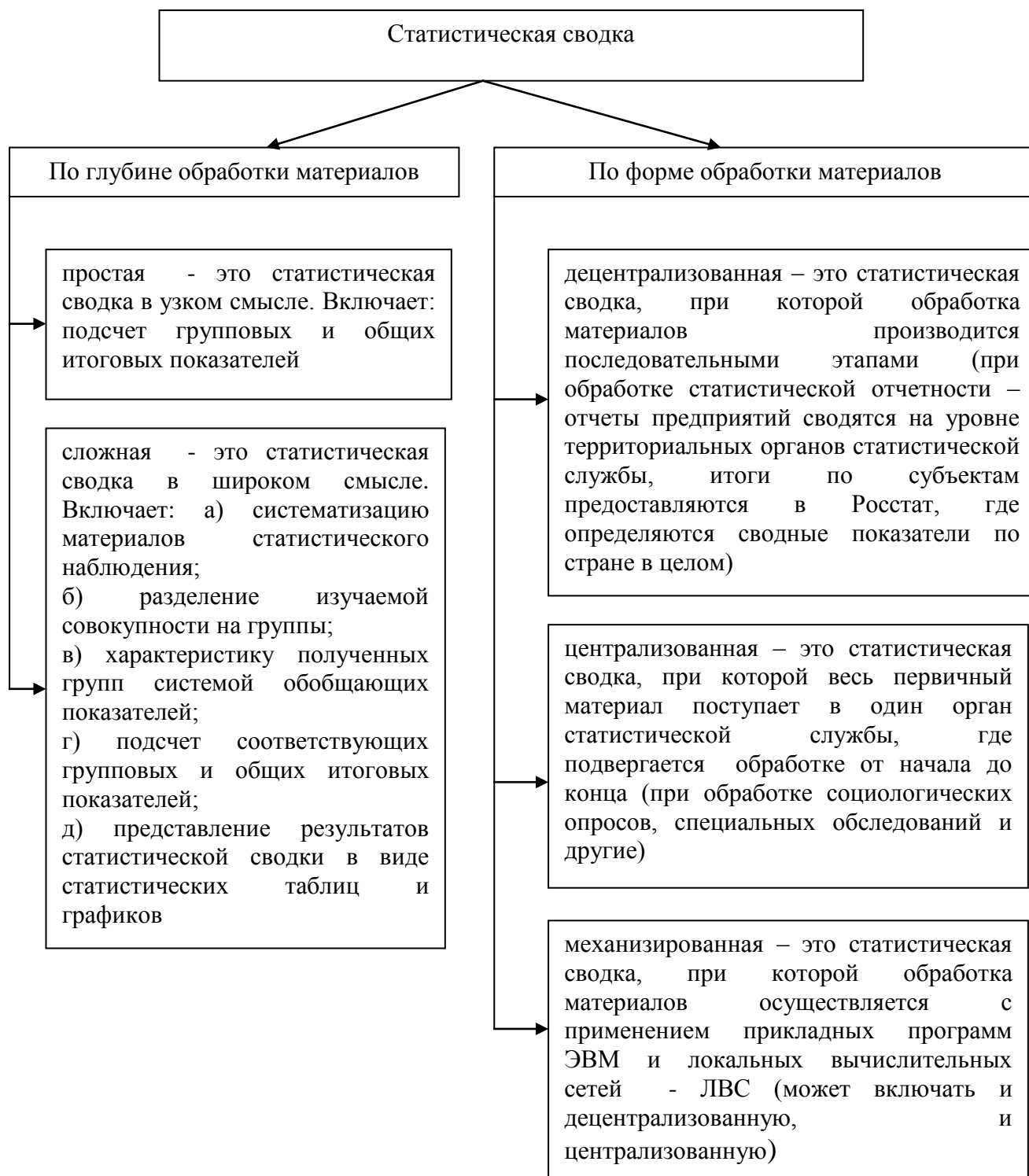


Рисунок 19 – Виды статистической сводки

Статистическая сводка состоит из двух элементов: подлежащее статистической сводки; сказуемое статистической сводки (табл. 12).

Статистическая сводка

Подлежащее статистической сводки - это перечень изучаемых явлений, групп и подгрупп, на которые разделена изучаемая совокупность (по горизонтали):	Сказуемое статистической сводки – это перечень показателей, которыми охарактеризованы изучаемое явление, объект, каждая группа, подгруппа и совокупность в целом, то есть - подлежащее таблицы, (по вертикали)			
	сказуемое	сказуемое	сказуемое	...
подлежащее				
подлежащее				
подлежащее				
....				

Проведению статистической сводки предшествует разработка ее программы, которая включает 1) выбор группировочных признаков; 2) определение порядка формирования групп; 3) разработку системы статистических показателей для характеристики групп и совокупности в целом; 4) разработку системы макетов статистических таблиц для представления результатов статистической сводки.

5.2 Сущность и виды статистической группировки. Статистические классификации

При статистической сводке в широком смысле материалы статистического наблюдения упорядочиваются, систематизируются, на основе чего образуются новые статистические совокупности. Отдельные единицы статистической совокупности объединяются в группы при помощи метода статистической группировки. Это позволяет сжать информацию, полученную в результате статистического наблюдения, на основе чего выявить закономерности развития изучаемого явления.

Следовательно, статистическая группировка – это разделение статистической совокупности на однородные группы по существенному варьирующему признаку.

Статистические группировки имеют следующее значение в статистических исследованиях: 1) раскрывают объективное состояние общественных явлений; 2) выявляют существенные черты и свойства изучаемых общественных явлений; 3) дают информацию о размерности отдельных групп и соотношении их в общей совокупности; 4) выявляют причинно-следственные связи между изучаемыми показателями; 5) выявляют вид и направление взаимосвязи между изучаемыми показателями.

Статистические группировки подразделяются на виды по следующим классификационным признакам (рис. 20).

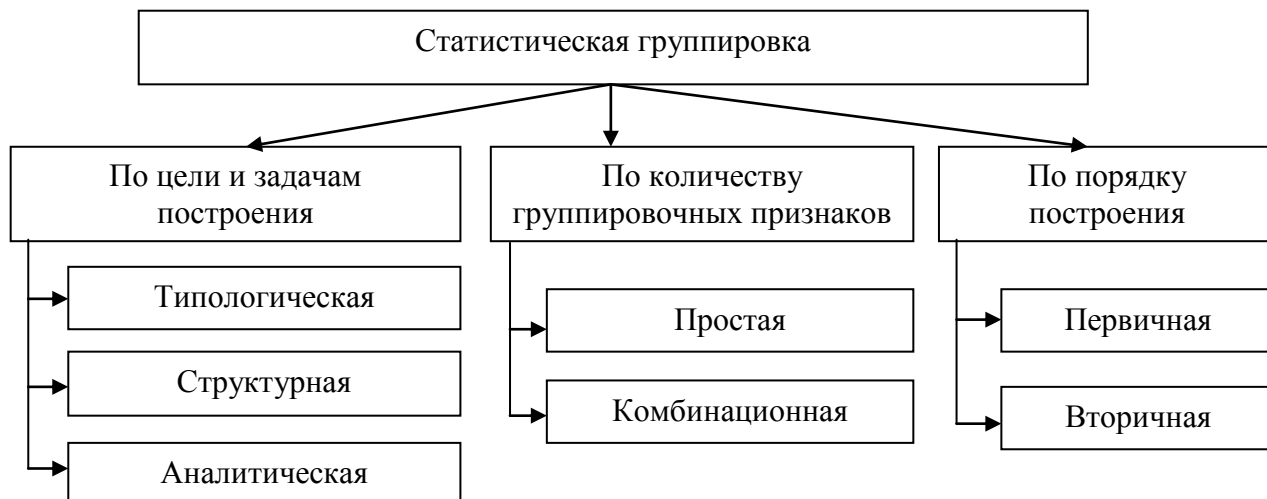


Рисунок 20 – Виды статистических группировок

Основные задачи, решаемые при построении статистических группировок: 1) выделение социально-экономических типов явлений; 2) изучение состава изучаемого явления и изменений в его структуре; 3) выявление взаимосвязи между изучаемыми признаками. В соответствии с этими задачами статистические группировки бывают трех видов: типологические; структурные; аналитические.

Типологическая группировка используется для разделения статистической совокупности на качественно однородные группы для выделения однородных социально-экономических типов явлений (группировки предприятий по формам собственности, группировка предприятий по

организационно-правовой форме хозяйствования, группировка населения по социальному статусу и др.). В основу типологической группировки, как правило, положен качественный группировочный признак.

Структурная группировка характеризует состав изучаемого явления и структурные сдвиги в развитии явлений и процессов (разделение населения по полу, возрасту трудоспособности, образованию и др. с характеристикой удельного веса каждой группы в общем объеме совокупности). В основу структурных группировок может быть положен как качественный, так и количественный группировочный признак. При сопоставлении структурных группировок за ряд периодов оцениваются структурные сдвиги в составе изучаемой совокупности за этот период.

Аналитическая группировка используется для выявления взаимосвязи и зависимости между явлениями и различными их признаками (зависимость уровня рождаемости от уровня доходов населения, зависимость уровня производительности труда работников от их заработной платы и др.). В основу аналитической группировки, как правило, положен количественный группировочный признак.

При построении аналитической группировки основная задача – это правильность выбора факторного признака и результативного признака. Существует несколько вариантов зависимостей: 1) признак-фактор количественный, признак-результат качественный (признак-фактор стаж работы (количественный) влияет на признак-результат квалификацию работника (качественный)); 2) признак-фактор качественный, признак-результат количественный (признак-фактор квалификация работников (качественный) влияет на признак-результат производительность труда работников (количественный)); 3) признак-фактор и признак-результат качественные (признак-фактор уровень образования (качественный) влияет на признак-результат квалификацию работников (качественный)); 4) признак-фактор и признак-результат количественные (признак-фактор

производительность труда работников (количественный) оказывает влияние на признак-результат заработную плату работников (количественный)).

По количеству группировочных признаков статистические группировки бывают двух видов: простые; комбинированные (комбинационные).

Простая группировка строится по одному группировочному признаку, положенному в ее основу. При этом, статистическая совокупность разделяется на группы по данному группировочному признаку. Комбинационная группировка строится по двум или нескольким группировочным признакам. При этом, статистическая совокупность разделяется на группы по первому группировочному признаку, которые затем в свою очередь разделяются на подгруппы по второму группировочному признаку и т.д.

По порядку построения статистические группировки бывают двух видов: первичные; вторичные.

Первичная – это первоначально построенная группировка по материалам статистического наблюдения. Вторичная – это перегруппировка ранее сгруппированных данных, которая производится для лучшей характеристики изучаемого явления, либо для приведения к сопоставимому виду группировок с различными интервалами с целью их сравнения. Вторичная группировка строится на основе первичной группировки двумя способами: а) способом долевого перегруппировки – переносом части единиц из одной группы в другую; б) способом преобразования (укрупнения или уменьшения) интервалов первичной группировки.

Особым видом статистической группировки являются классификации. Классификация – это устойчивая номенклатура классов и групп, которая имеет фундаментальное значение для проведения всех статистических работ.

Особенности классификации в отличие от статистической группировки следующие: а) в основу классификации положен в основном качественный признак; б) классификация является узаконенным стандартом, который устанавливается органами государственной, ведомственной и международной статистики; в) классификация является нормативом и едина для любого

исследования, независимо от цели его проведения и органов его проводящих;

г) классификация устойчива, то есть остается неизменной в течение длительного времени, а изменения и дополнения в нее вносятся в установленном порядке официальными органами статистической службы.

Разновидностью аналитических группировок являются корреляционные группировки и ряды распределения.

Корреляционные группировки - это разделение совокупности на однородные группы и подгруппы по двум или нескольким взаимосвязанным варьирующим признакам. То есть корреляционная группировка представляет собой аналитическую комбинационную группировку.

Ряды распределения - это группировки, в которых группы охарактеризованы одним показателем – числом единиц в группе (частотой) или удельным весом каждой группы во всей изучаемой совокупности (частостью).

5.3 Статистические ряды распределения, их виды и основные характеристики

Статистический ряд распределения – это упорядоченное распределение единиц совокупности по определенному варьирующему признаку.

В зависимости от признака, положенного в основу ряда распределения, они бывают двух видов: атрибутивные; вариационные.

Атрибутивный ряд распределения – это ряд, построенный по качественному признаку, который характеризует состав совокупности, то есть долю отдельных групп в общем объеме совокупности (группировка предприятий по форме собственности, охарактеризованная числом предприятий каждой формы собственности и их удельным весом в общей численности предприятий).

Вариационный ряд распределения – это ряд, построенный по количественному признаку.

Вариационный ряд распределения состоит из двух элементов:

- 1) варианты – это отдельные значения признака, которые он принимает в вариационном ряду;
- 2) частоты – это число единиц совокупности, имеющих одинаковое значение признака, то есть это показатель, который характеризует число встреч одной и той же варианты в данном ряду распределения. Если частоты выражаются в процентах (%) к итогу, то они называются частостями.

В зависимости от характера вариации признака вариационные ряды распределения бывают трех видов: дискретный; интервальный; ранжированный.

Дискретный вариационный ряд распределения характеризует распределение единиц совокупности по признаку, изменяющемуся прерывно, то есть через определенное число единиц. В дискретном ряду распределения значения признака выражаются, как правило, целыми числами (табл. 13 на основании данных табл.7).

Таблица 13

Ряд распределения семей по числу детей (дискретный)

Число детей, чел. (варианта), X	Количество семей (частота), f
1	3500
2	4000
3	1700
4	600
5	200
Всего	10000

Для графического изображения дискретного вариационного ряда используется специфическая диаграмма - полигон (рис. 21).

Полигон – это ломаная линия, отрезки которой соединяют точки с координатами значений вариант (группировочного признака) - откладываются по оси X и соответствующих им частот – откладываются по оси Y.

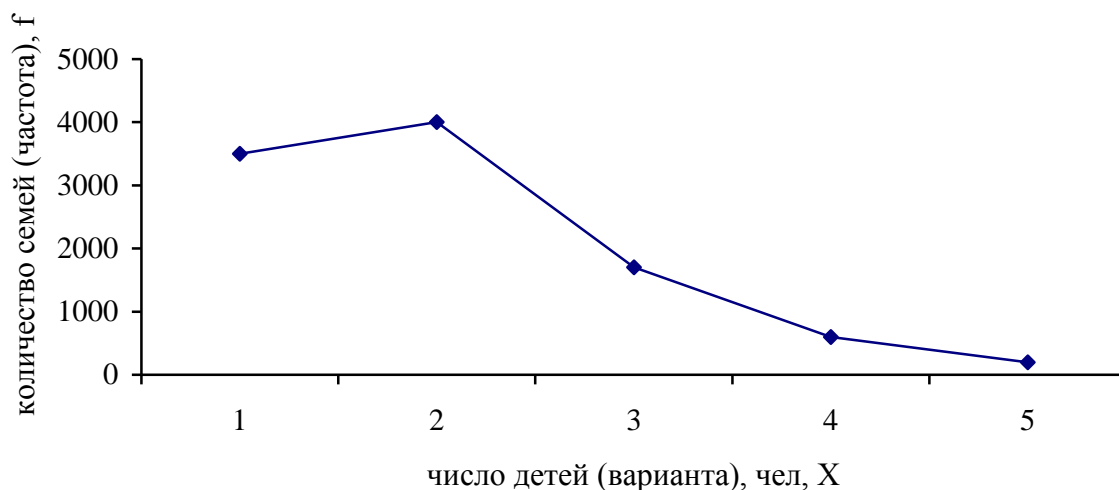


Рисунок 21 - Полигон распределения семей по числу детей

Интервальный вариационный ряд распределения характеризует распределение единиц совокупности по признаку, изменяющемуся непрерывно, то есть принимающему любые значения в пределах определенного промежутка (интервала). В интервальном ряду распределения значения признака выражаются любыми дробными числами (табл. 14 на основе данных табл. 8).

Таблица 14

Ряд распределения предприятий по численности работников

Среднесписочная численность работников предприятий, чел. (варианты) , X	Число предприятий, (частоты), f
20-100	12
100-180	9
180-260	5
260-340	4
Итого	30

Интервальный ряд распределения изображается графически в виде специфической диаграммы - гистограммы (рис. 22). Гистограмма – это ступенчатая фигура, состоящая из прямоугольников, основаниями которых являются величины интервалов (значения группировочного признака), - откладываются по оси X, а высотами - соответствующие частоты – откладываются по оси Y.

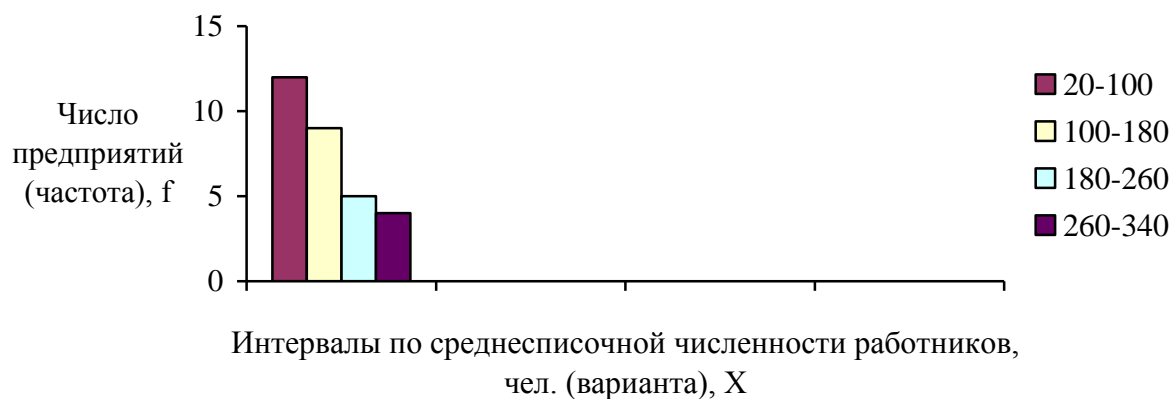


Рисунок 22 - Гистограмма распределения предприятий по среднесписочной численности работников

Интервальный ряд распределения может преобразовываться в дискретный путем определения середин интервальных значений признака (табл. 15).

Таблица 15

Распределение предприятий по среднесписочной численности работников

Среднесписочная численность работников предприятий, чел. (варианты), X	Число предприятий, (частоты), f	Средины интервалов, чел. (варианты), X
20-100	12	60
100-180	9	140
180-260	5	220
260-340	4	300
Итого	30	-

В соответствии с преобразованием интервального ряда распределения в дискретный гистограмма может преобразовываться в полигон путем соединения ломаной линией точек пересечения значений середин интервалов и соответствующих им частот.

Кроме указанных графических изображений, ряды распределения могут наглядно изображаться в виде кумуляты накопленных частот. Кумулята – это ломаная линия, отрезки которой соединяют значения вариант и

соответствующих им накопленных частот (табл. 16). Значения вариант откладываются по оси X , значения накопленных частот – по оси Y (рис. 23).

Таблица 16

Расчет накопленных частот

Число детей, чел. (варианта), X	Количество семей (частота), f	Количество семей (накопленные частоты), S
1	3500	3500
2	4000	7500
3	1700	9200
4	600	9800
5	200	10000
Всего	10000	-

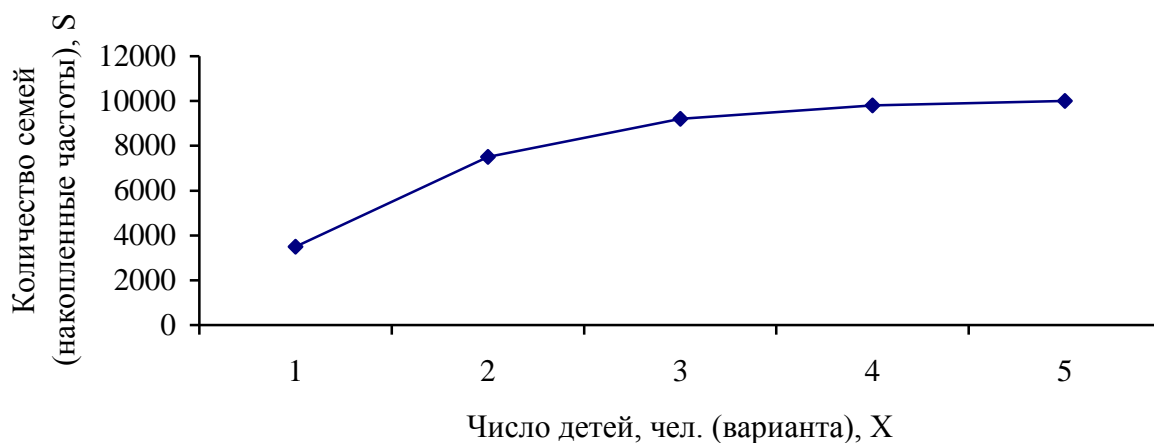


Рисунок 23 - Кумулята распределения семей по числу детей

Ранжированный ряд распределения является производным от дискретного или интервального ряда распределения.

Ранжированный ряд распределения – это ряд, в котором единицы расположены в порядке возрастания или убывания группировочного признака. Наглядно ранжированный ряд распределения изображается в виде специфической диаграммы - огивы Гальтона. Огива Гальтона – это постоянно возрастающая (убывающая) ломаная линия, отрезки которой соединяют номера единиц совокупности и значения признака (рис.26).

В вариационных рядах распределения может прослеживаться определенная зависимость между изменениями значений вариант и значений

частот – частоты изменяются закономерно в соответствии с изменениями значений вариант. Такая закономерность изменения часто называется закономерностью распределения.

Выявление закономерности распределения является одной из основных задач при статистическом анализе рядов распределения. Чтобы выявление закономерности было возможно, необходимо соблюдение следующих требований: 1) соблюдение достаточно большой численности статистических совокупностей; 2) определение оптимального числа групп и величин интервалов, при которых более отчетливо проявляется статистическая закономерность.

Рассмотренный ранее полигон, наглядно изображающий дискретный ряд распределения, можно считать эмпирической кривой распределения, которая отражает общее изменение признака под влиянием всех факторов.

Эмпирическая кривая распределения это – графическое изображение изменения частот, функционально связанного с изменением значений вариант вариационного ряда, в виде непрерывной линии.

Теоретическая кривая распределения это – непрерывная линия, отражающая общую закономерность данного распределения, складывающуюся под влиянием систематического фактора, при исключении влияния случайных (остаточных) факторов.

Кривые распределения в статистике бывают двух видов: 1) одновершинные – характерные для однородных статистических совокупностей (симметричные, умеренно симметричные, крайне асимметричные); 2) многовершинные – характерные для неоднородных статистических совокупностей.

В статистике анализируются, как правило, однородные совокупности. Поэтому, при многовершинности кривой распределения необходимо провести перегруппировку данных с целью выделения более однородных групп.

При нормальном распределении (симметричном распределении) наблюдается: 1) равенство частот любых двух вариант, равноотстоящих в обе

стороны от центра распределения; 2) равенство средней арифметической, моды и медианы.

Основными задачами статистики при анализе закономерности распределения являются: 1) оценка однородности характера распределения путем расчета показателей вариации; 2) определение показателей асимметрии и эксцесса.

Относительный показатель асимметрии определяется по формуле [19]

$$As = \frac{\bar{X} - Mo}{\sigma} \quad (51)$$

или

$$As = \frac{\bar{X} - Me}{\sigma} \quad (52)$$

где \bar{X} - средняя арифметическая;

Mo - мода;

σ - среднее квадратическое отклонение;

Me - медиана.

Относительный показатель асимметрии может принимать положительное и отрицательное значение. Если As имеет положительный знак, то распределение, по которому он определен, является правосторонним (имеет правостороннюю асимметрию), при этом наблюдается следующее неравенство

$$Mo > Me > \bar{X} \quad (53)$$

Если As имеет отрицательный знак, то распределение, по которому он определен, является левосторонним (имеет левостороннюю асимметрию), при этом наблюдается следующее неравенство

$$Mo < Me < \bar{X} \quad (54)$$

Для определения асимметрии можно использовать упрощенную формулу Линдберга [19]

$$As = P - 50 \quad (55)$$

где P – доля или удельный вес вариант, значение которых выше значения средней арифметической, в общей численности вариант данного ряда распределения (%);

50 – доля или удельный вес вариант, значение которых выше значения средней арифметической в нормальном ряду распределения.

Значение относительного показателя асимметрии изменяется в пределах от 0 до ± 1 . Если $As > 0,5$ (независимо от знака), то асимметрию в данном ряду распределения можно считать значительной. Если $As < 0,25$ (независимо от знака), то асимметрию в данном ряду распределения можно считать незначительной.

Для симметричных распределений определяется показатель эксцесса, по формуле

$$E_k = \frac{\mu_4}{\sigma^4} - 3 \quad (56)$$

где μ_4 – центральный момент четвертого порядка;

σ – среднее квадратическое отклонение.

По упрощенной формуле Линберга эксцесс равен [19]

$$E_k = P - 38,29 \quad (57)$$

где P – доля или удельный вес вариант, находящихся в границах отклонения, равного половине значения σ от значения средней величины в ту или другую сторону в общем числе вариант данного ряда распределения (%);

38,29 – доля или удельный вес вариант, находящихся в границах отклонения, равного половине значения σ от значения средней величины в ту или другую сторону в общем числе вариант нормального ряда распределения (%).

Значение показателя эксцесса может принимать положительное и отрицательное значение. Если E_k имеет положительное значение, то данное распределение можно считать островершинным. Если E_k имеет отрицательное значение, то данное распределение можно считать плосковершинным. Если $E_k = 0$, то данное распределение можно считать нормальным.

5.4 Порядок построения статистической группировки по количественному признаку

Статистическая группировка строится последовательно в несколько этапов.

Первый этап – выбор группировочного признака. Основной особенностью данного этапа является строгое соблюдение требований к выбору группировочного признака.

Группировочный признак – это существенный, варьирующий признак, положенный в основу группировки.

Группировочный признак является основанием группировки и от правильности его выбора зависит правильность построения всей группировки в целом и достоверность результатов исследования.

Поэтому выбор группировочного признака должен осуществляться на основе выполнения следующих требований:

- 1) группировочный признак должен соответствовать цели и задачам группировки;
- 2) группировочный признак должен быть существенным, то есть отражать основные черты изучаемого явления;
- 3) группировочный признак должен присутствовать у каждой единицы изучаемой совокупности;
- 4) группировочный признак должен иметь достаточную вариацию.

Группировочные признаки, положенные в основу группировки подразделяются на виды по следующим классификационным признакам:

- 1) форме выражения;
- 2) характеру отражения данных;
- 3) характеру колеблемости;
- 4) роли во взаимосвязи изучаемых явлений (рис. 24).



Рисунок 24 – Виды группировочных признаков

Второй этап - построение статистического ряда распределения. Основной особенностью данного этапа является то, что распределение единиц совокупности на группы по количественному признаку, как правило, связано с построением интервального ряда распределения.

Для построения интервального ряда распределения необходимо определить

- 1) число групп;
- 2) величину интервала;
- 3) границы групп;
- 4) число единиц в группах.

Число групп, на которые делится изучаемая совокупность, зависит от двух факторов: а) от размаха вариации, то есть от разницы между максимальным и минимальным значениями группировочного признака. Зависимость следующая: чем больше размах вариации, тем на большее число групп можно поделить изучаемую совокупность; б) от объема изучаемой совокупности, то есть от числа единиц, составляющих совокупность. Зависимость следующая: чем больше объем совокупности, тем на большее число групп можно ее разделить.

При определении числа групп необходимо соблюдать следующие требования:

- 1) каждая группа должна характеризовать однородные типы явлений, то есть не допускается резких различий группировочного признака внутри групп;
- 2) в результате статистической группировки должно получаться достаточно большое число групп (не менее трех групп) для обеспечения возможности выявления закономерности развития изучаемого явления;
- 3) число единиц в группах должно быть достаточно большим (не менее трех единиц) для соблюдения принципа массовости статистической совокупности.

Число групп в достаточно больших совокупностях (20 и более единиц) определяется по формуле Стерджесса

$$n = 1 + 3,322 \lg N, \quad (58)$$

где n – число групп;

N – число единиц совокупности.

Показатели числа групп и величины интервала являются обратно зависимыми – чем больше число групп, тем меньше величина интервала, и наоборот.

Величина интервала в группах зависит от их вида. Интервалы в статистической группировке построенной по интервальному признаку могут быть четырех видов (рис. 25).

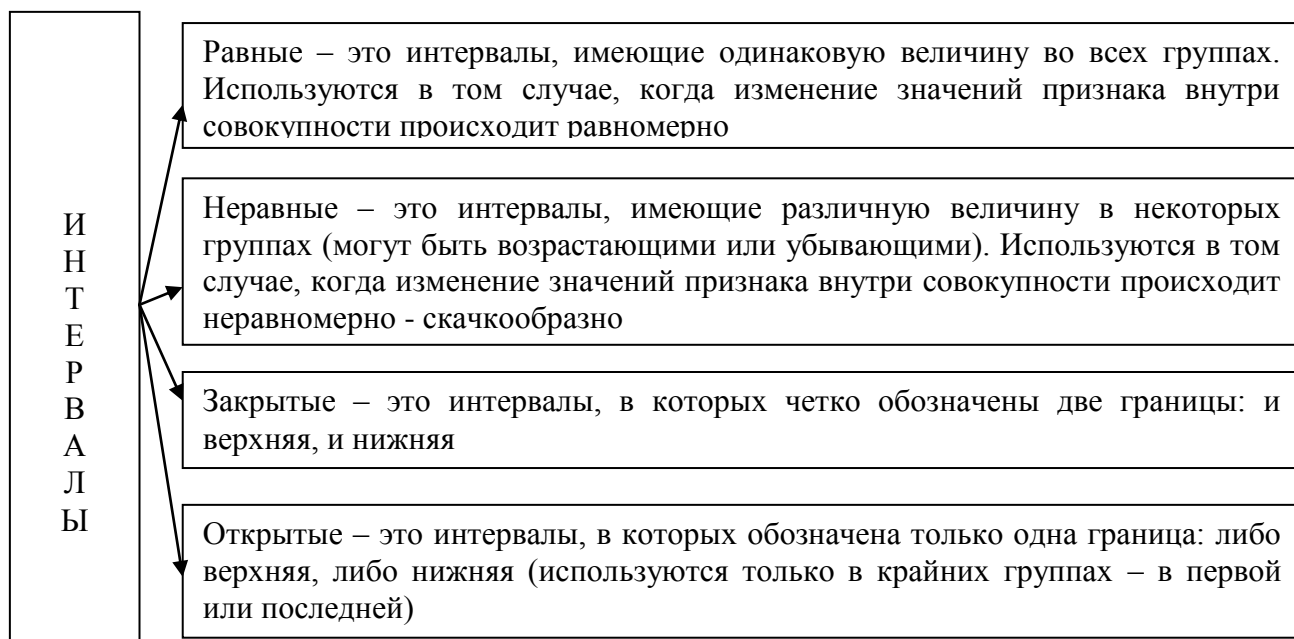


Рисунок 25 – Виды интервалов в статистических интервальных рядах распределения

Наиболее часто используются равные интервалы. Величина равного интервала определяется по формуле

$$i = \frac{X_{\max} - X_{\min}}{n}, \quad (59)$$

где X_{\max} – максимальное значение группировочного признака;

X_{\min} – минимальное значение группировочного признака.

Границы групп зависят от характера вариации группировочного признака: 1) в дискретном ряду распределения границы групп могут быть четко обозначены целым числом; 2) в интервальном ряду распределения границы групп могут выражаться дробными числами.

Границы групп определяются следующим образом: нижней границей первой группы является минимальное значение признака (X_{\min}); для определения верхней границы первой группы к нижней ее границе прибавляется величина интервала; верхняя граница первой группы одновременно является нижней границей второй группы; для определения

верхней границы второй группы к ее нижней границе опять прибавляется величина интервала и.т.д.:

$$\text{Первая группа: } X_{\min} + i \text{ (от } X_{\min} \text{ до } X_{\min} + i), \quad (60)$$

$$\text{Вторая группа: } X_{\min} + 2i \text{ (от } X_{\min} + i \text{ до } X_{\min} + 2i), \quad (61)$$

$$\text{Третья группа: } X_{\min} + 3i \text{ (от } X_{\min} + 2i \text{ до } X_{\min} + 3i), \quad (62)$$

$$\text{Четвертая группа: } X_{\min} + 4i \text{ (от } X_{\min} + 3i \text{ до } X_{\min} + 4i), \quad (63)$$

$$\text{Пятая группа: } X_{\min} + 5i \text{ (от } X_{\min} + 4i \text{ до } X_{\min} + 5i). \quad (64)$$

Число единиц в группах определяется подсчетом тех единиц, значение группировочного признака у которых попадает в рассчитанные границы групп.

Третий этап - статистическая сводка. Заключается в следующем: 1) отбор абсолютных показателей, которыми следует охарактеризовать выделенные группы; 2) подсчет общих и групповых итоговых величин по отобранным показателям; 3) оформление статистической сводки в узком смысле в виде таблицы со сводными абсолютными показателями; 4) отбор относительных показателей, которыми следует охарактеризовать выделенные группы; 5) расчет относительных показателей на основании величин итоговых абсолютных показателей; 6) оформление результатов статистической сводки и группировки в виде итоговой статистической таблицы с относительными показателями; 7) формулировка выводов по полученным результатам.

Пример построения статистической группировки по количественному признаку.

Построение статистической аналитической группировки по количественному признаку производится на основании исходных цифровых данных по двадцати сельскохозяйственным предприятиям, полученных в итоге статистического наблюдения (табл. 17). На основе имеющихся цифровых данных, характеризующих каждое сельскохозяйственное предприятие (табл.17), строится аналитическая группировка по факторному признаку – качество почв в баллах. Полученные группы следует охарактеризовать относительными показателями: средним качеством почв и средней урожайностью овощей.

Таблица 17

Качество почв, посевная площадь и валовой сбор овощей по предприятиям области

№	Качество почв, баллы	Валовой сбор овощей, т.	Посевная площадь овощей, га
1	68	1600	100
2	45	600	50
3	45	630	70
4	90	1680	60
5	48	520	40
6	75	1440	80
7	50	910	65
8	51	650	50
9	78	1540	70
10	55	960	60
11	58	700	50
12	70	900	50
13	89	1150	50
14	80	1260	70
15	85	1200	60
16	87	1610	70
17	50	900	60
18	88	1160	40
19	80	967,5	45
20	95	1732,5	55
Сумма	1387	22110	1195

Первый этап – определение группировочного признака. Группировочным признаком является факторный признак – качество почв в баллах. По группировочному признаку строится ранжированный ряд – единицы статистической совокупности располагаются по возрастанию группировочного признака (табл. 18).

Таблица 18

Ранжированный ряд распределения предприятий по качеству почв

№ предприятия	Качество почв, баллы	Валовой сбор овощей, т	Посевная площадь овощей, га
1	2	3	4
1	45	600	50
2	45	630	70
3	48	520	40
4	50	900	60
5	50	910	65
6	51	650	50
7	55	960	60
8	58	700	50
9	68	1600	100

Продолжение таблицы 18

1	2	3	4
10	70	900	50
11	75	1440	80
12	78	1540	70
13	80	1260	70
14	80	967	45
15	85	1200	60
16	87	1610	70
17	88	1160	40
18	89	1150	50
19	90	1680	60
20	95	1732	55

Ранжированный ряд изображается графически в виде огивы Гальтона (рис. 26)

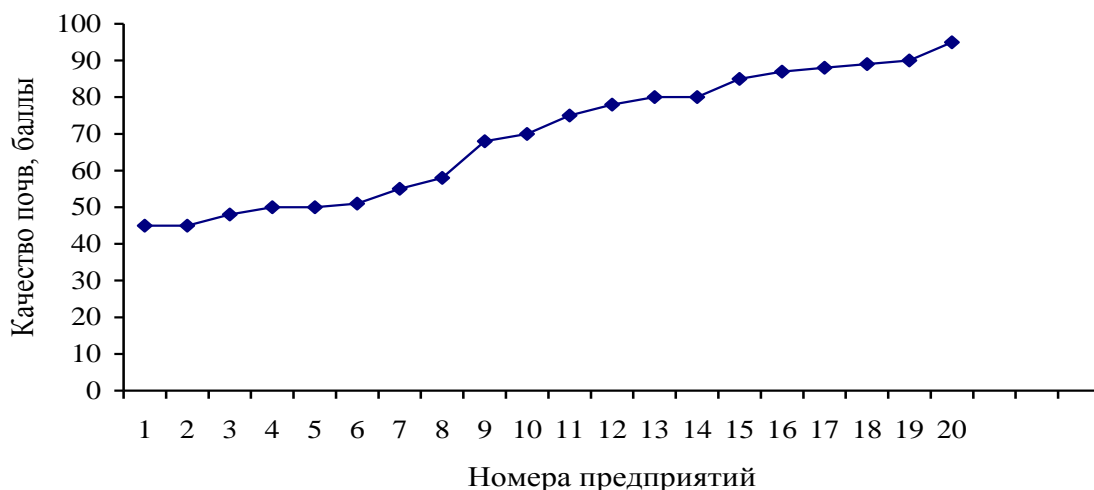


Рисунок 26 - Ранжированный ряд распределения предприятий по качеству почв

Число групп определяется по формуле Стерджесса (58) при $N=20$,
 $\lg N = \lg 20 = 1,3$

Число групп: $n = 1 + 3,322 \lg N = 1 + 3,322 \times 1,3 \approx 5$

Величина интервала определяется по формуле 59:

$$i = \frac{X_{\max} - X_{\min}}{n} = \frac{95 - 45}{5} = 10 \text{ баллов}$$

Границы групп:

- первая группа $45 + 10 = 55$;

- вторая группа $55 + 10 = 65$;
- третья группа $65 + 10 = 75$;
- четвертая группа $75 + 10 = 85$;
- пятая группа $85 + 10 = 95$.

Интервальный ряд распределения строится путем определения частот – путем подсчета единиц совокупности, вошедших в каждую группу (табл. 19).

Таблица 19

Интервальный ряд распределения предприятий по качеству почв

Интервалы по качеству почв, баллы	Число предприятий в группе
От 45 до 55	7
От 55 до 65	1
От 65 до 75	3
От 75 до 85	4
От 85 до 95	5
Всего по совокупности	20

На основании расчетов делается статистическая сводка, для этого строится таблица с общими итоговыми показателями по группам и по совокупности в целом.

В данном случае необходимо обратить внимание на основные требования к построению статистической группировки. Так как в первичной группировке во вторую группу вошло всего одно предприятие, то нарушается принцип массовости. Поэтому следует построить вторичную группировку (сделать перегруппировку ранее сгруппированных данных) способом укрупнения интервалов – объединения первой и второй групп и принятия открытого интервала «до 65» и построения нового интервального ряда распределения (табл. 20).

Таблица 20

Интервальный ряд распределения предприятий по качеству почв

Интервалы по качеству почв, баллы	Число предприятий в группе
До 65	8
От 65 до 75	3
От 75 до 85	4
От 85 до 95	5
Всего по совокупности	20

Интервальный ряд распределения изображается графически в виде гистограммы (рис.27):

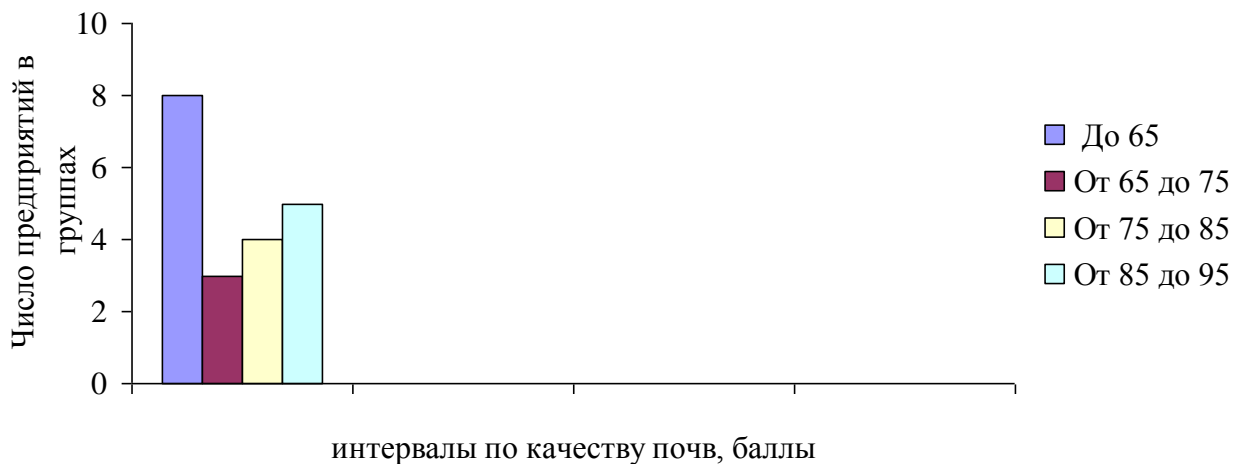


Рисунок 27 - Гистограмма интервального ряда распределения предприятий по качеству почв

На основании расчетов делается статистическая сводка, для этого строится таблица с абсолютными показателями - с общими итоговыми данными по группам и по совокупности в целом. Общие итоговые показатели определяются путем суммирования по группам и по совокупности в целом исходных абсолютных показателей: качества почв, валового сбора овощей, посевной площади овощей (табл. 21).

Таблица 21

Сводные данные по группам предприятий (абсолютные показатели)

Интервалы по качеству почв, баллы	Число предприятий в группе	Качество почв, баллы	Валовой сбор овощей, т.	Посевная площадь овощей, га
До 65	8	402	5870	445
65 – 75	3	213	3940	230
75 – 85	4	323	4967	245
85 – 95	5	449	7332	275
Всего по совокупности	20	1387	22110	1195

На основании сводной таблицы рассчитываются относительные показатели, характеризующие выделенные группы: средняя балльная оценка качества почв, средняя урожайность овощей по формулам:

$$\text{Средняя балльная оценка качества почв} = \frac{\text{Сумма баллов качества почв}}{\text{Число предприятий в группе (по совокупности)}} \quad (65)$$

$$\text{Средняя урожайность овощей} = \frac{\text{Сумма валового сбора овощей}}{\text{Сумма посевных площадей овощей}} \quad (66)$$

Результаты расчетов оформляются в таблице 22.

Таблица 22

Характеристика групп (относительные показатели)

Интервалы по качеству почв, баллы	Число хозяйств в группе	Средняя балльная оценка почв, балл	Средняя урожайность овощей, т/га
До 65	8	50,25	13,19
65 – 75	3	71,00	17,13
75 – 85	4	80,75	20,27
85 – 95	5	89,80	26,66
В среднем по совокупности	20	69,35	18,50

Аналитическая группировка показала, что между средней балльной оценкой почв и средней урожайностью овощей наблюдается прямая зависимость. Так, при увеличении средней балльной оценки почв от первой к четвертой группе средняя урожайность овощей также увеличивается.

6 СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ И ГРАФИКИ КАК ОСНОВНОЙ СПОСОБ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКИХ РАБОТ

6.1 Статистические таблицы как способ представления и обобщения статистического материала. Составные элементы статистических таблиц

Наиболее рациональным способом представления результатов статистической сводки и группировки являются статистические таблицы. Табличная форма изложения результатов статистической сводки и группировки стала использоваться в XVIII веке и впервые применены русским географом и статистиком И.К.Кирилловым. До 1727г. результаты статистического исследования оформлялись в виде громоздких текстовых описаний.

Табличная форма представления цифровой информации – это форма, при которой числовое значение располагается на пересечении четко сформулированного заголовка по вертикальному столбцу и сформулированного названия по соответствующей горизонтальной строке.

Статистическая таблица – это форма рационального изложения цифрового материала, представляющая собой совокупность взаимопересекающихся горизонтальных и вертикальных линий, образующих по горизонтали строки, по вертикали - графы.

Каждая строка и графа статистической таблицы имеют свое наименование, которое соответствует содержанию показателей, отражаемых в таблице. В образовавшиеся внутри таблицы клетки записывается цифровая информация. Цифровой материал может быть представлен абсолютными, относительными и средними величинами.

Значение статистических таблиц заключается в следующем:

- 1) статистические таблицы являются одним из основных результатов статистической деятельности;
- 2) статистическая таблица позволяет охватить материалы статистической сводки и группировки в целом;
- 3) статистическая

таблица является системой мыслей, то есть рассказом об исследуемом объекте и его особенностях, излагаемых в виде цифр на основе определенного порядка расположения систематизированной информации; 4) в статистической таблице наглядно проявляется связь между признаками изучаемого явления; 5) характеризующие в статистической таблице показатели можно объединить под общим заголовком; 6) в статистической таблице дается сводная характеристика статистической совокупности, подводятся итоги.

Статистическая таблица включает следующие основные составные элементы (табл. 23): 1) макет таблицы – составленная, но не заполненная цифровой информацией таблица; 2) общий заголовок таблицы – это название таблицы, которое отражает ее основное содержание, определяет место и время, к которому относятся данные таблицы; 3) подлежащее статистической таблицы – это перечень изучаемых явлений, групп и подгрупп, на которые разделена изучаемая совокупность; 4) сказуемое статистической таблицы (численные значения и характеристики изучаемого явления) – это перечень показателей, с помощью которых характеризуется изучаемое явление, объект (подлежащее таблицы); 5) боковые заголовки – это заголовки подлежащего статистической таблицы; 6) верхние заголовки – это заголовки сказуемого таблицы.

Таблица 23

Название (общий заголовок)

Заголовок подлежащего	Заголовки сказуемого				
	Наименование граф (верхние заголовки)	Наименование граф (верхние заголовки)	Наименование граф (верхние заголовки)	...	Итоговая графа
1	2	3	4	5	6
Наименование строк (боковые заголовки)					
Наименование строк (боковые заголовки)					
...					
Итоговая строка					

* Примечания к таблице

6.2 Виды статистических таблиц

Виды статистических таблиц подразделяются по двум основным классификационным признакам (рис. 28).

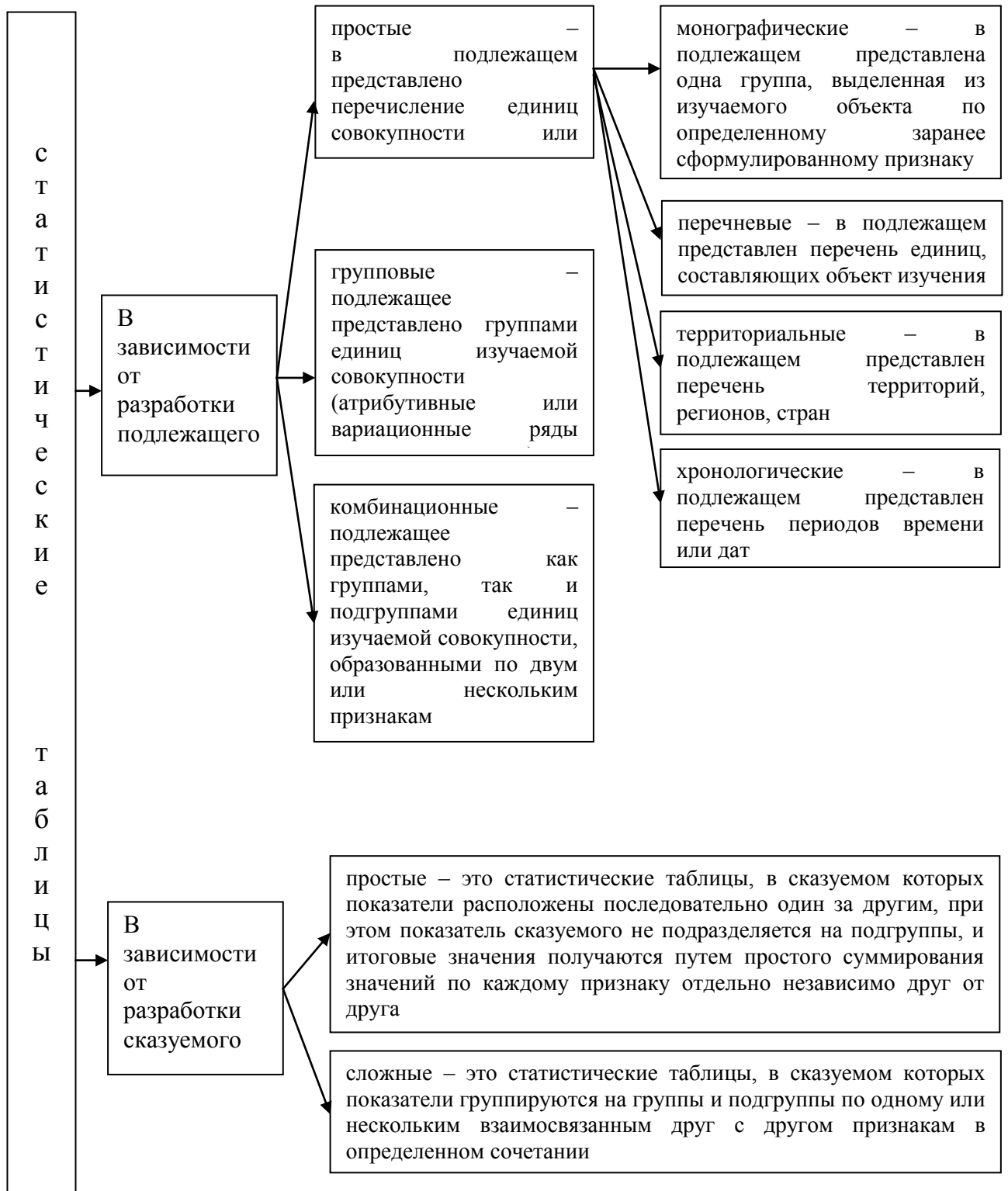


Рисунок 28 – Основные виды статистических таблиц

В статистическом анализе, наряду с основными перечисленными видами статистических таблицы, в зависимости от содержания и назначения используются также следующие виды:

1) матрица – это прямоугольная таблица числовой информации, которая состоит из m -количества строк и n -количества столбцов. Матрица является основой матричного моделирования и используется для изучения взаимосвязей между изучаемыми явлениями (составление балансово-нормативных моделей соотношения результатов производства и производственных затрат; построение межотраслевого баланса, системы национальных счетов и др.). Матрица бывает двух видов: а) прямоугольная – с размерностью $m \times n$; б) квадратная – число строк равно числу столбцов. При этом, если все элементы квадратной матрицы, стоящие вне главной диагонали, равны нулю, то она называется диагональной; если все элементы диагонали равны единице, – единичной, если все элементы диагонали равны нулю, – нулевой;

2) таблица сопряженности – это таблица, которая содержит сводную числовую характеристику изучаемой совокупности по двум и более атрибутивным (качественным) признакам или комбинации количественных и атрибутивных признаков (используются при изучении социальных явлений и процессов – общественного мнения, уровня и образа жизни населения, общественно-политического строя и др.).

6.3 Правила построения статистических таблиц

В целях наиболее эффективного представления статистической информации и результатов статистического исследования, построение статистических таблиц должно осуществляться с соблюдением следующих правил.

1. Таблица должна составляться компактной, небольшой, легко обозримой и содержать только те исходные данные, которые непосредственно отражают изучаемое явление.

2. Цифровой материал следует располагать таким образом, чтобы при анализе статистической таблицы сущность изучаемого явления раскрывалась чтением строк слева направо и сверху вниз.

3. Общий заголовок таблицы должен четко и кратко выражать ее основное содержание, указывать место и время, к которым относятся представленные данные, единицы измерения, если они относятся ко всем показателям, представленным в таблице.

4. Графы таблицы должны иметь четкие и лаконичные заголовки, при этом, если названия отдельных граф повторяются между собой, содержат повторяющиеся термины или несут единую смысловую нагрузку, то им следует присваивать общий объединяющий заголовок.

5. Слова в статистической таблице не сокращаются, кроме общепринятых сокращений (единиц измерения и пр.).

6. При отсутствии общей единицы измерения, относящейся ко всем представленным показателям, в каждой строке или графе должны быть единицы измерения, соответствующие каждому статистическому показателю.

7. Таблица должна содержать итоговые строки и графы.

8. Графы и строки таблицы нумеруются способом последовательной нумерации.

9. В графах и строках таблицы может быть приведен порядок расчета взаимосвязанных показателей (например: $гр. 3 = гр.2 : гр.1$).

10. Для всех числовых значений в таблице в пределах одной и той же строки или графы должна соблюдаться одинаковая степень точности, т.е. все числовые значения должны иметь одинаковое число знаков после запятой (если после запятой знаков нет, ставится «0»).

11. Если данные в таблице заимствованы, то под таблицей следует указать источник исходной информации.

12. К таблице может быть дано примечание, в котором раскрывается методика расчета показателей, а также даются подтверждающие пояснения по тем показателям, которые могут вызвать вопросы или показаться сомнительными.

13. При заполнении статистических таблиц используются следующие условные обозначения: «-» - явление отсутствует; «...» - нет информации; «нет сведений» - отсутствуют данные; «X» - значение признака не имеет смысла; «0,0» - информация имеет очень малое числовое значение, меньшее принятой в таблице точности.

6.4 Чтение и анализ статистических таблиц

Чтение и анализ статистических таблиц осуществляется в определенной последовательности.

Чтение статистической таблицы предполагает 1) усвоение содержания таблицы; 2) формулировку первоначальных суждений об изучаемом объекте; 3) уяснение назначения таблицы; 4) понятие содержания таблицы в целом; 5) оценку явления или процесса, отраженного таблицей.

Анализ статистических таблиц, как метод статистического исследования, путем разбивки предмета изучения на части включает два приема: 1) структурный анализ; 2) содержательный анализ.

Структурный анализ предполагает анализ строения таблицы, характеристику представленных в таблице следующих элементов: 1) изучаемой совокупности и единиц, которые ее формируют; 2) признаков и их комбинаций, формирующих подлежащее и сказуемое таблицы; 3) признаков количественных и атрибутивных; 4) соотношения признаков подлежащего с показателями сказуемого; 5) вида таблицы: простая или сложная (групповая или комбинационная); 6) решаемых задач – анализ структуры, типов явлений или их взаимосвязей.

Содержательный анализ предполагает изучение внутреннего содержания таблицы, которое включает 1) анализ отдельных групп подлежащего по соответствующим признакам сказуемого; 2) выявление соотношения и пропорций между группами явлений по одному и разным признакам; 3) сравнительный анализ и формулировку выводов по отдельным группам и по всей совокупности в целом; 4) установление закономерностей и определение резервов развития изучаемого явления.

Анализ статистических таблиц включает следующие элементы:

- 1) анализ данных таблицы - производится по каждому признаку в отдельности, затем – в логическом сочетании всей совокупности признаков в целом;
- 2) анализ отдельных признаков и групп - начинается с изучения абсолютных величин, затем – связанных с ними относительных и средних величин. При этом числовые характеристики рассматриваются в динамике за весь период изучения;
- 3) Построение статистических графиков на основании данных статистических таблиц - для получения более полного и наглядного представления об изучаемых явлениях и процессах.

Для правильного анализа статистических таблиц проверяется достоверность и научная обоснованность представленной в ней числовой информации, достоверность и надежность источника статистической информации. Указанные аспекты достигаются при помощи двух видов проверки: 1) логическая проверка; 2) счетная проверка.

Логическая проверка заключается в возможности определения конкретных признаков определенным числовым значением (например, нелогично, если средний возраст студентов очного обучения составит 60 лет). Счетная проверка заключается в выборочном расчете значений отдельных взаимосвязанных показателей, отдельных значений признаков по группе, либо итоговых значений строк или граф.

6.5 Статистические графики как способ представления статистического материала. Значение графического способа в статистике

Полученный в результате статистического наблюдения, сводки, группировки и анализа статистический материал нуждается в наглядном изображении при помощи построения статистических графиков.

Статистический график – это наглядное изображение статистических величин при помощи геометрических линий, фигур и знаков. То есть статистический график – это чертеж, на котором при помощи условных геометрических фигур изображаются статистические данные.

Статистические графики, наряду со статистическими таблицами, являются важнейшим средством выражения, представления и анализа статистических данных. Этим объясняется широкое применение статистических графиков для популяризации статистического материала.

Значение статистических графиков:

- 1) статистические графики облегчают восприятие статистических данных;
- 2) статистические графики позволяют охватить всю совокупность изучаемых явлений;
- 3) статистические графики позволяют выявить наиболее характерные соотношения и связи между изучаемыми явлениями;
- 4) статистические графики позволяют выявить основные тенденции развития изучаемых явлений;
- 5) статистические графики делают статистический материал более доходчивым, понятным и запоминающимся;
- 6) статистические графики являются методом обобщения статистической информации и дают новое знание об изучаемом явлении;
- 7) статистические графики позволяют осуществить контроль достоверности статистических показателей, так как представленные на графике

они более ярко отражают имеющиеся неточности, связанные либо с наличием ошибок наблюдения, либо с сущностью изучаемого явления;

8) статистические графики позволяют выявить причинно-следственные связи между изучаемыми явлениями, особенно на стадии установления первоначальных статистических гипотез;

9) статистические графики широко используются для отражения состава изучаемых явлений, их изменения во времени и пространстве.

При построении статистического графика следует выполнять следующие требования: 1) статистический график должен быть наглядным, так как смысл графического изображения как метода статистического анализа заключается в наглядном изображении статистических показателей; 2) статистический график должен быть выразительным, доходчивым и понятным; 3) статистический график должен включать всю совокупность обязательных составляющих его элементов.

6.6 Основные элементы статистического графика

Статистический график содержит следующие основные обязательные элементы (рис. 29).

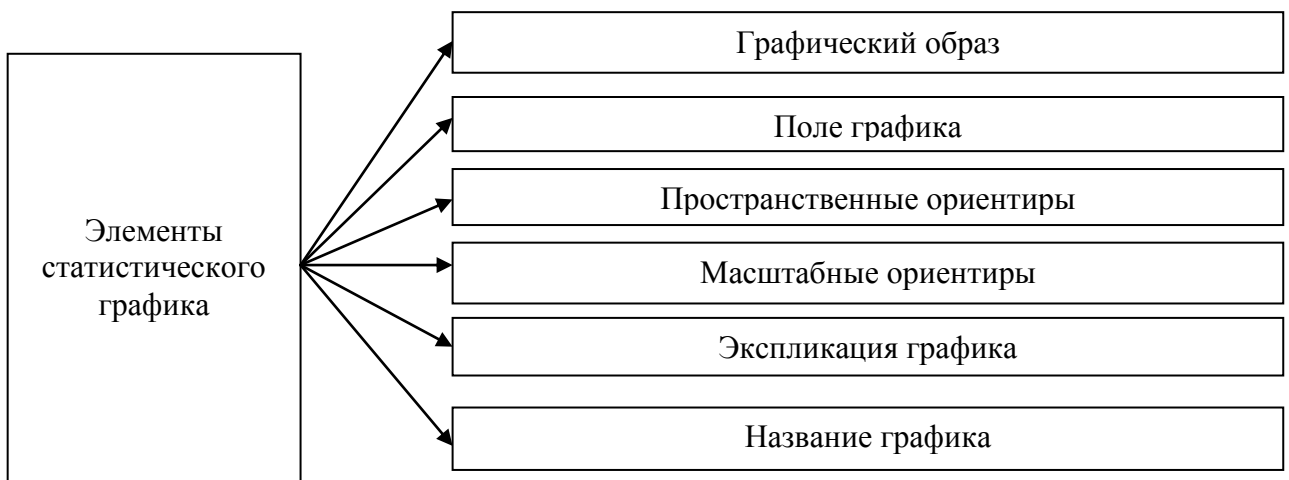


Рисунок 29 – Составные элементы статистического графика

Графический образ – это основа любого статистического графика, которая включает в себя геометрические знаки, совокупность точек, линии, фигуры и так далее, то есть – символические знаки, с помощью которых изображаются статистические величины. Графический образ характеризует язык статистического графика. Одни и те же статистические данные можно изобразить при помощи различных графических образов. Поэтому при построении статистического графика важен правильный выбор графического образа.

Поле графика – это место, на котором строится график (лист бумаги, географическая карта и т.д.). Поле графика характеризуется размером и пропорциями. Размер поля зависит от назначения графика. Пропорция поля – это соотношение его сторон (наиболее оптимальным для зрительного восприятия является график, выполненный на поле прямоугольной формы с соотношением сторон от 1:1,3 до 1:1,5, либо с равными сторонами).

Пространственные ориентиры – это параметры, которые определяют размещение графических образов на поле графика. Пространственные ориентиры задаются координатной сеткой или контурными линиями и делят поле графика на части, соответствующие значениям изучаемых показателей. В статистических графиках чаще всего применяется система прямоугольных (декартовых) координат.

Масштабные ориентиры – это параметры, которые придают графическим образам количественную характеристику, которая определяется при помощи системы масштабных шкал. Масштабные ориентиры включают масштаб графика и масштабную шкалу.

Масштаб графика – это условная мера перевода числовой величины в графическую величину (например: в 1 см – 1 кг). Чем длиннее отрезок линии, принятой за числовую единицу (например: в 1 см – 10 кг, 100 кг и т.д.), тем крупнее масштаб графика.

Масштабная шкала – это линия, отдельные точки которой могут быть прочитаны в соответствии с принятым масштабом графика, как определенные

значения статистических показателей. В статистических графиках обычно применяются прямолинейные масштабные шкалы, которые располагаются по осям координат.

Масштабные шкалы бывают равномерные и неравномерные (в статистике обычно используются равномерные масштабные шкалы). Равномерная масштабная шкала – это шкала, в которой равным графическим отрезкам соответствуют равные числовые величины. Масштабом равномерной шкалы является длина отрезка (графический интервал), принятого за единицу измерения в каких-либо мерах. Неравномерная масштабная шкала – это шкала, в которой равным графическим отрезкам соответствуют неравные числовые величины.

Правильность построения статистического графика зависит от правильности определения масштабной шкалы, которая зависит от выполнения следующих требований: шкала, по которой отсчитываются уровни изучаемых показателей, должна начинаться с 0; последнее число, наносимое на масштабную шкалу, должно немного превышать максимальный уровень изучаемого показателя; допускается разрыв масштабной шкалы в случае изображения статистических данных, имеющих значения только в определенных пределах. То есть, если на каком-то участке данные отсутствуют, то масштабная шкала может быть прервана и дальше продолжаться уже с того значения, которое имеется. При этом применяется условное обозначение – разрыв масштабной шкалы - \approx .

Экспликация графика – это пояснение содержания графика, которое включает в себя объяснение масштабных шкал; условные обозначения графического образа; условные обозначения отдельных элементов графического образа, которые могут быть расположены как в поле графика, так и под ним, но они должны быть размещены до названия графика.

Название графика – это краткое и четкое пояснение основного содержания изображаемых на графике данных, указание сущности отражаемого показателя, места и времени, к которому относятся отражаемые данные.

Графическое изображение это всегда рисунок, следовательно, статистический график всегда оформляется как рисунок и его название располагается под графиком.

6.7 Виды графических изображений и способы их построения

Статистические графики делятся на виды по следующим классификационным признакам: по способу построения; по характеру решаемых задач; по форме применяемого графического образа.

Классификация статистических графиков по способу построения и характеру решаемых задач представлена на рисунке 30.

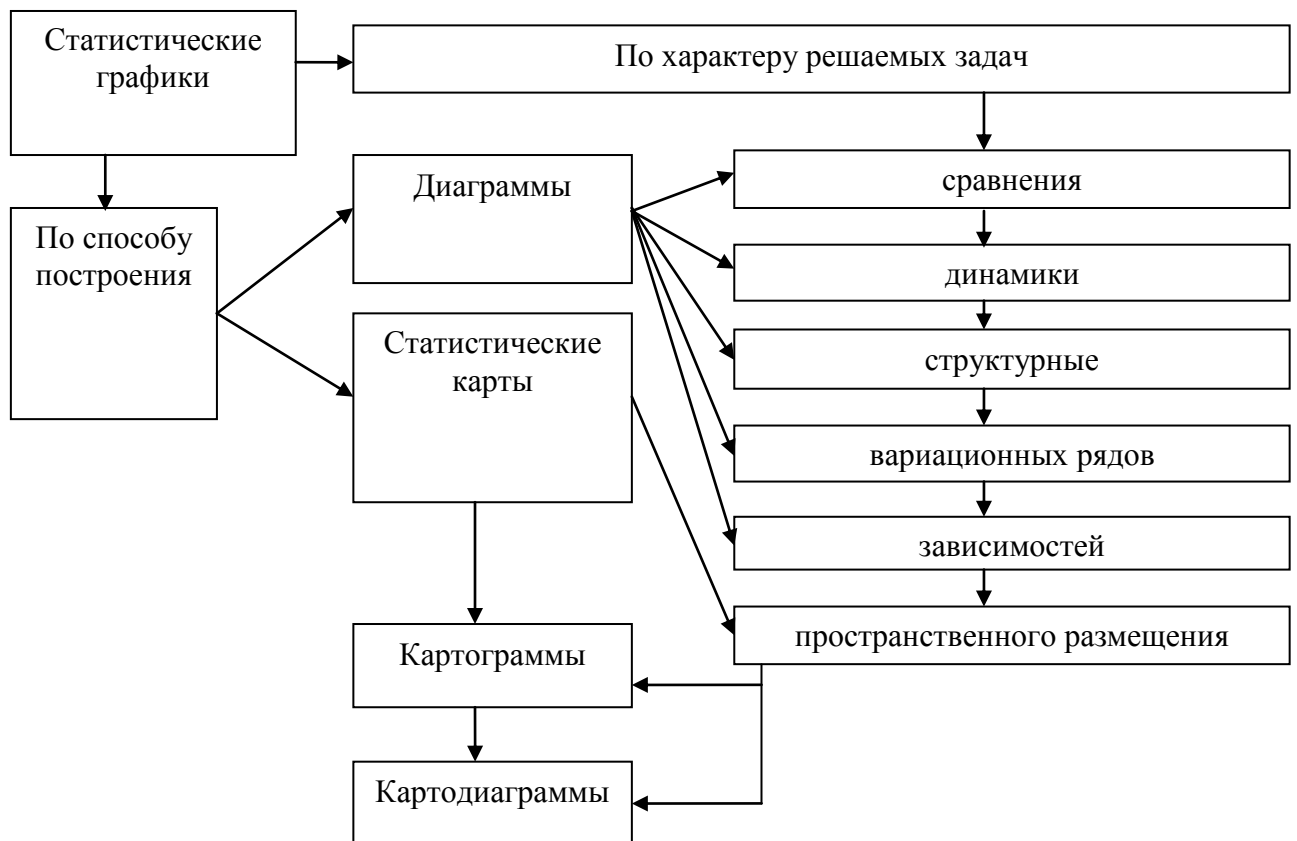


Рисунок 30 - Виды статистических графиков по способу построения и характеру решаемых задач

Диаграмма – это статистический график, на котором статистическая информация изображается посредством геометрических линий, фигур или символов. В зависимости от используемых геометрических фигур и символов диаграммы бывают следующих видов:

Линейная – это диаграмма, имеющая вид линии (статистической кривой), проходящей через пересечения координатной шкалы. При построении линейной диаграммы применяется система прямоугольных координат (на оси абсцисс (X) откладываются варианты изучаемого показателя или показатели времени, по оси ординат (Y) откладываются частоты или величины изучаемого показателя. Преимуществом линейной диаграммы является то, что на одном и том же поле графика можно изобразить несколько показателей и сравнить их взаимное развитие (рис.31).

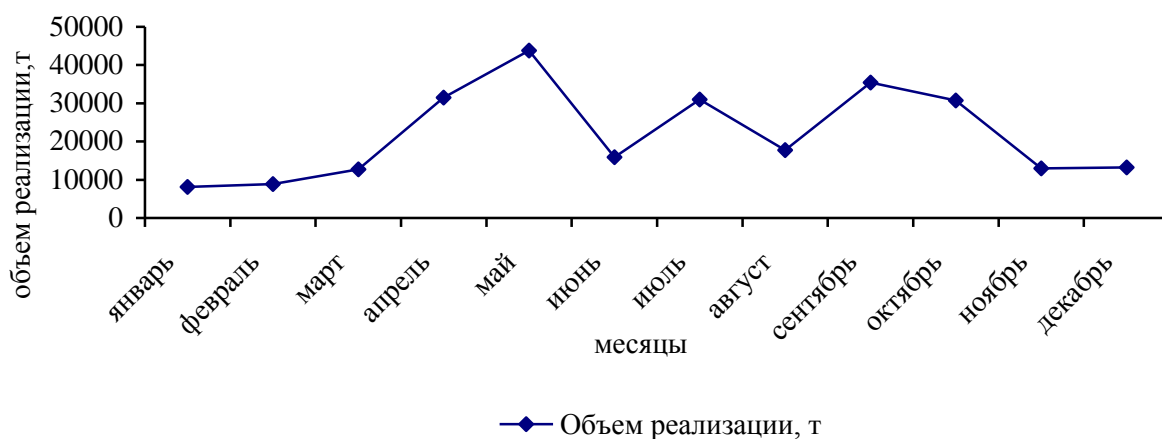


Рисунок 31 - Динамика объема реализации нефтепродуктов за 2014г.

Столбиковая – это диаграмма, которая имеет вид вытянутых по вертикали прямоугольников. При построении столбиковой диаграммы используется система прямоугольных координат. По оси абсцисс (X) откладываются основание столбиков, по оси ординат (Y) откладываются изображаемые статистические величины в соответствии с принятым масштабом (рис.32,33). Количество столбиков по оси X в диаграмме определяется числом изучаемых показателей. Высота столбика по оси Y должна начинаться с 0; шкала по оси Y должна быть непрерывной; основания столбиков должны быть равны между

собой; столбики могут быть размещены на одинаковом расстоянии друг от друга, вплотную друг к другу или наплывом; у основания столбиков дается название изучаемого показателя или показатель времени; величины значений изображаемых показателей помещаются внутри столбиков.

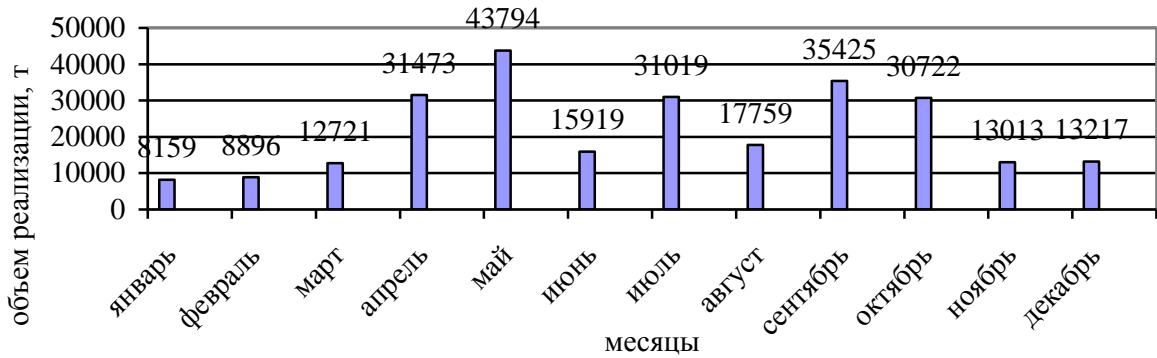


Рисунок 32 - Динамика объема реализации нефтепродуктов за 2014г.

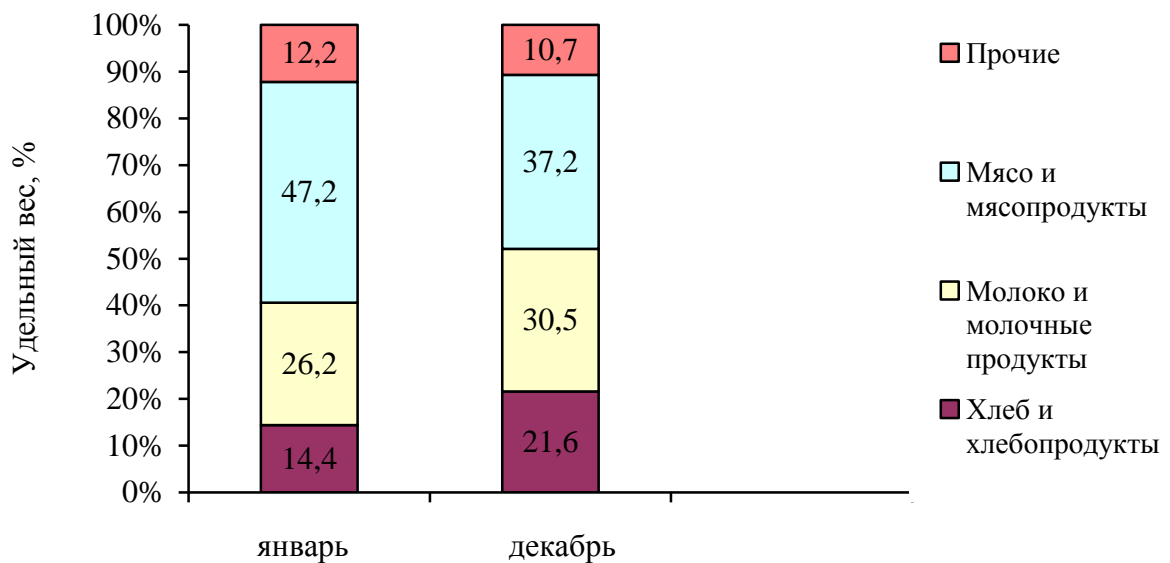


Рисунок 33 - Структура товарооборота магазина продуктов в январе и декабре 2014 г.

Ленточная – это диаграмма, которая имеет вид вытянутых по горизонтали прямоугольников. Специфика построения ленточной диаграммы аналогична построению столбиковой диаграммы, только основания прямоугольников

располагаются по вертикали (по оси Y), а масштабная шкала – по горизонтали (по оси X). Ширина полос ленточной диаграммы должна быть одинаковой, длина полос соответствует величине изображаемых показателей (рис. 34,35).

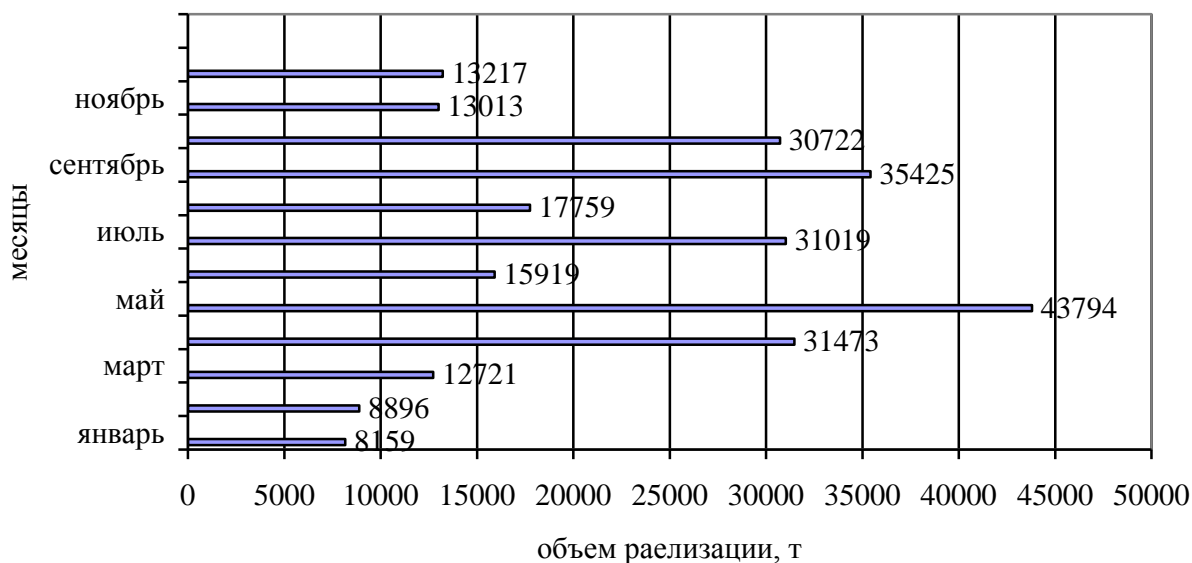


Рисунок 34 - Динамика объема реализации нефтепродуктов за 2014г.

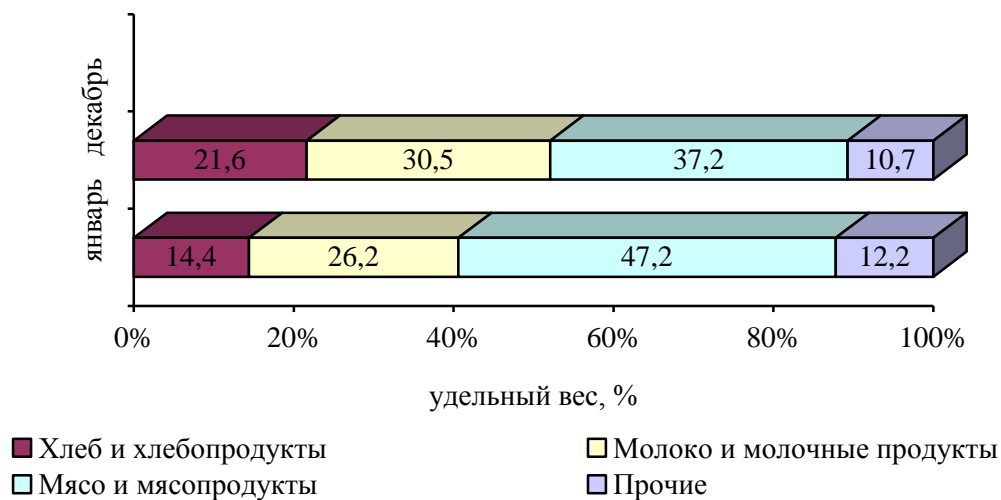


Рисунок 35 - Структура товарооборота в январе и декабре 2014 г.

Квадратная – это диаграмма, которая имеет вид квадрата. Величина изображаемого показателя представлена площадью квадрата. Чтобы изобразить квадратную диаграмму, необходимо из изучаемых статистических величин

извлечь квадратные корни и построить квадраты со сторонами, равными полученным результатам.

Круговая – это диаграмма, которая имеет вид круга. Круговая диаграмма строится аналогично квадратной диаграмме, при этом изображаемый показатель отражается площадью круга, пропорциональной квадратным корням из изображаемых величин.

Радиальная – это диаграмма, которая строится на базе полярных осей координат (например, для изображения повторяющихся годовых циклов с помесечными или поквартальными данными). Началом отсчета служит центр окружности, а масштабные шкалы располагаются на радиусах круга. Радиальные диаграммы бывают замкнутыми и спиральными (рис.36).

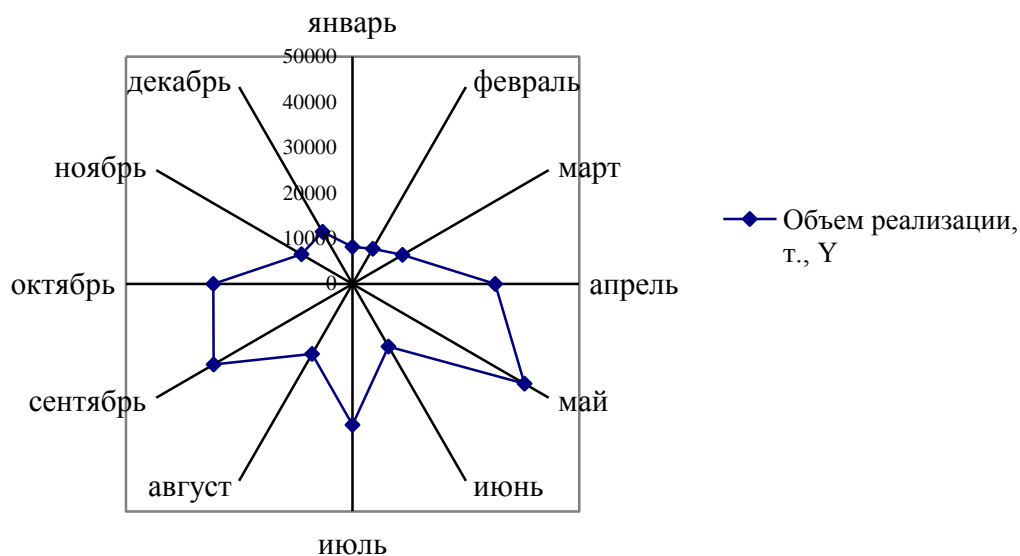


Рисунок 36 - Динамика объема реализации нефтепродуктов за 2014г.

Фигурная – это диаграмма, которая имеет вид определенных рисунков или символов, которые в наибольшей степени соответствуют существу изображаемых явлений.

Секторная – это диаграмма, которая имеет вид круга, разделенного на части – сектора (используется обычно при изображении структуры изучаемого явления). При построении секторной диаграммы площадь круга принимается за величину всего изучаемого явления, а площади отдельных секторов круга отображают долю его составных частей пропорционально их удельному весу (рис.37).

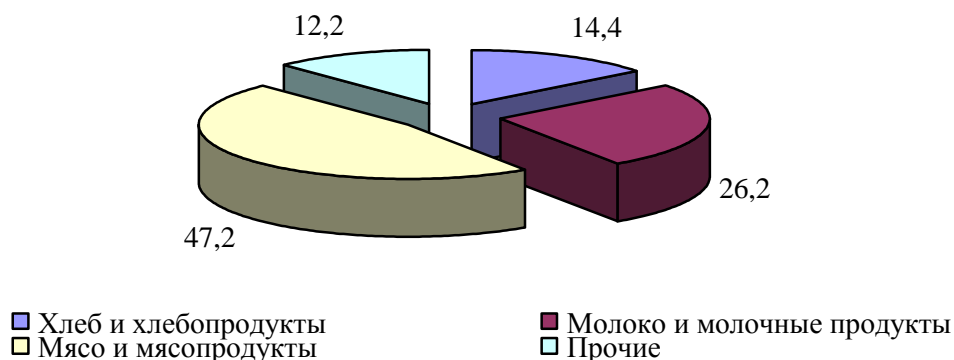


Рисунок 37 - Структура товарооборота в январе 2014 г., %

Знак Варзара – это диаграмма, имеющая вид прямоугольника, у которого один сомножитель принят за основание прямоугольника, другой сомножитель – за высоту прямоугольника, а вся площадь прямоугольника равна произведению двух сомножителей. Диаграмма знак Варзара используется для графического изображения сопоставления трех изучаемых величин. При этом, три изучаемые статистические величины взаимосвязаны между собой таким образом, что одна величина является произведением двух других величин.

Полигон – это диаграмма, имеющая вид ломаной линии, отрезки которой соединяют точки пересечения значений вариант и частот (по оси X откладываются значения вариант, по оси Y откладываются значения частот). Полигон используется для графического изображения дискретного ряда распределения (рис.21).

Гистограмма – это диаграмма, которая имеет вид ступенчатой фигуры, состоящей из прямоугольников, основаниями которых являются значения

вариант, а высотами – значения частот. Гистограмма используется для графического изображения интервального ряда распределения (рис.22).

Кумулята (кумулятивная кривая) – это диаграмма, имеющая вид ломаной линии, отрезки которой соединяют точки пересечения значений вариант и значений накопленных частот (итогов частот). Кумулята используется для графического изображения вариационных рядов с накопленными частотами (рис 23).

Кривая концентрации (кривая Лоренца) – это диаграмма, имеющая вид ломаной линии, отрезки которой соединяют значения нарастающих итогов значений признака и значения накопленных частот (рис. 38). При этом, по оси Х откладываются значения накопленных частот, по оси У откладываются нарастающие итоги значений признака. Кривая Лоренца может быть преобразована в график Лоренца, представляющая собой квадрат, в котором на осях координат расположены масштабные шкалы от 0 до 100%. Диагональ квадрата соответствует нормальному распределению и отсутствию концентрации. Чем больше кривая Лоренца отклоняется от диагонали нормального распределения, тем больше неравномерность распределения, и, следовательно, выше концентрация.

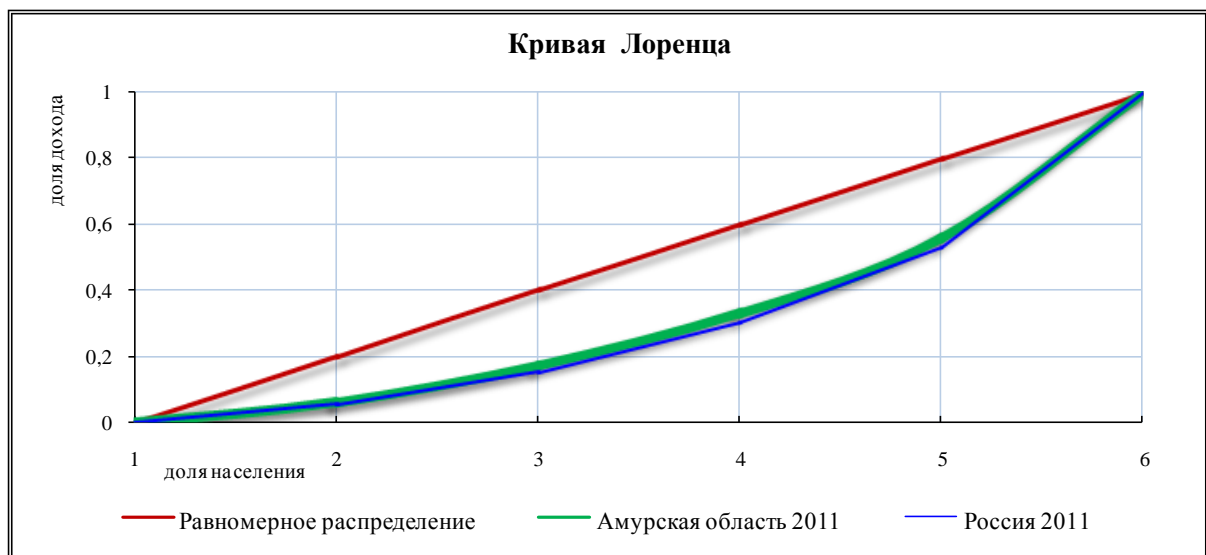


Рисунок 38 – Кривая Лоренца распределения населения по уровню доходов [16]

Статистические карты – это статистические графики, нанесенные на контурные географические карты. Статистические карты бывают двух видов: картограммы; картодиаграммы.

Картограмма – это схематическая контурная карта, либо план местности, на которой отдельные территории в зависимости от величины изображаемого показателя обозначаются при помощи графических символов (штриховки, точек и т.д.) (рис. 39). Картограммы бывают следующих видов: фоновая – это картограмма, в которой территории с различной величиной изучаемого показателя имеют различную штриховку; точечная – это картограмма, в которой в качестве графического символа используются точки одинакового размера, размещенные в пределах определенных территорий, при этом, чем больше точек на данной территории, тем больше значение признака.



Рисунок 39 – Картограмма размещения населения РФ в 2013г. [15]

Картодиаграмма – это сочетание контурной карты и диаграммы. Принцип построения картодиаграммы заключается в том, что на контурной карте составные части какой-либо диаграммы размещаются на площади определенной территориальной единицы.

По характеру решаемых задач статистические графики делятся на следующие виды: графики сравнения; графики структуры; графики динамики; графики пространственного размещения; графики вариационных рядов распределения; графики зависимости.

Графики сравнения решают задачу отображения сопоставления двух или нескольких показателей во времени и пространстве. К ним относятся столбиковые, ленточные, квадратные, круговые и фигурные диаграммы, а также отдельные картограммы и картодиаграммы.

Графики структуры решают задачу отображения состава изучаемого явления и структурных сдвигов. К ним относятся секторные диаграммы, а также столбиковые, ленточные, полосовые и другие, в которых показано разделение изучаемого явления на составные части.

Графики динамики решают задачу отображения изменения изучаемого явления во времени. К ним относятся все виды линейных диаграмм.

Графики пространственного размещения решают задачу отображения пространственной распространенности изучаемого явления. К ним относятся все виды картограмм и картодиаграмм.

Графики вариационных рядов распределения решают задачу отображения различного вида рядов распределения. К ним относятся гистограмма, полигон, кумулята, огива, кривая концентрации.

Графики зависимости решают задачу отображения степени взаимосвязи между двумя или несколькими изучаемыми признаками. К ним относятся все виды точечных диаграмм.

Классификация статистических графиков по форме применяемого графического образа представлена на рисунке 40.

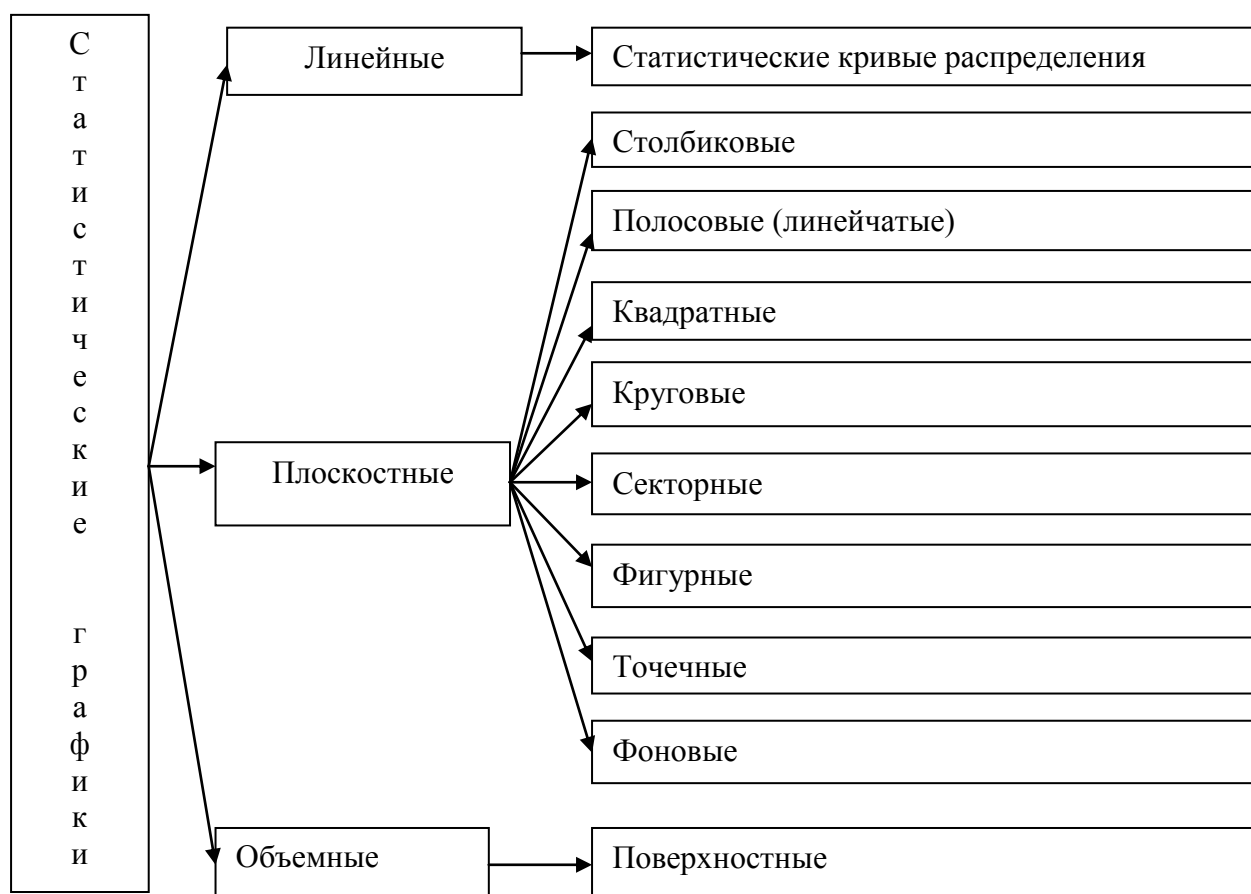


Рисунок 40 - Классификация статистических графиков по форме применяемого графического образа

Точечный – это график, в котором в качестве графического образа применяется совокупность точек (например: график корреляционной зависимости) (рис.41).

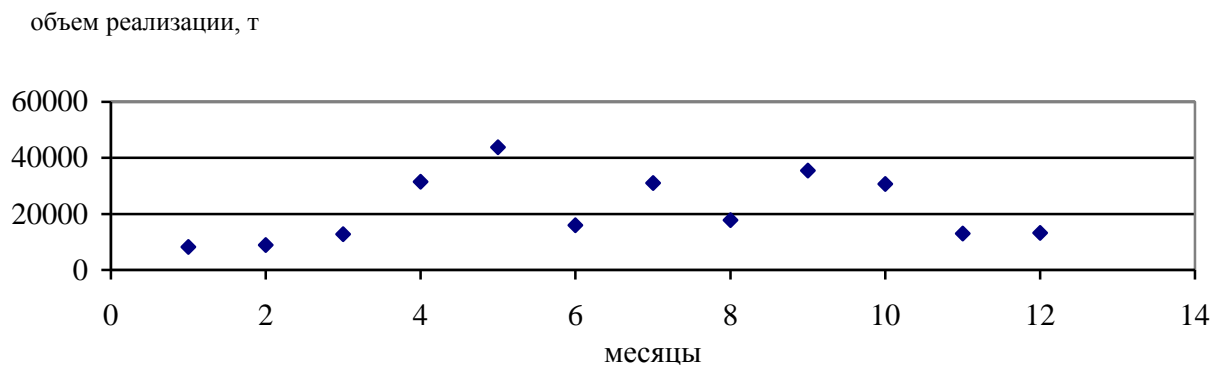


Рисунок 41 - Зависимость объема реализации нефтепродуктов от времени (месяца) за 2014г.

Линейный – это график, в котором в качестве графического образа применяется геометрическая линия (рис.31, 38).

Плоскостной – это график, в котором в качестве графического образа применяются геометрические – прямоугольники, столбики, линейки, полосы, квадраты, круги, точки, штриховка и не геометрические фигуры (рис. 32, 33, 34, 35, 36, 37 и др.).

Объемный – это график, в котором в качестве графического образа применяются объемные поверхностные распределения (рис.42).

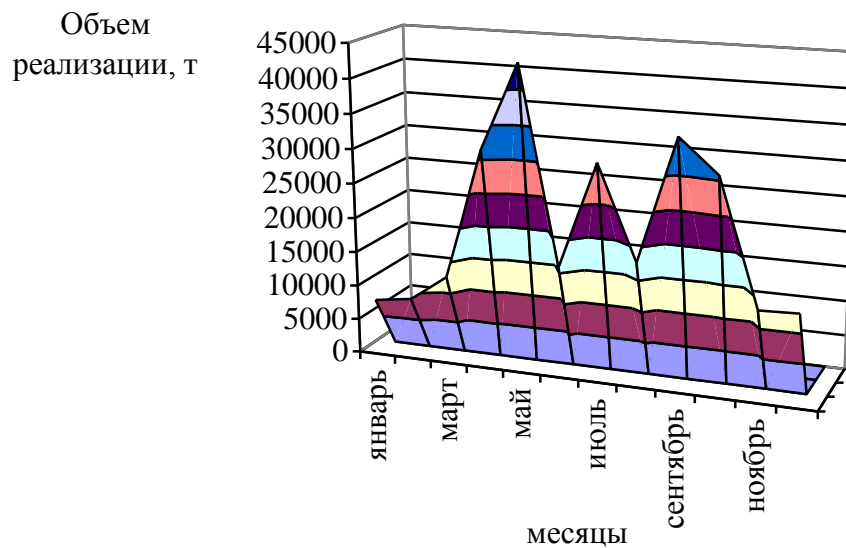


Рисунок 42 - Динамика объема реализации нефтепродуктов в 2014г.

7 СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА РЯДОВ ДИНАМИКИ. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ АНАЛИЗА РЯДОВ ДИНАМИКИ

7.1 Сущность, составные элементы, задачи и виды рядов динамики

Динамика – это изменение явления во времени. Основная цель статистического изучения динамики заключается в выявлении и измерении закономерности развития явления во времени. Эта цель достигается посредством построения и анализа статистических рядов динамики.

Ряд динамики в статистике – это ряд последовательно расположенных в хронологическом порядке количественных показателей, которые характеризуют изменение явления во времени.

Ряд динамики состоит из двух элементов: показателя времени (условное обозначение - t) и уровня ряда динамики (условное обозначение – Y).

1. Показатель времени t это – момент по состоянию на который определяется изучаемый показатель, либо период, за который определяется изучаемый показатель. Показатель времени может выражаться конкретной датой, либо интервалом времени – месяцем, годом и т.д.

2. Уровень ряда динамики Y – соответствующий количественный показатель, характеризующий состояние явления на определенный момент времени, либо развитие явления за определенный период времени. Уровень ряда динамики может выражаться абсолютными, относительными и средними величинами.

При построении и анализе рядов динамики в статистике решаются следующие задачи:

1. Характеристика интенсивности отдельных изменений изучаемого явления внутри ряда динамики от уровня к уровню путем расчета индивидуальных показателей ряда динамики.

2. Характеристика изменения изучаемого явления в целом за весь период, включенный в анализ путем расчета средних показателей ряда динамики.
3. Выявление закономерности и основной тенденции развития изучаемого явления во времени при помощи выравнивания ряда динамики и трендового анализа.
4. Прогнозирование изменения изучаемого явления на перспективу при помощи определения точечных и интервальных прогнозных значений.
5. Выявление степени влияния факторов, которые обуславливают изменение изучаемого явления при помощи корреляционно-регрессионного анализа в рядах динамики.

Ряды динамики делятся на виды по следующим классификационным признакам (рис. 43).

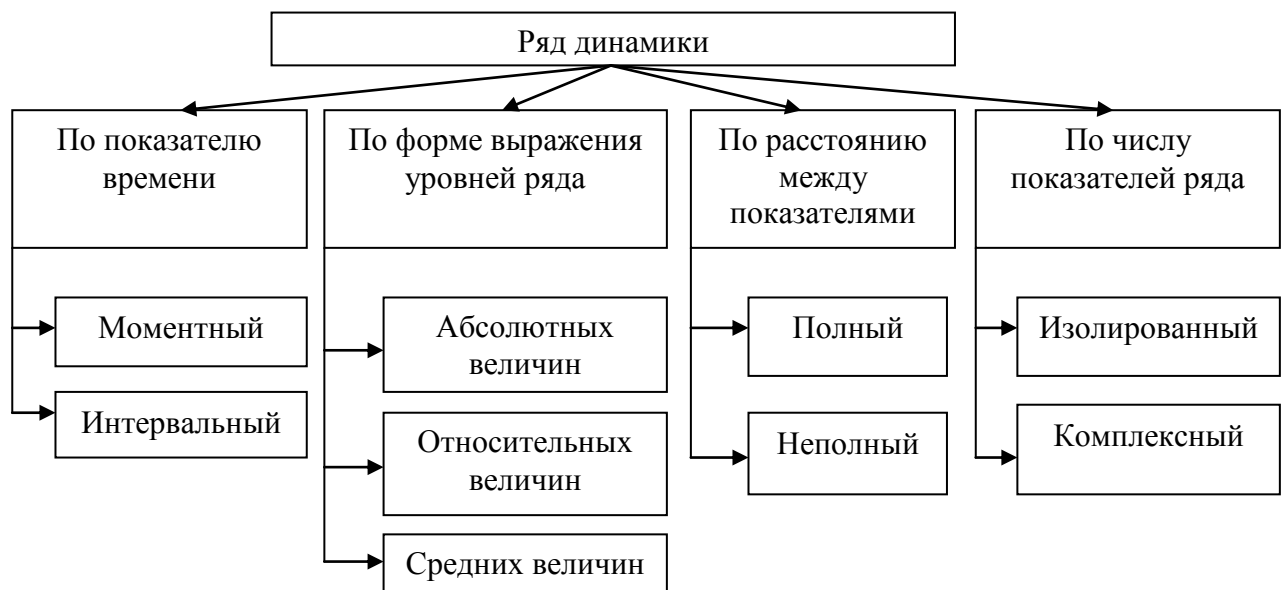


Рисунок 43 – Основные виды рядов динамики

Деление рядов динамики на моментные и интервальные зависит от содержания самого изучаемого показателя. Если изучаемый показатель – моментный (численность населения, остаток денежных средств на вкладах в банке, остаток материальных оборотных средств на складе и др.), то и ряд динамики, который отражает изменение данного показателя – моментный. Если изучаемый показатель интервальный (объем производства продукции за год,

сумма затрат на производство продукции, сумма выручки от реализации продукции и др.), то и ряд динамики, отражающий изменение данного показателя – интервальный.

Моментный ряд динамики показывает наличие изучаемого явления на определенный момент времени – конкретную дату (ряды динамики численности населения – на 1 января каждого года, величины запасов оборотных средств – на 1 число каждого месяца, суммы остатков денежных средств на вкладах в банке – на 1 число каждого месяца, поголовье сельскохозяйственных животных – на 1 января каждого года и др.) (табл. 24).

Таблица 24

Динамика численности населения Амурской области [16]

На 1 января года, t	Численность населения, тыс.чел., Y
2004	894
2005	888
2006	881
2007	874
2008	869
2009	864
2010	835
2011	829
2012	822
2013	817
2014	811

Особенность моментного ряда динамики заключается в том, что в его уровни входят одни и те же единицы изучаемой совокупности, поэтому при суммировании уровней моментного ряда динамики возникает повторный счет и сумма уровней моментного ряда динамики не имеет реального содержания. Особенность (табл. 24) проявляется следующим образом: например, в численность населения на 1 января 2012г. входит и часть населения, бывшая в 2011г.; на 1 января 2013г. – входит часть населения, бывшая в 2011г. т в 2012г. и т.д.

Интервальный ряд динамики показывает результат, накопленный или вновь произведенный за определенный период (интервал) времени (ряды динамики объема произведенной продукции, числа родившихся, объема

товарооборота торгового предприятия, суммы выданных кредитов банком и др. в целом за месяц, квартал, год и т.д.) (табл. 25).

Таблица 25

Динамика числа родившихся в Амурской области [16]

Годы, t	Число родившихся, чел., Y
2004	11092
2005	11006
2006	10661
2007	11807
2008	11512
2009	11218
2010	11479
2011	11211
2012	11740
2013	11453

Особенность интервального ряда динамики заключается в том, что каждый его уровень складывается из величин, образованных за более короткие промежутки времени, и путем суммирования уровней интервального ряда динамики получают новые ряды динамики за укрупненные периоды. Сумма уровней интервального ряда динамики имеет реальное содержание. Особенность проявляется следующим образом: сумма числа родившихся за 2004 – 2013г.г. характеризует общее число родившихся за изучаемый период.

Деление рядов динамики на ряды динамики абсолютных величин, ряды динамики относительных величин и ряды динамики средних величин зависит от формы выражения самих изучаемых показателей.

При этом ряды динамики абсолютных величин (табл. 24, 25), как правило, являются исходными, а ряды динамики относительных (табл. 26) и средних величин – производными.

Таблица 26

Динамика уровня рождаемости в Амурской области [16]

Годы, t	Коэффициент рождаемости, промилле, Y
1	2
2004	10,0
2005	12,4
2006	12,4
2007	12,1

Продолжение таблицы 26

1	2
2008	13,5
2009	12,9
2010	13,8
2011	13,6
2012	14,3
2013	14,1

Деление рядов динамики на полные и неполные зависит от наличия исходной информации.

Полные – это ряды динамики с равно отстоящими во времени уровнями, в которых даты или периоды времени следуют один за другим с равными промежутками (выше представленные ряды динамики были полными).

Неполные – это ряды динамики с неравно отстоящими во времени уровнями, в которых имеются разрывы между показателями времени и принцип равных промежутков не соблюдается.

Статистическому изучению, как правило, подвергаются полные ряды динамики. Если исходный ряд динамики является неполным, то для статистического анализа он преобразуется в полный методом условных расчетов недостающих уровней.

Деление рядов динамики на изолированные и комплексные зависит от цели анализа. Изолированные – это ряды динамики, которые характеризуют изменение во времени только одного показателя. Комплексные – это ряды динамики, которые характеризуют изменение во времени нескольких взаимосвязанных между собой показателей.

Выше представленные ряды динамики в отдельности являются изолированными.

Если их объединить в одну статистическую таблицу, то получится комплексный ряд динамики, который характеризует изменение во времени взаимосвязанных демографических показателей (табл. 27).

Комплексный ряд динамики [16]

Годы, t	Численность населения, тыс.чел., У	Число родившихся, чел., Y	Коэффициент рождаемости, промилле, Y
2004	894	11092	10,0
2005	888	11006	12,4
2006	881	10661	12,4
2007	874	11807	12,1
2008	869	11512	13,5
2009	864	11218	12,9
2010	835	11479	13,8
2011	829	11211	13,6
2012	822	11740	14,3
2013	817	11453	14,1

7.2 Правила построения рядов динамики

Достоверность результатов анализа изменения изучаемого явления во времени зависит от правильности построения исходного ряда динамики. Все правила построения рядов динамики можно объединить в четыре группы: 1.Периодизация. 2. Сопоставимость исходных данных. 3. Соответствие величин временных интервалов интенсивности изменения изучаемых явлений. 4.Упорядочение числовых уровней во времени.

Периодизация – это разделение всего периода изучения на однородные этапы, в пределах которых изменение изучаемого явления подчиняется единому закону развития. Периодизация производится тремя методами:

1) исторический метод. При историческом методе разделение изучаемого периода на однородные этапы осуществляется на основе значимых дат и событий (социалистическая экономика, переходная экономика, рыночная экономика и др.);

2) метод параллельной периодизации. При методе параллельной периодизации изменение изучаемого показателя, по которому может отсутствовать исходная информация, ставится в зависимость от изменения

другого показателя, который тесно взаимосвязан с изучаемым, по которому информация есть в полном объеме;

3) метод многомерного статистического анализа. При методе многомерного статистического анализа используется несколько взаимосвязанных показателей, определяется степень их зависимости и в пределах изучаемого периода производится типологическая группировка.

Сопоставимость исходных данных обеспечивается по следующим параметрам:

1) по территории. Сопоставимость по территории достигается тем, что данные по объектам, границы которых изменились, условно пересчитываются в рамки старых границ (если поменялись границы региона в результате объединения более мелких, то в динамике показатель изучается в границах региона до объединения);

2) по совокупности. Сопоставимость по совокупности достигается тем, что данные по совокупностям, численность которых изменилась, условно пересчитываются в рамки старой совокупности с предыдущим числом единиц. Это производится при помощи смыкания рядов динамики путем пересчета абсолютных уровней ряда динамики в относительные, либо в условные абсолютные уровни. (если увеличились размеры предприятия путем включения в него других подразделений, то анализ динамики объема производства продукции производится с учетом изменения числа подразделений предприятия).

Например, в первый год предприятие работало в прежних размерах, во втором году число торговых точек увеличилось, в третьем году предприятие работало уже в новых размерах. Для получения сопоставимых данных определяется коэффициент соотношения уровней $K = 630 / 450 = 1,4$. Полученный коэффициент умножается на объем товарооборота первого года, получается недостающий уровень $Y = 432 \times 1,4 = 604,8$. В результате получается новый ряд динамики при новых размерах торгового предприятия (табл. 28).

Объем розничного товарооборота торгового предприятия

Показатели	1-й год	2-й год	3-й год
Исходные размеры предприятия	432	450	-
Новые размеры предприятия	-	630	622,5
Сопоставимый ряд	6004,8	630	622,5

3) по времени регистрации. Сопоставимость по времени регистрации достигается путем регистрации исходной информации об изучаемых явлениях с учетом критерия минимальной мобильности;

4) по единицам измерения. Сопоставимость по единицам измерения достигается путем пересчета изучаемых показателей с разными единицами измерения в единую размерность, либо в условные единицы;

5) по ценам. Сопоставимость по ценам достигается путем пересчета стоимостных показателей в сопоставимые цены – в условные, либо в цены фиксированного периода;

6) по методологии расчета. Сопоставимость по методологии расчет достигается путем соблюдения единой методики расчета показателей за разные периоды времени.

Соответствие величины временных интервалов интенсивности изменения изучаемого показателя заключается в том, что чем больше вариация изучаемого показателя во времени, тем короче должны быть разрывы между показателями времени. Это достигается тем, что для менее мобильных явлений устанавливаются более крупные промежутки между уровнями ряда динамики, а для более мобильных явлений устанавливаются короткие промежутки между уровнями ряда динамики (перепись населения проводится через десятилетние промежутки времени, анализ урожайности сельскохозяйственных культур проводится через промежутки в 1 год, анализ цен производится через промежутки в 1 месяц и т.д.).

Упорядочение числовых уровней во времени заключается в том, что не допускаются неравномерные разрывы между уровнями ряда динамики. Для

проведения статистического анализа необходимо неполные ряды динамики преобразовывать в полные путем условных расчетов методом интерполяции.

7.3 Индивидуальные показатели ряда динамики

Первая задача построения и анализа рядов динамики заключается в характеристике интенсивности отдельных изменений изучаемого явления внутри ряда динамики от уровня к уровню путем расчета индивидуальных показателей ряда динамики.

Индивидуальные показатели ряда динамики получаются путем сравнения его уровней. В зависимости от способа сравнения индивидуальные показатели ряда динамики определяются в двух вариантах:

- 1) на постоянной базе сравнения – базисные показатели;
- 2) на переменной базе сравнения – цепные показатели.

Схематично расчет цепных и базисных индивидуальных показателей ряда динамики можно представить следующим образом (рис. 44).

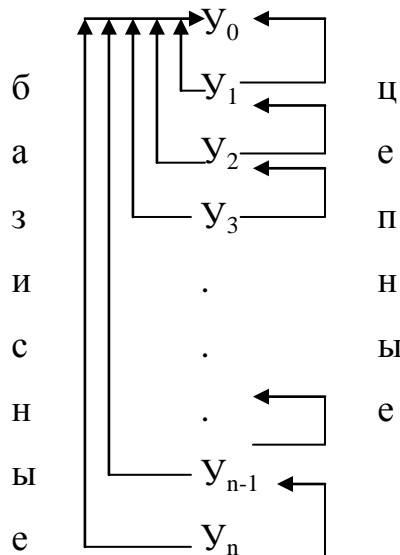


Рисунок 44 - Схема расчета индивидуальных показателей ряда динамики

При расчете индивидуальных показателей ряда динамики на постоянной базе сравнения (базисных показателей) каждый последующий уровень ряда динамики сравнивается с одним и тем же, принятым за базу сравнения. Как правило за базу сравнения принимается начальный – первый уровень ряда динамики, либо уровень, с которого начинается новый этап в развитии изучаемого явления. Базисные показатели ряда динамики характеризуют окончательные результаты всех изменений в уровнях ряда динамики от базисного уровня до анализируемого i -того уровня.

При расчете индивидуальных показателей ряда динамики на переменной базе сравнения (цепных показателей) каждый последующий уровень ряда динамики сравнивается с предыдущим. Цепные показатели ряда динамики характеризуют интенсивность изменения изучаемого явления от уровня к уровню ряда.

К индивидуальным показателям ряда динамики относятся:

1. Абсолютный прирост. 2. Коэффициент роста. 3. Темп роста. 4. Темп прироста. 5. Абсолютное значение 1% прироста. 6. Темп наращивания.

Абсолютный прирост – это разница между каждым текущим уровнем ряда динамики и предыдущим уровнем или базисным уровнем. Условное обозначение - Δ . Определяется по формулам:

- базисный абсолютный прирост

$$\Delta_{\sigma} = Y_i - Y_{\sigma} ; \quad (67)$$

- цепной абсолютный прирост

$$\Delta_{\eta} = Y_i - Y_{i-1} ; \quad (68)$$

где Y_i – текущий уровень ряда динамики,

Y_{σ} – базисный (начальный) уровень ряда динамики,

Y_{i-1} – предыдущий уровень ряда динамики.

Абсолютный прирост показывает, на сколько единиц каждый текущий уровень ряда динамики больше или меньше предыдущего уровня либо базисного уровня. Критерием оценки абсолютного прироста является его знак.

Если абсолютный прирост имеет положительный знак, то он характеризует рост (увеличение) изучаемого явления, если абсолютный прирост имеет отрицательный знак, то он характеризует уменьшение (снижение) изучаемого явления.

Между базисными и цепными абсолютными приростами существует следующая взаимосвязь: сумма цепных абсолютных приростов равна базисному абсолютному приросту последнего уровня ряда динамики

$$\Sigma \Delta_{\text{ц}} = \Delta_{\text{б}(n)} \quad (69)$$

Коэффициент роста – это отношение между каждым текущим уровнем ряда динамики и предыдущим уровнем или базисным уровнем. Условное обозначение – K_p . Определяется по формулам

- базисный коэффициент роста

$$K_{p(\text{б})} = \frac{Y_i}{Y_{\text{б}}} ; \quad (70)$$

- цепной коэффициент роста

$$K_{p(\text{ц})} = \frac{Y_i}{Y_{i-1}} ; \quad (71)$$

Коэффициент роста показывает во сколько раз каждый текущий уровень ряда динамики больше либо меньше предыдущего уровня или базисного уровня. Критерием оценки коэффициента роста является единица (1). Если коэффициент роста больше 1, то он характеризует рост (увеличение) изучаемого явления, если коэффициент роста меньше 1, то он характеризует снижение (уменьшение) изучаемого явления.

Между цепными и базисными коэффициентами роста существует следующая взаимосвязь: произведение цепных коэффициентов роста дает базисный коэффициент роста последнего уровня

$$K_{\text{ц}1} \times K_{\text{ц}2} \times K_{\text{ц}3} \times \dots \times K_{\text{ц}n-1} = K_{\text{б}n}. \quad (72)$$

Темп роста – это коэффициент роста, выраженный в процентах (%). Условное обозначение – T_p . Определяется по формулам

- базисный темп роста

$$T_{p(\text{б})} = K_{p(\text{б})} \times 100 = \frac{Y_i}{Y_{\text{б}}} \times 100 ; \quad (73)$$

- цепной темп роста

$$T_{p(u)} = K_{p(u)} \times 100 = \frac{Y_i}{Y_{i-1}} \times 100. \quad (74)$$

Темп роста показывает, сколько процентов (%) составляет каждый текущий уровень ряда динамики к предыдущему уровню или базисному уровню. Критерием оценки темпа роста является 100%. Если темп роста больше 100%, то он характеризует рост (увеличение) изучаемого явления, если темп роста меньше 100%, то он характеризует снижение (уменьшение) изучаемого явления.

Темп прироста – это разница между темпом роста и 100%. Условное обозначение – $T_{пр}$. Определяется по формулам

- базисный темп прироста

$$T_{np(b)} = T_{p(b)} - 100 \quad (75)$$

или

$$T_{np(b)} = \frac{\Delta i_b}{Y_b} \times 100; \quad (76)$$

- цепной темп прироста

$$T_{np(u)} = T_{p(u)} - 100 \quad (77)$$

или

$$T_{np(u)} = \frac{\Delta i_u}{Y_{i-1}} \times 100. \quad (78)$$

Темп прироста показывает, на сколько процентов каждый текущий уровень ряда динамики больше либо меньше предыдущего уровня или базисного уровня. Критерием оценки темпа прироста является его знак. Если темп прироста имеет положительный знак, то он характеризует рост (увеличение) изучаемого явления, если темп прироста имеет отрицательный знак, то он характеризует снижение (уменьшение) изучаемого явления.

Абсолютное значение 1% прироста (снижения) – это сотая доля от предыдущего уровня ряда динамики. Рассчитывается только по цепному варианту. Условное обозначение – 1%_{пр}. Определяется по формулам

$$1\%_{np} = \frac{Y_{i-1}}{100} \quad (79)$$

или

$$1\%_{np} = \frac{\Delta u}{T_{np(u)}}. \quad (80)$$

Абсолютное значение 1% прироста (снижения) показывает, сколько абсолютных единиц содержится в 1% прироста, либо снижения изучаемого показателя.

Темп наращивания – это отношение абсолютных приростов к уровням ряда динамики. Условное обозначение – T_n . Определяется только по цепному варианту по формуле

$$T_{n(y)} = \frac{\Delta i(y)}{Y_{\bar{o}}} \times 100. \quad (81)$$

Темп наращивания показывает развитие или затухание интенсивности изменения изучаемого явления. Критерием оценки темпа наращивания является 100%. Если темп наращивания больше 100%, то он характеризует развитие интенсивности изменения изучаемого явления, если темп наращивания меньше 100%, то - затухание интенсивности изменения изучаемого явления.

Пример расчета индивидуальных показателей ряда динамики. На основании выше представленных формул, расчет индивидуальных показателей ряда динамики представлен в таблице 29.

Таблица 29

Динамика численности населения Амурской области [16]

На 1 января года, t	Численность населения, тыс.чел., Y	Абсолютный прирост, тыс.чел.		Темп роста, %		Темп прироста, %		Абсолютное значение 1 % прироста, тыс.чел.
		базисный	цепной	базисный	цепной	базисный	цепной	
2004	894	-	-	-	-	-	-	-
2005	888	-6,00	-6,00	99,33	99,33	-0,67	-0,67	8,94
2006	881	-13,00	-7,00	98,55	99,21	-1,45	-0,79	8,88
2007	874	-20,00	-7,00	97,76	99,21	-2,24	-0,79	8,81
2008	869	-25,00	-5,00	97,20	99,43	-2,80	-0,57	8,74
2009	864	-30,00	-5,00	96,64	99,42	-3,36	-0,58	8,69
2010	835	-59,00	-29,00	93,40	96,64	-6,60	-3,36	8,64
2011	829	-65,00	-6,00	92,73	99,28	-7,27	-0,72	8,35
2012	822	-72,00	-7,00	91,95	99,16	-8,05	-0,84	8,29
2013	817	-77,00	-5,00	91,39	99,39	-8,61	-0,61	8,22
2014	811	-83,00	-6,00	90,71	99,26	-9,29	-0,74	8,17
ИТОГО	9383	-	-84,00	-	-	-	-	-

Расчет показал, что, например, на 1 января 2012г. численность населения составила 822 тыс.чел. По сравнению с 2004г. численность населения уменьшилась на 72 тыс.чел., а по сравнению с 2011г. уменьшилась на 7 тыс.чел. При этом, к численности населения 2004г. численность населения 2012г. составила 91,95%, а к численности населения 2011г. она составила 99,16%, то есть наблюдается снижение численности населения по сравнению с 2004г. на 8,05%, а по сравнению с 2011г. – на 0,84%. В каждом проценте данного снижения содержится 8,29 тыс.чел.

7.4 Средние показатели рядов динамики

Вторая задача построения и анализа рядов динамики заключается в характеристике интенсивности изменения изучаемого явления в целом за весь период, включенный в анализ путем расчета средних показателей ряда динамики.

Средние показатели ряда динамики являются обобщающими характеристиками индивидуальных показателей ряда динамики. Особенность расчета средних показателей ряда динамики с целью их достоверности заключается в том, что они могут быть определены только за период со стабильными условиями развития. Если изучаемый период включает в себя разнородные промежутки, то средние показатели следует рассчитывать отдельно по каждому промежутку, в которых соблюдается принцип стабильности развития.

К средним показателям ряда динамики относятся: 1) средний уровень ряда динамики; 2) средний абсолютный прирост; 3) средний коэффициент роста; 4) средний темп роста; 5) средний темп прироста.

Средний уровень ряда динамики – это обобщающая характеристика исходных уровней ряда динамики. Условное обозначение - \bar{Y} .

В зависимости от вида изучаемого ряда динамики, средний уровень определяется в четырех вариантах:

1) средний уровень интервального ряда динамики с равно отстоящими во времени уровнями определяется по формуле средней арифметической простой

$$\bar{Y} = \frac{\sum Y}{n}, \quad (82)$$

где $\sum Y$ - сумма исходных уровней интервального ряда динамики;

n – число уровней интервального ряда динамики.

2) средний уровень интервального ряда динамики с неравно отстоящими во времени уровнями определяется по формуле средней арифметической взвешенной

$$\bar{Y} = \frac{\sum Y \times t}{\sum t}, \quad (83)$$

где Y – уровни ряда динамики;

t – частота (статистический вес) – число одинаковых уровней ряда динамики.

3) средний уровень моментного ряда динамики с равно отстоящими во времени уровнями определяется по формуле средней хронологической простой

$$\bar{Y} = \frac{0,5 \times Y_1 + Y_2 + Y_3 + \dots + Y_{n-1} + 0,5 \times Y_n}{n - 1}. \quad (84)$$

где $Y_1, Y_2, Y_3, \dots, Y_{n-1}, Y_n$ – уровни моментного ряда динамики;

n – число уровней моментного ряда динамики (число дат).

4) средний уровень моментного ряда динамик с неравно отстоящими во времени уровнями определяется по формуле средней хронологической взвешенной

$$\bar{Y} = \frac{(Y_1 + Y_2) \times t_1 + (Y_2 + Y_3) \times t_2 + \dots + (Y_{n-1} + Y_n) \times t_{n-1}}{2 \times \sum t}, \quad (85)$$

где t – число моментов времени (дат), в которых уровни остаются неизменными.

Средний уровень ряда показывает, сколько единиц составляет изучаемый показатель в среднем за весь период изучения.

Средний абсолютный прирост – это обобщающая характеристика индивидуальных абсолютных приростов. Условное обозначение - $\bar{\Delta}$.

В зависимости от вида исходной информации средний абсолютный прирост определяется в двух вариантах:

1) если имеет исходная информация об уровнях ряда динамики, то на основании исходных уровней ряда динамики средний абсолютный прирост определяется по формуле

$$\bar{\Delta} = \frac{Y_n - Y_1}{n - 1}, \quad (86)$$

где Y_n – конечный уровень ряда динамики;

Y_1 – начальный уровень ряда динамики;

n – число уровней ряда динамики.

2) если имеется исходная информация об индивидуальных цепных абсолютных приростах изучаемого показателя, то на основании индивидуальных цепных абсолютных приростов средний абсолютный прирост определяется по формуле

$$\bar{\Delta} = \frac{\sum \Delta_{\eta}}{n - 1} \quad (87)$$

или

$$\bar{\Delta} = \frac{\Delta_{n(\bar{o})}}{n - 1}, \quad (88)$$

где Δ_{η} - цепные индивидуальные абсолютные приросты.

$\Delta_{n(\bar{o})}$ - базисный абсолютный прирост последнего уровня.

Средний абсолютный прирост характеризует скорость изменения уровней ряда динамики и показывает, на сколько единиц в среднем от уровня к уровню уменьшается, либо увеличивается изучаемое явление. Критерием оценки среднего абсолютного прироста является его знак. Если средний абсолютный прирост имеет положительный знак, то он характеризует средний рост (увеличение) изучаемого явления, если средний абсолютный прирост имеет

отрицательный знак, то он характеризует среднее уменьшение (снижение) изучаемого явления.

Средний коэффициент роста – это обобщающая характеристика индивидуальных коэффициентов роста. Условное обозначение - \bar{K}_p .

В зависимости от вида исходной информации средний коэффициент роста определяется по формуле средней геометрической в двух вариантах:

1) если имеется исходная информация об уровнях ряда динамики, то на основании исходных уровней ряда динамики средний коэффициент роста определяется по формуле

$$\bar{K}_p = \sqrt[n-1]{\frac{Y_n}{Y_1}}. \quad (89)$$

2) если имеется исходная информация об индивидуальных цепных коэффициентах роста изучаемого показателя, то на основании индивидуальных цепных коэффициентов роста средний абсолютный прирост определяется по формуле

$$\bar{K}_p = \sqrt[n-1]{K_1 \times K_2 \times K_3 \times \dots \times K_n} \quad (90)$$

или

$$\bar{K}_p = \sqrt[n-1]{K_{n(б)}}, \quad (91)$$

где $K_1, K_2, K_3, \dots, K_n$ – индивидуальные цепные коэффициенты роста;

$K_{n(б)}$ – базисный коэффициент роста конечного уровня ряда динамики.

Средний коэффициент роста характеризует интенсивность изменения изучаемого явления и показывает, во сколько раз происходит возрастание, либо убывание изучаемого явления в среднем от уровня к уровню. Критерием оценки среднего коэффициента роста является единица (1). Если $\bar{K}_p > 1$, то он характеризует тенденцию роста изучаемого явления, если $\bar{K}_p < 1$, то он характеризует тенденцию снижения изучаемого явления.

Средний темп роста – это обобщающая характеристика индивидуальных темпов роста. Условное обозначение - \bar{T}_p . Определяется по формуле

$$\bar{T}_p = \bar{K} \times 100. \quad (92)$$

Средний темп роста показывает, сколько процентов (%) в среднем составляет каждый последующий уровень ряда динамики к предыдущему. Критерием оценки среднего темпа роста является 100%. Если средний темп роста больше 100%, то оно показывает тенденцию роста (увеличения) изучаемого явления в среднем за изучаемый период. Если средний темп роста меньше 100%, то он показывает тенденцию уменьшения (снижения) изучаемого явления в среднем за изучаемый период.

Средний темп прироста – это обобщающая характеристика индивидуальных темпов прироста. Условное обозначение - \bar{T}_{np} . Определяется по формуле

$$\bar{T}_{np} = \bar{T}_p - 100. \quad (93)$$

Средний темп прироста показывает, на сколько процентов (%) увеличивается, либо уменьшается изучаемый показатель в среднем от уровня к уровню. Критерием оценки среднего темпа прироста является его знак. Если средний темп прироста имеет положительный знак, то он показывает увеличение (рост) изучаемого явления в среднем от уровня к уровню ряда динамики, если средний темп прироста имеет отрицательный знак, то он показывает уменьшение (снижение) изучаемого явления в среднем от уровня к уровню.

Пример расчета средних показателей ряда динамики на основании данных таблицы 29.

Средний уровень ряда динамики в данном случае рассчитывается по формуле средней хронологической простой, так как представленный ряд динамики является моментным с равноотстоящими во времени уровнями:

$$\bar{y} = \frac{0,5 \times 894 + 888 + 881 + 874 + 869 + 864 + 835 + 829 + 822 + 817 + 0,5 \times 811}{11 - 1} = \frac{8531}{10} = 853,1 \text{ тыс. чел.}$$

Расчет показал, что в среднем ежегодно численность населения составляла 853,1 тыс. чел.

Средний абсолютный прирост

- на основании исходных уровней ряда динамики: $\bar{\Delta} = \frac{811-894}{11-1} = -8,3$ тыс.чел.
- на основании цепных абсолютных приростов: $\bar{\Delta} = \frac{-83}{11-1} = -8,3$ тыс.чел.

Расчет показал, что в среднем ежегодно численность населения снижалась на 8,3 тыс.чел.

Средний коэффициент роста

- на основании исходных уровней ряда динамики: $\bar{K}_p = \sqrt[11-1]{\frac{811}{894}} = 0,9969$
- на основании индивидуальных цепных коэффициентов роста:
 $\bar{K}_p = \sqrt[10-1]{0,9933 \times 0,9921 \times 0,9921 \times 0,9943 \times 0,9942 \times 0,9664 \times 0,9928 \times 0,9916 \times 0,9939 \times 0,9926} = 0,9969$

Расчет показал, что, поскольку значение коэффициента роста меньше единицы, то можно сделать вывод о тенденции снижения численности населения в целом за изучаемый период с 2004г. по 2014г.

Средний темп роста: $\bar{T}_p = 0,9969 \times 100 = 99,69\%$

Расчет показал, что в среднем ежегодно численность населения составляла 99,69% к каждому предыдущему уровню.

Средний темп прироста: $\bar{T}_{np} = 99,69 - 100 = -0,31\%$

Расчет показал, что в среднем ежегодно численность населения снижалась на 0,31%.

7.5 Способы выявления основной тенденции на основе выравнивания рядов динамики. Статистическое моделирование на основе анализа рядов динамики

Третья задача построения и анализа рядов динамики заключается в выявлении закономерности и основной тенденции развития изучаемого явления во времени при помощи выравнивания ряда динамики и трендового анализа.

В некоторых рядах динамики закономерность - тренд (общая тенденция развития явления) явно отображается самими уровнями ряда и прослеживается

визуально при оценке исходного ряда динамики. Более часто встречаются ряды динамики, в которых уровни ряда под влиянием различных факторов изменяются неравномерно (либо в сторону роста, либо в сторону снижения) и общая тенденция развития явления явно не прослеживается. В таких случаях для выявления общей тенденции (тренда) развития явления используются специфические приемы выравнивания рядов динамики. Приемы выравнивания рядов динамики делятся на две группы: 1. Механические. 2. Математические.

Механические приемы основаны на преобразовании исходного ряда динамики в производный ряд, в котором показатели времени представлены более крупными промежутками (если исходный ряд динамики представлен показателями времени величиной в 1 год, то преобразованный ряд динамики представлен показателями времени величиной в 3, 5 и т.д. лет). Уровни ряда динамики в данном случае представлены средними уровнями за укрупненные промежутки времени (трехлетия, пятилетия и т.д.). При этом в средних уровнях происходит сглаживание различий в индивидуальных уровнях, которые обусловлены влиянием случайных факторов. Механические приемы включают два основных приема: 1) прием укрупнения интервалов ряда динамики; 2) прием сглаживания ряда динамики по скользящей средней.

Прием укрупнения интервалов заключается в объединении показателей времени в более крупные промежутки и расчете средних уровней за эти укрупненные промежутки. При этом исходный ряд динамики преобразуется в новый – ряд динамики средних величин, на основании которых возможно проследить основную тенденцию изменения явления (табл. 30).

Прием укрупнения периодов применяется достаточно - в том случае, когда исходный ряд динамики имеет большое число уровней – более 15 уровней, так как, при малом числе уровней исходного ряда динамики, новый преобразованный ряд динамики становится коротким, чем не соблюдается принцип массовости, необходимый для выявления закономерности развития изучаемого явления.

Прием укрупнения интервалов

На 1 января года	Численность населения, тыс.чел. [16], Y	Укрупненные периоды (трехлетия)	Средний уровень численности населения за укрупненные периоды, тыс.чел.
2004	894	2004-2006	887,7
2005	888	2007-2009	869,0
2006	881	2010-2012	828,7
2007	874	-	-
2008	869	-	-
2009	864	-	-
2010	835	-	-
2011	829	-	-
2012	822	-	-
2013	817	-	-
2014	811	-	-

Прием сглаживания ряда динамики по скользящей средней заключается в объединении показателей времени в более крупные промежутки и расчете средних уровней за эти укрупненные промежутки. При этом исходный ряд динамики преобразуется в новый – ряд динамики средних величин, на основании которых возможно проследить основную тенденцию изменения изучаемого явления. В отличие от приема укрупнения интервалов при приеме сглаживания по скользящей средней укрупненные периоды формируются методом скользящего, то есть, - путем последовательного исключения первого уровня ряда и включения следующего (табл. 31).

В результате сглаживания по скользящей средней новый преобразованный ряд динамики получается короче исходного на число уровней:

$$n = l - 1, \quad (94)$$

где l – число уровней периода скользящего.

Прием сглаживания по скользящей средней применяется достаточно часто, так как является более аналитичным. В результате применения данного приема в полученных средних уровнях сглаживается влияние случайных факторов на индивидуальные уровни ряда динамики и получается более

плавная линия тренда, которая характеризует основную закономерность развития изучаемого явления.

Таблица 31

Прием сглаживания по скользящей средней

На 1 января года	Численность населения, тыс.чел. [16], Y	Периоды скользящего (3 - летия)	Середина периода скользящего	Трехлетняя скользящая сумма	Трехлетняя скользящая средняя численность населения, тыс.чел.
2004	894	2004-2006	2005	2663	887,7
2005	888	2005-2007	2006	2643	881,0
2006	881	2006-2008	2007	2624	874,7
2007	874	2007-2009	2008	2607	869,0
2008	869	2008-2010	2009	2568	856,0
2009	864	2009-2011	2010	2528	842,7
2010	835	2010-2012	2011	2486	828,7
2011	829	2011-2013	2012	2468	822,7
2012	822	2012-2014	2013	2450	816,7
2013	817	-	-	-	-
2014	811	-	-	-	-

Математические приемы выравнивания ряда динамики являются наиболее точными и эффективными. Математические приемы заключаются в применении математической функции для отражения основной тенденции развития изучаемого явления. При этом уровни исходного ряда динамики выражаются в виде функции времени

$$\hat{Y}(t) = f(t), \quad (95)$$

где $\hat{Y}(t)$ - выровненные (теоретические) уровни ряда динамики;

$f(t)$ - функция времени – любая адекватная математическая функция.

Определение математической функции производится на основе предварительного анализа закономерностей развития изучаемого явления, которое бывает нескольких типов. Типы развития явления во времени могут быть следующие:

1) равномерное развитие - данному типу соответствуют постоянные абсолютные приросты

$$\Delta_i - const. \quad (96)$$

При равномерном типе развития явления, для отражения основной тенденции используется прямолинейная функция:

$$\hat{Y}(t) = a_0 + a_1 \times t, \quad (97)$$

где a_0 – параметр уравнения – средний уровень ряда (смысловой нагрузки как правило не несет);

a_1 – параметр уравнения - коэффициент регрессии, который показывает направление развития изучаемого явления. Если a_1 – положительный, то наблюдается тенденция роста, если a_1 – отрицательный, то наблюдается тенденция снижения. При этом, величина изменения изучаемого явления в среднем от уровня к уровню ряда динамики отражается числовым значением параметра a_1 ;

2) равноускоренное (равнозамедленное) развитие - данному типу соответствуют постоянные темпы прироста

$$T_{np_i} = const. \quad (98)$$

При равноускоренном (равнозамедленном) типе развития явления используется криволинейная функция – парабола второго порядка

$$\hat{Y}(t) = a_0 + a_1 \times t + a_2 \times t^2, \quad (99)$$

где a_0 и a_1 – аналогично прямолинейной функции;

a_2 – параметр уравнения, который характеризует изменение интенсивности развития явления во времени. Если a_2 - положителен, то он показывает ускорение тенденции, если a_2 – отрицателен, то он показывает замедление тенденции. При этом, существенное значение имеет знак параметров a_1 и a_2 – если знак параметра a_2 совпадает со знаком параметра a_1 , то тенденция, отражаемая параметром a_1 усиливается, и если знак параметра a_2 не совпадает со знаком параметра a_1 , то тенденция, отражаемая параметром a_1 ослабевает.

3) развитие с переменным ускорением (замедлением). Данному типу соответствуют различные абсолютные приросты и темпы прироста. При

развитии с переменным ускорением используется криволинейная функция - парабола третьего порядка

$$\hat{Y}(t) = a_0 + a_1 \times t + a_2 \times t^2 + a_3 \times t^3, \quad (100)$$

где a_0, a_1, a_2 – аналогично предыдущим функциям;

a_3 – параметр уравнения, который характеризует интенсивность развития ускорения (замедления). Если a_3 – положителен, то ускорение интенсивности увеличивается, если a_3 – отрицателен, то ускорение интенсивности тенденции уменьшается. При этом, существенное значение имеет знак параметров a_2 и a_3 – если знак параметра a_2 совпадает со знаком параметра a_3 , то ускорение (замедление), отражаемое параметром a_2 усиливается, и если знак параметра a_2 не совпадает со знаком параметра a_3 , то ускорение (замедление), отражаемое параметром a_2 ослабевает.

4) развитие по экспоненте - данному типу соответствуют постоянные темпы роста

$$T_p - const. \quad (101)$$

При развитии по экспоненте используется уравнение показательной функции

$$\hat{Y}(t) = a_0 \times a_1^t, \quad (102)$$

где a_1 – параметр уравнения, который характеризует темп роста (снижения) изучаемого явления в единицу времени – интенсивность развития. Если a_1 – положителен, то он характеризует рост интенсивности развития явления, если a_1 – отрицателен, то он характеризует снижение интенсивности развития явления. При этом, параметр a_1 может трактоваться как коэффициент эластичности, который в данном случае показывает, на сколько процентов (%) увеличивается, либо уменьшается в среднем от уровня к уровню изучаемое явление с ростом показателя времени на 1 процент (%).

Кроме указанных функций используются любые математические функции

- логарифмическая функция: $\hat{Y}(t) = a_0 + a_1 \times \lg t$; (103)

- степенная функция: $\hat{Y}(t) = a_0 \times t^{a_1}$; (104)

- гиперболическая функция: $\hat{Y}(t) = a_0 + \frac{a_1}{t}$; (105)

- и другие.

Задача исследования заключается в выборе из множества математических функций ту, которая наиболее точно отражает тенденцию изучаемого явления в каждом конкретном случае, то есть – наиболее адекватную линию тренда.

Выбор наиболее адекватной математической функции основан на способе наименьших квадратов, который заключается в достижении минимальности суммы квадратов отклонений фактических (эмпирических) уровней ряда динамики от теоретических (выровненных)

$$\Sigma(Y_i - \hat{Y}(t))^2 \rightarrow \min \quad (106)$$

Наиболее оптимальной является та математическая функция, сумма квадратов отклонений которой является минимальной по сравнению с другими функциями.

Выбор оптимальной функции производится на основе предварительного анализа при помощи прикладных программ ЭВМ – Microsoft Word и Microsoft Excel графическим методом с использованием приложения «Мастер диаграмм».

Методика выбора линии тренда в Microsoft Word и Microsoft Excel:

- 1) выбрать вкладку «Вставка»; 2) выбрать и активировать вкладку «Диаграмма»; 3) вставить в ячейки данные по исходному ряду динамики; 4) во вкладке «Диаграмма» выбрать «тип диаграммы» - «линейная»; 5) во вкладке «Диаграмма» выбрать «параметры диаграммы» - «добавить линию тренда»

последовательно все представленные функции, отметив во вкладке «параметры линии тренда» - «показывать уравнение на диаграмме» и «поместить на диаграмму величину достоверности аппроксимации».

Пример выбора оптимальной математической функции - (рис. 45,46,47,48, 49 на основании данных табл. 29).

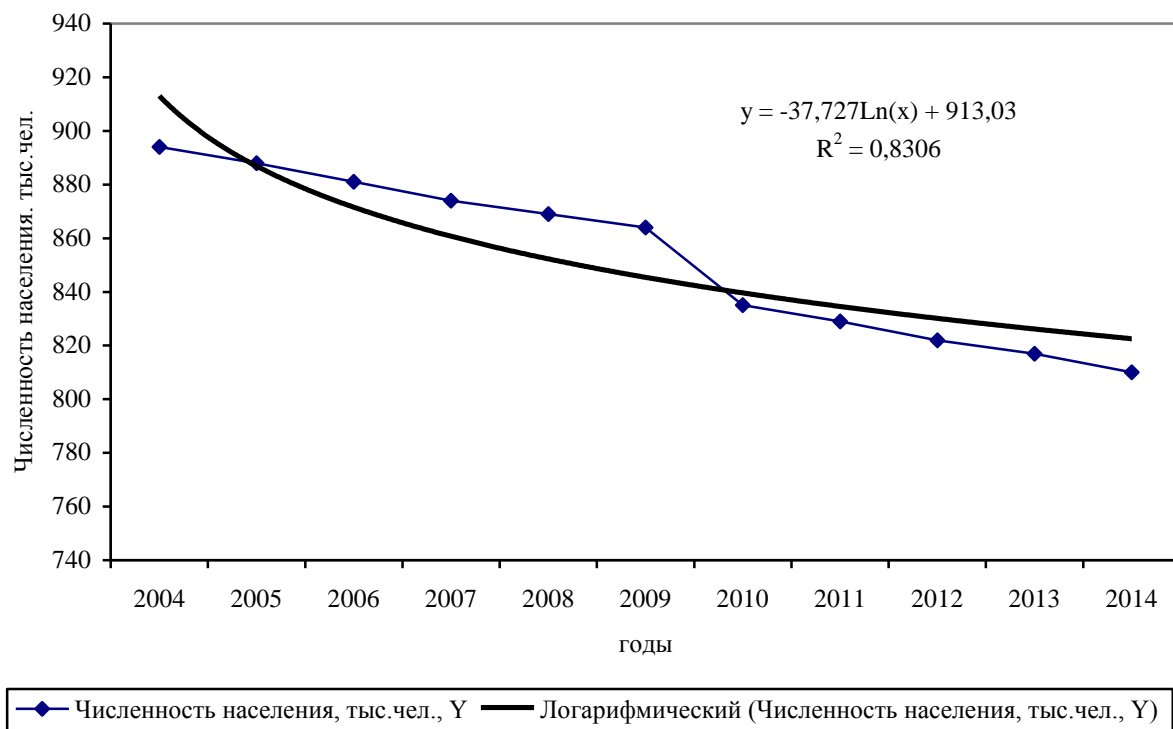


Рисунок 45 - Логарифмический тренд

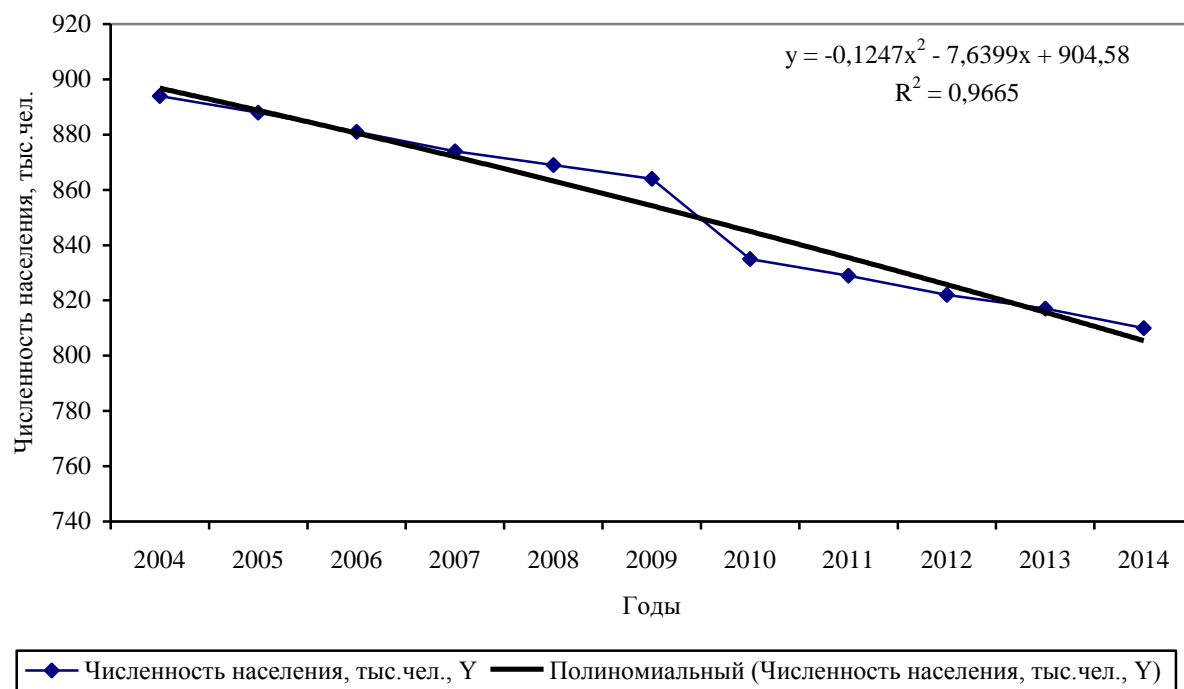


Рисунок 46 - Полиномиальный тренд

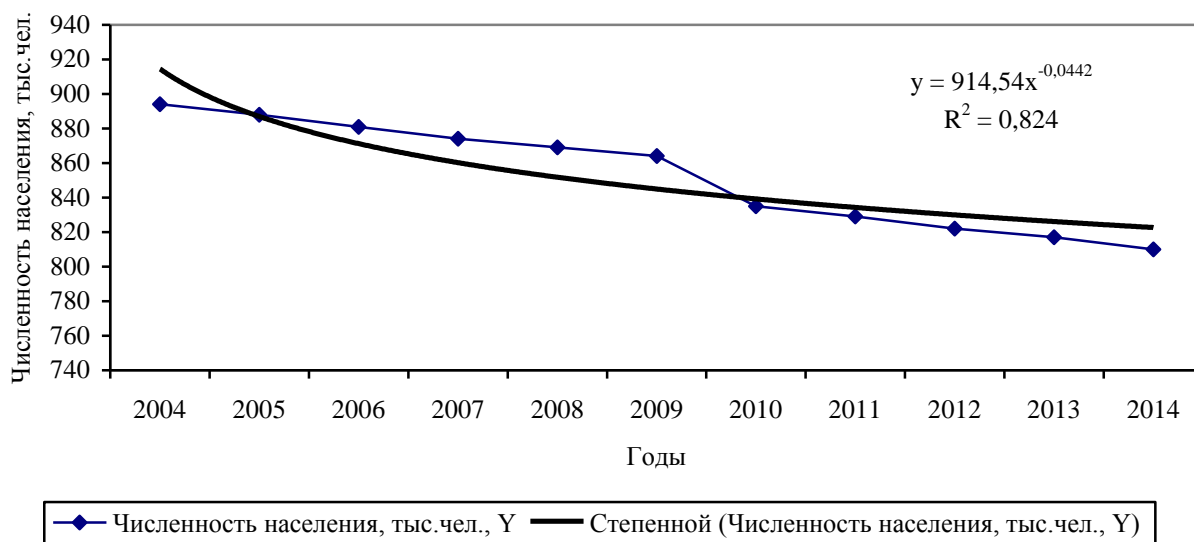


Рисунок 47 - Степенной тренд

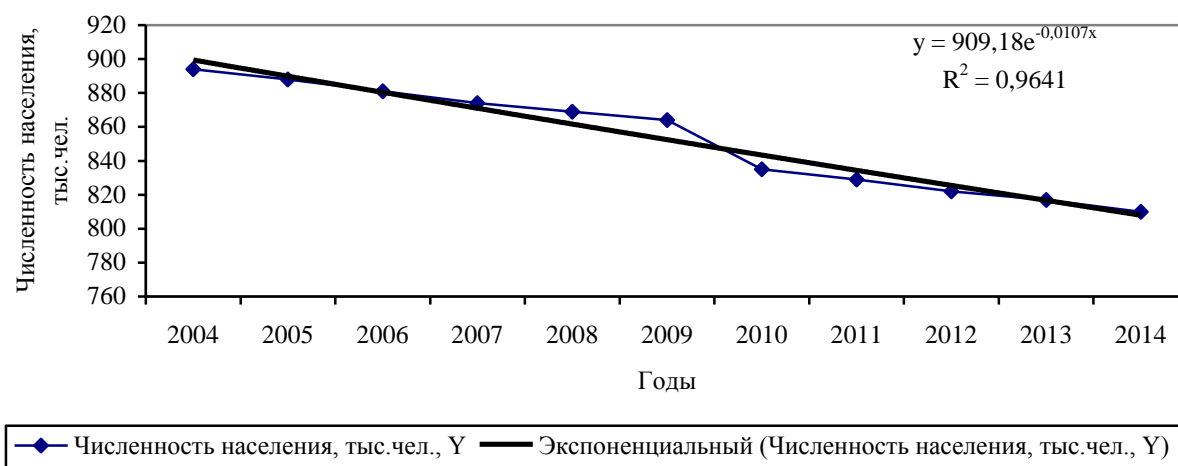


Рисунок 48 - Экспоненциальный тренд

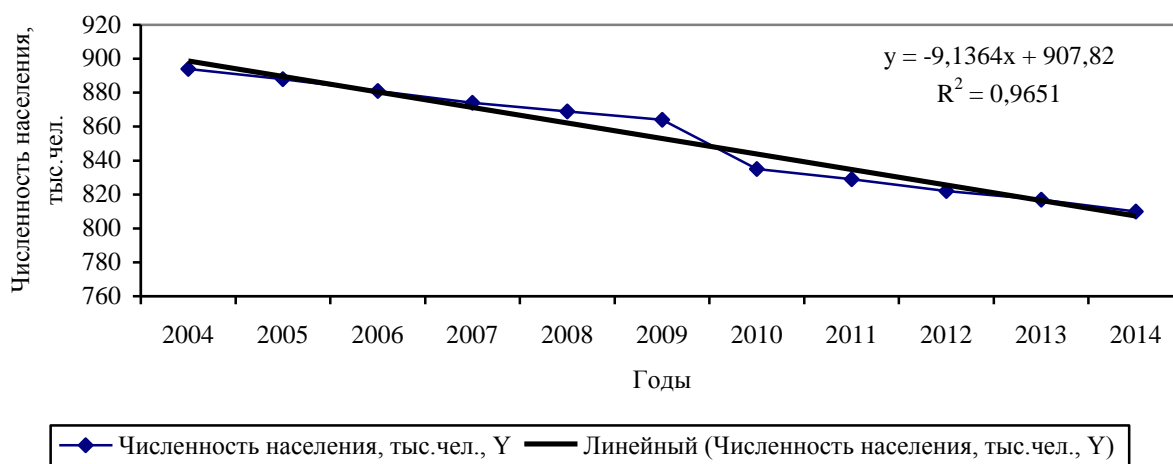


Рисунок 49 - Линейный тренд

Оптимальной является функция, которой соответствует максимальное значение коэффициента детерминации R^2 (табл. 32).

Таблица 32

Обобщение трендов

Линия тренда	Уравнение	Коэффициент детерминации
Логарифмический тренд	$\hat{Y}_t = 913,03 - 37,727Lnt$	$R^2 = 0,8306$
Полиномиальный тренд	$\hat{Y}_t = 904,58 - 7,6399t - 0,1247t^2$	$R^2 = 0,9665$
Степенной тренд	$\hat{Y}_t = 914,54t^{-0,0442}$	$R^2 = 0,8240$
Экспоненциальный тренд	$\hat{Y}_t = 909,18e^{-0,0107t}$	$R^2 = 0,9641$
Линейный тренд	$\hat{Y}_t = 907,82 - 9,1364t$	$R^2 = 0,9651$

Из таблицы 32 видно, что наиболее оптимальным является полиномиальный тренд, который имеет наибольший коэффициент детерминации $R^2 = 0,9665$. Полученное уравнение тренда характеризует ускоряющуюся тенденцию снижения численности населения. Так, в среднем ежегодно, численность населения сокращается на 7,6399 тыс.чел., с ускорением снижения в среднем ежегодно на $0,1247t^2$ тыс.чел. На основании уравнения тренда определяются теоретические уровни численности населения путем подстановки в полученное уравнение тренда каждого значения независимой переменной t (табл. 33). При этом, правильность расчета параметров уравнения тренда проверяется равенством сумм по графе Y и $Y(t)$.

Таблица 33

Расчет теоретических уровней ряда динамики

На 1 января года	Численность населения (фактическая), тыс.чел., Y	t	Численность населения (выровненная), тыс.чел., $Y(t)$
2004	894	1	896,82
2005	888	2	888,80
2006	881	3	880,54
2007	874	4	872,03
2008	869	5	863,26
2009	864	6	854,25
2010	835	7	844,99
2011	829	8	835,48
2012	822	9	825,72
2013	817	10	815,71
2014	811	11	805,45
Итого	9383	66	9383,05

На основе способа наименьших квадратов параметры вышеуказанных уравнений определяются с помощью решения систем из нормальных уравнений. Системы нормальных уравнений включают столько уравнений, сколько параметров математической функции необходимо определить.

Например, для прямолинейной функции, в которой имеется два параметра уравнения тренда a_0 и a_1 , они определяются путем решения системы из двух уравнений

$$\begin{cases} \sum Y = a_0 n + a_1 \sum t \\ \sum Yt = a_0 \sum t + a_1 \sum t^2 \end{cases} \quad (107)$$

где Y – уровни ряда динамики;

t – независимая переменная – показатель времени, который задается порядковым номером;

n – число уровней ряда динамики.

Для упрощения решения данной системы уравнений в статистике применяется метод отсчета от условного начала. При этом условное начало с порядковым номером 0 переносится на середину изучаемого периода, а показателям времени на убывание - задаются отрицательные значения, показателям времени на возрастание - задаются положительные значения. При нечетном числе уровней ряда динамики значения переменной t задаются через интервал равный единице. При четном числе уровней ряда динамики, условное начало приходится на полугодие между двумя серединными годами, следовательно, значение переменной t , равное 0 отсутствует. Значения переменной t начинаются с -1 – на убывание уровней и с 1 – на возрастание уровней. При этом значения переменной t задаются через интервал равный двум. Вследствие применения способа отсчета от условного начала, $\sum t$ приравнивается к 0. В результате, исходная система уравнений упрощается и принимает следующий вид

$$\begin{cases} \sum Y = a_0 n \\ \sum Yt = a_1 \sum t^2 \end{cases} \quad (108)$$

Отсюда параметры уравнения определяются по формулам

$$a_0 = \frac{\sum Y}{n}; \quad (109)$$

$$a_1 = \frac{\sum Yt}{\sum t^2}. \quad (110)$$

Следует учитывать, что упрощенная схема определения параметров уравнения тренда применяется только при использовании прямолинейной математической функции (при равномерном типе развития изучаемого явления).

После определения параметров линии тренда в полученное уравнение подставляется каждое значение независимой переменной t и определяются выровненные (теоретические) уровни ряда динамики $\hat{Y}(t)$. Критерием оценки достоверности анализа является равенство сумм фактических и теоретических уровней ряда динамики

$$\sum Y = \sum \hat{Y}(t). \quad (111)$$

Пример аналитического выравнивания ряда динамики по уравнению прямолинейной функции по упрощенной схеме (на основании данных таблицы 29) представлено в таблице 34.

Таблица 34

Расчет параметров уравнения прямолинейной функции

На 1 января	Численность населения (фактическая), тыс. чел., Y	Порядковый номер года, t	t^2	$Y \times t$	Численность населения выровненная (теоретическая), тыс. чел. $\hat{Y}(t)$
2004	894	-5	25	-4470	898,70
2005	888	-4	16	-3552	889,56
2006	881	-3	9	-2643	880,42
2007	874	-2	4	-1748	871,28
2008	869	-1	1	-869	862,14
2009	864	0	0	0	853,00
2010	835	1	1	835	843,86
2011	829	2	4	1658	834,72
2012	822	3	9	2466	825,58
2013	817	4	16	3268	816,44
2014	811	5	25	4055	807,30
Итого	9383	0	110	-1000	9383,00

Подставляя данные итоговой строки таблицы 34 в вышеуказанные формулы, определяются параметры прямолинейной функции

$$a_0 = \frac{\sum Y}{n} = \frac{9383}{11} = 853,0;$$

$$a_1 = \frac{\sum Yt}{\sum t^2} = \frac{-1000}{110} = -9,09.$$

В результате расчетов уравнение прямолинейного тренда примет вид $\hat{Y}(t) = 853,0 - 9,09 \times t$.

Расчет показал, что параметр уравнения $a_1 = -9,09$ имеет отрицательный знак, следовательно показывает тенденцию снижения численности населения в целом за изучаемый период. Так, в среднем ежегодно численность населения снижается на 9,09 тыс.чел.

Если сравнить полученную тенденцию, определенную по упрощенной схеме с тенденцией, полученной при помощи графического способа на основании приложения «Мастер диаграмм» Microsoft Word и Microsoft Excel (рис. 47), то видно, что тенденция совпадает и по направлению и по значению (незначительные расхождения могут произойти за счет разной точности расчетов). Поэтому, в современных условиях расчеты при использовании математических методов выравнивания ряда динамики производятся при помощи пакета прикладных статистических программ, в частности, при помощи программ Microsoft Word, Microsoft Excel.

7.6 Оценка устойчивости выявленной тенденции

На изменение явления во времени оказывают влияние 2 группы факторов:

1. Систематические факторы, которые обуславливают основную тенденцию развития явления (тренд). (На рисунках 45, 46, 47, 48, 49 влияние систематических факторов отражено линией тренда и характеризует общую тенденцию снижения численности населения).

2. Остаточные случайные факторы, которые обуславливают колеблемость фактических уровней относительно тренда. (На рисунках 45, 46, 47, 48, 49 влияние остаточных случайных факторов отражено фактической эмпирической линией и характеризует либо рост, либо снижение численности населения относительно тренда).

Для измерения степени воздействия остаточных случайных факторов на изменение изучаемого явления используются показатели колеблемости (устойчивости), которые характеризуют вариацию изучаемого явления относительно определенного тренда. К основным показателям устойчивости относятся:

- 1) среднее квадратическое отклонение;
- 2) коэффициент вариации;
- 3) средняя ошибка аппроксимации (приближения).

Среднее квадратическое отклонение в рядах динамики определяется по формуле

$$\sigma_t = \sqrt{\frac{\sum(Y - \hat{Y}(t))^2}{n}}, \quad (112)$$

где Y – фактические (эмпирические) уровни ряда динамики;

$\hat{Y}(t)$ - выровненные (теоретические) уровни ряда динамики;

n – число уровней ряда динамики.

Среднее квадратическое отклонение показывает, на сколько единиц в среднем отклоняются фактические уровни ряда динамики (эмпирическая линия) в ту или другую сторону от выровненных уровней ряда динамики (линии тренда).

Коэффициент вариации является относительным показателем устойчивости и определяется по формуле

$$\nu_t = \frac{\sigma_t}{\bar{Y}} \times 100, \quad (113)$$

где \bar{Y} - средний уровень ряда динамики.

Коэффициент вариации является критерием устойчивости выявленной тенденции и показывает степень влияния остаточных случайных факторов на изменение изучаемого явления. Если коэффициент вариации меньше 25% (в отдельных случаях – меньше 10%), то выявленную тенденцию можно считать устойчивой, а изучаемое явление можно считать стабильно развивающимся во времени. Если коэффициент вариации больше 25% (10%), то выявленную тенденцию можно считать неустойчивой, а изучаемое явление – не стабильно развивающимся во времени за счет сильного влияния случайных (остаточных) факторов.

Средняя ошибка аппроксимации (приближения) является относительным показателем устойчивости и определяется по формуле

$$\bar{A} = \frac{1}{n} \times \sum \frac{|Y_i - \tilde{Y}(t)|}{Y_i} \times 100. \quad (114)$$

Средняя ошибка аппроксимации показывает, насколько близко аналитическая функция выравнивания (линия тренда) приближена к значениям исходного ряда динамики (эмпирической линии). Если значение средней ошибки аппроксимации меньше 10%, то выявленную тенденцию можно считать устойчивой, а уравнение тренда можно считать удачным для достоверного прогнозирования. Если значение средней ошибки аппроксимации больше 10%, то выявленную тенденцию можно считать неустойчивой, а выбранное уравнение тренда можно считать недостаточно удачным для достоверного прогнозирования. В этом случае необходимо выбрать другую линию тренда с меньшей ошибкой аппроксимации. Если это не представляется возможным, то прогноз на перспективу можно проектировать с большей долей допущения и с меньшей достоверностью.

Пример расчета показателей устойчивости. Для расчета показателей устойчивости полиномиального тренда строится вспомогательная таблица 35.

Расчет показателей устойчивости

На 1 января года	Численность населения (фактическая), тыс.чел., Y	Численность населения (выровненная), тыс.чел., $Y(t)$	$y_i - y_t$	$(y_i - y_t)^2$	$\frac{ Y - Y(t) }{\bar{Y}}$
2004	894	896,82	-2,82	7,93	0,0031
2005	888	888,80	-0,80	0,64	0,0009
2006	881	880,54	0,46	0,21	0,0005
2007	874	872,03	1,97	3,90	0,0023
2008	869	863,26	5,74	32,91	0,0066
2009	864	854,25	9,75	95,04	0,0114
2010	835	844,99	-9,99	99,81	0,0118
2011	829	835,48	-6,48	41,99	0,0078
2012	822	825,72	-3,72	13,84	0,0045
2013	817	815,71	1,29	1,66	0,0016
2014	811	805,45	5,55	30,80	0,0069
Итого	9383,00	9383,05	-	328,73	0,0574

Среднее квадратическое отклонение $\sigma_t = \sqrt{\frac{\Sigma(Y - \hat{Y}(t))^2}{n}} = \sqrt{\frac{328,73}{11}} = 5,47$ тыс.чел.

Коэффициент вариации $v_t = \frac{\sigma_t}{\bar{Y}} \times 100 = \frac{5,47}{853,1} \times 100 = 0,64\%$

Средняя ошибка аппроксимации $\bar{A} = \frac{1}{11} \times 0,0574 \times 100 = 0,5218\%$

Расчет показал, что фактические уровни ряда динамики отклоняются от выровненных в ту или другую сторону в среднем на 5,47 тыс.чел. или на 0,64%. Поскольку коэффициент вариации меньше 25%, а средняя ошибка аппроксимации меньше 10%, то выявленную на основании уравнения полиномиального тренда ускоряющую тенденцию снижения численности населения можно считать устойчивой, а численность населения – стабильно уменьшающейся.

Для расчета показателей устойчивости прямолинейного тренда, полученного упрощенным способом строится вспомогательная таблица 36.

Расчет показателей устойчивости

На 1 января года	Численность населения (фактическая), тыс.чел., Y	Численность населения (выровненная), тыс.чел., $Y(t)$	$y_i - y_t$	$(y_i - y_t)^2$	$\frac{ Y - Y(t) }{\bar{Y}}$
2004	894	898,70	-4,70	22,09	0,0055
2005	888	889,56	-1,56	2,43	0,0018
2006	881	880,42	0,58	0,34	0,0007
2007	874	871,28	2,72	7,40	0,0032
2008	869	862,14	6,86	47,06	0,0080
2009	864	853,00	11,00	121,00	0,0129
2010	835	843,86	-8,86	78,50	0,0104
2011	829	834,72	-5,72	32,72	0,0067
2012	822	825,58	-3,58	12,82	0,0042
2013	817	816,44	0,56	0,31	0,0006
2014	811	807,30	3,70	13,69	0,0046
Итого	9383,00	9383,00	-	338,36	0,0586

Среднее квадратическое отклонение $\sigma_t = \sqrt{\frac{\Sigma(Y - \hat{Y}(t))^2}{n}} = \sqrt{\frac{338,36}{11}} = 5,55$ тыс.чел.

Коэффициент вариации $\nu_t = \frac{\sigma_t}{\bar{Y}} \times 100 = \frac{5,55}{853,1} \times 100 = 0,65\%$

Средняя ошибка аппроксимации $\bar{A} = \frac{1}{11} \times 0,0586 \times 100 = 0,5327\%$

Расчет показал, что фактические уровни ряда динамики отклоняются от выровненных в ту или другую сторону в среднем на 5,55 тыс.чел. или на 0,65%. Поскольку коэффициент вариации меньше 25%, а средняя ошибка аппроксимации меньше 10%, то выявленную на основании уравнения прямолинейного тренда тенденцию снижения численности населения можно считать устойчивой, а численность населения – стабильно уменьшающейся.

7.7 Оценка сезонных колебаний

Кроме колеблемости уровней ряда динамики относительно тренда, при анализе рядов динамики важное значение имеет выявление сезонных колебаний изучаемого явления.

Сезонные колебания – это устойчивые изменения уровней ряда динамики по внутригодовым периодам: месяцам, кварталам и пр. Для выявления сезонных колебаний анализируются месячные, квартальные и прочие уровни ряда динамики за год или несколько лет при помощи расчета специфических показателей – индексов сезонности.

Индекс сезонности определяется по формуле

$$I_c = \frac{Y_i}{\bar{Y}} \times 100, \quad (115)$$

где Y_i – уровень каждого месяца (квартала и пр.);

\bar{Y} – средний уровень ряда динамики.

При использовании данных только за один год существенное влияние на значение индекса сезонности оказывают случайные факторы. Поэтому для достоверности выявления закономерностей колеблемости явления используются данные за ряд лет (как правило, за 3 года).

При этом индекс сезонности определяется по формуле

$$\bar{I}_c = \frac{\bar{Y}_i}{\bar{Y}} \times 100, \quad (116)$$

где \bar{Y}_i – средний уровень за 3 года для каждого месяца;

\bar{Y} – общий среднемесячный уровень за 3 года.

Индекс сезонности показывает степень отклонения уровней ряда динамики за каждый месяц от среднегодового уровня.

Расчет индексов сезонности производится на основании исходных данных по транспортному предприятию о величине пробега за три года (табл. 36) и рассчитанных средних величин (табл.37).

Таблица 36

Данные по транспортному предприятию о пробеге автомобилей за три года

Месяц	Величина пробега автомобилей, тыс.км		
	1-й год	2-й год	3-й год
Январь	60,0	70,0	102,0
Февраль	62,0	77,4	105,0
Март	66,4	78,2	107,0
Апрель	70,0	80,0	110,5
Май	78,4	88,4	113,7
Июнь	80,0	89,5	115,0
Июль	80,3	90,3	116,4
Август	86,5	94,6	120,0
Сентябрь	79,0	94,0	118,7
Октябрь	76,4	92,0	115,0
Ноябрь	75,0	90,0	107,5
Декабрь	70,4	85,5	98,2

Для определения среднемесячных и среднегодовых уровней строится вспомогательная таблица 37.

Таблица 37

Средние уровни пробега автомобилей

Месяц	Средняя величина пробега, тыс.км			
	1-й год	2-й год	3-й год	В среднем за 3 года
Январь	60,0	70,0	102,0	77,3
Февраль	62,0	77,4	105,0	81,5
Март	66,4	78,2	107,0	83,9
Апрель	70,0	80,0	110,5	86,8
Май	78,4	88,4	113,7	93,5
Июнь	80,0	89,5	115,0	94,8
Июль	80,3	90,3	116,4	95,7
Август	86,5	94,6	120,0	100,4
Сентябрь	79,0	94,0	118,7	97,2
Октябрь	76,4	92,0	115,0	94,5
Ноябрь	75,0	90,0	107,5	90,8
Декабрь	70,4	85,5	98,2	84,7
За год	73,7	85,8	110,7	90,1

На основании данных таблицы 37 индекс сезонности определяется по формуле 116 (табл. 38).

Индексы сезонности пробега автомобилей

Месяц	Индексы сезонности
Январь	$I_{сез} = \frac{77,3}{90,1} = 0,858$
Февраль	$I_{сез} = \frac{81,5}{90,1} = 0,904$
Март	$I_{сез} = \frac{83,9}{90,1} = 0,931$
Апрель	$I_{сез} = \frac{86,8}{90,1} = 0,963$
Май	$I_{сез} = \frac{93,5}{90,1} = 1,038$
Июнь	$I_{сез} = \frac{94,8}{90,1} = 1,052$
Июль	$I_{сез} = \frac{95,7}{90,1} = 1,062$
Август	$I_{сез} = \frac{100,4}{90,1} = 1,114$
Сентябрь	$I_{сез} = \frac{97,2}{90,1} = 1,079$
Октябрь	$I_{сез} = \frac{94,5}{90,1} = 1,049$
Ноябрь	$I_{сез} = \frac{90,8}{90,1} = 1,008$
Декабрь	$I_{сез} = \frac{84,7}{90,1} = 0,940$

Расчет индекса сезонности (табл. 38) показал, что с мая по ноябрь величина пробега выше среднегодового уровня, а с ноября по май – меньше среднегодового уровня.

Наглядно сезонные колебания изображаются при помощи диаграммы сезонной волны (рис. 50):

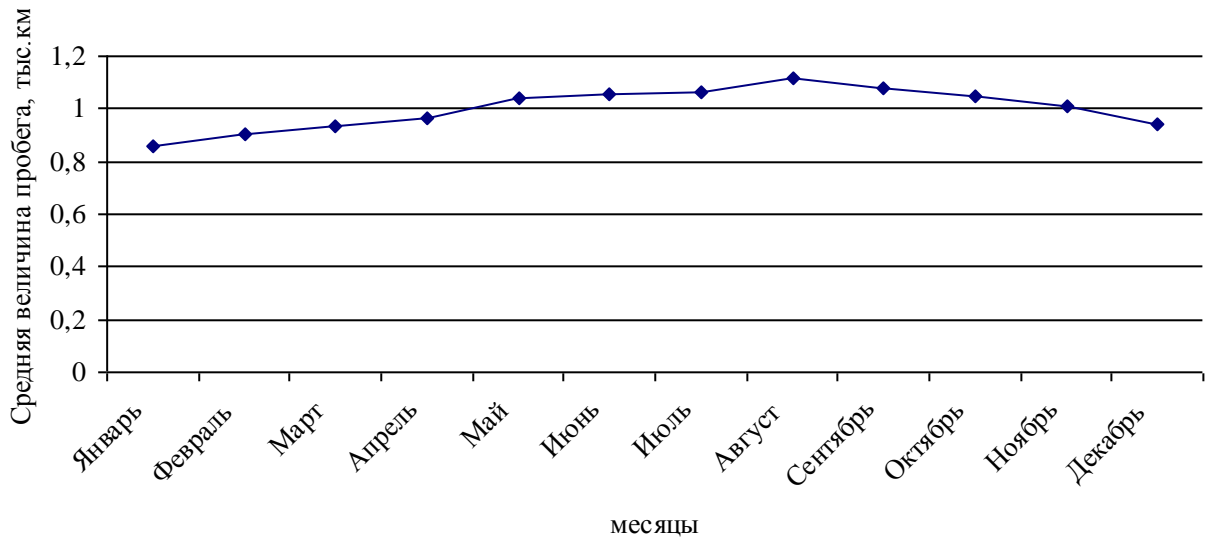


Рисунок 50 - Сезонная волна грузоперевозок

7.8 Прогнозирование развития явления на основе анализа рядов динамики и статистического моделирования в рядах динамики

Четвертая задача построения и анализа рядов динамики заключается в прогнозировании изменения изучаемого явления на перспективу, которое заключается в статистической оценке возможной меры развития социально-экономического явления на будущие периоды.

Статистическое прогнозирование производится методом экстраполяции, который заключается в распространении выявленной в анализе ряда динамики закономерности развития изучаемого явления на будущее. В практическом смысле экстраполяция – это нахождение уровней за пределами изучаемого ряда динамики, то есть – продление исходного ряда динамики на перспективу на основе выявленной закономерности внутри изучаемого ряда динамики.

Основой прогнозирования методом экстраполяции является статистическая гипотеза (предположение) о том, что закономерность, действующая внутри анализируемого базового ряда динамики, сохранится и в дальнейшем.

Применение метода экстраполяции предполагает выполнение следующих требований:

- 1) исходный, базовый ряд динамики должен иметь достаточно большое число уровней, то есть чем больше число уровней имеет исходный ряд динамики, тем на более продолжительный период можно прогнозировать развитие изучаемого явления;
- 2) срок прогнозирования должен быть пропорционален продолжительности исходного базового ряда динамики, то есть при числе уровней исходного ряда динамики не более 10 лет период прогноза может быть не более 3 лет;
- 3) установление сроков прогноза зависит от цели и задач исследования, при этом следует учитывать, что чем короче сроки прогноза, тем достовернее результаты экстраполяции;
- 4) достоверность результатов экстраполяции зависит от степени стабильности изучаемого явления и степени устойчивости выявленной тенденции, которая оценивается при помощи расчета показателей устойчивости.

Экстраполяция может проводиться следующими способами:

1. На основе среднего абсолютного прироста.
2. На основе среднего коэффициента роста.
3. На основе уравнения (линии) тренда.

Экстраполяция на основе среднего абсолютного прироста производится при постоянных цепных абсолютных приростах

$$\Delta i_y - const. \quad (117)$$

При этом прогнозируемый уровень определяется по формуле:

$$Y_{n+1} = Y_n + \bar{\Delta} \times l, \quad (118)$$

где Y_{n+1} – экстраполируемый (прогнозируемый) уровень ряда динамики;

Y_n – конечный уровень исходного базового ряда динамики;

l – срок прогноза;

$\bar{\Delta}$ – средний абсолютный прирост.

Экстраполяции на основе среднего коэффициента роста производится при постоянных цепных темпах роста

$$Tri_{it} - const. \quad (119)$$

При этом прогнозируемый уровень определяется по формуле:

$$Y_{n+1} = Y_n \times \bar{K}_p^l, \quad (120)$$

где \bar{K}_p - средний коэффициент роста.

Экстраполяция на основе уравнения (линии) тренда производится при изменяющихся абсолютных приростах и темпах роста. Результаты экстраполяции на основе аналитического выравнивания являются наиболее достоверными. При способе экстраполяции на основе уравнения (линии) тренда для определения прогнозируемых уровней необходимо продолжить значения независимой переменной t .

Применение вышеуказанных способов позволяет получить точечный прогноз. Для большей достоверности точечный прогноз подвергается статистической оценке при помощи расчета интервального прогноза. Интервальный прогноз предполагает определение возможных пределов (границ прогноза), в которых может находиться прогнозное значение с определенным уровнем вероятности. Для определения интервального прогноза рассчитываются средняя и предельная ошибки прогноза.

Средняя ошибка прогноза определяется по формуле

$$\mu_y = \sqrt{\frac{\sigma^2}{n}}, \quad (121)$$

где σ - среднее квадратическое отклонение;

n – число уровней ряда динамики.

Средняя ошибка прогноза показывает, на сколько единиц отклоняется прогнозируемый уровень от рассчитанного точечного прогнозного значения.

Предельная ошибка прогноза определяется по формуле

$$\Delta_y = \mu_y \times t, \quad (122)$$

где μ_y - средняя ошибка прогноза;

t - коэффициент кратности ошибки, который зависит от принятого уровня вероятности прогноза (как правило, уровень вероятности прогноза принимается 0,95, при котором значение t определяется по таблицам распределения Стьюдента $t_{кр(табл)}(\alpha; \nu = n - k - 1)$).

При анализе рядов динамики наряду с определением уровней ряда динамики за его пределами выполняется задача нахождения недостающих уровней внутри изучаемого ряда динамики. Это производится при помощи метода интерполяции.

Интерполяция – это определение недостающих уровней внутри изучаемого ряда динамики с целью преобразования неполного ряда динамики в полный. Интерполяция производится теми же способами, которые используются при экстраполяции: на основе среднего абсолютного прироста, на основе среднего коэффициента роста, на основе уравнения (линии) тренда. При этом за основу принимается полный период до интерполируемого уровня.

Пример прогнозирования на основе анализа ряда динамики численности населения.

Поскольку изучаемый исходный (базовый) ряд динамики включает одиннадцать уровней, то период прогноза может быть принят не более трех лет.

Экстраполяция на основе среднего абсолютного прироста:

- прогноз на 2015 год $Y_{n+1} = 811 + (-8,3) \times 1 = 802,7$ тыс.чел.
- прогноз на 2016 год $Y_{n+2} = 811 + (-8,3) \times 2 = 794,4$ тыс.чел.
- прогноз на 2017 год $Y_{n+3} = 811 + (-8,3) \times 3 = 786,1$ тыс.чел.

Прогноз на основании среднего абсолютного прироста показал, что при условии сохранения выявленной тенденции снижения к 2017 году численность населения может составить 786,1 тыс.человек.

Экстраполяция на основе среднего коэффициента роста:

- прогноз на 2015 год $Y_{n+1} = 811 \times 0,9969^1 = 808,48$ тыс.чел.
- прогноз на 2016 год $Y_{n+2} = 811 \times 0,9969^2 = 805,98$ тыс.чел.
- прогноз на 2017 год $Y_{n+3} = 811 \times 0,9969^3 = 803,48$ тыс.чел.

Прогноз на основании среднего коэффициента роста показал, что при условии сохранения выявленной тенденции снижения к 2017 году численность населения может составить 803,48 тыс.человек.

Экстраполяция на основе уравнения (линии) полиномиального тренда:

- прогноз на 2015 год $\hat{Y}(t) = 904,58 - 7,6399 \times 12 - 0,1247 \times 12^2 = 794,49$ тыс.чел.

- прогноз на 2016 год $\hat{Y}(t) = 904,58 - 7,6399 \times 13 - 0,1247 \times 13^2 = 784,19$ тыс.чел.

- прогноз на 2017 год $\hat{Y}(t) = 904,58 - 7,6399 \times 14 - 0,1247 \times 14^2 = 773,18$ тыс.чел.

Прогноз на основании уравнения полиномиального тренда показал, что при условии сохранения выявленной ускоряющейся тенденции снижения к 2017 году численность населения может составить 773,18 тыс.человек. (средний прогноз).

Экстраполяция на основе уравнения (линии) прямолинейного тренда, полученного упрощенным способом:

- прогноз на 2015 год $\hat{Y}(t) = 853,0 - 9,14 \times 7 = 789,02$ тыс.чел.

- прогноз на 2016 год $\hat{Y}(t) = 853,0 - 9,14 \times 9 = 770,74$ тыс.чел.

- прогноз на 2017 год $\hat{Y}(t) = 853,0 - 9,14 \times 11 = 752,46$ тыс.чел.

Прогноз на основании уравнения прямолинейного тренда показал, что при условии сохранения выявленной тенденции снижения к 2017 году численность населения может составить 752,46 тыс.человек. (средний прогноз).

Полученные точечные прогнозы на 2017 год на основании уравнений (линий) тренда следует дополнить интервальным прогнозом при уровне вероятности прогноза 0,95. Для этого определяется величина средней и предельной ошибок.

Для прогноза по полиномиальному тренду:

- средняя ошибка прогноза $\mu_y = \sqrt{\frac{5,47^2}{11}} = 1,65$ тыс.чел.

Для определения предельной ошибки определяется коэффициент кратности ошибки по таблицам распределения Стьюдента (табл.71).

$t_{кр(табл)}(\alpha; \nu = n - k - 1)$. При вероятности прогноза 0,95 ($\alpha = 0,05$) и числе

степеней свободы $\nu = n - k - 1 = 11 - 1 - 1 = 9$, коэффициент кратности ошибки составит $t_{кр(табл)} = 2,2622$

- предельная ошибка прогноза $\Delta_y = 1,65 \times 2,2622 = 3,7326$ тыс.чел.

Для прогноза по прямолинейному тренду:

- средняя ошибка прогноза $\mu_y = \sqrt{\frac{5,55^2}{11}} = 1,67$ тыс.чел.

- предельная ошибка прогноза $\Delta_y = 1,67 \times 2,2622 = 3,7779$ тыс.чел.

Пределы интервального прогноза на 2017 год по полиномиальному тренду

- нижний предел (пессимистический прогноз):

$$773,18 - 3,7326 = 769,45 \text{ тыс.чел.}$$

- верхний предел (оптимистический прогноз):

$$773,18 + 3,7326 = 776,91 \text{ тыс.чел.}$$

Интервальный прогноз по полиномиальному тренду показал, что с вероятностью 0,95 можно утверждать, что при условии сохранения выявленной ускоряющейся тенденции снижения на перспективу, к 2017 году численность населения может составить от 769,45 тыс.человек до 776,91 тыс.человек.

Пределы интервального прогноза на 2017 год по прямолинейному тренду

- нижний предел (пессимистический прогноз):

$$752,46 - 3,7779 = 748,68 \text{ тыс.чел.}$$

- верхний предел (оптимистический прогноз):

$$752,46 + 3,7779 = 756,24 \text{ тыс.чел.}$$

Интервальный прогноз по прямолинейному тренду показал, что с вероятностью 0,95 можно утверждать, что при условии сохранения выявленной тенденции снижения на перспективу, к 2017 году численность населения может составить от 748,68 тыс.человек до 756,24 тыс.человек.

8 СТАТИСТИЧЕСКИЕ ИНДЕКСЫ

8.1 Сущность и значение индексного метода в статистическом анализе социально-экономических явлений. Виды и свойства индексов

Одним из видов обобщающих статистических показателей, относящихся к сложным относительным величинам, является индекс. Статистические индексы используются при анализе массовых общественных явлений по следующим направлениям:

- 1) для соизмерения непосредственно несоизмеримых явлений;
- 2) для характеристики развития изучаемых явлений во времени;
- 3) для характеристики изменения изучаемых явлений в пространстве;
- 4) для характеристики структуры изучаемого явления;
- 5) для оценки взаимосвязей между изучаемыми явлениями, при их функциональной зависимости;
- 6) для выявления степени влияния факторов на изменение сложного социально-экономического явления;
- 7) Для сравнения изучаемого явления с определенным эталоном.

Индекс в статистике – это относительный обобщающий показатель сравнения двух сложных явлений, состоящих из элементов, которые не поддаются непосредственному суммированию.

Например, при сравнении общего объема производства продукции двух предприятий, относящихся к различным отраслям, в натуральном выражении сравнение невозможно, так как объем разнородных видов продукции не поддается непосредственному суммированию в натуральных измерителях в силу их разнокачественности и несоизмеримости. С целью соизмерения разнокачественных видов продукции, приведения их в сопоставимый вид и впоследствии – сравнения, используются статистические индексы.

Статистические индексы подразделяются на виды по следующим классификационным признакам (рис. 51).

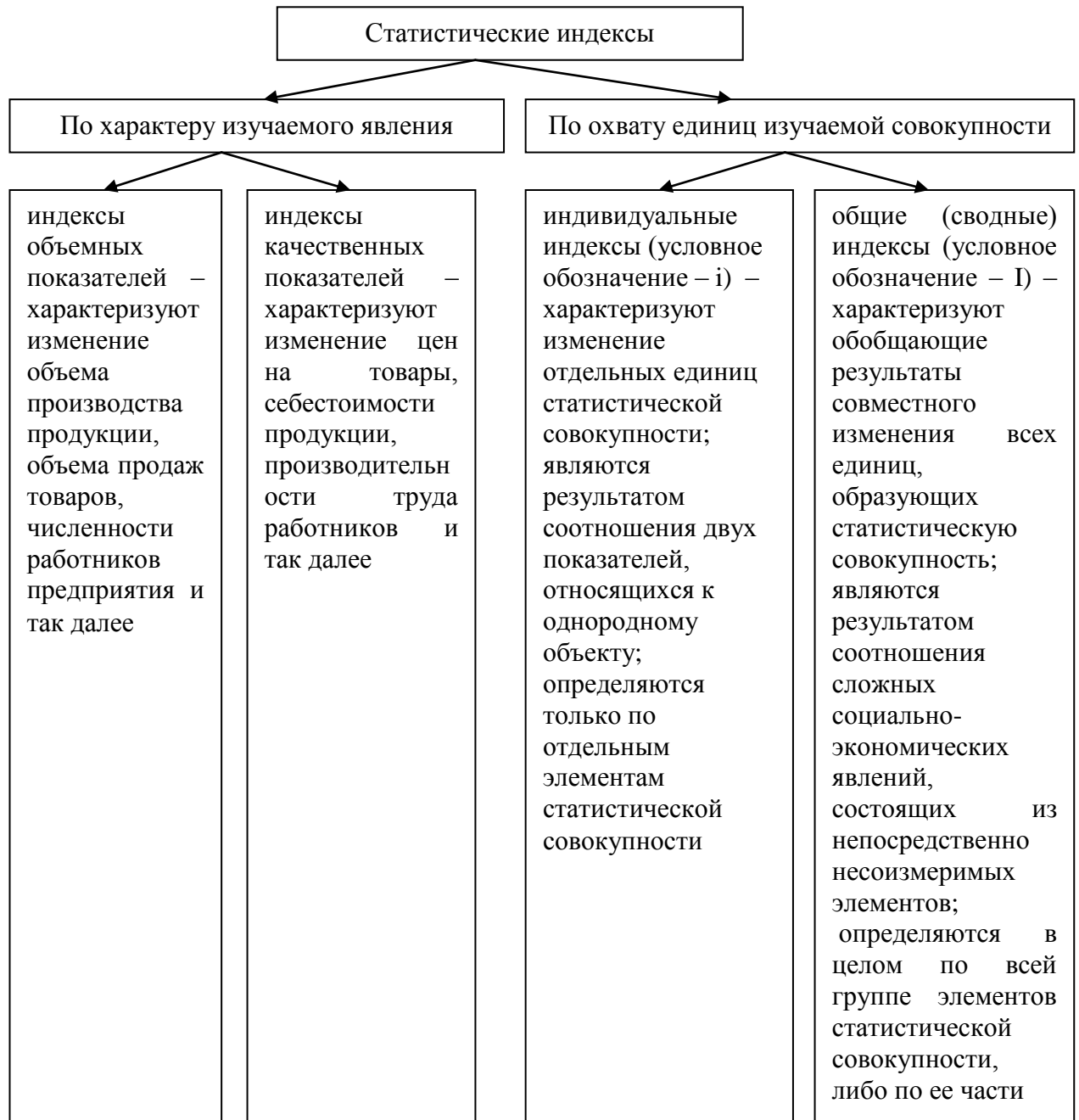


Рисунок 51 – Виды статистических индексов

Статистические индексы обладают следующими свойствами:

1. Синтетические свойства - посредством индексного метода производится соединение в единое целое разнородных единиц статистической совокупности.

2. Аналитические свойства - посредством индексного метода производится разложение изучаемого явления на составные элементы и оценивается влияние отдельных элементов и факторов на его изменение.

Для расчета статистического индекса необходимо произвести соотношение двух величин: числитель индексного соотношения (сравниваемый показатель) – это текущий (отчетный) уровень изучаемого явления (условное обозначение – 1); знаменатель индексного соотношения (показатель, с которым сравнивают) – это база сравнения, то есть – базисный уровень изучаемого явления (условное обозначение – 0). Индексное соотношение можно выразить следующей общей формулой

$$\text{Статистический индекс} = \frac{\text{Сравниваемый показатель} - \text{текущий (отчетный уровень) (1)}}{\text{База сравнения} - \text{базисный уровень (0)}} \quad (123)$$

При построении индексных соотношений используется система условных обозначений: физический объем производства продукции (продаж товаров) – q; цена единицы товара – p; себестоимость единицы произведенной продукции – z; трудоемкость производства единицы продукции – t; урожайность сельскохозяйственных культур – У; посевная площадь сельскохозяйственных культур – П; численность работников – Т и прочие.

Результат расчета индексного соотношения может быть выражен в коэффициентах или процентах (%). Критерием оценки индекса является 1 или 100%. Если полученный индекс больше 1 или 100%, то он показывает увеличение изучаемого показателя – текущего (отчетного) уровня по сравнению с базисным уровнем. Если полученный индекс меньше 1 или 100%, то он показывает уменьшение изучаемого показателя текущего (отчетного) уровня по сравнению с базисным уровнем. На сколько увеличился или уменьшился изучаемый показатель определяется путем вычитания из полученного индекса, выраженного в процентах (%) ста процентов.

8.2 Составные элементы, формы и правила построения основных статистических индексов

Составные элементы статистического индекса:

1. Индексируемая величина.
2. Соизмеритель (статистический вес).

Индексируемая величина – это показатель, изменение которого определяется с помощью данного индекса (в индексе цен – это цена, в индексе себестоимости – себестоимость, в индексе физического объема – объем производства или продаж и т.д.).

Соизмеритель (статистический вес) – это сомножитель индексируемой величины - показатель, который используется для приведения разнородных элементов статистической совокупности в сопоставимый вид.

Индивидуальный индекс состоит только из одного элемента – индексируемой величины. Общий (сводный) индекс состоит из двух элементов – индексируемой величины и соизмерителя. В числителе и знаменателе общего (сводного) индекса находятся произведения индексируемой величины и соизмерителя, которые отражают определенные экономические категории.

Соизмеритель применяется для перехода от натуральных единиц измерения разнородных элементов изучаемого явления к общим для них сопоставимым единицам измерения (например, для исчисления сводных (общих) индексов необходимо привести составные элементы изучаемого явления в сопоставимый вид. Для оценки изменения объема производства разнородной продукции в двух сравниваемых периодах (либо по двум разным предприятиям, регионам, странам) необходимо для приведения в однородный вид в качестве соизмерителя принять одинаковые цены. Для оценки изменения уровня цен в целом на все разнородные проданные товары необходимо для приведения их в сопоставимый вид, принять в качестве соизмерителя одинаковый объем продаж этих товаров.

В качестве соизмерителя индексируемых величин используются показатели, которые тесно взаимосвязаны с показателем индексируемой величины, чтобы в результате их произведения получались отдельные экономические категории (товарооборот, производственные затраты и пр.).

В зависимости от вида индекса используются два варианта соизмерителей: 1) в индексах объемных показателей соизмерители фиксируются, как правило, на базисном уровне; 2) в индексах качественных показателей соизмерители фиксируются, как правило, на отчетном уровне.

Статистические общие (сводные) индексы принимают три формы: агрегатную форму; форму среднего арифметического индекса; форму среднего гармонического индекса.

Агрегатная форма является основной формой общих (сводных) индексов.

Основные статистические индексы в агрегатной форме:

1) индивидуальный индекс физического объема

$$i_q = \frac{q_1}{q_0}, \quad (124)$$

где q_1 – объем производства продукции (продаж товаров) отчетного периода;

q_0 – объем производства продукции (продаж товаров) базисного периода;

2) индивидуальный индекс цены единицы товара

$$i_p = \frac{p_1}{p_0}, \quad (125)$$

где p_1 – цена единицы товара отчетного периода;

p_0 – цена единицы товара базисного периода.

3) общий (сводный) индекс физического объема

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}, \quad (126)$$

4) общий (сводный) индекс цен

$$I_p = \frac{\sum q_1 p_1}{\sum q_0 p_1}, \quad (127)$$

Форма среднего арифметического индекса является производной от агрегатной формы индекса физического объема. Используется форма среднего

арифметического индекса в том случае, когда отсутствует информация об объеме производства продукции за отчетный период, а имеется только информация об объеме производства продукции базисного периода и индивидуальном индексе физического объема. При этом физический объем производства продукции отчетного периода выражается из индивидуального индекса физического объема

$$i_q = \frac{q_1}{q_0}; \quad (128)$$

$$\downarrow$$

$$q_1 = i_q \times q_0. \quad (129)$$

В результате полученное выражение объема производства продукции подставляется в агрегатную форму общего (сводного) индекса физического объема

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{\sum i_q q_0 p_0}{\sum q_0 p_0}. \quad (130)$$

Полученная формула представляет собой среднюю арифметическую из индивидуальных индексов физического объема, взвешенную по стоимости продукции базисного периода. Интерпретация сводного индекса в форме среднего арифметического такая же, как и интерпретация сводного индекса физического объема в агрегатной форме.

Форма среднего гармонического индекса является производной от агрегатной формы индекса цен. Используется форма среднего гармонического индекса в том случае, когда отсутствует информация о цене единицы товара базисного периода, а имеется только информация о цене отчетного периода и индивидуальных индексах цен. При этом цена единицы товара базисного периода выражается из индивидуального индекса цен

$$i_p = \frac{p_1}{p_0}; \quad (131)$$

$$\downarrow$$

$$p_0 = \frac{p_1}{i_p}. \quad (132)$$

В результате полученное выражение цены единицы товара базисного периода подставляется в агрегатную форму общего (сводного) индекса цен

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{\sum q_1 p_1}{\sum \frac{q_1 p_1}{i_p}}. \quad (133)$$

Интерпретация сводного индекса в форме среднего гармонического такая же, как и интерпретация сводного индекса цен в агрегатной форме.

8.3 Взаимосвязь индексов

Индексы в статистике взаимосвязаны между собой. Взаимосвязь индексов проявляется в трех вариантах:

1. Взаимосвязь цепных и базисных индексов.
2. Взаимосвязь индексов переменного и фиксированного состава.
3. Взаимосвязь основных экономических индексов.

Взаимосвязь цепных и базисных индексов используется при изучении сложных социально-экономических явлений, когда производится сравнение более чем за два периода. Индексы рассчитываются в двух вариантах (аналогично цепным и базисным показателям динамики): 1) базисные индексы – на постоянной базе сравнения; 2) цепные индексы – на переменной базе сравнения. Поэтому взаимосвязь цепных и базисных индексов проявляется в двух вариантах:

1) произведение цепных индивидуальных индексов равно базисному индексу последнего уровня

$$i_{2/1} \times i_{3/2} \times i_{4/3} \times i_{5/4} \dots \times i_{n/n-1} = i_{n/1}; \quad (134)$$

2) произведение цепных сводных индексов с постоянными весами равно базисному сводному индексу последнего уровня

$$I_{2/1} \times I_{3/2} \times I_{4/3} \times I_{5/4} \dots \times I_{n/n-1} = I_{n/1}. \quad (135)$$

Взаимосвязь индексов переменного и фиксированного состава используется при изучении сложных социально-экономических явлений при оценке степени влияния на изменение изучаемого показателя не только значений индексируемой величины, но и соизмерителя. Поэтому индексы рассчитываются в двух вариантах: 1) индексы фиксированного состава (индексы с постоянными весами); 2) индексы переменного состава (индексы с переменными весами).

Индексы фиксированного состава (индексы с постоянными весами) – это индексы в числителе и знаменателе которых соизмеритель фиксируется на уровне одного и того же периода. При этом индексируемая величина изменяется, а соизмеритель остается неизменным. Например, индекс цен фиксированного состава

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}. \quad (136)$$

Индексы переменного состава (индексы с переменными весами) – это индексы в числителе и знаменателе которых изменяется не только индексируемая величина, но и соизмеритель. Индексы переменного состава являются соотношением средних уровней изучаемого явления и характеризуют общее изменение изучаемого показателя в результате изменения двух факторов: индивидуальных значений индексируемой величины и структуры совокупности (соизмерителя). Например, индекс цен переменного состава

$$I_{\bar{p}} = \frac{\bar{p}_1}{\bar{p}_0} = \frac{\sum q_1 p_1}{\sum q_1} \div \frac{\sum q_0 p_0}{\sum q_0}. \quad (137)$$

Взаимосвязь индексов фиксированного и переменного состава проявляется следующим образом – индекс переменного состава равен произведению индекса фиксированного состава и индекса структуры

$$I_{\text{переменного состава}}(I_{\bar{p}}) = I_{\text{фиксированного состава}}(I_p) \times I_{\text{структуры}}(I_{стр}), \quad (138)$$

$$\frac{\sum q_1 p_1}{\sum q_1} \div \frac{\sum q_0 p_0}{\sum q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times \left(\frac{\sum q_1 p_0}{\sum q_1} \div \frac{\sum q_0 p_0}{\sum q_0} \right). \quad (139)$$

Взаимосвязь основных экономических индексов используется при изучении изменений взаимосвязанных между собой экономических показателей. Поэтому взаимосвязь основных экономических индексов проявляется на основании взаимосвязи отражаемых ими экономических показателей:

1) товарооборот является произведением цены единицы каждого товара и объема продаж каждого вида товара. Взаимосвязь индексов цен, физического объема продаж и товарооборота проявляется следующим образом

$$I_{\text{товарооборота}}(I_{qp}) = I_{\text{цен}}(I_p) \times I_{\text{физического объема продаж}}(I_q), \quad (140)$$

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum q_1 p_1}{\sum q_1 p_0} \times \frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad (141)$$

2) производственные затраты являются произведением себестоимости единицы продукции и объема производства каждого вида продукции. Взаимосвязь индексов себестоимости, объема производства и производственных затрат проявляется следующим образом

$$I_{\text{производственных затрат}}(I_{qz}) = I_{\text{себестоимости}}(I_z) \times I_{\text{физического объема производства}}(I_q), \quad (142)$$

$$\frac{\sum z_1 q_1}{\sum z_0 q_0} = \frac{\sum q_1 z_1}{\sum q_1 z_0} \times \frac{\sum q_1 z_0}{\sum q_0 z_0}; \quad (143)$$

3) взаимосвязь индексов трудоемкости единицы продукции, объема производства продукции и затрат труда

$$I_{\text{затрат труда}}(I_{qt}) = I_{\text{трудоемкости}}(I_t) \times I_{\text{физического объема производства}}(I_q), \quad (144)$$

$$\frac{\sum t_1 q_1}{\sum t_0 q_0} = \frac{\sum q_1 t_1}{\sum q_1 t_0} \times \frac{\sum q_1 t_0}{\sum q_0 t_0}; \quad (145)$$

4) взаимосвязь индексов валового сбора, посевных площадей и урожайности сельскохозяйственных культур

$$I_{\text{валового сбора}}(I_{УП}) = I_{\text{урожайности}}(I_{У}) \times I_{\text{посевных площадей}}(I_{П}), \quad (146)$$

$$\frac{\sum Y_1 P_1}{\sum Y_0 P_0} = \frac{\sum P_1 Y_1}{\sum P_1 Y_0} \times \frac{\sum P_1 Y_0}{\sum P_0 Y_0}; \quad (147)$$

5) взаимосвязь индексов фонда оплаты труда, средней заработной платы одного работника и численности работников

$$I_{\text{фонда оплаты труда}}(I_{ТЗ}) = I_{\text{средней заработной платы}}(I_{З}) \times I_{\text{численности работников}}(I_{Т}), \quad (148)$$

$$\frac{\sum Z_1 T_1}{\sum Z_0 T_0} = \frac{\sum T_1 Z_1}{\sum T_1 Z_0} \times \frac{\sum T_1 Z_0}{\sum T_0 Z_0}; \quad (149)$$

б) и другие.

8.4 Построение основных экономических индексов

Основные экономические индексы отражают изменение основных экономических явлений (товарооборота, производственных затрат, производительности труда и т.д.).

Индексы товарооборота, цен и физического объема продаж включают следующую систему индексов.

Индивидуальные индексы цен единицы товара

$$i_p = \frac{p_1}{p_0}. \quad (150)$$

Индивидуальный индекс цен показывает, на сколько процентов или во сколько раз увеличилась, либо уменьшилась цена каждого товара в отчетном периоде по сравнению с базисным.

Индивидуальные индексы физического объема продаж единицы товара

$$i_q = \frac{q_1}{q_0}. \quad (151)$$

Индивидуальный индекс физического объема продаж показывает, на сколько процентов или во сколько раз увеличился, либо уменьшился объем продаж каждого вида товара в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс товарооборота

$$I_{qp} = \frac{\sum p_1 q_1}{\sum p_0 q_0}. \quad (152)$$

Общий индекс товарооборота показывает, на сколько процентов или во сколько раз увеличился, либо уменьшился общий объем товарооборота по всей группе товаров в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс цен

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} \quad (153)$$

Общий (сводный) индекс цен показывает, на сколько процентов или во сколько раз увеличились, либо уменьшились цены в целом на всю группу товаров в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс физического объема продаж

$$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}. \quad (154)$$

Общий (сводный) индекс физического объема продаж показывает, на сколько процентов или во сколько раз увеличились, либо уменьшились объемы продаж в целом по всей группе товаров в отчетном периоде по сравнению с базисным.

На основании исходных индексных соотношений определяются абсолютные изменения товарооборота всего, и в том числе за счет влияния факторов - изменения цен и физического объема продаж.

На основании общего индекса товарооборота определяется абсолютное изменение товарооборота

$$\Delta qp = \sum q_1 p_1 - \sum q_0 p_0. \quad (155)$$

Абсолютное изменение показывает, на сколько единиц увеличилась либо уменьшилась выручка от продаж (товарооборот) в целом по всей группе товаров в отчетном периоде по сравнению с базисным.

На основании общего индекса цен определяется абсолютное изменение товарооборота за счет изменения цен на товары

$$\Delta qp(p) = \sum q_1 p_1 - \sum q_1 p_0. \quad (156)$$

Абсолютное изменение показывает, на сколько единиц увеличилась либо уменьшилась выручка от продаж (товарооборот) в целом по всей группе товаров в отчетном периоде по сравнению с базисным за счет изменения цены товаров.

На основании общего индекса физического объема продаж определяется изменение товарооборота за счет изменения объема продаж товаров

$$\Delta qp(q) = \sum q_1 p_0 - \sum q_0 p_0. \quad (157)$$

Абсолютное изменение показывает, на сколько единиц увеличилась либо уменьшилась выручка от продаж (товарооборот) в отчетном периоде по сравнению с базисным за счет изменения объема продаж в целом по всей группе товаров.

Взаимосвязь индексов

$$I_{qp} = I_p \times I_q. \quad (158)$$

Взаимосвязь абсолютных изменений:

$$\Delta qp = \Delta qp(p) + \Delta qp(q). \quad (159)$$

Пример расчета выше представленных индексов товарооборота, цен и физического объема продаж представлен на основании условных исходных данных о реализации продукции (табл. 39).

Индивидуальные индексы цен:

- товар А $i_A = p_1/p_0 = 24/22 = 1,0909$;
- товар Б $i_B = p_1/p_0 = 28/30 = 0,9333$;
- товар В $i_B = p_1/p_0 = 35/40 = 0,8750$.

Расчет показал, что в отчетном периоде по сравнению с базисным периодом цена на товар А увеличилась на 9,09%, на товар Б - снизилась на 6,67%, на товар В - снизилась на 12,5%.

Таблица 39

Исходные условные данные для расчета индексов

Наименование товара	Базисный период		Отчетный период		Сума товарооборота, рублей		
	Цена за 1 единицу, рублей	Объем продаж, единиц	Цена за 1 единицу, рублей	Объем продаж, единиц	Базисный период	Отчетный период	Условный
	p_0	q_0	p_1	q_1	$q_0 p_0$	$q_1 p_1$	$q_1 p_0$
А	22	35	24	25	770	600	550
Б	30	20	28	30	600	840	900
В	40	50	35	40	2000	1400	1600
Итого	-	-	-	-	3370	2840	3050

Индивидуальные индексы физического объема продаж:

- товар А $i_A = q_1/q_0 = 25/35 = 0,7143$;
- товар Б $i_B = q_1/q_0 = 30/20 = 1,5$;
- товар В $i_V = q_1/q_0 = 40/50 = 0,8$.

Расчет показал, что в отчетном периоде по сравнению с базисным периодом физический объем продаж товара А уменьшился на 28,57%, товара Б - увеличился на 50%, товара В - снизился на 20%.

Общий индекс цен постоянного (фиксированного) состава

$$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0} = \frac{2840}{3050} = 0,9311.$$

Расчет показал, что цены в целом на всю группу товаров в отчетном периоде по сравнению с базисным периодом снизились на 6,89 %.

$$\text{Общий индекс физического объема продаж } I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0} = \frac{3050}{3370} = 0,9050.$$

Расчет показал, что в отчетном периоде по сравнению с базисным периодом объем продаж всей группы товаров уменьшился на 9,5 %.

$$\text{Общий индекс товарооборота } I_{qp} = \frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{2840}{3370} = 0,8427.$$

Расчет показал, что общая сумма товарооборота (выручки от продаж) по данной группе товаров в отчетном периоде по сравнению с базисным периодом уменьшилась на 15,73 %.

Абсолютное изменение товарооборота (выручки от реализации)

$$\Delta qp = \Sigma q_1 p_1 - \Sigma q_0 p_0 = 2840 - 3370 = -530 \text{ рублей};$$

- в том числе за счет изменения цен

$$\Delta qp(p) = \Sigma q_1 p_1 - \Sigma q_1 p_0 = 2840 - 3050 = -210 \text{ рублей};$$

- в том числе за счет изменения физического объема продаж

$$\Delta qp(q) = \Sigma q_1 p_0 - \Sigma q_0 p_0 = 3050 - 3370 = -320 \text{ рублей}.$$

Взаимосвязь индексов

$$I_{qp} = I_p \times I_q = 0,8427 = 0,9311 * 0,9050.$$

Взаимосвязь абсолютных изменений

$$\Delta qp = \Delta qp(p) + \Delta qp(q) = -530 = -210 + (-320).$$

Расчет показал, что выручка от реализации данной группы товаров (товарооборот) в отчетном периоде по сравнению с базисным периодом уменьшилась на 530 рублей, в том числе снижение цен привело к уменьшению выручки на 210 рублей, а снижение объема продаж привело к уменьшению выручки от реализации на 320 рублей

Индексы производственных затрат, себестоимости и физического объема производства включают следующую систему индексов.

Индивидуальные индексы себестоимости единицы продукции

$$i_z = \frac{z_1}{z_0}. \quad (160)$$

Индивидуальный индекс себестоимости показывает, на сколько процентов или во сколько раз увеличилась, либо уменьшилась себестоимость каждого вида продукции в отчетном периоде по сравнению с базисным;

Индивидуальные индексы физического объема производства единицы продукции

$$i_q = \frac{q_1}{q_0}. \quad (161)$$

Индивидуальный индекс физического объема производства показывает, на сколько процентов или во сколько раз увеличился, либо уменьшился объем производства каждого вида продукции в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс производственных затрат

$$I_{qz} = \frac{\sum z_1 q_1}{\sum z_0 q_0}. \quad (162)$$

Общий индекс производственных затрат показывает, на сколько процентов или во сколько раз увеличилась, либо уменьшилась общая сумма затрат на производство всей группы продукции в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс себестоимости

$$I_z = \frac{\sum q_1 z_1}{\sum q_1 z_0}. \quad (163)$$

Общий (сводный) индекс себестоимости показывает, на сколько процентов или во сколько раз увеличилась, либо уменьшилась себестоимость в целом всей группы продукции в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс физического объема производства

$$I_q = \frac{\sum q_1 z_0}{\sum q_0 z_0}. \quad (164)$$

Общий (сводный) индекс физического объема производства показывает, на сколько процентов или во сколько раз увеличился, либо уменьшился объем производства в целом по всей группе продукции в отчетном периоде по сравнению с базисным.

На основании исходных индексных соотношений определяются абсолютные изменения производственных затрат всего, и в том числе за счет влияния факторов: изменения себестоимости единицы продукции и изменения физического объема производства.

На основании общего индекса производственных затрат определяется их абсолютное изменение

$$\Delta qz = \Sigma q_1 z_1 - \Sigma q_0 z_0. \quad (165)$$

Абсолютное изменение показывает, на сколько единиц увеличились либо уменьшились затраты на производство всей группы продукции в отчетном периоде по сравнению с базисным.

На основании общего индекса себестоимости определяется абсолютное изменение производственных затрат за счет изменения себестоимости единицы продукции

$$\Delta qz(z) = \Sigma q_1 z_1 - \Sigma q_1 z_0. \quad (166)$$

Абсолютное изменение показывает, на сколько единиц увеличились либо уменьшились затраты на производство всей группы продукции в отчетном периоде по сравнению с базисным за счет изменения себестоимости единицы продукции.

На основании общего индекса физического объема производства определяется изменение производственных затрат за счет изменения объема производства продукции

$$\Delta qz(q) = \Sigma q_1 z_0 - \Sigma q_0 z_0. \quad (167)$$

Абсолютное изменение показывает, на сколько единиц увеличились либо уменьшились затраты на производство в отчетном периоде по сравнению с базисным за счет изменения объема производства всей группы продукции.

Взаимосвязь индексов

$$I_{qz} = I_z \times I_q. \quad (168)$$

Взаимосвязь абсолютных изменений

$$\Delta qz = \Delta qz(z) + \Delta qz(q). \quad (169)$$

Пример расчета индексов производственных затрат, себестоимости и объема производства продукции произведен на основании условных данных о производстве и себестоимости продукции (табл. 40).

Исходные условные данные для расчета индексов

Виды изделий	Произведено		Себестоимость единицы, рублей		Затраты на производство, рублей		
	базисный период	отчетный период	базисный период	отчетный период	базисный период	отчетный период	условные
	q_0	q_1	z_0	z_1	$q_0 z_0$	$q_1 z_1$	$q_1 z_0$
А, кг.	280	210	20	22	5600	4620	4200
Б, м.	160	300	30	15	4800	4500	9000
В, шт.	320	180	40	28	12800	5040	7200
Итого	-	-	-	-	23200	14160	20400

Индивидуальные индексы себестоимости единицы продукции:

$$\text{- изделие А } i_{z(A)} = \frac{z_1}{z_0} = \frac{22}{20} = 1,10;$$

$$\text{- изделие Б } i_{z(B)} = \frac{z_1}{z_0} = \frac{15}{30} = 0,50;$$

$$\text{- изделие В } i_{z(B)} = \frac{z_1}{z_0} = \frac{28}{40} = 0,70.$$

Расчет показал, что в отчетном периоде по сравнению с базисным периодом себестоимость единицы продукции А возросла на 10%, себестоимость единицы продукции Б снизилась на 50%, себестоимость единицы продукции В снизилась на 30%.

Индивидуальные индексы физического объема производства:

$$\text{- изделие А } i_{q(A)} = \frac{q_1}{q_0} = \frac{210}{280} = 0,75;$$

$$\text{- изделие Б } i_{q(B)} = \frac{q_1}{q_0} = \frac{300}{160} = 1,875;$$

$$\text{- изделие В } i_{q(B)} = \frac{q_1}{q_0} = \frac{180}{320} = 0,562.$$

Расчет показал, что в отчетном периоде по сравнению с базисным периодом объем производства продукции А снизился на 25%, объем производства продукции Б увеличился на 87,5%, объем производства продукции В снизился на 43,8%.

Общий индекс затрат на производство продукции

$$I_{qz} = \frac{\Sigma q_1 z_1}{\Sigma q_0 z_0} = \frac{14160}{23200} = 0,61.$$

Расчет показал, что общие затраты на производство всей группы продукции в отчетном периоде по сравнению с базисным периодом снизились на 39%.

Общий индекс себестоимости продукции

$$I_z = \frac{\Sigma q_1 z_1}{\Sigma q_1 z_0} = \frac{14160}{20400} = 0,694.$$

Расчет показал, что себестоимость в целом по всей группе продукции в отчетном периоде по сравнению с базисным периодом снизилась на 30,6%.

Общий индекс физического объема производства

$$I_q = \frac{\Sigma q_1 z_0}{\Sigma q_0 z_0} = \frac{20400}{23200} = 0,879.$$

Расчет показал, что объем производства в целом по всей группе продукции в отчетном периоде по сравнению с базисным периодом снизился на 12,1%.

Абсолютное изменение производственных затрат:

- всего $\Delta qz = \Sigma q_1 z_1 - \Sigma q_0 z_0 = 14160 - 23200 = -9040$ рублей.

- в том числе за счет изменения себестоимости единицы продукции

$\Delta qz(z) = \Sigma q_1 z_1 - \Sigma q_1 z_0 = 14160 - 20400 = -6240$ рублей.

- в том числе за счет изменения объема производства продукции

$\Delta qz(q) = \Sigma q_1 z_0 - \Sigma q_0 z_0 = 20400 - 23200 = -2800$ рублей.

Взаимосвязь индексов

$$I_{qz} = I_z \times I_q = 0,61 = 0,694 * 0,879.$$

Взаимосвязь абсолютных изменений

$$\Delta qz = \Delta qz(z) + \Delta qz(q) = -9040 = -6240 + (-2800).$$

Расчет показал, что в отчетном периоде по сравнению с базисным затраты на производство всей группы продукции уменьшились на 9040 рублей, в том числе снижение себестоимости единицы продукции привело к уменьшению производственных затрат на 6240 рублей, а уменьшение объема производства

по всей группе продукции привело к снижению производственных затрат на 2800 рублей.

Индексы производительности труда - индексы затрат труда, трудоемкости продукции и физического объема производства.

Индексы производительности труда имеют особенность расчета по сравнению с другими экономическими индексами. Особенность обусловлена несопоставимостью трудовых затрат на производство разнородных видов продукции. С целью приведения их в сопоставимый вид индексы производительности труда, как правило, определяются не на основании прямых показателей – уровней производительности труда, а на основе обратных показателей – трудоемкости производства продукции. Поскольку в основу положены обратные показатели, то индексы также определяются в обратном порядке – в числителе индексного соотношения принимаются базисные уровни трудоемкости продукции, а в знаменателе индексного соотношения принимаются отчетные уровни трудоемкости продукции. При этом, выводы делаются об изменении производительности труда.

Индексы затрат труда, трудоемкости и физического объема производства включают следующую систему индексов.

Индивидуальные индексы трудоемкости отдельного вида продукции

$$i_t = \frac{t_0}{t_1}. \quad (170)$$

Индивидуальный индекс трудоемкости показывает, на сколько процентов или во сколько раз увеличилась, либо уменьшилась производительность труда на производстве каждого вида продукции в отчетном периоде по сравнению с базисным.

Индивидуальные индексы физического объема производства отдельного вида продукции

$$i_q = \frac{q_1}{q_0}. \quad (171)$$

Индивидуальный индекс физического объема производства показывает, на сколько процентов или во сколько раз увеличился, либо уменьшился объем производства каждого вида продукции в отчетном периоде по сравнению с базисным.

Общий (сводный) индекс затрат труда

$$I_{qt} = \frac{\sum t_1 q_1}{\sum t_0 q_0}. \quad (172)$$

Общий индекс затрат труда показывает, на сколько процентов или во сколько раз увеличились, либо уменьшились затраты труда на производстве всей группы продукции в отчетном периоде по сравнению с базисным.

Общий (сводный) – натурально-трудовой индекс производительности труда

$$I_t = \frac{\sum q_1 t_0}{\sum q_1 t_1}. \quad (173)$$

Общий (сводный) индекс производительности труда показывает, на сколько процентов или во сколько раз увеличилась, либо уменьшилась производительность труда в целом на производстве всей группы продукции.

Общий (сводный) индекс физического объема производства

$$I_q = \frac{\sum q_1 t_0}{\sum q_0 t_0}. \quad (174)$$

Общий (сводный) индекс физического объема производства показывает, на сколько процентов или во сколько раз увеличился, либо уменьшился объем производства в целом по всей группе продукции.

На основании исходных индексных соотношений определяются абсолютные изменения затрат труда, производительности труда и физического объема производства.

На основании общего индекса затрат труда определяется их абсолютное изменение

$$\Delta qt = \sum q_1 t_1 - \sum q_0 t_0. \quad (175)$$

Абсолютное изменение показывает, на сколько человеко-часов увеличились либо уменьшились затраты труда на производство всей группы продукции в отчетном периоде по сравнению с базисным.

На основании общего индекса производительности труда определяется абсолютная экономия или перерасход рабочего времени на производстве всей группы продукции за счет изменения производительности труда

$$\Delta qt(t) = \Sigma q_1 t_0 - \Sigma q_1 t_1. \quad (176)$$

Абсолютное изменение показывает, на сколько человеко-часов увеличились либо уменьшились затраты рабочего времени на производство всей группы продукции в отчетном периоде по сравнению с базисным за счет изменения производительности труда (трудоемкости единицы продукции).

На основании общего индекса физического объема производства определяется изменение затрат труда за счет изменения объема производства продукции

$$\Delta qt(q) = \Sigma q_1 t_0 - \Sigma q_0 t_0. \quad (177)$$

Абсолютное изменение показывает, на сколько человеко-часов увеличились либо уменьшились затраты труда в отчетном периоде по сравнению с базисным за счет изменения объема производства всей группы продукции.

Взаимосвязь индексов

$$I_{qt} = I_t \times I_q. \quad (178)$$

Взаимосвязь абсолютных изменений:

$$\Delta qt = \Delta qt(t) + \Delta qt(q). \quad (179)$$

Пример расчета индексов затрат труда, производительности труда и физического объема производства продукции производится на основании условных данных о производстве продукции и затратах труда (табл. 41).

Исходные условные данные для расчета индексов

Вид продук ции	Базисный период		Отчетный период		Затраты труда, человеко-часов		
	произве дено, шт.	трудоемкос ть единицы, человеко- часов	произведе но, шт.	трудоемкос ть единицы, человеко- часов	базисный период	отчетный период	условн ые
	q ₀	t ₀	q ₁	t ₁	q ₀ t ₀	q ₁ t ₁	q ₁ t ₀
А	357	0,85	392	0,82	303,45	321,44	333,20
Б	135	0,93	158	0,91	125,55	143,78	146,94
В	238	1,21	215	1,15	287,98	247,25	260,15
Итого	-	-	-	-	413,53	712,47	740,29

Индивидуальные индексы производительности труда определяются отдельно по каждому виду продукции

$$\text{- изделие А } i = \frac{t_0}{t_1} = 0,85/0,82 = 1,0366,$$

$$\text{- изделие Б } i = \frac{t_0}{t_1} = 0,93/0,91 = 1,0219,$$

$$\text{- изделие В } i = \frac{t_0}{t_1} = 1,21/1,15 = 1,0522.$$

Расчет показал, что в отчетном периоде по сравнению с базисным производительность труда на производстве изделия А увеличилась на 3,66%, на производстве изделия Б – на 2,19%, на производстве изделия В – на 5,22%.

Общий натурально-трудовой индекс производительности труда

$$I_q = \frac{\sum t_0 q_1}{\sum t_1 q_1} = \frac{740,29}{712,47} = 1,0390.$$

Расчет показал, что в целом по всей группе продукции производительность труда в отчетном периоде по сравнению с базисным увеличилась на 3,9%.

Общий индекс затрат труда

$$I_q = \frac{\sum t_1 q_1}{\sum t_0 q_0} = \frac{712,47}{413,53} = 1,7229.$$

Расчет показал, что затраты труда на производстве всей группы продукции в отчетном периоде по сравнению с базисным увеличились на

72,29%.

Индекс физического объема производства продукции

$$I_q = \frac{\sum t_0 q_1}{\sum t_0 q_0} = \frac{740,29}{413,53} = 1,7902.$$

Расчет показал, что в отчетном периоде по сравнению с базисным физический объем производства по всей группе продукции увеличился на 79,02%.

Абсолютное изменение затрат труда в целом по всей группе продукции

$$\Delta qt = \Sigma q_1 t_1 - \Sigma q_0 t_0 = 712,47 - 413,53 = +298,94 \text{ человеко-часов.}$$

Расчет показал, что в отчетном периоде по сравнению с базисным общая сумма затрат труда увеличилась на 298,94 человеко-часов.

Экономия или перерасход затрат труда в результате изменения производительности труда

$$\mathcal{E}(\Pi) = \Sigma t_0 q_1 - \Sigma t_1 q_1 = 740,29 - 712,47 = +27,82 \text{ человеко-часов.}$$

Расчет показал, что в отчетном периоде по сравнению с базисным произошла экономия затрат труда на 27,82 человеко-часов в результате роста производительности труда.

Абсолютное изменение затрат труда за счет изменения физического объема производства

$$\Delta qt(q) = \Sigma q_1 t_0 - \Sigma q_0 t_0 = 740,29 - 413,53 = +326,76 \text{ человеко-часов.}$$

Расчет показал, что в отчетном периоде по сравнению с базисным затраты труда на производство всей группы продукции увеличились на 326,76 человеко-часов за счет увеличения объема производства продукции.

Взаимосвязь индексов

$$I_{qt} = I_t \times I_q = 1,7229 = 1,0390 * 1,7902.$$

Взаимосвязь абсолютных изменений

$$\Delta qt = \Delta qt(t) + \Delta qt(q) = +298,94 = -27,82 + 326,76.$$

Обобщение методики построения основных экономических индексов представлено в таблице 42.

Схема построения основных экономических индексов

Элементы агрегатного индекса	Индекс физического объема	Индекс цен	Индекс себестоимости	Индекс производительности труда
1.Индексируемая величина:				
- базисного периода	q_0	p_0	z_0	t_0
- отчетного периода	q_1	p_1	z_1	t_1
Индивидуальный индекс - i	$i_q = \frac{q_1}{q_0}$	$i_p = \frac{p_1}{p_0}$	$i_z = \frac{z_1}{z_0}$	$i_t = \frac{t_1}{t_0}$
2.Соизмеритель	p_0	q_1	q_1	q_1
Агрегатный индекс - I	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}$	$I_z = \frac{\sum q_1 z_1}{\sum q_1 z_0}$	$I_t = \frac{\sum q_1 t_1}{\sum q_1 t_0}$

Разновидности агрегатных индексов по используемым соизмерителям (статистическим весам) представлены в таблице 43. [22]

Агрегатные индексы

Название формулы	Вид индекса	
	Индекс физического объема и других объемных показателей	Индекс цен и других качественных показателей
Формула Ласпейреса (по базисным весам)	$I_{q(L)} = \frac{\sum q_1 p_0}{\sum q_0 p_0}$	$I_{p(L)} = \frac{\sum q_0 p_1}{\sum q_0 p_0}$
Формула Паше (по отчетным весам)	$I_{q(P)} = \frac{\sum q_1 p_1}{\sum q_0 p_1}$	$I_{p(P)} = \frac{\sum q_1 p_1}{\sum q_1 p_0}$
Формула Фишера	$I_q = \sqrt{I_{q(L)} \times I_{q(P)}}$	$I_p = \sqrt{I_{p(L)} \times I_{p(P)}}$

9 СТАТИСТИЧЕСКИЕ МЕТОДЫ ОРГАНИЗАЦИИ И ОЦЕНКИ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ

9.1 Содержание, задачи и значение выборочного метода

Любое статистическое исследование может осуществляться на основании данных а) сплошного наблюдения; б) несплошного наблюдения.

Несплошное наблюдение является основой применения выборочного метода. При выборочном методе обследованию подвергается часть всей изучаемой совокупности (как правило, 5-10% в зависимости от цели исследования) – выборочная совокупность. Выборочный метод включает совокупность математических обоснований, используемых при осуществлении несплошного наблюдения.

Следовательно, выборочный метод – это метод статистического исследования, при котором обобщающие показатели изучаемой (генеральной) совокупности с определенной вероятностью устанавливаются по ее части (выборочной совокупности).

Значение выборочного метода:

- 1) исследование проводится в более короткие сроки;
- 2) исследование проводится с минимальными затратами труда и средств;
- 3) повышается оперативность статистической информации;
- 4) уменьшается вероятность ошибок регистрации;
- 5) при помощи выборочного метода проверяется точность данных сплошного наблюдения;
- 6) минимальная численность обследуемых единиц позволяет провести исследование более тщательно и квалифицированно;
- 7) выборочный метод обеспечивает принцип равных возможностей попадания в обследование для каждой единицы изучаемой совокупности, что исключает возможность формирования совокупности только за счет лучших или худших

единиц и предупреждает возможность появления систематических ошибок наблюдения;

8) выборочный метод позволяет расширить программу статистического наблюдения;

9) в проведении некоторых исследований выборочный метод является единственно возможным (контроль качества продукции, если при этом продукция повреждается или уничтожается).

Содержание статистической методологии выборочного метода включает

а) способы определения ошибок выборки при различных приемах формирования выборочной совокупности; б) способы распространения характеристик выборки на всю изучаемую совокупность.

Цель выборочного метода:

1) расчет с заданной вероятностью пределов возможных отклонений характеристик всей исходной изучаемой совокупности от соответствующих характеристик выборки;

2) определение вероятности, с которой размер возможных отклонений характеристик всей исходной совокупности от соответствующих характеристик выборки не превысит заданного предела.

Выборочное наблюдение используется для решения следующих задач.

1. Оценка и проверка статистических гипотез. Решение задачи оценки и проверки статистических гипотез связано с решением общетеоретических проблем и проведением экспериментов. Включает а) поиск оптимальных выборочных параметров для отображения свойств изучаемой генеральной совокупности; б) формулировку статистических гипотез о характеристиках генеральной совокупности и последующую их проверку при помощи результатов выборочного метода.

2. Принятие производственных и управленческих решений. Решение задачи принятия управленческих решений связано с практическими интересами в области управления технологическими процессами, качеством продукции, работ и услуг. Включает а) измерение параметров и разработку показателей

качества; б) установление номинального режима, отклонение от которого подлежит статистической оценке; в) поиск оптимального режима и альтернативных технологий на базе анализа производимых замеров; г) управление допущениями при помощи контрольных карт Шухарта или кумулятивных сумм, разработанных на основе стандартов.

3. Отраслевые и социально-экономические исследования. Решение задач отраслевых и социально-экономических исследований связано с изучением использования оборудования, рабочего времени и других параметров.

4. Исследования в сфере предпринимательской деятельности. Решение задач исследования в сфере предпринимательской деятельности связано с прогнозированием объемов производства и продаж, оценкой рынков сбыта продукции и маркетинговых исследований.

5. Демографическое обследование населения. Решение задач демографического обследования населения связано с расширением программы переписей населения, с получением предварительных результатов переписей, с контролем качества проведения переписей населения, с получением важнейших показателей в межпереписной период.

6. Обследование социально-экономического положения населения. Решение задач обследования социально-экономического положения населения связано с изучением доходов населения и источников их формирования, уровня потребления материальных благ и услуг, жилищных условий, занятости населения, а также с выборочным обследованием бюджетов семей, характера использования рабочего и нерабочего времени, перемещений по территории страны и за ее пределами, заболеваемости, политической активности и др.

7. Исследование потребительского спроса на товары и услуги. Решение задачи исследования потребительского спроса на товары и услуги связано с использованием анкетирования, изучением сегментации рынка, позиционирования отдельных товаров и услуг и другие.

9.2 Генеральная и выборочная совокупности, их характеристики

При выборочном наблюдении изучению подвергаются два вида статистических совокупностей: 1) генеральная совокупность; 2) выборочная совокупность.

Генеральная совокупность – это вся совокупность единиц, которая изучается и из которой производится выборка. Численность генеральной совокупности обозначается N .

Выборочная совокупность – это часть генеральной совокупности, которая отобрана и подвергается выборочному обследованию. Численность выборочной совокупности обозначается n .

Цель выборочного метода заключается в получении правильного представления о показателях всей генеральной совокупности на основе показателей выборочной совокупности.

При выборочном методе определяется и изучается две категории обобщающих показателей: 1) относительные величины; 2) средние величины.

Относительные величины применяются для обобщающей характеристики совокупности по альтернативному признаку. К относительным величинам в данном случае относится доля единиц совокупности, которые обладают изучаемым признаком (например, при изучении совокупности студентов можно определить долю студентов-отличников, долю студентов, получающих стипендию и т.д.). Для генеральной совокупности этот показатель называется генеральной долей (долей в генеральной совокупности). Для выборочной совокупности этот показатель называется выборочной долей. Цель выборочного метода в данном случае заключается в определении с заданной вероятностью значения генеральной доли на основании значения выборочной доли.

Средние величины применяются для измерения среднего значения варьирующего признака (например, средней цены реализации товара, среднего возраста населения страны, среднего балла успеваемости студентов и др.). Для

генеральной совокупности этот показатель называется генеральной средней. Для выборочной совокупности этот показатель называется выборочной средней. Цель выборочного метода в данном случае заключается в определении с заданной вероятностью значения генеральной средней на основании значения выборочной средней.

Количественный и альтернативный признаки могут также быть охарактеризованы генеральной и выборочной дисперсией, генеральным и выборочным средним квадратическим отклонением и другими статистическими характеристиками.

Основные характеристики параметров выборочной и генеральной совокупностей представлены в таблице 44.

Таблица 44

Характеристики параметров генеральной и выборочной совокупностей

Характеристика	Генеральная совокупность	Выборочная совокупность
Объем совокупности (численность единиц)	N	n
Численность единиц, обладающих изучаемым признаком	M	m
Доля единиц, обладающих изучаемым признаком	P	w
Средний размер признака	\bar{X}	\tilde{X}
Дисперсия количественного признака	$\sigma_{\bar{X}}^2$	$\sigma_{\tilde{X}}^2$
Дисперсия альтернативного признака (дисперсия доли)	σ_p^2	S_w^2

9.3 Понятие репрезентативности выборки. Способы отбора, обеспечивающие репрезентативность

Состав выборочной совокупности не совпадает с составом генеральной совокупности, поэтому выборочные характеристики отличаются от соответствующих генеральных характеристик. Следовательно, основная цель выборочного метода заключается в максимальном приближении значений характеристик выборочной совокупности к значениям характеристик

генеральной совокупности. Данная задача выполняется путем соблюдения репрезентативности выборочной совокупности (репрезентативности выборки).

Репрезентативность (представительность) – это способность отображать, воспроизводить.

Для соблюдения принципа репрезентативности в статистике применяются различные способы, методы и виды формирования выборочной совокупности (рис. 52, 53, 54).

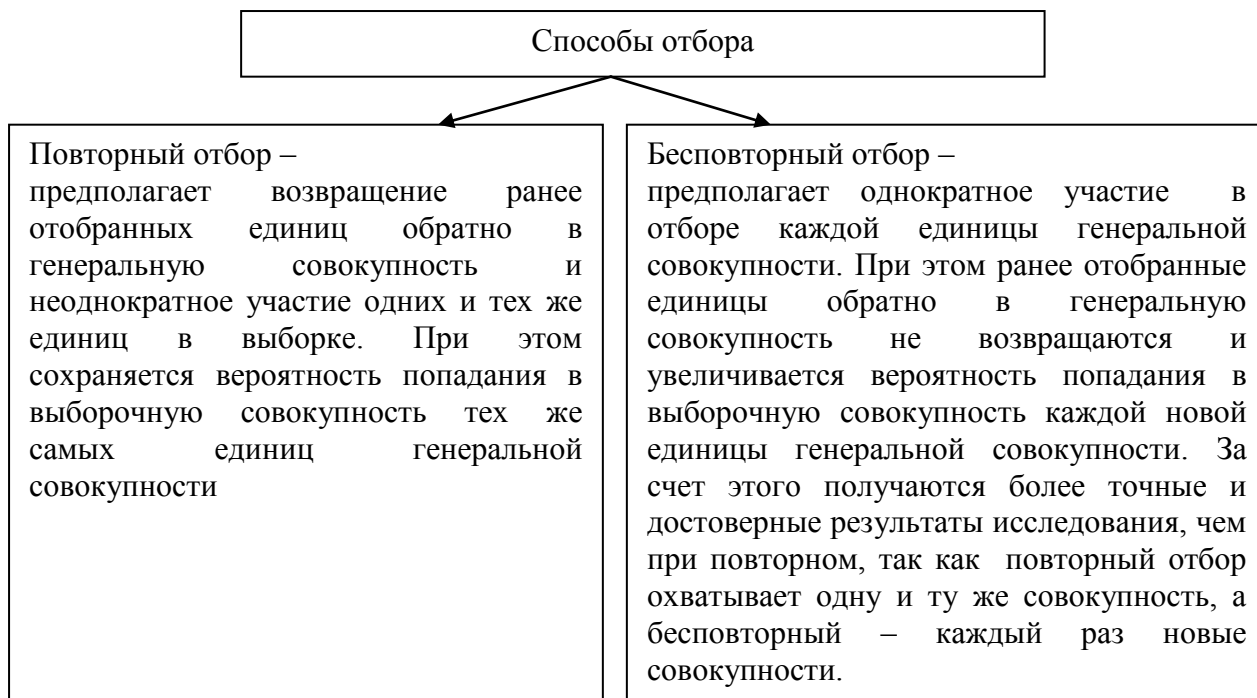


Рисунок 52 – Способы отбора, обеспечивающие репрезентативность выборки



Рисунок 53 – Виды отбора, обеспечивающие репрезентативность выборки

Метод отбора – это система организации отбора единиц в выборочную совокупность из генеральной совокупности. Методы отбора определяются правилами формирования выборочной совокупности и включают следующие (рис. 54).

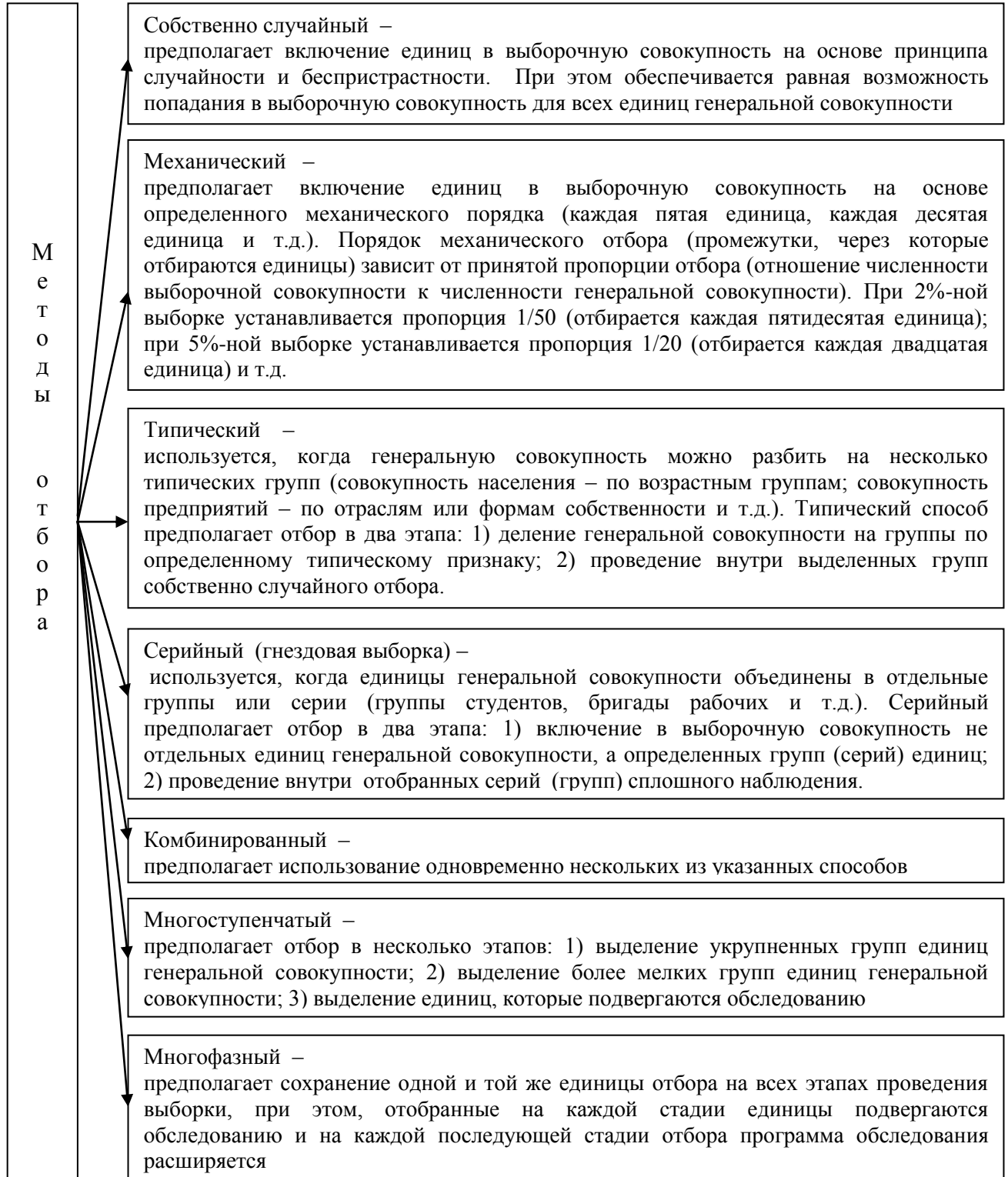


Рисунок 54 – Методы отбора, обеспечивающие репрезентативность выборки

При соблюдении принципа одинаковых возможностей попадания в выборку для каждой единицы генеральной совокупности наиболее достоверные результаты дает собственно случайный способ отбора.

9.4 Ошибки выборки

Ошибки выборки (ошибки репрезентативности) – это возможные пределы отклонений характеристик выборочной совокупности от характеристик генеральной совокупности. Существует две основных причины возникновения ошибок выборки (репрезентативности): 1) отличие состава выборочной совокупности от состава генеральной совокупности; 2) различие в вариации признака в выборочной и генеральной совокупностях.

Величина ошибки выборки (ошибки репрезентативности) – это разность между величиной характеристики выборочной совокупности и величиной характеристики генеральной совокупности.

Ошибки выборки (репрезентативности) возникают под влиянием следующих факторов: 1) степень вариации изучаемого признака; 2) численность выборочной совокупности; 3) методы отбора; 4) принятый уровень достоверности результатов исследования (принятый уровень вероятности).

Ошибки выборки (ошибки репрезентативности) принимают два основных вида (рис. 55).

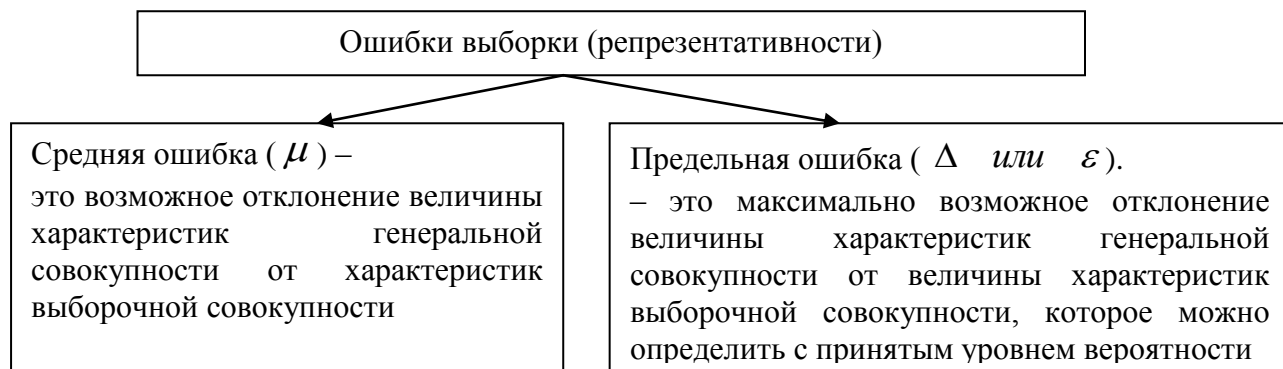


Рисунок 55 – Виды ошибок выборки (репрезентативности)

Величина средней ошибки выборки зависит от двух факторов:

- 1) от численности выборочной совокупности (зависимость обратная – чем больше численность выборочной совокупности, тем меньше величина средней ошибки выборки);
- 2) от степени вариации изучаемого признака, которая характеризуется дисперсией (зависимость прямая – чем больше степень вариации признака, тем больше величина средней ошибки выборки).

Средняя ошибка рассчитывается отдельно для генеральной средней и для генеральной доли. При этом средняя ошибка рассчитывается по-разному в зависимости от метода и способа отбора (табл. 45).

Таблица 45

Формулы расчета средней ошибки выборки при некоторых способах и методах формирования выборочной совокупности

Метод отбора	Способ отбора			
	повторный		бесповторный	
	для средней	для доли	для средней	для доли
Собственно-случайный и механический	$\sqrt{\frac{\sigma^2}{n}}$	$\sqrt{\frac{w(1-w)}{n}}$ или $\sqrt{\frac{qp}{n}}$	$\sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$
Типический	$\sqrt{\frac{\bar{\sigma}_i^2}{n}}$	$\sqrt{\frac{w_i(1-w_i)}{n}}$	$\sqrt{\frac{\bar{\sigma}_i^2}{n} \left(1 - \frac{n}{N}\right)}$	$\sqrt{\frac{w_i(1-w_i)}{n} \left(1 - \frac{n}{N}\right)}$
Серийный	$\sqrt{\frac{\delta_x^2}{r}}$	$\sqrt{\frac{\delta_w^2}{r}}$	$\sqrt{\frac{\delta_x^2}{r} \left(\frac{R-r}{R-1}\right)}$	$\sqrt{\frac{\delta_w^2}{r} \left(\frac{R-r}{R-1}\right)}$

где σ^2 - дисперсия;

n – численность выборочной совокупности;

N – численность генеральной совокупности;

w – доля единиц, обладающих изучаемым признаком;

qp – дисперсия альтернативного признака;

$\bar{\sigma}^2$ - средняя из внутригрупповых дисперсий;

δ^2 - внутригрупповая дисперсия;

R – общее число серий (групп);

r - число отобранных серий.

Величина предельной ошибки зависит от двух факторов:

- 1) от величины средней ошибки выборки;
- 2) от уровня принятой вероятности (уровня значимости).

Предельная ошибка определяется по формуле

$$\Delta(\varepsilon) = \mu \times t, \quad (180)$$

где t – коэффициент кратности ошибки выборки (коэффициент доверия).

Коэффициент доверия t зависит от уровня вероятности, с которой можно утверждать, что предельная ошибка не превысит t -кратную среднюю ошибку. Также коэффициент кратности ошибки называется нормированным отклонением, которое функционально взаимосвязано с вероятностью. С увеличением t увеличивается предельная ошибка выборки и увеличивается вероятность, с которой определяется величина генеральной характеристики. Следовательно, чем больше пределы, в которых допускается возможная ошибка, тем с большей вероятностью можно установить ее величину. Коэффициент кратности определяется по таблицам интегральной функции Лапласа (таблица 46). [22]

Таблица 46

Отдельные значения коэффициента t

Вероятность, $\Phi(t)$	Коэффициент кратности, t
0,683	1
0,866	1,5
0,954	2
0,988	2,5
0,997	3
0,999	3,5

9.5 Статистические оценки характеристик генеральной совокупности

Конечной целью выборочного метода является перенос его результатов на всю генеральную совокупность. Следовательно, статистические оценки – это распространение результатов выборки на генеральную совокупность. При

статистическом исследовании для обеспечения достоверности статистических оценок производится оценка результатов выборочного наблюдения с точки зрения возможности их распространения на генеральную совокупность. Возможность распространения зависит от качества выборки.

Качество выборки оценивается по следующим критериям: 1) полнота выборки; 2) относительная ошибка выборки; 3) представительность (репрезентативность) выборки.

Полнота выборки – это наличие (представленность) всех типов или групп единиц генеральной совокупности в выборочной совокупности.

Относительная ошибка выборки определяется по формулам

$$\text{- для средней} \quad \Delta_{\%} = \frac{\Delta_{\tilde{X}}}{\tilde{X}} \times 100, \quad (181)$$

$$\text{- для доли} \quad \Delta_{\%} = \frac{\Delta_w}{w} \times 100, \quad (182)$$

где $\Delta_{\%}$ - относительная предельная ошибка выборки;

$\Delta_{\tilde{X}}$ - предельная ошибка для средней;

Δ_w - предельная ошибка для доли;

\tilde{X} - выборочная средняя;

w – выборочная доля.

Если величина относительной ошибки не превышает заранее установленного для данного обследования предельного значения, то результаты выборочного обследования могут быть распространены на генеральную совокупность.

Представительность (репрезентативность) выборки определяется на основе относительной ошибки выборки. Если величина относительной ошибки не превышает заранее установленного для данного обследования предельного значения, то выборка является представительной (репрезентативной). Если выборка не представительна, то производится восстановление ее пропорций на основе исходной информации при помощи корректировки выборки.

Способы корректировки включают два способа: 1) способ отсечения; 2) способ взвешивания.

Способ отсечения заключается в том, что при недостаточной представленности отдельных типов или групп единиц производятся дополнительные расчеты, на основе которых определяется число хорошо представленных единиц, часть которых исключается из выборки с целью сохранения пропорций генеральной совокупности. При этом определяются характеристики выборки до и после отсечения. Для проверки используется принцип оценки точности выборки. Если расхождения между характеристиками, рассчитанными до и после отсечения не превышают $\pm 5\%$, итоги корректировки считаются удовлетворительными.

Способ взвешивания заключается в том, что часть формуляров, численность которых мала, используется многократно на основе специально рассчитанных весов.

Статистические оценки производятся на основании данных выборки, в которой соблюдаются основные принципы формирования выборочной совокупности:

- 1) принцип равных возможностей попадания в выборку для всех единиц генеральной совокупности;
- 2) принцип соблюдения достаточно большого числа единиц выборочной совокупности.

Статистические оценки производятся двумя способами распространения характеристик выборочной совокупности на генеральную совокупность: 1) способом прямого пересчета; 2) способом коэффициентов.

Способ прямого пересчета заключается в умножении характеристик выборочной совокупности на число единиц (объем) генеральной совокупности. При этом значения выборочных характеристик распространяются на генеральную совокупность с учетом ошибок выборки.

Способ коэффициентов заключается в том, что в данные сплошного обследования вносятся соответствующие поправки, относительный размер

которых устанавливается по результатам выборочного обследования. При этом данные поправки вносятся на основании сопоставления данных сплошного и выборочного обследования. В результате использования данного способа проверяются и уточняются результаты сплошного наблюдения.

При уточнении данных сплошного наблюдения на основе контрольных выборочных мероприятий определяется поправка на недоучет (в небольших совокупностях, объем которых не превышает несколько сотен или тысяч единиц). Численность совокупности с поправкой на недоучет определяется по формуле

$$N_1 = N_0 \times \frac{n_1}{n_0}, \quad (183)$$

где N_1 – численность совокупности с поправкой на недоучет;

N_0 – первоначальная численность совокупности;

n_1 – численность совокупности в контрольных точках по первоначальным данным;

n_0 – численность совокупности в контрольных точках по данным контрольных мероприятий.

В условиях существования большого числа факторов, влияющих на точность выборочного обследования, использование точечной оценки при распространении выборочных характеристик на генеральную совокупность не дает достаточно достоверных результатов. Поэтому для достижения достоверности результатов статистических оценок используется интервальная оценка. Интервальная оценка производится путем распространения результатов выборочного обследования на генеральную совокупность на основании определения доверительных интервалов. Доверительный интервал – это пределы, в которых с установленной степенью вероятности может принимать значение генеральная характеристика. Доверительный интервал определяется путем корректировки значений выборочных характеристик на величину предельной ошибки выборки.

Определение доверительного интервала предполагает расчет его границ. Границы доверительного интервала определяются по формулам

$$\text{- для средней} \quad \bar{X} = \tilde{X} \pm \Delta_{\bar{x}}, \quad (184)$$

$$\text{- для доли} \quad p = w \pm \Delta_w, \quad (185)$$

где \bar{X} - генеральная средняя;

\tilde{X} - выборочная средняя;

$\Delta_{\bar{x}}$ - предельная ошибка для средней;

Δ_w - предельная ошибка для доли;

P – генеральная доля;

w – выборочная доля.

Следовательно, доверительный интервал примет вид

$$\text{- для средней} \quad \tilde{X} - \Delta_{\bar{x}} \leq \bar{X} \leq \tilde{X} + \Delta_{\bar{x}}, \quad (186)$$

$$\text{- для доли} \quad w - \Delta_w \leq P \leq w + \Delta_w. \quad (187)$$

9.6 Определение необходимой численности выборочной совокупности

При проектировании выборочного наблюдения сначала определяется необходимая численность выборочной совокупности (объем выборки).

Необходимая численность выборочной совокупности (объем выборки) зависит от четырех основных факторов: 1) величины допустимой ошибки выборки; 2) вероятности, с которой можно гарантировать величину устанавливаемой ошибки выборки; 3) способа отбора; 4) уровня точности выборочной совокупности с заданной вероятностью.

При определении необходимой численности выборки с целью обеспечения ее представительности (репрезентативности) рассчитывается показатель вариации изучаемого признака.

Вариация количественного признака определяется приближенно при помощи показателя среднего квадратического отклонения, рассчитанного по

величине предполагаемого размаха вариации или среднего линейного отклонения (при условии максимального приближения фактического распределения к нормальному). Среднее квадратическое отклонение определяется по формулам

$$\sigma = \frac{R}{6} \quad (188)$$

или

$$\sigma = 1,25 \times \bar{d}, \quad (189)$$

где σ - среднее квадратическое отклонение;

R – размах вариации;

\bar{d} - среднее линейное отклонение.

Вариация качественного признака рассчитывается по альтернативному признаку. При необходимости неальтернативный признак преобразуется в альтернативный при условии, что доля неучтенных единиц во всей совокупности незначительна. Дисперсия при этом определяется по формуле

$$\sigma_w^2 = w \times (1 - w), \quad (190)$$

где w – доля единиц, обладающих изучаемым признаком в выборочной совокупности (приблизительно принимается 0,5)

Необходимая численность выборки определяется по-разному для средней и для доли (табл. 47).

Формулы расчета необходимой численности выборки (табл. 47) показывают, что:

1) с увеличением допустимой ошибки выборки необходимый объем выборки уменьшается;

2) с увеличением вероятности, с которой гарантируется характеристика генеральной совокупности, необходимый объем выборки увеличивается.

Формулы для определения необходимой численности выборки при некоторых способах формирования выборочной совокупности [22]

Метод формирования выборочной совокупности	Способ отбора	
	повторный отбор	бесповторный отбор
Собственно-случайная выборка: - для средней	$n = \frac{t^2 \times \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \times \sigma_{\bar{x}}^2 \times N}{\Delta_{\bar{x}}^2 \times N + t^2 \times \sigma_{\bar{x}}^2}$
- для доли	$n = \frac{t^2 \times w \times (1 - w)}{\Delta_w^2}$	$n = \frac{t^2 \times w \times (1 - w) \times N}{\Delta_w^2 \times N + t^2 \times w \times (1 - w)}$
Механическая выборка	То же	То же
Типическая выборка - для средней	$n = \frac{t^2 \times \bar{\sigma}_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$n = \frac{t^2 \times \bar{\sigma}_{\bar{x}}^2 \times N}{\Delta_{\bar{x}}^2 \times N + t^2 \times \bar{\sigma}_{\bar{x}}^2}$
- для доли	$n = \frac{t^2 \times w \times (1 - w)}{\Delta_w^2}$	$n = \frac{t^2 \times w \times (1 - w) \times N}{\Delta_w^2 \times N + t^2 \times w \times (1 - w)}$
Серийная выборка - для средней	$r = \frac{t^2 \times \delta_{\bar{x}}^2}{\Delta_{\bar{x}}^2}$	$r = \frac{t^2 \times \delta_{\bar{x}}^2 \times R}{\Delta_{\bar{x}}^2 \times R + t^2 \times \delta_{\bar{x}}^2}$
- для доли	$r = \frac{t^2 \times w_r \times (1 - w_r)}{\Delta_w^2}$	$r = \frac{t^2 \times w_r \times (1 - w_r) \times R}{\Delta_w^2 \times R + t^2 \times w_r \times (1 - w_r)}$

9.7 Особенности расчетов при малой выборке

От численности выборочной совокупности (объема выборки) зависит величина пределов, которые с заданной вероятностью не превзойдет ошибка выборки, а также способы определения этих пределов. По теореме А.М.Ляпунова – при большом числе единиц выборочной совокупности ($n \geq 100$) распределение случайных ошибок выборочной средней нормально или приближается к нормальному по мере увеличения числа наблюдений. Вероятность выхода ошибки за установленные пределы оценивается на основе таблиц интеграла Лапласа.

В статистических исследованиях зачастую используются небольшие по объему выборки – малые выборки. Малая выборка - это выборочное наблюдение, численность единиц которого не превышает 30 единиц.

Разработана теория малой выборки английским статистиком В.С.Госсетом (псевдоним – Стьюдент), который доказал, что оценка расхождения между средней малой выборки и генеральной средней имеет особый закон распределения. При оценке результатов малой выборки величина генеральной дисперсии в расчетах не используется. Поэтому для определения возможных пределов ошибки выборки в малых выборках используется критерий Стьюдента, который определяется по формуле

$$t = \frac{\tilde{X} - \bar{X}}{\mu_{\text{мб}}}, \quad (191)$$

где $\mu_{\text{мб}}$ - средняя ошибка малой выборки (мера случайных колебаний выборочной средней в малой выборке);

Средняя ошибка малой выборки в данном случае определяется по формуле

$$\mu_{\text{мб}} = \frac{\sigma}{\sqrt{n-1}}. \quad (192)$$

где σ - среднее квадратическое отклонение.

Среднее квадратическое отклонение в малой выборке определяется по формуле

$$\sigma = \sqrt{\frac{\sum (X_i - \tilde{X})^2}{n}}. \quad (193)$$

Предельная ошибка малой выборки определяется по формуле

$$\Delta_{\text{мб}} = t \times m_{\text{мб}}. \quad (194)$$

Коэффициент кратности ошибки t в данном случае определяется по таблицам распределения Стьюдента (табл. 48).

Распределение вероятности в малых выборках [22]

Коэффициент доверия, t	Численность выборочной совокупности, n									
	4	5	6	7	8	9	10	15	20	∞
0,5	0,348	0,356	0,362	0,366	0,368	0,370	0,372	0,376	0,378	0,383
1,0	0,608	0,626	0,636	0,644	0,650	0,654	0,656	0,666	0,670	0,683
1,5	0,770	0,792	0,806	0,816	0,832	0,828	0,832	0,846	0,850	0,865
2,0	0,860	0,884	0,908	0,908	0,914	0,920	0,924	0,936	0,940	0,954
2,5	0,933	0,946	0,955	0,959	0,963	0,966	0,968	0,975	0,978	0,988
3,0	0,942	0,960	0,970	0,976	0,980	0,938	0,984	0,992	0,992	0,977

Таблица вероятностей Стьюдента также может быть представлена в другом виде, с учетом числа степеней свободы. Числом степеней это - разность между количеством наблюдений и количеством уже рассчитанных независимо друг от друга коэффициентов (констант). (табл. 49).

Таблица 49

Значения t-критерия распределения Стьюдента [22]

Число степеней свободы, k	Значения t-критерия			
	для одностороннего интервала		для двустороннего интервала	
	P=0,95	P=0,99	P=0,95	P=0,99
3	2,35	4,54	3,18	5,84
4	2,13	3,75	2,78	4,60
5	2,01	3,37	2,57	4,03
6	1,94	3,14	2,45	3,71
7	1,89	3,00	2,36	3,50
8	1,86	2,90	2,31	3,36
9	1,83	2,82	2,26	3,35
10	1,81	2,76	2,23	3,17
15	1,75	2,60	2,13	2,95
20	1,73	2,53	2,09	2,85
30	1,70	2,46	2,04	2,75
60	1,67	2,39	2,00	2,66
∞	1,64	2,33	1,96	2,58

Число степеней свободы определяется по формуле

$$k = n - 1. \quad (195)$$

В результате определяется доверительный интервал малой выборки – область значений генеральной характеристики, выход за пределы которой имеет очень малую вероятность. Данная вероятность определяется по формуле

$$q = 1 - P. \quad (196)$$

В качестве доверительной вероятности, как правило, при двустороннем интервале используется $P = 0,95$ или $P = 0,99$. Следовательно, вероятности случайного выхода оцениваемой характеристики генеральной совокупности за пределы доверительного интервала соответственно будут равны $q = 0,05$ и $q = 0,01$. То есть – маловероятны.

Расчет средней и предельной ошибок в малых выборках аналогичен расчетам в больших выборках (формулы представлены в табл. 45).

9.8 Расчет оценок генеральной совокупности по характеристикам выборочной совокупности

Пример расчета характеристик генеральной совокупности на основе характеристик выборочной совокупности при повторном способе отбора.

В процессе анализа качества продукции молочного комбината в результате случайной повторной выборки взята 221 проба молока на жирность. Заданная вероятность результатов анализа $P = 0,997$. При этом получены следующие данные (табл. 50).

Таблица 50

Исходные данные по выборочной совокупности

Средний процент жира в молоке, %	3,0	3,2	3,4	3,8	4,0	4,5	Итого
Количество проб молока	32	48	65	35	26	15	221

Вспомогательные расчеты для определения выборочных характеристик: средней и среднего квадратического отклонения представлены в таблице 51.

Выборочная средняя (средний процент жира в молоке по данным выборочной совокупности) определяется по формуле средней арифметической взвешенной. Выборочная средняя составила

$$\bar{X} = \frac{\sum X \times f}{\sum f} = X \frac{3,0 * 32 + 3,2 * 48 + 3,4 * 65 + 3,8 * 35 + 4,0 * 26 + 4,5 * 15}{32 + 48 + 65 + 35 + 26 + 15} = \frac{775,1}{221} = 3,51\%,$$

где X – процент жира в молоке;

f – число проб.

Таблица 51

Расчет выборочной средней и среднего квадратического отклонения

Средний процент жира в молоке, %	Число проб, f	X^2	$X^2 \cdot f$
3,0	32	9,00	288,00
3,2	48	10,24	491,52
3,4	65	11,56	751,40
3,8	35	14,44	505,40
4,0	26	16,00	416,00
4,5	15	20,25	303,75
Итого	221	-	2756,07

Среднее квадратическое отклонение в выборочной совокупности составило $\sigma = \sqrt{\frac{\sum X^2 \times f}{\sum f} - (\bar{X})^2} = \sqrt{\frac{2756,07}{221} - (3,51)^2} = 0,388 \%$.

Пределы, в которых заключается средний процент жира в молоке в генеральной совокупности (генеральная средняя) определяются по формуле 186, для чего определяются средняя и предельная ошибки выборки.

Средняя ошибка выборки для средней при случайном повторном отборе составила $\mu = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{0,1505}{221}} = 0,0261 \%$.

Предельная ошибка для средней составила $\Delta = \mu \times t = 0,0261 \times 3 = 0,0783 \%$, где t – коэффициент кратности ошибки (определяется по таблице 46, при вероятности 0,997 $t = 3$).

Пределы генеральной средней составили

- нижний предел $X_r = 3,51 - 0,0783 = 3,4317\%$;

- верхний предел $X_r = 3,51 + 0,0783 = 3,5883\%$.

Расчет показал, что с вероятностью 0,997 можно утверждать, что в генеральной совокупности средний процент жира в молоке составит от 3,4317% до 3,5883%.

Доля проб молока со средним процентом жирности менее 3,5% по данным выборочной совокупности (выборочная доля) составила:

$$w = (32 + 48 + 65) / 221 = 0,656.$$

Средняя ошибка выборки для доли при случайном повторном отборе составила $\mu = \sqrt{\frac{w \times (1 - w)}{n}} = \sqrt{\frac{0,656 \times (1 - 0,656)}{221}} = 0,032$,

где n – численность выборочной совокупности.

Предельная ошибка для доли составила $\Delta = \mu \times t = 0,032 \times 2 = 0,064$,

где t – коэффициент кратности ошибки (определяется по таблице 46, при вероятности 0,9545 $t = 2$).

Пределы генеральной доли составили

- нижний предел $p = 0,656 - 0,064 = 0,592$;

- верхний предел $p = 0,656 + 0,064 = 0,720$.

Расчет показал, что с вероятностью 0,954 можно утверждать, что доля проб молока с жирностью менее 3,5% в генеральной совокупности составит от 0,592 до 0,720 (от 59,2% до 72%).

Необходимая численность выборки для определения генеральной средней при случайном повторном отборе определяется по формуле

$$n = \frac{t^2 \times \sigma_{\bar{x}}^2}{\Delta_{\bar{x}}^2}. \quad (197)$$

Необходимая численность выборки составила

$$n = \frac{3^2 \times 0,388^2}{0,0783^2} = \frac{1,354896}{0,006131} = 221 \text{ проба.}$$

Необходимая численность выборки для определения генеральной доли при случайном повторном отборе определяется по формуле

$$n = \frac{t^2 \times w \times (1 - w)}{\Delta_w^2}. \quad (198)$$

Необходимая численность выборки составила

$$n = \frac{2^2 \times 0,656 \times (1 - 0,656)}{0,064^2} = \frac{0,902656}{0,004096} \approx 221 \text{ проба.}$$

Расчет показал, что для того, чтобы генеральная средняя не превысила полученные пределы с вероятностью 0,997, необходимая численность выборки должна составлять 221 пробу. Для того, чтобы генеральная доля не превысила полученные пределы с вероятностью 0,954, необходимая численность выборки должна составлять 221 пробу. Следовательно, проведенная выборка репрезентативна и ее результаты возможно с достаточной точностью перенести на всю генеральную совокупность.

Пример расчета характеристик генеральной совокупности на основе характеристик выборочной совокупности при бесповторном способе отбора.

В процессе анализа возрастного состава студентов вуза в порядке случайной бесповторной выборки обследован возраст 100 человек студентов из общего числа 2000 человек. Результаты обработки данных приведены в таблице 52.

Таблица 52

Исходные данные по выборочной совокупности

Возраст, лет.	17	18	19	20	21	22	23	Итого
Число студентов, чел.	11	13	18	23	17	10	8	100

Средний возраст студентов по данным выборочной совокупности (выборочная средняя) определяется по формуле средней арифметической взвешенной. Средний возраст студентов в выборочной совокупности составил

$$\bar{X} = \frac{\sum X \times f}{\sum f} = \frac{17 \times 11 + 18 \times 13 + 19 \times 18 + 20 \times 23 + 21 \times 17 + 22 \times 10 + 23 \times 8}{11 + 13 + 18 + 23 + 17 + 10 + 8} = \frac{1984}{100} = 19,84 \text{ года.}$$

где X – возраст;

f – число студентов.

Среднее квадратическое отклонение определяется по формуле:

$$\sigma = \sqrt{\frac{\sum X^2 \times f}{\sum f} - (\bar{X}^2)} \quad (199)$$

Для определения выборочного среднего квадратического отклонения строится вспомогательная таблица 53.

Таблица 53

Расчет выборочного среднего квадратического отклонения

Возраст, лет, X	Число студентов, чел., f	X ²	X ² *f
17	11	289	3179
18	13	324	4212
19	18	361	6498
20	23	400	9200
21	17	441	7497
22	10	484	4840
23	8	529	4232
Всего	100	-	39658

Среднее квадратическое отклонение составило

$$\sigma = \sqrt{\frac{39658}{100} - (19,84)^2} = 1,71 \text{ лет.}$$

Пределы, в которых заключается средний возраст студентов в генеральной совокупности (генеральная средняя) определяются по формуле

$$\bar{X} - \Delta \leq \bar{X}_z \leq \bar{X} + \Delta \quad (200)$$

где Δ - предельная ошибка выборки.

Предельная ошибка выборки определяется по формуле

$$\Delta = \mu \times t \quad (201)$$

где t – коэффициент кратности ошибки (определяется по таблице 46, при вероятности 0,954 $t = 2$).

Средняя ошибка выборки для средней при случайном бесповторном отборе определяется по формуле

$$\mu = \sqrt{\frac{\sigma^2}{n} \times \left(1 - \frac{n}{N}\right)} \quad (202)$$

где n – численность выборочной совокупности;

N – численность генеральной совокупности.

Средняя ошибка выборки составила

$$\mu = \sqrt{\frac{1,71^2}{100} \times \left(1 - \frac{100}{2000}\right)} = 0,167 \text{ лет.}$$

Предельная ошибка выборки составила

$$\Delta = 0,167 \times 2 = 0,334 \text{ лет.}$$

Пределы генеральной средней составили

$$X_r = 19,84 - 0,334 = 19,506 \text{ лет.}$$

$$X_r = 19,84 + 0,334 = 20,174 \text{ лет.}$$

Расчет показал, что с вероятностью 0,954 можно утверждать, что в генеральной совокупности средний возраст студентов составит от 19,506 лет до 20,174 лет.

Доля студентов не старше 20 лет по данным выборочной совокупности (выборочная доля) составила

$$w = 11 + 13 + 18 + 23 / 100 = 0,65$$

Пределы генеральной доли определяются по формуле

$$w - \Delta \leq p_z \leq w + \Delta \quad (203)$$

где Δ - предельная ошибка выборки.

Предельная ошибка выборки определяется по формуле

$$\Delta = \mu \times t \quad (204)$$

где t – коэффициент кратности ошибки (определяется по таблице 46, при вероятности 0,683 $t = 3$).

Средняя ошибка выборки для доли при случайном бесповторном отборе определяется по формуле

$$\mu = \sqrt{\frac{w \times (1 - w)}{n} \times (1 - \frac{n}{N})} \quad (205)$$

где n – численность выборочной совокупности

Средняя ошибка для доли составила

$$\mu = \sqrt{\frac{0,65 \times (1 - 0,65)}{100} \times (1 - \frac{100}{2000})} = 0,0465$$

Предельная ошибка для доли составила

$$\Delta = 0,0465 \times 3 = 0,1395$$

Пределы генеральной доли составили

$$p = 0,65 - 0,1395 = 0,5105$$

$$p = 0,65 + 0,1395 = 0,7895$$

Расчет показал, что с вероятностью 0,683 можно утверждать, что в генеральной совокупности доля студентов не старше 20 лет составит от 0,5105 до 0,7895 (от 51,05% до 78,95%).

Необходимая численность выборки для определения средней при случайном бесповторном отборе определяется по формуле

$$n = \frac{t^2 \times \sigma_{\bar{x}}^2 \times N}{\Delta_{\bar{x}}^2 \times N + t^2 \times \sigma_{\bar{x}}^2} \quad (206)$$

Необходимая численность выборки составила

$$n = \frac{2^2 \times 1,71^2 \times 2000}{0,334^2 \times 2000 + 2^2 \times 1,71^2} = \frac{23392,8}{223,112 + 11,6964} = \frac{23392,8}{234,8084} \approx 100 \text{ человек.}$$

Необходимая численность выборки для доли при случайном бесповторном отборе определяется по формуле

$$n = \frac{t^2 \times w \times (1-w) \times N}{\Delta^2 \times N + t^2 \times w \times (1-w)} \quad (207)$$

Необходимая численность выборки для определения доли составила

$$n = \frac{3^2 \times 0,65 \times (1-0,65) \times 2000}{0,1395^2 \times 2000 + 3^2 \times 0,65 \times (1-0,65)} = \frac{4095}{40,9675} \approx 100$$

Расчет показал, что для того, чтобы генеральная средняя не превысила полученные пределы с вероятностью 0,954, необходимая численность выборки должна составлять 100 человек. Для того, чтобы генеральная доля не превысила полученные пределы с вероятностью 0,683, необходимая численность выборки должна составлять 100 человек. Следовательно, проведенная выборка репрезентативна и ее результаты возможно с достаточной точностью перенести на всю генеральную совокупность.

10 СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗЕЙ И ЗАВИСИМОСТЕЙ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ АНАЛИЗА ВЗАИМОСВЯЗЕЙ И ЗАВИСИМОСТЕЙ

10.1 Понятие взаимосвязей и зависимостей в статистике. Задачи статистического изучения взаимосвязей и зависимостей

В соответствии с законом диалектики, все явления взаимосвязаны между собой, обуславливают друг друга и находятся в постоянном движении и развитии. Поэтому объективно изучить определенное явление возможно только на основе исследования его во взаимосвязи с окружающими явлениями. Особенность взаимосвязей заключается в том, что их закономерный характер проявляется только в массе явлений – в их совокупности. Следовательно, одной из основных задач статистики является оценка степени объективно существующих связей между изучаемыми явлениями.

Задачи статистического изучения связей:

- 1) раскрытие причинно-следственных отношений между изучаемыми явлениями – связи явлений и процессов, при которой изменение одного явления (причины - совокупности условий и обстоятельств, действие которых приводит к появлению следствия) ведет к изменению другого явления (следствия);
- 2) выявление факторов, оказывающих основное влияние на вариацию изучаемых явлений и процессов;
- 3) выявление временной последовательности причинно-следственных связей;
- 4) выявление главных причин и абстрагирование от второстепенных.

Статистическое изучение взаимосвязей и зависимостей проводится в три основных этапа: 1) качественный анализ изучаемого явления (на основе методов экономической теории, социологии и др.); 2) построение модели связи

(на основе методов статистических группировок, средних величин, статистических таблиц и др.); 3) интерпретация результатов анализа (на основе анализа качественных особенностей изучаемого явления).

При статистическом изучении взаимосвязей, признаки делятся на два класса:

1. Факторные (факторы) это - признаки, которые обуславливают изменение других, связанных с ними признаков.
2. Результативные (результат) это – признаки, которые изменяются под действием факторных признаков.

Взаимосвязи между явлениями проявляются в трех основных формах (рис. 56).

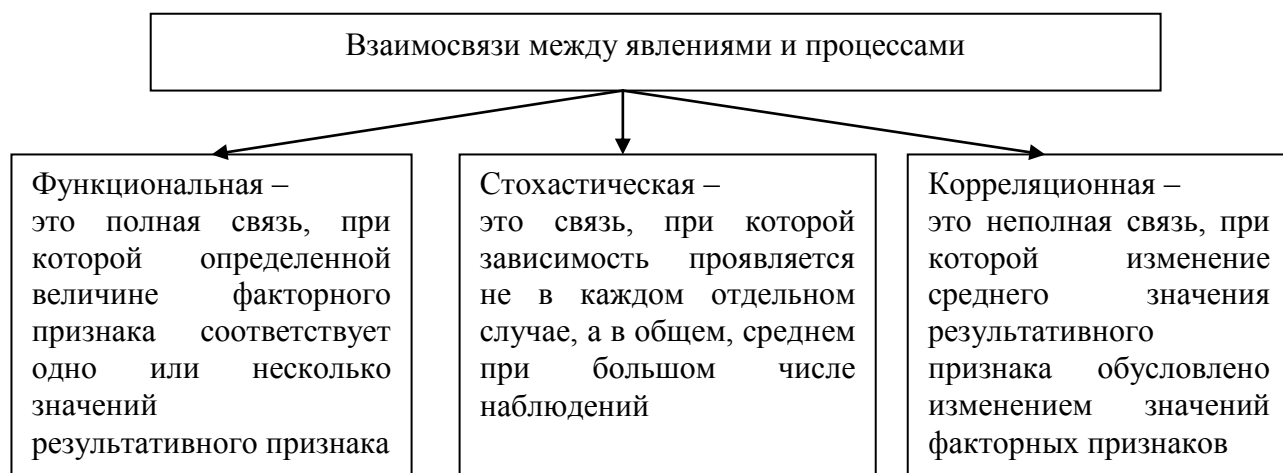


Рисунок 56 – Формы взаимосвязи между социально-экономическими явлениями и процессами

Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы совокупности. При функциональной связи величина результативного признака полностью определяется факторными признаками (например, фонд оплаты труда равен произведению численности работников на среднюю заработную плату одного работника и т.д.). Поэтому функциональная связь выражается в виде мультипликативной модели

$$Y = X \times Z, \quad (208)$$

где Y – результативный признак;

X – первый факторный признак;

Z – второй факторный признак.

Функциональная связь проявляется с одинаковой силой у всех единиц совокупности, независимо от изменения других признаков (например, независимо от формы собственности, вида деятельности и размеров предприятия, фонд заработной платы определяется произведением численности работников на среднюю заработную плату одного работника). Поэтому функциональную зависимость можно установить на основе единичного случая (отдельного элемента совокупности).

При стохастической связи зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений. Частным случаем стохастической связи является корреляционная связь.

При корреляционной связи изменение результативного признака необязательно зависит от изменения одного или нескольких факторных признаков. Корреляционная зависимость проявляется в среднем для массовых наблюдений. При этом определенному значению факторного признака соответствует некоторый ряд вероятных значений результативного признака, так как на результат влияют множество неучтенных случайных факторов (например, на среднюю заработную плату работников предприятия могут влиять множество факторов – экономическое положение в стране, регионе, отрасль экономики, к которой относится предприятие, и др.).

Корреляционная зависимость проявляется в среднем в массе случаев. Поэтому каждому значению факторного признака соответствует множество значений результативного признака, расположенное в некотором интервале. Корреляционные связи являются необратимыми и выражаются в виде аддитивной модели

$$Y = a + b \times X, \quad (209)$$

где a – свободный член;

b – коэффициент;

X – факторный признак.

Следовательно, корреляционная зависимость – это статистическая зависимость между случайными величинами, не имеющими строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой величины.

10.2 Методы статистического изучения взаимосвязей и зависимостей

Статистика, при изучении взаимосвязей и зависимостей решает следующие задачи: 1) дает количественную оценку наличия связей; 2) определяет направление связей; 3) дает характеристику силы связи; 4) дает характеристику формы влияния факторных признаков на результат.

Для решения указанных задач в статистике применяются следующие методы.

1. Для изучения функциональных зависимостей применяются балансовый метод, индексный метод, метод сопоставления параллельных рядов.
2. Для изучения корреляционных связей между качественными признаками применяются метод взаимной сопряженности, метод ранжирования и др.
3. Для изучения корреляционных связей между количественными признаками применяются метод параллельных рядов, графический метод, метод аналитических группировок, корреляционно-регрессионный анализ.

Наиболее аналитичным является метод корреляционно-регрессионного анализа, который объединяет в себе два приема, выполняющих разные, но взаимодополняющие задачи: 1) корреляционный анализ; 2) регрессионный анализ.

Задачи корреляционного анализа: 1) количественное измерение тесноты (силы) связи между результативным и факторными признаками; 2) определение неизвестных причинных связей; 3) оценка факторов, оказывающих наибольшее влияние на результативный признак.

Основной предпосылкой применения корреляционного анализа является необходимость подчинения совокупности значений всех факторных признаков и результативного признака нормальному закону распределения или близость к нему.

Задачи регрессионного анализа: 1) определение аналитического выражения связи, в котором результативный признак обусловлен влиянием факторных признаков, а множество прочих факторов, также оказывающих влияние на результат, принимается за постоянные и средние значения; 2) установление направления и формы зависимости; 3) определение функции регрессии (регрессионной модели); 4) использование уравнения регрессии (регрессионной модели) для оценки неизвестных значений зависимой переменной.

Основной предпосылкой регрессионного анализа является подчинение результативного признака нормальному закону распределения, а факторные признаки могут иметь произвольный закон распределения. При этом заранее подразумевается наличие причинно-следственных связей между результативным и факторными признаками.

Корреляционно-регрессионный анализ в целом включает измерение тесноты (силы) и направления связи между результатом и факторами (корреляционный анализ) и установление аналитического выражения связи (регрессионный анализ). Поэтому эти два приема объединены в один статистический метод.

Основное условие достоверности анализа заключается в том, что статистическая модель связи изучаемых явлений может быть достаточно адекватной моделируемому явлению (аппроксимирующая).

Достаточность аппроксимации (приближения) статистической модели к эмпирическим данным достигается путем соблюдения следующих основных требований к ее построению:

1) совокупность исходных данных должна быть однородной и математически описываться непрерывными функциями;

- 2) возможность описания моделируемого явления одним или несколькими уравнениями причинно-следственных связей;
- 3) все факторные признаки должны иметь количественное выражение;
- 4) исследуемая совокупность должна иметь достаточно большой объем (число наблюдений должно быть в 5-6 раз больше числа факторов, включенных в модель);
- 5) причинно-следственные связи должны описываться линейной или приводимой к линейной формой зависимости;
- 6) отсутствие количественных ограничений на параметры модели;
- 7) территориальная и временная структура изучаемой совокупности должна быть постоянной.

Теоретическая обоснованность моделей связи, построенных на основе корреляционно-регрессионного анализа, обеспечивается соблюдением следующих условий:

- 1) все признаки и их совместные распределения должны подчиняться нормальному закону распределения;
- 2) дисперсия моделируемого признака должна все время оставаться постоянной при изменении величины и значений факторных признаков;
- 3) отдельные наблюдения должны быть независимыми, то есть результаты, полученные при наблюдении, не должны быть связаны с предыдущими наблюдениями и содержать информацию о последующих наблюдениях, а также влиять на них;
- 4) число факторов, включаемых в модель, должно быть оптимальным, то есть число факторов, включаемых в модель, должно соответствовать объему изучаемой совокупности (число факторных признаков должно быть в 5-6 раз меньше численности единиц изучаемой совокупности).

10.3 Формы и виды корреляционно-регрессионной зависимости

Корреляционная связь проявляется в нескольких формах и видах (рис. 57).



Рисунок 57 – Виды и формы корреляционной зависимости

По направлению корреляционные связи принимают две формы: прямая зависимость; обратная зависимость.

Прямая (положительная корреляция) – это зависимость, при которой изменение результативного и факторных признаков происходит в одном направлении. То есть, при увеличении (уменьшении) факторного признака результативный признак также увеличивается (также уменьшается).

Обратная зависимость (отрицательная корреляция) – это зависимость, при которой изменение результативного и факторных признаков происходит в противоположном направлении. То есть, при увеличении (уменьшении) факторного признака результативный признак уменьшается (увеличивается).

По количеству признаков, включенных в модель, корреляционные связи принимают три формы: парная корреляция; частная корреляция; множественная корреляция.

Парная корреляция – это зависимость между двумя признаками. При этом, один – факторный, второй – результативный; либо один – факторный и второй факторный.

Частная корреляция – это зависимость между одним результативным и несколькими факторными признаками. При этом, один результативный, второй – один из факторных. Значение остальных факторных признаков в данном случае фиксируется на среднем уровне.

Множественная корреляция – это зависимость между одним результативным и двумя или несколькими факторными признаками. В отличие от парной и частной корреляции, которые могут быть линейными и криволинейными, множественная корреляция в конечном итоге сводится только к линейной форме связи.

По степени зависимости корреляционные связи принимают три формы: непосредственная связь; косвенная связь; ложная связь.

Непосредственная связь – это зависимость, при которой факторный и результативный признаки взаимодействуют между собой непосредственно напрямую без участия каких-либо других факторов.

Косвенная связь – это зависимость, при которой факторный и результативный признак взаимодействуют между собой не напрямую, а через участие какого-либо третьего фактора, опосредующего связь между изучаемыми признаками.

Ложная связь – это зависимость, при которой связь устанавливается только формально (математически), подтверждается только количественными оценками и не имеет качественной основы.

По форме связи принимают два вида: линейная зависимость; нелинейная зависимость.

Линейная зависимость (линейная регрессия) выражается прямолинейной функцией (уравнением прямой линии), которое имеет вид

$$Y(X) = a_0 + a_1 \times X, \quad (210)$$

где $Y(X)$ – результативный признак;

X – факторный признак;

a_0, a_1 – параметры уравнения регрессии.

Линейная регрессия используется в том случае, когда при изменении факторного признака происходит непрерывное равномерное изменение результативного признака.

Нелинейная зависимость (нелинейная регрессия) - выражается уравнением любой криволинейной функции (параболы, гиперболы, логарифмической функции и т.д.)

$$Y(X) = a_0 + a_1 \times X + a_2 \times X^2, \quad (211)$$

Нелинейная регрессия используется в том случае, когда при изменении факторного признака изменение результативного признака происходит неравномерно, либо направление его изменения вообще меняется на обратное.

10.4 Определение параметров уравнения регрессии, их интерпретация

В настоящее время традиционные методы корреляционно-регрессионного анализа широко представлены в разного рода пакетах прикладных статистических программ ЭВМ, что избавляет от решения задачи сложных математических расчетов. В этих условиях задачи исследования включают:

- 1) правильную подготовку исходной информации;
- 2) отбор изучаемых факторов на основе определения причинно-следственных связей;
- 3) обеспечение достаточного объема совокупности для получения несмещенных оценок;
- 4) правильную интерпретацию результатов анализа.

Правильная подготовка исходной информации и отбор изучаемых факторов производится на основе предварительного анализа, который предполагает определение наличия связи между признаками; определение того, какой признак является результатом и какой – фактором; установление вида и

формы связи между признаками. Предварительный анализ проводится при помощи следующих методов: аналитических группировок; графического метода (построение графика корреляционной зависимости – корреляционного поля); табличного метода (построение корреляционных таблиц); метода перебора различных уравнений (проверка каждого выбранного уравнения связи на основе t-критерия Стьюдента и F-критерия Фишера); метода экспертных оценок (расчет и анализ непараметрических показателей связи – коэффициентов Спирмена, Кенделла и конкордации); метода шаговой регрессии – шагового регрессионного анализа (последовательное включение факторов в уравнение регрессии и последующая проверка их значимости на основе критериев t-критерия Стьюдента и F-критерия Фишера).

Определение параметров уравнения регрессии (регрессионной модели) осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений изучаемой совокупности.

Сущность метода наименьших квадратов заключается в нахождении параметров регрессионной модели, при которых минимизируется сумма квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии (регрессионной модели). Формально метод наименьших квадратов можно представить следующим образом

$$\Sigma(Y(X) - Y_i)^2 \rightarrow \min, \quad (212)$$

где $Y(X)$ – теоретические значения результативного признака;

Y_i – фактические (эмпирические) значения результативного признака.

Для прямой зависимости

$$S = \Sigma(y - a_0 - a_1x)^2 \rightarrow \min. \quad (213)$$

Сумма квадратов отклонений S является функцией параметров a_0 и a_1 . Минимальность значения этой суммы осуществляется путем решения системы нормальных уравнений. Метод наименьших квадратов предполагает число нормальных уравнений в системе равно числу неизвестных коэффициентов уравнения регрессии.

При парной (двухфакторной) модели и линейной зависимости уравнение регрессии имеет вид

$$Y(X) = a_0 + a_1 X_1, \quad (214)$$

При парной регрессионной модели система нормальных уравнений включает два уравнения – по числу неизвестных параметров в регрессионной модели и имеет вид

$$\begin{cases} \sum Y = a_0 n + a_1 \sum x \\ \sum Yx = a_0 \sum x + a_1 \sum x^2 \end{cases}, \quad (215)$$

где a_0 - свободный член, который показывает усредненное влияние на результативный признак неучтенных (не выделенных для исследования) факторов;

a_1 - коэффициент чистой регрессии, который показывает, на сколько единиц своего измерения увеличится (или уменьшится) в среднем результативный признак при увеличении факторного признака на единицу своего измерения. Если a_1 положителен - наблюдается прямая связь между признаками, если a_1 отрицателен – наблюдается обратная связь. При этом, числовое значение коэффициента регрессии характеризует величину изменения результативного признака под влиянием факторного признака.

При множественной (трехфакторной) модели уравнение регрессии имеет вид

$$Y(X) = a_0 + a_1 X_1 + a_2 X_2, \quad (216)$$

Для определения параметров в данном случае решается система из трех уравнений, по числу неизвестных параметров в регрессионной модели и имеет вид

$$\begin{cases} \sum Y = a_0 n + a_1 \sum X_1 + a_2 \sum X_2 \\ \sum YX_1 = a_0 \sum X_1 + a_1 \sum X_1^2 + a_2 \sum X_1 X_2, \\ \sum YX_2 = a_0 \sum X_2 + a_1 \sum X_1 X_2 + a_2 \sum X_2^2. \end{cases} \quad (217)$$

Для нахождения параметров уравнения регрессии (регрессионной модели) решается вышеуказанная система уравнений, либо используются следующие формулы.

Коэффициент регрессии определяется по формуле

$$a_i = \frac{\Sigma(X_i - \bar{X}) \times (Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \quad (218)$$

или

$$a_i = r_{xy(i)} \times \frac{\sigma_y}{\sigma_{x(i)}}, \quad (219)$$

где X_i – индивидуальные значения факторного признака;

\bar{X} – среднее значение факторного признака;

Y_i – индивидуальные значения результативного признака;

\bar{Y} – среднее значение результативного признака;

$r_{xy(i)}$ – коэффициент парной линейной корреляции между результативным и факторным признаком;

σ_y – среднее квадратическое отклонение результативного признака;

$\sigma_{x(i)}$ – среднее квадратическое отклонение факторного признака.

Параметр a_0 определяется по формуле

$$a_0 = \bar{Y} - a_i \times \bar{X}. \quad (220)$$

Полученное уравнение регрессии выражает функциональную зависимость между результативным и факторным признаком (факторными признаками), которую можно допустить при условии, что все остаточные факторы не оказывают существенного влияния на результат.

Такое условие, как правило, не соблюдается, что снижает качество уравнения регрессии, и результаты анализа обладают ошибками упрощения. Причины возникновения ошибок упрощения: 1) неполнота охвата единиц изучаемой совокупности; 2) включение в модель не всех, а только части факторов, оказывающих влияние на результат; 3) неполная адекватность выбранного уравнения связи (регрессионной модели); 4) наличие мультиколлинеарности (тесной зависимости между самими факторными

признаками - парные коэффициенты корреляции превышают 0,8) в уравнениях множественной регрессии.

Интерпретация результатов анализа может быть затруднена, так как факторные признаки различны по своей сущности и имеют разные единицы измерения. Для приведения коэффициентов регрессии в сопоставимый вид и к единым единицам измерения коэффициенты регрессии приводятся в стандартизованный вид путем их выражения в относительных величинах. Для этого определяются частные коэффициенты: коэффициент эластичности - \mathcal{E} ; бета-коэффициент - β ; частный коэффициент детерминации - d .

Коэффициент эластичности определяется по формуле

$$\mathcal{E}_i = a_i \times \frac{\bar{X}_i}{\bar{Y}}, \quad (221)$$

где \mathcal{E}_i – коэффициент эластичности;

a_i – коэффициент регрессии;

\bar{X}_i - среднее значение факторного признака;

\bar{Y} - среднее значение результативного признака.

Коэффициент эластичности показывает, на сколько процентов (%) увеличится либо уменьшится в среднем значение результативного признака при увеличении значения факторного признака на 1 процент (%).

Бетта-коэффициенты используются для определения факторов, в развитии которых заложены наиболее крупные резервы изменения результативного признака. Характеризуют различия в степени вариации включенных в модель признаков. Определяются по формуле

$$\beta_i = a_i \times \frac{\sigma_{x_i}}{\sigma_y}, \quad (222)$$

где β_i - коэффициент;

a_i – коэффициент регрессии;

σ_{x_i} - среднее квадратическое отклонение факторного признака;

σ_y - среднее квадратическое отклонение результативного признака.

Бетта-коэффициент показывает, на какую часть своего среднего квадратического отклонения увеличится либо уменьшится значение результативного признака при увеличении значения факторного признака на одну величину своего среднего квадратического отклонения.

Проведение корреляционно-регрессионного анализа в современных условиях обеспечено прикладными программами ЭВМ. Одной из программ является «Пакет анализа» стандартного приложения Microsoft Excel.

Пример расчета коэффициентов регрессии и корреляции при парной зависимости – однофакторной модели представлен на основании данных таблицы 54.

Таблица 54

Качество почв и урожайность овощей по предприятиям области

№	Качество почв, баллы, X	Урожайность овощей, т/га, Y
1	68,00	16,00
2	80,00	18,00
3	50,00	15,00
4	45,00	12,00
5	87,00	23,00
6	85,00	20,00
7	45,00	9,00
8	48,00	13,00
9	95,00	31,50
10	50,00	14,00
11	70,00	18,00
12	80,00	21,50
13	89,00	23,00
14	78,00	22,00
15	55,00	16,00
16	75,00	18,00
17	88,00	29,00
18	51,00	13,00
19	58,00	14,00
20	90,00	28,00

При помощи построения регрессионной модели определяется степень зависимости урожайности овощей (результативный признак – Y) от качества почв (факторный признак – X). Форма связи определяется на основании предварительного анализа путем построения графика корреляционной зависимости – корреляционного поля. Корреляционное поле можно построить с

использованием приложения «Мастер диаграмм» Microsoft Word или Microsoft Excel следующим путем: вкладка «вставка», «диаграмма», ввести в ячейки исходные данные – в первую графу значения факторного признака X (качество почв, баллы), во вторую графу значения результативного признака Y (урожайность овощей, т/га); «тип диаграммы» - выбрать «точечная», «добавить линию тренда» - «линейный», во вкладке «параметры» выбрать «показывать уравнение на диаграмме» и «поместить на диаграмму величину достоверности аппроксимации» (рис. 58).

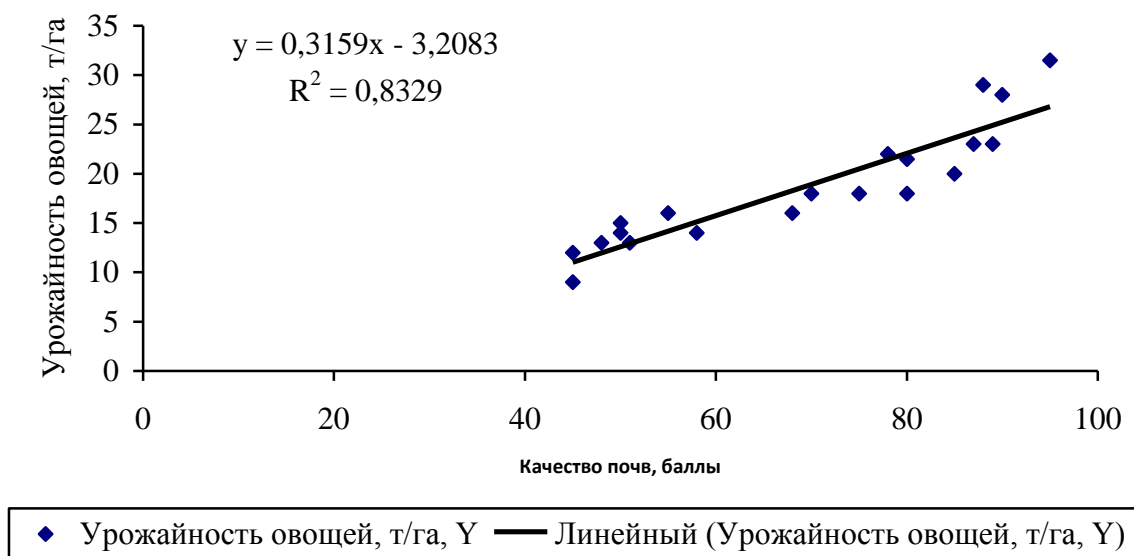


Рисунок 58 - Зависимость урожайности овощей от качества почв

Из графика корреляционной зависимости – корреляционного поля (рис. 58) видно, что между урожайностью овощей и качеством почв наблюдается линейная зависимость, так как точки в основном группируются близко к линейному тренду. При этом видно, что связь между урожайностью овощей и качеством почв имеет прямое направление, так как линия тренда направлена вверх. Также можно сказать, что между урожайностью овощей и качеством почв наблюдается тесная связь, так как точки, отражающие эмпирические значения результативного признака достаточно близко сгруппированы относительно линии тренда.

Кроме этого, на графике уравнение регрессии имеет вид:
 $Y(X) = -3,2083 + 0,3159 \times X$.

Параметр $a_1 = 0,3159$ в полученном уравнении регрессии – это коэффициент чистой регрессии, который показывает, что при увеличении качества почв на 1 балл, урожайность овощей в среднем увеличивается на 0,3159 т/га.

Проведенный анализ графическим способом показал предварительные результаты. Для полного анализа можно использовать приложение «Пакет анализа» Microsoft Excel. Алгоритм работы в данном приложении включает следующие шаги: внести исходные данные в ячейки листа Microsoft Excel; во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «регрессия», выделить диапазон входного интервала исходных данных результативного признака Y и диапазон входного интервала факторного признака X , ввести данные; отметить «уровень надежности 0,95», Для вывода результатов - «ОК». (табл. 55).

Таблица 55

Результаты регрессионного анализа

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
Y-пересечение	-3,20828	2,381644	-1,34709	0,194667	-8,21193	1,795365	-8,21193	1,795365
Переменная X 1	0,315909	0,033356	9,470809	2,05E-08	0,24583	0,385987	0,24583	0,385987

В таблице 55 сгенерированы значения коэффициента регрессии a_1 и его статистические оценки. В таблице 55 столбцы имеют следующую интерпретацию:

- коэффициенты – значения коэффициента a_i ;
- стандартная ошибка – стандартная ошибка коэффициента a_1 ;
- t – статистика – расчетные значения t – критерия (Стьюдента), вычисленные по формуле

$$t - \text{статистика} = \text{Коэффициенты} / \text{Стандартная ошибка}; \quad (223)$$

- Р - значения – значения уровня значимости, соответствующие вычислительным значениям t_i . Определяется с помощью функции: СТЬЮДРАСП (t_p ; $n - m - 1$);

- нижние 95% и верхние 95% - соответственно нижние и верхние границы доверительных интервалов для коэффициента регрессии a_1 . Для нахождения границ доверительных интервалов с помощью функции СТЬДРАСПОБР (α ; $n - m - 1$) рассчитываются критические значения t – критерия (Стьюдента), а затем вычисляются соответственно нижние и верхние границы доверительных интервалов по формулам

$$\text{нижние 95\%} = \text{Коэффициент} - \text{Стандартная ошибка} \times t_{кр}; \quad (224)$$

$$\text{верхние 95\%} = \text{Коэффициент} + \text{Стандартная ошибка} \times t_{кр}. \quad (225)$$

Из расчетов, представленных в таблице 55, уравнение регрессии имеет вид: $Y(X) = -3,20828 + 0,315909 \times X$.

Параметр $a_1 = 0,315909$ в полученном уравнении регрессии – это коэффициент чистой регрессии, который показывает, что при увеличении качества почв на 1 балл, урожайность овощей в среднем увеличивается на 0,315909 т/га.

Теоретические уровни урожайности овощей $Y(X)$ рассчитываются путем подстановки в полученное уравнение (регрессионную модель) каждого значения независимой переменной X . При использовании приложения «Пакет анализа» Microsoft Excel к выше указанным шагам необходимо добавить шаг «остатки». Для вывода результатов - «ОК». (табл. 56).

В таблице 56 теоретические значения результативного признака – значения урожайности овощей, возможные при каждом имеющемся значении качества почв, рассчитанные по полученной регрессионной модели, представлены в графе «Урожайность овощей (теоретическая), т/га., $Y(X)$ Предсказанное Y ». Значения в графе «остатки» определяется как разница между каждым фактическим Y и каждым теоретическим $Y(X)$ значением

результативного признака и показывают влияние случайных факторов в каждом наблюдении.

Таблица 56

Вывод остатка

Наблю дение	Урожайность овощей (фактическая), т/га, Y	Урожайность овощей (теоретическая), т/га., $Y(X)$. Предсказанное \hat{Y}	Остатки
1	16,00	18,27352	-2,27352
2	18,00	22,06443	-4,06443
3	15,00	12,58716	2,412837
4	12,00	11,00762	0,992382
5	23,00	24,27579	-1,27579
6	20,00	23,64397	-3,64397
7	9,00	11,00762	-2,00762
8	13,00	11,95534	1,044655
9	31,50	26,80306	4,696936
10	14,00	12,58716	1,412837
11	18,00	18,90534	-0,90534
12	21,50	22,06443	-0,56443
13	23,00	24,90761	-1,90761
14	22,00	21,43261	0,567388
15	16,00	14,16671	1,833293
16	18,00	20,48489	-2,48489
17	29,00	24,5917	4,408299
18	13,00	12,90307	0,096929
19	14,00	15,11443	-1,11443
20	28,00	25,22352	2,776481
Итого	374,00	373,99998	0,000007

Согласно методу наименьших квадратов, суммы теоретических и эмпирических значений результативного признака равны, а сумма значений остатков равна или стремится к нулю.

Для приведения коэффициентов чистой регрессии в сопоставимый вид, определяются стандартизованные коэффициенты регрессии по формулам 221, 222. Для этого следует сделать вспомогательные расчеты. При использовании приложения «Пакет анализа» Microsoft Excel, алгоритм расчета следующий: внести исходные данные в ячейки листа Microsoft Excel, во вкладке «данные» выбрать «анализ данных»; в меню выбрать инструмент «однофакторный дисперсионный анализ», выделить во «входной интервал» ячейки с исходными

данными, для вывода результатов нажать «ОК». Для вывода результатов - «ОК». (табл. 57).

Столбцы таблицы 57 имеют следующую интерпретацию:

- столбец «счет» - показывает число наблюдений;
- столбец «сумма» - показывает общие суммы по графе X и по графе Y;
- столбец «среднее» - показывает значение статистических средних по признаку X и по признаку Y;

- столбец «дисперсия» - показывает значение дисперсий по признаку X и по признаку Y;

- столбец df – показывает число степеней свободы. Для строки «между группами» число степеней свободы определяется количеством факторных признаков m в уравнении регрессии; для строки «внутри групп» число степеней свободы определяется разностью между общим числом значений признака по X и по Y – $2n$ и количеством переменных в уравнении регрессии m ; для строки «итого» число степеней свободы определяется суммой степеней свободы для строк «между группами» и «внутри групп».

- столбец SS – показывает сумму квадратов отклонений. Для строки «между группами» - это сумма квадратов отклонений теоретических данных от среднего; для строки «внутри групп» - это сумма квадратов отклонений эмпирических данных от теоретических; для строки «итого» - это сумма квадратов отклонений эмпирических данных от среднего.

- столбец MS - показывает дисперсии, рассчитываемые по формуле $MS = SS/df$. Для строки «между группами» - это факторная дисперсия; для строки «внутри групп» - это остаточная дисперсия.

- столбец F – показывает расчетное значение F - критерия Фишера, вычисляемое по формуле

$$F_p = MS (\text{между группами}) / MS (\text{внутри групп}). \quad (226)$$

- столбец $\text{Значимость } F$ – значение уровня значимости, соответствующее вычислительному значению F_p . Определяется с помощью функции: $FPACП(F_p; df (\text{между группами}); df (\text{внутри групп}))$.

Результаты расчетов однофакторного дисперсионного анализа

Однофакторный дисперсионный анализ						
ИТОГИ						
<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
Столбец 1, X	20	1387	69,35	303,8184		
Столбец 2, Y	20	374	18,7	36,40526		
Дисперсионный анализ						
<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
Между группами	25654,23	1	25654,23	150,808	8,45E-15	4,098172
Внутри групп	6464,25	38	170,1118			
Итого	32118,48	39				

На основании расчетных данных таблицы 57 стандартизованные коэффициенты регрессии составят:

- коэффициент эластичности $\Theta_i = a_i \times \frac{\bar{X}_i}{\bar{Y}} = 0,315909 \times \frac{69,35}{18,7} = 1,1717$;

- бетта - коэффициент $\beta_i = a_i \times \frac{\sigma_{x_i}}{\sigma_y} = 0,315909 \times \frac{\sqrt{303,8184}}{\sqrt{36,40526}} = 0,315909 \times \frac{17,43}{6,03} = 0,9131$.

Расчет коэффициента эластичности показал, что при увеличении качества почв на 1%, урожайность овощей возрастает в среднем на 1,1717%, при этом, так как значение коэффициента эластичности больше единицы, то зависимость урожайности овощей от качества почв можно считать эластичной, то есть – результативный признак изменяется большими темпами, чем факторный признак. Расчет бетта-коэффициента показал, что при увеличении качества почв на одну величину своего среднего квадратического отклонения, урожайность овощей увеличивается в среднем на 0,89 величины своих средних квадратических отклонений.

Пример расчета коэффициентов регрессии и корреляции при множественной зависимости – двухфакторной модели представлен на основании данных таблицы 58.

Таблица 58

Качество почв, доза внесения удобрений и урожайность овощей по предприятиям области

№	Урожайность овощей, т/га, Y	Качество почв, баллы, X_1	Доза внесения удобрений, т/га, X_2
1	16,00	68,00	0,25
2	18,00	80,00	0,35
3	15,00	50,00	0,20
4	12,00	45,00	0,10
5	23,00	87,00	0,70
6	20,00	85,00	0,50
7	9,00	45,00	0,10
8	13,00	48,00	0,15
9	31,50	95,00	0,90
10	14,00	50,00	0,25
11	18,00	70,00	0,32
12	21,50	80,00	0,50
13	23,00	89,00	0,66
14	22,00	78,00	0,69
15	16,00	55,00	0,25
16	18,00	75,00	0,41
17	29,00	88,00	0,80
18	13,00	51,00	0,10
19	14,00	58,00	0,20
20	28,00	90,00	0,80

При помощи построения регрессионной модели определяется степень зависимости урожайности овощей (результативный признак – Y) от качества почв (первый факторный признак – X_1) и от дозы внесения удобрений (второй факторный признак X_2). Для анализа можно использовать приложение «Пакет анализа» Microsoft Excel. Алгоритм работы в данном приложении включает следующие шаги: внести исходные данные в ячейки листа Microsoft Excel; во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «регрессия», выделить отдельно диапазон входного интервала результативного признака Y и вместе диапазон входного интервала факторных

признаков X_1 и X_2 , отметить «уровень надежности 0,95», Для вывода результатов - «ОК». (табл. 59).

Таблица 59

Результаты регрессионного анализа

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	7,846596	2,300323	3,411084	0,003327	2,993338	12,69985	2,993338	12,69985
Переменная X 1	0,038209	0,050049	0,763423	0,455676	-0,06739	0,143803	-0,06739	0,143803
Переменная X 2	19,93594	3,312169	6,018999	1,38E-05	12,94788	26,92401	12,94788	26,92401

В таблице 59 сгенерированы значения коэффициентов регрессии a_1 и a_2 и их статистические оценки. В таблице 59 столбцы имеют следующую интерпретацию:

- коэффициенты – значения коэффициентов регрессии a_i ;
- стандартная ошибка – стандартная ошибка коэффициентов регрессии a_i ;
- t – статистика – расчетные значения t – критерия (Стъюдента), вычисленные по формуле

$$t - \text{статистика} = \text{Коэффициенты} / \text{Стандартная ошибка}; \quad (227)$$

- P – значения – значения уровня значимости, соответствующие вычислительным значениям t_i . Определяется с помощью функции: СТЬЮДРАСП (t_p ; $n - m - 1$);

- нижние 95% и верхние 95% - соответственно нижние и верхние границы доверительных интервалов для коэффициентjd регрессии a_i . Для нахождения границ доверительных интервалов с помощью функции = СТЬДРАСПОБР (α ; $n - m - 1$) рассчитываются критические значения t – критерия (Стъюдента), а затем вычисляются соответственно нижние и верхние границы доверительных интервалов по формулам

$$\text{нижние 95\%} = \text{Коэффициент} - \text{Стандартная ошибка} \times t_{кр}; \quad (228)$$

$$\text{верхние 95\%} = \text{Коэффициент} + \text{Стандартная ошибка} \times t_{кр}. \quad (229)$$

Из расчетов, представленных в таблице 59, уравнение регрессии имеет вид: $Y(X) = 7,846596 + 0,038209 \times X_1 + 19,93594 \times X_2$.

Параметр $a_1 = 0,038209$ в полученном уравнении регрессии – это коэффициент чистой регрессии при первом факторном признаке, который показывает, что при увеличении качества почв на 1 балл, урожайность овощей в среднем увеличивается на 0,038209 т/га, при условии, что доза внесения удобрений зафиксирована на среднем уровне; а параметр $a_2 = 19,93594$ – это коэффициент чистой регрессии при втором факторном признаке, который показывает, что при увеличении дозы внесения удобрений на 1 т/га, урожайность овощей в среднем увеличивается на 19,93594 т/га, при условии, что качество почв зафиксировано на среднем уровне.

Теоретические уровни урожайности овощей $Y(X)$ рассчитываются путем подстановки в полученное уравнение (регрессионную модель) каждого значения независимых переменных X_1 и X_2 . При использовании приложения «Пакет анализа» Microsoft Excel к выше указанным шагам необходимо добавить шаг «остатки». Для вывода результатов - «ОК». (табл. 60).

В таблице 60 теоретические значения результативного признака – значения урожайности овощей, возможные при каждом имеющихся значениях качества почв и доз внесения удобрений, рассчитанные по полученной регрессионной модели, представлены в графе «Урожайность овощей (теоретическая), т/га., $Y(X)$ Предсказанное Y ». Значения в графе «остатки» определяется как разница между каждым фактическим Y и каждым теоретическим $Y(X)$ значением результативного признака и показывают влияние случайных факторов в каждом наблюдении.

Таблица 60

Вывод остатка

<i>Наблюдение</i>	Урожайность овощей (фактическая), т/га, Y	Урожайность овощей (теоретическая), т/га., $Y(X)$. Предсказанное Y	<i>Остатки</i>
1	2	3	4
1	16,00	15,42876	0,571236
2	18,00	17,88086	0,119139
3	15,00	13,74421	1,255788

Продолжение таблицы 60

1	2	3	4
4	12,00	11,55958	0,440425
5	23,00	25,1259	-2,1259
6	20,00	21,06229	-1,06229
7	9,00	11,55958	-2,55958
8	13,00	12,671	0,329002
9	31,50	29,41876	2,081242
10	14,00	14,74101	-0,74101
11	18,00	16,9007	1,099303
12	21,50	20,87125	0,628748
13	23,00	24,40488	-1,40488
14	22,00	24,58266	-2,58266
15	16,00	14,93205	1,067948
16	18,00	18,88597	-0,88597
17	29,00	27,1577	1,842297
18	13,00	11,78883	1,211173
19	14,00	14,04988	-0,04988
20	28,00	27,23412	0,76588
Итого	374,00	373,99999	0,000011

Согласно методу наименьших квадратов, суммы теоретических и эмпирических значений результативного признака равны, а сумма значений остатков равна или стремится к нулю.

Для приведения коэффициентов чистой регрессии в сопоставимый вид, определяются стандартизованные коэффициенты регрессии по формулам 221, 222. Для этого следует сделать вспомогательные расчеты. При использовании приложения «Пакет анализа» Microsoft Excel, алгоритм расчета следующий: внести исходные данные в ячейки листа Microsoft Excel, во вкладке «данные» выбрать «анализ данных»; в меню выбрать «однофакторный дисперсионный анализ», выделить во «входной интервал» ячейки с исходными данными, для вывода результатов нажать «ОК». Для вывода результатов - «ОК». (табл. 61).

Столбцы таблицы 61 имеют следующую интерпретацию:

- столбец «счет» - показывает число наблюдений;
- столбец «сумма» - показывает общие суммы по графам X_1 , X_2 и по графе Y ;
- столбец «среднее» - показывает значение статистических средних по признакам X_1 , X_2 и по признаку Y ;

- столбец «дисперсия» - показывает значение дисперсий по признакам X_1 , X_2 и по признаку Y ;
- столбец df – показывает число степеней свободы. Для строки «между группами» число степеней свободы определяется количеством факторных признаков m в уравнении регрессии; для строки «внутри групп» число степеней свободы определяется разностью между общим числом значений признака по X_1 , X_2 и по Y – $3n$ и количеством переменных в уравнении регрессии m ; для строки «итого» число степеней свободы определяется суммой степеней свободы для строк «между группами» и «внутри групп».
- столбец SS – показывает сумму квадратов отклонений. Для строки «между группами» - это сумма квадратов отклонений теоретических данных от среднего; для строки «внутри групп» - это сумма квадратов отклонений эмпирических данных от теоретических; для строки «итого» - это сумма квадратов отклонений эмпирических данных от среднего.
- столбец MS - показывает дисперсии, для строки «между группами» - это факторная дисперсия; для строки «внутри групп» - это остаточная дисперсия, рассчитываемые по формуле

$$MS = SS/df. \quad (230)$$

- столбец F – показывает расчетное значение F - критерия Фишера, вычисляемое по формуле

$$F_p = MS (\text{между группами}) / MS (\text{внутри групп}). \quad (231)$$

- столбец $\text{Значимость } F$ – значение уровня значимости, соответствующее вычислительному значению F_p . Определяется с помощью функции: $FPACП(F_p; df (\text{между группами}); df (\text{внутри групп}))$.

На основании расчетных данных таблицы 61 стандартизованные коэффициенты регрессии составят:

- коэффициенты эластичности: $\mathcal{E}_1 = a_1 \times \frac{\bar{X}_1}{\bar{Y}} = 0,038209 \times \frac{69,35}{18,7} = 0,1417$;

$$\mathcal{E}_2 = a_2 \times \frac{\bar{X}_2}{\bar{Y}} = 19,93594 \times \frac{0,4115}{18,7} = 0,4387$$

- бетта – коэффициенты:

$$\beta_2 = a_2 \times \frac{\sigma_{x_2}}{\sigma_y} = 0,038209 \times \frac{\sqrt{303,8184}}{\sqrt{36,40526}} = 0,038209 \times \frac{17,4304}{6,03} = 0,1104.$$

$$\beta_2 = a_2 \times \frac{\sigma_{x_2}}{\sigma_y} = 19,93594 \times \frac{\sqrt{0,069371}}{\sqrt{36,40526}} = 19,93594 \times \frac{0,2634}{6,03} = 0,8708$$

Таблица 61

Результаты расчетов однофакторного дисперсионного анализа

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Столбец 1, У	20	374	18,7	36,40526		
Столбец 2, X ₁	20	1387	69,35	303,8184		
Столбец 3, X ₂	20	8,23	0,4115	0,069371		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	51016,06	2	25508,03	224,877	9,03E-28	3,158843
Внутри групп	6465,568	57	113,431			
Итого	57481,62	59				

Расчет коэффициентов эластичности показал, что при увеличении качества почв на 1%, урожайность овощей возрастает в среднем на 0,1417%, при условии, что доза внесения удобрений фиксируется на среднем уровне; а при увеличении дозы внесения удобрений на 1%, урожайность овощей возрастает в среднем на 0,4387%, при условии, что качество почв фиксируется на среднем уровне. При этом, так как значение коэффициентов эластичности меньше единицы, то зависимость урожайности овощей от качества почв и дозы внесения удобрений можно считать неэластичной, то есть – результативный признак изменяется меньшими темпами, чем факторные признаки.

Расчет бетта-коэффициентов показал, что при увеличении качества почв на одну величину своего среднего квадратического отклонения, урожайность овощей увеличивается в среднем на 0,1104 величины своих средних квадратических отклонений, при условии, что доза внесения удобрений фиксируется на среднем уровне; а при увеличении дозы внесения удобрений ан

одну величину своего среднего квадратического отклонения, урожайность овощей увеличивается в среднем на 0,8707 величины своего среднего квадратического отклонения. Путем сравнения значений бетта-коэффициентов можно выявить, какой фактор оказывает наибольшее влияние на результативный признак. Так, из расчетов видно, что значение бетта-коэффициента при втором факторе больше значения бетта-коэффициента при первом факторе, это свидетельствует о том, что по данной совокупности доза внесения удобрений оказывает большее влияние на изменение урожайности овощей, чем качество почв.

10.5 Показатели тесноты связи

Задачей статистики при анализе взаимосвязей и зависимостей является определение наличия связи между признаками и оценка ее тесноты (силы) – степени ее приближения к функциональной зависимости. Оценка тесноты (силы) связи между признаками предполагает определение меры соответствия вариации результативного признака вариации факторных признаков.

Для измерения тесноты связи используются следующие показатели:

1. Эмпирическое корреляционное отношение.
2. Индекс корреляции (теоретическое корреляционное отношение).
3. Линейный коэффициент корреляции.

Эмпирическое корреляционное отношение определяется в процессе предварительного анализа на основе аналитической группировки. Используется при нелинейной зависимости. Определяется на основании расчета дисперсий результативного признака по формуле

$$\eta = \sqrt{\frac{\sigma^2 - \bar{\sigma}^2}{\sigma^2}} \quad (232)$$

или

$$\eta = \sqrt{1 - \frac{\bar{\sigma}^2}{\sigma^2}}, \quad (233)$$

или

$$\eta = \sqrt{\frac{\bar{\sigma}^2}{\sigma^2}}, \quad (234)$$

где η - эмпирическое корреляционное отношение;

$\bar{\sigma}^2$ - средняя из частных (групповых) дисперсий;

σ^2 - общая дисперсия;

$\bar{\delta}^2$ - межгрупповая дисперсия (дисперсия групповых средних).

Эмпирическое корреляционное отношение принимает значения в пределах от нуля до единицы и показывает тесноту связи между факторным и результативным признаками: чем ближе его значение к нулю, тем связь слабее, чем ближе его значение к единице, тем связь сильнее. (Пример расчета рассмотрен в главе 5).

Теоретическое корреляционное отношение (индекс корреляции) также используется при нелинейной зависимости, определяется по формуле

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}, \quad (235)$$

где η - теоретическое корреляционное отношение;

σ^2 - дисперсия эмпирических (фактических) значений результативного признака;

δ^2 - дисперсия теоретических (рассчитанных по уравнению регрессии) значений результативного признака.

Расчет корреляционного отношения основан на правиле сложения (разложения) дисперсий, которое заключается в том, что общая дисперсия равна сумме внутригрупповой дисперсии (обусловленной остаточными факторами) и межгрупповой дисперсии (обусловленной основным изучаемым фактором):

$$\sigma^2 = \bar{\delta}^2 + \bar{\sigma}^2 \quad (236)$$

Теоретическое корреляционное отношение принимает значения в пределах от нуля до единицы и показывает тесноту связи между факторным и

результативным признаками: чем ближе его значение к нулю, тем связь слабее, чем ближе его значение к единице, тем связь сильнее.

Пример расчета теоретического корреляционного отношения (индекса корреляции). На основании данных таблицы 56 индекс корреляции рассчитывается при помощи использования приложения «Пакет анализа» Microsoft Excel инструмента «описательная статистика». Алгоритм расчета следующий: во вкладке «данные» выбрать вкладку «анализа данных»; ввести в ячейки листа Microsoft Excel исходные данные (табл.56); выбрать инструмент «описательная статистика»; выделить входной интервал в диапазоне обеих граф; отметить «итоговая статистика» и «уровень надежности (0,95)». Для вывода результатов - «ОК». (табл. 62).

В таблице 62 сгенерированы все результаты расчета: степенных средних арифметических (среднее), структурных средних (медиана, мода), характеристик ряда распределения (эксцесс, асимметричность), показателей вариации (стандартное отклонение, дисперсия выборки, размах вариации - интервал), максимальное и минимальное значение признака в ряду распределения (максимум, минимум), итог по графам (сумма), число наблюдений (счет), значение коэффициента кратности при заданном уровне надежности (уровень надежности 0,95%).

Таблица 62

Результаты расчетов описательной статистики

<i>Столбец1 (Урожайность овощей (фактическая), т/га, Y)</i>	<i>Значения</i>	<i>Столбец2 (Урожайность овощей (теоретическая), т/га, Y(X). Предсказанное Y)</i>	<i>Значения</i>
1	2	3	4
Среднее	18,7	Среднее	18,7
Стандартная ошибка	1,349171	Стандартная ошибка	1,231272
Медиана	18	Медиана	19,69512
Мода	18	Мода	22,06443
Стандартное отклонение	6,033677	Стандартное отклонение	5,506415
Дисперсия выборки	36,40526	Дисперсия выборки	30,3206
Эксцесс	-0,23094	Эксцесс	-1,63136
Асимметричность	0,620535	Асимметричность	-0,1255
Интервал	22,5	Интервал	15,79544
Минимум	9	Минимум	11,00762
Максимум	31,5	Максимум	26,80306

Продолжение таблицы 62

1	2	3	4
Сумма	374	Сумма	374
Счет	20	Счет	20
Уровень надежности (95,0%)	2,823848	Уровень надежности (95,0%)	2,577081

На основании данных таблицы 62, теоретическое корреляционное отношение (индекс корреляции) составит:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{\frac{30,3206}{36,40526}} = 0,9126$$

Расчет теоретического корреляционного отношения (индекса корреляции) показал, что между урожайностью овощей и качеством почв наблюдается сильная по тесноте связь, так как значение теоретического корреляционного отношения (индекса корреляции) достаточно приближено к единице.

Линейный коэффициент корреляции является частным случаем индекса корреляции (введен Пирсоном, Эджвортом и Велдоном) и характеризует тесноту и направление связи между признаками при наличии между ними линейной зависимости. Определяется по формуле

$$\tau = \frac{X\bar{Y} - \bar{X} \times \bar{Y}}{\sigma_x \times \sigma_y}, \quad (237)$$

где $X\bar{Y}$ - средняя из произведений факторного и результативного признаков;

\bar{X} - среднее значение факторного признака;

\bar{Y} - среднее значение результативного признака;

σ_x - среднее квадратическое отклонение факторного признака;

σ_y - среднее квадратическое отклонение результативного признака.

При изучении совокупностей малого объема линейный коэффициент корреляции определяется по формуле

$$\tau = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2 \times \sigma_x \times \sigma_y} \quad (238)$$

Линейный коэффициент корреляции принимает значения в пределах от -1

до +1. Чем ближе его значение к нулю, тем связь между признаками слабее, чем ближе его значение к (+;-) единице, тем связь сильнее.

Интерпретация коэффициента линейной корреляции представлена в таблице 63.

Таблица 63

Оценка линейного коэффициента корреляции [22]

Значение коэффициент парной линейной корреляции	Характер связи	Интерпретация связи
$r = 0$	отсутствует	-
$0 < r < 1$	прямая	с увеличением X увеличивается Y
$-1 < r < 0$	обратная	с увеличением X уменьшается Y
$r = 1$	функциональная	каждому значению факторного признака строго соответствует одно значение результативного признака

Оценка степени тесноты связи на основе коэффициента парной линейной корреляции производится по таблице значений Чэддока (табл. 65). Для удобства можно использовать основные значения коэффициента парной линейной корреляции представленные в таблице 64.

Таблица 64

Значения коэффициента парной линейной корреляции [22]

Значение коэффициента парной линейной корреляции	Теснота связи
$r > 0,7$	Связь сильная (тесная)
$0,3 < r < 0,7$	Связь средняя
$r < 0,3$	Связь слабая

Таблица 65

Шкала Чэддока [11]

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

Пример расчета парного коэффициента корреляции. Расчет парного линейного коэффициента корреляции производится для определения тесноты

(силы) и направления связи между урожайностью овощей и качеством почв на основании данных таблицы 54. При использовании приложения «Пакет анализа» Microsoft Excel, алгоритм расчетов включает следующие шаги: во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «корреляция», ввести исходные данные в строку «входной интервал», выделив ячейки листа Microsoft Excel со значениями результативного признака и факторного признака вместе. Для вывода результатов - «ОК». (табл. 66). Результаты расчетов в таблице 66 представлены в виде корреляционной матрицы.

Таблица 66

Результаты расчетов коэффициента парной корреляции с использованием инструмента «корреляция»

	<i>Столбец 1</i>	<i>Столбец 2</i>
Столбец 1	1	
Столбец 2	0,912613	1

Также коэффициент парной корреляции можно определить при помощи инструмента «регрессия». Алгоритм расчетов включает следующие шаги: во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «регрессия», ввести исходные данные – значения результативного признака в окно «входной интервал Y» и исходные данные – значения факторного признака в окно «входной интервал X», отметить «уровень надежности 0,95». Для вывода результатов - «ОК». (табл. 67).

Таблица 67

Результат расчетов коэффициента корреляции с использованием инструмента «регрессия»

ВЫВОД ИТОГОВ	
<i>Регрессионная статистика</i>	
Парный R	0,912613
R-квадрат	0,832863
Нормированный R-квадрат	0,823578
Стандартная ошибка	2,534304
Наблюдения	20

Данные таблицы 67 соответствуют следующим статистическим показателям:

- множественный R – коэффициент корреляции R;
- R-квадрат – коэффициент детерминации R^2 ;
- стандартная ошибка – остаточное стандартное отклонение;
- наблюдения – число наблюдений, n.

Расчет коэффициента корреляции показал, что между урожайностью овощей и качеством почв наблюдается сильная по тесноте связь, так как значение коэффициента корреляции больше 0,7 (табл.64); и высокая по шкале Чэддока (табл. 65). При этом по направлению связь прямая, так как коэффициент корреляции имеет положительный знак.

Для измерения тесноты связи при множественной корреляционной зависимости определяется множественный (совокупный) и частные коэффициенты корреляции.

Множественный коэффициент корреляции определяется по формуле

$$R_{yX_1X_2 \dots} = \sqrt{\frac{\delta^2}{\sigma^2}} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma^2}}, \quad (239)$$

где $\sigma_{ост}^2$ - остаточная дисперсия;

σ^2 - общая дисперсия результативного признака;

δ^2 - дисперсия теоретических значений результативного признака, рассчитанных по уравнению множественной регрессии.

При оценке тесноты связи между результативным и двумя факторными признаками коэффициент множественной корреляции определяется по формуле, состоящей из коэффициентов парной линейной корреляции

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \times r_{yx_2} \times r_{x_1x_2}}{1 - r_{x_1x_2}^2}}, \quad (240)$$

Коэффициент множественной корреляции принимает значения в пределах от 0 до +1. Чем ближе его значение к 1, тем связь между признаками сильнее, чем ближе его значение к 0, тем связь слабее.

При небольшом числе наблюдений (менее 20) величина коэффициента множественной корреляции корректируется по формуле

$$\hat{R}_{yX_1X_2\dots X_k} = \sqrt{1 - (1 - R^2) \times \frac{n-1}{n-k-1}}, \quad (241)$$

где \hat{R} - скорректированное значение коэффициента множественной корреляции;

n – число единиц наблюдения;

k – число факторных признаков.

Пример расчет коэффициента множественной корреляции. Расчет коэффициентов парной корреляции при множественной модели связи производится на основании исходных данных таблицы 58. При использовании приложения «Пакет анализа» Microsoft Excel, алгоритм расчетов включает следующие шаги: во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «корреляция», ввести исходные данные в строку «входной интервал», выделив ячейки листа Microsoft Excel со значениями результативного признака и всех факторных признаков вместе. Для вывода результатов - «ОК». (табл. 68). Результаты расчетов в таблице 68 представлены в виде корреляционной матрицы.

Таблица 68

Результаты расчетов коэффициентов парной корреляции при множественной модели связи

	Столбец 1, Y	Столбец 2, X_1	Столбец 3, X_2
Столбец 1, Y	1		
Столбец 2, X_1	0,912613	1	
Столбец 3, X_2	0,972003	0,921843	1

Расчет показал, что между урожайностью овощей и качеством почв наблюдается прямая по направлению и сильная по тесноте связь, так как значение коэффициента парной корреляции между Y и X_1 достаточно приближено к единице и имеет положительный знак; между урожайностью овощей и дозой внесения удобрений также наблюдается прямая по направлению и сильная по тесноте связь, так как значение коэффициента

парной корреляции между Y и X_1 достаточно приближено к единице и имеет положительный знак. При этом, из сравнения коэффициентов парной корреляции видно, что между урожайностью овощей и дозой внесения удобрений связь сильнее, чем между урожайностью овощей и качеством почв, так как значение коэффициента корреляции между Y и X_2 больше его значения между Y и X_1 .

Также коэффициент парной корреляции можно определить при помощи инструмента «регрессия». Алгоритм расчетов включает следующие шаги: во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «регрессия», ввести исходные данные – значения результативного признака в окно «входной интервал Y » и исходные данные – значения факторных признаков вместе в окно «входной интервал X », отметить «уровень надежности 0,95». Для вывода результатов - «ОК». (табл. 69).

Данные таблицы 69 соответствуют следующим статистическим показателям:

- множественный R – коэффициент корреляции R ;
- R -квадрат – коэффициент детерминации R^2 ;
- стандартная ошибка – остаточное стандартное отклонение;
- наблюдения – число наблюдений, n .

Таблица 69

Результат расчетов множественного коэффициента корреляции

ВЫВОД ИТОГОВ	
<i>Регрессионная статистика</i>	
Множественный R	0,972944
R -квадрат	0,94662
Нормированный R -квадрат	0,94034
Стандартная ошибка	1,473749
Наблюдения	20

Расчет показал, что между урожайностью овощей, качеством почв и дозой внесения удобрений совместно наблюдается сильная по тесноте связь,

так как значение коэффициента корреляции больше 0,7 (табл.64); и высокая по шкале Чэддока (табл. 65).

При множественной корреляционной зависимости переменные, включенные в модель анализа, взаимосвязаны между собой. Теснота такой зависимости определяется при помощи частных коэффициентов корреляции, которые характеризуют степень влияния одного из факторов на результат при условии, что все остальные факторы зафиксированы на постоянном (среднем) уровне.

В зависимости от того, влияние скольких факторов исключается, частные коэффициенты корреляции могут быть двух порядков: 1) если исключается влияние одного фактора, определяется частный коэффициент корреляции первого порядка; 2) если исключается влияние двух факторов, определяется частный коэффициент корреляции второго порядка и т.д.

Значение частного коэффициента корреляции обычно не соответствует значению парного линейного коэффициента корреляции. Частный коэффициент корреляции первого порядка при исключении влияния второго фактора определяется по формуле на основании парных линейных коэффициентов корреляции между соответствующими признаками

$$r_{YX_1(X_2)} = \frac{r_{YX_1} - r_{YX_2} \times r_{X_1X_2}}{\sqrt{(1 - r_{YX_2}^2) \times (1 - r_{X_1X_2}^2)}}. \quad (242)$$

Для оценки степени зависимости вариации результативного признака от уровня вариации факторных признаков, включенных в модель, определяется коэффициент детерминации по формуле

$$d^2 = r^2 \times 100\% \quad (243)$$

или

$$D^2 = R^2 \times 100\%, \quad (244)$$

или

$$\eta^2 = \frac{\delta^2}{\sigma^2} \times 100\%, \quad (245)$$

где d^2 - коэффициент детерминации при парной корреляционной зависимости;

- r - коэффициент парной линейной корреляции;
 D^2 - коэффициент детерминации при множественной корреляционной зависимости;
 η^2 - коэффициент детерминации при нелинейной зависимости;
 R – коэффициент множественной корреляции;
 σ^2 - общая дисперсия;
 δ^2 - межгрупповая дисперсия (дисперсия групповых средних).

Коэффициент детерминации показывает, какая доля вариации результативного признака обусловлена вариацией факторных признаков, включенных в модель, и какая доля вариации результативного признака обусловлена вариацией остаточных, неучтенных в данной модели факторов.

При оценке степени влияния вариации каждого отдельного фактора на вариацию результативного признака совокупный коэффициент детерминации раскладывается на частные коэффициенты детерминации.

Частный коэффициент детерминации показывает, на сколько процентов вариация результативного признака обусловлена вариацией факторного признака, включенного в модель. Определяется по формуле

$$d_{x_i} = r_{yx_i} \times \beta_{x_i} \times 100\% , \quad (246)$$

где d_{x_i} - частный коэффициент детерминации;

r_{yx_i} - парный линейный коэффициент корреляции;

β_{x_i} - β -коэффициент.

Коэффициенты детерминации при парной и множественной моделях рассчитаны при помощи приложения «пакет анализа» Microsoft Excel в таблицах 67 и 69 (R-квадрат).

Коэффициент детерминации при парной модели $R^2 = 0,832863$ показывает, что 83,2863% изменения урожайности овощей обусловлены изменением качества почв, а остальные 16,7137% обусловлены изменением остаточных, неучтенных в данном случае факторов.

Коэффициент детерминации при множественной модели $R^2 = 0,94662$ показывает, что 94,662% изменения урожайности овощей обусловлены совместным изменением качества почв и дозы внесения удобрений, а остальные 5,338% обусловлены изменением остаточных, неучтенных в данном случае факторов.

10.6 Корреляция качественных признаков

Методы корреляционно-регрессионного анализа основаны на изучении взаимосвязи и зависимости количественных признаков. При их использовании определяются основные параметры распределения, поэтому эти методы называются параметрическими методами.

Кроме этого задача статистики также заключается в оценке тесноты связи между признаками, которые не имеют количественной характеристики и для которых невозможно рассчитать какие-либо параметры, то есть – между качественными (атрибутивными) признаками. Применение параметрических методов для качественных признаков в их обычном виде невозможно. Поэтому в статистике разработаны специальные методы, с помощью которых возможно определить тесноту связи между качественными признаками, не используя при этом количественные значения признака. Такие методы называются непараметрическими.

При наличии соотношения между вариацией качественных признаков определяется их взаимосвязанность или - ассоциация. Для расчета коэффициентов связи между качественными признаками строятся таблицы взаимной сопряженности, на основе которых определяются следующие показатели (рис. 59).

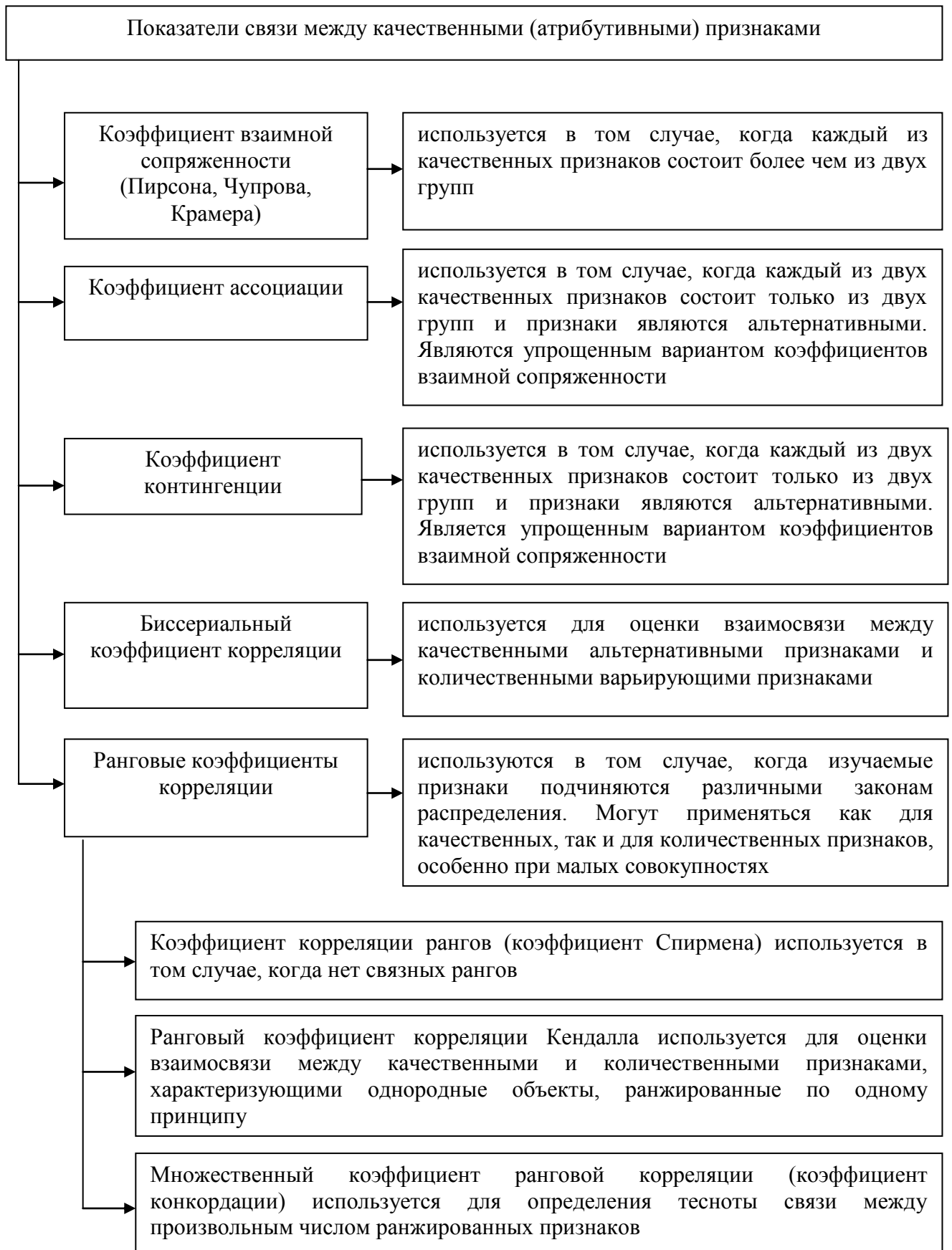


Рисунок 59 – Виды показателей связи между качественными (атрибутивными) признаками

Коэффициент взаимной сопряженности определяется по формулам [22]

- коэффициент Пирсона
$$C = \sqrt{\frac{\varphi^2}{1 + \varphi^2}}; \quad (247)$$

- коэффициент Чупрова
$$K_q = \sqrt{\frac{\varphi^2}{\sqrt{(K_1 - 1) \times (K_2 - 1)}}} \quad (248)$$

или

$$K_q = \sqrt{\frac{\chi^2}{n + \chi^2}}, \quad (249)$$

или

$$K_q = \sqrt{\frac{\chi^2}{n \times \sqrt{(k_1 - 1) \times (k_2 - 1)}}}, \quad (250)$$

- коэффициент Крамера
$$K_q = \sqrt{\frac{\chi^2 \div n}{\min\{k_1 - 1, k_2 - 1\}}}, \quad (251)$$

где φ^2 - показатель средней квадратической сопряженности;

K_1 - число значений (групп) первого признака;

K_2 - число значений (групп) второго признака;

n - число наблюдений;

k_1 - число строк в таблице взаимной сопряженности;

k_2 - число граф в таблице взаимной сопряженности;

χ^2 - критерий согласия (критерий Пирсона), используемый для проверки статистической гипотезы о виде распределения.

Показатель средней квадратической сопряженности определяется по формуле [22]

$$\varphi^2 = \Sigma \frac{f^2}{f_1 \times f_2}, \quad (252)$$

где f_1 и f_2 - частоты по каждому из признаков.

Коэффициент взаимной сопряженности принимает значения в пределах от нуля до + единицы. Чем ближе его значение к единице, тем теснее связь между изучаемыми признаками, чем ближе его значение к нулю, тем связь между изучаемыми признаками слабее.

Коэффициент контингенции определяется по формуле [22]

$$K_{\kappa} = \frac{ad + bc}{\sqrt{(a+b) \times (c+d) \times (a+c) \times (b+d)}}, \quad (253)$$

где a, b, c, d – частоты альтернативных признаков.

Коэффициент контингенции принимает значения в пределах от нуля до +единицы. Если $K_{\kappa} \geq 0,3$, то связь между альтернативными признаками считается подтвержденной.

Коэффициент ассоциации определяется по формуле [22]

$$K_a = \frac{ad - bc}{ad + bc}. \quad (254)$$

Коэффициент ассоциации принимает значения в пределах от нуля до +единицы. Если $K_{\kappa} \geq 0,5$, то связь между альтернативными признаками считается подтвержденной.

Значение коэффициента контингенции всегда меньше значения коэффициента ассоциации, так как коэффициент контингенции дает более приблизительную оценку тесноты связи.

Биссериальный коэффициент корреляции определяется по формуле [22]

$$r = \frac{|\bar{Y}_2 - \bar{Y}_1|}{\sigma_y} \times \frac{pq}{Z}, \quad (255)$$

где \bar{Y}_2 и \bar{Y}_1 – средние в группах;

σ_y – среднее квадратическое отклонение фактических значений признака от среднего уровня;

p – доля первой группы;

q – доля второй группы;

Z – табличные значения z -распределения в зависимости от p .

Биссериальный коэффициент корреляции принимает значения в пределах от нуля до + единицы. Чем ближе его значение к единице, тем теснее связь между изучаемыми признаками, чем ближе его значение к нулю, тем связь слабее.

Ранжирование – это упорядочение объектов изучения, которое выполняется на основе предпочтения. Ранг – это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин.

Принцип нумерации значений изучаемых признаков является основой непараметрических методов изучения взаимосвязи между изучаемыми как качественными, так и количественными признаками при условии, если их значения возможно упорядочить или проранжировать по степени убывания или возрастания признака. При этом используются следующие коэффициенты: 1) коэффициент корреляции рангов (коэффициент Спирмена); 2) ранговый коэффициент корреляции Кендалла; 3) множественный коэффициент ранговой корреляции (коэффициент конкордации).

Расчет ранговых коэффициентов выполняется в несколько этапов:

- 1) значения факторного признака X ранжируются в порядке возрастания или убывания;
- 2) значения результативного признака Y располагаются в порядке, соответствующем значениям признака X ;
- 3) для каждого ранга Y определяется число следующих за ним рангов, превышающих его величину;
- 4) полученные числа суммируются и определяется величина P – мера соответствия последовательностей рангов по X и по Y , которая учитывается со знаком «+»;
- 5) для каждого ранга Y определяется число следующих за ним рангов, меньших его величины;
- 6) суммарная величина обозначается Q и фиксируется со знаком «-»;
- 7) упорядоченная последовательность является исходной для построения квадратной матрицы размерностью $(n \times n)$; 8) определяется сумма баллов по всем членам ряда.

Коэффициент корреляции рангов Спирмена определяется по формуле [22]

$$\rho_{x/y} = 1 - \frac{6\sum d_i^2}{n \times (n^2 - 1)}, \quad (256)$$

где d_i^2 - квадрат разности рангов признаков X и Y;

n – число наблюдений (число пар рангов).

Коэффициент Спирмена принимает значения в пределах от минус единицы до плюс единицы. Чем ближе его значение к единице, тем связь между изучаемыми признаками сильнее. Знак коэффициента свидетельствует о направлении связи, если знак «+» (положительная дробь), то связь прямая, если знак «-» (отрицательная дробь), то связь обратная.

Ранговый коэффициент Кендалла определяется по формуле (когда нет связных рангов) [22]

$$\tau = \frac{2S}{n \times (n - 1)}, \quad (257)$$

где n – число наблюдений;

S – сумма элементов построенной квадратной матрицы, расположенных выше главной диагонали.

Коэффициент Кендалла принимает значения в пределах от минус единицы до плюс единицы. Чем ближе его значение к единице, тем связь между изучаемыми признаками сильнее. Знак коэффициента свидетельствует о направлении связи, если знак «+», то связь прямая, если знак «-», то связь обратная.

Коэффициент Кендалла, как правило, меньше коэффициента Спирмена (рассчитанных по одним и тем же данным). При достаточно большом объеме совокупности значения этих коэффициентов соотносятся следующим образом:

$$\tau = \frac{2}{3} \times \rho_{x/y}. \quad (258)$$

Связь между изучаемыми признаками можно считать статистически значимой, если значения коэффициентов Спирмена и Кендалла больше 0,5. Если связь между изучаемыми признаками прямая, то с увеличением ранга признака X, ранг признака Y также возрастает. При обратной связи между

признаками возрастанию рангов признака X соответствует убывание рангов признака Y . При тесной связи между признаками ранги X и Y в основном совпадают. Если в результате ранжирования никакой порядок возрастания или убывания рангов не обнаруживается, то связь между признаками отсутствует.

Коэффициент конкордации определяется по формуле [22]

$$W = \frac{12S}{m^2 \times (n^3 - n)}, \quad (259)$$

где m – количество факторов;

n – число наблюдений;

S – отклонение суммы квадратов рангов от средней квадратов рангов.

Коэффициент конкордации принимает значения в пределах от минус единицы до плюс единицы. Чем ближе его значение к единице, тем связь между изучаемыми признаками сильнее. Знак коэффициента свидетельствует о направлении связи, если знак «+», то связь прямая, если знак «-», то связь обратная.

10.7 Корреляция в рядах динамики

Комплексные (многомерные) ряды динамики включают несколько взаимосвязанных показателей. На их основе отражается зависимость результативного признака от одного или нескольких факторных признаков. Такие ряды называются связными рядами динамики. Поэтому при статистическом анализе развития явления во времени возникает необходимость оценить степень взаимосвязи в изменениях двух или более различных, но связанных между собой рядов динамики. Данная задача решается при помощи методов коррелирования следующих элементов: 1) уровней ряда динамики; 2) отклонений фактических уровней ряда динамики от тренда; 3) последовательных разностей путем исчисления парного коэффициента корреляции.

Коррелирование уровней ряда динамики достоверно характеризует тесноту связи между рядами динамики в том случае, если в каждом из рядов динамики, включенных в модель, отсутствует автокорреляция их уровней. В этом случае парный коэффициент корреляции определяется по формуле

$$r = \frac{\overline{xy} - \bar{x} \times \bar{y}}{\sigma_x \times \sigma_y}, \quad (260)$$

где X – уровни факторного ряда динамики;

Y – уровни результативного ряда динамики.

Исходя из условия отсутствия автокорреляции, прежде чем определять степень корреляционной зависимости в рядах динамики, необходимо проверить каждый из, включенных в модель рядов, на наличие или отсутствие в них автокорреляции.

В основной массе рядов динамики, между их уровнями, которые близко расположены друг от друга, существует зависимость. Данная зависимость может быть представлена в виде корреляционной зависимости между рядом $Y_1, Y_2, Y_3, \dots, Y_n$ и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени: $Y_{1+h}, Y_{2+h}, Y_{3+h}, \dots, Y_{n+h}$. Указанное временное смещение называется сдвигом (лагом), а явление сдвига - автокорреляцией.

Автокорреляция является наиболее существенной между смежными уровнями ряда динамики – последующим и предыдущим и бывает двух видов:

- 1) автокорреляция в наблюдениях за одной и более переменными;
- 2) автокорреляция ошибок (автокорреляция в отклонениях от тренда), которая приводит к искажению величин средних квадратических ошибок коэффициентов регрессии, что затрудняет построение доверительных интервалов для коэффициентов регрессии и проверки их значимости.

Автокорреляция измеряется при помощи нециклического коэффициента автокорреляции. Порядок коэффициента корреляции определяет временной лаг (L): коэффициент автокорреляции первого порядка $L = 1$; коэффициент автокорреляции второго порядка $L = 2$ и т.д.

Наиболее сильные искажения результатов анализа возникают при корреляции между исходными уровнями ряда динамики и теми же уровнями, сдвинутыми на одну единицу времени ($L = 1$), поэтому чаще используется коэффициент автокорреляции первого порядка, который определяется по формуле

$$r_a = \frac{\overline{Y_t \times Y_{t+1}} - \bar{Y}_t \times \bar{Y}_{t+1}}{\sigma_{et} \times \sigma_{et+1}}, \quad (261)$$

где σ_{et} - среднее квадратическое отклонение ряда динамики Y_t ;

σ_{et+1} - среднее квадратическое отклонение ряда динамики Y_{t+1} ;

Если значение последнего уровня ряда динамики мало отличается от первого уровня, то сдвинутый ряд не укорачивается и его можно условно дополнить, приняв $Y_n = Y_1$. При этом $Y_t = Y_{t+1}$ и $\sigma_{et} = \sigma_{et+1}$, так как они рассчитываются для одного и того же ряда динамики.

При такой замене коэффициент автокорреляции определяется по формуле

$$r_a = \frac{\overline{Y_t \times Y_{t+1}} - (\bar{Y}_t)^2}{\sigma_{et}^2}. \quad (262)$$

Для оценки наличия автокорреляции фактическое значение коэффициента автокорреляции сопоставляется с табличным (критическим) значением при 5%-м или 1%-м уровне значимости.

Если фактическое значение коэффициента автокорреляции меньше табличного (критического) значения, то автокорреляция в данном ряду динамики отсутствует.

Если фактическое значение коэффициента автокорреляции больше табличного (критического) значения, то автокорреляция в данном ряду динамики существует.

Табличные (критические) значения коэффициента автокорреляции определяются по таблице Р.Андерсена (табл. 70).

Значения коэффициентов автокорреляции [22]

Объем выборки (число уровней ряда динамики)	Положительные значения		Отрицательные значения	
	5%-й уровень	1%-й уровень	5%-й уровень	1%-й уровень
5	0,253	0,297	-0,753	0,798
6	0,345	0,447	-0,708	-0,863
7	0,370	0,510	-0,674	-0,799
8	0,371	0,531	-0,625	-0,764
9	0,366	0,533	-0,593	-0,737
10	0,360	0,525	-0,564	-0,705
11	0,353	0,512	-0,539	-0,679
12	0,348	0,505	-0,516	-0,655
13	0,341	0,495	-0,497	-0,634
14	0,335	0,485	-0,479	-0,615
15	0,328	0,475	-0,462	-0,597
20	0,299	0,432	-0,399	-0,524

В случае наличия автокорреляции между уровнями ряда динамики она должна быть устранена, либо уменьшена. В целях устранения или уменьшения автокорреляции используются три основных способа: 1) способ включения времени в качестве дополнительного фактора; 2) способ последовательных разностей; 3) способ авторегрессионных преобразований.

Способ включения времени в качестве дополнительного фактора предполагает расчет коррелирования отклонений. Коррелирование отклонений от выровненных уровней (тренда) заключается в том, что коррелируются не сами уровни ряда динамики, а отклонения фактических уровней от теоретических (тренда), то есть – остаточные величины. Производится в несколько этапов:

- 1) каждый ряд динамики выравнивается по адекватному для него тренду;
- 2) определяются отклонения фактических уровней от теоретических, рассчитанных по уравнению тренда по формулам

$$d_x = X_i - X_t \quad (263)$$

и

$$d_y = Y_i - Y_t, \quad (264)$$

- 3) определяется теснота связи между рассчитанными отклонениями по формуле

$$r_{dxdy} = \frac{\Sigma d_x \times d_y}{\sqrt{\Sigma d_x^2 \times \Sigma d_y^2}}. \quad (265)$$

Способ последовательных разностей предполагает расчет коррелирования последовательных разностей. Коррелирование последовательных разностей заключается в вычитании из каждого уровня ряда динамики предшествующего ему уровня, то есть определяются разности уровней по формуле

$$\Delta = Y_i - Y_{i-1}. \quad (266)$$

В результате, при переходе от абсолютных уровней ряда динамики к их разностям, исключается влияние общей тенденции на колеблемость уровней. Если использован линейный тренд, то определяются первые разности, если использован нелинейный тренд, то определяются n-ные разности.

После этого, коэффициент корреляции, отражающий тесноту связи между разностями, определяется по формуле

$$r_{\Delta x \Delta y} = \frac{\Sigma \Delta_x \times \Delta_y}{\sqrt{\Sigma \Delta_x^2 \times \Sigma \Delta_y^2}}. \quad (267)$$

Для периода изучения, включающего длительный временной промежуток, следует рассчитывать серию скользящих коэффициентов корреляции для отдельного однородного интервала времени.

10.8 Оценка существенности показателей связи

Корреляционно-регрессионный анализ может являться основой прогнозирования изучаемых явлений. Для достижения достоверности прогнозных значений необходимо оценить существенность (значимость) следующих показателей связи: коэффициентов регрессии; линейного коэффициента корреляции; коэффициента множественной корреляции; коэффициентов корреляции качественных признаков.

Проверка адекватности моделей, построенных на основе уравнений регрессии, заключается в проверке существенности (значимости) каждого коэффициента регрессии с помощью t-критерия Стьюдента и всей регрессионной модели с помощью F-критерия Фишера.

t-критерий Стьюдента определяется по формуле

$$t_p = \frac{|a_i|}{\sqrt{\sigma_{a_i}^2}}, \quad (268)$$

где a_i – коэффициент регрессии;

$\sigma_{a_i}^2$ – дисперсия коэффициента регрессии.

Дисперсия коэффициента регрессии определяется по формуле

$$\sigma_{a_i}^2 = \frac{\sigma_y^2}{k} \quad (269)$$

или

$$\sigma_{a_i}^2 = \frac{\sigma_y \times \sqrt{1 - R^2}}{\sigma_{x_i} \times \sqrt{n} \times \sqrt{1 - R_i}}, \quad (270)$$

где σ_y^2 – дисперсия результативного признака;

k – число факторных признаков в уравнении регрессии;

R_i – множественный коэффициент корреляции.

Коэффициент регрессии признается статистически значимым, если соблюдается условие

$$t_p > t_{кр(табл)}(\alpha; \nu = n - k - 1), \quad (271)$$

где α – уровень значимости критерия проверки гипотезы о равенстве нулю коэффициентов регрессии – отсутствию связи (статистическая существенность связи утверждается при отклонении нулевой гипотезы об отсутствии связи). Задается в зависимости от уровня вероятности (обычно 0,05 при вероятности 0,95);

$\nu = n - k - 1$ – число степеней свободы (характеризует число свободно варьирующих элементов изучаемой совокупности).

Критическое (табличное) значение t-критерия определяется по таблицам в зависимости от вероятности и числа степеней свободы (табл. 71).

Критические значения t-критерия Стьюдента при уровне значимости 0,10; 0,05 и 0,01 (двухсторонний) [22]

Число степеней свободы, ν	Уровень значимости, α			Число степеней свободы, ν	Уровень значимости, α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3634	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	1,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0793	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	0,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Пример расчета t-критерия Стьюдента на основании корреляционно-регрессионного анализа, проведенного в параграфе 11.4. Фактические значения t-критерия определяются при помощи приложения «Пакет анализа» Microsoft Excel и представлены в графах t – статистика по строке «переменная X_1 » таблицы 55 – по парной регрессионной модели и по строкам «переменная X_1 » и «переменная X_2 » таблицы 59 - по множественной регрессионной модели.

Из таблицы 55 видно, что фактическое значение t – критерия Стьюдента составило 9,470809. Табличное значение t – критерия Стьюдента определяется по таблице 71 и при уровне значимости $\alpha = 0,05$ числе степеней свободы $\nu = 20 - 1 - 1 = 18$ составило 2,1009. Расчет показал, что $t_p = 9,470809 > t_{кр(табл)} = 2,1009$, следовательно коэффициент регрессии $a_1 = 0,315909$ (парная регрессионная модель) является статистически значимым. То

есть теоретические уровни результативного признака Y статистически надежны и сформировались под неслучайным воздействием фактора X_1 .

Из таблицы 59 видно, что фактические значения t – критерия Стьюдента составили: при переменной X_1 0,763423; при переменной X_2 6,018999. Табличное значение t – критерия Стьюдента определяется по таблице 71 и при уровне значимости $\alpha = 0,05$ числе степеней свободы $\nu = 20 - 2 - 1 = 17$ составило 2,1098. Расчет показал, что 1) при факторе X_1 $t_p = 0,763423 < t_{кр(табл)} = 2,1098$, следовательно коэффициент регрессии $a_1 = 0,038209$ (парная регрессионная модель) является статистически незначимым. То есть теоретические уровни результативного признака Y статистически ненадежны и сформировались под случайным воздействием фактора X_1 ; 2) при факторе X_2 $t_p = 6,018999 > t_{кр(табл)} = 2,1098$, следовательно коэффициент регрессии $a_2 = 19,93594$ является статистически значимым. То есть теоретические уровни результативного признака Y статистически надежны и сформировались под неслучайным воздействием фактора X_1 .

Проверка адекватности всей регрессионной модели осуществляется при помощи расчета F -критерия Фишера и средней ошибки аппроксимации.

F -критерий Фишера определяется по формуле

$$F = \frac{\frac{1}{k+1} \times \sum \hat{Y}_k^2}{\frac{1}{n-k-1} \times \sum (Y_i - \hat{Y}_k)^2}, \quad (272)$$

где $\hat{Y}_{i(1,2,3,...k)}$ – теоретические значения результативного признака, рассчитанные по уравнению регрессии;
 n – объем изучаемой совокупности;
 k – число факторных признаков, включенных в модель.

Модель (уравнение) регрессии признается адекватным (статистически значимым), если соблюдается условие

$$F_p > F_{кр(табл)} (\alpha = 0,05; 0,01 : \nu_1 = k + 1; \nu_2 = n - k - 1). \quad (273)$$

F-критическое (табличное) определяется по таблицам в зависимости от уровня значимости (0,05 или 0,01) и числа степеней свободы ν_1 и ν_2 (табл. 72).

Таблица 72

Таблица значений F-критерия Фишера при уровне значимости $\alpha = 0,05$ [22]

ν_2	ν_1												
	1	2	3	4	5	6	8	10	12	15	20	30	∞
1	161	200	216	225	230	234	239	242	244	246	248	250	254
2	18,5	19,0	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,84	8,79	8,74	8,70	8,66	8,62	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,91	5,86	5,80	5,75	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,68	4,62	4,56	4,50	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	4,00	3,94	3,87	3,81	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,57	3,51	3,44	3,38	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,28	3,22	3,15	3,08	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	3,07	3,01	2,94	2,86	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,91	2,85	2,77	2,70	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85	2,79	2,72	2,65	2,57	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,75	2,69	2,62	2,54	2,47	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,67	2,60	2,53	2,46	2,38	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,53	2,46	2,39	2,31	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54	2,48	2,40	2,33	2,25	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,42	2,35	2,28	2,19	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,45	2,38	2,31	2,23	2,15	1,96
18	4,41	3,53	3,16	2,93	2,77	2,66	2,51	2,41	2,34	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38	2,31	2,23	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,28	2,20	2,12	2,04	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,32	2,25	2,18	2,10	2,01	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30	2,23	2,15	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,27	2,20	2,13	2,05	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,25	2,18	2,11	2,03	1,94	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,24	2,16	2,09	2,01	1,92	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,22	2,15	2,07	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,20	2,13	2,06	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,19	2,12	2,04	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,18	2,10	2,03	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,09	2,01	1,93	1,84	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,10	2,04	1,97	1,88	1,80	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	2,00	1,92	1,84	1,74	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,01	1,95	1,87	1,81	1,71	1,44
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,98	1,85	1,79	1,67	1,58	1,26
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,95	1,82	1,71	1,60	1,51	1,18
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,83	1,75	1,67	1,57	1,46	1,00

Пример расчета F-критерия Фишера на основании корреляционно-регрессионного анализа, проведенного в параграфе 11.4. Для расчета F –

критерия Фишера можно использовать приложение «Пакет анализа» Microsoft Excel. Алгоритм расчета для парной регрессионной модели включает следующие шаги: внести исходные данные в ячейки листа Microsoft Excel; во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «регрессия», выделить диапазон входного интервала исходных данных результативного признака Y и диапазон входного интервала факторного признака X , ввести данные; отметить «уровень надежности 0,95», Для вывода результатов - «ОК». (табл. 73). Алгоритм расчета для множественной регрессионной модели включает следующие шаги: внести исходные данные в ячейки листа Microsoft Excel; во вкладке «данные» перейти по ссылке «анализ данных», выбрать из меню инструмент «регрессия», выделить диапазон входного интервала исходных данных результативного признака Y и диапазон входного интервала факторных признаков X_1 и X_2 вместе, ввести данные; отметить «уровень надежности 0,95», Для вывода результатов - «ОК». (табл. 74).

Таблица 73

Результаты расчетов при парной регрессионной модели

Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	1	576,0915	576,0915	89,69623	2,05E-08
Остаток	18	115,6085	6,422695		
Итого	19	691,7			

Таблица 74

Результаты расчетов при множественной регрессионной модели

Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	2	654,7771	327,3886	150,7359	1,52E-11
Остаток	17	36,92289	2,171935		
Итого	19	691,7			

Из таблицы 73 видно, что фактическое значение F – критерия Фишера составило 89,69623. Табличное значение F – критерия Фишера определяется по таблице 72 и при уровне значимости $\alpha = 0,05$ числе степеней свободы

$\nu_1 = 2 + 1 = 3; \nu_2 = 20 - 1 - 1 = 18$ составило 3,16. Расчет показал, что $F_p = 89.69623 > F_{кр(табл)} = 3.16$, следовательно парная регрессионная модель в целом является статистически значимой. То есть теоретические уровни результативного признака Y статистически надежны и сформировались под неслучайным воздействием фактора X_1 .

Из таблицы 74 видно, что фактическое значение F – критерия Фишера составило 150,7359. Табличное значение F – критерия Фишера определяется по таблице 71 и при уровне значимости $\alpha = 0,05$ числе степеней свободы $\nu_1 = 3 + 1 = 4; \nu_2 = 20 - 2 - 1 = 17$ составило 2,96. Расчет показал, что $F_p = 150,7359 > F_{кр(табл)} = 2,96$, следовательно множественная регрессионная модель в целом является статистически значимой. То есть теоретические уровни результативного признака Y статистически надежны и сформировались под неслучайным совместным воздействием факторов X_1 и X_2 .

Средняя ошибка аппроксимации определяется по формуле

$$\bar{\varepsilon} = \frac{1}{n} \times \sum \frac{|Y - \bar{Y}_{1,2,\dots,k}|}{Y} \times 100. \quad (274)$$

Модель (уравнение) регрессии признается адекватным (статистически значимым), если значение средней ошибки аппроксимации не превышает 12-15%.

Пример расчета средней ошибки аппроксимации на основании проведенного корреляционно-регрессионного анализа в параграфе 11.4. При парной регрессионной модели средняя ошибка аппроксимации может быть определена графическим способом (рис. 55) или расчетным способом с использованием приложения «Пакет анализа» Microsoft Excel (табл. 67). Исходя из показанного на рисунке 55 коэффициента детерминации $R^2 = 0,8329$ и представленного в таблице 67 коэффициента детерминации $R^2 = 0,832823$, средняя ошибка аппроксимации составит $\bar{\varepsilon} = R^2 - 1 = 0,8329 - 1 = 0,1671$; $\bar{\varepsilon} = R^2 - 1 = 0,832823 - 1 = 0,167137$ Расчет показал, что средняя ошибка аппроксимации составила 16,71% (6,7137%), ее значение выходит за пределы

порогового значения 12-15%, но незначительно. Поэтому, с определенным уровнем допущения можно сказать, что полученную парную регрессионную модель можно считать достаточно удачной для последующего прогнозирования.

При множественной регрессионной модели средняя ошибка аппроксимации может быть определена расчетным способом с использованием приложения «Пакет анализа» Microsoft Excel (табл. 69). Исходя из представленного в таблице 69 коэффициента детерминации $R^2 = 0,94662$, средняя ошибка аппроксимации составит $\bar{\varepsilon} = R^2 - 1 = 0,94662 - 1 = 0,05338$. Расчет показал, что средняя ошибка аппроксимации составила 5,338%, ее значение не выходит за пределы порогового значения 12-15%. Поэтому, с определенным уровнем допущения можно сказать, что полученную множественную регрессионную модель можно считать достаточно удачной для последующего прогнозирования.

Значимость линейного коэффициента корреляции проверяется на основании расчета t-критерия Стьюдента, который определяется по формулам

- при малых совокупностях (численности изучаемой совокупности менее 50)

$$t_p = \sqrt{\frac{r^2}{1-r^2} \times (n-2)} = \frac{|r|}{\sqrt{1-r^2}} \times \sqrt{n-2}; \quad (275)$$

- при больших совокупностях (численности изучаемой совокупности более 100)

$$t_p = \frac{|r|}{\sqrt{1-r^2}} \times \sqrt{n}. \quad (276)$$

Линейный коэффициент корреляции признается статистически значимым, если соблюдается условие

$$t_p > t_{кр(табл)}(\alpha, k = n - 2). \quad (277)$$

Критическое (табличное) значение t-критерия определяется по таблицам в зависимости от вероятности и числа степеней свободы (табл. 71).

Коэффициент множественной корреляции считается значимым, если соблюдается условие

$$\frac{n-k}{k} \geq 20 \quad (278)$$

или

$$\frac{R}{\sigma_R} > 3. \quad (279)$$

Проверка значимости коэффициента множественной корреляции осуществляется на основе F-критерия Фишера, который определяется по формуле

$$F_p = \frac{\frac{1}{2} \times R^2_{Y/X_1X_2}}{\frac{1}{n-3} \times (1 - R^2_{Y/X_1X_2})}. \quad (280)$$

Множественный коэффициент корреляции признается статистически значимым, если соблюдается условие

$$F_p > F_{кр(табл)}(\alpha, \nu_1 = 2; \nu_2 = n - 3). \quad (281)$$

Критическое (табличное) значение F-критерия определяется по таблицам в зависимости от вероятности и числа степеней свободы (табл. 72).

Значимость коэффициентов корреляции качественных признаков.

Значимость коэффициента корреляции рангов Спирмена проверяется на основании t-критерия Стьюдента, расчетное значение которого определяется по формуле

$$t_p = \rho_{X/Y} \times \sqrt{\frac{n-2}{1-\rho_{X/Y}^2}}, \quad (282)$$

где $\rho_{X/Y}$ - коэффициент корреляции рангов;

n – число наблюдений (число пар рангов).

Значение коэффициента корреляции рангов Спирмена считается статистически существенным, если соблюдается условие

$$t_p > t_{кр(табл)}(\alpha; k = n - 2). \quad (283)$$

Критическое (табличное) значение критерия определяется по таблицам t-распределения Стьюдента в зависимости от уровня значимости и числа степеней свободы (табл. 71).

Значимость коэффициента конкордации проверяется на основе расчета критерия Пирсона - χ^2 , расчетное значение которого определяется по формуле

$$\chi_p^2 = \frac{12S}{m \times n \times (n-1)}, \quad (284)$$

где S – отклонение суммы квадратов рангов от вредней квадратов рангов;

m – число факторов;

n – число наблюдений.

Коэффициент конкордации считается статистически значимым, если соблюдается условие

$$\chi_p^2 < \chi_{кр(табл)}^2 (\alpha = 0,05; \nu = n-1). \quad (285)$$

Табличное (критическое) значение критерия χ^2 определяется по таблицам значений критерия Пирсона в зависимости от уровня значимости и числа степеней свободы (табл. 73).

Таблица 73

Таблица критических точек распределения Пирсона [22]

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	2	3	4	5	6	7
1	6,63490	5,02389	3,84146	0,00393	0,00098	0,00016
2	9,21034	7,37776	5,99146	0,10259	0,05064	0,02010
3	11,34487	9,34840	7,81473	0,35185	0,21580	0,11483
4	13,2767	11,14329	9,48773	0,71072	0,48442	0,29711
5	15,08627	12,8325	11,0705	1,14548	0,83121	0,55430
6	16,81189	14,44938	12,59159	1,63538	1,23734	0,87209
7	18,47531	16,01276	14,06714	2,16735	1,68987	1,23904
8	20,09024	17,53455	15,50731	2,73264	2,17973	1,64650
9	21,66599	19,02277	16,91898	3,32511	2,70039	2,08790
10	23,20925	20,48318	18,30704	3,94030	3,24697	2,55821
11	24,72497	21,92005	19,67514	4,57481	3,81575	3,05348
12	26,21697	23,33666	21,02607	5,22603	4,40379	3,57057
13	27,68825	24,7356	22,36203	5,89186	5,00875	4,10692
14	29,14124	26,11895	23,68479	6,57063	5,62873	4,66043
15	30,57791	27,48839	24,99579	7,26094	6,26214	5,22935
16	31,99993	28,84535	26,29623	7,96165	6,90766	5,81221

Продолжение таблицы 73						
1	2	3	4	5	6	7
17	33,40866	30,19101	27,58711	8,67176	7,56419	6,40776
18	34,80531	31,52638	28,86930	9,39046	8,23075	7,01491
19	36,19087	32,85233	30,14353	10,11701	8,90652	7,63273
20	37,56623	34,16961	31,41043	10,85081	9,59078	8,26040
21	38,93217	35,47888	32,67057	11,59131	10,2829	8,89720
22	40,28936	36,78071	33,92444	12,33801	10,98232	9,54249
23	41,63840	38,07563	35,17246	13,09051	11,68855	10,19572
24	42,97982	39,36408	36,41503	13,84843	12,40115	10,85636
25	44,31410	40,64647	37,65248	14,61141	13,11972	11,52398
26	45,64168	41,92317	38,88514	15,37916	13,84391	12,19815
27	46,96294	43,19451	40,11327	16,15140	14,57338	12,87850
28	48,27824	44,46079	41,33714	16,92788	15,30786	13,56471
29	49,58788	45,72229	42,55697	17,70837	16,04707	14,25645
30	50,89218	46,97924	43,77297	18,49266	16,79077	14,95346
31	52,19139	48,23189	44,98534	19,28057	17,53874	15,65546
32	53,48577	49,48044	46,19426	20,07191	18,29076	16,36222
33	54,77554	50,72508	47,39988	20,86653	19,04666	17,07351
34	56,06091	51,96600	48,60237	21,66428	19,80625	17,78915
35	57,34207	53,20335	49,80185	22,46502	20,56938	18,50893
36	58,61921	54,43729	50,99846	23,26861	21,33588	19,23268
37	59,89250	55,66797	52,19232	24,07494	22,10563	19,96023
38	61,16209	56,89552	53,38354	24,8839	22,87848	20,69144
39	62,42812	58,12006	54,57223	25,69539	23,65432	21,42616
40	63,69074	59,34171	55,75848	26,5093	24,43304	22,16426
41	64,95007	60,56057	56,94239	27,32555	25,21452	22,90561
42	66,20624	61,77676	58,12404	28,14405	25,99866	23,65009
43	67,45935	62,99036	59,30351	28,96472	26,78537	24,39760
44	68,70951	64,20146	60,48089	29,78748	27,57457	25,14803
45	69,95683	65,41016	61,65623	30,61226	28,36615	25,90127
46	71,20140	66,61653	62,82962	31,43900	29,16005	26,65724
47	72,44331	67,82065	64,00111	32,26762	29,95620	27,41585
48	73,68264	69,02259	65,17077	33,09808	30,75451	28,17701
49	74,91947	70,22241	66,33865	33,93031	31,55492	28,94065
50	76,15389	71,42020	67,50481	34,76425	32,35736	29,70668

10.9 Прогнозирование на основе статистического моделирования взаимосвязей и зависимостей

Методы прогнозирования используются для предвидения изменений и последствий влияния внешней и внутренней среды на изменение социально-экономических явлений и процессов. Методы прогнозирования подразделяются на количественные и качественные.

К качественным методам прогнозирования относятся в основном методы предвидения спроса, такие как мнение потребителей, мнение покупателей, мнение опытных менеджеров, рыночные тесты. С помощью этих методов определяют, как изменится объем и структура продаж в зависимости от цены товара, местонахождения и уровня доходов клиентов и других факторов.

К количественным методам прогнозирования относится наряду с прогнозированием на основе анализа рядов динамики (рассмотрено в параграфе 8.8) и прогнозирование на основе корреляционно-регрессионного анализа.

В статистике широко используются казуальные методы прогнозирования (корреляционно-регрессионный анализ; метод ведущих индикаторов; метод обследования намерений потребителей и др.). Казуальные методы прогнозирования включают разработку и использование прогнозных моделей, в которых изменения в уровне изучаемого явления являются результатом изменения одной и более переменных.

Казуальные методы прогнозирования требуют определения факторных признаков, оценки их изменений и установления зависимости между их значениями и значением результативного признака. Прогнозирование на основе корреляционно-регрессионного анализа является одним из основных и наиболее точных из группы казуальных методов прогнозирования. Корреляционно-регрессионный метод позволяет учитывать в прогнозных расчетах влияние различных факторов.

При корреляционно-регрессионном анализе строится регрессионная модель, в которой в качестве факторных признаков могут быть выбраны

переменные, оказывающие влияние на результативный признак.

Множественная регрессионная модель в общем виде имеет вид

$$Y (X_1; X_2; \dots; X_n) = a_0 + a_1 \times X_1 + a_2 \times X_2 + \dots + a_n \times X_n, \quad (286)$$

где Y - прогнозируемый (результативный) показатель;

$X_1; X_2; \dots; X_n$ - факторы (независимые переменные);

n - количество независимых переменных;

a_0 - свободный член уравнения регрессии;

$a_1; a_2; \dots; a_n$ - коэффициенты регрессии, измеряющие отклонение результативного признака от его средней величины при отклонении факторного признака на единицу его измерения.

При прогнозировании на основе регрессионной модели выполняется следующая последовательность этапов:

- 1) предварительный отбор независимых факторов, которые по убеждению исследователя оказывают влияние на результативный признак. Эти факторы должны быть либо известны, либо легко определяемы;
- 2) сбор данных по независимым переменным по изучаемой статистической совокупности, при этом, необходимо, чтобы каждая независимая переменная была представлена не менее чем 20 наблюдениями;
- 3) определение вида, формы и направления связи между каждой независимой переменной и результативным признаком, при этом связь должна быть линейной, если связь нелинейная, то необходимо произвести линеаризацию регрессионной модели путем замены или преобразования величины факторного признака;
- 4) построение регрессионной модели - расчет коэффициентов регрессии, и проверка их значимости;
- 5) если построенная регрессионная модель не является статистически значимой, то производится повторение выше указанных этапов (с 1 по 4) до тех пор, пока не будет получена статистически значимая модель;

- б) сравнение степени влияния различных факторов на результативный показатель при помощи расчета коэффициентов эластичности и бетта-коэффициентов.
- 7) прогнозирование на основе корреляционно-регрессионного анализа - определение вероятного значения факторного признака при изменении одного или нескольких результативных признаков, включенных в модель, при условии достаточной значимости и достоверности коэффициентов регрессии и корреляции;
- 8) определение значения точечного прогноза на основании регрессионной модели;
- 9) определение интервального прогноза при помощи расчета средней и предельной ошибок.

Пример прогнозирования на основе корреляционно-регрессионного анализа при парной зависимости. Прогнозирование на основе корреляционно-регрессионного анализа производится на основании построенной ранее регрессионной модели зависимости урожайности овощей от качества почв

$$Y(X) = -3,20828 + 0,315909 \times X$$

Прогноз можно выполнить в трех вариантах: 1) оптимистический – при наивысшем качестве почв по данной совокупности; 2) средний – при среднем качестве почв по данной совокупности; 3) пессимистический – при наименьшем качестве почв по данной совокупности.

Оптимистический прогноз урожайности овощей на основе парной регрессионной модели строится при условии, что качество почв принимается на самом высоком уровне по данной совокупности 95 баллов. На основании регрессионной модели возможная урожайность овощей, которую могут достигнуть предприятия данной совокупности при доведении качества почв до наиболее высокого по совокупности уровня, составит

$$Y(X) = -3,20828 + 0,315909 \times 95 = 26,80 \text{ т/га.}$$

Средний прогноз на основе парной регрессионной модели строится при условии, что качество почв принимается на среднем уровне по данной

совокупности 69,35 баллов. На основании регрессионной модели возможная урожайность овощей, которую могут достигнуть предприятия данной совокупности при доведении качества почв до среднего по совокупности уровня, составит

$$Y(X) = -3,20828 + 0,315909 \times 69,35 = 18,70 \text{ т/га.}$$

Пессимистический прогноз на основе парной регрессионной модели строится при условии, что качество почв принимается на наименьшем уровне по данной совокупности 45 баллов. На основании регрессионной модели возможная урожайность овощей, которую могут достигнуть предприятия данной совокупности при наименьшем по совокупности качестве почв, составит

$$Y(X) = -3,20828 + 0,315909 \times 45 = 11,01 \text{ т/га.}$$

Прогноз на основании парной регрессионной модели показал, что при условии наименьшего качества почв, урожайность овощей в предприятиях данной совокупности может составить 11,01 т/га; при условии среднего качества почв, предприятия данной совокупности могут иметь урожайность овощей 18,70 т/га; а при условии доведения качества почв до наивысшего уровня, предприятия данной совокупности могут достигнуть урожайности овощей 26,80 т/га.

С целью достижения большей достоверности прогноза, его точечное значение дополняется интервальной оценкой. Для определения интервальной оценки рассчитываются средняя и предельная ошибки прогноза по формулам 121 и 122. Интервальная оценка дается с вероятностью 0,954.

$$\text{Средняя ошибка прогноза: } \mu = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{6,03^2}{20}} = 1,35 \text{ т/га.}$$

Для определения предельной ошибки определяется коэффициент кратности ошибки по таблицам распределения Стьюдента (табл.71). $t_{кр(табл)}(\alpha; \nu = n - k - 1)$. При вероятности прогноза 0,95 ($\alpha = 0,05$) и числе степеней свободы $\nu = n - k - 1 = 20 - 1 - 1 = 18$, коэффициент кратности ошибки составит $t_{кр(табл)} = 2,1009$.

Предельная ошибка прогноза: $\Delta = 1,35 \times 2,1009 = 2,8362$ т/га.

Пределы интервального прогноза урожайности овощей при наивысшем качестве почв (оптимистический прогноз):

- нижний предел $Y(X) = 26,8 - 2,8362 = 23,9638$ т/га,
- верхний предел $Y(X) = 26,8 + 2,8362 = 29,6362$ т/га.

Пределы интервального прогноза урожайности овощей при среднем качестве почв (средний прогноз):

- нижний предел $Y(X) = 18,7 - 2,8362 = 15,8638$ т/га,
- верхний предел $Y(X) = 18,7 + 2,8362 = 21,5362$ т/га.

Пределы интервального прогноза урожайности овощей при низком качестве почв (пессимистический прогноз):

- нижний предел $Y(X) = 11,01 - 2,8362 = 8,1738$ т/га,
- верхний предел $Y(X) = 11,01 + 2,8362 = 13,8462$ т/га.

Интервальный прогноз показал, что с вероятностью 0,954 можно утверждать, что предприятия данной совокупности при наивысшем качестве почв могут достигнуть урожайности овощей от 23,9638 т/га до 29,6362 т/га; при среднем качестве почв предприятия данной совокупности могут достигнуть урожайности овощей от 15,8638 т/га до 21,5362 т/га; а при низком качестве почв урожайность овощей в предприятиях данной совокупности составит от 8,1738 т/га до 13,8462 т/га. Полученные прогнозные значения урожайности овощей по оптимистическому и среднему прогнозу в основном выше, чем средние уровни урожайности по совокупности, что свидетельствует о том, что предприятия имеют внутренние резервы повышения урожайности овощей за счет улучшения качества почв путем применения эффективных агротехнических мероприятий.

Прогнозирование на основании корреляционно-регрессионного анализа производится также на основании множественной регрессионной модели зависимости урожайности овощей от качества почв и дозы внесения удобрений $Y(X) = 7,846596 + 0,038209 \times X_1 + 19,93594 \times X_2$.

Прогноз урожайности овощей на основании множественной

регрессионной модели строится в трех вариантах: 1) оптимистический прогноз - при условии, что качество почв и доза внесения удобрений принимается на наивысшем в изучаемой совокупности уровне; 2) средний прогноз - при условии, что качество почв и доза внесения удобрений принимается на среднем по изучаемой совокупности уровне; 3) пессимистический прогноз - при условии, что качество почв и доза внесения удобрений принимается на низшем по изучаемой совокупности уровне.

Оптимистический прогноз урожайности овощей на основе множественной регрессионной модели строится при условии, что качество почв и доза внесения удобрений принимается на самом высоком уровне по данной совокупности 95 баллов и 0,90 т/га. На основании регрессионной модели возможная урожайность овощей, которую могут достигнуть предприятия данной совокупности при доведении качества почв и дозы внесения удобрений до наиболее высокого по совокупности уровня, составит

$$Y(X) = 7,846596 + 0,038209 \times 95 + 19,93594 \times 0,9 = 29,42 \text{ т/га.}$$

Средний прогноз на основе множественной регрессионной модели строится при условии, что качество почв и доза внесения удобрений принимается на среднем уровне по данной совокупности 69,35 баллов и 0,4115 т/га. На основании регрессионной модели возможная урожайность овощей, которую могут достигнуть предприятия данной совокупности при доведении качества почв и дозы внесения удобрений до среднего по совокупности уровня, составит

$$Y(X) = 7,846596 + 0,038209 \times 69,35 + 19,93594 \times 0,4115 = 18,70 \text{ т/га.}$$

Пессимистический прогноз на основе множественной регрессионной модели строится при условии, что качество почв и доза внесения удобрений принимается на наименьшем уровне по данной совокупности 45 баллов и 0,10 т/га. На основании регрессионной модели возможная урожайность овощей, которую могут достигнуть предприятия данной совокупности при наименьшем по совокупности качестве почв и дозе внесения удобрений, составит

$$Y(X) = 7,846596 + 0,038209 \times 45 + 19,93594 \times 0,1 = 11,56 \text{ т/га.}$$

Прогноз показал, что при условии доведения качества почв и дозы внесения удобрений до максимального по совокупности уровня, урожайность овощей может достигнуть 29,42 т/га; при условии доведения качества почв и дозы внесения удобрений до среднего уровня по изучаемой совокупности можно достигнуть урожайности овощей 18,7 т/га; а при условии качества почв и дозы внесения удобрений на наименьшем уровне по изучаемой совокупности урожайность овощей может составить 11,56 т/га.

С целью достижения большей достоверности прогноза, его точечное значение дополняется интервальной оценкой. Для определения интервальной оценки рассчитываются средняя и предельная ошибки прогноза по формулам 121 и 122. Интервальная оценка дается с вероятностью 0,954.

$$\text{Средняя ошибка прогноза: } \mu = \sqrt{\frac{6,03^2}{20}} = 1,35.$$

Для определения предельной ошибки определяется коэффициент кратности ошибки по таблицам распределения Стьюдента (табл.71). $t_{кр(табл)}(\alpha; \nu = n - k - 1)$. При вероятности прогноза 0,95 ($\alpha = 0,05$) и числе степеней свободы $\nu = n - k - 1 = 20 - 2 - 1 = 17$, коэффициент кратности ошибки составит $t_{кр(табл)} = 2,1098$.

$$\text{Предельная ошибка прогноза: } \Delta = 1,35 \times 2,1098 = 2,8482.$$

Пределы интервального прогноза урожайности овощей по наивысшему уровню качества почв и дозы внесения удобрений (оптимистический прогноз):

- нижний предел $Y(X) = 29,42 - 2,8482 = 26,5718$ т/га,
- верхний предел $Y(X) = 29,42 + 2,8482 = 32,2682$ т/га.

Пределы интервального прогноза урожайности овощей по среднему уровню качества почв и дозы внесения удобрений (средний прогноз):

- нижний предел $Y(X) = 18,7 - 2,8482 = 15,8518$ т/га,
- верхний предел $Y(X) = 18,7 + 2,8482 = 21,5482$ т/га.

Пределы интервального прогноза урожайности овощей по наименьшему уровню качества почв и дозы внесения удобрений (пессимистический прогноз):

- нижний предел $Y(X) = 11,56 - 2,8482 = 8,7118$ т/га,

- верхний предел $Y(X) = 11,56 + 2,8482 = 14,4082$ т/га.

Интервальный прогноз показал, что при достижении предприятиями изучаемой совокупности, которые находятся в приблизительно равных производственных условиях, наибольшего уровня качества почв и дозы внесения удобрений урожайность овощей может достигнуть от 26,5718 т/га до 32,2682 т/га; при условии сохранения уровня качества почв и дозы внесения удобрений на среднем уровне по изучаемой совокупности предприятия могут достигнуть урожайности овощей от 15,8518 т/га до 21,5482 т/га; а при наименьшем уровне качества почв и дозы внесения удобрений в предприятиях данной совокупности, урожайность овощей может составить от 8,7118 т/га до 14,4082 т/га.

11 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Статистика как наука. Основные понятия и категории статистической науки.
2. Предмет статистики, его особенности.
3. Задачи статистики.
4. Метод статистики, его основные особенности. Теоретические и методологические основы статистики.
5. Различие и общность статистических дисциплин.
6. Связь статистики с другими науками, ее место в системе наук.
7. Современная организация статистики в России.
8. Понятие о статистическом наблюдении, задачи и значение наблюдения в статистическом исследовании.
9. Цель, объект и единица статистического наблюдения.
10. Программа статистического наблюдения, ее содержание и правила построения.
11. Организационный план, место, время, сроки и критический момент статистического наблюдения.
12. Виды статистического наблюдения.
13. Основные организационные формы статистического наблюдения.
14. Способы проведения статистического наблюдения.
15. Ошибки статистического наблюдения, способы их устранения.
16. Сущность, значение, виды и свойства обобщающего статистического показателя.
17. Сущность, значение, виды и единицы измерения абсолютных показателей.
18. Понятие об относительных показателях, значение и способы их выражения.
19. Виды относительных показателей.
20. Статистическая средняя величина как статистический показатель, ее сущность и значение.
21. Свойства статистической средней величины.

22. Степенные средние, их основные виды и формы, область их применения в статистических исследованиях.
23. Структурные средние, их виды и область применения в статистических исследованиях.
24. Вариация признаков. Абсолютные и относительные показатели вариации.
25. Дисперсионный анализ на основе вариации сгруппированных данных.
26. Понятие о статистической сводке, ее виды, задачи и основное содержание.
27. Статистические ряды распределения, их виды, правила построения и основные характеристики.
28. Понятие о статистической группировке. Сущность метода статистических группировок, правила построения статистических группировок.
29. Виды статистических группировок, их значение в статистических исследованиях.
30. Классификации, применяемые в статистике.
31. Статистические таблицы как способ отражения статистических данных, их виды и правила построения.
32. Статистические графики как способ отражения статистических данных, их составные элементы.
33. Виды графических изображений в статистике, способы их построения и условия применения.
34. Статистические ряды динамики, их основные составные элементы и задачи анализа.
35. Правила построения статистических рядов динамики.
36. Индивидуальные показатели рядов динамики, методика их расчета.
37. Средние показатели рядов динамики, методика их расчета.
38. Основные приемы выявления тенденции развития социально-экономических явлений.
39. Измерение колеблемости уровней динамического ряда и их устойчивости.
40. Изучение сезонных колебаний на основе анализа рядов динамики.

41. Понятие о статистических моделях, модели тенденции и динамические модели.
42. Понятие прогнозирования. Прогнозирование на основе анализа рядов динамики и трендовых моделей.
43. Ограничения и ошибки прогноза. Интервальное значение прогноза.
44. Индекс как статистический показатель, его сущность и значение в статистическом исследовании.
45. Виды индексов. Основные элементы общего (сводного) индекса.
46. Формы индексов, применяемые в статистике.
47. Система взаимосвязанных индексов, применяемая для характеристики динамики сложного явления.
48. Содержание и порядок построения основных экономических индексов.
49. Сущность выборочного метода, необходимость его применения в исследовании массовых общественных явлений.
50. Способы и методы отбора, обеспечивающие репрезентативность выборки.
51. Основные характеристики генеральной и выборочной совокупности.
52. Статистические оценки. Ошибки выборки, методы их расчета.
53. Точечная и интервальная оценки параметров генеральной совокупности при больших и малых выборках.
54. Виды связей и методы их анализа в статистике.
55. Понятие о корреляционной связи, ее виды.
56. Основные этапы корреляционно-регрессионного анализа.
57. Определение параметров уравнения регрессии, их интерпретация.
58. Стандартизированные коэффициенты регрессии, их интерпретация.
59. Показатели тесноты связи, их интерпретация.
60. Корреляционный анализ в рядах динамики. Понятие об автокорреляции.
61. Корреляционный анализ качественных признаков.
62. Непараметрические коэффициенты корреляции.
63. Статистическая оценка значимости коэффициентов регрессии и корреляции.
64. Понятие о статистических моделях, факторные и пространственные модели.

65. Понятие прогнозирования. Прогнозирование на основе факторных моделей.
66. Интервальное значение прогноза. Ограничения при построении прогнозных факторных моделей.

ЗАКЛЮЧЕНИЕ

В современном обществе статистика является одним из важнейших инструментов управления экономикой как на макроуровне, так и на микроуровне. Статистика призвана обеспечить сбор, обработку и представление важной цифровой информации об уровне и возможностях развития государства, отдельных его регионов, секторов, отраслей экономики и отдельных предприятий.

Статистические данные являются одним из определяющих ориентиров социально-экономической политики, способствуют выработке объективного и научно обоснованного курса социально-экономического развития на всех уровнях. В этих условиях особое место отводится такой отрасли статистической науки, как общая теория статистики, которая обеспечивает теоретическую и методологическую подготовку экономистов высшей квалификации, финансистов, коммерсантов, бухгалтеров и аудиторов и др.

В результате освоения дисциплины «Статистика» обучающийся должен демонстрировать следующие результаты образования:

Знать:

- место и роль статистики в изучении, прогнозировании и регулировании социально-экономических явлений и процессов на уровне предприятий (организаций), регионов, отраслей и секторов экономики, экономики и общества в целом;
- нормативно-правовую базу, регулирующую организацию государственной и ведомственной статистики в России, а также международный опыт организации государственной и ведомственной статистики.
- основные понятия и категории статистической науки;
- основные виды и методы статистического наблюдения, классификации и группировки, применяемые в современной российской и международной статистике; основные классификации, применяющиеся для проведения всех видов статистических работ;

- основы построения, расчета и анализа современной системы показателей, основные виды и типы статистических показателей, которые применяются в анализе состояния, динамики и эффективности хозяйствующих субъектов, отраслей и секторов экономики и национальной экономики в целом;
- статистические методы исследования и моделирования связей между изучаемыми явлениями и процессами;
- статистическую методологию анализа динамики экономических явлений и процессов;
- основные формы (универсальные и специализированные) статистической отчетности хозяйствующих субъектов российской экономики и их содержание.

Уметь:

- анализировать во взаимосвязи экономические явления, процессы и институты на микро- и макроуровне;
- рассчитывать на основе типовых методик и действующей нормативно-правовой базы экономические и социально-экономические показатели;
- использовать источники экономической, социальной, управленческой информации;
- анализировать и интерпретировать финансовую, бухгалтерскую и иную информацию, содержащуюся в отчетности предприятий различных форм собственности, организаций, ведомств и т.д. и использовать полученные сведения для принятия управленческих решений;
- анализировать и интерпретировать данные отечественной и зарубежной статистики о социально-экономических процессах и явлениях, выявлять тенденции изменения социально-экономических показателей;
- осуществлять поиск информации по полученному заданию, сбор, анализ данных, необходимых для решения поставленных экономических задач;
- осуществлять выбор инструментальных средств для обработки экономических данных в соответствии с поставленной задачей, анализировать результаты расчетов и обосновывать полученные выводы;

- осуществлять поиск, сбор, обработку и анализ информации, необходимой для решения экономических задач; организовать и провести статистическое наблюдение, систематизировать массивы статистических данных и представлять статистические данные в табличной и графической форме;
- работать с информацией в глобальных компьютерных сетях; использовать для сбора и обработки статистической информации, а также решения задач статистического исследования современные технические средства и информационные технологии;
- формировать систему показателей, исчислять и интерпретировать статистические показатели с учетом их содержания;
- выявлять и оценивать причинно-следственные связи; четко и экономически грамотно формулировать аналитические выводы;
- строить статистические модели, анализировать и содержательно интерпретировать полученные результаты;
- прогнозировать на основе статистических моделей поведение экономических агентов, развитие экономических процессов и явлений, на микро- и макроуровне.

Владеть:

- методологией экономического исследования;
- современными методами сбора, обработки и анализа экономических и социальных данных;
- современными методиками расчета и анализа социально-экономических показателей, характеризующих экономические процессы и явления на микро- и макроуровне;
- понятийно-терминологическим языком статистики;
- компьютерными средствами поиска, обработки, использования и управления информацией;
- статистическими методами комплексной оценки деятельности предприятия (организации) и его внешней среды;

- статистическими методами построения прогнозов развития социально-экономических явлений на различных уровнях экономической деятельности.

Следовательно, в результате изучения курса теории статистики студенты должны овладеть теоретическими знаниями общих основ статистической науки и общими практическими навыками проведения статистического исследования и анализа его результатов. Студенты должны получить знания о научных принципах организации статистических служб РФ; принципах и методах проведения сбора статистической информации и обработки результатов статистического наблюдения; о сущности, видах и правилах построения обобщающих статистических показателей; об основах анализа динамики изучаемых явлений и взаимосвязей между изучаемыми явлениями; об основах анализа статистических данных и статистического моделирования и прогнозирования. Студенты должны научиться организовать и провести статистическое наблюдение; представлять результаты в форме построения статистических таблиц и графиков; исчислять различные обобщающие статистические показатели; анализировать статистические данные и формулировать обоснованные выводы на основе проведенного анализа.

СПИСОК ИСПОЛЬЗОВАННОЙ И РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Федеральный закон от 29.12.2012г. N 273-ФЗ «Об образовании в Российской Федерации» (с изм. и доп. от: 7 мая, 7 июня, 2, 23 июля, 25 ноября 2013 г., 3 февраля, 5 мая 2014 г.) [Электронный ресурс] <http://www.consultant.ru/>
2. Федеральный Закон «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» от 16 ноября 2007 года (в ред. Федеральных законов от 19.10.2011 N 285-ФЗ, от 16.10.2012 N 171-ФЗ, от 02.07.2013 N 171-ФЗ, от 23.07.2013 N 251-ФЗ) [Электронный ресурс] <http://www.consultant.ru/>
3. Федеральная целевая программа «Развитие государственной статистики России в 2007-2011 годах» утв. постановлением Правительства Российской Федерации от 2 октября 2006 года N 595 (с изменениями на 21 марта 2012 года) [Электронный ресурс] www.gks.ru
4. Концепция долгосрочного социально-экономического развития Российской Федерации на период до 2020 года» (утв. распоряжением Правительства РФ от 17 ноября 2008 г. N 1662-р) (в ред. распоряжения Правительства РФ от 08.08.2009 N 1121-р) [Электронный ресурс] base.garant.ru
5. Постановление Правительства РФ «О федеральной службе государственной статистики» от 2 июня 2008г. № 420 ((в ред. Постановлений Правительства РФ от 07.11.2008 N 814, от 27.01.2009 N 43, от 15.06.2010 N 438, от 28.01.2011 N 39, от 24.03.2011 N 210, от 26.01.2012 N 17, от 19.06.2012 N 601, от 15.04.2013 N 337, от 24.09.2013 N 843, от 02.11.2013 N 988) [Электронный ресурс] <http://www.consultant.ru/>
6. Указ Президента РФ от 12 мая 2008г. №724 «Вопросы системы и структуры федеральных органов исполнительной власти» (в ред. Указов Президента РФ от 30.05.2008 N 863, от 24.07.2008 N 1114, от 06.09.2008 N 1315, от 07.10.2008 N 1445, от 14.10.2008 N 1476, от 03.12.2008 N 1715, от 25.12.2008 N 1847, от 31.12.2008 N 1883, от 11.09.2009 N 1033, от 05.10.2009 N 1107, от 19.01.2010 N 82, от 04.03.2010 N 271, от 14.05.2010 N 589, от 23.06.2010 N 780, от 25.08.2010 N

1060, от 27.08.2010 N 1074, от 24.01.2011 N 86, от 08.02.2011 N 155, от 04.03.2011 N 270, от 24.05.2011 N 673, от 17.01.2012 N 79, от 21.05.2012 N 636, от 30.09.2013 N 742) [Электронный ресурс] <http://www.consultant.ru/>

7. Положение «Об условиях предоставления в обязательном порядке первичных статистических данных и административных данных субъектам официального статистического учета» утв. Постановлением Правительства РФ от 18.08.2008 № 620. [Электронный ресурс] <http://www.consultant.ru/>

8. Положение «О порядке представления статистической информации, необходимой для проведения государственных статистических наблюдений», утвержденным постановлением Госкомстата России от 15.07.2002г. № 154 Порядок представления форм статистической отчетности [Электронный ресурс] <http://www.gosfinansy.ru/>

9. Постановление Правительства РФ от 16.02.2008 № 79 «О порядке проведения выборочных статистических наблюдений за деятельностью субъектов малого и среднего предпринимательства» [Электронный ресурс] <http://www.assessor.ru/>

10. Приказ от 29.08.2012г. №470 «Об утверждении статистического инструментария для организации федерального статистического наблюдения за деятельностью предприятий» [Электронный ресурс] <http://www.assessor.ru/>

11. Елисеева, И.И. Статистика: [текст] учеб. / И.И.Елисеева, А.В.Изотов, Е.Б.Капралова [и др.]; под.ред. И.И.Елисеевой. – М.: КНОРУС, 2006. – 552с. 12.

12. Зинченко, А.П. Статистика. [текст] – М.: КолосС, 2007. – 568 с.

13. Минашкин В.Г. Статистика [текст]: учебник для бакалавров. 448 с. – 2014 г.

14. Отчётность в Росстат [Электронный ресурс] <http://nicolbuh.ru/>

15. Официальный сайт Федеральной службы государственной статистики Российской Федерации www.gks.ru

16. Официальный сайт Территориального органа Федеральной службы государственной статистики Российской Федерации по Амурской области www.amurstat.ru

17. Практикум по статистике: Учеб.пособие / А.П. Зинченко, А.Е.Шибалкин, О.Б. Тарасова, Е.В. Шайкина. – М.: колос, 2003. – 392 с.
18. Пастушенко С.Б. Статистика [текст]: учеб. пособие; рек. ДВ РУМЦ / С.Б. Пастушенко, издательство ДальГАУ, 232 с.
19. Статистика [текст]: учебник под ред. И.И. Елисеевой. – М.6 Высшее образование, 652 с. – 2014 г.
20. Статистика: Учебное пособие [текст] / Под. Ред. проф. М.Р. Ефимовой. – М.: ИНФРА-М. 2006. – 336с. – (Вопрос-ответ).
21. Статистика. Отчетность 2012, 2013 бланки, форма в Госкомстат [Электронный ресурс] <http://www.assessor.ru/>
22. Теория статистики: Учебник [текст] / Под. Ред. Р.А. Шмойловой. – 2-е изд., доп. и перераб. – М.: Финансы и статистика, 1998. – 576с.

СОДЕРЖАНИЕ

Введение	3
1 ПРЕДМЕТ, МЕТОД И ЗАДАЧИ СТАТИСТИКИ	7
1.1 Значение и основное содержание статистики	7
1.2 Объект изучения, теоретические основы и предмет статистики	9
1.3 Основные категории статистической науки	10
1.4 Методологическая основа и метод статистики	13
2.4 Методологическая основа и метод статистики	13
1.5 Общность и различие статистических дисциплин. Связь статистики с другими науками	16
1.6 Задачи статистики	20
1.7 Организация статистики в Российской Федерации	23
2 СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ МАССОВЫХ ОБЩЕСТВЕННЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ	29
2.1 Сущность, значение и задачи статистического наблюдения	29
2.2 Этапы проведения статистического наблюдения	30
2.3 Основные организационные формы статистического наблюдения	32
2.4 Виды статистического наблюдения	36
2.5 Способы проведения статистического наблюдения	39
2.6 Методологические аспекты проведения статистического наблюдения	40
2.7 Ошибки статистического наблюдения и меры борьбы с ними	43
3 ОБОБЩАЮЩИЕ СТАТИСТИЧЕСКИЕ ПОКАЗАТЕЛИ. АБСОЛЮТНЫЕ И ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ	46
3.1 Обобщающие статистические показатели, их свойства, виды и формы	46
3.2 Абсолютные величины, их виды и единицы измерения	48
3.3 Относительные величины, их виды и единицы измерения	50
4 СРЕДНИЕ ВЕЛИЧИНЫ И ПОКАЗАТЕЛИ ВАРИАЦИИ	58
4.1 Сущность, значение и свойства средней величины	58
4.2 Степенные средние величины	60
4.3 Структурные средние величины	64

4.4 Понятие вариации признаков. Показатели вариации	67
4.5 Вариация сгруппированных данных. Виды дисперсий, правило их сложения	79
5 СВОДКА И ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ	84
5.1 Сущность, виды и правила построения статистической сводки	84
5.2 Сущность и виды статистической группировки. Статистические классификации	86
5.3 Статистические ряды распределения, их виды и основные характеристики	90
5.4 Порядок построения статистической группировки по количественному признаку	98
6 СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ И ГРАФИКИ КАК ОСНОВНОЙ СПОСОБ ПРЕДСТАВЛЕНИЯ РЕЗУЛЬТАТОВ СТАТИСТИЧЕСКИХ РАБОТ	108
6.1 Статистические таблицы как способ представления и обобщения статистического материала. Составные элементы статистических таблиц	108
6.2 Виды статистических таблиц	110
6.3 Правила построения статистических таблиц	111
6.4 Чтение и анализ статистических таблиц	113
6.5 Статистические графики как способ представления статистического материала. Значение графического способа в статистике	115
6.6 Основные элементы статистического графика	116
6.7 Виды графических изображений и способы их построения	119
7 СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА РЯДОВ ДИНАМИКИ. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ АНАЛИЗА РЯДОВ ДИНАМИКИ	130
7.1 Сущность, составные элементы, задачи и виды рядов динамики	130
7.2 Правила построения рядов динамики	135
7.3 Индивидуальные показатели ряда динамики	138
7.4 Средние показатели рядов динамики	143

7.5 Способы выявления основной тенденции на основе выравнивания рядов динамики. Статистическое моделирование на основе анализа рядов динамики	148
7.6 Оценка устойчивости выявленной тенденции	160
7.7 Оценка сезонных колебаний	165
7.8 Прогнозирование развития явления на основе анализа рядов динамики и статистического моделирования в рядах динамики	168
8 СТАТИСТИЧЕСКИЕ ИНДЕКСЫ	174
8.1 Сущность и значение индексного метода в статистическом анализе социально-экономических явлений. Виды и свойства индексов	174
8.2 Составные элементы, формы и правила построения основных статистических индексов	177
8.3 Взаимосвязь индексов	180
8.4 Построение основных экономических индексов	183
9 СТАТИСТИЧЕСКИЕ МЕТОДЫ ОРГАНИЗАЦИИ И ОЦЕНКИ ВЫБОРОЧНОГО НАБЛЮДЕНИЯ	198
9.1 Содержание, задачи и значение выборочного метода	198
9.2 Генеральная и выборочная совокупности, их характеристики	201
9.3 Понятие репрезентативности выборки. Способы отбора, обеспечивающие репрезентативность	202
9.4 Ошибки выборки	205
9.5 Статистические оценки характеристик генеральной совокупности	207
9.6 Определение необходимой численности выборочной совокупности	211
9.7 Особенности расчетов при малой выборке	213
9.8 Расчет оценок генеральной совокупности по характеристикам выборочной совокупности	216
10 СТАТИСТИЧЕСКОЕ ИЗУЧЕНИЕ ВЗАИМОСВЯЗЕЙ И ЗАВИСИМОСТЕЙ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ЯВЛЕНИЙ. СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ АНАЛИЗА ВЗАИМОСВЯЗЕЙ И ЗАВИСИМОСТЕЙ	223

10.1 Понятие взаимосвязей и зависимостей в статистике. Задачи статистического изучения взаимосвязей и зависимостей	223
10.2 Методы статистического изучения взаимосвязей и зависимостей	226
10.3 Формы и виды корреляционно-регрессионной зависимости	229
10.4 Определение параметров уравнения регрессии, их интерпретация	231
10.5 Показатели тесноты связи	249
10.6 Корреляция качественных признаков	260
10.7 Корреляция в рядах динамики	266
10.8 Оценка существенности показателей связи	270
10.9 Прогнозирование на основе статистического моделирования взаимосвязей и зависимостей	281
КОНТРОЛЬНЫЕ ВОПРОСЫ	289
ЗАКЛЮЧЕНИЕ	293
СПИСОК ИСПОЛЬЗОВАННОЙ И РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ	297
СОДЕРЖАНИЕ	300